Quantum Chemical Modeling of Asymmetric Enzymatic Reactions

Maria E. S. Lind
Abstract

Computational methods are very useful tools in the study of enzymatic reactions, as they can provide a detailed understanding of reaction mechanisms and the sources of various selectivities. In this thesis, density functional theory has been employed to examine four different enzymes of potential importance for biocatalytic applications. The enzymes considered are limonene epoxide hydrolase, soluble epoxide hydrolase, arylmalonate decarboxylase and phenolic acid decarboxylase. Besides the reaction mechanisms, the enantioselectivities in three of these enzymes have also been investigated in detail. In all studies, quite large quantum chemical cluster models of the active sites have been used. In particular, the models have to account for the chiral environment of the active site in order to reproduce and rationalize the experimentally observed selectivities.

For both epoxide hydrolases, the calculated enantioselectivities are in good agreement with experiments. In addition, explanations for the change in stereochemical outcome for the mutants of limonene epoxide hydrolase, and for the observed enantioconvergency in the soluble epoxide hydrolase are presented.

The reaction mechanisms of the two decarboxylases are found to involve the formation of an enediolate- or a quinone methide intermediate, supporting thus the main features of the proposed mechanisms in both cases. For arylmalonate decarboxylase, an explanation for the observed enantioselectivity is also presented.

In addition to the obtained chemical insights, the results presented in this thesis demonstrate that the quantum chemical cluster approach is indeed a valuable tool in the field of asymmetric biocatalysis.
List of Publications

The results presented in this thesis are based on the following publications, referred to in the text by their Roman numerals I-IV.

I. Quantum Chemistry as a Tool in Asymmetric Biocatalysis: Limonene Epoxide Hydrolase Test Case
   Maria E. S. Lind and Fahmi Himo

II. Theoretical Study of Reaction Mechanism and Stereoselectivity of Arylmalonate Decarboxylase
    Maria E. S. Lind and Fahmi Himo

III. Quantum Chemical Modeling of Enantioconvergency in Soluble Epoxide Hydrolase
    Maria E. S. Lind and Fahmi Himo
    Manuscript

IV. Theoretical Study of the Reaction Mechanism of Phenolic Acid Decarboxylase
    Xiang Sheng, Maria E. S. Lind and Fahmi Himo
    Submitted for publication

Reprints were made with permission from the publishers.
Paper I: © 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim
Paper II: Copyright 2014 American Chemical Society
Publication not included in this thesis:

i. Ribonucleotide reductase inhibition by metal complexes of Triapine (3-aminopyridine-2-carboxaldehyde thiosemicarbazone): A Combined Experimental and Theoretical Study
Ana Popović-Bijelić, Christian R. Kowol, Maria E. S. Lind, Jinghui Luo, Fahmi Himo, Éva A. Enyedy, Vladimir B. Arion and Astrid Gräslund

Parts of this thesis are based on, and further developed from, the following licentiate thesis.

Quantum Chemical Modeling of Asymmetric Enzymatic Reactions Applications to Limonene Epoxide Hydrolase and Arylmalonate Decarboxylase
Contents

List of Publications v

Abbreviations ix

1 Introduction 1

2 Basic theory 3
  2.1 Density functional theory 3
    2.1.1 Dispersion 5
    2.1.2 B3LYP functional 5
  2.2 Enzyme kinetics 8
  2.3 Transition state theory 9
  2.4 Selectivity in enzymatic reactions 10

3 Quantum Chemical Modeling of Enzymatic Reactions 13
  3.1 Constructing an active site model 13
  3.2 Surrounding effects 15
  3.3 Computational details 15

4 Mechanisms and Selectivities of Epoxide Hydrolases 17
  4.1 Selectivity aspects in EH reactions 17
  4.2 Limonene epoxide hydrolase 18
    4.2.1 Quantum chemical studies of LEH (Paper I) 19
  4.3 Soluble epoxide hydrolase 23
    4.3.1 Proposed reaction mechanism 24
    4.3.2 Quantum chemical studies of sEH (Paper III) 24

5 Mechanisms and Selectivities of Decarboxylases 31
  5.1 Arylmalonate decarboxylase 31
    5.1.1 Proposed reaction mechanism 32
    5.1.2 Arylmalonate decarboxylase (Paper II) 32
5.2 Phenolic acid decarboxylase .......................... 37
5.2.1 Proposed reaction mechanism ....................... 37
5.2.2 Phenolic acid decarboxylase (Paper IV) ............ 38

6 Concluding Remarks ........................................ 43

Sammanfattning på svenska .................................. 45

Acknowledgments .............................................. 47

References ..................................................... 49
Abbreviations

AMDase  | Arylmalonate Decarboxylase
BsPAD   | Bacillus subtilis Phenolic Acid Decarboxylase
CPCM    | Conductor-like Polarizable Continuum Model
DFT     | Density Functional Theory
ee      | enantiomeric excess
EH      | Epoxide Hydrolase
GGA     | Generalized Gradient Approximation
HF      | Hartree-Fock
LEH     | Limonene Epoxide Hydrolase
L(S)DA  | Local (Spin) Density Approximation
PAD     | Phenolic Acid Decarboxylase
PCA     | p-coumaric acid
PDB     | Protein Data Bank
QC      | Quantum Chemical
rac-SO  | racemic Styrene Oxide
sEH     | soluble Epoxide Hydrolase
SMD     | Solvation Model Density
SO      | Styrene Oxide
StEH1   | Solanum tuberosum Epoxide Hydrolase
TS      | Transition State
WT      | Wild-type
ZPE     | Zero Point Energy
Amino acid abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Amino Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/Ala</td>
<td>Alanine</td>
</tr>
<tr>
<td>C/Cys</td>
<td>Cysteine</td>
</tr>
<tr>
<td>D/Asp</td>
<td>Aspartate</td>
</tr>
<tr>
<td>E/Glu</td>
<td>Glutamate</td>
</tr>
<tr>
<td>F/Phe</td>
<td>Phenylalanine</td>
</tr>
<tr>
<td>G/Gly</td>
<td>Glycine</td>
</tr>
<tr>
<td>H/His</td>
<td>Histidine</td>
</tr>
<tr>
<td>I/Ile</td>
<td>Isoleucine</td>
</tr>
<tr>
<td>K/Lys</td>
<td>Lysine</td>
</tr>
<tr>
<td>L/Leu</td>
<td>Leucine</td>
</tr>
<tr>
<td>M/Met</td>
<td>Methionine</td>
</tr>
<tr>
<td>N/Asn</td>
<td>Asparagine</td>
</tr>
<tr>
<td>P/Pro</td>
<td>Proline</td>
</tr>
<tr>
<td>Q/Gln</td>
<td>Glutamine</td>
</tr>
<tr>
<td>R/Arg</td>
<td>Arginine</td>
</tr>
<tr>
<td>S/Ser</td>
<td>Serine</td>
</tr>
<tr>
<td>T/Thr</td>
<td>Threonine</td>
</tr>
<tr>
<td>V/Val</td>
<td>Valine</td>
</tr>
<tr>
<td>W/Trp</td>
<td>Tryptophan</td>
</tr>
<tr>
<td>Y/Tyr</td>
<td>Tyrosine</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Many important industrial and biochemical reactions do not occur on a reasonable time scale and must therefore be catalyzed in some way. The development of catalytic reactions is thus of great interest in many different fields of chemistry, and in all of them the aim is the same: to achieve rate enhancement by lowering the energy barrier of the rate-determining step. Moreover, enantiopure compounds are becoming increasingly important in industrial applications, primarily in the production of new pharmaceuticals, which demands the development of new and highly enantioselective catalysts.

Enzymes are biological molecules with the ability to catalyze a wide array of chemical transformations with high efficiency and selectivity. Owing to these facts, enzymes are attractive as complements or substitutes to metal catalysts in asymmetric synthetic chemistry. In addition, the use of enzymes as biocatalysts offers an environmentally sustainable route for chemical production. However, many enzymes suffer from low activity or poor stereoselectivity for an extended substrate scope, which is limiting the application in synthetic organic chemistry. Nowadays, these problems can generally be addressed by protein engineering techniques with the introduction of mutations.

A detailed understanding of the reaction mechanism and the sources of different selectivities is of great use in order to control or manipulate an enzymatic reaction. Such knowledge allows for rational protein engineering of enzymes to fit specific needs, and could also aid in the design of inhibitors. However, the understanding of the underlying factors that control these properties, in particular the stereoselectivity, is often quite limited.

Computational chemistry has become an important tool for studying enzymatic reaction mechanisms, and also for studying different kinds of selectivities. Quantum chemical models have until quite recently mainly been employed for the former, that is investigating different mechanistic proposals.
1. Introduction

Nowadays, it is possible to treat larger systems with a higher accuracy than before. Along with an increased model size, more complex questions can be addressed, such as enantioselectivity and the effects of mutating residues not directly involved in the chemical step.

In the present thesis, density functional theory has been employed to study the reaction mechanisms and selectivities of two different types of enzymatic reactions, namely epoxide hydrolysis and decarboxylation. One of the main objectives has been to rationalize the selectivities of these enzymes using large active site models. In this context, an accuracy on the order of 1 kcal/mol in relative transition-state energies is needed and therefore, these studies also represent cases in which the limitations of the modeling approach can be evaluated.

The thesis is structured as follows. In Chapter 2, the basic theory of density functional theory is presented, together with some enzymatic kinetics and the concepts used for relating the calculated energies to experiments. The modeling approach employed in this thesis is described in Chapter 3. Chapters 4 and 5 present the results from the calculations. Finally, the thesis work is summarized in Chapter 6, and some main conclusions are presented.
The work presented in this thesis is based on density functional theory (DFT) calculations with the hybrid functional B3LYP, which is a cost-effective and widely used approach to model large molecular systems. In this chapter, a brief introduction to DFT and B3LYP will be given. In addition, some basics of enzyme kinetics are presented together with the concepts used for relating the calculated activation barriers and selectivities to experimental data. A more extensive description of these aspects can be found in standard textbooks.7,8,9,10

2.1 Density functional theory

Central in DFT are the two theorems by Hohenberg and Kohn.11 The first theorem states that the electron density $\rho(r)$ of a system determines its ground-state electronic energy. The electronic energy can thus be expressed as a functional of the electron density, which in turn is a function of three spatial coordinates. The second theorem states that the electron density obeys the variational principle, i.e. any trial density that satisfies the given boundary conditions will always give an energy higher or equal to the exact ground state energy ($E_0$).

$$E[\rho'(r)] \geq E_0 \quad (2.1.1)$$

The ground-state electronic energy is expressed as a sum of the kinetic energy of the electrons ($T$), the nuclear-electron attraction ($V_{Ne}$), and the electron-electron repulsion ($V_{ee}$). The functional describing the electron-electron repulsion contains two parts, namely the classical Coulomb repulsion ($J$) and a term accounting for the non-classical contributions, referred to as the exchange-correlation functional ($E_{xc}$).

$$E[\rho(r)] = T[\rho(r)] + V_{Ne}[\rho(r)] + V_{ee}[\rho(r)] \quad (2.1.2)$$
2. Basic theory

\[ E[\rho(\mathbf{r})] = T[\rho(\mathbf{r})] + V_{Ne}[\rho(\mathbf{r})] + J[\rho(\mathbf{r})] + E_{xc}[\rho(\mathbf{r})] \quad (2.1.3) \]

Kohn and Sham\textsuperscript{12} introduced a formalism in which the kinetic energy functional \((T)\) is divided into two parts, one representing the kinetic energy of a fictitious system of non-interacting electrons \(T_{ni}\) and one small correction term. The kinetic energy of the non-interacting system can be determined exactly and is calculated by introducing orbitals. The correction term accounts for the difference in kinetic energy between the system of non-interacting electrons and the real system \((T - T_{ni})\), and is included in the exchange-correlation term \((E_{xc})\), according to equation \((2.1.5)\).

\[ E[\rho(\mathbf{r})] = T_{ni}[\rho(\mathbf{r})] + V_{Ne}[\rho(\mathbf{r})] + J[\rho(\mathbf{r})] + E_{xc}[\rho(\mathbf{r})] \quad (2.1.4) \]

\[ E_{xc}[\rho(\mathbf{r})] = (T[\rho(\mathbf{r})] - T_{ni}[\rho(\mathbf{r})]) + (V_{ee}[\rho(\mathbf{r})] - J[\rho(\mathbf{r})]) \quad (2.1.5) \]

All terms in equation \((2.1.4)\) can be determined explicitly, except for the exchange-correlation part. Consequently, only approximate solutions can be obtained, although the DFT theory is exact. The various DFT methods differ thus in the choice of the functional describing the exchange-correlation energy, and the accuracy of a particular method therefore depends on its exchange-correlation functional. Generally, this functional is divided into two terms that separately account for the exchange and correlation energies.

One of the simplest approximations of the exchange-correlation functional is the Local (Spin) Density Approximation, L(S)DA. Within this approximation, the electron density is treated as uniform electron gas and the exchange-correlation energy depends only on the local density where it is evaluated. The expression for the L(S)DA exchange functional was proposed by Slater,\textsuperscript{13} and some corresponding correlation functionals have been derived from quantum Monte Carlo methods. One example is the correlation functional by Vosko, Wilk and Nusair.\textsuperscript{14}

The electron density in a molecule is in general far from uniform and a better alternative for treating molecular systems is therefore to use a generalized gradient approximation (GGA) functional. Such functionals also depend on the gradient of the electron density. A variety of GGA exchange and correlation functionals have been developed, where two popular ones are the exchange functional by Becke (B)\textsuperscript{15} and the correlation functional by Lee, Yang and Parr (LYP).\textsuperscript{16} Some functionals also incorporate a percentage of the exact Hartree-Fock (HF) exchange energy, that is the exact exchange energy for the system of non-interacting electrons. These types of functionals are referred to as hybrid functionals. One of the most popular hybrid-GGA functionals is the B3LYP functional,\textsuperscript{17,18} which will be discussed in section 2.1.2.
2.1. Density functional theory

### 2.1.1 Dispersion

Dispersion is a long-range attractive force arising from electron correlation. Many density functionals, such as B3LYP, fail to describe these interactions due to their local dependence on the electron density. There are, however, several ways to incorporate the dispersion interactions in the density functional.\(^\text{19}\) One method is to parametrize the density functional against an experimental set of data including non-covalent interactions or against high-level calculations. Examples of such functionals are the Minnesota family of functionals.\(^\text{20,21}\) Another alternative is to add an empirical long-range attractive term to the existing functional.\(^\text{19,22,23}\) In such a scheme, a correction term is added to the DFT energy, according to equation (2.1.6).

\[
E_{\text{total}} = E_{\text{DFT}} + E_{\text{disp}} \quad (2.1.6)
\]

In the work presented in this thesis, the dispersion corrections \((E_{\text{disp}})\) were calculated using the DFT-D2 method by Grimme.\(^\text{22}\) Within this approach, \(E_{\text{disp}}\) is based on an atom pairwise correction of the form \(C_6/R^6\) given by:

\[
E_{\text{disp}} = -s_6 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{C_{ij}^6}{R_{ij}^6} f_{\text{dmp}}(R_{ij}) \quad (2.1.7)
\]

The sum is over all atom pairs, where \(C_{ij}^6\) is the dispersion coefficient for atom pair \(ij\), \(R_{ij}\) is the distance between atoms \(i\) and \(j\), and \(N\) is the number of atoms. Furthermore, the dispersion-correction term includes a scaling factor \(s_6\), which is specific for each density functional, and a damping function \(f_{\text{dmp}}\) to correct for the behavior at short interatomic distances.

#### 2.1.2 B3LYP functional

The B3LYP functional provides a good trade-off between accuracy and computational cost, and has therefore become one of the most widely used density functionals in chemistry.\(^\text{24}\) The exchange-correlation functional is defined as:

\[
E_{xc}^{B3LYP} = (1-a)E_x^{LSDA} + aE_x^{HF} + bE_x^{B88} + (1-c)E_c^{LSDA} + cE_c^{LYP} \quad (2.1.8)
\]

where \(E_x^{LSDA}\) is the LSDA exchange, \(E_x^{HF}\) is the Hartree-Fock exchange, \(E_x^{B88}\) is the gradient correction term to the LSDA exchange by Becke,\(^\text{15}\) \(E_c^{LSDA}\) is the LSDA correlation functional by Vosko, Wilk and Nusair,\(^\text{14}\) and \(E_c^{LYP}\) is the correlation functional of Lee, Yang and Parr.\(^\text{16}\) The three empirical parameters \(a\), \(b\) and \(c\) are set to 0.20, 0.72 and 0.81, respectively.
The performance of B3LYP has been evaluated in various benchmark tests, in which calculated properties are compared to accurate experiments or high-level calculations. A large number of benchmark tests have been published, and the results from some of these are presented here. The performance of B3LYP for geometries and various energies are summarized in Tables 2.1 and 2.2. In an early test, B3LYP with the 6-31G(d) basis set was tested against the geometries within the G2 test set (55 small molecules of first and second row atoms) and was shown to have an average absolute error of 0.013 Å in bond lengths, 0.62° in bond angles and 0.35° in dihedral angles.\textsuperscript{25} Employing the larger 6-311+G(3df,2p) basis set did not result in any significant increase in the accuracy of the bond lengths and angles.\textsuperscript{25} Similar accuracy was reported in a more recent benchmark test (containing 44 small organic molecules), in which B3LYP with the 6-31+G(d) basis set was shown to have an average absolute error of 0.009 Å in bond lengths and 1.29° in bond angles.\textsuperscript{26} The performance of B3LYP in the calculations of hydrogen bonds and van der Waals interaction distances has also been evaluated recently. The results show that hydrogen bonds are accurately described with an average absolute error of 0.02 Å.\textsuperscript{27} For van der Waals interaction distances, the results are less accurate with an average absolute error of 0.599 Å.\textsuperscript{28}

Table 2.1 Average absolute deviation of B3LYP for geometries from various benchmark tests.

<table>
<thead>
<tr>
<th>Geometries</th>
<th>Result</th>
<th>Basis set</th>
</tr>
</thead>
<tbody>
<tr>
<td>G2 test set of 55 small molecules:\textsuperscript{25}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.013</td>
<td>6-31G(d)</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>0.62</td>
<td>6-31G(d)</td>
</tr>
<tr>
<td>Dihedral angles (°)</td>
<td>0.35</td>
<td>6-31G(d)</td>
</tr>
<tr>
<td>Test set of 44 small organic molecules:\textsuperscript{26}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.009</td>
<td>6-31+G(d)</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.29</td>
<td>6-31+G(d)</td>
</tr>
<tr>
<td>HB4/04 test set of 4 hydrogen bonding dimers:\textsuperscript{27}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hydrogen-bond lengths (Å)</td>
<td>0.02</td>
<td>6-31+G(d,p)</td>
</tr>
<tr>
<td>Test set of 10 van der Waals complexes:\textsuperscript{28}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>van der Waals interactions (Å)</td>
<td>0.599</td>
<td>aug-cc-pVTZ</td>
</tr>
</tbody>
</table>

For the atomization energies within the G2 test set, B3LYP with the 6-311+G(3df,2p) basis set was shown to have an average absolute error of 2.20 kcal/mol. Using the larger G3/05 test set containing 454 experimental energies of different types, an average absolute error of 4.14 kcal/mol was reported.\textsuperscript{29} A more recent benchmark study reported a similar average absolute error of 4.2 kcal/mol in the calculations of barrier heights of 24 different reactions.\textsuperscript{30} In the same study, B3LYP was found to have an average
Table 2.2 Average absolute deviation (kcal/mol) of B3LYP for energies from various benchmark tests.

<table>
<thead>
<tr>
<th>Energies</th>
<th>Result</th>
<th>Basis set</th>
</tr>
</thead>
<tbody>
<tr>
<td>G2 test set of 55 small molecules:</td>
<td>2.20</td>
<td>6-311+G(3df,2p)</td>
</tr>
<tr>
<td>Atomization energies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G3/05 test set of 454 experimental energies:</td>
<td>4.14</td>
<td>6-311+G(3d2f,2p)</td>
</tr>
<tr>
<td>Various energies(^a)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DBE18 test set of 18 diverse bond energies:</td>
<td>6.7</td>
<td>different for each subset(^b)</td>
</tr>
<tr>
<td>Bond energies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DBH24 test set of 24 diverse barrier heights:</td>
<td>4.2</td>
<td>MG3S</td>
</tr>
<tr>
<td>Barrier heights</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HB23 test set of 23 hydrogen bonding dimers:</td>
<td>1.60</td>
<td>aug-cc-pVTZ</td>
</tr>
<tr>
<td>Interaction energies</td>
<td>0.44 (B3LYP-D3)</td>
<td></td>
</tr>
<tr>
<td>Test set of 10 van der Waals complexes:</td>
<td>1.35</td>
<td>aug-cc-pVTZ</td>
</tr>
<tr>
<td>Interaction energies</td>
<td>0.82 (B3LYP-D2)</td>
<td></td>
</tr>
<tr>
<td>TKNC306 database of 306 data points:</td>
<td>3.47</td>
<td>different for each subset(^b)</td>
</tr>
<tr>
<td>Complete set(^c)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GMTKN24 database of 731 data points:</td>
<td>5.4</td>
<td>(aug)-def2-QZVP</td>
</tr>
<tr>
<td>Complete set(^c)</td>
<td>3.6 (B3LYP-D2)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Enthalpies of formation, ionization energies, electron affinities, proton affinities and hydrogen-bond energies. \(^b\) See reference for details. \(^c\) Complete set consisting of subsets for thermochemistry, kinetics and non-covalent interactions.

2.1. Density functional theory

Table 2.2 Average absolute deviation (kcal/mol) of B3LYP for energies from various benchmark tests.

For non-covalent interaction energies, B3LYP showed an average absolute error of 1.60 kcal/mol for hydrogen bonds\(^{31}\) and 1.35 kcal/mol for van der Waals interactions.\(^{28}\) It is worth noting that these errors were reduced to 0.44 and 0.82 kcal/mol, respectively, by adding the empirical dispersion correction according to the methods of Grimme.\(^{22,23}\)

Larger benchmark test sets, consisting of subsets for thermochemistry, kinetics and non-covalent interactions, have been used to assess the performance of various density functionals for a broader set of data. Two such databases are the TKNC306 (containing 306 data points)\(^{32}\) and the larger GMTKN24 (containing 731 data points).\(^{33}\) B3LYP showed an overall average absolute error of 3.47\(^{32}\) and 5.4 kcal/mol\(^{33}\) for the complete sets within the TKNC306 and GMTKN24 datasets, respectively. Notably, the average absolute error in the GMTKN24 dataset was reduced to 3.6 kcal/mol by adding
2. Basic theory

the empirical dispersion correction to the energies according to the DFT-D2 method.\textsuperscript{33}

On the basis of the results from the benchmark tests presented here, the B3LYP functional can be considered to be accurate enough to address mechanistic questions, especially when including the empirical dispersion correction. Employing a medium-sized basis set, such as the 6-31G(d), usually gives molecular geometries of sufficient accuracy. For energies, on the other hand, the results are in general more sensitive to the size of the basis set. A common procedure is therefore to optimize the geometry using a medium-sized basis set after which a single-point energy calculation using a larger basis set is performed. It is important to remember that in the studies of selectivities, the point of interest lies in the calculation of relative energies. Here, an improved accuracy can be expected due to cancellation of systematic errors.

2.2 Enzyme kinetics

Enzymes are large biomolecules that catalyze a wide range of biochemical transformations, often with high efficiency. Like any other catalyst, enzymes achieve rate enhancement by lowering the free energy of activation ($\Delta G^\ddagger$) without altering the reaction free energy. The exact origin of the catalytic efficiency of enzymes is a matter of debate. According to one theory, the main contribution comes from electrostatics, in that the enzyme active site is pre-organized to stabilize the transition state more than the ground state.\textsuperscript{34,35} The result of this transition-state stabilization is a lowered activation barrier and thus a higher reaction rate compared to the uncatalyzed reaction.

Michaelis-Menten kinetics is the simplest model by which to describe enzyme-catalyzed reactions. The enzyme (E) and substrate (S) are in equilibrium with an enzyme-substrate complex (E · S). Converting the substrate to product (P) is assumed to be rate-determining with a rate constant, $k_{\text{cat}}$. The final step corresponds to the product release.

$$\begin{align*}
E + S & \underset{k_{-1}}{\xrightarrow{k_1}} E \cdot S \underset{k_{\text{cat}}}{\xrightarrow{k_1}} E \cdot P \rightleftharpoons E + P
\end{align*}$$

(2.2.1)

The rate of product formation is given by equation (2.2.2).

$$v = \frac{d[P]}{dt} = k_{\text{cat}}[E \cdot S]$$

(2.2.2)

According to the steady-state approximation, the concentration of the enzyme-substrate complex is constant since the rates of its formation and its breakdown are equal.

$$\frac{d[E \cdot S]}{dt} = k_1[E][S] - (k_{-1} + k_{\text{cat}})[E \cdot S] = 0$$

(2.2.3)
2.3. Transition state theory

\[ [E \cdot S] = \frac{k_1[E][S]}{(k_{-1} + k_{cat})} \] (2.2.4)

Equation (2.2.4) can be simplified by defining the Michaelis constant \( K_M \), and by expressing the concentration of the enzyme in terms of the initial concentration of enzyme and the concentration of enzyme-substrate complex, according to \([E] = [E]_0 - [E \cdot S]\).

\[ [E \cdot S] = \frac{[E][S]}{K_M} = \frac{([E]_0 - [E \cdot S])[S]}{K_M} \] (2.2.5)

\[ K_M = \frac{k_{-1} + k_{cat}}{k_1} \]

Combining the final expression for \([E \cdot S]\) with equation (2.2.2) results in the Michaelis-Menten equation (2.2.6).

\[ v = k_{cat}[E \cdot S] = \frac{k_{cat}[E]_0[S]}{K_M + [S]} \] (2.2.6)

At low substrate concentrations, \([S] \ll K_M\), most active sites can be considered as unoccupied by substrate, and the initial concentration of enzyme can be approximated by the concentration of free enzyme \(([E]_0 = [E])\). Consequently, the reaction in (2.2.6) appears as overall second order with a rate constant of \(k_{cat}/K_M\).

\[ v = \frac{k_{cat}}{K_M}[E][S] \] (2.2.7)

At high substrate concentrations, \([S] \gg K_M\), the enzyme will be saturated with substrate \(([E \cdot S] = [E]_0)\) and the reaction rate is equal to the maximum velocity.

\[ v = V_{max} = k_{cat}[E]_0 \] (2.2.8)

\(k_{cat}\) is the overall rate constant and will reflect the binding of the transition state relative to the binding of the ground state. In cases where \(k_{-1} \gg k_{cat}\), \(K_M\) corresponds to the dissociation constant of the enzyme-substrate complex, and is thus an estimation of the substrate affinity for the active site. The specificity constant \(k_{cat}/K_M\) is a measure of the catalytic efficiency since it accounts for both how well the substrate binds to the active site and how fast the substrate is converted to product.

### 2.3 Transition state theory

The energy of a molecule as a function of its geometry can be visualized by an energy surface. Reactants, intermediates, products and transition states are
all stationary points on this surface, and are characterized by their second
derivatives of the energy with respect to the nuclear coordinates, i.e. the
force constants. Reactants, intermediates and products correspond to energy
minima with all force constants positive. Transition states, on the other hand,
are saddle points on the energy surface and have one negative force constant
corresponding to the reaction coordinate. Consequently, a transition state is
the highest point on the lowest energy path interconverting two minima.

All stationary points, including transition states, can be located computa-
tionally. From these structures, information about the barriers associated
with going from one minimum to another can be obtained. Transition state
theory relates the rate constant $k$ to the activation barrier, according to equa-
tion (2.3.1).

$$k = \frac{k_B T}{h} e^{-\Delta G^+/RT}$$  \hspace{1cm} (2.3.1)

Where $k_B$ is the Boltzmann constant, $T$ is the absolute temperature, $h$ is the
Planck constant and $R$ is the gas constant. $\Delta G^+$ in equation (2.3.1) is the
free energy of activation. At room temperature (298.15 K), a 1.4 kcal/mol
change in barrier corresponds to a change in the rate constant of one order
of magnitude.

Using transition state theory, different reaction pathways can be evalu-
at on the basis of their corresponding energy barriers. It should be noted
that using only the enthalpy part ($\Delta H^+$) of the free energy ($\Delta G^+ = \Delta H^+ -
\Delta S^+$) in many cases is a good approximation, since the entropy effects in
the chemical steps of enzymatic reactions are expected to be rather small.
Recent studies employing quantum mechanical/molecular mechanical simu-
lations have obtained entropy effects of about 1 kcal/mol.$^{36,37,38}$

2.4 Selectivity in enzymatic reactions

There are many different kinds of selectivities, such as stereo- and regio-
selectivities. The selectivity of an enzymatic reaction can be assessed from
the difference in free energy between the transition states ($\Delta \Delta G^+$) resulting
in the formation of different products. The larger the free energy difference
is, the greater the selectivity of the reaction. An example of a free energy
profile for a reaction resulting in two different products is shown in Figure
2.1, and the rate of formation of $P_1$ and $P_2$ can be expressed according to
equations (2.4.1) and (2.4.2).
2.4. Selectivity in enzymatic reactions

Figure 2.1 Free energy profile of a chemical reaction in which one substrate can be converted into two different products.

\[ v_1 = \frac{d[P_1]}{dt} = \frac{k_{\text{cat},1}}{K_M} [E][S] \]  
\[ v_2 = \frac{d[P_2]}{dt} = \frac{k_{\text{cat},2}}{K_M} [E][S] \]  

The selectivity of the reaction is determined by the ratio of the products, which in turn is determined by the ratio of the individual rates.

\[ \frac{v_1}{v_2} = \frac{k_{\text{cat},1} K_M}{k_{\text{cat},2} K_M} = e^{(\Delta G_2^\dagger - \Delta G_1^\dagger)/RT} = e^{\Delta \Delta G^\dagger / RT} \]  

In the case of \( P_1 \) and \( P_2 \) being enantiomers, the degree of selectivity (or the enantiopurity of the product) can be expressed in terms of the enantiomeric excess (ee), which is calculated according to equation (2.4.4).

\[ \% ee = \frac{[P_1] - [P_2]}{[P_1] + [P_2]} \cdot 100 \]  

(2.4.4)
Chapter 3

Quantum Chemical Modeling of Enzymatic Reactions

Enzymes are large and complex biomolecules, typically consisting of thousands of atoms. Despite the size, the catalysis usually takes place in a quite small and well-defined region called the active site. In most cases, catalysis is primarily governed by direct involvement of amino acid residues, metal ions and organic cofactors in the active site. In the modeling approach applied in this thesis, reaction mechanisms and selectivities are studied by considering only the residues of the active site using density functional theory. The effect of the parts of the enzyme that are not explicitly included in the model are approximated with a polarizable continuum model and a coordinate-locking scheme. This methodology is often referred to as the quantum chemical cluster approach and has been employed in a large number of studies concerning the reaction mechanisms of many different enzymes. In this chapter, the methodology and computational protocol used in this thesis for studying enzymatic reactions are presented.

3.1 Constructing an active site model

The procedure for designing an active site model is illustrated in Figure 3.1. The structural information used to construct the model is usually obtained from available X-ray crystal structures, from which a limited selection of active site residues is extracted, together with any cofactors or water molecules that are considered important for the catalysis. For instance, functional groups responsible for the chemical transformations and residues with explicit interactions with the substrate are always included in the model. Depending on the chemical problem, additional residues that define the bind-
Construction of an active site model usually starts with a crystal structure from which important active site residues are extracted and truncated. An inhibitor bound to the active site gives useful guidelines of how to insert the substrate. Atoms that are kept fixed during the geometry optimizations are indicated by an asterisk.

Peptide backbones are sometimes also kept, in particular if they form hydrogen bonds to the substrate or if an increased flexibility of the side chain is necessary. Hydrogen atoms are added and the protonation states of the
3.2 Surrounding effects

The effects of the surrounding enzyme can be considered to influence the reaction in two ways, namely by electrostatic polarization effects and by imposing steric strain. The electrostatic polarization effects are taken into account by approximating the surrounding enzyme with a homogenous polarizable medium with some dielectric constant (ε), which is typically set to 4 when modeling enzyme environments. The solvation effect is then calculated as a single-point energy on the geometries optimized in gas phase. The effect of the steric strain imposed by the enzyme is taken into account by fixing certain atomic coordinates to their crystallographic positions, typically where the truncations were made. In some cases, fixing additional coordinates is necessary in order to prevent unrealistic movements of the residues. While maintaining the overall features of the active site, it is still important to allow enough flexibility for the residues to adapt to the geometrical changes during the course of the reaction.

3.3 Computational details

All calculations presented in this thesis were carried out using the hybrid DFT functional B3LYP\textsuperscript{17,18} as implemented in the Gaussian03 program package.\textsuperscript{43} Geometries were optimized employing the 6-31G(d,p) basis set. More accurate energies were obtained by single-point calculations on the optimized structures using the larger 6-311+G(2d,2p) basis set. The solvation effects of the protein surrounding were taken into account by single-point calculations at the 6-31G(d,p) level of theory using either the conductor-like polarizable continuum model (CPCM)\textsuperscript{44,45,46,47} or the solvation model density (SMD),\textsuperscript{48} both with a dielectric constant ε = 4. Analytical frequency calculations were carried out at the same level of theory as the geometry optimizations to obtain the zero-point energies (ZPE). The final energies reported here are thus calculated using the large basis set and are corrected for solvation, ZPE and dispersion effects, according to the DFT-D2 method.\textsuperscript{22}

Additional computational details

The entropy effects in the chemical steps of an enzymatic reaction are in general expected to be small and do not have any implications on the mecha-
Therefore, the entropy is usually not considered when modeling enzymatic reactions with the quantum chemical cluster approach. On the other hand, significant entropy effects are expected in the event of binding or release of a gas molecule during the reaction. Such a case is the decarboxylation reaction, in which a gaseous carbon dioxide is eliminated (Papers II and IV). In accordance with previous quantum chemical studies, the entropy gain for such a process is estimated to be equal to the translational entropy of the free molecule being released. The translational entropy for carbon dioxide is calculated to be 11.1 kcal/mol at room temperature and this value is thus added to the reaction energy of the step in which carbon dioxide is released.
Chapter 4

Mechanisms and Selectivities of Epoxide Hydrolases

Epoxides and vicinal diols are valuable building blocks in the preparation of pharmaceuticals.\textsuperscript{50,51,52} By using epoxide hydrolases (EHs), these compounds can be prepared in high enantiomeric purity, starting from cheap racemic epoxides.\textsuperscript{52,53} In this chapter, the results from the quantum chemical calculations on two EHs are summarized. The main objective of these studies has been to identify and rationalize the sources of enantioselectivity. In the first study, the desymmetrization of a \textit{meso}-epoxide catalyzed by limonene epoxide hydrolase (LEH) was investigated. The second study concerns the enantioconvergent hydrolysis of a racemic epoxide by a soluble EH from potato (\textit{Solanum tuberosum} EH1, StEH1). In addition to the chemical aspects of LEH and StEH1, these studies also represent two interesting cases in which the usefulness of the cluster approach in reproducing experimental selectivities can be evaluated.

4.1 Selectivity aspects in EH reactions

Many enzymatic reactions are enantioselective, which means that one enantiomer out of a racemic substrate is faster converted into product. The enantiopurity of the residual substrate depends thus on the enantioselectivity of the particular enzyme. In the hydrolysis of epoxides, each enantiomer of substrate can furthermore be attacked at two different carbon centers. As a result, the enantiopurity of the diol product is dictated by both the enantioselectivity and regioselectivity of the oxirane ring-opening.\textsuperscript{54}

The intrinsic reactivities of the epoxide carbons depend on the substituent pattern. Bulky substituents generally promote attack at the less hindered car-
bon. However, since the transition state involves the development of a partial positive charge, electron donating substituents may overcome the unfavorable steric effects and instead steer the attack to the more substituted carbon. Furthermore, the reactivity of the epoxide carbons can vary depending on the reaction conditions. For instance, the hydrolysis of styrene oxide in water under acidic conditions occurs exclusively at the benzylic carbon (C1)\textsuperscript{55,56} due to the partial positive charge developed at this carbon upon protonation of the epoxide. Under basic conditions (3M KOH solution), on the other hand, the epoxide is opened at both C1 and C2 in a 51:49 ratio\textsuperscript{55} (Scheme 4.1).

\begin{center}
\textbf{Scheme 4.1} Regioselectivity in the hydrolysis of styrene oxide under acidic conditions\textsuperscript{55,56} and basic conditions.\textsuperscript{55}
\end{center}

In an enzyme, these principles are not easily applied since the regioselectivity will be influenced also by the electronic and steric factors specific to the particular active site. Depending on the EH, the ring opening can thus involve nucleophilic attack at either of the oxirane carbons.

Furthermore, some EHs show opposite regioselectivity for different enantiomers, which allows for the formation of enantiopure diols in a so-called enantioconvergent process\textsuperscript{57,58}. Such a scenario results in the conversion of a racemic epoxide to a single enantiomeric product with a theoretical 100% yield. This process has thus a clear advantage over a kinetic resolution, in which the limiting yield of the enantiopure diol is 50%.

\section*{4.2 Limonene epoxide hydrolase}

LEH catalyzes the hydrolysis of limonene-1,2-epoxide to generate the corresponding limonene-1,2-diol. LEH is one of few known EHs that are found to act through a mechanism consisting of one single step. In this one-step mechanism, a water molecule is activated by an aspartate (Asp132) and attacks at one of the oxirane carbons. Simultaneously, another aspartate (Asp101) is delivering a proton to the evolving oxyanion (Scheme 4.2 B).\textsuperscript{59}

LEH has been shown to catalyze the hydrolysis of a range of different epoxides.\textsuperscript{60} However, the stereoselectivity is generally poor for substrates other than the natural limonene epoxide. The enzyme shows for example an ee of only 14% in favor of the (R,R)-product in the hydrolysis of
4.2. Limonene epoxide hydrolase

**Scheme 4.2** A) LEH-catalyzed hydrolysis of cyclopentene oxide. B) Reaction mechanism of LEH. (Figure from Paper I).

cyclopentene oxide (Scheme 4.2 A). Applying directed evolution with iterative saturation mutagenesis, Zheng and Reetz engineered LEH mutants that show an improved enantioselectivity toward the (R,R)-diol or an inverse enantioselectivity toward the (S,S)-diol in the hydrolysis of cyclopentene oxide, see Figure 4.1.

**4.2.1 Quantum chemical studies of LEH (Paper I)**

The enantioselectivity of the wild-type (WT) LEH and each mutant in the (R,R)- and (S,S)-selective pathways in Figure 4.1 have been investigated using large cluster models of the active site. The active site model of the WT LEH was constructed on the basis of the X-ray crystal structure in complex with a heptanamide (PDB 1NWW). The model consists of 259 atoms and includes the Asp132-Arg99-Asp101 catalytic triad, Tyr53, Asn55 and the crystallographically observed water molecule. In addition, the residues that were subjected to mutagenesis in the directed evolution experiments were included, namely Met32, Leu35, Leu74, Met78, Ile80, Val83, Leu103, Leu114 and Ile116 (Figure 4.2). The models of each mutant were constructed by modifying the WT LEH active site model according to the experimental mutations in Figure 4.1.
Analyzing the enantioselectivity

The optimized transition states for attack at C1 (TS_{C1}) and C2 (TS_{C2}) in WT LEH are shown in Figure 4.3. Notably, the substrate is displaced in one direction or the other, depending on at which carbon the attack occurs. The substrate displacements in TS_{C1} and TS_{C2} are associated with similar energy penalties, as reflected by their almost isoenergetic activation barriers (calculated to be 15.7 and 15.6 kcal/mol for TS_{C1} and TS_{C2}, respectively). The relative energy of the transition states is calculated to be 0.1 kcal/mol and is a measure of the enantioselectivity. Consequently, the (R,R)- and (S,S)-products are formed in similar quantities, in agreement with the very small experimental ee of 14%.

The calculated differences in activation barriers for WT and all mutants are compared to the experimentally determined enantioselectivities in Figure 4.4. As seen in the comparison, each mutant in the (R,R)- and (S,S)-selective pathways are calculated to show the correct enantioselectivity. The trend within each pathway is, however, not perfectly accounted for.

By examining the transition-state structures, it turns out that the change in enantioselectivity can be explained by steric hindrance. In the (R,R)-pathway, the mutations result in a less hindered attack at C2, thus fa-
4.2. Limonene epoxide hydrolase

**Figure 4.2** Optimized structure of the active site model of wild-type LEH with the residues of the catalytic triad shown in yellow. Atoms indicated by asterisks were kept fixed to their crystallographically observed positions. (Figure from Paper I).

**Figure 4.3** Optimized transition states for attack at C1 and C2 in the wild-type, resulting in the (S,S)- and (R,R)-diols, respectively. For clarity, only the catalytic triad is shown together with the nucleophilic water molecule and the substrate. (Figure from Paper I).
Figure 4.4 Comparison between the calculated and experimentally observed enantioselectivities for the wild-type LEH and all mutants. (Figure from Paper I).

voring the formation of the \((R,R)\)-product. For example, both the Leu74Ile and Ile80Cys mutations introduce smaller side chains which result in some release of steric hindrance on one side of the active site that will favor the substrate displacement in the attack at C2. The same reasoning holds for the mutants in the \((S,S)\)-pathway, where the introduced mutations instead result in a less hindered attack at C1 or a more hindered attack at C2, thus favoring the formation of the \((S,S)\)-product. The Leu114Cys mutation will open up space on one side of the active site in a way that favors the displacement of the substrate when the attack occurs at C1. With the Ile80Phe mutation, a larger side chain is introduced on the other side of the active site compared to Leu114Cys, and will result in a less favored substrate displacement in the TS for attack at C2. Notably, the Met32Cys mutation appears in both the \(R1 \rightarrow R2\) and \(S2 \rightarrow S3\) mutations. As seen in Figure 4.4, the effect of this mutation is not well captured by the calculations and its role remains unclear.

Summary and conclusions

The results presented here show that a quantum chemical cluster model is able to describe the effects of the mutations in very good agreement with experiments. The enantioselectivity can be rationalized by how much steric hindrance the mutations release or introduce to allow or prevent the substrate displacement in the transition states for attack at either oxirane carbon. The trend in enantioselectivity within each pathway is not reproduced,
and could be a consequence of using the WT crystal structure as the basis for the models of the mutants. For example, introducing mutations in the active site might result in larger conformational changes than what can be captured by the WT active site model. Another source of error is the mutations at the periphery of the model, which may not be properly described. One such example is the Met32Cys mutation that appears in both the $R_1 \rightarrow R_2$ and $S_2 \rightarrow S_3$ mutations. It is possible that a Cys in this position is hydrogen bonding to a residue that is not explicitly included in the active site model, and the effect of the Met32Cys mutation is therefore not well reproduced in the calculations.

4.3 Soluble epoxide hydrolase

An interesting feature found for some soluble epoxide hydrolases (sEH) is the ability to operate in an enantioconvergent one-enzyme fashion, in which a single enantiomer of product is obtained from a racemic substrate.\textsuperscript{58,61,62,63,64} The degree of enantioconvergency is described by the regioselectivity coefficients,\textsuperscript{65} where a coefficient of 50% implies no regioselectivity, whereas 100% indicates full regioselectivity.

The soluble EH from potato (StEH1) has been shown to catalyze the enantioconvergent conversion of racemic styrene oxide ($rac$-SO) derivatives to the corresponding ($R$)-diol in good yields and with high ee.\textsuperscript{66} From the experiments, ($S$)-SO was found to preferably be attacked at the more substituted carbon (C1) in a 98:2 ratio. ($R$)-SO, on the other hand, was attacked at the less substituted carbon (C2) in a 93:7 ratio, see Scheme 4.3.\textsuperscript{66} The enzyme thus yields similar regioselectivity for ($S$)-SO as in solution under acidic conditions, whereas the opposite regioselectivity is observed for ($R$)-SO.

![Scheme 4.3 Enantioconvergent hydrolysis of $rac$-SO in StEH1 with experimental regioselectivity coefficients.\textsuperscript{66}](image-url)
4.3.1 Proposed reaction mechanism

The proposed reaction mechanism of StEH1 is shown in Scheme 4.4 and occurs through the formation of a covalent alkyl-enzyme intermediate, resulting from the nucleophilic attack by an aspartate (Asp105) at one of the oxirane carbons (step 1, Scheme 4.4). Two active site tyrosines (Tyr154 and Tyr235) have been shown important for catalysis and are proposed to assist in the epoxide ring-opening. The ester of the alkyl-intermediate is then hydrolyzed by a water molecule (steps 2 and 3, Scheme 4.4), which is activated by a histidine-aspartate charge relay (His300-Asp265). Furthermore, the residues constituting an oxyanion hole are suggested to be involved in the stabilization of the tetrahedral intermediate formed in the hydrolysis of the ester (Int2, Scheme 4.4).

One matter of debate with regard to the reaction mechanism shown in Scheme 4.4 is whether the catalytic His300 is protonated or neutral during the alkylation half-reaction (step 1, Scheme 4.4).

4.3.2 Quantum chemical studies of sEH (Paper III)

In order to establish the sources of the selectivity, the details of the full reaction mechanism have to be understood, and in particular the protonation state of His300 has to be resolved. Therefore, two quantum chemical cluster models of the active site were used in which His300 is treated as neutral (Model-N) or positively charged (Model-P). Both models were designed on the basis of the X-ray crystal structure of StEH1 in complex with the inhibitor valpromide (PDB 2CJP) and contain the catalytic triad (Asp105-His300-Asp265), the two active site tyrosines (Tyr154 and Tyr235), the residues of the oxyanion hole (backbone amide bond of Asp105-Trp106 and Gly32-Phe33) and a crystallographically observed water molecule. Additional residues defining the active site were also included since they are expected to be important for the selectivity, see Figure 4.5. Both models were used to study the full reaction mechanism and the enantioconvergency of StEH1 with rac-SO.

Due to the shape of the active site, each enantiomer of substrate can be envisioned to bind with the phenyl group positioned toward the inside of the binding site (“in”) or towards the more open part of the active site (“out”), see Figure 4.5. For each binding mode, attack at both C1 and C2 were considered.

Reaction mechanism of StEH1

Two representative energy profiles from Model-N and Model-P are shown in Figure 4.6, and correspond to the attack at C1 of (S)-SO with the phenyl oriented “in”. The main difference between the two models is the protonation
4.3. Soluble epoxide hydrolase

Scheme 4.4 Suggested reaction mechanism of soluble epoxide hydrolases (residue numbering from StEH1).

state of His300 during the alkylation half-reaction (step 1, Scheme 4.4). Since His300 has to be neutral in order to activate the water molecule in the first step of the hydrolysis (step 2, Scheme 4.4), the extra proton in Model-P was moved from His300 to the oxyanion (represented by the extra step $\text{Int1} \rightarrow \text{Int1}'$ in Figure 4.6 B) that is formed in $\text{Int1}$. As a consequence, the subsequent stationary points will therefore differ in the protonation state of the intermediate.
The energy profiles are calculated to be rather similar and give general support to the proposed reaction mechanism, in which a relatively stable covalent intermediate is formed and subsequently hydrolyzed in a two-step process. In both models, the rate limiting-step is calculated to correspond to the dissociation of the tetrahedral intermediate ($\text{Int1} \rightarrow \text{TS3}$), in accordance with experiments. The barriers are calculated to be 20.6 and 17.4 kcal/mol in Model-N and Model-P, respectively, as compared to the experimental barrier of ca. 16 kcal/mol.

In Figure 4.6, the free energy of the overall reaction between the free substrate and water is also indicated, and is calculated to be $-4.4$ kcal/mol. This value determines at which energy the next catalytic cycle starts, but can also be used to estimate the energetics of the step for product release and regeneration of the active site ($\text{Prod} \rightarrow \text{React}$).

In Model-N (Figure 4.6 A), the product release and regeneration of the active site is estimated to cost 11.6 kcal/mol, which will have implications on the first barrier in the following cycle. Considering the energy cost for closing the catalytic cycle, the barrier for $\text{TS1}$ in Model-N is calculated to be ca. 16 kcal/mol, which is in good agreement with the experimentally determined alkylation barrier of ca. 15 kcal/mol.

In the case of Model-P (Figure 4.6 B), the energy of the enzyme-product complex is calculated to be 2.3 kcal/mol higher than the free energy of the overall reaction. Here, the product release and regeneration of the active site is instead calculated to be exothermic by ca. 2 kcal/mol, and will therefore be

---

**Figure 4.5** Optimized structures of the enzyme-substrate complexes in Model-N with (S)-SO positioned toward the interior (React-N-S-in) and toward the exterior of the active site (React-N-S-out). All non-polar hydrogens are omitted for clarity.
4.3. Soluble epoxide hydrolase

Figure 4.6 Calculated potential energy profiles for the hydrolysis of (S)-SO at C1 in A) Model-N and B) Model-P with the phenyl positioned "in". All energies are relative to the corresponding enzyme-substrate complex of lowest energy, React-N-S-out and React-P-S-out for Model-N and Model-P, respectively.

not affect the barriers in the following cycle. The barrier for TS1 in Model-P is thus calculated to be 4.9 kcal/mol, which is considerably lower compared to both the experimental barrier for TS1 (of ca. 15 kcal/mol)\textsuperscript{83} and the corresponding barrier in Model-N (of ca. 16 kcal/mol).

It is not possible to completely rule out any of the mechanisms simply on the basis of the energies presented here, and the protonation state of His300 thus remains elusive. Furthermore, since the barriers for TS3 are calculated to be of similar energy (in Model-N) or of higher energy (in Model-P) compared to the barriers of TS1, both transition states can influence the selectivity. Consequently, both TS1 and TS3 were considered in the analysis of the enantioconvergency.
Enantioconvergency of StEH1

The resulting TS1 and TS3 barriers for both the (S)- and (R)-SO substrates in Model-N are summarized in Table 4.1, where the selectivity-determining barriers for the formation of each enantiomer of product are shown in bold face. As indicated in Table 4.1, each substrate orientation can be attacked at either C1 or C2, resulting in inversion or retention of configuration. The absolute configuration of the resulting diol products are also indicated.

Table 4.1 Calculated barriers (kcal/mol) for nucleophilic attack (TS1) and dissociation of tetrahedral intermediate (TS3) in Model-N. Selectivity-determining steps are indicated in bold face.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Orientation</th>
<th>Relative energy(^a)</th>
<th>Attacked carbon</th>
<th>Resulting diol</th>
<th>TS1(^a)</th>
<th>TS3(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(S)-SO</td>
<td>&quot;in&quot;</td>
<td>+1.6</td>
<td>C1 (R)</td>
<td>+4.4</td>
<td>+4.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C2 (S)</td>
<td>+7.6</td>
<td>+3.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&quot;out&quot;</td>
<td>0.0</td>
<td>C1 (R)</td>
<td>+10.2</td>
<td>+11.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C2 (S)</td>
<td>+5.6</td>
<td>+5.3</td>
<td></td>
</tr>
<tr>
<td>(R)-SO</td>
<td>&quot;in&quot;</td>
<td>+0.5</td>
<td>C1 (S)</td>
<td>+8.1</td>
<td>+12.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C2 (R)</td>
<td>+5.2</td>
<td>+7.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&quot;out&quot;</td>
<td>+1.9</td>
<td>C1 (S)</td>
<td>+6.6</td>
<td>+8.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C2 (R)</td>
<td>+7.2</td>
<td>+7.8</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) The energies are given relative to the binding mode of lowest energy.

In the case of the (S)-SO substrate, the path with lowest energy barriers resulting in the formation of the (R)-diol corresponds to the attack at C1 of React-N-S-in. For the generation of the (S)-diol, the path with lowest energy barriers corresponds to the attack at C2 of React-N-S-out. The other paths are associated with barriers of higher energies and will therefore not contribute. The selectivity-determining barriers are TS3 of 4.7 kcal/mol and TS1 of 5.6 kcal/mol, for the formation of the (R)- and (S)-diol, respectively (Table 4.1). For the (S)-SO substrate, the formation of the (R)-diol (attack at C1) is thus favored over the formation of (S)-diol (attack at C2), in agreement with experiments. The calculated energy difference of 0.9 kcal/mol corresponds to a ratio of ca. 80:20, which is somewhat underestimated compared to the experimental selectivity of 98:2.

For the (R)-SO substrate, a similar analysis shows that three paths have selectivity-determining barriers of comparable energies, while the forth path has higher barriers. The two paths of lower barriers correspond to the attack at C2 of React-N-R-in and React-N-R-out, and are calculated to be 7.9 and 7.8 kcal/mol, respectively (Table 4.1). The third path corresponds to the attack at C1 of React-N-R-out with a barrier of 8.0 kcal/mol. Here, a Boltz-
mann distribution of these three barriers was used to deduce a selectivity of ca. 70:30, in favor of the (R)-diol. The calculated selectivity is in agreement with the experimental selectivity of 93:7, although underestimated also for this substrate.

Some trends in geometries that explain the different selectivities are observed when analyzing the optimized structures of TS1. For the (S)-SO substrate, the TS1 barrier of lowest energy corresponds to the nucleophilic attack at C1 of React-N-S-in (TS1-N-S-in-C1, Figure 4.7). Here, the phenyl ring is almost completely co-planar with the C1-C2 bond of the oxirane ring (C2-C1-C3-C4 dihedral angle $\phi = 1^\circ$), and allows thus for a maximal stabilization of the partial positive charge developed at C1. In the other structures of TS1 for attack at C1, the active site residues restrict the phenyl ring from adapting this co-planar alignment, which results in higher energy barriers.

For the (R)-SO substrate, the barrier for TS1 of lowest energy corresponds to nucleophilic attack at C2 of React-N-R-in (TS1-N-R-in-C2, Figure 4.7). In this case, the phenyl ring is oriented in a way to instead minimize the steric repulsions with the other substituents on the C1 carbon (dihedral angle $\theta = 39^\circ$). In the structures of TS1 for attack at C2 of higher energies, these repulsive steric interactions cannot be avoided due to the restrictions from the surrounding enzyme.

In the case of Model-P, the comparison of all barriers (not presented here) shows that the nucleophilic attack is preferred at C1 for both the (S)- and (R)-SO substrates, in disagreement with the experimental observations. The calculations indicate thus that Model-P fails to reproduce the enantiocon-
Summary and conclusions

The reaction mechanism and enantioconvergent hydrolysis of rac-SO by StEH1 have been investigated using two active site models that differ mainly in the protonation state of His300. The results from Model-N (neutral His300) are in somewhat better agreement with the experimental observations compared to the results from Model-P (protonated His300), in particular since only Model-N can reproduce the observed enantioconvergency.

In Model-N, the nucleophilic attack at the benzylic C1 position of the (S)-SO substrate is favored as the active site allows the phenyl ring to orient in a co-planar fashion relative to the C1-C2 bond of the oxirane ring, resulting in a stabilization of the developing positive charge on C1. For the (R)-SO substrate, on the other hand, the nucleophilic attack at C2 is favored by minimizing the steric repulsion between the phenyl ring and the substituents on C1.
Chapter 5

Mechanisms and Selectivities of Decarboxylases

Decarboxylation is an important and widespread process in biological systems and recently the reaction has also become increasingly important in biocatalytic applications. Decarboxylases can be used in the synthesis of, for example, carboxylic acids and terminal olefins, using abundant and renewable starting materials. In addition, many decarboxylases have also been shown to catalyze additional reactions, such as C-C bond formation, racemization, and hydration.

This chapter concerns the results from the mechanistic studies on two decarboxylases: arylmalonate decarboxylase (AMDase) and phenolic acid decarboxylase from Bacillus subtilis (BsPAD). In the case of AMDase, the issues of the origins of enantioselectivity have also been addressed by studying the decarboxylation of two different substrates. As opposed to many other decarboxylases, both AMDase and BsPAD are independent of cofactors. Instead, the reactions are dependent on specific hydrogen bonds and the ability of the substrate to delocalize the free electron pair that results from the decarboxylation.

5.1 Arylmalonate decarboxylase

AMDase catalyzes the asymmetric decarboxylation of α-arylmalonates to produce the corresponding carboxylic acids with high enantioselectivity (Scheme 5.1). The enzyme is cofactor-independent, robust and shows a high enantioselectivity, and is therefore interesting for use in biocatalysis to prepare chiral carboxylic acids by enzymatic decarboxylation.
However, AMDase suffers from a limited substrate scope and accepts only compounds with an \( \alpha \)-aryl or \( \alpha \)-alkenyl group.\(^{87}\)

\[
\begin{align*}
\text{COOH} & \quad \text{COOH} \quad \text{CO}_2 \\
\text{AMDase} & \quad \text{COOH}
\end{align*}
\]

**Scheme 5.1** Reaction catalyzed by AMDase.

### 5.1.1 Proposed reaction mechanism

The reaction mechanism of AMDase has been proposed to involve a selective decarboxylation of the pro-\( \text{R} \) carboxylate group to form a planar enediolate intermediate.\(^{97}\) Once formed, the intermediate is suggested to undergo protonation at the \( \text{Si} \) face by a cysteine residue, and the resulting (\( \text{R} \))-product is formed with inversion of configuration (Scheme 5.2).\(^{97,98}\)

\[
\begin{align*}
\text{O} & \quad \text{Ar} \quad \text{R} \quad \text{CO}_2 \\
\text{Enediolate} & \quad \text{H} \quad \text{S} \\
\text{Cys}_{188} & \quad \text{R-product}
\end{align*}
\]

**Scheme 5.2** Proposed reaction mechanism of AMDase.

Two neighboring oxyanion holes (referred to as the dioxyanion hole) located in the active site have been suggested to be important for substrate binding and stabilization of the enediolate intermediate formed during decarboxylation.\(^{87,99}\) Furthermore, since the enzyme only accepts malonate substrates with an aryl or alkenyl substituent in \( \alpha \)-position, the conjugated \( \pi \)-electron system formed in the enediolate intermediate appears to be important for stabilizing the negative charge developed at the \( \alpha \)-carbon. In Figure 5.1 A, the crystal structure of the active site of AMDase in complex with a benzylphosphonate inhibitor is shown.\(^{87}\) Apart from the dioxyanion hole, the active site contains two binding cavities, one large solvent-exposed and one small hydrophobic.\(^{87,99}\) On the basis of a docking study and the crystal structure of AMDase with the benzylphosphonate inhibitor, the substrate was proposed to bind to the active site according to Figure 5.1 B.\(^{87,99}\)

### 5.1.2 Quantum chemical study of AMDase (Paper II)

The reaction mechanism and selectivity of AMDase shown in Scheme 5.2 have been studied using two quantum chemical cluster models of different
sizes. Both models were designed on the basis of the crystal structure of AMDase in complex with benzylphosphonate (PDB 3IP8). A smaller model (model I, 81 atoms) was used first in order to characterize all transition states, using the \( \alpha \)-methyl-\( \alpha \)-phenylmalonate substrate. This model consists of the catalytically important residues proposed to constitute the dioxyanion hole (Gly74, Thr75, Ser76, Tyr126, and Gly189)\(^{87,99}\) together with the residue responsible for the protonation (Cys188), see Figure 5.2. The larger model (model II, 223 atoms) contains additional residues defining the large and small active site cavities, which are assumed to be important for the enantioselectivity, see Figure 5.2. In the case of model II, another substrate was also considered, namely the slightly smaller \( \alpha \)-methyl-\( \alpha \)-vinylmalonate.

**Reaction mechanism of AMDase**

Before investigating the reaction mechanism, several different modes of binding between the substrate and the active site models were evaluated. Interestingly, the proposed orientation of the substrate, where the pro-\( S \) carboxylate alone is stabilized by hydrogen bonds to the dioxyanion hole (Figure 5.1 B), could not be optimized. In fact, only structures in which both carboxylate moieties interact with the dioxyanion hole were found. Similar results were obtained in the calculations with both models I and II. The optimized enzyme-substrate complexes of lowest energy in model I (React-I-R) and model II (React-II-R) are shown in Figure 5.2. As a result of the additional groups included in model II, the hydrogen-bond pattern differs slightly be-
between the two models. Also, one extra hydrogen bond is observed between the substrate and the Ser76-Leu77 amide backbone in model II.

![Figure 5.2](image.png)

**Figure 5.2** Optimized structures of the enzyme-substrate complexes in active site model I and II that result in the formation of the (R)-product. All non-polar hydrogens are omitted for clarity. Fixed atoms are indicated by an asterisk.

Decarboxylation of α-methyl-α-phenylmalonate from both **React-I-R** and **React-II-R** was indeed shown to occur in two separate steps, i.e. the decarboxylation of the pro-R carboxylate followed by the protonation at the Si face by Cys188. Notably, neither a TS for decarboxylation of the pro-S carboxylate, nor a TS for concerted decarboxylation and protonation could be optimized in any model. In Figure 5.3, the optimized stationary points from model I are shown. The rate-limiting step of the reaction is calculated to be the C-C bond cleavage (**TS1**), which is associated with a barrier of 19.5 (in model I) or 16.2 kcal/mol (in model II). Thus, the larger model resulted in a 3 kcal/mol lower activation barrier. One contributing factor to the energy difference is that an extra hydrogen bond is observed in **TS1** of model II, which is not present in model I.

The results from both models show that the proposed reaction mechanism is energetically feasible. The barrier of 16.2 kcal/mol from model II is in good agreement with the experimentally determined barrier of ca. 14-16 kcal/mol, as determined from the available rate constants.\(^{87,95,100}\)
5.1. Arylmalonate decarboxylase

As pointed out above, model II includes some additional residues that define the binding pockets and is therefore better able to account for the chiral environment of the active site, compared to model I. In fact, the calculations show that model II is able to reproduce and rationalize the stereoselectivity. In Figure 5.4, the schematic drawings of the optimized enzyme-substrate complexes resulting in the (R)- and (S)-products are shown. The binding modes

**Figure 5.3** Optimized stationary points for the decarboxylation of α-methyl-α-phenylmalonate in active site model I. (Figure from Paper II).

**Enantioselectivity of AMDase**

As pointed out above, model II includes some additional residues that define the binding pockets and is therefore better able to account for the chiral environment of the active site, compared to model I. In fact, the calculations show that model II is able to reproduce and rationalize the stereoselectivity. In Figure 5.4, the schematic drawings of the optimized enzyme-substrate complexes resulting in the (R)- and (S)-products are shown. The binding modes
differ in the positions of the methyl- and phenyl substituents. In React-II-R, the phenyl substituent is located in the large binding cavity, whereas it in React-II-S instead is pointing toward the small binding cavity. Assuming that no rearrangement of the enediolate intermediate occurs, the protonation of the resulting intermediate will occur at the opposite face of React-II-S, as compared to React-II-R. Thus, the stereochemical outcome of the reaction will be determined by the orientation of the substituents relative to Cys188. It is known from experiments that the reaction with α-methyl-α-phenylmalonate and α-methyl-α-vinylmalonate generates the (R)-product with an ee of >99%\(^\text{95}\) and 99%,\(^\text{87}\) respectively, corresponding to a difference in transition state energies for the formation of the (R)- and (S)-products of at least 3 kcal/mol.

**Figure 5.4** Schematic drawings of the optimized enzyme-substrate complexes in model II resulting in the (R)-product (React-II-R) and the (S)-product (React-II-S).

The energy difference between the two enzyme-substrate complexes in Figure 5.4 was calculated to be 14.1 kcal/mol, in favor of React-II-R, and is a result of the steric clashes between the phenyl substituent and the residues of the small cavity. In the following TS for decarboxylation (TS1), the enantiodiscrimination is even greater. The large energy difference between the (R)- and (S)-selective pathways indicate an exclusive formation of the (R)-product, in accordance with experiments.\(^\text{95}\)

The energetic preference for the (R)-selective pathway is likely to be overestimated due to the rigidity of the model, in particular the rigidity of the small binding cavity. It was therefore of interest to compare the calculated enantioselectivity of the current substrate with the enantioselectivity of a smaller one. The calculations on α-methyl-α-vinylmalonate using model II show that this substrate can in fact bind in both orientations, as the energy between them differ by only 0.6 kcal/mol (still in favor of React-II-R). The energetic preference increases to 1.7 kcal/mol in the following rate-limiting TS1. Thus, the enantioselectivity in the case of α-methyl-α-vinylmalonate...
is not determined in the binding but rather in the following TS for C-C bond cleavage. Although calculated to be smaller, the energy difference still originates from the repulsive steric interactions between the vinyl substituent and the residues of the small binding cavity. The relative energy of 1.7 kcal/mol in TS1 is in qualitative agreement with the experimental ee of 99%. 87

Summary and conclusions

Two models of different sizes have been used in order to gain insight into the reaction mechanism and the enantioselectivity of AMDase. Both models support the two-step mechanism proposed previously, in which decarboxylation is followed by a stereoselective protonation at the Si face by the Cys188 residue. The enantioselectivity, on the other hand, could only be reproduced and rationalized when the residues of the small and large binding cavities were included in the model. These cavities have thus implications on the stereochemical outcome of the reaction since they restrict the binding and orientation of the substrate during the reaction. The enantioselectivity for both substrates is governed by the steric repulsion between the vinyl- or phenyl substituent and the residues defining the small binding cavity. In the case of α-methyl-α-phenylmalonate, the calculations show that the enantioselectivity is determined already at the binding, whereas for α-methyl-α-vinylmalonate, also the following TS is found to contribute to the enantioselectivity.

5.2 Phenolic acid decarboxylase

Phenolic acid decarboxylases (PADs) catalyze the decarboxylation of p-hydroxycinnamic acid derivatives to generate the corresponding p-vinyl phenols. 101,102,103,104 These enzymes have gained attention as they could serve as biocatalysts in the preparation of styrene derivatives and flavor compounds using phenolic acids from renewable sources, such as lignin. 88,105 Similarly to other cofactor-free decarboxylases, PADs have a relatively narrow substrate specificity and are restricted to cinnamic acid derivatives with a hydroxyl group in para position. 101,106

5.2.1 Proposed reaction mechanism

A two-step reaction mechanism has been proposed on the basis of the substrate specificity and structural data in combination with the results from mutational experiments, see Scheme 5.3 (with residue numbering from BsPAD). 101,107,108 The substrate is suggested to bind to the active site with the
carboxylate group forming hydrogen bonds to two tyrosine residues (Tyr11 and Tyr13) and with the \( p \)-hydroxyl group positioned toward a glutamate residue (Glu64). The first step of the reaction involves the formation of a \( p \)-quinone methide intermediate, resulting from the deprotonation of the phenolic hydroxyl by Glu64 and a concurrent protonation by a third tyrosine residue (Tyr19) at the C2 of the substrate. The second step is the C-C bond cleavage to form the \( p \)-vinyl phenol and carbon dioxide products (Figure 5.3). Furthermore, mutational analysis indicates that also an arginine residue (Arg41) is important for catalysis.\(^{107,108}\)

\[ \begin{align*}
\text{Glu}64 & \quad \text{Tyr}19 \\
\text{Tyr}11 & \quad \text{Tyr}13 \\
\text{R} = \text{H}, \text{OCH}_3 \text{ or OH}
\end{align*} \]

\[ \begin{align*}
\text{Scheme 5.3} & \text{ Previously proposed reaction mechanism of PAD (residue numbering from BsPAD).}^{101,107,108}
\end{align*} \]

### 5.2.2 Quantum chemical study of BsPAD (Paper IV)

The reaction mechanism has been investigated using a large quantum chemical cluster model, which was designed on the basis of the X-ray crystal structure of the Tyr19Ala mutant of BsPAD in complex with \( p \)-hydroxycinnamic acid (\( p \)-coumaric acid, PCA) (PDB 4ALB\(^{107}\)). The model consists of the PCA substrate, the residues proposed to be important for catalysis (with Ala19 restored to Tyr), and the residues constituting the active site. In addition, two water molecules were also included in the model, see Figure 5.5. Different substrate orientations were evaluated, and for each one different protonation states of both the substrate and Glu64 were examined. Furthermore, the possibility of other residues than Tyr19 serving as proton donors in the first step of the reaction was also considered.

**Reaction mechanism of BsPAD**

The reaction mechanism that turned out to have the most plausible energies is shown in Scheme 5.4. Notably, the suggested enzyme-substrate complex
Figure 5.5 Optimized enzyme-substrate complex of BsPAD. Non-polar hydrogens of the active site residues are omitted for clarity. Atoms indicated by an asterisk were kept fixed during the geometry optimizations.

(React) does not correspond to the binding mode of PCA that was observed in the crystal structure and is therefore also inconsistent with the proposed reaction mechanism in Scheme 5.3. Here, the \( p \)-hydroxyl group instead forms hydrogen bonds to Tyr11 and Tyr13, and the carboxyl group is interacting with Glu64 and a water molecule (Figure 5.5). Previously, the \( p \)-hydroxyl group was proposed to be deprotonated in the first step of the reaction mechanism.\(^{107}\) However, in the substrate orientation in React, no such general base in the vicinity of the hydroxyl group could be observed. Nonetheless, it can be assumed that the proton is lost upon substrate binding, as a result of the lowered \( \text{p}K_a \) of the hydroxyl group caused by the interactions with Tyr11 and Tyr13. In React, the \( p \)-hydroxyl group is thus in its deprotonated state, whereas the carboxyl group of PCA was found to be protonated, see Figure 5.5.

In the first step of the reaction, the proton on the PCA carboxyl group is transferred to Glu64 (React \( \rightarrow \text{Int1} \), Scheme 5.4). The energy of \textbf{Int1} is calculated to be +5.6 kcal/mol relative to React. The optimized structures of \textbf{Int1} and the following stationary points in the reaction mechanism are shown in Figure 5.6. In the subsequent step, the \( p \)-quinone methide intermediate is formed through a proton transfer from Glu64 (\textbf{Int1} \( \rightarrow \text{TS1} \rightarrow \text{Int2} \), Scheme 5.4). For this step, the energy barrier and the energy of the formed intermediate are calculated to be 16.0 and +6.9 kcal/mol relative to React.
respectively. The last step involves the C-C bond cleavage to generate the \( p \)-vinyl phenolate and carbon dioxide (\( \text{Int2} \rightarrow \text{TS2} \rightarrow \text{Prod} \)). The energy barrier of this step is calculated to be 15.9 kcal/mol, and results in the formation of \( \text{Prod} \) with an energy of +1.6 kcal/mol, relative to \( \text{React} \) (including the correction for the entropy).

\[
\text{React} \xrightarrow{\text{TS1}} \text{Int1} \xrightarrow{\text{TS2}} \text{Int2} \xrightarrow{\text{CO}_2} \text{Prod}
\]

**Scheme 5.4** BsPAD reaction mechanism as proposed by the calculations.

From the calculations, the energy barriers for the substrate protonation \( \text{TS1} \) and the C-C bond cleavage \( \text{TS2} \) are very close and it is therefore not possible to determine which step is rate-limiting. Nevertheless, an overall barrier of 16.0 kcal/mol agrees well with the barrier range of 12-15 kcal/mol, as determined from the experimental rate constants for PADs from different organisms.\(^8,109,110\)

At \( \text{Int2} \), a hydrogen bond is observed between the carboxylate of the sub-
strate and Tyr19 (Figure 5.6). This hydrogen bond turns out to be important, since the intermediate without it is calculated to be 3.2 kcal/mol higher in energy, and the following transition state for C-C bond cleavage is higher by 5.0 kcal/mol.

The reaction mechanism presented in Scheme 5.4 is found to be consistent with the results from the mutational experiments on BsPAD. In this study, the Glu64Ala variant was shown to be inactive, which supports the proposal that this residue is responsible for the protonation of the substrate in TS1. Previously, Tyr19 was proposed to serve as proton donor. However, the Tyr19Ala mutant still retained 4% relative activity which shows that the
residue is important, but not critical for activity. In the mechanism proposed here, Tyr19 is instead involved in hydrogen bonding with the carboxylate group of the substrate in the \( p \)-quinone methide intermediate to facilitate the C-C bond cleavage in TS2. As mentioned above, the barrier in absence of this hydrogen bond is calculated to be 5.0 kcal/mol higher in energy, which thus accounts for the decrease in activity observed in the Tyr19Ala variant.

The Tyr11Phe/Tyr13Phe variant was shown to be inactive. From the calculations, Tyr11 and Tyr13 form hydrogen bonds to the hydroxyl group of the substrate throughout the whole reaction. In addition, the residues are involved in lowering the \( pK_a \) of the hydroxyl group, which is of importance for the quinone formation. Finally, also the Arg41Ala variant was shown to be inactive. According to the calculations, Arg41 plays a role in the substrate binding and in the stabilization of the negative charge on the carboxylate group of the substrate in Int1 and TS1, and of Glu64 in Int2, TS2 and Prod.

Several alternative pathways were also investigated, but turned out to be associated with higher energies. For example, the structure of the \( p \)-quinone methide intermediate in the previously proposed mechanism (Scheme 5.3) could not be optimized since the proton of the \( p \)-hydroxyl group could not be transferred to Glu64. The resulting intermediate formed after the protonation at C2 is calculated to be +36.2 kcal/mol, relative to the corresponding enzyme-substrate complex in Scheme 5.3.

**Summary and conclusions**

The calculations give general support to a reaction mechanism involving a quinone methide intermediate. However, the substrate is found to bind in a different orientation as compared to that in the crystal structure, with the \( p \)-hydroxyl group rather than the carboxyl group of the substrate interacting with Tyr11 and Tyr13. Furthermore, the calculations indicate that Glu64 is responsible for the protonation of the substrate in the first step of the reaction. The previously proposed proton donor Tyr19 is instead found to be involved in hydrogen bonding to the carboxylate of the substrate, which results in a lower activation barrier of the following TS for C-C bond cleavage. The barriers for the protonation and the C-C bond cleavage are essentially equal in energy, and the calculated barrier of ca. 16 kcal/mol is in good agreement with the experimental value. In addition, the agreement with the experimental mutational analysis provides further support to the mechanism proposed here.
Chapter 6

Concluding Remarks

In the present thesis, the quantum chemical cluster approach has been employed to model four different enzymes, two epoxide hydrolases and two decarboxylases. The studies presented here involve the reaction mechanisms, and in three of these enzymes the experimentally observed selectivities were also addressed. Large active site models were used that include essentially all residues constituting the binding pockets, which is of particular importance when modeling enantioselectivities.

The results of the calculations on the two epoxide hydrolases give detailed insights into the sources of the experimentally observed selectivities. The enantioselectivity of the desymmetrization reactions catalyzed by the limonene epoxide hydrolase variants were found to be controlled by the steric hindrance introduced or relieved by the different mutations. A steric reasoning was also used to explain the opposite regioselectivity observed in the soluble epoxide hydrolase from potato for the two enantiomers of styrene oxide.

For the two decarboxylases, the calculations support the main features of the previously proposed reaction mechanisms. For phenolic acid decarboxylase, the calculated mechanism involves a different substrate binding and an alternative acid responsible for the protonation compared to what was proposed previously. In the case of arylmalonate decarboxylase, details about the factors influencing the enantioselectivity were obtained from the studies on two prochiral substrates with different substituents.

As demonstrated in this thesis, quantum chemical calculations on enzymatic reactions can be a very useful tool for providing valuable information regarding the reaction mechanisms and the origins of different kinds of selectivities. From a methodological point of view, the results presented here are very promising and show that DFT is accurate enough to study enantioselectivities in quite large models. However, more studies have
to be carried out in order to further assess the strengths and limitations of this approach, in particular in terms of reproducing stereoselectivities.
Beräkningskemiska metoder är mycket användbara för att studera enzymatiska reaktioner och kan bidra till en ökad förståelse för reaktionsmekanismer samt för uppkomsten av olika typer av selektivitet. I avhandlingen har tätthetsfunktionalteori använts för att studera fyra olika enzym, vilka samtliga har möjliga tillämpningar inom biokatalys.

De enzym som har undersökt är två epoxidhydrolas och två dekarboxylas. Detaljstudier av reaktionsmekanismerna har utförts och i tre av enzymen har även enantioselektiviteten undersöks. I samtliga studier har stora kvantkemiska modeller av de aktiva centren använts, vilket är särskilt viktigt vid beräkningen av enantioselektiviteter då modellerna måste kunna representera den kirala miljön hos respektive enzym.

Beräkningarna på epoxidhydrolasen visade god överensstämmelse mot experiment och enantioselektiviteten hos samtliga mutanter samt uppkomsten av enantiokonvergens kunde förklaras.

Från studierna av de två dekarboxylasen framgick det att reaktionsmekanismerna involverar bildandet av en endiolat- eller kinonmetidintermediär. I båda fallen visade sig beräkningarna till stor del stödja de tidigare föreslagna reaktionsmekanismerna, samt förklarade uppkomsten av enantioselektivitet i det enzym där detta undersöks.

Avhandlingen har bidragit till ytterligare kemisk insikt om de studerade enzymen. Resultaten visar även att användandet av kvantkemiska beräkningar är ett mycket bra verktyg för att undersöka asymmetriska enzymatiska reaktioner.
Acknowledgments

I would like to express my gratitude to the following people:

My supervisor Professor Fahmi Himo, for accepting me as a graduate student in your research group. Thank you for all your support, enthusiasm and patience.

Professor Pher Andersson, for showing interest in my research.

Professor Adrian Mulholland, for my time at the University of Bristol and for valuable comments on the epoxide hydrolase projects.

Professor Astrid Gräslund and Dr. Ana Popović-Bijelić, for the collaboration in the study of triapine.

Professor Per Siegbahn and Professor Margareta Blomberg, for generously sharing your knowledge and experience.

Professor Margareta Blomberg, Henrik Daver, Sara Moa, Mikael Nordin, Dr. Xiang Sheng and Maria Lagergren, for reading, discussing and improving this thesis.

NSC, PDC and Lunarc, for computing resources.

All past and present members of the FH-group (Liao, Mikael, Stefano, Oscar, Lynn, Bianca, Henrik, Sara, Jacob, Genping, Marcin, Karim, Andrés, Xiang). Stefano, for being my unofficial second supervisor and organic chemistry guru, and for teaching me everything there is to know about italian love songs. Jacob, for being more and less abazing. I miss your "word of today" lessons and all your crazy drawings on the whiteboard. Karim, for live music and skattkistor. Marcin, for kinetics lessons and for showing me the importance of having a pair of Finland-ferry glasses. Sara, for all support and for
sharing my interest in resolution, el diablos, super-slim lines, late night dancing at Calexico’s etc… You are the star of floor 5! Andrés, for throwing away that yellow jacket. Mikael, for having opinions about absolutely everything and not being afraid of sharing them. G-P, for all the times you had to listen to my problems, and for always betting against me. I still owe you two beers. Henrik, for always having the other opinion. (I just want to say that… IHT… TYWM for all solutions to the PEBCAK. WWIDWY?) Liao, for everything you have taught me since day one. Xiang, for being excellent in the chicken and at work! Thank you for the collaboration on PAD, I am sorry for my short memory. Bianca, for creating the best office atmosphere and for always making time when you have none!

Sara, Henrik and Anuja, for the cover picture ♠

All the people at the department.

The TA-staff.

Financial support from Knut & Alice Wallenbergs Stiftelse.

Friends from outside the University and my family ♡

Samuel, for being the kindest and most caring person I have ever met.
References


