Quantum Chemical Modeling of Enzymatic Reactions: Applications to the Tautomerase Superfamily

Robin Sevastik

Department of Theoretical Chemistry
Royal Institute of Technology
Stockholm, Sweden, 2008
Abstract

In this thesis, quantum chemical methods are used to investigate enzymatic reaction mechanisms. The Density functional theory, in particular the hybrid B3LYP functional, is used to model two enzymes belonging to the tautomerase superfamily: 4-Oxalocrotonate Tautomerase (4-OT) and cis-Chloroacrylic Acid Dehalogenase (cis-CAAD). The methodology is presented and new mechanistic insights for the two enzymes are discussed. For 4-OT, two different models are built and the potential energy curves are computed. This allows the methodology to be evaluated. The results give new insight into the energetics of the 4-OT reaction, indicating that the charge-separated intermediate is quite close in energy to the reactant species. The models also make it possible to perform in silico mutations to investigate the role of active site groups. Excellent agreement is found between the calculations and site-directed mutagenesis experiments, further substantiating the validity of the models.

For cis-CAAD, the uncatalyzed reaction is first considered and excellent agreement is found between the calculated barrier and the measured rate constant. The enzymatic reaction is then studied with a quite large active site model and a reaction mechanism is proposed.
Acknowledgements

First, I would like to thank my supervisor Dr. Fahmi Himo. I really appreciate all your support, encouragement and wise words. Thank you so much for always having time to discuss both work related issues and more philosophical topics. You are a great supervisor. Second, I would like to thank Prof. Hans Ågren for giving me this opportunity and providing a great working environment. I am also very grateful to have had the opportunity to travel to conferences and participate in professional scientific settings.

Big thanks to the entire group at theoretical chemistry. I would especially like to thank Kathrin for proof reading my thesis and thanks to Peter, Polina, Chen, Tomasso, Liao, Sugata and Francesca for a very nice catalysis group.

I would also like to thank my family. Especially my parents and sister, but also all of my grandparents for the wisdom and love you have provided.

Finally, thank you Sophie for always supporting me and being by my side!
List of Papers

I. *Quantum Chemical Modeling of Enzymatic Reactions: The Case of 4-Oxalocrotonate Tautomerase*
   Robin Sevastik and Fahmi Himo

II. *Reaction Mechanism of cis-Chloroacrylic Acid Dehalogenase - A Theoretical Study*
    Robin Sevastik and Fahmi Himo
    *In Manuscript.*
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-OT</td>
<td>4-Oxalocrotonate Tautomerase</td>
</tr>
<tr>
<td>CAAD</td>
<td>Chloroacrylic Acid Dehalogenase</td>
</tr>
<tr>
<td>CHMI</td>
<td>5-(CarboxyMethyl)-2-Hydroxymuconate Isomerase</td>
</tr>
<tr>
<td>CPCM</td>
<td>Conductor-like Polarizable Continuum Model</td>
</tr>
<tr>
<td>DFT</td>
<td>Density Functional Theory</td>
</tr>
<tr>
<td>GGA</td>
<td>Generalized Gradient Approximation</td>
</tr>
<tr>
<td>GTO</td>
<td>Gaussian Type Orbitals</td>
</tr>
<tr>
<td>HF</td>
<td>Hartree-Fock</td>
</tr>
<tr>
<td>LDA</td>
<td>Local Density Approximation</td>
</tr>
<tr>
<td>MM</td>
<td>Molecular Mechanics</td>
</tr>
<tr>
<td>MSAD</td>
<td>Malonate Semialdehyde Decarboxylase</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PPT</td>
<td>Phenylpyruvate Tautomerase</td>
</tr>
<tr>
<td>QM</td>
<td>Quantum Mechanics</td>
</tr>
<tr>
<td>TS</td>
<td>Transition State</td>
</tr>
<tr>
<td>TST</td>
<td>Transition State Theory</td>
</tr>
<tr>
<td>ZPV</td>
<td>Zero Point Vibrations</td>
</tr>
</tbody>
</table>
# Amino Acid Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Symbol</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Ala</td>
<td>alanine</td>
</tr>
<tr>
<td>C</td>
<td>Cys</td>
<td>cysteine</td>
</tr>
<tr>
<td>D</td>
<td>Asp</td>
<td>aspartic acid</td>
</tr>
<tr>
<td>E</td>
<td>Glu</td>
<td>glutamic acid</td>
</tr>
<tr>
<td>F</td>
<td>Phe</td>
<td>phenylalanine</td>
</tr>
<tr>
<td>G</td>
<td>Gly</td>
<td>glycine</td>
</tr>
<tr>
<td>H</td>
<td>His</td>
<td>histidine</td>
</tr>
<tr>
<td>I</td>
<td>Ile</td>
<td>isoleucine</td>
</tr>
<tr>
<td>K</td>
<td>Lys</td>
<td>lysine</td>
</tr>
<tr>
<td>L</td>
<td>Leu</td>
<td>leucine</td>
</tr>
<tr>
<td>M</td>
<td>Met</td>
<td>methionine</td>
</tr>
<tr>
<td>N</td>
<td>Asn</td>
<td>asparagine</td>
</tr>
<tr>
<td>P</td>
<td>Pro</td>
<td>proline</td>
</tr>
<tr>
<td>Q</td>
<td>Gln</td>
<td>glutamine</td>
</tr>
<tr>
<td>R</td>
<td>Arg</td>
<td>arginine</td>
</tr>
<tr>
<td>S</td>
<td>Ser</td>
<td>serine</td>
</tr>
<tr>
<td>T</td>
<td>Thr</td>
<td>threonine</td>
</tr>
<tr>
<td>V</td>
<td>Val</td>
<td>valine</td>
</tr>
<tr>
<td>W</td>
<td>Trp</td>
<td>tryptophan</td>
</tr>
<tr>
<td>Y</td>
<td>Tyr</td>
<td>tyrosine</td>
</tr>
</tbody>
</table>
## Contents

1 Introduction ...................................................................................................... 1

2 Computational Methods .................................................................................. 3
   2.1 Density Functional Theory ............................................................................. 3
   2.2 Accuracy of B3LYP ....................................................................................... 4

3 Enzyme Modeling ............................................................................................. 7
   3.1 Constructing the Model ................................................................................ 7
   3.2 Surrounding Effects .................................................................................... 9
   3.3 Transition State Theory .............................................................................. 9
   3.4 Computational Details ............................................................................... 11

4 Applications to the Tautomerase Superfamily ............................................ 13
   4.1 Modeling of 4-Oxalocrotonate Tautomerase ................................................. 14
   4.2 Modeling of Chloroacrylic Acid Dehalogenase ........................................... 24

5 Conclusions ..................................................................................................... 35

References ........................................................................................................ 36
Chapter 1

Introduction

Enzymes make life on our planet possible. They are present in all living organisms and act as catalysts to speed up and control vital chemical reactions. Failure of critical enzymes can lead to serious diseases. In fact, many drugs act by inhibiting the function of specific target enzymes. Enzymes are in general very specific and selective, targeting one or a few substrates that are transformed to products. Like other catalysts, enzymes are not consumed and are regenerated after a full reaction cycle, allowing one enzyme to treat several thousands of substrates [1]. Enzymes are today also used as synthetic tools in large-scale industrial processes for production of basic and fine chemicals. In many cases, this technology has advantages over standard techniques. Therefore, it is of interest for both fundamental science and also for industrial applications to understand how enzymes catalyze their reactions. This understanding of the catalytic apparatus by which enzymes carry out their reactions can have large implications. For example, it can help in the rational design of new drugs that have improved selectivity and potency properties. It can also work as an inspiration to synthesize biomimetic catalytic complexes for industrial applications.

Computational chemistry is today a very important tool in the study of enzyme mechanisms. In particular, density functional theory (DFT) calculations have in the recent years become a very powerful method for analyzing reaction mechanisms. As the computer power grows, larger and larger systems can be treated at higher and higher accuracy. Today, calculations involving 100 atoms are almost performed on routine. However, enzymes consist of tens of thousands of atoms and the question is how one can model these large systems using active site models consisting of around 100 atoms. The basic idea is to use accurate DFT methods to treat the critical parts of the enzyme, where the reaction takes place and bonds are formed and broken. The rest of the enzyme is treated at a lower level of theory and various simplifying assumptions are made. This methodology has been used in numerous investigations in recent years, many times with very large success in explaining the reaction mechanisms and the factors contributing to the catalysis [2, 3].
In this thesis, the methodology of using DFT calculations to model enzymatic reaction will be applied to two enzymes that belong to the tautomerase superfamily, 4-Oxalocrotonate Tautomerase (4-OT) and cis-Chloroacrylic Acid Dehalogenase (cis-CAAD). In Chapter 2, a brief introduction to the theoretical methods will be given. In Chapter 3, the modeling strategy will be discussed, and in Chapter 4, the specific results for the two enzymes will be presented. Finally, conclusions are given in Chapter 5.
Chapter 2
Computational Methods

Computational chemistry is a very dynamic research field with constantly evolving theories and methods. Computational chemists have access to a number of computational approaches that are capable of solving a wide range of chemical problems. Molecular modeling has become a very practical and vital instrument in many different research areas, as diverse as biochemistry, surface chemistry, drug design, and homogenous catalysis. Like other computational areas, theoretical chemistry has gained from the enormous advancement in computer power. The computers have become faster and cheaper and more challenging problems have been able to be addressed.

The present thesis is concerned with the application of quantum chemical methods to study enzymatic reactions. The great majority of studies in this field use density functional theory (DFT). Therefore, a brief introduction to DFT is in place here.

2.1 Density Functional Theory

Modern density functional theory is based on two fundamental theorems by Hohenberg and Kohn [4]. These state that the electron density $\rho(r)$ of a system uniquely defines its ground state energy, and that there exists a variational principle for the density-dependent functional. In DFT the basic quantity is the electron density, in contrast to the wavefunction in wavefunction-based methods. The total energy can be expressed as a functional of the electron density, $E[\rho]$. The problem of determining the 3N-dimensional wavefunction reduces this way to finding the 3-dimensional electron density. The exact analytical form of the energy functional is, however, unknown.

In the Kohn-Sham formulation of DFT [5], the energy functional is expressed as follows:

$$E_{\text{DFT}}[\rho] = T_S[\rho] + E_{\text{xc}}[\rho] + J[\rho] + E_{\text{ne}}[\rho]$$
where $T_{\phi}[^\rho]$ is the kinetic energy, $E_{ne}[^\rho]$ is the nuclei-electron interaction, $J[^\rho]$ is the classical Coulomb integral, and $E_{xc}[^\rho]$ is the exchange-correlation term, which is the only unknown term in this expression.

Different expressions have been used for $E_{xc}[^\rho]$ over the years. In the Local Density Approximation (LDA) the electron density is treated as a uniform electron gas [6,7]. However, LDA gives poor description for molecules since the electron density is not very uniform. An alternative to the LDA method is the Generalized Gradient Approximation (GGA) method [8], where in addition derivatives of the electron density are taken into account, giving a more realistic description of the density.

A further improvement is adding some of the Hartree-Fock (HF) exchange to the functional together with empirical values, this gives the B3LYP functional [9].

$$E_{xc}^{B3LYP} = (1 - A)E_x^{LDA} + AE_x^{HF} + BE_x^{Becke} + (1 - C)E_x^{VWN} + CE_x^{LYP}$$

where $E_x^{LDA}$ is the exchange energy from the LDA method, $E_x^{HF}$ is the Hartree-Fock exchange energy, $E_x^{Becke}$ is Becke’s exchange functional [10], $E_x^{VWN}$ is the correlation functional introduced by Vosko, Wilk, Nusair [11], and $E_x^{LYP}$ is the correlation functional introduced by Lee, Yang and Parr [12]. The coefficients are $A = 0.20$, $B = 0.72$ and $C = 0.81$, that were determined empirically by a linear least-squares fit to 116 experimentally-determined energies [13].

### 2.2 Accuracy of B3LYP

When performing computational chemistry it is of importance that the accuracy of any chosen method is known. In the case of DFT the accuracy has been tested against a set of experimentally determined results using the G2 and G3 test sets [14, 15]. Structural parameters, such as bond distances, bond angles, and dihedral angles were computed using the B3LYP functional with varying basis sets and compared to experimentally-determined geometries for 53 molecules [14]. Using a medium-sized basis set, 6-31G(d,p), an average error of 0.013 Å was observed for bond distances, an average error of 0.62° was observed for bond angles, and an average error of 0.35° was observed for dihedral angles.
For energies, a comparison to determine the accuracy of obtained atomization energies using the B3LYP functional was performed on a total of 55 molecules using the G2 test set. With the 6-31G(d,p) basis set, an average error of 5.18 kcal/mol was obtained compared to experimental results. With a larger basis set, 6-311+G(3df,2p), the error was reduced to 2.20 kcal/mol. This accuracy is very high and comparable to the most accurate ab initio methods. In particular, the trade-off between accuracy and speed is excellent, making it a useful tool in the study of enzymatic reactions.
Chapter 3
Enzyme Modeling

As discussed above, DFT methods, in particular the B3LYP functional, provide an excellent tool to accurately calculate potential energy surfaces of chemical reactions. The method allows for treatment of more than one hundred atoms quite routinely using the computer power of today. However, enzymes usually consist of many thousands of atoms and the question is how to study the reactivity of enzymes using relatively small quantum chemical cluster models. In this chapter, the methodology and the ideas behind the modeling strategies will be discussed briefly.

3.1 Constructing the Model
The first step when performing enzyme modeling is the construction of a model that will reflect the main features of the active site of the enzyme. The usual procedure is to start from an experimentally-determined x-ray crystal structure of the enzyme. The coordinates are commonly deposited in online libraries, such as the Protein Data Bank (PDB).
Starting from the crystal structure, the active site of the enzyme is located and important amino acid residues are selected to be included in the computational model (Figure 1). Due to the resolutions of protein crystal structures, hydrogen atoms are not visible and have to be added manually to the model. For most groups this is quite simple. However, some cases can be problematic, such as when histidine residues are present in the active site. In such situations considerations based on the p$K_a$ values and the pH of the solution have to be made to assign the protonation state of the group. If no such results are available one might design separate models with different protonation states and test the effects on the barriers and energies of the reactions.
Figure 1. When constructing a computational model the active site of the enzyme is located and important residues are chosen to be part of the model.

The enzyme substrate is not always present in the crystal structure. In some cases inhibitors or products are part of the structure instead. In such cases the substrate can either be superimposed based on the position of the inhibitor or product. In other cases the substrate is built into the model based on knowledge of the active site and optimizations of the substrate keeping all other atoms fixed.

The focus when constructing a model is on the active site where the actual chemistry takes place. Choosing which residues to include is not always straightforward but most often there is experimental support to what residues are specifically important in the reaction mechanism. In other cases educated guesses have to be made by studying the active site visually and considering which residues could interact with the substrate. To keep the size of the model as small as possible truncations have to be made to residues that are included in the model. The residues are truncated so that they are big enough to keep their chemical properties but small enough to keep the size of the model at its minimum.

Keeping the model as small as possible is important not only to cut computational costs but also because of the problem of local minima. When optimizing bigger models there is a higher risk of the optimization converging in local minima that are not actually global minima. This can be very hard to see and the easiest way to avoid this problem is to keep the models as small as possible but still retaining the chemical features that are important for the reaction.

A common procedure when making a model is to construct a small model, conduct calculations, evaluate the results and then go back to the crystal structure and construct a
bigger model. By using this scheme information about the enzyme of interest can be obtained at a low computational cost and a larger model can be made based on the results from a smaller model. Even though small, a model can give useful mechanistic information. Also, when results from two different model sizes are obtained importance of specific residues in the larger model can be seen in a more direct way.

3.2 Surrounding Effects
The active site is only a very small part of the enzyme, it is where the actual reaction takes place and therefore it is natural to construct a model with the active site in focus. The surrounding enzyme does, however, influence the reaction in two main ways, by steric and polarization effects.

The enzyme surrounding keeps the active site groups in certain positions relative to each other and relative to the substrate. In the model, this can be taken into account by locking certain atoms, typically were the truncations are made, to their crystallographic positions. This way, very large and unrealistic movements of the various groups are avoided. However, the procedure can sometimes lead to a too rigid model. As the models become larger and longer chains are kept in the model, this problem becomes less important.

The polarization effects caused by the surrounding are modeled using the conductor-like polarizable continuum model (CPCM) [16]. This theory treats the enzyme surrounding as a continuous homogenous electrostatic medium with a dielectric constant ($\varepsilon$) that can be chosen depending on the situation at hand. Most often $\varepsilon$ is chosen to ($\varepsilon = 4$) which in several cases has been shown to give an accurate solvent effect when modeling enzymes. PCM calculations are performed as single point calculations on already optimized geometries in gas phase (vacuum). Usually the PCM calculations have an effect of a few kcal/mol on the relative energy of the reaction. Very large solvation effects usually indicate that the given DFT model is too small and missing some important groups.

3.3 Transition State Theory
To determine whether a reaction mechanism is feasible or not, one has to calculate the potential energy profile for it. This includes calculating the energies of all transition states and intermediates. The transition states are extremely short-lived and cannot be isolated experimentally. In computational chemistry a transition state can be found by optimizing a
reaction at the highest point on a potential energy curve. Many times, when searching for a transition state a so-called linear transit scheme can be very useful (Figure 2). A stepping procedure is performed starting from the reactant, going past the transition state and finally finishing at the product structure. During the linear transit scan a reaction coordinate is chosen and kept fixed while all other degrees of freedom are minimized. From the highest point of the curve, a full transition state optimization can then be started.

The calculated activation barrier of the reaction can be related to the experimental rate constants by using classical transition state theory [7]:

\[ k = \frac{k_B T}{h} \exp \left( \frac{-\Delta G^*}{RT} \right) \]

\( k \) is the rate constant, \( k_B \) is the Boltzmann’s constant (1.38 * 10^{-23} \text{ J/K}), \( T \) is the temperature in Kelvin, \( h \) is the Planck’s constant (6.626 * 10^{-34} \text{ Js}), and \( R \) is the gas constant (8.314 \text{ J/K}).

This is a very useful relation that allows us to judge whether a reaction mechanism is viable or not. For example, one can from this relation deduce that at room temperature, a barrier of ca 18 kcal/mol corresponds to a rate constant of 1 s^{-1}. Also, an increase of 1.4 kcal/mol of the barrier is equivalent to a decrease of the rate with one order of magnitude.

![Potential energy curve for a typical linear transit scheme.](image)
3.4 Computational Details

Throughout the thesis, the hybrid density functional theory method of B3LYP as implemented in the Gaussian03 program package [17], is used. Geometry optimization are performed using the medium-sized double-\(\zeta\) basis set 6-31G(d,p), which gives quite accurate geometries. To obtain accurate energies, however, the larger basis set 6-311+G(2d,2p) is used. These calculations are performed as single points based on the optimized geometries.

Hessian calculations are then performed to confirm the nature of the stationary point, with no negative eigenvalues for minima, and one negative eigenvalue for transition states. The Hessian calculations also give the zero-point vibrational energies which are added to the final energies. For very large models, Hessian calculations are many times not possible to calculate. In these cases, zero-point vibrational effects can be calculated for smaller models and then transferred to the large models.

Solvation corrections are calculated using the cavity techniques discussed above. The conductor-like polarizable continuum model (CPCM) is used at the same level of theory as the geometry optimizations.
Chapter 4

Applications to the Tautomerase Superfamily

The tautomerase superfamily originally got its name from the functions of the original members, 4-Oxalocrotonate Tautomerase (4-OT), 5-(CarboxyMethyl)-2-Hydroxymuconate Isomerase (CHMI) and Phenylpyruvate Tautomerase (PPT), that all catalyze a keto-enol tautomerization reaction. Most of the enzymes included in the family are bacterial enzymes that are part of the degradation pathway of aromatic carbons that allow bacteria to use aromatic carbons as their sole source of carbon energy [18]. The tautomerase superfamily has recently been expanded to include new members, Malonate Semialdehyde Decarboxylase (MSAD), Chloroacrylic Acid Dehalogenase (CAAD) and cis-Chloroacrylic Acid Dehalogenase (cis-CAAD). These new enzymes differ from the traditional ones in that they do not catalyze keto-enol tautomerizations. CAAD and cis-CAAD catalyze the dehalogenation of the two different isomers of chloroacrylic acid to semialdehyde and Cl⁻ [19]. MSAD catalyzes the decarboxylation of malonate semialdehyde to acetaldehyde and carbon dioxide [20]. CAAD, cis-CAAD and MSAD are involved in the pathway of decomposing poisonous halocarbon compounds to digestible carbon molecules in soil-living bacteria [21].

All of the enzymes belonging to the tautomerase superfamily have a similar structural β-α-β fold and they all have a conserved N-terminal proline (Pro-1) that plays an important catalytic role [22]. Pro-1 has been shown to have a low pKₐ in several of the enzymes, allowing it to act as a base in the keto-enol tautomerizations [23]. Another common denominator of the tautomerase superfamily is that the active sites of the enzymes are predisposed to bind and polarize a carboxylate group [22,23]. Most commonly, a positively charged arginine group is present in the active site and binds the carboxylate group of the substrate.
4.1 Modeling of 4-Oxalocrotonate Tautomerase

4.1.1 Introduction

4-Oxalocrotonate Tautomerase (4-OT) is a bacterial enzyme found in the soil living bacterium *Psedomonas putida mt-2*. It catalyzes the isomerization of unconjugated α-keto acids to their conjugated isomers (Scheme 1). The unconjugated α-keto acids that act as substrates for 4-OT are a product from the ring opening process of aromatic hydrocarbons [18]. Bacteria that use 4-OT in the breakdown process of aromatic hydrocarbon compounds can literally survive by digesting benzene and other refined crude oil products. This opens the possibility of these bacteria being used to clean up oil spills or digest human oil based waste material [24].

Scheme 1. Reaction catalyzed by 4-oxalocrotonate tautomerase.

The structural composition of 4-OT is a homo-hexamer, composed of a trimer of dimers, with 62 amino acids per monomer [25,26]. There are six active sites located at the interfaces between the dimers [27,28]. When the substrate, 2-oxo-hexenedioate, enters the active site it is anchored by three arginine groups, Arg-61’, Arg-11’ that interact with the substrates carboxylate group and Arg-39” that hydrogen binds to the carbonyl oxygen of the substrate [29,30]. A general mechanism for the 4-OT reaction has been accepted (Scheme 2) in which a conserved terminal proline (Pro-1) with a $pK_a$ of 6.4 [31] acts as a general base to abstract a proton from the C3 position of the substrate and deliver it back to the C5 position. The mechanism is thus a simple proton abstraction/delivery process that takes place in two steps.
The mechanism of 4-OT is the result of a great number of studies, both experimental [32-37] and theoretical [38-42]. One specific theoretic study modeled the enzyme with a QM/MM theory (B3LYP/6-31G*//HF/3-21G) and presented a potential energy surface having a barrier of 14.5 kcal/mol for the first step [42]. The intermediate was found to lie in a shallow minimum, 13.5 kcal/mol higher than the reactant and only 2.9 kcal/mol below the second transition state.

### 4.1.2 Small Model

A model was constructed (Model A) to investigate if a DFT method, treating the model with a high level of theory and the surrounding effects with a PCM model, could reproduce the previous available QM/MM results for the 4-OT reaction [42].

The model was built using the crystal structure (PDB code 4OTA [28]) and included the substrate, Pro-1, Arg-39, Arg-61, Arg-39” and two water molecules. In the optimized reactant structure the proline nitrogen was in a very good position to accept a proton from the
substrate, with an H-N distance of 2.36 Å (React, Figure 3). The stationary points were optimized along the reaction pathway (React, TS1, Inter, TS2, Prod) as shown in Figure 3.

**Figure 3.** Optimized stationary points for the small model of 4-OT. Fixed atoms are indicated with arrows.
Figure 4. Potential energy surface for Model A.

The PCM energies for the stationary points, React, TS1, Inter, TS2 and Prod were computed using a number of different dielectric constants (Figure 4). This was done to investigate the influence of the dielectric constant on the energetics. As expected, solvent effects lowered the energies of both TS1 and the intermediate compared to the reactant. Using $\epsilon = 4$, the barrier decreased from 12.8 to 9.2 kcal/mol, and the energy of the intermediate was lowered from 9.8 to 3.4 kcal/mol. The quite large solvent effects were a result of the reaction steps being proton transfers, generating large charge separations within the model. Solvating the transition state and intermediate lowered the energy of the charge-separated complex compared to the reactant.

The obtained potential energy curve is thus quite different from the one obtained using the QM/MM method [42]. To investigate if this discrepancy is due to the size of the quantum chemical model a larger model of the active site was designed.

4.1.3 Large Model

A larger model (Model B) that consisted of 177 atoms was built based on the first model, with several additional residues (Figure 5). The side chain of Phe-50’, the backbone chains of Ile-7’, Leu-8’, Ser-37, and Ile-27 were all added to help orient the proline and the substrate within
the active site and to provide stabilization to the charges that develop during the reaction process.

In addition, the side chains of Arg-39”, Arg-61’, Arg-11’, and Pro-1 were kept longer in the bigger model, to grant more flexibility to these groups. In this model, as in the previous model, certain atoms were kept fixed in order to maintain the structure close to the x-ray structure. These positions are indicated by arrows in Figure 5. The basic features of the core region (substrate, arginines, and proline) were very similar to Model A. As before, the proline nitrogen was in good position to accept the proton from the substrate, with a H-N distance of 2.65Å.

The important structural parameters for the larger model did not differ much from the smaller model. However some parameters were somewhat different. The critical C3-H and H-N distances at TS1 (Figure 5) were 1.32Å and 1.44Å, compared to 1.47Å and 1.29Å for Model A, indicating an earlier transition state in Model B compared to Model A. TS2 however was later in Model B, N-H and H-C5 distances were 1.38Å and 1.34Å, (compared to 1.27Å and 1.48Å, in Model A).

The intermediate in Model B was calculated to be 5.7 kcal/mol, lower than the reactant. The specific solvent stabilization that was added in the model by explicit groups around the substrate and proline led to an additional stabilization of the intermediate by about 7 kcal/mol compared to the homogenous solvent stabilization of Model A (ε =80).
Figure 5. Schematic representation of important optimized bond distances of TS1, Inter, TS2, and Prod of Model B. For clarity, several groups were omitted in TS1, Inter, TS2 and Prod. Fixed atoms are indicated with arrows.

TS1 was calculated to lie at +6.9 kcal/mol compared to the reactant in the cluster model. TS2 was 4.0 kcal/mol higher than the intermediate and the final product was 2.9 kcal/mol lower than the reactant in the cluster model. A very important feature of the potential energy surface
obtained for Model B (Figure 6) was that inclusion of solvent effects, using a homogenous polarizable medium on top of the quantum chemical cluster model, had almost no effect on the energetics of the reaction. This shows that most of the solvation effect is already included explicitly in the quantum model.

Figure 6. Potential energy curves obtained for Model B.

4.1.4 Theoretical Mutations

The roles of the three arginine groups (Arg-11’, Arg-61’ and Arg-39”) in the enzymatic reaction of 4-OT have been the subject of recent debate. Theoretical mutations of Arg39”, Arg61’, Arg11’ to isostere citrullines (Cit) (Scheme 3) have been carried out by Metanis et al. using chemical synthesis techniques [43]. They showed that the electrostatic stabilization provided by the positively charged Arg39” was responsible for lowering the barrier of the 4-OT reaction. Arg11’ had a smaller effect on the barrier acting as an electron sink. The mutation of Arg61’ had no effect on the barrier. Recent molecular dynamics simulations, by Thiel et al. [44], questioned these results and instead argued that the mutational effect came from a distortion of the active site which altered the positioning of the substrate.
In order to resolve this issue, quantum chemical calculations on mutations of Arg39" to Cit-39", were performed using the two active site models described above. Models A and B were altered such that Arg-39" was changed to Cit39" (=NH$_2^+$ group was replaced by =O). The React and TS1 for each of the mutated models were re-optimized and the energies were calculated in order to compare the barriers between the mutated enzyme and the wild type enzyme.

The activation energy and the structural differences between the mutational analogues of Arg-39" and the corresponding un-mutated structures were compared for Model A and Model B. The main structural differences were the increased hydrogen bonding distances of the Cit to the substrate in both models (Figure 7). Due to the lack of positive charge in citrulline, the hydrogen bonding distances increased between 0.3-0.5 Å. As a result of this, the hydrogen bonds to the two other arginines (Arg-61’ and Arg-11’) were shorter in the mutant structures.
Figure 7. Optimized geometries of Arg39Cit mutant models. Fixed atoms are indicated with arrows.

The activation barriers for the mutated structures of Model A and Model B were plotted against the dielectric constant ($\varepsilon$) and compared to the barriers of the un-mutated models in Figure 8. The mutation of Arg39 in Model A gave an increased barrier of 2.2 kcal/mol ($\varepsilon=4$) and 4.2 kcal/mol ($\varepsilon=80$). For Model B the mutation gave a barrier increase of 3.2 kcal/mol ($\varepsilon=4$) and 2.8 kcal/mol ($\varepsilon=80$). These results correspond to a $k_{cat}$ decrease of approximately two orders of magnitude, which reproduces the experimental results nicely confirming the electrostatic hypothesis of Metanis et al. [43].
Figure 8. Calculated energy barriers for the first step of the 4-OT reaction for wild-type and Cit-39” models.

4.1.5 Conclusions.

The calculations using the two different models of the 4-OT active site gave support to the previously proposed reaction mechanism, in which Pro-1 acts as a general base to abstract a proton from the C3 position of the substrate and deliver it back to the C5 position. However, the quantum chemical models employed in the current study generate potential energy curves (Figure 4 and Figure 6) that are qualitatively different from those obtained by the previous QM/MM calculations [39,40,42]. This controversy can effectively be solved by considering the pKₐ differences of the substrate and Pro-1. Because the considered reaction step is a proton transfer, the energy difference can advantageously be compared to experimental relative acidities of the species involved. The pKₐ of the active site proline has experimentally been determined to 6.4 [31], which is ca 3 units lower than for proline in solution. The pKₐ of the substrate has also been determined experimentally to 10.5 [43b]. Binding of the substrate to the 4-OT active site is expected to lower this value, because of the stabilizing effect of the active site surrounding, mainly the positively-charged Arg-39” group. This suggests that the pKₐ values of the two groups are relatively close to each other, indicating that the energy of the intermediate is closer to the energy of the reactant and in disagreement with previous QM/MM results. Furthermore, the obtained theoretical mutational results reproduces the experimental mutations very nicely, adding further support to the quantum chemical models.
4.2 Modeling of Chlороacrylic Acid Dehalogenase

4.2.1 Introduction

cis-Chlороacrylic acid dehalogenase (cis-CAAD) is a bacterial enzyme that catalyzes the hydrolytic dehalogenation of cis-3-chloroacrylic acid to malonate semialdehyde and HCL (Scheme 4) [19,45,46]. Cis-CAAD is part of a degradation pathway in the soil bacterium *Pseudomonas pavonaceae* allowing the bacteria to utilize 1,3-dichloropropene, a pesticide that is rapidly degraded to cis-3-chloroacrylic acid, as their sole source of carbon energy [21]. It has been proposed that cis-CAAD is a newly evolved enzyme since it has only had a short exposure time to the substrate that does not exist naturally in nature. However it could also be that cis-CAAD catalyzes another natural substrate that has not yet been determined and that cis-CAAD reactivity towards cis-chloroacrylic acid is a promiscuous behavior [46].

![Scheme 4. Reaction catalyzed by cis-chloroacrylic acid dehalogenase.](image)

The cis-CAAD structure has many similarities with 4-OT and CAAD [22,23]. cis-CAAD has a similar structure as 4-OT and is composed of 3 monomers that each consists of two β-α-β strands [47]. Further, cis-CAAD has a terminal proline (Pro-1) that is catalytically important. Cis-CAAD also has two arginine groups in the active site, Arg-70 and Arg-73 that can interact with the substrates C1 carboxylate group.

Based on experimental studies [48-53] a general mechanism for the cis-CAAD reaction has been proposed (Scheme 5). A base in the active site activates a water molecule, which then attacks the substrate at the C3 position to form a tetrahedral intermediate. The tetrahedral intermediate either collapses to an enol intermediate (path a), which is followed by a keto/enol tautomerization to get the final product, or the tetrahedral intermediate abstracts a proton (path b) to form a chlorohydrin intermediate followed by release of $H^+$ and $Cl^-$ to get the product.
DFT models were employed in order to establish the reaction mechanism of cis-CAAD. First, the uncatalyzed reaction was studied using a very small model. Second, the enzymatic reaction was studied using a quite large model of the active site.

4.2.2 Uncatalyzed Reaction

The uncatalyzed water reaction of cis-CAAD was investigated using a small model. The model contained the substrate, cis-3-chloroacrylic acid, and two water molecules, for a total of 14 atoms. The first water molecule (W1) was added as the hydration water and the second water (W2) was added as an activator of W1 to shuttle the proton to the carboxylate group of the substrate. In the optimized reactant structure (1, Figure 9), a hydrogen bonding network was formed between the carboxylate of the substrate – W2 – W1 and the chloride of the substrate. The transition state for the nucleophilic attack was optimized (2, Figure 9). A tetrahedral intermediate where both the chloride and the hydroxyl were bound was not observed. Instead the chloride was displaced in a concerted step to the intermediate (3, Figure 9).
The chloride ion was then removed from the model and another water molecule (W3) was added to model the keto-enol tautomerization step. The total potential energy curve for the uncatalyzed reaction is shown in Figure 10. The first step is the rate-limiting step with a calculated barrier of 30.2 kcal/mol in gas phase and 31.8 kcal/mol using a dielectric constant of $\varepsilon = 80$. This small model reproduces remarkably well the experimental value of 33.3 kcal/mol determined by Wolfenden and co-workers [53]. The barrier for the second step, the keto-enol tautomerization, was calculated to be 19.9 kcal/mol ($\varepsilon = 80$).
4.2.3 Active Site Model

A model of the active site was built based on the native crystal structure of cis-CAAD (PDB code 2FLZ [47]). The model contained 9 residues: Pro-1, Arg-73, His-28, Tyr-103’, Glu-114, His-69, Leu-119, Thr-34 and Arg-70 together with a water molecule (W1) (A, Figure 11). The model had a total of 159 atoms and a total charge of 0.

In the reactant structure the two arginine groups and His-28 bind the carboxylate end of the substrate, positioning it for the nucleophilic attack (B, Figure 11). W1 forms a hydrogen bond to Pro-1 (2.43 Å) and Glu-114 (2.32 Å), indicating that both of these groups were in position to act as the base that activates the water molecule. W1 was in position to attack the substrate C3, with an O-C distance of 3.66 Å. In the optimized transition state structure the O-C distance between W1 and C3 was 1.76 Å (C, Figure 11). In the intermediate structure (D, Figure 11) the chloride had totally been displaced by W1 and no tetrahedral intermediate where both Cl and W1 were bound to the substrate was formed. Instead Cl was released in a concerted fashion eliminating the possibility of a chlorohydrin intermediate forming and excluding path b in Scheme 2. Glu-114 acts as the catalytic base in the reaction, but the proton is shuttled from the water though Pro-1.
The barrier for this first step using $\varepsilon=4$ is calculated to be 21.9 kcal/mol, which is somewhat overestimated compared to the experimental value of 16.6 kcal/mol obtained on CAAD [53]. In order to model the second part of the cis-CAAD reaction (keto-enol tautomerization) the chloride was removed and the model was re-optimized (A and B, Figure 12). A hydrogen bonding network was observed between the hydroxyl of the substrate and the C2 position of the substrate, going via Tyr-103’, Glu-114, and Pro-1. At first, a concerted transition state for the proton transfer was searched for but this was unsuccessful. Instead the tautomerization was found to proceed in a stepwise mechanism of three steps. First a proton was transferred from Glu-114 to Pro-1. The barrier of this first step was very small (0.6 kcal/mol using $\varepsilon=4$).
This step was followed by a concerted proton transfer from the substrate to Tyr-103’ coupled with a proton transfer from Tyr-103’ to Glu-114. The barrier for this step was 6.2 kcal/mol using $\varepsilon=4$. The last step was the proton transfer from Pro-1 to the substrates C2 position. This step had the significant barrier of 13.6 kcal/mol ($\varepsilon=4$). When looking at the total energy barrier of the tautomerization step the barrier was 15.6 kcal/mol (Figure 13). All the optimized structures are displayed in Figure 12. The barrier for the enzymatic tautomerization step was only c:a 4 kcal/mol lower than in the non-enzymatic step, which means that there is a possibility that the tautomerization reaction takes place outside of the enzyme.
Figure 12. Optimized stationary points for the enzymatic keto-enol reaction. A) Schematic drawing of the model. B) Enol intermediate. C) TS for proton transfer from Glu-114 to Pro-1. D) Intermediate with a protonated Pro-1. E) TS for proton transfer from Pro-1 to substrate. For clarity, several groups were omitted in the ball-and-stick structures (B-H).
4.2.4 Protonated Proline

Since the protonation state of Pro-1 was not clear from the experimental studies, a version of the enzymatic model was constructed where Pro-1 was chosen to be protonated. The model had a total of 160 atoms and a charge of +1. The reactant (A, B, Figure 14) transition state (C, Figure 14) and intermediate (D, Figure 14) structures were optimized and a potential energy was evaluated. The barrier was calculated to be 31.7 kcal/mol with $\varepsilon=4$, which is 9.8 kcal/mol higher than the model with a neutral Pro-1. This rules out the possibility of Pro-1 being in the protonated form.
Figure 14. Optimized geometries of the enzymatic model of cis-CAAD with a protonated Pro-1. A) Schematic drawing of the model. B) Reactant geometry. C) Transition state geometry. D) Product geometry. Optimized distances shown in Å. Stars indicate centers that are fixed to their crystallographic positions during geometry optimization. (In B, C and D certain parts of the model have been omitted for the sake of clarity).

4.2.5 Alternative Substrate-Water Binding

Another possibility that was investigated was the position of the nucleophilic water relative to the substrate. It could be positioned on either side of the substrate. An alternative model was built based on the active site model but positioning the water on the opposite side of the substrate (Figure 15). The reactant (A, B Figure 15), transition state (C, Figure 15), and intermediate (D, Figure 15) structures were optimized and a potential energy curve was calculated. We find that Glu-114 still acts as the catalytic base. However, the proton shuttling now is affected by Tyr-103' instead of Pro-1 in the previous case. The barrier was calculated to be 26.4 kcal/mol with $\varepsilon=4$, which is 4.5 kcal/mol higher than the previous scenario, making this possibility less likely.
4.2.6 Conclusions

The uncatalyzed reaction mechanism was first studied with a small model and gave very good results compared to the experimental rate constant. The calculations indicated that the nucleophilic attack of water and the release of the chloride ion took place in a concerted step. For the enzymatic reaction the same concerted mechanism was observed. Glu-114 acts as the catalytic base to abstract a proton from the water molecule. The role of Pro-1 seems to be to shuttle a proton from the water to Glu-114 (Scheme 6).

The keto-enol tautomerization then takes place to complete the reaction. It is possible that the Tyr-Glu-Pro triad in the active site shuttles the proton from the substrates hydroxyl group to
the C2 carbon. The barrier for this is calculated to be 15.6 kcal/mol, which is somewhat lower than the barrier of 19.9 computed in the non-enzymatic tautomerization reaction. The possibilities of Pro-1 being protonated and an alternative binding of the nucleophilic water have also been considered in this study. Both of these alternatives were shown to be less likely since they lead to higher barriers.

**Scheme 6.** Suggested cis-CAAD reaction mechanism based on the results from the calculations.
Chapter 5
Conclusions

This thesis has dealt with how density functional theory can be used to model enzymatic reactions. Specifically, two enzymes have been treated, namely 4-Oxalocrotonate Tautomerase (4-OT) and cis-Chloroacrylic Acid Dehalogenase (cis-CAAD). This has generated new insights into the reaction mechanisms and the roles of the various groups at the active sites.

In the case of 4-OT it was shown that a relatively small QM model with solvent effects gave results contradicting previous QM/MM results. The larger QM model confirmed this picture, with a first transition state that was rate-limiting and an intermediate that was much lower than previously estimated. Barriers were also calculated for mutated models showing that Arg-39” is catalytically important because it stabilized the charge that develops at the substrate during the course of the reaction.

For cis-CAAD the uncatalyzed reaction was studied using a very small model. The calculated reaction barrier for this model was very close to the experimentally-determined barrier. A large model of the enzyme active site gave a relatively good agreement with the energy barrier of the enzymatic reaction. The calculations indicated that the nucleophilic attack of water and the release of the chloride ion took place in one concerted step. Glu-114 is suggested to act as the catalytic base that abstracts a proton from the water molecule. The role of Pro-1 is proposed to be the shuttling a proton from the water to Glu-114.

The results of both studies testify to the usefulness of quantum chemical methods in the field of enzymology. This methodology is very likely to have an even greater impact on the field in the future, as the size of the models increases and experience is gained on how to handle very large systems.
References


