Density Functional Theory: Dispersion Interactions & Biological Applications

by

Alya A. Arabi

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
August 2012

DALHOUSIE UNIVERSITY

DEPARTMENT OF CHEMISTRY

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled “Density Functional Theory: Dispersion Interactions & Biological Applications” by Alya A. Arabi in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Dated: August 14, 2012

External Examiner: _____

Research Co-Supervisors: _____

_____

Examining Committee: _____

_____

Departmental Representative: _____

DALHOUSIE UNIVERSITY

DATE: August 14, 2012

AUTHOR: Alya A. Arabi

TITLE: Density Functional Theory: Dispersion Interactions & Biological Applications

DEPARTMENT OR SCHOOL: Department of Chemistry

DEGREE: PhD CONVOCATION: October YEAR: 2012

_____

Signature of Author

Dedication

This thesis is dedicated to my admirable family:

My beloved father Ali O. Arabi & my darling mother Nida H. El-Badaoui,

My dear siblings: Alaa A. Arabi, Ghayda A. Arabi and Judy A. Arabi, and

My lovely husband Khaled O. Shahin.

"What we have learned is like a handful of earth,
what we have yet to learn is like the whole world."

AVVAYIAR

Table of Contents

List of Tables

List of Figures

Abstract

London or dispersion interactions are weak van der Waals (vdW) interactions. They are important in determining the structure and properties of many chemical and biochemical systems. In this thesis, an optimizer using the nonempirical generalized gradient approximation (GGA) functional PW86+PBE+XDM, to capture van der Waals interactions, is presented. The work in this thesis covers the assessment of a variety of basis sets for their ability to reproduce accurate GGA repulsive and binding energies. Selected basis sets were then used to compute binding energies of 65 vdW complexes at equilibrium. This functional was also tested for binding energies of two sets of vdW complexes at distorted geometries. The last part deals with forces to investigate their accuracy using PW86+PBE+XDM in order to build an optimizer for vdW complexes using a nonempirical DFT method. Eventually, after confirming a high reproducibility of the optimizer on the geometries and binding energies, it was used in two biologically relevant applications. This optimizer is a unique tool to compute deformation energies with a nonempirical DFT method.

The second part of this thesis covers a biologically relevant application where a conventional DFT is used. This application is related to the carrier of the genetic codes in living cells, DNA. DNA undergoes harmful mutations under external perturbations such as applied external electric fields. In this study, DNA base pairs were first mimicked by a simpler model, namely, the formic acid dimer. The effect of applied external electric fields on the geometries of the formic acid dimer is studied. The effect of these applied fields on the potential energy surface, the barrier height and the frequency of the double proton transfer in the formic acid dimer are also investigated. The study was then repeated on DNA base pairs to study the effect of an external applied electric field on the tunneling corrected rate constants of the double proton transfer reactions in AT and GC.

List of Abbreviations Used

BR Model	Becke-Roussel model
BSSE	Basis set superposition error
CGTO	Contracted Gaussian type orbital
CI	Configuration interaction
cc	Correlation consistent
CC	Coupled cluster
CP	Counterpoise
DFT	Density functional theory
DZ	Double zeta
GGA	Generalized gradient approximation
GTO	Gaussian type orbital
HF	Hartree-Fock
KS	Kohn-Sham
LCAO	Linear Combination of atomic orbitals
LDA	Local density approximation
LSDA	Local spin-density approximation
MAE	Mean absolute error
MAPE	Mean absolute percent error
Max	Maximum error
ME	Mean error
Min	Minimum error
MPn	n^{th} order Møller-Plesset perturbation theory
pc	Polarization consistent
STO	Slater-type orbital
TZ	Triple zeta
UHF	Unrestricted Hartree-Fock
vdW	van der Waals
XDM	Exchange-hole dipole moment

Acknowledgments

The work presented in this thesis could not have been completed without the help and the continuous support of my supervisors Professors Axel D. Becke and Cherif F. Matta. I am greatly thankful for the invaluable scientific and intellectual insight they provide me with, their pedagogical suggestions and advices.

Thanks for the committee members Professors Russell J. Boyd and Donald F. Weaver for their useful comments during the committee meetings.

Thanks for my colleague, Felix O. Kannemann who collaborated in the study presented in Chapter 4 by providing me with the input files of the 65 complexes used in this study. Thanks, also, for the many discussions we had about research.

I would like to thank Dr. Erin R. Johnson for allowing me to use the figure of the holes from her PhD thesis and for providing me with the input geometries of the applications presented in Chapter 8.

I am grateful for the Natural Sciences and Engineering Research Council of Canada, Killam Trusts and Dalhousie University for their financial support.

Thanks to the Atlantic Computational Excellence Network for the powerful computer resources they provided me with.

Finally, I wish to thank my family for their continuous and endless support.

PART I

Chapter 1

Introduction

The purpose of this chapter is to introduce the electronic structure theories used in the research presented in this thesis. The main focus is on density functional theory (DFT). A brief overview of Møller-Plesset Perturbation Theory (MPn) is provided at the beginning of this chapter as MP2 was used for method-comparison purposes in the second part of this thesis.

Conventional wavefunction methods such as Hartree-Fock (HF), n^{th} order Møller-Plesset Perturbation Theory (MPn), Configuration Interaction (CI), Coupled Cluster (CC), etc. are quantum theories used to predict chemical structures and their properties.

HF theory is the basis of the wavefunction electronic structure methods. It takes into account correlations between parallel-spin electrons only, i.e., it gives the exchange energy. MPn, CI and CC are "correlated" wavefunction methods. These sophisticated methods use HF theory as a starting point. They then evaluate an energy-correction term known as the correlation energy. Correlation energy can be obtained by taking into account correlation between opposite-spin electrons. The sum of exchange and correlation energies gives the total "exchange-correlation" energy.

1.1 Møller-Plesset Perturbation Theory

Møller-Plesset perturbation theory is a correlated wavefunction method proposed by C. Møller and M. S. Plesset in 1934. MPn is a perturbation treatment of atoms and molecules. It perturbs (excites) electrons from occupied to "*virtual*" (unoccupied) orbitals. In MPn, perturbations are applied to a reference single Slater determinant which is an antisymmetric product of HF spin-orbitals [1, 2]. The reference wavefunction, i.e., the ground-state HF function, is an unperturbed zero order wavefunction. The Hamiltonian used to obtain the reference HF wavefunction is H^0. H^0 is a sum over one electron Fock operators f_i given by

$$f(i) = -\frac{1}{2} \bigtriangledown_i^2 - \sum_{A=1}^{M} \frac{Z_A}{\vec{r}_{iA}} + v^{HF}(i), \tag{1.1.1}$$

where Z_A is the atomic number of nucleus A and $\vec{r}_{iA}$ is the distance between the ith electron and A^{th} nucleus. In (1.1.1), the first term is the kinetic energy, the second term is the classical electron-nuclei Coulomb interaction energy, and the third term $v^{HF}(i)$ is the *average* potential experienced by the i^{th} electron due to the other electrons in the system [1]. MP2 corrects for the HF energy by adding a second order energy correction term,

$$E_0^{(2)} = \sum_{s \neq 0} \frac{|\left\langle \psi_s^{(0)} \mid H' \mid \Phi_0 \right\rangle|^2}{E_0^{(0)} - E_s^{(0)}}, \tag{1.1.2}$$

where Φ_0 is the zeroth order HF reference function, $\psi_s^{(0)}$ are excited-state Slater determinants of H^0, and H' is the difference between the Hamiltonian of the reference wavefunction H^0 and the Hamiltonian of the true wavefunction. "s" denotes how many excitations are introduced and from which occupied orbitals (i, j, k, etc.) to which virtual orbitals (a, b, c, etc.). For example, the doubly excited determinant Φ_{ij}^{ab} is built from the ground state reference determinant by replacing the two occupied spin orbitals u_i and u_j with the two virtual spin orbitals u_a and u_b. Szabo and Ostlund

explain in details that, for single excitations, the expectation value $\left\langle \Phi_i^a \mid H' \mid \Phi_0 \right\rangle$ in (1.1.2) vanishes. This is known as Brillouin's theorem. Similarly, this expectation value is zero in case of triple or higher excitations (see [1] for more details).

In general, the correlated wavefunction methods result in reliable energies. However, they are too expensive to be practical for large chemical and biochemical systems. Density functional theory, born in 1964, is a quantum mechanical approach used to solve electronic structure problems. DFT is gaining popularity over "conventional" wavefunction methods because of its good performance at an economical computational cost.

1.2 Density Functional Theory

The approach used in DFT to solve the Schrödinger equation is less computationally intense than the approach undertaken by the conventional wavefunction methods. DFT uses the electron density $\rho(\vec{r})$ which depends only on three variables. Conventional methods use the wavefunction Ψ which depends on 3N variables, where N is the number of electrons in the system.

The one-electron density is the probability of finding any of the N electrons of a system (regardless of spin) in a volume element $\mathrm{d}\vec{r}_1$ while the other N-1 electrons are elsewhere. The two-electron density, or pair density, is invaluable for the description of electron correlation because it is the probability of finding a pair of electrons (spin up or down) simultaneously in the volume elements $\mathrm{d}\vec{r}_1$ $\mathrm{d}\vec{r}_2$ while the other N-2 electrons are elsewhere.

1.3 The First Density Functional Models

Well before 1964 (when DFT was rigorously established), in 1927, Thomas [3] and Fermi proposed independently the first model that attempts to obtain information about an atom or a molecule from the electron density $\rho(\vec{r})$, instead of the wave-function Ψ [4]. The Thomas-Fermi model was the first density functional for the energy. It is based on a uniform electron gas. This functional treats only the kinetic energy quantum mechanically. Nuclear-electron and electron-electron interactions are treated using classical electrostatics. Exchange and correlation energies are neglected. The Thomas-Fermi model is a poor approximation as it fails to reproduce atomic shell structure. In 1928, Dirac introduced an expression for the exchange energy based on the uniform electron gas [5, 6]. Adding this exchange energy expression to the Thomas-Fermi model gave rise to *Thomas-Fermi-Dirac* model. In 1951, Slater developed the exchange functional "X_α" [7] which originally aimed to simplify the exchange energy in HF theory. The major breakthrough in DFT was in 1964 when Hohenberg and Kohn published their theorems [8].

1.4 The Hohenberg-Kohn Theorems

The first Hohenberg-Kohn theorem is a proof of existence. It states that for each different external potential V_{ext}, there exists one and only one corresponding non-degenerate ground state electron density ρ_0. Vice versa, any ground state electron density is uniquely mapped to a single external potential and uniquely determines the number of electrons N. Therefore ρ_0 uniquely defines all other properties as well:

$$\rho_0 \implies V_{ext} \implies \hat{H} \implies \Psi_0 \implies everything, \tag{1.4.1}$$

where $\hat{H}$ is the N-electron Hamiltonian operator with external potential V_{ext}.

The second Hohenberg-Kohn theorem is a variational principle. It states that the minimal energy of a chemical system is always *necessarily* the energy E resulting from the ground state electron density ρ_0. Any other electron density ρ' gives a higher energy E_0,

$$E(\rho') \geq E_0(\rho_0). \tag{1.4.2}$$

1.5 Kohn-Sham Formalism

One year later, in 1965, Kohn and Sham developed a new approach [9] to solve the the Schrödinger N-electron equation using the "Kohn-Sham" (KS) Orbitals. The sum of the squares of these orbitals, by definition, is equal to the exact ground state density ρ:

$$\rho = 2\sum_{i=1}^{N/2} \psi_i^2 \tag{1.5.1}$$

and each KS orbital satisfies

$$-\frac{1}{2} \bigtriangledown^2 \psi_i + v_{KS}\psi_i = \epsilon_i \psi_i, \tag{1.5.2}$$

where v_{KS} is the potential that delivers, by assumption, the exact ground state ρ through (1.5.1). The energies (eigenvalues) of the KS orbitals do not have the same physical meaning as the energies of HF orbitals. Unless the eigenvalue is that of the highest occupied orbital, the energies of KS orbitals do not correspond to ionization potentials as in Koopmans theorem for HF orbitals [10, 11].

In Kohn-Sham theory, the total energy is written in the form,

$$E_{tot}[\rho] = T_{KS}[\rho] + \int v_{nuc}\rho + J[\rho] + E_{XC}[\rho] \tag{1.5.3}$$

where:

1. $T_{KS}[\rho]$ is the kinetic energy of the Kohn-Sham orbitals[1]:

$$T_{KS}[\rho] = -\frac{1}{2}\sum_{i}^{N/2} 2 \int \psi_i \bigtriangledown^2 \psi_i d\vec{r}_i. \tag{1.5.4}$$

2. $\int v_{nuc}\rho$ is the nuclear-electron potential energy, where

$$v_{nuc} = -\sum_{A=1}^{M} \frac{Z_A}{\mid \vec{r}_i - \vec{R}_A \mid}, \tag{1.5.5}$$

   where Z_A is the atomic number of nucleus A, $\vec{r}_i$ is position vector of electron i, and $\vec{R}_A$ is the position vector of nucleus A.

3. $J[\rho]$ is the classical Coulomb electron-electron repulsion energy:

$$J[\rho] = \frac{1}{2}\int\int \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2. \tag{1.5.6}$$

4. $E_{XC}[\rho]$ is the "exchange-correlation" energy. E_{XC} is relatively much smaller than the other three terms. This term includes everything else that has not been accounted for in the first three terms. In other words, in addition to "exchange" and "correlation" effects, the exchange-correlation functional accounts for the difference between the "real" and the KS kinetic energies. This complicated functional has no explicit form. It has to be approximated, as will be discussed shortly.

Variation of the total energy (1.5.3) with respect to the KS orbitals, subject to orthonormality constraints and (1.5.1), gives (1.5.2) with

$$v_{KS} = v_{nuc} + v_{elec} + v_{XC}, \tag{1.5.7}$$

where

[1]Note that this is not the kinetic energy of the "real" system.

1. v_{nuc} is the nuclear potential given by (1.5.5).

2. v_{elec} is the Coulomb potential arising from the electrons. v_{elec} is given by

$$\int \frac{\rho(\vec{r}_2)}{r_{12}} d\vec{r}_2. \tag{1.5.8}$$

3. the exchange correlation potential v_{XC} is given by

$$v_{XC} = \frac{\delta E_{XC}}{\delta \rho}, \tag{1.5.9}$$

   where $\frac{\delta}{\delta \rho}$ denotes the "functional" derivative of E_{XC} with respect to the density ρ [4].

The Kohn-Sham formalism solves the Schrödinger equation iteratively in a self-consistent field (SCF) fashion. SCF is needed because the Coulomb potential (1.5.8) and the exchange-correlation potential (1.5.9) are functionals of the density which itself depends on the unknown orbitals.

In (1.5.3), the KS kinetic energy, the nuclear-electron potential energy and the Coulomb repulsion can be computed exactly because the KS orbitals define the exact ground state density[2]. The only term that cannot be computed exactly is E_{XC}. It has to be approximated. In theory, if the exchange-correlation energy and its functional derivative were known exactly, the Schrödinger equation would be solved *exactly*. Therefore, in DFT, the electronic structure theory problem reduces to finding approximate but accurate expressions for the E_{XC} term in (1.5.3). E_{XC} can be written in terms of an exchange-correlation "hole" h_{XC} as follows [12]:

$$E_{XC} = \frac{1}{2} \int \int \frac{\rho(\vec{r}_1)}{r_{12}} h_{XC}(\vec{r}_1, \vec{r}_2) d\vec{r}_1 d\vec{r}_2. \tag{1.5.10}$$

[2]In HF the nuclear-electron potential energy and the classical Coulomb electron-electron repulsion energy cannot be computed exactly because the HF orbitals give the ground state density of HF and not the real ground state density.

The exchange-correlation hole is a fundamental quantity in DFT. It is a measure of the depletion in density around a reference electron $\vec{r}_1$ due to exchange *and* correlation effects. The exchange-correlation hole is thus a combination of two separate holes: the exchange hole h_X and the correlation hole h_C. Accurate modeling of holes results in accurate models of E_{XC}. The exchange and correlation holes will be discussed along with their properties before summarizing the approaches used to approximate E_{XC}.

1.6 Exchange and Correlation Holes

1.6.1 The Fermi Hole

The Fermi or exchange hole is more important than the Coulomb hole. The exchange hole for electrons of spin σ is given by

$$h_{X\sigma}(\vec{r}_1, \vec{r}_2) = -\frac{1}{\rho_\sigma(\vec{r}_1)} \sum_{ij} \psi_{i\sigma}(\vec{r}_1)\psi_{j\sigma}(\vec{r}_1)\psi_{i\sigma}(\vec{r}_2)\psi_{j\sigma}(\vec{r}_2). \tag{1.6.1}$$

The exchange hole is also be given by a more interpretive format as follows:

$$h_{X\sigma}(\vec{r}_1, \vec{r}_2) = -\frac{|\ \rho_\sigma(\vec{r}_1, \vec{r}_2)\ |^2}{\rho_\sigma(\vec{r}_1)}, \tag{1.6.2}$$

where

$$\rho_\sigma(\vec{r}_1, \vec{r}_2) = \sum_i \psi_{i\sigma}(\vec{r}_1)\psi_{i\sigma}(\vec{r}_2). \tag{1.6.3}$$

It is clear from (1.6.2) that the exchange hole is negative everywhere. This exchange hole has two important properties:

$$\int h_{X\sigma}(\vec{r}_1, \vec{r}_2) d\vec{r}_2 = -1, \tag{1.6.4}$$

and at $\vec{r}_2 = \vec{r}_1$ (1.6.2) becomes

$$h_{X\sigma}(\vec{r}_1, \vec{r}_1) = -\rho_\sigma(\vec{r}_1) \tag{1.6.5}$$

Property (1.6.4) states that the exchange hole always integrates to -1, which means that the hole contains exactly minus one electron of spin σ at any reference point, $\vec{r}_1$. Consequently, each electron in the system interacts with $N-1$ other electrons. Therefore, the probability of an electron interacting with itself is null. In HF theory, the exchange hole exactly "cancels" the self-interaction energy J in one-electron systems that arises from a fictitious interaction of the electron with itself. This is not possible in DFT, even for atoms, because of the *approximated* exchange correlation energy term. In DFT, imposing that the sum of the classical Coulomb and exchange-correlation energies vanishes in one-electron systems requires a different potential for each orbital, which is challenging [13].

Property (1.6.5) states that the depth of the exchange hole at the position of the reference electron, $\vec{r}_1$, is $-\rho_\sigma(\vec{r})$ (see Figure 1.1)[3]. This means that none of the other electrons of spin σ can overlap with the reference electron. This is the "Pauli exclusion participle", which excludes the probability of finding two electrons of the same spin overlapping in space.

Recall the relationship between E_{XC} and h_{XC} given by (1.5.10). A similar relationship connects E_X and $h_{X\sigma}$:

$$E_{X\sigma} = \frac{1}{2}\int\int \frac{\rho_\sigma(\vec{r}_1)}{r_{12}} h_{X\sigma}(\vec{r}_1, \vec{r}_2) d\vec{r}_1 d\vec{r}_2. \tag{1.6.6}$$

[3]This figure was obtained from the PhD thesis of Erin Johnson [14], with her permission.

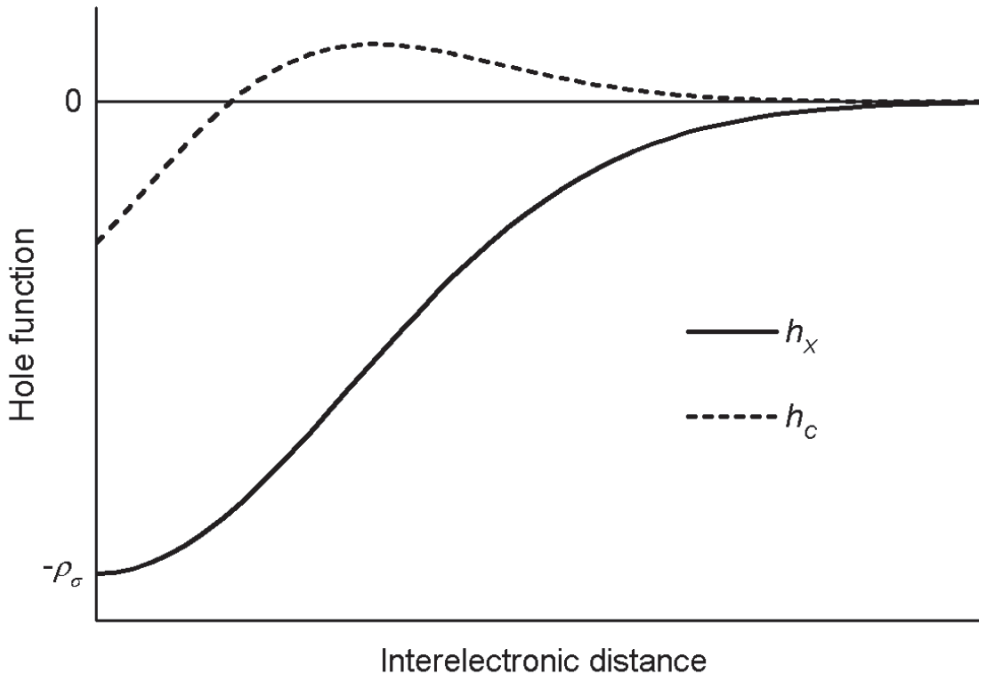


Figure 1.1: A plot of the exchange hole (h_X) for parallel-spin electrons, and the correlation hole (h_C) for opposite-spin electrons.

1.6.2 The Correlation Hole

Unlike the exchange hole $h_{X\sigma}$, the correlation hole h_C integrates to zero:

$$\int h_C(\vec{r}_1, \vec{r}_2) d\vec{r}_2 = 0 \tag{1.6.7}$$

This means that each spin-up electron interacts with all N spin-down electrons. As depicted in Figure 1.1, the correlation hole h_C is negative at small range, but is positive at long range. This means opposite-spin electrons repel each other until a certain separation r, after which the hole becomes positive so that h_C integrates to zero. Unlike the exchange hole, the correlation hole does not have a simply-determined depth at the reference electron. The hole has a cusp at $r_{12} = 0$ (see Figure 1.1). The depth of correlation hole at the reference electron, however, is related to its slope by the "interelectronic cusp condition" [15].

The rest of this chapter is a summary of approaches used to approximate exchange and correlation functionals to obtain E_{XC} in (1.5.3).

1.7 The Local Density Approximation

The simplest approximation for exchange-correlation energy is the local density approximation (LDA). In the LDA, E_{XC} is a functional of only the local density. E_{XC} is estimated by placing a uniform electron gas exchange-correlation hole at the reference point $\vec{r}$:

$$E_{XC}^{LDA}[\rho] = \int \rho(\vec{r}) \Big(-\frac{3}{4}\Big(\frac{3\rho(\vec{r})}{\pi}\Big)^{1/3} + \varepsilon_C[\rho(\vec{r})]\Big) d\vec{r}. \tag{1.7.1}$$

The term $-\frac{3}{4}\big(\frac{3\rho(\vec{r})}{\pi}\big)^{1/3}$ in (1.7.1) is the exchange energy per particle of a uniform gas [7, 16]. $\varepsilon_C[\rho(\vec{r})]$ is the correlation energy per particle. The exchange and correlation energies are weighted by the probability ρ of finding an electron at $\vec{r}$. $\varepsilon_C[\rho(\vec{r})]$ can be approximated several ways. For example, there are the PZ81 [13], the VWN [17] and the PW92 [18] LDA correlation functionals.

A similar equation applies to systems where the spin-α and the spin-β densities are not equal. The only difference is the use of ρ_α and ρ_β instead of ρ in (1.7.1). This spin dependent LDA is called the local spin-density approximation (LSDA). The LSDA is conceptually similar to unrestricted HF (UHF).

1.8 The Generalized Gradient Approximation

The major failure in LDA is the incapability of describing non-homogeneous electron densities. One way to improve this approximation is consider not only the electron density but also its gradient. In the *generalized gradient approximation* (GGA), density gradient corrections are added to the LDA. GGA exchange-correlation energy is

the sum of GGA exchange energy and GGA correlation energy:

$$E_{XC}^{GGA} = E_X^{GGA} + E_C^{GGA}. \tag{1.8.1}$$

The GGA exchange energy in the above equation is the LDA exchange energy [7, 16] [see (1.7.1)] with an additional correction term equal to:

$$-\sum_{\sigma} \int F\left(\frac{|\bigtriangledown \rho_\sigma(\vec{r})|}{\rho_\sigma^{4/3}(\vec{r})}\right) \rho_\sigma^{4/3}(\vec{r}) d\vec{r}, \tag{1.8.2}$$

where $F\left(\frac{|\bigtriangledown \rho_\sigma(\vec{r})|}{\rho_\sigma^{4/3}(\vec{r})}\right)$ is an "exchange enhancement factor" and $\frac{|\bigtriangledown \rho_\sigma(\vec{r})|}{\rho_\sigma^{4/3}(\vec{r})}$ is a dimensionless or "reduced" density gradient, symbolized by χ_σ. In LDA functionals, χ_σ vanishes to zero. In GGA functionals, $\chi_\sigma \to 0$ in the bonding regions where the gradient is minimal; and χ_σ gets more significant in low-density regions, e.g in exponential tails.

GGA atomization energies on the G2 set of molecules [19] are 10 times better than LSDA atomization energies on the same set of molecules. The range of error (compared to accurate reference energies in the G2 set) for GGA energies is 4 - 6 kcal mol^{-1} [20].

As will be shown in the following subsections, there are several GGA functionals reported in the literature to estimate E_X^{GGA} and E_C^{GGA} in (1.8.1).

1.8.1 GGA Exchange Functionals

The exchange enhancement factor in (1.8.2) is what has to be determined in a GGA exchange functional. In 1988, Becke developed the most widely used GGA exchange functional (B or B88), which has only one parameter, $\beta = 0.0042$ determined by least squares fit to exchange energies of rare-gas atoms. This functional gives the exact behavior of exchange energy density at the asymptotic limit. The form of this

functional is as follows:

$$F^B = \frac{\beta\chi_\sigma^2}{1 + 6\beta\chi_\sigma sinh^{-1}\chi_\sigma}, \tag{1.8.3}$$

where χ_σ is the reduced density gradient introduced in (1.8.2). Examples of other exchange GGA functionals are those of Perdew (1991) [21], Handy *et al.* (1993) CAM [22], Filatov and Thiel (1997) FT97 [23]. More exchange GGA functionals include B86 by Becke [24, 25], PW86 by Perdew and Wang [26], LG by Lacks and Gordon [27], PBE by Perdew, Burke and Ernzerhof [28], and revPBE by Zhang, Pan and Yang [29]. The exchange functional PW86 is particularly more attractive than the other functionals because it is a parameter-free functional and yet delivers accurate exchange energies [27, 29, 30, 31]. The PW86 exchange functional has the following form:

$$E_X = \int e_X^{LDA}[\rho(\vec{r})] g_X(s) d\vec{r}, \tag{1.8.4}$$

where e_X^{LDA} is given by

$$e_X^{LDA}[\rho(\vec{r})] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} \rho^{4/3}(\vec{r}) \tag{1.8.5}$$

and $g_X(s)$ is

$$g_X^{PW86}(s) = (1 + 1.296s^2 + 14s^4 + 0.2s^6)^{1/15}, \tag{1.8.6}$$

where s in this case is defined by

$$s = \frac{| \bigtriangledown\rho(\vec{r}) |}{2(3\pi^2)^{1/3}\rho^{4/3}(\vec{r})} \tag{1.8.7}$$

1.8.2 GGA Correlation Functionals

Two years prior to the development of B88, Perdew determined a GGA correlation functional (P86) [32], which was further improved, in 1991 by Wang, to a parameter free functional (PW91) [33]. Lee, Yang and Parr formulated a one parameter

GGA correlation functional (LYP) [34] that is based on the helium atom correlated wavefunction of Colle and Salvetti [35]. The GGA correlation functionals account for "dynamical" correlation within an atom, but not "non-dynamical" correlation between atoms. PBE [33, 28] is an attractive GGA correlation functional as it is (apart from the parameters of the LDA term) a parameter-free functional.

It is possible to combine any of the GGA exchange functionals with any of the GGA correlation functionals to form an exchange-correlation functional. As shown in this section, LDA functionals were improved by including the density gradient corrections in the GGA functionals. However, GGAs still have some disadvantages: they overbind multiple bonds by $\sim$20 kcal mol^{-1} [20], they underestimate the barrier heights of reactions involving radicals. The question to be answered in the following section is "Can the GGA family be further improved?".

1.9 Meta-GGA Functionals

GGA functionals can be further improved by including second order density gradients (or the Laplacian of the density) and the kinetic energy density. These two corrections were first suggested by Becke and Roussel in 1989 [36]. This family of functionals is known as "meta-GGA" functionals. Meta-GGAs describe the inhomogeneous distribution of electron density more accurately. It was not until 1994 that another meta-GGA was published by Proynov, Vela and Salahub [37]. More meta-GGA functionals have been developed since then [38, 39, 40, 41, 42, 43, 44, 45]. Meta-GGAs are only slightly better than GGAs in thermochemical tests [39].

1.10 Hybrid Functionals

"*Hybrid*" functionals (or *adiabatic connection* method, ACM, functionals) have a mixture of some exact exchange from HF with the DFT exchange from LDA/GGA functionals. In 1993, Becke proposed the first hybrid functional [46] by mixing some exact HF exchange with GGA exchange as follows:

$$E_{XC}^{B3} = E_{XC}^{LSD} + a(E_X^{HF} - E_X^{LSD}) + b\Delta E_X^{B88} + c\Delta E_C^{PW91} \tag{1.10.1}$$

The three parameters, a, b and c, were fitted to the atomization energies, ionization potentials and proton affinities of the G2 set. The first term in (1.10.1) is the local spin density exchange-correlation functional. This first term was corrected with two density gradient functionals for exchange and for correlation (the last two terms), each of which has a contribution weighed by a linear parametrized coefficient. The term in brackets [second term of (1.10.1)] is a measure of the importance of exact exchange. Its contribution to the B3 functional is determined by the parameter a. In other words, the parameter a is a measure of the HF:GGA ratio in the exchange energy.

In his paper [46], Becke proposed using the PW91 correlation functional in (1.10.1). One year later, in 1994, Stephens *et al.* [47] suggested using the LYP correlation functional (by Lee, Yang and Parr) instead of PW91, i.e., "B3LYP" hybrid functional rather than "B3PW91". The "3" in 'B3LYP' or 'B3PW91' refers to the 3 parameters of the functional as shown in (1.10.1). B3LYP is a widely used hybrid functional, the number of citations of [46] has exceeded 26 700. Hybrid functionals reduce the error in bond energies for the molecules in the G2 set from 4 - 6 kcal mol^{-1} with GGAs to 2 - 3 kcal mol^{-1} [12].

Despite its remarkable performance for many types of chemical systems, DFT cannot capture interactions weaker than hydrogen-bonding such as dispersion. This problem

is, however, solved as will be described in the following chapter.

Chapter 2

Dispersion Energy in DFT

Dispersion interactions are weak attractive van der Waals interactions. This type of interactions was first introduced by Fritz London in 1926 [48][1], hence the name "London dispersion interactions". These interactions are responsible for attraction between closed shell non-polar atoms or molecules such as rare-gas diatomics. Dispersion is ubiquitous in biological systems (DNA, RNA and protein folding, cell membranes, etc.), polymers, and atmospheric reactions. This non-classical type of interaction is due to long range dynamical correlation effects. It arises from the instantaneous fluctuation in the charge distribution. These spontaneous fluctuations create *instantaneous* dipole moments even though the permanent dipole moment of the atom/molecule is zero. For example, in a He_2 molecule, the instantaneous dipole moment in one He atom induces an instantaneous dipole moment in the second He atom. This dipole-induced-dipole interaction defines dispersion. The general form of dispersion is given by:

$$E_{disp} = -\frac{C_6}{R^6}, \tag{2.0.1}$$

where R is the internuclear distance and C_6 is a dispersion coefficient that depends on the polarizability of the atom/molecule. If higher-order multipole moments are

[1] His work was then translated into English in 1937 [49].

considered, higher order terms are included to the dispersion energy in (2.0.1) as follows:

$$E_{disp} = -\frac{C_6}{R^6} - \frac{C_8}{R^8} - \frac{C_{10}}{R^{10}} \dots \tag{2.0.2}$$

Dispersion interactions can be captured with correlated wavefunction methods such as Møller-Plesset Perturbation Theory (MPn) and Coupled Cluster theory (CC). The CC methods at complete basis set limits are considered as the gold standard method for obtaining reference binding energies. However, CC methods scale as N^7 with system size, i.e., they are much more computationally demanding. MP2 tends to overbind weakly-bound complexes [50]. To reduce basis-set dependence and basis-set superposition error (BSSE) in correlated wavefunction methods, "range separated hybrid" (RSH) methods can be used. These methods consist of a combination of MPn or CC theories with short-range DFT [51, 52, 53, 54].

Even though conventional wavefunction methods accurately describe weakly-bound systems, they are too expensive to be practical. Traditional DFT functionals (LDA, GGA, meta-GGA, hybrids) are incapable of modeling dispersion interactions [55] because of their nonlocal character. However, dispersion corrected DFT functionals are good practical alternatives to conventional wavefunction methods. In the past decade, there have been several attempts reported in the literature to develop DFT-based functionals that capture dispersion interactions. Some functionals are derived from first principles and therefore quite complicated and time consuming [56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66]. Others are based on fitting of empirical parameters [67, 68, 69, 70, 71, 72, 73].

An overview of DFT-based methods will be presented in this chapter along with their pros and cons. Several good reviews are available in the literature [50, 51, 62, 74, 75, 76, 77].

2.1 DFT-Based Dispersion Models

Some methods handle dispersion interactions purely empirically by including many parameters in DFT functionals, as in X3LYP, M05, M05-2X and M06-2X [78, 79, 80]. These empirically adjusted functionals may give poor results for weakly-bound complexes which are not in their training sets [81].

The empirical dispersion-corrected atom-centered potentials (DCACP) of Röthlisberger *et al.* [72, 82, 83] (used in plane wave codes) and the empirical dispersion correcting potentials (DCP) of Dilabio *et al.* [73, 84] are other approaches developed for dispersion interactions. In the DCP model, functionals can be parametrized to give good results with small basis sets, thus this is a convenient method for large systems. These functionals are compatible with electronic structure codes that include effective core potentials. Large grids are needed to get smooth potential energy surfaces [50]. The major drawback of DCP methods is the need for a new pseudo potential optimization for each DFT/basis-set combination, and for each element.

Another approach, DFT-dispersion (DFT-D), is based on adding an explicit dispersion term to DFT functionals. In other words, an explicit pairwise correction for dispersion interactions of the London form, $-\frac{C_{6,ij}}{R_{ij}^6}$ is added. An example of a DFT-D functional is the empirical GGA-based functional, B97-D, developed by Grimme in 2006 [85]. In B97-D, an empirical dispersion term is added to the GGA functional of Becke (B97) [86]. Many examples of DFT-D functionals are available in the literature [68, 69, 71, 85, 87, 88, 89, 90]. The major drawback of the empirical addition of the dispersion correction term is that dispersion coefficients and van der Waals radii have to be empirically chosen for each element. In addition, the dispersion coefficients do not sense different molecular environments, they are assumed to be constants. The latter problem was recently partially solved by Grimme *et al.* [91]. Dispersion coefficients in their revised DFT-D3 method are sensitive to the hybridization state of atoms in molecules. Their goal was achieved by employing a new concept called

"fractional coordination number". Despite their drawbacks, empirical DFT-D methods have been implemented in many computational packages because they are easy to implement and the forces required for geometry optimizations are easily computed. The computational cost for dispersion energy corrections is minor compared to the cost for SCF iterations.

The "ALL" functional of Andersson-Langreth-Lundqvist is derived from first principles [57]. Dobsin and Dinte have also derived this functional independently [92]. These functionals are applicable *only* to long-range interactions between *non- overlapping* systems, thus the system needs to be fragmented into non-overlapping groups [58, 61]. This problem was eliminated in more recent versions of ALL, thus it can capture *intra*-molecular interactions [93, 94, 95, 96]. A damping function for the ALL functional was derived by Kamiya *et al.* and by Sato *et al.* [58, 61]. Damping functions are necessary to avoid divergence of dispersion energies to minus infinity as internuclear distances approach zero. ALL was simplified by Silvestrelli *et al.* [77, 97, 98, 99] by using maximally localized Wannier functions.

The "vdW-DF" is also a nonempirical nonlocal van der Waals functional [93, 100, 101]. This "seamless" functional can capture dispersion interactions self-consistently [102]. This functional is seamless because dispersion energy is not distinguished from dynamical correlation energy, therefore dispersion is captured at the long-range limit and damping functions are not required at the short-range limit. Forces can also be evaluated efficiently for geometry optimization purposes. However, even though coarse grids are sufficient, double numerical integrations over space are needed [102]. Despite the difficulty in implementing this functional in electronic structure programs, it has been implemented in plane wave codes [100] and in Gaussian basis-set codes [102]. This method is more computationally expensive than DFT-D but less expensive than correlated *ab initio* methods (MPn and CC).

Sato and Nakai have developed another nonempirical method to calculate disper-

sion interactions [62]. Their local response dispersion (LRD) a DFT-D approach with dispersion coefficients derived from first principles. In other words, the LRD approach consists of adding an explicit nonempirical dispersion term to a DFT functional. In the LRD method, explicit frequency integrals of the Casimir-Polder type are used to evaluate dispersion coefficients from multipole polarizabilities of atoms within molecules. The multipole polarizabilities are obtained from the the local response approximation of Dobson and Dinte [92] where the dielectric model is modified according to Vydrov and van Voorhis [103]. A local response model for computing the C_6 terms has already been used in the ALL method [57]. With the LRD approach, however, there is no need to fragment the system into non-overlapping groups. In addition, with LRD, the frequency integrals are not as cumbersome as the numerical double integrations in the ALL and the vdW-DF methods. Other advantages are that the dispersion coefficients are system dependent and the method is computationally efficient (same cost as a KS calculation) without sacrificing accuracy. A limitation of the LRD method is the overestimation of higher-order dispersion coefficients, so only C_6 terms are included. Also, damping functions are required. The LRD adds the dispersion contribution non-self-consistently, which might be problematic for evaluating the necessary forces in a geometry optimization. This LRD functional shares many ideas in common with another preceding model, namely the exchange-hole dipole moment (XDM) model of Becke and Johnson [64, 65, 66].

Another parameter-free method has been developed by Tkatchenko and Scheffler [63]. Their method is similar to the LRD method. Both obtain C_6 dispersion coefficients derived, in part, from the ground state electron density. Integrals of Casimir-Polder type are used to obtain the C_6 coefficients, combined with reference C_6 and polarizability data for free atoms.

A nonempirical model developed prior to the nonempirical methods discussed above is the exchange-hole dipole moment (XDM) model of Becke and Johnson [64, 65,

66]. This model senses the environment of atoms in molecules, and it performs well for *inter*- and *intra*-molecular interactions [104] which are ubiquitous in biological systems. The computational cost of XDM is negligible even though not only C_6 terms are computed, but also higher-order C_8 and C_{10} terms. XDM is based on the position-dependent dipole moment of the exchange hole. It would be ideal to use the dipole moment of the exchange-correlation hole (discussed in Chapter 1). However, to avoid complexities, Becke and Johnson assume that the dipole of the correlation part is minor and can be ignored, and they consider only the exchange part. Hence the model is called *exchange-hole dipole moment* (XDM).

2.2 XDM Model

The exchange hole (or Fermi hole, section 1.6.1) is expressed as follows:

$$h_{X\sigma}(\vec{r}_1, \vec{r}_2) = -\frac{1}{\rho_\sigma(\vec{r}_1)} \sum_{ij} \psi_{i\sigma}(\vec{r}_1)\psi_{j\sigma}(\vec{r}_1)\psi_{i\sigma}(\vec{r}_2)\psi_{j\sigma}(\vec{r}_2), \tag{2.2.1}$$

where ρ_σ is the σ-spin density and $\psi_{i\sigma}$ are occupied Hartree-Fock or Kohn-Sham orbitals. They are assumed above to be real. This hole is a measure of depletion of probability (with respect to the total σ-spin electron density, ρ_σ) of finding an electron of spin σ at a point $\vec{r}_2$ next to a reference electron of same spin σ at point $\vec{r}_1$. The probability of finding two coincident electrons of the same spin (i.e., $\vec{r}_1 = \vec{r}_2$) is absolutely extinguished as imposed by the Pauli exclusion principle, i.e.,

$$h_{X\sigma}(\vec{r}_2 = \vec{r}_1) = -\rho_\sigma(\vec{r}_1). \tag{2.2.2}$$

The exchange hole is always negative and always contains one electron. The shape of this hole is a function of $\vec{r}_2$ and also depends on the position of the reference electron $\vec{r}_1$, i.e., it "follows" the reference electron. This hole is not spherically symmetric

around $\vec{r}_1$ unless $\vec{r}_1$ is *at the center* of a system with spherically symmetric total density. Thus the electron plus its hole have a position-dependent dipole moment even though the electron-hole combination is neutral overall. The $\vec{r}_1$-dependent dipole moment is obtained by integrating the exchange-hole expression over $\vec{r}_2$:

$$d_{X\sigma} = \left[\frac{1}{\rho_\sigma(\vec{r}_1)} \sum_{ij} \vec{r}_{ij\sigma} \psi_{i\sigma}(\vec{r}_1)\psi_{j\sigma}(\vec{r}_1)\right] - (\vec{r}_1) \tag{2.2.3}$$

$$\vec{r}_{ij\sigma} = \int \vec{r}_2 \psi_{i\sigma}(\vec{r}_2)\psi_{j\sigma}(\vec{r}_2) d^3\vec{r}_2. \tag{2.2.4}$$

Figure 2.1 is an illustration of an electron (e^-) in a spherical atom, the mean position of its hole (h^+) and the instantaneous dipole moment of the electron and its exchange hole ($d_{X\sigma}$) pointing towards the nucleus of the atom. The electron is at a distance r from the nucleus and at a solid angle Ω. The mean position of the hole is at a distance $r - d_{X\sigma}$ from the nucleus and at the same solid angle Ω. This is a simplified illustration because the hole is represented as a point located at the mean position of the hole.

This two-point (electron and hole) picture also generates higher order multipole moments $M_{l\sigma}$, with respect to the nucleus as origin:

$$M_{l\sigma} = -[r^l - (r - d_{X\sigma})^l]. \tag{2.2.5}$$

Higher multipole moments are also at a solid angle Ω. It is possible to define higher order multipole moments from the two-point picture (Figure 2.1) because the only variable required to calculate these multipole moments (2.2.5) is the magnitude of the exchange-hole *dipole* moment $d_{X\sigma}$. This ease in deriving higher order multipoles provides the XDM model with a great advantage over all other competing methods.

With XDM, the van der Waals theory has two variants, orbital based, i.e., (2.2.3) and (2.2.4), and density functional based. In the density-functional variant, $d_{X\sigma}$ can be

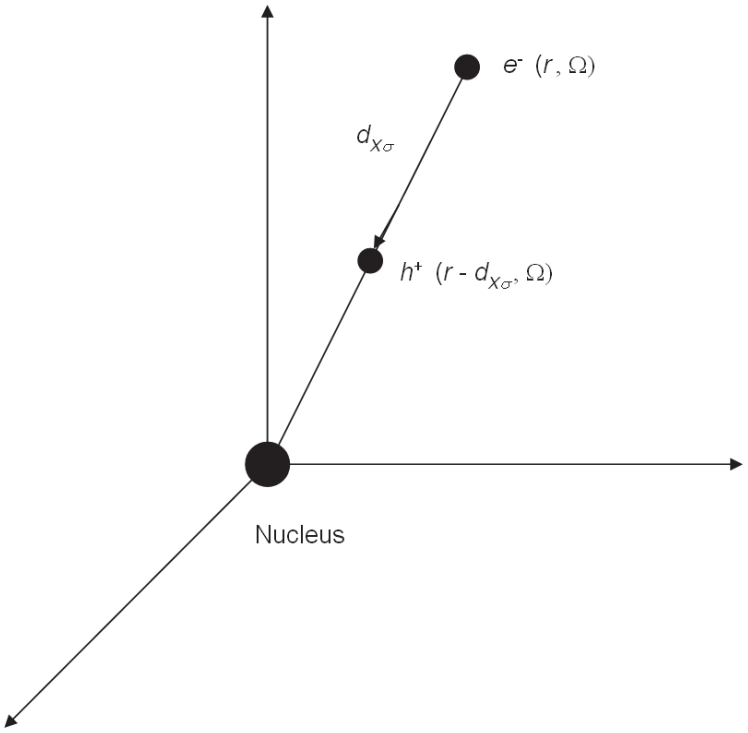


Figure 2.1: A representation of an electron (e^-) in a spherical atom, the mean position of its hole (h^+), and the dipole moment of the electron and its exchange hole ($d_{X\sigma}$) pointing towards the nucleus of the atom.

approximated using local densities and the Becke-Roussel (BR) exchange-hole model [36]. The BR exchange-hole model is based on a hydrogenic atom reference system instead of a uniform electron gas reference system. As represented in Figure 2.2, the BR model places a normalized exponential function of the form $-\frac{a^3}{8\pi}e^{-ar}$ at a distance b from an electron of spin σ.

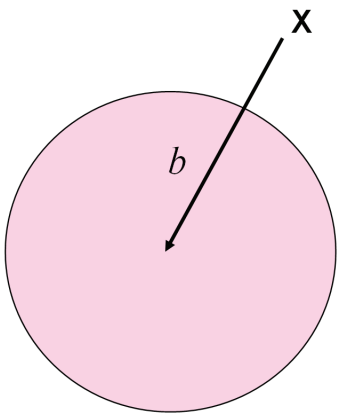


Figure 2.2: A representation of an electron at position X and an exponential function (hydrogenic hole) centered at a distance b from the reference electron.

This model shown in Figure 2.2 is exact for a hydrogen atom, but is an approximation for other systems. In the BR model the distance b (see Figure 2.2) is equal to $d_{X\sigma}$ (see Figure 2.1).

2.3 Becke-Roussel Exchange-Hole Model

The expression for the distance b (see Figure 2.2) is derived from the BR model as follows: in the BR model, an exponential function of the form $-Ae^{-ar}$ is placed at a distance b from a reference point (X in Figure 2.2). A is determined by normalizing the model hole to a unit charge. A is thus given by

$$A = \frac{a^3}{8\pi} \tag{2.3.1}$$

The unknowns a and b are determined by equating Taylor expansions of the model hole and the exact hole. The Taylor expansion of the spherically averaged hydrogenic BR model hole around the reference point $\vec{r}$ is:

$$h_{X\sigma}^{BR}(\vec{r}, s) = -Ae^{-x} - \frac{Aa^2}{6}\left(1 - \frac{2}{x}\right)e^{-x}s^2 + \ldots, \tag{2.3.2}$$

where $x = ab$. The Taylor expansion of the exact exchange hole is:

$$h_{X_\sigma}^{exact}(\vec{r}, s) = -\rho_\sigma - \frac{1}{6}\left(\bigtriangledown^2 \rho_\sigma - 2\tau_\sigma + \frac{1}{2}\frac{(\bigtriangledown\rho_\sigma)^2}{\rho_\sigma}\right)s^2 + \ldots, \tag{2.3.3}$$

where τ_σ is twice the positive definite kinetic energy density:

$$\tau_\sigma = \sum_i (\bigtriangledown\psi_{i\sigma})^2. \tag{2.3.4}$$

Comparing the first terms of (2.3.2) and (2.3.3), and using (2.3.1), gives the following relationship:

$$\rho_\sigma = Ae^{-x} = \frac{a^3}{8\pi}e^{-x} \tag{2.3.5}$$

Comparing the second terms of (2.3.2) and (2.3.3) and using (2.3.5) gives the following relationship:

$$Q_\sigma = \frac{\rho_\sigma}{6b}(a^2b - 2a) \tag{2.3.6}$$

where

$$Q_\sigma = \frac{1}{6}\left(\bigtriangledown^2 \rho_\sigma - 2\tau_\sigma + \frac{1}{2}\frac{(\bigtriangledown \rho_\sigma)^2}{\rho_\sigma} \right) \tag{2.3.7}$$

Solution for (2.3.5) and (2.3.6) is obtained by solving the following nonlinear equation (using the Newton-Raphson method):

$$\frac{xe^{-2x/3}}{x-2} = \frac{2}{3}\pi^{2/3}\frac{\rho_\sigma^{5/3}}{Q_\sigma}, \tag{2.3.8}$$

and then

$$b^3 = \frac{x^3e^{-x}}{8\pi\rho_\sigma} \tag{2.3.9}$$

is the equation used to determine b.

2.4 Dispersion Interactions between Free Atoms

As mentioned in section 2.2, the approximate model of the electron and its hole is important for generating dipole and higher order moments.

Figure 2.3 shows a spherically symmetric atom A placed at a large distance R from another spherically symmetric atom B. An electron at position $\vec{r}_A$ in atom A will generate multipole moments with respect to the nucleus A. These multipole moments are directed along $\vec{r}_A$ towards the nucleus, and they are evaluated according to (2.2.5). Similarly, an electron B at position $\vec{r}_B$ in atom B will generate multipole moments with

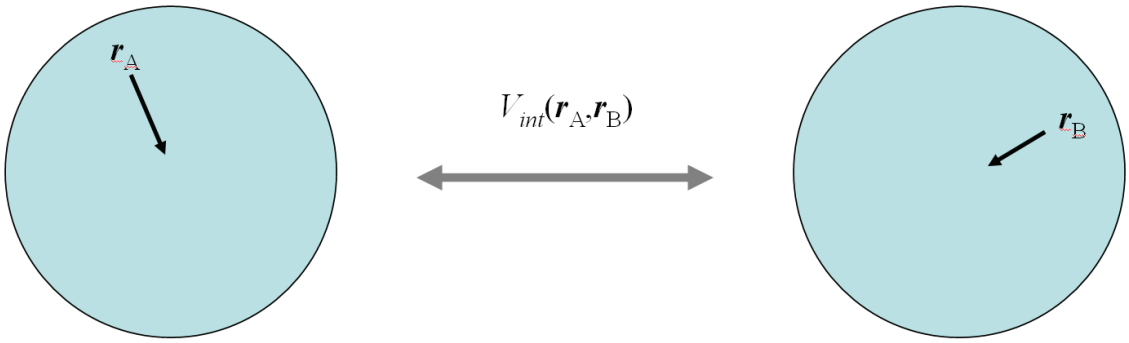


Figure 2.3: Interaction between two spherically symmetric atoms through multipole moments of an electron and its exchange hole in each of the atoms.

respect to the nucleus B directed along $\vec{r}_B$. The multipole moments of the electron and its hole at position $\vec{r}_A$ interact with the multipole moments of the electron and its hole at position $\vec{r}_B$, and create a multipole-multipole interaction between atoms A and B. This interaction, denoted as $V_{AB}(\vec{r}_A, \vec{r}_B)$ in Figure 2.3, is given by

$$\begin{aligned} V_{AB}(\vec{r}_A, \vec{r}_B) =& V_{dipole_A - dipole_B} + V_{dipole_A - quadrupole_B} + V_{quadrupole_A - dipole_B} \\ &+ V_{dipole_A - octopole_B} + V_{octopole_A - dipole_B} + V_{quadrupole_A - quadrupole_B} + \dots . \end{aligned} \tag{2.4.1}$$

The dispersion energy is derived using second-order ground-state perturbation theory in the "closure" or Unsöld approximation [2]. If the first-order ground-state energy correction is zero:

$$E^{(1)} = < V_{pert} > = 0, \tag{2.4.2}$$

then the second-order energy correction is approximately equal to:

$$E^{(2)} = -\frac{< V_{pert}^2 >}{\Delta E_{av}}, \tag{2.4.3}$$

where $< V_{pert}^2 >$ is the expectation value in the ground state, and ΔE_{av} is an average excitation energy. The expectation value $< V_{pert}^2 >$ can be evaluated by squaring

the multipole-multipole interaction $V_{AB}(\vec{r}_A, \vec{r}_B)$ (see (2.4.1) and Figure 2.3), then by integrating over all points $\vec{r}_A$ in atom A and all points $\vec{r}_B$ in atom B and weighted by the atomic spin densities. This is a "semiclassical" calculation of the expectation value because the dispersion energy is evaluated classically with multipole-multipole interactions, but the dipole moment of the exchange hole is a quantum effect.

The dispersion coefficients in

$$E_{disp} = -\frac{C_6}{R^6} - \frac{C_8}{R^8} - \frac{C_{10}}{R^{10}} - \ldots, \tag{2.4.4}$$

are obtained when evaluating the expectation value $< V_{pert}^2 >$. These coefficients are given by

$$C_6 = \frac{2}{3}\frac{< M_1^2 >_A < M_1^2 >_B}{\Delta E_{av}}, \tag{2.4.5}$$

$$C_8 = \frac{< M_1^2 >_A < M_2^2 >_B + < M_2^2 >_A < M_1^2 >_B}{\Delta E_{av}}, \tag{2.4.6}$$

and

$$C_{10} = \frac{4}{3}\frac{< M_1^2 >_A < M_3^2 >_B + < M_3^2 >_A < M_1^2 >_B}{\Delta E_{av}} + \frac{14}{5}\frac{< M_2^2 >_A < M_2^2 >_B}{\Delta E_{av}}. \tag{2.4.7}$$

The atomic moment integrals, $< M_l^2 >$, are given by

$$< M_l^2 > = \sum_\sigma \int \rho_\sigma(\vec{r})[r^l - (r - d_{X\sigma})^l]^2 d^3\vec{r}. \tag{2.4.8}$$

The average excitation energy, ΔE_{av}, is assumed to be the sum of average excitation energies of the atoms in the system. In this case

$$\Delta E_{av} = \Delta E_A + \Delta E_B, \tag{2.4.9}$$

where ΔE of each atom is given by

$$\Delta E = \frac{2}{3}\frac{< M_1^2 >}{\alpha}, \tag{2.4.10}$$

where α is the atomic dipole polarizability. Equation (2.4.10) is derived from second-order perturbation theory applied on individual atoms [66].

XDM is a nonempirical model for dispersion interactions depending *only* on atomic polarizabilities obtained from [105] and moment integrations. The dispersion coefficients evaluated using equations (2.4.5), (2.4.6) and (2.4.7) are in good agreement with *ab initio* reference data [106, 107, 108]. The mean absolute percent errors (MAPEs) for 21 pairs of the atoms H, He, Ne, Ar, Kr, Xe are 3.4%, 21.5% and 21.5% for C_6, C_8 and C_{10}, respectively [66].

2.5 Dispersion Interactions between Atoms in Molecules

The same concepts used for dispersion interactions between free atoms also apply to dispersion interactions between atoms in molecules. The Hirshfeld partitioning scheme [109] is used to partition the space of a molecular system into atomic components. Hirshfeld weight functions are defined by

$$w_i(\vec{r}) = \frac{\rho_i^{at}(\vec{r})}{\sum_n \rho_n^{at}(\vec{r})}, \tag{2.5.1}$$

where ρ_i^{at} is the spherical free atomic density of each atom i and the n summation is over all atoms in the molecular system. The sum of the weight functions at any point

$\vec{r}$ over all atoms adds up to one:

$$\sum_i w_i(\vec{r}) = 1, \tag{2.5.2}$$

The weight function $w_i(\vec{r})$ approaches one at points near nucleus i and approaches zero everywhere else. The moment integration $< M_l^2 >_i$ is partitioned as follows:

$$< M_l^2 >_i = \sum_\sigma \int w_i(\vec{r})\rho_\sigma(\vec{r})[r^l - (r - d_{X\sigma})^l]^2 d^3\vec{r}. \tag{2.5.3}$$

The effective atom-in-molecule polarizability α_i needed in (2.4.10) is given by

$$\alpha_i = \frac{< r^3 >_i}{< r^3 >_{i,free}} \alpha_{i,free}, \tag{2.5.4}$$

where $< r^3 >_i$ is an effective volume integration for an atom in a molecule given by

$$< r^3 >_i = \int r^3 w_i(\vec{r})\rho(\vec{r}) d^3\vec{r}, \tag{2.5.5}$$

and $< r^3 >_{i,free}$ is an effective volume integration for a free atom given by

$$< r^3 >_{i,free} = \int r^3 \rho_{i,free}(\vec{r}) d^3\vec{r}. \tag{2.5.6}$$

It is possible to express polarizabilities in terms of volumes due to the well-established qualitative relationship between these two quantities [110, 111, 112]. Thus, for each atom in a molecule, the effective "atomic" excitation energy is given by

$$\Delta E_i = \frac{2}{3} \frac{< M_l^2 >_i}{\alpha_i} \tag{2.5.7}$$

Assuming that each of the three intermolecular dispersion coefficients (C_m, m= 6, 8,

10) is a sum of interatomic dispersion coefficients, C_m is thus given by

$$C_m = \sum_i^A \sum_j^B C_{m,ij}, \tag{2.5.8}$$

Where i denotes an atom in molecule A and j denotes an atom in molecule B. Therefore, using (2.5.7) and substituting A with i and B with j, equations (2.4.5), (2.4.6) and (2.4.7) can now be rewritten as:

$$C_{6,ij} = \frac{\alpha_i \alpha_j < M_1^2 >_i < M_1^2 >_j}{< M_1^2 >_i \alpha_j + < M_1^2 >_j \alpha_i}, \tag{2.5.9}$$

$$C_{8,ij} = \frac{3}{2} \frac{\alpha_i \alpha_j \big(< M_1^2 >_i < M_2^2 >_j + < M_2^2 >_i < M_1^2 >_j \big)}{< M_1^2 >_i \alpha_j + < M_1^2 >_j \alpha_i}, \tag{2.5.10}$$

$$\begin{aligned} C_{10,ij} =& 2 \frac{\alpha_i \alpha_j \big(< M_1^2 >_i < M_3^2 >_j + < M_3^2 >_i < M_1^2 >_j \big)}{< M_1^2 >_i \alpha_j + < M_1^2 >_j \alpha_i} \\ &+ \frac{21}{5} \frac{\alpha_i \alpha_j < M_2^2 >_i < M_2^2 >_j}{< M_1^2 >_i \alpha_j + < M_1^2 >_j \alpha_i}. \end{aligned} \tag{2.5.11}$$

MAPEs of C_6, C_8 and C_{10} in atom-molecule and molecule-molecule pairs are 12.7%, 16.5% and 21.2%, respectively [113]. Test sets are detailed in [113] and reference data was obtained from [106, 108, 114].

2.6 Dispersion Energy Formula

The asymptotic *inter* and *intra*-molecular dispersion energy has the following general form:

$$E_{disp} = -\sum_{i>j} \left(\frac{C_{6,ij}}{R_{ij}^6} + \frac{C_{8,ij}}{R_{ij}^8} + \frac{C_{10,ij}}{R_{ij}^{10}} \right), \tag{2.6.1}$$

where R_{ij} is the distance between nuclei i and j. Equation (2.6.1) captures the proper physics at long range (as shown in Fig 2.4), but it diverges as $R_{ij} \to 0$. Therefore, the

terms of this asymptotic series need to be cut off at small R_{ij} by a damping function [56, 68, 115, 116]. The damped dispersion energy formula in XDM is given by

$$E_{disp} = -\sum_{i>j}\left(\frac{C_{6,ij}}{R_{vdW,ij}^6 + R_{ij}^6} + \frac{C_{8,ij}}{R_{vdW,ij}^8 + R_{ij}^8} + \frac{C_{10,ij}}{R_{vdW,ij}^{10} + R_{ij}^{10}}\right), \tag{2.6.2}$$

where $R_{vdW,ij}$ is an effective van der Waals separation that linearly depends on a critical radius $R_{c,ij}$:

$$R_{vdW,ij} = a_1 R_{c,ij} + a_2, \tag{2.6.3}$$

where a_1 and a_2 are two universal fit parameters, and the critical radius $R_{c,ij}$ is the separation at which the $\frac{C_m}{R^m}$ (m= 6, 8, 10) terms have roughly the same value:

$$\frac{C_{6,ij}}{R_{c,ij}^6} \approx \frac{C_{8,ij}}{R_{c,ij}^8} \approx \frac{C_{10,ij}}{R_{c,ij}^{10}}. \tag{2.6.4}$$

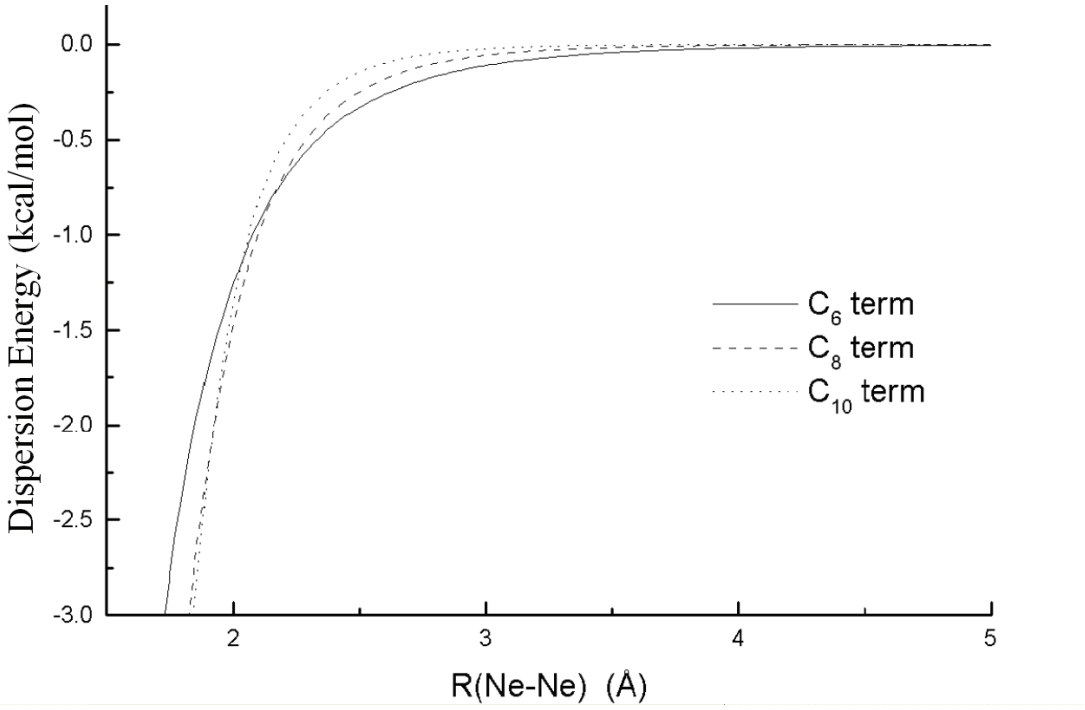


Figure 2.4: Dispersion energy from C_6, C_8 and C_{10} terms as a function of internuclear distance in the neon dimer

Figure 2.4 illustrates the equal contribution of the C_6, C_8 and C_{10} terms to the dispersion energy of the neon dimer at a specific internuclear distance (∼2.1 Å in this

case) called the critical radius.

Solving for $R_{c,ij}$, three solutions are available. $R_{c,ij}$ is chosen as the average of the three solutions:

$$R_{c,ij} = \frac{\left(\frac{C_{8,ij}}{C_{6,ij}}\right)^{1/2} + \left(\frac{C_{10,ij}}{C_{6,ij}}\right)^{1/4} + \left(\frac{C_{10,ij}}{C_{8,ij}}\right)^{1/2}}{3}. \tag{2.6.5}$$

The optimal values of the universal parameters, a_1 and a_2 in (7.2.3), were determined in [65] to be 0.83 and 1.55 Å, respectively [65]. These values were determined by minimizing the root mean square percent error of binding energies for 45 complexes. The MAPE of the energies calculated using the post-HF scheme, $E_{total} = E_X^{HF} + E_C^{BR} + E_{disp}^{XDM}$, is 14.1%. In this scheme E_X^{HF} is the HF exchange energy, E_C^{BR} is the dynamical correlation functional of Becke and Roussel [117] and E_{disp}^{XDM} is the XDM dispersion energy. The binding energies obtained from this functional were twice as good as those obtained from MP2 theory [14, 65].

The average contribution of each term to the total dispersion energy of the 45 complexes at their equilibrium geometry is roughly 60 %, 30 % and 10 % for C_6, C_8 and C_{10}, respectively. This indicates that higher order terms should not be neglected as their contribution to total dispersion energies is significant.

Chapter 3

Benchmark Study of Basis Sets Used for Computing Binding and Repulsive Energies of Weakly-Bound Systems

3.1 Abstract

In the previous chapter, the XDM model, which captures dispersion interactions in the DFT framework, was introduced. The next chapter presents how XDM was used in the development of a nonempirical GGA functional that accurately captures van der Waals interactions. The exchange-correlation part of this functional is computed self consistently with the basis-set-dependent program, G09 [118]. The purpose of the study presented in this chapter is to assess the ability of a variety of basis sets to reproduce accurate exchange-correlation energies in rare-gas diatomics. This chapter covers a detailed introduction about basis sets and their characteristics. The introduction is followed by a quantitative assessment of basis sets on the repulsive energies

of weakly-bound rare-gas diatomcs. In this study, the counterpoise correction and the grid size on the binding energies of ten rare-gas diatomic systems are also tested.

3.2 Introduction

A basis set is a set of functions used to describe molecular orbitals. Molecular orbitals ϕ are built from a linear combination of the atomic orbitals χ (LCAO) as follows:

$$\phi = \sum_{i}^{n} c_i \chi_i, \tag{3.2.1}$$

where c_i is a coefficient that determines the weight of the contribution of χ to the molecular orbital ϕ.

The LCAO formulation was proposed by Roothaan in 1951 to solve the SCF procedure for each one electron equation [119]. The advantage of the LCAO approach is that it converts the 3D, multicenter, differential Schrödinger equation to a generalized eigenvalue problem, namely the *secular* equation.

Slater-type orbitals (STO) are good representations of atomic orbitals. STO are given by

$$N r^{n-1} e^{-\zeta r} Y_{lm}(\theta, \phi). \tag{3.2.2}$$

where N is the normalization constant, n is the principle quantum number of the orbital, ζ is the orbital exponent which is related to the effective charge of the nucleus, r is the distance from the electron to the atomic nucleus and $Y_{lm}(\theta, \phi)$ is the angular part of the nucleus.

STOs have a cusp at the nucleus as depicted in Figure 3.1. This cusp is problematic in evaluating integrals within the Coulomb and exchange-correlation potentials mentioned in section 1.5. Pople and coworkers solved this problem in 1969 by replacing STOs with Gaussian type orbitals (GTOs) [120, 121, 122]. GTOs are Gaussian

continuous functions with no cusps as they are exponential functions given by

$$\chi^{GTO} = Nx^i y^j z^k e^{-\alpha r^2}, \tag{3.2.3}$$

where N is a normalization factor; α is the exponent of the function, and the sum l of i, j and k determines the angular momentum. α determines how diffuse the GTO is (when α is small) or how compact the GTO is is (at large values of α). Figure 3.1 depicts one STO and three GTOs with different values of α.

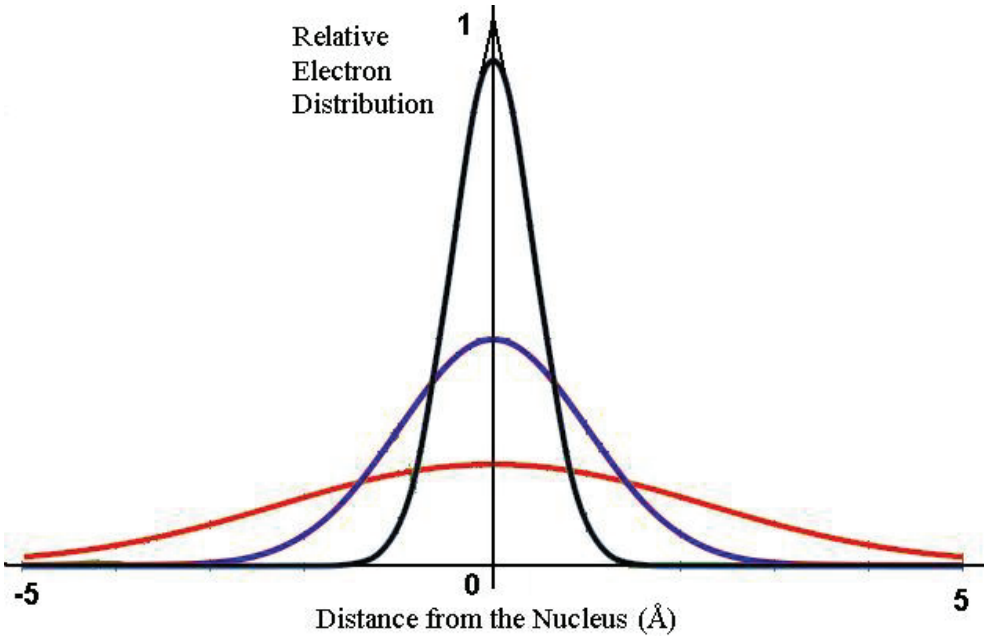


Figure 3.1: A plot of the Slater type orbital with the cusp, and three Gaussian type orbitals with different values of α.

The linear combination of multiple Gaussian functions (or primitives, g_p) forms one basis function χ_i, or equivalently, a *contracted Gaussian type orbital* (CGTO) given by

$$CGTO = \sum_p d_p g_p = \chi_i \tag{3.2.4}$$

The contribution of each primitive g_p to the basis function χ_i is determined by the contraction coefficient d_p. A minimal basis set is built from a minimal number of basis functions necessary to define each electron in a system. For example, STO-3G is a

minimal basis set where 3 primitives are contracted to form each STO basis function. When an atom forms a chemical bond it undergoes deformations to adapt itself to the molecular environment. Therefore, an "extended basis set" is needed to allow radial and angular flexibilities [2].

The radial flexibility allows contraction or expansion of the atomic orbital. Expanded atomic orbitals are mandatory for systems with a diffuse electron density as in anions and hydrogen-bonded complexes. The radial flexibility can be introduced by replacing each basis function of a minimal basis set by two (double zeta (DZ) basis set [123]), three (triple zeta (TZ) basis set) or n basis functions (n zeta basis set). Thus, the minimal basis set becomes a *split valence* basis set.

The angular flexibility is needed to account for the shape-distortion (polarization) of the atom. Angular flexibility is introduced by adding one or more polarization functions. A polarization function is a basis function that has an angular momentum l one order higher than the highest angular momentum in a minimal basis set. For example, p is the highest angular momentum in the minimal basis set of a carbon atom. Adding a polarization function on the carbon atom is equivalent to adding a set of d-orbitals.

3.3 Pople Basis Sets

Pople basis sets are denoted by A-XYZG. For example 6-311G means that the core and the valence electrons are considered separately. The number before the dash represents the number of the contracted primitives (six in this case) for each core electron. If there are three numbers after the dash, the basis set is triple zeta split-valence basis set. If there are two numbers after the dash, the basis set is a double zeta basis set [123]. The value of each number after the dash represents the number of primitives contracted to form each basis function.

Polarization functions can be added to Pople basis sets. They are denoted by a an asterisk *. In the double asterisks ** notation, the first asterisk is for heavy atoms and the second asterisk is for hydrogen atoms. Polarization functions can also be written explicitly. For example, 6-311G(d,p) is equivalent to 6-311G**. In this particular basis set, a set of d functions is added on the heavy atoms and a set of p functions is added on H atoms. It is possible to add more than one polarization function per atom to describe its distortion in the molecular bond more precisely. For example, the largest Pople basis set available is 6-311G++(3df,3pd), where three sets of d functions with one set of f functions are added on the heavy atoms, and three sets of p functions with one set of d functions are added on H atoms.

Diffuse functions can also be added to Pople basis sets. The are denoted by "++". The first + is for the heavy atoms and the second + is for the hydrogen atoms.

3.4 Dunning Basis Sets

In 1989, Dunning built basis sets that capture correlation effects (in MPn, CI, CC) in a consistent fashion [124]. "Dunning" basis sets give more reliable results than Pople basis sets because they have ∼six times more contracted Gaussian functions [125, 126, 127, 128]. In Dunning basis sets, basis functions that contribute the same amount of correlation energy are grouped together. Dunning basis sets are denoted by cc-pVXZ (*correlation consistent, polarized valence X zeta*), X can be double (D), triple (T), quadruple (Q), quintuple (5) and even sextuple (6) [129]. In aug-cc-pVXZ, "aug" stands for augmented, i.e., diffuse functions. The cc-pVDZ has 3s2p1d basis functions for first row elements of the periodic table. The cc-pVTZ has 4s3p2d1f orbitals for first row elements.

3.5 Polarization Consistent Basis Sets

"Polarization consistent" (pc) basis sets were proposed by Jensen in 2001 [130]. In analogy to Dunning basis sets, pc basis sets are built so that the basis functions, which contribute the same amount of energy, are grouped together. The pc basis sets were designed based on the importance of the polarization functions to the energy. Calculations using pc basis sets are faster than calculations using cc Dunning basis sets [130]. Polarization consistent basis sets are denoted by pc-x where x can be 0, 1, 2, 3 or 4. A pc-2 basis set is equivalent to a *triple* zeta (not DZ) Dunning basis set. The basis sets pc-0, pc-1, pc-2, pc-3 and pc-4 have 3s, 4s1p, 6s2p1d, 9s4p2d1f and 11s6p4d2f1g basis functions, respectively. Diffuse functions are also denoted by "aug", e.g., aug-pc2.

3.6 Basis Set Superposition Error and Counterpoise Correction

To calculate the binding energy of a reaction A+ B $\rightarrow$ AB, energies of reactants E_A and E_B should be computed using the same basis set used to compute the energy of the product E_{AB}. The binding energy is then calculated as follows:

$$E_{Binding} = E_{AB} - E_A - E_B. \tag{3.6.1}$$

E_A and E_B in (3.6.1) are understabilized compared to E_{AB}. This is because the reactant A has basis functions only for one atom A. Similarly, the reactant B has basis functions only for one atom B. However, the product AB has basis functions for two atoms A and B. This means the electrons in AB have more flexibility than the electrons in A or in B. This is known as the "basis set superposition error" (BSSE). To avoid this problem, a dummy atom X can be placed next to A to form an AX

molecule. Similarly, a dummy atom is placed next to B to form a BX molecule. Dummy atoms are added to allow the addition of basis functions. This way, the electrons in AX and the electrons in BX have the same flexibility as the electrons in AB. This correction is known as the "counterpoise" (CP) correction [131].

3.7 Grids in DFT

The integrals in the exchange-correlation potential in DFT functionals (GGA, and meta-GGA) are too complicated to be solved analytically. Therefore, the integrals of the exchange-correlation potential have to be evaluated numerically [132]. The orientation of the angular grids has to be defined so that the energy is independent of the molecule rotation. *Fine* grids are needed to get reliable results and smooth potential energy curves [50]. Ultrafine grids are sometimes necessary because a minor deviation in the gradient of energy is sufficient to get wrong frequencies as the error is amplified by taking second derivatives.

Grids are denoted by xxx*yyy, where xxx is the number of radial shells surrounding each nucleus and yyy is the number of angular points distributed evenly on each shell. To compromise between computational cost and efficiency, the size of the mesh of grids can be modified based on which region of the atom is covered. This technique is called "grid pruning". A smaller mesh is good only near the nucleus where the electron density is more spherical; and at very large distances from the nucleus where the electron density is irrelevant. Larger grids are *necessary* for regions in between where the density is not spherical or irrelevant.

3.8 Testing Basis Sets For Repulsive Energies of vdW Complexes

The effect of basis sets on the repulsive energies of rare-gas diatomics is presented in this section. The repulsive energies are calculated by computing single point energies of each monomer separately. The energy of the complex is also calculated at the experimental internuclear separations determined by Tang-Toennies [133]. In this study, the correlation consistent Dunning basis sets [aug-cc-pVXZ (X=D, T and Q)] and the polarization consistent basis sets by Jensen [aug-pc-n (n=0, 1, 2, 3)] were tested. The calculations were performed using Gaussian 03 [134]; with and without a counterpoise correction using several grids (default or 075302 pruned, ultrafine, i.e., 99590, pruned and unpruned, 056302 unpruned, 100302 unpruned and 099590 unpruned).

Dunning basis sets employed in this study are implemented in Gaussian 03. However, Jensen basis sets are not. They were obtained from the "Basis Set Exchange" website[1] [135, 136]. Other basis sets, such as DGDZVP-G03, DZVP DFT orbital, Def2-QZVP, Def2-SV-P and Def2TZVP, were also obtained from the same website.

3.8.1 Results and Discussion

The basis sets DGDZVP-G03, DZVP DFT orbital, Def2-QZVP, Def2-SV-P and Def2TZVP were ignored for the rest of this study because they did not produce good repulsive energies. Among the series of Dunning and Jensen basis sets tested, only aug-cc-pVDZ, aug-cc-pVTZ, pc-1, pc-2 and pc-3 were chosen because, based on our results, smaller basis sets were not reliable, and larger basis sets were very time consuming.

The results of the tests are summarized by reporting the mean percent errors (MPEs),

[1] https://bse.pnl.gov/bse/portal

the mean absolute percent errors (MAPEs) and the maximum absolute percent error (MaxAPE) of the binding energies of the ten rare-gas dimers He_2, He-Ne, He-Ar, He-Kr, Ne_2, Ne-Ar, Ne-Kr, Ar_2, Ar-Kr, Kr_2.

First, repulsive energies from the exchange GGA functional PW86 were compared to the exact HF repulsive energies computed on the basis-set free NUMOL program [137, 138]. The reference exact HF repulsive energies are obtained from [30]. The GGA exchange functional PW86 was chosen because it reproduces the exact HF exchange energies most accurately as addressed by Lacks and Gordon [27], Kannemann and Becke [30] and Murray, Lee, and Langreth [31].

Second, the exchange-correlation energies, calculated with the GGA functionals PW86 +PBE, were compared to reference energies calculated at the same level of theory with large basis sets, aug-pc-3 and aug-cc-pVQZ. The reference energies were computed using the 200590 unpruned grid without CP correction.

3.8.2 Comparison of PW86 and HF Repulsive Energies

The PW86 repulsive energies calculated with aug-cc-pVTZ and aug-cc-pVDZ with and without counterpoise (CP) correction using the unpruned grid 200590 are compared to the numerical HF energies and the PW86 energies computed using aug-cc-pV5Z with counterpoise correction obtained from [30]. The results are reported in Table 3.1 and Figure 3.2.

The PW86 repulsive energies computed with aug-cc-pVTZ without CP are much similar to the PW86 repulsive energies computed with aug-cc-pV5Z with CP. The MAPEs compared to the numerical HF repulsion energies are only 10% and 13% with quintuple and triple zeta basis sets, respectively. Without CP correction, triple zeta basis produces better repulsive energies than the double zeta basis set. CP corrections with the double zeta basis set significantly improve the repulsive energies

Table 3.1: Repulsion energies (in μH) of ten rare-gas diatomics.

	HF numerical	PW86 aug-cc-pV5Z/CP	PW86 aug-cc-pVTZ (CP)	PW86 aug-cc-pVDZ (CP)
He_2	28	29	34 (36)	28 (38)
He-Ne	47	52	41 (51)	13 (40)
He-Ar	85	87	96 (101)	76 (105)
He-Kr	103	98	104 (112)	88 (116)
Ne_2	96	117	114 (135)	19 (80)
Ne-Ar	190	198	196 (211)	117 (182)
Ne-Kr	227	219	206 (222)	136 (200)
Ar_2	435	503	519 (537)	482 (573)
Ar-Kr	538	626	632 (655)	621 (711)
Kr_2	674	796	786 (818)	807 (897)
MPE		8	9 (18)	-21 (12)
MAPE		10	13 (19)	30 (22)
MaxAPE		22	21 (41)	80 (36)

of diatomics which include the Ne atom. The conclusion from this section is that aug-cc-pVTZ without CP has virtually the same accuracy in repulsion energies obtained with aug-cc-pV5Z with CP, thus it is very reliable.

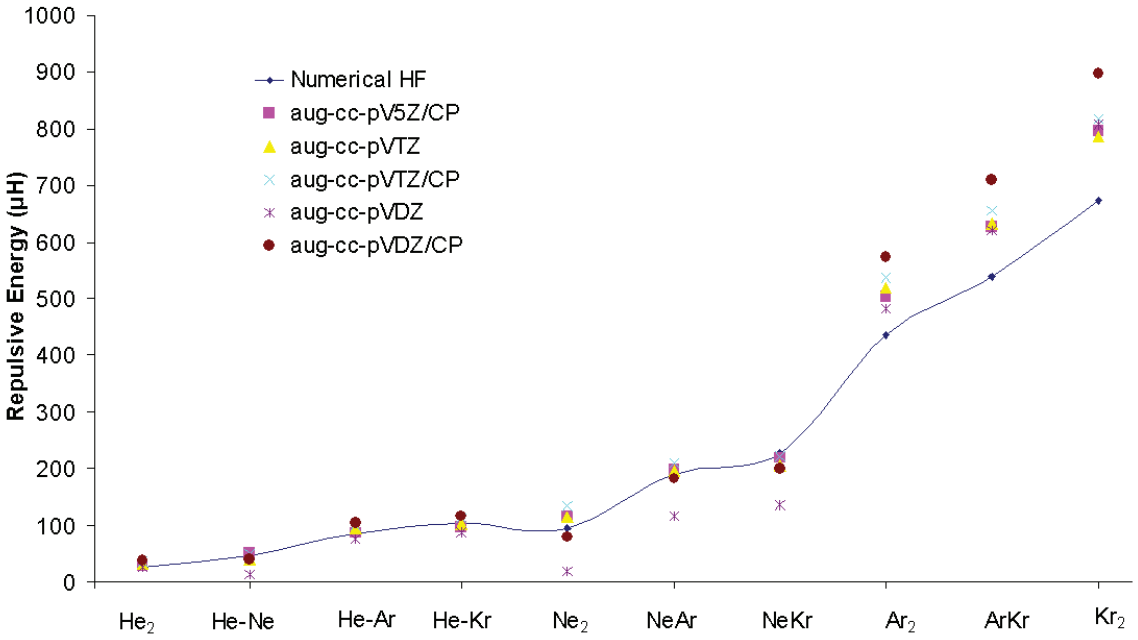


Figure 3.2: Repulsive energies of ten rare-gas diatomic systems with numerical HF (solid line) and with PW86 using different basis sets.

The Effect of the Grid Size on the PW86 Energies

The statistical results, with and without CP correction, for the exchange energies calculated with PW86 are listed in Table 3.2. From this table, it is observed that the three grids are equivalent when PW86 repulsive energies of the ten rare-gas dimers are computed.

Table 3.2: Statistical analysis of the PW86 repulsive energies using multiple basis sets and grid sizes with and without counterpoise correction. P and U stand for pruned and unpruned, respectively.

		Without CP Correction			With CP Correction		
Basis Set	Grid	MPE	MAPE	MaxAPE	MPE	MAPE	MaxAPE
aug-cc-pVDZ	Default	-30	33	70	11	22	34
	Ultrafine P	-22	31	81	12	22	35
	Ultrafine U	-22	31	82	11	22	35
	200590 U	-21	30	81	12	22	34
aug-cc-pVTZ	Default	-2	11	32	17	17	29
	Ultrafine P	8	13	21	18	18	38
	Ultrafine U	8	13	21	18	18	38
	200590 U	9	13	20	18	18	40
aug-pc-1	Default	-49	57	176	25	35	69
	Ultrafine P	-55	79	228	26	33	79
	Ultrafine U	-55	79	229	26	33	79
	200590 U	-54	79	227	27	33	78
aug-pc-2	Default	-12	12	24	36	36	67
	Ultrafine P	-1	5	12	38	38	70
	Ultrafine U	-1	5	13	38	38	70
	200590 U	0	4	10	39	39	72
aug-pc-3	Default	-7	11	27	8	9	20
	Ultrafine P	6	6	16	10	10	19
	Ultrafine U	6	6	16	10	10	18
	200590 U	7	7	16	10	10	21

In case of non-CP-corrected energies, repulsive energies obtained with the double zeta Dunning basis set are much more reliable than repulsive energies obtained with the equivalent Jensen double zeta basis set. This observation, however, is not valid with triple zeta basis sets. The Jensen triple zeta basis set gives slightly better statistics

than the Dunning triple zeta basis set. However, there are two major drawbacks associated with using the Jensen triple zeta basis set, aug-pc-2 basis. First, it is not applicable to all atoms, e.g., the Kr atom. In this study, dimers involving the Kr atom were not included when this particular basis set was used. Second, it is not readily available in the Gaussian package.

With Dunning basis sets, larger grids do not improve the MAPEs. However, with pc basis sets, larger grids reduce the MAPEs and the MaxAPEs by a factor of $\sim$2. Therefore, when no CP corrections are included, default grid can be used with cc basis sets, while ultrafine pruned grids are recommended with the pc basis sets.

3.8.3 The Effect of the Grid Size on the PW86+PBE Energies in vdW Complexes

In this section, exchange-correlation energies computed with PW86 and PBE are compared to exchange-correlation energies obtained from PW86+PBE with *large* quadrupole zeta basis sets, aug-pc-3 and aug-cc-pVQZ. Statistical analysis for the PW86+PBE energies are listed in Table 3.3.

From Table 3.3, it is deduced that increasing the number of angular points increases the computational cost without improving the statistical data. However, increasing the number radial shells drastically improves the statistical data. This is clear when e.g. results from the 056302 grid are compared to those from the 100302 grid.

Correlation consistent basis sets, especially aug-cc-pVTZ, obviously give better MAPEs than polarization consistent basis sets when exchange-correlation energies are calculated without CP correction. Thus the use of the aug-cc-pVTZ basis set is recommended for computing accurate PW86+PBE energies.

Table 3.3: Statistical analysis of the PW86+PBE energies using multiple basis sets and grid sizes with and without counterpoise correction.

		Witout CP Correction			With CP Correction		
Basis Set	Grid	MPE	MAPE	MaxAPE	MPE	MAPE	MaxAPE
aug-cc-pVDZ	056302 U	-324.70	397	1752	97.29	104	376
	100302 U	-174.54	183	683	100.42	102	383
	099590 U	-177.13	185	672	100.54	102	397
aug-cc-pVTZ	056302 U	-176.52	259	1129	87.19	87	329
	100302 U	-16.52	29	130	89.92	90	343
	099590 U	-21.72	32	137	89.71	90	355
aug-pc-1	056302 U	-921.36	1003	3146	117.37	136	242
	100302 U	-800.41	803	2884	119.82	124	194
	099590 U	-808.73	811	2896	118.85	123	193
aug-pc-2	056302 U	-343.36	709	3146	185.89	161	466
	100302 U	-250.51	251	682	189.08	189	481
	099590 U	-257.24	257	689	186.47	186	472

3.9 Conclusions

Double and triple zeta Dunning basis sets are the most reliable basis sets to use for computing repulsive and binding energies in rare-gas diatomics. Under certain conditions, polarization consistent basis sets give slightly better statistical values than correlation consistent basis sets. However, the differences are not worth importing the basis set from an external reference. In addition, pc basis sets are not available for all atoms. Therefore, for the next study presented in the following chapter, double and triple zeta Dunning basis sets will be used with the unpruned 200590 grid. The counterpoise correction for the basis set superposition error is impractical and thus will not be used in the following studies.

Chapter 4

Assessment of the PW86+PBE+XDM Density Functional on van der Waals Complexes at Equilibrium Geometries

A. D. Becke, A. A. Arabi, and F. O. Kannemann, Can. J. Chem. **88**, 1057, (2010).

4.1 Abstract

The purpose of this chapter is to describe a new nonempirical GGA functional which can accurately capture interactions weaker than hydrogen-bonding, e.g., dispersion interactions. The basis set study assessed (in Chapter 3) is considered in the development of this nonempirical functional. The functional developed is a sum of three

terms: PW86 functional for exchange energy [26], PBE functional for correlation energy [28] and XDM model for dispersion energy [65, 66, 113]. This choice of the PW86+PBE+XDM functionals comprises a nonempirical density-functional theory of high accuracy for thermochemistry and van der Waals complexes, as demonstrated by Kannemann and Becke [74]. In [74], Kannemann and Becke tested the functional with self-consistent LDA orbitals using the basis-set-free NUMOL code [137, 138]. In this study, the functional is tested with self-consistent GGA orbitals obtained from the basis-set dependent G09 package [118].

4.2 Introduction

The accuracy of functionals that capture weak interactions is greatly dependent on the choice of the exchange functional [29], especially at the limit of large reduced density gradient ($s_\sigma = \frac{|\bigtriangledown\rho|}{\rho^{4/3}}$). Comparing the available exchange functionals in the literature to the HF exact exchange, PW86 and "B86b" (by Becke [24]) were reported to reproduce the HF repulsive energies most accurately (with PW86 being slightly superior to the B86b) [27, 29, 30, 31]. As demonstrated in the previous chapter, the mean absolute percent error of the self-consistent PW86 with a triple zeta basis set is 13% when compared to the numerical HF repulsive energies. Therefore, PW86 was used to account for the exchange energies. The PBE correlation functional was chosen because it is a parameter free functional and performs well when combined with PW86 and XDM [30, 74]. The XDM model was used [66, 139], also because it is a nonempirical functional and is relatively simple compared to other models (as discussed in Chapter 2). The nonempirical XDM model exploits non-sphericity of the exchange hole around its reference point to generate C_6, C_8, and C_{10} interatomic dispersion coefficients.

Kannemann and Becke tested the binding energy curves using PW86+PBE+XDM for

ten rare gas diatomics [30] and 65 vdW complexes [74]. They demonstrated that this functional reproduces excellent binding-energy curves. However, they computed the post-LDA energies, i.e., they used the numerical program NUMOL, which computes the exchange and correlation energies self-consistently with the LDA orbitals. Then, by using a "post-LDA" code to extract the necessary information from the wavefunction files, dispersion energies are calculated and added to the exchange-correlation energies. In this study, "post-GGA" binding-energy curves are assessed. This means the exchange-correlation energies are computed self-consistently using the Gaussian package, then the necessary data, e.g., dispersion coefficients, are extracted from the wavefunction file to compute the dispersion energies. The "post-GGA" method has advantages over the "post-LDA" method. With "post-GGA", forces suitable for geometry optimization of intermolecular complexes can be obtained. However, forces cannot be obtained from the NUMOL "post-LDA" method. Self-consistent NUMOL GGA calculations are not yet feasible because NUMOL numerical grid noise in the GGA exchange-correlation potential is too large for application to dispersion-bound systems.

The total energy in this study is given by

$$E_{total}^{GGA} = E_X^{PW86} + E_C^{PBE} + E_{Disp}^{XDM}, \tag{4.2.1}$$

where the exchange-correlation energy is calculated with G09 and the dispersion is calculated using the post-GGA code.

4.3 Binding Energies of Intermolecular Complexes

Although XDM generates asymptotic dispersion coefficients without empirical fit parameters, two (universal) fit parameters are required in order to damp the dispersion energy at small internuclear separations (as explained in Chapter 2). A brief review

of the damped dispersion energy is presented in this section.

$$E_{disp}^{XDM} = -\frac{1}{2}\sum_{i\neq j}\left(\frac{C_{6,ij}}{R_{vdW,ij}^{6}+R_{ij}^{6}}+\frac{C_{8,ij}}{R_{vdW,ij}^{8}+R_{ij}^{8}}+\frac{C_{10,ij}}{R_{vdW,ij}^{10}+R_{ij}^{10}}\right) \tag{4.3.1}$$

with the dispersion coefficients $C_{6,ij}$, $C_{8,ij}$, $C_{10,ij}$ [66] and [64] given by (2.5.9), (2.5.10) and (2.5.11) in section 2.5. The "Becke-Roussel", BR, variant of XDM [64] is used as the BR variant was found to be superior to the "exact exchange", XX, variant [74]. The effective vdW interatomic separations, $R_{vdW,ij}$, are linearly related to "critical" interatomic separations $R_{c,ij}$ by

$$R_{vdW,ij} = a_1 R_{c,ij} + a_2 \tag{4.3.2}$$

with the critical separation $R_{c,ij}$ given by the average of $(C_{8,ij}/C_{6,ij})^{1/2}$, $(C_{10,ij}/C_{6,ij})^{1/4}$ and $(C_{10,ij}/C_{8,ij})^{1/2}$. The two damping parameters, a_1 and a_2, were fit to benchmark binding energies of 65 intermolecular complexes obtained from [74]. The reference binding energies range from 0.022 kcal mol^{-1} (He_2) to 20.65 kcal mol^{-1} (uracil dimer), i.e., spanning three orders of magnitude.

The set of 65 complexes is listed in Table 4.1 along with their reference binding energies. This set includes the ten rare-gas diatomics with reference data from Tang and Toennies [133], 21 complexes from the test set of Johnson and Becke [65], the 22 complexes of the "S22" biochemical benchmark set of Jurecka *et al.* [140], and 12 complexes from Zhao and Truhlar [141, 142]. Cartesian coordinate files of all the complexes and their monomers are available as supplementary data to [74].

With the "post-GGA" approach, self-consistent PW86+PBE orbitals were computed with G09 using aug-cc-pVDZ and aug-cc-pVTZ basis sets, without CP corrections, and an unpruned grid size of 200 radial x 590 angular points as motivated in Chapter 3. The XDM dispersion energy was then computed non-self-consistently by an interface program, "post-G09", [143] that reads G09 orbital information from "wfn" files. As

supported by Kong *et al.* [144], ignoring self-consistency in the E_{disp}^{XDM} term is of negligible consequence since the dispersion energy is a very small fraction of the total energy.

The binding energies are calculated using (4.2.1). Binding energies calculated numerically [74], with a double zeta, and with a triple zeta are listed in Table 4.1. Binding energies calculated in this study using Dunning basis sets are quite similar to those of the numerical post-LDA method by Kannemann and Becke [74]. The present mean absolute percent errors (MAPEs) with the aug-cc-pVDZ and aug-cc-pVTZ basis sets are 16.6 % and 11.5 %, respectively. MAPE using the numerical method is 12.6 % [74]. Best-fit values (minimum RMS percent error) of the damping parameters are

a_1=0.80 , a_2=1.49 Å (SCF aug-cc-pVDZ)

a_1=0.79 , a_2=1.36 Å (SCF aug-cc-pVTZ)

a_1=0.82 , a_2=1.16 Å (numerical post-LDA [74]).

The damping parameters, especially a_1, are similar for the basis-set and the numerical results. The quality of the double-zeta results, though not as good overall as triple-zeta, is pleasantly surprising. Even the rare-gas diatomic binding energies involving Ne are in good agreement with reference data, despite the notably poor repulsion energies discussed in Chapter 3. The under-repulsive tendency of the DZ basis set compared to TZ (see Table 3.1) is apparently compensated by a slightly larger best-fit a_2 parameter. The only significant qualitative failure of the double-zeta method is the parallel benzene dimer (the last entry in Table 4.1), for which only one third of the reference binding energy is obtained. This failure is not observed in the numerical or triple-zeta results.

Table 4.1: Binding energies (in kcal mol^{-1}) of vdW intermolecular complexes.

Complex	Data Set[a]	Reference[b]	Numerical[c] (post-LDA)	TZ[d] (post-GGA)	DZ[e] (post-GGA)
He-He	TT	0.022	0.018	0.021	0.015
He-Ne	TT	0.041	0.048	0.051	0.050
He-Ar	TT	0.059	0.059	0.058	0.052
He-Kr	TT	0.063	0.065	0.064	0.055
Ne-Ne	TT	0.084	0.100	0.088	0.097
Ne-Ar	TT	0.132	0.143	0.133	0.132
Ne-Kr	TT	0.141	0.164	0.154	0.148
Ar-Ar	TT	0.285	0.255	0.245	0.217
Ar-Kr	TT	0.333	0.311	0.302	0.256
Kr-Kr	TT	0.400	0.381	0.378	0.311
He-N2 L shaped	JB	0.053	0.040	0.046	0.035
He-N2 T shaped	JB	0.066	0.062	0.066	0.082
He-FCl	JB	0.097	0.077	0.080	0.096
FCl-He	JB	0.182	0.157	0.181	0.191
CH4-C2H4	JB	0.50	0.67	0.67	0.75
CF4-CF4	JB	0.78	0.70	0.76	0.82
SiH4-CH4	JB	0.81	0.88	0.90	1.13
CO2-CO2	JB	1.37	1.15	1.19	1.07
OCS-OCS	JB	1.40	1.38	1.35	1.19
C10H8-C10H8 parallel	JB	3.78	4.5	4.55	3.98
C10H8-C10H8 parallel crossed	JB	5.28	5.85	5.92	5.28
C10H8-C10H8 T shaped	JB	4.34	4.46	4.62	4.68
C10H8-C10H8 T shaped crossed	JB	3.09	3.50	3.59	3.63
CH4-NH3	JB	0.73	0.97	0.96	1.12
SiH4-HF	JB	0.73	0.62	0.62	0.91
CH4-HF	JB	1.65	1.76	1.79	1.94
C2H4-HF	JB	4.47	5.16	5.21	5.27
CH3F-CH3F	JB	2.33	2.16	2.16	2.17
H2CO-H2CO	JB	3.37	2.99	3.00	3.09
CH3CN-CH3CN	JB	6.16	6.13	6.18	6.2
HCN-HF	JB	7.30	7.71	7.83	7.91

Complex	Data Set[a]	Reference[b]	Numerical[c] (post-LDA)	TZ[d] (post-GGA)	DZ[e] (post-GGA)
(NH3)2 [C2h]	S22	3.17	3.24	3.23	3.29
(H2O)2 [Cs]	S22	5.02	5.23	5.22	5.25
Formic acid dimer [C2h]	S22	18.61	19.13	19.14	18.99
Formamide dimer [C2h]	S22	15.96	16.09	16.08	16.20
Uracil dimer [C2h]	S22	20.65	20.28	20.45	20.70
2-pyridoxine-2-aminopyridine [C1]	S22	16.71	17.44	17.53	17.89
Adenine-thymine WC [C1]	S22	16.37	16.68	16.89	17.17
(CH4)2 [D3d]	S22	0.53	0.54	0.53	0.77
(C2H4)2 [D2d]	S22	1.51	1.36	1.38	1.55
Benzene-CH4 [C3]	S22	1.50	1.48	1.48	1.50
Benzene dimer parallel displaced [C2h]	S22	2.73	2.63	2.64	2.63
Pyrazine dimer [Cs]	S22	4.42	3.69	3.62	3.64
Uracil dimer stack [C2]	S22	10.12	8.78	8.71	8.47
Indole-benzene stack [C1]	S22	5.22	4.27	4.23	4.16
Adenine-thymine stack [C1]	S22	12.23	10.07	9.96	9.82
Ethene-ethyne [C2v]	S22	1.53	1.71	1.76	1.92
Benzene-H2O [Cs]	S22	3.28	3.17	3.19	3.12
Benzene-NH3 [Cs]	S22	2.35	2.22	2.24	2.24
Benzene-HCN [Cs]	S22	4.46	4.15	4.30	4.26
Benzene dimer T shaped [C2v]	S22	2.74	2.40	2.52	2.79
Indole-benzene T shaped [C1]	S22	5.73	4.95	5.16	5.42
Phenol dimer [C1]	S22	7.05	6.52	6.55	6.65
HF-HF	ZT	4.57	4.91	4.86	4.90
NH3-H2O	ZT	6.41	7.12	7.17	7.45
H2S-H2S	ZT	1.66	2.02	2.25	2.35
HCl-HCl	ZT	2.01	2.43	2.45	2.50
H2S-HCl	ZT	3.35	4.43	4.53	4.76
CH3Cl-HCl	ZT	3.55	4.06	4.08	4.16
HCN-CH3SH	ZT	3.59	4.20	4.21	4.44
CH3SH-HCl	ZT	4.16	6.14	6.22	6.59
CH4-Ne	ZT	0.22	0.14	0.17	0.24
C6H6-Ne	ZT	0.47	0.35	0.35	0.37
C2H2-C2H2	ZT	1.34	1.64	1.49	1.57
C6H6-C6H6 parallel	ZT	1.81	1.49	1.50	0.56

a) Tang-Toennies (TT) [133], Johnson-Becke (JB) [65], S22 [140], Zhao-Truhlar (ZT) [141, 142].

b) Reference binding energy from respective data set.

c) Numerical post-LDA binding energy from [74].

d) Present aug-cc-pVTZ post-GGA binding energy.

e) Present aug-cc-pVDZ post-GGA binding energy.

Table 4.2: Statistical errors on the binding energies of the 65 vdW intermolecular complexes.

Statistics	Numerical[a] (post-LDA)	TZ (post-GGA)	DZ
MAE[b](kcal mol^{-1})	0.33	0.34	0.39
MAPE[c] (%)	12.60	11.50	16.60
MaxAPE[d] (%)	47.6	49.5	69.1

a) Numerical post-LDA statitstics from [74].
b) MAE is mean absolute error.
c) MAPE is mean absolute percent error.
d) MaxAPE is maximum absolute percent error.

4.4 Conclusions

The PW86+PBE+XDM is a promising DFT for vdW complexes with a wide range of binding energies from 0.022 kcal mol^{-1} for helium dimer to 20.65 kcal mol^{-1} in uracil dimer. The mean absolute error on the 65 complexes is 0.34 kcal mol^{-1} (i.e., 11.5 % error) with the augmented Dunning triple zeta basis set. The post-LDA approach [30, 74], while useful for assessing the performance of density functionals on energetics, cannot provide forces for optimization of geometries. In this study, a methodology was tested with the promise to provide energies, and forces (as shall be shown in a later chapter), at reasonable computational cost.

Chapter 5

Assessment of the PW86+PBE+XDM Density Functional on van der Waals Complexes at Non-Equilibrium Geometries

5.1 Abstract

As mentioned in the introductory chapter, the deficiency of conventional DFT in properly describing van der Waals (especially dispersion-bound) complexes has been extensively addressed in the past decade. There are now several new methods published in the literature that are capable of accurately capturing weak dispersion inter-

actions in complexes at *equilibrium* geometries. However, the performance of these new methods at *non-equilibrium* geometries remains to be assessed. As mentioned in the previous chapters, PW86+PBE+XDM for exchange + correlation + dispersion, respectively, is a highly accurate functional for general thermochemistry [74] and vdW complexes at equilibrium geometries [145]. Here we show that this nonempirical functional also performs well for vdW complexes at compressed and stretched intermonomer separations. The mean absolute relative error (MARE) is 9.4% overall for vdW complexes in the "S22x5" database incorporating compressed and stretched geometries [146]. The largest MARE on the S22x5 database is 13.3% on the compressed geometry set.

5.2 Introduction

Dispersion is a weak interaction resulting from a long-range electron correlation effect. Conventional DFT methods (LDA, GGA, meta-GGA, or "hybrid functionals") are unable to describe van der Waals (vdW) interactions [55, 147]. Efforts have been made in the past decade to properly describe dispersion interactions as they are ubiquitous in materials sciences, biological systems and many chemical reactions. Many attempts to capture dispersion are purely empirical; some are rigorously derived from first principles. The former are conventional functionals with a large number of empirical parameters, e.g., X3LYP [78] and M05-2X/M06-2X [148]. The latter are physically derived, e.g., the vdW-DF nonlocal correlation functional of Langreth *et al.* [93, 100], the local response dispersion (LRD) model of Sato and Nakai [62] and the exchange-hole dipole moment (XDM) dispersion model of Becke and Johnson [64, 65, 66]. The functional PW86+PBE+XDM is a nonempirical GGA-based functional of high accuracy for general thermochemistry (mean absolute error of atomization energies of 222 molecules in the G3/99 data set is 10 kcal mol^{-1} [74]) and for *inter*- as well as *intra*-molecular vdW interactions [66, 74, 145]. As shown in the previous

chapter, the mean absolute error for the binding energies of 65 vdW complexes is 0.34 kcal mol^{-1} (MARE is 11.50%) when an augmented Dunning triple zeta basis set is used [145]. This result is remarkably good as these 65 complexes include a wide variety of complexes (H-bonded, dispersion-bound and mixed) with binding energies spanning three orders of magnitude, from 0.022 kcal mol^{-1} (He dimer, the weakest dispersion interaction possible) to 20.65 kcal mol^{-1} (uracil dimer) [13,14]. In addition, this nonempirical method achieves high accuracy at a negligible computational cost. The cost of the computation is basically from the GGA part, PW86+PBE, of the functional. The XDM dispersion term is computed in a "post-GGA" manner.

This functional, like the majority of other vdW functionals in the literature, has been tested on vdW complexes at equilibrium geometries. It is important, however, to consider vdW interactions at non-equilibrium compressed and stretched intermonomer separations. As pointed out by Molnar *et al.* [149] and Hobza *et al.* [146, 150], accurate description of dispersion in complexes at non-equilibrium geometries is important. In graphite, for instance, dispersion plays a role not only between consecutive graphene sheets, but non-consecutive sheets also [151]. In addition, in biological macromolecules such as DNA, vdW interactions are not limited to consecutively-stacked base pairs only.

The purpose of this study is to test the performance of the nonempirical vdW functional PW86+PBE+XDM on vdW complexes at non-equilibrium geometries. The database used for this study is the S22x5 set of Hobza *et al.* [146]. The S22x5 set is an extension of the S22 benchmark set [140] where the 22 complexes are considered at five different intermonomer separations: equilibrium, compressed to 90%, and stretched to 120%, 150% and 200% of the equilibrium separations.

5.3 Method and Computational Details

The PW86+PBE+XDM functional is the focus of all calculations is this study. The PW86+PBE exchange-correlation energy is calculated self-consistently using the G09 program [118]. After extracting orbital data from the wavefunction (wfn) file generated by G09, the XDM dispersion energy is calculated non-self-consistently in a post-GGA manner. In other words, the XDM dispersion energy is added perturbatively to the PW86+PBE exchange-correlation energy. The perturbative addition of dispersion to a conventional DFT functional has been shown to be a valid approximation [100, 144].

5.3.1 Choice of the Basis Set

As shown in Chapter 3, the Dunning triple zeta basis set, augmented with diffuse functions, is a good basis set for computing binding energies of vdW complexes with the PW86+PBE+XDM functional [145]. We used aug-cc-pVTZ with 200x590 atomic grids and an SCF convergence criterion of 10^{-6} throughout.

5.3.2 Choice of the Damping Parameters

Damping functions are needed to prevent divergences in the dispersion energy at small internuclear separations. In the XDM model, the damped dispersion energy is given by [66]:

$$E_{disp}^{XDM} = -\frac{1}{2}\sum_{i\neq j}\left(\frac{C_{6,ij}}{R_{vdW,ij}^{6}+R_{ij}^{6}}+\frac{C_{8,ij}}{R_{vdW,ij}^{8}+R_{ij}^{8}}+\frac{C_{10,ij}}{R_{vdW,ij}^{10}+R_{ij}^{10}}\right), \tag{5.3.1}$$

where $R_{vdW,ij}$ is linearly related to a "critical" internuclear separation $R_{c,ij}$ by:

$$R_{vdW,ij} = a_1 R_{c,ij} + a_2, \tag{5.3.2}$$

and a_1 and a_2 are two universal fit parameters. The critical radius $R_{c,ij}$ is the internuclear separation where the three dispersion terms are approximately equal:

$$\frac{C_{6,ij}}{R_{c,ij}^6} \approx \frac{C_{8,ij}}{R_{c,ij}^8} \approx \frac{C_{10,ij}}{R_{c,ij}^{10}}. \tag{5.3.3}$$

$R_{c,ij}$ is taken to be the average of $\left(\frac{C_{8,ij}}{C_{6,ij}}\right)^{1/2}$, $\left(\frac{C_{10,ij}}{C_{6,ij}}\right)^{1/4}$ and $\left(\frac{C_{10,ij}}{C_{8,ij}}\right)^{1/2}$.

The damping parameters can be fit by minimizing the root mean square percent error (RMSPE) of binding energies to suitable reference data. As presented in the previous chapter, at equilibrium geometries, the damping parameters are reasonably transferable regardless of the basis set used or whether the PW86+PBE part is calculated with LDA or GGA orbitals (i.e., "post-LDA" versus "post-GGA") [74, 145]. In this study, the transferability of these parameters was examined using several databases covering a wide variety of complexes at equilibrium and non-equilibrium separations. The damping parameters a_1 and a_2 were fit to multiple sets; namely S22 [140], S22x5 [146] and to the 65 complexes from [74]. For the remainder of this chapter, the parameters will be referred to as the Fit22, Fit22x5 and Fit65 parameters. The S22 set is a benchmark database of 22 vdW complexes of bio-organic type. It is a popular database commonly used for parameterizing or validating new computational methods for weak interactions. The S22x5 set was recently introduced by Hobza's group [74] as well. It includes the 22 complexes of the S22 set, but at five different intermonomer separations (one compressed by a factor of 0.9 relative to the equilibrium distance, one at equilibrium, and three stretched by factors of 1.2, 1.5 and 2.0). In other words, the S22x5 database includes five points along the dissociation curve of each complex. The set of 65 complexes used here was introduced by Kannemann and Becke [74]. It includes a wide variety of vdW complexes (dispersion, H-bonded and mixed complexes) with a wide range of binding energies (0.022 kcal mol^{-1} for He-He through 20.65 kcal mol^{-1} for the uracil dimer). See [74] for the reference binding

energies and geometry files of these 65 complexes.

In this work, fits denoted by "S22" are not fit at the reference geometries reported in [140], but rather at equilibrium geometries given by PW86+PBE+XDM. The binding energies of each complex are computed at the five intermonomer separations given in S22x5, then plot a potential energy curve for each complex. Using cubic splines, the potential energy curves were then interpolated to determine the minimum PW86+PBE+XDM binding energy. A similar interpolation scheme has been employed in [152, 64, 65, 153].

5.3.3 Choice of Methods for Comparison Purposes

Our PW86+PBE+XDM method is compared to several others, some of which are wavefunction theory (WFT) based and some density-functional theory (DFT) based.

Table 5.1: A list of selected DFT methods to be compared with our method PW86+PBE+XDM.

DFT Methods	Classification
TPSS/LP[a]	Nonempirical meta-GGA
M06-2X/aDZ[b]	Empirical meta-GGA
BLYP-D/TZVP	Pure GGA + Empirical D
PBE-D/TZVP	Pure GGA + Empirical D
B2PLYP/TZVPP	Double hybrid GGA
B2PLYP-D/TZVPP	Double hybrid GGA + Empirical D

a) LP is largest Pople basis set: 6-311++G(3df,3pd).
b) aDZ is augmented Dunning double zeta basis set, aug-cc-pVDZ.

Table 5.1 lists the DFT methods for comparison, from a larger list of DFTs tested by Hobza's group [146]. The wavefunction theory (WFT) methods selected are spin-component-scaled SCS-CCSD, SCS-MP2, and MP2.5. The reference binding energies from [74] are computed at the CCSD(T) level at the complete basis set limit. We report the mean absolute relative error (MARE) as indicative of the overall performance

of the method:

$$MARE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{BE_{method} - BE_{ref}}{BE_{ref}}\right| * 100, \tag{5.3.4}$$

N is the number of complexes in the set used, and BE_{ref} is the reference binding energy. We also report the mean relative error (MRE) to reveal the direction of the error with each method:

$$MRE = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{BE_{method} - BE_{ref}}{BE_{ref}}\right) * 100, \tag{5.3.5}$$

The root-mean-square error (RMSE) and root mean square percent error (RMSPE) is calculated as follows:

$$RMSPE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{BE_{method} - BE_{ref}}{BE_{ref}}\right)^2} * 100, \tag{5.3.6}$$

The mean absolute error (MAE) is commonly reported in the literature. For completeness, we therefore report MAEs as well:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|BE_{method} - BE_{ref}\right|, \tag{5.3.7}$$

5.4 Results and Discussion

5.4.1 Transferability of the Damping Parameters a_1 and a_2

The parameters a_1 and a_2 were shown in [145] to be transferable with use of different basis sets or the use of post-GGA versus post-LDA procedures. In this section, we test the transferability of the damping parameters with respect to fit sets which differ either by the complexes included in the set (S22 versus the 65 complexes), or by the intermonomer separations of the complexes in the set (S22 versus S22x5). As

explained above, the S22x5 set includes the S22 complexes at equilibrium, at one compressed, and at three stretched intermonomer separations.

Table 5.2: List of the values for the parameters a_1 and a_2 fit to three different sets.

Fit Set	a_1	a_2(Å)	RMSPE
S22	0.80	1.39	6.95
S22x5	0.66	1.58	15.93
65	0.79	1.36	13.72

Table 5.2 shows that the Fit22 and the Fit65 parameters are almost identical. This result confirms the transferability of the damping parameters for complexes at equilibrium, regardless of the complexes included in the fit set. This result is expected as the XDM model is derived from first principles and the dispersion computed using this model is environment dependent. The parameters change slightly when fit to the S22x5 set of complexes at non-equilibrium geometries. The parameter a_1 decreases from 0.80 to 0.66 and a_2 increases from 1.39 to 1.58 Å. The skew in the parameters arises mainly from the complexes at compressed intermonomer separations. As will be shown below, binding energies of complexes at compressed separations change dramatically with minor changes in the damping parameters. This high sensitivity to minor changes in the parameters is a result of the very steep potential energy curve at compressed separations.

5.4.2 Comparison of Methods

In this section, all of the statistics reported for the WFT methods, the DFT methods listed in Table 5.1, and the CCSD(T)/CBS reference binding energies are obtained from the supplementary information of [146] and [74]. Figure 5.1 displays the MARE (%) for the binding energies of the S22x5 complexes. The equilibrium distances are

scaled by factors of 0.9, 1.0, 1.2, 1.5 and 2.0 to get the compressed, equilibrium and three stretched separations.

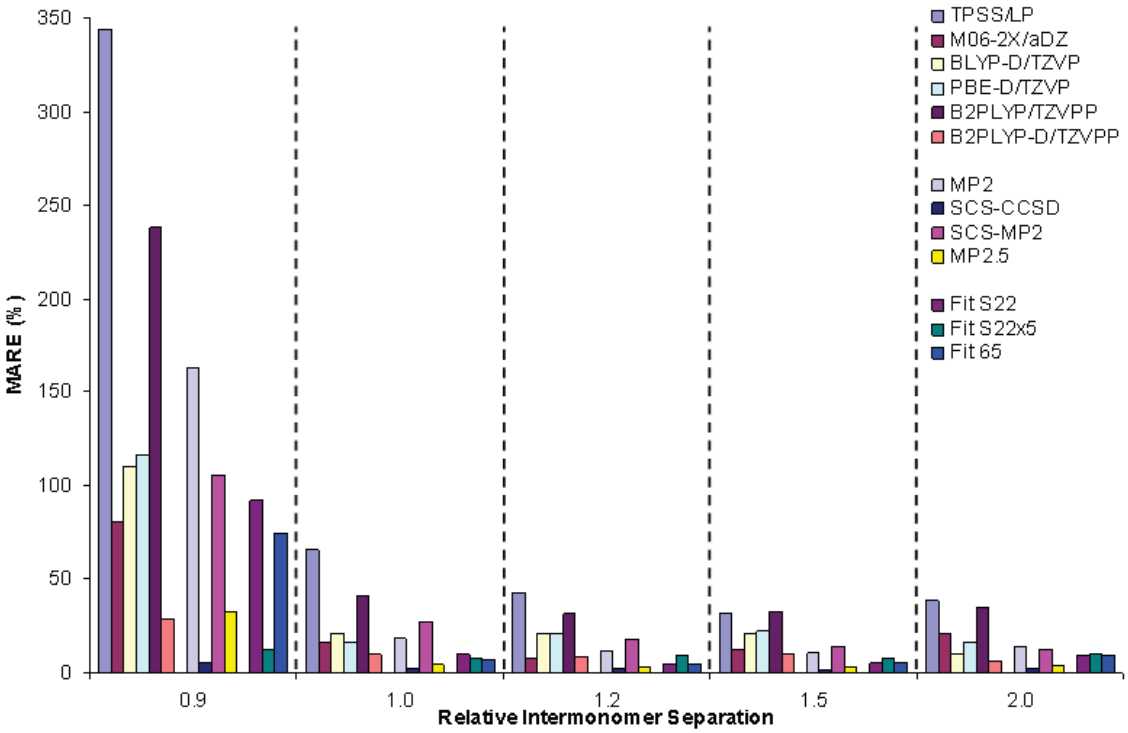


Figure 5.1: Mean absolute relative error [MARE (%)] for many methods used to compute the binding energies of 22 complexes at five different intermonomer separations.

Figure 5.1 clearly depicts poor performance for all methods at compressed separations, noting that the monomers are compressed only by 10% of the equilibrium separation. The worst situation is observed with the TPSS functional which is not designed to capture dispersion interactions. The MARE is 343.4%. The only two exceptions are the SCS-CCSD method (MARE is 4.7%), and our method PW86+PBE+XDM, with the Fit22x5 parameters, (MARE is 13.3%). The large MARE at shorter separations is not surprising since the potential energy curve is very steep in this region, and it is difficult to mimic the Hartree-Fock (HF) repulsion with a GGA functional [30]. The binding energies of the complexes at compressed intermonomer distances have almost 100% MARE with the Fit65 or FitS22 parameters.

Table 5.3 shows how the binding energies for the dispersion-bound complexes are severely below the reference binding energies with the Fit65 and Fit22 parameters.

Table 5.3: Binding energies for the S22 complexes at the compressed separations. The binding energies are reported using three sets of damping parameters, S22, S22x5 and the 65 complexes. HB, S and T stand for hydrogen-bonded, stacked and T shaped.

	Ref BE	BE with Fit 65 parameters	BE with Fit 22 parameters	BE with Fit 22x5 parameters	Complex
HB	2.410	2.507	2.456	2.682	ammonia dimer
	4.319	4.537	4.498	4.658	water dimer
	16.337	16.592	16.464	16.992	formic acid dimer
	14.141	14.125	13.996	14.543	formamide dimer
	18.729	18.474	18.334	18.924	uracil dimer HB
	15.126	15.861	15.695	16.387	2-pyridoxine-aminopyridine
	15.021	15.334	15.154	15.882	adenine-thymine WC
DISP	0.337	0.170	0.110	0.374	methane dimer
	0.681	0.203	0.088	0.648	ethene dimer
	1.088	0.863	0.752	1.235	benzene-methane
	0.148	-1.221	-1.598	0.174	benzene dimer S
	1.686	-0.755	-1.129	0.638	pyrazine dimer
	6.763	3.485	2.999	5.225	uracil dimer S
	1.429	-0.727	-1.229	1.119	indole-benzene S
	7.991	2.415	1.734	4.909	adenine-thymine S
MIX	1.174	1.528	1.476	1.716	ethene-ethyne
	3.007	2.712	2.605	3.08	benzene-water
	2.040	1.715	1.604	2.098	benzene-ammonia
	4.018	3.536	3.409	4.002	benzene-HCN
	2.204	1.557	1.395	2.143	benzene dimer T
	4.995	3.847	3.637	4.587	indole-benzene T
	6.422	5.631	5.462	6.213	phenole dimer

In contrast, when using the Fit22x5 parameters, the binding energies of the dispersion-bound complexes are reproduced quite well (except for the pyridine dimer and the stacked AT base pair). For the hydrogen-bonded complexes, the binding energies are reproduced accurately regardless of the fit. For the mixed complexes, the binding energies computed using S22 and 65 fits are acceptable. They are, however, more accurate using Fit22x5 parameters.

A quantitative comparison of the MARE of binding energies using the Fit22, Fit22x5

and Fit65 damping parameters is presented in Table 5.4. The statistics are classified by the relative intermonomer separation.

Table 5.4: MARE (%) for the binding energies computed with PW86+PBE+XDM with damping parameters fit to S22, S22x5 and the database of 65 complexes.

Relative Intermonomer Distances	Fit S22	Fit 65	Fit S22x5
0.9	91.87	74.13	13.25
1.0	9.71	7.09	7.52
1.2	4.00	4.52	8.86
1.5	4.73	5.48	7.14
2.0	9.05	9.28	9.99
All	23.87	20.10	9.35

As Table 5.4 quantifies, at compressed separations, Fit22x5 parameters significantly reduces the MARE of the binding energies by 60.88% and 78.62% compared to MARE with Fit65 and Fit22 parameters, respectively. With the Fit22x5 parameters, the MARE does not exceed 13.25% at all separations. For all three fits, the MARE remains almost unaltered at equilibrium and stretched geometries. With fits S22 and 65 (which have almost identical parameters), the MAREs are slightly better than those with Fit S22x5, especially at a relative separation of 1.2. The MAREs with the Fit65 parameters are only slightly better (compared to MAREs with the S22x5 fit) by 0.4, 4.3, 1.7 and 0.7 % at 1.0, 1.2, 1.5 and 2.0 relative separations, respectively (See Table 5.4). For completeness, mean absolute errors (MAE) are reported in Table 5.5.

Figure 5.2 is a zoom-in on Figure 5.1 in which MARE (%) is depicted for binding energies only at equilibrium and stretched separations (i.e., compressed separations are omitted).

This figure illustrates how, at each different separation, all three fits for the PW86+PBE

Table 5.5: MAE (kcal mol^{-1}) for the binding energies computed with PW86+PBE+XDM with damping parameters fit to S22, S22x5 and the database of 65 complexes.

Relative Intermonomer Distances	Fit S22	Fit 65	S22x5
0.9	1.13	0.98	0.51
1.0	0.53	0.44	0.35
1.2	0.01	0.01	0.02
1.5	0.10	0.12	0.02
2.0	0.04	0.04	0.04
All	0.40	0.36	0.28

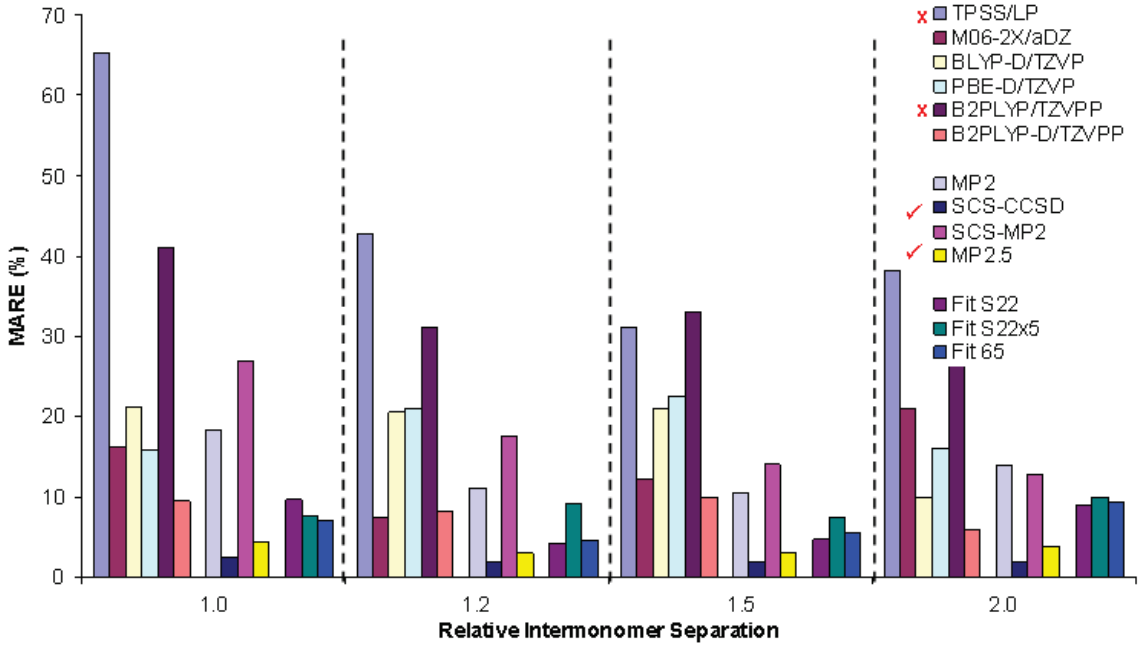


Figure 5.2: Mean absolute relative error [MARE (%)] for many methods used to compute the binding energies of 22 complexes at only equilibrium and stretched intermonomer separations. The "X" and the "check mark" indicate the two least and the two most reliable methods.

+XDM method have very similar MARE (%). For instance, at twice the equilibrium distances, the MAREs are 9.1, 10.0 and 9.3 % when the damping parameters are fit to S22, S22x5 and the 65 databases, respectively. The MAREs vary slightly at different separations, e.g., with the S22x5 fit, the MAREs are 7.5, 8.9, 7.1 and 10.0 for inter-

monomer separations scaled by 1.0, 1.2, 1.5 and 2.0, respectively. At all separations, the MARE is the smallest with SCS-CCSD and PW86+PBE+XDM (parameters fit to S22x5) while it is the greatest with TPSS and B2PLYP. The MARE obtained with PW86+PBE+XDM (parameters fit to S22x5) is 2.8 times greater than the MARE with the SCS-CCSD method. B2PLYP augmented with an empirical dispersion term (B2PLYP-D) is the only DFT method that has a MARE comparable with MP2.5 (second best among WFT methods), as can be seen in Figure 5.2.

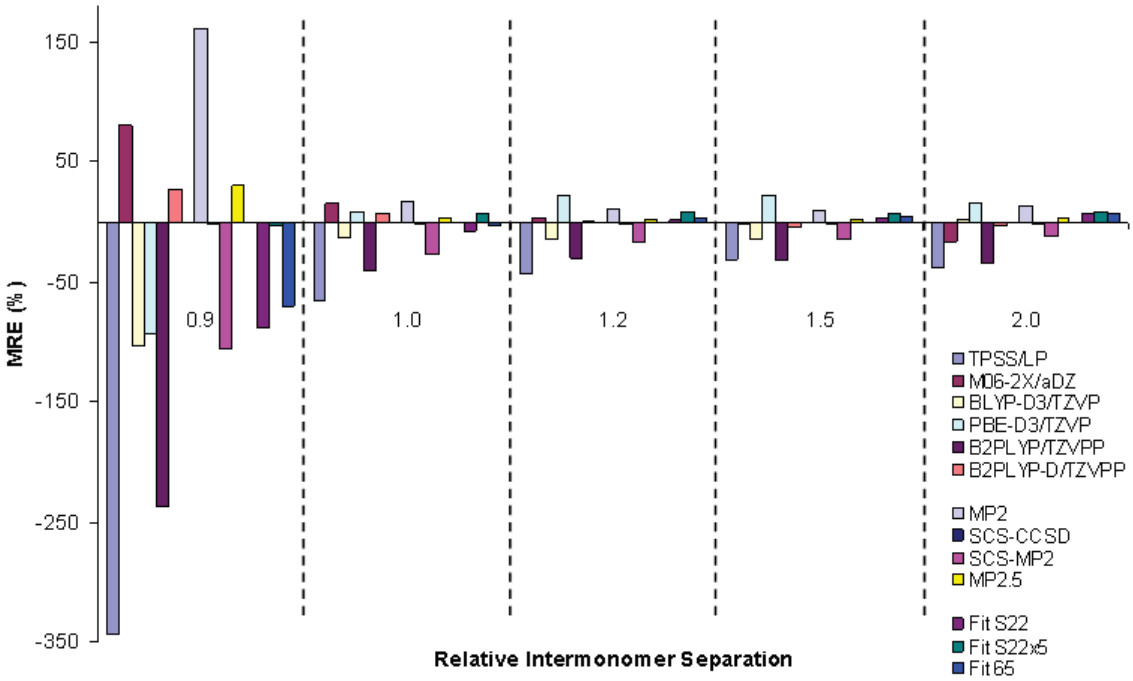


Figure 5.3: Mean relative error [MRE (%)] for many methods used to compute the binding energies of 22 complexes at five different intermonomer separations. Refer to [146] for details about the basis sets used with the wavefunction methods.

Figure 5.3 reveals the direction of the error for each method at the five separations. The largest MREs are observed at the compressed separations. Among the DFT methods, TPSS and B2PLYP gave the largest error, while MP2 and SCS-MP2 are the WFT methods that gave the largest errors. At compressed intermonomer separations, our method PW86+PBE+XDM has a MRE less than zero for all fits, the smallest (in absolute value) being -3.32% with the S22x5 fit. The negative MRE

might be due to the over-repulsion of the PW86 exchange functional at short distances [30]. At stretched separations, the binding energies are consistently slightly overestimated (1.53% to 9.09%). MP2, a strongly basis-set dependent method, consistently overestimates the binding energies at equilibrium and distorted separations. This is a well known result, namely at equilibrium geometries of dispersion-bound complexes. MP2.5 is the average of MP2 and MP3 methods. MP2.5 usually gives good results by an inherent cancellation of errors [146]. A positive MRE can be due to the dominant overestimation of MP2 over the underestimation of MP3. SCS-CCSD is a spin-component-scaled coupled cluster method, where the correlation energy is decomposed into parallel and anti-parallel spin components, each of which is multiplied by a parameter. SCS-CCSD gives very small MRE (1.2 - 1.7%), it is the best method among all methods tested in this study. SCS-MP2 [154] also scales the anti-parallel and the parallel spin correlation energies separately. However, the MRE with SCS-MP2 is large as this method was originally parameterized for reaction energies rather than intermolecular interactions. The SCS-CCSD method was also originally parameterized for reaction energies. Nevertheless, CCSD, to begin with, is more accurate than MP2 for intermolecular interactions. B2PLYP is a double hybrid functional: the exchange part is a hybrid of B88 [155] and HF; and the correlation part is a hybrid of LYP [34] and MP2. While the MP2 method is known to overbind dispersion-bound complexes, the GGA exchange functional B88 is over-repulsive especially at shorter separations. If the over-repulsion of the B88 functional dominates the overbinding of MP2, the MRE should be less than zero as it is in this study. The much smaller MRE with B2PLYP-D compared to B2PLYP is attributed to the dispersion correction added to B2PLYP-D. TPSS is a nonempirical meta-GGA functional that has no proper physics for capturing dispersion interactions. The MRE of TPSS is consistently underestimated, but it is the largest among all other methods considered. This result is expected as TPSS is not designed to capture dispersion interactions. M06-2X is a highly parameterized meta-GGA, which makes it difficult to explain any

of its trends. M06-2X overestimates dispersion, except for the complexes at 200% of the equilibrium separation.

5.5 Conclusions

In conclusion, the nonempirical functional PW86+PBE+XDM is an accurate functional for not only general thermochemistry and vdW complexes at equilibrium geometries, as previously shown [74, 145], but also for vdW complexes at non-equilibrium (compressed and stretched) intermonomer separations. The authors of [146] state, based on the DFT methods they tested, that DFT in general is not capable of properly describing vdW interactions at non-equilibrium geometries. However, this chapter demonstrates that (PW86+PBE+XDM) can indeed capture vdW interactions properly at equilibrium and non-equilibrium geometries. Binding energies of complexes at separations shorter than equilibrium separations are the most challenging to compute, but even in these compressed cases the mean absolute relative error can be as small as 13.3% for the complexes in the S22x5 database. Accurate binding energies at stretched geometries, however, are more important especially in biological systems. For stretched geometries, our method (with Fit22 parameters) can give mean relative error as small as 1.5, 4.1 and 7.3 at intermonomer distances scaled by a factor of 1.2, 1.5 and 2.0, respectively. This work can be extended to the new database, S66x8, released recently also by Hobza's group [150]. This database is a larger set of 66 complexes at 8 different intermonomer separations.

Chapter 6

Assessment of the PW86+PBE+XDM Density Functional on a Balanced Database of van der Waals Complexes at Non-Equilibrium Geometries

6.1 Abstract

In the previous chapter, the performance of the nonempirical functional for vdW complexes, PW86+PBE+XDM, was assessed on the S22x5 database. In this chapter, the purpose is still to assess the performance of this functional on vdW complexes at non-equilibrium geometries. However, the database used in this study, S66x8, is larger and more balanced. The S66x8, published by Hobza's group [150], contains 66 complexes at eight different separations ranging from compressing to 90% of the equilibrium geometry to stretching to 200% of the equilibrium geometry. The overall

root mean square percent error (RMSPE) on this lager database using aug-cc-pVTZ is 14.58%.

6.2 Introduction

In the past decade, the focus for benchmarking methods that capture weak interactions has been on databases with vdW complexes only at equilibrium geometries. For example, the S22 database [140] is one of the most popular databases for benchmarking newly developed methods. Recently, there has been interest in databases which include complexes not only at equilibrium geometries, but also at non-equilibrium geometries. It is important to be able to capture dispersion interactions properly in vdW complexes that are at non-equilibrium geometries as pointed out by Molnar *et al.* [149] and Hobza *et al.* [150, 146]. In biological systems, for instance in a folded protein, there are some *intra*-molecular dispersion interactions between amino acids that are not necessarily at equilibrium positions with respect to one another. S22x5 [146] and S66x8 [150] are examples of databases that include complexes at non-equilibrium geometries.

In the previous chapter, it was shown that the nonempirical functional PW86+PBE +XDM was the only DFT method, amongst many others (TPSS, M06-2X, BLYP-D, PBE-D, B2LYP and B2PLYP) to perform well at compressed geometries of the 22 complexes in the S22x5 database. Compressed vdW complexes are more challenging than stretched complexes as demonstrated in the previous chapter. Our smallest mean absolute relative error (MARE) on 22 compressed complexes (from the S22x5 database [146]) was 13.3%.

In this chapter, a larger, more balanced database (S66x8) [150] will be used to further test the accuracy of PW86+PBE+XDM on binding energies of vdW complexes at non-equilibrium geometries. The S66x8 has 66 complexes combined from 14 different

monomers. This database is an extension of the S22x5 which is, itself, an extension of the S22 database [140]. S66x8 does not include dimers with a binding energy smaller than 1.5 kcal mol^{-1}. The range of compression and stretching of intermonomer separations in S22x5 and S66x8 is the same. As in S22x5, the compression is no more than 10% and the stretching is no longer than twice the equilibrium separation. The difference, however, is in the 44 extra complexes in S66x8 and the number of intermonomer separations considered for each complex (5 and 8 for S22x5 and S66x8, respectively). The eight different separations are: equilibrium, compressed to 90% and 95%, and stretched to 105%, 110%, 125%, 150% and 200% of the equilibrium separations. The S66x8 is a more balanced database as it includes equal numbers of complexes in each of the three subcategories (23 hydrogen-bonding complexes, 23 dispersion-bound complexes and 20 mixed, i.e., electrostatic/dispersion, complexes). The 44 extra complexes (compared to S22x5) include some single hydrogen-bonding complexes which are otherwise absent in S22x5. In S66x8, the dispersion-bound complexes are not only $\pi - \pi$ stacked complexes (10) (as in S22x5) but there are also dispersion interactions between aliphatic hydrocarbon chains, which include aromatic aliphatic (8) and aliphatic aliphatic (5) interactions. The S66x8 is not heavily weighted toward nucleic acid-like structures as in S22 and S22x5. In the S66x8, the same basis set was used systematically for a final optimization of all complexes, resolution of the identity-MP2, with an augmented triple zeta basis set and with counterpoise correction, RI-MP2/aug-cc-pVTZ (cp). Binding energies were computed using the CCSD(T)/CBS method.

6.3 Method and Computational Details

In this study, the DFT exchange and correlation energies were computed with PW86 and PBE, respectively, using the G09 program [118]. The XDM dispersion correction was then added perturbatively to the exchange-correlation energy.

6.3.1 Choice of the Basis Set

For consistency, the basis set, grid and convergence criterion used for all calculations are identical set to what was used in the previous study on the S22x5. Thus, as in Chapter 3 and Ref [145], aug-cc-pVTZ was used with 200x590 atomic grids and an SCF convergence criterion of 10^{-6}.

6.3.2 Damping Parameters

As presented in Chapter 2, the dispersion energy can diverge at small separations. It is thus necessary to add a damping function to avoid this divergence. Grimme has shown that the form of the damping function is not as important as correctly parameterizing it [156]. As reiterated in the previous chapters, in the XDM model the damped dispersion energy is given by [66]:

$$E_{disp}^{XDM} = -\frac{1}{2}\sum_{i\neq j}\left(\frac{C_{6,ij}}{R_{vdW,ij}^6 + R_{ij}^6} + \frac{C_{8,ij}}{R_{vdW,ij}^8 + R_{ij}^8} + \frac{C_{10,ij}}{R_{vdW,ij}^{10} + R_{ij}^{10}}\right), \tag{6.3.1}$$

where the linear relationship between $R_{vdW,ij}$ and the"critical" internuclear separation $R_{c,ij}$ is given by:

$$R_{vdW,ij} = a_1 R_{c,ij} + a_2, \tag{6.3.2}$$

where a_1 and a_2 are universal damping parameters. The critical radius $R_{c,ij}$ is the internuclear separation where the three dispersion terms contribute equally to the dispersion energy:

$$\frac{C_{6,ij}}{R_{c,ij}^6} \approx \frac{C_{8,ij}}{R_{c,ij}^8} \approx \frac{C_{10,ij}}{R_{c,ij}^{10}}. \tag{6.3.3}$$

$R_{c,ij}$ is taken to be the average of $\left(\frac{C_{8,ij}}{C_{6,ij}}\right)^{1/2}$, $\left(\frac{C_{10,ij}}{C_{6,ij}}\right)^{1/4}$ and $\left(\frac{C_{10,ij}}{C_{8,ij}}\right)^{1/2}$.

In this study, as in previous chapters, the damping parameters are determined by minimizing the root mean square percent error (RMSPE) of binding energies of complexes

in a given database. We have observed that there is a strong correlation between a_1 and a_2 regardless of the database. Figure 6.1 depicts this relationship.

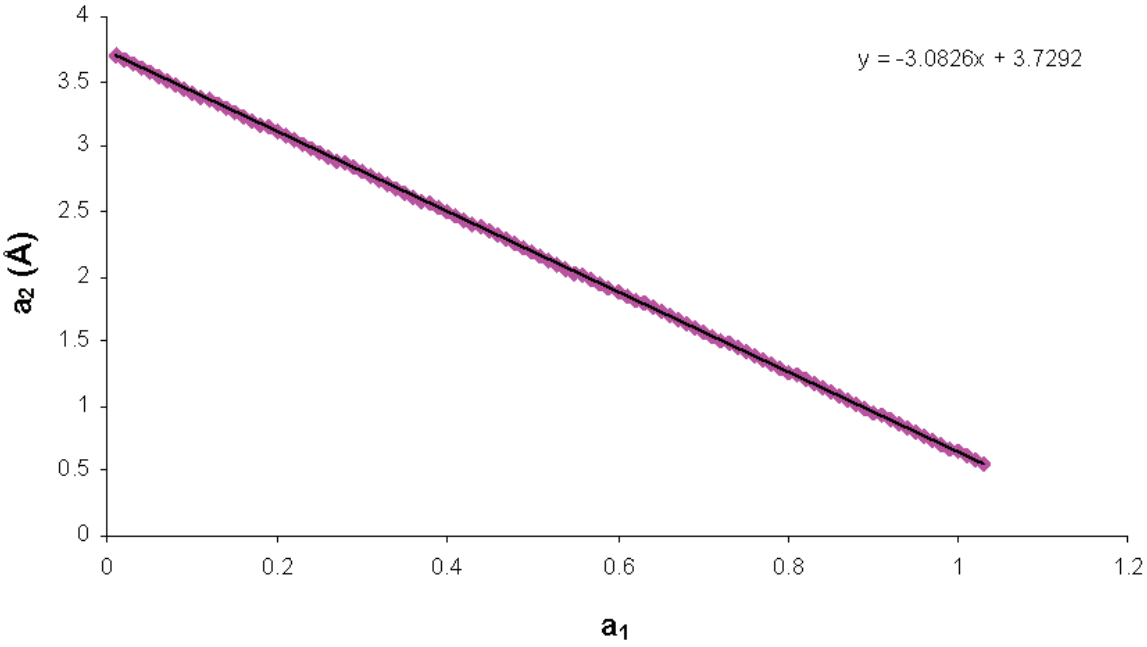


Figure 6.1: Plot of a_2 (Å) with respect to a_1. The parameters are determined by minimizing the RMSPE of the binding energies of the S22 complexes.

Figure 6.1 shows a strong linear relationship between a_1 and a_2 described by the following equation:

$$a_2 = -3.0826a_1 + 3.729 \tag{6.3.4}$$

The relationship between a_1 and a_2 determined by minimizing the RMSPE of the binding energies of the S22x5 complexes is given by:

$$a_2 = -3.1482a_1 + 3.668 \tag{6.3.5}$$

If a_1 and a_2 were determined by minimizing the RMSPE of the "interpolated" equilibrium binding energies from the S22x5 database, the relationship is given by:

$$a_2 = -3.0804a_1 + 3.852 \tag{6.3.6}$$

By "interpolated", it is meant that a potential energy surface for each of the 22 complexes was created by using the energies at five different intermonomer separations. Using cubic splines, the curve was interpolated to get the minimum PW86+PBE+XDM energy. A similar interpolation scheme has been employed in [64, 65, 153, 152].

These rather strong relationships (note the similar equations regardless of the database chosen) demonstrate how a_2 decreases as a_1 increases. Fitting the damping parameters to S66x8 resulted in a RMSPE of 14.58% with $a_1 = 0.60$ and $a_2 = 1.84$ Å. These damping parameters are comparable to those obtained by fitting to S22 [140], S22 interpolated (Arabi A. A. and Becke A. D, submitted to JCP), S22x5 [146] and the 65 complexes of Kannemann and Becke [74] as shown in Table 6.1.

Table 6.1: List of the values for the fit parameters a_1 and a_2.

	Fit Set	a_1	a_2(Å)	MARE
Numerical[a]	65	0.82	1.16	12.6
aug-cc-pVTZ	65	0.79	1.36	11.9
aug-cc-pVDZ	65	0.80	1.49	16.6
aug-cc-pVTZ	S22	0.70	1.57	5.3
aug-cc-pVTZ	S22 interpolated	0.80	1.39	5.6
aug-cc-pVTZ	S22x5	0.66	1.58	9.8
aug-cc-pVTZ	S66	0.72	1.57	5.8
aug-cc-pVTZ	S66x8	0.60	1.84	9.1

a) All calculations are post-GGA except for this one which is post-LDA from [74].

The equations for mean absolute relative error (MARE), mean relative error (MRE), root mean square error (RMSE) and root mean square percent error (RMSPE) have been already explicitly written in Chapter 5.

6.4 Results and Discussion

The 66 complexes (at equilibrium) in S66x8 along with their reference binding energies, the binding energies computed with PW86+PBE+XDM/aug-cc-pVTZ and their classification based on the dominant type of interaction are summarized in Table 6.2. The damping parameters used for computing the S66 (at equilibrium) complexes are $a_1 = 0.72$ and $a_2 = 1.57$ (as shown in Table 6.1). For this set of 66 complexes at equilibrium, RMSPE is 7.24% and MARE is 5.82 %.

Table 6.2: List of the 66 complexes at equilibrium with binding energies (in kcal mol^{-1}) and classification based on type of interaction. E, D and M stand for Electrostatic, Dispersion and Mixed, respectively. Reference BEs are obtained from [150].

Complex	Ref BE	Calculated BE	Classification
Water-Water	4.894	5.130	E
Water-MeOH	5.569	5.687	E
Water-$MeNH_2$	6.875	7.619	E
Water-Peptide	8.075	8.095	E
MeOH-MeOH	5.745	5.865	E
MeOH-$MeNH_2$	7.540	8.194	E
MeOH-Peptide	8.220	8.460	E
MeOH-Water	4.997	5.200	E
$MeNH_2$-MeOH	3.035	3.163	M
$MeNH_2$-$MeNH_2$	4.152	4.276	M
$MeNH_2$-Peptide	5.406	5.304	M
$MeNH_2$-Water	7.245	7.718	E
Peptide-MeOH	6.180	6.011	E
Peptide-$MeNH_2$	7.451	7.782	E
Peptide-Peptide	8.624	8.582	E
Peptide-Water	5.118	5.022	E
Uracil-Uracil, BP	17.182	17.544	E
Water-Pyridine	6.834	7.526	E
MeOH-Pyridine	7.404	8.079	E
AcOH-AcOH	19.090	19.816	E
$AcNH_2$-$AcNH_2$	16.260	16.552	E

Complex	Ref BE	Calculated BE	Classification
AcOH-Uracil	19.491	19.908	E
$AcNH_2$-Uracil	19.190	19.374	E
Benzene-Benzene, pi-pi	2.736	2.735	D
Pyridine-Pyridine, pi-pi	3.825	3.550	D
Uracil-Uracil, pi-pi	9.821	8.732	M
Benzene-Pyridine, pi-pi	3.369	3.212	D
Benzene-Uracil, pi-pi	5.705	4.907	D
Pyridine-Uracil, pi-pi	6.814	5.937	M
Benzene-Ethene	1.412	1.645	D
Uracil-Ethyne	3.737	3.543	M
Pyridine-Ethene	1.861	1.992	D
Pentane-Pentane	3.774	3.890	D
Neopentane-Pentane	2.611	2.845	D
Neopentane-Neopentane	1.772	2.132	D
Cyclopentane-Neopentane	2.405	2.673	D
Cyclopentane-Cyclopentane	2.997	3.234	D
Benzene-Cyclopentane	3.569	3.555	D
Benzene-Neopentane	2.894	3.030	D
Uracil-Pentane	4.839	4.413	D
Uracil-Cyclopentane	4.132	3.801	D
Uracil-Neopentane	3.700	3.510	D
Ethene-Pentane	1.989	2.031	D
Ethyne-Pentane	1.749	1.911	D
Peptide-Pentane	4.244	4.023	D
Benzene-Benzene, TS	2.866	2.648	D
Pyridine-Pyridine, TS	3.533	3.180	D
Benzene-Pyridine, TS	3.324	3.055	D
Benzene-Ethyne, CH-pi	2.864	2.733	M
Ethyne-Ethyne, TS	1.523	1.755	M
Benzene-AcOH, OH-pi	4.704	4.372	M
Benzene-$AcNH_2$, NH-pi	4.358	4.154	M
Benzene-Water, OH-pi	3.268	3.185	M
Benzene-MeOH, OH-pi	4.187	4.047	M
Benzene-$MeNH_2$, NH-pi	3.231	3.109	D
Benzene-Peptide, NH-pi	5.280	4.967	M
Pyridine-Pyridine, CH-N	4.146	3.984	M

Complex	Ref BE	Calculated BE	Classification
Ethyne-Water, CH-O	2.847	2.941	E
Ethyne-AcOH, OH-pi	4.860	5.161	E
Pentane-AcOH	2.879	2.845	D
Pentane-$AcNH_2$	3.505	3.483	D
Benzene-AcOH	3.801	3.495	D
Peptide-Ethene	2.988	2.928	M
Pyridine-Ethyne	3.990	4.525	E
$MeNH_2$-Pyridine	3.966	3.884	M

In Table 6.2, the maximum percent error is 20.3 % for the aliphatic-aliphatic dispersion-bound Neopentane-Neopentane complex.

6.4.1 Comparison of Results Between S22x5 and S66x8

In this subsection, the results from S22x5 (from the previous chapter) and S66x8 will be compared by looking at the mean absolute relative error of the 22 and 66 complexes at different intermonomer separations.

Table 6.3: MARE (%) for the binding energies of S22x5 and S66x8 computed with PW86+PBE+XDM.

Relative Intermonomer Distances	S66x8	S22x5
0.90	15.81	13.25
0.95	7.99	NA
1.00	7.54	7.52
1.05	8.03	NA
1.10	8.53	NA
1.20	NA	8.86
1.25	8.79	NA
1.50	7.51	7.14
2.00	8.17	9.99
All	9.05	9.35

The damping parameters used are (as displayed in Table 6.1) $a_1 = 0.66$ and $a_2 = 1.58$ for the S22x5 and $a_1 = 0.60$ and $a_2 = 1.84$ for the S66x8 databases. If the damping parameters are interchanged, the MARE for S66x8 would increase (by 3%) to 12.1% and the RMSPE for S22x5 increases (also by 3%) to 12.6%.

Figure 6.2 illustrates the difference in the MARE of binding energies in S22x5 compared to the MARE in S66x8, at the various non-equilibrium geometries.

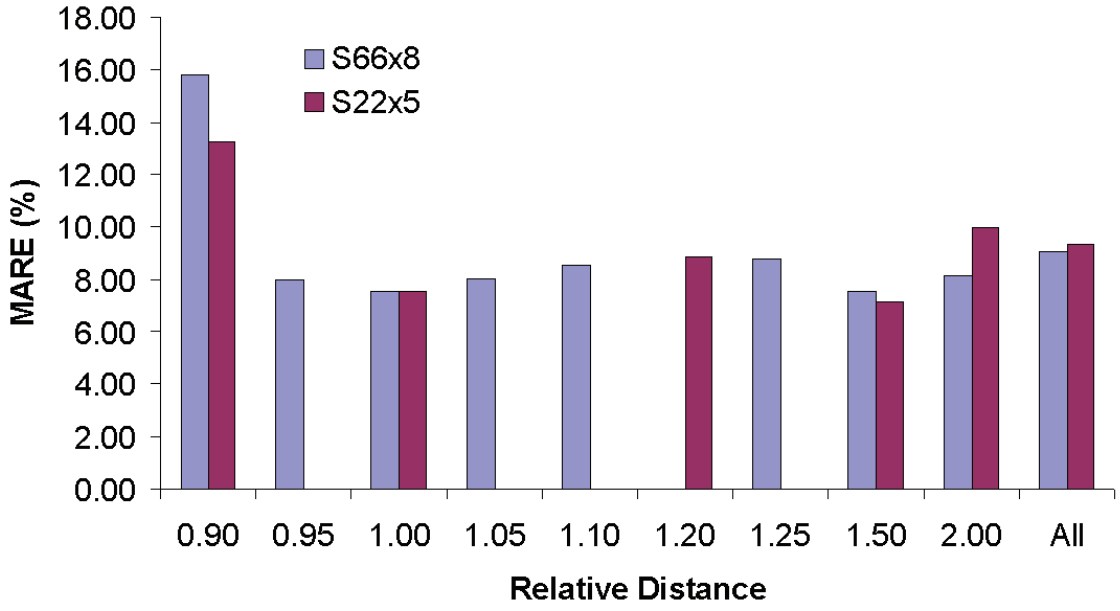


Figure 6.2: Mean absolute relative error [MARE (%)] for the binding energies of S22x5 and S66x8 complexes at different intermonomer separations.

As discussed in Chapter 5, and as shown in Table 6.3 and Figure 6.2, the binding energies of vdW complexes at compressed distances is more challenging than equilibrium and stretched distances. The test on the S66x8 database demonstrates that for vdW complexes at equilibrium and those compressed by 5%, the statistical error on the binding energies is almost identical (see Table 6.3 and Figure 6.2). This means the challenge starts when complexes are compressed beyond 5%. This challenge arises from the steep repulsive behavior of the exchange functional at shorter distances [30]. Figure 6.2 illustrates that the statistical errors for the binding energies in S22x5 and S66x8 are comparable, with a MARE slightly smaller overall for the

S66x8. Despite the larger error at the shortest separations, the overall error is close to what is observed at equilibrium geometries. This demonstrates that the damping parameters are roughly transferable at equilibrium and stretched geometries but not at compressed geometries as in this region a minuscule change in parameters causes a significant change in the binding energies.

6.4.2 Consequences of Fitting to S66x8

By fitting the parameters to S66x8, a higher accuracy is achieved for the S66x8 overall including the challenging compressed cases. However, there is a slight sacrifice in the accuracy of binding energies of complexes in the other databases as shown in Table 6.4. In this table, the MARE listed for each database is for the damping parameters fit to S66x8, i.e., $a_1 = 0.60$ and $a_2 = 1.84$ (see Table 6.1). ΔMARE denotes the difference between the smallest MARE achieved by fitting the parameters to the set itself (i.e., parameters listed in Table 6.1) and the MARE listed in this table (i.e., using parameters fit to S66x8).

Table 6.4: MARE (and difference in MARE) for the binding energies computed with PW86+PBE+XDM using different parameters.

	MARE at FitS66x8 parameters	a_1	a_2	ΔMARE
65	12.02	0.79	1.36	1.71
S22	5.79	0.70	1.57	0.53
S22x5	12.61	0.66	1.59	2.78
S66	7.54	0.72	1.57	1.72
S66x8	9.05	0.60	1.84	0.00

The increase in the statistical error (ΔMARE) can be as small as 0.53% (for S22) and as large as 2.78% for the S22x5 database. Nevertheless, even 2.78% is not a very large difference in error.

6.4.3 Comparison of Methods

In this section, the performance of PW86+PBE+XDM is compared to that of wave-function theory (WFT) methods as depicted in Figure 6.3. The RMSE for the WFT methods were obtained from the supplementary information of [150].

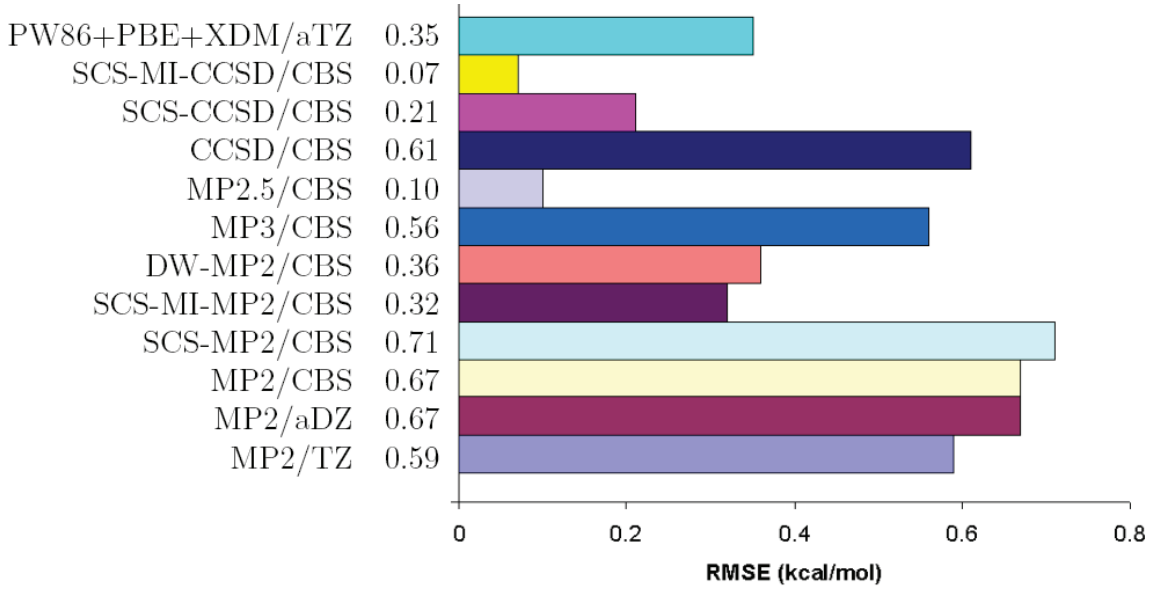


Figure 6.3: RMSE (kcal mol^{-1}) for many methods used to compute the binding energies of the S66x8 complexes.

Figure 6.3 depicts the RMSE for each method, the statistical error is explicitly displayed for each method. This figure illustrates how PW86+PBE+XDM/aug-cc-pVTZ outperforms all MP2 methods listed except for the SCS-MI-MP2/CBS which has an RMSE smaller by only 0.03 kcal mol^{-1}. Our DFT functional also outperforms the MP3/CBS and CCSD/CBS methods by 0.21 and 0.31 kcal mol^{-1}, respectively. SCS-MI-CCSD/CBS, MP2.5/CBS and SCS-CCSD/CBS are the only three methods that give RMSE smaller than PW86+PBE+XDM. SCS-MP2/CBS results in the largest statistical error.

The basis-set dependent MP2 method is the cheapest computationally compared to the other WFTs listed, but it is well known to overbind dispersion-bound complexes. DFT has the advantage of being even less computationally expensive than MP2 and

gives more accurate binding energies for the S66x8 (and S22x5 as shown in the previous chapter) vdW complexes. MP3, in contrast to MP2, underestimated dispersion interactions. The MP2.5, which is an average of MP2 and MP3 gives a well-balanced description of vdW complexes (RMSE of only 0.1 kcal mol^{-1}) due to the cancellation of the overestimation and the underestimation of MP2 and MP3, respectively. The dispersion weighted MP2, is a weighted average of MP2/CBS and SCS-MP2/CBS (spin component scaled MP2, scaling for opposite-spin and parallel-spin MP2 correlation). The idea behind DW-MP2 is similar to that behind the MP2.5 method as the overbinding MP2 is combined with the underbinding SCS-MP2 to reach some systematic cancellation of systematic opposite errors. The better performance of SCS-CCSD and SCS-MI-CCSD compared to CCSD is because both were parametrized using S22 [150].

6.5 Conclusions

In conclusion, the DFT functional PW86+PBE+XDM for exchange, correlation and dispersion, respectively, is an accurate method for computing energies of vdW complexes at equilibrium and distorted geometries. This DFT method even competes with wavefunction theory methods, e.g., MP2 and some of its derivatives. The mean absolute relative error for the S66x8 database (of 66 complexes at eight different separations each) is 9.1%. The largest statistical error was for the complexes compressed by 10%, MARE is 15.8 % (compared to 13.3% in S22x5). The MARE for the well-balanced S66x8 (9.05% overall) is 0.3% smaller compared to the MARE for S22x5 complexes (9.35%). The damping parameters a_1 and a_2 are strongly linearly correlated with each other and they are roughly universal with the exception of complexes compressed by more than 5% (i.e., 10%). At such short distances, the potential energy curve is very steep and a minor change in the parameters results in a large deviation in the binding energies. The good performance of PW86+PBE+XDM at

equilibrium as well as distorted geometries leads to a promising DFT for simultaneous optimization of *inter*- and *intra*-molecular geometries in complex chemical systems.

Chapter 7

Dispersion Forces From the PW86+PBE+XDM Density Functional for van der Waals Complexes

7.1 Abstract

As presented in the previous chapters, the nonempirical functional PW86+PBE+XDM comprises an accurate functional for thermochemistry and for van der Waals (vdW) complexes at equilibrium and non-equilibrium geometries [74, 145, 157]. This sets the stage for the further goal of optimizing the geometries of vdW complexes with this functional. In order to optimize molecular geometries, forces must be computed by differentiating the energy with respect to nuclear coordinates. We will assume that our dispersion coefficients are constant with respect to changes in the nuclear coordinates. This approximation allows us to avoid the complexity of taking the derivatives of the dispersion coefficients with respect to the density. The purpose of this chap-

ter is to validate this assumption by checking if, along a diatomic potential energy surface (PES), the internuclear distance at which the energy is minimum is commensurate with the distance at which the force vanishes to zero. This test was done on ten rare-gas diatomic systems using various integration grids and SCF convergence criteria.

7.2 Introduction

Dispersion is estimated to account for a thousandth of the correlation energy in chemical systems [64]. This is why conventional exchange-correlation GGA functionals fail to describe van der Waals (vdW) complexes [55, 147, 158] or, at best, give erratic results. In DFT, a functional can properly treat vdW interactions if a fully nonlocal dispersion correction is added.

The exchange-hole dipole moment (XDM) model proposed by Becke and Johnson [64, 65, 66] can accurately capture dispersion interactions, yet it is a simple model. This model is applicable to any DFT-D type method. For instance, it can be combined with PW86 [26] (a GGA functional for exchange) and PBE [28] (a GGA functional for correlation) to obtain PW86+PBE+XDM which comprises a nonempirical DFT method for vdW complexes. This functional (PW86+PBE+XDM) is of high accuracy for ordinary thermochemistry. Its mean absolute error (MAE) for the atomization energies of the 222 molecules in the G3/99 thermochemistry dataset is 10 kcal mol^{-1} [74, 145]. In addition, as presented in Chapter 4, this functional predicts accurate binding energies for vdW complexes. With an augmented triple zeta Dunning basis set, the mean absolute percent error (MAPE) for the binding energies of complexes in the S22 set is 5.5 % [145], a statistical value to be compared with 2.4 % and 20.4 % using CCSD(T*)/F12/AVDZ [159] and MP2/CBS [160], respectively. Chapter 5 also showed the ability of this functional to properly describe complexes at non-

equilibrium intermonomer separations. The mean absolute relative error (MARE) for the binding energies on the S22x5 set [146] is 9.4 % on average with a maximum of 13.3 % at compressed separations, i.e., the most challenging case [157]. The reader is reminded here that the S22x5 set of complexes includes the 22 complexes of the well known S22 database [140] at five different intermonomer separations scaled by factors of 0.9, 1.0, 1.2, 1.5 and 2.0 relative to the equilibrium distances.

The choice of the exchange functional in DFT-D [74, 145] is not obvious as the performance of common exchange GGA functionals can range anywhere from superficial overbinding to extreme over-repulsion. Among the common GGA functionals in the literature, PW86 reproduces the exact HF exchange energy in vdW systems[1] [137, 138] most accurately [30]. The correlation functional PBE was chosen as it is a nonempirical GGA functional. XDM is also nonempirical, it is an add-on functional that corrects for nonlocal correlation effects [64, 65, 66]. Hence we obtain the present DFT-D functional PW86+PBE+XDM.

As discussed in Chapter 2, the XDM model is based on the position-dependent dipole moment of the exchange hole given by

$$h_{X\sigma}(\vec{r}_1,\vec{r}_2) = -\frac{1}{\rho_\sigma(\vec{r}_1)}\sum_{ij}\psi_{i\sigma}(\vec{r}_1)\psi_{j\sigma}(\vec{r}_1)\psi_{i\sigma}(\vec{r}_2)\psi_{j\sigma}(\vec{r}_2), \tag{7.2.1}$$

where ρ_σ is the σ-spin density and $\psi_{i\sigma}$ are occupied Hartree-Fock or Kohn-Sham orbitals. Orbitals are assumed to be real. This hole is a measure of the depletion of probability (with respect to the total σ-spin electron density, ρ_σ) of finding an electron of spin σ at a point $\vec{r}_2$ next to a reference electron of same spin σ at point $\vec{r}_1$.

[1]Exact exchange repulsion energy for rare-gas diatomic systems was computed with the basis set free NUMOL code.

The damped [56, 68, 115, 116] dispersion energy in XDM is given by

$$E_{disp} = -\sum_{i>j}\left(\frac{C_{6,ij}}{R_{vdW,ij}^6 + R_{ij}^6} + \frac{C_{8,ij}}{R_{vdW,ij}^8 + R_{ij}^8} + \frac{C_{10,ij}}{R_{vdW,ij}^{10} + R_{ij}^{10}}\right), \tag{7.2.2}$$

where $R_{vdW,ij}$ is an effective van der Waals internuclear separation that linearly depends on a critical separation $R_{c,ij}$:

$$R_{vdW,ij} = a_1 R_{c,ij} + a_2, \tag{7.2.3}$$

where a_1 and a_2 are two universal fit parameters, and the critical separation $R_{c,ij}$ is the separation at which the $\frac{C_m}{R^m}$ (m= 6, 8, 10) terms have roughly the same value:

$$\frac{C_{6,ij}}{R_{c,ij}^6} \approx \frac{C_{8,ij}}{R_{c,ij}^8} \approx \frac{C_{10,ij}}{R_{c,ij}^{10}}. \tag{7.2.4}$$

As shown in equation 7.2.2, the XDM model includes not only C_6 coefficients, but also C_8 and C_{10} coefficients, all of which are environment (i.e., density) dependent. This allows the functional to be widely applicable to accurate descriptions of *inter*- or even *intra*-molecular dispersion interactions. We use the values of a_1= 0.79 and a_2=1.36 Å [145] in this work.

The dispersion coefficients in the XDM model are functionals of the electron spin density through the exchange-hole function (eq 7.2.1), and expectation values of atomic moment integrals and atomic polarizabilities (Chapter 2). Since XDM depends on electron spin densities, the dispersion coefficients are geometry dependent.

However, when taking the derivatives of the PW86+PBE+XDM energy with respect to x, y and z coordinates, it is convenient to treat the dispersion coefficients as constants. The purpose of this chapter is to validate this approximation. This goal can be achieved by verifying that forces computed with PW86+PBE+XDM vanish at equilibrium geometries.

7.3 Computational Details

In this study, calculations are performed using PW86+PBE+XDM in a post-GGA fashion. The GGA part of this functional PW86+PBE is computed self-consistently with the G09 package [118]. The XDM part is then added as a perturbative correction to the self-consistent GGA exchange-correlation energy. When forces are computed (i.e., total energy derivatives with respect to nuclear coordinates), our dispersion coefficients are assumed to be constants at the given geometry, even though in principle they are not.

The binding energies and the force curves of ten rare-gas diatomic systems (from He, Ne, Ar and Kr noble gases) are computed. Single-point calculations are performed on the diatomic systems at the experimental equilibrium internuclear separation (obtained from [133]) with plus and minus 40 points at increasing and decreasing distance increments of 0.01 Å. The forces are the sum of the G09 PW86+PBE force and the derivative of eq. (7.2.2) assuming that the dispersion coefficients are constants.

Augmented Dunning double and triple zeta basis sets are tested. The aug-cc-pVTZ gives accurate binding energies, but the aug-cc-pVDZ basis set is more economical for geometry optimizations [145]. Two SCF convergence criteria were tested, 10^{-6} and the tighter criterion 10^{-8} as the latter is becoming common practice nowadays. For the GGA part, two different grids are considered 99590 and 200590. The grid notation xxxyyy means there are xxx radial shells surrounding each nucleus and yyy angular points distributed on each shell. For all calculations where the grid is not explicitly mentioned, the 200590 grid is used.

7.4 Results

Using PW86+PBE+XDM, the energies and forces are tested on closed shell rare-gas diatomics as they are the prototype of vdW interactions.

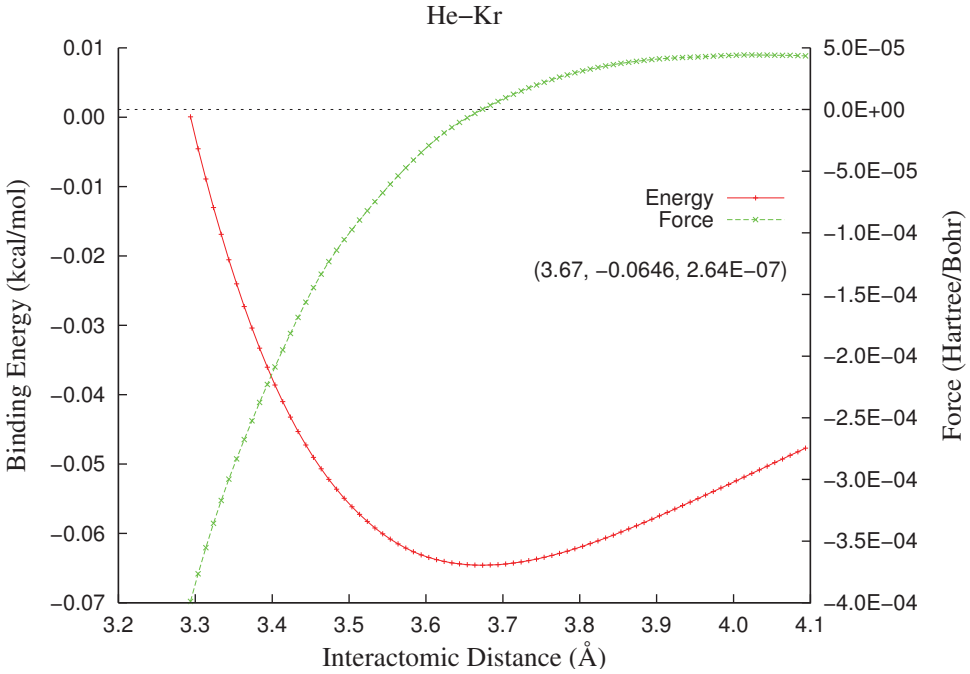


Figure 7.1: Binding energy of He-Kr in kcal mol^{-1} (left side x-axis) and the force in He-Kr in atomic units (right x-axis) as a function in internuclear separation in Å.

The forces are total forces (GGA+XDM forces) acting on each nucleus. Since these systems are only diatomic, the force on nucleus A is equal in magnitude to the force on nucleus B, but opposite in sign. This means we have only one unambiguous force at each geometry. In the case of triatomic or polyatomic molecules, the analysis might be ambiguous depending on the choice of the atom on which the force is exerted.

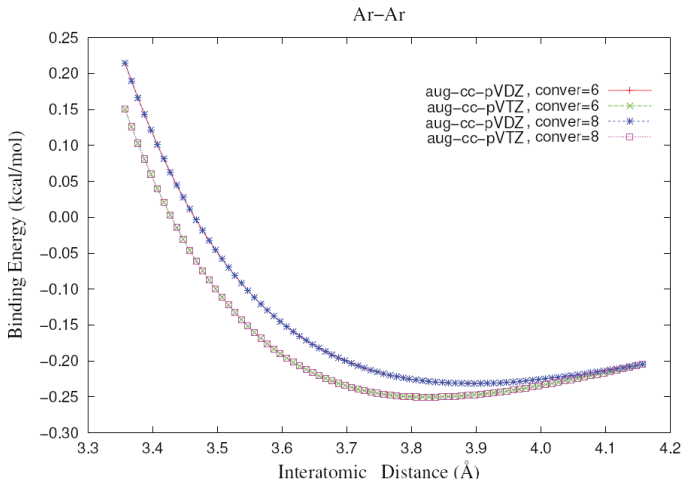
Figure 7.1 is a graph combining two plots for the binding energy and the force in He-Kr as a function of the internuclear separation (those are calculations using aug-cc-pVTZ, a 200590 grid and a convergence criterion of 10^{-8}). This figure is a representative of ten similar graphs for the ten rare-gas diatomics considered in this study. The binding energy curve and the force curve are displayed in the same graph to investigate if the zero-force geometry is commensurate with the minimum-energy geometry along the

PES. Binding energies are plotted rather than absolute energies for scaling purposes. This graph illustrates how at distance 3.67 Å, the binding energy of He-Kr is minimum (-0.0646 kcal mol^{-1}) and the force vanishes to 2.64E-7 atomic units (au) (which is essentially zero). The calculated equilibrium separation is in great agreement with the experimental equilibrium separation (3.69 Å [74]), with only 0.5% difference. The functional PW86+PBE+XDM slightly overbinds He-Kr by 3.4% (experimental binding energy is -0.0625 kcal mol^{-1} [74]). For all ten rare-gas diatomics considered in this study, the distance of minimum binding energy is commensurate with the distance of vanishing force. At worst, the distance at which the force is zero is of order 0.01 Å greater than the distance at which the binding energy is at a minimum. These results confirm that negligible errors are introduced by excluding dispersion-coefficient derivatives in the total-energy derivative calculations.

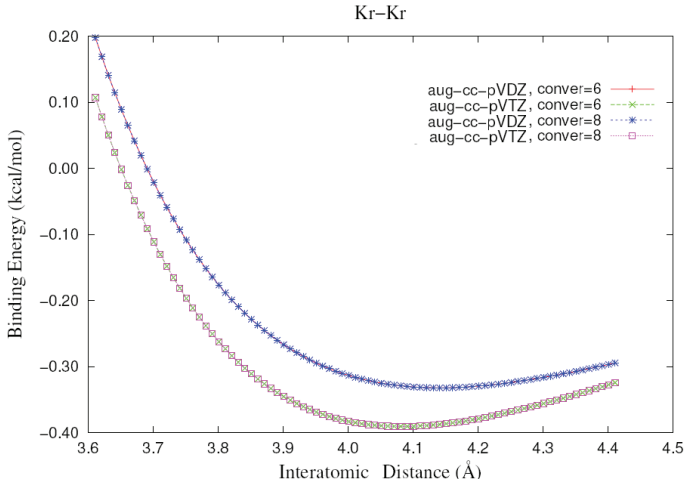
The most important issue behind this study has now been addressed. However, there are other minor, yet essential, details that should be considered when computing energies and forces. These details include the choice of a suitable SCF convergence criterion necessary grid size for the GGA computations.

7.4.1 Assessment of the SCF Convergence Criterion: Binding Energies

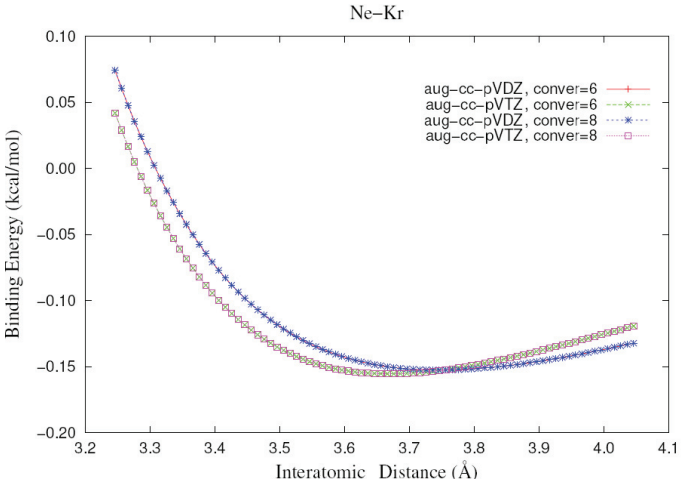
While an SCF convergence criterion of 10^{-8} (the default in some software packages) ensures a tight threshold, it might impede SCF convergence in many complexes. The purpose of this section is to investigate the effect of lowering the SCF convergence criterion from 10^{-8} to 10^{-6} on the binding energies of the ten rare-gas diatomics. This assessment will be done for both augmented double and triple zeta Dunning basis sets.



(a) GraphA



(b) GraphB



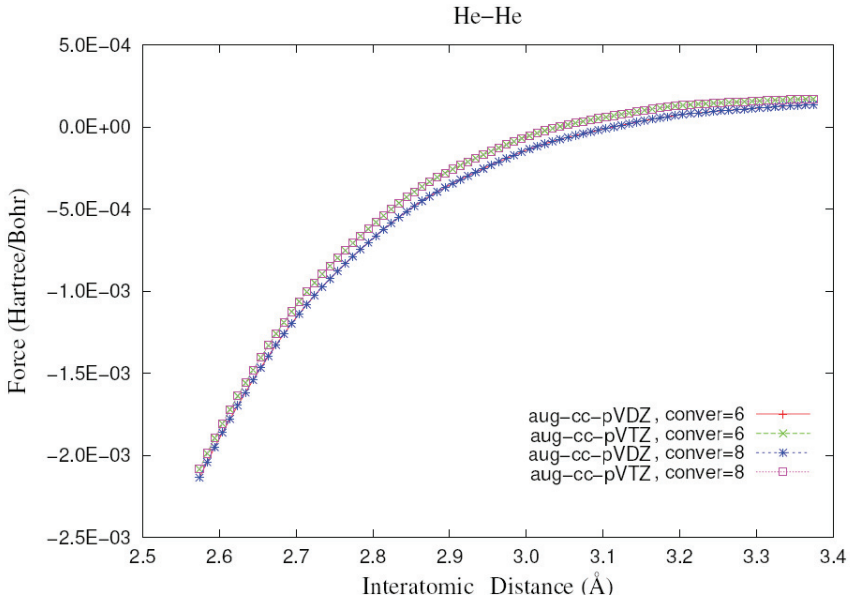
(c) Graph C

Figure 7.2: Binding energy curves of three rare-gas diatomic systems, Ar-Ar (graph A), Kr-Kr (graph B) and Ne-Kr (graph C) using augmented double and triple Dunning basis sets with SCF convergence criteria of 10^{-6} and 10^{-8}.

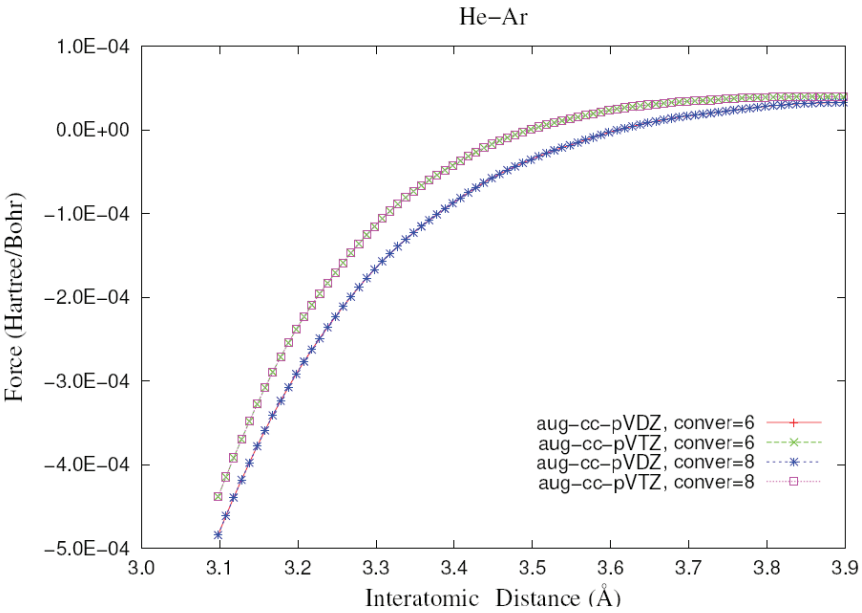
Figure 7.2 shows representative plots of the binding energies with respect to interatomic distances of the ten rare-gas diatomics using different basis sets and SCF convergence criteria. It is evident from Figure 7.2 that, for a given basis set, different convergence criteria make negligible difference in the binding energies along the whole range of internuclear separations. With either basis set, and for all ten diatomic systems, the maximum difference in binding energy between SCF convergence criteria of 10^{-6} and 10^{-8} is in the order of μkcal mol^{-1}. This means a convergence criterion of 10^{-6} allows high accuracy of binding energies even for the weakest cases such as the helium dimer. These results give us confidence in using an SCF convergence criterion of 10^{-6} to appreciably facilitate SCF convergence without concern for the accuracy of the binding energies. These graphs also illustrate how a double zeta basis set binds the rare-gas diatomic systems less than a triple zeta basis set. In Ne-Kr (graph C) and molecules similar to it, at a certain distance beyond the equilibrium separation, this trend is flipped, i.e., the molecules are more bound with a double zeta basis set rather than the triple zeta basis set. This might be due to basis set superposition error. Nevertheless, it is worth mentioning that the difference in binding energies between aug-cc-pVDZ and aug-cc-pVTZ is only of order 0.1 kcal mol^{-1} (with both convergence criteria).

7.4.2 Assessment of the SCF Convergence Criterion: Forces

Graphs in Figure 7.3 are similar to graphs in Figure 7.2, except that the forces are now assessed rather than the binding energies. For all rare-gas diatomic systems, the nuclear derivatives computed with aug-cc-pVDZ are smaller than those computed with aug-cc-pVTZ. As depicted in Graphs A and B, the only difference between the ten diatomic systems is the separation between the double and the triple zeta curves. Again the differences are minuscule. With the convergence criterion of 10^{-6}, the maximum difference in the force between aug-cc-pVDZ and aug-cc-pVTZ is 100 μau.



(a) GraphA



(b) GraphB

Figure 7.3: Force curves of two rare-gas diatomic systems, He-He (graph A), He-Ar (graph B) using augmented double and triple Dunning basis sets with SCF convergence criteria of 10^{-6} and 10^{-8}.

This difference drops to 90 μau with a convergence criterion of 10^{-8}.

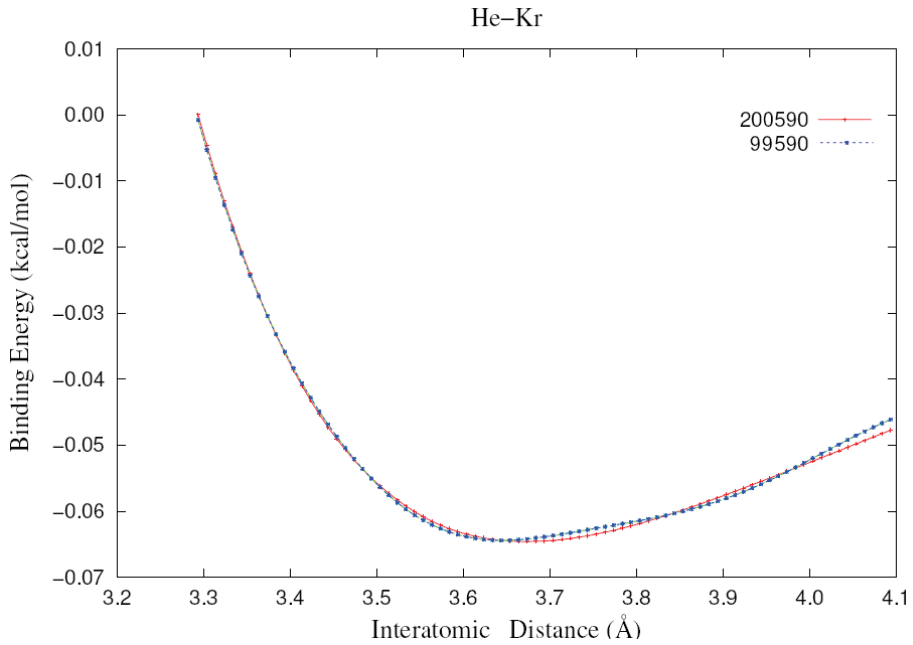
The message to be drawn from this figure is that the choice of the SCF convergence criterion (10^{-6} versus 10^{-8}) does not affect the values of the forces regardless of the basis set used. There is a maximum of $\sim$0.1 μau difference arising from different convergence criteria used at either basis set.

7.4.3 Assessment of the Grids

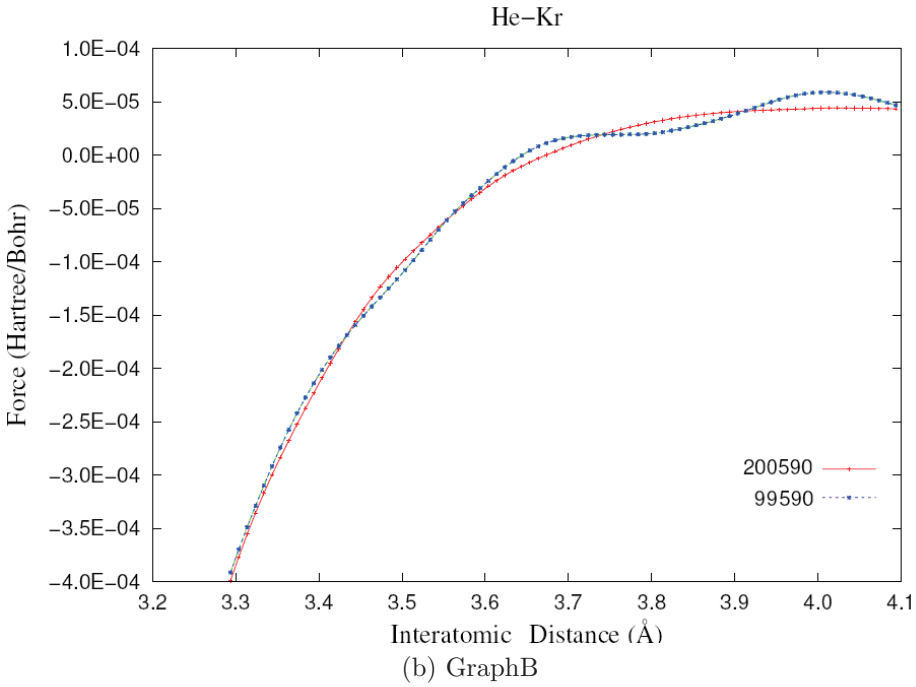
The integrals in the exchange-correlation potential in DFT GGA functionals are too complicated to be solved analytically. They have to be evaluated numerically on a real space grid [132]. As Johnson *et al.* have shown, fine grids are needed to get reliable results and smooth potential energy curves with meta-GGA functionals [161]. This section is an investigation of the grids required for computations with our post-GGA functional.

Graphs in Figure 7.4 show that changing the grid for the GGA part changes the shape of the binding energy and force curves. The 99590 grid gives sporadic results, with oscillations in the force curve of He-Kr (the change in the binding energy is much less pronounced). To remedy the situation, a refined grid, 200590, should be used. This refined grid includes more radial shells around each nucleus and gives smooth binding energies and force curves.

As to observations made with the rest of the diatomic systems: He-Ar is sensitive to the grids used (just as He-Kr is), He-He, He-Ne and Ne-Ne are less sensitive; while Ar-Kr, Ar-Ar, Kr-Kr, Ne-Ar and Ne-Kr are not sensitive at all to the change of grids.



(a) GraphA



(b) GraphB

Figure 7.4: Binding energy (graph A) and force (graph B) curves of He-Kr using 99590 grid (ultrafine grid) versus 200590 grid in G09.

7.5 Conclusions

For all ten rare-gas diatomics, the internuclear distance at which the binding energy is minimal along the potential energy surface is commensurate with the distance at which the force vanishes. This validates the assumption of constant dispersion coefficients in the computation of forces. For smooth potential energy surfaces and accurate binding energies and forces, a very fine grid must be used. In this chapter we find that the 200590 grid is more reliable than the 99590 grid. The use of an SCF convergence criterion of 10^{-6} is sufficient, as the difference between SCF convergence criteria of 10^{-6} and 10^{-8} in the evaluation of binding energies and forces in rare-gas diatomics is very negligible.

Chapter 8

Optimization of van der Waals Complexes Using the PW86+PBE+XDM Density Functional

8.1 Abstract

All the work presented in the previous chapters can be combined towards building an optimizer for van der Waals complexes including *inter*- and *intra*-molecular interactions. This optimizer for vdW complexes with a DFT method is achieved by using the "external" keyword in the Gaussian package (i.e., using the Berny optimizer). This new optimizer with the GGA density functional PW86+PBE+XDM will be a compromise between accurate geometries and binding energies for weakly-bound complexes and reasonable computational cost. This chapter will cover testing the optimizer and using it in some applications.

8.2 Introduction

Throughout the chapters of this thesis, it was demonstrated that the GGA functional PW86+PBE+XDM gives accurate thermochemistry, binding energies on vdW complexes (including hydrogen/electrostatic bonding, dispersion-bound complexes and mixed complexes) at equilibrium and distorted geometries (as tested on S22 [140], the 65 complexes of Kannemann and Becke [74, 145], S22x5 [146] and S66x8 [150]). The accuracy of this functional in computing forces by taking the derivative with respect to infinitesimal nuclear changes has been also demonstrated in the previous chapter. The ultimate goal is to have an optimizer for vdW complexes using an accurate nonempirical DFT method. The importance of such an optimizer is the full optimization of vdW complexes while accounting for the deformation geometry and, subsequently, the deformation energy. The majority of newly developed methods for vdW interactions are benchmarked against reference binding energies which do not account for deformation energies. This means the complexes are optimized at one particular method (e.g. MP2) then single point calculations are performed at a higher level of theory and a larger basis set (an extrapolation to a complete basis set limit in most cases). This is a valid approach for economizing the computational cost. However, the problem is in the monomers which are taken as fixed geometries. The structures of the monomers taken directly from the optimized dimer results in missing the deformation energy. This is where the importance of the optimizer presented in this chapter shines.

The Berny optimizer in the Gaussian package can be used with many wavefunction methods. The disadvantage, however, is the expensive computational cost, which limits the system under study to a small finite system. The Berny optimizer can be also used with empirical DFT functionals for vdW complexes. The disadvantage in this case is the empiricism, which limits the applications to systems identical or similar to what is included in the training sets. The optimizer presented in this chapter

compensates between accuracy and computational cost. The setup of this optimizer is displayed in Figure 8.1.

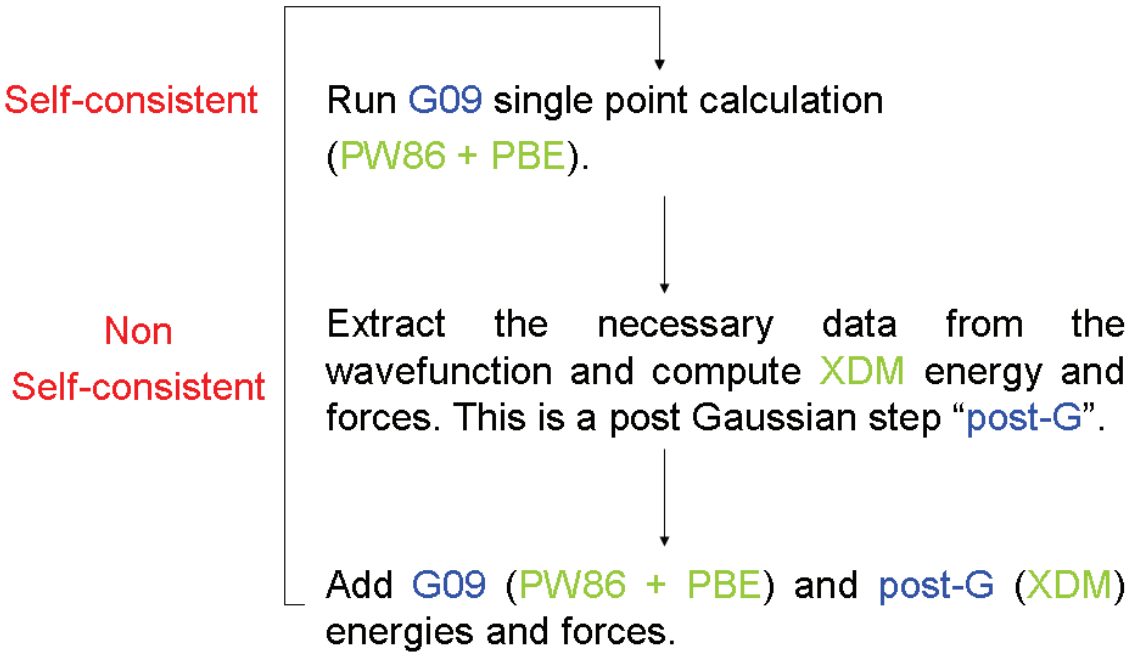


Figure 8.1: A schematic diagram of the steps undertaken by the optimizer for vdW complex using the "external" keyword in the Gaussian package.

In the first step Gaussian runs a single point calculation where the exchange and correlation energies are calculated self-consistently using PW86+PBE. In the next step, the necessary data for computing the XDM energy according to the equations presented in Chapter 2 is extracted from the wavefunction file of the single point calculation. This step is done non-self-consistently with a post Gaussian code. In the last step, the dispersion energy is added as a perturbative correction to the exchange + correlation energies and the XDM forces are added to the PW86+PBE forces to get a new geometry and start a new cycle until geometry convergence is achieved.

8.3 Assessment of the Reproducibility of the Optimizer

In this section, the reproducibility of the optimizer is tested on dispersion-bound rare-gas diatomic systems, which are the prototype of dispersion interactions. The assessment will be also conducted on a more practical example, the stacked benzene dimer. In this section, PW86+PBE+XDM /aug-cc-pVTZ is used.

8.3.1 rare-gas Diatomic Systems

In this section, the test is conducted on ten rare-gas diatomic systems from the He, Ne, Ar and Kr monomers. The experimental equilibrium intermonomer distances of these molecules were obtained from [133]. To test the optimizer, several input files were prepared by compressing and stretching the experimental distances by 2.2 Å in alternating increments of 0.01 and 0.1 Å as shown in Figure 8.2.

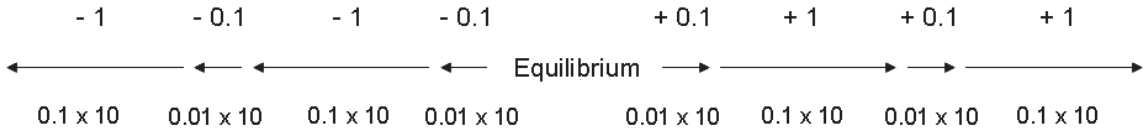


Figure 8.2: Steps of compressing and stretching the experimental distances of the ten rare-gas diatomic systems. The numbers shown under the arrows are the increment (in Å) x number of times, to give a total shift (in Å) displayed above the arrows.

The results after submitting the input files for an optimization with PW86+PBE+XDM /aug-cc-pVTZ ($a_1 = 0.79$ and $a_2 = 1.36$) are summarized in Table 8.1.

As shown in Table 8.1, the equilibrium distance is recovered with a maximum RMSE of 0.07 Å in Ar-Ar, Ar-Kr and Kr-Kr. The maximum RMSPE is 1.79% for Ne_2. This table also displays the range of minimum to maximum distances recovered after optimization. The results of this test demonstrate the good reproducibility of the optimizer on the most sensitive dispersion-bound complexes, ten rare-gas diatomics.

Table 8.1: Statistical results for the optimization of the ten rare-gas diatomic systems starting at different geometries.

	He-He	He-Ne	He-Ar	He-Kr	Ne-Ne	Ne-Ar	Ne-Kr	Ar-Ar	Ar-Kr	Kr-Kr
Ref	2.97	3.05	3.50	3.69	3.09	3.48	3.65	3.76	3.89	4.01
Avg	2.99	3.04	3.50	3.67	3.15	3.51	3.67	3.82	3.96	4.08
Max	3.07	3.09	3.55	3.70	3.18	3.54	3.69	3.84	3.96	4.08
Min	2.90	3.00	3.46	3.64	3.11	3.49	3.65	3.81	3.95	4.07
RMSPE	1.05	0.49	0.32	0.58	1.79	1.06	0.64	1.77	1.74	1.71
RMSE	0.03	0.01	0.01	0.02	0.06	0.04	0.02	0.07	0.07	0.07

8.3.2 Stacked Benzene Dimer

In this subsection, the optimizer will be tested on the stacked benzene dimer which is dispersion-bound through a π-π stacking interaction. The steps and increments of the increase and decrease in intermonomer distance are depicted in Figure 8.3.

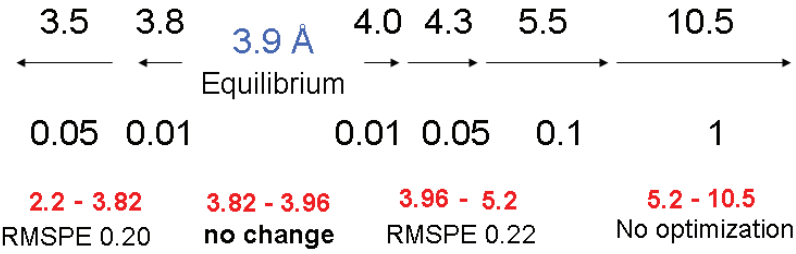


Figure 8.3: Steps of compressing and stretching the experimental distances of the stacked benzene dimer. The numbers shown under the arrows are the increment (in Å) and the total shift is displayed (in Å) above the arrows. The last row summarizes the results.

As the intermonomer distance varies from 3.82 to 3.96 Å in increments of 0.01 Å, the optimized geometry recovers the same initial separation. For separations ranging from 2.20 to 3.82 and from 3.96 to 5.2 Å, the benzene dimer is optimized to 3.9 Å with RMS errors of 0.20 to 0.22 %, i.e., 0.009 to 0.008 Å. Beyond these ranges, no optimization takes place because at the larger separations the benzene monomers are too far to interact and the optimization does not take place.

When the benzene dimer is optimized without dispersion (i.e. PW86+PBE without XDM), the percent error in the binding energy increases by one order of magnitude

from 1.03 to 36.64% [the reference binding energy is 1.70 kcal mol^{-1}, it is computed at the CCSD(T)/aug-cc-pVQZ* level of theory (the star * indicates a modified number of basis functions was used for the hydrogen atoms) [162].]. This particular example shows the great power of an optimizer that can account for dispersion.

In this section, with the tests on the rare-gas diatomic systems and the stacked benzene dimer, it was demonstrated that the optimizer gives reproducible geometrical parameters with highly accurate binding energies, which include deformation energies.

8.4 Application to Simple Systems

This section includes biologically relevant applications on DNA base pair complexes (to show the importance of dispersion in stacked nucleobases) and on conformational energies of substituted tetrahydropyrans and cyclohexanes (to study the anomeric effect). All calculations in this section were performed with PW86+PBE+XDM/aug-cc-pVTZ/W06. W06 is a density fitting set [163] of Ahlrichs *et al.* With density fitting, the two-electron integrals needed to compute the Coulomb interaction are approximated. Density fitting expands the density in an auxiliary basis set of atom-centered functions. With W06, by expressing the four-index integrals in terms of three-index integrals, time is saved as the storage requirements are reduced (much less I/O to read). Density fitting is used in this chapter to verify its reliability in computing geometries and energies. Density fitting is useful because it accelerates the PW86+PBE GGA computations significantly, which is an important feature for studying large complexes most efficiently without loosing accuracy.

8.4.1 Nucleic Acid Bases Interactions

In this subsection, nucleic acid interactions between DNA bases adenine, thymine, guanine and cytosine are studied. As depicted in Figure 8.4, a total of four complexes

are considered in this study, two are Watson and Crick hydrogen-bonded GC and AT and the other two are stacked bases.

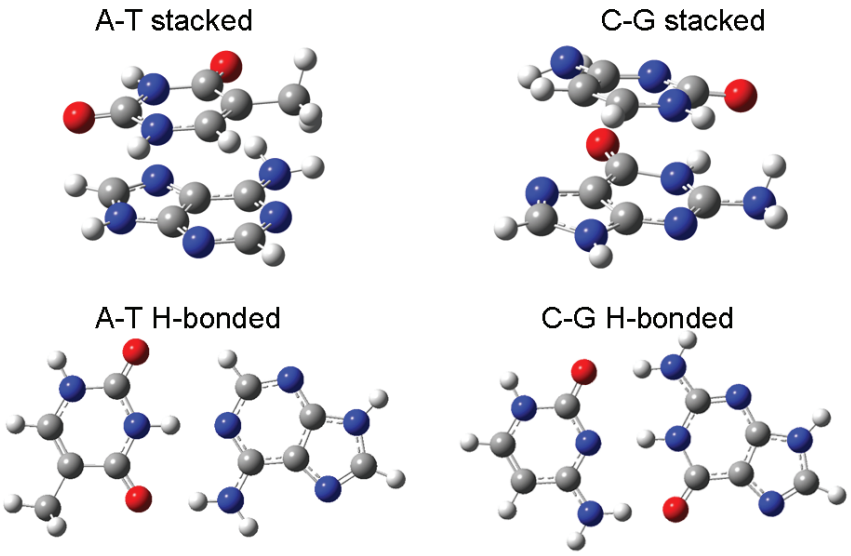


Figure 8.4: Stacked and hydrogen-bonded nucleobase pairs AT and GC obtained from [164].

The geometries obtained from [164] were optimized using RI-MP2/TZVPP. The reference binding energies were calculated at MP2/CBS + $(\Delta E_{CCSD(T)} - \Delta E_{MP2})_{cc-pVDZ}$. There are more updated reference binding energies, e.g., in [140]. However, the current particular geometries and reference binding energies were chosen for comparison purposes with what has been previously published in [104] on the same systems using HF/aug-cc-pVTZ+BR+XDM. In [104], the exchange energy is computed using HF, the Becke-Roussel (BR) [36] and XDM are used for dynamical correlation and dispersion energies, respectively.

The binding energies for these four complexes are computed with and without dispersion. Single point calculations were performed on the reference geometries, then binding energies were reevaluated after optimization using the optimizer presented above.

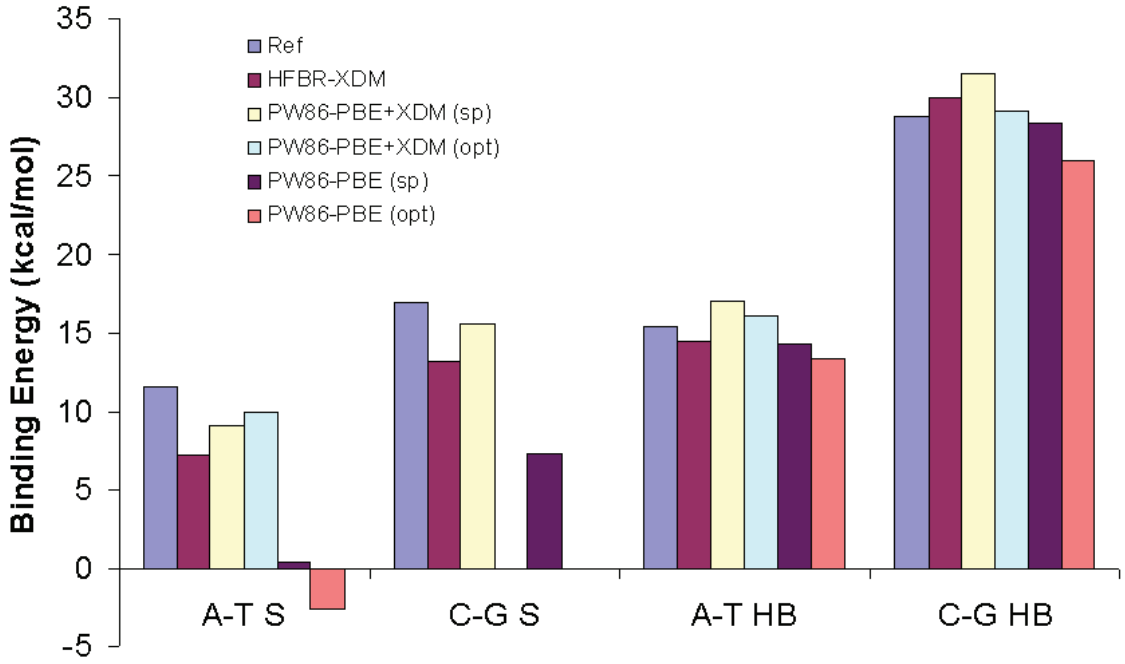


Figure 8.5: Bar graph for the binding energies of stacked and hydrogen-bonded nucleobase pairs AT and GC. The binding energies are calculated with and without dispersion from a single point calculation and after an optimization. Reference binding energies are obtained from [164].

Figure 8.5 displays the binding energies of the four complexes. This figure clearly depicts the repulsive energy of the stacked AT complex when optimized without dispersion. The hydrogen-bonded complexes do not show this repulsive behavior even when optimized without XDM, as in these complexes the electrostatic hydrogen bonding is dominant. GGAs can capture the HB interactions. This figure also shows the greater binding energy of hydrogen-bonding complexes compared to the stacked complex. Table 8.2 lists quantitatively the binding energies along with a statistical analysis.

From the first two rows of Table 8.2, it is obvious that XDM is crucial for the stacked complexes. Even if the complex was optimized with a wavefunction method, a single point calculation without XDM gives a very poor binding energy. This binding energy gets even worse as the complex is optimized without dispersion. With the hydrogen-

Table 8.2: Binding energies (in kcal mol^{-1}) and statistical errors for four nucleobase pairs. S and HB stand for stacked and hydrogen-bonding, respectively.

	PW86+PBE +XDM (sp[a])	PW86+PBE +XDM (opt[b])	PW86+PBE (sp[a])	PW86+PBE (opt[a])	HF+BR +XDM[d]	Ref[c]
A-T S	9.10	9.99	0.37	-2.54	7.20	11.6
C-G S	15.51	NA	7.26	NA	13.20	16.9
A-T HB	17.02	16.11	14.25	13.35	14.50	15.4
C-G HB	31.52	29.15	28.30	25.98	30.00	28.8
RMSPE	13.54	8.46	56.33	71.02	22.19	
RMSE	2.14	1.03	7.43	8.41	2.97	
MAE	2.06	0.89	5.63	6.34	2.55	
ME	0.11	-0.18	-5.63	-6.34	-1.95	
Min Error	-2.50	-0.18	-0.50	-2.05	-0.90	
Max Error	2.72	-1.61	-11.23	-14.14	-4.40	

a) sp means the binding energy was calculated from a single point calculation.
b) opt means the binding energy was calculated after an optimization.
c) The reference binding energies are obtained from [164].
d) These binding energies are obtained from [104].
In this table, root mean square error (RMSE), mean absolute error (MAE), mean error (ME), Min and Max error (in absolute value) in binding energies are given in kcal mol^{-1}. Root mean square percent error (RMSPE) is a percent error.

bonded complexes, the difference in binding energies with and without dispersion or in single point versus optimization is not as pronounced as in the stacked complexes. There is a small difference in the statistical errors for the PW86+PBE+XDM versus the HF+BR+XDM. This means that as long as a proper functional is chosen for exchange and for correlation and as long as dispersion is accounted for with XDM, the binding energy evaluated will be accurate.

The RMSPE for all complexes drops from 13.5% to almost half (8.5%) as an optimization is performed in the presence of XDM. This difference is due to the deformation energy, which is accounted for with optimization, but not with a single point calculation. As mentioned above, with the latter approach, the geometry of the complex is optimized at the MP2 level, however, the monomers are at fixed geometries, they are obtained directly from the optimized complex geometry. With the optimization calculation, both the complex and the monomers are optimized, which allows computing

the deformation energy. As dispersion is turned off, the error increases to more than 55%. In this case, as opposed to the case in the presence of XDM, the optimization causes an increase in the error from 56.3 to 71.0%. This is not unexpected, as after an optimization in the absence of XDM, the binding energy is computed for a complex at a less stable geometry. Thus it is better to compute single point energies at geometries "pre-optimized" at a level of theory that includes long-range nonlocal correlation effects rather than computing energies at geometries optimized with PW86+PBE. The mean error shows that in all cases the energies are underestimated, the only exception being PW86+PBE+XDM (sp). The minimum and maximum errors in binding energies show that PW86+PBE+XDM (opt) gives the smallest range of errors and the largest errors are in the absence of XDM (both sp and opt) especially for the stacked complexes.

The reason why there are no results for the optimized stacked GC (with and without dispersion) is that the complex will optimize to a hydrogen-bonding complex in both cases. This result is not surprising when the dispersion is turned off as stacked GC is basically a dispersion-bound complex. Therefore, the stacked complex will not exist, it will form instead the hydrogen binding GC. What was not expected is a hydrogen-bound GC after the optimization of the stacked GC even when XDM is turned on. Since the optimizer is doing a proper job of capturing the deformation energy in stacked AT, this means the optimizer is giving reliable results. Therefore, the conclusion in this particular case is that the naked stacked GC, in this particular orientation of G with respect to C, is not stable enough to exist; the more stable hydrogen-bonding GC is rather the final product after optimization. The stacked base pair would be stable if the dipole moments of the monomers are opposite to each other. The steps of the optimization in the presence of XDM for stacked GC are displayed in Figure 8.6.

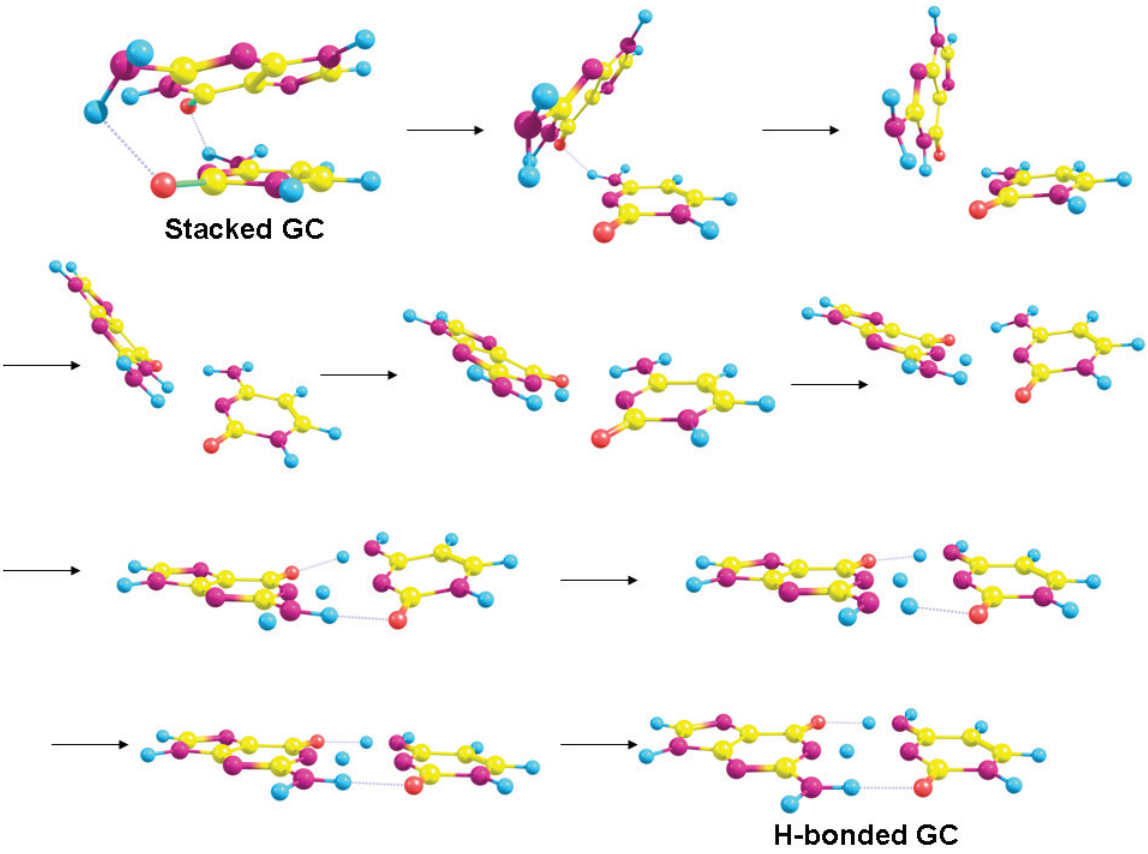


Figure 8.6: Optimization steps of stacked GC using PW86+PBE+XDM.

8.4.2 Anomeric Effect

The anomeric effects are of great biological importance. This effect is also known as Edward-Lemieux, named after the scientist who originally discovered this effect in 1955. In a cyclic molecule, e.g., a mono-substituted cyclohexane, the functional group on the equatorial position is preferred over the axial position. This is explained, in organic chemistry, by the steric hindrance effect where a functional group in the equatorial position is less hindered compared to the axial position, and is thus more stable. The anomeric effect predicts the axial position for a functional group attached to the anomeric carbon (C1) of a heterocyclic cyclohexane to be more stable than the equatorial position. This can be explained by a better orbital overlap in the axial orientation leading thus to a more stable molecule. Another explanation is the dipole cancellation of the heteroatoms in the axial stereoisomer. An example of substituted

heterocyclic hexanes is the hexoses which are ubiquitous in biological systems. If a wrong stereoisomer of a hexose is predicted computationally, then the stereospecificity of enzymes involved in the metabolic pathways of carbohydrates might not catalyze the reaction. This will lead to a metabolic malfunctioning in the body. Therefore, given the importance of correctly predicting the more stable diastereomer, it is necessary to have methods that describe energies accurately. Examples of computational and experimental studies on the anomeric effects can be found in [165, 166, 167, 168].

In this subsection, four mono-substituted cyclohexanes and four mono-substituted tetrahydropyrans (at the anomeric carbon) will be considered to test the accuracy of the PW86+PBE+XDM method in predicting the anomeric effect. Figure 8.7 depicts the molecules considered in this study.

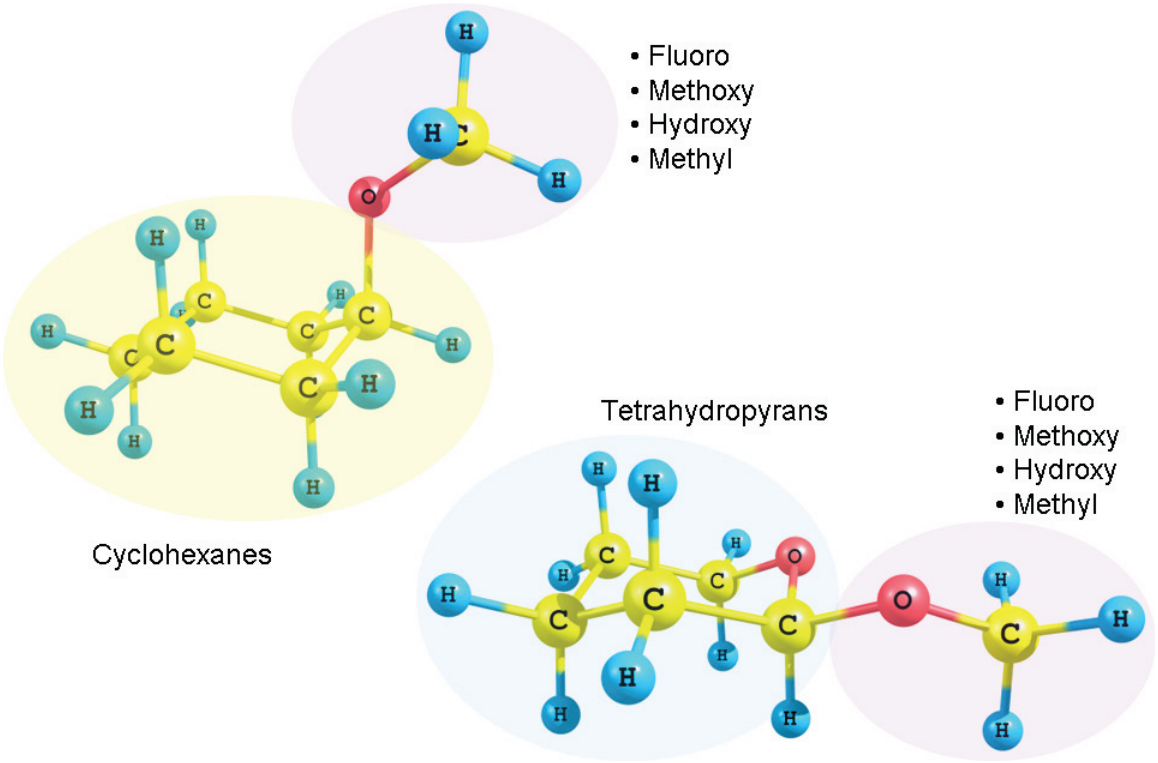


Figure 8.7: Molecular structures of cyclohexane mono-substituted at the axial position and mono-substituted tetrahydropyran at the equatorial position of the anomeric carbon. In both molecules the substituent shown is a methoxy group.

As calculated in [169] and interpreted in [104] for this particular example of molecules, dispersion (specifically *intra*-molecular dispersion) plays a vital role. The calcula-

tions in this study were repeated with and without XDM, with single point and optimization procedures. The basis set limit CCSD(T) reference energies were obtained from [169] and the geometries were obtained from a private communication with Dr. Erin Johnson. Using these particular geometries allows for a valid comparison of the current single point calculations and those in [104] calculated using HF+BR+XDM. Augmented triple zeta Dunning basis set is used to compute the HF energy in HF+BR+XDM, the BR and XDM were then computed numerically in a "post-HF" procedure. The results of this study are summarized in Table 8.3.

Table 8.3: Difference in energies (in kcal mol^{-1}) between axial and equatorial forms of four mono-substituted cyclohexanes and tetrahydropyrans along with a statistical errors analysis.

	PW86+PBE +XDM (sp[a])	PW86+PBE +XDM (opt[b])	PW86+PBE (sp[a])	PW86+PBE (opt[a])	HF+BR +XDM[d]	Ref[c]
			Tetrahydropyrans			
2-fluoro	2.56	2.92	2.40	2.78	2.18	2.45
2-hydroxy	0.62	0.80	0.36	0.56	0.57	0.86
2-methoxy	0.89	1.05	0.45	0.65	0.94	1.27
2-methyl	-2.79	-2.76	-3.19	-3.15	-2.96	-2.82
			Cyclohexanes			
fluoro	-0.37	-0.19	-0.53	-0.33	-0.18	-0.20
hydroxy	-0.82	-0.65	-1.10	-0.90	-0.58	-0.56
methoxy	-0.56	-0.35	-1.04	-0.73	-0.33	-0.21
methyl	-1.79	-1.76	-2.20	-2.15	-1.90	-1.75
RMSPE	69.26	26.51	158.11	96.68	26.00	
RMS	0.23	0.20	0.54	0.40	0.20	
MAE	0.20	0.13	0.49	0.37	0.17	
ME	-0.16	0.00	-0.49	-0.29	-0.16	

a) sp means the binding energy was calculated from a single point calculation.
b) opt means the binding energy was calculated after an optimization.
c) The reference energies are obtained from [169].
c) These energies are obtained from [104].
The energy difference is calculated using $E_{equatorial} - E_{axial}$.
Root mean square error (RMSE), mean absolute error (MAE), mean error (ME), Min and Max error in binding energies are given in kcal mol^{-1}. Root mean square percent error (RMSPE) is a percent error.

A positive energy means the axial position is more stable relative to the equatorial

position, which is what should be expected with the tetrahydropyrans, while the opposite should be expected for cyclohexanes. The molecule 2-methyl tetrahydropyran does not have an anomeric affect, it is governed by the steric rules. Table 8.3 shows the expected relative stability for all molecules at all methods. The accuracy of each method is significantly different. Similar to the case with DNA nucleobases, excluding the XDM from the calculation causes a significant increase in the error, by a factor of about 2 - 3. An optimization with XDM lowers the error from 69.3 to 26.5% (with the single point calculation), with a remarkable mean error. This improvement in accuracy indicates that the current optimization is more accurate than the MP2 optimization. While the RMSPE of HF+BR+XDM is 0.5% smaller than that of PW86+PBE+XDM (opt), the MAE and the ME are smaller in the latter. A consistent negative mean error is an indicative of an overstabilization of the equatorial position.

8.5 Conclusions

An optimizer for vdW complexes using the nonempirical DFT method PW86+PBE +XDM is presented in this chapter. This optimizer was built using the "external" keyword in the package Gaussian (03 or 09). This optimizer proved to be very reproducible for the weakest dispersion-bound complexes (ten rare-gas diatomic systems from the four monomers, He, Ne, Ar and Kr) and for a parallel stacked benzene dimer. The complexes are reproducibly optimized to final geometries that are in excellent agreement with the reference geometries. In addition, the computed binding energies are remarkably accurate. The use of this optimizer on biologically relevant studies proved the accuracy of this optimizer and its reliability for using it in further applications including *inter*- or *intra*-molecular interactions.

PART II

Chapter 9

Introduction

This part of the thesis investigates the effect of strong static electric fields on the double proton transfer reaction in Watson and Crick base pairs. Proton transfer reactions are common in biochemistry, accounting for the very high electrical conductivity of water, "proton wires" affecting proton transport across biological membranes [170, 171, 172, 173, 174], proton pumps such as bacteriorhodopsin [175, 176, 177], carbonic anhydrase [178, 179], photosynthetic reactions [180, 181, 182], ATP synthase [183, 184, 185], and has been invoked in the Löwdin mechanism of mutation and aging [186, 187, 188, 189, 190, 191, 192, 193].

The literature documents a number of biological effects of electric fields [194, 195, 196, 197]. For example, exposures of *Drosophila melanogaster* females to static electric fields of 1500 V cm^{-1} and 3300 V cm^{-1} for 24-hours increase the frequency of sex-linked recessive lethal mutations [198]. In another study, calculations show that weak residual external electrostatic fields can be amplified by a factor of $\sim$4x10^6 near the sharp edges of the junction linking two dividing cells [199]. These locally amplified fields are thought to inhibit cell proliferation [199]. The enhancement of the chemotherapeutic effect by electric fields, electrochemotherapy [200, 201, 202], is believed to result from the synergy between the chemotherapeutic drug and field-

induced electroporation of metastatic cell membranes [202].

In double-helical DNA, adenine (A) pairs with thymine (T) through two hydrogen bonds (A=T) and guanine (G) pairs with cytosine (C) with three hydrogen bonds (G≡C). Each hydrogen bond is essentially a proton trapped into an asymmetric double well potential between two electronegative atoms. The resulting potential barrier is too elevated for the proton to hop to the complementary base at room temperature. When the quantum nature of the proton is taken into account, however, the proton has a small, but finite, probability of tunneling through the barrier.

If a single proton tunnels from one base to its complementary partner, charge separation will result in a zwitterionic base pair and hence is improbable. When protons hop in pairs in opposite directions, charge neutrality is preserved. As a result of the double proton hopping, the DNA bases are converted to their less stable tautomers (the tautomeric forms that do not normally exist in stable DNA double helices), denoted by the asterisks: A*=T* and G*≡C*. If such double hopping occurs during DNA synthesis, a rare tautomer in the template helix is hydrogen bonded to a different base than in the Watson-Crick pairing scheme. For example, the base C (instead of T) would be complementary to A*, and T (instead of C) would hydrogen bond to G* (see Figure 9.1). This is the essence of Löwdin's mechanism for spontaneous and induced mutations [186, 187, 189, 191, 192, 203, 204].

Fu *et al.* showed that among several possible DNA mutation processes, only Löwdin mechanism for double proton transfer could commonly explain the universal mutation bias [205]. The slow accumulation of such errors, as long as they are non-lethal, is a possible contributor to the process of ageing. This study explores the effect of very strong electric fields on the double proton transfer reaction in WC dimers of DNA. While DNA base pairs are surrounded by the polar/ionic sugar-phosphate backbone, counterions, and histone proteins with their cationic side chains, these sources and sinks of electrostatic field are likely balanced. However, ionic compounds that interact with DNA can generate very intense local fields.

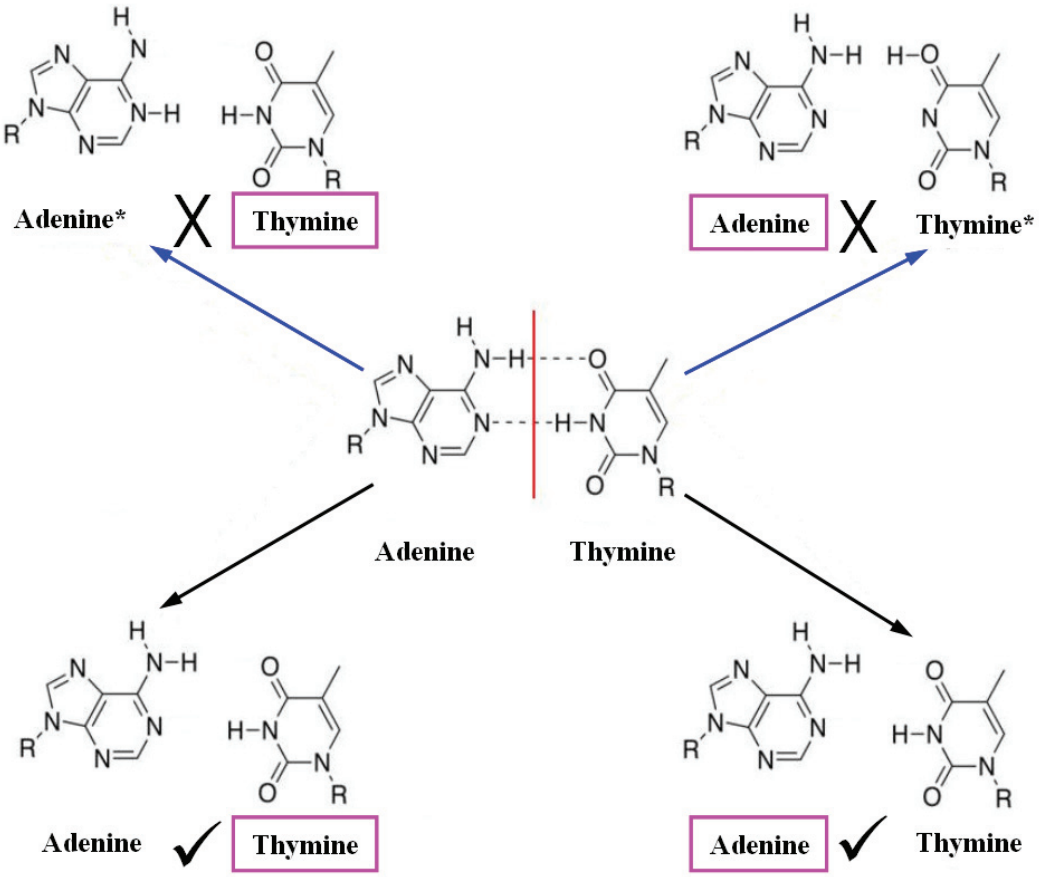


Figure 9.1: A schematic representation of a normal and mutated DNA replication at an AT base pair. In normal cases, the central AT unwinds to wild types A and T which will interact with a free T and A, respectively. This will form two normal replicates of AT (as shown at the bottom). With a mutation, the central AT unwinds to A* and T*. These rare tautomers cannot form a stable base pair with wild T and A, respectively (as shown in the above part of the sheme).

9.1 Field Strengths

The DNA molecule is surrounded by the highly charged phosphate backbone, histone proteins and metal cofactors. Even though these sources of charges are likely balanced, these charges create electric fields that vary in strength depending on the charges and the separation between them. In a WC double stranded B-DNA, the anomeric carbon atoms of two deoxyribose sugar groups are separated by ∼10 - 11 Å, and the diameter of the double helix is ∼20 Å. Therefore, 10 - 35 Å are typi-

cal separations to be considered between charges in the microenvironment of a DNA macromolecule. In vacuum, the electric field created at the mid-distance between two points separated by 10 Å and bearing a 0.5 au charge each (see Figure 9.2), is 5.14x10^9 V m^{-1}. This strength of the field decreases by an order of magnitude (i.e., to 5.14x10^8 V m^{-1}) if the charged points were separated by 34 Å. Such strong electric

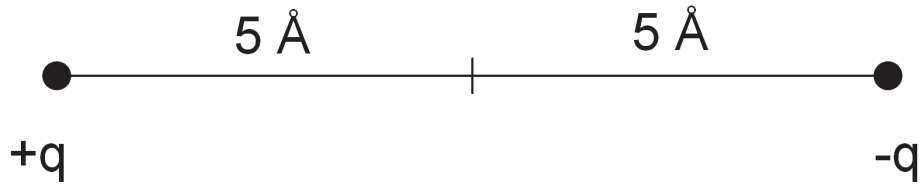


Figure 9.2: A schematic representation of two point charges separated by a distance of 12 Å. The equidistant point in the middle where the strength of the electric field is calculated.

fields are commonly encountered in biological systems. For example, phospholipid bilayers possess differently charged lipid heads creating aligned dipoles at the interfaces. These dipoles, in addition to the concentration gradient of ions at both sides of biomembranes, create electric fields in the order of 10^9 V m^{-1} [177]. Therefore, the strengths of the external electric field applied in this study are 5.14 to 51.42x10^8 V m^{-1}. These fields are two to three orders of magnitude stronger than electric fields encountered in a mass spectrograph using the field-induced droplet ionization (FIDI) techniques [206, 207]. On the other hand, the field strength in the tip-sample gap of a typical scanning tunneling microscope (STM) is $\sim$3x10^9 V m^{-1} for a tip-surface separation of 2 Å and a bias voltage of 0.6 V [208]. The field strengths in an STM, which fall in the range of field strengths examined in this study, are enough to overwhelm the tendency of water molecules to form hydrogen-bonded networks and to align their dipoles ([208] and the literature cited therein). It is well known that strong electric fields can strip atoms and molecules from an electron or more. Thus, the question of whether the field strengths used in our studies are sufficient to ionize the complex via tunneling ionization at the time scale of the proton transfer reaction has to be

addressed. Neglecting the orientation-dependence of the Stark shift of a molecule in an external electric field, the rate of tunneling ω of an electron from the complexes in the presence of an external electric field is approximately [209, 210]:

$$\omega(t) = 4\frac{\omega_0}{|E|}(2IP)^{\frac{5}{2}}exp\left[-\frac{2}{3}\frac{(2IP)^{\frac{3}{2}}}{|E|}\right], \tag{9.1.1}$$

where $\mid E \mid$ is the electric field strength, IP is the vertical ionization potential, both expressed in au, and the atomic unit of frequency is $\omega_0 = 4\text{x}10^{16}$ s^{-1}. The vertical IP of a complex is the total energy of the neutral dimer minus the total energy of the cation at the same geometry.

The inverse of the tunneling ionization rates is interpreted as the average time per ionization event, a characteristic time that decreases with field strength. Together, the tunneling ionization rate and the half-life of the reaction determine if an ionization event will take place at the time scale of the double proton transfer reaction.

The first chapter of this section tests the effect of static electric fields on the double proton transfer on formic acid dimer, which is a symmetric dimer. In this chapter, a benchmark study is also performed to assess the performance of the computationally inexpensive DFT B3LYP method compared to the wavefunction method MP2. The second chapter deals with the effects of electric fields on the DNA base pairs.

Chapter 10

Electric Field-Effects on Double Proton Transfer Kinetics in the Formic Acid Dimer

10.1 Abstract

Molecules can be exposed to strong local electric fields of the order of 10^8 - 10^{10} V m^{-1} in the biological milieu. The effects of such fields on the rate constant (k) of a model reaction, the double-proton transfer reaction in the formic acid dimer (FAD), are investigated. The barrier heights and shapes are calculated in the absence and presence of several static homogenous external fields ranging from 5.14x10^8 to 5.14x10^9 V m^{-1} using density functional theory (B3LYP) and second order Møller-Plesset perturbation theory (MP2) in conjunction with the 6-311++G(d,p) Pople basis set. Conventional transition state theory (CTST) followed by the Wigner tunneling correction is

then applied to estimate the rate constants at 25^{o}C. It is found that electric fields parallel to the long axis of the dimer (the line joining the two carbon atoms) lower the uncorrected barrier height, and hence increase the raw k. The fields also flatten the potential energy surface near the transition state region and, hence, decrease the multiplicative tunneling correction factor. The net result of these two opposing effects is that fields increase k(corrected) by a factor of ∼3 - 4 (DFT-MP2, respectively) compared to the field-free k. Field strengths of ∼3x10^9 V m^{-1} are found to be sufficient to double the tunneling-corrected double proton transfer rate constant at 25^{o}C. Field strengths of similar orders of magnitudes are encountered in the scanning tunneling microscope (STM), in the microenvironment of a DNA base-pair, in an enzyme active site, and in intense laser radiation fields.

10.2 Introduction

The double proton transfer in the formic acid dimer is studied as a model for a biochemical proton transfer reaction under the influence of a strong local electric field. The rate of double proton transfer in a dimer depends on the height of the barrier and the shape of the potential energy surface (PES). External electromagnetic fields have been shown to have considerable effects on the shape of the PES of gas-phase proton transfer reactions [211, 212]. Bandrauk *et al.* have shown that IR laser fields of intensity ranging from 1 to 3x10^{17} W m^{-2} (corresponding to maximal electric field strengths of ∼8.8x10^9 - 1.5x10^{10} V m^{-1}, respectively) induce considerable changes to the PES of the reactions of halogens with methane ($X_{radical}$ + CH_4 → HX + $CH_{3radical}$, X = F, Cl) [211]. The electric field, time dependent (as in the case of a laser pulse) or static, has been shown to couple to dipolar and polarizability terms in the effective potential along the reaction path, a coupling extremized sharply near the transition-state region. These extrema in field-dipole and field-polarization contributions can be made to interfere constructively by the proper choice of phase. This

constructive interference can lead to the disappearance or even the inversion of the barrier into a well. In the latter case, the transition state structure becomes a field-induced bound state [211].

As we show below, biological systems such as a DNA base pair can experience intense local electrostatic fields close to the orders of magnitudes of the fields studied by Bandrauk *et al.* [211]. These fields can arise from the surrounding medium including ionic/polar sugar-phosphate backbone, ions, and polar or charged histones intertwined with DNA. What are the effects of these local intense fields on the shape of the PES and the barrier? And how would these effects alter the rate constant of the double proton transfer? These are the principal questions addressed in this chapter.

The model reaction studied in this chapter, the double proton transfer in formic acid dimer (or related dimers), has been studied primarily in the absence of external fields. The reaction has been examined in dimers of acetic acid [213, 214], formamide [214], and formic acid [213, 214, 215, 216, 217, 218]. Chojnacki *et al.* calculated barrier heights, energies of dissociation and electrostatic molecular potentials of formic acid dimer using DFT and MP2 levels of theory [214]. Chocholousova *et al.* investigated the PES and the free energy surface of six formic acid dimer isomers using molecular dynamics and MP2 [215]. Fillaux addressed the issue of a stepwise double proton transfer versus a synchronous mechanism in formic acid dimer and analogues [216]. Scheiner and Kern also investigated the synchronous interchange of protons in AT and GC [217]. In addition, they calculated the quantum mechanical tunneling rates and the equilibrium constants of the proton interchange reaction [217]. Tsuzuki *et al.* report a benchmark study on the interaction energies of a series of simple hydrogen-bonding-complexes using several ab initio levels of theory [218]. Yavuz and Trindle studied the effect of counterpoise corrections to the basis set superposition error (BSSE) on geometries, binding energies and frequencies of *cis*- and *trans*-formic acid

dimers at the MP2 level of theory [219]. Vibrational properties of cyclic formic acid dimer were studied by Bosi *et al.* [220] and Fernandez *et al.* [221]. The Raman spectra of jet-cooled formic acid dimers and deuterated formic acid dimers have recently been reported for the first time by Zielke and Suhm [222].

The mechanism of double proton transfer in formic acid dimer has been studied by Ushiyama and Takatsuka [223]. These authors performed a full-dimensional classical dynamics study of the mechanism of the double-proton transfer in the formic acid dimer. They concluded that the double proton transfer occurs in a synchronous (simultaneous) step only within the adiabatic approximation, whereby the positions of all nuclei adjust instantaneously to the motion of the proton. This approximation is the one adopted to construct the PES and the minimum energy path connecting reactants and products. On the other extreme, that is, within the so-called "sudden approximation" whereby the motion of the proton is infinitely faster than that of the remainder of the nuclei, Ushiyama and Takatsuka concluded that the two protons move asynchronously, the second proton lagging by $\sim$8 fs behind the motion of the first proton [223].

An early semiempirical (AM1) quantum chemical study appears to be the only one incorporating the effect of an intense external electric field on double proton transfer (in DNA base pairs) [224]. This study shows that the single well preventing charge separation in adenine-thymine base pair (A^+-T^-) in vacuum is converted to a double-well potential under the influence of an externally-applied electric field and/or a polarizable (solvation) medium [224].

With the exception of this AM1 study, the effects of locally intense fields on proton transfer in DNA doe not appear to have been extensively investigated. The principal goals of the present study are: (1) to determine the effects of static external fields on the height and shape of the barrier of the double proton transfer reaction in the formic acid dimer and on the rate constant, (2) to determine the order of magnitude of the field strengths sufficient for causing a chemically-significant effect ($\geq$ 1 kcal mol^{-1}),

and (3) to compare results obtained at a relatively inexpensive method, B3LYP [155, 46, 34] with a more trustworthy, but computationally demanding, level such as the second-order Møller-Plesset perturbational approach (MP2) [225].

10.3 Field Strengths

As justified in the previous chapter, the field strength applied in this study range from 5.14x10^8 to 5.14x10^9 V m^{-1}. In this section, using eq.9.1.1 we will test if the dimer will ionize at the time-scale of the double proton transfer reaction in the presence of such strong fields. The IP of the formic acid dimer, is 11.0155 [MP2/6-311++G(d,p)] / 11.0162 eV [B3LYP/6-311++G(d,p)]. At field strengths of 5.14x10^8 V m^{-1} and 5.14x10^9 V m^{-1}, the calculated tunneling ionization rates are 1.1x10^{-191} s^{-1} and 7.6x10^{-3} s^{-1}, respectively, spanning 188 orders of magnitude. The inverse of the tunneling ionization rates is interpreted as the average time per ionization event, a characteristic time that decreases with field strength. The average time for a first ionization event of a formic acid dimer is calculated to be 130 s (∼2.2 minutes) for the strongest field considered in this study (5.14x10^9 V m^{-1}). As shown below, the half-life ($t_{1/2}$) of the double proton transfer calculated in this work is of the order of 10^1 - 10^3 picoseconds, at least 11 orders of magnitude shorter than an average field-induced tunneling ionization event. The proportion of formic acid dimers that ionizes under the influence of the strongest field is, thus, negligible at the time-scale of the double proton transfer, and hence is not considered.

10.4 Computational Details

10.4.1 Electronic Structure Calculations

A series of external static homogenous electric fields of different strengths were applied parallel to the three orthogonal directions (x, y, z) depicted in Fig. 10.1, in addition to a field-free set of calculations.

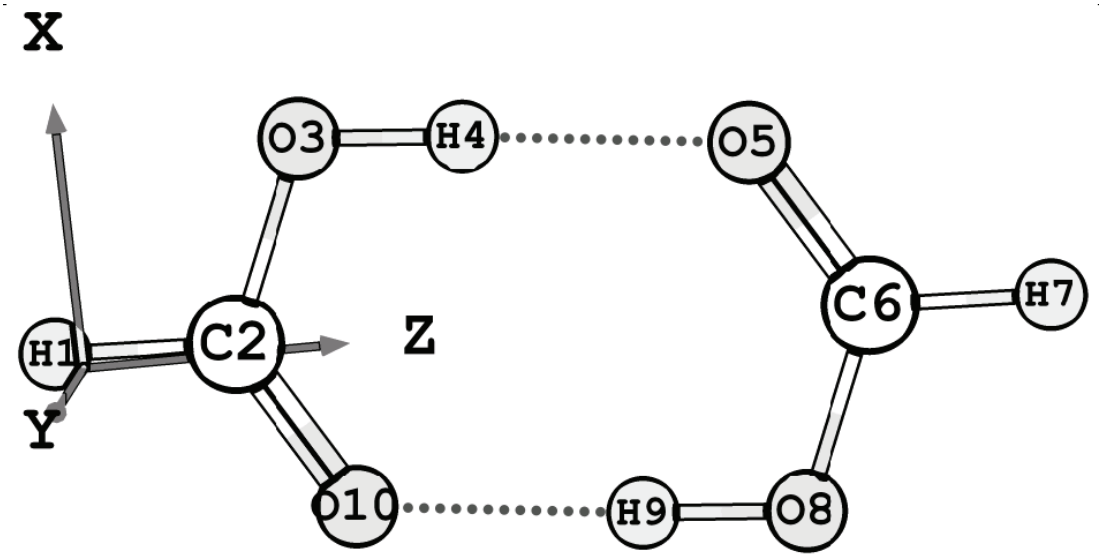


Figure 10.1: Ball-and-stick representation of the formic acid dimer showing the numbering scheme and the coordinate axes parallel to which the fields (**E**x, **E**y, **E**z) are applied. The distortion of the figure from the C_{2v} symmetry is an artifact of the perspective view and the tilting of the dimer to make the direction of the y-axis visible in the drawing (this applies also to Figs. 2 and 3).

The applied fields are of the following strengths (in 10^8 V m^{-1}): 5.14, 10.28, 20.56, 30.83, 41.12 and 51.40. Two levels of theory were used in this study: (1) DFT's hybrid functional B3LYP (three-parameter exchange functional of Becke and the correlation functional of Lee, Yang and Parr) with the 6-311++G(d,p) Pople basis set; and (2) MP2 using the same basis set. Gaussian 09 [118] was used for all electronic structure calculations. Relaxed scans of the potential energy surface were performed by varying the internuclear separation in one of the two hydroxyl groups (O3-H4, Fig. 10.1) from 0.7 to 2.5 Å in increments of 0.1 Å. Three scans were performed at every

field strength, one scan in each of the three perpendicular directions depicted in Fig. 10.1. The scans performed at the DFT level were used to locate the approximate position of the transition state (TS) at each combination of orientation and field strength. Each one of these approximate transition state geometries was subject to a TS optimization where the only constraint imposed on the optimized geometry is the presence of one and only one imaginary frequency. These TS optimizations starting from the approximate DFT guesses were performed at the two levels of theory (B3LYP and MP2) followed by a frequency calculation to confirm the existence of one and only one imaginary frequency as required for first-order saddle points. Unless mentioned otherwise, the energies (or energy differences) reported in this chapter are Gibbs energies (or Gibbs energy differences). The imaginary frequencies reported and tabulated in this chapter, and used in tunneling correction calculations, are the raw unscaled harmonic frequencies obtained directly from the Gaussian 09 output files [118].

10.4.2 Transition State Calculations and Proton Tunneling Correction

Conventional transition state theory (CTST) is used to provide estimates of the double proton transfer rate constant in the absence and presence of various field strengths. The (unimolecular) classical rate constant is expressed as:

$$k(T) = \kappa(T)\frac{k_B T}{h} exp\left(- \frac{\Delta G^{\ddagger}}{RT} \right) \tag{10.4.1}$$

where k_B, T, h, $\Delta G^{\ddagger}$, and R are the Boltzmann constant, absolute temperature, Planck's constant, Gibbs energy of activation, and the universal gas constant, respectively. The factor $\kappa(T)$ accounts for deviations from CTST due to mechanisms that tend to decrease the rate constant such as reflection, or those which tend to increase

it such as quantum mechanical leakage through the barrier (tunneling). Since the reaction studied involves the transfer of protons, the effect of tunneling on the reaction rate constant can be expected to be significant in view of the small mass of the proton. In this work, whereby we are primarily interested in orders of magnitudes and trends, the simple Wigner tunneling [226] correction suffices:

$$\kappa(T) = 1 + \frac{1}{24}\left[\frac{hIm(\nu^{\ddagger})}{k_B T}\right] \tag{10.4.2}$$

where $\mathrm{Im}(\nu^{\ddagger})$ is the imaginary part of the complex frequency along the reaction path characterizing the activated complex. To test the validity of the Wigner correction, Professor J. Raul Alvarez-Idaboy has communicated to us the recalculated κ (25^{o}C) for the field-free B3LYP using an Eckart barrier, $\kappa_{Eckart}(25^oC) = 4.39$. While larger than the corresponding Wigner correction [$\kappa_{Wigner}(25^oC) = 2.52$], it indicates that the Wigner correction is sufficient since it yields the correct order of magnitude and thus our conclusions should remain unchanged.

10.5 Conventions Used in this chapter

The two levels of theory B3LYP/6-311++G(d,p) and MP2/6-311++G(d,p) will be referred to as DFT and MP2. The DFT results are presented without brackets followed by MP2 results in brackets, e.g., 1.830 (1.832) Å. The field strengths of 5.14, 10.28, 20.56, 30.83, 41.12 and 51.40 (in 10^8 V m^{-1}) will be denoted, for simplicity, as 5, 10, 20, 30, 40 and 50, respectively, and the field-free case will be denoted as 0. Unless stated otherwise, when the results with and without field are compared, this implies a comparison between the field-free case and the case under the influence of the strongest field (i.e., field 50) directed parallel to the long axis of the dimer (z-axis), (Fig. 10.1).

10.6 Results and Discussion

10.6.1 Field-Effects on the Mechanism of the Double Proton Transfer in the Formic Acid Dimer

The potential energy scans in absence or presence of fields in all studied orientations and strengths revealed no intermediates, i.e., there exists only one transition state for each combination of orientation and strength of the external field (Fig. 10.2).

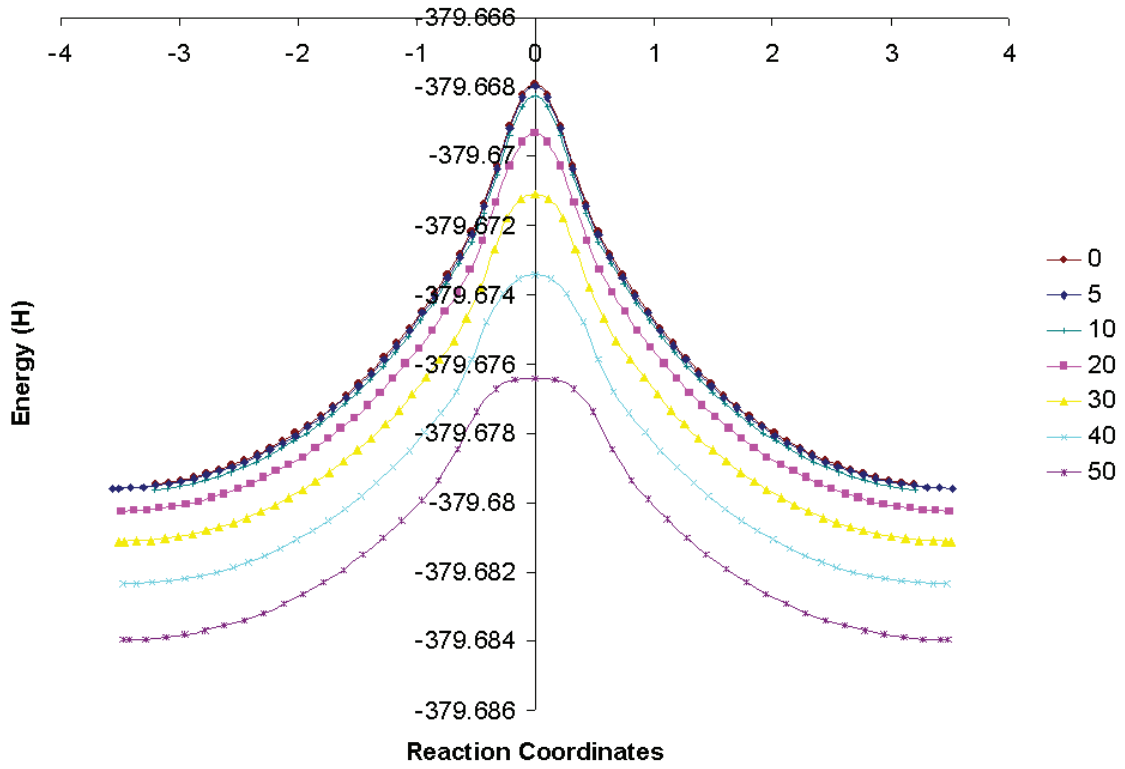


Figure 10.2: The potential energy surfaces of the double proton transfer reaction in formic acid dimer under the effect of electric fields with various strengths applied in the z direction.

The nature of each of these transition states was confirmed to exhibit one and only one imaginary frequency. Within the adiabatic approximation, the proton transfer is, thus, a concerted one step reaction in which both protons move in opposite directions simultaneously.

10.6.2 Preferred Orientation of the Formic Acid Dimer in an External Static Homogenous Electric Field

It is instructive to start by elucidating the preferred orientation of the formic acid dimer in the external field in the absence of constraints. The symmetry of the dimer reduces the number of unique principal orientations to be examined. For each field strength, there exists three principal orientations of the dimer (with the field parallel to the x-, y-, or z-axes). In every orientation, the energy of the dimer in the field is calculated before the double proton transfer reaction, and at the transition state. The most favorable orientation (in all field strengths and for both reactants and transition states) is found to be in the presence of an external electric field parallel to the z-axis, a field denoted by **E**z. In the remainder of this section, only maximal fields of maximum strength are considered, i.e., the fields **E**x, **E**y, and **E**z of strength equal to 5.14x10^9 V m^{-1}.

Let n_x, n_y, and n_z, represent the mole-fractions of dimers in the orientations defined by **E**x, **E**y, and **E**z, and let ΔG^o_{zx} and ΔG^o_{zy} be the difference in Gibbs energy of the system when the orientation of the field is switched from parallel to the z-axis to parallel to the x- and y-axis, respectively. The relative populations in the principal orientations with respect to the external field are given by

$$\frac{n_z}{n_i} = exp\bigg(- \frac{\Delta G^o_{zi}}{RT} \bigg), i = x, y \tag{10.6.1}$$

$$n_x + n_y + n_z = 1 \tag{10.6.2}$$

The relative populations are listed in Table 10.1. It is clear that the z-direction is the most favorable orientation at the global minima (reactants/products), as well as

at the activated complex (transition state) structure. At 25^{o}C, more than 97% of formic acid dimers at the transition state structure are orientated with their long axis parallel to the field. A smaller majority [80.9 (76.8) %] of dimers at the global minimum are also in this orientation with respect to the external field. This indicates that the system clearly favors this orientation aligning the field parallel/antiparallel to the direction of proton transfers.

Table 10.1: Mole fractions of formic acid dimers along the three principal orientations with respect to the external field in the global minimum (reactants/products) and at the activated complex (transition state) geometries. T = 298.15 K.

	Reactants		Transition state	
Mole fraction	DFT	MP2	DFT	MP2
n_z (%)	80.9	76.8	97.8	97.7
n_z (%)	12.0	15.7	1.6	1.7
n_z (%)	7.1	7.5	0.6	0.6

10.6.3 Field-Effects on the Geometry of the Formic Acid Dimer in the Reactants/Products and the Transition State of the Double Proton Transfer

Zero Field

The transition state structure of the formic acid dimer has a D_{2h} symmetry in the absence of an external electric field (Fig. 10.3).

The transferring protons, H4/H9, are equidistant 1.209 (1.201) Å from the two oxygen atoms, O3$\cdots$H4$\cdots$O5/O8$\cdots$H9$\cdots$O10, (see Table 10.2). The angles O3-C2-O10 and O5-C6-O8 are equal to 126.60^o (126.92)^o. The reactants/products of the double proton transfer in the absence of an externally applied field have a C_{2h} point group as can be seen from Fig. 10.3.

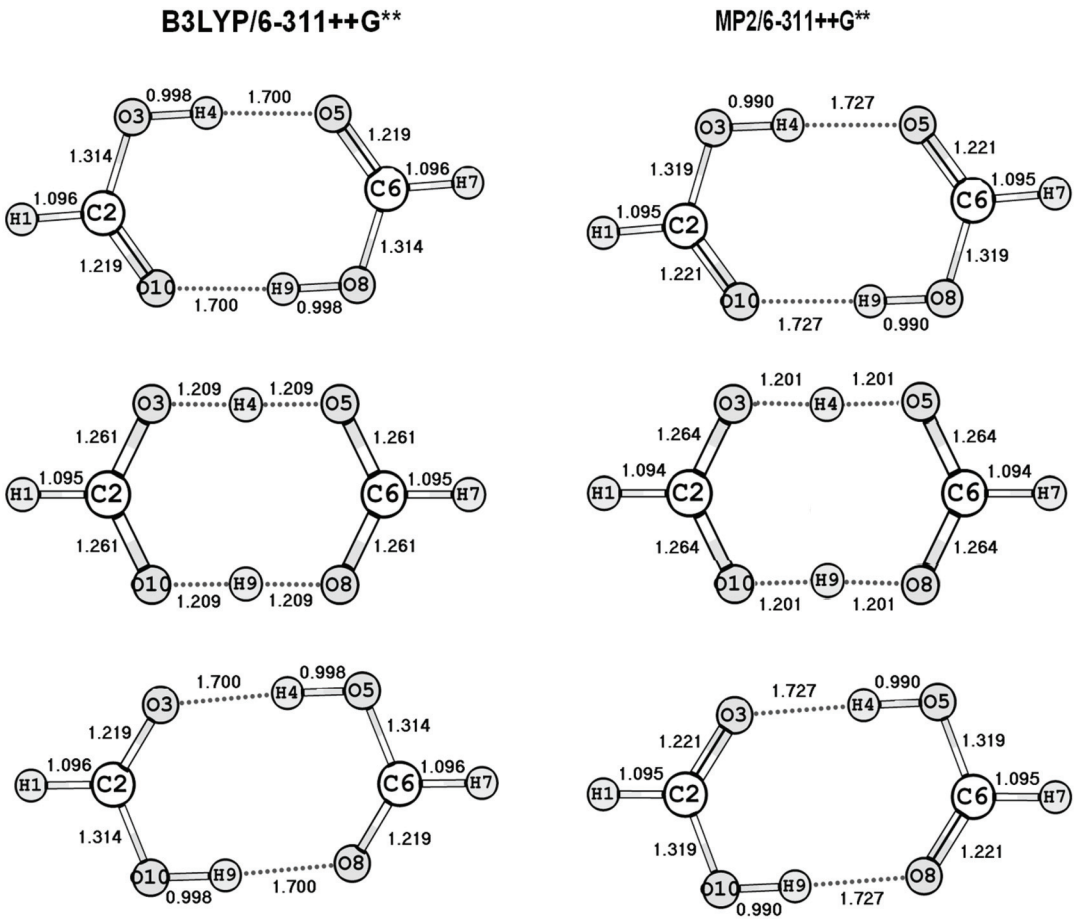


Figure 10.3: Ball-and-stick representation of reactants (top), transition state (middle), and products (bottom) of the double proton transfer reaction in the formic acid dimer in absence of external fields. Bond lengths are in Å.

Fields of Different Strengths in the z-Direction

Fields parallel to the z-axis affect the height and shape of the barrier considerably more than fields parallel to the x- or y-axes. Thus, emphasis will be placed on the geometric and energetic changes accompanying fields parallel to the z-axis.

Transition State

The transition state has a C_2 rotation axis aligned along the z-axis and belongs to the C_{2v} point group (Fig. 10.4).

In the presence of a field in the z-direction, bond angles remain essentially unchanged

Table 10.2: Interatomic distances (in Å) in the reactants, activated complex, and products of the double proton transfer reaction in the formic acid dimer in absence and presence of different field strengths parallel to the z-axis.

Field Strength ($x10^8$ Vm^{-1})	0		5.14		10.28		20.56		30.83		41.12		51.40	
Theory	DFT	MP2	DFT	MP2	DFT	MP2	DFT	MP2	DFT	MP2	DFT	MP2	DFT	MP2
	Transition State													
O3-H4	1.209	1.201	1.192	1.185	1.177	1.170	1.148	1.142	1.124	1.119	1.104	1.100	1.088	1.083
H4-O5	1.209	1.201	1.227	1.218	1.245	1.235	1.281	1.270	1.316	1.304	1.351	1.338	1.384	1.370
O8-H9	1.209	1.201	1.227	1.218	1.245	1.235	1.281	1.270	1.316	1.304	1.351	1.338	1.384	1.370
H9-O10	1.209	1.201	1.192	1.185	1.177	1.170	1.148	1.142	1.124	1.119	1.104	1.100	1.088	1.083
	Reactants													
O3-H4	0.998	0.990	0.996	0.988	0.995	0.987	0.992	0.985	0.989	0.982	0.987	0.981	0.985	0.979
H4-O5	1.700	1.727	1.715	1.741	1.729	1.755	1.758	1.785	1.789	1.817	1.822	1.851	1.857	1.887
O8-H9	0.998	0.990	1.000	0.991	1.002	0.993	1.006	0.996	1.010	1.000	1.015	1.004	1.021	1.009
H9-O10	1.700	1.727	1.687	1.713	1.673	1.699	1.646	1.672	1.620	1.645	1.593	1.619	1.566	1.593
	Products													
O3-H4	1.700	1.727	1.687	1.713	1.673	1.699	1.646	1.672	1.620	1.645	1.593	1.619	1.566	1.593
H4-O5	0.998	0.990	1.000	0.991	1.002	0.993	1.006	0.996	1.010	1.000	1.015	1.004	1.021	1.009
O8-H9	1.700	1.727	1.715	1.741	1.729	1.755	1.758	1.785	1.789	1.817	1.822	1.851	1.857	1.887
H9-O10	0.998	0.990	0.996	0.988	0.995	0.987	0.992	0.985	0.989	0.982	0.987	0.981	0.985	0.979

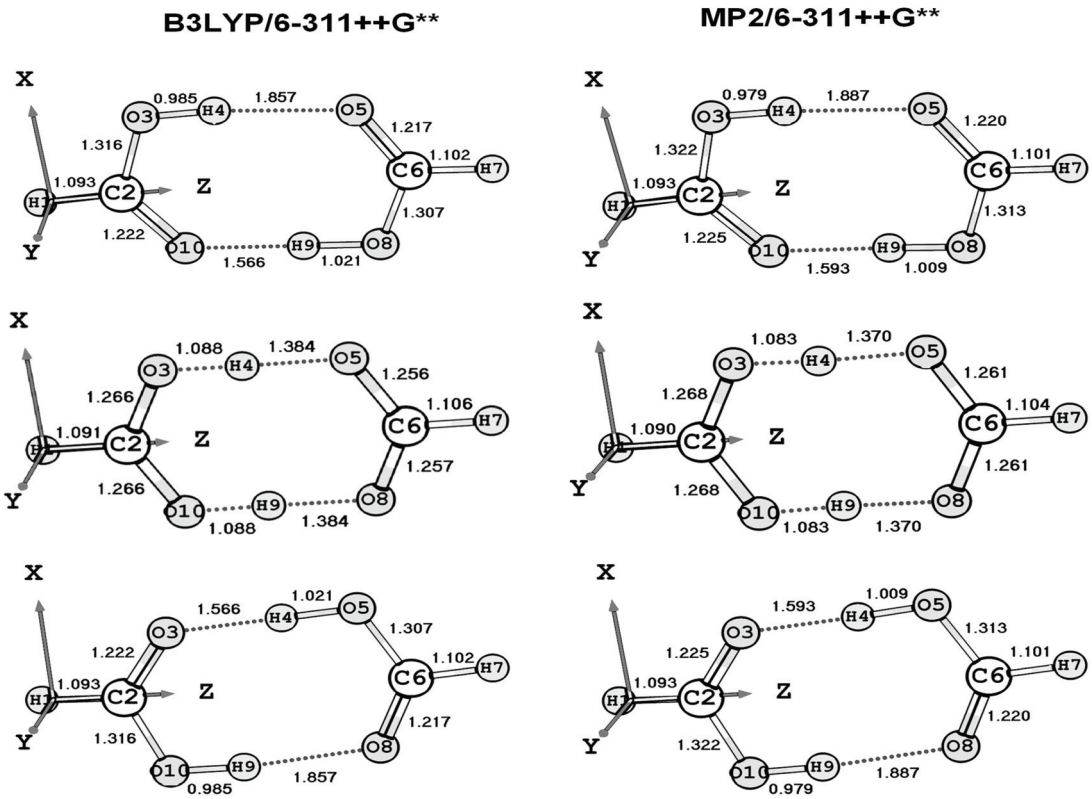


Figure 10.4: Ball-and-stick representation of reactants (top), transition state (middle), and products (bottom) of the double proton transfer reaction in the formic acid dimer in presence of $|\ \mathbf{E}z\ | = 5.14x10^9$ V m^{-1} (0.010 au). Bond lengths are in Å.

with O3-C2-O10 and H7-C6-O8 slightly decreasing by less than 1^o under the influence of the field. (The reader is reminded that when "field" is mentioned without specification of its strength or direction, this field is the strongest, i.e., the field labeled 50, along the z-direction).

Four bond lengths were significantly altered by the field. Distances O3-H4 and H9-O10 decrease from 1.209 (1.201) to 1.088 (1.083) Å, while distances H4-O5 and O8-H9 increase from 1.209 (1.201) to 1.384 (1.270) Å. Table 10.2 lists interatomic distances corresponding to bond lengths from the DFT and the MP2 calculations. A glance at this table reveals the close agreement between the two levels of theory in reproducing trends of field-effects on interatomic distances.

Reactants and Products

In the absence or presence of a field in the z-direction, the distances H4-O5, H9-O10, O3-H4 and O8-H9 in the reactants are equal to the distances O8-H9, O3-H4, H9-O10 and H4-O5 in the products (Table 10.2). In other words, the reactants and products are identical and related by a 180^o rotation around the z-axis for both the field free case (Fig. 10.3) and in the presence of **E**z (Fig. 10.3). We will thus adopt the notation "reactants/products" henceforth.

Bond angles for reactants/products remain nearly unchanged when fields of different strengths are applied in the z-direction. The field in the z-direction reduces the symmetry of the system. In the reactants/products, the field lengthens the H···O hydrogen bond and shortens the covalent O-H bond while inducing the opposite effect in the other arm of the dimer. The length of the hydrogen bond linking H4 and O5, dH4-O5, is 1.857 (1.887) Å, and the length of the O3-H4 bond, dO3-H4, equals 0.985 (0.979) Å (Table 10.2). The corresponding distances in the other hydrogen bonded arm of the dimer are dH9-O10 = 1.566 (1.593) Å and dO8-H9 = 1.021 (1.009) Å, respectively (Figs. 10.3, 10.4).

In the reactants/products, the presence of a field directed parallel to the z-axis has

minimal influence on the H1-C2 and C6-H7 distances. This field causes the C2-H1 to shorten by 0.003 Å from 1.096 (1.095) Å to 1.093 (1.092) Å; while C6-H7 is stretched by 0.007 Å from 1.096 (1.095) Å to 1.102 (1.101) Å.

The transition state structure undergoes similar changes under the influence of the field. Thus, bond angles remain essentially unchanged with the field. C-H distances dH1-C2 and dC6-H7 undergo only slight modifications: dH1-C2 decreases from 1.095 (1.094) to 1.091(1.090) Å; while dC6-H7 increases from 1.095 (1.094) to 1.106 (1.104) Å. The "stacked" O-H arms are equal, i.e., bond distances dO3-H4 = dH9-O10 = 1.088 (1.083) Å; while bond distances dH4-O5 = dO8-H9 = 1.384 (1.370) Å (Table 10.2).

The transition state has a higher symmetry than reactants/products. In the absence of a field, the point groups of the product/reactant and of the transition state are C_{2h} and D_{2h}, respectively. In the presence of a field, point groups of products/reactants and transition state structure are C_s and C_{2v}, respectively.

Furthermore, the distance between the monomers changes significantly in the transition state with respect to the reactants/products. The distance between the two monomers may be gauged by the inter-oxygen distance, i.e., the O3-O5 (or O8-O10) distance. In the absence or presence of all field strengths, the O-O distance in the reactants/products is ∼2.7 Å, to one decimal place, a distance significantly shortened in all of the respective transition state to only ∼2.4 Å, i.e., the monomers are closer in the transition state by ∼0.3 Å.

10.6.4 Field-Effects on the Barrier Height of the Double Proton Transfer

Fig. 10.4 depicts a ball-and-stick representation of the geometries (labeled with bond lengths) of the reactants, transition state, and products of the double proton transfer, calculated at the two levels of theory. Since, as discussed above, the reactants and

the products are identical whether in the presence or the absence of **E**z, the calculated forward and backward barriers are the same and their distinction will, thus, be dropped in the remainder of this chapter. The energies and barrier heights are listed in Table 10.3.

Table 10.3: Transition states imaginary frequencies, Gibbs energies (and their differences), raw and corrected reaction rate constants, and Wigner tunneling corrections in absence and presence of fields of varying strengths and directions.

Field Strengths /10^8 V m^{-1}	0.00	5.14	10.28	20.56	30.83	41.12	51.40
Field in the z-Direction (DFT)							
Im($\nu^{\ddagger}$) (cm^{-1})	1253.03i	1239.90i	1200.85i	1048.91i	816.18i	554.79i	342.12i
G‡ (au)	-379.6341	-379.6341	-379.6344	-379.6352	-379.6366	-379.6385	-379.6409
G(reactants) (au)$^{(a)}$	-379.6405	-379.6406	-379.6407	-379.6413	-379.6423	-379.6437	-379.6455
ΔG (kcal mol^{-1})	4.06	4.04	3.99	3.81	3.54	3.24	2.90
k_{298K} (s^{-1})	6.54x10^9	6.76x10^9	7.36x10^9	9.97x10^9	1.57x10^{10}	2.61x10^{10}	4.64x10^{10}
κ_{298K} tunnel	2.52	2.49	2.40	2.07	1.65	1.30	1.11
k_{298K} corrected (s^{-1})	1.65x10^{10}	1.68x10^{10}	1.76x10^{10}	2.06x10^{10}	2.59x10^{10}	3.39x10^{10}	5.16x10^{10}
$t_{1/2}$ (ps)$^{(b)}$	42.03	41.15	39.28	33.63	26.77	20.45	13.43
Field in the z-Direction (MP2)							
Im($\nu^{\ddagger}$) (cm^{-1})	1351.00i	1338.03i	1299.79i	1152.68i	926.64i	657.59i	420.12i
G‡ (au)	-378.6992	-378.6992	-378.6995	-378.7003	-378.7018	-378.7036	-378.7060
G(reactants) (au)$^{(a)}$	-378.7083	-378.7084	-378.7085	-378.7090	-378.7100	-378.7113	-378.7130
ΔG (kcal mol^{-1})	5.73	5.72	5.67	5.46	5.15	4.78	4.37
k_{298K} (s^{-1})	3.89x10^8	3.96x10^8	4.31x10^8	6.14x10^8	1.04x10^9	1.94x10^9	3.87x10^9
κ_{298K} tunnel	2.77	2.74	2.64	2.29	1.83	1.42	1.17
k_{298K} corrected (s^{-1})	1.08x10^9	1.08x10^9	1.14x10^9	1.41x10^9	1.90x10^9	2.75x10^9	4.53x10^9
$t_{1/2}$ (ps)$^{(b)}$	642.47	639.52	609.53	492.90	364.64	252.06	152.86

Table 10.3 continued ...

Field in the x-Direction, DFT							
Im($\nu^{\ddagger}$) (cm^{-1})	1253.03i	1253.04i	1253.07i	1253.20i	1253.41i	1253.71i	1254.09i
G‡ (au)	-379.6341	-379.6341	-379.6342	-379.6345	-379.6351	-379.6360	-379.6370
G(reactants) (au)	-379.6405	-379.6406	-379.6407	-379.6410	-379.6417	-379.6425	-379.6437
ΔG‡ (kcal mol^{-1})	4.06	4.06	4.06	4.07	4.10	4.13	4.17
Field in the x-Direction, MP2							
Im($\nu^{\ddagger}$) (cm^{-1})	1351.00i	1350.99i	1350.96i	1350.84i	1350.56i	1350.31i	1349.98i
G‡ (au)	-378.6992	-378.6992	-378.6993	-378.6996	-378.7002	-378.7011	-378.7022
G(reactants) (au)$^{(a)}$	-378.7083	-378.7083	-378.7084	-378.7088	-378.7094	-378.7103	-378.7115
ΔG‡ (kcal mol^{-1})	5.73	5.74	5.74	5.75	5.77	5.80	5.84
Field in the y-Direction, DFT							
Im($\nu^{\ddagger}$) (cm^{-1})	1253.03i	1253.07i	1253.19i	1253.68i	1254.50i	1255.66i	1257.17i
G‡ (au)	-379.6341	-379.6341	-379.6341	-379.6344	-379.6348	-379.6353	-379.6361
G(reactants) (au)$^{(a)}$	-379.6405	-379.64	-379.6406	-379.6410	-379.6415	-379.6423	-379.6432
ΔG‡ (kcal mol^{-1})	4.06	4.06	4.08	4.13	4.22	4.34	4.51
Field in the y-Direction, MP2							
Im($\nu^{\ddagger}$) (cm^{-1})	1351.00i	1350.95i	1351.10i	1351.66i	1352.66i	1354.06i	1355.91i
G‡ (au)	-378.6992	-378.6992	-378.6992	-378.6995	-378.6999	-378.7004	-378.7011
G(reactants) (au)$^{(a)}$	-378.7083	-378.7083	-378.7083	-378.7086	-378.7090	-378.7098	-378.7108
ΔG‡ (kcal mol^{-1})	5.73	5.73	5.71	5.69	5.75	5.87	6.05

(a)By symmetry, G(products)=G(reactants), for any given field.
(b)$t_{1/2}$= ln 2/k.

The energies of all species (reactants/products, and transition state) decrease as a function of the magnitude of **E**z. Fields stabilize the transition state more than they stabilize the reactants/products, and hence fields parallel to the z-direction decrease the barrier height.

The calculated Gibbs energies and barrier heights are collected in Table 10.3. As can be seen from this table, the trends from the MP2 and DFT results are remarkably consistent, especially with respect to the effect of fields on barrier heights. The agreement in trends between the two levels of theory can also be construed from Fig. 10.5.

The results in the Table 10.3 and Fig. 10.5 show that the most significant effect on

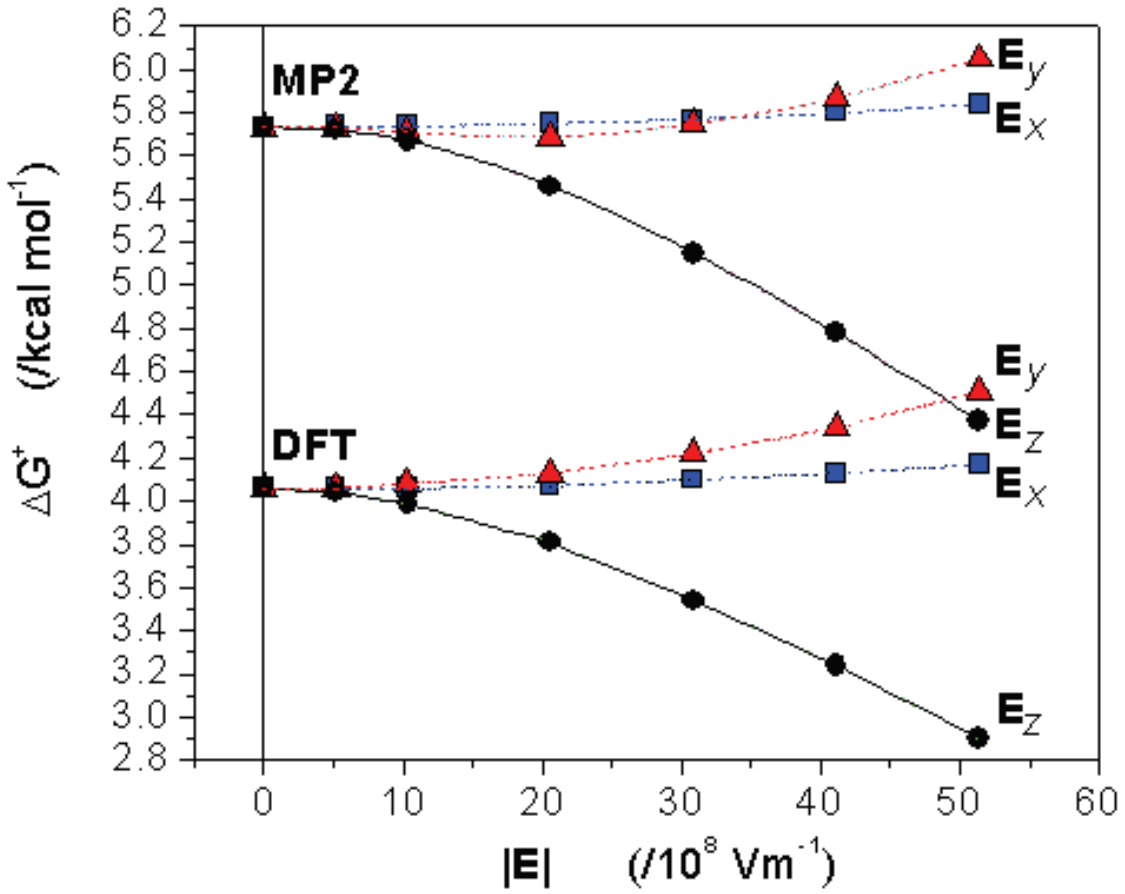


Figure 10.5: Barrier heights ($\Delta G^{\ddagger}$) as a function of field strength, orientation, and level of theory in kcal mol^{-1}. The fields parallel to the x-, y- and z-axes are labeled with the corresponding subscripts and symbolized by blue squares, red triangles and black circles, respectively. Interpolated curves linking the data points were added to guide the eye. The curves originating at the top-left are calculated at the MP2/6-311++G(d,p) level and those originating at the bottom left at the B3LYP/6-311++G(d,p) level.

the barrier height is achieved when the external field is parallel to the z-axis, and that this effect increases significantly with the field strength. It is important to note that the MP2 barriers are systematically higher than the corresponding barriers calculated at the DFT level of theory by ∼1.5 kcal mol^{-1}. This difference translates into more than one order of magnitude change in rate constants.

The field applied in the x-direction causes the barrier to increase from 4.06 (5.73) kcal mol^{-1} (in the absence of field) to 4.17 (5.84) kcal mol^{-1} at the maximum field intensity. Similar slight increase in the barrier height is also achieved by the application of fields

parallel to the y-axis (perpendicular to the plane of the molecule). The MP2 barrier height remains nearly constant (decreases insignificantly from 5.73 to 5.69 kcal mol^{-1}) with field strengths from 0 to 20, then increases gradually to 6.05 kcal mol^{-1} as the field increases from 20 to 50. In the case of DFT, the barrier height consistently increases from 4.06 to 4.51 kcal mol^{-1} for field strengths from 0 to 50. At the maximal field strength (field "50") in the z-direction, the barrier height is lowered to ∼76 (71) % of its field-free value.

10.6.5 Field-Effects on the Imaginary Vibrational Frequencies

The imaginary vibrational frequencies at the transition states (DFT and MP2) are listed in Table 10.3. The magnitudes of the imaginary frequencies obtained from DFT calculations are systematically lower by (∼100 cm^{-1}) than the corresponding frequencies at the MP2 level of theory, whether in absence or presence of field. In the field-free case, these frequencies are $1253.03i$ ($1351.00i$), at the DFT (MP2) levels, respectively. Fields oriented parallel to the x- or y- axes have a minor effect on the imaginary frequency at all strengths studied, as can be seen from Table 3. The averages (± standard deviation) of the magnitudes of the imaginary frequency in the presence of **E**x and **E**y are: $\left\langle Im(v^{\ddagger})_{\mathbf{Ex}} \right\rangle = 1253.42i \pm 0.38 (1350.61 \pm 0.37) \mathrm{cm}^{-1}$; and $\left\langle Im(v^{\ddagger})_{\mathbf{Ey}} \right\rangle = 1254.55i \pm 1.47 (1352.72 \pm 1.77) \mathrm{cm}^{-1}$.

Fields aligned (anti)parallel to the direction of proton transfer, **E**z, have considerably more significant effect on the magnitude of the imaginary vibration than **E**x or **E**y, the effect being most pronounced above a field strength threshold around 10^9 V m^{-1} (Fig. 10.6). Beyond this threshold, **E**z fields cause a significant decrease in the magnitude of the imaginary frequency.

Since a low $Im(v^{\ddagger})$ is associated with a small curvature of the PES at the transition state region [227], the decrease in the magnitude of the imaginary frequency with

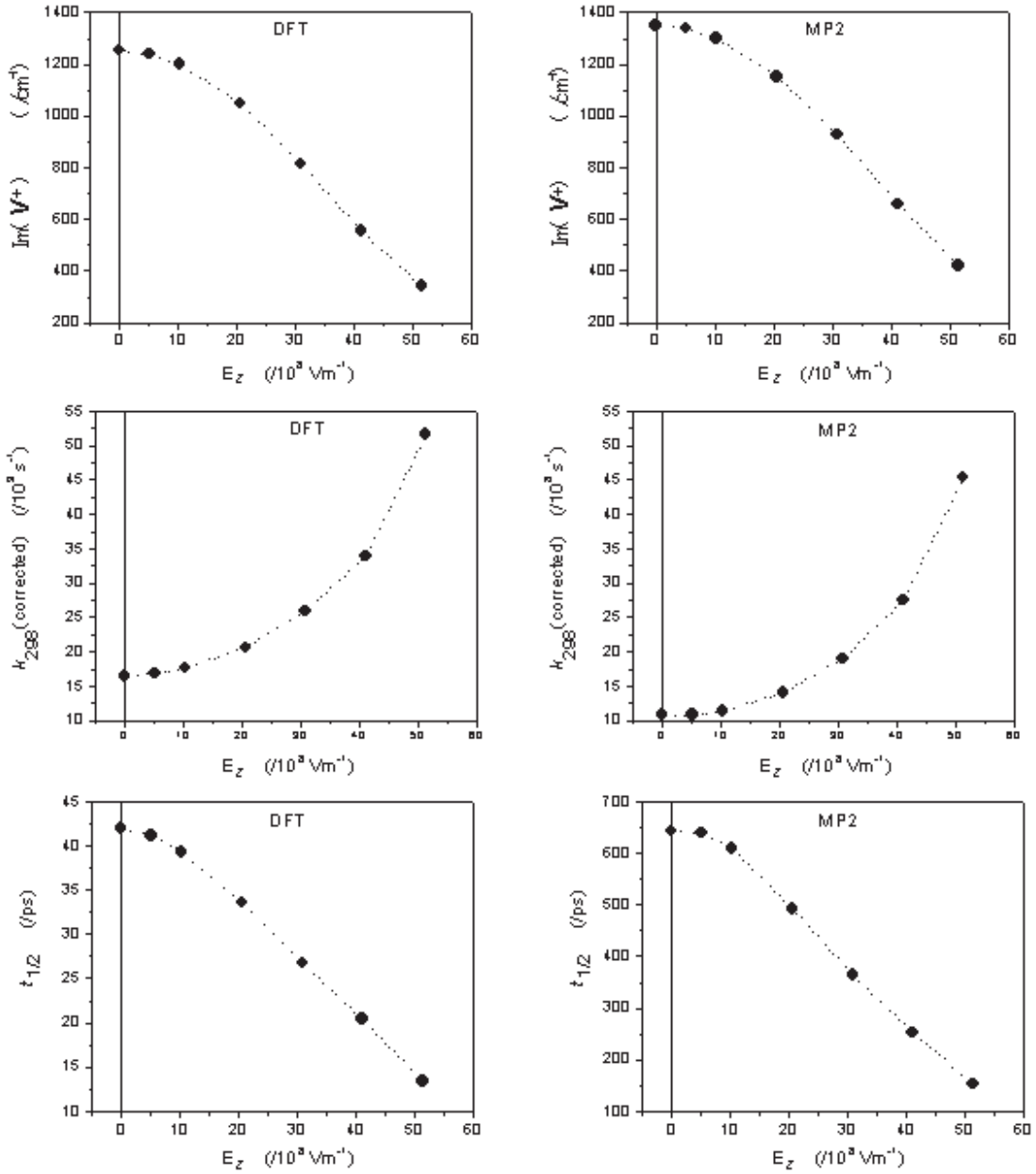


Figure 10.6: Magnitude of the imaginary frequency [Im($\nu^{\ddagger}$) / cm^{-1}] (top), first order rate constant (k / s^{-1}) at 25^{o}C (298.15 K) obtained from CTST [Eq. 11.4.1] after Wigner's tunneling correction [Eq. 11.4.2] (middle), and half life [$t_{1/2}$ / picoseconds] (bottom) of the double proton transfer in the formic acid dimer as a function of | **E**z |(in V m^{-1}) and level of theory. Plots on the left are based on B3LYP/6-311++G(d,p) calculations, those on the right derive from the MP2/6-311++G(d,p) results. (Note the one order of magnitude reduction in scale of the middle-right plot reporting k_{298} at the MP2 level compared to the corresponding plot obtained at the DFT level (middle-left).

increasing field strength indicates that strong z-fields flatten the potential energy surface considerably near the transition state region. As can be gleaned from Table 10.3 and Fig. 10.6, the imaginary frequency drops from its field-free value $1253i$ ($1351i$) to $342i$ ($420i$) cm^{-1} at maximal field strength ($5.14x10^9$ V m^{-1}). The magnitude of the imaginary frequency, thus, decreases by ∼75 (70) % under the influence of the maximal field intensity in the z-direction when compared to the zero-field case. The lowering of the barrier brought about by the application of intense **E**z is, thus, accompanied with a considerable field-induced flattening of the PES.

10.6.6 Field-Effects on the Unimolecular Reaction Rate Constant of the Double Proton Transfer

The first order reaction rate constants k(T) calculated from CTST at 25^{o}C (T = 298.15 K) in the absence and presence of all studied fields parallel to the z-axis (at both levels of theory) are collected in Table 10.3.

The entry "k_{298}" in Table 10.3 is the classical unimolecular rate constant calculated from Eq. (11.4.1), i.e., when the dimensionless non-classical correction factor is unity ($\kappa_{298K} = 1$). As mentioned above, the ∼1.5 kcal mol^{-1} systematic difference in the calculated barrier heights between MP2 and DFT is magnified, through the exponential function, to more than one order of magnitude difference in the value of the rate constant. Thus, the uncorrected CTST rate constants at the DFT level range from $6.59x10^9$ s^{-1} in the field-free case to $4.46x10^{10}$ s^{-1} at the limit of strongest field strength, while the corresponding values at the MP2 level are $3.89x10^8$ s^{-1} and $3.87x10^9$ s^{-1}, respectively (Table 10.3).

Using the magnitudes of the imaginary frequencies calculated at the DFT (MP2) levels at 25^{o}C, the Wigner correction [226] (κ_{298K}) ranges from 2.52 (2.77) for the field-free case, to 1.11 (1.17) for the case of maximal field intensity. While tunneling corrections can be more accurately predicted by more sophisticated treatments

such as that of Skodje and Truhlar [228] especially for $\kappa > 2$, this is not expected to give rise to a qualitatively different picture in this case since $\kappa < 2.8$. For small to moderate tunneling ($1.1 < \kappa < 4$) [229], the Wigner correction should be sufficiently accurate as has been confirmed by comparing the DFT field-free value with the one calculated from an Eckart potential.

With the inclusion of the Wigner tunneling correction, the corrected rate constants range from 1.65x10^{10} (1.08x10^{9}) s^{-1} for the field-free case to 5.16x10^{10} (4.53x10^{9}) s^{-1} in the case of the strongest field along the z-direction (Table 10.3 and Fig. 10.6). In other words, the tunneling-corrected net effect of the strongest **E**z field is to approximately triple (quadruple) the rate constant of the double proton transfer reaction in the formic acid dimer at 25^{o}C. The field-induced magnification of k_{298K}(corrected) is reflected on the half-life ($t_{1/2}$) of the dimer, a time that decreases quickly with field strength as can be seen from Table 10.3 and Fig. 10.6.

As can be anticipated from Fig. 5, the tunneling-corrected rate constant is exponentially-correlated to the z-field strength. The following equations represent the lines of best exponential fit to the data:

$$k_{298K}^{DFT}(corrected) \approx 1.312 \times 10^{10} exp(2.53 \times 10^{-10} \mid \mathbf{E}z \mid),$$
$$[n = 7, r^2 = 0.953] \tag{10.6.3}$$

and

$$k_{298K}^{MP2}(corrected) \approx 7.508 \times 10^{8} exp(3.39 \times 10^{-10} \mid \mathbf{E}z \mid),$$
$$[n = 7, r^2 = 0.964] \tag{10.6.4}$$

where the pre-exponential constants have the dimensions/units of k (s^{-1}) and the constants in the exponents have the dimensions/units of reciprocal electric field (m V^{-1}), and $\mid \mathbf{E}z \mid$ is in V m^{-1}. The strength of the above correlations is reflected

in both the r^2 and the root mean square deviation (RMSD) of the fitted and the calculated values of k_{298K}(corrected), being slightly stronger for in the case of the MP2 results (Eq. 10.6.4).

10.7 Conclusions

The double proton transfer reaction in the formic acid dimer has been studied in the presence of external static homogenous electric fields of various strengths in three perpendicular directions. The reaction path linking reactants and products is found to exhibit a single transition state without intermediate species in the absence or in the presence of all of the studied fields. The lack of a "double/multiple hump(s)" along the reaction path indicates that the double proton transfer in the formic acid dimer occurs synchronously within the adiabatic approximation.

The transition states have a higher symmetry than the reactants or products. When the applied field is parallel to the z-direction, the reactants and products belong to the C_s symmetry point group, while the transition states belong to the C_{2v} group. The relative Boltzmann orientations in an external field indicate a strong tendency for the formic acid dimers in vacuum-phase to align their long axis (the line passing through the carbons) parallel to the field, i.e., parallel to the z-direction. Mole fractions of ∼77% (MP2) - 81% (DFT) of reactants/products and ∼98% (MP2 and DFT) of transition states are aligned with the C-C axis parallel to the external field. Fields in the x- and y-direction (**E**x, **E**y, respectively) do not have a significant effect on the potential energy barrier height of this reaction and were not considered in detail. In contrast, fields parallel to the C-C axis (**E**z) have a considerably effect on the barrier height. The most intense **E**z fields ($|\ \mathbf{E}z\ | = 5.14 \times 10^9$ V m^{-1}) decrease the barrier height by ∼25% (MP2) - 30% (DFT). The same field also decreases the magnitude of imaginary frequency considerably by ∼70% (MP2) - 75% (DFT). Since the imaginary frequency measures the curvature of the potential energy surface along

the reaction path at the transition state region, this means that the **E**z significantly flattens the potential energy surface at or near the transition state region.

The flattening of the potential energy surface along the reaction path near the transition state is accompanied with a significant decrease in the rate of tunneling through the barrier. The flattening of the surface, thus, reduces the overall (corrected) rate constant. Therefore, **E**z elicits two opposing effects on the rate constant: On one hand it increases the raw rate constant (k), that is, k uncorrected for tunneling, by decreasing the barrier. On the other hand, it reduces the magnitude of the dimensionless tunneling multiplicative correction κ and hence tends to slow the decrease in k achieved through the lowering of the barrier. The combined effect of these two opposing contributions is a net relative increase of the tunneling-corrected k with the magnitude of **E**z, albeit around half as fast as the relative increase or raw (uncorrected) k with field strength (Table 10.3). Thus, the rate of reaction increases with increasing field strength in the z-direction, an effect particularly notable beyond a field-strength threshold of $\sim 10^9$ V m^{-1}. With the inclusion of tunneling corrections, the value of k (at 25^{o}C) increases by a factor between $\sim$3.1 (DFT) and $\sim$4.2 (MP2) by going from the field-free case to an $\mid$ **E**z $\mid$ of 51.40x10^8 V m^{-1} (the strongest field strength considered in this work).

The intensity of field necessary to double of the speed of the double proton transfer reaction is a little higher than $\sim$3x10^9 V m^{-1}. This field strength is typically encountered in the gap between the tip and the sample of an STM [208], and represents the order of magnitude of fields in the microenvironment of biological molecules such as DNA, enzymes active sites, or due to solvent fluctuations. It is recommended, thus, to consider the effect of strong local electric fields on the kinetics of reactions likely to take place in such electrical environments.

Chapter 11

Effects of Intense Electric Fields on the Double Proton Transfer in Watson-Crick DNA Base Pairs

11.1 Abstract

This chapter investigates the effect of strong electric fields encountered in the microenvironment of DNA on the double proton transfer reactions in DNA base pairs, adenine-thymine and guanine-cytosine. Fields in the order of 10^8 to 10^9 V m^{-1} were applied in two opposite directions along the movement of the protons. Using the density functional theory with a triple zeta Pople basis set, the results show that while fields applied in the positive direction elicit an increase of the tunneling-corrected rate of the double proton transfer in the forward direction, the opposite effect is observed for the reverse reaction. However, when fields are applied in the negative direction, the rate of both the forward and reverse reactions, in general, increases exponentially with stronger fields.

11.2 Introduction

External electromagnetic fields can considerably alter the topography of the potential energy surfaces of gas-phase proton- [212, 230], double proton- [231], hydrogen- [230, 232, 233, 234], or hydride-transfer reactions [212]. Intense electric fields 5.14x10^8 - 5.14x10^9 V m^{-1} can also determine the selectivity of a particular reaction channel of bond activation by porphine-based catalysts between two competing channels (whether the reaction proceeds to C-H hydroxylation or C=C epoxidation) [235].

There is also ample literature examining the biological effects of electric fields [194, 195, 196, 197, 231]. As an example of a direct observable biological field effects of static external electric fields, the frequency of sex-linked recessive lethal mutations in *Drosophila melanogaster* females exposed to static electric fields of 1.5 and 3.3x10^5 V m^{-1} for 24-hours increase significantly [198]. Further, weak residual external electrostatic fields to which a cell may be exposed can locally be amplified 4x10^6 times in proximity of sharp edges of the junction joining a pair of dividing cells [199]. Such locally-amplified fields are believed to inhibit cell proliferation [199]. Electrochemotherapy is an entire domain of biomedical research concerned with the enhancement of the chemotherapeutic effect of certain drugs by external electric fields [200].

Intense local electric fields ($\sim 10^8$ to 10^9 V m^{-1}) in the normal cellular microenvironment have been demonstrated to have significant effects on the kinetics of a model double proton transfer reaction [231]. In the previous chapter, the proton exchange within a formic acid dimer has been used as a simple model for a DNA Watson-Crick (WC) dimer where such proton exchange has long been postulated as a molecular basis for spontaneous and induced mutations [186, 187, 189, 190, 191, 192, 193, 203, 204]. In double-helical DNA, in the absence of external fields, the relatively high potential barrier for the forward reaction at room temperature ($\Delta G^{\ddagger} \geq 10$ kcal mol^{-1}) results in very little reaction ($K_{eq} \sim 10^{-8}$), which accounts for the great stability of DNA and

the fidelity with which the genetic code is preserved and transmitted to the daughter cells. The small remaining probability of reaction is a mechanism, first proposed by Löwdin, for the accumulation of genetic errors with the passage of time during the life of the organism leading to mutations that cause aging and possibly cancer.

The present study explores the effect of very strong electric fields on the double proton transfer reaction in WC dimers of DNA. While DNA base pairs are surrounded by the polar/ionic sugar-phosphate backbone, counterions, and histone proteins with their cationic side chains, these sources and sinks of electrostatic field are likely balanced. However, ionic compounds that interact with DNA can generate very intense local field.

Local microscopic electric fields are ubiquitous in biology and are central to nerve excitation, transport processes, and energy transduction [151, 236, 237, 238, 239, 240, 241, 242, 243, 244]. Several authors have shown an order of magnitude change in the rate of photosynthetic reactions under the influence of electric fields of the order of 10^8 V m^{-1} [245, 246, 247]. Murgida and Hildebrandt demonstrated that charge transfer across biological interfaces can be controlled by the modulation of activation energies through induced electric fields [248]. Multiple studies investigated the decrease in the enzymatic activity of *cytochrome c* oxidase under the influence of an electric field [249, 250, 251, 252]. Wackerbarth and Hildebrandt studied the structural (conformational) and dynamical changes in proteins under the influence of strong electric fields of strength $8x10^9$ V m^{-1} which are similar to the strengths of local electric fields in interfacial regions of biomembranes [253]. These authors studied the native state and another conformational state of *cytochrome c* adsorbed on the electrical double layer of a silver electrode/electrolyte interface. It was also shown that similar electric field strengths ($\sim 10^8$ V m^{-1}) affect the kinetics of charge recombination in bacterial reaction centers by an order of magnitude [245, 246, 247].

Other examples of the importance of strong local electric fields in a biological context include the effect of pH and local electric field generated from biomembranes on

proton binding energy in the light-activated proton pump bacteriorhodopsin [254]. Other studies have demonstrated the blocking of the transmembrane proton transfer though affecting the photocycle of oriented bacteriorhodopsin in the presence of an electric field [255, 256]. Choi and Moon investigated the effects of high electric fields on the structure of an ion-exchange membrane [257]. Simons studied the effects of electric fields on the proton transfer between membrane-bound amines and water [258]. The group of Zewail reported a strong dependence of the rate of proton transfer on polarity in the base pair in 7-azaindole (an analogue of AT base pair) in the condensed phase [259]. Electric fields are also well known to influence (enhance) the chemotherapeutic effect of certain drugs, in what became known as "electrochemotherapy" [202, 260, 261].

Strong electric fields are, thus, frequently encountered in the microenvironment of biomolecules and of biochemical reactions. This chapter reports an investigation of the effects of strong static homogenous electric fields on the rate constant of the double proton transfer in the two WC base pairs: AT and GC. So far, the dynamics of this reaction has been studied without the explicit inclusion of strong external fields into the calculations. There are several studies that were performed in the presence of (explicit or continuum) solvation and/or ions, and these calculations, naturally, include the field effects of the solvent or ions. However, as a sampling of the literature demonstrates, these studies do not generally study, systematically, the effects of field strengths and orientations on the kinetics. Gorb *et al.*, for instance, studied the double proton transfer in AT and GC in the gas phase and in the presence and absence of water molecules and metal ions [191]. These authors found (computationally using B3LYP/6-31G(d) and MP2/6-31G(d) levels of theory), that the rare tautomeric form A*T* does not exist while that of the equilibrium constant between GC and G*C* at room temperature [MP2/infinite//MP2/6-31G(d)] is estimated to be $2.0\text{x}10^{-6}$ s^{-1}. Water molecules were found to stabilize GC more than metal ions do. A comparison study of double proton transfer (DPT) in isolated GC versus embedded-GC in a dou-

ble stranded DNA sequence (GCGCG) was undertaken by Zoete and Meuwly [262]. They found that the barrier height is 18.8 and 18.7 kcal mol^{-1} for the isolated and embedded GC, respectively. Liu *et al.* investigated the electronic promotion effect of DPT on the conduction of DNA [263]. They found that the DPT lowers the ionization potentials (by 0.38 and 0.25 eV) for GC and AT, respectively), and it enhances the charge transfer along the DNA duplex. None of these studies was conducted in the presence of an explicit and well-characterized electric field with a single exception we are aware of: A semiempirical (AM1) investigation of the effect of strong electric fields on the double proton transfer in DNA base pairs has demonstrated the existence of a double-well potential under either strong external electric fields or when the base pairs are surrounded by a polarisable medium (continuum solvent) [224].

The previous chapter described the effect of explicit well characterized external electric fields on the kinetics of the double proton transfer in the formic acid dimer [231]. Two electronic structure methods have been explored in the aforementioned study, a perturbational approach (second order Møller-Plesset perturbation theory, or MP2) and a density functional approach (DFT) using the B3LYP functional. External fields of gradually increasing strength in the range (10^8 to 10^9 V m^{-1}) have been applied and their effects on the alignment of the free dimers, their geometries, and the potential energy surfaces (PES) of the DPT reaction have been studied. After tunneling corrections, it was found that the strongest applied field (5.14x10^9 V m^{-1}) enhances the rate constant by a factor of 3 - 4 at 25^{o}C. B3LYP and MP2 gave similar qualitative results, but there exists a systematic bias of $\sim$1.4 kcal mol^{-1} in the ΔG barriers (MP2 barriers being higher than the corresponding DFT barriers). This is enough to alter the vales of k and $t_{1/2}$ by an order of magnitude. In the present work, and in view of the relatively large size of the two WC dimers and the extensive calculations necessary to systematically alter the orientations field strengths, the DFT/B3LYP level is used.

The goal of the present study is to extend the previous one [231] by considering the

biochemical double proton transfer reaction in the hydrogen-bonded AT and GC base pairs. The influence of strong external electric fields on the rate constant of double proton transfer in base pairs is investigated to obtain insight about the effects of these fields on the rate/probability of mutation brought about by this mechanism.

11.3 Field Strengths

In a WC double stranded B-DNA, the anomeric carbon atoms of two deoxyribose sugar groups are separated by ∼10 - 11 Å, the diameter of the double helix is in the vicinity of 20 Å, while an enzyme active site can be even larger in size. The electric field at the center of a 10 Å dipole where the charges are of a magnitude of 0.5 au is 5.14x10^9 V m^{-1} in vacuum, a value that decreases by an order of magnitude if the distance between the charges is 34 Å. In view of these considerations, the chosen strengths of the external electric field in this study are (in 10^8 V m^{-1}): 0.00, 5.14, 12.86, 25.71, 36.00, 38.57, and 51.42.

To ensure that these fields strengths do not ionize the DNA dimers by field-induced tunneling of electron from the dimer, the frequency of tunneling (ω) of an electron is calculated according to Eq. 9.1.1 in the introductory chapter.

The electric fields are applied in the two opposite directions along the long axis of the dimer (±x directions) (see Fig. 11.1).

Table 11.1 lists the inverse ($1/\omega$) of the electron tunnelling. $1/\omega$ may be interpreted as the average (characteristic) time for an ionization event.

Under the strongest field the shortest calculated characteristic tunnel ionization time of the GC base pair is ∼8.4x10^{-8} s; the lifetime of GC is ∼3 ps (similar to the half life calculated in this study, for the reverse reaction in the absence of field, 5x10^{-13}). One study reports that the estimated frequency of double proton transfer in GC to be 10^{-7} to 10^{-9} s^{-1} [205]. Another study reports that the timescale for the double

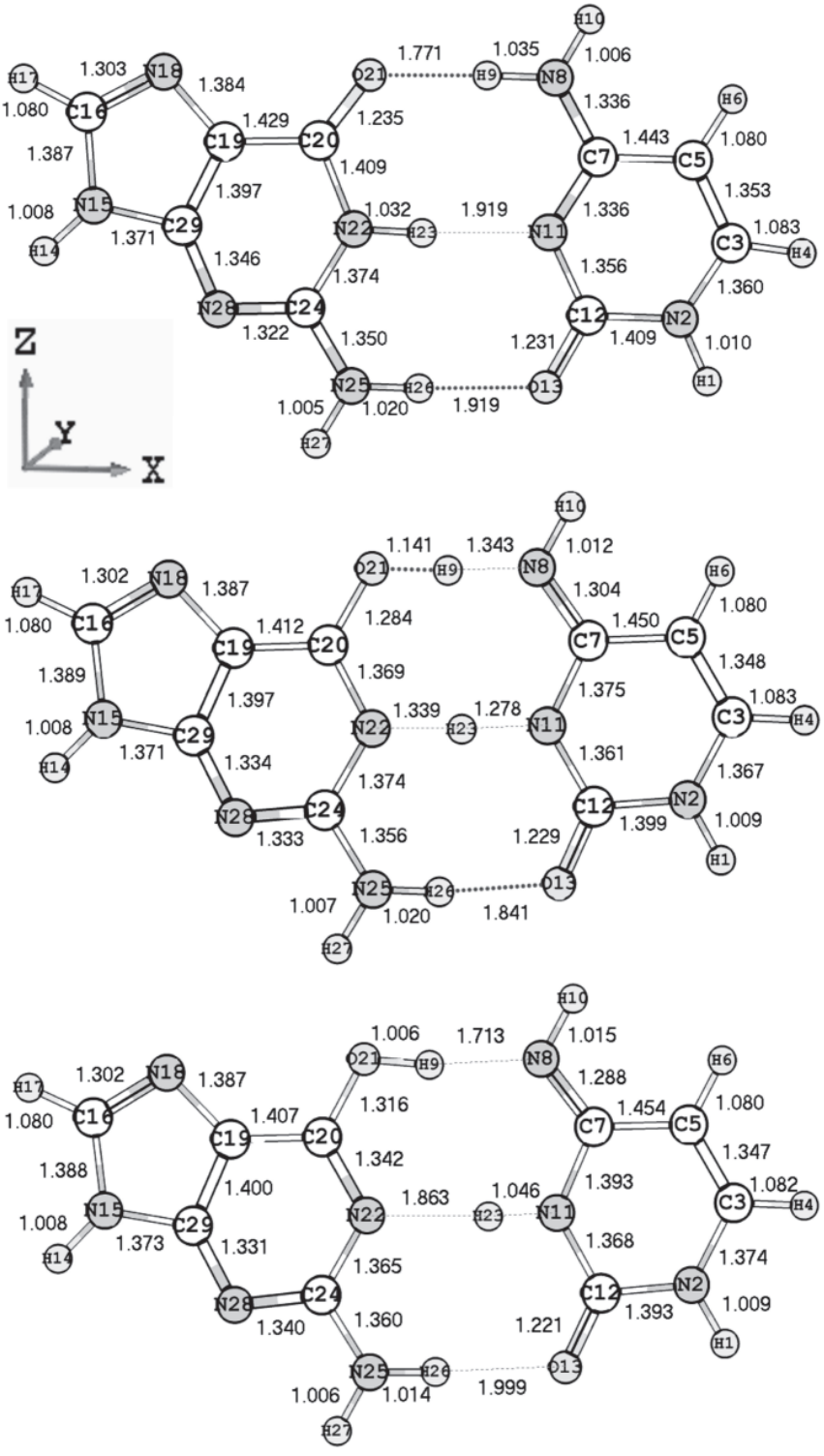


Figure 11.1: Ball-and-stick representation of the hydrogen-bonded GC in the absence of external fields. Reactants (top), transition state (middle) and products (bottom). The numbering scheme, bond lengths (Å), and the coordinate axes are also displayed.

Table 11.1: Reciprocal tunneling ionization rates (characteristic ionization times), in s, for the two forms of the GC base pair (GC and G*C*) in the presence of different strengths of electric fields applied in the x-direction.

| E | (V m^{-1})	5.14x10^{8}	1.29x10^{9}	2.57x10^{9}	3.60x10^{9}	3.86x10^{9}	5.14x10^{9}
| E | (au)	1.00x10^{-3}	2.50x10^{-3}	5.00x10^{-3}	7.00x10^{-3}	7.50x10^{-3}	1.00x10^{-2}
GC	1.07x10^{95}	4.90x10^{26}	1.20x10^{4}	4.77x10^{-3}	4.14x10^{-4}	8.38x10^{-8}
G*C*	7.26x10^{95}	1.05x10^{27}	1.74x10^{4}	6.20x10^{-3}	5.29x10^{-4}	1.00x10^{-7}

The calculated vertical IP of the GC-dimer is 0.2695 au (7.33 eV) and that of the G*C*-dimer is 0.2708 au (7.37 eV).

proton transfer in GC is in the 100 fs timescale [262]. We conclude that no significant population of an ensemble of base-pairs would ionize by tunnel ionization up to a threshold of approximately 38.57x10^{8} V m^{-1} (0.0075 au).

11.4 Computational Details

11.4.1 Electronic Structure Calculations and External Fields

All calculations were performed using the B3LYP functional [46, 34] with a 6-311++G(d,p) basis set. No solvation has been included in the calculations, which may be justifiable since the processes of replication and transcription of DNA occur either in the complete absence of water in the immediate (enzyme/protein) surrounding or in the presence of a partial hydration [264, 265].

The Gaussian 09 package [118] was used in all calculations. A "very tight" geometry optimization threshold has been imposed with a maximum (gradient) force on the nuclei of 2.0x10^{-6} hartrees/bohrs, a maximum root-mean-square (RMS) force of 2.0x10^{-6} hartrees/bohrs, a maximum displacement is 4.0x10^{-6} hartrees/bohrs, and a RMS displacement of 1.0x10^{-6} bohr. An ultrafine grid has been used throughout to maximize the accuracy of the subsequent harmonic frequency analysis. The type of the stationary point is confirmed by the frequency analysis where the presence of a

single imaginary frequency indicates a transition state (TS) structure.

Fields perpendicular to the motion of double proton transfer (in-plane and/or out-of-plane) have been shown to have significantly less influence on the kinetics of the double proton transfer in the formic acid dimer [231]. In view of the similarity of DNA dimers to the formic acid dimer (planar, double proton transfer across two hydrogen bonds), it may be reasonably assumed that this conclusion is also valid in the case of the DNA WC dimers. Hence, only fields parallel and antiparallel to the long axis of the dimers (±x) are considered in this study. Finally, the field strengths applied in the electronic structure calculations are (x10^8 V m^{-1}): 5.14, 12.86, 25.71, 36.00, 38.57 and 51.42.

11.4.2 The Reactions and their Potential Energy Surfaces

In the AT base pair, there is only one possible double proton transfer reaction between the purine (A) and the pyrimidine (T) as there are only two hydrogen bonds. In the GC base pair, however, there are three hydrogen bonds and, therefore, there are three possible double proton transfers reactions (H4/H8, H8/H9, and H4/H9). Three relaxed scans were performed by moving one of the three protons at a time. The most stable tautomer G*C* is that where H4 and H8 undergo the double proton transfer reaction. This result is consistent with other studies [262, 263]. Similar potential energy scans repeated in the presence of electric fields along led to similar conclusions, that the most stable G*C* is that where the double proton transfer involves H4 and H8, and this is the only form of G*C* that will be considered in this thesis.

Initial guesses of the transition state structures were obtained from the potential energy scans in the presence and absence of electric fields. These initial guesses were each subjected to a transition state optimization with the single constraint of the presence of one and only one imaginary frequency. The reaction path was then de-

termined by descending from the position of the TS on the potential energy surface (PES) toward the reactants and products by following the negative gradient, that is, the line of steepest descent. All traced reaction paths have been found to connect the correct reactant, TS, and products.

Unless mentioned otherwise, the energies (or energy differences) reported in this chapter are Gibbs energies obtained from the electronic structure calculation, and the imaginary frequencies are the raw unscaled harmonic frequencies.

11.4.3 Rate and Equilibrium Constants and Proton Tunneling Correction

The rate constant of the double proton transfer is estimated from conventional transition state theory (CTST) [266]:

$$k(T) = \kappa(T)\frac{k_B T}{h} exp\left(- \frac{\Delta G^{\ddagger}}{RT} \right) \tag{11.4.1}$$

where $\kappa(T)$ is a temperature-dependent (Wigner) tunneling correction factor, k_B Boltzmann constant, T the absolute temperature, h Planck's constant, $\Delta G^{\ddagger}$ Gibbs energy of activation, and R the universal gas constant. The tunneling correction used in this thesis is given by [226]:

$$\kappa(T) = 1 + \frac{1}{24}\left[\frac{hIm(\nu^{\ddagger})}{k_B T} \right] \tag{11.4.2}$$

where $\mathrm{Im}(\nu^{\ddagger})$ is the imaginary frequency.

The equilibrium constants were calculated by taking the ratio of the tunneling-corrected rate constant of the forward reaction over the tunneling-corrected rate constant of the reverse reaction ($k_{forward}/k_{reverse}$).

11.5 Conventions

The field strengths of 5.14, 12.86, 25.71, 36.00, 38.57 and 51.42 (in 10^8 V m^{-1}) are denoted in italics as Field 5, 13, 26, 36, 39 and 51, respectively, for simplicity. When the results with field are compared to those without, the comparison includes the reaction under the influence of the strongest field (i.e., field 51). The direction of the field will be reflected in the sign (as in $\pm$E). In some sections, the quantities reported for the forward reaction will be followed by the corresponding quantities for the reverse reaction (in parentheses). Otherwise the direction of the field will be explicitly specified.

11.6 Results and Discussion

11.6.1 Double Proton Transfer in the AT Base Pair

With the adenine-thymine base pair, the conclusions drawn analysing the electronic energies are different from those drawn using Gibbs energies at 25^{o}C or 37^{o}C. Considering the electronic energy, the double proton transfer occurs with a forward barrier height ranging from 13.68 (at 5.14x10^8 V m^{-1}) to 16.56 kcal mol^{-1} (at 5.14x10^9 V m^{-1}) (depending on the field strength), and a reverse barrier height ranging form 0.07 to 1.96 kcal mol^{-1}. The reverse barrier heights are deemed insignificant as they are smaller in magnitude than the average errors in estimating barrier heights by B3LYP. The field-free Gibbs energy surface at 25^{o}C and at 37^{o}C demonstrates that the reverse reaction has a negative energy of activation (i.e., the TS is more stable than the reactants) (refer to Table 11.2 for details).

We were unable to trace the reaction path linking the TS to the product in this case and hence we can only speculate that there is a small dip in the energy profile just before the TS (since the TS is well characterized with one imaginary frequency).

Table 11.2: Electronic and Gibbs energies ($\Delta G^{\ddagger}$) at 25^{o}C and 37^{o}C and imaginary frequencies [Im($\nu^{\ddagger}$)] of the double proton transfer reaction in the adenine-thymine DNA base pair.

Electronic Energy				
| E | (x10^8 V m^{-1})	0.00	5.14	25.71	51.42
E Reactant (au)	-921.7474	-921.7483	-921.7548	-921.7699
E Product (au)	-921.7258	-921.7263	-921.7327	-921.7493
E Transition State (au)	-921.7256	-921.7256	-921.7284	-921.7462
$\Delta G^{\ddagger}$ forward (kcal mol^{-1})	13.68	14.22	16.56	14.89
$\Delta G^{\ddagger}$ reverse (kcal mol^{-1})	0.07	0.42	2.69	1.96
Im ($\nu^{\ddagger}$) (cm^{-1})	372.22i	705.92i	1016.78i	1034.16i
ΔE (kcal mol^{-1})	13.60	13.80	13.86	12.93
Gibbs Energy at 25^{o}C				
Field Strength (x10^8 V m^{-1})	0.00	5.14	25.71	51.42
E Reactant (au)	-921.5665	-921.5672	-921.5737	-921.5897
E Product (au)	-921.5451	-921.5450	-921.5505	-921.5677
E Transition State (au)	-921.5468	-921.5473	-921.5516	-921.5688
$\Delta G^{\ddagger}$ forward (kcal mol^{-1})	12.37	12.50	13.87	13.10
$\Delta G^{\ddagger}$ reverse (kcal mol^{-1})	-1.03	-1.47	-0.68	-0.68
Im ($\nu^{\ddagger}$) (cm^{-1})	372.22i	705.92i	1016.78i	1034.16i
ΔE (kcal mol^{-1})	13.40	13.98	14.54	13.78
Gibbs Energy at 37^{o}C				
Field Strength (x10^8 V m^{-1})	0.00	5.14	25.71	51.42
E Reactant (au)	-921.4127	-921.4134	-921.4198	-921.4365
E Product (au)	-921.3914	-921.3906	-921.3954	-921.4131
E Transition State (au)	-921.3944	-921.3955	-921.4011	-921.4178
$\Delta G^{\ddagger}$ forward (kcal mol^{-1})	11.44	11.21	11.74	11.78
$\Delta G^{\ddagger}$ reverse (kcal mol^{-1})	-1.92	-3.08	-3.58	-2.95
Im ($\nu^{\ddagger}$) (cm^{-1})	372.22i	705.92i	1016.78i	1034.16i
ΔE (kcal mol^{-1})	13.36	14.29	15.32	14.73

The calculated vertical IP of the GC-dimer is 0.2695 au (7.33 eV) and that of the G*C*-dimer is 0.2708 au (7.37 eV).

Similar observation of a negative reverse barrier height (when considering Gibbs energies) has been already observed by Gorb *et al.* [191] and Ceron-Carrasco *et al.* [204]. These observations are generalized here to the same reaction in the presence of electric fields in the range of 10^8 - 10^9 V m^{-1}. The AT to A*T* will, hence, not be considered further in this thesis and will be the subject of a closer inspection in a future study.

11.6.2 Double Proton Transfer in the GC Base Pair

In the absence and presence of all studied field strengths in either direction (±x), the PES exhibits a single maximum along the reaction path, exhibiting different curvatures (including close to complete flatness) as a function of the field strength and orientation (Fig. 11.2).

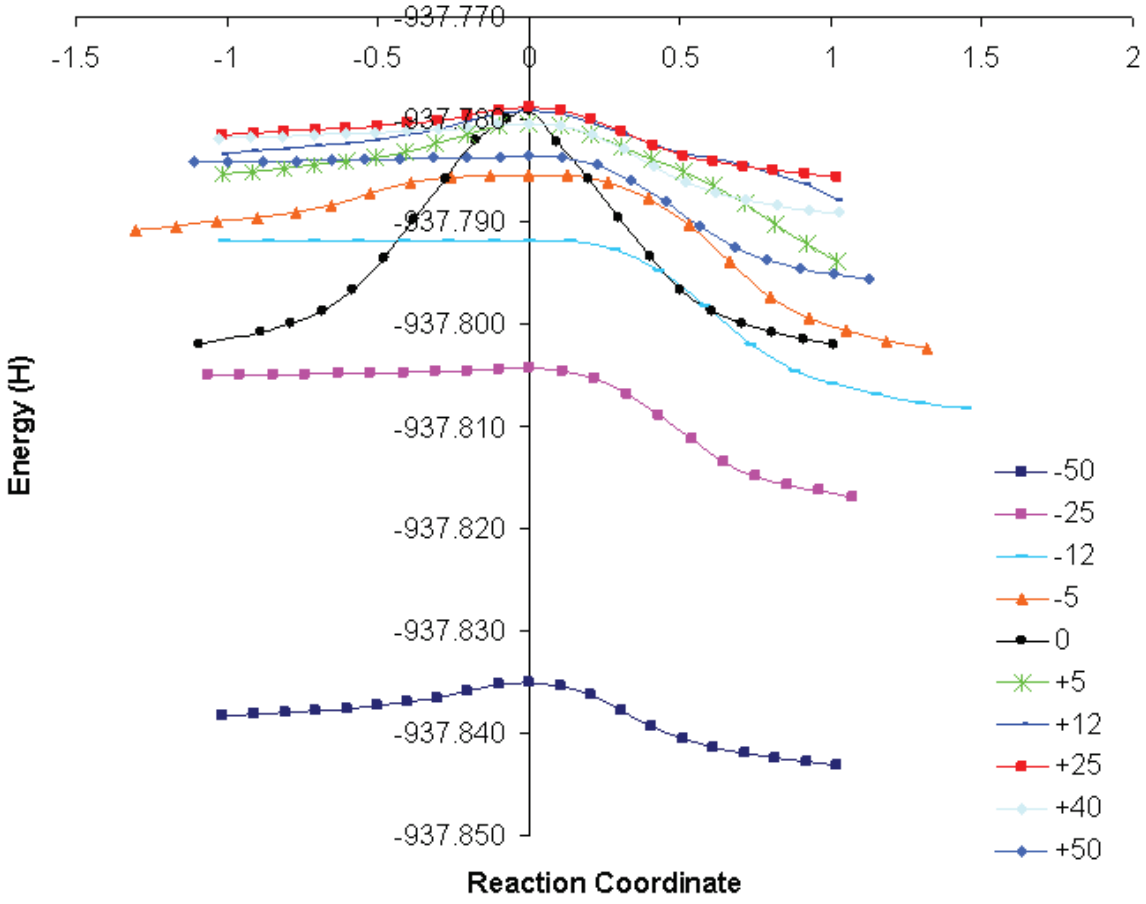


Figure 11.2: Potential energy curves (electronic energies) of the double proton transfer under different field strengths applied in the +x and -x direction.

Since PESs exhibit no intermediates, the double proton transfer is a synchronous one-step reaction irrespective of the field strength or direction, within the approximations of this work.

11.6.3 Field-Effects on the Geometry of the GC Complex

In the absence of fields, the inter-monomer N-N internuclear distance is 2.95 Å. As the field-free reaction progresses, the guanine and cytosine monomers get closer together, then the six-membered rings of each monomer are elongated along the x-axis. This elongation accompanies the shortening of the N-N distance to 2.62 Å in the TS structure. As the reaction proceeds, the N22-H23 bond stretches faster than the N8-H9 bond (Figures 11.1 and 11.3 and Table 11.3). In the products, the N-N bridge lengthens back, close to its original separation, to 2.91 Å. The O21-N8 bridge also shrinks in the TS structure. The TS is propeller-twisted by $\sim 11^o$, whether gauged by the dihedral angle d1(N11-N22-N25-O13) = -11^o or d2(N11-N22-O21-N8) = -11^o. The products exhibit a slight propeller twist of -5^o (d1) and -4^o (d2). The reactant dimer is perfectly planar.

Table 11.3: Selected distances that vary along the double proton transfer reaction in the absence of field. The reactant is perfectly planar.

	Reactant	Transition State	Product
N22-N11	2.95	2.62	2.91
O21-N8	2.81	2.48	2.71
N22-H23/N11-H23	1.03/1.92	1.34/1.28	1.86/1.05
N8-H9/O21-H9	1.03/1.77	1.34/1.14	1.71/1.00

In the strongest field in the +x direction, the monomers move very slightly closer to each other, but there is no elongation of the six-membered rings. The N-N distance changes from 3.11 Å (reactants), to 2.65 Å (TS), and to 3.11 Å (products). The transition state structure is considerably more propeller twisted than in the field-free case, the twist is -21^o = d1 = d2. The reactants in this case are also propeller twised, unlike the field-free case, (d1 = 6^o, d2 = 5^o) and the products (d1 = 6^o, d2 = 7^o). In the presence of the strongest field in the -x direction, the movement of the monomers with respect to each other is less significant than in the field-free case. The N-N dis-

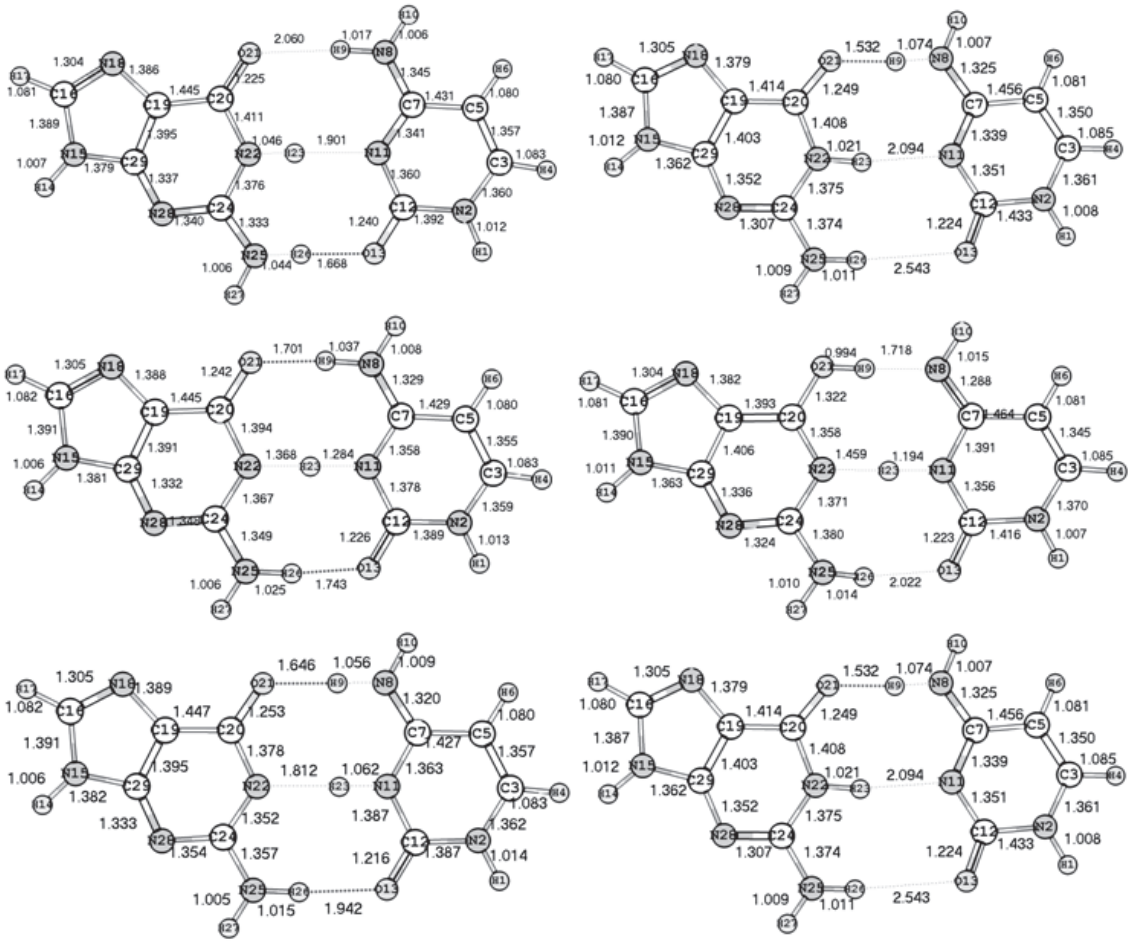


Figure 11.3: Ball-and-stick representation of the hydrogen-bonded GC in the presence of the strongest field (5.14x10^9 V m^{-1}) applied in the -x direction (left) and the +x direction (right) for the reactants (top) transition states (middle) and products (bottom). The numbering scheme and optimized bond lengths (Å) are also displayed.

tance varies from 2.95 Å (reactants), to 2.65 Å (TS), and back to 2.87 Å (products). The reactants are planar, while products have a slight propeller twist, d1 = 3^o and d2 = 2^o. The transition state structure exhibits a higher propeller twist of d1 = 11^o and d2 =13^o. The general observation is that the GC dimer twists as a function of the double proton transfer reaction coordinate with a maximal twist near the TS region.

11.6.4 Effects of Electric Fields on the Barrier Height and Reaction Energy of the Double Proton Transfer Reaction: GC to G*C*

Fields applied in the positive x-direction

Table 11.4 shows that for the forward reaction bathed in fields pointing to the positive x-direction, the (Gibbs) activation energy barrier first rises from its field-free value of 12.1 kcal mol^{-1} under the influence of the weakest field (5.14x10^{8} V m^{-1}) to 13.4 kcal mol^{-1}, and then the barrier generally decreases with the field strength reaching its lowest value of 10.3 kcal mol^{-1} under a 5.14x10^{9} V m^{-1} field. The small oscillations of the barrier heights in the range from ∼3 to 4x10^{9} V m^{-1} is below the accuracy of the electronic structure method (4.7 kcal mol^{-1} for barrier heights and 2.0 kcal mol^{-1} for reaction energies) [267] (Table 11.4). The observed behavior of the barrier as a function of field strength is consistent with a net result of two competing effects one dominant at relatively weak field strengths and one dominant at high intensities of the external field.

The reaction in the reverse direction has a much smaller barrier height compared to the barrier in the forward direction at zero and relatively weak fields. At zero field and up to ∼1.3x10^{9} V m^{-1}, the barrier of the reverse reaction remains around 2 kcal mol^{-1} then sharply rise to ∼12 kcal mol^{-1} for field strength of 2.6x10^{9} V m^{-1} and remains fairly constant beyond this threshold, almost equal (within the accuracy of the calculations) to the forward barrier. The calculated Gibbs energies of reaction indicate that the double proton transfer reaction is endergonic for zero- and low-fields up to 2.6x10^{9} V m^{-1}. This is reflected in the very small magnitudes of the calculated K_{eq} which are much smaller than 10^{-8} indicating a dominance of reactants over the products. The Gibbs energies of reaction in fields beyond 2.6x10^{9} V m^{-1} are exergonic (see Fig. 11.5), reflected in K_{eq} values that fall between ∼1 and 10.

Table 11.4: Gibbs energies, transition state imaginary frequencies [Im($\nu^{\ddagger}$)], forward and reverse barrier heights ($\Delta G^{\ddagger}$ forward/reverse), reaction energies (ΔE), Wigner tunneling corrections [K_{298K}], raw and corrected reaction rate constants [k_{298K} forward/ k_{298K} reverse and k_{298K} corrected forward/reverse] in addition to half-lives ($t_{1/2}$ forward/reverse) reactions, equilibrium constants (K_{eq}), concentration of rare tautomer after 1 second of applying the field ([G*T*]), percent tautomerization after one second of applying the field (%G*T*), and Boltzmann distributions (BD) for reactants, products and transition states; in absence and presence of fields of varying strengths in the ±x directions.

Fields Applied in the Positive Direction							
Field Strength (V m^{-1})	0.00	5.14x10^{8}	1.29x10^{9}	2.57x10^{9}	3.60x10^{9}	3.86x10^{9}	5.14x10^{9}
E Reactant (au)	-937.6359	-937.6346	-937.6315	-937.6285	-937.6302	-937.6308	-937.6318
E Product (au)	-937.6190	-937.6168	-937.6141	-937.6294	-937.6302	-937.6306	-937.6338
E Trasition State (au)	-937.6166	-937.6132	-937.6112	-937.6104	-937.6115	-937.6120	-937.6154
Im($\nu^{\ddagger}$) (cm^{-1})	1316.767i	1254.74i	1225.61i	1165.25i	1056.07i	1013.79i	612.99i
$\Delta G^{\ddagger}$ forward (kcal mol^{-1})	12.13	13.41	12.76	11.38	11.76	11.82	10.33
$\Delta G^{\ddagger}$ reverse (kcal mol^{-1})	1.52	2.26	1.85	11.92	11.71	11.70	11.58
ΔE (kcal mol^{-1})	10.60	11.16	10.91	-0.54	0.05	0.12	-1.25
κ_{298K} tunnel	2.68	2.53	2.46	2.32	2.08	2.00	1.36
k_{298K} forward (s^{-1})	8.03x10^{3}	9.15x10^{2}	2.74x10^{3}	2.83x10^{4}	1.48x10^{4}	1.35x10^{4}	1.68x10^{5}
k_{298K} forward corrected(s^{-1})	2.15x10^{4}	2.31x10^{3}	6.74x10^{3}	6.56x10^{4}	3.08x10^{4}	2.69x10^{4}	2.29x10^{5}
$t_{1/2}$ forward (s)	3.22x10^{-5}	3.00x10^{-4}	1.03x10^{-4}	1.06x10^{-5}	2.25x10^{-5}	2.58x10^{-5}	3.03x10^{6}
k_{298K} reverse (s^{-1})	4.77x10^{11}	1.38x10^{11}	2.73x10^{11}	1.14x10^{4}	1.61x10^{4}	1.65x10^{4}	2.02x10^{4}
k_{298K} reverse corrected (s^{-1})	1.28x10^{12}	3.48x10^{11}	6.72x10^{11}	2.64x10^{4}	3.36x10^{4}	3.30x10^{4}	2.75x10^{4}
$t_{1/2}$ reverse (s)	5.42x10^{-13}	1.99x10^{-12}	1.03x10^{-12}	2.63x10^{-5}	2.06x10^{-5}	2.10x10^{-5}	2.52x10^{5}
K_{eq}	1.68x10^{-8}	6.65x10^{-9}	1.00x10^{-8}	2.49	9.19x10^{-1}	8.14x10^{-1}	8.31
.[G*T*] after 1 sec	6.85x10^{-9}	6.55x10^{-9}	9.93x10^{-9}	7.13x10^{-1}	4.79x10^{-1}	4.49x10^{-1}	8.93x10^{-1}
%G*T* after 1 sec	0.00	0.00	0.00	71.32	47.88	44.88	89.26
BD Reactants	78.90	18.78	0.74	0.03	0.19	0.35	1.01
BD Products	0.00	0.00	0.00	0.86	1.94	3.16	94.04
BD TSs	76.30	2.07	0.24	0.11	0.34	0.56	20.39

Table 11.4 continued...

Fields Applied in the Negative Direction							
Field Strength (V m^{-1})	0.00	5.14x10^{8}	1.29x10^{9}	2.57x10^{9}	3.60x10^{9}	3.86x10^{9}	5.14x10^{9}
E Reactant (au)	-937.6359	-937.6380	-937.6421	-937.6509	-937.6598	-937.6622	-937.6761
E Product (au)	-937.6190	-937.6215	-937.6265	-937.6352	-937.6439	-937.6463	-937.6679
E Trasition State (au)	937.6166	937.6175	937.6228	937.6355	937.6469	937.6500	937.6665
Im($\nu^{\ddagger}$) (cm^{-1})	1316.76i	406.62i	306.53i	800.79i	1052.31i	1096.73i	1246.72i
$\Delta G^{\ddagger}$ forward (kcal mol^{-1})	12.13	12.86	12.07	9.65	8.04	7.67	6.01
$\Delta G^{\ddagger}$ reverse (kcal mol^{-1})	1.52	2.52	2.30	-0.18	-1.94	-2.30	0.82
ΔE (kcal mol^{-1})	10.60	10.34	9.77	9.82	9.98	9.96	5.18
κ_{298K} tunnel	2.68	1.16	1.09	1.62	2.07	2.17	2.51
k_{298K} forward (s^{-1})	8.03x10^{3}	2.31x10^{3}	8.82x10^{3}	5.28x10^{5}	7.94x10^{6}	1.48x10^{7}	2.46x10^{8}
k_{298K} forward corrected(s^{-1})	2.15x10^{4}	2.68x10^{3}	9.62x10^{3}	8.57x10^{5}	1.65x10^{7}	3.22x10^{7}	6.17x10^{8}
$t_{1/2}$ forward (s)	3.22x10^{-5}	2.59x10^{-4}	7.21x10^{-5}	8.09x10^{-7}	4.21x10^{-8}	2.15x10^{-8}	1.12x10^{-9}
k_{298K} reverse (s^{-1})	4.77x10^{11}	8.76x10^{10}	1.27x10^{11}	8.37x10^{12}	1.65x10^{14}	2.99x10^{14}	1.55x10^{12}
k_{298K} reverse corrected (s^{-1})	1.28x10^{12}	1.02x10^{11}	1.39x10^{11}	1.36x10^{13}	3.42x10^{14}	6.48x10^{14}	3.88x10^{12}
$t_{1/2}$ reverse (s)	5.42x10^{-13}	6.82x10^{-12}	4.99x10^{-12}	5.10x10^{-14}	2.03x10^{-15}	1.07x10^{-15}	1.79x10^{-13}
Keq	1.68x10^{-8}	2.63x10^{-8}	6.92x10^{-8}	6.31x10^{-8}	4.82x10^{-8}	4.96x10^{-8}	1.59x10^{-4}
.[G*T*] after 1 sec	6.85x10^{-9}	1.63x10^{-8}	5.92x10^{-8}	5.31x10^{-8}	3.82x10^{-8}	3.96x10^{-8}	1.59x10^{-4}
%G*T* after 1 sec	0.00	0.00	0.00	0.00	0.00	0.00	0.02
BD Reactants	0.00	0.00	0.00	0.00	0.00	0.00	100.00
BD Products	0.00	0.00	0.00	0.00	0.00	0.00	100.00
BD TSs	0.00	0.00	0.00	0.00	0.00	0.00	100.00

Fields applied in the negative x-direction

With fields applied in the negative direction (Fig. 11.4, Table 11.4), the barrier of the forward reaction rises slightly first (from the field-free value of 12.1 to 12.9 kcal mol^{-1} under the weakest field) then continues to drop to reach its lowest value of 6.0 kcal mol^{-1} under the strongest field, 50 % from of its field-free magnitude.

One concludes that strong fields (in excess of $\sim 10^{9}$ V m^{-1}) applied in the negative x-direction facilitate the classical (over-the-barrier) reaction.

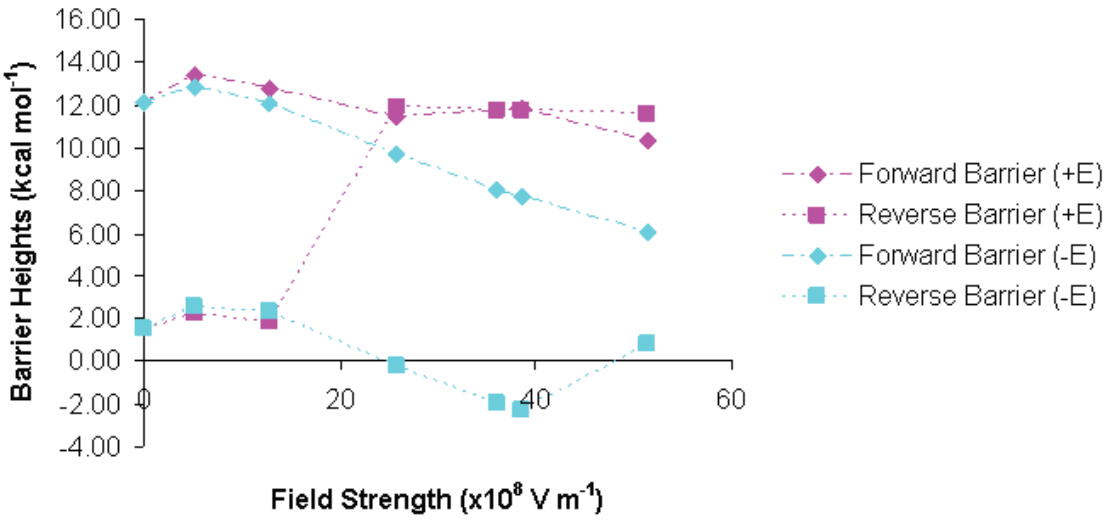


Figure 11.4: Barrier heights (in kcal mol^{-1}) as a function of the field strength (in absolute values). Barrier heights when fields are applied in the +x and -x directions are represented in pink and blue colors, respectively. The forward barrier heights are represented by the diamond shape and the reverse barrier heights are represented with squares. The dotted lines were added to guide the eye in linking the data points.

The field free barrier in the reverse reaction is low (1.5 kcal mol^{-1}) and only slightly affected by external fields for all studied field strengths. The barrier of the reverse reaction increases to 2.5 kcal mol^{-1} under the weakest field and vanishs at $\sim$3x10^9 V m^{-1} before reaching a minimum (of -2.3 kcal mol^{-1}) at 4x10^9 V m^{-1} and then rise again slightly above zero (0.82 kcal mol^{-1}) at the strongest field. The barrier heights are all within 2 kcal mol^{-1} from zero, the small fluctuations probably of little physical significance (the known uncertainty of the electronic structure method is 4.7 kcal mol^{-1} for barrier height and 2.0 kcal mol^{-1} for reaction energy) [267]. It is, hence, difficult to draw a clear conclusion in this case regarding the barrier of the reverse reaction with the fields applied in the negative x-direction.

The energy of (forward) reaction is $\sim$10 kcal mol^{-1} in all cases (except for the strongest field). In the latter case (E = -5.14x10^9 V m^{-1}), the energy of reaction drops abruptly to 5.2 kcal mol^{-1}. With stronger fields applied in the negative x-direction, the energies of reactants, TS species, and products always decrease (become more and more stable), as can be seen from Table 11.4 and Fig. 11.5.

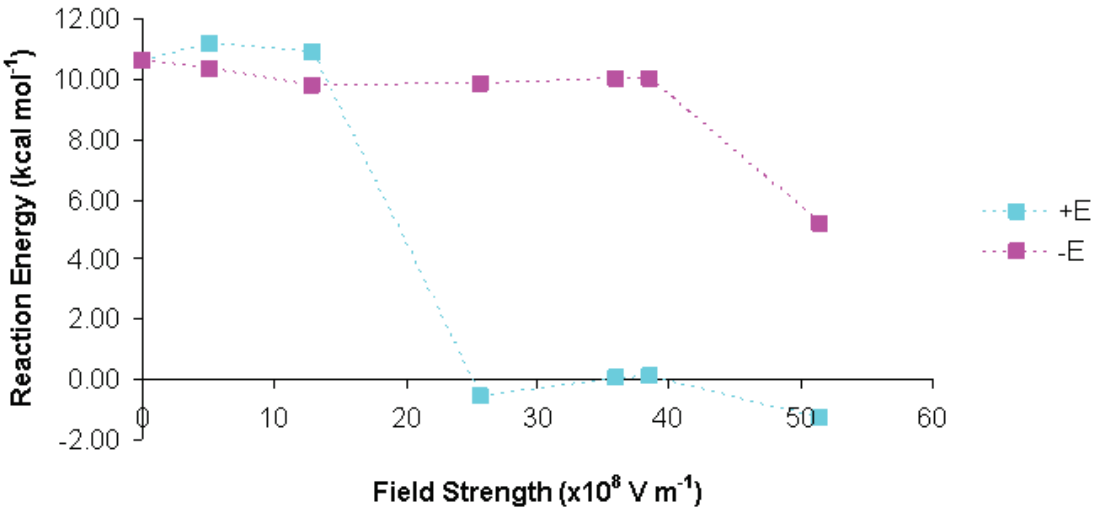


Figure 11.5: Reaction energies (kcal mol^{-1}) as a function of the field strength (in absolute values) applied in the +x and -x directions. The dotted lines are added to guide the eye in linking the data points.

This suggests that stronger fields applied in the -x-direction stabilize the base pair regardless of its position on the reaction path.

11.6.5 Effects of Fields on the Imaginary Vibrational Frequencies

As the strength of field applied in the positive x-direction increases, the magnitude of the (single) imaginary vibrational frequency at the TS structure decreases. The magnitude of this frequency decreases consistently from $1316.8i$, its zero-field value, to $612.9i$ cm^{-1} at the strongest field. This decrease in the magnitude of the imaginary frequency indicates a flattening of the potential energy surface near the transition state as can be gleaned from Fig. 11.2. This field-induced reduction of the magnitude of the imaginary frequency indicates a flattening of the PES, i.e., a thicker potential energy barrier, and hence a decrease in the quantum mechanical tunneling probability of protons (treated as quantum, non-classical particles). The Wigner correction factor κ(T) at 25^{o}C decreases from 2.68 (field-free case) to only 1.3 (5.14×10^9 V m^{-1}).

There is a small rise in the forward barrier from the field-free value of 12.1 kcal mol^{-1} to 13.4 kcal mol^{-1} for the weakest field, which then decrease (with small fluctuations) to reach 10.3 kcal mol^{-1} for the strongest field. At strong fields, thus, there is a decrease in the barrier, and hence a classical acceleration of the reaction but there is also a concomitant decrease in the magnitude of the quantum mechanical tunneling probability. The actual rate constant will, thus, be the result of these two competing tendencies, vide infra. This classical-quantum field effect has been recently described in the context of the double proton transfer reaction in the formic acid dimer [231]. When the field direction is reversed, i.e., oriented in the negative x-direction, the magnitude of the imaginary frequency decreases substantially first with weaker fields 5.14x10^8 - 1.29x10^9 V m^{-1} (406.6i, 306.5i, see Table 11.4 and Fig. 11.6) then rises again to react 1246.7i at the strongest studied field of 5.14x10^9 V m^{-1}. Because of this behavior of the imaginary frequency, the Wigner correction starts at 2.68 for the field-free case, reaches a minimum of 1.09 at 1.29x10^9 V m^{-1} to regain almost its free-field value at the strongest field (2.51). The substantial decrease in the activation energy barrier dominates $k_{forward}$ in this case (due to the exponential dependence of k on $\Delta G^{\ddagger}$) (Table 11.4).

The imaginary frequency for the transition state at field strength of 3.60x10^9 V m^{-1} is nearly the same regardless of the direction of the field applied. It is 1057i and 1052i cm^{-1} for 3.60x10^9 V m^{-1} and -3.60x10^9 V m^{-1}, respectively.

11.6.6 Field-Effects on the Rate Constant of the Double Proton Transfer

The rate constants were calculated using CTST. The rates constants of the first order double proton transfer reaction were calculated at 25^{o}C.

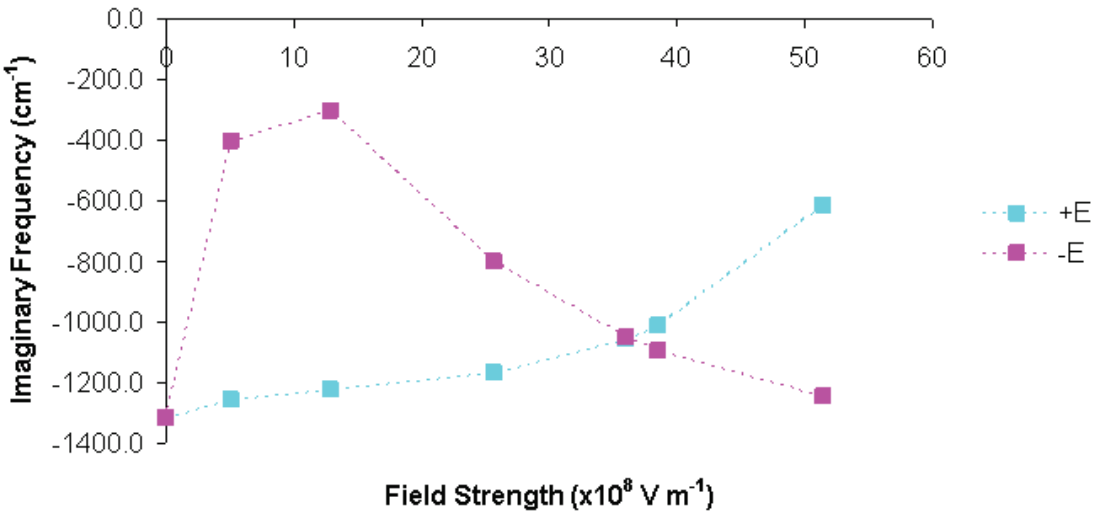


Figure 11.6: Imaginary frequencies (cm^{-1}) as a function of the field strength (in absolute values) applied in the +x and -x directions. The dotted lines are added to guide the eye in linking the data points.

Field applied in the positive x-direction

The tunneling correction decreases from 2.68 (field-free case) to 1.36 (strongest field, 5.14x10^9 V m^{-1}). Compared to the field-free case (k = 2.15x10^4 s^{-1}), the rate constant of the forward reaction with fields 5.14x10^8 V m^{-1} to 1.29x10^9 V m^{-1} decreases by an order of magnitude ($\sim 10^3$ s^{-1}). With fields of 2 .57x10^9 V m^{-1} to 3.86x10^9 V m^{-1}, the rate constant ($\sim 10^4$ s^{-1}) is similar to the field-free case. This rate constant increases by an order of magnitude (2.29x10^5 s^{-1}) under the strongest field.

For the reverse reaction, the corrected rate constant (1.28x10^{12} s^{-1}) is eight orders of magnitudes larger than of the forward reaction (2.15x10^4 s^{-1}) (refer to Fig. 11.7 in this section).

At fields 5.14x10^8 V m^{-1} to 1.29x10^9 V m^{-1}, k decreases by one order of magnitude ($\sim 10^{11}$ s^{-1}). Since, up to and including 1.29x10^9 V m^{-1} fields, the rates constants for the forward reaction are much smaller than those of the reverse reaction, this indicates that even if the rare tautomers form; the WC-base pair is recovered at a much faster

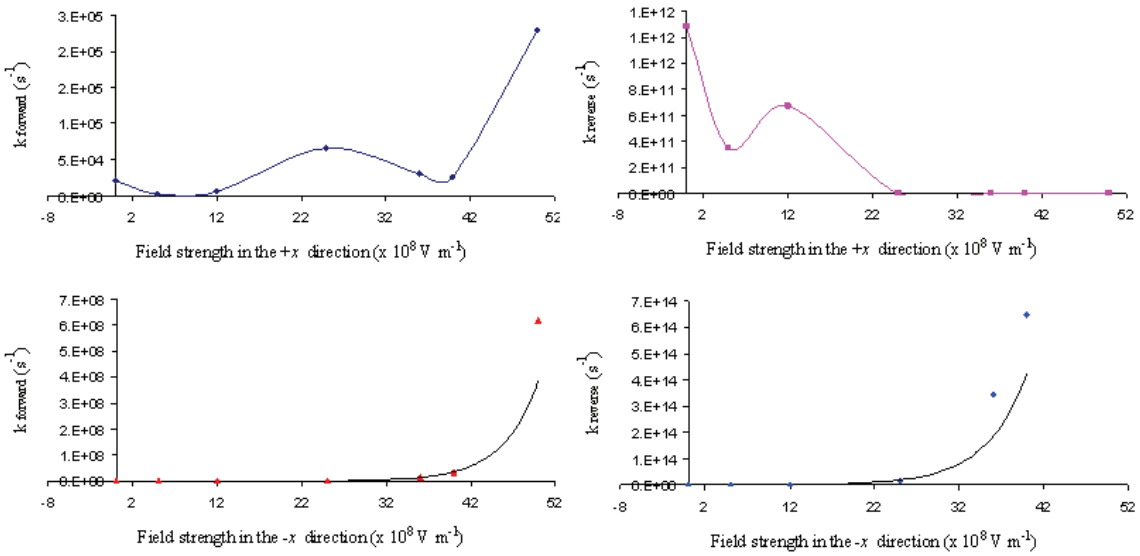


Figure 11.7: Tunneling-corrected rate constants at 25°C for the forward reactions (left) and the reverse reactions (right) in the presence of fields applied in the +x direction (top) and -x direction (bottom). The lines in the top plots are only to help guiding the eye while the lines in the bottom plots are best fit exponential functions.

rate provided that the reverse reaction is thermodynamically allowed. Furthermore, at fields up to and including 1.29x10^9 V m^{-1}, the reverse reaction is thermodynamically favored as the reactants are more stable than the products by ~11 kcal mol^{-1}. Second, the equilibrium constant K_{eq}, defined as the ratio of rates $k_{forward}/k_{reverse}$, is much smaller than one, it is ~10^{-8} - 10^{-9}. A Boltzmann distribution shows that reactants are more populated at weak fields: 79% of the population of reactants is in the field-free case,18% are populated at the weakest field, the other 3% are distributed at the stronger fields. At fields of 2.57x10^9 V m^{-1} and above, $k_{reverse}$ drops to ~10^4 s^{-1}, which is of the same order of magnitude as $k_{forward}$. More specifically, at fields 12.57 - 3.86x10^9 V m^{-1}, the forward and reverse rates are almost equal, which is a result of having equal forward and reverse barrier heights. The following three factors: (1) equal rates for forward and reverse reactions, (2) energies of reaction almost equal to zero, and (3) $0.8 < K_{eq} < 2.5$; suggest that, at this range of field strengths, the reactants and products are almost freely interchangeable at similar rates.

It is only at the strongest field that the rate of tautomerization is faster in the forward

direction (2.29x10^5 s^{-1}) compared to the reverse direction (2.75x10^4 s^{-1}), with a K_{eq} of 8.31 (i.e., well above one); and the tautomerization is exergonic favoring the products by 1.25 kcal mol^{-1} over the reactants. The calculated equilibrium constants (at 25^{o}C) suggest that rare tautomers form only at the strongest field applied in the +x direction. In other words, DNA base pairs undergo a double proton transfer spontaneously only in very strong fields (above 5.14x10^9 V m^{-1}). The percentage of reactants converted to products after a certain time (in the presence or absence of field) is given by the following equation [189]:

$$G^*T^*(t) = \frac{k_f[GT]_i - k_r[G^*T^*]_i}{k_f + k_r}\left\{1 - exp\big[-(k_r + k_f)t\big]\right\} \tag{11.6.1}$$

Where G*T* (t) is the concentration of the rare tautomer at time t, k_f and k_r are the forward and reverse rate constants, [GT]$_i$ is the initial concentration of the reactant, [G*T*]$_i$ is the initial concentration of the rare tautomer. Starting with 1.0 M of the reactants and 1.0 nM of rare tautomer, after 1 sec, the concentration of the rare tautomer remains at the nanomolar scale for fields +0 to +12.57x10^8 V m^{-1}. At fields stronger than +12.57 V m^{-1}, if applied for 1 sec, the concentration of the rare tautomers ranges from 0.48 to 0.89 M, i.e., 48 to 89% of the molecules will convert to the rare tautomers.

Fields applied in the negative x-direction

When fields are applied in the negative direction, the tunneling correction factor ranges approximately from 1 to 3. Initially the weakest field applied (-5) decelerates the forward reaction by an order of magnitude (10^3 s^{-1}) compared to the field-free case (10^4 s^{-1}). The reaction rate then increases steadily by one order of magnitude for each increment of 5 V m^{-1} until it reaches 6.0x10^8 s^{-1} at the strongest field. This is again reflected by the proportional decrease in the half-life. The half-life is shorter

by one order of magnitude as the reaction goes faster by one order of magnitude (SeeTable 11.4 and Fig. 11.8).

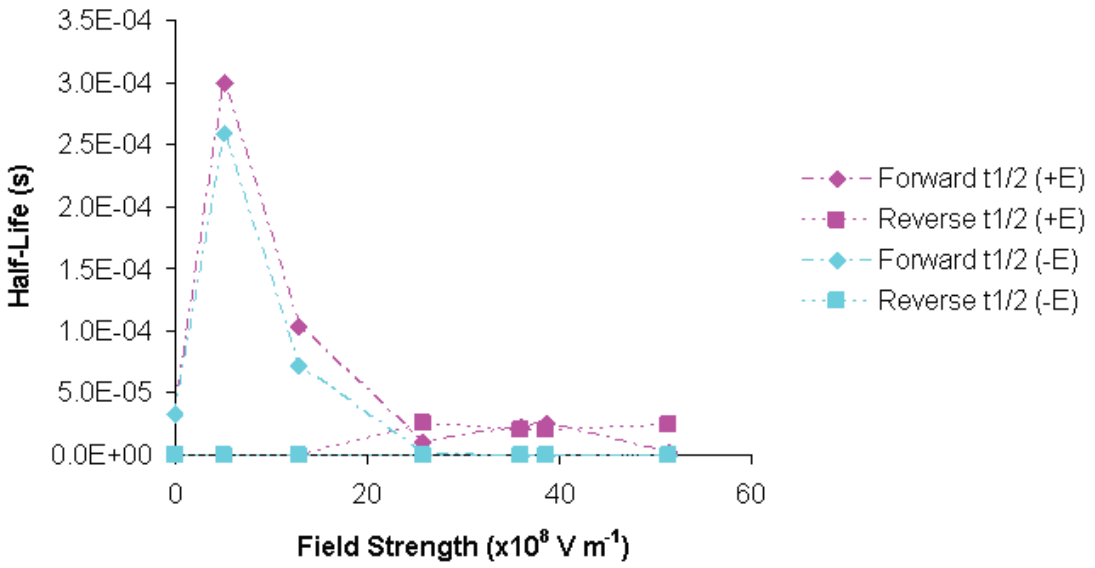


Figure 11.8: Half-lives (s) as a function of the field strength (in absolute values) applied in the +x and -x directions. Half-lives when fields are applied in the +x and -x directions are represented in pink and blue colors, respectively. The forward half-lives are represented by the diamond shape and the reverse half-lives are represented with squares. The dotted lines were added to guide the eye in linking the data points.

Even though the increase is not gradual, the rate constant of the reverse reaction increases with stronger fields in the -x direction to reach a maximum of 2.99×10^{14} s^{-1} at a field strength of -38.6×10^8 V m^{-1}. The only exception is at the strongest field where the rate constant drops by two orders of magnitudes (to 1.55×10^{12} s^{-1}).

For fields in the -x direction, the rates in the reverse direction are much greater than those in the forward direction. In addition, the equilibrium constants are almost zero in all cases, which is strongly favoring the reactants. Also, unlike the case with fields in the $\pm$x direction where, depending on the field applied for a one second, $\sim$50 to 90% of reactants are converted to products (provided that $[GT]_i$ = 1M and $[G^*T^*]_i$ =1 nM), with the field applied in the -x direction, none of the reactant is converted to products after applying the field for one sec (only 0.01% do at the strongest field). Among all fields; reactants, transition states and products have the lowest energy at

the strongest field. This is in agreement with the Boltzmann distribution analysis where the population of each is 100% at the strongest field; and with the K_{eq} which is 2x10^{-4} at the strongest field while $\sim 10^{-8}$ for the rest of the fields. In conclusion, the reactants are more favored than products and they are the most stable at the strongest field.

When the field is applied in the -x direction, the rate constants of the forward reaction fit an exponential function given by:

$$k_{forward}(-\text{x-}direction) \approx 2379 exp(2 \times 10^{-9} \mid \mathbf{E}_{-\text{x}} \mid),$$
$$[n = 7, r^2 = 0.931] \tag{11.6.2}$$

Excluding the data point at the strongest field applied in the -x direction, the rate constants of the reverse reaction fit an exponential function given by:

$$k_{reverse}(-\text{x-}direction) \approx 1 \times 10^{11} exp(2 \times 10^{-9} \mid \mathbf{E}_{-\text{x}} \mid),$$
$$[n = 7, r^2 = 0.802] \tag{11.6.3}$$

In the above equations, the constant multiplying the exponential function has the dimension of the rate constant (unit: s^{-1}) and the constant in the exponent has the dimension of inverse electric field (unit: mV^{-1}). The strong correlation between the data points is shown by the value of r^2 which is close to unity in both cases.

11.7 Conclusions

The double proton transfer in AT is not likely to occur in the absence or even the presence of an electric field. This work has, thus, focused on the GC base pair primarily.

In the presence of fields of any strength or direction, the double proton transfer occurs

synchronously. In all cases, the reactants and products are essentially planar and the TS is propeller twisted by $\sim 11^o$ to 21^o. As the reactants (in the forward reaction) or the products (in the reverse reaction) reach the transition state, the base pairs get closer to each other, as measured by the distance of the N-N bridge.

As the field applied in the -x direction increases, the tunneling-corrected rate constants of the forward and reverse reaction increase exponentially (note that k_{revese} at the strongest field in the positive direction is an outlier as mentioned in the discussion).

The rate constant oscillates as a function of the field strength applied in the positive direction. The general trend, however, is that k_{revese} decreases with increasing field strength while $k_{forward}$ increases with increasing field strengths.

Unlike the case of the formic acid dimer, where two contradictory effects take place simultaneously, with GC, under stronger electric fields applied in the negative direction, the barrier height decreases *and* the potential energy curve gets sharper at the transition state. These complementary effects (in the presence of -| E |) help increase the rate of the double proton transfer.

Chapter 12

Conclusion

This thesis consists of two major parts. The first deals with dispersion interactions in density functional theory. The second deals with effects of external static electric fields on the double proton transfer reaction in hydrogen-bonded DNA base pairs.

The first chapter is a general introduction to DFT. The history of functionals improvement for filling some gaps in DFT is also summarized. The second chapter presents the exchange-hole dipole moment model which is a nonempirical model for capturing dispersion interactions. Dispersion interactions are weak long-range interactions that are not properly described by the conventional DFT methods introduced in Chapter 1.

The first part of this thesis is mainly about the nonempirical van der Waals density functional PW86+PBE+XDM. With this functional the PW86+PBE exchange +correlation energy is computed self-consistently, while XDM is computed non-self-consistently. This is a "post-GGA" approach where dispersion is added as a perturbative correction to the total energy. The ultimate goal of this part is to have an optimizer for vdW complexes using a nonempirical DFT method.

The first study (Chapter 3) is a benchmark study for basis sets. From this study it was concluded that augmented Dunning double or triple zeta basis sets without

counterpoise correction and with a large grid is an optimal combination to get high accuracy with a reasonable computational cost.

The next study (Chapter 4) consists of testing the performance of this functional on a wide variety of 65 complexes (including hydrogen bonding, dispersion and mixed complexes). The test set includes complexes with binding energies ranging from 0.022 (helium dimer) to 22.65 (uracil dimer) kcal mol^{-1}. The mean absolute percent error for the binding energies of complexes in this database is 11.5 and 16.6% with aug-cc-pVDZ and aug-cc-pVTZ, respectively. This study confirmed the highly accurate performance of the functional on molecules at equilibrium geometry.

Being interested in an optimizer, it is necessary to assess the accuracy of binding energies at distorted geometries. For this purpose two studies were performed and presented in Chapters 5 and 6. In Chapter 5, a database of 22 vdW complexes at five different intermonomer separations was considered. The conclusion from this study was that compressed complexes are the most challenging ones to accurately describe. Nevertheless, PW86+PBE+XDM gives a mean absolute relative error of 9.4% on the binding energies of the entire set, and 13.3% on the binding energies of the compressed complexes. This study showed that this functional is the only density functional (amongst many others) that can accurately describe compressed geometries.

In Chapter 6, a larger database of 66 vdW complexes at eight different separations, S66x8, was considered. This database is more balanced in terms of the number of complexes in each subcategory (electrostatic, dispersion and mixed). S66x5 includes some new types of dispersion-bound complexes such as aliphatic-aliphatic and aliphatic-aromatic interactions. In addition a systematic method/basis set was used for the optimization and energy evaluation of all complexes in S66x8. With this database, the compressed complexes were also challenging, but the overall performance of the method on this database was very good, RMSPE is 14.6%. The MARE is 9.1% which is slightly smaller than that of the S22x5 database.

The following chapter (Chapter 7) deals with the accuracy of the computed XDM forces. These forces are evaluated by taking the derivative of the energy with respect to infinitesimally small changes in the nuclear coordinates. In this chapter, the distance at which the energy is minimum was tested to be commensurate with the distance at which the forces vanish to zero. The test was done for the ten rare-gas diatomic systems from the He, Ne, Ar and Kr elements. The results of this test were highly satisfying. In addition, in this chapter, the smoothness of the GGA potential energy surfaces and the force curves were found to be sensitive to the grid used.
The last chapter in this part (Chapter 8) presents the optimizer for vdW complexes with a nonempirical DFT method. This optimizer was built using the "external" keyword in the Gaussian package (G03 or G09). The optimizer was tested to be reproducible on ten rare-gas diatomic systems and on a stacked parallel benzene dimer. This optimizer was then used in two biologically relevant studies that involve DNA nucleotides and tetrahydropyrans/cyclohexanes to study the anomeric affects. From these applications, it was shown that the optimizer accounts properly for the deformation energy of vdW complexes. It improves the binding energies significantly compared to single point calculations on pre-optimized geometries (usually at the MP2 level of theory) using PW86+PBE+XDM. These applications also demonstrated the importance of including XDM when considering dispersion-bound complexes. When XDM is turned off, the errors increase by a factor of 2 - 10. Overall, PW86+PBE+XDM is a very accurate nonempirical DFT method for vdW complexes, it can be used in single point calculations and with an optimizer.

The second part of this thesis deals with the effect of external applied fields on the double proton transfer in DNA base pairs. This reaction in DNA base pairs produces the tautomeric form of the bases, which results in a complementarity mismatch of the DNA base pairs during the replication process as discussed in chapter 9. This phenomenon causes a mutation in the DNA, which can consequently cause aging and

cancer.

As a prelude study, away from the complexity of a non-symmetric system, the formic acid dimer was used as a simple symmetric model of a cyclic double hydrogen-bonding system. In this study (Chapter 10), fields were applied in three different directions with field strengths of 5x10^8 to 50x10^8 V m^{-1}, i.e., similar to fields encountered in a DNA microenvironment. Multiple conclusions were drawn from this study. The mechanism of this reaction occurs synchronously as one and only one transition state structure is found on the potential energy surface. Only the field applied along the movement of the double proton transfer makes a significant difference in frequencies, barrier heights and reaction rates. The strongest field causes a 75% decrease in the vibrational frequency, which results in a flatter potential energy surface around the transition state region. This causes the tunneling correction factor, estimated using Wigner approximation, to decrease, leading to a decrease in the reaction rate as the field strength gets stronger. On the other hand, more intense fields decrease the barrier height causing an acceleration of the reaction rate. As the field applied gets stronger by one order of magnitude from 5x10^8 to 50x10^8 V m^{-1}, the net effect of the two opposing factors is an increase in the reaction rate by a factor of 3 to 4 depending on the method used (B3LYP vs. MP2). The last conclusion was that MP2 and DFT produce identical trends for all properties. Therefore, despite the difference in the absolute values, B3LYP can be as reliable as MP2 for trends in the presence of applied electric fields.

In the next study, the effect of electric fields of the same magnitude is studied in DNA base pairs. In this study, the electric fields were applied only along the axis of the double proton transfer in AT and GC, but in two opposite directions due to the non-symmetry of the systems. Since the reactant is different than the product with DNA base pairs, both forward and reverse reactions were considered. For AT, the conclusion was that the reaction does not take place as the transition state is lower in energy compared to the products. The only way protons can be exchanged

between the base pairs is when the hydrogen atoms act as wave-packets and tunnel from one side to another (i.e., without overcoming a barrier height). In GC, there are three possible double proton transfer reactions as there are three hydrogen bonds between G and C, however, only one of them produces the major tautomers as shown in Chapter 11. Similar to the mechanism of the double proton transfer in FAD, that of GC is synchronous in the absence and presence of all field strengths studied (5x10^8 to 50x10^8 V m^{-1}). As the field applied in the -x direction increases, the tunneling-corrected rate constants of the forward and reverse reaction increase exponentially. With stronger fields applied in the +x-direction, the rate constant oscillates with a general trend: k_{revese} decreases while $k_{forward}$ increases. The rate of the double proton transfer increases as the field strength (applied in the -x-direction) increases. Two complementary effects help increase this DPT rate: lower barrier heights and sharper potential energy curve at the transition state.

From the work done in this thesis, there are plenty of directions for future work. One study of particular interest to conduct in the future is the study of the effect of electric fields on stacked DNA base pairs where dispersion interactions are major. This study can be a big future project as there are many issues to be addressed. First the performance of the optimizer presented in Chapter 8 has to be assessed in the presence of electric fields. In addition, the validity, as well as the accuracy, of doing all types of calculations (e.g., scan, transition state optimization, frequency analysis, computing Gibbs energies instead of electronic energies) has to be tested. For a frequency calculation, a new code has to be prepared and tested for the optimizer to compute not only the first derivatives (forces), but also the second derivatives. These calculations can be done using MP2 for geometries; however, as shown in Chapter 8, optimizing with MP2 and running single point calculations can result in significantly larger errors than optimizing and computing energies with PW86+PBE+XDM. This error is expected to be even more significant in larger systems where more dispersion

interactions are encountered. Second, there are numerous complexes to be included, AT-AT, AT-TA, AT-GC, AT-CG, GC-GC and GC-CG. The number of hydrogen bonds in these systems can vary from four in AT-AT or AT-TA to six in GC-GC or GC-CG. Therefore, there are many possible calculations to be conducted based on, first, the amount (one to four) and, second, the identity of protons to be moved in scan calculations. The number of calculations will also double as the electric fields will be applied in two opposite directions for each complex. This study can be further extended to include three stacked base pairs, which is still within the limit of what the optimizer can handle at a reasonable computational cost using augmented double or triple zeta Dunning basis sets and density fitting.

Another application, currently under study, is also a big project involving the effect of the DNA sequence of two and three stacked base pairs on the twist of the DNA molecule. The other aspect of this study is to investigate whether the twist of the DNA molecule is predominantly from the dispersion interaction of naked stacked base pairs or whether it is a result of the rigid structure of the phosphate backbone. This study can be further extended to investigations of different molecules intercalated between DNA base pairs.

Bibliography

[1] Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry, Introduction to Advanced Electronic Structure Theory*; Dover, New York, 1982.

[2] Levine, I. N. *Quantum Chemistry*, sixth ed.; Pearson/Prentice Hall, 2009.

[3] Thomas, L. H. *Proc. Camb. Phil. Soc.* **1927**, *23*, 542.

[4] Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press, New York, 1989.

[5] Dirac, P. A. M. *Proc. Roy. Soc. Lon. Ser. A* **1928**, *117*, 610.

[6] Dirac, P. A. M. *Proc. Roy. Soc. Lon. Ser. A* **1928**, *118*, 351.

[7] Slater, J. C. *Phys. Rev.* **1951**, *81*, 385.

[8] Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864.

[9] Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133.

[10] Koopmans, T. *Physica* **1934**, *1*, 104.

[11] Ladik, J.; Bogar, F.; Penke, B. *Int. J. Quantum Chem.* **2004**, *98*, 522.

[12] Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*, 2nd ed.; Wiley-VCH, Weinheim, 2001.

[13] Perdew, J. P.; Zunger, A. *Phys. Rev. B* **1981**, *23*, 5048.

[14] Johnson, E. R. Ph.D. thesis, Queen's University, 2007.

[15] Becke, A. D. *J. Chem. Phys.* **1988**, *88*, 1053.

[16] Dirac, P. A. M. *Math. Proc. Camb. Phil. Soc.* **1930**, *26*, 376.

[17] Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.

[18] Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244.

[19] Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, *94*, 7221.

[20] Becke, A. *Density Functional Theory of Electronic Structure, Lecture Notes for Chemistry 6301*, **2007**.

[21] Perdew, J. P. *Physica B* **1991**, *172*, 1.

[22] Laming, G. J.; Termath, V.; Handy, N. C. *J. Chem. Phys.* **1993**, *99*, 8765.

[23] Filatov, M.; Thiel, W. *Mol. Phys.* **1997**, *91*, 847.

[24] Becke, A. D. *J. Chem. Phys.* **1986**, *85*, 7184.

[25] Becke, A. D. *J. Chem. Phys.* **1986**, *84*, 4524.

[26] Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1986**, *33*, 8800.

[27] Lacks, D. J.; Gordon, R. G. *Phys. Rev. A* **1993**, *47*, 4681.

[28] Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.

[29] Zhang, Y. K.; Pan, W.; Yang, W. T. *J. Chem. Phys.* **1997**, *107*, 7921.

[30] Kannemann, F. O.; Becke, A. D. *J. Chem. Theory Comput.* **2009**, *5*, 719.

[31] Murray, E. D.; Lee, K.; Langreth, D. C. *J. Chem. Theory Comput.* **2009**, *5*, 2754.

[32] Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822.

[33] Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671.

[34] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

[35] Colle, R.; Salvetti, O. *Theo. Chim. Acta* **1975**, *37*, 329.

[36] Becke, A. D.; Roussel, M. R. *Phys. Rev. A* **1989**, *39*, 3761.

[37] Proynov, E. I.; Vela, A.; Salahub, D. R. *Chem. Phys. Lett.* **1994**, *230*, 419.

[38] Filatov, M.; Thiel, W. *Phys. Rev. A* **1998**, *57*, 189.

[39] Schmider, H. L.; Becke, A. D. *J. Chem. Phys.* **1998**, *109*, 8188.

[40] Van Voorhis, T.; Scuseria, G. E. *J. Chem. Phys.* **1998**, *109*, 400.

[41] Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *111*, 911.

[42] Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 5029.

[43] Perdew, J. P.; Kurth, S.; Zupan, A.; Blaha, P. *Phys. Rev. Lett.* **1999**, *82*, 2544.

[44] Proynov, E.; Chermette, H.; Salahub, D. R. *J. Chem. Phys.* **2000**, *113*, 10013.

[45] Adamo, C.; Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* **2000**, *112*, 2643.

[46] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

[47] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.

[48] London, F. *Zs. f. Phys.* **1926**, *39*, 322.

[49] London, F. *Trans. Faraday Society* **1937**, *189*, 8.

[50] Johnson, E. R.; Mackie, I. D.; DiLabio, G. A. *J. Phys. Org. Chem.* **2009**, *22*, 1127.

[51] Gerber, I. C.; Angyan, J. G. *J. Chem. Phys.* **2007**, *126*, 044103.

[52] Goll, E.; Leininger, T.; Manby, F. R.; Mitrushchenkov, A.; Werner, H.-J.; Stoll, H. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3353.

[53] Goll, E.; Werner, H.-J.; Stoll, H. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3917.

[54] Angyan, J. G.; Gerber, I. C.; Savin, A.; Toulouse, J. *Phys. Rev. A* **2005**, *72*, 012510.

[55] Kristyan, S.; Pulay, P. *Chem. Phys. Lett.* **1994**, *229*, 175.

[56] Wheatley, R. J.; Meath, W. J. *Mol. Phys.* **1993**, *80*, 25.

[57] Andersson, Y.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **1996**, *76*, 102.

[58] Kamiya, M.; Tsuneda, T.; Hirao, K. *J. Chem. Phys.* **2002**, *117*, 6010.

[59] Lotrich, V. F.; Bartlett, R. J.; Grabowski, I. *Chem. Phys. Lett.* **2005**, *405*, 43.

[60] Sato, T.; Tsuneda, T.; Hirao, K. *Mol. Phys.* **2005**, *103*, 1151.

[61] Sato, T.; Tsuneda, T.; Hirao, K. *J. Chem. Phys.* **2007**, *126*, 234114.

[62] Sato, T.; Nakai, H. *J. Chem. Phys.* **2009**, *131*, 224104.

[63] Tkatchenko, A.; Scheffler, M. *Phys. Rev. Lett.* **2009**, *102*, 073005.

[64] Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2005**, *123*, 154101.

[65] Johnson, E. R.; Becke, A. D. *J. Chem. Phys.* **2006**, *124*, 174104.

[66] Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2007**, *127*, 154108.

[67] Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149.

[68] Wu, Q.; Yang, W. *J. Chem. Phys.* **2002**, *116*, 515.

[69] Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463.

[70] Zimmerli, U.; Parrinello, M.; Koumoutsakos, P. *J. Chem. Phys.* **2004**, *120*, 2693.

[71] Jurecka, P.; Cerny, J.; Hobza, P.; Salahub, D. R. *J. Comput. Chem.* **2007**, *28*, 555.

[72] Lin, I.-C.; Coutinho-Neto, M. D.; Felsenheimer, C.; von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U. *Phys. Rev. B* **2007**, *75*, 205131.

[73] DiLabio, G. A. *Chem. Phys. Lett.* **2008**, *455*, 348.

[74] Kannemann, K. O.; Becke, A. D. *J. Chem. Theory Comput.* **2010**, *6*, 1081.

[75] Grafenstein, J.; Cremer, D. *J. Chem. Phys.* **2009**, *130*, 124105.

[76] Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5291.

[77] Dobson, J. F.; Gould, T. *J. Phys.: Condens. Matter* **2012**, *24*, 073201.

[78] Xu, X.; Goddard, W. A. *Proc. Natl. Acad. Sci.* **2004**, *101*, 2673.

[79] Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364.

[80] Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. C* **2008**, *112*, 4061.

[81] Johnson, E. R. *J. Chem. Phys.* **2011**, *135*, 234109.

[82] von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *Phys. Rev. B* **2005**, *71*, 195119.

[83] von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *Phys. Rev. Lett.* **2004**, *93*, 153004.

[84] Mackie, I. D.; DiLabio, G. A. *J. Phys. Chem. A* **2008**, *112*, 10968.

[85] Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787.

[86] Becke, A. D. *J. Chem. Phys.* **1997**, *107*, 8554.

[87] Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.

[88] Ortmann, F.; Bechstedt, F.; Schmidt, W. G. *Phys. Rev. B* **2006**, *73*, 205101.

[89] Schwabe, T.; Grimme, S. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397.

[90] Wu, X.; Vargas, M. C.; Nayak, S.; Lotrich, V.; Scoles, G. *J. Chem. Phys.* **2001**, *115*, 8748.

[91] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.

[92] Dobson, J. F.; Dinte, B. P. *Phys. Rev. Lett.* **1996**, *76*, 1780.

[93] Dion, M.; Rydberg, H.; Schroder, E.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **2004**, *92*, 246401.

[94] Hult, E.; Andersson, Y.; Lundqvist, B. I.; Langreth, D. C. *Phys. Rev. Lett.* **1996**, *77*, 2029.

[95] Puzder, A.; Dion, M.; Langreth, D. C. *J. Chem. Phys.* **2006**, *124*, 164105.

[96] Rydberg, H.; Dion, M.; Jacobson, N.; Schroder, E.; Hyldgaard, P.; Simak, S. I.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **2003**, *91*, 126402.

[97] Silvestrelli, P. L. *J. Phys. Chem. A* **2009**, *113*, 5224.

[98] Silvestrelli, P. L. *Phys. Rev. Lett.* **2008**, *100*, 053002.

[99] Silvestrelli, P. L.; Benyahia, K.; Grubisic, S.; Ancilotto, F.; Toigo, F. *J. Chem. Phys.* **2009**, *130*, 074702.

[100] Thonhauser, T.; Cooper, V. R.; Li, S.; Puzder, A.; Hyldgaard, P.; Langreth, D. C. *Phys. Rev. B* **2007**, *76*, 125112.

[101] Langreth, D. C.; Lundqvist, B. I.; Chakarova-Kack, S. D.; Cooper, V. R.; Dion, M.; Hyldgaard, P.; Kelkkanen, A.; Kleis, J.; Kong, L.; Li, S.; Moses, P. G.; Murray, E.; Puzder, A.; Rydberg, H.; Schroder, E.; Thonhauser, T. *J. Phys.: Condens. Matter* **2009**, *21*, 084203.

[102] Vydrov, O. A.; Wu, Q.; Van Voorhis, T. *J. Chem. Phys.* **2008**, *129*, 014106.

[103] Vydrov, O. A.; Van Voorhis, T. *J. Chem. Phys.* **2009**, *130*, 104105.

[104] Johnson, E. R.; Becke, A. D. *Chem. Phys. Lett.* **2006**, *432*, 600.

[105] *CRC Handbook of Chemistry and Physics*; Lide, D. R., Ed., 76th ed.; Boca Raton, 1995.

[106] Thakkar, A.; Hettema, H.; Wormer, P. *J. Chem. Phys.* **1992**, *97*, 3252.

[107] Tang, K.; Norbeck, J.; Certain, P. *J. Chem. Phys.* **1976**, *64*, 3063.

[108] Rijks, W.; Wormer, P. *J. Chem. Phys.* **1988**, *88*, 5704.

[109] Hirshfeld, F. L. *Theor. Chem. Acc.* **1977**, *44*, 129.

[110] Jackson, J. *Classical Electrodynamics*, 2nd ed.; Wiley: New York, 1975; p 60–62.

[111] Tabor, D. *Gases, Liquids and Solids*; Penguin: Harmondsworth, 1969; p 238–239.

[112] Brinck, T.; Murray, J.; Politzer, P. *J. Chem. Phys.* **1993**, *98*, 4305.

[113] Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2006**, *124*, 014104.

[114] Hettema, H.; Wormer, P.; Thakkar, A. *Mol. Phys.* **1993**, *80*, 533.

[115] Ahlrichs, R.; Penco, R.; Scoles, G. *Chem. Phys.* **1977**, *19*, 119.

[116] Mooij, W. T. M.; van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Eijck, B. P. *J. Phys. Chem. A* **1999**, *103*, 9872.

[117] Becke, A. *Int. J. Quantum Chem.* **1994**, *52*, 625.

[118] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Jr.,; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, 2009.

[119] Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69.

[120] Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1969**, *51*, 2657.

[121] Ditchfie, A.; Miller, D. P.; Pople, J. A. *J. Chem. Phys.* **1970**, *53*, 613.

[122] Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650.

[123] Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939.

[124] Dunning, Jr., T. H. *J. Chem. Phys.* **1989**, *90*, 1007.

[125] Wilson, A. K.; van Mourik, T.; Dunning, Jr., T. H. *J. Mol. Struct. Theochem* **1996**, *388*, 339.

[126] Wilson, A. K.; Woon, D. E.; Peterson, K. A.; Dunning, Jr., T. H. *J. Chem. Phys.* **1999**, *110*, 7667.

[127] Woon, D. E.; Dunning, Jr., T. H. *J. Chem. Phys.* **1994**, *100*, 2975.

[128] Woon, D. E.; Dunning, Jr., T. H. *J. Chem. Phys.* **1993**, *98*, 1358.

[129] van Mourik, T.; Dunning, Jr., T. H. *Int. J. Quantum Chem.* **2000**, *76*, 205.

[130] Jensen, F. *J. Chem. Phys.* **2001**, *115*, 9113.

[131] Chalasinski, G.; Szczesniak, M. M. *Chem. Rev.* **1994**, *94*, 1723.

[132] Krack, M.; Koster, A. M. *J. Chem. Phys.* **1998**, *108*, 3226.

[133] Tang, K. T.; Toennies, J. P. *J. Chem. Phys.* **2003**, *118*, 4976.

[134] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A. J.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M. W.; Gill, P. M.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian03, Revision B.03*, 2003.

[135] Feller, D. *J. Comput. Chem.* **1996**, *17*, 1571.

[136] Schuchardt, K. L.; Didier, B. T.; Elsethagen, T.; Sun, L.; Gurumoorthi, V.; Chase, J.; Li, J.; Windus, T. L. *J. Chem. Inf. Model.* **2007**, *47*, 1045.

[137] Becke, A. D. *Int. J. Quantum Chem.* **1989**, *36*, 599.

[138] Becke, A. D.; Dickson, R. M. *J. Chem. Phys.* **1990**, *92*, 3610.

[139] Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2005**, *122*, 154104.

[140] Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.

[141] Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 415.

[142] Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 5656.

[143] Schmider, H. L.; Johnson, E. R.; Becke, A. D. unpublished code.

[144] Kong, J.; Gan, Z.; Proynov, E.; Freindorf, M.; Furlani, T. R. *Phys. Rev. A* **2009**, *79*, 042510.

[145] Becke, A. D.; Arabi, A. A.; Kannemann, F. O. *Can. J. Chem.* **2010**, *88*, 1057.

[146] Grafova, L.; Pitonak, M.; Rezac, J.; Hobza, P. *J. Chem. Theo. Comput.* **2010**, *6*, 2365.

[147] Johnson, E. R.; Wolkow, R. A.; DiLabio, G. A. *Chem. Phys. Lett.* **2004**, *394*, 334.

[148] Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.

[149] Molnar, L. F.; He, X.; Wang, B.; Merz, Jr., K. M. *J. Chem. Phys.* **2009**, *131*, 065102.

[150] Rezac, J.; Riley, K. E.; Hobza, P. *J. Chem. Theo. Comput.* **2011**, *7*, 2427.

[151] Sarabadani, J.; Naji, A.; Asgari, R.; Podgornik, R. *Phys. Rev. B* **2011**, *84*, 155407.

[152] Johnson, E. R.; Becke, A. D. *J. Chem. Phys.* **2005**, *123*, 024101.

[153] Hesselmann, A. *J. Chem. Phys.* **2012**, *136*, 014104.

[154] Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095.

[155] Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

[156] Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comput. Chem.* **2011**, *32*, 1456.

[157] Arabi, A. A.; Becke, A. D. *J. Chem. Phys.* **2012**, *137*, 014104.

[158] Ruiz, E.; Salahub, D. R.; Vela, A. *J. Am. Chem. Soc.* **1995**, *117*, 1141.

[159] Marchetti, O.; Werner, H.-J. *J. Phys. Chem. A* **2009**, *113*, 11580.

[160] Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 289.

[161] Johnson, E. R.; Becke, A. D.; Sherrill, C. D.; DiLabio, G. A. *J. Chem. Phys.* **2009**, *131*, 034111.

[162] Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2004**, *108*, 10200.

[163] Weigend, F. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057.

[164] Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608.

[165] Bjornsson, R.; Arnason, I. *Phys. Chem. Chem. Phys.* **2009**, *11*, 8689.

[166] Csonka, G. I.; French, A. D.; Johnson, G. P.; Stortz, C. A. *J. Chem. Theo. Comput.* **2009**, *5*, 679.

[167] Lee Woodcock, H.; Moran, D.; Pastor, R. W.; MacKerell, Jr., A. D.; Brooks, B. R. *Biophys. J.* **2007**, *93*, 1.

[168] Wallevik, S. O.; Bjornsson, R.; Kvaran, A.; Jonsdottir, S.; Arnason, I.; Belyakov, A. V.; Baskakov, A. A.; Hassler, K.; Oberhammer, H. *J. Phys. Chem. A* **2010**, *114*, 2127.

[169] Weldon, A.; Vickrey, T.; Tschumper, G. *J. Phys. Chem. A* **2005**, *109*, 11073.

[170] Nagle, J. F.; Morowitz, H. J. *Proc. Natl. Acad. Sci. USA* **1978**, *75*, 298.

[171] Pomes, R.; Roux, B. *Biophys. J.* **1996**, *71*, 19.

[172] Pomes, R.; Roux, B. *J. Phys. Chem.* **1996**, *100*, 2519.

[173] Decornez, H.; Drukker, K.; Hammes-Schiffer, S. *J. Phys. Chem. A.* **1999**, *103*, 2891.

[174] Decornez, H.; Hammes-Schiffer, S. *Isr. J. Chem.* **1999**, *39*, 397.

[175] Mathies, R. A.; Lin, S.; Ames, J. B.; Pollard, W. T. *Ann. Rev. Biophys. Biophys. Chem.* **1991**, *491*, 20.

[176] Luecke, H.; Richter, H. T.; Lanyi, J. K. *Sience* **1998**, *280*, 1934.

[177] Gai, F.; Hasson, K. C.; McDonald, J. C.; Anfinrud, P. A. *Science* **1998**, *279*, 1886.

[178] Bertrand, P. *J. Biol. Inorg. Chem.* **2004**, *9*, 2.

[179] Shinobu, A.; Agmon, N. *J. Phys. Chem. A* **2009**, *113*, 7253.

[180] Okamura, M. Y.and Feher, G. *Ann. Rev. Biochem.* **1992**, *61*, 861.

[181] Cui, Q.; Karplus, M. *J. Phys. Chem. B.* **2003**, *107*, 1071.

[182] Smedarchina, Z.; Siebrand, W.; Fernandez-Ramos, A.; Cui, Q. *J. Am. Chem. Soc.* **2003**, *125*, 243.

[183] Rastogi, V. K.; Girvin, M. E. *Nature* **1999**, *402*, 263.

[184] Boyer, P. D. *Nature* **1999**, *402*, 247.

[185] Wang, H.; Oster, G. *Nature* **1998**, *396*, 279.

[186] Lowdin, P. O. *Rev. Mod. Phys.* **1963**, *35*, 724.

[187] Lowdin, P. O. *Adv. Quantum Chem.* **1965**, *2*, 213.

[188] Dogonadze, R. R.; Kharkats, Y. I.; Ulstrup, J. **1976**, *37*, 360.

[189] Gorb, L.; Podolyan, Y.; Leszczynski, J.; Siebrand, W.; Fernandez-Ramos, A.; Smedarchina, Z. *Biopolymers* **2002**, *61*, 77.

[190] Podolyan, Y.; Gorb, L.; Leszczynski, J. *Int. J. Mol. Sci.* **2003**, *4*, 410.

[191] Gorb, L.; Podolyan, Y.; Dziekonski, P.; Sokalski, W.; Leszczynski, J. *J. Am. Chem. Soc.* **2004**, *126*, 10119.

[192] Kryachko, E.; Sabin, R. J. *Int. J. Quantum Chem.* **2003**, *91*, 695.

[193] *Quantum indeterminism, mutation, natural selection, and the meaning of life. Quantum Biochemistry: Electronic Structure and Biological Activity.*; Stamos, D. N., Ed.; Wiley-VCH: Weinheim, 2010; p 837–871.

[194] Hermann, D. M.; Hossmann, K. A. *J. Neurol. Sci.* **1997**, *1*, 152.

[195] Verschaeve, L.; Maes, A. *Mutation Res.* **1998**, *410*, 141.

[196] Huuskonen, H.; Lindbohm, M. L.; Juutilainen, J. *Mutation Res.* **1998**, *410*, 167.

[197] Repacholi, M. H. *Bioelectromagnetics* **1998**, *19*, 1.

[198] Portnov, O. G.; Shakarnis, V. F.; Maiore, D. *Genetika* **1975**, *11*, 177.

[199] Sardari, D.; N. Verga, N. *Electromag. Biol. Med.* **2010**, *29*, 26.

[200] Yao, C. G.; Sun, C. X.; Mi, Y., X. L.; Wang, S. B. *IEEE Trans. Plasma Sci.* **2004**, *32*, 1626.

[201] Plotnikov, A.; Niego, B.; Ophir, R.; Korenstein, R.; Keisari, Y. *Prostate* **2006**, *66*, 1620.

[202] Jordan, D. W.; Uhler, M. D.; Gilgenbach, R. M.; Lau, Y. Y. *J. Appl. Phys.* **2006**, *99*, 094701.

[203] Mejia-Mazariegos, L.; Hernandez-Trujillo, J. *Chem. Phys. Lett.* **2009**, *482*, 24.

[204] Ceron-Carrasco, J. P.; Requena, A.; Zuniga, J.; Michaux, C.; Perpete, E. A.; Jacquemin, D. *J. Phys. Chem. A* **2009**, *113*, 10549.

[205] Fu, L.-Y.; Wang, G.-Z.; Ma, B.-G.; Zhang, H.-Y. *Biochem. Biophys. Res. Commun.* **2011**, *409*, 367.

[206] Grimm, R. L.; Beauchamp, J. L. *J. Phys. Chem. B* **2003**, *107*, 14161.

[207] Eow, J. S.; Ghadiri, M. *Chem. Eng. Process.* **2003**, *42*, 259.

[208] Hong, Y.; R., H. J.; Kanga, H. *J. Chem. Phys.* **1998**, *108*, 4367.

[209] Corkum, P. B.; Burnett, N. H.; Brunel, F. *Phys. Rev. Lett.* **1989**, *62*, 1259.

[210] Landau, L. D.; Lifshitz, E. M. *Quantum Mechanics*; Pergamon, New York, 1965.

[211] Bandrauk, A.; Sedik, E.; Matta, C. *J. Chem. Phys.* **2004**, *121*, 7764.

[212] Bandrauk, A. D.; Sedik, E. S.; Matta, C. F. *Mol. Phys.* **2006**, *104*, 95.

[213] Wojcik, M. J.; Hirakawa, A. Y.; Tsuboi, M. *Int. J. Quantum Chem.* **1986**, *30*, 133.

[214] Chojnacki, H.; Andzelm, J.; Nguyen, D. T.; Sokalski, W. A. *Comp. and Chem.* **1995**, *19*, 181.

[215] Chocholousova, J.; Vacek, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2002**, *4*, 2119.

[216] Fillaux, F. *Chem. Phys. Lett.* **2005**, *408*, 302.

[217] Scheiner, S.; Kern, C. W. *J. Am. Chem. Soc.* **1979**, *101*, 4081.

[218] Tsuzuki, S.; Uchimaru, T.; Matsumura, K.; Mikami, M.; Tanabe, K. *J. Chem. Phys.* **1999**, *110*, 11906.

[219] Yavuz, I.; Trindle, C. *J. Chem. Theory Comput.* **2008**, *4*, 533.

[220] Bosi, P.; Zerbi, G.; Clementi, E. *J. Chem. Phys.* **1977**, *66*, 3376.

[221] Fernandez, L. E.; Marigliano, A. C. G.; Varetti, E. L. *Vib. Spectrosc.* **2005**, *37*, 179.

[222] Zielke, P.; Suhm, M. A. *Phys. Chem. Chem. Phys.* **2007**, *9*, 4528.

[223] Ushiyama, H.; Takatsuka, K. *J. Chem. Phys.* **2001**, *115*, 5903.

[224] Longo, R. L.; Freitas, L. C. G. *Int. J. Quantum Chem.* **1990**, *38*, 35.

[225] Moller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

[226] Wigner, E. P. *J. Phys. Chem. B* **1932**, *19*, 203.

[227] *The Theory of Rate Processes*; Glasstone, S.; Laidler, K. J.; Eyring, H., Eds., 1st ed.; McGraw-Hill Book Company, Inc. , New York, 1941.

[228] Skodje, R. T.; Truhlar, D. G. *J. Phys. Chem* **1981**, *85*, 624.

[229] *The Tunnel Effect in Chemistry*; Bell, R. P., Ed.; Chapman and Hall, London, 1980.

[230] Bandrauk, A. D.; Sedik, E. S.; Matta, C. F. *J. Chem. Phys.* **2004**, *121*, 7764.

[231] Arabi, A. A.; Matta, C. F. *Phys. Chem. Chem. Phys.* **2011**, *13*, 13738.

[232] Ivanov, M. Y.; Matusek, D. R.; Wright, J. S. *Chem. Phys. Lett.* **1996**, *225*, 232.

[233] Matusek, D. R.; Ivanov, M. Y.; Wright, J. S. *Chem. Phys. Lett.* **1996**, *258*, 255.

[234] Ivanov, M. Y.; Matusek, D. R.; Wright, J. S. *Phys. Rev. A* **1996**, *54*, 5159.

[235] Shaik, S.; de Visser, S. P.; Kumar, D. *J. Am. Chem. Soc.* **2004**, *126*, 11746.

[236] Markin, V. S.; Liu, D. S.; Gimsa, J.; Strobel, R.; Rosenberg, M. D.; Tsong, T. Y. *J. Membr. Biol.* **1992**, *126*, 137.

[237] Tsong, T. Y. *J. Biol. Phys.* **2002**, *28*, 309.

[238] Jiang, X.; Engelhard, M.; Ataka, K.; Heberle, J. *J. Am. Chem. Soc.* **2010**, *132*, 10808.

[239] Armstrong, F. A.; Wilson, G. S. *Electrochim. Acta* **2000**, *45*, 2623.

[240] Ruzgas, T.; Wong, L.; Gaigalas, A. K.; Vilker, V. L. *Langmuir* **1998**, *14*, 7298.

[241] Friis, E.; Andersen, J.; Kharkats, Y.; Kuznetsov, A.; Nichols, R.; Zhang, J.; Ulstrup, J. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 1379.

[242] Gaigalas, A. K.; Ruzgas, T. *J. Electroanal. Chem.* **1999**, *465*, 96.

[243] Fedurco, M. *Chem. Rev.* **2000**, *209*, 263.

[244] Avila, A.; Gregory, B.; Niki, K.; Cotton, T. *J. Phys. Chem. B* **2000**, *104*, 2759.

[245] Franzen, S.; Goldstein, R. F.; Boxer, S. G. *J. Phys. Chem.* **1990**, *94*, 5135.

[246] Gopher, A.; Blatt, Y.; Schonfeld, M.; Okamura, M. Y.; Feher, G.; Montal, M. *Biophys. J.* **1985**, *48*, 311.

[247] Popovic, Z. D.; Kovacs, G. J.; Vincett, P. S.; Alegria, G.; Dutton, P. L. *J. Chem. Phys.* **1986**, *110*, 227.

[248] Murgida, D. H.; Hildebrandt, P. *J. Phys. Chem. B* **2002**, *106*, 12814.

[249] Moroney, P. M.; Scholes, T. A.; Hinkle, P. C. *Biochem.* **1984**, *23*, 4991.

[250] Gregory, L.; Fergusonmiller, S. *Biochemistry* **1989**, *28*, 2655.

[251] Sarti, P.; Antonini, G.; Malatesta, F.; Brunori, M. *Biochem. J.* **1992**, *248*, 123.

[252] Nicholls, P.; Butko, P. *J. Bioenerg. Biomembr.* **1993**, *25*, 137.

[253] Wackerbarth, H.; Hildebrandt, P. *ChemPhysChem* **2003**, *4*, 714.

[254] Bombarda, E.; Becker, T.; Ullmann, G. M. *J. Am. Chem. Soc.* **2006**, *128*, 12129.

[255] Nagel, G.; Kelety, B.; Mockel, B.; Buldt, G.; Bamberg, E. *Biophys. J.* **1998**, *74*, 403.

[256] Geibel, S.; Friedich, T.; Ormos, P.; Wood, P. G.; Nagel, G.; Bamberg, E. *Biophys. J.* **2001**, *81*, 2059.

[257] Choi, J.; Moon, S. *J. Colloid Interface Sci.* **2003**, *265*, 93.

[258] Simons, R. *Nature* **1979**, *280*, 824.

[259] Kwon, O.-H.; Zewail, A. H. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 8703.

[260] Mi, Y.; Sun, C. X.; Yao, C. G.and Xiong, L.; Wang, S. B.; Li, C. X.; Li, J.; Hu, L. N. *Proc. Ann. Int. Conf. IEEE Engineering in Med. and Biol. Soc.* **2005**, 4904.

[261] Plotnikov, A.; Niego, B.; Ophir, R.; Korenstein, R.; Keisari, Y. *Prostate* **2006**, *66*, 1620.

[262] Zoete, V.; Meuwly, M. *J. Chem. Phys.* **2004**, *121*, 4377.

[263] Liu, H.; Li, G.; Zhang, L.; Li, J.; Wang, M.; Bu, Y. *J. Chem. Phys.* **2011**, *135*, 134315.

[264] Topal, M. D.; Fresco, J. R. *Nature* **1976**, *263*, 285.

[265] Petruska, J.; Sowers, L. C.; Goodman, M. F. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 1559.

[266] Laidler, K. J. *Chemical Kinetics*, 3rd ed.; Cambridge, 1987.

[267] Goerigk, L.; Grimme, S. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670.

Appendix A

Copyright Permissions

Dear Dr. Arabi:

Thank you for requesting permission to reproduce material from American Institute of Physics publications.

Permission is granted - subject to the conditions outlined below - for the following:

Journal of Chemical Physics 137(1), 014104 (2012)

To be used in the following manner:

Reproduced in your thesis for submission to Dalhousie University, Halifax, Nova Scotia, Canada. It is understood that your thesis may be reproduced and distributed by the Library and Archives of Canada.

1. The American Institute of Physics grants you non-exclusive world rights in all languages and media.

2. This permission extends to all subsequent and future editions of the new work.

3. The following copyright notice must appear with the material (please fill in the information indicated by capital letters):
"Reprinted with permission from [FULL CITATION]. Copyright [PUBLICATION YEAR], American Institute of Physics."
Full citation format is as follows: Author names, journal title, Vol., Page, (Year of publication).
For an article, the copyright notice must be printed on the first page of the article or book chapter. For figures, photographs, covers, or tables, the notice may appear

with the material, in a footnote, or in the reference list.

4. This permission does not apply to any materials credited to sources other than the copyright holder.

Please let us know if you have any questions.

Sincerely,
Susann Brailey

Manager, Rights and Permissions
American Institute of Physics
Suite 1NO1
2 Huntington Quadrangle
Melville, NY 11747-4502

Phone: 1-516-576-2268
Fax: 1-516-576-2450
Email: sbrailey@aip.org

Dear Ms. Alya Arabi,

Please review: http://www.nrcresearchpress.com/page/authors

http://www.nrcresearchpress.com/page/authors/information/rights

As an author, you may reuse published material without paying copyright fees.

Permission is granted.

If you require a formal record, please go to Rightslink directly. There will be no charge for this.

Thank you for checking.

Regards,

Eileen Evans-Nantais
Client Service Representative
Canadian Science Publishing/Éditions Sciences Canada
1200 Montreal Road, Building M-55
Ottawa, ON K1A 0R6
P: 613-990-7873
F: 613-952-7656
eileen.evans-nantais@nrcresearchpress.com
www.nrcresearchpress.com

Dear Aliya

The Royal Society of Chemistry (RSC) hereby grants permission for the use of your paper(s) specified below in the printed and microfilm version of your thesis. You may also make available the PDF version of your paper(s) that the RSC sent to the corresponding author(s) of your paper(s) upon publication of the paper(s) in the following ways: in your thesis via any website that your university may have for the deposition of theses, via your university's Intranet or via your own personal website. We are however unable to grant you permission to include the PDF version of the paper(s) on its own in your institutional repository. The Royal Society of Chemistry is a signatory to the STM Guidelines on Permissions (available on request).

Please note that if the material specified below or any part of it appears with credit or acknowledgment to a third party then you must also secure permission from that third party before reproducing that material.

Please ensure that the thesis states the following:

Reproduced by permission of the PCCP Owner Societies

and include a link to the paper on the Royal Society of Chemistry's website.

Please ensure that your co-authors are aware that you are including the paper in your thesis.

Regards

Gill Cockhead
Publishing Contracts and Copyright Executive

Gill Cockhead (Mrs), Publishing Contracts and Copyright Executive
Royal Society of Chemistry, Thomas Graham House
Science Park, Milton Road, Cambridge CB4 0WF, UK
Tel +44 (0) 1223 432134, Fax +44 (0) 1223 423623
http://www.rsc.org