THESIS

ONIOM-BSSE SCHEME FOR OH⋯π SYSTEM AND

APPLICATION ON HIV-1 REVERSE TRANSCRIPTASE

RATTAPON HONGKRENGKAI

GRADUATE SCHOOL, KASETSART UNIVERSITY

2006

THESIS APPROVAL

GRADUATE SCHOOL, KASETSART UNIVERSITY

Master of Science (Chemistry)
DEGREE

Physical Chemistry | Chemistry
FIELD | **DEPARTMENT**

TITLE: ONIOM-BSSE Scheme for OH···π System and
Application on HIV-1 Reverse Transcriptase

NAME: Mr. Rattapon Hongkrengkai

THIS THESIS HAS BEEN ACCEPTED BY

_____ THESIS ADVISOR
(Associate Professor Supa Hannongbua, Dr.rer.nat.)

_____ COMMITTEE MEMBER
(Miss Pensri Bunsawansong, Ph.D.)

_____ COMMITTEE MEMBER
(Assistant Professor Utsanee Leerawat, Ph.D.)

_____ DEPARTMENT HEAD
(Assistant Professor Yerry Mahatumarattana, B.Sc.)

APPROVED BY THE GRADUATE SCHOOL ON ___11_/_4_/_00b____

_____ DEAN
(Associate Professor Vinai Artkongharn, M.A.)

THESIS

ONIOM-BSSE SCHEME FOR OH···π SYSTEM AND APPLICATION ON HIV-1 REVERSE TRANSCRIPTASE

RATTAPON HONGKRENGKAI

A Thesis Submitted in Partial Fulfillment of

the Requirements for the Degree of

Master of Science (Chemistry)

Graduate School, Kasetsart University

2006

The OH···π interaction in alcohol-ethylene systems using ethanol, propanol, n-butanol or iso-butanol for alcohol molecules have been investigated by MP2(full)/ 6-311+G(d,p), MP2(full)/6-31G(d) and B3LYP/6-31G(d) levels of calculation. Two optimization schemes, namely standard and counterpoise (CP) gradient optimizations, have been used to determine the structures and energetics of the alcohol-ethylene complexes. The interaction energies determined by MP2(full)/6-311+G(d,p) calculation, with and without BSSE correction, are in range of -3.26 to -3.35 kcal/mol and -2.12 to -2.23 kcal/mol, respectively. The ONIOM method has been tested for its applicability on the OH···π system to decrease the computational demand; ONIOM2M(MP2(full)/6-311+G(d,p):MP2(full)/6-31G(d)) and ONIOM2B(MP2(full) /6-311+G(d,p):B3LYP/6-31G(d)), have been applied to alcohol-ethylene systems where the OH···π region is treated at the high level of calculation. The comparison between the MP2 and the ONIOM results showed that the ONIOM method can provide reasonable geometries and binding energies for OH···π systems. The ONIOM method has been utilized in the study of the structure and the binding energy of nevirapine at the binding site of HIV-1 reverse transcriptase (RT) both in the wild type and the Y181C mutant type structures. The present results clearly indicate that the Y181C substitution is more electrostatic repulsive than the wild type RT which results in the decreasing of the stabilization energy of nevirapine and its binding pocket.

Rattapon   Hongkrengkai   Supa Hannongbua.   07 / 04 / 2006
_____   _____
Student's signature   Thesis Advisor's signature

ACKNOWLEDGEMENTS

Rattapon Hongkrengkai

March 2006

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES (Cont'd)

LIST OF FIGURES

LIST OF FIGURES (Cont'd)

ABBREVIATIONS

Ala	=	Alanine
AOs	=	Atomic Orbitals
Arg	=	Arginine
Asn	=	Asparagine
Asp	=	Aspartic acid
B1	=	B3LYP/6-31G(d)
B3LYP	=	Becke's three parameter hybrid functional using the LYP
BBF	=	Backbone fixed
BLYP	=	Beck-Lee-Yang-Parr functional
Cys	=	Cysteine
DFT	=	Density Functional Theory
EE	=	Ethanol-ethylene system
FC	=	frozen-core
Gln	=	Glutamine
Glu	=	Glutamic acid
Gly	=	Glycine
GTO	=	Gaussian-Type Orbital
HAF	=	Hetero atom fixed
HF	=	Hartree-Fock Theory
His	=	Histidine
HIV-1	=	Human Immunodeficiency Virus Type 1
HOMO	=	Highest Occupied Molecular Orbital
iBE	=	iso-Butanol-Ethylene system
Ile	=	Isoleucine
IMOMO	=	Integrated molecular orbital molecular mechanics
kDa	=	Kilo Dalton
KS	=	Kohn-Sham
LCAO	=	Linear Combination of Atomic Orbitals
LCAO-MO	=	Linear Combination of Atomic Orbitals to Molecular
LDA	=	Local Density Approximation

ABBREVIATIONS (Cont'd)

Leu	=	Luecine
LSD	=	Local Spin Density Approximation
LUMO	=	Lowest Unoccupied Molecular Orbital
LYP	=	Lee-Yang-Parr functional
Lys	=	Lysine
M1	=	MP2(full)/6-31G(d)
M11	=	MP2(full)/6-311+G(d,p)
Met	=	Methionine
MM	=	Molecular mechanics
MO	=	Molecular Orbital
MOs	=	Molecular Orbitals
MP2	=	The second-order Møller-Plesset perturbation theory
MT	=	Mutant Type
nBE	=	n-Butanol-ethylene system
Nevirapine	=	11-cyclopropyl-5,11-dihydro-4-methyl-6H-dipyrido[3,2-b:2', 3'-e(1,4)diazopin-6-one
ONIOM	=	our own n-layered integrated molecular orbital and molecular mechanics
PE	=	Propanol-Ethylene system
Phe	=	Phenylalanine
PM3	=	Parametric Method Number 3
Pro	=	Proline
QM	=	Quantum mechanics
QM/MM	=	Quantum mechanics and molecular mechanics
RNA	=	Ribonucleic acid
RT	=	Reverse Transcriptase
SCF	=	Self-Consistent Field
Ser	=	Serine
STO	=	Slater Type Orbital
STO-3G	=	Slater Type Orbital approximated by 3 Gaussian type orbitals

ABBREVIATIONS (Cont'd)

Thr	=	Threonine
TIBO	=	Tetrahydro-imidazo[4,5,1-*jk*] [1,4] –Benzodiazepin-2 (1H) - One and –thione analogues
Trp	=	Tryptophan
Tyr	=	Tyrosine
Val	=	Valine
WT	=	Wild Type

ONIOM-BSSE SCHEME FOR OH⋯π SYSTEM AND APPLICATION ON HIV-1 REVERSE TRANSCRIPTASE

INTRODUCTION

One of the major goals of chemistry in recent times has been the investigation and understanding of week intermolecular interactions. These interactions are very important in diverse fields ranging from organic chemistry, organometalic, biomolecular structure, supramolecular assembly, crystal packing, reaction selectivity, molecular recognition, drug-receptor interaction and so forth. On the basis of these interactions, not only theoretical studies but also experimental realizations of novel functional molecules, a nanomaterial and molecular device have become possible (Kim et al., 2005). The conventional H-bond (A-H⋯B), wherein both the donor A and acceptor B are electronegative atoms, has been experimentally and theoretically characterized in fine detail. These characterizations also include variants of the conventional H-bond like short strong H-bonds and dihydrogen bonds. However, little is known about weak interactions involving π systems such as the π⋯π and π⋯H interactions (Tarakeshwar et al., 2000).

π-H interactions involving π systems are interesting both from a practical and theoretical point of view. This is because they can be used as model systems to examine the nature of hydrophobicity at the molecular level and help distinguish the limits of what should or should not be considered as "hydrogen bonds". Although a number of experimental studies of weak interactions involving π systems have invoked the π-H interactions to explain conformational preferences in organic and bio-molecules, crystal packing, host-guest complexation, it has been difficult to estimate their intrinsic stability in vivo. Thus, theoretical calculations are the only recourse to obtain reliable estimates of their stability and understand their origin. However, the lack of detailed investigations of these interactions has given rise to different interpretations of the π H-bond with earlier studies attributing the formation of a π H-bond to be due to electrostatic interactions, with the π-cloud of the olefinic or aromatic system behaving as an electron donor or proton acceptor. As a result of this interpretation, the strength of the π⋯HY interaction has been entirely correlated to

the electronegativity of Y with the nature of the π system (olefinic or aromatic) having no role therein. Given the number of recent studies highlighting the importance of dispersive forces in governing the strength of interactions involving π systems, it becomes important to obtain a reliable estimate of the contributions of these individual interacting forces for π-H interactions of different π systems (Kim *et al*., 2000). This detailed understanding of the physical origin of the π-H interaction, apart from resolving many of the misleading notions of the π-H interaction, would also aid the development of force-fields capable of accurately representing these interactions.

Theoretical study intermolecular H-bond is difficult, and the most accurate techniques should be applied. Consider the structure evaluation and particularly the influence of the basis set saturation effect. Routinely the structure of a complex is optimized by a gradient technique on the basis set superposition error (BSSE)-uncorrected potential-energy surface and the final stabilization energy determined for the calculated structure by a posteriori applying BSSE corrections, e.g., the function counterpoise (CP) procedure proposed by Boys and Bernardi. Such a treatment is incorrect because both the final stabilized energy and also the structure and other properties (e.g., vibration frequencies) of a complex are affected by the BSSE correction. Therefore, this correction should be included in the structure and properties determination at least at the CP approximation (Hobza *et al*., 2000).

Nowadays, development of x-ray crystallography has provided more than 50 valuable crystal structures of HIV-1 RT, which are available in the Protein Data Bank (RCSB PDB, http://www.rcsb.org) (Berman *et al*., 2000). The crystallization structure of the wild-type of HIV-1 RT in complex with nevirapine inhibitor at 2.2 Å resolutions, displayed in Figure 1, has been reported previously (PDB code 1VRT) (Ren *et al*., 1995). Residues at the non-nucleoside binding site are mainly hydrophobic contacts. Structure of nevirapine inhibitor consists of two pyridine rings and cyclopropyl group that contact to the side chains of residues in the binding sites pocket. Interaction of pyridine ring and aromatic side chains of Tyr181, Tyr188,

Trp229, and Phe227 in the binding sites play an important role and these residues consist of aromatic side chains. It is found that mutations of these residues reduce the binding of inhibitor. However, despite intensive experimental investigations, the detailed origins of enzyme and inhibitor interaction remain unclear (Tantillo *et al.*, 1994). This question is of crucial importance. Better understanding is vital for the analysis of activities of mutant or designed proteins, and for the design of inhibitors as pharmaceutical lead compounds.
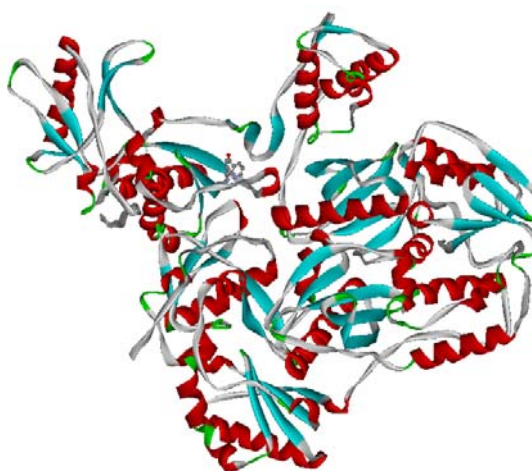
Figure 1 The structure of HIV-1 RT, coded 1VRT.pdb.

In theoretical investigation for the enzyme inhibitor interaction, the number of atoms in the molecular system makes it impossible to study with the accuracy of the method used and is limited by the computational effort required. Recently, the development of multilayer integration method in computational chemistry became more feasible to investigate the large molecular system. Now, the ONIOM method has been introduced and its efficiency has been applied over the year (Hannongbua *et al.*, 2003, 2005). In the ONIOM approach, a large molecular system can be partitioned into multilayer regions and treated different levels of theoretical methods. The inner layer or the region of the reaction occur was used a high level of *ab initio* or DFT (Density Functional Theory) methods (Morokuma *et al.*, 1996).

In this present study, quantum mechanics (QM) and our own n-layered integrated molecular orbital plus molecular mechanics method (ONIOM) are applied to the alcohol-ethylene and HIV-1 RT/nevirapine complexes with the aims of:

1. Investigate of the local structure of $OH \cdots \pi$ interaction between alcohol-ethylene systems

2. Comparing of the local structure of alcohol-ethylene systems, by using QM and ONIOM calculations.

3. Comparing of the local structure of alcohol-ethylene with and without BBSE optimization.

4. Propose of the ONIOM-BSSE scheme for binding energy correction of the ONIOM calculations.

5. Test of the applicability of the ONIOM-BSSE scheme on the HIV-1 reverse transcriptase at the binding pocket system.

LITERATURE REVIEWS

The H-bond plays a key role in chemistry, physics and biology and its consequences, such as the properties of liquid and solid water, were observed before the bond was identified and named. For a historical survey, dating back to the beginning of 20th century, the reader is referred to first chapters of recently published monographs on H-bonding: *An Introduction to Hydrogen Bonding* by Jeffrey, *The Weak Hydrogen Bond* by Desiraju and Steiner, and *Hydrogen Bonding* by Scheiner. The term "hydrogen bond" was probably used first by Linus Pauling in his paper (1931) on the nature of chemical bond.

The importance of H-bonds is enormous. They are responsible for the structure and properties of water, an essential compound for life, as a solvent and in its various phases. Further, H-bonds also play a key role in determining the shapes, properties, and functions of biomolecules.

Dewar, in 1946, suggested that π-complexes could be formed with the π-system acting as a proton acceptor. The earliest reported theoretical calculation of a "π hydrogen bond" (water-formaldehyde system) was carried out by Morokuma and coworkers (Morokuma *et al.*, 1971). However, the π-complexes formed through the involvement of polar π-bonds are not equilibrium structures, i.e., when the constraint of perpendicularity is removed, the complexes relax to a more stable dimer in which the hydrogen bonding occurs through a lone pair of electrons. On the other hand, the π-complexes formed by molecules possessing nonpolar π-bonds, such as ethene and benzene, are equilibrium structures and hence can be experimentally observed. Given this distinction, it is useful to explore the features of these π-complexes formed by nonpolar π-bonds. The nonpolar systems considered include the rare gases, the dimers of oxygen and nitrogen, the dihalogens, and various hydrocarbons. It should be noted that though phenol, benzonitrile, and a large number of purines and pyrimidines are π-systems, their interactions with various countermolecules are mediated through the

substituents constituting them rather than through the π-bonds (Schiefke *et al*., 1995; Watanebe *et al*., 1996; Matsuda *et al*., 1999).

Much experimental and theoretical effort has been directed toward obtaining a quantitative description of hydrogen bonding by studying water (Kim *et al*., 1992) and methanol clusters (Curtiss *et al*., 1979; Odutola *et al*., 1979; Castleman *et al*., 1985). Thus, detailed information of the hydrogen-bonding profiles of these systems has been obtained from high-resolution microwave and far-infrared spectra (Morokuma *et al*., 1971). Most of the theoretical investigations of these clusters have been restricted to the determination of the most stable structures, the corresponding interaction energies, and harmonic vibrational frequencies.

In the case of the interactions of π-systems with various water (Carney *et al*., 1999; Mons *et al*., 1999) or methanol clusters (Brutschy *et al*., 1991; Gruenloh *et al*., 1999), the H-bonding interactions existing between the monomers in these clusters are much stronger than the π-or σ-type of the interaction which exist between the clusters and the π-system. Thus, there are little changes in the gross structural features of the water clusters in the neutral state and in their complexes states with the π-system. This features is particularly exemplified in the infrared spectra and *ab initio* theoretical investigation of Bz-$(H_2O)_8$ (Courty *et al*., 1998), wherein it was shown that a cubic water octamer structure is π-bonded to the aromatic ring.

The earliest theoretical study of the interaction of water with nonpolar π-system was carried out by Del Bene in 1974 (Augspurger *et al*., 1993). Using the 4-31G basis set, the optimization of ethylene-water and acetylene-water at the Hartree-Fock level yielded minimum energy and geometries which are remarkably accurate for the level of calculations. However, the calculated interaction energies are expectedly low due to the non inclusion of correlation. In 1983, the first high-level calculations of the interaction of ethylene with first row hydrides were carried out. It was found that in ethylene-water, acetylene behaves as a proton donor. Around the same time, Kerlstrom carried out the first *ab initio* calculations of the interaction of

benzene with water. An intermolecular potential obtained from these calculations was employed in the Monte Carlo and molecular dynamics simulations of the interaction of water with nonpolar solvents (Linse *et al*., 1984). These theoretical studies were followed by the first matrix isolation studies of acetylene-water, ethylene-water and benzene-water by Engdahl and Nelander (Nelander *et al*., 1984). These IR investigations led to the first experimental confirmations of water being hydrogen-bonded to the π-orbital system of benzene and ethylene (Engdahl *et al*., 1985). In the case of the acetylene-water complex, it was confirmed that water behaves as a proton acceptor.

The existence of the OH$\cdots\pi$ H-bond in the benzene-water and benzene-methanol complexes initiated speculations about the existence of the CH$\cdots\pi$ H-bond. The formation of the CH$\cdots\pi$ bond was supported by experimental (Janda *et al*., 1975) and theoretical (Hobza *et al*., 1994) finding of an equilibrium structure of benzene dimer where the T-shaped arrangement of aromatic rings was believed to the stabilized by the adjoining C-H$\cdots\pi$ H-bond. The T-shaped arrangement of aromatic ring is rather common in a biological environment and stabilizes, such as the structure of phenylalanine (Hunter *et al*., 1991). This arrangement was also found in many crystal structures. Quantum chemical calculation showed that this arrangement is quite stable.

The benzene dimer was first investigating theoretically. The T-shaped structure of the dimer was optimized at the correlated MP2/6-31G(d) level. The C-H bond of the proton donor, which points to the center of the opposite benzene ring, is the shortest among all the C-H bonds. Improving the calculation did not change this result. Extending the basis set to 6-31G(d,p) level gave exactly the same result. The harmonic vibrational frequencies were evaluated at both basis sets mentioned above the proton-donor C-H stretch vibrational frequency did not predict the expected red shift typical for H-bonding but rather a large blue shift of 48 cm^{-1}. The C-H potential in the dimer is anharmonic, and therefore, there was doubt as to whether it is not the harmonic approximation that is responsible for the unexpected prediction of the blue

shift. The anharmonic C-H stretch vibration frequency was estimated at three different levels. The simplest model, considering the one-dimensional C-H anharmonicity, provided an even larger blue shift than the harmonic approximation. The two-dimensional model, which took into account the intermolecular benzene$\cdots$benzene stretch, provided a blue shift of 54 cm^{-1}. Finally, a model using effectively all of the remaining coordinates predicted a blue shift of 56 cm^{-1}, similar to the two-dimensional approach. The anharmonic calculations confirmed the surprising results of the C-H stretch frequency of the proton donor upon dimer formation (Hobza *et al*., 1998).

A new method for correcting the basis set superposition error (BSSE) in *ab initio* quantum calculations of hydrogen-bonded (H-bonded) molecular complexes were investigated by Muguet (*Muguet et. al.*, 1995). The Hartree-Fock molecular orbitals (MOs) were first localized which the localized MOs (LMOs) were then separately attributed to of the component or fragment molecules. They set to a zero value the LMO coefficients relative to the AOs belonging to all the other partner molecules or molecular fragments. After purification of the "off-fragment" coefficients, the LMOs were then reorthonormalized. The resulting wave function constituted a first level of approximation to a BSSE-corrected wave function. An iteration procedure was then implemented, comprising the following steps: HF MOs; localization; fragment attribution; off fragment purification and so on. The coverged wave function satisfied a self-consistent equation. The scheme could be extended to MCSCF wave functions. The MCSCF MOs were localized, and then off-fragment LMOs components were eliminated. The resulting LMOs were reorthonormalized to generate a MO basis for a CI computation.

The $(H_2O)_n$ and benzene-$(H_2O)_n$ (n=1, 2, 3) clusters had been characterized by means of the MP2 and density functional methods. The minimum-energy structures were optimized, and the harmonic frequencies were calculated for both the benzene-water clusters (BW_n) and the free water clusters (W_n). The W_n clusters were π-hydrogen-bonded to the benzene ring, with the strength of this interaction being greatest for BW_2. The geometries of the W_n portions of the BW_n clusters were found

to be close to those of the free water clusters, with the perturbation due to the interaction with the benzene being greatest for BW$_3$. For the OH stretching modes good arrement was found between the calculated and measured OH stretch IR spectra for both W$_n$ and BW$_n$ cluster. Both geometric and electronic effects were responsible for the significant changes in the OH stretch IR spectra brought about by the complexation of the water clusters with the benzene molecule which this were investigated by Fredericks and Jordan (Jordan *et al.*, 1996)

Medium-size basis sets were proposed to evaluate efficiently the dispersion interactions of hydrocarbon molecules. (Tsuzuki *et al.*, 1998)The aug(d,p)-6-311G** basis set was prepared by the augmentation of the diffuse d and p functions to the 6-311G** basis set. The aug(df,pd)-6-311G** basis set was prepared by the further augmentation of the diffuse f and d functions to the aug(d,p)-6-311G** basis set. The calculated MP2 and CCSD(T) intermolecular interaction energies of methane, ethane, propane, ethylene, acetylene, and benzene dimmer with these basis sets were compared with those calculated with Sadlej's basis set and Dunning's correlation-consistent basis sets. Although the aug(d,p)-6-311G** basis set was more compact than Sadlej's basis set, this basis set was more effective to evaluate the dispersion energy.The aug(df,pd)-6-311G** basis set was considerably smaller than Dunning's cc-pVQZ and cc-pV5Z basis sets. The calculated interaction energies with the aug(df,pd)-6-311G** basis set were close to those calculated with the nearly BSSE free cc-pVQZ and cc-pV5Z basis sets.

The internal rotational barrier heights of phenol and anisole were calculated using several basis sets up to cc-pVQZ with MP2-level electron correlation correction to evaluate the basis set effects. The calculations showed that the effects of the further improvement of the basis set beyond the cc-pVTZ were very small. Although the electron correlation substantially increased the barrier heights of the two molecules, the effects of the electron correlation beyond the MP2 method were not large (Tsuzuki *et al.*, 2000). The barrier heights calculated with the CCSD(T) method were close to those with the MP2 method. The internal rotational potentials of methoxy and hydroxyl groups of o-hydroxyanisole were calculated at the MP2/cc-pVTZ//HF/6-

311G** level. The calculated potentials were compared with those of phenol and anisole. o-hydroxyanisole preferred planar structure in which the hydroxyl group had an intramolecular hydrogen bond with the oxygen atom of the methoxy group. The calculated torsional potential of the methoxy group had the maximum (7.30 kcal/mol) when the methoxy group rotated 180^0 from the minimum energy structure, in which the hydroxyl group did not have the hydrogen bond. The barrier height of the methoxy group of o-hydroxyanisole was considerably larger than that of anisole (2.99 kcal/mol). The internal rotational barrier height of o-hydroxyanisole showed that the intramolecular hydrogen bond greatly stabilized the energy minimum structure and that the hydrogen bond strictly restricted the conformational flexibility of the methoxy group.

The nature and origin of the π-H interaction in both the ethane (olefinic) and benzene (aromatic) complexes of the first-row hydrides (BH_3, CH_4, NH_3, H_2O, and HF) had been investigated by carrying out high level *ab initio* calculations. The results indicated that the strength of the π-H interaction was enhanced as one progresses from CH_4 to HF. Unlike conventional H-bonds, this enhancement could not be simply explained by the increase in electrostatic interactions or the electronegativity of the atom bound to the π H-bonded proton. The contributions of each of the attractive (electrostatic, inductive, dispersive) and repulsive exchange components of the total binding energy were important (Kim *et al*., 2001). Thus, the inductive energy was highly correlated to the olefinic π-H interaction as they progress from CH_3 to HF. On the other hand, both electrostatic and inductive energies were important in the description of the aromatic π-H interaction. In either case, the contribution of dispersion energies is vital to obtain an accurate estimate of the binding energy.

Model chemistry for the evaluation of intermolecular interaction between aromatic molecules (AIMI Model) had been developed. The CCSD(T) interaction energy at the basis set limit had been estimated from the MP2 interaction energy near the basis set limit and the CCSD(T) correction term obtained by using a medium size

basis set. The calculated interaction energies of the parallel, T-shaped, and slipped-parallel benzene dimers were -1.48, -2.46, and -2.48 kcal/mol, respectively (Tsuzuki *et al*., 2001). The substantial attractive interaction in benzene dimer, even where the molecules were well separated, shows that the major source of attraction was not short-range interactions such as charge-transfer but long-range interactions such as electrostatic and dispersion. The inclusion of electron correlation increases attraction significantly. The dispersion interaction was found to be the major source of attraction in the benzene dimer. The orientation dependence of the dimer interaction was mainly controlled by long-range interactions. Although electrostatic interaction was considerably weaker than dispersion interaction, it was highly orientation dependent. Dispersion and electrostatic interactions were both important for the directionality of the benzene dimer interaction.

The interactions of the first-row hydrides (NH_3, H_2O, HF) with ethane had been investigated by carrying out calculations, at the second order Møller-Plesset (MP2) level of theory using both the 6-31+G* and aug-cc-pVDZ basis sets. Unlike previous investigations of these systems, the geometries and vibrational frequencies in the present study were obtained by carrying out explicit counterpoise correct optimizations (Tarakeshwar *et al*., 2002). In an effort to understand the nature of the H-π interactions prevalent in these complexes, the interaction energies were decomposed into individual energy components using the symmetry adapted perturbation theory. Given the goals of the present investigation, the geometries, vibrational frequencies and interaction energy components of the water dimer had also been evaluated. While the interaction energy of the conventional H-bond was dominated by electrostatic, dispersive and inductive interactions were important in the description of the π H-bond. An important distinction between conventional H-bonded complexes and the π H-bonded complexes was that the inductive interaction gets magnified at the MP2 level. Thus, the inclusion of electron correlation was an important prerequisite both for the magnification of the inductive interaction and to obtain an accurate estimate of the dispersion energies. It was observed that changes in various geometrical and vibrational parameters of these π H-bonded complexes could

be correlated to the magnitude of either the individual or a combination of various interaction energy components.

State-of-the-art electronic structure methods had been applied to the simplest prototype of aromatic π-π interactions, the benzene dimer (Sherrill *et al*., 2002). By comparison to results with a large aug-cc-pVTZ basis set, they demonstrate that more modest basis sets such as aug-cc-pvDZ were sufficient for geometry optimizations of intermolecular parameters at the second-order Møller-Plesset perturbation theory (MP2) level. However, basis sets even larger than aug-cc-pVTZ were important for accurate binding energies.

Geometry optimizations were carried out for the CN-H_2O, CN^--H_2O, NO-H_2O, HO-H_2O, and OH^-- H_2O intermolecular complexes on both the uncorrected and CP-corrected potential energy hypersufaces. Because of the correction of the basis set superposition error (BSSE) during the gradient optimization, CP-corrected gradient optimization was more prior than normal optimization in structure research. But there was no significant difference between CP-corrected gradient optimization and normal optimization at interaction energies and BSSE. The diffuse basis functions were necessary for all the present systems. 6-311+G** basis set was efficient to these systems for its good results with low time consumption which these had been investigated by Tian (Tian *et al*., 2003).

The ONIOM method was applied to the interaction of nevirapine with the HIV-1 reverse transcriptase binding site. The isolated complex of pyridine (part of nevirapine) and methyl phenol (part of Tyr181) was found at the MP2/631+G(d) level to have stacking interaction with 8.8 kcal/mol binding energy. Optimization of nevirapine and Tyr181 geometry in the pocket of 16 amino acid residues at the ONIOM3(MP2/6-31G(d):HF/3-21G:PM3) level gave the complex structure with weak hydrogen bonding but without stacking interaction. The binding energy of 8.9 kcal/mol comes almost entirely from the interaction of nevirapine with amino acid residues other than Tyr181 (Hannongbua *et al*., 2003).

Planar H-bonded and stacked structures of guanine···cytosine (G···C), adenine···thymine (A···T), 9-methylguanine···1-methylcytosine (mG···mC), and 9-methyladenine···1-methymine (mA···mT), were optimized at the RI-MP2 level using the TZVPP([5s3p2d1f/3s2p1d]) basis set. Planar H-bonded structures of G···C, mG···mC, and A···T correspond to the Watson-Crick (WC) arrangement, in contrast to mA···mT for which the Hoogsteen (H) structure was found. Stabilization energies for all structures were determined as the sum of the complete basis set limit of MP2 energies and a ($\Delta E^{CCSD(T)}$-ΔE^{MP2}) correction term evaluated with the cc-pVDZ (0.25.0.15) basis set. The complete basis set limit of MP2 energies was determined by two-point extrapolation using the aug-ccpVXZ basis sets for X=D and T and X=T and Q. This procedure was required since the convergency of the MP2 interaction energy for the present complexes was rather slow, and it was thus important to include the extrapolation to the complete basis set limit. For the MP2/aug-cc-pVQZ level of theory, stabilization energies for all complexes studied were already very close to the complete basis set limit. The much cheaper D→T extrapolation provided a complete basis set limit close (by less than 0.7 kcal/mol) to the more accurate T→Q term, and the D→T extrapolation could be recommended for evaluation of complete basis set limits of more extended complexes (e.g. larger motifs of DNA). The convergency of the ($\Delta E^{CCSD(T)}$-ΔE^{MP2}) term was known to be faster than that of the MP2 or CCSD(T) correlation energy itself, and the cc-pVDZ (0.25,0.15) basis set provided reasonable values for planar H-bonded as well as stacked structures. Inclusion of the CCSD(T) correction was essential for obtaining reliable relative values for planar H-bonding and stacking interactions; neglecting the CCSD(T) correction results in very considerable errors between 2.5 and 3.4 kcal/mol. Final stabilization energies (kcal/mol) for the base pairs studied were very substantial (A···T WC, 15.4; mA···mT H, 16.3; A···T stacked, 11.6; mA···mT stacked, 13.1 G···C WC, 28.8; mG···mc WC, 28.5; G···C stacked, 16.9; mG···mC stacked, 18.0), much larger than published previously(Hobza *et al*., 2003). On this basis of comparison with experimental date, they concluded that those values represented the lower boundary of the true stabilization energies. On the basis of error analysis, they expected the present H-bonding energies to be fairly close to the true values, while stacked energies were still expected to be about 10% too low. The stacking energy for the mG···mC pair was

considerably lower than the respective H-bonding energy, but it was larger than the mA··mT H-bonding energy. This conclusion could significantly change the present view on the importance of specific H-bonding interactions and nonspecific stacking interactions in nature, for instance, in DNA. Present stabilization energies for H-bonding and stacking energies represented the most accurate and reliable values and could be considered as new reference data.

Sandwich and T-shaped configurations of benzene dimer, benzene-phenol, benzene-toluene, benzene-fluorobenzene, and benzene-benzonitrile were studied by coupled-cluster theory to elucidate how substituents tune π-π interactions (Sherrill *et al*., 2004). All substituted sandwich dimers bind more strongly than benzene dimer, whereas the T-shaped configurations bind less favorably depending on the substituent. Symmetry-adapted perturbation theory (SAPT) indicated that electrostatic, dispersion, induction, and exchange-repulsion contributions were all significant to the overall binding energies, and all but inductions were important in determining relative energies. Models of π-π interactions based solely on electrostatics, such as the Hunter-Sanders rules, did not seem capable of explaining the energetic ordering of the dimers considered.

Particular interaction between efavirenz and the HIV-1 reverse transcriptase binding site was investigated, based on the B3LYP/6-31G(d,p) and ONIOM2 methods. The interaction between efavirenz and Lys101 was found to be the strongest interaction, typically -11.29 kcal/mol. The stability of this complex system leads to the foundation of the estimated binding energy of approximately -22.66 kcal/mol. Moreover, two hydrogen bonds between benzoxazin-2-one, and the backbone carbonyl oxygen and the backbone amino hydrogen of Lys101 were observed (Hannongbua *et al*., 2005). These hydrogen bond interactions play an important role in the bound efavirenz/HIV-1 RT complex.

The benzene dimer was the simplest prototype of π-π interactions and had been used to understand the fundamental physics of these interactions as they were observed in more complex systems (Sherrill *et al*., 2005). In biological systems,

however, aromatic rings were rarely found in isolated pairs; thus, it was important to understand whether aromatic pairs remain a good model of π-π interactions in clusters. In his study, *ab initio* methods were used to compute the binding energies of several benzene trimers and tetramers, most of them in 1D stacked configurations. The two-body terms changed only slightly relative to the dimer, and except for the cyclic trimer, the three- and four-body terms were negligible. This indicated that aromatic clusters did not feature any large nonadditive effects in their binding energies, and polarization effects in benzene clusters did not greatly change the binding that would be anticipated from unpertubed benzene-benzene interactions at least for the 1D stacked systems considered. Three-body effects were larger for the cyclic trimer, but for all systems considered, the computed binding energies are with in 10% of what would be estimated from benzene dimer energies at the same geometries.

The interaction between aromatic rings and sulfur atoms in the side chains of amino acids is a factor in the formation and stabilization of α-helices in proteins (Sherrill *et al*., 2005). They studied the H_2S-benzene dimer as the simplest possible prototype of sulfur-π interactions. High quality potential energy curves were obtained using coupled-cluster theory with single, double, and pertubative triple substitutions (CCSD(T)) and a large, augmented quadruple-ζ basis set (aug-cc-pVQZ). The equilibrium intermonomer distance for the hydrogens-down C_{2v} configuration is 3.8 Å with interaction energy of -2.74 kcal/mol. Extrapolating the binding energy to the complete basis set limit gives -2.81 kcal/mol. These binding energy was comparable to that of H_2O-benzene or of the dimer were also considered at less complete levels of theory. A considerable reduction in binding for the sulfur-down configuration, together with an energy decomposition analysis, indicated that the attraction in H_2S-benzene was best thought of as arising from a favorable electrostatic interaction between partially positive hydrogens in H_2S with the negatively charged π-cloud of the benzene.

The nature of interactions of phenol with various molecules (Y=HF, HCl, H_2O, H_2S, NH_3, PH_3, MeOH, MeSH) was investigated using *ab initio* calculations (Kim *et*

al., 2005). The optimized geometrical parameters and spectra for the global energy minima of the complexes match the available experimental data. The contribution of attractive (electrostatic, inductive, dispersive) and repulsive (exchange) components to the binding energy was analyzed. HF favors σ_o-type H-bonding, while H_2O, NH_3, and MeOH favor σ_H-type H-bonding, where σ_o-/σ_H-type were the case when an H-bond forms between the phenolic O/H atom and its interacting molecule. On the other hand, HCl, H_2S, and PH_3 favor π-type H-bonding, which were slightly favored over σ_o-, σ_H-, σ_H-type bonding, respectively. MeSH favors χ_H-type bonding, which had characteristics of the π and σ_H. The origin of this conformational preference depending on the type of molecules was elucidated. Finally, phenol-Y complexes were compared with water-Y complexes. In the water-Y complexes where σ_o-/σ_H-type involves the H-bond by the water O/H atom, HF and HCl favor σ_o-type, H_2O involves both σ_o-/σ_H-type, and H_2S, NH_3, PH_3, MeOH, and MeSH favor σ_H-type bonding. Except for HF, seven other species had larger binding energies with phenol molecules than a water molecules.

This study was the first step in the systematic investigation of substituted (carboxyl) polystyrene nanoparticles. Understanding the fundamental interactions between the p-carboxyl styrene monomers, where an ethyl group was used instead of a vinyl group (referenced, for convenience, as "p-carboxyl styrene"), provides the basic information needed to construct potentials for nanoparticles composed of these monomers(Gordon *et al*., 2006). In this work, low energy isomers of p-carboxyl styrene dimer were studied. The dimer structures and their relative and binding energies were determined using both Møller-Plesset second-order perturbation theory (MP2) and the general effective fragment potential (EFP2) method. Sections of the intermolecular potential energy surface (PES) of the p-carboxylated styrene dimer in its global minimum orientation were also determined. As expected, double hydrogen bonding between the two carboxylic groups provides the strongest interaction in this system, followed by isomers with a single H-bond and strong benzene ring-benzene ring (π-π) type interactions. Generally, the EFP2 method reproduces the MP2 geometries and relative energies with good accuracy, so it appears to be an efficient

alternative to the correlated *ab initio* methods, which are too computationally demanding to be routinely used in the study of the more-complex polymeric systems of interest.

METHODS OF CALCULATION

Models and Methods

1. **Ethanol and ethylene: A model for OH$\cdots\pi$ system**

In order to investigate the OH$\cdots\pi$ interaction, the MP2(full)/6-31G(d) level of calculations were applied to obtain the intermolecular interaction between ethanol-ethylene system. The geometries for isolated ethanol and ethylene molecules were optimized at the same level of calculation. During optimization for isolated molecules, the ethanol and ethylene monomer were constrained at C_s and D_{2h} symmetries, respectively. For the complex structure, there are three possible geometries for the ethanol-ethylene systems optimized with the same selected level of calculation based on Cs symmetry constrain and these possible geometries were shown in Figure 2. The geometrical parameters and the binding energies with and without basis set super position error (BSSE) of the ethanol-ethylene system will be dicussed in the section of results and disscussion.

2. **OH$\cdots\pi$ system for alcohol-ethylene complex**

2.1 The standard gradient optimization

The intermolecular interaction of ethanol-ethylene, propanol-ethylene, n-butanol-ethylene, and iso-butanol-ethylene systems were calculated by the standard gradient optimization with C_s symmetry constraint at the MP2(full)/6-311+G(d,p), MP2(full)/6-31G(d) and B3LYP/6-31G(d) levels of calculation. In this regard, the complex structures of all models were shown in Figure 3. The geometrical parameters and the binding energies with and without basis set superposition error (BSSE) of the alcohol-ethylene systems will be discussed in the section of results and discussion.

(a)

(b)

(c)
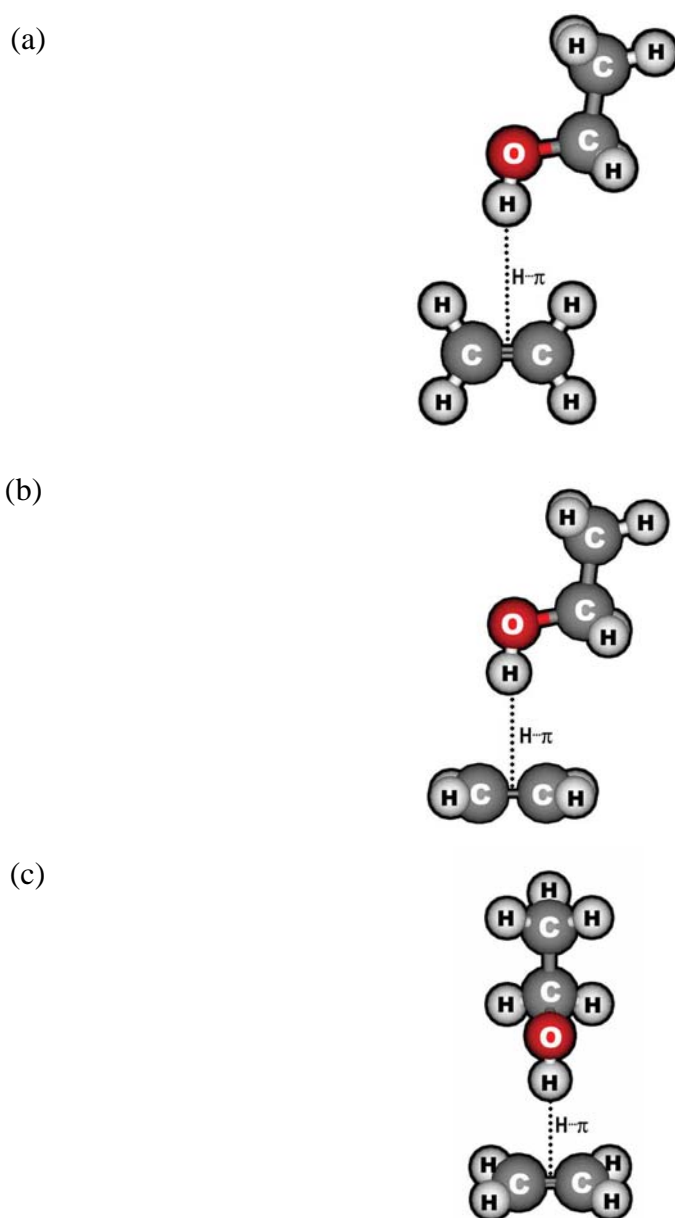
Figure 2 Starting geometries of the three possible geometries for testing models of ethanol-ethylene system at the MP2(full)/6-31G(d) level of calculation (a) model 1 (b) model 2 (c) model 3
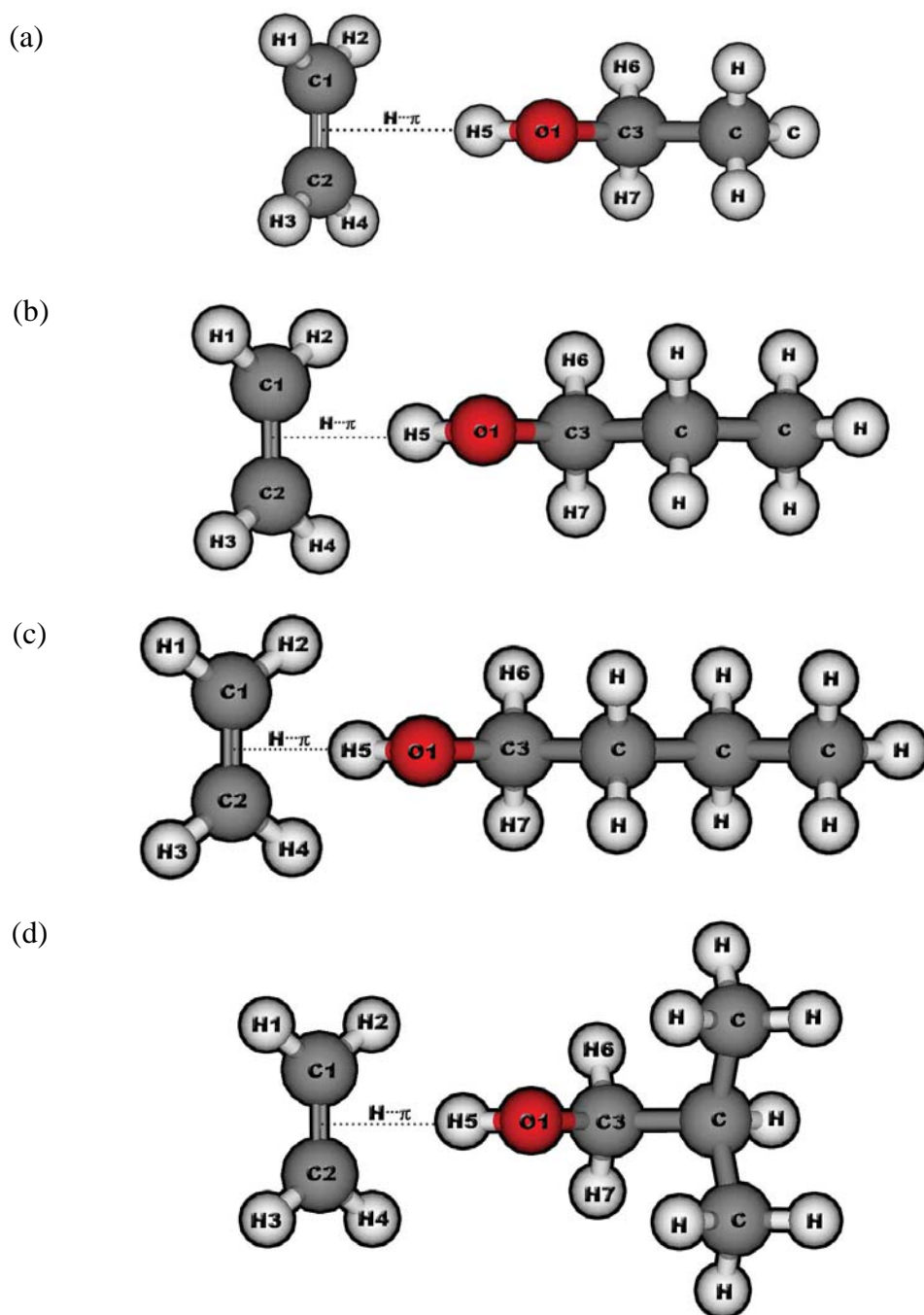
Figure 3 Starting geometries for alcohol-ethylene systems at the MP2(full)/6-311+G(d,p), MP2(full)/6-31G(d) and B3LYP/6-31G(d) levels of calculation (a) ethanol-ethylene system (b) propanol-ethylene system (c) n-butanol-ethylene system (d) iso-butanol-ethylene system

2.2 The CP-corrected gradient optimization

Fully CP-corrected gradient optimization of ethanol-ethylene, propanol-ethylene, n-butanol-ethylene and iso-butanol-ethylene systems were performed at the MP2(full)/6-311+G(d,p), MP2(full)/6-31G(d) and B3LYP/6-31G(d) levels of calculation using C_s symmetry constraint. The geometrical parameters and the binding energies with BSSE of the alcohol-ethylene systems will be discussed in the section of results and discussion.

2.3 The ONIOM optimization

The ethanol-ethylene, propanol-ethylene, n-butanol-ethylene and iso-butanol-ethylene systems were divided for the application of the OINIOM calculations and the applicability of the ONIOM-BSSE scheme. The ONIOM2 method stands for the two-layer ONIOM calculation. The partitions of the complex structures were shown in Figure 4. Each ONIOM method can be simply expressed its combined methods for the different layers by combination of method and basis sets. In this thesis, two ONIOM models were selected. Therefore, these two ONIOM models can be presented as:

ONIOM2M: MP2(full)/6-311+G(d,p):MP2(full)/6-31G(d)
ONIOM2B: MP2(full)/6-311+G(d,p):B3LYP/6-31G(d)

The total ONIOM energy of the entire system (AB) was obtained from three independent energy calculations in ONIOM2.

$$E^{ONIOM2} [AB] = E [High, A] + E [Low, AB] - E [Low, A]$$

(a)

(b)

(c)

(d)
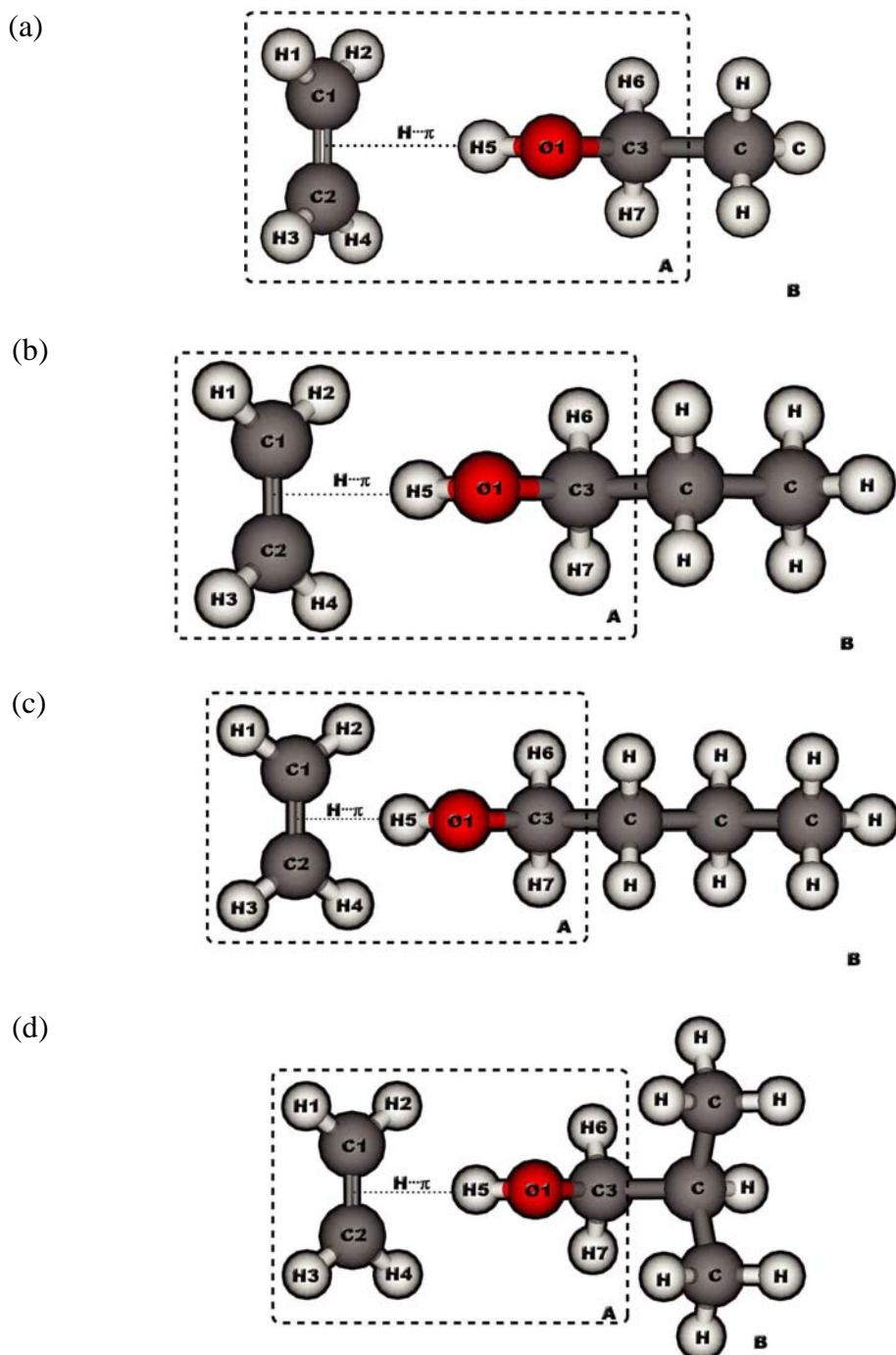
Figure 4 Starting geometries for alcohol-ethylene systems of ONIOM2M and ONIOM2B methods for region A using high level of theory and region B using low level of theory (a) ethanol-ethylene system (b) propanol-ethylene system (c) n-butanol-ethylene system (d) iso-butanol-ethylene system

3. Structure and binding energy of ethanol-ethylene complex

The H⋯π interaction following the B3LYP and MP2 calculations with a 6-31G(d,p) basis set were applied to obtain the intermolecular interaction between ethanol and ethylene in the complex. The MP2 method was used with both full and frozen core (FC) electron correlation. During optimization, the ethanol and ethylene monomers were kept at C_s and D_{2h} symmetries, respectively, while the complex was kept at C_s symmetry. The standard and CP-corrected gradient optimizations of the intermolecular interaction were performed. To reduce the computational demand in the next application, the two-layered ONIOM (ONIOM2) method was applied to examine the interaction between ethanol and ethylene in the complex with the same C_s symmetry configuration. The system was divided into two parts as shown in Figure 4(a). The inner layer or H⋯π interaction (region A) was treated at the MP2 level of calculation, while the outer layer (region B) was treated at the B3LYP level of calculation. In order to evaluate the intermolecular interaction, we also computed the basis set superposition error (BSSE) based on the counterpoise scheme.

4. Structure and binding energy of HIV-1 RT binding site/nevirapine complex

The starting model of the HIV-1 RT binding site/nevirapine complex in this study was obtained from the 2.2 Å resolved crystal structure of nevirapine bound to HIV-1 RT (1VRT.pdb for the wild type and 1JLB.pdb for Y181C mutant type). Based on this structure, we adopted the system consisting of 22 residues surrounding the non-nucleoside inhibitor binding pocket (NNIBP) with at least one atom interacting with any of the atoms of the nevirapine structure within a 7 Å diameter centered at nevirapine (Figure 5). All residues, assumed to be in their neutral form, were terminated, if not connected to another residue in the selected model by a link with an acetyl group (CH_3CO-) and a methyl group (-$NHCH_3$) at the N- and C-terminal ends of the cut residues, respectively (Figure 6), from the adjacent residues as presented in the backbone geometries of the X-ray structure. Hydrogen atoms were added to the geometrical structure to generate the complete structure of the model system and their

positions were optimized by the semiso-empirical PM3 method. This structure was used

(a)
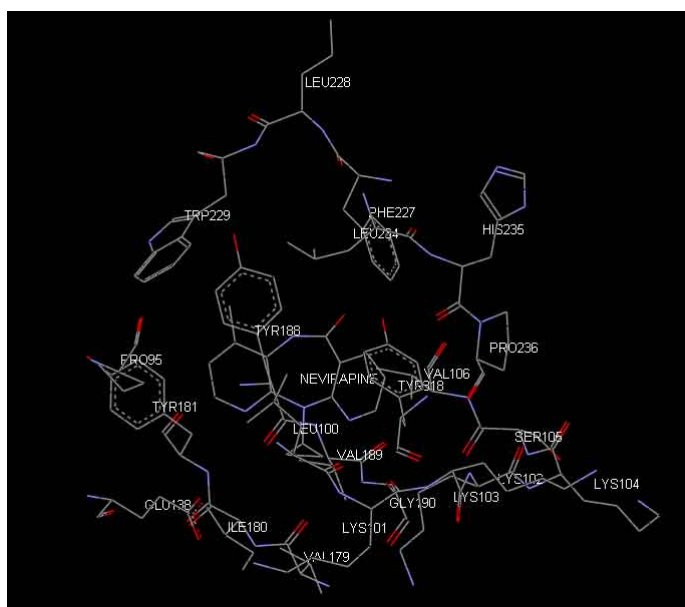
(b)
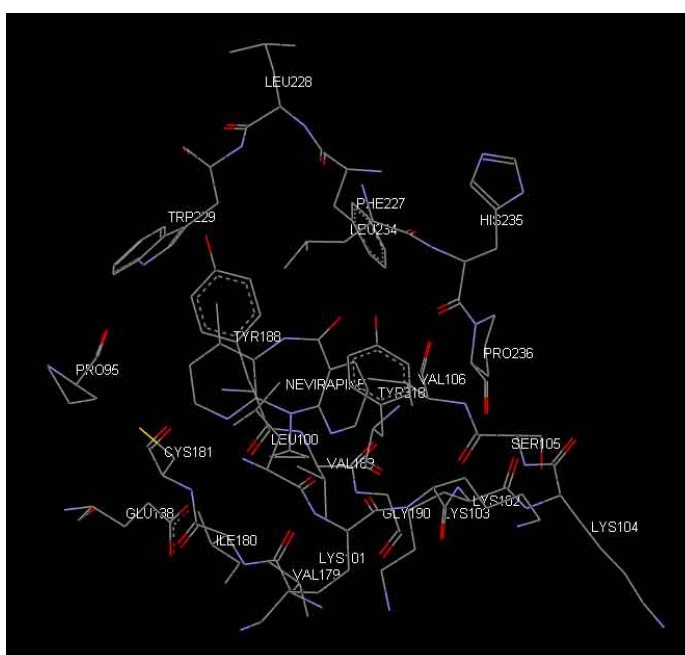
Figure 5 The starting model of the HIV-1 RT binding site/nevirapine complex adopted the system consisting of 22 residues surrounding the non-nucleoside inhibitor binding pocket (NNIBP) with at least one atom interacting with any of the atoms of the nevirapine structure within a 7 Å diameter centered at nevirapine (a) wild type (b) mutant type

as the starting geometry for all calculations. Therefore, the residues in this system are CH$_3$CO-Pro95-NHCH$_3$, CH$_3$CO-Lue100-Lys101-Lys102-Lys103-Lys104-Ser105-Val106-NHCH$_3$, CH$_3$CO-Val179-Ile180-Tyr181-NHCH$_3$, CH$_3$CO-Tyr188-Val189-Gly190-NHCH$_3$, CH$_3$CO-Phe227-Leu228-Trp229-NHCH$_3$, CH$_3$CO-Leu234-His235-Pro236-NHCH$_3$, CH$_3$CO-Tyr318-NHCH$_3$ of the p66 domain of RT and CH$_3$CO-Glu138-NHCH$_3$ of the p51 domain. The ONIOM3 model was used.
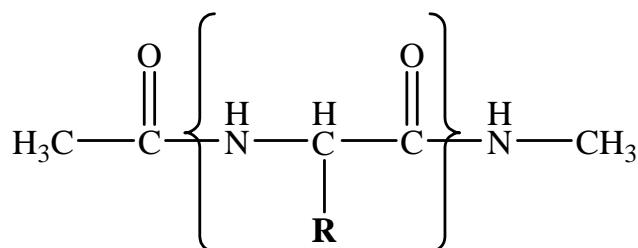
Figure 6 Capped groups of the terminal ends of chains.

In order to compute the BSSE-corrected binding energy in the ONIOM model, the scheme for the calculation of the counterpoise correction binding energy is expressed as following

$$BE_{ONIOM}^{CP} = BE_{HL}^{CP}(m1m2) + BE_{LL}^{CP}(M1M2) - BE_{LL}^{CP}(m1m2)$$

where CP, HL, LL, m and M denote counterpoise correction, high level of calculation, low level of calculation, small model in inner layer and real model in outer layer, respectively. The numbers 1 and 2 stand for the part of nevirapine and the part of an amino acid studied as defined in the ONIOM model system, respectively.

All calculations were carried out using the GAUSSIAN 03 package running on Linux PC 3.2 GHz.

RESULTS AND DISSCUSSION

1. <u>Ethanol and ethylene: A model for OH···π system</u>

In this section, the intermolecular interaction between ethanol-ethylene forming the OH-π system were calculated by the MP2 method using 6-31G(d) basis sets at C_s symmetry constrain. The geometries for isolated ethanol and ethylene molecules were optimized at the same level of calculation. During optimization for isolated molecules, the ethanol and ethylene monomer were constrained at C_s and D_{2h} symmetry, respectively. For the complex structure, there are three possible geometries for the ethanol-ethylene system optimized with the same selected level of calculation based on C_s symmetry constraint. The geometrical parameters and the binding energies for the isolated molecules and the ethanol-ethylene system are summarized in Table 1. The optimized geometries for the monomer and the complex structure are shown in Figure 7.

Firstly, the binding energies for three model systems optimized at MP2/6-31G(d) level of calculation will be discussed. The obtained results indicated that the uncorrected binding energy for the model 1, model 2 and model 3 are -0.64, -3.66 and -3.77 kcal/mol, respectively. The model 3 provides the lowest binding energy compared to those model in this study. The binding energies with BSSE correction are -2.01 and -1.89 kcal/mol for model 2 and model 3. It can be seen that only two model systems give the attractive binding energy in which the model 2 yields the binding energy lower than the model 3. Therefore, the only binding energy investigation can not adequate for determination in the selecting model system. Consequently, the geometrical parameter of the ethanol-ethylene system was analyzed.

Considering the geometry, the intermolecular distances between H-atom of methanol and the π-system of ethylene in the model 2 and model 3 are found to be 2.431 and 2.382 Å, respectively. Accordingly, the model 3 was applied to study in the further alcohol-ethylene system because this model yields to the appropriate on both reliable intermolecular distance and binding energy.
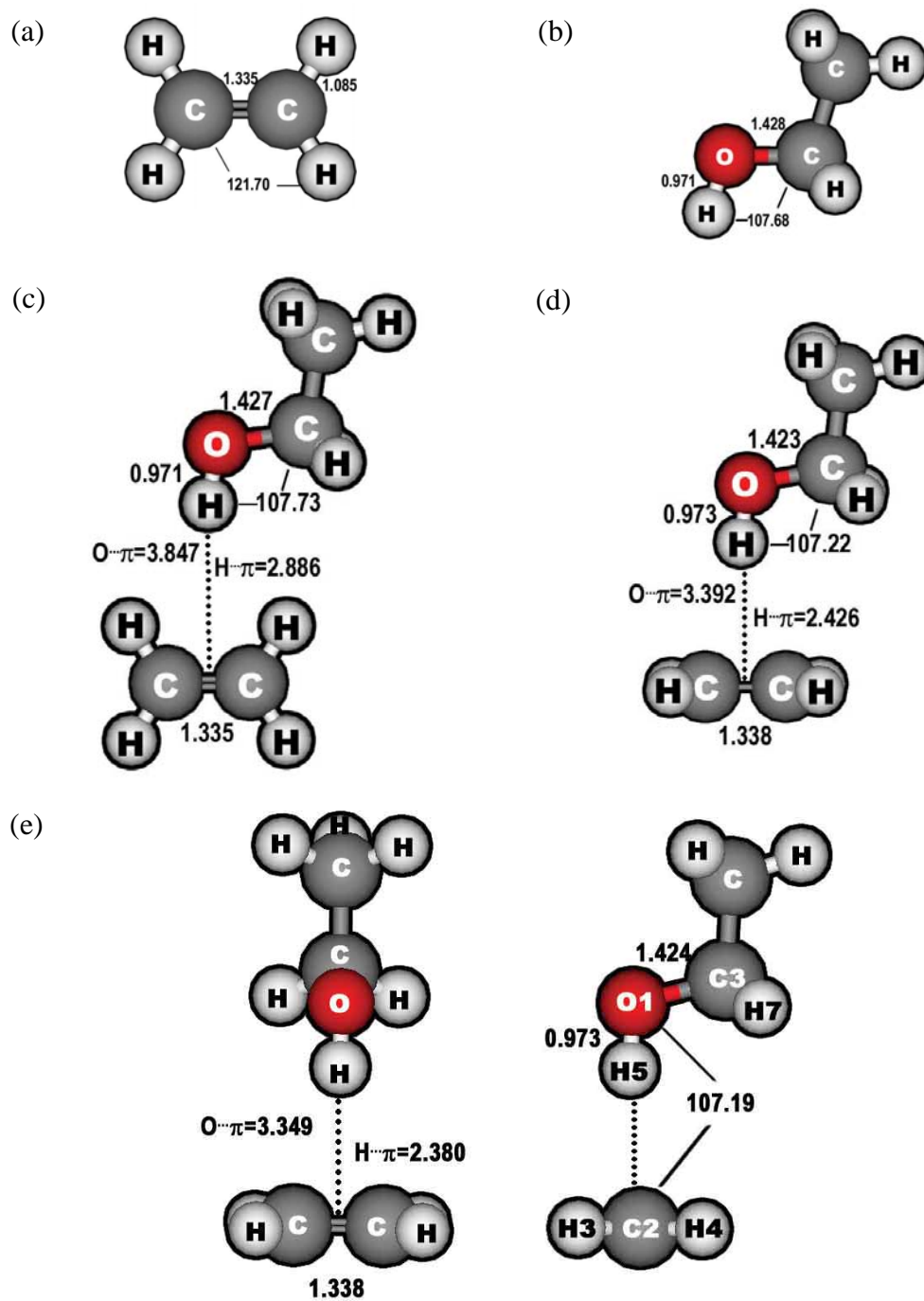
Figure 7 Optimized structure of the monomer and complex of the ethanol-ethylene system at the MP2(Full)/6-31G(d) level of theory (a) ethylene (b) ethanol (c) model 1 (d) model 2 (e) model 3

Table 1 Selected bond distances (Å) and binding energy (BE) of the optimized structure of the monomer and complex of the ethanol-ethylene system at the MP2(Full)/6-31G(d) level of calculation.

Ethanol-Ethylene complex	Ethanol	Ethylene	Model 1	Model 2	Model 3
Distance (Å)					
H$\cdots\pi$	-	-	2.886	2.426	2.380
O$\cdots\pi$	-	-	3.847	3.392	3.349
O-H	0.971	-	0.971	0.973	0.973
C-O	1.428	-	1.427	1.423	1.424
C=C	-	1.335	1.335	1.338	1.338
Angle (Degree)					
<C-O-H	107.68	-	107.73	107.22	107.19
Energy(kcal/mol)					
BE(uncorr.)	-	-	-0.64	-3.66	-3.77
BE(corr.)	-	-	0.09	-2.01	-1.89

Comparing the geometry and the binding energy of iso-butanol-ethylene system based on the model 2 and model 3 configurations, the obtained results shown that the uncorrected binding energy for model 2 and model 3 are -3.84 and -3.95 kcal/mol, respectively. Thus, the results indicated that the model 3 provides the lower energy than the model 2. After corrected the binding energy, the model 2 and model 3 produced the binding energies are -2.14 and -2.01 kcal/mol, respectively. It can be seen that two models system provide the similar energetic both uncorrected and corrected binding energies. Therefore, the geometry was investigated. The H$\cdots\pi$ in the model 2 and model 3 are found to be 2.409 and 2.362 Å, respectively. Accordingly, the model 2 and model 3 of iso-butanol-ethylene system present the same fashion of ethanol-ethylene system.

2. <u>OH⋯π system for alcohol-ethylene complex</u>

2.1 The standard gradient optimization

Geometrical optimization of ethanol-ethylene, propanol-ethylene, n-butanol-ethylene, and iso-butanol-ethylene systems were performed at B3LYP/6-31G(d), MP2(full)/6-31G(d) and MP2(full)/6-311+G(d,p) levels of calculation using C_s symmetry constraint. The calculated binding energies and selected geometrical parameters for alcohol-ethylene systems are summarized in Table 2. The standard gradient optimizations of the intermolecular interaction were observed. Considering the H5⋯π and O1⋯π distances of alcohol-ethylene systems, it was found that the MP2(full)/6-311+G(d,p), MP2(full)/6-31G(d) and B3LYP/6-31G(d) level of calculations give the distances about 2.362-2.433 Å for the H5⋯π distance and 3.333-3.397 Å for the O1⋯π distance. The MP2(full)/6-31G(d) level of calculation produce the H5⋯π and O1⋯π shorter distances than the obtained results from the MP2(full)/6-311+G(d,p) and B3LYP/6-31G(d) level of calculations. The schematic presentations of the optimized geometries at MP2(full)/6-311+G(d,p) level of calculations are shown in Figure 8. The effect of the basis set size on the geometry optimized at the MP2 method, it indicates that the intermolecular distances from 6-311+G(d,p) basis set yields the longer distances than the intermolecular distances from 6-31(d) basis set.

The uncorrected binding energies of ethanol-ethylene, propanol-ethylene, n-butanol-ethylene, and iso-butanol-ethylene systems at the MP2(full)/6-311+G(d,p) level of calculation are -3.35, -3.40, 3.44 and -3.49 kcal/mol, respectively. It was found that the binding energy of these alcohol-ethylene system increases from ethanol-ethylene system to iso-butanol-ethylene system. However, the BSSE corrected binding energy has an effect on the alcohol-ethylene systems more than 30%. Therefore, the BSSE-corrected binding energies should be included the OH⋯π system.
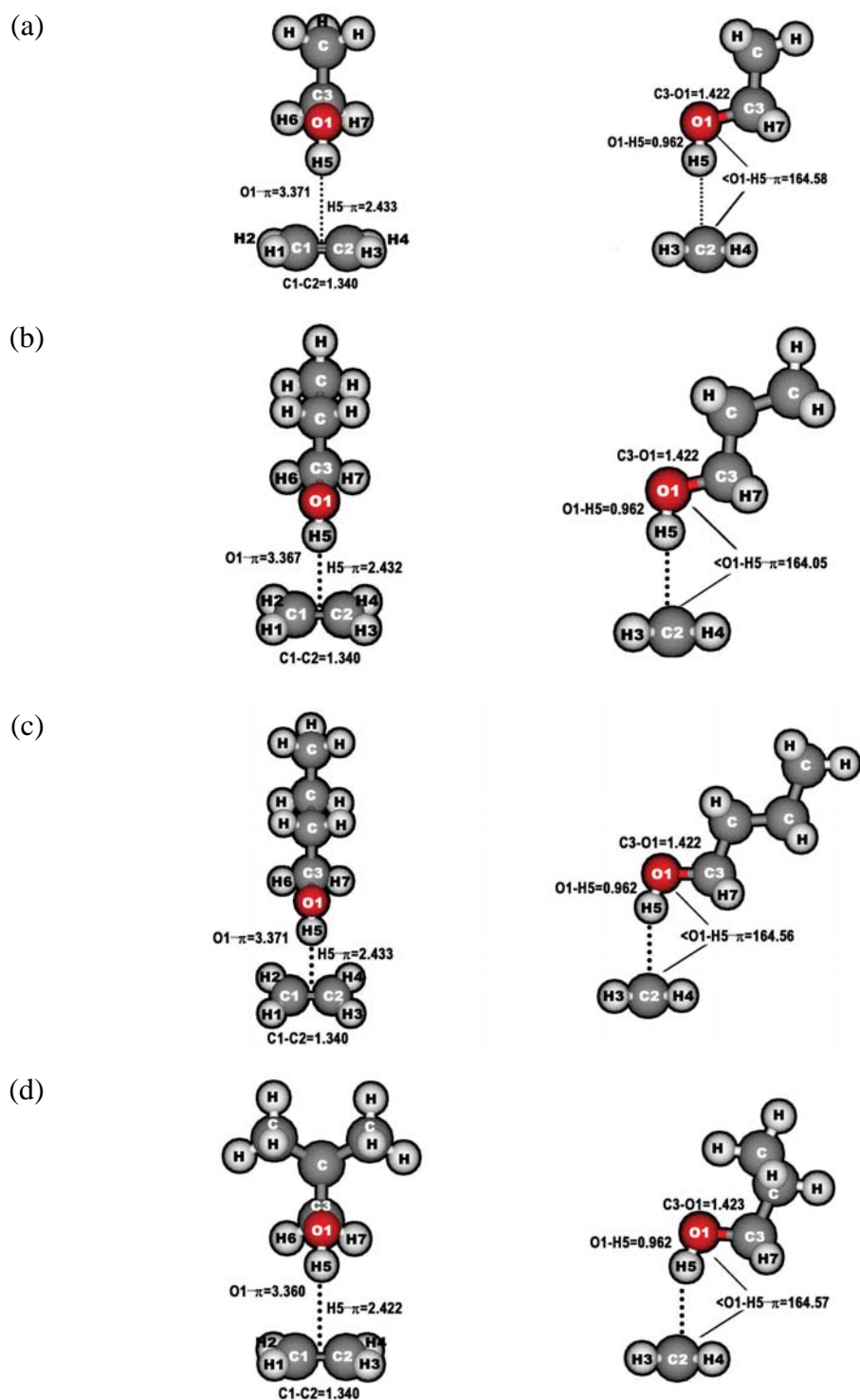
Figure 8 The standard gradient optimized structure of alcohol-ethylene systems at the MP2(Full)/6-311+G(d,p) level of calculation (a) ethanol-ethylene complex (b) propanol-ethylene complex (c) n-butanol-ethylene complex (d) iso-butanol-ethylene complex

Table 2 Selected bond distances (Å) and binding energy (BE) of the standard gradient optimized structure of alcohol-ethylene systems at the MP2(full)/6-311+G(d,p); (M11), MP2(full)/6-31G(d); (M1) and B3LYP/6-31G(d); (B1) levels of calculation

Parameter	Ethanol-ethylene			Propanol-ethylene			n-Butanol-ethylene			iso-Butanol-ethylene		
	M11	M1	B1	M11	M1	B1	M11	M1	B1	M11	M1	B1
Distance (Å)												
H5···π	2.433	2.380	2.426	2.432	2.379	2.421	2.433	2.380	2.422	2.422	2.362	2.415
O1···π	3.371	3.349	3.397	3.367	3.346	3.391	3.371	3.346	3.393	3.360	3.333	3.386
O1-H5	0.962	0.973	0.972	0.962	0.973	0.972	0.962	0.973	0.972	0.962	0.973	0.971
C3-O1	1.422	1.424	1.421	1.422	1.423	1.420	1.422	1.423	1.420	1.423	1.424	1.421
C1=C2	1.340	1.338	1.334	1.340	1.338	1.334	1.340	1.338	1.334	1.338	1.338	1.334
C1-H1	1.085	1.085	1.087	1.085	1.085	1.087	1.085	1.085	1.087	1.085	1.085	1.087
C2-H2	1.085	1.085	1.087	1.085	1.085	1.088	1.086	1.085	1.088	1.086	1.085	1.088
Angle(Degree)												
<O1-H5···π	164.58	173.48	176.23	164.05	172.30	176.27	164.56	171.74	177.29	164.57	176.27	178.57
<C3-O1···π	95.79	102.57	110.68	95.33	101.70	110.67	95.70	101.24	109.87	95.74	104.69	109.03
Energy(kcal/mol)												
BE(uncorr.)	-3.35	-3.77	-3.04	-3.40	-3.80	-3.05	-3.44	-3.80	-3.03	-3.49	-3.95	-3.14
BE(corr.)	-2.03	-1.89	-1.98	-2.07	-1.91	-1.98	-2.08	-1.90	-1.96	-2.15	-2.04	-2.03

2.2 The CP-corrected gradient optimization

Fully CP-corrected gradient optimized of ethanol-ethylene, propanol-ethylene, n-butanol-ethylene and iso-butanol-ethylene systems were performed at the B3LYP/6-31G(d), MP2(full)/6-31G(d) and MP2(full)/6-311+G(d,p) levels of calculation using C_s symmetry constraint. The calculated binding energies and selected geometrical parameters for alcohol-ethylene systems at the same levels of calculation are summarized in Table 3.

Comparing the MP2 and B3LYP methods with the 6-31G(d) basis set, the obtained results indicate that the intermolecular distances for H5$\cdots\pi$ and O1$\cdots\pi$ of ethanol-ethylene, propanol-ethylene, n-butanol-ethylene, and iso-butanol-ethylene systems by the MP2 method give the longer distance that of the B3LYP method. The alcohol-ethylene systems optimized at the MP2/6-311+G(d,p) level of calculation for the H5$\cdots\pi$ distances are 2.548, 2.546, 2.546 and 2.524 Å for ethanol-ethylene, propanol-ethylene, n-butanol-ethylene, and iso-butanol-ethylene systems, respectively. Considering the geometrical parameters between the standard gradient optimization and the CP-corrected gradient optimization in Table 2 and Table 3, it was found that the CP-corrected gradient optimization has an effect on the intermolecular interaction by the H5$\cdots\pi$ distances longer than the standard gradient optimization.

In the case of the binding energy, the obtained results show that the alcohol-ethylene systems calculated at the MP2(full)/6-311+G(d,p) level of theory based on the CP-corrected gradient optimized geometry at the same level produce the binding energy in the range of -3.26 to -3.35 kcal/mol without the BSSE-correction and in the range of -2.12 to -2.23 kcal/mol with the BSSE-correction. When comparing the uncorrected binding energies based on the standard gradient optimization and the CP-corrected gradient optimization (see Table 2 and Table 3), it was shown that the former provide the binding energy lower that of the later with the energy different less than 0.24 kcal/mol. However, the corrected binding energies results show that the CP-corrected gradient optimization geometry has an effect on the corrected binding energy.

Table 3 Selected bond distances (Å) and binding energy (BE) of the CP-corrected gradient optimized structure of alcohol-ethylene systems at the MP2(full)/6-311+G(d,p); (M11), MP2(full)/6-31G(d); (M1) and B3LYP/6-31G(d); (B1) levels of calculation

Parameter	Ethanol-ethylene			Propanol-ethylene			n-Butanol-ethylene			iso-Butanol-ethylene		
	M11	M1	B1	M11	M1	B1	M11	M1	B1	M11	M1	B1
Distance (Å)												
H5···π	2.548	2.579	2.527	2.546	2.575	2.525	2.546	2.576	2.525	2.524	2.561	2.519
O1···π	3.506	3.552	3.492	3.500	3.548	3.493	3.502	3.549	3.488	3.486	3.535	3.490
O1-H5	0.962	0.973	0.972	0.962	0.973	0.972	0.962	0.973	0.972	0.962	0.973	0.971
C3-O1	1.423	1.424	1.421	1.422	1.424	1.420	1.423	1.424	1.421	1.424	1.425	1.422
C1=C2	1.340	1.337	1.333	1.340	1.337	1.333	1.340	1.337	1.333	1.340	1.337	1.333
C1-H1	1.085	1.085	1.087	1.085	1.085	1.087	1.085	1.085	1.087	1.085	1.085	1.087
C2-H2	1.085	1.085	1.088	1.085	1.085	1.088	1.085	1.085	1.088	1.085	1.085	1.088
Angle (Degree)												
<O1-H5···π	173.34	179.60	172.34	171.25	179.67	174.19	172.08	179.45	171.19	178.63	179.47	176.93
<C3-O1···π	102.38	107.81	113.56	100.77	107.27	112.29	101.36	107.06	114.42	106.09	107.89	110.23
Energy(kcal/mol)												
BE(uncorr.)	-3.26	-3.54	-3.01	-3.31	-3.57	-3.02	-3.35	-3.57	-3.00	-3.34	-3.71	-3.10
BE(corr.)	-2.12	-2.14	-2.04	-2.16	-2.16	-2.04	-2.16	-2.16	-2.03	-2.23	-2.27	-2.10

2.3 The ONIOM optimization

In order to understand the OH⋯π system, the interaction between aliphatic alcohol (such as ethanol, propanol, n-butanol or iso-butanol) and ethylene was used as a model to examine the nature of these weak interactions. Therefore, the models of this interaction are ethanol-ethylene, propanol-ethylene, n-butanol-ethylene and iso-butanol-ethylene systems. The alcohol-ethylene system was divided for the application of the ONIOM calculations and proved the applicability of the ONIOM-BSSE scheme. Consequently, the ONIOM method was applied in the study of the structure and the binding energy of nevirapine at the binding site of HIV-1 reverse transcriptase both in the wild type and the Y181C mutant type structures.

The ONIOM2 method stands for the two-layer ONIOM calculation. Each ONIOM method can be simply expressed its combined methods for the different layers by combination of method and basis sets. In this thesis, two ONIOM methods were selected. The ONIOM2M and the ONIOM2B methods stand for the MP2(full)/6-311+G(d,p):MP2(full)/6-31G(d) and MP2(full)/6-311+G(d,p):B3LYP/6-31G(d) combined methods, respectively.

The calculated binding energies and selected geometrical parameters for the alcohol-ethylene systems by the ONIOM2M and ONIOM2B methods are summarized in Table 4. The geometrical optimization and selected distance between H atom of alcohol and the center of ethylene molecules forming OH⋯π systems optimized at the ONIOM2M level of calculation are depicted in Figure 9. The obtained results show that the ONIOM2M and ONIOM2B methods has a smaller effect to the H5⋯π and O1⋯π distances and also to the binding energies.

The H5⋯π distance obtained from the ONIOM2M method for ethanol-ethylene, propanol-ethylene, n-butanol-ethylene and iso-butanol-ethylene systems are 2.428, 2.424, 2.240 and 2.377 Å, respectively, according to the ONIOM-BSSE binding energies for these systems are -1.98, -2.00, -2.04 and -2.15 kcal/mol, respectively.

(a)

(b)

(c)

(d)
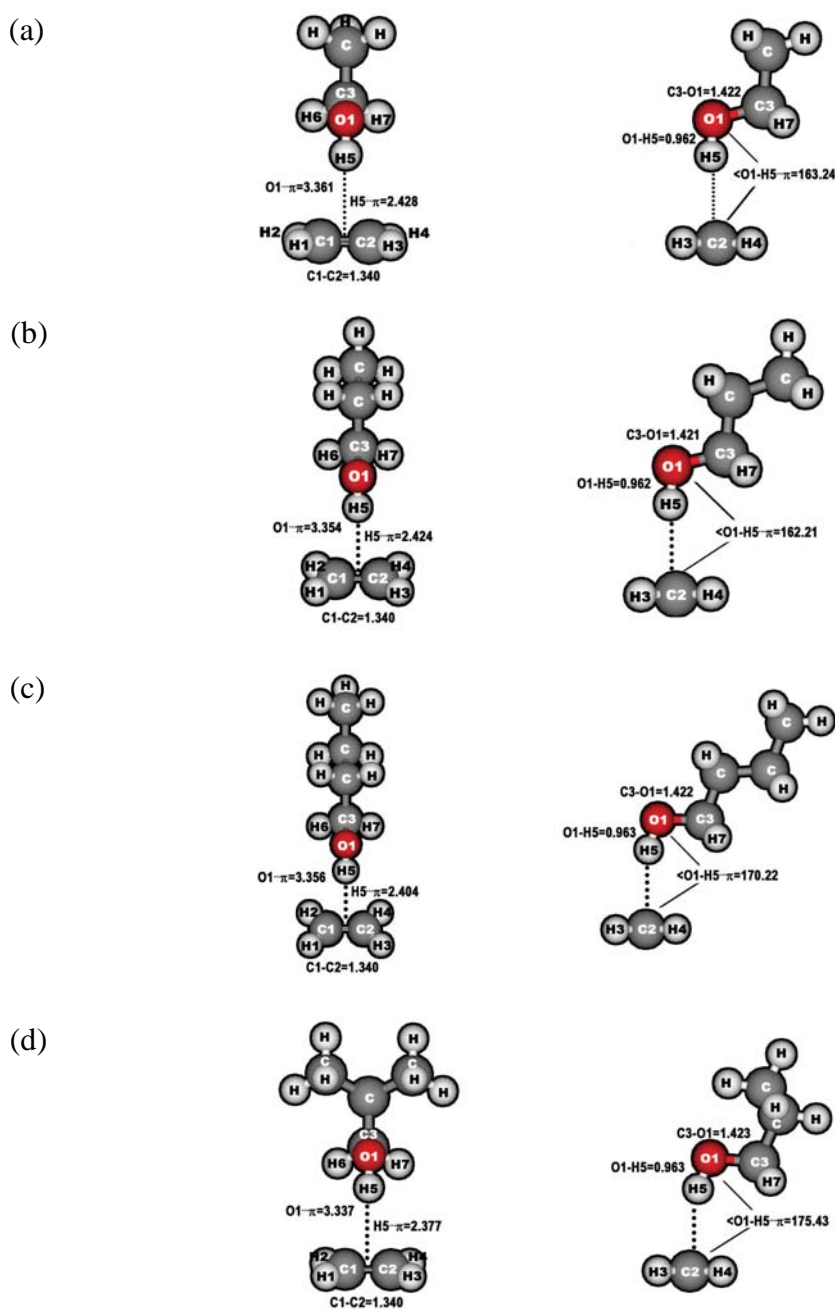
Figure 9 The ONIOM optimized structures of alcohol-ethylene systems at the ONIOM2M methods stand for the MP2(full)/6-311+G(d,p):MP2(full)/6-31G(d) combined methods (a) ethanol-ethylene complex (b) propanol-ethylene complex (c) n-butanol-ethylene complex (d) iso-butanol-ethylene complex

Table 4 Selected bond distances (Å) and binding energy (BE) of the ONIOM optimized structure of alcohol-ethylene systems at the ONIOM2M and ONIOM2B methods stand for the MP2(full)/6-311+G(d,p):MP2(full)/6-31G(d) and MP2(full)/6-311+G(d,p):B3LYP/6-31G(d) combined methods, respectively

Parameter	Ethanol-ethylene		Propanol-ethylene		n-Butanol-ethylene		iso-Butanol-ethylene	
	ONIOM2M	ONIOM2B	ONIOM2M	ONIOM2B	ONIOM2M	ONIOM2B	ONIOM2M	ONIOM2B
Distance (Å)								
H5···π	2.428	2.433	2.424	2.409	2.404	2.423	2.377	2.402
O1···π	3.361	3.375	3.354	3.364	3.356	3.375	3.337	3.362
O1-H5	0.962	0.962	0.962	0.962	0.963	0.962	0.963	0.961
C3-O1	1.422	1.423	1.421	1.423	1.422	1.423	1.423	1.424
C1=C2	1.340	1.340	1.340	1.340	1.340	1.340	1.340	1.340
C1-H1	1.085	1.085	1.085	1.085	1.085	1.085	1.085	1.085
C2-H2	1.085	1.086	1.085	1.085	1.085	1.085	1.085	1.085
Angle(Degree)								
<O1-H5···π	163.24	166.29	162.21	172.03	170.22	170.11	175.43	176.70
<C3-O1···π	94.99	97.18	94.26	101.32	100.03	99.76	103.86	104.88
Energy(kcal/mol)								
BE(uncorr.)	-3.23	-3.11	-3.27	-3.11	-3.24	-3.10	-3.37	-3.18
BE(corr.)	-1.98	-1.93	-2.00	-1.95	-2.04	-1.95	-2.15	-2.02

3. <u>Structure and binding energy of ethanol-ethylene complex</u>

In this thesis section, the intermolecular interaction between ethanol and ethylene system will be investigated in the detail for reduce the computational demand both in the quantum level of calculation and also in the ONIOM method for using in the next application of the interaction between nevirapine inhibitor and the HIV-1 reverse transcriptase binding site complex.

The calculated results obtained from the MP2 and B3LYP methods with a 6-31G(d,p) basis set are shown in Table 5. Considering the effect of the CP-corrected gradient optimization, this method affects the intermolecular distances. The CP-corrected optimized geometries produced a longer H···π distance when compared with the standard optimization. The results were caused by a nonphysical attraction between the two fragments introduced by the BSSE correction. The binding energy differences between the two optimization procedures are almost identical to the uncorrected and CP-corrected binding energies.

When comparing the full and frozen core (FC) electron correlations in the MP2 method, only small differences were found in terms of the geometrical parameters and the binding energies. Furthermore, the CP-corrected binding energy after optimization showed similar results. Thus, in the conclusion the CP-corrected binding energy is less sensitive to the CP-corrected geometry than the standard geometry. This should clearly be true as the binding energy obtained from the CP-corrected optimization must be higher than that at the standard optimized minimum.

Considering the results obtained from the ONIOM model, different ONIOM calculated results are shown in Table 5. It can be seen that the H···π interaction shows little sensitivity to the choice of high-level ONIOM model. Interestingly, the MP2 calculation of the H···π interaction in the high-level ONIOM method is acceptable for both the full and frozen core electron correlation. Therefore, the MP2 frozen core electron correlation will be applied in the high-level ONIOM method in the next section.

Table 5 Selected bond distances (Å) and binding energy (BE) of the ethylene-ethanol complex, optimized using different methods with 6-31G(d) basis set.

	MP2(Full)		MP2(FC)		B3LYP		ONIOM	
	Standard	CP-corr.	Standard	CP-corr.	Standard	CP-corr.	MP2(full):B3LYP	MP2(FC):B3LYP
Distance (Å)								
H5···π	2.386	2.569	2.397	2.569	2.425	2.524	2.393	2.404
O1···π	3.337	3.533	3.348	3.534	3.392	3.486	3.349	3.360
O1-H5	0.965	0.965	0.966	0.966	0.968	0.968	0.964	0.965
C3-O1	1.421	1.421	1.423	1.423	1.420	1.420	1.422	1.423
C1=C2	1.337	1.336	1.338	1.337	1.333	1.332	1.337	1.338
C1-H1	1.080	1.080	1.081	1.081	1.087	1.087	1.080	1.081
C2-H2	1.080	1.080	1.081	1.081	1.087	1.087	1.080	1.081
Angle(Degree)								
<O1-H5···π	168.68	176.95	168.07	177.00	177.32	172.47	171.38	170.60
<C3-O1···π	98.95	105.19	98.46	105.17	109.99	113.60	100.89	100.27
Energy(kcal/mol)								
BE(uncorr.)	-3.65	-3.42	-3.57	-3.37	-2.96	-2.91	-3.54	-3.46
BE(corr.)	-1.86	-2.09	-1.88	-2.09	-1.91	-1.97	-1.81	-1.82

4. Structure and binding energy of HIV-1 RT binding site/nevirapine complex

4.1 Interaction energy contribution based on the HAF and BBF

In the hetero atom fixed (HAF) all the heavy atoms of the amino acids in the binding pocket were fixed at the X-ray structure and therefore the position of the nevirapine inhibitor as well as the position of all hydrogen atoms were optimized, whiles the backbone atoms fixed (BBF), only the backbone atoms of the amino acids were fixed. The binding site in this study was obtained from X-ray structure consisting of 22 amino acids within a 10 Å diameter centered at nevirapine inhibitor. Hydrogen atoms were added to the X-ray structure to generate the complete model system of the HIV-1 RT binding site structure. Therefore, the positions of atoms were optimized with the semiempirical PM3 method. This structure was used as the starting geometry for all calculations.

The interaction energies between nevirapine inhibitor with the individual amino acid residues (called generally X_i) have been calculated for the optimized complex structures, obtained from the X-ray geometry optimized by PM3. The calculated interaction energies contributed from each residue that surrounds the binding pocket using HAF and BBF approximations are shown in Table 6.

The obtain results show that there are more attractive interaction between nevirapine inhibitor and the residues surrounding the binding site of the HIV1-reverse transcriptase. The interaction energies for Leu100, Tyr181 and Ty188 are the main contributors (-6.08, -5.81 and -7.63 kcal/mol at the HAF and -6.58, -5.14 and -8.87 kcal/mol at BBF, respectively). The BSSE-corrected interaction energies are -2.37, -2.67 and -4.52 kcal/mol at the HAF and are -3.09, -2.79 and -5.05 kcal/mol at BBF for the three main amino acid contributors, respectively. The HAF and the BBF give the interaction energy not clearly different. Therefore, these results can be useful and helpful for considering the inner layer of the ONIOM3 level of calculations which will be presented in the next section and demonstrated in Figure 10.

Table 6 Interaction energies of nevirapine with individual residues (X_i), calculated at the MP2/6-31G(d,p) level of calculation using HAF and BBF approximations

Residue	Interaction energy (kcal/mol)	
	HAF	BBF
	Uncorr.(corr.)	Uncorr.(corr.)
PRO095	-1.86(-1.10)	-1.39(-0.87)
LEU100	**-6.08(-2.37)**	**-6.58(-3.09)**
LYS101	-2.55(-1.46)	-2.29(-1.10)
LYS102	-0.44(-0.25)	-0.58(-0.37)
LYS103	-2.38(-1.36)	-2.04(-1.20)
LYS104	0.00(0.00)	0.02(0.02)
SER105	-0.09(-0.09)	-0.16(-0.16)
VAL106	-4.07(-1.70)	-2.60(-0.08)
VAL179	0.05(1.44)	0.41(1.47)
ILE180	-1.16(-0.70)	-0.73(-0.50)
TYR181	**-5.81(-2.67)**	**-5.14(-2.79)**
TYR188	**-7.63(-4.52)**	**-8.87(-5.05)**
VAL189	-0.97(-0.62)	-1.14(-0.64)
GLY190	-1.61(-0.39)	0.07(1.65)
PHE227	-2.71(-1.73)	-2.67(-1.74)
LEU228	-0.19(-0.19)	-0.20(-0.20)
TRP229	-4.61(-2.23)	-3.02(-1.19)
LEU234	-0.95(0.92)	-1.21(1.22)
HIS235	-3.80(-2.31)	-4.22(-2.39)
PRO236	-3.14(-1.56)	-2.83(-1.63)
TYR318	-2.99(-1.38)	-3.28(-1.80)
GLU138(b)	-0.20(-0.08)	-0.15(-0.09)
Total	-53.18(-24.32)	-48.59(-20.52)
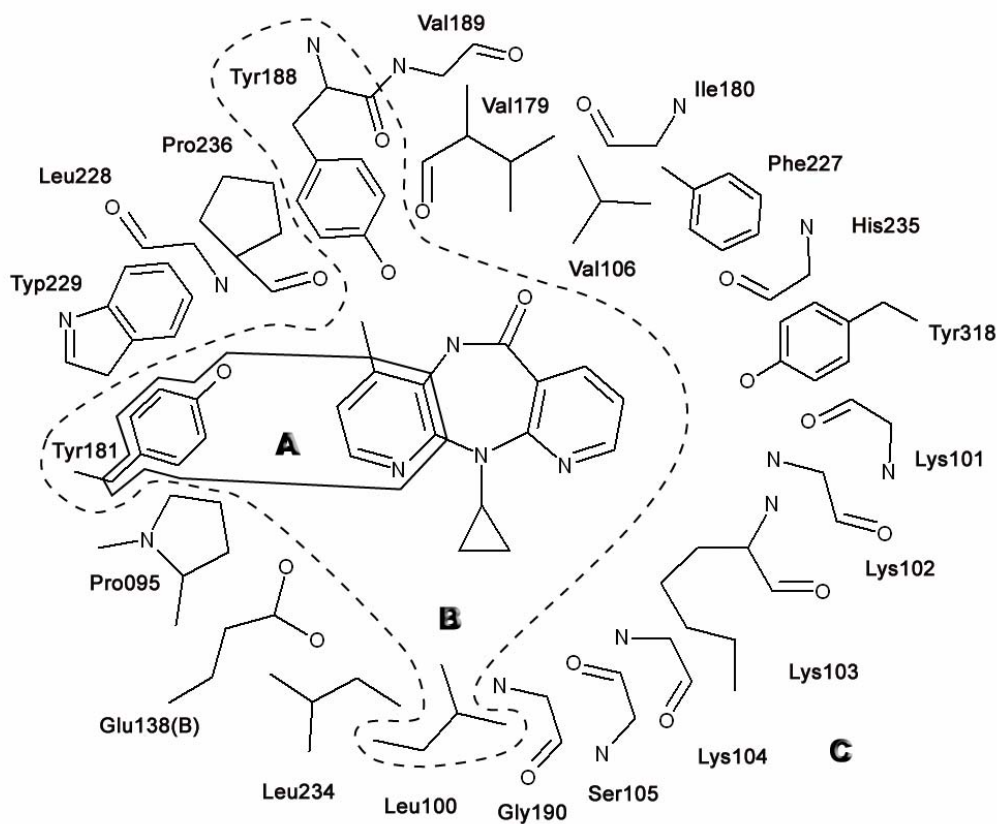
Figure 10 Adopted model system of nevirapine bound to the wild type of HIV-1 RT binding site. Layer partitioning is shown for ONIOM3 model.

4.2 Structure and binding energy for wild type

The main focus of this study is to investigate the interaction between nevirapine and the binding site of HIV-1 RT. However, this complex system is too large for high level (MP2) calculations. Therefore, we divided the model system into three parts and applied the three-layer ONIOM3 method as already described.

Table 7 shows the selected optimized distances between heavy atoms in the interacting core region treated at the MP2/6-31G(d,p) level of calculation, as well as the binding energy and its component of nevirapine bound to the binding site of HIV-1 RT. The optimization was performed using two approaches: hetero atom fixed (HAF) and backbone atom fixed (BBF). It was found that the binding energies are

-14.83 and -15.93 kcal/mol for HAF and BBF, respectively. When the basis set superposition error was corrected using the counterpoise method, the binding energies were reduced to -8.79 and -10.45 kcal/mol for HAF and BBF, respectively. The BBF optimization allows the side chain of residues to be relaxed in the complex and produces an increase in binding energies from 1.10 to 1.66 kcal/mol. It was found that the binding energy difference between the two approaches comes from $\Delta\Delta E$(Low, ABC-AB) of about 1.50 kcal/mol. These results imply that the relaxation of residues during optimization has less effect on the binding energy obtained from the medium-level and high-level calculations.

Table 7 Selected optimized inter-fragment distances (Å), binding energies (ΔE, kcal/mol) and components for the nevirapine and HIV-1 RT complex for the wild type, optimized using the ONIOM(MP2/6-31G(d,p):B3LYP/6-31G(d,p):PM3) method.

Bond distances (Å)	Expt.	HAF[a]	HAF	BBF
N1-C7	3.67	3.89	3.76	3.56
N1-C8	3.93	4.14	3.94	3.88
C2-C8	3.22	-	3.39	3.25
C2-C9	3.78	3.81	3.79	3.56
C2-C13	3.33	-	3.55	3.70
C3-C8	3.52	3.83	3.72	3.42
C3-C12	3.39	4.10	3.57	3.63
C3-C13	3.04	3.57	3.34	3.33
Energy Components				
ΔE		-8.85	-14.83(-8.79)	-15.93(-10.45)
ΔE (High, A)		-1.05	-3.62(-1.43)	-3.82(-1.45)
$\Delta\Delta E$ (Mid, AB-A)		1.21	-2.02(1.83)	-1.42(1.69)
$\Delta\Delta E$ (Low, ABC-AB)		-9.01	-9.19	-10.69

[a]Hannongbua *et al*., 2003, value in parenthesis represent corrected binding energy.

The HAF approach, previously using 16 residues (Hannongbua *et al*., 2003) and presently using 22 residues represented in the binding pocket, yields longer inter-fragment distances than the experimental results obtained from the X-ray crystallographic data. Surprisingly, the BBF optimized geometries produced shorter inter-fragment distances for N1-C7, N1-C8, C2-C9 and C3-C8 than the experimental results. The H3 attached to C3 of nevirapine points toward the inner region of the tyrosine ring close to the C8 of Tyr181 (Figure 11) and the distances between C2-C8 and C2-C9 are 3.25 and 3.56, respectively. Therefore, it is important to note that the hydrogen bonding between the nitrogen (N1) atom of nevirapine and the hydrogen atom of C beta (C7) in Tyr181 play an important role in the binding pocket of the HIV-1 RT/nevirapine complex and may also help form a facial H···π interaction via the H3 of the pyridine ring with Tyr181.

However, the HAF optimized kept the conformation of the binding pocket as found in the X-ray structure. In analyzing the binding energies components of the ONIOM method, the HAF gives -3.62, -2.02 and -9.19 kcal/mol from high level, medium level and low level of calculation, respectively. When the ONIOM-BSSE was corrected, the binding energies from high level and medium level reduce to -1.43 and 1.83 kcal/mol, respectively. Comparing the binding energies components between the HAF and the BBF, it was found that two approaches distribute the binding energies components from high level and medium level almost identical with the binding energies different less than 0.15 kcal/mol. Consequently, the ONIOM result presented here was helpful for considering in the section. Therefore, the HAF optimization procedures were applied on the structure, interaction energy contribution of nevirapine inhibitor with each residue in the Y181C mutant type and also the binding energy of nevirapine inhibitor complex with the HIV-1 reverse transcriptase biding site for the Y181C mutant type.

(a)

(b)

(c)
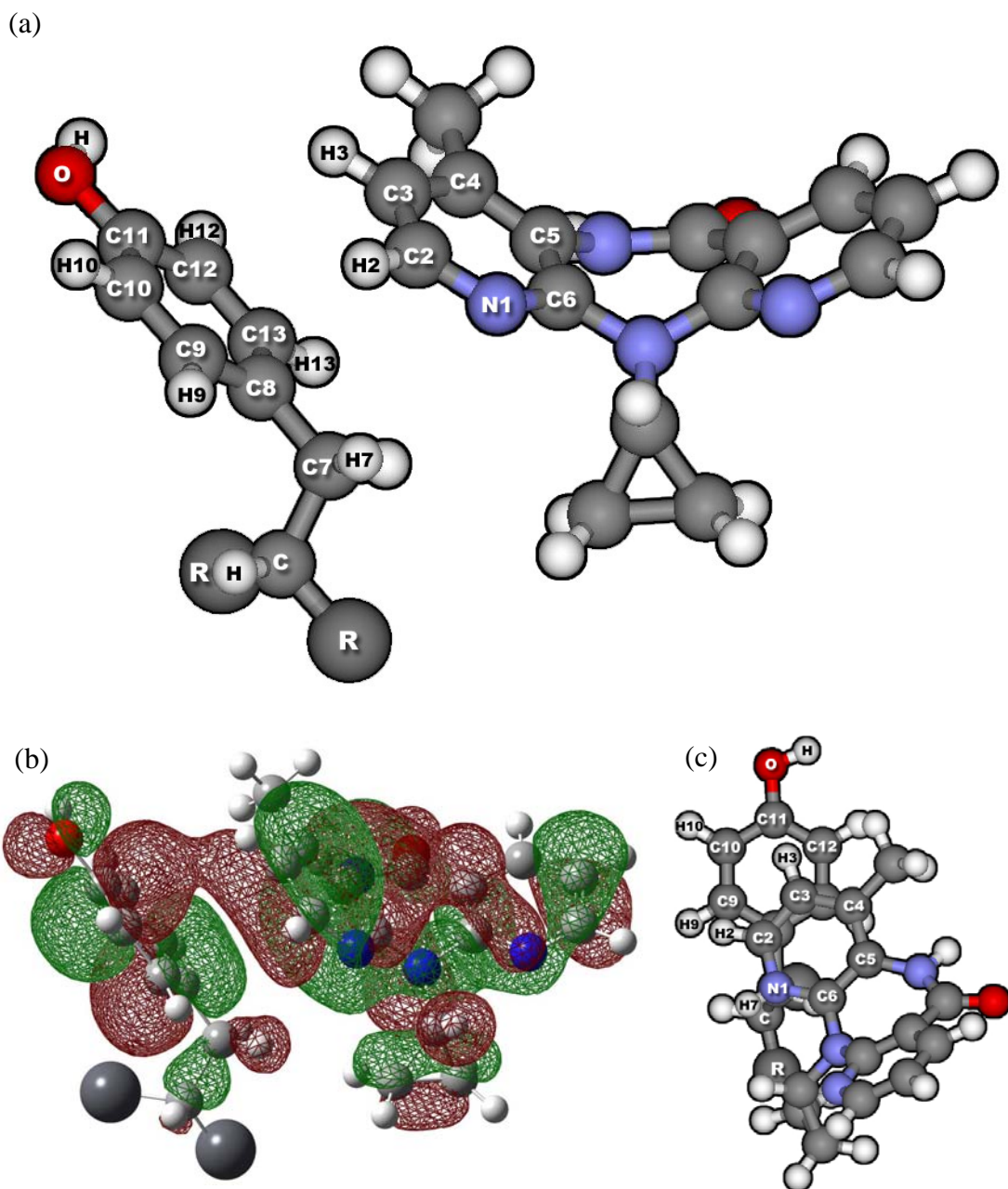
<u>Figure 11</u> Optimized structure of nevirapine and Tyr181 complex from ONIOM(MP2/6-31G(d,p):B3LYP/6-31G(d,p):PM3) (a) label of atom (b) overlap orbital of the H3 of pyridine ring with Tyr181 and (c) top view

4.3 Interaction energy contribution of nevirapine in the Y181C mutant type

        In order to generate the binding site of nevirapine bound to the HIV-1 reverse transcriptase in the Y181C mutant type starting geometry was obtained from the X-ray structure consisting of 22 amino acids within a 10 Å diameter centered at nevirapine inhibitor. Hydrogen atoms were added to the X-ray structure to generate the complete model system of the HIV-1 RT binding site structure. Therefore, the positions of atoms were optimized with the semiempirical PM3 method. This structure was used as the starting geometry for all calculations between the HIV-1 RT binding sites in the Y181C mutant type complex and nevirapine inhibitor.

        The interaction energies between nevirapine inhibitor and individual residues have been calculated based on the HAF optimized procedure of the X-ray structure. The interaction energy of each pair was calculated at the MP2/6-31G(d,p) level of calculation. The calculated interaction energies between nevirapine inhibitor and each amino acid in the binding site on the Y181C mutant type structure are shown in Table 8. It is clearly seen that there are more attractive interactions between nevirapine inhibitor and the residues surrounding the binding site of the HIV1-RT. The uncorrected interaction energies of Leu100, Tyr188 and Trp229 are -5.44, -6.14 and -4.96 kcal/mol, respectively. These three amino acids are the main contributors. Furthermore, the corrected interaction energies are -1.46, -1.75 and -2.85 kcal/mol for Leu100, Tyr188 and Trp229, respectively. Comparing the interaction energies contribution on each amino acid surrounding the nevirapine inhibitor between wild type and mutant type, it was found that the main contributor from two structures give one amino acid different. This amino acid is Tyr181 from the wild type structure and Trp229 from the mutant type structure. Therefore, these results can be helpful for considering the inner model layer of ONIOM3 calculations for the mutant type structure. The main focus of this study is to investigate the binding energy different between the wild type and the mutant type structure. Consequently, three amino acids were selected based on this investigation on the quantum level of calculation in the ONIOM method. These three amino acids are Leu100, Tyr181C and Tyr188. Accordingly, both the wild type and mutant type produce the identical ONIOM model.

Table 8 Interaction energies of nevirapine with individual residues (X_i), calculated at the MP2/6-31G(d,p) level of calculation using HAF approximation

Residue	Interaction energy (kcal/mol)	
	Uncorrected	Corrected
PRO095	-0.77	-0.66
LEU100	**-5.44**	**-1.46**
LYS101	-2.38	-1.38
LYS102	-0.53	-0.28
LYS103	-2.51	-1.40
LYS104	0.02	0.02
SER105	-0.17	-0.17
VAL106	-3.11	0.30
VAL179	-1.46	-0.64
ILE180	-0.56	-0.43
CYS181	**-3.40**	**-1.06**
TYR188	**-6.14**	**-1.75**
VAL189	-1.86	-0.87
GLY190	-1.35	0.20
PHE227	-3.82	-2.12
LEU228	-0.38	-0.38
TRP229	-4.96	-2.85
LEU234	-0.86	2.28
HIS235	-2.65	-1.77
PRO236	-0.58	1.03
TYR318	-3.40	-1.72
GLU138(b)	-1.03	-0.94
Total	-47.32	-16.03

4.4 Structure and binding energy for the Y181C mutant type

The main focus of this section is to investigate the interaction between nevirapine and the binding site of HIV-1 RT in the Y181C mutant type structure based on the HAF optimized procedure. However, this complex system is too large for high level (MP2) calculations. Therefore, we divided the model system into three parts and applied the three-layer ONIOM3 method similar to the previous section as shown Figure 12.

Table 9 shows the selected optimized distances for heavy atoms in the interacting core region treated at the MP2/6-31G(d,p) level of calculation, as well as the binding energy and its component of nevirapine bound to the binding site of HIV-1 RT. It was found that the binding energies are -5.59 kcal/mol. When the basis set superposition error was corrected using the counterpoise method, the binding energies were reduced to -4.65 kcal/mol (the values in parenthesis represent the corrected binding energy). The HAF approach of HIV-1 RT in the Y181C mutant type represented in the binding pocket, yields longer inter-fragment distances than the experimental results obtained from the X-ray crystallographic data.

Table 9 Selected optimized inter-fragment distances (Å), binding energies (ΔE, kcal/mol) and components for the nevirapine and HIV-1 RT complex for the Y181C mutant type, optimized using the ONIOM(MP2/6-31G(d,p):B3LYP/6-31G(d,p):PM3) method.

Bond distances (Å)	Expt.	HAF
N1-C7	3.42	3.62
N1-S8	3.77	4.18
C2-S8	3.02	3.28
C3-S8	3.59	3.89
Energy Components		
ΔE	-	-5.59(-4.65)
ΔE (High, A)	-	-0.44(1.31)
ΔΔE (Mid, AB-A)	-	3.86(3.36)
ΔΔE (Low, ABC-AB)	-	-9.32
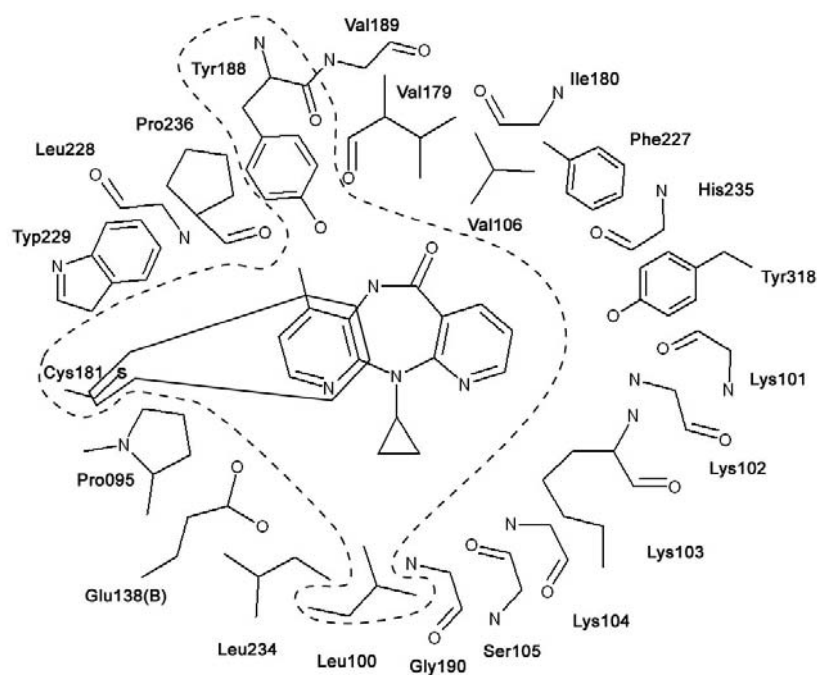
Figure 12 Adopted model system of nevirapine bound to the mutant type of HIV-1 RT binding site. Layer partitioning is shown for ONIOM3 model.

49

(a)
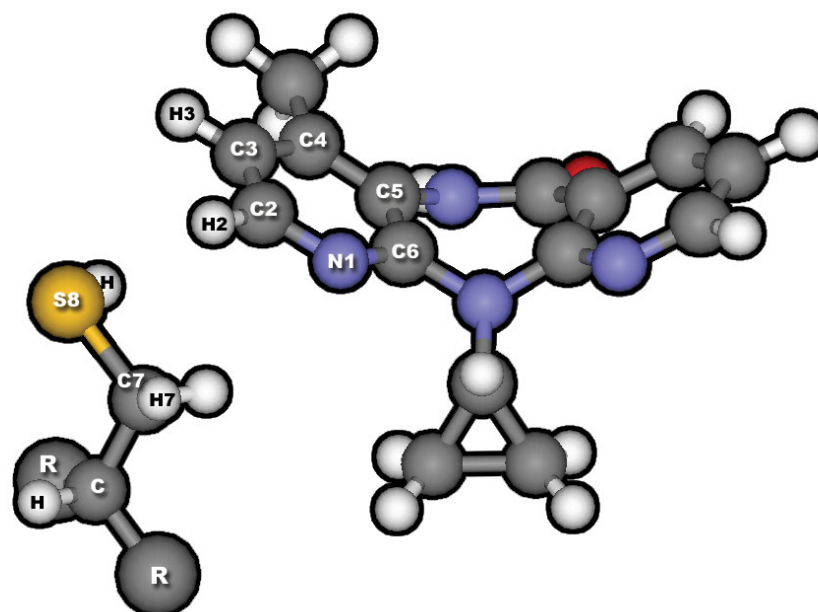
(b)
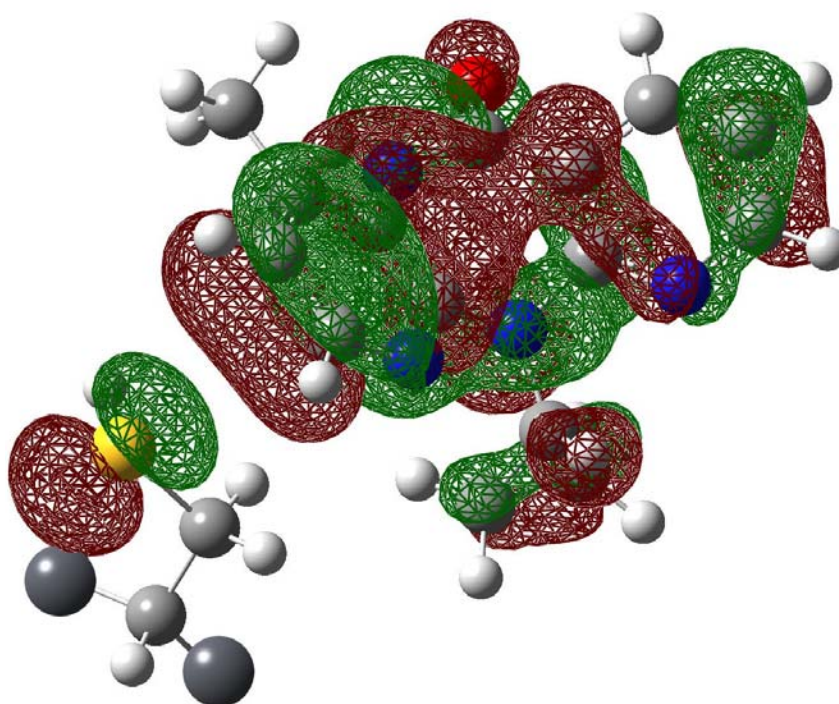
Figure 13 Optimized Structure of nevirapine and Tyr181 complex for the Y181C
mutant type from ONIOM(MP2/6-31G(d,p):B3LYP/6-31G(d,p):PM3) (a)
label of atom (b) the HOMO of nevirapine with Y181C mutant type

4.5 Comparison of wild type and Y181C mutant type of HAF

The interaction energies between nevirapine inhibitor with the individual residue have been calculated for the optimized complex structures. The interaction energy of each pair (nevirapine-each residue) used MP2/6-31G(d,p) level of calculation.

The calculated interaction energies between nevirapine and each residue that surrounds the binding pocket are shown in Table 10 for the wild type and the mutant type (Y181C). It is clearly seen that there are more repulsive interactions between nevirapine and each residue surrounding the binding pocket for the Y181C mutant than for the wild type. These repulsive interactions in the Y181C mutant are the origin of the reduced stability of nevirapine binding compared to the wild type.

Moreover, there are some significant residues that produce the large difference (about 2 kcal/mol) between the wild type and Y181C mutant type, which is the residues of Pro236 and Cys181 showing more influences on the interaction energies in the enzyme complexes. This effect decreases the stability of nevirapine upon binding in the Y181C/RT complex.

The interaction energies of Leu100 (approximately -6.08(-2.37) and -5.44 (-1.456) kcal/mol for wild type and Y181C mutant type, respectively) and Tyr188 (approximately -7.63(-4.52) and -6.14(-1.75) kcal/mol for wild type and Y181C mutant type, respectively) clearly indicated stronger interaction than the other residues which is the main contribution. Moreover, these energies of each residue are not significant different when compared between the wild type and the mutant type system. Results of these interactions were applied to ONIOM3 calculations on the next section which will be helpful to explain the efficiency drop of the inhibitor against Y181C of HIV-1 RT.

Table 10   Comparison of interaction energy between wild type and mutant type have
been used HAF approximation

Residue	Interaction energy (kcal/mol)		
	WT	MT	ΔΔE(kcal/mol)
	Uncorr.(corr.)	Uncorr.(corr.)	Uncorr.(corr.)
PRO095	-1.86(-1.10)	-0.77(-0.66)	-1.09(-0.44)
LEU100	**-6.08(-2.37)**	**-5.44(-1.46)**	**-0.64(-0.91)**
LYS101	-2.55(-1.46)	-2.38(-1.38)	-0.17(-0.08)
LYS102	-0.44(-0.25)	-0.53(-0.28)	0.09(0.03)
LYS103	-2.38(-1.36)	-2.51(-1.40)	0.13(0.04)
LYS104	0.00(0.00)	0.02(0.02)	-0.02(-0.02)
SER105	-0.09(-0.09)	-0.17(-0.17)	0.08(0.08)
VAL106	-4.07(-1.70)	-3.11(0.30)	-0.96(-2.00)
VAL179	0.05(1.44)	-1.46(-0.64)	1.51(2.08)
ILE180	-1.16(-0.70)	-0.56(-0.43)	-0.60(-0.27)
TYR181CYS	**-5.81(-2.67)**	**-3.40(-1.06)**	**-2.41(-1.61)**
TYR188	**-7.63(-4.52)**	**-6.14(-1.75)**	**-1.49(-2.77)**
VAL189	-0.97(-0.62)	-1.86(-0.87)	0.89(0.25)
GLY190	-1.61(-0.39)	-1.35(0.20)	-0.26(-0.59)
PHE227	-2.71(-1.73)	-3.82(-2.12)	1.11(0.39)
LEU228	-0.19(-0.19)	-0.38(-0.38)	0.19(0.19)
TRP229	-4.61(-2.23)	-4.96(-2.85)	0.35(0.62)
LEU234	-0.95(0.92)	-0.86(2.28)	-0.09(-1.36)
HIS235	-3.80(-2.31)	-2.65(-1.77)	-1.15(-0.54)
PRO236	-3.14(-1.56)	-0.58(1.03)	**-2.56(-2.59)**
TYR318	-2.99(-1.38)	-3.40(-1.72)	0.41(0.34)
GLU138(b)	-0.20(-0.08)	-1.03(-0.94)	0.83(0.86)
Total	-53.18(-24.32)	-47.32(-16.03)	-5.85(-8.30)

The calculation of the binding energies for nevirapine inhibitor with binding pocket in comparison between wild type and Y181C mutant type are the main application of this research. However, these complex systems are too large for high level (MP2) calculations. Therefore, the model systems have been divided into three parts and applied the three-layer ONIOM3 method. Table 11 shows the selected optimized distances between heavy atoms in the interacting core treated at the MP2/6-31G(d,p) level of calculation, as well as the binding energy and its component of nevirapine bound to the binding site of HIV-1 RT. As the difference of binding energy between wild type and Y181C mutant type is 9.24 kcal/mol and using BSSE corrected the binding energy between wild type and Y181C mutant type is 4.14 kcal/mol. The HAF optimization was found that the binding energy of nevirapine in the binding pocket of HIV-1 RT comes from $\Delta\Delta E$(Low, ABC-AB) for wild type and Y181C mutant type are -9.19 kcal/mol and -9.32 kcal/mol, respectively. These results clearly show that the Y181C mutation decreases the stabilization energy of nevirapine bound to its binding pocket when compared to the wild type which these affect form ΔE(High, A) reduce from -3.62(-1.43) kcal/mol to -0.44(1.31)kcal/mol for the wild type and Y181C mutant type, respectively and $\Delta\Delta E$(Mid, AB-A) reduce from -2.02(1.83) and 3.86(3.36) for the wild type and Y181C mutant type, respectively. This consideration of the binding energy leads to another strong argument for the explanation that nevirapine has higher biological activity to the wild type than the Y181C mutant enzyme.

Table 11 Comparison of energy component between wild type and mutant type have
been used HAF approximation

Energy Components	WT	MT	$\Delta\Delta E^a$
	ONIOM	ONIOM	
ΔE	-14.83(-8.79)	-5.59(-4.65)	9.24(4.14)
ΔE (High, A)	-3.62(-1.43)	-0.44(1.31)	3.18(2.74)
$\Delta\Delta E$ (Mid, AB-A)	-2.02(1.83)	3.86(3.36)	5.88(1.53)
$\Delta\Delta E$ (Low, ABC-AB)	-9.19	-9.32	0.13

[a]$\Delta\Delta E = |\Delta E_{WT} - \Delta E_{MT}|$ and the values in parenthesis represent the corrected binding energy.

CONCLUSIONS

This thesis was performed the ONIOM-BSSE scheme in order to investigate the geometry and the binding energy of the ethanol-ethylene, propanol-ethylene, n-butanol-ethylene, and iso-butanol-ethylene systems formed OH···π system complex and the nevirapine bound to the HIV-1 RT binding pocket complex.

Based on two optimization schemes, the obtained results of alcohol-ethylene systems indicate that the standard gradient optimization provides the lower binding energy than the CP-corrected gradient optimization with the energy different less than 0.24 kcal/mol. However, the BSSE-corrected binding energies results indicate that the latter geometry has less effect of the binding energy than the former geometry. In comparison of the non-ONIOM and ONIOM methods, it was found that the MP2 calculation of the OH···π interaction in the high-level ONIOM method is acceptable for both the full and frozen core electron correlation. Consequently, the BSSE-corrected binding energies are similar.

The complex structure between the HIV-1 RT binding site and the nevirapine inhibitor suggests that the nitrogen (N1) of the pyridine ring forms hydrogen bonding and produces the facial H···π interaction via nevirapine and Tyr181. The analysis of the ONIOM-BSSE binding energy shows that the binding energy in the interacting core region (small model) is very weak (about -1.4 kcal/mol) and the substantial binding energy comes almost exclusively from the interaction of nevirapine with other residues in the binding pocket The three-layered ONIOM (MP2/6-31G(d,p):B3LYP/6-31g(d,p):PM3) calculations have been successfully applied to determine the binding energy of nevirapine inhibitor bound to the HIV-1 RT binding pocket. In comparison between wild type and Y181C mutant type, the results clearly indicate that the Y181C substitution is more electrostatic repulsion than the wild type RT which affected to decrease the stabilization energy of nevirapine bound to its binding pocket. Finally, this thesis shown that it is possible to apply the combined high and low quantum chemical methods based on various approaches such as Møller-Plesset perturbation theory, Density Functional Theory (DFT) and

semiempirical methods studied bimolecular system and, accordingly, it is feasible for correction ONIOM energy by using the ONIOM-BSSE scheme. Moreover, the three-layered ONIOM method is recommended as best with compromised combined methods for future studies of similar inhibitor-enzyme interactions.

LITERATURE CITED

Augspurger, J.D., C.E. Dykstra and T.S. Zwier. 1993. Model study of the structures and stabilities of benzene-$(H_2O)_2$-12 complexes. **J. Phys. Chem.** 97: 980-984.

Berman, H.M., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov and P.E. Bourne. 2000. The Protein Data Bank. **Nucl. Acids. Res.** 28: 235-242.

Brutschy, B., J. Eggert, C. Janes, and H. Baumgaertel. 1991. Nucleophilic substitution reactions in molecular clusters following photoionization. **J. Phys. Chem**. 95: 5041-5050.

Carney, J.R., and T.S. Zwier, 1999. Infrared and ultraviolet spectroscopy of water-containing clusters of indole, 1-methylindole, and 3-Methylindole. **J. Phys. Chem. A.** 103: 9943-9957.

Castleman, A.W., and B.D. Kay. 1985. Molecular beam electric deflection study of the hydrogen bonded water, methanol, and ethanol cluster $(H_2O)N$, $(CH_3OH)N$, and $(C_2H_5OH)N$. **J. Phys. Chem**. 89: 4867-4868.

Courty, A., M. Mon, I. Dimicoli, F. Piuzzi, M.P. Gaigeot, V. Brener, P. de Pujo, and P. Millie. 1998. Quantum effects in the threshold photoionization and energetics of the bezene-H_2O and Benzene-D_2O complexes: experiment and simulation. **J. Phys. Chem. A.** 102: 6590-6600.

Duan, C., Y. Meng, Z. Zhou, B. Wang, and Q. Zhong. 2005. Non-convertional hydrogen bonding interaction of BH_3NH_3 complexes: a comparative theoretical study. **J. Mol. Struct. (Theochem)**. 713: 135-144.

Ettischer, I., U. Buck, 1998. Vibrational predissociation spectra of size selected methanol clusters: new experimental results. **J. Chem. Phys**. 108: 33-38.

Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morukuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Radbuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, J. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. 2003. **Gaussian 03**. Gaussian Inc., Pittsburgh, PA.

Fruip, D.J., L.A. Curtiss, and M. Blander. 1979. Studies of molecular association in H_2O and D_2O vapors by measurement of thermal conductivity. **J. Chem. Phys**. 71: 2703-2711.

Gordon, M.S., I. Admovic, H. Li, and M.H. Lamm. 2006. Modeling Styrene-Styrene interactions. **J. Phys. Chem. A**. 110:519-525.

Gruenloh, C.J., F.C. Hagemeister, J.R. Carney, and T.S. Zwier. 1999. Resonant ion-dip infrared spectroscopy of ternary benzene-(water)$_n$(methanol)$_m$ hydrogen-bonded clusters. **J. Phys. Chem. A**. 103: 503-513.

Hannongbua, S., M. Kuno, and K. Morokuma. 2003. Theoretical investigation on nevirapine and HIV-1 reverse transcriptase binding site interaction, based on ONIOM method. **Chem. Phys. Lett**. 380: 456-463.

Hannongbua, S., M. Kuno, S. Saen-oon, and P. Nunrium. 2005. Particular interaction between efavirenz and the HIV-1 reerse transcriptase binding site as explained by the ONIOM2 method. **Chem. Phys. Lett**. 405: 198-202.

_____, _____, and _____. 2005. Binding energy analysis for wild-type and Y181C mutant HIV-1 RT/8-C1 TIBO complex structures: quantum chemical calculations based on the ONIOM method. **Prot.: Struct. Func. Bioinf**. 61: 859-869.

Hobza, P. and P. Jurecka. 2003. True stabilization energies for the optimal planar hydrogen-bonded and stacked structures of guanine¨cytosine, adenine¨thymine, and their 9- and 1-methyl derivatives: Complete basis set calculations at the MP2 and CCSD(T) levels and comparison with experiment. **J. Am. Chem. Soc**. 125: 15606-15613.

_____, and E. Mrazkova. 2003. Hydration of sulfo and methyl groups in diethyl sulfoxid is accompanied by the formation of red-shifted hydrogen bonds and improper blue-shifted hydrogen bonds: an ab inito quantum chemical study. **J. Phys. A**. 107:1032-1039.

_____, and J. Pittner. CCSDT and CCSD(T) calculations on model H-bonded and stacked complexes. 2004. **Chem. Phys. Lett**. 390: 496-499.

_____, and K.M. Dethlefs. 2000. Noncovalent interactions: A challenge for experiment and theory. **Chem. Rev**. 100: 143-167.

_____, and Z. Havlas. 2000. Blue-shifting hydrogen bonds. **Chem. Rev**. 100:4253-4264.

_____, B.J. Veken, W.A. Herrebout, R. Szostak, D.N. Shchepkin, and Z. Havlas. 2001. The nature of improper, blue-shifting hydrogen bonding verified experimentally. **J. Am. Chem. Soc**. 123: 12290-12293.

Hobza, P., C. Riehn, A. Weichert, and B. Brutschy. 2002. Structure and binding energy of the phenol dimer: correlated ab initio calculations compared with results from rotational coherence spectroscopy. **Chem. Phys**. 283: 331-339.

_____, H.L. Selzle, and E.W. Schlag. 1996. Potential energy surface for the benzene dimer. Results of ab inito CCSD(T) calculations show two nearly isoenergetic structure: T-shaped and parallel displaced. **J. Phys. Chem**. 100: 18790-18794.

_____, J. Chocholousova, and J. Vacek. 2003. Acetic acid dimer in the gas phase, nonpolar solvent, microhydrated environment, and dilute and concentrated acetic acid: Ab initio quantum chemical and molecular dynamics simulations. **J. Phys. Chem. A**. 107: 3086-3092.

_____, _____, _____, F. Huisken, and O. Werhahn. 2002. Stacked structure of the glycine dimer is more stable than the cyclic plannar geometry with two O-H…O hydrogen bond: Concerted action of empirical, high-level nonempirical ab initio, and experimental studies. **J. Phys Chem. A**. 106:11540-11549.

_____, P. Jurecka, J. Sponer. 2004. Potential energy surface of the cytosine dimer: MP2 complete basis set limit interaction energies, CCSD(T) Correction term, and Comparison with the amber force field. **J. Phys. Chem. B**. 108: 5466-5471.

_____, V. Spirko, H.L. Selzle, and E.W. Schlag. 1998 Anti-hydrogen bond in the benzene dimer and other carbon proton donor complexes. **J. Phys Chem. A.** 102: 2501-2504.

Iwata, S., and H. Watanabe. 1996. Theoretical studies of geometric structures of phenol- water clusters and their infrared absorption spectra in the O-H stretching region. **J. Chem. Phys**. 105: 420-431.

Jordan, K.D., S.Y. Fredericks, and T.S. Zwier. 1996. Theoretical characterization of the structures and vibrational spectra of benzene-(H2O)n (n=1-3) clusters. **J. Phys. Chem**. 100: 7810-7821.

Kim, K.S., and P. Tarakeshwar. 2002. Comparison of the nature of p and conventional H-bonds: a theoretical investigation. **J. Mol. Struct**. 615: 227-238.

_____, B.J. Mhin, U. Choi, and K. Lee. 1992. Ab inito studies of the water dimer using large basis sets: The structure and thermodymamic energies. **J. Chem. Phys**. 97: 6649-6662.

_____, _____, and I. Bandyopadhyay. 2005. Phenol vs Water molecule interacting with various molecules: σ-type, π-type, and χ-type hydrogen bonds, interaction energy, and their energy components. **J. Phys. Chem**. 109: 1720-1728.

_____, P. Tarakeshwar, and J.Y. Lee. 2000. Molecular clusters of π-systems: Theoretical studies of structures, spectra and origin of interaction energies. **Chem. Rev**. 100: 4145-4185.

Linse, P., G. Karistroem, and B. Joensson. 1984 Monte Carlo studies of a dilute aqueous solution benzene. **J. Am. Chem. Soc**. 106: 4096-4102.

Lu, J., Z. Zhou, Q. Wu, and G. Zhao. 2005. Density functional theory study of the hydrogen bonding interaction in formamide dimer. **J. Mol. Struct. (Theochem)**. 724: 107-114.

Matsuda, Y., T. Ebata, and N. Mikami. 1999. Vibrational spectroscopy of 2-pyridone and its clusters in supersonic jets: Structures of the clusters as revealed by characteristic shifts of the NH and C=O bands. **J. Chem. Phys**. 110:8397-8407.

Mons, M., I. Dimicoli, B. Tradivel, F. Piuzzi, V. Brenner, and P. Millie. 1999. Site dependence of the binding energy of water to indole: microscopic approach to the side chain hydration of tryptophan. **J. Phys. Chem. A.** 103: 9958-9965.

Morokuma, K. 1971. Molecular orbital studies of hydrogen bonds. III. C=O··H-O hydrogen bond in H_2CO··H_2O and H_2CO··$2H_2O$. **J. Chem. Phys**. 55: 1236-1244.

_____, M. Svensson, S. Humbel, R.D.J. Froese, T. Matsubara, and S. Sieber. 1996. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alder Reactions and $Pt(P(t\text{-Bu})_3)_2 + H_2$ Oxidative Addition. **J. Phys. Chem**. 100: 19357-19363.

Muguet, F., and G.W. Robison. 1995. Towards a new correction method for the basis set superposition error: Applicati to the ammonia dimer. **J. Chem. Phys**.102: 3648-3654.

Nelander, B., and A. Engdahl. 1983. The acetylene-water complex: A matrix isolation study. **Chem. Phys. Lett.** 100: 129-132.

_____, and _____. 1985. A matrix isolation study of the benzene-water interaction. **J. Phys. Chem.** 89: 2860-2864.

Odutola, J.A., R. Viswanathan, and T.R. Dyke. 1979. Molecular beam electric deflection behavior and polarity of hydrogen-bonded complexes of ROH, RSH, and RNH. **J. Am. Chem. Soc**. 101: 4787-4792.

Ren, J, R. Esnouf, E. Garman, D. Somers, C. Ross, I. Kirby, J. Keeling, G. Darby, Y. Jones and D. Stuart. 1995. High resolution structures of HIV-1 RT from four RT-inhibitor complexes. **Nat. Struct. Biol.** 2: 293-302.

Schiefke, A., C. Deusen, C. Jacoby, M. Gerhards, M. Schmitt, and K. Kleinermanns. 1995. Structure and vibrations of the phenol-ammonia cluster. **J. Chem. Phys**. 102: 9197-9204.

Sherrill, C.D. and M.L. Abrams. 2005. Important configureations in configuration interaction and coupled-cluster wave functions. **Chem. Phys. Lett**. 412: 121-124.

_____, and M.O. Sinnokrot. 2003. Unexpected substituent effects in face-to-face π-stacking interactions. **J. Phys. Chem. A**. 107:8377-8379.

_____, and _____. 2004. Highly accurate coupled cluster potential energy curves for the benzene dimer: Sandwich, T-shaped, and Parallel-displaced configurations. **J. Phys. Chem. A**. 108: 10200-10207.

_____, and _____. 2004. Substituent effects in π-π interactions: Sandwich and T-shaped configurations. **J. Am. Chem. Soc**. 126: 7690-7697.

_____, and T.P. Tauer. 2005. Beyond the benzene dimer: An investigation of the additivity of p-p interactions. J. Phys. Chem. A. 109: 10475-10478.

_____, E.F. Valeev, M.L. Abrams, and T.D. Crawford. 2002. The equilibrium geometry, harmonic vibrational frequencies, and estimated ab initio limit for the barrier to planarity of the ethylene radical cation. **J. Phys. Chem. A**. 106: 2671-2675.

_____, _____, and M.O. Sinnokrot. 2002. Estimates of the ab inito limit for π-π interactions: the benzene dimer. **J. Am. Chem. Soc**. 124: 10887-10893.

_____, T.P. Tauer, and M.E. Derick. 2005. Estimates of the ab inito limit for sulfur-π interactions: The H_2S-Benzene dimer. **J. Phys. Chem. A**. 109: 191-196.

Tantillo, C., J. Ding, A. Jacobo-Molina, R.G. Nanni, P.L. Boyer, S.H. Hughes, R. Pauwels, K. Andries, P.A.J. Janssen and E. Arnold. 1994. Locations of anti-AIDS drug binding sites and resistance mutations in the three-dimensional structure of HIV-1 reverse transcriptase: implications for mechanisms of drug inhibition and resistance. **J. Mol. Biol.** 243(3): 369-387.

Tarakeshwar, P., H.S. Choi, and K.S. Kim. 2001. Olefinic vs Aromatic π-H interaction: A theoretical investigation of the nature of interaction of first-row hydrides with Ethene and benzene. **J. Am. Chem. Soc**. 123: 3323-3331.

_____, H. Houjou, Y. Nagawa, and K. Hiratani. 2000. High-level ab inito calculations of torsional potential of phenol, antisole, and *o*-hydroxyanisole: Effects of intramolecular hydrogen bond. **J. Phys. Chem. A**. 104: 1332-1336.

_____, K. Honda, T. Uchimaru, M. Mikami, and K. Tanabe. 1998. New medium-size basis sets to evaluate the dispersion interaction of hydrocarbon molecules. **J. Phys. Chem. A**. 102: 2091-2094.

_____, _____, _____, _____, and _____. 1999. High-level ab initio calculations of interaction energies of C_2H_4-CH_4 and C_2H_6-CH_4 dimer: A model study of CH/π interaction. **J. Phys Chem. A**. 103: 8265-8271.

_____, _____, _____, _____, and _____. 2000. The magnitude of the CH/π interaction between benzene and some model hydrocarbons. **J. Am. Chem. Soc**. 132: 3746-3753.

_____, _____, _____, _____, and _____. 2001. Origin of attraction and directionality of the π/π interaction: Model chemistry calculations of benzene dimer interaction. **J. Am. Chem. Soc**. 124: 104-112.

Tarakeshwar, P., K. Honda, T. Uchimaru, M. Mikami, and K. Tanabe. 2002. High level ab initio calculations of intermolecular interaction of propane dimer: Orientation dependence of interaction energy. **J. Phys. Chem. A**. 106: 3867-3872.

_____, _____, _____, _____, and _____. 2002. The interaction of benzene with cholro- and fluoromethanes: Effects of halogenation on CH/π interaction. **J. Phys. Chem. A**. 106: 4423-4428.

Wu, Q., H. Zhang, Z. Zhou, J. Lu, and G. Zhao. 2005. Study of the hydrogen bonding interaction of 1:1 complexes of serine with formamide using density fucional theory. **J. Mol. Sturct. (Theochem)**. 757: 9-18.

Zheng, W., W. Wang, X. Pu, A. Tian, and N.B. Wong. 2003. Effect of CP-corrected gradient optimization on the water-radical (anion) dimer hypersurface. **J. Mol. Struct. (Theochem).** 631: 171-180.

APPENDIX

APPENDIX A

<u>Theoretical background</u>

In all times throughout the history, people have tried to understand the nature. The best way to explore the laws of nature is by doing experiments. A couple of hundred years ago, people have tried to understanding about almost everything especially for big things and small things. Practical experiments can be hard to conduct on very small systems. In these cases calculations based on mathematical model can be helpful to increase the knowledge and understanding of microscopic phenomena.

With the fundamental laws of quantum mechanics and molecular mechanics, the great number of different calculation methods and even more calculation software of varying quality were used. Nowadays, the information technological revolution has provided faster and cheaper computer systems for the availability of high performance computers. This has yielded a tremendous increase of computer-based calculations in the field of atomic and molecular science. Computer based calculations have a corollary in molecular design, where the control of a single molecule, in particular its three-dimensional structure and its binding to an enzyme, is a crucial step in the molecular design process. Hence, computer-aided molecular design has a become part of the industrial pharmaceutical research.

Computational chemistry has to be applicable to a certain system size or a number of atoms per molecule or unit cell. Many molecules of practical interest have less than 100 atoms. The system size with 100 atoms per model system allows to study by many computational methods. For the larger systems, the standard computational method is not possible to be used because of the computational expensive. In this thesis, molecular interaction in biological system has been studied on a large number of atoms in molecular model system. Some part of this thesis has been focused on the small molecular model system that provides key interaction of the HIV-1 RT. Then, in order to get structural and energetic information of the system

studies, the choice of the methods has been must considered.

1. <u>Choices of methods</u>

The first choice for the decision is between quantum mechanical and molecular mechanical methods. If a large system with more than thousands of atoms and the process with no breaking or forming bond, molecular mechanical methods is preferable. However, one need to be aware that generality and transferability of force fields cannot be taken for granted and the results obtained from molecular mechanical calculations can be misleading.

Among quantum mechanical methods, the next choice is between *ab initio* method, such as HF (Hartree-Fock) or MP2 (The second-order Møller-Plesset perturbation theory) and semiempirical methods. The latter is preferable because of the speed. However, semiempirical methods need to be used with great care on the range of applicability of the empirical parameters and the natural approximation. The *ab initio* method typically provides reliable results, but the sensitivity of the results on the choice of computational parameters in particular the choice of the basis functions and the level of correlation requires attention. Now, density functional theory (DFT) tends to be stronger than the Hartree-Fock method in the sense of a reasonable choice of the basis functions and other parameters, such as the geometrical structures, the vibrational frequencies and the electronic structure. The obtained results from density functional theory calculations usually do not give large and unexpected deviation from experiment. These levels of methods are possible only for rather small systems.

In theoretical studies of larger systems, the combined quantum mechanics and molecular mechanics (QM/MM) method has been a major goal of computational chemistry over the past decade. In the QM/MM method the system is divided into two parts, a small part and a significantly larger part. The small part represents chemically active sites where a reaction will occur and a larger inactive part. The chemically active site is treated with QM, while the inactive part is treated with MM. The mechanical influence of the inactive part constrains the geometry of the active site

and therefore has an effect on its chemistry. The key to the success of a hybrid QM/MM technique is the manner in which the influence of the MM region is communicated to the QM region.

2. General methods

Quantum mechanical calculations on atomic systems can be done in many ways. One way is to make calculations that in many ways rely on experimentally determined parameters. This is called semiempirical calculations. Another approach to perform the calculations is to use foundations of quantum mechanics. The approach is called *ab initio* when it makes no use of empirical information, except for the fundamental constants of the nature such as mass of the electron, Planck's constant etc., which are required the numerical predictions. Appendix Table A1 attempts to show advantages and disadvantages of semiempirical, *ab initio* and density function theory (DFT) methods. According to quantum mechanics postulates, the state of a system is fully described by a wave function that depends on the position of the electrons and nuclei in the system.

$$H\Psi = E\Psi \qquad (1)$$

where H is the Hamiltonian operator which gives the kinetic, T, and potential, V, energies of the system that is;

$$H = T + V \qquad (2)$$

and

$$T = -\frac{\hbar^2}{2m}\nabla^2 \qquad (3)$$

then

$$H = -\frac{\hbar^2}{2m}\nabla^2 + V \tag{4}$$

Then, rewrite equation (1) is;

$$\left\{ -\frac{\hbar^2}{2m}\nabla^2 + V \right\}\Psi = E\Psi \tag{5}$$

where the Laplacian operator, ∇^2, is;

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{6}$$

$\hbar$ is Plank's constant divided by 2π. Ψ is the wave function which characterizes the particle's properties. E is the eigenenergy of the particle corresponding to wave function.

In order to simplify the Schrödinger equation, one can separate into the electronic and nuclear motion. This is the Born-Oppenheimer approximation. This means that the nuclei are seen as fixed and the Schrödinger equation is solved for the electrons moving in the field of the fixed nuclei. The Born-Oppenheimer Hamiltonian then looks like

$$H = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} \tag{7}$$

Even with these simplifications, the Schrödinger equation cannot be solved analytically for many-electron systems.

<u>Appendix Table A1</u> Advantages and disadvantage for the use of semiempirical, *ab initio*, density function theory (DFT) method.

Method	Advantages	Disadvantages
Semiempirical	Uses quantum physics	Requires experimental data or
	Uses experimentally derived	data from *ab inito* for
	empirical parameters	parameters
	Uses approximation extensively	Less rigorous than *ab inito*
	Less demanding computationally	methods
	than *ab initio* methods	
	Capable of calculating transition	
	state and excited states	
Ab initio	Uses quantum physics	Computationally expensive
	Mathematically rigorous	
	No empirical parameters	
	Uses approximation extensively	
	Useful for a broad range of	
	systems	
	Does not depend on experimental	
	data	
	Capable of calculating transition	
	state and excited states	
DFT	Uses quantum physics	Even gradient-corrected
	Mathematically rigorous	functionals apparently are
	No empirical parameters	unable to handle van der Waals
	Uses electron density	interactions
	Does not depend on experimental	For excited states will being
	data	developed.
	Working for ground-state	

To make calculations using Schrödinger equation possible, either approximated numerical or analytical wave functions can be introduced. The wave functions then generate the potential and from the potential, the set of wave functions can be refined. This procedure is repeated until a stable, self-consistent solution of the Schrödinger equation is obtained. A possible model for the trial wave functions is to construct it from molecular orbitals (MO). Molecular orbitals are often constructed by linear combinations of atomic orbitals (AO).

$$\psi_i^{MO} = \sum_{\upsilon} c_{\upsilon i} \phi_\upsilon^{AO} \tag{8}$$

3. **Hartree-Fock theory**

In Hartree-Fock–theory (HF) each electron is represented by a spin orbital that is a product of an orbital wave function and a spin function, and the total wave function *psi* (ψ) is assumed to be represented by a single Slater determinant, i.e., the antisymmetrized product of dpin orbitals. The electron moves in the field of the nuclei and the average field of the other electrons in the system. The spin orbital give the electron wave function that minimizes the Rayleigh ratio (E_{HF}) are found by variation theory.

$$E_{HF} = \min \frac{\int \psi^* H \psi \, d\tau}{\int \psi^* \psi \, d\tau} \tag{9}$$

where H is the Born-Oppenheimer-Hamiltonian. The lowest energy (E_{HF}) is the electronic energy for the system and is called the Hartree-Fock limit. This procedure leads to the Hartree-Fock equations for the individual spin orbitals. The HF equation for spin orbital $\phi(n)$ (where electron n is assigned to the spin orbital ϕ_a) is

$$f_n \phi_a(n) = \varepsilon \phi_a(n) \tag{10}$$

where ε is the orbital energy of the spin orbital and f_n is the Fock operator;

$$f_n = h_n + \sum_n \{J_u(n) - K_u(n)\} \tag{11}$$

Here h_n is the core Hamiltonian for electron n, the sum is over all spin orbitals u and the Coulomb operator J and the exchange operator K are defined as

$$J_u(1) = \int \phi_u^*(1) \frac{1}{r_{12}} \phi_u(2) d\vec{x}_2 \tag{12}$$

$$K_u(1) = \int \phi_u^*(1) \frac{1}{r_{12}} \phi_u(2) d\vec{x}_2 \tag{13}$$

Each spin orbital is then calculated by solving equation (10). Since f_n depends on the spin orbitals of all other electrons in the system, the solution must be iterated from a trial set of spin orbitals until the solution is self-consistent, hence the name of the process: self-consistent field (SCF). The Hartree-Fock theory can be divided into two different cases depending on how to incorporate the spin of the system into the calculations. If electrons with spin α are considered equal as electrons with spin β, it is called restricted HF (RHF). The restricted HF approach can not be used on systems with open electron shells. When different spins are treated as different spatial orbitals, it is called unrestricted HF (UHF). Treating the spins as different gives a double set of equations, matrices and integrals to compute and hence gives better values for calculated energies since it brings the possibility to splitting of the energy levels for spin α and spin β.

Electron correlation

Since in the Hartree-Fock theory each electron moves in the average field of all the other electrons, it does not take into account the quantum mechanical effects on electron distributions nor does it consider the instantaneous electrostatic interactions between electrons. In other words, the Hartree-Fock theory ignores electron

correlation. There are two major ways to incorporate electron correlation in the Hartree-Fock theory, perturbation theory and configuration interaction. Configuration interaction uses a linear combination of Salter determinants or configuration state functions corresponding to for excitations of electrons from the Hartree-Fock reference function to form the actual wave function, but this method are not used in this study and will not be discussed further.

4. <u>**Møller-Plesset (MP) Perturbation Theory**</u>

In Møller-Plesset perturbation theory, the zero-order Hamiltonian, H^0, is given by the sum of one-electron Fock operators from equation (11).

$$H^0 = \sum_i^n f_i \qquad (14)$$

Moreover, the perturbation, H^1, is given by

$$H^1 = H - H^0 \qquad (15)$$

H is the Born-Oppenheimer Hamiltonian. If the eigenenergy of H^0 and H^1 are E^0 and E^1 respectively. The second-order perturbation theory gives the first correction, E^2, to the ground-state energy

$$E^2 = \sum_{n \neq 0} \frac{\langle \phi_n | H^1 | \phi_0 \rangle \langle \phi_0 | H^1 | \phi_n \rangle}{E^0 - E^n} \qquad (16)$$

ϕ_0 is the Hartree-Fock ground-state, and ϕ_n is the n-th excited configuration. This specific method that takes use of the second-order perturbation theory is called MP2.

5. Density Functional Theory (DFT)

One major limitation with *ab initio* calculations is the computational difficulty of performing accurate calculations with large basis sets on molecules containing many atoms. Another family of theoretical methods that became increasingly popular is the density functional theory (DFT). The basic idea behind density functional theory is that the energy of the system can be expressed in terms of the electron probability density. The electron density, ρ, is a function of the position vector $\boldsymbol{r}$, $\rho=\rho(\boldsymbol{r})$. The energy E is said to be a functional of the electron density, $E=E[\rho]$. It has been proved that the electron density uniquely determines the ground-state energy.

Using the electron density, the energy functional can be written as

$$E[\rho] = T[\rho] + V_{ne}[\rho] + V_{ee}[\rho] \tag{17}$$

$$E[\rho] = C_F \int \rho^{5/3}(\vec{r})d\vec{r} - Z\int \frac{\rho(\vec{r})}{r}d\vec{r} + V_{ee}[\rho] \tag{18}$$

The subscripts for the potential energy terms stand for nuclei-electron and electron-electron. In Hartree-Fock theory, V_{ee} consists of a Coulomb part and a electron exchange energy. V_{ee} can be decomposed to

$$V_{ee}[\rho] = J[\rho] - K[\rho] \tag{19}$$

J is the Coulomb operator and K the exchange operator (see equations 12 and 13). In density functional theory the $K[\rho]$ term can be modified to include both exchange and correlation effects. How that is done is shown later on.

The first and simplest approach to replace a wave function with the electron density was the Thomas-Fermi atomic model that uses only the Coulomb operator J in the expression for V_{ee}. The model uses the electron gas assumption for calculating the

kinetic energy, T, of a system as a function of the Fermi energy, εF. The total energy functional within the Thomas-Fermi (TF) theory

$$E_{TF}\left[\rho(\vec{r})\right] = T + V_{ne} + V_{ee} \tag{20}$$

$$E_{TF}\left[\rho(\vec{r})\right] = C_F \int \rho^{5/3}(\vec{r})d\vec{r} - Z\int \frac{\rho(\vec{r})}{r}d\vec{r} + \frac{1}{2}\int\int \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}}d\vec{r}_1d\vec{r}_2 \tag{21}$$

where $\rho(\vec{r})$ is the electron density. Since exchange and correlation are not included in the Thomas-Fermi model, bonding in molecules is not predicted. Dirac added an exchange term for a uniform electron gas to the TF energy functional. The exchange term gives a new V_{ee} term on the form of equation (19) where

$$K\left[\rho(\vec{r})\right] = \frac{1}{4}\int \frac{1}{r_{12}}\left|\rho_1(\vec{r}_1,\vec{r}_2)\right|^2 d\vec{r}_1 d\vec{r}_2 \tag{22}$$

This results in a new total energy functional $E_{TFD}\left[\rho\right]$

$$E_{TFD}\left[\rho\right] = C_F \int \rho^{5/3}(\vec{r})d\vec{r} + \int \rho(\vec{r})\upsilon(\vec{r})d\vec{r} - Z\int \frac{\rho(\vec{r})}{r}d\vec{r} - C_X \int \rho^{4/3}(\vec{r})d\vec{r} \tag{23}$$

Even with the addition of the new exchange functional, the model works very poorly. Negative ions do not exist, molecules do not hold together, the energies for real atoms are far too low and $\rho(0) = \infty$. The main reason for these failures is the assumption that the potential is uniform or slowly varying, which is unrealistic. The next improvement of DFT came when Hohenberg and Kohn proved that the electron density could uniquely define the potential. Since the density determines the number of electrons and hence the wave functions for a non-degenerate ground state, it determines all properties of the system. The energy, E_ν, depending on the specific external potential, υ, used can be written

$$E_\upsilon[\rho] = T[\rho] + V_{ne}[\rho] + J[\rho] + V_{XC}[\rho]$$

Here $J[\rho]$ is the Coulombic term and $V_{XC}[\rho]$ is the exchange correlation energy functional. To find the exact density and the exact form of the exchange-correlation functional are the two main difficulties of DFT.

5.1 Kohn-Sham equations

One way to incorporate correlation only is the Kohn-Sham (KS) equations, which to some extent can be described as a correlation corrected Hartree-Fock approach. Their approach was to split the complicated kinetic energy functional, $T[\rho]$, into a part on the independent particle form and a remainder that adds to the unknown exchange-correlation energy, $E_{XC}[\rho]$. This is done in different ways for the spin-paired or closed shell situation and spin unrestricted or odd-electron situation. These models treat the density as slowly varying and therefore approximate the density as being constant locally. Thus the name of the procedure is local density approximation (LDA).

5.2 Local Density Approximation

The total electron density can be assumed to be expressed as a sum of contributions from a set of single-particle (so-called Kohn-Sham) orbitals,

$$\rho(\vec{r}) = \sum_i^N \sum_s |\psi_i(\vec{r}, s)|^2 \tag{24}$$

These orbitals form a wave function that exactly describes a system containing N non-interacting electrons. The corresponding simplified part of the kinetic energy, $T_S[\rho]$, can thereby be treated exactly in the form

$$T_S[\rho] = \frac{1}{2}\sum_i^N \left\langle \psi_i \left| -\nabla^2 \right| \psi_i \right\rangle \tag{25}$$

This treatment results in an approach more accurate than TF or TFD models, but also harder to compute. The total energy, $E[\rho]$, is then given by

$$E[\rho] = T_S[\rho] + J[\rho] + E_{XC}[\rho] + \int \upsilon[\vec{r}]\rho(\vec{r})d\vec{r} \tag{26}$$

Using the orthonormal of the wave functions and the fact that the derivative of the functional above should be zero for a minimum lead to the Kohn-Sham orbital equations

$$\left[-\frac{1}{2}\nabla^2 + \upsilon_{eff}\right]\psi_i = \varepsilon_i\psi_i \tag{27}$$

These are the DFT equivalence to the single particle Schrödinger equation in the Hartree-Fock theory. The effective potential, V_{eff}, used in equation 27 is

$$V_{eff} = V(\vec{r}) + \frac{\delta J[\rho]}{\delta\rho(\vec{r})} + \frac{\delta E_{XC}[\rho]}{\delta\rho(\vec{r})} \tag{28}$$

$$V_{eff} = V(\vec{r}) + \int \frac{\rho(\vec{r})}{r_{12}} + \upsilon_{XC}(\vec{r}) \tag{29}$$

The last term is the single-particle exchange-correlation potential. Since the Kohn-Sham equations are non-linear, they must be solved iteratively in a SCF way similar to that of Hartree-Fock-theory. The KS equations are easier to solve than the corresponding Hartree-Fock equations, as the latter contain an exchange potential operator in the one-electron Hamiltonian. The choice of Kohn-Sham orbitals has been a problem, but for practical purposes they resemble their Hartree-Fock equivalencies well enough, although there is not always a 1:1 correspondence.

5.3 Local Spin Density Approximation

If the electron shells are not closed there will be a difference in density between electrons with α-spin and β-spin. The total density is then $(\rho^{\alpha}+\rho^{\beta})$ and the spin density is $(\rho^{\alpha}-\rho^{\beta})$. This spin-polarized model, called the local spin density approximation (LSDA) is not only superior to the former LDA in odd-electron systems but also for closed shell systems. This is due to the fact that LSDA allows electrons with different spins to have different densities. The difference between LDA and LSDA is thus like their Hartree-Fock theory counterparts restricted and unrestricted. Separating the different spin densities gives a new Kohn-Sham kinetic energy

$$T_S\left[\rho^{\alpha},\rho^{\beta}\right]=\frac{1}{2}T_S^0\left[2\rho^{\alpha}\right]+\frac{1}{2}T_S^0\left[2\rho^{\beta}\right] \tag{30}$$

The superscript 0 denotes the spin-restricted case, in which $\rho^{\alpha}+\rho^{\beta}=\rho$. Introducing spin polarization of the exchange-correlation energy functional, and separating exchange and correlation contributions gives

$$E_{XC}\left[\rho\right]\rightarrow E_X\left[\rho^{\alpha},\rho^{\beta}\right]+E_C\left[\rho^{\alpha},\rho^{\beta}\right] \tag{31}$$

By using the spin-polarization, ζ,

$$\zeta=\frac{\rho^{\alpha}-\rho^{\beta}}{\rho^{\alpha}+\rho^{\beta}} \tag{32}$$

in the Dirac exchange functional, the exchange energy part becomes

$$E_X^{LSDA}\left[\rho^{\alpha},\rho^{\beta}\right]=C_X\int\rho^{4/3}\left((1+\zeta)^{4/3}+(1-\zeta)^{4/3}\right)d\vec{r} \tag{33}$$

$$E_X^{LSDA}\left[\rho^\alpha,\rho^\beta\right]=\int\rho\varepsilon_X\left(\rho,\zeta\right)d\vec{r} \qquad (34)$$

The exact value of the correlation energy, E_C, per particle is only known for a few special cases,

$$E_C^{LSDA}\left[\rho^\alpha,\rho^\beta\right]=\int\rho\varepsilon_C\left(\rho,\zeta\right)d\vec{r} \qquad (35)$$

Unlike the exchange energy, the correlation part cannot be decomposed into different spin contributions as electrons with equal spins interact as well as those with different spins. The first approach to approximated correlation energy was a model developed by Hedin and von Barth that is based on a random phase approximation (RPA). Their model uses the same form for the partitioning of the correlation part as for the exchange

$$\varepsilon_C^{BM}\left(\rho,\zeta\right)=\varepsilon_C^0\left(r_S\right)+\left(\varepsilon_C^1\left(r_S\right)-\varepsilon_C^0\left(r_S\right)\right)f\left(\zeta\right) \qquad (36)$$

where r_S is the radius of a sphere with the effective volume of an electron. The function $f\left(\zeta\right)$ is weight-factor depending on the spin polarization. The ε^0 and ε^1 terms are complicated analytical functions derived by Hedin and Lundqvist5. Janak, Moruzzi and Williams (JMW) improved the Hedin-von Barth functional by changing some of the parameters. Their functional is called the JMW-functional. Another suggestion to the correlation energy was made by Vosko, Wilk and Nusair (VWN). They improved the model by Hedin-Lundqvist with data fitted from a Monte Carlo calculation of correlation energies made by Ceperley and Adler. Their expression for the correlation energy is

$$E_C^{VWN}=\frac{A}{2}\left[\ln\frac{x^2}{X(x)}+\frac{2b}{Q}\tan^{-1}\frac{Q}{2x+b}-\frac{bx_0}{X(x_0)}\left(\ln\frac{(x-x_0)^2}{X(x)}+\frac{2(b+2x_0)}{Q}\tan^{-1}\frac{Q}{(2x+b)}\right)\right] \qquad (37)$$

where A, b, c, x_0 and Q are fitted constants and $X(x)$ is a second degree

polynomial. These complicated and highly accurate expressions form the VWN functional. An even better correlation-energy model, the PWC functional, was derived by Perdew and Wang as another model with fitted data from the calculations by Ceperley and Adler. This functional is parameterized and hard to write on an explicit form, but simplified it looks like

$$\varepsilon_c^{PWC}(r_s) = -ac_0(1+br_s)\ln\left[1+\left(ac_0\left(dr_s^{1/2}e^{-c_1/c_0} + fr_s e^{-2c_1/c_0} + gr_s^{3/2}e^{-3c_1/c_0}\right)\right)\right] \qquad (38)$$

and here a, b, c, d, f and g are fitted parameters and c_n are expansion terms for the correlation energy of the RPA above. The PWC functional is fitted to more data points than the VWN functional.

5.4 Non-local corrections

To further improve density functional theory, the fluctuation of the electron density needs to be taken into consideration. This can be done by using not only the electron densities but also the gradients of the densities. One simple way to incorporate the gradients is the generalized gradient approximation (GGA), proposed by Perdew and Wang. This approximation depends only on the density and its first spatial derivative, which makes it easy to evaluate comparing with the first non-local corrections that used the second derivatives of the electron density. With the GGA, Perdew and Wang have made a combined exchange-correlation functional, the PW91. For the spin-unrestricted case, the exchange part of the functional is

$$E_X^{PW91}[n_\uparrow, n_\downarrow] = \frac{1}{2}E_X^{PW91}[2n_\uparrow] + \frac{1}{2}E_X^{PW91}[2n_\downarrow] \qquad (39)$$

The restricted component used for each spin-orientation above is

$$E_X^{PW91}[n] = \int n\varepsilon_x(r_s, 0)F(s)d\vec{r} \qquad (40)$$

when

$$\varepsilon_x(r_s,0) = -\frac{3k_F}{4\pi}$$

$$s = \frac{|\nabla n|_s}{2nk_F}$$

$$k_F = \frac{1.9192}{r_s}$$

$F(s)$ is the complicated expression shown below.

$$F(s) = \frac{1 + 0.19645s \ \sinh^{-1}(7.7956s) + s^2\left(0.2743 - 0.1508e^{-100s^2}\right)}{1 + 0.19645s \ \sinh^{-1}(7.7956s) + 0.004s^4} \tag{41}$$

The correlation part of the PW91 is a bit more complicated but looks like

$$E_c^{PW91}[n_\uparrow, n_\uparrow] = \int n\left(\varepsilon_c(r_s,\zeta) + H(t,r_s,\zeta)d\vec{r}\right) \tag{42}$$

when

$$t = \frac{|\nabla n|}{2ngk_s}$$

$$g = \frac{\left[(1+\zeta)^{2/3} + (1-\zeta)^{2/3}\right]}{2}$$

$$k_s = \left(\frac{4k_F}{\pi}\right)^{1/2}$$

The function H used above consists of a lot of fitted parameters, but simplified it looks like

$$H = k_1 g^3 \ln\left[1 + \frac{t^2 + At^4}{k_1\left(1 + At^2 + A^2t^4\right)}\right] + k_2 g^3 t^2 e^{-100 g^4 t^2 \left(\frac{k_s}{k_F}\right)^2} \qquad (43)$$

Another much used non-local correction is the BLYP correction. It consists of an exchange functional (B88) derived by Becke, combined with a correlation correction by Lee, Yang and Parr. The B88 functional, see equation (44), is a gradient correction to the exchange energy given with LDA. It uses one parameter β that is fitted to exact atomic Hartree-Fock data and has the value of 0.0042 a.u.

$$E_X^{B88} = E_X^{LDA} - \beta \sum \int \rho_\sigma^{4/3} \frac{x_\sigma^2}{1 - 6\beta x_0 \sinh^{-1} x_\sigma} d^3 r \qquad (44)$$

when

$$x_\sigma = \frac{|\nabla \rho_\sigma|}{\rho_\sigma^{4/3}}$$

The LYP correlation functional, that is a correction of second order, looks for the spin restricted case like,

$$E_C^{LYP} = -a\int \frac{1}{1 + d\rho^{-1/3}}\left(\rho + b\rho^{-2/3}\left[C_F \rho^{5/3} - 2t_W + \left(\frac{1}{9}t_W + \frac{1}{18}\nabla^2 \rho\right)\right]e^{c\rho^{-1/3}}\right)d\vec{r} \qquad (45)$$

The Weizsacker kinetic energy density, t_W, that is used above instead of the normal Kohn-Sham kinetic energy seen in equation (26) in the expression above, is defined as

$$t_W(r) = \frac{1}{8}\frac{|\nabla \rho|^2}{\rho} - \frac{1}{8}\nabla^2 \rho \qquad (46)$$

The introduction of non-locality mostly enhances the accuracy of the calculations, but also increases the computational difficulties.

6. **Semiempirical methods**

Due to their greatly increased requirement for central processing unit time and storage space in the computer memory, *ab initio* quantum chemical methods were limited in their practical applicability. However, improvements in computer hardware and the availability of easy-to-use programs have helped to make *ab initio* methods a widely used computational tool. The approximate quantum mechanical methods require significantly less computational resources. Semiempirical methods are based on the Roothaan-Hall equations which for a closed-shell system is

$$FC = SCE \qquad (47)$$

In *ab initio* calculations all elements of the Fock matrix are calculated using equation (20), irrespective of whether the basis functions ϕ_μ, ϕ_λ, ϕ_σ and ϕ_ν are on the same atom, on atoms that are bounded or on atoms that are not formally bonded. The semiempirical methods consider the Fock matrix element in three groups: $F_{\mu\mu}$ (the diagonal elements); $F_{\mu\mu}$ (where ϕ_μ and ϕ_ν are on the same atom) and F (where ϕ_μ and ϕ_ν are on different atoms.)

The greatest proportion of the time required to perform an *ab initio* Hartree-Fock SCF calculation is invariably calculating and manipulating integrals. The most obvious way to reduce the computational effort is therefore to neglect or approximate some of these integrals. Semiempirical methods achieve this in part by explicit considering only the valence electrons of the system; the core electrons are assumed into the nuclear core. The overlap matrix, S (in equation (19)), is set equal to the identity matrix, I. The main implication of this is that the Roothaan-Hall equations are simplified: **FC = SCE** becomes **FC = CE**.

6.1 MNDO

Dewar and Thiel introduced the modified neglect of diatomic overlap (MNDO) method which was based on the neglect of diatomic differential overlap (NDDO); this theory only neglects differential overlap between atomic orbitals on different atoms.

6.2 AM1 and PM3

The Austin Model 1 (AM1) and Parametric Method Number 3 (PM3) are based on MNDO (the name derives from the fact that PM3 is the third parameterizations of NMDO, AM1 being considered the second). AM1 and PM3 modified the core-core repulsions just outside bonding distances.

7. Hybrid methods

The basis of the QM/MM method is that the process or subsystem of most interest is localized in a fairly small part of a larger system. Therefore only this small region requires quantum mechanical calculations. The bulk of the system is treated more simply by molecular mechanical methods. The combination of the efficiency and speed of the MM force field with the versatility and range of applicability of the QM method allows reactions in large systems to be studied. A number of hybrid QM/MM methods have been suggested to circumvent the problems of larger systems.

An approach originated by Morokuma group under the general name of the ONIOM (our own N-layered integrated molecular orbital and molecular mechanics) method is based on the principle of the extrapolation and is conceptually quite different although operationally similar in some special cases from the QM/MM methods. Originally, the MO and the MM methods were combined under the name of the integrated molecular orbital-molecular mechanics (IMOMM). Later the method was further expanded to an onion-like multi-layered method, ONIOM, with ONIOMn referring specifically to an n-layered version, and individual methods used are divided

by a colon, such as ONIOM2(MP2:MM3). Therefore, IMOMM is equivalent to ONIOM2(MO:MM) and IMOMO to ONIOM2(MO:MO). The ONIOM method has been implemented into the Gaussian98 program system (Frisch *et al.*, 1998). Additionally, this implementation can also determine integrated energy derivatives with respect to an electric field like dipole moments, polarizabilities, hyperpolarizabilities, infrared intensities and Raman intensities. In this study, the combined methods between MP2, DFT, HF and PM3 were performed. The comparison with the highest level of theory, MP2, was also taken into account.

7.1 ONIOM energy definition

The basic idea behind the ONIOM approach can be explained most easily when it is considered as an extrapolation scheme in a two-dimensional space, spanned by the size of the system on one axis and the level of theory on the other axis. Appendix Figure A1 shows the extrapolation procedure schematically. The goal is to describe the real system at the higher level of the theory, i.e., the approximation of the target E_4 (point 4) in a system partitioned into the two-layer ONIOM or E_9 (point 9) in a system consisting of the three layers. In the case of two layers, the extrapolated energy E_{ONIOM2} is then defined as:

$$E_{ONIOM2} = E_3 - E_1 + E_2 \tag{48}$$

where E_3 is the energy of the entire (real) system calculated at the low level method and E_1 and E_2 are the energies of the model system determined at the low and high level of theory, respectively. E_{ONIOM2} is an extrapolation to the true energy of the real system E_4:

$$E_4 = E_{ONIOM2} + D. \tag{49}$$

Thus, if the error D of the extrapolation procedure is constant for two different structures (e.g. between reactant and transition state), their relative energy ΔE_4 will be evaluated correctly by using the ONIOM energy ΔE_{ONIOM2}.
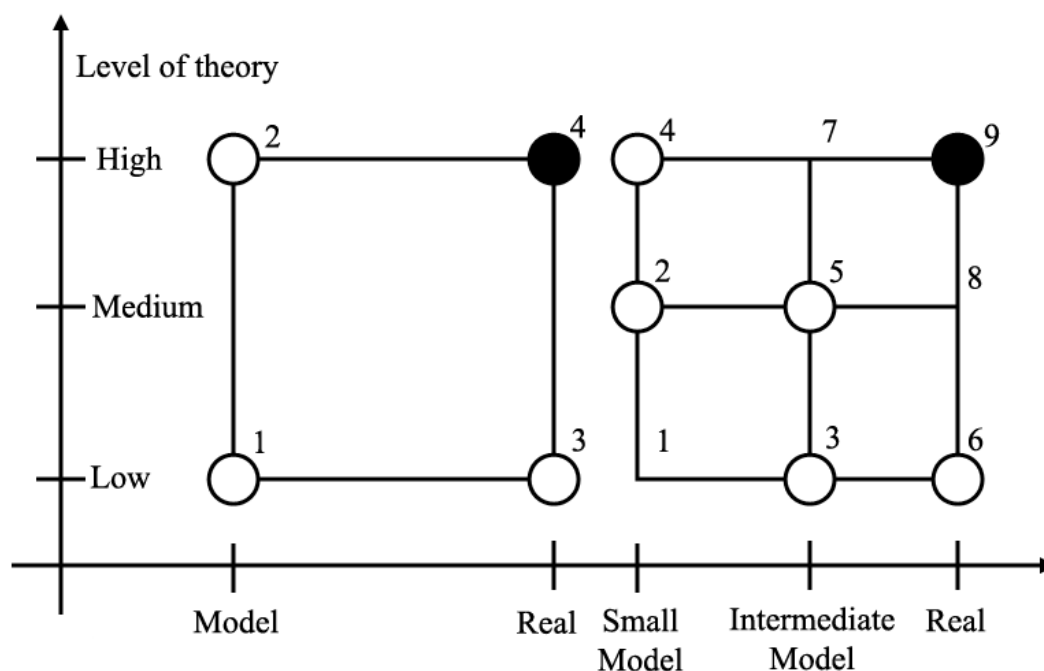
Appendix Figure A1 The ONIOM extrapolation scheme for a molecular system partitioned into two (left) and three (right) layers.

For a system partitioned into three different layers, the expression for the total energy E_{ONIOM3} as an approximation for E_9 reads:

$$E_{ONIOM3} = E_6 - E_3 + E_5 - E_2 + E_4 \tag{50}$$

Since the evaluation of E_1 (the smallest model system at the lowest level of theory) does not require much computational effort, its value can be used to determine the effect of the three-layer approach as compared to a two-layer partitioning with points1, 4 and 6. If the energy difference between the two- and three-layer extrapolation is constant, a two layer partitioning with the intermediate layer omitted would give comparably accurate results.

It should be noted that the layers need not be inclusive or contiguous. The so-called 'inner layer' does not have to be physically inside the 'outer layer'. The

layers can be any part of the system. Each layer does not have to be contiguous; it can consist of several separate regions of the system.

7.2 Treatment of link atoms

An important and critical feature of all the combination schemes is the treatment of link atoms. For the following discussion, the first introduce some useful definitions by adopting a two-layer ONIOM schemes as an example, as illustrated in Appendix Figure A2 (Morokuma *et al.*, 1999). The methodology in the case of a three-layer ONIOM is exactly the same and will not be discussed explicitly.

The atoms present both in the model system and the real system are called set 1 atoms and their coordinates are denoted by R_1. The set 2 atoms are the artificially introduced link atoms. They only occur in the model system and their coordinates are described by R_2. In the real system they are replaced by the atoms described by R_3. Atoms that belong to the outer layer and are not substituted by link atoms are called set 4 atoms with coordinates R_4. The geometry of the real system is thus described by R1, R3 and R4 and they are the independent coordinates for the ONIOM energy:

$$E_{ONIOM} = E_{ONIOM} (R_1, R_3, R_4) \tag{51}$$

In order to generate the model system, described by R_1 and the link atoms R_2, Morokuma *et al.* (1999) define R_2 as a function of R_1 and R_3:

$$R_2 = f (R_1, R_3,) \tag{52}$$

The explicit functional form of the R_2 dependency can be chosen arbitrarily. However, considering the fact that the link atoms are introduced to mimic the corresponding covalent bonds of the real system, they should follow the movement of the atoms they replace. If atom A belongs to set 1 and atom B belongs to set 3 (see Appendix Figure A2), the set 2 link atom (symbolized by H atom in Appendix Figure

A2) is replaced onto the bond axis A-B. Therefore, in the model calculations the link atoms are always aligned along the bond vectors of the real system. For the exact position r_2 of a single H atom along and A-B bond (r_3- r_1), they introduce a fixed scale factor (or distance parameter) g. Hence,

$$r_2 = r_1 + g(r_3 - r_1) \tag{53}$$

If the A-B bond distance $|\, r_3 - r_1\,|$ changes during a geometry optimization, the A-H bond distance $|\, r_2 - r_1\,|$ also changes.

Model System = inner layer + link atoms
Real System = inner layer + outer layer
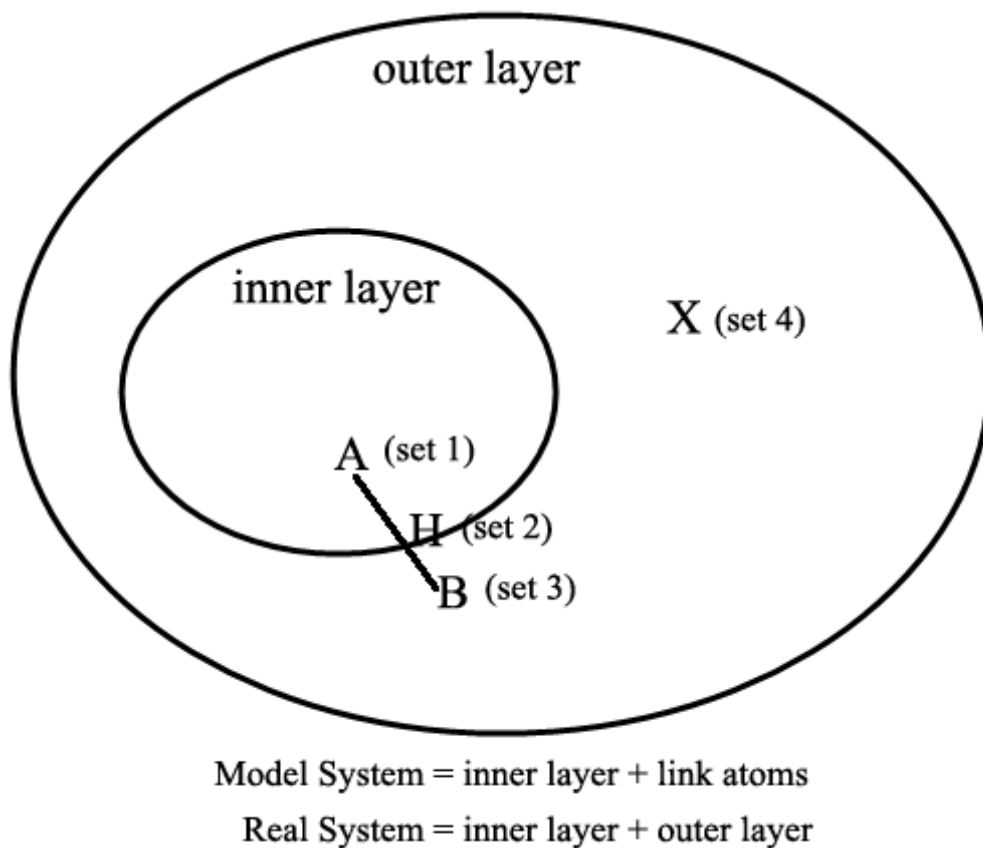
Appendix Figure A2 Definition of different atom sets within the ONIOM scheme.

How should one choose a reasonable value for the parameter g? Of course, this crucially depends on the nature of the cut bond, the atoms A, B and link atom. If

they want to substitute a C-C single bond by a C-H bond in the model calculations, a reasonable value for g would be a standard C-C bond length (1.084 Å) divided by a standard C-C bond length (1.528 Å) and g becomes 0.709. Another common example is the substitution of a P-C bond (typically 1.860 Å) by a P-H bond (1.403 Å) for modeling bulky phosphine groups by using PH_3 in the model calculations, which give 0.754 as a reasonable value for g. It should be noted that the optimal scale factor also depends on the level of theory used for the two layers which are connected by the cut bonds.

8. <u>Basis set superposition error (BSSE)</u>

Usually, in the case of two interacting monomers A and B, the interaction energy (ΔE) is calculated as

$$\Delta E = E_{AB}(AB) - E_A(A) - E_B(B) \tag{54}$$

Where subscripts denote the molecular species in the energy expressions, while the letters in parentheses refer to the basis used in the calculation. Because the unused basis functions of the second unit in the associated complex can augment the basis set of the first, thereby lowering its energy compared to a calculation of the first unit, alone. The first can cause a similar error on the second. So the basis set superposition error (BSSE) is introduced. While alternative methods are used to correct this error, the counterpoise (CP) correction proposed by Boys and Bernardi continues to be the most prominent means of correcting for BSSE. The CP method calculates each of the fragments with the basis functions of the other, using 'ghost orbitals'. The CP-corrected interaction energy can be defined as

$$\Delta E^{CP} = E_{AB}(AB) - E_A(AB) - E_B(AB) \tag{55}$$

Normally, one adds CP correction as a single-point correction to a previously optimized geometry of the complex. For the existence of interaction in the complex, a

little change of energy will make much change of geometry. So, some researchers proposed one should use CP to correct the optimized geometry, as well as the interactions energy. To do that, CP correction of energy is added in each step of gradient optimization.

The scheme for calculating the ONIOM counterpoise correction binding energy (BE) used the Boys-Bernadi and also includes the relaxation effects expressed as shown in Equation (56):

$$BE_{ONIOM2}^{CP} = BE_{HL}^{CP}(m1m2) + BE_{LL}^{CP}(M1M2) - BE_{LL}^{CP}(m1m2) \qquad (56)$$

where CP, HL, LL, m and M denote the counterpoise correction, high level of calculation, low level of calculation, small model in inner layer and real model in outer layer, respectively. The numbers, 1 and 2, stand for a part of monomer1 and a part of monomer2 studied as defined in the ONIOM model system, respectively.

9. Basis Set

The basis set most commonly used in quantum mechanical calculations are composed of atomic functions. The next approximation involves expressing the molecular orbitals as linear combinations of a pre-defined set of one-electron functions known as basis function. An individual molecular orbitals is defined as:

$$\phi_i = \sum_{\mu=1}^{N} c_{\mu i} \chi_{\mu} \qquad (57)$$

where the coefficients $c_{\mu i}$ are known as molecular orbital expansion coefficients. The basis function $\chi_1 ... \chi_N$ are also chosen to be normalized. Gaussian-type atomic functions were used as basis functions. Gaussian functions have the general form

$$g(\alpha, \vec{r}) = cx^n y^m z^l e^{-\alpha r^2} \tag{58}$$

where $\vec{r}$ is of course composed of x, y, and z. α is a constant determining the size (radical extent) of the function. In Gaussian function, $e^{-\alpha r}$ is multiplied by powers (possibly 0) of x, y, and z and a constant for normalization, so that:

$$\int_{\text{allspace}} g^2 = 1 \tag{59}$$

Thus, c depends on α, l, m, and n.

Here are three representative Gaussian functional (s, p_y and d_{xy} types, respectively):

$$g_s(\alpha, \vec{r}) = \left(\frac{2\alpha}{\pi}\right)^{3/4} e^{-\alpha r^2}$$

$$g_y(\alpha, \vec{r}) = \left(\frac{128\alpha^5}{\pi^3}\right)^{1/4} y e^{-\alpha r^2}$$

$$g_{xy}(\alpha, \vec{r}) = \left(\frac{2048\,\alpha^7}{\pi^3}\right)^{1/4} xy e^{-\alpha r^2} \tag{60}$$

Linear combinations of primitive gaussians like these are used to form the actual basis functions; the latter are called contracted Gaussians and have the form

$$\chi_\mu = \sum_P d_{\mu p} g_p \tag{61}$$

where the $d_{\mu p}$'s are fixed constants within a given basis set. Note that contracted functions are also normalized in common practice. A few commonly used basis sets are lists as following.

Minimal Basis Sets: Minimal basis sets contain the minimum number of basis functions needed for each atom, as in these examples:

H: 1s

C: 1s, 2s, $2p_x$, $2p_y$, $2p_z$

Minimal basis sets use fixed-size atomic-type orbitals. The STO-3G basis set is a minimal basis set (although it is not the smallest possible basis set). It used three gaussian primitives per basis function, which accounts for the "3G" in its name. "STO" stands for "Slater-type orbitals," and the STO-3G basis set approximates Slater orbitals with gaussian functions.

Split Valence Basis Sets
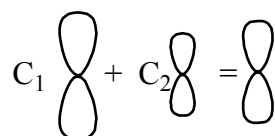
The first way that a basis set can be made larger is to increase the number of basis functions per atom. Split valence basis sets, such as 3-21G and 6-31G, have two (or more) sized of basis function for each valence orbital. For example, hydrogen and carbon are represented as:

H: 1s, 1s'

C: 1s, 2s, 2s', $2p_x$, $2p_y$, $2p_z$, $2p'_x$, $2p'_y$, $2p'_z$

where the primed and unprimed otbitals differ in size.

The double zeta basis sets, such as the Dunning-Huzinaga basis set (D95), form all molecular orbitals from linear combinations of two sized of functions for each atomic orbital. Similarly, triple split valence basis sets, like 6-311G, use three sizes of contracted functions for each orbital-type.

Polarized Basis Sets
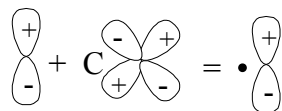
Split valence basis sets allow orbitals to change size, but not to change shape. Polarized basis sets remove this limitation by adding orbitals with angular momentum beyond what is required for the ground state to the description of each atom. For example, polarized basis sets add d functions to carbon atoms and f functions to transition metals, and some of them add p functions to hydrogen atoms.

So far, the only polarized basis set 6-31G(d) is used. Its name indicates that it is the 6-31G basis set with d functions added to heavy atoms. This basis set is becoming very common for calculations involving up to medium-sized systems. This basis set is also known as 6-31G*. Another popular polarized basis set is 6-31G(d,p), also known as 6-31G**, which adds p functions to hydrogen atoms in addition to the d functions on heavy atoms.

Diffuse Functions
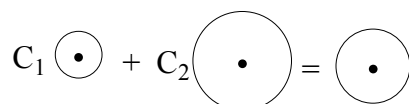
Diffuse functions are large-size versions of s- and p- type functions (as opposed to the standard valence-size functions) which allow orbitals to occupy a larger region of space. Basis sets with diffuse functions are important for systems where electrons are relatively far from the nucleus: molecules with lone pairs, anions and other systems with significant negative charge, systems in their excited states, systems with low ionization potentials, descriptions of absolute acidities. The 6-31+G(d) basis set is the 6-31G(d) basis set with diffuse functions added to heavy atoms. The double plus version, 6-31++G(d), adds diffuse functions to the hydrogen

atoms as well. Diffuse functions on hydrogen atoms seldom make a significant difference in accuracy.

High Angular Momentum Basis Sets

Even larger basis sets are now practical for many systems. Such basis sets add multiple polarization functions per atom to triple zeta basis set. For example, the 6-31G(2d) basis set adds two d functions per heavy atom instead of just one, while the 6-311++G(3df,3pd) basis set contains three sets of valence region functions, diffuse functions on both heavy atoms and hydrogens, and multiple polarization funtions:3 d functions and 1 f function on heavy atoms and 3 p functions and 1 d function on hydrogen atoms. Such basis sets are useful for describing the interactions between electrons in electron correlation methods.

APPENDIX B

Poster Contributions to Conferences and Publication

Poster Presentation

Hongkrengkai, R., M. Kuno, and S. Hannongbua, **The Applicability of the ONIOM-BSSE scheme of weak interation for H···π systems**. The 2nd Asian Pacific Conference on Theoretial & Computational Chemistry (APCTCC 2005), Chulalongkorn University, Bangkok, Thailand, 2-6 May 2005.

The Applicability of the ONIOM-BSSE scheme of weak interaction for H···π systems

Rattapon Hongkrengkai[a], Mayuso Kuno [a,*], Supa Hannongbua[b]
[a] Department of Chemistry, Faculty of Science, Srinakharinwirot University, Sukhumvit 23, Bangkok, 10110 Thailand
[b] Department of Chemistry, Faculty of Science, Kasetsart University, Jatujak, Bangkok, 10900 Thailand
Phone: (662)664-1000 ext. 8461, Fax: (662)259-2097, E-mail: mayuso@swu.ac.th

Introduction

H···π systems are interesting both the experimental and theoretical studies because they can be explain the conformational preferences in biological systems, crystal packing, host guest complexation. In this work, we investigated the H···π interaction of alcohol and ethylene systems. On a different note, the basis set superposition error (BSSE) not only effects the magnitude of the interaction energy, but also influences the intermolecular geometrical parameters. Therefore, the optimization with and without counterpoise (CP) minimization were taken into account.

Methods

The geometry optimization for all complexes were performed with C_s symmetry constrains based on the B3LYP, MP2 and ONIOM2 methods with several basis set using Gaussian 03 package. The optimization procedure with and without CP method was used at the MP2 and B3LYP levels of theory. The BSSE corrected binding energies for all calculations were computed using the CP method of Boys and Bernardi scheme and also includes the relaxation effects.

$$BE_{ONIOM}^{CP} = BE_{HL}^{CP}(m1m2) + BE_{LL}^{CP}(M1M2) - BE_{LL}^{CP}(m1m2)$$
$$+ \Delta E_{HL}^{relax}(m1) + \Delta E_{LL}^{relax}(M1) - \Delta E_{LL}^{relax}(m1)$$
$$+ \Delta E_{HL}^{relax}(m2) + \Delta E_{LL}^{relax}(M2) - \Delta E_{LL}^{relax}(m2)$$

$$BE^{CP}(AB) = BE^{CP}(AB^{AB}) - E^{AB}(A^{AB}) - E^{AB}(B^{AB})$$

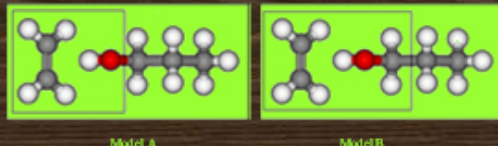$$\Delta E^{relax} = E^{A}(A^{AB}) - E^{A}(A^{A})$$

Model A   Model B

Figure The molecular system partitioned for ONIOM2 calculation

Conclusions

We have carried out MP2, B3LYP and ONIOM2 methods with 6-311+G** and 6-31G* basis sets to investigate the structure and the binding energy between alcohol and ethylene complexes. It was found that the basis sets has a smaller effected to binding energy at the MP2(full) level of calculation. The BSSE corrected binding energy with CP corrected geometries are stable than without CP corrected geometries about 0.1 kcal/mol. The ONIOM results derived in this study suggest that the ONIOM-BSSE scheme provides reasonable results for investigating the H···π systems.

Results

Table 1: The binding energies (kcal/mol) and selected bond distances (Å) for ethylene and alcohol complexes obtained from quantum chemical calculations with and without counterpoise (CP) minimization.

Method	MP2(full)		B3LYP	
System	6-311+G**	6-31G*	6-311+G**	6-31G*
MeOH–C₂H₄				
	-3.18	-3.72	-1.91	-3.10
	-1.97	-1.88	-1.11	-2.03
	-3.10	-3.51	-1.92	-3.07
	-2.08	-2.12	-1.47	-2.07
	2.427	2.382	2.503	2.417
	2.517	2.577	2.530	2.523
EtOH–C₂H₄				
	-3.35	-3.77	-1.80	-3.04
	-2.00	-1.84	-1.55	-1.95
	-3.24	-3.54	-1.82	-3.01
	-2.11	-2.12	-1.57	-2.01
	2.433	2.380	2.530	2.424
	2.548	2.578	2.52	2.527
PrOH–C₂H₄				
	-3.40	-3.80	-1.79	-3.05
	-2.04	-1.88	-1.54	-1.97
	-3.31	-3.57	-1.80	-3.02
	-2.14	-2.14	-1.58	-2.01
	2.431	2.380	2.521	2.421
	2.544	2.574	2.547	2.525

Table 2: The binding energies (kcal/mol) and selected bond distances (Å) for ethylene and alcohol complexes obtained from ONIOM2 calculations.

	MP2(full)/6-311+G**: MP2(full)/6-31G*	MP2(full)/6-311+G**: B3LYP/6-311+G**	MP2(full)/6-311+G**: B3LYP/6-31G*
Model A			
MeOH–C₂H₄			
	-3.14	-2.72	-2.84
	-2.01	-1.72	-1.87
	2.379	2.434	2.433
EtOH–C₂H₄			
	-3.18	-2.40	-2.79
	-2.00	-1.43	-1.83
	2.373	2.473	2.441
Model B			
EtOH–C₂H₄			
	-3.33	-3.02	-3.11
	-1.95	-1.88	-1.91
	2.429	2.440	2.433
PrOH–C₂H₄			
	-3.37	-3.01	-3.10
	-1.97	-1.91	-1.99
	2.424	2.433	2.409

Acknowledgements

References

-Tschumper, G.S., Morokuma, K., Gauging the applicability of ONIOM (MO/MO) methods to weak chemical interactions in large systems: hydrogen bonding in alcohol dimers., **J. Mol. Struct. Theochem.**, 2002, 592, 137-147

-Tarakeshwar, P., Kwang S.K., Comparison of the nature of π and conventional H-bonds: a theoretical investigation., **J. Mol. Struct. Theochem.**, 2002, 615, 227-238

Publication

ONIOM-BSSE scheme for H⋯π system and application on HIV-1 reverse transcriptase

Mayuso KUNO*, Rattapon Hongkrengkai, Supa Hannongbua

Date: Apr 05, 2006
To: "mayuso kuno" mayuso.kuno@gmail.com
From: "CPLETT" cplett@elsevier.com
Subject: CPLETT-06-430R1: Final Decision

Ms. No.: CPLETT-06-430R1
Title: ONIOM-BSSE scheme for H...¶ system and applications on HIV-1 reverse transcriptase
Corresponding Author: Dr. mayuso kuno
Authors: Rattapon Hongkrengkai; Supa Hannongbua, Dr.rer.nat

Dear Dr. kuno,

We are pleased to inform you that your manuscript referenced above has been accepted for publication in Chemical Physics Letters.

Your accepted paper is considered to be a significant contribution to the field, and we appreciate the opportunity to publish it in this journal. Thank you for submitting your work to Chemical Physics Letters, and we hope that you will continue to consider this journal as a primary source for publishing your future research.

With kind regards,

David C. Clary
Editor
Chemical Physics Letters

Chemical Physics Letters, Editorial Office
E-mail: cplett@elsevier.com

1

ONIOM-BSSE scheme for H···π system

and applications on HIV-1 reverse transcriptase

Mayuso Kuno [a*], Rattapon Hongkrengkai[b], Supa Hannongbua[b]

[a] Department of Chemistry, Faculty of Science, Srinakharinwirot University,

Sukhumvit 23, Bangkok, 10110, THAILAND

[b] Department of Chemistry, Faculty of Science, Kasetsart University, Cha-tu-chak,

Bangkok, 10900, THAILAND

Abstract

Intermolecular interactions between ethanol and ethylene forming H···π complex systems were investigated using B3LYP, MP2 and ONIOM methods with a 6-31G(d,p) basis set. All binding energies were corrected using the counterpoise method of Boys-Bernardi approach. The ONIOM-BSSE scheme was used for the binding energy calculations on the H···π systems of ethanol-ethylene and HIV-1 RT/nevirapine complexes. The ONIOM results derived from this study suggest that the ONIOM-BSSE scheme provides reasonable results for investigating the H···π systems.

* To whom correspondence should be addressed.

Tel. +662-6641000 ext 8461, Fax +662-2592097, e-mail: mayuso@swu.ac.th.

2

1. Introduction

Understanding attractive interactions involving π systems is one of the major goals of research into weak interactions [1]. The natures of these interactions have been investigated in both experimental and theoretical studies [2]. In particular, the interactions between alcohol and ethylene as prototypes for the H⋯π system have interested many fields of chemistry ranging from molecular biology to material science. The H⋯π bonded interaction can influence the three-dimensional structure of enzymatic reactions which are important for the molecular recognition process in biological systems such as drug receptor interactions [3-5].

In theoretical investigations of an enzyme inhibitor interaction, the number of atoms in the molecular system makes it impossible to study with great accuracy. Such study is also limited by the computational effort required. Recently, the developments of multilayer integration methods in computational chemistry have made investigation of a large molecular system more feasible [6]. Now, the ONIOM method has been introduced and its efficiency has been improved over the years [7-9]. In the ONIOM approach, a large molecular system can be partitioned into multilayer regions which are treated using different theoretical methods. The inner layer or the region, where the reaction occurs, uses a high-level *ab initio* or DFT (Density Functional Theory) method [10-12].

In this study, in order to understand the H⋯π system, the interaction between ethanol and ethylene was used as a model to examine the nature of weak interactions. The standard and counterpoise (CP) corrected gradient optimizations of this complex were applied [13-15]. The ethanol-ethylene complex was divided for the application of the ONIOM calculations and proved the applicability of the ONIOM-BSSE scheme. Consequently, the ONIOM method was applied in the study of the structure

3

and the binding energy of nevirapine at the binding site of HIV-1 reverse transcriptase (RT). In particular, this study on focused for improving the binding energy calculations of this complex system.

2. Model and methods

2.1 Ethanol-ethylene complex

In order to investigate the H···π interaction following the B3LYP and MP2 methods with a 6-31G(d,p) basis set were applied to obtain the intermolecular interaction between ethanol and ethylene in the complex. During optimization, the ethanol and ethylene were kept at C_s and D_{2h} symmetries, respectively, while the complex was kept at C_s symmetry. The standard and CP-corrected gradient optimizations of the intermolecular interaction were observed. In order to evaluate the intermolecular interaction, we also computed the basis set superposition error (BSSE) based on the counterpoise scheme of Boys-Bernardi [16] using the following equation Eq. (1):

$$E_{int}^{CP}(AB) = E(AB^{AB}) - E(A^{AB}) - E(B^{AB}) \qquad (1)$$

where $E(AB^{AB})$ is the energy of the complex AB with the basis set of AB, while $E(A^{AB})$ and $E(B^{AB})$ are the energies of monomer A and monomer B, respectively, with the basis set of AB at their respective geometries adopted from the complex of AB.

To reduce the computational demand in the next application of HIV-1 RT binding site/nevirapine complex, two-layered ONIOM2 (MP2/6-31G(d,p):B3LYP/ 6-31G(d,p)) method was applied to examine the interaction between ethanol and ethylene in the complex with the same C_s symmetry configuration. There are two possible H···π systems for ethanol-ethylene complex as shown in Fig. 1. The systems were divided into two parts. The inner layer or H···π interaction (region A) was

4

treated at the MP2 method, while the outer layer (region B) was treated at the B3LYP method. In order to estimate the applicability of the ONIOM-BSSE corrected binding energy in the ONIOM model, the counterpoise correction scheme is expressed as Eq. (2):

$$BE_{ONIOM}^{CP} = BE_{HL}^{CP}(m1m2) + BE_{LL}^{CP}(M1M2) - BE_{LL}^{CP}(m1m2) \qquad (2)$$

where CP, HL, LL, m and M denote counterpoise correction, high level of calculation, low level of calculation, small model in inner layer and real model in outer layer, respectively. Numbers 1 and 2 stands for the part of nevirapine and the part of an amino acid studied as defined in the ONIOM model system, respectively.

-- Insert Fig 1 --

2.2 HIV-1 RT binding site/nevirapine complex

The starting model of the HIV-1 RT binding site/nevirapine complex in this study was obtained from the 2.2 Å resolved crystal structure (1VRT.pdb) [17]. We adopted the system consisting of 22 residues surrounding the non-nucleoside inhibitor binding pocket (NNIBP) with at least one atom interacting with any of the atoms of nevirapine structure within a 7 Å diameter centered at nevirapine. All residues, assumed to be in their neutral form, were terminated, if not connected to another residue in the selected model by a link with an acetyl group (CH_3CO-) and a methyl group (-$NHCH_3$) at the N- and C-terminal ends of the cut residues, respectively, from the adjacent residues as presented in the backbone geometries of the X-ray structure. Hydrogen atoms were added to the geometrical structure to generate the complete structure of the model system and their positions were optimized by the semi-empirical PM3 method. This structure was used as the starting geometry for all

5

calculations. Therefore, the ONIOM3 (MP2/6-31G(d,p):B3LYP/6-31G(d,p):PM3) method was applied on the partitioned model system as shown in Fig. 2.

--Insert Fig 2--

Optimizations were performed taking into account two approximations, heavy atoms fixing (HAF) and backbone atoms fixing (BBF). In HAF, all heavy atoms of the amino acids in the binding pocket were fixed at the X-ray structure and, therefore, only the geometry and the position of the nevirapine inhibitor and all the hydrogen atoms were optimized. In BBF, only the backbone atoms of the amino acids were fixed. In order to evaluate the applicability of the ONIOM-BSSE corrected binding energy in the ONIOM model of HVI-1 RT, the counterpoise correction was computed following Eq. (2). All calculations were carried out using the GAUSSIAN 03 package [18] running on Linux PC 3.2 GHz.

3. Results and discussion
3.1 *Structure and binding energy of ethanol-ethylene complex*

The calculated results obtained from the MP2 and B3LYP methods are shown in Table 1. Considering the effect of the CP-corrected gradient optimization on the ethanol-ethylene complex, we can see that this method affects the calculated intermolecular distances H5···π and O1···π. The CP-corrected optimized geometries produced a longer H···π distance when compared with the standard optimization. The results were caused by a nonphysical attraction between the two fragments introduced by the BSSE correction. The binding energy differences between the two optimization procedures are less than 0.22 and 0.06 kcal/mol when using the MP2

6

and B3LYP methods, respectively. This effect is almost identical to the uncorrected and CP-corrected binding energies.

--Insert Table 1--

When comparing the model I and the model II of ethanol-ethylene complex, only small differences were found in terms of the geometrical parameters and the binding energies. Furthermore, the CP-corrected binding energy after optimization showed similar results. We found that the binding energy differences under the MP2 method on the model I are 1.44 and 1.20 kcal/mol and on the model II are 1.70 and 1.28 kcal/mol by standard and CP-corrected optimizations, respectively. Using B3LYP method, the corresponding energy differences are 0.91 and 0.85 kcal/mol on the model I and 1.05 and 0.93 kcal/mol on the model II, by the same optimizations. Thus, these indicate that the CP-corrected binding energy is less sensitive to the CP-corrected geometry than the standard geometry. This should clearly be true as the binding energy obtained from the CP-corrected optimization must be higher than that at the standard optimized minimum [14].

Considering the results obtained from the ONIOM model, the ONIOM results of the model I and the model II are shown in Table 1. It can be seen that the H⋯π interaction shows little sensitivity to the choice of high-level ONIOM model. Interestingly, the MP2 calculation of the H⋯π interaction and also the binding energy (both uncorrected and corrected binding energies) in the high-level of ONIOM method is acceptable for model I and model II of the ethanol-ethylene complex. Therefore, the ONIOM method was applied on HIV-1 RT binding site/nevirapine

7

complex in the next section and the binding energy correction was also computed using the ONIOM-BSSE scheme following Eq. (2).

3.2 *Structure and binding energy of HIV-1 RT binding site/nevirapine complex*

The main focus of our study is to investigate the interaction between nevirapine and the binding site of HIV-1 RT. However, this complex system is too large for high level (MP2) calculations. First, the interaction energies between nevirapine and individual residues were investigated at the MP2/6-31G(d,p) level of calculation. These interactions are given in Table 2. It is clearly seen that more attractive interactions are found. Interaction energies of Leu100, Ty181 and Tyr188 are main contributors as obtained by both HAF and BBF optimized procedures. Therefore, these results were considering the inner model layer of the ONIOM calculation. Furthermore, the interaction between nevirapine inhibitor and Tyr181 form H⋯π interaction, thus, this interaction were treated at the MP2 method. Next, we divided the model system into three parts and applied the three-layer ONIOM3 (MP2/6-31G(d,p):B3LYP/6-31G(d,p):PM3) method as already described and the binding energies were corrected based on the ONIOM-BSSE scheme.

Table 3 shows the selected optimized intermolecular distances and angles between heavy atoms in the interacting core region treated at MP2/6-31G(d,p) level, as well as the binding energy and its component of nevirapine bound to the binding site of HIV-1 RT. The optimization was performed using two approaches. It was found that the binding energies are -14.83 and -15.93 kcal/mol for HAF and BBF, respectively. Including the ONIOM-BSSE scheme, the binding energies were reduced to -8.79 and -10.45 kcal/mol for HAF and BBF, respectively. The BBF optimization produces an increase in binding energies from 1.10 to 1.66 kcal/mol. It was found that

8

the binding energy difference between the two approaches comes from $\Delta\Delta E$(Low, ABC-AB) of about 1.50 kcal/mol. These results imply that the relaxation of residues during optimization has less effect on the binding energy obtained from the medium-level and high-level calculations.

--Insert Table 2--

The HAF approach, previously using 16 residues [7] and presently using 22 residues represented in the binding pocket, yields longer inter-fragment distances than the experimental results obtained from the X-ray crystallographic data [17]. Surprisingly, the BBF optimized geometries produced shorter inter-fragment distances for N1-C7, N1-C8, C2-C9 and C3-C8 than the experimental results and effected to the intermolecular angle (C2-C3-C11) of about 10 degrees due to the movement of nevirapine and Tyr181. The H3 attached to C3 of nevirapine points toward the inner region of the tyrosine ring close to the C8 of Tyr181 (Fig. 3) and the distances between C2-C8 and C2-C9 are 3.25 and 3.56, respectively. Therefore, it is important to note that the hydrogen bonding between the nitrogen atom of nevirapine and the hydrogen atom of C beta (C7) in Tyr181 play an important role in the binding pocket of the HIV-1 RT/nevirapine complex and may also help to form a facial H··π interaction via the H3 of the pyridine ring with Tyr181.

--Insert Fig 3--

9

4. Conclusions

This work was performed the ONIOM-BSSE scheme in order to investigate the geometry and the binding energy of the ethanol and ethylene formed H···π system complex and the nevirapine bound to the HIV-1 RT binding pocket complex. The obtained results indicate that the counterpoise corrected gradient optimization affects the inter-molecular interaction compared to the standard optimization and the former yielded the longer inter-molecular fragments than the latter. In comparison of the non-ONIOM and ONIOM methods, it was found that the binding energies from both full and frozen core electron correlations at the MP2 level of calculation after BSSE correction are similar. The complex structure of the HIV-1 RT/nevirapine inhibitor suggests that the nitrogen (N1) of the pyridine ring forms hydrogen bonds and produces the facial H···π interaction via nevirapine and Tyr181. The analysis of the ONIOM-BSSE binding energy shows that the binding energy in the interacting core region (small model) is very weak at about -1.4 kcal/mol and the substantial binding comes almost exclusively from the interaction of nevirapine with other residues in the binding pocket.

10

References

[1]   K. S. Kim, P. Tarakeshwar, J. Y. Lee, Chem. Rev. 100 (2000) 4145.

[2]   K. Müller-Dethlefs, P. Hobza, Chem. Rev. 100 (2000) 143.

[3]   K. Kowski, W. Lüttke, P. Rademacher, J. Mol. Struct. 567-568 (2001) 231.

[4]   P. Tarakeshwar, K.S. Kim, J. Mol. Struct. 615 (2002) 227.

[5]   I. Bandyopadhyay, H.M. Lee, K.S. Kim, J. Phys. Chem. A. 109 (2005) 1720.

[6]   X. He, Y. Mei, Y. Xiang, D.W. Zhang, J.Z.H. Zhang, Prot.: Struct. Func. Bioinf.
      61 (2005) 423.

 [7]   M. Kuno, S. Hannongbua, K. Morokuma. Chem. Phys. Lett. 380 (2003) 456.

[8]   P. Nunrium, M. Kuno, S. Saen-oon, S. Hannongbua, Chem. Phys. Lett. 405
      (2005) 198.

[9]   S. Saen-oon, M. Kuno, S. Hannongbua, Prot.: Struct. Func. Bioinf. 61 (2005)
      859.

[10] F. Maseras, K. Morokuma, J. Comput. Chem. 16 (1995) 1170.

[11] M. Svensson, S. Humbel, R.D.J. Froese, T. Matsubara, S. Sieber, K. Morokuma,
      J. Phys. Chem. 100 (1996) 19357.

[12] S. Dapprich, I. Komaromi, K.S. Byun, K. Morokuma, M.J. Frisch, J. Mol.
      Struct. (Theochem) 461 (1999) 1.

[13] W. Zheng, W. Wang, X. Pu, A. Tian, N.B. Wong, J. Mol. Struct. (Theochem)
      631 (2003) 171.

[14] L. Pejov, Chem. Phys. 285 (2002) 177.

[15] F.F. Muguet, G.W. Robinson, J. Chem. Phys. 102 (1995) 3648.

[16] S.F. Boys, F. Bernardi, Mol. Phys. 19 (1970) 553.

[17] J. Ren, R. Esnouf, E. Garman, D. Somers, C. Ross, I. Kirby, J. Keeling, G.
      Darby, Y. Jones, D. Stuart, D. Stammers, Struct. Biol. 2 (1995) 293.

[18] M.J. Frisch et al., Gaussian03, Gaussian, Inc, Pittsburgh, PA, 2003.

11

TABLE CAPTION

Table 1 Selected bond distances (Å), angle (degree) and binding energy (ΔE, kcal/mol) of the ethanol-ethylene complex based on Model I and Model II, optimized using different methods with 6-31G(d,p) basis set.

Table 2 Interaction energies (IE, kcal/mol) of nevirapine with individual residues, calculated at the MP2/6-31G(d,p) level.

Table 3 Selected optimized inter-fragment distances (Å), binding energies (ΔE, kcal/mol) and its components for the HIV-1 RT/nevirapine complex, optimized at the ONIOM(MP2/6-31G(d,p):B3LYP/6-31G(d,p):PM3) method.

12

FIGURE CAPTION

Fig. 1. The MP2/6-31G(d,p) optimized structures for model I and model II of ethanol-ethylene complex forming H⋯π system. Layer partitioning is shown for ONIOM2 (A is the inner layer and B is the outer layer calculations.

Fig. 2. Adopted model system of nevirapine bound to the HIV-1 RT binding site. Layer partitioning is shown for ONIOM3 (A is the inner layer, B is the medium layer and C is the outer layer calculations.

Fig. 3. Optimized structure of the nevirapine and Tyr181 complex from ONIOM(MP2/6-31G(d,p):B3LYP/6-31G(d,p):PM3) based on BBF with selected atomic labels, bond distances and bond angle.

Table 1

	Methods	r(H5···π)	r(O1···π)	r(O1-H5)	r(C1=C2)	$\angle$(O1-H5-π)	$\angle$(C3-O1-π)	ΔE-Uncorr.[a]	ΔE-Corr.[b]
Model I									
Standard[c]	MP2	2.45	3.40	0.97	1.34	169.3	99.4	-3.45	-2.01
	B3LYP	2.48	3.45	0.97	1.33	179.7	108.3	-2.80	-1.89
	ONIOM2M[e]	2.45	3.41	0.97	1.34	170.6	100.3	-3.34	-1.93
CP-corr.[d]	MP2	2.57	3.54	0.97	1.34	175.7	104.2	-3.32	-2.12
	B3LYP	2.55	3.51	0.97	1.33	174.2	112.4	-2.77	-1.92
Model II									
Standard[c]	MP2	2.40	3.35	0.97	1.34	168.1	98.5	-3.57	-1.88
	B3LYP	2.43	3.39	0.97	1.33	177.3	110.0	-2.96	-1.91
	ONIOM2M[e]	2.40	3.36	0.97	1.34	170.6	100.3	-3.46	-1.82
CP-corr.[d]	MP2	2.57	3.53	0.97	1.34	177.0	105.2	-3.37	-2.09
	B3LYP	2.52	3.49	0.97	1.33	172.5	113.6	-2.91	-1.97

[a]Non-BSSE corrected binding energy. [b]BSSE Corrected binding energy. [c]Standard gradient optimization. [d]counterpoise corrected gradient optimizations. [e]Optimized at ONIOM(MP2/6-31G(d,p):B3LYP/6-31G(d,p)) level of calculation.

14

Table 2

Residue	Interaction energy (IE, kcal/mol)			
	HAF		BBF	
	IE-Uncorr.[a]	IE-Corr.[b]	IE-Uncorr.[a]	IE-Corr.[b]
PRO095	-1.86	-1.10	-1.39	-0.87
LEU100	**-6.08**	**-2.37**	**-6.58**	**-3.09**
LYS101	-2.55	-1.46	-2.29	-1.10
LYS102	-0.44	-0.25	-0.58	-0.37
LYS103	-2.38	-1.36	-2.04	-1.20
LYS104	0.00	0.00	0.02	0.02
SER105	-0.09	-0.09	-0.16	-0.16
VAL106	-4.07	-1.70	-2.60	-0.08
VAL179	0.05	1.44	0.41	1.47
ILE180	-1.16	-0.70	-0.73	-0.50
TYR181	**-5.81**	**-2.67**	**-5.14**	**-2.79**
TYR188	**-7.63**	**-4.52**	**-8.87**	**-5.05**
VAL189	-0.97	-0.62	-1.14	-0.64
GLY190	-1.61	-0.39	0.07	1.65
PHE227	-2.71	-1.73	-2.67	-1.74
LEU228	-0.19	-0.19	-0.20	-0.20
TRP229	-4.61	-2.23	-3.02	-1.19
LEU234	-0.95	-0.92	-1.21	1.22
HIS235	-3.80	-2.31	-4.22	-2.39
PRO236	-3.14	-1.56	-2.83	-1.63
TYR318	-2.99	-1.38	-3.28	-1.80
GLU138[c]	-0.20	-0.08	-0.15	-0.09

[a] Non-BSSE corrected interaction energy.

[b] BSSE Corrected interaction energy.

[c] All residues in this study taken from the p66 domain of RT, excepted GLU138 taken from the p51 domain of RT.

15

Table 3.

	Experiment[a]	HAF[b]	HAF[c]	BBF
Bond distances (Å)				
r(N1-C7)	3.67	3.89	3.76	3.56
r(N1-C8)	3.93	4.14	3.94	3.88
r(C2-C8)	3.22	-	3.39	3.25
r(C2-C9)	3.78	3.81	3.79	3.56
r(C2-C13)	3.33	-	3.55	3.70
r(C3-C8)	3.52	3.83	3.72	3.42
r(C3-C12)	3.39	4.10	3.57	3.63
r(C3-C13)	3.04	3.57	3.34	3.33
Bond angles (degree)				
∠(C2-C3-C11)	94.6	-	93.0	104.7
∠(C-C7-C8)	113.2	-	114.8	112.9
Binding energy (kcal/mol)				
ΔE	-	-8.85	-14.83 (-8.79)[d]	-15.93 (-10.45)[d]
ΔE (High, A)	-	-1.05	-3.62 (-1.43)[d]	-3.82 (-1.45)[d]
ΔΔE (Mid, AB-A)	-	1.21	-2.02 (1.83)[d]	-1.42 (1.69)[d]
ΔΔE (Low, ABC-AB)	-	-9.01	-9.19	-10.69

[a] The experimental results obtained from the X-ray crystallographic data in ref. [17].

[b] Heavy atoms fixing taken from ref. [7] using 16 residues.

[c] Heavy atoms fixing in this study using 22 residues and see text for definition.

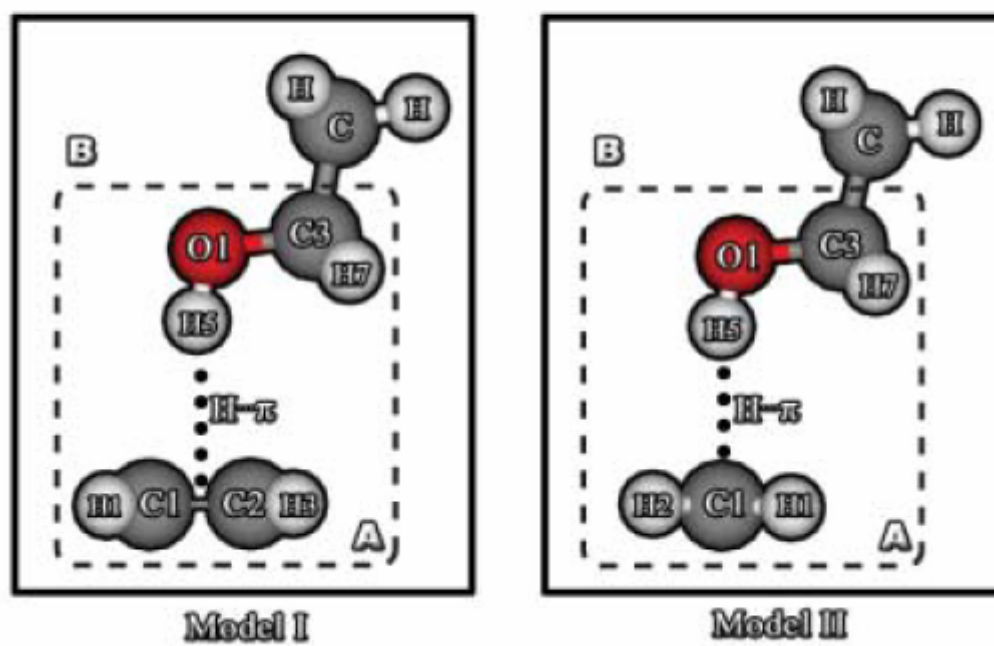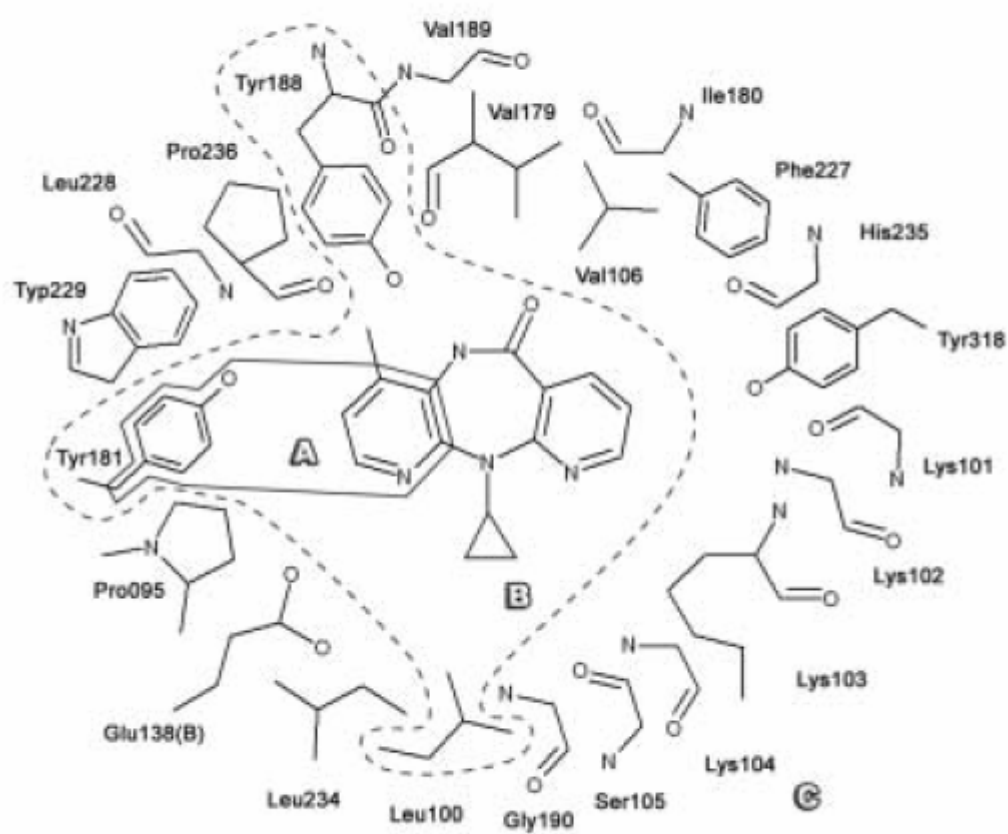[d] BSSE Corrected binding energy in parenthesis.

Fig. 1.

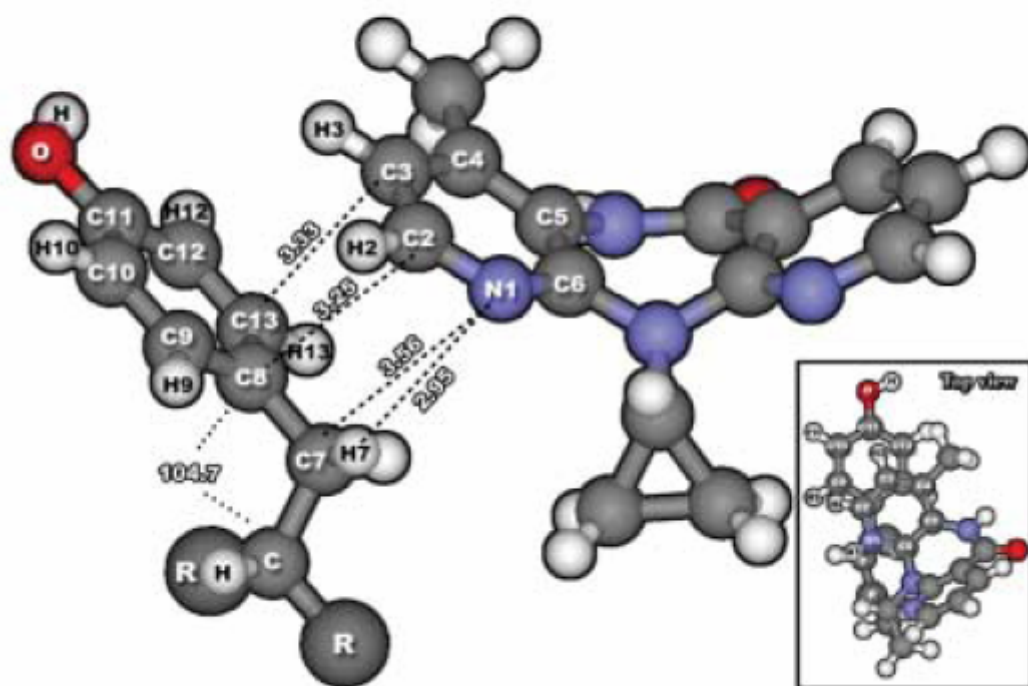17

Fig. 2.

18

Fig. 3.

CURRICULUM VITAE

NAME	:	Mr. Rattapon Hongkrenkai		
BIRTH DATE	:	August 25, 1979		
BIRTH PLACE	:	Bangkok, Thailand		
NATIONALITY	:	Thai		
EDUCATION	:	**YEAR**	**INSTITUTION**	**DEGREE/DIPLOMA**
	:	2002	Kasetsart Univ.	B. Sc. (General Science)
	:	2005	Kasetsart Univ.	M. Sc. (Physical Chemistry)

SCHOLARSHIP : Postgraduate Education and Research Program in Petroleum and Petrochemical Technology (MUA-ADB) 2002-2003

PRESENTATION :

Hongkrengkai, R., M. Kuno, and S. Hannongbua, **The Applicability of the ONIOM-BSSE scheme of weak interation for H···π systems**. The 2nd Asian Pacific Conference on Theoretial & Computational Chemistry (APCTCC 2005), Chulalongkorn University, Bangkok, Thailand, 2-6 May 2005.