Computational Modelling of Intermolecular Interactions in bio, organic and nano molecules

A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy in the Faculty of Engineering and Physical Sciences

2010

Anitha Ramraj

School of Chemistry

Contents

List of Tables

List of Figures

Abbreviations

AM1	Austin Model One
AMBER	Assisted Model Building with Energy Refinement
AO	Atomic orbital
B3LYP	Becke-3, Lee, Yang, Parr Functional
B86	Becke (1986) exchange functional
I.E.	Interaction Energy
B.E.	Binding Energy
BSSE	Basis Set Superposition Error
CBS	Complete Basis Set Limit
CCSD(T)	Coupled-Cluster with Single and Double and Triple excitations
CNDO	Complete neglect of differential overlap
DFT	Density Functional Theory
DFT-D	Dispersion Corrected Density Functional Theory
Disp	Dispersion Contribution
GAFF	General AMBER Force Field
GB	Generalized Born Solvation Model
GGA	Generalized Gradient Approximation
GIAO	Gauge Including Atomic Orbital
Glu	Glucose
GTO	Gaussian Type Orbital
HF	Hartree-Fock
HGGA	Hybrid Generalized Gradient Approximation
HMGGA	Hybrid meta-Generalized Gradient Approximation
I.F.	Imaginary Frequencies
INDO	Intermediate neglect of differential overlap
IR	Infrared
KS	Kohn Sham
LCAO	Linear Combination of Atomic Orbitals
LSDA	Local Spin density approximation
LYP	Lee-Yang-Parr correlation functional

MD	Molecular Dynamics
MGGA	Meta-Generalized Gradient Approximation
MM	Molecular Mechanics
MM-GBSA	Molecular Mechanics-Generalized Born Surface Area
MNDO	Modified neglect of differential overlap
MO	Molecular Orbital
MP2	Moller-Plesset Second Order Perturbation Theory
MUE	Mean Unsigned Error
MUSE	Manchester University Semi-Empirical Program
NDDO	Neglect of diatomic differential overlap
NMR	Nuclear magnetic resonance
NPT	Constant Particle Number, Constant Pressure and Constant Temperature
NVT	Constant Particle Number, Constant Volume and Constant Temperature
PBE	Perdew, Burke, and Enzerhof GGA exchange-correlation functional
PDB	Protein Data Bank
PM3	Parameterized Model 3
PM3-D	Dispersion Corrected Parameterized Model 3
PM3-D*	Modified Dispersion Corrected Parameterized Model 3
PME	Particle-Mesh Ewald method
QM	Quantum Mechanics
RHF	Restricted Hartree-Fock
SCF	Self Consistent Field
STO	Slater Type Orbital
TZV (2d, 2p)	Triple zeta Valence polarization basis set
VDW	van der Waals

Abstract

ABSTRACT OF THESIS submitted by ***Anitha Ramraj*** for the Degree of Doctor of Philosophy, entitled "Computational Modelling of Intermolecular interactions in bio, organic and nano molecules".
Date of Submission: 20/09/2010

We have investigated the noncovalent interactions in carbohydrate-aromatic interactions which are pivotal to the recognition of carbohydrates in proteins. We have employed quantum mechanical methods to study carbohydrate-aromatic complexes. Due to the importance of dispersion contribution to the interaction energy, we mainly use density functional theory augmented with an empirical correction for the dispersion interactions (DFT-D). We have validated this method with a limited number of high level *ab initio* calculations. We have also analysed the vibrational and NMR chemical shift characteristics using the DFT-D method. We have mainly studied the complexes involving β-glucose with 3-methylindole and *p*-hydroxytoluene, which are analogues of tryptophan and tyrosine, respectively. We find that the contribution for interaction energy mainly comes from CH/π and OH/π interactions. We find that the interaction energy of complexes involving CH/π and OH/π interactions is reflected in the associated blue and red shifts of vibrational spectrum. We also find that the interactions involving 3-methylindole are somewhat greater than those for *p*-hydroxytoluene. The C-H blue shifts are also in parallel with the predicted NMR proton shift. We have also tested different density functionals including both standard density functionals and newly developed M0x functionals and MP2 method for studying carbohydrate-aromatic complexes. The DFT-D method and M06 functionals of the M0x family are found to perform better, while B3LYP and BLYP functionals perform poorly. We find that the inclusion of a dispersion term to BLYP is found to perform better. The dispersion energy dominates over the interaction energy of carbohydrate-aromatic complexes. From the DFT-D calculations, we found that the complexes would be unstable without the contribution from dispersive energy. We have also studied the importance of noncovalent interactions in functionalization of nanotubes by nucleic acid bases and aromatic aminoacids by using semi-empirical methods with dispersion term such as PM3-D and PM3-D*. We find that the both semi-empirical schemes give reasonable interaction energies with respect to DFT-D interaction energies. We have also used PM3-D method to study the adsorption of organic pollutants on graphene sheet and on nanotubes. We found that the semi-empirical schemes, which are faster and cheaper, are suitable to study these larger molecules involving noncovalent interactions and can be used as an alternative to DFT-D method. We have also studied the importance of dispersion interaction and the effect of steric hindrance in aggregation of functionalized anthracenes and pentacenes. We have also employed molecular dynamics simulation methods to study the aggregation of anthracene molecules in toluene solution.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Anitha Ramraj

Copyright Statement

Acknowledgements

I thank my supervisors Prof. Ian H. Hillier and Dr. Neil A Burton for their support, valuable suggestions and guidance that helped me throughout this course. I would like to thank Dr. Mark Vincent for his valuable ideas, suggestions and technical support during my PhD. Thanks to all faculties, postdocs and students who I have worked closely. I would like to take this opportunity to thank Prof. P. Venuvanalingam for his guidance and support.

My special thanks to Mahesh who helped me financially to come to UK. I would also like to thank the past and present members of the computational chemistry group, Dr. Jonathan McNamara, Dr. Mahesh Sundararajan, Dr. Raman Sharma, Dr. Richard Dimelow, Dr. Rajeev Assary, Dr. Claudio, Dr. Prabha Jayapal, Dr. Chengkui Xiahou, Dr. Ganga Periyasamy, Dr. Shanthi Pandian, Dr. Slimane Dudou, Dr. Sheeba Jem, Dr. Rajesh Kumar, Dr. Xiao Shan, Krishnamoorthy, Ibrahim, Simon, Pavalina, Monica and Rafel for their support and help.

My heartiest thanks to my friends, Mahesh, Maheshwaran, Prabha, Shanthi, Sheeba, Krishnamoorthy, Priya, Karthik for accomadating me and making my stay in Manchester a memorable one which made me to feel like staying at home. And my special thanks to other department friends Gopal and Ganga for their company. I would like to express my thanks to Slimane and Sheeba for their help in setting up MD simulations and special thanks to Sheeba for helping me in writing codes. I would also like to express my gratitude to R. Thirumurugan, my school chemistry teacher who inspired me in first place to choose chemistry as my degree subject and for his continuous moral support.

I really thank my dear friend Cheng kui for her warm friendship, support, encouragement and company throughout my stay in Manchester. I also thank my friends

Lee, Jennifer, Kalai, Maha, Kalaiselvi, Pavi, Divya, Kings, Nandha, Subbu, Anil, Bibin for making me smile during stressfull times and friends from SRCM meditation centre.
I specially thank my dear friend Badri for his warm friendship, encouragement, moral support and advice which helped me to overcome all difficulties in my first year.

I thank my mom, dad, periya anna, china anna, periya anni, chinna anni and my nieces Swetha and Akshaya kutties for their support, prayers, wishes and their trust towards me for which I'm always grateful to them. I thank my best friend, better half Anand for his support, encouragement, patience and motivation throughout my PhD. I thank God for blessing me with Anand in my life. It is because of Anand's moral support and encouragement that I'm being here.

Thanks to University of Manchester, Overseas Research Scholarship for funding.

Last but not least, I would like to thank God for His blessings.

Those who work at a thing heart and soul not only achieve success in it but through their absorption in that they also realize the supreme truth—Brahman. Those who work at a thing with their whole heart receive help from God.

-Swami Vivekananda

Dedicated to

இன்று நான் உயர்ந்து நிற்க பாடு பட்டு

துன்பங்களில் துணையாய் நின்று

இன்பங்களை இணைந்து கொண்டாட

நான் தேடும் போது

இறைவியாய் ஆகி விட்டு

மேலே நின்று புன்முறுவல் செய்யும்

என் பாட்டிக்கு சமர்ப்பணம்....

என்றும் உன் நினைவுடன்,

உன் அனிதா

Publications

1. Rajesh K. Raju, **Anitha Ramraj**, Ian H. Hillier, Mark A. Vincent and Neil A. Burton, '*Carbohydrate-aromatic π interactions: A test of density functionals and the DFT-D method.'* ***Phys. Chem. Chem. Phys.*** 2009, **11**, 3411-3416.
2. Rajesh K. Raju, **Anitha Ramraj**, Mark A. Vincent, Ian H. Hillier and Neil A. Burton, '*Carbohydrate-protein recognition probed by density functional theory and ab initio calculation including dispersive interactions.'* ***Phys. Chem. Chem. Phys.*** 2008, **10**, 6500-6508.
3. **Anitha Ramraj**, Ian H. Hillier, Mark A. Vincent and Neil A. Burton, '*Assessment of approximate quantum chemical methods for calculating the interaction energy of nucleic acid bases with graphene and carbon nanotubes.'* ***Chem. Phys. Lett.*** 2010, **484**, 295-298.
4. **Anitha Ramraj** and Ian H. Hillier, '*Binding of pollutant aromatics on carbon nanotubes and graphite.'* ***J. Chem. Inf. Model.*** 2010, **50**, 585-588.
5. **Anitha Ramraj**, Rajesh K. Raju, Qiantao Wang, Ian H. Hillier, Richard A. Bryce, Mark A. Vincent, *'An evaluation of the GLYCAM06 and MM3 force fields, and the PM3-D* molecular orbital method for modelling prototype carbohydrate-aromatic interactions.'* ***J. Mol. Graph. Model.*** 2010,

Chapter 1

Introduction

1.0 Introduction to noncovalent interactions

A brief introduction to the noncovalent interactions is given in this chapter. We have also given an introduction to the role of noncovalent interactions in carbohydrate-protein recognition, functionalization of nanotubes and organic molecules aggregation through stacking.

Weak interactions such as hydrogen bonds or dispersive interactions, which are responsible for π-stacking interactions, are traditionally invoked. Noncovalent interactions are of pivotal importance in many areas of chemistry, biology, and material science and the interaction between π-aromatic molecules in particular has received much attention due to its important roles in a variety of phenomena such as base-pair stacking in DNA,[1] tertiary structure of proteins,[2, 3] self-assembly of synthetic molecules, [4,5]host-guest interactions including drug binding, and so forth. Intermolecular interactions between π systems are one of the principal noncovalent forces governing molecular recognition and biomolecular structure,[7-9] but are the least understood of the noncovalent interactions. It has been estimated that around 60% of aromatic side chains in proteins participate in π- π interactions.[2] While individually weak, the addictive power of these interactions leads to large effects, DNA structure is a well known example.[10] Recent studies on the stacked DNA bases pairs as well as stacked amino acids showed that the stabilization from stacking interactions can be surprisingly large and almost reaching stabilization due to hydrogen- bonding.[11] Stacking interactions is limited not only to the aromatic system but also for interaction of aromatic system with other delocalized π-electron systems like peptide bonds or even between with two

systems with delocalized electrons. The preference of aromatic guest molecules to orient in a fashion similar to that of sandwich or parallel-displaced stacked forms in biological or chemical receptors is evident.[7] Hunter et al. studied the orientational preferences of phenylalanine side chains in proteins using crystallographic data and observed that phenylalanine-phenylalanine interacting pairs are found in a wide range of T-shaped (edge-to-face) and parallel-displaced (offset-stacked) arrangements.[3] Arene-Arene interactions play a crucial role in protein-ligand recognition and concomitantly to drug design. The vast majority of X-ray crystal structures of protein complexes with small guest molecules or drugs reveal that the stabilization of these guest molecules or drugs in proteins comes from a variety of nonbonded interactions such as π-π stacking, OH/π, CH/π, NH/π and cation/π interactions, between aromatic side chains of proteins and guest/drug molecules.[2, 7] The complex of the enzyme acetylcholinesterase (Ache) with the drug E2020 (Aricept), which was developed to treat symptoms of Alzheimer's disease is one example of these interactions. Cryptolepine, a naturally occurring indoquinoline alkaloid used as an antimalarial drug intercalates into DNA and stacks between nucleic acid bases.[14] The cryptolepine molecule has no hydrogen-bonding contacts either with nucleic acid bases or with solvents. The absence of such interactions suggests that stacking forces alone provide the stabilizing mechanism of the complex.

Carbohydrate-protein interactions play crucial roles in biological mechanisms, for example in bacterial adhesions and toxins, viral glycoproteins, and in various amyloid forming proteins such as those associated with Alzheimer and Creutzfeldt diseases. A fundamental question involved in these kinds of interactions is how carbohydrates recognize a range of proteins. The recognition of carbohydrates in proteins is also mediated through nonbonded interactions between sugars and aromatic aminoacids.[15-18] The preference for carbohydrates to occupy stacked positions with aromatic side chains reveals the importance of carbohydrate-π interactions. Noncovalent interactions comprise interactions between permanent multipoles, between a permanent multipole and an induced multipole, and between two induced multipoles. Dispersion, electrostatic, induction and exchange-repulsion forces are the fundamental forces that gave stabilization to nonbonded interactions. Electrostatic interaction operates between

two charged dipoles and is a long-range interaction operating over few angstrom distances. The term induction refers to the interaction of the dipole moment of a polar molecule and the induced dipole moment of a nonpolar molecule and is also known as the Debye force. The London dispersion force is the weakest intermolecular interaction and it is a temporary attractive force that results when the electrons in two adjacent molecules occupy positions that make the atoms form temporary dipoles. This is sometimes referred to as an induced dipole-induced dipole interaction. Because of the constant motion of electrons around nuclei, an atom or molecule can be instantaneous polarized when its electrons around nuclei, an atom or molecule can be instantaneous polarized when its electrons are distributed unsymmetrically around the nucleus.

Carbon nanotube functionalized by biomolecules hold a great promise for molecular probes and sensors. An effective scheme to functionalize nanotubes-FET sensors was found to be simulataneously achieve robust, reproducible of nanotubes with molecular flexibility promising sensitivity to a wide spectrum of analytes. Nucleic acid biopolymers are found to be intriguing candidates for the molecular targeting layer since they can be engineered, using direct evolution, for affinity to a wide variety of targets, including small molecules and proteins. It has also been found that ss-DNA have high affinity for carbon nanotubes due to a favourable π-π stacking.[19] Due to aggregation of nanotubes, CNTs are very difficult to disperse homogeneously in solution. One of the approach to disperse nanotube is to separate the nanotube by attaching molecules which dissolve in solution. This process is known as nanotube functionalization. The attachment of a molecule to a nanotube can be done in two ways, ie. covalent attachment and noncovalent attachment.[20] Since nanotube consists of benzene aromatic rings, they can be functionalized by aromatic molecules through noncovalent interactions. This noncovalent functionalization of nanotubes is attractive because it offers the possibility of attaching molecules without affecting the electronic network of the tubes. The noncovalent interactions are based on π-π stacking, and it is controlled by thermodynamics. Stacking interactions between nanotubes and polynuclear species can aid the controlled placement of the carbon structures onto various surfaces and nanoparticles.[21] Pyrene modified oxide surfaces have been shown to be employed for the patterned self-assembly of single walled carbon nanomaterials.[22]

Aromatic interactions such as π-π stacking interactions play an important role in supramolecular chemistry. Molecules with conjugated bonds involve these π- π stacking interactions by overlapping of *p*-orbitals. Large polycyclic aromatic hydrocarbons such as anthracene, tetracene and pentacene are of basic molecules in designing thin-film transistors. Such molecules can be used to design organic thin-film transistors by making use of π-π stacking interactions between these molecules. These acene molecules are promising candidates for organic semiconductors because their planar shapes assist crystal packing and the extended π system over molecules enables the overlapping of π systems between molecules. In recent years, anthracene and pentacene oligomers were developed and studied to obtain face-to-face interaction over edge-to-face interaction in their crystal structures. Therefore the understanding of crystal structures of these acene molecules has become necessary in order to design more efficient organic thin-film transistors.

Among these interactions, the dispersive energy is described by the damped interatomic potential of the form C_6/R^6. But the accurate treatment of these interactions including dispersion is still computationally challenging. To calculate such interactions, high-level *ab initio* methods, at least at the level of MP2 with larger basis sets are required, with the associated computational expense. Density functional theory (DFT), which is an alternative to ab initio methods with less computational expense, however failed to describe dispersive interactions. An alternative strategy is to augment an explicit R^{-6} term, which is to describe interatomic dispersive interactions, with QM description of the system. This approach can be used with DFT and semi-empirical theory.[23-24]

1.1 Brief description of chapters

This thesis presents our computational studies involving noncovalent interactions in chemical and biological recognition.

Chapter 2 presents the theoretical background of our computational approaches we have used in our projects. This chapter explains the basic electronic structure theory of modern computational chemistry such as Hartree-Fock (HF) theory, density functional theory (DFT) theory and semi-empirical theory. The theory behind molecular dynamics simulations and various principles behind the data analysis is also explained.

In chapter 3, we discuss the importance of carbohydrate-aromatic interactions in carbohydrate-protein recognition. We present our results of electronic structure calculations on carbohydrate-aromatic interactions involving complexes of glucose and aromatic aminoacids, tyrosine and tryptophan. We have studied the structural and energetic aspects of sugar-aromatic complexes using mainly a density functional theory model with empirical corrections for dispersion interactions in detail. We have also explored the role of CH/π and OH/π interactions in the recognition process. We also present our analysis on changes in vibrational frequencies and NMR chemical shifts due to carbohydrate-aromatic interactions in these complexes. We have also discussed the performance of semi-empirical methods implemented with dispersion correction term such as PM3-D and PM3-D* to study carbohydrate-aromatic interactions.

In chapter 4, we present our continuation studies on carbohydrate-aromatic interactions. We present the performance of a number of density functionals and methods for the accurate description of carbohydrate-aromatic interactions. We also present the studies of simple models involving CH/π and OH/π interactions. We discuss the benchmark calculations and the performance of difference density functionals compared to benchmark calculations.

Chapter 5 presents the noncovalent functionalization of nanotubes by pollutant molecules. We discuss the role of noncovalent interactions in the adsorption of pollutant molecules on graphene sheets and nanotubes. The results of semi-empirical method PM3-D calculations on organic pollutants and C96, [6, 6] nanotubes are presented and discussed.

In chapter 6, we present our studies on noncovalent functionalization of graphene and nanotubes by nucleic acid bases and aromatic aminoacids. We have explored the size effects of graphene sheets and circumference effects of nanotubes by choosing different sizes of graphene sheets and different types of nanotubes. We discuss the performance of semi-empirical methods PM3-D and PM3-D*. We also present the results of calculations of density functional theory and M0x functionals.

In chapter 7, we discuss the role of noncovalent interactions and steric effects in functionalized anthracene and pentacene molecules. We have employed quantum mechanical methods to explore the effect of dispersion and steric hindrance to gain insights into the arrangement of functionalized anthracene and pentacene in their crystal forms. We also present the results of molecular dynamics simulations on anthracene aggregation in toluene solution.

1.2 References

1. Hunter, C. A.; Singh, J.; Thornton, J. M. *Journal of Molecular Biology*, **1991**, *218*, 837.
2. Saenger, W. *Prinicples of Nucleic Acid Structure; Springer-Verlag*, **1984**.
3. Burley, S. K.; Petsko, G. A. *Science*, **1985**, *v229*, 23.
4. Classens, C. G.; Stoddart, J. F. *Journal of Physical Organic Chemistry*, **1997**, *10*, 254.
5. Glaser, R.; Dendi, L. R.; Knotts, N.; Barnes, C. L. *Crystal Growth & Design*, **2003**, *3*, 291.
6. Hunter, C. A. *Chemical Society Reviews*, **1994**, *23*, 101.
7. Meyer, E. A.; Castellano, R. K.; Diederich, F. Angewandte *Chemie-International Edition*, **2003**, *42*, 1210.
8. Muller-Dethlefs, K.; Hobza, P. *Chemical Reviews*, **2000**, *100*, 143.
9. Hobza, P.; Sponer, J. *Chemical Reviews*, **1999**, *99*, 3247.
10. Watson, J. D.; Crick, F. H. C. *Nature*, **1953**,*171*, 737.
11. Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. *Physical Chemistry Chemical Physics*, **2006**, *8*, 1985.
12. Kryger, G.; Silman, I.; Sussman, J. L. *Structure*, **1999**, *7*, 297.
13. Brana, M. F.; Cacho, M.; Gradillas, A.; De Pascual-Teresa, B.; Ramos, A. *Current Pharmaceutical Design*, **2001**, *7*, 1745.
14. Lisgatren, J. N.; Coll, M.; Portugal, J.; Wright, C. W.; Aymani, J. *Nature Structural Biology*, **2002**, *9*, 57.
15. Sujatha, M. S.; Balaji, P. V. *Proteins: Structure, Function, and Genetics*, **2004**, *55*, 44.
16. Sujatha, M. S.; Sasidhar, Y. U.; Balaji, P. V. *Protein Science*, **2004**, *13*, 2502.
17. Sujatha, M. S.; Sasidhar, Y. U.; Balaji, P. V. *Biochemistry*, **2005**, *44*, 8554.
18. Sujatha, M. S.; Sasidhar, Y. U.; Balaji, P. V. *Journal of Molecular Structure: THIOCHEM*, **2007**, 814.
19. Staii, C.;Johnson, A. T. Jr., *Nano Lett*.,**2005,** *5*, 1774.
20. Tasis,D.; Tagmatarchis, N.; Bianco, A.; Prato, M. *Chem. Rev*. **2006,** *106*, 1105
21. Zhu, J.; Yudasaka, M.; Zhang, M.; Kasuya, D.; Iijima, S. *NanoLett.* **2003**, *3*, 1239
22. Zhu, J.; Yudasaka, M.; Zhang, M.; Iijima, S.*J. Phys. Chem. B* **2004**, *108*, 11317
23. . Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A*, 2005, **109**, 5656.
24. Grimme, S. *J. Comput. Chem.*, 2004, **25**, 1463.

Chapter 2

Theoretical Background

2.0 Introduction

Computational chemistry has become a powerful tool for the chemists in most of the research areas with the development of fast processors. The application of computational chemistry methods has emerged in many areas such as molecular modelling, nanotechnology, pharmaceutical chemistry and material science. The theoretical methods and principles behind these methods are described in this chapter. In our studies, we have mainly used ab initio methods, density functional theory methods, semi-empirical methods, and molecular dynamics simulation methods.

2.1 *Ab initio* methods

2.1.1 The Schrödinger equation

The fundamental principle of quantum mechanics is to find a solution of the time-independent, non-relativistic Schrödinger equation[1] for molecules and chemical systems.[2-7] The wavefunction, ψ, exists for any chemical system, and that appropriate operators (functions), which act upon ψ return the observable properties of the system. The operator, that returns the system energy, E, as an eigenvalue is called the Hamiltonian operator, $\hat{H}$. Therefore, by solving Schrödinger equation, electronic structure and properties of any molecule may be determined.

$$\hat{H}\,\Psi = E\,\Psi \qquad (2.01)$$

The typical form of the Hamiltonian operator takes into account five contributions to the total energy of a system: the kinetic energies of the electrons and nuclei, the attraction

of the electrons to the nuclei, and the interelectronic and internuclear repulsions. Expressing the Hamiltonian into mathematical notation,

$$H = -\sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 - \sum_k \frac{\hbar^2}{2m_k}\nabla_k^2 - \sum_i \sum_k \frac{e^2 Z_k}{r_{ik}} + \sum_{i\langle j} \frac{e^2}{r_{ik}} + \sum_{k\langle l} \frac{e^2 Z_k Z_l}{r_{kl}} \qquad (2.02)$$

In atomic units,

$$H = -\sum_i \frac{1}{2}\nabla_i^2 - \sum_k \frac{1}{2m_k}\nabla_k^2 - \sum_i \sum_k \frac{Z_k}{r_{ik}} + \sum_{i\langle j} \frac{1}{r_{ij}} + \sum_{k\langle l} \frac{Z_k Z_l}{r_{kl}}$$

where i and j run over electrons, k and l run over nuclei, $\hbar$ is Planck's constant divided by 2π, m_e is the mass of the electron, m_k is the mass of the nucleus k, ∇^2 is the Laplacian operator, e is the charge on the electron, Z is an atomic number, and r_{ab} is the distance between particles a and b. The last three potential energy terms appear exactly as in classical mechanics.

2.1.2 Variational theorem

To decide whether a trial wave function is better than other trial wave function, variational theorom is used, which states that '*for the ground state the energy calculated from an approximation to the true wave function will always be greater than the true energy E_0'*. Therefore the better the wave function the lower the energy, and for any wave function, Ψ, the energy, E, computed as the expectation value of the Hamiltonian, is an upper bound to the exact ground state energy, E_0. So the variational energy is defined as,

$$E = \frac{\int \Psi^* \hat{H} \Psi \, d\mathbf{r}}{\int \Psi^* \Psi \, d\mathbf{r}} \geq E_0 \qquad (2.03)$$

Equation (2.03) is the mathematical expression of the variation theorem. This theorem allows the calculation of an upper bound for the system's ground state energy. The

function Ψ is called a trial variation function, and the ratio of integrals in (2.03) is called the variational integral. To arrive at a good approximation to the ground-state energy E_0, many trial variation functions are employed in order to find the one that gives the lowest eigenvalue of the variational integral.[2-7]

2.1.3 Born-Oppenheimer approximation

Accurate wave functions for systems are extremely difficult to calculate because of the correlated motions of particles. In order to simplify the problem somewhat, it is good to consider some approximation. According to the Born-Oppenheimer approximation, the nuclei of molecular systems are moving much more slowly than electrons since nuclei are much heavier than electrons. Therefore, the electrons are considered to be moving in the field of fixed nuclei. As such, it is convenient to decouple these two motions, and compute electronic motions, and compute electronic energies for fixed nuclear positions. That is, the nuclear kinetic energy term is taken to be independent of the electrons, correlation in the attractive electron-nuclear potential energy term is eliminated, and the repulsive nuclear-nuclear potential energy term becomes a simply evaluated constant for a given geometry. According to this approximation, the kinetic energy of the nuclei can be neglected, and the repulsion between nuclei can be considered to be constant. An *electronic* Hamiltonian can be constructed which neglects the kinetic energy of the nuclei where the nuclear-nuclear repulsion is considered constant. The electronic Hamiltonian describing the motion of N electrons in a field of M nuclei, in atomic units is

$$\hat{H}_{elec} = -\sum_i \frac{1}{2}\nabla_i^2 - \sum_i \sum_k \frac{Z_k}{r_{ik}} + \sum_{i=1} \sum_{j>i} \frac{1}{r_{ij}} \tag{2.04}$$

The first term in the electronic Hamiltonian describes the kinetic energy of the electrons. The second term involves the electron-nuclear attractions, whilst the final term represents the electron-electron repulsions.[2-7] The classical nuclear–nuclear

repulsion term is constant within the Born–Oppenheimer approximation and is usually added to the electronic energy at the end of a calculation.[2-7]

$$E_{tot} = E_{elec} + \sum_{k=1} \sum_{l>k} \frac{Z_k Z_l}{r_{kl}} \tag{2.05}$$

2.1.4 Antisymmetry or Pauli exclusion principle

Pauli exclusion principle states that *'no two electrons can be characterized by the same set of quantum numbers'*. The electronic Hamiltonian of equation 2.04 depends only on the spatial coordinates of the electrons. However, the full description of an electron requires the specification of its spin. This is done by introducing two spin functions, $\alpha(\omega)$ and $\beta(\omega)$. In this formalism, an electron is not only described by the spatial coordinates q, but also by one spin coordinate ω. In addition, many-electron wave functions must be antisymmetric, that is, they muct change sign whenever the coordinate of two electrons are interachnged,[8]

$$\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_j, \ldots, \mathbf{x}_N) = -\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_j, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_N) \tag{2.06}$$

where $x = \{q, \omega\}$ denotes not only the three spatial coordinates but also the spin coordinate ω. The requirement imposed by equation 2.06, sometime called the antisymmetry principle,[9] is a very general statement of the familiar Pauli Exclusion Principle.[8]

2.1.5 Slater determinants

The requirement of antisymmetry can be achieved by constructing the wave function from a Slater determinant.[10] The columns in Slater determinants are single electron wave functions, *orbitals*, whilst the electron co-ordinates are along the rows. The

spatial distribution and spin of an electron is usually described by a *spin orbital*, $\chi(\mathbf{x})$. From each spatial orbital $\psi(\mathbf{r})$ two different spin orbitals can be obtained by multiplying by the α or β spin functions, respectively. For the general case of N electrons and N spin orbitals, the Slater determinant is

$$\Psi(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N) = (N!)^{-1/2}\begin{vmatrix} \chi_i(\mathbf{x}_1) & \chi_j(\mathbf{x}_1) & \ldots & \chi_N(\mathbf{x}_1) \\ \chi_i(\mathbf{x}_2) & \chi_j(\mathbf{x}_2) & \ldots & \chi_N(\mathbf{x}_2) \\ \ldots & \ldots & \ldots & \ldots \\ \chi_i(\mathbf{x}_N) & \chi_j(\mathbf{x}_N) & \ldots & \chi_N(\mathbf{x}_N) \end{vmatrix} \qquad (2.07)$$

The factor $(N!)^{-1/2}$ is a normalisation constant. A Slater determinant has N electrons occupying N spin orbitals ($\chi_i,\chi_j,\ldots,\chi_N$) without specifying which electron is in which orbital. Interchanging two rows of a Slater determinant corresponds to interchanging the co-ordinates of two electrons which changes the sign of the determinant, thus meeting the requirement of the antisymmetry principle.[2-7] If two electrons occupy the same spin orbital then two columns of the determinant are equal, and so the determinant is zero, satisfying the Pauli principle.[9] Slater determinants introduce *exchange correlation* in that the motion of two electrons with parallel spin is correlated, but the motion of electrons with opposite spins remains uncorrelated.[2-7]

2.1.6 The Hartree-Fock self-consistent field method

Fock first proposed the extension of Hartree's SCF procedure to Slater determinantal wave functions,[11] which is central to solving electronic structure problems. Just as with Hartree product orbitals, the HF MOs can be individually determined as eigenfunctions of a set of one-electron operators, and the interaction of each electron with the field of all other electrons includes exchange effects as well as Coulomb repulsion. Later Roothaan[12] described matrix algebraic equations that permitted HF calculations to be carried using a basis set representation of the MOs using a linear combination of atomic orbitals (LCAO). Restricted Hartree–Fock (RHF) methods are considered to be simplified solution of HF equations because of the simplification of considering closed shell electronic configuration. Restricted spin orbitals are constrained such that the

spatial function is the same for both α and β spin functions and this is referred to as *Restricted Hartree-Fock* (RHF) theory.[12] Detailed description of UHF (unrestricted Hartree-Fock) theory can be found elsewhere.[13, 14]

The HF equation is,

$$\hat{f}(i)\left|\chi_a\right\rangle = \varepsilon\left|\chi_a\right\rangle \tag{2.08}$$

where the one electron Fock operator, $\hat{f}(i)$ is defined for each electron *i* as

$$\hat{f}(i) = \hat{h}(i) + \hat{V}^{HF}(i) \tag{2.09}$$

where $\hat{h}(i)$ the core Hamiltonian operator containing kinetic energy and nuclear electronic energy terms,

$$\hat{h}(i) = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{Z_A}{r_{iA}} \tag{2.10}$$

and the final term, the HF potential, $\hat{V}^{HF}(i)$

$$\hat{V}^{HF}(i) = \sum_{b\neq a}\hat{J}_b(i) - \sum_{b\neq a}\hat{K}_b(i) \tag{2.11}$$

where the first term ($\hat{J}_b(i)$) is called Coulomb operator, defined as,

$$\hat{J}_b(1) = \int\left|\chi_b(2)\right|^2 r_{12}^{-1}dx_2 \tag{2.12}$$

which represents the average local potential at x_1 arising from an electron in the spin orbital χ_b, The second term is called exchange operator ($\hat{K}_b(i)$), comes from the antisymmetric nature of the determinantal wave function. This describes the effect of exchange interaction between like spin electrons, and is defined as,

$$\hat{K}_b(1)\chi_a(1) = \int [\chi_b(2)^* r_{12}^{-1} \chi_a(2) dx_2] \chi_b(1) \tag{2.13}$$

In RHF formalism, the HF equation can be written as,

$$f(1)\chi_i(1) = \varepsilon_i \chi_i(1) \tag{2.14}$$

The orbital χ_i is expanded as k finite number of functions using a linear expansion,

$$\chi_i = \sum_{\mu=1}^{k} C_{\mu i} \phi_\mu \tag{2.15}$$

If the $\{\phi_\mu\}$ was complete, this would be an exact expansion, but is limited to finite number due to the computational demand. Upon inserting equation 2.15 in equation 2.14 the HF equation is transformed in to the Roothan matrix equation,[12] which has the form,

$$\mathbf{FC} = \varepsilon \mathbf{SC} \tag{2.16}$$

where C is a k by k matrix of coefficients, ε is the diagonal matrix of orbital energies, S is the overlap matrix and F is the Fock matrix. Both overlap matrix and Fock matrix can be defined as,

$$S_{\mu\nu} = \int \phi_\mu^*(1)\phi_\nu(1) dx_1 \tag{2.17}$$

$$F_{\mu\nu} = H_{\mu\nu}^{core} + \sum_{\lambda\sigma}^{k} P_{\lambda\sigma} \left[(\mu\nu \mid \lambda\sigma) - \frac{1}{2}(\mu\lambda \mid \sigma\nu) \right] \tag{2.18}$$

where the P is the density matrix contains the elements,

$$P_{\mu\nu} = 2\sum_{a}^{N/2} C_{\mu a} C_{\nu a} \tag{2.19}$$

The notation $(\mu\nu \mid \lambda\sigma)$ also implies a specific integration, in this case

$$(\mu\nu \mid \lambda\sigma) = \iint \phi_\mu(1)\phi_\nu(1)\frac{1}{r_{12}}\phi_\lambda(2)\phi_\sigma(2)dr(1)dr(2) \tag{2.20}$$

At the same time, the process of solving equation 2.16 to find orbital energies and coefficients begins with a specified basis set and geometry and with a guess for the MO coefficients, which are used to construct the Fock matrix. The new Fock matrix is diagonalised to yield a new set of MO coefficients, which are used to form a new Fock matrix. The procedure is continued until the set of coefficients used to construct the Fock matrix are equal to those obtained from the diagonalisation. At self-consistency the potential generated by the electron density is identical to that produced by solving for the electron distribution.

2.2 Basis sets

A basis set is a set of functions used to express the molecular orbitals, which are expanded as a linear combination of such functions with the coefficients to be determined. Usually these functions are atomic orbitals, in that they are centered on atoms, but functions centered in bonds or lone pairs have been used as have pairs of functions centered in the two lobes of a p orbital. Additionally, basis sets composed of sets of plane waves down to a cutoff wavelength are often used, especially in calculations involving systems with periodic boundary conditions. In principle a complete set of basis functions must be used to represent the spatial orbitals exactly, giving an energy equal to that given in the variational expression [Eq. (2.03)]. This limiting energy is known as the *Hartree-Fock limit* and is not the exact ground-state energy since it still ignores the effects of electron correlation. However, it is not computationally feasible to use an infinite basis so a truncated expansion is used. The basis sets used in molecular orbital calculations are usually composed of atomic orbitals. One choice for the orbitals would be to use Slater type orbitals (STOs),[15]

commonly used in calculations of many electron atoms but their analytical solution is difficult. Boys[16] proposed an alternative to the use of STOs, which simplifies the radial decay of the STOs be changed from e^{-r} to e^{-r^2}. That is, the AO-like functions are chosen to have the form of the Gaussian function. The general functional form of the normalized Gaussian-type orbital (GTO) in atom-centred cartesian coordinates is

$$\phi(x,y,z;\alpha,i,j,k) = \sum_{a=1}^{M} c_a \phi(x^i y^j z^k e^{-\alpha(x^2+y^2+z^2)} \qquad (2.21)$$

where M is the number of Gaussians used in the linear combination, and the coefficients *c* are chosen to optimize the shape of the basis function sum and ensure normalization. When the basis function is defined as a linear combination of Gaussian, it is referred to as a *'contracted'* basis function, and the individual Gaussians from which it is formed are called *'primitive'* Gaussians. α is an exponent controlling the width of the GTO, and *i, j,* and *k* are the non-negative integers that dictate the nature of the orbital in a Cartesian sense. Thus, in a basis set of contracted GTOs, each basis function is defined by the contraction coefficients *c* and exponents α of each of its primitives. If all the three indices (*i, j, k*) are zero the GTO has spherical symmetry, and is called an s-type GTO. When only one of the indices is one, the function has an axial symmetry about a single axial Cartesian axis and is called *p*-type GTO. Therefore, there are three possible *p*-type GTOs namely p_x, p_y and p_z orbitals and when the sum of the indices is equal to two, the orbital is called a *d*-type GTO. Even though there are six different combinations of the index values, only five solutions of Schrödinger equation are found for hydrogen atom, which are sufficient enough to span all possible values of the z component of the orbital angular momentum of *l=2*. These five functions are usually referred to as, *xy*, *xz*, *yz*, x^2-y^2, and $3z^2$-r^2. First three of the canonical d functions are common with the cartesian d functions and latter two are the combination cartesian *d* functions. The basis sets used for the majority of our work are the Pople basis sets.[17, 18, 19] Depending on the number of basis functions used, basis sets can be classified into different types.

2.2.1 Minimal basis set

Minimal basis set is the one in which smallest number of basis functions required for describing all the electrons on each atom. The most common minimal basis set is STO-nG, where *n* value represents the number of Gaussian primitive functions comprising a single basis function. In these basis sets, the same number of Gaussian primitives comprise core and valence orbitals. Commonly used minimal basis sets of this type are: STO-3G, STO-4G, STO-6G, STO-3G* (Polarized version of STO-3G).

2.2.2 Double zeta (DZ) and Triple zeta (TZ) basis sets

Inadequate of minimal basis set to describe a non-spherical (anisotropic) electron distribution in molecules leads to the next improvement of basis sets which is doubling up the basis functions for each atomic orbital (AO). Hence, this basis set with two basis functions for each AO is called *Double Zeta* (DZ) type basis sets, where first basis function will be the contraction of first two primitive Gaussians and second basis function will be the third primitive Gaussian. Therefore this DZ basis set adds two s-functions for hydrogen (1s and 1s'), four s-functions (1s, 1s', 2s, and 2s') and two sets of p-functions (2p and 2p') for first row elements, and six s-functions and four p-functions for second row elements. Basis set which contains three times as many functions as the minimum basis, i.e. six s-functions and three sets of p-functions for the first row elements is known as *Triple Zeta* (TZ) basis set.

2.2.3 Split Valence Basis set

Basis sets in which there are multiple basis functions corresponding to each valence atomic orbital, are called valence double, triple, or quadruple-zeta basis sets. Since the different orbitals of the split have different spatial extents, the combination allows the electron density to adjust its spatial extent appropriate to the particular molecular environment. Minimum basis sets are fixed and are unable to adjust to different molecular environments. Basis sets in which there are multiple basis functions

corresponding to each atomic orbital, including both valence orbitals and core orbitals are called double, triple, or quadruple-zeta basis sets.

2.2.4 Polarization Basis set

The most common addition to minimal basis sets is probably the addition of polarization functions, denoted (in the names of basis sets developed by Pople) by an asterisk, *. Two asterisks, **, indicate that polarization functions are also added to light atoms (hydrogen and helium). These are auxiliary functions with one additional node. For example, the only basis function located on a hydrogen atom in a minimal basis set would be a function approximating the 1s atomic orbital. When polarization is added to this basis set, a p-function is also added to the basis set. This adds some additional needed flexibility within the basis set, effectively allowing molecular orbitals involving the hydrogen atoms to be more asymmetric about the hydrogen nucleus. For example, d-type functions can be added to a basis set with valence p orbitals, and f-functions to a basis set with d-type orbitals.

2.2.5 Diffuse Basis set

Diffuse functions can help to describe electron density far from the nucleus, where electrons may be loosely bound in anions or excited states.[2] These functions are of the same angular momentum as the valence shell orbitals, but have small exponents which give the orbital a large spatial extent. Normally, the diffuse basis sets are denoted as '+' or '++' with the first indicating one set of *s*- and *p*- diffuse functions on heavy atoms, whilst the second indicating that a diffuse *s*-function is also added to hydrogen.

2.3 Density functional theory

An alternative to the wave function based methods, which are computationally expensive, is density functional theory (DFT). DFT is based on determining the electron density, a physically observable quantity, rather than wave function. Unlike HF, in itself DFT contains no approximations. Unlike HF and post-HF methods, there is no a priori way to establish how good a given method is and no systematic way to improve upon it. It has however become clear that DFT methods often produce results of comparable quality to much more expensive post-HF methods. Density functional methods have their origins in the Thomas-Fermi-Dirac model.[20, 21]

2.3.1 Hohenberg-Kohn existence theorem

Electrons interact with one another and with an 'external potential'. Thus, in the uniform electron gas, the external potential is the uniformly distributed positive charge, and in a molecule, the external potential is the attraction to the nuclei. It is the ground-state density, which determines the Hamiltonian operator. Integration of the density gives the number of electrons, so all that remains to define the external potential (i.e., the charges and positions of the nuclei). The non-degenerate ground-state density must determine the external potential, and thus the Hamiltonian, and thus the wave function. But to be noted that the Hamiltonian determines not just the ground-state wave function, but all excited-state wave functions as well, so there is a tremendous amount of information coded in the density. This theorem states the ground state energy and all other electronic properties are uniquely determined by the electron density.[22] That is, the energy of an electronic system can be expressed in terms of the electron probability density $\rho(\mathbf{r})$ which represents the total electron density at a particular point in space **r**. The electronic energy $E[\rho(\mathbf{r})]$ is said to be a *functional* of the electron density. Therefore, for any given density there is a specific energy.

DFT attempts to compute the ground-state electronic energy E_0 and other ground-state molecular properties from the ground-state electron density ρ_0.[23] Hohenberg and Kohn[24] proved that for molecules with a nondegenerate ground state, the ground-state molecular energy, wave function, and all other molecular electronic properties are determined, in a unique way, by the ground-state electron probability density $\rho_0(x,y,z)$. Therefore, the ground-state electronic energy E_0 is a functional of ρ_0, that is, $E_o = E_\upsilon[\rho_0]$, where the υ subscript emphasizes the dependence of E_0 on the external potential $\upsilon(\mathbf{r})$, which differs for different molecules.[23]

Taking the average of the purely electronic Hamiltonian 2.04 for the ground state the expression $E = \overline{T} + \overline{V}_{Ne} + \overline{V}_{ee}$ is obtained. Each of these averages is a function of ρ_0:

$$E_0 = E_\upsilon[\rho_0] = \overline{T}[\rho_0] + \overline{V}_{Ne}[\rho_0] + \overline{V}_{ee}[\rho_0] \tag{2.22}$$

$\hat{V}_{Ne} = \sum_{i=1}^{n} \upsilon(\mathbf{r}_i)$, where $\upsilon(\mathbf{r}_i) = -\sum_\alpha Z_\alpha / r_{i\alpha}$ in atomic units, so the middle term on the r.h.s of 2.22 can be computed from

$$\overline{V}_{Ne}[\rho_0] = \left\langle \psi_0 \left| \sum_{i=1}^{n} \upsilon(r_i) \right| \psi_0 \right\rangle = \int \rho_0(\mathbf{r})\upsilon(\mathbf{r})d\mathbf{r} \tag{2.23}$$

where $\upsilon(\mathbf{r})$ is the nuclear-attraction potential-energy function for an electron located at point **r**. On the other side, the functionals $\overline{T}[\rho_0]$ and $\overline{V}_{ee}[\rho_0]$ are unknown, so we have

$$E_0 = E_\upsilon[\rho_0] = \int \rho_0(r)\upsilon(r)dr + F[\rho_0] \tag{2.24}$$

where the unknown functional $F[\rho_0]$, defined by $F[\rho_0] \equiv \overline{T}[\rho_0] + \overline{V}_{ee}[\rho_0]$, is independent of the external potential.

2.3.2 Hohenberg-Kohn variational theorem

As mentioned earlier, Hohenberg-Kohn existence theorem was unable to provide density of system. Later, Hohenberg-Kohn showed in a second theorem, just as with MO theory, the density obeys a variational principle. By choosing a candidate density, we can evaluate wave function and Hamiltonian by optimizing candidate density with variational principle. But still there is a problem choosing improved candidate densities rationally, and also there is no prescription to avoid solving the schordinger equation. From this theorem, it is known that there are mappings from the density onto the Hamiltonian and the wave function, and hence the energy. But there is no mechanical suggestion for it. This theorem states that for every trial density function $\rho(\mathbf{r})$ that satisfies $\int \rho_{tr}(\mathbf{r})d\mathbf{r} = n$ and $\rho_{tr}(\mathbf{r}) \geq 0$ for all $\mathbf{r}$, the following inequality holds: $E \leq E_{tr}$ where E is the exact energy.

2.3.3 Kohn-Sham method

Kohn and Sham devised a practical method for finding ρ_0 and for finding E_0 and ρ_0. They considered a fictitious reference system (denoted by the subscript s) of n noninteracting electrons, each experiencing the same external potential-energy function $\upsilon_s(\mathbf{r}_i)$, where $\upsilon_s(\mathbf{r}_i)$ is such as to make the ground-state electron probability density $\rho_s(\mathbf{r})$ of the reference system equal to the exact ground-state electron density $\rho_0(\mathbf{r})$ of the molecule under study. The Hamiltonian of the reference system is

$$\hat{H}_s = \sum_{i=1}^{n}\left[-\frac{1}{2}\nabla_i^2 + \upsilon_s(\mathbf{r}_i)\right] \equiv \sum_{i=1}^{n}\hat{h}_i^{KS} \tag{2.25}$$

$$\hat{h}_i^{KS} \equiv -\frac{1}{2}\nabla_i^2 + \upsilon_s(\mathbf{r}_i) \tag{2.26}$$

$\hat{h}_i^{KS}$ is the one-electron Kohn-Sham Hamiltonian. Since the reference system *s* consists of noninteracting particles, the ground-state wave function $\psi_{s,0}$ of the system is the Slater determinant of the lowest-energy Kohn-Sham spin-orbitals u_i^{KS} of the reference system, where the spatial part $\theta_i^{KS}(r_i)$ of each spin-orbital is an eigenfunction of the one-electron operator $\hat{h}_i^{KS}$, that is

$$\psi_{s,0} = \left| u_1 u_2 ... u_n \right\rangle, \qquad u_i = \theta_i^{KS}(\mathbf{r}_i)\sigma_i \tag{2.27}$$

$$\hat{h}_i^{KS}\theta_i^{KS} = \varepsilon_i^{KS}\theta_i^{KS} \tag{2.28}$$

where σ_i is a spin function and the ε_i^{KS}'*s* are the Kohn-Sham orbital energies.
Equation 2.24 can also be expressed as

$$E_0 = E_\upsilon[\rho] = \int \rho(\mathbf{r})\upsilon(\mathbf{r})dr + \overline{T}[\rho] + \frac{1}{2}\iint \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 + E_{xc}[\rho] \tag{2.29}$$

$$\Delta\overline{T}[\rho] \equiv \overline{T}[\rho] - \overline{T}_s[\rho] \tag{2.30}$$

$$\Delta\overline{V}[\rho] \equiv \overline{V}_{ee}[\rho] - \frac{1}{2}\iint \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 \tag{2.31}$$

$$E_{xc}[\rho] \equiv \Delta\overline{T}[\rho] + \Delta\overline{V}_{ee}[\rho] \tag{2.32}$$

The key to an accurate KS DFT calculation of molecular properties is to get a good approximation to E_{xc}. In order to evaluate the terms in equation 2.29, it is necessary to find the ground-state electron density, which in turn is given by

$$\rho = \rho_0 = \sum_{i=1}^{n} \left| \theta_i^{KS} \right|^2 \tag{2.33}$$

and equation 2.29 is evaluated as

$$E_0 = -\sum_{\alpha} Z_\alpha \int \frac{\rho(\mathbf{r}_1)}{r_{1\alpha}} d\mathbf{r}_1 - \frac{1}{2}\sum_{i=1}^{n} \left\langle \theta_i^{KS}(1) \middle| \nabla_1^2 \middle| \theta_i^{KS}(1) \right\rangle + \frac{1}{2}\iint \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 + E_{xc}[\rho] \qquad (2.34)$$

The KS orbitals are found from Equation 2.35, which can also be written as in Equation 2.36.

$$\left[-\frac{1}{2}\nabla_1^2 - \sum_{\alpha} \frac{Z_\alpha}{r_{1\alpha}} + \int \frac{\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_2 + \upsilon_{xc}(1) \right] \theta_i^{KS}(1) = \varepsilon_i^{KS} \theta_i^{KS}(1) \qquad (2.35)$$

$$\hat{h}^{KS}(1)\theta_i^{KS}(1) = \varepsilon_i^{KS}\theta_i^{KS}(1) \qquad (2.36)$$

where the exchange-correlation potential υ_{xc} is found as the functional derivative of the exchange-correlation energy E_{xc}:

$$\upsilon_{xc}(\mathbf{r}) \equiv \frac{\delta E_{xc}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} \qquad (2.37)$$

The problem in using the Kohn-Sham method to find ρ and E_0 is that the correct functional $E_{xc}[\rho]$ is not known, and therefore approximations to E_{xc} must be used.

2.4.4 The local-density approximation (LDA)

Within the LDA formalism, if ρ varies extremely slowly with position, then $E_{xc}[\rho]$ is accurately given by[23]

$$E_{xc}^{LDA}[\rho] = \int \rho(\mathbf{r}) \varepsilon_{xc}(\rho) d\mathbf{r} \qquad (2.38)$$

where $\varepsilon_{xc}(\rho)$ is the exchange plus correlation energy per electron in a homogeneous electron gas with electron density ρ.[25] Taking the functional derivative of $E_{xc}^{LDA}[\rho]$,

$$\upsilon_{xc}^{LDA} = \varepsilon_{xc}\big(\rho(r)\big) + \rho(r)\frac{\partial \varepsilon_{xc}(\rho)}{\partial \rho} \tag{2.39}$$

It can be shown that $\varepsilon_{xc}(\rho)$ can be written as $\varepsilon_{xc}(\rho) = \varepsilon_x(\rho) + \varepsilon_c(\rho)$, where $\varepsilon_x(\rho)$ is given by Equation 2.41 and $\varepsilon_c(\rho) = \varepsilon_c^{VWN}(\rho)$ being a very complicated function of ρ.[69] The particular expressions for υ_{xc}^{LDA} and E_x^{LDA} are given in Equations 2.40 and 2.42 respectively:

$$\upsilon_{xc}^{LDA} = \upsilon_x^{LDA} + \upsilon_c^{LDA}, \qquad \upsilon_x^{LDA} = -\big[(3/\pi)\rho(r)\big]^{1/3}, \qquad \upsilon_c^{LDA} = \upsilon_c^{VWN} \tag{2.40}$$

$$\varepsilon_x(\rho) = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\big(\rho(r)\big)^{1/3} \tag{2.41}$$

$$E_x^{LDA} \equiv \int \rho \varepsilon_x d\mathbf{r} = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\int \big[\rho(\mathbf{r})\big]^{4/3} d\mathbf{r} \tag{2.42}$$

For open-shell molecules and molecular geometries near dissociation the local-spin density approximation should be used (LSDA). Contrary to the LDA, where electrons with opposite spins paired with each other have the same spatial KS orbital, the LSDA allows these electrons to have different spatial orbitals $\theta_{i\alpha}^{KS}$ and $\theta_{i\beta}^{KS}$. In spin-DFT one deals separately with the electron density due to the spin-α electrons and the density of the spin-β electrons, so functionals such as E_{xc} are now expressed as $E_{xc} = E_{xc}\big[\rho^{\alpha}, \rho^{\beta}\big]$.[23]

2.4.5 The functionals E_x and E_c

As an aid in the development of approximate functionals for use in Kohn-Sham DFT, the functional E_{xc} is written as the sum of an exchange-energy functional E_x and a correlation-energy functional E_c:[23]

$$E_{xc} = E_x + E_c \tag{2.43}$$

E_x can be defined by the same formula used for the exchange energy in Hartree-Fock theory. For a closed-shell molecule we have

$$E_x = -\frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\langle \theta_i^{KS}(1)\theta_j^{KS}(2)\middle|1/r_{12}\middle|\theta_j^{KS}(1)\theta_i^{KS}(2)\right\rangle \tag{2.44}$$

Having defines E_x, the correlation-energy functional E_c is defined as the difference between E_{xc} and E_x, in other words, $E_c \equiv E_{xc} - E_x$, and Equation 2.43 follows.

2.4.7 Gradient-corrected and hybrid functionals

The LDA and LSDA are based on the uniform-electron-gas model, which is only appropriate for systems where ρ varies slowly with position. There is another formalism that goes beyond the LSDA and that is termed GGA, which stands for generalized-gradient approximation. In this approximation E_{xc}^{GGA} includes also in the integrand the gradients of ρ^{α} and ρ^{β}.[23] Thus,

$$E_{xc}^{GGA}\left[\rho^{\alpha},\rho^{\beta}\right] = \int f\left(\rho^{\alpha}(r),\rho^{\beta}(r),\nabla\rho^{\alpha}(r),\nabla\rho^{\beta}(r)\right)dr \tag{2.45}$$

E_{xc}^{GGA} is typically split into exchange and correlations parts, which are modelled separately, that is, $E_{xc}^{GGA} = E_x^{GGA} + E_c^{GGA}$. Some common gradient-corrected exchange functionals are the B88[26] and PW91[27] functionals. Commonly used gradient-corrected correlation functionals include the LYP,[28] P86[29, 30] and PW91[27] functionals. Any exchange functional can be combined with any correlation functional. For instance, the BLYP/6-31G* notation represents a density functional calculation done with the B88 exchange functional, with the KS orbitals expanded in a 6-31G* basis set. A kind of functional that is widely used is the hybrid exchange-correlation functional. A hybrid

functional mixes together the formula 2.44 for E_x with gradient-corrected E_x and E_c formulas. For example, the very popular B3LYP[31, 32] hybrid functional is defined by

$$E_{xc}^{B3LYP} = (1-\alpha_0-\alpha_x)E_x^{LSDA} + a_0 E_x^{exact} + a_x E_x^{B88} + (1-a_c)E_c^{VWN} + a_c E_c^{LYP} \qquad (2.46)$$

where E_x^{exact} is given by Equation 2.44, and where the parameter values a_0=0.20, a_x=0.72, and a_c=0.81 were chosen to give good fits to experimental molecular atomisation energies. Another prominent hybrid functional is the Becke half-and-half functional, termed BH&H.[33] This functional computes E_{xc} as

$$E_{xc}^{BH\&H} = 0.5E_x^{HF} + 0.5E_{xc}^{LSDA} \qquad (2.47)$$

2.4 Moller-Plesset perturbation methods

HF theory fails to adequately represent electron correlation. Moller-Plesset perturbation methods[34] include electron correlation effects by means of Rayleigh-Schrödinger perturbation theory (RS-PT), usually to second (MP2), third (MP3) or fourth (MP4) order. Perturbation theory essentially involves dividing the *true* Hamiltonian into two parts, a zero-order Hamiltonian ($\hat{H}_0$) and a perturbation ($\hat{V}$).

$$\hat{H} = \hat{H}_0 + \lambda\hat{V} \qquad (2.48)$$

$\hat{H}_0$ is exactly soluble, such that the eigenfunctions and eigenvalues are known and λ is a parameter that can vary between 0 and 1, and provides a scheme in which to gradually improve the eigenfunctions and eigenvalues of $\hat{H}_0$. The assumption that $\hat{V}$ is small compared to $\hat{H}_0$ allows the perturbed wave function and energy to be expressed as a power series in terms of the parameter λ.

$$\Psi_i = \Psi_i^{(0)} + \lambda\Psi_i^{(1)} + \lambda^2\Psi_i^{(2)} + \ldots = \sum_{n=0} \lambda^n \Psi_i^{(n)} \tag{2.49}$$

$$E_i = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \ldots = \sum_{n=0} \lambda^n E_i^{(n)} \tag{2.50}$$

where $E_i^{(1)}$ and $E_i^{(2)}$ are the first and second order corrections to the zero-order energy. To determine the corrections to the energy it is necessary to determine the wave functions to a given order. The energies are calculated from the following equations.

$$E_i^{(0)} = \int \Psi_i^{(0)} H_0 \, \Psi_i^{(0)} d\mathbf{x} \tag{2.51}$$

$$E_i^{(1)} = \int \Psi_i^{(0)} V \, \Psi_i^{(0)} d\mathbf{x} \tag{2.52}$$

$$E_i^{(2)} = \int \Psi_i^{(0)} V \, \Psi_i^{(1)} d\mathbf{x} \tag{2.53}$$

In Møller-Plesset perturbation theory (MPPT),[34] the zero-order Hamiltonian is taken to be the sum of the one-electron Fock operators for *N* electrons.

$$\hat{H}_0 = \sum_{i=1}^{N} \hat{f}_i = \sum_{i=1}^{N} \left(\hat{h}_i + \sum_{j=1}^{N} \left(\hat{J}_i - \hat{K}_i \right) \right) \tag{2.54}$$

The Hartree-Fock wavefunction $\Psi_i^{(0)}$ is an eigenfunction of $\hat{H}_0$ and thus the zero-order energy is taken to be a sum of the orbital energies of the occupied molecular orbitals.

$$E_0^{(0)} = \sum_{i=1}^{occ} \varepsilon_i \tag{2.55}$$

The true Hamiltonian is equal to the sum of the nuclear attraction terms and the electron repulsion terms.

$$\hat{H} = \sum_{i=1}^{N} \hat{h}_i + \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{r_{ij}} \tag{2.56}$$

However, the sum of the Fock operators counts the electron-electron repulsions twice, such that the perturbation becomes the exact electron-electron operator minus twice the average electron-electron repulsion operator.[2-7]

$$\hat{V} = \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} - \sum_{j=1}^{N}\left(\hat{J}_j - \hat{K}_j\right) \tag{2.57}$$

Thus, the first order energy $E_0^{(1)}$ is given by the average of the perturbation operator over the zero-order wavefunction [equation (2.52)].

$$E_0^{(1)} = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(J_{ij} - K_{ij}\right) \tag{2.58}$$

The sum of the zero-order and first-order energies is just the Hartree-Fock energy, which compares with equation for a closed-shell system.

$$E_0^{(0)} + E_0^{(1)} = \sum_{i=1}^{N}\varepsilon_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(J_{ij} - K_{ij}\right) \tag{2.59}$$

To improve the Hartree-Fock energy Møller-Plesset perturbation theory[34] must be used to at least second order. This level is referred to as MP2 and involves integrals of the type described by equation (2.53). The higher order wavefunction $\Psi_0^{(1)}$ is expressed as linear combinations of solutions to the zero-order Hamiltonian.

$$\Psi_0^{(1)} = \sum_j C_j^{(1)}\Psi_j^{(0)} \tag{2.60}$$

The $\Psi_j^{(0)}$ includes both single and double excitations, obtained by promoting electrons into the virtual orbitals obtained from a Hartree-Fock calculation. The second order energy for closed shell system in terms of spatial orbitals is

$$E_0^{(2)} = 2\sum_{ijab}^{N/2} \frac{\langle \psi_i \psi_j | \psi_a \psi_b \rangle \langle \psi_a \psi_b | \psi_i \psi_j \rangle}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} - \sum_{ijab}^{N/2} \frac{\langle \psi_i \psi_j | \psi_a \psi_b \rangle \langle \psi_a \psi_b | \psi_j \psi_i \rangle}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \quad (2.61)$$

where ψ_i and ψ_j are occupied orbitals and ψ_a and ψ_b denote virtual orbitals. According to Brillouin's theorem, which states *single excitations do not mix directly with the single determinant ground state,* $\Psi_0^{(0)}$, the above integrals will only be non-zero for double excitations. The second order correction to the energy, which is the first contribution to the correlation energy, therefore only involves a sum over doubly excited determinants. The summation is restricted so that each excitation is counted only once. Third-,[35] fourth-[36] and fifth-order[37] Møller-Plesset calculations (MP3, MP4 and MP5) calculations are also available.

2.5 Semi-empirical implementations of molecular orbital theory

In early days of computational chemistry, carrying out Hartree Fock calculations without further approximations even for small systems with minimal basis sets was a challenging task due to calculation of many electron integrals. So accepting the Hartree Fock theory as a framework several research groups turned their attention to implementing such theoretical methods to minimize the number of integrals and this turned into the development of semi–empirical methods. The basic approximations inherent in semi–empirical methods are summarized in this section.

2.5.1 CNDO formalisms

The secular equation can be written as,

$$\begin{vmatrix} F_{11}-ES_{11} & F_{12}-ES_{12} & \cdots & F_{1N}-ES_{1N} \\ F_{21}-ES_{21} & F_{22}-ES_{22} & \cdots & F_{2N}-ES_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ F_{N1}-ES_{N1} & F_{N2}-ES_{N2} & \cdots & F_{NN}-ES_{NN} \end{vmatrix} = 0 \qquad (2.62)$$

To replace the matrix elements in the HF secular equation 2.62, Pople and co-workers in 1965 developed the formalism called complete neglect of differential overlap (CNDO)[38] method by adopting following conventions, i) Just as in Extended Huckel Theory (EHT), the basis set is formed from valence STOs, one STO per valence orbital. In the original CNDO implementation, only atoms having s and p valence orbitals were addressed. In the secular determinant, overlap matrix elements are defined by

$$S_{\mu\upsilon} = \delta_{\mu\upsilon} \qquad (2.63)$$

where δ is the Kronecker delta. All two-electron integrals are parameterized according to the following scheme. To define,

$$(\mu\upsilon|\lambda\sigma) = \delta_{\mu\upsilon}\delta_{\lambda\sigma}(\mu\mu|\lambda\lambda) \tag{2.64}$$

Thus, the only integrals that are non-zero have μ and υ as identical orbitals on the same atom, and λ and σ also as identical orbitals on the same atom, but the second atom might be different than the first.

For the surviving two-electron integrals,

$$(\mu\mu|\lambda\lambda) = \gamma_{AB} \tag{2.65}$$

One-electron integrals for diagonal matrix elements are defined by

$$\left\langle \mu \left| -\frac{1}{2}\nabla^2 - \sum_k \frac{Z_K}{r_K} \right| \mu \right\rangle = -IP_\mu - \sum_k (Z_k - \delta_{Z_A Z_k})\gamma_{Ak} \tag{2.66}$$

The only terms remaining to be defined in the assembly of the HF secular determinant are the one-electron terms for off-diagonal matrix elements. These are defined as

$$\left\langle \mu \left| -\frac{1}{2}\nabla^2 - \sum_k \frac{Z_k}{r_k} \right| \upsilon \right\rangle = \frac{(\beta_A + \beta_B)S_{\mu\upsilon}}{2} \tag{2.67}$$

where μ and υ are centered on atoms A and B, respectively, the β values are semiempirical parameters, and $S_{\mu\upsilon}$ is the overlap matrix element computed using the STO basis set. While the CNDO method may appear to be moderately complex, it represents a vast simplification of HF theory. Equation 2.64 reduces the number of two-electron integrals having non-zero values from formally N^4 to simply N^2. A number of minor modifications to the conventions outlined above were explored, and the different methods had names like CNDO/1, CNDO/2, CNDO/BW, etc.

2.5.2 INDO formalism

Pople, Beveridge, and Dobosh in 1967 suggested modifications to the CNDO formalism to permit a more flexible handling of electron-electron interactions on the same center in order to model such spectroscopic transitions, and referred to this new formalism as 'intermediate neglect of differential overlap' (INDO). The key change is simply to use different values for the unique one-center two-electron integrals. When the atom is limited to a basis set of s and p orbitals, there are five such unique integrals

$$\begin{aligned}
&(ss|ss) = G_{ss} \\
&(ss|pp) = G_{sp} \\
&(pp|pp) = G_{pp} \\
&(pp|p'p') = G_{pp'} \\
&(sp|sp) = L_{sp}
\end{aligned} \tag{2.68}$$

The G and L values may be regarded as free parameters, but in practice they can be estimated from spectroscopic data. When the atomic valence orbitals include d and f functions, the number of unique integrals increases considerably, and the estimation of appropriate values from spectroscopy becomes considerably more complicated. One effect of the greater flexibility inherent in the INDO scheme is that valence bond angles are predicted with much greater accuracy than is the case for CNDO. These methods were improved by introducing MINDO/3 and SINDO by Bingham *et al.*[70]

2.5.3 NDDO formalism

The INDO model extends the CNDO model by adding flexibility to the description of the one-center two-electron integrals. In INDO, however, there continues to be only a single two-center two-electron integral, which takes on the value γ_{AB} irrespective of which orbitals on atoms A and B are considered. The neglect of diatomic differential overlap (NDDO) method relaxes the constraints on two-center two-electron integrals in

a fashion analogous to that for one-center integrals in the INDO method. Thus, all integrals $(\mu\upsilon|\lambda\sigma)$ are retained provided μ and υ are on the same atomic center and λ and σ are on the same atomic center. Most modern semi-empirical models are NDDO models.

2.5.4 Basic NDDO formalism

The INDO model extends the CNDO model by adding flexibility to the description of the one-center two-electron integrals. In INDO,[39] however, there continues to be only a single two-center two-electron integral, which takes on the value γ_{AB} irrespective of which orbitals on atoms A and B are considered. The neglect of diatomic differential overlap (NDDO) method relaxes the constraints on two-center two-electron integrals in a fashion analogous to that for one-center integrals in the INDO method. Thus, all integrals $(\mu\upsilon|\lambda\sigma)$ are retained provided μ and υ are on the same atomic center and λ and σ are on the same atomic center. Most modern semi-empirical models are NDDO models.

2.5.5 MNDO formalism

Dewar and Thiel[40, 41] reported a *Modified Neglect of Differential Overlap* (MNDO) method based on NDDO formalism for the elements C, H, O, and N. As with MINDO/3, Dewar and Thiel optimized the parameters of the MNDO model against a large set of molecular properties. General shortcomings of this method including the inaccurate description of hydrogen bonding and dispersion due to electron correlation. Recognizing the drawbacks of these models, Dewar and co-workers [42] modified the functional form of their model by including a nuclear- repulsion term. This approaches termed as *Austin Model 1* (AM1). The AM1 is originally described for the elements C, H, O and N and the nuclear repulsion energy between any two nuclei A and B is computed as,

$$V_N(A,B) = Z_A Z_B (s_A s_A \mid s_B s_B) + \frac{Z_A Z_B}{r_{AB}} \sum_{i=1}^{4} [a_{A,i} e^{-b_{A,i}(r_{AB}-c_{A,i})^2} + a_{B,i} e^{-b_{B,i}(r_{AB}-c_{B,i})^2}] \qquad (2.69)$$

AM1 Parameterization of elements such as B, F, Al, Si ,P, S, Cl, Zn, Ge, Br, I and Hg have been subsequently reported.[47] Further developments of AM1 have been carried out by Stewart using an optimizing algorithm with the option of two Gaussian functions instead of four found in Eqn. 2.69. This is called *Parametric Method 3* (PM3). Extending the MNDO method to *d* orbitals is a challenging task because of the explicit treatment of two-centre two-electron integrals that is required. Thiel *et al* have proposed a scheme to include the above integrals in AM1, PM3 and MNDO methods.[40]

2.6 DFT and PM3 methods augmented with an empirical dispersion term

In the DFT-D approach of Grimme[43] a pair-wise additive potential of the form C_6/R^6 is used to account for long-range dispersion effects that are in general poorly described with common density functionals. The dispersion corrected total energy is given by

$$E_{DFT-D} = E_{KS-DFT} + E_{disp} \qquad (2.70)$$

where $E_{KS\text{-}DFT}$ is the normal self-consistent Kohn-Sham energy obtained from a specific density functional and E_{disp} is an empirical term containing the dispersion correction:

$$E_{disp} = -s_6 \sum_{i=1}^{N_{at}-1} \sum_{j=j+1`}^{N_{at}} \frac{C_6^{ij}}{R_{ij}^6} f_{dmp}(R_{ij}) \qquad (2.71)$$

In this equation, N_{at} is the number of atoms in the system, C_6^{ij} is the dispersion coefficient for the pair of atoms *i* and *j* (calculated from the atomic C_6 coefficient of atoms *i* and *j*), S_6 is a scaling factor that depends on the density functional used and R_{ij} is the interatomic distance between atoms *i* and *j*. A damping function is used in order to avoid near singularities for small distances. This function is given by

$$f_{dmp}\left(R_{ij}\right)=\frac{1}{1+e^{-\alpha\left(R_{ij}/R_0-1\right)}} \qquad (2.72)$$

where R_0 is the sum of atomic van der Waals radii and α is a parameter determining the steepness of the damping function. In order to get consistent van der Waals radii R_0 for arbitrary elements they were derived from theoretical data, The radius of the 0.01 au electron density contours from ROHF/TZV computations of the atoms in the ground state scaled by a factor of 1.22 was taken. In order to obtain the composed dispersion coefficients C_6^{ij} a simple average of the form

$$C_6^{ij}=2\frac{C_6^i C_6^j}{C_6^i+C_6^j} \qquad (2.73)$$

is used. The atomic C_6 coefficients were taken from the original parameterization of the DFT-D method. Although other functionals may be used within the DFT-D formalism, we use here the simpler GGA functional used initially by Grimme (BLYP), in conjunction with a TZV (2d, 2p) basis set. The exponents of all polarization functions are taken from Dunning's cc-pVTZ basis set.[44] The corresponding value of S_6 of 1.4 was used, with the C_6 coefficients taken from Table 1 of Grimme's published paper. We do not consider basis set superposition errors (BSSE) in view of the quite large basis set. The value for S_6 is chosen to be 1.4 in line with the value used for the BLYP density functional for the PM3-D method as well. All PM3-D and PM3-D* calculations were performed with our own local semi-empirical molecular orbital program.[45]

2.7 GIAO method for NMR calculations

NMR calculations assess the energy difference between a system in the presence and absence of an external magnetic field. The chemical shift is proportional to the second derivative of the energy with respect to the external magnetic field and the internal magnetic moment of the nucleus. The inclusion of a perturbation in the form of magnetic field in the Hamiltonian makes the integrals to be evaluated more complex in nature. Unlike the electric field, which perturbs the potential energy term of the Hamiltonian, the magnetic field perturbs the kinetic energy term. It is the motion of the electrons that generates electronic magnetic moments. The nature of the perturbed kinetic energy operator is such that an origin must be specified defining a coordinate system for the calculation. This origin is termed as 'gauge origin'. There is considerable freedom in the choice of gauge origin and this freedom in the choice should not have any consequences since all physical quantities are required to be gauge-invariant. However, gauge invariance holds only for exact solutions to the Schrödinger equation and cannot be enforced for approximate wavefunctions using a limited number of basis functions. Hence, the choice of gauge is a practical problem for the calculations of magnetic properties. The Gauge Including Atomic Orbital (GIAO)[46-48] method is one of the methods used in computational chemistry to deal with the gauge problem. The GIAO method uses explicit field-dependent basis set functions for calculating magnetic properties. By a clever incorporation of the gauge origin into the basis function themselves, all matrix elements involving the basis functions can be arranged to be independent of it. The GIAO is thus a gauge transformed atomic orbital, with the new gauge origin at the nucleus. The GIAOs are better viewed as gauge transformation of the individual basis functions. Since the atomic orbitals are inherently local, such transformation has only a local field effect, ensuring an optimal gauge for the atomic orbital. The use of GIAOs eliminates any explicit reference to the global gauge origin. GIAOs are not proper gauge transformations neither of the wave function, nor of the molecular orbital and their use does not make the energy gauge invariant. GIAOs do however, ensure rapid basis set convergence for many second order magnetic properties, and thus resolve the gauge problem.

2.8 Molecular Dynamics

Molecular dynamics[49] is a principal theoretical technique used to simulate the dynamics of a molecular system based on Classical mechanics. In molecular dynamics, atoms and molecules are allowed to interact for a period of time by approximations of known physics. Integrating Newton's law of motion generates successive configurations in molecular dynamics. Molecular dynamics is frequently used in the study of proteins, biomolecules, as well as in atmospheric and material science.

2.8.1 Equations of Motion

The molecular dynamics simulation method is based on Newton's second law, the equation of motion, $F = ma$, where F is the force exerted on the particle, m is its mass and a is its acceleration. The force on the atom is

$$F = -\frac{\partial V}{\partial r} \qquad (2.74)$$

Therefore, the acceleration of the atom is,

$$a = \frac{F}{m} \qquad (2.75)$$

The acceleration of atoms determines how the system evolves in time. The successive configurations of the system are generated by integrating Newton's law of motion. The essential data involved in this approach is that the integration is broken down into many small stages, each separated in time by a fixed time δt. The total force on each particle in the configuration at a time t is calculated as the vector sum of the interactions with other particles. The force thus calculated is used to determine the accelerations of the particles, which are then combined with the positions and velocities at a time t to calculate the positions and velocities at a time $t+\delta t$. The force is assumed to be constant during the time step. The forces on this particles in their new positions are then

determined in a similar way, leading to new positions and velocities at time $t+\delta t$, and so on. The Verlet algorithm and leap-frog algorithm are the algorithms used mainly for integrating the equations of motion in molecular dynamics simulations.

The verlet algorithm uses the positions and accelerations at time t, and the positions from the previous step $r(t-\delta t)$, to calculate the new positions $r(t+\delta t)$. The following taylor expansions can be written for $r(t+\delta t)$ and $r(t-\delta t)$

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \ldots\ldots\ldots \quad (2.76)$$

and

$$r(t-\delta t) = r(t) - v(t)\delta t + \frac{1}{2}a(t)\delta t^2 - \ldots\ldots \quad (2.77)$$

where $v(t)$ is the velocity and $a(t)$ is the acceleration. Summing the two equations gives,

$$r(t+\delta t) = 2r(t) + a(t)\delta t^2 - r(t-\delta t)\ldots. \quad (2.78)$$

Eqn 2.78 is the basic equation of the verlet integrator.[50] It can be seen that in order to obtain the next set of coordinates, the current coordinates, current accelerations and the previous coordinates required. Accelerations are calculated using 2.79 and the force F can be acquired by taking the negative of the force-field potential energy with respect to the coordinates (Eqn. 2.78). In the leap-frog algorithm,[51] the next position is determined both by the current position, and the velocity at the next half-time interval. The new position at time $t+\delta t$ is equal to the old position at t plus the average velocity in the interval t to $t+\delta t$ multiplied by the time step. However, we note that the acceleration is locally constant, and so the average velocity in the interval t to $t+\delta t$ must be equal to the velocity midway in the interval at time $t+\delta t/2$. Therefore,

$$r(t+\delta t) = r(t) + v(t+\delta t/2)\delta t \quad (2.79)$$

By comparing Eqn. 2.76, the velocity at time $t+\delta t$ is given by

$$v\left(t + \delta t / 2\right)\delta t = v\left(t\right) + \frac{1}{2} a\left(t\right)\delta t \qquad (2.80)$$

Reversing the time interval δt to - δt,

$$v\left(t - \delta t / 2\right)\delta t = v\left(t\right) - \frac{1}{2} a\left(t\right)\delta t \qquad (2.81)$$

Substracting Eqn. 2.81 from Eqn. 2.80 gives the velocity at $t+\delta t$,

$$v\left(t + \delta t / 2\right)\delta t = v\left(t - \delta t / 2\right) + a\left(t\right)\delta t \qquad (2.82)$$

By using this method, the velocity at time $(t - \delta t / 2)$ leap-frog the coordinates at time t to give the new velocities at time $(t + \delta t / 2)$, according to Eqn. 2.82. Then the coordinates at time t leap-frog the new velocities at time $(t + \delta t / 2)$ to give the new coordinates at time t, according to Eqn. 2.79. The main advantage of leap-frog algorithm is that the numerical inaccuracy is reduced because it uses differences between smaller quantities.

2.8.2 Constant Temperature Dynamics

Molecular dynamics simulations at constant temperature are useful for studying the behaviour of systems at different temperatures. A molecular dynamics simulation will conserve the total energy of the system, providing no work is done on the system. However, reactions occur at constant temperature, and so the use of a thermostat is required to keep the temperature constant through during simulations. The kinetic energy of the system is related to the temperature T according to the equipartion theory,

$$\sum_{i=1}^{N} \frac{1}{2} m_i v_i^2 = \frac{3N - N_c}{2} k_B T \qquad (2.83)$$

where N is the total number of atoms in the system, N_c is the number of constraints applied to the system and k_B is the Boltzmann constant. The Langevin thermostat was used in our simulations. The Langevin thermostat follows the Langevin equations of motion instead of Newton's equations of motion. In the Langevin equation of motion,[52-54] a frictional force added to the conservative force acting on the particle is proportional to the velocity, and it adjusts the kinetic energy of the particle so that the temperature matches the set temperature.

$$ma = -\zeta v + f(r) + f' \qquad (2.84)$$

where m is the mass of a particle, a is the acceleration, $f(r)$ is the conservative force acting on the particle, v is the velocity of the particle, ζ is a frictional constant, and f' is a random force. The frictional force $-\zeta v$ decreases the temperature because ζ is a fixed positive value. The random force is determined because ζ is a fixed positive value. The random force is randomly determined from a Gaussian distribution to add kinetic energy to the particle, and its variance is the function of set temperature and time step. Therefore, the random force is balanced with the frictional force and maintains the system temperature at the set value. A simple leap-frog integrator is used to propogate the dynamics. The Berendsen thermostat[55] and the Andersen thermostat[56] are the other temperature regulation schemes available in AMBER.

2.8.3 Constant Volume Dynamics

It is often necessary to simulate a chemical or biological molecule surrounded by solvent molecules to investigate how the system behaves in its native environment. Therefore, biomolecular simulations often carried out in explicit solvent to mimic real conditions. To prevent the solvent molecules from evaporating away, periodic boundary conditions are applied. The system is placed in rectangular box of solvent

molecules, called the primary box. The primary box is then repeated an infinite number of times in each of the three directions, so each atom in the primary box have an image of itself in an equivalent position in each of the other boxes. This means that if an atom moves out of the primary box during simulation, then it is replaced with the image of this atom that moves into the primary box from the opposite side so that the number of molecules in the box is always stays constant. An example of periodic boundary box in two dimensions is shown in Fig. 2.1. The primary box is shown in grey. If atom A moves out of the primary box then it is replaced with its image A' that moves into the primary box from the opposite side.

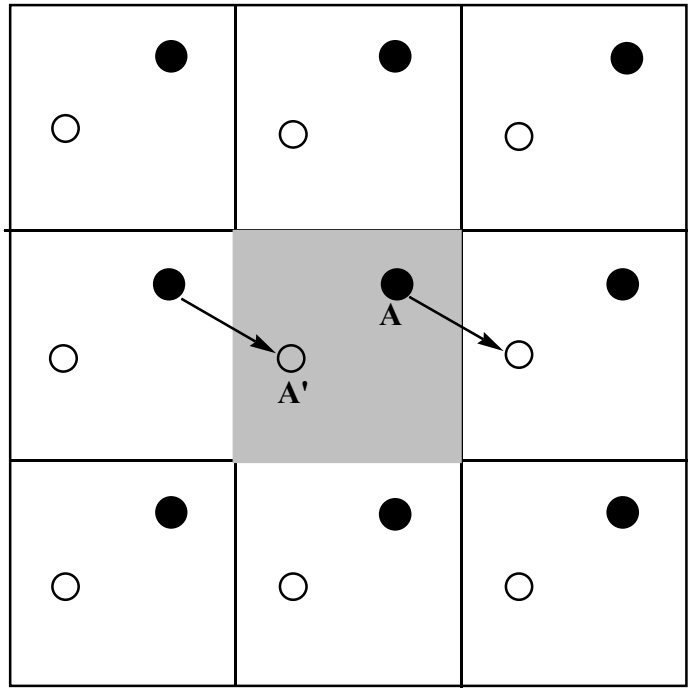


Figure 2.1 Periodic boundary conditions shown in two dimensions.

Nonbonded interactions such as van der Waals and the electrostatic interactions in principle extend to infinite separation. The evaluation of the interacting pairs remains a very challenging problem as there are an infinite number of boxes. This can be solved by choosing a reasonable cut-off for the nonbonded interactions. However, electrostatic interactions are long range, and may extend beyond the side of the potential cut off. The

Particle-Mesh Ewald method (PME)[57] is a common method used in MD packages to approximate the electrostatic interactions.

2.8.4 Particle-Mesh Ewald

The Particle-Mesh Ewald method is based on the original Ewald summation method devised by Ewald[58] in 1921 to study the energy of ionic crystals. In this method, a particle interacts with all other particles in the simulation box, and with all their images in the other boxes generated by the periodic boundary conditions. The charge-charge contribution to the potential energy of interaction in the central box which contains N particles can be written as

$$U = \frac{1}{2}\sum_{ij}^{N} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad (2.85)$$

where r_{ij} is the distance between the two particles i and j and q_i and q_j are their charges. For a box at a distance nL from the central box, where L is the length of the box, the contribution to the potential energy from the images of that box is

$$U = \frac{1}{2}\sum_{n}\sum_{ij}^{N} \frac{q_i q_j}{4\pi\varepsilon_0 \left| r_{ij} + nL \right|} \qquad (2.86)$$

Eqn. 2.86 calculates contribution from interactions in the central box as well as all its images. The sum in Eqn. 2.86 is a slow converging sum because the terms form a divergent series which does not have a finite sum. The principle of Ewald sum is to convert the summation into two series, each of which converges more rapidly. This is done by considering each charge to be surrounded by a neutralising charge distribution of equal magnitude but of opposite sign. Initial set of charges are surrounded by a Gaussian charge distribution (calculated in real space) to which is a cancelling charge distribution must be added (calculated in reciprocal space). The following final expression for the long range forces is used in the Ewald summation method,

$$U = \frac{1}{2}\sum_{ij}\left\{\begin{array}{c} \sum_{n=0}^{\infty}\frac{q_i q_j}{4\pi\varepsilon_0}\frac{erfc\left(\alpha\left|r_{ij}+nL\right|\right)}{\left|r_{ij}+nL\right|} \\ +\sum_{\mathrm{k}\neq 0}\frac{1}{\pi L^3}\frac{q_i q_j}{4\pi\varepsilon_0}\frac{4\pi^2}{k^2}\exp\left(-\frac{k^2}{4\alpha^2}\right)\cos\left(\mathrm{k}.r_{ij}\right), \\ -\frac{\alpha}{\sqrt{\pi}}\sum_{k=1}^{N}\frac{q_k^2}{4\pi\varepsilon_0}+\frac{2\pi}{3L^3}\left|\sum_{k=1}^{N}\frac{q_k}{4\pi\varepsilon_0}r_k\right|^2 \end{array}\right. \qquad (2.87)$$

where erfc is the complementary error function. The first term in the summation is performed in 'real space'. The contribution from the reciprocal charge distribution is given by the second term in Eqn. 2.87. The vectors **k** are called reciprocal vectors. The sum of Gaussian functions in real space includes the interaction of each Gaussian with itself and the third term corrects for this self-interaction. The fourth correction term may also be required, depending upon the medium that surrounds the sphere of simulation boxes. The first two terms are fast converging sums, and the last two terms are constant correction terms. This summation is computationally expensive, but using the Fast Fourier Transform to calculate the second term can make things considerably faster. This allowed the Ewald method to be used for studying highly polar and charged systems. There are many approaches which use this same basic idea of Ewald, one of which is the Particle-Mesh Ewald (PME) method.[57] The Particle-Mesh Ewald method considers gridding the charge density so that the nearest 27 points in three dimensions are used to calculate the potential due to the Gaussians at these points and therefore giving the potential and forces on each particle.

2.8.5 Constant Pressure Dynamics

Molecular dynamics simulations must be performed under conditions of constant pressure to mimic experimental conditions. Similar to keeping constant temperature during simulations, we use a pressure barostat for constant pressure dynamics. In constant pressure dynamics, the volume of the unit cell is adjusted by small amounts on each step to make the computed pressure approach the target pressure, usually atmospheric pressure. The thermodynamic term isothermal compressibility, κ,

measures how the volume V fluctuates as a pressure P change is applied according to the relation,

$$\kappa = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T \tag{2.88}$$

For two successive time steps at constant temperature, this compressibility parameter can be written as

$$\kappa = -\frac{1}{V}\frac{\Delta V}{\Delta P} = -\frac{1}{V}\frac{(\lambda V - V)}{(P_{new} - P_{curr})} \tag{2.89}$$

The pressure change can be defined differently when a pressure bath is applied,

$$P_{new} - P_{curr} = \frac{\Delta t}{\tau_P}(P_{bath} - P_{curr}) \tag{2.90}$$

where τ_P is the pressure coupling parameter.

Substituting Eqn. 2.90 in Eqn. 2.89 and subsequent rearrangement gives the volume scaling factor λ as

$$\lambda = 1 - \frac{\kappa \Delta t}{\tau_P}(P_{bath} - P_{curr}) \tag{2.91}$$

The simulation box is then scaled by a factor λ between two successive time steps, or equivalently all atomic Cartesian coordinates are scaled by a factor $\lambda^{1/3}$. The isothermal compressibility κ is present in the expression of the scaling factor λ but the exact value is not critical to the dynamics because the parameter τ_P is also present in the scaling factor and is usually determined by the user. The Anderson barostat[56] and the Parinello-Rahman barostat[59] are two methods to keep pressure constant in MD simulations.

2.8.6 Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) method

Generalized Born surface area is simply Generalized Born model including hydrophobic solvent accessible surface area SA term. The MM-GBSA approach is a postprocessing method which is implemented in the AMBER package, to evaluate free energies of binding or to calculate the absolute free energies of molecules in solution. The MM-GBSA[60] method employs molecular mechanics (MM), the generalized Born model (GB)[61-63] and solvent accessibily surface area (SA) to calculate the free energy terms. The binding free energy for the protein-ligand complex formation is given by

$$\Delta G_{bind} = \langle G_{complex} \rangle - \left(\langle G_{protein} \rangle + \langle G_{ligand} \rangle \right) \qquad (2.92)$$

$\langle G_{complex} \rangle$, $\langle G_{protein} \rangle$ and $\langle G_{ligand} \rangle$ are the free energies of the complex, protein and the ligand, respectively. The MM-GBSA approach uses a single molecular dynamics trajectory method, in which the coordinates of the complex, receptor (protein) and ligand are taken from the same trajectory. The free energy term includes contributions from gas phase, solvation and entropic effects.

$$G = E_{gas} + G_{sol} - TS \qquad (2.93)$$

E_{gas} is the gas phase energy given by the AMBER force field (MM energy) and is the sum of the internal energy (includes bond, angle and torsion energies), van der Waals and electrostatic energies.

$$E_{gas} = E_{\text{int}} + E_{vdw} + E_{ele} \qquad (2.94)$$

The total solvation free energy term include contributions from electrostatic and non-electrostatic terms.

$$G_{sol} = G_{ele} + G_{nonelec} \tag{2.95}$$

The electrostatic contribution to the solvation free energy is calculated by using the Generalized Born solvation model (GB). The GB model approximates G_{ele} by an analytical formula.[61, 64]

$$G_{ele} \approx G_{GB} = -\frac{1}{2}\sum_{ij}\frac{q_i q_j}{f_{GB}(r_{ij}, R_i, R_j)}\left(1 - \frac{e^{-\kappa f GB_{ij}}}{\varepsilon_w}\right) \tag{2.96}$$

where r_{ij} is the distance between atoms i and j, R_i and R_j are the effective Born radii of the atoms, and f_{GB} is a smooth function, of which a common form is:

$$f_{GB} = \left[r_{ij}^2 + R_i R_j \exp\left(-r_{ij}^2 \Big/ 4R_i R_j\right)\right]^{1/2} \tag{2.97}$$

The electrostatic screening effects of (monovalent) salt are incorporated via a Debye-Huckel screening parameter κ /Å$^{-1} \approx 0.316\sqrt{[salt]}$.[64] The function f_{GB} is designed to interpolate between the limit $r_{ij} \to 0$ when atomic spheres merge into one, and the opposite extreme $r_{ij} \to \infty$ when the ions can be treated as point charges obeying Coulomb's law. There are a number of modified GB models are available in AMBER.

The nonpolar contribution to the solvation free energy is given by

$$G_{nonelec} = G_{SA} = \gamma \times SASA + b \tag{2.98}$$

$G_{nonelec}$ is proportional to the total solvent-accessible surface area (SASA), with a proportionality constant derived from experimental solvation energies of small non-polar molecules.

2.8.7 Force field

The success of molecular dynamics simulations depend on the expression for the potential energy as a function of the atomic coordinates, known as the force field. Force field methods calculate the energy of a system as a function of nuclear positions only and they do not take any electronic effects into consideration. They are also called molecular mechanics methods (MM), and there are many assumptions that have to be made in order for these methods to be practical, taking into account that they do not deal with electrons in the system. Classical mechanical force field methods are made up of two parts, an analytical part describing the potential energy of the system in terms of the position of the nuclear coordinates and a set of empirical parameters. Most molecular mechanics force fields today have their parameters obtained from high-level quantum mechanical calculations. We have used AMBER[65] 2009 version for MD simulations. AMBER is designed to work with several simple types of force fields, although it is most commonly used with parameterizations developed by Peter Kollman and his co-workers. We have used a number ff03[66] and GAFF[67] force fields in our MD simulations.

The force field consists of bonded and nonbonded interactions. The total energy of the system given by a force field can be written as,

$$E_{Total} = E_{Bond} + E_{Angle} + E_{Dihedral} + E_{Elec} + E_{vdW} \qquad (2.99)$$

where E_{Bond} and E_{Angle} are the bond stretching and angle bending terms described by simple harmonic functions and $E_{Dihedral}$ is the torsional term. E_{Elec} is calculated by Coulomb electrostatic term and E_{vdW} is the van der Waals term described by a Lennard-Jones potential. The first three terms of r.h.s of Eqn. 2.99 collectively represent the bonded interactions and the final two terms describe the nonbonded interactions. The above equation can be expanded using the individual expression for different terms:

$$E_{Total} = \sum^{Bond} \frac{1}{2} K_r (r - r_0)^2 + \sum^{Angle} \frac{1}{2} k_\theta (\theta - \theta_0)^2 + \sum^{Dihedral} \frac{1}{2} V_n [1 + \cos(n\varphi - \varphi_0)] + \sum_{ij}^{Elec} \frac{q_i q}{4\pi\varepsilon_0 r_{ij}} + \sum_{ij}^{vdW} \frac{A_i}{r_{ij}^{12}} - \frac{B_{ij}}{r_i^6} \quad (3.00)$$

AMBER uses a simple harmonic potential for bond stretching energy. k_r is the force constant for bond stretching in kcal $mol^{-1} rad^{-2}$, θ is the angle formed between the three atoms and θ_0 is the equilibrium value of θ in radians. Torsion term represents the four body interactions involving atoms A, B, C and D where A is bonded to atom B. B is bonded to atom C and C is bonded to atom D. The dihedral φ is defined to be the angle between the planes ABC and BCD, and it is a measure of the torsion about the bond BC. The parameter *n* is called the periodicity and is the number of times the cosine potential passes through a minimum as the angle φ is turned through a complete revolution. V_n is a parameter that gives an indication to the barrier height of rotation and φ_0 is the phase angle that controls where the minima of the potential lies. This same formula is used to calculate the interactions for improper torsions and occurs when atoms A, B and D are all bonded to atom C, but not to each other. The improper dihedral interaction determines the interaction m between the planes ABC and BCD.

The electrostatic interactions are evaluated from the Columbic potential. q_i and q_j are the atomic charges on the atoms *i* and *j*, ε_0 is the dielectric constant and r_{ij} is the interatomic distance between atoms *i* and *j*. The term E_{vdW} is the van der Waals energy describing the repulsion or attraction between atoms that are not directly bonded. Together with the electrostatic term E_{Elec} they describe the non-bonded energy. E_{vdW} may be interpreted as the part of the interaction, which is not related to electrostatic energy due to charges. At large interatomic distances, E_{vdW} asymptotically approaches to zero, and for small distances, it becomes very repulsive. In quantum mechanical term this is due to the overlap of the electron clouds of the two atoms; the electrons repel each other as they are negatively charged. At intermediate distances, however there is slight attraction between two such electron clouds due to electron correlation phenomenon. The motion of electrons may create a slightly uneven distribution at given time and this instantaneous polarization will induce a charge polarization in the

neighbouring molecule, creating an attraction. It can be derived theoretically that the attraction varies as the inverse sixth power of the distance between the two atoms. Actually the induced dipole-dipole interaction is only one of such terms: there are also contributions from induced quadrupole-dipole, quadrupole-quadrupole etc. interactions. These vary as R^{-8}, R^{-10} etc. Several potential functions can be used to describe van der Waals interactions, but the one most commonly used is the Lennard-Jones 6-12 potential,[68] which has the form,

$$E_{vdW} = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} = 4\varepsilon_{ij}\left[(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{uj}})^{6} \right] \qquad (3.01)$$

where A_{ij} and B_{ij} are parameters describing the attractive and repulsive parts of the potential, respectively. σ_{ij} is the equilibrium bond distance between the two atoms for the van der Waals interaction and ε_{ij} is the binding energy of the van der Waals interaction. Each atom has its own van der Waals radius σ_i and van der Waals energy ε_i. For a pair of atoms *i* and *j*, these parameters can be combined to give the interaction parameters σ_{ij} and ε_{ij}. The AMBER program uses the Lorentz-Berthelot mixing rules for the parameters σ_{ij} and ε_{ij}. A geometric mean rule (the Berthelot rule) is applied for the well depth ε parameter and an arithmetic mean rule (the Lorentz rule) is used for σ. The following equations show the mathematical formulae for these rules for two atoms *i* and *j*,

$$\varepsilon_{ij} = \left(\varepsilon_{ii}\varepsilon_{jj}\right)^{1/2} \qquad (3.02)$$

$$\sigma_{ij} = \frac{1}{2}\left(\sigma_{ii} + \sigma_{jj}\right) \qquad (3.03)$$

2.9 References

1. Schrödinger, E. *Ann. Physik*, **1926**, *79*, 361.

2. Jensen, F. *Introduction to Computational Chemistry*, (Wiley, **1999**).

3. Foresman, B. J.; Frisch, A. *Exploring Chemistry with Electronic Structure Methods*, Second Edition, (Gaussian Inc. **1996**).

4. Leach, A. R. *Molecular Modelling, Principles and Applications*, (Longmann, **1996**).

5. Atkins, P. W.; Friedman, R. S. *Molecular Quantum Mechanics*, Third Edition, (Oxford, **1997**).

6. Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry, Introduction to Advanced Electronic Structure Theory*, (Dover, **1996**).

7. Levine, I. N. *Quantum Chemistry*, Fourth Edition, (Prentice-Hall, **1991**).

8. Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry*; Dover: New York, **1996**.

9. Pauli, W. *Phys. Rev.* **1940**, *58*, 719.

10. Slater, J. *Phys. Rev.*, **1929**, *34*, 1293.

11.Fock, V. *Z. Physik* **1930**, *62*, 795.

12.Roothaan, C.C.J. *Rev. Mod. Phys.*, **1951**, *23*, 69.

13.Pople, J.A. *Ann. Rev. Phys. Chem*., **1959**, *10*, 331.

14. McWeeny, R.; Dierksen, G. *J. Chem. Phys.*, **1968**, *49*, 4852.

15.Slater, J.C. *Phys. Rev.*, **1930**, *36*, 57.

16.Boys, S.F. *Proc. Roy. Soc. (London) A*, **1950**, *200*, 542.

17.Binkley, J.S.; Pople, J.A.; Hehre, W.J. *J. Am. Chem. Soc..*, **1980**, *102*, 939.

18.Hehre, W.J.; Ditchfield, R.; Pople, J.A. *J. Chem. Phys.*, **1972**, *56*, 2257.

19.Frisch, M.J.; Pople, J.A.; Binkley, J.S. *J. Chem. Phys.*, **1984**, *80*, 3265.

20.Block, F. *Z. Physik*, **1929**, *57*, 545.

21.Dirac, P.A.M. *Proc. Cambridge Phil. Soc.*, **1930**, *26*, 376.

22.Hohenberg, P.; Kohn, W. *Phys. Rev.*, **1964**, *136*, B864.

23.Szilagyi, R.K.; Metz, M.; Solomon, E.I. *J. Phys. Chem. A,* **2002**, *106*, 2994.

24. Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, 136, B864.

25. Vosko, S. H.; Wilk, L.; Nusair, M. Can. *J. Phys. Chem.*, **1980**, 58, 1200.

26. Becke, A. D. *Phys. Rev. A*., **1988**, 38, 3098.

27. Perdew, J. P.; Wang, Y. *Phys. Rev. B*., **1992**, 45, 13244.

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

29. Perdew, J. P. *Phys. Rev. B*., **1986**, 33, 8822.

30. Perdew, J. P. *Phys. Rev. B.*, **1986**, 34, 7046.

31.Becke, A. D. *J. Chem. Phys.*, **1993**, 98, 5648.

32. Stephens, P. J.; Devlin, F. J.; Chablowski, C. F.; Frisch, M. J. *J. Phys. Chem.,***1994**, *98*, 11623.

33. Becke, A. D. *J. Chem. Phys.*, **1993**, *98*, 1372.

34. C. Møller and M. S. Plesset, *Phys. Rev.*, **1934,** *46*, 618.

35. J. A. Pople, R. Seeger and R. Krishnan, *Int. J. Quant. Chem. Symp.*, **1977**, *11*, 149.

36.R. Krishnan, M. J. Frisch and J. A. Pople, *J. Chem. Phys.*, **1980**, *72*, 4244.

37. K. Raghavachari, J. A. Pople, E. S. Replogle and M. Head-Gordon, *J. Phys. Chem.*, **1990**, *94*, 5579.

38. Cramer, C. J. *Essentials of Computational Chemistry*; Wiley: Chichester, **2003**.

39. Pople, J.A.; Beveridge, D.L.; Dobosh, P.A. *J. Chem. Phys.*, **1967**, *47*, 2026.

40. Dewar, M.J.S.; Thiel, W. *Theor. Chim. Acta*, **1977**, *46*, 89.

41. Dewar, M.J.S.; Thiel, W. *J. Am. Chem. Soc.*, **1977**, *99*, 4899.

42. Dewar, M.J.S.; Zoebisch, E.G.; Healy, E.D.; Stewart, J.J.P. *J. Am. Chem. Soc.*, **1985**, *107*, 3902.

43. Grimme, S. *Journal of Computational Chemistry.,* **2004**, *25*, 1463.

44. Dunning, T. H. *The Journal of Chemical Physics.,* **1989**, *90*, 1007.

45. Mohr, M.; McNamara, J. P.; Wang, H.; Rajeev, S. A.; Ge, J.; Morgado, C. A.; Hillier, I. H. *Faraday Discussions.,* **2003**, *124*, 413.

46. Ditchfield, R. *Molecular Physics: An International Journal at the Interface Between Chemistry and Physics.,* **1974**, *27*, 789.

47. London, F. J. *Phys. Radium.,* **1937**, *8*, 397.

48. Wolinski, K.; Hinton, J. F.; Pulay, P. *Journal of the American Chemical Society*., **2002**, *112*, 8251.

49. Karplus, M.; Petsko, G. A. *Nature*., **1990**, *347,* 631.

50. Verlet, L. *Physical Review.,* **1967**, *159*, 98.

51. Hockney, R. W. *Methods in Computational Physics*., **1970**, *9*, 136.

52. He, S.; Scheraga, H. A. *The Journal of Chemical Physics*., **1998**, *108*, 271.

53. Izaguirre, J. A.; Catarello, D. P.; Wozniak, J. M.; Skeel, R. D. *The Journal of Chemical Physics*., **2001**, *114*, 2090.

54. Pastor, R. W.; Brooks, B. R.; Szabo, A. *Molecular Physics*., **1988**, *65*, 1409.

55. Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. V.; DiNola, A.; Haak, J. R. *The Journal of Chemical Physics*., **1984**, *81*, 3684.

56. Andersen, H. C. *The Journal of Chemical Physics*., **1980**, *72*, 2384.

57. Darden, T.; York, D.; Pedersen, L. *The Journal of Chemical Physics*., **1993**, *98*, 10089.

58. Ewald, P. P. *Annalen der Physik.,* **1921**, *369*, 253.

59.Parrinello, M.; Rahman, A. *Journal of Applied Physics*., **1981**, *52*, 7182.

60. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. *Accounts of Chemical Research*., **2000**, *33*, 889.

61. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *Journal of the American Chemical Society*., **2002**, *112*, 6127.

62. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chemical Physics Letters*., **1995**, *246*, 122.

63. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *The Journal of Physical Chemistry*., **1996**, *100*, 19824.

64. Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. *Theoretica Chimica Acta*., **1999**, *101*, 426.

65. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatam, T. E.; DeBolt, S.; Ferguson, D.; Siebel, G.; Kollman, P. *Computer Physics Communications*., **1995**, *91*, 1.

66. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *Journal of Computational Chemistry.,* **2003**, *24*, 1999.

67. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *Journal of Computational Chemistry*., **2004**, *25*, 1157.

68. Jones, J. E. *Proceedings of the. Royal Society London A*., **1924**, *106*, 463.

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

70. Bingham, R. C.; Dewar, M. J. S.; and Lo, D. H. *J. Am. Chem. Soc.*,**1975**,*97*, 1285.

71. Stewart, J. J. P. *J. Comput. Chem.*, **1989**, *10*, 209.

Chapter 3

Computational investigation of role of CH/π and OH/π interactions in carbohydrate-aromatic complexes

3.0 Abstract

Herein we have carried out a quantum mechanics study aimed at the investigation of CH/π and OH/π interactions in carbohydrate-aromatic complexes. The calculations have been performed, employing the DFT-D method using BLYP-D/TZV (2d, 2p) and semi-empirical methods PM3-D and PM3-D*, methods implemented with dispersion correction term. Glucose molecules can interact with aromatic amino acids in different orientations. The complexes chosen can be classified as lower-face complexes, upper-face complexes and side-face complexes based on the orientation. The calculations have been carried out to calculate the interaction energies between carbohydrate glucose and aromatic amino acids such as tyrosine and tryptophan modelled as 3-methyl indole and p-hydroxy toluene respectively. The dispersion contribution to the interaction energy seems to be main stabilization factor for carbohydrate-aromatic complexes. The semi-empirical method PM3-D, which has proven to be successful method for studying stacked complexes, overestimates the interaction energies of carbohydrate-aromatic complexes. Whereas PM3-D* method which includes modified core-core repulsion term seems to be performing better for carbohydrate-aromatic complexes. Among three methods, DFT-D performs better and producing interaction energies and inter-ring separation reasonably. Interaction energies of glucose with tryptophan are ranging from 9-10 kcal mol^{-1}. Interaction energies of glucose with tyrosine are ~2 kcal mol^{-1} less than interaction energies with tryptophan. The results obtained in the calculations show the importance of contributions from CH/π, NH/π and

OH/π interactions in stabilizing the carbohydrate-aromatic complexes. To support the fact the presence of CH/π and OH/π interactions, NMR and IR frequency shifts have been calculated using BLYP-D/TZV (2d, 2p). The large blue shifts have been observed for C-H bonds involving CH/π interaction. Likewise, red shifts have been observed for O-H bonds involving OH/π interaction. NMR chemical shifts support the fact that C-H bond is involving in carbohydrate-aromatic interaction.

3.1 Introduction

Carbohydrates play key roles in various biological and molecular recognition processes in bacterial and viral infections, cell recognition, cell adhesion in metastatis and inflammation, differentiation, development, regulation, intercellular communication, growth, fertilization and signal transduction events. Carbohydrates can serve as fuel and building materials. Carbohydrates, amphipathic molecules, interact with biomolecules through hydrogen bonds using hydroxyl groups as well as through nonpolar forces. Carbohydrates can interact with different proteins, enzymes and antibodies. Carbohydrates are also known to mediate the infection of cells by pathogens, the distribution and reactivity of protein within cells, and many aspects of immune response. The centre of all biological processes involving carbohydrates is weak noncovalent interactions such as hydrogen bonds, π-interactions, dispersive interactions and stacking interactions. Though there are many potential opportunities for application of carbohydrates in pharamaceutics, the development is slow because of problems in understanding mechanism of carbohydrates in biological processes. Therefore it is necessary to understand the mechanism of carbohydrate-protein recognition to tackle the problems in carbohydrate-based drug development. In order to understand the mechanism of biological processes involving carbohydrates, structural aspects of carbohydrates have to be revealed. Since carbohydrates are complex structures with different functionalities towards surroundings and the principles of sacchardies are not well understood, the process of designing and synthesizing receptors is a challenging task.

Site-directed mutagenesis of a protein molecule and specific modification of carbohydrate ligand have been used to understand the interactions between carbohydrates and proteins. However, a small structural alternation in carbohydrate-complex can lead to different interpretations due to the combined interaction arising from hydrogen bonds and van der Waals forces which will alter the binding mode of the ligand. This technical dilemma has been overcome by using affinity labelling combined with site-directed mutagenesis.

Lectins, sugar-binding proteins which are found in viruses, plants and human, bind to carbohydrates with specificity and play an important role in biological events. Lectins such as phytoagglutinins, phytohemaglutinins and hemagglutinins are derived from plants and characterized well with regard to their chemical structure and carbohydrate-binding properties. Although functions of lectins in plant tissues are of interest, it is their unique and interesting chemical and biological properties make them extremely useful for the study of host-guest interactions in biological processes. Lectin-carbohydrate interactions are of most interest among the various types of interactions occur at cell surface and these lectin-carbohydrate interactions can serve as a tool to study the process of specific cell recognition and adhesion.[1] Lectins can be classified as groups according to their specificities to carbohydrates such as Gal- specific and GalNAc- specific lectins.

X-ray crystallograpahic and site-directed mutagenesis experiments have shown the presence of aromatic amino acids in the binding sites of carbohydrates.[2- 6] It has been shown that the binding site of α-amylase isozyme (AMI) contains Trp278 and Trp279. It has also been shown that upon mutation of Trp279 by Alanine, binding affinity is reduced ten- and three-fold for starch granules and β-cyclodextrin respectively.[6] Studies on carbohydrate-binding module (CBM29)[7] from piromyces equi cellulose/hemicellulase complex, have showed the presence of aromatic amino acids in binding site. It has been observed that all pyranosides are in 4C1 chair conformation. These studies show that complex of CBM29-2 and cellohexaose contains three aromatic amino acids Trp24, Trp26 and Tyr46 forming the hydrophobic platform for cellohexaose to interact. It has also been shown that these aromatic amino acids Trp24,

Trp26 and Tyr46 interact with the α-face of the glucosyl moities of residues 6, 4 and 2 respectively. This clearly shows the interaction pattern as CH1, CH3 and CH5 bonds of glucose moities point the π-electron cloud of the aromatic ring. From crystal structure data analysis, it has been observed that stacking of galactose with aromatic amino acid is a common feature among galactose-specific proteins.[8] Mutation of His by Trp in asialoglycoprotein receptor increases the affinity for galactose. [9] In another site-directed mutagenesis study, it has been found that because of replacement of Tyr248 of ricin by His257 in site 2 of polygonatum multiflorum RIPm, the hydrophobic interaction between galactose and Tyr248 can no longer occur.[10] Similar finding has been observed by (Author) that upon mutation of Tyr248 by His248 in the B chain drastically reduces the galactose-binding activity.[11] These results show the necessity of Tyr248 for the galactose-binding site 2 of ricin.[12] These results also indicate the inactivity of PMRIPm towards Gal or GalNAc.

The Dolichos biflorus seed lectin (DBL) has high affinity for GalNAc over Gal due to the hydrophobic interaction between aromatic amino acids Trp132 and Tyr104 and acetyl group of GalNAc. DBL has low affinity for Gal since the lost of hydrophobic interaction due to the replacement of aromatic amino acid by aliphatic amino acid.[13] Studies on X6b from the clostridium thermocellum xylanase Xyn10B suggest the binding motif as shallow cleft which accommodate upto four xylose sugar moities binding to three aromatic amino acids Trp53, Tyr103 and Tyr134.[14] Analyses of the three-dimensional structures of several carbohydrate-binding proteins by Rini (1995) show the presence of hydrophobic stacking interaction between aromatic amino acids and apolar side of the galactose in galactose-binding lectins.[15] In 3D structure of galactose mutarotase from lactococcus lactis obtained at 1.9 Å resolution, though hydrogen bonds between Arg71, His96, His170, Asp243 and Glu304 and galactose moities play the main role for sugar-protein binding, the presence of aromatic amino acids Tyr172, Phe279 and Phe283 have also been found at the bottom of the binding cleft which implies the presence of hydrophobic stacking interaction between aromatic amino acids and sugar.[16] Crystallographical studies[17] on cyclodextrin glycosyl transferase (CGTase) suggest the importance of carbohydrate-aromatic interaction for the activity of CGTase. Replacement of Tyr100 by leucine significantly decreases the

activity of CGTase. Mutations of F183L/F259L result in losing of active centre at subsite +2. It has been shown that acarbose molecules which bound to Y100L mutant moved from the active centre toward the side chain of Tyr195, which results in lose of hydrogen bonding and hydrophobic interaction between acarbose and subsites.[17] The role of aromatic amino acids Trp22, Tyr24, Tyr8, Trp7 and Tyr12 were analysed by mutation with alanine in CBM10 from p.fluorescens. cellulosa. Xylanase A (Pf. Xyn 10A) with cellulose.[18] Trp22, Trp24 and Tyr8 were found to be important for the binding of Pf. Xyn10A to cellulose. Mutation of these aromatic amino acids by Alanine significantly reduces the binding affinity of protein to cellulose compared to wild-type protein. Mutation of Tyr12 had very little effect in binding affinity. Titration of mutants W22A, W24A and W7A with N-bromosuccinimide shows that aromatic amino acids Trp22 and Trp24 were located in the surface and Trp7 was buried. Although Trp7 is internal, it has been found that mutation of Trp7 by alanine greatly reduces the protein's capacity to bind to cellulose.

NMR spectroscopy is a useful tool to study carbohydrate-aromatic interactions. NMR techniques have been used along with computational techniques such as molecular simulations and ab initio calculations to study carbohydrate-aromatic interactions. Protons in the aromatic environment shift to upfield. Therefore, proton of C-H bond which involves in CH/π interaction with aromatic ring will move to upfield. This gives the information regarding binding motifs of carbohydrates and aromatic amino acid. Bernardi et al[19] have studied the interactions of carbohydrate-aromatic complexes to find the applications of these interactions to design functional mimics of the GM1 glycomimetics by using NMR and computational techniques. Studies on these functional mimics suggest the presence of stacking interaction between the phenyl ring and the GalNAc residue plays a main role in calculating the affinity of these mimics to cholera toxin (CT). In another study, Sophie et al[20] have investigated different monosaccharides such as α-, β- anomers of methylglucose, galactose, ribofuranose as well as α-methyl mannose and aromatic moieties such as phenol, phenyl alanine, tyrosine and tryptophan.

From these studies, it is apparent that aromatic amino acids are found nearby carbohydrates. Pyranose ring of carbohydrates form mainly two faces for the interaction with aromatic amino acids, i.e., polar face and apolar face. While hydroxyl groups form polar face, C-H bonds form apolar face of the sugar. In glucose, CH1, CH3 and CH5 form nonpolar face and these bonds interact with π-face of the aromatic amino acids through CH/π forces. Therefore possible mode of interaction between carbohydrates and aromatic amino acids mainly involve dispersion forces such as CH/π and sometimes through OH/π. Carbohydrate binding proteins could be classified into two types based on whole ligand-binding sites, i.e. buried or ligand-engulfing binding sites and Shallow ligand binding sites.[21] These two binding sites are determined by the type of aminoacids nearby carbohydrates. Despite of this binding site morphology, there are two forces governing the carbohydrate-protein recognition.

1. hydrogen bond between hydroxyl groups of carbohydrates and polar group of proteins
2. less polar force van der Waals interaction between C-H groups of carbohydrates and aromatic moieties of the amino acids[21]

Aromatic aminoacids bind to carbohydrates in shallow binding mode mainly through noncovalent interactions. The noncovalent interactions between carbohydrate and proteins may involve contributions from electrostatic forces, hydrogen bonding and dispersion forces CH/π and OH/π.

3.2.1 Structural studies

Sugar-binding proteins were discovered more than 100 years ago in plants and research has been going on for many years on their properties and functions.[22] Although conconavalin A, sugar-binding protein has been found to be first sugar binding protein in 1932 by Sumner&Howell[22], structural knowledge of carbohydrate-protein has been derived only in early 1990s by Quiocho et al[23]. X-ray structures of liganded form of L-Arabinose binding protein (1.7Å) and D-galactose binding protein (1.9 Å) have been obtained by Quicho et al.[23] Stablizing factors between protein and carbohydrate interactions have been found to be hydrogen bonds, van der Waals contacts involving

all the atoms of carbohydrates and stacking between carbohydrate and aromatic moiety of amino acids. It has been found that polar groups of sugar in the solvent inaccessible surface cleft enable the stacking to occur. This fresh insight of stacking interaction between aromatic amino acid and carbohydrate led many research to understand carbohydrate-protein recognition.[23] Spurlino et al, have obtained two mutant structures of maltodextrin binding protein (MdBP)-maltose complexes at 1.9 Å resolutions. The two mutants are obtained by replacing Trp230 and Trp232 by alanine. The role of aromatic amino acid tryptophan has been studied. Although both mutants show minor structural changes, Trp230Ala mutant shows nearly 12-fold association constant lower than the wild-type protein. This shows the importance of aromatic amino acids, here tryptophan for carbohydrate-protein recognition.[24]

3.2.2 Experimental studies

The role of aromatic amino acid in carbohydrate-protein recognition has been studied mainly by site-directed mutagenesis, covalent modification and by calorimetric methods.[25] Site-directed mutagenesis combined with affinity labeling has been found to be a useful technique to understand hydrogen bonding and van der waals interactions between carbohydrate and protein.[21] Zolotnitsky et al,[26] have conducted isothermal titration calorimetry (ITC) experiments of xylosacchardies for two xylanases. ITC has been found to measure binding enthalpy which allows the determination of the binding constant, entropy, and stoichiometry. The change in the heat capacity of binding ΔC_p was used to obtain the information regarding stacking interactions. It has been found to be negative for xylosaccharides which supports the presence of hydrophobic stacking interaction.[26]

In another study, Zolotnitsky et al have carried out ITC measurements for XynE and xylotriose and have calculated of -158 $calmole^{-1}K^{-1}$ of ΔCp for XynE and xylotriose, suggesting the presence of stacking interactions between sugar and protein.[27]
Huber et al have studied the role of Trp999 for the action of β-galactosidases by site-directed mutagenesis and it was shown to be a key residue for shallow mode binding. It

has also been shown that affinity of this residue towards inhibitors decreased upon mutation.[28] Recently Sophie et al (2008)[29] have shown the presence of CH/π interactions through NMR experiments. Spurlino et al have studied maltodextrin binding protein (MdBP) by site-directed mutagenesis and aromatic amino acids have been found in ligand-binding sites.[24] Later, two mutants of the MdBP-maltose complexes have been obtained by replacing Trp with Ala and association constant have been found to be less than wild-type.

3.2.3 Computational studies

Although there have been many experimental studies on carbohydrate-protein interaction nearly for 100 years, theoretical studies have started to emerged only from 2004. Structures and interactions of carbohydrate-protein recognition have been explored by experimental methods.[30] However, it is only recently that the energetic aspects of carbohydrate-protein recognition have begun to be explored.[31, 32] Simulation techniques are useful tools to generate structures similar as experimental data and to study structural properties. Given that three-dimensional structure of the carbohydrate and receptor molecule, it is now possible to calculate and model the interactions between carbohydrates and other biological macromolecules by computational methods. Free-energy perturbation (FEP) simulation methods are suitable methods to compute both absolute and relative binding free energies for protein-protein interactions.[33] This method has also been applied to study biotin-streptavidin binding motifs and relative binding energies are found to be in agreement with experiment.[34]

JP Simons et al[35] have investigated the carbohydrate-aromatic interactions by using IR and UV spectroscopic techniques. They have computed IR spectrum of uncomplexed monosaccharides α-, β-D-methylglucose, α-, β-D-methylgalactose, α-, β-L-methylfucose and α-, β-L-fucose and compared their spectrum with experimental spectrum of corresponding complexed monosaccharides with toluene. Since IR vibrational frequencies are sensitive to the hydrogen-bonded environment of their OH groups, they can be used as signatures of the presence or absence of OH/π hydrogen

bonding interactions. It can be predicted from the difference between IR spectra of free and complexed carbohydrates whether OH bond is involved in OH/π interaction or not by calculating red-shifts of the OH stretching frequency. IR stretching frequencies will be shifted to lower wave numbers for OH bonds which involve OH/π interactions. Frequencies will be shifted to higher wave numbers for C-H bonds which involve CH/π interactions and known as blue-shifts. Simons et al[35] have found that among all the complexes explored, dispersion forces involving CH/π interactions predominate OH/π interactions though specific OH/π interactions have been identified.

Fernandez et al (2005)[36] have studied fucose and benzene complexes using the MP2 method and NMR experiments. It was shown that CH/π interactions play an important role in stabilizing these complexes. Interaction energy has been calculated as 3.0 kcal mol^{-1} of fucose-benzene complex. It has been estimated that each CH/π contributes approximately ~1 kcal mol^{-1} to interaction energy since three C-H bonds were pointed towards aromatic ring. Spiwok et al (2004)[37] have reported the interaction energies for Glc- and Gal- with 3-methylindole as -5.2 to -2.7 kcal mol^{-1} using MP2/6-31+G (d). Later, Spiwok et al (2006),[38] have reported the interaction energies as -2.8 to -12.3 kcal mol^{-1} using MP2/6-311+G (d). Interaction energies have been calculated for galactose and aromatic amino acids as ranging from -2.4 to 2.8 kcal mol^{-1} using UB3LYP/6-31G** by Sujatha et al (2004)[39] and have also been reported that CH/π interactions in gal_aromatic complexes are comparable to hydrogen bonds in subsequent studies using MP2/6-311G++**. In another study,[40] Sujatha et al (2007) have reported the interaction energies ranging from 3.5-7.5 kcal mol^{-1} for Gal and three aromatic amino acid complexes at different positions and orientations using MP2/6-311G+(d, p). Jimnez-Barbero et al[41] have studied various oligosaccharides bind to hevein domains by NMR experiments and molecular modelling. All these studies indicate the importance of carbohydrate-π interactions and the role of aromatic amino acids in carbohydrate recognition by proteins.

From these studies, it becomes apparent that dispersive interactions play important role in carbohydrate-aromatic interactions and therefore inclusion of dispersive correction term become necessary to evaluate binding energies of carbohydrate-aromatic

complexes. Quantum mechanical (QM) methods can be useful to calculate binding energies, IR vibrational frequencies and NMR shifts and to reveal the structural properties. However, calculating binding energies of complexes for which dispersive interactions are predominant is a challenging task. Highly correlated quantum mechanical methods such as CCSD(T) are needed for the accurate prediction of dispersion contribution and these methods are expensive and can be applied only for small-sized systems such as molecules containing 10-15 atoms. MP2 methods are found to overestimate the interaction energies nearly 1-2 kcal mol^{-1} for complexes involving dispersion forces. Standard density functional methods (DFT) are efficient and cheap compared to high-level ab initio methods. However, DFT methods fail to predict the interaction energies for complexes involving dispersion forces. Grimme[42] has approached this problem by including a dispersion correction term explicitly to standard density functional methods. This empirical atom-atom dispersive correction term is C_6/R^6, where R^{-6} is distant dependent. This approach, inclusion of dispersion term into standard density functional methods, which can be denoted as DFT-D has been proven to be successful approach for calculating interaction energies of noncovalently bound complexes.[43, 44] The same approach has been applied by our group and developed PM3-D scheme by including dispersion correction term to semi-empirical method PM3.[45] This PM3-D method has been tested against a small training set of 22 complexes involving a range of noncovalent interactions. It has also been tested for a large database containing 156 biologically relevant molecules including hydrogen bonded DNA base pairs, interstrand base pairs, stacked base pairs and amino acid base pairs and found to give interaction energies with a mean unsigned error (MUE) of 1.2 kcal mol^{-1}.[45] However, the data set does not contain carbohydrate-aromatic complexes. When applied this method to study carbohydrate-aromatic complexes, it has been found that PM3-D is seriously overestimating the carbohydrate-aroamtic interactions. To tackle this problem, our group have developed PM3-D* method by including modified core-core repulsion term into PM3. PM3-D* method has been used to study carbohydrate-aromatic interactions and has been found to be performing better compared to PM3-D. We explore these three methods, DFT-D, PM3-D and PM3-D* to study the interactions of glucose molecules with tryptophan and tyrosine.

3.3 Computational details

The carbohydrate calculations have been carried out using locally modified version of Gaussian 03, following the DFT-D method developed by Grimme. In the DFT-D approach,[42] a pair-wise additive potential of the form C_6/R^6 is used to account for long-range dispersion effects that can be particularly poorly described both with some density functionals and with current semi-empirical methods (AM1, PM3).[46,47] For each model, the dispersion corrected total energy is given by:

$$E_{total} = E + E_{disp} \qquad 3\text{-}1$$

where E is the normal self-consistent DFT or semi-empirical energy (PM3) and E_{disp} is an empirical term containing the dispersion correction;

$$E_{disp} = -s_6 \sum_i \sum_j \frac{C_6^{ij}}{R_{ij}^6} f_{dmp}(R_{ij}) \qquad 3\text{-}2$$

Here, the summation is over all atom pairs, C_6^{ij} is the dispersion coefficient for the pair of atoms i and j (calculated from the atomic C_6 coefficients), s_6 is a scaling factor that depends on the density functional or semi-empirical method used and R_{ij} is the interatomic distance between atoms i and j. A damping function is used in order to avoid near singularities for small distances. This function is given by:

$$f_{dmp}(R_{ij}) = \frac{1}{1 + e^{-\alpha(R_{ij}/R_0 - 1)}} \qquad 3\text{-}3$$

where R_0 is the sum of the atomic van der Waals radii and α is a parameter determining the steepness of the damping function. We note that initially the DFT-D method used the combination rule:

$$C_6^{ij} = 2C_6^i C_6^j / (C_6^i + C_6^j) \qquad 3\text{-}4$$

Whereas more recently the geometric mean was employed;

$$C_6^{ij} = \sqrt{C_6^i C_6^j} \qquad 3\text{-}5$$

The DFT-D calculations reported in this work make use of the combination rule given in Equation 3-4. The values for the C_6, R_0, s_6 and α parameters were taken from the

respective parameterizations (Table 3.1), with one exception. In the original DFT-D formalism, atomic C_6 coefficients were quoted only for the elements H, C, N, O, F, and Ne. Therefore a corresponding coefficient for chlorine was determined using an algorithm proposed by Halgren giving a value of 8.0 J nm^6 mol^{-1}; a van der Waals radius of 1.82 Å was used for chlorine (Table 3.1).

Table 3.1 Atomic C_6 coefficients and van der waals radii, R_0

	C_6/**J** nm^6 mol^{-1}	R_0/Å
H	0.16	1.11
C	1.65	1.61
N	1.11	1.55
O	0.70	1.49
Cl	8.00	1.82

The DFT-D calculations reported herein have been performed using a locally modified version of GAUSSIAN 03[48] with the dispersion corrected BLYP method (BLYP-D) and the TZV (2d, 2p) basis set. We do not consider basis set superposition errors (BSSE) in view of the quite large basis sets employed. This is justified by the success of the DFT-D method in reproducing the interaction energies of a large database of interactions calculated by high level methods at the basis set limit. As in the case of the DFT-D calculations, the numerical integration was performed employing the weighting scheme of Becke along with grids of ultrafine quality.

3.4 Carbohydrate conformers

Carbohydrates can adopt different conformations because of the hydroxyl groups. Mainly pyranose rings can adopt chair or boat conformations. Chair conformations are generally favored over boat conformations, since boat conformations cause crowding of the groups. Glucose, which is pyranose, has two anomers called as α-anomer and β-anomer. The α-anomer of glucose will have O1 in an axial orientation and the β-anomer of glucose will have O1 in an equatorial orientation. The pyranose ring can adopt chair,

half-chair or boat conformations. Among chair conformations, $^4C_1$ conformation, where 4th Carbon and 1st Carbon will be up and down respectively, will be favored. Glucose, mannose and galactose differ in their chirality at the different carbon centres. In the $^4C_1$ conformation, OH2, OH3 and OH4 groups are in equatorial positions, in mannose; OH2 is axial, whereas in galactose, OH4 is axial. The hydroxyl groups can be classified as gauche and trans conformations dependent on the Hn-On-Cn-Cn+1 dihedral angle. Gauche conformations can be further classified as g+ and g-, where dihedral angles lying between 0 to 120 and 240 to 360 respectively. The exocyclic hydroxyl group can adopt G-, G+ and T conformations dependent on the O6-C6-C5-O5 dihedral angle. The terminal OH6 can adopt t, g+ or g- conformations dependent on the H6-O6-C6-C5 dihedral angle. In the different chair conformations, each group attached to the ring either projects outward from the edge of the pyranose ring in an equatorial position or points upward or downward from the ring in an axial position. Atoms in the axial positions are more likely to clash sterically. In the hexose residues that predominate in biological glycans, one of the chair conformations, designated $^4C_1$, is genetically favoured because this arrangement places the bulky C6–O6 group and most of the hydroxyl groups in equatorial positions. In this conformation, one face of the ring is relatively non–polar. This is normally called as B face of the sugar. The fact that pyranose rings behave as fixed units greatly simplifies description of the conformations of glycans. There are relatively few substituents and thus little of the side chain conformational variability that characterizes protein structures. However, rotation about the C5–O6 bond can occur. As with any single bond, staggered conformations are preferred because these maximize the distance between atoms at either end of the rotating bond. These rotations are described by the torsion angle ω.

C4 makes the boat conformation less stable than the chair conformations. The $^4C_1$ chair conformation results in the least crowding, because the exocyclic C6 and the hydroxyl groups 2, 3, and 4 are in an equatorial position, while in the $^1C_4$ chair conformation they are axial. $^4C_1$ and $^1C_4$ conformations are shown in Fig 3.1. Representations of possible conformations have been generated by rotation about a single bond. A staggered conformation is preferred to maximize the distance between atoms. Three possible staggered conformations can be created by rotations of 120°.

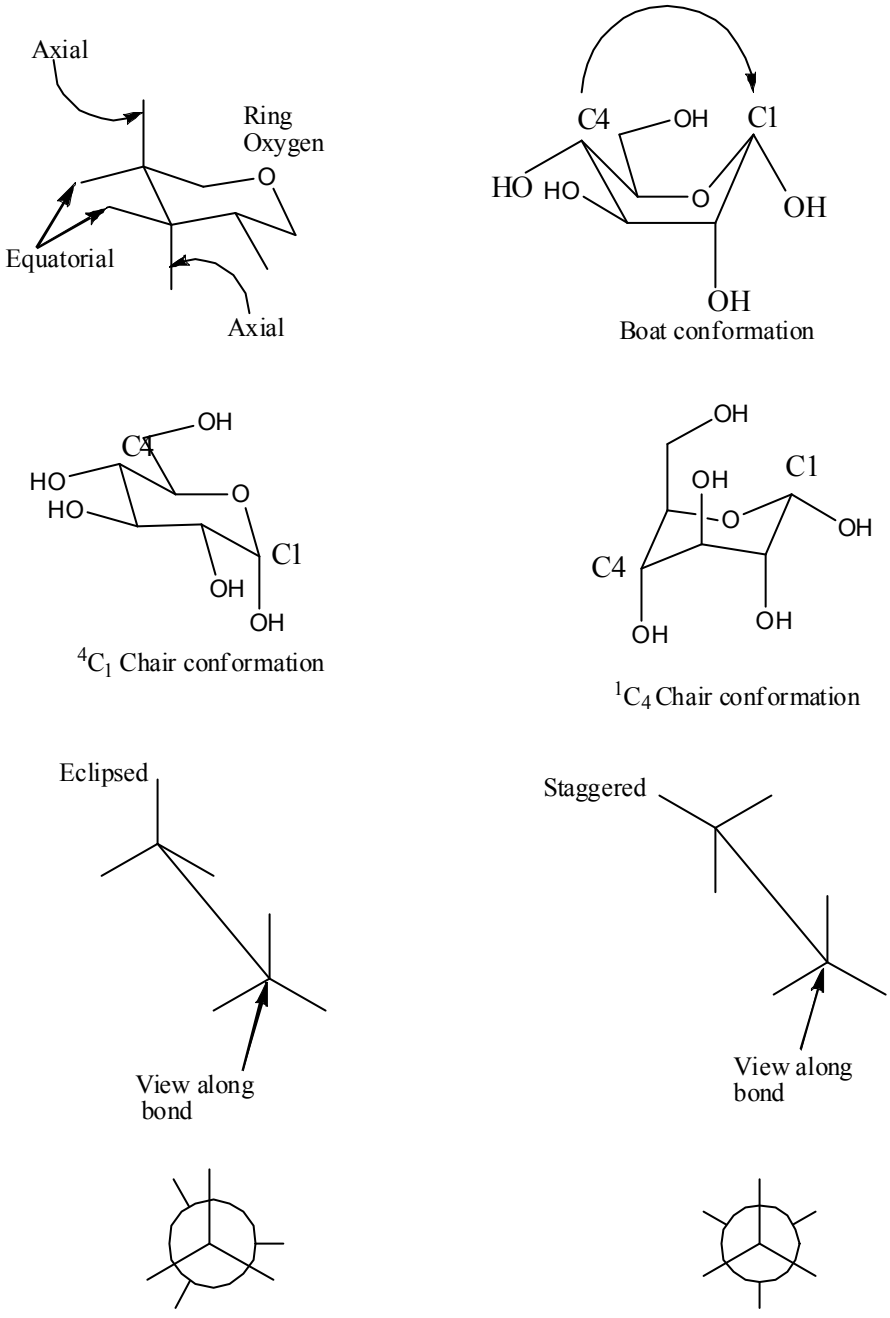


Figure 3.1 Pyranose conformations

3.4 Results and Discussion

3.4.1 Carbohydrate-aromatic complexes

We now explore the models of carbohydrate-aromatic complexes used in our study. We have chosen eleven complexes involving glucose-aromatic interactions from different crystal structures (Table 3.2). The conformers of glucose in these complexes differ (Table 3.2). We modelled OH2, OH3 and OH4 conformation as cc since this conformation involves internal hydrogen bonding and so energetically favorable. We have chosen six complexes in which glucose interacts with tryptophan and five complexes in which glucose interacts with tyrosine. The aromatic amino acids tryptophan and tyrosine are modelled as 3-methyl indole and p-hydroxy toluene respectively.

Since pyranose ring of glucose molecules have hydroxyl groups in their equatorial positions, the two surfaces above and below the pyranose ring are hydrophobic surfaces. Aromatic amino acids can interact with these two hydrophobic surfaces and hydrophilic surface through noncovalent interactions such as CH/π and OH/π interactions. Therefore there are different orientations carbohydrates interact with aromatic amino acids. The chosen complexes fall mainly in three different orientations which can be named as lower-bound complexes, upper-bound complexes and side-bound complexes. In lower-bound complexes, glucose interacts with aromatic amino acid mainly through CH1, CH3 and CH5 groups. In upper-bound complexes, glucose interacts with aromatic amino acid through CH2 and CH4 groups. In side-bound complexes, glucose can interact either through OH2, OH3 and OH4 or through external hydroxyl methyl group and OH1 group. The different orientations of interaction of glucose-aromatic complexes are shown in Fig 3.2.

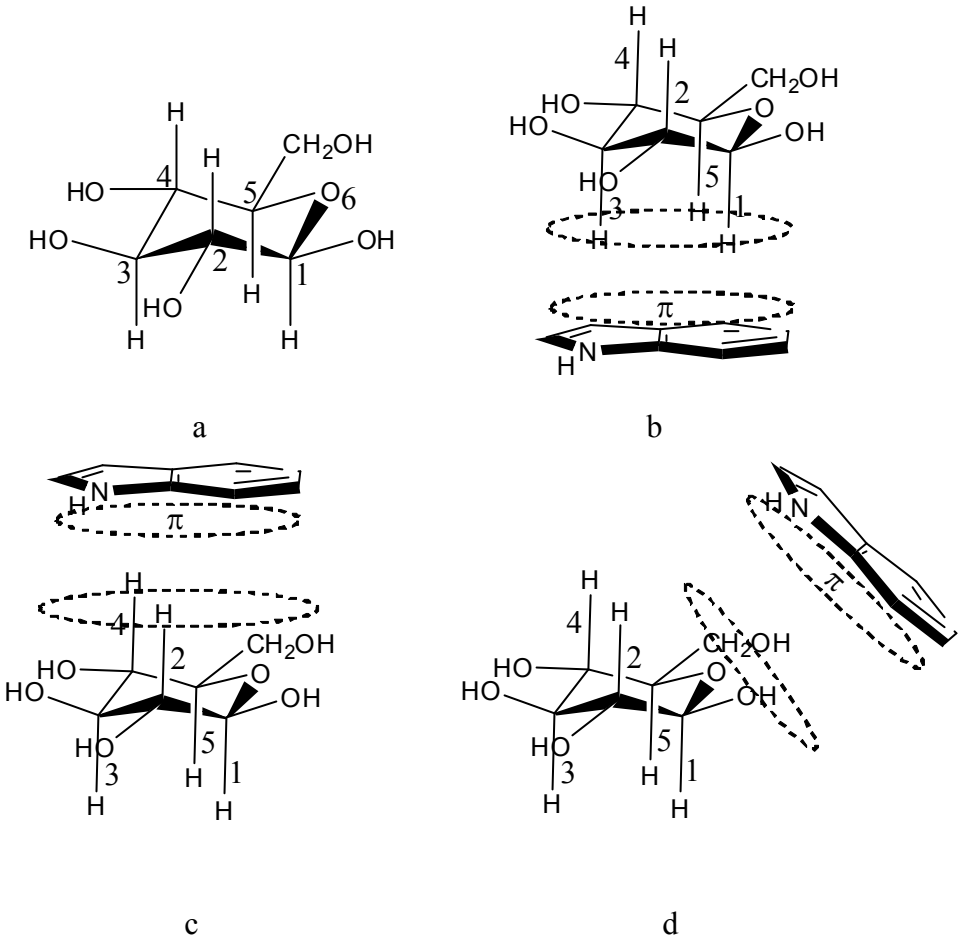


Figure 3.2 a. Structure of glucose, b, c, d. lower-bound, upper-bound and side-bound complex of glucose and tryptophan respectively

3.4.2 DFT-D interaction energies

We now discuss the structural and energetic results of carbohdrate-aromatic complexes obtained by DFT-D method. The eleven glucose-aromatic complexes are optimized by DFT-D method using the functional BLYP and basis set TZV (2d, 2p). The interaction energies of complexes along with IR stretching frequencies and NMR shifts are calculated. Interactions of glucose with 3-methyl indole and with p-hydroxy toluene are discussed in the following sections. The optimized complexes using DFT-D method are shown in Fig 3.3.

3.4.21 Glucose with tryptophan

A few different configurations of glucose and 3-methyl indole, the model of tryptophan were optimized and their binding and interaction energies were calculated. Table 3.2 lists the conformations of glucose and PDB codes from which the structures have been extracted. All of these complexes have same conformation as G-g+ for exocyclic methyl group. The conformation of OH2, OH3 and OH4 has been kept as cc (counter clockwise). The difference in interaction pattern between different complexes lies in the orientations of aromatic amino acid with respect to sugar.

Aromatic amino acid, 3-methyl indole in complex 1, 2, 5 and 6 has been found to interact with the lower face of the sugar. In complex 3 and 4, aromatic amino acid has been found to be interacting in side face of sugar. In complex 3, exocyclic methyl group of glucose is pointing towards the N-H of 3-methyl indole, whereas in complex 4 it is pointing towards the methyl group of 3-methyl indole. Complex 1 and 2 have three C-H bonds pointing towards aromatic ring whereas complex 3 and 4 have four C-H bonds pointing towards the aromatic ring. Complex 5 has three C-H bonds and one N-H bonds pointing towards the aromatic ring.

All interaction energies for the complexes of glucose with 3-methyl indole are range from 9.0 kcal mol^{-1} to 11 kcal mol^{-1} except complex 6. The higher interaction energy of complex 5 is due to the presence of NH/π interaction. Though complex 6 has NH/π interaction, the reason for less interaction energy of complex 6 is due to the change in conformation of glucose upon optimization after which NH/π interaction will be lost.

3.4.22 Glucose with tyrosine

Some configurations of glucose with tyrosine have been extracted from crystal structures. Complex of glucose and *p*-hydroxy toluene, the model of tyrosine, were optimized and their binding and interaction energies were calculated (Table 3.2). Similar to glucose molecules complexed with tryptophan, glucose molecules with tyrosine have G-g+ conformation for exocyclic group except in complex 9 and 10. The conformation of exocyclic methyl group of glucose molecules in complex 9 and 10 is Tg+. The conformation of OH2, OH3 and OH4 of glucose molecules has been kept as *cc*.

Here, the orientation of aromatic amino acid *p*-hydroxy toluene in complex 7 and 11 has been found to be in lower face of sugar, whereas in complex 8, it has been found to be in side face of sugar. In complex 9 and 10, aromatic amino acid has been found to be interacting in upper face of sugar. Glucose interacts with *p*-hydroxy toluene in complex 7 and 11 mainly through CH1, CH3 and CH5. In complex 9 and 10, apart from the hydrogen bond from O-H of *p*-hydroxy toluene, glucose interacts through CH2, CH4 and exocyclic methyl group. In complex 8, glucose interacts mainly through CH3 and CH5.

All interaction energies of complex of glucose and *p*-hydroxy toluene are ranging from 6.9 kcal mol^{-1} to 10.0 kcal mol^{-1}. Complex 7 and 11, which have three CH/π interactions through CH1, CH3 and CH5, have interaction energies as -7.2 and -7.3 kcal mol^{-1} respectively which are close to the value for similar sucrose-toluene complexes. No OH/π interaction has been found among these complexes apart from OH-O hydrogen bonding between glucose and *p*-hydroxy toluene. It has been found that the

interaction energies of glucose with tryptophan are higher than the interaction energies of glucose with tyrosine except complex 8 and 9, where OH-O hydrogen bonding is present. The reason for the higher interaction energies with tryptophan might be due to the larger π surface area which allows glucose molecules to interact more. The fact that CH/π and OH/π interactions stabilize carbohydrate-aromatic complexes has been supported by IR and NMR shifts of C-H and O-H bonds.

3.4.23 IR frequencies

The presence of CH/π and OH/π interactions has been supported by IR and NMR shifts. IR frequencies have been calculated at the level of BLYP-D/TZV (2d, 2p). IR shifts, the difference between O-H or C-H frequency of complex and corresponding frequency of the monomer, have been calculated (Table 3.3 and Table 3.4). All frequency calculations have been carried out on the minimized structures of complex and monomers. It is evident from large blue shifts from C-H vibrational modes that sugar in complex 1 interacts with aromatic amino acid through CH1, CH3 and CH5 bonds. There is no red shift from O-H bonds of sugar in complex 1. In complex 2, CH1, CH3 and CH5 frequncies are shifted to large wave numbers indicating the interaction through C-H bonds, particularly through CH5 bond which is shifted nearly 104 ppm. Similar to complex 1, there is no O-H shift in complex 2. Complex 3 and Complex 4 which are side-bound complexes involves interaction of CH1, CH5 and OH1 with 3-methyl indole. This fact is evident from the shifts of their vibrational modes. OH1 in complex 3 and complex 4 have got large red shifts as -100 and -99 ppm respectively. CH1 and CH5 bonds in complex 3 and 4 have got blue shifts as 57 and 58 ppm respectively. Complex 5, lower-bound complex is expected to be involving CH1, CH3 and CH5 interactions. Their vibrational modes have been shifted by 144, 78 and 16 ppm respectively which confirm their interactions with aromatic amino acid. Also in complex 5, there is a red shift of OH2 bond which forms partial hydrogen bond with N of aromatic amino acid. Hence, complex 5 altogether involves 3 CH/π and one OH/π interactions which give higher interaction energy -11.5 kcal mol^{-1}. In complex 6, CH1 and CH5 have shifted to 42 and 54 ppm respectively from their monomer vibrational modes and OH4 have shifted to lower wave number -78ppm which shows the partial

hydrogen bonding with N of aromatic amino acid. Although the interaction mode of complex 6 is similar to complex 5, the reason for the lower interaction energy than complex 5 is because of the loss of OH4-N hydrogen bonding during the optimization of sugar. Complex 7 is a lower-bound complex expected to be involving interaction through CH1, CH3 and CH5 bonds. It has been supported by their blue shifts which are 46, 52 and 106 ppm respectively. There is no OH shift found in complex 7. Complex 8, side-bound complex involves less interaction of CH/π and OH/π and mainly stabilized by OH-O hydrogen bond between sugar and aromatic amino acid. Complex 9, upper-bound complex is expected to be involving interactions mainly from CH2 and CH4 bonds and is evident from their blue shifts 107 and 46 ppm respectively. The large red shift from OH1 is due to hydrogen bond between OH1 and O of *p*-hydroxy toluene. Complex 10 which is side-bound complex is mainly stabilized by hydrogen bond between sugar and *p*-hydroxy toluene. Complex 11, lower-bound complex, involves CH1, CH3 and CH5 interactions and has been supported by their large blue shifts 65, 15 and 120 ppm respectively. There is no OH red shift. All these C-H and O-H shifts are clearly in agreement with the interaction energies of the complexes.

3.4.24 NMR shifts

These interaction energies and IR shifts are also supported by NMR data. NMR shifts have been calculated within GIAO approximation at the level of BLYP/TZV (2d, 2p) basis sets. The calculated proton NMR shifts are summarized in parentheses of Table 3.4. The absolute values indicate the interaction of C-H bonds with aromatic ring.

It is evident in all the complexes from complex 1 to 11 that large NMR shifts are in line with large IR blue shifts. Particularly Complex 1 and 5 have got large C-H blue shifts and their corresponding NMR shifts are 3.6 and 3.4 which support the fact of CH/π interaction. Complex 8 and 10 have got lower IR shifts and their corresponding NMR shifts are nearly 0 or less than zero. All other complexes have got significant C-H shifts from IR stretching frequencies and their corresponding NMR data also support this fact. Therefore, the presence of CH/π and OH/π interactions and their contributions to interaction energies have been clearly supported from IR and NMR shifts.

Table 3.2 Interaction energy (IE) and dispersion contribution (kcal mol^{-1}) for glucose-3-methylindole (W), and –*p*-hydroxytoluene(A) complexes (DFT-D).

Complex	PDB code	Complex	Conformer	IE	Dispersion energy
1	1KWF[49]	Glc401(B)-W205(A)	ccG-g+	–9.8	–10.6
2	1KWF	Glc402(B)-W132(A)	ccG-g+	–10.6	–9.1
3	1GWM[50]	Bgc1158(A)-W26(A)	ccG-g+	–10.2	–9.5
4	1GWM	Bgc1160(A)-W24(A)	ccG-g+	–10.3	–10.2
5	M[a]	β-glucose-W	g+ttG-g+	–11.5	–11.0
6	M[a]	β-glucose-W	ttg+G-g+	–10.0	–10.3
7	1KWF	Glc404(B)-Y372(A)	ccG-g-	–7.3	–9.2
8	1KWF	Glc404(C)-Y372(A)	ccG-g-	–10.6	–6.2
9	1KWF	Glc405(B)-Y277(A)	ccTg+	–9.8	–9.0
10	1KWF	Glc405(C)-Y277(A)	ccTg+	–9.0	–8.4
11	1GWM	Bgc1156(A)-Y46(A)	ccG-g-	–7.2	–8.6

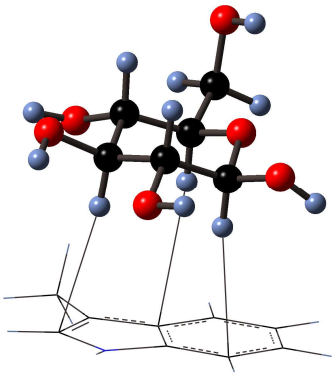
[a]Initial structures **5** and **6** were generated from the optimized structures 1 and 2, respectively, by changing the sugar conformation

Table 3.3 OH stretching frequencies and shifts from monomer values (cm^{-1}) for glucose-3-methylindole and –*p*-hydroxytoluene complexes (DFT-D).

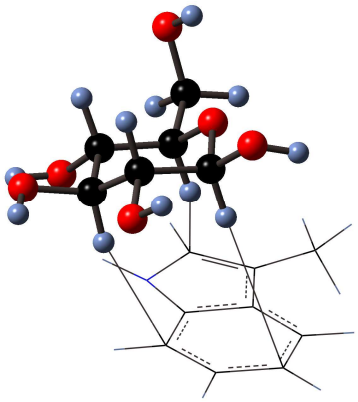
Complex	Frequency					shift				
	OH1	OH2	OH3	OH4	OH6	OH1	OH2	OH3	OH4	OH6
1	3671	3669	3676	3667	3656	-2	-8	5	-1	-4
2	3675	3676	3670	3662	3660	2	-1	-1	-6	0
3	3571	3668	3671	3667	3647	-100	-9	0	-1	-13
4	3573	3666	3671	3668	3652	-99	-11	0	0	-8
5	3667	3580	3640	3672	3655	4	-62	-5	-1	-6
6	3674	3677	3652	3590	3655	1	0	-19	-78	-5
7	3671	3675	3674	3653	3673	0	-1	2	-2	-1
8	3672	3676	3674	3640	3680	1	0	2	-15	6
9	3518	3659	3675	3661	3606	-110	0	-2	2	-6
10	3662	3677	3674	3662	3617	-10	1	1	2	6
11	3676	3673	3672	3653	3673	5	-3	0	-2	-1

Table 3.4 C-H stretching frequencies and shifts from monomer values (cm^{-1}) for glucose-3-methylindole and *–p*-hydroxytoluene complexes (DFT-D). The chemical shift (δ, ppm) for C-H protons, given in brackets, is calculated at the BLYP/TZV2D level .

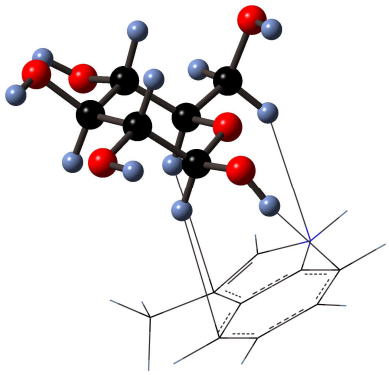
Complex	Frequency				
	CH1	CH2	CH3	CH4	CH5
1	3024	2948	2978	2956	2940
2	2907	2946	2961	2977	3010
3	2935	2950	2913	2962	2932
4	2936	2951	2911	2964	2931
5	2964	3016	2950	2970	2919
6	2920	2951	2910	3022	2959
7	2909	2952	2970	2945	2991
8	2876	2957	2944	2982	2915
9	2931	3016	2923	2966	2920
10	2893	2962	2924	2971	2917
11	2929	2961	2933	2947	3007
	Shift				
1	146 (3.6)	-4 (0.6)	62 (1.6)	-7 (0.4)	35 (0.5)
2	29 (1.6)	-5 (0.3)	45(-0.4)	14 (0.2)	104 (2.6)
3	57 (2.2)	-2 (0.6)	-3 (0.5)	-1 (0.5)	27 (1.2)
4	58 (2.8)	-1 (0.6)	-5 (0.6)	1 (0.4)	26 (1.6)
5	144 (3.4)	-4 (0.7)	78 (1.3)	8 (0.4)	16 (0.5)
6	42 (1.7)	-1 (0.5)	-6 (1.0)	59 (0.9)	54 (2.6)
7	46 (0.5)	-8 (0.2)	52 (0.1)	-2 (0.3)	106 (2.0)
8	13 (0.1)	-3 (0)	26 (-0.3)	35(-0.1)	29 (-0.5)
9	-18 (0.2)	107 (2.2)	-3 (0.3)	46 (0.1)	-19 (0.2)
10	18 (0)	5 (-0.3)	3 (0.1)	32 (-0.2)	9 (0.2)
11	65 (0.1)	1 (0.2)	15 (0.1)	0 (0.3)	120 (2.5)



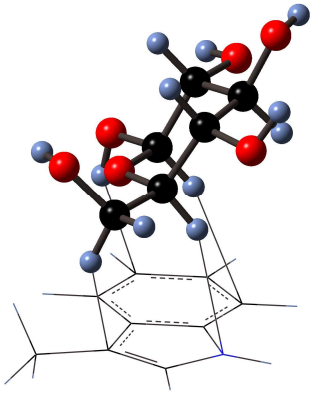
1



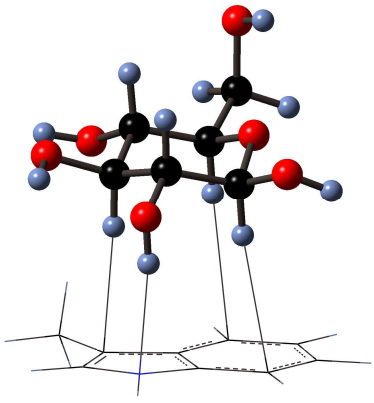
2



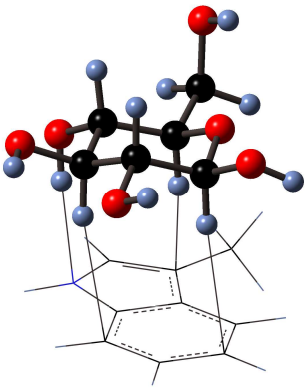
3



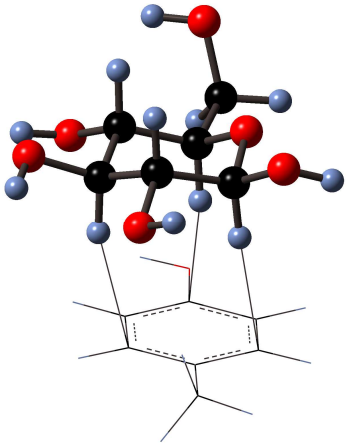
4



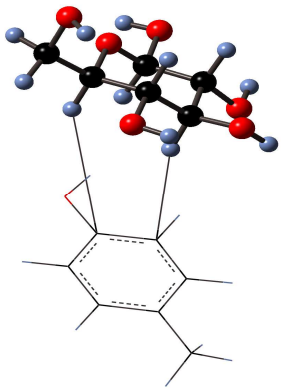
5



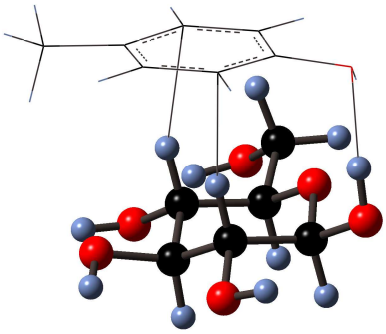
6



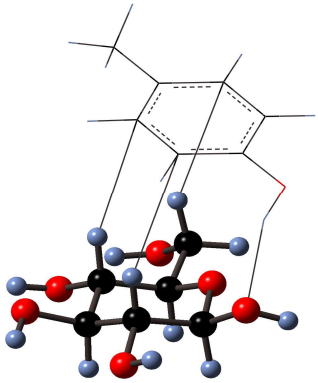
7



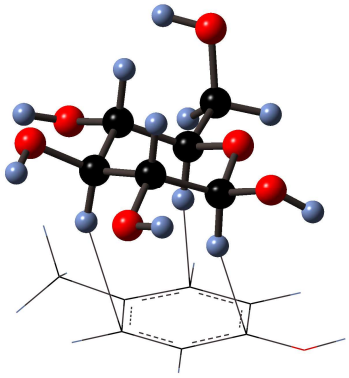
8



9



10



11

Figure 3.3 Optimized structures of glucose with 3-methyl indole and p-hydroxy toluene obtained by DFT-D method

3.4.3 PM3-D and PM3-D* interaction energies

We have evaluated the interaction energies of complexes using semi-empirical methods implemented with dispersion correction. PM3-D and PM3-D* interaction energies and dispersion energies are shown in Table 3.5. Inter molecular distances of complexes are shown in Table 3.6. We found that PM3-D is overestimating the interaction energies by 4-5 kcal mol^{-1} compared to DFT-D interaction energies, and underestimating ring-ring separation by 0.5 Å. The PM3-D optimized complexes are shown in Fig 3.4. Although PM3-D has been found successful for studying interactions involving aromatic systems, it is not accurate enough to study carbohydrate-aromatic interactions. To solve this problem, a new PM3-D method named PM3-D* has been developed by modifying core-core repulsion term. All the complexes were optimized using this new method PM3-D* and interaction energies were calculated. The optimized complexes using PM3-D* are shown in Fig 3.5. The interaction energies and distances obtained using PM3-D* have been found in good agreement with DFT-D values. The PM3-D* interaction energies and inter-ring separation are in the range of 0.7 kcal mol^{-1} and 0.16 Å of DFT-D values.

Table 3.5 Interaction energies (kcal mol^{-1}) of sugar-aromatic complexes

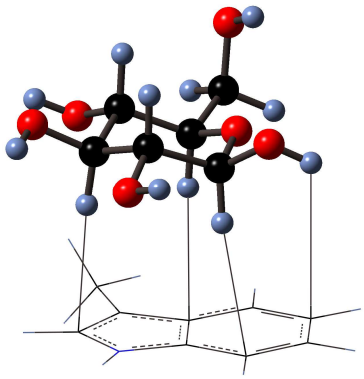
Complex	PM3-D//DFTD	PM3-D*//DFTD	PM3-D//PM3-D	PM3-D*//PM3-D*
1. β-glucose-Trp	-13.25	-9.83	-14.98 (-13.3)	-10.12 (-11.0)
2. β-glucose-Trp	-11.90	-10.52	-14.38 (-8.8)	-9.31 (-9.9)
3. β-glucose-Trp	-12.56	-10.83	-14.89 (-12.6)	-9.43 (-11.2)
4. β-glucose-Trp	-12.88	-10.47	-14.33 (-12.8)	-9.25 (-11.3)
5. β-glucose-Trp	-13.97	-12.16	-17.44 (-13.8)	-8.89 (-10.7)
6. β-glucose-Trp	-12.84	-11.88	-17.63 (-14.4)	-12.12 (-12.8)
7. β-glucose-Tyr	-11.35	-8.10	-13.14 (-12.0)	-9.19 (-10.2)
8. β-glucose-Tyr	-12.57	-9.57	-12.41 (-7.3)	-8.22 (-5.9)
9. β-glucose-Tyr	-11.62	-8.44	-15.71 (-13.0)	-8.68 (-11.2)
10. β-glucose-Tyr	-10.72	-9.88	-16.51 (-11.0)	-8.93 (-11.5)
11. β-glucose-Tyr	-10.51	-7.81	-12.86 (-11.6)	-7.17 (-10.1)
MUE	2.55	0.74	5.29	1.24

The dispersion energy is given in parentheses

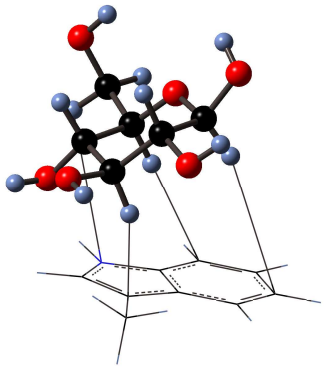
Table 3.6 Distance between sugar-aromatic complexes (defined by centre of mass)

Complex	PM3-D//PM3-D	PM3-D*//PM3-D*
1. β-glucose-Trp	3.95 (0.15)	4.32 (0.22)
2. β-glucose-Trp	4.95 (0.52)	4.38 (0.05)
3. β-glucose-Trp	4.35 (0.11)	4.37 (0.10)
4. β-glucose-Trp	4.17 (0.27)	4.40 (0.04)
5. β-glucose-Trp	3.93 (0.24)	4.29 (0.12)
6. β-glucose-Trp	3.99 (0.21)	4.04 (0.16)
7. β-glucose-Tyr	3.91 (0.20)	4.05 (0.06)
8. β-glucose-Tyr	5.19 (0.23)	5.59 (0.16)
9. β-glucose-Tyr	3.51 (0.55)	3.89 (0.18)
10. β-glucose-Tyr	3.98 (0.41)	3.84 (0.56)
11. β-glucose-Tyr	4.04 (0.11)	4.04 (0.11)
MUE	0.27	0.16

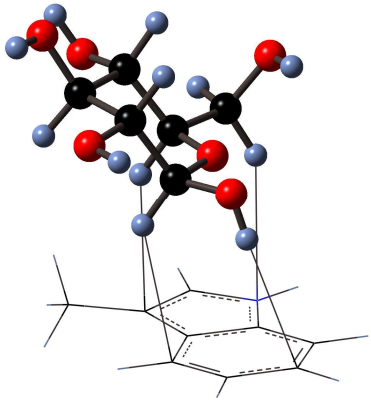
The deviation of the semi-empirical structure distance from DFT-D is given in parentheses



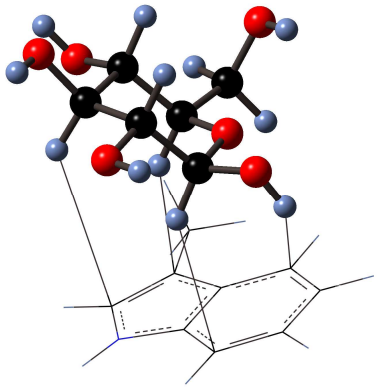
1



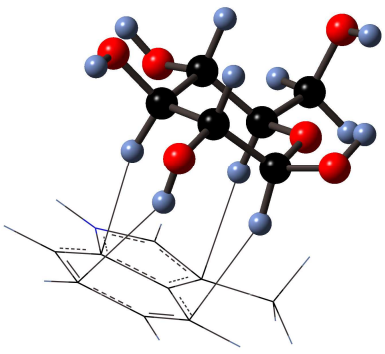
2



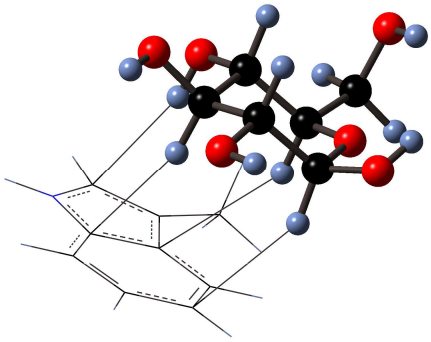
3



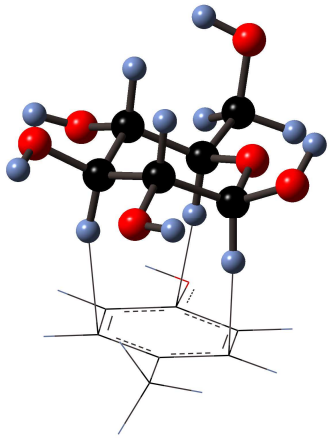
4



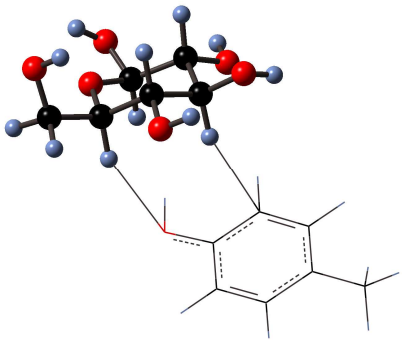
5



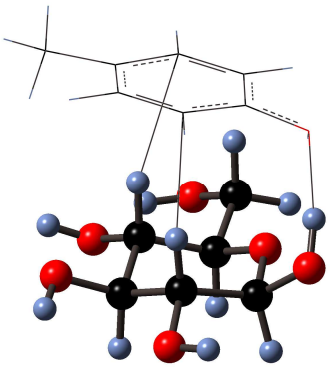
6



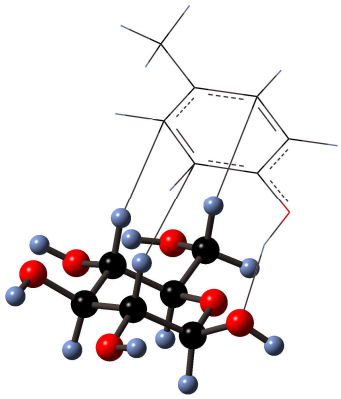
7



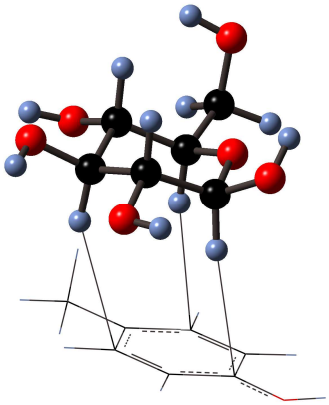
8



9

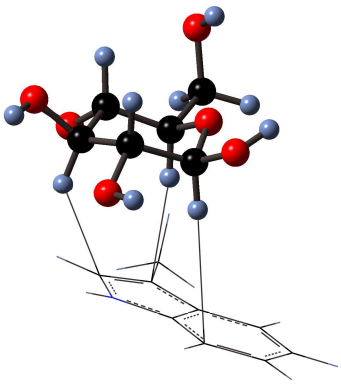


10

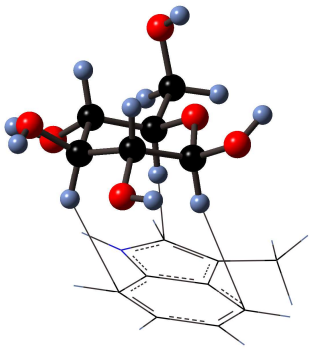


11

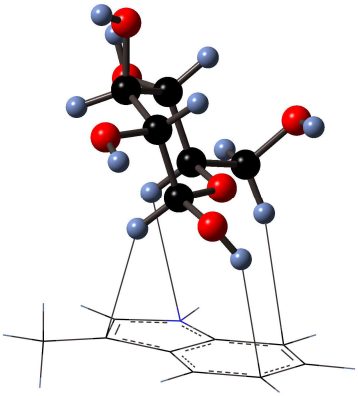
Figure 3.4 Optimized structures of glucose with 3-methyl indole and p-hydroxy toluene obtained by PM3-D method



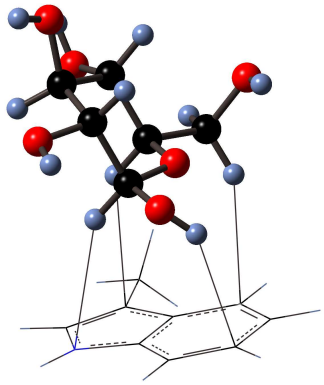
1



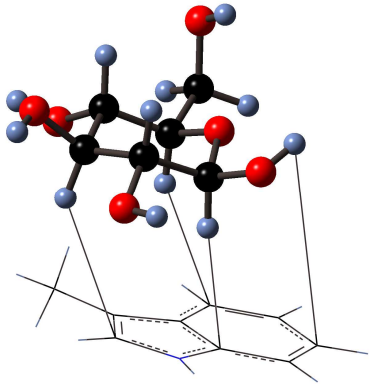
2



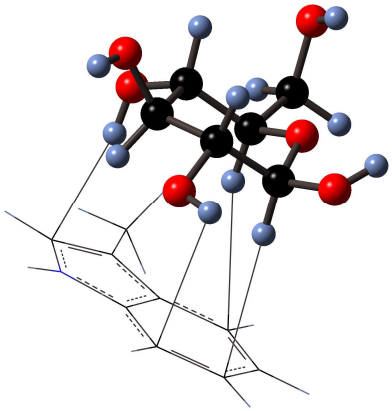
3



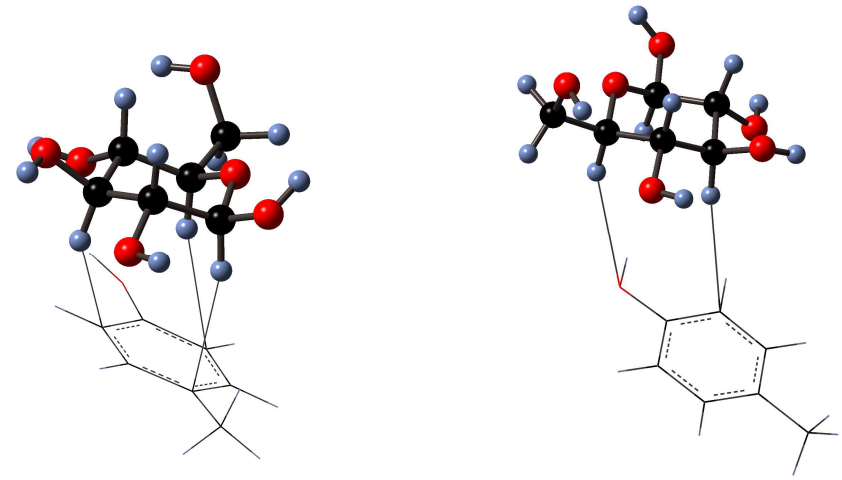
4



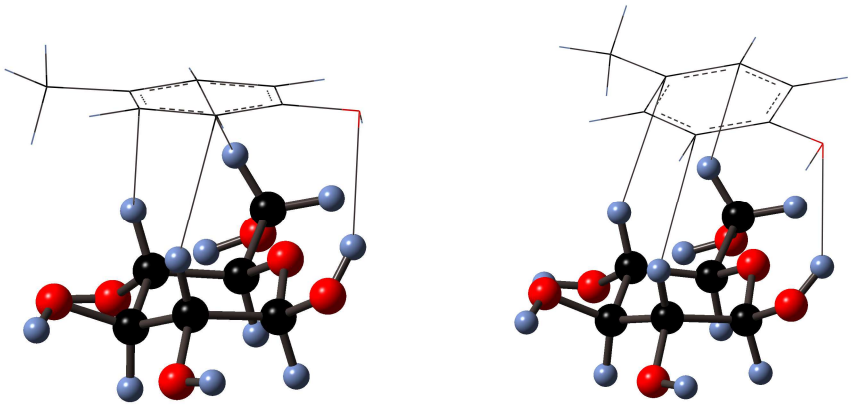
5



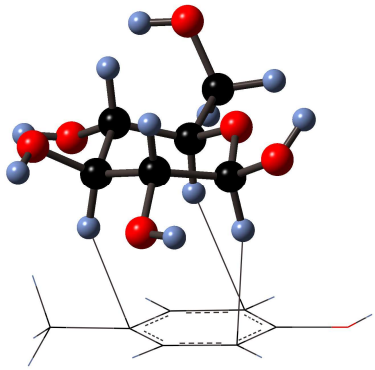
6



7 **8**



9 **10**



11

Figure 3.5 Optimized structures of glucose with 3-methyl indole and p-hydroxy toluene obtained by PM3-D* method

3.5 Conclusions

We have studied carbohydrate-aromatic interaction by density functional theory method implemented with dispersion correction. We have used BLYP functional with TZV (2d, 2p) basis sets. We have studied the structural and energetic aspects of carbohydrate-aromatic complexes in detail. We have also used semi-empirical methods implemented with dispersion correction known as PM3-D and PM3-D* to study these carbohydrate-aromatic complexes. We have found that PM3-D method which has been proven to be successful method in studying aromatic-aromatic interactions, is overestimating carbohydrate-aromatic interactions. Reparameterized PM3-D method known as PM3-D* which includes modified core-core repulsion term has been found to performing better for studying carbohydrate-aromatic interactions. We have studied the interaction of glucose with 3-methyl indole and p-hydroxy toluene, models of aromatic amino acids tryptophan and tyrosine respectively in different orientations. Lower-bound complexes involve mainly CH/π interactions through CH1, CH3 and CH5 bonds. Upper-bound complexes involve CH/π interactions through CH2 and CH4 bonds. Side-bound complexes involve mixture of CH/π and OH/π interactions through CH1, CH5 and OH1 bonds. The interaction energies of glucose with tryptophan are ~2 kcal mol^{-1} higher than with tyrosine. Since tryptophan has larger π surface area, glucose interact more with tryptophan. We found that the interaction of glucose with tyrosine through three C-H bonds is approximately 6-7 kcal mol^{-1}, where each C-H bond contributes ~2 kcal mol^{-1}. The interaction of glucose with tryptophan through three C-H bonds gives interaction energy approximately 9-10 kcal mol^{-1} which suggests that each C-H bond contributes ~3 kcal mol^{-1}. Carbohydrate-aromatic interaction mainly involves CH/π or OH/π or mixture of CH/π and OH/π interactions. This has been supported by IR frequency shifts and NMR proton shifts. These shifts are in good agreement with interaction energies of the complexes. NH/π interaction has also been found in glucose-3-methyl indole complexes. It has been found that dispersion contribution is the main stabilizing factor for carbohydrate-aromatic complexes. Dispersion energy is higher than overall interaction energy which means that complex would be unbound without dispersion

contribution. Therefore standard density functional theory implemented with dispersion correction term can be used to study carbohydrate-aromatic complexes.

3.6 References

1. David, L. S.; Donald, R, T.; Horace, H. L. *Life Sciences.* **1978**,*22*,727.

2. Robert, X.R.; Haser, T.E.; Gottschalk, F. R.; Driguez, B. H.; Svensson, N. A. *Structure.* **2003**,*11*, 973.

3. Boraston, A. B.; Bolam, D. N.; Gilbert, H. J.; Davies, G. J. *Biochem. J.* **2004**, *382*, 769.

4. Gibson, R.M.; Svensson, B. *Carlsberg Res. Commun.***1987**, *52*, 373.

5. Kadziola, A.; Søgaard,M.; Svensson,B.; Haser,R. *J. Mol. Biol*. **1998**, *278*, 205.

6. Søgaard, M.; Kadziola, A.; Haser, R.; Svensson, B. *J. Biol. Chem.***1993**, *268*, 22480.

7. Simon, J. C. *Proc. Natl. Acad. Sci.* **2002**, *99*,14077.

8. Sujatha, S. M.; Petety, V. B. *PROTEINS: Structure, Function, and Bioinformatics.* **2004**, *55,*44.

9. Susanne, T. I.; Kurt, D. **1994**, *269*, 15512.

10. Els, J. M.; Van, D.; Qiang, H.; Diana, C.; Annick, B.; Pierre, R.; Fred, V. L.; Peumans, W. J. *Eur. J. Biochem.* **2000**, *267*, 2746.

11. Lehar, S.M.; Pedersen, J.T.; Kamath, R.S.; Swimmer, C.; Goldmacher, V.S.; Lambert, J.M.; BlaÈttler, W.A.; Guild, B.C. *Protein Eng*. **1994**, *7*, 1261.

12. Rutenber, E.; Robertus, J.D. *Proteins*. **1991**, *10*, 260.

13. Thomas W. H.; Remy, L.; Julie, B.; Minh-Hoa, D. T.; Gerard, S.; Anne, I.; Elias F.; Lode, W.; Marilynn, E. E. *J. Mol. Biol.***1999**, *286*, 1161.

14. Simon, J. C.; David, N. B.; Johan, P. T.; Harry, J. G.; Luis, M. A.; Gideon, J. D.; Carlos, M. G.; Fontes, A. *Biochemistry*, **2000**, *39*, 5013.

15. James, M. R. *Annu. Rev. Biophys. Biomol. Struct.***1995**. *24,*551.

16. James, B. T.; Hazel, M. H. *Journal. Biol. Chem.* **2002**, *277*, 20854.

17. Keiko, H.; Ryuta, K.; Osamu, S.; Masanobu, A.; Kazuaki, H.; Kunio, Y. *J. Biochem.***2003**, *134*, 881.

18. Tama, P.; Lorand, S.; Tibor, N.; Laszlo, O.; Peter, J. S.; Michael, P. W.; Harry, J. G. *Biochemistry*. **2000**, *39*, 985.

19. Anna, B.; Daniela, A.; Donatella, P.; Inmaculada, S.; Silvia, M. F.; Javier, C.; Jesffls, J. *Chem. Eur. J.* **2004**, *10*, 4395.

20. Sophie, V. D.; Daz, M.; Carmen, F. A. Pan, W.; Vincent, S. P.; Cuevas,G.; Canada, F. J.; Jimenez-Barbero, J.; Bartik, K. *Chem. Eur. J.* **2008**, *14*, 7570.

21. Muraki, M.; Harata, K. *Biochemistry*. **2000**, *39*, 292.

22. Sumner, J. B.; Howell, S. F. *Journal of Bacteriology*, **1936**, *32*, 227.

23. Quiocho, F. A. *Protein-Carbohydrate interactions: Basic molecular features Pure Appl Chem*, **1989**, *61*,1293.

24. Spurlino, J.; Rodseth, L.; Quiocho, F. A. *J. mol. Biol.*, **1992**, *226,* 15.

25. Spiwok, V.; Lipovova, P.; Skalova, T.; Vondrackova, E.; Dohnalek, J.; Hasek, J.; Kralova, B. *J. Comp. Aided. Mol. Design*., **2006**, *19*, 887.

26. Zolotnitsky, G.; Cogan, U.; Adir, N.; Solomon, V.; Shoham, G.; Shoham, Y. *Proc. Nat. Acad. Sci.* **2004**, *101*, 11275.

27. Shulami, S.; Zaide, G.; Zolotnitsky, G.; Langut, Y.; Feld, G.; Sonenshein, A. L.; Shoham, Y. *Appl. Environ. Microb.*, **2007**, *73*, 874.

28. Huber, R. E.; Hakda, S.; Cheng, C.; Cupples, C. G.; Edwards, R. A. *Biochemistry,* **2003**, *42*, 1796.

29. Sophie, V. D.; Daz, M.; Carmen, F. A. Pan, W.; Vincent, S. P.; Cuevas,G.; Canada, F. J.; Jimenez-Barbero, J.; Bartik, K. *Chem. Eur. J.* **2008**, *14*, 7570.

30. Quiocho, F. A. *Ann. Rev. Biochem.*, **1986**, *55*, 287.

31. Chervenak, M. C.; Toone, E. J. *J. Am. Chem. Soc.*, **1994**, *116*, 10533.

32. Chervenak, M. C.; Toone, E. J. *Biochemistry*., **1995**, *34*, 5685.

33. Pearlman, D. A.; Kollman, P. A. *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications.* (W. van Gunsteren, P. K. Weiner, eds.), The Netherlands: ESCOM, **1989**, 101.

34. Miyamoto, S.; Kollman, P. *Proteins*., **1993**, *16*, 226.

35. Stanca-Kaposta, E. C.; Gamblin, D. P.; Screen, J.; Liu, B.; Lavina, C. S.; Davis, B. G.; Simons, J. P. *Phys. Chem. Chem. Phys.*, **2007**, *9*, 4444.

36. Fernandez-Alonso, M. C.; Canada, F. J.; Jimenez-Barbero, J.; Cuevas, G. *J. Am. Chem. Soc.*, **2005**, *127*, 7379.

37. Spiwok, V.; Lipovova, P.; Skalova, T.; Buchtelova, E.; Hasek, J.; Kralova, B. *Carbohydrate. Res.*, **2004**, *339,* 2275.

38. Spiwok, V.; Lipovova, P.; Skalova, T.; Vondrackova, E.; Dohnalek, J.; Hasek, J.; Kralova, B. *J. Comp. Aided. Mol. Design*., **2006**, *19*, 887.

39. Sujatha, M. S.; Sasidhar, Y. U.; Balaji, P. V. *Protein. Sci.*, **2004**, *13*, 2502.

40. Sujatha, M. S.; Sasidhar, Y. U.; Balaji, P. V. *J. Mol. Struct. THEOCHEM.*, **2007**, *814*, 11.

41. Sophie, V. D.; Daz, M.; Carmen, F. A. Pan, W.; Vincent, S. P.; Cuevas,G.; Canada, F. J.; Jimenez-Barbero, J.; Bartik, K. *Chem. Eur. J.* **2008**, *14*, 7570.

42. Grimme, S. *J. Comput. Chem.*, **2004**, *12*, 1463.

43. Grimme, S. *J. Comput. Chem.*, **2004**, *25*, 1463.

44. Grimme, S. *J. Comput. Chem.*, **2006**, *27*, 1787.

45. McNamara, J. P.; Hillier, I. H., *Phys. Chem. Chem. Phys.*, **2007**, *9*, 2362.

46. Dewar, M. J. S.; Zoebisch, E.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.*, **1993**, *115*, 5348.

47. Stewart, J. J. P.; *J. Comput. Chem.*, **1989**, *10*, 209.

48. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. GAUSSIAN 03 (Revision C.02), Gaussian, Inc., Wallingford, CT, **2004**.

49. Guerin, D. M. A.; Lascombe, M. B.; Costabel, H.; Souchon, V.; Lamzin, P.; Beguin, P.; Alzari, M. *J. Mol. Biol.*, **2002**, *316*, 1061.
50. Charnock, S. J.; Bolam, D. N.; Nurizzo, D.; Szabo, L.; McKie, V. A.; Gilbert, H. J.; Davies, G. J. *Proc. Natl. Acad. Sci. U. S. A.*, **2002**, *99*, 14077.

Chapter 4

Carbohydrate-aromatic interactions: Validation of different density functional methods and DFT-D method

4.0 Introduction

In this chapter, we discuss the results of different density functional including both standard and newly developed DFT functionals in describing carbohydrate-aromatic interactions. In previous chapter, we discussed the importance of carbohydrate-aromatic interactions in the recognition of carbohydrates in the binding motifs of proteins using a density functional method and semi-empirical methods both implemented with dispersion correction term. As we discussed earlier, nonbonded interactions play important roles in understanding structure-function of proteins and in designing rational drugs.[1, 2] Nonbonded interactions such as hydrogen bonds and dispersive interactions are essential to understand biological recognition and are needed an accurate description. Standard density functional methods fail to describe these noncovalent interactions. The high level methods beyond MP2, such as CCSD(T) can give an accurate description of dispersion interactions. These high level methods like CCSD(T) together with large basis sets and basis set superposition error (BSSE) corrections can be applied to study only small models contain approximately <15 atoms.[3] For larger models, these methods are not possible to use due to the computational cost. Still, these methods have been used to evaluate databases containing complexes with different range of such nonbonded interactions found in biological systems. To overcome the difficulties in treating dispersive interactions,

Grimme proposed an approach that is to implement the dispersive term explicitly to standard density functional method, which is known as DFT-D. Grimme's DFT-D method has been proved to be successful method to study systems involving noncovalent interactions in biological systems. In line with Grimme's approach,[4] Hillier et al have added dispersive correction term to semi-empirical methods (PM3-D and PM3-D*).[5, 6] These methods also have already been shown to be very successful methods to study noncovalent interactions. Recently, Zhao and Truhlar[7, 8-17] have introduced new set of DFT functionals which are known as M0x functionals to describe noncovalent interactions. The lack of studies in modelling of π-interactions involved in the recognition of carbohydrates in proteins lead us to study these interactions. In previous chapter, we have shown that the DFT-D method is successful in describing these noncovalent interactions in carbohydrate-aromatic complexes. We have discussed the DFT-D energetics, structural and vibrational features of a range of complexes involving monosachharides and analogues of aromatic amino acids. We have analysed the importance of CH/π and OH/π interactions in the recognition of carbohydrates in proteins by taking simple complexes like β-glucose-*p*-hydroxy toluene and β-glucose-3-methylindole, as model systems of β-glucose and aromatic amino acids tyrosine and tryptophan respectively. We have found these two CH/π and OH/π interactions make the complexes more stable, although the crystal structures display the preference of CH/π interactions for carbohydrate to interact with aromatic amino acids. The strengths of CH/π and OH/π interactions in the stability of carbohydrate-aromatic complexes are reflected in the shift of the corresponding C-H and O-H stretching frequency in gas phase IR vibrational spectrum. We have found that our DFT-D calculations are in agreement with experiment.[18, 19]

The main objective of this work presented in this chapter is to find the most appropriate quantum mechanical methods to quantify the noncovalent interactions (CH/π and OH/π) involved in carbohydrate-aromatic interactions. We have tested the accuracy of a variety of DFT functionals with a special focus to the new Truhlar functionals as well as the DFT-D approach, with reference to a small database of high level calculations of carbohydrate-aromatic complexes. In order to accurately quantify

these CH/π and OH/π interactions, we have also studied a simple model of methane and water with 3-methylindole complexes.

4.1 Computational Details

For database, MP2 energies of the complex and monomers are extrapolated to the basis limit by the use of the aug-cc-pVDZ and aug-cc-pVTZ basis sets giving the MP2/CBS energy. This value is then corrected for higher order correlation effects obtained by computing MP2 and CCSD(T) energies using a modified 6-31G** (0.25, 0.15) basis set followed by employing the difference between these energetic to obtain the correction. Basis set superposition errors (BSSE) are also taken into account in the calculation of the interaction energy.

4.1.1 DFT calculations

We have employed a number of DFT functionals including DFT-D method to test their performance to study carbohydrate-aromatic interactions. DFT functionals we used are as follows,

- GGA functionals: BLYP and PBE
- Hybrid GGA functionals: B3LYP, B97-2, B98, BH&H and BH&HLYP
- Meta GGA functionals: VSXC, M06-L
- Hybrid Meta GGA functionals: BMK, MPW1B95, MPWB1K, PW6B95, PWB6K, M05, M05-2X, M06 and M06-2X

The theories behind these functionals are described in chapter 2. We have also performed calculations with BLYP-D and also with MP2 method. Since the new Truhlar DFT functionals were not available in the Gaussian 03 program,[20] we have used these functionals using our implementation of these functionals within the Gaussian 03 code revision D. We have used TZV (2d, 2p) basis sets throughout our calculations without BSSE correction.

4.2 Results and Discussions

4.2.1 Methane-3-methylindole complexes: Simplest model for CH/π interaction

Since carbohydrate-aromatic complexes mainly involve CH/π and OH/π interactions, we have chosen two simple models such as methane-3-methylindole and water-3-methylindole which can describe CH/π and OH/π interactions for validation of different density functionals. We have optimized methane-3-methylindole complexes starting from random initial positions. We have located one stable structure having each hydrogen bond of water pointing towards each of the two rings of 3-methylindole. Ringer et al,[21] have studied methane-3-methylindole complexes at CCSD(T) level and have located number of structures. The most stable structure is found to be having one hydrogen atom pointing to each of the two rings with interaction energy of -2.08 kcal mol^{-1}. We have located similar structure as Ringer et al at DFT-D level with interaction energy of -1.69 kcal mol^{-1}. In order to study the performance of standard and newly developed DFT functionals, we took the DFT-D optimized structure of methane-3-methylindole complex and evaluated the interaction energies using various density functional methods and MP2 method. The interaction energies (I.E.) are shown in Table 4.1. For convenience, we have also plotted the I.E. values for different functionals and MP2 method and are shown in Fig 4.2. We have carried out benchmark *ab initio* calculation at CCSD(T) level for DFT-D optimized structure of methane-methylindole complex which yields the interaction value as -2.36 kcal mol^{-1}. We have also carried out the calculations of the functionals of Zhao and Truhlar. The I.E. values for the methane-indole complexes are all greater by ~0.5 kcal mol^{-1} than the corresponding ones for the methane-benzene complex[22] but for both complexes the trends in the values for different functionals are very similar. The DFT-D method gives interaction energy within ~0.7 kcal mol^{-1} of that of the benchmark value. The added dispersion correction (-3.12 kcal mol^{-1}) to the BLYP functional makes DFT-D perform better. PWB6K and PW6B95 show better performance in comparison with MPWB1K and MPW1B95 methods. Overall, Zhao and Truhlar

functionals give interaction energies within ~0.5 kcal mol^{-1} of that of the benchmark value. As expected, MP2 method gives interaction energy within ~0.9 kcal mol^{-1} of that of the benchmark value.

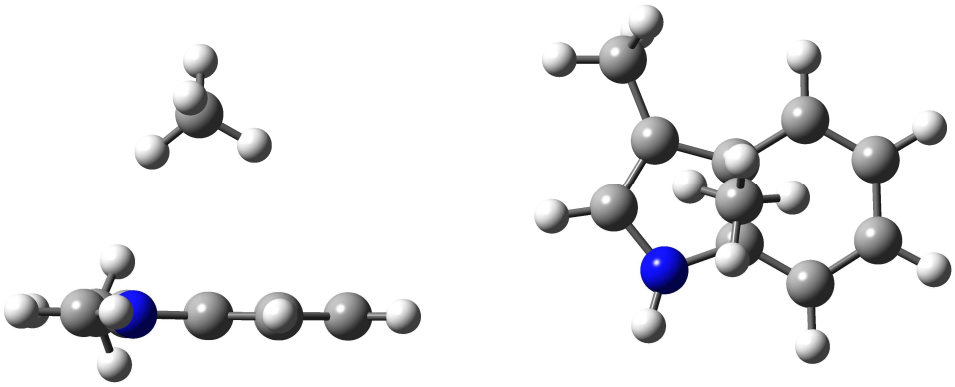


Figure 4.1 Two stereo views of the structure of methane-3-methylindole complexes

Table 4.1 Interaction energy (I.E., kcal mol^{-1}) for methane-3-methylindole complex

Functional	I.E.
DFT-D	-1.69
M05	-1.26
M05-2X	-1.89
M06	-2.01
M06-2X	-2.49
M06-L	-1.88
MPW1B95	-0.87
MPWB1K	-0.95
PW6B95	-1.16
PWB6K	-1.50
MP2	-3.25
Benchmark	-2.36

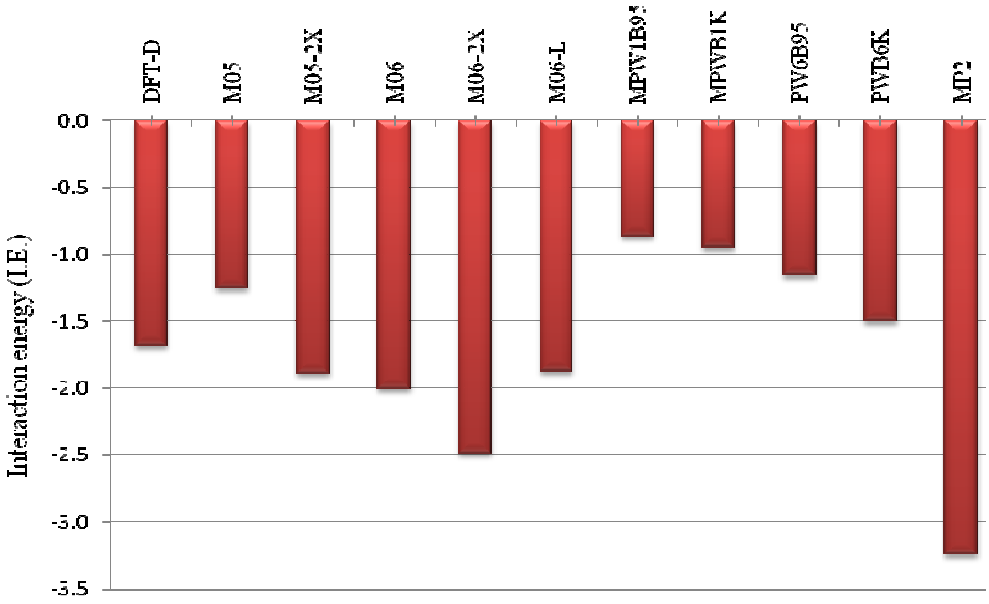


Figure 4.2 Interaction energy of methane-3-methylindole complex calculated using different density functional and the MP2 method

4.2.2 Water-3-methyl indole complexes: Simplest model for OH/π interaction

We have chosen water-3-methylindole complex as a simplest model of OH/π interaction for validating different density functional methods. We have optimized water-3-methylindole complex using DFT-D method starting from different initial positions in which two models involving each hydrogen bond of water pointing to each of the two rings of 3-methylindole and a hydrogen bond pointing to N-H of 3-methylindole. We have located one stable structure in which each hydrogen bond of water pointing to each of the two rings of 3-methylindole. We have carried out

calculations using different density functional methods and MP2 method by taking DFT-D optimized structure of water-3-methylindole complex. The interaction energy values obtained by these methods are summarized in Table 4.2. We have also plotted the I.E. values for the different density functional methods and MP2 method and are shown in Fig 4.4. The DFT-D method gives interaction energy of this complex as -5.31 kcal mol^{-1}. This interaction energy is in contrast to the structure given by B3LYP functional in which water binds to a single ring.[23] We have carried out benchmark calculation at CCSD(T) level using DFT-D optimized structure which yields the interaction energy as -4.83 kcal mol^{-1}. All I.E. values obtained by different methods are within 0.5 kcal mol^{-1} of the benchmark value, except for the M05-2X and M06-2X values, which are within 1.0 and 1.4 kcal mol^{-1} of the benchmark value. The M05, M06, M06-L, PW6B95, PWB6K, MPW1B95 and MPWB1K methods gives the interaction energy within ~0.5 kcal mol^{-1} of the benchmark value. MP2 method slightly overestimates the interaction energy by 1.6 kcal mol^{-1} from the benchmark value. Zhao *et al*.,[24] have investigated the performance of the MPW1B95, MPWB1K, PW6B95 and PWB6K functionals in describing the complexes of water with benzene and methylindole and have found similar interaction energies to the ones reported here.

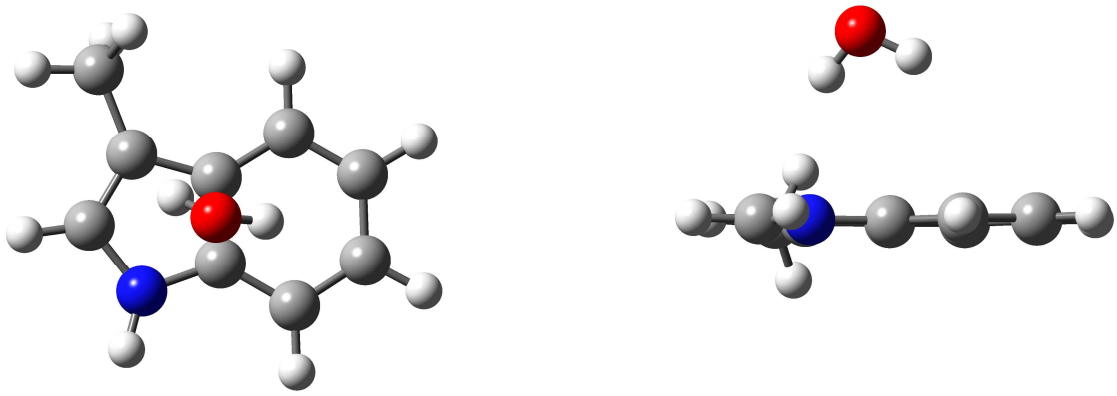


Figure 4.3 Two stereo views of the structure of water-3-methylindole complexes

Table 4.2 Interaction energy (I.E., kcal mol^{-1}) for water-3-methylindole complex

Functional	I.E.
DFT-D	-5.31
M05	-4.60
M05-2X	-5.86
M06	-5.16
M06-2X	-6.21
M06-L	-4.91
MPW1B95	-4.41
MPWB1K	-4.56
PW6B95	-4.63
PWB6K	-5.12
MP2	-6.42

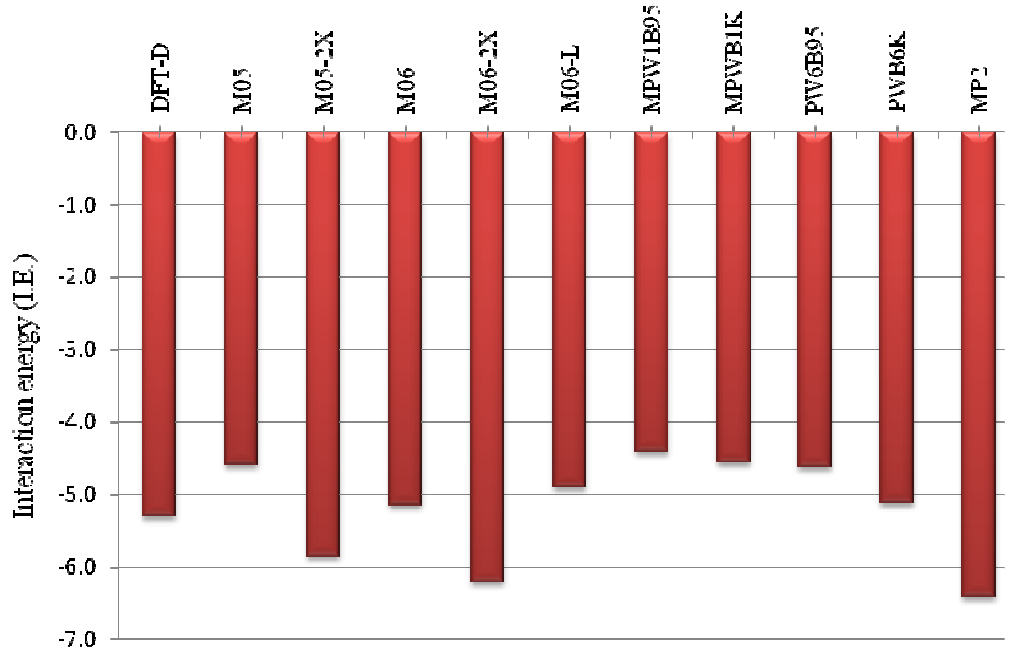


Figure 4.4 Interaction energy of water-3-methylindole complex calculated using different density functional and the MP2 method

4.2.3 Sugar-aromatic complexes-Test of different density functional methods

In the previous chapter, we have discussed the carbohydrate-aromatic interactions studied by DFT-D, PM3-D and PM3-D* methods. Since benchmark calculations are computationally expensive, we have chosen only one carbohydrate-aromatic complex ie. β-glucose-*p*-hydroxy toluene complex **7** (Fig.4.5) for benchmark calculation. The β-glucose-*p*-hydroxy toluene complex **7** is structure from our previous studies, which has three C-H-π interactions. The CCSD (T) interaction energy of complex 7 is -7.42 kcal mol^{-1}. We have performed different DFT functionals and MP2 method for this complex. The interaction energies from different functionals are tabulated in Table 4.3. The MUE values were taken with respect to CCSD (T) interaction energy value. We have plotted the MUE values for convenience (Fig. 4.6). It can be seen from Fig.4.6. that the DFT-D method and M06, M06-2X, M06-L along with BH&H perform the best with a MUE value ranging within ~0.9 kcal mol^{-1}. It is apparent that DFT-D method performs the best overall with very small MUE value of 0.15 kcal mol^{-1}. Both GGA functionals BLYP and PBE show very large MUE values of 9.33 and 5.42 kcal mol^{-1} respectively. Upon inclusion of dispersion correction with the BLYP functional reduced the MUE value from 9.33 to 0.15 kcal mol^{-1} tremendously which show the importance of dispersion term in its description. The most commonly used functional B3LYP gives the MUE value of 7.77 kcal mol^{-1} which shows the failure of functional to study complexes involving dispersion interactions. The Truhlar functionals MPW1B95, MPWB1K, PWB6K and M05 show MUE values in the range of 2.8-4.3 kcal mol^{-1} for this complex. The MUE value for MP2 interaction energy with respect to CCSD (T) interaction energy is 2.56 kcal mol^{-1} which shows the overestimation of MP2 method to study noncovalent interactions. The meta-GGA functional VSXC shows very large MUE of 21.93 kcal mol^{-1}.

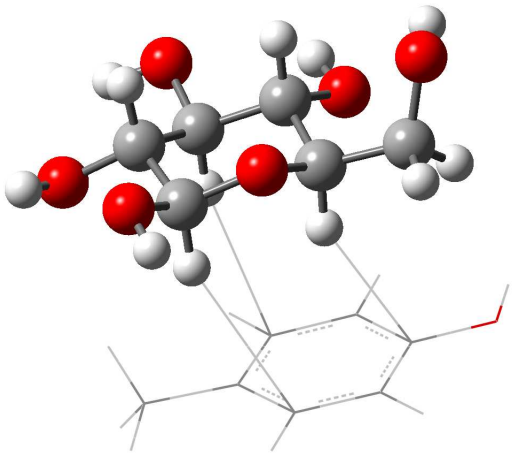


Figure 4.5 DFT-D optimized structure of Complex 7

Table 4.3 Mean Unsigned Error (MUE, kcal mol^{-1}) of interaction energies of Complex 7 for 18 DFT functionals and the MP2, compared to CCSD (T) value.

Functional	MUE	Functional	MUE
BLYP	9.33	MPWB1K	3.95
PBE	5.42	PW6B95	3.90
B3LYP	7.77	PWB6K	2.81
B97-2	7.90	M05	3.94
B98	5.84	M05-2X	1.46
BH&H	0.81	M06	0.63
BH&LYP	6.41	M06-2X	0.62
VSXC	21.93	M06-L	0.87
BMK	5.07	MP2	2.56
MPW1B95	4.34	DFT-D	0.15

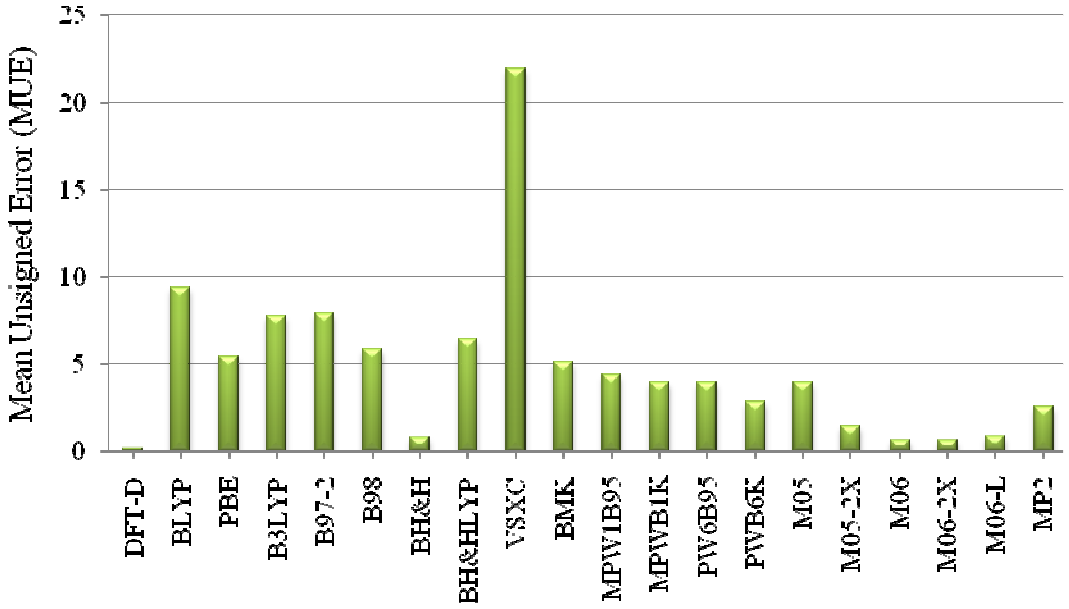


Figure 4.6 Mean Unsigned Error (MUE, kcal mol^{-1}) of interaction energies of complex 7 for 18 DFT functionals, DFT-D and MP2 method with respect to CCSD (T).

With our limited number of benchmark data, we have already shown that the DFT-D method could reproduce the interaction energies very close to that of the benchmark value with a MUE of ~0.6 kcal mol^{-1}.[22] So we decided to test the performance of different DFT functionals and the MP2 method with respect to DFT-D value. We have studied the performance of the 18 functionals and the MP2 method in calculating the interaction energies of the 11 different complexes. This includes the six β-glucose-3-methylindole complexes and five β-glucose-p-hydroxy toluene complexes (Figs, chapter 3). For all the calculations of different functionals, we have employed DFT-D optimized structures. The interaction energies calculated for the 11 complexes using the DFT-D method, different density functionals, and the MP2 method are given in Table 4.4. The mean unsigned errors (MUE) of I.E. values for different functionals and MP2 method with respect to the DFT-D values are given in Table 4.5. We have also plotted the MUE values against different functionals (Fig.4.7). We can see a similar trend in the MUE values when compared with the MUE values with respect to

CCSD (T) as shown in Fig 4.6. The functionals BLYP, PBE, B3LYP, B97-2, B98, BMK and BH&HLYP underestimate the interaction energy by 5-9 kcal mol^{-1}. VSXC overestimates the interaction energy by a larger amount, giving an MUE value of 20.44 kcal mol^{-1}. The Truhlar hybrid meta GGA functionals MPW1B95, MPWB1K, PW6B95 and M05 underestimate interaction energy values within 3.9-4.4 kcal mol^{-1}, which show some improvement in their description of nonbonded interactions. The hybrid meta GGA functional PWB6K shows better performance giving interaction energies with a MUE value of 2.85 kcal mol^{-1}. The functional BH&H is more successful, overestimating the interaction energy by only 1.57 kcal mol^{-1}. Among the M0x functionals which were developed by Zhao and Truhlar recently, except M05, all other functionals shows very good performance, giving interaction energies within ~1.3 kcal mol^{-1}. We find that the M06 group of functionals perform better compared to M05 group of functionals, with all three giving interaction energies within ~1 kcal mol^{-1}, with M06-2X and M06 performing better with MUE values of 0.46 and 0.63 kcal mol^{-1} respectively. Among all the functionals we used, VSXC is the functional which performs poor to describe noncovalent interactions with MUE value of 20.44 kcal mol^{-1}. Therefore VSXC is clearly not the functional to study noncovalent interactions.

Table 4.4 Interaction energies (kcal mol^{-1}) for the full dataset of 11 complexes, for 18 functionals, DFT-D and MP2 method

Complex	DFT-D	BLYP	PBE	B3LYP	B97-2	B98	BH&H	BH&HLYP	VSXC	BMK
1	-9.78	0.79	-3.57	-1.01	-0.82	-3.12	-10.94	-2.61	-36.28	-4.41
2	-10.55	-1.44	-5.29	-3.14	-2.74	-4.85	-12.52	-4.66	-30.05	-6.05
3	-10.20	-0.66	-4.56	-2.20	-1.88	-4.02	-11.16	-3.49	-28.60	-4.97
4	-10.33	-0.07	-4.31	-1.82	-1.46	-3.76	-11.69	-3.32	-32.62	-5.25
5	-11.46	-0.49	-5.09	-2.60	-2.30	-4.69	-13.71	-4.51	-36.31	-6.54
6	-9.96	0.08	-4.04	-1.84	-1.60	-3.81	-11.70	-3.54	-30.43	-4.98
7	-7.27	1.91	-2.00	0.35	0.48	-1.58	-8.23	-1.01	-29.35	-2.35
8	-10.56	-4.36	-7.29	-5.51	-4.81	-6.49	-13.15	-6.55	-21.19	-6.96
9	-9.77	-0.57	-4.42	-2.20	-2.38	-3.91	-11.61	-3.66	-30.24	-4.84
10	-9.04	-0.65	-4.17	-2.02	-1.41	-3.56	-10.64	-3.29	-29.06	-4.83
11	-7.16	1.40	-2.32	-0.14	-0.04	-1.95	-8.18	-1.50	-26.81	-2.43
Complex	MPW1B95	MPWB1K	PW6B95	PWB6K	M05	M05-2X	M06	M06-2X	M06-L	MP2
1	-5.03	-5.50	-5.47	-6.72	-5.35	-8.52	-9.45	-9.51	-9.08	-13.65
2	-6.18	-6.67	-6.54	-7.79	-6.74	-9.42	-10.21	-10.30	-9.68	-13.98
3	-5.24	-5.60	-5.66	-6.74	-5.99	-8.49	-8.86	-9.24	-8.42	-13.05
4	-5.53	-5.97	-5.94	-7.15	-5.86	-8.80	-9.28	-9.84	-8.78	-13.37
5	-6.53	-7.17	-6.86	-8.38	-6.75	-10.42	-11.19	-11.45	-10.43	-15.70
6	-4.87	-5.44	-5.27	-6.65	-5.81	-8.73	-9.66	-9.46	-8.76	-13.83
7	-3.08	-3.47	-3.52	-4.61	-3.48	-5.96	-6.79	-6.80	-6.55	-9.98
8	-7.40	-7.68	-7.64	-8.54	-7.81	-9.15	-9.39	-9.73	-9.43	-11.17
9	-5.26	-5.67	-5.70	-6.88	-5.94	-8.37	-9.23	-9.30	-8.95	-12.44
10	-5.14	-5.47	-5.57	-6.59	-5.46	-7.51	-8.40	-8.56	-8.47	-11.00
11	-3.24	-3.65	-3.66	-4.76	-3.70	-6.14	-6.72	-6.85	-6.54	-10.00

Table 4.5 Mean Unsigned error (MUE, kcal mol^{-1}) of interaction energies for full dataset (11 structures, Table 4.4), for 18 DFT functionals and the MP2, compared to DFT-D values.

Functional	MUE	Functional	MUE
BLYP	9.29	MPWB1K	3.97
PBE	5.37	PW6B95	4.03
B3LYP	7.66	PWB6K	2.85
B97-2	7.89	M05	3.93
B98	5.85	M05-2X	1.34
BH&H	1.57	M06	0.63
BH&LYP	6.18	M06-2X	0.46
VSXC	20.44	M06-L	1.00
BMK	4.77	MP2	2.92
MPW1B95	4.44		

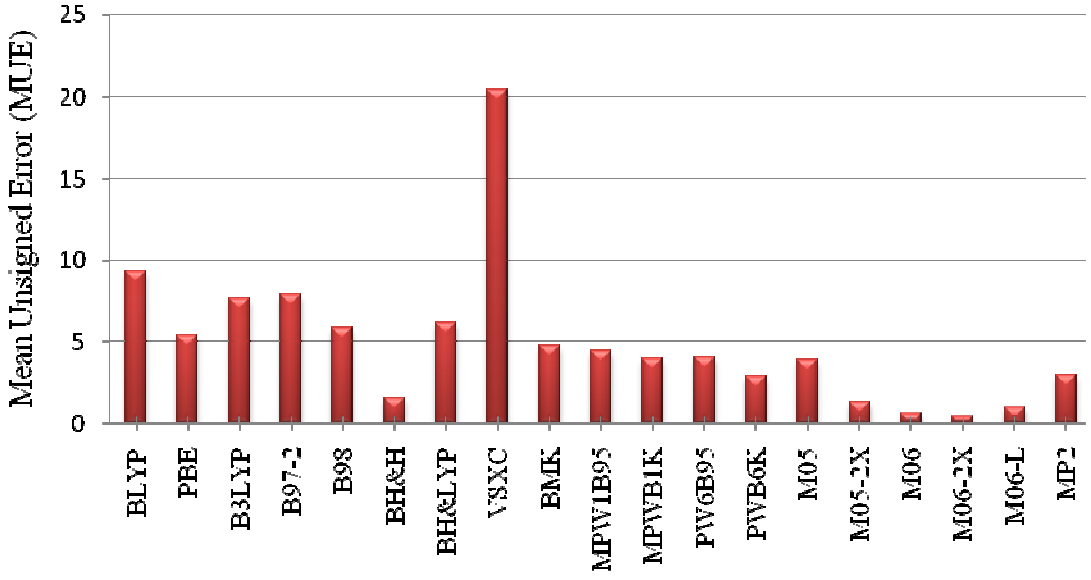


Figure 4.7 Mean Unsigned Error (MUE, kcal mol^{-1}) of interaction energies for a database of 11 different complexes, for 18 DFT functionals, and the MP2 method with respect to DFT-D values.

4.3 Conclusions

We have tested the performance of different density functionals for the accurate description of carbohydrate-aromatic π interactions. We have chosen a small database containing interaction energies of a small number of representative complexes methane-indole and water-indole involving CH/π and OH/π interactions with aromatic molecule respectively. We have shown the performance of both the DFT-D method and a variety of DFT functionals including the recently developed Truhlar functionals. From our studies, we have found that DFT-D method is particularly successful in describing carbohydrate-aromatic interactions with a MUE value of less than 1 kcal mol^{-1}. We have also found that the most commonly used standard density functionals like BLYP and B3LYP failed to describe noncovalent interactions. Among all hybrid GGA functionals such as B3LYP, B97-2, B98, BH&H and BH&HLYP, BH&H only perform best to describe noncovalent interactions. Upon inclusion of dispersion term, the performance of BLYP functional is improved and gives the interaction energy with MUE value within ~1 kcal mol^{-1}. It is evident from the MUE values that meta-GGA functional VSXC shows very large overestimation of the interaction and is not recommended to study noncovalent interactions. We have found that M0x functionals perform better overall, expecially M06 group of functionals shows better performance over M05 functionals. M06 functionals give interaction energies within MUE value of ~1 kcal mol^{-1}. The hybrid meta GGA functionals M06 and M06-2X can be used as an alternative to study the carbohydrate-aromatic interactions as they reproduce the interaction energy within 0.5-1.3 kcal mol^{-1}. It is evident that MP2 method overestimates the interaction energy by ~2-3 kcal mol^{-1}.

4.4 References

1. Meyer, E. A.; Castellano, R. K.; Diederich, F. *Ang. Chemie- Int. Ed.*, **2003**, *42*, 1210.

2. Gabius, H. J.; Siebert, H. C.; André, S.; Jiménez-Barbero, J.; Rüdiger, H. *Chem. Bio. Chem.*, **2004**, *5*, 740.

3. Jurečka, P.; Šponer, J.; Černy, J.; Hobza, P. *Phys. Chem. Chem. Phys.*, **2006**, *8*, 1985.

4. Grimme, S. *J. Comp.Chem.,* **2004**, *25*, 1463.

5. McNamara, J. P.; Hillier, I. H.; Shan, X. *Phys. Chem. Chem. Phys.*, **2007**, *9*, 2362.

6. Sharma, R.; McNamara, J. P.; Raju, R. K.; Vincent, M. A.; Hillier, I. H.; Margado, C. A. *Phys. Chem. Chem. Phys.*, **2008**, *10*, 2767.

7. Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.*, **2006**, *125*, 194101.

8. Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Phys.*, **2005**, *123*, 1.

9. Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory. Comp.*, **2006**, 2, 364.

10. Zhao, Y.; Tishchenko, O.; Truhlar, D. G. *J. Phys. Chem. B.*, **2005**, *109*, 19046.

11. Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A*, **2004**, *108*, 6908.

12. Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A.*, **2005**, *109*, 5656.

13. Zhao, Y.; Truhlar, D. G. *Phys. Chem. Chem. Phys.*, **2005**, *7*, 2701.

14. Zhao, Y.; Truhlar, D. G. *J. Chem. Theory. Comp.*, **2006**, *2*, 1009.

15. Zhao, Y.; Truhlar, D. G. *J. Chem. Theory. Comp.*, **2007**, *3*, 289.

16. Zhao, Y.; Truhlar, D. G. *Theo. Chem. Acta.*, **2008**, *120*, 215.

17. Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.*, **2008**, *41*, 157.

18. Screen, J.; Stanca-Kaposta, E. C.; Gamblin, D. P.; Liu, B.; Macleod, N. A.; Snoek, L. C.; Davis, B. G.; Simons, J. P. *Ange. Chemie-Int. Ed.*, **2007**, *46*, 3644.

19. Stanca-Kaposta, E. C.; Gamblin, D. P.; Screen, J.; Liu, B.; Snoek, L. C.; Davis, B. G.; Simons, J. P. *Phys. Chem. Chem. Phys.*, **2007**, *9*, 4444.

20. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota,

K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. GAUSSIAN 03 (Revision C.02), Gaussian, Inc., Wallingford, CT, **2004**.

21. Ringer, A. L.; Figgs, M. S.; Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A*., **2006**, *110*, 10822.

22. Rajesh, R. K.; Ramraj, A.; Hillier, I. H.; Vincent, M. A.; Burton, N. A. *Phys. Chem. Chem. Phys.*, **2009**, *11*, 3411.

23. Mourik, T. V. *Chem. Phys.*, **2004**, *304*, 317.

24. Zhao, Y.; Tishchenko, O.; Truhlar, D. G. *J. Phys. Chem. B.*, **2005**, *109*, 19046.

Chapter 5

Theoretical investigation of organic pollutants on graphene sheets and nanotubes

5.0 Abstract

We have investigated the role of π-π interaction in the adsorption of organic pollutants on graphene sheets and on nanotubes using semi-empirical parameterization scheme implemented with dispersion correction term (PM3-D). Energetic and structural aspects of adsorption of 10 different organic pollutants such as 1-naphthylamine, 1-naphthol, 2-naphthol, naphthalene, 2-phenylphenol, 2, 4-dichlorobenzne, 1, 3-dichlorobenzene, pyrogallol, catechol and phenol on C96, [6, 6] and [5, 0] have been studied in detail. The effect of aromaticity and -OH groups have been analysed in detail. We have calculated the intermolecular separations, interaction energies and dispersion contribution of the complexes. Interaction energies increase with the number of aromatic rings such as 1-naphthylamine> 2-naphthol> 2, 4-dichlorophenol. Interaction energies of -OH substituted pollutants are in the order of pyrogallol> catechol> phenol. Mulliken population analyses have been carried out for all complexes and a small degree of charge has been found to be transferring from adsorbates to π-surfaces. Dispersive contribution has been found to be dominant for stabilizing complexes. Without dispersive contribution, complexes would be unbound. The trend of dispersive contribution follows the same order of interaction energies. All our calculations suggest that the intermolecular forces such as π-π interactions indeed dominant in the adsorption of aromatic pollutants on graphene and nanotubes.

5.1 Introduction

Over the past three decades, there has been increasing concern and research about global warming and environmental pollutions which cause major changes in global warming. According to estimations from world health organization (WHO), ~25% of the diseases to human-being occur due to the prolonged exposure to environmental pollutions. Improper management of solid wastes which may be toxic, infectious or radioactive and lack of regulations and proper disposal of such solid waste are main causes to environmental pollutions and great risks to public health particularly in developing countries. The connections between environmental pollutions and global warming and health risks are yet to be well understood and analysed. Exposures to environmental pollutions are major challenges due to the lack of detailed monitoring and pollutants from higher number of sources. Environmental pollutants can be mainly classified as i) heavy metals and ii) organic pollutants. Heavy metals are in lower concentrations in nature. However, in contaminated environments, heavy metals are in higher concentrations. Upon deposition, these heavy metals are not easily degraded which persist in the environment for many years and poison humans through inhalation, ingestion and skin absorption. Heavy metals that are of concern include lead, mercury, cadmium, arsenic, chromium, zinc, nickel and copper.

Organic pollutants are also persistent and non-biodegradable compounds. Organic pollutants can be accumulated in the food chain especially in the fish and livestock which cause serious health risks to humans. Poor solubility of these organic pollutants and easy accumulation in food products and fatty tissues are main challenges in elimination of these organic pollutants. Organic pollutants include aldrin, dieldrin, dichlorodiphenyl-trichloroethane (DDT), dioxin, xylenes, endrin, heptachlor, toxaphene, mirex, pesticides, polychlorinated biphenyls, resorcinol and poly aromatic hydrocarbons (PAHs).

Carbon nanotubes are playing vital role in many areas of nanotechnological applications. They have been widely used as catalysts and biosensors in DNA bionanotechnology and as electrodes in batteries and supercapacitors. Recently, carbon

nanotubes (CNTs) have been considered as adsorbent to remove environmental contaminants due to their large surface area. These CNTs have been examined for removal of inorganic and organic contaminants from water and gases. The sorption of contaminants on CNTs affects the fact of environmental contaminants, as well as the properties of CNTs themselves. Therefore, it has become crucial to understand the properties of sorption of contaminants on CNTs. There are different kind of organic contaminants as mentioned earlier in environment such as contaminants with hydrophilic functional groups (for eg. cyclohexanol), with aromatic ring (for eg. naphthalene) or with both (for eg. naphthol). Hydrophobic interactions have been considered to play main role in the sorption of hydrophobic organic contaminants on CNTs. The sorptions of organic contaminants have been studied on activated carbonaceous materials, graphene sheets and on CNTs. The mechanism of sorption of contaminants on CNTs is believed to be involving many interactions such as hydrophobic and π-π interactions. Understanding these interactions has become necessary to use CNTs efficiently to remove environmental contaminants. There have been number of experimental studies to investigate the sorption of inorganic and organic contaminants such as butane, trihalomethanes, dioxin, chlorophenols and poly aromatic hydrocarbons (PAHs). In all these studies, CNTs have been found more efficient than activated carbon for soprtion of these contaminants. Aromatic compounds among the contaminants have been found to be of more interest and crucial.[14]

SWNT-ferrocene nanohybrid can be used as a sensor for detecting anionic species due to hydrogen bond interactions with anionic species. The detection of ionic pollutants is very important in the field of environmental chemistry.

5.2 Previous studies

Gadupudi Purnachandra Rao et al, have studied the sorption of divalent cations such as Pb^{2+}, Cd^{2+}, Cu^{2+}, Ni^{2+}, Zn^{2+} on CNTs. Langmuir adsorption and Frendulich isotherm have been used for these studies.[1] Grulke et al, have studied the adsorption of butane on multi walled carbon nanotube by using adsorption isotherm modified by BET equation at room temperature. It has been found that sorption of butane on outer nanotube with smaller diameter is higher than on nanotubes with bigger diameter.[2] Highly toxic compound dioxin and other related compounds have been studied by Yang et al by using temperature-programmed desorption (TPD).[3] The strong interaction between dioxin and CNTs has been found which enhances the removal of dioxin by CNTs. It has also been found that CNTs are more effective for the interaction with dioxin than activated carbon. Peng et al, have studied the interaction of 1, 2-dichloro benzene on graphitized carbon and on CNTs and found that the time needed for adsorption of 1, 2-dichloro benzene on CNT is only 40 mins. It has also been found that CNTs can be used in wide range of pH 3.0-10.0.[4] The adsorption of resorcinol on multi walled carbon nanotube (MWCNT) has been studied by Liao et al. Langmuir and Frendulich isotherms have been used to study the adsorption. Short contact time and wide pH range have been found for MWCNTs which makes them as more potential for these adsorptions. MWCNTs with acid groups have been found to be less potential for the adsorption of resorcinol because of the increased electrostatic repulsion. It has been found that carboxylic groups weaken the π-π interaction and water adsorption. It has also been found that the adsorption increases with the increasing number of hydroxyl groups.[5] Yang et al have investigated the adsorption of various poly aromatic hydrocarbons (PAHs) such as naphthalene, pyrene and phenantharene on fullerene, SWCNTs and MWCNTs. Adsorption affinities have been found to be increasing with the order of the size of the aromatic sheet i.e. naphthalene <phenanthrene <pyrene. It has been found that the adsorption of phenanthrene was 2-4 times higher than on fullerene following the order of SWCNT> MWCNT>> fullerene.[6] Xia et al, also have reported that the adsorption on nanotube is 2 order of magnitudes higher than natural soil/sediments.[7] In all these above experiments, the

main investigation has been studying the properties of adsorbents such activated carbon or CNTs or MWCNTs. Later only, the properties of adsorbates especially aromatic compounds has attracted the lot of attention of researchers since most of the organic contaminants are aromatic compounds. Therefore, the detailed understanding of these aromatic compounds has become necessary. However, very few experimental studies have been conducted to investigate the fundamental properties of these aromatic compounds.

Xia et al[7] and Crespo et al[8] have studied the aromatic and nonaromatic compounds in gas and aqueous phase. It has been found that benzene has higher affinity for CNTs than cyclohexane. Giannozzi[9] in his study has reported that the sorption of organic compounds to CNTs as physical interaction without charge transfer between the adsorbates and adsorbents. Suzana Gotovac at al,[10] have studied the liquid-phase adsorption of tetracene and phenanthrene on SWCNT by adsorption isotherms and X-ray photoelectron spectroscopy. The adsorption of tetracene was found to be six times greater than that of phenanthrene. The difference between the adsorption of tetracene and phenanthrene was caused by the curvature of the nanotube which reduces the contact between phenanthrene and nanotube. Higher frequency shift in radial breathing mode (RBM) of the Raman band of the SWCNT has also been observed which is caused by the adsorption of phenanthrene and tetracene and is the indication of π-π interaction between adsorbates and adsorbent. Zhang et al[11] have investigated the adsorption of anthracene and derivatives of anthracene on SWCNT. The adsorption of 9, 10- anthracenedicarbonitrile and 9-anthracene-methanol has been found to be strong. The adsorptions of anthracene and their derivatives on SWCNT have been found to be reversible, eventhough they are strong. Wang et al[12] have modified MWCNTs with poly [2-methoxy, 5-(2'-ethyl-hexyloxy)-*p*-phenylene vinylene] (MEH-PPV) and poly [vinylpyrrolidone] (PVP) to form two soluble MWCNTs/polymer supramolecular composites. It has been observed that MWCNTs/MEH-PPV exhibits much larger untrafast third-order optical nonlinearity than MWCNTs/PVP. The reason for the enhancement of MWCNTs/MEH-PPV is π-π interaction between the polymer MEH-PPV and the sidewall of carbon nanotubes. All these studies have indicated the presence of π-π interaction between adsorbates and

adsorbents. These aromatic interactions depend on many factors such as size of the aromatic ring, substituents on the ring and the shape of the aromatic ring. Substituent effect on the aromatic-aromatic interactions has been studied by Hunter et al.[13] Conformational studies have been carried out on the H-bonded supramolecular zipper complexes which is useful to lock the geometry of two aromatic rings in offset stacking conformation. Along with the electrostatic properties of the ring surfaces, electrostatic interactions of the ring substituents play an important role in the aromatic-aromatic interactions.

Recently, there have been two experimental studies on the adsorption of organic contaminants which contain hydroxyl and amino substitutions. Chen et al,[14] have studied the adsorption of hydroxyl and amino substituted aromatics on graphite and on nanotubes, lightly oxidized nanotubes and on deeply oxidized nanotubes by using adsorption isotherms. Adsorption affinities for adorbents increase in the following order graphite> nanotube> lightly oxidised nanotube> deeply oxidised nanotube. Molecules substituted with hydroxyl or amino group such as 2-naphthol and 1-naphthyl amine were found to be adsorbed more on nanotube than molecules without substitution such as naphthalene. Likewise, molecules with two aromatic rings such as naphthalene or 2-naphthol were found to be adsorbed more than molecules with one aromatic ring such as 1,3-dichloro benzene. It has been found that adsorption affinities of different adsorbates to nanotubes follow the order of 1, 3-dichlorobenzene/2,4-dichlorophenol/naphthalene<2-naphthol<<1-naphthyl amine.[14] It has been suggested that the mechanism behind this trend is caused by electron donor acceptor (EDA) interaction. This EDA π-π interaction arises between π electron-rich aromatic rings of adsorbates and π electron depleted region of nanotubes or graphite. It has also been reported that another important mechanism of strong adsorption of 1-naphthyl amine is lewis acid base interaction where NH2 plays as lewis base and O-containing groups on nanotubes play as lewis acid.

In another study by Lin et al,[15] role of aromaticity and subsitution of hydroxyl groups have been investigated with adsorbates containing different number of hydroxyl groups and aromatic rings by using adsorption isotherms. Cyclohexanol, phenol,

catechol, pyrogallol, 2-phenyl phenol, 1-naphthol and napthalene were selected for their study. Sorption affinities of these compounds to CNTs have been found to be increasing in the order of cyclohexanol < phenol < catechol < pyrogallol < 2-phenyl phenol < 1-naphthol.[15] It is apparent that sorption affinities increase with the increasing number of –OH substitutions and aromatic rings. It has been suggested that the main mechanism of the sorption is π-π interaction between adsorbates and CNTs. It has also been reported that the substitution in the aromatic ring can strenghthen the π-π interaction.

The above explanation of EDA interaction between adosrbates and graphite/nanotubes given by Chen et al,[14] was challenged by Jiang et al.[16] The π-π mechanism given for higher adsorption affinities of 2,4-dichloro phenol and 2-naphthol than 1,3-dichloro benzene and napthalene was commented by Jiang et al that there is another mechanism such as oxidative coupling for the enhance in adsorption affinities of phenolic compounds. Jiang et al suggested that at higher pH range, -OH will dissociate as O^- which will enhance the adsorption of 2-naphthol.[16]

In response to the comment of Jiang et al, Chen et al[17] argued that oxidative coupling could not be major mechanism for higher adsorption of 2-naphthol by conducting some experiments under oxic and anoxic conditions. Adsorption coefficient kd and mass recovery were compared in these both conditions at adsorption equilibrium. It has been argued that if oxidative coupling is significant, the kd value would be higher under oxic condition and mass recovery would be lower. However it has been found that kd value is similar at a given pH under oxic and anoxic conditions for both the compounds 2, 4-dichloro phenol and 2-naphthol. Therefore it has been suggested that oxidative coupling could not be major mechanism for the higher adsorption affinities of 2, 4-dichloro phenol and 2-naphthol and it has been suggested that further studies are needed to reveal key factors controlling oxidative coupling of phenolic compounds with nanotubes.[17]

There is only limited number of theoretical studies having been carried out to study the role of π-π interactions in the adsorption of organic contaminants on CNTs. This

shows the fundamental need for understanding the mechanism behind the adsorption of these pollutants on π-surfaces. Quantum mechanical (QM) calculations have become possible for studying these intermolecular interactions such as π-π interactions.

5.3 Results and Discussions

We have used quantum mechanical calculations to study the interaction of organic pollutants on graphene sheets and on nanotubes. We have chosen range of organic pollutants with functional groups such as hydroxyl, amino or chloro substituted aromatic molecules. We have used the models of finite clusters rather than using infinite periodic solids. The organic pollutants we have chosen were 1-naphthol (1-NATH), 2-naphthol (2-NATH), 2-phenylphenol (2-PHPH), 2,4-dichloro phenol (2,4-DCP), phenol (PH), catechol (CT), pyrogallol (PY), 1-naphthyl amine (1-NALA), 1,3-dichloro benzene (1,3-DCB) and naphthalene (NAPH). The organic pollutants are shown in Figure 5.1. We have studied the interaction of aromatic pollutants on larger graphene sheets, $C_{96}H_{24}$ as a model of graphene surface or graphite. Since the change in interaction energies of nucleic acid bases on graphene sheets where the size is bigger than C_{96} is less than 1 kcal/mol, $C_{96}H_{24}$ is a good model to study these interactions. We have used two different nanotubes such as [6, 6] arm-chair nanotube and [5, 0] zig-zag nanotube which have 300 and 140 carbon atoms respectively, with the length of 30 Å. We have studied the adsorption of 10 organic pollutants on these models. Initial structures were constructed by placing organic pollutants, with the centre of the ring for molecules with one ring or centre of the bridging bond for molecules with two rings above one of the six equivalent central carbon atoms of the graphene sheet or nanotube at 3.5-4.0 Å respectively. The geometries of the intermolecular complexes were minimized at the PM3-D level. The interaction energies, dispersion energies and intermolecular distances were calculated and reported in Table 5.1-5.4. We have also evaluated the interaction energies at the DFT-D level using PM3-D minimized structures.

5.3.1 Adsorption of organic pollutants on graphene sheet

We have found that the intermolecular distances between aromatic ring of the adsorbates and the central benzene ring of the graphene sheet are almost 3.30-3.34 Å except 2PHPH for which it is 3.68 Å. The minimized complexes of organic pollutants on graphene sheet are shown in Figure 5.2. In the minimized complexes, the centre of the ring of the organic pollutants have been found to be on top of the carbon of the graphene sheet, which structures have been found to be more stable. The interaction energies of organic pollutants on graphene are ranging from 18.0 kcal/mol to 26.4 kcal/mol except *p*H for which is bit lower, 15.8 kcal/mol. The interaction energies are found to be higher for pollutants with two aromatic rings following by molecules with one aromatic ring substituted by amino and hydroxyl groups. Thus the trend of interaction energies is as follows 1-NALA> 2-NATH≥ 1-NATH> 2-PHPH> NAPH> PY> 2, 4-DCP> 1, 3-DCB> CT> PH. In line with Chen et al study,[14] 1-NALA and 2-NATH interact stronger on graphene sheet than NAPH and 2-PHPH. The reason for the lower interaction energies of 2-PHPH compared to other naphthalene pollutants is because of one of the ring in 2-PHPH is almost perpendicular to surface. It has been found that molecules with higher number of –OH groups interact stronger than molecules with less number of –OH groups. Therefore, the trend of interaction energies of molecules with –OH groups, PH< CT< PY is in agreement with the finding of Lin et al.[15] We have also evaluated the interaction energies at DFT-D level using PM3-D minimized structures. We have found that the interaction energies of adsorbates on graphene sheets obtained at DFT-D level are only ~2 kcal/mol less than PM3-D interaction energies on graphene. Therefore, interaction energies of adsorbates on nanotubes have been obtained only by using PM3-D method.

5.3.2 Adsorption of organic pollutants on nanotube

The intermolecular distances between aromatic ring of the adsorbates and the central benzene ring of the nanotube are ranging from 3.1-3.7 Å. The intermolecular distances of adsorbates with –OH groups from nanotubes are found to be lower than from graphene. In minimized complexes, similar to graphene, adsorbates have been found to be interacting through the centre of the ring on top of the carbon or on centre of the bridging bond of benzene of nanotube. All optimized complexes of adsorbates on nanotubes [6, 6] and [5, 0] are shown in Figures 5.3 and 5.4 respectively. The interaction energies of adsorbates on [6, 6] nanotube are ranging from 13.1 kcal/mol to 17.9 kcal/mol. For adsorbates on [5, 0] nanotube, the interaction energies are ranging from 10.2 kcal/mol to 16.1 kcal/mol. The interaction energies on [5, 0] nanotube are approximately 1.5-2.0 kcal/mol lower than the interaction energies on [6, 6] nanotube. The higher interaction energies of adsorbates on [6, 6] nanotube compared to [5, 0] nanotube are because of the larger circumference of [6, 6] nanotube which allows the adsorbate molecule to interact stronger through π-π interaction. Since the interaction energies of adsorbates on graphene sheet obtained by DFT-D are only ~2 kcal/mol less than the interaction energies obtained by PM3-D, the interaction energies of adsorbates on nanotube are obtained only by PM3-D method. The trend of interaction energies of adsorbates on nanotube follows the order of 1-NALA> 2-NATH, 2-PHPH> 1-NATH> NAPH> 2, 4-DCP> 1, 3-DCB> PY> CT> PH.

We found that the dispersive interactions are dominant in all three systems. Without dispersion contribution, all other energies would be repulsive and so that complexes would be unbound. We have also noted that the trend of interaction energies of all adsorbates on three models follows the same that of the dispersive interactions. We have conducted Mulliken population analysis to investigate the charge transfer between adsorbates and π-surfaces. A small degree of electron transfer in the order of 10^{-3}e have been found to be transferred from the adsorbates to π-surface of nanotube or graphene sheet. There is no correlation between the quantity of charge transfer and the calculated interaction energy. Therefore, these calculations do not support the

suggestion that the degree of EDA is responsible for the variation in the interaction energy of the different adsorbates. Interaction energies of adsorbates on nanotubes have been found to be less than the interaction energies on graphene sheet. However, the trend of interaction energies is same for graphene sheet and both nanotubes [6, 6] and [5, 0]. The reason for higher interaction energies on graphene sheet is because of the flat surface which allows the aromatic molecules to interact with π-surface and enhances the π-π interaction.

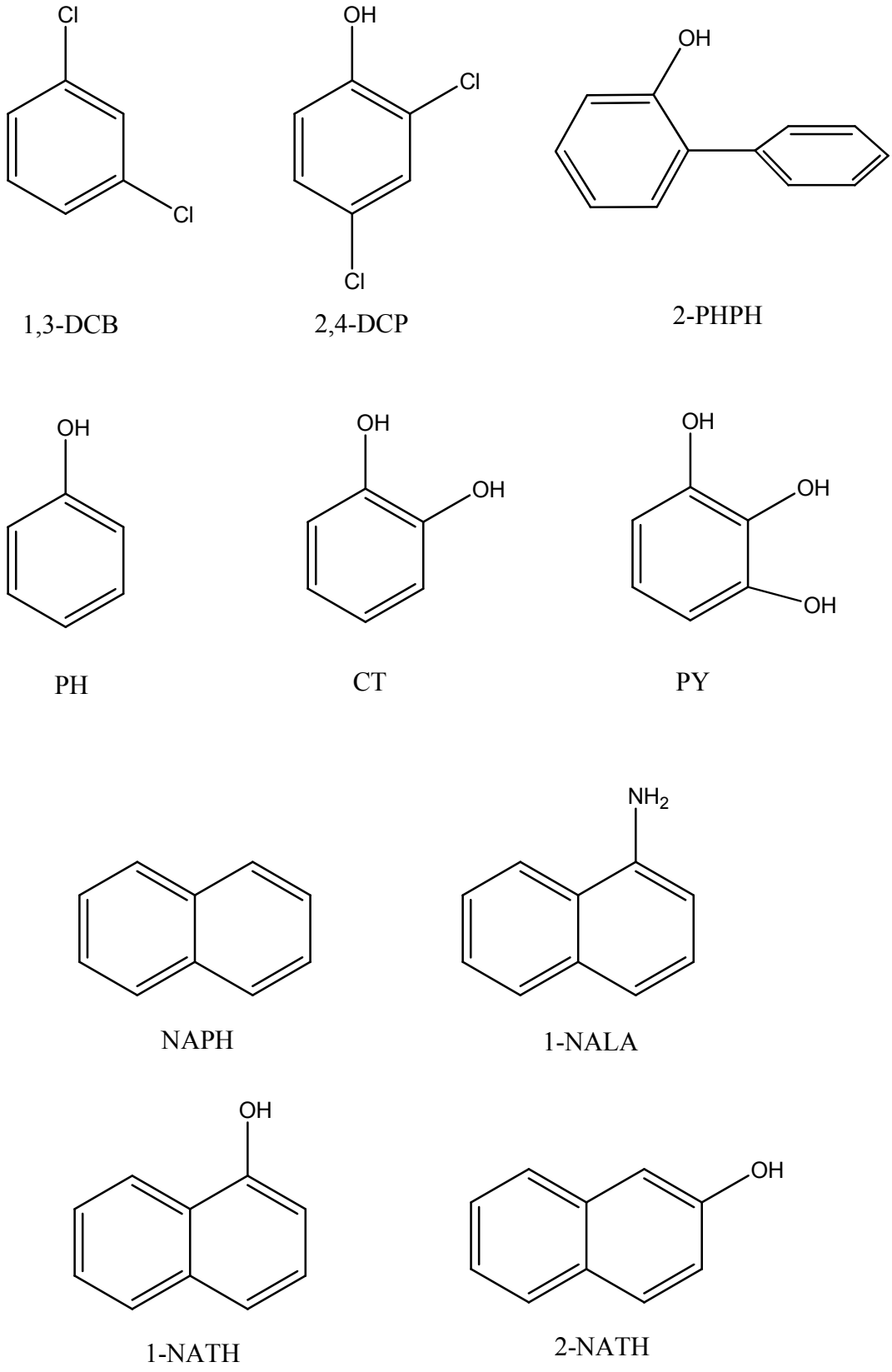
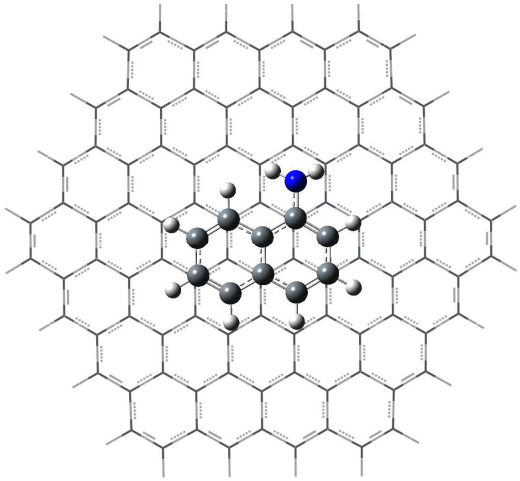
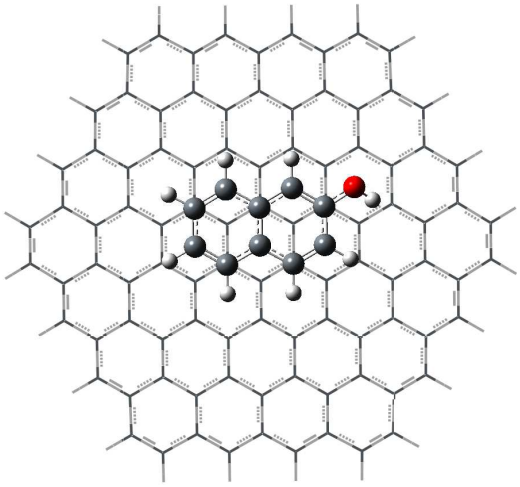


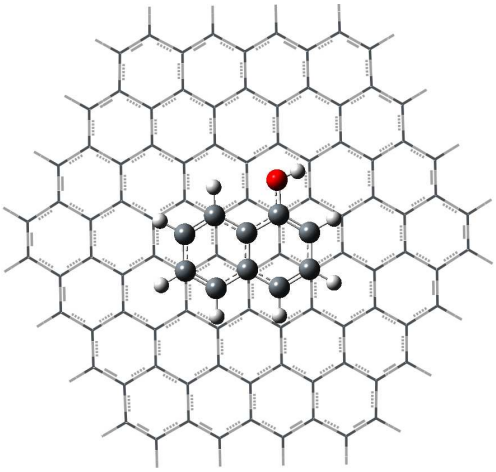
Figure 5.1 Aromatic pollutants



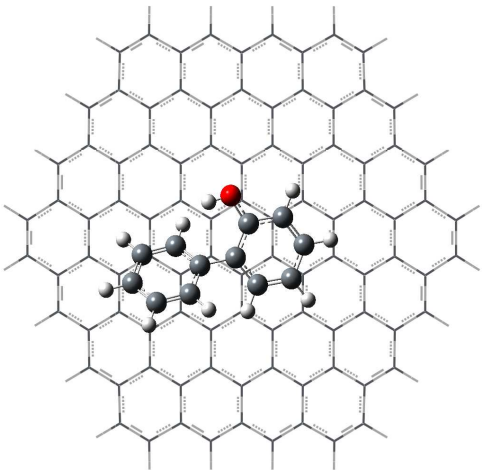
1-NALA-C96



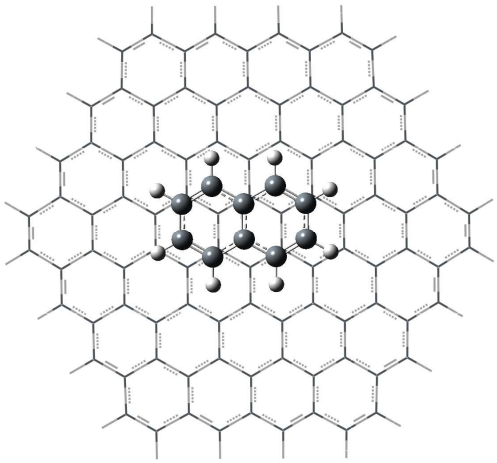
2-NATH-C96



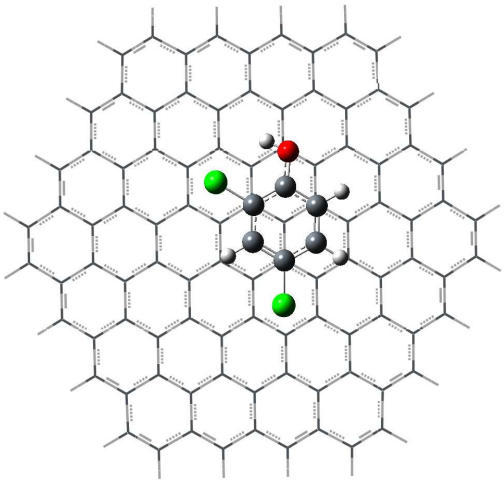
1-NATH-C96



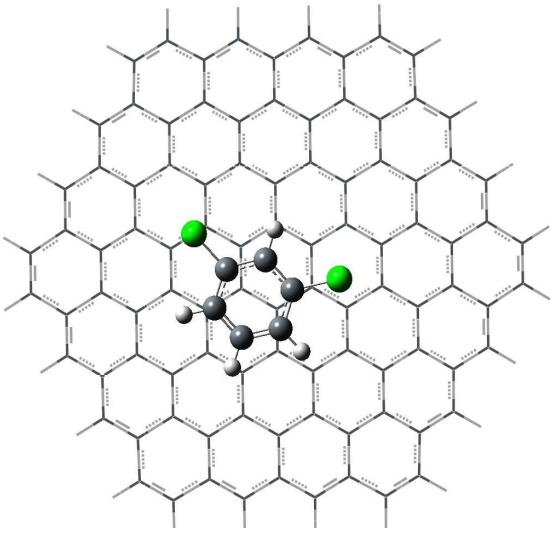
2-NATH-C96



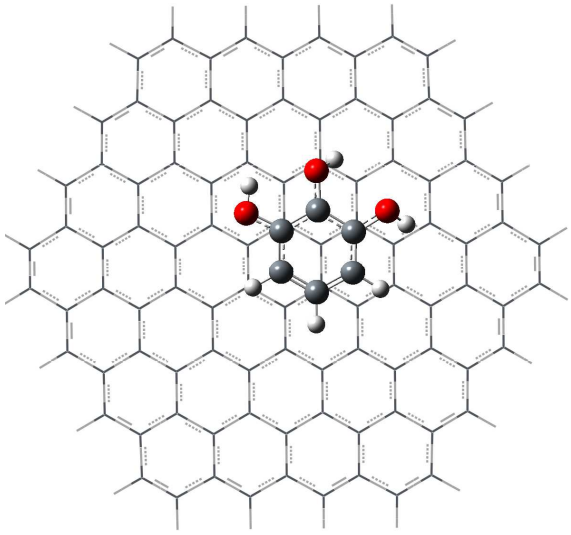
2-PHPH-C96



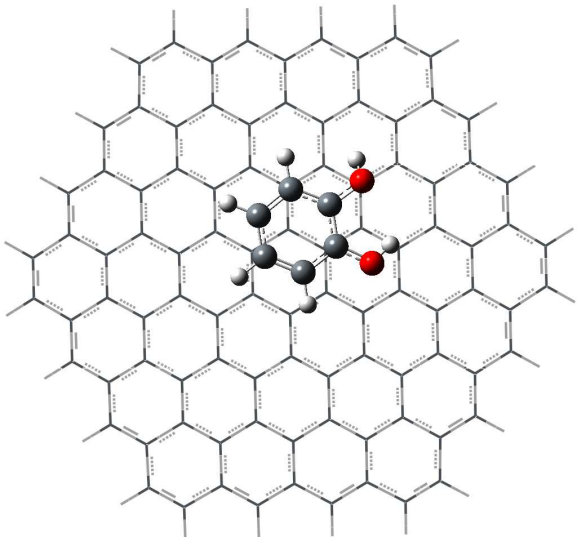
2, 4-DCP-C96



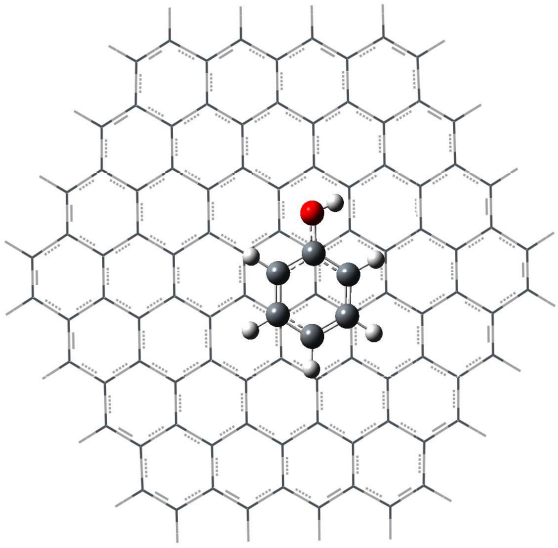
1, 3-DCB-C96



PY-C96

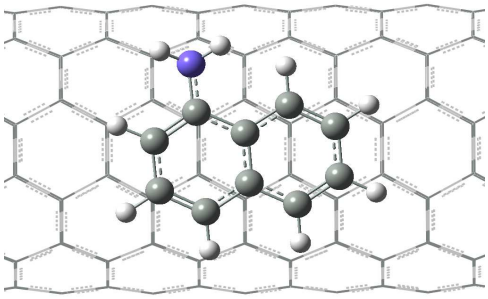


CT-C96

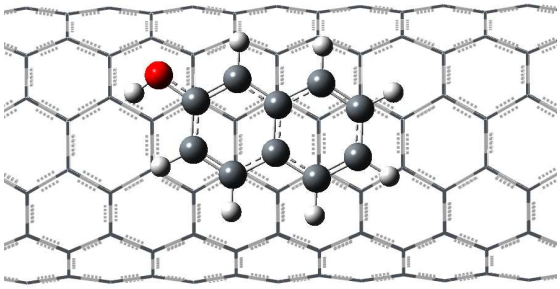


PH-C96

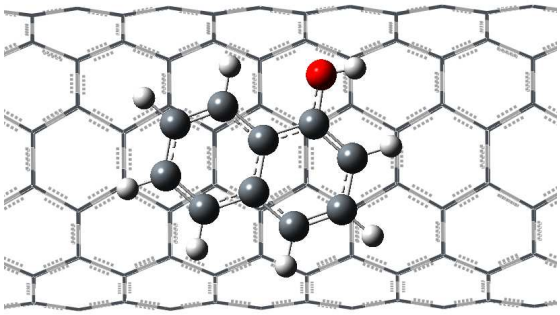
Figure 5.2 Optimized structures of aromatic pollutants on C96



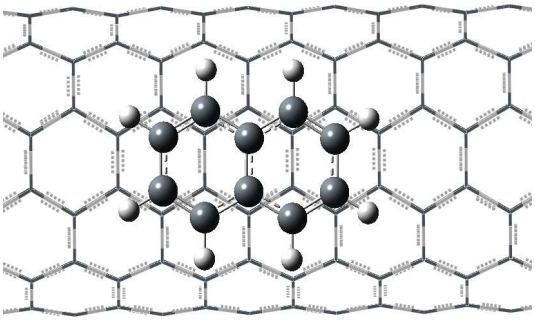
NAPH-[6, 6]



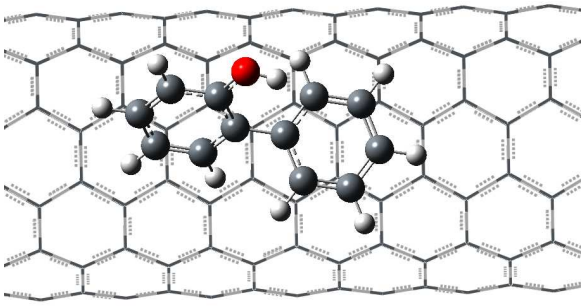
1-NATH-[6, 6]



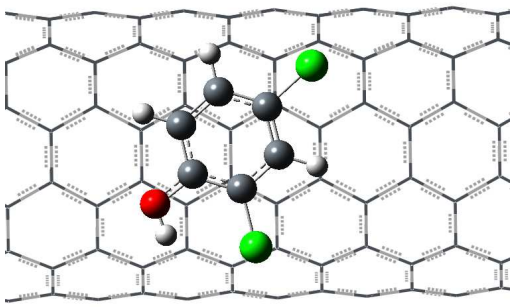
2-NATH-[6, 6]



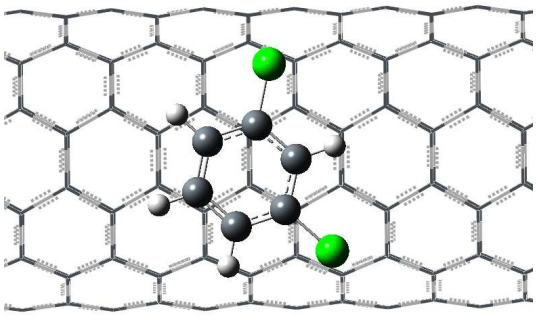
NAPH-[6, 6]



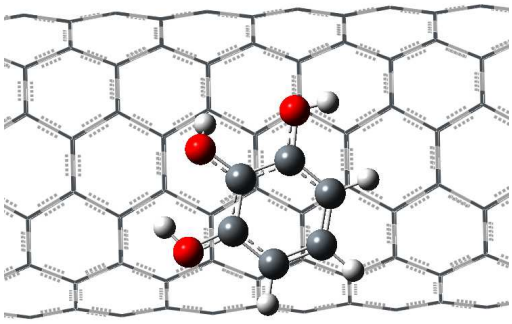
2-PHPH-[6, 6]



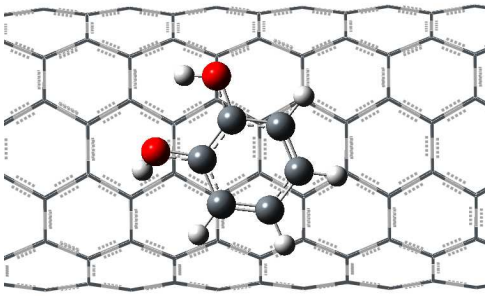
2, 4-DCP



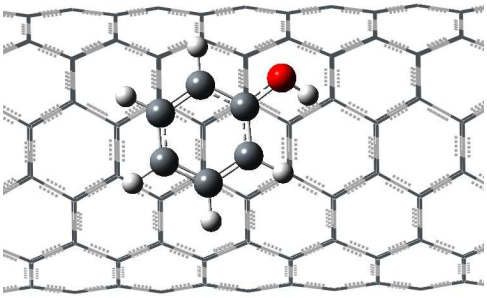
1, 3-DCB-[6, 6]



PY-[6, 6]

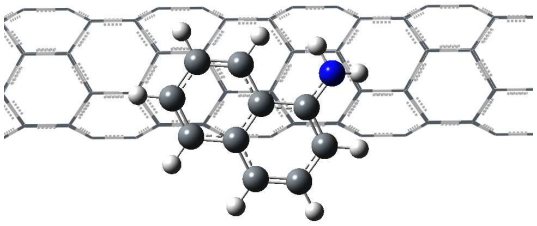


CT-[6, 6]

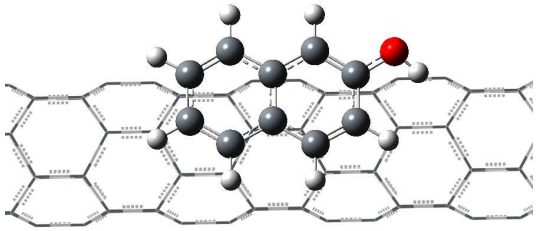


PH-[6, 6]

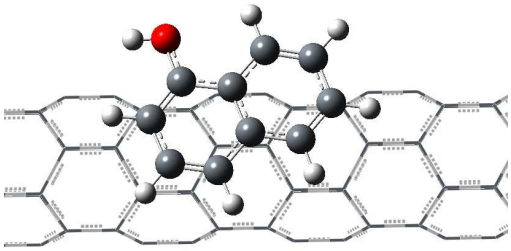
Figure 5.3 Optimized structures of aromatic pollutants on [6, 6]



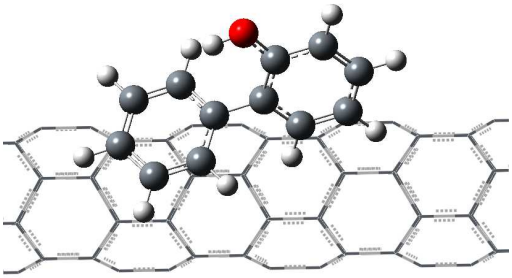
1-NALA-[5, 0]



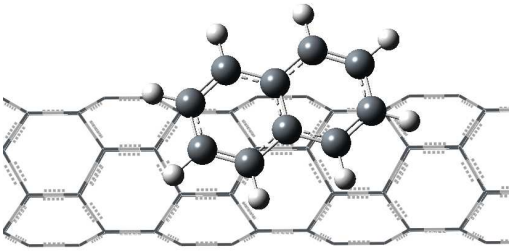
2-NATH-[5, 0]



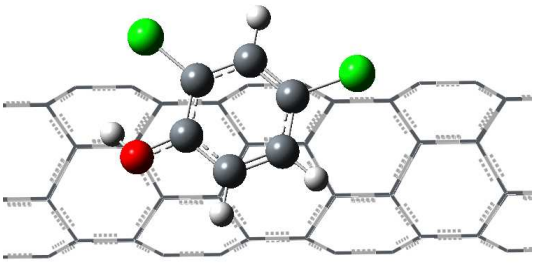
1-NATH-[5, 0]



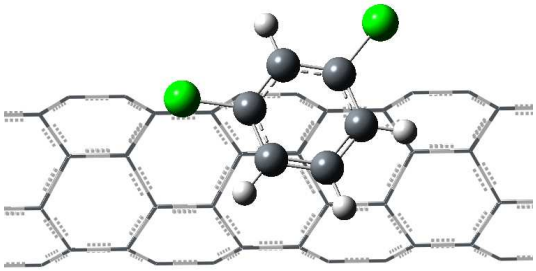
2-PHPH-[5, 0]



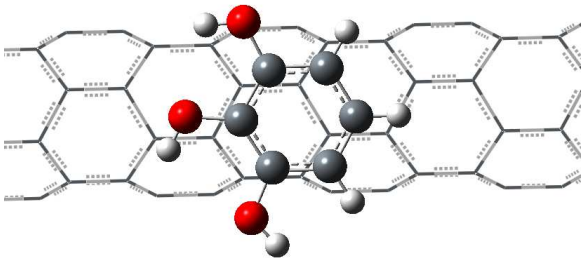
NAPH-[5, 0]



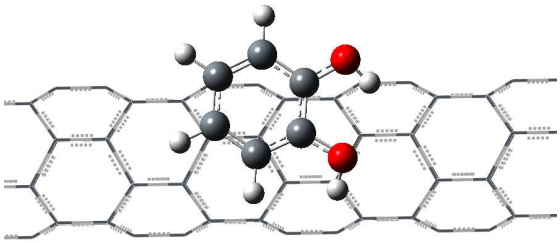
2, 4-DCP-[5, 0]



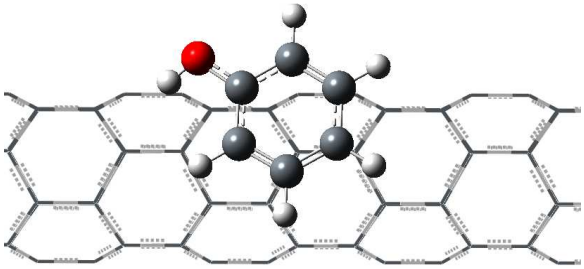
1, 3-DCB-[5, 0]



PY-[5, 0]



CT-[5, 0]



PH-[5, 0]

Figure 5.4 Optimized structures of aromatic pollutants on [5, 0]

Table 5.1 Structural data (Å)[a] for complexes of pollutants with –OH group on graphene, and [6, 6] and [5, 0] nanotubes

Size of graphene sheet	Model	1-NATH	2-NATH	2-PHPH	2,4-DCP	PH	CT	PY
C96	PM3-D	3.32	3.32	3.68	3.35	3.30	3.31	3.30
[6, 6]	PM3-D	3.43	3.40	3.70	3.25	3.24	3.23	3.15
[5, 0]	PM3-D	3.13	3.15	4.01	3.15	3.16	3.17	3.08

[a] Average distance of the pollutant heavy atoms from the RMS plane through the central benzene ring of the graphene sheet or nanotube.

Table 5.2 Structural data (Å)[a] for complexes of pollutants without –OH group on graphene, [6, 6] and [5, 0] nanotubes

Size of graphene sheet	Model	1-NALA	1,3-DCB	NAPH
C96	PM3-D	3.32	3.34	3.31
[6, 6]	PM3-D	3.41	3.42	3.45
[5, 0]	PM3-D	3.11	3.15	3.14

[a] Average distance of the pollutant heavy atoms from the RMS plane through the central benzene ring of the graphene sheet or nanotube.

Table 5.3 Interaction energies (kcal mol^{-1})[a] for complexes of pollutants with –OH group on graphene, and [6, 6] and [5, 0] nanotubes

Size of graphene sheet	Model	1-NATH	2-NATH	2-PHPH	2,4-DCP	PH	CT	PY
96	PM3-D	-24.5(-34.9)	-24.6(-34.9)	-24.1(-31.8)	-19.4(-27.5)	-15.8(-22.4)	-18.0(-24.6)	-19.8(-26.8)
	DFT-D//PM3-D	-22.1(-35.0)	-22.4(-35.0)	-18.3(-31.9)	-17.8(-27.6)	-14.3(-22.5)	-16.5(-24.7)	-18.8(-26.9)
[6, 6]	PM3-D	-17.3(-24.8)	-17.8(-25.4)	-17.8(-23.0)	-14.4(-20.5)	-11.7(-16.6)	-13.1(-17.9)	-13.8(-19.1)
[5, 0]	PM3-D	-15.0(-21.3)	-15.5(-21.8)	-15.5(-19.3)	-13.0(-18.2)	-10.2(-14.0)	-11.1(-15.3)	-11.8(-16.0)

[a] dispersion energy is given in parentheses

Table 5.4 Interaction energies (kcal mol^{-1}) for complexes of pollutants without –OH group on graphene, and [6, 6] and [5, 0] nanotubes

Size of graphene sheet	Model	1-NALA	1,3-DCB	NAPH
96	PM3-D	-26.4(-36.6)	-18.3(-25.7)	-22.8(-32.6)
	DFT-D//PM3-D	-22.7(-36.7)	-16.2(-25.7)	-19.6(-32.7)
[6, 6]	PM3-D	-17.9(-25.6)	-13.9(-19.5)	-16.6(-23.7)
[5, 0]	PM3-D	-16.1(-22.4)	-12.1(-16.8)	-14.3(-20.3)

[a] dispersion energy is given in parentheses

5.4 Conclusions

We have investigated the role of π-π interaction in the adsorption of aromatic pollutants on graphene sheet and on nanotubes using semi-empirical method implemented with dispersion correction term PM3-D. We have studied the energies and structures of 10 aromatic pollutants on C96, [6, 6] and [5, 0] nanotube in detail. All these calculations have been performed in gas phase using finite clusters rather than infinite periodic solids. We have also evaluated the interaction energies at DFT-D level using PM3-D optimized structures. The interaction energies are higher for pollutants with two aromatic rings than pollutants with one aromatic ring. Aromatic pollutants substituted with higher number of –OH groups have been found to be interacting stronger than pollutants with less number of –OH substituted groups. These findings are in good agreement with experiments. We found that the dispersive interactions are dominant. Without dispersion contribution, all other energies would be repulsive and so that complexes would be unbound. It has been found that dispersive contribution increases with the number of aromatic rings and degree of –OH substitution. It has also been noted that curvature of the π-surface plays an important role in π-π interaction. If the curvature of the nanotube is less, there is more possibility for a molecule to interact with π-surface. This is the reason for less interaction energies of adsorbates on [5, 0] nanotube than on less curved [6, 6] and on flat surface sheet graphene. However, the trend of interaction energies has been found to be same for both graphene sheet and nanotubes. Mulliken population analyses show a small degree of charge transfer from adsorbate to π-surface of the graphene sheet and nanotube, which indicate that there is no correlation between the charge transfer and the variation in the interaction energy. All these results suggest that the strength of the intermolecular interactions is indeed dominant in determining the strength of the adsorption.

5.5 References

1. Rao, G. P.; Lu, C.; Su, F. *Sep. Purif. Technol.*, **2007**, *58*, 224.
2. Hilding J.; Grulke, E. A.; Sinnott, S. B.; Qian, D. L.; Andrews, R.; Jagtoyen, M. *Langmuir*, **2001**, *17*, 7540.
3. Long, R. Q.; Yang, R. T. *J. Am. Chem. Soc.*, **2001**, *123*, 2058.
4. Peng, X. J.; Li, Y. H.; Luan, Z. K.; Di, Z. C.; Wang. H. Y.; Tian, B. H.; Jia, Z. P. *Chem. Phys. Lett.*, **2003**, *376*, 154.
5. Liao, Q.; Sun, J.; Gao, L. *Carbon*, **2008**, *46*, 553.
6. Yang, K.; Zhu, L. Z.; Xing, B. S. *Environ. Sci. Technol.*, **2006**, *40*, 1855.
7. Xia, G.; Ball. W.P. *Environ. Sci. Technol*. **1999**, *33*, 262.
8. Crespo, D.; Yang, R. T. *Ind. Eng. Chem. Res.*, **2006**, *45*, 5524.
9. Giannozzi, P. *Appl. Phys. Lett.*, **2004**, *84*, 3936.
10. Gotovac, S.; Honda, H.; Hattori, Y.; Takahashi, K.; Kanoh, H.; Kaneko, K., *Nano Lett*, **2007**, *7*, 583.
11. Zhang. J.; Lee, J. K.; Wu, Y.; Murray, R. W. *Nano Lett*, **2003**, *3*, 403.
12. Wang, Z. W.; Liu, C. L.; Liu, Z. G.; Xiang, H.; Li, Z.; Gong, Q. H. *Chem. Phys. Lett.*, **2005**, *407*, 35.
13. Cockroft, S. L.; Perkins, J.; Zonta, C.; Adams, H.; Spey, S. E.; Low, C. M. R.; Vinter, J. G.; Lawson, K. R.; Urch, C. J.; Hunter, C. A. *Org. Biomol. Chem.*, **2007**, *5*, 1062.
14. Chen, W.; Duan, L.; Wang, L.; Zhu, D. *Environ. Sci. Technol*. **2008**, *42*, 6862.
15. Lin, D.; Xing, B. *Environ. Sci. Technol*. **2008**, *42*, 7254.
16. Jiang, J.; Pang, S.-Y.; Ma, J. *Environ. Sci. Technol*. **2009,** *43,* 3398.
17. Chen, W.; Duan, L.; Wang, L.; Zhu, D. *Environ. Sci. Technol*. **2009,** *43,* 3400.

Chapter 6

Theoretical investigation of noncovalent interactions in functionalisation of nanotube and graphene sheets by biomolecules

6.0 Abstract

In this project, we have assessed quantum mechanical methods to study the importance of noncovalent interactions in functionalizaing nanotubes and graphene sheets by nucleic acid bases and by aromatic amino acids. We have mainly used semi-empirical methods implemented with dispersion correction (PM3-D and PM3-D*) to get equilibrium structures. We have studied structural and energetic aspects of these functionalized graphene and nanotube by nucleic acid bases, aromatic amino acids and tryptophan analogues in detail. Interaction energies, dispersion energies and intermolecular distances are calculated for all minimized complexes. We have evaluated the interaction energies on semi-empirical minimized structures of nucleic acid bases on graphene sheets by using density functional theory method implemented with dispersion correction and M0x methods. We have found that interaction energies obtained by semi-empirical methods are within ~2 kcal mol^{-1} of interaction energies obtained by DFT-D method. Among the family of M05 and M06, M06-2X method seems to give interaction energies close to the values of the DFT-D method for nucleic acid bases on the C24 sheet. Semi-empirical methods with dispersion corrected term have been found to be performing better and faster to study these molecules. Interaction energies of nucleic

acid bases on graphene sheets and on nanotubes follow the order of Guanine> Adenine> Thymine> Cytosine> Uracil. Nucleic acid bases are found to be interacting stronger on nanotubes with high diameter due to the larger circumference of π surface. We have also studied the interaction of amino acids on the graphene sheet C96. The interactions of aromatic amino acids on nanotubes [6, 6] and [5, 0] have also been studied. Aromatic amino acids have been found to be interacting stronger on π-surface than other amino acids. Interaction energies of aromatic amino acid follow the order of Trp> Tyr> Phe> His. We have also tested PM3-D* method to evaluate the interaction energies of tryptophan and their derivatives on [6, 6] nanotube. We have considered tryptophan and their derivatives such as 5-fluorotryptophan, 5-hydroxytryptophan, 7-azatryptophan and protonated 7-azatryptophan for this study. Interaction energies of these analogues are in the order of AW^+ > 5-HW> 5-FW> AW ≈ W. Semi-empirical methods implemented with dispersion term (PM3-D and PM3-D*) are found to give reasonable energies and structures with respect to DFT-D method.

6.1 Introduction

Carbon nanotubes were discovered by Sumio lijima[1] in 1991 although researchers were studying the synthesis and properties of carbon filaments for many years before the discovery of carbon nanotubes. Prior to the discovery of carbon nanotubes, C60 fullerene molecule was discovered by Kroto et al[2] in 1985. After these two discoveries, new fields of material research have been launched. Structure and electronic properties of carbon nanotubes have been studied by experimental techniques such as scanning tunneling microscopy (STM).[3] Carbon nanotubes (CNTs) are found to be long, thin, cylindrical molecules made up of carbon atoms. CNTs can be described as graphene sheets rolled-up into cylinders with nanometer size diameter. Graphene sheet, one-atom-thick layer of sp^2 bonded carbon atoms, was first extracted by Geim et al[4] (University of Manchester) from bulk graphite in 2004. Nanotubes can be formed by rolling up the sheet so that the end points O and A of the vector $C_h = na_1 + ma_2 = (n, m)$ coincide. C_h and T are chiral and translational vector respectively whereas a_1 and a_2 are unit vectors Fig. 6.1.[6] These nanotubes can be denoted as [*n*, *m*], where n and m are chiral numbers. If *n* and *m* are equal, nanotubes are called as armchair nanotubes, since the C-C bonding

pattern will be in armchair pattern. If *n* and *m* are not equal and *m* is zero, then nanotubes are called as zig-zag nanotubes, since the C-C bonding pattern will be in zig-zag pattern.

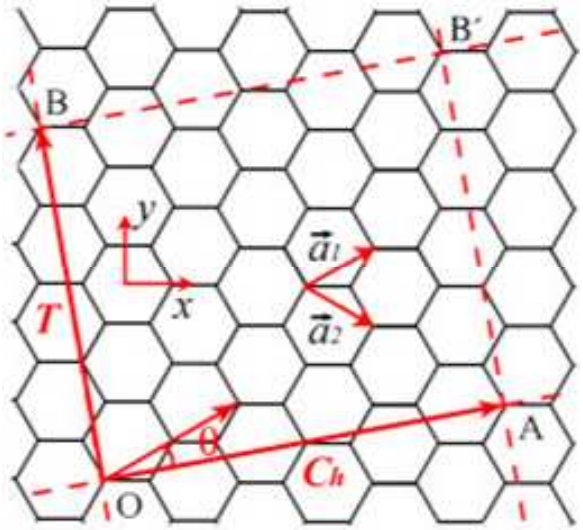


Figure 6.1 Hexagonal graphite layer showing chiral vector, chiral angle and unit vectors (Image is taken from ref 5)

CNTs are unique in electronic, mechanical and structural properties. This uniqueness of CNTs leads to the application of nanotubes in many research fields such as bionanotechnology, DNA nanotechnology, environmental nanotechnology, etc.[5] Particularly in bionanotechnology, nanotubes can be used as transporters to deliver the drug molecules to targets. However, the aggregation of nanotubes in solution or insolubility of nanotubes is a barrier to these applications. To overcome this barrier, i.e., to get a single nanotube to use its properties efficiently, nanotubes need to be

functionalized by functional groups, which can be soluble in solvents. There are mainly two ways to accomplish functionalisation, i.e., a. covalent functionalisation, adding the functional group to nanotube covalently, b. noncovalent functionalisation, adding the functional group to nanotube noncovalently, through which nanotube properties can be unaltered. Therefore, noncovalent functionalisation is advantageous to maintain the nanotube properties and use them efficiently. Noncovalent functionalisation mainly involves π-π stacking between nanotube and aromatic molecules. It becomes necessary to understand these π-π stacking interactions at atomic level quantitative and qualitatively. Investigation of these noncovalent interactions theoretically requires realistic models of nanotubes, which will be 30Å in length, (nearly 300 to 320 atoms). Investigating these noncovalent interactions among these big molecules at high level computation is a challenging task. High-level ab initio methods such as CCSD(T), coupled cluster single and double with peturbative triple excitation methods which can describe noncovalent interactions with high accuracy are limited to only small systems containing 10 to 15 atoms. Generally accepted MP2 (Moller-plesset perturbation theory) methods are also expensive and include electron correlation term, but they are found to overestimate the binding energies of noncovalent systems. The standard density functional theory methods such as DFT, which have been proven to be successful are also failed to describe the noncovalent interactions. Grimme proposed an approach that adding the C6/ R^6 term, which can describe dispersive interactions explicitly to standard DFT methods.

DNA nanotechnology is an emerging research discipline that researchers use the molecular recognition properties of DNA to design and synthesize self-assembled molecules of DNA with different properties at nanoscale. Nadrian C. Seeman in early 1980s has found that DNA can be used as structural material to crystallize biomolecules although DNA has been used as genetic materials for many decades before 1991. DNA (deoxyribo nucleic acid) is a polymer of repeating units of nucleotides which are made up of two segments i) the backbone of nucleoside i.e phosphate and sugar and ii) bases. A basic structure of DNA is made up of two nucleotide polymer chains which run in opposite direction. These two chains also known as strands form a double helical structure, which is stabilized by hydrogen bonds between the bases of nucleotides. Base

in one strand form a hydrogen bond with one type of base in the other strand which is called as complementary base-pairing. There are four nucleic acid bases in DNA called as Adenine, Thymine, Cytosine and Guanine. Adenine and Cytosine are known as pyrimidine bases whereas Thymine and Guanine are known as purine bases with one ring. Watson and Crick[23] are two scientists who discovered the rule of base-pairing between bases in DNA. The rule of base-pairing is that Adenine always pairs with thymine and guanine always pairs with cytosine to establish the hydrogen bonds. Nadrian C. Seeman et al (1991)[24] have published the synthesis of DNA cube, the first three dimensional nanoscale objects following by a DNA truncated octahedron.

6.2 Results and discussion

Biomolecules such as nucleic acid bases, proteins and carbohydrates have been shown as functionalizing CNTs through noncovalent interactions. Therefore it has been necessary to understand the interaction of these biomolecules on CNTs. Herein we discuss the previous studies having been carried out for the interaction between a) nucleic acid bases on graphene and CNTs, and b) aminoacids on graphene and CNTs.

6.2.1 Nucleic acid bases on CNTs

Stacking between nucleic acid bases (NAB) led the research towards functionalising of nanotubes by NAB through π-π stacking. There are several studies done on the immobilization of proteins and nucleic acids on nanotubes[1-3] and the attachement of DNA and RNA onto CNTs for improving the solubility and bioavailability of nanomaterials in aqueous solution.[8-9] Hwang et al showed that CNTs can be employed to utilize as generic nanobiomarkers for the precise detection of a particular gene with very high sensitivity and specificity.[13] Zheng et al[5] (2003) in their work have showed that CNTs can be dispersed in water medium in the presence of ssDNA strands and can be separated. This fact was supported by optical absorption and fluorescene spectroscopy and by atomic force microscopy measurements. It has also been shown that CNTs are wrapped by DNA bases through π-π stacking by molecular modeling

studies. Gowtham et al[6], have studied the interaction of five nucleic acid bases (Adenine (A), Thymine (T), Cytosine (C), Guanine (G) and Uracil (U)) on graphene and nanotubes using density functional methods. Grimme et al[7] have studied the interaction of these five nucleic acid bases on different sizes of flat sheets using DFT method implemented with dispersion term and found that the order of interaction is as follows G> A> T> C> U. Das et al[8], have studied the interaction of four nucleic acid bases (A, T, C, G) on [5, 5] nanotube using HF method and force field and they also found that the order of binding in solvent changes as follows G> T> A> C. Shtogun et al[9] studied the interaction of Adenine, thymine and their radicals on nanotubes using density functional theory and found that there are many orientations possible for interaction of these bases on nanotubes. Wang et al[10] showed in his studies that thymine disperse nanotube stronger than cytosine and adenine.

These findings have motivated us to investigate the role of noncovalent interactions in the absorption of nucleic acid bases on nanotubes. All of these studies have been done on the expense of dispersion or on the models used for studies. Since these calculations require more realistic models, which will obviously require high computational efficiency and accurate description of dispersion, these calculations become impossible to use high-level ab initio methods such as CCSD(T). Eventhough density functional theory implemented with dispersion term serves as an alternative to high level *ab initio* methods; still it is difficult to use these methods to study interaction on nanotube with realistic models (300 to 320 atoms). There arises the necessity of faster and cheaper semi-empirical methods such as PM3, AM1 implemented with dispersion term. PM3, AM1 methods have been implemented with dispersion term and named as PM3-D, AM1-D and PM3-D* (which is modified with core-core repulsion term by our group).[25] These methods have been tested against S22 and JSCH2005 database and found to be successful in reproducing the interaction energies within average of 0.5-2.5 kcal mol^{-1} of DFT-D and MP2 interaction energies. Therefore, these methods have been used to study the interaction of nucleic acid bases A, T, C, G and U on different sizes of flat sheets and different kinds of nanotubes.

The primary target of this project has been the application of semi-empirical methods implemented with dispersion term such as PM3-D and PM3-D* to study the interaction of nucleic acid bases on flat aromatic hydrocarbon sheets and on nanotubes. We have also used standard density functional method DFT-D and recent methods M0x to evaluate the energies of semi-empirical optimized structures. All geometry calculations have been performed using PM3-D and PM3-D* methods. The semi-empirical optimized structures were used to evaluate the energies at DFT-D level and by using M05 and M06 families of density functionals. Interaction energies with dispersion contribution are calculated along with intermolecular distances for optimized complexes.

We have chosen semi-empirical methods because to study π-π interactions on large graphene sheets and on nanotubes at high level theories is very expensive and time consuming. Semi-empirical methods are found to be two times faster than standard density functional methods and less expensive.[ref] Dispersion correction term implemented semi-empirical methods such as PM3-D and PM3-D* have been already proven to be successful methods to study π-π interactions in various systems.

Models

Lowest energy tautomers of nucleic acid bases Adenine, Thymine, Cytosine, Guanine and Uracil (A, T, C, G and U) have been considered for all the calculations. We have used two nanotubes both armchair nanotube [6, 6] and zig-zag nanotube [5, 0] with 30Å length which have 300 and 140 carbon atoms terminated by 24 and 10 hydrogens respectively. We have also used flat graphene sheets with different sizes such as C24, C54, C96, C150 and C216. The starting structures have been modelled by placing bio molecules like nucleic acid base or amino acid molecules on the surface of nanotube or flat sheet, having the center of the ring or the center of the bridging bond of the adsorbate molecule above one of the six equivalent central carbon atoms of the graphene sheet or nanotube. These starting structures then were optimized using PM3-D and PM3-D* methods. We have calculated interaction and binding energies for these intermolecular complexes.

6.2.2 Interaction of nucleic acid bases on polycyclic aromatic hydrocarbons

We have studied the interactions of nucleic acid bases Adenine, Thymine, Cytosine, Guanine and Uracil (A, T, C, G and U) on polycyclic aromatic hydrocarbons with different sizes from C24 to C216. The initial structures were constructed by placing pyrimidine bases and purine bases, with the centre of the ring or the centre of the bridging bond, respectively, above one of the six equivalent central carbon atoms of the ring at the approximately 3.5-4.0Å distance. All the structures are optimized using PM3-D and PM3-D* methods. Interaction and binding energies with dispersion contribution were calculated. Since there is no significant difference between interaction and binding energies, interaction energies are considered for further discussion.

There are number of stacking structures possible for nucleic acid bases on graphene sheets depending on the starting structures Fig. 6.2.[JMac] Structures can be classified as T, B and C depending on the position of bases on sheets. Structure T will have the centre of base placed on the top of the carbon of the sheet, whereas B and C will have the centre of base placed on the top of the bridge bond and on the centre of the ring respectively. The optimized structures of these complexes fall in the category of T and B being the stable structures. All base molecules are found to be parallel with the surface of the graphene sheet.

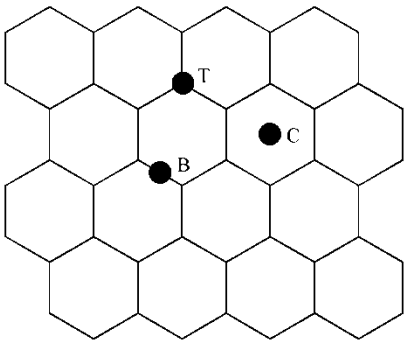


Figure 6.2 Possible stacking positions for bases on graphene sheet, T (top of a carbon atom), B (top of the centre of C-C bond), C (top of the centre of a carbon hexagon)

We have calculated intermolecular distances for all optimized complexes of bases on C24, C56 and C94 and are all in the range of 3.00-3.35 Å (Table 6.1). The average distance decreases with increasing size of the sheet. For the smallest sheet C24, all the distances are very close to the optimal value obtained by BLYP-D/TZV (d, p).[ref] The decrease in distance for larger sheets are minimal such as < 0.1 Å for PM3-D and PM3-D* structures whereas it has been slightly higher such as 0.3 Å for larger sheets at DFT-D level.

Table 6.1 Structural data (Å) for optimized graphene-base complexes

Size of graphene sheet	Model	A	T	C	G	U
24	PM3-D	3.33	3.35	3.31	3.33	3.33
	PM3-D*	3.24	3.23	3.21	3.24	3.23
	B97-D/TZV (d, p)[ref]	3.26	3.23	3.19	3.20	3.22
54	PM3-D	3.30	3.33	3.30	3.28	3.33
	PM3-D*	3.21	3.21	3.20	3.20	3.20
	B97-D/TZV (d, p)	3.11	3.10	3.04	3.05	3.09
96	PM3-D	3.31	3.33	3.29	3.23	3.30
	PM3-D*	3.21	3.20	3.19	3.20	3.19
	B97-D/TZV (d, p)	3.03	3.01	2.97	3.01	3.00

Note: Distance calculated by taking the average of the heavy nucleobase atoms from the RMS plane defined through centre benzene ring of the sheet or nanotube

We have started our calculations with the small flat sheet coronene, which contains 24 carbon atoms and 12 hydrogen atoms. We have found that two purine bases interact stronger on the surface than pyrimidine bases. It might be due to the larger surface area of purine bases, which allow them to interact stronger on the surface. The PM3-D interaction energies of Adenine (A) and Guanine (G) on C24 are –16.1 and –17.9 kcal mol^{-1}. The PM3-D interaction energies of Thymine (T), Cytosine (C) and Uracil (U) are –15.9, –14.7 and –12.9 kcal mol^{-1} respectively. Therefore, the trend of PM3-D interaction energies of bases on C24 is G> A> T> C> U. The interaction energies

obtained by PM3-D* also follow the similar trend to PM3-D interaction energies except Adenine. The difference may be attributed to the interaction of bases with edge of the sheet. All the interaction energies, except the interaction of adenine on C24, differ by less than 2 kcal mol^{-1} for both semi-empirical schemes. To explore the interaction energy trend in various sizes of the sheets, we have extended our calculations on C54, C96, C150, and C216 sheets too. Interaction energies of bases on sheets increase with increasing size of the sheet. However, the difference in the increment is not much over sheet size of C96. This might be because over 12Å radius there will not be any van der Waals interactions. We have then calculated the interaction energies for structures obtained from both PM3-D and PM3-D* methods by using DFT method implemented with dispersion term (DFT-D). Semi-empirical interaction energies differ by ~2 kcal mol^{-1} of BLYP-D interaction energies. BLYP-D interaction energies are within 2 kcal mol^{-1} of the B97-D/TZV (2d, 2p) values. Dispersion energies are found to be higher than overall interaction energies. This shows that the complexes would be unbound without dispersion contribution. Interaction energies on all sheets are higher for base guanine and are less for base uracil. We have found that BLYP-D interaction energies also follow the similar trend on all sheets, which is G> A> T> C> U. It is apparent that bases with larger π surface area interact stronger on aromatic hydrocarbons than bases with one aromatic ring. For convenience, we have also plotted the interaction energy of bases on different sheets obtained by PM3-D* method to show the trends of increasing energy with increasing sheet size and G> A> T> C> U (Fig.6.3).

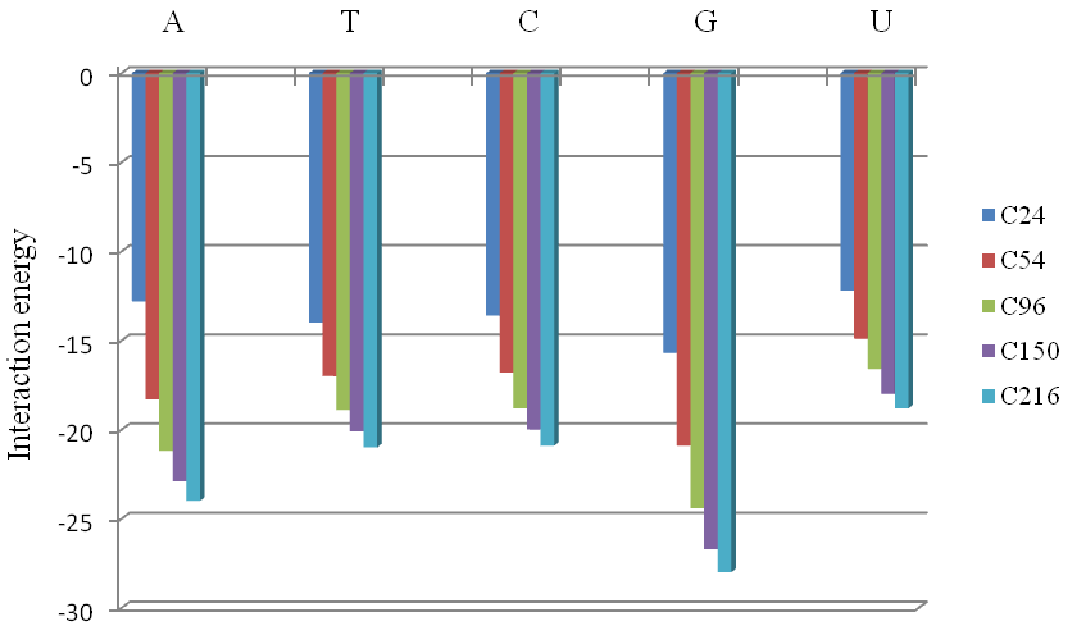


Figure 6.3 Interaction energies of bases on different sizes of flat sheets

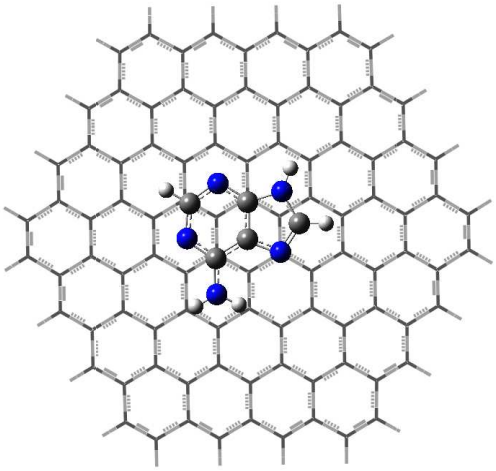
Table 6.2 Interaction energies (kcal mol^{-1}) for graphene-base complexes

Size of graphene sheet	Model	A	T	C	G	U
24	PM3-D	-16.1 (-23.2)	-15.9 (-20.2)	-14.7 (-19.0)	-17.9 (-23.5)	-12.9 (-16.8)
	PM3-D*	–12.8 (-24.4)	–14.0 (-21.2)	–13.6 (-19.9)	–15.7 (-24.8)	–12.2 (-18.3)
	DFT-D//PM3-D	-14.2 (-23.2)	-13.4 (-20.3)	-13.0 (-19.1)	-16.9 (-23.6)	-11.3 (-16.9)
	DFT-D//PM3-D*	-13.8 (-24.5)	-13.5 (-21.2)	-13.3 (-19.9)	-16.8 (-24.8)	-11.5 (-18.3)
	B97-D/TZV (2d, 2p)	-14.2 (-24.9)	-15.4 (-24.5)	-13.7 (-22.5)	-18.3 (-27.9)	-13.2 (-20.3)
54	PM3-D	-20.1 (-28.2)	-18.2 (-24.3)	-17.4 (-22.6)	-22.0 (-29.4)	-14.6 (-19.8)
	PM3-D*	-18.3 (-29.7)	-17.0 (-26.0)	-16.8 (-24.1)	-20.9 (-31.8)	-14.9 (-22.4)
	DFT-D//PM3-D	-20.4 (-28.2)	-16.3 (-24.3)	-15.8 (-22.6)	-21.3 (-29.4)	-13.6 (-19.8)
	DFT-D//PM3-D*	-18.4 (-29.8)	-16.9 (-26.1)	-16.0 (-24.1)	-21.8 (-31.8)	-14.6 (-22.5)
	B97-D/TZV (2d, 2p)	-19.3 (-31.7)	-18.0 (-29.8)	-17.8 (-28.0)	-23.3 (-36.2)	-15.4 (-24.8)
96	PM3-D	-21.0 (-29.3)	-18.5 (-25.2)	-17.4 (-23.6)	-22.9 (-30.7)	-15.0 (-21.2)
	PM3-D*	-21.2 (-31.1)	-18.9 (-27.4)	-18.8 (-25.5)	-24.4 (-33.7)	-16.6 (-23.7)
	DFT-D//PM3-D	-19.8 (-29.3)	-16.6 (-25.2)	-16.9 (-23.6)	-22.3 (-30.7)	-15.4 (-21.2)
	DFT-D//PM3-D*	-19.2 (-31.1)	-17.9 (-27.4)	-17.0 (-25.5)	-23.0 (-33.7)	-14.8 (-23.7)
	B97-D/TZV (2d, 2p)	-20.2 (-32.8)	-19.0 (-31.2)	-18.5 (-28.8)	-24.2 (-37.8)	-16.3 (-26.1)
C150	PM3-D*	-22.9 (-31.6)	-20.1 (-27.3)	-20.0 (-25.8)	-26.7 (-34.3)	-18.0 (-24.2)
C216	PM3-D*	-24.0 (-31.7)	-21.0 (-28.0)	-20.9 (-25.9)	-28.0 (-34.4)	-18.8 (-24.2)

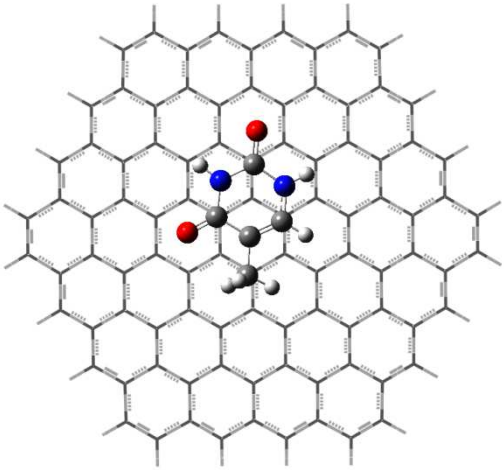
Binding energies are also evaluated by using M05 and M06 families of density functional methods on PM3-D* optimized structures of bases on C24, C54 and C96. We have not evaluated the interaction energies for bases on sheets larger than C96, since there is no effect of size of the sheet for larger sheets. To evaluate the binding energies on semi-empirical structures, BLYP-D functional has been used with TZV (2d, 2p) basis sets for DFT-D and M0x levels. We have used five different M0x methods such as M05, M05-2X, M06, M06-2X and M06-L. Binding energies and mean unsigned error (MUE) with reference to DFT-D values were calculated. For bases on C24 graphene sheet, M05 gives higher MUE value as 11.3 kcal mol^{-1}. M06-2X gives interaction energies within 1.0 kcal mol^{-1} difference of DFT-D interaction energies. We have found that M06 family is better than M05 family to calculate interaction energies. Among M06 family, M06-2X gives interaction energies with less deviation from DFT-D interaction energies than M06 and M06-L methods.

Table 6.3 Binding energies (kcal mol^{-1}) for graphene-base complexes using M05 and M06 functionals

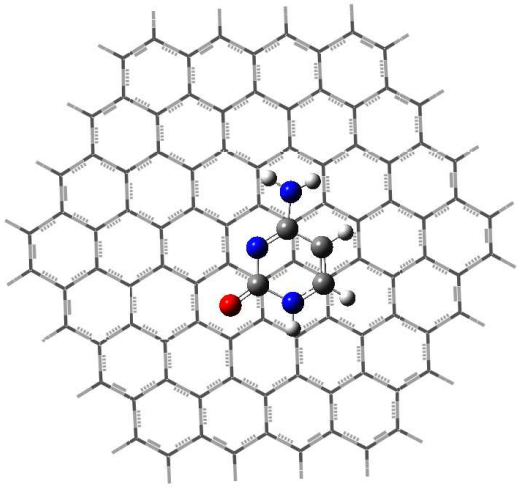
Size of graphene sheet	Functionals	A	T	C	G	U	MUE
24	M05	0.1	-2.4	-2.7	-3.3	-2.0	11.3
	M05-2X	-6.8	-8.6	-9.4	-11.1	-7.6	4.7
	M06	-8.1	-9.1	-9.1	-11.0	-7.6	4.4
	M06-2X	-11.2	-12.0	-12.8	-15.3	-10.3	1.0
	M06-L	-9.4	-10.0	-10.5	-12.8	-8.4	3.1
	DFT-D	-13.2(-24.5)	-13.0(-21.2)	-12.9(-19.9)	-16.6(-24.8)	-11.1(-18.3)	
54	M06-2X	-13.2	-13.0	-13.1	-16.6	-11.6	3.4
	DFT-D	-17.8(-29.8)	-16.1(-26.1)	-15.5(-24.2)	-21.2(-31.9)	-14.0(-22.5)	
96	M05	0.1	-1.3	-1.6	-2.1	-1.2	16.5
	M05-2X	-8.3	-8.8	-9.1	-11.7	-8.0	8.6
	M06	-8.7	-9.2	-8.8	-11.5	-7.6	8.6
	M06-2X	-13.0	-13.0	-13.1	-16.9	-11.5	4.2
	M06-L	-10.9	-11.1	-11.1	-14.3	-9.6	6.3
	DFT-D	-18.4(-31.1)	-16.9(-27.4)	-16.4(-25.5)	-22.2(-33.7)	-14.8(-23.7)	



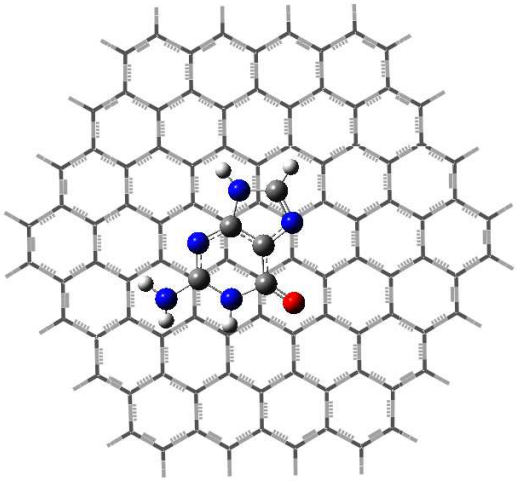
Adenine-C96



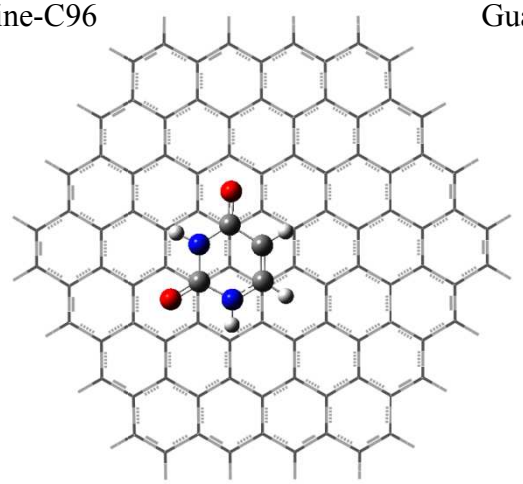
Thymine-C96



Cytosine-C96



Guanine-C96



Uracil-C96

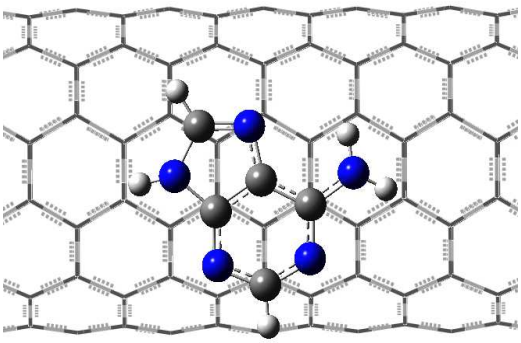
Figure 6.4 Structures of nucleic acid bases on C96 optimized by PM3-D*

6.2.3 Interaction of nucleic acid bases on carbon nanotubes

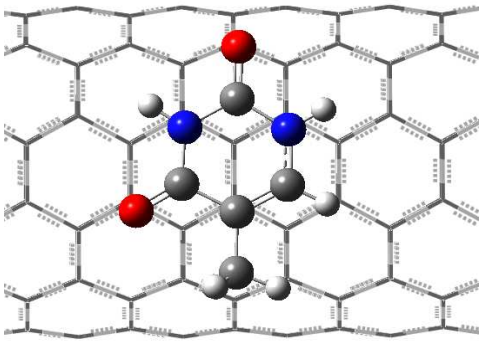
We have studied the interaction of nucleic acid bases (A, T, C, G and U) on carbon nanotubes. We have chosen two kinds of nanotubes, one armchair [6, 6] nanotube and one zig-zag [5, 0] and [10, 0] nanotubes. The diameters of the nanotubes [6, 6], and [5, 0] are 0.63 and 0.35Å respectively. The pyrimidine and purine bases are placed on the nanotubes, with the centre of the ring or the centre of the bridging bond, respectively, above one of the six equivalent central carbon atoms of the ring at the middle of the nanotube. The bases are placed on nanotube approximately at 3.5-4.0 Å respectively. The nanotubes have been chosen with 30 Å lengths and terminated with hydrogens. The [6, 6] nanotube has 300 carbon atoms and 24 hydrogens whereas [5, 0] nanotubes has 150 carbon atoms and 10 hydrogen atoms. All structures were optimized using both semi-empirical parameterizations PM3-D and PM3-D*. Interaction energies and dispersion contributions along with intermolecular distances are calculated and reported in Table 6.4. Interaction energies of bases on nanotubes follow the trend of G> A> T> C> U and are ranging from 12.0 to 21.0 kcal mol^{-1}. It has been found that the interaction energies of bases on [6, 6] nanotubes are higher than on [5, 0] nanotubes. This is because [6, 6] nanotube has larger circumference which allows the bases to interact stronger. The interaction energies of bases on [5, 0] nanotubes are lower than other nanotubes. This is because [5, 0] nanotube has small diameter which is highly curved.

Table 6.4 Interaction energies (kcal mol^{-1}) and structural data (Å) for nanotube-base complexes. The total interaction energies is given, with the dispersion contribution in paranthesis, followed by the intermolecular distance in square brackets, defined as in Table 6.1.

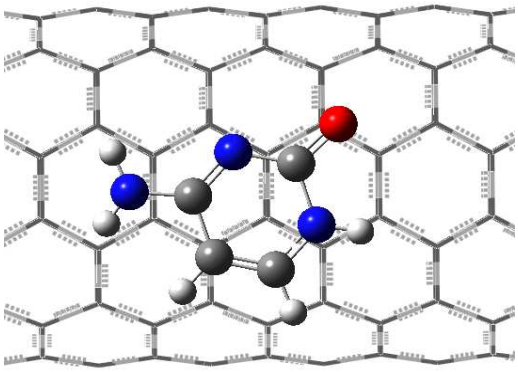
Nanotube	Model	A	T	C	G	U
[6, 6]	PM3-D	-15.3(-21.6) [3.18]	-13.6(-18.6) [3.47]	-12.6(-17.5) [3.21]	-16.8(-22.9) [3.23]	-10.7(-15.7) [3.25]
	PM3-D*	-18.8(-23.7) [3.04]	-16.3(-20.1) [3.10]	-15.9(-19.0) [3.12]	-21.3(-24.7) [3.13]	-14.7(-17.7) [3.11]
[5, 0]	PM3-D	-13.1(-18.1) [3.08]	-11.8(-16.0) [3.43]	-11.0(-14.6) [3.53]	-13.4(-19.1) [3.52]	-9.4 (-13.4) [3.14]
	PM3-D*	-15.1(-20.2) [3.32]	-13.2(-17.4) [3.34]	-13.4(-16.8) [3.29]	-16.0(-20.9) [3.00]	-12.1(-15.2) [3.35]



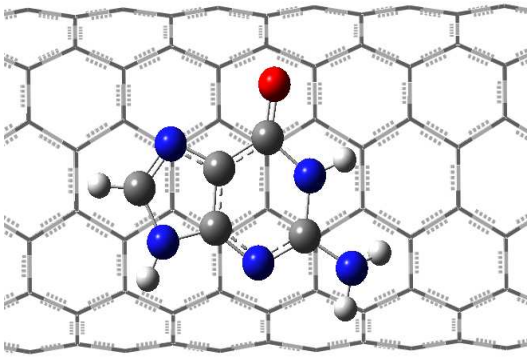
Adenine-[6, 6]



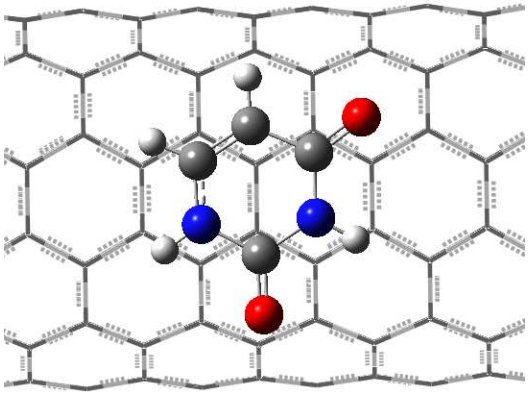
Thymine-[6, 6]



Cytosine-[6, 6]

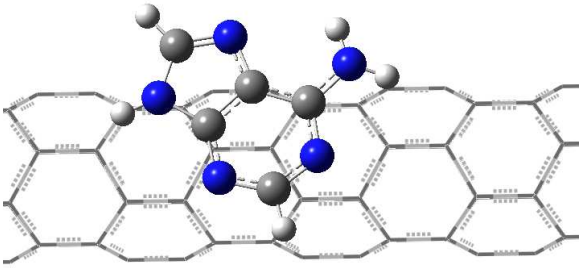


Guanine-[6, 6]

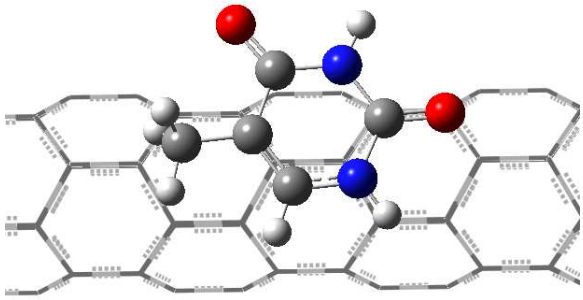


Uracil-[6, 6]

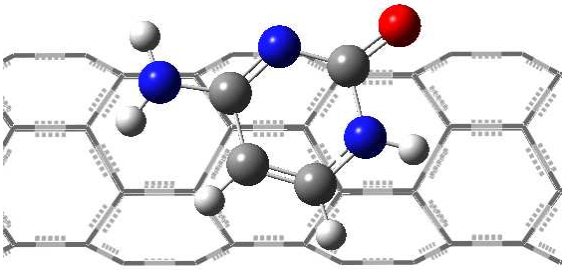
Figure 6.5 Structures of nucleic acid bases on [6, 6] nanotube optimized by PM3-D*



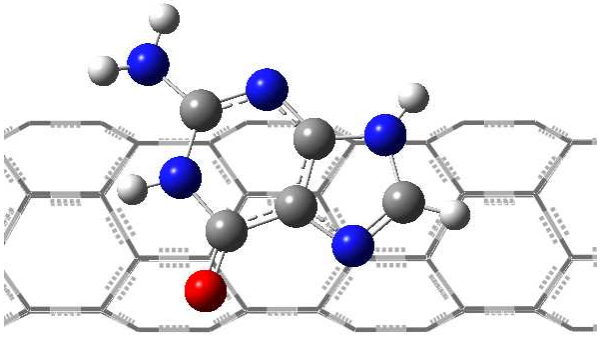
Adenine-[5, 0]



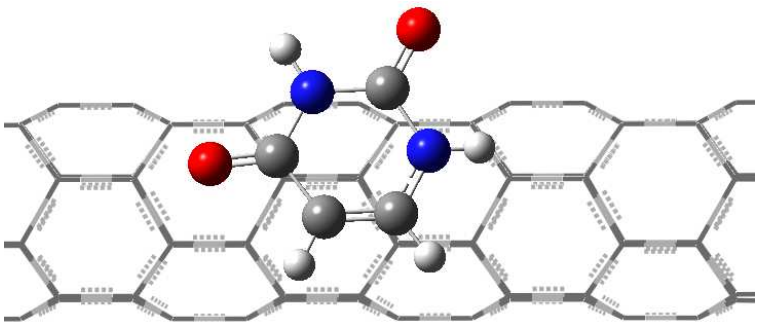
Thymine-[5, 0]



Cytosine-[5, 0]



Guanine-[5, 0]



Uracil-[5, 0]

Figure 6.6 Structures of nucleic acid bases on [5, 0] nanotubes optimized by PM3-D*

6.2.4 Interaction of aromatic amino acids on nanotubes

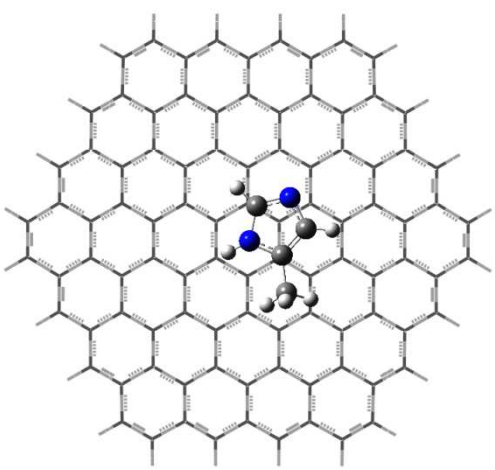
Su et al, have studied a range of peptides on nanotubes and found that histidine has high affinity binding for range of nanotubes. This study indicates that there is possibility of enhancing binding affinity by modifying the peptides or aromatic amino acids of peptides. Jon Mc et al studied the interaction of different aromatic amino acids on small graphene sheet such as pyrene. Continuing to that, we have studied the interaction of aromatic amino acids on graphene sheet with size C96 which is sufficient size to model π-π interactions. Therefore we have studied the aromatic amino acids in different orientations on graphene sheets. We have optimized the structures of graphene sheet with series of aromatic amino acids glycine (GLY), alanine (ALA), histidine (HIS), tyrosine (TYR), phenylalanine (PHE), tryptophan (TRP), leucine (LEU) and lysine (LYS). We have modeled the different orientation of aromatic amino acids to interact with graphene sheets. These different orientations are labeled as T-shaped or S-shaped where the backbone conformation is T-shaped or stacked. For His and Phe, there are two more conformations labeled as TR or SR, where the ring is T-shaped or stacked. In T-shaped structures, acid group which is hydrophilic has been modeled in the way pointing outwards from the surface to be solvated by the real systems. We have calculated the interaction energies and dispersion contributions and reported in Table 6.5. In Figure 6.7, we show the optimized complexes of amino acids and graphene sheets. The interaction energy plot is quite flat since all the energies are almost close. We have found that stacked structures are being most stable structures for ala and gly where T-shaped structures are in higher energy than stacked structures. For LEU, T-shaped structure is preferred with the lowest energy. For HIS and PHE, we have used four possible structures. For PHE, stacked structure is favorable structure where the ring is in stacked position on the graphene sheet with the large dispersive interaction. HIS structure follows the similar trend as PHE with the stacked structure as favorable structure. As expected the interaction energies of aromatic amino acids on C96 are higher than on pyrene. Aromatic amino acids have been found to interacting stronger on C96. All these energies on graphene sheet follow the similar trend as on pyrene sheet. Interaction energies are higher for graphene sheet compared to pyrene sheet. We have validated PM3-D method with DFT-D method for pyrene sheet. Therefore, PM3-D

values here have been considered to be correct and found to be performing better for interaction of aromatic amino acid on graphene sheets.

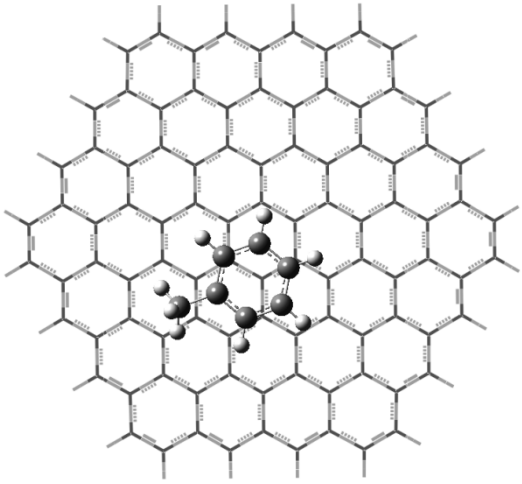
We have studied the interaction of aromatic amino acids HIS, ALA, TRP and TYR on carbon nanotubes [6, 6] and [5, 0]. The starting structures were constructed by placing aromatic amino acids, with the centre of the ring on the top of the atom of the six equivalent carbons at the middle of the nanotube. All structures were optimized using both semi-empirical parameterizations PM3-D and PM3-D* methods. Interaction and dispersion energies are calculated and reported in Table 6.5. We have shown the optimized structures of aromatic aminoacids on [6, 6] nanotube in Figure 6.8. Interaction energies of aromatic amino acids on [6, 6] carbon nanotube are higher than on [5, 0] nanotube. Interaction energies are ranging from 14.0 to 23.5 kcal mol^{-1} on C96 and differ by ~2 kcal mol^{-1} of MP2 values on C48. Interaction energies on [6, 6] nanotube are ranging from 11.5 to 17.8 kcal mol^{-1} of MP2 values on [5, 5] nanotube. The trend of interaction energies on both C96 and [6, 6] follows as His> Phe> Tyr> Trp. Since tryptophan is aromatic amino acid with larger π surface area, it interacts stronger on C96 and nanotube than other amino acids.

Table 6.5 Interaction energies and dispersion in paranthesis between C96, [6, 6] and aromatic amino acids

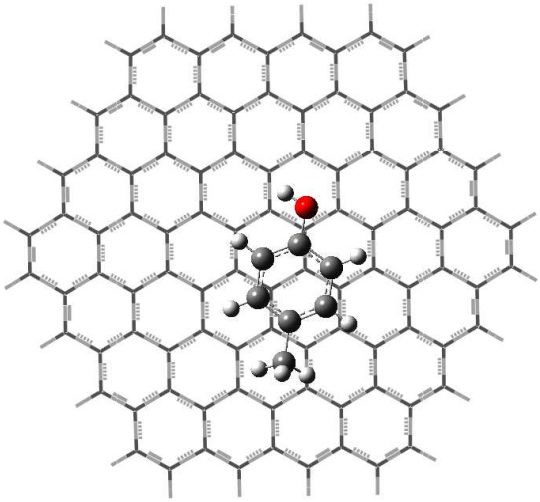
Complex	Model	His	Phe	Tyr	Trp	MUE
C96	PM3-D*	-14.0 (-20.2)	-16.9 (-24.6)	-19.0 (-27.1)	-23.5 (-33.8)	2.4
	MP2 (C48)	-12.7	-14.3	-17.5	-19.3	
[6, 6]	PM3-D*	-11.5 (-15.5)	-12.7 (-18.7)	-14.7 (-20.5)	-17.8 (-25.3)	1.0
	MP2 [5, 5]	-9.2	-12.4	-14.5	-16.6	



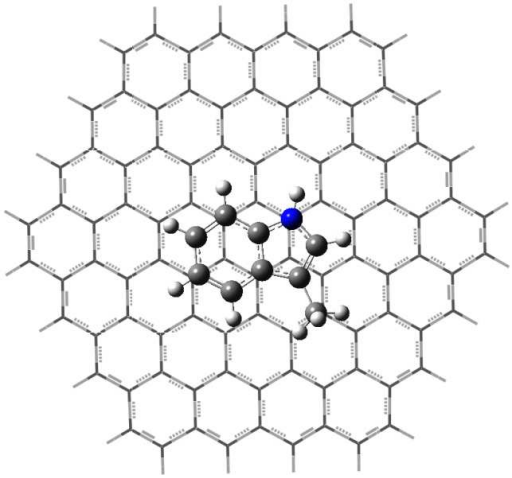
His-C96



Phe-C96

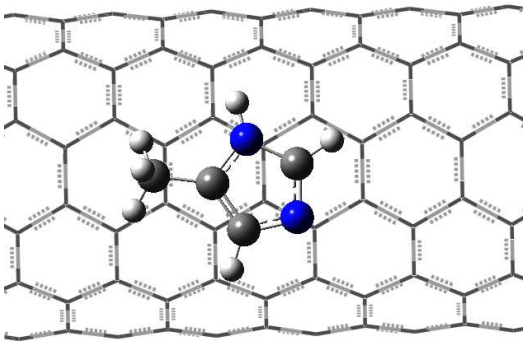


Tyr-C96

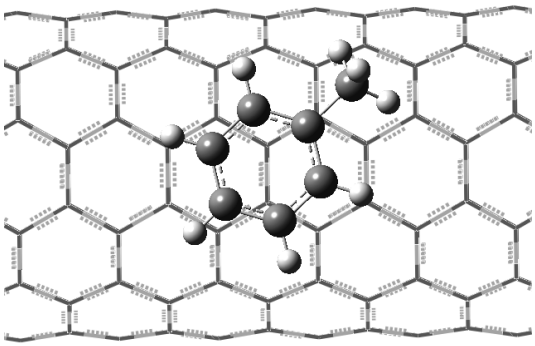


Trp-C96

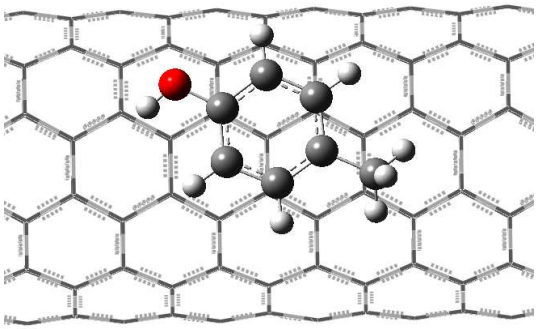
Figure 6.7 Structures of amino acids on C96 optimized by PM3-D*



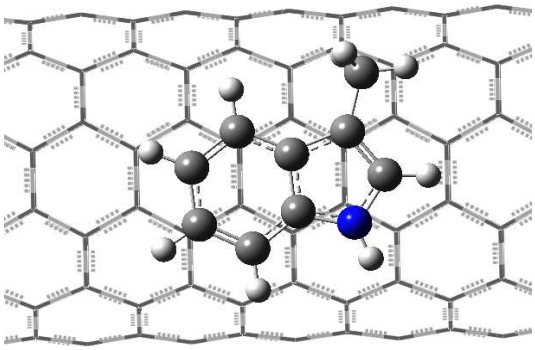
His-[6, 6]



Phe-[6, 6]



Tyr-[6, 6]



Trp-[6, 6]

Figure 6.8 Structures of amino acids on [6, 6] nanotube optimized by PM3-D*

6.3 Conclusions

We have used dispersion corrected semi-empirical methods PM3-D and PM3-D* to study absorption of biomolecules like nucleic acid bases, aminoacids and substituted tryptophan analogs on graphene and nanotubes. We have validated semi-empirical methods by comparison with a number of calculations of DFT-D and M0x functionals for a number of complexes involving graphene sheets. We find that PM3-D and PM3-D* reasonably perform better in line with DFT-D method. The PM3-D* method which includes modified core-core repulsion term in PM3-D and reparameterized specifically to study carbohydrate-aromatic interactions was found to give better interaction energies for biomolecules-nanotube complexes. We have studied the structure and energetics of nucleic acid bases, aminoacids and tryptophan derivatives on various graphene sheets and on nanotubes in detail. We have studied the absorption of biomolecules by taking two kind of nanotubes arm-chair [6, 6] and zig-zag [5, 0] nanotubes. We find that the interaction of nucleic acid bases follow the order of Guanine> Adenine> Thymine> Cytosine> Uracil on graphene sheets and nanotubes. We have tested M0x functionals to study the complexes involving C24, C54 and C96 with nucleic acid bases. We find that the M06-2X gives interaction energies close to DFT-D values compared to other functionals of M0x family. We also find that there is a considerable increase in the error for the larger sheets, which can be attributed to the failure of these functionals to properly describe the longer range dispersive interactions which are predicted to be important by the DFT-D treatment. The interaction energies are close to those predicted by the more computationally demanding DFT-D and MP2 methods. We have studied the aromatic aminoacids (His, Phe, Tyr, Trp) absorption on C96 graphene sheet and [6, 6] nanotube. The interaction energy of aromatic aminoacids follows the order of Trp> Tyr> His> Phe. The PM3-D* method gives the interaction energy of aromatic aminoacids on C96 and [6, 6] nanotube with MUE value of 2.4 and 1.0 kcal mol^{-1} with respect to MP2 values respectively. We find that the PM3 based interaction energies are superior to those obtained by the periodic plane-wave pseudopotential approach within both the GGA and LDA approximation of density functional theory. We find that the dispersion contribution is greater than overall binding energy in all studied complexes. Thus, the stabilization of complexes is dependent on the inclusion of dispersion term.

The semi-empirical schemes PM3-D and PM3-D* methods give the interaction energies close to DFT-D method, which have been shown to be successful method to study noncovalent interactions. These semi-empirical methods are also faster and cheaper compared to most computationally expensive methods like MP2. Therefore PM3-D and PM3-D* methods can be used as an alternative to DFT-D method to study the complexes involving noncovalent interactions.

6.4 References

1. Iijima, S. *Nature.*, **1991**, *345*, 56.
2. Kroto, H. W.; Heath, J. B.; O'Brien, S. C.; Curl, R. F.; Smalley, R. E. *Nature.*, **1985**, *318*, 162.
3. Zhang, Z.; Charles M. L. *Appl. Phys. Lett.*, **1993**, *62*, 22.
4. Geim, A. K.; Novoselov, K. S. *Nat.Mater.,* **2007**, *6***,** 183.
5. Zheng, M.; Jagota, A.; Semke, E. D.; Diner, B. A.; Mclean, R. S.; Lustig, S. R.; Richardson, R. E.; Tassi, N. G. *Nat. Mater.*, **2003**, *2*, 338.
6. Saito, R.; Fujita, M.; Dresselhaus, G.; Dresselhaus, M. S. *Appl. Phys. Lett.*, **1992**, *60*, 2204.
7. Gowtham, S.; Scheicher, R. H.; Pandey, R.; Karna, S. P.; Ahuja, R. *Nanotechnology.*, **2008**, *19*, 125701.
8. Antony, J.; Grimme, S. *Phys. Chem. Chem. Phys.*, **2008**, *10*, 2722.
9. Cousins, B. G.; Das, A. K.; Sharma, R.; Li, Y.; McNamara, J. P.; Hillier, I. H.; Kinloch, R.; Rein, V. *Small.*, **2009**, *5*, 587.
10. Wang, Y. *J. Chem. Phys.*, **2008**, *112*, 14297.
11. Tu, X.; Manohar, S.; Jagota, A.; Zheng, M. *Nature*, **2009**, *460*, 250.
12. Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.*, **2006**, *8*, 1985.
13. Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.*, **2008**, *120*, 215.
14. Morgado, C. A.; Vincent, M. A.; Hillier, I. H.; Shan, X. *Phys. Chem. Chem. Phys.*, **2007**, *9*, 448.
15. McNamara, J. P.; Hillier, I. H. *Phys. Chem. Chem. Phys.*, **2007**, *9*, 2362.
16. Sharma, R.; McNamara, J. P.; Raju, R. K.; Vincent, M. A.; Hillier, I. H.; Morgado, C. A. *Phys. Chem. Chem. Phys.*, **2008**, *10*, 2767.
17. Grimme, S. *J. Comput. Chem.*, **2004**, *25*, 1463.
18. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.;

Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. GAUSSIAN 03 (Revision C.02), Gaussian, Inc., Wallingford, CT, **2004**.

19. Antony, J.; Grimme, S. *Phys. Chem. Chem. Phys.*, **2008**, *10*, 2722.
20. Rajesh, C.; Majumder, C.; Mizuseki, H.; Kawazoe, Y.; *J. Chem. Phys.*, **2009**, *130*, 124911.
21. Hohenstein, E. G.; Chill, S. T.; Sherill, C. D. *J. Chem. Theory. Comput.*, **2008**, *4*, 1996.
22. Raju, R. K.; Ramraj, A.; Hillier, I. H.; Vincent, M. A.; Burton, N. A. *Phys. Chem. Chem. Phys.*, **2009**, *11*,3411.
23. Watson, J. D.; Crick, F. H. C. *Nature.*, **1953**, *171*, 737.
24. Chen, J.; Seeman, N. C. *Nature.*, **1991**, *350*, 6319.
25. McNamara, J. P.; Berrigan, S. D.; Hillier, I. H. *J. Chem. Theory Comput.*, **2007**, *3*, 1014.

Chapter 7

Theoretical study of crystal structures of thin film transistors: Anthracene and Pentacene derivatives

7.0 Introduction

Understanding the properties and practical applications of organic thin film transistors (OTFTs) in various technologies has become crucial in recent years. Although inorganic semi-conductors have been improved markedly, certain aspects such as critical limit for miniaturization and integration lead to development of organic thin film transistors.[1] Organic semiconductors are light, flexible, easy to fabricate over a large area and can be printed as thin films, whereas inorganic semiconductors lack these properties. Eventhough inorganic semiconductors have higher charge career mobility; some organic semiconductors such as pentacene or rubrene have achieved high mobility with values larger than 1 $cm^2V^{-1}s^{-1}$. However there are certain challenges in processing OTFTs such as, 1. Low solubility, 2. High volatility and 3. Less stability which would make synthesis and deposition of organic thin-films more difficult.[2-9]

Polyacene molecules fused of n-benzene rings are promising candidates as organic semiconductors. Due to the planarity and π-π interactions of acene molecules, densely packed crystalline structures can be easily obtained. Therefore it becomes necessary to understand the crystal packing of these organic semiconductors to develop compact thin-film transistors with high charge career mobility. Electronic and optical properties

of polyacenes such as naphthalene (n=2), anthracene (n=3), tetracene (n=4) and pentacene (n=5) have been investigated. Although pentacene is most promising candidate due to high charge career mobility among polyacenes, poor solubility in organic solvents and instability to acid resist practical applications of this organic semiconductor. Moreover molecular arrangement in the crystal forms is very important to achieve more π-stacking. For example, since pentacene has herringbone crystal structure form where one molecule inclined to other, π-π interaction between the molecules will be less (Figure 7.0).[10] In order to achieve more π-stacking, it becomes important to obtain crystal structure where molecules will be planar to each other. The strong correlation between molecular ordering and TFT mobility shows the importance of understanding the driving force behind the molecular ordering in crystal structures. The challenge in molecular ordering and increasing field-effect mobility leads to researchers to find an approach to tackle this problem. One approach is to functionalize the organic molecules to form molecular crystals with increased π-orbital overlap.[11-14] Adding bulky functional groups at 6, 13 positions of pentacene has found to be discouraging edge-to-face interactions and encouraging face-to-face molecular interactions. This approach has been found to improve π-orbital overlap.[15]

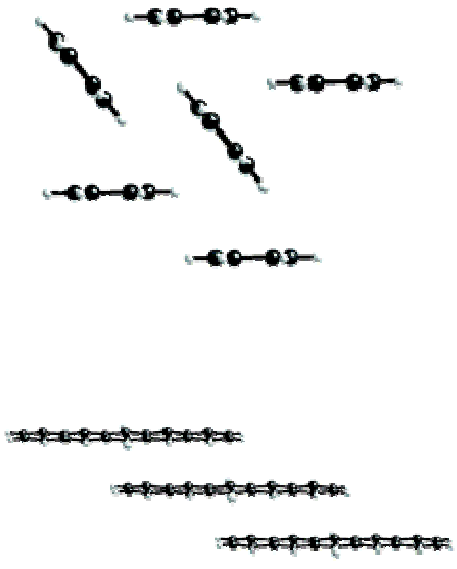


Figure 7.1 Herringbone (top) and π-stacking arrangement of acenes

7.1 Previous studies

In last decade, researchers have been attracted to p-type organic semiconductors, such as derivatives of polycyclic aromatic hydrocarbons anthracene, pentacene, rubrene and thiophene for their applications in high performance organic thin-film transistors.[16] Anthracenes have high oxidation potential, small π-surface, and high fluroscence quantum yields. The covalent adding of anthracene decreases the oxidation potential of the molecule while simultaneously increasing potential area of intermolecular overlap.[17] Synthesis and characterization of series of functionalized anthracenes were reported by adding different substituents. Fluoro and aryl Subsitituents were added to anthracene and tested for hole mobility.[18,19] The addition of alkyl groups led to material with significantly improved performance, both due to the subtle change in packing of the molecules, leading to both edge-to-face and face-to-face interactions.[20] In order to create amorphous materials to minimize aromatic interactions in the solid state, molecules need to be designed that crystallize with their chromophores completely isolated by electronically insulating functional groups. One approach to achieve this is to add bulky alkyl silylethyne groups to anthracene at 9, 10 positions. Likewise, the functionalized pentacenes were developed to design thin film transistors. The first functionalized pentacene studied for transistor applications was 2, 3, 9, 10-tetramethyl pentacene. There is no difference found in the arrangement of crystals between this compound and unsubstituted pentacene. Both molecules showed the herringbone arrangement in crystal form. Varying the nature and location of substituent groups around pentacene has led to an array of solid-state arrangements. The detailed study of these arrangements elucidated the effect of intermolecular order on charge transport. Among variety of intermolecular arrangements, thienlyl functionalized pentacene showed the long-range π-stacking order. Main reason to add property-modifying substituents to the aromatic ring of an acene is to hold them from the ring by a rigid, sterically undemanding spacer. One of the substituent to achieve that effectively is found to be alkyne substituent. Altering the size of the substituent on the alkyne provides good control over the π-stacking order. Spherical substituents such as tert-butyl, trialkyl-silyl for which the substituent diameter less than half the length of the acene leads to a 1-D, “slipped-stack” arrangement. Substituent diameter very close to half the length of the acene, the molecules show a 2-D “bricklayer” arrangement. If the

substituent size is increased further, edge-to-face interactions between the subsituent and the aromatic chromophore begin to dominate, since the volume of the substituent is able to cover the aromatic surface and the herringbone arrangement becomes the preferred solid-state arrangement. Performance of thin-film transistor is related to the crystal packing of the molecules. Among the functionalized pentacene molecules, materials with silyl substituents which show 2-D face-to-face interactions yield high-performance thin-film devices.

7.2 Results and Discussions

The aim of this project is to understanding the packing of anthracene and pentacene and their derivatives in crystal structures and finding out the ways to develop potential TFTs. We studied silyl substituted crystal structures of anthracene and pentacene using quantum mechanical methods and molecular dynamics methods. This chapter has been classified mainly into two parts, 1. Quantum mechanical studies of silyl substituted anthracene and pentance 2. Molecular dynamics studies of anthracene and substituted anthracenes in different solutions.

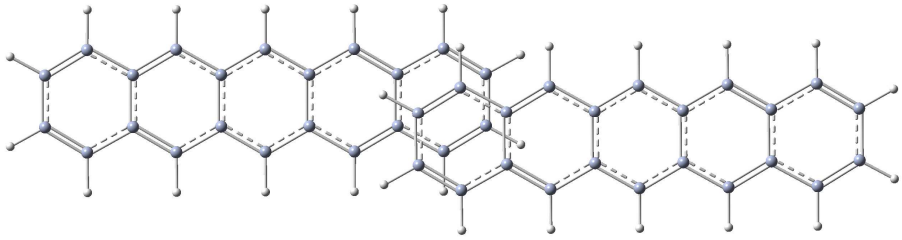
7.2.1 Quantum mechanical studies of silyl substituted anthracene and pentacene

This part has been classified further into two sections i.e. 1. anthracene and their derivatives and 2. pentacene and their derivatives. We started this project in order to find an answer for the different orientations of stacked structures. The crystal structures of anthracene and derivatives of anthracene form different stacked structures. The question arises cause of different stacked structures is whether due to dispersion forces or due to steric effects the way they are. Therefore we performed quantum mechanical calculations to find different stacked structures and the lowest energy structures. To find insights into the stacked structures, we started to perform the calculations by taking smallest model i.e., dimer where dispersion forces are high and highly possible structures can be obtained. Then we increased the number of molecules to analyse the steric effect. We have performed quantum mechanical calculations from dimer to tetramer for different anthracenes and pentacenes. We introduced the bulky silyl groups

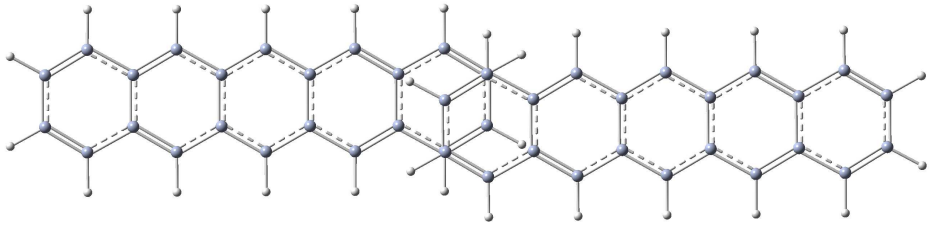
in the middle of anthracene and pentacene molecules on both sides. We also introduced the bulky vinyl groups in the middle of anthracene molecules on both sides. We calculated the interaction energies along with dispersion contribution using semi-empirical method implemented with dispersion term, i.e., PM3-D method.

7.2.2 Labelling the overlapped structures

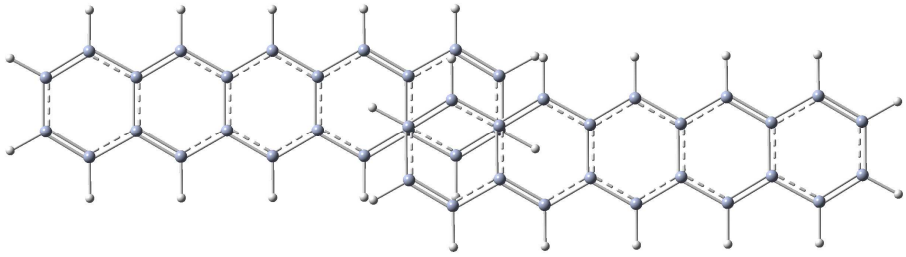
The overlap pattern of stacked anthracene and pentacene molecules will be summarized by a label identifying the number of overlapped rings and the way that they overlap. Some examples of the naming convention are shown in Fig.7.1. They are ***n*b**, ***n*c** and ***n*o**, where ***n*** represents number of phenyl ring (***n***=1-5) and **b**, **c** and **o** represent centre of phenyl ring of one pentacene molecule is on top of the **bond** of n^{th} phenyl ring of other anthracene/pentacene, centre of phenyl ring of one anthracene/pentacene molecule pointed towards the **carbon** of n^{th} phenyl ring of other anthracene/pentacene, centre of phenyl ring of one anthracene/pentacene molecule is **offset** from n^{th} phenyl ring of other anthracene/pentacene, respectively.



1b (bond of the phenyl ring)



1c (carbon of the phenyl ring)



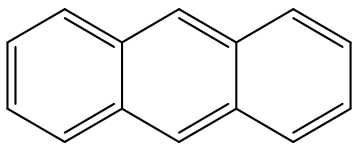
1o (ring is offset with respect to other)

Figure 7.2 Examples for 1b, 1c and 1o structures

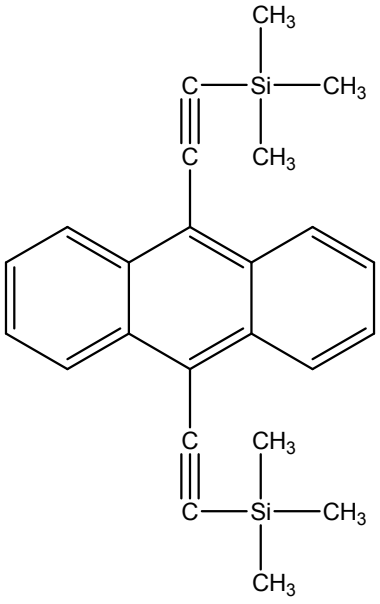
7.2.3 Anthracene and their derivatives

The studied molecules are 1) Anthracene, 2) 9, 10-trimethylsilyl anthracene, 3) 9, 10-triethylsilyl anthracene, 4) 9, 10-triisopropylsilyl anthracene, 5) 9, 10-(di(dimethylbenzene)ethynyl) anthracene, 6) 9, 10-(di(propylbenzene)ethynyl) anthracene. The chem draw picture of these structures is shown in Figure 7.2.

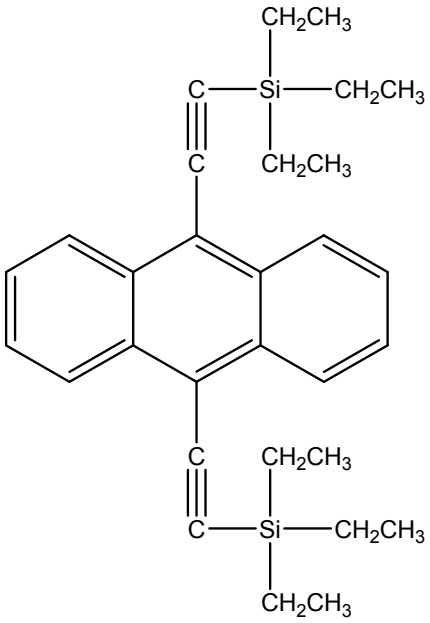
The initial models were built by placing one molecule on top of another molecule at possible different stacking positions. We have optimized these initial structures by using PM3-D method. We have calculated the interaction energy of stacked structure along with dispersion contribution. Throughout the calculations, we have used semi-empirical method with dispersion correction term PM3-D, which has already been proved to be a successful method to study noncovalent interactions.



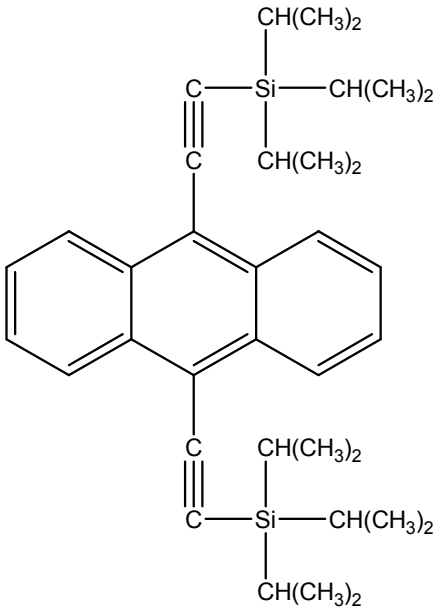
1



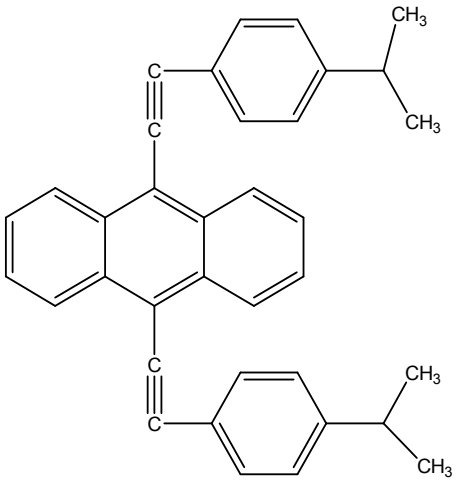
2



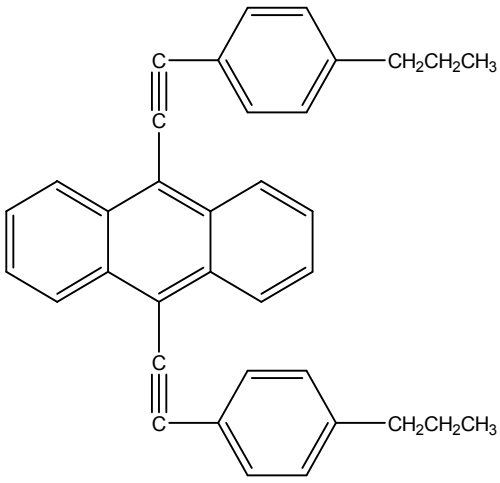
3



4



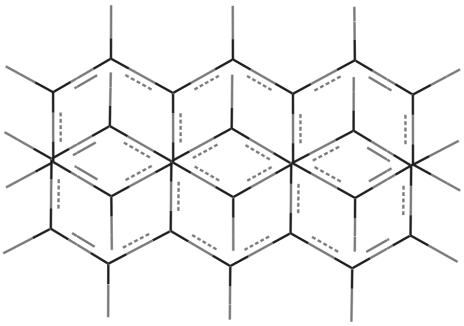
5



6

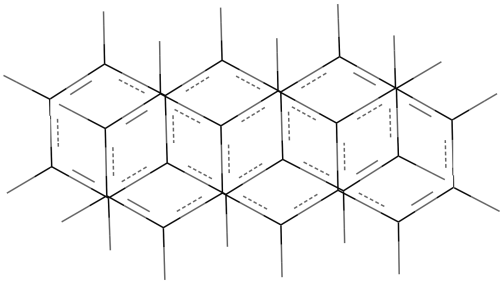
Figure 7.3 Chemical structures of anthracene and their derivatives

For anthracene, we have obtained four different stacked structures. The four stacked structures are i) Cross-stacked, (CS) ii) Parallel-displaced, (PD) iii) Graphite-type (GT) and iv) T-stacked (TS) structures. The three structures (i-iii) (Fig. 7.3) are very close in energy, differ by less than 0.3 kcal mol^{-1} and T-stacked structure is lowest stable structure with interaction energy –12.8 kcal mol^{-1}. The interaction energy of parallel-displaced structure is –13.4 kcal mol^{-1}, which is 0.1 kcal mol^{-1} less than graphite-type structure, for which interaction energy is –13.5 kcal mol^{-1}. T-shaped structure is being the least stable among four of them having –12.8 kcal mol^{-1} as interaction energy. Cross-stacked has been found to be most stable structure having –13.7 kcal mol^{-1}. All these structures were shown in Figure 7.3.



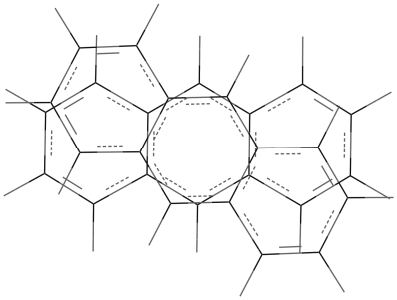
Parallel-displaced structure

I.E.=-13.4 kcal mol^{-1}



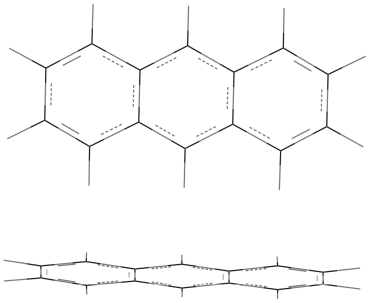
Graphite-type structure

I.E.=-13.5 kcal mol^{-1}



Cross-stacked structure

I.E.=-13.7 kcal mol^{-1}



T-shaped structure

I.E.=-12.8 kcal mol^{-1}

Figure 7.4 Different stacked structures of anthracene

7.2.31 Anthracene and silyl substituted anthracenes

The dimers of **1**, **2**, **3**, **4**, **5** and **6** are modelled by placing one molecule on top of another molecule in different overlapping points such as **1b**, **1c**, **1o**, **2b**, **2c**, **2o, 3b, 3c**, **3o** and optimized using PM3-D method. We have calculated the interaction energy and dispersion energy for all the stacked structures. We have tabulated all these energies in Table 7.0. Dispersion energy plays an important role in all these stacked structures. The major contribution comes from dispersion energy to interaction energy in most of these stacked structures except few structures. Without dispersion energy, these structures are destabilized. Dimer **2** exhibited three stacked structures 3c-GT, CS and TS which are named according to overlapping method. Among these three structures, cross-stacked structure shows higher interaction energy (-51.3 kcal mol^{-1}) than other two structures. T-stacked structure is least stable one similar to T-stacked structure of anthracene with -17.4 kcal mol^{-1} interaction energy. We have obtained two cross-stacked structures for dimer **3** with interaction energy of -53.2 and -64.1 kcal mol^{-1} respectively. Dimer **4** gives two stacked structures 1b and 1c with -50.4 and -55.0 kcal mol^{-1} respectively. In both these dimers, one molecule on top of another stacks through by interacting only one ring. In **1b** structure, C-C bond of one anthracene molecule points towards the centre of the ring of another anthracene molecule. In **1c** structure, carbon of one anthracene molecule points towards the centre of another molecule. The lack of dispersion in these two structures is apparent from the less dispersion contribution to interaction energy. We have obtained only cross-stacked structure for dimer **5** with interaction energy of -62.1 kcal mol^{-1}. We have located three structures for dimer **6** such as CS, 3c-GT and 3o with interaction energy of -63.9, -58.6 and -56.7 kcal mol^{-1} respectively. We have found that cross-stacked structure is most stable compared to other structures. T-stacked structure is least stable structure due to the lack of dispersion interaction compared to other structures. The reason for the less distribution energy of structures **1b** and **1c** of dimer 4 may arise from the bulky spherical substituent isopropyl group preventing aromtatic rings to stack with each other.

Table 7.1 Interaction energies and dispersion contributions (in kcal mol^{-1}) of anthracene and their derivatives dimers

Molecules	Types	ΔE_{int} (ΔE_{disp})
1	Parallel-displaced	-13.4 (-21.4)
	Graphite-type	-13.5 (-21.6)
	Cross-stacked	-13.7 (-21.7)
	T-shaped	-12.8 (-12.0)
2	3c-GT	-40.9 (-42.7)
	CS	-51.3 (-55.4)
	TS	-17.4 (-15.8)
3	CS1	-53.2 (-54.5)
	CS2	-64.0 (-63.6)
4	1b	-50.4 (-40.0)
	1c	-55.0 (-44.3)
5	CS	-62.1 (-66.9)
6	CS	-63.9 (-67.9)
	3c-GT	-58.6 (-64.1)
	3o	-56.7 (-63.7)

7.2.4 Pentacene and silyl substituted pentacenes

We studied three crystal structures of substituted pentacene derivatives to explore the influence of dispersion and steric effects in crystal packing. The crystal structures of following have been considered.

1. Pentacene
2. 6, 13-bis ((triisopropylsilyl) ethynyl) pentacene (K-TIPS),
3. 1, 4, 8, 11-methyl, 6, 13-bis ((triethylsilyl) ethynyl) pentacene (TMTES),
4. 1, 4, 8, 11-methyl, 6, 13-bis ((triisopropylsilyl)ethynyl) pentacene (TMTIPS)

It is assumed that steric effects and dispersion contribution have major roles in crystal packing of these crystal structures. Since pentacenes have achieved high electron career mobility, they have received much attention in research of organic thin film transistors. Because of herringbone packing of pentacene molecules, where one pentacene is inclined to other pentacene, one can expect only less π-stacking between these molecules. Therefore, the functional groups, particularly ethynyl silyl groups have been introduced to centre ring of the pentacene molecules to make them parallel and π-stacked. However, upon introducing the bulky substituent groups such as ethynyl silyl with methyl groups, the steric effects will also arise. The dispersion or π-stacking between the pentacene molecules, which will bind them together and make them stable, will be higher if higher number of phenyl rings of one pentacene π-stacked with phenyl rings of other pentacene. In these crystal structures, only two to three phenyl rings of one pentacene are π-stacked above the phenyl rings of other pentacene. Therefore to explore the overlap pattern, we attempted to study these crystal structures theoretically.

7.2.41 Pentacene dimers

The semi-empirical method PM3 implemented with a dispersion term, i.e., PM3-D, which can describe dispersion interaction in reasonable agreement with high-level methods, has been used throughout this study. First of all, as a matter of interest, π-stacked pentacene dimer models 5b, 5c, 5o, 4b, 4c, 4o, 3b, 3c and 3o have been

modelled and energy minimized using PM3-D. Interaction energies have been calculated on these dimers and summarized in Table 7.1.

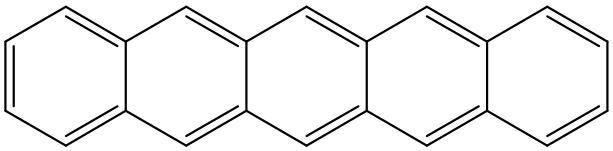
Table 7.2 Interaction energies and dispersion contribution (in kcal mol^{-1}) of pentacene dimers

Dimer models	I.E. (Disp)
5b	-22.48 (-35.28)
5c	-22.52 (-36.10)
5o	-22.27 (-34.15)
4b	-20.57 (-29.86)
4c	-19.73 (-29.73)
4o	-21.04 (-30.57)
3b	-16.56 (-24.47)
3c	-15.48 (-22.82)
3o	-17.36 (-24.92)

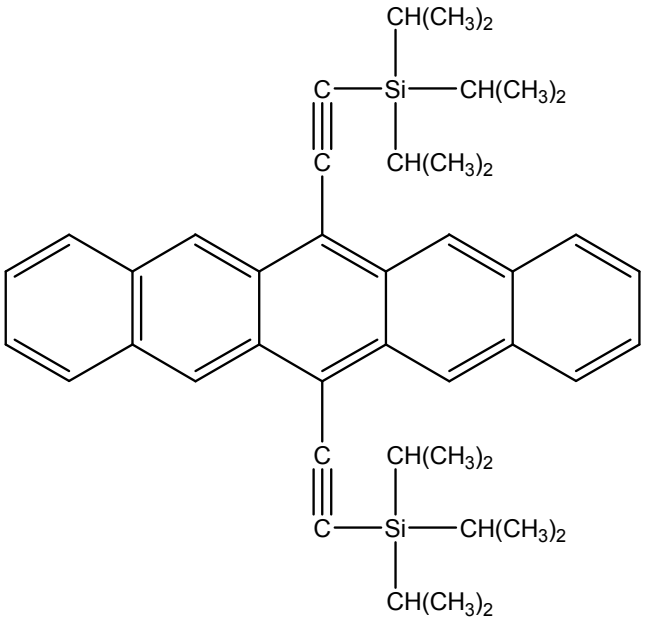
As we expected, 5b, 5c, 5o dimer models have got higher interaction energies than other models. The interaction energies decrease with the number of phenyl rings decrease.

7.2.42 Ring overlap in the crystal structures

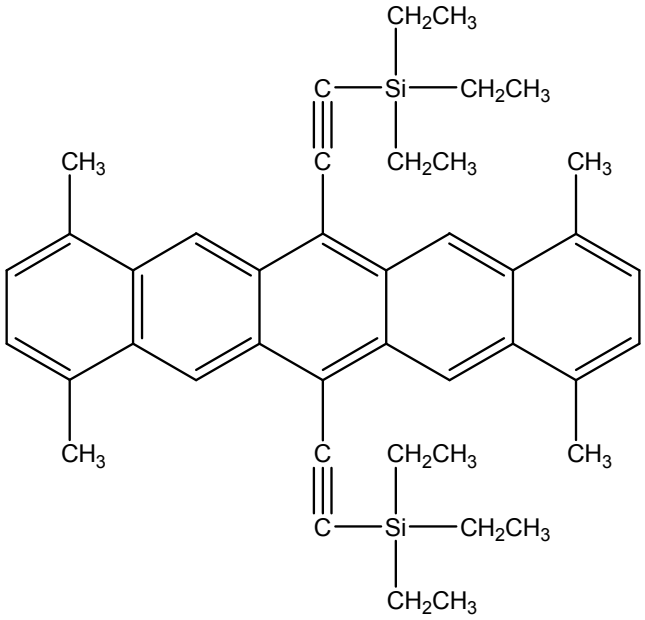
In the K-TIPS crystal structure, the overlapping pattern is **3c**, since 1st ring of one pentacene molecule is stacked above the **3**rd phenyl ring of the other molecule. In the TMTES crystal structure, the overlapping pattern is **3c**, since the 1st ring of one pentacene molecule is stacked above the **3**rd benzene ring of the other molecule. In the TMTIPS crystal structure, the overlapping pattern is 2c, since 1st ring of one pentacene molecule is stacked above the **2**nd ring of the other pentacene molecule. Even though there is a possibility for stacking of all the phenyl rings of one pentacene molecule with phenyl rings of other pentacene molecule to enhance π-stacking between these molecules, they stack through only two or three phenyl rings of pentacene. The possibility for more π-stacking and the role of steric effect due to the bulky silyl substituent groups have been analysed theoretically.



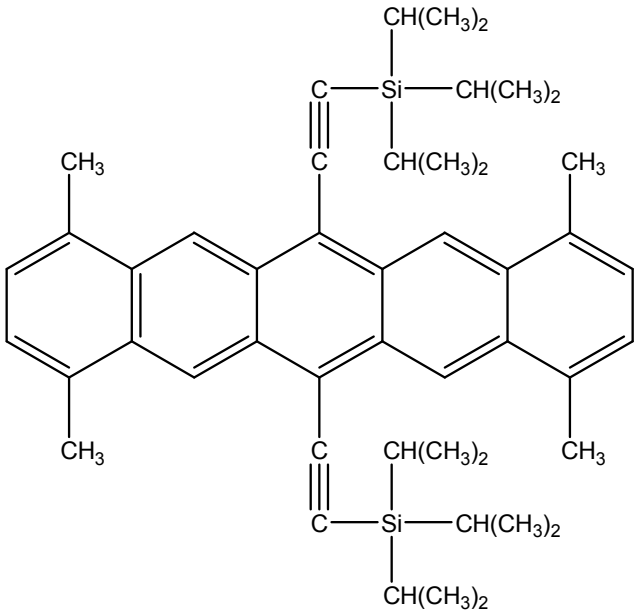
1



2



3

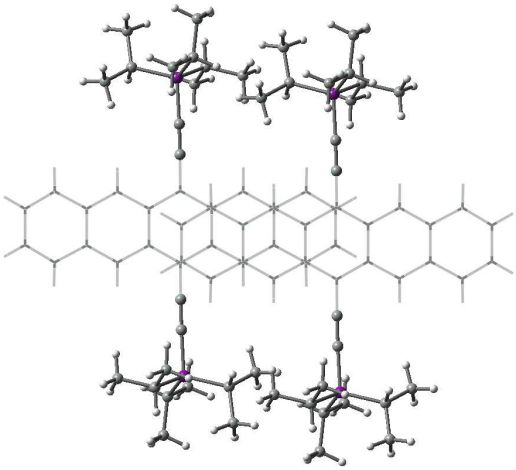


4

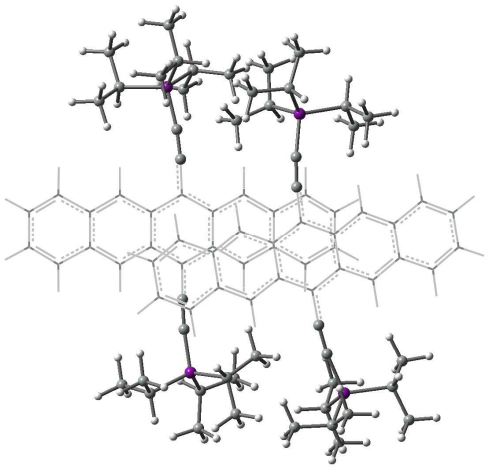
Figure 7.5 Chemical structures of pentacene and their derivatives

7.2.43 Dimer models

First all the calculations have been carried out on dimer models taken from these crystal structures. The dimers of crystal structures and the corresponding minimized dimer structures, interaction energies and dispersion contribution (in parenthesis) are shown in Fig. 7.5.

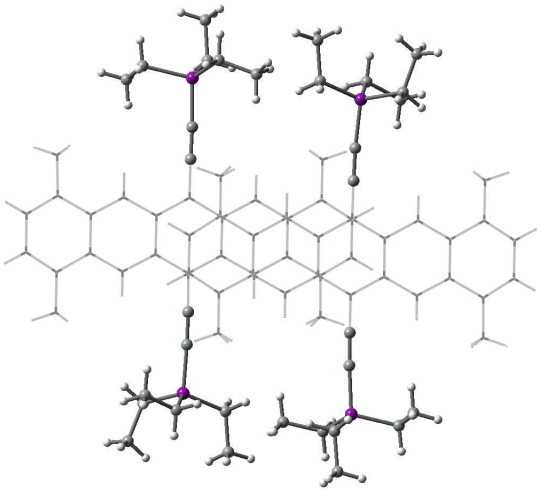


K-TIPS crystal structure

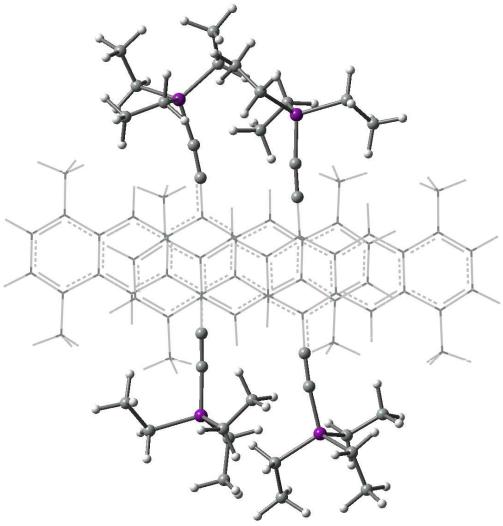


I.E.=67.6 (-65.1) kcal mol^{-1}

PM3-D K-TIPS

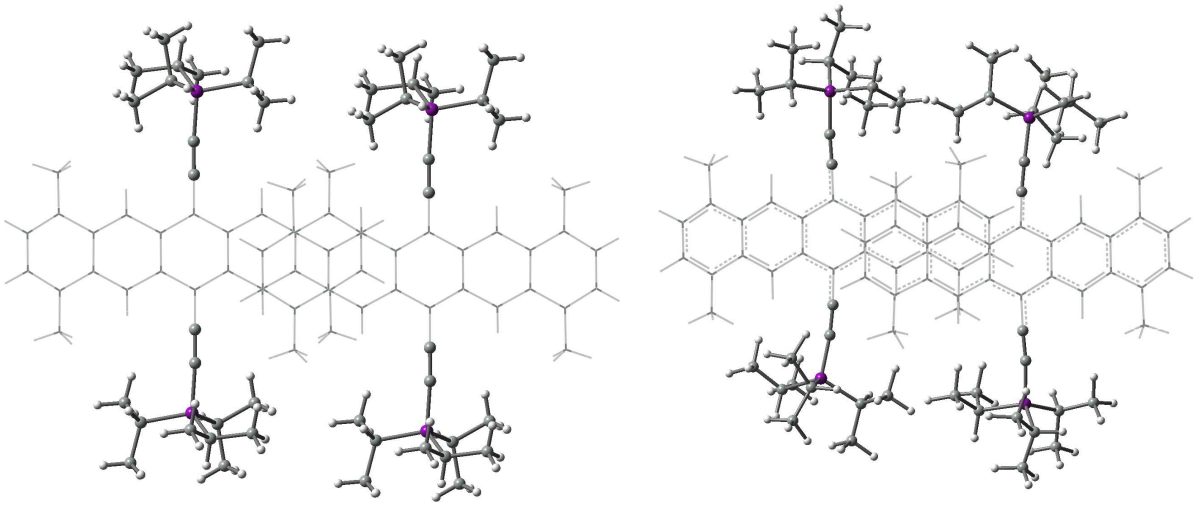


TMTES Crystal structure



I.E.=-83.1 (-81.1) kcal mol^{-1}

PM3-D TMTES



TMTIPS Crystal structure

I.E.= -65.8 (-59.7) kcal mol^{-1}

PM3-D TMTIPS

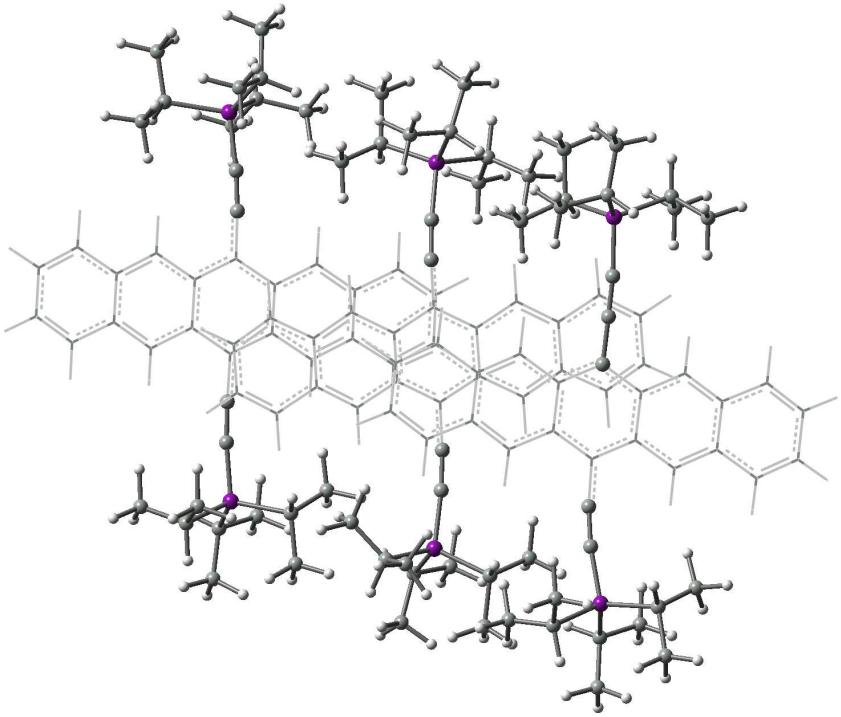
Figure 7.6 Dimers from crystal structures and corresponding PM3-D energy minimized structures

It has been found that K-TIPS tend to overlap to get displaced 3o structure and TMTES tend to overlap to get 4c structure by overcoming the steric effect of bulky groups, in turn with bending of ethynyl silyl chain of pentacene. For TMTIPS, dimer has been taken from crystal structure and modelled as 1o for initial structure and energy minimized. The energy minimized structure is 2o, which is more or less similar to crystal structure. For K-TIPS and TMTES, irrespective of the bulky groups, dimers tend to overlap upto 3^{rd} or 4^{th} phenyl ring of pentacene.

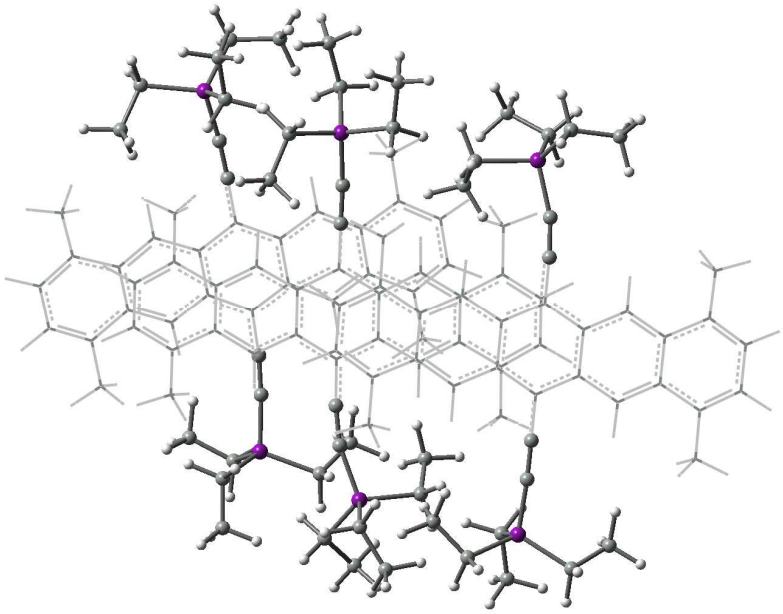
7.2.44 Trimer models

To explore the steric effect, trimers have been taken from these crystal structures and energy minimized. In K-TIPS minimized trimer, two molecules are in 3c position similar to crystal structure, while one molecule is in 4c. In TMTES minimized trimer, two molecules stay in the pattern of 3c, similar to crystal structure, while one molecule

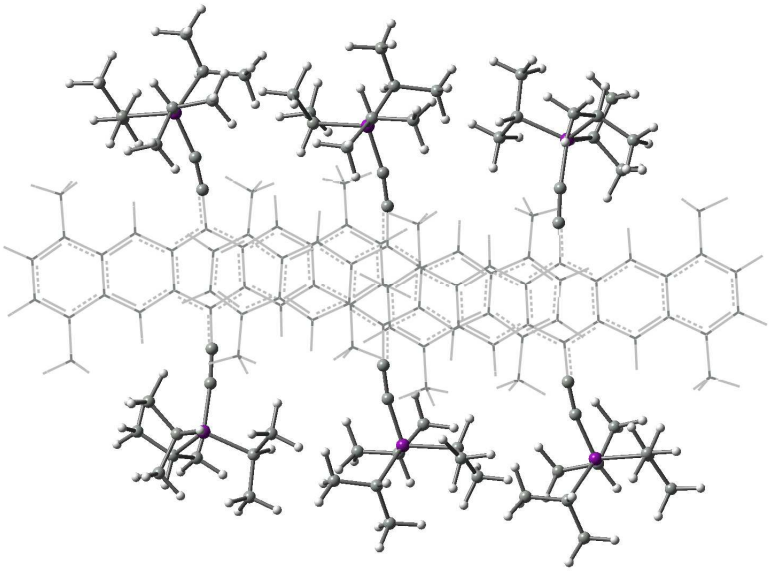
is in 4c and becomes slightly crossed to centre molecule. In TMTIPS minimized trimer, all molecules tend to be in 3c position, while crystal structure is in 2c pattern. It has been found that adding another molecule to dimer prevents the tendency of molecules to go to 4c position i.e., towards more dispersion. With increasing number of molecules or in other words, with increasing steric groups, the structures become more and more similar to crystal structures. Therefore it clearly seems to be the effect of steric bulky groups. The minimized trimer structures, interaction energies and dispersion contribution (in parenthesis) have been shown in Fig. 7.6.



K-TIPS I.E.= -135.9 (-129.1) kcal mol^{-1}



TMTES I.E.= -143.1 (-135.7) kcal mol^{-1}



TMTIPS I.E.= -152.1 (-153.6) kcal mol^{-1}

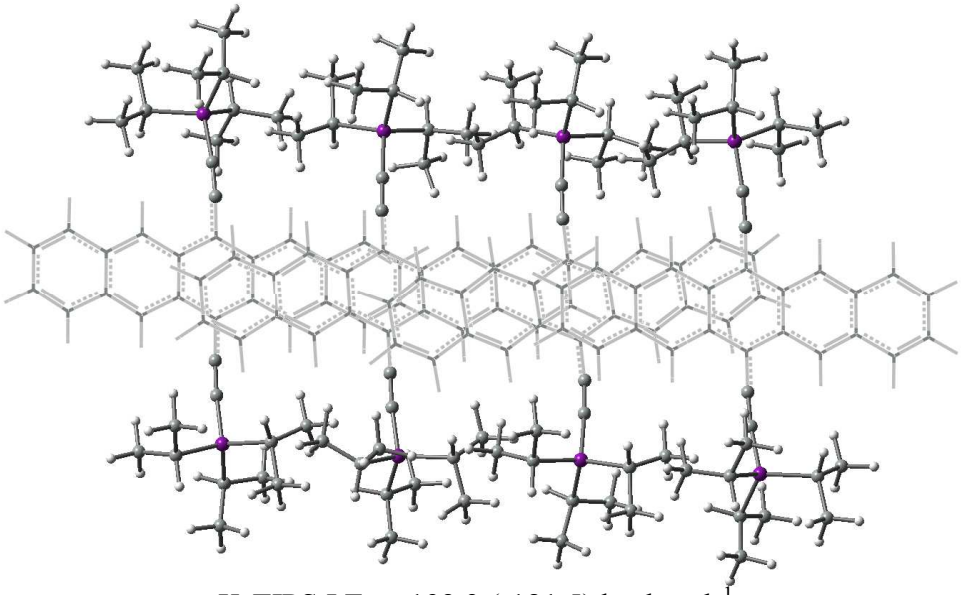
Figure 7.7 PM3-D structures of trimers

7.2.45 Tetramer models

Tetramer has also been taken to support the finding of steric effect. After full relaxation of K-TIPS tetramer, all molecules are in 3c position, similar to crystal structure. It supports the finding that steric effect prevents the molecule to disperse more. It can be concluded that with increasing number of molecules or bulky groups, the molecules tend to overlap less and less and steric effect becomes competitive to dispersion by preventing the molecules to overlap more. It is convincing that crystal structures show the less dispersion pattern since to maintain the balance between steric effect and dispersion. The minimized tetramer structure of K-TIPS, interaction energy and dispersion contribution (in parenthesis) have been shown in Fig. 7.7. We have tabulated the summary of the overlap patterns of crystal and PM3-D structures of functionalized pentacene in Table 7.2.

Table 7.3 Summary of the overlap patterns of crystal and PM3-D pentacene structures

Structures	K-TIPS	TMTES	TMTIPS
Crystal	3c	3c	2c
Dimer	3o	4c	2o
Trimer	3o, displaced 3o	3c, displaced 4o	3c, displaced 3c
Tetramer	3c, 3c, 3b	–	–



K-TIPS I.E.= -192.3 (-181.5) kcal mol^{-1}

Figure 7.8 PM3-D structure of K-TIPS

7.3 Molecular dynamics of Anthracene

We have also employed molecular dynamics simulation to gain insights into the structural and dynamic features of anthracene crystal structure. We have simulated the anthracene dimer in toluene solution at various concentrations. The starting structure for MD setup was obtained from quantum mechanical calculations. We have chosen the PM3-D minimized anthracene dimer for molecular dynamics simulation. The force field parameters of anthracene were obtained using the antechamber module in AMBER, the atomic charges were derived using RESP program. We have used ff03 forcefield for toluene molecules and GAFF forcefield for anthracene molecules. Our main aim of this MD simulation is to find out the existence of anthracene dimer throughout the simulation. Therefore, we have built four different models with same box size of 60 Å and with different number of anthracene molecules such as 2, 4, 8 and 16 molecules. The systems were then minimized to remove the bad contacts in the crystal in two steps.

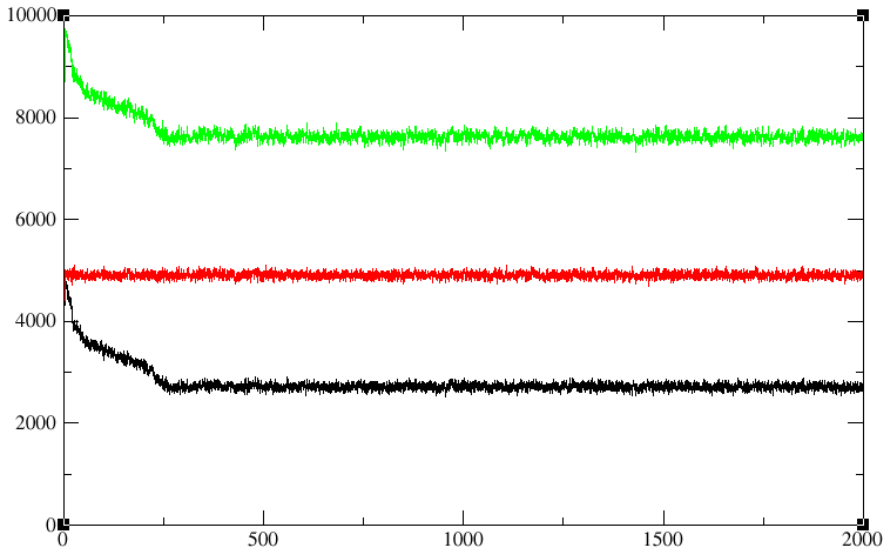


Figure 7.9 Plot of energy vs time, where E_{ktot} refers kinetic energy (green), E_{tot} refers total energy (red), E_{ptot} refers potential energy (black).

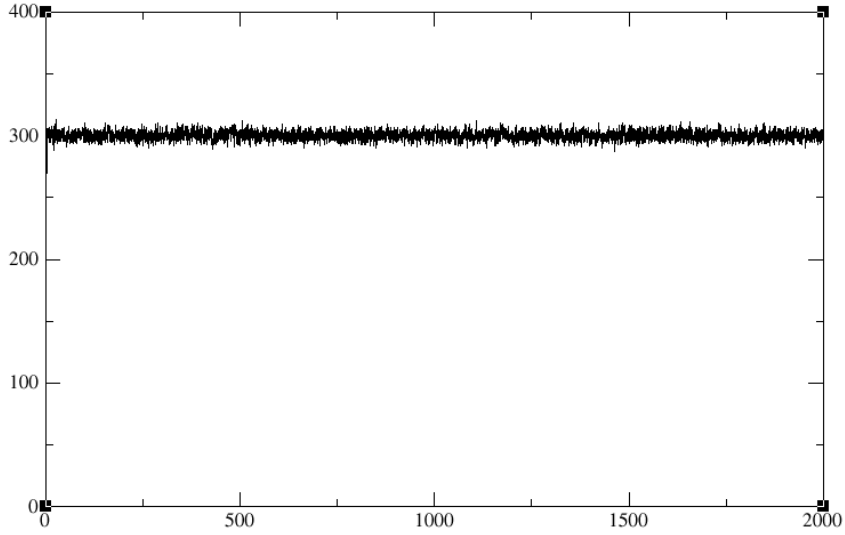


Figure 7.10 Plot of temperature vs time

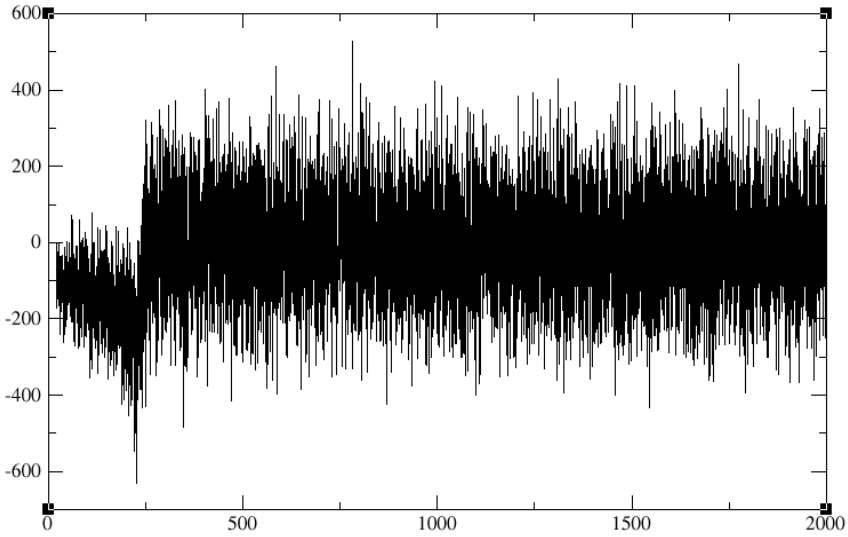


Figure 7.11 Plot of pressure vs time

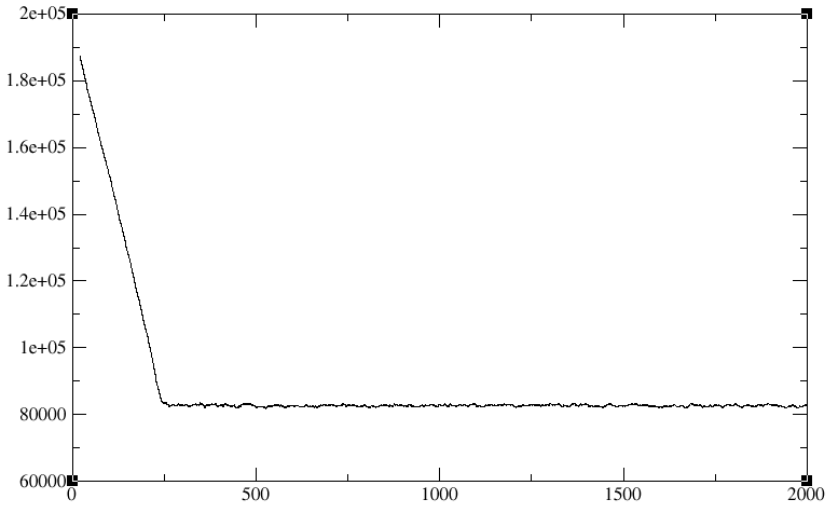


Figure 7.12 Plot of volume vs time

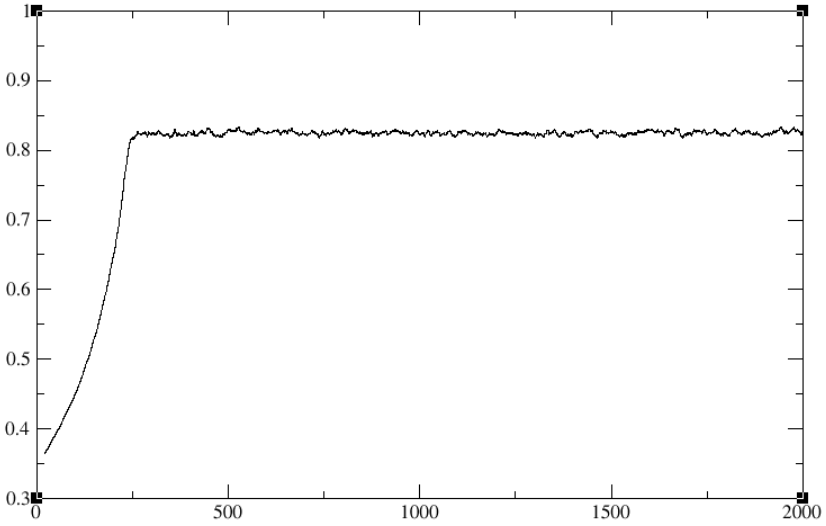


Figure 7.13 Plot of density vs time

First, the solvent molecules were minimized for 500 steps of steepest descent and 500 steps of conjugate gradient while restraining the whole solute molecules with a force constant of 10 kcal mol^{-1} $Å^2$. In the second step, the whole system was minimized for 2500 steps each of steepest and conjugate gradient without any constraints. The same procedure was carried out for all systems of anthracene in toluene solution.

All MD simulations were performed with AMBER9 programme package. The particle mesh Ewald (PME) method was used to treat long-range electrostatic interactions, and a 12 Å cutoff was chosen for nonbonded interactions. The SHAKE method was applied to constrain all of the covalent bonds involving hydrogen atoms. Periodic boundary conditions were applied to all dimensions. The system was then heated from 0 to 300K for first 20ps with weak restraints on anthracene molecules while keeping volume as constant. Then the system was equilibrated to 2ns by keeping pressure as constant. The time steps were 1 fs during equilibration and 2 fs for production dynamics. The successful equilibration has been assessed with change in energy, temp, pressure, volume and density over 2ns. We have shown the change in energy, temp, pressure, volume and density plots to show the successful equilibration of system 4in60Å (Figure

7.8-7.12). We have collected a total number of 200 snapshots at 25ps intervals from the 2ns. We have also calculated the equilibrium distances of anthracene molecules for all snapshots collected. The above procedures are applied for all four systems. For convenience, the four systems are named as 2in60Å (2 anthracene molecules in 60Å box), 4in60Å (4 anthracene molecules in 60Å box), 8in60Å (8 anthracene molecules in 60Å box) and 16in60Å (16 molecules in 60Å box).

We have calculated the distances between centers of mass of the two anthracene molecules for all the systems. For two molecules in 60Å box, there is only one pair of distances available and the distances were calculated for 200 snapshots collected from system 2in60Å. We have plotted the distances in Figure 7.13. Since the distance of stacked anthracene is ~4Å around calculated by theoretical methods,[24] we are interested in the structures with distances below 5Å. From the plot, it can be seen that there is only few points available below 5Å, which implies the existence of few dimers in the beginning of the few ps simulation. And there are few more dimers can be seen around 400ps simulation. However, it is evident that most of the time during simulation, molecules stay apart with distance of more than ~5Å which show the existence of less number of dimers.

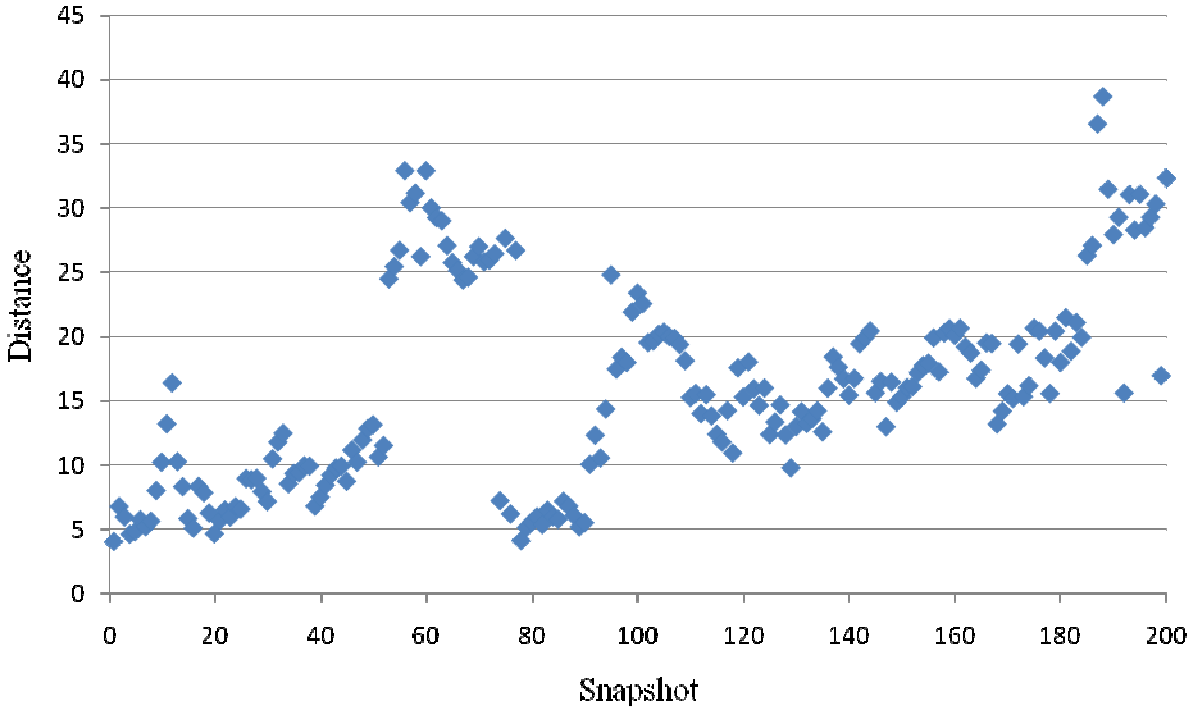


Figure 7.14 Distance between the centers of mass of two anthracene molecules (in 2in60Å)

The system with 4 anthracene molecules in 60Å box was equilibrated and simulated for 2ns. From the 2ns simulated system, we have collected 200 snapshots at the interval of 25ps. Since there are four anthracene molecules, the possible number of distances between two anthracene molecules is six. These six pairs of distances between anthracene molecules were calculated and plotted (Figure 7.14). There are some points below 5Å which shows the existence of dimer. Compared to the system 2in60Å, the probability of finding a dimer is increased, since more number of points below 5Å can be seen until ~1.5ns simulation.

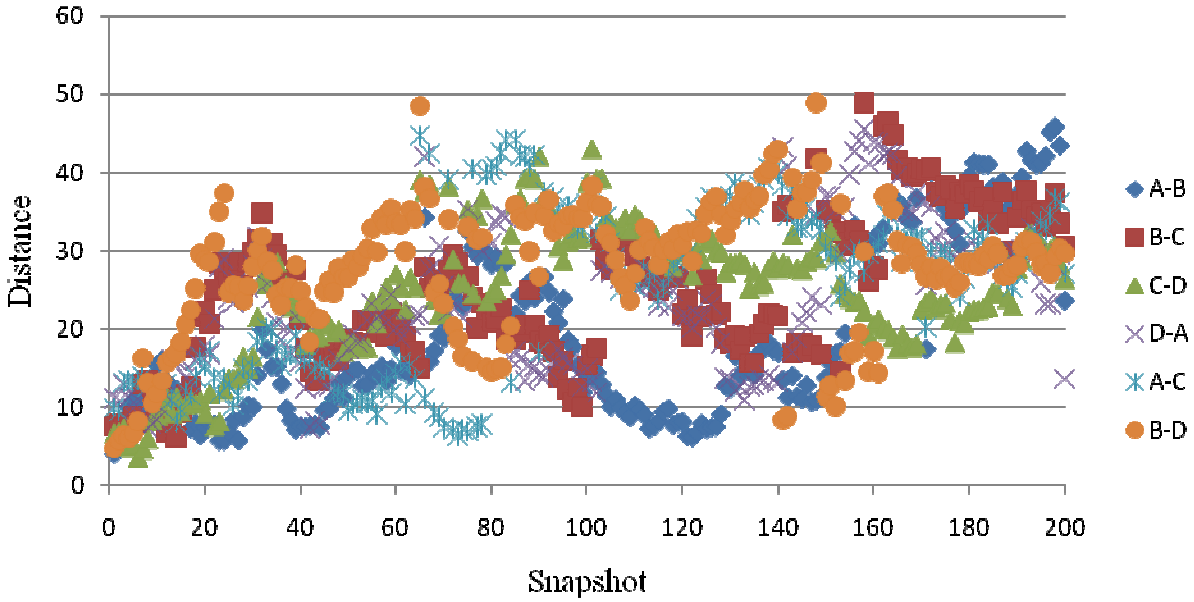


Figure 7.15 Distance between the centers of mass of two anthracene molecules (in 4in60Å)

We have applied the same procedure for the system 8in60Å, as for 2in60Å and 4in60Å. We have equilibrated and simulated for 2ns simulation and collected the snapshots. For 8 numbers of molecules, the number of pair of distances available is 28. Since the plot of 28 numbers of distances for 200 snapshots is not clear for analysis, we have taken the number of counts of dimers below 5Å distance for each snapshot and plotted which can be seen in Fig. 7.15. It can be seen that there is minimum one dimer existing until the simulation of ~1000ps. When we increase the number of molecules of anthracene in

60Å box from four to eight molecules, the probability of finding a dimer is also increased which is evident from the graph (Figure 7.15).

For the system 16in60Å, we have applied the same procedure as other systems and collected 200 snapshots at 25ps interval from 2ns simulation. For 16 anthracene molecules, the number of distances can be obtained is 120. For the convenience, the numbers of dimers for each snapshot below 5Å were plotted and are shown in Fig. 7.16. From this graph, it can be seen that there is maximum 2-5 numbers of distances available below 5Å throughout the simulation of ~1100ps. From the plot, it is also evident that there is always a dimer exist throughout the simulation.

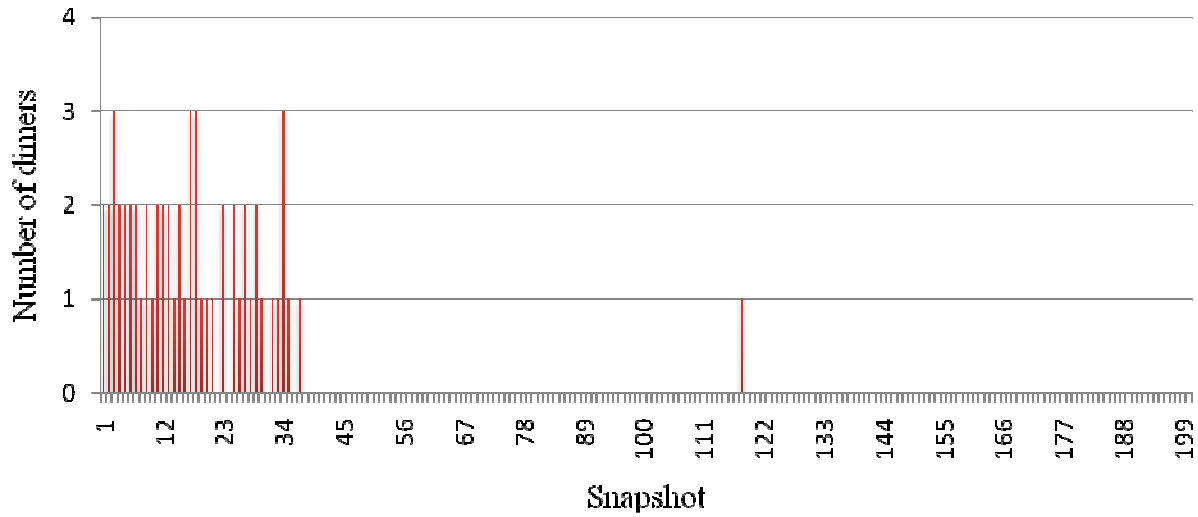


Figure 7.16 Distance between the centers of mass of two anthracene molecules (in 8in60Å)

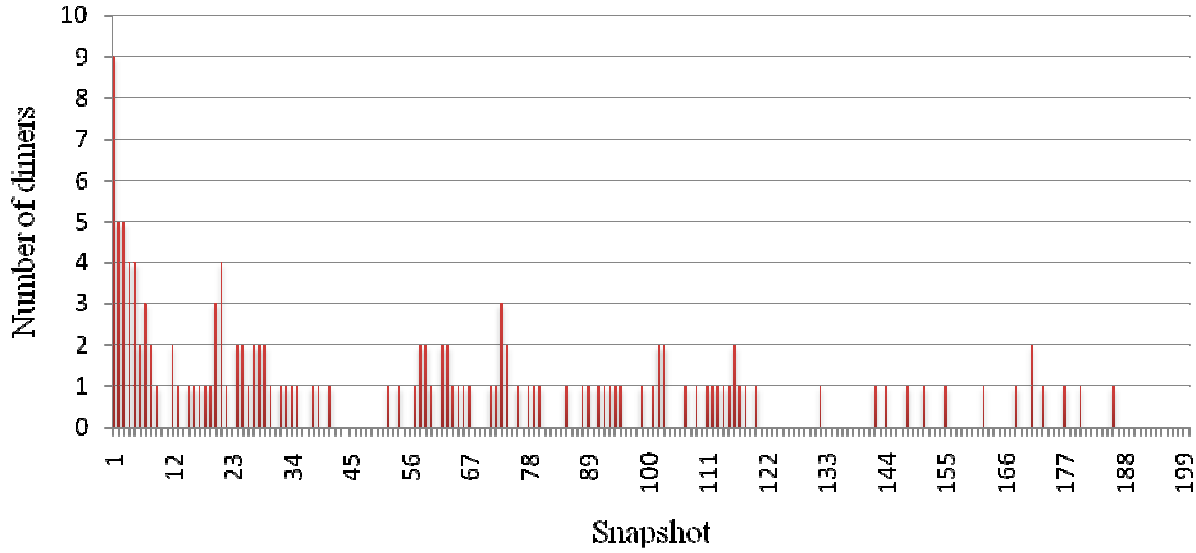


Figure 7.17 Distance between the centers of mass of two anthracene molecules (in 16in60Å)

From all these simulations of different number of anthracene molecules in a box, it can be seen that the probability of dimer formation increases with increasing number of molecules.

7.4 Conclusions

We have studied the functionalized anthracene and pentacene by using semi-empirical method implemented with dispersion correction term. We have chosen functionalized anthracene and pentacene with different sizes of bulky groups attached to the middle of the anthracene and pentacene molecules. For functionalized pentacene, we have chosen three crystal structures of substituted pentacene derivaties. We have explored the importance of dispersion contribution and the effect steric hindrance in π-π stacking of functionalized anthracene and pentacene molecules. We have employed only PM3-D method to study these interactions, since PM3-D method has been shown to be successful method to study the noncovalent interactions. We have studied the dimer models of functionalized anthracenes. We find that the anthracene molecules without any substituents interact through more π-π stacking interaction by interacting through all three rings of anthracene. We find that the functionalized anthracene molecule **4** with bulky isopropyl groups interact through only one ring of anthracene. This can be attributed to the steric effect of isopropyl group which prevent the anthracene molecules to disperse. We have also studied the three functionalized pentacene crystal structures and pentacene by choosing models of dimer, trimer and tetramer to explore the steric effect. It has been found that with increasing molecules, i.e., with increasing bulky groups, steric effects increase and stacking interactions decrease. In other words, these crystal structures balance stacking and steric effects in order to maintain compact crystal packing. Dispersion plays an important role in stacking when the number of molecules is less. Steric effect overcomes the dispersion effect with increasing number of molecules in stacking. This same Conclusions has been drawn for anthracene and pentacene molecules. We have also studied different number of anthracene molecules in toluene solution by using molecular dynamics methods. We have built four systems of anthracene molecules in 60Å box of toluene solution. We have heated the four systems to 300K and equilibrated for 2ns. Equilibrations have been achieved successfully and the snapshots were collected from 2ns. We have calculated the distances between the anthracene molecules for all snapshots by taking the centers of mass of the molecules. For the system 2in60Å, the distance graph shows the formation of dimer only in the beginning of the simulation. For the system 16in60Å, the distance graph shows the

formation of dimer complexes throughout the simulation. Therefore, from the distance graphs, it is evident that with increasing number of molecules, the probability of forming dimer complexes increases.

7.5 References

1. Anthony, J. E. *Chem. Rev.* **2006**, *106*, 5028.
2. Sheats, J. R. *J. Mater. Res.* **2004**, *19*, 1974.
3. Gundlach, D. J.; Nichols, J. A.; Zhou, L.; Jackson, T. N. *Appl. Phys. Lett.* **2002**, *80*, 2925.
4. Müllen, K.; Wegner, G. *Electronic Materials: The Oligomer Approach*; Wiley-VCH: New York, **1998**.
5. Valdes-Aguilera, O.; Neckers, D. C. *Acc. Chem. Res.* **1989**, *22*, 171.
6. Kazmaier, P. M.; Hoffmann, R. *J. Am. Chem. Soc.* **1994**, *116*, 9684.
7. Bre´das, J.-L.; Calbers, J. P.; da Silva Filho, D. A.; Cornil, *J. Proc. Nat. Acad. Sci. U.S.A.* **2002**, *99*, 5804.
8. Cheng, Y. C.; Silbey, R. S.; da Silva Filho, D. A.; Calbert, J. P.; Cornil, J.; Bre´das, J. L. *J. Chem. Phys.* **2003**, *118*, 3764.
9. Kwon, O.; Coropceanu, V.; Gruhn, N. E.; Durivage, J. C.; Laquindanum, J. G.; Katz, H. E.; Cornil, J.; Bre´das, J. L. *J. Chem. Phys.* **2004**, *120*, 8186.
10. Curtis, M. D.; Cao, J.; Kampf, J. W. *J. Am. Chem. Soc.* **2004**, *126*, 4318.
11. Kim, Y.H.; Shin, D.C.; Kim, S.H.; Ko, C.H.; Yu, H.S.; Chae,Y.-S.; Kwon, S.K. *Adv. Mater.* **2001**, *13*, 1690.
12. Shi, J.; Tang, C. W. *Appl. Phys. Lett.* **2002**, *80*, 3201.
13. Duong, H. M.; Bendikov, M.; Steiger, D.; Zhang, Q.; Sonmez, G.;Yamada, J.; Wudl, F. *Org. Lett.* **2003**, *5*, 4433.
14. Xu, Q.; Duong, H. M.; Wudl, F.; Yang, Y. *Appl. Phys. Lett.* **2004**, *85*, 3357.
15. Landis, C. A.; Parkin, S. R.; Anthony, J. E. *Jpn. J. Appl. Phys.* **2005**,*44*, 3921.
16. Lloyd, M. T.; Mayer, A. C.; Tayi, A. S.; Bowen, A. M.; Kasen, T.G.; Herman, D. J.; Mourey, D. A.; Anthony, J. E.; Malliaras, G. G. *Org. Electron.* **2006**, *7*, 243.
17. Swartz, C. R.; Parkin, S. R.; Bullock, J. E.; Anthony, J. E.; Mayer, A. C.; Malliaras, G. G. *Org. Lett.* **2005**, *7*, 3163.
18. Anthony, J. E.; Swartz, C. R.; Landis, C. A.; Parkin, S. R. *Proc.SPIE* **2005**, *5940*, 2.

19. Jiang, J.; Kaafarani, B. R.; Neckers, D. C. *J. Org. Chem.* **2006**, *71*,2155.

20. Payne, M. M.; Delcamp, J. H.; Parkin, S. R.; Anthony, J. E. *Org.Lett.* **2004**, *6*, 1609.

21. Anthony, J. E.; Payne, M. M.; Parkin, S. R. Manuscript in preparation.

22. Sheraw, C. D.; Jackson, T. N.; Eaton, D. L.; Anthony, J. E. *Adv.Mater.* **2003**, *15*, 2009.

23. Martin, D. C.; Chen, J. H.; Yang, J. Y.; Drummy, L. F.; Kubel, C.*J. Polym. Sci., Part B: Polym. Phys.* **2005**, *43*, 1749.

24. Podeszwa, R.; Szalewicz, K., *Phys. Chem. Chem. Phys.*, **2008**, *10*, 2735.

Chapter 8

Summary

We have investigated the role of noncovalent interactions in carbohydrate-aromatic interactions, in functionalization of nanotube by nucleic acid bases, aromatic aminoacids and organic molecules, and in aggregation of functionalized anthracene and pentacene. We have studied the importance of noncovalent interactions using electronic structure methods in carbohydrate-aromatic complexes involving β-glucose and 3-methylindole, *p*-hydroxy toluene as models of tryptophan and tyrosine respectively. We have used dispersion corrected density functional theory (DFT-D) and semi-empirical methods (PM3-D and PM3-D*) to study the noncovalent interactions involving dispersion. We have validated the DFT-D method by comparison with high level ab initio calculations (MP2/cc-pVDZ and CCSD(T)/CBS limit). We have found that DFT-D perform best overall compared to semi-empirical methods. We have also found that the semi-empirical PM3-D method, which has been shown to be successful method to study molecules involving noncovalent interaction, overestimates the carbohydrate-aromatic interaction energy. The PM3-D* method, which is a modified version of PM3-D reparameterized specifically for carbohydrate-aromatic interactions, was found to give better results for carbohydrate-aromatic complexes. We have studied the structural and energetic aspects of the carbohydrate-aromatic complexes in detail. We have also analysed their vibrational frequencies and NMR chemical shift characteristics. We find that the stabilization of carbohydrate-aromatic complexes is dependent on the inclusion of dispersion correction energy for the DFT-D calculations. Thus, without dispersion contribution, complex would be unstable. We find that the carbohydrate binds stronger on 3-methylindole than on *p*-hydroxy toluene. We also find that the aromatic aminoacid bind on either side of the β-glucose. We find hydrophobic face of the carbohydrate mainly interact through CH/π interactions with aromatic aminoacids. The polar surface of the carbohydrate interacts through OH/π interactions with aromatic aminoacids. From the interaction energy values, we have found that the binding energy of CH/π bound complexes is approximately 6 kcal mol^{-1}, suggesting each CH/π interaction contributes

close to ~2 kcal mol^{-1}. The computed blue shifts of C-H stretch vibrational modes suggest the interaction through C-H bonds. The red shifts of O-H stretch vibrational modes suggest the interaction through O-H bonds. We also find the complexes involving OH….N hydrogen bonding interactions associated with O-H shifts. Calculated binding energies suggest that the individual CH/π interactions are stronger for tryptophan than for toluene which is reflected in the larger computed shifts in the C-H frequencies, while the OH/π interactions are comparable in both sets of complexes. This probably accounts for our finding that both OH/π and CH/π complexes have very similar energy for the tryptophan complexes. We have not found any OH/π interactions in the complexes of glucose and *p*-hydroxytoluene except OH-O hydrogen bonding in structures **9** and **10**. The proton NMR calculations also show the largest chemical shifts for the protons of those CH groups which have the greatest interaction with the aromatic system, as predicted by the shifts in the vibrational frequencies and which is also supported by the experimental NMR calculations.

We have then tested the performance of a number of DFT-D functionals including both standard and newly developed functionals for the description of carbohydrate-aromatic interactions. We have studied small models, such as water-3-methylindole and methane-3-methylindole involving OH/π and CH/π interactions respectively, as representative molecules for carbohydrate-aromatic interactions. We have computed interaction energies by using 18 different density functionals and the DFT-D and MP2 methods. We have also computed CCSD(T) interaction energy for a structure of glucose-*p*-hydroxy toluene. We have also evaluated the interaction energy for this complex by using 18 density functionals and MP2 method and DFT-D method. We have found that DFT-D give interaction energy with a MUE value of 0.15 kcal mol^{-1} with respect to CCSD(T) interaction energy. Therefore, we have chosen DFT-D energies as benchmark for testing other functionals. We have employed all these different density functionals to study small models of OH/π and CH/π interactions and 11 complexes of carbohydrate-aromatic interactions. The results of all these calculations suggest that the DFT-D method perform better than any other method. The standard density functionals like BLYP and B3LYP give very poor performance for studying these complexes. We find that BH&H perform better among hybrid GGA functionals. All other hybrid GGA

functionals such as B3LYP, B97-2, B98 and BH&HLYP give interaction energy with larger MUE value with respect to DFT-D method and are not suitable for studies involving noncovalent interactions. The meta-GGA functional VSXC shows very large overestimation of the interaction energy. The recently developed M0x functionals show better performance compared to other standard density functionals. Among M0x functionals except M05, all other functionals give interaction energy values within ~1.5 kcal mol^{-1}. The better performance of M06 and M06-2X suggest that these functionals can be used as an alternative to the DFT-D method as they reproduce the interaction energy values within 0.5-0.9 kcal mol^{-1} of the DFT-D values.

We have studied the role of noncovalent interactions in complexes of organic pollutants on graphene sheets and nanotubes using semi-empirical method implemented with dispersion correction term, PM3-D. We have studied the structural and energetic aspects of complexes of C96 and [6, 6] with a number of organic molecules with aromatic ring and different substituents in detail. The organic molecules chosen are as follows, 1,3-dichlorobenzene (1,3-DCB), 2,4-dichlorophenol (2,4-DCP), naphthalene (NAPH), 2-naphthol (2-NATH), 1-naphthylamine (1-NALA), phenol (PH), 2-phenylphenol (2-PHPH), 1-naphthol (1-NATH), catechol (CT), pyrogallol (PY). We found that the largest interaction energies are for the group of molecules involving naphthalene and its derivatives. The interaction energies of these complexes are greater than for the complexes involving single ring systems. The order of interaction energy for organic molecules 1, 3-DCB, 2, 4-DCP, NAPH is as follows 1,3-DCB<2,4-DCP<NAPH, which is in agreement with the experiment. We have noted that molecules with higher number of –OH substituents interact stronger on graphene sheet and nanotube. For the series of hydroxyl molecules, both the PM3-D and DFT-D calculations predict the affinity to be in the order PH(1-OH)<CT(2-OH)<PY(3-OH), in agreement with experiment. For all adsorbates we see that the variation in the interaction energy follows that of the dispersive interaction. We have used a simple Mulliken population analysis to investigate electron transfer between the adsorbate and the surfaces. We find a small degree of electron transfer, of the order of 10^{-3}e, from the adsorbate. However, there is no correlation between this quantity and the calculated interaction energy. Thus, these calculations do not support the suggestion that the degree of EDA is responsible for the

variation in the interaction energy of the different adsorbates. We have found that PM3-D method, which is some two orders of magnitude faster than DFT-D calculations, yield relative adsorption affinities for a range of aromatic hydrocarbons on both graphene and CNTs which are in good agreement with experiment. As with other studies, we found that the dispersive interactions are dominant and the complexes would be unstable without its contribution. We do not find any correlation between the interaction energy and the very small degree of charge transfer which our calculations predict. With increase in the number of –OH substitution and number of aromatic rings, we find increase in the interaction energy of these complexes.

We have investigated the importance of noncovalent interactions in functionalization of nanotubes by nucleic acid bases and aromatic amino acids. We have studied the complexes involving different sizes of graphene sheets and nanotubes with five nucleic acid bases and four aromatic aminoacids. We have employed semi-empirical methods PM3-D and PM3-D* methods to study these complexes. We have also employed DFT-D and M0x functionals for complexes of graphene sheets and nucleic acid bases. We have tested the performance of other methods with respect to DFT-D interaction values. We have studied the structure and energetic aspects of nucleic acid bases on graphene sheets and nanotubes in detail using PM3-D and PM3-D* methods. The size of carbon nanotube to represent realistic model restrict the use of computationally demanding expensive methods like ab initio methods and DFT-D methods to study these interactions on nanotubes. Thus, we have applied only semi-empirical methods to study these interactions on nanotubes. Similar to other studies, we found that the dispersive interactions are dominant in these complexes as well. We find that the order of binging energy of nucleic acid bases, G>A>T>C>U is same for all the graphene sheets and nanotubes, except on C24. We have also found that the interaction energy values are higher for [6, 6] nanotubes than for [5, 0] nanotubes. This can be attributed to the larger π surface area of [6, 6] nanotube which allows the molecules interact stronger than [5, 0] nanotube. We have shown that our semiempirical PM3 schemes predict both the structures and the interaction energies of the purine and pyrimidine bases with both planar π systems and carbon nanotubes which are close to those predicted by the more computationally demanding DFT-D and MP2 methods. With respect to DFT-D results,

the M06-2X member of the new M05 and M06 families of functionals yield interaction energies close to the DFT-D values for the C24 graphene sheet. However, there is a considerable increase in the error for the larger sheets, which can be attributed to the failure of these functionals to properly describe the longer range dispersive interactions which are predicted to be important by the DFT-D treatment. We have also studied the interaction of four aromatic amino acids His, Phe, Tyr and Trp on C96 and on [6, 6] nanotube. We have found that the interaction energies of His, Phe, Tyr and Trp with $C_{96}H_{24}$ and with a [6, 6] nanotube are of similar accuracy like interaction energy of nucleic acid bases. Thus, the PM3-D and PM3-D* parameterizations can be used as a rapid way of estimating these interactions to gain insights into functionalization of nanotubes, which are larger and restrict the use of computationally expensive methods.

We have investigated the noncovalent interactions in functionalized anthracene and pentacene molecules using semi-empirical method implemented with dispersion term. We find that the dispersion and steric effect are in balance and decide the arrangement of anthracene and pentacene molecules. We have also studied the dynamics of anthracene molecules in toluene solution using various number of anthracene molecules. We find that the formation of complex increases with the increasing number of anthracene molecules.