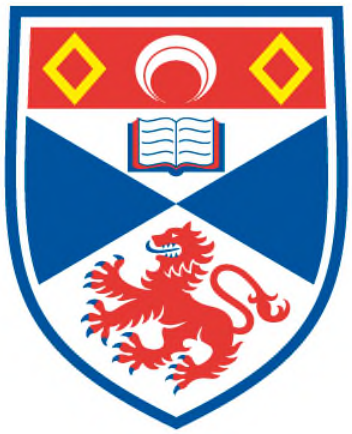
ASSESSMENT OF DENSITY FUNCTIONAL METHODS FOR COMPUTING STRUCTURES AND ENERGIES OF ORGANIC AND BIOORGANIC MOLECULES

Jie Cao

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2011

Assessment of Density Functional Methods for Computing Structures and Energies of Organic and Bioorganic Molecules

A thesis submitted for the degree of Ph.D.

by

Jie Cao

Supervised by Dr. Tanja van Mourik

University of St Andrews

Submitted June 2011

I, Jie Cao, hereby certify that this thesis, which is approximately 26500 Words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.
I was admitted as a research student in September 2007 and as a candidate for the degree of Doctor of Philosophy in January 2009; the higher study for which this is a record was carried out in the university of St Andrews between 2007 and 2011.

Date…… signature of candidate……....

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date……. signature of supervisor……..

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Access to all of printed copy but embargo of all of electronic publication of thesis for a period of 2 years on the following ground:

publication would preclude future publication;

Date…… signature of candidate……....
Date……. signature of supervisor……..

ACKNOWLEDGEMENT

First of all, I would like to thank my supervisor Dr. Tanja van Mourik for offering me a position in her group on this project so that I can have the opportunity to get an insight into the realm of computational chemistry, in which I did not have much experience before the project. Also I appreciate her consistent help throughout my work including discussion with me about every problem encountered, giving suggestions about relative issues, language corrections about my thesis, etc.

In addition, I would like to thank Dr. Herbert Früchtl for helping me with all those software problems including system configuration, software installation, etc. Many thanks to Dr. Wuzong Zhou for giving me advices concerning the research ideas and thesis writing experience. I will attribute my achievements to my group member for sharing their ideas and experience with me generously.

At last, I'd like to thank EaStCHEM for the hardware support for my calculations.

I have enjoyed a great pleasure in my research and obtained a great deal of knowledge which I will never forget in my life.

ABSTRACT

The work in this thesis mainly focuses on the assessment of density functional methods for computing structures and energies of organic and bioorganic molecules.

Previous studies found dramatic conformational and stability changes from B3LYP to MP2 geometry optimization for some Tyr-Gly conformers. Possible reasons could be large intramolecular basis set superposition errors (BSSEs) in the MP2 calculations and the lack of dispersion in the B3LYP calculations.

The fragmentation method and three kinds of rotation methods were used to investigate intramolecular BSSE. It is concluded that the rotation method cannot be used to correct intramolecular BSSE along a rotation profile.

Another methodology is to employ modern density functionals. We focused on M06-L with the Tyr-Gly conformer 'book6'. Potential energy profiles were determined by computing the energy for geometries optimized at various fixed values of a distance that controls the degree of foldedness of the structure. M06-L manifested itself as a very promising method to investigate the potential energy surface of small peptides containing aromatic residues.

To predict Tyr-Gly structures, 108 potential conformers were created with a Fortran program. The geometry optimizations were done using M06-L/6-31G(d) and M05-2X/6-31+G(d). Two schemes were employed and the most stable conformers were compared to the 20 stable conformers found by B3LYP. Both schemes found 10 conformers similar to one of the B3LYP stable conformers, as well as several newly found conformers. The study of a missing B3LYP stable conformer showed that the possible reason of missing conformers may be the lack in dispersion in B3LYP theory.

To study the hydration effect, we studied the conformations of neutral and zwitterionic 3-fluoro-γ-aminobutyric acid (3F-GABA) in solution using different solvation models, mainly the explicit water molecule models. Zwitterionic forms of

3F-GABA are preferred in solution. M06-2X performs better in calculating transition energy profiles than MP2.

Table of Content

Chapter 1 General Introduction

1.1 Quantum chemistry

As a branch of theoretical chemistry, quantum chemistry addresses chemistry problems by applying quantum mechanics and quantum field theory [1]. One of the areas that quantum chemistry has been widely applied to is the electronic behaviour of atoms and molecules relative to their chemical reactivity. Significant contributions have been made by scientists from both chemistry and physics because quantum chemistry lies on the border between them, and has a strong and active overlap with the field of atomic physics and molecular physics, as well as physical chemistry [2].

Quantum chemistry usually describes the fundamental behaviour of matter mathematically at the molecular scale, but it can span from elementary particles such as electrons and photons to the cosmos such as star-formation. In principle, it is possible to describe all chemical systems using this theory. However, in practice, it is only the simplest chemical systems that may be investigated in purely quantum mechanical terms, and approximations must be made for most practical purposes such as Hartree-Fock, post Hartree-Fock or density functional theory. Therefore, it is not necessary to understand the details of quantum mechanics for most chemistry, since the important implications of the theory, principally the orbital approximation, can be understood and applied in simpler terms.

The first step in solving a quantum chemical problem is usually solving the Schrödinger equation (or Dirac equation in relativistic quantum chemistry) with the electronic molecular Hamiltonian, which is called determining the electronic structure of the molecule. From the electronic structure of a molecule or crystal, its chemical properties can be implied.

In quantum mechanics, the sum of two operators can be used to describe the

Hamiltonian, or the physical state, of a particle, one corresponding to kinetic energy and the other to potential energy. In the Schrödinger wave equation used in quantum chemistry, the Hamiltonian does not contain terms for the spin of the electron.

From the solutions of the Schrödinger equation for the hydrogen atom, the form of the wave function for atomic orbitals and the relative energy of the various orbitals can be derived. The orbital approximation can be used to understand the other atoms such as helium, lithium and carbon.

1.2 The central theme of the thesis

Proteins (also known as polypeptides) are organic compounds made of amino acids arranged in a linear chain and folded into a globular form. By the peptide bonds, the amino acids in a polymer are joined together between the carboxyl and amino groups of adjacent amino acid residues. Proteins are the most attracting molecular devices when analyzing the complex structure of a biological system, and are likely involved in all processes of a living organism and responsible for behavioural changes in the cells. Because of the significant role of proteins in biological systems, it is very interesting for molecular biologists to look for the function of each protein, aiming to understand how they can change the state and behaviour of a cell and to use their functions to treat diseases with particular drugs if possible.

So far there are many chemical approaches to determine the structure of a protein. Historically, X-ray crystallography was the first one [3], which appeared in 1934 when Bernal and Crowfoot took the first X-ray photograph of a crystalline globular protein. After that, Wüttrich introduced nuclear magnetic resonance (NMR) in protein structure and dynamic analysis [4]. Both of these methods are reliable, however, to yield a complete definition of structure, they all take a long period of time as well as high costs. Besides, X-ray crystallography can be applied only if it is possible to crystallize a protein into a regular lattice, whereas NMR works with proteins in a

solvated environment. All these factors will cause the protein to take different conformations, leading to many difficulties in determining a single good protein model.

Peptides are short polymers of amino acids linked by peptide bonds, which have the same chemical structure as proteins, with shorter length though. Dipeptides are the shortest peptides and consist of two amino acids joined by a single peptide bond.

It is of significant interest to illustrate the most stable conformations of a peptide. The reason is that peptides have strong relevance to the protein folding problem. Although normally protein folding processes happen in a solution environment, it is still of great relevance to have a good understanding of the conformational preference of gas-phase peptides so as to understand how the structure changes from the gas-phase to a solvated environment. In this project, we focus on gas-phase peptides.

The search for the most stable peptide conformers is a complex issue due to the fact that one of the most prominent features of peptides is their extensive flexibility which means a large number of possible conformers have to be considered to identify the most stable of these, preventing the routine use of high-accuracy but computational expensive electronic structure methods. However, with the development of computer architecture, high-accuracy computational methods are becoming feasible to study peptides, such as the recently prevailing density functional theory.

In the past few years, different methodologies were explored to investigate the conformational features of peptides, with a variety of electronic structure methods being employed. The structure of the dipeptide Tyr-Gly was studied in a previous study by Toroz and van Mourik, in which two methodologies were considered. However, even the best methods used in this study (B3LYP/6-31+G(d) and MP2/6-31+G(d)) sometimes failed to predict the correct structures [5-6].

The theme of the main project of this thesis is to develop a more reliable method to predict the most stable conformations of small peptides.

1.3 The relations between the different projects presented in this thesis

From the previous researches on the Tyr-Gly conformation [5-6], some problems were found in the applied methodology during the conformational analysis of the Tyr-Gly dipeptide, displayed by dramatic conformational and stability changes from B3LYP to MP2 geometry optimization for some of the Tyr-Gly conformers. Two of these conformers, labelled book4 and book6, were studied in more detail over the last three years [5-8]. For “book4”, the B3LYP/6-31+G(d) and MP2/6-31+G(d) optimized structures mainly differed in the value of one particular dihedral angle. For “book6”, B3LYP/6-31+G(d) predicted an “open book” structure while MP2/6-31+G(d) predicted a more folded “closed book” structure. Looking into these problems, two causes have been discovered: Firstly, too folded conformers or a distorted potential energy surface with the minima missing may result from large intramolecular basis set superposition errors (BSSEs) in the MP2 calculations; Secondly, the missing conformers on the B3LYP potential energy surface may result from the lack of dispersion in the B3LYP calculations.

Evidently, intramolecular interactions with aromatic residues in the peptide can affect the conformation of peptides and proteins to a large extent. Particularly, the intramolecular BSSE can influence the interactions a lot, especially in cases when aromatic rings exist. The normal method for BSSE correction, the counterpoise (CP) procedure proposed by Boys and Bernardi [9], can only eliminate intermolecular BSSE and there is no direct way to solve the problem caused by intramolecular BSSE. In this thesis, we consider two methods proposed to eliminate intramolecular BSSE, which are the fragmentation method and the rotation method.

Considering the problems caused by lack of dispersion for certain computational

methods, a more reliable method is needed in the conformational search to reduce the possibility of obtaining false peptide structures, with relatively low computational expense. Therefore, the performance of the density functional M06-L for determining the structure of a folded Tyr-Gly conformer is studied. Potential energy profiles have been created for transition from the B3LYP to the MP2 conformer by optimizing the Tyr-Gly structure at fixed values of the R_{CC} distance that controls the degree of foldedness of the conformer. These are compared with a reference profile taken from previous work [6], which was obtained with the df-LCCSD(T0)/aug-cc-pVTZ method. The results show that M06-L yields results that excellently agree with the reference profile which enables itself a promising functional to study the potential energy surface of small aromatic peptides.

In the main project, the peptides are studied in the gas-phase. However, in a biological environment, water is present. Therefore, ultimately the structure prediction of peptides has to include water molecules into the system, which makes the study of hydration effects important. The third project presented in this thesis provides an initial investigation of solvation effects. In this project, the conformations of neutral and zwitterionic 3-fluoro-γ-aminobutyric acid (3F-GABA) are studied in solution using different solvation models in order to understand better the structure of 3F-GABA in a physiological environment and the relative accuracy of different solvation models. In this work, explicit water molecule models are built by surrounding zwitterionic 3F-GABA conformers with different numbers of explicit water molecules. The structure optimizations and energy calculations are carried out using the density functional M06-2X, which originates, like M06-L, from the Truhlar group.

Chapter 2 General Background Theory

2.1 Introduction

2.2 Computational chemistry methods

There are three kinds of electronic structure methods: ab initio methods, semi-empirical methods and density functional theory.

Ab initio quantum chemistry methods are computational chemistry methods based on quantum chemistry [10]. The term ab initio indicates that the calculation is from first principles and that no empirical data are used (other than a number of physical constants, such as the speed of light).

The simplest type of ab initio electronic structure calculation is the Hartree-Fock (HF) scheme, in which the correlation between electrons of opposite spin is not specifically taken into account. Only its average effect (mean field) is included in the calculation. This is a variational procedure, therefore the obtained approximate energies, expressed in terms of the system's wave function, are always equal to or greater than the exact energy, and tend to a limiting value, called the Hartree-Fock limit, as the size of the basis is increased [11]. Many types of calculations begin with a Hartree-Fock calculation and are subsequently corrected for electronic correlation. Møller-Plesset perturbation theory (MPn) and coupled cluster theory (CC) are examples of these post-Hartree-Fock methods [11-12]. In some cases, particularly for bond breaking processes, the Hartree-Fock method is inadequate and this single-determinant reference method is not a good basis for post-Hartree-Fock methods. It is then necessary to start with a wave function that includes more than one determinant such as Configuration Interaction methods (CI), Multi-configuration Self-consistent Field (MCSCF) methods or Multi-reference Configuration Interaction

(MRCI) methods [11].

Another kind of computational chemistry approach includes the semi-empirical methods, whose main difference from ab initio methods is that they make many approximations and obtain some parameters from empirical data.

Density functional theory (DFT) is based on the proof by Hohenberg and Kohn that the ground-state electronic energy is determined completely by the electron density [13]. DFT describes a molecular system directly via its density, without first finding the wavefunction. It models electron correlation via functionals of the electron density.

2.2.1 Ab initio methods

2.2.1.1 Hartree-Fock Method (HF)

The Hartree-Fock (HF) method is an approximate method for the determination of the wave function and energy of a quantum system.

In order to solve the electronic Schrödinger equation [14]:

$$H^{elec}\psi^{elec}(\vec{r},\vec{R}) = E^{eff}(\vec{R})\psi^{elec}(\vec{r},\vec{R}) \qquad (2.1)$$

Here $\vec{r},\vec{R}$ represent the positions of electrons and the nuclei. H^{elec} is the electronic Hamiltonian which, compared to the full Hamiltonian, neglects the kinetic energy term of the nuclei, and ψ^{elec} is the electronic wavefunction. E^{eff} represents the effective nuclear potential function. Some approximations need to be made. These will lead to the Hartree-Fock method (which is the simplest ab initio method).

The first approximation is to decompose Ψ into a combination of molecular orbitals (MOs) as (for a two-electron system):

$$\psi(\vec{r}) = \phi_1(\vec{r}_1)\phi_2(\vec{r}_2) \qquad (2.2)$$

where the Φ_n are the one-electron wavefunctions. However, this is not a good wavefunction, as wavefunctions need to be antisymmetric which means swapping the

coordinates of two electrons should lead to a sign change. Therefore we create a good form of the two-electron wavefunction as:

$$\psi(\vec{r}) = \phi_1(\vec{r}_1)\phi_2(\vec{r}_2) - \phi_1(\vec{r}_2)\phi_2(\vec{r}_1) \tag{2.3}$$

In general, the antisymmetry of an N-electron wavefunction can be achieved by constructing the wavefunction as a Slater Determinant :

$$\psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_2(1) & \cdots & \phi_N(1) \\ \phi_1(2) & \phi_2(2) & \cdots & \phi_N(2) \\ \cdots & \cdots & \cdots & \cdots \\ \phi_1(N) & \phi_2(N) & \cdots & \phi_N(N) \end{vmatrix} \tag{2.4}$$

Here Φ_i are "spinorbitals", which contain also the spin of the electron. Then the second approximation is that the Hartree-Fock wavefunction consists of a single Slater Determinant. After that, we use the third approximation: the MOs Φ_i are written as a linear combination of pre-defined one-electron functions (basis functions or AOs):

$$\phi_i = \sum_{\mu=1}^{N} c_{\mu i}\chi_\mu \tag{2.5}$$

Here $c_{\mu i}$ are the expansion coefficients, and χ_μ are the AOs or basis functions. This is the LCAO (linear combination of atomic orbitals) approximation. The exact form of the wavefunction depends on the coefficients $c_{\mu i}$. The Hartree-Fock method aims to find the optimal wavefunction.

According to the Variation Principle (which holds for a variational method like Hartree-Fock): "The energy calculated from an approximation to the true wavefunction will always be greater than the true energy". So we just need to find the coefficients $c_{\mu i}$ that give the lowest energy. This leads to the Hartree-Fock equations

[15], which can be solved by the Self-Consistent Field (SCF) method.

The main weakness of the Hartree-Fock method is that it neglects electron correlation. In Hartree-Fock theory each electron moves in an average field of all the other electrons. Instantaneous electron-electron repulsions are ignored.

2.2.1.2 Configuration Interaction (CI) [16]

The Hartree-Fock method determines the best one-determinant wavefunction for a given basis set, lacking electron correlation. In order to overcome this weakness, additional Slater Determinants can be added to the wavefunction. By obtaining the additional determinants using single, double, triple, etc. excitations from the optimized HF determinant and putting them into a linear combination, we get the Configuration Interaction method:

$$\psi_{CI} = a_0\phi_{HF} + \sum_S a_S\phi_S + \sum_D a_D\phi_D + \cdots \qquad (2.23)$$

Here, Φ_{HF} is the Hartree-Fock determinant, Φ_S is the Hartree-Fock determinant with one Molecular Orbital (MO) swapped with a virtual orbital ("single excitation"), Φ_D is the Hartree-Fock determinant with two MOs swapped with virtual orbitals ("double excitation") and a_0, a_S, a_D are CI expansion coefficients. The Configuration Interaction method aims to find the optimal CI expansion coefficients a_0, a_S, a_D, …by using the variation principle (find those coefficients that yield the wavefunction with the lowest energy). When all possible excited Slater Determinants are included, it becomes Full CI which is very expensive and not feasible for any but the smallest systems.

2.2.1.3 Multi-configuration Self Consistent Field (MCSCF) [11-12]

When not only the CI expansion coefficients a_i, but also the MOs in the determinants (the MO coefficients) are optimized by the variation principle, the CI

becomes an MCSCF. It is usually used for cases where HF does not give a qualitatively correct description.

In a special case, the Molecular Orbitals are divided into active and inactive spaces. The active space includes some of the highest occupied and some of the lowest unoccupied MOs. A full CI is performed for the active MOs. This case is called Complete Active Space SCF (CASSCF).

Usually when HF does not give a correct reference wavefunction, for example for molecular ground states that are almost degenerate with low-lying excited states, in bond breaking situations, or for describing excited states, MCSCF will be employed. It recovers some electron correlation, but mostly “static electron correlation” resulting from the additional flexibility required to qualitatively describe the system, and not so much “dynamic electron correlation” (energy lowering caused by correlating the motions of the electrons).

2.2.1.4 Multi-reference Configuration Interaction (MRCI)

Multi-reference Configuration Interaction is a CI with a MCSCF wavefunction as the reference. However, it is very computationally demanding.

2.2.1.5 Møller-Plesset Perturbation theory (MP)

Møller-Plesset perturbation theory is one of several quantum chemistry post-Hartree-Fock ab initio methods in the field of computational chemistry. It improves on the Hartree-Fock method by adding electron correlation effects by means of Rayleigh-Schrödinger perturbation theory (RS-PT), usually to second (MP2), third (MP3) or fourth (MP4) order. Its main idea was published as early as 1934 [17]. Nowadays, MP3 and MP4 are not much used anymore, as the MP series has been shown to not always converge [18].

2.2.1.5.1 Rayleigh-Schrödinger perturbation theory (RS-PT)

The MP-theory [17] is a special application of Rayleigh-Schrödinger perturbation theory. In RS-PT one considers an unperturbed Hamiltonian operator $\hat{H}_0$ to which is added a small (often external) perturbation $\hat{V}$:

$$\hat{H} = \hat{H}_0 + \lambda\hat{V} \tag{2.6}$$

where λ is an arbitrary real parameter. In MP-theory the zeroth-order wave function is an exact eigenfunction of the Fock operator, which thus serves as the unperturbed operator. The perturbation is the correlation potential.

In RS-PT the perturbed wave function and perturbed energy are expressed as a power series in λ:

$$\Psi = \lim_{n\to\infty}\sum_{i=0}^{n}\lambda^i\Psi^{(i)} \tag{2.7}$$

$$E = \lim_{n\to\infty}\sum_{i=0}^{n}\lambda^i E^{(i)} \tag{2.8}$$

Substitution of these series into the time-independent Schrödinger equation gives a new equation:

$$(\hat{H}_0 + \lambda V)(\sum_{i=0}^{n}\lambda^i\Psi^{(i)}) = (\sum_{i=0}^{n}\lambda^i E^{(i)})(\sum_{i=0}^{n}\lambda^i\Psi^{(i)}) \tag{2.9}$$

Equating the factors of λ^k in this equation gives an kth-order perturbation equation, where k=0,1,2, ..., n.

2.2.1.5.2 Møller-Plesset Perturbation Theory

The MP-energy corrections are obtained from Rayleigh-Schrödinger (RS) perturbation theory with the perturbation (correlation potential):

$$\hat{V} \equiv H - F - \langle \Phi_0 \mid H - F \mid \Phi_0 \rangle \tag{2.10}$$

Where the normalized Slater determinant Φ_0 is the lowest eigenfunction of the Fock operator:

$$F\Phi_0 \equiv (\sum_{k=1}^{N} f(k))\Phi_0 = 2(\sum_{i=1}^{N/2} \varepsilon_i)\Phi_0 \tag{2.11}$$

Here N is the number of electrons of the molecule under consideration, H is the usual electronic Hamiltonian, f(1) is the one-electron Fock operator, and ε_i is the orbital energy belonging to the doubly occupied spatial orbital φ_i. The shifted Fock operator serves as the unperturbed (zeroth-order) operator:

$$\hat{H}_0 \equiv F + \langle \Phi_0 \mid H - F \mid \Phi_0 \rangle \tag{2.12}$$

The Slater determinant Φ_0 being an eigenfunction of F, it follows readily as:

$$F\Phi_0 - \langle \Phi_0 \mid F \mid \Phi_0 \rangle \Phi_0 = 0 \Rightarrow \hat{H}_0 \Phi_0 = \langle \Phi_0 \mid H \mid \Phi_0 \rangle \Phi_0 \tag{2.13}$$

So that the zeroth-order energy is the expectation value of H with respect to Φ_0, i.e., the Hartree-Fock energy:

$$E_{MP0} \equiv E_{HF} = \langle \Phi_0 \mid H \mid \Phi_0 \rangle \tag{2.14}$$

Since the first-order MP energy

$$E_{MP1} \equiv \langle \Phi_0 \mid \hat{V} \mid \Phi_0 \rangle = 0 \tag{2.15}$$

is zero, the lowest-order MP correlation appears in second order. This result is the Møller-Plesset theorem [17]: the correlation potential does not contribute in first-order to the exact electronic energy. In order to obtain the MP2 formula for a closed-shell molecule, the second-order RS-PT formula is written on the basis of doubly-excited Slater determinants. (Singly-excited Slater determinants do not contribute because of the Brillouin theorem). After application of the Slater-Condon rules for the simplification of N-eletron matrix elements with Slater determinants in bra and ket and integrating out spin, the MP2 energy becomes:

$$E_{MP2} = \sum_{i,j,a,b} \langle \varphi_i(1)\varphi_j(2) \mid r_{12}^{-1} \mid \varphi_a(1)\varphi_b(2) \rangle \times \frac{2\langle \varphi_a(1)\varphi_b(2) \mid r_{12}^{-1} \mid \varphi_i(1)\varphi_j(2) \rangle - \langle \varphi_a(1)\varphi_b(2) \mid r_{12}^{-1} \mid \varphi_j(1)\varphi_i(2) \rangle}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \tag{2.16}$$

where φ_i and φ_j are canonical occupied orbitals and φ_a and φ_b are canonical virtual orbitals. The quantities ε_i, ε_j, ε_a, and ε_b are the corresponding orbital energies. Clearly, through second-order in the correlation potential, the total electronic energy is given by the Hartree-Fock energy plus second-order MP correction: $E \approx E_{HF} + E_{MP2}$. The solution of the zeroth-order MP equation (which by definition is the Hartree-Fock equation) gives the Hartree-Fock energy. The first non-vanishing perturbation correction beyond the Hartree-Fock treatment is the second-order energy.

Equivalent expressions are obtained by a slightly different partitioning of the Hamiltonian, which results in a different division of energy terms over zeroth- and first-order contributions, while for second- and higher-order energy corrections the two partitionings give identical results. This difference is due to the fact, well-known in Hartree-Fock theory, that

$$\langle \Phi_0 \mid H - F \mid \Phi_0 \rangle \neq 0 \Leftrightarrow E_{HF} \neq 2\sum_{i=1}^{N/2} \varepsilon_i \tag{2.17}$$

(The Hartree-Fock energy is not equal to the sum of occupied-orbital energies). In the alternative partitioning one defines,

$$\hat{H}_0 \equiv F \quad , \quad \hat{V} \equiv H - F \tag{2.18}$$

Clearly in the partitioning,

$$E_{MP0} = 2\sum_{i=1}^{N/2} \varepsilon_i \quad , \quad E_{MP1} = E_{HF} - 2\sum_{i=1}^{N/2} \varepsilon_i \tag{2.19}$$

Obviously, the Møller-Plesset theorem does not hold in the sense that $E_{MP1} \neq 0$. The solution of the zeroth-order MP equation is the sum of orbital energies. The zeroth plus first order correction yields the Hartree-Fock energy. As with the original formulation, the first non-vanishing perturbation correction beyond the Hartree-Fock treatment is the second-order energy. The second- and higher-order corrections are the same in both formulations.

2.2.1.6 Coupled cluster theory (CC) [19-21]

While perturbation theory adds all types of corrections (S, D, T, Q, …) to the reference wavefunction to a given order (first, second, …), coupled cluster theory adds all corrections of a certain type to infinite order.

The coupled cluster wavefunction can be written as :

$$\Psi cc = e^{T} \ \phi_0 \qquad (2.20)$$

Where e^{T} is the exponential operator and ϕ_0 is the HF wavefunction. The exponential operator can be expanded into a Taylor series:

$$e^{T} = 1 + T + 1/2\ T^2 + 1/6\ T^3 + \ldots \qquad (2.21)$$

The cluster operator is written in the form:

$$T = T_1 + T_2 + T_3 + \ldots + T_n \qquad (2.22)$$

Here T_1 represents the operator of all single excitations, T_2 represents the operator of all double excitations and so forth. When all operators up to T_n are included, it becomes full CI, which is not feasible for any but the smallest systems. Therefore, for large systems the cluster operator must be truncated at some excitation level. For example, when $T = T_1 + T_2$, it is the CCSD method; when $T = T_1 + T_2 + T_3$, it is the CCSDT method; when the triples are derived based on perturbation theory, then it becomes the CCSD(T) method, which is a very accurate method.

2.2.2 Semi-empirical quantum chemistry methods [22-23]

Based on the Hartree-Fock formalism, semi-empirical quantum chemistry methods apply some approximations and get some parameters from empirical data.

They are very important in computational chemistry for treating large molecules where the full Hartree-Fock method without the approximations is too expensive. The use of empirical parameters appears to allow some inclusion of electron correlation effects into the methods. Semi-empirical calculations are much faster than their ab initio counterparts. However, their results can be very wrong if the molecule being computed is not similar enough to the molecules in the database used to parametrize the method. They have been most successful in the description of organic chemistry, where only a few elements are used extensively and molecules are of moderate size.

2.2.3 Density Functional Theory (DFT) [24-25]

2.2.3.1 Introduction to Density Functional Theory

During the past decades, the density functional theory has emerged as a powerful methodology for the simulation of chemical systems. It is built based on the premise that the energy of an electronic system can be defined in terms of its electron probability density, which means the electronic energy is regarded as a functional of the electron density, in the sense that there exists a one-to-one correspondence between the electron density of a system and its energy.

The advantage of the DFT treatment over a more pure approach based on the notion of a wavefunction can be best demonstrated according to the following: given a system comprising n electrons, its wavefunction would have three coordinates for each electron and one more per electron if the spin is included, i.e., a total of 4n coordinates, whereas the electron density depends only on three coordinates, independently of the number of electrons that constitute the system [26]. Therefore, the electron density maintains the same number of variables, independent of the system size while the complexity of the wavefunction increases with the number of electrons.

Many interesting reviews on DFT have been published by now [27-33]. They

focused on a variety of aspects including the theory, methodological developments, and the practical application of DFT to specific problems. However, the range of possibilities in this field has been extended dramatically with the computational development in the past decades. For many years B3LYP has been the most widely used functional [34-35], but nowadays a great number of density functionals at different levels of sophistication have become available such as the famous Minnesota functionals from the Truhlar group [36-40].

2.2.3.2 Basic Principles: The Hohenberg-Kohn Theorem

The first time when the concept of density functional came out was in the late 1920s, developed by E. Fermi [41] and L. H. Thomas [42], who introduced the idea of expressing the energy of a system as a function of total electron density. In 1951, J. C. Slater [43] applied a very similar idea into the development of the Hartree-Fock-Slater method, which was initially considered as an approximate methodology to the Hartree-Fock theory, but thought of as a predecessor theory of DFT nowadays.

However, these theories were only able to relate (albeit with several limitations) the energy and other properties of the system with the electron density, until a formal proof of this notion came only in the 1960s, when P. Hohenberg and W. Kohn published a theorem [44] illustrating that the ground-state energy of a nondegenerate electronic system and the corresponding electronic properties are uniquely defined by its electron density. Nevertheless, although the existence of a functional relating the electron density and the energy of a system has been confirmed by the Hohenberg-Kohn theorem, the form of such a functional still cannot be told. The search for functionals able to connect these two quantities remains one of the goals of DFT methods.

2.2.3.3 The Kohn-Sham Formalism

In 1965, with the introduction of atomic orbitals, W. Kohn and L. Sham [45] developed a formalism that was the foundation of the current application of DFT in the computational chemistry field. A practical way was yielded by this formalism to solve the Hohenberg-Kohn theorem for a set of interacting electrons, starting from a virtual system of non-interacting electrons which have an overall ground-state density equal to the density of some real system of chemical interest where electrons do interact. The difficulty in representing the kinetic energy of the system was the main problem behind initial DFT formalisms. The main assumption in the Kohn-Sham approach is that the kinetic energy functional of a system can be split into two parts: one part that can be calculated exactly and that considers electrons as noninteracting particles and a small correction term accounting for electron-electron interaction.

2.2.3.4 Classification of DFT Methods

DFT functionals can be classified into the following categories: the Local Density Approximation (LDA), the Generalized Gradient Approximation (GGA), hybrid GGA, hybrid-meta GGA and meta-GGA.

LDA assumes that the density locally can be treated as a uniform electron gas, or equivalently that the density is a slowly varying function; GGA assumes that the density functional depends on the up and down spin densities and their reduced gradient; Hybrid GGA contains a percentage of exact Hartree-Fock exchange; Hybrid-meta GGA in addition explicitly depends on the kinetic energy density. Meta-GGA has no Hartree-Fock exchange and depends on the up and down spin kinetic energy densities.

Hybrid GGA functionals include B3LYP [46-47], X3LYP [48], B97-1 [49] and BHandH (or BH&H) [50]. By far B3LYP is the most popular functional, accounting for most occurrences in the literature over 1990-2006 [51]. The X3LYP functional favors noncovalent interactions [48], describes hydrogen bonding accurately [52],

but it cannot be used for stacking interactions [53]. B97-1 tends to give the best results for a combination of thermochemical kinetics and nonbonded interactions [39] and is the best non-meta functional in a study on H-bonded and stacked structures of formic acid tetramers and formamide tetramers [54]. BHandH can obtain good results for dispersion systems, but it overestimates hydrogen-bonding interactions [55].

Hybrid meta GGA includes the following functionals: PWB6K, M05, M05-2X, M06-HF, M06 and M06-2X. PWB6K was used for thermochemistry and nonbonded interactions [56] and consistently performed well for noncovalent interactions [57-58]. The M05 and M05-2X, M06, M06-2X and M06-L functionals were developed by Truhlar et al. and are part of the M05 [59] and M06 [40, 60-61] series. The M06 suite was developed on the experience obtained from the M05 functionals and supercedes these essentially [40]. M06-L is a non-hybrid (local) functional, therefore, it is less computationally demanding. This is the reason why we investigate if it gives sufficiently accurate results in this thesis.

2.3 Basis sets [10-11, 62-64]

2.3.1 Introduction

In chemistry, a basis set is a set of functions used to generate the molecular orbitals, which can be expanded as a linear combination of basis functions with certain weights or coefficients to be determined. Generally these functions are centered on atoms, named atomic orbitals. Alternatively, they can be centered on bonds or lone pairs. In addition, basis sets can be composed of sets of plane waves down to a cutoff wavelength. These are often used in calculations involving systems with periodic boundary conditions.

When performing molecular calculations, a basis composed of a finite number of atomic orbitals is commonly used. The wavefunctions under consideration in those

situations are all represented as a linear combination of the basis functions in the basis set used. If the finite basis is expanded towards an infinite complete set of functions, it is said that those calculations are approaching the basis set limit [65].

Atomic orbitals may be Slater orbitals, which are functions that decay exponentially with distance from the nuclei. These Slater-type orbitals (STOs) can in turn be approximated by a linear combination of Gaussian-type orbitals (GTOs), as accomplished by Frank Boys. Since it is easier to compute overlap and other integrals with Gaussian basis functions, a great deal of computational savings can be realized using GTO-based basis sets.

Nowadays there are hundreds of basis sets composed of Gaussian-type orbitals. The smallest ones of these are called minimal basis sets, which are composed of the minimum number of basis functions required to represent all of the electrons on each atom.

The addition of polarization functions is probably the most common addition to minimal basis sets, denoted by an asterisk, *, in Pople-type basis sets. If two asterisks are added, the polarization functions are also added to light atoms (hydrogen and helium). These are auxiliary functions with one additional node. For example, in a minimal basis set, the only basis function positioned on a hydrogen atom would be a function approximating the 1s atomic orbital. When adding a polarization function to this basis set, additional flexibility is added to the basis set which effectively enables the creation of molecular orbitals that are more asymmetric about the hydrogen nucleus. Similarly, d-type functions can be added to a basis set containing s and p orbitals, and so on. A more precise notation than the * and * * notation indicates exactly which and how many functions are added to the basis set. For example, (p, d), which is equivalent to * *.

The addition of diffuse functions is another common addition to the basis sets, which is represented by a plus sign, +, in Pople-type sets, and by “aug” (from “augmented”) in Dunning-type correlation consistent basis sets. Two plus signs means

that diffuse functions are added to light atoms (hydrogen and helium) as well. These are very shallow Gaussian basis functions. The effect of additional diffuse basis functions can be significant concerning anions and other large, “soft” molecular systems.

2.3.2 Minimal basis sets

The most common minimal basis set is STO-nG, where n is an integer. N means the number of Gaussian primitive functions comprising a single basis function. In this situation, the core and valence orbitals are composed of the same number of Gaussian primitive orbitals. Minimal basis sets are typically insufficient to give accurate results for researches, but they are much cheaper in computational expense. Examples of minimal basis sets are:

STO-3G

STO-4G

STO-6G

2.3.3 Split-valence basis sets [66-67]

Chemical bonding is mainly created due to valence electrons during most molecular bonding situations. Therefore, it is necessary to represent valence orbitals by more than one basis function. Basis sets in which there are multiple basis functions corresponding to each valence atomic orbital are called valence double, triple, quadruple-zeta, and so on, basis sets. Because the different orbitals of the split have different spatial extents, the combination allows the electron density to adjust its spatial extent appropriate to the special molecular environment. The fixed minimum basis sets cannot adjust to varing molecular environments.

2.3.4 Pople basis sets [68-69]

The Pople-type split-valence basis sets can be denoted as X-YZg. X means the number of primitive Gaussians comprising each core atomic orbital basis function. The Y and Z mean that two basis functions are composing the valence orbitals; the first one is composed of a linear combination of Y primitive Gaussian functions, the other one is composed of a linear combination of Z primitive Gaussian orbitals. Therefore, two numbers after the hyphens mean that it is a split-valence double-zeta basis set. The forms of split-valence triple- and quadruple-zeta basis sets are denoted as X-YZWg and X-YZWVg, etc. The following forms are a list of frequently used split-valence basis sets of this type:

3-21G

3-21G* - Polarized

3-21+G - Diffuse functions

3-21+G* - With polarization and diffuse functions

4-21G

4-31G

6-21G

6-31G

6-31G*

6-31+G*

6-31G(3df, 3pd)

6-311G

6-311G*

6-311+G*

2.3.5 Correlation-consistent basis sets

Dunning and coworkers [70-73] have developed some of the most widely used basis sets, which are designed to converge systematically to the complete-basis-set (CBS) limit using extrapolation techniques. For atoms in the first and second rows of the periodic table, the notation of the correlation consistent basis sets is cc-pVNZ where N=D, T, Q, 5, 6, … (D=double, T=triple, etc.). The “cc-p” means “correlation-consistent polarized” and the “V” (valence) indicates that these are split-valence basis sets. Larger shells of polarization (correlating) functions (d, f, g, etc.) are included successively. Presently these “correlation-consistent polarized” basis sets have been widely used. See the following as examples:

cc-pVDZ – Double-zeta

cc-pVTZ - Triple-zeta

cc-pVQZ - Quadruple-zeta

cc-pV5Z - Quintuple-zeta, etc.

aug-cc-pVDZ, etc. - Augmented versions of the preceding basis sets with added diffuse functions.

In terms of second-row atoms, it is necessary to add more functions; these are the cc-pV(N+d)Z basis sets. Even larger atoms require the cc-pVNZ-PP and cc-pVNZ-DK families of basis sets, where DK and PP stand for Douglas-Kroll and pseudopotential (A simplified potential with an attempt to replace the complicated effects of the motion of the core electrons of an atom and its nucleus with an effective potential, or pseudopotential, so that the Schrödinger equation contains a modified effective potential term instead of the Coulombic potential term for core electrons normally found in the Schrödinger equation).

2.4 Basis Set Superposition Errors (BSSE) and the Counterpoise Procedure (CP)

In quantum chemistry, calculations of interaction energies are susceptible to basis set superposition error if they use finite basis sets. If the atoms of interacting molecules (or of different parts of the same molecule) approach one another, each individual monomer “borrows” functions from other nearby components, effectively increasing its basis set and decreasing its energy. If the total energy is minimized as a function of the system geometry, the short-range energies from the mixed basis sets in the system must be compared with the long-range energies from the unmixed sets, and this mismatch introduces an error, that is, BSSE.

Two methods exist to eliminate this problem. One is the chemical Hamiltonian approach (CHA) [74], which replaces the conventional Hamiltonian with one designed to prevent basis set mixing a priori, by removing all the projector-containing terms which would allow basis set extension.

The other method is the counterpoise approach (CP), which was proposed by Boys and Bernardi in 1970 [9]. In this procedure, the monomer calculations have the same flexibility as what is available to them in the dimer calculation, which means the monomer energies are evaluated in the complete dimer basis set.

The interaction energy ΔE is given by:

$$\Delta E(R) = E_{AB}(R) - E_A - E_B \tag{2.24}$$

Here $E_{AB}(R)$ is the energy of the AB dimer at geometry R, E_A and E_B are the energies of the separate monomers, which equals the energy of AB when $R=\infty$. The counterpoise-corrected interaction energy is given by the following equation:

$$\Delta E^{CP}(R) = E_{AB}(R) - E_A^{\{AB\}}(R) - E_B^{\{AB\}}(R) \tag{2.25}$$

The superscripts {AB} indicate that the monomer energies are computed in the dimer basis set, A ∪ B. The {B} basis in the $E_A^{\{AB\}}$ calculation and the {A} basis in the $E_B^{\{AB\}}$ calculation are usually referred to as "ghost" basis sets. Equation 2.24 represents the uncorrected interaction energy, referred to as ΔE^{noCP}. ΔE_{BSSE} is usually defined as the difference between ΔE^{CP} and ΔE^{noCP}:

$$\Delta E_{BSSE} = \Delta E^{noCP} - \Delta E^{CP} = E_A^{\{AB\}}(R) + E_B^{\{AB\}}(R) - E_A^{\{A\}} - E_B^{\{B\}} \tag{2.26}$$

Equation 2.25 is valid for the interaction between two atoms A and B, with interfragment distance R. If A and B are molecules, they are brought to the complex geometry R together with their geometry deformed to the same geometry as they have in the particular complex geometry one is studying. The energy required to bring monomer A to a particular intramolecular geometry r_A is given by the deformation energy:

$$\Delta U_A^{def}(r_A) = E_A^{\{A\}}(r_A) - E_A^{\{A\}}(r_e) \tag{2.27}$$

The deformation energy of B has a similar expression. Here, r_e means the equilibrium geometry of molecule A. The deformation energy is a property of the monomer, therefore, it is computed in the monomer basis set, indicated by the superscripts {A} in Equation 2.27. The total counterpoise-corrected interaction energy with intermolecular geometrical parameters R and intramolecular parameters r_A and r_B can then be written as:

$$\Delta E^{CP}(R, r_A, r_B) = E_{AB}(R, r_A, r_B) - E_A^{\{AB\}}(R, r_A, r_B) - E_B^{\{AB\}}(R, r_A, r_B) + \Delta U_A^{def}(r_A) + \Delta U_B^{def}(r_B) \quad (2.28)$$

Usually the deformation energy is very small at the complex's optimized geometry. Therefore, neglecting it will only introduce a small error in the computed binding energy. However, if considering geometries further away from the equilibrium structure, the deformation energy must be taken into account.

As for the intramolecular BSSE, we can use the fragmentation method to perform the counterpoise correction and calculate the BSSE energy, which will be illustrated in detail in chapter 3.

2.5 Computational methods used in this work

In the "Fragmentation method and Rotation method for intramolecular BSSE correction" chapter, the B3LYP and MP2 levels of methods are performed with the software packages Gaussian (version 03) and NWChem [75].

In the "Performance of M06-L density functional for a folded Tyr-Gly conformer" chapter, the M05-2X/6-31+G* and M06-L/6-31G* optimizations, B3LYP, M05-2X and M06-L single-point calculations are performed with NWChem; whereas the mPW2-PLYP, B3LYP-D and mPW2-PLYP-D calculations were done with ORCA [76].

In the "Predicting the structure of flexible peptides" chapter, all the different levels of methods are performed with the software package Gaussian (version 09).

In the "Modelling zwitterions in solution: 3-fluoro-γ-aminobutyric acid (3F-GABA)" chapter, all the different levels of methods are performed with the software package Gaussian (version 09).

Chapter 3 Fragmentation Method and Rotation Method for Intramolecular BSSE Correction

3.1 Abstract

Two different methods to correct the intramolecular BSSE were studied, namely, the fragmentation method and the rotation method, in order to have a better understanding of their differences and accuracies. The rotation method was further investigated by adding more ghost atoms to the system. Energy profiles were created by rotating one of the central carbon-carbon bonds. It was shown that propanediol was not a good model system to apply the fragmentation method while a new model system, $C_{11}H_8O_4$, with scheme2 (generating the dimer for the fragmentation method by removing the central five carbon atoms) solved this problem. Three kinds of rotation method were applied (the original rotation method, Palermo's extended rotation method and a third rotation method proposed by us). However, the original and the third rotation method yielded BSSE energies with too negative values while Palermo's extended rotation method obtained negative as well as positive BSSE values. The results show that the rotation method cannot be used to correct intramolecular BSSE along a rotational profile.

3.2 Introduction

One of the most magnificent factors that lead to the failure of predicting correct peptide structures is the difficulty of correctly describing the intramolecular interactions occurring in peptides. It is evident that intramolecular interactions involving aromatic residues play an important role in determining peptide and protein

conformation. However, the interactions may be affected a lot by intramolecular basis set superposition errors (BSSE), especially in cases when there are aromatic rings involved. Because the standard method for BSSE correction, the counterpoise (CP) procedure created by Boys and Bernardi [9], can only be applied to eliminate intermolecular BSSE, there is no direct way to solve the artificial attraction problem in the above situations.

To get an estimate of the intramolecular BSSE in a molecular system, one way is to break the molecule into two fragments, and use the CP method to compute the intermolecular BSSE in the complex system. But there is one problem with this approach, that is, to obtain non-radical fragments, the broken bonds may need to be saturated with additional hydrogens and some atoms around the broken bond may need to be removed in order to avoid overlapping or very close atoms.

Recently Palermo et al. published an alternative approach for intramolecular BSSE correction, called the rotation method [77], which is displayed in Fig 3.1. The aim of the procedure is to get the interaction between the two ellipses in C. By rotating the central bond to that extent where the interaction between them can be excluded we get the non-interaction reference geometry, like D shows. But we cannot simply get the interaction energy by computing the difference of the energies of C and D because of the possible BSSE in C. That is why ghost atoms in both A and B are introduced according to the other's conformation. Palermo et al. proposed that the CP-corrected interaction energy can be obtained using the following equation [77]:

$$\Delta E^{CP,rotation} = E_A - E_B \tag{3.1}$$

Here the assumption is made that the calculation of the interacting geometry and the reference geometry should use the same size basis set. The interaction energy without CP-correction can be calculated using the following equation:

$$\Delta E^{noCP} = E_C - E_D \tag{3.2}$$

The BSSE is computed as the difference between the sum of the energies of the two geometries with ghost functions and the sum of the energies of the structures without the ghost functions:

$$\partial_{BSSE}^{rotation} = E_A + E_B - E_C - E_D \qquad (3.3)$$

Note that the BSSE according to equation 3.3 is not equal to ($\Delta E^{noCP} - \Delta E^{CP}$).

However, this method is not correct (it underestimates BSSE), as it only corrects for BSSE in one of the ellipses [78]. This was shown by comparison of the rotation method with the counterpoise method to correct for intermolecular BSSE.

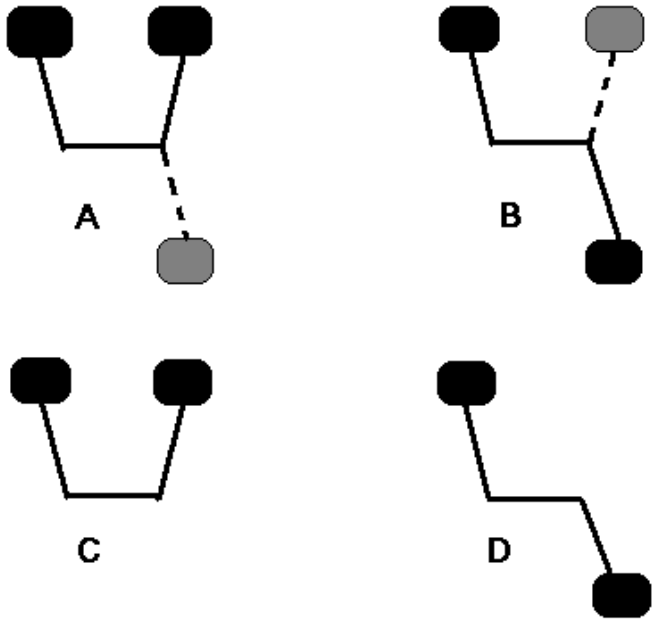


Figure 3.1 Schematic drawing of the molecular fragments that define the rotation method. Solid lines, dark gray ellipses denote the molecular system; dashed lines, light-gray ellipses denote representation by ghost orbitals. A: optimal geometry with ghost atoms; B: non-interacting geometry with ghost atoms; C: optimal geometry; D: non-interacting geometry.

In a response to van Mourik's paper, Palermo defined an extended rotation method, as illustrated in Figure 3.2:

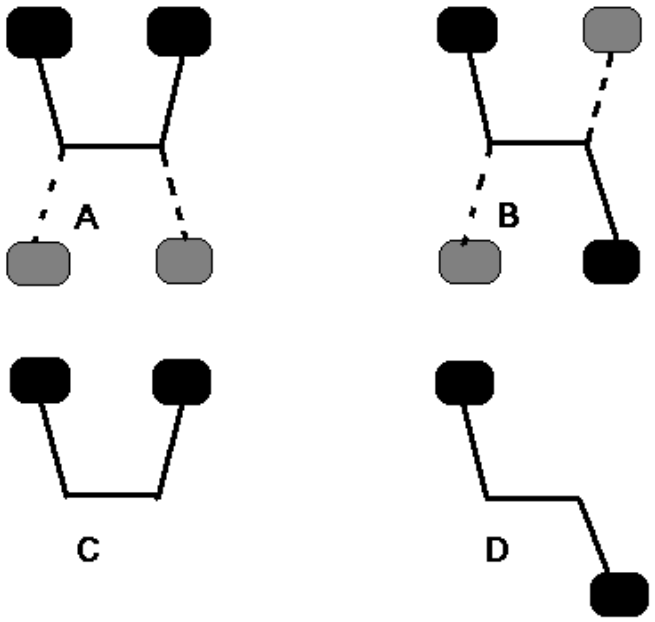


Figure 3.2 Schematic drawing of the molecular fragments that define the rotation method using extra ghost atoms. Solid lines, dark gray ellipses denote the molecular system; dashed lines, light-gray ellipses denote representation by ghost orbitals. A: optimal geometry with extra ghost atoms; B: non-interacting geometry with extra ghost atoms; C: optimal geometry; D: non-interacting geometry.

In the extended rotation method, Palermo et al. argued that extra ghost atoms were needed to make the CP-corrected interaction energy more accurate (see Fig 3.2). This should be the more correct way of correcting for intramolecular BSSE [79].

In the extended rotation method proposed by Palermo, the CP-corrected interaction energy is obtained using the following equation (again, with the assumption that the calculation of the interacting geometry and the reference geometry should use the same size basis set):

$$\Delta E^{CP,rotation} = E_A - E_B \tag{3.4}$$

The interaction energy without CP-correction is calculated using the following equation:

$$\Delta E^{noCP} = E_C - E_D \tag{3.5}$$

The BSSE is computed as the difference between the interaction energy without CP-correction and the CP-corrected interaction energy:

$$\partial_{BSSE}^{rotation} = \Delta E^{noCP} - \Delta E^{CP,rotation} = (E_C - E_D) - (E_A - E_B) \tag{3.6}$$

Inspection of the individual terms in this equation shows no reason why the BSSE energy, defined like this, should always be negative. The fact that the BSSE energy calculated according to Eq. 3.6 can be positive shows that this energy quantity is not comparable to the definition of intermolecular BSSE (which can only be negative).

Thus, because the extended rotation method proposed by Palermo is not equivalent to the counterpoise procedure for intermolecular BSSE, which is demonstrated by equation (2.25). The reason is the common misconception that the size of the basis set should be the same in the interacting and non-interacting cases.

Therefore, we propose a third rotation method that is equivalent to equation (2.25) for the intermolecular case, where the CP-corrected interaction energy is calculated using the equation:

$$\Delta E^{CP,rotation} = E_C - E_B \tag{3.7}$$

The interaction energy without CP-correction is calculated using the equation:

$$\Delta E^{noCP} = E_C - E_D \tag{3.8}$$

The BSSE is calculated as the difference between the interaction energy without CP-correction and the CP-corrected interaction energy:

$$\partial_{BSSE}^{rotation} = \Delta E^{noCP} - \Delta E^{CP,rotation} = E_B - E_D \tag{3.9}$$

3.3 Methodology

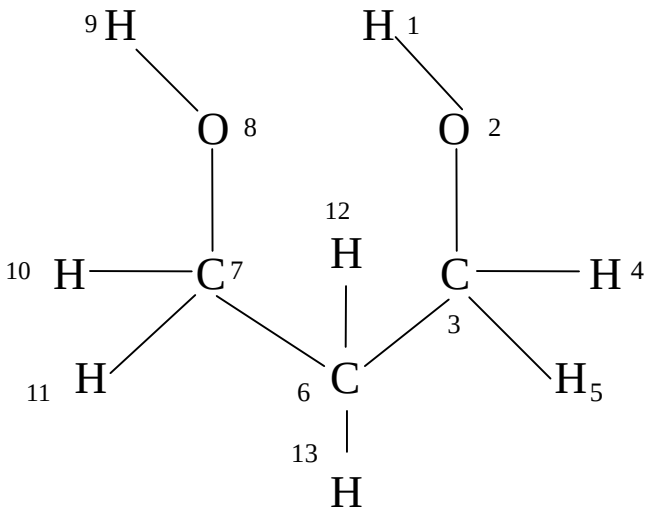


Figure 3.3 Atom labeled structure of propanediol

In order to see how well the rotation method corrects the interaction energy, some simple molecule models were used. One of the models is propanediol whose structure is shown in Figure 3.3.

First, the structure of propanediol was optimized with Gaussian 03 using the B3LYP method with the 6-31+G* basis set. Then by rotating the C3-C6 bond with steps of 20˚, the propanediol's energy was computed using the MP2 method with the 6-31+G* basis set.

Then the rotation method was applied. At each rotation step, the energy of propanediol with ghost atoms at the non-interacting geometry and the energy of propanediol at the non-interacting geometry with ghost atoms replacing the interacting part were also calculated using the MP2 method and the same basis set, like A and B shown in Fig 3.1. The interaction energy was then obtained by computing the difference of these two energies. After that, the fragmentation method was used. At each rotation step, the middle CH_2 group was removed which led to two monomers. The broken bond was saturated by a hydrogen atom of which the position

was optimized using the B3LYP method with the 6-31+G* basis set. The intramolecular BSSE energy, the interaction energy and the energy of the resulting methanol dimer were calculated afterwards using the counterpoise procedure with the MP2 method and the 6-31+G* basis set.

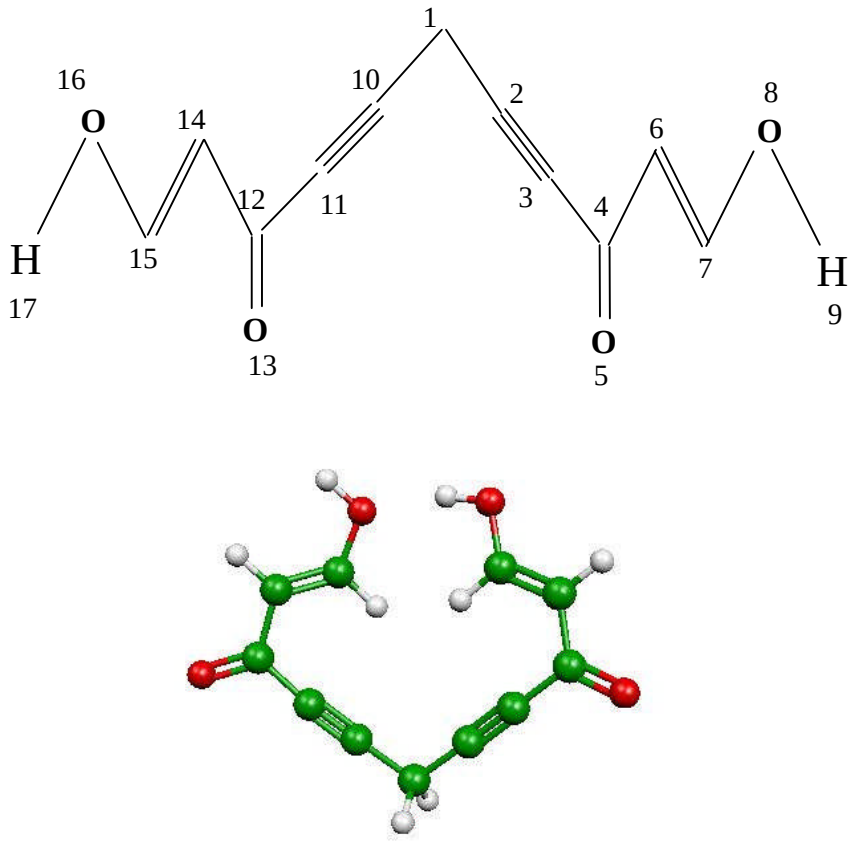


Figure 3.4 Atom labeled structure and three-dimensional plot of $C_{11}H_8O_4$

The same procedure of the rotation method was implemented on a bit more complex molecule, $C_{11}H_8O_4$, of which the atom-labeled structure and three-dimensional plot are shown in Figure 3.4. The rotation bond was C1-C10. For the employment of the fragmentation method on the new molecule, two schemes were

performed. Scheme1 was to remove the middle CH_2 group and saturate the broken bonds and calculate the intramolecular BSSE energy, the interaction energy and the $C_5O_2H_4$ dimer energy using the counterpoise procedure as done for propanediol. Scheme2 was almost similar to Scheme1, the only difference was that instead of removing one carbon atom, the middle five carbon atoms were removed. The following computation was the same as above.

Also, the extended rotation method with extra ghost atoms included was investigated. In this case, two kinds of BSSE calculation method were employed, as illustrated by equations 3.4-3.9.

3.4 Results and Discussion

3.4.1 Propanediol as the model system

First, propanediol was used as the model to calculate the interaction and intramolecular BSSE energy both using the original rotation method and the fragmentation method.

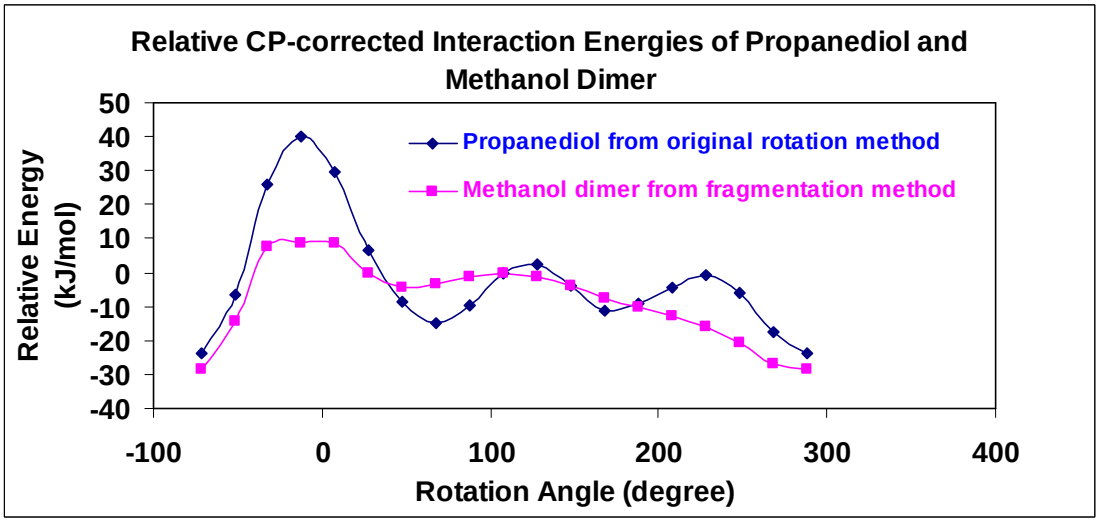


Figure 3.5 The relative CP-corrected interaction energies of propanediol and methanol dimmer. The energy at rotation angle = 108˚ was used as the reference for the relative energies.

Figure 3.5 shows the CP-corrected relative interaction energies of propanediol obtained from the rotation method and the methanol dimer obtained by the fragmentation method at each rotation step. The rotation method yielded a curve of CP-corrected relative interaction energy with three maxima and three minima while the fragmentation method only gave two obvious maxima and minima.

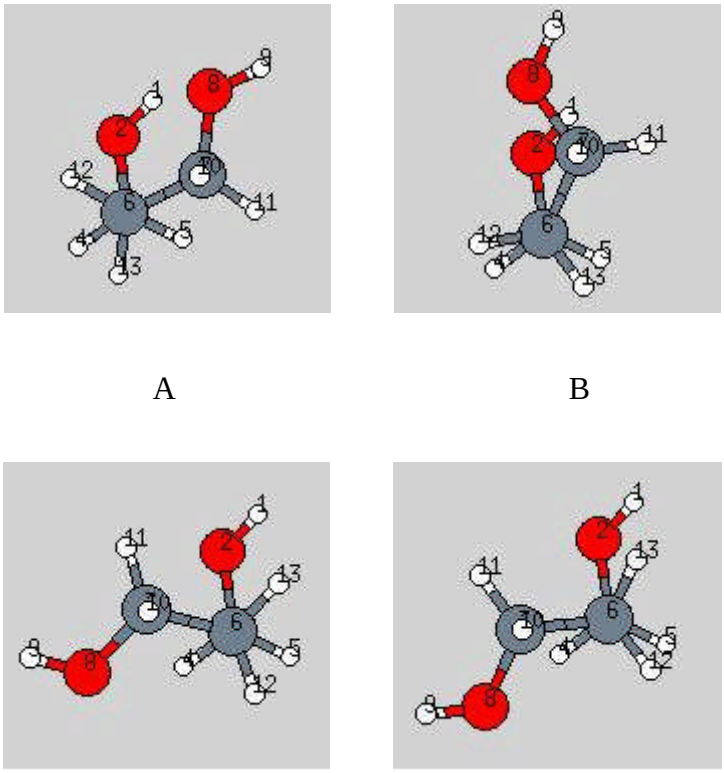


Figure 3.6 Newman projections of propanediol at different stages of the rotation procedure

The curve obtained by the rotation method can be explained as follows: at the beginning, when the rotation angle (the O4-C1-C2-C3 torsion angle) is about -72˚, the two oxygen atoms are close to each other, as shown in Figure 3.6 A. When the C1-C2 bond is rotated, hydrogens H4 and H12 get closer to each other, so do H5 and H13, which makes the interaction between them more repulsive, as shown in Figure 3.6 B. The maximum at -12˚ occurs when the Newman projection turns into the eclipsed form. After that, the interaction becomes less repulsive when the hydrogen atoms get further from each other in the two hydrogen pairs and reaches the first minimum when the hydrogen atoms are staggered, like Figure 3.6 C shows. In the following few steps, H12 and H5 get closer, as well as O2 and H13, therefore the interaction becomes

more repulsive again and reaches the second maximum when the hydrogen atoms are eclipsed again. Because there is an attraction between O2 and H13, which reduces the overall repulsion energy of propanediol, the second maximum is much smaller than the first one. As the rotation continues, the interaction again decreases in repulsion and increases to the third maximum. The reason why the third maximum is much smaller than the first one is the same as above, when the hydrogen atoms are eclipsed. The curve of the fragmentation method is flatter and has only the first two maxima, probably due to the removal of the middle CH_2 group.

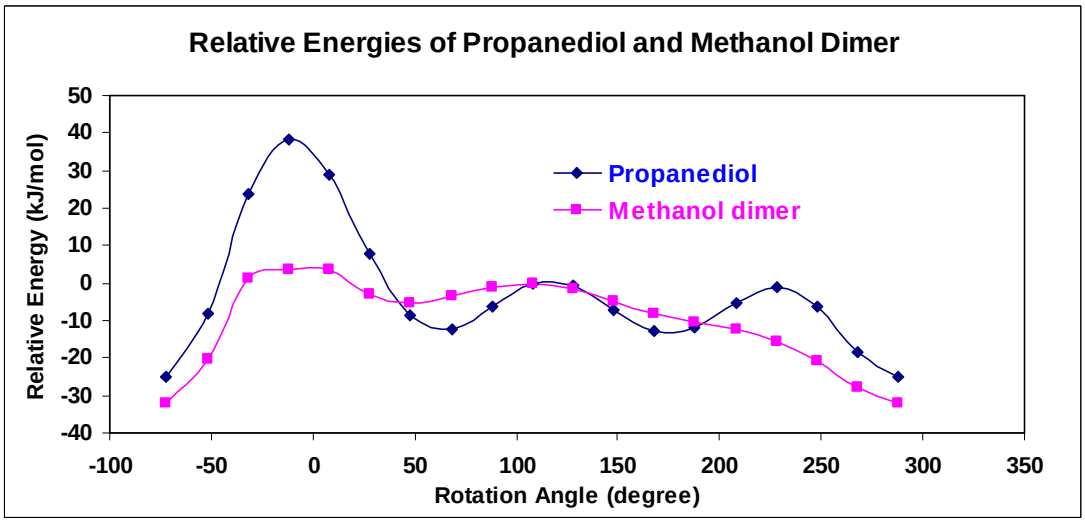


Figure 3.7 The uncorrected relative energies of propanediol and methanol dimmer. The energy at rotation angle = 108˚ was used as the reference for the relative energies.

The relative energies of propanediol and methanol dimer at each rotation step are shown in Figure 3.7. Both curves are very similar to those displayed in Figure 3.5. The relative energy of the methanol dimer is quite different from the relative energy of propanediol which means propanediol is not a good model to test the fragmentation method. The reason could be that the molecule is rather small which causes the two interaction parts, mainly the oxygen atoms, to be close to the middle carbon atom which may therefore affect the interaction, but which is removed in the fragmentation

method. Thus, removal of the middle carbon atom resulted in extraordinary change of the curve of the model energy.

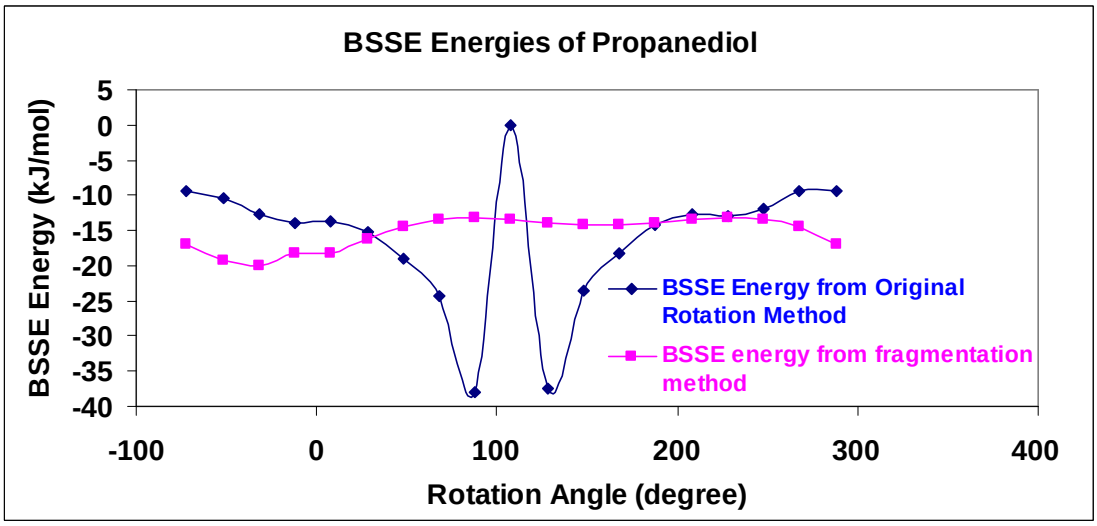


Figure 3.8 BSSE energies of propanediol using the original rotation method

The BSSE energies calculated from the original rotation method and the fragmentation method are very different from each other, as shown in Figure 3.8. The fragmentation method yields BSSE energies between -20 and -10 kJ/mol, while the rotation method gives BSSE energies between 0 and -40 kJ/mol, with the left part of the curve decreasing and the right part going up with increasing rotation angles. In the middle point of the curve, which is the reference position, the BSSE energy jumps to 0, caused by the complete overlapping of the ghost atoms and the real atoms which makes the ghost atoms' effect disappear. Note that the BSSE as defined for the original rotation method (Eq. 3.3) does not equal the difference between the uncorrected and CP-corrected interaction energies. Thus, the BSSE curve displayed in Figure 3.8 is not the same as difference between the uncorrected and CP-corrected interaction energy curves in Figure 3.5 and 3.7.

In addition, as mentioned above, there is a problem with the use of propanediol as

the model system to study the rotation methods, that is, the molecule is small. As a result, the two ends are too close to the central CH_2 group which means that the interaction between those two ends and the middle group cannot be neglected and the removal of the middle group therefore leads to inaccurate results. In order to avoid this problem, a new model, $C_{11}H_8O_4$, was utilized, in which the two interaction parts are far from the middle carbon atoms.

3.4.2 $C_{11}H_8O_4$ as the model system

3.4.2.1 Scheme1 (Removal of central CH_2 group)

For scheme1, the $C_5H_4O_2$ dimer was generated by removing the central CH_2 group.

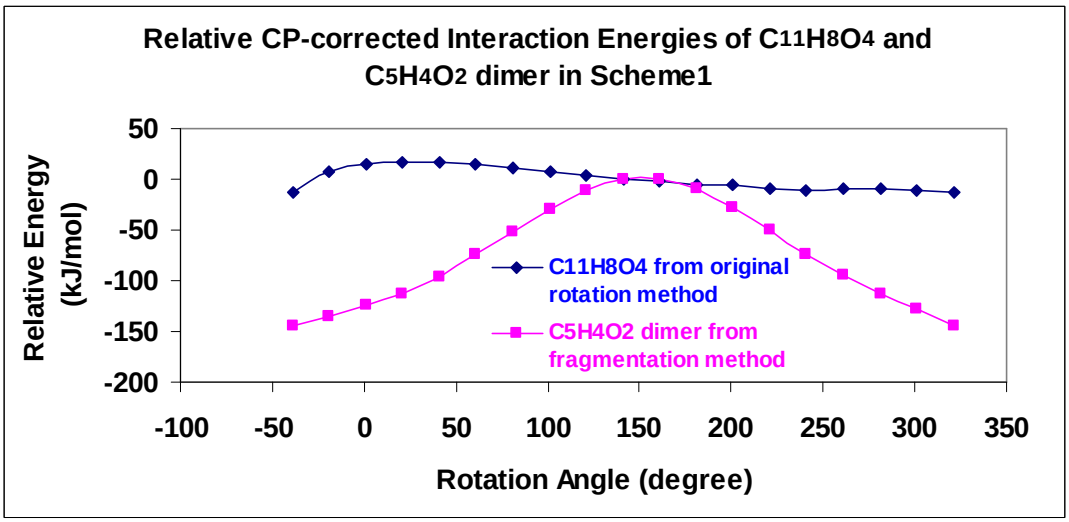


Figure 3.9 The relative CP-corrected interaction energies of $C_{11}H_8O_4$ and $C_5H_4O_2$ dimer in scheme1 (Removal of central CH_2 group). The energy at rotation angle = 141˚ was used as the reference for the relative energies.

Figure 3.9 shows the CP-corrected relative interaction energies of $C_{11}H_8O_4$ obtained from the original rotation method and the fragmentation method at each

rotation step. It can be seen from the graph that the curve obtained from the rotation method is flat, with the energies varying between -14 kJ/mol and -18 kJ/mol. However, the curve obtained by the fragmentation method is totally different. There is a maximum at a rotation angle of ~150˚.

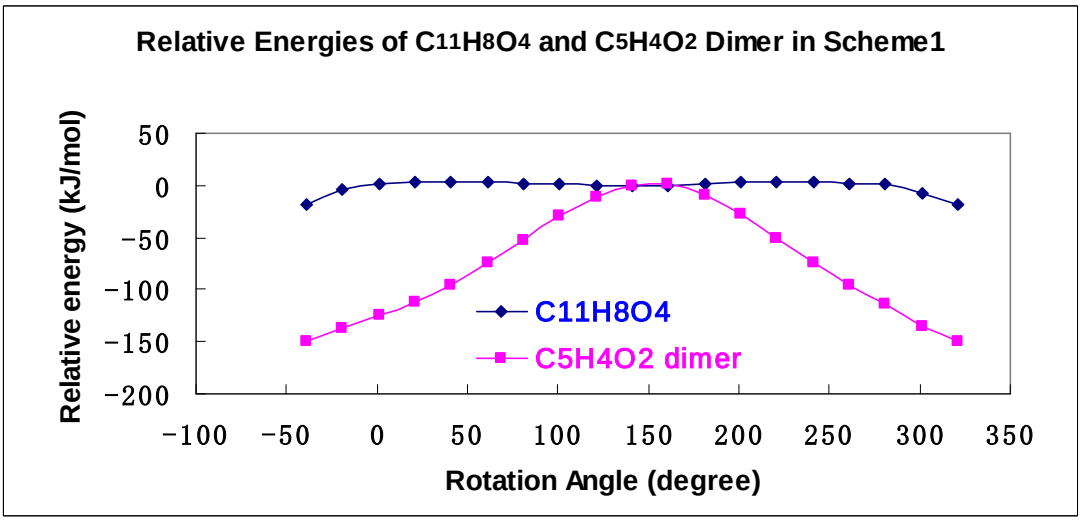


Figure 3.10 The uncorrected relative energies of $C_{11}H_8O_4$ and $C_5H_4O_2$ dimer in scheme1 (Removal of central CH_2 group). The energy at rotation angle = 141˚ was used as the reference for the relative energies.

Figure 3.10 shows the curve of the relative energies of $C_{11}H_8O_4$ and the $C_5H_4O_2$ dimer in scheme1. The same difference exists between the two curves as was observed for the CP-corrected interaction energies. Therefore, $C_{11}H_8O_4$ in connection with scheme1 is not a good model system to investigate the BSSE using the fragmentation method.

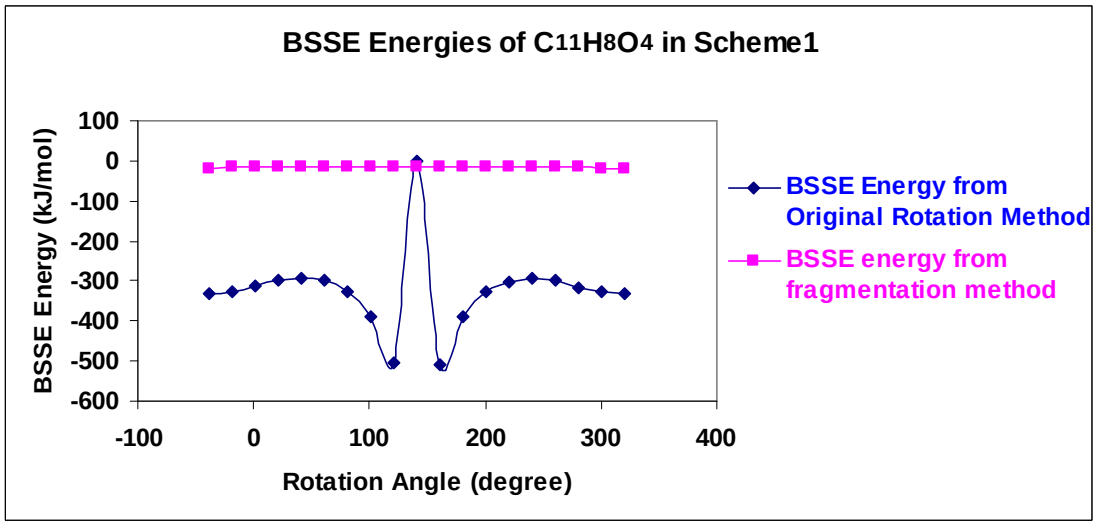


Figure 3.11 BSSE energies of $C_{11}H_8O_4$ in scheme1 (Removal of central CH_2 group) using the original rotation method

Fig 3.11 shows the BSSE curves of the original rotation method and the fragmentation method for $C_{11}H_8O_4$ in scheme1. For the fragmentation method, the BSSE energies vary between -21 kJ/mol and -13 kJ/mol, while for the rotation method, the BSSE energies are much more negative, with the same curve shape as in the propanediol case.

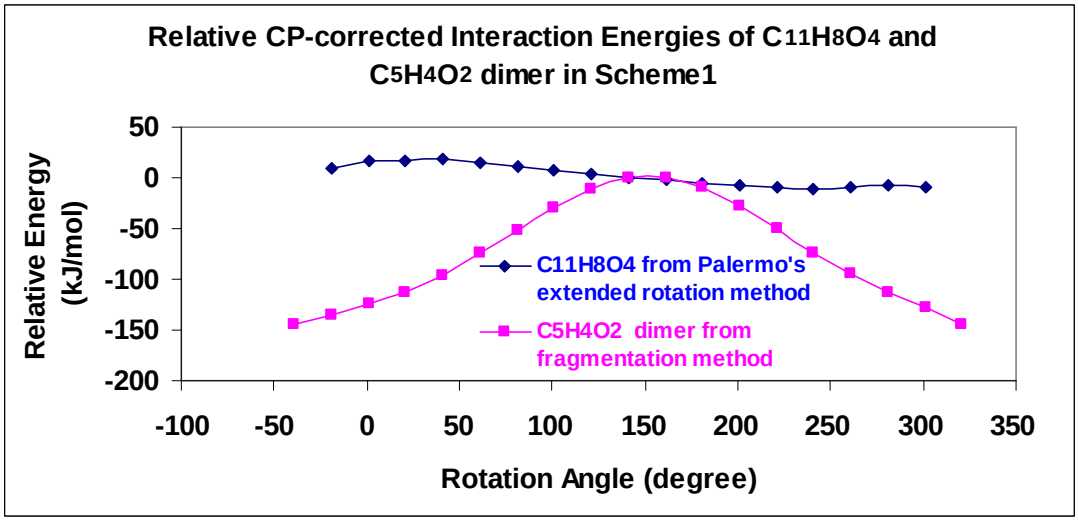


Figure 3.12 Relative CP-corrected interaction energies of $C_{11}H_8O_4$ and $C_5H_4O_2$ dimer in scheme1(Removal of central CH_2 group) using Palermo's extended rotation method. The energy at rotation angle = 141˚ was used as the reference for the relative energies.

Next, we applied the extended rotation method proposed by Palermo et al.

Figure 3.12 shows the relative CP-corrected interaction energies of $C_{11}H_8O_4$ and $C_5H_4O_2$ in scheme1 using Palermo's extended rotation method and the fragmentation method at each rotation step. From the graph, it can be seen that the rotation method yields energies between -11 kJ/mol and 18 kJ/mol, while the dimer's energies from the fragmentation method vary in a range of much larger magnitude, with a maximum at a rotation angle of ~150˚.

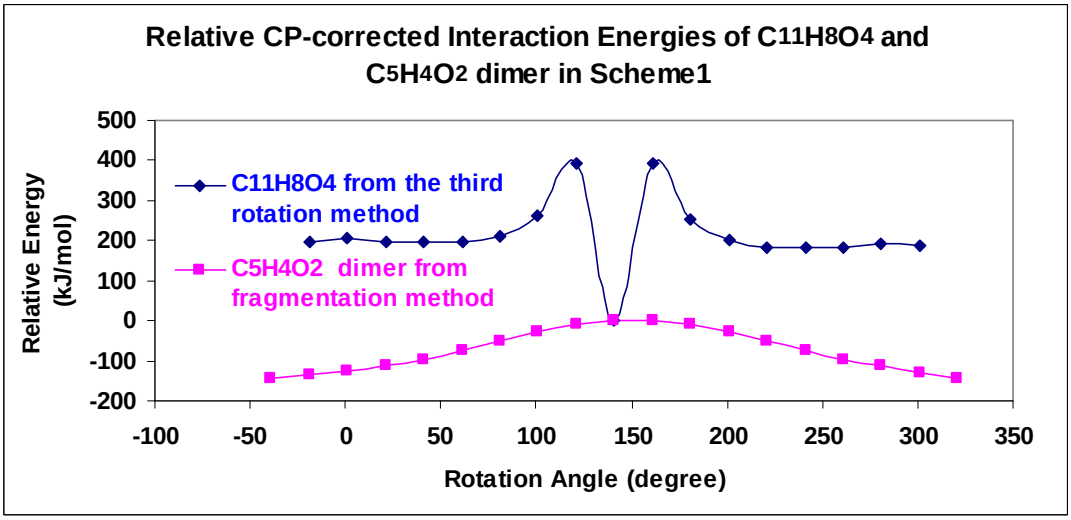


Figure 3.13 The relative CP-corrected interaction energies of $C_{11}H_8O_4$ and $C_5H_4O_2$ dimer in scheme1 (Removal of central CH_2 group) using the third rotation method. The energy at rotation angle = 141˚ was used as the reference for the relative energies.

Figure 3.13 shows the relative CP-corrected interaction energies of $C_{11}H_8O_4$ and $C_5H_4O_2$ in scheme1 using the third rotation method and the fragmentation method at each rotation step. The dimer energy curve from the fragmentation method is the same as that in Figure 3.12. However, the energy curve from the third rotation method displays a strange shape with the left part increasing and the right part decreasing. At the reference position in the middle the energy suddenly drops to 0.

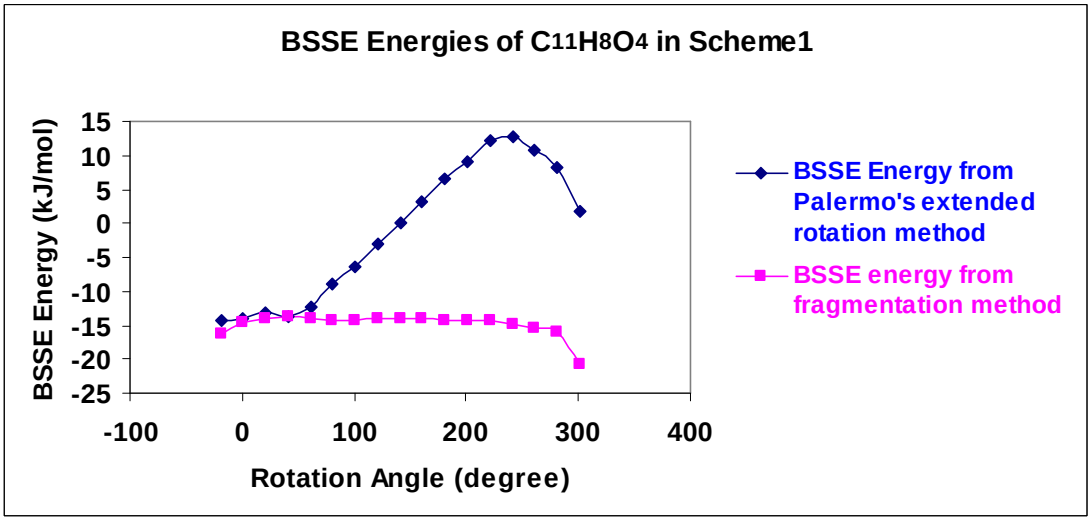


Figure 3.14 BSSE energies of $C_{11}H_8O_4$ in scheme1 (Removal of central CH_2 group) using Palermo's extended rotation method

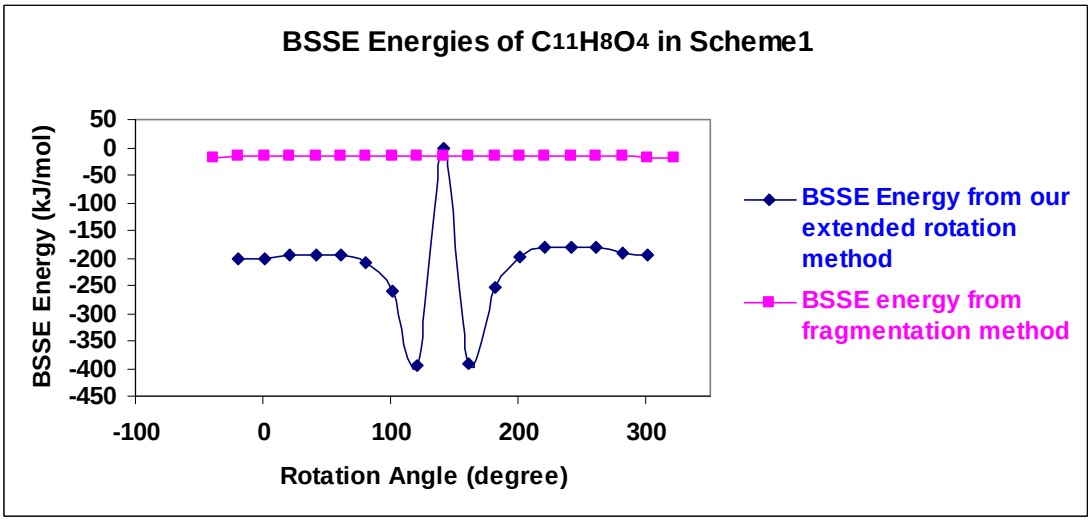


Figure 3.15 BSSE energies of $C_{11}H_8O_4$ in scheme1 (Removal of central CH_2 group) using the third rotation method

Figure 3.14 shows the BSSE energies of $C_{11}H_8O_4$ in scheme1 using Palermo's extended rotation method. The energies obtained with this method vary between -15 and 13 kJ/mol, with the left part below 0 kJ/mol and the right part above 0 kJ/mol with a maximum at about 250˚. As mentioned above, the definition of BSSE in the

extended rotation method (Eq. 3.6) allows the BSSE to be both negative and positive. This indicates that the BSSE as defined in the extended rotation method is not equivalent to the usual definition in the intermolecular counterpoise method. This shows that, compared to the counterpoise method, Palermo's extended rotation method does not give the correct BSSE energies. However, by using the third rotation method proposed by us (see Fig. 3.15), the BSSE curve changes to the same curve shape as that obtained with the original rotation method (see Fig. 3.11), but with the BSSE values much less negative.

As noted above, scheme1 is not a good model for the fragmentation method. It was thought that removing one carbon atom was not enough. This led to scheme2 in which the middle five carbon atoms were removed.

3.4.2.2 Scheme2 (removal of central 5 carbon atoms)

For scheme2, a $C_3H_4O_2$ dimer was generated by removing the middle five carbon atoms.

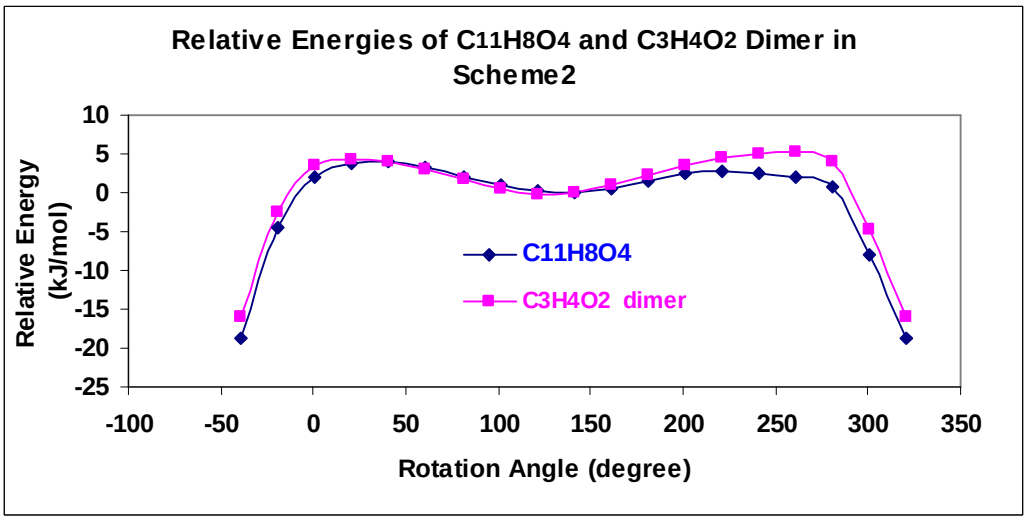


Figure 3.16 The relative energies of $C_{11}H_8O_4$ and $C_3H_4O_2$ dimer in scheme2 (removal of central 5 carbon atoms). The energy at rotation angle = 141˚ was used as the reference for the relative energies.

Figure 3.16 shows the curves of relative energies of $C_{11}H_8O_4$ and the $C_3H_4O_2$ dimer in scheme2. In the graph, the dimer energy curve is quite similar to that of the relative energy of $C_{11}H_8O_4$, both of them having two maxima. This means scheme2 is acceptable for investigating the fragmentation method.

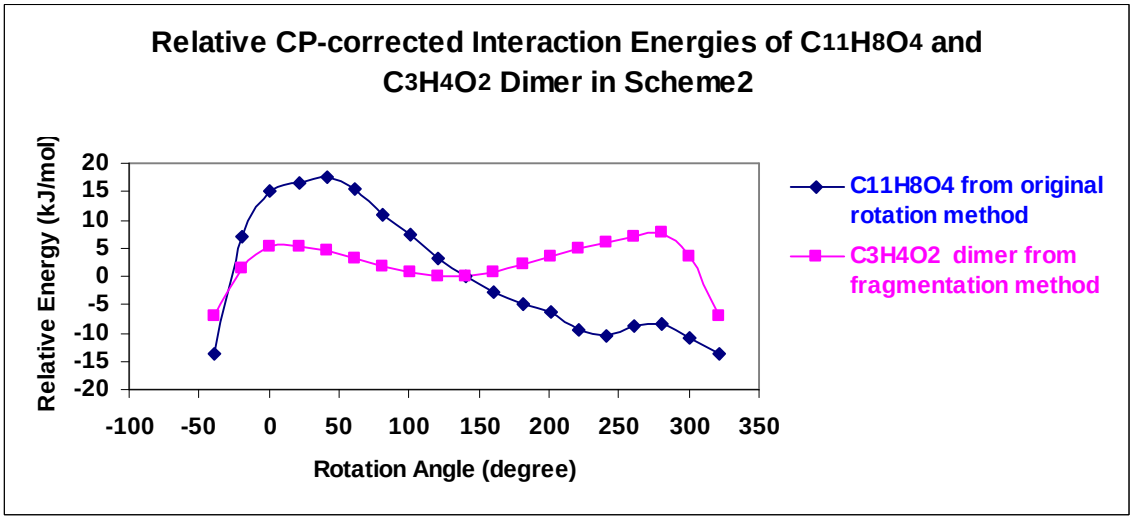


Figure 3.17 The CP-corrected relative interaction energies of $C_{11}H_8O_4$ and $C_3H_4O_2$ dimer in scheme2 (removal of central 5 carbon atoms) using original rotation method. The energy at rotation angle = 141˚ was used as the reference for the relative energies.

Figure 3.17 shows the CP-corrected relative interaction energies of $C_{11}H_8O_4$ and the $C_3H_4O_2$ dimer in scheme2 using the original rotation method. From the graph we can see that both curves have two maxima. The rotation curve's left maximum is much higher than the right one while the heights of the fragmentation curve's maxima are nearly the same.

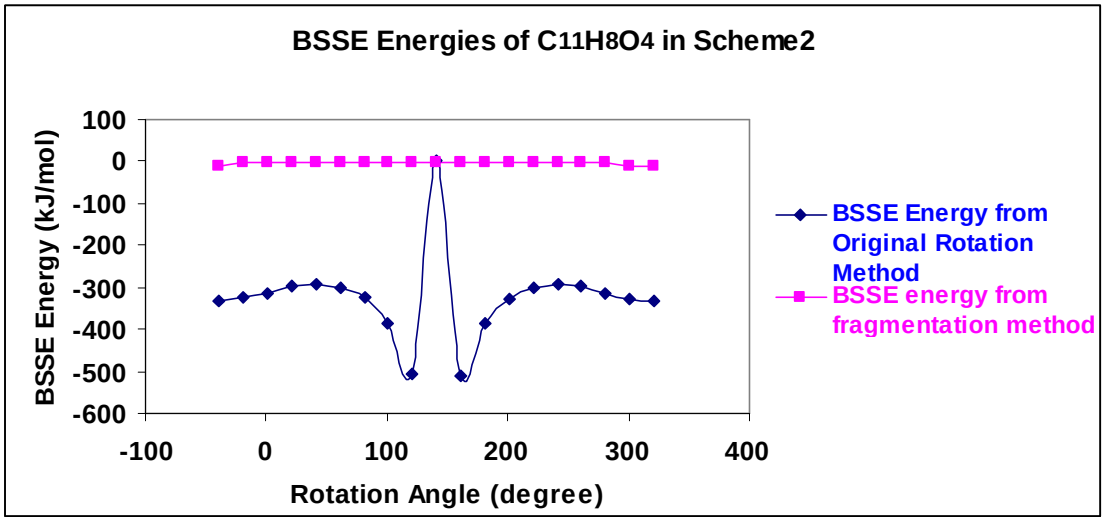


Figure 3.18 BSSE energies of $C_{11}H_8O_4$ in scheme2
(removal of central 5 carbon atoms) using original rotation method

Figure 3.18 shows the BSSE energies of $C_{11}H_8O_4$ from the original rotation method and the fragmentation method in scheme2. The BSSE energies are similar to those obtained in scheme1. Compared to scheme1, the BSSE energies from the fragmentation method in scheme2 are less negative, varying between -10 kJ/mol and 0 kJ/mol. The reason for this could be that with the interacting parts further apart, the superposition effect gets lower. The BSSE energies from the rotation method stay the same as those in Figure 3.11 because there is no change for this method.

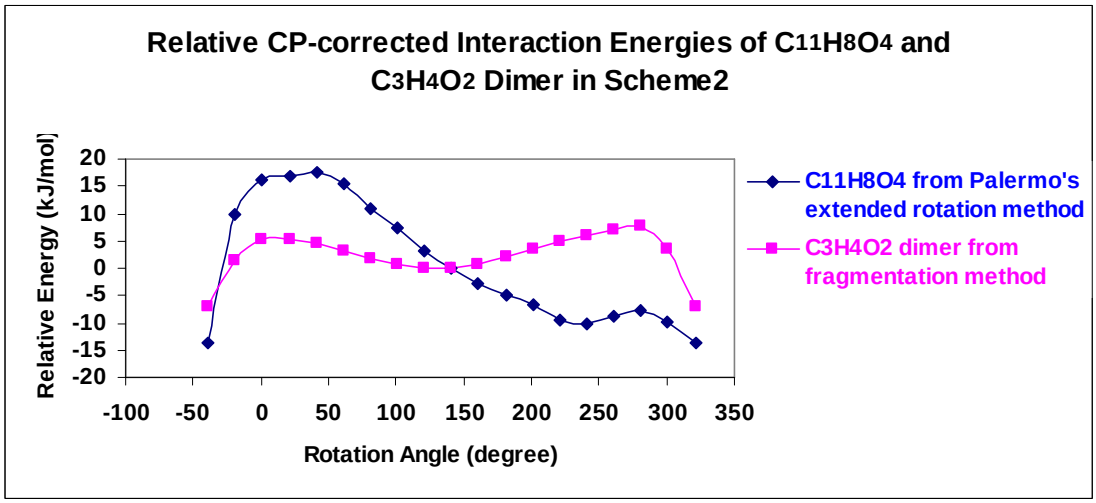


Figure 3.19 The relative CP-corrected interaction energies of $C_{11}H_8O_4$ and $C_3H_4O_2$ dimer in scheme2 (removal of central 5 carbon atoms) using Palermo's extended rotation method. The energy at rotation angle = 141˚ was used as the reference for the relative energies.

Figure 3.19 shows the relative CP-corrected interaction energies of $C_{11}H_8O_4$ and the $C_3H_4O_2$ dimer in scheme2 using Palermo's extended rotation method. The curve from the fragmentation method is the same as in Figure 3.17 while there is a slight difference between the rotation method curve and the one in Figure 3.17.

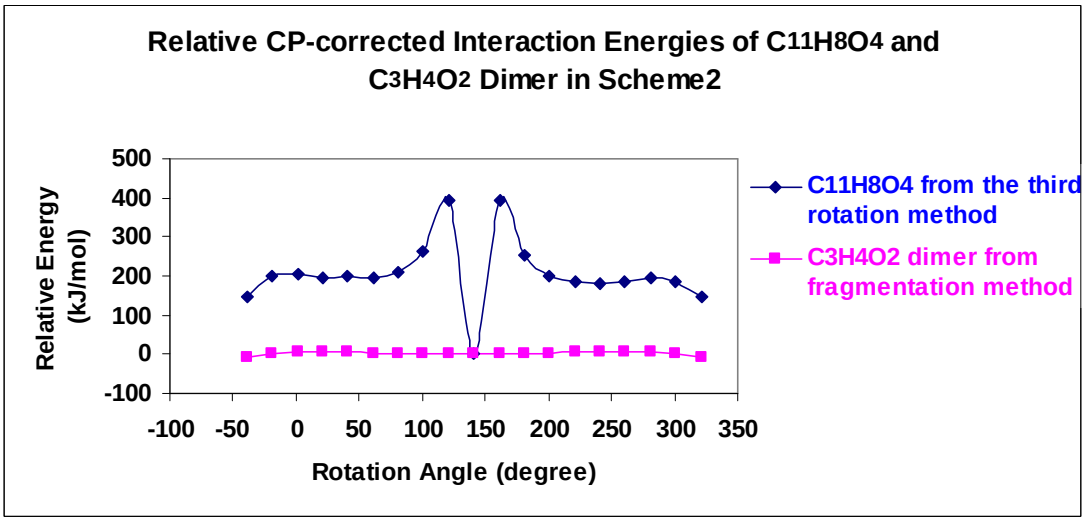


Figure 3.20 The relative CP-corrected interaction energies of $C_{11}H_8O_4$ and $C_3H_4O_2$ dimer in scheme2 (removal of central 5 carbon atoms) using the third rotation method. The energy at rotation angle = 141˚ was used as the reference for the relative energies.

Figure 3.20 shows the relative CP-corrected interaction energies of $C_{11}H_8O_4$ and $C_5H_4O_2$ in scheme2 using the third rotation method and the fragmentation method at each rotation step. The dimer energy curve from the fragmentation method is the same as that in Figure 3.19. The energy curve from the third rotation method displays a strange shape with the left part increasing and the right part decreasing. At the reference position in the middle the energy suddenly drops to 0.

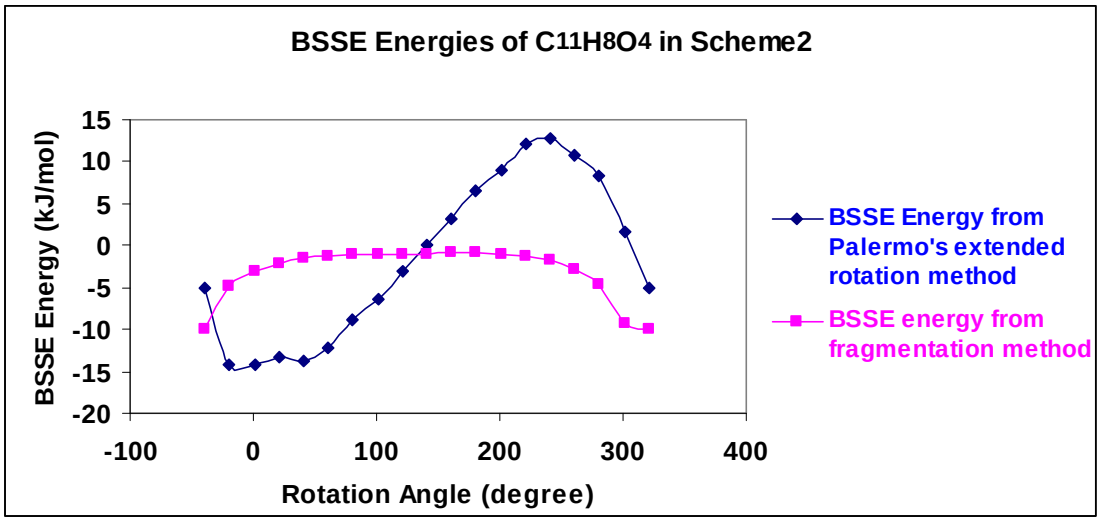


Figure 3.21 BSSE energies of $C_{11}H_8O_4$ in scheme2 (removal of central 5 carbon atoms) using Palermo's extended rotation method

Figure 3.21 shows the BSSE energies of $C_{11}H_8O_4$ in scheme2 using Palermo's extended rotation method. The BSSE energies from the fragmentation method vary between -11 kJ/mol and 0 kJ/mol with a flat curve shape. The BSSE energies from the extended rotation method vary between -15 kJ/mol and 13 kJ/mol and the curve has its left part below 0 kJ/mol and its right part above 0 kJ/mol, with a similar shape to the corresponding curve for $C_{11}H_8O_4$ in scheme1. The fact that some BSSE energies are positive means the extended rotation method gives wrong BSSE energies.

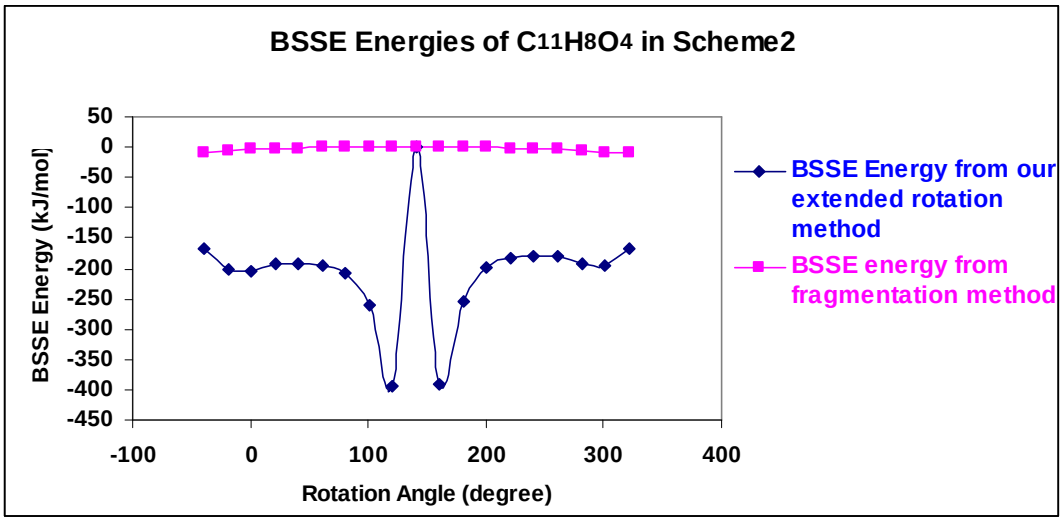


Figure 3.22 BSSE energies of $C_{11}H_8O_4$ in scheme2 (removal of central 5 carbon atoms) using the third rotation method

Figure 3.22 shows the BSSE energies of $C_{11}H_8O_4$ in scheme2 using the third rotation method. The curves are very similar to those in Figure 3.18, except that the BSSE energies calculated with the rotation method are smaller.

In scheme2, the relative energies of $C_{11}H_8O_4$ and $C_3H_4O_2$ dimer are very similar to each other, indicating it is a good model for the fragmentation method. However, Figure 3.17 shows an obvious difference between the CP-corrected energy curves from the rotation method and the fragmentation method. The BSSE energies from the rotation method appear very negative, as shown in Figure 3.18. With Palermo's extended rotation method, the relative interaction energies and CP-corrected interaction energies of $C_{11}H_8O_4$ show similar curve shapes. But the BSSE energy curve obtained from the extended rotation method has its right part above 0 kJ/mol, which is wrong. The BSSE energy curve obtained using the third rotation method proposed by us shows a similar shape as in the original rotation method, but with somewhat smaller BSSE values.

3.5 Conclusion

The intramolecular BSSE was investigated for two model systems (propanediol and $C_{11}H_8O_4$) using two different methods: the fragmentation method and the rotation method. The application of the rotation method was further studied using Palermo's extended rotation method and a third rotation method proposed by us. The results assess their reliability and accuracy.

For the propanediol model system, the CP-corrected interaction energy curves of propanediol and the dimer created by removing the central CH_2 group (see Fig. 3.5) were very similar to the interaction energy curves without CP-correction (see Fig. 3.7). The BSSE energy curves calculated by the fragmentation method and the rotation method were remarkably different. In addition, the large difference between the two relative energy curves of propanediol indicate that propanediol is not a good model to test the fragmentation method.

For the $C_{11}H_8O_4$ model system, two schemes were used: one was to generate the dimer by removing the central CH_2 group; the other was to generate the dimer by removing the central five carbon atoms.

In the first scheme, the relative energy curves of $C_{11}H_8O_4$ and the dimer generated by removing the central CH_2 group are very different from each other (see Fig. 3.10), like for the propanediol model system, meaning that the scheme is not good for the fragmentation method either. The BSSE energies from the rotation method are very negative, with the value jumping to 0 kJ/mol in the middle of the curve (see Fig. 3.11), showing a prominent difference to the BSSE curve from the fragmentation method. With Palermo's extended rotation method, the BSSE values vary within a reasonable range, but with the right part of the curve above 0 kJ/mol (see Fig. 3.13), indicating it is a wrong method for the BSSE energy calculation. By applying our third rotation method, the BSSE energy curve is similar to that obtained with the original rotation method, still having exceptionally negative values (see Fig. 3.14).

In the second scheme, the relative energy curve of $C_{11}H_8O_4$ shows good resemblance to that of the $C_3H_4O_2$ dimer (see Fig. 3.15), indicating it is the best scheme for the fragmentation method. However, the BSSE energies from the rotation method still present very negative values with a similar shape to that in scheme1 (see Fig. 3.17). With Palermo's extended rotation method, the same problem occurs that the right part of the BSSE energy curve is positive (see Fig. 3.19). With the third rotation method proposed by us, the BSSE energy curve resembles that in the original rotation method, indicating that it is still not a good method to obtain accurate BSSE energies.

In summary, the results show that the intramolecular BSSE cannot be corrected by the rotation methods discussed above.

Of course, the fragmentation method is also not an exact method to eliminate BSSE, as broken bonds may need to be saturated with additional hydrogen atoms and some atoms around the broken bond may need to be removed to avoid overlapping atoms. One way to verify the accuracy of the fragmentation method may be by comparison to results from plane-wave DFT methods (which do not produce BSSE) with a sufficiently high energy cut-off and a sufficiently large unit cell to avoid intermolecular interactions. We have not attempted this in this work.

Chapter 4 Performance of the M06-L Density Functional for a Folded Tyr-Gly Conformer

4.1 Abstract

The routine B3LYP and MP2 methods do not always yield qualitatively correct structures for flexible organic systems containing π systems, particularly when small to medium-sized basis sets are used. This is due to large basis set superposition errors in the MP2 calculations and the lack of dispersion in the B3LYP calculations. Here we study the ability of several recently developed DFT methods to predict the conformation of a conformer of the Tyr-Gly dipeptide, for which B3LYP/6-31+G(d) and MP2/6-31+G(d) geometry optimizations give strikingly different structures. B3LYP yields an open structure while MP2 yields a more folded structure. The Tyr-Gly geometry was optimized with fixed values (in the range from 3.0 to 8.0 Å) of a distance that controls the degree of foldedness of the structure, using three levels of theory: B3LYP/6-31+G(d), M06-L/6-31G(d) and M05-2X/6-31+G(d). Afterwards, a series of DFT methods were employed to calculate the single-point energies and obtain the transition energy profiles. The local meta functional M06-L is found to yield results in close agreement with the reference CCSD(T) profile, which manifests itself to be a very promising method to investigate the potential energy surface of small peptides containing aromatic residues [80].

4.2 Introduction

To predict the most stable conformations of a peptide is a significant and complex problem. This topic's importance arises from its strong relevance to the protein

folding problem. Even though the protein folding process occurs in solution, a good understanding of the conformational preferences of gas-phase peptides is still worth investigating. The interior of proteins contain mainly hydrophobic side chains and are therefore shielded from solvent interactions containing mainly hydrophobic side chains. Therefore, using gas phase peptides as models for the hydrophobic core may be more appropriate than considering fully hydrated peptides only. In addition, to find out the influence of solvent on the three-dimensional structure of peptides and proteins, an understanding of the structural constraints resulting from the non-covalent intramolecular interactions is required. This can only be obtained by investigating isolated gas phase molecules.

It is difficult to study conformations of peptides due to their extensive flexibility, which means that a great deal of possible conformers have to be to considered to find the most stable ones. By applying geometry optimization and energy calculations of all possible conformers at a high level of theory such as electronic structure methods, the problem can in principle be solved satisfactorily. Nevertheless, the large computational cost of these theories is an obstacle. In order to overcome this problem, alternative computationally cheaper methods must be considered such as the density functional theories.

During the past decades, density functional theory (DFT) has emerged as a powerful methodology for the simulation of chemical systems. It is built based on the premise that the energy of an electronic system can be defined in terms of its electron probability density, which means the electronic energy is regarded as a functional of the electron density, in the sense that there exists a one-to-one correspondence between the electron density of a system and its energy. One of the most popular density functionals is the B3LYP functional [46-47, 81], which is available in most commonly used quantum chemistry program packages. About 80% of all occurrences of the names of functionals in journal titles and abstracts over 1990-2006 were attributed to B3LYP [51]. Thus, an extensive experience in using this functional has

been gained by the computational chemistry community. However, given that B3LYP and many other DFT methods are unable to describe London dispersion interactions [82-84], the simplest correlated ab initio method MP2 (second-order Møller-Plesset perturbation theory) is often utilized instead in situations where dispersion is thought to be important (such as interactions with π-electron clouds). In previous studies on gas-phase neurotransmitters and small peptides, a DFT method was used to optimize the geometry of the target, whereas the relative energies were evaluated using MP2 single-point calculations [85-87]. This approach was also used in a recent study on 20 conformers of the tyrosine-glycine (Tyr-Gly) dipeptide that were identified as stable conformers according to the hierarchical selection scheme [88]. In this study, MP2 optimization was also performed on a selection of the Tyr-Gly conformers using the same basis set, 6-31+G(d), as in the B3LYP optimizations. Surprisingly, for some of the conformers, B3LYP and MP2 gave very different optimized geometries. The MP2 single-point calculations (using B3LYP-optimized geometries) showed that the six most stable conformers all have a folded “book” conformation. However, B3LYP predicted the extended conformations to be more stable. After optimization using the MP2 method, the six book conformers increased their degree of foldedness dramatically. The fourth and sixth conformers, dubbed “book4” and “book6” gave the largest difference between B3LYP and MP2 optimized geometries. The B3LYP and MP2 optimized structures of book6 are shown in Figure 4.1.

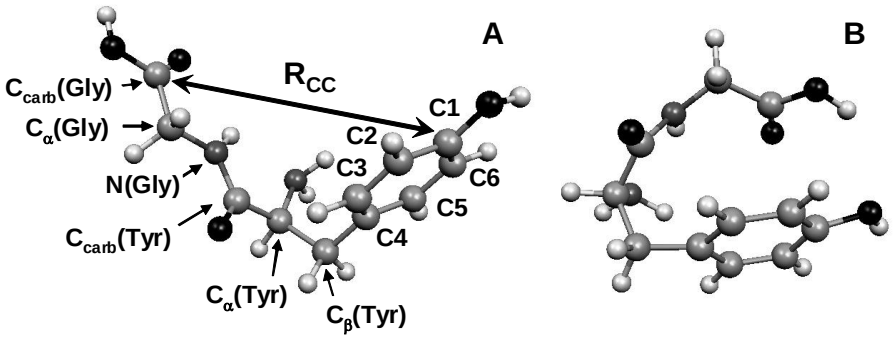


Figure 4.1: B3LYP /6-31+G(d) and MP2/6-31+G(d) optimized geometries of the Tyr-Gly conformer book6. A: B3LYP structure; B: MP2 structure.

The different geometries derived by B3LYP and MP2 could be interpreted by two probable explanations: Missing dispersion interactions in the B3LYP calculations and large intramolecular basis set superposition error (BSSE) effects in the MP2 calculations. These deficiencies can be solved by two methods: Including dispersion in the B3LYP calculations or reducing the BSSE in the MP2 calculations. Furthermore, other more accurate methods can be used to overcome this problem. One way to solve the geometry difference problems is to use a local variant of the MP2 method which reduces BSSE values significantly compared to classic MP2 [17-18, 89-97]. The other way is to employ density functionals that do describe dispersion interactions such as the M0x (x=5, 6, 8) series of functionals [39, 59-60, 98-99] from the Truhlar group.

In this work, we take the M06-L [100] functional as our method, which is a local (non-hybrid) functional and computationally more efficient than the meta-hybrid functionals in the M0x families. For one of the problematic Tyr-Gly conformers, book4, M06-L was shown to obtain much improved results compared to B3LYP [8]. Also, a lot of good performances of M06-L have been observed in other researches for a range of systems, such as zeolities [99], magnetic properties [101-102], water

clusters [103-104] and non-covalent interactions [40, 60, 105-106].

We take book6 conformer as the research target. Transition energy profiles have been created from the B3LYP to the MP2 conformer with the Tyr-Gly structure optimized at fixed values of the R_{CC} distance which controls the degree of foldedness of book6 (see Figure 4.1 for its definition). A reference profile, computed with df-LCCSD(T0)/aug-cc-pVTZ, was taken from previous work [6]. Local correlation methods are known to produce much smaller BSSE values than their canonical counterparts [97]. It was shown that df-LMP2/aug-cc-pVTZ yields small BSSE values for this Tyr-Gly conformer (less than 2 kJ/mol). It is therefore expected that the df-LCCSD(T0) reference profile is not significantly affected by BSSE. It is found that the B3LYP minimum is essentially correct, which leads to the conclusion that the closed book structure obtained by MP2 calculations is due to an artificial attraction between the tyrosine and glycine residues resulting from the intramolecular BSSE.

Through the investigation of the transition energy profiles, M06-L is shown to yield results that excellently agree with the reference profile. Therefore, we consider it a promising functional for investigating the potential energy surface of small aromatic peptides.

4.3 Methodology

First of all, we optimized the Tyr-Gly geometry at fixed values of the distance between the C_{carb}(Gly) and C1 atoms (R_{CC}), in the range from 3.0 to 8.0 Å (see Figure 4.1 for the atom labelling). Subsequently optimization procedures were carried out using three levels of theory: B3LYP/6-31+G(d), M05-2X/6-31+G(d) and M06-L/6-31G(d). The first one is the most widely used density functional [46-47, 81]. M05-2X [39] and M06-L [60] are from the M05 and M06 families of density functionals developed by Zhao, Truhlar and Schultz, which are becoming more and more popular in modern computational chemistry. Given the fact that we are looking

for a low-cost method to predict the relative stabilities of peptide conformers reliably, and that the inclusion of diffuse functions in the basis set would be computationally expensive in DFT calculations, we employed the 6-31G(d) basis set in the M06-L calculations.

The transition energy profiles were subsequently calculated by employing single-point calculations for the three sets of partially optimized geometries at several levels of theory, including M05-2X/6-31+G(d), M06-L with the 6-31G(d) and 6-31+G(d) basis sets, B3LYP/6-31+G(d), B3LYP-D with the 6-31+G(d) and aug-cc-pVDZ basis sets, mPW2-PLYP/aug-cc-pVDZ and mPW2-PLYP-D/aug-cc-pVDZ, which is a double hybrid functional containing a portion of MP2 correlation energy computed using the hybrid DFT orbitals [107]. For the B3LYP-D and mPW2-PLYP-D methods, the postfix "-D" means the functionals are augmented with an empirical dispersion term [108-109]. A previous study found that mPW2-PLYP-D with the basis set aug-cc-pVDZ gave the best performance for demonstrating the three minima along the ϕ_{Gly} torsional profile of Tyr-Gly conformer book4 [8].

For the B3LYP/6-31+G(d) optimized structures, M05-2X/6-31+G(d), M06-L/6-31G(d) and M06-L/6-31+G(d) single-point calculations were performed using NWChem [75]. For the M06-L/6-31G(d) optimized structures, B3LYP/6-31+G(d) and M05-2X/6-31+G(d) calculations were done with NWChem. For the M05-2X/6-31+G(d) optimized structures, B3LYP/6-31+G(d), M06-L/6-31G(d), M06-L/6-31+G(d) calculations were done with NWChem while B3LYP-D/6-31+G(d), B3LYP-D/aug-cc-pVDZ, mPW2-PLYP/aug-cc-pVDZ and mPW2-PLYP-D/aug-cc-pVDZ calculations were performed using ORCA [76]. We took the profile calculated with df-LCCSD(T0) (density-fitting local coupled cluster theory with single, double, and perturbative local triple calculations) [90, 110-112] and the aug-cc-pVTZ basis set, taken from Ref. [6] as the reference profile. The "xfine" grid (125 radial and 1454 angular shells) was used in the NWChem

computations while “grid 6” (default GaussChebyshev radial grid coupled with 590 angular Lebedev points) was used in the ORCA computations.

4.4 Results

4.4.1 Performance of density functionals

In previous work [6], the energy profiles of Tyr-Gly conformer book6 for transition between the MP2 and B3LYP minima were computed using partially optimized structures (at fixed values of R_{CC}) computed with B3LYP/6-31+G(d). However, the partially optimized structures may differ depending on the level of theory employed for the geometry optimization. For the Tyr-Gly book4 conformer, geometry optimizations at fixed values of the ϕ_{Gly} Ramachandran angle obtained more compact structures with MP2 than with B3LYP, particularly in the ϕ_{Gly} region of 120-130º [8]. As the intramolecular BSSE is particularly large in this region [5], the more compact structures predicted by MP2 are presumably due to a combination of intramolecular BSSE in the MP2 calculations and missing dispersion in the B3LYP calculations.

The M05-2X functional, which describes dispersion but does not suffer from the large BSSE values that plague MP2 calculations, gave geometries that are more compact than the B3LYP geometries, but do not show the sharp increase in compactness around ϕ_{Gly} = 120-130º, and was therefore deemed to be the most accurate of the three sets of geometries. As the shape of the transition energy profiles depends on the set of partially optimized geometries employed [8], we investigated the effect of the choice of geometry optimization method on the transition energy profiles of book6.

First, three sets of optimized geometries have been employed to assess the

performance of the different density functionals, as shown in Figure 4.2, 4.3 and 4.4. Also shown in the figures is the df-LCCSD(T0)/aug-cc-pVTZ reference profile, obtained using B3LYP-optimized geometries [6].

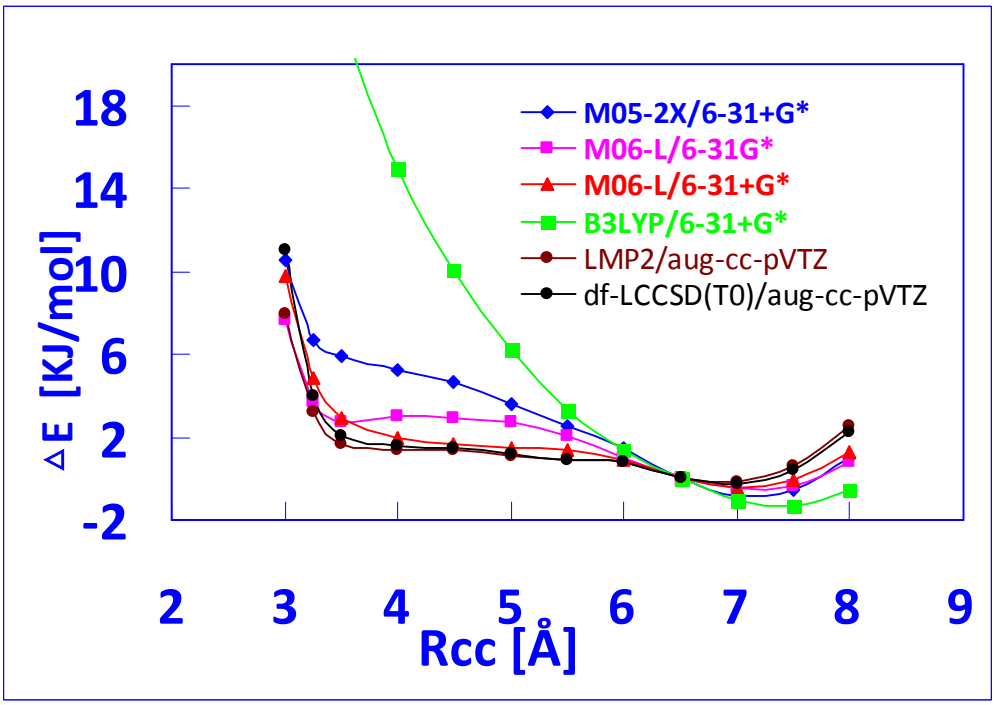


Figure 4.2 Transition energy profiles computed with different DFT methods using geometries optimized with B3LYP/6-31+G(d) at fixed R_{CC} distances. The reference df-LCCSD(T0) profile, taken from Ref. [6], is obtained using geometries optimized with B3LYP/6-31+G(d) at fixed R_{CC} distances. The energy at R_{CC} = 6.5 Å was used as the reference for the relative energies ΔE.

The transition energy profiles calculated with different DFT methods using geometries optimized with B3LYP/6-31+G(d) at fixed R_{CC} distances are shown in Figure 4.2. It can be seen that the df-LCCSD(T0) profile displays a minimum at R_{CC} = 7.0 Å with a flat part between about 4 and 6 Å. Also shown in Figure 4.2 is an LMP2/aug-cc-pVTZ curve taken from Ref. [6]. The LMP2/aug-cc-pVTZ curve resembles the reference curve very well. There is only one minimum in the M05-2X/6-31+G(d) and M06-L/6-31+G(d) transition energy profiles. The M06-L/6-31G(d) profile exhibits a shallow minimum at about 3.5 Å. This could result

from the BSSE which is expected to be still present in the DFT calculations.

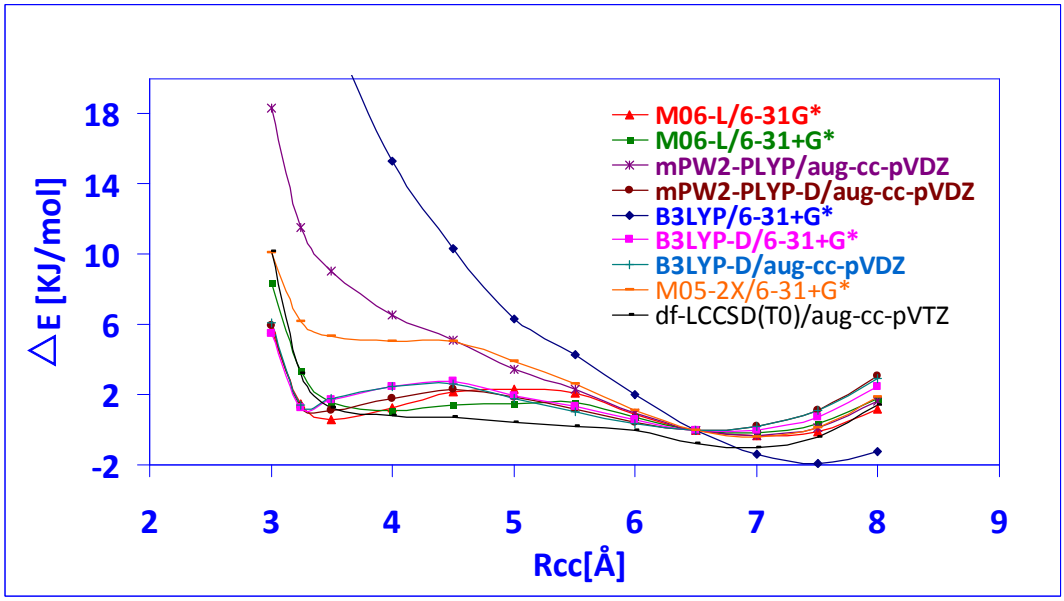


Figure 4.3: Transition energy profiles computed with different DFT methods using geometries optimized with M05-2X/6-31+G(d) at fixed R_{CC} distances. The reference df-LCCSD(T0) profile, taken from Ref. [6], is obtained using geometries optimized with B3LYP/6-31+G(d) at fixed R_{CC} distances. The energy at R_{CC} = 6.5 Å was used as the reference for the relative energies ΔE.

Figure 4.3 shows the transition energy profiles calculated with different DFT methods using geometries optimized with M05-2X/6-31+G(d) at fixed R_{CC} distances. The B3LYP curve displays a steep rise at short distances with a minimum at a slightly larger distance (7.5 Å). The possible reason could be the lack of intramolecular dispersion. Only a minor improvement is achieved by the mPW2-PLYP double hybrid functional as compared to B3LYP although it includes a percentage of MP2 correlation energy and is expected to recover part of the intramolecular dispersion energy. The M05-2X functional does not improve the results much either. It is relatively flat in the 3.5-4.5 Å region and rises steeply at short distances proving underestimated dispersion. But with an empirical dispersion term added to B3LYP

and mPW2-PLYP, profiles in much better agreement with the reference df-LCCSD(T0) curve are obtained, which means that the unsatisfactory curves derived by B3LYP and mPW2-PLYP are essentially caused by lack of dispersion (although the empirical dispersion term cannot be interpreted as pure dispersion energy [108, 113]). Also the M06-L method derives the transition energy profile in good agreement with the reference profile, which is especially impressive given the local nature of this functional and the small basis set employed. One point needed to be mentioned is that both the two DFT-D profiles and the two M06-L profiles depict a shallow minimum at about 3.5 Å, which is different from the reference profile. This can be explained by the different geometries employed in the generation of the reference profile.

4.4.2. Tyr-Gly geometries along the energy profile

To study the effect of the choice of the geometry optimization method, we computed the profiles using the same method using the three sets of geometries.

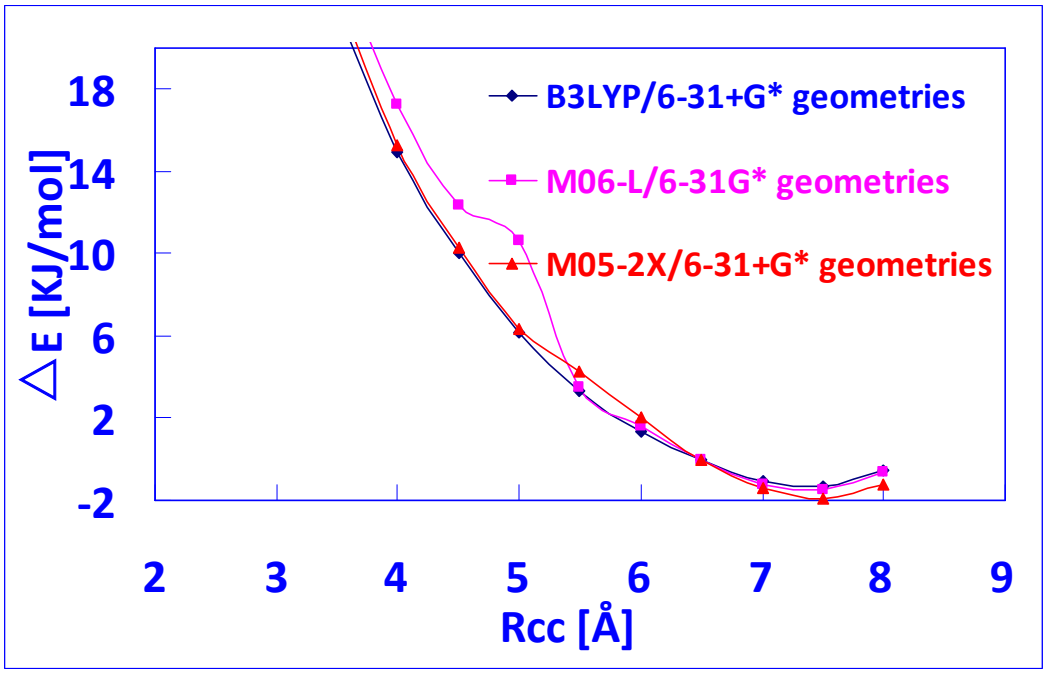


Figure 4.4: Transition energy profiles computed with B3LYP/6-31+G(d) using geometries optimized with M06-L/6-31G(d), M05-2X/6-31+G(d) and B3LYP/6-31+G(d), at fixed R_{CC} distances. The energy at R_{CC} = 6.5 Å was used as the reference for the relative energies ΔE.

First of all, we take B3LYP/6-31+G(d) as the method to calculate the transition energy profiles using the three sets of geometries. Figure 4.4 shows that the B3LYP/6-31+G(d) and M05-2X/6-31+G(d) geometries yield curves that are very similar to each other. The curve obtained using the M06-L/6-31G(d) geometries has a sudden change in the region between about 4.5 and 5.5 Å.

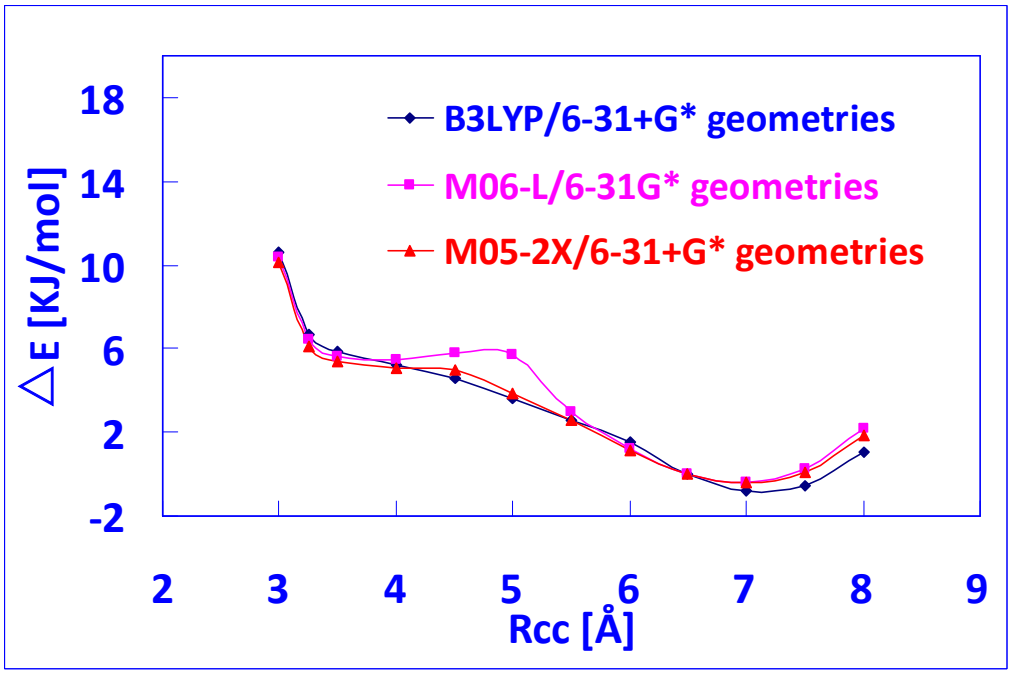


Figure 4.5: Transition energy profiles computed with M05-2X/6-31+G(d) using geometries optimized with M06-L/6-31G(d), M05-2X/6-31+G(d) and B3LYP/6-31+G(d), at fixed R_{CC} distances. The energy at R_{CC} = 6.5 Å was used as the reference for the relative energies ΔE.

Figure 4.5 shows the transition energy profiles computed with M05-2X/6-31+G(d) using geometries optimized with M06-L/6-31G(d), M05-2X/6-31+G(d) and B3LYP/6-31+G(d), at fixed R_{CC} distances. The three M05-2X transition energy curves are also very similar but again the curve obtained using the M06-L geometries lies above the other two in the region between 4.5 and 5.5 Å.

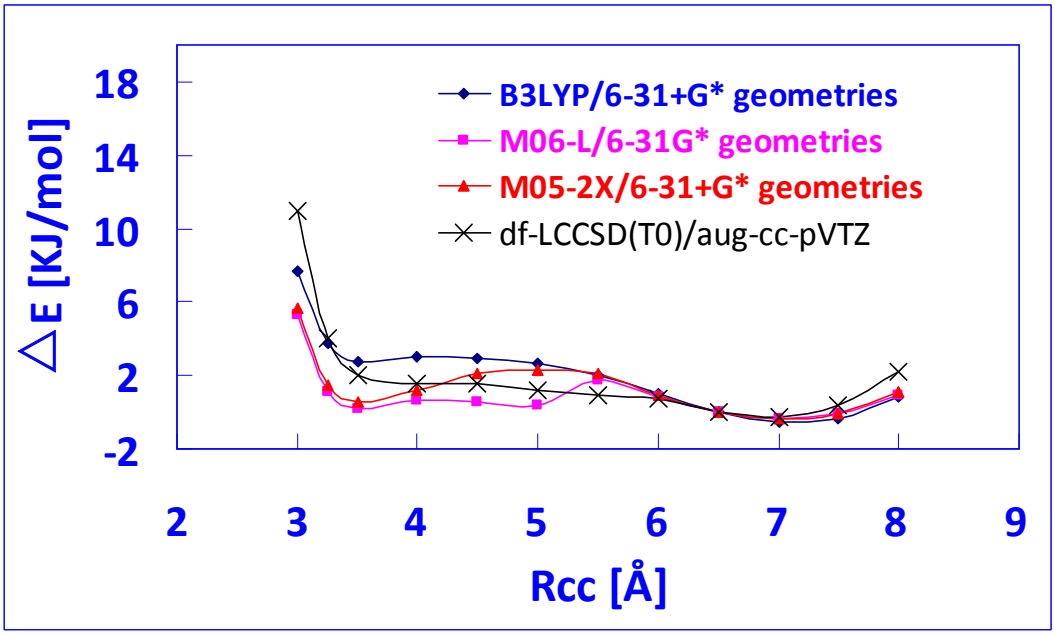


Figure 4.6: Transition energy profiles computed with M06-L/6-31G(d) using geometries optimized with M06-L/6-31G(d), M05-2X/6-31+G(d) and B3LYP/6-31+G(d), at fixed R_{CC} distances. The reference df-LCCSD(T0) profile, computed using the B3LYP geometries, is taken from Ref. [6]. The energy at R_{CC} = 6.5 Å was used as the reference for the relative energies ΔE.

The three M06-L/6-31G(d) transition energy profiles computed using the three sets of geometries (obtained at the B3LYP/6-31+G(d), M05-2X/6-31+G(d) and M06-L/6-31G(d) levels of theory) are shown in Figure 4.6. The sudden change still exist in the M06-L/6-31G(d) curve. Overall, the three profiles are similar to each other, though there are some differences prominently in the 3-5 Å range.

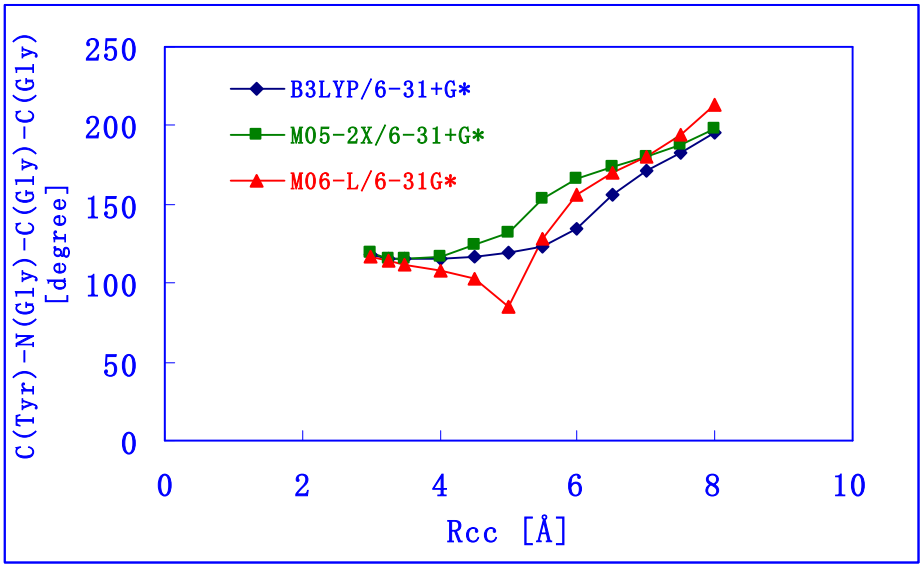


Figure 4.7: The ϕ_{Gly} torsion angle as a function of R_{CC} in the partially optimized structures obtained with M06-L/6-31G(d), M05-2X/6-31+G(d) and B3LYP/6-31+G(d).

Figure 4.7 shows that the M06-L/6-31G(d) optimized geometries mainly differ in the orientation of the C-terminus compared to the other two sets of optimized geometries, which can be quantified by the ϕ_{Gly} Ramachandran angle (the C_{carb}(Tyr)-N(Gly)-C_{α}(Gly)-C_{carb}(Gly) torsion angle). In the R_{CC} = 4-5.25 Å region, the ϕ_{Gly} value is prominently smaller in the M06-L curve compared to the M05-2X and B3LYP curves. Particularly, at R_{CC} = 5.0 Å, M06-L/6-31G(d) gives a value of ϕ_{Gly} = 86º, whereas M05-2X/6-31+G(d) predicts a value of 133º.

The improved version of the M05-2X functional, namely the M06-2X functional [98], was also employed to optimize the geometry using the 6-31+G(d) basis set. Two geometries were derived at R_{CC} = 5.0 Å: one was similar to the M05-2X structure (ϕ_{Gly} = 136 º), the other was similar to the M06-L structure (ϕ_{Gly} = -78º). The structure with the ϕ_{Gly} value of 136º was about 0.11 kJ/mol lower in energy. A structure with a ϕ_{Gly} value of 115º was derived by M06-L/6-31+G(d) geometry optimization at fixed R_{CC} = 5.0 Å. Therefore, it appears that it is the functional and basis set applied that affects the exact conformation of the C-terminus, and there is some possibility that the M06-L yielded geometry may be correct. In addition, because the differences in

geometry and energy are small, further investigation has not been employed.

4.5 Conclusions

Herein, we present a study of recently developed DFT methods for the prediction of the conformations of a conformer of the Tyr-Gly dipeptide, namely book6, for which B3LYP/6-31+G(d) and MP2/6-31+G(d) geometry optimizations give very different geometries.

The study employed the recently developed DFT methods: M06-L and M05-2X as well as the most popular B3LYP functional. The same procedure could be used to test other DFT methods.

Potential energy profiles for transition from the MP2 to the B3LYP structure have been computed using geometries optimized using M06-L/6-31G(d), M05-2X/6-31+G(d) and B3LYP/6-31+G(d) at a range of fixed values of the distance (R_{CC}) between two carbon atoms located at the outer ends of the molecule. M06-L/6-31G(d) and M06-L/6-31+G(d) gives profiles excellently resembling the reference df-LCCSD(T0)/aug-cc-pVTZ profile, while with the B3LYP, M05-2X and mPW2-PLYP functionals, the profiles rise too steeply at short distances. In addition, with an empirical dispersion term added into the B3LYP and mPW2-PLYP functionals, these two methods generate profiles closely resembling the reference profile, proving the underestimation of intramolecular dispersion in the performance of the B3LYP, M05-2X and mPW2-PLYP functionals.

One particular point needed to be mentioned is that the geometries optimized with M06-L at fixed R_{CC} values show a somewhat differently oriented C-terminus in the 4.00-5.25 Å region compared to B3LYP and M05-2X. Nevertheless, given that this slight change is at a very flat part of the potential energy surface and the comparison with geometries optimized with the M06-2X functional indicate that both C-terminal arrangements may be feasible, we consider M06-L a practical applicable method for investigating potential energy surfaces.

In conclusion, the M06-L functional has proved itself as a very promising method in the investigation of potential energy surface of small peptides containing aromatic residues, with the results even more accurate than the meta-hybrid functional M05-2X and double hybrid functional mPW2-PLYP which are computationally more expensive. That is why to use the M06-L functional to perform a more accurate scan of the Tyr-Gly potential energy surface than the B3LYP-based scan from former studies (see Chapter 5).

Chapter 5 Predicting the Structure of Flexible Peptides

5.1 Abstract

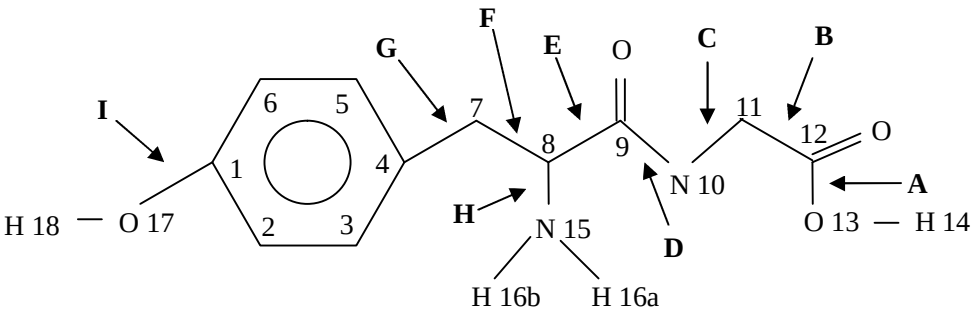
Given the fact that B3LYP and MP2 are not good enough to predict the structures of the tyrosine-glycine dipeptide from the experience of previous works [5-6], we applied a more suitable level of theory, M06-L from the Minnesota density functional theories, to investigate this problem. In this work, 108 potential conformers were created with a Fortran program using a recursive procedure taken from previous work [88]. Two schemes were employed: one was to optimize the first 30 conformers ranked according to the M06-L/6-31G+(d) single-point energies; the other was to optimize all of the 108 conformers. The geometry optimizations were done using M06-L/6-31G(d) and M05-2X/6-31+G(d). The most stable conformers were compared to the 20 stable conformers found by B3LYP geometry optimizations in the previous study [88]. Both schemes found 10 conformers similar to one of the B3LYP stable conformers, as well as several newly found conformers. Relative transition energy profiles of a missing B3LYP stable conformer were calculated using different levels of theory. The results showed that the missing of some of the B3LYP conformers may be due to the lack of dispersion in B3LYP theory.

5.2 Introduction

Peptides are short polymers of amino acids linked by peptide bonds, which have the same chemical structure as proteins, but with shorter length. The shortest peptides are dipeptides which consist of two amino acids joined by a single peptide bond, which therefore have an amino end and a carboxylic end. One of the most prominent features of peptides is their extensive flexibility. This makes it difficult to perform

conformational studies, because a large number of possible conformers have to be considered to identify the most stable conformers, preventing the routine use of high-accuracy but computationally expensive electronic structure methods. However, with the development of computer architecture, high-accuracy computational methods are becoming feasible to study peptides, such as the prevailing density functional theory.

In a previous study by Toroz and van Mourik [88], two different methodologies were explored to study the conformational features of the neutral (non-zwitterionic) dipeptide Tyr-Gly. In the first method, a ‘stepwise rotation’ procedure was employed, whereas in the second method, a ‘hierarchical selection’ scheme was performed.



A: 14-13-12-11	**D: 11-10-9-8**	**G: 8-7-4-5**
B: 13-12-11-10	**E: 10-9-8-7**	**H: 16-15-8-7**
C: 12-11-10-9	**F: 9-8-7-4**	**I: 18-17-1-2**

Figure 5.1: Atom labelling and definition of the dihedral angles

In the first methodology, named ‘stepwise rotation’, the main analysis was performed via a sequential variation of the nine dihedral angles shown in Figure 5.1, labelled (A-I). In the first step, dihedral angles A and B were varied simultaneously and all the possible conformers were optimized using HF/6-31+G* and subsequently

at the B3LYP/6-31+G* level of theory. In the second step, dihedral angles C-H were varied sequentially through six steps. In every step, B3LYP/6-31+G* was used to optimize the conformers and all of them were considered in the subsequent step except for the last stage. Because the dihedral angle 'I' only has two values, 0° and 180°, turning over the OH group is the same as flipping the whole aromatic ring, which means variation of dihedral I should not be necessary.

For the other methodology a procedure, dubbed the hierarchical selection method, was employed to study the conformational energy landscape of Tyr-Gly and to identify the most stable conformers of it, by using a combination of electronic structure methods of varying accuracy and computational cost. First of all, all possible conformers were created and sorted according to their number of intramolecular H-bonding interactions. Next, the conformers were sorted according to their single-point or optimized energies at increasingly more accurate levels of theory. Only the most stable conformers were taken through to the next level of calculation. Twenty conformers of Tyr-Gly were identified as the most stable ones by using B3LYP/6-31+G* as the geometry optimization method and computing single-point energies using MP2/6-31+G*. These 20 conformers are shown in Figure 5.2.

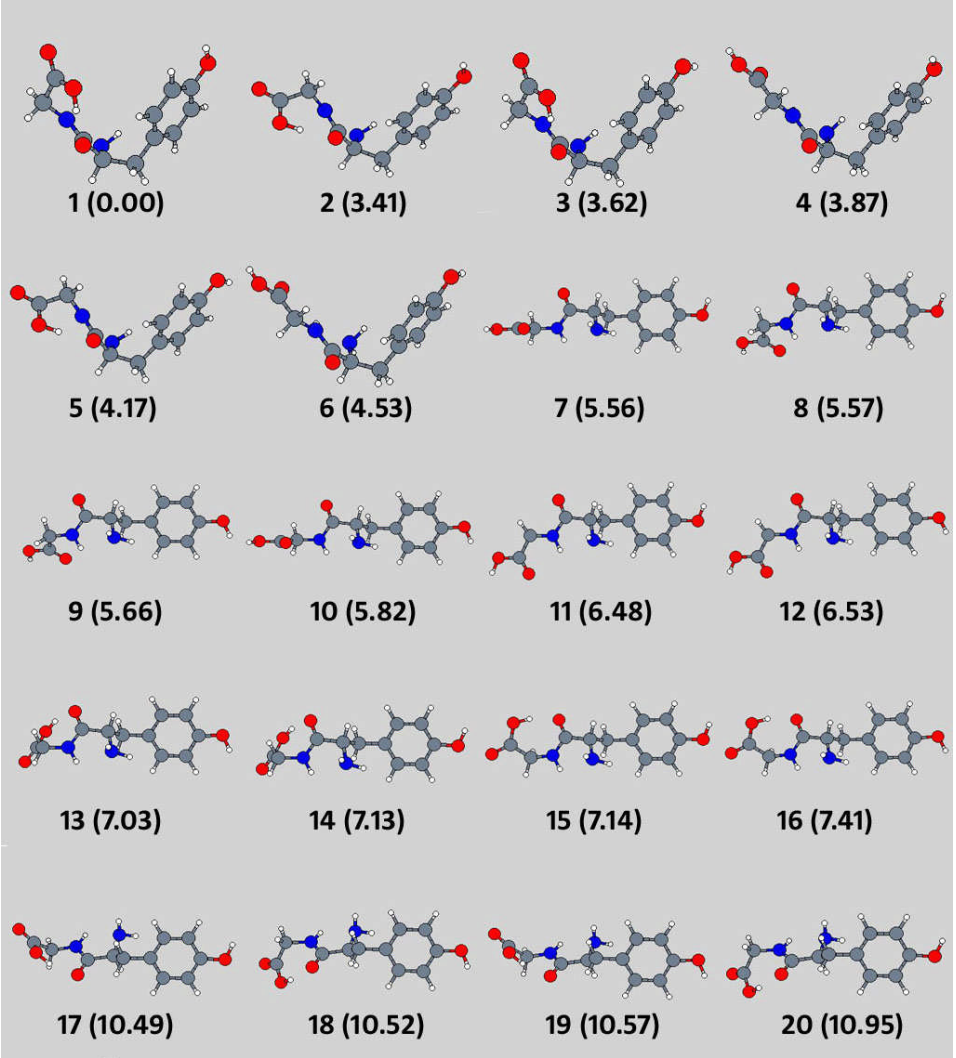


Figure 5.2: B3LYP/6-31+G(d) optimized structures of the 20 most stable Tyr-Gly conformers. Relative energies (from single-point MP2/6-31+G(d) calculations with inclusion of scaled (0.976) B3LYP/6-31+G(d) ZPEs) are given in kJ/mol.

However, it was observed that B3LYP/6-31+G* and MP2/6-31+G* calculations predict markedly different structures for some of the Tyr-Gly conformers. B3LYP geometry optimization gave an extended global minimum while single-point MP2 calculations significantly changed the order of stability of the conformers, yielding a folded global minimum. In addition, after full MP2 optimization, the book-type (when dihedral angle F < 90˚ in Figure 5.1) conformers increased their degree of foldedness even more prominently. A subsequent study by Holroyd and van Mourik [5]

investigated this in detail for one of the 20 most stable conformers, namely conformer 4, which only differed in the orientation of one torsion angle between B3LYP and MP2 optimized structures. Relaxed potential energy profiles for rotation around the bond corresponding to that torsion angle were created by geometry optimization at fixed values using step sizes of ~10-20°. The profiles were computed at various levels of theory such as HF/6-31+G*, B3LYP/6-31+G* and MP2/6-31+G*. After that, single-point energy calculations were performed at the MP2-optimized structures with df-MP2 (density fitting MP2 [114]), df-SCS-MP2 (density fitting spin-component-scaled MP2 [115]) and df-LMP2 (density fitting local MP2 [116]) with the aug-cc-pVnZ (n=D, T, Q) basis sets. It was suggested that the very different structures obtained by these two methods resulted from the dispersion underestimation by B3LYP and large BSSE in the MP2 calculations.

Another study by Shields and van Mourik [6] showed that for conformer 6, B3LYP/6-31+G* and MP2/6-31+G* optimizations also yielded strikingly different structures: MP2 predicted a folded “closed-book” conformer, while B3LYP predicted a more open conformer [6]. After further investigation using different levels of theory, including the local electron correlation methods LMP2 (local MP2) and LCCSD(T0) (local coupled cluster with single, double and noniterative local triple excitations) and large basis sets (aug-cc-pVnZ, n=D, T, Q), it was proven that the folded MP2 minimum is an artefact caused by large intramolecular BSSE effects in the MP2/6-31+G* calculations.

Therefore, it appears that B3LYP and MP2 (with small to medium-sized basis sets) are not good enough to predict the structures of Tyr-Gly and there could be additional stable conformers besides the 20 stable conformers already found. That is why we employed the advanced level of theory M06-L/6-31G(d), aiming to see the reliability of these conformers and if there are more stable conformers missed by the B3LYP method.

5.3 Methodology

In the investigation of predicting all the stable structures of Tyr-Gly, 108 potential conformers were created by a Fortran program using a recursive procedure taken from previous work [88]. A variety of different levels of computational methods were performed with the software package Gaussian (version 09) on the EaStCHEM Research Computing Facility.

5.3.1 Conformational creation procedure

The main aim of this process was to perform a conformational creation via a simultaneous variation of nine dihedrals (labelled A-I, as shown in Figure 5.1). Since different conformers primarily differ in the values of the dihedrals of the internal bonds, a different dihedral sequence will result in different conformer structures. Therefore, all the potential conformers can be created by this procedure.

In order to create a subset of possible conformers that is not too big to study in more detail, we chose the values and step sizes of the dihedrals of each bond to be rotated based on the 20 most stable conformers found in the work by Toroz and van Mourik. In the current work, the initial Tyr-Gly structure was built using Molden [117], the initial values of dihedral A, B, C, D, E, F, G, H and I were set to 0˚, 60˚, 60˚, 180˚, 120˚, 60˚, 90˚, 30˚ and 0˚, respectively. For dihedrals B, C and F a step size of 120˚ was employed. Dihedrals A and G were varied with a step size of 180˚. The other dihedrals were not varied, as these have a similar value in all 20 conformers. Therefore, a total of 108 potential conformers were built through this process. None of these contained overlapping atoms, and thus, no conformers were discarded by the program. We know that the original search employed a not-so-suitable functional (B3LYP), and we therefore want to see if other conformers are created when using a better functional. A comparison of the M06-L method with the B3LYP and MP2

methods for predicting the structure of conformer 6 found that the M06-L functional performs better than the other two methods (see chapter 4) [80]. Therefore, in this work, we use the M06-L method to predict the energetically favored structures of Tyr-Gly.

5.3.2 Predicting the structure of Tyr-Gly

5.3.2.1 Scheme 1

Single-point energies were calculated using M06-L/6-31+G(d) for all the 108 possible conformers created by the recursive procedure, which were thereafter sorted according to increasing energy. The calculations employed the software package NWChem [75].

The 30 conformers in the fore part of the order were selected and optimized using M06-L/6-31G(d). The optimized structures were sorted from the most stable conformer to the least one, with the relative energies calculated in kJ/mol by using the energy of the most stable one as the reference. The conformers with relative energies within 12 kJ/mol were further investigated and compared with the 20 stable conformers found by B3LYP/6-31+G(d) optimizations in the former study by Toroz and van Mourik [88], aiming to see the differences between the two sets of stable conformers found by M06-L/6-31G(d) and B3LYP/6-31+G(d) optimizations.

The same procedure was employed for the selected 30 conformers optimized using M05-2X/6-31+G(d), in order to see the differences caused by optimizations using different levels of methods and to verify the reliability of M06-L/6-31G(d).

5.3.2.2 Scheme 2

All 108 possible conformers created by the recursive procedure of the hierarchical selection scheme were optimized using M06-L/6-31G(d). They were then

sorted according to their optimized energies and the relative optimized energies were calculated in kJ/mol using the energy of the most stable conformer as the reference. The conformers within 12 kJ/mol energy difference of the most stable one were selected and their structures were analyzed in terms of the dihedrals. A comparison was performed between these and the 20 most stable conformers found by B3LYP/6-31+G(d) optimizations in the previous study by Toroz and van Mourik [88], to identify new conformers and conformers not found by M06-L. For the missing ones, their structures optimized by B3LYP were used as starting structures for M06-L/6-31G(d) and M05-2X/6-31+G(d) geometry optimizations, followed by dihedral angle analysis to attempt to find the reason why M06-L/6-31G(d) did not find these conformers.

For one of the missing conformers, conformer7, it was found that the geometries optimized using different methods mainly differ in torsion angle C (C9-N10-C11-C12), which is -108˚ in the B3LYP/6-31+G(d) optimized geometry, 179.8˚ in the M06-L/6-31G(d) optimized geometry and -76.5˚ in the M05-2X/6-31+G(d) optimized geometry. Therefore, we calculated the transition energy profiles for rotation around the N10-C11 bond by geometry optimization using M05-2X/6-31+G(d) at fixed values of torsion angle C, using step sizes of 20˚. The profiles were calculated at the M06-L/6-31G(d), M05-2X/6-31+G(d), B3LYP/6-31+G(d) and B3LYP-D/6-31+G(d) levels of theory.

All the calculations in Scheme 2 were performed using the software package Gaussian 09.

5.4 Results

5.4.1 Scheme 1

Table 5.1: Relative optimized energies △E (in kJ/mol), based on M06-L/6-31G* optimization calculations of the 30 conformers selected in scheme 1; values of the dihedrals E, D, C and B (in degrees); structural features and similarity to the 20 most stable conformers found by B3LYP optimizations [88]. OHO : OH ⋯ O hydrogen-bonding interaction of the hydroxyl group and the carbonyl oxygen of tyrosine. Book: when dihedral F < 90˚. Syn/anti : tyrosine OH syn or anti with respect to the glycine NH.

Conf No.	△E	E	D	C	B	structural features	Similar B3LYP conformer
95	0.000	2.33	176.95	75.52	-62.80	OHO,book,anti	1
77	3.697	6.02	174.93	76.39	-60.17	OHO,book,syn	3
103	5.009	9.51	-174.71	-76.88	58.01	OHO,book,anti	2
85	5.805	8.30	-174.72	-76.70	58.02	OHO,book,syn	5
100	6.109	13.24	177.37	-176.36	-178.90	book,anti	4
82	6.818	11.96	178.06	-177.08	-179.05	book,syn	6
98	7.502	20.22	-158.55	78.84	3.24	book,anti	new
92	7.503	20.33	-158.56	78.62	3.25	book,anti	same to 98
80	8.391	20.54	-160.78	78.22	1.64	book,syn	new
74	8.394	20.51	-160.74	78.21	1.75	book,syn	same to 80
106	8.763	12.05	167.08	-83.78	-178.24	book,anti	new
88	9.701	11.31	167.43	-84.25	-178.15	book,syn	new
40	10.836	30.51	-160.11	77.88	178.01	anti	new
58	11.014	30.77	-160.15	77.88	178.04	syn	new
41	11.109	17.42	172.02	77.65	-58.59	OHO,anti	19
59	11.259	17.60	172.17	77.66	-58.70	OHO,syn	17
49	11.380	15.47	-177.50	-76.12	58.66	OHO,anti	18
46	11.744	16.36	176.25	-177.14	-179.48	anti	new
108	11.846	12.70	170.47	-91.78	11.86	book,anti	new
67	11.865	15.34	-177.41	-76.11	58.51	OHO,syn	20

Table 5.1 lists the 20 conformers with relative M06-L/6-31G(d) optimized energies under 12 kJ/mol obtained from the 30 most stable conformers based on single-point M06-L/6-31+G* calculations. Their relative optimized energies using the most stable one as the reference, values of four selected torsion angles, structural

features and similarity to the 20 most stable conformers found by the previous B3LYP optimizations [88] are listed in the table. Only 10 conformers are found to be similar to one of the 20 B3LYP conformers. The new conformers, among which five are folded and the others are extended, are shown in Figure 5.3 (since conformer 92 and 74 are identical to 98 and 80 respectively, they are not shown in Figure 5.3). As seen from the graph, conformer 98 is similar to 80 apart from the flipping of the OH group; conformer 106 is similar to 88 apart from the flipping of the OH group; and conformer 40 is similar to 58 apart from the flipping of the OH group.

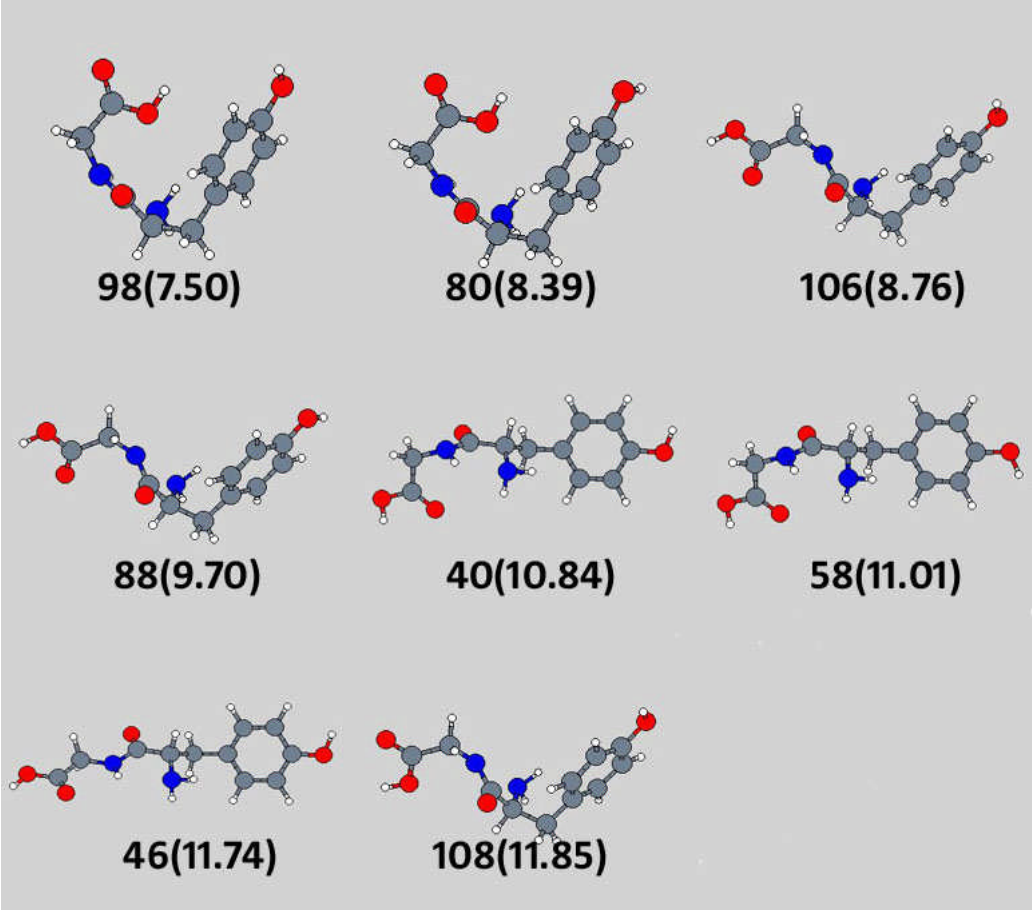


Figure 5.3: New conformers found by M06-L/6-31G* optimization calculations of the 30 conformers selected in scheme 1

Table 5.2: Relative optimized energies △E (in kJ/mol), based on M05-2X/6-31+G* optimization calculations upon the 30 conformers selected in scheme 1; values of the dihedrals E, D, C, B (in degrees); structural features and similarity to the 20 most stable conformers found by B3LYP optimizations [88]. OHO : OH ⋯ O hydrogen-bonding interaction of the hydroxyl group and the carbonyl oxygen of tyrosine. Book: when dihedral F < 90˚. Syn/anti : tyrosine OH syn or anti according to glycine NH's direction.

Conf No.	△E	E	D	C	B	structural features	Similar B3LYP conformer
95	0.000	2.65	175.02	75.77	-62.99	OHO,book,anti	1
103	3.032	10.20	-175.74	-77.18	58.00	OHO,book,anti	2
77	3.844	6.03	172.86	76.48	-60.74	OHO,book,syn	3
85	3.907	8.69	-175.46	-77.25	57.89	OHO,book,syn	5
100	4.576	12.30	177.83	-179.07	-178.80	book,anti	4
82	5.334	11.28	178.40	-179.57	-178.95	book,syn	6
106	6.102	11.57	168.14	-73.88	172.58	book,anti	new
88	7.074	10.94	168.45	-73.77	172.17	book,syn	new
59	7.954	15.24	172.72	78.11	-58.96	OHO,syn	17
41	7.982	14.93	172.58	78.09	-58.91	OHO,anti	19
49	8.366	13.42	-177.01	-76.82	58.67	OHO,anti	18
40	8.448	25.93	-167.51	73.00	-173.33	anti	new
58	8.475	26.24	-167.43	72.86	-173.25	syn	new
67	8.752	13.44	-176.93	-76.84	58.64	OHO,syn	20
46	9.133	14.58	177.21	178.81	-179.33	anti	new
64	9.314	14.76	177.19	178.50	-179.42	syn	new
108	9.489	10.65	172.37	-68.29	-16.71	book,anti	new
92	9.553	21.48	-167.27	70.32	15.59	book,anti	new
74	9.676	20.77	-170.35	71.42	14.03	book,syn	new
90	10.682	10.44	172.43	-69.59	-15.18	book,syn	new

Table 5.2 lists the 20 conformers with relative M05-2X/6-31+G(d) optimized energies under 12 kJ/mol obtained from the 30 most stable conformers based on single-point M06-L/6-31+G* calculations. Their relative optimized energies using the most stable one as the reference, values of four selected torsion angles, structural features and similarity to one of the 20 most stable conformers found by the B3LYP optimizations [88] are included in this table. As found by the M06-L/6-31G(d) optimizations, 10 conformers are similar to one of the 20 B3LYP stable conformers.

The new conformers found are shown in Figure 5.4. In the graph, conformer 106 is similar to 88 except for the flipping of the OH group, so are conformer 40 and 58; conformer 46 and 64; conformer 108 and 90; and conformer 92 and 74. Compared to Figure 5.3, both methods found the new conformers 106, 88, 40, 58 and 46 with similar optimized structures. For the new conformer 108, M05-2X/6-31+G(d) gave different torsion angles of C9-N10-C11-C12 and N10-C11-C12-O13 compared to the M06-L/6-31G(d) optimized structure (-91.78˚ and 11.86˚ in M06-L optimized structure compared to -68.29˚ and -16.71˚ in M05-2X optimized structure).

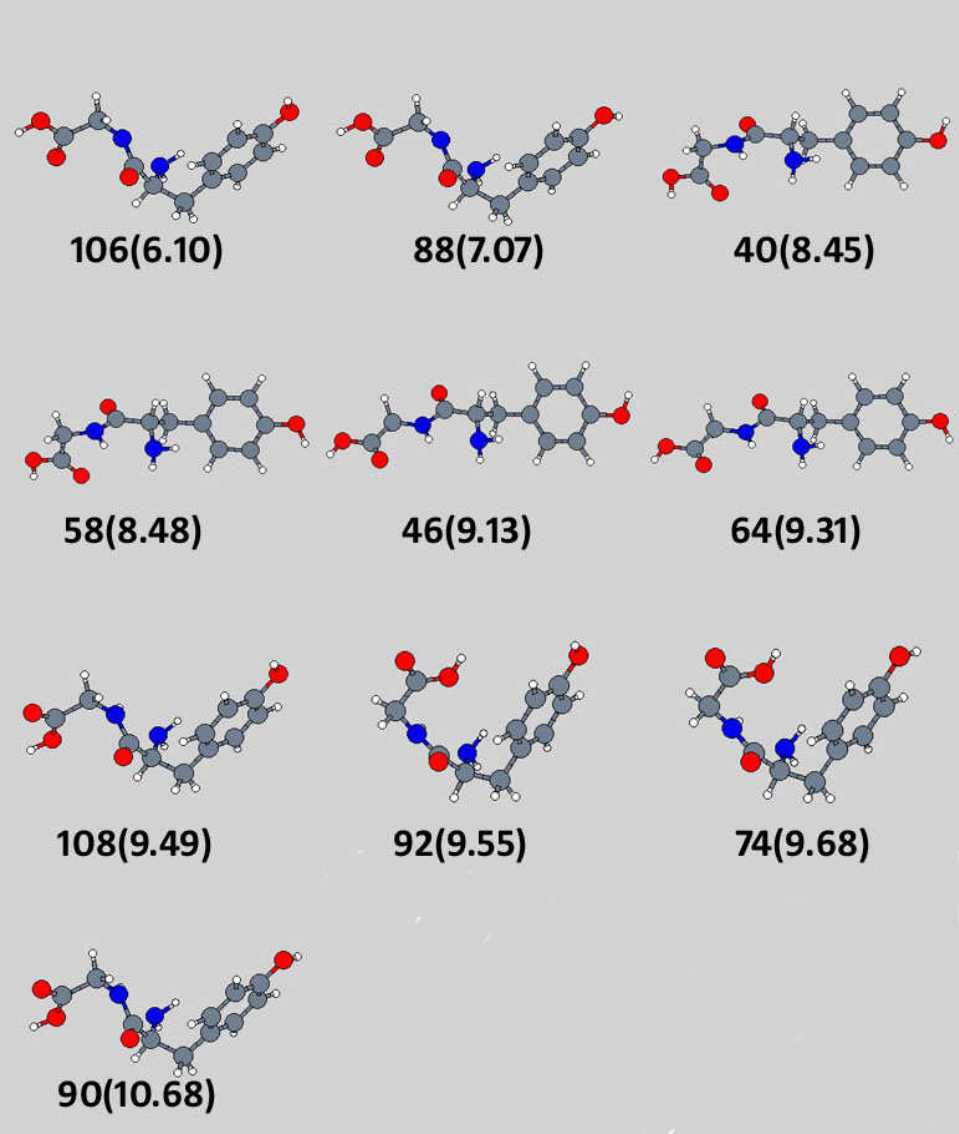


Figure 5.4: New conformers found by M05-2X/6-31+G* optimization calculations of the 30 conformers selected in scheme 1

According to Table 5.1 and Table 5.2, both sets of stable conformers derived by M06-L/6-31G(d) and M05-2X/6-31+G(d) optimization have only found 10 conformers similar to one of the 20 stable conformers optimized by B3LYP/6-31+G(d) displayed in Figure 5.2. The B3LYP conformers 7, 8, 9, 10, 11, 12, 13, 14, 15 and 16 are missing in the two sets of 30 optimized conformers. This may be caused by M06-L and M05-2X finding new stable conformers that were not identified by B3LYP, or not predicted to be among the 20 most stable ones. The relative stability order

changed between the two methods. In Table 5.1, conformer 67 is the least stable conformer while in Table 5.2, six conformers are less stable than this one, namely conformer 46, 64, 108, 92, 74 and 90. In addition, in Table 5.1 conformer 40 and 58 are more stable than conformer 59, 49 and 41, which are similar to B3LYP conformers 17, 18 and 19, whereas in Table 5.2, they are less stable than those three conformers. This means the non-book conformers 17, 18 and 19 are favoured by M05-2X compared to the MP2 method which was used to calculate the single-point energies of the 20 B3LYP conformers.

In conclusion, by using scheme 1, we only found 10 conformers similar to one of the 20 B3LYP conformers. Both of the two methods predict the top six B3LYP conformers to be the six most stable conformers. However, this is in disagreement with a study by Ali Abo-Riziq et al. [118]. In that work, the conformational structures of Tyr-Gly in the gas phase were investigated by UV-UV and IR-UV double resonance spectroscopy and density functional theory calculations. By comparing the results of frequency calculations with the experimentally obtained spectra, the possible conformers that best match the experiment could be chosen with some confidence. In the study, 12 stable conformers were calculated with B3LYP/6-31+G(d), with the B3LYP conformers 10, 7, 11 and 12 listed in the fourth, fifth, sixth and seventh place. The above four conformers were shown to be the ones observed experimentally. The six most stable conformers found by us were not observed. This could mean that either even more accurate levels of theory are needed to correctly rank the different conformers or the six most stable conformers found computationally may only exist for a short period under the experimental conditions used, as suggested by the authors of Ref. [118], and may therefore not be detected in the experiment. M06-L/6-31G(d) optimizations find eight new conformers more stable than B3LYP conformer-20 while with M05-2X/6-31+G(d) optimizations four new conformers are found to be more stable than B3LYP conformer-20. Both methods find the book conformers 106 and 88 to be stable which are missing in the 20 B3LYP

conformers. The reason could be that the book conformers are not favoured by B3LYP optimizations and converge to more extended geometries.

5.4.2 Scheme 2

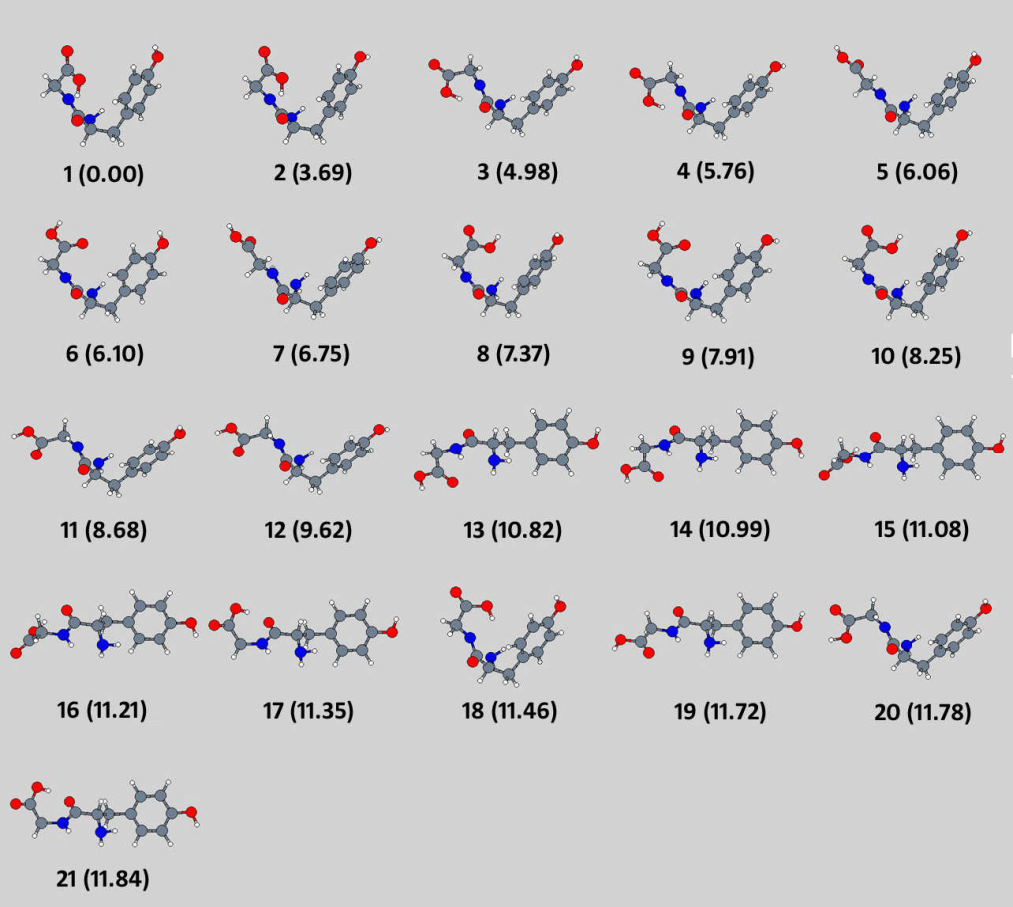


Figure 5.5: M06-L/6-31G* structures of the 21 most stable Tyr-Gly conformers identified in this work. Relative energies (from M06-L/6-31G* optimized energies) are given in kJ/mol.

Table 5.3: Relative energies △E (in kJ/mol), based on single-point M06-L/6-31G* calculations upon the M06-L/6-31G* optimized geometries; values of the dihedrals E, D, C, B (in degrees); structural features and similarity to the 20 most stable conformers found by B3LYP optimizations [88]. OHO : OH⋯O hydrogen-bonding interaction of the hydroxyl group and the carbonyl oxygen of tyrosine. Book: when dihedral F < 90°. Syn/anti : tyrosine OH syn or anti according to glycine NH's direction.

conf	△E	E	D	C	B	structural features	Similar B3LYP conformer
1-M06	0.000	1.61	177.24	75.34	-63.31	OHO,book, anti	1
2-M06	3.688	5.37	174.92	76.22	-60.32	OHO,book, syn	3
3-M06	4.981	9.48	-174.95	-76.57	57.98	OHO,book, anti	2
4-M06	5.759	8.51	-174.77	-76.66	58.08	OHO,book, syn	5
5-M06	6.056	13.15	177.71	-177.43	-179.12	book, anti	4
6-M06	6.104	17.20	-176.17	79.33	179.43	book, anti	new
7-M06	6.746	12.03	178.01	-176.80	-179.37	book, syn	6
8-M06	7.368	19.36	-159.49	80.91	2.15	book, anti	new
9-M06	7.905	21.60	-171.37	80.47	178.61	book, syn	new
10-M06	8.250	20.41	-160.89	78.07	1.38	book, syn	new
11-M06	8.677	11.89	167.19	-85.04	-177.90	book, anti	new
12-M06	9.616	11.27	167.39	-85.04	-178.04	book, syn	new
13-M06	10.819	30.44	-160.98	78.37	178.13	anti	new
14-M06	10.993	30.69	-161.10	78.32	178.28	syn	new
15-M06	11.076	17.53	172.09	77.60	-58.50	OHO, anti	19
16-M06	11.215	17.50	172.33	77.58	-58.77	OHO, syn	17
17-M06	11.352	15.19	-177.40	-76.09	58.49	OHO, anti	18
18-M06	11.458	-0.14	174.25	108.06	-14.92	book, anti	new
19-M06	11.718	17.20	177.22	177.01	-179.13	anti	new
20-M06	11.776	13.00	170.16	-90.11	11.20	book, anti	new
21-M06	11.839	15.18	-177.30	-76.10	58.44	OHO, syn	20

Table 5.3 lists the 21 most stable conformers (based on M06-L/6-31G* geometry optimizations of the 108 conformers created by the recursive procedure), in which conformer 1-M06, 2-M06, 3-M06, 4-M06, 5-M06, 7-M06, 15-M06, 16-M06, 17-M06 and 21-M06 are similar to one of the 20 conformers optimized by the B3LYP/6-31+G* method [88] and the rest are newly found stable structures. Three conformers, 6-M06, 9-M06, 18-M06, are missed by scheme 1. The M06-L/6-31G* optimized structures of all the 21 stable conformers and their relative optimized

energies are shown in Figure 5.5.

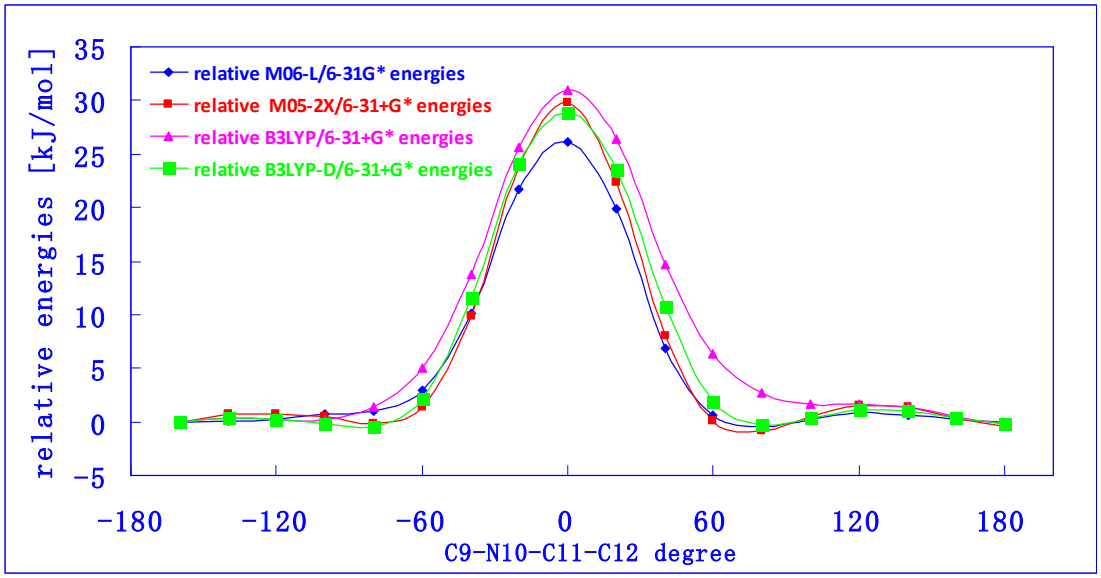


Figure 5.6: Relative transition energy profiles of conformer 7 for rotation around the N10-C11 bond using different levels of theory (based on the M05-2X/6-31+G(d) optimized geometry for each rotation step)

For the missing B3LYP conformers 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 20, we applied M06-L/6-31G(d) optimization calculations and found that only conformers 7 and 10 converged to the structures with large differences to the original ones, all the others almost kept the same structures. Therefore, we tried to investigate conformer 7 to a further extent. Since we observed that the geometries optimized using various methods mainly differ in the torsion angle C (C9-N10-C11-C12), we calculated the transition energy profiles for rotation around the N10-C11 bond at different levels of theory. Figure 5.6 presents the relative transition energy profiles of conformer 7 with respect to rotation around the N10-C11 bond using the M06-L/6-31G*, M05-2X/6-31+G*, B3LYP/6-31+G* and B3LYP-D/6-31+G* method. The M06-L, M05-2X and B3LYP-D curves have minima at 80˚ and 180˚. Only the B3LYP curve does not present a minimum at 80˚, which could result from the lack of dispersion in this functional. However, the B3LYP curve shows another minimum at about -108˚,

which is the reason why in B3LYP conformer 7 the dihedral C equals -108˚. Both the M05-2X and B3LYP-D curves have another minimum at -80˚, and this explains the finding that after optimization of conformer 7 using M05-2X/6-31+G*, the dihedral C equalled -80˚. Full geometry optimization of this conformer with B3LYP-D/6-31+G*, starting from a structure with the torsion angle C set to -80°, resulted in a structure with a torsion angle value of -82.3°, indicating that B3LYP misses this particular conformer due to the lack of dispersion in B3LYP theory.

5.5 Conclusion

From the former studies we realize that the B3LYP and MP2 methods (with small basis sets like 6-31+G*) do not correctly predict the structure and energy order of tyrosine-glycine conformers [5-6]. In this work, the M06-L functional from the Minnesota density functional theories was applied to solve the problem. First, 108 potential conformers were created with a Fortran program using a recursive procedure taken from previous work [88]. After that, we employed two schemes. The first one was to sort all the conformers according to their M06-L/6-31+G(d) single-point energies and optimize the first 30 conformers. The second one was to implement the optimization on all the 108 conformers. The optimization calculations were calculated using M06-L/6-31G(d) or M05-2X/6-31+G(d). For both schemes, the most stable conformers were selected according to the relative optimized energies and compared to the 20 stable conformers found by B3LYP optimized structures in previous work [88]. 10 stable conformers similar to one of the B3LYP stable conformers were found by both schemes, with the rest as newly found conformers. Three newly found conformers from scheme 2, conformer 6-M06, 9-M06 and 18-M06, were missed by scheme 1, indicating that these dropped down in energy more than other conformers upon geometry optimisation. We also applied relative transition energy profile calculations on a missing B3LYP stable conformer using different levels of theory.

The results indicate that B3LYP misses the conformer found by the other methods. This is due to the lack of dispersion in B3LYP theory, as also indicated by full geometry optimization using B3LYP-D, which did yield the conformer missed by B3LYP.

Chapter 6 Modelling Zwitterions in Solution: 3-fluoro-γ-amino-butyric acid (3F-GABA)

6.1 Abstract

The conformations of neutral and zwitterionic 3-fluoro-γ-aminobutyric acid (3F-GABA) in solution are studied using different solvation models, aiming to give more detailed knowledge of the structure of 3F-GABA in a physiological environment and to provide a better understanding of the relative accuracy of different solvation models. Extended zwitterions are obtained from all the explicit water models. No single-water molecule models can stabilize folded zwitterions. All the zwitterions considered are stable in five-water molecule models, indicating that zwitterionic forms of 3F-GABA are preferred in solution. All the relative stability energy differences of the explicit water molecules models are much larger than that from the calculation using the PCM continuum solvation model. The more explicit water molecules added in, the smaller the relative energy difference becomes. It is found that M06-2X/6-31+G* performs better in calculating transition energy profiles than MP2/6-31+G* [119].

6.2 Introduction

As the chief inhibitory amino acid in the mammalian central nervous system [120], γ-aminobutyric acid (GABA), H_2N-CH_2-CH_2-CH_2-COOH, plays an important role in maintaining normal neuronal activity by regulating neuronal excitability. It is also directly responsible for the regulation of muscle tone in humans. GABA, having both an amino and a carboxylic acid group, can exist in either zwitterionic or neutral

(nonzwitterionic) forms, which are represented in Figure 6.1. 3F-GABA is an analogue of the neurotransmitter GABA, having the similar feature as GABA by showing the different forms when existing in different environments. In aqueous solution 3F-GABA exists as a zwitterion, but it is a neutral (non-zwitterionic) molecule in the gas phase. Therefore, solvent effects must be accounted for to investigate the zwitterionic form of 3F-GABA computationally. Unlike GABA, 3F-GABA contains a chiral atom (the central C atom) and 3F-GABA therefore exists in two enantiomeric forms, (R) and (S). In this work we considered the (S) enantiomer.

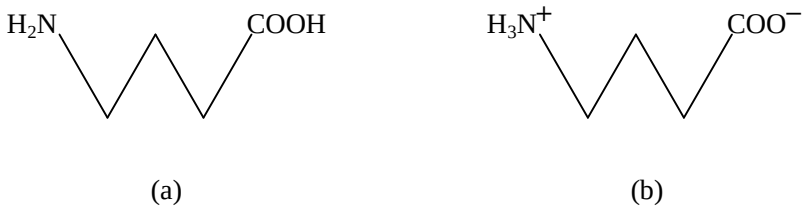


Figure 6.1: Structures of GABA in: (a) Neutral form (b) Zwitterionic form

The zwitterionic form of 3F-GABA in solution can be modelled in two different ways: (1) by using a so-called continuum solvation model, where the solvent is modelled as a continuum of a dielectric constant; (2) by including explicit water molecules in the calculation as 3F-GABA hydrate, that is, putting water molecules around 3F-GABA. However, in a recent study [121] in which the continuum solvation model PCM (polarisable continuum model [122]) was used to study the different conformations of 3F-GABA in solution, the most stable conformer according to the calculations did not conform to the experimentally observed structure. In a study on neutral and zwitterionic GABA [123], the stabilization of GABA zwitterions was investigated at the HF/6-31+G*, B3LYP/6-31+G*, and MP2/6-31+G* levels of theory using explicitly bound water molecule models. The only approaches yielding

zwitterionic structures consistent with experiment use the explicit water models GABA·2H_2O or GABA·5H_2O. Therefore, an accurate description of aqueous phase GABA requires both explicit interaction with at least two water molecules and long-range dielectric interactions with the solvent. Also it was found that explicit water molecules are required to obtain reliable stabilities of the different conformers.

In the current work, we investigate the performance of different solvation models (explicit and continuum solvation) to study the neutral and zwitterionic conformations of 3F-GABA to see if explicit waters change their relative stability. Here we look at the relative stability of two conformers, the extended conformer F and folded conformer B.

It is shown that explicit two-water molecule models cannot stabilize zwitterionic 3F-GABA. In transition energy profiles for a GABA dihydrate (for hydrogen transfer between the carboxylic acid oxygen and the amino nitrogen), M06-2X/6-31+G* gives results in excellent agreement with the CCSD(T) reference profile, while MP2/6-31+G* falsely stabilizes the zwitterionic GABA dihydrate model, thereby indicating the influence of BSSE in the calculation of transition energy profiles and manifesting M06-2X as an adequate functional for exploring the stability of explicit water molecule models for zwitterionic 3F-GABA.

6.3 Methodology

In the investigation of different explicit water models, both an extended and a folded stable zwitterionic 3F-GABA conformer were considered, conformer F and conformer B, in order to illustrate how the explicit water molecules bind to 3F-GABA conformers with different foldedness and to study the variation of their stability in solution.

Firstly, only one water molecule was used to form the models. Seven starting structures were built for the extended 3F-GABA conformer F. The water molecule

was simply put near either of its ends or near the fluorine atom, in a position where one or two potential hydrogen bonds could be formed. For example, the oxygen atom of the water molecule was put near the NH_3 group to form a potential N-H⋯O bond, or near the fluorine atom to form an O-H⋯F bond. When put near the two oxygen atoms, it was placed adjacent to one of the oxygen atoms or in the middle of them where two O-H⋯O bonds could be formed, as seen in Figure 6.2 (es4, es5, es6). Another six starting structures were built for the folded 3F-GABA conformer B, in which the water molecule was also put in a position aiming to form potential hydrogen bonds. But this time, since the NH_3 group is near the two-oxygen group, a new starting geometry can be created in which the water molecule is placed between the NH_3 group and an oxygen atom, forming a potential N-H⋯O and O-H⋯O bond, as indicated in Figure 6.2 (fs1).

Secondly, two-water molecule models were used. For these, the number of possible structures of zwitterionic 3F-GABA dihydrate increased significantly compared to the single-water molecule models. We considered 12 starting structures for each of the extended conformer F and folded conformer B. Based on the single-water molecule models, the second water molecule can be put near the first one or the two water molecules can be placed separately, near either of the ends or near the fluorine atom. The main rule was still to form potential hydrogen bonds in the models, as shown in Figure 6.7 and 6.10.

Thirdly, observing that two water molecules cannot keep folded 3F-GABA in its zwitterionic form after optimization, three water molecules were considered. At this stage not every possible position for the water molecules to be placed was taken into account. Only several starting models were built, which tended to form the largest number of potential hydrogen bonds.

After building the explicit water models, M06-2X/6-31+G* was used to optimize their structures. M06-2X is a meta-hybrid functional, a member of the M06 family of density functionals developed by Zhao, Truhlar and Schultz, which show good

capability for dispersion-dominated interaction calculations [40, 124].

For all explicit water models, counterpoise-corrected energies [9] were also calculated for comparison with the regular optimized energies, to investigate to what extent the BSSE affects the results.

A previous study by Crittenden et al. obtained nine stable zwitterionic GABA • $5H_2O$ conformers in the gas phase using MP2/6-31+G* optimizations [123]. Based on these structures, we created explicit five-water molecule models by changing GABA to (S)-3F-GABA enantiomer and optimized these models using M06-2X/6-31+G*.

Furthermore, in the previous study, seven stable zwitterionic GABA dihydrates were obtained in the gas phase using MP2/6-31+G*, denoted as zw1-zw7, proving that two water molecules were sufficient to keep folded GABA in its zwitterionic form after optimization [123], which, as our study shows, cannot be realized for folded zwitterionic 3F-GABA with the same structure using M06-2X/6-31+G*. In order to establish the intrinsic reason that caused the different results, hydrogen transition profiles (for hydrogen transfer between the carboxylic acid oxygen and the amino nitrogen) were calculated for one of the seven zwitterionic GABA dihydrates, zw7, by optimizing the transition state structures using MP2/6-31+G* and computing single-point energies using different methods: MP2/6-31+G* (both with and without intermolecular counterpoise correction), MP2/aug-cc-pVDZ, MP2/aug-cc-pVTZ, M06-2X/6-31+G* and CCSD(T)/aug-cc-pVDZ.

6.4 Results

6.4.1 Single-water molecule models

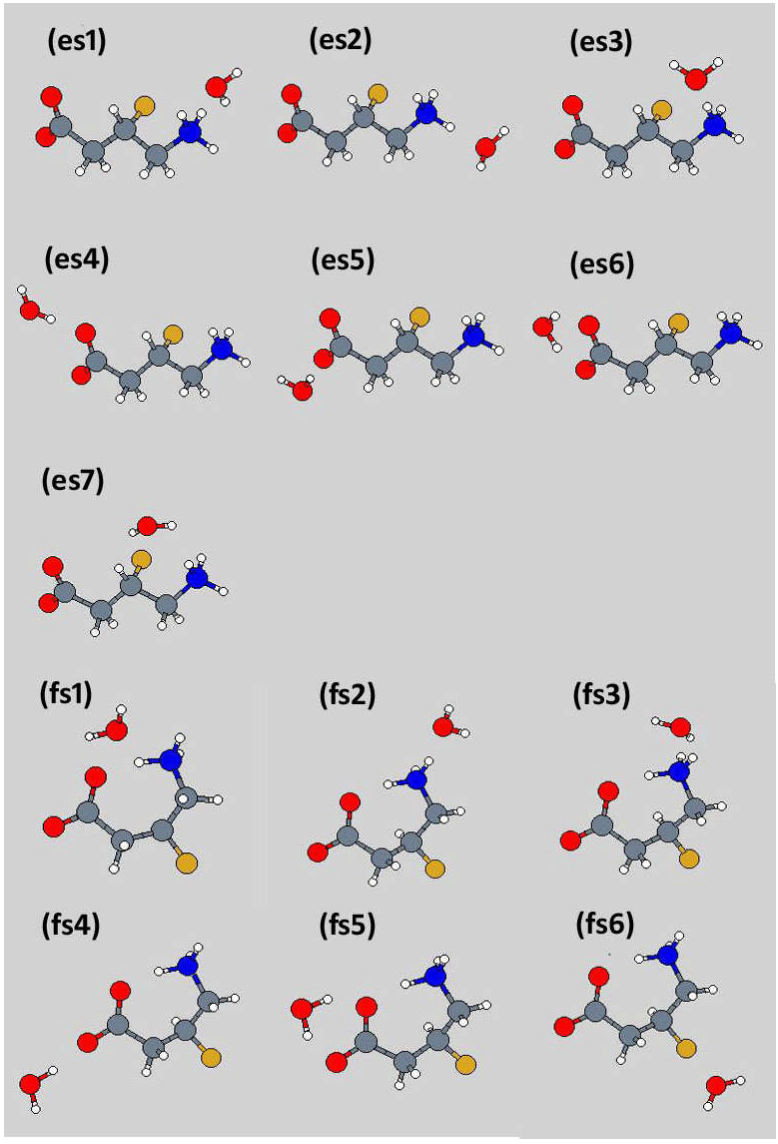


Figure 6.2: Starting structures of single-water molecule models for zwitterionic 3F-GABA

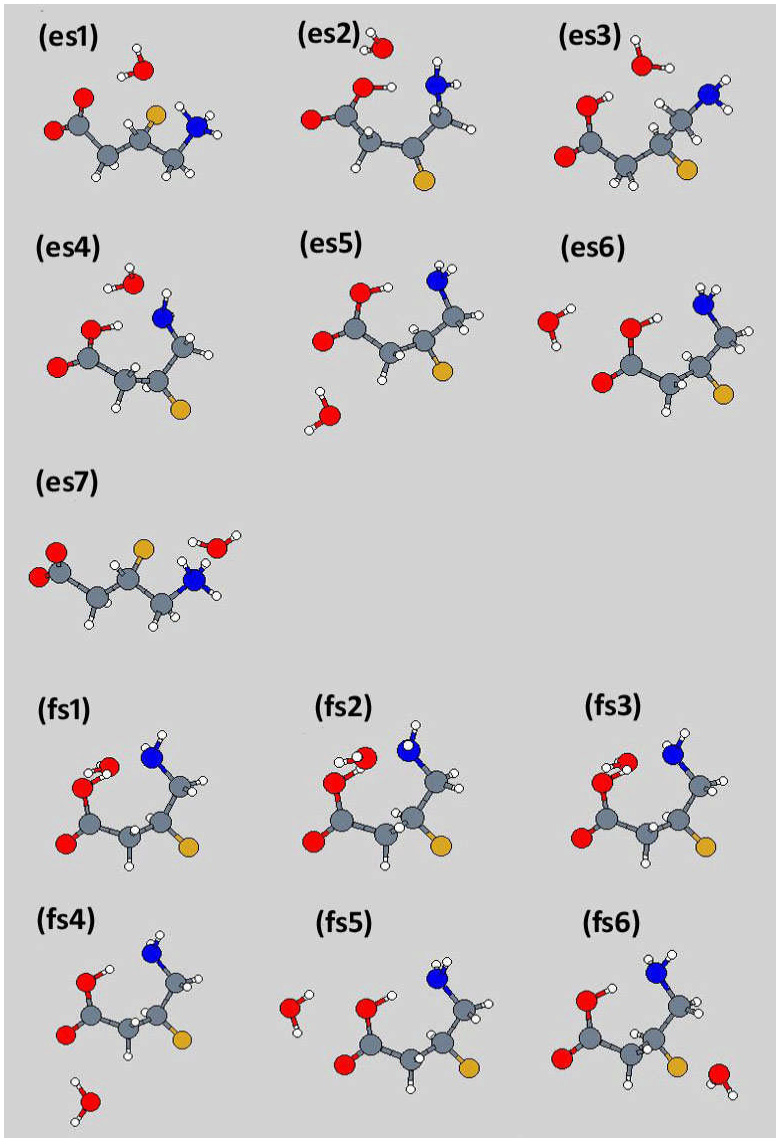
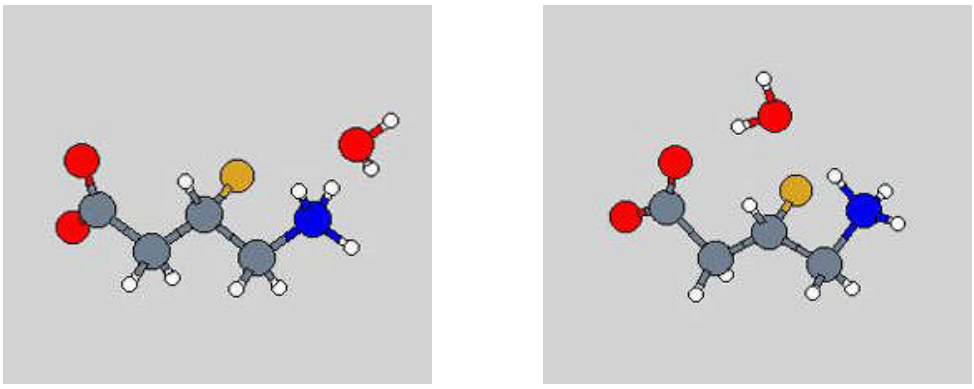
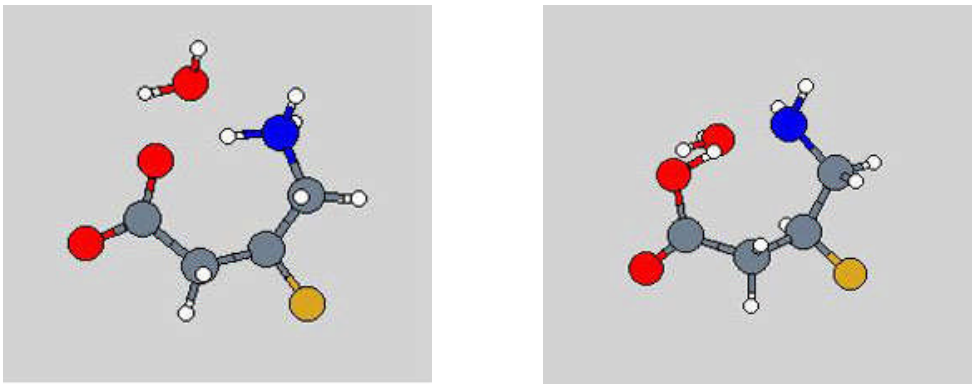


Figure 6.3: Optimized structures of single-water molecule models for zwitterionic 3F-GABA



(a) (b)

Figure 6.4: Single-water molecule model of extended zwitterionic 3F-GABA (es1)
(a) Before optimization (b) After optimization



(a) (b)

Figure 6.5: Single-water molecule model of folded zwitterionic 3F-GABA (fs1)
(a) Before optimization (b) After optimization

Figures 6.2 and 6.3 show the starting and optimized structures of the single-water molecule models for zwitterionic 3F-GABA respectively, in which es1-es7 are extended models and fs1-fs6 are folded ones. Es1 and es7 remained in their zwitterionic form in the optimization process, as is shown in Figure 6.4 by taking es1 as an example. All the other conformers do not stay in zwitterionic forms with one water molecule and converged to folded neutral forms, as shown in Figure 6.5 by taking fs1 as an example.

TABLE 6.1: Relative optimized energies △E, relative counterpoise-corrected optimized energies △CPE, BSSE values and structural features of the single-water molecule models. All energies are in kJ/mol and are relative to the lowest optimized energy model at the M06-2X/6-31+G* level of theory.

3F-GABA-H_2O	Starting structures	△E	△CPE	BSSE	Struct. features[a]
S1	fs1, fs3	0.00	0.00	-5.52	N, F
S2	fs2	8.74	8.21	-4.98	N, F
S3	es2, es4	9.05	8.27	-4.74	N, F
S4	es3	5.35	8.45	-8.62	N, F
S5	es5	16.28	13.35	-3.18	N, F
S6	fs4	15.69	13.35	-3.18	N, F
S7	es6, fs5	17.68	16.86	-4.70	N, F
S8	fs6	21.82	19.62	-3.32	N, F
S9	es1	168.88	171.13	-7.76	ZW, E
S10	es7	202.24	203.11	-6.39	ZW, E

[a] N = neutral; ZW = zwitterionic; F = folded; E = extended

Table 6.1 lists the relative optimized energies, relative counterpoise-corrected energies, BSSE values and structural features of the single-water molecule models. The unique optimized structures are labeled S1-S10 and are ordered according to increasing counterpoise-corrected optimized energy. The “starting structures” column shows which models converged to these. From the optimized energies and structures it is indicated that fs1 is the same as fs3, so are es2 and es4; es6 and fs5. In total we obtained eight folded neutral structures and two extended zwitterions. Relative counterpoise corrected energies were calculated for these conformers and compared to the relative optimized energies, as shown in Table 6.1 and Figure 6.6, indicating that counterpoise correction does not change the relative energies significantly even though it changes the order of stability of some of the conformers. It can be seen that after optimization the neutral forms are preferred. S1 is the most stable conformer. Apparently the neutral optimized conformers are significantly more stable than the zwitterionic optimized conformers S9 and S10. From the BSSE values it appears that S4 and S9’s BSSE is much larger than those of other conformers, which probably

results from the fact that among those conformers S4 and S9 have the water molecule between the two-oxygen group and the NH_3 group forming a hydrogen-bonded bridge.

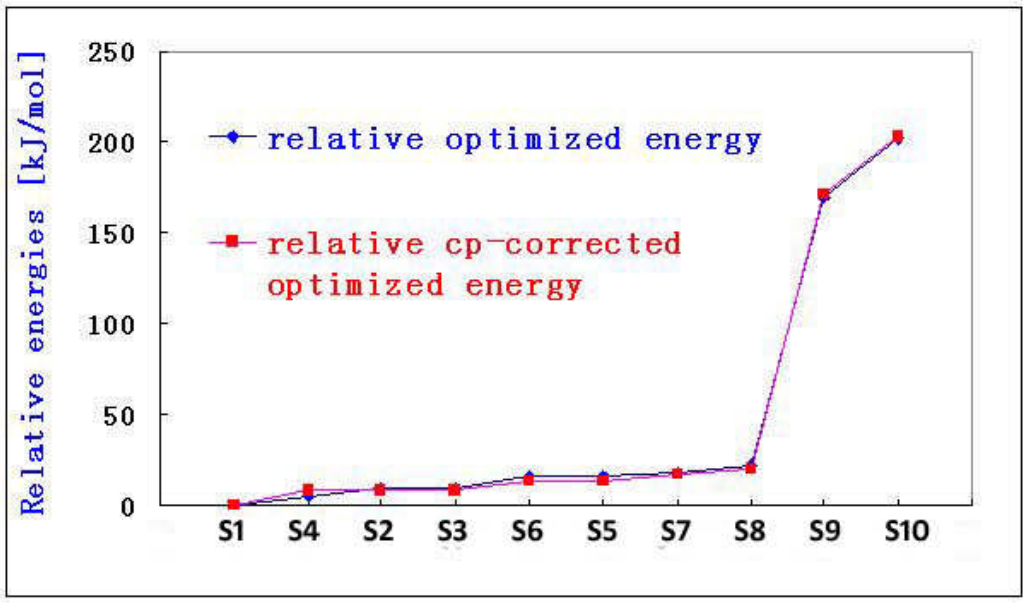


Figure 6.6: Comparison of relative optimized energies and relative counterpoise corrected energies of single-water molecule models

TABLE 6.2: Dihedral analysis of the unique zwitterions obtained from the single-water molecule models, compared with the extended conformer F and folded conformer B, optimized with B3LYP/6-31+G*/PCM [122].

3F-GABA • H_2O models	O-C1-C2-C3 dihedral angle	C1-C2-C3-F dihedral angle	C1-C2-C3-C4 dihedral angle	C2-C3-C4-N dihedral angle
conformer F	-34.65	-66.60	175.72	176.73
conformer B	-26.69	-169.22	74.21	-80.54
S9, extended	-15.73	-70.07	166.01	-179.24
S10, extended	-27.28	-78.11	163.66	173.59

Table 6.2 lists the dihedrals of the zwitterions obtained from the single-water molecule model optimizations, as well as those of the extended conformer F and

folded conformer B optimized with B3LYP/6-31+G*/PCM. S9 and S10 are similar to conformer F, whereas no zwitterion has a similar structure as conformer B.

6.4.2 Two-water molecule models

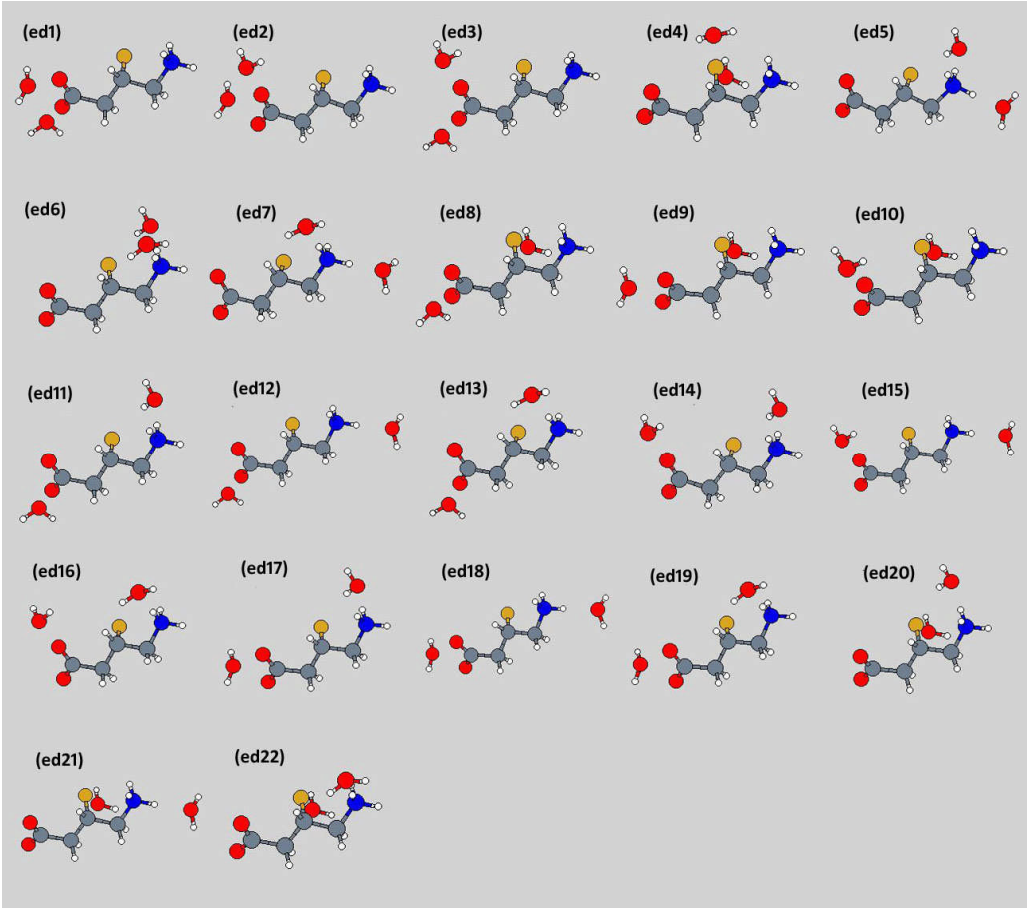


Figure 6.7: Starting structures of extended two-water molecule models for zwitterionic 3F-GABA

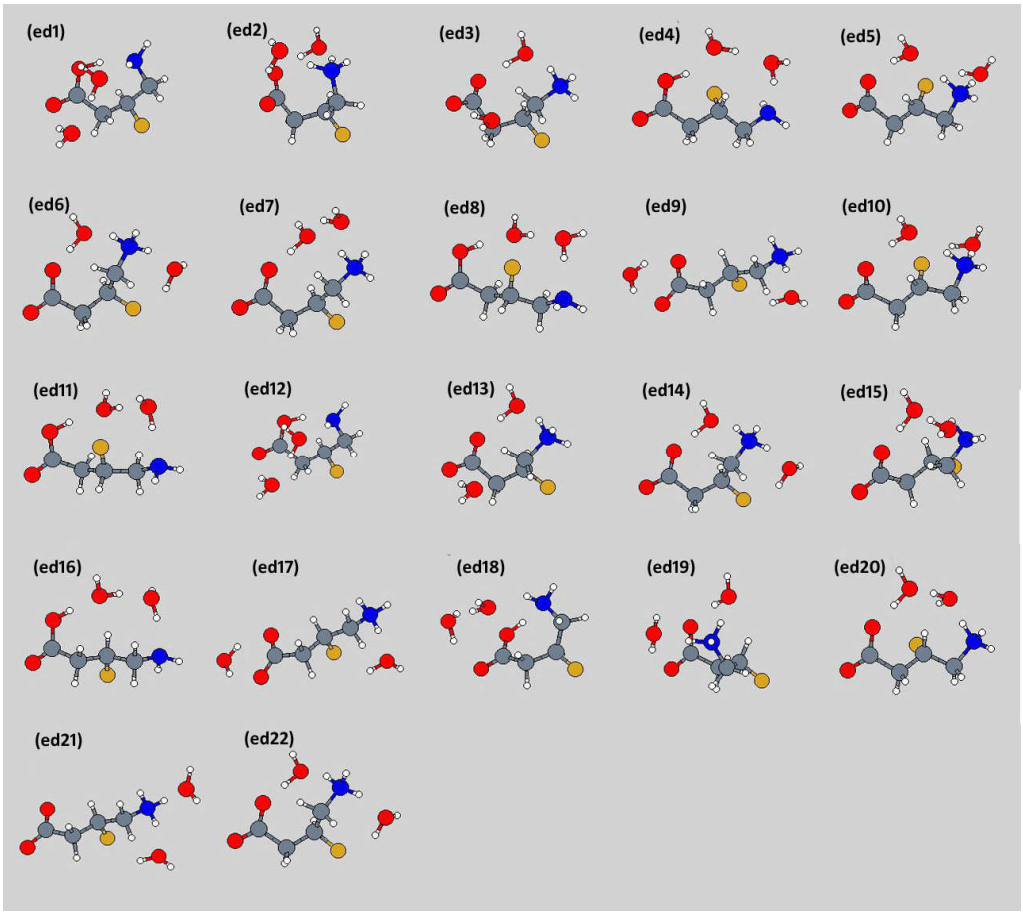
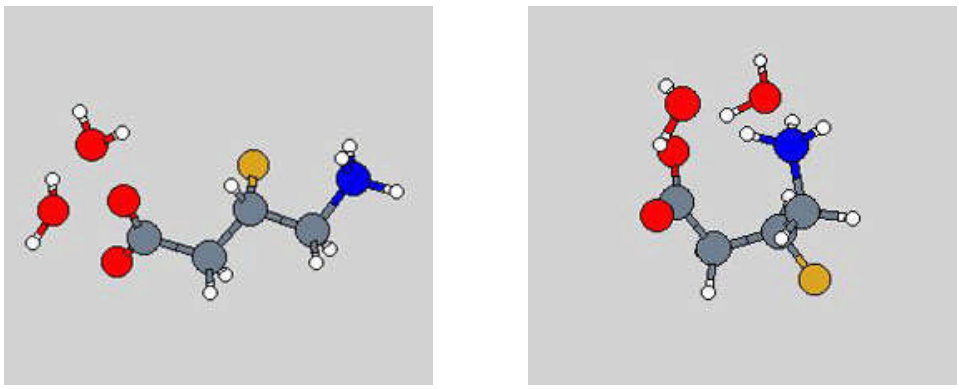


Figure 6.8: Optimized structures of extended two-water molecule models for zwitterionic 3F-GABA



(a) (b)

Figure 6.9: Two-water molecule model of extended zwitterionic 3F-GABA (ed2)
(a) Before optimization (b) After optimization

Figures 6.7 and 6.8 show the starting and optimized structures of the extended two-water molecule models for zwitterionic 3F-GABA. Only seven of them, ed1, ed4, ed8, ed11, ed12, ed16 and ed18, converged to neutral forms after optimization while the other conformers all remained in their zwitterionic forms. Therefore, two water molecules are sufficient to keep extended zwitterionic 3F-GABA from converging to neutral forms, as also shown in Figure 6.9 by taking ed2 as an example. From Figure 6.8 it can be seen that, although most of the 3F-GABAs remained in zwitterionic form after optimization (though being much less stable than the neutral conformers, see table 6.3), their structures have not stayed the same. Ed5, ed9, ed10, ed17, ed20 and ed21 still represent fully extended structures, but with some bonds rotated. Ed3, ed6, ed7, ed13, ed14, ed15 and ed22 are “bent” (partially folded), extended structures. The remaining two structures, ed2 and ed19, are folded, with the water molecules forming a hydrogen-bonded bridge between the two-oxygen group and the NH_3 group, showing the most significant change that the optimization procedure may incur.

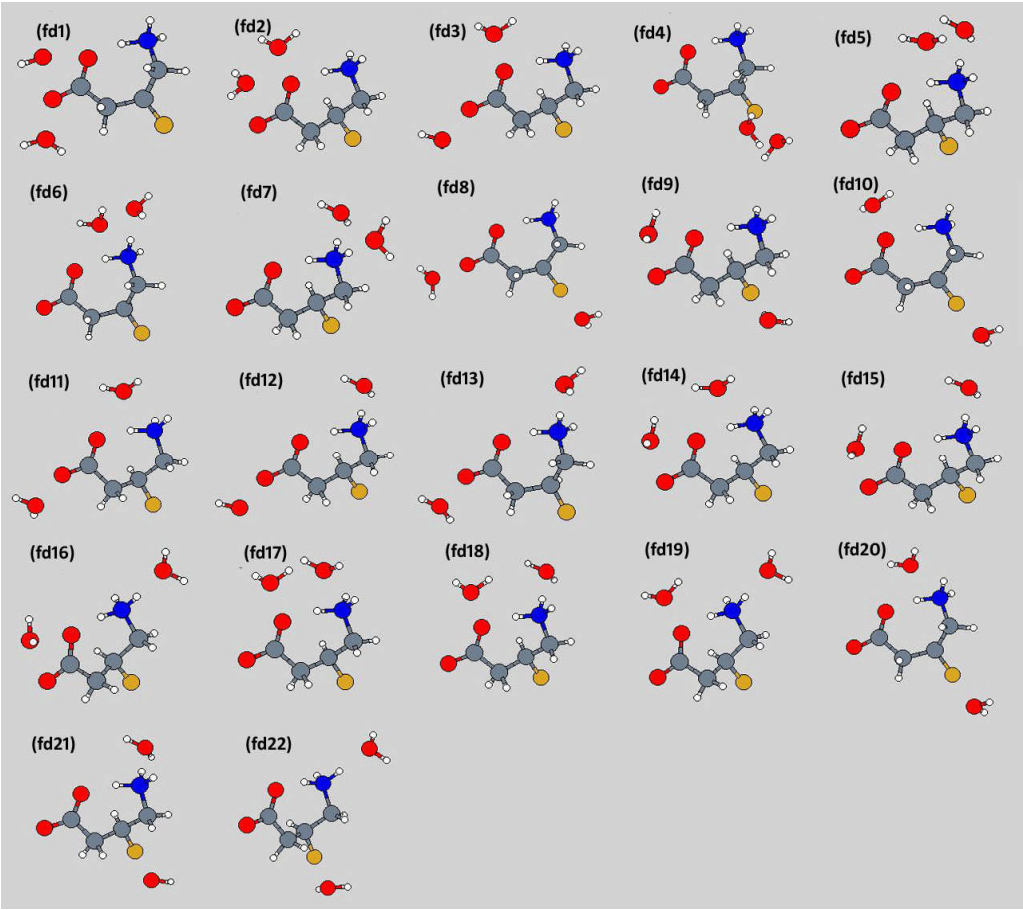


Figure 6.10: Starting structures of folded two-water molecule models for zwitterionic 3F-GABA

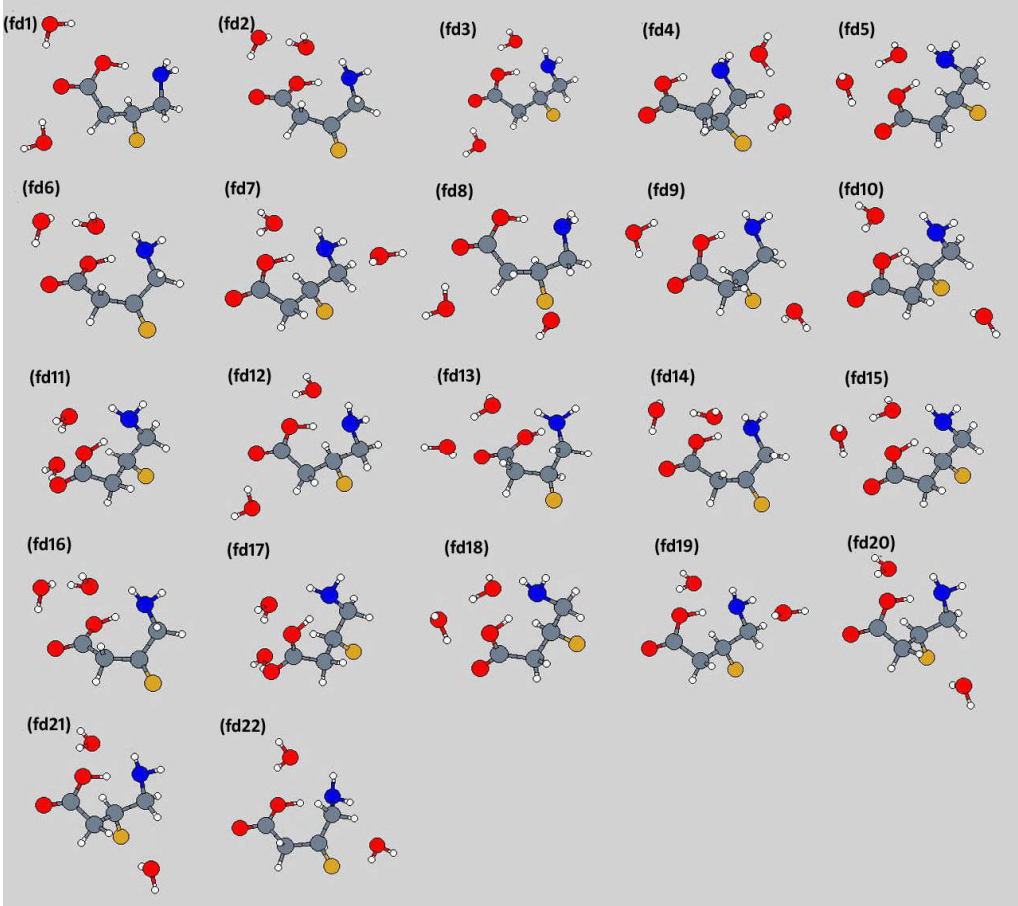
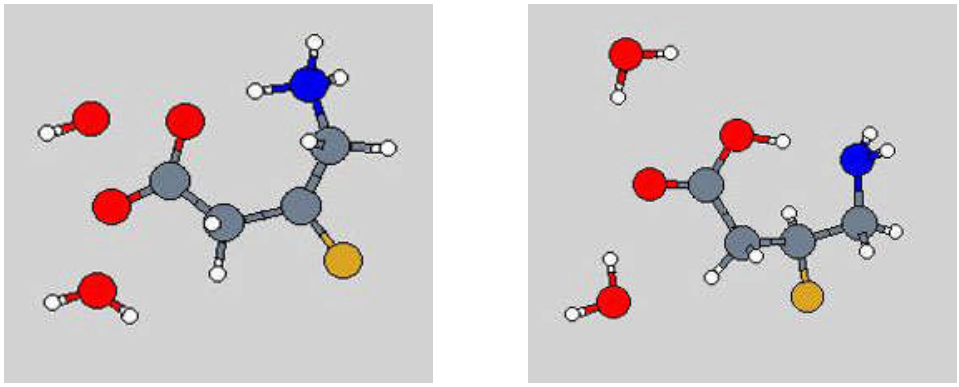


Figure 6.11: Optimized structures of folded two-water molecule models for zwitterionic 3F-GABA



(a) (b)

Figure 6.12: Two-water molecule model of folded zwitterionic 3F-GABA (fd1)
(a) Before optimization (b) After optimization

Figures 6.10 and 6.11 show the starting and optimized structures of the folded two-water molecule models for zwitterionic 3F-GABA. The results show that none of these remained zwitterionic and all converged to neutral forms, which indicate that it is difficult for two water molecules to stabilize the structure of this folded zwitterionic 3F-GABA conformer, as shown in Figure 6.12 by taking fd1 as an example. However, two water molecules are able to stabilize another folded 3F-GABA structure, see ed2 in Figure 6.8.

TABLE 6.3: Relative optimized energies △E, relative counterpoise-corrected optimized energies △CPE, BSSE values and structural features of the two-water molecule models. All energies are in kJ/mol and are relative to the lowest optimized energy model at the M06-2X/6-31+G* level of theory.

3F-GABA-2H_2O	Starting structures	△E	△CPE	BSSE	Struct. features[a]
D1	fd5, fd15, fd18	0.00	0.00	-13.98	N, F
D2	ed18, fd2, fd6, fd14, fd16	4.83	4.63	-13.78	N, F
D3	fd11, fd17	10.43	9.08	-12.63	N, F
D4	ed16	7.70	11.17	-17.45	N, E
D5	ed1, ed12, fd13	13.89	12.34	-12.43	N, F
D6	ed11	18.34	22.69	-18.33	N, E
D7	ed4	23.08	25.61	-16.52	N, E
D8	ed8	22.56	26.13	-17.55	N, E
D9	fd4	29.16	26.27	-11.09	N, F
D10	fd3, fd12	32.42	27.06	-8.62	N, F
D11	fd10, fd21, fd20	37.11	31.92	-8.79	N, F
D12	fd7, fd19	39.70	34.43	-8.71	N, F
D13	ed2, ed19	32.08	35.30	-17.20	ZW, F
D14	fd22	42.45	36.68	-8.20	N, F
D15	fd8	49.35	42.28	-6.91	N, F
D16	fd1	49.41	43.65	-8.23	N, F
D17	fd9	52.68	46.87	-8.17	N, F
D18	ed7, ed15	78.42	78.97	-14.54	ZW
D19	ed3, ed13	87.84	86.87	-13.01	ZW
D20	ed6, ed14, ed22	86.63	88.81	-16.16	ZW
D21	ed20	135.40	140.35	-18.91	ZW, E
D22	ed5, ed10	164.30	164.36	-14.08	ZW, E
D23	ed9, ed17	202.20	199.68	-11.49	ZW, E
D24	ed21	212.90	209.84	-10.89	ZW, E

[a] N = neutral; ZW = zwitterionic; F = folded; E = extended

Table 6.3 lists the relative optimized energies, relative counterpoise-corrected optimized energies, BSSE values and structural features of the two-water molecule models. There are 24 unique optimized structures, labeled as D1-D24 and ordered according to increasing counterpoise-corrected optimized energy. The “starting structures” column shows which models converged to these. By comparison of the optimized energies and structures, we found that among all those zwitterionic

optimized structures, ed9 and ed17 are similar, so are ed2 and ed19; ed3 and ed13; ed5 and ed10; ed7 and ed15; ed6, ed14 and ed22. In total we obtained 16 neutral structures and eight zwitterions, of which two are “bent”, three are extended and one is folded. Relative counterpoise corrected energies were calculated for these conformers and compared to the relative optimized energies, as shown in Table 6.3 and Figure 6.13. As in the situation of the one-water molecule models, the counterpoise correction does not change the relative energies significantly, while it does change the order of stability of some of the conformers. From the “structured features” column, it can be seen that the neutral forms are preferred after optimization as the most stable conformer, D1, is a folded neutral form.

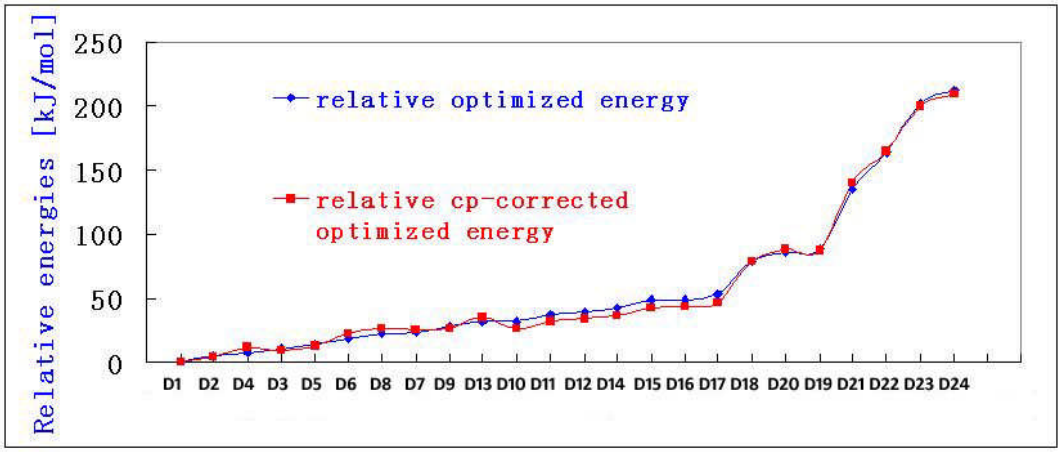


Figure 6.13: Comparison of relative optimized energies and relative counterpoise corrected energies of two-water molecule models

TABLE 6.4: Dihedral analysis of the unique zwitterions obtained from the two-water molecule models, compared with the extended conformer F and folded conformer B, optimized with B3LYP/6-31+G*/PCM [122].

3F-GABA • $2H_2O$ models	O-C1-C2-C3 dihedral angle	C1-C2-C3-F dihedral angle	C1-C2-C3-C4 dihedral angle	C2-C3-C4-N dihedral angle
conformer F	-34.65	-66.60	175.72	176.73
conformer B	-26.69	-169.22	74.21	-80.54
D13, folded	69.86	163.42	46.03	-85.77
D20, “bent” extended	3.85	-173.70	67.39	-154.45
D19, “bent” extended	44.02	-178.39	62.05	-159.16
D21, extended	-19.42	-65.15	176.26	-170.39
D22, extended	-15.96	-67.39	169.60	-178.72
D23, extended	-29.66	-74.81	167.22	172.64

Table 6.4 lists the dihedrals of the zwitterions obtained from the two-water molecules model optimizations, as well as those of the extended conformer F and folded conformer B optimized with B3LYP/6-31+G*/PCM. The obtained zwitterions D21, D22 and D23 are all similar to the extended conformer F whereas no zwitterion has a structure close to the folded conformer B.

6.4.3 Three-water molecule models

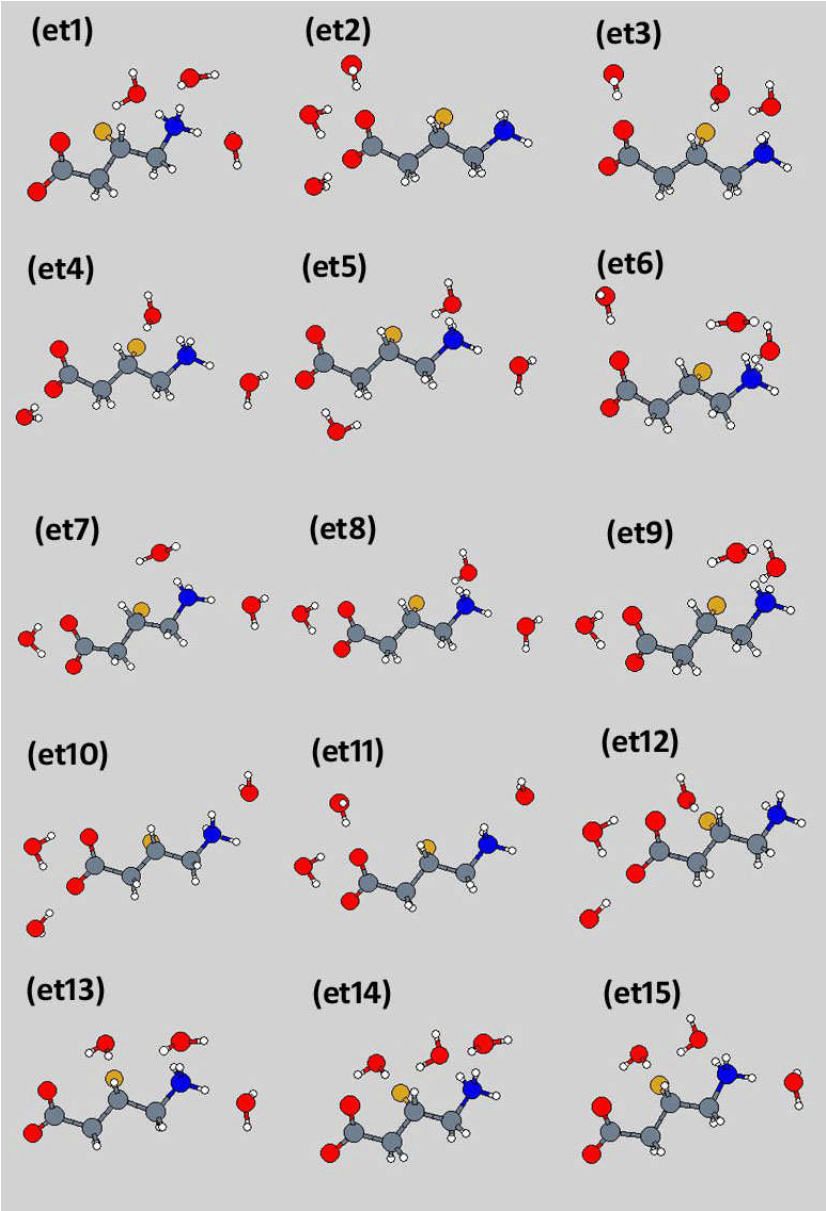


Figure 6.14: Starting structures of extended three-water molecule models for zwitterionic 3F-GABA

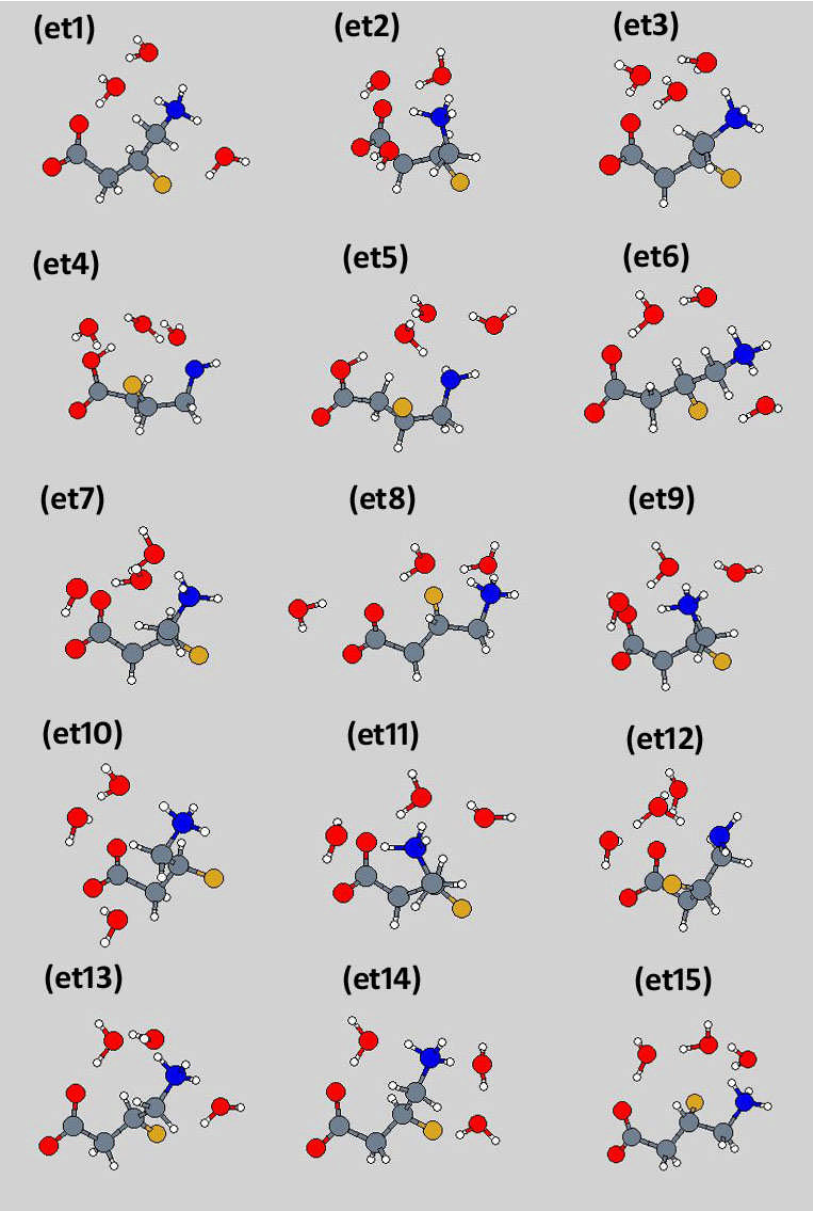
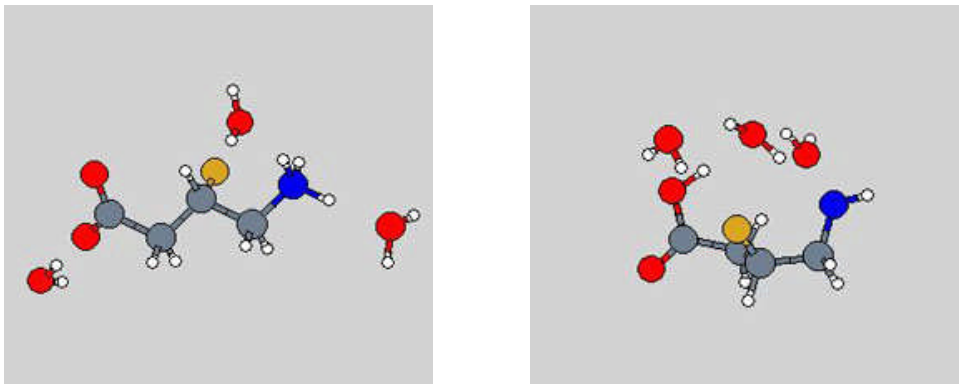


Figure 6.15: Optimized structures of extended three-water molecule models for zwitterionic 3F-GABA



(a) (b)

Figure 6.16: Three-water molecule model of extended zwitterionic 3F-GABA (et4) (b) Before optimization (b) After optimization

Figures 6.14 and 6.15 show the starting and optimized structures of the extended three-water molecule models for zwitterionic 3F-GABA. Only two of them, et4 and et5, became neutral after optimization, as is shown in Figure 6.16 by taking et4 as an example. From Figure 6.9 we can see that after optimization most of the zwitterionic 3F-GABAs have changed to some extent, as in the two-water molecule models' situation. For example, et6, et8 and et15 still represent fully extended structures, but with some bonds rotated; et1, et3, et7, et10, et12, et13, et14 are "bent", extended structures (with a hydronium ion is created in et12). The remaining three structures, et2, et9 and et11, are folded, with the water molecules forming a hydrogen-bonded bridge between the two-oxygen group and the NH_3 group, which is the most significant change that the optimization procedure may incur.

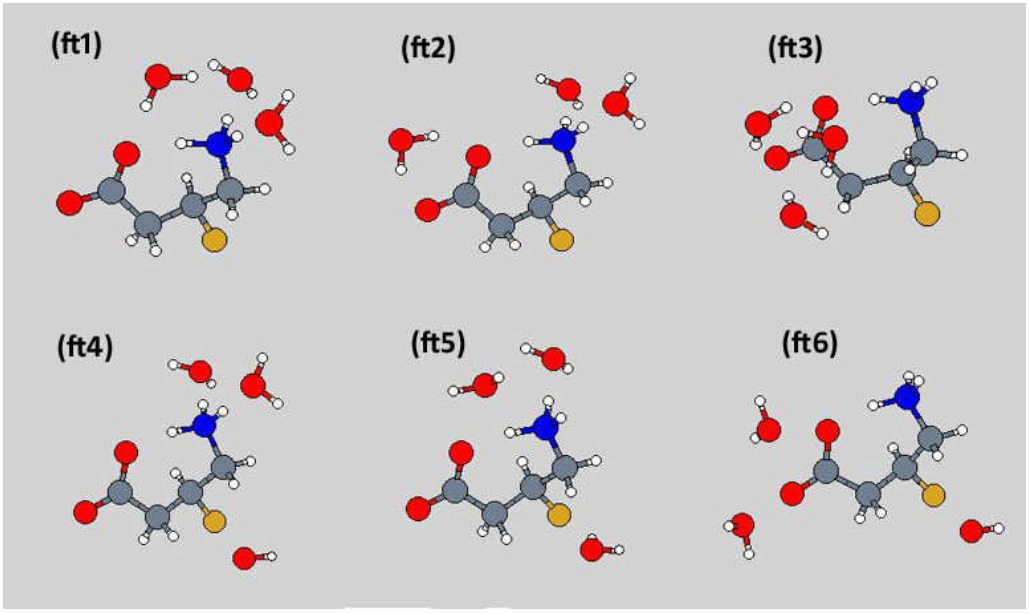


Figure 6.17: Starting structures of folded three-water molecule models for zwitterionic 3F-GABA

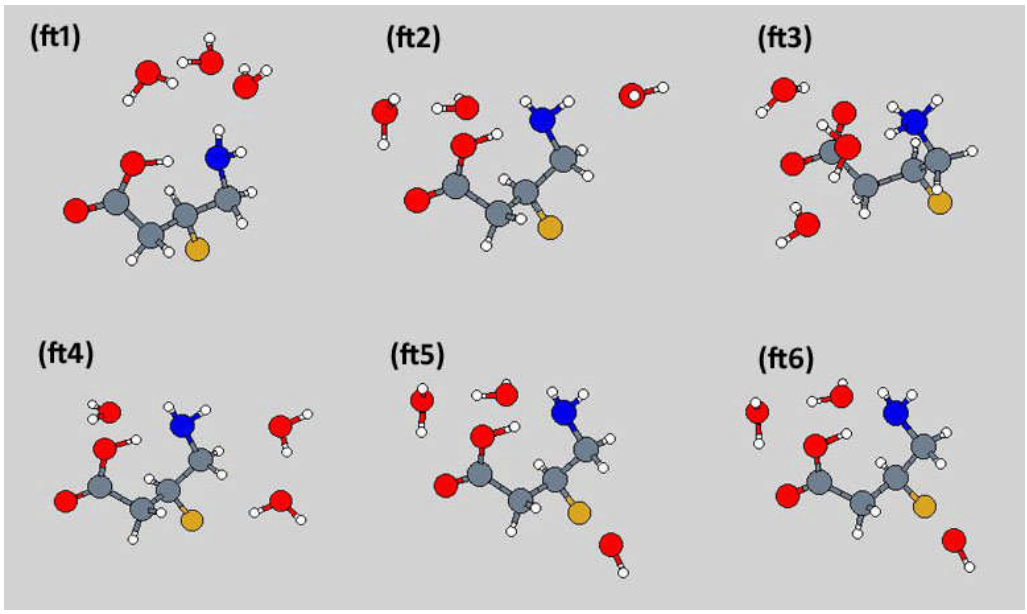


Figure 6.18: Optimized structures of folded three-water molecule models for zwitterionic 3F-GABA

Figures 6.17 and 6.18 show the starting and optimized structures of the folded

three-water molecule models for zwitterionic 3F-GABA. The ft1 and ft2 models started with two water molecules near the NH_3 group, hoping to form hydrogen bonds with it to prevent the hydrogen transition from NH_3 to one of those oxygen atoms. However, the optimized results showed the other way. On the other hand, ft3, with one water molecule between the oxygen atom and the NH_3 group and the other two water molecules near the oxygen atoms, proved to maintain the zwitterionic form of 3F-GABA, which may be due to the hydrogen-bonded chain of three water molecules. The other three models, ft4-ft6 with one water molecule near the fluorine atom, proved to be unstable and converged to neutral forms after optimization.

TABLE 6.5: Relative optimized energies △E, relative counterpoise-corrected optimized energies △CPE, BSSE values and structural features of three-water molecule models. All energies are in kJ/mol and are relative to the lowest optimized energy model at the M06-2X/6-31+G* level of theory.

3F-GABA-3H_2O	Starting structures	△E	△CPE	BSSE	Struct. features[a]
T1	ft3	0.00	0.00	-26.13	ZW, F
T2	ft1	8.17	0.01	-17.96	N, F
T3	et2	7.57	5.10	-23.66	ZW, F
T4	ft5, ft6	14.17	5.44	-17.39	N, F
T5	et4	7.88	5.68	-23.92	N
T6	ft2	17.35	8.70	-17.47	N, F
T7	et5	14.43	10.24	-21.93	N
T8	et9, et11	20.68	17.22	-22.66	ZW, F
T9	et7	21.31	20.84	-25.66	ZW
T10	ft4	36.51	24.25	-13.88	N, F
T11	et3	29.29	29.62	-26.46	ZW
T12	et12	49.63	53.49	-29.99	ZW
T13	et1, et13	60.09	54.23	-20.27	ZW
T14	et10	65.15	61.38	-22.36	ZW
T15	et14	81.28	77.95	-22.80	ZW
T16	et6	88.67	87.14	-24.60	ZW, E
T17	et15	100.53	99.34	-24.93	ZW, E
T18	et8	146.95	139.67	-18.85	ZW, E

[a] N = neutral; ZW = zwitterionic; F = folded; E = extended

Table 6.5 lists the relative optimized energies, relative counterpoise-corrected optimized energies, BSSE values and optimized structures of the three-water molecule models. The unique optimized structures are labeled T1-T18 and are ordered according to increasing counterpoise-corrected optimized energy. The “starting structures” column shows which models converged to these. The optimized energies indicate that ft5 and ft6 are similar, so are et1 and et13, et9 and et11. In total we obtained six neutral structures and 12 zwitterions. Relative counterpoise-corrected energies were calculated for these conformers and compared to the relative optimized energies, as is shown in Table 6.5 and Figure 6.19, indicating that counterpoise correction does change the relative energies significantly for some of the conformers.

Looking at Table 6.5, we can see the most stable conformer is a folded zwitterionic structure optimized from ft3, which contains a hydrogen-bonded triangle ring of water molecules.

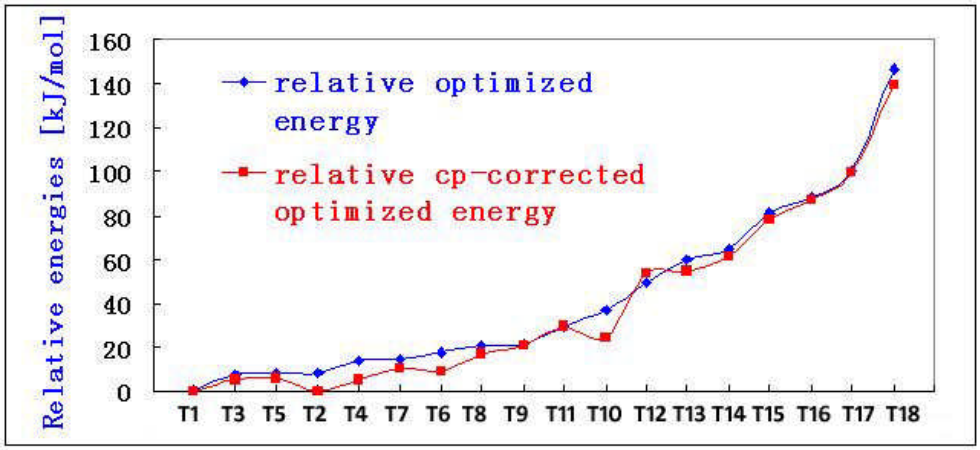


Figure 6.19: Comparison of relative optimized energies and relative counterpoise corrected energies of three-water molecule models

TABLE 6.6: Dihedral analysis of the unique zwitterions obtained from the three-water molecule models, compared with the extended conformer F and folded conformer B, optimized with B3LYP/6-31+G*/PCM [122].

3F-GABA • $3H_2O$ models	O-C1-C2-C3 dihedral angle	C1-C2-C3-F dihedral angle	C1-C2-C3-C4 dihedral angle	C2-C3-C4-N dihedral angle
conformer F	-34.65	-66.60	175.72	176.73
conformer B	-26.69	-169.22	74.21	-80.54
T1, folded	10.26	-169.64	73.72	-64.40
T3, folded	65.81	170.54	53.58	-83.99
T8, folded	64.60	166.23	48.52	j-84.61
T11, “bent” extended	55.23	173.42	54.54	-169.13
T12, “bent” extended	72.12	35.07	-84.86	152.75
T13, “bent” extended	4.17	-172.06	68.43	-160.77
T14, “bent” extended	62.44	175.55	55.55	-172.83
T15, “bent” extended	4.38	-173.68	68.11	-158.97
T16, extended	-54.04	-96.07	144.67	-173.25
T17, extended	-27.17	-66.28	175.72	172.06
T18, extended	-15.77	-67.13	170.25	-179.67

Table 6.6 lists the dihedrals of the zwitterions obtained from the three-water molecules model optimizations, as well as those of the extended conformer F and folded conformer B optimized with B3LYP/6-31+G*/PCM. It shows that T17 and T18 are quite similar to conformer F, whereas only one folded zwitterion T1 is similar to conformer B.

6.4.4 Five-water molecule models

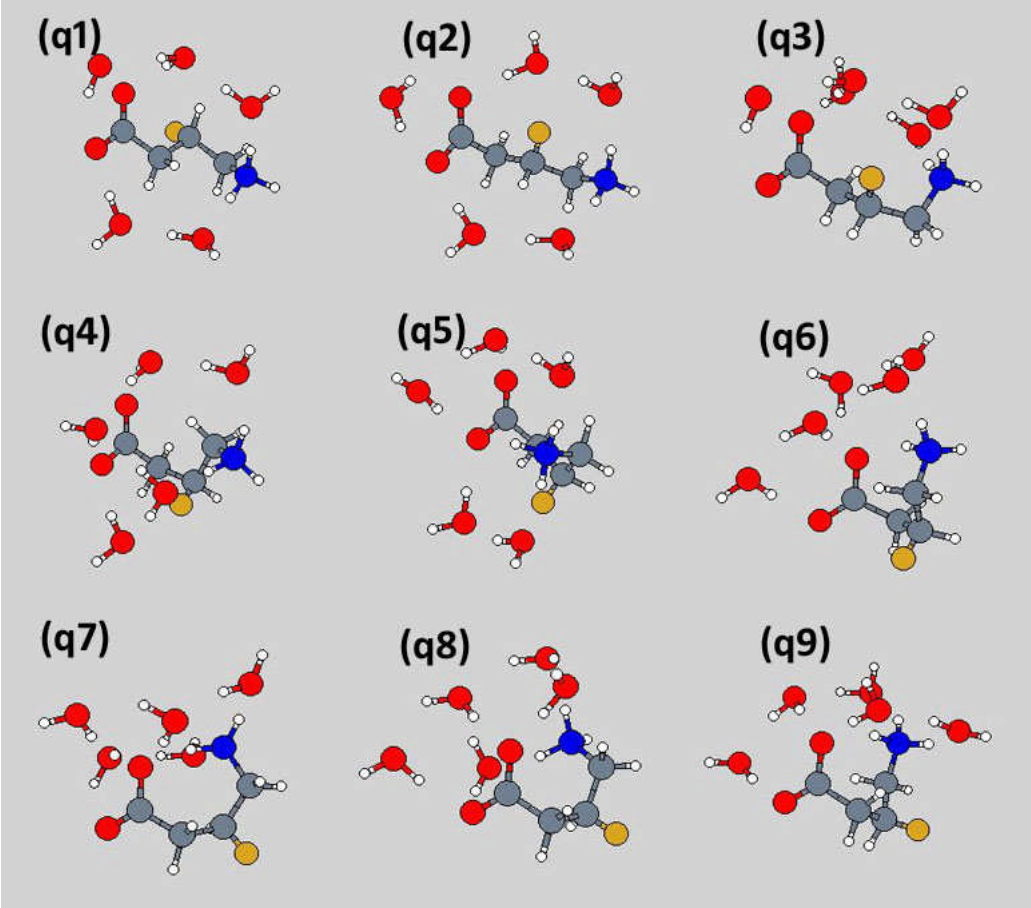


Figure 6.20: Starting structures of five-water molecule models for zwitterionic 3F-GABA

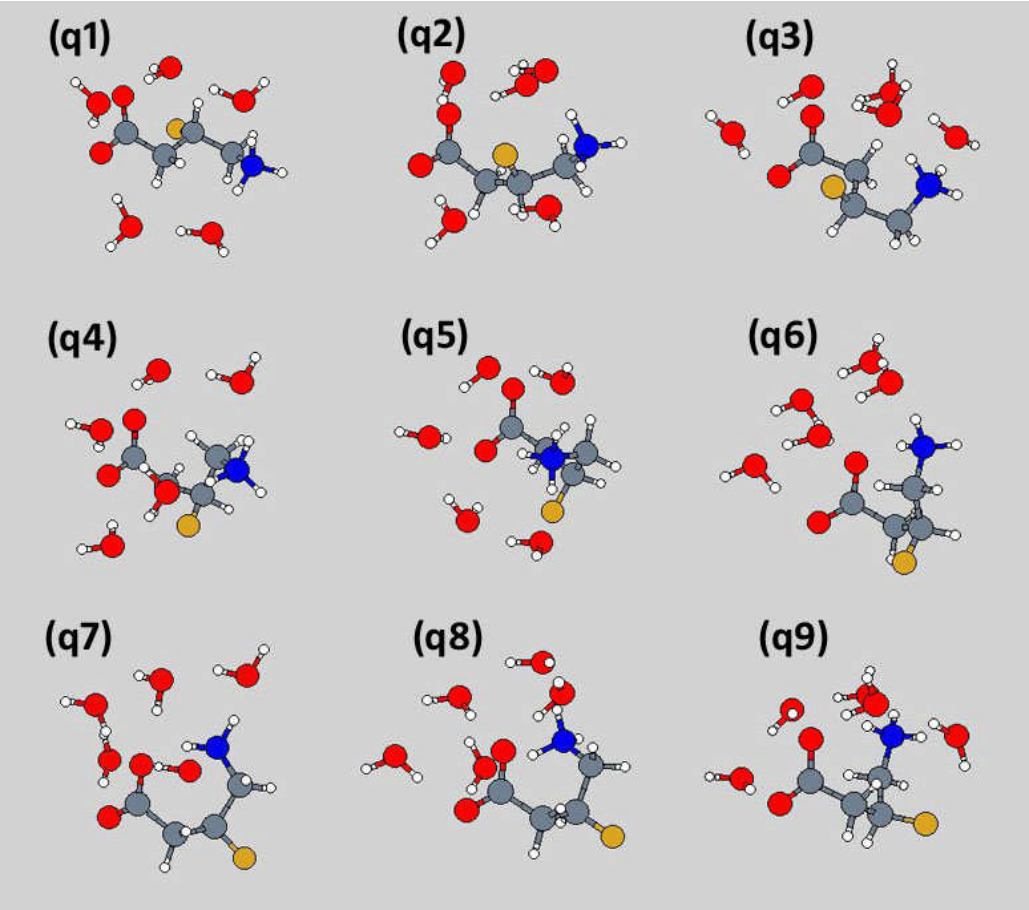


Figure 6.21: Optimized structures of five-water molecule models for zwitterionic 3F-GABA

Figures 6.20 and 6.21 show the starting and optimized structures of the five-water molecule models for zwitterionic (S)-3F-GABA created based on the nine MP2/6-31+G* optimized stable zwitterionic GABA • $5H_2O$ conformers from a previous study by Crittenden et al. [123]. After optimization using M06-2X/6-31+G*, all the 3F-GABAs kept their zwitterionic form. Only q2 and q3 changed geometries after optimization; all the other structures remained almost the same.

TABLE 6.7: Relative optimized energies △E, relative counterpoise-corrected optimized energies △CPE, BSSE values and structural features of five-water molecule models. All energies are in kJ/mol and are relative to the lowest optimized energy model at the M06-2X/6-31+G* level of theory.

3F-GABA-5H_2O	Starting structures	△E	△CPE	BSSE	Struct. features[a]
Q1	q9	0.00	0.00	-37.85	ZW, F
Q2	q8	-2.22	1.26	-41.34	ZW, F
Q3	q7	-0.85	1.67	-40.37	ZW, F
Q4	q5	5.63	7.99	-40.21	ZW, F
Q5	q6	18.11	18.76	-38.49	ZW, F
Q6	q4	16.76	22.49	-43.58	ZW
Q7	q2	55.65	59.74	-41.93	ZW, E
Q8	q3	58.42	62.30	-41.72	ZW
Q9	q1	58.96	62.95	-41.83	ZW

[a] N = neutral; ZW = zwitterionic; F = folded; E = extended

Table 6.7 lists the relative optimized energies, relative counterpoise-corrected energies, BSSE values and structural features of the five-water molecule models. The unique optimized structures are labeled Q1-Q9 and are ordered according to increasing counterpoise-corrected optimized energy. In total we obtained five folded zwitterions, three “bent” extended zwitterions and one extended zwitterion. Relative counterpoise corrected energies were calculated for these conformers and compared to the relative optimized energies, as shown in Table 6.7 and Figure 6.22, indicating that counterpoise correction does change the relative energies to some extent due to the relatively large intramolecular BSSE for the five-water molecule models, as shown in the BSSE value column of Table 6.7. It can be seen that after optimization the folded forms are preferred. The most stable conformer is a folded zwitterion.

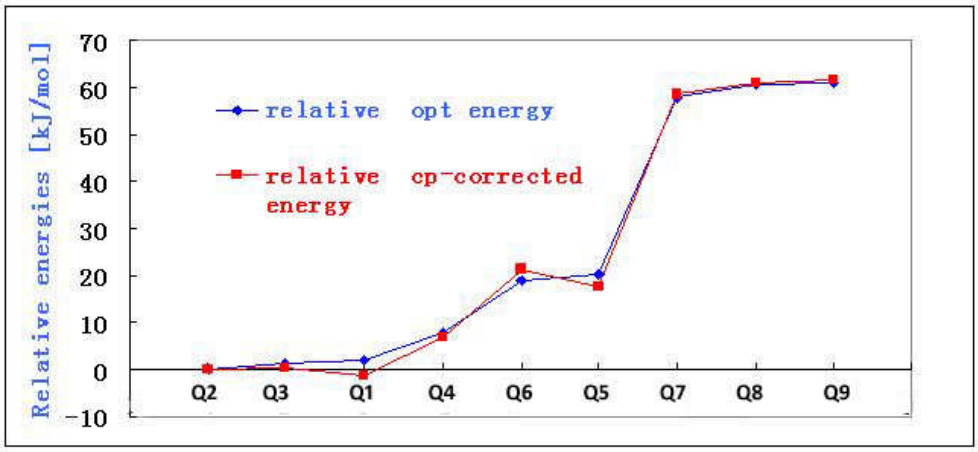


Figure 6.22: Comparison of relative optimized energies and relative counterpoise corrected energies of five-water molecule models

6.4.5 Relative stability comparison of different solvation models

TABLE 6.8: Relative stability comparison of conformer F and conformer B optimized with B3LYP/6-31+G*/PCM [122] and explicit water models similar to them.

Models	CP-corrected optimized energies	energy difference [kJ/mol]
PCM model:		
conformer F	-462.343240	19.01
conformer B	-462.335999	
$2H_2O$ models:		
D13, folded	-614.9134757	164.38
D23, extended	-614.8508677	
$3H_2O$ models:		
T1, folded	-691.3225154	99.34
T17, extended	-691.2846805	
$5H_2O$ models:		
Q3, folded	-844.1356654	58.07
Q7, extended	-844.1135490	

Table 6.8 lists the CP-corrected energies of the zwitterions obtained from the two, three and five-water molecule models which are similar to conformer F and conformer B, as well as those of conformer F and conformer B optimized with B3LYP/6-31+G*/PCM. From the comparison of the relative energy differences, we

can see that for both solvation models, folded conformers are more stable than the extended ones; all the relative energy differences of the explicit water molecules models are much larger than that from the calculation using PCM; however, the more explicit water molecules added in, the smaller the relative energy difference becomes. Therefore, it may be that many more waters need to be included to obtain converged results.

6.4.6 Hydrogen transition profiles

In previous work, Crittenden et al. found that folded zwitterionic GABA can be stabilized using two-water molecule models, by optimization using MP2/6-31+G*. However, my study shows that two-water molecule models are not sufficient to keep the folded conformer B in its zwitterionic form using M06-2X/6-31+G* as the optimization method. A further investigation shows that with M06-2X/6-31+G* optimization, folded zwitterionic GABA two-water molecule models also cannot be stabilized. To investigate this further, we calculated the hydrogen transition profiles of one of the GABA dihydrate conformers, zw7, using a variety of methods.

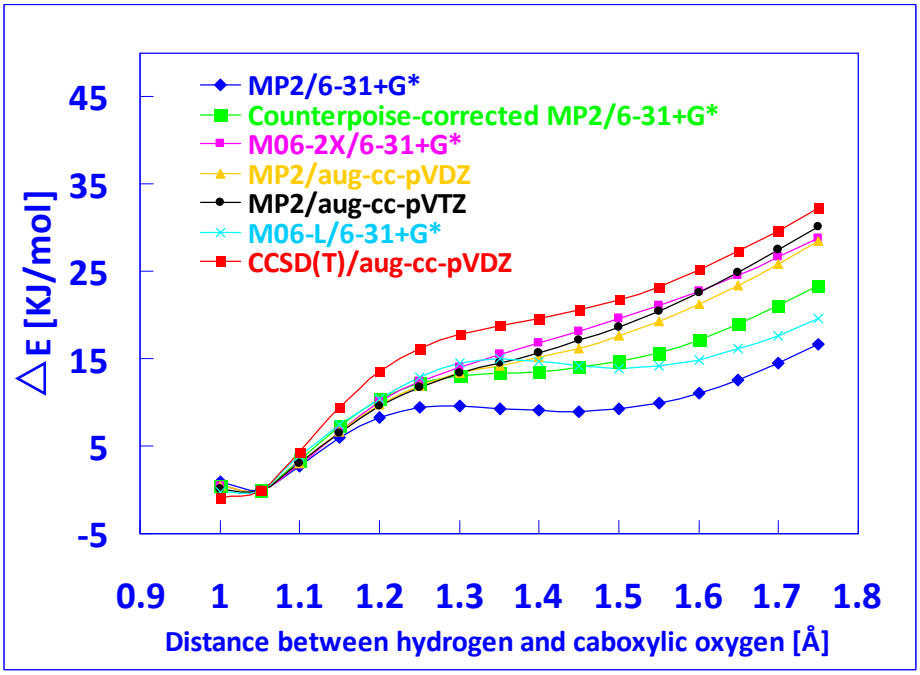


Figure 6.23: Hydrogen transition profile of GABA zw7

Figure 6.23 shows the hydrogen transition profiles of GABA conformer zw7. All the profiles present a minimum at about 1.05 Å, corresponding to the zwitterionic stable form of GABA. In addition, the MP2/6-31+G* curve shows another shallow minimum at about 1.45 Å, indicating the existence of a second stable geometry which corresponds to the neutral form. However, after counterpoise correction or applying larger basis sets such as aug-cc-pVDZ and aug-cc-pVTZ, the minimum at about 1.45 disappears, which means the second minimum is not real, caused by BSSE and the use of limited basis sets. As seen from the graph, the counterpoise correction removes the MP2/6-31+G* curve's second minimum, but compared with the other curves, it still appears to be much flatter near the former minimum. All the other MP2 curves with larger basis sets present similar gradients along all the distances, indicating that the basis set has more influence than counterpoise correction on the transition energy profile. It can also be seen that the MP2 curves with larger basis sets are in good agreement with the M06-2X/6-31+G* curve, indicating that even with the smaller basis set, M06-2X still yields the proper result. In addition, the M06-L/6-31+G* curve

shows a minimum on the right part similar to MP2/6-31+G*, which may be due to the non-hybrid character of M06-L. By comparing the curves with the reference transition energy profile, CCSD(T)/aug-cc-pVDZ, it can be concluded that the M06-2X/6-31+G* curve is very close to the reference curve and therefore is sufficiently accurate to calculate the transition energy profile in this circumstance.

6.5 Conclusions

We have optimized single-, two-, three- and five-water 3F-GABA models and extracted their energies after optimization.

It is shown that for the single-water models, the most stable structure is a neutral folded conformer. There are no stable folded zwitterions and all neutral complexes are more stable than the zwitterionic ones.

The optimized structures and energies of the two-water models reveal that the most stable structure is a neutral folded conformer as well. The most stable zwitterion has a folded structure which contains an NH…OH…OCO…HO…HN H-bonding cycle, but with no intermolecular H-bond. All other zwitterions are less stable than this one.

Three-water molecule models show that the most stable structure is a zwitterionic folded conformer. All the extended or “bent” extended zwitterions are less stable than the folded zwitterions.

Five-water molecule models were built based on the MP2/6-31+G* optimized GABA • $5H_2O$ zwitterionic structures from a previous study by Crittenden et al. [123] After M06-2X/6-31+G* optimizations, two models, q2 and q3, change geometries, although all of them still remain zwitterionic. This shows that the optimized geometries will differ according to the levels of theory used for the geometry optimizsation.

MP2/6-31+G* falsely yields folded zwitterionic GABA dihydrates due to BSSE

which means the basis set does affect the transition energy profiles.

M06-2X/6-31+G* gives better transition energy profiles than MP2/6-31+G* and its counterpoise-corrected profiles.

Chapter 7 Conclusion

In this thesis density functional methods were assessed for computing structures and energies of organic and bioorganic molecules.

Previous studies on the Tyr-Gly conformation had found that B3LYP and MP2 geometry optimizations yield dramatically different structures for some of the Tyr-Gly conformers [5-6]. Particularly two of the conformers (book4 and book6) have been studied in detail in former works, leading to an explanation for the optimization discrepancy: large intramolecular basis set superposition errors (BSSEs) in the MP2 calculations and the lack of dispersion in the B3LYP calculations may lead to wrong structures.

In the current work we investigated intramolecular BSSE using the fragmentation method and three kinds of rotation methods (the original method, Palermo's extended rotation method and a third rotation method proposed by us). The BSSE energies from the original and the third rotation method turned out to be too negative while that from Palermo's extended rotation method appeared to have negative as well as positive values (which was expected, based on the definition of BSSE in this method). Because we assume BSSE should be negative, the rotation methods cannot be used to correct intramolecular BSSE along a rotation profile.

In addition, we employed modern density functionals to overcome the deficiencies of B3LYP and MP2 using Tyr-Gly conformer 'book6' as the research target. First, the conformer was optimized at various fixed values of a distance controlling the degree of foldedness of the structure. Then the potential energy profiles were calculated using different methods. The results showed that the M06-L/6-31G(d) and M06-L/6-31+G(d) methods yielded profiles that excellently agree with the reference df-LCCSD(T0)/aug-cc-pVTZ profile, which entitles M06-L a very promising method to study the potential energy surface of small peptides containing aromatic residues.

To predict the gas-phase structure of Tyr-Gly, we created 108 potential conformers with a Fortran program using a recursive procedure taken from previous work. After that, two schemes were performed: the first one was to optimize the first 30 conformers ranked according to the M06-L/6-31+G(d) single-point energies; the second one was to optimize all of the 108 conformers. The methods used were M06-L/6-31G(d) and M05-2X/6-31+G(d). We selected the most stable conformers from both schemes and compared them to the 20 stable conformers found by B3LYP geometry optimizations in the previous study. The results show that both schemes find 10 conformers similar to one of the B3LYP stable conformers, as well as several newly found conformers. Relative transition energy profiles using various levels of theory were calculated for one of the missing B3LYP conformers. The results indicate that the missing of some B3LYP conformers may be due to the lack of dispersion in B3LYP theory.

For now the investigation of peptides was in the gas phase in this work, but since in biological environments water is present, the structure prediction of peptides ultimately has to take water into account. As an initial investigation of the effect of hydration on a flexible molecule, we studied the hydration effect by investigating the conformations of neutral and zwitterionic 3-fluoro-γ-aminobutyric acid (3F-GABA) in aqueous solution using explicit water molecule models. From the results we conclude that no single-water molecule models can stabilize folded zwitterions. Nevertheless, in five-water molecule models, all the zwitterions considered were stable, meaning that with more water molecules included, the zwitterionic forms of 3F-GABA tend to become more stable. All the relative stability energy differences of the explicit water molecule models are much larger than that from the calculation using the PCM continuum solvation model. However, the more explicit water molecules added in, the smaller the relative energy difference becomes. Finally, it was found that M06-2X/6-31+G(d) yield better results in calculating transition energy profiles than MP2/6-31+G(d).

In conclusion, in this thesis we have performed investigations to solve the problem of the differences between B3LYP and MP2 optimized structures for some of the Tyr-Gly conformers and the hydration effect of 3F-GABA in solution in order to assess different density functional theories. First of all, the investigation of intramolecular BSSE showed that none of the rotation methods tested correctly correct BSSE along a rotation profile. In addition, it appeared that M06-L with basis sets 6-31G(d) or 6-31+G(d) can obtain satisfying transition energy profiles for some Tyr-Gly conformers. Besides, with M06-L and M05-2X, we found several new stable conformers compared to the former 20 stable conformers found by B3LYP. Finally, with the investigation of 3F-GABA using different solvation models, it was found that M06-2X/6-31+G(d) showed a better performance in calculating transition energy profiles than MP2/6-31+G(d).

Chapter 8 References

[1] P. W. Atkins and R. Friedman, *Molecular Quantum Mechanics* 4ed. Oxford: Oxford University Press, 2005.

[2] P. W. Atkins, *Physical Chemistry* Oxford: Oxford University Press, 2002.

[3] J. Bernal and D. Crowfoot, *Nature*, vol. 133, p. 794, 1934.

[4] K. Wüthrich, *J. Biol. Chem.*, vol. 265, p. 22059, 1990.

[5] L. F. Holroyd and T. van Mourik, *Chem. Phys. Lett.*, vol. 442, p. 42, 2007.

[6] A. E. Shields and T. van Mourik, *J. Phys. Chem. A*, vol. 111, p. 13272, 2007.

[7] T. van Mourik, P. G. Karamertzanis, and S. L. Price, *J. Phys. Chem. A*, vol. 110, p. 8, 2006.

[8] T. van Mourik, *J. Chem. Theor. Comput.*, vol. 4, p. 1610, 2008.

[9] S. F. Boys and F. Bernardi, *Mol. Phys.*, vol. 19, p. 553, 1970.

[10] I. N. Levine, *Quantum Chemistry*. Englewood Cliffs, New jersey: Prentice Hall, 1991.

[11] C. J. Cramer, *Essentials of Computational Chemistry*. Chichester: John Wiley & Sons, Ltd., 2002.

[12] F. Jensen, *Introduction to Computational Chemistry*. Chichester, England: John Wiley and Sons, 2007.

[13] G. Vignale and M. Rasolt, *Phys. Rev. Lett.*, vol. 59, p. 2360, 1987.

[14] E. Schrödinger, *Phys. Rev.*, vol. 28(6), p. 1049, 1926.

[15] W. G. Richards and D. L. Cooper, *Ab Initio Molecular orbital Calculations for Chemists*. Oxford: Clarendon Press, 1983.

[16] C. D. Sherrill and H. F. Schaefer III, *"The Configuration Interaction Method: Advances in Highly Correlated Approaches."* vol. 34. New York: Academic press, 1999.

[17] C. Møller and M. S. Plesset, *Phys. Rev.*, vol. 46, p. 618, 1934.

[18] T. Helgaker, P. Jorgensen, and O. Jeppe, *Molecular Electronic Structure Theory*. Chichester: Wiley, 2000.

[19] J. Čížek, *J.Chem. Phys.*, vol. 45, p. 4256, 1966.

[20] O. Sinanoğlu and K. Brueckner, *Three approaches to electron correlation in atoms.* New Haven, United States: Yale Univ. Press, 1971.

[21] O. Sinanoğlu, *J. Chem. Phys.*, vol. 36, p. 706, 1962.

[22] Hückel, *Theory for Organic Chemists*. London: Academic Press, 1978.

[23] A. Streitwieser, *Molecular Orbital Theory for Organic Chemists.* New York: Wiley, 1961.

[24] R. Dreizler and E. Gross, *Density Functional Theory* New York: Plenum Press, 1995.

[25] J. Kohanoff, *Electronic Structure Calculations for Solids and Molecules: Theory and Computational Methods* Cambridge University Press, 2006.

[26] F. Jensen, *Introduction to Computational Chemistry*. Chichester, England: John Wiley and Sons, 2004.

[27] T. Ziegler, *Chem. Rev.*, vol. 91, p. 651, 1991.

[28] R. G. Parr, *Annu. Rev. Phys. Chem.*, vol. 34, p. 631, 1983.

[29] H. C. Chermette, *Chem. Rev.*, vol. 180, p. 699, 1998.

[30] X. Gonze, *Phys. Rev. A*, vol. 52, p. 1096, 1995.
[31] E. J. Baerends and O. V. Gritsenko, *J. Phys. Chem. A,* vol. 101, p. 5383, 1997.
[32] W. Kohn, A. D. Becke, and R. G. Parr, *J. Phys. Chem.*, vol. 100, p. 12974, 1996.
[33] R. G. Parr and W. T. Yang, *Annu. Rev. Phys. Chem.*, vol. 46, p. 701, 1995.
[34] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, *J. Phys. Chem.*, vol. 98, p. 11623, 1994.
[35] K. Kim and K. D. Jordan, *J. Phys. Chem.*, vol. 98(40), p. 10089, 1994.
[36] Y. Zhao, B. J. Lynch, and D. G. Truhlar, *J. Phys. Chem. A,* vol. 108, p. 2715, 2004.
[37] Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A,* vol. 108, p. 6908, 2004.
[38] E. E. Dahlke and D. G. Truhlar, *J. Phys. Chem. B,* vol. 109, p. 15677, 2005.
[39] Y. Zhao, N. E. Schultz, and D. G. Truhlar, *J. Chem. Theor. Comput.*, vol. 2, p. 364, 2006.
[40] Y. Zhao and D. G. Truhlar, *Acc. Chem. Res.*, vol. 41, p. 157, 2008.
[41] E. R. Fermi, *Accad. Naz. Lincei*, vol. 6, p. 602, 1927.
[42] L. H. Thomas, *Proc. Cambridge Philos. Soc.* , vol. 23, p. 542, 1927.
[43] J. C. Slater, *Phys. Rev.*, vol. 81, p. 385, 1951.
[44] P. Hohenberg and W. Kohn, *Phys. Rev. B,* vol. 136, p. 864, 1964.
[45] W. Kohn and L. J. Sham, *Phys. Rev. A,* vol. 140, p. 1133, 1965.
[46] A. D. Becke, *J. Chem. Phys.*, vol. 98, p. 5648, 1993.
[47] C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B,* vol. 37, p. 785, 1988.
[48] X. Xu and W. A. Goddard III, *Proc. Natl. Acad. Sci.*, vol. 101, p. 2673, 2004.
[49] A. D. Becke, *J. Chem. Phys.*, vol. 107, p. 8554, 1997.
[50] A. D. Becke, *J. Chem. Phys.*, vol. 98, p. 1372, 1993.
[51] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, *J. Phys. Chem. A,* vol. 111, p. 10439, 2007.
[52] B. Santra, A. Michaelides, and M. Scheffler, *J. Chem. Phys.*, vol. 127, p. 184104, 2007.
[53] J. Cerny and P. Hobza, *Phys. Chem. Chem. Phys.*, vol. 7, p. 1624, 2005.
[54] Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A,* vol. 109, p. 6624, 2005.
[55] J. Csontos, N. Y. Palermo, R. F. Murphy, and S. Lovas, *J. Comput. Chem.*, vol. 29, p. 1344, 2008.
[56] Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A,* vol. 109, p. 5656, 2005.
[57] Y. Zhao and D. G. Truhlar, *Phys. Chem. Chem. Phys.*, vol. 7, p. 2701, 2005.
[58] Y. Zhao, O. Tischemko, and D. G. Truhlar, *J. Phys. Chem.*, vol. 109, p. 19046, 2005.
[59] Y. Zhao, N. E. Schultz, and D. G. Truhlar, *J. Chem. Phys.*, vol. 123, p. 161103, 2005.
[60] Y. Zhao and D. G. Truhlar, *J. Chem. Phys.*, vol. 125, p. 194101, 2006.
[61] Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A,* vol. 110, p. 13126, 2006.
[62] F. Jensen, *Introduction to Computational Chemistry*. Chichester, England: John Wiley and Sons, 1999.
[63] A. R. Leach, *Molecular Modelling: Principles and Applications*: Singapore: Longman, 1996.
[64] W. J. Hehre, *A Guide to Molecular Mechanics and Quantum Chemical Calculations*. Irvine, California: Wavefunction, Inc., 2003.
[65] M. B. Roman, *J. Chem. Phys.*, vol. 132, p. 231101, 2010.
[66] T. H. Dunning Jr. and P. J. Hay, *Modern Theoretical Chemistry* vol. 3. New York: H.F. Schaefer, Plenum, 1977.

[67] R. C. Binning and L. A. Curtiss, *J. Comp. Chem.*, p. 1206, 1990.
[68] W. J. Hehre, R. F. Stewart, and J. A. Pople, *J. Chem. Phys.*, vol. 51, p. 2657, 1969.
[69] W. J. Hehre, R. F. Stewart, and J. A. Pople, *J. Chem. Phys.*, vol. 52, p. 2769, 1970.
[70] T. H. Dunning, *J. Chem. Phys.*, vol. 90, p. 1007, 1989.
[71] D. E. Woon and J. T. H. Dunning, *J. Chem. Phys.*, vol. 98, p. 1358, 1993.
[72] D. E. Woon and J. T. H. Dunning, *J. Chem. Phys.*, vol. 100, p. 2975, 1994.
[73] A. K. Wilson, D. E. Woon, K. A. Peterson, and J. T. H. Dunning, *J. Chem. Phys.*, vol. 110, p. 7667, 1999.
[74] I. Mayer and P. Valiron, *J. Chem. Phys.*, vol. 109, p. 3360, 1998.
[75] M. Valiev, E. J. Bylaska, and N. Govind, *Comput. Phys. Commun*, vol. 181, p. 1477, 2010.
[76] F. Neese, *ORCA - an ab initio, density functional and semiempirical program package version 2.6, Revision 35*.
[77] N. Y. Palermo, J. Csontos, M. C. Owen, R. F. Mruphy, and S. Lovas, *J. Comput. Chem.*, vol. 28, p. 1208, 2007.
[78] T. van Mourik, *J. Comput. Chem.*, vol. 28, p. 1208, 2007.
[79] J. Csontos, N. Palermo, R. Murphy, and S. Lovas, *J. Comput. Chem.*, vol. 10, p. 20868, 2007.
[80] J. Cao and T. van Mourik, *Chem. Phys. Lett.*, vol. 485, p. 40, 2010.
[81] A. D. Becke, *Phys. Rev. A*, vol. 38, p. 3098, 1988.
[82] T. Van Mourik and R. J. Gdanitz, *J. Chem. Phys.*, vol. 116, p. 9620, 2002.
[83] W. Kohn, Y. Meir, and D. E. Makarov, *Phys. Rev. Lett.*, vol. 80, p. 4153, 1998.
[84] A. Millet, T. Korona, R. Moszynski, and E. Kochanski, *J. Chem. Phys.*, vol. 111, p. 7727, 1999.
[85] M. Gerhards, C. Unterberg, A. Gerlach, and A. Jansen, *Phys. Chem. Chem. Phys.*, vol. 6, p. 2682, 2004.
[86] R. J. Graham, R. T. Kroemer, M. Mons, E. G. Robertson, L. C. Snoek, and J. P. Simons, *J. Phys. Chem. A*, vol. 103, p. 9706, 1999.
[87] N. A. MacLeod and J. P. Simons, *Chem. Phys.*, vol. 283, p. 221, 2002.
[88] D. Toroz and T. van Mourik, *Mol. Phys.*, vol. 104, p. 559, 2006.
[89] S. Saebø, W. Tong, and P. Pulay, *J. Chem. Phys.*, vol. 98, p. 2170, 1993.
[90] C. Hampel and H.-J. Werner, *J. Chem. Phys.*, vol. 104, p. 6286, 1996.
[91] J. M. Pedulla, F. Vila, and K. D. Jordan, *J. Chem. Phys.*, vol. 105, p. 11091, 1996.
[92] B. Hartke, M. Schütz, and H.-J. Werner, *Chem. Phys.*, vol. 239, p. 561, 1998.
[93] M. Schütz, G. Rauhut, and H.-J. Werner, *J. Phys. Chem. A*, vol. 102, p. 5997, 1998.
[94] N. Runeberg, M. Schütz, and H.-J. Werner, *J. Chem. Phys.*, vol. 110, p. 7210, 1999.
[95] L. Magnko, M. Schweizer, G. Rauhut, M. Schütz, H. Stoll, and H.-J. Werner, *Phys. Chem. Chem. Phys.*, vol. 4, p. 1006, 2002.
[96] T. Hrenar, G. Rauhut, and H.-J. Werner, *J. Phys. Chem. A*, vol. 110, p. 2060, 2006.
[97] R. A. Mata and H.-J. Werner, *J. Chem. Phys.*, vol. 125, p. 184110, 2006.
[98] Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.* , vol. 120, p. 215, 2008.
[99] Y. Zhao and D. G. Truhlar, *J. Phys. Chem. C*, vol. 112, p. 6860, 2008.
[100] Y. Zhao and D. G. Truhlar, *J. Phys. Chem.*, vol. 125, p. 194101, 2006.
[101] R. Valero, R. Costa, I. D. P. R. Moreira, D. G. Truhlar, and F. Illas, *J. Chem. Phys.*, vol. 128, p.

114103, 2008.
[102] Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A*, vol. 112, p. 6794, 2008.
[103] E. E. Dahlke, R. M. Olson, H. R. Leverentz, and D. G. Truhlar, *J. Phys. Chem. A*, vol. 112, p. 3976, 2008.
[104] V. S. Bryantsev, M. S. Diallo, A. C. T. van Duin, and W. A. Goddard III, *J. Chem. Theor. Comput.*, vol. 5, p. 1016, 2009.
[105] D. Benitez, E. Tkatchouk, I. I. Yoon, J. F. Stoddart, and W. A. Goddard III, *J. Am. Chem. Soc.*, vol. 130, p. 14928, 2008.
[106] H. R. Leverentz and D. G. Truhlar, *J. Phys. Chem. A*, vol. 112, p. 6009, 2008.
[107] S. Grimme, *J. Chem. Phys.*, vol. 124, p. 034108, 2006.
[108] S. Grimme, *J. Comput Chem.*, vol. 25, p. 1463, 2004.
[109] S. Grimme, *J. Comput. Chem.*, vol. 27, p. 1787, 2006.
[110] M. Schütz, *J. Chem. Phys.*, vol. 113, p. 9986, 2000.
[111] M. Schütz and H.-J. Werner, *Chem. Phys. Lett.*, vol. 318, p. 370, 2000.
[112] M. Schütz and H.-J. Werner, *J. Chem. Phys.*, vol. 114, p. 661, 2001.
[113] O. Bludský, M. Rubeš, P. Solán, and P. Nachtigall, *J. Chem. Phys.*, vol. 128, p. 114102, 2008.
[114] H.-J. Werner, F. R. Manby, and P. J. Knowles, *J.Chem. Phys*, vol. 118, p. 8149, 2003.
[115] F. Grimme, *J. Chem. Phys.*, vol. 118, p. 9095, 2003.
[116] G. Hetzer, P. Pulay, and H.-J. Werner, *Chem. Phys. Lett.*, vol. 290, p. 143, 1998.
[117] G. Schaftenaar and J. H. Noordik, *J. Comput.-Aided Mol. Design*, vol. 14, p. 123, 2000.
[118] A. Abo-Riziq, L. Grace, B. Crews, M. P. Callahan, T. van Mourik, and M. S. de Vries, *J. Phys. Chem. A*, vol. 10, p. 1021, 2010.
[119] J. Cao, R. Bjornsson, M. Bühl, W. Thiel, and T. Van Mourik, *Chem. - Eur. J.*, Accepted, 2011
[120] H. P. Rang, M. M. Dale, and J. M. Ritter, *Pharmacology*. Edinburgh: Churchill Livingstone, 1999.
[121] G. Deniau, A. M. Z. Slawin, J. P. Issberner, T. van Mourik, K. T. Sillar, J. J. Lambert, and D. O'Hagan, *ChemBioChem*, vol. 8, p. 2265, 2007.
[122] S. Miertus, E. Scrocco, and J. Tomasi, *Chem. Phys.*, vol. 55, p. 117, 1981.
[123] D. L. Crittenden, M. Chebib, and M. J. Jordan, *J. Phys. Chem. A*, vol. 108, p. 203, 2004.
[124] Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, vol. 120, p. 215, 2008.

Publications:

1. J. Cao and T. van Mourik, “Performance of the M06-L density functional for a folded Tyr-Gly conformer”, Chem. Phys. Lett., vol. 485, p. 40, 2010.

2. J. Cao, R. Bjornsson, M. Bühl, W. Thiel, and T. Van Mourik, “Modelling zwitterionic in solution: 3-fluoro-γ-aminobutyric acid (3F-GABA)”, Chem. - Eur. J., Accepted, 2011.

Posters:

1. J. Cao and T. van Mourik, “The good performance of the M06-L functional for the structure prediction of a Tyr-Gly conformer.” ScotCHEM Computational Chemistry Symposia at Heriot-Watt University (Edinburgh), 13 May 2009.
2. J. Cao and T. van Mourik, “Modelling zwitterions in solution: 3-fluoro- γ -aminobutyric acid (3F-GABA).” ScotCHEM Computational Chemistry Symposia at University of Strathclyde (Glasgow), 8 April 2010.
3. J. Cao and T. van Mourik, “Modelling zwitterions in solution: 3-fluoro-gamma-aminobutyric acid (3F-GABA).” Graduate student meetings at University of Nottingham, 7 July 2010.

Presentations:

1. J. Cao, “Elucidation of the most stable conformations of the gas-phase tyrosine-glycine dipeptide.” Computation in Chemistry seminar at University of St Andrews, 14 December 2009.
2. J. Cao, “Modelling zwitterions in solution: 3-fluoro- γ -aminobutyric acid (3F-GABA).” Computation in Chemistry seminar at University of St Andrews, 15 November 2010.