HERIOT-WATT UNIVERSITY

Advances in the Theoretical Determination of Molecular Structure with Applications to Anion Photoelectron Spectroscopy

by

Zibo G. Keolopile

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in the
School of Engineering and Physical Sciences
Institute of Chemical Sciences

December 2013

ACADEMIC REGISTRY
Research Thesis Submission

Name:	Zibo Goabaone Keolopile		
School/PGI:	Engineering and Physical Sciences		
Version: (*i.e, Frst, Resubmission, Final*)	First	Degree Sought (Award **and** Subject area)	PhD Theoretical and Computational Chemistry

I, ZIBO G. KEOLOPILE, declare that this thesis titled, 'ADVANCES IN THE THEORETICAL DETERMINATION OF MOLECULAR STRUCTURES WITH APPLICATION TO ANION PHOTOELECTRON SPECTROSCOPY' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.

- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.

- Where I have consulted the published work of others, this is always clearly attributed.

- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.

- I have acknowledged all main sources of help.

- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

"Quantum mechanics describes nature as absurd from the point of view of common sense. And yet it fully agrees with experiment. So I hope you can accept nature as She is - absurd."

Richard P. Feynman

Abstract

Doctor of Philosophy

by Zibo G. Keolopile

This Dissertation is focussed primarily on development of methods aiming at the determination of molecular structures with application to systems with intra and intermolecular hydrogen bonds. I have developed and demonstrated usefulness of Potential Energy Surface Scanning Tool (PESST) by performing a systematic search for the most stable structures of neutral and anionic phenylalanine and tyrosine molecules using electronic structure methods. I have found out that tautomers resulting from the proton transfer from the carboxylic OH to phenyl ring determine the structure of the most stable anions of phenylalanine, but double proton transfer from the carboxylic and hydroxyl groups determine structures of the most stable anions of tyrosine. The most stable conformer of these valence anions remained adiabatically unbound with respect to the canonical neutral in case of phenylalanine but bound in case of tyrosine. Valence anions identified in this report have recently been observed experimentally.

Acetoacetic acid (AA), equipped with neighbouring carboxylic and keto groups, is a promising system for studies of intramolecular proton transfer. The results of my computational search for the most stable tautomers and conformers of the neutral and anionic AA were used to interpret anion photoelectron and electron energy-loss spectroscopy measurements. The valence anion was identified in photoelectron spectroscopy experiments and the measured electron vertical detachment energy is in good agreement with my computational predictions. My computational results allow rationalizing these experimental findings in terms of the co-existence of various conformers of AA.

I considered stability of dimers formed by molecules that can exist in different conformational states. I have developed a protocol that allows the dissection of the total stabilisation energy into one-body conformational and deformational components and the two-body interaction energy term. Interplay between these components determines the overall stability of the dimer. The protocol has been tested on the dimers of oxalic acid. The global minimum stability results from a balancing act between a moderately attractive two-body interaction energy and small repulsive one-body terms. I have analysed zero-point vibrational corrections to the stability of various conformers of oxalic acid and their dimers. I have found that minimum energy structures with the most stabilising sets of hydrogen bonds have the largest zero-point vibrational energy, contrary to a naive anticipation based on red shifts of OH stretching modes involved in hydrogen bonds.

My computational results demonstrated an unusual electrophilicity of oxalic acid (OA), the simplest dicarboxylic acid. The electrophilicity results primarily from the bonding carbon-carbon interaction in the SOMO orbital of the anion, but it is further enhanced by intramolecular hydrogen bonds. The well-resolved structure in the photoelectron spectrum has been reproduced theoretically, based on Franck-Condon factors for the

vibronic anion→neutral transitions. The excess electron binding energies in the dimer and trimer of OA become very significant due to intermolecular proton transfer, with the corresponding vertical detachment energy (VDE) values of approximately 3.3 and 4.6 eV. I have postulated a mechanism of excess electron mobility along molecular linear chains supported by cyclic hydrogen bonds.

Searches for the most stable molecular conformer are frustrated by energy barriers separating minima on the potential energy surface (PES). I have suggested that the barriers might be suppressed by subtracting selected force field terms from the original PES. The resulting deformed PES can be used in standard molecular dynamics (MD) or Monte Carlo simulations. The MD trajectories on the original and deformed PESs of ethanolamine differ markedly. The former gets stuck in a local minimum basin while the latter moves quickly to the global minimum basin.

Acknowledgements

Contents

List of Figures

List of Tables

Abbreviations

AA	**A**cetoacetic **A**cid
AEA	**A**diabatic **E**lectron **A**ffinity
ADZ	**A**ugumented **D**ouble **Z**eta
ADZ	**A**ugumented **T**riple **Z**eta
CBS	**C**omplete **B**asis **S**et
DFT	**D**ensity **F**unctional **T**heory
EBE	**E**lectron **B**inding **E**nergy
EKE	**E**lectron **K**inetic **E**nergy
FA	**F**ormic **A**cid
FC	**F**rank–**C**ondon
FF	**F**orce **F**ield
GM	**G**lobal **M**inimum
GOT	**G**aussian **O**utput **T**ools
HOMO	**H**ighest **O**ccupied **M**olecular **O**rbital
LM	**L**ocal **M**inima
LUMO	**L**owest **U**noccupied **M**olecular **O**rbital
MC	**M**onte-**C**arlo
MD	**M**olecular **D**ynamics
OA	**O**xalic **A**cid
Phe	**Ph**enylalanine
PES	**P**otential **E**nergy **S**urface
PESST	**P**otential **E**nergy **S**urface **S**canning **T**ool
QEq	**Ch**arge **Eq**ulibration
RMSD	**R**oot **M**ean **S**quare **D**eviation
SOMO	**S**ingly **O**occupied **M**olecular **O**rbital

SSC **S**ystematic **S**creening **C**onformers

TS **T**ransition **S**tate

Tyr **Tyr**osine

VDE **V**ertical **D**etachment **E**nergy

Physical Constants

Speed of Light	c	$=$	$2.997\ 924\ 58 \times 10^8$ ms^{-1} (exact)
Planck constant	h	$=$	$6.62606957 \times 10^{-34}$ m^2 kg / s
Reduced Planck constant	$\hbar$	$=$	$1.054571726(47) \times 10^{-34}$ m^2 kg / s
Photon	hc	$=$	$1.98644568 \times 10^{25} Jm$

Symbols

r	distance	m
ω	angular frequency	rads^{-1}
μ	reduced mass	kg
E	energy	kcal/mol
$\tilde{E}$	energy	cm^{-1}
I	moment of inertia	kg m^2
T	kinetic energy	J

Dedicated to my family, Saodah, Goabaone and friends.

List of Papers

1 Zibo G. Keolopile, Maciej Gutowski, Maciej Haranczyk,
  Discovery of Most Stable Structures of Neutral and Anionic Phenylalanine through Automated Scanning of Tautomeric and Conformational Spaces, *The Journal of Chemical Theory and Computation* doi:10.1021/ct400531a

2 Zibo G. Keolopile, Maciej Gutowski, Maciej Haranczyk,
  Adiabatically Bound Valence Anion of Parent Tyrosine: A Remarkable Stability and Unusual Structure, *in preparation*

3 Zibo G. Keolopile, Maciej Gutowski, (Heriot-Watt University)
  Angela Buonaugurio, Evan Collins, Xinxing Zhang, Jeremy Erb, Thomas Lectka, Kit Bowen (John Hopkins University)
  Michael Allan (Switzerland)
  Intramolecular Proton Transfer in Acetoacetic Acid Induced by a π^* Excess Electron, *in preparation*

4 Zibo G. Keolopile, Matthew R. Ryder, Maciej Gutowski,
  Submitted to *The Journal of Computational Chemistry*

5 Zibo G. Keolopile, Matthew R. Ryder, Benjamin Calzada, Maciej Haranczyk, Maciej Gutowski (Heriot Watt University)
  Angela Buonaugurio, Jacob Graham, Allyson Buytendyk, Kit Bowen (John Hopkins University)
  Intrinsic Electrophilicity of Oxalic Acid Monomer Is Enhanced in the Dimer and Trimer by Intermolecular Proton Transfer *in preparation*

6 Sanliang Ling, Zibo G. Keolopile, Alexander V. Abramov, Maciej Gutowski
  Suppressing Energy Barriers between Conformers on Molecular Potential Energy Surfaces *Submitted to the Journal of Computational Chemistry*

Chapter 1

Introduction

This dissertation is focused primarily on the development of new methods for the determination of molecular structures and applications in the area of neutral and anionic systems with *intra-* and *inter*-molecular hydrogen bonding. Molecular structures are built from atoms. An atom is a system of electrons revolving around a dense nucleus located at the center of an atom whereas a molecule is a system of atoms. The electrons of an atom are bound to the nucleus by the Coulomb force. These bound electrons possess a set of stable discrete energy levels loosely associated with orbitals. A molecule is a group of atoms bound to each other by chemical bonds (e.g. covalent or ionic) based on the Coulomb force. The arrangement of atoms defines the molecular structure. When electrons absorb or emit photons with energy equivalent to the difference in energy levels, they undergo transitions between them and remain within the bound state (discrete energy levels), but if the absorbed photon energy exceeds the electron binding energy, then the electron ends up in the continuum of energy levels where it becomes a free electron and the excess photon energy is converted to the kinetic energy of photoelectron ejected from the molecule. This is the essence of photoelectron spectroscopy. The principles of quantum mechanics can be used to model these spectra, as illustrated in this thesis.

The introductory chapter introduces the concept of molecular structure and summarises current methods for its determination. The methods I focused attention on are of two types. The first one relies on enumeration of "all possible" structures and determination of their stability (level of fitness). This approach is named further on as a "brute force" approach. The second one is based on molecular dynamics on suitably deformed potential energy surface. The theoretical predictions in this thesis run parallel with experimental verifications, in particular, the experiments on anion photoelectron spectroscopy. This experimental technique is also being discussed.

1.1 Molecular Structure Models

We have seen advances in prediction of molecular structure in terms of the concepts of valence bond theory[1] and molecular orbital theory[1] early in the twentieth century. Within these theories, the properties of the molecules were explained in terms of their electronic structure. The establishment of covalent bonding in terms of shared pairs of electrons was first introduced by G.N Lewis[2] in 1916 where he proposed that an atom shares a certain number of valence electrons with one or more other atoms in order to have an electron system which is relatively stable, and Linus Pauling[3] suggested that those shared valence electrons oscillate between two atoms participating in a bond formation. A shared electron pair forms a single bond, two and three shared electron pairs forms a double and triple bond, respectively, whereas an unshared pair of valence electrons of an atom is called a lone pair. This lone pair does not participate in bonding but it influences the shape of the molecular structure and its chemical properties.

1.1.1 Lewis structures Model

Lewis proposed the octet rule.[2, 4] In the octet rule, atoms of low atomic numbers (less than 20) share electron pairs with neighbouring atoms until they have acquired a total of eight valence electrons (an octet), as a noble gas, in exception of a hydrogen atom with only two to resemble helium. Only the valence electrons on each individual atom are used in the Lewis structure. Although the Lewis structure shows the pattern of bonds and lone pairs, it does not depict the shape/geometry of the molecule. In addition, it provides inadequate description of the molecule in terms of bond lengths. For example, in ozone, O_3, one O–O bond is different from the other, which is incorrect. This is rectified by the concept of resonance, in which the actual structure is considered to be a superposition of all possible Lewis structures. The valence-bond theory is a quantum mechanical way of expressing Lewis' ideas in terms of wave functions. Suppose there are two widely separated hydrogen atoms which can be described by a wave function $\psi_1 = \phi_a(1)\phi_b(2)$, with ϕ_a as $1s$ orbital of atom a. Then, the indistigushability of electrons allow us to equally describe the same system with a wave function $\psi_2 = \phi_a(2)\phi_b(1)$ with electron 2 sitting on atom a and electron 1 on atom b. The system will then be described well by a linear combination of ψ_1 and ψ_2, $\Psi_{\binom{s}{a}} = \phi_a(1)\phi_b(2) \pm \phi_a(2)\phi_b(1)$, where (s) (symmetric) goes with $+$ and (a) (antisymmetric) goes with $-$. The superimposed wave functions of separate fragments of the molecule describe the electron pair. In Ψ_s, the wave function describing an atom a interferes constructively with the wave function of atom b, then the amplitude of the overall wave function is enhanced in the internuclear region. With Ψ_s being symmetric with respect to permutation of electrons 1 and 2, it has to be combined

with an antisymmetric spin function, e.g. $\Psi_{spin,a} = \frac{1}{\sqrt{2}}\Big(\alpha(1)\beta(2) - \alpha(2)\beta(1)\Big)$. It applies to electrons with opposite spins. In short, spin pairing of electrons in two contributing atomic orbitals forms the valence bond function $\Psi_{VB}(1,2)$ and this wave function, $\Psi_{VB}(1,2) = N\Big(\phi_a(1)\phi_b(2) + \phi_a(2)\phi_b(1)\Big) \times \frac{1}{\sqrt{2}}\Big(\alpha(1)\beta(2) - \alpha(2)\beta(1)\Big)$, describes a single bond. The bond is formed because the pair of electrons is localised primarily between the two nuclei and hence binds them together by the Coulomb force. A similar approach can be applied to homonuclear diatomic molecules and polyatomic molecules. This is regarded as blending of structures. Resonance between Lewis structures distributes the bonding character of electrons over the molecule and lowers the energy.[1]

Bond length and strength are the key properties of bonds. The most widely used bond length is the equilibrium bond length and it can be approximated from the covalent radii of the elements participating in a bond formation. The covalent radii of elements vary through the periodic table,[5] likewise, the metallic and ionic radii.[5] The strength of the bond is measured by its dissociation enthalpy.[6]

The distribution of electrons to give the lowest energy structure can be deduced in terms of formal charge.[1] A formal charge is the charge assigned to an atom in a molecule, assuming that electrons in a chemical bond are shared equally between atoms, regardless of relative electronegativity. The Lewis structures having low formal atomic charges typically have the lowest energy and contribute the most to the total VB wave function. The oxidation character of the bond is regarded as the charge that an atom would have if it completely has ionic bonds with other atoms of different elements. So, each atom is assigned a number corresponding to its oxidation state. Likewise the ionic structure.

Valence-Shell Electron Pair Repulsion Model. The extension of Lewis's ideas by including the molecular shape came in 1940 by Nevil Sidgwick and Herbert Powell.[7] Their suggestions were then put into practise by Ronald Gillespie and Ronald Nyholm[8] and this is a well known valence-shell electron pair repulsion model (VSEPR model). In VSEPR model, bonding pairs, lone pairs or any electrons associated with any bond take up positions as far apart as possible. This minimises the energy of the molecule by minimising the repulsion between electrons and the shape is identified by referring to the positions of atoms in the resulting structure. This basic shape is modified by the strength of the electrostatic repulsion between lone pairs and the bonding pairs. Lone pairs repel other most strongly followed by lone pair:bonding pair repulsion and lastly the bonding pair:bonding pair repulsion.

1.2 Mathematical Models

The physical properties of a molecule are associated with its molecular structure. The Lewis or VSEPR models are not sufficient to predict specific molecular properties, for example, the value of dipole moment. More complete physical models, typically based on quantum mechanics, are needed to predict these properties. In the current age of information and visualisation, I expect numerical molecular characteristics, such as bond lengths and angles, orbitals, to be converted to graphical representations which can be rotated and viewed at different angles. This eases visualisation of complex structures which are hard to draw by hand. There are two major types of mathematical models that will be discussed in this thesis, the molecular mechanics model and the quantum mechanical model.

Molecular Mechanics Model. The molecular model is discussed in detail in chapter 2. In this model, atoms are represented by balls and bonds represented by springs. The potential energy is calculated using force fields.[9, 10] The force field has a form of mathematical functions involving parameters. This model is calibrated with data from experiments and high-level quantum mechanical calculations (e.g. bond lengths, angles, etc.).

Quantum Chemical Model. The quantum mechanical model is based on pure theory of quantum physics and considers molecules as positively charged atomic nuclei surrounded by electron clouds. The kinetic and potential energy operators are constructed using the rules of quantum mechanics and the Coulomb law. The approach starts from the pure quantum mechanical principles and does not require the use of experimental data for parametrisation as in molecular mechanics, thus called the Ab Initio (from the beginning). These Ab Initio methods are discussed in detail in chapter 3.

The outcome of this method could produce accurate representation of molecular properties by solving the Schrödinger equation for charged nuclei in a field of electrons. However, practical solutions of the Schrödinger equation on multi-atoms systems became achievable when the nuclei were approximated as fixed particles (the Born–Oppenheimer approximation) and the orbiting electrons were treated as independent moving particles (the Hartree-Fock model). Born and Oppenheimer[11] showed in 1927 that to a very good approximation the nuclei are much more sluggish than the electrons because they are much more massive therefore they can be treated as stationary relative to the orbiting electrons in a molecule. In principle, the nuclei are not stationary, but vibrate about the equilibrium positions and by the "fixed" nuclear positions I mean any set of instant nuclear positions. The stationary nuclei are the essence of the Born–Oppenheimer approximation. In this approximation, the Schrödinger equation is separable into an

electronic and a nuclear equation. The electronic equation can be solved for any molecular geometry and the associated electronic energy is obtained. This leads to the concept of potential energy surfaces described in section 1.3.

Solving the Schrödinger equation for polyatomic molecules proved to be computationally demanding. Thus the field benefited from the growth in computer technology. This in turn encouraged scientists to develop many-electron theories more complete than the Hartree-Fock theory, for example, the coupled cluster theory, the density functional theory, etc..

Current electronic structure methods achieved remarkable accuracy. Today, obtaining preliminary computational information about a new molecule takes minutes, while synthesising and measuring its properties might takes weeks or months. The current electronic structure packages have well developed tools for the qualitative and quantitative characterisation of wave functions and electronic charge distributions. One can also calculate molecular properties, spectroscopic parameters, and characterise transition states for elementary chemical processes.

1.3   The Potential Energy Surface.

The potential energy surface (PES) is a mathematical/graphical relationship between the molecular energy and its corresponding geometry. The molecular energy comes from solving the electronic Schrödinger equation in the Born–Oppenheimer approximation. Hence, the PES is the plot of electronic energy (supplemented with the Coulomb nuclear-nuclear repulsion energy) as a function of molecular geometric parameters. The concept of the PES first came in 1931 by Eyring and Polanyi[12, 13], where they discussed and plotted the response of the energy with respect to the change in nuclear coordinates. The examples of the common one-dimensional PES are the infinite square well, the simple harmonic potential and the Morse potential[14]. The infinite square well potential is zero inside the box and infinite outside the well (see figure 1.1). The solutions to the Schrödinger equation yield the discrete, quantized energy eigenvalues $E_n = (h^2/8mL^2)n^2$, with L as the width of the box, m as the particle mass, n as the principal quantum number ($n > 0$) and h as the Planck's constant. The spacing of energy levels increases with n. This model is solved analytically without approximations.

Allow us to illustrate the concept of the harmonic oscillator PES through the use of a diatomic molecule, AB. Let us treat atoms A and B as connected balls so that my discussion is limited to a simple picture of Hooke's law: $F = -k_f(r - r_e) = -k_f x$, where
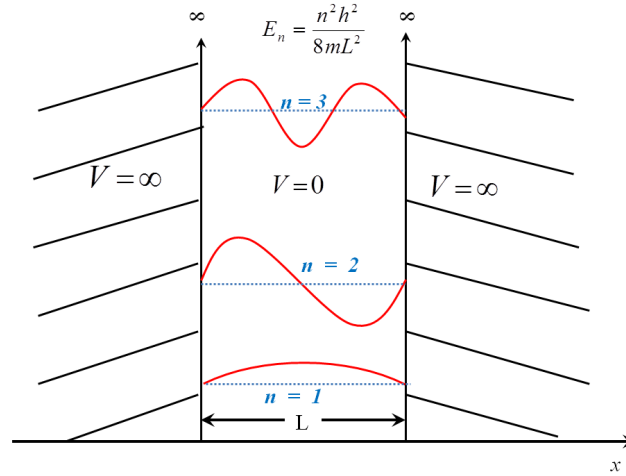
FIGURE 1.1: *The square-well potential.*

F is the force, k_f is the restoring force constant and $x = r - r_e$ is the displacement from the equilibrium position, r_e. The restoring force is a function of the displacement of the spring from its undistorted length and it is directly proportional to the displacement. By definition, compressing or stretching the spring increases energy of the system. Suppose further that no gravitational force is acting on the masses of the atoms A and B, then the energy in the system by default is only the potential energy due to the spring. From Newton's equations with Hooke's law, I can write two equations of motion, one for each atom;

$$m_1 \frac{d^2 x_1}{dt^2} - k(x_2 - x_1 - r_e) = 0 \tag{1.1}$$

$$m_2 \frac{d^2 x_2}{dt^2} + k(x_2 - x_1 - r_e) = 0 \tag{1.2}$$

I can treat the above two body problem as the one body problem. Firstly, addition of the above two equation leads us to

$$\frac{d^2}{dt^2}(m_1 x_1 + m_2 x_2) = 0. \tag{1.3}$$

The center of mass coordinate given is by $X = \frac{m_1 x_1 + m_2 x_2}{M}$ where $M = m_1 + m_2$. Then, equation 1.3 reduces to one-body equation of the form $M \frac{d^2 X}{dt^2} = 0$. By rearranging equations 1.1 and 1.2 and doing simple algebra, I can arrive at

$$\mu \frac{d^2 x}{dt^2} + kx = 0 \tag{1.4}$$

after subtracting the resulting equation 1.2 from equation 1.1. This tells us that the motion of two-body problem depends on the relative coordinate $x = x_2 - x_1 - r_e$. The reduced mass μ is defined such that $\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2}$. The general solution to equation 1.4

is given by $x(t) = c_1 \sin \omega t + c_2 \cos \omega t$ with natural frequency ω defined as $\omega = \sqrt{k_f/\mu}$. If the spring is displaced by A and released, then the spring oscillate back and forth with angular frequency $\omega = 2\pi v_f$ and the plot of displacement against time is sinusoidal with amplitude equals to A. v_f is the frequency of oscillation. In physics, the potential energy is related to a force by $F(x) = -\frac{dV}{dx}$ and together with Hooke's law give the potential energy associated with the simple harmonic oscillator,

$$V(x) = \frac{1}{2}k_f x^2. \tag{1.5}$$

Inspecting the energy change by increasing and reducing the internuclear distance yields the parabolic potential energy curve illustrated in figure 1.2.
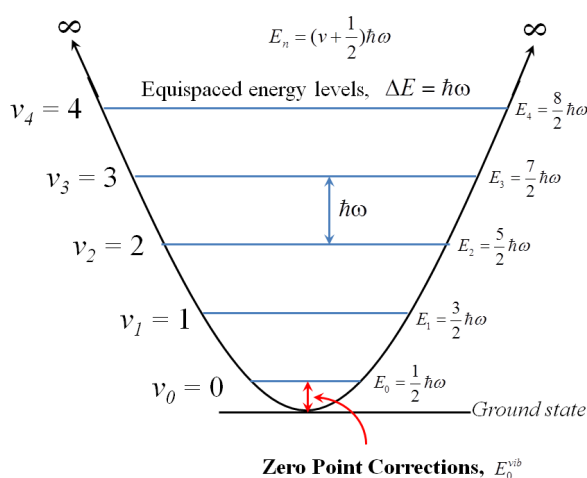
FIGURE 1.2: *The harmonic oscillator potential.*

The parabolic potential is the cornerstone of the harmonic oscillator. I will revisit the importance of the harmonic-oscillator approximation resulting from the expansion of the internuclear potential around its minimum after discussions of the quantum mechanical harmonic oscillator. The Schrödinger equation for a one dimensional harmonic oscillator is given as

$$-\frac{\hbar^2}{2\mu}\frac{d^2\Psi}{dx^2} + V(x)\Psi(x) = E\Psi(x) \tag{1.6}$$

with $V(x)$ defined in equation 1.5. Substituting for $V(x)$ and solve the resulting equation for energies, one obtain $E_v = \hbar\sqrt{k/\mu}(v+\frac{1}{2}) = (v+\frac{1}{2})\hbar\omega$, with v as the vibrational energy levels ($v = 0, 1, 2, \cdots$). The energies are quantised and the energy levels are equispaced: $E_{v+1} - E_v = \hbar\omega$. From the experimental point of view, the transition between the adjacent energy level allow us to predict the frequency of radiation, which should satisfy the Bohr frequency condition $\Delta E = hv_{obs}$, where the observed frequency $\tilde{v}_{obs} = \frac{1}{2\pi c}\left(\frac{k}{\mu}\right)^{\frac{1}{2}}$. Tilde declares units of wavenumbers. I can now express the vibrational energy in units

of wavenumbers as:

$$G(v) = \left(v + \frac{1}{2}\right)\tilde{v} \qquad v = 0, 1, 2, \cdots \tag{1.7}$$

where $G(v)$ is called the vibrational term given as $G(v) = E_v/hc$. These transitions are subject to selection rule which states that $\Delta v = \pm 1$ and there should be only one line in the infra red region corresponding to this transition. In addition, it is observed that the energy of the harmonic oscillator at its lowest quantum energy level $v = 0$, is not zero, i.e. $E_0 = \frac{1}{2}\hbar\omega$. This energy is termed the *zero-point energy* E_0^{vib}. By the Heisenberg's uncertainty principle, the stationary system should not have exactly defined momentum and position. A molecular potential energy surface is well approximated by a parabolic potential around the bottom of the well but for large displacements, especially near the dissociation limit, the parabolic approximation fails. A better description of the actual potential is the Morse potential, see figure 1.3.
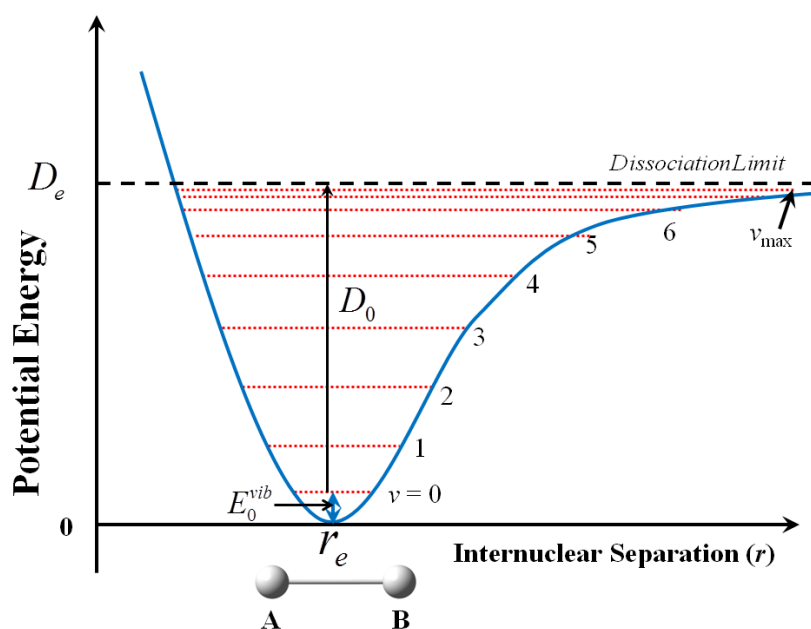
FIGURE 1.3: *Illustration of the Morse potential.*

The Morse potential[14] for a diatomic molecule is illustrated in figure 1.3 where the potential energy is plotted as a function of nuclear separation distance. The Morse potential accounts for the anharmonicity of the bond and the expression for the potential is as follows; $V_{Morse}(\Delta r) = hcD_e(1 - e^{-a\Delta r})^2$ with D_e as the depth of the potential minimum and a is related to the force constant through, $a = \left(\frac{k_f}{2hcD_e}\right)^{\frac{1}{2}}$, with k_f having the same description as in the harmonic oscillator problem. The Morse potential is expanded through Taylor series ($e^{-a\Delta r} \approx 1 - a\Delta r$) and normally the first two terms are retained where the first term becomes $V_{Morse}(\Delta r) = hcD_e a^2 \Delta r^2$. By inserting a into the expression yields the harmonic oscillator potential discussed above.

The allowed solutions to the Schrödinger equation for the vibrational energy levels are given as $E_v = (v + \frac{1}{2})\hbar\omega - (v + \frac{1}{2})^2\hbar\omega x_e$, with $\omega x_e = \frac{a^2\hbar}{2\mu}$ and x_e as the anharmonicity constant. The first term is identical as for the harmonic oscillator. The second term in the expression is negative and quadratic in v, thus increases faster than the first term and hence the expression for E_v converges at high quantum numbers. The convergence implies a finite number of vibrational energy levels, $v = 0, 1, 2, \cdots, v_{max}$, as shown in figure 1.3. The dissociation energy D_0 of the molecule is calculated as the difference between the depth of the potential D_e and the zero-point energy, $D_0 = D_e - E_0^{vib}$.
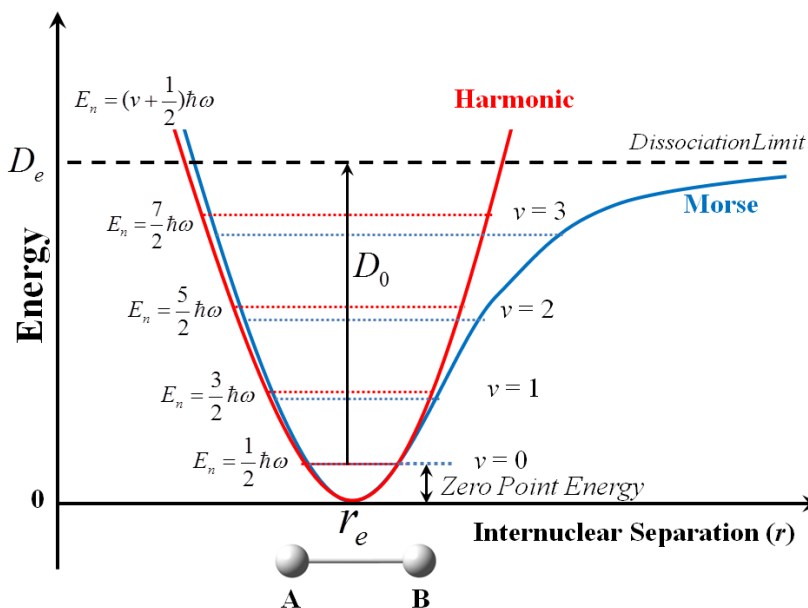
FIGURE 1.4: *Comparison between the harmonic and the Morse potentials.*

If I expand the potential in equation 1.5 by considering the Taylor series $V(x) = V(x_0) + V'x + \frac{1}{2}V''x^2 + \frac{1}{6}V'''x^3 + \cdots$ and restrict myself to small displacements, then I can neglect higher order terms beyond quadratic and see that the harmonic potential coincides well with the Morse potential as shown in figure 1.4.

Let us briefly discuss the electronic excitation from one electronic eigenstate to another as illustrated in figure 1.5. Suppose an electron is excited from the electronic ground state into an electronic excited state. Based on the Boltzmann statistics, the ground vibrational state of the lower electronic level is the most probable. There are several vibrational states in the excited electronic state as in the electronic ground state. The electron can fall into any vibrational state with the wave function that matches the initial vibrational state wave function. The match between the initial and the final vibrational states is measured by the overlap integral $S(v_f, v_i)$; $S = 1$ for absolute match and $S = 0$ for no match. It is assumed that the initial and the final geometrical coordinates of the nuclei are the same in the course of electronic transitions. The electronic transition
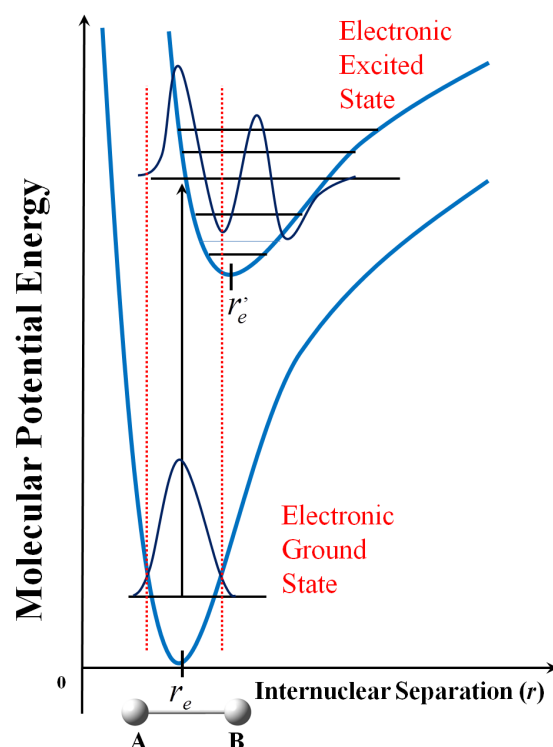
FIGURE 1.5: *Illustration of the quantum mechanical version of Franck-Condon princi-
ple.*

that occurs without change in nuclear geometry is termed the vertical transition. This
vertical transition of the electrons is known as the Frank-Condon (FC) principle[15, 16],
which reads; the electronic transition from one eigenstate to another takes place very
much faster than the nuclei can respond because the nuclei are so much heavier than
the electrons.[17] The most intense vibronic transition, $0 < S < 1$, is from the ground
vibrational state, $v_i = 0$ to an upper vibrational state lying vertically above it with
the wave function having a significant amplitude where the maximum of $v_i = 0$ is. The
intensity is proportional to the square modulus of of the overlap integral, $|S(v_f, v_i)|^2$, also
known as the Frank-Condon factor. An application of this principle will be presented in
chapter 8.

The transition of electrons from one eigenstate into another results in electron density
building up from one region to another. This in turn results in a new force being
experienced by the stationary nuclei, hence the nuclei change positions in response to
this new force, as depicted in figure 1.5 (the position of the minimum in electronic
ground state is different from the position of the minimum in the electronic excited
state). Eventually the molecule might become vibrationally excited in the course of an
electronic excitation, and the probability of the excitation is controlled by the Frank-
Condon factor.

The Stationary points. There are $3N - 5(6)$ internal geometric degrees of freedom for an $N-$atom molecule. The PES is a function of these $3N - 5(6)$ coordinates, hence it becomes nontrivial for $N > 3$. Still, the concept of PES is very important because it helps in visualising and understanding the relationship between the potential energy and the corresponding molecular geometry. Any point on the PES from which the motion in any direction leads to a higher energy is called a minimum. The deepest minimum is named "global" and other minima are "local". A local minimum corresponds to a stable structure with a finite lifetime and it is identified by zero gradients on all geometrical coordinates with all positive curvatures. Different minima correspond to different stable conformations or structural isomers of the molecule. One will never know if the "global minimum" is found unless performing a thorough search in all geometrical parameters. The transition states on the PES are identified by zero slopes with all curvatures positive but one. These transition state structures are very important because they determine the barrier heights for elementary molecular processes. Higher order transition states are not very important to the field of physical chemistry and they will not be discussed in this thesis.

Methods for locating stationary points. Stationary points are located through geometry optimisation. The stationary point of interest might be a minimum or a transition state. Geometry optimisation for finding a minimum energy structure is a computational procedure for determining a stable structure. It is often referred to as minimisation. The iterative procedure relies on forces and curvatures, which can be obtained from classical force fields or in the course of quantum chemical calculations. In minimisation, the molecular geometry is modified slightly by moving atoms in the direction of forces acting on them and the energy for the new geometry is compared to the energy from the previous iteration. This is repeated until minimum stationary point is found where forces on each atom are zero. Likewise, in the transition state optimisation, atoms are moved in the direction of forces except one coordinate whereby the movement is against the direction of force. Performing geometry optimisation alone is not sufficient enough to give us information about the nature of the stationary point, therefore it is necessary to perform second derivatives (Hessian) calculations in order to identify the curvatures of the PES. The detailed discussions of geometry optimisation is presented in section 2.2. These methods often locate stationary points that are closest in geometry to the starting geometry. So, finding the most stable minimum (global minimum) is a challenging task because the optimisation might get trapped in the nearest local minimum. In practice, computational scientists attempt to suppress this problem by starting geometry optimisations from different initial structures. This method is not bulletproof and a more reliable approach will be presented in Chapter 5 and further tested in chapter 6.

The global minimum structure is of high importance to the field of physical chemistry. Let us consider the simplest phase, gas phase, and the ideal gas model. One needs to know the most stable molecular structures to determine electronic, vibrational and rotational energy levels. This in turn will allow to determine thermodynamic properties of the gas. In more general terms, experimental chemists frequently rely on the so-called structure-property relationships. These are relations between specific structural features and molecular properties.

Computational studies have been carried out in the gas phase and in solid state to find the most stable molecular structures.[18, 19] In some computational studies, the most stable conformers were determined by manually rotating dihedral angles[20] of the molecule and this process becomes labour intensive when dealing with systems comprised of several atoms. In addition, rotating a specific dihedral angle affects other neighbouring atoms attached to the same bond that is being rotated. This greatly distorts the molecule, hence the number of geometrical optimisation steps needed to converge to a local minimum might become significant. The most serious deficiency is that the whole potential energy surface is not explored; some potentially relevant regions of the PES might be missed.

In other computational studies,[18, 19] the initial molecular structures were taken from a crystallographic database and used for determination of gas-phase molecular structures. Unfortunately these initial structures are biased in favour of intermolecular interactions and might be unsuitable for gas phase problems.[18, 19] Hence, there have been many ingenious efforts to develop methods for finding the most stable molecular structures. These efforts have been grouped, for the purpose of this thesis, into two families: deterministic and stochastic methods.

1.4 Deterministic Models for molecular structures

There have been developments in the Combinatorial Computational Cheminformatics C^3 approaches recently. These brute-force method have been applied to numerous search algorithms to find the most stable molecular tautomers[21, 22], congeners[23] and conformers[24–26]. For example, an algorithm that creates a library of tautomers, TauTGen[21] requires a molecular framework of heavy atoms. The code distributes hydrogens onto probable sites on this molecular framework. Each distribution represents an initial geometrical structure. The inability of the tool to generate conformers of each tautomer generated is a disadvantage. A tool that constructs a library of congeners of a given molecule has also been developed, ConGENER[23]. Likewise, a tool that

constructs a library of conformers for a given molecule, the SSC[25], has also been developed. The SSC tool rotates a whole molecular fragment around selected rotatable bonds. Thus it is more robust than changing selected dihedral angles. This tool generates only conformers of a tautomer specified on the input file. For a molecule that can undergo tautomerisation, this is a disadvantage because an initial input file for each tautomer has to be created and then ran separately. OMEGA[24] and CAESAR[26] generate library of conformers by dividing the molecule into smaller fragments and then reattach them in all possible ways. Unfortunately these latter tools might miss other conformers because the fragments might have and/or prefer different conformation.

In the past, our group developed a tool that performs a combinatorial/computational search for the most stable tautomers,[21] and the tool was applied to the anions of nucleic acid bases.[27] More recently, our group developed another tool for a combinatorial/computational search of the most stable conformers and it was applied to nucleosides.[25] In chapter 5, I will discuss the capabilities of my newly developed Potential Energy Surface Scanning Tool (PESST) that searches *simultaneously* for the most stable tautomers and conformers[28] and I will demonstrate its usefulness on the neutral and anionic phenylalanine. The demonstration will be extended to the search of the most stable valence anions of tyrosine, discussed in chapter 6

A silent advantage of deterministic methods is that they ignore barriers on the PES. Thus there is no problem with overcoming barriers when searching for the global minimum. This problem is a deficiency of stochastic methods discussed in section 1.5

1.5 Stochastic Methods for molecular structures

One experiences a combinatorial explosion of molecular structures as the number of rotatable bonds increases. For example, arginine has ten rotatable bonds, see fig 1.6. If the PES is scanned, then the PESST[29] or SSC[25] tool will create 1679616 initial structures, although the problem can be minimised through activation of only few rotational degrees of freedom. Performing energy minimisation for a large number of structures quickly becomes computationally infeasible.

Many ingenious theoretical methods and algorithms aiming to determine the most stable molecular structures have been developed.[30–35] The most common methods are finite temperature Monte Carlo,[32] molecular dynamics,[33] basin-hopping[34] methods and genetic algorithm[35]. The first two methods are implemented in various simulated annealing algorithms. In standard simulated annealing protocols, the temperature of the system and the PES are decided *a priori*. The simulation is run at an *acceptably*

Colour Codes
Nitrogen : Blue
Carbon   : Grey
Oxygen   : Red
Hydrogen: White

Rotatable bonds   : a b c d e f g h i j
Number of conformers: $2 \times 3 \times 6 \times 6 \times 6 \times 6 \times 6 \times 3 \times 6 \times 2 = 1679616$
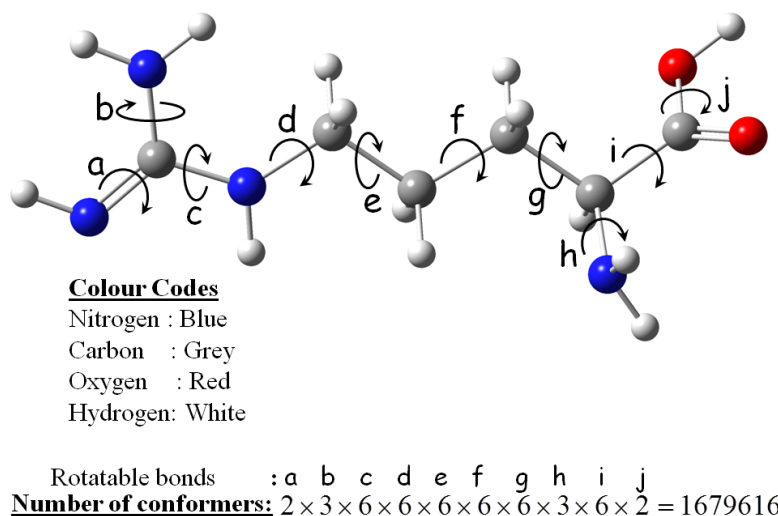
FIGURE 1.6: *Conformational space of arginine.*

high temperature (T) for a *sufficiently long time*. The reason being that the system's energy is speculated to be high enough to overcome potential energy barriers. Then the system's temperature is decreased so that it settles on the minimum. Although successful application of this methods have been reported, the minimum might be "local" rather than "global".

There is temptation to run simulations at high temperatures to overcome barriers which results in dissociation of molecular systems. At non-excessive temperature simulations, the system might not have enough energy to overcome barriers quickly. A long run would be needed to witness such a transition. As a result, the computational runtime become significant and the PES is not explored efficiently, which are disadvantages.

There have been past efforts to suppress energy barriers. These include the diffusion equation method (DEM) which transforms the original PES with multiple minima into a new PES with only one minimum.[36] Some other methods include the distance scaling method,[37] the shift method,[38] the Gaussian density annealing method,[39, 40] and the ant-lion strategy for changing the range of the potential.[41] All these above named methods require an analytical expression for the original PES and they cannot be applied to the PES determined using electronic structure methods. Another successful method, the basin hoping method,[34] requires geometrical energy minimisation for each stochastically selected molecular structure, in which the PES is transformed into a collection of inter-penetrating staircases. This is disadvantageous (minimising a large number of initial structures which converge into the same local minimum) when executed on electronic structure PES even though successful applications have been reported.[42, 43]

Here, I propose to change the PES and the temperature during the course of the simulation and run calculations at reasonably low temperatures to address the problem of dissociation of the molecules. In this case, the dynamics of the system is executed on deformed PES and the deformation protocols are initiated and terminated "on-the-fly" depending on the local information gathered during the course of the trajectory. In addition, the system's temperature might also be temporarily increased to increase its kinetic energy if it is trapped in a local minimum. In chapter 10 I will demonstrate how the PES can be deformed in molecular dynamics simulations. The dynamics will be executed on the deformed PES resulting from subtracting selected force field terms that I found to be responsible for barriers.

1.6 Experimental Determination of Molecular Structures

Spectroscopy is the study of the interaction of electromagnetic radiation with atoms and molecular structures. Important information about molecular structures can be deduced by analysing the absorption properties at various regions of the electromagnetic spectrum. The absorption in the microwave region, commonly known as microwave spectroscopy, is used to study the rotational motion of the molecules which in turn provide the moments of inertia which can be used to deduce bond lengths. Absorption in the infra red region is used to study the vibrational motion of molecules, hence the name infra red spectroscopy. The information we gather from this spectroscopy is the rigidity of the bonds which in turn can be used to map the potential energy surface of the molecule.

Anion Photoelectron Spectroscopy. The anion photoelectron spectroscopy is based on the photoelectric effect explained by Einstein in 1905.[44] It is a powerful tool to study the geometries, vibrational modes, thermochemistry, and electronic structure of anions and neutrals.[45] In anion photoelectron experiments, a mass-selected beam of atomic or molecular anions AB^- is prepared and a fixed frequency photon source hv is used to photodetach electrons from the anion, $AB^- + hv \rightarrow AB + e$. The kinetic energy of the resulting photodetached electrons KE_e is measured and the technique is governed by the energy-conserving relationship $hv = KE_e + EBE$. The electronic binding energy, EBE, is obtained by subtracting the electron kinetic energy from the photon energy. This technique has been applied to a wide range of systems (atoms, organic molecules, etc.) through the use of different sources of anions and photons of different energy.[45, 46] An example of the photoelectron spectrum of oxalic acid (OA) anion is shown in figure 1.7.
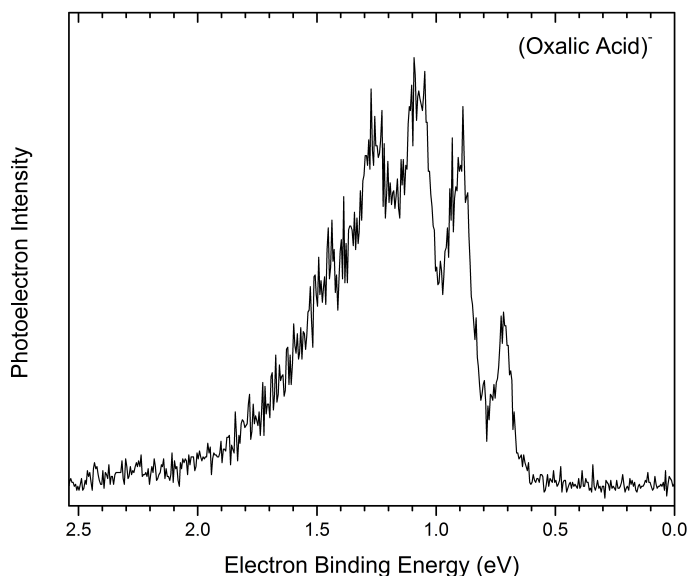
FIGURE 1.7: *Anion photoelectron spectrum of oxalic acid monomer discussed in chapter 9.*

In photoelectron experiments, the photocurrent (photoelectrons emitted per second) is plotted against the electron binding energy (see figure 1.7). The peaks in the spectrum show the electron binding energies corresponding to different quantum states of the neutral. This type of photoelectron spectra can be reproduced using theory but the geometry of the anion has to be known. The theoretical results greatly help to interpret the experimental spectrum by assigning experimental peaks to the computed transitions. Results of such a study are reported in chapter 8 for the anions of oxalic acid, monomer OA^-, dimer $(OA)_2^-$ and trimer $(OA)_3^-$. Photoelectron spectroscopy is a handy method to validate the computed anionic molecular structures. The spectra obtained using this technique has been discussed extensively in chapters 6, 7 and 9.

Microwave Spectroscopy. Molecular systems undergo rotations and vibrations in the gas phase whereas the molecular rotations are restricted in a liquid and solid phase due to intermolecular interactions between neighbouring molecular species. The rotation is relative to the set of orthogonal axis, x, y, z, passing through the system's center of mass. The rotational properties of the molecule about each rotational axis are correlated with a set of quantised energy levels which depend on angular momentum quantum number l and the moment of inertia I about that particular axis. The moment of inertia is the product of mass and square of its perpendicular distance from the axis of rotation: $I = \sum_i m_i r_i^2$.

Consider two masses (m_1, m_2) attached by rigid rod of length r free to rotate about its
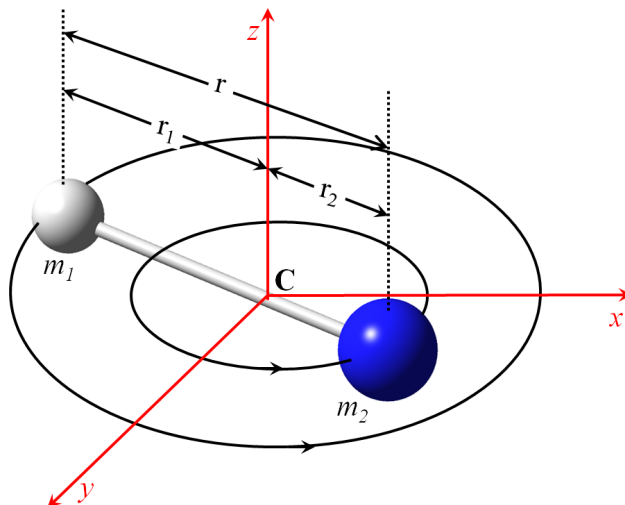
FIGURE 1.8: *Molecular system undergoes vibrations and rotations about axis passing through its center of mass C.*

center of mass, perpendicular to the molecular axis, as illustrated in figure 1.8. Masses m_1 and m_2 are located at distance r_1 and r_2 from its center of mass C, respectively. The kinetic energy possessed by the system is $T = \frac{1}{2}m_1\dot{r}_1^2 + \frac{1}{2}m_2\dot{r}_2^2 = \frac{1}{2}\omega^2\sum_i m_i r_i^2 = \frac{1}{2}I\omega^2$. Note that the velocity $\dot{r}_i = 2\pi r_i v_{rot} = r_i\omega$ with v_{rot} as the number of rotational cycles per second. By definition, the moments about point C should be equal, that is, $m_1 r_1 = m_2 r_2$. From this definition, one may arrive at the following two set of equations, $r_1 = \frac{m_2}{m_1+m_2}r$ and $r_2 = \frac{m_1}{m_1+m_2}r$. I can now rewrite the moment of inertia as $I = m_1\frac{m_2^2}{(m_1+m_2)^2}r^2 + m_2\frac{m_1^2}{(m_1+m_2)^2}r^2$ or in terms of the reduced mass μ as $I = \mu r^2$. Again, two-body problem simplifies into a one-body problem.[47–49]

The distance r depends on the three perpendicular axis $(x, y$ and $z)$, so are the moments of inertia. Then I can classify, by convention, the moments of inertia as I_x, I_y and I_z. The rotational axes are chosen such that $I_x \leq I_y \leq I_z$. The moment of inertia I depends on the molecular geometry and the masses of the atoms associated with the geometry. For a rigid rotor free from any force, its potential energy is constant and may be set to zero. As a result, the total energy comprises of only the kinetic energy $T = \frac{L^2}{2I}$, where $L = \sqrt{l_x^2 + l_y^2 + l_z^2}$ is the magnitude of angular momentum expressed in terms of angular frequency as $L = I\omega$. I now define the total rotational energy about three axes as $E = \frac{l_x^2}{2I_x} + \frac{l_y^2}{2I_y} + \frac{l_z^2}{2I_z}$.

From the above illustration, I can assume that the molecular structures do not deform under rotational stress and treat them as rigid rotors[47, 50, 51]. Although molecular structures vibrate as they rotate, the amplitude of vibration is insignificant in comparison with the atomic bond lengths, therefore to a good approximation, bond lengths can be fixed. Rigid rotors are categorised into spherical, symmetric, linear and asymmetric rotors. In the spherical rotors, the distance from the center of the sphere (rotational axis)

is the same in all directions hence equal moments of inertia. Such examples of molecular systems are NH_4^+, CH_4, SF_6. The classical expression for the rotational energy reduces to $E = \frac{l_x^2 + l_y^2 + l_z^2}{2I} = \frac{L^2}{2I}$. Here L is the magnitude of the angular momentum with the corresponding quantum mechanical expression given by $L^2 = l(l+1)\hbar^2$, where $l = 0, 1, 2, \cdots$. This implies that the rotational angular velocities are quantised accordingly: $\omega = \sqrt{l(l+1)}\frac{\hbar}{I}$. So is the total energy, with discrete values,

$$E(l) \equiv T = l(l+1)\frac{\hbar^2}{2I}. \tag{1.8}$$

In most cases, the total energy is expressed in terms of the rotational constant B and further in units of wavenumbers, denoted as $\tilde{E}(l)$ or $F(l)$ (which is called the *rotational term*), so that the discussion of vibration-rotation spectra become convenient.[47–49] Thus $\tilde{E}(l) = \frac{E(l)}{hc} = l(l+1)\frac{h}{8\pi^2 cl} = \tilde{B}l(l+1)$. The value of $l = 0$ is allowed, therefore the rotor may have zero energy and the separation of adjacent energy levels is given by $\tilde{E}(l+1) - \tilde{E}(l) = 2\tilde{B}(l+1)$ as illustrated in figure 1.9. The requirement for transition between two quantum states is $\Delta l = \pm 1$. The energy difference between two adjacent states l and $l - 1$ is simply $\Delta E = \frac{\hbar^2}{I}l$ and the frequency at which the absorption transition occur is deduced from Bohr frequency and expressed as $\tilde{v} = \frac{h}{4\pi^2 I}l = 2Bl$. This frequency lie on the microwave region, hence the name microwave spectroscopy. Given the transition frequencies from the spectrum, I can calculate the moment of inertia from the expression for the rotational constant which in turn leads to the bond distance from $I = \mu r$ if masses are known.

I can now combine equations 1.7 and 1.8 to get the vibrational and rotational energy of the diatomic molecule in the rigid rotor harmonic oscillator approximation:

$$\tilde{E}_v = G(v) + F(l) = (v + \frac{1}{2})\tilde{v} + \tilde{B}l(l+1) \qquad \begin{array}{l} v = 0, 1, 2, \cdots \\ l = 0, 1, 2, \cdots \end{array} \tag{1.9}$$

Each vibrational transition is accompanied by a rotational transition which both satisfy the selection rules, $\Delta v = +1$ and $\Delta l = \pm 1$ for absorption. Figure 1.9 illustrate the rotational energy levels associated with the ground vibrational state.

Symmetric rotors have two equal I's but different from the third. In this discussion, I shall treat the two equal I's as perpendicular to the rotational axis, $I_\perp$, and the third unique I as parallel, $I_\parallel$, to the axis. If $I_\parallel > I_\perp$, then the rotor is classed as oblate. Such example of the molecular species is the benzene C_6H_6. If $I_\parallel < I_\perp$, then the rotor is classed as prolate, with CH_3Cl as an example. The expression for the energy reduces to: $E(J) = \frac{L^2 - l_x^2}{2I_\perp} + \frac{l_x^2}{2I_\parallel} = \frac{L^2}{2I_\perp} + l_x^2\left(\frac{1}{2I_\parallel} - \frac{1}{2I_\perp}\right)$. The component of the angular momentum $\hat{L}$ operator about any axis is restricted to the eigenvalues $a = l(l+1)\hbar$ and
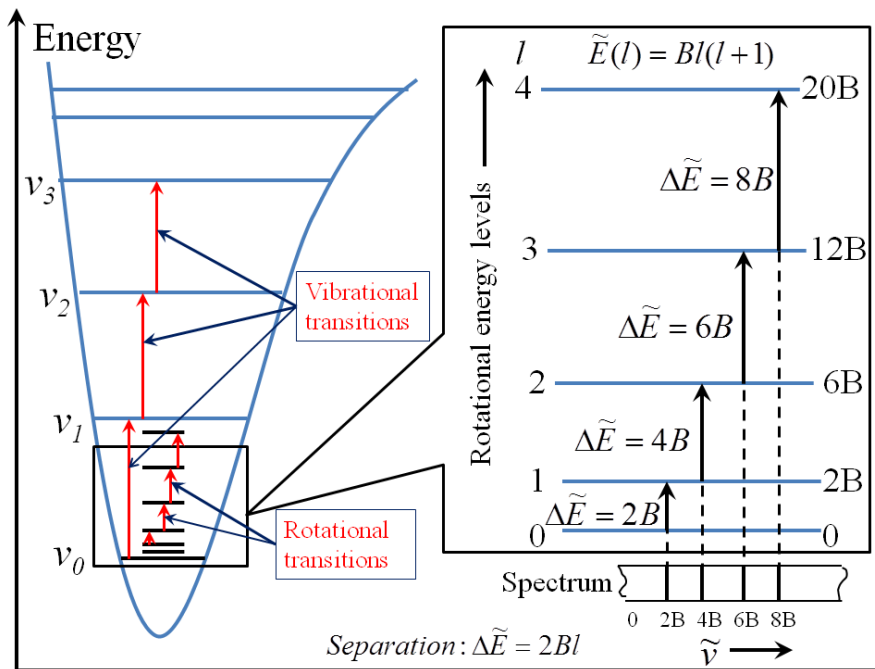
FIGURE 1.9: *An illustration of rotational energy levels of the molecule associated with ground vibrational state for diatomic molecule. Note that each vibrational energy level has rotational energy levels associated with it. Energies are in wavenumbers*

$\hat{L}_z$ operator to the eigenvalues $b = m_l\hbar$, with magnetic quantum number m_l taking values $m_l = 0, \pm 1, \ldots, \pm l$. It follows that the rotational terms are given by $\tilde{E}(l, m_l) = Bl(l + 1) + (A - B)m_l^2$ and the rotational constants A and B are defined as $A = \frac{\hbar}{4\pi c I_\parallel}$ and $B = \frac{\hbar}{4\pi c I_\perp}$, respectively.

In linear rotors, the moment of inertia about the molecular axis is equal to zero, where nuclei masses are taken as mass points lying along the rotational axis (see figure 1.8). The components of angular momentum is equivalent to zero and therefore $m_l \equiv 0$ in the expression above, hence $\bar{E}(l) = Bl(l + 1)$. This solution is similar to the solution of the spherical rotor except that the spherical rotor has $A = B$. Examples of linear rotors are CO_2 and HCl.

Asymmetric rotors have all the three moments of inertia different. Such molecular examples are H_2O and CH_3OH. These systems are too complicated to derive their analytical solutions. The rotational energy depends on three molecular constants (the rotational constants), which depend upon the moment of inertia, which in turn depend on the molecular structure.

X-ray Diffraction. X-ray diffraction is a method for investigating the atomic and molecular structure of matter and the technique exploits the properties of electromagnetic (EM) waves.[52]. Laue[53] discovered that crystalline matter diffract X-rays such a way that the molecular structure is revealed. The X-rays are electromagnetic radiation

with shorter wavelength than light in the range of 0.5-2.5 Å, whereas the wavelength of visible light is of the order of 600 Å. They are produced when charged particles are decelerated rapidly, in particular, the electrons. These electrons strike the anode target with very high velocity and, at the point of impact, x-rays are produced and radiated in different directions with a continuous range of wavelengths called Bremsstrahlung. In addition there are sharp peaks superimposed on the continuum which are character-istics of metal target used and they are referred to as K, L, M, etc classed according to the increase in their wavelengths. These sharp peaks results from collision of irradi-ated high velocity electrons with electrons in the K shell leaving the atom of the metal target in an excited state. One of the higher energy electron in the outer shell imme-diately falls into the vacancy in the K shell emitting excess energy in the process as an x-ray photon. The electromagnetic waves exhibit wave-like behaviour, that is, they undergo interference and diffraction.[52, 53] Diffraction is a phenomenon occurring when a wave encounters an obstacle and the resulting pattern of varying intensities is called the diffraction pattern.[53, 54] The pattern occurs when the waves encounter an object with dimensions comparable to its wavelength. X-rays, for example, are form of electro-magnetic radiation having a wavelengths in the range of 0.01 to 10 nanometers, which is comparable to the lattice spacing, therefore, they are used in the determination of the crystalline molecular structures by analysing the diffraction pattern produced.[53, 54]

The region where the interference of waves is constructive (peaks coincide) is seen by enhanced amplitudes and the region of destructive interference is recognised by dimin-ished amplitudes (peaks coincides with troughs). The intensity of the EM waves is proportional to the square of the amplitude of the wave.

Atoms in a crystalline structure are arranged in semi-transparent layers. Each layer reflects incident ray such that the angle of reflection is equal to angle of incident (Snell's law). The multiple reflections from these layers are described by Bragg's law;[55–57] $2d\sin\theta = n\lambda$, where d is the separation distance between the layers, θ is the incident angle (angle between the incident ray and the reflecting layers), n is the reflection pat-tern with $n = 2, 3, 4, \ldots$ called 2nd order, 3rd order, 4th order and so on, corresponding to path-length differences of $2, 3, 4, \ldots$, wavelengths. λ is the wavelength of the incident wave. The pattern produced by Bragg diffraction gives information about the sepa-rations, angles for coherent and incoherent scattering from atoms in a crystal. Bragg proposed that the intense peak will be produced if the multiple reflections from atoms interfere constructively. The Bragg (intense) peaks of the reflected radiation were ob-served at specific λ and θ and the constructive interference was when the phase shift is a multiple of 2π.[48] The separation distance may be readily calculated from the known values of θ.

X-ray crystallography is the most common and powerful technique for the determination of the atomic and molecular structures of crystals.[58, 59] The atoms in a crystalline structures cause the x-ray beam to scatter. The scattering is caused by the oscillations in the electrons of an atom generated by the incident EM wave. Atoms with larger number of electrons scatter x-rays more strongly than atoms with fewer number of electrons. The dependence of the scattering on the number of electrons is expressed in terms of the scattering factor, f, of an atom which is related to the electron density distribution $\rho(r)$ through $f = 4\pi \int_0^\infty \rho(r) \frac{\sin kr}{kr} r^2 dr$ with $k = \frac{4\pi}{\lambda} \sin \theta$.[48] The 3-D picture of the electron density is constructed from the measured angles θ and intensities of the scattered beams and then the determination of the average positions of the atoms as well as chemical bonds is deduced from the constructed electron density.

Electron Diffraction. Crystal structures of solids are also being studied by electron diffraction experiments. These experiment follow Louis de Broglie's ideas[60, 61] that electrons exhibit both particle and wave properties. His concept was based on the fact that, analogous to photons, electrons should have well defined wavelength $\lambda = h/p$, with h as the Plank's constant and p as the momentum of the electron $p = mv$ in the non-relativistic limit, $p = \sqrt{2m(KE_e)}$, where m is the electron rest mass and KE_e is its kinetic energy. The electrons are given kinetic energy by accelerating them through a measurable electric potential, $KE_e = eV$. Electron diffraction technique is used to study matter by analysing interference pattern of the electrons reflected from the sample since the wavelength $(\lambda = h/\sqrt{2meV})$ used is of order of the lattice spacing in crystals. Diffraction pattern is expected to appear for electrons with energy range of few keV.

Chapter 2

Classical Methods

2.1 Force Fields

Majority of the problems we encounter in computational chemistry are too large to be tackled by electronic structure methods even if core electrons are ignored. In force fields methods, the energy of the system is approximated as a function of only nuclear positions and the electrons are ignored. In this case, the force fields can provide answers to large systems that cannot be obtained through quantum mechanics and in some cases accurate solutions can be obtained.

The total energy of the force fields E_{FF} in use nowadays is interpreted in terms of four component picture of the intra- and two component non-bonded interactions and each component describes the energy required to distort a molecule in a specific fashion. The general functional form is;

$$E_{FF} = \underbrace{E_{str} + E_{bend} + E_{tors} + E_{cross}}_{\text{intra-}} + \underbrace{E_{el} + E_{vdw}}_{\text{non-bonded}} \qquad (2.1)$$

E_{str} is the energetic contributions due to bond stretch between two atoms, E_{bend} is the energetic penalty for bending an angle from the equilibrium value, E_{tors} is the torsional energy for rotating around the bond whereas E_{cross} describes the coupling between the aforementioned terms. The E_{el} and the E_{vdw} components describes the non-bonded atom-atom interactions. The contributions are schematically represented in figure 2.1.

The optimal energies and geometries (stable structures) can be determined by minimisation methods described in the next section.
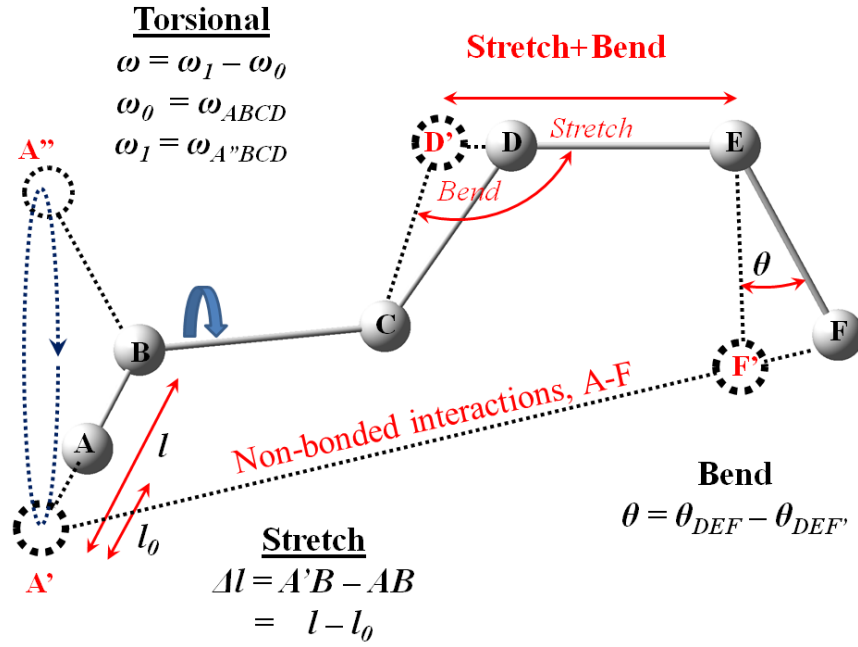
FIGURE 2.1: *Illustration of the components of the force field energy.*

2.1.1 The stretch energy

The potential energy function for a bond between two atoms A and B can be written as a Taylor expansion around an equilibrium bond length.

$$E_{str}(l^{AB} - l_0^{AB}) = E_0 + \frac{dE}{dl}(l^{AB} - l_0^{AB}) + \frac{1}{2}\frac{d^2E}{dl^2}(l^{AB} - l_0^{AB})^2 + \cdots \qquad (2.2)$$

The derivatives are evaluated at $l = l_0$ with E_0 term set to zero (zero point for the energy scale), as a result, the second term becomes zero and therefore the expansion becomes;

$$E_{str}(l^{AB} - l_0^{AB}) = k_2^{AB}(l^{AB} - l_0^{AB})^2 + \cdots$$
$$= k_2^{AB}(\Delta l^{AB})^2 + \cdots \qquad (2.3)$$

where k_2^{AB} is the force constant for bond A - B. If the equation is terminated at the second order, then the equation assumes the form of a harmonic oscillator which is sufficient for determining equilibrium geometries. It should be noted that additional terms are needed for the determination of anharmonic vibrational frequencies.

$$E_{str}(l^{AB} - l_0^{AB}) = k_2^{AB}(\Delta l^{AB})^2 + k_3^{AB}(\Delta l^{AB})^3 + k_4^{AB}(\Delta l^{AB})^4 + \cdots \qquad (2.4)$$

Following this fact, if the function is terminated at the third order, then the cubic anharmonicity constant k_3 (which is negative) will make the energy to go towards $-\infty$ if

the calculations are started from geometries with long bond lengths and the atoms may end up flying apart. This problem is rectified in Morse potential[14]:

$$E_{Morse} = D_e \left(1 - e^{-a(l-l_0)} \right)^2 \tag{2.5}$$

where D_e is the depth of the potential minimum and a is related to the force constant k through equation;

$$a = w\sqrt{\frac{\mu}{2D_e}} = \sqrt{\frac{k}{2D_e}} \tag{2.6}$$

with μ as the reduced mass and $\omega = \sqrt{\frac{k}{\mu}}$ as the frequency of the bond vibration. The Morse potential reproduces the actual behaviour quite accurately but the disadvantage with this potential is that it cannot be amended easily and it requires three parameters to be specified for each bond, this is the reason why Taylor's expansion is implemented in most force fields. Moreover, the bond lengths do not deviate significantly from equilibrium values during molecular mechanics simulations using Taylor's expansion. Such force fields that uses both the Taylor expansion and Morse potential are the Universal Force Field (UFF)[62] and DREIDING[63]. The Optimized Potentials for Liquid Simulations (OPLS-AA)[64] force field uses the standard form of equation 2.3 for the bond stretching term.

2.1.2 The bending energy

Consider three atoms A, B and C forming an angle A-B-C with A-B and B-C as chemical bonds. Small displacements from equilibrium bond angle are modelled by polynomial expansions and the penalty for straining an angle is calculated using Taylor's expansion around the equilibrium angle. Similarly, the expansion also takes the form of the harmonic oscillator if terminated at second order and this is adequate for most applications. The general form of angle strain energy E_{bend} is written as;

$$\begin{aligned} E_{ABC-bend} = \frac{k_{ABC}}{2}(\theta_{ABC} - \theta_{ABC,eq})^2[1 - k'_{ABC}(\theta_{ABC} - \theta_{ABC,eq})- \\ k''_{ABC}(\theta_{ABC} - \theta_{ABC,eq})^2 - k'''_{ABC}(\theta_{ABC} - \theta_{ABC,eq})^3] \end{aligned} \tag{2.7}$$

where θ is the angle between bonds AB and BC. The DREIDING[63] force field uses the above functional form but the most preferred is the harmonic cosine form

$$E_{ABC-bend} = \frac{1}{2}k_{ABC}[\cos\theta_{ABC} - \cos\theta_{ABC,eq}]^2 \tag{2.8}$$

The OPLS-AA[64] uses the standard form described in equation 2.7 whereas the UFF[62] uses the bending term described with cosine Fourier expansion in angle θ;

$$E_{UFF-bend} = \sum_{n=0}^{m} C_n \cos(n\theta) \tag{2.9}$$

with coefficients C_n satisfying appropriate boundary conditions.

2.1.3   The torsional energy

If I consider four-atoms sequence A-B-C-D where there is a bond between A and B, B and C and between C and D, then the torsional angle ω is defined as the angle between the plane spanned by atoms A-B-C and the plane spanned by atoms B-C-D. This can be viewed as the angle between bond AB and bond CD when looking along bond BC. The torsional energy E_{tors} function is periodic in the angle ω and in principle, the energy should return to the same value when rotated by 360°. To fulfil this criterion, the E_{tors} is written as a Fourier series.

$$E_{tors}(\omega) = k_{ABCD} \sum_{n=0}^{m} V_n cos(n\omega) \tag{2.10}$$

where n defines the periodicity and $n = 1$ describes a rotation that is periodic by 360°, then for n = 2, 3, 4 and so on, the function is respectively periodic by 180°, 120°, 90° and so on. V_n contribute to the barrier height. It should be noted that other terms in the force field such as non-bonded interactions between atom A and D also contribute to the barrier height as the bond is rotated.

Here UFF[62] and the DREIDING[63] force field uses torsional periodicities and minima modelled using Fourier representation in equation 2.10 with $V_0 = 1$, $V_n = -\cos(n\omega_0)$ with ω_0 as the equilibrium angle and $k_{ABCD} = 1/2V_\omega$ yielding;

$$E_{UFF/DREIDING}(\omega) = \frac{1}{2}V_\omega[1 - \cos(n\omega_0)\cos(n\omega)] \tag{2.11}$$

whereas the OPLS-AA uses the functional form;

$$E_{OPLS-AA-tors} = \frac{1}{2}V_1(1 + \cos\omega) + \frac{1}{2}V_2(1 + \cos 2\omega) +$$
$$\frac{1}{2}V_3(1 + \cos 3\omega) + \frac{1}{2}V_4(1 + \cos 4\omega) \tag{2.12}$$

2.1.4 Cross terms: Class 1, 2 and 3 Force Fields

The aforementioned terms couple with one another. They are not treated as individual coordinates. This can be understood by expanding the total potential energy function in a multidimensional Taylor series where the products of the first order expansion are components of the cross terms. The example of such cross terms are listed below;

$$E_{str/bend} = k_{ABC}(\theta_{ABC} - \theta_{ABC,eq})[(l_{AB} - l_{AB,eq}) - (l_{BC} - l_{BC,eq})] \tag{2.13a}$$

$$E_{str/str} = k_{ABC}(l_{AB} - l_{AB,eq})(l_{BC} - l_{BC,eq}) \tag{2.13b}$$

$$E_{bend/bend} = k_{ABCD}(\theta_{ABC} - \theta_{ABC,eq})(\theta_{BCD} - \theta_{BCD,eq}) \tag{2.13c}$$

$$E_{str/tor} = k_{ABCD}(l_{AB} - l_{AB,eq})cos(n\omega_{ABCD}) \tag{2.13d}$$

$$E_{bend/tor} = k_{ABCD}(\theta_{ABC} - \theta_{ABC,eq})cos(n\omega_{ABCD}) \tag{2.13e}$$

$$E_{bend/tors/bend} = k_{ABCD}(\theta_{ABC} - \theta_{ABC,eq})(\theta_{BCD} - \theta_{BCD,eq})cos(n\omega_{ABCD}) \tag{2.13f}$$

In principle, in most force field, the constant k depends on all atoms in the sequence but the central atom. For example, the stretch/bend constant k_{ABC} is usually taken as K_B in common force fields.

The classification of force fields is done based on these cross terms. Force fields that do not include these terms are regarded as class I and those with explicit cross terms are viewed as class II force fields. The inclusion of these terms improves the ability of the force fields to predict the structural properties accurately. The cross terms are very important for optimal representation of the PES and for reproducing the vibrational frequencies accurately. In general, the $E_{str/bend}$ term is the most important among the all cross terms. The class III force fields take into account the chemical effects and other features such as electronegativity and hyperconjugation.

2.1.5 The van der Waals energy

The attraction or repulsion between two non-bonded atoms is modelled by van der Waals function. The van de Waals function has two terms, the dispersive interactions and the exchange-repulsive interactions. The attractive contributions are due to dispersive forces which results from instantaneous dipoles caused by fluctuation in the electron clouds. The repulsive contributions are quantum mechanical in origin and can be understood through Pauli principle which prohibits any two electrons in a systems from having the same set of quantum numbers. In this case, the pairs of electrons cannot occupy the same region of space. The attractive dispersion interactions varies as $\frac{1}{r^6}$ and dominates at large distances. The repulsive term dominates at short distances and is modelled as either $\frac{1}{r^{12}}$ or Ae^{-Br}.
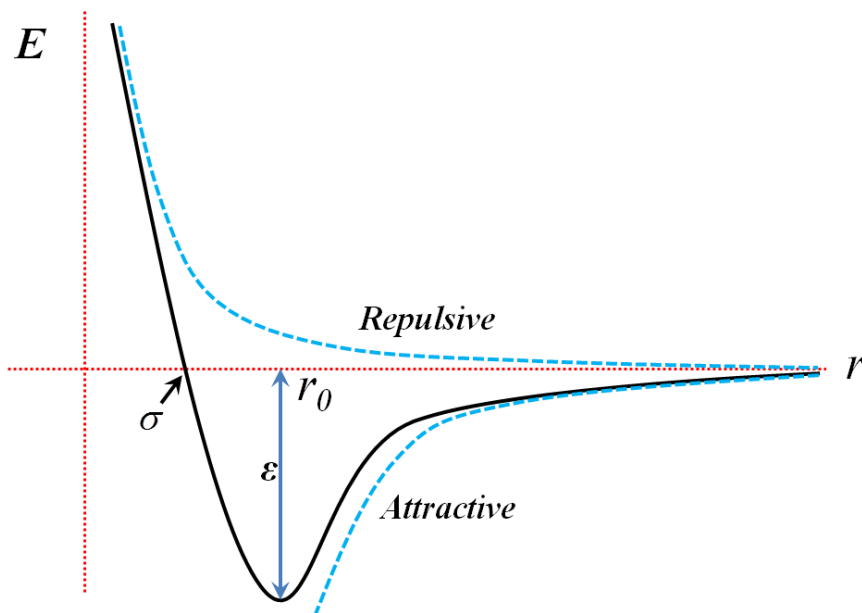
FIGURE 2.2: *Illustration of the Lennard-Jones potential.*

Quantum mechanics is used to model the dispersive and exchange-repulsive interactions between pairs of atoms and molecules and the calculations are demanding as large basis sets and electron correlation are required. For fast calculations, Lennard-Jones 12-6 function has been used to model these interactions. It has an attractive part that varies as r^{-6} and a repulsive part that varies as r^{-12} and the total potential is defined as follows:

$$E_{LJ}(r) = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right] \tag{2.14}$$

where σ is the collision diameter (the separation for which the energy is zero) and ε is the well depth. These two parameters can be adjusted. The $E_{LJ}(r)$ is positive at small separations with slightly negative minimum at equilibrium bond distance and approaches zero at large separation distances as shown in figure 2.2.

Amber[65], OPLS[66] and OPLS-AA[64] uses the above standard Lennard-Jones 12-6 potential. The Merck Molecular Force Field (MMFFs)[67, 68] uses a buffered 14-7 Lennard-Jones potential with exponents and empirical constants derived from experimental data;

$$E_{buf14-7}(r) = \varepsilon \left(\frac{1.07 r_0}{r + 0.07 r_0} \right)^{7} \left(\frac{1.12 r_0^7}{r^7 + 0.12 r_0^7} - 2 \right) \tag{2.15}$$

Some molecular force fields such as the MM2[69, 70] and MM3[71] uses the 'Buckingham' or 'Hill' (also known as the exponential-6 (X6)) type potential;

$$E_{X6}(r) = A e^{-Br} - \frac{C}{r^6} \tag{2.16}$$

or sometimes written as;

$$E_{X6} = \varepsilon \left[\frac{6}{\alpha - 6} e^{(1 - r/r_0)} - \frac{\alpha}{\alpha - 6} \left(\frac{r_0}{r} \right)^6 \right] \tag{2.17}$$

with A,B and C as constants and R_0 as the minimum energy distance. The DREIDING force field[63] uses both the Lennard-Jones 12-6 type potential and the exponential-6 form.

Hydrogen bonding in force fields. Hydrogen bonds are formed between a protic hydrogen atom and the lone pair of the electronegative atom such as nitrogen and oxygen, that is, from the positively charged hydrogen and the negatively charged heteroatom. The hydrogen bonding parameters are substituted into the van de Waals potential for only pairs of atoms which are in vicinity of each other and able to participate in the hydrogen bonding interactions. They are modelled by short bond length R_0 and large ε but in some force fields the modified Lennard-Jones potential is used;

$$E_{H-bond}(R) = \varepsilon \left[5 \left(\frac{r_0}{r} \right)^{12} - 6 \left(\frac{r_0}{r} \right)^{10} \right] \tag{2.18}$$

In some force fields, the hydrogen bonding functional form is multiplied by a factor $(1 - \cos \omega^{XHY})$ or $(\cos \omega^{XHY})$ to give it some directionality, where ω^{XHY} is the angle between the hydrogen atom H bonded to atom X and the hydrogen bond formed between the H and the electronegative atom Y.

The MM2[69, 70] and MM3[71] uses the angle-independent 12-10 Lennard potential described above. The DREADING force field uses CHARMM[72] like hydrogen bonding potential described by the following equation;

$$E_{HB} = D_{HB} \left[5 \left(\frac{r_{HB}}{r_{XY}} \right)^{12} - 6 \left(\frac{r_{HB}}{r_{XY}} \right)^{10} \right] \cos^4(\omega_{XHY}) \tag{2.19}$$

whereas the Amber[65], OPLS[66], OPLS-AA[64] and MMFFs[67, 68] have no explicit hydrogen bonding term.

The total non-bonded energy is the sum of this term, the van de Waals, E_{vdw}, and the electrostatic, E_{el}, term discussed in the next section.

2.1.6 The electrostatic energy

Quantum mechanical description of electron density leads to accumulation of excess charge in some regions and depletion in others. This is also due to electronegative elements attracting more electrons than less electronegative elements thus uneven charge distributions in the molecule. The excess charges interact with each other through Coulomb interaction. The most convenient way of describing the electrostatic interactions is to assign each atom a partial charge and to estimate the electrostatic interactions between two molecules or parts of the same molecule using the Coulomb potential. This is calculated as the sum of the interactions between pairs of point charges;

$$E_{el} = \sum_{i=1}^{N} \frac{1}{4\pi\varepsilon_0\varepsilon} \frac{q_i q_j}{r_{ij}} \qquad (2.20)$$

where q_i and q_j are point charges, ϵ is the dielectric constant used to model the effect of surrounding molecules whereas ε_0 is the permittivity of free space and N is the total number of atoms. Good partial charges are obtained through Molecular Electrostatic Potential (MEP) fitting from electronic structure methods through minimization of the error function defined below.

$$ErrF(q) = \sum_{r}^{N_{points}} \left(\phi_{esp}(r) - \sum_{a}^{N_{atoms}} \frac{q_a(R_a)}{|R_a - r|} \right)^2 \qquad (2.21)$$

The electrostatic potential ϕ_{esp} at point r is then calculated from the following equation.

$$\phi_{esp}(r) = \sum_{a}^{N_{nuc}} \frac{Z_a}{|R_a - r|} - \int \frac{\Psi^2(r')}{|r' - r|} dr' \qquad (2.22)$$

where Z_a are the nuclear charges and Ψ as the electronic wave function. The partial charges obtained through minimisation of the error functions are too sensitive to small details in the fitting data and this is primarily due to charge fittings using only surface atoms. This is often rectified by adding a hyperbolic penalty term for having non-zero partial charges. This Restrained ElectroStatic Potential (RESP)[73] fitting scheme has been used in AMBER force field[65].

The OPLS-AA uses the standard form of Coulomb potential, described in equation 2.20. Both the UFF and the DREIDING uses the slightly modified Coulomb potential constant;

$$E_{UFF} = 332.0637 \frac{q_i q_j}{\varepsilon r_{ij}} \qquad (2.23)$$

for the electrostatic interactions with charges obtained using Charge Equilibration (QEq) scheme[74].

2.2   Geometry Optimization, (Mathematical methods of)

The potential energy surface of many-body system is multi-dimensional. Studies on "how the energy varies with the coordinate system" are complicated because for any system with N atoms the energy is thus a function of $3N$ Cartesian coordinates system or $3N-5$ internal coordinates for linear systems or $3N-6$ for non-linear systems. Visualisation of this kind of potential energy surface becomes impossible except for small systems where energy is a function of a few coordinates. It is necessary to understand the potential energy surface because the stationary points are important for thermodynamics and kinetics. These points can be identified through the use of optimisation algorithms.

Consider a function f which depends on several more independent variable $r_1, r_2, \ldots, r_i$. The information about the forces and curvature of the energy function can be obtained through derivatives. The first derivatives of the function with respect to each variable should be zero at the minimum and the second derivatives should be positive i.e.

$$\frac{\partial f}{\partial r_i} = 0 \qquad\qquad \frac{\partial^2 f}{\partial r_i^2} > 0 \qquad\qquad (2.24)$$

At the minimum, all frequencies are real but at the transition state, there is only one imaginary frequency with all others real and the second derivatives are positive except for one direction.

$$\underbrace{\frac{\partial^2 f}{\partial r_n^2} < 0}_{\text{for only one r}} \qquad\qquad \underbrace{\frac{\partial^2 f}{\partial r_n^2} > 0}_{\text{for other r's}} \qquad\qquad (2.25)$$

Most minimisation algorithms locate the nearest minimum from the starting point by going downhill on the energy surface. The direction of nearest minimum is obtained from the first derivative of the energy (gradient). The force acting on each atom is equal to the negative of the gradient and each atom can be moved in response to these forces in an attempt to lower the energy of the system. The potential energy function for a multidimensional case is written as a Tylor's expansion about point r_n

$$f(r) = f(r_n) + (r - r_n)f'(r_n) + (r - r_n)^{\mathbf{T}} \cdot f''(r_n) \cdot \frac{1}{2}(r - r_n) + \cdots \qquad (2.26)$$

For any potential energy function with $3N$ Cartesian coordinates, the variable r is a simplification of a $3N$ dimensional vector r. Each component is a partial derivative of f with respect to appropriate coordinate, $\frac{\partial f}{\partial r_n}$ and the gradient at point n will be denoted as g_n. Each component (i,j) of matrix $f''_{i,j}(r_n)$ is the partial second order derivative of the potential energy function with respect to 2 coordinates r_i and r_j, $\frac{\partial^2 f}{\partial r_i \partial r_j}$.

The highest-order derivative used in the optimisation method determines its class. Those which only use the first-order derivatives are classed as first-order methods and those which use both first and second order derivatives are classed as second-order derivative methods. The most frequently used first-order minimisation method is the steepest descents method, which is popular in minimising structures far from the minimum, and the second-order minimisation method is the Newton-Raphson method, which is widely used for structures close to the minimum.

2.2.1 The Steepest Descent Method

In the steepest descent method, the direction of movement of particles is parallel to the direction of net force, downhill, that is, in the direction opposite to the gradient $\mathbf{g}$. We start from the initial position r_n followed by deciding the step size λ along the gradient and then descend towards minimum through the use of the following equation:

$$r_{n+1} = r_n - \lambda_n \mathbf{g}_n \tag{2.27}$$

For search method, we then proceed along the direction of forces until minimum is found and then recompute the gradient $\mathbf{g}_{n+1}$ at the located minimum and repeat the process. The step size λ is governed by the directional derivative;

$$\frac{d}{d\lambda_n} f(r_{n+1}) = \nabla f(r_{n+1})^{\mathbf{T}} \cdot \frac{d}{d\lambda_n} r_{n+1} \tag{2.28}$$

$$= -\nabla f(r_{n+1})^{\mathbf{T}} \cdot \mathbf{g}_n \tag{2.29}$$

in which both $-\nabla f(r_{n+1})^{\mathbf{T}}$ and $\mathbf{g}_n$ are orthogonal to each other when the equation is set to zero. The succeeding search is orthogonal to the preceding search resulting in a zig-zag pattern. Since the points are taken in a linear fashion, the movement is known as line search and the method itself is stable, fast, simple and easy to apply. It is fast in terms of initial convergence but become slower towards minimum because the step size decreases as it approaches minimum.

The procedure is outlined below;

1. Calculate the gradient of the function: $\boldsymbol{g}_n = \nabla f_n(r)$

2. Determine the direction of movement; $\boldsymbol{d}_n = -\boldsymbol{g}_n$

3. Calculate the step size λ; $min_{\lambda_n > 0} f(r_n + \lambda_n d_n)$

4. Calculate the new point from equation 2.27; $r_{n+1} = r_n + \lambda \boldsymbol{d}_n$

5. Repeat the procedure at this new point by calculating the gradient $\boldsymbol{g}_{n+1} = \nabla f(r_{n+1})$ and the direction $\boldsymbol{d}_{n+1} = -\boldsymbol{g}_{n+1}$ at point r_{n+1} and determine other points until convergence.

2.2.2 The Newton-Raphson Method

This is the most robust second-order minimisation method. The potential energy function is written as a Taylor's expansion around point r_n as described by equation 2.26. For $N-$atom molecule, the coordinates r of the atoms have $3N-$dimensions with elements $\{r_{n,k}\}$;

$$
r_n = \begin{bmatrix} r_{n1} \\ r_{n2} \\ r_{n3} \\ \vdots \\ r_{3N} \end{bmatrix}
\tag{2.30}
$$

and with $3N-$dimensional gradients(first derivatives) g which can be calculated analytically. The first derivative gives the direction of the vector. Let us pay attention on the nth coordinate so that I follow its direction. The gradient of the nth coordinate will be given by

$$
\mathbf{g}_n = \nabla_n f(r_n) = \begin{bmatrix} \frac{\partial f(r)}{\partial r_{n1}} \\ \frac{\partial f(r)}{\partial r_{n2}} \\ \vdots \\ \frac{\partial f(r)}{\partial r_{3N}} \end{bmatrix}
\tag{2.31}
$$

with the elements $\{g_{n,k}\}$ collected in a column vector $\mathbf{g}_n$ and the Hessian (second derivative), H, is $3N \times 3N$ matrix whose elements are defined by;

$$
H_{ij}^{r_n} = \frac{\partial^2 f(r)}{\partial r_i \partial r_j}\bigg|_{r=r_n} = \begin{bmatrix} \frac{\partial^2 f(r)}{\partial r_1 \partial r_1} & \frac{\partial^2 f(r)}{\partial r_1 \partial r_2} & \cdots & \frac{\partial^2 f(r)}{\partial r_1 \partial r_{3N}} \\ \frac{\partial^2 f(r)}{\partial r_2 \partial r_1} & \frac{\partial^2 f(r)}{\partial r_2 \partial r_2} & \cdots & \frac{\partial^2 f(r)}{\partial r_2 \partial r_{3N}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(r)}{\partial r_{3N} \partial r_1} & \frac{\partial^2 f(r)}{\partial r_{3N} \partial r_2} & \cdots & \frac{\partial^2 f(r)}{\partial r_{3N} \partial r_{3N}} \end{bmatrix}_{r=r_n}
\tag{2.32}
$$

Here the second derivative gives forces and curvature of the directional vector therefore the minimum can be guessed along the line searched. Suppose I use the unitary matrix U that diagonalises H_n such that the diagonalised Hessians are given by λ;

$$
(U^{\mathbf{T}} H_n U)_{k,l} = \delta_{k,l} \lambda_k \qquad here, \qquad UU^{\mathbf{T}} = 1
\tag{2.33}
$$

and λ is a diagonal matrix with eigenvalues of the Hessian matrix;

$$\lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_{3N} \end{bmatrix} \tag{2.34}$$

There will be $3N - 5|6$ eigenvalues λ which are non-zero. By rewriting equation 2.26 in terms of $\mathbf{g}_n$ and H_n and then insert the unitary matrix into the resulting equation, one obtain the following equation truncated at the second order;

$$f(r) = f(r_n) + (r - r_n)\mathbf{g}_n^{\mathbf{T}} + (r - r_n)^{\mathbf{T}} \cdot H_n \cdot \frac{1}{2}(r - r_n) \tag{2.35}$$

$$= f(r_n) + UU^{\mathbf{T}}(r - r_n)\mathbf{g}_n^{\mathbf{T}} + \frac{1}{2}(r - r_n)^{\mathbf{T}}UU^{\mathbf{T}}H_nUU^{\mathbf{T}}(r - r_n) \tag{2.36}$$

I can write equation 2.36 in terms of steps along the $3N - 5|6$ directions corresponding to non-zero Hessian eigenvalues in a more compact form by assigning $U^{\mathbf{T}}(r - r_n) = Q$ and $U^{\mathbf{T}}\mathbf{g}_n = G$ and by using the fact that $(AB)^{\mathbf{T}} = B^{\mathbf{T}}A^{\mathbf{T}}$;

$$f = \underbrace{f(r_n)}_{\text{constant},C} + G^{\mathbf{T}}Q + \frac{1}{2}Q^{\mathbf{T}}\lambda Q \tag{2.37}$$

$$= C + \sum_m G_m Q_m + \frac{1}{2}\sum_m Q_m \lambda Q_m \tag{2.38}$$

where I have introduced the component of the step, $Q_m = \sum_k U_{mk}^{\mathbf{T}}(r - r_n)_k$ and the component of the gradient $G_m = \sum_k U_{mk}^{\mathbf{T}}\mathbf{g}_{n,k}$ along the mth eigenvector of the Hessian matrix H. I focused my attention on the kth component of the vector. At the minimum, I seek the value $r = r_{n+1}$ so that the first derivative is zero, $f'(r_{n+1}) = 0 = \mathbf{g}$, therefore I minimise the energy function in equation 2.38 with respect to Q_m and write;

$$\frac{\partial f}{\partial Q_m} = G_m + \lambda_m Q_m = 0 \tag{2.39}$$

This implies that the information about geometry optimisation step and its direction is contained in Q_m;

$$Q_m = \frac{-G_m}{\lambda_m} \tag{2.40}$$

which is Newton-Raphson method. In geometry optimisation, Newton-Raphson method is able to find the minimum in one step for a quadratic function and the step may be written as

$$\Delta \mathbf{x}' = (\Delta x_1', \Delta x_2', \Delta x_3', \cdots, \delta x_N') \tag{2.41}$$

$$\Delta x' = -\frac{g_n}{\varepsilon_n} \tag{2.42}$$

with the projected gradient $\mathbf{g}_n$ along the Hessian eigenvector with eigenvalue ε_n. The step is controlled such that the direction is correct and total length of the step does not exceed the region in which the 2nd order Taylor expansion is valid. This is done through the use of shift parameter γ;

$$\Delta x'_n = -\frac{f_n}{\varepsilon_n - \gamma} \tag{2.43}$$

The correct step direction is given by all values of γ below the lowest Hessian eigenvalue and those γ values approaching $-\infty$ gives almost zero step size. The inverted Hessian matrix H^{-1} is computationally demanding for large systems.

However, most of the energy functions are not quadratic therefore the minimum cannot be determined in one step. In this case, I apply the procedure iteratively following the below outline;

1. Calculate the first derivative of the function, the gradient: $\mathbf{g}_n = \nabla f(r_n)$.

2. Calculate the second derivative, the Hessian: $\mathbf{H}_n = \nabla^2 f(r_n)$.

3. Determine the step length and the direction of search from equation 2.40

4. Calculate the coordinates of the new step:

5. Repeat the procedure at the new coordinated found in step 4 until convergence.

The Newton-Raphson method is slow at the beginning for very distorted geometries but converges very quickly when geometries are close to the minimum unlike the steepest descent method with rapid convergence at the beginning but very slow towards the convergence. The solution to this problem is to combine these two algorithms, starting with steepest descent and then Newton-Raphson method towards convergence as done in Gaussian electronic structure suite of programmes[75] under quadratic convergence.

Normal modes of vibration. An expansion of potential energy V around the minimum energy structure and truncated after quadratic terms leads to the following expression for the total classical energy of the molecule;

$$E_{tot} = E_{kin} + E_{pot} = \frac{1}{2}\sum_i m_i \dot{x}_i^2 + \frac{1}{2}\sum_{ij} K_{ij} x_i x_j \tag{2.44}$$

where

$$K_{ij} = \frac{\partial^2 V}{\partial x_i \partial x_j} \tag{2.45}$$

Vibrational normal modes are defined in terms of mass-weighted coordinates q_i;

$$q_i = m_i^{1/2} x_i \tag{2.46}$$

where m_i is the mass of the atom displaced by x_i. The kinetic energy becomes ;

$$E_{kin} = \frac{1}{2} \sum_i \dot{q}_i^2 \qquad (2.47)$$

and the potential energy takes the form

$$E_{pot} = \frac{1}{2} \sum_{ij} K_{ij} q_i q_j \qquad (2.48)$$

with

$$K_{ij} = \frac{k_{ij}}{(m_i m_j)^{1/2}} = \frac{\partial^2 V}{\partial q_i q_j} \qquad (2.49)$$

The total classical energy takes this form

$$E_{tot} = \frac{1}{2} \sum_i \dot{q}_i^2 + \frac{1}{2} \sum_{ij} K_{ij} q_i q_j \qquad (2.50)$$

In order to have the potential energy in the diagonal form one introduces eigenvectors of the K_{ij} as new coordinates labelled Q_i and corresponding eigenvalues labelled κ_i. The procedure is analogous to the one presented in equations 2.36/2.38 , but now it applies to the mass-weighted Hessian rather than to Hessian. In these coordinate the total classical energy takes the form

$$E_{tot} = \frac{1}{2} \sum_i \dot{Q}_i^2 + \frac{1}{2} \sum_i \kappa_i Q_i^2 \qquad (2.51)$$

This leads to the following quantum mechanical Hamiltonian

$$\hat{H} = \sum_i \hat{H}_i \qquad \hat{H}_i = -\frac{1}{2} \hbar^2 \frac{\partial^2}{\partial Q_i^2} + \frac{1}{2} \kappa_i Q_i^2 \qquad (2.52)$$

Since the Hamiltonian is additive in normal modes Q_i the vibrational wave function of the molecule is a product of wave functions for each mode

$$\Psi_{vib} = \prod_i \Psi_{v_i}(Q_i) \qquad (2.53)$$

and the total vibrational energy is the sum of the energies associated with each mode

$$E_{vib} = \sum_i E_{v_i} \qquad (2.54)$$

$$\infty$$

$$E_n = (v + \frac{1}{2})\hbar\omega$$

Equispaced energy levels, $\Delta E = \hbar\omega$

$v_4 = 4$ $\qquad\qquad E_4 = \frac{8}{2}\hbar\omega$

$v_3 = 3$ $\qquad\qquad E_3 = \frac{7}{2}\hbar\omega$

$\hbar\omega$

$v_2 = 2$ $\qquad\qquad E_2 = \frac{5}{2}\hbar\omega$

$v_1 = 1$ $\qquad\qquad E_1 = \frac{3}{2}\hbar\omega$

$v_0 = 0$ $\qquad\qquad E_0 = \frac{1}{2}\hbar\omega$

Ground state

Zero Point Corrections, E_0^{vib}
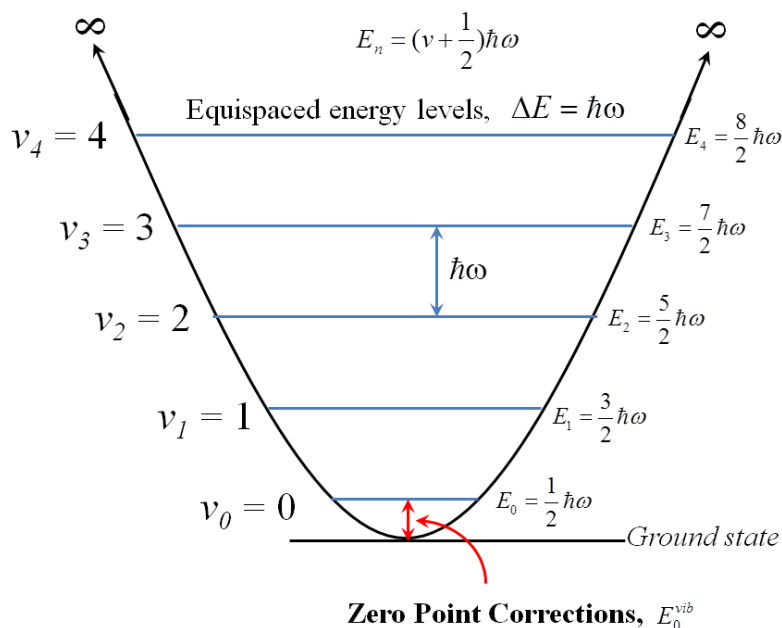
FIGURE 2.3: *Illustration of the vibrational energy levels of the molecule.*

The solution Ψ_{v_i}, E_{v_i} are known from the theory of one-dimensional harmonic oscillator

$$E_{v_i} = (v_i + \frac{1}{2})\hbar\omega_i \qquad \omega_i = k_i^{\frac{1}{2}} \qquad v_i = 0, 1, 2, \ldots \tag{2.55}$$

$$\Psi_{v_i} = N_{v_i} H_{v_i}(y_i) e^{-y_i^2/2} \qquad y_i = \left(\frac{\omega}{\hbar}\right)^{\frac{1}{2}} Q_i \tag{2.56}$$

where v_i is the vibrational quantum number, ω_i is the vibrational frequency, $H_{v_i}(y_i)$ are the Hermite polynomials of the variable y_i. The total vibrational energy is given by

$$E_{vib} = \sum_i (v_i + \frac{1}{2})\hbar\omega_i \tag{2.57}$$

The vibrational ground state of a polyatomic molecule is of special interest as it is the only quantum vibrational state available at zero Kelvin. The associated energy E_0 is the termed "zero-point" vibrational energy

$$E_0 = \frac{1}{2}\sum_i \kappa\omega_i \tag{2.58}$$

This is illustrated in figure 2.3.

2.3 Molecular Dynamics Algorithms/Schemes

The most common classical molecular simulation search methods used are Molecular Dynamics[76, 77] and Monte-Carlo[78] methods. Molecular Dynamics (MD) simulation method addresses the time evolution of system, therefore, one can easily examine each MD step individually. Its is possible to predict the outcome of the MD simulation at any time in the past and in the future. The total energy of the system is made up of kinetic energy and also contributions from potential energy. For MD calculations, sampling is performed mostly under microcanonical ensemble. Monte-Carlo (MC) samples under canonical ensemble. The MD simulation search method use differential equations embodied in Newton's second law to calculate forces and accelerations of particles for the next geometry in the trajectory. Most MD codes compute forces from derivatives of empirical force fields discussed in the preceding section. In real situation, the forces should change instantly whenever the position of the particle changes and even when the nearest neighbour changes its position. Even though the parameters used in force fields are computed from electronic structure calculations, the end results are not exact, therefore, for better results, forces and energy can be obtained from Born-Oppenheimer electronic energy surface which replaces the potential energy determined by a force field.

Consider a set of N-particles assuming a set of coordinates $q_0 = (r_1, r_2, \ldots, r_{3N})$ with velocity expressed as $(\frac{dr}{dt})_0$, the vector of masses m and the total energy E which comprised of the kinetic energy Ξ and the potential energy V. Suppose the particle displaces by Δx in time duration Δt, such that the new generated coordinates $x(t + \Delta t)$ and velocities $\frac{dq}{dt}(t + \Delta t)$ at time $(t + \Delta t)$ is expressed as,

$$x(t + \Delta t) = x_0 + (\frac{dx}{dt})_0 \Delta t \qquad (2.59)$$

and

$$\frac{dx}{dt}(t + \Delta t) = \left(\frac{dx}{dt}\right)_0 - \Delta t \left[\frac{(\frac{\delta V}{\delta x})_0}{m}\right] \qquad (2.60)$$

where $-(\frac{\delta V}{\delta x})_0$ is the initial geometry force along the x-coordinate. Then one can express the momenta $p = (p_1, p_2, \ldots p_{3N})$ of the particles as,

$$p_x = m_x \frac{dx}{dt} \qquad (2.61)$$

Having a new set of coordinates and velocities, one can generate another set using Newton's equations and the procedure can be repeated for a finite number of steps. The behaviour of particles can then be investigated from these sequence of coordinates. In MD (or Newtonian dynamics) calculations, the potential energy of the system comes from a force field and contributes to the Hamiltonian of the system. The Hamiltonian

function $H(p, r)$ consist of terms such as the kinetic energies, $\Xi(p)$, of the particles and the potential energy, $V(r)$. The potential energy is the energy possessed by a mass due to its position within a force field. The Hamiltonian as function of momenta and position can then be written as;

$$H(p, r) = \sum_{i=1}^{3N} \frac{p_i^2}{2m_i} + V(r) \tag{2.62}$$

One can write Hamilton's equations of motion for a set of particles assuming positions r with momenta p as;

$$\dot{r}_1 = \frac{\partial H}{\partial p_i}(p, r) \tag{2.63}$$

$$\dot{p}_i = -\frac{\partial H}{\partial r_i}(p, r) \tag{2.64}$$

Newton's equations of motion can be solved by substituting equations 2.62 into 2.63 and 2.64. The outcome of the analysis gives the following set of equations;

$$\dot{r}_i(t) = v_i(t) \tag{2.65a}$$

$$\ddot{r}_i = \dot{v}_i(t) = a_i = \frac{f_i(t)}{m_i} \tag{2.65b}$$

$$f_i(t) = -\frac{\partial}{\partial r_i} V \tag{2.65c}$$

where a_i is the acceleration of particle i and $f_i(t)$ is the force exerted on the particle i. $V(r_i(t))$ is the potential energy and $r_i(t)$ is the position vector of the ith particle at time t. Equation 2.65b is the Newton's second order differential equation with respect to $r(t)$, and can be written as,

$$\ddot{r}(t) = \frac{1}{m_i} f_i(r(t)) \tag{2.66}$$

This implies that only positions of particles are important for solving equation 2.66, hence most molecular dynamics codes involve the use of $x-, y-, z-$coordinate systems, that is, the positions to evaluate forces at every step rather than using potential energy as defined in equation 2.65c. The idea behind equation 2.66 is that it is cheaper in-terms of computational time as compared to equation 2.65c.

Computer algorithms used in simulations include Verlet[79] and Beeman[80] schemes and these algorithms are used to integrate above Newton's equations of motion. I will discuss both algorithms in the following sections. The algorithms use the position, velocity and force components.

2.3.1 The Verlet algorithm

Let's use Taylor expansion to show how the positions and the velocities of particles change in Δt at time t_n (time at the current position), where $\Delta t = t_{n+1} - t_n = t_n - t_{n-1}$. One obtains the following equations at time $t_n + \Delta t$ (next position after time step Δt) and $t_n - \Delta t$ (the previous position, i.e. before the time step Δt);

$$r_{n+1} = r_n + \dot{r}_n \Delta t + \frac{1}{2!}\ddot{r}_n(\Delta t)^2 + \frac{1}{3!}\dddot{r}_n(\Delta t)^3 + O((\Delta t)^4) \tag{2.67}$$

and

$$r_{n-1} = r_n - \dot{r}_n \Delta t + \frac{1}{2!}\ddot{r}_n(\Delta t)^2 - \frac{1}{3!}\dddot{r}_n(\Delta t)^3 + O((\Delta t)^4) \tag{2.68}$$

Adding equation 2.67 to equation 2.68, one obtain;

$$r_{n+1} + r_{n-1} = 2r_n + \ddot{r}_n(\Delta t)^2 + O((\Delta t)^4) \tag{2.69}$$

Substituting equation 2.66 into equation 2.69 and rearrange terms, one obtains the Verlet algorithm[79];

$$r_{n+1} = 2r_n - r_{n-1} + \frac{1}{m}f_n(\Delta t)^2 + O((\Delta t)^4) \tag{2.70}$$

One can do the same for the velocity by subtracting equation 2.68 from equation 2.67 and substitute equation 2.66 into the resulting equation. Rearranging the equation, one obtain the expression for the velocity of particles;

$$v_n = \frac{1}{2\Delta t}(r_{n+1} - r_{n-1}) - \frac{1}{6m}\dot{f}_n(\Delta t)^2 + O((\Delta t)^4) \tag{2.71}$$

At this point, one may think of using time-reversal invariant expression for $\dot{f}_n$, where,

$$\dot{f}_n = \frac{1}{2\Delta t}(f_{n+1} - f_{n-1}) \tag{2.72}$$

Note that if the terms having $\dot{f}_n$ are omitted, then one obtains the standard velocity Verlet algorithm, equation 2.73, and by substituting equation 2.72 into equation 2.71, one obtains velocity Verlet algorithm,[81] equation 2.74.

$$v_n = \frac{1}{2\Delta t}(r_{n+1} - r_{n-1}) \tag{2.73}$$

$$v_n = \frac{1}{2\Delta t}(r_{n+1} - r_{n-1}) - \frac{1}{12m}(f_{n+1} - f_{n-1})\Delta t + O((\Delta t)^4) \tag{2.74}$$

2.3.2 Leapfrog algorithm

To get the velocity in Verlet algorithm, positions of the particles have to be calculated first. This is disadvantageous because there is no velocity term in the equation. As a result, leapfrog algorithm[81] was then developed. In the leapfrog method, the velocities for the next time step $v(t_n + \Delta t)$ are calculated from the velocities from the previous time step $v(t_n - \Delta t)$. The derivation of leapfrog algorithm is similar to the Verlet algorithm. In this case, the positions or velocities are expanded in Tylor series at time $(t_n + \frac{1}{2}\Delta t)$ and $(t_n - \frac{1}{2}\Delta t)$ in $\frac{1}{2}\Delta t$ at time t_n. Then I get,

$$r_{n+\frac{1}{2}} = r_n + \frac{1}{2}\dot{r}_n\Delta t + \frac{1}{8}\ddot{r}_n(\Delta t)^2 + \frac{1}{48}\dddot{r}_n(\Delta t)^3 + O((\Delta t)^4) \qquad (2.75)$$

and

$$r_{n-\frac{1}{2}} = r_n - \frac{1}{2}\dot{r}_n\Delta t + \frac{1}{8}\ddot{r}_n(\Delta t)^2 - \frac{1}{48}\dddot{r}_n(\Delta t)^3 + O((\Delta t)^4) \qquad (2.76)$$

The same procedure as for Verlet algorithm is followed. Equation 2.76 is subtracted from equation 2.75 to get,

$$r_{n+\frac{1}{2}} - r_{n-\frac{1}{2}} = \dot{r}_n\Delta t + \frac{1}{24}\dddot{r}_n(\Delta t)^3 + O((\Delta t)^4) \qquad (2.77)$$

Rearranging equation 2.77 and using equation 2.65b and 2.66 ($\ddot{r} = \frac{\partial \ddot{r}}{\partial r} = \frac{1}{m}\frac{\partial f}{\partial r} = \frac{1}{m}\dot{f}$) and then shifting the time points by $\frac{1}{2}\Delta t$, one obtains the expression for leapfrog algorithm;

$$r_{n+1} = r_n + v_{n+\frac{1}{2}}\Delta t + \frac{1}{24m}\dot{f}_{n+\frac{1}{2}}(\Delta t)^3 + O((\Delta t)^5) \qquad (2.78)$$

I can simplify the above equation by eliminating the function $\dot{f}_{n+\frac{1}{2}}$ from above expression using its time-reversal invariant expression and obtain;

$$r_{n+1} = r_n + v_{n+\frac{1}{2}}\Delta t + \frac{1}{24m}(f_{n+1} - f_n)(\Delta t)^2 + O((\Delta t)^5) \qquad (2.79)$$

where;

$$\dot{f}_{n+\frac{1}{2}} = \frac{1}{\Delta t}(f_{n+1} - f_n) \qquad (2.80)$$

Similarly, the expression for velocity can be obtained by expanding the velocity at time $t_n + \frac{1}{2}\Delta t$ and $t_n - \frac{1}{2}\Delta t$. The above procedure for positions is followed to get,

$$v_{n+\frac{1}{2}} = v_{n-\frac{1}{2}} + \frac{1}{m}f_n\Delta t + \frac{1}{24m}\ddot{f}_n(\Delta t)^3 + O((\Delta t)^5) \qquad (2.81)$$

Just like Verlet algorithm, omitting the terms containing $\dot{f}$ and $\ddot{f}$ leads to the standard Leap-Frog algorithm;

$$r_{n+1} = r_n + v_{n+\frac{1}{2}}\Delta t + O((\Delta t)^3) \qquad (2.82)$$

and the velocity;

$$v_{n+\frac{1}{2}} = v_{n-\frac{1}{2}} + \frac{1}{m}f_n\Delta t + O((\Delta t)^3) \tag{2.83}$$

The velocity at t_n can be computed by averaging the velocities $v(t_n + \Delta t)$ and $v(t_n - \Delta t)$ and the positions are computed from equation 2.82.

The equation 2.81 can be simplified further and one can also maintain its accuracy by substituting $\ddot{f}_n$ with its time-reversal invariant expression, equation 2.84, to obtain velocity Leap-Frog expression, equation 2.85.

$$\ddot{f}_n = \frac{1}{(\Delta t)^2}(f_{n+1} - 2f_n + f_{n-1}) \tag{2.84}$$

Velocity Leap-Frog expression;

$$v_{n+\frac{1}{2}} = v_{n-\frac{1}{2}} + \frac{1}{m}f_n\Delta t + \frac{1}{24m}(f_{n+1} - 2f_n + f_{n-1})\Delta t + O((\Delta t)^5) \tag{2.85}$$

2.3.3 The velocity Verlet scheme

The Verlet equations do not have the velocity term, therefore, it is difficult to get velocities of the particles without knowing the positions first. The velocities are obtained only after the computation of new positions. These new position are obtained from current position $r(t)$ and previous positions $r(t - \Delta t)$. The simulation requires some calculation of $r(t - \Delta t)$ at the beginning $(t = 0)$ for the computations to start. There were some developments on the Verlet algorithm later on, as a result of this deficiency. Velocity Verlet[82] method was developed and it computes positions, velocity and acceleration at the same time. The velocity Verlet scheme[83] is as follows;

1    $\vec{x}(t + \Delta t) = \vec{x}(t) + \vec{v}(t)\Delta t + \frac{1}{2}\vec{a}(t)(\Delta t)^2$

2    $\vec{v}(t + \frac{\Delta t}{2}) = \vec{v}(t) + \frac{\vec{a}(t)\Delta t}{2}$

3    $\vec{a}(t + \Delta t) = -\frac{1}{m}\vec{\nabla}V(r(t + \Delta t))$

4    $\vec{v}(t + \Delta t) = \vec{v}(t + \frac{\Delta t}{2}) + \frac{\vec{a}(t+\Delta t)\Delta t}{2}$

2.3.4 The Beeman Scheme

Another computational algorithm that is well known in the computational simulations is the Beeman algorithm[80] and it is similar to the velocity Verlet algorithm. The scheme is moderately computationally more expensive and has better accuracy than the Verlet algorithm. The position and the velocity expressions, respectively, are calculated from;

$$r_i(t + \Delta t) = r_i(t) + v_i(t)\Delta t + \frac{2}{3}a_i(t)\Delta t^2 - \frac{1}{6}a_i(t - \Delta t)\Delta t^2 + O(\Delta t^4) \qquad (2.86)$$

and

$$v_i(t + \Delta t) = v_i(t) + \frac{1}{3}a_i(t + \Delta t)\Delta t + \frac{5}{6}a_i(t)\Delta t - \frac{1}{6}a_i(t - \Delta t)\Delta t + O(\Delta t^3) \qquad (2.87)$$

The equations of motion can also be integrated following the modified Beeman[84] scheme.

1 $r(t + \Delta t) = r(t) + \dot{r}(t)\Delta t + \frac{\Delta t^2}{6}[4\ddot{r}(t) - \ddot{r}(t - \Delta t)]$

2 $\dot{r}_p(t + \Delta t) = \dot{r}(t) + \frac{\Delta t}{2}[3\ddot{r}(t) - \ddot{r}(t - \Delta t)]$

3 $\ddot{r}(t + \Delta t) = F[r_i(t + \Delta t), \dot{r}_i^p(t + \Delta t), \quad i = 1, \ldots, N]/m$

4 $\dot{r}_c(t + \Delta t) = \dot{r}(t) + \frac{\Delta t}{6}[2\ddot{r}(t + \Delta t) + 5\ddot{r}(t) - \ddot{r}(t - \Delta t)]$

5 Replace $\dot{r}_p$ with $\dot{r}_c$ and go to 3. Iterate to convergence.

6 Replace t with $t + \Delta t$ and go to step 1 to determine the parameters of $t + \Delta t$.

Since r is the dynamic variable, then $\dot{r}_p$ and $\dot{r}_c$ respectively are the predicted and corrected velocities. N is the total number of particles studied and F is the total force acting on the particle. The total force is calculated using equation 2.65c with the potential energy, $V(r(t))$, of the system calculated from the force field. Accelerations and velocities of the particles are calculated from equation 2.66 or the expressions described above. From the above scheme, only the positions of the particles are important in computations of all parameters.

2.4 Explicit Models for Tautomeric/Conformational Search

Many ingenious theoretical methods and algorithms aiming to determine the most stable molecular structures have been developed.[30, 31] The most common methods are finite temperature Monte Carlo,[32] molecular dynamics,[76] basin-hopping[34] methods, as well as genetic algorithms.[35] These search methods are classed as stochastic because there is no guarantee that they will identify the global minimum. These stochastic search methods are applicable to large and complex systems, they typically rely on classical force fields and therefore their accuracy is limited. In addition, there are brute force methods based on systematic scanning the potential energy surface (PES), also known as the Combinatorial Computational Cheminformatics C^3. Some of them specialize in

molecular tautomers[21, 22, 85] and others in conformers[24–26] but in this dissertation, I present a method/tool that would deal with tautomers and conformers on equal footing (chapter 5).

2.4.1 Stochastic Search Methods

The stochastic methods follow common procedure of performing random sampling on the potential energy surface followed by evaluation of fitness of the randomly generated structures.

Monte-Carlo. Monte Carlo (MC) methods generate a sequence of configurations and addresses them. These configurations are generated randomly and they depend fully on their immediate predecessor but there is no relation between the current step and previous steps, hence, time irreversible. Unlike MD, the MC methods are not time dependent but one can follow and analyse the structure of the system as it evolves hence the sampling is performed under different ensembles. This makes it possible to study even complex structures with MC methods. Here, the geometry of the molecule is changed randomly based on the degrees of freedom the molecule has. Then the decision is made whether to accept the new step based on the Metropolis criterion.[32]

Configurations are generated by Cartesian moves or torsional moves depending on the choice of the researcher and the type of the molecule. In Cartesian moves, a single atom/molecule is randomly moved. For large systems, several atoms/molecules are moved. In torsional moves, the dihedral angles are changed. At every MC iteration, three random numbers ranging between 0 and 1 are being generated and the new cartesian coordinates of the atom are given by;

$$x_{new} = x_{old} + rand * \delta r \tag{2.88a}$$

$$y_{new} = y_{old} + rand * \delta r \tag{2.88b}$$

$$z_{new} = z_{old} + rand * \delta r \tag{2.88c}$$

In case of the torsional moves;

$$\phi_{new} = \phi_{old} + rand * \delta\phi \tag{2.89}$$

where $rand$ is the random number generated, δr and $\delta\phi$ are maximum possible displacement/angle in any direction. These implies that the maximum change is governed by δr and $\delta\phi$.

Some algorithms implemented in MC codes perform minimisation after every MC step. The idea is to relax the strained bonds and angles and then to calculate the energy based on any of the local minimum converged to. This is followed by screening the current configuration based on its energy. The configurations having lower energies than preceding configurations are automatically accepted and then used as the starting point for the next MC step. Otherwise, a random number is generated between 0 and 1 and the Boltzmann factor P is calculated

$$P = exp\left(-\frac{\Delta V(r)}{k_B T}\right) \qquad \Delta V = V(r_{new}) - V(r_{old}) \qquad (2.90)$$

and then compared to this random number. ΔV is the change of potential energy of the system which depends only on the positions of the atoms. If the Boltzmann factor is greater than the random number, then the step is accepted and the current configuration is kept for the next MC step. But if the Boltzmann factor is less that the random number, then the step is rejected and the previous configuration is retained and used in the next MC step. The rate of acceptance depends on the maximum displacement/angle. The value should be chosen wisely. If the value is very small, then most of the steps will be accepted and the system will take more time to cover the phase space because the current state and the previous state will be similar. If the value is large, then most of the steps will be rejected since the energy change will be large and therefore, improper sampling of the system. The above algorithm leads to the sampling consistent with the Boltzmann distribution.

Simulated annealing. In MD and MC methods discussed so far, the system temperature is set as a guiding parameter for sampling the potential energy surface. The phase space is thought to be well sampled for sufficiently long runs at elevated temperatures because the molecule will have high enough kinetic energy to overcome barriers on the potential energy surface. If the temperature of the system is allowed to decrease during the course of the simulation, starting from elevated temperatures, then the system might settle into the local/global minimum structure. For sufficiently long MD or MC run with infinitely slowly cooling, the resulting minimum is the global minimum. This idea of tweaking the system's temperature coupled to the MD or MC run is called the simulated annealing[86] technique. This method is very slow and more disadvantageous because at low system's temperature, the molecule might settle in a local minimum whereas elevated temperatures might lead to dissociation of the molecule which result in a failed search. This technique is covered in chapter 9.

Chapter 3

Molecular Orbital Theory

Quantum mechanics is an important tool for investigation of a wide range of problems in molecular physics. The advancements in supercomputers and computational methods has enabled breakthrough in solving chemical problems that were beyond reach only few years ago. These problems are related to the determination of molecular structure and reactivity, and the spectroscopic parameters that can be used to interpret the experimental spectroscopic data. This chapter will cover the important advanced electronic structure theories[10, 87–91] used in this study. I focused mainly on the non relativistic molecular Born-Oppenheimer electronic Schrödinger equation and paid particular attention to the systems of identical particles and constructed their wave functions. The eigenfunctions of the Hamiltonian are then defined based on the particles in the system and yield the corresponding eigenvalues, the energies. These methods have been implemented in the quantum chemistry codes and they start from the Hartree-Fock method and extend to coupled cluster method. This chapter will also covers density functional theory. These methods have higher accuracy than the force fields discussed in the previous chapter.

3.1   Fundamental Aproximations

3.1.1   The Schrödinger Equation

Particles obey laws of quantum mechanics. The description of the dynamics of a system comprised of N particles can be understood from generalisation of the dynamics of a single particle. The evolution of the wave function Ψ with time according to the general

form of time-dependent Schrödinger equation is mathematically expressed as;

$$i\hbar \frac{\partial}{\partial t} \Psi(\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_N, t) = \hat{H} \Psi(\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_N, t) \tag{3.1}$$

where the Hamiltonian $\hat{H}$ of one-particle system $\hat{P}^2/(2m) + \hat{V}(\vec{r})$ is generalized to N-particle system. This form of the Schrödinger equation is often separated into equations that depend explicitly on time and those that depend explicitly on space variation of the wave function Ψ. In most cases, calculations based on this formulation is possible when the potential energy is time-independent. The one dimensional form of equation 3.1 is

$$\hat{H} = \sum_{j=1}^{N} \frac{\hat{P}_j^2}{2m_j} + \hat{V}(\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_N) = -\sum_{j=1}^{N} \frac{\hbar^2}{2m_j} \nabla_j^2 + \hat{V}(\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_N) \tag{3.2}$$

where m_j and P_j are the mass and the momentum of the jth-particle, respectively and the operator $\hat{V}$ corresponds to the total potential energy. This form of equation allows techniques of separation of variables to be used, in which a trial solution is of the form;

$$\Psi(\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_N, t) = \psi(\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_N)\theta(t). \tag{3.3}$$

If I inset this equation 3.3 into the Schrödinger equation 3.1 and use the Hamiltonian defined by equation 3.2, then one obtains;

$$-\sum_{j=1}^{N} \frac{\hbar^2}{2m_j} \theta \frac{d^2\psi}{dr^2} + \hat{V}(\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_N)\psi\theta = i\hbar\psi \frac{d\theta}{dt} \tag{3.4}$$

which can be simplified by diving by $\psi\theta$ on both sides;

$$-\sum_{j=1}^{N} \frac{\hbar^2}{2m_j} \frac{1}{\psi} \frac{d^2\psi}{dr^2} + \hat{V}(\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_N) = i\hbar \frac{1}{\theta} \frac{d\theta}{dt} \tag{3.5}$$

The l.h.s. of the equation is a function of coordinates $\vec{r}$ and any change in coordinates does not affect the r.h.s. of the equation. This implies that the l.h.s. can be equated to a constant E and the time dependent equation separates into the following two differential equations.

$$-\sum_{j=1}^{N} \frac{\hbar^2}{2m_j} \frac{d^2\psi}{dr^2} + \hat{V}(\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_N)\psi = E\psi \tag{3.6}$$

$$i\hbar \frac{d\theta}{dt} = E\theta \tag{3.7}$$

and the latter(time dependent) has the solution $\theta \propto e^{-iEt/\hbar}$. In many cases $\hat{V}$ in equation 3.2 does not depend on time and then the state of the N-particle system is described by the complete wave-function

$$\Psi(\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_N, t) = \psi(\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_N)e^{-iEt/\hbar} \tag{3.8}$$

with ψ as the solution to the time-independent Schrödinger equation and the exponential part is the time dependent part with E as the total energy of the system. These are termed stationary states. Schrödinger equation can be solved for many different systems since the Hamiltonian of the system caters for different interactions within a system. The Hamiltonian of the system includes the kinetic energy of the electrons $\hat{T}_e$ and those due to nuclei $\hat{T}_n$, the interaction potential energy between electrons and nuclei $\hat{V}_{e-n}$, electrons-electrons $\hat{V}_{e-e}$ and also nuclei-nuclei $\hat{V}_{n-n}$. To be more precise, for a system with N atomic nuclei $(a, b, ...)$ with atomic numbers $(Z_a, Z_b, ...)$ and n electrons $(i, j, ...)$, the expanded non-relativistic Hamiltonian would be expressed as,

$$\hat{H} = \hat{T}_e + \hat{T}_n + \hat{V}_{e-n} + \hat{V}_{e-e} + \hat{V}_{n-n}$$
$$= -\frac{\hbar^2}{2m_e}\sum_i \nabla_i^2 - \frac{\hbar^2}{2}\sum_a \frac{\nabla_a^2}{m_a} - ke^2 \sum_{i,a} \frac{Z_a}{|r_i - r_a|}+$$
$$+ \frac{ke^2}{2}\sum_{i \neq j} \frac{1}{|r_i - r_j|} + \frac{ke^2}{2}\sum_{a \neq b} \frac{Z_a Z_b}{|r_a - r_b|} \tag{3.9}$$

where $k = \frac{1}{4\pi\varepsilon_0}$, ε_0 is the permittivity of free space, e is the elementary charge, m_e and m_a are the mass of the electron and nucleus, respectively, and Z is the charge of the nucleus.

Construction of many-electron wave function from spin-orbitals. The exact eigenstates of the many particle system described by equation 3.2 are generally difficult to obtain. One may deduce some of their properties through symmetry schemes. Consider an exchange operator P_{ij} (also known as the permutation operator) that when operating on an $N-$identical particle wave function $\psi(\chi_1, \ldots, \chi_i, \ldots, \chi_j, \ldots, \chi_N)$ it interchanges the coordinates of the ith and the jth-particle and χ_i is the spatial and spin coordinates of the particles i. Then,

$$P_{ij}\psi(\chi_1, \ldots, \chi_i, \ldots, \chi_j, \ldots, \chi_N) = \psi(\chi_1, \ldots, \chi_j, \ldots, \chi_i, \ldots, \chi_N) \tag{3.10}$$

and

$$P_{ij}^2\psi(\chi_1, \ldots, \chi_i, \ldots, \chi_j, \ldots, \chi_N) = P_{ij}\psi(\chi_1, \ldots, \chi_j, \ldots, \chi_i, \ldots, \chi_N) \tag{3.11a}$$
$$= \psi(\chi_1, \ldots, \chi_i, \ldots, \chi_j, \ldots, \chi_N) \tag{3.11b}$$

From equation 3.11, we see that two successive application of exchange operator P_{ij} leave the wave function unchanged. Let us consider the states that are eigenstates of all permutation operator in this system with $N-$particles. All the eigenstates that satisfy the Schrödinger equation $H|\Psi\rangle = \varepsilon|\Psi\rangle$ are also eigenstates of the permutation operator P_{ij} and the states must satisfy $P_{ij}|\Psi\rangle = c_{ij}|\Psi\rangle$. Here P_{ij} is self adjoint/Hermitian and unitary, $P_{ij} = P_{ij}^\dagger = P^{-1}$ so that $P_{ij}^\dagger P_{ij} = 1$ and I assumed that $|\Psi\rangle$ is an eigenvector of all permutations so that it is the eigenvector of all transpositions. Then

$$P_{ij}|\Psi = c_{ij}|\Psi\rangle \tag{3.12}$$

requires

$$P_{ij}^2|\Psi\rangle = c_{ij}^2|\Psi\rangle = |\Psi\rangle \tag{3.13}$$

so that the eigenvalue c_{ij}^2 of $I = P_{ij}^2$ are simply

$$c_{ij}^2 = 1 \qquad and \qquad c_{ij}^{transpose} = (\pm 1)^{n_{ij}}, \tag{3.14}$$

hence two eigenvalues. n_{ij} is the number of transpositions in which P^{th} permutation can be splitted. Then I can rewrite the wave function as;

$$P_{ij}\psi(\chi_1, \ldots, \chi_i, \ldots, \chi_j, \ldots, \chi_N) = \pm\psi(\chi_1, \ldots, \chi_i, \ldots, \chi_j, \ldots, \chi_N) \tag{3.15}$$

I used the adjectives *symmetric* and *antisymmetric* for wave functions corresponding to eigenvalue $+1$ (Ψ_s) and -1 (Ψ_a) respectively, all with respect to the interchange of the pair (i, j).

$$\psi_s(\chi_1, \ldots, \chi_i, \ldots, \chi_j, \ldots, \chi_N) = \psi_s(\chi_1, \ldots, \chi_j, \ldots, \chi_i, \ldots, \chi_N) \tag{3.16}$$

$$\psi_a(\chi_1, \ldots, \chi_i, \ldots, \chi_j, \ldots, \chi_N) = -\psi_a(\chi_1, \ldots, \chi_j, \ldots, \chi_i, \ldots, \chi_N) \tag{3.17}$$

System of distinguishable non-interacting particles. If I consider a system of N non-interacting, distinguishable particles, each particle having different mass m_i and experiencing different potential $\hat{V}_i(\chi_i)$, then the potential is given by;

$$\hat{V}(\chi_1, \chi_2, \ldots, \chi_N) = \sum_{i=1}^{N} \hat{V}_i(\chi_i) \tag{3.18}$$

and the Schrodinger equations by;

$$\hat{H}\psi_{n_1,n_2,\cdots,n_N}(\chi_1, \chi_2, \ldots, \chi_N) = E_{n_1,n_2,\cdots,n_N}\psi_{n_1,n_2,\cdots,n_N}(\chi_1, \chi_2, \ldots, \chi_N) \tag{3.19}$$

which separates into N one particle equations

$$\left[-\frac{\hbar^2}{2m_i}\nabla_i^2 + \hat{V}_i(\chi_i)\right]\psi_{n_i}(\chi_i) = \varepsilon_{n_i}\psi_{n_i}(\chi_i) \tag{3.20}$$

and the solution to these equations yield one-particle energies ε_i with the total energy of the system equal to the sum of these one-particle energies

$$E_{n_1,n_2,\cdots,n_N} = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_N = \sum_{i=1}^{N}\varepsilon_{n_i} \tag{3.21}$$

and the total wave function as the product of single particle states ψ_i (spin orbitals)

$$\psi_{n_1,n_2,\cdots,n_N}(\chi_1,\chi_2,\ldots,\chi_N) = \psi_{n_1}(\chi_1)\psi_{n_2}(\chi_2)\cdots\psi_{n_N}(\chi_N) = \prod_{i=1}^{N}\psi_{n_i}(\chi_i) \tag{3.22}$$

Symmetric and asymmetric functions In quantum particles, the wave functions are either totally symmetric or totally antisymmetric under the interchange of particles for systems of N identical particles. That is, the identical particles lose their identity as described by equations 3.16/3.17. This can be understood by constructing symmetric $\psi_s(\chi_1,\chi_2)$ and asymmetric $\psi_a(\chi_1,\chi_2)$ wave functions from any wave function $\psi(\chi_1,\chi_2)$

$$\psi_s(\chi_1,\chi_2) = \frac{1}{\sqrt{2}}\left[\psi(\chi_1,\chi_2) + \psi(\chi_2,\chi_1)\right] \tag{3.23}$$

$$\psi_a(\chi_1,\chi_2) = \frac{1}{\sqrt{2}}\left[\psi(\chi_1,\chi_2) - \psi(\chi_2,\chi_1)\right] \tag{3.24}$$

and the normalisation factor is given by $\frac{1}{\sqrt{2}}$.

System of identical non-interacting particles For a system of N non-interacting identical particles with equal masses $m_i = m$, and experiencing the same potential $\hat{V}_i(\chi_i)$, the Schrödinger equation is similar to the equation 3.20 where the system is written as N separate identical one-body equations

$$\left[-\frac{\hbar^2}{2m}\nabla_i^2 + \hat{V}_i(\chi_i)\right]\psi_{n_i}(\chi_i) = \varepsilon_{n_i}\psi_{n_i}(\chi_i) \tag{3.25}$$

A symmetric or asymmetric wave functions for a system of two identical, non-interacting particles can be constructed in terms of one body wave functions, equations 3.23 and 3.24

$$\psi_s(\chi_1,\chi_2) = \frac{1}{\sqrt{2}}\left[\psi_{n_1}(\chi_1)\psi_{n_2}(\chi_2) + \psi_{n_1}(\chi_2)\psi_{n_2}(\chi_1)\right] \tag{3.26}$$

$$\psi_a(\chi_1,\chi_2) = \frac{1}{\sqrt{2}}\left[\psi_{n_1}(\chi_1)\psi_{n_2}(\chi_2) - \psi_{n_1}(\chi_2)\psi_{n_2}(\chi_1)\right] \qquad n_1 \neq n_2 \tag{3.27}$$

If $n_1 = n_2 = n$, then $\psi_s(\chi_1, \chi_2) = \psi_n(\chi_1)\psi_n(\chi_2)$ and $\psi_a(\chi_1, \chi_2) = 0$. The general form of symmetric ψ_s and asymmetric ψ_a two body wave functions are given as;

$$\psi_s(\chi_1, \chi_2) = \frac{1}{\sqrt{2!}} \sum_P \hat{P}\psi_{n_1}(\chi_1)\psi_{n_2}(\chi_2) \tag{3.28a}$$

$$\psi_a(\chi_1, \chi_2) = \frac{1}{\sqrt{2!}} \sum_P (-1)^P \hat{P}\psi_{n_1}(\chi_1)\psi_{n_2}(\chi_2) \tag{3.28b}$$

where $\hat{P}$ is the permutation operator and $(-1)^P$ is equal to $+1$ for an even permutation and equal to -1 for an odd permutation. The sum is over all possible permutations. I can use the same logic and construct a wave function of identical, non-interacting three-body system. The symmetric and asymmetric wave functions respectively are given as;

$$\psi_s(\chi_1, \chi_2, \chi_3) = \frac{1}{\sqrt{3!}} \sum_P \hat{P}\psi_{n_1}(\chi_1)\psi_{n_2}(\chi_2)\psi_{n_3}(\chi_3) \tag{3.29a}$$

$$\psi_a(\chi_1, \chi_2, \chi_3) = \frac{1}{\sqrt{3!}} \sum_P (-1)^P \hat{P}\psi_{n_1}(\chi_1)\psi_{n_2}(\chi_2)\psi_{n_3}(\chi_3) \tag{3.29b}$$

and the generalized wave function from equations 3.28 and 3.29 for N identical, non-interacting system as

$$\psi_s(\chi_1, \chi_2, \cdots, \chi_N) = \frac{1}{\sqrt{N!}} \sum_P \hat{P}\psi_{n_1}(\chi_1)\psi_{n_2}(\chi_2)\cdots\psi_{n_N}(\chi_N) \tag{3.30}$$

$$\psi_a(\chi_1, \chi_2, \cdots, \chi_N) = \frac{1}{\sqrt{N!}} \sum_P (-1)^P \hat{P}\psi_{n_1}(\chi_1)\psi_{n_2}(\chi_2)\cdots\psi_{n_N}(\chi_N). \tag{3.31}$$

The latter, known as the **Slater determinant**[92], can be rewritten in an $N \times N$ determinantal form involving spin orbitals only,

$$\psi_a(\chi_1, \chi_2, \cdots, \chi_N) = \frac{1}{\sqrt{N!}} \begin{bmatrix} \chi_1(\mathbf{x}_1) & \chi_1(\mathbf{x}_2) & \cdots & \chi_1(\mathbf{x}_N) \\ \chi_2(\mathbf{x}_1) & \chi_2(\mathbf{x}_2) & \cdots & \chi_2(\mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_N(\mathbf{x}_1) & \chi_N(\mathbf{x}_2) & \cdots & \chi_N(\mathbf{x}_N) \end{bmatrix} \tag{3.32}$$

$$\equiv |\chi_1\chi_2\cdots\chi_N| \tag{3.33}$$

If any two particle occupy the same one particle state, then the determinant vanishes due to two rows being identical. This is known as the Pauli exclusion principle. I will revisit the Slater determinants in section 3.2. Approximations are required when solving complex differential equations such as equation 3.1 with Hamiltonian defined by equation 3.9 and the wave function represented by equations 3.30 and 3.31.

3.1.2 The Born-Oppenheimer approximation

The atomic nuclei are much heavier than the orbiting electrons. Any movement made by the nuclei result in electrons adjusting themselves instantly to the current positions of the nuclei because they move faster than the nuclei. One can approximate the nuclei to be stationary with respect to the orbiting electrons. Within this approximation, the kinetic energy of the nuclei in equation 3.9 is set to zero, $T_n = 0$ and the nuclei-nuclei interactions can be considered to be constant, $\hat{T}_{n-n} = k$. Note that any constant has no effect on the operator eigen functions, thus equation 3.9 reduces to;

$$\hat{H}_{elec} = -\frac{\hbar^2}{2m_e} \sum_i \nabla_i^2 - ke^2 \sum_{i,a} \frac{Z_a}{|r_i - r_a|} + \frac{ke^2}{2} \sum_{i \neq j} \frac{1}{|r_i - r_j|} \qquad (3.34)$$

which is the electronic Hamiltonian $\hat{H}_{elec}$ and the electronic wave function $\psi_{elec} = \psi_{elec}(r_i; r_A)$ is the solution to a Schrödinger equation $\hat{H}_{elec}\psi_{elec} = \xi_{elec}\psi_{elec}$. The electronic wave function depends explicitly on the electronic coordinates and parametrically on the nuclear coordinates whereas the electronic energy depends only on nuclear coordinates $\xi_{elec} = \xi_{elec}(r_a)$. I can now write the total energy of the system ξ_{tot} for a fixed nuclei as;

$$\xi_{tot} = \xi_{elec} + \frac{ke^2}{2} \sum_{a \neq b} \frac{Z_a Z_b}{|r_a - r_b|} \qquad (3.35)$$

which has the additional term, the nuclear repulsion term, $\hat{V}_{n-n}$. I can use the same assumptions and solve for the nuclear motion in the average field of the electrons. Then equation 3.9 would be rewritten as;

$$\hat{H}_{nucl} = \xi_{elec}(r_a) + \hat{V}_{n-n} + \hat{T}_n \qquad (3.36)$$

$$= \xi_{tot}(r_a) - \frac{\hbar^2}{2} \sum_a \frac{\nabla_a^2}{m_a} \qquad (3.37)$$

where the potential energy of the nuclei $\hat{V}_{n-n}$ has been embedded on the total energy $\xi_{tot}(r_a)$. Here, the electronic energy plays a major role and the Schrödinger equation $\hat{H}_{nucl}\psi_{nucl} = \xi\psi_{nucl}$ leads to the energy levels for molecular vibrations and rotations which in turn are fundamental in spectroscopy. The nuclear wave function depends on nuclear coordinates $\psi_{nucl} = \psi_{nucl}(r_a)$ and ξ is the Born-Oppenheimer approximation[11] to the total energy given by equation 3.1. In the context of Born-Oppenheimer approximation, the nuclei move on the electronic potential energy surface described by electronic Schrödinger equation. If these corrections are introduced in the adiabatic approximations, then the result obtained will be a good approximation and this is the Born-Oppenheimer approximation. From now onwards, I will consider only the electronic Hamiltonians and electronic wave functions unless otherwise stated.

3.1.3   Trial Wave Functions (Variational Principle)

Let's start from the trial wave function $\tilde{\Phi}$ which can be written in terms of eigenfunctions of the Hamiltonian ψ_i forming a linear combination, with c_i as the coefficient;

$$\tilde{\Phi} = \sum_i c_i \psi_i \tag{3.38}$$

and the wave function $\tilde{\Phi}$ also satisfies the normalization condition,

$$\langle \tilde{\Phi} | \tilde{\Phi} \rangle = \sum_{ij} c_i^* c_j \langle \psi_i | \psi_j \rangle = \sum_i c_i^* c_i = \sum_i |c_i|^2 = 1 \tag{3.39}$$

The expectation value for the electronic Hamiltonian $\hat{H}_{elec}$ discussed above

$$\langle \tilde{\Phi} | \hat{H}_{elec} | \tilde{\Phi} \rangle = \left\langle \sum_i c_i \psi_i | \hat{H}_{elec} | \sum_j c_j \psi_j \right\rangle = \left\langle \sum_i c_i \psi_i | \sum_j c_j E_i \psi_j \right\rangle \tag{3.40a}$$

$$= \sum_{ij} E_i c_i^* c_j \langle \psi_i | \psi \rangle = \sum_i E_i |c_i|^2 \tag{3.40b}$$

Let us compare $\langle \tilde{\Phi} | \hat{H}_{elec} | \tilde{\Phi} \rangle$ with the exact ground state solution E_0;

$$\langle \tilde{\Phi} | \hat{H}_{elec} | \tilde{\Phi} \rangle - E_0 = \sum_i E_i |c_i|^2 - E_0 \sum_i |c_i|^2$$

$$= \sum_i |c_i|^2 (E_i - E_0) \geq 0 \tag{3.41}$$

The variational principle state that the expectation value E_0 is a lower limit for $\langle \tilde{\Phi} | \hat{H}_{elec} | \tilde{\Phi} \rangle$. Any trial wave function used yields energy greater than the ground state energy. In general form

$$\frac{\langle \tilde{\Phi} | \hat{H}_{elec} | \tilde{\Phi} \rangle}{\langle \tilde{\Phi} | \tilde{\Phi} \rangle} \geq E_0 \tag{3.42}$$

3.2   The Hartree-Fock Theory (Principles of)

Given the single determinant defined by equation 3.32, $|\Psi_0\rangle = |\chi_1 \chi_2 \cdots \chi_a \chi_b \cdots \chi_N\rangle$, the energy is minimized with respect to the spin orbitals, χ_i, provided the spin orbitals remain orthogonal and normalised i.e. $\int dx_1 \chi_a^*(1) \chi_b(1) = \langle a|b \rangle = \delta_{ab}$

$$L[\chi_a] = E_0[\chi_a] - \sum_{a=1}^N \sum_{b=1}^N \varepsilon_{ba} (\langle a|b \rangle - \delta_{ab}) \tag{3.43}$$

with the condition $\langle a|b \rangle - \delta_{ab} = 0$ and the first variation in L goes to zero, $\delta L = 0$ for infinitesimal change in spin orbitals, $\chi_a \to \chi_a + \delta\chi_a$.

$$\delta L = \delta E_0 - \sum_{a=1}^{N} \sum_{b=1}^{N} \varepsilon_{ba} \delta\langle a|b \rangle = 0 \tag{3.44}$$

My goal is to determine the Fock operator using equation 3.44 and I will start by expanding the δE_0 term. Note that equation 3.34 can be written as a sum of terms comprising single particle operators, each of which act on a single electronic coordinate and pair interaction operators acting on pair of electrons.

$$\hat{H}_{elec} = \sum_i \hat{h}_1(\chi_i) + \frac{1}{2} \sum_{i \neq j} \hat{h}_2(\chi_i, \chi_j) \tag{3.45}$$

where $\hat{h}_1(\chi_i)$ is the sum of the kinetic energy term and the nuclei-electron interaction term and $\hat{h}_2(\chi_i, \chi_j) = \frac{1}{r_{12}}$ is the pair interaction operator, the electron-electron interaction term. I can write the energy E_0 in equation 3.44 as;[87](pp. 118)

$$E_0[\chi_a] = \sum_{a=1}^{N} \langle a|\hat{h}|a \rangle + \frac{1}{2} \sum_{a=1}^{N} \sum_{b=1}^{N} \langle aa|bb \rangle - \langle ab|ba \rangle \tag{3.46}$$

using equation 3.45 where I have conveniently defined my integrals as;

$$\langle a|h|a \rangle = \int dx_1 \chi_a^*(1) h_1 \chi_b(1) \tag{3.47}$$

$$\langle aa|bb \rangle = \int dx_1 dx_2 \chi_a^*(1) \chi_a(1) r_{12}^{-1} \chi_b^*(2) \chi_b(2) \tag{3.48}$$

$$\langle ab|ba \rangle = \int dx_1 dx_2 \chi_a^*(1) \chi_b(1) r_{12}^{-1} \chi_b^*(2) \chi_a(2) \tag{3.49}$$

The Hartree-Fock equations are obtained from minimisation of the energy $E_0[\chi_a]$, so the variance of the energy δE_0 is given by;

$$\begin{aligned}
\delta E_0 = &\sum_{a=1}^{N} \langle \delta\chi_a|\hat{h}|\chi_a \rangle + \langle \chi_a|h|\delta\chi_a \rangle \\
&+ \frac{1}{2} \sum_{a=1}^{N} \sum_{b=1}^{N} \langle \delta\chi_a\chi_a|\chi_b\chi_b \rangle + \langle \chi_a\delta\chi_a|\chi_b\chi_b \rangle + \langle \chi_a\chi_a|\delta\chi_b\chi_b \rangle + \langle \chi_a\chi_a|\chi_b\delta\chi_b \rangle \\
&- \frac{1}{2} \sum_{a=1}^{N} \sum_{b=1}^{N} \langle \delta\chi_a\chi_b|\chi_b\chi_a \rangle + \langle \chi_a\delta\chi_b|\chi_b\chi_a \rangle + \langle \chi_a\chi_b|\delta\chi_b\chi_a \rangle + \langle \chi_a\chi_b|\chi_b\delta\chi_a \rangle
\end{aligned} \tag{3.50}$$

which reduces to two unique two electron integrals

$$\delta E_0 = \sum_{a=1}^{N} \langle \delta\chi_a|\hat{h}|\chi_a \rangle + \sum_{a=1}^{N} \sum_{b=1}^{N} \langle \delta\chi_a\chi_a|\chi_b\chi_b \rangle - \langle \delta\chi_a\chi_b|\chi_b\chi_a \rangle + C^* \tag{3.51}$$

The other variance I am interested on is $\varepsilon_{ba}\delta\langle a|b\rangle$. From equation 3.43, there is no variation in constant δ_{ab}, so

$$\delta\langle a|b\rangle = \langle\delta\chi_a|\chi_b\rangle + \langle\chi_a|\delta\chi_b\rangle \tag{3.52}$$

and then $\varepsilon_{ba}\delta\langle a|b\rangle$ can be approximated as,

$$
\begin{aligned}
\sum_{ab}\varepsilon_{ba}\delta\langle a|b\rangle &= \sum_{ab}\varepsilon_{ba}(\langle\delta\chi_a|\chi_b\rangle + \langle\chi_a|\delta\chi_b\rangle) \\
&= \sum_{ab}\varepsilon_{ba}\langle\delta\chi_a|\chi_b\rangle + \sum_{ab}\varepsilon_{ab}\langle\chi_b|\delta\chi_a\rangle \\
&= \sum_{ab}\varepsilon_{ba}\langle\delta\chi_a|\chi_b\rangle + \sum_{ab}\varepsilon_{ba}^*\langle\delta\chi_a|\chi_b\rangle^* \\
&= \sum_{ab}\varepsilon_{ba}\langle\delta\chi_a|\chi_b\rangle + C*
\end{aligned}
\tag{3.53}
$$

Substituting equation 3.51 and 3.53 into equation 3.44, one obtains

$$\delta L = \sum_{a=1}^{N}\langle\delta\chi_a|\hat{h}|\chi_a\rangle + \sum_{a=1}^{N}\sum_{b=1}^{N}\langle\delta\chi_a\chi_a|\chi_b\chi_b\rangle - \langle\delta\chi_a\chi_b|\chi_b\chi_a\rangle - \sum_{ab}\varepsilon_{ba}\langle\delta\chi_a|\chi_b\rangle$$
$$+ C^* = 0 \quad (3.54)$$

I can rewrite the equation 3.54 in terms of the coulomb operator $\hat{J}_b(1)$

$$\hat{J}_b(1)\chi_a(1) = \left[\int dx_2\chi_b^*(2)r_{12}^{-1}\chi_b(2)\right]\chi_a(1) \tag{3.55}$$

and exchange operator $\hat{K}_b(1)$

$$\hat{K}_b(1)\chi_a(1) = \left[\int dx_2\chi_b^*(2)r_{12}^{-1}\chi_a(2)\right]\chi_b(1) \tag{3.56}$$

$$= \left[\int dx_2\chi_b^*(2)r_{12}^{-1}\hat{P}_{12}\chi_b(2)\right]\chi_a(1) \tag{3.57}$$

defined by its effect when operating on a spin orbital $\chi_a(1)$. The operator $\hat{P}_{12}$ interchanges electron 1 and electron 2 when operating to the right.

$$\delta L = \sum_{a=1}^{N}\int dx_1\delta\chi_a^*(1)\left[\hat{h}(1)\chi_a(1) + \sum_{b=1}^{N}(\hat{J}_b(1) - \hat{K}_b(1))\chi_a(1) - \sum_{b=1}^{N}\varepsilon_{ba}\chi_b(1)\right]$$
$$+ C^* = 0 \quad (3.58)$$

The quantity in bracket must be zero because $\delta\chi_a^*$ is arbitrary.

$$\hat{h}(1)\chi_a(1) + \sum_{b=1}^{N}(\hat{J}_b(1) - \hat{K}_b(1))\chi_a(1) - \sum_{b=1}^{N}\varepsilon_{ba}\chi_b(1) = 0 \qquad (3.59)$$

$$or \qquad \underbrace{\left[\hat{h}(1) + \sum_{b=1}^{N}(\hat{J}_b(1) - \hat{K}_b(1))\right]}_{\text{Fock operator}}\chi_a(1) = \sum_{b=1}^{N}\varepsilon_{ba}\chi_b(1) \qquad (3.60)$$

The Fock operator $\hat{f}$ is defined above by the quantity in brackets and it is the sum of the core Hamiltonian operator $\hat{h}(1)$ and an effective one electron potential, the so called Hatree-Fock potential v^{HF};

$$v^{HF}(1) = \sum_b \hat{J}_b(1) - \hat{K}_b(1) \qquad (3.61)$$

$$= \sum_b \int dx_2\chi_b^*(2)r_{12}^{-1}(1 - \hat{P}_{12})\chi_b(2) \qquad (3.62)$$

The Fock operator in short form,

$$\hat{f}(1) = \hat{h}(1) + v^{HF} \qquad (3.63)$$

The above equation 3.60 reduces to

$$\hat{f}|\chi_a\rangle = \sum_{b=1}^{N}\varepsilon_{ba}|\chi_b\rangle \qquad (3.64)$$

There are several solutions for the above equation, each solution corresponding to a different set of ε_{ba}. The above equation may be written in canonical form using solutions of ε_{ba} that satisfies $\varepsilon_{ba} = \delta_{ba}\varepsilon_a$. I can use the matrices U that satisfies the relation $U^\dagger = U^{-1}, U^\dagger U = 1$ (orthonormal) and consider a new set of spin orbitals χ_a'. These new set of spin orbitals should satisfy $\chi_a' = \sum_b \chi_b U_{ba}$. From the fact that the matrix U is unitary, $|\det(U)|^2 = 1$, $\det(U) = e^{i\phi}$, and the transformed determinant should differ from the original determinant by only phase factor. I can now work on its effect on the $\hat{f}(1)$ and on the $\varepsilon_b a$. The coulomb and exchange terms of the Fock operator are the only terms that depend on the spin orbitals. The transformed sum of the coulomb operators

is

$$\sum_a \hat{J}'_a(1) = \sum_a \int dx_2 \chi_a^*(2) r_{12}^{-1} \chi_a(2)$$

$$= \sum_a \int dx_2 \sum_b U_{ba}^* \chi_b^*(2) r_{12}^{-1} \sum_c U_{ca} \chi_c(1)$$

$$= \sum_{bc} \left(\sum_a U_{ba}^* U_{ca} \right) \int dx_2 \chi_b^*(2) r_{12}^{-1} \chi_c(2)$$

$$= \sum_{bc} \delta_{bc} \int dx_2 \chi_b^*(2) r_{12}^{-1} \chi_c(2) = \sum_b \hat{J}_b(1) \tag{3.65}$$

where I have used the fact that $\sum_a U_{ba}^* U_{ca} = \delta_{bc}$. I will now deal with the transformed exchange operator $\hat{K}'_a$

$$\sum_a \hat{K}'_a(1) \chi'_d(1) = \sum_a \int dx_2 \chi_a^*(2) r_{12}^{-1} \chi'_d(2) \chi'_a(1)$$

$$= \sum_a \int dx_2 \sum_b U_{ba}^* \chi_b^*(2) r_{12}^{-1} \chi_d(2)' \sum_c U_{ca} \chi_c(1)$$

$$= \sum_{bc} \left(\sum_a U_{ba}^* U_{ca} \right) \int dx_2 \chi_b^*(2) r_{12}^{-1} \chi_d(2)' \chi'_c(1)$$

$$= \sum_{bc} \delta_{bc} \int dx_2 \chi_b^*(2) r_{12}^{-1} \chi_d(2)' \chi_c(1) = \sum_b \hat{K}_b(1) \chi'_d(1) \tag{3.66}$$

Equation 3.65 and equation 3.66 implies that $f'(1) = f(1)$. The next term I paid attention to is the Lagrange multiplier ε_{ba} where I multiplied equation 3.64 by $\langle \chi_c |$ and realised that ε_{ba} are matrix elements of the Fock operator.

$$\langle \chi_c | f | \chi_a \rangle = \sum_b \varepsilon_{ba} \langle \chi_c | \chi_b \rangle = \varepsilon_{ca} \tag{3.67}$$

So, the effect of unitary transformation on the ε_{ba}

$$\varepsilon'_{ba} = \int dx_1 \chi_a^*(1) f(1) \chi'_b(1)$$

$$= \sum_{cd} U_{ca}^* U_{db} \int dx_1 \chi_c^*(1) f(1) \chi'_d(1)$$

$$= \sum_{cd} U_{ca}^* \varepsilon_{cd} U_{db} \tag{3.68}$$

$$\boldsymbol{\varepsilon}' = \mathbf{U}^\dagger \boldsymbol{\varepsilon} \mathbf{U} \qquad in \; matrix \; form \tag{3.69}$$

From equation 3.69, I can choose the unique matrix U which diagonalises ε and the unique set of spin orbitals χ'_a are the so called canonical orbitals. I can then rewrite the

Hartree-Fock equations in the canonical form as

$$\hat{f}|\chi_a'\rangle = \varepsilon_a'|\chi_a'\rangle \tag{3.70}$$

or simply

$$\hat{f}|\chi_a\rangle = \varepsilon_a|\chi_a\rangle \tag{3.71}$$

3.2.1 The Roothan equations

Let us start by expanding the unknown basis functions ψ in terms of linear combination of a set of K known basis function $\phi_\mu(r)$ where $\mu = 1, 2, \ldots, K$.

$$\psi_i = \sum_{\mu=1}^{K} C_{\mu i}\phi_\mu \qquad i = 1, 2, \ldots, K \tag{3.72}$$

The Fock operator defined in equation 3.60,

$$\hat{f}(1) = \hat{h}(1) + \sum_{b=1}^{N}(J_b(1) - K_b(1)) \tag{3.73}$$

can be written in a matrix form in the basis $[\phi_\mu]$

$$F_{\mu v} = \int dr_1 \phi_\mu^*(1)\hat{f}(1)\phi_v(1) \tag{3.74a}$$

$$= \int dr_1 \phi_\mu^*(1)\hat{h}(1)\phi_v(1) + \sum_{a}^{N/2} \int \phi_\mu^*(1)[2J_a(1) - K_a(1)]\phi_v(1) \tag{3.74b}$$

$$= H_{\mu v}^{core} + \sum_{a}^{N/2} 2\langle \mu v|aa\rangle - \langle \mu a|av\rangle \tag{3.74c}$$

Here, one-electron part of the Fock matrix (core Hamiltonian matrix) is fixed for a given basis set and it is defined as

$$H_{\mu v}^{core} = \int dr_1 \phi_\mu^*(1)\hat{h}(1)\phi_v(1) \tag{3.75}$$

and the two electron integrals as

$$\langle \mu v|aa\rangle = \int dr_1 dr_2 \phi_\mu^*(1)\phi_v(1)r_{12}^{-1}\phi_a^*(2)\phi_a(2) \tag{3.76}$$

$$\langle \mu a|av\rangle = \int dr_1 dr_2 \phi_\mu^*(1)\phi_a(1)r_{12}^{-1}\phi_a^*(2)\phi_v(2) \tag{3.77}$$

I can use my definition of ψ in equation 3.72 to simplify the two-electron part $G_{\mu v}$ of Fock matrix

$$
\begin{aligned}
F_{\mu v} &= H_{\mu v}^{core} + \sum_a^{N/2} \sum_{\lambda \sigma} C_{\lambda a} C_{\sigma a}^* [2\langle \mu v | \sigma \lambda \rangle - \langle \mu \lambda | \sigma v \rangle] \\
&= H_{\mu v}^{core} + \sum_{\lambda \sigma} P_{\lambda \sigma} [\langle \mu v | \sigma \lambda \rangle - \frac{1}{2} \langle \mu \lambda | \sigma v \rangle] \\
&= H_{\mu v}^{core} + G_{\mu v}
\end{aligned}
\tag{3.78}
$$

It is clear that the matrix G depends on the two-electron integrals and it also depends on the density matrix P, respectively defined below;

$$
\langle \mu v | \lambda \sigma \rangle = \int dr_1 dr_2 \psi_\mu^*(1) \psi_v(1) r_{12}^{-1} \psi_\lambda^*(2) \psi_\sigma(2)
\tag{3.79}
$$

$$
P_{\mu v} = 2 \sum_a^{N/2} C_{\mu a} C_{va}^*
\tag{3.80}
$$

and the Fock matrix $\mathbf{F}$ depends on the density matrix $\mathbf{P}$, more specific, the coefficients $\mathbf{c}$, that is $\mathbf{F} = \mathbf{F}(\mathbf{P}) = \mathbf{F}(\mathbf{C})$. I can use the form of eigenvalue equation and write the Roothaan equations $\mathbf{F}(\mathbf{C})\mathbf{C} = \mathbf{SC}\varepsilon$ which can be solved iteratively because of their non-linearity. The eigenvector $\mathbf{C}$ and the eigenvalues ε are obtained from diagonalization of the Fock matrix $\mathbf{F}$. I can now write,

$$
\mathbf{FC} = \mathbf{SC}\varepsilon
\tag{3.81}
$$

and the transformed Roothaan equations[93],

$$
\mathbf{F'C'} = \mathbf{C'}\varepsilon
\tag{3.82}
$$

where the transformed matrix $\mathbf{X}$ is defined such that it satisfies the relation $\mathbf{X}^\dagger \mathbf{FX} = \mathbf{F'}$, and the coefficients $\mathbf{C'}$ were related to the old $\mathbf{C}$ through $\mathbf{C'} = \mathbf{X}^{-1}\mathbf{C}$ *or* $\mathbf{C} = \mathbf{XC'}$.

3.2.2 Self-Consistent Field Method (SCF)

The computational procedure for solving the Hartree-Fock equations is called the self-consistent field, and follows;

1. Specify set of molecular coordinates

2. Calculate all required molecular integrals, $S_{\mu v}, H_{\mu v}^{core}, (\mu v | \lambda \sigma)$

3. Calculate transformation matrix $\mathbf{X}$ by diagonalization of the overlap matrix $\mathbf{S}$

4. Generate the initial guess at the density matrix $\mathbf{P}$

5. Use density matrix $\mathbf{P}$ and $(\mu v|\lambda\sigma)$ to calculate the matrix $\mathbf{G}$ of equation 3.78

6. Combine $\mathbf{G}$ and $\mathbf{H}^{core}$ to get Fock matrix $\mathbf{F} = \mathbf{H}^{core} + \mathbf{G}$

7. Transform the Fock matrix, $\mathbf{F'} = \mathbf{X}^{\dagger}\mathbf{F}\mathbf{X}$

8. Diagonalise $\mathbf{F'}$ to obtain $\mathbf{C'}$ and ε.

9. Calculate $\mathbf{C} = \mathbf{X}\mathbf{C'}$

10. Use C to form a new matrix P using equation 3.80

11. Evaluate if the difference between the new density matrix $\mathbf{P}$ in step 10 and the matrix in step 5 falls below predetermined threshold. If not, go to step 5 and repeat the procedure with the new density matrix from step 10

12. If converged, calculate the expectation values and other quantities of interest from the converged solution.

3.2.3 Electron Correlation Energy

The Fock operator described by equations 3.60/3.73 is used to find one-electron molecular orbitals. The Hartree-Fock wave function is limited to a single Slater determinant. Lower electronic energies can be obtained by constructing an approximate wave function as a linear combination of Slater determinants corresponding to different electronic configuration;

$$\Psi = c_0\Psi_{HF} + c_1\Psi_1 + c_2\Psi_2 + \cdots \tag{3.83}$$

The weights c ensures normalisation. The c_0 coefficient is typically larger than any other coefficients in the combination, thus the HF wave function dominates in the linear combination described above. In the Hartree-Fock picture, the correlated motion of each electron with every other electron is ignored, thus the dynamical character of the electron-electron interaction is missing. A multi-determinantal wave function allows to capture the correlated motion of the electrons. In short, for any given electronic state, the correlation energy E_{corr} is the difference between the exact non-relativistic energy E_{exact} and the restricted Hartree-Fock energy E_{HF} obtained in a complete basis set.

$$E_{corr} = E_{exact} - E_{HF} \tag{3.84}$$

The E_{corr} is always negative. The correlated methods based on multi determinantal wave function give better predictions than the Hartree-Fock method.

3.3 Perturbation theory

This is a mathematical method suitable for systems with Hamiltonian $\hat{H}$ very close to a model Hamiltonian $\hat{H}_0$, for which solutions of the Schrödinger equation are known. The Hamiltonian operator $\hat{H}$ can then be split into two parts, the unperturbed Hamiltonian $\hat{H}_0$ and perturbation operator $\hat{W}$ which is very small compared to $\hat{H}_0$.

$$\hat{H} = \hat{H}_0 + \hat{W} \tag{3.85}$$

The perturbation $\hat{W}$ is encountered in systems subjected to weak electric or magnetic fields or in theory of electron correlation and comes about its effects on the energy spectrum. The most explicit ways of writing the perturbation $\hat{W}$ is through the use of real parameter λ such that $0 \leq \lambda \leq 1$ and therefore $\hat{W} = \lambda \hat{H}_p$. The parameter λ controls the magnitude of perturbation and then the total Hamiltonian would be written as;

$$\hat{H} = \hat{H}_0 + \lambda \hat{H}_p \tag{3.86}$$

Thus the eigenvalue problem becomes

$$\hat{H}|\psi_n\rangle = E_n|\psi_n\rangle \tag{3.87a}$$

$$(\hat{H}_0 + \lambda \hat{H}_p)|\psi_n\rangle = E_n|\psi_n\rangle \tag{3.87b}$$

Non-degenerate Perturbation Theory. The following scheme applies to situations in which $E_n^{(0)}$ is non-degenerate.

$$\hat{H}_0|\phi_n\rangle = E_n^{(0)}|\phi_n\rangle \tag{3.88}$$

The main idea behind perturbation theory is that the perturbed eigenvalues and eigenstates can both be expanded in power series in the parameter λ;

$$E_n = E_n^{(0)} + \lambda E_n^{(1)} + \lambda E_n^{(2)} + \cdots \tag{3.89}$$

$$|\psi_n\rangle = |\phi_n\rangle + \lambda|\psi_n^{(1)}\rangle + \lambda^2|\psi_n^{(2)}\rangle + \cdots \tag{3.90}$$

The parameters $E_n^{(k)}$ and the kets $|\psi_n^{(k)}\rangle$, respectively, represent the kth-order correction to the eigen-energies and eigenvectors. It is worth noting that only one or two terms in these expansions are kept as the series might be divergent[91].

If λ is set to zero, then the expressions 3.89 and 3.90 when substituted into equation 3.87 yield the unperturbed solutions $E_n = E_n^{(0)}$ and $|\psi_n\rangle = |\phi_n\rangle$, equation 3.88. I can determine the values of $E_n^{(1)}$, $E_n^{(2)}$, $E_n^{(3)}$ and $|\psi_n^{(1)}\rangle$ by substituting equations 3.89 and

3.90 into equation 3.87.

$$\left(\hat{H}_0 + \lambda\hat{H}_p\right)\left(|\phi_n\rangle + \lambda|\psi_n^{(1)}\rangle + \lambda^2|\psi_n^{(2)}\rangle + \cdots\right) =$$
$$\left(E_n^{(0)} + \lambda E_n^{(1)} + \lambda^2 E_n^{(2)} + \cdots\right)\left(|\phi_n\rangle + \lambda|\psi_n^{(1)}\rangle + \lambda^2|\psi_n^{(2)}\rangle + \cdots\right) \quad (3.91)$$

Considering the fact the the coefficient of successive powers of λ on both sides of equation 3.91 are equal, then I can deduce these equations;

- Zero order in λ

$$\hat{H}_0|\phi_n\rangle = E_n^{(0)}|\phi_n\rangle \quad (3.92)$$

- First order in λ

$$\hat{H}_0|\psi_n^{(1)}\rangle + \hat{H}_p|\phi_n\rangle = E_n^{(0)}|\psi_n^{(1)}\rangle + E_n^{(1)}|\phi_n\rangle \quad (3.93)$$

- Second order in λ

$$\hat{H}_0|\psi_n^{(2)}\rangle + \hat{H}_p|\psi_n^{(1)}\rangle = E_n^{(0)}|\psi_n^{(2)}\rangle + E_n^{(1)}|\psi_n^{(1)}\rangle + E_n^{(2)}|\phi_n\rangle \quad (3.94)$$

- "i" order in λ and $i > 0$

$$\hat{H}_0|\psi_n^i\rangle + \hat{H}_p|\psi_n^{i-1}\rangle = \sum_{j=0}^{i} E^{(j)}\psi_n^{(i-j)} \quad (3.95)$$

Here, my goal is to find the expression for the perturbation corrections $E_n^{(k)}$. At this stage, it is often convenient to use the intermediate normalisation condition obtained by multiplying equation 3.90 by $\langle\phi_n|$;

$$\langle\phi_n|\psi_n\rangle = \langle\phi_n|\phi_n + \lambda\psi_n^{(1)} + \lambda^2\psi_n^{(2)} + \cdots\rangle \quad (3.96)$$
$$= \underbrace{\langle\phi_n|\phi_n\rangle}_{=1} + \lambda\underbrace{\langle\phi_n|\psi_n^{(1)}\rangle}_{=0} + \lambda^2\underbrace{\langle\phi_n|\psi_n^{(2)}\rangle}_{=0} + \cdots \quad (3.97)$$
$$\langle\phi_n|\psi_n\rangle = 1 \quad (3.98)$$

where $\langle\phi_n| = \langle\psi_n^{(0)}|$ and the normalisation condition is expressed as;

$$\langle\psi_n^{(0)}|\psi_n^{(k)}\rangle = \delta_{0k} \quad (3.99)$$

I can multiply the equation 3.93 by $\langle\phi_n|$;

$$\langle\phi_n|\hat{H}_0|\psi_n^{(1)}\rangle + \langle\phi_n|\hat{H}_p|\phi_n\rangle = \langle\phi_n|E_n^{(0)}|\psi_n^{(1)}\rangle + \langle\phi_n|E_n^{(1)}|\phi_n\rangle \quad (3.100)$$

and use the facts that $\langle\phi_n|\hat{H}_0|\psi_n^{(1)}\rangle$ and $\langle\phi_n|\psi_n^{(1)}\rangle$ are both equal to zero and $\langle\phi_n|\phi_n\rangle = 1$ and obtain

$$E_n^{(1)} = \langle\phi_n|\hat{H}_p|\phi_n\rangle \tag{3.101}$$

The set of unperturbed states $|\phi_n\rangle$ form a complete and orthonormal basis ($\langle\phi_n|\phi_k\rangle = \delta_{nk}$), therefore I can expand $|\psi_n^{(1)}\rangle$ in the $|\phi_n\rangle$ basis such that

$$\begin{aligned}
|\psi_n^{(1)}\rangle &= \left(\sum_m |\phi_m\rangle\langle\phi_m|\right)|\psi_n^{(1)}\rangle \\
&= \sum_{m\neq n} \langle\phi_m|\psi_n^{(1)}\rangle|\phi_m\rangle \\
&= \sum_{m\neq n} b_m \phi_m
\end{aligned} \tag{3.102}$$

Note that $\langle\phi_n|\psi_n^{(1)}\rangle = 0$ and the coefficient $b_m\left(= \langle\phi_m|\psi_n^{(1)}\rangle\right)$ can be obtained by multiplying equation 3.93 by $\langle\phi_m|$ on both sides

$$\langle\phi_m|\left(\hat{H}_0|\psi_n^{(1)}\rangle + \hat{H}_p|\phi_n\rangle\right) = \langle\phi_m|\left(E_n^{(0)}|\psi_n^{(1)}\rangle + E_n^{(1)}|\phi_n\right) \tag{3.103a}$$

$$\langle\phi_m|\hat{H}_0|\psi_n^{(1)}\rangle + \langle\phi_m|\hat{H}_p|\phi_n\rangle = E_n^{(0)}\langle\phi_m|\psi_n^{(1)}\rangle + E_n^{(1)}\langle\phi_m|\phi_n\rangle \tag{3.103b}$$

$$E_m^{(0)}\langle\phi_m|\psi_n^{(1)}\rangle + \langle\phi_m|\hat{H}_p|\phi_n\rangle = E_n^{(0)}\langle\phi_m|\psi_n^{(1)}\rangle + 0 \tag{3.103c}$$

Thus,

$$b_m = \langle\phi_m|\psi_n^{(1)}\rangle = \frac{\langle\phi_m|\hat{H}_p|\phi_n\rangle}{E_n^{(0)} - E_m^{(0)}} \tag{3.104}$$

I can then substitute the above equation into equation 3.102 to get

$$|\psi_n^{(1)}\rangle = \sum_{m\neq n} \frac{\langle\phi_m|\hat{H}_p|\phi_n\rangle}{E_n^{(0)} - E_m^{(0)}}|\phi_m\rangle \tag{3.105}$$

and I can then determine the eigenfunction $\psi_n^{(1)}$ of $\hat{H}$ to first order in $\lambda\hat{H}_p$ through substitution of the above equation into equation 3.90

$$|\psi_n\rangle \approx |\phi_n\rangle + \sum_{m\neq n} \frac{\langle\phi_m|\hat{W}|\phi_n\rangle}{E_n^{(0)} - E_m^{(0)}}|\phi_m\rangle \tag{3.106}$$

The second-order corrections to the total energy can be determined from multiplication of equation 3.94 by $\langle\phi_n|$ on both sides and then use the facts that $\langle\phi_n|\psi_n^{(1)}\rangle = \langle\phi_n|\psi_n^{(2)}\rangle = 0$ and $\langle\phi_n|\phi_n\rangle = 1$ to obtain

$$E_n^{(2)} = \langle\phi_n|\hat{H}_p|\psi_n^{(1)}\rangle \tag{3.107}$$

through which equation 3.105 is inserted on the above equation to get the second-order correction to the energy

$$E_n^{(2)} = \sum_{m \neq n} \frac{|\langle \phi_m | \hat{H}_p | \phi_n \rangle|^2}{E_n^{(0)} - E_m^{(0)}} \tag{3.108}$$

The total energy of the system up to the second-order correction can be obtained by inserting equations 3.108 and 3.101 into the equation 3.89:

$$E_n = E_n^{(0)} + \langle \phi_n | \hat{H}_p | \phi_n \rangle + \sum_{m \neq n} \frac{|\langle \phi_m | \hat{H}_p | \phi_n \rangle|^2}{E_n^{(0)} - E_m^{(0)}} + \cdots \tag{3.109}$$

Calculations up to the second-order correction are typically satisfactory. In principle, I also expect

$$\left| \frac{\langle \phi_m | \hat{H}_p | \phi_n \rangle}{E_n^{(0)} - E_m^{(0)}} \right| \lll 1 \qquad n \neq m \tag{3.110}$$

The degenerate perturbation theory is discussed next.

Degenerate Perturbation Theory. In this section, I will discuss perturbation theory for degenerate energy level of $\hat{H}_0$. The eigenvalue problem is written as;

$$\hat{H} | \psi_n \rangle = E_n | \psi_n \rangle \tag{3.111a}$$

$$\left(\hat{H}_0 + \hat{H}_p \right) | \psi_n \rangle = E_n | \psi_n \rangle \tag{3.111b}$$

Considering the *f*-fold degenerate system described by different eigenstates $|\phi_{n_\alpha}\rangle$ corresponding to the same eigenvalues $E_n^{(0)}$, I write:

$$\hat{H}_0 | \phi_{n_\alpha} \rangle = E_n^{(0)} | \phi_{n_\alpha} \rangle \qquad \alpha = 1, 2, 3, \cdots, f \tag{3.112}$$

Suppose the states $|\phi_{n_\alpha}\rangle$ are orthonormal with respect to quantum number α, the eigenfunction $|\psi_n\rangle$ can be expanded as a linear combination in the $|\phi_{n_\alpha}\rangle$ basis, in the zeroth-order approximation:

$$|\psi_n\rangle = \sum_{\alpha=1}^{f} a_\alpha |\phi_{n_\alpha}\rangle \tag{3.113}$$

and the coefficients a_α should satisfy the relation:

$$\langle \psi_n^{(0)} | \psi_n^{(0)} \rangle = \sum_{\alpha,\beta} a_\alpha^* a_\beta \delta_{\alpha,\beta} = \sum_{\alpha=1}^{f} |a_\alpha|^2 = 1 \tag{3.114}$$

through the orthonormality condition $\langle \phi_{n_\alpha} | \phi_{n_\beta} \rangle = \delta_{\alpha,\beta}$. To get these coefficients and the first-order corrections to the energy, equations 3.112 and 3.113 are substituted into equation 3.111 and then the resulting equation is multiplied by $\langle \phi_{n_\beta} |$ on both sides and

use the fact that $\langle \phi_{n,\beta} | \phi_{n,\alpha} \rangle = \delta_{\beta,\alpha}$ to get the following equations:

$$\langle \phi_{n_\beta} | \left(\sum_\alpha \Big[E_n^{(0)} | \phi_{n_\alpha} \rangle + \hat{H}_p | \phi_{n_\alpha} \rangle \Big] a_\alpha \right) = \langle \phi_{n_\beta} | \left(E_n \sum_\alpha a_\alpha | \phi_{n_\alpha} \rangle \right) \tag{3.115}$$

$$\sum_\alpha a_\alpha \Big[E_n^{(0)} \delta_{\alpha,\beta} + \langle \phi_{n_\beta} | \hat{H}_p | \phi_{n_\alpha} \rangle \Big] = E_n \sum_\alpha a_\alpha \delta_{\alpha,\beta} \tag{3.116}$$

or

$$a_\beta E_n = a_\beta E_n^{(0)} + \sum_{\alpha=1}^{f} a_\alpha \langle \phi_{n\beta} | \hat{H}_p | \phi_{n_\alpha} \rangle \tag{3.117}$$

which can be rewritten as a system of $f-$homogeneous linear equations in $E_n^{(1)}$ for coefficients a_α, thus having f different real roots:

$$\sum_{\alpha=1}^{f} \Big(\hat{H}_{p\beta\alpha} - E_n^{(1)} \delta_{\alpha,\beta} \Big) a_\alpha = 0 \qquad \beta = 1, 2, 3, \cdots, f \tag{3.118}$$

where $H_{p\beta\alpha} = \langle \phi_{n\beta} | \hat{H}_p | \phi_{n_\alpha} \rangle$ and $E_n^{(1)} = E_n - E_n^{(0)}$. The energies $E_n^{(1)}$ are the 1st order correction to the eigenvalue, $E_n^{(0)}$. The coefficients a_α are non-vanishing only when the determinant $|\hat{H}_{p\alpha\beta} - E_n^{(1)} \delta_{\alpha,\beta}|$ is zero:

$$\begin{bmatrix} \hat{H}_{p11} - E_n^{(1)} & \hat{H}_{p12} & \hat{H}_{p13} & \cdots & \hat{H}_{p1f} \\ \hat{H}_{p21} & \hat{H}_{p22} - E_n^{(1)} & \hat{H}_{p23} & \cdots & \hat{H}_{p2f} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{H}_{pf1} & \hat{H}_{pf2} & \hat{H}_{pf3} & \cdots & \hat{H}_{pff} - E_n^{(1)} \end{bmatrix} = 0 \tag{3.119}$$

In general, the roots $E_{n_\alpha}^{(1)}$ are different therefore the eigenvalues $\hat{H}$ are not degenerate, hence the f-fold degenerate level of the unperturbed problem $E_n^{(0)}$ is split into f different levels E_{n_α}

$$E_{n_\alpha} = E_n^{(0)} + E_{n_\alpha}^{(1)} \qquad \alpha = 1, 2, 3, \cdots, f \tag{3.120}$$

I can find the eigenfunctions $|\psi_n\rangle$, of $\hat{H}$ in the zeroth-order approximation from equation 3.113 but firstly, I need to substitute these roots into equation 3.118 and then solve the resulting equation for the coefficients α.

3.4 Moller-Plesset Perturbation Theory

Here I present a theory that improves the Hartree-Fock energy by adding the missing part, the correlation energy, which can be obtained through partitioning of the Hamiltonian and the perturbation procedure discussed in the previous sections. I can write

the partitioned Hamiltonian as;

$$\hat{H} = \hat{H}_0 + V \tag{3.121}$$

where $\hat{H}_0$ is the Hartree-Fock Hamiltonian which can be expressed in terms of Fock operators $f(i)$ (equal to one-electron operator $h(i)$ plus electron-electron effective potential $v^{HF}(i)$);

$$\hat{H}_0 = \sum_i f(i) \tag{3.122a}$$

$$= \sum_i [h(i) + v^{HF}(i)] \tag{3.122b}$$

and the potential energy operator V is defined by;

$$V = \sum_{r<s} r_{rs}^{-1} - V^{HF} \tag{3.123a}$$

$$= \sum_{r<s} r_{rs}^{-1} - \sum_i v^{HF}(i) \tag{3.123b}$$

The energy of the system is estimated by applying the above Hamiltonian on the Schrödinger equation;

$$
\begin{aligned}
E_{HF} &= \langle \Phi_{HF} | \hat{H}_0 + V | \Phi_{HF} \rangle \\
&= \langle \Phi_{HF} | \hat{H}_0 | \Phi_{HF} \rangle + \langle \Phi_{HF} | V | \Phi_{HF} \rangle \\
&= \langle \Phi_{HF} | \sum_i f(i) | \Phi_{HF} \rangle + \langle \Phi_{HF} | \sum_{i<j} r_{ij}^{-1} | \Phi_{HF} \rangle - \langle \Phi_{HF} | \sum_i v^{HF}(i) | \Phi_{HF} \rangle \\
&= \sum_a \varepsilon_a + \frac{1}{2} \sum_{ab} \langle ab \| ab \rangle - \sum_a \langle a | v^{HF} | a \rangle \\
&= \sum_a \varepsilon_a - \frac{1}{2} \sum_{ab} \langle ab \| ab \rangle
\end{aligned}
\tag{3.124}
$$

Note that the Hartree-Fock wave function Φ_{HF} is an eigenfunction of $\hat{H}_0$ therefore the eigenvalue $E_0^{(0)}$ is equal to the sum of the orbital energies $\sum_a \varepsilon_a$, that is,

$$E_0^{(0)} = \sum_a \varepsilon_a \tag{3.125}$$

and the total energy is the sum of the zeroth and first-order energies,

$$E_0 = E_0^{(0)} + E_0^{(1)} \tag{3.126a}$$

$$= \sum_a \varepsilon_a - \frac{1}{2} \sum_{ab} \langle ab \| ab \rangle \tag{3.126b}$$

The second-order correction to the Hartree-Fock energy has been derived in the previous section,

$$E_0^{(2)} = \sum_n \frac{|\langle 0|V|n\rangle|^2}{E_0^{(0)} - E_n^{(0)}} \tag{3.127}$$

where the summation is over the full set of excited determinants. Singly excited detrminant Ψ_i^a do not contribute to $E_0^{(2)}$ because

$$\langle \Phi_{HF}|V|\Psi_i^a\rangle = \langle \Phi_{HF}|\hat{H} - \hat{H}_0|\Psi_i^a\rangle \tag{3.128a}$$

$$= \langle \Phi_{HF}|\hat{H}|\Psi_i^a\rangle - \langle \Phi_{HF}|\hat{H}_0|\Phi_i^a\rangle \tag{3.128b}$$

$$= \langle \Phi_{HF}|\hat{H}|\Psi_i^a\rangle - \langle i|f|a\rangle \tag{3.128c}$$

$$= 0 \tag{3.128d}$$

The first term disappears according to the Brillouin's theorem[87] and the second term because $\langle i|f|a\rangle = \langle i|fa\rangle = \varepsilon_a\langle i|a\rangle = 0$. The triply (and higher) excited determinants also do not contribute to $E_0^{(2)}$ because of the two-particle nature of V. The effect of $\hat{H}_0$ operating on Ψ_{ij}^{ab} is

$$\hat{H}_0|\Psi_{ij}^{ab}\rangle = \left(E_0^{(0)} - (\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b)\right)|\Psi_{ij}^{ab}\rangle \tag{3.129a}$$

$$= E_n^{(0)}|\Psi_{ij}^{ab}\rangle \tag{3.129b}$$

This result will be used in the denominator of equation 3.127. I can rewrite the Moller-Plesset second-order correction to the energy as;

$$E_0^{(2)} = \sum_{i<j}^{occ} \sum_{a<b}^{virt} \frac{|\langle \Psi_0| \sum_{r<s} r_{rs}^{-1}|\Psi_{ij}^{ab}\rangle|^2}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \tag{3.130a}$$

$$= \sum_{i<j}^{occ} \sum_{a<b}^{virt} \frac{|\langle ij\|ab\rangle|^2}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \tag{3.130b}$$

The explicit expression for the total MP2 energy is given by equations 3.126 and 3.130;

$$E_{MP2} = E_0^{(0)} + E_0^{(1)} + E_0^{(2)} \tag{3.131a}$$

$$= E_{HF} + E_0^{(2)} \tag{3.131b}$$

The third and forth order corrections can also be made in a similar but more complicated way.

3.5 Coupled-Cluster Theory

The coupled-cluster method is the most successful method used to describe the many-electron molecular wave functions accurately. It starts from a single determinant reference function and recovers large fraction of the correlation energy and providing accurate correction to the Hartree-Fock description. This method can be applied to systems with few tens of atoms which are dominated by a single electronic configuration.

The coupled-cluster wave function. In the independent-particle model, the wave function for the motion of non-interacting fermions, which corresponds to a product of creation operators $a_i^\dagger$ working on the vacuum state $|vac\rangle$, is described by Slater determinant of occupied spin orbitals $i = 1, N$.

$$|\Phi_{HF}\rangle = \prod_{i=1}^{N} a_i^\dagger |vac\rangle \tag{3.132}$$

If two interacting electron are occupying the spin orbitals i and j, then their motion will be impeded by their instantaneous interactions. As a result, the electrons will get excited to different sets of virtual spin orbitals a and b and the process may be described as

$$a_i^\dagger a_j^\dagger \to a_i^\dagger a_j^\dagger + \sum_{a>b} t_{ij}^{ab} a_a^\dagger a_a^\dagger \tag{3.133}$$

where t_{ij}^{ab} are the so called amplitudes. The right hand side of the equation 3.133 is named a two electron cluster (a pair cluster). The "correlation process" of these electrons is described by operator:

$$\hat{\tau}_{ij}^{ab} = a_a^\dagger a_i a_b^\dagger a_j \tag{3.134}$$

with a_i's as the annihilating operators and one can assume that i and j are unoccupied in vacuum state $|vac\rangle$ and write the pair-cluster ij as;

$$\left[\prod_{a>b} \left(1 + t_{ij}^{ab} \hat{\tau}_{ij}^{ab} \right) \right] a_i^\dagger a_j^\dagger |vac\rangle = a_i^\dagger a_j^\dagger |vac\rangle + \sum_{a>b} t_{ij}^{ab} a_a^\dagger a_b^\dagger |vac\rangle \tag{3.135}$$

using the relation $t_{ij}^{ab} \hat{\tau}_{ij}^{ab} = 0$. If I allow each pair of electrons to interact and rewrite equation 3.135 for all pairs of occupied spin orbitals, then one can show that the expression for pair-correlated electronic state is:

$$|CCD\rangle = \left[\prod_{a>b, i>j} \left(1 + t_{ij}^{ab} \hat{\tau}_{ij}^{ab} \right) \right] |\Phi_{HF}\rangle \tag{3.136}$$

Note that all $\hat{\tau}_{ij}^{ab}$ commute with each other and the above expression is known as the

coupled-cluster doubles (CCD) wave function in which only double excitations are allowed. The pair clusters account for the majority of the contributions to the description of the correlated motion of interacting electrons. This is due to the fact that the correlation motion of electrons is important for electrons that are close to each other and at most two electron may coincide in space. The accurate description can be obtained by considering clusters of all sizes:

$$\hat{\tau}_{ijk\cdots}^{abc\cdots} = a_a^\dagger a_i a_b^\dagger a_j a_c^\dagger a_k \cdots = \hat{\tau}_\mu \tag{3.137}$$

and the generalised coupled-cluster wave function is expressed as:

$$|CC\rangle = \left[\prod_\mu \left(1 + t_\mu \hat{\tau}_\mu \right) \right] |\Phi_{HF}\rangle \tag{3.138}$$

The spin-orbital excitation operators satisfy the relation $\hat{\tau}_\mu^2 = 0$, thus the correlating operators may be written as exponentials of the excitation operators.

$$1 + t_\mu \hat{\tau}_\mu = e^{t_\mu \hat{\tau}_\mu} \tag{3.139}$$

thus, the coupled-cluster function 3.138 can be rewritten as;

$$|CC\rangle = e^{\hat{T}} |\Phi_{HF}\rangle \tag{3.140}$$

where excitation operator $\hat{T} = \sum_\mu t_\mu \hat{\tau}_\mu$ and the expression 3.140 is known as the **exponential ansatz**. I may rewrite the excitation operator as a linear combination of single, double, triples, ... ,excitations up to $N-$fold excitations for N electron system.

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots + \hat{T}_N \tag{3.141}$$

where

$$\hat{T}_1 = \sum_i^{occ} \sum_a^{virt} t_i^a a_a^\dagger a_i = \sum_{ai} t_i^a \tau_i^a \tag{3.142}$$

$$\hat{T}_2 = \sum_{i>j}^{occ} \sum_{a>b}^{virt} t_{ij}^{ab} a_a^\dagger a_i a_b^\dagger a_j = \frac{1}{4} \sum_{aibj} t_{ij}^{ab} \tau_{ij}^{ab} \tag{3.143}$$

$$\hat{T}_3 = \sum_{i>j>k}^{occ} \sum_{a>b>c}^{virt} t_{ijk}^{abc} a_a^\dagger a_i a_b^\dagger a_j a_c^\dagger a_k = \frac{1}{6} \sum_{aibjck} t_{ijk}^{abc} \tau_{ijk}^{abc} \tag{3.144}$$

etc. with the coefficients $t_i^a, t_{ij}^{ab}, t_{ijk}^{abc}, \ldots$ as the amplitudes. The exponential operator can be expressed as a Taylor's expansion

$$e^{\hat{T}} = \sum_{N=0}^{\infty} \frac{\hat{T}^N}{N!} \tag{3.145}$$

$$= 1 + \hat{T} + \frac{1}{2}\hat{T}^2 + \frac{1}{3!}\hat{T}^3 + \frac{1}{4!}\hat{T}^4 + \cdots \tag{3.146}$$

$$= 1 + (\hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots) + (\hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots)^2 + (\hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots)^3 + \cdots \tag{3.147}$$

therefore, I can gather similar orders of excitation and rewrite the equation as;

$$e^{\hat{T}} = 1 + \hat{T}_1 + \left(\hat{T}_2 + \frac{1}{2}\hat{T}_1^2\right) + \left(\hat{T}_3 + \hat{T}_1\hat{T}_2 + \frac{1}{6}\hat{T}_1^3\right) +$$
$$\left(\hat{T}_4 + \frac{1}{2}\hat{T}_2^2 + \frac{1}{2}\hat{T}_1^2\hat{T}_2 + \hat{T}_1\hat{T}_3 + \frac{1}{24}\hat{T}_1^4\right) + \cdots \tag{3.148}$$

Note that there is only one way to generate single excited determinant, $\hat{T}_1$, but two ways to generate double excitations. Either through generation of pure double excited determinant, $\hat{T}_2$ or generation of two successive single excitations, $\hat{T}_1\hat{T}_1 = \hat{T}_1^2$. The $\hat{T}_2$ is the connected cluster whereas the two successive single excitations $\hat{T}_1\hat{T}_1 = \hat{T}_1^2$ are the disconnected clusters. Similarly, there are three triply excited determinants that can be generated. One may be through single pure triple, $\hat{T}_3$, or through single and double, $\hat{T}_1\hat{T}_2 = \hat{T}_3$, or through generation of three singles, $\hat{T}_1\hat{T}_1\hat{T}_1 = \hat{T}_1^3$. The pure triple $\hat{T}_3$ is the connected cluster whereas $\hat{T}_1\hat{T}_2 = \hat{T}_3$ and $\hat{T}_1\hat{T}_1\hat{T}_1 = \hat{T}_1^3$ are the disconnected clusters. I can relate the coupled-cluster method to the Configuration Interaction (CI) method and write the lowest-order configuration operators for each classes of excitation as;

$$\hat{C}_0 = 1 \qquad \textit{zero excitations} \tag{3.149a}$$

$$\hat{C}_1 = \hat{T}_1 \qquad \textit{single excitations} \tag{3.149b}$$

$$\hat{C}_2 = \hat{T}_2 + \frac{1}{2}\hat{T}_1^2 \qquad \textit{double excitations} \tag{3.149c}$$

$$\hat{C}_3 = \hat{T}_3 + \hat{T}_1\hat{T}_2 + \frac{1}{6}\hat{T}_1^3 \qquad \textit{triple excitations} \tag{3.149d}$$

$$\hat{C}_4 = \hat{T}_4 + \hat{T}_1\hat{T}_3 + \frac{1}{2}\hat{T}_2^2 + \frac{1}{2}\hat{T}_1^2\hat{T}_2 + \frac{1}{24}\hat{T}_1^4 \qquad \textit{four fold excitations} \tag{3.149e}$$

and so on where the operators $\hat{C}_i$ pertain to the CI method. The generalised coupled-cluster exponential operator can now be written in terms of CI operators as;

$$e^{\hat{T}} = \hat{C}_0 + \hat{C}_1 + \hat{C}_2 + \hat{C}_3 + \hat{C}_4 + \cdots \tag{3.150}$$

The above exact coupled-cluster wave function satisfies the Schrödinger equation

$$\hat{H}\exp(\hat{T})|\Phi_{HF}\rangle = E\exp(\hat{T})|\Phi_{HF}\rangle \tag{3.151}$$

for a given orbital basis but the truncated wave function cannot satisfy this equation exactly. Firstly, the coupled-cluster wave function is being optimised and the optimised wave function that satisfies the above equation is projected onto the Hartree-Fock state and onto the excited manifold $\langle\mu| = \langle\Phi_{HF}|\tau_\mu^\dagger$. I then get the following set of coupled-cluster equations, the energy and the amplitudes, respectively;[88]

$$\langle\Phi_{HF}|\hat{H}\exp(\hat{T})|\Phi_{HF}\rangle = E \tag{3.152}$$

$$\langle\mu|\hat{H}\exp(\hat{T}|\Phi_{HF}\rangle = E\langle\mu|\exp(\hat{T})|\Phi_{HF}\rangle \tag{3.153}$$

The most convenient way of expressing the projected coupled-cluster equations is to multiply the above Schrödinger equation 3.151 by the operator $\exp(-T)$ from the left

$$\exp(-\hat{T})\hat{H}\exp(\hat{T})|\Phi_{HF}\rangle = E|\Phi_{HF}\rangle \tag{3.154}$$

and then project the resulting equation 3.154 against the same determinants used in above equations 3.152/3.153 to get the following equations for coupled-cluster energy and amplitudes, respectively.

$$\langle\Phi_{HF}|\exp(-\hat{T})\hat{H}\exp(\hat{T})|\Phi_{HF}\rangle = E \tag{3.155}$$

$$\langle\mu|\exp(-\hat{T})\hat{H}\exp(\hat{T})|\Phi_{HF}\rangle = 0 \tag{3.156}$$

The non-Hermitian *similarity-transformed Hamiltonian*[88] $\hat{H}^T = \exp(-\hat{T}\hat{H}\exp(\hat{T})$ operator (from equation 3.154) can be expanded in terms of the commutators and yield no higher than the quartic in the amplitudes

$$\exp(-\hat{T})\hat{H}\exp(\hat{T}) = \hat{H} + [\hat{H},\hat{T}] + \frac{1}{2!}[[\hat{H},\hat{T}],\hat{T}] + \frac{1}{3!}[[[\hat{H},\hat{T}],\hat{T}],\hat{T}]$$
$$+ \frac{1}{4!}[[[[\hat{H},\hat{T}],\hat{T}],\hat{T}],\hat{T}] \tag{3.157}$$

and the resulting projected amplitudes from equation 3.156 will yield at most quartic equations in the cluster amplitudes. This termination of the expansion is due to the special structure of the singles (equation 3.142) and doubles (equation 3.143) which are linear combinations of commuting excitation operators. The expressions for the Coupled-Cluster Singles and Double (CCSD) amplitudes equations[88](pp. 661) are as

follows;

$$\langle\mu_1|\hat{H}|\Phi_{HF}\rangle + \langle\mu_1|[\hat{H},\hat{T}_1]|\Phi_{HF}\rangle + \langle\mu_1|[\hat{H},\hat{T}_2]|\Phi_{HF}\rangle + \frac{1}{2}\langle\mu_1|[[\hat{H},\hat{T}_1],\hat{T}_1]|\Phi_{HF}\rangle$$
$$+ \langle\mu_1|[[\hat{H},\hat{T}_1],\hat{T}_2]|\Phi_{HF}\rangle + \frac{1}{6}\langle\mu_1|[[[\hat{H},\hat{T}_1],\hat{T}_1],\hat{T}_1]|\Phi_{HF}\rangle = 0 \quad (3.158)$$

$$\langle\mu_2|\hat{H}|\Phi_{HF}\rangle + \langle\mu_2|[\hat{H},\hat{T}_1]|\Phi_{HF}\rangle + \langle\mu_2|[\hat{H},\hat{T}_2]|\Phi_{HF}\rangle$$
$$+ \frac{1}{2}\langle\mu_2|[[\hat{H},\hat{T}_1],\hat{T}_1]|\Phi_{HF}\rangle + \langle\mu_2|[[\hat{H},\hat{T}_1],\hat{T}_2]|\Phi_{HF}\rangle$$
$$+ \frac{1}{2}\langle\mu_2|[[\hat{H},\hat{T}_2],\hat{T}_2]|\Phi_{HF}\rangle + \frac{1}{6}\langle\mu_2|[[[\hat{H},\hat{T}_1],\hat{T}_1],\hat{T}_1]|\Phi_{HF}\rangle$$
$$+ \frac{1}{2}\langle\mu_2|[[[\hat{H},\hat{T}_1],\hat{T}_1],\hat{T}_2]|\Phi_{HF}\rangle + \frac{1}{24}\langle\mu_2|[[[[\hat{H},\hat{T}_1],\hat{T}_1],\hat{T}_1],\hat{T}_1]|\Phi_{HF}\rangle = 0 \quad (3.159)$$

The Coupled-Cluster Energies. The coupled-cluster energy is obtained from the Schrödinger equation. Here I start by subtracting the Hartree-Fock energy from the Schrödinger equation (on both sides) and the equation reads;

$$(\hat{H} - E_{HF})\Psi_0 = (E - E_{HF})\Psi_0 \quad (3.160)$$

where my intermediate normalised full CI wave function looks like

$$|\Psi_0\rangle = |\Phi_0\rangle + \sum_{\substack{a<b \\ r<s}} C_{ab}^{rs}|\Phi_{ab}^{rs}\rangle + \sum_{\substack{a<b<c<d \\ r<s<t<u}} C_{abcd}^{rstu}|\Phi_{abcd}^{rstu}\rangle + \cdots \quad (3.161)$$

I can now define the new Hamiltonian $\hat{H}_N = \hat{H} - E_{HF}$ and the correlation energy as $\Delta E = E - E_{HF}$ and then the new equation becomes;

$$\hat{H}_N\Psi_0 = \Delta E\Psi_0 \quad (3.162)$$

If I substitute the coupled-cluster ansatz for the wave-function, multiply the equation with Hartree-Fock determinant on both sides and integrate the equation, I simply get;

$$\langle\Phi_{HF}|\hat{H}_N|(\hat{C}_0 + \hat{C}_1 + \hat{C}_2 + \hat{C}_3 + \hat{C}_4 + \cdots)\Phi_{HF}\rangle =$$
$$\Delta E\langle\Phi_{HF}|(\hat{C}_0 + \hat{C}_1 + \hat{C}_2 + \hat{C}_3 + \hat{C}_4 + \cdots)\Phi_{HF}\rangle \quad (3.163)$$

The Brillouins theorem requires $\langle \Phi_{HF}|\hat{H}_N|\hat{C}_1\Phi_{HF}\rangle = 0$ and also taking into account the above definitions of $\hat{H}_N$, and E_{HF}, I can rewrite the equation as;

$$\Delta E = \langle \Phi_{HF}|\hat{H}_N|\hat{C}_2\Phi_{HF}\rangle$$
$$= \langle \Phi_{HF}|\hat{H}_N|(\hat{T}_2 + \frac{1}{2}\hat{T}_1^2)\Phi_{HF}\rangle \tag{3.164}$$

I can expand T_1 and T_2 in equation 3.164 using equations 3.142 and 3.143;

$$\hat{T}_2\Phi_{HF} = \sum_{i>j}^{occ}\sum_{a>b}^{virt} t_{ij}^{ab}\Phi_{ij}^{ab} \tag{3.165}$$

$$\hat{T}_1^2\Phi_{HF} = \hat{T}_1\sum_{i}^{occ}\sum_{a}^{virt} t_i^b\Phi_i^a = \sum_{i}^{occ}\sum_{a}^{virt} t_i^a\hat{T}_1\Phi_i^a = \sum_{i}^{occ}\sum_{a}^{virt} t_i^a\sum_{j}^{occ}\sum_{b}^{virt} t_i^b\Phi_{ij}^{ab} \tag{3.166}$$

I then eliminate the unrestricted summations on equation 3.166 and use the symmetry of the doubly excited determinant, $\Phi_{ij}^{ab} = -\Phi_{ji}^{ab} = -\Phi_{ij}^{ba} = \Phi_{ji}^{ba}$ and obtain;

$$\hat{T}_1^2\Phi_{HF} = 2\sum_{i>j}^{occ}\sum_{a>b}^{virt}(t_i^a t_j^b - t_i^b t_j^a)\Phi_{ij}^{ab} \tag{3.167}$$

Substitution of equations 3.165 and 3.167 into equation 3.164 leads to;

$$\Delta E = \langle \Phi_{HF}|\hat{H}_N|\left(\sum_{i>j}^{occ}\sum_{a>b}^{virt} t_{ij}^{ab}\Phi_{ij}^{ab} + \sum_{i>j}^{occ}\sum_{a>b}^{virt}\left(t_i^a t_j^b - t_i^b t_j^a\right)\Phi_{ij}^{ab}\right)\rangle \tag{3.168a}$$

$$= \langle \Phi_{HF}|\hat{H}_N|\sum_{i>j}^{occ}\sum_{a>b}^{virt}\left(t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a\right)\Phi_{ij}^{ab}\rangle \tag{3.168b}$$

$$= \sum_{i>j}^{occ}\sum_{a>b}^{virt}\left(t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a\right)\langle ia||jb\rangle \tag{3.168c}$$

Thus, the coupled-cluster energy is given by $E_{CC} = E_{HF} + \Delta E$ with the integral expressed as;

$$\langle ia||jb\rangle = \int d\tau(1)d\tau(2)\psi_i^*(1)\phi_a^*(2)g(1,2)(1 - \hat{P}_{12})\psi_j(1)\psi_b(2) \tag{3.169}$$

Note that only the singles and doubles excitation amplitudes contributes directly to the coupled cluster energy whereas the higher-order excitation amplitude contribute indirectly irrespective of the truncation level of the cluster operator. Therefore, majority of the coupled-cluster energy is from the singles and doubles. The important special case of the coupled-cluster theory- **the coupled-cluster singles and doubles (CCSD) model** will be studied in more details in the next section and significant number of calculations in this dissertation were performed at this level.

3.5.1   The closed-shell CCSD model

In this model, the cluster operator is restricted to contain only the singles and doubles operators of singlet symmetry. The singles and doubles amplitudes contribute a lot to the total coupled-cluster energy and the doubles amplitudes contribute significantly over the singles amplitudes but the singles are important for the description of the molecular properties.

The singlet CCSD state is generated from the closed-shell Hartree-Fock state by applying the exponential cluster operator and the terms that transform as singlet in $\hat{T}_1$ and $\hat{T}_2$ are retained in the cluster operator:

$$|CC\rangle = e^{\hat{T}_1 + \hat{T}_2}|\Phi_{HF}\rangle \tag{3.170}$$

The singlet cluster operators should satisfy the spin tensors as defined below and this will impose constraints on the singles and doubles cluster amplitudes.

$$[\hat{S}_z, \hat{T}_i] = 0 \tag{3.171}$$

$$[\hat{S}_\pm, \hat{T}_i] = 0 \tag{3.172}$$

Let us deal with operator $\hat{T}_1$ by substituting equation 3.142 into equation 3.171 to get the following requirement

$$[\hat{S}_z, \hat{T}_1] = \sum_{a\sigma i\tau} t_{i\tau}^{a\sigma}[\hat{S}_z, a_{a\sigma}^\dagger a_{i\tau}] \tag{3.173a}$$

$$= \sum_{a\sigma i\tau} t_{i\tau}^{a\sigma}\left([\hat{S}_z, a_{a\sigma}^\dagger]a_{i\tau} + a_{a\sigma}^\dagger[\hat{S}_z, a_{i\tau}]\right) \tag{3.173b}$$

$$= \sum_{a\sigma i\tau}(\sigma - \tau)t_{i\tau}^{a\sigma}a_{a\sigma}^\dagger a_{i\tau} = 0 \tag{3.173c}$$

here σ and τ run over $-\frac{1}{2}$ and $\frac{1}{2}$. To satisfy the above equality condition, σ must be equal to τ. The other requirement is obtained from substitution of the same equation (eqn 3.142) into equation 3.172 to get two sets of equations.

$$[\hat{S}_+, \hat{T}_1] = \sum_{ai}(t_{i\beta}^{a\beta} - t_{i\alpha}^{a\alpha})a_{a\alpha}^\dagger a_{i\beta} = 0 \tag{3.174}$$

$$[\hat{S}_-, \hat{T}_1] = \sum_{ai}(t_{i\alpha}^{a\alpha} - t_{i\beta}^{a\beta})a_{a\beta}^\dagger a_{i\alpha} = 0 \tag{3.175}$$

here α and β amplitude must be identical to satisfy the above equalities. I can write the final form of the single cluster operator as a linearly independent combination of the

singlet one-electron excitation operators, E_{ai};

$$\hat{T}_1 = \sum_{ai} t_i^a E_{ai} \tag{3.176}$$

where $E_{ai} = a_{a\alpha}^\dagger a_{i\alpha} + a_{a\beta}^\dagger a_{i\beta}$ is the singlet one-electron excitation operator. I can follow the same procedure outlined above to determine the singlet doubles operator $\hat{T}_2$ using the above requirements and it should also satisfy the symmetry relation $t_{ij}^{ab} = t_{ji}^{ba}$, thus:

$$\hat{T}_2 = \frac{1}{2} \sum_{abij} t_{ij}^{ab} E_{ai} E_{bj} \tag{3.177}$$

The CCSD Energies. The coupled-cluster electronic energy is calculated from equation 3.155

$$E_{CC} = \langle \Phi_{HF} | e^{-\hat{T}} \hat{H} e^{\hat{T}} | \Phi_{HF} \rangle \tag{3.178}$$

and for the CCSD electronic energy, the above equation 3.178 can be rewritten in the form of commutator expansion (equation 3.157)

$$E_{CCSD} = E_{HF} + \frac{1}{2} \langle \Phi_{HF} | [[\hat{H}, \hat{T}_1], \hat{T}_1] | \Phi_{HF} \rangle + \langle \Phi_{HF} | [\hat{H}, \hat{T}_2] | \Phi_{HF} \rangle \tag{3.179}$$

I will deal with the last two terms (the correlation corrections from the singles and doubles). I can now substitute the CCSD operators defined by equations 3.176 and 3.177 in the expression 3.179 and obtain

$$E_{CCSD} = E_{HF} + \frac{1}{2} \sum_{aibj} t_i^a t_j^b \langle \Phi_{HF} | [[\hat{H}, E_{ai}], E_{bj}] | \Phi_{HF} \rangle +$$
$$\frac{1}{2} \sum_{aibj} t_{ij}^{ab} \langle \Phi_{HF} | [\hat{H}, E_{ai} E_{bj}] | \Phi_{HF} \rangle \tag{3.180}$$

which can be rewritten as;

$$E_{CCSD} = E_{HF} + \frac{1}{2} \sum_{aibj} \left(t_{ij}^{ab} + t_i^a t_j^b \right) \langle \Phi_{HF} | [[\hat{H}, E_{ai}], E_{bj}] | \Phi_{HF} \rangle \tag{3.181}$$

where I have used the following commutator relations for the doubles and higher order excitations;

$$\langle \Phi_{HF} | \hat{H}, E_{ai} E_{bj}] | \Phi_{HF} \rangle = \langle \Phi_{HF} | [[\hat{H}, E_{ai}], E_{bj}] | \Phi_{HF} \rangle \tag{3.182}$$

3.5.2   The CCSD(T) model

I have seen from the previous sections that the CCSD wave function is corrected to the second order in $\hat{T}_1$ and $\hat{T}_2$. The CCSD(T) method consists of, in addition to the CCSD energies, the corrections from non-iterative fourth order connected triples. Therefore, I can express the CCSD(T) energy as;

$$E^{CCSD(T)} = E^{CCSD} + \Delta E^{CCSD(T)} \tag{3.183a}$$

where $\Delta E^{CCSD(T)}$ is the CCSD(T) correction to the CCSD energy expressed as;[88](pp. 793)

$$\Delta E^{CCSD(T)} = \langle \bar{t}|[\hat{\Phi}, {}^*\hat{T}_3^{(2)}]|\Phi_{HF}\rangle \tag{3.184}$$

and these triples $\hat{T}_3^{(2)}$ correspond to the second-order amplitudes generated from CCSD amplitudes and those amplitudes in $\langle \bar{t}|$ are also the CCSD amplitudes defined by the following equations;[94]

$$\langle \bar{t}| = \sum_{ai} \bar{t}_i^a \left\langle {}^{\bar{a}}_{i} \right| + \frac{1}{2} \sum_{abij} \bar{t}_{ij}^{ab} \left\langle {}^{\overline{ab}}_{ij} \right| \tag{3.185}$$

$$|\bar{t}\rangle = \sum_{ai} t_i^a \left| {}^{a}_{i} \right\rangle + \frac{1}{2} \sum_{abij} t_{ij}^{ab} \left| {}^{ab}_{ij} \right\rangle \tag{3.186}$$

Note that the bra and the ket amplitudes are not identical because of the biorthogonal representation. I can deduce this two relations, $\bar{t}_i^a = 2t_i^a$ and $\bar{t}_{ij}^{ab} = 4t_{ij}^{ab} - 2t_{ji}^{ab}$ for the CCSD(T) amplitudes and then rewrite the CCSD(T) correction as;[88] (pp. 794)

$$\Delta E^{CCSD(T)} = \sum_{ai} \bar{t}_i^a \left\langle {}^{\bar{a}}_{i} \right| [\hat{\Phi}, {}^*\hat{T}_3^{(2)}] \left| \Phi_{HF} \right\rangle + \frac{1}{2} \sum_{abij} \bar{t}_{ij}^{ab} \left\langle {}^{\overline{ab}}_{ij} \right| [\hat{\Phi}, {}^*\hat{T}_3^{(2)}] \left| \Phi_{HF} \right\rangle \tag{3.187}$$

where the triples amplitudes ${}^*\hat{T}_3^{(2)}$ of the CCSD(T) calculation are also explained well from ref[88].

3.6   Density Functional Theory

Electronic wave functions for $N-$electron depend on the $4N$ coordinates. The singlet wave function has a form of Slater determinant that depends on $N-$spin orbitals and has $N!$ terms. The correlated wave functions for medium size molecules (~ 10 atoms) and reasonable basis set (aug-cc-pVnZ family) can easily require millions of determinants. There is need for an alternate formulation of electronic structure theory that would not lead to these computationally cumbersome wave functions.

The electronic Hamiltonian operator give us the information about the positions of the nuclei, their atomic numbers and the total number of electrons. The total number of electrons gives us information about the key variable, the electron density ρ, which can be obtained by integrating $|\Psi|^2$ over all electronic coordinates $\mathbf{r}$ except for one in which the integral is summed over its spin coordinate σ;

$$\rho(r) = \sum_{\sigma=\frac{1}{2},\frac{1}{2}} \int d\tau_2 d\tau_3 \cdots d\tau_N |\Psi(r,\sigma_1,r_2,\sigma_2,\cdots,r_N,\sigma_N)|^2 \qquad (3.188)$$

From this definition, $\rho(r)$ is independent of the label of the electron omitted from the integral therefore the integral of the electron density over all space return the total number of electrons N,

$$N = \int \rho(r) d^3r \qquad (3.189)$$

This is a 3 dimensional integral, with variables x, y, and z, the result of integration is the total number of electrons. DFT calculations are centralised around the electron density. This is more advantageous over the wave function methods because the dimensionality is reduced to 3D regardless of the number of electrons in the system, therefore, calculations can be performed on large systems.

Thomas-Fermi approach. The idea of determining energy using purely electron density was first illustrated by Thomas and Fermi in 1920s. In their model (Thomas-Fermi Model),[95] the kinetic energy term of the electrons T_e defined in equation 3.9/3.36 was made to depend exclusively on the electron density.

$$T[\rho] = C_F \int \rho^{5/3}(r) dr, \qquad with \qquad C_F = \frac{3}{10}(2\pi^2)^{2/3} = 2.871 \qquad (3.190)$$

The last two terms of equation 3.34, the electron-nucleus and the electron-electron interactions were also written in terms of electron density and the total energy of the system was defined as

$$E[\rho] = C_F \int \rho^{5/3}(r) dr - Z\frac{\rho(r)}{r}dr + \frac{1}{2}\int\int \frac{\rho(r_1)\rho(r_2)}{|r_1-r_2|}dr_1 dr_2 \qquad (3.191)$$

Hohenberg-Kohn Theorem. The fundamentals of the density functional theory were validated in 1964 by Hohenberg and Kohn (Hohenberg-Kohn(HK))[96] in which the electronic Hamiltonian was expressed as a functional of electronic density $\rho(r)$ based on two key theorems; The first theorem demonstrated that one can obtain external potential $v(r)$ from the electronic density and then the Hamiltonian can be expressed as a functional of electronic density. There is an agreement between the external potential and the electronic density and the electronic density and electronic wave function $\Psi(r)$ can be used interchangeably to give a full description of the system. The second

theorem demonstrated how the ground state electronic energy/density can be obtained variationally for a given external potential and a number of electrons.

$$E_v(\rho) \geq E_v(\rho_0) = E_0 \tag{3.192}$$

Consider a system of electrons enclosed in a box subjected to a known external potential $v(r)$ with the ground state electron density $\rho_0(r)$ and electronic energy E_0. If there is another external potential $v'(r)$ different from $v(r)$ that can give the same ground state electron density $\rho_0(r)$, then the Hamiltonian $\hat{H}$ and $\hat{H}'$ must differ, so is the normalised wave function Ψ and Ψ' for these two states. Then, the energy of these states respectively, are given by;

$$\begin{aligned} E_0 < \langle \Psi' | \hat{H} | \Psi' \rangle &= \langle \Psi' | \hat{H}' | \Psi' \rangle + \langle \Psi' | \hat{H} - \hat{H}' | \Psi' \rangle \\ &= E_0' + \int \rho(r)[v(r) - v'(r)]dr \end{aligned} \tag{3.193}$$

and

$$\begin{aligned} E_0' < \langle \Psi | \hat{H}' | \Psi \rangle &= \langle \Psi | \hat{H} | \Psi \rangle + \langle \Psi | \hat{H}' - \hat{H} | \Psi \rangle \\ &= E_0 - \int \rho(r)[v(r) - v'(r)]dr \end{aligned} \tag{3.194}$$

Summing up equation 3.193 and 3.194 yield contradicting equation, $E_0 + E_0' < E_0' + E_0$, demonstrating non-existence of two different external potentials that can yield the same electronic density $\rho(r)$. Then, for any given external potential $v(r)$, the electronic energy E_v can be expressed as a functional of electron density $\rho(r)$.

$$\begin{aligned} E_v[\rho(r)] &= T[\rho(r)] + V_{ne}[\rho(r)] + V_{ee}[\rho(r)] \\ &= \int \rho(r)v(r)dr + F_{HK}[\rho(r)] \end{aligned} \tag{3.195}$$

where $T[\rho(r)]$ is the kinetic energy of the many electron system, $V_{ee}[\rho(r)]$ is the electron-electron interaction energy where their sum gives $F_{HK}[\rho(r)]$, the universal functional of charge density $\rho(r)$. F_{HK} is independent from any external potential $v(r)$. The interaction of electron density with the external potential are accounted in the integral. If a system is described by a ground state wave function Ψ, and electron density $\rho(r)$, then the external potential $v(r)$ is uniquely being determined by $\rho(r)$. For any different wave function Ψ' and its related electron density $\rho'(r)$, the energy is expressed as;

$$E[\rho'(r)] = \int \rho'(r)v(r)dr + F_{HK}[\rho'(r)] \geq E[\rho(r)] \tag{3.196}$$

Thus, from the second Hohenberg-Kohn theorem, the ground state electron density can

be found by using the variational principle, $E_0[\rho_0] \leq E_v[\rho]$, and the minimum energy can be obtained only from the exact ground-state electron density.

Kohn-Sham formalism. Improvements of Thomas-Fermi model came in 1965 by Kohn and Sham[97] where they described the kinetic energy of the system as if the electrons were not interacting and moving in an effective potential. In Kohn-Sham formalism, the universal functional of charge density $F_{HK}[\rho(r)]$ is written as the sum of the kinetic energy of non-interacting electrons T_s, the Hartree energy E_H and the exchange and correlation energy E_{xc} which account for all the many-body quantum effects. From equations 3.195/3.196,

$$
\begin{aligned}
E[\rho(r)] &= \int \rho(r)v(r)dr + F_{HK}[\rho(r)] \\
&= \int \rho(r)v(r)dr + T_s[\rho(r)] + E_H[\rho(r)] + E_{xc}[\rho(r)] \\
&= \int \rho(r)v(r)dr + T_s[\rho(r)] + \frac{1}{2}\int \frac{\rho(r)\rho(r')}{|r-r'|}drdr' + E_{xc}[\rho(r)]
\end{aligned}
\tag{3.197}
$$

The kinetic energy T_s does not give a correct description for a system with interacting particles, hence the wave functions used are not true wave functions but the Pseudo-functions, the Kohn-Sham orbitals.

$$
T_s[\Psi_0] = \langle\Psi_0|\hat{T}\Psi_0\rangle = -\frac{1}{2}\sum_i \langle\phi_i|\nabla^2\phi_i\rangle = -\frac{1}{2}\sum_i \int \phi_i^*(r)\nabla^2\phi_i(r)dr
\tag{3.198}
$$

Therefore, I can minimise the above energy functional, equation 3.197, and require the variation in total energy $\delta E[\rho(r)]$ with respect to the variation of the spin-orbitals $\delta\phi_i^*$ to be zero.

$$
\frac{\delta}{\delta\phi_i^*}\left(E[\rho(r)] - \varepsilon_i\left[\int \rho(r)dr - N\right]\right) = 0
\tag{3.199}
$$

The integral minus N, $\int \rho(r)dr - N$, is the normalisation constraint on $\rho(r)$ with ε_i as the Lagrange multipliers. Here, the non-interacting electrons are moving in an effective potential v_{eff} which depends solely on the electron density $\rho(r)$,

$$
\begin{aligned}
V_{eff} &= \frac{\delta\{\int \rho(r)v(r)dr + E_H[\rho(r)] + E_{xc}[\rho(r)]\}}{\delta\rho(r)} \\
&= v(r) + \int \frac{\rho(r')}{|r-r'|}dr' + v_{xc}(r)
\end{aligned}
\tag{3.200}
$$

with v_{xc} defined as,

$$
v_{xc}(r) = \frac{\delta E_{xc}[\rho(r)]}{\delta\rho(r)}
\tag{3.201}
$$

One obtains a system of equations analogous to the Hartree-Fock equations, the one-electron Schrödinger like equations central to the Kohn-Sham DFT.

$$\left[-\frac{1}{2}\nabla^2 + \int \frac{\rho(r')}{|r - r'|}dr' + v_{xc}(r) + v(r) \right]\phi_i(r) = \varepsilon_i\phi_i(r) \tag{3.202}$$

$$\left[-\frac{1}{2}\nabla^2 + V_{eff} \right]\phi_i(r) = \varepsilon_i\phi_i(r) \tag{3.203}$$

with the electron density related to the Kohn-Sham orbitals through

$$\rho(r) = \sum_{i=1}^{N} |\phi_i|^2 \tag{3.204}$$

with the $\varepsilon_i's$ as the Kohn-sham one-electron orbital energies and the equation has to be solved iteratively. The total energy of the system can be found from the resulting density through substitution of equation 3.198 into equation 3.197 and the $-\frac{1}{2}\nabla^2\phi_i$ term is from equation 3.202.

$$E[\rho(r)] = \sum_i \int \phi_i^*(r)\left[\varepsilon_i - \int \frac{\rho(r')}{|r - r'|}dr' - v_{xc}(r) - v(r) \right]\phi_i(r)dr$$
$$+ \int \rho(r)v(r)dr + \frac{1}{2}\int\int \frac{\rho(r)\rho(r')}{|r - r'|}drdr' + E_{xc}[\rho(r)] \tag{3.205}$$

which simplifies to

$$E[\rho(r)] = \sum_i^N \varepsilon_i - \int \rho(r)v_{xc}(r)dr - \frac{1}{2}\int\int \frac{\rho(r)\rho(r')}{|r - r'|}drdr' + E_{xc}[\rho(r)] \tag{3.206}$$

These are the Kohn-Sham equations, 3.200, 3.203 and 3.206 which must be solved iteratively, with $\varepsilon_i's$ in equation 3.206 as the lowest eigenvalues. The procedure is as follows;

1. Obtain an initial guess of electron density

2. Construct the effective potential V_{eff} from equation 3.200

3. Obtain Kohn-Sham orbitals from equation 3.203 using this V_{eff}

4. Obtain new electron density from equation 3.204

5. Calculate the enrgy using this new electron density from equation 3.206

6. Check for convergence. Determine if the new electron density in step 4 is the same as the electron density from the previous step or within a specified criterion. If not converged, then return to step (2) with the new electron density

7. If converged, use the last electron density and the energy as the final calculated parameters

3.6.1 The exchange correlation functional

Early approximations to the exchange-correlation functional were derived from the homogeneous electron gas (HEG) model.[97] In this model, N interacting electrons were placed in space volume V with positively charged background to neutralise the system. The electron density $\rho = N/V$ was sampled by varying both N and V to infinity, keeping ρ finite. The exchange-correlation energy functional is a function of electronic density at each point in space. The energy value of this functional depend solely upon the value of the local density where the functional is evaluated, hence the approximation is termed the local density approximation (LDA). This exchange-correlation energy functional is not a functional of the derivatives of the density or the Kohn-Sham orbitals and it can be decomposed into exchange E_x and correlation E_c terms linearly.'

$$E_{xc} = E_x + E_c \tag{3.207}$$

The general expression for the exchange-correlation energy for spin unpolarised system is written as,

$$E_{xc}^{LDA}[\rho] = \int \rho(r)\varepsilon_{xc}(\rho)dr \tag{3.208}$$

and for the local spin density approximation (LSDA) as;

$$E_x^{LSDA} = -\frac{3}{4}\left(\frac{6}{\pi}\right)^{1/3} \int \left[(\rho^\alpha)^{4/3} + (\rho^\beta)^{4/3}\right] dr \tag{3.209}$$

The commonly used LDA exchange energy functional include Vosko-Wilk-Nusair (VWN), Perdew-Zunger (PZ81), Cole-Perdew (CP), Perdew-Wang (PW92) etc and they will not be discussed in this dissertation.

The other class of approximations to the exchange-correlation energy functional is the generalised gradient approximations (GGA). This approximation is considered to be an upgrade to the LDA. In addition to the electron density in the LDA, it also uses the gradient of the electron density to correct the variation of the electron density with variation of positions. The expression for the GGA exchange-correlation energy functional is as follows,

$$E_{xc}^{GGA} = E_{xc}^{GGA}[\rho(r), \nabla\rho(r)] \tag{3.210}$$

The GGA exchange-correlation energy functional can also be decomposed into the exchange part, E_x^{GGA} and the correlation part, E_c^{GGA};

$$E_{xc}^{GGA} = E_x^{GGA} + E_c^{GGA} \tag{3.211}$$

Here I have Perdew-Wang (PW86),[98] Becke (B88),[99] Lee-Yang-Parr (LYP),[100] Perdew-Burke-Ernzerhof (PBE),[101] etc. functionals.

The development of density functionals is still ongoing because more and more different DFT problems have to be addressed.[102–108] Some problems includes studies of the systems where dispersion interactions are important[102–105]. In some cases, incorrect descriptions for the transition metal oxides have been observed.[106–108] New functionals[109, 110] have been developed to address this problems and they are termed meta GGA functionals. This meta-GGA functionals depends on, in addition to the ρ and $\nabla\rho$, the kinetic energy density τ or $\nabla^2\rho$, where τ is expressed as;

$$\tau = \frac{1}{2}\sum_i^{occ} |\nabla\psi_i|^2. \tag{3.212}$$

New better performing functionals which mix Hartree-Fock exact exchange into pure GGA and meta GA functioanls were developed and they were named hybrids functionals. Their improved performance is due to the nonlocality of the exact exchange. Examples include B3LYP,[100, 111–113] M06-L,[114] M06-2X,[110] etc. In this dissertation, I will only briefly discuss the B3LYP functional because it was extensively used for the screening purposes. This functional has been extensively tested and seemed to perform very well in many cases. The expression of the B3LYP exchange-correlation functional is as follows;

$$E_{xc}^{B3LYP} = (1 - a_0 - a_x)E_x^{LSDA} + a_0 E_x^{HF} + a_x E_x^{B88} + (1 - a_c)E_c^{VWN} + a_c E_c^{LYP} \tag{3.213}$$

with constants $a_0 = 0.20$, $a_x = 0.72$ and $a_c = 0.81$ obtained by parameter fitting to the experimental molecular atomisation energies.

3.7   Correlation consistent basis sets

In this dissertation, I will only discuss the correlated consistent family of basis sets developed by Dunning and co-workers[115] purposely for recovering the correlation energy of a chemical system and accomplish the convergence of the electronic energies by extrapolation to the complete basis set (CBS) limit. The design of these basis sets were based on previous work by Almlöf and Taylor[116] and by Jankowski *et al*[117].

Almlöf and Taylor efficiently recovered the correlation energy of the system by truncating set of atomic natural orbitals as basis functions.

Jankowski *et al* performed high-level correlated calculations on fluoride and analysed the effects of adding polarisation functions to the basis sets. They added even-tempered sets of s and p polarisation functions and optimised the parameters of each set with each additional functions added. The exponents of these basis functions were determined through:

$$\alpha_n = \frac{\alpha_0}{q^n}, \qquad n = 1, 2, ... \tag{3.214}$$

with α_0 being the smallest exponent of the $s-$ or $p-$functions in the basis set and q is the geometric progression constant. In their analysis, they figured that certain sets of functions recovered roughly equal amounts of correlation energy and devised a truncation scheme which allowed addition of groups of functions which recovered similar amounts of the correlation energy.

Then Dunning generalised their ideas and performed correlation calculations in order to optimise the even-tempered sets of polarisation functions.[115] He started with oxygen atom but constrained $1s-$orbitals to double occupancy because they were unable to correlated with valence electrons. Dunning adopted Jankowski's scheme and started grouping these functions. At each grouping stage, one function of the next highest angular momentum is added to the existing set to form a new sets of basis functions. For example, $4s3p2d$ is superseded by $5s4p3d2f$ which is then replaced by $6s5p4d3f2g$. This ensured that the correlation energy is recovered consistently at each stage as all the functions contribute roughly equally, so is the name *correlation consistent* basis sets. These basis sets were labelled "$N-$tuple zeta" where N is the cardinal number representing both the maximum value of l in the set and the number of sets of Gaussians per valence orbital. The cardinal number is well represented as N=D,T,Q,5,6,$\cdots$ ("double-, triple-, $\cdots$") zeta and Dunning's first and second row basis sets were named **cc-pVNZ**.[118–120] The 'cc-p', stands for '*correlation-consistent polarized*' and the 'V' indicates they are valence-only basis sets as noted earlier. The basis sets constructed later included 'tight' functions which were optimised while allowing core-valence correlation. These basis sets were named the **C**ore-**V**alence; **cc-pCVNZ**.[120]

Studies of the dipole-bound anions, Rydberg states require basis set with extra diffuse functions because they often involve more diffuse orbitals. The calculations require functions with small exponents to correctly describe these systems. From the cc-pVDZ basis sets described above, Dunning and co-workers extended them by adding single Gaussian function of each symmetry and optimised its exponent. These type of basis sets were named '**aug**mented-**cc-pVDZ**.[121] The augmented correlation-consistent basis sets used in the remaining chapters of this dissertation are listed in table 3.1. Although

| | Contracted functions | |
Basis set	First row elements	Hydrogen atom
aug-cc-pVDZ	4s, 3p, 2d	3s, 2p
aug-cc-pVTZ	5s, 4p, 3d, 2f	4s, 3p, 2d
aug-cc-pVQZ	6s, 5p, 4d 3f, 2g	5s, 4p, 3d, 2f

TABLE 3.1: Composition of some common correlated consistent basis sets

these basis sets are incomplete, one may add extra diffuse functions with low exponents. These exponents should form an even-tempered set of basis functions calculated from equation 3.214 with α_0 as the minimum exponent from the standard basis set which undergoes geometric progression.[122]

Extrapolation to the complete basis set limit. The systematic convergence of Dunning's basis sets to the CBS limit suggests an empirical extrapolation techniques. At the CBS limit, the cardinal number N is infinity. Many suggestions about extrapolations were formulated.[123, 124] The Helgaker[123] formulations were based on extrapolating the Hartree-Fock energy and the correlation energy. In principle, the extrapolation of the Hartree-Fock energy can be achieved easily as a finite number of basis functions could describe the single-determinant Hartree-Fock wave function but the correlated calculations require arbitrary large basis.[123] The three-point extrapolation scheme applied to the Hartree-Fock and correlated calculation from this paper complied with Feller's suggestions.[124] Their extrapolation model is as follows;

$$E_N^{HF} = E_{CBS}^{HF} + Be^{-AN} \tag{3.215}$$

with E_N^{HF} as the energy calculated with the $N-$tuple zeta basis set, E_{CBS}^{HF} as the Hartree-Fock energy at the complete CBS limit whereas A and B are fitting parameters.

The correlation energy with respect to the basis set size has the form[123]

$$E_N^{corr} = E_{CBS}^{corr} + AN^{-3} \tag{3.216}$$

The above two formulations, Bachorz *et al* proposed the expressions for the Hartree-Fock energy

$$E_{CBS}^{HF} = \frac{E_{N-2}E_N - E_{N-1}^2}{E_{N-2} - 2E_{N-1} + E_N} \tag{3.217}$$

and correlated energy

$$E_{CBS}^{corr} = \frac{E_{N-1}(N-1)^3 - E_N N^3}{(N-1)^3 - N^3} \tag{3.218}$$

at the basis set limit.[125]

Basis-set superposition error

As mentioned earlier, the localised basis set are incomplete. The problem become significant when dealing with molecular complexes. Suppose I wish to calculate the interaction energy between monomer X and monomer Y. The interaction energy E_{int} is usually calculated (using the supermolecular calculation) as the difference between the energy of the complex $E_{XY}(G)$ at any optimal geometry G and the sum of the energies of the monomers;

$$E_{int} = E_{XY}(G) - (E_X(G_X) + E_Y(G_Y)) \tag{3.219}$$
$$= E_{XY}(G) - E_X(G_X) - E_Y(G_Y) \tag{3.220}$$

$E_X(G_X)$ and $E_Y(G_Y)$ are the energies of the monomer X and Y at their optimal geometries G_X and G_Y, respectively, on the monomer complete basis set (MCBS). This might look obvious at first glance. Let us consider the situation when operating in the basis set approximation. The monomer X has a set of basis function $\|X)$, and Y has set of basis functions $\|Y)$, and the dimer has both sets of functions $\|X + Y)$.

$$E_{int} = E_{XY}(G\|X + Y) - E_X(G_X\|X) - E_Y(G_Y\|Y) \tag{3.221}$$

Here the dimer energy $E_{XY}(G\|X + Y)$ is calculated using a different and larger set of basis functions $\|X+Y)$ than the energies of the monomers, $E_X(G_X\|X)$ and $E_Y(G_Y\|Y)$. In addition, in the dimer, the monomer X can benefit from the basis functions $\|Y)$ of the monomer Y and vice-versa. As a result, their energies are lowered in the dimer. However, in the monomer energy calculation, the monomer X has only access to its own set of basis function $\|X)$, so is monomer Y. The artificial lowering of the dimer energy with respect to the monomers is termed "the basis set superposition error, BSSE".[89] The overlap of the set of basis function $\|Y)$ with the set of basis function $\|X)$ depends on the intermolecular distance between the monomers, X and Y in the dimer. If the monomer X is so close to the monomer Y such that their set of basis functions overlap, then monomer X will benefit from the set of basis function $\|Y)$ of the monomer Y and the more the closer the monomer X is, to the monomer Y, the more the overlap and the more the BSSE. In short, the magnitude of the BSSE is dependent on the intermolecular distance.

The counterpoise correction to the basis set superposition error. So, let us estimate the energy of the dimer $E_{XY}(G)$ at any geometry G using monomer energies;

$$E_{XY}(G) = E_X(G_X) + E_Y(G_Y) + E_X^{1bt}(G) + E_Y^{1bt}(G) + E_{XY}^{2bt}(G) \tag{3.222}$$

where $E_X^{1bt}(G)$ and $E_Y^{1bt}(G)$ are the interaction energies of one-body term $1bt$ in the dimer complete basis set calculated as

$$E_W^{1bt}(G) = E_W(G) - E_W(G_W) \qquad (3.223)$$

with $E_W(G)$ as the energy of the monomer W at its optimal geometry G in the dimer complete basis set (DCBS) and $E_W(G_W)$ as the energies of the monomer W at its optimal geometry on the MCBS. $E_{XY}^{2bt}(G)$ is the interaction energy of the two-body term $2bt$ in the DCBS calculated as; Energy of two-body term;

$$E_{XY}^{2bt}(G) = E_{XY}(G) - E_X(G) - E_Y(G) \qquad (3.224)$$

The $E_{XY}^{2bt}(G)$ term has to be counterpoise corrected. The Boys and Bernardi prescription;[234]

$$E_{XY}^{2bt} = E_{XY}(G\|X+Y) - E_X(G\|X+Y) - E_Y(G\|X+Y) \qquad (3.225)$$

and the counterpoise corrected total dimer energies are given by;

$$E_{XY}^{CP} = E_{XY}(G\|X+Y) + E_X(G\|X) - E_X(G\|X+Y) + E_Y(G\|Y) - E_Y(G\|X+Y) \qquad (3.226)$$

$$= E_{XY}(G\|X+Y) - BSSE_X(G) - BSSE_Y(G) \qquad (3.227)$$

Here the basis set superposition error (BSSE) is defined as the energy of the monomer due to the presence of "ghost" basis functions.

$$BSSE_X(G) = E_X(G\|X+Y) - E_X(G\|X) \qquad (3.228)$$

$$BSSE_Y(G) = E_Y(G\|X+Y) - E_Y(G\|Y) \qquad (3.229)$$

therefore, I define the interaction energy of the two-body system as;

$$E_{XY}^{2bt} = E_{XY}^{CP}(G) - E_X(G\|X) - E_Y(G\|Y) \qquad (3.230)$$

which is exactly the same as;

$$E_{XY}^{2bt} = E_{XY}(G\|X+Y) - E_X(G\|X+Y) - E_Y(G\|X+Y) \qquad (3.231)$$

3.8   Statistical Thermodynamic Properties

In classical physics, a system observable is the property of the system state that can be determined by some sequence of physical operations. Suppose I want to understand the system having N-particles. Any observable that need to be measured, e.g. energy, depends on the coordinates and momenta of the particles in the system. Therefore, the observable will be a functional of momenta $p(t)$ and positions $r(t)$. For example, consider an observable A, its instantaneous value for a system having N-particles at time t would be $A[p^N(t), r^N(t)]$. Particles interact with each other, as a result, the value of the observable fluctuates and what is measured would be the average value. If for example, the time for calculations is long enough such that time t approaches infinity, then the value of the observable will approach exact average value. That is;

$$A_{ave} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A[p^N(t), r^N(t)] dt \tag{3.232}$$

For a system comprised of a very large number of particles, the calculations are not feasible. Therefore, the average value of an observable is replaced by ensemble average. The idea was to assume that a system is made up of several replicas, each replica representing a possible state that the real system might be in. Then the ensemble average would be the average of those replicas considered all at once. It is regarded as the expectation value of the observable.

$$\langle A \rangle = \int \int dp^{3N} dr^{3N} A[p^{3N}, r^{3N}] \rho[p^{3N}, r^{3N}] \tag{3.233}$$

$\rho[p^N, r^N]$ is the probability density in the ensemble. For a system with a fixed number of particles, the probability density is a well known Boltzmann distribution.

$$\rho(p, r) = \frac{e^{[-E(p,r)/k_B T]}}{Q} \tag{3.234}$$

where E is the energy, k_B is the Boltzmann's constant, T is the system's temperature and Q is the partition function. For a system having a constant number of particles, N, volume, V, and temperature,T, (the canonical ensemble), the partition function is expressed as,

$$Q_{NVT} = \frac{1}{N!} \frac{1}{h^{3N}} \int \int dp^N dr^N exp \left[-\frac{H(p^N, r^N)}{k_B T)} \right] \tag{3.235}$$

$N!$ is due to indistinguishability of the particles. H is the Hamiltonian function, which is the total energy of the system. It is equal to the sum of the kinetic energy, $\Xi(p)$, and the potential energy, $V(r)$.

$$H(p^N, r^N) = \Xi(p^N) + V(r^N) \tag{3.236}$$

There are several observables that can be determined by a computer simulation under different ensembles (e.g. canonical, microcanonical, grandcanonical, isothermal-isobaric ensemble etc.). One way to quantify these observables is by comparison with the experimental data. Internal energy is the observable that is mostly determined in order to get information about the system. The values that are usually obtained during the course of the simulations are ensemble averages;

$$\langle E \rangle = \frac{1}{M} \sum_{i=1}^{M} E_i \tag{3.237}$$

The other observable that is also determined most frequently, is the simulation temperature, T. It is directly related to the kinetic energy of the system by the following expression ;

$$\Xi(p^N) = \sum_{i=1}^{3N} \frac{m_i v_i \cdot m_i v_i}{2m_i} = \sum_{i=1}^{3N} \frac{P_i^2}{2m_i} = \frac{k_B T}{2}(3N - N_c) \tag{3.238}$$

where m_i is the mass of particle i and N_c is the number of constraints on the system. Kinetic energy depends only on the momentum of the particles. It can also be argued that it depends on the velocities of the particles, $v_i = p_i/m_i$. Pressure is another observable. I am not going to show the derivation of the gas laws but it is related to the system's kinetic energy or temperature by the following expression for perfect gas; $PV = Nk_BT$. The equation is slightly modified for real gases.

Here I discuss the most important equations in thermodynamics. These are based on a non-interacting system of particles applicable only to ideal gas. The reason being that my discussions in chapter 5 onwards are based on the calculations performed in the gas phase conditions. The observables that I am interested in are the contributions to the entropy, energy and heat capacity resulting from translational, electronic, rotational and vibrational motions.

For a system of identical, indistinguishable particles, the number of available molecular states is much greater than the number of particles with its canonical partition function $Q(N, V, T)$ given by;

$$Q(N, V, T) = \frac{[q(V,T)]^N}{N!} \tag{3.239}$$

which simplifies a many-body problem to a one-body problem and $q(V, T)$ is the molecular partition function. Let us consider an ideal gas which can undergo translational, rotational, vibrational, electronic and nuclear degrees of freedom. Each degrees of freedom can be treated as separate quantity to a very good approximation, if I use the rigid rotor-harmonic oscillator approximation, therefore I can write the Hamiltonian as sum

of independent terms.

$$q(V,T) = \sum_j e^{-E_j/k_B T} = \sum_{i,j,k,...} e^{-(\varepsilon_i + \varepsilon_j + \varepsilon_k + ...)/k_B T} \qquad (3.240)$$

where E_j are the energy states of $N-$body problem. In this case, the $N-$body energy is $E_{i,j,k,..} = \varepsilon_i + \varepsilon_j + \varepsilon_k + \cdots$. From the fact that the electronic and nuclear degrees of freedom can be written separately, that is, $\hat{H} = \hat{H}_t + \hat{H}_r + \hat{H}_v + \hat{H}_e + \hat{H}_n$, with subscripts t, r, v, e and n for translation, rotation, vibration, electronic and nuclear, respectively. This implies that the total energy is the sum of the contributions of each energy component, $\varepsilon = \varepsilon_t + \varepsilon_r + \varepsilon_v + \varepsilon_e + \varepsilon_n$. So, from equation 3.240, I have;

$$q(V,T) = q_t q_r q_v q_e q_n \qquad (3.241)$$

Each contribution will be studied in the next section in the context of entropy, energy and the heat capacity. The entropy contributions S is calculated from[126]

$$S = Nk_B + Nk_B \ln\left(\frac{q(V,T)}{N}\right) + Nk_B T\left(\frac{\partial \ln q}{\partial T}\right)_V \qquad (3.242)$$

In Gaussian electronic structure code, the Nk_B term is substituted with R $(R = Nk_B)$ and the molar masses are used such that $n = N/N_A$, hence $N = 1$ after moving the first term in the natural log as e.

$$\begin{aligned}
S &= R + R\ln(q(V,T)) + RT\left(\frac{\partial \ln q}{\partial T}\right)_V \\
&= R\ln(q(V,T)e) + RT\left(\frac{\partial \ln q}{\partial T}\right)_V \\
&= R\left(\ln(q_t q_e q_r q_v e) + T\left(\frac{\partial \ln q}{\partial T}\right)_V\right) \qquad (3.243)
\end{aligned}$$

I can also use the partition function to calculate the total energy of the system through,

$$E = Nk_B T^2\left(\frac{\partial \ln q}{\partial T}\right)_V \qquad (3.244)$$

and then the heat capacity,

$$C_V = \left(\frac{\partial E}{\partial T}\right)_{N,V} \qquad (3.245)$$

3.8.1 Contributions from translational motion

I will not dwell much into the derivation of the translational partition function q_v, but it is given by

$$q_t = \left(\frac{2\pi m k_B T}{h^2}\right)^{3/2} \cdot V \tag{3.246}$$

for the ideal monoatomic gas and by

$$q_t = \left(\frac{2\pi (m_1 + m_2) k_B T}{h^2}\right)^{3/2} \cdot V \tag{3.247}$$

for the ideal diatomic gas.[127](pp. 85/93) This can be generalised to polyatomic molecules and rewrite the translational term as;

$$q_t = \left(\frac{2\pi M k_B T}{h^2}\right)^{3/2} \cdot V \tag{3.248}$$

with M as the total mass of the system. The internal energy of the system from translations can be obtained by calculating the partial derivative of q_t with respect to T,

$$\left(\frac{\partial \ln q_t}{\partial T}\right)_V = \frac{3}{2T} \tag{3.249}$$

From ideal gas law, $PV = NRT = \frac{n}{N_A} N_A k_B T$, and $V = \frac{k_B T}{P}$, equation 3.248 become,

$$q_t = \left(\frac{2\pi M k_B T}{h^2}\right)^{3/2} \cdot \frac{k_B T}{P} \tag{3.250}$$

Then from equations 3.243 and 3.249 I can write the translational entropy as,

$$S_t = R\left(\ln(q_t e) + T\left(\frac{3}{2T}\right)\right)$$
$$= R(\ln q_t + 1 + 3/2) \tag{3.251}$$

using Stirling's approximation. I therefore, write the contributions due to the internal thermal energy from translational motion as,

$$E_t = N k_B T^2 \left(\frac{\partial \ln q}{\partial T}\right)_V$$
$$= RT^2 \left(\frac{3}{2T}\right)$$
$$= \frac{3}{2} RT \tag{3.252}$$

and the heat capacity at constant volume from equation 3.245 as,

$$
\begin{aligned}
C_t &= \frac{\partial E_t}{\partial T} \\
&= \frac{3}{2}R
\end{aligned}
\tag{3.253}
$$

3.8.2 Contributions from electronic motion

The electronic partition function defined by the following equation,[127](pp. 96)

$$
q_e = \omega_0 e^{-\epsilon_0/k_B T} + \omega_1 e^{-\epsilon_1/K_B T} + \omega_2 e^{-\epsilon_2/k_B T} + \cdots
\tag{3.254}
$$

ω is the degeneracy of the energy levels and the ϵ_n is the energy of the nth level. The electronic partition function for the ground state system (with $\varepsilon_0 = 0$ as the reference state) with assumptions that higher energy levels are not accessible is given by;

$$
q_e = \omega_0
\tag{3.255}
$$

Therefore the entropy of the molecule due to electronic motion from equation 3.243 simplifies to,

$$
\begin{aligned}
S_e &= R\left(\ln q_e + T\left(\frac{\partial \ln q_e}{\partial T}\right)_V\right) \\
&= R(\ln q_e + 0)
\end{aligned}
\tag{3.256}
$$

The last term vanished because there is no temperature dependent terms in the partition function.

3.8.3 Contributions from rotational motion

Molecular rotations can be categorised into linear polyatomic molecules and non-linear polyatomic molecules. For single atom, the contributions from the rotational motion does not depend on temperature $q_r = 1$, thus the contributions to the internal energy, heat capacity and entropy are all zero. For linear polyatomic molecules, the contributions are given by the rotational partition function[127](pp. 133),

$$
q_r = \frac{8\pi^2 I k_B T}{h^2} = \frac{1}{\sigma_r}\left(\frac{T}{\Theta_r}\right)
\tag{3.257}
$$

with I as the moment of inertia and σ as the symmetry number. The σ for a polyatomic molecule is simply the number of ways that the molecule can be rotated into itself. Here

q_r is temperature dependent quantity, therefore the entropic contributions from equation 3.243 are given by,

$$
\begin{aligned}
S_r &= R\left(\ln q_r + T\left(\frac{\partial \ln q_r}{\partial T}\right)_V\right) \\
&= R(\ln q_r + 1)
\end{aligned}
\tag{3.258}
$$

and the contribution from the internal thermal energy from equation 3.244 is given by,

$$
\begin{aligned}
E_r &= RT^2\left(\frac{\partial \ln q_r}{\partial T}\right)_V \\
&= RT^2\left(\frac{1}{T}\right) \\
&= RT
\end{aligned}
\tag{3.259}
$$

and the heat capacity from equation 3.245 reduces to

$$
\begin{aligned}
C_r &= \left(\frac{\partial E_e}{\partial T}\right)_V \\
&= R
\end{aligned}
\tag{3.260}
$$

For non-linear polyatomic molecules, the problem is a little bit complicated and the contribution from rotational partition function is derived in ref[127](pp. 136)

$$
q_r = \frac{\pi^{1/2}}{\sigma_r}\left(\frac{T^3}{(\Theta_{r,x}\Theta_{r,y}\Theta_{r,z})}\right)^{\frac{1}{2}}
\tag{3.261}
$$

From equations 3.243, 3.249 and 3.261, the contributions from the entropical effects simplifies to,

$$
\begin{aligned}
S_r &= R\left(\ln q_r + T\left(\frac{\partial \ln q_r}{\partial T}\right)_V\right) \\
&= R\left(\ln q_r + T\left(\frac{3}{2T}\right)\right) \\
&= R\left(\ln q_r + \frac{3}{2}\right)
\end{aligned}
\tag{3.262}
$$

with the contributions from the internal thermal energy due to rotational motion of the molecule

$$
\begin{aligned}
E_r &= RT^2 \left(\frac{\partial \ln q_r}{\partial T} \right)_V \\
&= RT^2 \left(\frac{3}{2T} \right) \\
&= \frac{3}{2} RT
\end{aligned}
\tag{3.263}
$$

obtained by simplifying equation equation 3.244 by substitution of equation 3.261 with the help from equation 3.249. Finally, equation 3.245 for the heat capacity reduces to;

$$
\begin{aligned}
C_r &= \left(\frac{\partial E_r}{\partial T} \right)_V \\
&= \frac{3}{2} R
\end{aligned}
\tag{3.264}
$$

3.8.4 Contributions from vibrational motion

All molecules are completely not at rest. Corrections to the zero-point energy vibrations are measured from the bottom of the potential well and each nuclei vibration is treated as a function of $3n-6$ coordinates system. An ideal diatomic molecule vibrates like two masses attached by a spring, therefore the potential of the system can be approximated by a quantum harmonic oscillator approximation with the energy that varies with the square of the displacement from equilibrium. The vibrational energy levels within the quantum harmonic oscillator are quantised and defined by,

$$
\varepsilon_n = \left(n + \frac{1}{2} \right) \hbar v \qquad n = 0, 1, 2, \cdots
\tag{3.265}
$$

with $\hbar = h/2\pi$ where h is the plank constant. The natural frequency is given by $v = \frac{1}{2\pi} \sqrt{k/\mu}$ with k as the bond force constant and μ as the reduced mass approximated as;

$$
\mu = \frac{m_1 m_1}{m_1 + m_2}
\tag{3.266}
$$

The vibrational partition function q_v for diatomic molecule is from McQuarrie[127](pp. 96)

$$q_v(T) = \sum_n e^{-\beta \varepsilon_n} \tag{3.267}$$

$$= e^{-\beta h v/2} \sum_{n=0}^{\infty} e^{-\beta h v n} \tag{3.268}$$

$$= \frac{e^{-\Theta_{v,i}/2T}}{1 - e^{\Theta_{v,i}/T}} \tag{3.269}$$

with $\beta = 1/k_B T$ and $\Theta_{v,i} = h v k_B$. For polyatomic molecule, with known normal frequencies $v_1, v_2, \cdots, v_\alpha$, the partition function is given by;

$$q_v(T) = \prod_{i=1}^{\alpha} \frac{e^{-\Theta_{v,i}/2T}}{1 - e^{\Theta_{v,i}/T}} \tag{3.270}$$

where $\alpha = 3n - 6$. This is applicable to results printed out from Gaussian electronic structures code. Of course the zero of energy $V = 0$ can be taken as the first vibrational energy level $n = 0$, as shown in figure 3.1. In this case the partition function for each and the overall vibrational levels, respectively, are given by;

$$q_v(T) = \frac{1}{1 - e^{\Theta_{v,i}/T}} \qquad for\ each\ vibrational\ level \tag{3.271}$$

$$q_v(T) = \prod_{i=1}^{\alpha} \frac{1}{1 - e^{\Theta_{v,i}/T}} \qquad for\ overall\ vibrational\ levels \tag{3.272}$$

and the energies referenced to this energy level are also printed in the Gaussian output file.

The contribution to the entropy is also calculated from equation 3.243;

$$S_r = R \left(\ln q_v + T \left(\frac{\partial \ln q_v}{\partial T} \right)_V \right) \tag{3.273}$$

The natural logarithm of contribution from the vibrational partition function q_v;

$$\ln q_v = \sum_{i=1}^{\alpha} \left(\frac{\Theta_{v,i}}{2T} + \ln \left(1 - e^{-\theta_{v,i}/T} \right) \right) \tag{3.274}$$
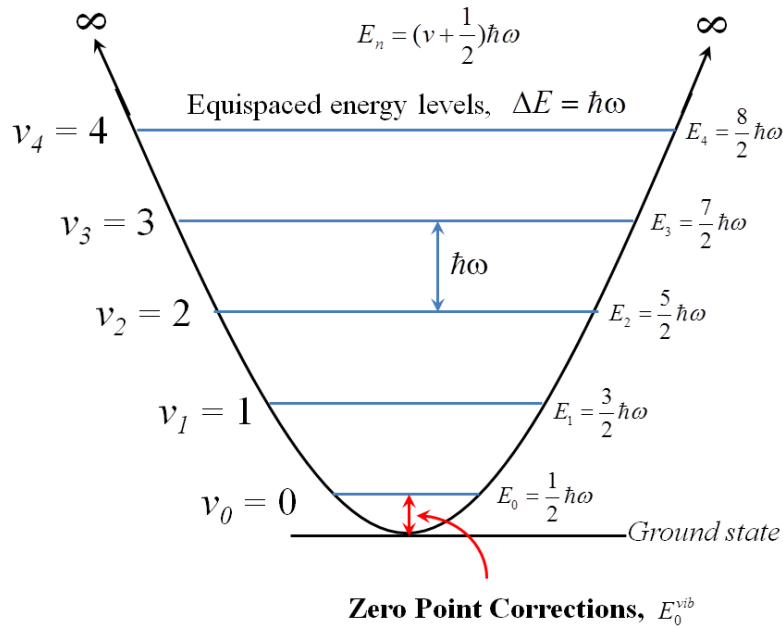
FIGURE 3.1: *The Harmonic potential well with zero point energies corrections shown.*

is then substituted to equation 3.243/3.273 with its first partial derivative and the equation leads to;

$$s_v = R\left(\sum_{i=1}^{\alpha}\left(\frac{\Theta_{v,i}}{2T} + \ln\left(1 - e^{-\theta_{v,i}/T}\right)\right) + T\left(\sum_{i=1}^{\alpha}\frac{\Theta_{v,i}}{2T^2} + \sum_{i=1}^{\alpha}\frac{(\Theta_{v,i}/T^2)e^{-\Theta_{v,i}/T}}{1 - e^{-\Theta_{v,i}/T}}\right)\right)$$

$$= R\left(\sum_{i=1}^{\alpha}\ln\left(1 - e^{-\Theta_{v,i}/T}\right) + \left(\sum_{i=1}^{\alpha}\frac{(\Theta_{v,i}/T)e^{-\Theta_{v,i}/T}}{1 - e^{-\Theta_{v,i}/T}}\right)\right) \tag{3.275}$$

which can be simplified by multiplying with $\frac{e^{-\Theta_{v,i}/T}}{e^{-\Theta_{v,i}/T}}$;

$$S_v = R\sum_{i=1}^{\alpha}\left(\frac{\Theta_{v,i}/T}{e^{\Theta_{v,i}/T} - 1} - \ln\left(1 - e^{-\Theta_{v,i}/T}\right)\right) \tag{3.276}$$

The contribution to internal thermal energy from molecular vibrations from equation 3.244 is given by;

$$E_v = NK_B\sum_{i=1}^{\alpha}\left(\frac{\Theta_{v,i}}{2} + \frac{\Theta_{v,i}e^{-\Theta_{v,i}/T}}{1 - e^{-\Theta_{v,i}/T}}\right) \tag{3.277}$$

$$= R\sum_{i=1}^{\alpha}\Theta_{v,i}\left(\frac{1}{2} + \frac{1}{e^{\Theta_{v,i}/T} - 1}\right) \tag{3.278}$$

3.9 Remarks

The methods discussed in this chapter will be used extensively in the remaining sections of the dissertation.

Chapter 4

Electron Binding Energies

4.1 Koopmans' Theorem

Consider the neutral N-electron closed-shell system represented by the optimised Hartree-Fock state $|\Phi_{HF}\rangle = |\chi_1 \chi_2 \cdots \chi_a \chi_b \cdots \chi_N\rangle$ of energy, equation 3.101,

$$E_{HF}^N = \langle \Phi_{HF} | \hat{H} | \Phi_{HF} \rangle \tag{4.1}$$

The Fock operator $\hat{f}$ defined in equation 3.60/3.73 is a well-defined Hermitian operator with infinite number of eigenfunctions and each solution $|\chi_i\rangle$ will have a spin orbital energy ε_i such that equation 3.71 reads;

$$f|\chi_i\rangle = \varepsilon_i |\chi_i\rangle \qquad i = 1, 2, \ldots, \infty \tag{4.2}$$

The N spin orbitals with the lowest orbital energies are just the occupied spin orbitals of $|\Phi_{HF}\rangle$ whereas the remaining infinite number of spin orbitals are virtual spin orbitals. Suppose we annihilate the occupied spin orbitals such that the ionised system is described by single-configuration state $a_{i\sigma}|\Phi_{HF}\rangle$, where σ is the spin coordinates, then the expanded cationic wave-function in the configuration generated from the Hartree-Fock state would be represented as;[88]

$$|N-1\rangle = \sum_i C_i a_{i\sigma} |\Phi_{HF}\rangle \tag{4.3}$$

The ionised state

$$|N-1\rangle_{i\sigma} = a_{i\sigma} |\Phi_{HF}\rangle \tag{4.4}$$

can be generated from the Hartree-Fock state by applying the associated annihilation operator, thus reducing to a single term. However, the cationic states can be determined

through variational principle by solving

$$H^{N-1}C = E^{N-1}C \tag{4.5}$$

where E^{N-1} is the the energy of the cationic system and the matrix elements of the Hamiltonian operator given by

$$H_{ij}^{N-1} = \langle \Phi_{HF} | a_{i\sigma}^{\dagger} \hat{H} a_{j\sigma} | \Phi_{HF} \rangle \tag{4.6}$$

therefore, we can rewrite the matrix elements as;

$$H_{ij}^{N-1} = \langle \Phi_{HF} | a_{i\sigma}^{\dagger}, [\hat{H}, a_{j\sigma}] | \Phi_{HF} \rangle + \langle \Phi_{HF} | a_{i\sigma}^{\dagger} a_{j\sigma} \hat{H} | \Phi_{HF} \rangle \tag{4.7}$$

$$= -\langle \Phi_{HF} | [a_{i\sigma}^{\dagger}, [a_{j\sigma}, \hat{H}]]_+ | \Phi_{HF} \rangle + \delta_{ij} \langle \Phi_{HF} | \hat{H} | \Phi_{HF} \rangle \tag{4.8}$$

$$= -f_{ji} + \delta_{ij} E_{HF}^{N} \tag{4.9}$$

where f_{ji} are the elements of the Fock matrix given as;[88]

$$f_{ji} = \frac{1}{2} \sum_{\sigma} \langle cs | [a_{i\sigma}^{\dagger}, [a_{j\sigma}, \hat{H}]]_+ | cs \rangle \tag{4.10}$$

for a closed-shell electronic state $|cs\rangle$. Thus we can set up the cationic eigenvalue equations and solve them in any basis that satisfy the variational conditions for the neutral system and the cationic (eigenvalue) energy would be given by;

$$E_i^{N-1} = E_{HF}^{N} - \varepsilon_i \tag{4.11}$$

By rearranging the above equation, we can define the ionization potential (IP) of a closed-shell molecular system at Koopmans' theorem as the negative orbital energies of the canonical orbitals;

$$IP_i^{KT} = E_i^{N-1} - E_{HF}^{N} = -\varepsilon_i \tag{4.12}$$

We can follow the same procedure and create an electron on the virtual orbital, and then find the solution for the corresponding eigenvalue equations. The anionic wavefunction will be

$$|N+1\rangle = \sum_a C_a a_{a\sigma}^{\dagger} | \Phi_{HF} \rangle \tag{4.13}$$

and the anionic state after application of associated creation operators would be

$$|N+1\rangle_{a\sigma} = a_{a\sigma}^{\dagger} | \Phi_{HF} \rangle \tag{4.14}$$

The resulting Hamiltonian matrix elements of the Hamiltonian operator will look like

$$H_{ab}^{N+1} = f_{ab} + \delta_{ab} E_{HF}^N \qquad (4.15)$$

with the anionic energy given by

$$E_a^{N+1} = E_{HF}^N + \varepsilon_a \qquad (4.16)$$

So, the electron affinities (EA) of a closed shell molecular system at Koopmans' level is given by the negative orbital energies of the canonical orbitals

$$EA_a^{KT} = E_{HF}^N - E_a^{N+1} = -\varepsilon_a \qquad (4.17)$$

The Koopmans' theorem states that for a given N-electron Hartree-Fock single determinant $|N\rangle$ with occupied and virtual spin orbital energies ε_i and ε_a, the energy required to produce $(N-1)-$electron single determinant $|N-1\rangle$ is equal to $-\varepsilon_i$, and $-\varepsilon_a$ to produce $(N+1)-$electron single determinant $|N+1\rangle$, all having identical spin orbitals with those of the $N-$electron system. This approximation assumes that there is no spin orbital relaxation in $(N\pm1)-$electron states. This "frozen orbital" approximations tends to produce too positive an ionisation potential or too negative an electron affinity and the correction to this is from allowing orbital relaxation and correlation effects one obtains when going beyond the Hartree-Fock approximations. Orbital relaxation accounts for the majority of these correction and it is reproduced at the Hartree-Fock level. Although the correlation effects are large for systems with large number of electrons, this cancels the relaxation error for ionisation potentials but *vice-versa* for electron affinities, thus the Koopmans' ionisation potentials might be good approximation to the experimental ionisation potential.

4.2 Electronic Affinities

It is well established that structures with dipole moments above ca. 1.625 D can bind an extra electron primarily due to the electrostatic dipole potential of the underlying neutral system[128–131]. The binding energy associated with the excess electron requires size extensive methods if estimated as the difference between the optimal energies of the neutral and the anion. Such problem can be addressed through the use of the second order Moller-Plesset (MP2)[132] perturbation theory, the coupled-cluster method with singles and doubles and non-iterative triples (CCSD(T))[94], etc. and the electron binding energies are analysed within a perturbative framework designed for dipole-bound

anions and the solvated electrons. The electron binding energies are normally estimated at Koopmans theorem with and without orbital relaxation and the electron correlation contributions are also estimated at these level of theory. Consider a neutral molecule (M) weakly interacting with the loosely bound electron (lbe), the total Hamiltonian can be described as:

$$H = H^0 + \lambda W^M + \eta V^{lbe} \tag{4.18}$$

λ and η have physical values equal to unity and they define the order of perturbation theory. The orbitals of the M are used in the determination of the Fork operator and the ground state (*zeroth-order*) Hamiltonian is estimated as the sum of Fork operators for all electrons in the anion.

$$H^0 = F^M + F^{lbe} \tag{4.19}$$

with H^0 corresponding to the Hartree Fork level of theory for the M and KT level of theory for the lbe and W^M and V^{lbe} respectively are the fluctuation operators for the neutral molecule and the lbe. The fluctuation-interaction operator V^{lbe} is expressed in terms of the coulomb (J_M) and the exchange (K_M) operators for M.

$$V^{lbe} = \sum_{i \varepsilon M} \frac{1}{r_{lbe,i}} - (J_M(lbe) - K_M(lbe)) \tag{4.20}$$

where $r_{lbe,i}$ is the distance between the *ith* electron of the neutral and the lbe. The energies of the anions are determined by applying double perturbation on equation 4.18 and the expansion should yield:

$$E = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \varepsilon^{(kl)} \tag{4.21}$$

with $\varepsilon^{(kl)}$ as the kth order in W^M and l^{th} order in V^{lbe}. If we neglect both orbital relaxation and electron correlation effects, the sum of the three lowest-order terms reproduces the SCF energy of M and the binding energy (E^{KT}) at Koopmans theory is given by.

$$\varepsilon^{(00)} + \varepsilon^{(10)} + \varepsilon^{(01)} = E_M^{SCF} - E_{bind}^{KT} \tag{4.22}$$

If we account for the polarization of the neutral by excess electron, the effect of back polarization, the orbital relaxation effect and the induction effects, then we can estimate the induced SCF binding energy from the difference in the SCF energies of the neutral and the anionic species

$$\Delta E_{ind}^{SCF} = E_{bind}^{SCF} - E_{KT}^{SCF} \approx -\varepsilon_{ind}^{(02)} \tag{4.23}$$

where,

$$E_{bind}^{SCF} = E_M^{SCF} - E_A^{SCF} \tag{4.24}$$

$$\varepsilon_{ind}^{(02)} = \varepsilon^{(02)} - \varepsilon_{disp}^{(02)} \tag{4.25}$$

where $\varepsilon_{disp}^{(02)}$ is the dynamic correlation between the *lbe* and the electron of the M and it is approximated by $\Delta E_{bind}^{MP2-disp}$ which takes proper permutational symmetry for all electrons in the anion. We therefore approximate the dynamic correlation as:

$$\varepsilon_{disp}^{(02)} \approx \sum_{a\varepsilon M} \sum_{r<s} \frac{|\langle \phi_a \phi_{lbe} \| \phi_r \phi_s \rangle|^2}{e_a + e_{lbe} - e_r - e_s} = -\Delta E_{disp}^{MP2} \tag{4.26}$$

The terms, ϕ_a and ϕ_{lbe} are the spin orbitals occupied in the ground state wave function, ϕ_r and ϕ_s are unoccupied spin orbitals, e_r and e_s are the corresponding orbital energies. The total MP2 contribution to the binding energy is described in terms of dispersion and non-dispersion terms:

$$\Delta E_{bind}^{MP2} = E_{bind}^{MP2} - E_{bind}^{SCF} = E_{bind}^{MP2(disp)} + \Delta E_{bind}^{MP2(no-disp)} \tag{4.27}$$

The latter expression is dominated by the correlation to the static coulomb interaction between the *lbe* and the charge distribution of M. The overall contributions to higher order MP contributions can be estimated as;

$$\Delta E_{bind}^{MPn} = E_{bind}^{MPn} - E_{bind}^{MP(n-1)}, \qquad n = 3, 4 \tag{4.28}$$

The MP4 results can be subtracted from the CCSD results in order to estimate the correlation contributions beyond the fourth-order. This is estimated by subtracting the DQ, SDQ and SDTQ MP4 energies from the D, SD and SD(T) coupled cluster binding energies, respectively.

$$\Delta E_{bind}^{CC} = E_{bind}^{CC} - E_{bind}^{MP4} \tag{4.29}$$

4.2.1 Vertical Attachment/Detachment Energies

Suppose we start from a neutral structure M defined by the geometry G. The energy of the anion M^- at a geometry $\mathbf{G}$ can be written as;

$$E_{M^-}(G) = E_M(G_M) + \Delta E_M(G) - EBE(G) \tag{4.30}$$

where $E_M(G_M)$ is the energy of the neutral M at its optimal geometry G_M,

$$\Delta E_M(G) = E_M(G) - E_M(G_M) \tag{4.31}$$
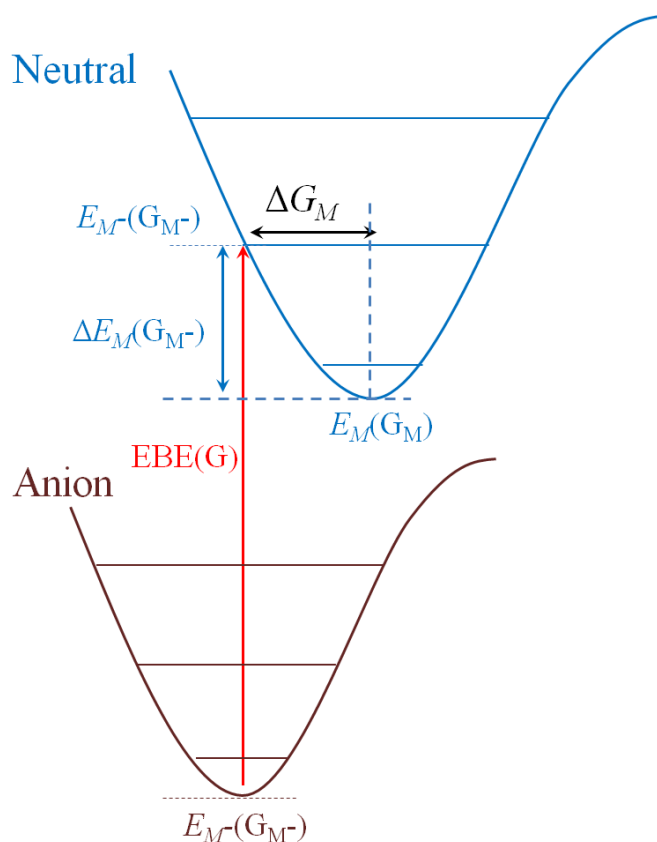
FIGURE 4.1: *Illustration of the adiabatic electron affinity given by equation 4.34.*

represent an increase of the energy of the neutral M associated with its geometrical deformation from G_M to G, and $EBE(G)$ is the vertical electron binding energy at the geometry G;

$$EBE(G) = E_M(G) - E_{M^-}(G) \tag{4.32}$$

The values of EBE are positive for vertically bound anionic states and negative for resonances. The electron vertical attachment energy (VAE) is equivalent to $EBE(G_M)$. The electron vertical detachment energy (VDE) is equivalent to $EBE(G_{M^-})$ and the adiabatic electron affinity is defined as;

$$AEA = E_M(G_M) - E_{M^-}(G_{M^-}) \tag{4.33}$$

where G_{M^-} is the optimal geometry on the anion of M. Notice that;

$$AEA = -\Delta E_M(G_{M^-}) + EBE(G_{M^-}) = -\Delta E_M(G_{M^-}) + VDE \tag{4.34}$$

This equation is illustrated in figure 4.1.

Chapter 5

Discovery of Most Stable Structures of Neutral and Anionic Phenylalanine through Automated Scanning of Tautomeric and Conformational Spaces

5.1 Abstract

I have developed a software tool for combinatorial generation of tautomers and conformers of small molecules and demonstrated it by performing a systematic search for the most stable structures of neutral and anionic phenylalanine (Phe) using electronic structure methods. For the neutral canonical tautomer I found out that the conformers *with* and *without* the intramolecular $(O)H \cdots NH_2$ hydrogen bond are similarly stable, within the error bars of my method. A unique IR signature of the conformer without the hydrogen bond has been identified. I also considered anions of Phe, both valence and dipole-bound type. I have found out that tautomers resulting from proton transfer from the carboxylic OH to the phenyl ring do support valence anions that are vertically strongly bound, with electron vertical detachment energies (VDE) in a range of 3.2-3.5 eV. The most stable conformer of these valence anions remains adiabatically unbound

with respect to the canonical neutral by only 2.17 kcal/mol at the CCSD(T)/aug-cc-pVDZ level. Based on past efforts with valence anions of nucleic acid bases, I suggest that the valence anions of Phe identified in this report can be observed experimentally. The most stable conformer of canonical Phe is characterized by an adiabatic electron affinity of 53 meV (a dipole-bound state).

5.2 Introduction

Recent years brought significant advances in non-destructive transfer of low vapor pressure compounds to the gas phase.[133–136] The resulting intact molecules are probed with different spectroscopic methods and an interpretation of these experimental results frequently invokes comparisons with computed spectroscopic characteristics.[137] These characteristics can be routinely determined for molecules with a few tens of atoms at density functional or second-order Møller-Plesset (MP2) theory levels. The only caveat is that the gas phase molecular structure has to be known.

Information about gas phase structure is frequently limited even for neutral molecules. Computational chemists are tempted to use condensed-phase structures (primarily from crystallographic data bases), but these structures might be incorrect in gas phase conditions. The reason is that intermolecular interactions in crystalline lattices might favor conformers, or even tautomers, different from the most stable gas phase structures.[18, 25, 138, 139] It has also been recognized that solute-solvent interaction can favor specific conformers[140] and tautomers, e.g., zwitterionic amino acids in polar solutions.[141] Charged molecular systems, e.g., ions of nucleic acid bases, bring additional challenges as they might favor unexpected tautomers.[142–144]

There are brute force methods based on systematic scanning the potential energy surface (PES). Some of these methods specialize in molecular tautomers[21, 22, 85] and others in conformers,[24–26] but there is no tool that would deal with tautomers and conformers on equal footing. In the past, my group developed a tool that performs a combinatorial/computational search for the most stable tautomers,[21] and the tool was applied to the anions of nucleic acid bases.[27] More recently, a tool for a combinatorial/computational search of the most stable conformers was developed and it was applied to nucleosides.[25] Here I discuss capabilities of my newly developed Potential Energy Surface Scanning Tool (PESST) that searches *simultaneously* for the most stable tautomers and conformers[28] and I demonstrate its usefulness on the neutral and anionic phenylalanine, see Figure 5.1.
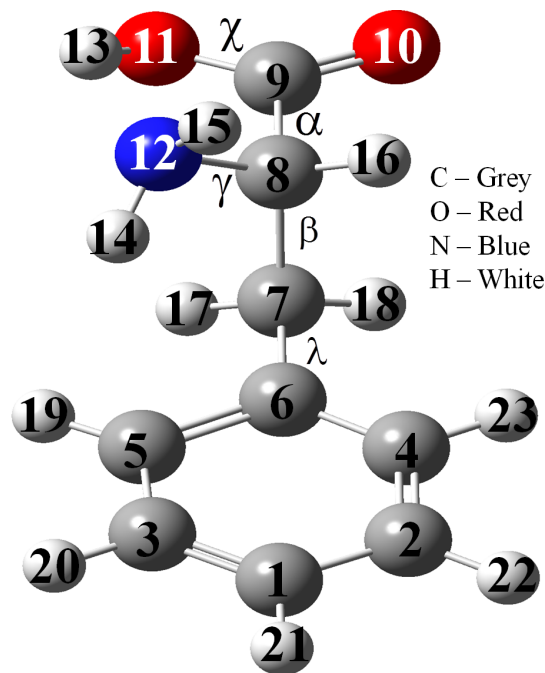
FIGURE 5.1: *Molecular structure of phenylalanine with labelled atoms and rotatable bonds.*

Valence anions of molecular systems are more challenging for structural determination than their neutral counterparts. They might favour tautomers that would be quite unstable for the neutral system. For example, past studies of valence anions of nucleic acid bases demonstrated that the most stable anionic tautomers result from proton transfer from a typical proton donor (e.g., NH) to carbon sites, rather than to conventional proton acceptors involving highly electronegative atoms (O or N).[142–144] The valence anions observed experimentally were characterized by significant electron vertical detachment energies (VDE) of 1-3 eV, though computational results demonstrated that these anionic tautomers might be adiabatically unbound by a few kcal/mol with respect to the most stable conventional neutral tautomers.[27] Apparently, the anionic tautomers are separated by sufficiently high barriers from these regions of PES where the valence anion is weakly bound or even unbound. Notice that the six-member ring of phenylalanine might support unusual tautomers of its valence anion, which would result from proton transfer from the carboxylic group to the phenyl ring. Identification of these tautomers and their most stable conformers is the main goal of this project.

Amino acids typically take the canonical HOOC-CHR-NH$_2$ form in the gas phase.[145] However, an excess electron bound by the dipolar potential of a neutral amino acid can promote intramolecular proton transfer and is responsible for the development of a zwitterionic minimum, which remains local in the case of the anion of glycine,[146] but becomes global in the case of the anion of arginine.[147, 148] Further stabilisation of these anions by water molecules has also been considered.[149, 150] In this report I

will search for dipole-bound anions of phenylalanine associated with its canonical and zwitterionic forms.

L-Phenylalanine (Phe) is one of the twenty standard amino acids found in nature. It is an essential, nonpolar, α-amino acid and is a precursor for another amino acid, tyrosine. Previous studies on Phe focused on electronic spectra using fluorescence spectroscopy in a supersonic jet and five conformers of Phe were stabilised in low-temperatures.[151] The study was extended by Meijer and coworkers,[152] using mid- and far infra-red spectra, who reported six conformers in the gas phase. Their work showed good agreement between the experimental spectrum and results computed at the B3LYP/6-311++G(2d,p) and MP2/6-311+G(2df,2p) levels of theory, with vibrational frequencies scaled uniformly by 0.98. Fausto and coworkers[153] reported a matrix-isolation IR spectrum and concluded that entropic effects play a major role in the relative abundance of conformers. Based on their B3LYP/6-311++G(d,p) results they confirmed that only six structures contribute to the spectrum of Phe.

Many computational studies addressed the structural preference of Phe. Purushotham *et al*[154] used the M052X functional[155] and extended basis sets to characterize the conformational space and noncovalent interactions in the neutral canonical Phe, as well as in ions resulting from its protonation and deprotonation. Lin and coworkers[156] performed a conformational search on neutral Phe at the B3LYP/6-311++G(d,p) level and characterised the intramolecular hydrogen bonding using the Atoms in Molecules approach.[157] Lee *et al* reported relative energies of the six most stable neutrals and cations of Phe at the B3LYP and MP2 levels of theory with 6-31+G* basis sets.[158] They paid particular attention to the dependence of vertical ionisation energies on intramolecular hydrogen bonding. The neutral structures were similar to those from Refs. [154, 156]. Other studies were focused on the cations resulting from protonation of Phe.[159]

In the current study I used the neutral and anionic Phe to demonstrate how the PESST tool can be used to identify stable molecular structures. The anions I have considered here result from the attachment of an excess electron, but not from the deprotonation of Phe. I considered both valence- and dipole-bound anions. The latter might be supported by polar canonical or zwitterionic structures of the neutral.

5.3   Description of PESST Program

The purpose of PESST is to construct a library of molecular structures representing various tautomers and their conformers. PESST was designed having in mind organic
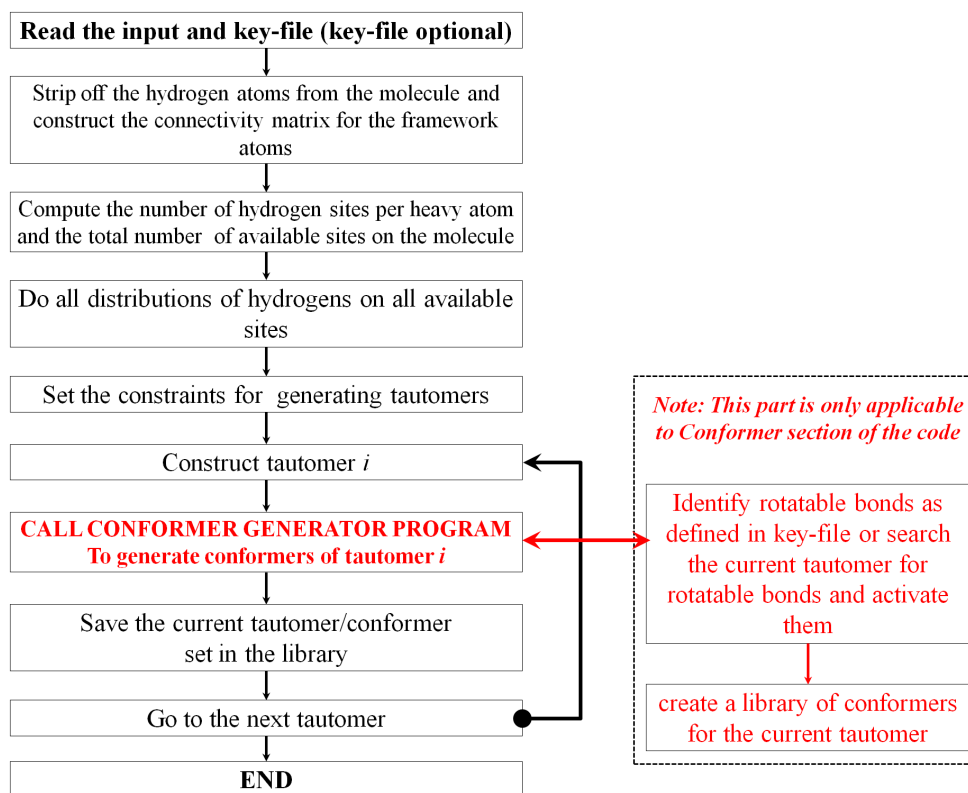
FIGURE 5.2: *The working procedure of PESST. The generation of conformers for each tautomer is marked in red.*

molecules with heteroatoms. Thus it assumes that a non-hydrogen atom can form up to four chemical bonds.

PESST reads an initial molecular structure provided by the user and is capable of creating quite a complete library of tautomers and conformers using its default settings. The user, however, would typically like to impose some constraints on the construction of tautomers, specify rotatable bonds and the magnitude of an angular step, with which each rotatable bond will be probed. This information is provided by the user in an optional "key-file". A library of molecular structures representing acceptable tautomers and conformers is then created.

Each constructed structure is used as an initial structure for the following geometry optimization at a level of theory specified by the user (a force field or an electronic structure model). The pre-screening procedure is automated to a large extent and involves creation of complete input files. The number of computational runs might be quite significant, therefore I suggest selecting an appropriate level of theory and using a submission script included in the PESST package.[29] Output files are analysed using Gaussian Output Tools (GOT) to extract converged geometries, corresponding energies, dipole moments, etc.[160] The GOT scripts arrange converged structures according to their increasing energy. One should keep in mind that different initial structures might

	C1	C2	C3	C4	C5	C6	C7	C8	C9	O10	O11	N12	H13	H14	H15	H16	H17	H18	H19	H20	H21	H22	H23	
C1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	C1
C2	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	C2
C3	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	C3
C4	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	C4
C5	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	C5
C6	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	C6
C7	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	C7
C8	0	0	0	0	0	0	**0**	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	C8
C9	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	C9
O10	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	O10
O11	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	O11
N12	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	N12
H13	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	H13
H14	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	H14
H15	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	H15
H16	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H16
H17	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H17
H18	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H18
H19	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H19
H20	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H20
H21	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H21
H22	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H22
H23	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H23
	C1	C2	C3	C4	C5	C6	C7	C8	C9	O10	O11	N12	H13	H14	H15	H16	H17	H18	H19	H20	H21	H22	H23	

Disconnected bond set to zero

FIGURE 5.3: *Connectivity matrix of the phenylalanine molecule after disconnecting atoms **C7** and **C8** (zero in bold). The fragment connected to **C8**, which is to be rotated, is highlighted in red with the remaining fragment connected to **C7** in blue.*

have converged to the same final structure and the resulting library of pre-screened structures might be redundant. It can be easily spotted due to the similarity of converged energies. Thus I recommend inspecting visually the similarly of the stable structures, using e.g., Molden[161], and eliminating redundant structures. The converged geometries represent refined molecular structures and the fittest (most stable) structures are typically further examined using a more advanced theoretical model than the one invoked at the pre-screening stage.

5.3.1 Analysis of Initial Geometry

The program reads in the total number of atoms NA in the considered molecule followed by the list of atoms with their Cartesian coordinates. Next it counts the number of hydrogen atoms (NHA) in the molecule and strips them off. The number of atoms

in the remaining framework (NFA) is also determined, as well as distances between all pairs of atoms. A symmetric connectivity matrix is then constructed based on the distance matrix and the atomic radii of the atoms adopted by the Cambridge Structural Database[5], i.e., each atom is assigned an atomic radius R_{cov}. The values of $R_{cov}(i)$ are retrieved from a data file, which can be amended by the user. Two framework atoms are connected if and only if their separation distance R_{ij} satisfies the criterion:

$$R_{ij} < R_{cov}(i) + R_{cov}(j) + T \tag{5.1}$$

where T is a tolerance which I set to 0.4 Å. The connectivity matrix for Phe is illustrated in Figure 5.3, with 1's and 0's for the connected an unconnected pairs, respectively.

The atom type determines the maximum number of single bonds ($NMAX[i]$) the atom i can form, e.g., 4 for C or N, and this information is retrieved from a data file, which can be amended by the user. Next, the program counts the number of framework atoms already connected to each framework atom $i(NBOND[i])$. The maximum number of hydrogen atoms that can be connected to the atom i is given by:

$$NHMAX[i] = NMAX[i] - NBOND[i] \qquad (i = 1, \ldots, NFA) \tag{5.2}$$

See Figure 5.4 for an illustration of the $NMAX, NBOND$, and $NHMAX$ values.

The total number of slots available to the NHA hydrogen atoms is:

$$NSLOTS = \sum_i NHMAX[i] \tag{5.3}$$

An upper limit to the total number of created tautomers is given by a binomial coefficient $\binom{NSLOTS}{NHA}$, i.e., the total number of distributions of NHA atoms over $NSLOTS$. The binomial coefficient is the rigorous number of created tautomers for molecules in which any framework atom can bind one hydrogen atom only ($NHMAX[i] = 1$ for any i).For molecules in which framework atoms can bind more than one hydrogen atom the total number of tautomers is significantly reduced, *vide infra*. The number of tautomers can be further reduced by imposing constraints specified in the key-file.

5.3.2 Constraining the Number of Tautomers

The key-file is used to provide additional control of the program. Here I will discuss constraints imposed on the library of tautomers. The keywords are **R**, for **RANGE**, **XC** for **EXCLUDE**, **NoSI** for **NO STEREO-ISOMERS**. Each keyword is preceded by the number of framework atoms it applies to (Figure 5.5). The list of atoms and
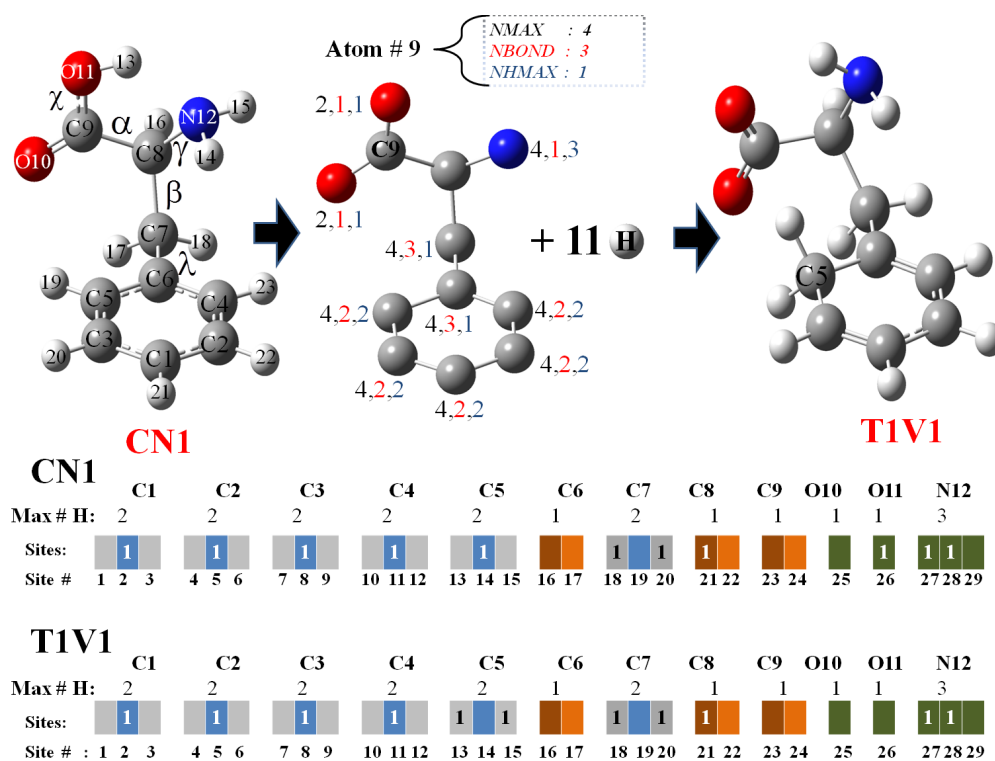
FIGURE 5.4: *Molecular structure of phenalalanine and a set of sites for hydrogen atoms. The grey-blue for atoms with $NMAX = 4$ and $NHMAX = 2$. The blue and grey sites are occupied when one and two hydrogen atoms, respectively, are attached. The $NMAX = 4, NBOND = 3$ cases are labelled in dark and light brown. They are above and below the molecular framework and only one of them can be occupied for a given tautomer. The terminal framework atoms have sites labelled in green and they can accommodate 0-3 hydrogens.*

specific constraints follow. For example, if I consider **"RANGE"** as my input keyword, then the user should list each atom followed by the minimum and maximum number of hydrogen atoms that could be attached to this framework atom. For the **"EXCLUDE"** keyword, the label of a framework atom (m) is followed by one integer (n) indicating a number of hydrogen atoms. The program will exclude all tautomers having n hydrogen atoms attached to the mth framework atom.

Special care is taken of these framework atoms that are already connected to three others. Such an atom can accommodate one hydrogen atom only, either above or below the plane spanned by the framework atoms. The user might conclude that occupation of either of these sites would lead to equivalent chemical structures and then an elimination of one of them is desirable. This is what the **NoSI** keyword provides.

An example of the key-file is given in Figure 5.5 and the labelling of atoms of Phe is the same as in Figure 5.1. The 12th framework atom (nitrogen) is connected to one other framework atom. By default, it can accommodate 0-3 hydrogen atoms. By using the R keyword followed by 12 1 3 I accept only those tautomers that have 1-3 hydrogen

FIGURE 5.5: *Illustration of the key-file used to control the program. This file is optional.*

atoms attached to atom 12. The same result can be achieved using the **XC** keyword followed by 12 0. The **NoSI** keyword is applied to the 6th atom (C), a framework atom connected to three other framework atoms.

5.3.3   Determining Sites for Hydrogen Atoms

The procedure of determining positions of hydrogen atoms around a framework atom i depends on the value of $NBOND[i]$. The case of $NBOND[i] = 2$ is probably the most common and will be illustrated for **C7**, which is connected to **C6** and **C8** (Figure 5.6,a). The **C7** site can attach one or two hydrogens and their sites will be labelled **H1** and **H2, H3**, respectively. A unit vector attached to **C7** and parallel to **C7-C8** is defined. Its end defines a point **A**. Similarly, a unit vector attached to **C7** and parallel to **C7-C6** defines a point **B**. A point **C** is obtained by bisecting **A-B**. Lastly, a vector $\vec{C}$ is defined by **C-C7**. When translated to **C7** and scaled in length to the sum of atomic radii of C and H, its end defines **H1**. In case of two hydrogens attached to **C7**, the above procedure is extended further. I labelled the **A-B** vector as $\vec{D}$ and I translated it to **C7**. The site **H1** is then rotated by 54.5° about and then by −54.5° to give the sites **H2** and **H3**, respectively (Figure 5.6,b). Here I used standard equations to rotate
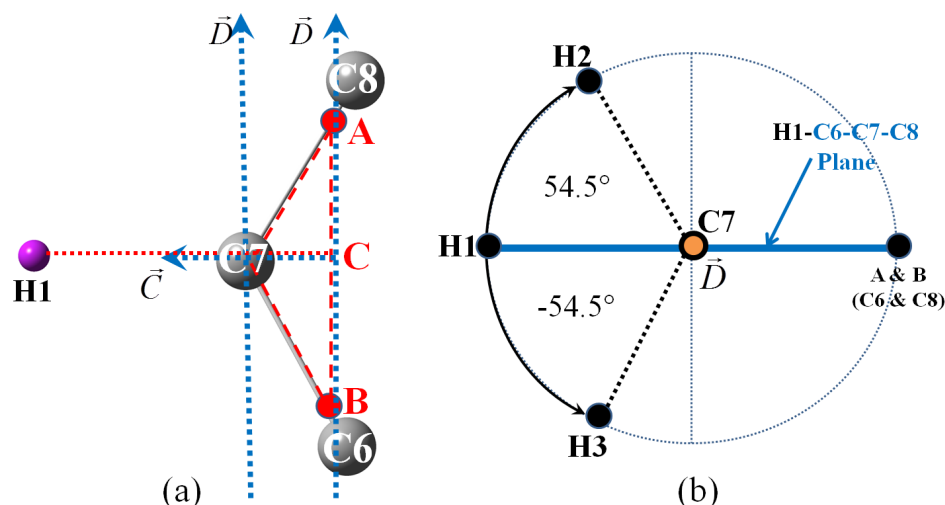
FIGURE 5.6: *Generation of hydrogen sites for an atom connected to two other framework atoms, here **C7** connected to **C6** and **C8**. Part (a) demonstrates the construction of **H1**. The points **A** and **B** are each unit length from atom **C7** along bonds **C7-C8** and **C7-C6**, respectively. Point **C** results from the bisection of **A-B**. Site **H1** is calculated along a line defined by **C-C7**. The length of **C7-H1** is equal to the sum of atomic radii of carbon and hydrogen. Part (b) demonstrates the construction of **H2** and **H3**; a side view along the **C6-C7-C8** plane is shown with vector going into the page through **C7**.*

a point about a vector,[162] Equations 5.4.

$$P1_x = (a(v^2 + w^2) - u(bv + cw - ux - vy - wz))(1 - \cos\theta) + x\cos\theta$$
$$+ (-cv + bw - wy + vz)\sin\theta \tag{5.4a}$$

$$P1_y = (b(u^2 + w^2) - v(au + cw - ux - vy - wz))(1 - \cos\theta) + y\cos\theta$$
$$+ (cu - aw + wx - uz)\sin\theta \tag{5.4b}$$

$$P1_z = (c(u^2 + v^2) - w(au + bv - ux - vy - wz))(1 - \cos\theta) + z\cos\theta$$
$$+ (-bu + av - vx + uy)\sin\theta \tag{5.4c}$$

Caption to Equation 5.4: There are ten variables $x, y, z, a, b, c, u, v, w$, and θ, where x, y and z are the $x-, y-, z-$coordinates of point to be rotated (*point **A** with unit length from atom **C7** or current atom*) about a line passing through atom defined by coordinates a, b and c with direction (*unit vector $\vec{B}$*) vector u, v and w (where $u^2 + v^2 + w^2 = 1$). Variables a, b and c are simply the $x-, y-, z-$coordinates of the reference atom (*atom **C7***) and lastly the angle of rotation θ, in radians.

The site defined by **H1** is assigned a blue colour in Figure 5.4, whereas the sites **H2** and **H3** are assigned a grey colour. The same procedure is applicable to atoms in the ring with $NBOND[i] = 2$.

Next I considered a terminal framework atom, i.e., $NBOND[i] = 1$, and the procedure will be illustrated for **O11** (Figure 5.7,a). I defined unit vectors $\vec{E}$ and $\vec{F}$ along **C9-O11**
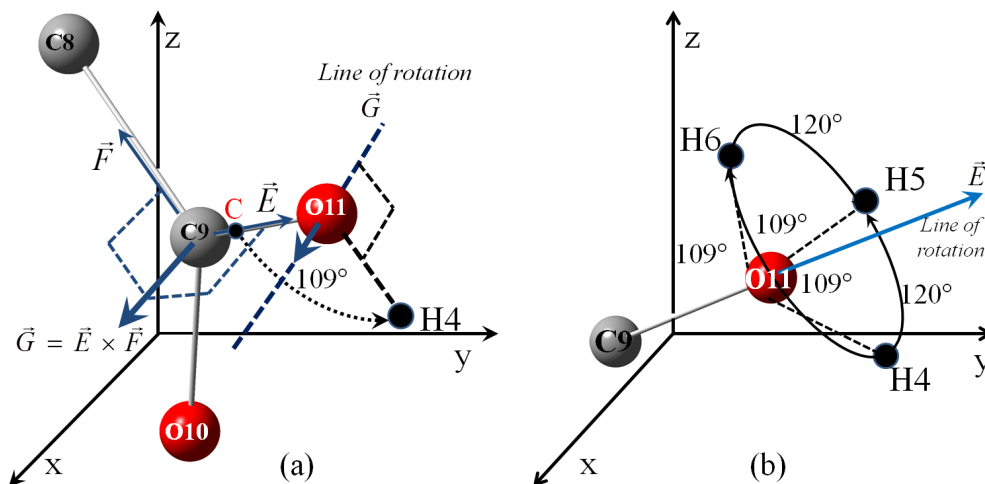
FIGURE 5.7: *Generation of hydrogen sites for a terminal framework atom, here **O11**. A procedure to define a site for (a) the first hydrogen atom (**H4**) and (b) the second and third hydrogen atom (**H5** and **H6**) is illustrated.*

and **C9-C8**, respectively. The cross product of $\vec{E}$ and $\vec{F}$ yields a vector $\vec{G}$, which is translated such that it passes through **O11**. The point **C** which results from scaling the bond **O11-C9** to the desired OH bond length is rotated by 109° about the translated vector to give a point **H4**. The dihedral angle **C8-C9-O11-H4** is set to 180° and the coordinates of **H4** are assigned to the first hydrogen atom. In case of two or three hydrogens attached to the terminal atom, the additional sites **H5** and **H6** are defined by rotating **H4** by 120° and then by −120° about the vector $\vec{E}$ (Figure 5.7,b). The hydrogen sites on a terminal atom were assigned a green colour in Figure 5.4.

The case of $NBOND[i] = 3$ will be illustrated for **C6**, which is connected to **C7, C4**, and **C5** (Figure 5.8). Two sites for hydrogen atoms will be created, **H7** and **H8**. First, unit vectors are calculated along the three bonds **C6-C7, C6-C5** and **C6-C4**. Next, an average vector is constructed from three cross products illustrated in Figure 5.8 and its length is scaled to the desired length of a CH bond yielding **H7**. The reverse vector defines **H8**, which is generated by default, but can be eliminated with the **NoSI** keyword applied to **C6**. The sites **H7** and **H8** are assigned a brown colour in Figure 5.4.

The hydrogen sites, illustrated in the bottom part of Figure 5.4, can be occupied in many ways representing various tautomers. The process of creation of tautomers is illustrated in this Figure using **CN1** and **T1V** as an example.

5.3.4   Conformers

PESST can create conformers for each tautomer considered. The procedure is an extension of past approach described in Ref.[25]. Here, rotations of molecular fragments are performed using Equation 5.4. Firstly, the code identifies rotatable bonds. The bond

$$C6 - H7 = N(\vec{A} \times \vec{B} + \vec{B} \times \vec{C} + \vec{C} \times \vec{A})$$

$$C6 - H8 = -(C6 - H7)$$
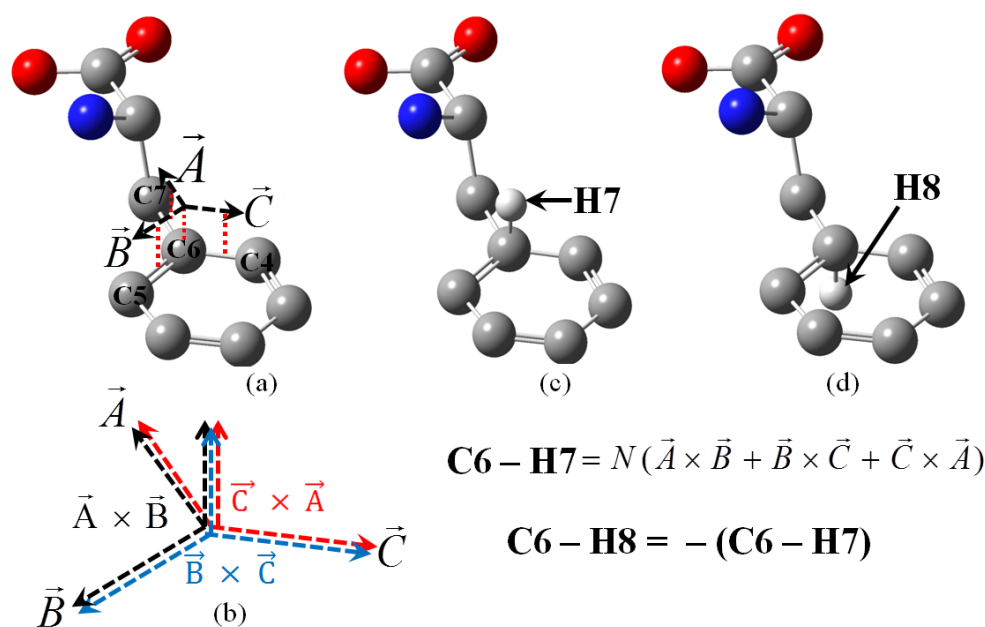
FIGURE 5.8: *Generation of hydrogen sites for an atom connected to three other framework atoms, here **C6** connected to **C4**, **C5** and **C7**. Part (a) construction of three unit vectors along the framework bonds; (b and c) three cross products of unit vectors are averaged and the resulting vector is scaled in length to the sum of proper atomic radii to define the **H7** site (above the ring). (d) The reverse yields a site on the other side of the ring (**H8**).*

being checked for rotation is disconnected on the connectivity matrix, the program proceeds by counting the number of atoms in both fragments and identifies a smaller branch (a red branch in Figure 5.3). If there is only one atom returned, then the bond is the terminal bond. If the number of atoms returned is equal to the number of atoms in the molecule, then the bond is within the ring. If the number of atoms returned is not equal to one or the total number of atoms, then the bond is rotatable and the smaller fragment will be rotated, see Figure 5.9, with increments of $360°/n_{rot}$, where n_{rot} is an integer defined for each rotatable bond.

The user can control the construction of conformers by specifying the **CONF** keyword in the key-file followed by a list of bonds and their n_{rot} values, see Figure 5.5. The code will construct conformers associated with these bonds, while all other conformers will be ignored. If the **CONF** keyword is missing the code will activate all rotatable bonds for each tautomer and the default values of n_{rot} will be assumed, e.g., $n_{rot} = 2$ for X–OH bonds, $n_{rot} = 3$ for X–NH$_2$, and $n_{rot} = 6$ for C–C bonds. An upper limit to the total number of created conformers, n_{total}, is a product of n_{rot} values for all activated bonds:

$$n_{total} = \prod_{i}^{active\ bonds} n_{rot,i} \tag{5.5}$$
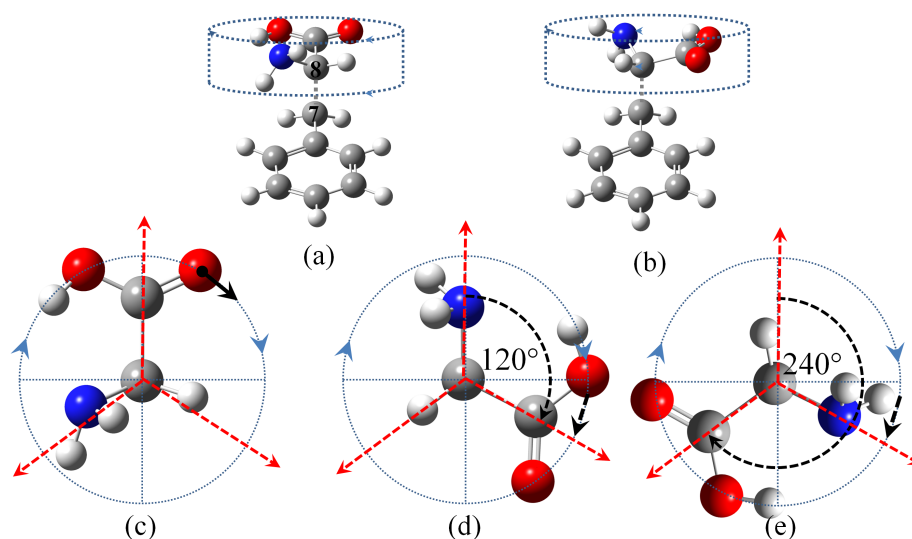
FIGURE 5.9: *(a) The bond to be rotated is defined by **C7C8**. The smaller fragment involves atoms connected to **C8** (see Figure 5.3). (b) a conformer resulting from rotation by 120°. (c)-(e) are top views along **C7C8** for rotations by 0°, 120°, and 240°.*

The value of n_{total} increases rapidly with the size of the molecule[25]. Therefore I strongly recommend the use of **CONF** keyword and specifying rotatable bonds suitable for the chemical problem studied.

The code identifies potential clashes that might develop between various parts of the molecule in the course of systematic rotations. If the distance between any two atoms is less than the sum of their atomic radii minus T from Equation 5.1, then the conformer is discarded, otherwise stored in the library. The construction of conformers in PESST is illustrated in Figure 5.10.

5.4 Computational Approach

The libraries involving tautomers and conformers of Phe were generated with PESST.[29] An overview of PESST is shown in Figure 5.2. It creates conformers for each tautomer resulting in a double-loop structure of the programme. The pre-screening of initial structures for the neutral and valence anions of Phe was conducted through geometry optimizations with the B3LYP exchange-correlation functional[100, 111–113] and standard Poples basis sets.[163] This selection of methodology for the pre-screening step was validated and used in previous studies focused on identification of most stable tautomers of small biomolecules such as nucleosides.[142–144] However, these basis sets and electronic structure model are not sufficient for dipole-bound states.[164] Electronic structure calculations for dipole-bound anions require basis sets with very diffuse basis functions.[122, 129] Here I have used Dunnings aug-cc-pVDZ[121](ADZ) basis set
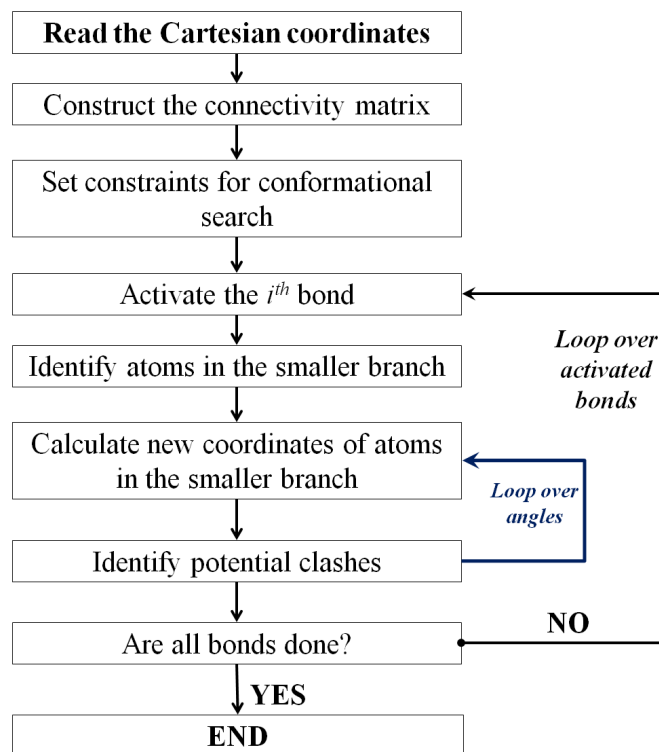
FIGURE 5.10: *Construction of conformers with PESST.*

supplemented with extra 7 $s-$ and 7 $p-$functions centered on the nitrogen atom. The exponents of these functions form an even-tempered series initiated from the lowest exponent in the standard basis set and advance with the geometric progression constant of 3.2^{-1}.[165] I will use a label DF for these additional diffuse functions and ADZ+DF for the combined basis set. Pre-screening for dipole-bound anions was performed at the Hartree-Fock/ADZ+DF level.

For the neutral Phe I identified only canonical tautomers, the zwitterions did not support local minima. The conformational search was performed by activating dihedral angles $\chi, \alpha, \beta, \gamma$ and λ illustrated in Figure 5.1. The n_{rot} values were $n_\chi = 2, n_\alpha = 12, n_\beta = 12, n_\gamma = 3$ and lastly λ was probed at $0°$ and $90°$. PESST created 1728 initial structures which were fully optimised at the B3LYP/6-311G* level and the resulting optimised structures converged into 45 local minima. The three most stable conformers were then subjected to further MP2/ADZ+DF reoptimization and frequency calculations followed by single-point coupled cluster calculations with single, double and non-iterative triple excitations (CCSD(T)) on the MP2 optimised geometry.[166]

In the case of valence anions I engaged PESST to construct a library of tautomers and conformers. The constraints set on the program are shown in Table 5.1. It created 50 tautomers that were subjected to the B3LYP/6-31G* conformational search based on

Constraints keyword	Atoms No.	Number of hydrogens	
NoSI	6,8	N/A	
XC	8	0	
	9,10	1	
		Minimum	Maximum
R	1	0	2
	2,4	1	2
	3,5	0	1
	7	2	2

TABLE 5.1: The constraints set when generating valence anions of Phe

all rotatable bonds and the default settings for n_{rot}'s. The most stable tautomers/conformers within the energy range of 5 kcal/mol were reoptimised and frequencies were determined at the MP2/ADZ level. Three families of valence anions have been identified. The single point CCSD(T)/ADZ calculations have been performed for the most stable conformer from each family.

Search for the most stable dipole-bound (**DB**) anionic conformers was performed for the canonical (**CDB**) and zwitterionic tautomers (**ZDB**). The conformational search for **CDB**'s was performed with the following values of $n_{rot} : n_\chi = 2, n_\alpha = 6, n_\beta = 6, n_\gamma = 3$ and λ was probed at $0°$ and $90°$. These constraints resulted in 432 initial structures. The conformational search for **ZDB**'s was performed with $n_\beta = 6$ and probing γ at $0°$ and $90°$, and α at $0°$ and $90°$ and γ at $0°$ and $60°$. These constraints led to 48 initial structures. The most stable structures (three **CDB**'s and two **ZDB**'s) were subjected to the MP2/ADZ+DF optimizations and frequency calculations and vertical and adiabatic electron binding energies were determined at the CCSD(T)/ADZ+DF level.

To summarize, my final results are based on CCSD(T) energies determined at the optimal MP2 geometries. The basis set is ADZ for valence anions and ADZ+DF for neutrals and dipole-bound anions. The zero-point vibrational corrections and thermal contributions to energy and entropy are determined in the rigid rotor – harmonic oscillator approximation, assuming $T = 298\ K$ and $p = 1$ atm, based on the MP2 geometries and harmonic frequencies. The calculations were performed using the Gaussian 09 suite of programs.[75] The singly-occupied molecular orbitals of the anions of Phe were visualized using the Visual Molecular Dynamics[167] (VMD) package and the orbital contour values were determined with the OpenCubMan[168] package. These contour values were obtained for 80% of the orbital electron density.
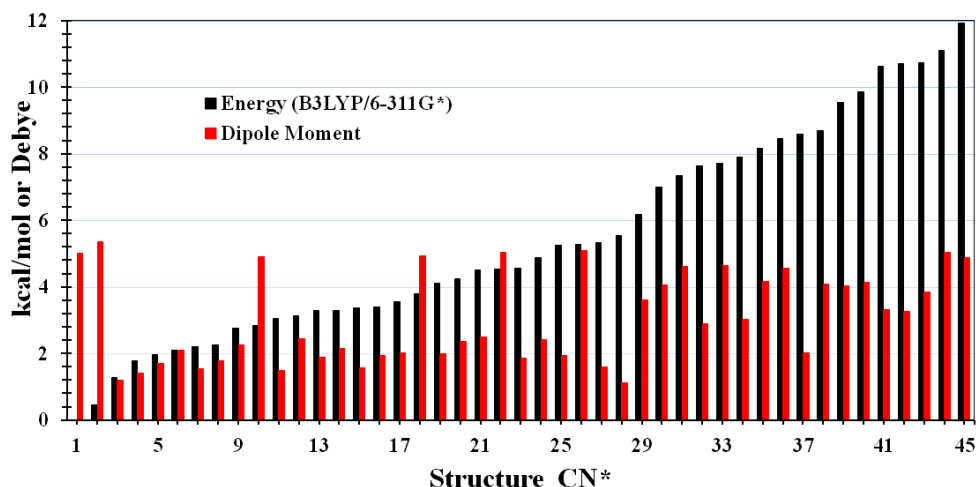
FIGURE 5.11: *Relative energies and dipole moments of Phe calculated at the B3LYP/6-311G* level of theory. The energy of **CN1** is set to zero.*

5.5 Results

5.5.1 Neutral Phenylalanine

The conformers of canonical neutral (**CN**) tautomer were labelled **CNm** ($1 \leq m \leq 45$), with ordering based on the B3LYP/6-311G* electronic energies of fully optimized structures. Their relative energies and dipole moments are illustrated in Figure 5.11. The MP2/ADZ+DF optimized structures and relative energies of the three most stable conformers are shown in Figure 5.12 (see Figures A.2 for a more complete set of conformers). At the MP2 and CCSD(T) levels of theory, the relative energies of **CN1-CN3** span ca. 2 kcal mol^{-1}. Thus determination of the most stable conformer becomes a computationally demanding task. My most accurate CCSD(T) electronic energies corrected for harmonic, zero-point, MP2 vibrational terms indicate that **CN3** is less stable than **CN1** by only 0.19 kcal/mol, and it becomes the most stable conformer in terms of Gibbs free energy. The results presented in Figure 5.12 demonstrate that entropic effects contribute to the relative stability of the most stable conformers, in agreement with earlier observations.[153]

There is an important structural difference between **CN3** and the two remaining conformers **CN1** and **CN2**, which leads to a unique feature in the IR spectrum of **CN3** (Figure 5.13). The **CN1** and **CN2** conformers support a hydrogen bond between the OH site of the carboxylic group and the NH$_2$ site, with the (O)H$\cdots$N distances of 1.862 and 1.903 Å, respectively, but the OH site in the **CN3** conformer is not engaged in any hydrogen bond (see Figure 5.12). In consequence, the OH stretching mode is strongly

System	MP2		CCSD(T)			
	μ	E_{el}	E_{el}	$E_{el}+E_0^{vib}$	H	G
CN1	4.95	0.00	0.00	0.00	0.00	0.00
CN2	5.28	2.02	1.49	1.24	1.42	0.65
CN3	1.16	0.73	0.60	0.19	0.42	-0.14
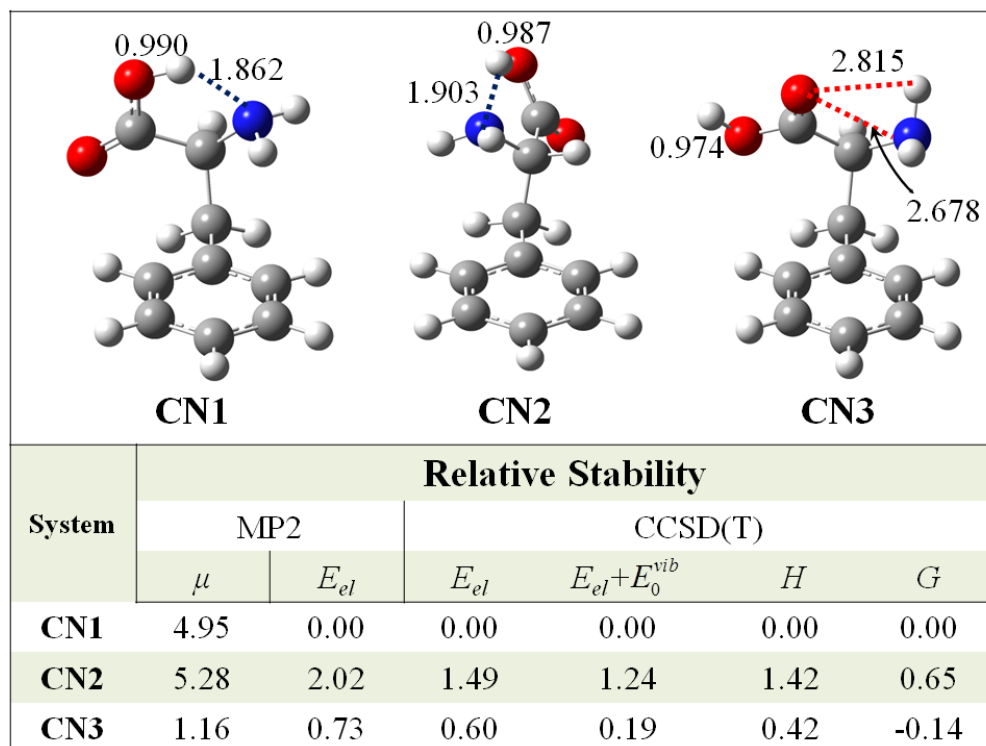
FIGURE 5.12: *The MP2/ADZ+DF structures of the three most stable conformers of the neutral canonical Phe with distances in and dipole moments in D. The intramolecular hydrogen bonds marked with dashed lines with the colour of the proton acceptor atom. The relative stability (kcal/mol) determined at the MP2 and CCSD(T) levels (E_{el}, E_0^{vib}, H, and G for electronic energy, zero-point vibrational correction, enthalpy, and Gibbs free energy, respectively). The CCSD(T) energies determined at the optimal MP2 geometries. The zero-point and thermal corrections determined at the MP2/ADZ+DF level.*

red shifted in the **CN1** and **CN2** conformers, as illustrated by the MP2 harmonic frequency of 3406 and 3463 cm^{-1}, respectively, but its frequency is much higher in the **CN3** conformer, $\tilde{v} = 3727$cm^{-1} (Figure 5.13). The OH stretching mode of the carboxylic group routinely carries the largest IR intensity and the simulated IR spectra in the 2500-4000 cm^{-1} region are presented in Figure 5.13. The most intense IR peak for the **CN1** and **CN2** conformers, i.e., the OH stretching mode, has a lower frequency than the symmetric and asymmetric stretching NH modes. On the other hand, the most intense IR peak in **CN3**, again the OH stretching mode, has the largest frequency. I believe this difference between **CN3** and the pair **CN1** and **CN2** will help to discriminate contributions from different conformers to the IR spectra of gas phase Phe recorded in different temperatures.

It was counterintuitive for me to find out that the conformer *without* the intramolecular (O)H$\cdots$N hydrogen bond, i.e., **CN3**, might be the most stable. One reason might be that formation of this hydrogen bond in **CN1** and **CN2** is associated with intramolecular strain (a destabilizing effect). The results obtained with the MM3 force field[169, 170]
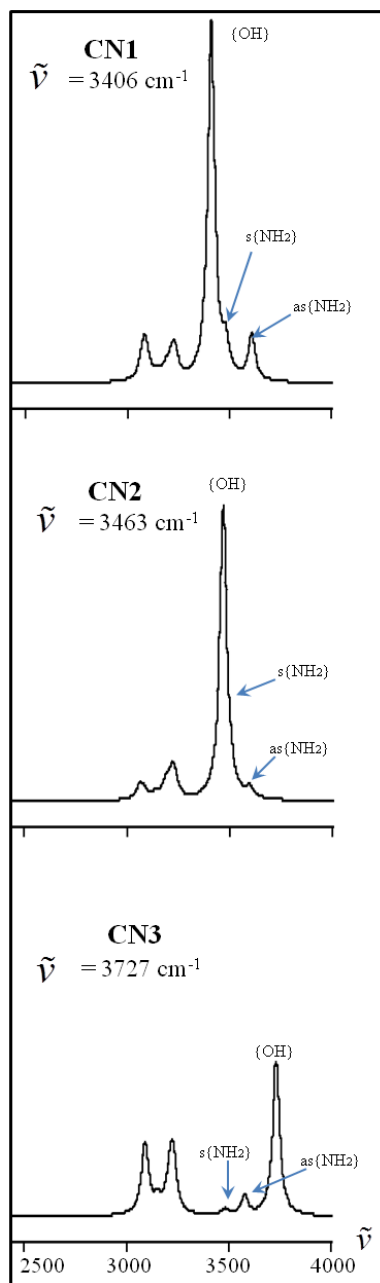
FIGURE 5.13: *The stretching modes of* **CN1, CN2** *and* **CN3** *calculated at the MP2/ADZ+DF level of theory. Keywords: s symmetric, as asymmetric. Lonrentzian vibrational spectra with scaling factor of 1.0 and half-width of* $20.0 cm^{-1}$.

confirm this hypothesis. The sum of angle bending, torsional and bond stretching terms is more attractive in **CN3** than in **CN1** by 2.8 kcal/mol.

The stability of tautomers and conformers in polar environments is strongly dependent on the molecular dipole moment. Moreover, molecules with dipole moments exceeding 1.625 D can bind an excess electron.[129] The dipole moments of 45 neutral conformers of Phe, characterised at the B3LYP/6-311G* level, lie in a wide range from 1.1 to 5.4 D (Figure 5.11). The five conformers of Phe with the largest dipole moments (in excess of 5 D) follow the order **CN2 >CN26 >CN22 >CN44 >CN1**. These results suggest that the **CN2** and **CN1** will be more stabilized in polar environments (or by a dipole-bound electron) than the **CN3** conformer, which is characterized by a dipole moment of 1.16 D only.

5.5.2   Valence Anions of Phenylalanine

I did not identify any bound valence anion when considering the canonical tautomer of Phe. The result was not surprising as bound valence anionic states are not known for either glycine or toluene, which might be considered as building blocks of Phe.[171] Using the PESST tool I explored non-canonical tautomers and their conformers. In the course of this search I identified three families of valence anions, which involve proton transfer from COOH to the phenyl ring. The first family, **T1V** (tautomer No. 1, valence anion) involves proton transfer to either **C4** or **C5**. The second, **T2V**, involves proton transfer to **C1**, and the third, **T3V**, to **C2** or **C3**.

The most stable conformers for each family, i.e., **T1V1, T2V1, T3V1**, are characterized in Figure 5.14 (see Figures A.3 and A.4 for a more complete set of conformers). The VDE values for these valence anions exceed 3.2 eV at the projected MP2 and CCSD(T) levels. The excess electron is described by a π^* orbital localized primarily on the protonated phenyl ring (Figure 5.15). One protic H(N) is involved in a hydrogen bond with COO$^-$ and the second H(N) "solvates" electron density on the phenyl ring (Figure 5.2 and Figure 5.15).

The **TnV1** structures ($n = 1 - 3$) do not make much "chemical sense" for the neutral Phe, which is reflected in dipole moments exceeding 15 D and large values of VDE for these valence anions. However, the instability of these anions with respect to the most stable canonical neutral, **CN1**, is surprisingly small. For **T1V1**, it drops from 4.35 kcal/mol in terms of the CCSD/ADZ electronic energy to 2.17 kcal/mol in terms of electronic energy corrected for zero point vibrations, and further to 1.67 kcal/mol in terms of Gibbs free energy. This instability is comparable with error bars of my theoretical model, CCSD(T)/ADZ plus zero-point and thermal corrections.[172] I believe

System	$\mu_{neutral}$	VDE		Relative Stability *w.r.t* CN1				
	MP2	MP2	CCSD(T)	PMP2	CCSD(T)			
				E_{el}	E_{el}	$E_{el}+E_0^{vib}$	H	G
T1V1	15.13	3.35	3.25	8.63	4.35	2.17	2.41	1.67
T2V1	17.05	3.30	3.22	12.07	7.53	5.28	5.65	4.22
T3V1	18.20	3.52	3.50	11.98	7.64	5.37	5.66	4.64
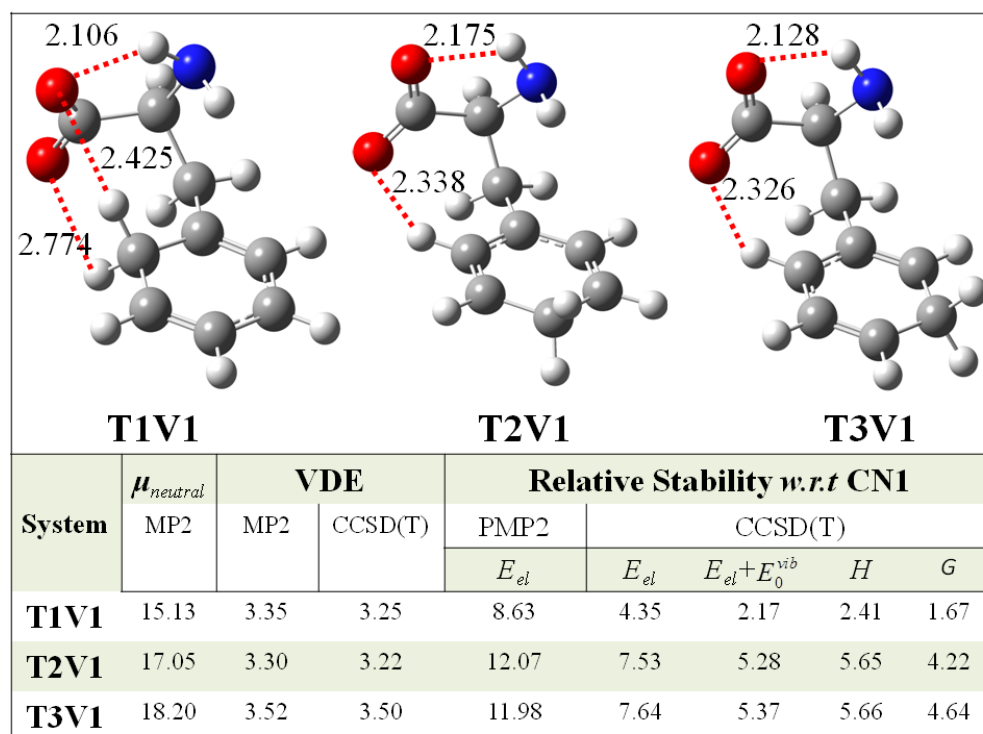
FIGURE 5.14: *The most stable conformers for three families of valence anions of Phe. The dipole moment of neutral Phe at the optimal anionic geometry, $\mu_{neutral}$, in D; the electron vertical detachment energy (VDE) in eV; the relative stability of anions with respect to **CN1** in kcal/mol. The anionic structures (with distances in Å) optimized at the MP2/ADZ level of theory. The remaining quantities the same as in Figure 5.12.*

that valence anions of Phe can be formed in experimental conditions similar to those used in the past to identify valence anions of nucleic acid bases (see Appendix A).[27]

5.5.3 Dipole-Bound Anions of Phenylalanine

The three most stable conformers of dipole-bound anions of Phe based on the canonical tautomer are characterized in Figure 5.16. They are labelled **CDB1**, **CDB2**, and **CDB10** as they are associated with the neutral canonical conformers **CN1**, **CN2**, and **CN10**. These neutrals are characterized by MP2 dipole moments of 4.95, 5.28, and 4.54 D, respectively. Recall that the dipole moment of **CN3** is only 1.16 D, which is too small to support a dipole-bound state. The molecular frameworks distort only slightly upon binding an excess electron and the dipole moment of the neutral increases by 0.05, 0.06, and 0.19 D, upon these geometrical distortions (Figures 5.12 and 5.16). The geometrical distortions can be viewed as nascent zwitterionization, because the (O)H$\cdots$N distance shortens by 0.02 Å, and the OH distance elongates by 0.002 Å.

The excess electron binding energies are in the $30-65$ meV range and the vertical attachment and detachment energies barely differ (Figure 5.16), which is common
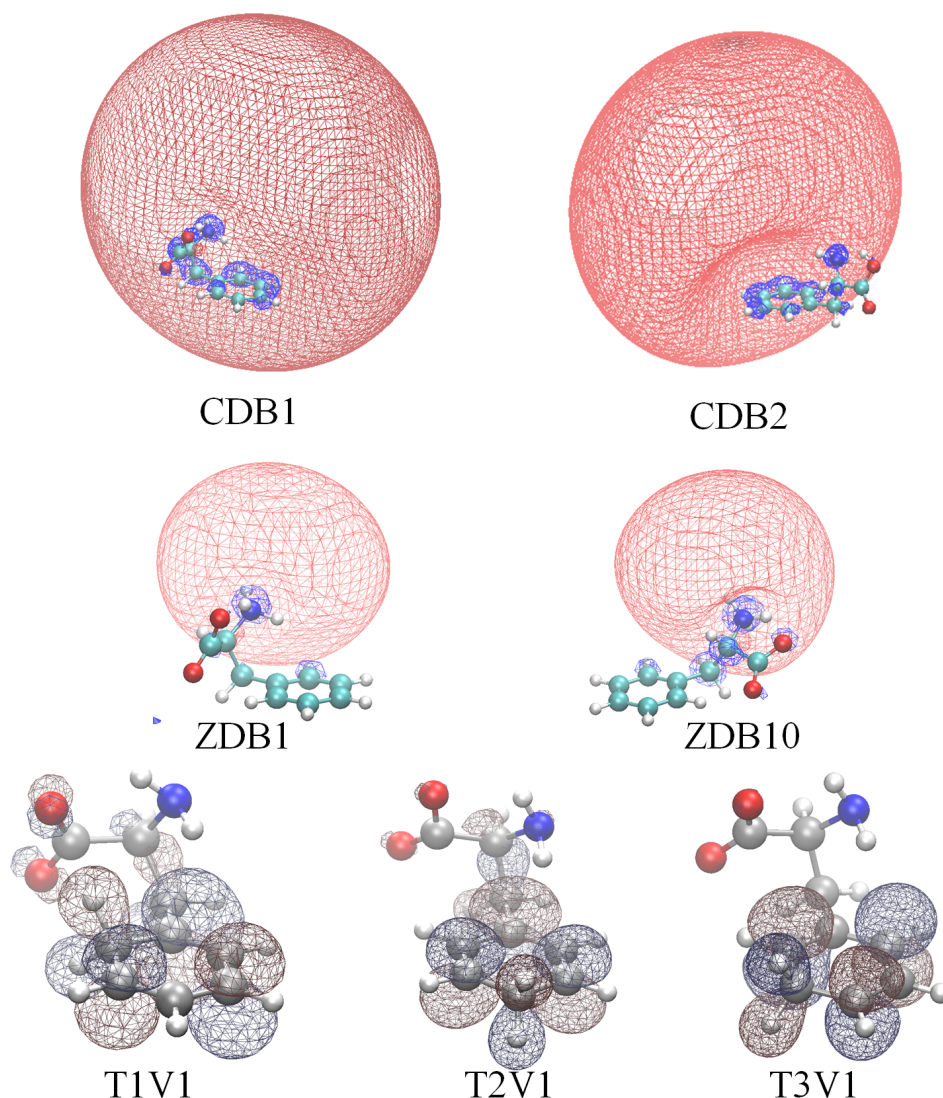
FIGURE 5.15: *Highest occupied molecular orbital of the dipole- and valence-bound anions of phenylalanine generated with contour values determined with OpenCubMan programme 57 (80% of the orbital electron density).*

for dipole-bound anions.[129] Despite their small values, these electron binding energies are measurable in anion photoelectron spectroscopy and Rydberg electron transfer experiments.[137] Based on the $E_{el} + E_0^{vib}$ stability of **CDB1** with respect to **CN1**, I conclude that the gas phase neutral Phe is characterized by an adiabatic electron affinity of 53 meV. Notice lack of correlation between the electron vertical binding energies and the values of dipole moment of the neutral, with **CDB2** having apparently a too small electron binding energy for its dipole moment. The reason is the relative orientation of its dipole and the phenyl ring. The excess electron in typical dipole-bound anions is localized in the region where the electrostatic potential produced by the dipole of the neutral is the most attractive. The electron still fulfils the Pauli exclusion principle and the anionic singly occupied molecular orbital (SOMO) remains orthogonal to the

System	$\mu_{neutral}$	VAE	VDE		Relative Stability *w.r.t* CN1				
	MP2	MP2	MP2	CCSD(T)	MP2			CCSD(T)	
					E_{el}	E_{el}	$E_{el}+E_0^{vib}$	H	G
CDB1	5.00	20.37	23.61	52.72	-0.52	-1.16	-1.23	-1.22	-1.65
CDB2	5.34	11.48	12.36	33.16	1.74	0.76	0.48	0.07	-0.59
CDB10	4.73	23.05	31.18	64.43	4.38	2.96	2.63	2.75	1.63
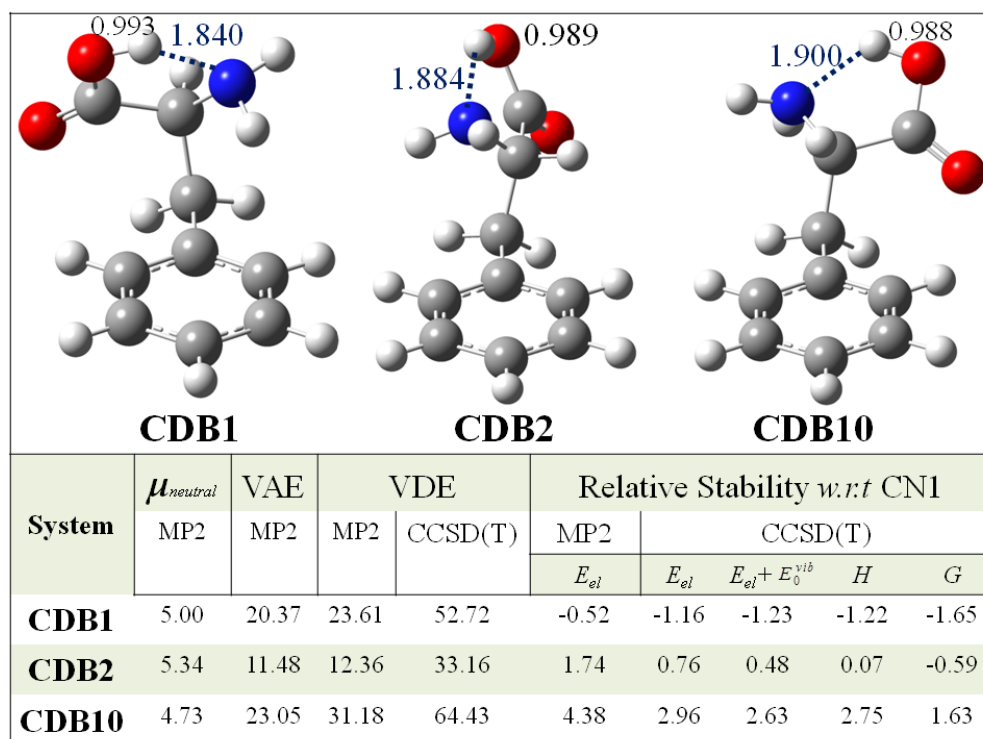
FIGURE 5.16: *The three most stable conformers of dipole-bound anions of Phe based on the canonical tautomer. The dipole moment of neutral Phe at the optimal anionic geometry, , in D; the electron vertical attachment (VAE) and detachment energy (VDE) in meV; the relative stability of anions with respect to **CN1** in kcal/mol. The anionic structures (with distances in Å) optimized at the MP2/ADZ+DF level of theory. The remaining quantities the same as in Figure 5.12.*

occupied orbitals of the neutral, which requires some nodal surfaces in the regions of atoms localized close to the positive pole of the dipole. In the case of **CDB2** the phenyl ring happens to be in the region where the dipolar electrostatic potential would favour localization of the excess electron. The SOMO needs to develop several nodal surfaces that are preceded by a dent visible in Figure 5.15. The multitude of nodal surfaces destabilizes the anion, which is reflected in relatively small values of electron vertical binding energies.

Even though the gas phase neutral Phe does not support zwitterionic minima, local minima of this type exist on the anionic potential energy surface and two most stable, **ZDB1** and **ZDB2**, are illustrated in Figure 5.17. **ZDB1** might be viewed as a product of proton transfer from COOH to NH_2 initiated from **CDB1**. The dipole moment of the neutral increases upon this proton transfer by 3.62 D and the **ZDB1** minimum is characterized be a VDE of 0.293 eV. This is a five-fold increase in comparison with the VDE of **CDB1**. **ZDB1** is, however, unstable with respect to **CN1** by ca. 19.6 kcal/mol, thus adiabatically more unbound than **T1V1**. Similarly, **ZDB2** results from proton transfer initiated in either **CDB2** or **CDB10**. In fact, geometrically it resembles

System	$\mu_{neutral}$	VDE		Relative Stability *w.r.t* CN1				
	MP2	MP2	CCSD(T)	PMP2	CCSD(T)			
				E_{el}	E_{el}	$E_{el}+E_0^{vib}$	H	G
ZDB1	8.62	149.57	292.51	24.51	19.50	19.59	23.92	24.63
ZDB10(2)	8.82	249.04	395.73	29.61	23.31	23.10	27.60	27.62
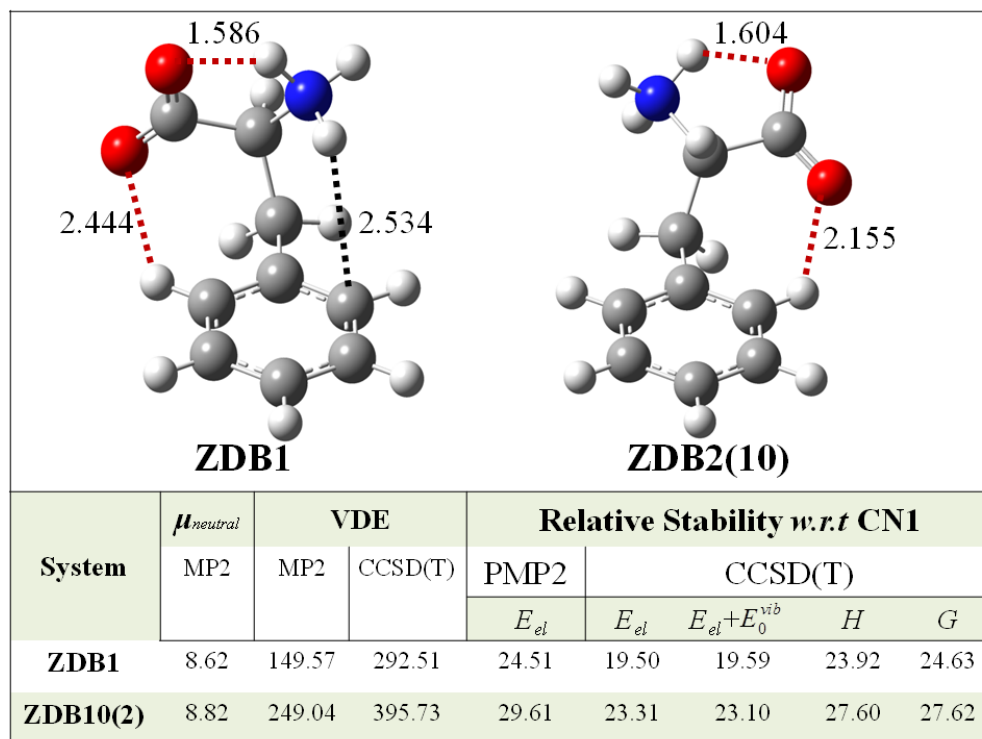
FIGURE 5.17: *The two most stable conformers of dipole-bound anions of Phe based on the zwitterionic tautomer. The dipole moment of neutral Phe at the optimal anionic geometry, $\mu_{neutral}$, in D; the electron vertical detachment energy (VDE) in meV; the relative stability of anions with respect to **CN1** in kcal/mol. The anionic structures (with distances in Å) optimized at the MP2/ADZ+DF level of theory. The remaining quantities the same as in Figure 5.12.*

more **CDB10** than **CDB2** (Figures 5.16 and 5.17). The difference in dipole moments of the neutral at the **ZDB2** and **CDB2** geometries is 3.48 D, and the VDE value for **ZDB2** is 0.396 eV. Again, **ZDB2** is unstable with respect to **CN1** by 23.1 kcal/mol. Notice a significant contraction of the SOMO orbital, attached to the nitrogen terminus of the molecular framework, upon proton transfer from COOH to NH_2 (Figure 5.15).

5.6   Summary and Conclusions

The Potential Energy Surface Scanning Tool (PESST)[28] allows automated scanning of molecular tautomeric and conformational spaces. Its performance has been tested on the neutral and anionic phenylalanine (Phe). Molecular ions frequently favour structures different from those of the neutral species, thus the anions of Phe have been particularly challenging. The prescreening has been performed with lower level electronic structure methods (B3LYP exchange-correlation functional for the neutral and valence anions and Hartree-Fock for dipole-bound anions), while final energies were determined at the CCSD(T) level with basis sets of aug-cc-pVDZ quality.

Three conformers of the neutral canonical Phe have been found differing in stability by less than 1.3 kcal/mol. It was counterintuitive for me to find out that conformers *with* and *without* the intramolecular $(O)H \cdots NH_2$ hydrogen bond might be similarly stable. It has been interpreted as a manifestation of structural strain, which develops upon formation of the hydrogen bond. I have identified a unique IR signature of the conformer without the intramolecular hydrogen bond.

I have also identified three tautomeric families of valence anions of Phe resulting from proton transfer from COOH to the phenyl ring. They are characterized by significant values of VDE, in a range of 3.2-3.5 eV. At 0 K, the most stable tautomer/conformer of valence anions is adiabatically unbound with respect to the most stable canonical neutral by only 2.17 kcal/mol and its stability further improves at elevated temperatures. Based on past efforts with valence anions of nucleic acid bases[27] I suggested that the valence anions of Phe identified in this report can be observed experimentally and indeed they were observed later on and the experimental photoelectron spectrum is shown in Figure A.1.

I characterized dipole-bound anions supported by the canonical and zwitterionic tautomers of Phe. These based on the canonical tautomer are relatively weakly bound, with electron vertical binding energies in a range of $30 - 65$ meV. Still, these might be the lowest anionic states in the system and I predict an adiabatic electron affinity of 53 meV for the most stable conformer of the neutral. The dipole-bound anions remain vertically bound upon intramolecular proton transfer from COOH to NH_2. Local zwitterionic minima develop for these anions with electron vertical binding energies in a range of 290-400 meV. These minima are adiabatically unbound with respect to the canonical neutral by ca. 19-23 kcal/mol. I observed that the phenyl ring of Phe can interfere with the SOMO orbital of its dipole-bound anion and thus reduce stability of the anion.

Chapter 6

Adiabatically Bound Valence Anion of Parent Tyrosine: A Remarkable Stability and Unusual Structure

6.1 Abstract

The parent anions of tyrosine (doublets) and anions resulting from deprotonation of tyrosine (singlets) were studied in a joint computational and anion photoelectron spectroscopy effort. I searched for the most stable structures using a combinatorial-computational approach. Extended libraries of tautomers and conformers were constructed and the most stable structures were identified in the course of a multi-stage selection process. The most stable valence anions of tyrosine are adiabatically bound with respect to the most stable canonical neutral by ca. 0.6 eV and are characterized by electron vertical detachment energy of ca. 3.8 eV. These anionic tautomers result from a double proton transfer from the carboxylic and hydroxyl groups to the ring, with transferred protons occupying the neighbouring ortho and meta positions with respect to the amino acid residue. I also identified a few other families of anionic tautomers and characterized their relative stability and values of VDE. Two families, which might be kinetically easy to form, have VDE values matching positions of distinct peaks in the PES spectrum.

6.2 Introduction

The amino acids such as L-Phenylalanine, L-Tyrosine are important intermediates in biochemical processes.[173–175] These are two of the 21 standard amino acids found in nature. They are the building blocks of proteins and play major role in molecular physiology.[176, 177] L-Phenylalanine is a precursor to L-Tyrosine, which in turn, is a precursor to a neurotransmitter, dopamine. I have recently studied L-Phenylalanine in detail[29] and also others.[154] Here I turn attention to L-Tyrosine (Tyr) (Fig. 6.1), which plays an important role in protein redox processes.[178, 179] It acts as an intermediate in an electron transfer process by promoting signals from the surface of the cell to the nucleus.[174, 178, 179] The excess charge might induce structural changes in Tyr. The structural modifications might explain the mechanism of signal transmission.

Valence anions of molecular systems are more challenging for structural determination than their neutral counterparts. They might favour tautomers that would be quite unstable for the neutral system. For example, past studies of valence anions of nucleic acid bases demonstrated that the most stable anionic tautomers result from proton transfer from a typical proton donor (e.g., NH) to carbon sites, rather than to conventional proton acceptors involving highly electronegative atoms (O or N).[27, 29, 142–144] The valence anions observed experimentally were not necessarily adiabatically bound with respect to the most stable neutral, though they displayed significant electron vertical detachment energies (VDE) of 1-3 eV. Apparently, the anionic tautomers are separated by sufficiently high barriers from these regions of potential energy surface where the valence anion is weakly bound or even unbound. Notice that the six-member ring of Tyr might support unusual tautomers of its valence anion, which would result from proton transfer from the carboxylic and/or hydroxyl groups to the phenol ring.

I reported a proton transfer from the carboxylic group to the phenyl ring for the most stable tautomers of valence anions of L-Phenylalanine in chapter 5.[29] These valence anions were vertically strongly bound, with VDEs in a range of 3.2–3.5 eV. Even though the most stable conformer of these valence anions remained adiabatically unbound with respect to the canonical neutral by ca. 2 kcal/mol at the CCSD(T)/aug-cc-pVDZ level of theory, these valence anionic states have recently been observed experimentally.[180] Tyr is more challenging than L-Phenylalanine for exploration of possible tautomers due to the presence of the hydroxyl group, a potential proton donor site. A comprehensive exploration of its structural space requires an advanced tool for enumeration of possible tautomers and conformers. I have recently developed this advanced tool, Potential Energy Surface Scanning Tool (PESST) described in chapter 5,[181] and I will use it in the course of this project.

Physiochemical properties of Tyr have been probed in mass spectrometry[182, 183], FT-IR spectroscopy,[178] ion-beam scattering,[184] electron spray ionisation,[182] and electron paramagnetic resonance spectroscopy[185, 186] studies. The computational studies addressed the conformational preference of the neutral, radical cation, and deprotonated species[182, 187]. The latter anions differ from the main target of my study, radical anions of complete (parent) Tyr. The neutral and deprotonated species were studied in the combined anion photoelectron spectroscopy and computational study.[182] The main and striking observation was that the gas phase sample of deprotonated species was represented by a 70:30% mixture of phenoxide and carboxylate species, though carboxylate anions were dominant in solutions. Thus it is easier to deprotonate gas phase Tyr on the hydroxyl than carboxylic site. This finding will have bearing on my study of valence radical anions of complete Tyr. The dynamics of collisions of deprotonated Tyr with $O_2[a^1\Delta g]$ was studied in Ref. 184.

In this study, I will search for the most stable tautomers and conformers of the neutral and anionic Tyr. My computational efforts will start from systematic constructions of molecular libraries[29] and the most stable structures will be determined using state-of-the-art electronic structure methods. The parallel anion photoelectron spectroscopy measurements will verify correctness of my predictions for molecular anions and will be described in details in the Appendix B.

6.3   Computational Methodology

The most stable structures of the neutral and anionic tyrosine were identified using a combinatorial-computational approach [22] and a versatile tool for creation of molecular conformers and tautomers, PESST.[29] I considered several tautomers of Tyr and some representative structures are illustrated in Fig. 6.1. I found that the canonical tautomer is the most stable for the neutral Tyr, but it does not support vertically bound valence anions. Thus it is labelled N from now on. The tautomer D results from a double proton transfer to the ring. One proton comes from the carboxylic group, another from the hydroxyl group and I considered the carbon atoms 2, 3, 5, and 6 as potential binding sites for these protons. Another tautomer, labelled D^N, results from a double proton transfer from the same proton donor sites, but one of the protons is attached to the nitrogen, another to the phenyl ring. There are two families for single proton transfer to the ring: SH results from proton transfer from the hydroxyl group and SC from the carboxylic group. For the two latter families the carbon atoms 2 and 3 are considered as potential binding sites.
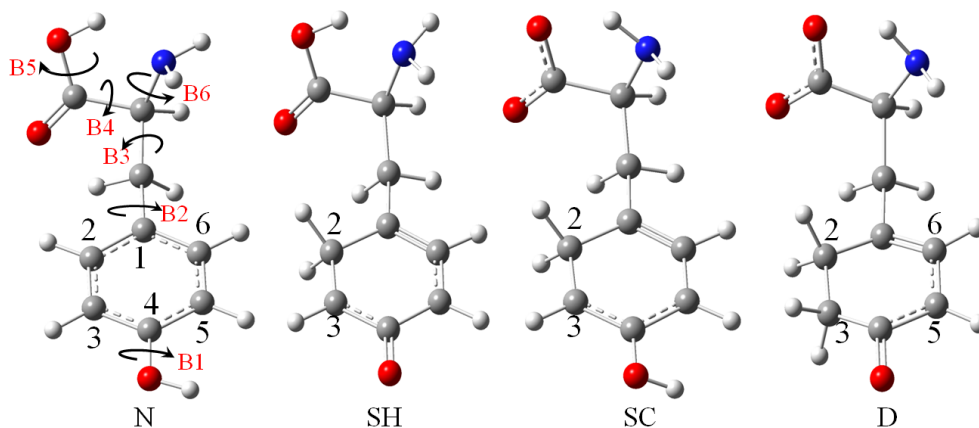
FIGURE 6.1: *Tautomers of Tyr considered in this study. Rotatable bonds B1 – B6 considered when generating conformers of neutral or anionic Tyr.*

		n_{rot} value				
Tautomer	B1	B2	B3	B4	B5	B6
N	2	0 & 90°	6	6	2	3
$D^{k,l}$	–	4	6	6	–	3
$D^{N,k}$	–	4	6	6	–	–
SH^k	–	4	6	6	2	3
SC^k	2	4	6	6	–	3

TABLE 6.1: The n_{rot} values for bonds B1 – B6 of the $N, D^{k,l}, D^{N,k}, SH^k$ and SC^k systems considered in this report.

The rotational bonds activated when searching for the most stable conformers are labelled B1 – B6 in Fig. 6.1. I created libraries of conformers by rotating the Bn bonds with increments $360°/n_{rot}$, where n_{rot} is an integer defined for each rotatable bond Bn. The detailed information about the active rotatable bonds and the n_{rot} values is summarized in Table 6.1. These procedures led to 864 initials structures for N, 432 for D, 144 for D^N, and 864 for SH and SC.

With the active B2 rotatable bond, the carbon site 2 is equivalent to 6 and 3 to 5. When dealing with D–type structures I considered protons occupying the following k,l positions: 2,3 or 2,5 or 2,6 or 3,5. These structures will be labelled $D^{k,l}$ and various conformers of these tautomers will be labelled $D^{k,l}_m$, with the stability decreasing as m increases. Similarly, $D^{N,k}_m$ denotes a specific member of the D^N family: it is the m-th most stable conformer of the tautomer resulting from one proton transferred to nitrogen and another to the k'th carbon atom of the phenyl ring. Finally, SC^k_m and SH^k_m denotes the mth most stable conformer of the tautomer resulting from a single proton transfer to the kth carbon atom in the phenyl ring from the carboxylic and hydroxyl group, respectively. The mth conformer of the anion resulting from deprotonation of Tyr at the hydroxyl and carboxylic site will be labelled DP^H_m and DP^C_m, respectively.

The initial structures created by PESST were pre-screened by performing full geometry optimisations using Becke, three-parameter, Lee-Yang-Parr (B3LYP)functional[100, 111–113] with Pople's type 6-31++G**[163] basis set. The lowest energy structures from each tautomer that spanned an energy range of 5 kcal/mol were re-optimised at B3LYP and MP2 levels using Dunning's ADZ[115, 121] basis set, followed by frequency calculations. Finally, single point CCSD(T)[188]/ADZ energy calculations were carried out at the MP2 optimised geometries for the three most stable conformers of each tautomer. To summarize, my most accurate energies of the neutral and anionic systems are based on the CCSD(T) electronic energies and zero-point vibrational energies determined at the MP2 level, both obtained with the ADZ basis set. The electronic structure calculations reported in this study were performed using the Gaussian 09[75] suite of programmes. The orbitals occupied by an excess electron were generated with the Molden[161] programme using a contour value of 0.5 au. The GaussView[189] package was used to draw molecular structures.

6.4   Computational Results

The most stable conformers of the neutral Tyr and their relative stabilities are illustrated in Fig. 6.2. My search failed to identify locally stable conventional zwitterionic structures, with the carboxylic proton transferred to the NH_2 group. All geometry optimizations initiated from initial zwitterionic structures collapsed to the canonical N tautomer. The $N1--N3$ conformers span a narrow energy range of 1.6 kcal/mol and the relative stabilities calculated at the CCSD and CCSD(T) levels differ by less than 0.21 kcal/mol (see the supporting information). All three conformers support a reasonably short $OH \cdots N$ intramolecular hydrogen bond. $N2$ differs from $N1$ by rotation of the hydroxyl group only, but I illustrate it here to emphasize a significant increase of the dipole moment from 3.7 to 6.2 D. This makes the $N2$ conformer suitable to support a dipole-bound state[191]. The most stable conformers I identified agree with the most stable conformers reported in Refs. 182, 187.

An overview of stability of various tautomers of tyrosine in the valence anionic state is illustrated in Fig. 6.3. These relative B3LYP/ADZ electronic energies are reported with respect to the most stable neutral conformer, $N1$. The results should be viewed with a grain of salt, because the stability of valence anions is frequently overestimated at the B3LYP level of theory.[29] Still, they suggest that the valence anions of Tyr might be adiabatically bound provided at least one proton is transferred to the phenyl ring. They also justify my selection of $D^{2,3}$, SH^3, $D^{3,5}$, SC^2, and $D^{N,3}$ for further, more accurate, considerations and discarding the $D^{2,6}$ and SH^2 families. The $D^{N,2}$ family

N1 0.00
$\mu = 3.68$

N2 0.31
$\mu = 6.20$

N3 1.62
$\mu = 4.98$
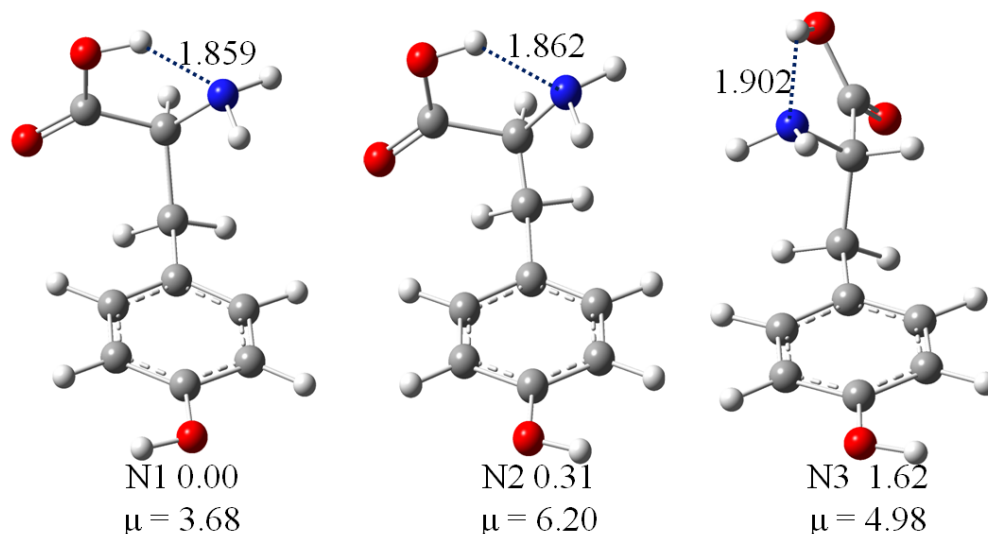
FIGURE 6.2: *Structures and relative stability (in kcal/mol) of low-lying conformers of neutral Tyr. The intra-molecular hydrogen bond parameters are in Å. Dipole moments in Debyes.*

is not reported because the anionic structures collapsed to the SH^2 structures. The viability of double proton transfer involving the carboxylic and hydroxy groups of Tyr was already suggested by the results of Ref. 182. My results demonstrate that the double proton transfer to the phenyl ring carbons 2 and 3 creates the most stable molecular framework for the valence anion of Tyr.

From now on, the presented stabilities of valence anions will be based on electronic energies determined at the CCSD(T) level and molecular geometries and zero-point vibrational corrections determined at the MP2 level; all results obtained with the ADZ basis set. The energies of the anions are reported with respect to the energy of the neutral $N1$, thus they are equivalent to adiabatic electron affinities.

The three most stable conformers from the $D^{2,3}$ family are illustrated in Fig. 6.4. The stability of $D_1^{2,3}$ is approximately 0.56 eV, thus much larger than the uncertainty of my theoretical model. I concluded that these three (and nine other $D^{2,3}$ conformers reported in the supporting information) are most probably adiabatically bound with respect to $N1$. The two protons transferred to the phenyl ring interact with lone electron pairs of oxygens from the carboxylate group. The NH$_2$ group is also involved in hydrogen bonding with the carboxylate group. The phenyl ring is strongly buckled at the protonated carbons. In $D_1^{2,3}$, the (1,2,3,4) dihedral angle is 49.0° and very similar values were found for the analogous dihedral angles in $D_2^{2,3}$ and $D_3^{2,3}$. The molecular orbitals occupied by the unpaired electron are illustrated in the bottom part of Fig 6.4. The electron density is localized on the side of the ring opposite to the transferred protons. The bonding pattern is analogous in all three conformers: a bonding interaction between C1 and the

$D_1^{2,3} : -14.17$ $D_1^{3,5} : -2.08$ $D_1^{2,5} : -1.55$ $D_1^{2,6} : +50.46$

$D_1^{N,3} : +0.44$ $SH_1^3 : -7.01$ $SH_1^2 : +6.16$ $SC_1^2 : -1.05$ $SC_1^3 : +0.32$
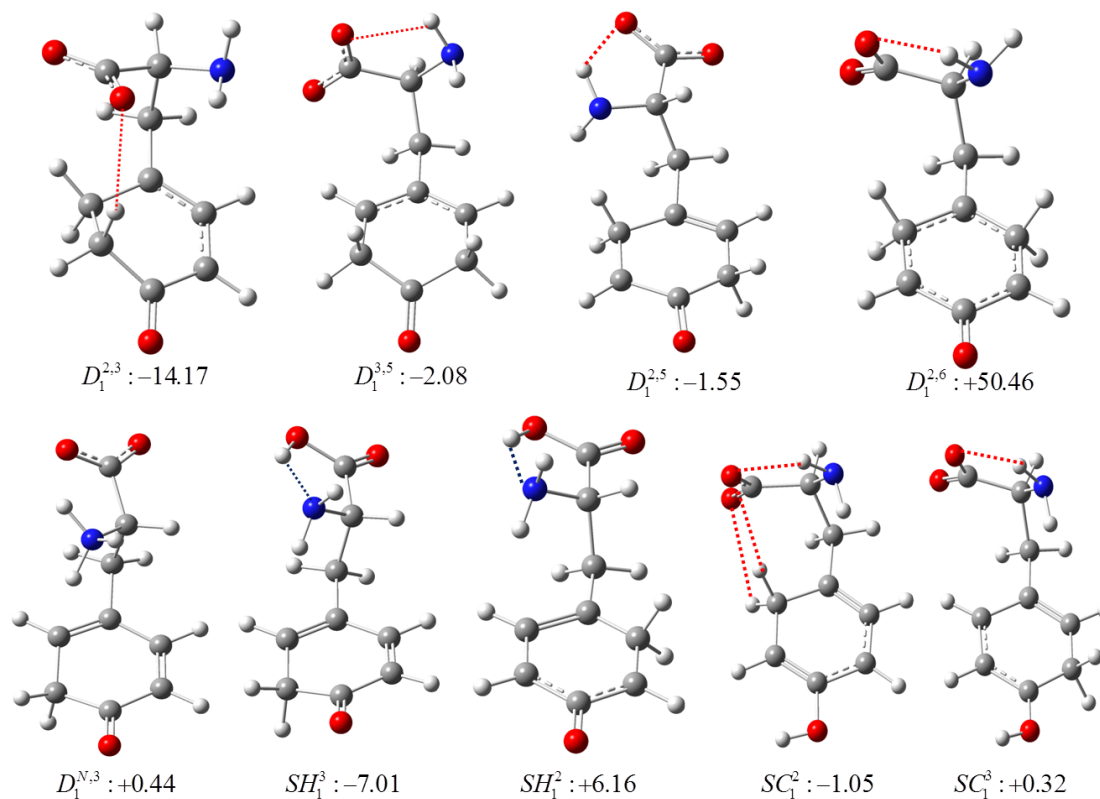
FIGURE 6.3: *The most stable conformers of the families of valence anions of tyrosine considered in this study. The relative energies (in kcal/mol) with respect to the most stable canonical neutral $N1$, were calculated at B3LYP/ADZ level.*

next carbon followed by an antibonding interaction. The electron vertical detachment energies are significant, all in the 3.76 – 3.80 eV range.

The three most stable conformers from the $D^{3,5}$ family are illustrated in Fig. 6.5. Their stability with respect to $N1$ is still substantial, ca. -0.1 – -0.2 eV. The NH_2 group remains involved in hydrogen bonding with the carboxylate group. The electron density of unpaired electron is localized on C2 and C6, but not on C1 and C4. The electron vertical detachment energies are significant, all in the 3.65 – 3.90 eV range.

The two most stable conformers from the $D^{N,3}$ family are illustrated in Fig. 6.6. They are unstable with respect to $N1$ by ca. 0.3 – 0.4 eV. I report their characteristics despite their instability because formation of these doubly-proton-transferred anions is kinetically the most straightforward: the carboxylic proton transfers to nitrogen and the hydroxyl proton transfers to C3. The electron density is localized on the side of the ring opposite to the transferred proton. The bonding pattern is analogous in both conformers: a bonding interaction between C4 and the next carbon followed by an antibonding interaction. The electron vertical detachment energies are much smaller than for the $D^{2,3}$ and $D^{3,5}$ families, ca. 1.6 – 1.7 eV.
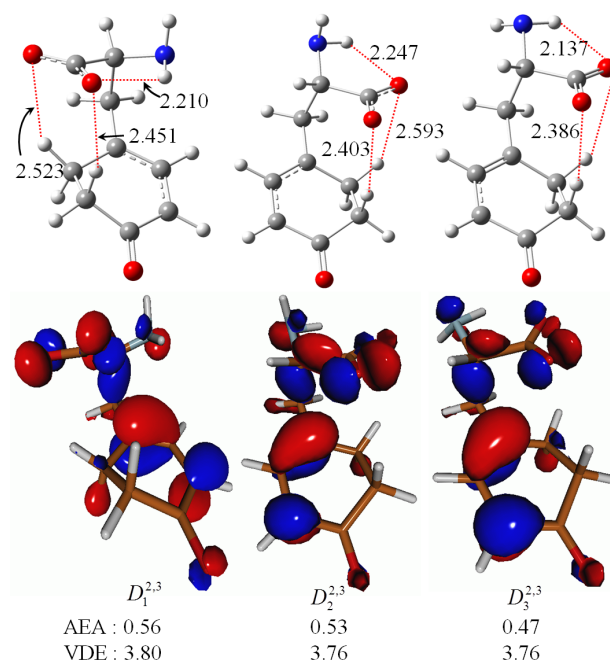
FIGURE 6.4: *Three most stable conformers of $D^{2,3}$ family. AEA with respect to the most stable N1. The geometrical parameters are in Å, AEA and VDE are in eV.*
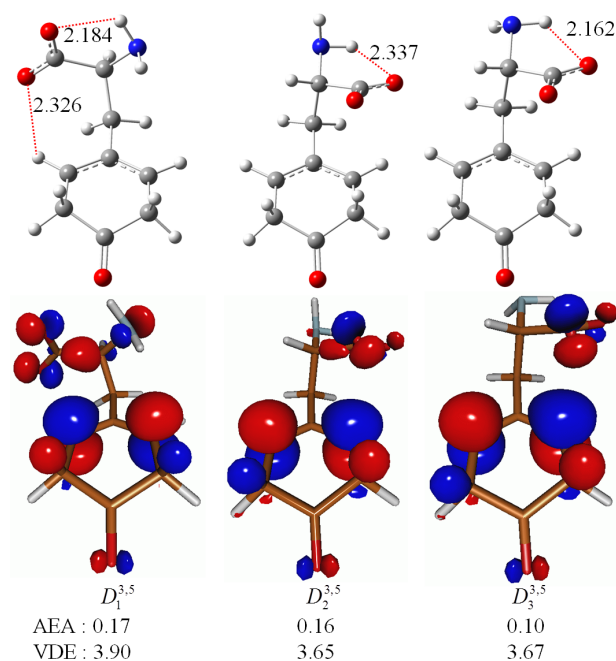
FIGURE 6.5: *Three most stable conformers of $D^{3,5}$ family. AEA with respect to the most stable N1. The geometrical parameters are in Å, AEA and VDE are in eV.*

$$D_1^{N,3} \qquad\qquad D_2^{N,3}$$

AEA :	-0.30	-0.36
VDE :	1.73(1.80)	(1.63)
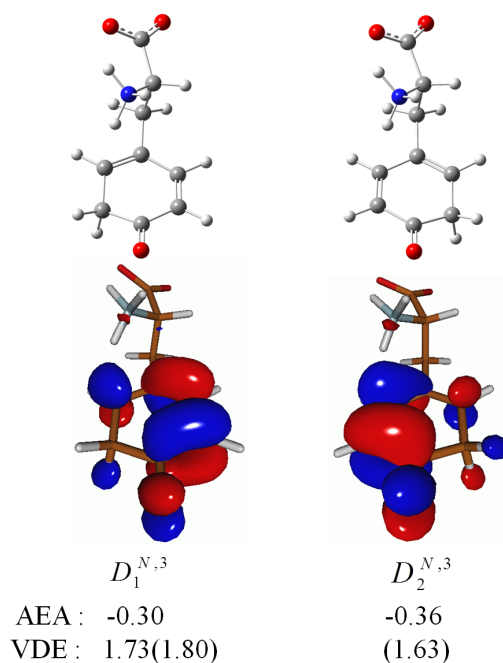
FIGURE 6.6: *Two most stable conformers of $D^{N,3}$ family. AEA with respect to the most stable $N1$. The geometrical parameters are in Å, AEA and VDE are in eV.*

The three most stable conformers from the SH^3 family are illustrated in Fig. 6.7. Their stability with respect to $N1$ is still substantial, ca. -0.08 – -0.10 eV. SH_1^3 and SH_2^3 are very similar, they differ by rotation of the phenyl ring by 180° around the B2 bond. The OH from the carboxylic group forms a relatively short hydrogen bond with the lone electron pair of nitrogen. In SH_3^3, however, the same OH interacts with the excess electron density localized on the ring, while the NH_2 groups acts as a symmetric double proton donor interacting with the CO proton acceptor site. The electron density is localized on the side of the ring opposite to the transferred proton. The bonding pattern is analogous in three conformers: a bonding interaction between C4 and the next carbon followed by an antibonding interaction. The electron vertical detachment energies are even smaller than in the $D^{N,3}$ family, ca. 1.0-1.2 eV.

The three most stable conformers from the SC^2 family are illustrated in Fig. 6.8. They are unstable with respect to $N1$ by ca. 0.06 – 0.11 eV. I also report their characteristics despite their instability because formation of these singly-proton-transferred anions is also kinetically straightforward: the carboxylic proton transfers to the neighbouring C2. The NH proton donor and the protonated carbon from the phenyl ring interact with the carboxylate group. The excess electron density is delocalized over the whole phenyl ring, still favouring the side opposite to the transferred proton. The bonding pattern is analogous in all conformers: a bonding interaction between C1 and the next carbon followed by a series of antibonding interactions. The electron vertical detachment energies are substantial, all in the 3.3-3.5 eV range.

	SH^3_1	SH^3_2	SH^3_3
AEA:	0.08	0.09	0.10
VDE:	1.02	1.07	1.23
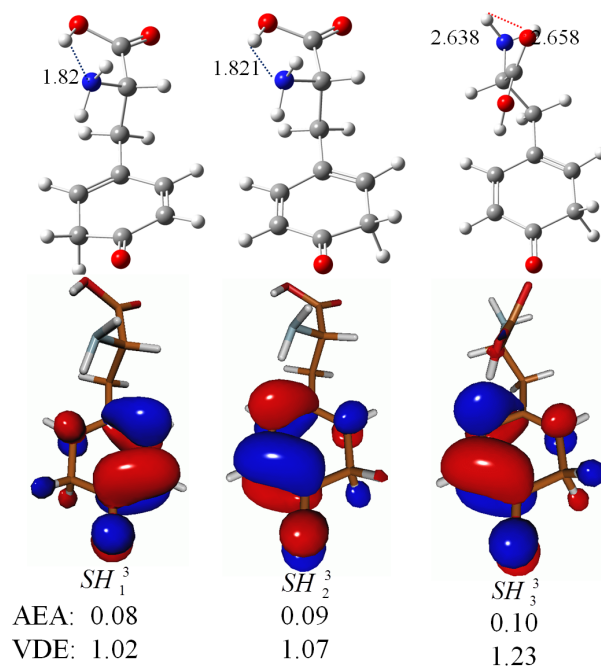
FIGURE 6.7: *Three most stable conformers of SH^3 family. AEA with respect to the most stable N1. The geometrical parameters are in Å, AEA and VDE are in eV.*

	SC^2_1	SC^2_2	SC^2_3
AEA :	-0.06	-0.09	-0.11
VDE :	3.34(3.43)	(3.46)	(3.26)
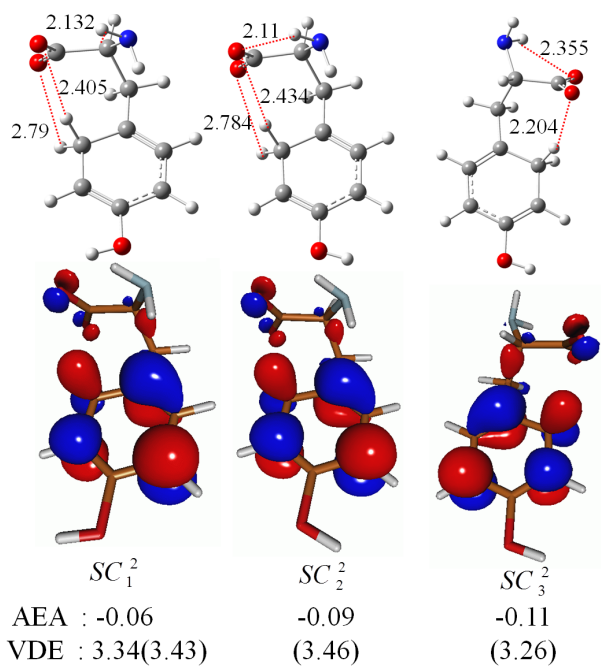
FIGURE 6.8: *Three most stable conformers of SC^2 family. AEA with respect to the most stable N1. The geometrical parameter are in Å, AEA and VDE are in eV.*

The interpretation of PES spectra reported in Sec. 6.5 required elementary information about deprotonated tyrosine, DP. The results of my B3LYP and MP2 calculations were consistent with the earlier results of Tian et al.[182] I did not find any more stable conformer than they have done. The range of MP2 VDE's from the three most stable DP^C and DP^H conformers was 3.9 – 5.5 and 2.8 – 3.1 eV, respectively. My detailed results on the conformers of DP^C and DP^H are available in the supporting information.

6.5   Anion Photoelectron Spectroscopy Results

The PES spectra of the tyrosine parent anion in figure 6.9 was recorded with 4.86 eV photons and laser ablation pulses of different power are presented by my experimental collaborators. The detail experimental procedure can be found in Appendix B. The spectrum involves broad and overlapping peaks, which are labelled A – D. The intensity of B at a EBE of ca. 2.8 eV was sensitive to the power of laser ablation pulses. I expect that this peak is associated with the deprotonated Tyr (DP), rather than the tyrosine parent anion. The samples of Tyr used by my collaborators involve naturally occurring isotopes, thus the anionic beam with the molecular mass of $C_9H_{11}NO_3$ might contain up to 9% of DP, i.e., $C_9H_{10}NO_3^-$, assuming that one ^{12}C is replaced with ^{13}C. The position of peak B coincides with my values of VDE for the DP^H family, 2.8 – 3.1 eV. Also the authors of Ref. 182 reported a peak with EBE between 2.6 and 3.0 eV in their study on DP. Apparently, the efficiency of formation of DP in our source of anions is sensitive to the laser ablation power. This allowed me to discriminate B from other peaks associated with the tyrosine parent anion.

Next, I focus my attention on other features in the PES spectrum from Fig. 6.9. Its onset at ca. 1.3 eV is larger than the AEA value of 0.56 eV reported by me. It indicates that the 0–0 transition has a negligibly small intensity in the PES spectrum. I associate peak A at ca. 1.8 eV with the $D^{N,3}$ family. This family is kinetically easily accessible and the calculated VDE values of 1.6-1.7 eV match the position of peak A. I associate peak C at ca. 3.3 eV with the SC^2 family. This family is also kinetically easily accessible and the calculated VDE values of 3.3-3.5 eV match the position of peak C. Finally, I associate a broad peak D at 3.9 eV with the $D^{2,3}$ and $D^{3,5}$ families. The calculated VDE values of 3.7-3.9 eV match the position of this broad peak.

The PES spectrum of DP recorded with 3.49 eV photons is presented in Fig. 6.10. I associate peak B at 2.82 eV with the DP^H family (the calculated VDE's in the 2.8-3.1 eV range). I believe that the same peak contaminates the PES spectrum of tyrosine parent anion in Fig. 6.9. I am unable to provide a convincing interpretation of the A peak at 1.56 eV. Per analogy with B in Fig. 6.9, it may be associated with $C_9H_9NO_3^-$,
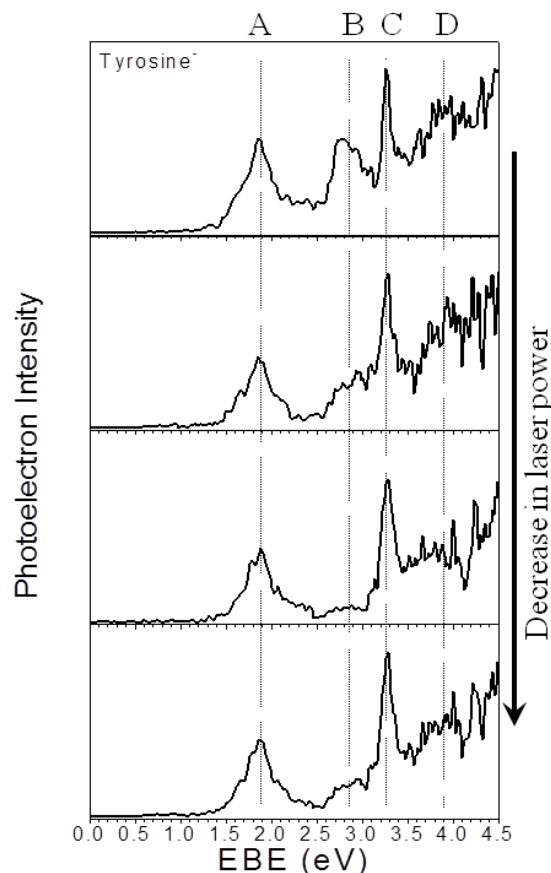
FIGURE 6.9: *Photoelectron spectra of Tyr⁻ recorded with 4.86 eV and laser ablation pulses of different power. The procedure for obtaining the spectra is outline in appendix B.*

assuming that one ^{12}C is replaced with ^{13}C. So far I have failed to identify a molecular anion of this composition with a VDE of 1.56 eV. It should be pointed out that the 3.49 eV photons used in my collaborators' experiments are unable to unravel a high EBE part of the spectrum for DP. My calculated VDE values for the DP^C family are in the 3.9 – 5.5 eV range. In fact, the PES spectrum of DP reported in Ref. 182 with 4.66 eV photons has a peak between 3.4 and 4.0 eV, thus consistent with my predictions for the DP^C family. Further experimental and computational work is needed to understand the origin of peak A in Fig. 6.10.

6.6 Summary and conclusions

Structural modifications of tyrosine upon attachment of an excess electron and formation a valence anionic state were the main focus of this study. My electronic structure calculations were paralleled by an anion photoelectron spectroscopy study of the tyrosine parent anion and deprotonated tyrosine. I searched for the most stable neutral
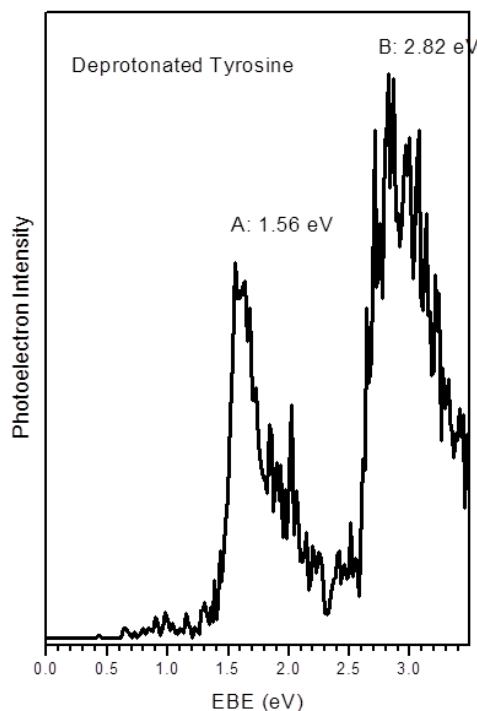
FIGURE 6.10: *Photoelectron spectrum of deprotonated Tyr recorded with 3.49 eV photons.*

and anionic structures of tyrosine using a combinatorial-computational approach.[22] Extended libraries of tautomers and conformers of tyrosine were constructed using the potential energy surface scanning tool[29] and the most stable structures were identified in the course of a multi-stage selection process. Based on past efforts with nucleic acid bases[27] and phenylalanine[29] I considered anionic tautomers resulting from proton transfer to the phenyl ring. Both single and double proton transfer tautomerizations were considered involving the carboxylic and hydroxyl proton donors.

My results provide evidence that the most stable valence anions of tyrosine are adiabatically bound with respect to the most stable canonical neutral by ca. 0.6 eV and are characterized by electron vertical detachment energy of ca. 3.8 eV. These anionic tautomers result from a double proton transfer from the carboxylic and hydroxyl groups to the ring, with transferred protons occupying the neighbouring ortho and meta positions with respect to the amino acid residue. I identified a few other families of anionic tautomers and characterized their relative stability and values of VDE. Two families, which might be kinetically easy to form, have VDE values matching positions of distinct peaks in the PES spectrum.

The PES spectrum of the parent anion of tyrosine was contaminated with a peak associated with deprotonated tyrosine. Some molecules of deprotonated tyrosine are present in the anionic beam of the properly mass-selected parent anion due to the natural abundance of the ^{13}C isotope. Other peaks in the PES spectrum were assigned to different families of valence anions identified in the course of my computational search. The measured PES spectrum of deprotonated tyrosine has two distinct features at 1.56 and 2.82 eV. The latter was interpreted in terms of anions resulting from deprotonation of the hydroxyl group. The former has not been interpreted yet and my ongoing search is focused on anions that might be confused with deprotonated tyrosine in consequence of mass leaks (^{13}C vs ^{12}C).

Chapter 7

Intramolecular Proton Transfer in Acetoacetic Acid Induced by a π^* Excess Electron

7.1 abstract

Acetoacetic acid (AA), equipped with neighbouring carboxylic and keto groups, is a promising system for studies of intramolecular proton transfer. The results of my computational search for the most stable tautomers and conformers of the neutral and anionic AA allowed to guide and interpret the parallel anion photoelectron and electron energy-loss spectroscopy measurements. I identified several conformers of the canonical neutral AA within a narrow energy range of 1 kcal/mol. The conformer with the intramolecular hydrogen bond, and thus susceptible to intramolecular carboxylic-to-keto proton transfer, proved to be the third most stable and the most polar, with a dipole moment of 5.4 D. The polar neutral conformer supports a dipole-bound anion with an electron vertical detachment energy of 57 meV. The valence anions of AA all exist in the enol form. The distinct global minimum is more stable than other minima by more than 12 kcal/mol. It is supported by a short intramolecular hydrogen bond between the enol and carboxylate groups. The valence anion is characterized by a significant electron vertical detachment energy of 2.38 eV, but a modest adiabatic electron affinity of 0.33 eV. The inherent instability of AA with respect to acetone and CO_2 is suppressed in the valence anionic state. The valence anion was identified in photoelectron spectroscopy experiments and the measured electron vertical detachment energy of 2.30 eV is very good agreement with my computational predictions. I conclude that binding an excess electron in a π^* valence orbital changes localization of the proton in AA. The results of

electron energy-loss spectroscopy demonstrate that AA is able to quasithermalize electrons with energies in the 1–2 eV range. Its efficiency is, however, lower than the that of the dimer of formic acid. My computational results allow rationalizing these experimental findings in terms of the co-existence of various conformers of AA. Only one, and not the most stable, conformer is prearranged for intramolecular proton transfer and thus quasithermalization of electrons. All other low-lying conformers undergo electron autodetachment with or without vibrational excitations.

7.2 Introduction

The phenomenon of proton transfer is coupled with electronic structure of molecules involved. In particular, an excess electron can drive redistribution of protons.[192] Proton motion coupled with electron transfer has been identified long ago as the basic mechanism of bioenergetic conversion.[193] Recent experimental and theoretical research on electron-driven processes has been primarily driven by two factors: (i) damage of DNA and RNA by low-energy electrons[194] and (ii) the potential of electron beam lithography to fabricate the next generation of micro- and nano-electromechanical devices.[195]

Recent experimental gas phase studies on electron-induced proton transfer in fragments of DNA and proteins[148, 192] have been hampered by low vapor pressure of these molecules. This motivated me to indentify model systems in which analogous fundamental chemical physics processes take place, but which display a larger vapor pressure than typical biomolecules. When dealing with intermolecular proton transfer induced by a π^* excess electron I suggested that the dimer of formic acid may serve as a model system.[196] It displays many similarities with intermolecular proton transfer in anionic complexes of nucleic acid bases with weak acids.[196] Similarly, a hydrogen bonded complex of ammonia and hydrogen chloride may serve as a model system for intermolecular proton transfer induced by a σ excess electron.[197]

The electron-induced proton transfer in formic acid dimer was invoked to interpret differences in the results of electron energy loss spectroscopy experiments on the monomer and dimer of formic acid.[198] A peak of very low energy electrons was found to be much stronger in the dimer than in the monomer. The dramatic increase in the efficiency of the dimer to quasithermalise electrons arriving in the 1-2 eV energy range was interpreted in terms of intermolecular proton transfer that quenches the fast autodetachment channel. It was concluded that the phenomenon of electron-driven proton transfer can be ubiquitous and it may be responsible for rapid slowing down of excess electrons.

The anion of formic acid dimer has recently been experimentally characterized using Ar-tagged vibrational predissociation and electron autodetachment spectroscopies.[199] These results confirmed that excess electron attachment occurs with barrier-free transfer of one of the protons across the H-bonded bridge. The study corroborated that the anion of formic acid dimer composes of a largely intact formate anion attached to the dihydroxymethyl radical through a symmetrical, double O-H bonded bridge.

Much less information is available on intramolecular proton transfer induced by excess electron and the results are available primarily for low-vapour pressure molecules, e.g., nucleotides.[200] For example, the anion photoelectron spectrum of 2'-deoxyadenosine-5'-monophosphate[200] has been interpreted through intramolecular proton transfer from a hydroxyl group of the phosphate to the N3 position of the adenine.[201] Fundamental studies of intramolecular proton transfer induced by an excess electron would benefit from model systems with well defined intramolecular hydrogen bonds and vapour pressures significantly larger than those in biomolecules.

Here I present computational and experimental results on the neutral and anionic acetoacetic acid (AA), see Fig. 7.1. Upon formation of proper conformers, the neighbouring carboxylic and keto group can engage in intramolecular hydrogen bond providing a bridge for intramolecular proton transfer. AA is the simplest beta-keto acid and is thermodynamically unstable with respect to decomposition to acetone and CO_2. However, the experimental half-file of 140 minutes has been reported for a water solution of AA at 37 °C.[202] Thus the molecule can be probed experimentally provided care is taken to operate at low temperatures. Some computational information is available on the ketonic decarboxylation of AA[203–205]. Huang et al.[203] reported a two step process of unimolecular decarboxylation with barriers of 25.7 and 55.9 kcal/mol at the MP2/6-31G* level of theory.

My computational study is focused on the tautomers and conformers of the neutral and anionic AA. The anion photoelectron spectroscopy measurements allow probing bound anionic states of AA and verify computational predictions. The electron energy loss spectroscopy measurements allow probing the anionic resonance state of this molecule. Based on past efforts with the dimer of formic acid[198], particular attention is paid to the possibility of quasithermalisation of electrons arriving in the 1-2 eV energy range,

7.3   Computational Methods

The tautomers and conformers of neutral and anionic AA were explored using my in-house tool, PESST[29]. Only two tautomers were considered, the canonical *keto* tautomer and a zwitterionic *enol* tautomer resulting from proton transfer from O2 to O1, both shown in Fig. 7.1. The *keto* tautomer supports a dipole-bound anionic state. The *enol* tautomer supports a strongly bound valence anion in addition to the dipole bound anion. Henceforth I will use **DB** and **VB** to refer to the dipole- and valence-bound anions, respectively. The conformational space of *keto* was explored through systematic rotations around bonds α, β, γ, and η, see Fig. 7.1, while ϵ replaces α for *enol*.

When probing the conformational space of *keto*, the bond α was rotated with a step size of 180°. The bonds β and γ were rotated with a step size of 60° each whereas η was fixed at 0° and 60°. All combinations of these rotatable bonds α, β, γ, and η resulted in 144 initial structures for screening purposes. When probing the conformational space of *enol*, γ and ε were rotated with a step-size of 60° and 180°, respectively, and β was kept fixed at 0° and 90° and η fixed at 0° and 60° yielding 42 initial structures.

Electronic structure calculations for dipole-bound anions require basis sets with very diffuse basis functions. In this study, I used the standard Dunning's ADZ[115, 121] basis set supplemented with extra diffuse basis functions centred on the carbon of (-CH$_3$). The exponents α_n of these basis functions were determined through: $\alpha_n = \frac{\alpha_0}{q^n}$, $n = 1, 2, \ldots$, initiated from the lowest exponent α_0 of the s-, p- or d-functions in the standard ADZ basis set, and advanced with $q = 3.2$.[165] I limited the extra diffuse set to 5 s-, 5 p-, and 2 d-functions based on the values of the LCAO coefficients of the singly occupied molecular orbital in the dipole-bound state calculated at the geometry of the neutral.[122, 129] This procedure led to the lowest exponents of $1.3256 \times 10^{-4} au$, $1.1421 \times 10^{-4} au$ and $1.4437 \times 10^{-2} au$ for the s-, p- and d-symmetries, respectively. The dipole moment of the neutral increases upon the intra-molecular proton transfer from O2 to O1, thus the above set of additional diffuse functions should be sufficient for all geometries considered.

The initial structures of the neutral and valence anions of acetoacetic acid were pre-screened at the density functional level of theory with the B3LYP exchange-correlation functional.[100, 111–113] All minima were re-optimised at the coupled-cluster singles and doubles[206] (CCSD) level of theory; harmonic frequencies were calculated at the same level. The minimum energy structures contained within an energy range of 10 kcal/mol were subjected to single-point coupled-cluster calculations with single, double and non-iterative triple excitations[188] (CCSD(T)). Thus my most accurate energies of the neutral and anionic systems are based on the CCSD(T) electronic energies and
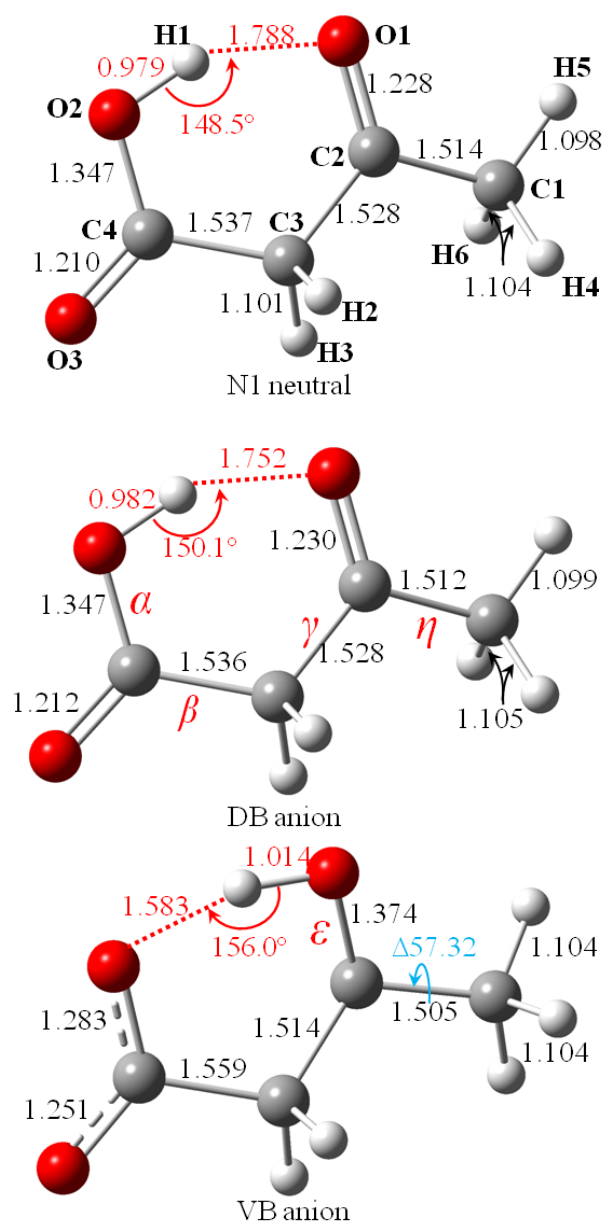
FIGURE 7.1: Molecular structures of AA considered in this study. The principal geometrical parameters (in Å) were characterized at CCSD/ADZ + DF level of theory. Bonds $\alpha, \beta, \gamma, \eta$ and ε were rotated for conformational searches. The change Δ in the O1–C2–C1–H5 dihedral angle (blue) from the N1 to VB geometry is in ° (deg).
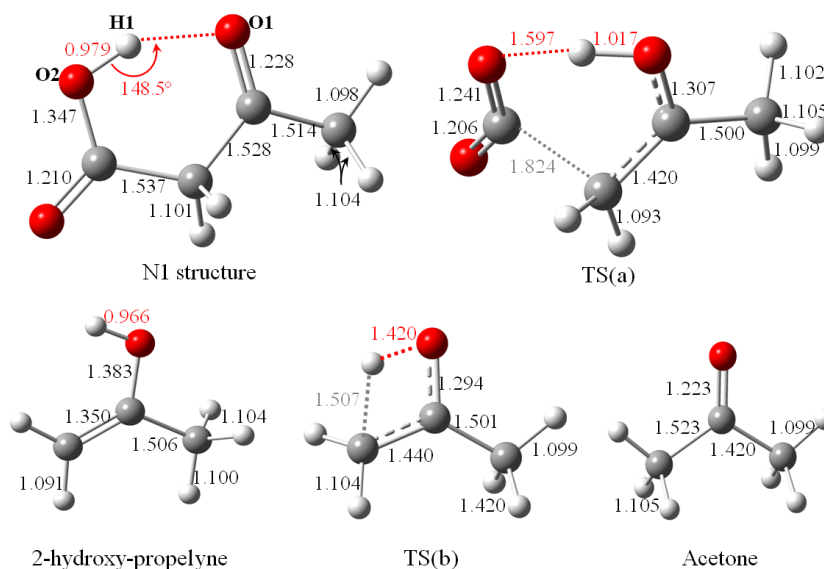
FIGURE 7.2: The unimolecular decomposition path of AA characterised at the CCS-D/ADZ level of theory. The bond H1 – O1 was scanned and the transition states TS(a) and TS(b) illustrate the unimolecular decomposition path.

zero-point vibrational energies determined at the CCSD level, both obtained with the ADZ+DF basis set.

I identified several low-lying conformers of the neutral *keto* AA and the energy barriers separating them. The search for transition states was performed in two stages. First, I constructed the energy profiles (MP2 or B3LYP) for rotations around α or β by performing partial geometry optimizations for several fixed values of the selected dihedral angle. The geometries associated with maxima on these profiles were used as starting geometries for transition state searches using the Berny algorithm.[207–209] I also considered the decomposition of neutral AA into CO_2 and 2-hydroxy-propelyne (the latter further transforms to acetone, see Fig. 7.2). Again, initially I constructed energy profiles for the relevant proton transfer reactions followed by transition state searches using the Berny algorithm.[207–209]

The **N1** conformer of neutral AA (see Fig. 7.3) was the most promising candidate for supporting a dipole-bound anionic state: (i) it is the most stable neutral conformer in terms of electronic energy, (ii) its dipole moment of 5.4 D exceeds dipole moments of other low-energy conformers by 1.8 D. Thus my discussion of the dipole-bound anionic state is limited to this conformer, though I also considered evolution of this anionic state along the proton transfer coordinate.

The vertical excess electron binding energies were calculated it two ways. In "indirect" approaches, the energy of the anion was subtracted from the energy of the neutral and the procedure was executed at the SCF, MP2, CCSD, and CCSD(T) levels of theory.
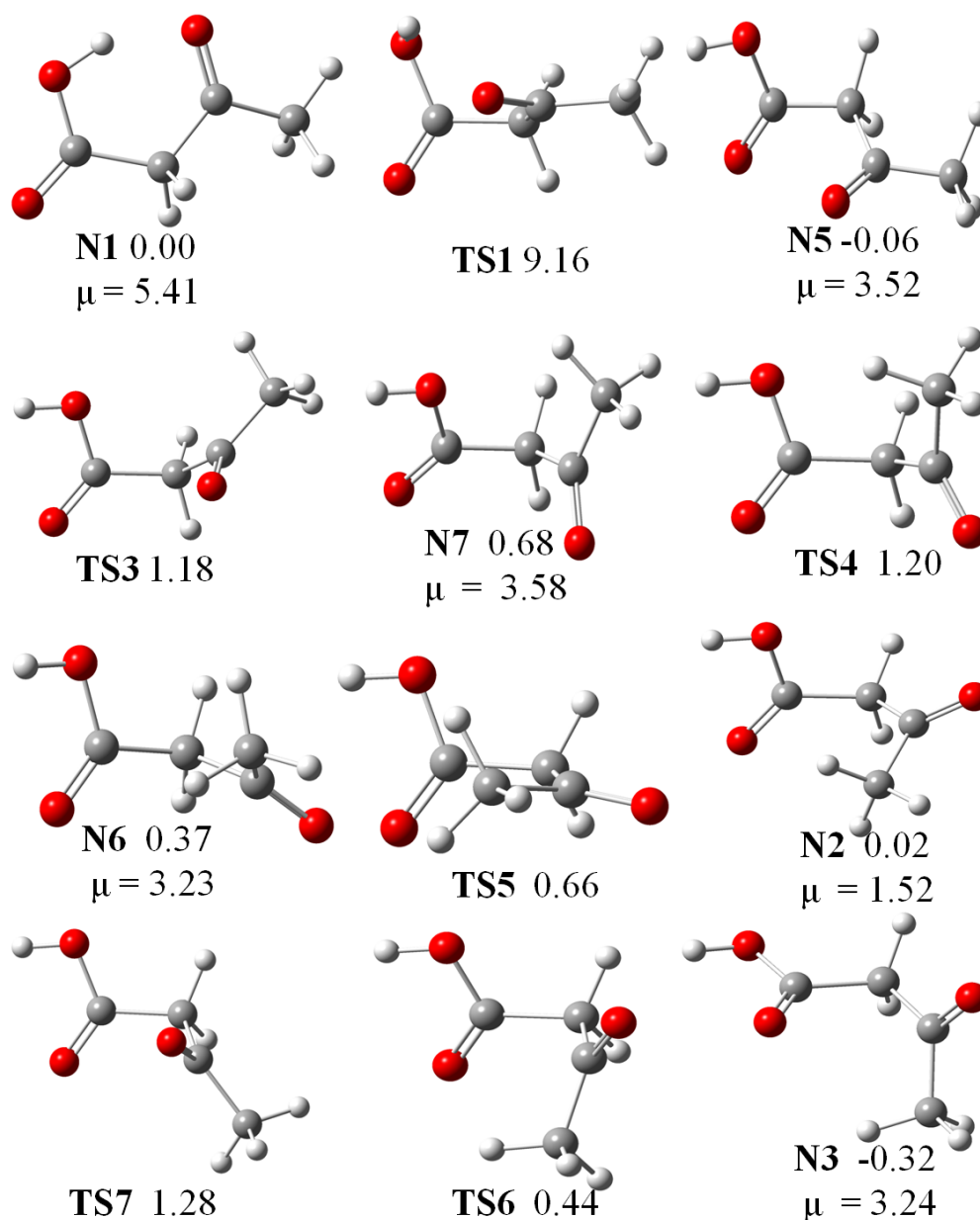
FIGURE 7.3: The low-lying neutral conformers of AA with the interconnecting barriers separating them. The relative energies (in kcal/mol) were obtained at the CCSD(T)/ADZ level of theory and corrected for MP2/ADZ zero-point vibration energies; the dipole moments μ are in Debyes.

The vertical excess electron binding energy can be also calculated "directly". Here I used the Electron Propagator Theory (EPT) (with P3 propagator)[210] applicable to both electron detachment and attachment processes in which the frozen-orbitals, single-determinant picture of Koopmans' theorem (KT)[211] is valid. One can calculate the vertical excess electron binding energy as either Electron Affinity (EA) of the neutral or Ionization Potential (IP) of the anion. I will use *EPT* to refer to third-order electron binding energies.

In order to illustrate the evolution of the **DB** and **VB**, anionic states as a function of intramolecular proton transfer, I constructed $\lambda - 1$ intermediate structures. By using the same set of geometric variables β_k (inter-atomic distances, angles and dihedral angles, $k = 1, \ldots, 3N - 6$), I defined the set of intermediate structures as:

$$\beta_k^m = \beta_k^i + m\delta_k \qquad m = 1\ldots, \lambda - 1 \tag{7.1a}$$

$$\delta_k = \frac{\beta_k^f - \beta_k^i}{\lambda}, \tag{7.1b}$$

where β_k^f and β_k^i are the values of the kth variable for the final (**VB**) and initial (**DB**) structure, respectively.

The energy of the anion M^- at a geometry **G** can be written as;

$$E_{M^-}(G) = E_M(G_M) + \Delta E_M(G) - EBE(G), \tag{7.2}$$

where $E_M(G_M)$ is the energy of the neutral M at its optimal geometry G_M,

$$\Delta E_M(G) = E_M(G) - E_M(G_M) \tag{7.3}$$

represents an increase of the energy of the neutral M associated with its geometrical deformation from G_M to G, and $EBE(G)$ is the vertical electron binding energy at the geometry G:

$$EBE(G) = E_M(G) - E_{M^-}(G). \tag{7.4}$$

The values of *EBE* are positive for vertically bound anionic states and negative for resonances. The electron vertical attachment energy (VAE) is equivalent to $EBE(G_M)$. The electron vertical detachment energy (VDE) is equivalent to $EBE(G_{M^-})$ and the electronic part of the adiabatic electron affinity is defined as:

$$AEA = E_M(G_M) - E_{M^-}(G_{M^-}), \tag{7.5}$$

where G_{M^-} is the optimal geometry on the anion of M. Notice that

$$AEA = -\Delta E_M(G_{M^-}) + EBE(G_{M^-}) = -\Delta E_M(G_{M^-}) + VDE. \qquad (7.6)$$

Further extensions of this notation are needed for AA, which supports a valence and dipole-bound anions. In this study I will characterize three characteristic molecular structures. In addition to G_M I will consider $G_{M^-_{dbs}}$ and $G_{M^-_{VB}}$ which are the lowest energy structures for the dipole- and valence-bound anions of M, respectively. The labels of VAE, VDE and AEA will be supplemented with a subscript *dbs* or VB to discriminate between the two anionic states. I will also consider the quantity:

$$\Delta E_{VB \rightarrow dbs} = E_{M^-_{dbs}}(G_{M^-_{VB}}) - E_{M^-_{VB}}(G_{M^-_{VB}}), \qquad (7.7)$$

which is the vertical excitation energy from the valence to the dipole-bound anion at the optimal geometry of the M^-_{VB}. This quantity might be important in future experiments on the anion of acetoacetic acid.

All electronic structure calculations reported in this study were completed with the Gaussian 2009 package[75]. The orbitals occupied by an excess electron were generated with the Visual Molecular Dynamics[167] package and the contour values used in the plots were calculated with the OpenCubeMan[212] tool using a fraction of electron (F_e) density equal to 0.6. The GaussView[189] package was used to draw molecular structures.

Experimental procedure to obtain experimental photoelectron spectrum is discussed in appendix C.

7.4   Computational Results

7.4.1   Neutral AA

My results indicate that the neutral AA supports minima for the *keto* tautomer only (Fig. 7.3). My attempts to identify an *enol* minimum energy structure with H1 bound to O1 failed: the H1 proton either transferred back to O2 or the molecule broke into 2-hydroxyl-propelyne and CO_2 (Fig. 7.2). Notice that the proton-transferred structure, with the H1 proton bound to O1, is the global minimum for the anion of AA (the bottom part of Fig. 7.1).

The low-lying minimum energy structures of the canonical neutral ($\mathbf{N1} - \mathbf{N7}$) and transition states that separate the minima ($\mathbf{TS1} - \mathbf{TS7}$) are illustrated in Fig. 7.3
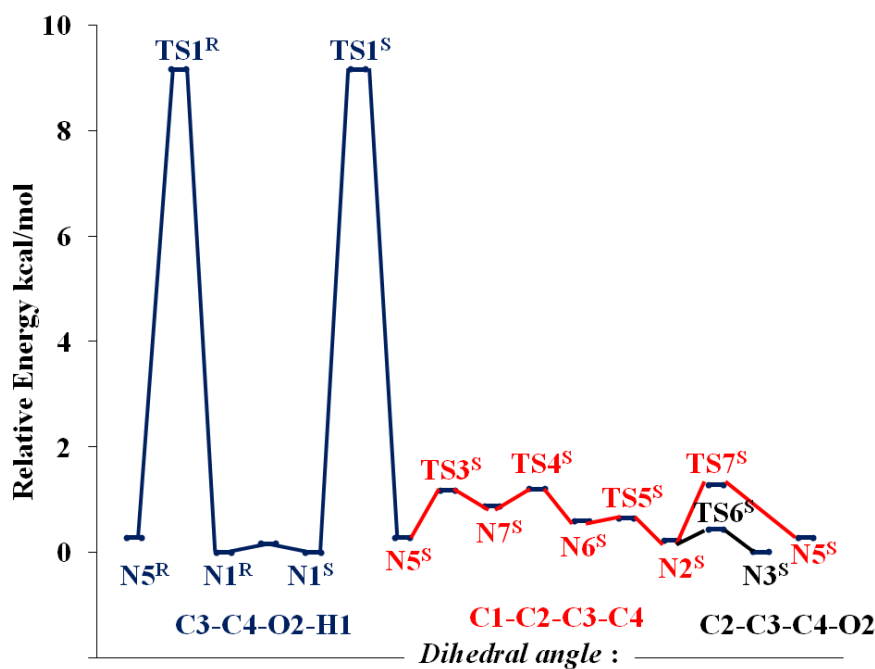
FIGURE 7.4: Energy profile connecting neutral structures of AA. The energies (in kcal/-mol) were calculated at the CCSD(T)/ADZ level using the CCSD/ADZ geometries.

and a simplified landscape of the potential energy surface is illustrated in Fig. 7.4. The stationary points exist in two enantiomeric forms. One could anticipate the **N1** conformer to be uniquely stable due to the intramolecular O2H1$\cdots$O1 hydrogen bond, but the relative energies listed in Fig. 7.3 indicate that **N1** is not necessarily the most stable conformer. There are at least five neutral conformers within a narrow energy range of 1 kcal/mol. **N1** is unique not so much in its stability but in the energy barrier (**TS1** at 9.2 kcal/mol) that separates it from **N5** and the remaining conformers. This barrier can be associated with breaking of the intramolecular hydrogen bond. The barriers separating the conformers **N2** − **N7** are much smaller and do not exceed 1.1 kcal/mol.

There are at least two factors that oppose the stabilizing effect of the intramolecular hydrogen bond in **N1**. First, **N1** is the most polar conformer, with a dipole moment exceeding 5.4 D. The remaining conformers have dipole moments smaller by more than 1.8 D. High polarity typically increases electronic energy of a neutral molecule by raising the energy of the highest occupied orbital. Second, formation of the O2H1$\cdots$O1 hydrogen bond is associated with intramolecular strain. The results obtained with the Amber[221] force field confirm this hypothesis. The sum of the angle bending, torsional and bond stretching terms is larger in **N1** than in e.g., **N5**, by ca. 2.8 kcal/mol, an energy term comparable with the stabilizing effect of the hydrogen bond.

The neutral AA is unstable with respect to the decomposition to acetone and CO_2 (Fig.
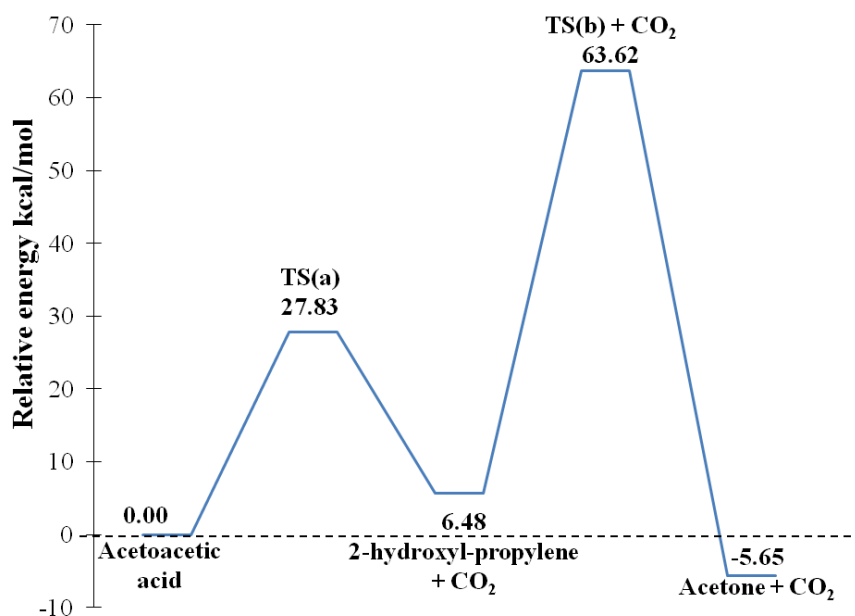
FIGURE 7.5: Dissociation energy profile of neutral AA. The relative energies were computed at CCSD(T)/ADZ level of theory using CCSD/ADZ optimal geometries.

7.5). The products are stable with respect to the reactant by 5.7 kcal/mol in terms of electronic energies corrected for zero-point vibrations. The entropic effects will further enhance the stability of products. Here I considered only unimolecular decomposition channels. The first step of decomposition is associated with intramolecular proton transfer and leads to CO_2 plus 2-hydroxyl-propylene. The barrier height (**TS(a)** in Figures 7.2 and 7.5) is 27.8 kcal/mol and is comparable with the previous MP2 value of 25.7 kcal/mol.[203]. The unimolecular transformation of 2-hydroxyl-propylene to acetone encounters a high barrier of 57.1 kcal/mol (the past MP2 result was 55.9 kcal/mol[203]). The decomposition of AA in the liquid phase is not limited to unimolecular channels and is expected to proceed with lower barriers. The thermodynamic instability of AA is consistent with experimental findings.[202]

7.4.2 Anionic AA

Neutral molecules with dipole moments exceeding 1.625 D can bind an excess electron.[191] **N1** is the most promising conformer to host a dipole-bound anionic state in view of its competitive stability and dominant polarity ($\mu = 5.4$ D). In addition to the dipole-bound state, AA can support a valence anion.

An overview of anionic states of AA is presented in Figure 7.6 using a set of geometries defined by Eq. 7.1. This series of geometries connects the dipole-bound minimum based on the **N1** structure of the neutral with the minimum of the valence-bound anion. The
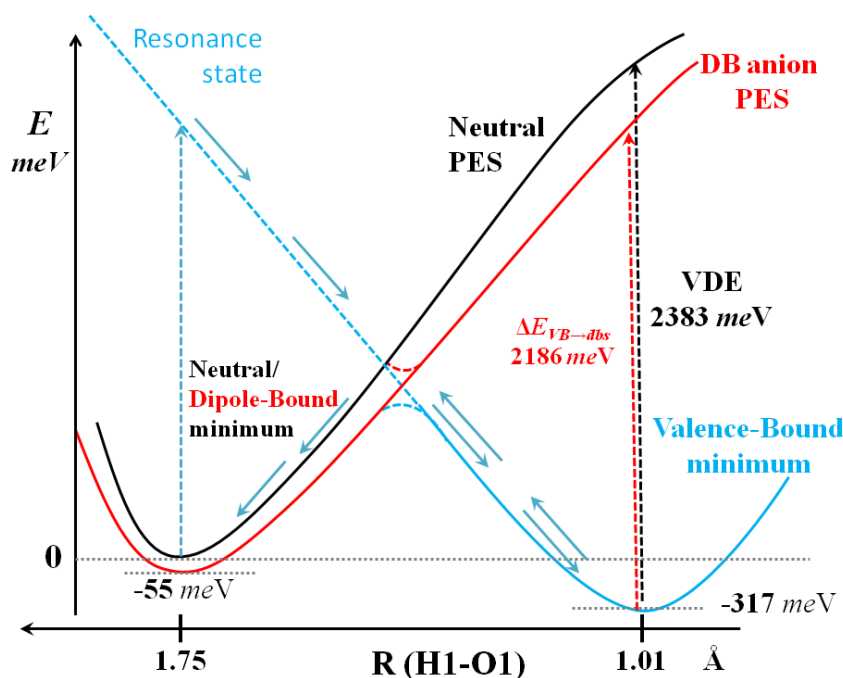
FIGURE 7.6: Energy profile depicting the neutral, dipole-bound and valence anionic potential energy surfaces of AA. The energies (meV) computed at the CCSD(T)/ADZ+DF level.

potential energy curve of the neutral is repulsive upon transferring the proton from O2 to O1. The dipole moment of the neutral increases upon proton transfer from 5.4 D at the minimum of the neutral to 10.0 D at the minimum of the valence anion. Thus the dipole-bound anion remains bound upon proton transfer. The valence anion is unbound at the minimum energy structure of the neutral; it can be probed as a resonance state with a finite lifetime in electron energy loss experiments (see Section 7.5.2). The energy of the resonance quickly decreases upon intramolecular proton transfer, crosses the potential energy surface of the neutral, and undergoes an avoided crossing with the dipole-bound anion. With the proton transferred from O2 to O1, the valence anion supports a minimum characterized by a significant electron vertical detachment energy of 2383 meV. The valence anion is adiabatically bound with respect to the neutral by 317 meV (in terms of electronic energies). I re-emphasise that the dipole-bound anion remains bound at the minimum of the valence-bound anion and the electron excitation energy from the valence- to dipole-bound state is 2186 meV. The singly-occupied molecular orbitals in the dipole-bound and valence anions, each at its minimum energy structure, are illustrated in Figure. 7.9. The former orbital is very diffuse and localized on the positive pole of the molecular dipole, which proves to be a hydrophobic group CH_3. The latter is a π orbital localized around the C2 atom, with bonding interactions involving the neighbouring carbon atoms and an antibonding interaction between C2 and O1.
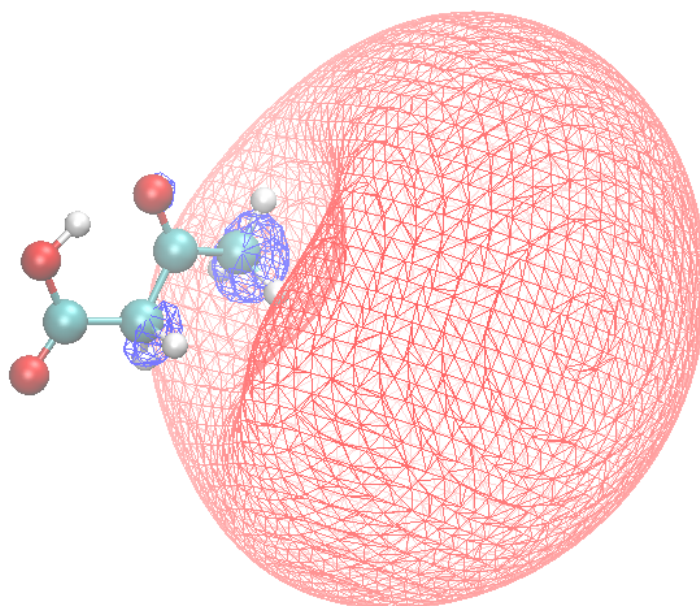
FIGURE 7.7: Dipole bound
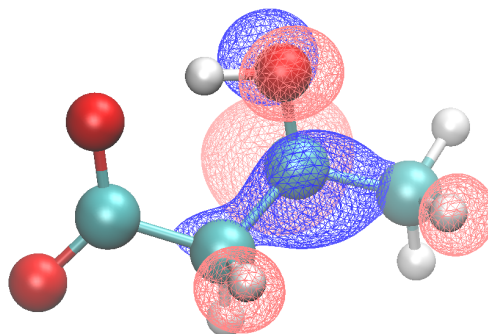
FIGURE 7.8: Valence bound

FIGURE 7.9: The orbital occupied by an excess electron in the DB and VB anions of AA. The plots were generated using VMD with contour values calculated using the OpenCubeMan package from 60% of electron density of these respective orbitals.

The minimum energy structure of dipole-bound anion differs from the neutral by shortening the O1$\cdots$H1 distance by 0.036 Å, increasing the O1–H1–O2 angle by 1.6° and an elongation of the O2$\cdots$H1 distance by 0.003 Å, see Figure 7.1. Overall it is a small step towards proton transfer, a nascent zwitterionization. The dipole moment of the neutral increases by 0.16 D upon these geometrical distortions, a typical increase for dipole-bound anionic states.

The vertical electron binding energies for the dipole-bound anionic state, obtained with indirect and direct methods, are reported in Table 7.1 for the minimum energy geometries of **N1**, and dipole- and valence-bound anions. The "indirect" electron binding energies are reported at the SCF and correlated levels of theory. The SCF contribution represents only 46-65% of the CCSD(T) results, illustrating the role of electron correlation effects in dipole-bound anionic states. The differences between the CCSD and CCSD(T) results do

Method		G_M	$G_{M_{dbs}^-}$	$G_{M_{VB}^-}$
			EBE	
Indirect	SCF	24.06	26.89	127.68
	MP2	32.81	36.31	217.72
	CCSD	52.34	56.59	196.85
	CCSD(T)	52.19	56.55	197.04
Direct EA	KT	21.59	24.33	113.30
	EPT	43.63	47.46	245.07
Direct IP	KT	26.32	29.78	135.79
	EPT	45.27	50.45	247.84

TABLE 7.1: Vertical electron binding energies (meV) of the dipole-bound anionic state at the G_M, $G_{M_{dbs}^-}$ and the $G_{M_{VB}^-}$ geometries using the CCSD/ADZ+DF optimal geometries.

	AEA (electronic)				ΔE_{vib}	AEA	
	HF	MP2	CCSD	CCSD(T)	CCSD	CCSD(T)	
DB	8.60	38.88	54.15	54.56	2.63	57.19	
VB	-182.80	286.26	289.55	317.02	17.27	334.29	

TABLE 7.2: Adiabatic electron affinities (meV) of the DB and VB anions calculated at different levels of theory with the CCSD optimal geometries. All calculations with the ADZ+DF basis set.

not exceed 0.2 meV suggesting methodological saturation of my results. The electron binding energies obtained in "direct" approaches start from the Koopmans' theorem estimations, thus they neglect orbital relaxation and electron correlation effects. The differences between the KT and SCF "indirect" results are small indicating that orbital relaxation effects are relatively small for the dipole-bound anionic state. The *EPT* results are in good agreement with the "indirect" CCSD(T) results, further confirming the reliability of electron binding energies for this anionic state. Various estimations of adiabatic electron affinity for this anionic state are reported in Table 7.2. The CCSD(T) electronic energy result of 54.56 meV is methodologically converged and the zero-point vibrational correction further stabilizes the anion by 2.63 meV. This dipole-bound anion is amenable to experimental investigations using Rydberg electron transfer methods[222]. Probing this anionic state in anion photoelectron spectroscopy experiments[223] might be more problematic due to the possibility of collapsing to the global minimum of valence anion (Fig. 7.6). Finally, the decomposition of electronic component of AEA into the $\Delta E_M(G_{M^-})$ and VDE terms (Eq. 7.6, Table 7.3), illustrates a very small geometrical distortion of the molecular framework upon the excess electron attachment and the dominant role of VDE.

Str.	6-311G*		ADZ + DF			
	E_{elec}^{B3LYP}	E_{elec}^{B3LYP} $+E_{0,vib}^{B3LYP}$	E_{elec}^{B3LYP}	E_{elec}^{B3LYP} $+E_{0,vib}^{B3LYP}$	E_{elect}^{MP2}	E_{elec}^{MP2} $+E_{0,vib}^{MP2}$
VB1	0.00	0.00	0.00	0.00	0.00	0.00
VB2	14.87	14.56	13.56	13.50	12.65	12.56
VB3	16.73	16.36	15.33	15.14	15.84	15.68

TABLE 7.4: The Relative Electronic Energies (kcal/mol) of valence anions of acetoacetic acid calculated at different levels of theory.

	$-\Delta E_M(G_{M^-})$	VDE	AEA
DB	-1.99	56.55	54.56
VB	-2065.69	2382.71	317.02

TABLE 7.3: The electronic component (CCSD(T)/ADZ+DF) of AEA (meV) decomposed into the $\Delta E_M(G_{M^-})$ term and VDE.

In contrast to the neutral AA, which supports several low-lying minimum energy structures, the valence anion of AA supports one distinct global minimum illustrated in the bottom of Fig. 7.1. The other local minima were less stable by more than 12 kcal/mol, and are characterized in table 7.4 and figure 7.13. The H1 proton is bound to O1 and the O1H1$\cdots$O2 hydrogen bond is very short, R(H1O2)=1.583 Å, in the global minimum structure of valence anion. The intramolecular hydrogen bond is more linear than in the neutral or dipole-bound anion by 6–8°. When compared with the N1 neutral, there is a significant elongation of the C2O1 distance by 0.146 A and a shortening of C4O2 by 0.064 Å, consistent with a redistribution of double bonds upon tautomerization. Finally, the CH$_3$ group is rotated by ca. 58° in comparison with N1.
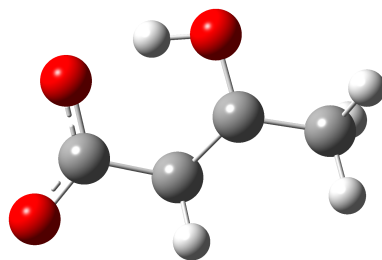
FIGURE 7.10: VB1 -8.35, 178.2°
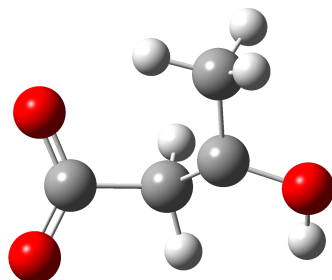
FIGURE 7.11: VB2 4.21, $-62.8°$
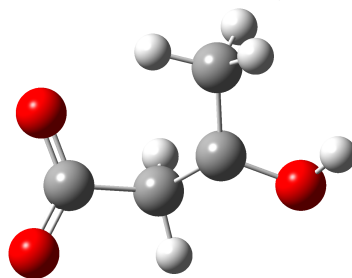
FIGURE 7.12: VB3 7.33, $-59.8°$

FIGURE 7.13: The principal geometrical parameters (the C1-C2-C3-C4 dihedral angle in °) and the relative energies (in kcal/mol) of the structures considered in this study, with respect to **N1** structure, characterized at MP2/ADZ + DF level of theory.

The values of electron VDE obtained using "indirect" methods (including SCF) and "direct" EPT span a reasonably narrow range of 2300–2600 meV (Table 7.5). The KT(EA) value is only 713 meV while the KT(IP) value is 4200 meV, demonstrating that orbital relaxation effects are critically important for this anionic state. Electron correlation effects, on the other hand, do not contribute much to the VDE value. The anion should be readily formed in common experimental sources, because its adiabatic electron affinity is modest but positive, 334 meV after inclusion of the zero-point vibrational correction (Table 7.2). The value of electronic contribution to AEA can be analysed in terms of Eq. 7.6 (Table 7.3). The proton transfer from O2 to O1 is accompanied by a significant increase of the energy of the neutral ($\Delta E_M(G_{M_{VB}^-}) = 2066 \ meV$). This energy increase is, however, outweighed by the VDE of 2383 meV. The modest electronic contribution to AEA of 317 meV results from cancellation of the $\Delta E_M(G_{M_{VB}^-})$ and VDE terms. Thus the situation is different than in the dipole-bound anion, for which the $\Delta E_M(G_{M_{dbs}^-})$

Method	VDE	
Indirect	SCF	2400.48
	$MP2$	2347.34
	$CCSD$	2482.36
	$CCSD(T)$	2382.71
Direct EA	KT	713.43
	EPT	2304.34
Direct IP	KT	4200.09
	EPT	2602.50

TABLE 7.5: The values of VDE (meV) for the valence anion of AA.

term was small in comparison with the VDE value. The anion of AA supports more than one bound electronic state at the geometry $G_{M_{VB}^-}$. The electronic excitation energy from the valence to dipole-bound state is 2186 meV. In future experiments, one might want to probe the molecular dynamics of AA^- upon this electronic excitation.

Contrary to my experience with the neutral AA (Figures 7.2 and 7.5), I did not witness any propensity of the valence anion of AA to unimolecular decomposition. The decomposition products CO_2 plus acetone are thermodynamically less stable than AA^-, and the AEA of CO_2 is negative[224]. Acetone, with a dipole moment of 2.91 D, supports only a very weakly bound state (AEA ¡ 2 meV).[225] Unless other decomposition channels prove to be thermodynamically favourable, I suggest that AA can be stabilized with respect to the spontaneous decomposition[202] through formation of the valence anion. Whether such an anion can exist in condense phases will be considered in my future studies.

7.5 Experimental results

7.5.1 Photoelectron spectroscopy results for AA^- and $(AA)_2^-$

Parent anions of AA were prepared by my collaborators using two rather different source environments. Figure 7.14 presents a representative mass spectrum of the anions produced.

The photoelectron spectrum of the AA parent anion measured on the continuous photoelectron spectrometer is presented in Figure 7.15. This spectrum consists of a broad, featureless band with an onset at EBE ~1.2 eV and an intensity maximum at EBE 2.04 eV. This latter quantity corresponds to the VDE of the AA parent anion. The calculated adiabatic electron affinity of 0.33 eV is much smaller than the onset indicating that the
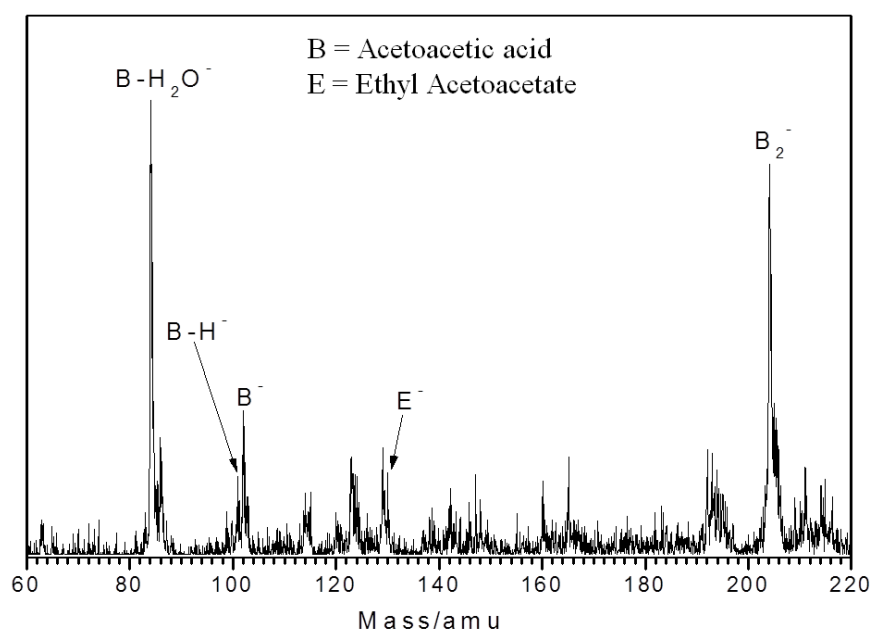
FIGURE 7.14: Mass spectrum of anions observed in these experiments.
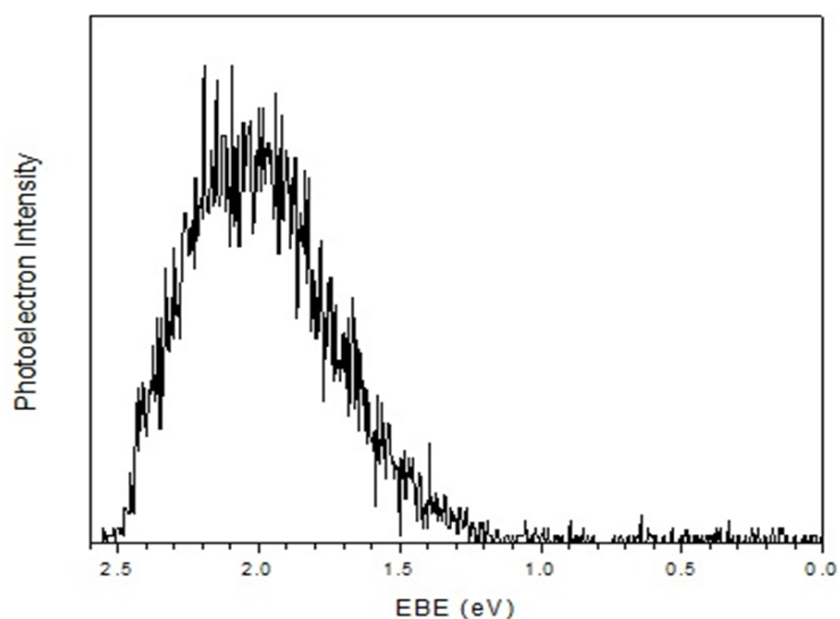
FIGURE 7.15: Photoelectron spectrum of the AA parent anion recorded with 2.54 eV photons on the continuous photoelectron spectrometer.

0-0 transition has a very small intensity. The calculated VDE of 2.38 eV is larger by 0.34 eV than the measured intensity maximum. The source of this discrepancy will be analysed further on.

The photoelectron spectrum of the AA parent anion measured on the pulsed photoelectron spectrometer is presented in Figure 7.16. This spectrum also consists of a broad,
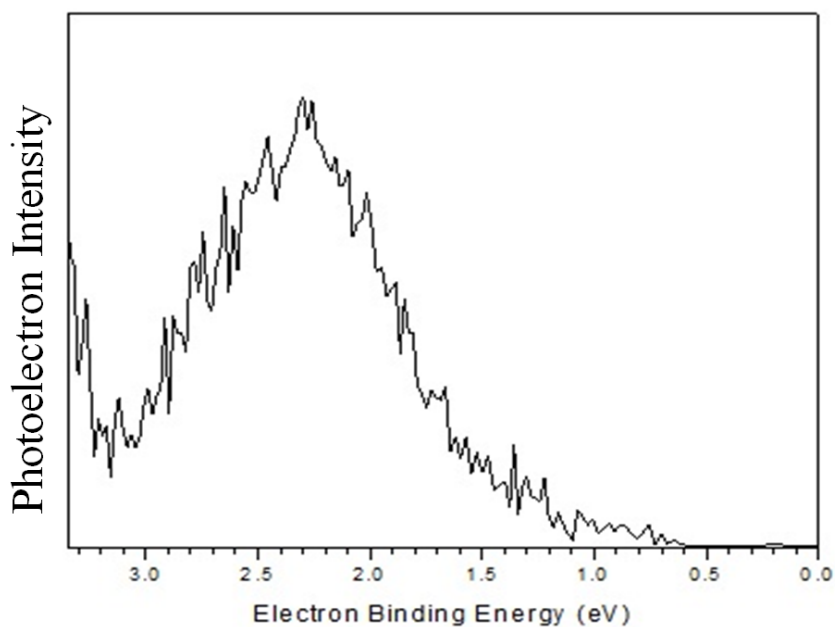
FIGURE 7.16: Photoelectron spectrum of the AA parent anion recorded with 3.49 eV photons on the pulsed photoelectron apparatus.

featureless band with an onset at EBE 1.2 eV but with an intensity maximum at EBE 2.30 eV. This latter quantity again corresponds to the VDE of the AA parent anion. This value of VDE is in excellent agreement with my calculated value of 2.38 eV.

For the most part the two photoelectron spectra are the same, indicating that both source environments produced the same AA parent anion species. The main difference lies in the slightly different VDE values obtained on different apparatus. The difference between the two measured VDE values is no doubt due to the electron transmission function roll-off inherent for low kinetic energy (high EBE) electrons. This effect would come into play more strongly for a lower versus a higher photon energy, since a lower photon energy puts more of the spectrum in the low EKE region of the spectrum. On the continuous apparatus, whose photon energy was limited to 2.54 eV/photon, this had the effect of cutting into the still rising, photoelectron intensity at the top of the peak, making its intensity maximum appear to occur at a slightly lower EBE value. By contrast, the pulsed apparatus, which utilized a photon energy of 3.49 eV/photon, was not significantly affected in this way. Therefore, I believe that the better value for the VDE of the AA parent anion is 2.3 eV, as predicted by theory.

My collaborators also measured the photoelectron spectrum of the parent dimer anion of AA, and it is presented in Figure 7.17. This spectrum also exhibits a single broad band, and it has a similar width to that of the monomeric AA anion. Interestingly, its VDE value, at 1.7 eV, is considerably lower than that of the monomer, indicating that
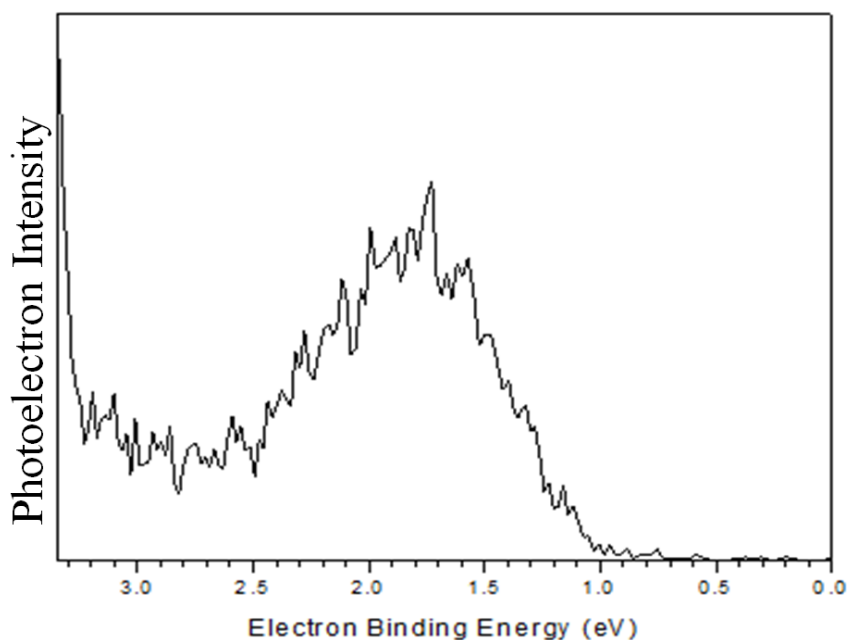
FIGURE 7.17: Photoelectron spectrum of the AA dimer (parent) anion recorded with 3.49 eV photons on the pulsed photoelectron apparatus.

the additional AA molecule is not simply solvating the AA anion, i.e., it is not a simple anion-molecule complex.

7.5.2 Electron energy loss spectroscopy results for AA

Figure 7.18 shows an electron energy-loss spectrum recorded by varying the incident electron energy and collecting slow scattered electrons with residual energies of 25 meV (bottom), 50 meV (middle), and 1.8 eV (top). The spectrum is plotted as a function of the energy loss – the elastically scattered electrons appear on the left; electrons leaving with smaller than incident energies on the right. The numerous peaks in the left part of the plots are associated with excitations of specific vibrational levels of the neutral target. For the residual energies of 25 and 50 meV, notice a peak of very low energy electrons. The observation where electrons leave nearly all of their energy in the internal modes of a polyatomic target without exciting any specific vibrational modes was made already in the formic acid monomer.[226]. Here, however, the relative intensity of this peak, when compared with the peaks associated with excitations of specific vibrational modes, is significantly larger than in the formic acid monomer, but significantly smaller than in the formic acid dimer. The existence of this peak indicates that this system is capable to quasithermalise electrons arriving in the 1-2 eV energy range.

The appearance of this peak can be explained in terms of the results of quantum chemical calculations reported in Sec. 7.4.2. The initial resonant state of the anion formed
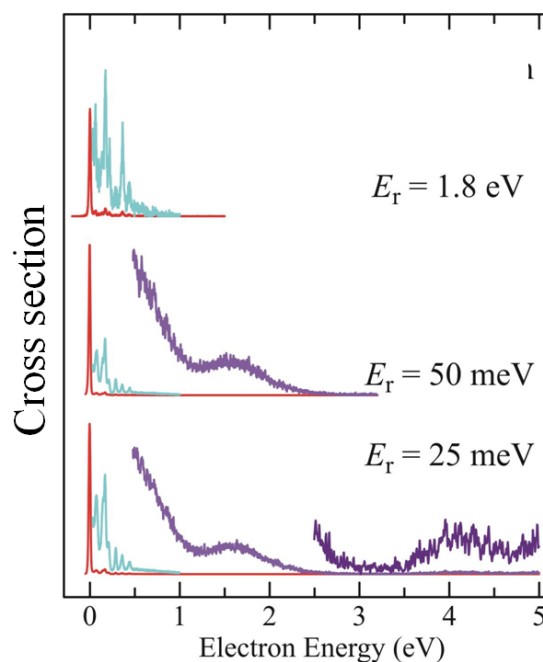
FIGURE 7.18: Cross sections for the excitation of specific vibrational modes (red and blue) and electron loss energy spectra (magenta) as a function of the incident electron energy.

by a vertical electron attachment is stabilised by a rapid proton transfer (Figure 7.6), thus quenching the fast autodetachment channel, which would otherwise dominate the scattering and lead to elastic scattering and selective excitation of vibrational levels. The system may either lose a slow electron at the crossing of the anionic and neutral surfaces or proceed toward the minimum of the valence anion (see blue arrows in Figure 7.6). The anion has enough vibrational energy to statistically return back to the crossing point, where a detachment of a slow electron may occur again. This mechanism thus represents a source of slow electrons and is capable of rationalizing the experimental observation reported in Figure 7.18. The proton transfer is particularly efficient in competing with the very fast electron autodetachment, because it proceeds without any activation barrier (Figure 7.6) and because the low mass causes protons to move rapidly.

The relative intensity of the peak associated with low-energy electrons is clearly smaller than in the formic acid dimer but larger than in the formic acid monomer. I will interpret these findings in terms of my computational results for the neutral AA reported in Section 7.4.1. Among several low-lying conformers only **N1** supports the intramolecular hydrogen bond. This conformer is prearranged for the intramolecular proton transfer from O2 to O1 upon formation of the valence resonance state (Figure 7.3). All other low-lying neutral conformers undergo primarily elastic collisions with excess electrons or inelastic collisions with excitations of specific vibrational modes, but the lifetime of the valence anionic resonance is too short to allow for intramolecular proton transfer

that quenches the autodetachment channel and leads to ejection of very slow electrons. The quasithermalisation of excess electrons was more efficient in the formic acid dimer than in AA, because the global minium of the former is supported by a system of cyclic hydrogen bonds prearranged for intermolecular proton transfer and other minima, not prearranged for intermolecular proton transfer, are much less stable than the global mimimum, thus not populated in the beam of neutral dimers crossed with electrons. On the other hand, the presence the **N1** conformer in the beam of neutral AA explains why the peak associated with low energy electrons is more intense in AA than in the formic acid monomer.[198, 226]

7.6 Summary

I performed search for the most stable tautomers and conformers of the neutral and anionic acetoacetic acid using a potential energy surface scanning tool (PESST).[29] I identified several conformers of the canonical neutral within a narrow energy range of 1 kcal/mol. The conformer with the intramolecular hydrogen bond, and thus susceptible to intramolecular carboxylic-to-keto proton transfer, proved to be the third most stable and the most polar, with a dipole moment of 5.4 D. This polar conformer is separated from other conformers by a relatively high barrier exceeding 9 kcal/mol. The barriers separating other conformers are much smaller and do not exceed 1.1 kcal/mol. The neutral acetoacetic acid was found unstable by 5.7 kcal/mol with respect to acetone and CO_2.

The polar neutral conformer supports a dipole-bound anion with an electron vertical detachment energy of 57 meV. The dipole-bound anion remains bound upon the intramolecular carboxylic-to-keto proton transfer but its energy increases steadily and I failed to identify a minimum for the zwitterionic enol structure.

The valence anions of acetoacetic acid all exist in the enol form. The distinct global minimum is more stable than other minima by more than 12 kcal/mol. It is supported by a short (H$\cdots$O distance of 1.58 Å) intramolecular hydrogen bond between the enol and carboxylate groups. I conclude that binding an excess electron on a π^* valence orbital changes localization of the proton in acetoacetic acid. The valence anion is characterized by a significant electron vertical detachment energy of 2383 meV, but a modest adiabatic electron affinity of 334 meV. The electronic excitation energy from the valence to dipole-bound state is 2186 meV. Thus the dynamics of the anion excited to the repulsive wall of the dipole-bound state can be explored in future experiments. The valence anion of acetoacetic acid is stable, in terms of electronic energies corrected for zero-point vibrations, with respect to acetone and the anion of CO_2.

Parent anions of acetoacetic acid were successfully prepared using two different source environments. The photoelectron spectra obtained on the continuous and pulsed photoelectron spectrometers were obtained with 2.54 and 3.49 eV photons, respectively. The PES spectra consist of a broad, featureless band with an onset at electron binding energy of ca. 1.2 eV. The intensity maxima on the continuous and pulsed photoelectron spectrometers differ by 0.26 eV (2.04 vs. 2.30 eV). Only the latter is in good agreement with my computed VDE of 2.38 eV. The difference between the two measured VDE values is due to the electron transmission function roll-off inherent to low kinetic energy electrons. The calculated adiabatic electron affinity of 0.33 eV is much smaller that the onset of the spectra indicating that the 0-0 transition has a very small intensity. The reported spectra do not provide evidence for the dipole-bound anion. The photoelectron spectrum of the parent dimer anion of acetoacetic acid shows a vertical detachment energy at 1.7 eV, thus considerably lower than that of the monomer, indicating that the additional monomer is not simply solvating the monomeric anion.

The electron energy-loss spectrum of acetoacetic acid shows a peak of low energy electrons. The relative intensity of this peak, in comparison with the peaks associated with excitations of specific vibrational levels of the neutral target, is significantly larger than for the formic acid monomer,[198, 226] but significantly smaller than in the formic acid dimer.[198] My computational results allow to rationalize why the quasithermalization of electrons is less efficient in acetoacetic acid than in the formic acid dimer. Among several low-lying conformers of the neutral molecule only the high-dipole conformer is prearranged for intramolecular proton transfer and quasithermalization of electrons. All other low-lying conformers scatter electrons with or without vibrational excitations, but the lifetimes of valence anion resonances are too short to allow for intramolecular proton transfer that quenches the autodetachment channel and leads to ejection of very slow electrons.

In the future, I will study the strength of coupling between the dipole- and valence-bound anionic states in the avoided crossing region (Fig. 7.6). I will also explore the chemical nature of parent anions of: dimer acetoacetic acid, ethyl acetoacetate, and dehydrated acetoacetic acid. All these species were observed in the anion mass spectrum (Fig. 7.14). Finally, I will consider stability of the valence anion of AA in condense phases.

Chapter 8

Intermolecular Interactions between Molecules in Various Conformational States: the Dimer of Oxalic Acid

8.1 Abstract

I considered stability of dimers formed by molecules that can exist in different conformational states. I have developed a protocol that allows the dissection of the total stabilisation energy into one-body conformational and deformational components and the two-body interaction energy term. Interplay between these components determines the overall stability of the dimer. The protocol has been tested on the dimers of oxalic acid. The global minimum energy structure I identified is stabilised by two inter- and four intramolecular hydrogen bonds, while the most stable structure identified in previous studies is supported by two inter- and three intramolecular hydrogen bonds. The latter structure proves to be less stable by 25 meV than the former. The global minimum stability results from a balancing act between a moderately attractive two-body interaction energy and small repulsive one-body terms. I have analysed zero-point vibrational corrections to the stability of various conformers of oxalic acid and their dimers. I have found that minimum energy structures with the most stabilising sets of hydrogen bonds have the largest zero-point vibrational energy, contrary to a naive anticipation based on red shifts of OH stretching modes involved in hydrogen bonds.

8.2 Introduction

Numerous computational studies have been devoted to the relative stability of molecular conformers of isolated molecules.[25, 29] Here I will consider the relative stability of $X_m Y_n$ complexes formed by the mth conformer of X and the nth conformer of Y. The stability of the $X_m Y_n$ complex results from an interplay between one-body terms (the monomer energies) and the two-body interaction energy term. The goal of this report is to formulate a computational protocol for the analysis of the relative stability of the $X_m Y_n$ complexes.

The protocol will be illustrated on dimers formed by oxalic acid (OA). The monomer of OA can exist in various conformational and tautomeric forms, [227–231] with the most stable forms **A** and **B** illustrated in Fig. 8.1. Many groups have found **A** to be more stable than **B**,[227–231] due to a more favourable system of intramolecular hydrogen bonds. Here I consider the dimer of OA. Many dimer structures are possible resulting from pairing miscellaneous conformers of OA through various sets of intermolecular hydrogen bonds.[228] Blair and Thakkar have recently studied 69 local minima on the potential energy surfaces of (OA)$_2$.[228] Their extensive search was carried out using a multi-stage search-and-screen method described in Ref. 232, with their final spin-component-scaled MP2 energies extrapolated to the basis set limit. Their findings were that combinations of the **A** and **B** monomers are very favourable; 6 of the 19 lowest-energy dimers were of this type.[228] Their most stable structure, labelled **AB** in the middle of Fig. 8.2, has two intermolecular and three intramolecular hydrogen bonds.
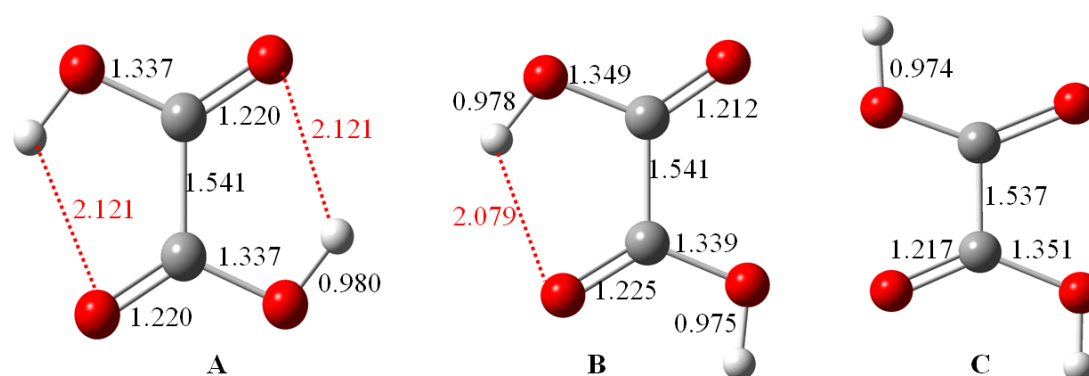
FIGURE 8.1: The principal geometrical parameters (in Å) of the **A**, **B**, and **C** conformers of the oxalic acid determined at the MP2/aug-cc-pVDZ level.

I agree with Blair and Thakkar that cyclic (double) hydrogen bonded structures are the main suspects for the global minimum of (OA)$_2$. Notice that each monomer can engage a proton donor and a proton acceptor site from either the same or neighbouring carboxylic group(s). The **BB** dimer (Fig. 8.2 bottom) exemplifies the first case; it
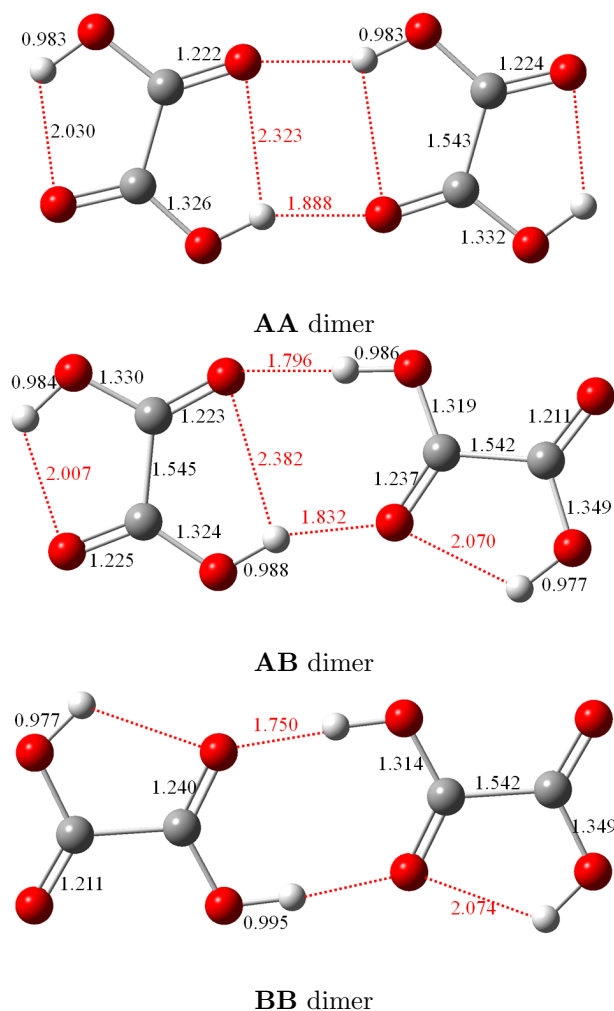
FIGURE 8.2: The principal geometrical parameters (in Å) of the oxalic acid dimers **AA**, **AB**, and **BB** determined at the MP2/aug-cc-pVDZ level.

resembles the well known structural motif of the formic acid dimer.[196, 199] The second case is exemplified by the **AA** dimer[1] (Fig. 8.2, top); it illustrates a unique ability of this dicarboxylic acid to engage in cyclic hydrogen bonds. Finally, the **AB** dimer (Fig. 8.2, middle), identified as the global minimum in Ref.228, represents a hybrid case.

In this report I will formulate a computational protocol for the analysis of stability of binary complexes formed by the monomers that can exist in various conformational states. I will analyse three dimers of OA: **AA**, **AB**, and **BB** (Fig. 8.2). The overall stability will be discussed in terms of the relative stability of monomeric conformers, the energetic penalty for deformation of these conformers to the geometry of the dimer, and the two-body interaction energy between the deformed conformers. This analysis unravels that the minimum on the potential energy surface associated with the **AA**

---

[1]Notice that my **AA** structure is different from the **AA** structure presented in Fig. 8.2 of Ref.228. My **AA** structure has two intermolecular hydrogen bonds, but the structure from Ref. 228 has one intermolecular hydrogen bond.

dimer is deeper by 25 meV than the minimum associated with the **AB** dimer. The inclusion of harmonic zero-point vibrations does not change the ordering, the **AA** dimer is the most stable form of $(OA)_2$ I identified.

8.3 Methods

I considered binary complexes formed by the molecules X and Y that can exist in various forms: X can exist as $X_0, X_1, X_3, \ldots$, and Y can exist as $Y_0, Y_1, Y_3, \ldots$, see Fig. 8.3. I assume that the stability decreases in each sequence, thus X_0 and Y_0 are the most stable forms of X and Y, respectively. In this report I will have in mind various conformers of X and Y, but the approach is also applicable to the tautomers of X and Y. The discussion presented here is limited to these geometries G of the $X + Y$ system that can be viewed as X interacting with Y, i.e., reactive parts of the potential energy surface are not considered. Then, for any geometry G, it makes sense to identify the X_m that resembles most the X monomer of the geometry G, i.e., $X(G)$. In the majority of cases, this identification is straightforward, in dubious cases one can use measures of similarity between $X(G)$ and various X_m's.[2] An analogous procedure allows for identifying the Y_n that resembles most $Y(G)$. (The monomers identified in Fig. 8.3 are X_1 and Y_2). This procedure allowed me to consider the $X \cdots Y$ complex at the geometry G as a deformed X_m interacting with a deformed Y_n and to label this system as $X_m Y_n(G)$ $\left(X_1 Y_2(G_{X_1 Y_2}) \right.$ in Fig. 8.3$\left. \right)$.

The total energy of $X_m Y_n(G)$ can be represented as:

$$E_{X_m Y_n}(G) = E_{X_0}(G_{X_0}) + E_{Y_0}(G_{Y_0}) + E^{stab}_{X_m Y_n}(G), \tag{8.1}$$

where $E_{X_0}(G_{X_0})$ and $E_{Y_0}(G_{Y_0})$ are the energies of the most stable forms of X and Y at their respective optimal geometries and $E^{stab}_{X_m Y_n}(G)$ is the stabilisation energy calculated as:

$$E^{stab}_{X_m Y_n}(G) = E^{1b}_{X_m}(G) + E^{1b}_{Y_n}(G) + E^{2b}_{X_m Y_n}(G). \tag{8.2}$$

$E^{1b}_{Z_n}(G)$ $(Z = X, Y)$ is the one-body (1b) terms related to the monomer Z:

$$E^{1b}_{Z_n}(G) = E_{Z_n}(G) - E_{Z_0}(G_{Z_0}), \tag{8.3}$$

where $E_{Z_n}(G)$ is the energy of Z at the geometry G. $E^{1b}_{Z_n}(G)$ is the energy penalty for distorting the monomer Z from the global minimum G_{Z_0} to the geometry G. The one-body term $E^{1b}_{Z_n}(G)$ can be further split into a term that describes a conformational

---

[2]One can superimpose two molecular structures to minimise the RMSD value. The remaining value of RMSD is a measure of dissimilarity between the two structures
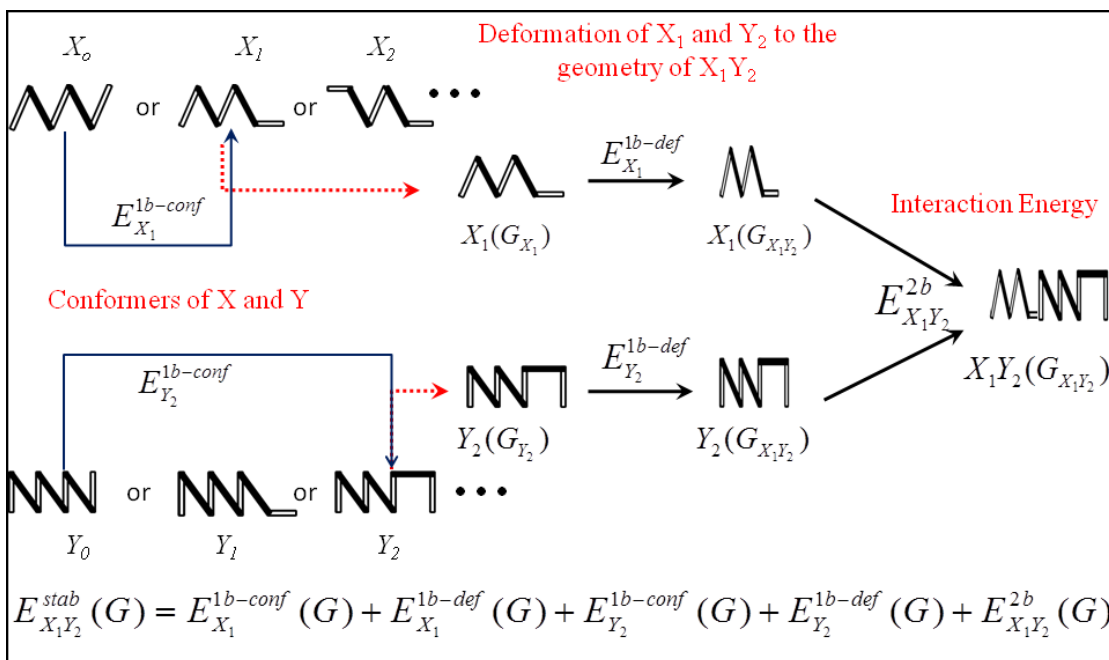
FIGURE 8.3: Illustration of the components of the stabilisation energy in the X_1Y_2 complex. Left: one-body conformational components; Center: one-body deformation components; Right: two-body interaction term.

or tautomeric change from 0 to n, and the deformation of the Z_n structure from G_{Z_n} to G (from now on I will focus discussion on the conformers of X and Y):

$$E_{Z_n}^{1b}(G) = E_{Z_{0 \to n}}^{1b-conf} + E_{Z_n}^{1b-def}(G), \tag{8.4}$$

with the conformation $E_{Z_{0 \to n}}^{1b-conf}$ and deformation $E_{Z_n}^{1b-def}(G)$ penalties defined, respectively, as:

$$E_{Z_{0 \to n}}^{1b-conf} = E_{Z_n}(G_{Z_n}) - E_{Z_0}(G_{Z_0}), \tag{8.5}$$

$$E_{Z_n}^{1b-def}(G) = E_{Z_n}(G) - E_{Z_n}(G_{Z_n}), \tag{8.6}$$

and illustrated in the left and middle parts of Fig. 8.3. $E_{X_mY_n}^{2b}(G)$ is the two-body ($2b$) interaction energy calculated as:

$$E_{X_mY_n}^{2b}(G) = E_{X_mY_n}(G) - E_{X_m}(G) - E_{Y_n}(G), \tag{8.7}$$

and illustrated in the right part of Fig. 8.3. The total energy of the dimer is referenced with respect to the energies of the most stable structures of X and Y, i.e., X_0 and Y_0, see Eq. 8.1. Thus $E_{X_mY_n}^{stab}(G)$ is a measure of stability of the complex at the geometry G.

The next, practical step is to determine the values of $E^{stab}_{X_m Y_n}(G)$ as accurately as possible. When solving the electronic Schrödinger equation one faces the challenges of electron correlation and incompleteness of one-electron basis sets. The dimer energies are contaminated with the so called basis set superposition error, which has to be dealt with when determining the equilibrium geometries and frequencies and the values of $E^{2b}_{X_m Y_n}(G)$.[233–235]

The approach I adopt here to obtain accurate electronic energies was developed in the group of Hobza.[236] The geometries of the isolated conformers **A** and **B** of the OA monomer and the counterpoise-corrected geometries[235] of the **AA**, **BB**, and **AB** dimers were optimised at the MP2 level of theory[132] using the aug-cc-pVDZ[121] basis set. In my first approach, the single point energies of the monomers and the counterpoise-corrected dimers were extrapolated to the complete basis set (CBS) limit at the SCF and MP2 levels of theory using the aug-cc-pVNZ[121] basis sets ($N = D, T$ and Q and stands for 2, 3 and 4, respectively). The stabilisation energies and their components were obtained from the extrapolated energies of the monomers and dimers. In my second approach, the stabilisation energies were calculated at the SCF and MP2 levels for the same sequence of basis sets. Next, these stabilisation energies were extrapolated to the basis set limit. Formally, the stabilisation energies obtained in these two approaches should be the same and the numerical discrepancy might be viewed as a measure of uncertainty in the SCF and MP2 stabilisation energies.

At the SCF level I followed an exponential functional form proposed by Halkier et al.[237] and adapted for a three-point extrapolation by Bachorz et al.:[125]

$$E^{\infty}_{HF} = \frac{E_{N-2}E_N - E^2_{N-1}}{E_{N-2} - 2E_{N-1} + E_N}. \tag{8.8}$$

The MP2 correlation energy

$$E_{corr-MP2} = E_{MP2} - E_{HF} \tag{8.9}$$

was extrapolated to the CBS limit using a N^{-3} functional form proposed by Helgaker et al.[123] and adapted for a two-point extrapolation by Bachorz et al.:[125]

$$E^{\infty}_{corr} = \frac{E_{N-1}(N-1)^3 - E_N N^3}{(N-1)^3 - N^3}. \tag{8.10}$$

Higher-order electron correlation effects were estimated by performing single-point calculations for the monomers and the counterpoise-corrected dimers at the CCSD(T) level of theory[238] using the aug-cc-pVDZ basis set. These corrections were split into the

CCSD contribution from single and double excitations

$$\Delta E_{CCSD} = E_{CCSD} - E_{MP2} \tag{8.11}$$

and the CCSD(T) contribution from triple excitations

$$\Delta E_T = E_{CCSD(T)} - E_{CCSD}. \tag{8.12}$$

These two terms were added to the sum of the CBS SCF and MP2-correlation terms to obtain my final electronic energies. Harmonic frequencies and zero-point vibrational corrections were obtained at the MP2 level using the aug-cc-pVDZ basis set; for the dimers the counterpoise-corrected surface was used.[235]

The counterpoise-corrected dimer energy, $E_{XY}^{CP}(G)$, is defined as:[233–235]

$$E_{XY}^{CP} = E_{XY}(G\|X + Y) - BSSE_X(G) - BSSE_Y(G). \tag{8.13}$$

I have dropped the m and n subscripts for the monomers, $\|X + Y)$ denotes the dimer centered basis set,[239] and $BSSE_X(G)$ and $BSSE_Y(G)$ denote basis set superposition errors for X and Y, respectively, at the geometry G:

$$BSSE_X(G) = E_X(G\|X + Y) - E_X(G\|X), \tag{8.14}$$

$$\tag{8.15}$$

and analogously for $BSSE_Y(G)$. Here $\|X)$ denotes the monomer basis set centered on X.

The counterpoise-corrected two-body interaction energy is calculated as:

$$E_{XY}^{2b-CP} = E_{XY}^{CP}(G) - E_X(G\|X) - E_Y(G\|Y), \tag{8.16}$$

which is equivalent to the interaction energy obtained from the Jansen-Ros-Boys-Bernardi counterpoise procedure: [233, 234]

$$E_{XY}^{2b-CP} = E_{XY}(G\|X + Y) - E_X(G\|X + Y) - E_Y(G\|X + Y). \tag{8.17}$$

8.4 Results

My results based on the CBS extrapolated E_{HF}^{∞} and $E_{corr-MP2}^{\infty}$ energies of the monomers and dimers with the CCSD(T) correction are summarised in Table 8.1. In terms of electronic energies, the **AA** dimer is more stable than the **AB** and **BB** dimers by 25

	AA		AB		BB	
$E_{Z_{0 \to n}}^{1b-conf}$	0.0	0.0	0.0	123.7	123.7	123.7
$E_{Z_n}^{1b-def}(G)$	36.7	36.7	65.4	19.4	40.2	40.2
$E_{Z_n}^{1b}(G)$	36.7	36.7	65.4	143.1	163.9	163.9
$E_{X_m Y_n}^{2b}$	-491.6		-601.8		-691.9	
$E_{X_m Y_n}^{stab}$	-418.2		-393.3		-364.1	
ΔE_0^{vib}	31.4		35.1		38.2	
$E_{X_m Y_n}^{stab} + \Delta E_0^{vib}$	-386.8		-358.2		-325.9	

TABLE 8.1: The total stabilisation energy and its components for the **AA**, **AB**, and **BB** dimers. The corrections from zero-point vibrations are included in the bottom rows. The energies are in meV

and 54 meV, respectively. The **BB** and **AB** dimers are characterised by more attractive 2-body terms (more negative than that for **AA** by 200 and 110 meV, respectively), but these favourable 2-body interactions are unable to compensate for the less favourable 1-body terms. The **AA** dimer remains more stable than **AB** and **BB** by 29 and 61 meV when zero-point vibrational corrections are included. The above results based on extrapolation of the monomer and dimer energies (Method **1**) differ by less than 1.7 meV from the stabilisation energies extrapolated the SCF and MP2 limits (Method **2**), see Table D.12 in appendix D.

The structure of $E_{A \to B}^{1b-conf}$ is summarised in Table 8.2. It is dominated by the HF component and further enhanced by the correlation MP2 component, with the ΔE_{CCSD} and ΔE_T components being one order of magnitude smaller and of opposite signs. The complete value of $E_{A \to B}^{1b-conf}$, 123.7 meV, is significant. It is larger than the differences in stabilisation energy between the **AA, AB**, and **BB** complexes. $E_{A \to B}^{1b-conf}$ might be interpreted as the energy needed to break the intramolecular hydrogen bond in **A**.

	$E_{\mathbf{A} \to \mathbf{B}}^{1b-conf}$
E_{HF}^{∞}	95.6
$E_{corr-MP2}^{\infty}$	31.7
ΔE_{CCSD}	-7.8
ΔE_T	4.2
$\boldsymbol{E_{Total}}$	**123.7**
ΔE_0^{vib}	-6.9
$E_{Total} + \Delta E_0^{vib}$	116.8

TABLE 8.2: The structure of the $E_{\mathbf{A} \to \mathbf{B}}^{1b-conf}$ component. All energies are in meV.

The structure of $E_{A \ or \ B}^{1b-def}$ terms for all three complexes is illustrated in Table 8.3. This

	A@AA	A@AB	B@AB	B@BB
	$E_{A\ or\ B}^{1b-def}$			
E_{HF}^{∞}	19.1	50.9	20.5	50.1
$E_{corr-MP2}^{\infty}$	17.2	13.7	-3.0	-12.2
ΔE_{CCSD}	-2.0	-1.0	2.3	4.3
ΔE_{T}	2.4	1.8	-0.4	-2.0
E_{Total}	36.7	65.4	19.4	40.2

TABLE 8.3: The structure of the $E_{A\ or\ B}^{1b-def}$ components. All energies are in meV.

	$E_{\mathbf{AA}}^{2b}$	$E_{\mathbf{AB}}^{2b}$	$E_{\mathbf{BB}}^{2b}$
E_{HF}^{∞}	-358.3	-457.0	-525.8
$E_{corr-MP2}^{\infty}$	-132.4	-139.4	-156.5
ΔE_{CCSD}	25.6	24.0	24.4
ΔE_{T}	-26.5	-29.5	-33.9
E_{Total}	-491.6	-601.9	-691.8

TABLE 8.4: The structure of two-body interaction energy in the **AA**, **AB**, and **BB** dimers. All energies are in meV.

repulsive component of the stabilisation energy is dominated by the HF term. The MP2 component is dominant among the correlation contributions. It is repulsive for the **A** monomer in the **AA** and **AB** complexes and attractive for the **B** monomer in the **AB** and **BB** complexes. It is comparable in magnitude to the HF term for the **AA** complex, but smaller in absolute value in the remaining complexes. The correlation contributions ΔE_{CCSD} and ΔE_{T} are smaller than the MP2 contribution. For all three complexes notice a sign alteration along the corr-MP2, CCSD, and T series. The total values of $E_{A\ or\ B}^{1b-def}$ terms are comparable in magnitude to the differences in stabilisation energy between the **AA**, **AB**, and **BB** complexes.

The structure of two-body interaction terms is illustrated in Table 8.4. The total values are expected to be attractive, because I am dealing with minimum energy structures of the dimer. The HF term is dominant for each dimer followed by a substantial corr-MP2 term (ca. 30% of the HF term). The ΔE_{CCSD} and ΔE_{T} components are one order of magnitude smaller than the HF and corr-MP2 components and they cancel out to a large extent. The two-body interaction term is the most attractive in the **BB** dimer, which is reflected in the shortest OH...O distance (1.750 Å) and the most significant elongation of the OH proton donor bond (0.020 Å) (see Figs. 8.1 and 8.2). The same geometric parameters reflect that this interaction is the weakest in **AA** (see Figs. 8.1 and 8.2). Still, we remember that the **AA** dimer is the global minimum for the (OA)$_2$ complex (Table 8.1).

System	Mode	Frequency (cm^{-1})
A	Symmetric IAH	3635
	Asymmetric IAH	3639
B	IAH	3662
	NoH	3727
C	Symmetric NoH	3736
	Asymmetric NoH	3736
AA	Symmetric IAH & IEH	3566
	Asymmetric IAH & IEH	3572
	Asymmetric IAH	3588
	Symmetric IAH	3591
AB	IAH & IEH(**A**)	3449
	IEH(**B**)	3520
	IAH(**A**)	3573
	IAH(**B**)	3673
BB	Symmetric IEH	3275
	Asymmetric IEH	3356
	Symmetric IAH	3680
	Asymmetric IAH	3681

TABLE 8.5: Harmonic MP2/aug-cc-pVDZ frequencies of the O–H stretching modes in OA and (OA)$_2$. IAH – intramolecular hydrogen bond, IEH – intermolecular hydrogen bond, and NoH – no hydrogen bond

In the bottom lines of Tables 8.1 and 8.2 I illustrate the effect of correction for zero-point vibrations to the stabilisation energy and $E_{A \to B}^{1b-conf}$, respectively. When dealing with intra- or intermolecular hydrogen bonds it is natural to pay attention to the proton donor stretching modes, in my case OH's in the OH...O configuration, which are expected to be redshifted in comparison with the OH stretches not involved in hydrogen bonds.

The frequencies of the OH stretching modes in the **A**, **B**, and **C** monomers and the **AA**, **AB** and **BB** dimers are presented in Table 8.5. Looking at the monomers, the highest and lowest frequencies are observed for the **C** and **A** monomers, respectively, reflecting the engagement of both OH's in intramolecular hydrogen bonds in **A**, but none in **C** (see Fig. 8.1). Not surprisingly, the **B** monomer has one OH stretch similar to these in **A** and another similar to those in **B**. Based on the frequencies of the OH stretches alone one would expect a higher zero-point energy in **B** than in **A**, but the result presented in the bottom rows of Table 8.2 indicates otherwise. I verified that modes other than OH stretches typically blueshift when progressing from **C** through **B** to **A** (see Table 8.6). The same intramolecular hydrogen bonds that are responsible for the softening of OH stretches rigidify molecular structures with respect to the bending and puckering motions. The total zero-point vibrational correction decreases from **A** through **B** to **C** (bottom rows of Table 8.2, Table 8.6).
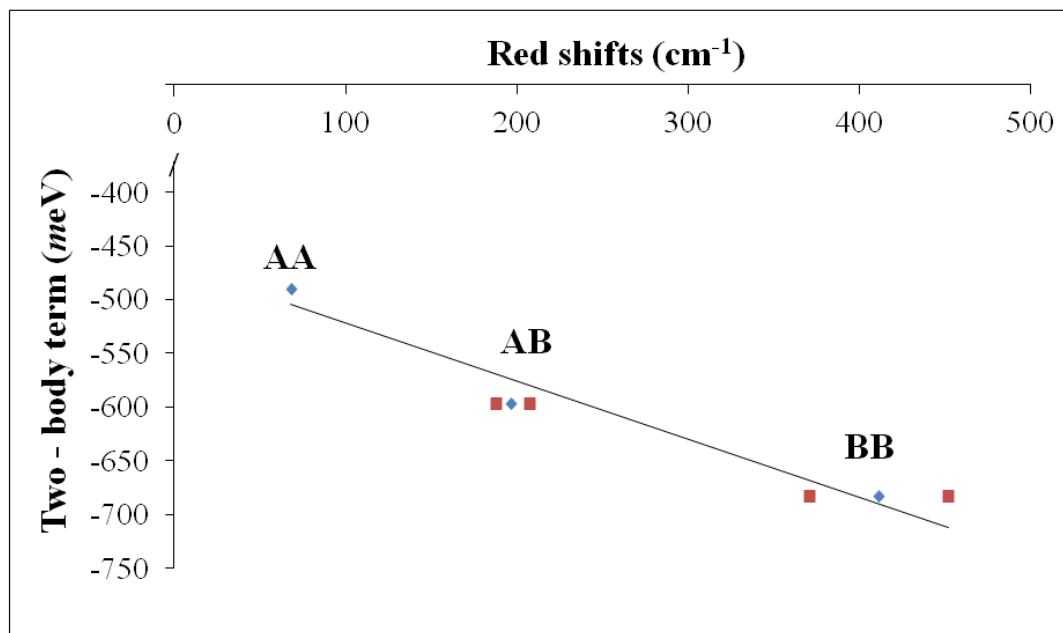
FIGURE 8.4: Correlation between the two-body interaction energy terms and the red shifts of the O–H's involved in intermolecular hydrogen bonds. Red squares for the red shifts of individual modes and blue diamonds for average values for each dimer. The results obtained at the MP2/aug-cc-pVDZ level.

Next I considered the **AA**, **AB** and **BB** dimers, each stabilised by two intermolecular hydrogen bonds (Figure 8.2). The frequencies of the OH's involved in intermolecular hydrogen bonds are redshifted with respect to the monomers, with the largest effects in the **BB** dimer (370 and 450 cm^{-1}) and the smallest in the **AA** dimer (65 and 71 cm^{-1}). The average values of these shifts correlate well with the values of $E^{2b}_{X_mY_n}$, see Fig. 8.4. The OH stretches not involved in the intermolecular hydrogen bonds are redshifted for **A** in **AA** and **AB** (by 40 – 60 cm^{-1}), and blueshifted for **B** in **AB** and **BB** (by 10 – 20 cm^{-1}). The opposite pattern for the **A** and **B** monomers might be related to the opposite pattern of correlation contributions to the one-body deformation terms reported in Table 8.3. Based on the frequencies of OH stretches alone one would expect a higher zero-point energy in **AA** than in **AB** or **BB** (Tables 8.5 and 8.6), but the results presented in Table 1 indicate otherwise. I verified that intramonomer modes other than OH stretches typically blueshift with the enhancement of $E^{2b}_{X_mY_n}$, i.e., from **AA** through **AB** to **BB** (Table 8.6). In addition, soft intermolecular modes become stiffer as $E^{2b}_{X_mY_n}$ increases (Table 8.6). The same intermolecular hydrogen bonds that are responsible for the overall softening of the OH stretches rigidify dimer structures with respect to other intramolecular and intermolecular distortions. The total zero-point vibrational correction increases from **AA** through **AB** to **BB** (Tables 8.1 and 8.6).

System	Contribution	Value (meV)
A	O–H stretches	451.0
	Other modes	885.7
	Total E_0^{vib}	1336.7
B	O–H stretches	458.1
	Other modes	871.8
	Total E_0^{vib}	1329.9
C	O–H stretches	463.2
	Other modes	860.3
	Total E_0^{vib}	1323.5
AA	O–H stretches	887.6
	Other intramolecular modes	1791.7
	Intermolecular modes	25.6
	Total E_0^{vib}	2704.9
AB	O–H stretches	881.4
	Other intramolecular modes	1799.0
	Intermolecular modes	28.3
	Total E_0^{vib}	2708.7
BB	O–H stretches	867.5
	Other intramolecular modes	1812.8
	Intermolecular modes	31.4
	Total E_0^{vib}	2711.7

TABLE 8.6: Contributions from O–H stretches, other intramolecular modes, and intermolecular modes to the zero-point vibrational energy in OA and $(OA)_2$. All results obtained at the MP2/aug-cc-pVDZ level.

8.5 Summary

I discussed the case of dimers formed by molecules that can exist in different conformational states. I have developed a protocol that allows dissecting the total stabilisation energy into one-body conformational and deformational components and the two-body interaction energy term. Interplay between these components determines the overall stability of the dimer. The stabilisation SCF and MP2 energies were extrapolated to the basis set limit using a series of aug-cc-pVNZ basis sets ($N = D - Q$). In the first approach, the stabilisation energies and their components were obtained from the extrapolated monomer and dimer energies. In the second approach, the stabilisation energies were calculated at the SCF and MP2 levels for the same sequence of basis sets. Next, these stabilisation energies were extrapolated to the basis set limit. Higher-order electron correlation effects were calculated at the CCSD(T) level of theory using the aug-cc-pVDZ basis set. All dimer energies were counterpoise-corrected.

The protocol has been tested on the dimers of oxalic acid. The monomer can exist in various conformational forms and a variety of intermolecular hydrogen bonds is also

possible.[228] I studied two dimers, **AB** and **BB**, discussed earlier by Blair and Thakkar and a *new* dimer, **AA**, that I suspected to be the global minimum (see Fig. 8.2). I have demonstrated that the **AA** dimer is more stable than the **AB** and **BB** dimers by 25 and 54 meV, respectively. The stabilisation energies obtained through the two extrapolation schemes differed by less than 1.7 meV. The **AA** dimer is stabilised by two inter- and four intramolecular hydrogen bonds, while the most stable dimer identified by Blair and Thakkar, **AB**, is supported by two inter- and three intramolecular hydrogen bonds. The **BB** and **AB** dimers are characterised by more attractive two-body interaction energies than **AA**, but these favourable two-body interactions are unable to compensate for the less favourable one-body interactions. I concluded that that overall global stability of **AA** results from a balancing act between the moderately attractive two-body interaction energy and small repulsive one-body terms.

I have demonstrated that understanding of zero-point vibrational corrections to the stability of various conformers of oxalic acid and their dimers cannot be limited to the contributions from the OH stretching modes that can become strongly redshifted if engaged in hydrogen bonds. The same hydrogen bonds rigidify molecular structures with respect to bending, puckering, and intermolecular distortions. When contributions from all vibrational modes are included, the oxalic acid monomer or dimer with the most stabilising sets of hydrogen bonds have the largest zero-point vibrational energy, contrary to the naive anticipation based on red shifts of OH stretching modes involved in hydrogen bonds.

Chapter 9

Intrinsic Electrophilicity of Oxalic Acid Monomer Is Enhanced in the Dimer and Trimer by Intermolecular Proton Transfer

9.1 Abstract

My computational results together with experimental results from my experimental collaborators demonstrate an unusual electrophilicity of oxalic acid (OA), the simplest dicarboxylic acid. The monomer is characterized by an electron vertical detachment energy (VDE) and adiabatic electron affinity of 1.10 and 0.72 eV, respectively. The electrophilicity results primarily from the bonding carbon-carbon interaction in the SOMO orbital of the anion, but it is further enhanced by intramolecular hydrogen bonds. The well-resolved structure in the photoelectron spectrum is reproduced theoretically, based on Franck-Condon factors for the vibronic anion→neutral transitions. The excess electron binding energies in the dimer and trimer of OA become very significant due to intermolecular proton transfer, with the corresponding VDE values of approximately 3.3 and 4.6 eV. I have postulated a mechanism of excess electron mobility along molecular linear chains supported by cyclic hydrogen bonds.

9.2 Introduction

Stable, closed-shell organic molecules with heteroatoms, such as monocarboxylic acids (formic, acetic), nucleic acid bases and amino acids, usually do not support bound valence anionic states in the neighbourhood of the optimal geometry of the neutral species.[171] These molecules still support metastable (resonant) anionic states, with finite lifetimes and energies higher than the energy of the neutral,[198, 240, 241] but they are not able to bind an excess electron in a valence orbital. The electrophilicity of these molecules is typically enhanced upon specific geometric distortions, including tautomerizations.[27, 224] In consequence, valence anionic states are frequently characterised by positive values of electron vertical detachment energy (VDE), while adiabatic electron affinities (AEA) of the corresponding neutrals might remain negative (CO_2[224]), approach zero (canonical uracil[125]), or settle at positive values (unconventional tautomers of G[144, 242] and A,[243] nucleotides[200, 244]). Here I report a significant electrophilicity of the oxalic acid (OA) monomer. It is the simplest dicarboxylic acid, see Figure 9.1, which might be viewed as a product of condensation of two formic acid (FA) molecules (with the release of H_2). Let me reemphasize that the FA monomer does not support a bound valence anionic state.
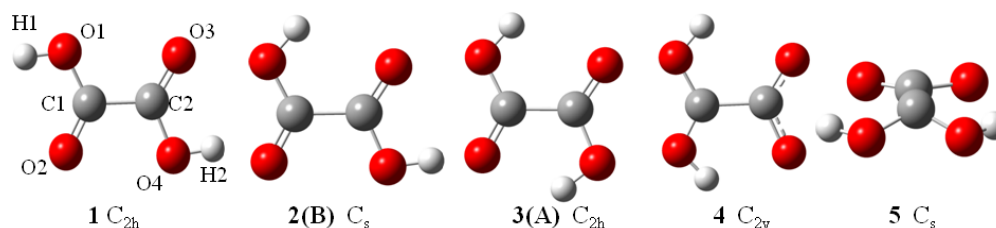
FIGURE 9.1: *Conformers and tautomers of the oxalic acid monomer. The naming scheme for atoms shown for **1**. The structures **3** and **2** were labelled in Ref.[193] as A and B, and I will use their notation when discussing the dimer of OA.*

The neutral OA molecule can exist in three conformational forms (structures **1-3** in Figure 9.1), it also has a local minimum for a "rare tautomer" (structure **4**). These minimum energy structures differ in the extent of intramolecular hydrogen bonding; this phenomenon attracted attention of many experimental[245–250] and computational[228–231, 251–253] groups. Indeed, the gas phase structure of the neutral OA monomer has been studied by electron diffraction,[245] infrared and Raman spectroscopy,[245, 246] matrix-isolation,[247–249] UV absorption,[250] microwave spectroscopy,[254] and theoretically at various levels of theory.[228–231, 251–253] Here I report a photoelectron spectrum (PES) of the OA monomer anion, which unravels a significant electrophilicity of the neutral: the main features extend from 0.5 to 2.5 eV (Figure 9.2). My computational results provide an interpretation of this well-resolved photoelectron spectrum.
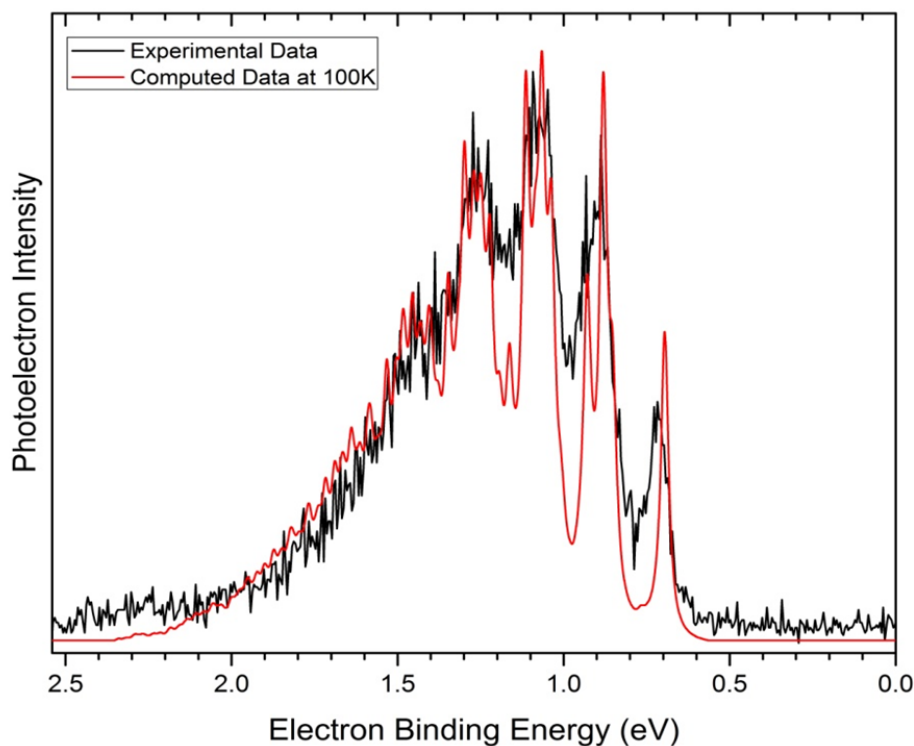
FIGURE 9.2: *Photoelectron spectrum of recorded with 2.54 eV photons (black). Computed spectrum based on the CCSD(T)/ATZ electronic energies and CCSD/ATZ geometries and hessians (red).*

It has been demonstrated in the past that electrophilicity of hydrogen bonded dimers can be enhanced by intermolecular proton transfer.[255, 256] Again, formic acid serves as a handy example.[196, 199] Its dimer undergoes intermolecular proton transfer and supports a valence anion with a VDE of 2.35 eV, but an AEA close to zero ($|AEA| < 20$ meV) and the structure that resembles the formate anion engaged in a symmetrical double hydrogen-bonded bridge with the dihydroxymethyl radical. I suspected that a similar mode of stabilisation of valence anion will take place in the dimer of OA, but I expected an even greater value of VDE and a positive value of AEA. The PES spectra of $(OA)_2^-$ and $(OA)_3^-$ reported here demonstrate a significant increase of the VDE values in comparison with OA^-. The computational results for $(OA)_2^-$ confirm intermolecular proton transfer and the positive values of AEA.

9.3 Computational Details

The minimum energy structures and harmonic frequencies for **1-5** of OA and OA^- were determined at the CCSD level of theory[188] and single-point energies were determined at the CCSD(T) level.[188] These calculations were performed with the ADZ[121] basis set. For the most stable neutral and anionic structure **3**, the calculations were repeated with

the aug-cc-pVTZ[121] (ATZ) basis set. For the neutral and anionic $(OA)_2$, the minimum energy structures and harmonic frequencies were determined at the MP2 level of theory and single-point energies were determined at the CCSD(T) level. All calculations for the dimer were performed with the ADZ basis set. The electronic structure calculations have been carried out with the Gaussian 09[75] and Molpro[257, 258] codes. Molecular structures and orbitals were plotted with the GMolden program.[161]

The Franck-Condon (FC) factors, i.e., the squares of overlap integrals between vibrational wave functions, were calculated for the anionic and neutral OA in harmonic approximation with molecular structures and Hessians determined at the CCSD/ATZ level. Both geometrical equilibrium parameters as well as curvatures are affected by excess electron attachment and the resulting FC factors may contribute to vibrational structure in the photoelectron spectrum. The polyatomic FC factors were calculated using the recursion relations of Doktorov.[259, 260] The simulations were performed for different temperatures of the anionic beam ($25\ K < T < 300\ K$). The energy of 0-0 transition was determined from the CCSD(T)/ATZ electronic energies and the CCSD/ATZ zero-point harmonic frequencies. The intensity for the 0-0 transition was normalized to 1 and all other intensities were scaled accordingly. The calculated FC factors were convoluted with Lorentzian line shapes (full width at half maximum = 218 cm^{-1}) and the simulated spectrum is presented in Figure 9.2.

9.4   Electron Attachment to the Monomer of Oxalic Acid

The landscape of potential energy surface for neutral and anionic OA is summarized in Figure 9.5. The conformer **3** is stabilized by two intramolecular hydrogen bonds, thus the stability decreases along the series **3→2→1**, see also Ref.[251] The structures of the minima and transitions states TS1-TS3 for the neutral and anion are detailed in Figures 9.3-9.4 and the geometrical coordinates in Tables E.1-E.2 in appendix E. The CCSD(T) barriers determined at the CCSD transition states separating neutral conformers are large enough (0.602 eV for **3→2** and 0.519 eV for **2→1**) to support a few vibrational states for **2** and **1**. The "rare tautomer" **4**, which might be viewed as a product of intramolecular proton transfer, is a local minimum on the potential energy surface of the neutral, but it is less table than **3** by 1.010 eV and the barrier for **4→3** is only 0.020 eV.

The OA monomer supports a bound valence anionic state: all anionic minima (**1-3**, and **5**) are more stable than the most stable neutral **3** by a few tenths of an eV (Figure 9.5). **3** is the most stable anionic conformer (Figure 9.5). The barriers separating conformers **1-3** are smaller for the valence anion than for the neutral. The tautomer **4** is strongly
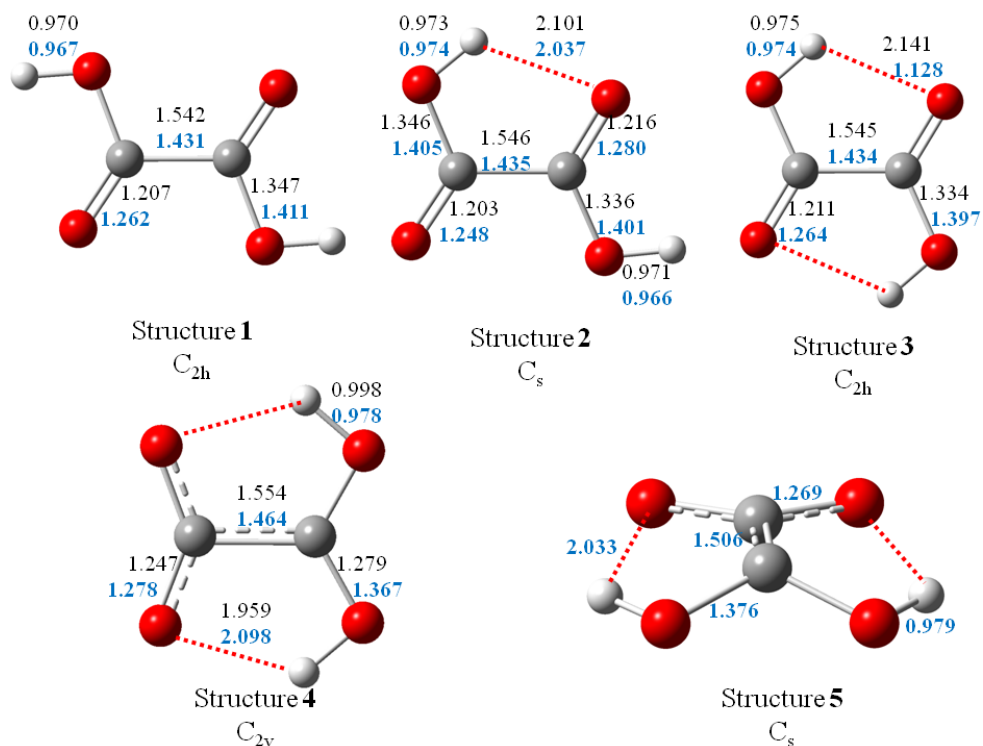
FIGURE 9.3: *The CCSD/ADZ structures for the neutral and anion of oxalic acid. The geometrical parameters (in) for the neutral and anion in black and blue, respectively. All structures are minima but the anionic 4, which is a transition state.*
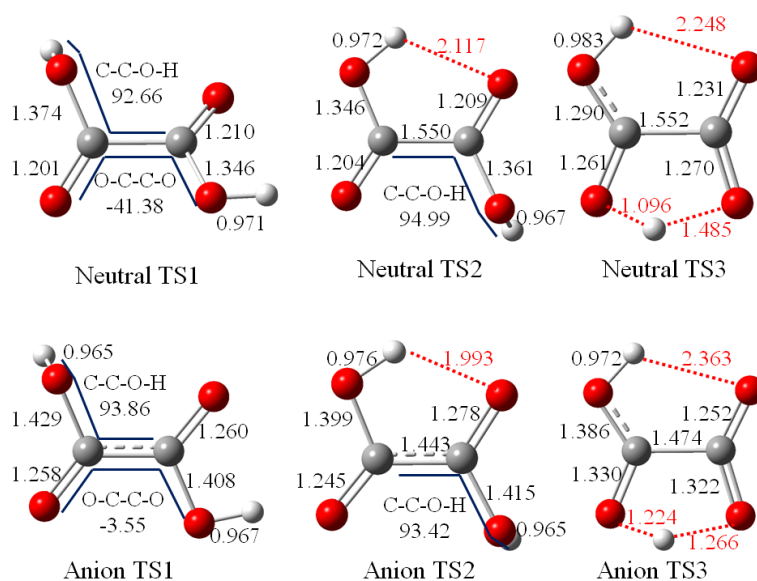

FIGURE 9.4: *The CCSD/ADZ structures of the neutral and anionic transition states TS1-TS3 of oxalic acid. The geometrical parameters are in Å, °.*
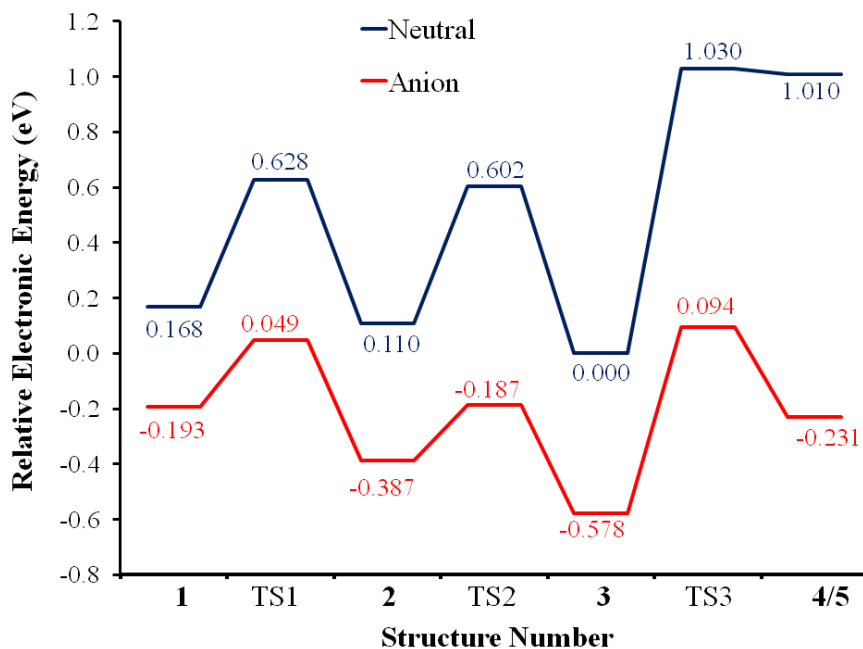
FIGURE 9.5: *Energetics of stationary points (minima and transition states) on the potential energy surface of the neutral and anionic monomer of oxalic acid, where the zero of energy is set to the energy of the neutral **3**.*

stabilized by the excess electron attachment. The molecular framework for the anion lowers symmetry from C_{2v} (structure **4**, with an imaginary frequency for a **b1** mode) to C_s (structure 5) and the CCSD(T) barrier for **5→3** becomes 0.325 eV, thus one order of magnitude larger than the barrier **4→3** for the neutral. I concluded that the anion **5** might be sufficiently long-lived to be observed in experimental conditions.

Which factors contribute to the stability of valence anions of the OA monomer? The singly occupied molecular orbital (SOMO) is of π symmetry (Figure 9.6). The SOMO is characterised by a bonding C-C interaction and antibonding C-O interactions, which is clearly illustrated for the C_{2h} structures **3** and **1**. A similar pattern holds for **4** and **5**, though the carbon atoms are not equivalent due to lower symmetries. I believe that the unique electrophilicity of OA results primarily from the proximity of the carboxylic groups, which allows for the bonding C-C interaction in the SOMO.

There are also secondary factors that contribute to the stability of valence anions of OA. The CCSD(T) values of VDE (Fig. 9.7a) increase from **1** to **3** demonstrating that intramolecular hydrogen bonding stabilises the anion. An even greater increase of VDE is brought by intramolecular proton transfer as the value of VDE increases by 0.605 eV from **3** to **4**. Finally, a buckling of the molecular framework further increases the value of VDE (**4→5**) by 0.591 eV.
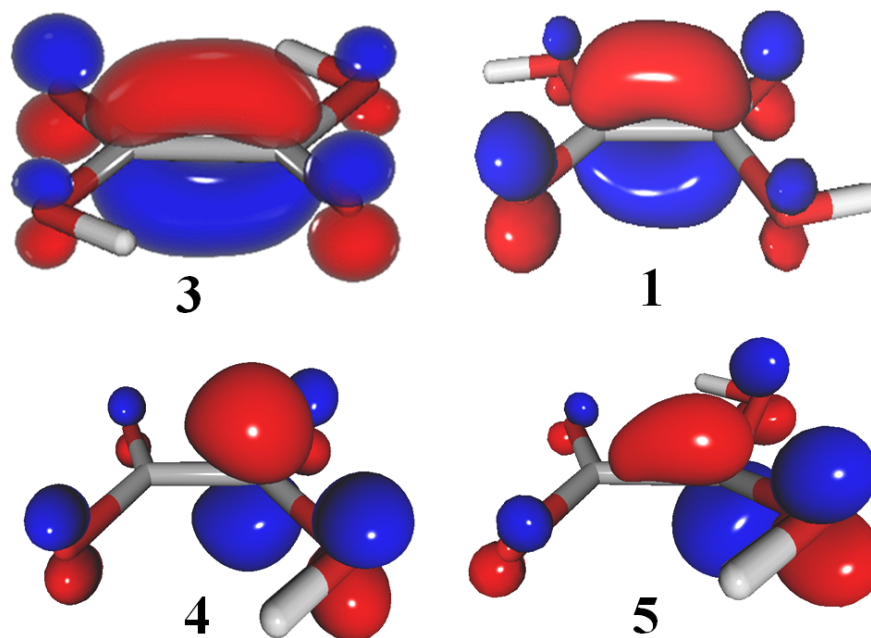
FIGURE 9.6: *The SOMO of anionic OA structures plotted with a contour value of 0.1 au.*

The buckling of molecular frameworks upon binding an excess electron on a π^* orbital is a common phenomenon in valence anions of nucleic acid bases.[27] In the case of the most stable valence anion of OA, **3**, the framework remains unbuckled and the C_{2h} symmetry is maintained. Notice, however, that the buckling mode of **bg** symmetry is softer by 201.7 cm^{-1} for the anion than for the neutral (Table 9.1). It requires the intramolecular proton transfer (**3**→**4**) to change the sign of curvature of the buckling mode. Upon intramolecular proton transfer the unpaired electron becomes localized on the $C(OH)_2$ fragment (bottom of Figure 9.6) and the C atom forms an apex of the buckled structure (Figure 9.1). The stationary point, **4**, is a transition state for the valence anion and the b_1 mode has an imaginary frequency of $\sim 300i$ cm^{-1}. This mode morphs into an a' symmetry mode of **5** with a frequency of ~ 800 cm^{-1}.

The electron binding energies (VDE, AEA) are summarised in Figure 9.7. The AEA values are corrected for the energies of zero-point vibrations and are reported with respect to the corresponding neutral ("local" AEA, Figure 9.7b), and the most stable neutral **3** ("global" AEA, Figure 9.7c). For each structure, I report significant differences between the VDE and "local" AEA values, which must be associated with geometric distortions, such as changes of bond lengths and angles, upon binding an excess electron. These will be critical for the discussion of the PES spectrum of OA$^-$. Notice that the "global" AEA values (Figure 9.7c) remain positive for all structures.

The experimental PES spectrum of OA$^-$ is presented in black in Figure 9.2. It spreads
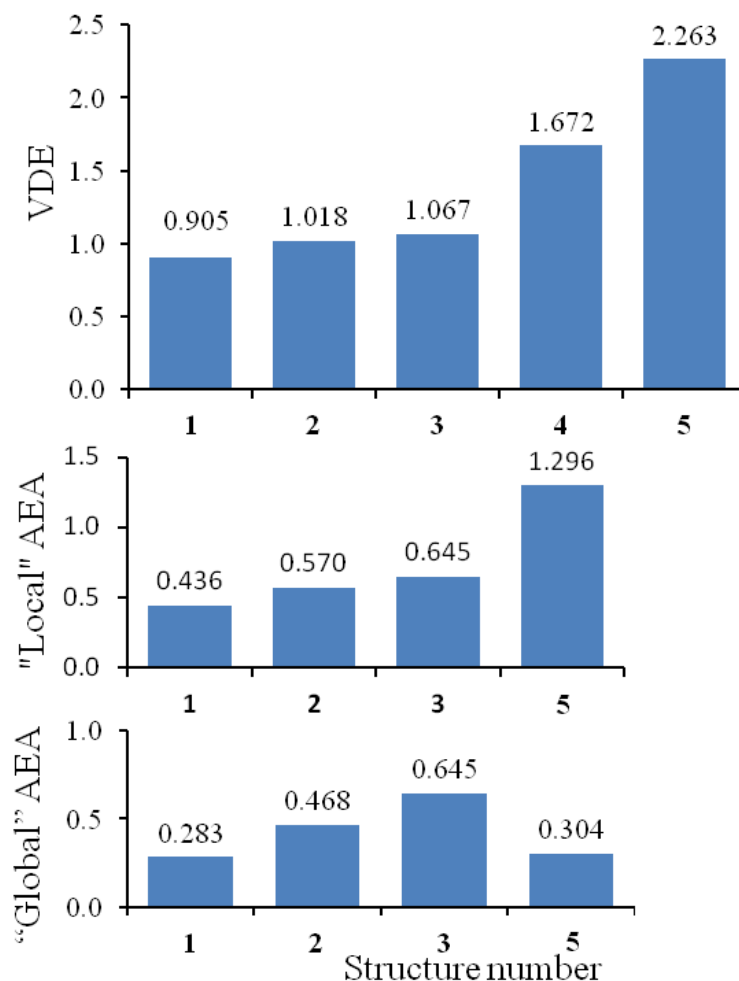
FIGURE 9.7: *Excess electron binding energies (eV) for the oxalic acid monomer. The AEA values are reported with respect to the corresponding neutral ("Local" AEA), and with respect to the most stable neutral 3 ("Global" AEA).*

from approximately 0.5 to 2.5 eV, with well defined peaks at 0.72, 0.92, 1.10, 1.29, and 1.40 eV. In view of the fact that the anionic minimum **3** is more stable than other minima by more than 0.2 eV (Figure 9.5), I focused my attention on FC factors for the structure **3** anion→neutral vibronic transitions, and I assumed that contributions from other anionic structures, to the experimental PES spectrum, are less probable. Two observations support this assumption: (i) the position of the highest peak in the experimental spectrum (1.10 eV) coincides with the calculated value of VDE for **3** of 1.13 eV, and (ii) the position of the first peak in the spectrum at 0.72 eV coincides with the calculated values of AEA for **3** of 0.70 eV.

The calculated FC factors and signal intensities are presented in Table E.1 of the SI and the resulting computed spectrum is presented in red in Figure 9.2. The best match with the experimental spectrum was found for $T = 100\ K$. A very good agreement between the computed and experimental spectra suggests that the experimental anionic beam is

| Mode	Symmetry	Nature	Frequency (cm^{-1})	
			Anion	Neutral
1	Au	C-C Rotation	122.7	190.0
2	Bu	C-C-O Bend	269.4	255.0
3	Au	Buckling	481.5	403.5
4	Ag	C-C Stretch	418.6	411.8
5	Ag	C-C-O Bend	572.7	603.5
6	Bu	C-C-O Bend	683.3	612.6
7	Bg	C-O(H) Rotation	686.5	615.9
8	Au	C-O(H) Rotation	691.2	633.3
9	Bg	Buckling	855.5	653.8
10	Ag	C-C Stretch	848.1	836.3
11	Bu	C-O(H) Stretch	1233.5	1082.5
12	Ag	C-O-H Bend	1262.1	1278.5
13	Bu	C-O-H Bend	1368.4	1322.6
14	Ag	C-O(H) Stretch	1490.3	1424.2
15	Bu	C=O Stretch	1894.8	1628.2
16	Ag	C=O Stretch	1884.5	1780.1
17	Bu	O-H Stretch	3737.4	3732.6
18	Ag	O-H Stretch	3733.6	3735.9

TABLE 9.1: Nature of vibrational modes and harmonic frequencies, calculated at the CCSD/ATZ level for structure 3. The "buckling" mode in bold, the modes contributing primarily to the vibronic structure reported in Figure 9.2 in blue.

indeed dominated by the most stable anionic structure **3**. Notice, however, a nonzero photoelectron intensity at approximately 2.2-2.3 eV, where the computed spectrum has no intensity. This weak feature might result from a small fraction of structure **5** in the anionic beam, with the calculated VDE of 2.26 eV (Fig. 9.7a).

Further analysis of harmonic frequencies (Table 9.1) and geometric parameters (Table 9.2) of the neutral and anion of **3** is needed to understand the origin of vibronic structure reported in Figure 9.2 and Tabulated figure E.1. The C-C bond contracts and the C-O bonds elongate upon excess electron attachment. These are significant distortions, exceeding 0.05 Å. In addition, the C-C=O and C-O-H angles expand and contract, respectively, by 4-5°. The geometric distortion from the anion to the neutral can be accomplished by displacements along the fully symmetric modes 10, 12, 14 and 16. All these are C-O or C-C stretching modes with an exception of 12, which is a C-O-H bending mode (Table 9.1). The geometric changes are consistent with the nature of SOMO in the anion, which is bonding in the C-C region and antibonding in the C-O regions (Figure 9.6). The C-O stretching modes 14 and 16 are strongly red-shifted upon an excess electron attachment by 66 and 104 cm^{-1}, respectively (Table 9.1). Finally, perusal of data from Tabfig E.1 confirms that the largest FC factors are associated with the 0-0 transition at 0.70 eV, and vibrational excitations involving the modes 12, 14 and

Parameter	Neutral	Anion
C1-C2	1.538	1.422
C1-O1	1.321	1.382
C1=O2	1.200	1.253
O1-H1	0.970	0.969
H1$\cdots$O4	2.118	2.099
C2-C1-O1	113.58	113.94
C2-C1=O2	121.20	125.41
C1-O1-H1	107.19	102.70

TABLE 9.2: Geometric parameters for the neutral and anionic structure **3** at the CCSD/ATZ level.
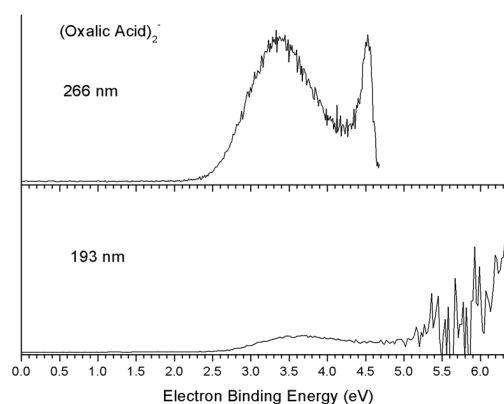
FIGURE 9.8: *Photoelectron spectrum of $(OA)_2^-$ recorded with 4.66 and 6.42 eV photons.*

16, which are responsible for the developments of PES peaks at 0.92, 1.10, 1.29, and 1.40 eV.

9.5 Excess Electron Binding Energies in the Dimer and Trimer of Oxalic Acid

The PES spectra of anionic dimer and trimer of OA are presented in Figures 9.8 and 9.9, respectively. The VDE value for the anionic dimer is 3.3 eV, an increase by 2.2 eV in comparison with OA$^-$. This value of excess electron binding energy is comparable with that of F$^-$ (3.40 eV) and Br$^-$ (3.36 eV), but smaller than for Cl$^-$ (3.61 eV).[171] The VDE value for $(OA)_3^-$ is further increased to 4.6 eV.

My computational results for the neutral and anionic dimer of OA are presented in Figure 9.10 and the complete molecular structures in Table E.3-E.4 of appendix E. The oxalic acid monomer can engage in cyclic hydrogen bonding in more than one way. The proton
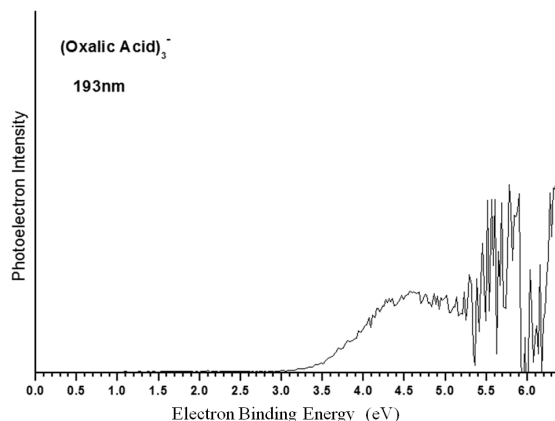
FIGURE 9.9: *Photoelectron spectrum of $(OA)_3^-$ recorded with 6.42 eV photons.*

donor and acceptor sites come from either the same or the neighbouring carboxylic group(s). In addition, each monomer can form various intramolecular hydrogen bonds (Figure 9.1). These factors lead to a large number of potentially stable $(OA)_2$ structures, which have recently been studied by Blair and Thakkar.[228] I will follow their notation and the monomer structures **3** and **2** become **A** and **B**, respectively. My three most stable neutral dimers are illustrated in the top part of Figure 9.10. In addition to the **AB** and **BB** structures,[228] I identified the AA structure, which proves to be the most stable. One should note, however, that the differences in stability between the three neutral dimers do not exceed 0.05 eV, therefore I have considered various anionic structures.

The three most stable anions I identified are illustrated in the bottom part of Figure 9.10. The anionic structures might be traced back to the three most stable neutral dimers, **AA**, **AB**, and **BB**. They all undergo intermolecular proton transfer upon the excess electron attachment reminiscent of a double hydrogen-bonded bridge in the anionic dimer of FA.[255] There is, however, an important difference between the anionic dimers of FA and OA. The former is characterized by a slightly negative AEA[255] but the latter is stable by more than 1.5 eV with respect to the **AA** neutral dimer. **BB-** and **AB-** are similarly stable ("global" AEA≈ 1.65 eV), while **AA-** is less stable by approximately 0.1 eV. The calculated VDE values for $(OA)_2^-$ are in the 3.06-3.32 eV range, thus they coincide with the maximum of the PES spectrum (Figure 9.8) and are approximately 1 eV greater than the VDE for the anion of FA dimer.[196, 199] The increase of cluster size from two to three OA monomers increases VDE by 1.3 eV. This increase is larger than for FA, where the VDE values for anionic trimer and dimer differ by approximately 0.8 eV.[180]
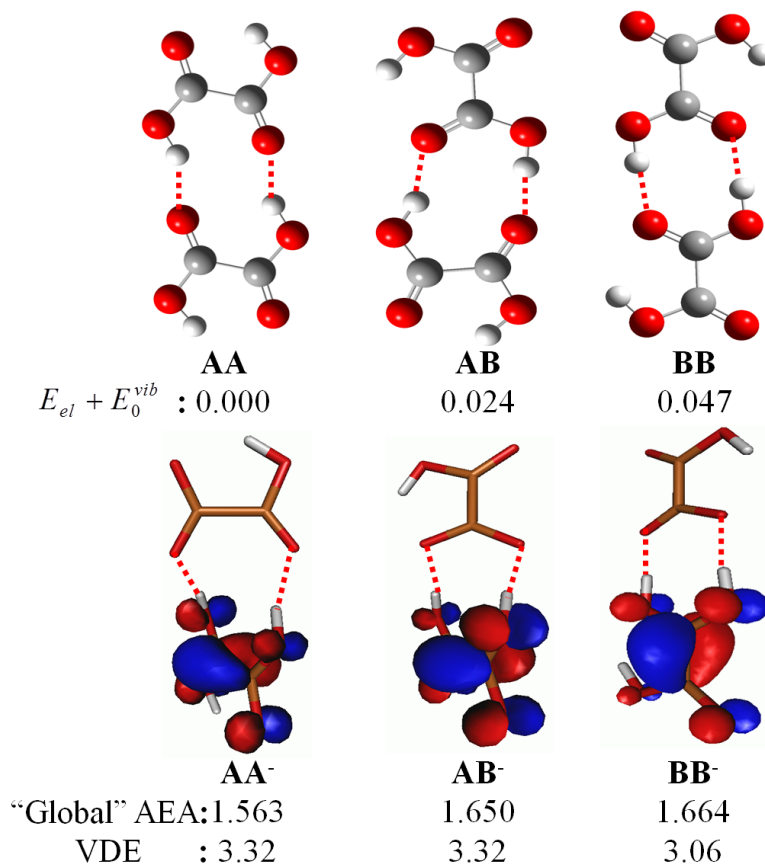
	AA	**AB**	**BB**
$E_{el} + E_0^{vib}$:	0.000	0.024	0.047

	AA⁻	**AB⁻**	**BB⁻**
"Global" AEA:	1.563	1.650	1.664
VDE :	3.32	3.32	3.06

FIGURE 9.10: *Relative stabilities (eV) of the neutral and anionic dimers of OA, along with "Global" AEA and VDE values (eV).*

9.6   Discussion and Summary

My computational results and photoelectron spectra demonstrate strong electrophilicity of OA. The monomer is characterized by a VDE of 1.10 eV and AEA of 0.72 eV. For comparison, the monomer of FA does not support a bound valence anion in the neighbourhood of the global minimum of the neutral. The unique electrophilicity of OA results primarily from the proximity of the carboxylic groups, which allows for the bonding C-C interaction in the SOMO of the anion. The intramolecular hydrogen bonds also contributes to the overall stability of the anion.

The PES spectrum of OA⁻ displays a well resolved structure and my computational results allowed for interpreting this structure in terms of vibronic anion→neutral transitions involving the most stable anionic and neutral conformers of C_{2h} symmetry. The distinct vibrational structure results from differences in potential energy surfaces of the neutral and the anion. The C-C bond contracts and the C-O bonds elongates upon excess electron attachment. In addition, the C-C=O and C-O-H angles expand and contract, respectively. The geometric changes are consistent with the nature of SOMO in the anion, which is bonding in the C-C region and antibonding in the C-O regions.
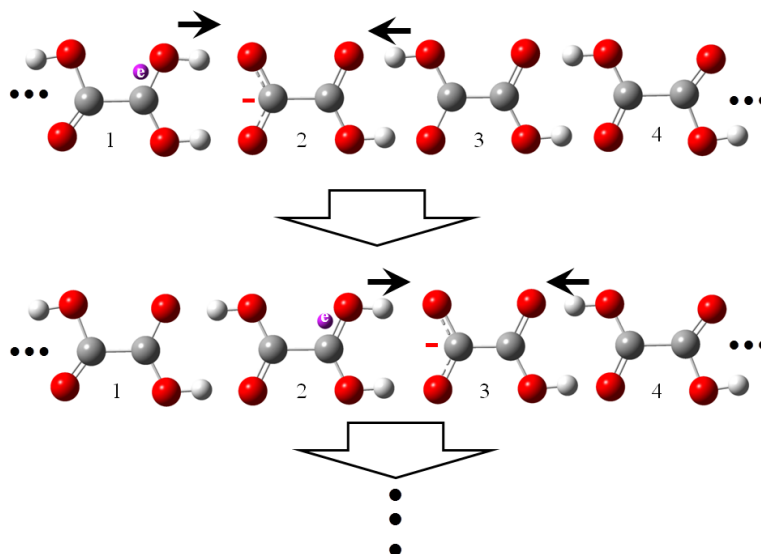
FIGURE 9.11: *Postulated mechanism of excess electron mobility along a linear chain of oxalic acid molecules. The arrows illustrate the motion of protons that would facilitate the excess electron mobility. The unpaired electron is represented by a violet dot and the excess charge by a red minus.*

The geometric distortion from the anion to the neutral can be accomplished by displacements along fully symmetric modes. The largest FC factors are associated with these vibrational modes and the 0-0 transition.

The experimental VDE for the anionic dimer and trimer of OA is 3.3 and 4.6 eV respectively. The computational results unravel that the neutral dimer is supported by cyclic hydrogen bonds. There are at least three types of dimers within a narrow energy range of 0.05 eV. All three structures undergo intermolecular proton transfer upon the excess electron attachment and are characterized by AEA exceeding 1.5 eV and VDE in the range of 3.1-3.3 eV. The computed values of VDE are in good agreement with the maximum of the PES peak at 3.3 eV.

A comparison of the VDE values for the $(OA)_n^-$ and $(FA)_n^-$ systems allows for the appreciation of intrinsic electrophilicity of OA. The values for $(OA)_n^-$ are 3.3 and 4.6 eV for $n = 2, 3$, respectively, with the corresponding values for $(FA)_n^-$ being 2.4, and 3.2 eV.[180]

Due to its dicarboxylic nature, OA can form linear chains supported by a series of cyclic hydrogen bonds. A model chain of OA molecules is illustrated in Figure 9.11. Based on my results for small anionic clusters of OA I postulated a mechanism of excess electron mobility along such a chain. I expect that the mobility of electrons will be supported by a synchronized motion of protons, as is common in proton-coupled electron transfer.[261–263] I envisage that protons will shuttle back and forth in the linear chain of

OA molecules, as dictated by the network of hydrogen bonds. The motion of an unpaired electron will follow the evolution of the electrophilic binding site, $-[C(OH)_2]^+$. The propagation step of the postulated mechanism is illustrated in Figure 9.11. In the top part, the unpaired electron is localized on the monomer 1 and the excess negative charge on monomer 2 (a deprotonated). Upon the shuttling motion of protons, the monomer 2 becomes the electrophilic binding site promoting transfer of the unpaired electron from 1 to 2, while the excess charge becomes localized on monomer 3. The phenomenon may be general for molecular chains supported by cyclic hydrogen bonds. The rate of electron transfer can be determined form the Marcus-Hush theory.[264] The problem is amenable to surface science experiments involving measurements of conductance[265] of molecular chains with cyclic hydrogen bonds.

Chapter 10

Suppressing Energy Barriers between Conformers on Molecular Potential Energy Surfaces

10.1 Abstract

Searches for the most stable molecular conformer are frustrated by energy barriers separating minima on the potential energy surface (PES). I have suggested that the barriers might be suppressed by subtracting selected force field terms from the original PES. The resulting deformed PES can be used in standard molecular dynamics (MD) or Monte Carlo simulations. The MD trajectories on the original and deformed PESs of ethanolamine differ markedly. The former gets stuck in a local minimum basin while the latter moves quickly to the global minimum basin. A concept of generalized simulated annealing protocols has been formulated.

10.2 Introduction

The conformational space of a molecule is spanned by its internal rotational bonds.[9] Specific combinations of intramolecular torsional angles give birth to molecular conformers, which are represented by minima on the molecular potential energy surface (PES). Each conformer has its own (in general different) electronic energy, vibrational energy levels and rotational constants. According to the Boltzmann distribution, conformers

with lower energies dominate the overall population, and therefore the molecular properties measured in experiments. The number of local minima on the molecular PES quickly increases with molecular size, and the most common optimization algorithms (driven by calculated forces) typically identify a minimum which is close to the initial structure, rather than the most stable conformer. It is therefore of ultimate importance to develop methods and algorithms for determination of the most stable molecular conformers.

I am primarily interested in PESs determined using reliable electronic structure methods. This is a broad range of methods from semi-empirical models, through density functional theory (DFT) methods, to Møller-Plesset and coupled cluster treatments of electron correlation. These methods are more time consuming than commonly used force fields,[266] but they offer a broad spectrum of accuracy, with the most advanced methods taking into account all practically important chemical and physical interactions and being able to break chemical bonds. The scaling of these methods is, however, quite steep, e.g. $N^3 - N^4$ for DFT, N^5 for the second-order Møller-Plesset method, etc., where N is the number of basis functions. The applicability of these reliable methods in structural predictions hinges on efficient global minimum search algorithms.

Finding the global minimum on a PES is frustrated by barriers that separate local minima. The most straightforward approaches that ignore barriers, such as scanning the PES with discrete geometrical increments along all $3N - 6(5)$ internal degrees of freedom or systematic searches focused on rotational bonds only,[9, 145, 267] suffer from combinatorial explosions of required calculations. These brute force methods provide, however, the most reliable benchmarks for all other, perhaps faster but less reliable, methods.

The most common methods for finding the most stable molecular structures are finite temperature Monte Carlo[32] and molecular dynamics (MD)[33] methods as well as genetic algorithms.[35, 267] The molecular dynamics method is of particular relevance here, and it is implemented in various simulated annealing algorithms.[86] The main idea is that a simulation initiated from any initial molecular structure and performed at a *sufficiently high temperature* (T) for a *sufficiently long time* has a chance to probe the basins of the global minimum and other low energy minima. The performance of the method depends critically on the height of energy barriers separating local minima and on the overall simulation time.[268, 269] One could think that by increasing the simulation temperature one would provide enough energy to overcome any existing barrier. Unfortunately, molecules decompose into smaller fragments in the course of high T simulations when reliable PESs are used.[270, 271] Performing extremely long simulations at a low T is not an attractive option due to significant computational runtimes. In practice, a compromise is made between the simulation temperature and the overall

simulation time and various protocols are used that define T as a function of time. Many valuable results were obtained using simulated annealing algorithms, though the identified low energy structure might be a local minimum rather than the global minimum. Again, the failure of the method is associated with the height of an energy barrier which separates the local minimum from the global minimum.

The past efforts to suppress energy barriers on a PES include the diffusion equation method (DEM), which transforms the original PES with multiple minima into a new PES with only one (the global) minimum.[36] Some other methods which deform the original potential energy surface in a different way include the distance scaling method,[37] the shift method,[38] the Gaussian density annealing method,[39, 40] and the so-called ant-lion strategy of changing the range of the potential.[41] While these methods proved to be robust for complex systems, they require an analytical expression for the original PES. Thus, they are not applicable to PESs determined using electronic structure methods. Another successful approach, the basin-hopping method,[34] in which the PES is transformed into a collection of interpenetrating staircases, requires a geometrical energy minimization for each stochastically selected molecular structure. This feature disfavours the method when combined with PESs determined using electronic structure methods, though successful applications have been reported.[42, 43]

Here I proposed another method to suppress, or even eliminate, barriers that separate various conformers on a PES determined by an electronic structure method. The deformed potential energy surface is obtained by subtracting selected *force field* terms from the original PES. First I will demonstrate that typical conformational barriers result from the torsional and van der Waals (vdW) force field terms. Next, I will compare constant temperature molecular dynamics trajectories for ethanolamine on the original PES (determined at the DFT level of theory), and on the deformed PES with the torsional and vdW terms subtracted. I will demonstrate that the MD trajectory moves quickly to the global minimum basin on the deformed PES, but it gets stuck in a local minimum basin on the original PES. Finally, I will discuss a concept of generalized simulated annealing protocols, in which both the temperature and potential energy can change as time progresses.

10.3 Methods

I started from a potential energy function, E_{elec}, defined by a reliable electronic structure method. In the current case I used density functional theory with the B3LYP hybrid exchange-correlation functional[100, 113] and 6-31G* atomic basis sets.[272] The global and local minima were identified using our Systematic Screening of Conformers

(SSC) tool.[25] DFT/B3LYP is a relatively fast electronic structure method, but still prohibitively slow for structural predictions based on long MD trajectories. In parallel to the electronic structure method I used a potential energy E_{FF} defined by a reliable force field (FF). The potential energy in a typical FF is given by the following expression:

$$E_{FF} = E_{str} + E_{bend} + E_{tors} + E_{cross} + E_{vdw} + E_{coul} \tag{10.1}$$

where E_{str} is the bond stretching energy, E_{bend} is the bond bending energy, E_{tors} is the torsional energy for rotations of molecular fragments around bonds, E_{cross} describes cross terms, e.g., bond-bond angle, and finally E_{vdw} and E_{coul} are the van der Waals (vdW) energy and the Coulomb energy, respectively.[90] There are two advantages resulting from using a FF. First, the time associated with a computation of E_{FF} is negligible in comparison with the time required to compute E_{elec}. Second, Equation 10.1 offers a dissection of the total potential energy into distinct components with clear chemical interpretations. I will use a very popular FF, AMBER,[221] but other FFs might be used as well.

My approach hinges on an assumption that *a priori* known terms of Equation 10.1 contribute to the barriers which separate local minima representing various molecular conformers. I defined a modified, or deformed, PES as

$$E_M(\mathbf{x}) = E_{elec}(\mathbf{x}) - \sum_i c_i E_{FF,i}(\mathbf{x}) \tag{10.2}$$

in which E_M indicates the modified potential energy, $E_{FF,i}$ is a specific FF term from the right hand side of Eq. 10.1, c_i is a linear coefficient (most typically $c_i = \pm 1.0$), and $\mathbf{x}$ represents a molecular geometry. The modified potential energy E_M is also used to calculate the modified gradient,

$$\mathbf{g}_M(\mathbf{x}) = \nabla E_M(\mathbf{x}), \tag{10.3}$$

which requires the gradients of E_{elec} and all $E_{FF,i}$s involved in Equation 10.2. This gradient will be used to represent forces acting on atoms moving on the modified PES, as dictated by the conventional MD equations.

The molecular dynamics simulations were performed for ethanolamine, Fig. 1, at 200 K and 400 K, with a time step of 0.5 fs, for a total of 20000 steps on both the original and modified PES. The Beeman algorithm[80] was used to predict accelerations, velocities and coordinates of nuclei. The resulting trajectories were illustrated by the root mean square deviations (RMSD) of positions of nuclei with respect to a reference structure,

which might be a local or the global minimum structure:

$$RMSD(t) = \sqrt{\frac{\sum_{i=1}^{3N}[x_i(t) - x_{i,ref}]^2}{3N}} \qquad (10.4)$$

in which $x_i(t)$ represents the ith geometrical coordinate at simulation time t, and $x_{i,ref}$ represents the ith geometrical coordinate at the reference structure. The current and reference structures were aligned to resemble each other using the VMD[167] command "measure fit". While advancing a MD trajectory I collected information about the most stable molecular structures and their complete potential energies. After the MD trajectory is completed I used the identified structures as initial geometries for standard geometry optimizations based on calculated forces.
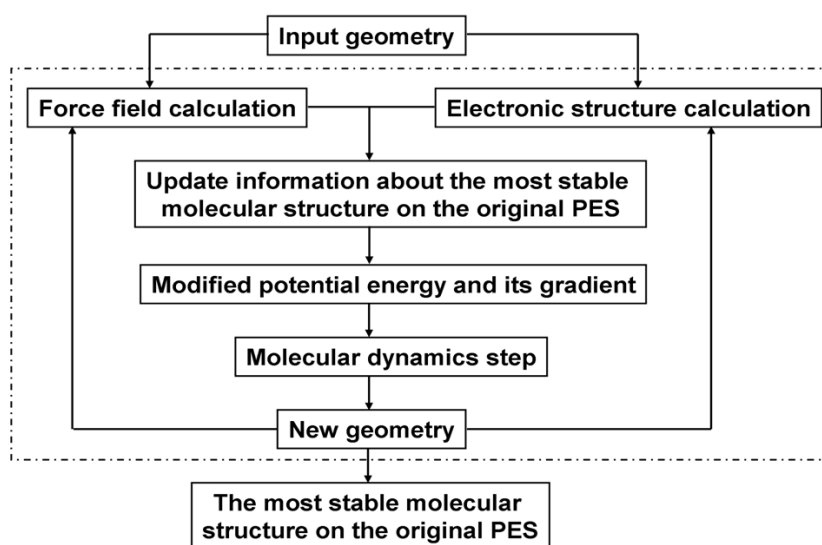

FIGURE 10.1: *A flowchart of the interface between an electronic structure method and a force field.*

All B3LYP calculations were performed with Gaussian 03 (G03).[75] Molecular mechanics calculations were performed either with G03 or Tinker.[273] I have developed an interface between G03 and Tinker that allowed propagating MD equations on the original and modified PES, see Fig. 10.1. The charge equilibration (QEq) method[74] was used to assign charges on atoms to account for Coulomb energy in the AMBER FF calculations performed with G03. Default atomic charges were used in Tinker.

10.4 Results

Ethanolamine has three rotational degrees of freedom and the OH and NH_2 groups can form an intramolecular hydrogen bond, which determines two chiral global minima GM1
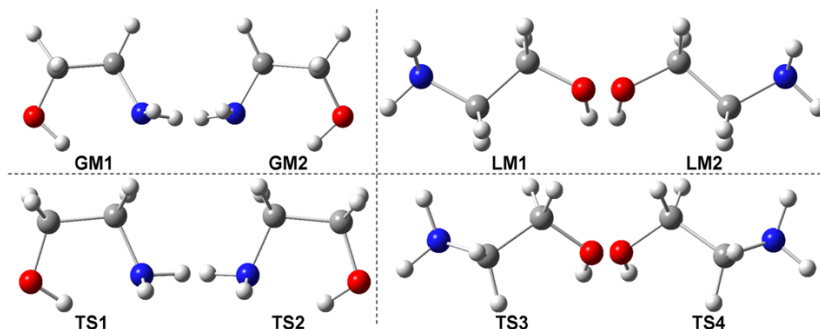
FIGURE 10.2: *Different minima and transition state structures of ethanolamine. Colour code: grey – carbon, white – hydrogen, red – oxygen, blue – nitrogen.*

and GM2 (see Fig. 10.2). In Fig. 10.3 I present a potential energy curve for motion around the O-C-C-N dihedral angle. The dihedral angle was increased or decreased, when starting from GM1 and GM2 respectively, with steps of 5°, and partial geometry optimizations were performed for each fixed value of the angle. Altogether, 72 structures were considered. This procedure unravelled six new stationary points on the PES (Figs. 10.2 and 10.3): one pair of chiral local minima (LM1 and LM2) and two pairs of chiral transition states (TS1 and TS2, TS3 and TS4). Subsequent intrinsic reaction coordinate[274] calculations were performed for fully optimized transition state structures. It was confirmed that both TS1 and TS2 connect GM1 and GM2, TS3 connects LM1 and GM1, and TS4 connects LM2 and GM2. It should be pointed out that there is also a Cs symmetry stationary point (denoted as SP0) between TS1 and TS2, but it was abandoned because of two negative curvatures.
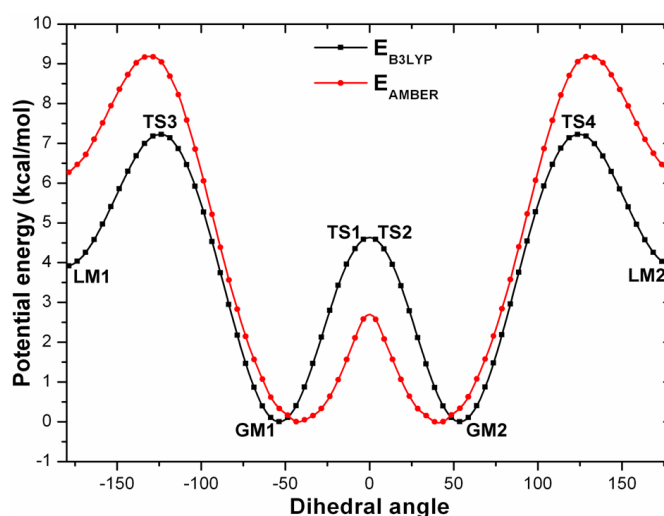

FIGURE 10.3: *Original potential energy curve for ethanolamine as a function of the O-C-C-N dihedral angle. The energies are determined at the B3LYP and AMBER force field levels.*

Based on the 72 partially optimized geometries, I performed single point AMBER FF
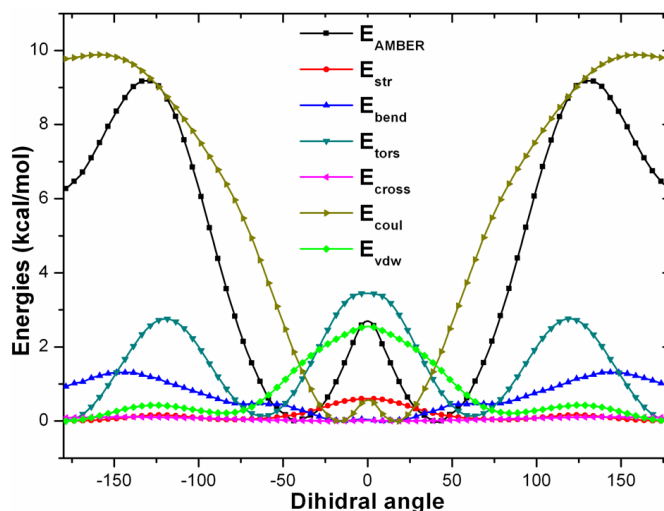
FIGURE 10.4: *The AMBER force field energy and its components as a function of the O-C-C-N dihedral angle.*

calculations. A comparison between the B3LYP and AMBER potential energy curves is presented in Fig. 10.3 and a dissection of the AMBER energy, given by Equation 10.1, is illustrated in Fig. 10.4. In Figs. 10.3 and 10.4 the zero of energy was set to the global minimum of each curve. Fig. 10.3 illustrates that the AMBER curve properly reproduces the main features of the B3LYP curve. The angles at which stationary points develop are similar, though some of the barrier heights are inaccurate by $\pm 2 - 3$ kcal/mol.

The results presented in Fig. 10.4 illustrate that the barriers on the total potential energy curve result primarily from the torsional and vdW terms, whereas the Coulomb term displays the largest overall variation as a function of the dihedral angle. The last finding is consistent with the dominant role of the intramolecular $OH \cdots NH_2$ hydrogen bond in the GM1 and GM2 structures.

In Fig. 10.5 I compared potential energy curves on the original and modified potential energy surface. The modified surface was obtained by subtracting the AMBER torsional and vdW terms from the B3LYP electronic energy (Fig. 10.5a) or from the total AMBER energy (Fig. 10.5b). As visualized in Fig. 10.5a, the transition barrier between LM1 and GM1 is suppressed from 3.3 to 0.4 kcal/mol. Similarly, the transition barrier from GM1 to GM2 is suppressed from 4.6 to 0.5 kcal/mol. Thus both energy barriers were suppressed by almost 90%. There are two sources of the remaining barriers. First, other energy terms contribute to the barriers, e.g., the Coulomb term. Second, the residual barriers reflect inherent inconsistencies between the electronic structure and FF models. An important observation is that the positions of the global minima are similar on the original and modified PESs.
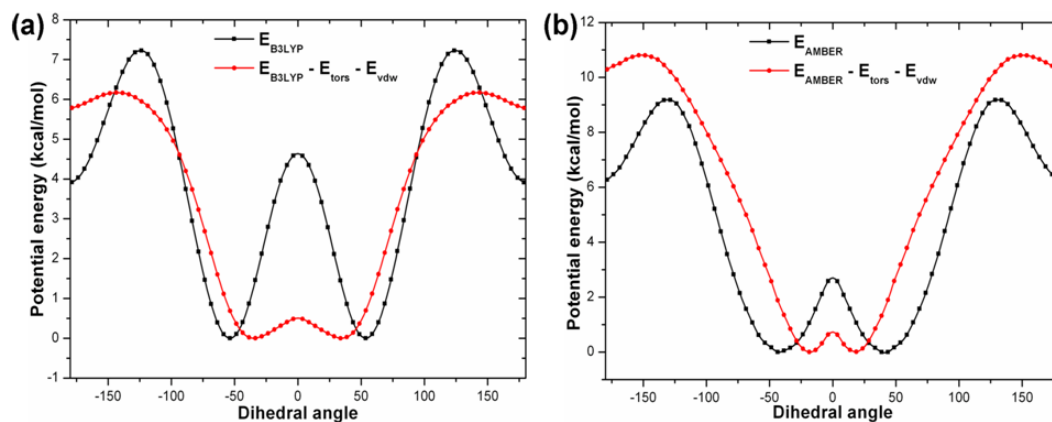
FIGURE 10.5: *Deformation of the original PES of ethanolamine at the (a) B3LYP and (b) AMBER levels.*

The results of MD simulations for ethanolamine on the original and modified PES are summarized in Fig. 10.6 for T = 200 and 400 K. The simulations were initiated from the local minimum geometry and the presented RMSD values were calculated with respect to the local (Figs. 10.6 a,c) and global (Figs. 10.6 b,d) minimum structures. For the T=400 K trajectory on the deformed PES I also calculated the RMSD values with respect to SP0, i.e., the C_s symmetry stationary point with two negative curvatures, which separates GM1 and GM2, see Fig. 10.6d. The RMSD values at T=200 K (Figs. 10.6 a,b) demonstrate that the trajectory on the original PES gets trapped in the local minimum. On the other hand, the trajectory on the modified PES settles in the basin of the global minimum after less than 7 ps. For simulations performed at 400 K, the trajectory on the modified PES moves to the basin of the global minimum within 1 ps. On the original PES, it takes about 7 ps. On the deformed PES, the RMSD values calculated with respect to GM2 show large amplitude oscillations, which are finally dampened (a black curve in Fig. 10.6d). This pattern is associated with initial oscillations between the GM1 and GM2 structures followed by localization in the basin of GM2. This interpretation is further strengthened by the plot of RMSD calculated with respect to SP0 (a blue curve in Fig. 10.6d).

10.5   Discussion

I have demonstrated that the energy barriers separating different conformers of ethanolamine can be suppressed by subtracting the torsional and vdW terms from the total potential energy, Fig. 10.5. This has beneficial effects when running MD trajectories on the deformed PESs. These trajectories move quickly to the global minimum basin, Fig. 10.6. Trajectories run on the original PESs might get stuck in the local minimum basin.
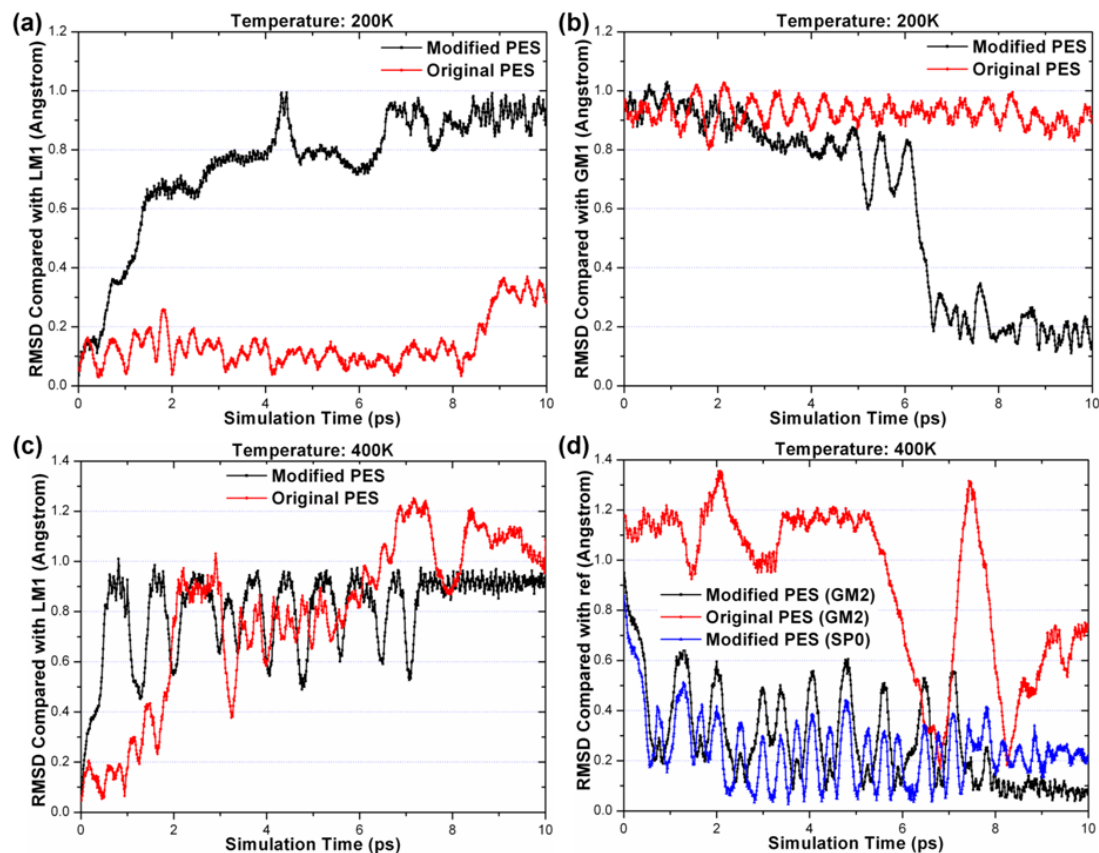
FIGURE 10.6: *The RMSD values as a function of time for trajectories on the modified and original PES at 200 K (a,b) and 400 K (c,d) and calculated with respect to LM1 (a,c), GM1 (b) and GM2 and SP0 (d).*

Increasing the simulation temperatures helps to overcome the barrier, but there is always some risk associated with this procedure. I found that for MD trajectories run on the original PES at T=2000 K, one observes thermal decomposition of ethanolamine, i.e., a thermal dissociation of the C-H bonds. This is clearly an undesirable scenario. One could suppress these unwelcome events by strengthening the stretching intramolecular degrees of freedom. This would correspond to deformation of the original PES according to Equation 10.2, but this time I would add some stretching terms to the original PES, i.e., $c_{FF,str} < 0$. One could think about a complex deformation, in which the torsional and vdW terms are subtracted to suppress energy barriers separating various conformers, the intramolecular stretching terms are added to avoid thermal decomposition, and the MD calculations are performed at a temperature as high as a few thousands of K.

The results presented in Fig. 10.5 demonstrate that after removal of the torsional and vdW terms there is a deep basin for the global minimum. The existence of this basin results from the intramolecular OH$\cdots$NH$_2$ hydrogen bond. What would happen if a molecule was complex enough to support several distinct conformers, each with its own hydrogen bond and, therefore, favourable Coulomb interactions? There would be

several basins on the PES and the deformations described so far would not suppress the barriers separating them. However, one could suppress these barriers by subtracting the Coulomb term from the original PES, i.e., $c_{FF,coul} > 0$ in Equation 10.2. This reasoning leads to a concept of generalized simulated annealing protocols. In standard protocols one *a priori* defines temperature for each time interval of the MD simulation. Here I suggest extending this concept by allowing the PES *and* temperature to change as time progresses. In addition, I suggest that these changes should be made "*on the fly*" based on local information available from the advancing trajectory. For example, to create an opportunity for an MD trajectory to move from one basin of favourable Coulomb interactions to another, I suggest a temporary removal of the Coulomb term from the original PES. I expect the code to recognize the fact that a basin of attractive Coulomb interactions has been encountered. In response, the code should modify the PES by removing those Coulomb interactions that are characteristic of the encountered basin, and the trajectory should be propagated for a sufficiently long time to leave the basin. It would make sense during this interval to temporarily have the temperature increased, the stretching degrees of freedom strengthened, and the torsional and vdW barriers suppressed. Later, while the trajectory explores a new basin of favorable Coulomb interactions, the Coulomb term of the former basin should be restored; the temperature lowered, etc. The idea of generalized dynamically modified simulated annealing protocols based on deformed PESs will be explored in this ongoing project.

10.6   Summary

I have defined a modified potential energy surface as a linear combination of the original potential energy surface and the terms typically encountered in molecular force fields, such as stretching, bending, torsional, Coulomb, and van der Waals. Depending on the circumstances I deform the surface in different ways. For example, to suppress barriers separating different conformers, I recommend subtracting the torsional and vdW terms from the original PES. To avoid thermal decomposition into smaller molecular fragments in simulations performed at elevated temperatures, I recommend adding stretching terms to the original PES. To create an opportunity for a MD trajectory to get out of a basin of attractive Coulomb interactions, I recommend a temporary removal of selected Coulomb terms.

I formulated a concept of generalized simulated annealing protocols, in which both the temperature and potential energy can change as time progresses. The changes are made "*on the fly*" based on local information from the advancing trajectory rather than decided *a priori*, before starting the MD trajectory. I developed an interface between

Gaussian 03 and Tinker, which allows running MD simulations on the original PES, which typically results from electronic structure calculations, and on an arbitrarily deformed PES.

I performed a detailed study of the conformational space of ethanolamine. I demonstrated that the torsional and van der Waals energies contribute most to energy barriers separating different minima. A deformed PES, with the torsional and vdW terms subtracted, proved to be useful in MD simulations. The MD trajectory moved quickly to the global minimum basin on the deformed PES, but it got stuck in a local minimum basin on the original PES.

Chapter 11

Summary and Future Work

In this work, I focused my attention on developing new methods for the determination of molecular structures, and their applications to systems with intra and intermolecular hydrogen bonds. My targets were molecular structures of various sizes. I have developed a software tool for combinatorial generation of tautomers and conformers of molecules, PESST. I also developed a technique which is based on temporary modifications of the potential energy surface in order to improve efficiency of molecular dynamics and promote even sampling of the PES.

I have demonstrated usefulness of PESST by performing a systematic search for the most stable structures of neutral and anionic phenylalanine (Phe) using electronic structure methods. For the neutral canonical tautomer I found out that the conformers *with* and *without* the intramolecular (O)H$\cdots$NH$_2$ hydrogen bond are similarly stable, within the error bars of my method. A unique IR signature of the conformer without the hydrogen bond has been identified. I also considered anions of Phe, both valence type and dipole-bound. I have found out that tautomers resulting from proton transfer from the carboxylic OH to the phenyl ring do support valence anions that are vertically strongly bound, with electron vertical detachment energies (VDE) in a range of 3.2-3.5 eV. The most stable conformer of these valence anions remained adiabatically unbound with respect to the canonical neutral by only 2.17 kcal/mol at the CCSD(T)/aug-cc-pVDZ level. The valence anions of Phe identified in this report have recently been observed experimentally[180]. The most stable conformer of canonical Phe is characterized by an adiabatic electron affinity of 53 meV (a dipole-bound state).

Although PESST tool allows to perform rigorous conformational and tautomeric searches, one experiences a combinatorial explosion of molecular structures as the number of rotatable bonds increases. The bottleneck arises at the screening level when electronic structure methods are employed as many initial structures has to be considered. This

problem can be minimised through activation of few bonds but this approach might lead to incomprehensive search. The most diligent way of tacking this problem is to perform minimisation using force fields at the screening level and then the resulting structures can be reoptimised at higher level of theory for accuracy. This approach will be implemented in the PESST tool as an optional feature in the near future.

Searches for the most stable molecular conformer are frustrated by energy barriers separating minima on the potential energy surface (PES). I have suggested that the barriers might be suppressed by subtracting selected force field terms from the original PES. The resulting deformed PES can be used in standard molecular dynamics (MD) or Monte Carlo simulations. The MD trajectories on the original and deformed PESs of ethanolamine differ markedly. The former gets stuck in a local minimum basin while the latter moves quickly to the global minimum basin. A concept of generalized simulated annealing protocols has been formulated. In this protocol, the changes are made "*on the fly*" based on local information from the advancing trajectory rather than decided *a priori*, before starting the MD trajectory. I also propose to invert the potential energy surface if the molecular dynamics are stuck on the local minima. This procedure will convert local minima to transition states and *vice versa* and those simulation stiuck on the local minima will found themselves on the transition states and this will speed up the search. Note that the inversion will be temporary.

I considered stability of dimers formed by molecules that can exist in different conformational states. I have developed a protocol that allows the dissection of the total stabilisation energy into one-body conformational and deformational components and the two-body interaction energy term. Interplay between these components determines the overall stability of the dimer. The protocol has been tested on the dimers of oxalic acid. The global minimum stability results from a balancing act between a moderately attractive two-body interaction energy and small repulsive one-body terms. I have analysed zero-point vibrational corrections to the stability of various conformers of oxalic acid and their dimers. I have found that minimum energy structures with the most stabilising sets of hydrogen bonds have the largest zero-point vibrational energy, contrary to a naive anticipation based on red shifts of OH stretching modes involved in hydrogen bonds.

My computational results along side with experimental results from my collaborators demonstrated an unusual electrophilicity of oxalic acid (OA), the simplest dicarboxylic acid. The electrophilicity results primarily from the bonding carbon-carbon interaction in the SOMO orbital of the anion, but it is further enhanced by intramolecular hydrogen bonds. The well-resolved structure in the photoelectron spectrum has been reproduced

theoretically, based on Franck-Condon factors for the vibronic anion→neutral transitions. The excess electron binding energies in the dimer and trimer of OA become very significant due to intermolecular proton transfer, with the corresponding VDE values of approximately 3.3 and 4.6 eV. I have postulated a mechanism of excess electron mobility along molecular linear chains supported by cyclic hydrogen bonds. Due to its dicarboxylic nature, OA can form linear chains supported by a series of cyclic hydrogen bonds. A model chain of OA molecules is illustrated in Figure 9.11. Based on my results for small anionic clusters of OA I postulated a mechanism of excess electron mobility along such a chain. I expect that the mobility of electrons will be supported by a synchronized motion of protons, as is common in proton-coupled electron transfer.[261–263] I envisage that protons will shuttle back and forth in the linear chain of OA molecules, as dictated by the network of hydrogen bonds. The motion of an unpaired electron will follow the evolution of the electrophilic binding site, $-[C(OH)_2]^+$. The propagation step of the postulated mechanism is illustrated in Figure 9.11. In the top part, the unpaired electron is localized on the monomer 1 and the excess negative charge on monomer 2 (a deprotonated). Upon the shuttling motion of protons, the monomer 2 becomes the electrophilic binding site promoting transfer of the unpaired electron from 1 to 2, while the excess charge becomes localized on monomer 3. The phenomenon may be general for molecular chains supported by cyclic hydrogen bonds. The rate of electron transfer can be determined form the Marcus-Hush theory.[264] The problem is amenable to surface science experiments involving measurements of conductance[265] of molecular chains with cyclic hydrogen bonds.

Searches for the most stable molecular conformer are frustrated by energy barriers separating minima on the potential energy surface (PES). I have suggested that the barriers might be suppressed by subtracting selected force field terms from the original PES. The resulting deformed PES can be used in standard molecular dynamics (MD) or Monte Carlo simulations. The MD trajectories on the original and deformed PESs of ethanolamine differ markedly. The former gets stuck in a local minimum basin while the latter moves quickly to the global minimum basin.

I formulated a concept of generalized simulated annealing protocols, in which both the temperature and potential energy can change as time progresses. The changes are made "on the fly" based on local information from the advancing trajectory rather than decided a priori, before starting the MD trajectory. Only interactions that contribute to the are subtracted. "Diagnostic tools" that identify origins of barriers have recently been implemented in my code.

Appendix A

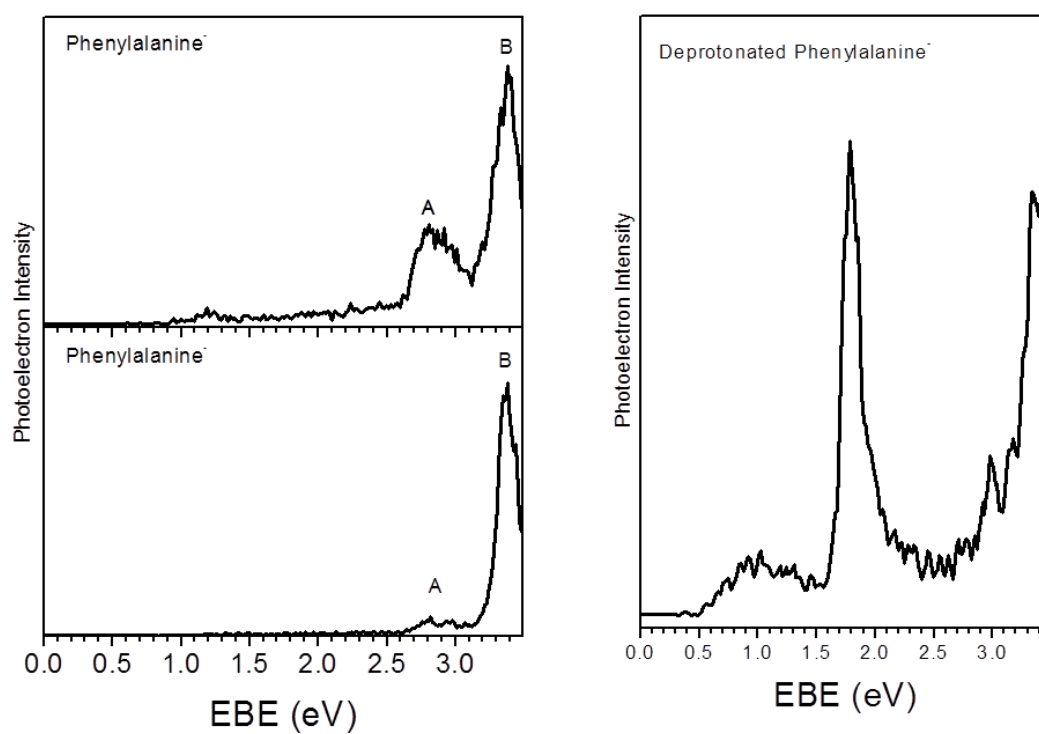# Supporting Information to Chapter 5

FIGURE A.1: *Photoelectron spectra of intact Phenylalanine anion and deprotonated phenylalanine anion*
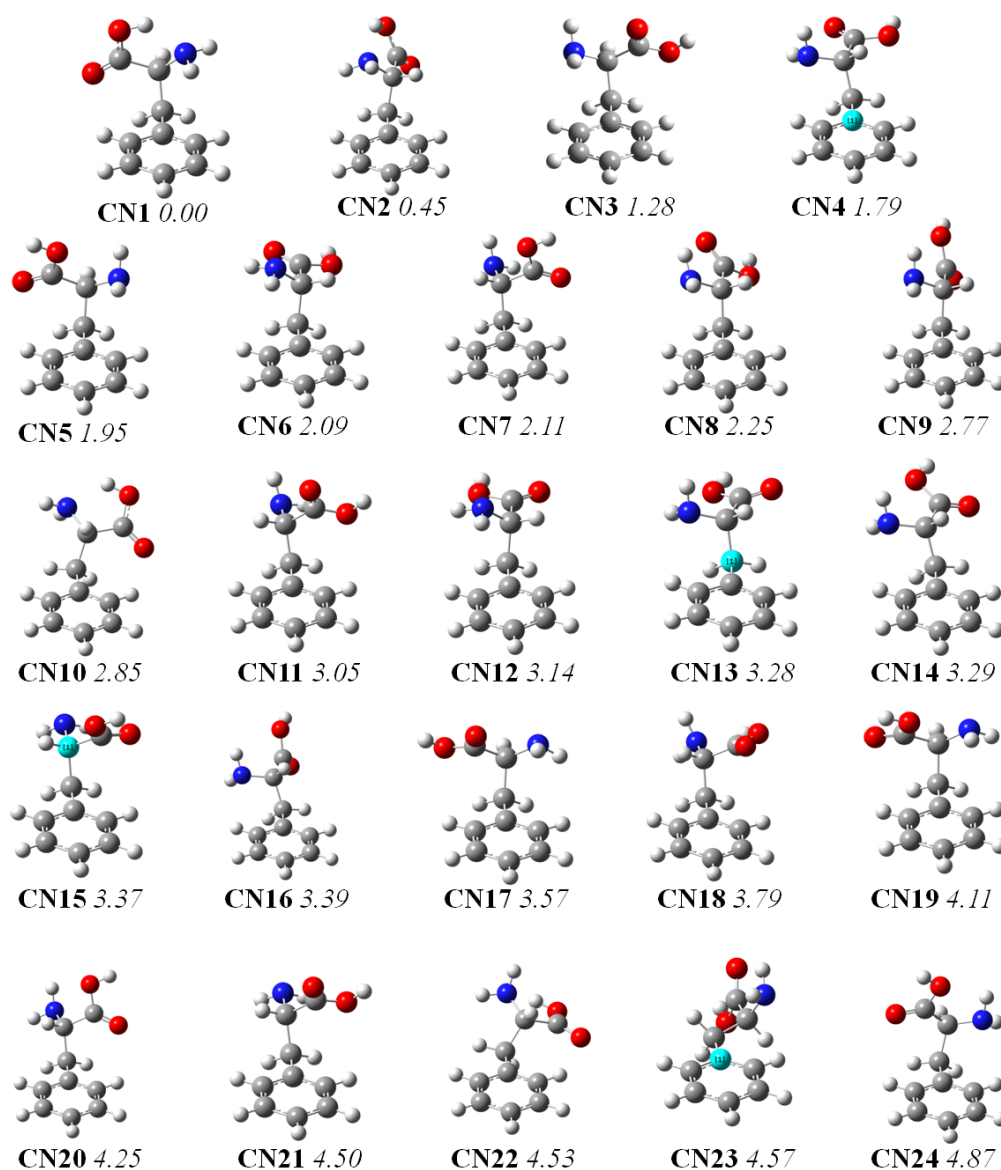
FIGURE A.2: *The most stable conformers of neutral phenylalanine. The relative electronic energies (kcal/mol) with respect to the most stable canonical neutral **CN1** and geometrical parameters (Å) characterised at the MP2/ADZ level of theory. The prescreening was done at DFT(B3LYP)/6-311G* level of theory.*
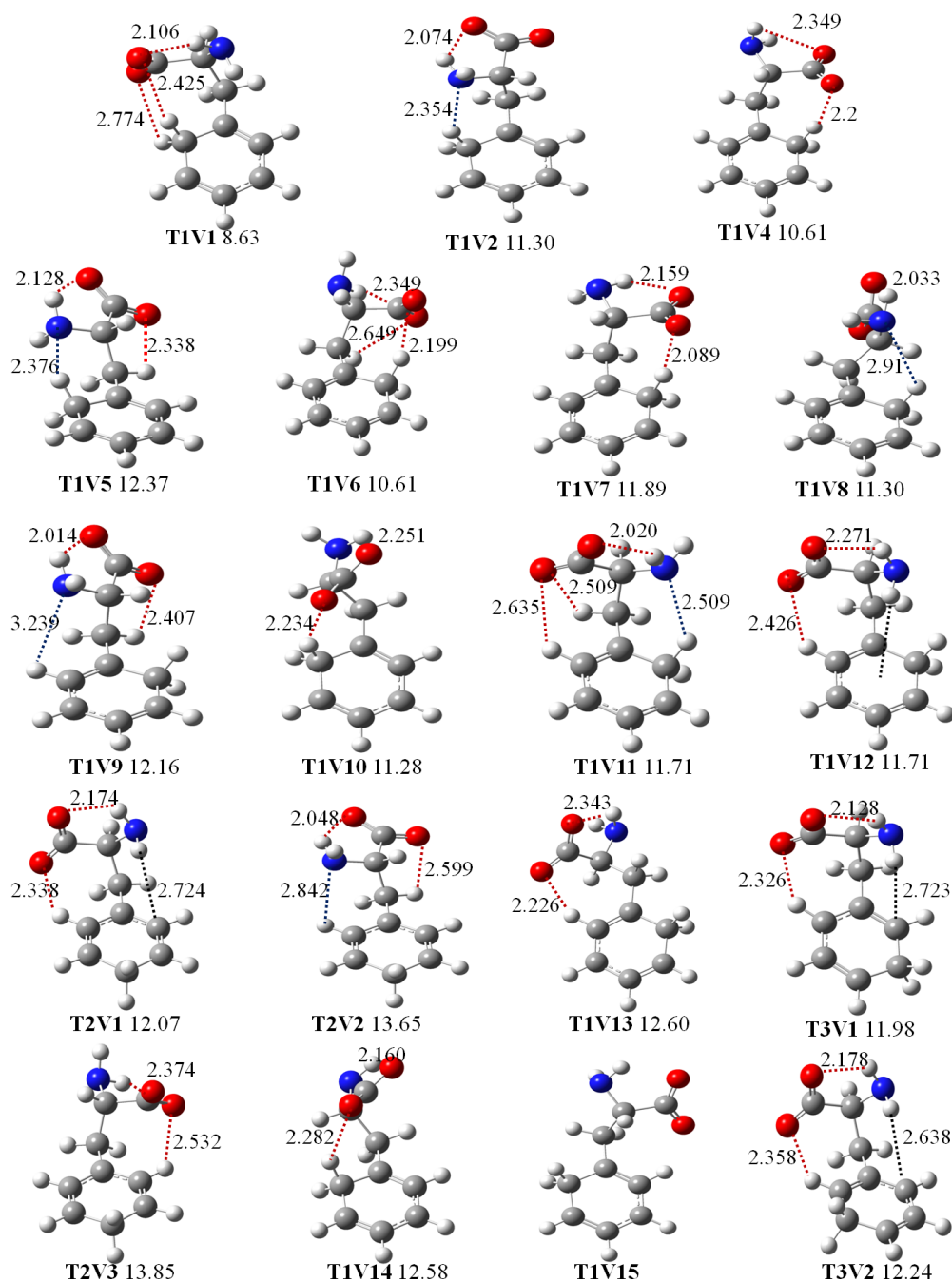
FIGURE A.3: *The most stable tautomers/conformers of valence anions of phenylalanine. The relative electronic energies (kcal/mol) with respect to the most stable canonical neutral **CN1** and geometrical parameters (Å) characterised at the MP2/ADZ level of theory. The pre-screening was done at DFT(B3LYP)/6-31G* level of theory.*
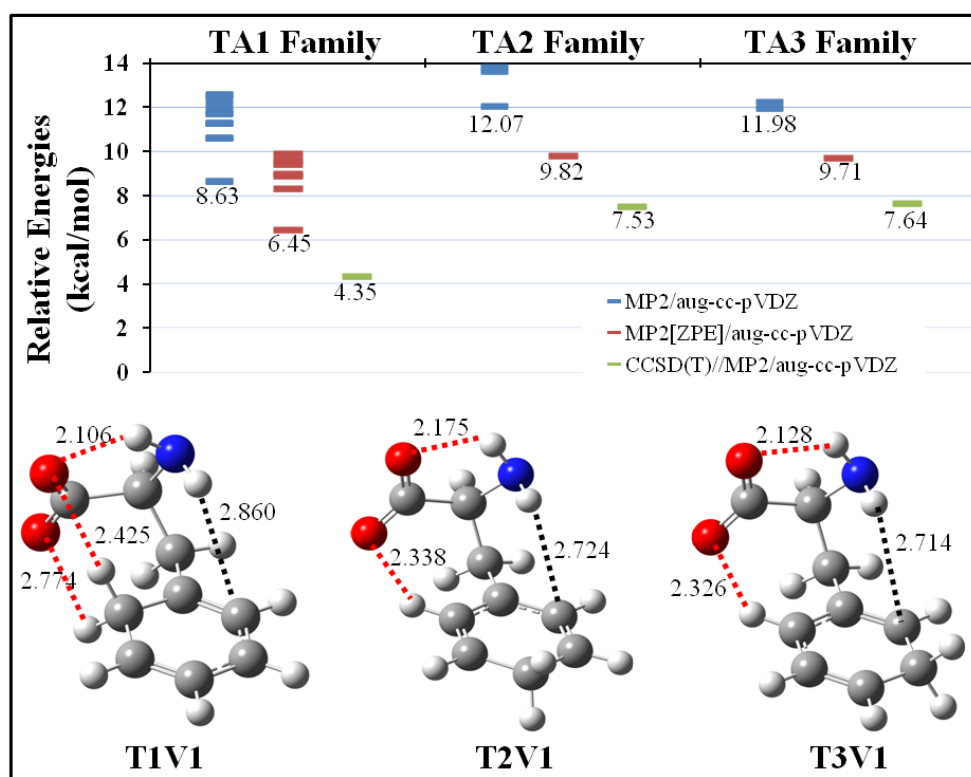
FIGURE A.4: *Three families of valence anions of phenylalanine, **T1V**, **T2V**, and **T3V**. The spread of electronic energies within each family is illustrated in the top-left part (MP2/ADZ level of theory). Structures of the most stable conformer within each family in the bottom part with distances in Å.*

Appendix B

Supporting Information to Chapter 6

B.1   Anion Photoelectron Spectroscopy Methodology

Negative ion photoelectron spectroscopy (PES) is conducted by crossing a mass-selected beam of negative ions with a fixed frequency photon source and energy analyzing the resultant photodetached electrons. This technique is governed by the energy-conserving relationship $hv = EKE + EBE$, where hv is the photon energy, EKE is the measured electron kinetic energy, and EBE is the electron binding energy. Briefly, both mass spectra and photoelectron spectra were collected on an apparatus consisting of a laser vaporization source employing a Nd:YAG laser, a linear time-of-flight mass spectrometer for mass analysis and selection, a second Nd:YAG laser used for photo-detachment, and a magnetic bottle used for electron energy analysis. The details of our apparatus have been described elsewhere.[190] The photoelectron spectra of Tyr$^-$ and (Tyr-H)$^-$ were measured with 4.86 and 3.49 eV photons, respectively.

All valence anions in these experiments are formed using a laser vaporization ion source. This source consisted of a rotating, translating Tyr-coated metal rod (Cu or Ag), a laser beam entrance port, a pulsed gas valve to feed pulses of helium into the laser-sample interaction region, and a gas expansion exit nozzle. Typically, helium gas at 4 bar was expanded in synchronization with laser ablation pulses. The Tyr-coated rods were prepared by pressing Tyr powder directly onto the metal rod to form a thin layer on its surface. The Tyr coating was then ablated at very low laser power with the second harmonic (532 nm) of a Nd:YAG laser. We speculate that the role of the metal rod was to supply photoemitted electrons. No mixed metal/Tyr cluster anions were observed.
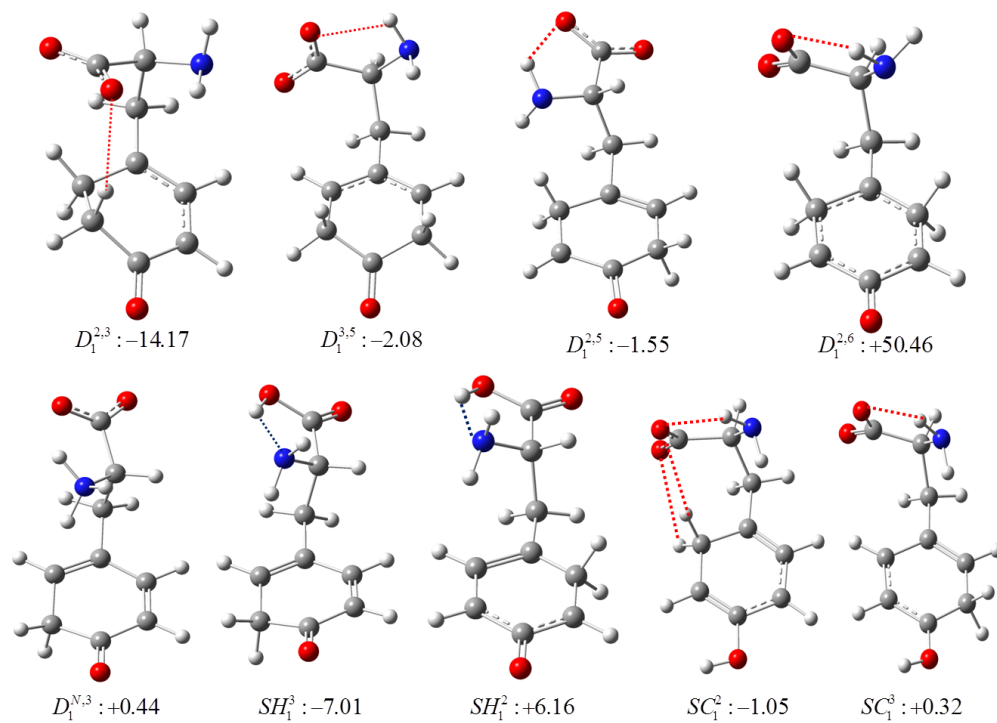
FIGURE B.1: *Families of valence anions of tyrosine, $D^{2,3}, D^{3,5}, D^{2,5}, D^{2,6}, SH^2, SH^3, SC^2$ and SC^3 considered in this study. The relative energies (in kcal/mol) with respect to the most stable canonical neutral N1, were calculated at B3LYP/aug-cc-pvdz. Each structure is the most stable conformer within its family.*
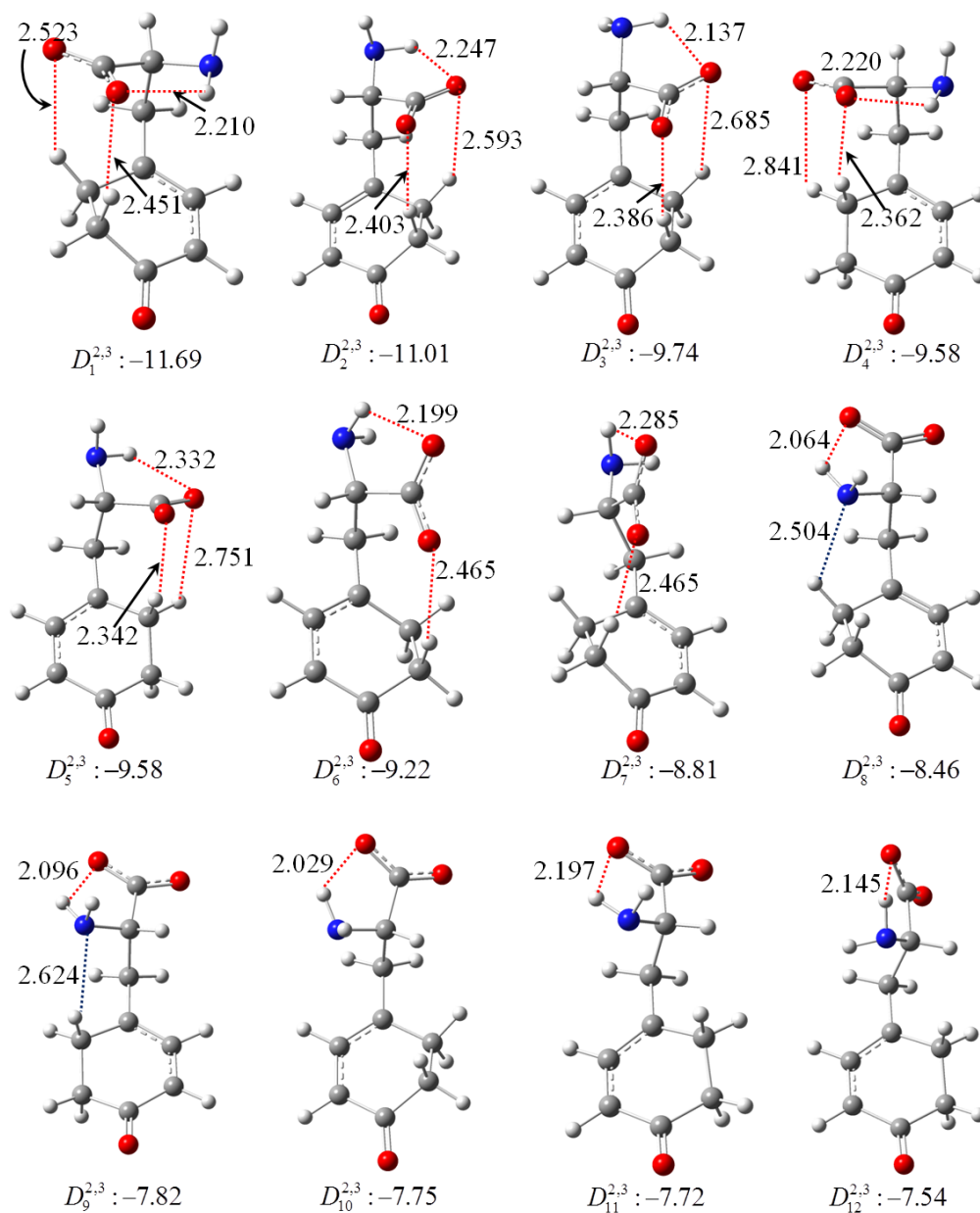
FIGURE B.2: *The principal geometrical parameters (in Å) of various doubly proton transferred anionic conformations of tyrosine characterised at MP2/aug-cc-pVDZ level of theory. The protons were transferred from the carboxylic and the hydroxyl groups into the ring. The CCSD(T) relative energies (in kcal/mol) were calculated at MP2 optimal geometries.*

| Anions | Relative Energies *w.r.t.* Neutral(N1), (kcal/mol) | | | | | |
| | 6-311+G* | aug-cc-pVDZ | | | | |
	B3LYP	B3LYP	B3LYP $+ E_0^{vib}$	MP2	MP2 $+ E_0^{vib}$	CCSD(T)
$D_8^{2,3}$	-23.51	-15.00	-17.97	-2.02	-3.39	-8.46
$D_2^{2,3}$	-22.79	-14.53	-17.38	-4.83	-6.02	-11.01
$D_9^{2,3}$	-22.93	-14.43	-17.45	-1.22	-2.65	-7.82
$D_{10}^{2,3}$	-22.62	-14.31	-17.42	-1.35	-2.84	-7.75
$D_1^{2,3}$	-22.45	-14.17	-16.95	-5.77	-6.96	-11.69
$D_{11}^{2,3}$	-22.39	-14.07	-17.12	-1.39	-2.82	-7.72
$D_6^{2,3}$	-21.94	-13.80	-16.72	-2.50	-3.85	-9.22
$D_7^{2,3}$	-21.63	-13.62	-16.54	-2.09	-3.50	-8.81
$D_{12}^{2,3}$	-21.82	-13.47	-16.50	-1.11	-2.58	-7.54
$D_3^{2,3}$	-21.50	-13.46	-16.37	-3.55	-4.83	-9.74
$D_{13}^{2,3}$	-21.67	-13.44	-16.51	-0.95	-2.38	-7.25
$D_4^{2,3}$	-21.62	-13.40	-16.31	-3.40	-4.77	-9.58
$D_5^{2,3}$	-21.60	-13.29	-16.20	-2.28	-3.54	-9.58

TABLE B.1: *Relative stabilities (in kcal/mol, with respect to N1 structure) at different levels of theory for the doubly proton transferred anionic conformations of tyrosine. The protons were transferred from the carboxylic and the hydroxyl groups on to the ring.*

	Relative Energies *w.r.t* Neutral(N1), (kcal/mol)				
	6-311+G*	aug-cc-pVDZ			
Anions	B3LYP	B3LYP	B3LYP $+ E_0^{vib}$	MP2	MP2 $+ E_0^{vib}$
SH_1^3	-11.28	-7.01	-10.15	3.87	2.00(-0.04
SH_2^3	-11.11	-6.90	-10.08	4.13	2.07(-0.04)
SH_5^3	-8.51	-4.03	-8.73	6.06	3.59(2.66)
SH_4^3	-8.21	-4.03	-7.31	6.00	3.86(1.76)
SH_6^3	-8.20	-4.65	-8.19	5.53	3.89
SH_{11}^3	-8.03	-3.88	-7.08	6.27	4.34
SH_3^3	-7.92	-4.26	-7.86	5.79	2.17(1.34)
SH_8^3	-7.84	-4.53	-8.22	6.77	4.21
SH_7^3	-7.66	-4.04	-7.31	6.10	4.14
SH_{10}^3	-6.91	-3.58	-7.43	7.09	4.30
SH_9^3	-6.84	-3.20	-6.55	6.55	4.23
SH_{12}^3	-5.53	-1.90	-5.47	8.21	6.09
SH_{13}^3	-5.43	-1.61	-5.31	10.21	7.66
SH_{14}^3	-5.27	-1.47	-5.13	10.09	7.72

TABLE B.2: *Relative stabilities (in kcal/mol with respect to N1 structure) at different levels of theory for the singly proton transferred conformations of anionic tyrosine. The proton was transferred from the hydroxyl group on to the ring.*
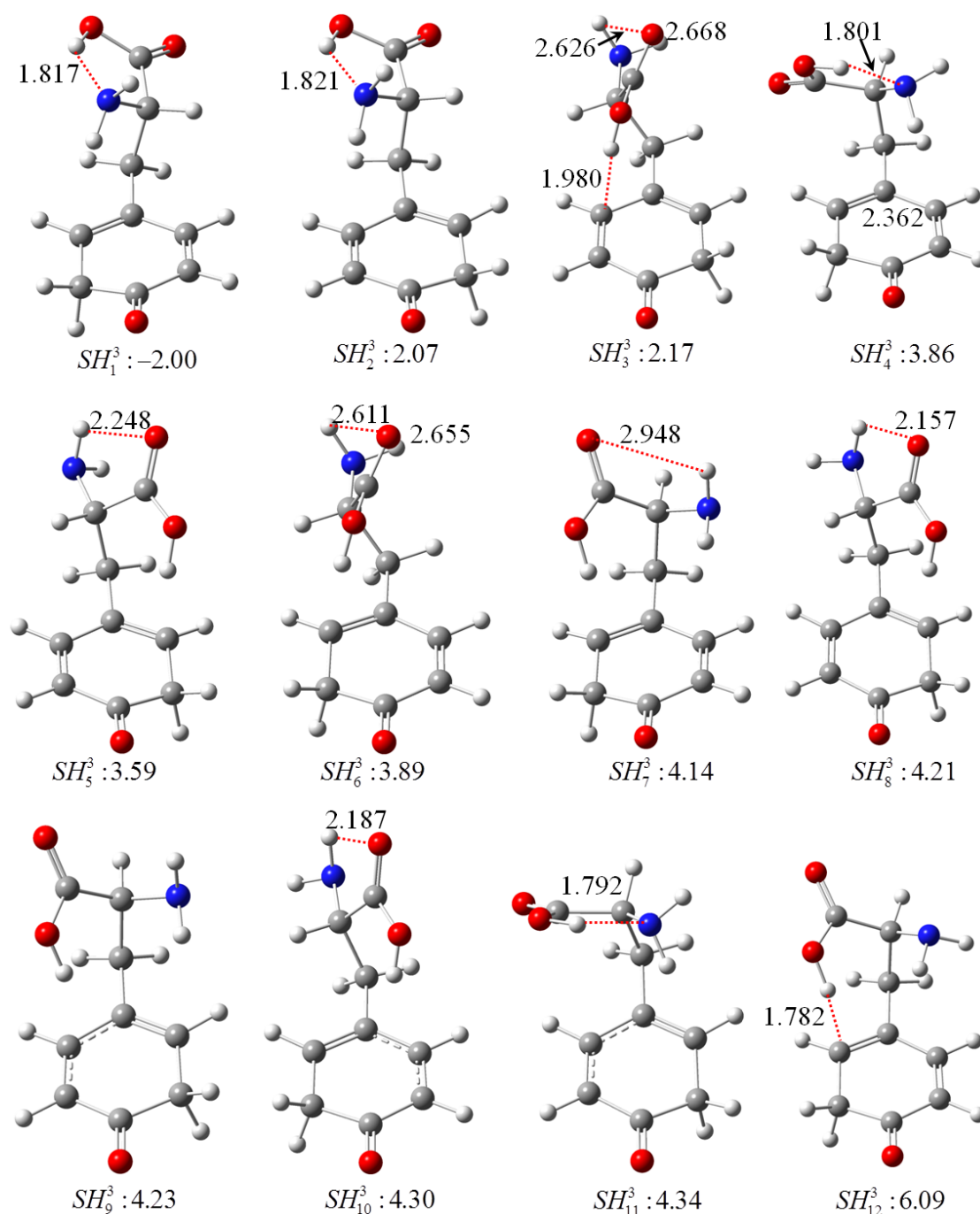
FIGURE B.3: *The principal geometrical parameters (in Å) and the MP2 relative energies (in kcal/mol, corrected for zero point vibrations) with respect to N1 structure, of various singly proton transferred conformations of anionic tyrosine characterised at MP2/aug-cc-pVDZ level of theory. The protons were transferred from the carboxylic and the hydroxyl groups on to the ring.*
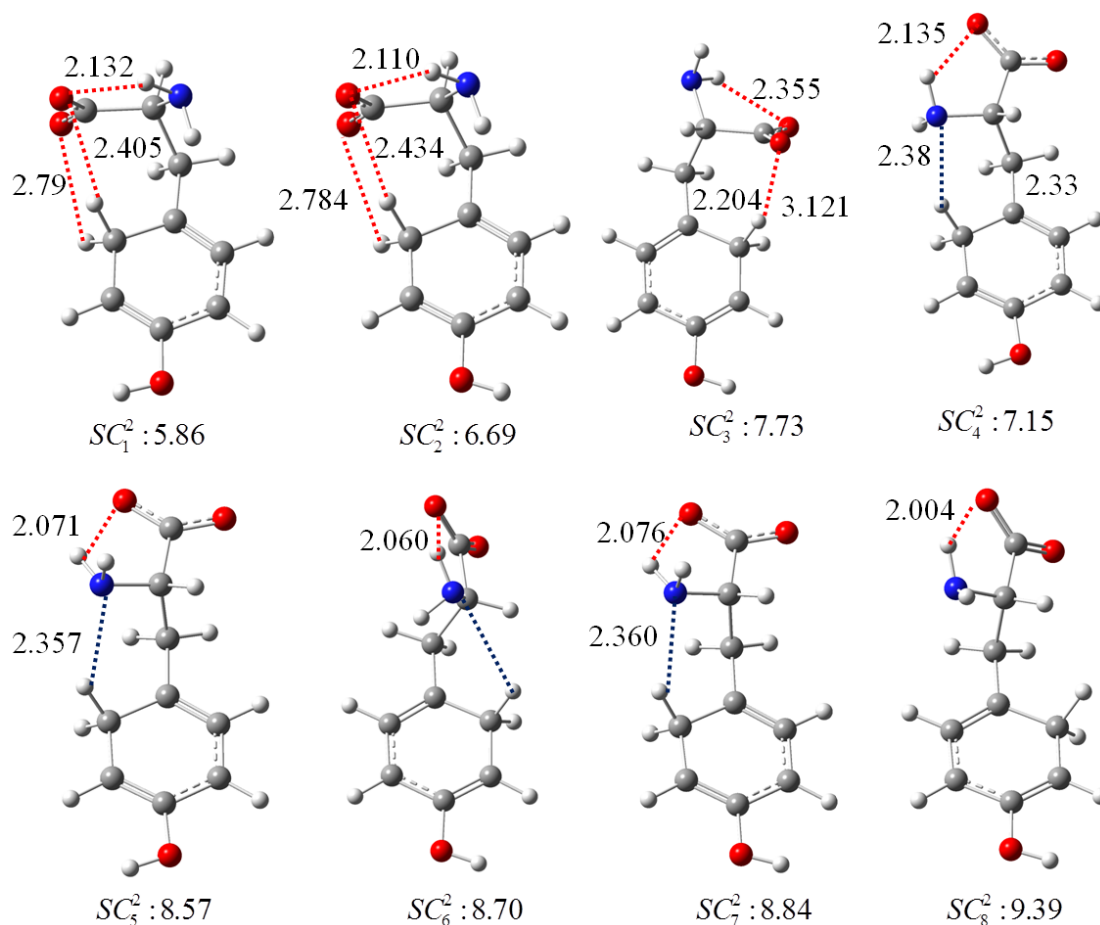
FIGURE B.4: *The principal geometrical parameters (in Å) and the MP2 relative energies (in kcal/mol, corrected for zero point vibrations) with respect to N1 structure of various singly proton transferred conformations of anionic tyrosine characterised at MP2/aug-cc-pVDZ level of theory. The proton was transferred from the carboxylic group on to the ring.*

	Relative Energies *w.r.t* Neutral(N1), (kcal/mol)				
	6-311+G*	aug-cc-pVDZ			
Anions	B3LYP	B3LYP	B3LYP + E_0^{vib}	MP2	MP2 + E_0^{vib}
SC_5^2	-4.92	-1.30	-4.94	10.88	8.57
SC_1^2	-4.30	-1.05	-4.54	8.02	5.86(3.46)
SC_7^2	-4.23	-0.83	-4.64	11.48	8.84
SC_3^2	-3.95	-0.61	-4.16	9.98	7.73(5.13)
SC_4^2	-3.70	-0.16	-3.82	9.57	7.15
SC_8^2	-3.58	-0.26	-4.06	11.84	9.39
SC_6^2	-3.48	-0.20	-3.85	11.04	8.70
SC_2^2	-3.06	-0.07	-3.76	9.20	6.69(4.58)

TABLE B.3: *Relative stabilities (in kcal/mol with respect to N1 structure) at different levels of theory for the singly proton transferred conformations of anionic tyrosine. The proton was transferred from the hydroxyl group on to the ring.*

	Vertical Detachment Energies,(eV)				
	aug-cc-pVDZ				
Structure	B3LYP	MP2	Structure	B3LYP	MP2
$D_8^{2,3}$	3.81	4.25	SH_1^3	1.43	1.07
$D_2^{2,3}$	3.65	3.97	SH_2^3	1.47	1.12
$D_9^{2,3}$	3.76	4.23	SH_5^3	1.63	1.49
$D_{10}^{2,3}$	3.92	4.25	SH_4^3	1.36	0.99
$D_1^{2,3}$	3.83	3.96	SH_6^3	1.66	1.34
$D_{11}^{2,3}$	3.93	4.41	SH_{11}^3	1.32	0.70
$D_6^{2,3}$	3.56	3.72	SH_3^3	1.64	0.99
$D_7^{2,3}$	3.47	3.55	SH_8^3	1.62	1.48
$D_{12}^{2,3}$	4.09	4.49	SH_7^3	1.70	1.57
$D_3^{2,3}$	3.81	3.98	SH_{10}^3	1.65	1.25
$D_{13}^{2,3}$	4.15	4.55	SH_9^3	1.66	1.33
$D_4^{2,3}$	3.87	4.09	SH_{12}^3	1.73	1.63
$D_5^{2,3}$	3.78	4.19	SH_{13}^3	1.23	0.87

TABLE B.4: *Electronic binding energies (in eV) of **SH**m and **D**m structures at B3LYP/MP2/aug-cc-pVDZ levels of theory.*

VDE, (eV) aug-cc-pVDZ		
Anions	B3LYP	MP2
SC_5^2	3.02	3.40
SC_1^2	3.23	3.43
SC_7^2	3.06	3.43
SC_3^2	3.03	3.26
SC_4^2	3.15	3.56
SC_8^2	3.27	3.57
SC_6^2	3.44	3.84
SC_2^2	3.25	3.46

TABLE B.5: *Electronic binding energies (in eV) of **SH**m structures at B3LYP/MP2/aug-cc-pVDZ levels of theory.*

	Relative Energies *w.r.t* Neutral(N1), (kcal/mol)								
	6-31++G**	aug-cc-pVDZ							
	E^{B3LYP}	E^{B3LYP}	B3LYP $+ E_0^{vib}$	E^{MP2}	$E^{MP2} + E_0^{vib}$	$E^{CCSD(T)}$	$E^{CCSD(T)}$ E_0^{vib}	VDE (MP2)	μ_{ne}^{M}
$D_1^{2,3}$	-15.75	-14.17	-16.95	-5.77	-6.96	-11.69	-12.84	3.96(3.80)	9.
$D_2^{2,3}$	-16.06	-14.53	-17.38	-4.83	-6.02	-11.01	-12.16	3.97(3.76)	9.
$D_3^{2,3}$	-14.78	-13.46	-16.37	-3.55	-4.83	-9.74	-10.98	3.98(3.76)	10
$D_1^{3,5}$	-3.38	-2.08	-5.83	5.02	2.16	-1.15	-3.99	4.04(3.90)	11
$D_2^{3,5}$	-2.79	-1.58	-5.44	5.44	2.58	-0.79	-3.62	3.78(3.65)	9.
$D_3^{3,5}$	-1.44	0.48	-4.52	6.76	3.83	0.54	-2.37	3.79(3.67)	11
$D_1^{N,3}$	-0.19	0.44	-3.05	9.86	8.04	8.63	6.81	1.80(1.73)	7.
$D_2^{N,3}$	0.52	1.09	-2.68	10.60	8.36	8.33	6.09	1.63	7.
SH_1^3	-7.38	-7.01	-10.15	3.87	2.00	-0.04	-1.91	1.07(1.02)	2.
SH_2^3	-7.27	-6.90	-10.08	4.13	2.07	-0.04	-2.10	1.12(1.07)	0.
SH_3^3	-4.12	-4.26	-7.86	5.79	2.17	1.34	-2.28	0.99(1.23)	3.
SC_1^2	-1.09	-1.05	-4.54	8.02	5.86	3.46	1.33	3.43(3.34)	14
SC_2^2	0.04	-0.07	-3.76	9.20	6.69	4.58	2.10	3.46	15
SC_3^2	–0.76	-0.61	-4.16	9.98	7.73	5.13	2.52	3.26	13

TABLE B.6: *Relative stabilities (in kcal/mol) of $D^{2,3}, D^{3,5}, D^{N,3}, SH^3$ and SC^2 structures with respect to $N1$ structure calculated at different levels of theory. The CCSD(T) energies were determined at MP2/ADZ optimal geometries and then corrected for MP2 E_0^{vib}. The MP2 (CCSD(T) in brackets) VDEs are in eV and the dipole moments of the neutrals at anion geometries are in Debyes.*

DP_1^H :1.45

VDE : 2.94

DP_2^H :3.43

VDE : 3.10

DP_3^H :3.58

VDE : 2.84

DP_1^C : 0.00

VDE : 5.47

DP_2^C : 0.69

VDE :

DP_3^C : 3.21

VDE : 3.92
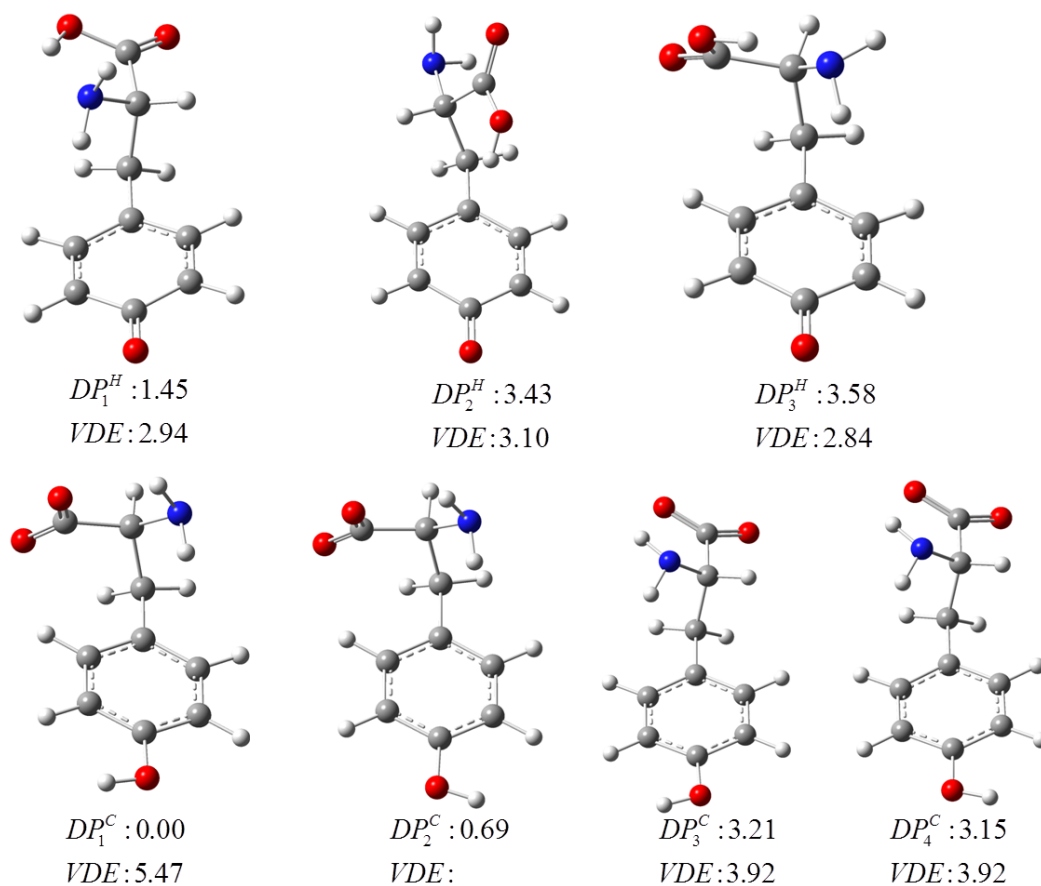
DP_4^C : 3.15

VDE : 3.92

FIGURE B.5: *Low-lying deprotonated anions of tyrosine optimised at the MP2/ADZ level of theory. The relative stabilities (in kcal/mol) are with respect to DP_1^C structure and the VDEs are in eV. The remaining quantities are the same as in Fig. 1. (Zibo, is there experimental information on the proton affinity of (Tyr-H)-? It would be equivalent to the deprotonation enthalpy of Tyr.)*

Appendix C

Supporting Information to Chapter 11

C.1 Experimental methods

C.1.1 Synthesis of Acetoacetic Acid

Our synthesis followed the protocol described in Ref.213. Sodium hydroxide (0.16 mol) was added to aqueous ethyl acetoacetate (0.15 mol) on ice. The reaction was stirred and allowed to react overnight. The resulting solution was saturated with ammonium sulfate. Sulfuric acid was then used to acidify the solution. The solution was extracted three times with diethyl ether (200 ml) and dried over magnesium sulfate. After removing the ether using a rotary evaporator and drying using a rotary vacuum pump, a yellowish gel of acetoacetic acid was obtained. Since batches of acetoacetic acid were readily subject to decomposition, care was taken to ensure that the solutions did not become warmer than 30 °C, and they were utilized in photoelectron and electron energy loss spectroscopy experiments as quickly as possible.

C.1.2 Photoelectron Spectroscopy

In the first approach, the anions of acetoacetic acid were generated with a nozzle-ion source and their photoelectron spectra were measured with a continuous anion photoelectron apparatus.[214] In the second approach, the anions were generated with a photo-induced electron emission, pulsed source and their photoelectron spectra were measured with a pulsed anion photoelectron apparatus.[190] In both instruments, anion photoelectron spectroscopy was conducted by crossing a mass-selected negative ion beam with a

fixed energy photon beam and energy analysing the resultant photodetached electrons. This technique is governed by the energy conserving relationship, $hv = EKE + EBE$, where hv is the photon energy, EKE is the measured electron kinetic energy, and EBE is the electron binding energy.

C.1.2.1 Continuous Anion Photoelectron Spectrometer

In this type of anion photoelectron spectrometer, the ion source and all of the other components operate continuously.[214] The ion source was a biased (–500 V), supersonic expansion nozzle-ion source, in which acetoacetic acid samples were placed inside its stagnation chamber. There, due to the warming of the source by the adjacent hot filament, some of the sample evaporated and was expanded through a 25 μm diameter nozzle orifice into $\sim 10^{-4}$ torr vacuum by argon gas which was maintained at a pressure of 1–2 atm. in the stagnation chamber. Anions were formed by injecting low energy electrons from a hot, even more negatively-biased, thoriated iridium filament into the expanding jet, where a weak external magnetic field helped to form a micro-plasma. The nascent anions were then extracted from this region and mass-selected by a 90° magnetic sector, mass spectrometer with a mass resolution of $\sim$400. Upon mass-selection, the beam of acetoacetic acid parent anions was crossed with an argon ion laser beam (operated intra-cavity), where electrons were photodetached. These were then energy-analysed by a hemispherical electron energy analyser having a resolution of 30 meV. Photoelectron spectra were recorded with 2.54 eV photons and calibrated against the well-known photoelectron spectrum of O^-.[215]

C.1.2.2 Pulsed Anion Photoelectron Spectrometer

In this type of anion photoelectron spectrometer, the ion source and all of the other components operate in a pulsed mode.[190] Anions were generated by the interaction of laser-generated photoelectrons with a pulsed jet of helium gas containing a small amount of acetoacetic acid vapor. The photoelectrons were produced by pulsed laser irradiation (Nd:YAG laser operating at 2.33 eV/photon) of a rotating, translating copper rod which was mounted inside a grounded housing having a laser beam entrance port, a pulsed gas valve, and an exit nozzle. A small amount of acetoacetic acid sample was placed inside the pulsed gas valve and together with 4 atm. of helium gas, its vapour was expanded in synchronization with the laser pulses. The resulting anions were entrained in the ensuing jet which itself was directed through a skimmer and a subsequent differential pumping chamber into the Wiley-McLaren extraction plates of a linear, time of flight mass spectrometer (mass resolution 2,000). After mass selection by a mass gate

and deceleration via a momentum decelerator, the anions of interest were irradiated by a second pulsed laser beam (Nd:YAG laser operating at 3.49 eV/photon) which photodetached electrons from them. The photodetached electrons were then energy analysed by a magnetic bottle, electron energy analyser with a resolution of 35 meV at $EKE = 1$ eV. The photoelectron spectrum was calibrated against the well-known atomic lines of Cu^-.[216]

C.1.3 Electron Energy Loss Spectroscopy

The measurements were performed using a spectrometer with hemispherical analysers.[217, 218]. The energy resolution was about 14 meV in the energy-loss mode, corresponding to about 10 meV in the incident electron beam, at a beam current of around 250 pA. The energy of the incident beam was calibrated on the 19.365 eV ^2S resonance in helium[219] and is accurate to within $\pm$ 10 meV. The analyser response function was determined by elastic scattering of helium. The sample was introduced through a 0.25 mm diameter effusive nozzle kept at 35 °C. Absolute values of the cross sections were determined by comparison with the theoretical helium elastic cross section[220], as described in Ref .218. Absolute differential elastic and vibrational excitation cross sections have been measured at 135° from threshold to 5 eV. The scattering angle of 135° has been chosen to emphasize the resonant phenomena and reduce the amount of direct dipole excitation which peaks in the forward direction.

	ADZ					
Str.	E_{elect}^{MP2}	E_{elec}^{MP2} $+E_{0,vib}^{MP2}$	E_{elec}^{CCSD}	$E_{elec}^{CCSD(T)}$	$E_{elec}^{CCSD(T)}$ $+E_{0,vib}^{MP2}$	μ^{CCSD}
N1	0.00	0.00	0.00	0.00	0.00	5.41
N2	0.03	-0.17	-0.04	0.23	0.03	1.52
N3	0.04	-0.29	-0.41	0.01	-0.32	3.24
N4	0.10	0.04	—	—	—	—
N5	0.24	-0.10	-0.13	0.28	-0.06	3.52
N6	0.43	0.20	0.36	0.60	0.37	3.23
N7	0.74	0.54	0.55	0.88	0.68	3.58

TABLE C.1: The Relative Electronic Energies (kcal/mol) of canonical conformers of neutral acetoacetic acid calculated at different levels of theory followed by their dipole moments calculated at MP2/ADZ level of theory. Single point CCSD(T) energy calculation on CCSD optimal geometries

Neutral@Geometry	Dipole moments, μ		
	HF	MP2	CCSD
N	5.63	5.42	5.41
$N^{(MP2)}$	6.15	5.42	5.54
$N@G_M$	6.01	5.31	5.41
DBA	5.71	5.59	5.57
$DBA^{(MP2)}$	6.32	5.59	5.70
$N@G_{M_{dbs}^-}$	6.18	5.47	5.57
VBA	11.09	9.69	10.00
$VBA^{(MP2)}$	10.72	9.69	9.85
$N@G_{M_{VB}^-}$	10.87	9.84	10.00

TABLE C.2: The dipole moments of the neutral (N) at HF, MP2 and CCSD level of theory at the three representative geometries; (i) optimal geometry of N, (G_M), (ii) optimal geometry of DB, $(G_{M_{dbs}^-})$, (iii) optimal geometry of VB, $(G_{M_{VB}^-})$. All the geometries were optimized at the CCSD/ADZ+DF level.

FF-comp	Amber		
	N1	N3	N5
Bond Stretching	1.3976	0.8843	0.954
Angle Bending	3.5055	1.8546	1.6753
Improper Torsion	0.0419	0.0074	0.0009
Torsional Angle	1.955	0.7796	1.4466
Van der Waals	2.2484	1.1863	1.769
Charge-Charge	-18.9634	-22.9145	-18.3529
Total	-18.9634	-18.202	-12.5068

TABLE C.3: Contributions from components of the force fields to **N1**, **N3** and **N5** energies calculated using Amber, UFF and Dreiding force fields implemented on Gaussian suite of programmes.

Method	EBE		
	G_M	$G_{M_{dbs}^-}$	$G_{M_{VB}^-}$
$KT(EA)$	21.77	24.22	113.47
ΔE_{bind}^{SCF}	2.30	2.68	14.21
$\Delta E_{bind}^{MP2-disp}$	24.32	26.69	172.36
$\Delta E_{bind}^{MP2-no-disp}$	-15.57	-17.27	-82.32
ΔE_{bind}^{CCSD}	19.52	20.27	-20.87
ΔE_{bind}^{T}	-0.15	-0.04	0.19
Total	52.19	56.55	197.04

TABLE C.4: Incremental electronic binding energies of the dipole bound anionic state at G_M, $G_{M_{dbs}^-}$ and $G_{M_{VB}^-}$ geometries using the CCSD/ADZ+DF optimal geometries.

	N1 Neutral		
C	2.516605	-0.648014	-0.164126
C	1.211565	0.093479	0.037571
O	1.153675	1.319670	0.001508
C	-0.021125	-0.764763	0.316302
C	-1.404072	-0.176482	-0.005698
O	-1.514892	1.165542	-0.016913
O	-2.354505	-0.900934	-0.197982
H	-0.618422	1.551878	0.063044
H	0.056655	-1.735423	-0.197071
H	-0.012395	-0.985494	1.401474
H	2.664376	-1.383781	0.645036
H	2.464521	-1.209737	-1.113638
H	3.353198	0.063007	-0.196045

	DB anion				**VB** anion		
C	2.528004	-0.637843	-0.134706	C	2.511211	-0.655154	-0.074065
C	1.212397	0.090863	0.024596	C	1.147644	-0.034892	-0.222404
O	1.141073	1.318560	-0.006574	O	1.132583	1.317090	0.021988
C	-0.019652	-0.780706	0.265877	C	-0.065077	-0.805140	0.256521
C	-1.405710	-0.177533	-0.005056	C	-1.433281	-0.102430	0.001885
O	-1.503532	1.166336	-0.000845	O	-1.405513	1.180419	0.029601
O	-2.373320	-0.887120	-0.175374	O	-2.430087	-0.838424	-0.166454
H	-0.596109	1.536352	0.061769	H	0.139806	1.518589	0.058054
H	0.057258	-1.721840	-0.300587	H	-0.089964	-1.811827	-0.193167
H	0.003643	-1.067548	1.336289	H	0.002606	-0.947620	1.364299
H	2.662289	-1.361688	0.689045	H	2.818750	-0.697998	0.998761
H	2.509216	-1.215975	-1.075889	H	2.519055	-1.687175	-0.467913
H	3.359697	0.079805	-0.152549	H	3.270896	-0.060956	-0.612740

TABLE C.5: Cartesian coordinates of **N1** neutral, **DB** anion and **VB** anion. Optimal CCSD/ADZ+DF geometries.

Appendix D

Supporting Information to Chapter 7

		Structure	A	
	C	-0.100004	1.452790	0.000000
	O	2.142895	2.617973	0.000000
	O	-2.142895	2.522237	0.000000
	C	0.100004	-1.452790	0.000000
	O	-2.142895	-2.617973	0.000000
	O	2.142895	-2.522237	0.000000
	H	3.420243	1.277323	0.000000
	H	-3.420243	-1.277323	0.000000

	B				C		
C	-0.103498	-1.519182	0.000000	C	-0.000210	-1.452478	0.000000
O	2.234239	-2.537434	0.000000	C	0.000000	1.452476	0.000000
O	-2.041389	-2.739748	0.000000	O	-1.895611	2.753111	0.000000
C	0.000000	1.391823	0.000000	O	-2.399284	-2.326089	0.000000
O	-2.301417	2.444545	0.000000	O	1.895632	-2.752993	0.000000
O	2.013326	2.535590	0.000000	O	2.399371	2.325981	0.000000
H	-2.048273	4.269650	0.000000	H	2.266801	4.162590	0.000000
H	3.431199	-1.129123	0.000000	H	-2.266404	-4.162654	0.000000

TABLE D.1: Cartesian coordinates of **A**, **B** and **C**. Optimal MP2/ADZ geometries.

Structure	AA		
C	3.012217	4.265209	0.000000
O	4.409611	6.109001	0.000000
C	4.183783	1.594077	0.000000
O	6.701141	1.641044	0.000000
O	0.508705	4.384606	0.000000
O	2.934772	-0.347392	0.000000
O	-2.934772	0.347392	0.000000
C	-4.183783	-1.594077	0.000000
O	-6.701141	-1.641044	0.000000
C	-3.012217	-4.265209	0.000000
O	-0.508705	-4.384606	0.000000
O	-4.409611	-6.109001	0.000000
H	-0.239449	2.684034	0.000000
H	7.163701	3.439756	0.000000
H	0.239449	-2.684034	0.000000
H	-7.163701	-3.439756	0.000000

AB				BB		
C	2.657441	-2.225376	0.000000	C	0.179275	-3.659410	0.000000
O	0.687855	-3.751595	0.000000	O	2.444324	-2.640101	0.000000
O	2.602913	0.110776	0.000000	O	-1.866311	-2.514930	0.000000
O	-3.441648	-0.507144	0.000000	C	0.213180	-6.573652	0.000000
C	-4.551904	1.519877	0.000000	O	2.135300	-7.816563	0.000000
O	-7.056230	1.727923	0.000000	O	-2.135300	-7.563960	0.000000
C	-3.218087	4.117365	0.000000	O	1.866311	2.514930	0.000000
O	-0.716116	4.118249	0.000000	C	-0.179275	3.659410	0.000000
O	-4.523972	6.028833	0.000000	O	-2.444324	2.640101	0.000000
H	-0.851329	-2.701018	0.000000	C	-0.213180	6.573652	0.000000
H	0.000000	2.393256	0.000000	O	2.135300	7.563960	0.000000
H	-7.399291	3.555029	0.000000	O	-2.135300	7.816563	0.000000
C	5.210077	-3.630099	0.000000	H	-3.325887	-6.153075	0.000000
O	5.407436	-5.910937	0.000000	H	-2.266801	0.767431	0.000000
O	7.186447	-2.019781	0.000000	H	2.266801	-0.767431	0.000000
H	6.491956	-0.308456	0.000000	H	3.325887	6.153075	0.000000

TABLE D.2: Cartesian coordinates of **AA**, **AB** and **BB**. Counterpoise corrected MP2/ADZ geometries.

Mode	A	B	C
1	123.607	87.151	21.1894
2	258.659	263.778	261.9709
3	403.787	419.802	424.8354
4	461.287	440.474	427.4107
5	559.702	544.231	528.1462
6	656.005	641.773	626.089
7	691.564	659.728	640.5536
8	699.084	669.801	674.6756
9	821.023	803.459	789.6835
10	825.716	824.176	822.8306
11	1192.281	1156.445	1130.5523
12	1224.933	1208.645	1208.697
13	1331.599	1333.311	1330.3802
14	1449.839	1443.178	1425.1552
15	1780.342	1756.03	1775.3356
16	1808.369	1811.712	1789.8683
17	3635.964	3662.35	3735.5508
18	3639.927	3727.354	3736.2524

TABLE D.3: Harmonic MP2/ADZ frequencies (cm^{-1}) of the monomers **A, B** and **C**.

Mode	AA	AB	BB	Mode	AA	AB	BB
1	17.0672	27.8245	30.5123	22	775.1238	818.8559	828.8528
2	29.9133	31.8696	41.9236	23	829.6556	824.762	828.9879
3	69.8338	78.6683	120.1046	24	835.3086	836.6523	831.2621
4	93.3522	87.9664	93.2379	25	835.8026	841.7894	904.1634
5	95.466	121.7411	95.4087	26	835.8991	858.3734	941.2647
6	106.4619	108.4781	125.1741	27	1201.1566	1172.8163	1174.2605
7	114.5557	88.7529	80.139	28	1212.5695	1214.1713	1175.1683
8	132.3666	127.8147	103.3316	29	1227.3352	1239.996	1279.5879
9	260.1689	272.3119	291.0229	30	1231.8449	1255.6911	1285.617
10	264.3439	290.7507	309.9501	31	1317.2535	1325.2103	1358.3455
11	409.0231	416.4866	429.2537	32	1331.8589	1349.1454	1377.1289
12	414.2823	435.3743	447.2087	33	1447.2325	1453.6441	1493.6178
13	460.0051	462.5256	468.9468	34	1462.6132	1487.1377	1507.1716
14	463.7264	470.1698	482.7704	35	1768.3116	1730.7687	1695.4713
15	556.0654	554.2937	552.6827	36	1772.2614	1766.9121	1741.454
16	562.1927	564.1807	569.7961	37	1801.4773	1805.6334	1811.6881
17	665.2621	656.7428	650.0362	38	1813.4255	1812.0823	1812.0588
18	671.7102	658.31	651.0011	39	3566.1234	3449.8377	3275.6488
19	741.4182	677.2045	663.4845	40	3572.3094	3520.1854	3356.4603
20	741.8903	757.61	673.9942	41	3588.9855	3573.8454	3681.0172
21	746.1642	794.2069	824.1387	42	3591.377	3673.8063	3680.9255

TABLE D.4: Harmonic, MP2/ADZ, counterpoise-corrected frequencies (cm^{-1}) of the dimers **AA, AB** and **BB**.

Structure		E_{SCF}	E_{MP2}	E_{MP2}^{corr}
	ADZ	-376.432612205	-377.48124344130	-1.0486312363
A	ATZ	-376.521206245	-377.79314019141	-1.2719339464
	AQZ	-376.545579436	-377.8986605574	-1.3530811214
	ADZ	-376.429077345	-377.47706952308	-1.0479921781
B	ATZ	-376.517756536	-377.78862033752	-1.2708638015
	AQZ	-376.54210275	-377.89405856305	-1.351955813

TABLE D.5: Electronic energies (in Hartrees) of the oxalic acid monomers **A** and **B** calculated at the SCF and MP2 levels of theory with ANZ basis sets for the optimal MP2/ADZ geometry.

	A	B
E_{HF}^{∞}	-376.5548295851	-376.5513163473
$E_{corr-MP2}^{\infty}$	-1.412296627	-1.411131065
ΔE_{CCSD}	-0.0074814587	-0.0077695769
ΔE_T	-0.0416671	-0.0415113
E_{Total}	-378.0162747713	-378.011728289
E_0^{vib}	0.049127	0.048874
$E_{Total} + E_0^{vib}$	-377.9671477713	-377.962854289
R.E	0.00 meV	116.8325231

TABLE D.6: Extrapolated energies (in Hartrees) to the complete basis set limit of the monomers **A** and **B** for the optimal MP2/ADZ geometries. The relative energies are in meV.

Structure		E_{SCF}	E_{MP2}	E_{MP2}^{corr}
	ADZ	-752.876255288921	-754.9757994	-2.0995440790
AA	ATZ	-753.054121098974	-755.6009006	-2.5467795233
	AQZ	-753.102946837941	-755.8124207	-2.7094738848
	ADZ	-752.875133842846	-754.9747310	-2.0995971817
AB	ATZ	-753.053041223151	-755.5997376	-2.5466963368
	AQZ	-753.10185603891	-755.8112728	-2.7094167457
	ADZ	-752.8732276256	-754.9735851	-2.1003574961
BB	ATZ	-753.051060455934	-755.5984728	-2.5474123343
	AQZ	-753.099838208325	-755.8099902	-2.7101520230

TABLE D.7: Counterpoise-corrected single point electronic energies (in Hartrees) of the oxalic acid **AA**, **AB** and **BB** dimers calculated at the SCF and MP2 levels with the ANZ basis sets for the optimal MP2/ADZ geometries.

	AA	AB	BB
E_{HF}^{∞}	-753.1214213534	-753.1203147801	-753.1182742837
$E_{corr-MP2}^{\infty}$	-2.8281967973	-2.8281586657	-2.8289080121
ΔE_{CCSD}	-0.0141689065	-0.0143183211	-0.0143302645
ΔE_T	-0.0841281271	-0.0842115488	-0.0844162572
E_{Total}	-756.0479151843	-756.0470033158	-756.0459288175
E_0^{vib}	0.099404	0.099544	0.099657
$E_{Total} + E_0^{vib}$	-755.9485111843	-755.9474593158	-755.9462718175

TABLE D.8: Extrapolated counterpoise-corrected SCF and MP2 electronic energies (in Hartrees) of the oxalic acid **AA**, **AB** and **BB** dimers. Further corrections from CCSD(T) and zero-point energies also included.

E	E(AB) – E(AA)	E(BB) – E(AA)
E_{HF}^{∞}	30.11	85.64
$E_{corr-MP2}^{\infty}$	1.04	-19.35
ΔE_{CCSD}	-4.07	-4.39
ΔE_T	-2.27	-7.84
E_{Total}	**24.81**	**54.06**
ΔE_0^{vib}	3.81	6.88
$E_{Total} + \Delta E_0^{vib}$	28.62	60.94

TABLE D.9: Relative extrapolated counterpoise-corrected SCF and MP2 electronic energies (in meV) of the oxalic acid **AA**, **AB** and **BB** dimers. The energies are with respect to the **AA** dimer. Further corrections from CCSD(T) and zero-point energies are also included.

Structure		E_{SCF}	E_{MP2}	E_{MP2}^{corr}
A@AA	ADZ	-376.431840748	-377.47990769899	-1.048066951
	ATZ	-376.520518407	-377.79180033328	-1.2712819263
	AQZ	-376.544889758	-377.89733141847	-1.3524416605
A@AB	ADZ	-376.430685532	-377.47893225385	-1.0482467218
	ATZ	-376.519371689	-377.79078875081	-1.2714170618
	AQZ	-376.543731968	-377.89630462089	-1.3525726529
B@AB	ADZ	-376.428357285	-377.47642707378	-1.048069788
	ATZ	-376.517034984	-377.78799220596	-1.270957222
	AQZ	-376.541364528	-377.89342430268	-1.3520597747
B@BB	ADZ	-376.427338259	-377.4758231904	-1.0484849314
	ATZ	-376.515983196	-377.78730210761	-1.2713189116
	AQZ	-376.540291362	-377.89269807756	-1.3524067156

TABLE D.10: Electronic energies (in Hartrees) of the distored oxalic acid monomers calculated at the SCF and MP2 levels with the monomer-centered basis ANZ basis sets. The monomer geometries are extracted from the optimal MP2/ADZ dimer geometries.

	A@AA	**A@AB**	**B@AB**	**B@BB**
E_{HF}^{∞}	-376.5541262194	-376.5529572318	-376.5505633422	-376.5494756422
$E_{corr-MP2}^{\infty}$	-1.411666331	-1.411794300	-1.411242719	-1.411578897
ΔE_{CCSD}	-0.007555301	-0.0075170462	-0.0076845262	-0.0076134096
ΔE_T	-0.0415774	-0.0416014	-0.0415261	-0.0415845
E_{Total}	-378.0149252518	-378.0138699784	-378.0110166869	-378.0102524487

TABLE D.11: Extrapolated electronic energies (in Hartrees) of the distorted **A** and **B** monomers. The monomer geometries are extracted from the optimal MP2/ADZ dimer geometries.

	Structure of the total stabilisation energy $E^{stab}_{X_m Y_n}$					
	AA		**AB**		**BB**	
	Method **1**	Method **2**	Method **1**	Method **2**	Method **1**	Method **2**
E^∞_{HF}	-320.1	-321.1	-290.0	-291.3	-234.4	-236.2
$E^\infty_{corr-MP2}$	-98.1	-98.0	-97.0	-97.0	-117.4	-117.4
ΔE_{CCSD}	21.6	21.6	17.5	17.5	17.2	17.2
ΔE_T	-21.6	-21.6	-23.8	-23.8	-29.4	-29.4
E^{stab}_{Total}	-418.1	-419.1	-393.3	-394.6	-364.1	-365.8

TABLE D.12: Comparison of two extrapolaton methods for the SCF and corr-MP2 components of the stabilisation energy. Method **1**: the stabilisation energy obtained from the extrapolated monomer and dimer energies. Method **2**: the stabilisation energies were calculated at the SCF and MP2 levels for the sequence of aug-cc-pVNZ ($N = D - Q$) basis sets. Next, these stabilisation eneriges were extrapolated to the basis set limit. Higher-order corrections obtained at the CCSD(T) level of theory. All energies in meV.

Appendix E

Supporting Information to Chapter 9

E.1   Experimental Details

The experimental spectrum was done in collaboration with the experimental group led by Kit Bowen. (Angela Buonaugurio, Jacob Graham, Allyson Buytendyk, Kit Bowen)

Negative ion photoelectron spectroscopy is conducted by crossing a mass-selected beam of parent negative ions with a fixed-frequency photon beam and energy-analysing the resultant photo-detached electrons. This process is governed by the energy conserving relationship $hv = EKE + EBE$, where hv is the photon energy, EKE is the measured electron kinetic energy, and EBE is the electron binding energy. OA anions were generated in a biased (-500V) supersonic expansion nozzle-ion source, in which the OA sample was heated between 80-100 °C and co-expanded with approximately several atmospheres of argon gas through a 10 μm diameter nozzle. Low energy electrons were injected directly into the expanding jet by a hot and even more negatively biased thoriated iridium filament, in the presence of a weak external magnetic field where the microsplasma was formed. The anions were then extracted and transported by an ion optics series through a 90° magnetic sector, mass spectrometer with a typical mass resolution of ∼400. The mass-selected OA anion beam was then crossed with an intracavity run argon ion laser beam. The resultant photo-detached electrons were energy-analyzed in a hemispherical electron energy analyser with a resolution of ∼30 meV. The photoelectron spectrum reported here was recorded with 2.54 eV photons (488 nm) for the monomer, and 4.66 eV (266 nm) and 6.42 eV (193 nm) photons for the dimer and trimer. The spectra were calibrated against the well-known photoelectron spectrum of the O$^-$ anion.[215]

Str.		Neutral				Neutral		
1	C	0.000000	0.000000	0.000000	C	0.000000	0.000000	0.000000
	O	0.000000	0.000000	1.347237	O	0.000000	0.000000	1.410901
	O	0.982973	0.000000	-0.701148	O	1.111248	0.000000	-0.598095
	C	-1.445226	0.000000	-0.538566	C	-1.315462	0.000000	-0.564242
	O	-1.445226	0.000000	-1.885803	O	-1.315462	0.000000	-1.975143
	O	-2.428199	0.000000	0.162582	O	-2.426710	0.000000	0.033853
	H	-2.375931	0.000000	-2.160378	H	-2.263433	0.000000	-2.164223
	H	0.930705	0.000000	1.621812	H	0.947971	0.000000	1.599980
2	C	0.000000	0.000000	0.000000	C	0.000000	0.000000	0.000000
	C	0.000000	0.000000	1.545546	C	0.000000	0.000000	1.435223
	O	1.038406	0.000000	2.178347	O	1.054823	0.000000	2.160888
	O	1.248928	0.000000	-0.502446	O	1.307705	0.000000	-0.513327
	O	-0.999174	0.000000	-0.669697	O	-0.968228	0.000000	-0.787615
	O	-1.230807	0.000000	2.065282	O	-1.253784	0.000000	2.060719
	H	-1.134011	0.000000	3.031641	H	-1.010901	0.000000	2.996169
	H	1.868349	0.000000	0.248091	H	1.858923	0.000000	0.289720
3	C	0.000000	0.000000	0.000000	C	0.000000	0.000000	0.000000
	C	0.000000	0.000000	1.545275	C	0.000000	0.000000	1.433856
	O	1.035141	0.000000	2.173612	O	1.029863	0.000000	2.167474
	O	1.220308	0.000000	-0.539951	O	1.275118	0.000000	-0.570281
	O	-1.035141	0.000000	-0.628337	O	-1.029863	0.000000	-0.733618
	O	-1.220308	0.000000	2.085226	O	-1.275118	0.000000	2.004137
	H	1.858435	0.000000	0.196827	H	1.859498	0.000000	0.208324
	H	-1.858435	0.000000	1.348448	H	-1.859498	0.000000	1.225532
4	C	0.000000	0.000000	0.000000	C	0.000000	0.000000	0.000000
	O	0.000000	0.000000	1.279197	O	0.000000	0.000000	1.367353
	O	1.091336	0.000000	-0.667331	O	1.229188	0.000000	-0.598958
	C	-1.355194	0.000000	-0.759816	C	-1.241769	0.000000	-0.776275
	O	-2.311569	0.000000	0.039809	O	-2.312542	0.000000	-0.078220
	O	-1.171901	0.000000	-1.992883	O	-1.083271	0.000000	-2.044626
	H	-0.978854	0.000000	1.475342	H	-0.965073	0.000000	1.53023
	H	0.748027	0.000000	-1.604757	H	0.952871	0.000000	-1.537864
5		**N/A**			C	0.000000	0.000000	0.000000
					O	0.000000	0.000000	1.376150
					O	1.255539	0.000000	-0.563393
					C	-1.039949	0.856719	-0.673194
					O	-2.030448	1.163356	0.058656
					O	-0.777753	1.163360	-1.876500
					H	-0.880453	0.393846	1.547854
					H	1.051739	0.393847	-1.436974

TABLE E.1: Atomic positions in direct coordinates.

Str.			Neutral				Neutral	
TS1	C	0.000000	0.000000	0.000000	C	0.000000	0.000000	0.000000
	O	0.000000	0.000000	1.373996	O	0.000000	0.000000	1.408050
	O	0.985759	0.000000	-0.686448	O	1.114021	0.000000	-0.587928
	C	-1.419965	-0.000149	-0.591831	C	-1.318245	-0.015025	-0.568172
	O	-1.523351	0.824962	-1.649875	O	-1.324339	0.069034	-1.995016
	O	-2.300484	-0.708814	-0.160060	O	-2.410541	-0.089681	0.051494
	H	-2.430969	0.741564	-1.985418	H	-1.353796	-0.837411	-2.324055
	H	0.042112	-0.904733	1.712862	H	0.947453	0.006284	1.599989
TS2	C	0.000000	0.000000	0.000000	C	0.000000	0.000000	0.000000
	C	0.000000	0.000000	1.549845	C	0.000000	0.000000	1.443090
	O	1.043819	0.000000	2.159779	O	1.073719	0.000000	2.136488
	O	1.233182	0.142139	-0.521457	O	1.301271	0.060228	-0.509984
	O	-0.997214	-0.141800	-0.659399	O	-0.961679	-0.056090	-0.788631
	O	-1.232208	0.060222	2.125433	O	-1.262529	-0.005839	2.082589
	H	-1.560480	-0.822281	2.344851	H	-1.484004	-0.928737	2.258687
	H	1.862902	0.197293	0.217327	H	1.844836	0.076291	0.300128
TS3	C	0.000000	0.000000	0.000000	C	0.000000	0.000000	0.000000
	O	0.000000	0.000000	1.552249	O	0.000000	0.000000	1.385725
	O	1.160651	0.000000	2.044199	O	1.145824	0.000000	-0.675211
	O	1.218865	0.000000	-0.356721	C	-1.069921	0.685586	-0.747712
	O	-1.104011	0.000000	-0.545348	O	-2.183348	0.983202	-0.259642
	O	-1.083242	0.000000	2.253218	O	-0.608299	0.913655	-1.965371
	H	1.651146	0.000000	1.063933	H	-0.904350	0.283476	1.599776
	H	-1.808702	0.000000	1.589434	H	0.527920	0.476839	-1.618668

TABLE E.2: Atomic positions in direct coordinates.

Peak	Assignment	Position (cm^{-1})	Rel. FC factor	Signal
1	1_0^0	5612.612	1	1
2	$1_0^1 1_1^0$	5545.326	0.950	0.062
3	$1_0^1 12_0^1 1_1^0$	6807.456	0.595	0.039
4	$1_0^1 12_0^1 14_0^1 1_1^0$	8297.722	0.977	0.063
5	$1_0^1 12_0^1 14_0^2 1_1^0$	9787.988	0.810	0.053
6	$1_0^1 12_0^1 14_0^3 1_1^0$	11278.253	0.453	0.029
7	$1_0^1 12_0^1 16_0^1 1_1^0$	8691.909	0.542	0.035
8	$1_0^2 14_0^1$	7348.336	0.038	0.038
9	$1_0^2 14_0^2$	8838.602	0.031	0.031
10	$1_0^1 14_0^1 1_1^0$	7035.592	1.537	0.100
11	$1_0^1 14_0^2 1_1^0$	8525.858	1.257	0.082
12	$1_0^1 14_0^3 1_1^0$	10016.124	0.693	0.045
13	$1_0^2 14_0^1 16_0^1$	9232.789	0.033	0.033
14	$1_0^1 14_0^1 16_0^1 1_1^0$	8920.044	1.351	0.088
15	$1_0^1 14_0^1 16_0^1 1_1^0$	10804.497	0.669	0.043
16	$1_0^1 14_0^2 16_0^1 1_1^0$	10410.310	1.031	0.067
17	$1_0^1 14_0^2 16_0^2 1_1^0$	12294.762	0.480	0.031
18	$1_0^1 14_0^3 16_0^1 1_1^0$	11900.576	0.530	0.034
19	$1_0^1 16_0^1 1_1^0$	7429.779	0.893	0.058
20	$1_0^1 16_0^2 1_1^0$	9314.231	0.471	0.031
21	$2_0^1 14_0^1 2_1^0$	7117.246	1.607	0.041
22	$2_0^1 14_0^2 2_1^0$	8607.512	1.315	0.034
23	$2_0^1 14_0^1 16_0^1 2_1^0$	9001.698	1.412	0.036
24	$2_0^1 14_0^2 16_0^1 2_1^0$	10491.964	1.078	0.027
25	5_0^1	6185.324	0.031	0.031
26	$5_0^1 12_0^1 14_0^1$	8937.719	0.033	0.033
27	$5_0^1 14_0^1$	7675.590	0.049	0.049
28	$5_0^1 14_0^2$	9165.855	0.039	0.039
29	$5_0^1 14_0^1 16_0^1$	9560.042	0.052	0.052
30	$5_0^1 14_0^1 16_0^2$	11444.494	0.030	0.030
31	$5_0^1 14_0^2 16_0^1$	11050.308	0.039	0.039
32	$5_0^1 16_0^1$	8069.776	0.034	0.034
33	$10_0^1 12_0^1 14_0^2$	10703.411	0.027	0.027
34	12_0^1	6874.742	0.626	0.626
35	12_0^2	8136.872	0.198	0.198
36	12_0^3	9399.002	0.042	0.042
37	$12_0^1 14_0^1$	8365.008	1.028	1.028
38	$12_0^1 14_0^2$	9855.273	0.852	0.852
39	$12_0^1 14_0^3$	11345.539	0.476	0.476
40	$12_0^1 14_0^4$	12835.805	0.202	0.202
41	$12_0^1 14_0^5$	14326.070	0.069	0.069

42	$12_0^2 14_0^1$	9627.138	0.330	0.330
43	$12_0^2 14_0^2$	11117.403	0.277	0.277
44	$12_0^2 14_0^3$	12607.669	0.157	0.157
45	$12_0^2 14_0^4$	14097.935	0.067	0.067
46	$12_0^3 14_0^1$	10889.267	0.071	0.071
47	$12_0^3 14_0^2$	12379.533	0.061	0.061
48	$12_0^3 14_0^3$	13869.799	0.035	0.035
49	$12_0^1 14_0^1 16_0^1$	10249.460	0.875	0.875
50	$12_0^1 14_0^1 16_0^2$	12133.912	0.422	0.422
51	$12_0^1 14_0^1 16_0^3$	14018.365	0.150	0.150
52	$12_0^1 14_0^1 16_0^4$	15902.817	0.044	0.044
53	$12_0^1 14_0^2 16_0^1$	11739.726	0.677	0.677
54	$12_0^1 14_0^2 16_0^2$	13624.178	0.306	0.306
55	$12_0^1 14_0^2 16_0^3$	15508.630	0.103	0.103
56	$12_0^1 14_0^2 16_0^4$	17393.083	0.028	0.028
57	$12_0^1 14_0^3 16_0^1$	13229.991	0.353	0.353
58	$12_0^1 14_0^3 16_0^2$	15114.444	0.150	0.150
59	$12_0^1 14_0^3 16_0^3$	16998.896	0.047	0.047
60	$12_0^1 14_0^4 16_0^1$	14720.257	0.139	0.139
61	$12_0^1 14_0^4 16_0^2$	16604.709	0.055	0.055
62	$12_0^2 14_0^1 16_0^1$	11511.590	0.272	0.272
63	$12_0^2 14_0^1 16_0^2$	13396.042	0.127	0.127
64	$12_0^2 14_0^1 16_0^3$	15280.494	0.044	0.044
65	$12_0^2 14_0^2 16_0^1$	13001.856	0.213	0.213
66	$12_0^2 14_0^2 16_0^2$	14886.308	0.094	0.094
67	$12_0^2 14_0^2 16_0^3$	16770.760	0.031	0.031
68	$12_0^2 14_0^3 16_0^1$	14492.121	0.113	0.113
69	$12_0^2 14_0^3 16_0^2$	16376.573	0.046	0.046
70	$12_0^2 14_0^4 16_0^1$	15982.387	0.045	0.045
71	$12_0^3 14_0^1 16_0^1$	12773.720	0.057	0.057
72	$12_0^3 14_0^2 16_0^1$	14263.985	0.045	0.045
73	$12_0^1 16_0^1$	8759.194	0.570	0.570
74	$12_0^1 16_0^2$	10643.647	0.292	0.292
75	$12_0^1 16_0^3$	12528.099	0.110	0.110
76	$12_0^1 16_0^4$	14412.551	0.034	0.034
77	$12_0^2 16_0^1$	10021.324	0.175	0.175
78	$12_0^2 16_0^2$	11905.776	0.087	0.087
79	$12_0^2 16_0^3$	13790.229	0.032	0.032
80	$12_0^3 16_0^1$	11283.454	0.036	0.036
81	14_0^1	7102.878	1.618	1.618
82	14_0^2	8593.144	1.323	1.323
83	14_0^3	10083.409	0.729	0.729
84	14_0^4	11573.675	0.305	0.305
85	14_0^5	13063.941	0.103	0.103
86	14_0^6	14554.206	0.029	0.029
87	$14_0^1 16_0^1$	8987.330	1.421	1.421
88	$14_0^1 16_0^2$	10871.782	0.704	0.704
89	$14_0^1 16_0^3$	12756.235	0.257	0.257
90	$14_0^1 16_0^4$	14640.687	0.077	0.077
91	$14_0^2 16_0^1$	10477.596	1.085	1.085
92	$14_0^2 16_0^2$	12362.048	0.505	0.505
93	$14_0^2 16_0^3$	14246.501	0.174	0.174
94	$14_0^2 16_0^4$	16130.953	0.049	0.049
95	$14_0^3 16_0^1$	11967.862	0.558	0.558
96	$14_0^3 16_0^2$	13852.314	0.243	0.243
97	$14_0^3 16_0^3$	15736.766	0.079	0.079
98	$14_0^4 16_0^1$	13458.127	0.217	0.217
99	$14_0^4 16_0^2$	15342.580	0.089	0.089
100	$14_0^4 16_0^3$	17227.032	0.027	0.027
101	$14_0^5 16_0^1$	14948.393	0.068	0.068
102	16_0^1	7497.064	0.940	0.940
103	16_0^2	9381.517	0.495	0.495
104	16_0^3	11265.969	0.192	0.192
105	16_0^4	13150.421	0.061	0.061

FIGURE E.1: *Franck-Condon factors and signal intensities for the anion→neutral vibronic transitions based on the most stable conformers of OA and OA⁻ (structure **3**).*

C	-0.007614	0.067617	-0.044657
O	-0.012404	0.102824	1.220484
C	1.430454	-0.046910	-0.638850
O	1.722756	-0.008152	-1.830042
O	-0.971451	0.112203	-0.862059
O	2.345220	-0.190228	0.335938
O	-1.067600	-1.093192	-3.099364
C	-1.213968	-0.321504	-4.163519
C	-0.598784	0.943727	-4.477379
O	-0.925245	1.527363	-5.542639
O	-2.010997	-0.828822	-5.133761
O	0.307658	1.487700	-3.635062
H	-0.924316	-0.550697	-2.243407
H	1.748211	-0.135540	1.135138
H	0.615710	0.838994	-2.958716
H	-1.960606	-0.119753	-5.816024

C	3.192225	-0.094134	-0.098631
C	1.638803	-0.139342	-0.001531
O	1.137132	0.856519	0.625760
O	3.681852	1.022401	0.502695
O	3.904585	-0.925892	-0.630612
O	1.029234	-1.106965	-0.525141
O	-1.303592	-1.417824	0.493601
C	-2.320673	-0.631766	0.177439
O	-3.534297	-1.216348	0.315818
C	-2.368571	0.756340	-0.218645
O	-3.498652	1.247324	-0.488844
O	-1.255007	1.507797	-0.314193
H	2.853772	1.461347	0.817989
H	-0.422650	-1.173634	0.041175
H	-4.129831	-0.471653	0.058881
H	-0.442033	1.101247	0.097487

C	-0.007538	0.000000	-0.275186
O	-1.183460	0.000000	-0.783432
C	-0.007061	0.000000	1.282165
O	-1.275286	0.000000	1.771338
O	1.096921	0.000000	-0.874240
O	0.977782	0.000000	1.997323
O	1.013266	0.000000	-3.453830
C	-0.186689	0.000000	-4.075503
C	-0.191033	0.000000	-5.518625
O	-1.181213	0.000000	-6.251380
O	-1.341835	0.000000	-3.409643
O	1.086119	0.000000	-6.067878
H	-1.816254	0.000000	0.945166
H	-1.209808	0.000000	-2.413351
H	0.962148	0.000000	-2.437479
H	1.706819	0.000000	-5.317922

TABLE E.3: The MP2/ADZ minimum energy structures for the neutral and anionic dimers of OA. Atomic positions in direct coordinates.
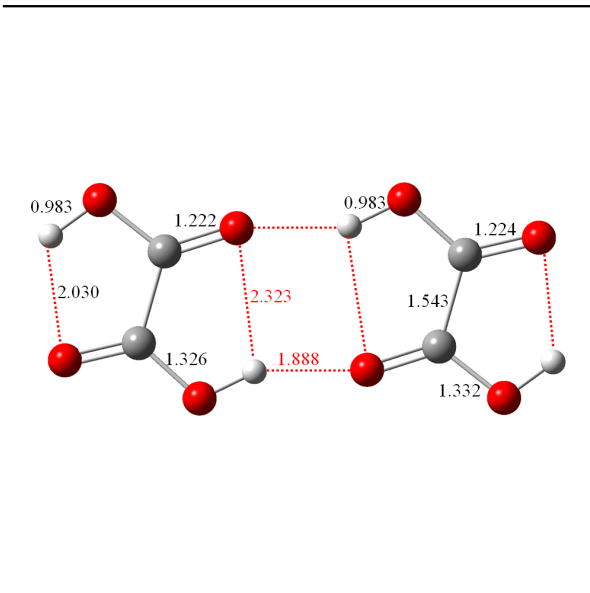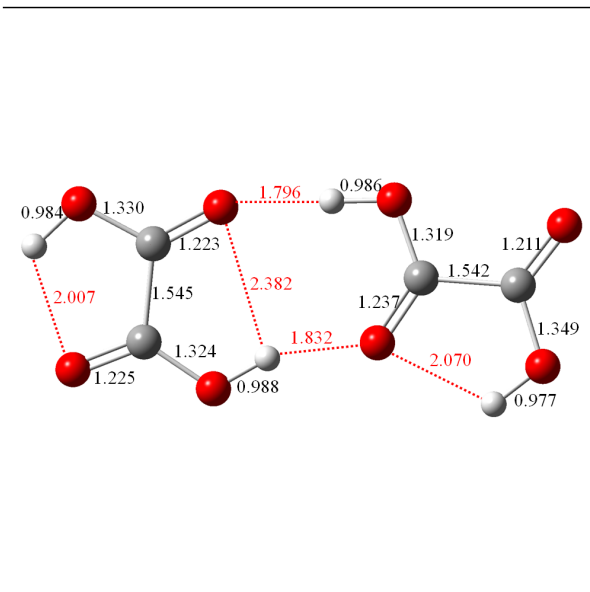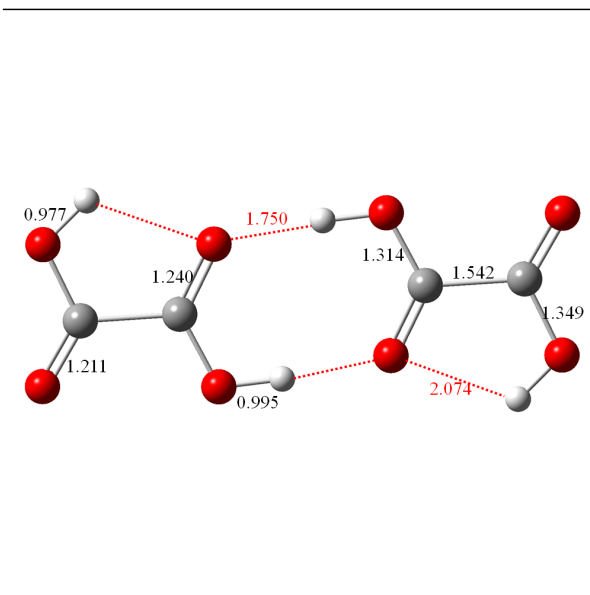
C	1.905763	-1.634008	0.007058
O	2.870748	-2.388132	-0.004465
C	2.148046	-0.109419	-0.001022
O	3.442837	0.202120	-0.020224
O	0.642822	-2.033343	0.026110
O	1.251304	0.720147	0.009174
O	-1.004680	2.280474	0.036381
C	-2.267617	1.881139	0.055612
O	-3.232606	2.635264	0.066764
C	-2.509904	0.356549	0.063560
O	-3.804699	0.045010	0.082541
O	-1.613165	-0.473018	0.053370
H	0.023807	-1.268275	0.032589
H	3.918726	-0.658218	-0.024308
H	-0.385662	1.515407	0.029970
H	-4.280585	0.905351	0.086659

C	1.406257	-1.177618	0.000000
O	0.363997	-1.985258	0.000000
O	1.377402	0.058620	0.000000
O	-1.821241	-0.268369	0.000000
C	-2.408763	0.804284	0.000000
O	-3.733995	0.914377	0.000000
C	-1.702938	2.178815	0.000000
O	-0.378952	2.179283	0.000000
O	-2.393982	3.190320	0.000000
H	-0.450504	-1.429317	0.000000
H	0.000000	1.266456	0.000000
H	-3.915535	1.881241	0.000000
C	2.757053	-1.920965	0.000000
O	2.861490	-3.127932	0.000000
O	3.802903	-1.068822	0.000000
H	3.435394	-0.163228	0.000000

C	0.094868	-1.936476	0.000000
O	1.293480	-1.397081	0.000000
O	-0.987609	-1.330843	0.000000
C	0.112810	-3.478626	0.000000
O	1.129952	-4.136346	0.000000
O	-1.129952	-4.002674	0.000000
O	0.987609	1.330843	0.000000
C	-0.094868	1.936476	0.000000
O	-1.293480	1.397081	0.000000
C	-0.112810	3.478626	0.000000
O	1.129952	4.002674	0.000000
O	-1.129952	4.136346	0.000000
H	-1.759983	-3.256066	0.000000
H	-1.199539	0.406107	0.000000
H	1.199539	-0.406107	0.000000
H	1.759983	3.256066	0.000000

TABLE E.4: The MP2/ADZ minimum energy structures for the neutral and anionic dimers of OA. Atomic positions in direct coordinates.

Bibliography

[1] P. Atkins, T. Overton, J. Rourke, M. Weller, and F. Armstrong. *Shriver and Atkins' Inorganic Chemistry*. Oxford University Press, 2010.

[2] Gilbert N. Lewis. The atom and the molecule. *Journal of the American Chemical Society*, 38(4):762–785, 1916. doi: 10.1021/ja02261a002. URL http://pubs.acs.org/doi/abs/10.1021/ja02261a002.

[3] Linus Pauling. The nature of the chemical bond. iv. the energy of single bonds and the relative electronegativity of atoms. *Journal of the American Chemical Society*, 54(9):3570–3582, 1932. doi: 10.1021/ja01348a011. URL http://pubs.acs.org/doi/abs/10.1021/ja01348a011.

[4] R. Abegg. Die valenz und das periodische system. versuch einer theorie der molekularverbindungen. *Zeitschrift fr anorganische Chemie*, 39(1):330–380, 1904. ISSN 1521-3749. doi: 10.1002/zaac.19040390125. URL http://dx.doi.org/10.1002/zaac.19040390125.

[5] Cambridge structural database, May 2012. URL http://www.ccdc.cam.ac.uk/products/csd/radii/.

[6] M. Nic, J. Jirat, and B. Kosata. Iupac, compendium of chemical terminology,, April 1997. URL http://goldbook.iupac.org/B00699.html.

[7] N. V. Sidgwick and H. M. Powell. Bakerian lecture. stereochemical types and valency groups. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 176(965):153–180, 1940. doi: 10.1098/rspa.1940.0084. URL http://rspa.royalsocietypublishing.org/content/176/965/153.abstract.

[8] R. J. Gillespie and R. S. Nyholm. Inorganic stereochemistry. *Q. Rev. Chem. Soc.*, 11:339–380, 1957. doi: 10.1039/QR9571100339. URL http://dx.doi.org/10.1039/QR9571100339.

[9] Andrew R. Leach. *Molecular Modelling: Principles and Applications*. Pearson Education. Prentice Hall, 2001. ISBN 9780582382107. URL http://books.google.co.uk/books?id=kB7jsbV-uhkC.

[10] Christopher J Cramer. *Essentials of computational chemistry: theories and models.* Wiley. com, 2005.

[11] M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. *Annalen der Physik*, 389(20):457–484, 1927. ISSN 1521-3889. doi: 10.1002/andp.19273892002. URL http://dx.doi.org/10.1002/andp.19273892002.

[12] H. Eyring and M. Polanyi. Concerning simple gas reactions. *Z. Phys. Chem. Abt. B*, 12:279, 1931.

[13] Henry Eyring. The activated complex in chemical reactions. *The Journal of Chemical Physics*, 3(2):107–115, 1935. doi: 10.1063/1.1749604. URL http://link.aip.org/link/?JCP/3/107/1.

[14] Philip M. Morse. Diatomic molecules according to the wave mechanics. ii. vibrational levels. *Phys. Rev.*, 34:57–64, Jul 1929. doi: 10.1103/PhysRev.34.57. URL http://link.aps.org/doi/10.1103/PhysRev.34.57.

[15] J. Franck and E. G. Dymond. Elementary processes of photochemical reactions. *Trans. Faraday Soc.*, 21:536–542, 1926. doi: 10.1039/TF9262100536. URL http://dx.doi.org/10.1039/TF9262100536.

[16] Edward U. Condon. Nuclear motions associated with electron transitions in diatomic molecules. *Phys. Rev.*, 32:858–872, Dec 1928. doi: 10.1103/PhysRev.32.858. URL http://link.aps.org/doi/10.1103/PhysRev.32.858.

[17] P.W. Atkins and R.S. Friedman. *Molecular Quantum Mechanics.* OUP Oxford, fifth edition, 2011. ISBN 9780199541423. URL http://books.google.co.uk/books?id=yk7lRAAACAAJ.

[18] Frank H. Allen, Stephanie E. Harris, and Robin Taylor. Comparison of conformer distributions in the crystalline state with conformational energies calculated by ab initio techniques. *Journal of Computer-Aided Molecular Design*, 10(3):247–254, 1996.

[19] Aurora J. Cruz-Cabeza, John W. Liebeschuetz, and Frank H. Allen. Systematic conformational bias in small-molecule crystal structures is rare and explicable. *CrystEngComm*, 14:6797–6811, 2012. doi: 10.1039/C2CE25585E. URL http://dx.doi.org/10.1039/C2CE25585E.

[20] David A. Gallagher. Conformatonal searching. Online material. URL http://chemed.chem.pitt.edu/cacc/Training%20Pitt/Conformational%20Searching_University%20of%20Pittsburgh.pdf.

[21] Maciej Haraczyk and Maciej Gutowski. Quantum mechanical energy-based screening of combinatorially generated library of tautomers. tautgen: a tautomer generator program. *J Chem Inf Model*, 47(2):686–694, 2007. doi: 10.1021/ci6002703. URL http://dx.doi.org/10.1021/ci6002703.

[22] Maciej Haranczyk and Maciej Gutowski. Combinatorial-computational-chemoinformatics (c3) approach to finding and analyzing low-energy tautomers. *J Comput Aided Mol Des*, 24(6-7):627–638, Jun 2010. doi: 10.1007/s10822-010-9344-6. URL http://dx.doi.org/10.1007/s10822-010-9344-6.

[23] Maciej Haranczyk, Tomasz Puzyn, and Pawe Sadowski. Congener a tool for modeling of the congeneric sets of environmental pollutants. *QSAR & Combinatorial Science*, 27(7):826–833, 2008. ISSN 1611-0218. doi: 10.1002/qsar.200710149. URL http://dx.doi.org/10.1002/qsar.200710149.

[24] Paul C. D. Hawkins, A. Geoffrey Skillman, Gregory L. Warren, Benjamin A. Ellingson, and Matthew T. Stahl. Conformer generation with omega: Algorithm and validation using high quality structures from the protein databank and cambridge structural database. *Journal of Chemical Information and Modeling*, 50(4): 572–584, 2010. doi: 10.1021/ci100031x. URL http://pubs.acs.org/doi/abs/10.1021/ci100031x.

[25] Sanliang Ling and Maciej Gutowski. Ssc: a tool for constructing libraries for systematic screening of conformers. *J Comput Chem*, 32(9):2047–2054, Jul 2011. doi: 10.1002/jcc.21774. URL http://dx.doi.org/10.1002/jcc.21774.

[26] Jiabo Li, Tedman Ehlers, Jon Sutter, Shikha Varma-O'Brien, and Johannes Kirchmair. Caesar: a new conformer generation algorithm based on recursive buildup and local rotational symmetry consideration. *Journal of Chemical Information and Modeling*, 47(5):1923–1932, 2007. doi: 10.1021/ci700136x. URL http://pubs.acs.org/doi/abs/10.1021/ci700136x. PMID: 17691764.

[27] Xiang Li, Kit H Bowen, Maciej Haranczyk, Rafa A Bachorz, Kamil Mazurkiewicz, Janusz Rak, and Maciej Gutowski. Photoelectron spectroscopy of adiabatically bound valence anions of rare tautomers of the nucleic acid bases. *J Chem Phys*, 127(17):174309, Nov 2007. doi: 10.1063/1.2795719. URL http://dx.doi.org/10.1063/1.2795719.

[28] Zibo Keolopile, Maciej Gutowski, and Maciej Haranczyk. Discovery of most stable structures of neutral and anionic phenylalanine through automated scanning of tautomeric and conformational spaces. *Journal of Chemical Theory and Computation*, February 2013. URL http://sourceforge.net/projects/pesst/.

[29] Zibo G Keolopile, Maciej Gutowski, and Maciej Haranczyk. Discovery of most stable structures of neutral and anionic phenylalanine through automated scanning of tautomeric and conformational spaces. *Journal of Chemical Theory and Computation*, 0(ja):null, 0. doi: 10.1021/ct400531a. URL http://pubs.acs.org/doi/abs/10.1021/ct400531a. doi:10.1021/ct400531a.

[30] Charles L. Brooks, Jos N. Onuchic, and David J. Wales. Taking a walk on a landscape. *Science*, 293(5530):612–613, 2001. doi: 10.1126/science.1062559. URL http://www.sciencemag.org/content/293/5530/612.short.

[31] D. Wales. *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses.* Cambridge Molecular Science. Cambridge University Press, 2003. ISBN 9780521814157. URL http://books.google.co.uk/books?id=YQrB6s3LALEC.

[32] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953. doi: 10.1063/1.1699114. URL http://link.aip.org/link/?JCP/21/1087/1.

[33] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids.* Oxford University Press, Oxford, 1987.

[34] David J. Wales and Jonathan P. K. Doye. Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms. *The Journal of Physical Chemistry A*, 101(28):5111–5116, 1997. doi: 10.1021/jp970984n. URL http://pubs.acs.org/doi/abs/10.1021/jp970984n.

[35] John H. Holland. Genetic algorithms. 267(1):66–72 (Intl. ed. 44–50), July 1992. ISSN 0036-8733 (print), 1946-7087 (electronic). doi: http://dx.doi.org/10.1038/scientificamerican0792-66. URL http://www.nature.com/scientificamerican/journal/v267/n1/pdf/scientificamerican0792-66.pdf.

[36] Lucjan Piela, Jaroslaw Kostrowicki, and Harold A. Scheraga. On the multiple-minima problem in the conformational analysis of molecules: deformation of the potential energy hypersurface by the diffusion equation method. *The Journal of Physical Chemistry*, 93(8):3339–3346, 1989. doi: 10.1021/j100345a090. URL http://pubs.acs.org/doi/abs/10.1021/j100345a090.

[37] Jaroslaw Pillardy and Lucjan Piela. Molecular dynamics on deformed potential energy hypersurfaces. *The Journal of Physical Chemistry*, 99(31):11805–11812, 1995. doi: 10.1021/j100031a003. URL http://pubs.acs.org/doi/abs/10.1021/j100031a003.

[38] Jaroslaw Pillardy, Krzysztof A. Olszewski, and Lucjan Piela. Performance of the shift method of global minimization in searches for optimum structures of clusters of lennard-jones atoms. *The Journal of Physical Chemistry*, 96(11):4337–4341, 1992. doi: 10.1021/j100190a042. URL http://pubs.acs.org/doi/abs/10.1021/j100190a042.

[39] Jianpeng Ma and John E. Straub. Simulated annealing using the classical density distribution. *The Journal of Chemical Physics*, 101(1):533–541, 1994. doi: 10.1063/1.468163. URL http://link.aip.org/link/?JCP/101/533/1.

[40] Chiachin Tsoo and Charles L. Brooks III. Cluster structure determination using gaussian density distribution global minimization methods. *The Journal of Chemical Physics*, 101(8):6405–6411, 1994. doi: 10.1063/1.468397. URL http://link.aip.org/link/?JCP/101/6405/1.

[41] Frank H. Stillinger and Dorothea K. Stillinger. Cluster optimization simplified by interaction modification. *The Journal of Chemical Physics*, 93(8):6106–6107, 1990. doi: 10.1063/1.459003. URL http://link.aip.org/link/?JCP/93/6106/1.

[42] Satya Bulusu and X. C. Zeng. Structures and relative stability of neutral gold clusters: Au[sub n] (n = 15–19). *The Journal of Chemical Physics*, 125(15): 154303, 2006. doi: 10.1063/1.2352755. URL http://link.aip.org/link/?JCP/125/154303/1.

[43] De-en Jiang, Michael Walter, and Sheng Dai. Gold sulfide nanoclusters: A unique core-in-cage structure. *Chemistry A European Journal*, 16(17):4999–5003, 2010. ISSN 1521-3765. doi: 10.1002/chem.201000327. URL http://dx.doi.org/10.1002/chem.201000327.

[44] A. Einstein. ber einen die erzeugung und verwandlung des lichtes betreffenden heuristischen gesichtspunkt. *Annalen der Physik*, 322(6):132–148, 1905. ISSN 1521-3889. doi: 10.1002/andp.19053220607. URL http://dx.doi.org/10.1002/andp.19053220607.

[45] O. David Sparkman. *Mass spectrometry desk reference*. Pittsburgh: Global View Pub., 2000.

[46] Phil Price. Standard definitions of terms relating to mass spectrometry. *Journal of the American Society for Mass Spectrometry*, 2(4):336–348, 1991. ISSN 1044-0305. doi: 10.1016/1044-0305(91)80025-3. URL http://dx.doi.org/10.1016/1044-0305%2891%2980025-3.

[47] D.A. McQuarrie and J.D. Simon. *Physical Chemistry: A Molecular Approach*. University Science Books, 1997. ISBN 9780935702996. URL http://books.google.de/books?id=f-bje0-DEYUC.

[48] P.W. Atkins, J. De Paula, and R. Friedman. *Quanta, matter, and change: a molecular approach to physical chemistry*. Oxford University Press, 2009. ISBN 9780199206063. URL http://books.google.co.uk/books?id=r2YvAQAAIAAJ.

[49] P. Atkins and J. de Paula. *Atkins' Physical Chemistry*. OUP Oxford, 2010. ISBN 9780199543373. URL http://books.google.co.uk/books?id=nQx_LgEACAAJ.

[50] Rigid rotor approximation, accessed 2012. URL http://en.wikipedia.org/wiki/Rigid_rotor.

[51] W. Gordy. *Microwave Molecular Spectra in Technique of Organic Chemistry*, volume IX. New York: Interscience.

[52] Arthur H. Compton. A quantum theory of the scattering of x-rays by light elements. *Phys. Rev.*, 21:483–502, May 1923. doi: 10.1103/PhysRev.21.483. URL http://link.aps.org/doi/10.1103/PhysRev.21.483.

[53] von Laue M. Concerning the detection of x-ray interferences. *Nobel Lectures, Physics*, Retrieved 2009-02-18., 19011921.

[54] W. Friedrich, P. Knipping, and M. von Laue. Interferenz-erscheinungen bei rntgenstrahlen. *Sitzungsberichte der Mathematisch-Physikalischen Classe der Kniglich-Bayerischen Akademie der Wissenschaften zu Mnchen*, page 303, 1912.

[55] W. L. Bragg. The specular reflection of x-rays. *nature*, 90:410, December 1912. doi: 10.1038/090410b0.

[56] W. L. Bragg. The diffraction of short electromagnetic waves by a crystal. *Proceedings of the Cambridge Philosophical Society*, 17:43, 1913.

[57] W. L. Bragg. Die reflexion der rntgenstrahlen. *Jahrbuch der Radioaktivitt und Elektronik*, 11:350, 1914.

[58] Protein data bank, 2013. URL http://www.pdb.org.

[59] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000. doi: 10.1093/nar/28.1.235. URL http://nar.oxfordjournals.org/content/28/1/235.abstract.

[60] L. de Broglie. Recherches sur la thorie des quanta (researches on the quantum theory), thesis (paris), 1924; l. de broglie. *Annual Physics (Paris)*, 3:22, 1925.

[61] Louis Broglie. The reinterpretation of wave mechanics. *Foundations of Physics*, 1(1):5–15, 1970. ISSN 0015-9018. doi: 10.1007/BF00708650. URL http://dx.doi.org/10.1007/BF00708650.

[62] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff. Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society*, 114(25):10024–10035, 1992. doi: 10.1021/ja00051a040. URL http://pubs.acs.org/doi/abs/10.1021/ja00051a040.

[63] Stephen L. Mayo, Barry D. Olafson, and William A. Goddard. Dreiding: a generic force field for molecular simulations. *The Journal of Physical Chemistry*, 94(26): 8897–8909, 1990. doi: 10.1021/j100389a010. URL http://pubs.acs.org/doi/abs/10.1021/j100389a010.

[64] William L. Jorgensen, David S. Maxwell, and Julian Tirado-Rives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45):11225–11236, 1996. doi: 10.1021/ja9621760. URL http://pubs.acs.org/doi/abs/10.1021/ja9621760.

[65] Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules j. am. chem. soc. 1995, 117, 51795197. *Journal of the American Chemical Society*, 118(9):2309–2309, 1996. doi: 10.1021/ja955032e. URL http://pubs.acs.org/doi/abs/10.1021/ja955032e.

[66] William L. Jorgensen and Julian. Tirado-Rives. The opls [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6):1657–1666, 1988. doi: 10.1021/ja00214a001. URL http://pubs.acs.org/doi/abs/10.1021/ja00214a001.

[67] Thomas A. Halgren. Merck molecular force field. i. basis, form, scope, parameterization, and performance of mmff94. *Journal of Computational Chemistry*, 17(5-6):490–519, 1996. ISSN 1096-987X. doi: 10.1002/(SICI)1096-987X(199604)17:5/6⟨490::AID-JCC1⟩3.0.CO;2-P. URL http://dx.doi.org/10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P.

[68] Thomas A. Halgren. Mmff vi. mmff94s option for energy minimization studies. *Journal of Computational Chemistry*, 20(7):720–729, 1999. ISSN

1096-987X. doi: 10.1002/(SICI)1096-987X(199905)20:7⟨720::AID-JCC7⟩3.0.CO;
2-X. URL `http://dx.doi.org/10.1002/(SICI)1096-987X(199905)20:7<720::AID-JCC7>3.0.CO;2-X`.

[69] Norman L. Allinger. Conformational analysis. 130. mm2. a hydrocarbon force field utilizing v1 and v2 torsional terms. *Journal of the American Chemical Society*, 99(25):8127–8134, 1977. doi: 10.1021/ja00467a001. URL `http://pubs.acs.org/doi/abs/10.1021/ja00467a001`.

[70] Norman L. Allinger, Randall A. Kok, and Mita R. Imam. Hydrogen bonding in mm2. *Journal of Computational Chemistry*, 9(6):591–595, 1988. ISSN 1096-987X. doi: 10.1002/jcc.540090602. URL `http://dx.doi.org/10.1002/jcc.540090602`.

[71] Norman L. Allinger, Young H. Yuh, and Jenn Huei Lii. Molecular mechanics. the mm3 force field for hydrocarbons. 1. *Journal of the American Chemical Society*, 111(23):8551–8566, 1989. doi: 10.1021/ja00205a001. URL `http://pubs.acs.org/doi/abs/10.1021/ja00205a001`.

[72] A. D. MacKerell, D. Bashford, Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wirkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102(18):3586–3616, 1998. doi: 10.1021/jp973084f. URL `http://pubs.acs.org/doi/abs/10.1021/jp973084f`.

[73] Christopher I. Bayly, Piotr Cieplak, Wendy Cornell, and Peter A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model. *The Journal of Physical Chemistry*, 97(40):10269–10280, 1993. doi: 10.1021/j100142a004. URL `http://pubs.acs.org/doi/abs/10.1021/j100142a004`.

[74] Anthony K Rappe and William A Goddard III. Charge equilibration for molecular dynamics simulations. *The Journal of Physical Chemistry*, 95(8):3358–3363, 1991.

[75] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin,

V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. R. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian 09 Revision A.1, 2009.

[76] M. P. Allen. *Computer simulation of liquids.* Clarendon Press Oxford University Press, Oxford England New York, 1989. ISBN 0198556454.

[77] B. J. Alder and T. E. Wainwright. Studies in molecular dynamics. i. general method. *The Journal of Chemical Physics*, 31(2):459–466, 1959. doi: 10.1063/1. 1730376. URL http://link.aip.org/link/?JCP/31/459/1.

[78] Dirk Kroese. *Handbook of Monte Carlo methods.* Wiley, Hoboken, N.J, 2011. ISBN 0470177934.

[79] Loup Verlet. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159:98–103, Jul 1967. doi: 10. 1103/PhysRev.159.98. URL http://link.aps.org/doi/10.1103/PhysRev.159. 98.

[80] D. Beeman. Some multistep methods for use in molecular dynamics calculations. *Journal of Computational Physics*, 20(2):130 – 139, 1976. ISSN 0021-9991. doi: http://dx.doi.org/10.1016/0021-9991(76)90059-0. URL http://www. sciencedirect.com/science/article/pii/0021999176900590.

[81] Michel A. Cuendet and Wilfred F. van Gunsteren. On the calculation of velocity-dependent properties in molecular dynamics simulations using the leapfrog integration algorithm. *The Journal of Chemical Physics*, 127(18):184102, 2007. doi: 10.1063/1.2779878. URL http://link.aip.org/link/?JCP/127/184102/1.

[82] M.Germana Paterlini and David M Ferguson. Constant temperature simulations using the langevin equation with velocity verlet integration. *Chemical Physics*, 236(13):243 – 252, 1998. ISSN 0301-0104. doi: http://dx.doi.org/10. 1016/S0301-0104(98)00214-6. URL http://www.sciencedirect.com/science/ article/pii/S0301010498002146.

[83] William C. Swope, Hans C. Andersen, Peter H. Berens, and Kent R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters.

The Journal of Chemical Physics, 76(1):637–649, 1982. doi: 10.1063/1.442716. URL http://link.aip.org/link/?JCP/76/637/1.

[84] Keith Refson. Moldy: a portable molecular dynamics simulation program for serial and parallel computers. *Computer Physics Communications*, 126(3):310 – 329, 2000. ISSN 0010-4655. doi: http://dx.doi.org/10.1016/S0010-4655(99)00496-8. URL http://www.sciencedirect.com/science/article/pii/S0010465599004968.

[85] Pavel Pospisil, Patrick Ballmer, Leonardo Scapozza, and Gerd Folkers. Tautomerism in computeraided drug design. *Journal of Receptors and Signal Transduction*, 23(4):361–371, 2003. doi: 10.1081/RRS-120026975. URL http://informahealthcare.com/doi/abs/10.1081/RRS-120026975. PMID: 14753297.

[86] Scott Kirkpatrick, D. Gelatt Jr., and Mario P. Vecchi. Optimization by simmulated annealing. *Science*, 220(4598):671–680, 1983.

[87] A Szabo and N S Ostlund. *Modern Quantum Chemistry*. McGraw Hill, New York, 1982.

[88] T Helgaker, P Jørgensen, and J Olsen. *Molecular Electronic Structure Theory*. John Wiley & Sons, LTD, Chichester, 2000.

[89] Frank Jensen. *Introduction to Computational Chemistry*. Wiley, 2 edition, 2006. ISBN 0470011874.

[90] Lucjan Piela. *Ideas of Quantum Chemistry*. Elsevier, 2007.

[91] Nouredine Zettili. *Quantum Mechanics: Concepts and Applications*. Wiley, 2 edition, February 2009. ISBN 0470026782.

[92] J. C. Slater. The theory of complex spectra. *Phys. Rev.*, 34:1293–1322, Nov 1929. doi: 10.1103/PhysRev.34.1293. URL http://link.aps.org/doi/10.1103/PhysRev.34.1293.

[93] C. C. J. Roothaan. New developments in molecular orbital theory. *Rev. Mod. Phys.*, 23:69–89, Apr 1951. doi: 10.1103/RevModPhys.23.69. URL http://link.aps.org/doi/10.1103/RevModPhys.23.69.

[94] Krishnan Raghavachari, Gary W. Trucks, John A. Pople, and Martin Head-Gordon. A fifth-order perturbation comparison of electron correlation theories. *Chemical Physics Letters*, 157(6):479 – 483, 1989. ISSN 0009-2614. doi: 10.1016/S0009-2614(89)87395-6. URL http://www.sciencedirect.com/science/article/pii/S0009261489873956.

[95] L. H. Thomas. The calculation of atomic fields. *Mathematical Proceedings of the Cambridge Philosophical Society*, 23:542–548, 1 1927. ISSN 1469-8064. doi: 10.1017/S0305004100011683. URL http://journals.cambridge.org/article_S0305004100011683.

[96] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, Nov 1964. doi: 10.1103/PhysRev.136.B864. URL http://link.aps.org/doi/10.1103/PhysRev.136.B864.

[97] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, Nov 1965. doi: 10.1103/PhysRev.140.A1133. URL http://link.aps.org/doi/10.1103/PhysRev.140.A1133.

[98] John P. Perdew and Wang Yue. Accurate and simple density functional for the electronic exchange energy: Generalized gradient approximation. *Physical Review B*, 33(12):8800–8802, June 1986. doi: 10.1103/physrevb.33.8800. URL http://dx.doi.org/10.1103/physrevb.33.8800.

[99] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A*, 38(6):3098–3100, September 1988. doi: 10.1103/physreva.38.3098. URL http://dx.doi.org/10.1103/physreva.38.3098.

[100] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37:785–789, Jan 1988. doi: 10.1103/PhysRevB.37.785. URL http://link.aps.org/doi/10.1103/PhysRevB.37.785.

[101] John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical review letters*, 77(18):3865, 1996.

[102] J. M. Perez-Jorda and A. D. Becke. A density-functional study of van der waals forces: rare gas diatomics. *Chemical Physics Letters*, 233(1):134–137, 1995. doi: doi:10.1016/0009-2614(94)01402-H. URL http://www.ingentaconnect.com/content/els/00092614/1995/00000233/00000001/art01402.

[103] Jose M. Perez-Jorda, Emilio San-Fabian, and Angel J. Perez-Jimenez. Density-functional study of van der waals forces on rare-gas diatomics: Hartree–fock exchange. *The Journal of Chemical Physics*, 110(4):1916–1920, 1999. doi: 10.1063/1.477858. URL http://link.aip.org/link/?JCP/110/1916/1.

[104] Yingkai Zhang, Wei Pan, and Weitao Yang. Describing van der waals interaction in diatomic molecules with generalized gradient approximations: The role of the

exchange functional. *The Journal of Chemical Physics*, 107(19):7921–7925, 1997. doi: 10.1063/1.475105. URL http://link.aip.org/link/?JCP/107/7921/1.

[105] David C. Patton and Mark R. Pederson. Application of the generalized-gradient approximation to rare-gas dimers. *Phys. Rev. A*, 56:R2495–R2498, Oct 1997. doi: 10.1103/PhysRevA.56.R2495. URL http://link.aps.org/doi/10.1103/PhysRevA.56.R2495.

[106] Joachim Sauer and Jens Dobler. Structure and reactivity of v2o5: bulk solid, nanosized clusters, species supported on silica and alumina, cluster cations and anions. *Dalton Trans.*, pages 3116–3121, 2004. doi: 10.1039/B402873B. URL http://dx.doi.org/10.1039/B402873B.

[107] K. Terakura, T. Oguchi, A. R. Williams, and J. Kübler. Band theory of insulating transition-metal monoxides: Band-structure calculations. *Phys. Rev. B*, 30:4734–4747, Oct 1984. doi: 10.1103/PhysRevB.30.4734. URL http://link.aps.org/doi/10.1103/PhysRevB.30.4734.

[108] Vladimir I. Anisimov, Jan Zaanen, and Ole K. Andersen. Band theory and mott insulators: Hubbard *U* instead of stoner *I*. *Phys. Rev. B*, 44:943–954, Jul 1991. doi: 10.1103/PhysRevB.44.943. URL http://link.aps.org/doi/10.1103/PhysRevB.44.943.

[109] Jianmin Tao, John P. Perdew, Viktor N. Staroverov, and Gustavo E. Scuseria. Climbing the density functional ladder: Nonempirical meta˘generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.*, 91:146401, Sep 2003. doi: 10.1103/PhysRevLett.91.146401. URL http://link.aps.org/doi/10.1103/PhysRevLett.91.146401.

[110] Yan Zhao and Donald G. Truhlar. Density functionals with broad applicability in chemistry. *Accounts of Chemical Research*, 41(2):157–167, 2008. doi: 10.1021/ar700111a. URL http://pubs.acs.org/doi/abs/10.1021/ar700111a. PMID: 18186612.

[111] S. H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics*, 58:1200–1211, 1980. doi: 10.1139/p80-159.

[112] Axel D. Becke. Density-functional thermochemistry. iii. the role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648–5652, 1993. doi: 10.1063/1.464913. URL http://link.aip.org/link/?JCP/98/5648/1.

[113] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density

functional force fields. 1994. URL http://pubs.acs.org/doi/abs/10.1021/j100096a001.

[114] Yan Zhao and Donald G. Truhlar. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *The Journal of Chemical Physics*, 125(19):194101, 2006. doi: 10.1063/1.2370993. URL http://link.aip.org/link/?JCP/125/194101/1.

[115] Jr. Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *The Journal of Chemical Physics*, 90(2):1007–1023, 1989. doi: 10.1063/1.456153. URL http://link.aip.org/link/?JCP/90/1007/1.

[116] Jan Almlof and Peter R. Taylor. General contraction of gaussian basis sets. i. atomic natural orbitals for first- and second-row atoms. *The Journal of Chemical Physics*, 86(7):4070–4077, 1987. doi: 10.1063/1.451917. URL http://link.aip.org/link/?JCP/86/4070/1.

[117] Karol Jankowski, Ralf Becherer, Peter Scharf, Heinz Schiffer, and Reinhart Ahlrichs. The impact of higher polarization basis functions on molecular ab initio results. i. the ground state of f[sub 2]. *The Journal of Chemical Physics*, 82 (3):1413–1419, 1985. doi: 10.1063/1.448464. URL http://link.aip.org/link/?JCP/82/1413/1.

[118] David E. Woon and Jr. Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. iii. the atoms aluminum through argon. *The Journal of Chemical Physics*, 98(2):1358–1371, 1993. doi: 10.1063/1.464303. URL http://link.aip.org/link/?JCP/98/1358/1.

[119] David E. Woon and Jr. Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. v. core-valence basis sets for boron through neon. *The Journal of Chemical Physics*, 103(11):4572–4585, 1995. doi: 10.1063/1.470645. URL http://link.aip.org/link/?JCP/103/4572/1.

[120] Angela K. Wilson, David E. Woon, Kirk A. Peterson, and Jr. Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. ix. the atoms gallium through krypton. *The Journal of Chemical Physics*, 110(16):7667–7676, 1999. doi: 10.1063/1.478678. URL http://link.aip.org/link/?JCP/110/7667/1.

[121] Rick A. Kendall, Jr. Thom H. Dunning, and Robert J. Harrison. Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions. *The*

Journal of Chemical Physics, 96(9):6796–6806, 1992. doi: 10.1063/1.462569. URL http://link.aip.org/link/?JCP/96/6796/1.

[122] Piotr Skurski, Maciej Gutowski, and Jack Simons. How to choose a one-electron basis set to reliably describe a dipole-bound anion. *International Journal of Quantum Chemistry*, 80(4-5):1024–1038, 2000. ISSN 1097-461X. doi: 10.1002/1097-461X(2000)80:4/5⟨1024::AID-QUA51⟩3.0.CO;2-P. URL http://dx.doi.org/10.1002/1097-461X(2000)80:4/5<1024::AID-QUA51>3.0.CO;2-P.

[123] Trygve Helgaker, Wim Klopper, Henrik Koch, and Jozef Noga. Basis-set convergence of correlated calculations on water. *The Journal of Chemical Physics*, 106 (23):9639–9646, 1997. doi: 10.1063/1.473863. URL http://link.aip.org/link/?JCP/106/9639/1.

[124] David Feller. The use of systematic sequences of wave functions for estimating the complete basis set, full configuration interaction limit in water. *The Journal of Chemical Physics*, 98(9):7059–7071, 1993. doi: 10.1063/1.464749. URL http://link.aip.org/link/?JCP/98/7059/1.

[125] Rafal A. Bachorz, Wim Klopper, and Maciej Gutowski. Coupled-cluster and explicitly correlated perturbation-theory calculations of the uracil anion. *The Journal of Chemical Physics*, 126(8):085101, 2007. doi: 10.1063/1.2436890. URL http://link.aip.org/link/?JCP/126/085101/1.

[126] Joseph W. Ochterski. Thermochemistry in gaussian, April 2000. URL http://www.gaussian.com/g_whitepap/thermo.htm.

[127] D.A. McQuarrie. *Statistical Mechanics*. University Science Books, 2000. ISBN 9781891389153. URL http://books.google.co.uk/books?id=itcpPnDnJMOC.

[128] E. Fermi and E. Teller. The capture of negative mesotrons in matter. *Phys. Rev.*, 72:399–408, Sep 1947. doi: 10.1103/PhysRev.72.399. URL http://link.aps.org/doi/10.1103/PhysRev.72.399.

[129] Maciej Gutowski, Kenneth D. Jordan, and Piotr Skurski. Electronic structure of dipole-bound anions. *The Journal of Physical Chemistry A*, 102(15):2624–2633, 1998. doi: 10.1021/jp980123u. URL http://pubs.acs.org/doi/abs/10.1021/jp980123u.

[130] Maciej Gutowski and Piotr Skurski. Theoretical study of the dipole-bound anion (hf)[sub 2][sup -]. *The Journal of Chemical Physics*, 107(8):2968–2973, 1997. doi: 10.1063/1.474654. URL http://link.aip.org/link/?JCP/107/2968/1.

[131] Maciej Gutowski and Piotr Skurski. Dispersion stabilization of solvated electrons and dipole-bound anions. *The Journal of Physical Chemistry B*, 101(45):9143–9146, 1997. doi: 10.1021/jp972600z. URL http://pubs.acs.org/doi/abs/10.1021/jp972600z.

[132] C. Møller and M. S. Plesset. Note on an approximation treatment for many-electron systems. *Phys. Rev.*, 46:618–622, Oct 1934. doi: 10.1103/PhysRev.46.618. URL http://link.aps.org/doi/10.1103/PhysRev.46.618.

[133] Gerard Meijer, Mattanjah S. de Vries, Heinrich E. Hunziker, and H. Russell Wendt. Laser desorption jet-cooling spectroscopy of para-amino benzoic acid monomer, dimer, and clusters. *The Journal of Chemical Physics*, 92(12):7625–7635, 1990. doi: 10.1063/1.458200. URL http://link.aip.org/link/?JCP/92/7625/1.

[134] Isabel Hnig, Christian Pltzer, Kai A. Seefeld, Dennis Lwenich, Michael Nispel, and Karl Kleinermanns. Cover picture: Photostability of isolated and paired nucleobases: Nh dissociation of adenine and hydrogen transfer in its base pairs examined by laser spectroscopy (chemphyschem 9/2004). *ChemPhysChem*, 5(9): 1257–1257, 2004. ISSN 1439-7641. doi: 10.1002/cphc.200490047. URL http://dx.doi.org/10.1002/cphc.200490047.

[135] Weijun Zheng, Shoujun Xu, Dunja Radisic, Sarah Stokes, Xiang Li, and Jr. Kit H. Bowen. On the interaction of electrons with betaine zwitterions. *The Journal of Chemical Physics*, 122(10):101103, 2005. doi: 10.1063/1.1871912. URL http://link.aip.org/link/?JCP/122/101103/1.

[136] Ilko Bald, Iwona Dbkowska, and Eugen Illenberger. Probing biomolecules by laser-induced acoustic desorption: Electrons at near zero electron volts trigger sugarphosphate cleavage. *Angewandte Chemie International Edition*, 47 (44):8518–8520, 2008. ISSN 1521-3773. doi: 10.1002/anie.200803382. URL http://dx.doi.org/10.1002/anie.200803382.

[137] J.P. Schermann. *Spectroscopy and Modeling of Biomolecular Building Blocks*. Elsevier Science, 2007. ISBN 9780080558226. URL http://books.google.co.uk/books?id=N-5QUjjsQSMC.

[138] Carolyn Pratt Brock and Robin P. Minton. Systematic effects of crystal-packing forces: biphenyl fragments with hydrogen atoms in all four ortho positions. *Journal of the American Chemical Society*, 111(13):4586–4593, 1989. doi: 10.1021/ja00195a010. URL http://pubs.acs.org/doi/abs/10.1021/ja00195a010.

[139] D.W.H. Rankin, N. Mitzel, and C. Morrison. *Structural Methods in Molecular Inorganic Chemistry*. Inorganic Chemistry: A Textbook Series. Wiley, 2013. ISBN 9781118462881. URL http://books.google.co.uk/books?id=FIwg4qz0Z8UC.

[140] Jose M Hermida-Ramn, Enrique M Cabaleiro-Lago, and Jess Rodrguez-Otero. Computational study of the dissociation of oxalic acid in water clusters. *Chemical Physics*, 302(13):53 – 60, 2004. ISSN 0301-0104. doi: http://dx.doi.org/10.1016/j.chemphys.2004.02.021. URL http://www.sciencedirect.com/science/article/pii/S0301010404000886.

[141] Charles R Cantor and 1940-(joint author.) Schimmel, Paul R. (Paul Reinhard). *Biophysical chemistry*. San Francisco : W.H. Freeman and Company, 1980. ISBN 0716711907 (pbk. : pt. 2). Includes bibliographies and indexes.

[142] Rafal A. Bachorz, Janusz Rak, and Maciej Gutowski. Stabilization of very rare tautomers of uracil by an excess electron. *Phys. Chem. Chem. Phys.*, 7:2116–2125, 2005. doi: 10.1039/B503745J. URL http://dx.doi.org/10.1039/B503745J.

[143] Maciej Harańczyk, Janusz Rak, and Maciej Gutowski. Stabilization of very rare tautomers of 1-methylcytosine by an excess electron. *J Phys Chem A*, 109(50): 11495–11503, Dec 2005. doi: 10.1021/jp0535590. URL http://dx.doi.org/10.1021/jp0535590.

[144] Maciej Harańczyk and Maciej Gutowski. Finding adiabatically bound anions of guanine through a combinatorial computational approach. *Angew Chem Int Ed Engl*, 44(40):6585–6588, Oct 2005. doi: 10.1002/anie.200501671. URL http://dx.doi.org/10.1002/anie.200501671.

[145] Sanliang Ling, Wenbo Yu, Zhijian Huang, Zijing Lin, Maciej Haraczyk, and Maciej Gutowski. Gaseous arginine conformers and their unique intramolecular interactions. *The Journal of Physical Chemistry A*, 110(44):12282–12291, 2006. doi: 10.1021/jp0645115. URL http://pubs.acs.org/doi/abs/10.1021/jp0645115.

[146] Maciej Gutowski, Piotr Skurski, and Jack Simons. Dipole-bound anions of glycine based on the zwitterion and neutral structures. *Journal of the American Chemical Society*, 122(41):10159–10162, 2000. doi: 10.1021/ja001658f. URL http://pubs.acs.org/doi/abs/10.1021/ja001658f.

[147] Piotr Skurski, Janusz Rak, Jack Simons, and Maciej Gutowski. Quasidegeneracy of zwitterionic and canonical tautomers of arginine solvated by an excess electron. *Journal of the American Chemical Society*, 123(44):11073–11074, 2001. doi: 10.1021/ja011358d. URL http://pubs.acs.org/doi/abs/10.1021/ja011358d.

[148] Shoujun Xu, Weijun Zheng, Dunja Radisic, and Jr. Kit H. Bowen. The stabilization of arginine's zwitterion by dipole-binding of an excess electron. *The Journal of Chemical Physics*, 122(9):091103, 2005. doi: 10.1063/1.1864952. URL http://link.aip.org/link/?JCP/122/091103/1.

[149] Shoujun Xu, J. Michael Nilles, and Jr. Kit H. Bowen. Zwitterion formation in hydrated amino acid, dipole bound anions: How many water molecules are required? *The Journal of Chemical Physics*, 119(20):10696–10701, 2003. doi: 10.1063/1.1620501. URL http://link.aip.org/link/?JCP/119/10696/1.

[150] Maciej Haranczyk and Maciej Gutowski. Effect of excess electron and one water molecule on relative stability of the canonical and zwitterionic tautomers of glycine. *J Chem Phys*, 128(12):125101, Mar 2008. doi: 10.1063/1.2838910. URL http://dx.doi.org/10.1063/1.2838910.

[151] Selso J. Martinez-III, Joseph C. Alfano, and Donald H. Levy. The electronic spectroscopy of the amino acids tyrosine and phenylalanine in a supersonic jet. *Journal of Molecular Spectroscopy*, 156(2):421 – 430, 1992. ISSN 0022-2852. doi: http://dx.doi.org/10.1016/0022-2852(92)90243-H. URL http://www.sciencedirect.com/science/article/pii/002228529290243H.

[152] G. von Helden, I. Compagnon, M. N. Blom, M. Frankowski, U. Erlekam, J. Oomens, B. Brauer, R. B. Gerber, and G. Meijer. Mid-ir spectra of different conformers of phenylalanine in the gas phase. *Phys. Chem. Chem. Phys.*, 10: 1248–1256, 2008. doi: 10.1039/B713274C. URL http://dx.doi.org/10.1039/B713274C.

[153] A. Kaczor, I. D. Reva, L. M. Proniewicz, and R. Fausto. Importance of entropy in the conformational equilibrium of phenylalanine:a matrix-isolation infrared spectroscopy and density functional theory study. *The Journal of Physical Chemistry A*, 110(7):2360–2370, 2006. doi: 10.1021/jp0550715. URL http://pubs.acs.org/doi/abs/10.1021/jp0550715. PMID: 16480295.

[154] Uppula Purushotham, Dolly Vijay, and G. Narahari Sastry. A computational investigation and the conformational analysis of dimers, anions, cations, and zwitterions of l-phenylalanine. *Journal of Computational Chemistry*, 33(1):44–59, 2012. ISSN 1096-987X. doi: 10.1002/jcc.21942. URL http://dx.doi.org/10.1002/jcc.21942.

[155] Yan Zhao, Nathan E. Schultz, and Donald G. Truhlar. Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions. *Journal of Chemical Theory and Computation*, 2(2):364–382, 2006. doi: 10.1021/ct0502763. URL http://pubs.acs.org/doi/abs/10.1021/ct0502763.

[156] Zhijian Huang, Wenbo Yu, and Zijing Lin. Exploration of the full conformational landscapes of gaseous aromatic amino acid phenylalanine: An ab initio study. *Journal of Molecular Structure: {THEOCHEM}*, 758(23):195 – 202, 2006. ISSN 0166-1280. doi: http://dx.doi.org/10.1016/j.theochem.2005.10.043. URL http://www.sciencedirect.com/science/article/pii/S0166128005008158.

[157] Friedrich Biegler-Knig and Jens Schnbohm. Update of the aim2000-program for atoms in molecules. *Journal of Computational Chemistry*, 23(15):1489–1494, 2002. ISSN 1096-987X. doi: 10.1002/jcc.10085. URL http://dx.doi.org/10.1002/jcc.10085.

[158] Kang Taek Lee, Jiha Sung, Kwang Jun Lee, Seong Keun Kim, and Young Dong Park. Conformation-dependent ionization of l-phenylalanine: structures and energetics of cationic conformers. *Chemical Physics Letters*, 368(34):262 – 268, 2003. ISSN 0009-2614. doi: 10.1016/S0009-2614(02)01850-X. URL http://www.sciencedirect.com/science/article/pii/S000926140201850X.

[159] K. Y. Baek, M. Hayashi, Y. Fujimura, S. H. Lin, and S. K. Kim. Investigation of conformation-dependent properties of l-phenylalanine in neutral and radical cations by using a density functional taking into account noncovalent interactions. *The Journal of Physical Chemistry A*, 114(28):7583–7589, 2010. doi: 10.1021/jp1024153. URL http://pubs.acs.org/doi/abs/10.1021/jp1024153.

[160] Maciej Haranczyk and Sanliang Ling. Got. gaussian output tools. *Available at:*. URL http://gaussot.sf.net.

[161] G. Schaftenaar and Jan H. Noordik. Molden: a pre- and post-processing program for molecular and electronic structures. *Journal of Computer-Aided Molecular Design*, 14(2):123–134, 2000.

[162] G. Murray. Rotation about an arbitrary axis in 3 dimensions., May 2011. URL http://inside.mines.edu/~gmurray/ArbitraryAxisRotation.

[163] R. Ditchfield, W. J. Hehre, and J. A. Pople. Self-consistent molecular-orbital methods. ix. an extended gaussian-type basis for molecular-orbital studies of organic molecules. *The Journal of Chemical Physics*, 54(2):724–728, 1971. doi: 10.1063/1.1674902. URL http://link.aip.org/link/?JCP/54/724/1.

[164] Janusz Rak, Piotr Skurski, and Maciej Gutowski. An ab initio study of the betaine anion–dipole-bound anionic state of a model zwitterion system. *The Journal of Chemical Physics*, 114(24):10673–10681, 2001. doi: 10.1063/1.1373689. URL http://link.aip.org/link/?JCP/114/10673/1.

[165] Maciej Gutowski and Jack Simons. Double-rydberg anions: Ground-state electronic and geometric stabilities. *The Journal of Chemical Physics*, 93(6):3874–3880, 1990. doi: 10.1063/1.458773. URL http://link.aip.org/link/?JCP/93/3874/1.

[166] Istvn Hargittai. Ronald j. gillespie; the vsepr model; and molecular symmetry. *Structural Chemistry*, 20(2):155–159, 2009. ISSN 1040-0400. doi: 10.1007/s11224-009-9439-7. URL http://dx.doi.org/10.1007/s11224-009-9439-7.

[167] W. Humphrey, A. Dalke, and K. Schulten. Vmd: Visual molecular dynamics. *J Mol Graph*, 14(1):33–8, 27–8, Feb 1996.

[168] Maciej Haranczyk and Maciej Gutowski. Visualization of molecular orbitals and the related electron densities. *Journal of Chemical Theory and Computation*, 4(5):689–693, 2008. doi: 10.1021/ct800043a. URL http://pubs.acs.org/doi/abs/10.1021/ct800043a.

[169] Jenn Huei Lii and Norman L. Allinger. Molecular mechanics. the mm3 force field for hydrocarbons. 2. vibrational frequencies and thermodynamics. *Journal of the American Chemical Society*, 111(23):8566–8575, 1989. doi: 10.1021/ja00205a002. URL http://pubs.acs.org/doi/abs/10.1021/ja00205a002.

[170] Jenn Huei Lii and Norman L. Allinger. Molecular mechanics. the mm3 force field for hydrocarbons. 3. the van der waals' potentials and crystal data for aliphatic and aromatic hydrocarbons. *Journal of the American Chemical Society*, 111(23):8576–8582, 1989. doi: 10.1021/ja00205a003. URL http://pubs.acs.org/doi/abs/10.1021/ja00205a003.

[171] NIST. Chemistry webbook. standard reference database number 69. URL http://webbook.nist.gov/chemistry/.

[172] Rafal A. Bachorz, Wim Klopper, Maciej Gutowski, Xiang Li, and Kit H. Bowen. Photoelectron spectrum of valence anions of uracil and first-principles calculations of excess electron binding energies. *The Journal of Chemical Physics*, 129(5):054309, 2008. doi: 10.1063/1.2965128. URL http://link.aip.org/link/?JCP/129/054309/1.

[173] Jens Z. Pedersen and Alessandro Finazzi-Agró. Protein-radical enzymes. {*FEBS*} *Letters*, 325(12):53 – 58, 1993. ISSN 0014-5793. doi: http://dx.doi.org/10.

1016/0014-5793(93)81412-S. URL http://www.sciencedirect.com/science/article/pii/001457939381412S.

[174] Steven Y Reece, Justin M Hodgkiss, JoAnne Stubbe, and Daniel G Nocera. Proton-coupled electron transfer: the mechanistic underpinning for radical transport and catalysis in biology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1472):1351–1364, 2006. doi: 10.1098/rstb.2006. 1874. URL http://rstb.royalsocietypublishing.org/content/361/1472/1351.abstract.

[175] JoAnne Stubbe. Radicals with a controlled lifestyle. *Chem. Commun.*, pages 2511–2513, 2003. doi: 10.1039/B307617M. URL http://dx.doi.org/10.1039/B307617M.

[176] Washington Mutatu, Karin L. Klettke, Clifton Foster, and Kevin D. Walker. Unusual mechanism for an aminomutase rearrangement: retention of configuration at the migration termini. *Biochemistry*, 46(34):9785–9794, 2007. doi: 10.1021/bi700691p. URL http://pubs.acs.org/doi/abs/10.1021/bi700691p. PMID: 17676876.

[177] Markus Nett, Tobias A. M. Gulder, Andrew J. Kale, Chambers C. Hughes, and Bradley S. Moore. Function-oriented biosynthesis of -lactone proteasome inhibitors in salinispora tropica. *Journal of Medicinal Chemistry*, 52(19):6163–6167, 2009. doi: 10.1021/jm901098m. URL http://pubs.acs.org/doi/abs/10.1021/jm901098m. PMID: 19746976.

[178] Kevin Range, Idelisa Ayala, Darrin York, and Bridgette A. Barry. Normal modes of redox-active tyrosine: conformation dependence and comparison to experiment. *The Journal of Physical Chemistry B*, 110(22):10970–10981, 2006. doi: 10.1021/jp061503f. URL http://pubs.acs.org/doi/abs/10.1021/jp061503f. PMID: 16771350.

[179] B A Barry and G T Babcock. Tyrosine radicals are involved in the photosynthetic oxygen-evolving system. *Proceedings of the National Academy of Sciences*, 84(20):7099–7103, 1987. URL http://www.pnas.org/content/84/20/7099.abstract.

[180] K. H. Bowen and et al. *Private communication*, 2013.

[181] Zibo G Keolopile, Maciej Gutowski, and Maciej Haranczyk. Potential energy scanning tool (pesst), June 2012. URL http://sourceforge.net/projects/pesst/. Source Code.

[182] Zhixin Tian, Xue-Bin Wang, Lai-Sheng Wang, and Steven R. Kass. Are carboxyl groups the most acidic sites in amino acids? gas-phase acidities, photoelectron

spectra, and computations on tyrosine, p-hydroxybenzoic acid, and their conjugate bases. *Journal of the American Chemical Society*, 131(3):1174–1181, 2009. doi: 10.1021/ja807982k. URL http://pubs.acs.org/doi/abs/10.1021/ja807982k.

[183] Jang Sook Kwon, Chang Min Choi, Hwan Jin Kim, Nam Joon Kim, Joonkyung Jang, and Mino Yang. Combined theoretical modeling of photoexcitation spectrum of an isolated protonated tyrosine. *The Journal of Physical Chemistry A*, 113(12): 2715–2723, 2009. doi: 10.1021/jp809573a. URL http://pubs.acs.org/doi/abs/10.1021/jp809573a. PMID: 19245232.

[184] Fangwei Liu, Yigang Fang, Yun Chen, and Jianbo Liu. Reactions of deprotonated tyrosine and tryptophan with electronically excited singlet molecular oxygen (a1g): A guided-ion-beam scattering, statistical modeling, and trajectory study. *The Journal of Physical Chemistry B*, 116(22):6369–6379, 2012. doi: 10.1021/jp303022b. URL http://pubs.acs.org/doi/abs/10.1021/jp303022b.

[185] R. J. Hulsebosch, J. S. van den Brink, S. A. M. Nieuwenhuis, P. Gast, J. Raap, J. Lugtenburg, and A. J. Hoff. Electronic structure of the neutral tyrosine radical in frozen solution. selective 2h-, 13c-, and 17o-isotope labeling and epr spectroscopy at 9 and 35 ghz. *Journal of the American Chemical Society*, 119(37):8685–8694, 1997. doi: 10.1021/ja9707872. URL http://pubs.acs.org/doi/abs/10.1021/ja9707872.

[186] Anna Lisa Maniero, Vasile Chis, Alfonso Zoleo, Marina Brustolon, and Alberto Mezzetti. Three different tyrosyl radicals identified in l-tyrosine hcl crystals upon -irradiation: Magnetic characterization and temporal evolution. *The Journal of Physical Chemistry B*, 112(12):3812–3820, 2008. doi: 10.1021/jp710220u. URL http://pubs.acs.org/doi/abs/10.1021/jp710220u. PMID: 18318523.

[187] K. Y. Baek, Y. Fujimura, M. Hayashi, S. H. Lin, and S. K. Kim. Density functional theory study of conformation-dependent properties of neutral and radical cationic l-tyrosine and l-tryptophan. *The Journal of Physical Chemistry A*, 115(34):9658–9668, 2011. doi: 10.1021/jp200826z. URL http://pubs.acs.org/doi/abs/10.1021/jp200826z.

[188] Rodney J. Bartlett and Monika Musiał. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.*, 79:291–352, Feb 2007. doi: 10.1103/RevModPhys.79.291. URL http://link.aps.org/doi/10.1103/RevModPhys.79.291.

[189] Roy Dennington, Todd Keith, and John Millam. Gaussview Version 5. Semichem Inc. Shawnee Mission KS 2009.

[190] M. Gerhards, O. C. Thomas, J. M. Nilles, W.-J. Zheng, and Jr. Bowen, K. H. Cobalt-benzene cluster anions: Mass spectrometry and negative ion photoelectron spectroscopy. *Journal of Chemical Physics*, 116(23):10247–10252, June 2002.

[191] Jack Simons. Molecular anions. *The Journal of Physical Chemistry A*, 112(29): 6401–6511, 2008. doi: 10.1021/jp711490b. URL http://pubs.acs.org/doi/abs/10.1021/jp711490b.

[192] Anil Kumar and Michael D. Sevilla. Proton-coupled electron transfer in dna on formation of radiation-produced ion radicals. *Chemical Reviews*, 110(12):7002–7023, 2010. doi: 10.1021/cr100023g. URL http://pubs.acs.org/doi/abs/10.1021/cr100023g.

[193] Christopher J. Chang, Michelle C.Y. Chang, Niels H. Damrauer, and Daniel G. Nocera. Proton-coupled electron transfer: a unifying mechanism for biological charge transport, amino acid radical initiation and propagation, and bond making/breaking reactions of water and oxygen. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1655(0):13 – 28, 2004. ISSN 0005-2728. doi: http://dx.doi.org/10.1016/j.bbabio.2003.08.010. URL http://www.sciencedirect.com/science/article/pii/S0005272803002159. ¡ce:title¿Special issue dedicated to Jerry Babcock¡/ce:title¿.

[194] L. Sanche. Low energy electron-driven damage in biomolecules. *The European Physical Journal D - Atomic, Molecular, Optical and Plasma Physics*, 35(2):367–390, 2005. ISSN 1434-6060. doi: 10.1140/epjd/e2005-00206-6. URL http://dx.doi.org/10.1140/epjd/e2005-00206-6.

[195] M. A. McCord and M. J. Rooks. *SPIE Handbook of Microlithography, Micromachining and Microfabrication*, volume 1. Cornell NanoScale Science & Technology Facility,, 2000. URL http://www.cnf.cornell.edu/cnf_spietoc.html.

[196] Rafał A. Bachorz, Maciej Harańczyk, Iwona Dabkowska, Janusz Rak, and Maciej Gutowski. Anion of the formic acid dimer as a model for intermolecular proton transfer induced by a pi[sup *] excess electron. *The Journal of Chemical Physics*, 122(20):204304, 2005. doi: 10.1063/1.1899144. URL http://link.aip.org/link/?JCP/122/204304/1.

[197] Soren N. Eustis, Dunja Radisic, Kit H. Bowen, Rafa A. Bachorz, Maciej Haranczyk, Gregory K. Schenter, and Maciej Gutowski. Electron-driven acid-base chemistry: Proton transfer from hydrogen chloride to ammonia. *Science*, 319(5865):936–939, 2008. doi: 10.1126/science.1151614. URL http://www.sciencemag.org/content/319/5865/936.abstract.

[198] Michael Allan. Electron collisions with formic acid monomer and dimer. *Phys. Rev. Lett.*, 98:123201, Mar 2007. doi: 10.1103/PhysRevLett.98.123201. URL http://link.aps.org/doi/10.1103/PhysRevLett.98.123201.

[199] Helen K. Gerardi, Andrew F. DeBlase, Christopher M. Leavitt, Xiaoge Su, Kenneth D. Jordan, Anne B. McCoy, and Mark A. Johnson. Structural characterization of electron-induced proton transfer in the formic acid dimer anion, (hcooh)[sub 2][sup -], with vibrational and photoelectron spectroscopies. *The Journal of Chemical Physics*, 136(13):134318, 2012. doi: 10.1063/1.3693271. URL http://link.aip.org/link/?JCP/136/134318/1.

[200] Sarah T. Stokes, Andrej Grubisic, Xiang Li, Yeon Jae Ko, and Kit H. Bowen. Photoelectron spectroscopy of the parent anions of the nucleotides, adenosine-5[sup [prime]]-monophosphate and 2[sup [prime]]deoxyadenosine-5[sup [prime]]-monophosphate. *The Journal of Chemical Physics*, 128(4):044314, 2008. doi: 10.1063/1.2823001. URL http://link.aip.org/link/?JCP/128/044314/1.

[201] Monika Kobylecka, Jiande Gu, Janusz Rak, and Jerzy Leszczynski. Barrier-free proton transfer in the valence anion of 2-deoxyadenosine-5-monophosphate. ii. a computational study. *The Journal of Chemical Physics*, 128(4):044315, 2008. doi: http://dx.doi.org/10.1063/1.2823002. URL http://scitation.aip.org/content/aip/journal/jcp/128/4/10.1063/1.2823002.

[202] R. W. Hay and M. A. Bond. Kinetics of decarboxilation of acetoacetic acid. *Aust. J. Chem.*, 20(9):1823–1828, 1967. doi: doi:10.1071/CH9671823.

[203] Chun-Liang Huang, Chen-Chang Wu, and Min-Hsiung Lien. Ab initio studies of decarboxylations of the -keto carboxylic acids xcoch2cooh (x = h, oh, and ch3). *The Journal of Physical Chemistry A*, 101(42):7867–7873, 1997. doi: 10.1021/jp9712664. URL http://pubs.acs.org/doi/abs/10.1021/jp9712664.

[204] Kai Julius Pedersen. The ketonic decomposition of beta-keto carboxylic acids. *Journal of the American Chemical Society*, 51(7):2098–2107, 1929. doi: 10.1021/ja01382a016. URL http://pubs.acs.org/doi/abs/10.1021/ja01382a016.

[205] S. Hoz and A. J. Kresge. The enol of acetoacetic acid: A computational study of the relative stabilities of the ketone and carboxylic acid isomers. *Journal of Physical Organic Chemistry*, 10(3):182–186, 1997. ISSN 1099-1395. doi: 10.1002/(SICI) 1099-1395(199703)10:3⟨182::AID-POC873⟩3.0.CO;2-N. URL http://dx.doi.org/10.1002/(SICI)1099-1395(199703)10:3<182::AID-POC873>3.0.CO;2-N.

[206] George D. Purvis-III and Rodney J. Bartlett. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *The Journal of Chemical*

Physics, 76(4):1910–1918, 1982. doi: 10.1063/1.443164. URL http://link.aip.org/link/?JCP/76/1910/1.

[207] H. B. Schlegel. Optimization of equilibrium geometries and transition structures. *J. Comp. Chem.*, 3:214–218, 1982.

[208] H. B. Schlegel. Estimating the hessian for gradient-type geometry optimizations. *Theor. Chem. Acc.*, 66:333–340, 1984.

[209] J. Simons, P. Jørgensen, H. Taylor, and J. Ozment. Walking on potential energy surfaces. *J. Phys. Chem.*, 87:2745–2753, 1983.

[210] W. von Niessen, J. Schirmer, and L.S. Cederbaum. Computational methods for the one-particle green's function. *Computer Physics Reports*, 1(2):57 – 125, 1984. ISSN 0167-7977. doi: 10.1016/0167-7977(84)90002-9. URL http://www.sciencedirect.com/science/article/pii/0167797784900029.

[211] T Koopmans. ber die zuordnung von wellenfunktionen und eigenwerten zu den einzelnen elektronen eines atoms. *Physica*, 1(16):104 – 113, 1934. ISSN 0031-8914. doi: 10.1016/S0031-8914(34)90011-2. URL http://www.sciencedirect.com/science/article/pii/S0031891434900112.

[212] Maciej Haranczyk and Maciej Gutowski. Visualization of molecular orbitals and the related electron densities. *Journal of Chemical Theory and Computation*, 4(5): 689–693, 2008. doi: 10.1021/ct800043a. URL http://pubs.acs.org/doi/abs/10.1021/ct800043a.

[213] Robert C. Krueger. Crystalline acetoacetic acid. *Journal of the American Chemical Society*, 74(21):5536–5536, 1952. doi: 10.1021/ja01141a521. URL http://pubs.acs.org/doi/abs/10.1021/ja01141a521.

[214] J. V. Coe, J. T. Snodgrass, C. B. Freidhoff, K. M. McHugh, and K. H. Bowen. Photoelectron spectroscopy of the negative ion seo. *The Journal of Chemical Physics*, 84(2):618–625, 1986. doi: http://dx.doi.org/10.1063/1.450608. URL http://scitation.aip.org/content/aip/journal/jcp/84/2/10.1063/1.450608.

[215] D. M. Neumark, K. R. Lykke, T. Andersen, and W. C. Lineberger. Laser photodetachment measurement of the electron affinity of atomic oxygen. *Phys. Rev. A*, 32:1890–1892, Sep 1985. doi: 10.1103/PhysRevA.32.1890. URL http://link.aps.org/doi/10.1103/PhysRevA.32.1890.

[216] Joe Ho, Kent M. Ervin, and W. C. Lineberger. Photoelectron spectroscopy of metal cluster anions: Cun, agn, and aun. *The Journal of Chemical Physics*, 93

(10):6987–7002, 1990. doi: http://dx.doi.org/10.1063/1.459475. URL http://scitation.aip.org/content/aip/journal/jcp/93/10/10.1063/1.459475.

[217] M Allan. Measurement of absolute differential cross sections for vibrational excitation of o 2 by electron impact. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 28(23):5163, 1995. URL http://stacks.iop.org/0953-4075/28/i=23/a=021.

[218] Michael Allan. Measurement of the elastic and v = 0 1 differential electronn 2 cross sections over a wide angular range. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 38(20):3655, 2005. URL http://stacks.iop.org/0953-4075/38/i=20/a=003.

[219] A. Gopalan, J. Bömmels, S. Götte, A. Landwehr, K. Franz, M.-W. Ruf, H. Hotop, and K. Bartschat. A novel electron scattering apparatus combining a laser photoelectron source and a triply differentially pumped supersonic beam target: characterization and results for the resonance. *The European Physical Journal D - Atomic, Molecular, Optical and Plasma Physics*, 22(1):17–29, 2003. ISSN 1434-6060. doi: 10.1140/epjd/e2002-00219-7. URL http://dx.doi.org/10.1140/epjd/e2002-00219-7.

[220] R. K. Nesbet. Variational calculations of accurate e^--he cross sections below 19 ev. *Phys. Rev. A*, 20:58–70, Jul 1979. doi: 10.1103/PhysRevA.20.58. URL http://link.aps.org/doi/10.1103/PhysRevA.20.58.

[221] Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 1995. doi: 10.1021/ja00124a002. URL http://pubs.acs.org/doi/abs/10.1021/ja00124a002.

[222] C. Desfrancois, S. Carles, and J. P. Schermann. Weakly bound clusters of biological interest. *Chemical Reviews*, 100(11):3943–3962, 2000. doi: 10.1021/cr990061j.

[223] A. W. Castleman and K. H. Bowen. Clusters: structure, energetics, and dynamics of intermediate states of matter. *The Journal of Physical Chemistry*, 100(31):12911–12944, 1996. doi: 10.1021/jp961030k. URL http://pubs.acs.org/doi/abs/10.1021/jp961030k.

[224] Gennady L. Gutsev, Rodney J. Bartlett, and Robert N. Compton. Electron affinities of co[sub 2], ocs, and cs[sub 2]. *The Journal of Chemical Physics*, 108(16):6756–6762, 1998. doi: 10.1063/1.476091. URL http://link.aip.org/link/?JCP/108/6756/1.

[225] C. Desfrançois, H. Abdoul-Carime, N. Khelifa, and J. P. Schermann. From $\frac{1}{r}$ to $\frac{1}{r^2}$ potentials: Electron exchange between rydberg atoms and polar molecules, Oct 1994.

[226] Michael Allan. Study of resonances in formic acid by means of vibrational excitation by slow electrons. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 39(14):2939, 2006. URL http://stacks.iop.org/0953-4075/39/i=14/a=002.

[227] Mathew Ryder. *Quantum chemical characterisation of hydrogen bonded systems upon addition of an excess electron*. MChem report, Heriot-Watt University, 2013.

[228] Shamus A. Blair and Ajit J. Thakkar. How many intramolecular hydrogen bonds does the oxalic acid dimer have? *Chemical Physics Letters*, 495(46):198 – 202, 2010. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/j.cplett.2010.07.019. URL http://www.sciencedirect.com/science/article/pii/S000926141000936X.

[229] Gerhard Raabe. The use of quantum-chemical semiempirical methods to calculate the lattice energies of organic molecular crystals. part ii: The lattice energies of - and -oxalic acid (cooh)2. *Zeitschrift fur Naturforschung A*, 57:961–966, 2002.

[230] A. Mohajeri and N. Shakerin. The gas-phase acidity and intramolecular hydrogen bonding in oxalic acid. *Journal of Molecular Structure: {THEOCHEM}*, 711(13):167 – 172, 2004. ISSN 0166-1280. doi: http://dx.doi.org/10.1016/j.theochem.2004.10.002. URL http://www.sciencedirect.com/science/article/pii/S0166128004006736.

[231] Jee-Gong Chang, Hsin-Tsung Chen, Shucheng Xu, and M. C. Lin. Computational study on the kinetics and mechanisms for the unimolecular decomposition of formic and oxalic acids. *The Journal of Physical Chemistry A*, 111(29):6789–6797, 2007. doi: 10.1021/jp069036p. URL http://pubs.acs.org/doi/abs/10.1021/jp069036p.

[232] Amlan K. Roy, James R. Hart, and Ajit J. Thakkar. Clusters of glycolic acid and 16 water molecules. *Chemical Physics Letters*, 434(46):176 – 181, 2007. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/j.cplett.2006.12.010. URL http://www.sciencedirect.com/science/article/pii/S0009261406018033.

[233] H.B. Jansen and P. Ros. Non-empirical molecular orbital calculations on the protonation of carbon monoxide. *Chemical Physics Letters*, 3(3):140 – 143, 1969. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/0009-2614(69)80118-1. URL http://www.sciencedirect.com/science/article/pii/0009261469801181.

[234] S.F. Boys and F. Bernardi. The calculation of small molecular interactions by the differences of separate total energies. some procedures with reduced errors. *Molecular Physics*, 19(4):553–566, 1970. doi: 10.1080/00268977000101561. URL http://www.tandfonline.com/doi/abs/10.1080/00268977000101561.

[235] Sílvia Simon, Miquel Duran, and J. J. Dannenberg. How does basis set superposition error change the potential surfaces for hydrogen-bonded dimers? *The Journal of Chemical Physics*, 105(24):11024–11031, 1996. doi: 10.1063/1.472902. URL http://link.aip.org/link/?JCP/105/11024/1.

[236] Kevin E. Riley, Michal Pitonak, Petr Jurecka, and Pavel Hobza. Stabilization and structure calculations for noncovalent interactions in extended molecular systems based on wave function and density functional theories. *Chemical Reviews*, 110 (9):5023–5063, 2010. doi: 10.1021/cr1000173. URL http://pubs.acs.org/doi/abs/10.1021/cr1000173.

[237] Asger Halkier, Trygve Helgaker, Poul Jrgensen, Wim Klopper, and Jeppe Olsen. Basis-set convergence of the energy in molecular hartreefock calculations. *Chemical Physics Letters*, 302(56):437 – 446, 1999. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/S0009-2614(99)00179-7. URL http://www.sciencedirect.com/science/article/pii/S0009261499001797.

[238] PeterR. Taylor. Coupled-cluster methods in quantum chemistry. In BjrnO. Roos, editor, *Lecture Notes in Quantum Chemistry II*, volume 64 of *Lecture Notes in Chemistry*, pages 125–202. Springer Berlin Heidelberg, 1994. ISBN 978-3-540-58620-3. doi: 10.1007/978-3-642-57890-8_3. URL http://dx.doi.org/10.1007/978-3-642-57890-8_3.

[239] Maciej Gutowski, Grzegorz Chalasinski, and Jeanne Van Duijneveldt-Van De Rijdt. Effective basis sets for calculations of exchange-repulsion energy. *International Journal of Quantum Chemistry*, 26(6):971–982, 1984. ISSN 1097-461X. doi: 10.1002/qua.560260602. URL http://dx.doi.org/10.1002/qua.560260602.

[240] K. Aflatooni, G. A. Gallup, and P. D. Burrow. Electron attachment energies of the dna bases. *The Journal of Physical Chemistry A*, 102(31):6205–6207, 1998. doi: 10.1021/jp980865n. URL http://pubs.acs.org/doi/abs/10.1021/jp980865n.

[241] A. M. Scheer, P. Mozejko, G. A. Gallup, and P. D. Burrow. Total dissociative electron attachment cross sections of selected amino acids. *The Journal of Chemical Physics*, 126(17):174301, 2007. doi: 10.1063/1.2727460. URL http://link.aip.org/link/?JCP/126/174301/1.

[242] Maciej Haranczyk, Maciej Gutowski, Xiang Li, and Kit H Bowen. Adiabatically bound valence anions of guanine. *J Phys Chem B*, 111(51):14073–14076, Dec 2007. doi: 10.1021/jp077439z. URL http://dx.doi.org/10.1021/jp077439z.

[243] Maciej Haraczyk, Maciej Gutowski, Xiang Li, and Kit H Bowen. Bound anionic states of adenine. *Proc Natl Acad Sci U S A*, 104(12):4804–4807, Mar 2007. doi: 10.1073/pnas.0609982104. URL http://dx.doi.org/10.1073/pnas.0609982104.

[244] Monika Kobylecka, Jerzy Leszczynski, and Janusz Rak. Valence anion of thymine in the dna π-stack. *Journal of the American Chemical Society*, 130(46):15683–15687, 2008. doi: 10.1021/ja806251h. URL http://pubs.acs.org/doi/abs/10.1021/ja806251h. PMID: 18954049.

[245] Z. Nahlovska, B. Nahlovsky, and T. G. Strand. Molecular structure of gaseous oxalic acid from electron diffraction and ir data. *Acta Chemica Scandinavica*, 24:2617–2628, 1970.

[246] B.C. Stace and C. Oralratmanee. Infrared and raman spectra of gaseous oxalic acid. *Journal of Molecular Structure*, 18(2):339 – 342, 1973. ISSN 0022-2860. doi: http://dx.doi.org/10.1016/0022-2860(73)85238-X. URL http://www.sciencedirect.com/science/article/pii/002228607385238X.

[247] R.L. Redington and T.E. Redington. Infrared matrix-isolation spectra of monomeric oxalic acid. *Journal of Molecular Structure*, 48(2):165 – 176, 1978. ISSN 0022-2860. doi: http://dx.doi.org/10.1016/0022-2860(78)80019-2. URL http://www.sciencedirect.com/science/article/pii/0022286078800192.

[248] J. Nieminen, M. Rasanen, and J. Murto. Matrix-isolation and ab initio studies of oxalic acid. *The Journal of Physical Chemistry*, 96(13):5303–5308, 1992. doi: 10.1021/j100192a024. URL http://pubs.acs.org/doi/abs/10.1021/j100192a024.

[249] Ermelinda M. S. Maas, Rui Fausto, Mika Pettersson, Leonid Khriachtchev, and Markku Rsnen. Infrared-induced rotamerization of oxalic acid monomer in argon matrix. *The Journal of Physical Chemistry A*, 104(30):6956–6961, 2000. doi: 10.1021/jp000634s. URL http://pubs.acs.org/doi/abs/10.1021/jp000634s.

[250] R. A. Back. The ultraviolet absorption spectrum of oxalic acid vapor. *Canadian Journal of Chemistry*, 62(8):1414–1428, 1984. doi: 10.1139/v84-241. URL http://www.nrcresearchpress.com/doi/abs/10.1139/v84-241.

[251] Cheng Chen and Shuang-Fuh Shyu. Conformers and intramolecular hydrogen bonding of the oxalic acid monomer and its anions. *International Journal of Quantum Chemistry*, 76(4):541–551, 2000. ISSN 1097-461X. doi: 10.1002/(SICI) 1097-461X(2000)76:4⟨541::AID-QUA5⟩3.0.CO;2-U. URL http://dx.doi.org/10.1002/(SICI)1097-461X(2000)76:4<541::AID-QUA5>3.0.CO;2-U.

[252] C. Van Alsenoy, V.J. Klimkowski, and Lothar Schfer. Ab initio studies of structural features not easily amenable to experiment: Part 37. structural and conformational investigations of the dicarbonyls glyoxal, biacetyl and oxalic acid. *Journal of Molecular Structure: {THEOCHEM}*, 109(34):321 – 330, 1984. ISSN 0166-1280. doi: http://dx.doi.org/10.1016/0166-1280(84)80016-0. URL http://www.sciencedirect.com/science/article/pii/0166128084800160.

[253] Giuseppe Buemi. Dft study of the hydrogen bond strength and ir spectra of formic, oxalic, glyoxylic and pyruvic acids in vacuum, acetone and water solution. *Journal of Physical Organic Chemistry*, 22(10):933–947, 2009. ISSN 1099-1395. doi: 10.1002/poc.1543. URL http://dx.doi.org/10.1002/poc.1543.

[254] Peter D. Godfrey, Mathew J. Mirabella, and Ronald D. Brown. Structural studies of higher energy conformers by millimeter-wave spectroscopy: oxalic acid. *The Journal of Physical Chemistry A*, 104(2):258–264, 2000. doi: 10.1021/jp992499t. URL http://pubs.acs.org/doi/abs/10.1021/jp992499t.

[255] M. Gutowski, I. Dabkowska, J. Rak, S. Xu, J.M. Nilles, D. Radisic, and K.H. Bowen Jr. Barrier-free intermolecular proton transfer in the uracil-glycine complex induced by excess electron attachment. *The European Physical Journal D - Atomic, Molecular, Optical and Plasma Physics*, 20(3):431–439, 2002. ISSN 1434-6060. doi: 10.1140/epjd/e2002-00168-1. URL http://dx.doi.org/10.1140/epjd/e2002-00168-1.

[256] Kamil Mazurkiewicz, Maciej Haraczyk, Maciej Gutowski, Janusz Rak, Dunja Radisic, Soren N Eustis, Di Wang, and Kit H Bowen. Valence anions in complexes of adenine and 9-methyladenine with formic acid: stabilization by intermolecular proton transfer. *J Am Chem Soc*, 129(5):1216–1224, Feb 2007. doi: 10.1021/ja066229h. URL http://dx.doi.org/10.1021/ja066229h.

[257] Hans-Joachim Werner, Peter J. Knowles, Gerald Knizia, Frederick R. Manby, and Martin Schtz. Molpro: a general-purpose quantum chemistry program package.

Wiley Interdisciplinary Reviews: Computational Molecular Science, 2(2):242–253, 2012. ISSN 1759-0884. doi: 10.1002/wcms.82. URL http://dx.doi.org/10.1002/wcms.82.

[258] Hans-Joachim Werner, Peter J. Knowles, Gerald Knizia, Frederick R. Manby, Martin Schütz, et al. Molpro, version 2012.1, a package of ab initio programs. URL http://www.molpro.net.

[259] E.V. Doktorov, I.A. Malkin, and V.I. Man'ko. Dynamical symmetry of vibronic transitions in polyatomic molecules and the franck-condon principle. *Journal of Molecular Spectroscopy*, 56(1):1 – 20, 1975. ISSN 0022-2852. doi: http://dx.doi.org/10.1016/0022-2852(75)90199-X. URL http://www.sciencedirect.com/science/article/pii/002228527590199X.

[260] Dong-Sheng Yang, Marek Z. Zgierski, David M. Rayner, Peter A. Hackett, Ana Martinez, Dennis R. Salahub, Pierre-Nicholas Roy, and Jr. Tucker Carrington. The structure of nb[sub 3]o and nb[sub 3]o[sup +] determined by pulsed field ionization–zero electron kinetic energy photoelectron spectroscopy and density functional theory. *The Journal of Chemical Physics*, 103(13):5335–5342, 1995. doi: 10.1063/1.470568. URL http://link.aip.org/link/?JCP/103/5335/1.

[261] Sharon Hammes-Schiffer. Introduction: Proton-coupled electron transfer. *Chemical Reviews*, 110(12):6937–6938, 2010. doi: 10.1021/cr100367q. URL http://pubs.acs.org/doi/abs/10.1021/cr100367q.

[262] Sharon Hammes-Schiffer and Alexei A. Stuchebrukhov. Theory of coupled electron and proton transfer reactions. *Chemical Reviews*, 110(12):6939–6960, 2010. doi: 10.1021/cr1001436. URL http://pubs.acs.org/doi/abs/10.1021/cr1001436.

[263] Andrew Sirjoosingh and Sharon Hammes-Schiffer. Proton-coupled electron transfer versus hydrogen atom transfer: Generation of charge-localized diabatic states. *The Journal of Physical Chemistry A*, 115(11):2367–2377, 2011. doi: 10.1021/jp111210c. URL http://pubs.acs.org/doi/abs/10.1021/jp111210c.

[264] Paul F. Barbara, Thomas J. Meyer, and Mark A. Ratner. Contemporary issues in electron transfer research. *The Journal of Physical Chemistry*, 100(31):13148–13168, 1996. doi: 10.1021/jp9605663. URL http://pubs.acs.org/doi/abs/10.1021/jp9605663.

[265] Fang Chen, Joshua Hihath, Zhifeng Huang, Xiulan Li, and N.J. Tao. Measurement of single-molecule conductance. *Annual Review of Physical Chemistry*, 58(1):535–564, 2007. doi: 10.1146/annurev.physchem.58.032806.

104523. URL http://www.annualreviews.org/doi/abs/10.1146/annurev.physchem.58.032806.104523. PMID: 17134372.

[266] Jay W Ponder and David A Case. Force fields for protein simulations. *Adv Protein Chem*, 66:27–85, 2003.

[267] Janusz Rak, Piotr Skurski, Jack Simons, and Maciej Gutowski. Low-energy tautomers and conformers of neutral and protonated arginine. *Journal of the American Chemical Society*, 123(47):11695–11707, 2001. doi: 10.1021/ja011357l. URL http://pubs.acs.org/doi/abs/10.1021/ja011357l.

[268] J. E. Straub and D. Thirumalai. Theoretical probes of conformational fluctuations in s-peptide and rnase a/3'-ump enzyme product complex. *Proteins:Struct., Funct., Bioinf.*, 15:360–373, 1993.

[269] J. E. Straub and D. Thirumalai. Exploring the energy landscape in proteins. *Proceding of National Academy of Science USA*, 90:809–813, 1993.

[270] Stephan Irle, Guishan Zheng, Marcus Elstner, and Keiji Morokuma. Formation of fullerene molecules from carbon nanotubes: a quantum chemical molecular dynamics study. *Nano Letters*, 3(4):465–470, 2003. doi: 10.1021/nl034023y. URL http://pubs.acs.org/doi/abs/10.1021/nl034023y.

[271] John S. Tse and Dennis D. Klug. Structure and dynamics of liquid sulphur. *Phys. Rev. B*, 59:34–37, Jan 1999. doi: 10.1103/PhysRevB.59.34. URL http://link.aps.org/doi/10.1103/PhysRevB.59.34.

[272] W. J. Hehre, R. Ditchfield, and J. A. Pople. Self—consistent molecular orbital methods. xii. further extensions of gaussian—type basis sets for use in molecular orbital studies of organic molecules. *The Journal of Chemical Physics*, 56(5):2257–2261, 1972. doi: 10.1063/1.1677527. URL http://link.aip.org/link/?JCP/56/2257/1.

[273] J. W. Ponder. Tinker: Software for molecular design, v5.1. *Saint Louis, MO*, 2003.

[274] Kenichi Fukui. The path of chemical reactions - the irc approach. *Accounts of Chemical Research*, 14(12):363–368, 1981. doi: 10.1021/ar00072a001. URL http://pubs.acs.org/doi/abs/10.1021/ar00072a001.