High-Level Computational Methods: Application to Drug Design, Catalytic Atmospheric Reactions and Thermochemistry

Farzaneh Sarrami M.Sc.
I, Farzaneh Sarrami, certify that:

This thesis has been substantially accomplished during enrolment in the degree. The thesis comprises only my original work towards the PhD except where indicated in the preface. Due acknowledgement has been made in the text to all other materials used. This thesis does not contain any material previously published or written by another person, except where due references have been made in the text.

The work described in this thesis was funded by Australian Postgraduate Award (APA) and Australian Research Council (ARC) Discovery Early Career Researcher Award (Project Number: DE140100311).

This thesis contains published works and some of them have been co-authored.

Signature: [Blank]

Date: 9/5/2019
Abstract

High-level quantum chemical methods play an important role in our understanding of chemical mechanisms underlying catalytic/biochemistry reactions. These methods are particularly useful in cases where experimental data missing or the experimental error is large. This thesis focuses on applications of quantum chemical methods, such as density functional theory (DFT), double-hybrid DFT, and composite ab initio procedures, to solving catalytic reaction mechanisms relevant to biochemistry and atmospheric chemistry. In addition, I use composite ab initio procedures for designing bio-inspired antioxidants based on the endogenous antioxidant carnosine. In particular, I focus on catalytic processes involving proton transfers in biologically relevant species (e.g., carnosine and vitamin E) and atmospherically relevant species (e.g., glyoxal and Criegee intermediates). In some of these reaction mechanisms the proton transfers are coupled with the formation or breaking of covalent bonds between heavy atoms (e.g., C–O and C≡C). I find that these reactions can be efficiently catalysed by molecules that act as proton bridges (e.g., water and sulphuric acid). Finally, in the last chapter of the thesis I use high-level ab initio composite methods for investigating the thermochemical properties of a series of dicarboranes. I evaluated the performance of a range of lower cost Gn and CBS composite ab initio procedures. The combination of these approaches gives us the depth understanding reaction mechanism and thermochemical properties of the compounds using high-level quantum chemical methods.
Overview and Thesis Structure

This PhD thesis designed as a series of publications according to the UWA guidelines.

Quantum chemical methods have various applications ranging from anti-cancer drug development and discovery to alternative energy resource development and materials development. In this thesis, I applied these procedures, including density functional theory (DFT), *ab initio*, and composite methods for investigating a wide range of biochemistry and atmospheric chemistry phenomena. The biochemistry systems considered in this dissertation undergo intermolecular Br⁺ transfer in carnosine and ring opening rearrangement in vitamin E. The atmospheric chemistry investigation includes catalytic reaction mechanisms in important biogenic compounds in the atmosphere. In the last part of this computational study, I obtain accurate thermochemical properties for dicarborane isomers.

In chapter I, I give a brief overview of the theoretical methods most commonly used in this thesis. Namely, Hartree–Fock, density functional theory (DFT), double hydride DFT (DHDF), and composite ab initio methods and the description of functionals and basis sets associated with the use of these methods. After this brief overview of the theoretical methods, the topics covered in this thesis will be presented in four chapters.

In the first part of Chapter II, I explore a key step underlying the antioxidant activity of carnosine using the high-level G4(MP2) composite method. Namely, the intramolecular X⁺ transfer between the imidazole nitrogen and the terminal primary amine. Based on this reaction mechanism, I proceed to explore structure-function relationships for lowering the reaction barrier height for the intramolecular X⁺ shift in carnosine (X = Cl and Br) in order to design more effective antioxidants. In the last part of this chapter I review experimental and computational studies on first 'OH scavenging of α-tocopherol (the most biologically active form of the vitamin E) then I explore the second 'OH scavenging reaction mechanism following the rearrangement reaction using double-hybrid density functional theory (DHDF) method. Based on
these results, I design novel candidates for improving the antioxidant activity through reducing the size of the heterocyclic ring.

Chapter III describes briefly secondary organic aerosol (SOA) formation from the gas phase precursors in atmospheric systems. In the first project, I use the high-level ab initio G4(MP2) procedure to investigate the uncatalysed, water-catalysed and sulphuric acid-catalysed unimolecular 1,4 H-shift reactions in methyl Criegee intermediate (CI) and in two biogenic compound (isoprene and α-pinene derived CIs). In the second project, I use the same procedure to study the catalytic reaction mechanism of glyoxal with ethanol to form hemiacetal. I consider both the uncatalysed and catalysed reaction mechanisms. The considered catalysts are H$_2$O, HCOOH, and H$_2$SO$_4$.

In chapter IV, in this project, I use the high-level ab initio procedure to investigate and re-evaluates key thermochemical properties of dicarborane isomers. This project was done in collaboration with Dr. Li-Juan Yu.

Finally, chapter V presents an overview of the research and conclusions are drawn.
Chapter 1 General Introduction

1.1 Theory background

1.1.1 Schrödinger Equation
1.1.2 Born-Oppenheimer Approximation
1.1.3 Hartree-Fock Theory
1.1.4 Configuration Interaction (CI) Theory
1.1.5 Møller–Plesset perturbation Theory (MP)
1.1.6 Coupled Cluster Theory (CC)

1.2 Basis Sets

1.2.1 Slater and Gaussian Type Orbitals
1.2.2 Classification of Basis set
1.2.2.1 Pople style basis sets

1.3 Density Functional Theory (DFT)

1.3.1 Introduction of DFT methods in drug design
1.3.2 Introduction of DFT methods in catalytic reactions
1.3.3 Current DFT Methods
1.3.3.1 The Hohenberg-Kohn Theorems
1.3.3.2 The Kohn-Sham Theory

1.3.4 Various DFT Functionals

1.3.4.1 The Local Density Approximation (LDA)
1.3.4.2 Generalized Gradient Approximation (GGA)
1.3.4.3 Meta-GGA
1.3.4.4 Hybrid GGA (HGGGA)
1.3.4.5 Double Hybrid Density Functional Approximation (DHDFA)
1.3.5 Dispersion Correction
# Table of Contents

### 1.4 Composite Methods

1.4.1 Gaussian-4 Theory

1.4.1.1 G4MP2 Theory

1.4.2 Weizmann-n (Wn) Theories

### Chapter 2 Antioxidant design application

#### 2.1 Introduction

2.1.1 Introduction to the antioxidant design based on Carnosine

2.1.2 Introduction to the antioxidant design based on vitamin E

#### 2.2 Series of papers

2.2.1 Computational design of carnosine-based HOBr antioxidants

2.2.2 Mechanistic insights into the water-catalysed ring-opening reaction of vitamin E by means of double-hybrid density functional theory

### Chapter 3: Atmosphere chemistry application

#### 3.1 Introduction

3.1.1 Introduction to the sulphuric acid catalysed 1,4-hydrogen transfer reaction in Criegee intermediates

3.1.2 Introduction to the sulphuric acid-catalysed formation of hemiacetal from glyoxal and ethanol

#### 3.2 Series of papers

3.2.1 A computational investigation of the sulphuric acid catalysed 1,4-hydrogen transfer in higher Criegee intermediates

3.2.2 Sulphuric acid-catalysed formation of hemiacetal from glyoxal and ethanol

### Chapter 4: Thermochemistry of icosahedral closo-dicarboranes

#### 4.1 Introduction

4.1.1 Introduction to Thermochemistry of icosahedral closo-dicarboranes

4.1.2 Thermochemistry of icosahedral closo-dicarboranes: a composite ab initio quantum-chemical perspective

### Chapter 5: Summary and Conclusions

### Chapter 6 Appendices
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Appendix A</td>
<td>94</td>
</tr>
<tr>
<td>6.2 Appendix B</td>
<td>99</td>
</tr>
<tr>
<td>6.3 Appendix C</td>
<td>102</td>
</tr>
<tr>
<td>6.4 Appendix D</td>
<td>106</td>
</tr>
<tr>
<td>6.5 Appendix E</td>
<td>112</td>
</tr>
<tr>
<td>6.6 Appendix F</td>
<td>119</td>
</tr>
<tr>
<td>6.7 Appendix G</td>
<td>121</td>
</tr>
<tr>
<td>6.8 Appendix H</td>
<td>123</td>
</tr>
<tr>
<td>6.9 Appendix I</td>
<td>125</td>
</tr>
<tr>
<td>6.10 Appendix J</td>
<td>127</td>
</tr>
<tr>
<td>6.11 Appendix K</td>
<td>129</td>
</tr>
<tr>
<td>References</td>
<td>131</td>
</tr>
</tbody>
</table>
Firstly, I would like to express my sincere gratitude to my supervisor Assoc/Prof. Amir Karton for the continuous support of my PhD study and related research, for his patience, motivation, and encouragement. His guidance helped me in the past three and half years of research and writing of this thesis. I could not have a better advisor and mentor for my PhD study.

I would like to thank my co-supervisor Assoc/Prof Keith Stubbs for his support and guidance. I also need to thank Prof Dylan Jayatilaka and Assoc/ Prof Graham Chandler for their provoking dissections I have had about computational chemistry and my research projects.

My sincere thanks also goes to the following scientific collaborators for their invaluable contribution Dr Dino Spagnoli (UWA), Dr Heng Chooei (UWA), and Prof Masoud Salavati-Niasari (University of Kashan), Dr Ali Salehabadi (University of Kashan), Assoc/Prof Mohsen Shahlaei (University of Kermanshah), who provided me an opportunity to join their team as intern, and who gave access to the research facilities. Without they precious support it would not be possible to conduct this research.

I would like to thank computing resources from the National Computational Infrastructure (NCI), which is supported by the Australian Government. The high performance computing (HPC) system at NCI, which is named Raijin currently comprises 84,656 cores (Intel Xeon Sandy Bridge 2.6 GHz, Broadwell 2.6 GHz) in 4416 compute nodes, 300 Terabytes of main memory and 8 Petabytes of high-performance operational storage capacity. We also acknowledge the system administration support provided by the Faculty of Science at the University of Western Australia (UWA) to the Linux cluster of the Karton group. The HPC cluster of Karton group is designed for carrying out large memory intensive ab initio calculations. The cluster comprises of 396 cores in the compute nodes; ~5 TB of main memory; and ~86 TB of local scratch disk. We gratefully acknowledge the provision of an Australian Postgraduate Award (APA) and Australian Government Research Training Program (RTP) for financial support for my research.
Acknowledgements

I thank my fellow lab mates in for the incredible discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last three and half years.

Last but not the least, I would like to thank my family: especially my beloved daughter ARISA and my husband, supporting me spiritually during my PhD and writing this thesis. Words cannot express how grateful I am to my parents for all of their support so far and I could not have achieved without them. Your prayer for me was what sustained me thus far.
# Authorship Declaration: Co-authored Publications

Publications included in this thesis

   
   Location in the thesis: Chapter 4
   
   Overall contribution by Sarrami: 75%

   
   Location in the thesis: Chapter 3 Section
   
   Overall contribution by Sarrami: 75%

   
   Location in the thesis: Chapter 2 Section
   
   Overall contribution by Sarrami: 75%

   
   Location in the thesis: Chapter 3 Section
   
   Overall contribution by Sarrami: 75%

   
   Location in the thesis: Chapter 2 Section
   
   Overall contribution by Sarrami: 75%
I, Amir Karton certify that the student statements regarding their contribution to each of the works listed above are correct.

Date: 9/5/2019

Student signature: 

Date: 9/5/2019

Coordinating supervisor signature: 

Date: 9/5/2019
Other Co-authored Publications During the PhD Candidature

Only the first pages of these publications are included in this thesis

   Location in the thesis: Chapter 6 Section

   Location in the thesis: Chapter 6 Section

   Location in the thesis: Chapter 6 Section

   Location in the thesis: Chapter 6 Section

   Location in the thesis: Chapter 6 Section

Location in the thesis: Chapter 6 Section
List of Figures

Figure 1.1 HF-SCF approach.

Figure 1.2 Examples of single (S), double (D), triple (T), and quadruple (Q) excited determinant generated from an HF reference.

Figure 1.3 Jacob’s ladder of DFT functionals.

Figure 2.1 Schematic representation of oxidative stress.

Figure 2.2 Schematic representation of the intramolecular rearrangements involved in the X⁺ (X = Cl and Br) transfer from the imidazole nitrogen (N₂) to the terminal amino nitrogen (N₃).

Figure 2.3 The reaction mechanism for the antioxidant activity of α-tocopherol (AT*), through the reaction with two hydroxyl radicals. The ring-opening reaction leading to ATQ is the subject of the present work and shown in a dashed circle.

Figure 3.1 Formation of vinylhydroperoxides (VHPs) via ozonolysis followed by a 1,4 H-shift. Attack of ozone on the C=C double bond leads to a primary ozonide (POZ), which rapidly decomposes, yielding Criegee Intermediates (CIs), then a 1,4 H-shift from the syn methyl group leads to the formation of VHPs.

Figure 4.1 ortho (1), meta (2) and (3) para-Carborane.
General Symbols and Abbreviations

$\mathcal{H}$ Hamiltonium operator
$\Psi$ Wavefunction
$E$ Energy
$\hat{T}_e$ Kinetic energy operator for electrons
$\hat{T}_n$ Kinetic energy operator for nuclei
$\hat{V}_{ee}$ Electron-electron electrostatic interaction operator
$\hat{V}_{en}$ Electron-nuclei electrostatic interaction operator
$\hat{V}_{nn}$ Nuclei-nuclei electrostatic interaction operator
$\nabla$ Laplacian
$Z$ Nuclear charge
$M$ Nuclear mass
$BO$ Born-Oppenheimer
$r$ Position vector for electrons
$R$ Position vector for nuclei
$\psi_e$ Electronic wavefunction
$\psi_n$ Nuclear wavefunction
$PES$ Potential energy surface
$GTO$ Gaussian type orbitals
$STO$ Slater type orbitals
$MO$ Molecular orbital
$MAD$ Mean absolute deviation
$MSD$ Mean signed deviation
$RMSD$ Root mean square deviation
$SPE$ Single point energy
$ZPE$ Zero-point energy
$VB$ Valence band
$M_{basis}$ Number of basis functions
$HF$ Hartree-Fock
$SCF$ Self consistent field
$CI$ Configuration interaction
$CIS$ Configuration interaction with single excitations
$CISD$ Configuration interaction with single and double excitations
$CISDT$ Configuration interaction with single, double and triples excitations
$FCI$ Full configuration interaction
$MBPT$ Many body Peturbation theory
$MP_n$ $n^{th}$ order Møller-Plesset Peturbation theory
$CCSD$ Coupled cluster with singles and double excitations
$CCSDT$ Coupled cluster with single, double and triples excitations
$CCSD(T)$ Coupled cluster with single, double and Peturbation triples
$DFT$ Density functional theory
$LDA$ Local density approximation
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSDA</td>
<td>Local spin density approximation</td>
</tr>
<tr>
<td>GGA</td>
<td>Generalized gradient approximation</td>
</tr>
<tr>
<td>MGGA</td>
<td>Meta-GGA</td>
</tr>
<tr>
<td>HGGA</td>
<td>Hybrid-GGA</td>
</tr>
<tr>
<td>HMGGA</td>
<td>Hybrid meta-GGA</td>
</tr>
<tr>
<td>RS</td>
<td>Range separated</td>
</tr>
<tr>
<td>DH</td>
<td>Double hybrid</td>
</tr>
<tr>
<td>$G_n$</td>
<td>Gaussian-$n$ theories</td>
</tr>
<tr>
<td>CBS</td>
<td>Complete basis set</td>
</tr>
<tr>
<td>$W_n$</td>
<td>Weizmann-$n$ theories</td>
</tr>
<tr>
<td>ccCA</td>
<td>Correlation consistent composite approach</td>
</tr>
<tr>
<td>HEAT</td>
<td>Heat accuracy extrapolated \textit{ab initio} thermochemistry</td>
</tr>
</tbody>
</table>
1.1 Theory background

1.1.1 Schrödinger Equation

The time-independent Schrödinger equation plays a central role in computational chemistry and physics for computing properties of molecules such as optimal geometries, electronic, vibrational and rotational energy levels and reaction barriers. The Schrödinger equation is:

\[ \hat{H} \Psi = E \Psi \]  \hspace{1cm} (1.1)

Where \( \hat{H} \) is the Hamiltonian operator, which operates on the wavefunction \( \Psi \), and provides the energy \( E \) for a given system. \( \hat{H} \) can be written as the sum of the kinetic and potential energy operator of the isolated system of \( N \) electrons:

\[ \hat{H} = \hat{T}_e + \hat{T}_N + \hat{V}_{ee} + \hat{V}_{eN} + \hat{V}_{NN} \]  \hspace{1cm} (1.2)

Where \( \hat{T}_e \) is the electronic kinetic energy operator, \( \hat{T}_N \) is the nuclear kinetic energy operator, \( \hat{V}_{ee} \) is the potential energy of the electron-electron, \( \hat{V}_{eN} \) is the nuclear-electron and \( \hat{V}_{NN} \) is the nuclear-nuclear interactions. These operations can be written as:

\[ \hat{T}_e = - \sum_i^N \frac{1}{2} \nabla_i^2 \]  \hspace{1cm} (1.3)

\[ \hat{T}_N = - \sum_i^M \frac{1}{2M} \nabla_i^2 \]  \hspace{1cm} (1.4)

\[ \hat{V}_{ee} = \sum_{i>j}^N \frac{1}{r_{ij}} \]  \hspace{1cm} (1.5)

\[ \hat{V}_{eN} = - \sum_i^N \sum_A^M \frac{Z_A}{r_{Ai}} \]  \hspace{1cm} (1.6)
\[ \nabla_{NN} = -\sum_{B>A}^{N} \frac{Z_A^2 e^2}{r_{AB}} \]  
\hspace{1cm} (1.7)

Where \( \nabla \) is the Laplacian, \( i \) and \( j \) denote electrons, \( A \) and \( B \) represent the nuclei with \( Z \) charge, \( M_A \) is the mass of the nucleus \( A \), and \( r \) is the distance between two particles. Apart from hydrogen atoms and hydrogen-like atoms, Schrödinger equations of other atomic or molecular systems, which involve three or more ‘particles’, cannot be solved analytically. In order to solve the equations for these systems, namely known as ‘many-body systems’, an approximation must be made.

### 1.1.2 Born-Oppenheimer Approximation

An essential part of solving the Schrödinger equation is the Born–Oppenheimer approximation.\(^1\) In this approximation, the coupling between the nuclei and electronic motion is neglected by assuming that the nuclei are infinitely heavier than the electrons. This allows the electronic part to be solved with the nuclear positions as parameters. The approximation also allows for the separation of electronic and nuclear motion. This simplifies the Hamiltonian operator since \( \hat{T}_N \) is neglected and \( \nabla_{NN} \) becomes a constant:

\[ \hat{H}_e = \hat{T}_e + \nabla_{ee} + \nabla_{eN} + \text{constant} \]  
\hspace{1cm} (1.8)

The solution describing a system of atoms would be a function of the form \( \Psi(R,r) \) where the symbol \( r \) is used to collectively label the co-ordinates of all the electrons \( (r_1, r_2, r_3, \ldots) \) and \( R \) to label the co-ordinates of the nuclei \( (R_1, R_2, R_3, \ldots) \). The electronic and nuclear Schrödinger equations are written as:

\[ \hat{H}_{tot}(R,r)\Psi_{tot}(R,r) = E_{tot}\Psi_{tot}(R,r) \]  
\hspace{1cm} (1.9)

\[ \hat{H}_e(R,r)\Psi(R,r) = E_e\Psi(R,r) \]  
\hspace{1cm} (1.10)

\[ [\hat{T}_n + E_e(R)]\psi(R) = E\psi(R) \]  
\hspace{1cm} (1.11)

Based on this approximation, the nuclei are stationary from an electronic point of view.
Chapter 1: General Introduction

The electronic wave function thus depends parametrically on the nuclear coordinates, and therefore the electronic energy eigenvalue $E_e$ depends on the chosen position $R$ of the nuclei. Solving the electronic Schrödinger equation, by varying these positions $R$ in a series of small steps gives $E_e$, which is known as the potential energy surface (PES).

The dynamics of a many-electron system is very complex, and consequently requires elaborate computational methods. A significant simplification can be obtained by introducing independent-particle models, where the motion of one electron is considered to be independent of the dynamics of all other electrons. In an independent-particle model the interactions between the particles is approximated, either by neglecting all but the most important one, or by taking all interactions into account in an average function. This approximation has an acceptable accuracy, and is called the Hartree–Fock (HF) method.

### 1.1.3 Hartree-Fock Theory

The Hartree-Fock (HF) method was developed by Hartree $^2$ and Fock $^4$ soon after the introduction of quantum mechanics. Briefly, its aim is to find a set of molecular orbitals (MOs) that minimize the energy of a wave function $\Psi$ (a Slater determinant). The HF method finally considers the ‘mean field’ approximation, which means that each electron does not experience other individual electrons, but a mean field from their charges. Thus, the HF method reduces an $N$-electron problem to $N$ one-electron problems. The solution of the HF equations is not trivial as the Fock operator is a function of the molecular orbitals themselves, which correspond to the solutions of the eigenvalues equations. The algorithm shown in Figure 1.1, used to solve the HF equations, is known as the Self Consistent Field (SCF) procedure.$^6$ According to this algorithm, an initial guess for the Fock matrix is built using the molecular orbitals obtained at a lower level of theory (for instance Extended Huckel $^5$ or semiempirical methods, not discussed here).
Chapter 1: General Introduction

The main assumptions of the HF-SCF approach are:

1) The Born-Oppenheimer approximation, which assumes the true wavefunction, is a function of the coordinates of each of the nuclei.
2) The momentum operator, which is assumed to be completely classical.
3) The representation of molecular orbitals, which is assumed to be a finite number of one-electron wavefunctions (molecular orbitals) which would need a complete (infinite) representation.
4) The energy eigenfunctions (wavefunctions) which are assumed to be products of one-electron functions (Slater determinants). The effects of electron correlation, beyond that of exchange energy resulting from the anti-symmetry of the wavefunction, are neglected.6

The HF method uses an effective potential energy term that accounts for an electron’s interaction with the averaged electron density of the remaining electrons. The HF
wavefunction includes the effective potential that accounts for correlation between electrons of the same spin, but neglects the correlation from electrons of opposite spin, known as the electron correlation \( E_{\text{corr}} \), which is defined as the difference between the exact energy and the HF energy.

\[
E_{\text{corr}} = E_{\text{exact}} - E_{\text{HF}}
\] (1.12)

Electron correlation is the phenomenon of the motion of pairs of electrons in atoms or molecules being correlated. The purpose of post-Hartree-Fock calculations (correlated calculations) is to treat such correlated motion in a better way. In the Hartree-Fock treatment, electron electron repulsion is handled by having each electron move in a smeared-out, average electrostatic field due to all the other electrons and the probability that an electron will have a particular set of spatial coordinates at some moment is independent of the coordinates of the other electrons.

The HF method generally recovers 99% of the energy of a system, and the remaining 1% of the total energy is the electron correlation. The electron correlation energy is necessary for an accurate description of chemical properties, as it leads to large deviations with respect to experimental results. A number of approaches, usually denoted as post-HF methods, have been devised to include electron correlation in multi-electron wavefunctions. One of these approaches, known as the Configuration Interaction (CI), expands the multi-electron wavefunction using a linear combination of Slater determinants, by requiring that the energy be at a minimum known as the Configuration Interaction (CI).

### 1.1.4 Configuration Interaction (CI) Theory

The Configuration Interaction (CI) theory is similar to the HF theory. In order to recover electron correlation energy, the interaction of the HF determinant and excited determinants (replacing one or more occupied orbitals within the HF determinant by virtual orbitals outside the HF determinant) must be included. Therefore, a linear combination of N-electron configurations, and optimise the coefficients of the different configurations, using the linear variational method. In addition to the HF determinant, all excited configurations from (occupied to virtual orbitals) or
determinants (Figure 1.2) must be included to achieve the exact solution to the Schrödinger equation. Figure 1.2 shows the single (S), double (D), triple (T), and quadruple (Q) excited determinants that are generated from the HF ground determinant.

\[ \Psi_{FCI} = a_0 \Psi_{HF} + \sum_i a^a_i \Psi_i^a + \sum_{ij} a_i^{ab} \Psi_{ij}^{ab} + \sum_{ijk} a_{ijk}^{abc} \Psi_{ijk}^{abc} + \cdots \]  

(1.13)

Where \( i, j, k \) are occupied orbitals, \( a, b, c \) are virtual or unoccupied molecular orbitals (MOs), and \( \Psi_i^a, \Psi_{ij}^{ab} \) and \( \Psi_{ijk}^{abc} \) are possible determinants found by performing single, double and triple excitations from the reference determinant (\( \Psi_{HF} \)), respectively. While the full CI is able to achieve the exact solution to the Schrödinger equation, the calculation of electronic energy is limited by the binomial computational scaling with respect to the size of the system.\(^6\)

As the number of electrons grow, the basis set calculations become more complex due to the huge number of configurations. This limits the application of FCI
Chapter 1: General Introduction

method to all but small systems, such as small molecules (a few light atoms) with a relatively small basis set. Therefore, a truncation of the CI expansion is needed to lower the cost of the calculations. Truncating the excitation level at one (CI with singles (CIS)) does not improve the HF result, as it is equal to HF for the ground state energy. However, the CI approach does include single and the double excited configurations (CISD) and does recover between 82-90% of the correlation energy for molecules built from atoms of the first row. The only CI method that is generally applicable for a large variety of systems is CISD. For computationally feasible systems (i.e. medium-size molecules and basis sets), CISD typically recovers 80–90% of the available correlation energy.6

1.1.5 Møller–Plesset perturbation Theory (MP)

The Møller-Plesset perturbation theory (MP), first was published in 1934,12 is one of the most popular quantum chemistry post-HF methods. This perturbation theory states the approximate solution differs from the exact solution by one small correction term. The electronic Hamiltonian can be written as:

\[ \hat{H} = \hat{H}_0 + \lambda \hat{H}' \]  

(1.14)

Where \( \hat{H}' \) is the perturbation of the reference \( \hat{H}_0 \) multiplied by a variable parameter \( \lambda \). Perturbation methods can be used in quantum mechanics for adding corrections to solutions that employ an independent-particle approximation, and the theoretical framework is then called the Many-Body Perturbation Theory (MBPT).6

Let us assume that the Schrödinger equation for the reference Hamiltonian operator is solved as:

\[ H_0 \Phi_i = E_i \Phi_i \quad i = 0, 1, 2, \ldots, \infty \]  

(1.15)

Where \( i \) is a variable parameter, determining the strength of the perturbation. The solutions for the unperturbed Hamiltonian operator form a complete set (since \( H_0 \) is Hermitian) which can be chosen to be orthonormal. Notably, only cases where the perturbation is time-independent are considered and the reference wavefunction is non-
Chapter 1: General Introduction

degenerate. The perturbed Schrödinger equation is:

\[ \hat{H}\Psi = W\Psi \]  \hspace{1cm} (1.16)

The new energy and wavefunction can then be written as a Taylor expansion:

\[ W = \lambda^0 W_0 + \lambda^1 W_1 + \lambda^2 W_2 + \lambda^3 W_3 + \cdots \]  \hspace{1cm} (1.17)

\[ \Psi = \lambda^0 \Psi_0 + \lambda^1 \Psi_1 + \lambda^2 \Psi_2 + \lambda^3 \Psi_3 + \cdots \]  \hspace{1cm} (1.18)

For \( \lambda = 0 \), \( \Psi_0 = \Phi_0 \) and \( W_0 = E_0 \), and this is the unperturbed, or zeroth-order wave function and energy. The \( \Psi_1, \Psi_2, \ldots \) and \( W_1, W_2, \ldots \) are the first-order, second-order, etc., corrections. The perturbation parameter \( \lambda \) will eventually be set to \( \lambda \), and the \( n \)th-order energy or wave function becomes a sum of all terms up to order \( n \). With the expansions (Equation (1.17) and (1.18)), the Schrödinger equation (1.16) becomes:

\[ (\hat{H}_0 + \lambda \hat{R}')(\lambda^0 \Psi_0 + \lambda^1 \Psi_1 + \lambda^2 \Psi_2 + \lambda^3 \Psi_3 + \cdots) = (\lambda^0 W_0 + \lambda^1 W_1 + \lambda^2 W_2 + \lambda^3 W_3 + \cdots) (\lambda^0 \Psi_0 + \lambda^1 \Psi_1 + \lambda^2 \Psi_2 + \lambda^3 \Psi_3 + \cdots) \]  \hspace{1cm} (1.19)

Since this holds for any value of \( \lambda \), the terms can be collected terms with the same power of \( \lambda \) to give:

\[ \begin{align*}
\lambda^0: \hat{H}_0 \Psi_0 &= W_0 \Psi_0 \\
\lambda^1: \hat{H}_0 \Psi_1 + \hat{R}' \Psi_0 &= W_0 \Psi_1 + W_1 \Psi_0 \\
\lambda^2: \hat{H}_0 \Psi_2 + \hat{R}' \Psi_1 &= W_0 \Psi_2 + W_1 \Psi_1 + W_2 \Psi_0 \\
& \vdots \\
\lambda^n: \hat{H}_0 \Psi_n + \hat{R}' \Psi_{n-1} &= \sum_{i=0}^{n} W_i \Psi_{n-i}
\end{align*} \hspace{1cm} (1.20)

The first-order equation contains two unknowns: the first-order correction to the energy, \( W_1 \), and the first-order correction to the wave function, \( \Psi_1 \). The \( n \)th-order energy correction can be calculated by multiplying from the left by \( \Phi_0 \) and integrating, and using the “turnover rule” \( \langle \Phi_0 | \hat{R}_0 | \Psi_i \rangle = \langle \Psi_i | \hat{R}_0 | \Phi_0 \rangle^* \).

So far, the MP theory has been described generally. However if the perturbation theory is to be applied to the calculation of correlation energy, the unperturbed Hamiltonian operator must be used. The most common choice is to take this as a sum
over Fock operators, leading to Møller–Plesset (MP) perturbation theory. The second-order Møller–Plesset correction is:

\[ E(\text{MP2}) = \sum_{\text{occ}} \sum_{\text{vir}} \frac{\langle \phi_i | \phi_j | \phi_a | \phi_b \rangle - \langle \phi_i | \phi_j | \phi_b | \phi_a \rangle}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \]  

(1.21)

Where \( \varepsilon \) is the orbital energy for \( i \) and \( j \) occupied orbitals, and \( a \) and \( b \) represent the virtual orbitals. Once the two-electron integrals over MOs are known, the second-order energy correction can be calculated as a sum over such integrals. There are of the fourth-order of basis functions (\( M^4_{\text{basis}} \)) integrals, thus the calculation of the energy increases as the order with the system size.

MP2 is an \( M^5_{\text{basis}} \) method, but fairly inexpensive as not all two electron integrals over MOs are required. Only those corresponding to the combination of two occupied and two virtual MOs are needed. In practical terms, this means that the MP2 energy for systems with a few hundred basis functions can be calculated at a cost similar to, or less than, what is required for calculating the HF energy. MP2 typically accounts for 80–90% of the correlation energy, and it is the most economical method for including electron correlation. The MP2 method is often used to treat systems where electronic correlation effects play a key role. For instance, weak interactions like dispersion (London) forces or \( \pi \)-stacking interactions between aromatic compounds.

MP3 describes the interaction between electron pairs as well. The formula for calculating the contribution from third-order correction involves a computational effort that formally increases as \( M^6_{\text{basis}} \). This correction energy accounts for about 90–95% of the correction energy. The computational cost of the fourth-order correction energy without the contribution from triple excited determinants. MP4(SDQ), increases also as \( M^6_{\text{basis}} \), while the triple (T) contribution increases as \( M^7_{\text{basis}} \). Generally, the triples contribution to MP4 takes roughly the same amount of time as the SDQ contributions, but the triples are the most important contribution at forth order. The full forth-order energy typically accounts for around 95–98% of the correlation energy.
1.1.6 Coupled Cluster Theory (CC)

More popular variants of the CISD type methods are the coupled cluster (CC) family of methods (CCSD, CCSDT, CCSDTQ, etc.) in which excitations are generated via an exponential excitation operator ansatz. Coupled-cluster (CC) theory was introduced in the late 1950s by Coester and Kümmel and because of relatively-soft computational scaling, has since been used to extend \( \text{ab initio} \) computations into the medium-mass region of the nuclide chart. Coupled-cluster methods usually give more accurate energies than their CI counterparts without a substantial increase in the runtime.

Perturbation methods add all types of corrections (S,D,T,Q, etc.) to the reference wave function to a given order (2, 3, 4, etc.). In Coupled Cluster (CC) methods, all corrections of a given type are included to infinite order. Introducing the electron correlation through the use of a cluster operator \( \mathcal{T} \), gives:

\[
\mathcal{T} = \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3 + \mathcal{T}_4 + \cdots + \mathcal{T}_n
\]  

(1.22)

The \( \mathcal{T}_i \) operator acting on an HF reference wave function \( (\Phi_0) \) generates all \( i \)th excited Slater determinants:

\[
T_1 \Phi_0 = \sum_i^{\text{oocc}} \sum_{a}^{\text{vir}} T_i^a \Phi_i^a
\]  

(1.23)

\[
T_2 \Phi_0 = \sum_{i<j}^{\text{oocc}} \sum_{a<b}^{\text{virir}} T_{ij}^{ab} \Phi_{ij}^{ab}
\]  

(1.24)

In CC theory, term amplitudes are used for the expansion coefficients. Using intermediate normalization, a CI wave function can be generated by allowing the excitation operator to work on an HF wave function:

\[
\Psi_{CI} = (1 + \mathcal{T}) \Phi_i = (1 + \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3 + \mathcal{T}_4 + \cdots) \Phi_0
\]  

(1.25)

The corresponding coupled cluster wave function is defined as:
\( \Psi_{CI} = e^T \Phi_0 \)  \hspace{1cm} (1.26)

\[ e^T = 1 + \hat{\mathcal{T}} + \frac{1}{2} \hat{\mathcal{T}}^2 + \frac{1}{3} \hat{\mathcal{T}}^3 + \cdots + \sum_{k=0}^{\infty} \frac{1}{k!} \hat{\mathcal{T}}^k \] \hspace{1cm} (1.27)

From Eqs. (1.22) and (1.27), the exponential operator may be written as:

\[
e^T = 1 + \hat{T}_1 + \left( \hat{T}_2 + \frac{1}{2} \hat{T}_1^2 \right) + \left( \hat{T}_3 + \hat{T}_2 \hat{T}_1 + \frac{1}{6} \hat{T}_1^3 \right) + \left( \hat{T}_4 + \hat{T}_3 \hat{T}_1 + \frac{1}{2} \hat{T}_1^2 + \frac{1}{2} \hat{T}_1^2 \hat{T}_2 + \frac{1}{24} \hat{T}_1^4 \right) + \cdots \] \hspace{1cm} (1.28)

The first term generates the reference HF and the second term generates all singly excited states. The first parenthesis \( \left( \hat{T}_2 + \frac{1}{2} \hat{T}_1^2 \right) \) generates all doubly excited states. The second parenthesis \( \left( \hat{T}_3 + \hat{T}_2 \hat{T}_1 + \frac{1}{6} \hat{T}_1^3 \right) \) generates all triply excited states. The quadruple excited states can be viewed as composed of five terms, a true quadruple \( \hat{T}_4 \) and four product terms \( \left( \hat{T}_3 \hat{T}_1 + \frac{1}{2} \hat{T}_1^2 + \frac{1}{2} \hat{T}_1^2 \hat{T}_2 + \frac{1}{24} \hat{T}_1^4 \right) \). By comparison with the CI wave function in Eq. (1.25), the CC wave function at each excitation level contains additional terms arising from products of excitations.

With the CC wave function in Eq. (1.26) the Schrödinger equation becomes:

\[ \hat{H} e^T \Phi_0 = E e^T \Phi_0 \] \hspace{1cm} (1.29)

Multiplying from the left by \( \Phi_0^* \) and integrating gives Eq. (1.30).

\[ \langle \Phi_0 | \hat{H} | e^T \Phi_0 \rangle = E \langle \Phi_0 | e^T \Phi_0 \rangle \] \hspace{1cm} (1.30)

\[ e^T \Phi_0 = (1 + \hat{\mathcal{T}} + \cdots) \Phi_0 = \Phi_0 + \hat{T} \Phi_0 + \frac{1}{2!} \hat{T}^2 \Phi_0 + \cdots \] \hspace{1cm} (1.31)

Since \( \langle \Phi_0 | e^T \Phi_0 \rangle = 1 \), Eq. (1.30) becomes:

\[ \langle \Phi_0 | \hat{H} | e^T \Phi_0 \rangle = E \] \hspace{1cm} (1.32)

If all cluster operators up to \( \hat{T}_n \) are included in \( \hat{T} \), all possible excited determinants are generated and the CC wave function is equivalent to the full CI, which is impossible for
all but the smallest systems. When the $\mathcal{T}$ operator is truncated, some of the terms in the amplitude equations will become zero, and the amplitudes derived from these approximate equations will no longer be exact. How severe the approximation is depends on the number of terms are included in $\mathcal{T}$. Including only the $\mathcal{T}_1$ operator does not give any improvement over HF, as matrix elements between the HF and singly excited states are zero. The lowest level of approximation is therefore $\mathcal{T} = \mathcal{T}_2$, referred to as Coupled Cluster Doubles (CCD). Compared with the number of doubles, there are relatively few singly excited states. Using $\mathcal{T} = \mathcal{T}_1 + \mathcal{T}_2$ gives the Coupled Cluster Singles and Doubles (CCSD), which is only slightly more demanding than CCD, and is a more complete model. Both CCD and CCSD involve a computational effort that scales as $M_{\text{basis}}^6$ in the limit of a large basis set. The next higher level has $\mathcal{T} = \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3$, giving the CCSDT model. This involves a computational effort that scales as $M_{\text{basis}}^8$. A widely–used CC method is CC with single, double, and quasi-perturbative triple excitations (CCSD(T)). Known as 'the gold standard' of quantum chemistry, CCSD(T) and is often used to calculate reference values for benchmarking simpler methods due to its ability to account for over 99% of the correlation energy. Recent developments in local coupled-cluster methods allow the treatment of molecules in the regime of about 100 atoms. Given a sufficiently large basis set, however, the CCSD(T) method is able to meet the goal of a chemical accuracy of \(\sim 4 \text{ kJ/mol} \) (\(\sim 1 \text{ kcal/mol}\)) for most systems.

### 1.2 Basis Sets

#### 1.2.1 Slater and Gaussian Type Orbitals

Since an analytical expression for eigenfunctions of Eq. (1.2) will never be found, it makes sense to look for approximate solutions that are linear combinations of atomic orbitals. In these solutions, the spatial orbitals in the Slater determinants are constructed from what is known of the orbitals from the isolated hydrogen atom, by expressing them as linear combinations of atomic eigenfunctions. Although an infinite series would be required for an exact treatment, the basis set approximation allows for
construct finite combinations for each spatial orbital. Basis functions differ in the way they account for the radial term. Two types of basis functions are Slater type orbitals (STOs) and Gaussian type orbitals (GTOs) \(^{28}\), which have the following functional form:

\[
\chi_{\ell,n,l,m} \left( r, \theta, \varphi \right) = N Y_{\ell,m} (\theta, \varphi) r^{n-1} e^{-\zeta r}
\]  \hspace{1cm} (1.33)

Where \( N \) is a normalization constant, \( Y_{\ell,m} \) are spherical harmonic functions which depend on the angular momentum of \( l \) and \( m \) and \( n \) is the principle quantum number. The exponent of STO basis functions is often denoted by the Greek letter \( \zeta \). The exponential dependence on the distance between the nucleus and electron mirrors the exact orbitals for the hydrogen atom. STOs are primarily used for atomic and diatomic systems where high accuracy is required. They can also be used with density functional methods that do not include exact exchanges and where the Coulomb energy is calculated by fitting the density into a set of auxiliary functions. Gaussian type orbitals (GTOs) can be written in terms of polar or Cartesian coordinates \(^{29}\) as shown by:

\[
\chi_{\ell,n,l,m} \left( r, \theta, \varphi \right) = N Y_{\ell,m} (\theta, \varphi) r^{2n-2-l} e^{-\zeta r^2}
\]  \hspace{1cm} (1.34)

\[
\chi_{\ell,l_x,l_y,l_z} \left( x, y, z \right) = N x^{l_x} y^{l_y} z^{l_z} e^{-\zeta r^2}
\]  \hspace{1cm} (1.35)

The sum of \( l_x, l_y \) and \( l_z \) determines the type of orbital (for example \( l_x + l_y + l_z = 1 \) is a p-orbital). In terms of computational efficiency, GTOs are therefore preferred and are used almost universally as basis functions in electronic structure calculations. However, more GTOs are required to achieve certain accuracy compared with STOs.

### 1.2.2 Classification of Basis set

The most important factor to consider in the choosing STO or GTO functions is the number of functions to be used. The smallest number of functions possible is called the minimum basis set, which includes a single for hydrogen or helium. For the first row elements of the periodic table, this basis set requires two \( s \)-functions \( 1s, 2s \) and one set
of \( p \)-functions (2\( p_x \), 2\( p_y \), 2\( p_z \)). Lithium and beryllium only require two \( s \)-functions, but also a set of \( p \)-functions.

The next improvement of the basis set is a doubling of all basis functions to produce a *double zeta* (DZ) type basis. The term ‘zeta’ stems from the exponent of the STO basis functions, which is often denoted by the Greek letter \( \zeta \). For the elements on the first row, a DZ basis set employs two \( s \)-functions for hydrogen (1\( s \) and 1\( s' \)). Doubling the number of basis functions thus allows for a much better description of different electron distribution in different directions. The basis set can be extended to *triple zeta* (TZ), with three times as many functions as the basis, also *quadruple zeta* (QZ), *quintuple zeta* (5Z) and so on. However, large basis sets are often given explicitly in terms of the number of basis functions of each type.

So far, only the number of \( s \)-functions and \( p \)-functions for each atom has been discussed. In most cases, higher angular momentum functions as *polarization functions* are also important. If methods including electron correlation are used, these polarisation functions are essential. These functions are added to the chosen \( sp \)-basis. Adding a single set of polarization functions (\( s \)-functions on hydrogens and \( d \)-functions on heavy atoms) to the DZ basis forms a *double zeta plus polarization* (DZP) type basis. Similar to the \( sp \)-basis sets, multiple sets of *polarization functions* with different exponents may also be added. If two sets of polarization functions are added to a TZ \( sp \)-basis, a *triple zeta plus Polarization* (TZ2P) type basis is obtained. For larger basis sets with many polarization functions the explicit composition in terms of number and type of function, is usually given.

Usually, too much effort is required to calculate a DZ for every orbital. Instead, the calculation is simplified by calculating a DZ only for the valence orbital. Since the inner-shell electrons are not vital to the calculation, they are described with a single Slater Orbital, which is called *split-valence basis set* method. The \( n - ijG \) or \( n - ijkG \) split-valence basis sets were developed by Pople and co-workers,\(^{30,31,32}\) where \( n \) is the number of primitives summed to describe the inner shells and \( ij \) or \( ijk \) is the number of primitives for contractions in the valence shell. In the 6-31G basis, for example, the core orbitals are described by a contraction of six GTOs, whereas the inner part of the valence orbitals is a contraction of three GTOs and the outer part of the valence is represented by one GTO.
Including polarization functions give more angular freedom so that the basis is able to represent bond angles more accurately, especially in strained ring molecules. For example for the 6-31G** basis set, the first asterisk indicates the addition of $d$ functions on the non-hydrogen atoms and the second asterisk indicates a $p$ function on the hydrogen atom. Other functions that can be added to the basis set for a better description of the wavefunction far from the nucleus are the diffuse functions, which are additional GTOs with small exponents. The diffuse functions are usually indicated with a "+". For example, in the 6-31++G basis set, the "++" indicates the addition of a set of $s$ and $p$ functions to the heavy atoms, with an additional $s$ diffuse GTO for hydrogen.

For correlated calculations, the basis set requirements are more demanding since the polarizations of the charge distribution must be described and an orbital space suitable for recovering correlation effects must also be provided. The correlation consistent basis sets are very suited to this purpose and usually denoted as cc-pVXZ, where the “cc” represents the correlation-consistent basis, and indicates the functions were optimized for best performance with correlated calculations. The “p” indicates that polarization functions are included on all atoms. The “VXZ” stands for valence with the cardinal number $X = D, T, Q, \ldots$ indicate double-, triple- or quadruple-zeta respectively. The inclusion of diffuse functions, which can improve the flexibility in the outer valence region, leads to the augmented correlation-consistent basis sets (aug-cc-pVXZ) where one set of diffuse functions is added to the cc-pVXZ basis. There are two widely used families of basis sets: the Pople style and correlation-consistent basis sets.

### 1.2.2.1 Pople style basis sets

**STO-nG basis sets**: These minimum type basis sets are determined by fitting to the single Slater type orbital (STO), rather than optimisation by a variational procedure. These sets have developed by Pople and n is an integer between two and six.

**k-nlmG basis sets**: These split valence basis sets were developed by Pople and coworkers, where $k$ indicates the number of primitive GTOs (PGTOs) used to represent the core orbitals and G (Gaussian) indicates the $s$- and $p$- functions in the
basis. The three values (nlm) indicate the triple split valence (such as, 3-21G, 6-31G, 6-311G).

To each of these basis sets can be added diffuse and/or polarization functions such as 3-21G* (with polarization function), 6-31G*, 6-311G*. Diffuse functions are normally s- and p- functions and consequently go before the G. They are denoted by + or ++, with the first + indicating one set of diffuse s- and p-functions on heavy atoms, and the second + indicating an additional diffuse s-functions added to hydrogen. The argument for only adding diffuse functions on non-hydrogen atoms is the same for only adding polarization functions on non-hydrogen atoms.

Polarization functions are indicated after the G, with a separate designation for heavy atoms and hydrogen. The 6-31+ G(d) is a split valence basis with one set of diffuse sp-functions on heavy atoms only and a single d-type polarization function on heavy atoms. The largest standard Pople style basis set is 6-311++ G(3df,3pd).

The currently most employed coupled-cluster method is the non-iterative CCSD(T) approach \(^42\) (where the triples are added perturbatively after an iterative CCSD calculation). The advantage of this approach lies in its relatively low computational cost while the accuracy remains relatively high. The CCSD(T) method can give results of very high accuracy, provided that sufficiently large basis sets are used. Unfortunately, the combination of highly correlated methods, such as CCSD(T), and large basis sets means that such calculations are computationally expensive. \(^6\)

1.3 Density Functional Theory (DFT)

The idea to use electron density instead of a wave function dates back to Thomas and Fermi in 1927.\(^{43,44}\) In 1964,\(^{45}\) Hohenberg and Kohn used this idea to develop Density Functional Theory (DFT) which the ground stated electronic energy is determined completely by the electron density \(\rho\), and that a trial electron density must give an energy greater than or equal to the true energy.\(^{46}\) Kohn-Sham then introduced orbitals for the use of DFT in computational chemistry.\(^{47}\) In the Kohn-Sham strategy the energy of a system in formulated as a deviation from the energy of an idealized system with non-interacting electrons. From the energy equation, by minimizing the energy with respect to the Kohn-Sham orbitals, the Kohn-Sham equations can be derived, in a
similar way to the Hartree-Fock equations. Finding good functionals connecting the electron density and energy is the main approach in DFT. The main problem with DFT is its inability to systematically improve the results, and to describe certain important features, such as van der Waals interactions.

### 1.3.1 Introduction of DFT methods in drug design

There has been a constant rise in the need for growing efficiency in the field of medicinal chemistry and drug design. When working at the quantum mechanical level, the study of biological system requires a careful selection of the computational methods. As electron correlation is excluded in HF methods, it may not be able to describe certain properties precisely. MPn and CCSD(T) methods consider electron correlation, but their high computational costs limit their use. Consequently, DFT has become the most popular approach for treating chemical systems of realistic sizes due to its attractive accuracy-to-computational-cost ratio.

The present work focuses on the application of DFT for the study of pharmaceutically relevant molecules. DFT offers promising applications in drug design and a number of such applications are currently being employed. However, there are some other studies, which point out its limitations. The superior results of DFT, in general, make it an attractive method to use in medicinal studies. When accurate functionals and basis sets are used for the study of potential drug molecules, DFT can be utilized to predict relative conformational energies, binding energies, electron affinities, ionization energies, drug molecule geometries, transition barriers, metal-ligand bond strengths and transition metal reaction pathways accurately. However, the limitations of DFT in successfully determining some properties such as atomization energy, as discussed formerly, must not be ignored. It is crucial to be aware of the accuracy of DFT for a specific issue in an attempt to identify the extent to which its predictions can be depended upon. For every system under consideration, the performance of DFT must be calibrated in order to assess whether DFT meets required accuracy. An essential point to note while carrying out a DFT analysis is the selection of the exchange correlation functional. Owing to the extensive use and availability of broad validation studies, the B3LYP hybrid functional is commonly alluring. Nevertheless, more exact functionals may be accessible and must be taken into
account. Concluding, when suitable validity analyses are executed and appropriate functionals are implemented, DFT has a lot of potential to become a very valuable tool for addressing challenging problems existing in medicinal chemistry.

### 1.3.2 Introduction of DFT methods in catalytic reactions

Catalysis in general, and heterogeneous catalysis in particular, is critical to most industrial processes, including production of chemicals, pharmaceuticals, cosmetics, foods, and polymers. Catalysis is also central to the generation of clean energy and to the protection of the environment. At present, catalysts are used in over 80% of all chemical industrial processes. Foundation of catalysis depends on chemical kinetics which is a science studying the reaction rates of chemical reactions, taking into account their reaction mechanism. In the last few decades, DFT has emerged as an attractive tool for the computational study of chemical reactions. DFT calculations of catalytic reactions on catalyst surfaces can provide insights about reactivity and mechanisms and the determining factors in the catalytic activity of the materials. For example, it has been found that the electronic structure of the catalytic surface determines its properties. A number of general relations between these factors and a catalytic activity have been developed and used for the research the design of catalytic materials for different applications.

### 1.3.3 Current DFT Methods

#### 1.3.3.1 The Hohenberg-Kohn Theorems

The first Hohenberg-Kohn theorem is a proof of existence. It states that for each different external potential $V_{ext}$ there exists one and only one corresponding non-degenerate ground state electron density $\rho_0$. Vice versa, any ground state electron density is uniquely mapped to a single external potential and uniquely determines the number of electrons $N$. Therefore $\rho_0$ uniquely defines all other properties:

$$\rho_0 \Rightarrow V_{ext} \Rightarrow \hat{H} \Rightarrow \Psi_0 \Rightarrow everything$$
Where $\hat{H}$ is the N-electron Hamiltonian operator with external potential $V_{\text{ext}}$. The first Hohenberg-Kohn theorem, then, says that any ground state property of a molecule is a function of the ground state electron density function, e.g. for the energy.

$$E_0 = E[\rho_0] = F[\rho_0]$$ \hspace{1cm} (1.37)

which says that a functional $F$ exists, but does not tell us how to find it; this omission is the main problem with DFT. The significance of this theorem is that it assures us that there is, in principle, a way to calculate molecular properties from electron density.

The second Hohenberg-Kohn theorem\(^{59}\) is a variational principle which states that the minimal energy of a chemical system is always necessarily the energy $E$ resulting from the ground state electron density $\rho_0$. Any other electron density $\rho_t$ gives a higher energy $E_0$. The second theorem can thus be state as

$$E[\rho_t] \geq E[\rho_0]$$ \hspace{1cm} (1.38)

Where $\rho_t$ is the trial electronic density and is the true energy value of the ground state.

The trial density must satisfy the conditions $\int \rho_t(r)dr = n$, where $n$ is the number of electrons in the molecule. The Hohenberg-Kohn theorems were originally proved only for non-degenerate ground states, but have been shown to be valid for degenerate ground states too.\(^{60}\)

### 1.3.3.2 The Kohn-Sham Theory

An important step toward applying DFT was made by Kohn and Sham (KS) in 1965. They calculated the molecular properties from the electron density, suggesting this might yield a way to calculate energy.\(^{61}\) There are two ideas in KS theory: (1) To express the molecular energy as a sum of terms, only one of which, a relatively small term, involves the “unknown” functional and (2) To use an initial guess of the electron density $\rho$ in the KS equations to calculate an initial guess of the KS orbitals and energy levels (below); this initial guess is then used to iteratively refine these orbitals and energy levels. The final KS orbitals are used to calculate an electron density that in turn
Chapter 1: General Introduction

is used to calculate the energy. The KS model is closely related to the HF method, sharing identical formulas for the kinetic, electron–nuclear and Coulomb electron–electron energies.

\[ E[\rho] = T_e[\rho] + V_{en}[\rho] + V_{ee}[\rho] + J[\rho] \]  \hspace{1cm} (1.39)

where \( T_e[\rho] \) gives the kinetic energy, \( V_{en}[\rho] \) yields the Coulomb attraction between electrons and nuclei, \( V_{ee}[\rho] \) describes the electron-electron interaction and \( J[\rho] \) which gives the coulomb part. The \( J[\rho] \) can be described by their classical expressions.

\[ J[\rho] = \frac{1}{2} \int \frac{\rho(r)\rho(r')}{|r-r'|} \, dr \, dr' \]  \hspace{1cm} (1.40)

Kohn and Sham introduced the calculation of the kinetic energy via a fictitious reference system of non-interacting quasi-particles, which is supposed to have the same density as the true system. This Kohn–Sham (KS) is now the most commonly used approach to DFT (KS-DFT). The missing kinetic energy for the independent compared to the correlated electrons, as well as the overall correlation and the exchange effects are described by the exchange-correlation functional \( E_{XC} \). This term corrects for the classical self-interaction energy and for the difference in kinetic energy between the fictitious non-interacting system and the real one. The sum of all these contributions is the total DFT energy.

\[ E^{DFT} = T_{KS}[\phi] + V_{en}[\rho] + V_{ee}[\rho] + J[\rho] + E_{XC}[\rho] \]  \hspace{1cm} (1.41)

\[ T_{KS}[\phi] = -\frac{1}{2} \sum_i^N \langle \phi_i | \nabla_i^2 | \phi_i \rangle \]  \hspace{1cm} (1.42)

The electronic energy and the respective set of KS-orbitals are obtained iteratively by solving the KS-equations. So the Kohn–Sham operator (\( h_{i KS} \)), an effective one-electron operator, is given by:

\[ h_{i KS} = -\frac{1}{2} \nabla_i^2 - \sum_k^nuclei \frac{z_k}{|r-r'|} + \int \frac{\rho(r')}{|r-r'|} \, dr' + V_{XC} \]  \hspace{1cm} (1.43)
\( V_{xc} \) describes exchange correlation potential that is the derivative of the \( E_{xc} [\rho] \) functional with respect to the density. The Kohn-Sham equation is a practical tool for solving many-body problems.

### 1.3.4 Various DFT Functionals

The difference between various DFT methods is the choice of functional form for the exchange–correlation energy. The accuracy of a density functional has been described by Perdew \(^{63}\) as a 'Jacob’s ladder’, which ascends from the ‘Hartree-hell’ to the ‘heaven of chemical accuracy’. Each step of the ladder can be classified for density functionals. The higher the step the more information is used from the systems in the functional and the more expensive it becomes. Below we will focus on the functionals based on Jacob’s ladder.

1) The local density approximation (LDA);
2) The generalized gradient approximation (GGA);
3) Meta-GGA (MGGA);
4) Hybrid meta-GGA (HMGGGA);
5) Double-hybrid density functionals (DHDF)

Figure 1.3 shows the five generations of DFT functionals. \(^{64,65,66,67,68,69,70}\)
1.3.4.1 The Local Density Approximation (LDA)

The simplest version of DFT, the local density approximation (LDA), is based on the assumption that electron densities vary slightly, and that it behaves like the uniform electron gas. The exchange energy for a uniform electron gas is given by:

\[
E_{\text{LDA}}^{\text{LDA}}[\rho] = -C_x \int \rho^{4/3}(\mathbf{r}) \, d\mathbf{r}
\]

(1.44)

\[
\varepsilon_{\alpha}^{\text{LDA}} = -C_x \rho^{4/3}
\]

(1.45)

Where the densities are not equal has been virtually abandoned and replaced by the Local Spin Density Approximation (LSDA), which is given as the sum of the individual densities (\(\alpha\) and \(\beta\)) raised to \(\frac{4}{3}\) the power.

\[
E_{\text{LDA}}^{\text{LDA}}[\rho] = -2^{1/3} C_x \left( \int \rho_{\alpha}^{4/3} + \rho_{\beta}^{4/3} \right) \, d\mathbf{r}
\]

(1.46)

The elaboration of the LDA method to the LSDA assigns electrons of \(\alpha\) and \(\beta\) spin to different spatial KS orbitals \(h_{\alpha}^{KS}\) and \(h_{\beta}^{KS}\), with different electron density
functions $\rho_a$ and $\rho_R$. This “spin-density theory”, LSDA, has the advantages that it can handle systems with one or more unpaired electrons, like radicals, and systems in which electrons are becoming unpaired.\(^7\)

LSDA methods often provide results with accuracy similar to Hartree–Fock methods. For molecular systems, the LSDA approximation underestimates the exchange energy by ~10%, thereby creating errors that are larger than the whole correlation energy.\(^6\) As a consequence bond strengths are overestimated, often by ~100 kJ/mol. LSDA functionals are useful in solid-state physics or the systems with metals, where the approximation of a slowly varying electron density is quite valid.\(^6\) But molecular calculations have been largely replaced by higher rungs of the ladder so LSDA functionals are little used nowadays. They have been largely replaced by methods which use gradient-corrected (“nonlocal”) functionals.

### 1.3.4.2 Generalized Gradient Approximation (GGA)

A more sophisticated approach to use the Generalized Gradient Approximation (GGA)\(^7\) in which the exchange and correlation energies are dependent not only on the electron density (LDA) but also on derivatives of the density ($\nabla \rho$). This general approximation is based on an LDA description modified by an enhancement factor $\epsilon_{xc}$ which depends on both the electron density and its gradient.

$$V_x^{GGA}[\rho] = \int \rho(r)\epsilon_{xc}(\rho(r), |\nabla \rho(r)|)dr$$  \hspace{1cm} (1.47)

In the case of LDA a unique $\epsilon_{xc}$ is available. For GGA however, because the density gradient can be implemented in various ways, several versions exist. Comparison of this approximation has shown lattice constants calculated using LDA are in general 1 – 3% smaller than the experimental constants, although the lattice parameters calculated match in most cases quite well. Common GGA functionals are the PBE exchange and correlation functional by Perdew, Burke, and Ernzerhof,\(^7\) Beckes B88 exchange functional\(^7\) and the LYP correlation functional by Lee, Yang, and Parr.\(^7\)
### 1.3.2.3 Meta-GGA

As discussed previously, the first derivative of the density functional function is used in GGA functionals. These functionals represent the third rung of Jacob ladder and include higher order derivatives of the electron density, such as the electron density Laplacian \( \nabla^2 \rho \). Additional of these derivatives seems to offer some improvement, but Laplacian \( \rho \) is numerically unstable and thus, meta-GGA functional can not depend on \( \rho \) itself but on the kinetic energy density \( \tau \) instead. This is obtained by summing the squares of the gradients of the Kohn-Sham MOs:

\[
\tau(r) = \frac{1}{2} \sum_{l=1}^{\text{occ}} |\nabla_{\phi_l}(r)|^2
\]

The kinetic energy density and the Laplacian of the density essentially carry the same information as the variable leads to the meta-GGA. Various kinds of MGGA functionals have been developed including \( \tau \) HCTH,\(^{78}\) BB98,\(^{79}\) and TPSS.\(^{80}\) A detailed discussion of the theory and mathematics behind MGGA functionals is provided by Kurth and Perdew.\(^{81}\)

### 1.3.4.4 Hybrid GGA (HGGA)

The fourth lung of Jacob’s ladder of DFT functionals contains the hybrid GGA. In this approach, exchange-correlation energy \( E_{xc}(\rho) \) can be taken as a weighted sum of the DFT exchange-correlation energy and HF exchange energy, also called the 'exact' exchange.

\[
E_{xc}^{\text{hybrid}} = E_{xc}^{\text{meta-GGA}} + (1 - a_x)E_{xc}^{\text{meta-GGA}} + a_x E_{xc}^{\text{HF}}
\]

The percentage of HF exchange energy used is the main distinguishing characteristic of these hybrid functionals. The first successful hybrid method was B3LYP, used in this thesis, is given by:

\[
E_{xc}^{\text{B3LYP}} = (1 - a)E_{xc}^{\text{LSDA}} + a E_{xc}^{\text{exact}} + b \Delta E_{xc}^{\text{B88}} + (1 - c)E_{xc}^{\text{LYP}}
\]
Chapter 1: General Introduction

Where $a$, $b$ and $c$ are optimised to 0.20, 0.72 and 0.81, respectively. The PBE functional has also been improved in the B3LYP method by addition of exact exchange to give the PBE0 functional (also denoted PBE1PBE in the literature). Inclusion of this exact HF exchange is often found to improve the calculated results, although the optimum fraction to include depends on the specific property of interest. The improvement of new functionals by inclusion of a suitable fraction of exact exchange is now a standard feature of B3LYP. At least part of the improvement may arise from reducing the self-interaction error, since HF theory is completely self interaction-free.

Another class of functionals exists in the fourth rung of Jacob’s ladder. These functionals are known as hybrid meta-GGA functionals, which can be generated by adding Hartree-Fock exchange to the meta-GGA. Hybrid MGGA (HMGGA) uses the first derivative of $\rho$ and its second derivative, or the kinetic energy density. The strongpoint of HMGGA seems to be “an improvement over the previous formalisms in barrier heights and atomization energies. All the calculation in this thesis are done in non-periodic condition. In addition, we use Berny algorithm (Keyword OPT=TS) for finding transition structure (TS). Apart from that we also perform intrinsic reaction coordinate (IRC) calculations for confirming the connectivities of the TSs.

1.3.4.5 Double Hybrid Density Functional Approximation (DHDF)

Functionals on the fifth and last rung of the ladder take the virtual KS-orbitals into account when calculating the correlation energy. Several approaches to accomplish this have been published, including the perturbation methods described by Göorling and Levy, and random phase approximation (RPA) methods. The most widely used approach is the double-hybrid density functional (DHDF) proposed by Grimme which is given by:

$$E^{DHDF}_{XC} = (1 - a_X)E^{(\text{meta-})\text{GGA}}_X + a_X E^{HF}_X + (1 - a_C)E^{(\text{meta-})\text{GGA}}_C + a_C E^{MP2}_C$$

In DHDFs, a part of the correlation energy is computed by second order Møller-Plesset perturbation theory (MP2) from the KS-orbitals of a preceding hybrid functional.
Chapter 1: General Introduction

SCF calculation. The most accurate are those containing high amounts of Fock exchange and much smaller amounts of non-local correlation. For example, the B2-PLYP\textsuperscript{86} functional employs 53\% exact exchange ($\alpha_X = 0.53$) and 27\% MP2 correlation ($\alpha_C = 0.27$).

The idea of DHDF theory is based on a mixing of standard GGA exchange and correlation functionals with HF exchange and the perturbative second-order correlation part of MP2.\textsuperscript{87} The virtual orbital-dependent part accounts for nonlocal electron correlation effects that are very difficult to treat accurately by density-only (local) functionals.\textsuperscript{88} Several studies have convincingly shown that DHDF belong to the most accurate quantum chemical approaches for large molecules. They are very accurate for thermodynamic data yielding the smallest mean absolute deviation (MAD) ever obtained by a density functional for the full G3/05 test set.\textsuperscript{88,89} Examples of popular DH functionals are B2-PLYP,\textsuperscript{86} B2GP-PLYP,\textsuperscript{90} DSD-BLYP,\textsuperscript{91} DSD-PLYP,\textsuperscript{92} DSD-PBEP86,\textsuperscript{93} and PWPB95.\textsuperscript{94}

1.3.5 Dispersion Correction

Despite the many successes of DFT methods, the long-range electron correlations that describe the dispersion forces (part of van der Waals type interactions) are poorly represented in the current DFT methods.\textsuperscript{95,96,97} Over the recent years, many different approaches to treat London dispersion have been proposed.\textsuperscript{98,99,100} In the following, only the DFT-D scheme introduced by Grimme and coworkers will be discussed in more detail as it was used extensively in this thesis. The DFT-D scheme provides a semi-classical dispersion energy $E_{dis}^D$ that can simply be added to any converged DFT, HF calculations. The energy correction is generally given as:

$$E_{dis} = \frac{1}{2} \sum_{A \neq B} \sum_{n=6,8,10} S_n \frac{C_n^{AB}}{R_{AB}^n} f_{disp}(R_{AB})$$  \hspace{1cm} (1.52)

The $C_n^{AB}$ denotes the system dependent $n^{th}$ order dispersion considered in an all atom pairs ($AB$) system, which depends on the coordination number, and have the interatomic distance of $R$. $S_n$ are the global scaling factors. DFT accounts for some correlation energy in the short- to medium-range, and the sum in Equation 2.43 diverges
for small $R_{AB}$. Thus, the dispersion energy needs to be damped for small distances to avoid divergence of energy expression.

The initially proposed damping function is very close to the damping function used in DFT-D1/2. As the dispersion energy vanishes for short distances, this damping is called zero damping and the correction scheme is thus called D1/2(0). However, a problem with is that the zero damping is that it may in unphysical repulsion forces. Becke and Johnson (BJ) introduced a rational damping function $R_{AB}^n$ that leads to a constant contribution of the dispersion energy to the total correlation energy from spatially close, directly bonded, pair of atoms. This BJ-damping function is the default in the DFT-D3 scheme. The best version of dispersion correction DFT-D3 is used in this thesis. One advantage of the D3 correction is that it, does not depend on the density and thus, the electronic structure is not directly affected. In addition, the D3 scheme is available for almost all elements in the periodic table and can simply be added to any DFT, HF or semiempirical calculation without the necessity for specific implementation into a program.

1.4 Composite Methods

Composite methods use a series of less-computationally intensive (i.e., reduced CPU time, memory, and disk space requirements) to approximate calculations at high levels of theory with large basis sets, such as CCSD(T)/CBS. In order to achieve reliable energetics, composite methods must account for physical effects arising from electron correlation, relativistic effects, and anharmonic zero point energies. Well-known ab initio composite methods include, Weizmann-n (Wn) the complete basis set (CBS-n) method developed by Petersson and co-workers. The High Accuracy Extrapolated ab initio Thermochemistry (HEAT), created by Stanton and coworkers, and the correlation consistent composite approach (ccCA) and the Gaussian-n (Gn) methods. These Gn approaches combine a series of lower-level quantum computations to estimate the result of high-level accuracy. The advantage of these methods is the use of high correlation level and big basis sets. This series began in 1989 with G1 continued with G2 (1991) and G3 (1998) and the publication for G4 (2007). The most popular Gaussian high-
accuracy methods at present are probably G4 and G3 and their faster but nearly as accurate variants, G4(MP2). In the following section, we briefly discuss about G4 theory and G4MP2 which be used extensively in this thesis.

The validation of such methods requires a set of highly accurate experimental data experimental thermochemical and kinetic data has been used for this purpose, however, the available experimental data is not always sufficiently diverse and/or sufficiently accurate. Composite ab initio methods offer an alternative since they can achieve accuracies of at least one order of magnitude better than the best contemporary DFT functionals. The use of benchmark experimental data is additionally complicated by the fact that it has to be converted into nonrelativistic electronic energies for a direct comparison with the DFT data.

1.4.1 Gaussian-4 Theory

The G4 theory contains several modifications not included in G3 theory lead to improvement of performance. The G4 energy is given by:

\[
E_0(G4) = E \left( CCSD(T), 6 - 31G(d) \right) + E(plus) + E(2df,p) \\
+ E(\Delta G3LXP)E(HF/limit) + E(SO) \\
+ E(HLC) + E(ZPE)
\]

The \( E(plus) \) is the effect of diffuse function obtained the difference between MP4/6-311G(d,p) and MP4/6-311+G(d,p). The term \( E(2df,p) \), which is higher polarization function is obtained as the difference between MP4/6-311G(d,p) and MP4/6-311G(2df,p). The \( (\Delta G3LXP) \) is given by:

\[
E(\Delta G3LXP) = E[MP2(Full)/G3LargeXP] - E[MP2/6 - 31G(2df,p)] \\
- E[MP2/6 - 31 + G(d)] \\
+ E[MP2/6 - 31G(d)]
\]
Chapter 1: General Introduction

The spin-orbit correction $E(SO)$ is included only for atomic species. The higher level correction $E(HLC)$ equals to $-An_\beta - B(n_\alpha - n_\beta)$ for molecules (with $A = 6.9471, B = 2.4409$ mhartrees) and $-Cn_\beta - D(n_\alpha - n_\beta)$ for atoms.

1.4.1.1 G4MP2 Theory

G4(MP2) theory is based on G4 theory.\(^{128}\) It uses B3LYP/6-31G(2df,p) optimized geometries for a series of single point energy calculations at higher levels of theory. This basis set was found to perform well for most geometries with some exceptions hydrogen bonded complexes and Jahn-Teller distorted systems. The zero-point energy, $E(ZPE)$, is based on B3LYP/6-31G (2df,p) frequencies scaled by 0.9854, the same as in G4 theory. We have previously analyzed the use of scaled vibrational frequencies in the Gn methods and found that they work well over a range of molecules.\(^{133}\) The first energy calculation is at the triples-augmented coupled cluster level of theory, CCSD(T)/ 6-31G(d). This energy is then modified by a series of energy corrections to obtain a total energy $E_0$,

$$E_0[G4(MP2)] = E[CCSD(T)/6 - 31(d)] + \Delta E(MP2) + \Delta E(HF) + \Delta E(HLC) + \Delta E(SO) + E(ZPE)$$

The correction at the second-order Moller-Plesset level (MP2) basis-set-correction term is given by

$$\Delta E(MP2) = E[MP2/G3MP2LargeXP] - E[MP2/6 - 31(d)]$$

and the $\Delta E(HF)$ term corrects for the HF/CBS (complete-basis-set) limit energy extrapolated from truncated versions of the aug-cc-pVTZ and aug-cc-pVQZ basis sets. The HLC is added to account for remaining deficiencies in the energies and has the same form as in G4 theory.

In this thesis, we used the G4(MP2) as the Gaussian method of choice, as it represents a good compromise between accuracy and speed. This method has average absolute deviations from experiment of 3.5 kJ mol\(^{-1}\). Overall, the G4(MP2) method provides an accurate and economical method for thermodynamic predictions.
1.4.2 Weizmann-n (Wn) Theories

The primary goal of designing the Weizmann-n (Wn) procedures\textsuperscript{134} was to achieve accuracy in the sub-kcal/mol range. Wn methods (W1 and W2) theories that incorporate relativistic effects have MADs of 1.3 and 1.0 kJmol\textsuperscript{-1} respectively for molecular heats of formation of first- and second-row compounds. Attempting to extend the applicability of W1 and W2 methods, Karton and Martin developed explicitly correlated versions, denoted as W1-F12. The F12 theory is based on the fact that the basis set limit of correlation energy is difficult to compute precisely. I briefly describe the steps involved in W1-F12 since this method has been used in the following project in Chapter 4.

(i) Reference geometry is obtained at B3LYP/cc-pV(T+d)Z level of theory, where the zero-point vibrational energy (ZPVE) contribution is scaled by 0.985;

(ii) The SCF energy extrapolated from the HF/VDZ-F12 and HF/VTZ-F12 energies with an exponent of 5.0 in W1-F12.

(iii) The valence CCSD is extrapolated from the V\{D,T\} Z-F12 basis set pair with an exponent of 5.0 in W1-F12.

(iv) The valence (T) contribution is extrapolated from A’V\{D,T\}Z basis set pair by using the two-point extrapolation formula with $\alpha$ equals to 3.22 in W1-F12.

(v) The inner-shell contribution is obtained from standard CCSD and (T) calculations. CCSD is calculated with PWCVTZ basis set in W1-F12.

(vi) Scalar relativistic correlation is obtained as the difference between non-relativistic CCSD(T)/A’VDZ.

(vii) The spin-orbit coupling is taken from the experimental fine structure.
Chapter 2: Antioxidant design application

2.1 Introduction

Oxidative stress is defined as the imbalance between the production of reactive oxygen species (ROS) and the capability of the cell to elicit an effective antioxidant response. Oxidative stress has been implicated in a number of diseases, including fibrosis and cancer.\(^\text{135}\) It also plays a pivotal role in the progression of Alzheimer's disease (AD), the most common neurodegenerative disorder.\(^\text{136}\) ROS are generally defined as oxygen-containing small species including superoxide anion radical (O\(_2^−\)), hydroxyl radical (OH\(^−\)), hydroxyl ion (OH\(^+\)), hydrogen peroxide (H\(_2\)O\(_2\)), singlet oxygen (1O\(_2\)), and ozone (O\(_3\)).\(^\text{137, 138}\) These species can be generated from either extracellular/intracellular sources and leads to damages biological molecule, including DNA, RNA, protein enzymes, unsaturated lipids, etc.\(^\text{139}\)

Recently dietary supplementation with antioxidants has gained much attention,\(^\text{140, 141, 142, 143}\) as potential therapeutic intervention for its ability to fight oxidative stress. The main function of antioxidants is to scavenge or neutralize free radical formation and to inhibit the deleterious effects of ROS. Figure 2.1

Many experiments show that carnosine and vitamin E can be effective antioxidants depending on their concentration and mode of action. However, there is lack of computational studies of their properties which could help us better understand the mechanism of their antioxidant activity. We performed the computations methods both in the gas phase and in water solution. We hope that our computational investigation will inspire further experimental investigations of the antioxidant activity of the proposed antioxidants, which can inhibit their oxidative damage leading to ROS.
production.

**Figure 2.1** Schematic representation of oxidative stress

The following section gives an overview of an investigation into the computational antioxidant design that could provide insights to modern drug discovery.
2.1.1 Introduction to the antioxidant design based on Carnosine

Carnosine (b-alanyl-L-histidine) is an endogenous dipeptide produced in the body by the enzyme carnosine synthase. Vladimir Gulevitsch, who isolated and purified this compound and established its chemical structure, described it as a component of beef extract as early as 1900. It was discovered that carnosine is accumulated in vertebrate brain and muscles in amounts proportional to their functional activity. Carnosine can also play a regulatory role in inhibiting biological damage mediated by HOX (X = Cl, Br).

Myeloperoxidase (MPO), is most abundant enzyme in neutrophils and inflict oxidative damage on invading pathogens to prevent and control microbial infections. At the site of inflammation Myeloperoxidase (MPO), uses H₂O₂ to catalyse the oxidation of X⁻ ions to form strong oxidizing agents HOX. HOX plays a key role against invading pathogens by inflicting massive oxidative damage. However, reactive HOX species also damage host tissue and contribute to inflammatory injury when produced at the wrong place, time, and/or concentration. This has been implicated in numerous inflammatory diseases including atherosclerosis, arthritis, and some cancers. Therefore, it is important to identify the mechanisms by which carnosine can effectively scavenge HOX (X = Cl, Br) and as a result rationally design exogenous agents with higher efficacy.

Carnosine has two imidazole nitrogens (N₁ and N₂), an amino nitrogen (N₃), and an amido nitrogen (N). (Figure 2.2) The N₂ imidazole nitrogen is readily chlorinated by HOX (X = Cl, Br). Kinetic measurements suggest that, subsequent to the initial chlorination at the N₂ imidazole nitrogen, a rapid intramolecular X⁺ transfer occurs, in which the X⁺ migrates to the terminal primary amine (N₃).
Figure 2.2 Schematic representation of the intramolecular rearrangements involved in the X⁺ (X = Cl and Br) transfer from the imidazole nitrogen (N₂) to the terminal amino nitrogen (N₃).

In this work we investigate the intramolecular Br⁺ transfer in carnosine using the high-level G4(MP2) variant of the Gaussian-4 (G4) composite thermochemical protocol.¹⁶³,¹⁶⁴ First part of the project shows that the barrier for the intramolecular Br⁺ shift is similar to that of the Cl⁺ transfer. I proceed to explore structure-function relationships for lowering the reaction barrier height for the intramolecular X⁺ shift in carnosine (X = Cl and Br) as this step is definitely rate determining step (RSD) of overall process. So for the rest of the investigation, I focused in this step. I find that increasing the length of the β-alanyl-glycyl side chain significantly reduces the reaction barrier height for the Br⁺ transfer. Then proceed to investigate electronic effects on the reaction barrier height and show that substituting the imidazole ring with electron donating groups reduces the reaction barrier for the intramolecular X⁺ shift (X = Cl and Br), whilst substitution with electron withdrawing groups increases the reaction barrier. Finally, I show that combining the electronic and structural effects leads to very low barrier heights for the X⁺ shifts. For our best antioxidants I obtain reaction barrier heights of 41.4 and 27.8 kJ mol⁻¹, respectively, for the Cl⁺ and Br⁺ shifts (cf. to reaction barrier heights of 109.5 and 97.2 kJ mol⁻¹ in carnosine, respectively).
2.1.2 Introduction to the antioxidant design based on vitamin E

Free radical peroxidation of unsaturated chains in bio-membrane lipids disrupts the important structural function of membranes and is associated with a variety of significant pathological events.\textsuperscript{165} Barclay\textsuperscript{165} has extensively studied the kinetics of inhibition of lipid peroxidation by α-tocopherol (AT), the most biologically active form of the vitamin E, and found that the rate constant for scavenging peroxyl radicals is smaller in the membranes by 2–3 orders of magnitude than in solution.\textsuperscript{166} This indicates that there is always a need for designing more effective antioxidant based on vitamin E in the membrane. Without more detailed knowledge of oxidative mechanism, this is mainly speculation.

Each AT molecule can trap two radicals,\textsuperscript{167} to best of our knowledge; second radical scavenging process leading to α-tocopherylquinone (ATQ) has not been understood computationally until today. 2.3 illustrates the complete radical scavenging mechanism of AT. The tocopheroxyl radical that is formed is sufficiently stable to be unable to continue the chain and, instead, is removed from the cycle by reaction with another radical to form an inactive, non-radical product.\textsuperscript{168}
Chapter 2: Antioxidant design application

Figure 2.3 The reaction mechanism for the antioxidant activity of α-tocopherol (AT*), through the reaction with two hydroxyl radicals. The ring-opening reaction leading to ATQ is the subject of the present work and shown in a dashed circle.

In the present work, double-hybrid density functional theory (DHDDFT) are employed to obtain accurate geometries and energies of transition states (TSs), reactive intermediates, and products, thus producing a potential energy surface (PES) that identifies key reaction pathways of the second ‘OH scavenging reaction mechanism of a simplified model of (AT) here referred AT*. The present work includes two parts: First, we explore the mechanism of ring opening rearrangement reaction following 2nd ‘OH scavenging of AT* to ATQ, which is highlighted in figure 2.3 We explore an uncatalysed, one water-catalysed and 2 water-catalysed pathways.

With no catalyst and catalyst with one and two explicit water molecules, which have not previously been considered in the antioxidant activity of vitamin E. Secondly, based on the results obtained from the first part of this study, we systematically explored several candidates to design novel candidates with improved antioxidant activity of the AT. Our results suggest that the reducing the atom number of heterocyclic ring is a suitable way to synthesize the novel antioxidants with improved antioxidant activity than vitamin E.
2.2 Series of papers
2.2.1 Computational design of carnosine-based HOBr antioxidants

Presentation of the article

Title: Computational design of bio-inspired carnosine-based HOBr antioxidants.
Authors: Sarrami, F., Yu, L.J. and Karton, A
DOI: https://dx.doi.org/ 10.1007/s10822-017-0060-3
Date of Publication: October 2017

Graphical TOC
Computational design of bio-inspired carnosine-based HOBr antioxidants

Farzaneh Sarrami1 · Li-Juan Yu1 · Amir Karton1

Received: 20 May 2017 / Accepted: 31 August 2017
© Springer International Publishing AG 2017

Abstract During a respiratory burst the enzyme myeloperoxidase generates significant amounts of hypohalous acids (HOX, X = Cl and Br) in order to inflict oxidative damage upon invading pathogens. However, excessive production of these potent oxidants is associated with numerous inflammatory diseases. It has been suggested that the endogenous antioxidant carnosine is an effective HOCl scavenger. Recent computational and experimental studies suggested that an intramolecular Cl⁻ transfer from the imidazole ring to the terminal amine might play an important role in the antioxidant activity of carnosine. Based on high-level ab initio calculations, we propose a similar reaction mechanism for the intramolecular Br⁻ transfer in carnosine. These results suggest that carnosine may be an effective HOBr scavenger. On the basis of the proposed reaction mechanism, we proceed to design systems that share similar structural features to carnosine but with enhanced HOX scavenging capabilities for X = Cl and Br. We find that (i) elongating the β-alanylglycyl side chain by one carbon reduces the reaction barriers by up to 44%, and (ii) substituting the imidazole ring with strong electron-donating groups reduces the reaction barriers by similar amounts. We also show that the above structural and electronic effects are largely additive. In an antioxidant candidate that involves both of these effects the reaction barriers are reduced by 71%.

Keywords Molecular design · Antioxidant design · Computational chemistry · CCSD(T) · G4(MP2) theory

Introduction

The most abundant enzyme in neutrophils (white blood cells) is myeloperoxidase (MPO). MPO plays a key role in the antimicrobial oxygen-dependent activity of neutrophils. This heme-containing enzyme uses H₂O₂ to catalyze the oxidation of X⁻ ions (X = Cl, Br) to form strong oxidizing agents HOX [1–3]. The normal function of MPO is to inflict oxidative damage on invading pathogens as part of an immune response for preventing and controlling microbial infections [4, 5]. However, reactive HOX species also damage host tissue and contribute to inflammatory injury when produced at the wrong place, time, and/or concentration. The relationship between oxidative damage mediated by HOX and various chronic diseases has been extensively documented. It is now understood that HOX play a pivotal role in many inflammatory (and other) diseases including atherosclerosis, cancer, diabetes, asthma and Alzheimer’s disease [4–13]. Major targets for HOX oxidation include sulfur-containing amino acids (e.g., Cys, Met), amino acids containing reactive nitrogens (e.g., Lys, His, Arg), nucleobases (e.g., cytosine, thymine, guanine, adenine) and other biologically important purine bases (e.g., xanthine, hypoxanthine, uric acid) [14–31].

Carnosine (β-alanyl-l-histidine, Fig. 1) is an endogenous dipeptide produced in the body by the enzyme carnosine synthase, and is concentrated in muscle, heart, and brain tissues [32]. It has been shown to play a preventative role in many inflammatory diseases (e.g., cardiac disorders, cancer, diabetes, Alzheimer’s, and Parkinson’s diseases) [33]. It has been suggested that carnosine may play a role in inhibiting
biological damage mediated by HOCl by effectively removing the chlorinating equivalents [34–37]. A recent computational investigation proposed that an intramolecular Cl⁺ shift in carnosine is responsible for stabilizing the chlorinated carnosine product and therefore may contribute to the antioxidant activity of carnosine [37]. This study found that increasing the length of the β-alanyl-glycy1 side chain in carnosine reduces the reaction barrier height for the intramolecular Cl⁺ shift. In particular, increasing the length of the side chain by one and two carbon atoms resulted in reaction barrier heights of 66.4 and 49.4 kJ mol⁻¹, respectively, cf. to a reaction barrier height of 109.5 kJ mol⁻¹ in carnosine. Based on these findings this study designed a number of antioxidants with longer β-alanyl-glycyl side chains.

It has been shown experimentally that the rate constants for the reaction of HOBr with nitrogen-containing functional groups increases in the order of amide< free amine < imidazole [26]. This experimental result suggests that the intramolecular Br⁻ transfer in carnosine is expected to be a key step in the antioxidant activity of carnosine since it converts the kinetic product (brominated at the imidazole N₂ nitrogen) into the thermodynamic product (brominated at the amine N₂ nitrogen, Fig. 1) [15]. We note that initial bromination at the imidazole ring is unlikely to affect the kinetics of the overall process since it has been found that the reaction of HOBr with the imidazole ring in 4-imidazoleacetic acid proceeds very rapidly and is too fast to follow directly by stopped-flow methods at 10 °C [26]. The ability of carnosine to act as an effective HOBr scavenger, in addition to being an HOCl scavenger, is particularly significant in light of the fact that HOBr reacts 30–5000 times faster than HOCl with many amino acids residues (e.g., Ala, Gly, Lys, Trp, Tyr) and other biologically relevant functional groups such as substituted imidazoles [23, 26].

In this work we investigate the intramolecular Br⁻ transfer in carnosine. We show that the barrier for the intramolecular Br⁻ shift is similar to that of the Cl⁺ transfer. We proceed to explore structure–function relationships for lowering the reaction barrier height for the intramolecular X⁺ shift in carnosine (X = Cl and Br). We find that increasing the length of the β-alanyl-glycyl side chain significantly reduces the reaction barrier height for the Br⁻ transfer. We proceed to investigate electronic effects on the reaction barrier height and show that substituting the imidazole ring with electron donating groups reduces the reaction barrier for the intramolecular X⁺ shift (X = Cl and Br), whilst substitution with electron withdrawing groups increases the reaction barrier. Finally, we show that combining the above electronic and structural effects leads to very low barrier heights for the X⁺ shifts. For our best antioxidants we obtain reaction barrier heights of 41.4 and 27.8 kJ mol⁻¹, respectively, for the Cl⁺ and Br⁻ shifts (cf. to reaction barrier heights of 109.5 and 97.2 kJ mol⁻¹ in carnosine, respectively).

**Computational methods**

The geometries of all structures were optimized using the B3LYP-D3 density functional theory (DFT) exchange-correlation functional in conjunction with the 6-31G(2df,p) Pople-style basis set [38–41]. Empirical D3 dispersion corrections [42, 43] were included using the Becke-Johnson [44] damping potential as recommended in ref. [41] (denoted by the suffix-D3). Bulk solvent effects in aqueous solution were included using the charge-density-based SMD continuum solvation model [45]. The resulting level of theory is denoted by SMD(water)-B3LYP-D3/6-31G(2df,p). Harmonic vibrational analyses have been performed at the same level of theory to confirm each stationary point as either an equilibrium structure (i.e., all real frequencies) or a transition structure (i.e., with one imaginary frequency). Zero-point vibrational energies (ZPVEs), thermal correction to the enthalpy, and entropic corrections have been obtained from these calculations. The connectivity of the local minima and first-order saddle points was confirmed by performing intrinsic reaction coordinate calculations [46, 47]. In order to ensure that the equilibrium structures are in their minimum-energy conformations we carried out systematic conformational searches using the MMFF94 molecular mechanics force-field with the Avogadro program [48]. The geometries of the low-lying conformations were then re-optimized at the SMD(water)-B3LYP-D3/6-31G(2df,p) level of theory and the lowest-energy conformation was selected.

Gas-phase electronic energies were obtained using the high-level G4(MP2) variant of the Gaussian-4 (G4)
composite thermochemical protocol [49, 50] using the SMD\textsubscript{water}-B3LYP-D3/6-31G(2df,p) optimized geometries. The G4(MP2) protocol is an efficient composite procedure for approximating the CCSD(T) (coupled cluster energy with singles, doubles, and quasiperturbative triple excitations) energy in conjunction with a large triple-ζ-quality basis set [49–51]. G4(MP2) theory has been found to produce gas-phase thermochemical properties (such as reaction energies, bond dissociation energies, and enthalpies of formation) with a mean absolute deviation (MAD) of 4.4 kJ mol\textsuperscript{−1} from the experimental energies of the G3/05 test set [52]. It has been found that G4(MP2) shows a similarly good performance for reaction barrier heights [53–56]. The electronic G4(MP2) energies were converted to Gibbs free energies at 298 K using the above-mentioned ZPVE, enthalpic, and entropic corrections. Corrections for bulk solvent effects in aqueous solution were added to the gas-phase G4(MP2) Gibbs free energies using the SMD model at the M05-2X/6-31G(d) level of theory as recommended by Marenich, Cramer, and Truhlar [45].

In summary, we follow here the same computational approach that was used in ref. [37], namely, the Gibbs free energies in aqueous solution (∆G\textsubscript{298,aq}) are computed according to the following equation:

\[ \Delta G_{298,\text{aq}} = \Delta E_a + \Delta ZPVE + \Delta (H_{298} - H_0) - \Delta S^0 + \Delta E_{\text{uvb}}(\text{SMD}\textsubscript{water} - \text{M05-2X}/6-31G(d)) \]

where the ZPVE, enthalpic temperature (H\textsubscript{298} – H\textsubscript{0}), and entropic (∆S) corrections have been obtained within the rigid rotor-harmonic oscillator approximation from the SMD\textsubscript{water}-B3LYP-D3/6-31G(2df,p) geometries and harmonic frequencies. The single-point energy calculations involved in the electronic energies (∆E\textsubscript{uvb}(G4(MP2))) and bulk solvent corrections [∆E\textsubscript{uvb}(SMD\textsubscript{water} - M05-2X/6-31G(d))] were also calculated using the SMD\textsubscript{water}-B3LYP-D3/6-31G(2df,p) optimized geometries. We note that it was found in ref. [37] that the SMD and conductor-like polarizable continuum model (CPCM) [57] solvation models predict similar solvation energies and lead to the same qualitative trends. All the DFT and ab initio calculations were carried out with the Gaussian 09 suite of programs [58].

Result and discussion

Reaction mechanism for the intramolecular Br\textsuperscript{+} transfer

We begin by exploring the potential energy surface (PES) for the intramolecular Br\textsuperscript{+} transfer in carnosine using high-level ab initio G4(MP2) calculations in a simulated aqueous environment (see “Computational methods”). The equilibrium and transition structures that were located along the reaction path are shown in Fig. 2 along with a schematic representation of the Gibbs free-energy surface (∆G\textsubscript{298,SMID\textsubscript{water}}-G4(MP2), kJ mol\textsuperscript{−1}). The lowest energy conformation of brominated carnosine (REAC, Fig. 2) was chosen as the zero-energy reference point of the reaction profile. The Gibbs-free energy PES for the bromine transfer is similar to that for the chlorine migration reported in ref. [37].

The first reaction intermediate (INT1, Fig. 2) lies higher in energy than the minimum-energy conformation by 14.1 kJ mol\textsuperscript{−1}. In this conformation of REAC the carboxylic acid adopts the higher-energy trans configuration, allowing the OH group to form a hydrogen bond with the adjacent imidazole nitrogen N2 (with an OH–N2 distance of 1.638 Å, and an OHN angle of 168.4°). The first step involves a proton transfer from the carboxylic acid oxygen to the imidazole nitrogen (N2, Fig. 1). The barrier for this step is merely 20.0 kJ mol\textsuperscript{−1}. This step is of key importance for the following halide transfer step since it creates a net positive charge on the imidazole ring, which makes the halide transfer more facile. This is demonstrated by looking at the atomic polar tensor (APT) [59, 60] charges of the halide and the imidazole ring. The APT charges on the bromine atom is positive in both REAC (0.047 a.u.) and INT2 (0.174 a.u.). However, whilst in REAC the net charge on the imidazole ring is slightly negative (−0.111 a.u.) hence resulting in an electrostatic attraction between the two moieties, in INT2 the net charge on the imidazole ring is very positive (+0.882 a.u.). A similar trend is observed for X = Cl (for a list of all the APT charges see Table S1 of the Supporting Information). The second step, the key for the antioxidant activity of carnosine, is the intramolecular migration of Br\textsuperscript{+} from the kinetically favored site (imidazole ring) to the thermodynamically preferred site (terminal amine) [14–18, 37]. The transition structure for this step (TS2) lies 97.2 kJ mol\textsuperscript{−1} above the initial reactant (REAC). This reaction barrier height is lower by 12.3 kJ mol\textsuperscript{−1} than that for the Cl\textsuperscript{+} transfer [37]. Thus, carnosine should be a more effective HOBs-antioxidant than an HOCl-antioxidant. The last step is a proton shift from the terminal amine to the carboxylic acid oxygen. The transition structure for the second proton transfer (TS3) lies 23.2 kJ mol\textsuperscript{−1} above the reactant. The resulting reaction intermediate (INT4) lies 12.9 kJ mol\textsuperscript{−1} above the energy of the reactant (REAC). However, the minimum-energy conformation of INT4 lies 33.6 kJ mol\textsuperscript{−1} below the reactant (Fig. 2). Thus, the final product (PRODUCT) results in effective trapping of the bromine atom.

The relative Gibbs free energy of the TS for the Br\textsuperscript{+} shift (TS2) is significantly higher in energy than that for the proton shifts, thus this step is clearly the rate-determining step (RDS) for the overall process. We will therefore focus in the next sections primarily on this step. We also note that TS2
has a substantially larger dipole moment ($\mu = 13.5$ D) than the reactant ($\mu = 2.2$ D). Therefore the Br-transfer reaction is expected to proceed with greater facility as the polarity of the solvent in which the reaction is conducted is increased. We hope that this study would inspire experimental measurements for testing this computational prediction.

It should be noted that in aqueous solution under physiological conditions the reactant (REAC) would be expected to exist predominantly in its zwitterionic form ($\text{RNH}_3^+\text{R}^\prime\text{COO}^-$), which does not allow for a unimolecular bromine transfer to proceed. However, the proposed mechanism can proceed through the portion of the population in the non-zwitterionic form. Based on DFT geometry, energy, and NMR chemical shift calculations in a simulated aqueous environment Diez et al. suggested that the nonzwitterionic form of carnosine should exist in appreciable amounts in aqueous solution [61]. For the brominated form of carnosine we obtain that the zwitterionic form of REAC lies 20.5 kJ mol$^{-1}$ below REAC at the SMD(water)-B3LYP-D3/6-31G(2df,p) level of theory. These results are consistent with those obtained for the chlorinated form of carnosine [37].

Rational design of HOX antioxidants

It has been previously found that increasing the length of the $\beta$-alanyl-glycyl side chain in carnosine significantly reduces the reaction barrier height for the intramolecular Cl$^+$ shift [37]. In "Structural modifications" we will show that similar structural modifications also reduce the reaction barrier height for the intramolecular Br$^+$ transfer. In "Effects of substituents on the imidazole ring" we will explore the effects of substituents on the imidazole ring on the intramolecular X$^+$ shift, X = Cl and Br.

Structural modifications

Figure 3 shows a schematic representation of the modified carnosine structures that were considered in which the length of the $\beta$-alanyl-glycyl side chain has been: (i) shortened by one carbon (Gly-His) or (ii) increased by one carbon (homocarnosine).

The calculated Gibbs free-energy reaction barrier heights ($\Delta G^*_{298}$) and energies ($\Delta G_{298}$) for the intramolecular Br$^+$ transfer reactions are summarized in Table 1. The
reaction for carnosine is associated with $\Delta G_{298}^\ddagger = 97.2$ and $\Delta G_{298} = -33.6$ kJ mol$^{-1}$. Shortening the $\beta$-alanyl-glycyl side chain by one carbon (Gly-His) results in a higher barrier of $\Delta G_{298}^\ddagger = 111.3$ and a more exergonic reaction energy of $\Delta G_{298} = -43.5$ kJ mol$^{-1}$. On the other hand, increasing the length of the side chain by one carbon (homocarnosine) decreases the barrier by a significant amount ($\Delta G_{298}^\ddagger = 54.3$) and the overall reaction becomes much more exergonic ($\Delta G_{298} = -59.8$ kJ mol$^{-1}$). These results are consistent with the results obtained for the Cl$^+$ transfer (see Table 1 and ref. [37]).

Figure 4 shows the optimized transition structures for the intramolecular bromine transfer in carnosine and its structural derivatives. A close inspection of Fig. 4 reveals that increasing the length of the $\beta$-alanyl-glycyl side chain results in TS$_2$ for the Br$^+$ transfer becoming earlier (i.e., less distorted) and therefore lower in energy. This is demonstrated by the angle between the plane of the imidazole ring and the N–Br bond decreasing in the order: 52.3° (Gly-His), 39.5° (carnosine), and 27.9° (homocarnosine). In addition, the length of the N$_2$–Br bond decreases in the same order, namely it is 1.933 (Gly-His), 1.893 (carnosine), and 1.863 (homocarnosine) Å. These results indicate that increasing the length of the $\beta$-alanyl-glycyl side chain allows for the bromine atom to be positioned closer to the imidazole ring in TS$_2$. Consequently, the imidazole ring becomes more aromatic and TS$_2$ becomes more stable. These results are consistent with the results for the Cl$^+$ transfer discussed in ref. [37].

We also find that the strong hydrogen bond formed between the $\text{HN}^+$ and $\text{COO}^-$ moieties in TS$_2$ is another factor contributing to the increased stability of TS$_2$ in the order: Gly-His < carnosine < homocarnosine. In particular, the length of this hydrogen bond decreases in the order: 2.335 (Gly-His), 1.922 (carnosine), and 1.607 (homocarnosine) Å (Fig. 4). These H-bond distances indicate that the H-bond becomes stronger as the length of the $\beta$-alanyl-glycyl side chain increases. The ability to form a more effective

![Fig. 3](image1)

**Fig. 3** Structurally modified candidates considered for HOBr-antioxidants with potentially reduced barriers for the intramolecular Br$^+$ transfer.

![Fig. 4](image2)

**Fig. 4** Transition structures obtained at the SMD(water)-B3LYP-D3/6-31G(d,p) level of theory for the intramolecular Br$^+$ transfer in N-brominated carnosine and the related systems shown in Fig. 3. Atomic color scheme: H white; C gray; N blue; O red; Br dark red.

The angle ($\alpha$) between the plane of the imidazole ring and the N–Br bond is given in degrees. The length of the N–Br bond and the COO$^-$–HN$^+$ H-bond are given in Å.
Effects of substituents on the imidazole ring

The results described above for N-brominated carnosine (and in ref. [37] for N-chlorinated carnosine) demonstrate that structural modifications of the β-alanyl-glycyl side chain can significantly reduce the barrier for the intramolecular X⁺ transfer (X = Cl and Br). In this Section we will consider substituent effects on the imidazole ring for achieving the same goal. Figure 5 shows the ring-substituted forms of N-halogenated carnosine that are considered in the present work. In these derivatives the carbon of the imidazole ring situated between the two nitrogen atoms is substituted with electron-accepting (R = CHO and NO₂) or electron-donating (R = OH and NH₂) groups.

Table 2 gives the Gibbs free-energy barriers (ΔG²E) and reaction free energies (ΔG₂F) for the overall reaction pathway for the systems in Fig. 5. Inspection of Table 2 shows that the reaction barriers decrease as the σ-electron-donating capability of the substituent increases, as reflected in the Hammett σₚ constants [62]. For the Cl⁺ transfer there is a linear correlation between the Gibbs-free activation energy and the σₚ values, with a squared correlation coefficient of R² = 0.9598. The highest reaction barrier of 130.9 kJ mol⁻¹ (R = NO₂) is reduced to 80.3 kJ mol⁻¹ for R = NH₂. A similar picture is obtained for the Br⁺ transfer, namely the highest barrier of 110.0 kJ mol⁻¹ (R = NO₂) is reduced to merely 53.2 (R = NH₂) kJ mol⁻¹. Thus, in both cases changing the R-group from a strong electron acceptor to a strong electron donor reduces the reaction barrier for the intermolecular X⁺ transfer by more than 50 kJ mol⁻¹.

Table 3 gives the APT charges on the N₁ and N₂ nitrogens of the imidazole ring and the X⁺ halide in TS2. The electron-donating or withdrawing capability of the substituent group has a significant effect on the charges of the imidazole nitrogens. The negative charges on N₁ and N₂ increase dramatically with the electron donating capability of the substituent group. For example, in the transition structure for the Cl⁺ transfer the negative charge on N₁ increases monotonically from -0.26 (R = NO₂) to -0.80 (R = NH₂), whilst the negative charge on N₂ increases from -0.31 (R = NO₂) to -0.69 (R = NH₂).

H-bond in TS2 is attributed to the greater conformational flexibility as the length of the β-alanyl-glycyl side chain increases.

### Table 2: Gibbs free-energy barrier heights (ΔG²E) and reaction free energies (ΔG₂F) for the overall reaction pathway at the SMD(water)-G4(MP2) level of theory for the intramolecular X⁺ transfer in the N-halogenated derivatives of carnosine shown in Fig. 5 (X = Cl and Br) (in kJ mol⁻¹)

<table>
<thead>
<tr>
<th>Rᵦ</th>
<th>Cl⁺</th>
<th>Br⁺</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔG²E</td>
<td>ΔG₂F</td>
<td>ΔG²E</td>
</tr>
<tr>
<td>NO₂ (0.78)</td>
<td>130.9</td>
<td>-69.5</td>
</tr>
<tr>
<td>CHO (0.42)</td>
<td>122.0</td>
<td>-57.1</td>
</tr>
<tr>
<td>H (0.0)</td>
<td>109.5ᵇ</td>
<td>-38.3ᵇ</td>
</tr>
<tr>
<td>OH (0.37)</td>
<td>100.1</td>
<td>-41.5</td>
</tr>
<tr>
<td>NH₂ (0.66)</td>
<td>80.3</td>
<td>-30.4</td>
</tr>
</tbody>
</table>

*ᵃσₚ constants are given in parenthesis
*ᵇTaken from ref. [37]

### Table 3: APT charges (in a.u.) on the N₁ and N₂ atoms of the imidazole ring and the halide atom in the transition structure for the intramolecular halide transfer (TS2)

<table>
<thead>
<tr>
<th>Rᵦ</th>
<th>Cl⁺</th>
<th>Br⁺</th>
</tr>
</thead>
<tbody>
<tr>
<td>N₁</td>
<td>N₂</td>
<td>X</td>
</tr>
<tr>
<td>NO₂ (0.78)</td>
<td>-0.26</td>
<td>-0.31</td>
</tr>
<tr>
<td>CHO (0.42)</td>
<td>-0.33</td>
<td>-0.37</td>
</tr>
<tr>
<td>H (0.0)</td>
<td>-0.48</td>
<td>-0.73</td>
</tr>
<tr>
<td>OH (0.37)</td>
<td>-0.76</td>
<td>-0.87</td>
</tr>
<tr>
<td>NH₂ (0.66)</td>
<td>-0.80</td>
<td>-0.80</td>
</tr>
</tbody>
</table>

For atom numbers see Fig. 1
*ᵇσₚ constants are given in parenthesis
to $-0.80 \ (R = NH_2)$ a.u. Similar trends are observed for the transition structure for the Br$^+$ transfer (Table 3). The general increase in the negative charge on the nitrogen atoms when going from electron-withdrawing to electron-donating substituents may partly explain the reduction in the barrier height in the same order. That is, pushing electron density into the positively charged imidazole ring should stabilize the transition structure. Inspection of Table 3 reveals that $\sigma$-electron-donating groups also have a significant effect on the atomic charge of the migrating $X^+$ center. In particular, for $X = Cl$ we obtain atomic charges of $+0.96 \ (H)$, $+0.82 \ (OH)$, and $+0.66 \ (NH_2)$ a.u. A similar trend is observed for $X = Br$, namely we obtain atomic charges of $+0.81 \ (H)$, $+0.69 \ (OH)$, and $+0.59 \ (NH_2)$ a.u. The reduction in the positive charge on the imidazole moiety may also contribute to stabilizing the transition structure for the halide transfer.

Overall, the APT charges in Table 3 indicate that electron-donating substituents push electron density into the $N_2$-$X^+$ bond, which is being broken in TS2. In particular, the sum of the charges on the $N_2$ and $Cl^+$ atoms are: $+0.65 \ (NO_2)$, $+0.59 \ (CHO)$, $+0.23 \ (H)$, $-0.04 \ (OH)$, and $-0.14 \ (NH_2)$ a.u. A similar trend is observed for the $N_2$ and $Br^+$ atoms. Thus, the sum of the atomic charges changes from a large positive charge (for $R = NO_2$ and CHO) to a slightly negative charge (for $R = OH$ and NH$_2$). We also note that there is a fairly good linear correlation between these charges and the Hammett substituent constants. In particular, we obtain a squared correlation coefficient of $R^2 = 0.9675$ for $X = Cl$ and $R^2 = 0.9740$ for $X = Br$.

Rational design of HOCl and HOBr antioxidants

In “Structural modifications” we have shown that the reaction barrier for the $X^+$ transfer is reduced by over 40 kJ mol$^{-1}$ by increasing the length of the $\beta$-alanyl-glycyl side-chain by one carbon. In “Effects of substituents on the imidazole ring” we have shown that these barriers can be reduced by over 40 kJ mol$^{-1}$ by substituting the imidazole ring with a strong electron-donating group. It is of interest to see whether these structural and electronic effects can be combined in the same system. Or in other words, are these effects additive? The ability to combine the two effects in an additive (or partially additive) manner would allow the rational design HOCl-antioxidants with very low barriers for the intramolecular $X^+$ transfer. As a proof-of-principle, we consider the $X^+$ transfer in a derivative of carnosine in which the length of the $\beta$-alanyl-glycyl side chain is increased by one carbon and the imidazole ring is substituted with a strong electron-donating amine group (Fig. 6). The reaction energies and barrier heights for the $X^+$ transfer for this antioxidant are presented in Table 4. We obtain very low reaction barrier heights of $\Delta G^{298}_{298} = 41.4 \ \text{and} \ 27.8 \ \text{kJ mol}^{-1}$ for the Cl$^+$ and Br$^+$ transfers, respectively. These barrier heights are lower by as much as 68.1 and 69.4 kJ mol$^{-1}$ for the Cl$^+$ and Br$^+$ transfers in carnosine, respectively.

Figure 7 shows the optimized transition structures for the intramolecular $X^+$ transfer in the electronically and structurally modified HOCl-antioxidant in Fig. 6. Inspection of the transition structures shows that the angle between the imidazole ring and the $N-X$ bond decreases in the order 41.5$^\circ$ (Cl$^+$) and 54.6$^\circ$ (Br$^+$). As discussed in “Structural modifications” these structural parameters indicate that the transition structure for the bromine transfer is earlier (and hence energetically more stable) relative to the transition structure for the chlorine transfer. In addition, the $\text{COO}^-$...$\text{HN}^+$ H-bond distance in the transition structure for the bromine transfer (1.671 Å) is slightly shorter than that in the transition structure for the chlorine transfer (1.714 Å). Thus, the trends in the Gibbs free-energy barriers may be associated with the formation of a more effective intramolecular hydrogen bond between the charged carboxylate and imidazole groups (see also discussion in “Structural modifications”).

Table 4 Gibbs free-energy barrier heights ($\Delta G^{298}_{298}$) and reaction energies ($\Delta G^{298}_{298}$) calculated at the SMD(water)-G4(MP2) level of theory for the intramolecular $X^+$ transfer in the $X$-halogenated derivative of carnosine shown in Fig. 6 ($X = Cl$ and Br) (in kJ mol$^{-1}$)

<table>
<thead>
<tr>
<th>$R^+$</th>
<th>Cl$^+$</th>
<th>Br$^+$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\Delta G^{298}_{298}$</td>
<td>$\Delta G^{298}_{298}$</td>
</tr>
<tr>
<td>Carnosine, $R = H$</td>
<td>109.5$^a$</td>
<td>-38.3$^a$</td>
</tr>
<tr>
<td>Homocarnosine, $R = NH_2$</td>
<td>41.4</td>
<td>-40.6</td>
</tr>
</tbody>
</table>

Values taken from ref. [37]
Conclusions

On the basis of our computational experiments, we propose a mechanism for the intramolecular Br⁺ shift in carnosine that emerges from a unique structural relationship between three adjacent functional groups, namely, the imidazole ring, carboxylic acid, and primary amine moieties. The Br⁺ transfer is found to be the rate-determining step, with a barrier of 97.2 kJ mol⁻¹ relative to the minimum-energy conformation of N-brominated carnosine. Structurally related systems that include the same key functional groups were also examined. We find that increasing the length of the β-alanyl-glycyl side-chain leads to a substantial reduction in the barrier height for the Br⁺ shift. These results are consistent with the results of previous theoretical and experimental investigations that considered the Cl⁺ transfer in carnosine and its derivatives.

We proceed to investigate electronic effects on the reaction barrier heights for the intramolecular Cl⁺ and Br⁺ shifts. We find that substitution of the imidazole ring with electron-donating substituents leads to substantial reductions in the barrier heights for both the Cl⁺ and Br⁺ shifts.

Finally, we show that the above structural and electronic effects are largely additive. For example, a derivative of carnosine, in which the length of the β-alanyl-glycyl side chain is increased by one carbon and the imidazole ring is substituted with a strong electron-donating NH₂ group, results in barriers of 41.4 and 27.8 kJ mol⁻¹, respectively, for the Cl⁺ and Br⁺ transfers. These findings pave the way for the rational design of carnosine-based antioxidants with very low reaction barrier heights for the intramolecular Cl⁺ and Br⁺ shifts and therefore to antioxidants with potentially greater HOCl and HOBBr-scavenging capabilities. We hope that our computational investigation will inspire further experimental investigations of the antioxidant activity of the proposed antioxidants.

Acknowledgements This work is dedicated to our colleague and friend Dr. Ming Wen Shi, who tragically passed away earlier this year. This research was undertaken with the assistance of resources from the National Computational Infrastructure (NCI), which is supported by the Australian Government. We also acknowledge the system administration support provided by the Faculty of Science at the University of Western Australia to the Linux cluster of the Karlon group. We gratefully acknowledge the provision of an Australian Postgraduate Award (to F.S.), and an Australian Research Council (ARC) Discovery Early Career Researcher Award (to A.K., Project No. DE140100311). We would also like to thank the reviewers of the manuscript for their valuable comments and suggestions.

References

2.2.2 Mechanistic insights into the water-catalysed ring-opening reaction of vitamin E by means of double-hybrid density functional theory

Presentation of the article

Title
Mechanistic insights into the water-catalysed ring-opening reaction of vitamin E by means of double-hybrid density functional theory.

Authors
Sarrami, F., Kroeger, A.A. and Karton, A

Journal
Chemical Physics letters 2018. 708, 123-129.

DOI
https://doi.org/10.1016/j.cplett.2018.07.036

Date of Publication
September 2018

Graphical TOC
Mechanistic insights into the water-catalysed ring-opening reaction of vitamin E by means of double-hybrid density functional theory

Farzaneh Sarrami, Asja A. Kroeger, Amir Karton*

School of Molecular Sciences, The University of Western Australia, Perth, WA 6009, Australia

ARTICLE INFO

Article history:
Received 23 June 2018
In final form 14 July 2018
Available online 26 July 2018

Abstract

The potent antioxidant α-tocopherol is known to trap two hydroxyl radicals leading to the formation of the experimentally observed α-tocopherylquinone product. Based on double-hybrid density functional theory calculations, we propose for the first time, a reaction mechanism for the conversion of α-tocopherol to α-tocopherylquinone. We find that a water-catalysed ring-opening reaction plays a key role in this conversion. The water catalysts act as proton shuttles facilitating the proton transfers and reducing the ring strain in the cyclic transition structures. On the basis of the proposed reaction mechanism, we proceed to design an antioxidant with potentially enhanced antioxidant properties.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Oxidative damage from reactive oxygen species such as the hydroxyl or hydroperoxyl radical to DNA, proteins, and lipids can lead to a broad range of chronic diseases such as Parkinson’s and Alzheimer’s disease [1–4]. Biological systems protect themselves from such oxidative damage with radical scavenging antioxidants, of which some are hydrophilic and some are lipophilic [5–7]. Vitamin E is an important lipid-soluble antioxidant occurring in lipoproteins and membranes [8,9]. It may exist in a number of different forms, each containing a chemically reactive hydroxyl group that is responsible for antioxidant activity and an unreactive lipophilic alkyl side-chain [10]. The most active form of vitamin E, α-tocopherol (AT), is illustrated in Fig. 1. Numerous studies on the biological activities of this antioxidant have found that it plays an important part in disease prevention. Research by Maydan showed that α-tocopherol may inhibit oxidation of low-density lipoprotein (LDL) cholesterol in plasma, thereby supporting the treatment of cardiovascular diseases [11,12]. It was further found to reduce cancer cell proliferation in breast cancers [13] and has shown anti-androgen activity in prostate carcinoma cells, making it a potent chemopreventive agent of androgen-dependent diseases [14]. These promising results have encouraged a wide range of experimental as well as theoretical studies investigating the antioxidant activity of AT aimed at gaining a better understanding on its mechanism of action [7,15–17].

It was found that a single molecule of AT may trap two hydroxyl radicals leading to the major oxidation product α-tocopherylquinone (ATQ) [18]. Fig. 2 illustrates the proposed mechanism of this reaction. A hydrogen abstraction from the hydroxyl group by the first hydroxyl radical leads to the tocopherol radical. Tocopherol radicals are resonance-stabilised free radicals that are important intermediates in the antioxidant activity of vitamin E, since they may participate in radical-radical termination reactions. Addition of a second hydroxyl radical to the carbon neighboring the endocyclic oxygen and subsequent ring-opening lead to the α-tocopherylquinone (ATQ) product.

A number of computational studies investigated the reaction of AT with a single hydroxyl or hydroperoxyl radical to give α-tocopherone [19–21]. Navarrete et al. [19] investigated the hydrogen abstraction from the phenolic hydrogen by hydroxyl and hydroperoxyl radicals [20]. Chen et al. [21] considered the bond dissociation energies associated with the first hydrogen abstraction of AT and structure function relationships in related structures. They showed that reducing the size of the aromatic ring is likely to give better antioxidant activity. Replacing the oxygen atom in AT with S or Se, on the other hand, appeared to reduce antioxidant activity. Wright et al. [22] investigated substituent effects on the O–H bond strength of phenolic antioxidants related to vitamin E. To the best of our knowledge, the mechanism of the second radical scavenging process leading to the final ATQ oxidation product has not been studied computationally. Here we use high-level double-hybrid DFT (DHDFT) methods to investigate the reaction pathway of the rearrangement of α-tocopherone to form ATQ. We consider both the uncatalysed and water-catalysed reaction pathways. Based on our proposed reaction
mechanism, we proceed to design vitamin E-related antioxidants with improved antioxidant properties.

2. Computational details

The geometries of all structures were optimized using the B3LYP-D3 density functional theory (DFT) exchange-correlation functional in conjunction with the 6-31G(2df,p) Pople-style basis set [23–25]. Empirical D3 dispersion corrections [26] were included using the Becke-Johnson [27] damping potential as recommended in Ref. [25] (denoted by the suffix-D3). Bulk solvent effects in aqueous solution were included using the charge-density-based CPCM polarizable solvation model [28]. Harmonic vibrational analyses have been performed at the same level of theory to confirm each stationary point as either an equilibrium structure (i.e., all real frequencies) or a transition structure (i.e., with one imaginary frequency). Zero-point vibrational energies and thermal correction to the enthalpy have been obtained from these calculations. We note that the main text primarily looks at $\Delta H_{298}$ values to allow a discussion of the energetic effects in the uncatalysed and catalysed reaction mechanisms, however, similar trends are observed on the Gibbs free energy ($\Delta G_{298}$) reaction profiles, which are given in Figs. S1–S3 and Table S1 of the Supporting Information. The connectivity of the local minima and first-order saddle points was confirmed by performing intrinsic reaction coordinate calculations [29]. High-level DHF/D calculations using the spin-component-scaled DSD-PBEP86-D3 functional were performed in order to obtain accurate electronic energies for the equilibrium and transition structures located along the uncatalysed and catalysed reaction pathways [30]. These DHF/D calculations involve second-order Møller–Plesset perturbation theory to overcome limitations of traditional DFT methods and display excellent performance for challenging chemical problems [31–35]. DHF/D has been found to produce thermochemical properties (such as reaction energies and barrier heights) with mean absolute deviations (MADs) approaching the threshold of 'chemical accuracy' (arbitrarily defined as 4.2 kcal mol$^{-1}$). The DHF/D calculations, which inherit the slow basis-set convergence of MP2, are carried out with the quadruple-zeta Def2-QZVP basis set [36]. Corrections for bulk solvent effects in aqueous solution were added to the gas-phase DHF/D energies using the CPCM model at the HF/6-31+G(d) level of theory as recommended in Ref. [37]. The resulting level of theory is denoted by CPCM(water)-DSD-PBEP86. The Gaussian16 suite of programs was used for all the DFT and DHF/D calculations [38].

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.cplett.2018.07.036.

3. Results and discussion

We begin by considering the possible reactions of the tocopherol radical with the hydroxyl radical. Fig. 2 shows the four resonance structures of the tocopherolyl radical along with the
four possible products of the radical coupling reaction (denoted A, B, C, and D). Table 1 gives the reaction enthalpies for these four species relative to the free reactant (AT). As expected, intermediates B, C, and D are vastly favoured over intermediate A, which contains a reactive peroxide group. Intermediate C, the only structure that can undergo rearrangement to ATQ, is thermodynamically more stable than intermediates B and D by 21.2–21.8 kJ mol$^{-1}$. We note that similar trends are observed on the $\Delta G_{298}$ potential energy surface (for further details see Table S1). In the following sections we will consider the uncatalysed and water-catalysed rearrangement reactions for the formation of ATQ from intermediate C.

3.1. Rearrangement reaction of $\alpha$-tocopherol to quinone

3.1.1. Uncatalysed reaction mechanism

The reaction profile for the uncatalysed ring-opening of $\alpha$-tocopherol is given in Fig. 3. The reaction mechanism involves a proton transfer from O to O, concomitant with the breaking of the Cl–O bond and the formation of a double bond between C and O. This concerted ring-opening reaction is associated with a fairly high activation enthalpy of 159.9 kJ mol$^{-1}$, which is largely attributed to the highly strained four-membered cyclic transition structure. The ATQ product lies 7.1 kJ mol$^{-1}$ below $\alpha$-tocopherol. We note that on the $\Delta G_{298}$ potential energy surface the reaction is exergonic by 18.7 kJ mol$^{-1}$ and the activation Gibbs free energy is 159.9 kJ mol$^{-1}$ (for further details see Fig. S1).

3.1.2. Water-catalysed reaction mechanism

The participation of catalytic water molecules in reactions involving proton transfers is well documented and can significantly reduce activation barriers by reducing strain energies involved in the TS [39-45]. Here, we examine the catalytic effect of one and two explicit water molecules on the rearrangement reaction of $\alpha$-tocopherol to ATQ. The reaction profile for the H$_2$O-catalysed reaction is shown in Fig. 4 (red line). In the complex (RC), the water molecule is hydrogen bonded to the hydroxyl group of the tocopherone via one hydrogen bond (1.806 Å). As a result, the reactant complex is more stable than the free reactants by 25.5 kJ mol$^{-1}$. Participation of water in the transition structure reduces the activation enthalpy of the rearrangement by 58.1 kJ mol$^{-1}$, from 159.9 (uncatalysed) to 101.8 (H$_2$O-catalysed) kJ mol$^{-1}$ by expanding the transition structure from a four-membered ring TS to a less strained six-membered ring TS. We note that the activation energy for the catalysed reaction is 127.3 kJ mol$^{-1}$ relative to the RC. The product complex (PC) is stabilised by two hydrogen bonds, one between water and quinone (1.859 Å) and a longer hydrogen bond (1.915 Å) between the newly formed hydroxyl group and water (Fig. 4). As a result the PC is 32.6 kJ mol$^{-1}$ more stable than the free reactants.

The inclusion of two explicit water molecules expands the hydrogen bond network facilitating the proton transfer. As illustrated in Fig. 4, in the RC the two water molecules form a bridge between the exocyclic hydroxyl group and the oxygen of the chromosome moiety via three hydrogen bond interactions. The lengths of these hydrogen bonds are 1.781 (H–H–O), 1.781 (H–H–O), and 1.916 (H–H–O) Å (Fig. 4). Accordingly, the RC lies 59.0 kJ mol$^{-1}$ below the energy of the free reactants. In the subsequent transition structure the two water molecules act as proton shuttles, facilitating the proton transfer via an eight-membered cyclic TS. The TS lies 110.3 kJ mol$^{-1}$ above the RC. Thus, compared to the uncatalysed pathway, participation of two water molecules decreases the activation energy towards the rearrangement by as much as 49.6 kJ mol$^{-1}$. Similarly to the RC, the PC formed between the product and the two water catalysts involves three strong hydrogen bonds (1.790 (H–H–O), 1.750 (H–H–O), and 1.786 (H–H–O) Å, Fig. 4) and lies 68.2 kJ mol$^{-1}$ below the energy of the free reactants. Finally, we note that on the Gibbs free energy surface similar reaction barriers are obtained for the reactions catalysed by one and two water molecules (Fig. S2 of the Supporting Information). This is consistent with previous results for water-catalysed proton transfers, which show that on the $\Delta G_{298}$ surface the most significant reduction in the barrier height is provided by the first water catalyst and inclusion of a second water catalyst has a minor effect [32].

Further inspection of the transition structures of the catalysed and uncatalysed pathways, as depicted in Fig. 5, indicates significant changes in geometries upon catalysis. Most notably, the inclusion of explicit catalytic water molecules improves the trajectories of the proton transfers. While the transition structure of the uncatalysed reaction leads to a highly unfavourable angle of \( \alpha = \angle \text{O} - \text{H} - \text{O} = 125.6^\circ \) for the first proton transfer, this improves to \( 158.0^\circ \) in the H$_2$O-catalysed pathway. Near-ideal angles of \( 171.8^\circ \) (\( \angle \text{O} - \text{H} - \text{O} \), \( 165.3^\circ \)) (\( \angle \text{O} - \text{H} - \text{O} \)), and \( 176.6^\circ \)) (\( \angle \text{O} - \text{H} - \text{O} \)) are observed in the H$_2$O-catalysed pathway (Fig. 5). It is well known that in reactions at the carbonyl group an ideal angle of the incoming or leaving substituent relative to the carbonyl improves the stabilising overlap of the highest occupied molecular orbital (HOMO) of the nucleophile and the unoccupied carbonyl $^\pi$ orbital while minimising steric interactions. Deviation from the ideal trajectory of \( \approx 107^\circ \)–\( 110^\circ \), leads to a rapid increase in activation energies [46]. While in the transition structure of the uncatalysed pathway this trajectory is highly unfavourable with an angle of \( \alpha = \angle \text{O} - \text{C} - \text{O} = 93.3^\circ \),

![Fig. 3](image_url)

**Fig. 3.** Reaction profile ($\Delta H_{298}$, CCDC/water-D3D-PRB98, kJ mol$^{-1}$) for the uncatalysed rearrangement reaction of $\alpha$-tocopherol to ATQ. The bonds being broken and formed in the transition structures are represented by black dashed lines.

<table>
<thead>
<tr>
<th>Intermediate</th>
<th>$\Delta H_{298}$ (kJ mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-114.2</td>
</tr>
<tr>
<td>B</td>
<td>-281.3</td>
</tr>
<tr>
<td>C</td>
<td>-302.5</td>
</tr>
<tr>
<td>D</td>
<td>-280.7</td>
</tr>
</tbody>
</table>

*Table 1: Reaction enthalpies ($\Delta H_{298}$) calculated at the CCDC/water-D3D-PRB98 level for the four possible products (A, B, C, and D) of the radical coupling reaction between the tocopherol radical and H$_2$O (Fig. 2). Enthalpies in kJ mol$^{-1}$ are given relative to the free reactants.*
participation of the water catalysts allows for this angle to increase to $\angle \text{O}^2-\text{C}^1-\text{O}^3 = 105.3^\circ$ (1 H$_2$O catalyst) and $\angle \text{O}^2-\text{C}^1-\text{O}^3 = 111.0^\circ$ (2 H$_2$O catalysts). Further, it is instructive to note that the opening six-membered heterocyclic ring assumes favourable chair-conformations in the transition structures of the water-catalysed pathways, whereas a twisted boat-like geometry is observed for the uncatalysed pathway.

3.2. Rational design of vitamin E based antioxidants

It is well known that the presence of a saturated heterocycle is essential for the antioxidant activity of chromanols [47]. Several previous studies aimed at designing novel antioxidants have found that changing the atom number of the heterocyclic ring as well as substituting the heteroatom of the chromanol moiety affects antioxidant activity with regard to the first radical scavenging reaction. There is general agreement in the literature that reducing the number of atoms in the heterocyclic ring has the potential to improve antioxidant activity [21,48,49]. In order to test if a similar trend can be observed for the scavenging of a second hydroxyl radical, derivatives of α-tocopherol with a modified heterocyclic ring are considered in this work and depicted in Scheme 1.

We begin with examining the effect of the size of the heterocyclic ring in order to reduce the reaction barrier for the intramolecular ring-opening step. Fig. 6 shows the enthalpic energy surfaces for the rearrangements of candidate 1.
Scheme 1. Structures of the α-tocopherol derivatives investigated in this work. The number in brackets indicates the number of atoms in the heterocyclic.

Fig. 6. Reaction profile (ΔH_{298} (PCM/water) DSD-PPSEPS, kJ mol^{-1}) for the uncatalysed (black line), H_2O-catalysed (red line), and 2H_2O-catalysed (blue line) rearrangement reaction in the proposed antioxidant candidate 1 (Scheme 1). A schematic representation of the TS is shown in which bonds being broken and formed are represented by dashed lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Similar to the findings presented in Section 3.1, participation of catalytic water molecules as proton shuttles reduces the energy barriers towards ring-opening. In particular, the barrier height for the H_2O-catalysed reaction is reduced from 127.3 (α-tocopherone) to 89.2 (candidate 1) kJ mol^{-1}, note that these reaction barrier heights are taken relative to the RCs (Figs. 4 and 6). The reduction in the barrier height is less pronounced for the reaction catalysed by two water molecules. Namely, the barrier heights are 110.3 (α-tocopherone) and 87.3 (candidate 1) kJ mol^{-1}. We note that similar results are obtained on the Gibbs free energy surface, that is inclusion of a second water catalyst does not provide further catalytic enhancement (see Fig. 53 of the Supporting Information).

An important aspect in the improved antioxidant activity of candidate 1 is that it provides a strong thermodynamic driving force for the overall reaction. The ring-opening reaction is exothermic by 83.6 (uncatalysed) and 65.6 (H_2O catalyst), and 61.3 (2 H_2O catalysts) kJ mol^{-1} (Fig. 6). These reaction enthalpies should be compared to reaction enthalpies of merely 7–9 kJ mol^{-1}
obtained for the uncatalysed and catalysed ring-opening reactions in \( \alpha \)-tocopherone (Figs. 3 and 4).

Figs. 54 and 55 of the Supporting Information give the reaction profiles for the uncatalysed, \( \text{H}_2\text{O} \)-catalysed, and \( 2\text{H}_2\text{O} \)-catalysed rearrangements for candidates 2 and 3 (Scheme 1). These candidates, however, do not lead to a significant improvement over \( \alpha \)-tocopherone in either the reaction barrier height or in the thermodynamic driving force for the reaction.

4. Conclusions

The antioxidant \( \alpha \)-tocopherol (a.k.a. vitamin E) is known to react with two OH radicals to form \( \alpha \)-tocopherylquinone. The first reaction with OH is a hydrogen-abstraction reaction which forms a tocopheroxyl radical. This reaction has been the subject of theoretical investigations. However, the subsequent reaction leading to the formation of \( \alpha \)-tocopherylquinone has not been studied computationally. On the basis of DFT computational modelling, we propose a mechanism for the intramolecular ring-opening reaction of \( \alpha \)-tocopherol leading to the final \( \alpha \)-tocopherylquinone. We find that the uncatalysed intramolecular ring-opening proceeds via a strained four-membered transition structure. Accordingly, this reaction is associated with a high activation enthalpy of 159.9 kJ mol\(^{-1}\). Involvement of one and two water catalysts in the TS reduce the reaction barrier height to 127.3 and 110.3 kJ mol\(^{-1}\), respectively. This catalytic efficiency can be rationalised by (i) the reduction in strain energy in the TSs in the order uncatalysed, 1 \( \text{H}_2\text{O} \) catalyst, and 2 \( \text{H}_2\text{O} \) catalysts, and (ii) improved trajectories for the proton transfers from the hydroxyl group to the heterocyclic oxygen.

We find that the proposed reaction mechanism also applies to structurally-related systems in which the size of the heterocyclic ring is reduced and the heterocyclic oxygen is replaced with a sulfur atom. Our results show that reducing the size of the heterocyclic ring of the \( \alpha \)-tocopherone to a four-membered ring not only reduces the barrier for the ring-opening reaction but also significantly increases the thermodynamic driving force for the overall reaction. In particular, for the four-membered-ring analogue of \( \alpha \)-tocopherol we obtain reaction barrier heights of 89.2 and 87.3 kJ mol\(^{-1}\) for the reactions catalysed by one and two water molecules, respectively (cf. with reaction barriers of 127.3 and 110.3 kJ mol\(^{-1}\) for \( \alpha \)-tocopherol). Importantly, the reaction with \( \alpha \)-tocopherol is exothermic by merely 7.1–9.2 kJ mol\(^{-1}\) (for the uncatalysed and catalysed reactions), however, for the four-membered-ring analogue the reaction is exothermic by 61.3–83.6 kJ mol\(^{-1}\) (for the uncatalysed and catalysed reactions). The significant reduction in the reaction barrier height and increase in the reaction exothermicity suggest that the four-membered-ring analogue of \( \alpha \)-tocopherol could be a more efficient antioxidant than vitamin E for neutralising two equivalents of OH radicals.

Acknowledgments

This research was undertaken with the assistance of resources from the National Computational Infrastructure (NCI), which is supported by the Australian Government. We also acknowledge the system administration support provided by the Faculty of Science at the University of Western Australia to the Linux cluster of the Karion group. We gratefully acknowledge the provision of an Australian Postgraduate Award (to F.S.), a Forrest Research Foundation Scholarship and an Australian Government Research Training Program Stipend (to A.A.K.), and an Australian Research Council (ARC) Future Fellowship (to A.K.; Project No. FT170100373).

References


Chapter 3: Atmosphere chemistry application

3.1 Introduction

The importance of hydrocarbons in the Earth’s troposphere has been recognized since the 1950s, when their contribution (along with that of NOx i.e., NO and NO₂ to urban (Los Angeles) “smog” formation was first outlined. While air quality in urban areas is still a major issue, it is also now recognized that about 90% of the hydrocarbons emitted to the atmosphere arise from biogenic sources. It is also now recognized that the impacts of hydrocarbon chemistry are not limited to the Earth’s surface, but also reach to the upper troposphere. Thus, it is clear that there is a great need to study hydrocarbon chemistry in detail.

Organic aerosols play an important role in global climate and human health. The physical and chemical processing of volatile gas-phase hydrocarbons leads to the significant fraction of organic aerosol material, which leads to the less volatile products that condense in the particulate phase; this is referred to as secondary organic aerosol (SOA). There is growing evidence suggesting that, like sulfate, SOA is formed through aqueous-phase reactions in clouds, fogs and aerosols. The prediction of the growth of atmospheric organic aerosols (OA) represents a major challenge in current modelling efforts due to the high uncertainties in sources, properties, reaction pathways and product volatility distribution of organics. In particular, the formation of OA from reaction products of volatile precursors is insufficiently described in models and their amount is underestimated by factors of 2–100. The reasons for this discrepancy could both be due to the incomplete knowledge of precursors or insufficient understanding of the processes that convert organic gas phase precursors into low-volatility products that contribute to SOA formation.
As most of the experimental values for intermediates of ozonolysis path have uncertainty, much of the understanding ozonolysis mechanism comes indirectly from measurements of stable products interfaced with computational studies for potential energy surface mapping. The following section gives an overview of computational investigation into the two catalytic reaction mechanisms in the atmospheric system, which may contribute to SOA formation. I hope that our computational investigation will inspire further experimental investigations of the new catalyst in the atmospheric reaction pathways, which can inhibit SOA formation.

### 3.1.1. Introduction to the sulphuric acid catalysed 1,4-hydrogen transfer reaction in Criegee intermediates

Criegee intermediates (CIs) are carbonyl oxides formed in the ozonolysis of unsaturated hydrocarbons; their atmospheric fate plays a critical role in determining the oxidative efficacy of the atmosphere and its capacity for free radical generation and secondary organic aerosol formation.

Despite extensive experimental work on the ozonolysis mechanism in the atmosphere, considerable uncertainty remains due to the extremely short lifetimes of Criegee intermediate species. CIs have been experimentally detected from the gas-phase ozonolysis of alkenes, however much of our understanding of the ozonolysis mechanism comes indirectly from measurements of stable products interfaced with computational studies for potential energy surface mapping. The CI can experience a few unimolecular processes, most notably, decomposition into a hydroxyl radical (OH) to form the stabilized Criegee intermediate (SCI). In addition to their role as sources of OH, exploring the unimolecular reactions of Criegee intermediates with water is especially critical as they are the key processes in the formation of aerosols, which scatter sunlight and affect earth’s radiative balance. In particular, the oxidation of sulfur dioxide via stabilized Criegee intermediates is an important source of sulfuric acid, which plays an important role in aerosol nucleation.
Due to water being the dominant trace species in the troposphere, its potential catalytic influence on the bimolecular chemistry of CIs has been examined numerous times in the literature (for a recent review see Vereecken et al.\textsuperscript{202} and references therein). The reaction of CI with water is predicted to predominantly form hydroxyl-alkyl hydroperoxides, however, research has shown that a water molecule can catalyse a H-migration to form a VHP in terpene-derived CI.\textsuperscript{203,204,205}

In the present work we use the high-level ab initio G4(MP2) procedure\textsuperscript{206} to study the sulphuric acid-catalysed unimolecular 1,4 H-shift reactions in methyl CI and in isoprene derived and α-pinene derived CIs. (Figure 3.1) To make comparisons with the previous theoretical investigations, the uncatalysed and water-catalysed reactions are also considered using the same theoretical description. The reaction barrier heights for the sulphuric acid-catalysed reactions are 16.4 (methyl CI), 24.5 (isoprene CI), and 8.4 (α-pinene CI) kJ mol\textsuperscript{-1}, relative to the reactant complexes. Where in all cases the energies of the transition structures are well bellow the energies of the isolated CI and H\textsubscript{2}SO\textsubscript{4} catalyst. Conversely, water was found to exhibit only a minor catalytic effect for these 1,4 H-shift reactions. We find that a sulphuric acid catalyst affords a barrierless reaction pathway for the 1,4-H tautomerisation in CI, suggesting the pathway to be competitive with unimolecular decay and reaction with water loss processes in regions with high sulphuric acid concentrations. We note that, according to the Arrhenius equation, a change of 5.7 kJ mol\textsuperscript{-1} in the barrier corresponds to a change of 1 order of magnitude in the reaction rate at 298 K.

**Figure 3.1** Formation of vinylhydroperoxides (VHPs) via ozonolysis followed by a 1,4 H-shift. Attack of ozone on the C=C double bond leads to a primary ozonide (POZ), which rapidly decomposes, yielding Criegee Intermediates (CIs), then a 1,4 H-shift from the syn methyl group leads to the formation of VHPs.
3.1.2 Introduction to the sulphuric acid-catalysed formation of hemiacetal from glyoxal and ethanol

Recent studies suggest that there are large, unidentified pathways of SOA formation both in urban areas and in the free troposphere. Formaldehyde, acetaldehyde, and glyoxal are three of the most abundant tropospheric aldehydes. None of these compounds would form much aerosol in the atmosphere due to their high vapour pressures and gas-phase reactivities. Over the past few years, glyoxal has gained great attention as a possible SOA precursor in the aqueous phase. Because of the large number of reactions that can take place leading to different products and the difficulty in isolating intermediates, it can be a challenge to evaluate the viability of different reaction mechanisms. Computational chemistry is a powerful tool in this regard because we can calculate both the energetic stability of a wide range of potential intermediates and the activation energy barriers for different mechanistic pathways.

Glyoxal, the simplest α-dicarbonyl, is formed in the atmosphere through the oxidation of volatile organic compounds, which are emitted via biogenic and anthropogenic processes. Glyoxal can react in the atmosphere with itself and with other small molecules, such as water and alcohols to form oligomers that are involved in SOA formation. One potential mechanism for this can be explained that aldehyde functional groups can further react in the aerosol phase via hemiacetal/acetal formation with alcohols. Acid catalysis may also be an important mechanism for accelerating glyoxal reactions in this scenario. A principal candidate for atmospheric acid catalyst is sulfuric acid that produced through the oxidation of SO2 emitted from fossil fuel combustion. The conversion of glyoxal to hemiacetal is of particular importance since the enhanced hydrogen bonding associated with the polar OH groups in the hemiacetal significantly lowers the vapour pressure compared to the parent glyoxal molecule. Therefore, hemiacetal can partition more easily into the particle phase and contribute to the growth of SOAs.

The goal of our current study is to use the high-level ab initio G4(MP2)
procedure to investigate the most thermodynamically favourable reaction pathways and intermediates in the catalytic reaction mechanism of glyoxal with ethanol to form hemiacetal. We consider both water and sulphuric acid catalysts. We find that a sulphuric acid catalyst can significantly reduce the barrier for this reaction. These results are consistent with previous computational investigations that showed that sulphuric acid can effectively catalyse proton shifts. However, whereas the previous studies focused on intramolecular proton transfers, the present work shows that H$_2$SO$_4$ can efficiently catalyse an intermolecular proton transfer which is coupled with the formation of a covalent C–O bond between glyoxal and ethanol. The present findings provide valuable insights into the catalytic properties of sulphuric acid and suggest a potential new pathway for the formation of hemiacetal in the gas phase.