Role of Hydrogen Bonding Interactions in Biologically Important Macromolecules

By

Swati Panigrahi

Enrolment No. LIFE05200704002

Saha Institute of Nuclear Physics, Kolkata

A thesis submitted to the Board of Studies in Life Sciences

In partial fulfillment of requirements For the Degree of

DOCTOR OF PHILOSOPHY

of

HOMI BHABHA NATIONAL INSTITUTE

May, 2012

Homi Bhabha National Institute

Recommendations of the Viva Voce Board

As members of the Viva Voce Board, we certify that we have read the dissertation prepared by Swati Panigrahi entitled "Role of hydrogen bonding interactions in biologically important macromolecules" and recommend that it may be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

_____ **Date:**

Chairman – Prof. Dipak Dasgupta

_____ **Date:**

Guide / Convener – Prof. Dhananjay Bhattacharyya

_____ **Date:**

Member 1 – Prof. Rahul Banerjee

_____ **Date:**

Member 2 – Prof. Sangam Banerjee

_____ **Date:**

Member 3 – Prof. Samita Basu

_____ **Date:**

External Examiner – Prof. Manju Bansal

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to HBNI. I hereby certify that I have read this dissertation prepared under my direction and recommend that it may be accepted as fulfilling the dissertation requirement.

Date:

Place:

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Swati Panigrahi

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.

Swati Panigrahi

Dedicated to

My Beloved Parents

ACKNOWLEDGEMENTS

I express my deep sense of respect and sincere gratitude to my thesis supervisor Prof. Dhananjay Bhattacharyya for his continuous support in my study and research and also for his emotional support in my hard times. He has been a constant source of inspiration for me during the course of my thesis work. He always encouraged me with his immense patience, motivation, enthusiasm and knowledge. I also thank Director, Saha Institute of Nuclear Physics for providing me such a good place to work. I am also extremely grateful to Professor Dipak Dasgupta, the Head of Biophysics Division, for providing me all the facility to accomplish my research works. I also thank to Prof. Sangam Banerjee, Prof. Prabal Kumal Maiti, Prof. A. K. Sood, Prof. Abhay Sankar Chakrabarty and Prof. Slawomir J. Grabowski, for their valuable suggestions and discussions. I am also thankful to my seniors, Dr. Ashim Roy, Dr. Malyasri Bhattacharyya, Dr. Debashree Bandyopadhyaya and my collaborators Rahul Pal, Rajat Pal, M. Santosh, A. Bhattacharya. I also take this opportunity to thank my labmates Sanchita, Sukanya, Manas, Angana, Sangeeta di and my friend Soumini for their constant support and motivation. I am also thankful to all the members of Biophysics Division for helping me whenever needed. I am really grateful to my parents for being supporting me all the time; I also thank my twin sisters Dipa and Puja and brother-in-law Vicky for with me always. I am also thankful to my mother-in-law, especially to my late father-in-law, without their blessings I could not be able to complete my research work successfully. Last but not the least I am extremely thankful to my husband Prasant for his endless encouragement and support. Without him I will not be able to be in a position where I am. Finally, I am grateful to God, for showing me the right path always.

Contents

Synopsis

Title of the proposed thesis:

"Role of Hydrogen Bonding Interactions in Biologically Important Macromolecules"

Submitted by Ms. Swati Panigrahi, Biophysics Division, Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata 700064

Background:

The structures of biologically important macromolecules like DNA, RNA, and proteins are largely determined by the non-covalent interactions. These interactions play crucial role in molecular self-assembly, folding, molecular recognition etc. Non-covalent interaction comprises contributions from hydrogen bond (arise between polar molecules), ionic interaction (arise between charged residues), hydrophobic interaction (between non-polar group) and dispersion interaction. Dispersion interaction arises as a major attractive interaction between non-polar molecules. Between the two interactions, covalent interactions between the atoms and van der Waals between the molecules, lies an intermediate known as hydrogen bonding interaction, which addresses different physical, chemical and biochemical processes[1]. Traditional H-bonds in biomolecules are formed between the polar molecular groups, O-H, N-H, which carry a hydrogen atom known as donors and between O/N atoms of another molecule or another part of same molecule, known as acceptors. In generally hydrogen bond is represented as D-H…A, where D-H is known as donor group, associated with strong/large dipole moment and A is known as

acceptor group, which has lone pair of anti-bonding orbital. Hydrogen bond can be classified into strong and weak depending on the strength of hydrogen bonds[1]. The weak hydrogen bonded systems involve C-H…O/N, C-H…C, C-H…π types of interactions, where the proton donor/acceptor are non-electronegative in nature[2].

Nucleic acid bases interact with each other through face to face "stacking interaction" and edge to edge "hydrogen bonding interaction". These two forces are considered to be the intrinsic factor for helix stability. The bases are held together by hydrogen bonds between the complementary bases A=T/A=U and G≡C types as proposed by Watson and Crick (WC), which maintains the helicoidal conformation of DNA, making it the most stable biomolecule in nature. While presence of –OH group in the RNA makes it more flexible, as this substitution may have important biological effects. RNA crystallography reveals that in addition to the usual Watson-Crick basepairing, bases of RNA can also undergo pairing through Hoogsteen edge/sugar edge.[3] Non-WC basepairs occurring within secondary structural blocks are the key determinant for the folding of RNA, RNA recognition by ligands, proteins, ions and antibodies and of immense importance.

Non-covalent interactions of various biomolecules (like DNA/RNA) with graphene and carbon nanotube (CNT) draw special attentions in recent years due to their unique chemical, mechanical and electronic properties. Both nano-graphene and CNT are used as carriers of different drugs and biomolecules. Different molecules can interact with the CNT and graphene sheet through hydrogen bonding interaction and dispersion interaction. Hydrogen bonding interaction also plays significant roles in stabilizing Gold

nano particle (GNP)-drug complexes, which can have several therapeutic applications in nano-medicines.

Computational chemistry is an important technique to evaluate the molecular geometries, rates of equilibrium and other physical and chemical properties of the molecules. The tools of computational chemistry are force fields (molecular mechanics), *ab initio*, semi empirical and density functional methods.[4] *Ab initio* method is based on approximate solution of Schrödinger equation. The important part of solving the Schrödinger equation is the Born-Oppenheimer approximation, where coupling between electron and nuclear motion is neglected. The movement of electrons is considered differently by different methods. Hartree-Fock method considers the average electron interaction. Density functional theory (DFT), which takes in account of the Kohn-Sham equation, is superior to this as it considers both the electron correlation and electron exchange in its integral. Different levels of DFT are available to deal with different types of molecules and properties. The HF and DFT methods also used some approximation in the form of basis sets, where molecular orbital are given in terms of set of known functions. Energy of a molecule can be accurately derived by all the methods, but it is the optimized geometry which matters depending on the method and basis set employed.

Objectives:

I have attempted to understand the following through different *ab-initio* quantum chemical procedures.

1. Structure and energy of non-canonical basepairs: comparison of various computational chemistry methods with crystallographic ensembles

2. Structure, stability and energetic of rarely occurring base pairs in nature: A quantum chemical approach

3. RNA non-canonical basepair database: A complete analysis of the energetic, structural features, dynamics of the possible basepairs in their optimization form

4. Binding of Quercetin with gold nano-particle

5. Wetting Property of the edges of monoatomic step on graphite

6. Interaction of nucleobases with wrinkled graphene surface

7. Unzipping of Small interfering RNA (siRNA) on single walled Carbon Nanotube and graphene

Work Abstract:

1. Structure and Energy of Non-canonical Basepairs: Comparison of Various Computational Chemistry Methods with Crystallographic Ensembles

Different types of non-canonical basepairs, in addition to the Watson-Crick ones, are observed quite frequently in RNA. Their importance in the three dimensional structure is not fully understood, but their various roles have been proposed by different groups. We have analyzed the energetics and geometry of 32 most frequently observed basepairs in the functional RNA crystal structures using different popular empirical, semi-empirical and *ab initio* quantum chemical methods and compared their optimized geometry with the crystal data. These basepairs are classified into three categories: polar, non-polar and sugar-mediated, depending on the types of atoms involved in hydrogen bonding. In case of polar basepairs, most of the methods give rise to optimized structures close to their initial geometry. The interaction energies also follow similar trends, with

the polar having more attractive interaction energies. Some of the C-H…O/N hydrogen bond mediated non-polar basepairs are also found to be significantly stable in terms of their interaction energy values. Few polar basepairs, having amino or carboxyl groups not hydrogen bonded to anything, such as G:G H:W C, show large flexibility. Most of the non-polar basepairs, except A:G s:sT and A:G w:sC, are found to be stable; indicating C-H…O/N interaction also plays a prominent role in stabilizing the basepairs. The sugar mediated basepairs show variability in their structures, due to the involvement of flexible ribose sugar. These presumably indicate that the most of the polar basepairs along with few non-polar ones act as seed for RNA folding while few may act as some conformational switch in the RNA.

2: Structure, stability and energetic of rarely occurring base pairs in nature: A quantum chemical approach

Many frequently occurring non-canonical basepairs play important role in three-dimensional RNA structure. However, there are some base pairs, which occur very rarely in the RNA crystal structures. Considering all such base pairs together, they appear as about 3.4% of total basepairing types in the non-redundant data set as obtained from recent database (http://www.saha.ac.in/biop/www/HD-RNAS.html). Our main aim is to study the role of these rare base pairs and to understand why these are so rarely seen in nature. We have modeled these base pairs using an in-house program NUCGEN, which uses base pair orientation parameters, such as propeller, buckle, etc. for generating the coordinates of a base pair. These parameters are related to another set of six parameters in helical sense, which can be applied identically to both the bases before the helical transformation (Propeller and Stretch equivalents) is applied to one of the bases. We have

modeled 37 rare base pairs and 14 sugar-mediated base pairs using this procedure. Base pairs without sugar-mediated interactions are categorized into 3 types, (i) with two polar H-bonds (ii) with one polar H-bond and another is C-H…O/N contact and (iii) both H-bonds are C-H…O/N mediated. Base pairs, with N-H…O/N types H-bond, are found to be very stable. In some cases, the interaction energy values of C-H…O/N mediated basepairs are found to be as high as around -4 Kcal/mol, associated with larger H-bond lengths around 2.4 Å. We have also carried out the NBO charge analysis and AIM analysis on the optimized geometry, to study the amount of charge transfer and electron density through the hydrogen bonded systems. The sugar-mediated basepairs are observed to be very unstable and their optimized geometry differ significantly from their expected structures, due to involvement of flexible ribose sugar and backbone torsion angle, which provides extra flexibility to the geometry of the basepairs.

3: RNA non-canonical database: A complete analysis of the energetic, structural features, dynamics of the possible basepairs in their optimization form

RNA crystallography reveals that in addition to the usual Watson-Crick basepairing, RNA has ample varieties of base pairing geometry. Among the observed RNA structures, 60% are held together by regular Watson-Crick type of basepairing, while rest are engaged in other types of base pairing edges known as non-Watson-Crick/non-canonical basepairs. All the purine and pyrimidine bases interact with each other in plane through three different edges. These are known as (i) Watson-Crick edge (ii) Hoogsteen edge (iii) sugar edge. We have collected a set of non-redundant PDB files from the HD-RNAS data set (http://www.saha.ac.in/biop/www/HD-RNAS.html) and ran

BPFIND and NUPARM on them, which give us possibility of different basepairs in RNA. Presently our database is available in our institute website, which is as follows, http://www.saha.ac.in/biop/www/db/local/BP/rnabasepair.html. The numbers of geometries that can occur between two specific edges are hyperlinked. From our data set one can find out frequency of each type of basepairs. We have also hyperlinked optimized geometry (optimized by B3LYP/6-31G** approach) of the best structure of the specific type, along with its hydrogen bond lengths, basepair parameters, interaction energy, associated glycosidic angles and C1'-C1' distances. The distribution of the base pair parameters as obtained from the crystal structures is hyperlinked with the duplex frequency. Some of the basepairs are observed as part of base triads also. So we have also calculated the frequency of the triads depending on the base pairing edge involved. We have hyperlinked representatives of each type of base triplets, associated with the duplex, along with their PDB ID and residue numbers. We have also given the detail information about the basepairs which are occurring rarely or with zero frequency in nature by modeling them with NUCGEN. Our data set also gives information about the protonated basepairs also, which are detected by BPFIND. The primary reason behind developing the database is to gather a complete knowledge about the possible basepairs, along with the base triplets, which plays crucial role in riboswitch and to understand their structure and flexibility.

4: Binding of Quercetin with gold nano-particle

Gold nano-particles can be used in chemical science, technologies and drug delivery applications as they are chemically inert in nature and have no side effects in biological systems. Quercetin is an important flavonoid possessing strong antioxidant property. It is

used in the treatment of several oxidative stress-associated diseases. Gold nanoparticles were synthesized by reducing chloroauric acid with trisodium citrate by our experimental collaborator.[5] The aim of the present study is to investigate the binding of quercetin with citrate capped gold nano-particles theoretically. Theoretically we have used GGA:PW91/TZ2P approach along with the zeroth order regular approximation of relativistic Hamiltonian (ZORA) using ADF package. We have taken a three gold with citrate-capped system along with one quercetin molecule, which acts as a proper miniature model for the theoretical calculation. The total interaction energy of the citrate capped gold nano-structure was obtained as -16.79 Kcal/mol along with two stable O-H…O bonds, which is sufficient to stabilize the complex system and comparable to the interaction. Our results support the experimental finding of formation of stable quercetin-gold nano complex. We have also calculated normal mode frequencies of the systems and compared those with observed IR spectroscopic results, which also proved that GNP normally remains capped by citrate ions.

5: Wetting Property of the edges of monoatomic step on graphite

An arbitrarily cut graphene mostly shows two distinct edges, the zig-zag (Trans-edge) and the arm-chair (Cis-edge). We have carried out quantum chemical calculations on model systems of fully reduced nano graphene to characterize the hydrogen bonding capacity of C-H bond associated with these two edges of graphene with water. Analysis of the optimized geometry gives a pair of H-bond like interactions between the C-H groups of both graphene edges with the nearest water molecule. Trans-graphene edge water system is found to be more stable by 2 Kcal/mol than that of the Cis-edge. Charge transfer analysis using NBO approach demonstrates that the energy of charge transfer

between the lone pair of electrons of the oxygen to the anti-bonding orbital of the closest C-H bond is more in case of Trans-graphene edge than that of the Cis-graphene edge. Similarly, the change in NBO charges of the carbon, hydrogen and oxygen of the complex and components in Trans-graphene edge is more significant. This implies that the Trans-graphene edge interacts with water more strongly than that of the cis-graphene edge. This was found to be in agreement with the Atomic force Microscope (AFM) experiments by our collaborators on highly oriented pyrolytic graphite (HOPG), which detected two distinct frictional properties at the edges on varying humid conditions, giving rise to different AFM signature.

6: Interaction of nucleobases with wrinkled graphene surface

Graphene-nucleobase interaction is gaining importance due to its possible therapeutic applications. We have carried out detailed quantum chemical calculations of complexes of nano-graphene sheet and the nucleobases of DNA and RNA using dispersion corrected density functional theory. Binding energies show a trend as observed earlier by different theoretical and experimental measurements.[6] However, in our present investigation the optimized structures of the complex as well as isolated graphene show significant curvature[7], similar features are also observed by our collaborators in Atomic Force Microscopic studies. Analysis of NBO charges indicates possibility of weak hydrogen bond like interactions involving pyramidal amino groups of the nucleobases and π-center of the nano-graphene.

7: Small interfering RNA (siRNA) wrapping on single walled Carbon Nanotube and graphene

Recently it has been found that siRNA is quite important in different therapeutic applications. Molecular dynamics simulations of siRNA along with carbon nanotube (CNT) as well as of CNT-DNA complex for same sequence, has been carried out to understand how the CNT molecules can help in intracellular delivery. In order to understand the reason behind the differential binding of RNA and DNA to CNT, we have carried out *ab initio* quantum chemical calculations using DFT-D method with ωD97XD/6-31G** basis set. The interaction energy of uridine with CNT is found to be around -18.72 Kcal/mol where as in thymidine, it is obtained as -16.25 Kcal/mol. We have also carried out charge transfer analysis of the systems and we observed that uridine undergoes maximum charge transfer with the terminal carbon atoms of the CNT forming three stable hydrogen bonds, however we did not observe any such effect in case of thymidine complex. So this analysis gave us a proper explanation that RNA has stronger binding energy with CNT than that of DNA, which in turn initiates the wrapping of the siRNA on the CNT, which are in agreement with the MD simulation studies by our collaborators. We have also studied interactions of components/fragments of si-RNA with the graphene sheet by modeling miniature models of graphene and nucleosides complexes and optimizing them by DFT-D followed by frequency calculation. We observed that unpaired uracil residue makes strongest contacts with the graphene molecule through van der Waals and specific H-bonding interaction involving 2'-OH group of the ribose sugar. The thermodynamic analysis of the system also shows ΔG value of the graphene+uridine complex is more negative. So these interactions can be the

dragging force for the double helical si-RNA to unzip. Our Quantum chemical analyses are in agreement with that of MD simulation studies performed by our collaborators.

References:

1. Jeffrey, G. A., "*An introduction to hydrogen bonding"*. Oxford University Press New York (1997).

2. Desiraju, G. R.; Steiner, T., "*The weak hydrogen bond: in structural chemistry and biology"*, Oxford University Press, USA (2001).

3. Leontis, N. B.; Westhof, E., "Geometric nomenclature and classification of RNA base pairs". *RNA***,***7*, 499 (2001).

4. Jensen, F., "*Introduction to computational chemistry"*, Wiley (2007).

5. Pal, R.; Chakraborti, A. S. "*Preparation Of Gold Nanoparticle Quercetin Complexes By Citrate Reduction Method", Conference proceedings in ICANN,* 283-287 (2010).

6. Umadevi, D.; Sastry, G. N., "*Quantum Mechanical Study of Physisorption of Nucleobases on Carbon Materials: Graphene versus Carbon Nanotubes*", *J. Phys. Chem. Lett.***,** *2*, 1572-1576 (2011).

7. Meyer, J. C.; Geim, A.; Katsnelson, M.; Novoselov, K.; Booth, T.; Roth, S., The structure of suspended graphene sheets. *Nature***,** *446*, 60-63(2007).

List of publications

Publication in refereed journals:

1. A. Roy, ***S. Panigrahi***, M. Bhattacharyya and D. Bhattacharyya (2008) Structure, stability, and dynamics of canonical and noncanonical base pairs: quantum chemical studies, *J. Phys. Chem. B*, 112, 3786-96.

2. ***S. Panigrahi***, A. Bhattacharya, D. Bandyopadhyay, S. J. Grabowski, D. Bhattacharyya and S. Banerjee (2011) Wetting Property of the Edges of Monoatomic Step on Graphite: Frictional-Force Microscopy and *ab initio* Quantum Chemical Studies, *J. Phys. Chem. C*, 115, 14819-14826.

3. ***S. Panigrahi***, R. Pal and D. Bhattacharyya (2011) Structure and Energy of Non-canonical Basepairs: Comparison of Various Computational Chemistry Methods with Crystallographic Ensembles, *J. Biomol. Struct. Dyn.*, 29, 541-556.

4. ***S. Panigrahi***, A. Bhattacharya, S. Banerjee and D. Bhattacharyya (2012), Interaction of nucleobases with wrinkled graphene surface: Dispersion corrected DFT and AFM studies, *J. Phys. Chem. C*, 116, 4374–4379.

5. M. Santosh, ***S. Panigrahi***, D. Bhattacharyya, A. K. Sood and P. K. Maiti (2012), Unzipping and binding of siRNA with single walled Carbon Nanotube: a platform for siRNA delivery, *J. Chem. Phys*. 136, 047206.

6. M. Santosh, ***S. Panigrahi***, D. Bhattacharyya, A. K. Sood and P. K. Maiti (2012), siRNA Unzipping on Graphene, *J. Chem. Phys.*, 137, 054903.

7. R. Pal, ***S. Panigrahi***, D. Bhattacharyya, A. S. Chakraborti (2013) Characterization of citrate capped gold nanoparticle-quercetin complex: Experimental and quantum chemical approach. *J. Mol. Struct*., 1046, 153-163

Poster Presentations

1. "Comparison of B3LYP/6-31G** and HF/cc-pVDZ methods in structure and energetics of hydrogen bonded nucleotide basepairs", presented at "National Symposium on Quantum Chemistry, soft computing and Optimization", IACS, Kolkata, April 4-5, 2008.

2. "Variation of sugar pucker in RNA structures: Crystal database and quantum chemical studies", poster presented at "National Symposium on Cellular and Molecular Biophysics, CCMB, Hyderabad, January 22-24, 2009.

3. "Wetting Properties of Planar Nano Graphene: A AFM and Quantum Chemical Approach", poster presented at " International Symposium on Facets of Weak Interactions in Chemistry, at Saha Institute of Nuclear Physics, Kolkata, organized by Calcutta University, from 13-15 January, 2011.

4. "Binding of nucleobases with wrinkled graphene surface: Dispersion corrected DFT approach and AFM studies", poster presented at "2nd International Conference on Advanced Nanomaterials and Nanotechnology (ICANN-2011)", Dec 8-10, 2011 at Indian Institute Of Technology, Guwahati.

5. "Study of Interactions of Nanomaterials with Nucleotides: A Quantum Chemical Approach", poster presented at " Recent Advances in Chemical and Physical Biology" at Saha Institute of Nuclear Physics, Kolkata, organized by Saha Institute of Nuclear Physics (SINP), India and The Mechanobiology Institute (MBI), NUS, Singapore, from 5-7 March, 2012.

Schools/Symposium attained

1. "International Conference on Physics Biology Interface", December 13-16, 2009, at Saha Institute of Nuclear Physics, Kolkata, India

2. One day satellite symposium on "Recent trends in Peptide research", on January 8, 2010 at Bose Institute Kolkata

3. Schools on Computational Techniques in Soft Matter 2010 (CTAM10), December 6-10, 2010, at S.N. Bose National Centre of Basic Sciences, Kolkata.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

3D	Three Dimensional
ADF	Amsterdam Density Functional
AFM	Atomic Force Microscope
AIM	Atom in Molecule
bp	Basepair
BCP	Bond Critical Point
BD	Bonding Orbital
BSSE	Basis Set Superposition Error
CCD	Dunning Correlation Consistent Basis Set
CI	Configuration Interaction
CNT	Carbon Nanotube
DFT	Density Functional Theory
DFT-D	Dispersion Corrected Density Functional
DNA	Deoxy ribonucleic Acid
GGA	Generalized Gradient Approximation
GNP	Gold Nano Particle
GNPQ	Gold Nano Particle Quercetin Complex
H-bond	Hydrogen Bond
HF	Hartree-Fock
IR	Infra Red

LP	Lone Pair
MD	Molecular Dynamics Simulation
MO	Molecular Orbital
MP2	Moller-Plesset Perturbation Theory to Second Order
NBO	Natural Bond Orbital
ncRNA	Non-coding RNA
NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank
QM	Quantum Mechanics
RMSD	Root Mean Square Deviation
RNA	Ribonucleic Acid
SCF	Self Consistent Field
siRNA	Small Interfering RNA
WC	Watson-Crick Basepairs
ZORA	Zeroth Order Regular Approximation Hamiltonian

Chapter 1

An Overview of Fundamental Interactions Observed in Biomolecules and Nanomaterials

1.1 Brief Summary of Major Forces Stabilizing the Biomolecules

All the biological processes, in virus, bacteria, small insects to human are controlled by different interactions and interconversion of several biological macromolecules. These macromolecules are organic compounds and hence have many covalent bonds. Apart from covalent bonding interactions there are other forces also taking place in controlling the cellular machineries.

The monomeric subunits of different biomolecules and many of chemical compounds are formed by covalent bonds. Covalent bond is formed when two subsystems with unfilled electronic shells start to overlap. At that point electron density between them increases and a strong bond is formed. Covalent bond involves sharing of a pair of valence electrons by two atoms. In other words the electronic rearrangement of charges with an appreciable energy gain is known as covalent interactions. It is the origin of formation of many molecules. These are short range interactions and are highly directional in nature. The simplest example is the formation of hydrogen molecule. Hydrogen atom has one proton surrounded by one electron, two hydrogen atoms are attracted to each other in such a way that their individual electron clouds mix together forming a single cloud occupied by both the electrons with opposite spins, keeping the protons separated by well-defined distances. Enthalpies of covalent interactions are of the order of -5 eV/-115.3 kcal/mol (Maréchal, 2007).

In some of the molecules, two or three covalent bonds are formed between a pair of atoms, such as ethylene or acetylene. Some molecules also possess multiple double bonds such as benzene, where the double bonds are delocalized. However in case of supramolecular complexes, in addition to the covalent interaction, non-covalent

interactions come into play. Covalent interaction leads to the formation of classical molecule, while non-covalent interactions lead to the formation of molecular cluster. The realm of non-covalent interactions is extensive and covers a lot many of the chemical processes (Černý and Hobza, 2007). The structures of biologically important macromolecules like DNA, RNA, proteins etc. are also largely stabilized by the non-covalent interactions. They play crucial role in molecular self-assembly, folding, molecular recognition etc. Covalent interactions mainly belong to the field of chemistry, while non-covalent interactions play significant roles in biologically important macromolecules and are considered important by condensed matter physics also. Non-covalent interaction comprises contributions mainly from ionic interaction (arise between charged residues), hydrogen bond (arise between polar molecules), hydrophobic interaction (between non-polar groups) etc.

1.1.1 Ionic/electrostatic interaction

Charged atoms interact with each other through ionic interactions. Coulombic ionic forces are long range in nature and its magnitude varies as $1/R^2$, where R is the distance between the two atoms. Although the magnitudes of enthalpies of the ionic interactions are comparable to those of covalent interactions, they are barely directional. Electrostatic interactions are especially important for proteins, ion-pairs, and salt-bridges etc., which play important role in stabilizing the structures.

In general electrostatic potential between two molecules are calculated as sum of interactions between the pairs of point charges using the Coulomb's law, which define the potential energy as

$$V = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad \text{.....[1.1]}$$

Where N_A and N_B are the number of point charges in the molecules A and B, and r_{ij} is the distance between the point charges i and j.

Electrostatic interactions are the fundamental interactions which determine the structures, dynamics and functions of various biomolecules. They are more significant in protein folding, ligand-receptor interactions, molecular recognition etc. (Hunenberger et al., 2001).

1.1.2. van der Waals Interaction

When two uncharged molecules (such as noble gas) come in close contact to each other, their surrounding electron clouds influence each other, though neither of them have permanent charge/dipole moment. Random variations in the positions of the electrons around one nucleus create a transient opposite electric dipole moment in the other atom. This causes a weak attractive interaction between the two atoms by minimizing the electron-electron repulsion between them; such instantaneous dipoles-induced dipole-dipole interaction is known as London dispersion interaction (van Holde et al., 2006). These are of the order of 0.23 kcal/mol for small molecules. The magnitude of the attractive potential is dependent on the volume and the number of polarized electrons in each interacting group. The attractive potential is short range in nature. However, the two atoms cannot come very close to each other to minimize the potential energy as at very short distances repulsive force come into play, which is combination of repulsion of the electron clouds of the two atoms in addition to nuclear-nuclear repulsion. These two forces together forbid the two atoms to come very close to each other. Both the attractive

short range London dispersion and repulsive potential give rise to the complete van der Waals or Lennard-Jones potential, this has the following form

$$V(r)= 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \qquad \text{.....[1.2]}$$

The first term in the above equation represents the repulsive term and the second term presents the attractive term. Collision diameter and wall-depth are denoted by σ and ε respectively. These two opposing energies, attractive and repulsive are equal at equilibrium distance, which is known as van der Waals radii. The typical van der Waals energy curve is given in Figure 1.1. The graph shows that the interaction energy is zero at infinite distance and it decreases when separation decreases, passed through a minimum, then again energy increases on decreasing the distance further. The attractive force is long range order, whereas the repulsive force is short range in nature.

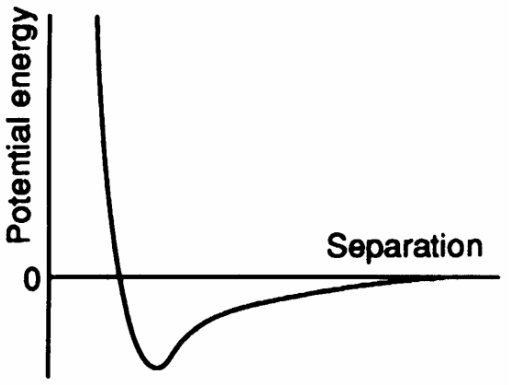


Figure 1.1. A typical van der Waals Energy curve

There is also other form of potential known as Buckingham potential which describes the repulsive exchange repulsion that originates from the Pauli Exclusion Principle by a more realistic exponential function of distance. The van der Waals interactions between atoms

and molecules play important roles in many chemical systems. They are in detailed balance with electrostatics and exchange repulsion interaction. They play important role in structures of DNA and proteins, packing of crystals, formation of aggregates etc.

Dispersion interaction arises as a major attractive interaction between non-polar molecules (Paton and Goodman, 2009). Their importance in the unsaturated organic groups is more significant, and termed as π- π stacking (Pyykko, 1997). Dispersive forces between stacked DNA basepairs and stacked amino acids are large, and have substantial contribute to the stabilizing energy. This is the only attractive force between noble gas atoms and it comes from London dispersion energy, which was theoretically derived on the basis of quantum mechanics. The dispersion energy is proportional to the product of subsystem polarization and the sixth power of reciprocal of the distances. Recent studies (Müller-Dethlefs and Hobza, 2000) shows that dispersion energy between aromatic systems with delocalized electron is comparable to that of hydrogen bonded complex.

1.1.3. Hydrogen Bonding

1.1.3.1. Historical Origin

Between the two interactions, covalent interaction between the atoms and van der Waals interaction between the atoms and molecules exist an intermediate known as hydrogen bonding interaction, which addresses different physical, chemical and biochemical processes. It is a topic of interest since a long time. Many eminent scientists postulated many theories about the definition of hydrogen bond (H-bond). Pauling had proposed that “Under certain conditions hydrogen atoms are attracted by rather strong

forces to two atoms, instead of only one, so that it may be considered to be acting as a bond between them, this is called hydrogen bond". Also he added that "A hydrogen bond is formed only between the most electronegative atoms and the second and weaker connections of H-atom are mostly electrostatic in nature of its interactions". Even though historically, electrostatics interactions are considered as the origin of the H-bond, depending on the strengths of the interactions, other contributions are also assumed to be present. According to Desiraju and Steiner (Desiraju and Steiner, 2001), the H-bonds are electrostatic in nature, but this character can vary. They define hydrogen bond as a complex interaction of electrostatic, polarization, van der Waals and covalent forces. A simple electrostatic model can account for a weak H-bond; whereas participation of covalent bonds is required for explanation of strong H-bond. So H-bond is formed in the extreme region of electrostatic and covalent bonds and it is the topic of interest in my research area. According to Pimentel and McClellan (Pimentel and McClellan, 1960) "A hydrogen bond is said to exist when (i) there is evidence of a bond and (ii) there is an evidence that this bond sterically involves hydrogen atoms already bonded to another atom". However, this theory remains silent about the nature of donors and acceptors.

There are different methods to detect hydrogen bonding experimentally. IR spectroscopy appeared in 1936 is found to be an efficient method to detect and observe hydrogen bonds. After 1950, discoveries of X-ray and neutron scattering established a fundamental property of hydrogen bonding; i.e. hydrogen bonds are directional in nature. After the prediction of alpha helix by Pauling and DNA structure by Watson and Crick, hydrogen bonding emerges as an important topic of research.

1.1.3.2. Geometry of Hydrogen Bond

Traditional H-bonds are formed between the polar molecular groups, O-H, N-H which carry a hydrogen atom known as donors and oxygen or nitrogen atoms of another molecule or another part of the molecule known as acceptors. The donors possess characteristics dipole moments and the acceptor is characterized by at least one non-bonding orbital or lone pair electrons that can point towards the hydrogen atom of the donor group. The donating molecular groups retain their identity upon formation of hydrogen bonds. In general hydrogen bond is represented as D-H…A, where D-H is known as donor group and A is known as acceptor group. The D-H and A are atoms of different molecules or part of the same molecules. The hydrogen bond of the type D-H…A can be described in terms of d, D, θ and φ respectively (Figure 1.2).

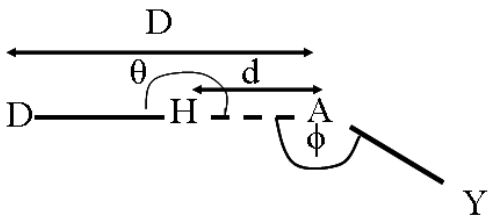


Figure 1.2. Schematic representation of hydrogen bond

Three out of the four parameters are independent. Where D denotes the distance between the heavy atoms, d is the hydrogen bond lengths H…A and θ denotes the D-H…A angle. The values of these parameters vary depending on the nature of the hydrogen bond. If the hydrogen bond is extended on the acceptor side, then the acceptor angle H…A-Y can also be defined as angle φ.

1.1.3.3. Nature of Hydrogen Bond

Hydrogen bond interaction can be described as contributions from electrostatic, polarization, exchange repulsion, charge transfer and dispersion interaction, among these exchange repulsion is repulsive in nature, whereas all the other interactions are attractive. Electrostatics, polarization, exchange repulsion and charge transfer vary with types of hydrogen bonds and represent its specific and directional property, while van der Waals interaction including dispersion signifies its non-directional property. Although electrostatic plays a dominant role among these interactions, in case of weak hydrogen bond interaction the contribution from electrostatic is found to be less significant, and dispersion interaction dominates (van Mourik and van Duijneveldt, 1995).

Hydrogen bond can be classified into strong and weak depending on the strength of hydrogen bonds. According to Jeffrey (Jeffrey, 1997) energy range for strong hydrogen bonds is 15-40 kcal/mol, for moderate hydrogen bond 4-15 kcal/mol, whereas for week H-bond 1-4 kcal/mol. So the upper limit of the energy for strong hydrogen bond is comparable to that of covalent bond, while the lower limit of the energy of the weak hydrogen bonds is comparable to the van der Waals interaction. The strength of the hydrogen bonds are also dependent on the geometry of the D-H…A orientation. In case of strong H-bond, the distance between H…A must be smaller than 2.5 Å, while the angles are close to 180^{o}. While large deviation from the above limit can be observed in case of weak hydrogen bonded system. So in case of weak hydrogen bonds, special care has to be taken when relating the geometry with the interaction energy.

Hydrogen bond can also be classified into conventional and non-conventional depending on the nature of hydrogen bond (Maréchal, 2007). Conventional hydrogen

bonds involve traditional donors such as N-H, O-H, which have appreciable dipole moment and the acceptors used to have lone pair of electron in its non-bonding orbital. Whereas, hydrogen bonds involving C-H group as the donors, which possess almost negligible dipole moment, are known as non-conventional type. In general most of the strong hydrogen bonds are conventional type, whereas most of the weak H-bonds are non-conventional type.

Hydrogen bond of the type D-H…A can also be classified as intermolecular and intramolecular H-bond. In case of intermolecular H-bond the D-H and A belong to two independent molecular groups, while in case of intramolecular H-bond, the donors D-H groups and acceptors A belong to same molecule. Intermolecular H-bonds are most commonly observed in nature, but intramolecular H-bonds also play important role in secondary structure formation of proteins or RNA.

1.1.3.4. Weak Hydrogen Bond

In case of strong hydrogen bond, the H-bond distance is smaller than the sum of van der Waals radii of the H and A atoms, as these are mostly electrostatic in nature. However, such criteria is not applicable for the weak or non-conventional hydrogen bonds (Grabowski, 2006). Such types of systems involve C-H…O/N, C-H…C, C-H…π etc., where the proton donors/acceptors are not strongly electronegative in nature. The results of spectroscopy and crystallography need not match well for the weak H-bond, as these are flexible in nature and may deform easily. The ability of the C-H group to act as proton donor also depends on the hybridization of carbon atom (Allerhand and Von Rague Schleyer, 1963) and also with that of adjacent electronic groups with the order C(sp) > C (sp^2) > C (sp^3). The H…O distance decreases significantly with increasing the

acidity of the C-H group. The strengths of the C-H bond donor capacity also vary, such as C-H group associated with methyl groups are weak hydrogen bond donors, while those associated with aromatic compounds are found to be stronger hydrogen bond donors.

1.1.3.5. Biological Significances of Weak Hydrogen Bond

The importance of weak hydrogen bonds in the biological macromolecules has been observed by various groups (Sutor, 1963). In biological molecules there are several C-H groups which can act as potential H-bond donors and many π-centers which can act as H-bond acceptors are present both on the surfaces and in the interior of the macromolecules and involve in the weak hydrogen bond formation. They play crucial role in molecular recognition and structural stabilization of molecular conformations. The side chains of different amino acids, which are the building blocks of the proteins show diversity in the possible types of hydrogen bond donors as discussed above. Amino acids like tyrosine, phenylalanine, tryptophan etc. also contain weak hydrogen bond acceptors. The C-H…O bonding patterns are frequently observed in parallel and antiparallel β-sheet and help in stabilizing those structures (Derewenda et al., 1995). In addition to the C-H…O/N type of weak hydrogen bonds, N/O-H…π hydrogen bonds are also observed in the protein and these mostly play dominant role in ligand-protein types of interactions, having energy ranges 2-4 kcal/mol. The involvement of weak hydrogen bonds in the enzymatic activity has been studied by various groups (Derewenda et al., 1994). The C-H…O hydrogen bonds in addition to the O-H…O hydrogen bonds are also frequently observed in carbohydrates, which are one of the most important biological macromolecules.

Stabilization to the entire non-covalent complex is due to favorable energy. However, these depend on the environment effects and mostly in water phase, some other forces also come into play, among them hydrophobic interactions play an important role.

1.1.4. Hydrophobic Interaction

Importance of hydrophobic effect has been widely studied in various areas of chemistry and biology. It is the most important intermolecular force of nature (Voet and Voet, 1994). The term **'**Hydrophobic effects' in general refers to the poor solubility of non-polar solutes in water, while 'hydrophobic interaction' refers to the association of two non-polar moieties in water (Southall et al., 2002). Two non-polar residues are driven to associate in water by the reduction in the surface area of solute-water contacts. X-ray experiments and structural studies by different groups confirmed that polar groups are localized at micellar surface and non-polar groups organize themselves at the interior of proteins (Harkins et al., 1946). Temperature dependency of the hydrophobic effect has been shown by various groups (Xu and Dill, 2005). Hydrophobic interaction plays significant role in the stability of globular conformation of protein (Tanford, 1962), and also in many physiological processes, such as protein folding, protein-protein binding, protein-DNA binding, and membrane and micelle formation, nucleic acid interactions etc. (Stojanovi and Zari 2009, Tanford, 1987). Different groups also proposed that hydrophobic interaction is not only temperature dependent, but also it depends on the size and shape of the solute (Southall et al., 2002).

1.2. Structural Hierarchy of Biomolecules and Importance of Hydrogen Bonding Interactions

1.2.1. An Overview of Nucleic Acids

All the genetic information of a living organism is encoded in deoxy Ribonucleic acid (DNA) in the higher organism living cells. It preserves and transmits genetic characters. DNA is the control center of each and every cell. It is present in each cell in the form of number of chromosomes. The entire assembly of the DNA containing chromosomes is known as nucleus. Transcription of DNA forms another class of biomolecules known as Ribonucleic acids (RNA), which usually carry the code for protein synthesis. The relation between DNA, RNA and protein can be presented by "central dogma" of molecular biology: DNA directs its own replication; it directs the formation of RNA by transcription, which in turn directs formation of protein by translation. Nucleotides form the building blocks of DNA and RNA, which consists of three distinct moieties, (i) a purine or pyrimidine base (ii) pentose sugar (de-oxy ribose in case of DNA and ribose sugar in case of RNA) (iii) monophosphate group. Polymerization occurs by the formation of phopsphodiester linkages between the nucleotides. Four nitrogenous bases are found in DNA, pyrimidines (Cytosine and thymine) are the six membered rings, while purines (Adenine and guanine) each of which is a fused system of five and six membered rings (Figure 1.3).

The hydrophilic sugar and the phosphate form the backbone, while the hydrophobic bases remain in the central region. The two chains of DNA double helical structure run in opposite directions. The chain terminates with a phosphate group linked

to the sugar carbon is denoted as C5'; whereas the chain terminated with free hydroxyl group linked to the sugar carbon is denoted as C3'. Both the chains run from C5' to C3' direction. The canonical Watson-Crick model is the most widely observed double helical right handed DNA, with ten nucleotides per unit turn, separated by 3.4 Å translational rise along the helix axis in each of the two chains and two constituents chains of the DNA run in mutually antiparallel directions.

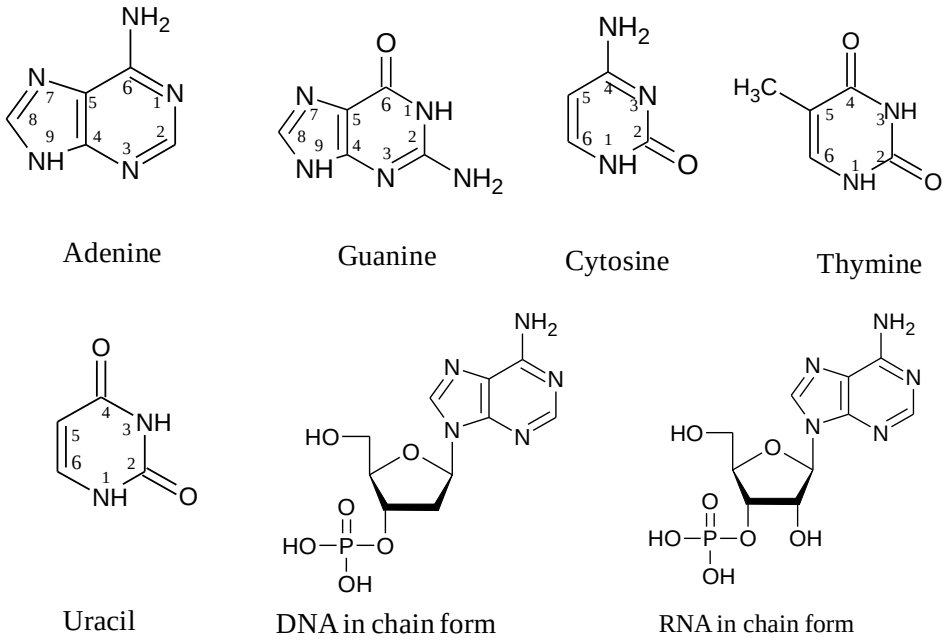


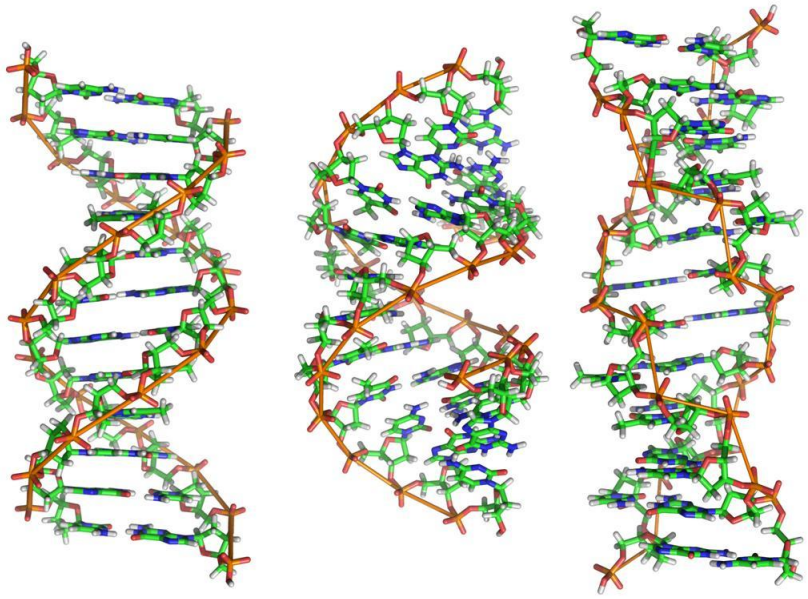
Figure 1.3. Schematic diagram of five major bases of nucleic acids and a representative of nucleic acid (adenine) in DNA and RNA in their chain form

The B-form DNA (Figure 1.4a) is found to be nearest to the original proposed Watson-Crick model with ten nucleotides per unit turn and is observed in high relative humidity. In addition to the most commonly observed B-form, DNA molecule also has many variants, which depends on the base sequence and environment and these different structural polymorphs of DNA have several biological significances. These different

variants of DNA are described excellently by Ghosh and Bansal (Ghosh and Bansal, 2003). The structural features of DNA have been experimentally observed by X-ray diffraction and several other spectroscopy techniques. A-DNA is the right handed double helical DNA (Figure 1.4b) observed under low humid conditions with 11 residues per unit turns. Another form known as Z-DNA is observed to be left handed double helical DNA with a dinucleotide repeat unit (Figure 1.4c). It has six dinucleotides per unit turns, with zig-zag backbone. Protonation at low pH also lead to formation of unusual DNA duplex, triplex and quadruplex geometries. Other forms of DNA such C-DNA, G-DNA, H-DNA, I-DNA etc. are observed under different conditions, which have been elaborately discussed by Ghosh and Bansal (Ghosh and Bansal, 2003).

When double helical structures are formed by stacking of the basepairs, the gap between the ribose sugars (attached with the bases) form continuous indentations on the surface, termed as groove. Through these grooves the basepairs are directly accessible to the external environments comprising of ions, ligands, drugs, proteins etc. These grooves are unequal; the larger one in B-DNA is the major and smaller is the minor groove. The C1'-N9 in case of purine and C1'-N1 in pyrimidine are on the minor groove side, while C6/N7 (purine) and C4 (Pyrimidine) base atoms are on the major groove side. Groove width can be defined as the perpendicular distance between phosphate groups on opposite strands minus the van der Waals diameter of a phosphate group, while groove depths are normally defined in terms of the differences in cylindrical polar radii between phosphorous and N2 guanine or N6 adenine atoms for minor and major groove respectively (Neidle, 2007). The B-DNA has wide major groove, which are rich in O6, N6 of purines and N4, O4 of pyrimidine. Most of the DNA-binding proteins prefer the

major grooves. The minor groove is found to be an important target for some regulatory and structural proteins, small ligands etc. However, in case of A-DNA one can observe narrow major groove and shallow minor groove due to its different structural arrangements. Presence of different donors, acceptors and hydrophobic groups alters the functionality of the groove.



(a) (b) (c)

Figure 1.4. Three main variants of DNA (a) B-DNA (b) A-DNA (c) Z-DNA

Polymeric RNA has same chemical form/structure as that of DNA, but has two modifications; (i) thymine is substituted by uracil base in RNA (ii) presence of one hydroxyl group in sugar of RNA. The hydroxyl group, which is very flexible in nature, plays an important role in many enzymatic reactions. RNA can adopt large number of helical and folded conformations, where such variations are much less in case of DNA. The nucleic acids are involved in several biochemical functions. Depending on the

chemical modifications, adenosine plays important roles in several biochemical processes, adenosine tri phosphate (ATP) is known as the energy currency of the cells in the entire living organism. It plays an important role in many enzymatic processes, mediates many activities of peptide hormones, also form a part of coenzyme A and NAD^+. RNA polymer also contains large number of modified bases; some of them have antibiotic activities. Along with ATP, guanosine triphosphate (GTP) is also very important in different biochemical reactions.

1.2.1.1. Torsion Angles and Sugar Pucker

We have to first understand the structures of the biomolecules in detail to recognize its different biochemical properties. Two terms known as torsion angle or dihedral angles are important in describing the three dimensional structure of a complex molecule. Torsion angle between the atoms A-B-C-D is defined as the angle between the planes containing the atoms A-B-C and B-C-D, and dihedral angle is defined as angles between normal to the planes containing the atoms A-B-C and B-C-D (Figure 1.5).

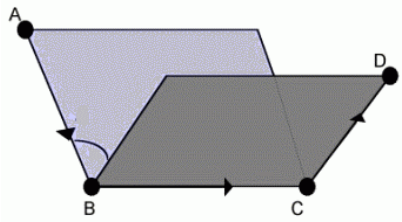


Figure 1.5. Schematic diagram of the torsion angle

Although torsion angles ranges between 0^o to 360^o or -180^o to $+180^o$, still all the conformations are not favorable, due to steric hindrance. The different conformations are

(i) Syn ~0^o, (2) anti ~ 180^o (3) synclinal $\pm 60^o$ (4) anti-clinal $\pm 120^o$, among them syn and anti-clinal are generally not allowed.

Conformation of the nucleotides is determined also by the sugar pucker and the orientation of the base and phosphate respectively. The poly nucleotide backbone has the sequences P-O5'-C5'-C4'-C3'-O3'-P (Figure 1.6). All the bonds are single covalent bonds, so nearly free rotations about the bonds are possible. These rotations are measured as the torsion angles defined as α, β, λ, δ, ε, ξ and the endocyclic torsion angles of sugar are defined as τ_0, τ_1, τ_2, τ_3, τ_4 depending on the rotation of the specific bond as given in the Table 1.1. Orientation of the base relative to the sugar is defined by χ. The base can adopt two distinct mode of orientations with respect to the sugar moiety about the glycosyl C1'-N link, (i) *Syn* (ii) *Anti*. In the slightly preferred *anti* conformation, the bulky part of the base is oriented away from the sugar O4', thereby avoiding the steric contacts. In purine, the six membered purine rings and in case of pyrimidine O2 is pointed away from the sugar and in the *syn* conformation they are oriented towards the sugar (Figure 1.7).

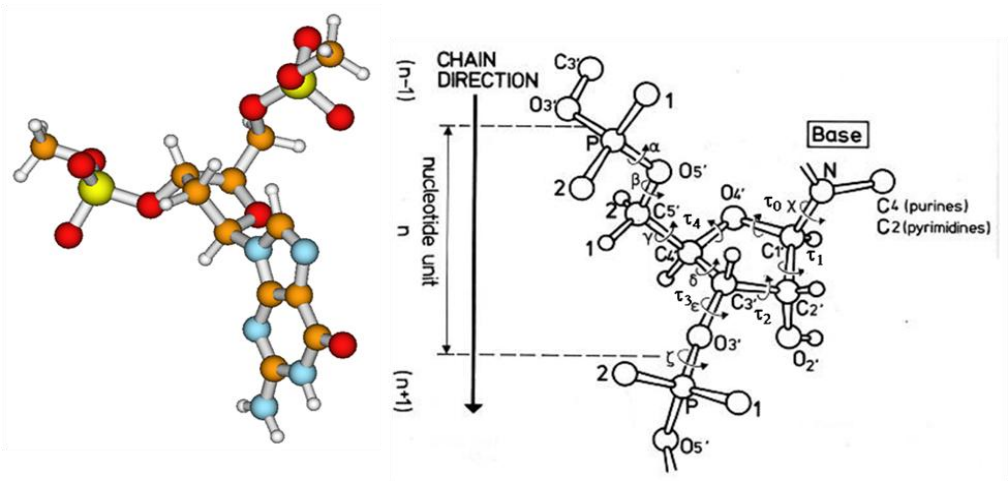
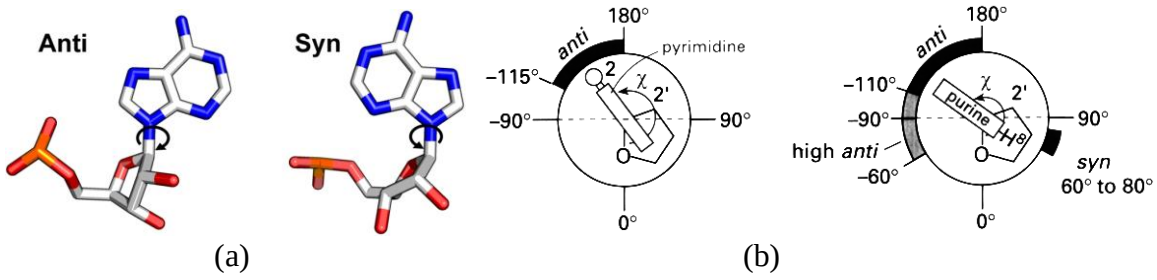


Figure 1.6. Definition of torsion angles for a polyribonucleotide chain



(a) (b)

Figure 1.7. (a) Schematic diagram shows the *anti* and *syn* orientation of the nucleoside and (b) *anti* and *syn* conformational ranges for glycosydic bonds in pyrimidine (left) and purine (right) nucleosides (Blackburn, 2006)

Table 1.1. Definition of Torsion angles in Nucleotides

Torsion Angle	Atoms involved
α	O3'-P-O5'-C5'
β	P-O5'-C5'-C4'
γ	O5'-C5'-C4'-C3'
δ	C5'-C4'-C3'-O3'
ε	C4'-C3'-O3'-P
ξ	C3'-O3'-P-O5'
χ	O4'-C1'-N1-C2 (Pyrimidines) O4'-C1'-N9-C4 (Purines)
τ_0	C4-O4'-C1'-C2'
τ_1	O4'-C1'-C2'-C3'
τ_2	C1'-C2'-C3'-C4'
τ_3	C2'-C3'-C4'-O4'
τ_4	C3'-C4'-O4'-C1'

The five member sugar rings observed in DNA and RNA are inherently non-planar in nature, this non-planarity is termed as sugar pucker. As the planar funarose ring is energetically unfavorable in general, all the torsions angles are zero, so all the carbon atoms are fully eclipsed in nature. The system tends to reduce its energy by puckering. When one of the ring atoms remain out of plane by 0.5 Å of the four, then the pucker type is known as envelop (E). In case of twist form (T), two adjacent atoms are displaced on opposite sides of the plane, whereas the other three atoms remain in plane. Atoms displaced in the same side of the C5' are designed as endo, whereas atoms displaced in the opposite sides of the C5' are in exo conformation.

In general, there is a continuum of interconvertible puckers which are produced by systematic changes in the ring torsional angles and are separated by energy barriers (Figure 1.8a). The puckers can be defined by two parameters P and τ_m, where P is the phase angle of rotation which is defined in terms of five torsion angles τ_0 to τ_m (Figure 1.8 b) is defined by the following equation (Neidle, 2007)

$$\tan P = \frac{(\tau_4 + \tau_1) - (\tau_3 + \tau_0)}{2\tau_2(\sin 36^o + \sin 72^o)} \qquad \text{.....[1.3]}$$

Whereas τ_m, the maximum degree of puckering is defined as

$$\tau_m = \frac{\tau_2}{\cos P} \qquad \text{.....[1.4]}$$

The value of P for furanose rings can be divided into 10 major classes of envelop puckering that lie around a full cycle of 360^o (Altona and Sundaralingam, 1972) as shown in Figure 1.8a. The interconversion of puckers depends on the barrier heights, which in turn depend on the route around the pseudorotation cycle. In case of nucleic

acids, two ranges of pseudorotations are generally observed (Figure 1.9), (i) C3'-Endo, where 0 ^{o}< P < 36^{o} (ii) C2'-Endo, where the 144^{o} < P < 190^{o}.

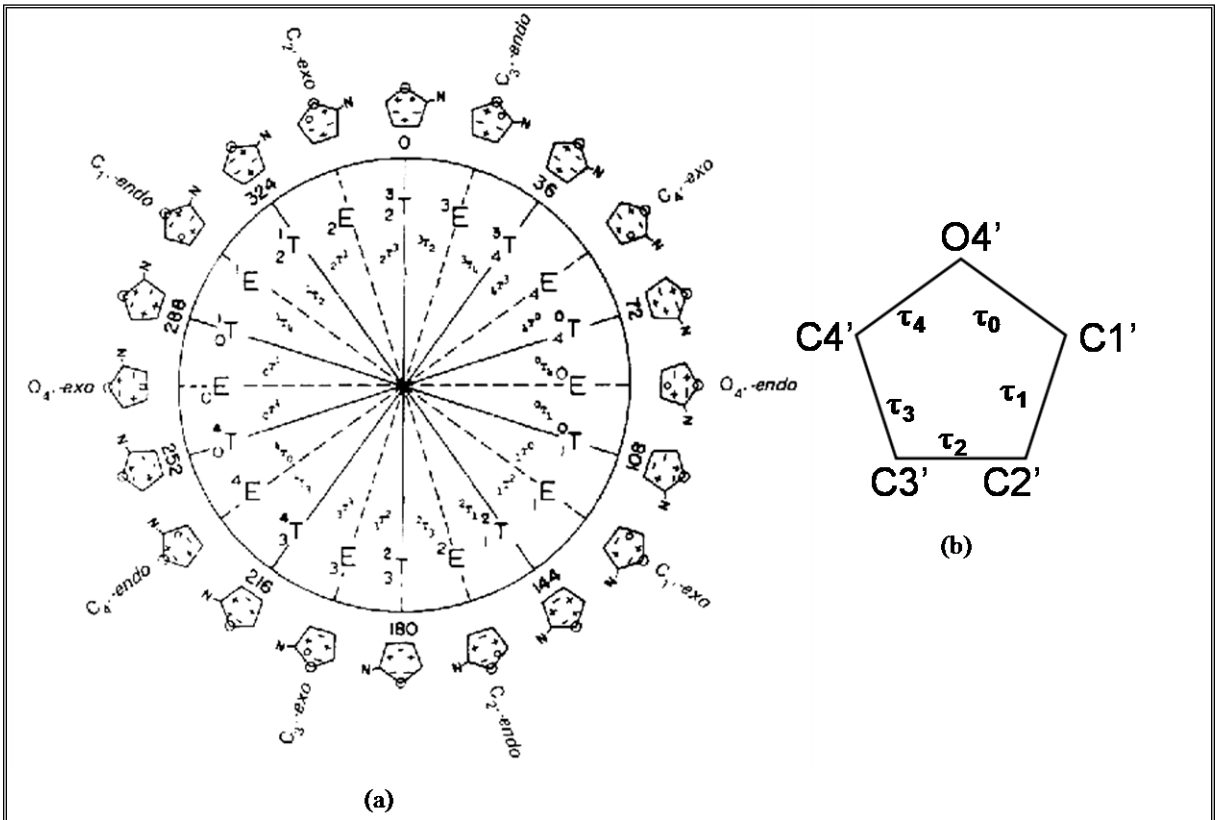


Figure 1.8. (a) Pseudorotation cycle of the furanose ring in the nucleosides (Saenger, 1984) (b) The five internal torsion angles in a ribose ring

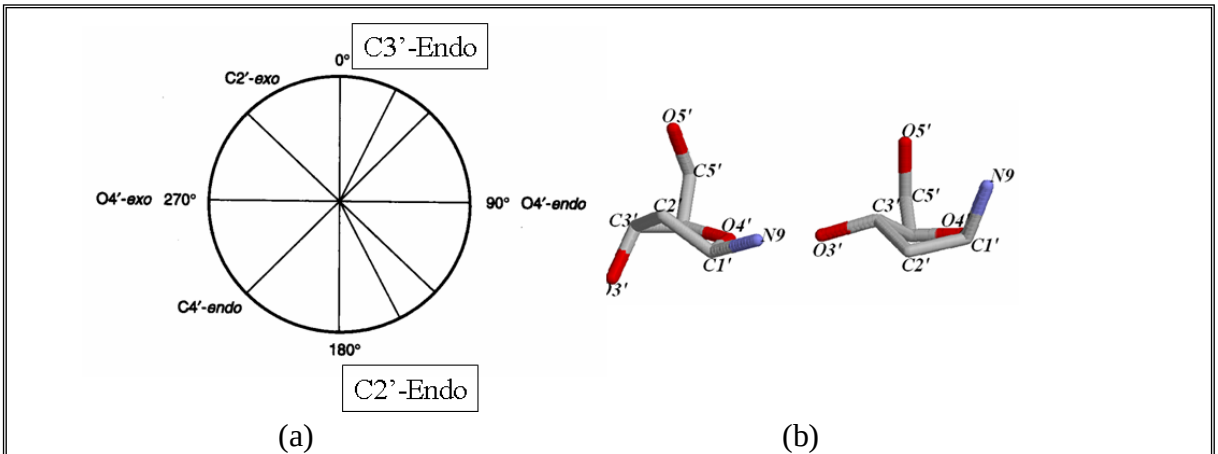


Figure 1.9. (a) Representation of pseudorotation phase angle P (b) Schematic diagram of the C2'-Endo (left) and C3'-Endo sugar pucker (right) conformations

1.2.1.2. Interactions Stabilizing Nucleic Acid Structures

The negatively charged phosphate groups, which are present at the exterior of the helix, are readily available for physical and chemical interactions with solvent water molecules and ions present in the cells. Nucleic acid bases interact with each other through face to face "stacking interaction" and edge to edge "hydrogen bonding interaction". These two forces are considered to be the intrinsic factors for helix stability. Among them, stacking interaction is more necessary for the overall helix stability. The bases interact through the edges in a very specific manner through formation of hydrogen bonds in most double helical forms, forming basepairs. Stacking interaction mainly holds the tertiary structure of nucleic acid and helps in folding mechanism, while hydrogen bonding provides directionality and specificity. The bases are held together by forming hydrogen bonds between the complementary bases. While adenine forms two hydrogen bonds with its complementary bases thymine or uracil, guanine forms three hydrogen bonds with its complementary base cytosine (Figure 1.10). Only these types of basepairing patterns are observed in DNA, commonly known as Watson-Crick/canonical basepairing. The stable hydrogen bond patterns formed by the basepairs are capable of maintaining the helicoidal conformation of DNA, making it the most stable biomolecule in nature. Presence of –OH group in the RNA makes it more flexible. This substitution may have important biological effects and make the life time of RNA limited, as there is possibility of hydrolysis by the –OH group. RNA crystallography reveals that in addition to the usual Watson-Crick basepairing, RNA has ample varieties of basepairing geometry. This is the main reason behind different structural motifs present in RNA, while they are not found in DNA.

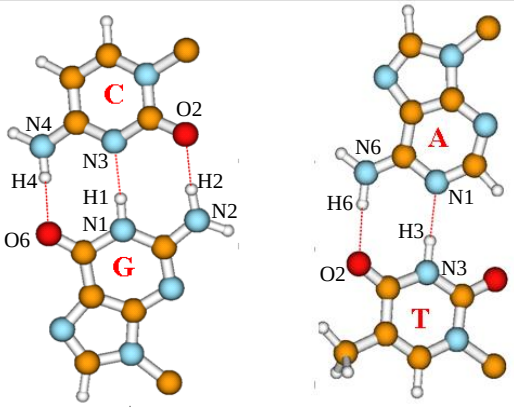


Figure 1.10. Hydrogen bonds observed between the complementary bases of DNA

Leontis and Westhof have done an excellent classification of RNA structures (Leontis and Westhof, 2001). They observed that among the total RNA structures, 60% are held together by regular Watson-Crick type of basepairing, while rest are engaged in other types of basepairing edges. All the other types of hydrogen bonding patterns except the regular Watson-Crick type give rise to non-canonical basepairs. All the purine and pyrimidine bases interact with each other in plane through three different edges. These are known as (i) Watson-Crick edge (ii) Hoogsteen edge (iii) sugar edge (Figure 1.11a). For each pair of edges, nucleotides can pair in two distinct ways, designated as cis and trans. In the cis configuration, glycosidic bonds of the nucleotides are oriented on the same side of the axis joining the base centres, while in trans orientation they are oriented in the opposite sides (Figure 1.11b). Leontis and Westhof have derived 12 distinct edge to edge interactions in RNA as given in Table 1.2**.** These classifications demonstrate the basepairing edges, along with the associated glycosidic angles and local strand orientations. The preferred *anti* configuration denotes the antiparallel orientation whereas the rare *syn* conformations denote the parallel orientation (Table 1.2).

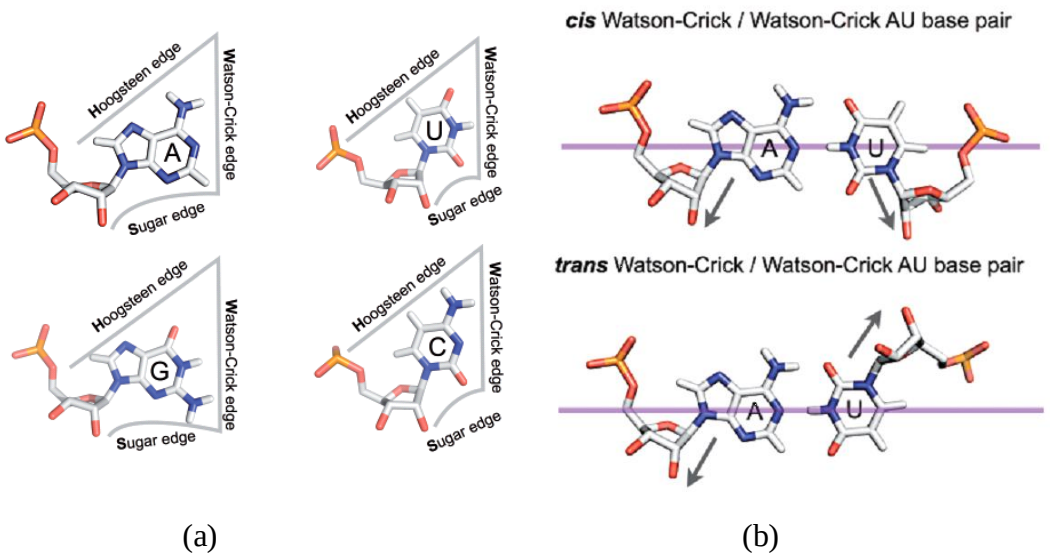


Figure 1.11. (a) Represents the three base pairing edges of the RNA nucleotides (b) Cis and trans orientation of glycosidic bonds (Almakarem et al., 2012).

1.2.1.3. Importance of Non-Watson Crick Basepairs

RNA molecules are key players in all steps of gene expressions. Except their functions as messenger RNA, they are also involved in several regulatory processes, play important roles in m-RNA splicing, transport and translations (Gesteland et al., 2006, Hermann and Westhof, 1999). Most cellular RNAs work in concert with protein partners either in permanent complex such as ribosome or in transient associations such as m-RNA splicing machinery. Non-Watson Crick (non WC) basepairs in RNA provide important sites for the specific interactions of RNA folds with proteins.

Table 1.2. The 12 main families of RNA basepairs between nucleic acids along with the glycosidic bond and local strand orientation (Leontis and Westhof, 2001)

No	Glycosidic bond Orientation	Interacting edges	Local standard Orientation
1	Cis	Watson-Crick/Watson-Crick	Antiparallel
2	Trans	Watson-Crick/Watson-Crick	Parallel
3	Cis	Watson-Crick/Hoogsteen	Parallel
4	Trans	Watson-Crick/Hoogsteen	Antiparallel
5	Cis	Watson-Crick/Sugar edge	Antiparallel
6	Trans	Watson-Crick/Sugar edge	Parallel
7	Cis	Hoogsteen/Hoogsteen	Antiparallel
8	Trans	Hoogsteen/Hoogsteen	Parallel
9	Cis	Hoogsteen/Sugar edge	Parallel
10	Trans	Hoogsteen/Sugar edge	Antiparallel
11	Cis	Sugar edge/Sugar edge	Antiparallel
12	Trans	Sugar edge/Sugar edge	Parallel

Regular A-form helices with WC basepairs are the basic building blocks of RNA architecture. RNA secondary structures involve Watson-Crick basepairing between complementary bases (A-U) and (G-C), also several non WC basepairs including the most common wobble G:U basepairs (Varani and McClain, 2000). The G:U wobble basepairs are the most common and highly conserved non-Watson-Crick basepairs in RNA, having biological importance in living organism, play important role in codon degeneracy (Das and Lyngdoh, 2012, Mangang and Lyngdoh, 2001, Voet and Voet, 1994). The thermodynamic stability of the G:U wobble basepairs is comparable to that of the Watson-Crick basepairs. They can provide unique recognition site also. In Watson-Crick basepairs the potential sites of hydrogen bonding are engaged in the basepair

interactions, so no sites are left for other specific interactions. In contrast in non WC basepairs, we have several potential sites available, which provide specific sites of hydrogen bonding. So non-WC basepairs are the key determinant for the folding of RNA, RNA recognition by ligands, proteins, ions and antibodies. The G:U and G:A are the most common types of non-Watson Crick basepairs found in large RNA. Non-WC interactions are also most commonly found in triplets (Mukherjee et al., 2006), where a third base forms hydrogen bond with the basepairing systems. Some kinds of non-WC basepairs are incorporated into stacked RNA stems without disrupting the helical structure, however in some cases the non WC basepairs increase the flexibility of the RNA backbone, widening the deep groove to accommodate the protein domains and regular secondary structural elements such as β turns and α helices (Hermann and Westhof, 1999).

1.2.1.4. RNA Diversity

There are mainly three major classes of RNA.

1. Ribosomal RNA (rRNA)
2. Transfer RNA (tRNA)
3. Messenger RNA (mRNA)

All of these different types of RNA are synthesized by the transcription process from DNA, each of which are associated with specific functions, tRNA molecules carry amino acids and deposit them in correct order, mRNA mediate translation of the hereditary information from DNA into protein, rRNA are involved in protein synthesis. The structural genes on DNA are transcribed into complementary strands of messenger RNA (mRNA), these are essentially a series of consecutive three nucleotide segments knows as

codons, each of which specifies a particular amino acid, by binding to tRNA on the ribosome.

1.2.1.5. Elements of RNA Structure

RNA structure is divided into three fundamental levels of organization.

(i) Primary
(ii) Secondary
(iii) Tertiary

Primary structure refers to the nucleotide sequence of RNA that is obtained to a first approximation from the DNA sequence of the gene encoding the RNA.

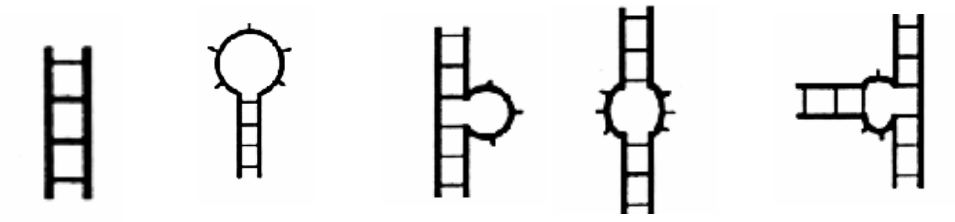


Figure 1.12. Secondary structural elements of RNA (Chastain and Tinoco, 1991)

The **secondary structure** of RNA is presented as a two dimensional representation of its Watson-Crick basepairs and unpaired regions. These folded RNA structures are imperfect due to non-complementary bases, thus form secondary structural elements such as duplex and loop regions that can be divided into six different types: duplexes, single-stranded regions, hairpins, internal loops or bubbles, bulge loops or

bulges and junctions. The various types of secondary structural motifs are presented in Figure 1.12.

Duplexes: RNA is a single stranded polynucleotide chain, however it can fold upon itself forming double stranded polynucleotide chain through formation of hydrogen bonds between the complementary bases.

Hairpin loops: It mainly connects the 3' and 5' ends of the double helix. The loop length can vary from 2 to 14 nucleotides (Figure 1.13). Tetraloops are the most common features of the hairpin loops. Among the tetraloops, GNRA (N can be A/G/C/U, R can be purines such as A/G), UNCG, ANYA (Y can be pyrimidine such as C/U) loops are characterized by their sequence and conserved structures. In ribosomal RNAs about 70% of tetraloops belong to either GNRA or the UNCG families and are observed to be thermodynamically stable structures. GNRA tetraloop (Figure 1.13) is more frequently observed in the available RNA structures (Jucker and Pardi, 1995). The secondary structure of tRNA is organized into 3 hairpin structures, D- , T- and anticodon arms and the acceptor stem, which is represented as the classic cloverleaf structure shown in Figure 1.14.

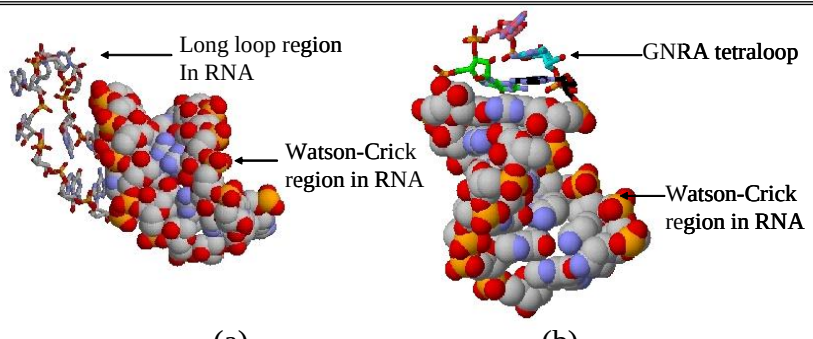


Figure 1.13. (a) Represents the long loop region (b) shows the presence of GNRA tetraloop along with the Watson-Crick region in RNA structure

Bulge: Bulge loops interrupt helices by insertion of non-canonical basepairs in one strand.

Internal Loops: Internal loops separate double helical RNA into two segments by inclusion of residues that are not Watson-Crick paired in at least one strand of the duplex. The non-canonical base pairing is common in internal loops. Internal loops can be of two types, symmetric, if same numbers of nucleotides are inserted on the both strands or asymmetric, if different numbers of nucleotides are inserted on the both strands.

Junction Loops: Junction loops are formed by the insertion of three or more double helices. Junction loops are observed in tRNA and in the hammerhead ribozyme.

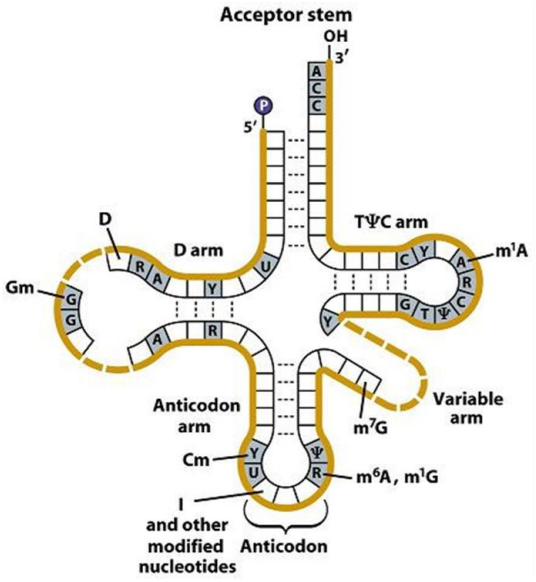
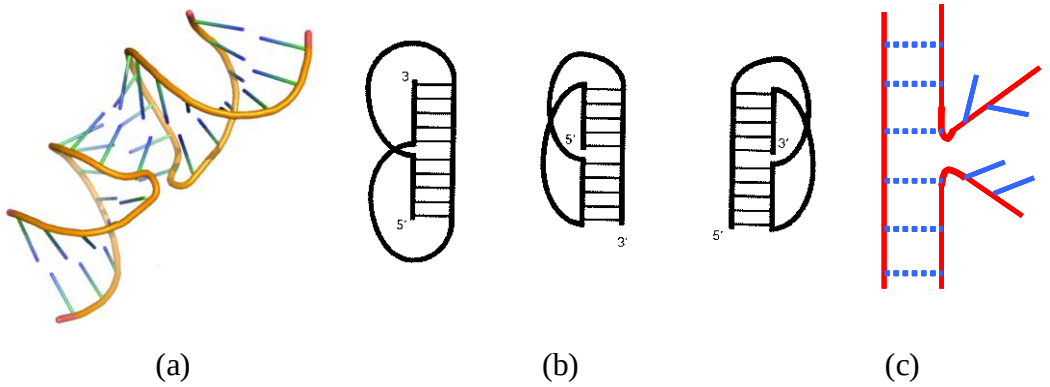


Figure 1.14. Secondary Clover leaf structure of tRNA with all its components

Other types of motifs like binding motifs such as metal ion binding motifs are also observed. Metal binding not only stabilize the RNA, but also perform many catalytic functions.

Tertiary structure elements in RNA involve interaction between distinct secondary structural elements. These involve coaxial helices, kissing hairpin loops, A-minor motif, pseudoknots, ribose zipper and so forth (Figure 1.15). Coaxial stacking of helical regions, the most fundamental method by which RNA achieves higher order organization, is a consequence of the highly favorable energetic contributions of stacking interactions between the π-electron systems of the nucleotide bases to the overall stability of nucleic acid structure (Saenger, 1984). Coaxial stacking can also be termed as junction loop. The adenosine platform (A-platform), which occurs in three separate locations in this RNA, consists of two sequential adenosine residues arranged side-by-side to create a pseudo base pair. A-platforms are observed in many other large RNAs. RNA pseudoknots have a stretch of nucleotides within a hairpin loop, that pairs with nucleotides, external to that loop. In other words pseudoknots are formed when all the bases in one helix are pseudoknotted to all the bases in the other helix. Pseudoknots help in predicting secondary structures of RNA. Conformational changes can be induced upon formation of these tertiary interactions.



(a) (b) (c)

Figure 1.15. Tertiary structural elements of RNA (a) kissing loop (b) pseudoknots (c) coaxial stacks

There are several other classes of RNA, which do not code for protein, but play various functional roles in living organism, these are commonly known as non-coding RNA (nc-RNA). Their size ranges from 20 to 10000 long nucleotides, they play essential role in regulatory functions and controls gene action. Other RNA with characteristics features also come into picture, such as micro RNA (miRNA) and small interfering RNA (siRNA), associated with unique biological processes. They play important roles in gene silencing process by binding to 3' untranslated region of mRNA. Due to their specific functions, siRNA can have important therapeutic applications.

1.2.1.6. RNA World

RNA world mainly refers to the important role of RNA in the origin of life. According to the RNA world hypothesis, RNA once both encoded genetic information and possessed the ability to replicate that information faithfully, as well as catalyzed the synthesis of essentially chemical building blocks (Joyce, 1989). Discovery of ribosome and its role in the peptidyl bond formation highlights the importance of RNA in the origin of life (Steitz and Moore, 2003). Discovery of a few classes of naturally occurring catalytic RNA, known as ribozyme, enhance the speculation about its important role as a biocatalyst. Some self-cleaving ribosome are observed, such as hepatitis delta virus ribozyme, hairpin, hammerhead ribozyme, which perform essential biological functions by accelerating the rate of site specific RNA strand cleavage. Aptamers are mainly DNA/RNA molecules that fold into higher ordered structures and form complexes with specific ligands. RNA plays important role as novel receptors and enzymes. Naturally occurring aptamers that bind to small molecules have been found as part of RNA genetic material control elements, known as riboswitch (Mandal and Breaker, 2004). Riboswitch

shows high affinity and selectivity in binding specific targets. Riboswitches control expression of the mRNA via two domains acting in concert: the aptamer domain and expression platform. The aptamer domain directly binds small molecule ligands and is responsive to intracellular ligand concentrations. Genetic regulation by riboswitches requires communication of ligand binding by the aptamer domain to the downstream expression platform. The three way junction elements in riboswitch, which form the ligand binding site, are defined by a series of non-canonical basepair interactions, as formation of unusual base triplets, which forms the 'roof' and 'block' elements of the ligand binding pocket. All these theories predict RNA to be the most important genetic material in prebiotic life (Gesteland et al., 2006).

1.2.1.7. Other Nucleic Acid Conformations

DNA is observed most commonly in its double helical form, while (Williamson, 1994) frequently it used to wrap around histones and packed as chromatin. Nucleic acids can form other complex structures such as triplex, quadruplex, Holliday junction etc. (Gilbert and Feigon, 1999), which play important roles in natural regulation and control. These modified structures may lead to invention of some novel therapeutics.

Triple Helix

Double helical DNA is common but, under certain circumstances, DNA can form triple helix (Figure 1.16a). A third DNA strand can bind into the major groove of a homopurine duplex DNA to form a DNA triple helix. The 1957 discovery of a three-strand DNA triple helix by two pyrimidine strands and one purine strand was determined when solutions of poly(A) and poly(U) were mixed in appropriate proportions, forming

1:1:1 three stranded polynucleotide complex. This discovery remained a curiosity till the recognition in 1987 when it was found that a third strand of DNA can actually recognize the base sequence of the double helix even without opening it (Felsenfeld et al., 1957, Le Doan et al., 1987). Owing to its enhanced stability, triple helical DNA can affect activities such as gene expression, DNA replication and others requiring DNA opening, and associated with therapeutic potentials (Jain et al., 2008). At ambient temperatures, the double helix is formed with classical Watson–Crick base pairing, while the third strand forms non-classical Hoogsteen or (reverse Hoogsteen) base pairing with one of the other two (Frank-Kamenetskii and Mirkin, 1995, Bloomfield et al., 2000). The triple helix can also be formed with DNA–RNA (Roberts and Crothers, 1992) and DNA–peptide nucleic acid (PNA), whose uncharged peptide backbone helps in the stabilization of the triplet structure (Betts et al., 1995). Studies (Frank-Kamenetskii and Mirkin, 1995) show that structure of triplexes may vary substantially as (i) triplexes may consist of two pyrimidine and one purine strand (YR*Y) or of two purine and one pyrimidine strand (YR*R) (ii) triplexes can be built from RNA or DNA chains or their combinations (iii) triplexes can be formed within a single polymer molecule (Intramolecular triplexes) or by different polynucleotides (Intermolecular triplexes) (iv) For special DNA sequences consisting of clustered purines and pyrimidines in the same strand, triplex formation may occur by a strand-switch mechanism (alternate strand triplexes). Examples of such triplex are CG*C, TA*T, CG*G etc. (Figure 1.16b,c) (Scaria et al., 1995). An important feature of the YR*Y triplexes is that formation of the CG*C triplex requires the protonation of the N3 of cytosine in the third strand. Thus, such triplexes are favorable under acidic conditions, termed as $CG*C^{+}$ type. To form such triplex, the third strand

must be located in the major groove of the double helix that is forming Hoogsteen hydrogen bonds with the purine strand of the duplex.

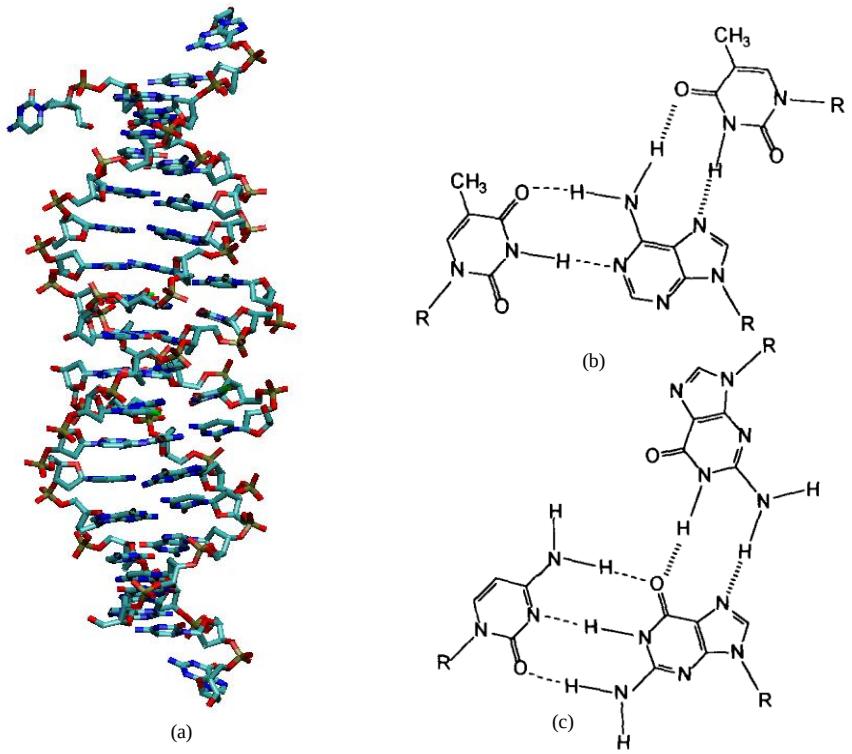


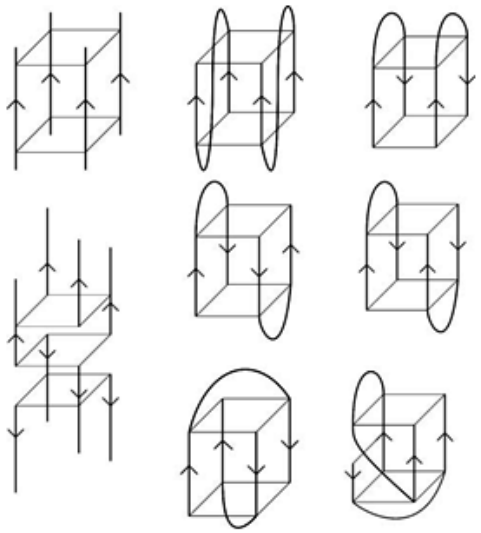
Figure 1.16. (a) An example of triple helical DNA (Rhee et al., 1999) (PDB ID 1D3R) (b) Triplex TA*T, involving Hoogsteen basepairing of thymine (Scaria et al., 1995) (c) Triplex CG*G involving Hoogsteen edge of guanine (Scaria et al., 1995)

Sequence specific triplex formation can be applied for various purposes in biotechnology like gene targeting, gene silencing and mutagenesis (Praseuth et al., 1999, Knauert and Glazer, 2001). There are several other structural motifs in RNA such as C-motif, A-motif which involve interaction of distant bases with basepairs leading to the formation of higher order structures. Several such examples are C-motif, A-motif where

there bases interact to form base triples (Holbrook, 2005). Triple helix also forms important components of riboswiches.

Quadruplex

Guanine quadruplex (G-quadruplex) is a four strand DNA composed of non-Watson Crick basepairing between stacked sets of four guanines (Sen and Gilbert, 1988). Each guanine tetrad is commonly known as G-quartet, which is formed of four guanosine bases, arranged in a square co-planar array, where each base can act both as donor and acceptor held together by hydrogen bonds involving Hoogsteen and regular Watson-Crick edges of the bases. The constituents guanine can be of a single nucleic acid chain (Intramolecular) or multiple strands (Intermolecular), and the strand can be oriented in parallel or antiparallel orientations (Figure 1.17). In the parallel quadruplex, all the glycosidic angles are oriented in *anti* conformation, whereas antiparallel quadruplexes, the glycosidic angles can have both the *syn* or *anti* conformations (Keniry, 2000). The negatively charged cavities located between the G-tetrads (formed by the guanine O_6 carbonyl groups) are stabilized by the inclusion of cations and the stability of the G-quadruplex structures is highly dependent on the size and charge of cation. G-quadruplex structures exhibit a remarkable dependency on the alkali cations such as Na^+ and K^+ for their formation and stabilization, hence physiological buffer conditions favor their formation. Among the ions, K^+ shows more potential for binding than that of the Na^+ due to better co-ordinations of the K^+ with the eight carbonyl oxygen atoms present in the adjacent stacked tetrads (Figure 1.18).

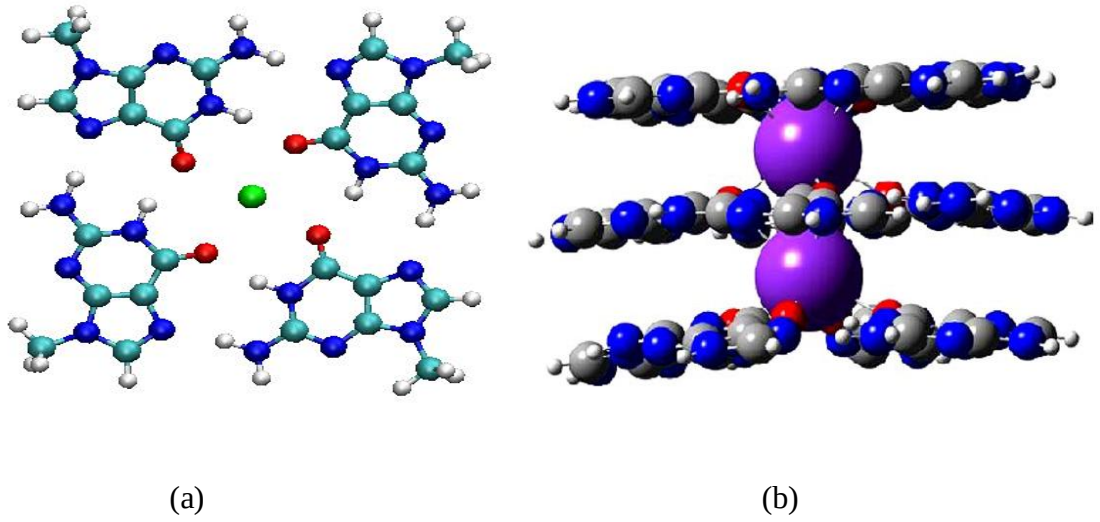


(a) (b) (c)

Figure 1.17. Some possible topologies for simple tetramolecular (a) and bimolecular quadruplexes (b and c). Strand polarities are shown by arrows (Burge et al., 2006).

Quadruplex are mainly stabilized by base stacking, hydrogen bonding, electrostatic and van der Waals interactions. Owing to their compact geometry, G-quadruplexes are found to be highly stable under physiological conditions, having melting temperature higher than that of the duplex B-DNA structure. Studies by Capra et. al (Capra et al., 2010) also proposed that quadruplex DNA sequences are evolutionary conserved and associated with distinct genomic features. These are most commonly found in the biologically significant region (Lipps and Rhodes, 2009) such as eukaryotic telomeres (Williamson, 1994). G-quadruplex based inhibitors of telomerase may be relevant to cancer therapy (Read et al., 1999). G-rich regions are also observed in other parts of the human genome, such as gene promoter regions, recombination sites, immunoglobulin switch regions and DNA tandem repeats. Therefore quadruplex formation may also be important for other diseases than cancer (Gilbert and Feigon, 1999). There have been number of

investigations carried out for these quadruplex structures by NMR, X-ray crystallography, Raman spectroscopy to understand its structural properties.



(a) (b)

Figure 1.18. (a) An example of G-quartet structure with potassium ion at the center (PDB ID: 156D) (b) Structure of the stacked G-quartets with potassium ions (Clay and Gould, 2005)

The remarkable polymorphism of nucleic acid quadruplexes is a result of subtle balance between several forces. The structure of a quadruplex is primarily determined by the condensation of guanine residues around a monovalent cation, while there are several other forces such as base stacking interactions, hydrogen bonding interactions, hydrophobic forces also play significant roles in stabilizing the quadruplex geometry. In addition to the G-quadruplex structures, an unusual stable Uracil quartets (U-quartets) is observed by parallel strands of r(UG_4U) through NMR spectroscopy (Cheong and Moore, 1992). These U-quartets are found to be stabilized by C-H…O contacts only, implying C-H…O interactions also play prominent roles in stabilizing the higher order structures (Figure 1.19).

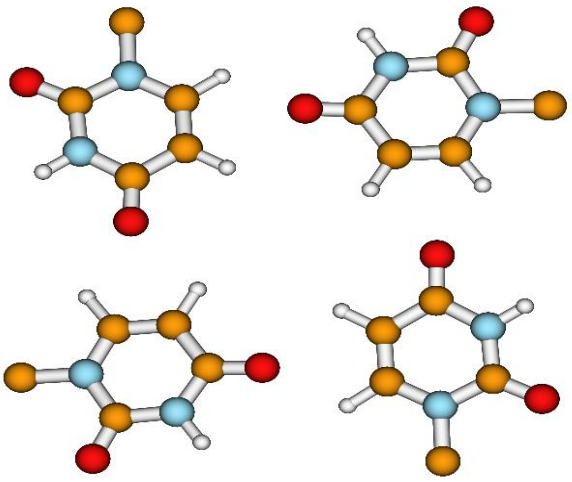


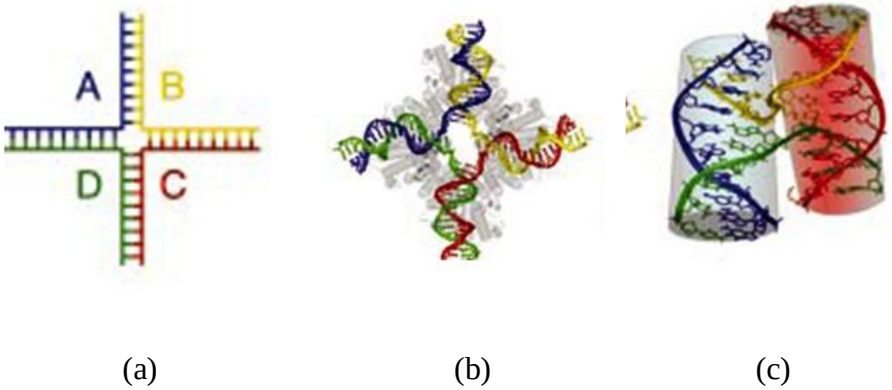
Figure 1.19. Initial structure of the uracil quartet with C-H…O interactions (PDB ID: 1rau)

Computational studies by different groups have been carried out to understand the interactions of guanine and uracil quartets with metal ions by different level of DFT functional. Study by Meyer et. al (Meyer et al., 2001) shown that G-quartets structures are found to be more stable than that of the U-quartets. All these structures have provided significant insights into the possible folding pathways of the nucleic acid structures.

Four Way Junction/Holliday Junction

DNA recombination is one of the fundamental biological processes required for genetic diversity and maintaining genomic integrity in living organism (Ortiz-Lombardía et al., 1999). Holliday junction, a four way DNA junction is the central intermediate in homologous recombination and is essential in maintaining genomic stability (Holliday, 1964). In addition, it also play important role in repair of and replication through DNA

lesions. Holliday junction also play important role in diagnosing certain disease conditions (Karow et al., 2000). Different groups using different experimental techniques explained the characteristics features of the Holliday junction (Ortiz-Lombardía et al., 1999, Broker, 1973). Two forms of Holliday junctions are in general observed (Figure 1.20). In the absence of cations that is in low salt concentration we observe open X-form, where the junction is believed to be extended, with its arms unstacked, and have four fold planar orientations. In the physiological conditions, in presence of metal ions, the arms stacked pairs and arrange in a two-fold non-planar junction, termed as stacked X-form. In the stacked X-form, the arms show nearly continuous B-DNA duplexes, with two G-A mismatches at the junction crossovers (Ortiz-Lombardía et al., 1999).



(a) (b) (c)

Figure 1.20. (a) Structure of Holliday junction (b) Open-X Holliday junction (c) Stacked-X Holliday junction (Watson et al., 2004)

1.2.1.8. Dynamics of the Basepairs

The architecture of nucleic acid basepairs can be described in terms of number of rotational and translational parameters, which has been designed in the Cambridge convention (Dickerson et al., 1989, Olson et al., 2001). These parameters are defined

from the spatial location of bases, where sugar phosphate backbone is not taken into account. The basepairs in nucleic acids are not really planar; the deviation from planarity can be described on the nature of bases and basepairs, and dependent on their stacking environment. The intra basepair parameters are calculated using the base edge specific axis system for the two bases forming the basepairs. Where the long axis of a basepair (y-axis) can be defined by the line joining the C6 of pyrimidine to the C8 of a purine or by an alternate line from C6 of a pyrimidine to a hypothetical C8* atom on the purine, chosen in such a way that C6-C8* vector is nearly parallel to the C1'-C1' vector (Figure 1.21a). The x-axis is directed towards the major groove along the pseudo two fold axes of an idealized Watson-Crick basepair. The z-direction is perpendicular to the plane of the pair. The intra basepair parameters can be described in terms of three rotational (Buckle, open angle and propeller) and three translational parameters (Stagger, shear and stretch) (Figure 1.21b). These intra basepair parameters indicate the planarity and proximity of the association between two bases. They have a direct resemblances to the three dimensional conformation of the basepairs. Propeller twist indicates twisting of the two bases along the long axis. Buckle is defined as the dihedral angle between the bases, along their short axes. It gives the value of amount of cusp formation. Open angle describes angle between the two bases in a basepair plane. Stagger indicates the out of plane motion of one base with respect to the other, shear defines sliding of one base with respect to the other in a basepair plane and stretch indicates separation of two bases relating to the hydrogen bond distances. Among these parameters open angle, shear and stretch relate to hydrogen bonding patterns and proximity, while buckle, propeller and stagger describe overall non-planarity of the basepairs (Figure 1.20b).

In addition to the intra basepair parameters, there are well defined inter basepair parameters, such as tilt, roll, twist, shift, slide, rise, which describe local conformation of a double helix at every basepair step. There are several methodology and software available, such as CURVES (Lavery and Sklenar, 1988), FREEHELIX (Dickerson and Chiu, 1997), CEHS (Lu et al., 1997) , 3DNA (Lu and Olson, 2003) , NUPARM (Mukherjee et al., 2006, Bansal et al., 1995) etc. for calculation of base and basepair morphology. NUPARM has been developed by our groups, and is used extensively in my calculations for analyzing the flexibility of the basepairs.

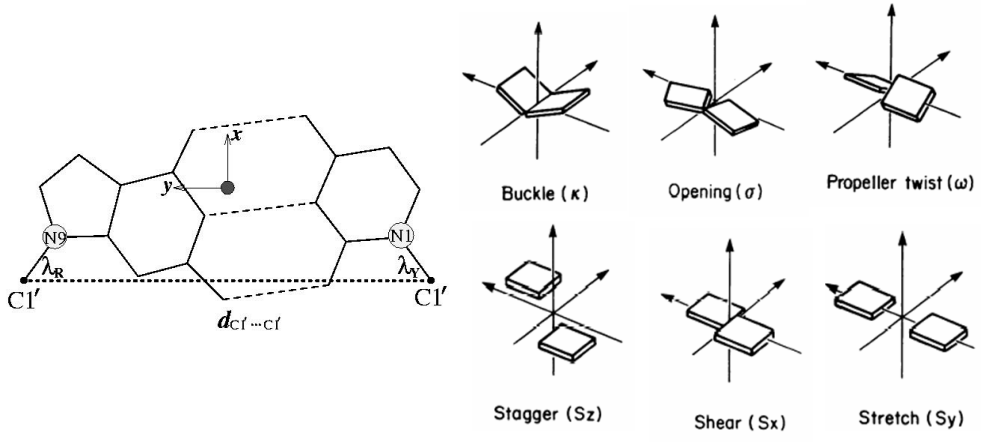


Figure 1.21. (a) Reference frame for idealized helical DNA, showing the axis origin at (●) (Clowney et al., 1996) (b) Schematic view of intra basepair parameters, the figures at the top row denote the rotational parameters, while the bottom row represents the translational parameters of a basepair (Dickerson et al., 1989)

1.2.2. An Overview of Proteins

Proteins form another classes of biomaterials. They play crucial role in performing biological functions, so also triggers many physiological processes. They are the basic components of many organs and are deeply involved in the bio-reactivity, such as enzymes are mostly made of proteins. They also provide fundamental services of storage and transport, also responsible for the structures and metabolism of many tissues, for example myoglobin and haemoglobin play important roles in transport and storage respectively. Amino acids linked together by peptide linkage are the building blocks of the protein. The protein backbone consists of hundreds of single covalent bonds and rotation around these bonds give rise to number of conformations, however every protein perform specific chemical and structural function. This suggests that each protein has a specific three dimensional structure, which in turns is determined by the nucleotide sequence of structural genes.

1.2.2.1. Structural Hierarchy in Proteins and Forces Need for Stabilization

Primary Structure

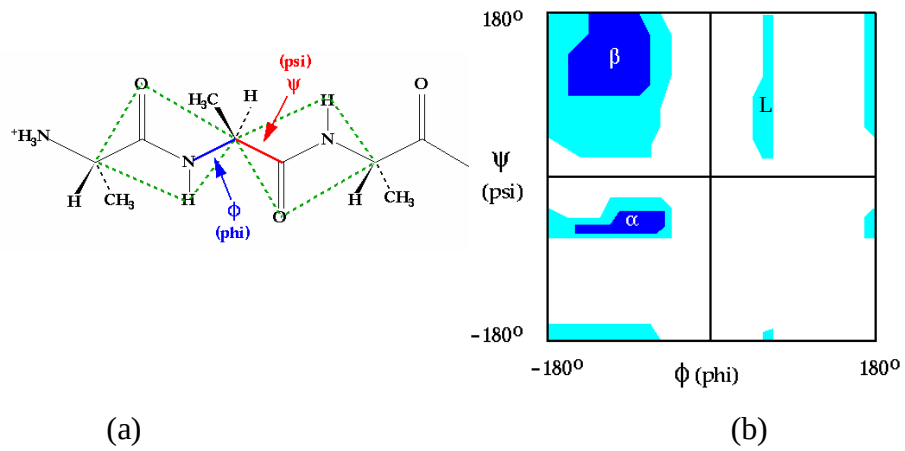
Proteins and polypeptide chains composed of amino acids linked together by peptide bonds are known as primary structures. Each amino acid consists of a central tetrahedral carbon (C^{α}) which in turn is connected to (1) a hydrogen atom (2) amino group (3) carboxyl group (4) distinguishable side chain or R group. There are 20 different types of naturally occurring amino acids having different R groups. All the amino acids observed in nature are in L-isomer; however why this isomeric property has been chosen by nature

is still unknown. A broad classification of amino acids divides them into 3 groups characterized by the chemical nature of the side-chain.

1. NPO: Amino acids with strictly non-polar side chains
2. CPO: Amino acids with charged polar residues
3. UPO: Amino acids with uncharged polar side chain

During protein synthesis $-NH_2$ group of 1^{st} amino acid interacts with the –COOH group of the 2^{nd} amino acid, forming a peptide linkage (C-N bond) with release of one water molecule. This process is repeated as long as the sequence of amino acids elongates. The mRNA sequence drives the synthesis of protein in ribosome. The X-ray diffraction studies reveal that the peptide C-N bond is shorter than that of the normal C-N bond as observed in simple amines. So also the atoms involved in the formation of the peptide bonds (C-N) are coplanar in nature. This leads to the conclusion that C-N bonds in amino acids are associated with some double bond character. Therefore their free rotations are prohibited. Rotations are permitted only through the $N\text{-}C^{\alpha}$ and $C^{\alpha}\text{-}C'$ bonds. Thus the backbone of polypeptide chains is made of rigid planes known as peptide plane. By convention, the dihedral angle φ signifies rotation around $N\text{-}C^{\alpha}$ bond and ψ denotes rotation around $C^{\alpha}\text{-}C'$ bond (Figure 1.22a). In this way the amino acid residues are associated with two conformational angles φ and ψ. In principle φ and ψ can take any values between -180^{o} to $+180^{o}$, however many are prohibited by the steric interactions between the atoms of polypeptide backbone and amino acid side chains. The allowed values of φ and ψ are given in the Ramachandran plot (Figure 1.22b). The major allowed

area are occupied by right-handed α helices, β strand and left handed α helices, the conformations of these structure are discussed below.



(a) (b)

Figure 1.22. (a) Schematic diagram of peptide backbone with backbone dihedral angles φ and ψ (b) Ramachandran plot showing sterically allowed φ and ψ angles for α-helices, β-sheet, and left handed structures (designed as L), the blue color region represents the fully allowed region while the sky color presents the partially allowed region (taken from http://iop.vast.ac.vn/theor/conferences/smp/1st/kaminuma/UCSFComputerGraphicsLab/AAA.html)

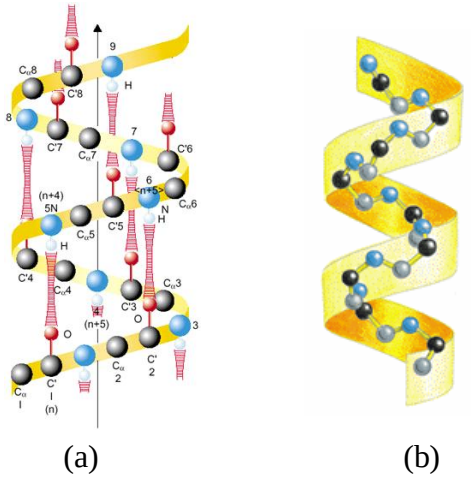
Secondary structure

Secondary structures are some regular and repetitive structures in protein molecules, which provide a rigid and stable framework to the protein. In a protein φ and ψ may adopt some well-defined values that allow the establishment of interchain or intrachain hydrogen bonds. These hydrogen bonds make the corresponding conformations more stable than the other ones. Such stabilized conformation constitutes the secondary structure of the proteins.

α-helices

A systematic establishment of hydrogen bonds between the C=O of residue i with N-H of residue i+4 form the α-helices (Figure 1.23a, b). This results in a strong hydrogen

bond that has nearly optimum N-O distance around 2.8 Å. A regular α-helix in a protein is associated with (φ,ψ) value around (-60°, -50°) respectively which corresponds to the allowed region of the bottom left quadrant of the Ramachandran plot as shown by Figure 1.22b. Each helical turn includes 3.6 amino acids residues. So all the N-H and C=O groups are involved in the hydrogen bonds except the amino terminus of the helix, which has a positive charge and carboxyl end which has a negative charge. These provide a dipole moment to the helix. So the ends of the α-helix always lie at the surface of the protein molecule. The dipole moment corresponds to 0.5-0.7 unit charge at each end of the α-helix, which in turn helps in binding the charged ligands. In α-helical structure, the side chain of each amino acid remains exposed, pointing radially outwards from the helical axis.



(a) (b)

Figure 1.23. (a) Formation of a right handed α-helix, showing hydrogen bonds (denoted by red lines) between the C=O of residue i with N-H of residue i+4 (b) schematic diagram of α-helix

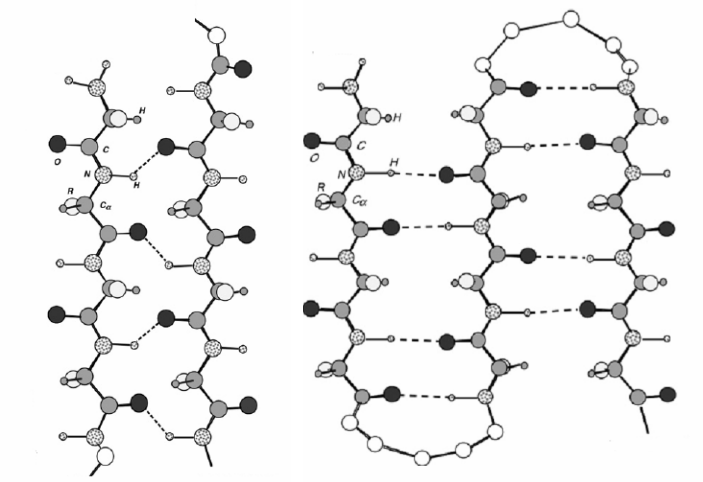
Some of the amino acids like glutamine, methionine favor α-helix formation, while valine, glycine serine destabilize the α-helix, so also residue with bulky side chains are poor in forming alpha-helices. Proline residues do not have N-H group and hence is

incapable to form H-bond between i and i+4th residue, hence it acts as helix breaker. So the tendency of a given segment of a polypeptide chain to form an α-helix depends on the identity and sequence of amino acids residues within the segment. The 3_{10} and π helices are observed as the variants of the α-helix. In case of 3_{10} the hydrogen bond is formed between residues i and i+3, with (φ,ψ) angles be (-50^o,-25^o). These 3_{10} helices are observed occasionally in proteins, because of the steric interference of the R-groups involved. Where as in case of π helices hydrogen bonds are observed between residues i and i+5 residues with φ and ψ angles be -60^o and -70^o respectively. These occur rarely and observed as segments of longer helices, since the backbone atoms are loosely packed, they leave an axial hole, thus reduces its stability relative to more closed packed conformations. The left handed α-lelices as shown by the upper right of Ramachandran plot are occasionally found in proteins.

β-sheet

β-sheet is the more extended conformation of the polypeptide chain and are usually 5-10 residues long, corresponds to the allowed region of the upper left quadrant of the Ramachandran plot. In contrast to the α-helices, where H-bonds are always between nearby residues, in β-sheet hydrogen bonds are formed between adjacent segments of polypeptide chains, and also with the distant segments also. These are of two kinds, parallel and antiparallel β-sheet. In parallel β-sheet the hydrogen bonded chains are extend in the same directions. In this case the amino acids of the two polypeptides adopt φ and ψ values around -110^o and $+120^o$ respectively (Figure 1.24a), the tetrapeptides belong to the same proteins but are separated along the sequence of protein by an

appreciable number of amino acids, which may involve in the formation of the β-sheet, held together by bent hydrogen bonds between C=O and N-H groups. In case of antiparallel β-sheet, the neighboring hydrogen bonded polypeptide chains extend in the opposite directions, with φ and ψ be -130^{o} and +145^{o} respectively (Figure1.24b). The antiparallel β-sheet has narrowly spaced hydrogen bond pairs that alternate with widely spaced pairs. In most of the cases, four or more strands align side by side forming parallel or antiparallel β-sheet to form an extended sheet. These sheets are often twisted or rolled to give a three dimensional shape. The side chains of alternate residues are projected in the same face of the β-sheet.



(a) (b)

Figure 1.24. The β conformation of the polypeptide chain (a) parallel β-sheet (b) antiparallel β-sheet (Maréchal, 2007) showing hydrogen bonds between amino terminal and carboxyl terminal

Hydrogen bonds are in less constrained configurations in the antiparallel β-sheets than that of the parallel β-sheet. These antiparallel β-sheets are comparatively more stable and more easily formed than that of the parallel β-sheet. So also the turns those separated the

two sequences of such conformations can be shorter in case of antiparallel β-sheet, than that of the parallel ones (Figure 1.24).

The secondary structural elements are joined by loops in a protein. The exposed C=O and N-H groups of the loop regions are generally exposed to the solvent and can form hydrogen bonds with the solvent molecules. They can serve as a binding site of the ligands, so also form the active sites of many enzymes. Two adjacent antiparallel β-strands are joined together by hairpin loops. Combinations of secondary structural elements in specific geometric patterns give rise to formation of motifs, also called as supersecondary structures of the protein. Some of the motifs perform specific biological functions; ex. Helix turn helix motif is specific for DNA binding.

Tertiary structure of proteins

The α-helix and β-sheet are arranged in a definite order in a protein to make it active. This typical ordering forms the tertiary structure of the proteins. In other words tertiary structures refer to the complete 3-dimensional structure of a protein molecule. Domain is structural unit of tertiary structure of proteins, which is defined as a polypeptide chain or part of polypeptide chain, that can fold independently into a stable tertiary structure. The different combinations of secondary structural elements and motifs form the building block of the domains. These domains are classified into three categories. (i) α-domain (ii) β-domain (iii) α/β domain (iv) α+β domain. In case of α-domain the core is built of α-helix. In β-domain, the antiparallel β-sheet forms the core and the α/β structures are combinations of β-α-β motifs. Another fourth class is the α+β domain. The interactions of different domains within the tertiary structure of a protein are governed by several forces; those include (i) hydrogen bonding (ii) electrostatic

interaction (iii) van der Waals forces (iv) hydrophobic interactions (v) formation of disulphide bonds. Different proton donors and acceptors present at the protein backbone and at the side-chains of the amino acids and they can form hydrogen bonds with the surroundings water molecules. I have previously discussed about the classifications of amino acids according to the hydrophobic/hydrophilic nature of their side chains. These characteristics nature of the side chains of amino acids play an important role in shaping the protein structure. The electrostatic interactions like charge-charge interactions between the oppositely charged R-groups and the charge-dipole interactions between ionized R-groups of the amino-acids with the water molecule play an important role. Although van der Waals forces are weak in nature compared to the other forces required for protein folding, it plays crucial role in maintaining the protein structure. van der Waals interaction are predominant in the interior of proteins as hydrophobic side chain of valine, leucine, isoleucine, phenylalanine etc. are mostly packed there and these side chains are incapable to form hydrogen bond.

Quaternary Structure

The arrangement of two or more non-covalently linked polypeptide chains in a three dimensional proteins form a quaternary structure. In other words we can say some proteins require the presence of other proteins or ions to be active and to perform some specific functions, this form is known as the quaternary structure of the protein. Hydrophobic interactions, salt bridges, hydrogen bonds act as main stabilizing forces in maintaining the quaternary structure of the proteins. The most common example is the haemoglobin, it consists of four polypeptide chains, where each of the four subunits

contain an oxygen binding heme groups, which are arranged symmetrically. So also several virus coats are made of many proteins, which are quaternary structure.

1.2.3. Structural Features of other Biomolecules

In addition to DNA, RNA and protein several other biomolecules are found in nature and they play important roles in many biological processes. Carbohydrates are the most abundant class of biological molecules and are involved in wide range of process from molecular recognition, cell signaling, protein stabilization etc (Dashnau et al., 2005). The basic units of carbohydrates are monosaccharaides, such as glucose, ribulose etc (Figure 1.25a). Polysaccharides such as cellulose (Figure 1.25b), starch consist of many covalently linked monosaccharide units. The general chemical composition of carbohydrate is CH_2O. Monosaccharide can be classified depending on the chemical nature of the carbonyl group and number of carbon atoms. The chemical composition within a class of monosaccharide is identical; however, the orientation of the hydroxyl groups can differ across stereoisomers.

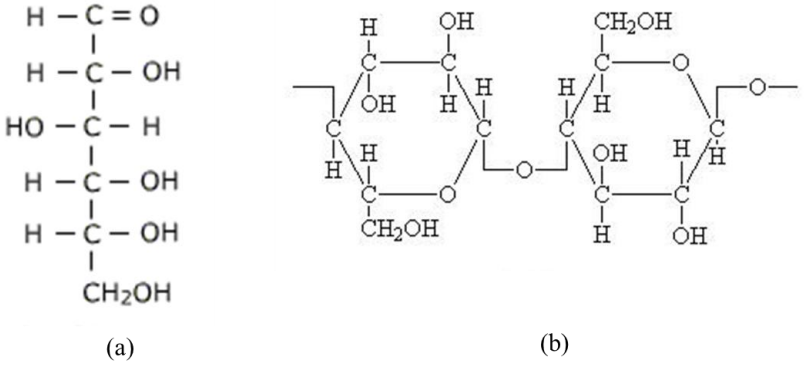


Figure 1.25. General Structures of (a) Glucose (b) Cellulose

The variations in the orientation of the hydroxyl groups can account for different physical and chemical properties of the sugars. With increase in the number of hydroxyl groups in carbohydrate can enhance the interaction of carbohydrates with other proteins and solvent. Hydrogen bonding interaction also plays crucial role for the specificity of carbohydrate-carbohydrate self-assembly in natural systems (de la Paz et al., 1997). Many synthetic receptors also employ sugar hydroxyl groups as hydrogen bonding interaction in recognition processes. Arrangement of the hydroxyl groups into intramolecular H-bond networks can lead to a phenomenon known as cooperatively, which paly crucial role in molecular recognition.

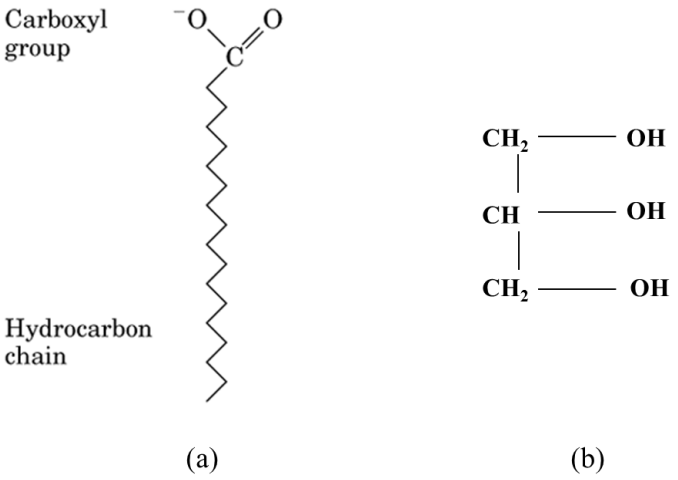


Figure 1.26. Structural features of (a) Fatty acid (b) Glycerol

Lipids form the important component of biological membranes. They are found to be integral component of vitamins, hormones, oils etc. Fatty acids, triacylglycerol, glycerophospholipids are some of the classes of lipids. Fatty acids are made of carboxylic acids with long hydrocarbon chains, while triacylglycerol are fatty acid triesters of

glycerol (Figure 1.26). The greater varieties of different lipid membranes with modifications in hydrocarbon chain, polar groups, and backbone suggests that many of these lipids may play unique role in structure and functions of membrane (Boggs, 1987). The hydrogen bonding at the aqueous lipid bilayer interface plays significant role in many biological processes. Lipid has the ability to interact with other lipid molecules by intermolecular hydrogen bonding interactions with hydrogen bond donating groups such as --P--OH,--OH, and--NH3+ and hydrogen bond accepting groups such as --P--O- and –COO (Boggs, 1986). Hydrogen bonding interactions depend on the ionization state of the lipid; they can be altered by changes in the environment which affect the pK of the ionizable groups.

There are several others biomolecules or molecules that take part in a biological cellular processes, such as natural antibiotics, artificial synthetic molecules are also used for the treatment of disordered/diseases cells. In this thesis I have analyzed few such molecules.

Chapter 2

Importance of Quantum Mechanics to Study Different Properties associated with Biomolecular Structure and Recognitions

2.1. Introduction to Computational Quantum Chemistry

Computational chemistry is an important technique to evaluate the molecular geometries, rates of reactions and other physical and chemical properties of the molecules. Computational chemistry is widely used to explore the interaction of potential drugs with the biomolecules. It is also helpful in studying different properties such as conducting, magnetic properties of different metals. Computational chemistry includes quantum mechanics, molecular mechanics, molecular simulation etc. to understand properties of the molecular systems. Quantum mechanics explicitly considers electronic distribution, so useful in calculating molecular properties which depends on the electron distribution. Quantum chemistry is useful in understanding molecules at the sub atomic levels in details. Different chemical properties of the molecules like electronic spectra, nuclear magnetic resonance can also be studied by quantum mechanics (QM). It gives us the information about the total energy of the systems (Lewars, 2011). By simulation and modeling, we can predict several properties of a system, which is hard to study by experiments. There are certain computational properties, which do not correspond to physical observables, such as aromaticity, bond-order, amount of charge transfer between hydrogen bonded systems and are best studied by quantum mechanics. There are number of quantum theories available for treating the molecules; the principal approaches involve *ab initio* method, and semiempirical approach. The major advantage of *ab initio* quantum chemical calculations is its applicability in absence of any empirical parameters. Semiempirical quantum mechanics are also used to study the biomolecules, but they are not suitable for molecular interaction (Hobza et al., 1997).

2.2. Tools of Quantum Chemistry and Schrödinger Equation

The tools of computational chemistry are force fields (molecular mechanics), *ab initio*, semi empirical and density functional methods. In the Force Field method, the electronic energy is given as the parametric function of the nuclear coordinates and then fitting the parameters to experimental or higher level computational data. Force field method is fast and suitable for large molecules.

For non-relativistic atoms, Schrödinger equation is sufficient to calculate all the properties related to the system. For heavier elements, some relativistic effect has to be introduced using relativistic pseudo potentials.

Schrödinger equation gives analytic solutions for few simple systems, whereas for many particle complex systems, numerical approaches have been used to solve Schrödinger equations. Solving Schrödinger requires incorporation of some level of approximation to the wave function. The basic method used in QM calculations is Hartree-Fock approximation, which solves the time-independent Schrodinger equation by assuming that electron moves in the average fields of the other electrons (Cramer, 2004).

$$H\psi = E\psi \qquad \text{.....[2.1]}$$

giving us energy and wave function, the latter is a mathematical function used to calculate electron distribution, without appeal to fitting to the experiments. Where, H is known as the Hamiltonian operator and can be written as sum of the kinetic and potential energy of the nucleus and electrons.

$$H_{Total} = T_{el} + V_{el\text{-}el} + T_{nuc} + V_{nuc\text{-}nuc} + V_{el\text{-}nuc} \qquad \text{.....[2.2]}$$

T_{el} and T_{nuc} represent kinetic energy of the electron and nucleus respectively, whereas $V_{el\text{-}el}$, $V_{nuc\text{-}nuc}$, $V_{el\text{-}nuc}$ represent the potential energy terms.

2.3. Born-Oppenheimer Approximation

Calculations of accurate wave function of many particle systems are extremely difficult, as the Hamiltonian contains many correlated attractive and repulsive terms. In principle, solving Schrödinger Equation is not possible for many-electron atoms or molecules due to many-body problem. So an approximation has been introduced known as Born-Oppenheimer approximation, where coupling between electron and nuclear motion is neglected. It is based on the hypothesis that as masses of the nuclei are considerably heavier than that of the electrons, so nuclei moves much slowly than that of the electrons. The electronic wave function thus depends on the position of nuclei and not on their momentum. Born-Oppenheimer approximation assume electronic and nucleus motion are independent of each other, considering study of the electrons in the field of frozen nuclei. Under the Born-Oppenheimer approximation, the total Hamiltonian of the molecules is written in the following form.

$$H_{approx} = T_{el} + V_{el\text{-}el} + V_{el\text{-}nuc} = H_{el} \quad \text{.......................... [2.3]}$$

$V_{el\text{-}nuc}$ = electron-nucleus cross term which fix nuclear positions

The electronic Schrödinger equation can be written as

$$H_e\,(r;R)\psi_e\,(r;R) = E_e\,(R)\psi_e\,(r;R) \quad \text{..............................[2.4]}$$

which is solved for all nuclear coordinate R to map out the potential energy surface. Born-Oppenheimer approximation usually gives good results in ground electronic state. However in exited states, problem arises due to non-adiabatic interactions.

2.4. Variation Theorem and Orbital Approximation

The average total energy for a state specified by a particular ψ is given as expectation values of H given as

$$E[\psi] = \int \psi^* \hat{H} \psi dr \equiv \langle \psi \mid \hat{H} \mid \psi \rangle \qquad \text{.....[2.5]}$$

The notation [Ψ] emphasizes that energy is a *functional* of the wave function. The energy is higher than that of the ground state unless Ψ corresponds to Ψ_0 – which is known as *variational theorem*;

$$E[\Psi] \geq E_o \qquad \text{..... [2.6]}$$

The dynamics of many electron systems are very complicated. For simplification of the $V_{el\text{-}el}$ term, we have incorporated another level of approximation in many electron system. This is known as orbital approximation by introducing the independent particle model, where the motion of an electron is considered to be independent of the dynamics of the other electron. Independent particle model assumes that each electron moves in its own orbital, and it ignores the correlation with other electrons.

2.5. Linear Combination of Atomic Orbitals (LCAO) Approximation

The eigenfunctions of eqn. 2.4 for a molecular system are called molecular orbitals (MO). The energy of the electron in that particular orbital can be obtained from the energy eigen values associated with each molecular orbital. The eqn. 2.4 can be solved by constructing a guess wave function ϕ, which is expressed as a linear combination of atomic wave functions φ. This approach is known as linear combination of atomic orbitals (LCAO).

$$\phi = \sum_{i=1}^{N} ci\,\varphi i \qquad \text{..... [2.7]}$$

Where *ci* are the coefficients, their values can be obtained by minimizing the energies.

2.6. Huckel Theory

Huckel theory is another approach to understand the molecular orbital approximation methods. Huckel theory is limited to conjugated π-systems and capable to explain the non-additive nature of certain properties of aromatic compounds. Huckel theory employs some of the mathematical concepts and techniques; those are useful in current high level *ab initio* orbital approximation also. According to Huckel theory, the total wave function of a molecule can be written as a product of its σ and π parts.

$$\psi_{total} = \psi_\sigma\, \psi_\pi \qquad \text{.....[2.8]}$$

ψ_σ is the wave function describing the electronic sigma orbital and ψ_π is the wave function describing the electrons in the pi orbital. The wave function of the pi electron can be described as the product of all the pi molecular orbitals.

$$\psi_\pi = \psi_1\, \psi_2\, \psi_3\, \psi_4 \ldots\ldots\ldots\, \psi_N$$

where the 1….N represents the number of molecular orbitals. The ψ_1 to ψ_N are the wave functions for each electron moving in the electrostatic field of nuclei and the other electrons. The π electrons are thus considered to move in the field of nuclei and core sigma electrons. Each molecular orbital ψ_i is constructed as a linear combination of atomic orbitals according to LCAO approach.

$$\psi_j = \sum_{r=1}^{N} C_{jr}\varphi_r \qquad \text{..... [2.9]}$$

With j = 1… N , φ_r is the $2p_z$ atomic orbital on atom r of the conjugated π systems. C_{jr} are the coefficients associated with each atomic orbital. Huckel theory gives insight about chemical reactivity and other associate properties about certain compounds.

Huckel theory has certain limitations.

1. Huckel theory has some drawbacks as it is restricted to π- systems.
2. It ignores the electron-electron repulsion.
3. It is useful for qualitative study of the conjugated systems, but because of its limitation it is not useful for quantitative purposes.

2.7. Extended Huckel Theory

Extended Huckel theory has certain improvement over Huckel theory to take in account of the orbital approximation. It considers all the valence electrons in the molecular orbital calculation. Extended Huckel theory can be useful in calculating the molecular structure, transition states for reactant species etc. Whereas it has also certain limitations, it fails to take in account of the electron spin, so cannot distinguish between different multiplets.

2.8. Hartree-Product Wave Function

The solution to the Schrödinger equation with non-interacting Hamiltonian can be given as the product of the electronic wave functions given as

$$\Psi_{HP} = \psi_1 \psi_2 \psi_3 \psi_4 \ldots\ldots\ldots \psi_N \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [2.10]$$

Where the one electron wave functions are called orbitals, and in case of molecules, this is called molecular orbital, where each molecular orbital can be expanded in a basis set of atomic functions.

$$\psi_i(r) = \sum_{i=1}^{nbasis} C_{ij}\varphi_j(r) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad [2.11]$$

The wave function formed Ψ_{HP} is known as the Hartree-product wave function.

$$H\,\Psi_{HP} = (\sum_{i=1}^{N} \varepsilon_i)\ \Psi_{HP} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad [2.12]$$

2.9. Introduction to Slater Determinant and Hartree-Fock Equation

So the energy eigen values of many electron wave function can be written as sum of one electron eigen values. This represents a non-interacting system of electrons. Hartree product assumes that specific electrons have been assigned to specific orbitals, whereas the antisymmetric principle requires that electrons are indistinguishable in nature. According to Pauli Exclusion Principle, no two electrons can be characterized by same sets of four quantum numbers. The electronic wave functions must change sign whenever the coordinates of the electrons are interchanged, such wave functions are known as antisymmetric wave functions. For a closed shell system, where all electrons are paired, antisymmetric wave function can be represented by Slater determinant (SD) of the spin orbital; Slater determinant is the simplest form of the orbital wave function that satisfies antisymmetric principle.

Slater determinant for the many electronic systems can be represented as

$$\psi(x_1, x_{2,}, ..., x_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(x_1) & \chi_2(x_1) & \cdots & \chi_N(x_1) \\ \chi_1(x_2) & \chi_2(x_2) & \cdots & \chi_N(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(x_N) & \chi_2(x_N) & \cdots & \chi_N(x_N) \end{vmatrix}$$

…….. [2.13]

Where $\chi(x_i)$ indicates product of spatial wave function and spin orbital.

So for many electron system,

$$\psi(x_1, x_2 .... x_N) = A[\sum_{i=1}^{N} \chi_i(x_i)]$$

Where $\chi_i(x_i) = \psi_i \alpha$ or $\chi_i(x_i) = \psi_i \beta$

α and β are the spin coordinates representing up and down spins respectively.

$$\psi_i(r_i) = \sum_{j=1}^{nbasis} C_{ij} \varphi_{ij}(r_i) \qquad \text{…… [2.14]}$$

The coefficient C_{ij} can be obtained by solving the Schrödinger wave equation.

$$\hat{H}_{ele} \psi(x) = E_{el} \psi(x) \qquad \text{……[2.15]}$$

This equation can be solved by the variation principle. We have to choose a guess wave function first and then solve it iteratively until we get the lowest energy. This procedure is known as self-consistent field method.

Energy of a Slater determinant can be obtained as

$$E_{SD} = \int \psi^*_{SD} \hat{H}_{ele} \psi_{SD} dx \qquad \text{…… [2.16]}$$

On solving the equation explicitly, we obtain energy of Slater determinant as

$$E_{SD} = \sum_{i=1}^{N} h_i + \sum_{i=1}^{N}\sum_{j>i}^{N} (J_{ij} - K_{ij}) \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad [2.17]$$

Where K_{ij} is called exchange integral, it appears in accordance with the Pauli exclusion principle, which postulates two electron can exchange their position.

J_{ij} the Coulomb integral, represents the columbic interaction between the electrons on orbital *i* with an electron in orbital *j*

Therefore we can calculate the energy of a Slater determinant wave function, and optimize the wave function according to the variational principle; this is called as the famous Hartree-Fock equation.

$$E_{HF} = V_{NN} + \sum_{i=1}^{N} h_i + \sum_{i=1}^{N}\sum_{j>i}^{N} (J_{ij} - K_{ij}) \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad [2.18]$$

V_{NN} represents the potential energy due to nuclear-nuclear repulsion.

2.10. Associated Properties of Molecules Obtained from *ab Initio* Quantum Mechanics

Quantum mechanics gives good results for the hydrogen bonded systems. The very important advantage is QM calculation can be done on any selected geometry, even in regions, which cannot be studied by experiments. Accuracy of *ab initio* calculations is determined by the size of basis set of atomic orbital which is used to construct the molecular orbital and incorporation of correlation effects. Results from the high level *ab initio* QM methods are comparable to the best physico-chemical experiments (Sponer, 2006). Thermodynamics and structural properties of molecules can be well studied by *ab initio* quantum chemical methods. Calculation of electric multipoles gives us the distribution of charge in a molecule. We can calculate dipole, quadruple and other higher

order electric multipole moment through *ab initio* quantum studies. Population analysis helps in calculating the atomic charge on each nucleus. Different methods such as Mulliken analysis, Natural population analysis, Atoms in molecules are available to study the population analysis. Electrostatic potential has contributions from both the nuclei and electrons; it is useful in molecular recognition processes.

2.11. Drawbacks and Improvements to Hartree-Fock Theory

The most significant drawbacks of Hartree-Fock theory are that it fails to estimate adequate electron correlation due to consideration of independent particle model. The Hartree-Fock energy with even infinite basis set and accuracy is significantly more than the exact energy. This perhaps comes from the correlation. London dispersion interaction arises from the instantaneous dipole-induced dipole interaction, are the consequences of intermolecular correlation of the electrons. Correlation of electrons has to be taken into account to obtain accurate values of charge distributions. The electrons are assumed to move in an average potential of the other electrons, so the instantaneous position of the electron is not influenced by the presence of neighboring electrons, so gives the energy higher than the exact energy.

$$E\ (\text{correlation}) = E\ (\text{exact}) - E\ (\text{HF})$$

Electron correlation plays significant role in studying the intermolecular interaction. Geometries obtained through Hartree-Fock methods are often reliable but the energies are not reliable. There are number of ways in which correlation effect can be incorporated in Hartree-Fock method, such as configuration interaction and perturbation theory.

2.11.1. Configuration Interaction

In configuration interaction (CI) approach, exited states are included in the description of electronic state. Wave function is described as a linear combination of the ground state and exited state wave functions. In this approach, the complete wave function is represented as

$$\psi = \psi_{HF} + \psi_S + \psi_D + \psi_T + \ldots \quad \text{..... [2.19]}$$

whereas S represents all single excitations, D represents all double excitations and so on. Complete CI gives exact wave functions for the given atomic basis, but it is possible for small molecules, while impossible for large molecules due to increase in the number of configurations involved.

2.11.2. Many Body Perturbation Theory

Moller and Plesset proposed perturbation theory to tackle the problem of electron correlation. This method is based on adding successive improvements to both energy and wave functions to the HF description. In this case the true Hamiltonian operator is expressed as sum of the zeroth order Hamiltonian Ho and a perturbation ν, which is expressed as

$$H = Ho + \nu \quad \text{.....[2.20]}$$

To obtain an improvement of Hartree-Fock energy, it is necessary to obtain Moller-Plesset perturbation theory (Møller and Plesset, 1934) to at least second order, commonly known as MP2 level of theory. Higher orders of calculations such as MP3, MP4 are also possible, but those are computationally expensive, inappropriate for complex species.

They are often restricted to calculate the single point energy calculation on geometries obtained through some lower level of theory.

2.12. A Brief Introduction about Basis Set

Basis sets are composed of atomic functions from which molecular wave functions are constructed (Leach, 1996). The atomic wave functions can be Slater type or Gaussian type functions. The Slater functions have the form $e^{-\alpha r}$ and are used in molecular orbital calculations. It has certain drawbacks like when the atomic orbitals are based on the different atoms, and it is time consuming also. These drawbacks can be removed by introducing Gaussian type wave functions, which has the form

$$\phi_{nlm}^{GTO}(x,y,z) \propto x^n y^l z^m e^{-\alpha r^2} \qquad \text{.....[2.21]}$$

Where α determines the width of the radial function, smaller value of α gives large spread, while greater value of α does not spread very far. The integers n, l, m are non-negative integers which determines the nature of the orbitals. When all the integers are zero, we get the s-type orbital, sum equal to 1 gives the p_x, p_y and p_z types of orbitals depending on the symmetry about the given Cartesian axes, when the sum of the indices are two, we get d-type Gaussian orbital. The advantage of Gaussian function is that product of two Gaussians can be represented as a single Gaussian along the line, joining the centers of two Gaussians. The Gaussian function has certain drawbacks, it decays towards zero more quickly, than that of the Slater orbital, so could not describe the nature of the exponential tail properly, underestimating the long-range overlap between the atoms. This problem can be solved if we represent a single Gaussian function as a linear combination of Gaussian function, which now decays slowly like that of the Slater

function. This form of Gaussian is known as contracted basis function, and the individual Gaussian from which they are formed are known as primitive or uncontracted Gaussian. The contraction length is the number of terms used in the expansion.

A contracted Gaussian orbital is represented as

$$\phi_{nlm}^{CGTO}(x,y,z) = \sum_i C_i \phi_{nlm,i}^{GTO}(x,y,z) \qquad \text{.....[2.22]}$$

Where C_i represents the contraction coefficient associated with each basis functions. Different types of basis sets are used in the ab-initio quantum chemical methods, depending on the nature of atoms in the systems.

The primary one is the minimal basis set, which has one basis function for each atomic orbital in the real atom. In STO-nG, n numbers of GTO are used to mimic the STO. Minimal basis set often give good geometries for the smaller system, but it gives poor results for other energy related quantities and other properties and lack the flexibility when involved in bond formation with other atoms. Minimal basis sets are unable to describe the non-spherical aspects of the electron distributions. The double zeta polarized basis sets (DZ), where two atomic orbitals are used for each atomic orbital in an atom, but insufficient to describe proper energy of the system. To overcome these problems split valence basis sets come in to the picture, which are generally presented by 3-21G, 6-31G etc. Where in case of 6-31G basis set, 6 Gaussian functions are used to describe the core orbital, the valence electrons are described by 4 Gaussian, the contracted part of the 3 Gaussian and the diffusion part by one Gaussian. The charge distribution about an atom in a molecule is usually perturbed in comparison with the isolated atom, like when a hydrogen atom form molecule, it acquires some p-type of characteristics in addition to the s-character, and first and second rows elements of the

periodic table adopt d-orbitals characteristics when involved in molecule formation. So polarization functions are added into the basis set, which are normally presented by (*), 6-31G** basis is equivalent to 6-31G basis set, with inclusion of polarization functions on both the hydrogen and other non-hydrogen atoms, useful when hydrogen atoms are involved in formation of bonds. Diffusion functions are necessary in the basis sets of the species, like anions, and lone pair containing molecules, where significant electron density away from the nucleus. The diffusion functions are denoted by adding "+" to the basis sets of the atoms. The other basis sets most widely used are the Dunning correlation consistent basis sets, ranging from double zeta plus polarization to hextuple zeta plus polarization plus diffuse, which is best choice for weakly bound species get accurate results.

2.13. Density Functional Theory

Density functional theory (DFT) (Kohn, 1999) considers electron correlation explicitly; it is more economic than MP2 or other post Hartree-Fock method. In Hartee-Fock theory, the multi-electron wave function is expressed as a Slater determinant which is constructed from a set of N-electron wave functions. The DFT method is based on evaluation of electronic energy which in terms of depends on the overall electronic density. Study by Hohenberg and Kohn (Hohenberg and Kohn, 1964) shown that the ground state electronic energy and other properties of the systems were uniquely defined by electron density. In DFT, the energy functional is written as the sum of two terms given by

$$E[\rho(r)] = \int V_{ext}\rho(r)dr + F[\rho(r)] \quad \text{.....[2.23]}$$

Where the first term represents interactions of electrons with an external potential V_{ext} , the explanation of the second term was properly given by Kohn and Sham(Kohn and Sham, 1965), which was given by

$$F[\rho(R)] = E_{K.E}[\rho(r)] + E_H[\rho(r)] + E_{XC}[\rho(r)] \qquad \text{.....[2.24]}$$

The first term represents the kinetic energy, second term represents Coulombic energy and the third term represents contributions from exchange and correlation functions. Different methods have been proposed to take in account of the exchange-correlation functional. The simplest way is the "Local Density Approximation (LDA)", where the electron density seems to be constant throughout the space. The other well-known method is the "Generalized Gradient Approximation (GGA)", which depends on the gradient of the density on each point of the space, and not just in its value. The most common GGA functions were developed by Becke (Becke, 1993). The other known example is LYP functional as developed by Lee, Yang and Parr (Lee et al., 1988). Some of the DFT functionals contain empirical terms and are tested to give good results, than that of the non-empirical functionals like PBE, PW91. The most widely accepted method is B3LYP (Gill et al., 1992) functional, which is a combination of gradient corrected exchange correlation functional with certain portion of the exact exchange from Hartree-Fock calculations, thus gives accurate results in case of hydrogen bonded systems.

2.14. Dispersion Corrected Density Functional Theory

Density functional theory is very widely used theoretical approach to molecular structure. It includes electron correlation in an approximate manner. However it is unable to describe the long range electron correlation, which plays significant roles in describing

the stacked complexes (Kristyán and Pulay, 1994, Cybulski et al., 2002). Latter on different methods arise, which takes in account of the dispersion interaction, by considering the long-range van der Waals interactions can be done by introducing the dispersion interaction either by proposing a suitable non-local functional or by combing the standard functional with empirical dispersion terms (van Mourik and Gdanitz, 2002, Elstner et al., 2001). The empirical potential has been added in the form of C^6R^{-6}. After adding the dispersion correction the total dispersion corrected energy takes the form of

$$E_{DFT\text{-}D} = E_{KS\text{-}DFT} + E_{dis} \quad \text{.....[2.25]}$$

Where the 1st term in the right hand side represents the self-consistent Kohn-Sham energy obtained from the chosen DFT functional and the 2nd term represents the empirical dispersion correction, which has the form as given below

$$E_{dis} = -S_6 \sum_{i=1}^{N_{at}-1} \sum_{j=i+1}^{N_{at}} \frac{C_6^{ij}}{R_{ij}^6} f_{dmp}(R_{ij}) \quad \text{.....[2.26]}$$

Where N_{at} represents the number of atoms in the system, R is the interatomic distances, and C_6 are the dispersion coefficients, the damping function is represented as $f_{dmp,}$ which has the following expression

$$f_{dmp}(R_{ij}) = \frac{1}{1+e^{-d(R_{ij}/R_r - 1)}} \quad \text{.....[2.27]}$$

Where R_r is the sum of the atomic van der Waal radii.

The most widely used hybrid functional which include long range dispersion interaction is ωB97XD(Chai and Head-Gordon, 2008) functional, which has been incorporated in a commercial software package Gaussian g09, which has been extensively used by me in my thesis work (Frisch et al., 2009). Zhao and Truhlar (Zhao and Truhlar, 2006) have developed the M06 family of local (M06-L) and hybrid (M06, M06-2X) meta GGA

functional, which gives good results for non-covalent interactions as studied by various groups (Umadevi and Sastry, 2011).

2.15. Dirac equation and relativistic effect

Heavy elements in periodic table contain rather distinct characteristics and follow Dirac equation. Special relativity has basic postulates (i) speed of light is constant in any reference frame (ii) laws of physics are invariant to choice of reference frame. In general Schrödinger equation does not follow the above rules, so are non-relativistic in nature. However, Dirac equation follows the two above criteria, so relativistic in nature. Dirac equation always gives interaction energy lower to that of Schrödinger equation for a simple system like one electron system and with increase in the charge of the nucleus (For many electron system), the differences between the relativistic and non-relativistic energy becomes bigger. Many properties of the atoms and molecules can be explained with the help of non-relativistic quantum mechanical approach. However when the nuclei get heavier, with increase in the atomic number, the relativistic effect plays a dominant role in the chemical process. According to the theory of relativity, the relativistic mass increase as

$$m = \frac{m_0}{(1-(v^2/c^2))^{1/2}} \qquad \text{..... [2.28]}$$

Where m_o being the rest mass and v be the speed of the electron. The effective Bohr radius also decreases for the inner electrons, which move with large average speeds, according to the following equation.

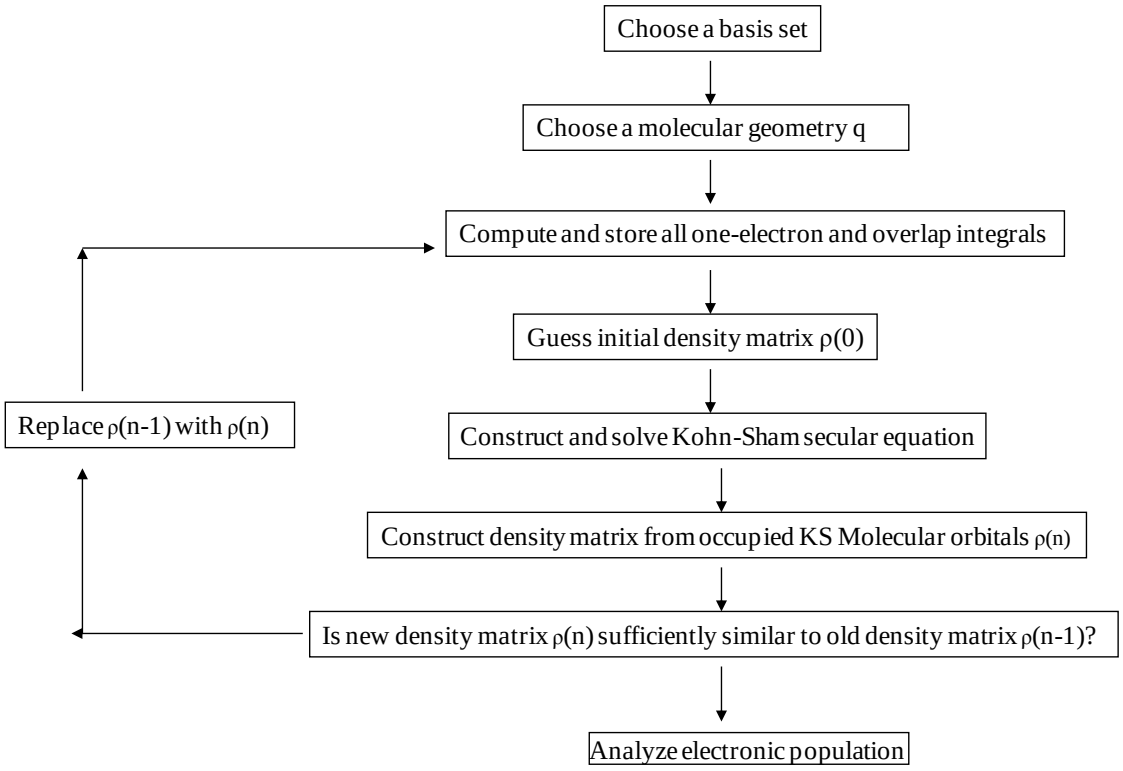
$$B_0 = (4\pi\varepsilon_0)(\hbar^2/me^2)$$

As atomic nucleus charge (Z) increases, electrons that orbits close to the nucleus (the s electrons) increase their average velocity. This result in energy stabilization and relative radial contraction of the s and p orbital those lie closer to the nucleus, while d and f orbital are destabilized and tend to expand radially and outwardly (Pyykko, 1988). However the higher s-shells also suffer similar contraction, as they are orthogonal to the lower ones. Relativity has another influence beyond stabilizing electrons, known as spin-orbit coupling. These contractions, decontractions, stabilizations, destabilizations and spin-orbit splitting are called relativistic effect. With increase in the size of the atom, especially for the atoms belonging to the 4d transition metal or heavier, the relativistic effect plays significant role in stabilizing the systems. Some relativistics effects can be included quite easily using relativistic pseudo potential through "Effective core potential (ECP)" functional. There are numbers of pseudopotentials available, among them LANL2DZ-ECP (Hay and Wadt, 1985, Cao and Dolg, 2010) , Stuttgart-ECP (Cao and Dolg, 2010) are most popular. LANL2DZ basis use effective core for all atoms larger than Neon, and are most widely used in analyzing the heavy elements in the periodic table.

2.16. Applications of QM to Study the Properties of the Biomolecular Systems

Quantum mechanics is the best tools to study the hydrogen bond interactions in different biological macromolecules as well different organic molecules, since hydrogen bond is also partially electrostatic in nature, so methods which include electron correlation and exchange are going to give good classifications of hydrogen bonds. Optimization of the systems using HF/DFT or other higher order theory, employing

specific basis sets (which depend on the atoms present in the system), results in stable geometry, and the different physical and associated properties of a system are analyzed at the minimum energy configuration of the system. The flow chart describing the optimization process is given below.



2.16.1. Interaction Energy and Basis Set Superposition Error

Different types of hydrogen bond interactions can be studied in terms of their bond lengths and bond angles. Stability of a complex molecule can be best studied by calculating its interaction energy. Energy of formation of biomolecular complex can be calculated by the energy of the dimer and then subtracting the energy of the isolated monomers from the dimer. The interaction energy of the dimer XY can be calculated as

$$E_{int} = E_{XY} - E_{Xo} - E_{Yo} \quad \text{.....[2.29]}$$

Where E_{XY} denotes energy of the complex systems, and E_{Xo} and E_{Yo} denote the energy of the two subsystems in isolation. However the energy obtained by such approach is found to be overestimated than the true value. This discrepancy arises due to an error known as basis set superposition error (BSSE). The BSSE in general arises due to overlapping of the orbitals of the two subsystems. Inclusion of BSSE is needed in the computation of hydrogen bonded complex. This error arises as the basis functions of the complex formed by the H-bonded interactions of the monomers are different from those of the isolated monomers. The larger basis set used for the complex yields an energy which is comparatively lower than the sum of the energies of the isolated monomers, with their basis set. BSSE introduces a non-physical attraction between the two monomers. BSSE can be best estimated by the MOROKUMA method used in GAMESS-US (Schmidt et al., 1993) and Boys and Bernardi function counterpoise methods incorporated in Gaussian 09 (Frisch et al., 2009). One can compare the energy of XY with that of X, with extra basis functions provided by Y plus the energy of Y, with extra basis function provided by X. This method of correcting the energy by adding the extra basis function is known as counterpoise corrections. In other words, the energy of individual component X is calculated in the presence of ghost orbital of Y, and the energy of individual component of Y can be calculated in the presence of ghost orbital X, there are called ghost orbital as only basis functions are taken into the consideration, while they are not accompanied by real atom, nuclei, and electrons. The value of the basis set super position error (BSSE) is dependent on the basis set chosen.

So, equation 2.29 becomes

$$E_{int} = E_{XY} - E_{Xo} - E_{Yo} + BSSE \qquad \text{.....[2.30]}$$

after taking in account of the basis set superposition error.

2.16.2. Natural Bond Orbital Analysis

To study the amount of charge transfer and orbital interaction, Natural Bond orbital (NBO) provides greatest tools (Carpenter and Weinhold, 1988). It is somehow alternate to the Morokuma analysis, but while Morokuma analysis represents H-bond as electrostatic, this emphasizes on importance of nonclassical donor-acceptor interactions. The NBO analysis includes all possible interactions between donor Lewis-type NBOs and acceptor non-Lewis NBOs. Energy transfers from the donor bonding orbital to the acceptor anti-bonding orbital were estimated by second-order perturbation theory to quantify the extent of hydrogen bond formation. Electronic properties, like occupancy of the natural orbitals, energy values for charge transfer, natural charges for the relevant atoms and bonds, which are involved in hydrogen bonded interactions are best studied by this method.

2.16.3. Atom in Molecule (AIM) analysis

The quantum theory of atoms in molecules (QTAIM) has been developed by Prof. Richard F. W. Bader and his groups (Bader, 1991). AIM characterize the chemical bonding of a molecule, which is in general based on the topological analysis of electron charge density $\rho(r)$ and Laplacian of the electron density and play crucial role in understanding the non-covalent interaction. Generally the characteristics features of covalent bond are when the electron density has large magnitude and Laplacian is found to be negative. The critical point in the electron density is the point in space, where the

first derivative of the electron density vanishes. One can discriminate between local minimum/maximum, saddle point by considering the second derivative of the electron density. In general there are four types of stable critical points are observed in the chemical structure of an element, (3, -3), (3, -1), (3, +1), (3, +3). Among these (3, -1) is known as bond critical point, and the value of electron density at the bond critical point can be characterized to determine the strength of the bond.

AIM can be best used to characterize the strengths of the hydrogen bonds between pair of atoms can be analyzed in terms of their electron densities (ρ_c) and their Laplacians ($\nabla^2\rho_c$) at the bond critical point (BCP). So also the associative energy parameters, like total electronic energy at the BCP (H_c), total kinetic energy at the BCP (G_c) and local potential energy at the BCP (V_c), also contributes in differentiating the types of hydrogen bonds involved. One can apply the topological parameters following equations 2.31 and 2.32, which relate the energy parameters and the Laplacian of electron density at the bond critical point ($\nabla^2\rho_c$).

$$2G_c+V_c = (1/4)\ \nabla^2\rho_c \qquad \text{..... [2.31]}$$

$$H_c = V_c+G_c \qquad \text{..... [2.32]}$$

Depending on the sign of $\nabla^2\rho_c$ and H_c, one can differentiate strong, medium and weak hydrogen bonds (Rozas et al., 2000). For weak hydrogen bond both $\nabla^2\rho_c$ and H_c are found to be > 0, for medium hydrogen bond $\nabla^2\rho_c$ > 0, and H_c <0, while for strong hydrogen bond $\nabla^2\rho_c$ and H_c < 0. I have employed the above approaches throughout my thesis work to analyze the strengths of hydrogen bonds in different biomolecules.

Aim and Scope of This Study

With the knowledge on structures of different biological macromolecules and the development of theoretical methods to understand their dynamics and interaction between themselves, it is high time now to follow up those studies. The work presented in this thesis mainly deals with study of different properties of biomolecules and nanomaterials by quantum chemical approach. The 3rd chapter deals with study of hydrogen bonding interaction between canonical and non-canonical RNA basepairs. I have classified the 3rd chapter into three sub-sections. The objectives of this study are as follows

(i) To analyze the structures and dynamics of frequently observed canonical and non-canonical basepairs in RNA by different computational chemistry methods

(ii) To study the conformational changes occurring in the RNA basepairs upon optimization by different quantum chemical methods.

(iii) I have modeled the hypothetical but possible basepairs and try to study their properties.

(iv) I have developed a new RNA database, "RNA non-canonical basepair database", which includes structures and enegetics of every possible type of basepairing pattern in RNA.

The relevance of this study lies in understanding the stability of the different basepairs and their potential role in the structure and functions of RNA.

In the 4th chapter, I have explained the importance of hydrogen bonding interaction in stabilizing goldnano-quercetin complex, which can be used in the therapeutic applications. I have attempted to study the following issues.

(i) Importance of capping agents in goldnano particle

(ii) Mode of interaction of quercetin with the citrate capped gold nanomaterials

This study can help in understanding the actually geometry and bonding of the gold citrate complex and its complexation with quercetin. My theoretical results are well correlated with the experimental finding of my collaborators.

Graphene is emerging as a new class of nano material with unique physical and chemical properties, with appearance of distinct electronic properties at the edges. Wetting property of different edges of the graphene has been discussed in the 5^{th} chapter, along with sufficient experimental evidences. I have addressed the explanations to the following queries.

(i) How the different edges interact with water?

(ii) Which edge of graphene is more hydrophilic?

(iii) What kind of interactions is observed in maintaining the stable graphene water complex?

Chapter 6 has three sub-sections; I have studied interaction of graphene and carbon nanotubes, with different nucleobases and nucleotides. I have tried to address the following issues.

(i) Inclusion of dispersion corrected density functional method (DFT-D) in optimizing the graphene_nucleobases/CNT_nucleobases and graphene_nucleosides/ CNT_nucleoside system.

(ii) Understanding change in the structural features of the graphene sheet

(iii) Analyze the interactions observed between graphene and CNT complex systems.

In the last chapter, chapter 7, I have summarized my present thesis works and also highlight future prospective which can be carried out for better understandings of the biomolecules and nano materials.

These studies appear helpful in understanding the importance of non-canonical basepairs in RNA. The quantum chemical analysis of nanomaterials such as graphene, carbon nanotube, gold clusters may helpful in designing nanoparticle based drug delivery vehicle.

Chapter 3

Theoretical Analysis of RNA basepairs:

Quantum Chemical Approach

Section 3.1

Structure, Stability and Dynamics of Canonical and Non-Canonical Basepairs: Quantum Chemical Studies

3.1.1. Introduction

Ribonucleic acid (RNA), the most important biomolecule in living organism, are the integral components of cellular machinery, plays important role in protein synthesis, transport, as well have many regulatory and catalytic functions. Structure and dynamics of macromolecules like RNA are influenced by a varieties of contributions, among them hydrogen bonding interaction between bases being one of the most important. Other relevant forces are arising from base-base stacking interaction, base-backbone or backbone-backbone interactions etc. In general, the regular stretches of A-form RNA are often defined by contiguous canonical Watson-Crick basepairs, formed through hydrogen bonds between complementary bases adenine-uracil, guanine-cytosine and sometimes guanine-uracil (Wobble basepair), which are also the major structural motifs of various functional RNAs such as ribosome, ribozyme, tRNA etc (Yang et al., 2003). In addition to Watson-Crick basepairs, a large numbers of non-Watson-Crick basepairs are also observed in RNA. These non-canonical basepairs play important role in maintaining the secondary and tertiary structures of RNA. Non-Watson-Crick basepairs often utilize the H-bonding capabilities of the ribose 2' hydroxyl group, which forms the highly specific fundamental biocatalytic functions. Each base can interact with the other base to form hydrogen bonds through three edges (i) Watson-Crick edge, (ii) Hoogsteen or "C-H" edge and (iii) sugar edge. Two nucleobases in RNA can interact with each other through any of these three edges in either cis or trans orientation with respect to the hydrogen bonds of the sugar rings. We defined a shorthand nomenclature to indicate the base pairing edge as well as orientation and have explained that in Table 3.1.1 (Das et al., 2006). This leads to a total of 12 different base edges families with 168 possible

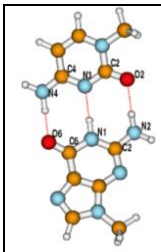
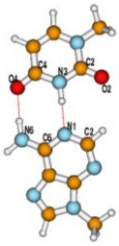
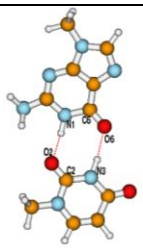
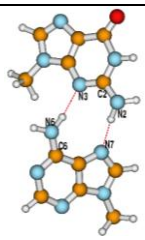
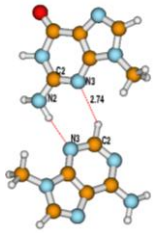
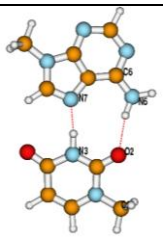
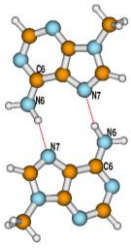
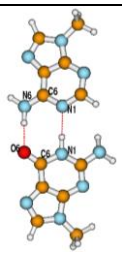
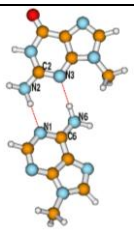
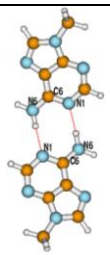
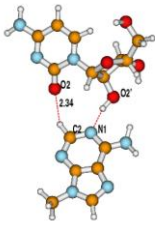
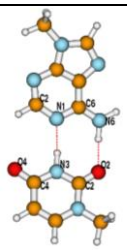
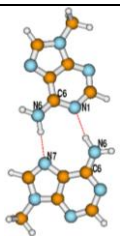
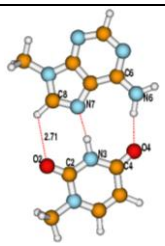
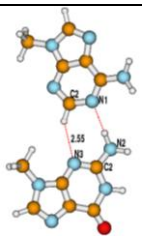
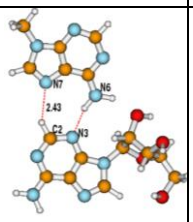
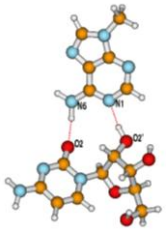
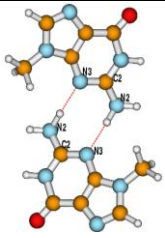
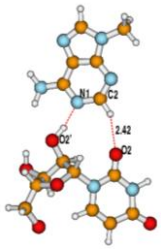
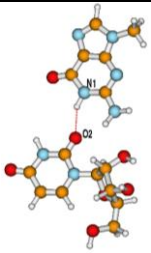
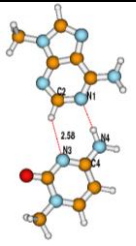
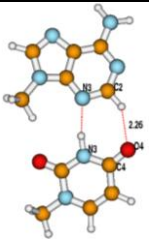
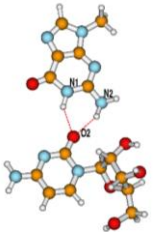
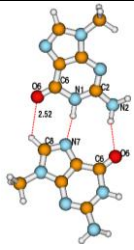
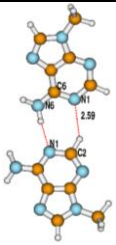
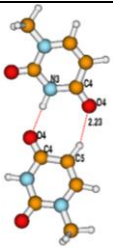
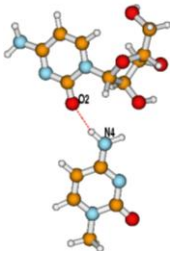
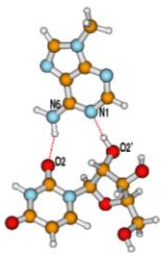
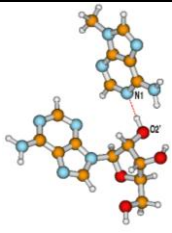
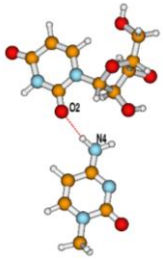
basepairing patterns (Leontis and Westhof, 1998). Besides, water mediated and protonated basepairs have also been identified recently (Leontis et al., 2002) and these perhaps give rise to extra stability to the folded three-dimensional structures. BPFIND (Das et al., 2006) detects all canonical and non-canonical basepairs occurring in nucleic acid three-dimensional structures only when the basepairs are bonded by two hydrogen bonds through any edges. The direct hydrogen bonds can be of polar (N-H…O/N) or non-polar (C-H…O/N) types. Though the C-H…O/N mediated basepairs involve intrinsically weak non-polar force and very often assume non-planar geometry but many of them still contribute substantially to the RNA structural stability as they are seen to occur quite frequently in RNA crystal structures (Das et al., 2006). The basepairs involving hydrogen bond to the sugar residues have additional flexibility due to single covalent glycosidic bond and can be non-planar even with perfectly linear hydrogen bonds. A similar detection criterion is also followed by other method, which however considers basepairs mediated by single H-bond also (Yang et al., 2003, Huang et al., 2003). Since a biomolecular structure is ultimately governed by free energy changes, a thorough understanding of structure-energy relationship in RNA non-WC basepairs is essential. Among the non-canonical basepairs the most frequent ones are trans A:G Hoogsteen/Sugar Edge (H:S T) and trans A:U Hoogsteen/WC (H:W T), whose frequencies of occurrences are comparable to that of the wobble G:U basepair (W:W C). These high frequencies of occurrences of non-canonical basepairs certainly indicate that these have an important role in the stability of structural RNA. They stabilize tertiary contacts between remote segments of the RNA macromolecules forming specific RNA motifs. They are frequently involved in the formation of internal RNA loops and specific

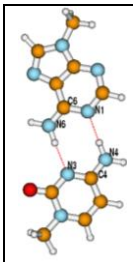
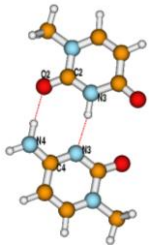
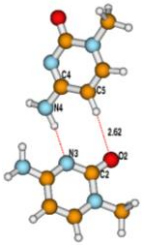
segments containing several contiguous non-WC basepairs. These non-canonical basepairs often occur within secondary structural blocks and play important role in the evolution and folding of RNA architectures and in building up the complex three-dimensional structures of RNAs (Hendrix et al., 2005). Attempts have been made by different groups to classify the non-canonical basepairs and detecting their potential role in maintaining the three dimensional structure of RNA. Leontis and Westhof have characterized the non-canonical basepairs in terms of their isosteric parameters (Leontis and Westhof, 1998) and was found useful in modeling RNA structures from sequence conservation. The dynamics of the basepairs can be studied in terms of three-rotational and three-translational parameter. It was observed that most of the basepairs in cis orientation have large negative propeller twist, which is nearly zero for the trans basepairs (Mukherjee et al., 2006). The basepairs involving sugar edges were generally found to be non-planar, presumably to avoid steric clash involving bulky sugar moiety. Study of the frequently occurring basepairs may highlight some new aspects of structural properties of the basepairs and their biological relevance.

Different papers have been devoted on quantum chemical energy calculation of the canonical and non-canonical basepairs (Sponer et al., 2004, Sponer et al., 2005b, Sponer et al., 2005c, Sponer et al., 2007, Bhattacharyya et al., 2007), however there are not much efforts in systematically characterizing H-bonding stabilities of all frequently observed basepairs in RNA. In this present study, we have considered 33 basepairs (as shown in Figure 3.1.1) which occurred frequently in RNA crystal structures. We have carried out geometry optimization of all the basepairs by using Density functional theory (DFT) with hybrid B3LYP functional and quantitatively analyzed the alteration of

structures during optimization. We have determined hydrogen bonding strengths corresponding to each type of H-bond, such as N-H…O, N-H…N, O-H…N etc, using a statistical procedure. The basepairs considered are classified in to three types (i) basepairs stabilized by multiple polar H-bonds (ii) basepairs stabilized by one polar and one non-polar C-H...O/N type of interaction (iii) basepairs involving the sugar O2' or H2'. We have analyzed structure and dynamics in 33 basepairs, where direct H-bond exists between the bases through different approaches such as changes in calculated infrared (IR) frequency and intensity. We have also attempted to understand basepair dynamics for these 33 basepairs by calculating intrinsic flexibility of each of the basepairs from the normal mode frequencies and calculated standard deviations of the intra basepair parameters, viz. propeller, buckle, open-angle, stagger, shear and stretch. These standard deviation values are in excellent agreement with those obtained from crystal data (Mukherjee et al., 2006). The analysis of basepair dynamics indicates that movements of most basepairs are highly favorable along their propeller or buckle directions. Such movements maintain hydrogen-bonding strength by generally modifying pyramidalization of the H-bond forming amino groups. The other movements along open, shear and stretch are more hindered as they are associated with distortion of the H-bonds between the bases. The results also indicate that the basepairs stabilized by two or more polar H-bonds are stable with equivalent strength whereas those having C-H...O/N are rather weak.

Figure 3.1.1. Structures of the 33 basepairs optimized by B3LYP/6-31G** basis set

					
1. G:C W:W C	2. A:U W:W C	3. G:U W:W C	4. A:G H:S T	5. A:G s:s T	6. A:U H:W T
					
7. A:A H:H T	8. G:A W:W C	9. G:A S:W T	10. A:A W:W T	11. A:C w:s T	12. A:U W:W T
					
13. A:A H:W T	14. A:U H:W C	15. A:G w:s C	16.A:A h:s T	17. A:C w:s C	18. G:G S:S T
					
19. A:U w:s T	20. G:U w:s C	21. A:C w:w C	22. A:U s:w C	23. G:C w:s C	24. G:G H:W T
					
25. A:A w:w C	26. U:U h:w T	27. C:C h:s T	28. A:U w:s C	29.A:A s;s T	30. C:U h:s T

					
31. A:C W:W T	32. C:U W:W T	33. C:C w:h C			

3.1.2. Choice of the Basepairs

We have selected 33 frequent basepairs of usual Watson-Crick type as well as non-canonical type (Table 3.1.1) from crystal structures of RNA as indicated in our website http://www.saha.ac.in/biop/bioinformatics.html. These basepairs were detected by BPFIND software with considerable frequency in 145 functional RNA crystal structures. These structures are selected from the Protein data bank solved by X-ray crystallography at 3.5 Å or better resolution as available in September 2004. Structures having more than 30 nucleotides are considered in this database. Proton mediated basepairs are not considered in our present study. Basepairs stabilized by two N-H...N/O type of hydrogen bonding are denoted by 'W' (Watson-Crick), 'H' (Hoogsteen) or 'S' (sugar) and those stabilized by one C-H…N/O types hydrogen bonding are denoted by lower case letters 'w', 'h' or 's'. These characters, i.e. W, H etc. are used to mention the edge of a base through which it forms H-bonds (Das et al., 2006). The last letter in basepair nomenclature indicates cis (C) or trans (T) orientation of the glycosidic bonds with respect to the virtual H-bonds.

3.1.3. Computational Details and Optimization

We used MOLDEN software (Schaftenaar and Noordik, 2000) for model building hydrogen atomic positions and graphical visualization of the RNA basepairs. We have replaced the sugar phosphate moieties by methyl groups for each of the 23 basepairs to avoid any artificial dipole moment arises due to the N-H bond, and this reduces the computational cost also. For the third variety of sugar mediated basepairs, we have considered sugars attached to one of the bases, terminating by -OH groups at 5' and 3' positions. Geometry optimization of the basepairs without any constraints was carried out using using GAMESS-US (Schmidt et al., 1993) with hybrid B3LYP functional (Lee et al., 1988, Becke, 1993) with 6-31G (2d, 2p) basis set. Basis set superposition errors (BSSE) were calculated by MOROKUMA method (Kitaura and Morokuma, 1976) using Hartree-Fock procedure (HF) with 6-31G (2d,2p) basis set for the fully relaxed geometry optimized structures. Normal mode analyses of the optimized structures were performed by numerical Hessian method.

The interaction energy ΔE of a dimer A…B is defined as the electronic energy difference between the dimer and isolated optimized monomers (E_A and E_B) where monomer energies are obtained assuming the geometries of the optimized dimer and using the basis set of the dimer. Mathematically this is given as

$$\Delta E = E_{XY}^{opt} - E_X^{opt} - E_Y^{opt} \qquad \text{.....[3.1.1]}$$

Where E_{XY}^{opt} is the energy of the optimized basepair and E_X^{opt} and E_Y^{opt} are the energy of the bases X, Y in the optimized basepair geometry. We also calculated the deformation energy, which is a repulsive contribution due to changes of monomer geometries upon

complex formation and is defined, as the energy required in deforming the isolated and optimized bases to the geometry they assume in the basepair. The deformation energy is hence calculated as

$$E_{def} = (E_X^{opt} - E_X^{o}) + (E_Y^{opt} - E_Y^{o}) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [3.1.2]$$

where E_X^{o} and E_Y^{o} are the energy of the optimized isolated bases. Taking into consideration of the basis set superposition error (BSSE) and deformation energy the total interaction energy of a basepair is given by

$$\Delta E_{cor} = E_{XY}^{opt} - (E_X^{opt} + E_Y^{opt}) + BSSE + E_{def.} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots [3.1.3]$$

Putting the expression for deformation energy in this equation the total interaction energy of a basepair stands out as

$$\Delta E_{cor} = E_{XY}^{opt} - E_X^{o} - E_Y^{o} + BSSE. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots [3.1.4]$$

Intra basepair parameters, viz. buckle, open-angle, propeller, stagger, shear and stretch, were calculated using the NUPARM software (Mukherjee et al., 2006) version 2.0.

Table 3.1.1. The basepairs studied along with their source and frequency of occurrence in RNA structure database, sorted according to frequency.

No	Basepair	Leontis/Westhof Nomenclature	Frequency	PDB ID and residue names	Environment
1	G:C W:W C	Cis Watson-Crick/Watson-Crick (canonical)	21593	1DFU 97 (M)-79 (N)	Double helical
2	A:U W:W C	Cis Watson-Crick/Watson-Crick (canonical)	6862	1ASY 607 (S)- 666 (S)	Double helical
3	G:U W:W C	Cis Watson-Crick/Watson-Crick (wobble)	2749	1ASY 610 (S)-625 (S)	Double helical
4	A:G H:S T	Trans Hoogsteen/Sugar Edge	2303	1MZP 21 (B)-44 (B)	Double helical; Triplet
5	A:G s:s T	Trans Sugar Edge/Sugar Edge	1496	1FFK 243 (O)-274 (O)	Triplet
6	A:U H:W T	Trans Hoogsteen/Watson-Crick	1164	1ASY 614 (S)-608 (S)	Isolated

7	A:A H:H T	Trans Hoogsteen/Hoogsteen	444	1ASY 609 (R)- 623 (R)	Triplet
8	G:A W:W C	Cis Watson-Crick/Watson-Crick	411	1QVG 2596 (O)-2582 (O)	Double helical
9	G:A S:W T	Trans Sugar Edge/Watson-Crick	335	1FFK 166 (O)-924 (O)	Triplet
10	A:A W:W T	Trans Watson-Crick/Watson-Crick	253	1QVG 2566 (O)-2699 (O)	Parallel helix
11	A:C w:s T	Trans Watson-Crick/Sugar Edge	214	1HNW 696(A)-797(A)	Quadruple
12	A:U W:W T	Trans Watson-Crick/Watson-Crick	210	1ASZ 615 (R)-648 (R)	Isolated
13	A:A H:W T	Trans Hoogsteen/Watson-Crick	205	1FFK 460 (O)-455 (O)	Triplet.
14	A:U H:W C	Cis Hoogsteen/Watson-Crick	202	1FFZ 2470 (A)-2277 (A)	Triplet
15	A:G: w:s C	Cis Watson-Crick/Sugar Edge	180	1SM1 991 (O)-2020 (O)	Triplet
16	A:A h:s T	Trans Hoogsteen/Sugar Edge	170	1GTR 13(B)-22(B)	Triplet
17	A:C w:s C	Cis Watson-Crick/Sugar Edge	134	1HNX 1280(A)-1149(A)	Triplet
18	G:G S:S T	Trans Sugar Edge/Sugar Edge	133	1FFK 315 (O)-336 (O)	Isolated
19	A:U w:s T	Trans Watson-Crick/Sugar Edge	107	4TNA 8-21	Triplet
20	G:U w:s C	Cis Watson-Crick/Sugar Edge	89	1S72 196(O)-425(O)	Quadruple
21	A:C w:w C	Cis Watson-Crick/Watson-Crick	81	1M5K 20 (B)-63 (B)	Double helical
22	A:U s:w C	Cis Sugar Edge/Watson-Crick	76	1M1K 329 (A)-346 (O)	Double helical
23	G:C w:s C	Cis Watson-Crick/Sugar Edge	66	1FFK 1302(O)-1353(O)	Double helix
24	G:G H:W T	Trans Hoogsteen/Watson-Crick	65	1QVG 868 (O)-775 (O)	Triplet
25	A:A w:w C	Cis Watson-Crick/Watson-Crick	63	1FFK 635 (O)-620 (O)	Double helical
26	U:U h:w T	Trans CHO Edge/Watson-Crick	49	1W2B 2781 (O)- 2791 (O)	Parallel helix
27	C:C h:s T	Trans Hoogsteen/Sugar Edge	56	1JBR 8(C)-22(F)	Double helix
28	A:U w:s C	Cis Watson-Crick/Sugar Edge	52	1F1T 11(A)-26(A)	Isolated
29	A:A s:s T	Trans Sugar Edge/Sugar Edge	52	1HNW 50(A)-360(A)	Partial Stacking
30	C:U h:s T	Trans Hoogsteen/Sugar Edge	44	1S72 2690(O)-2704(O)	Double helix
31	A:C W:W T	Trans Watson-Crick/Watson-Crick	34	1QVG 1742 (O)-2037 (O)	Double helical
32	C:U W:W T	Trans Watson-Crick/Watson-Crick	19	1JJ2 1394 (O)-1432 (O)	Isolated
33	C:C w:h C	Cis Watson-Crick/CHO-Edge	18	1Q82 1176 (A)-1196 (A)	Isolated

3.1.4. Results

3.1.4.1. Analysis of the Interaction Energy

Optimized structures of all the 33 selected Watson-Crick as well as non-canonical basepairs are shown in Figure 3.1.1. Interaction energy of the analyzed basepairs in RNA is found to vary between -5 to -27 kcal/mol (Table 3.1.2), indicating wide range of stabilization provided by these frequently occurring basepairs. It is observed that basepairs stabilized by C-H mediated H-bond are less favorable (ΔE_{corr} > -10 kcal/mol) whereas all the basepairs with two polar H-bonds, except few (G:G S:S T; A:G H:S T; A:A H:H T) are highly favorable (ΔE_{corr} < –10 kcal/mol). Compared to all other basepairs, the G:C basepair is stabilized maximally and this also explain their significantly higher frequency of occurrence in any RNA structure. Canonical A:U basepair and G:U wobble basepair have interaction energy around –15 kcal/mol, which is comparable to the interaction energy of many non-canonical basepairs. The interaction energy of G:C W:W C and A:T W:W C basepairs were reported by Sponer et al. (Sponer et al., 2004) as –27.5 kcal/mol and –15.0 kcal/mol respectively whereas these were calculated as –25.2 kcal/mol for G:C W:W C basepair and –12.2 kcal/mol for A:T W:W C basepair (Mo, 2006). From our results it is shown that interaction energy of G:C W:W C basepair is –26.51 kcal/mol whereas for A:U W:W C basepair it is -14.43 kcal/mol which are in good agreement with the previous reports. It should be noted that both these groups calculated the interaction energy of basepairs by considering the bases only and did not consider sugar part or replaced sugar part by a methyl group. On the other hand we have replaced sugar part of a nucleotide by a methyl group (to reduce artificial polarity of the bases) and thus we have taken into consideration of two possible relative

orientations of basepair with respect to sugar moiety. It is further noted that interaction energy is not always related to the frequency of occurrence although G:C W:W C has highest frequency with best stabilization energy. The 24th basepair G:G H:W T is very stable (ΔE = -17.58 kcal/mol) but rarely observed basepair. On the other hand the A:G s:s T basepair, even with small interaction energy (ΔE = -5.71 kcal/mol), is observed very frequently. This higher frequency might be due to possibility of formation of a third H-bond involving 2-amino group of guanine with the sugar moiety of adenine. Deformation energy of these basepairs has positive values except in few cases, which have negligibly small negative value. Surprisingly, the basepairs having very small deformation energy component are all having one non-polar interaction. This, possibly indicate that only one strong H-bond does not significantly alter structures of the optimized bases, which multiple H-bonds may do. Deformation energies of the bases with sugar are generally higher, indicating more variability of the sugar part.

In general most of the basepairs having multiple polar H-bonds have high interaction energy ranging from –10 kcal/mol to –27 kcal/mol. Few basepairs with two polar H-bonds namely A:G H:S T (4th basepair), A:A H:H T (7th basepair) and G:G S:S T (18th basepair) are found to have weaker interaction energy as exception. The modes of interactions for these basepairs are found to be only a pair of N-H…H type hydrogen bonds. The basepairs stabilized by one polar and one non-polar interaction are all found to have interaction energies ranging between –5.71 kcal/mol (for A:G s:s T) to –9.43 kcal/mol (for A:U s:w C). Optimized structures of few basepairs involving sugar have single H-bond and as a result even smaller interaction energy.

Table 3.1.2. Interaction energy components (kcal/mol) of the basepairs

No.	Basepair	ΔE_{cor}	ΔE	E_{def}
1	G:C W:W C	-26.51	-33.80	3.61
2	A:U W:W C	-14.43	-18.29	1.02
3	G:U W:W C	-14.68	-19.03	1.73
4	A:G H:S T	-9.50	-13.29	1.51
5	A:G s:s T	-5.71	-7.79	0.29
6	A:U H:W T	-13.31	-17.79	1.53
7	A:A H:H T	-9.76	-12.69	0.97
8	G:A W:W C	-15.00	-20.05	2.13
9	G:A S:W T	-10.61	-14.22	1.33
10	A:A W:W T	-12.10	-14.88	0.66
11	A:C w:s T	-8.76	-14.32	2.60
12	A:U W:W T	-12.71	-17.25	1.39
13	A:A H:W T	-11.28	-14.09	0.82
14	A:U H:W C	-14.21	-18.66	1.47
15	A:G w:s C	-6.02	-7.78	0.18
16	A:A h:s T	-7.41	-28.27	17.96
17	A:C w:s C	-15.48	-21.34	2.70
18	G:G S:S T	-8.12	-12.06	1.79
19	A:U w:s T	-13.13	-43.56	28.00
20	G:U w:s C	-9.69	-12.48	0.55
21	A:C w:w C	-6.28	-8.44	0.37
22	A:U s:w C	-9.43	-12.37	0.10
23	G:C w:s C	-8.13	-15.99	5.28
24	G:G H:W T	-17.58	-21.38	1.45
25	A:A w:w C	-6.07	-7.71	-0.11
26	U:U h:w T	-9.33	-11.58	-0.12
27	C:C h:s T	-5.22	-11.72	4.55
28	A:U w:s C	-16.33	-19.35	0.03
30	C:U h:s T	-7.40	-9.49	0.16
31	A:C W:W T	-14.32	-18.19	1.61
32	C:U W:W T	-10.44	-14.79	1.22
33	C:C w:h C	-8.28	-11.69	0.92

* The components, namely $\Delta E_{cor,}$ ΔE and E_{def} are calculated using eqn 3.1.4, 3.1.1 and 3.1.2 respectively

3.1.4.2. Structures of the Basepairs in terms of Relative Orientation Parameters

The different orientation parameters (Dickerson, 1989) of the basepairs have been well described in chapter1. Among these six-parameters buckle, propeller and stagger describe the overall non-planarity of a basepair while open-angle, shear and stretch are

related to the hydrogen bonding pattern and proximity. The flexibility and dynamics of the basepairs can be easily visualized from the parameter values. The structural intra-base-pair parameters of these base pairs in crystal geometry as well as energy minimized geometry, as calculated by NUPARM 2.0, are shown in Table 3.1.3.

The first three canonical (W:W C type) basepairs, A:U, G:C and G:U adopt near planar geometry in their optimized geometry configuration. Previous studies of Watson-Crick A:T or G:C basepairs by Hartee-Fock method also reported planar structures of the basepairs (Sponer et al., 1996, Gould and Kollman, 1994, Basu et al., 2005, Brameld et al., 1997) whereas non-planar structure (Ogawa et al., 2003, Danilov and Anisimov, 2005) was detected by MP2/6-31G (d) or MP/-6-31G (d,p) methods. However, quantitative analyses of the basepairs non-planarity were not reported earlier.

Table 3.1.3. Intra basepair parameter of the basepairs in crystals and optimized geometries. First line for each entry gives parameter values in crystal geometry while the second line gives those in optimized geometry

No	Basepair	Buckle (°)	Open angle (°)	Propeller (°)	Stagger (Å)	Shear (Å)	Stretch (Å)
1	G:C W:W C	-1.04	0.45	-2.54	-0.35	0.29	2.85
		-0.3	-3.43	2.21	0.12	0.17	2.92
2	A:U W:W C	7.52	0.27	-9.33	0.33	-0.13	2.79
		-0.98	0.03	0.35	0.00	0.09	2.86
3	G:U W:W C	-1.85	-2.38	-2.52	0.05	-2.31	2.72
		3.94	4.11	-1.30	-0.15	-2.42	2.84
4	A:G H:S T	-8.52	-7.45	-12.13	-0.16	2.14	3.29
		6.42	-20.31	-40.88	0.05	1.95	3.24
5	A:G s:s T	-13.64	30.40	-8.76	0.05	1.57	3.12
		1.52	23.45	-33.6	-0.90	1.38	3.20
6	A:U H:W T	14.00	1.95	4.88	0.27	0.04	2.71
		-3.20	7.49	1.91	0.00	0.15	2.79
7	A:A H:H T	3.34	-3.81	-6.87	0.46	2.66	2.86
		-0.06	- 0.06	-32.08	0.01	2.67	2.88
8	G:A W:W C	3.69	-0.72	-9.66	0.09	0.61	2.78
		6.14	-2.93	-12.75	-0.21	-0.01	2.94
9	G:A S:W T	16.46	-12.22	-33.27	- 0.33	1.71	3.28

		6.74	-18.41	-17.40	-0.07	1.82	3.32
10	A:A W:W T	15.42	-1.67	7.20	-0.37	2.87	2.88
		4.40	0.00	-12.06	0.00	2.27	2.98
11	C:A s:w T	45.45	27.17	-20.56	-0.76	-0.18	3.50
		20.24	37.07	-5.81	1.29	0.49	3.66
12	A:U W:W T	-7.28	-0.91	4.10	-0.15	-0.1 5	2.88
		-0.67	-7. 96	-0.57	0.03	-0.06	2.81
13	A:A H:W T	1.02	6.22	21.72	1.04	2.55	2.77
		-1.40	0.95	17.58	0.00	2.50	2.94
14	A:U H:W C	-3.17	-2.51	-18.60	0.34	-0.23	2.82
		-0.24	2.23	-2.63	-0.01	-0.13	2.79
15	A:G w:s C	-8.41	43.94	-42.52	0.55	1.94	2.78
		-4.35	34.47	-16.38	0.04	2.23	3.07
16;	A:A h:s T	-6.06	-3.64	2.18	-0.14	2.28	2.85
		-6.17	-3.60	2.13	0.13	2.28	2.85
17	A:C w:s C	1.16	-32.88	-20.09	1.02	-0.61	4.51
		41.62	-37.10	-10.15	-1.21	-0.86	4.21
18	G:G S:S T	-19.42	-0.87	27.07	0.38	1.72	3.53
		-12.95	0.08	30.85	0.00	1.17	3.53
19	A:U w:s T	5.77	-33.38	-15.74	0.31	-0.51	4.09
		-1.87	-42.52	-16.09	-1.19	0.28	4.04
20	G:U w:s C	23.36	-27.57	3.18	-0.84	2.44	3.71
		-7.15	-32.74	-4.68	0.03	4.15	3.26
21	A:C w:w C	-19.67	0.02	-7.73	-0.67	3.00	2.39
		-4.42	13.55	-4.84	-0.01	2.60	2.70
22	A:U s:w C	31.27	-6.37	-4.82	-0.15	0.03	3.01
		2.34	-6.63	2.83	0.03	0.16	2.98
23	G:C w:s C	12.18	-32.23	-15.72	-0.24	3.55	3.26
		-8.60	-36.80	4.47	0.03	4.35	3.18
24	G:G H:W T	24.14	-7.80	-2.81	-0.40	0.17	2.75
		10.26	-9.91	3.48	0.03	0.17	2.85
25	A:A w:w C	10.15	2.24	-15.67	-0.76	2.58	2.71
		8.96	14.47	-0.22	0.09	2.50	2.75
26	U:U h:w T	4.87	14.60	-1.88	- 0.11	2.29	2.94
		-2.05	9.07	2.52	0.01	2.70	3.04
27	C:C h:s T	-10.01	-9.87	-7.22	0.46	4.53	2.57
		-10.96	-12.74	-13.11	-0.15	5.19	2.89
28	A:U w:s C	2.90	-27.86	-4.84	0.47	-0.26	4.16
		-45.45	-30.68	-20.69	-1.29	0.56	4.20
29	A:A s:s T	-66.22	14.01	0.67	1.88	0.98	2.50
		-62.28	41.18	-43.59	-0.09	-2.29	3.02
30	C:U h:s T	1.21	-3.39	-1.07	0.32	5.00	2.94
		-11.39	-13.50	-8.16	-1.05	5.15	2.83
31	A:C W:W T	6.06	-1.63	-29.72	-0.01	2.61	2.97
		-1.10	-0.78	-1.93	0.03	2.33	2.94
32	C:U W:W T	-9.70	1.35	-17.55	0.16	-0.35	3.01
		-0.83	1.95	-15.81	0.29	-0.25	2.95
33	C:C w:h C	-13.94	-17.69	-6.66	0.12	- 1.92	3.30
		-5.13	-13.06	-31.91	-0.32	-1.86	3.26

Our analysis suggests various degree of non-planarity of the basepairs by DFT theory, which incorporates both electron exchange and electron correlation in its functional. The planar orientation of the canonical basepairs may be due to neglect of the sugar moiety and stacking between neighboring basepairs in the crystal environment (Table 3.1.1). Most of the non-canonical basepairs also tend to adopt near planar orientation with perfect hydrogen bond geometry (Table 3.1.3). The basepairs interacting through sugar show sufficient non-planarity in most cases.

Stagger values of most of the basepair become zero or close to zero on optimization. This clearly indicates that most of the basepairs have a tendency to assume planer geometry on optimization as the influence of stacked basepairs disappeared during their optimizations. The value of open-angle, shear and stretch, which describe H-bonding pattern, mostly retain their initial conformation. These values are also close to the corresponding average values obtained from analysis of 145 crystal structures of functional RNA (Mukherjee et al., 2006). This possibly justifies choice of the initial structures from a big pool of similar conformations. This further indicates that energy minimization did not perturb the hydrogen bond geometry of the initial structures, which was reported to occur earlier for some basepairs with single or poor H-bonds (Bhattacharyya et al., 2007, Sponer et al., 2005b).

Basepairs with H-bonds involving primary amino group (N2 of Guanine, N4 of Cytocine and N6 of Adenine) tend to adopt large propeller due to pyramidalization of the amino groups, while those having interaction through secondary amino group (N3 of Uracil, N1 of Guanine) are more planar with small value of propeller twist. Our previous study (Mukherjee et al., 2005) of different amino group containing biomolecular systems

by different quantum chemical methods, namely MP2/6-31G (2d,2p), HF/6-31G (2d,2p), HF-6-311G (2d,2p) or HF/6-311+G(2d, 2p) also suggested larger pyramidalization of the primary amino groups as compared to secondary amino group. It is seen that all the basepairs stabilized by two primary amino groups, such as N6 of adenine, N4 of cytosine or N2 of guanine have large propeller twist. Examples of such basepairs are A:G H:S T (4^{th}), A:A H:H T (7^{th}), G:A S:W T (9^{th}), A:A W:W T (10^{th}), A:A H:W T (13^{th}) and G:G S:S T (18^{th}). On the other hand H-bond involving a secondary amino group tends to reduce the propeller twist in G:C W:W C (1^{st}), A:U W:W C (2^{nd}), G:U W:W C (3^{rd}), A:U H:W T (6^{th}), A:U W:W T (12^{th}), A:U H:W C (14^{th}) and G:G H:W T (24^{th}). It may also be noted that it was reported earlier that these pyramidalization of primary amino groups are their inherent tendency and possibility of a weak intramolecular hydrogen bond also cannot make them planer (Sen et al., 2006), which is also observed in most of the basepairs having N-H…N type of interaction. The canonical basepairs, on the other hand, have strong H-bonding between primary amino groups and oxygen as acceptor. Such strong H-bonds may be sufficient to reduce pyramidalization of the primary amino groups.

Some basepairs undergo considerable movement during optimization. We tried to understand the reasons behind these changes and found that these basepairs were either (a) stabilized by additional sugar mediated hydrogen bonding, (b) stacking interaction with neighboring basepairs of the folded RNA macromolecules as in A:A W:W T (10^{th}), A:C w:w C (21^{st}), A:U s:w C (22^{nd}), A:A w:w C (25^{th}) and A:C W:W T (31^{st}) basepairs or (c) by other stabilizing interactions, such as base triple formation as in A:G H:S T (4^{th}), A:A H:H T (7^{th}), G:A S:W T (9^{th}), A:A H:W T (13^{th}), A:U H:W C (14^{th}), A:G w:s C

(15th), A:G s:s T (5th) and G:G H:W T (24th) basepairs, which were considered as basepairs only in our study. Few basepairs with one H-bond through sugar hydroxyl group and another C-H…O interaction, such as C:C h:s T and C:U h:s T (27th and 30th, respectively) optimize to structures with single H-bond, as the initial C-H…O interaction is too weak to sustain during optimization. In addition, there are crowding effect arising due to C1' or H1' atoms. The A:A s:s T (29th) basepair underwent considerable movement to w:s T orientation upon optimization as the initial structure had only two C-H…O interactions while the later is stabilized by a significantly stronger N-H…O H-bond. For this reason we have not analyzed its interaction energies. Such variation was also noted earlier (Sponer et al., 2005a). The basepairs with sugar edge, even without explicit sugars, remain non-planar mainly due to crowding effect involving methyl groups at C1' position, which is also the expected scenario in normal RNA where the bulky ribose sugar might have larger number of contacts. The last basepair (C:C w:h C) basepair adopts large propeller possibly due to weak C-H…O interaction as well as pyramidal amino group. The C:U W:W T (32nd) basepair retain reasonably large propeller possibly to avoid electrostatic repulsion between the two O2 groups of the two pyrimidines. The A:U H:W C and A:U H:W T basepairs tried to adopt a planer geometry with reduced propeller twist possibly to get additional stability by a C-H…O type interaction between O2 atom of uracil and H8 atom of adenine.

From Table 3.1.1 it is found that there are five basepairs in our set, which were isolated basepairs in crystal structure also. They were expected to retain similar structural parameters upon optimization and these are found to be true for trans A:U WC/WC (W:W T), trans G:G SE/SE (S:S T) and trans C:U WC/WC (W:W T) basepairs. As

indicated above the A:U H:W T basepair became flat after optimization, due to involvement of a secondary amino group in H-bonding and also to improve a C-H…O hydrogen bond distance. On the other hand C:C w:h C basepair, adopt a non-planar geometry upon optimization, presumably due to its extra flexibility.

3.1.4.3. Strengths of Different Types of H-Bonds

The most common types of hydrogen bonds observed in biomolecules are namely N-H...O, N-H...N, S-H...O, S-H...N etc. Recent analysis of crystal structures of organic or bimolecular fragments indicated that perhaps C-H…O or C-H...N forms another class of hydrogen bond (Desiraju and Steiner, 2001) and play significant role. Following these observations some other groups characterized these as blue shifting hydrogen bond having substantially smaller interaction energy (Vargas et al., 2000). However, such H-bonds may still be important for understanding ligand receptor interactions and hence in structure based drug design also. We have therefore attempted to estimate the strengths of the different types of H-bonds from our interaction energy values.

The selected basepairs have a number of H-bonds of wide variety, namely 20 N-H…O type, 29 N-H...N type, 4 O-H…N type, 10 C-H...O type and 5 C-H...N type and all the basepair interaction energies are at least sum of two such energy components.

$$E_{\text{int}}^{i} = n_i^{NHO} E^{NHO} + n_i^{NHN} E^{NHN} + n_i^{OHN} E^{OHN} + n_i^{CHO} E^{CHO} + n_i^{CHN} E^{CHN} + \delta_i \qquad \text{…….[3.1.5]}$$

Where n_i^{NHO}, n_i^{NHN}, n_i^{CHO} and n_i^{CHN} are number of hydrogen bonds of the corresponding type in a basepair and E^{NHO}, E^{NHN} etc. are the average interaction energies in forming a H-bond of that type. In most cases n_i^{X}, where X is any of N-H…O or N-H…N etc., are

one or zero while some basepairs have two N-H...O or N-H...N type of hydrogen bonds. As for example, in case of G:C W:W C basepair, n^{NHO} is 2, N^{NHN} is 1 while n^{CHO} and n^{CHN} are both zero. In order to determine the value of the energy components, viz E^{NHO}, E^{NHN} etc, we have followed a least squares fit like procedure, whereby minimizing $\Sigma\delta_i^2$ (eqn 3.1.5) by optimizing it with respect to E^{NHO}, E^{NHN} etc. Thus we differentiate eqn 3.1.6. by E^{NHO}. E^{NHN}, E^{CHO} and E^{CHN} and equate those to zero. This gives us a set of four linear simultaneous equations as shown in eqn 3.1.7.

$$\sum_i \delta_i^2 = \sum_i \left(E_{\text{int}}^i - n_i^{NHO}E^{NHO} - n_i^{NHN}E^{NHN} - n_i^{OHN}E^{OHN} - n_i^{CHO}E^{CHO} - n_i^{CHN}E^{CHN}\right)^2 \text{..[3.1.6]}$$

$$\begin{aligned}
&E^{NHO}\sum_i \left(n_i^{NHO}\right)^2 + E^{NHN}\sum_i n_i^{NHO}n_i^{NHN} + E^{OHN}\sum_i n_i^{NHO}n_i^{OHN} + E^{CHO}\sum_i n_i^{NHO}n_i^{CHO} + E^{CHN}\sum_i n_i^{NHO}n_i^{CHN} = \sum_i n_i^{NHO}E_{\text{int}}^i \\
&E^{NHO}\sum_i n_i^{NHO}n_i^{NHN} + E^{NHN}\sum_i \left(n_i^{NHN}\right)^2 + E^{OHN}\sum_i n_i^{NHN}n_i^{OHN} + E^{CHO}\sum_i n_i^{NHN}n_i^{CHO} + E^{CHN}\sum_i n_i^{NHN}n_i^{CHN} = \sum_i n_i^{NHN}E_{\text{int}}^i \\
&E^{NHO}\sum_i n_i^{NHO}n_i^{OHN} + E^{NHN}\sum_i n_i^{NHN}n_i^{OHN} + E^{OHN}\sum_i \left(n_i^{OHN}\right)^2 + E^{CHO}\sum_i n_i^{CHO}n_i^{OHN} + E^{CHN}\sum_i n_i^{CHN}n_i^{OHN} = \sum_i n_i^{OHN}E_{\text{int}}^i \\
&E^{NHO}\sum_i n_i^{NHO}n_i^{CHO} + E^{NHN}\sum_i n_i^{NHN}n_i^{CHO} + E^{OHN}\sum_i n_i^{CHO}n_i^{OHN} + E^{CHO}\sum_i \left(n_i^{CHO}\right)^2 + E^{CHN}\sum_i n_i^{CHO}n_i^{CHN} = \sum_i n_i^{CHO}E_{\text{int}}^i \\
&E^{NHO}\sum_i n_i^{NHO}n_i^{CHN} + E^{NHN}\sum_i n_i^{NHN}n_i^{CHN} + E^{OHN}\sum_i n_i^{CHN}n_i^{OHN} + E^{CHO}\sum_i n_i^{CHO}n_i^{CHN} + E^{CHN}\sum_i \left(n_i^{CHN}\right)^2 = \sum_i n_i^{CHN}E_{\text{int}}^i
\end{aligned}$$

…….[3.1.7]

We have solved the above simultaneous equations (Eqn 3.1.7) by GNU-octave (version 2.1.33) using E^i_{int} values from Table 3.1.2 and obtained the values of energy components as listed in Table 3.1.4. These energy component values can also be used to recalculate the total interaction energy values and these are highly correlated to the original data with correlation coefficient = 0.82 (As shown in Figure. 3.1.2). It is found from Table 3.1.4 that the different types of H-bonds lead to different amount of interaction energies and classification of H-bonds just into two types, viz. polar and non-polar, is not sufficient to explain the varieties. It is found that N-H…O and N-H…N interactions are significantly distinct while O-H…N interactions are strongest among these. Earlier reports (Vargas et

al., 2000, Stepanian et al., 1985, Sheina et al., 1987) on C-H…O type interaction using MP2/aug-cc-pVTZ method is also in agreement with our results as their reported limiting value of interaction is about –2 kcal/mol as compared to our average value of –0.88 kcal/mol. Furthermore, it is expected that a basepair stabilized by two N-H…O bonds would be significantly more stable (by about –4 kcal/mol) than a basepairs stabilized by two N-H…N H-bonds.

We have also attempted to fit an equation

$$E_{\text{int}}^{i} = n_i^{NHO} E^{NHO} + n_i^{NHN} E^{NHN} + n_i^{OHN} E^{OHN} + n_i^{CHO} E^{CHO} + n_i^{CHN} E^{CHN} + K + \delta_i \quad \ldots\ldots[3.1.8]$$

using the same method. The fitted data for E^{NHO}, E^{NHN}, E^{OHN}, E^{CHO}, E^{CHN} along with K are given in Table 3.1.4. The value of K is very small (–0.16 kcal/mol) and the new set of interaction energy components gives almost identical correlation coefficient. The small value of K suggests that all the basepairs might have a small base independent attraction. Thus, it appears that the bases do not have any default non-specific attractive interaction.

Table 3.1.4. Energy components (in kcal/mol) due to formation of different type of H-bonds

Type of Hydrogen Bond	Energy components	Energy components fitting equation 3.1.8*
NH…O	-7.12	-7.04
NH…N	-5.72	-5.64
OH…N	-7.24	-7.16
CH…O	-1.67	-1.64
CH…N	-0.57	-0.49

* The constant term describing non-specific interaction, as obtained from fit, is –0.16 kcal/mol

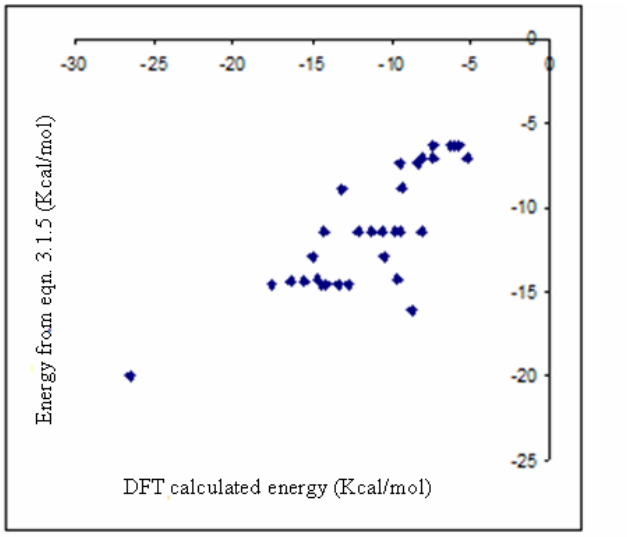


Figure 3.1.2. Correlation between interaction energy values calculated by combined DFT/HF method and those obtained from solving equation 3.1.7 are shown here.

3.1.4.4. Vibration Frequency of Basepairs

Infrared (IR) spectroscopy is one of the important tools for the detection of hydrogen bond. Hydrogen bond formation is most unambiguously determined by analysis of alteration of IR frequency and intensity upon complexation and such changes have been found to be in very good agreement with the theoretically derived spectra. We have calculated the theoretical IR spectra of the basepairs as well as for four methylated bases by numerical Hessian method. Infrared frequencies for the amino N-H bond stretching are in excellent agreement to those measured by different experiments. The calculated values are, however, systematically larger than the experimental values (Stepanian et al.,

1985, Santamaria et al., 1999). In case of basepairs stabilized by NH…O/N type (Table 3.1.5) hydrogen bonding red shift of the N-H stretching is accompanied by sharp increase of I/I_o ratio by at least a factor of 10 where I_o is the intensity of a normal mode vibration in the monomer and I is the intensity of the same vibration in the complex.

Often intensity of one of the N-H bond vibration increases 50-60 times as compared to the same vibration of the isolated base. The A:G H:S: T basepair, which is the most frequently observed non-canonical basepair does not show any strong H-bond with high intensity change. This is also reflected in rather weak interaction energy of this basepair with only two N-H…N hydrogen bonds and large propeller twist of its optimized structure. These H-bond distances fall within the normal H-bond distance range and D-H…A angle differ slightly from linearity.

Table 3.1.5. Hydrogen bond geometry and alterations in calculated IR frequencies and intensities for the basepairs stabilized by polar H-bonds only

No.	Basepair	Hydrogen bond	Length (Å)	Angle (^{o})	D…A (Å)	IR Freq. (cm^{-1})*	δυ (cm^{-1})	I/I_o
1	G:C W:W C	N4-H…O6 (2)	2.77	178.8	1.73	3170 (3591)	421	58.0
		N1-H…N3 (1)	2.92	177.3	1.89	3115 (3621)	506	10.1
		N2-H…O2 (1)	2.93	176.2	1.91	3370(3555s)	185	42.8
2	A:U W:W C	N6-H…O4 (1)	2.86	174.9	1.84	3332 (3583s)	251	24.7
		N3-H…N1 (2)	2.86	177.7	1.81	2979 (3589s)	610	41.8
3	G:U W:W C	N1-H…O2 (1)	2.77	172.6	1.74	3303 (3621)	318	54.1
		N3-H…O6 (2)	2.92	172.2	1.89	3180 (3589)	409	15.6
4	A:G H:S T	N6-H…N3 (1)	3.06	171.2	2.05	3389 (3583s)	194	20.5
		N2-H…N7 (2)	2.98	177.6	1.96	3308 (3555s)	247	28.3
6	A:U H:W T	N6-H…O2 (1)	2.79	169.9	2.12	3490 (3583s)	93	8.6
		N3-H…N7 (1)	3.13	169.2	1.55	2996 (3589)	593	43.2
7	A:A H:H T	N6-H…N7 (2)	3.02	164.9	2.02	3400 (3583s)	183	25.7
		N6-H…N7 (1)	3.02	164.9	2.02	3369 (3583s)	214	10.3
8	G:A W:W C	N1-H…N1 (1)	2.95	177.1	1.91	3083 (3621)	538	39.7
		N6-H…O6 (2)	2.83	177.1	1.80	3291 (3583s)	292	35.4
9	G:A S:W T	N6-H…N3 (1)	2.99	175.6	1.97	3255 (3583s)	328	6.2
		N2-H…N1 (2)	2.99	177.3	1.96	3303 (3555s)	252	80.6
10	A:A W:W T	N6-H…N1 (1)	2.99	176.6	1.97	3252 (3583s)	331	12.1
		N6-H…N1 (2)	2.99	176.6	1.97	3312 (3583s)	271	47.2

12	A:U W:W T	N3-H…N1 (1)	2.81	177.8	1.76	2918 (3589)	671	49.5
		N6-H…O2 (2)	3.06	172.1	2.05	3439 (3583s)	144	12.0
13	A:A H:W T	N6-H…N7 (1)	3.01	167.2	2.00	3352 (3583s)	231	42.1
		N6-H…N1 (2)	2.99	177.5	1.97	3312 (3583s)	271	6.6
14	A:U H:W C	N6-H…O4 (2)	2.98	171.1	1.97	3468 (3583s)	115	13.2
		N3-H…N7 (1)	2.79	174.1	1.75	2924 (3589)	665	44.6
18	G:G S:S T	N2-H…N3 (2)	3.00	178.3	1.98	3265 (3555s)	290	10.0
		N2-H…N3 (1)	3.00	178.2	1.98	3314 (3555s)	241	63.6
24	G:G H:W T	N2-H…O4 (1)	3.16	164.3	2.17	3540 (3555s)	15	15.5
		N1-H…N7 (1)	2.85	170.6	1.82	3258 (3621)	363	46.9
31	A:C W:W T	N4-H…N1 (1)	2.92	179.2	1.89	3181(3591)	410	3.0
		N6-H…N3 (2)	2.96	176.8	1.93	3237(3583s)	346	60.1
32	C:U W:W T	N4-H…O2 (1)	2.90	175.3	1.88	3359 (3591)	232	25.2
		N3-H…N3 (2)	2.97	170.8	1.95	3187 (3589)	402	21.5

* The values in parenthesis are the vibration frequencies of the free bases. Symmetric stretching of both amino N-H bonds are indicated by "s" after the frequency values

The non polar H-bonds in basepairs stabilized by C-H…O/N (Table 3.1.6) mediated H-bonds however, show a blue shift of C-H stretching and the I/I_o ratios generally decrease on hydrogen bonding. Although, C-H stretching of U:U h:w T basepair undergo a red shift of 45 cm^{-1} and I/I_o ratio increase by a factor of 385.7 but it is not significant as the intensity of the C-H stretching in the isolated bases was too small and we consider it a blue shifting H-bond. These H-bonds are rather elongated and D-H..A angles also differ significantly from linearity.

Table 3.1.6. Hydrogen bond geometries and alterations in calculated IR frequencies and intensities of basepairs stabilized by C-H…N/O interaction along with polar hydrogen bonds

No.	Basepair	Hydrogen Bond	Length (Å)	Angle (°)	D...A (Å)	IR Freq. (cm^{-1})*	δυ (cm^{-1})	I/I_o
5	A:G s:s T	N2-H…N3 (2)	3.00	177.28	1.98	3352 (3555s)	203	32.1
		C2-H…N3 (1)	3.65	124.2	2.92	3168 (3143)	-25	0.10
15	A:G w:s C	N2-H…N1 (1)	3.03	172.87	2.01	3340 (3555s)	215	37.4
		C2-H…N3 (2)	3.51	147.3	2.55	3191 (3143)	-48	0.3
21	A:C w:w C	N4-H…N1 (1)	2.99	178.13	1.97	3308 (3591)	283	33.6
		C2-H…N3 (2)	3.52	143.80	2.58	3184 (3143)	-41	0.7
22	A:U s:w C	N3-H…N3 (2)	2.98	170.35	1.95	3151 (3589)	438	26.1
		C2-H…O4 (1)	3.32	146.86	2.25	3199 (3143)	-56	1.1
25	A:A w:w C	N6-H…N1 (1)	3.05	174.3	2.03	3399 (3583s)	184	22.8
		C2-H…N1 (2)	3.51	142.75	2.59	3160 (3143)	-17	0.4
26	U:U h:w T	N3-H…O4 (1)	2.87	178.05	1.84	3324 (3589)	265	25.8
		C5-H…O4 (2)	3.25	156.65	2.22	3210 (3255)	45	385.7
33	C:C w:h C	N4-H…N3 (1)	2.98	159.45	1.84	3470 (3591s)	121	9.3
		C5-H…O2 (1)	3.72	164.58	2.26	3239 (3222)	-17	5.5

3.1.4.5. Basepair Opening Flexibility

Basepair opening involves breakage of hydrogen bonds holding the bases in a pair and movement of at least one base out of a helical stack. It is an important part of many biochemical transformations like replication, transcription and enzyme catalyzed DNA modifications and in case of selective methylation. Since both canonical and non-canonical basepair are held within the double helix by hydrogen bonding as well as base-stacking interactions, basepair opening require much higher activation energy and it is an energetically unfavorable process. Up to date many groups have investigated the feasibility of different basepair opening in term of energy from both theoretical and experimental studies (Seibert et al., 2003, Varnai and Lavery, 2002, Giudice et al., 2003, Ramstein and Lavery, 1988). In this work we have calculated the opening and H-bonding dynamics assuming vibration of a basepair is completely harmonic. We have estimated

the intra basepair parameters buckle, open-angle, stagger, shear, propeller and stretch at the two extreme points of the normal mode vibration and tried to realize the feasibility of different basepair opening from the differences in the parameters at these two limits. These analyses were done for the first two types of basepairs, 23 in number. As the third category involves H-bonding with the sugar moieties, their dynamics are largely dominated by torsional flexibility and sugar pucker variations, which could be non-harmonic as well. Moreover, their complete characterization requires comparison of sugars of C2'-Endo geometry or deoxyribose sugars, which are beyond the scope of the present study.

We know that most molecules including the basepairs have 3N-6 degrees of freedom and one expects to determine vibrational frequencies about each of these degrees of freedoms by normal mode analysis. Changes in frequencies about some X-H bonds are indicative of H-bonding, which have already been discussed. When two bases are loosely bound to each other by H-bonds, there would be some vibrations and rotations about these loose bonds also. In order to detect frequencies of these vibrations we have generated coordinates of all atoms of a basepair, which undergoes modification due to such vibration. The amplitude of vibration (A) in the Cartesian coordinate system corresponding to each frequency are used to obtain the altered geometry of the basepair,

$$X_{max}^{i} = X_{o}^{i}\{1 + A \sin(90)\} \text{ and } X_{min}^{i} = X_{o}^{i}\{1 + A \sin(-90)\} \quad \text{.....[3.1.9]}$$

where X_{o}^{i} are position vectors for i^{th} atom of a basepair in the optimized geometry. We calculated all the intra basepair parameters from these pairs of structures (X_{min}^{i} and X_{max}^{i}) corresponding to each frequency and estimated differences between the parameters. Thus, in principle, we obtain 3N-6 sets of differences in basepair parameters. All the

differences are negligible for high frequency motions (above 3000 cm^{-1}) where only C-H or N-H stretching vibrations take place. The differences in basepair parameters are often significant corresponding to frequencies in the 1000 to 2000 cm^{-1} region as major C-C, C-N bond vibrations occur in these regions, which also can affect calculation of shear, open-angle etc. As for example a change in bond angle C2 (Py)-N1-C1' causes large differences in basepair parameter shear or open-angle hence we do not consider those as intra basepair motion. Large differences in parameters at low frequency region generally correspond to inter base motion. Sometimes these are coupled to two or more base motions also, such as coupled open-stagger movement in the 9^{th} basepair, G:A S:W T, A:A H:W T (13^{th}) or G:G S:S T (18^{th}) basepairs. Table 3.1.7, gives a list of such changes in basepair parameters and their corresponding frequencies, only when the changes are significantly large and they are not coupled to any intramolecular vibration such as bond length stretching, angle bending etc. We have selected only those motions for which a rotational parameter, viz, propeller, buckle or open angle, alter by 4^{o} or more or for which a translational parameter, such as shear, stagger or stretch, changes by 0.3 Å or more. Thus, we can say now that a G:C W:W C basepair undergoes a motion along buckle with a frequency of 43.2 cm^{-1}, along propeller with a frequency of 31.0 cm^{-1} and along stagger with a frequency of 73.4 cm^{-1}.

We have further calculated force constants from the frequency values as these are related through the equation $w = \sqrt{k/\mu} = 2\pi c \bar{\upsilon}$ where $\bar{\upsilon}$ are the frequencies in cm^{-1}, c is the velocity of light, k is the force constant and μ is the reduced mass of the basepair at the respective frequency. We have used the reduced mass μ calculated by GAMESS for calculating the force constant corresponding to all the translational motion, namely

stagger and shear. We have used reduced moments of inertia for calculating force constants corresponding to rotational motions about propeller, buckle and open angles. In order to calculate a reduced moment of inertia of a basepair about a rotational axis, for example $\overline{Z}_m$ along which propeller motion takes place (Mukherjee et al., 2006), we have oriented a basepair such that its mean axes X_m, Y_m and Z_m are along the laboratory reference frame X, Y and Z. We have then calculated moments of inertia of both the bases about Z_m axis, for example, as

$$I_{zz}^{1} = \sum_{i=1}^{N1} (m_i x_i^2 + m_i y_i^2) \text{ and } I_{zz}^{2} = \sum_{i=1}^{N2} (m_i x_i^2 + m_i y_i^2) \qquad \text{.....[3.1.10]}$$

where N1 and N2 are the number of atoms in the first base and second base respectively. The reduced moment of inertia I is calculated as $I = I_1 I_2/(I_1 + I_2)$. Finally, force constant corresponding to the rotational degree of freedom are calculated as

$$k = 4\pi^2 c^2 \bar{\upsilon}^2 I \qquad \text{.....[3.1.11]}$$

These force constants values would be useful for molecular dynamics simulations of RNA treating each base as a mass, in a lattice model type simulation. We can also use these force constants (k) to predict probability distribution ρ(α) for different values of a parameter through

$$\rho(\alpha)/\rho(\alpha_o) = e^{-\frac{1}{2}k(\alpha-\alpha_o)^2/k_B T} \qquad \text{.....[3.1.12]}$$

where α_0 is equilibrium value of a parameter α, k_B is Boltzman constant and T is absolute temperature. The half-width at half maximum from this probability distribution corresponds to standard deviation (σ), which were also calculated earlier from x-ray

crystallographic data for some of the basepairs (Mukherjee et al., 2006). Hence we have measured the intrinsic standard deviations (σ_{calc}) as

$$\sigma_{calc} = \sqrt{2 \ln 2 . k_B T / k} \qquad \text{.....[3.1.13]}$$

The calculated values of standard deviations of different basepair parameters for the 23 basepairs (Table 3.1.7) are in excellent agreement to those reported earlier for some of the basepairs from analysis of 145 functional RNA crystal structures as shown in Figure 3.1.3.

The σ_{calc} values for open are quite small compared to those for propeller or buckle. Same trend is also seen for basepairs in crystal structures. We did not notice any movement about the stretch directions for any basepair and hence we can speculate their values as zero. Very small values of σ_{expt} for stretch of all the basepairs were noticed earlier from analysis of RNA crystal structures (Mukherjee et al., 2006).

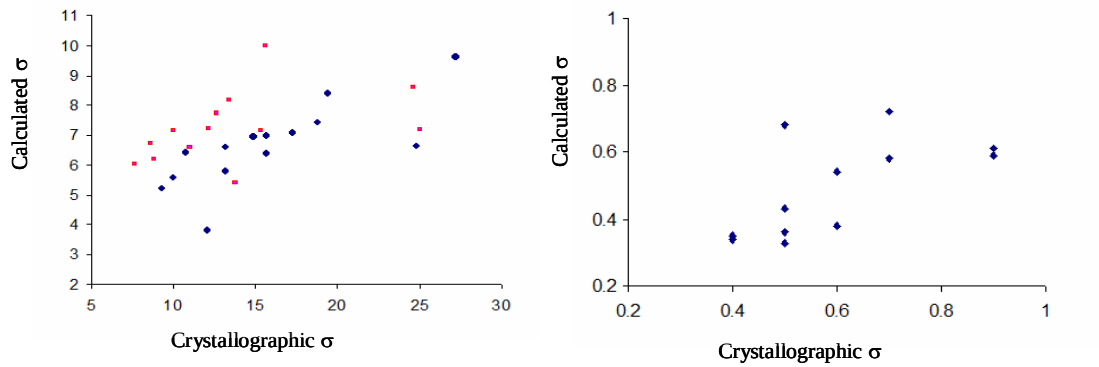


Figure 3.1.3. Correlation between calculated and measured standard deviations for (a) Buckle and Propeller and (b) Stagger for different basepairs. Filled circles represent values of Buckle (Blue) and those of Propeller are represented by filled squares (Red) in Figure 3.1.3 (a).

Table 3.1.7. Calculation of dynamics of the basepairs along the five intra basepair parameter directions are shown.

No	Basepair	Buckle	Open angle	Propeller	Stagger	Shear
1	G:C W:W C	43.19; 82.5	-	31.00; 26.2	73.36; 3.18	-
2	A:U W:W C	29.84; 38.0	-	38.74; 30.6	73.38; 3.0	100.8; 6.0
3	G:U W:W C	32.04; 43.7	60.09; 22.3	31.55; 32.5	73.57; 3.0	-
4	A:G H:S T	20.49; 28.9	38.22; 119.5	38.22; 40.2	64.26; 2.5	-
5	A:G s:s T	12.47; 12.7	24.81; 56.2	-	44.3; 0.97	-
6	A:U H:W T	25.0; 24.4	48.24; 120.0	33.72; 23.2	67.39; 3.0	93.57; 5.5
7	A:A H:H T	20.06; 24.4	44.34; 139.9	29.45; 23.1	75.3; 2.8	60.79; 2.0
8	G:A W:W C	16.88; 21.4	-	25.27; 17.6	59.74; 1.9	74.96; 3.0
9	G:A S:W T	14.59;16.8	37.72; 133.0	26.37; 11.8	37.72; 0.78	101.49; 5.9
10	A:A W:W T	19.33; 28.6	59.6; 306.5	28.16; 15.8	42.07; 1.1	97.32; 5.6
12	A:U W:W T	19.71; 27.3	64.76; 335.7	35.65; 22.7	69.77; 2.9	107.3; 7.5
13	A:A H:W T	22.86; 35.1	48.04; 181.6	30.13; 22.9	48.04; 1.2	100.8; 6.6
14	A:U H:W C	24.63; 23.6	65.49; 221.5	36.61; 27.3	74.4; 3.2	-
15	A:G w:s C	18.78; 27.0	30.87; 94.4	23.22; 16.0	41.5; 1.0	61.43; 2.1
18	G:G S:S T	-	36.28; 131.1	24.2; 19.7	36.28; 0.7	102.5; 5.5
21	A:C w:w C	26.98; 30.2	-	27.51; 29.6	39.01; 0.8	65.92; 2.38
22	AU s:w C	40.6; 71.6	110.2; 672.4	31.71; 24.7	73.17; 3.0	-
24	G:G H:W T	13.94; 15.1	40.9; 157.2	33.67; 45.6	52.96; 1.8	70.2; 3.3
25	A:A w:w C	20.07; 29.9	34.97; 111.2	15.89; 10.9	44.08; 1.1	82.00; 3.7
26	U:U h:w T	45.75; 83.3	93.39; 488.1	30.83; 15.3	67.86; 2.9	75.89; 3.5
30	A:C W:W T	53.32; 127.2	68.3; 288.3	34.56; 24.6	53.32; 1.6	-
32	C:U W:W T	37.90; 45.7	119.9; 592.4	22.2; 9.5	68.67; 2.8	95.81; 6.1
33	C:C w:h C	47.71; 99.3	57.61; 177.5	39.41; 41.4	65.53; 2.1	-

In the table 3.1.7, the sixth parameter, stretch, remains static in all the basepairs. The four parameters in each cell represent the (a) normal mode frequency (cm^{-1}) and (b) force constant to the corresponding vibration (k) (kcal/mol for rotational parameters or kcal/mol/$Å^2$ for translational parameters), respectively.

It appears from the Table 3.1.7 that buckle, propeller and stagger motions are most feasible for all type of basepairs and these motions occur mainly at the lower frequency region. This may indicate that by changing buckle, propeller and stagger, the energy of a basepair does not alter significantly while it is always difficult to pull two bases from each other along the stretch direction, irrespective of polarity of Hydrogen bonds and type of basepair. Furthermore, previous studies (Dickerson, 1989) using energy calculation for basepairs having different H-bond lengths indicated a rather steep and non-harmonic energy landscape. This also indicates that movement of a basepair in stretch direction is a costly affair. The measured values of standard deviations of open-angles from RNA crystal structures are generally small and our calculated σ's for open-angles also follow the same trend. Another hindered motion is that about open angle, as a change in open angle always causes alteration of H-bond geometry. The calculated values of standard deviations σ_{calc} for buckle; stagger and propeller are in general with close agreement with the corresponding experimental values. As for example σ_{calc} for buckle for G:C W:W C, A:U W:W C, G:U W:W C and A:A H:W T are small in both sets of data. On the other hand σ_{calc} as well as σ_{expt} for buckle of A:G s:s T, A:G W:S T are large. These agreements are reflected in large correlation coefficients (R=0.76 for buckle, 0.52 for propeller and 0.72 for stagger) for these parameters (Figure 3.1.3).

In general our σ_{calc} are always smaller than σ_{expt} by a factor of two. This regular difference may be due to several effects: (i) many of the basepairs in crystal are positioned in a stacked geometry with attached sugar; (ii) most of the basepairs in RNA macromolecular environment have different non-local interactions, such as base triple formation, perturbing their structures and giving rise to more variability; (iii) it was

shown earlier that the theoretically calculated IR frequencies are generally systematically larger than corresponding frequencies from IR experiments (Santamaria et al., 1999), giving rise to smaller σ_{calc} and (iv) the crystal data analysis was done earlier by considering all available structures better than 3.5 Å resolution and even considering homology related structures. This might have introduced some artificial extra structural variability to the basepairs. It is expected that a fresh analysis of structural parameters from non-redundant set of RNA crystal structures would possibly give a better agreement of the intrinsic σ_{calc} values with observed σ_{expt} values for most of the basepairs.

Considering the intrinsic standard deviations and force-constants, we are tempted to speculate that during basepair opening, i.e. melting of DNA or RNA, the basepairs may first undergo a transformation to increase propeller, buckle or stagger and subsequently increase open angle or shear and lastly the hydrogen bonds break to increase separation between the two involved bases.

3.1.5. Discussion

On analyzing the structures and dynamics of the frequently occurring basepairs, we observed that from most of the polar H-bonded basepairs are significantly more stable than the non-polar basepair mediated by C-H…O/N kind of interactions. The high interaction energy values of the polar basepairs indicate that they are highly stable and may help in prediction of RNA secondary structure. We also observed that energy contributions due to three types of polar H-bonds, namely N-H…O, O-H…N and N-H…N, are notably different while energy contribution due to non-polar interactions are insignificant as found by the quantum chemical calculations.

We have for the first time analyzed basepair opening dynamics of different canonical as well as non-canonical basepairs using normal mode analysis, followed by calculation of basepair parameters. Our calculated intrinsic flexibility of the basepairs is in excellent agreement with the experimental data for which large number of basepairs were found in RNA crystal structures. The calculated flexibilities, for which sufficient data is not available in X-ray crystal structure database, thus can be useful for modeling RNA three-dimensional structures. We have seen that some basepair's structure and interaction energies are affected by environment to some extent but occasionally isolated basepairs (in crystal environment) also deform considerably upon geometry optimization. This possibly indicate that crystal environment have relatively less importance in formation and structure of a basepair.

Some of the basepairs among the 145 RNA basepairs which are occurred frequently are found to be occasional in the non-redundant database. So optimization of the frequently occurred basepairs by different popular empirical, semi-empirical and *ab initio* quantum chemical methods, may help in understanding the features of the RNA basepairs in more details, which we will follow in the next section.

Section 3.2

Theoretical Analysis of Non-Canonical Basepairs: Comparison of Various Computational Chemistry Method with Crystal Database

3.2.1. Introduction

Different aspects of canonical and non-canonical basepairs are described in section 3.1. A number of softwares are available like BPFIND (Das et al., 2006), find-pair (Lu and Olson, 2003, Lu and Olson, 2008), FR3D (Sarver et al., 2008), Basepair viewer by NDB server (http://ndbserver.rutgers.edu/services/BPviewerm), MC-Annotate (Lemieux and Major, 2002) etc. for detection of the canonical and non-canonical basepairs in DNA or RNA molecules. Obviously Different groups adopt different algorithms. While find-pair (Lu and Olson, 2003, Lu and Olson, 2008) emphasizes on planarity of two bases and any hydrogen bond between the nucleotide residues, the FR3D and MC-Annotate methods add hydrogen atoms explicitly and look for probability of hydrogen bond formation, similar to an earlier approach from our laboratory to detect H-bonds in protein-DNA complex crystal structures (Mukherjee et al., 2005). However the latter methods FR3D and MC-Annotate do not consider the sugar-mediated basepairs. The BPFIND (Das et al., 2006) algorithm on the other hand looks for at least two hydrogen bonds and their linearity and co-planarity of the two involved bases are considered together using few well-defined pseudo-angles. Unlike the other methods, BPFIND can detect protonated basepairs also (Das et al., 2006, Chawla et al., 2011). Although a single hydrogen bond between two bases can lead to considerably strong attraction, even after formation of a strong hydrogen bond the bases can rotate freely about the vector through the H…Acceptor bond, thus may attain non-planar geometry, which may not stack well within a double helix. However, for specificity in molecular recognition, at least two hydrogen bonds are required. Therefore, BPFIND was designed to detect basepairs having at least two hydrogen bonds between the bases or one between

the bases and another between a base and the sugar of another residue. This also considers C-H…O/N interaction mediated basepairs, as importance of C-H…O/N type of hydrogen bonds in different types of systems has been reported by various groups (Desiraju and Steiner, 2001). Considering three basepairing edges of each base along with the Cis and Trans orientations with respect to each other, there is a possibility of 288 different types of canonical and non-canonical basepairs. However consideration of at least two hydrogen bonds between the bases reduces the number of duplex basepairs to 126. This also includes some of the protonated basepairs, where absence of an unusual protonation may lead to electrostatic repulsion.

Structure and interaction energy of different types of DNA or RNA basepairs optimized by different quantum chemical methods have been studied by various groups. It was observed that most of the frequently observed basepairs are found to be quite stable in terms of their interaction energy components. Obviously, it was noted that non-canonical basepairs having two polar hydrogen bonds between them have high stability, similar to those of the canonical basepairs, while non-polar basepairs having C-H…O/N type interactions have significantly low interaction energies. On the other hand, basepairs having single hydrogen bond may show instability as often these basepairs deform significantly upon optimization. Optimizations of these basepairs generally require a few constraints to maintain their desired orientation (Sharma et al., 2010b, Mládek et al., 2009, Sponer et al., 2005b). Structure of a basepair can be easily characterized by six relative degrees of freedom between the two bases and following IUPAC-IUB (Olson et al., 2001) conventions (More elaborately discussed in chapter 1), a few algorithms were developed for calculation of basepair parameters, e.g., propeller, shear, open-angle etc

(Lu and Olson, 2003, Lu and Olson, 2008, Ravishanker et al., 1989, Lavery and Sklenar, 1988, Mukherjee et al., 2006, El Hassan and Calladine, 1995, Babcock and Olson, 1994). It was shown earlier that the canonical as well as the non-canonical basepairs observed in RNA mostly possess coplanar orientations, as measured by the relative orientation parameters, particularly stagger, propeller and buckle. These parameters calculated using unified axis definition (Lu and Olson, 2008, Lu and Olson, 2003) however do not give a true indication of non-planarity, particularly for non-canonical basepairs (Halder and Bhattacharyya, 2010). It may further be noted that there are few other methods to quantitate basepair orientation, such as isostericity index, which also indicate need of base pairing edge specific definition to compare structures of non-canonical basepairs (Stombaugh et al., 2009).

Geometry optimization, by nature, is an approximate method and its outcome depends on the choice of the approximations. There are some reports on structures of basepairs optimized by HF/6-31G** (Kabelác et al., 2000, Bhattacharyya et al., 2007, Sharma et al., 2008), but there were questions regarding this method as it does not include electron correlation effect. The most widely accepted method now is B3LYP/6-31G**, though few other methods are also becoming popular due to advancement in computer technology and development of new basis set and density functionals (Sponer et al., 2004). Several groups of workers carried out characterization of structures of different biomolecular fragments or their complexes through optimization by empirical, semi-empirical and *ab initio* quantum chemical methods of different level (Mallajosyula et al., 2005, Danilov et al., 2005, Kurita et al., 2005, Starikov and Steiner, 1997, Hobza et al., 1997, Hobza and Sponer, 1999, Sponer and Hobza, 2003, Jayaram et al., 2006, Paton

and Goodman, 2009, Riley et al., 2009). However, quantitative comparisons of the optimized structures were rarely addressed. Hobza *et. al* (Hobza et al., 1997) studied diverse DNA basepairs using different methods but usage of thymine bases reduced its relevance in RNA structure and function. The recent study by Paton and Goodman also was restricted to compare energies rather than quantitative geometry comparison (Paton and Goodman, 2009). It is well known that most of the quantum chemical methods have some or other artifacts due to their inherent approximation, e.g. Hartree-Fock methods are incapable of considering electron correlation, Density Functional Theory based methods are inadequate to estimate dispersion interaction etc. Nucleic acid basepairs have also been optimized by few post Hartree-Fock methods by which one can consider effect of electron correlation, but these are extremely costly for many atomic systems (Starikov and Steiner, 1997, Riley et al., 2009, Danilov and Anisimov, 2005, Sponer et al., 2010) and hence is not adopted regularly. A thorough comparison of the optimized structures against some experimental structures is necessary to understand quality of the methods. Geometry comparison of the optimized basepairs to a specific crystal structure is also questionable as there can be subtle distortions in experimental structures, leading to artifacts or context dependent variability, whose effects can be removed by considering many examples or an ensemble of experimental structures in an unbiased way. Moreover it is a general trend that one optimizes a structure using a faster method (smaller basis set or semi empirical methods, for examples (Jayaram et al., 2006, Jensen, 2007) and calculates single point energy using a more rigorous method, assuming the optimization by a faster method may lead to correct structure. It was also pointed out that even semi-empirical methods give good optimized structures or saddle point structures while their

energy values may not be accurate (Jensen, 2007). This concept has not been questioned earlier as comparisons of only energies of basepairs are generally done. On availability of huge structural data, now time is ripe enough to evaluate this common theoretical technique.

There are few attempts now to classify the available RNA crystal structures and thereby determining a non-redundant set of RNA crystal structures (Stombaugh et al., 2009, Ray et al., 2009, Murthy and Rose, 2003, Klosterman et al., 2002, Ray et al., 2012). This RNA crystallographic data can be analyzed to get structural features of the non-canonical basepairs in an unbiased way. We have analyzed the non-redundant set of RNA crystal structures and detected all Watson-Crick and non-canonical basepairs in this set. We have further determined basepair orientation parameters of these basepairs following IUPAC-IUB convention. We have thereafter chosen 32 basepairs, which occur frequently in the functional RNA crystal structures and optimized their structures in gas phase using most of the widely accepted computational chemistry methods. The basepair orientation parameters of the optimized basepairs have been compared with crystallographic averages to determine efficiency of the methods. Most of the basepairs stabilized by a pair of polar hydrogen bonds are optimized to structures close to the crystal ensembles by most of the methods, indicating their high stability. The non-polar basepairs and the basepairs involving sugar-mediated hydrogen bonds are generally found to be rather flexible.

3.2.2. Methods

3.2.2.1. Basepair Coordinates Extraction and Model Building

The best representatives of each functional class of RNA crystal structures were collected from all available RNA structures in PDB (Berman et al., 2000) in April 2010, classified by HD-RNAS (Ray et al., 2012) (http://www.saha.ac.in/biop/www/HD-RNAS.html), which contains best representative structures from each type of classes obtained through best resolution, smaller R-factor and larger lengths nucleotides. This non-redundant data set of RNA crystal structures corresponds to 107 PDB files having at least 30 residue long RNA chains. The BPFIND (Das et al., 2006) software was used to detect all possible basepairs in these structures and NUPARM (Mukherjee et al., 2006) was used to calculate their basepair parameters, such as propeller, buckle, etc. using hydrogen bonding edge specific axis system. We have selected 32 highly frequent basepairs and chosen their best representatives as the initial geometries for optimization. Initial structures of most of the basepairs were chosen from the best representative within the non-redundant set and are given in Table 3.2.1.

Table. 3.2.1. Mean and standard deviations (within parenthesis) of intra basepair parameters from crystal structure database analysis. We have tabulated only frequent basepairs in the non-redundant dataset (frequency greater than 10).

No	Basepair (Frequency /frequency in triplets) for the first 32 basepairs)	Buckle (^{O})	Open (^{O})	Propeller (^{O})	Stagger (Å)	Shear (Å)	Stretch (Å)
1	G:C W:W C (5676/264)	-5.39(9.3)	0.49(3.8)	-6.51(7.9)	-0.14(0.4)	-0.07(0.4)	2.86(0.1)
2	A:U W:W C (1711/81)	-1.96(8.3)	3.27(4.5)	-6.56(8.3)	-0.04(0.4)	0.11(0.3)	2.79(0.1)
3	G:U W:W C (825/28)	-1.51(8.5)	-0.61(5.8)	-6.55(7.3)	-0.17(0.4)	-2.22(0.4)	2.81(0.2)
4	A:G H:S T (531/36)	-2.12(14.0)	13.11(5.7)	1.07(13.1)	0.14(0.5)	2.24(0.4)	3.31(0.2)
5	A:U H:W T (401/36)	-1.41(13.6)	-0.39(6.6)	-1.17(9.8)	-0.04(0.5)	0.17(0.5)	2.82(0.2)
6	A:G W:W C (142/0)	-10.77(14.2)	1.27(5.8)	-8.55(12.0)	-0.41(0.4)	0.07(0.5)	2.8(0.2)
7	A:A H:H T (106/3)	-9.63(15.2)	-4.66(3.3)	5.81(13.2)	-0.35(0.3)	2.51(0.4)	2.86(0.2)
8	G:A S:W T (86/7)	13.04(19.0)	13.62(7.0)	0.27(16.6)	0.16(0.6)	1.85(0.4)	3.32(0.2)
9	A:A H:W T (80/24)	1.84(12.8)	5.7(5.2)	-2.88(19.6)	0.04(0.5)	2.31(0.4)	2.92(0.2)
10	A:U H:W C (77/14)	-3.59(18.1)	2.93(6.8)	-4.53(12.4)	0.15(0.6)	-0.22(0.4)	2.8(0.2)
11	U:U W:W C (76/0)	-10.11(8.6)	-2.33(5.3)	-12.64(8.0)	-0.15(0.4)	-2.36(0.4)	2.86(0.2)
12	A:C H:W T (75/3)	-1.26(20.5)	4.3(6.4)	-5.68(15.6)	-0.18(0.5)	2.37(0.4)	2.91(0.2)
13	G:C W:W T (63/1)	-4.35(15.4)	7.34(7.3)	-6.67(10.7)	-0.09(0.4)	-2.27(0.5)	2.88(0.2)
14	A:A W:W T (59/25)	7.87(11.3)	-7.58(7.5)	-2.04(30.2)	-0.14(0.7)	2.19(0.5)	2.87(0.2)

15	A:U W:W T (59/3)	0.05(12.7)	-1.9(7.4)	0.13(13.3)	0.03(0.6)	-0.24(0.4)	2.83(0.2)
16	G:G H:W C (48/2)	6.43(12.6)	-3.86(6.2)	-0.31(12.1)	0.08(0.6)	-2.94(0.4)	2.88(0.2)
17	G:G H:W T (39/0)	9.29(12.9)	1.77(6.6)	-0.64(8.4)	-0.13(0.3)	-0.02(0.6)	2.89(0.2)
18	G:G S:S T (32/4)	4.64(22.1)	-4.54(3.6)	13.25(18.5)	-0.66(0.5)	1.27(0.3)	3.46(0.2)
19	U:U W:W T (20/0)	-6.22(16.0)	-4.52(3.2)	-4.36(10.2)	0.11(0.4)	-2.35(0.4)	2.79(0.2)
20	A:G s:s T (285/30)	15.84(26.9)	27.99(7.2)	-7.44(13.6)	0.22(0.9)	1.71(0.3)	3.02(0.2)
21	G:U s:h C (53/0)	-11.75(18.2)	-1.22(5.5)	0.15(7.4)	-0.1(0.6)	0.99(0.4)	3.44(0.1)
22	A:A h:s T (49/0)	-5.65(13.5)	-1.39(6.0)	1.66(9.4)	-0.13(0.5)	2.38(0.3)	2.75(0.2)
23	A:G w:s C (48/6)	-19.31(28.9)	45.49(12.6)	-6.41(24.2)	-0.1(0.9)	2.07(0.4)	2.84(0.2)
24	A:U s:w C (18/2)	-21.67(10.6)	-3.7(4.2)	11.59(12.3)	0.14(0.3)	0.15(0.3)	2.95(0.2)
25	U:U w:h T (17/0)	15.73(18.7)	-13.61(6.9)	-9.00(15.1)	-0.18(0.7)	2.59(0.3)	2.91(0.2)
26	A:C w:w C (14/0)	-10.02(12.4)	14.26(10.4)	-8.08(6.6)	0.05(0.9)	2.45(0.4)	2.44(0.4)
27	A:A w:w C (12/0)	-16.84(15.7)	21.66(13.2)	-15.44(12.8)	-0.77(0.2)	-2.14(0.9)	2.48(0.2)
28	A:A W:S C (86/0)	-4.49(13.8)	-29.14(10.1)	-12.72(12.2)	-0.38(0.4)	-2.51(0.5)	3.52(0.3)
29	A:U w:s T (69/1)	-1.57(21.0)	-58.71(12.1)	-18.61(14.6)	-0.33(0.9)	-0.4(0.5)	4.06(0.3)
30	A:C W:S C (61/0)	-4.44(22.8)	-34.74(10.3)	-21.29(14.5)	-0.47(0.6)	0.3(0.6)	4.23(0.4)
31	A:G S:S C (35/5)	1.6(23.2)	-15.89(20.7)	-8.36(30.3)	-0.13(1.0)	2.36(0.4)	3.82(0.5)
32	A:U W:S C (18/0)	-17.64(37.2)	-31.48(8.3)	-1.47(13.4)	-0.59(0.5)	0.44(0.4)	4.26(0.3)

33	A:C +:W C (29)	-2.59(18.0)	6.92(9.8)	-9.51(7.4)	0.04(0.4)	-2.44(0.7)	2.78(0.2)
34	C:U W:W C (22)	7.01(22.3)	-5.14(9.0)	-16.44(14.2)	0.14(0.7)	0.59(0.8)	2.94(0.3)
35	G:C S:S C (22)	0.42(21.9)	-22.19(7.1)	-13.71(22.0)	-0.63(1.0)	0.77(0.5)	4.16(0.4)
36	A:C S:W C (21)	0.96(21.6)	-27.6(5.6)	-17.53(22.0)	-0.55(0.6)	2.58(0.3)	3.48(0.1)
37	G:U S:S C (20)	21.75(7.0)	-21.75(4.2)	-14.57(12.5)	-1.00(0.6)	0.82(0.5)	4.19(0.2)
38	A:U h:s T (19)	-2.23(12.9)	-54.33(11.2)	10.28(14.2)	-0.89(0.8)	0.8(0.7)	3.65(0.4)
39	A:A s:s C (18)	-27.28(22.0)	-32.16(5.2)	-3.25(21.7)	-0.76(1.0)	-2.56(0.4)	2.92(0.3)
40	G:C S:W T (16)	0.88(25.3)	14.28(8.6)	-10.51(20.8)	-0.10(0.7)	1.77(0.5)	3.24(0.2)
41	G:C h:h T (12)	5.92(15.4)	-6.33(7.5)	1.19(6.5)	0.12(0.6)	3.35(0.7)	2.94(0.1)
42	G:U S:W C (11)	-15.28(4.7)	9.72(4.4)	21.46(9.3)	0.87(0.4)	0.51(0.3)	2.97(0.2)
43	G:C W:+ C (11)	0.42(9.1)	-2.44(11.2)	-7.48(10.0)	-0.08(0.5)	-2.3(0.5)	3.04(0.3)
44	A:C W:W T (10)	3.64(8.6)	8.63(12.9)	-27.68(8.8)	-0.3(0.8)	2.44(0.5)	2.81(0.2)

Atomic co-ordinates of the selected basepairs were extracted from the corresponding PDB files (Table 3.2.2). The selected basepairs can be categorized as (i) those stabilized by N-H...N/O type of hydrogen bond (polar basepairs), (ii) those stabilized by one N-H...N/O hydrogen bond and another C-H...N/O type of interaction (non-polar basepairs), (iii) those stabilized by one polar/non-polar hydrogen bond between the bases and another hydrogen bond between one of the base and the O2' of hydroxyl group atom of the sugar connected to another base (sugar mediated basepairs). We have represented the basepairing edges of the non-polar basepairs with lower case letters throughout this manuscript following Das *et. al* (Das et al., 2006). The sugar phosphate moieties at N9 position of purines or N1 of pyrimidines were replaced by methyl groups, instead of hydrogen to remove the unnatural dipole moments or hydrogen bonding. Modeling of all the hydrogen atoms were done with the help of the MOLDEN (Schaftenaar and Noordik, 2000).

Table 3.2.2. Source of the frequent basepairs along with the residue numbers.

No	Basepair	PDB ID	Residue No.
1	G:C W:W C	1DFU	97(M)-79(N)
2	A:U W:W C	1ASY	607(S)-666(S)
3	G:U W:W C	1ASY	610(S)-625(S)
4	A:G H: S T	1MZP	21(B)-44(B)
5	A:U H:W T	1ASY	614(S)-608
6	A:G W:W C	1QVG	2596(O)-2582(O)
7	A:A H:H T	1ASY	609(R)-623(R)
8	G:A S:W T	1FFK	166(O)-924(O)
9	A:A H: W T	1FFK	460(O)-455(O)
10	A:U H:W C	1FFZ	2470(A)-2277(A)
11	U:U W:W C	2QAO	1856(B)-1886(B)
12	A:C H:W T	1JSE	1055(A)-1200(A)
13	G:C W:W T	1H4S	15(T)-48(T)
14	A:A W:W T	1QVG	2566(O)-2699(O)
15	A:U W:W T	1ASZ	615(R)-648(R)
16	G:G H:W C	2QAL	1053(A)-1057(A)
17	G:G H:W T	1QVG	868(O)-775(O)
18	G:G S:S T	1FFK	315(O)-336(O)
19	U:U W:W T	2J02	956(A)-960(A)
20	A:G s:s T	1FFK	243(O)-274(O)
21	G:U s:h C	3CME	1917(O)-1918(O)
22	A:A h:s T	1GTR	13(B)-22(B)
23	A:G w:s C	1SM1	991(O)-2020(O)
24	A:U s:w C	1M1K	329(A)-346(O)
25	U:U w:h T	1W2B	2781(O)-2791
26	A:C w:w C	1M5K	20(B)-63(B)
27	A:A w:w C	1FFK	635(O)-620(O)
28	A:A W:S C	1N78	89(D)-96(D)

29	A:U w:s T	4TNA	8-21
30	A:C W:S C	1HNX	1280(A)-1149(A)
31	A:G S:S C	3CME	1797(O)-2483(O)
32	A:U W:S C	1F1T	11(A)-26(A)

3.2.2.2. Computational Details

Geometry optimization of the basepairs were performed without any constrains by different empirical, semi-empirical (AM1, PM3) (Dewar et al., 1985, Stewart, 1989) and *ab initio* quantum chemical methods namely MP2/6-31G** (Second order Møller–Plesset perturbation) (Møller and Plesset, 1934), HF/cc-pVDZ (Dunning correlation consistent basis set) (Dunning Jr, 1989, Kendall et al., 1992, Peterson et al., 1993), B3LYP/6-31G** (Becke, 1993, Lee et al., 1988) and GGA:PW91/DZP (Perdew et al., 1996) using AMBER (Case et al., 2004), GAMESS-US (Schmidt et al., 1993), GAUSSIAN-03 (Frisch et al., 2003) and ADF softwares (Fonseca Guerra et al., 1998, Te Velde et al., 2001). Huge variations of *ab-initio* methods in terms of both DFT functionals and basis sets are used commonly. We have chosen only those popular ones, which consider electron exchange and co-relation interactions, important for modeling the hydrogen-bonded systems. We did not consider dispersion correction, as its contribution is generally more significant in stacking interactions. We have optimized the basepairs using sander module of AMBER both in the gaseous medium and in solvated condition using Cornell *et. al* force field (Cornell et al., 1995). In the solvated conditions we have used periodic boundary condition along with Particle Mesh Ewald method for long-range electrostatics potential calculation.

The total interaction energy is calculated by the following method

$$E_{int} = E_{xy} - E_x - E_y \quad \text{.....[3.2.1]}$$

Where E_{xy}, E_x and E_y are total potential energies of the complex and the two component bases, respectively in their optimized geometry. The interaction energy calculated in this fashion is often overestimated due to basis set superposition error (BSSE) that arise due to overlapping of orbitals of component molecules in a complex system. We have calculated the single point BSSE corrected interaction energy by Boys-Bernardi function counterpoise method (Boys and Bernardi, 1970) using the method and basis set, which was used to optimize the structures. Therefore our calculated interaction energies are equivalent to MP2/6-31G**//MP2/6-31G**, HF/cc-pVDZ//HF/cc-pVDZ, etc., according to standard chemical notation and it also incorporates deformation component. Orientation parameters of the optimized basepairs were calculated by NUPARM (Mukherjee et al., 2006). The *ab initio* calculations by various groups predict that in the gas phase the amino groups of the nucleobases are intrinsically non-planar, due to a partial sp^3 hybridization of the nitrogen atoms. Pyramidalization generally means that the amino group's hydrogens deviate from the nucleobase plane in one direction, while the amino nitrogen moved to the opposite direction, this amino group nitrogen can also form very efficient out of plane hydrogen bonds, in addition to the base pairing interaction. Pyramidalization of the amino groups also arises due to intermolecular interactions. It is a general tendency of the amino groups attached to the nucleotide bases to undergo pyramidalisation for positioning of the lone pairs of electrons. However as we have optimized the structures by different methods, so to study whether these different methods maintain these degree of pyramidalisation equally or differently, we have also

calculated the amount of pyramidalisation of amino groups, following our earlier publications (Mukherjee et al., 2005), where the pyramidalization amount is calculated as

$$\theta = \cos^{-1}\left[\left(\overline{x_C} - \overline{x_N}\right)\cdot\{(\overline{x_{H1}} - \overline{x_N})\times(\overline{x_{H2}} - \overline{x_N})\}\right] - 90^{\circ} \qquad \text{.....[3.2.2]}$$

Where $\overline{x_{H1}}$, $\overline{x_{H2}}$, $\overline{x_N}$ and $\overline{x_C}$ are the coordinate vectors of the hydrogen, nitrogen and ring carbon atoms respectively.

The lone pairs of the amino groups are involved in conjugation with π-electrons clouds of the nucleotide rings. Due to this kind of delocalization the single bond between the carbon and nitrogen tends to acquire some amount of double bond characteristics, and to quantify the amount of double bond behavior of the C-N bond of the amino group, we have carried out bond-order analysis of the C-N bond. The bond-order values of the ring-nitrogen bonds were calculated by GAMESS using B3LYP/6-31G**, HF/cc-pVDZ and MP2/6-31G** methods from their corresponding optimized geometries. We have drawn the electro-static potential surfaces for all the basepairs optimized by B3LYP/6-31G** method using MOLDEN. The multipole derived method with contour value 0.02 is used for this purpose.

3.2.3 Results

3.2.3.1. Analysis of RNA Basepairs Crystal Structures

The program BPFIND has identified 11935 basepairs having two or more hydrogen bonds in the suggested non-redundant data set, which consists of 107 PDB files of RNA crystal structures. Out of these basepairs almost 67% is found to be canonical and the rest is non-canonical. Among all the 128 possible basepairs, we found 6 non-canonical basepairs, G:U W:W C (system 3), A:G H:S T (system 4), A:U H:W T (system

5), A:G W:W C (system 6), A:A H:H T (system 7) and A:G s:s T (system 20) have extremely high frequencies. The basepair formed by Hoogsteen edge of adenine and sugar edge of guanine in Trans orientation (A:G H:S T, system 4) is a remarkable example of purine-purine basepair having a percentage of occurrence as high as 14% of all non-canonical basepairs and about 5% of all basepairs found by BPFIND. It is observed that 81% non-canonical pairs are stabilized through two polar hydrogen bonds, while the remaining 19% involve at least one C-H...N/O type of hydrogen bond. Among the non-polar basepairs, A:G s:s T is found maximally, around 7.3% of total non-canonical basepairs. Numbers of occurrences of few less frequent basepairs are too small for any statistical analysis and we could not determine their best representative from the database. Hence, we have not considered these for further analysis through crystal structure database or quantum chemical studies.

We have analyzed structures of the frequently observed basepairs using six intra-basepair parameters (Table 3.2.1). According to IUPAC-IUB convention buckle and shear values of a Cis basepair undergo sign reversal when calculated from the other side while for a Trans basepair the sign of open-angle and stagger alter when they are calculated from opposite end (Mukherjee et al., 2006). We observe that most of the Cis basepairs have negative propeller twist, except few exceptions such as A:U s:w C, where large positive propeller twist is associated with large buckle also, probably due to involvement of out of plane sugar moiety. In most of the Trans basepairs, we found propeller twists are closer to zero, with exceptions in few basepairs like G:G S:S T (system 18) , U:U w:h T (system 25) and A:U w:s T (system 29). We get exceptionally high amount of propeller twist for G:G S:S T (system 18) and A:U w:s T (system 29)

basepairs, probably due to involvement of the bulky sugar edge in the basepair formation. Average values and standard deviations of buckle are found to be high in all the non-polar basepairs, probably due to less strengths of association. In A:G w:s C (system 23) basepair in addition to the high buckle values, variation in open-angle is also significantly higher. The open-angle values of the sugar-mediated basepairs are found to be generally higher, which is probably inherent due to their axis definition (Mukherjee et al., 2006). In case of G:U s:h C basepair (system 21), the crystal data shows high stretch values. A closer look at the examples of this basepair types found in our database indicates that these are generally present between consecutive residues of the same chain. In these dinucleotide platform basepairs, the backbone constraints probably force the bases to stay little far from each other, which have recently been analyzed by various groups (Sharma et al., 2010b, Lu et al., 2010b). Shear values are true indicators of hydrogen bonding type and its values are somewhat discrete – either close to zero or close to around +3 Å or –3 Å. We noticed many of the observed basepairs are parts of base triplets (Table 3.2.1). Significant frequencies of such triplets are formed by A:A H:W T (system 9), A:U H:W C (system 10) and A:A W:W T (system 14) basepairs and this induces alteration in the structures of the basepairs. As a result of triplet formation, propeller, buckle and stagger values of these basepairs assume more variability as reflected by their larger standard deviations.

Our final set has 19 polar, 8 non-polar C-H…O/N mediated and 5 sugar-mediated basepairs.

3.2.3.2. Computational Analysis

Evaluation of variation of geometry by different methods

Similar to as described in section 3.1, we observed that most of the basepairs maintain close to their initial geometry even after unconstrained optimization by B3LYP/6-31G** as shown in Figure 3.2.1. This possibly indicates that at least two hydrogen bonds between the molecules may be a necessity and sufficient between planar moieties for molecular recognition. It was also observed that the other quantum chemical methods, namely HF/cc-pVDZ, MP2/6-31G**, GGA:PW91/DZP and PM3 also show similar behavior. Most of the basepairs deviate significantly after minimization by AMBER in presence of water. It was observed that the bases form hydrogen bonds with water more strongly than with the counter base. As the force field parameters are obtained by comparison with well-studied Watson-Crick basepaired double helices, the empirical molecular mechanics method is capable to give meaningful properties for similar molecules. A more extensive tuning of the force fields considering the non-Watson Crick basepaired system is perhaps required. Although structures of the PM3 optimized basepairs generally follow the trend, the sugar-mediated basepairs are rather unusual.

The sugar pucker values of the structures optimized by both the semi-empirical methods are mostly found to be in the disallowed O4'-exo regions, associated with very small amplitude of pucker. Among the *ab initio* optimized structures, B3LYP/6-31G** gives relatively smaller amplitude of pucker (around 35^{o}) as compared to the other methods as well as that of crystallographic average, but all the methods retain the sugars in the usual C3'-endo region (Table 3.2.3).

Table 3.2.3. Sugar-pucker pseudorotation amplitude and phase of the involved sugars in the sugar mediated basepair systems along with crystallographic average and standard deviation.

		(System 28)	(System 29)	(System 30)	(System 31)	(System 32)
	Method	A:A W:S C	A:U w:s T	A:C W:S C	A:G S:S C	A:U W:S C
Amplitude	MP2/6-31G**	39.3	39.4	38.2	42.1	39.2
Phase		9.1	13.2	16.5	8	15.6
Amplitude	B3LYP/6-31G**	35.2	35.1	33.8	38.7	34.1
Phase		12.1	14.5	18.8	8.6	17.2
Amplitude	HF/cc-pVDZ	36.9	36.6	36.4	39.1	40.1
Phase		13.7	18.2	18.6	11.8	17.1
Amplitude	GGA:PW91/DZP	40.6	39.4	40.3	44.1	43.4
Phase		12.1	11	11.3	17.7	10.1
Amplitude	AM1	8.7	18.1	9.5	9.7	18.5
Phase		-48.4	-16.1	-30.9	-70.9	3.3
Amplitude	PM3	5.2	14.1	8.7	8.8	26
Phase		78.5	38.9	11.5	104.7	33.1
Amplitude	AMBER	45.3	37.8	36.6	39	41.7
Phase		22.1	22.7	16.1	30.1	11.2
Amplitude	Crystal Data Set	38.1 (3.6)	39.4 (5.2)	38.4 (2.7)	36.9 (3.8)	39.2 (3.3)
Phase	Avg. (Std.)	13.9 (10.4)	20.2 (28.3)	14.5 (4.9)	41.4 (55.4)	18.5 (7.5)

Structures of all the basepairs optimized by B3LYP/6-31G** method, along with their electrostatic potential surface are shown in Figure 3.2.1. The A-form double helical structures of RNA, with only Watson-Crick basepairs, generally expose their monotonous minor groove only. Presence of the non-canonical basepairs can introduce variation to such surfaces due to alteration in their shape, size and unusual functional groups. These basepairs now expose their unusual edge (such as Watson-Crick edge) also to the solvent or incoming ligand, which can be important for molecular recognition.

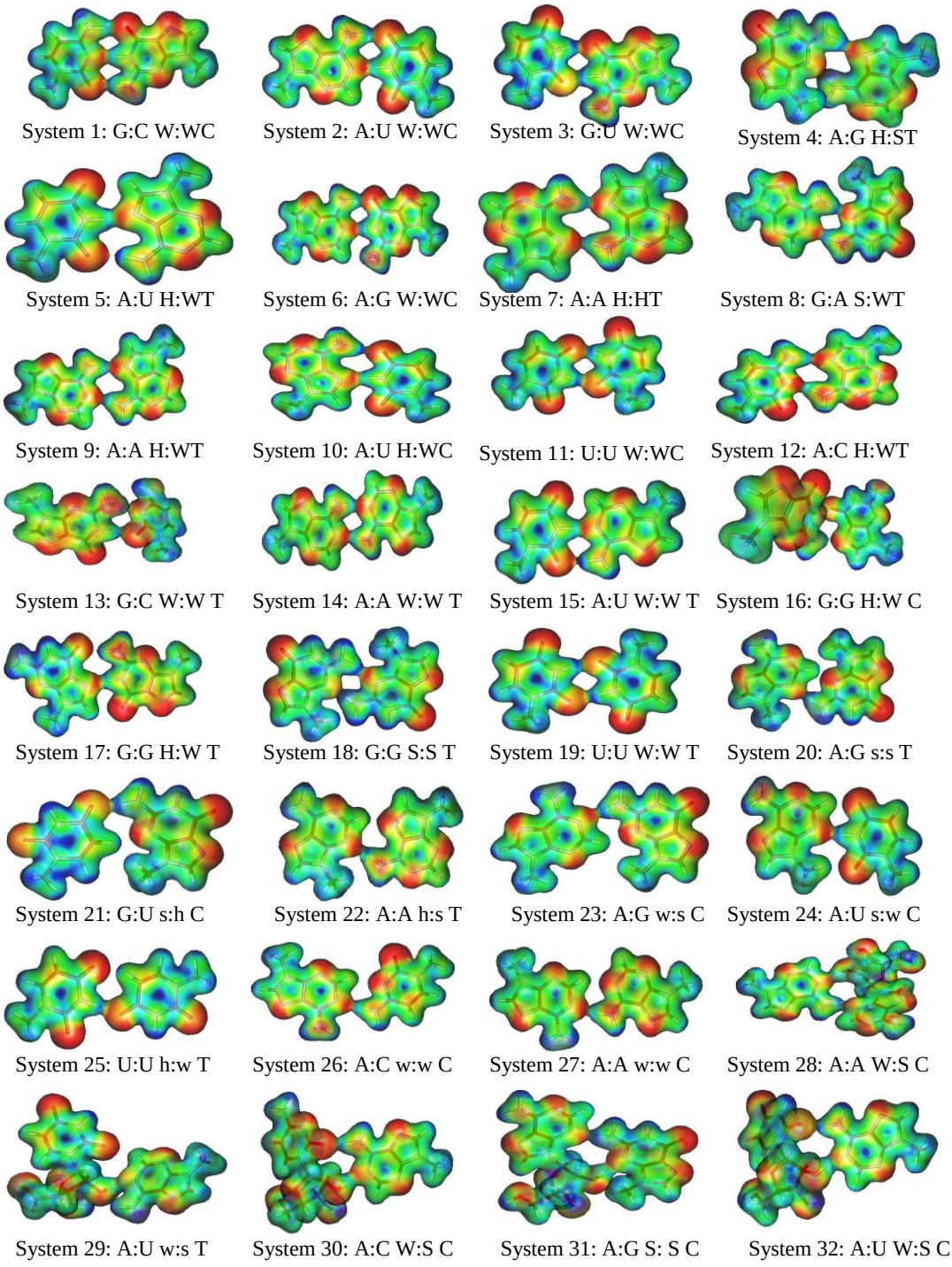


Figure 3.2.1. Electro-static potential surfaces of the 32 basepairs optimized by B3LYP/631G** method, where potential < -0.10 (red), -0.0500 (yellow), 0.0000 (green), 0.0500 (light blue), >0.1000 (dark blue).

Some of these, e.g., A:G W:WC (system 6), A:G w:s C (system 23), A:C w:w C (system 26), A:G S:SC (system 31), expose up to six free hydrogen bonding donor or acceptor sites in the major groove, as compared to only three such sites for canonical ones (Figure 3.2.1). In addition, some of the C-H groups also have high electropositive potential, capable of formation of hydrogen bonds. Overlapping of the molecular orbitals is more prominent in case of N-H…N/O type of hydrogen bonding whereas it is negligible is case of C-H…O/N type of hydrogen bonding where the H…O/N distances are comparatively larger.

All the hydrogen bond lengths and angles for each of the basepairs were calculated from the optimized geometry and their average values are given in Table 3.2.4. For the polar hydrogen bonds only those are tabulated where the hydrogen to acceptor distances are smaller than 2.5 Å and angles at hydrogen are larger than 150^{o}. However in case of C-H…O/N type interactions, we have relaxed our criteria for the analysis (Distance cutoff = 4.0 Å and angle cutoff = 90^{o}), as these are normally long. It is observed that the polar hydrogen bond lengths and angles are very similar by all the empirical, semi empirical and *ab initio* methods. The hydrogen bond lengths of the HF/cc-pVDZ optimized structures are slightly larger than the other methods, while GGA:PW91/DZP method gives shortest hydrogen bonds. The average hydrogen bond angles for polar types are generally close to 170^{o}. The angles at the weak C-H…O/N interactions are mostly smaller than 150^{o}, although linear hydrogen bond like interaction is also observed.

Table 3.2.4. Average and standard deviation of the different types of H-bond lengths (Å) and angles (^{o}) derived from the geometry-optimized structures obtained by different methods, the standard deviation values of the respective parameters are given in parenthesis.

Method	N-H...O (20)#	Angle	N-H…N (30)#	Angle	O-H…N (5)#	Angle	C-H…O (3)#	Angle	C-H…N (6)#	Angle
MP2/6-31G**	1.90 (0.1)	170.6 (6.4)	1.95 (0.09)	172.2 (6.2)	1.81 (0.07)	160.4 (9.7)	2.28 (0.14)	152.1 (13.3)	2.40 (0.04)	159.5 (11.0)
HF/cc-pVDZ	2.01 (0.10)	170.0 (5.1)	2.13 (0.09)	173.2 (6.0)	1.97 (0.04)	166.4 (2.6)	2.35 (0.04)	151.0 (4.4)	2.67 (0.06)	165.7 (2.3)
B3LYP/6-31G**	1.89 (0.11)	171.9 (4.4)	1.95 (0.12)	173.1 (5.6)	1.78 (0.04)	168.7 (4.1)	2.33 (0.17)	143.4 (15.2)	2.60 (0.16)	168.0 (4.3)
GGA:PW91/DZP	1.82 (0.08)	172.1 (4.1)	1.88 (0.09)	171.4 (6.3)	1.72 (0.04)	163.8 (1.1)	2.23 (0.08)	151.1 (8.5)	2.41 (0.08)	163.8 (1.1)
AMBER	1.87 (0.06)	169.0 (8.1)	1.98 (0.06)	170.1 (5.5)	1.91 (0.03)	159.9 (3.6)	2.52 (0.02)	138.5 (8.9)	2.69 (0.05)	159.9 (3.6)
AMBER W$^{\$}$	1.93 (0.08)	161.8 (7.8)	2.05 (0.09)	163.0 (7.7)	2.03 (0.01)	151.8 (3.2)				
PM3	1.81 (0.01)	170.2 (7.3)	1.82 (0.04)	169.0 (8.3)	1.96 (0.31)	167.8 (2.9)	2.53 (0.07)	141.7 (16.4)	2.36 (0.41)	167.9 (3.0)

$^{\$}$ Denotes the values are obtained in solvated medium

# Denotes no of such hydrogen bond observed in the optimized structures.

It is a general tendency of the hydrogen atoms of primary amino groups attached to the nucleotide bases to adopt pyramidal geometry for positioning of the lone pairs of

electrons. As formation of H-bond causes electron transfer between the amino groups and the acceptor atoms, it may affect double bond characters of the C-N bonds and hence pyramidalisation (Mukherjee et al., 2005). We have therefore analyzed the pyramidal values (out of plane positioning of the hydrogen atoms) for the hydrogen bonded amino groups and compared with those of the free amino groups (Table 3.2.5). As reported earlier (Mukherjee et al., 2005), guanine 2-amino group shows maximum pyramidalisation in all the optimization methods possibly due to proximity of a secondary amino group at N1 position. The MP2/6-31G** method gives rise to more pyramidal amino groups to all the three isolated bases. The pyramidal character of the amino groups generally reduces when it is engaged in strong hydrogen bonding. The amino groups retain significant non-planarity in some of the optimized basepairs, eg. A:G H:S T (system 4), G:C W:W T (system 13) G:G H:W C (system 16), G:G S:S T (system 18), etc. Pyramidalization of the 2-amino group of guanine in A:G W:W C basepair increases from its value in isolated condition, possibly to avoid steric clash with the nearby C2H group of adenine. In most of the basepairs, pyramidalisation values obtained from the B3LYP/6-31G**, MP2/6-31G** and HF/cc-pVDZ are found to be highly correlated with each other with correlation coefficients close to 1. The lone pair electrons of the nitrogen of the amino groups are also involved in extended conjugation with π–electron clouds of the nucleotide rings. Due to this kind of delocalization the single bonds between the carbon and nitrogen tend to acquire some amount of double bond characteristics. We have carried out analysis of the bond-order of C-N bonds to quantify the amount of double bond characters of the C-N bonds connecting the amino group to the base ring. As expected the bond order and pyramidalisation values are anti-correlated to each other.

The double bond characters are more in the C-N bond in the HF/cc-pVDZ optimized structures, which signifies significant amount of electron transfer from the lone pairs of nitrogen to the ring carbons. Thus all the *ab initio* methods are highly reliable in predicting electron delocalization and hence structures of the basepairs, although the degree of delocalization depends on the methods to some extent.

Table. 3.2.5. Pyramidalization and bond order values of the bases and basepairs obtained through various quantum chemical methods

	Base/basepair	MP2/6-31G**		HF/cc-pVDZ		B3LYP/6-31G**		GGA: PW91/ DZP[#]
		Bond Order	Pyr (^{O})	Bond Order	Pyr (^{O})	Bond Order	Pyr (^{O})	Pyr (^{O})
	Adenine (isolated)	1.066	32.28	1.337	29.02	1.213	28.54	23.25
	Guanine (isolated)	1.032	45.87	1.235	40.28	1.143	42.7	42.51
	Cytosine (isolated)	1.066	34.23	1.292	28.17	1.199	30.38	27.59
1	G:C W:WC	1.107	30.91	1.337	19.35	1.250	16.27	0.02
		1.191	8.88	1.436	3.54	1.354	0.99	0.02
2	A:U W:W C	1.129	14.69	1.427	0.05	1.315	0.66	2.36
3	G:U W:W C[$]	1.049	39.96	1.257	35.30	1.158	36.86	35.58
4	A:G H: S T	1.107	28.52	1.378	22.84	1.279	18.40	0.71
		1.099	38.67	1.300	32.41	1.230	32.17	27.51
5	A:U H:W T	1.105	16.46	1.401	10.99	1.272	3.46	4.29
6	A:G W:W C	1.151	22.34	1.437	20.36	1.338	5.93	8.54
	[$]	1.033	44.89	1.244	40.88	1.151	42.03	43.03
7	A:A H:H T	1.112	29.45	1.382	24.66	1.285	17.31	7.11
		1.113	29.54	1.382	24.75	1.285	17.31	5.90
8	G:A S:W T	1.101	36.97	1.307	31.42	1.240	28.37	28.35
		1.123	17.87	1.412	14.86	1.298	6.28	16.11
9	A:A H: W T	1.118	24.16	1.401	18.31	1.305	11.25	14.52
		1.133	24.44	1.414	18.43	1.299	11.34	6.95
10	A:U H:W C	1.113	11.78	1.412	0.33	1.295	2.76	3.19
12	A:C H:W T	1.146	9.40	1.450	0.01	1.329	0.11	1.00
		1.144	16.32	1.407	0.04	1.301	0.07	10.19
13	G:C W:W T	1.095	50.87	1.300	29.80	1.189	38.63	31.67
	[$]	1.080	40.41	1.321	22.09	1.220	22.71	18.44
14	A:A W:W T	1.138	16.06	1.434	1.69	1.310	9.94	1.85
		1.138	15.95	1.434	13.65	1.310	9.83	1.41
15	A:U W:W T	1.120	18.16	1.420	0.99	1.283	0.64	6.95
16	G:G H:W C	1.080[$]	35.60	1.295[$]	28.35	1.165[$]	41.64[$]	27.44

		1.080[$]	0.44	1.296[$]	28.29	1.186	37.71	38.02[$]
17	G:G H:W T	1.076	33.81	1.299	26.61	1.201	28.90	28.49
	[$]	1.039	44.60	1.244	39.39	1.153	41.23	40.78
18	G:G S:S T	1.093	35.91	1.302	26.69	1.225	29.21	26.96
		1.093	35.89	1.303	26.76	1.225	29.07	32.18
20	A:G s:s T [$]	1.059	35.60	1.356[$]	24.76	1.225	22.80	9.21
		1.041	44.42	1.274	37.09	1.194	38.82	38.41
21	G:U s:h C	1.089	38.11	1.302	32.35	1.226	31.90	27.15
22	A:A h:s T	1.102	23.92	1.385	19.28	1.275	15.16	5.61
	[$]	1.071	30.67	1.346	27.01	1.218	26.15	18.87
23	A:G w:s C [$]	1.065	26.09	1.344	26.22	1.212	28.35	26.85
		1.072	41.70	1.278	35.78	1.212	30.41	32.08
24	A:U s:w C [$]	1.072	30.44	1.344	27.95	1.217	28.02	19.78
26	A:C w:w C [$]	1.053	37.66	1.317	34.73	1.198	32.24	31.08
		1.117	29.01	1.349	21.15	1.279	5.96	17.63
27	A:A w:w C	1.119	19.24	1.400	19.64	1.283	16.21	9.25
	[$]	1.061	33.99	1.332	31.34	1.205	26.90	28.17
28	A:A W:S C	1.126	28.57	1.407	25.02	1.288	23.71	12.52
	[$]	1.071	28.85	1.354	25.93	1.223	28.94	18.48
29	A:U w:s T[$]	1.118	32.05	1.350	25.56	1.227	22.35	20.07
30	A:C W:S C	1.144	19.66	1.453	14.83	1.326	7.73	5.70
	[$]	1.075	32.17	1.310	29.94	1.210	28.44	23.44
31	A:G S:S C [$]	1.079	28.40	1.361	24.20	1.226	23.00	14.50
		1.080	30.81	1.307	31.07	1.221	30.34	22.01
32	A:U W:S C	1.130	21.53	1.410	22.71	1.297	14.48	11.91

[$] Represents the free amino groups

[#] Bond order values could not be calculated by this method

Interaction Energies of the Basepairs

The BSSE and deformation corrected interaction energies are analyzed to understand the strength of association and stability of the basepairs. We find that all the adopted methods give very similar trends in their interaction energy values for most of the basepairs (Table 3.2.6). The weaker interaction strengths are well correlated with larger hydrogen bond lengths for the HF/cc-pVDZ optimized structures. As expected the non-polar basepairs have poor interaction energy as compared to most of the polar

basepairs. In most of the basepairs the interaction energies calculated by the B3LYP/6-31G** and MP2/6-31G** methods follow similar trend, the latter being slightly stronger in all the cases. Several non-canonical basepairs, such as A:U H:W C (system 10), A:C H:W T (system 12), G:C W:W T (system 13) etc. are found to be very stable by all the adopted methods, with interaction energy values even better than those of the usual A:U W:W C and G:U W:W C basepairs. Some of the polar basepairs like A:G H:S T (system 4), A:A H:H T (system 7) and G:G S:S T (system 18) show comparatively poor interaction energies as obtained by all the methods. There are two reasons for poor interaction energy values of G:G S:S T (system 18) basepair: i) absence of the stronger N-H...O or O-H…N hydrogen bonds and ii) steric contacts between N-methyl hydrogen and amino hydrogen of the other base. In G:G S:S T and A:G H:S T basepairs methyl group of one of the bases come in close contact with the amino group of the other base, which forces the structures to be slightly deformed (Figure 3.2.1). These recurrently found basepairs might get additional stability from a third hydrogen bond involving sugar (Halder and Bhattacharyya, 2010), which we did not consider in the present study. In case of A:A H:H T (system 7) basepair, there is no scope of steric contact but presence of only two N-H...N hydrogen bonds could not provide sufficient strength to its stability. The G:G H:W C basepair (system 16) is found to be highly stable in case of MP2/6-31G** and HF/cc-pVDZ methods. Furthermore this basepair has the tendency to convert to the W:W T type forming two stronger N-H…O type hydrogen bonds.

Most of the C-H…O/N mediated non-polar basepairs remain in the low stability regions as obtained by all the methods. This is due to the fact that C-H…O bond is weak in nature and has energy component lower by about 6 kcal/mol than that of the N-H…N

hydrogen bond (Roy et al., 2008). The interaction energies, calculated by different methods, for the non-polar basepairs differ significantly (Table 3.2.6). This indicates that strength of the C-H…O interaction is probably considered unequally by the quantum chemical methods. Furthermore, acidity of the C-H groups may also depend on environment (Banerjee and Bhattacharyya, 2008). Among the sugar mediated basepairs, A:C W:S C (system 30) is found to be stable by all the methods. For the other sugar-mediated basepairs, interaction energies calculated by the all the adopted methods, do not show consistent nature. As for example, the interaction energy of A:U W:S C (system 32) basepair obtained by MP2/6-31G**, HF/cc-pVDZ, B3LYP and GGA:PW91/DZP methods are -9.0, -6.0, -14.6 and -8.6 kcal/mol, respectively. This is possibly due to extra flexibility of the sugar moiety and glycosidic torsion angle, leading to different optimized geometries by different methods. The MP2/6-31G** method results in non-planar geometry, associated with high buckle value (the final optimized structure does not maintain any hydrogen bond between the constituent bases). However optimizations by HF/cc-pVDZ, GGA:PW91/DZP and B3LYP/6-31G** methods maintain the expected hydrogen bonding patterns through N6-H6…O2 and O2'-H2'…N1, but the structures do not remain planar.

Table. 3.2.6. Interaction Energy (in kcal/mol) of the optimized basepairs by different methods

No	Basepair	MP2/631G**	HF/cc-pVDZ	B3LYP /6-31G**	GGA:PW91/DZP	PM3	AMBER
1	G:C W:W C	-24.46	-22.05	-24.18	-25.57	-15	-26.11
2	A:U W:W C	-12.21	-8.81	-11.64	-13.56	-6.9	-12.44
3	G:U W:W C	-12.83	-10.69	-12.20	-12.73	-6.99	-14.26
4	A:G H: S T	-9.44	-5.79	-8.06	-10.96	-3.69	-10.69
5	A:U H:W T	-12.8	-9.28	-11.39	-13.22	-7.76	-12.81
6	A:G W:W C	-14.61	-10.81	-13.63	-15.71	-8.87	-15.66
7	A:A H:H T	-9.73	-5.86	-8.49	-11.45	-4.04	-9.76
8	G:A S:W T	-10.08	-6.6	-8.98	-11.68	-5.34	-11.09
9	A:A H: W T	-10.54	-6.59	-9.61	-12.68	-5.4	-10.72
10	A:U H:W C	-13.03	-9.4	-12.06	-13.88	-7.82	-13.40
11	U:U W:W C	-9.99	-7.70	-9.64	-9.85	-4.98	-11.14
12	A:C H:W T	-13.91	-8.16	-12.71	-15.21	-8	-14.14
13	G:C W:W T	-13.08	-14.45	-11.83	-11.96	-5.60	-16.98
14	A:A W:W T	-11.08	-7.18	-10.27	-13.33	-5.68	-11.54
15	A:U W:W T	-11.7	-8.45	-10.51	-12.25	-6.75	-11.88
16	G:G H:W C	-21.38	-20.21	-11.33	-11.67	-6.02	-15.55
17	G:G H:W T	-16.79	-14.25	-15.25	-16.43	-9.21	-17.06
18	G:G S:S T	-7.78	-4.65	-6.50	-9.88	-2.38	-9.96
19	U:U W:W T	-9.77	-7.64	-9.00	-8.98	-5.08	-11.11
20	A:G s:s T	-7.53	-3.95	-4.56	-7.15	-2.5	-11.38
21	G:U s:h C	-6.94	-5.04	-6.01	-7.11	-3.93	-7.39
22	A:A h:s T	-5.39	-3.08	-4.01	-6.58	-1.98	-10.48
23	A:G w:s C	-6.31	-4.01	-4.61	-6.72	-2.59	-8.10
24	A:U s:w C	-8.58	-5.71	-6.89	-8.55	-4.12	-9.08
25	U:U w:h T	-7.62	-5.75	-6.85	-6.95	-3.58	-8.01
26	A:C w:w C	-7.01	-4.91	-5.34	-7.22	-2.17	-5.01
27	A:A w:w C	-5.78	-3.44	-4.53	-6.42	-1.84	-5.42

28	A:A W:S C	-15.81	-10.40	-14.76	-20.35	-7.18	-10.03
29	A:U w:s T	-11.8	-9.22	-10.64	-11.93	-6.13	-8.80
30	A:C W:S C	-14.17	-10.72	-10.00	-14.37	-8.96	-15.01
31	A:G S:S C	-9.33	-5.18	-7.78	-12.43	-3.66	-16.10
32	A:U W:S C	-9.02	-6.07	-14.55	-8.55	-5.67	-13.84

Superposition and RMSD Calculation

The variations in interaction energies of a basepair optimized by different methods can be due to the difference in the geometry of the optimized structures or the method itself. We have therefore analyzed these structural variations in terms of the Root Mean Square Deviations (RMSD) of the basepairs (Table 3.2.7), using CHARMM (Brooks et al., 1983). Few representatives of polar, non-polar and sugar-mediated basepairs optimized by different methods are superimposed on their respective B3LYP/6-31G** optimized geometry as given in Figure 3.2.2. We notice that all the RMSD values are generally small for the polar basepairs with few exceptions.

We observed large RMSD between the HF/cc-pVDZ and GGA:PW91/DZP, and between the GGA:PW91/DZP and MP2/6-31G** for A:G H:S T (system 4) and A:A H:H T (system 7) basepairs. These two basepairs also have poor interaction energies and highly non-planar amino groups. It is possible that the hydrogen bonds in these basepairs are weaker compared to the other polar ones, which allowed them to rotate freely. The degrees of non-planarity of the two bases in the basepairs also differ significantly in these cases. Among the polar basepairs G:C W:W T (system 13) and G:G H:W C (system 16)

show unexpectedly high RMSD values. This is correlated to their dissimilar interaction energy value trends.

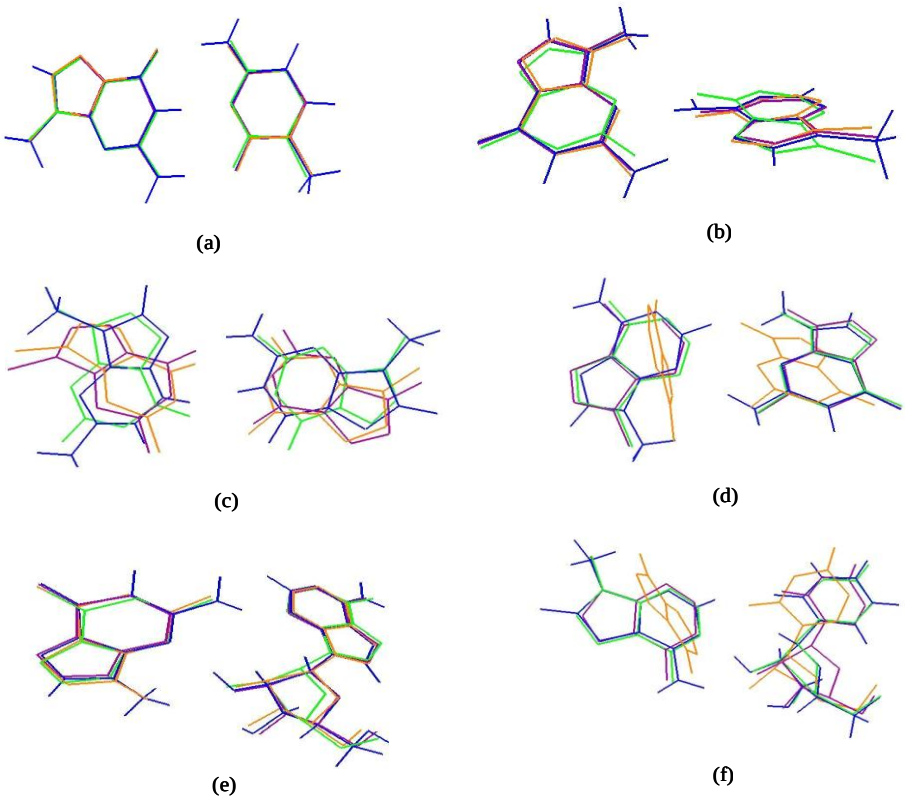


Figure 3.2.2. Superposition of the (a) G:C W:W C (b) A:G H:S T (c) G:G H:W C (d) A:G s:s T (e) A:G S:S C (f) A:U w:s T basepairs optimized by B3LYP/6-31G** (blue), HF/cc-pVDZ (purple), GGA/PW91 (green), MP2/6-31G** (orange).

The RMSD values for the non-polar basepairs are found to be larger than those of the polar ones. In case of A:G s:s T (system 20) basepair, we get large RMSD when the MP2/6-31G** optimized structure participates in superposition.

Table. 3.2.7. RMSD (Å) of the optimized basepairs obtained by different methods

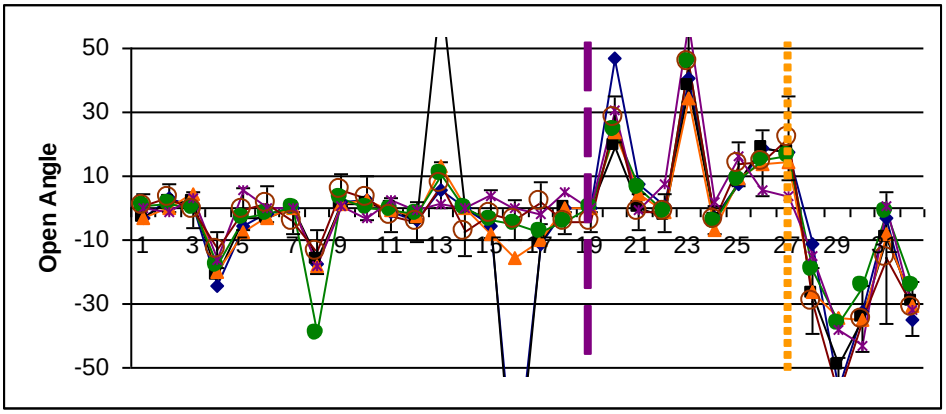
Serial No	Basepair	B3LYP/6-31G**—HF/cc-pVDZ	B3LYP/6-31G**—GGA:PW91	B3LYP/6-31G**—MP2/6-31G	HF/cc-pVDZ—GGA:PW91	HF/cc-pVDZ—MP2/6-31G	GGA:PW91—MP2/6-31G
1	G:C W:W C	0.13	0.15	0.19	0.12	0.01	0.13
2	A:U W:W C	0.07	0.09	0.10	0.12	0.11	0.12
3	G:U W:W C	0.10	0.12	0.15	0.12	0.12	0.09
4	A:G H: S T	0.11	0.34	0.19	0.39	0.21	0.51
5	A:U H:W T	0.15	0.14	0.13	0.09	0.08	0.05
6	A:G W:W C	0.18	0.11	0.15	0.17	0.09	0.11
7	A:A H:H T	0.17	0.32	0.29	0.43	0.17	0.56
8	G:A S:W T	0.11	0.22	0.11	0.24	0.12	0.13
9	A:A H: W T	0.11	0.10	0.20	0.16	0.15	0.18
10	A:U H:W C	0.09	0.05	0.05	0.10	0.11	0.08
11	U:U W:W C	0.13	0.06	0.13	0.13	0.10	0.09
12	A:C H:W T	0.06	0.17	0.06	0.19	0.09	0.16
13	G:C W:W T	0.85	0.34	2.06	1.00	2.12	2.22
14	A:A W:W T	0.19	0.17	0.17	0.13	0.11	0.11
15	A:U W:W T	0.12	0.11	0.10	0.13	0.13	0.05
16	G:G H:W C	1.60	1.12	1.35	1.84	0.58	1.92
17	G:G H:W T	0.08	0.14	0.34	0.20	0.32	0.44
18	G:G S:S T	0.18	0.24	0.18	0.20	0.11	0.22
19	U:U W:W T	0.08	0.13	0.02	0.17	0.08	0.14
20	A:G s:s T	0.35	0.29	1.61	0.20	1.84	1.71
21	G:U s:h C	0.15	0.35	0.08	0.41	0.22	0.36
22	A:A h:s T	0.35	0.26	0.24	0.14	0.25	0.18
23	A:G w:s C	0.45	0.48	0.50	0.74	0.14	0.74
24	A:U s:w C	0.18	0.24	0.28	0.16	0.13	0.11
25	U:U w:h T	0.08	0.10	0.08	0.12	0.09	0.05
26	A:C w:w C	0.18	0.18	0.24	0.11	0.10	0.09
27	A:A w:w C	0.15	0.11	0.18	0.22	0.29	0.12
28	A:A W:S C	0.10	0.70	0.92	0.64	0.86	0.29
29	A:U w:s T	0.78	0.24	1.25	0.76	1.60	1.38
30	A:C W:S C	0.36	0.69	1.06	0.44	0.80	0.60
31	A:G S:S C	0.19	0.53	0.22	0.41	0.32	0.52
32	A:U W:S C	0.12	0.84	0.82	0.80	0.86	1.50

Analysis of basepair parameters

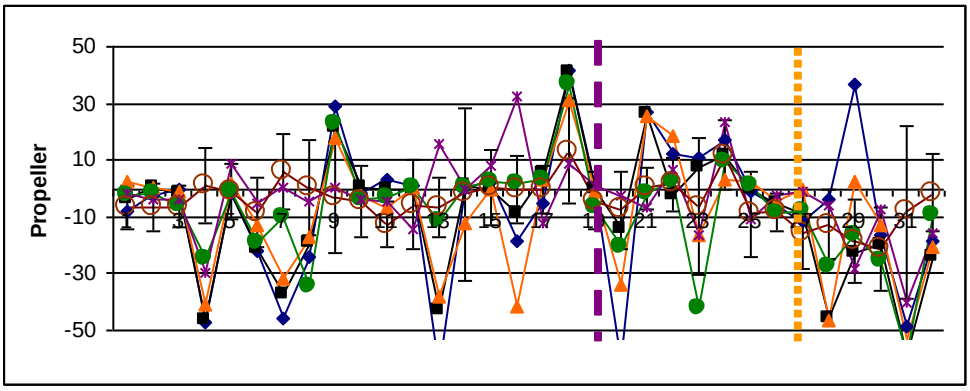
We have analyzed the basepair orientation parameters, namely buckle, open-angle, propeller, stagger, shear and stretch, for all the basepairs optimized by different methods (Table 3.2.8). Comparisons of buckle and stagger values indicate their random nature. The remaining parameters namely propeller, open-angle, shear and stretch follow well defined nature of variations and found to be close to the average values as obtained from crystal structure analysis. Their values reflect nature and type of hydrogen bonds holding the basepairs as shown in Figure 3.2.3. The propeller values for the three usual canonical basepairs G:C W:W C, A:U W:W C and G:U W:W C are close to zero and slightly negative in nature.

Among the optimized basepairs, we get large positive propeller values in case of A:A H:W T (system 9) and G:G S:S T (system 18) basepairs as obtained by all the methods. The shear values are indicative of positions of the hydrogen bonding atoms in a basepair and hence they vary systematically within a wide range between -3 Å and +3 Å. Comparatively larger values of open-angle and shear are often indicative of hydrogen bonding characteristics of the basepairs and not of their distorted geometry. Two different types of basepairing patterns of U:U W:W T basepair are possible (Figure 3.2.4), but one of them has been observed more frequently in the crystal structures of RNA.

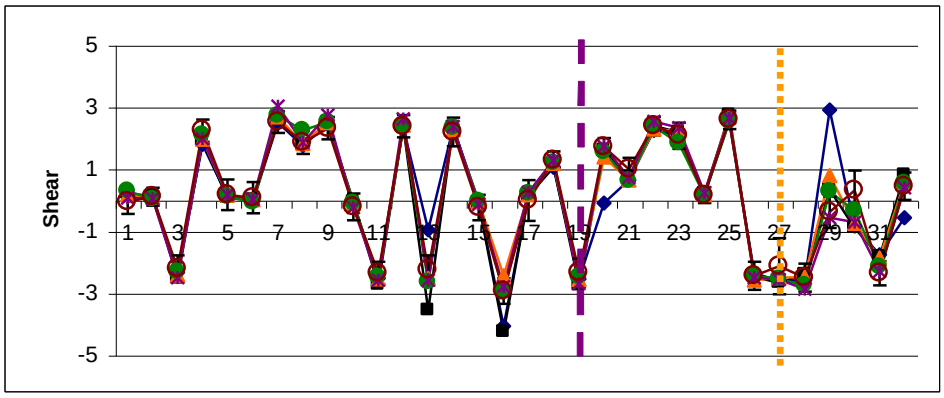
(a)


Open Angle
50
30
10
-10
-30
-50
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
31

(b)


Propeller
50
30
10
-10
-30
-50
1
3
5
9
15
21
25
29
31

(c)


Shear
5
3
1
-1
-3
-5
1
3
5
7
9
11
15
17
21
23
25
27
29
31

(d)

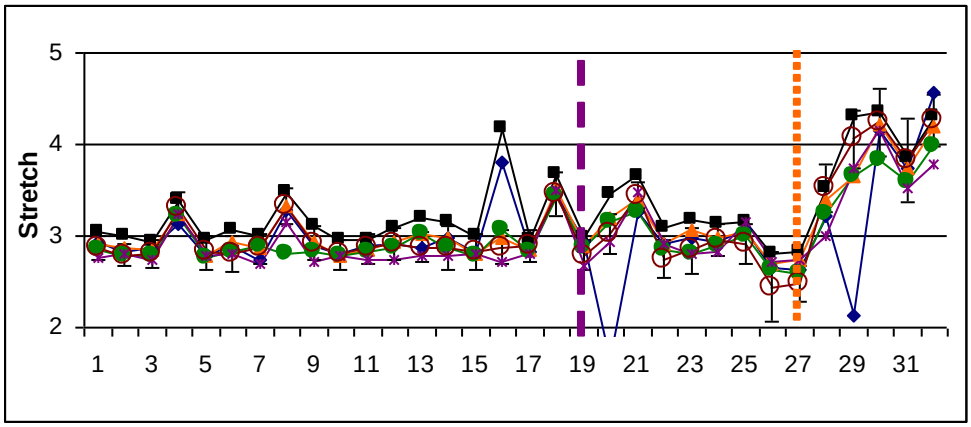


Figure 3.2.3. Graphs of (a) Open-angle (°), (b) Propeller (°), (c) Shear (Å) and (d) Stretch (Å) of the 32 optimized basepairs obtained by MP2/6-31G** (blue♦), HF/cc-pVDZ (black --■--), B3LYP/6-31G** (red ▲), GGA:PW91/DZP (green ●), PM3 (violet *), crystal average (maroon ○) , standard deviation of the crystal average (error bars), where up to the broken violet line (systems 1-19) we have the N-H…O/N mediated basepairs, up to the dotted orange line we have the C-H…O/N mediated basepairs (systems 20-27), rest are the sugar mediated basepairs (systems 28-32).

We have modeled and optimized the other one as shown in Figure 3.2.4. Both the structures have two N-H…O hydrogen bonds and large shear values of opposite signs. The probable alteration in shear for transition of U:U W:W T basepair from one to another form was however not observed. The open-angle and stretch values also follow similar trend with little wider variability. Stretch values of most of the polar and C-H…O mediated basepairs lie around 3 Å, but in case of some sugar mediated basepairs stretch values are found to be greater than 3.8 Å. Stretch values obtained from the HF/cc-pVDZ optimized structures are found to be comparatively larger than the other methods for most of the basepairs due to the larger hydrogen bond lengths predicted by this method. The MP2/6-31G** optimized structures of the A:G s:s T (system 20) and A:U w:s T (system 29) basepairs are found to be significantly different from rest of the methods. The shear

value of A:G s:s T (system 20), for example, is close to zero for the MP2/6-31G** optimized structure while proper hydrogen bonding demands that its shear be near 2 Å. There is a possibility of formation of one C-H…N type of hydrogen bond in this basepair and MP2/6-31G** method apparently considered this interaction in a repulsive manner. This correlates well with its RMSD values also. Propeller values of the A:G w:s C (system 23) basepair is found to be flexible: in some methods like in MP2/6-31G** and HF/cc-pVDZ method it has positive propeller values, while in case of B3LYP/6-31G** and GGA:PW91/DZP method it has negative propeller values. This is probably due to weaker C-H…N interaction as predicted in the previous section. Among the sugar mediated basepairs, propeller values of the MP2/6-31G** optimized structures of A:U w:s T (system 29) and A:G S:S C (system 31) show large deviation from the crystal average values, indicating flexibility of the ribose sugar.

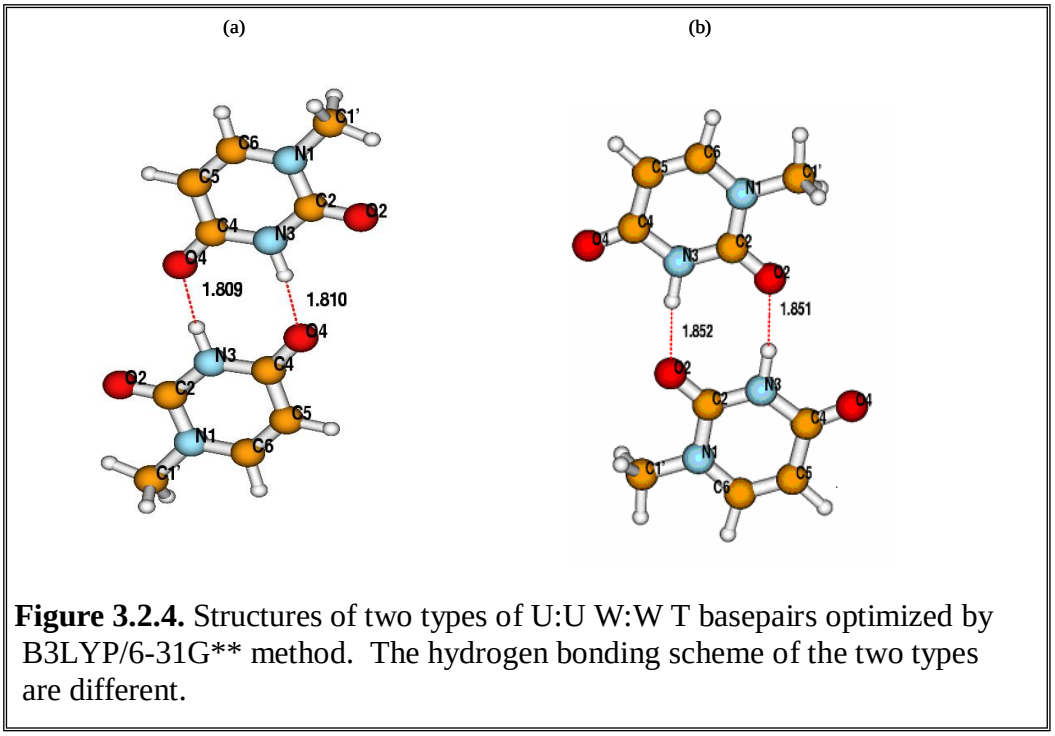


Figure 3.2.4. Structures of two types of U:U W:W T basepairs optimized by B3LYP/6-31G** method. The hydrogen bonding scheme of the two types are different.

Two quite frequent polar basepairs, namely G:C W:W T (system 13) and G:G H:W C (system 16) show unusual geometry after optimization by various methods, as denoted by their RMSD values also. In G:G H:W C (system 16), for example, one of the O6G does not form any hydrogen bond. Large negative charge of this oxygen can perturb the structure significantly and we noticed alternate hydrogen bonding scheme of the optimized structures, particularly by MP2/6-31G** and HF/cc-pVDZ. Optimizations by these methods lead to W:W T orientation having two N-H…O hydrogen bonds involving both the O6G atoms. The basepair parameters considering H:W C basepairing edge specific axis system are very large (-94.54^{o} open-angle and 4.18 Å stretch, for example, for the HF/cc-pVDZ optimized structure) while the parameters considering W:W T edge specific axis system are small, indicating fairly strong basepairing in the optimized states (Table 3.2.8).

Table. 3.2.8. Intra Basepair Parameters of the Basepairs in the optimized geometries obtained by different methods and the average and standard deviations of the basepair parameters in the crystal dataset.

Serial No.	Basepair	Method	Buckle (^{O})	Open Angle (^{O})	Propeller (^{O})	Stagger (Å)	Shear (Å)	Stretch (Å)
1	G:C W:W C	MP2/6-31G**	-4.92	-3.08	-7.88	-0.14	0.12	2.93
		HF/cc-pVDZ	-4.2	-3.04	-3.63	-0.09	0.21	3.04
		B3LYP/6-31G**	-0.3	-3.43	2.21	0.12	0.17	2.92
		PW91/DZP	-1.5	0.48	-2.37	-0.37	0.29	2.85
		AM1	51.24	0.47	1.66	-0.69	0.2	2.98
		AMBER	-3.28	-2.21	2.54	0	0.34	2.91
		AMBER_WATER	18.13	6.29	19.8	0.74	-0.64	2.75
		PM3	-3.98	-0.22	-1.81	-0.48	0.06	2.77
		CRYSTAL_AVG	-5.39	0.49	-6.51	-0.14	-0.07	2.86
		CRYSTAL_STDV	9.31	3.77	7.88	0.37	0.38	0.13
2	A:U W:W C	MP2/6-31G**	-4.42	3.12	-2.24	0.08	0.08	2.83
		HF/cc-pVDZ	0.03	2.03	0.03	0	0.13	2.99
		B3LYP/6-31G**	-0.98	0.03	0.35	0	0.09	2.86
		PW91/DZP	4.11	1.23	-1.73	-0.03	0.03	2.78
		AM1	13.22	-8.76	3.35	0.1	0.06	3.44
		AMBER	6.2	-1.65	-3.98	-0.11	0.21	2.93
		AMBER_WATER	71.64	-36.6	-30.18	-0.2	0.46	4.1
		PM3	-5.17	-1.03	-3.66	0.21	0.05	2.81
		CRYSTAL_AVG	-1.96	3.27	-6.56	-0.04	0.11	2.79
		CRYSTAL_STDV	8.3	4.5	8.27	0.36	0.29	0.12
3	G:U W:W C	MP2/6-31G**	-4.03	0.66	-0.09	0.13	-2.4	2.86
		HF/cc-pVDZ	-0.52	1.71	-1.94	-0.08	-2.54	2.94
		B3LYP/6-31G**	3.94	4.11	-1.3	-0.15	-2.42	2.84
		PW91/DZP	-0.87	0.29	-5.78	-0.04	-2.33	2.79
		AM1	-16.78	6.12	3.76	-0.47	2.23	2.99
		AMBER	0.85	0.18	-1.91	0.02	-2.54	2.86
		AMBER_WATER	18.72	0.35	7.6	-0.11	-2.55	2.86
		PM3	-0.03	3.26	-3.57	-0.21	-2.52	2.74
		CRYSTAL_AVG	-1.51	-0.61	-6.55	-0.17	-2.22	2.81
		CRYSTAL_STDV	8.48	5.8	7.32	0.41	0.43	0.15
4	A:G H:S T	MP2/6-31G**	13.55	-24.57	-47.02	0.19	1.79	3.13
		HF/cc-pVDZ	4.49	-21.1	-46.7	0.15	1.9	3.39
		B3LYP/6-31G**	6.42	-20.3	-40.9	0.05	1.95	3.24
		PW91/DZP	-13	-18.15	-24.78	0.23	2.09	3.21
		AM1	-48.92	11.42	17.95	0.91	2.7	3.23
		AMBER	-17.57	-25.05	-49.97	0.11	1.9	3.18
		AMBER_WATER	-38.3	-11.5	-13.68	-0.48	2.09	3.43
		PM3	32.89	-16.04	-29.79	0.56	1.98	3.24
		CRYSTAL_AVG	-2.12	-13.1	1.07	-0.14	2.24	3.31
		CRYSTAL_STDV	14.1	5.7	13.1	0.5	0.36	0.17
5	A:U H:W T	MP2/6-31G**	6.24	-5.41	-1.93	-0.09	0.11	2.79
		HF/cc-pVDZ	4.21	-3.28	-1.63	-0.06	0.07	2.95
		B3LYP/6-31G**	-3.2	-7.49	1.52	0	0.15	2.79
		PW91/DZP	4.9	-2.93	-0.91	-0.03	0.11	2.77
		AM1	16.81	11.36	-2.99	0.02	0.14	3.46
		AMBER	6.97	1.46	-0.46	0	0.02	2.93
		AMBER_WATER	52.24	-4.58	28.68	-1.35	0.18	2.66
		PM3	-6.11	5.63	8.66	0.22	0.16	2.79

		CRYSTAL_AVG	-1.41	-0.39	-1.17	-0.04	0.17	2.82
		CRYSTAL_STDV	13.6	6.63	9.75	0.46	0.49	0.19
6	A:G W:W C	MP2/6-31G**	-0.24	-1.6	-21.9	-0.13	-0.1	2.89
		HF/cc-pVDZ	-3.85	-2.85	-21	-0.2	0.01	3.06
		B3LYP/6-31G**	6.14	-2.93	-12.8	-0.21	-0.01	2.94
		PW91/DZP	3.72	-1.84	-19.01	-0.31	-0.08	2.82
		AM1	11.18	-8.26	1.19	0.27	-0.36	3.39
		AMBER	4.56	-2.6	-6.64	-0.15	-0.08	2.95
		AMBER_WATER	-18.3	-7.92	20.25	0.07	0.31	3.02
		PM3	-12.7	0.06	-5.09	-0.03	0.03	2.8
		CRYSTAL_AVG	-10.8	1.27	-8.55	-0.41	0.07	2.8
		CRYSTAL_STDV	14.2	5.84	12.05	0.41	0.5	0.19
7	A:A H:H T	MP2/6-31G**	15.21	0	-45.8	0	2.48	2.75
		HF/cc-pVDZ	11.2	-0.01	-37.5	0	2.6	3.01
		B3LYP/6-31G**	-0.06	-0.06	-32.1	0.01	2.67	2.88
		PW91/DZP	2.3	-0.01	-10.44	-0.02	2.72	2.9
		AM1	16.84	-0.26	-7.71	-0.07	3.57	3.28
		AMBER	0.62	0.23	-10.65	-0.01	2.65	2.94
		AMBER_WATER	37.77	3.16	3.16	0.36	2.83	2.84
		PM3	-0.03	-0.01	0.05	-0.53	3.03	2.69
		CRYSTAL_AVG	-9.63	-4.66	5.81	-0.35	2.51	2.86
		CRYSTAL_STDV	15.8	3.28	13.25	0.29	0.35	0.17
8	G:A S:W T	MP2/6-31G**	10.55	-17.39	-24.4	-0.03	1.74	3.28
		HF/cc-pVDZ	11.35	-16	-19.5	-0.05	1.76	3.48
		B3LYP/6-31G**	6.74	-18.4	-17.4	-0.07	1.82	3.32
		PW91/DZP	4.05	-39.08	-34.67	0.01	2.24	2.81
		AM1	59.36	8.72	-2.08	-0.86	2.49	3.35
		AMBER	-14.84	-19.73	-25.88	0.1	1.79	3.28
		AMBER_WATER	8.9	-16.1	22.67	0.58	1.72	3.26
		PM3	21.81	-17.95	-4.62	0	1.86	3.15
		CRYSTAL_AVG	13.04	-13.6	0.27	-0.16	1.85	3.32
		CRYSTAL_STDV	19	7.02	16.55	0.58	0.37	0.2
9	A:A H:W T	MP2/6-31G**	-10.37	1.87	28.94	0.01	2.45	2.89
		HF/cc-pVDZ	-7.74	1.11	21.05	0.01	2.51	3.11
		B3LYP/6-31G**	-1.4	0.95	17.58	0	2.5	2.94
		PW91/DZP	-0.13	2.91	22.63	0.14	2.51	2.83
		AM1	-10.41	5.65	31.66	-0.61	8.5	-2.3
		AMBER	10.64	1.05	9.17	-0.07	2.55	2.94
		AMBER_WATER	60.6	-7.48	59.44	-0.72	2.38	2.63
		PM3	-2.3	0.32	0.68	0.44	2.73	2.72
		CRYSTAL_AVG	1.84	5.7	-2.88	0.04	2.31	2.92
		CRYSTAL_STDV	12.8	5.23	19.57	0.5	0.36	0.19
10	A:U H:W C	MP2/6-31G**	-4.51	2.42	-1.76	0.04	-0.1	2.8
		HF/cc-pVDZ	0.12	1.24	0.06	0	-0.06	2.96
		B3LYP/6-31G**	-0.24	2.23	-2.63	-0.01	-0.13	2.79
		PW91/DZP	0.49	-0.17	-3.62	0	-0.11	2.79
		AM1	12.59	-11.03	2.72	-0.02	-0.22	3.47
		AMBER	2.47	-4.48	-3.77	-0.2	0.06	2.92
		AMBER_WATER	117.8	-39.5	-7.38	0.97	0.69	3.23
		PM3	-6.61	-2.96	-4.11	0.21	-0.24	2.78
		CRYSTAL_AVG	-3.59	2.93	-4.53	0.15	-0.22	2.8
		CRYSTAL_STDV	18.1	6.78	12.38	0.6	0.43	0.16
11	U:U W:W C	MP2/6-31G**	-1.5	-1.17	2.86	0	2.54	2.89
		HF/cc-pVDZ	-0.01	-0.22	0	0	2.74	2.95
		B3LYP/6-31G**	2.27	-0.66	-6.57	0.03	2.55	2.85

		PW91/DZP	0.07	-0.57	-2.82	-0.02	2.48	2.82
		AM1	0.95	1.47	-1.05	0.00	2.97	2.97
		AMBER	-2.55	-1.75	-0.95	0	-2.8	2.83
		AMBER_WATER	2.77	-0.16	9.64	0.14	-2.89	2.82
		PM3	-2.12	2.73	-4.3	0.09	2.6	2.74
		CRYSTAL_AVG	-10.1	-2.33	-12.6	-0.15	-2.36	2.86
		CRYSTAL_STDV	8.56	5.29	8.02	0.4	0.37	0.16
12	A:C H:W T	MP2/6-31G**	2.39	-3.64	1.28	-0.23	2.33	2.92
		HF/cc-pVDZ	0.01	-3.59	0	0	2.37	3.09
		B3LYP/6-31G**	-0.35	-2.73	0.09	0	2.41	2.93
		PW91/DZP	15.61	-2.05	0.19	0	2.41	2.86
		AM1	62.03	-30.34	13.82	0.73	2.35	3.67
		AMBER	1.95	-3.35	3.53	0.02	2.25	2.92
		AMBER_WATER	2.47	-7.19	-10.61	0.39	2.26	2.9
		PM3	17.29	-0.69	-14.6	-0.03	2.61	2.74
		CRYSTAL_AVG	-1.26	-4.3	-5.68	0.18	2.37	2.91
		CRYSTAL_STDV	20.5	6.37	15.61	0.48	0.36	0.18
13	G:C W:W T	MP2/6-31G**	-164.44	5.65	-63.64	0.13	-0.97	2.87
		HF/cc-pVDZ	-30.58	68.21	-43.03	0.15	-3.56	3.19
		B3LYP/6-31G**	-17.61	13.42	-37.98	0.06	-2.54	3.02
		PW91/DZP	-16.06	10.64	-11.47	0.2	-2.65	3.02
		AM1	-14.33	82.47	-25.94	-0.64	-4.05	2.74
		AMBER	-1.85	-16.16	-3.18	0.01	-3.24	2.90
		AMBER_WATER	-38.25	17.63	59.91	1.09	-1.35	2.64
		PM3	16.6	1.37	16.07	-0.31	-2.59	2.79
		CRYSTAL_AVG	-4.35	7.34	-6.67	-0.09	-2.27	2.88
		CRYSTAL_STDV	15.35	7.3	10.69	0.42	0.48	0.16
14	A:A W:W T	MP2/6-31G**	0.1	-0.02	0.22	0.33	2.33	2.98
		HF/cc-pVDZ	-0.84	0.03	1.24	0	2.34	3.16
		B3LYP/6-31G**	4.4	0	-12.1	0	2.27	2.98
		PW91/DZP	-1.18	-0.18	0.07	-0.08	2.33	2.88
		AM1	-0.23	0	0.35	-0.01	2.86	3.53
		AMBER	12.35	0.17	8	0.02	2.33	2.96
		AMBER_WATER	28.87	23.3	76.09	1.01	1.93	2.68
		PM3	-0.75	-0.04	-0.02	-0.43	2.36	2.78
		CRYSTAL_AVG	7.87	-7.58	-2.04	-0.53	2.19	2.87
		CRYSTAL_STDV	11.3	7.53	30.17	0.49	0.47	0.23
15	A:U W:W T	MP2/6-31G**	-5.73	-5.62	2.77	0.1	-0.1	2.83
		HF/cc-pVDZ	0.28	-3.58	-0.3	-0.01	-0.13	2.99
		B3LYP/6-31G**	-0.67	-7.96	-0.57	0.03	-0.06	2.81
		PW91/DZP	-4.69	-3.74	2.62	0.04	-0.05	2.78
		AM1	-0.75	-10.25	1.68	-0.01	-0.14	3.47
		AMBER	-4.74	-1.02	-1.37	0.03	-0.16	2.93
		AMBER_WATER	-23.8	-5.81	6.23	-0.07	-0.29	2.9
		PM3	-6.16	3.85	7.75	0.22	-0.14	2.8
		CRYSTAL_AVG	0.05	-1.9	0.13	0.03	-0.24	2.83
		CRYSTAL_STDV	12.7	7.42	13.34	0.55	0.41	0.19
16	G:G H:W C	MP2/6-31G**	-51.38	-86.65	-18.78	-0.75	-4.07	3.81
			-49.34[#]	0.35[#]	25.86[#]	0.06[#]	1.88[#]	2.72[#]
		HF/cc-pVDZ	-14.52	-94.54	-8.67	-0.32	-4.25	4.18
			-16.19[#]	0.04[#]	4.87[#]	0.01[#]	2.17[#]	2.87[#]
		B3LYP/6-31G**	-49.98	-15.60	-41.41	0.08	-2.44	2.98
		PW91/DZP	13.27	-4.84	1.75	0.16	-2.99	3.06
		AM1	64.95	6.80	35.59	0.21	-3.73	2.64
		AMBER	7.54	-6.09	-0.59	-0.14	-3.81	2.90

		AMBER_WATER	41.76	1.43	7.79	0.85	3.55	2.73
		PM3	22.52	0.11	32.7	0.08	-2.79	2.71
		CRYSTAL_AVG	6.43	-3.86	-0.31	0.08	-2.94	2.88
		CRYSTAL_STDV	12.62	6.2	12.13	0.61	0.41	0.18
17	G:G H:W T	MP2/6-31G**	-10.94	-11.38	-5.16	-0.11	0.15	2.83
		HF/cc-pVDZ	7.49	-10.4	4.94	0.08	0.14	2.96
		B3LYP/6-31G**	10.26	-9.91	3.48	0.03	0.17	2.85
		PW91/DZP	20.52	-7.64	2.86	-0.01	0.2	2.83
		AM1	29.14	-10.73	-2.39	-0.78	-0.38	3.22
		AMBER	28.84	-3.69	0.8	-0.54	0.08	2.87
		AMBER_WATER	-0.77	-32.2	-49.07	1.78	-2.21	2.52
		PM3	-12.89	-1.76	-12.01	0.05	0.24	2.81
		CRYSTAL_AVG	9.29	1.77	-0.64	-0.13	-0.02	2.89
		CRYSTAL_STDV	12.9	6.59	8.36	0.28	0.65	0.17
18	G:G S:S T	MP2/6-31G**	-11.51	-0.1	41.89	-0.02	1.05	3.45
		HF/cc-pVDZ	-7.71	0.09	40.71	0	1.09	3.67
		B3LYP/6-31G**	-13	0.08	30.85	0	1.17	3.53
		PW91/DZP	0.96	-4.17	36.56	-0.43	1.19	3.43
		AM1	-67.72	-18.84	-15.13	-0.66	1.92	3.63
		AMBER	24.24	-0.8	53.1	-0.05	1.13	3.47
		AMBER_WATER	-123	18.67	2.05	2.25	2.55	2.36
		PM3	-44.78	5.06	8.71	-0.63	1.28	1.28
		CRYSTAL_AVG	4.64	-4.54	13.3	-0.66	1.27	3.46
		CRYSTAL_STDV	22.1	3.6	18.48	0.46	0.29	0.24
19	U:U W:W T	MP2/6-31G**	0.31	-0.01	0.38	0	-2.6	2.93
		HF/cc-pVDZ	-0.31	0.14	-0.66	-0.01	-2.75	2.97
		B3LYP/6-31G**	-0.34	0.02	-0.89	0	-2.58	2.89
		PW91/DZP	-11.5	0.1	-6.55	0.02	-2.49	2.86
		AM1	-0.29	-0.02	-0.07	0.00	-3.05	2.94
		AMBER	-11.02	-0.13	-8.5	0.02	-2.73	2.88
		AMBER_WATER	16.27	1.81	20.85	0.26	-2.51	2.93
		PM3	0.79	0.01	1.26	0.28	-2.69	2.67
		CRYSTAL_AVG	-6.22	-4.52	-4.36	0.11	-2.35	2.79
		CRYSTAL_STDV	16	3.25	10.24	0.45	0.45	0.16
20	A:G s:s T	MP2/6-31G**	97.13	47.17	-61.45	-2.56	-0.1	1.64
		HF/cc-pVDZ	-9.47	19.51	-14.3	-0.55	1.6	3.46
		B3LYP/6-31G**	1.52	23.45	-33.6	-0.9	1.38	3.2
		PW91/DZP	-10.6	24.27	-20.69	-0.56	1.55	3.15
		AM1	-22.03	23.91	-2.91	-1.01	2.30	3.42
		AMBER	-102.7	49.3	80.69	2.75	-0.4	1.58
		AMBER_WATER	-7.42	-7.42	34.68	-0.13	1.2	3.42
		PM3	-9.31	30.88	-2.52	-0.23	1.72	2.93
		CRYSTAL_AVG	-15.8	27.99	-7.44	0.22	1.71	3.02
		CRYSTAL_STDV	26.9	7.23	13.6	0.88	0.29	0.21
21	G:U s:h C	MP2/6-31G**	3.12	7.74	27.12	0.48	0.63	3.27
		HF/cc-pVDZ	4.18	0.19	26.19	0.46	0.74	3.65
		B3LYP/6-31G**	3.61	4.6	25.57	0.47	0.63	3.39
		PW91/DZP	-5.65	6.24	-1.47	0.14	0.61	3.30
		AM1	-151.35	12.53	83.67	5.50	-0.67	-0.33
		AMBER	-1.75	-0.73	0.09	-0.06	0.99	3.54
		AMBER_WATER	-19.4	-3.27	9.42	-0.76	0.96	3.47
		PM3	-13.95	-0.67	-6.7	0.19	0.71	3.48
		CRYSTAL_AVG	-11.8	-1.22	0.15	-0.1	0.99	3.44
		CRYSTAL_STDV	18.2	5.48	7.45	0.56	0.37	0.14
22	A:A h:s T	MP2/6-31G**	-13.38	0.45	12.11	0.18	2.25	2.91

		HF/cc-pVDZ	-9.97	-2.23	-2.79	0.28	2.39	3.09
		B3LYP/6-31G**	4.32	-1.05	18.61	-0.44	2.29	2.92
		PW91/DZP	-6.49	-0.95	2.42	0.01	2.36	2.85
		AM1	-19.09	-9.09	7.28	0.39	3.08	3.17
		AMBER	-99.06	-36.54	84.24	-2.59	0.78	1.33
		AMBER_WATER	43.61	33.02	61.18	-0.15	1.7	3.62
		PM3	-21.65	7.69	6.33	0.24	2.53	2.91
		CRYSTAL_AVG	-5.65	-1.39	1.66	-0.13	2.38	2.75
		CRYSTAL_STDV	13.5	6.01	9.41	0.52	0.3	0.2
23	A:G w:s C	MP2/6-31G**	15.39	40.75	10.49	0.4	2.1	2.97
		HF/cc-pVDZ	13.41	37.95	7.47	0.49	2.19	3.18
		B3LYP/6-31G**	-4.35	34.47	-16.4	0.04	2.23	3.07
		PW91/DZP	12.78	45.91	-42.34	-0.73	1.84	2.81
		AM1	-63.96	18.13	-2.31	1.04	-3.02	2.85
		AMBER	89.19	46.9	-50.66	-2.23	0.59	2.02
		AMBER_WATER	31.8	40.84	-22.56	-1.46	1.94	2.79
		PM3	23.58	58.14	-16.57	0.78	2.32	2.8
		CRYSTAL_AVG	-19.3	45.49	-6.41	-0.1	2.07	2.84
		CRYSTAL_STDV	28.9	12.6	24.19	0.92	0.43	0.25
24	A:U s:w C	MP2/6-31G**	-14.57	-3.24	17.21	0.24	0.18	2.94
		HF/cc-pVDZ	-9.3	-3.84	11.43	0.17	0.21	3.12
		B3LYP/6-31G**	2.34	-6.63	2.83	0.03	0.16	2.98
		PW91/DZP	-19.9	-4.37	9.29	0.22	0.14	2.9
		AM1	6.60	-5.34	1.90	0.03	-0.33	3.05
		AMBER	-25	-1.85	-1.3	0.62	0.21	2.94
		AMBER_WATER	-11.6	-5.87	-65.02	-0.6	0.19	3.04
		PM3	-8.6	1.69	23.38	-0.09	0.23	2.83
		CRYSTAL_AVG	-21.7	-3.7	11.6	0.14	0.15	2.95
		CRYSTAL_STDV	10.6	4.23	12.27	0.32	0.26	0.17
25	U:U w:h T	MP2/6-31G**	0.72	7.7	-1.09	0.01	2.54	3.07
		HF/cc-pVDZ	0.07	8.13	-0.07	0	2.74	3.16
		B3LYP/6-31G**	-2.05	9.07	2.52	0.01	2.7	3.04
		PW91/DZP	3.9	8.77	1.01	0	2.54	3.01
		AM1	0.56	3.99	1.38	0.01	2.67	3.18
		AMBER	6.39	15.07	1.91	-0.14	3.03	3.09
		AMBER_WATER	-7.91	4.58	-15.64	1.37	2.9	2.74
		PM3	-7.46	16.05	-10.77	0.09	2.63	3.15
		CRYSTAL_AVG	15.73	13.61	-9	0.18	2.59	2.91
		CRYSTAL_STDV	18.7	6.89	15.07	0.67	0.3	0.22
26	A:C w:w C	MP2/6-31G**	21.14	18.68	-6.29	-0.22	-2.4	2.66
		HF/cc-pVDZ	18.18	18.57	-4.92	-0.09	-2.49	2.81
		B3LYP/6-31G**	4.42	13.55	-4.84	0.01	-2.6	2.7
		PW91/DZP	17.17	15.25	-8.41	-0.06	-2.44	2.63
		AM1	71.17	1.69	6.98	-1.01	-3.38	2.54
		AMBER	12.31	10.95	-5.11	0.18	-2.51	2.75
		AMBER_WATER	-16.1	-1.02	-31.03	-1.81	-2.45	2.73
		PM3	13.99	5.78	-2.44	-0.35	-2.51	2.71
		CRYSTAL_AVG	10.02	14.26	-8.08	0.05	-2.45	2.44
		CRYSTAL_STDV	12.4	10.4	6.61	0.94	0.45	0.38
27	A:A w:w C	MP2/6-31G**	0.17	17.78	-11.53	0.15	2.61	2.64
		HF/cc-pVDZ	2.21	17.69	-8.78	0.14	2.67	2.82
		B3LYP/6-31G**	8.96	14.47	-0.22	0.09	2.5	2.75
		PW91/DZP	6.63	15.95	-8.07	0.07	2.56	2.58
		AM1	6.61	23.24	-12.85	0.42	3.20	2.89
		AMBER	-7.39	15.03	-24.01	-0.72	-2.53	2.66

		AMBER_WATER	30.68	6.1	-18.36	0.03	-2.86	2.9
		PM3	-15.26	4.06	-1.01	-0.38	2.58	2.73
		CRYSTAL_AVG	-16.8	21.66	-15.4	-0.77	-2.14	2.48
		CRYSTAL_STDV	15.7	13.2	12.76	0.21	0.91	0.19
28	A:A W:S C	MP2/6-31G**	36.53	-10.99	-3.97	-0.33	-2.60	3.21
		HF/cc-pVDZ	-13.34	-26.85	-46.15	0.00	-2.40	3.53
		B3LYP/6-31G**	-18.29	-26.14	-46.70	-0.21	-2.49	3.40
		PW91/DZP	28.33	-19.49	-27.51	-0.24	-2.75	3.25
		AM1	-85.49	-35.53	-44.98	-0.95	-2.98	3.50
		AMBER	-29.78	-19.76	34.35	-0.87	-2.35	3.01
		AMBER_WATER	-34.75	-37.75	41.92	-1.96	-2.02	2.57
		PM3	19.74	-15.04	-5.63	-0.54	-2.88	3.01
		CRYSTAL_AVG	-4.49	-29.14	-12.72	-0.38	-2.51	3.52
		CRYSTAL_STDV	13.85	10.12	12.24	0.40	0.46	0.26
29	A:U w:s T	MP2/6-31G**	74.76	-56.77	36.37	-2.97	2.9	2.14
		HF/cc-pVDZ	-22.88	-49.46	-22.95	-0.71	0.33	4.3
		B3LYP/6-31G**	31.63	-34.6	2.77	-1.2	0.79	3.66
		PW91/DZP	31.47	-36.06	-16.15	-1.1	0.26	3.66
		AM1	13.26	-34.82	-13.83	-0.90	0.44	3.63
		AMBER	-11.21	-149.7	-105.9	3.57	0.37	-0.88
		AMBER_WATER	21.71	-31.7	-28.68	-0.1	0.39	3.93
		PM3	-65.03	-37.98	-28.52	-1.62	-0.58	3.75
		CRYSTAL_AVG	-1.57	-58.7	-18.6	-0.33	-0.4	4.06
		CRYSTAL_STDV	21	12.1	14.55	0.88	0.51	0.31
30	A:C W:S C	MP2/6-31G**	46.82	-32.19	-16.43	-1.1	-0.9	4.15
		HF/cc-pVDZ	26.99	-34.5	-19.8	-0.93	-0.9	4.35
		B3LYP/6-31G**	42.14	-35.2	-13.1	-1.23	-0.76	4.21
		PW91/DZP	-19.5	-24.35	-25.65	-0.99	-0.36	3.83
		AM1	47.39	-43.82	-8.24	-1.67	-1.52	4.20
		AMBER	5.24	-28.33	-10.61	-0.69	-0.95	3.97
		AMBER_WATER	-54.7	-14.3	-43.43	-1.33	-0.2	3.42
		PM3	37.64	-42.97	-7.5	-1.76	-0.73	4.16
		CRYSTAL_AVG	-4.44	-34.7	-21.3	-0.47	0.3	4.23
		CRYSTAL_STDV	22.8	10.3	14.47	0.62	0.64	0.37
31	A:G S:S C	MP2/6-31G**	-9.51	-3.2	-48.48	-0.65	-1.8	3.74
		HF/cc-pVDZ	-3.21	-9.3	-59.4	-0.02	-1.83	3.84
		B3LYP/6-31G**	-15.4	-8.22	-53.8	-0.2	-1.87	3.73
		PW91/DZP	20.34	-1.13	-59.39	-0.62	-2.16	3.58
		AM1	53.08	-2.15	-14.86	-1.06	-3.03	3.37
		AMBER	107.02	20.68	-63.88	-0.87	-2.05	3.27
		AMBER_WATER	-48	-6.39	0.97	0.56	-2	3.67
		PM3	34.06	0.63	-40.04	-0.28	-2.3	3.52
		CRYSTAL_AVG	-1.6	-15.8	-8.36	-0.13	-2.36	3.82
		CRYSTAL_STDV	23.2	20.7	30.3	0.95	0.39	0.46
32	A:U W:S C	MP2/6-31G**	-83.93	-35.11	-18.25	-0.77	-0.6	4.57
		HF/cc-pVDZ	-42.61	-29.2	-23.8	-1.01	0.8	4.3
		B3LYP/6-31G**	-45.45	-30.7	-20.7	-1.29	0.56	4.2
		PW91/DZP	6.59	-24.65	-9.28	-0.74	0.54	3.98
		AM1	71.92	-28.53	-25.99	-1.78	-0.52	4.25
		AMBER	1.76	-13.47	20.28	0.68	0.84	3.42
		AMBER_WATER	-80.5	-1.48	-45.97	-2.99	1.18	2.59
		PM3	-45.88	-32.03	-15.68	-2.2	0.4	3.78
		CRYSTAL_AVG	-17.6	-31.5	-1.47	-0.59	0.44	4.26
		CRYSTAL_STDV	37.2	8.33	13.42	0.5	0.45	0.28

[#] Represents the basepair parameters when it is in W:W T conformations.

Similar orientation of dangling ketone oxygen in G:C W:W T (system 13) basepair forces unusual optimized structure by some of the methods. The MP2/6-31G** method, for example, gives a stacked geometry of this basepair. The G:G H:W C basepair is an integral part of G-quartet in telomeric DNA and some untranslated regions of mRNA and is quite stable with K^+ ion in the center of the quartet. More details on the quartet structures are given in chapter 1. We have therefore optimized the G-quartet structure (PDB ID:156D) using the popular *ab initio* quantum chemical methods. Basepairs of this quartet have comparatively large propeller values in cases of MP2/6-31G** and B3LYP/6-31G** (Figure 3.2.5) optimized structures, while larger shear values for the HF/cc-pVDZ optimized structure. This indicates that these basepair has a tendency to acquire non-planar geometry and may need an ion, some topological strain due to loop and bead formation or ligand for stabilization. Optimizations of the quartet along with a K^+ ion lead to more planar base quartet orientations (Table 3.2.9). Each guanine residue utilizes both N1 and N2 atoms of one face and O6 and N7 atoms of the other face to form hydrogen bonds with the other residues of the G-quartet. Therefore, the residues do not have much flexibility to alter orientation as compared to the isolated baepairs. However, the instability in the quadruplex comes from the arrangement of the guanine O6 carbonyl groups central to the G-quadruplex. This negatively charged cavities located between the G-quartet needs to be stabilized by the co-ordination of the cations. When the K^+ ions are located centrally, this stabilizes the system by favorable electrostatic interaction, neutralizes partial negative charge of the O6 atoms. This indicates that the strong non-hydrogen bonded functional groups can be involved in molecular recognition and in the absence of the ligand the non-canonical basepairs can

adopt a different structure, thereby acting as a switch. This indicates that the non-hydrogen bonded polar functional groups can be involved in molecular recognition and in the absence of the specific ligand the non-canonical basepairs can adopt a different structure, thereby acting as a conformational switch. The non-polar and sugar mediated basepairs also presumably play similar role in different biochemical processes, such as in A-minor motif (Nissen et al., 2001).

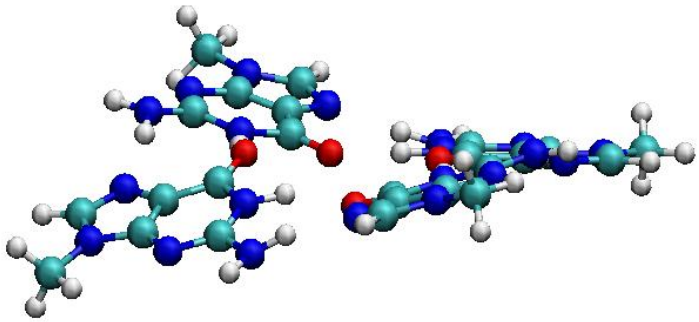


Figure 3.2.5. Represents structure of G-quadruplet optimized by B3LYP/6-31G** method, without ion

Amount of pyramidalisation is found to be directly proportional to the value of propeller twist. The polar basepairs, e.g. A:G H:S T (system 4)**,** A:G W:W C (system 6), A:A H:H T (system 7)**,** G:C W:W T (system 13), etc. have highly pyramidal amino groups associated with large propeller, while the A:U W:W C, A:U H:W T (system 5), A:U H:W C (system 10) basepair systems have planar amino group and small propeller. This out of plane motion of the amino groups can give rise to tertiary interaction with the neighboring basepairs, by forming hydrogen bond between the amino group of the

guanine with the oxygen of the adjacent basepair, associated with comparably high stagger and propeller values (Sponer et al., 2003, Mokdad et al., 2006, Bandyopadhyay and Bhattacharyya, 2006)

Table. 3.2.9. The basepair parameters of the G- quadruplet structures optimized by different quantum chemical methods in absence of ion and in presence of K^+ ion (in parenthesis)

Method	Base pair	Buckle (^{O})	Open (^{O})	Propeller (^{O})	Stagger (Å)	Shear (Å)	Stretch (Å)
MP2/6-31G**	G1-G2	23.73 (-9.03)	-9.45 (-9.35)	-26.08 (6.36)	-0.06 (-0.27)	2.70 (-1.73)	2.93 (2.54)
	G3-G4	20.60 (13.22)	-10.27 (2.58)	-29.45 (-38.97)	-0.04 (-0.75)	2.69 (-1.74)	2.92 (2.54)
	G1-G4	0.79 (-13.7)	-7.28 (-6.93)	23.68 (12.13)	0.21 (-1.00)	-2.77 (2.65)	2.93 (2.79)
	G2-G3	-4.34 (-31.59)	-7.92 (-2.92)	26.83 (-8.39)	0.15 (0.08)	2.78 (-2.17)	2.91 (3.1)
HF/cc-pVDZ	G1-G2	7.30	-3.41	-15.19	-0.33	3.48	2.97
	G3-G4	-3.79	-6.40	5.48	0.02	3.26	3.06
	G1-G4	16.33	-6.79	1.40	0.01	-4.11	2.94
	G2-G3	4.08	-1.61	12.53	0.43	3.84	2.87
B3LYP/6-31G**	G1-G2	-0.41 (15.08)	-8.92 (-3.63)	-25.09 (-0.45)	0.12 (-0.09)	3.01 (-2.71)	2.84 (2.98)
	G3-G4	11.93 (16.53)	-6.09 (-3.48)	-16.05 (2.14)	-0.13 (-0.07)	2.90 (-2.7)	2.92 (2.98)
	G1-G4	10.45 (-12.88)	-7.26 (-3.59)	28.20 (5.55)	0.06 (-0.02)	-2.83 (2.69)	2.86 (2.98)
	G2-G3	-0.79 (15.48)	-5.62 (-3.44)	16.25 (4.09)	0.19 (-0.02)	2.95 (-2.69)	2.91 (2.98)
GGA:PW91/DZP	G1-G2	0.37 (3.09)	-5.55 (-4.53)	-19.05 (-3.68)	-0.12 (-0.16)	2.83 (-2.72)	2.86 (2.92)
	G3-G4	13.59 (11.99)	-4.86 (-4.1)	-4.67 (-8.56)	0.03 (-0.22)	2.90 (-2.74)	2.88 (2.91)
	G1-G4	3.64 (-11.17)	-5.58 (-4.99)	15.72 (10.14)	0.06 (-0.09)	-2.83 (2.72)	2.86 (2.90)
	G2-G3	10.82 (6.27)	-5.40 (-4.95)	2.42 (1.26)	0.04 (-0.11)	2.83 (-2.74)	2.88 (2.92)

G1, G2, G3 and G4 represent the individual components of the guanine quadruplet.

3.2.4. Discussion

In this section, I have analyzed frequently observed RNA basepairs from experimental data as well as computer simulation with different levels of rigor and accuracy. All the methods are known to be very reliable and accurate for predictions of structures, interactions, thermodynamics and other properties of bio-molecules and organic molecules. I observed that the results from these popular approximate methods are quite consistent with each other. These are indicative of important contribution of non-canonical basepairs in RNA structure, stability, flexibility and recognition. Considering ensemble of crystallographic structures of different RNA basepairs to be true representatives of experimental structures, especially in the absence of any better data, we have compared structures of different basepairs optimized by different methods with the crystallographic ensembles. I found that most of the *ab initio* quantum chemical methods optimize the structures in a similar manner. The structures of the non-polar or sugar mediated basepairs, optimized by the rigorous methods, however are sometimes quite different. Few polar basepairs, such as G:G H:W C and G:C W:W T also follow the above criteria. In these cases, free carboxyl groups of guanine of the initial geometry sometimes involved itself in hydrogen bonding interaction in altered geometry. I believe that these base pairings depend on environment such as ion, water or protein mediated interactions. Thus, in absence of the above ligands, the basepair remain in one form while in presence of the appropriate mediator as discussed above, the basepair adopts a different form. Thus, these may be termed as conformational switch. Variation in the structures could best be characterized in terms of six basepair parameters using basepairing edge specific axis system. The variations in basepair structures are sometimes

not reflected in their interaction energies. As for example, MP2/6-31G** and B3LYP/6-31G** energies of most of the polar basepairs are extremely similar but their structures are often significantly different.

Section 3.3

RNA Non-Canonical Basepair Database: A Complete Analysis of the Energetic, Structural Features and Dynamics of the Possible Basepairs

3.3.1. Introduction

Ribonucleic acids (RNA) play crucial role in gene expressions (Gesteland et al., 2006, Leontis and Westhof, 2003, Hermann and Westhof, 1999). They are involved in several biological processes (chapter 1). In addition to the regular helical structure, RNA undertakes several secondary and tertiary modifications such as hairpin loop, conjunction loop, pseudoknots etc (Hendrix et al., 2005). Watson-Crick (WC) basepairing are the most common type of interaction observed in RNA, but analysis of available structure of RNA solved by x-ray crystallography reveals that in addition to WC base pairing, RNA has ample varieties of base contacts, such as non-Watson-Crick base pairing (non-WC), triplets (Broitman et al., 1987), quadruplets (Lipps and Rhodes, 2009), basepairing in protonated form, water or ion mediated basepairs etc. which play important roles in maintaining the three dimensional structures of RNA. Non-WC basepairs also act as recognition elements in RNA-RNA and RNA-protein interaction (Hermann and Westhof, 1999). Interaction between double stranded helix and unpaired regions are always guided by non-WC basepairs. The Non-WC basepairs are also involved in specific interactions with proteins and ligands due to exposure of unusual atoms and groups in major and minor grooves for binding. It has been shown in the previous section (section 3.2) that some of the non-canonical basepairs, such as G:G H:W C and G:C W:W T basepairs have propensity to function as conformational switch as their optimized structures often depend on environment or the method used for optimization.

Non-Watson Crick basepairs are also the key members for formation of DNA or RNA triple helices, where two strands are held together by WC basepairing, while the third strand is attached to the purine strand through non-WC basepairing scheme. Owing

to the enhanced stability of the triplex (Frank-Kamenetskii and Mirkin, 1995), which can affect activities such as gene expression, the DNA triple helix furthered new hopes in therapeutic applications (Jain et al., 2008). Triple helix in RNA also has many biological applications; it forms important components of riboswitches (Mandal and Breaker, 2004), which are the novel type of genetic control elements, controlling gene expression.

Leontis and Westhof (Leontis and Westhof, 2001) derived 12 distinct base edge to edge interactions in RNA. These classifications demonstrate the basepairing edges, along with the associated glycosidic angles and local strand orientations. Among the observed RNA structures, 60% are held together by regular Watson-Crick type of basepairing, while rests are engaged in non-Watson-Crick through non-canonical way. There are some types of basepairs observed in RNA, where two electronegative atoms are quite close to each other, giving rise to possible electrostatic repulsion. These repulsions can become strong attractive interactions if one of the electronegative atoms, such as imino nitrogen or carboxyl oxygen, becomes protonated. These kinds of basepairs are known as protonated basepairs (Leontis et al., 2002, Chawla et al., 2011, Das et al., 2006). Analyzing the non-canonical basepairs in more detail may open up some new insights in understanding RNA folding and its three dimensional architecture.

I have already discussed about different softwares for the detection and classifications of the canonical and non-canonical basepairs in section 3.2. I have carried out all the calculations using BPFIND (Das et al., 2006), which has been developed by our groups. It detects a basepair when at least two hydrogen bonds are found between the residues, where the hydrogen bonding heavy atoms belong to the base moiety or sugar O2' atom. Apart from the regular edges of the bases, BPFIND also takes in account of the

protonated bases. Several basepairs between chemically modified bases can be detected by this algorithm; however they are replaced by their nearest counterpart.

Different databases have been developed on classification of canonical and non-canonical basepairs depending on the basepairing patterns, isostericity of the basepairs etc, however each of them use specific methods for defining the basepair patterns. NCIR (Nagaswamy et al., 2002) (A database of non-canonical interactions in known RNA structures), BPS (Xin and Olson, 2009) (A database of RNA base-pair structures) give us information about three dimensional structures of RNA. The JSCH-2005 and S22 database (Jurečka et al., 2006) contain energies and geometry of 137 molecular complexes obtained by MP2 and CCSD(T) CBS calculations, which includes DNA and RNA basepairs, amino acid complexes, and model basepairs, for which hydrogen bonding and dispersion interactions play important roles. NCIR database gives the geometry of the basepairs, base triplets, and base quadruplets as found in the crystal structures along with detail literature survey. However, it does not give any quantitative information about the stability and dynamics of the basepairs, such as interaction energy, base pair geometrical parameters etc. The BPS database provides quantitative information about the basepairs, triplets, and higher order base interactions along with isosteric basepairs. It gives statistical analysis of the basepair parameters, their isosteric parameters, and hydrogen bonding patterns. However BPS uses 3DNA package for basepair identification and contains many base-base contacts formed by single H-bond. A very recent study by Leontis and co-workers provides information about the possible base triplets. They have classified the triplets in 68 families, here the basepairs belonging to

the same triple family are found to be geometrically similar in nature, and they have their glycosidic bonds oriented in the same directions (Almakarem et al., 2012).

Here, we are introducing a new database known as RNA non-canonical basepairs database (http://www.saha.ac.in/biop/www/rnabasepair.html), which provides both qualitative and quantitative information about the possible basepairs of RNA. Structural and dynamical parameters of each of the basepair such as interaction energy, basepair orientational parameters, isosteric values along with types of triplets associated with each basepair can be obtained from the database. This present database is unique in the sense that all the quantitative parameters associated with each of the basepair have been derived from the optimized geometry. We also represent the statistical analysis of the basepair parameters taken from non-redundant set of RNA crystal structures in graphical representation.

3.3.2. Description about Database

All the PDB files were collected from the non-redundant data set from HD-RNAS database (Ray et al., 2012) as available in April 2010, which contains best representative structures from each type of functional classes obtained through best resolution, smallest R-factor and larger lengths of nucleotide chains (Ray et al., 2012). The dataset contains 109 PDB files solved by X-ray crystallography with 3.5 Å or better resolution and 30 nucleotides or larger length. By running BPFIND on all the PDB files, every possibility of different types of basepairing patterns in RNA was obtained. I find many basepairs in the database frequently, while some basepairs occur very rarely in the RNA crystal structures, few with zero frequency also, but PBFIND hypothesized them as

such basepairing types are possible. Since improvement of RNA synthesis, X-ray crystallography, and NMR spectroscopy resources lead to increase in high resolution crystal structures with each passing day, possibly in near future these rare basepairs will appear with high frequency and possibly with specific functional roles in biological processes.

Theoretically, a total of 126 basepairing types are possible by RNA bases having two or more number of hydrogen bonds between the bases and sugars, whereas DNA molecules show generally two. However, the number of occurrences of these non-Watson-Crick basepairs is limited in crystal structures database, thus also limiting our extent of study. Our group has developed a software NUCGEN v2.0 (Halder et al.), which uses the basic principle as employed in its previous version (Bansal et al., 1995, Bhattacharyya and Bansal , 1990). It is capable of generating non-canonical basepairs based on a user-given geometry in the form of a set of six basepair orientational parameters, such as buckle, open angle, propeller, stagger, shear, stretch. The NUCGEN v2.0 (Halder et al.), uses a relation between local parameters and helical parameters as derived earlier (Bhattacharyya and Bansal, 1990) to obtain the parameters in helical sense.

I have made use of NUCGEN v2.0 software to generate coordinates corresponding to the non-canonical basepairs, which are not found frequently. Our database includes all possible canonical, non-canonical, protonated and triplets stabilized by at least two hydrogen bonds, which includes the model hypothetical basepairs also. Our database includes a total of 126 types of basepairs, among them 46 are frequently observed for which proper representatives can be detected for further calculations. There are

64 are hypothetical which were proposed by BPFIND as two or more hydrogen bonds would have been possible in such orientations but these were not found in the crystal structures, and 16 protonated basepairs (Table 3.3.1). I have optimized all the possible basepairs using B3LYP/6-31G (2d, 2p) method using Gaussian g09 (Frisch et al., 2009) and studied their structures and stability features in more details. The structural parameters of the optimized basepairs have been calculated by NUPARM (Mukherjee et al., 2006, Bansal et al., 1995) using basepairing edge specific axis system. Such an axis system was found to be important for describing proper geometry of a non-canonical basepair (Berman et al., 1992, Lu and Olson, 2008, Halder and Bhattacharyya, 2010) as the values of propeller, open angle, buckle, shear etc. calculated using standard axis system defined for Watson-Crick basepairs becomes huge, and ineffective for these non-canonical basepairs.

The home page of the database (Figure 3.3.1) contains a symmetric matrix which gives us the possible types of pairing patterns, involving different edges (i) Watson-Crick edge (ii) Hoogsteen edge (iii) sugar edge of the two bases. Thus the matrix dimension is (12 ×12).

Table 3.3.1. The information about all possible 126 basepairs, their frequency in duplex and triplet, the numbers of types of base triplets, associated with each duplex are given. The hydrogen bond and angles, the BSSE corrected interaction energy of the DFT optimized structures are also reported.

Sl No	Basepair Frequency of occurrences in (duplex/triplets)	Types of triplets	H-bond atoms	H-bond length	Angle	Int.Energy (kcal/mol)
Polar basepair						
1	G:C W:W C (6056/518)	13	N4-H4...O6 N1-H1...N3 N2-H2...O2	1.73 1.89 1.91	178.8 177.3 176.2	-26.51
2	A:U W:W C (1822/110)	13	N6-H6...O4 N3-H3...N1	1.84 1.81	174.9 177.7	-14.43
3	G:U W:W C (847/39)	6	N1-H1...O2 N2-H3...O6	1.75 1.89	172.6 172.2	-14.68
4	A:G H: S T (531/78)	8	N6-H6...N3 N2-H2...N7	2.05 1.96	171.2 177.6	-9.50
5	A:U H:W T (417/65)	10	N6-H6...O2 N3-H3...N7	2.13 1.75	169.9 169.2	-13.31
6	A:A H:H T (110/9)	3	N6-H6...N7 N6-H6...N7	2.02 2.02	164.9 164.9	-9.76
7	A:G W:W C (151/0)		N1-H1...N1 N6-H6...O6	1.91 1.81	177.1 177.1	-15.00
8	G:A S:W T (88/9)	1	N6-H6...N3 N2-H2...N1	1.98 1.96	175.6 177.3	-10.61
9	A:A H: W T (84/32)	5	N6-H6...N7 N6-H6...N1	2 1.97	167.2 177.5	-11.28
10	A:U H:W C (79/14)	4	N6-H6...O4 N3-H3...N7	1.98 1.75	171.1 174.1	-14.21
11	U:U W:W C (84/0)		N3-H3...O4 N3-H3...O2	1.83 1.83	172.17 172.18	-9.64
12	A:C H:W T (78/9)	3	N4-H4...N7 N6-H6...N3	1.96 1.94	177.48 168.02	-12.71
13	G:C W:W T (64/5)	2	N2-H2…N3 N1-H1…O2	2.19 1.92	179.95 168.95	-11.83
14	A:A W:W T (61/39)	5	N6-H6...N1 N6-H6...N1	1.97 1.97	176.6 176.6	-12.10
15	A:U W:W T (62/8)	4	N3-H3...N1 N6-H6...O2	1.76 2.07	177.8 172.1	-12.71
16	G:G H:W C (51/3)	1	N1-H1…O6 N2-H2…N7	2.03 2.04	175.23 165.68	-11.33

17	G:G H:W T (42/1)	1	N2-H2...O6 N1-H1...N7	2.17 1.83	164.3 170.6	-17.58
18	G:G S:S T (32/7)	1	N2-H2...N3 N2-H2...N3	1.98 1.98	178.3 178.2	-8.12
19	U:U W:W T (20/0)		N3-H3...O2 N3-H3...O2	1.85 1.85	171.39 171.31	-9.00
20	G:C S:W T (16/0)		N2-H2…N3 N4-H4…N3	1.87 2.02	179.35 175.23	-12.25
21	G:U S:W C (11/0)		N2-H2…O4 N3-H3…N3	1.84 2.02	172.18 173.01	-10.27
22	C:U W:W C (22/0)		N4-H4…O4 N3-H3…N3	1.822 1.913	176.27 171.20	-10.95
23	G:G W:W T (7/3)	1	N1-H1...O6 N1-H1....O6	1.72 1.72	172.50 172.50	-21.77
24	U:U W:W T (1/0)		N3-H3...O4 N3-H3...O4	1.81 1.81	173.12 173.05	-10.2
25	G:U W:W T (4/0)		N1-H1...O4 N3-H3...O6	1.75 1.78	174.00 176.26	-14.36
26	A:C W:W T (4/0)		N6-H6...N3 N4-H4....N1	1.92 1.93	174.87 178.43	-12.87
27	C:C W:W T (2/0)		N4-H4...N3 N4-H4...N3	1.89 1.89	174.50 174.45	-16.78
28	C:U W:W T (2/0)		N4-H4...O2 N3-H3...N3	1.82 2.02	177.30 169.88	-8.27
29	G:U S:W T (4/0)		N2-H2..O2 N3-H3...N3	1.87 2.02	172.76 172.38	-9.82
30	A:G H:W C (9/3)	2	N6-H6...O6 N1-H1...N7	1.86 1.88	168.75 173.67	-12.98
31	G:G W:S C (0/0)		N6-H6...O6 N1-H1...N3	1.81 1.93	171.80 175.17	-11.98
Non_polar basepair						
32	A:G s:s T (289/38)	5	N2-H2...N3 C2-H2...N3	2.01 2.74	175.77 127.81	-5.71
33	G:U s:h C (57/3)	1	N2-H2...O4 C5-H5...N3	1.91 2.45	170.52 158.27	-6.01
34	A:A h:s T (51/2)	1	N6-H6...N3 C2-H2...N7	2.04 2.51	170.78 144.10	-4.01
35	A:G w:s C (48/6)	1	N2-H2...N1 C2-H2...N3	2.02 2.55	172.87 147.3	-6.02
36	A:U s:w C (19/2)	2	N3-H3...N3 C2-H2...O4	1.96 2.26	170.35 146.86	-9.43
37	U:U h:w T (17/0)		N3-H3...O4 C5-H5...O4	1.84 2.23	178.05 156.65	-9.33
38	A:C w:w C (14/0)		N4-H4...N1 C2-H2...N3	1.97 2.58	178.13 143.8	-6.28
39	A:A w:w C (12/3)	1	N6-H6...N1 C2-H2...N1	2.04 2.59	174.3 142.75	-6.07
40	G:C h:h T (12/2)	1	N4-H4…N7 C5-H5…O6	2.12 2.35	170.66 178.25	-8.22

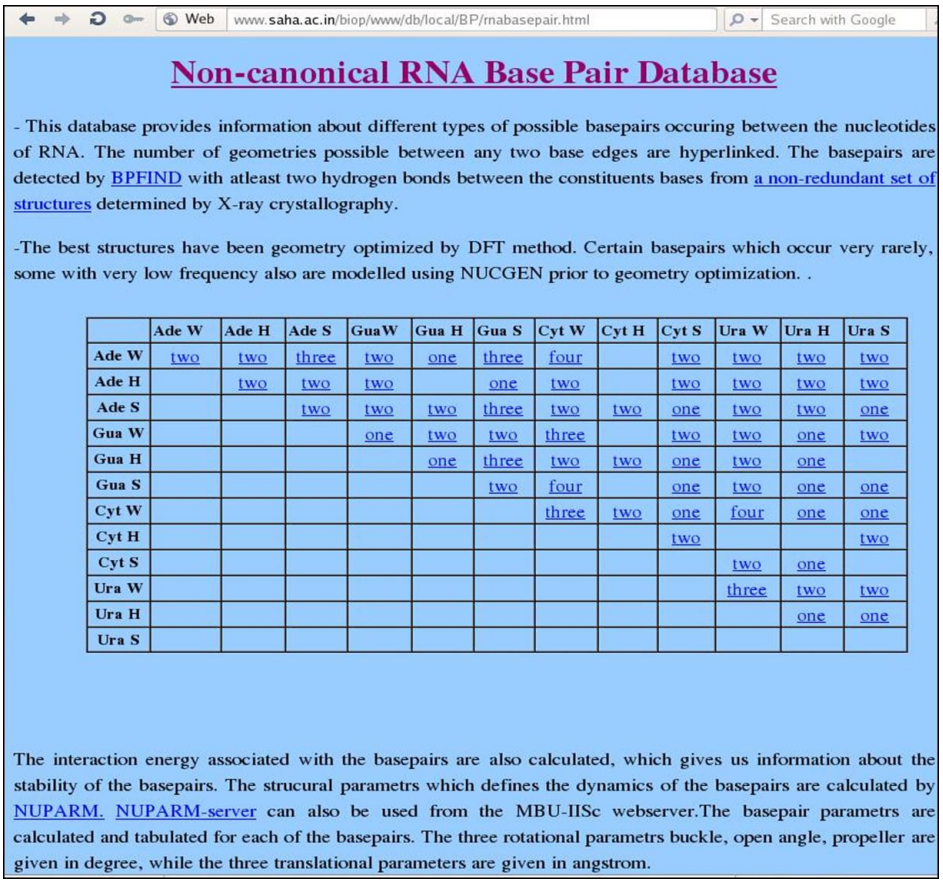
41	A:G h:w T (0/0)		N1-H1...N7 C8-H8...O6	1.83 2.39	172.92 131.37	-9.84
42	A:A s:w T (7/0)		N6-H6...N3 C2-H2...N1	2.02 2.51	174.02 145.27	-4.38
43	A:C s:w T (0)		N4-H4...N3 C2-H2...N3	2.01 2.48	174.00 141.33	-4.88
44	A:G w:w T (0)		N1-H1...N1 C2-H2...O6	1.87 2.54	173.91 137.40	-7.94
45	G:U w:h T (8/0)		N1-H1...O4 C5-H5...O6	1.88 2.19	162.05 167.24	-11.48
46	G:U h:w T (2/0)		N3-H3...N7 C8-H8...O4	1.95 2.22	164.74 134.65	-6.81
47	C:C w:h T (8/1)	1	N4-H4...O2 C5-H5...N3	1.96 2.61	175.15 154.23	-9.01
48	A:U s:w T (9/2)	1	N3-H3...N3 C2-H2...O2	1.96 2.36	172.52 145.43	-7.28
49	A:A h:w C (0/0)		C8-H8...N1 N6-H6...N7	2.37 2.02	142.21 171.34	-6.22
50	A:U w:h C (1/0)		N6-H6...O4 C5-H5...N1	1.93 2.35	173.29 165.00	-6.82
51	A:A h:h C (0)		N6-H6...N7 C8-H8...N7	2.03 2.35	173.67 145.45	-5.75
52	A:C h:w C (0)		N4-H4...N7 C8-H8...N3	2.04 2.25	169.13 138.56	-8.75
53	A:U h:h C (8/0)		N6-H6...O4 C5-H5...N7	1.96 2.33	158.32 170.20	-6.09
54	G:G h:s C (0)		N6-H6...N7 C8-H8...N3	1.92 2.66	174.10 127.80	-7.64
55	G:C h:h C (0)		N4-H4...O6 C5-H5...N7	2.00 2.51	163.93 175.51	-8.17
56	G:U h:w C (9/1)	1	N3-H3...N7 C8-H8...O2	1.97 2.25	164.46 132.61	-5.88
57	C:C w:h C (5/0)		N4-H4...N3 C5-H5...O2	2.09 3.72	148.84 139.36	-8.95
58	C:U w:h C (0)		N4-H4...O4 C5-H5...N3	1.91 2.23	176.39 161.12	-9.25
59	U:U w:h C (7/0)		N3-H3...O4 C5-H5...O2	1.89 2.25	179.95 157.96	-6.16
60	A:G s:w C (1/0)		N1-H1...N3 C2-H2...O6	1.96 2.45	174.54 138.97	-5.41
C-H…O mediated basepair						
61	A:U h:h T (8/0)	1	C5-H5...N7 C8-H8...O4	2.41 2.22	159.18 155.96	-3.94
62	A:A s:s T (6/0)		C2-H2...N3 C2-H2...N3	2.58 2.58	143.33 143.33	-0.72
63	A:U w:h T (0)		C5-H5...N1 C2-H2...O4	2.40 2.39	161.81 159.47	-1.81
64	G:G h:h T (0)		C8-H8...N7 C8-H8...N7	2.38 2.38	136.94 136.96	-5.6
65	G:U h:h T (0)		C5-H5...N7 C8-H8...O4	2.31 2.25	159.45 151.12	-5.31
66	U:U h:h T (0)		C5-H5...O4 C5-H5...O4	2.23 2.23	175.32 175.32	-4.5
67	A:U s:h C		C2-H2...O4	2.39	158.36	-1.34

	(2/0)		C5-H5...N3	2.41	161.44	
68	A:A s:w C (1/0)		C2-H2...N1 C2-H2...N3	2.53 2.64	147.84 144.25	-0.79
Sugar_mediated_polar basepair						
69	A:A W:S C (86/0)	0	N6-H6…N3 O2'-H2'…N1	2.02 1.76	159.66 171.88	-14.76
70	A:C W:S C (62/2)	2	N6-H6...O2 O2'-H2'...N1	1.82 1.76	173.2 165.8	-10.00
71	A:G S:S C (35/5)	2	N2-H2...N3 O2'-H2'...N3	1.99 1.81	160.98 170.92	-7.78
72	A:U W:S C (18/0)	0	N6-H6...O2 O2'-H2'...N1	1.92 1.76	161.3 168.8	-14.55
73	A:A H:S C (23/0)		N6-H6…N3 O2'-H2'…N7	2.17 1.86	152.50 172.75	-10.17
74	A:C S:W C (21/5)	1	N4-H4…N3 O2'-H2'…N3	2.10 1.78	153.24 176.26	-16.11
75	G:C S:S C (23/1)	1	N2-H2…O2 O2'-H2'…N3	1.86 1.82	155.54 171.38	-8.58
76	G:U S:S C (21/0)	0	N2-H2…O2 O2'-H2'…N3	1.96 1.80	150.91 173.22	-6.66
77	A:C H:S C (5/0)		N6-H6…O2 O2'-H2'…N7	1.84 1.81	169.50 159.84	-8.17
78	A:U H:S C (0)		N6-H6…O2 O2'-H2'…N7	2.03 1.76	152.55 164.48	-7.24
79	A:G S:W C (1/1)	1	N2-H2…O2' N1-H1…N3	1.989 2.217	163.92 161.85	-9.26
80	G:C W:S C (6/0)		N1-H1…O2 N2-H2…O2'	1.83 2.25	173.15 152.44	-1.45
81	G:U W:S C (8/0)		N2-H2…O2' N1-H1…O2 N3-H3…O6	2.42 1.734 1.874	160.09 168.93 161.29	-12.96
82	G:G S:S C (6/0)		N2-H2…N3 O2'-H2…N3	2.13 1.86	149.98 167.13	-7.29
83	G:C S:W C (4/0)		N4-H4…N3 O2'-H2'…N3	2.100 1.859	149.92 167.41	-10.91
84	C:U S:W C (5/0)		N3-H3…O2 O2'-H2'…O2	1.76 1.79	166.70 167.09	-7.51
85	U:U W:S C (0)		N3-H3…O4 N3-H3…O2 O2'-H2'…O2	2.154 1.704 1.961	148.25 171.59 165.88	-5.35
86	G:G W:S T (0)		N4-H4…N7 O2'-H2'…O6	2.257 1.746	158.74 157.45	-10.51
87	G:C W:S T (6/0)		N2-H2…O2 O2'-H2'…O2	1.865 1.723	156.04 160.68	-7.96
88	G:U W:S T (5/0)		N2-H2…O2 O2'-H2'…O6 O3'-H3'…O6	1.987 1.872 2.070	170.42 145.33 145.32	-15.48
89	G:G H:S T (0)		N2-H2…O6 O2'-H2'…O6 N2-H2…O3'	2.110 1.983 2.155	154.87 146.80 169.49	-8.73
90	C:U S:W T (2/0)		N3-H3…O2 O2'-H2'…O4	1.747 1.763	168.07 165.52	-8.62
91	U:U W:S T (0)		N3-H3…O2 O2'-H2'…O4	1.718 1.915	170.89 165.66	-6.14

Sugar mediated nonpolar basepair						
92	A:U w:s T (70/1)	1	C2-H2...O2 O2'-H2'-N1	2.54 1.74	153.86 166.21	-10.64
93	A:A s:s C (18/4)	2	C2-H2…N3 O2'-H2'…N3	2.637 1.763	143.11 170.08	-10.19
94	A:G s:s C (40/5)		O2'-H2'…N3 C2-H2…N3	1.79 2.71	174.47 112.33	-8.71
95	A:C s:s C (137/0)		C2-H2…O2 O2'-H2'…N3	2.192 1.795	153.24 172.51	-3.77
96	A:U s:s C (56/0)		O2'-H2…N3 C2-H2…O2	1.770 2.35	169.75 151.72	-10.00
97	A:G w:s T (85/0)		O2'-H2'…N1 N2-H2…N3	1.79 2.31	154.07 175.23	-0.24
98	A:U h:s T (19/0)		N4-H4…O2'	2.03	165.56	-2.48
99	A:C s:h C (8/0)		N4-H4…N3 C5-H5…O2'	2.11 2.56	160.22 141.56	-6.36
100	C:C h:s C (9/7)	1	N4-H4…O2 C5-H5…O2'	1.953 2.741	159.10 125.82	-1.00
101	C:U h:s C (3/0)		N4-H4…O2 C5-H5…O2'	2.022 2.565	171.54 152.86	-4.63
102	A:C w:s T (3/0)		N6-H6…O3' O2'-H2'…N1	2.12 1.76	172.29 165.79	-4.31
103	A:C h:s T (0)		C8-H8…O2 O2'-H2'…N7 N6-H6…O2'	2.49 1.76 2.04	119.97 153.93 154.88	-1.81
104	A:C s:h T (0)		N4-H4…O2' C5-H5…N3	2.100 2.506	167.82 158.74	-5.97
105	A:U s:h T (2/2)	1	C5-H5…N3 O2'-H2'…O4	2.364 1.841	166.43 171.86	-3.13
106	G:C h:s T (0)		O2'-H2'…N7 C8-H8…O2	1.775 2.029	171.85 157.11	-8.58
107	C:C h:s T (9/2)	1	O3'-H3…O2 N4-H4…O2'	1.86 2.03	175.32 153.12	-8.74
108	C:U h:s T (6/0)		N4-H4…O2' C5-H5…O2	2.138 2.366	151.62 160.27	-4.60
109	C:U s:h T (0)		C5-H5…O2 O2'-H2'…O4	2.66 1.96	119.47 149.16	-7.82
110	U:U h:s T (0)		O2'-H2'…O4 C5-H5…O2	1.72 2.23	177.47 145.57	-7.12

Protonated basepair						
111	A:G +:H C (4)		N6-H6…O6 N1-H1..N7	1.733 1.837	173.26 172.10	-40.90
112	A:C +:W C (30)		N6-H6…N3 N1-H1…O2	2.017 1.628	177.00 178.63	-39.84
113	A:C W :+ T (0)		N6-H6…O2 N3-H3…N1	2.034 1.587	164.56 178.30	-30.75

114	G:G H:z T (2)		N3-H3…O6 N2-H2..N7	1.683 1.884	159.58 179.90	-47.02
115	G:C W:+ C (11)		N3-H3…O6 N1-H1…O2	1.453 2.019	168.51 171.00	-39.97
116	G:C H:+ C (0)		N4-H4…O6 N3-H3…N7	1.725 1.718	168.15 174.50	-44.95
117	G:C z:W C (16)		N2-H2…N3 N3-H3…O2	2.023 1.595	176.65 177.27	-43.78
118	G:C H:+ T (4/2)	2	N3-H3…O6 N4-H4…N7	1.677 1.860	173.52 170.92	-40.87
119	C:C W:+ C (8)		N4-H4…N3 N3-H3…O2	1.995 1.644	177.35 173.05	-38.51
120	C:C W:+ T (1)		N4-H4…O2 N3-H3…N3 N4-H4…O2	1.646 1.768 1.960	177.47 179.09 171.90	-47.88
121	C:U +:W C (1)		N3-H3…O4 N3-H3…O2	1.549 2.126	166.70 171.09	-29.42
122	C:U +:W T (0)		N3-H3…O2 N3-H3…O2	1.584 2.122	172.51 165.63	-26.68
123	A:G z:h C (0)		N3-H3…N7 C8-H8…O2’	1.788 2.257	171.69 156.10	-34.79
124	G:C S:+ T (0)		N2-H2…O2 N3-H3…N3	1.824 2.003	179.04 171.67	-10.33
125	A:G z:h T (0)		N3-H3…O6 C2-H2…N7	2.015 2.818	164.89 133.47	-10.73
126	C:U +:S T (0)		N3-H3…O2 N3-H3…O2	1.835 1.793	174.82 170.92	-3.46



Web www.saha.ac.in/biop/www/db/local/BP/rnabasepair.html Search with Google

Non-canonical RNA Base Pair Database

- This database provides information about different types of possible basepairs occuring between the nucleotides of RNA. The number of geometries possible between any two base edges are hyperlinked. The basepairs are detected by BPFIND with atleast two hydrogen bonds between the constituents bases from a non-redundant set of structures determined by X-ray crystallography.

-The best structures have been geometry optimized by DFT method. Certain basepairs which occur very rarely, some with very low frequency also are modelled using NUCGEN prior to geometry optimization. .

	Ade W	Ade H	Ade S	GuaW	Gua H	Gua S	Cyt W	Cyt H	Cyt S	Ura W	Ura H	Ura S
Ade W	two	two	three	two	one	three	four		two	two	two	two
Ade H		two	two	two		one	two		two	two	two	two
Ade S			two	two	two	three	two	two	one	two	two	one
Gua W				one	two	two	three		two	two	one	two
Gua H					one	three	two	two	one	two	one	
Gua S						two	four		one	two	one	one
Cyt W							three	two	one	four	one	one
Cyt H									two			two
Cyt S										two	one	
Ura W										three	two	two
Ura H											one	one
Ura S												

The interaction energy associated with the basepairs are also calculated, which gives us information about the stability of the basepairs. The strucural parametrs which defines the dynamics of the basepairs are calculated by NUPARM. NUPARM-server can also be used from the MBU-IISc webserver.The basepair parametrs are calculated and tabulated for each of the basepairs. The three rotational parametrs buckle, open angle, propeller are given in degree, while the three translational parameters are given in angstrom.

Figure 3.3.1. RNA non-canonical basepair database home page

The total numbers of geometries that can occur between two specific edges are hyper-linked to details of these basepairs (Figure 3.3.2). Some of the matrix elements are blank as basepairing by the formation of two H-bonds are not possible between these edges. Each of these pages gives information about the type and orientation of the basepairs, their frequency in the duplex and base triplets, example of one of such kind of basepair, whose structural parameters are closest to their respective average values along with the residue number and PDB file name and the hydrogen boding schemes for the basepairs.

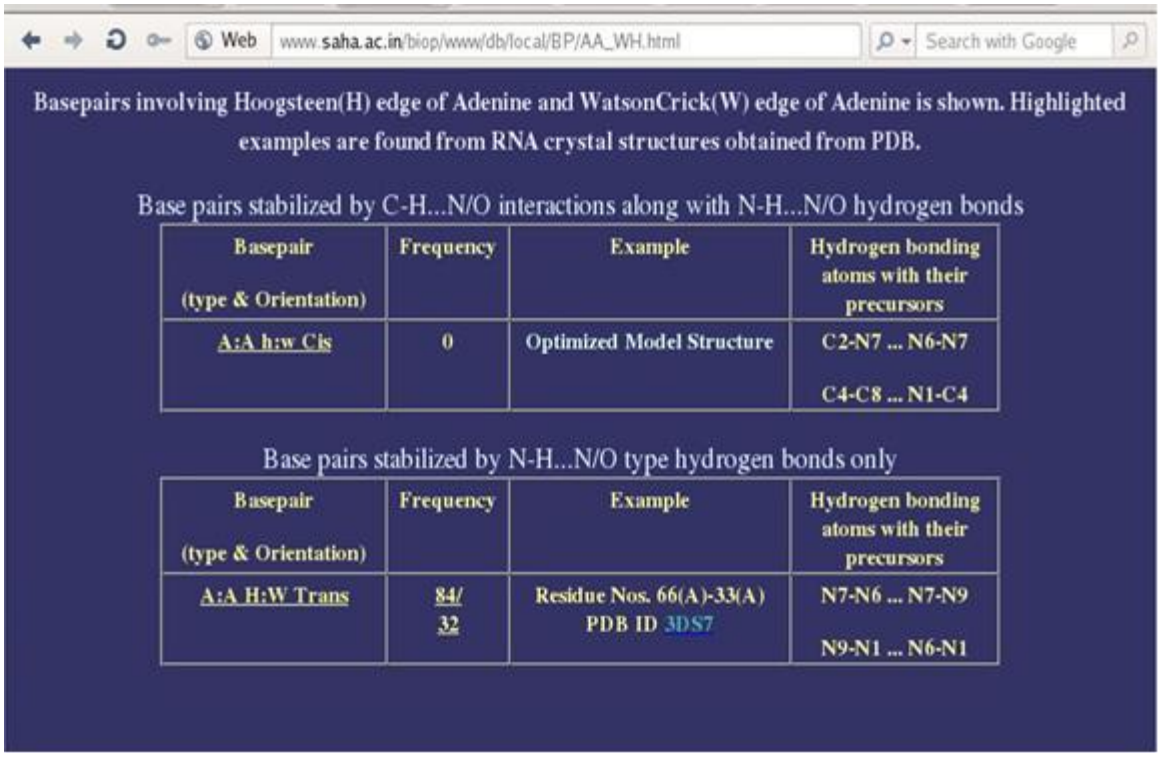


Figure 3.3.2. The hyperlinked page gives information about possible basepairing involving different edges of the basepairs along with the frequency (both in duplex and triplets), one example of such basepair along with PDB ID and residue number, and their hydrogen bonding precursors

Each basepair type is hyper-linked to its optimized geometry along with hydrogen bond lengths, basepair parameters and BSSE corrected interaction energy (Panigrahi et al., 2011b, Roy et al., 2008, Chawla et al., 2011). It also gives us information about the isosteric basepair parameters of the optimized structures, such as open angle, glycosidic angles, C1'-C1' distances and C8-C6 distances (Figure 3.3.3).

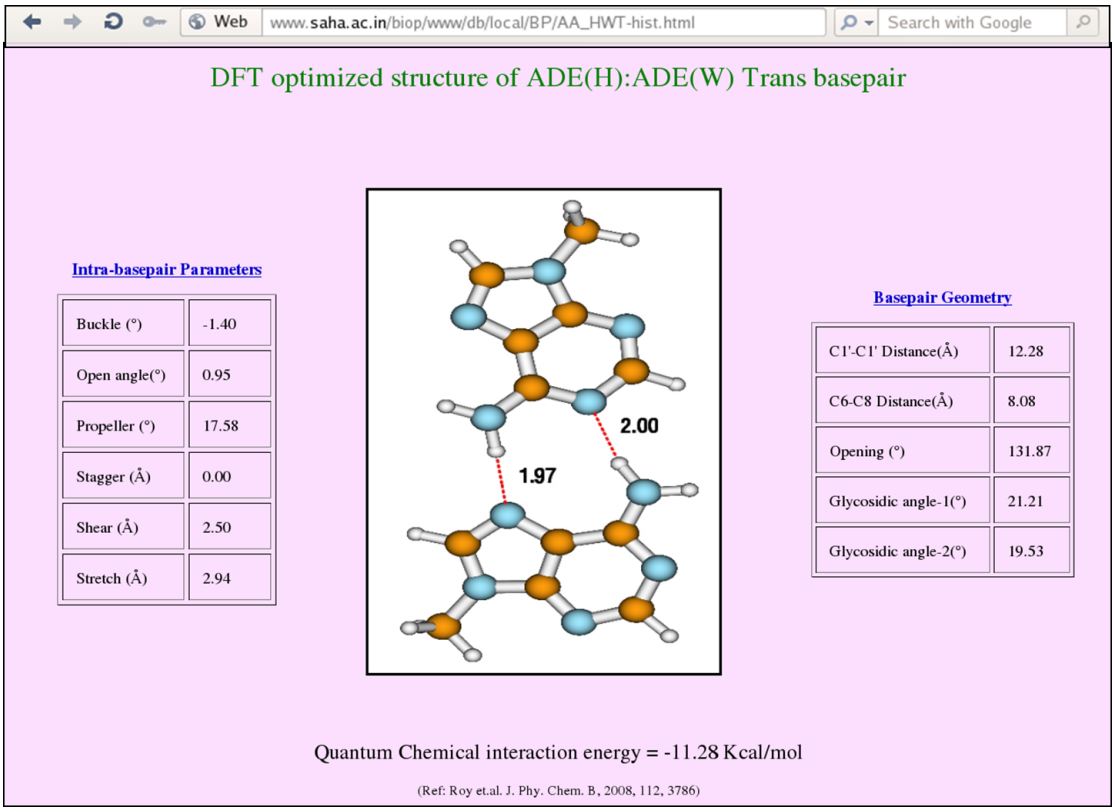


Figure 3.3.3. Structure of the optimized geometry along with the basepair parameter values, isosteric parameters, interaction energy

Frequency of the duplex (which has frequency greater than 30) is hyper-linked to the statistical analysis of the basepair parameters as calculated from the non-redundant crystal data set in the form of histogram plot (Figure 3.3.4). I have also given the histogram plot of the distribution of E-value of the basepair parameters following Das et. al (Das et al., 2006) using the following equation

$$E = \sum_i (d_i - 3.0)^2 + 1/2 \sum_j (\theta_j - \pi)^2 \qquad \text{.....[3.3.1]}$$

Where d_i is defined as the distance between the donor and acceptor atoms of a hydrogen bond between two bases under consideration, θ_j are the pseudo angles in radian formed by the precursor atoms of the bases.

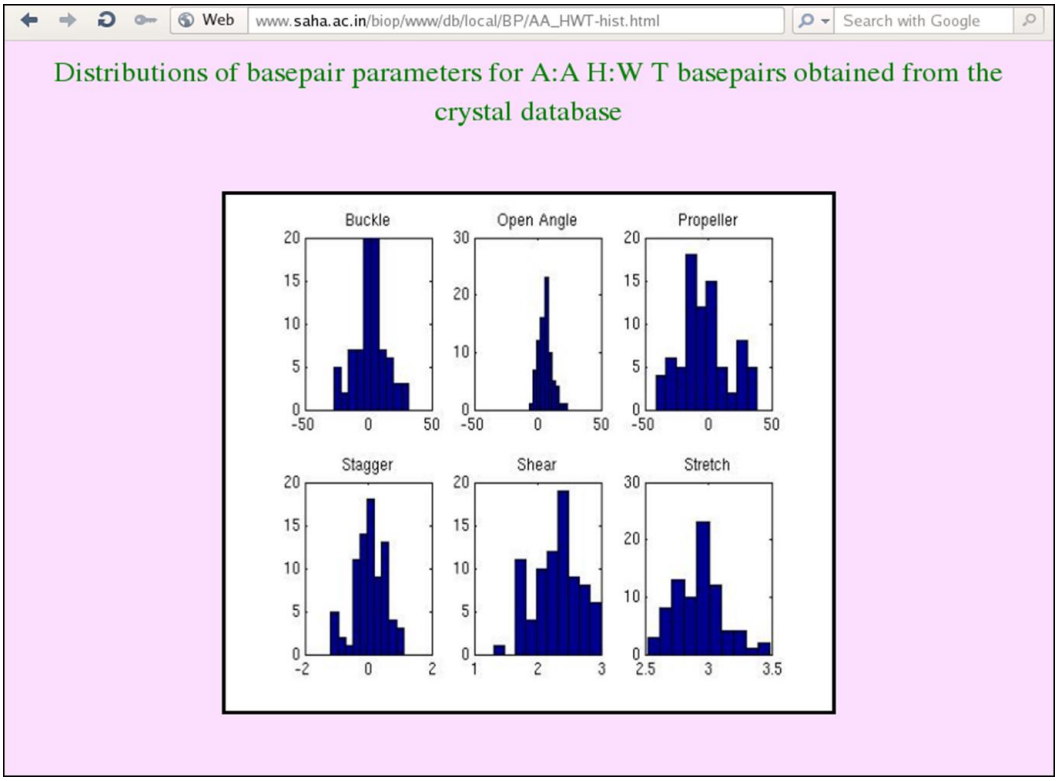


Figure 3.3.4. Distribution of basepair parameters for A:A H:W T basepair (From Crystal database)

Many of the observed basepairs are parts of base triplets. Basepairs bonded through hydrogen bonding interaction through specific edges, expose their other edges to interact with other bases through hydrogen bonding interaction forming triplets (Figure 3.3.5). I have calculated total number of base triplets associated with each basepair type. The total number of triplets is hyperlinked to crystallographic structures of representatives of each

type of base triplets associated with the basepair, along with the PDB ID and residue number from where it is taken. It also gives the information about frequency of occurrences of such types of triplets associated with the particular duplex. Our database also includes information about the protonated basepairs detected by BPFIND, where unusual protonation at some particular sites of the involved bases leads to strong hydrogen bonds between the basepairs with significant stabilization energy.

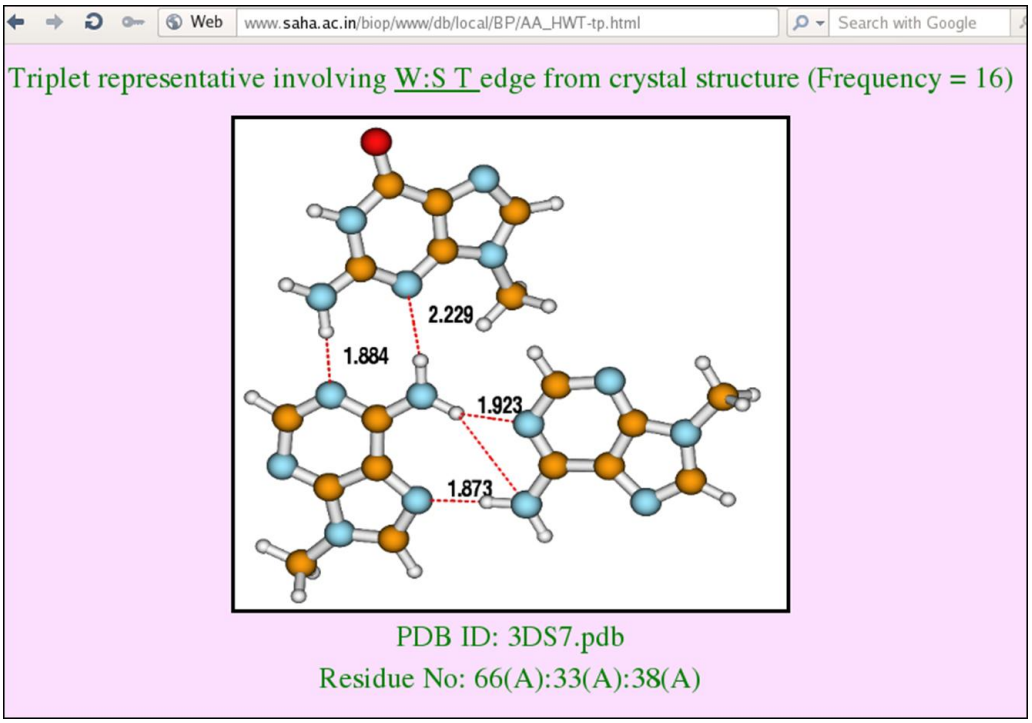


Figure 3.3.5. Triplet representatives of A:A H:W T basepair

3.3.3. Analysis

All the possible 126 basepairs (Table 3.3.1) can be classified in 5 categories depending on the nature of hydrogen bonding between them, basepairs stabilized by (i)

both N-H…O/N hydrogen bonds (polar basepairs), basepairs stabilized by one N-H…O/N and another C-H…O/N hydrogen bonds (non-polar basepairs), (iii) basepairs stabilized by both C-H…O/N types of hydrogen bonds (CHO mediated basepairs) (iv) basepairs stabilized by protonation of one of the bases (protonated basepairs) (v) basepairs stabilized by one hydrogen bond between the bases and another hydrogen bond involving 2'-OH group of ribose sugar and another base. The frequencies of the occurrences of the basepairs are given in Table 3.3.1. I observed that in addition to the frequent Watson-Crick basepairs, several other non-canonical basepairs also occur frequently (the frequency of occurrences are given in parenthesis), among them G:U W:W C (847), A:G H:S T (531), A:U H:W T (417), A:A H:H T (110), and A:G W:W C (151) are observed with frequency value greater than 100.

The C-H...O mediated basepairs are generally very rare and I have modeled all of them, since it is difficult to get a proper representative. We had studied some C-H…O mediated basepairs in the earlier chapters. I observed that some of the non-polar sugar mediated basepairs also occur frequently, such as A:C s:s C (137), A:G :w s T (85), A:U w:s T (70) in addition to the polar sugar mediated baseairs A:A W:S C (86), A:C W:S C (62), A:G S:S C (35). For the protonated basepairs, I am presenting the data as calculated by Chawla et. al, however some of the basepairs which are found to be highly frequent by Chawla et. al, are found to be less frequent in the non-redundant dataset, however their database is found to be different from our non-redundant database. I have also modeled three protonated basepairs, G:C S:+ T, A:G z:h T and C:U +:S T and analyzed them, which were not considered earlier.

I observed the variation in bond lengths and angles of all the optimized basepairs. However, in this section, I have given more emphasis on the variation of C-H…O/N bond lengths in the non-polar and C-H…O mediated optimized basepairs due to availability of large number of theoretical data. The average and standard deviation of the C-H…O, C-H…N, N-H…O and N-H…N bond lengths obtained from the optimized geometry are given in table 3.3.2 and their distributions are presented in Figure 3.3.6.

Table 3.3.2. Average and standard deviations (SDV) of hydrogen bond lengths

Nature of Hydrogen bond (No. of data)	Average (SDV)
C-H…O (26)	2.38 (0.31)
C-H…N (32)	2.49 (0.14)
N-H…O (73)	1.85 (0.14)
N-H…N (76)	1.97 (0.12)

I observed C-H…O bond lengths ranges between 2.0 Å to 2.5 Å, except an outlier value 3.72 Å, which corresponds to bond length value for the C:C w:h C basepair, which is one of the model basepairs. The initial geometry of this basepairs have two H-bonds between N4-H4…N3 and C5-H5…O2, with proper H-bond distances and angle. However, its optimization leads to a much distorted geometry, where amino group ($–NH_2$) acts both as a donor and acceptor, with lengthening of the C-H…O bond. The C-H…N bond length varies widely from 2.0 Å to 2.8 Å. I observed that N-H…O and N-H…N hydrogen bond lengths are generally smaller. These bond lengths are anti-correlated to the hydrogen bonding strength as reported earlier (Section 3.2). The H-bonding energy for N-H…O hydrogen bond is strongest and such high energy brings the hydrogen atoms

closest to the acceptor oxygen atoms. On the other hand C-H…N bond is the weakest and hence the hydrogen atoms remain far from the acceptor nitrogen atoms.

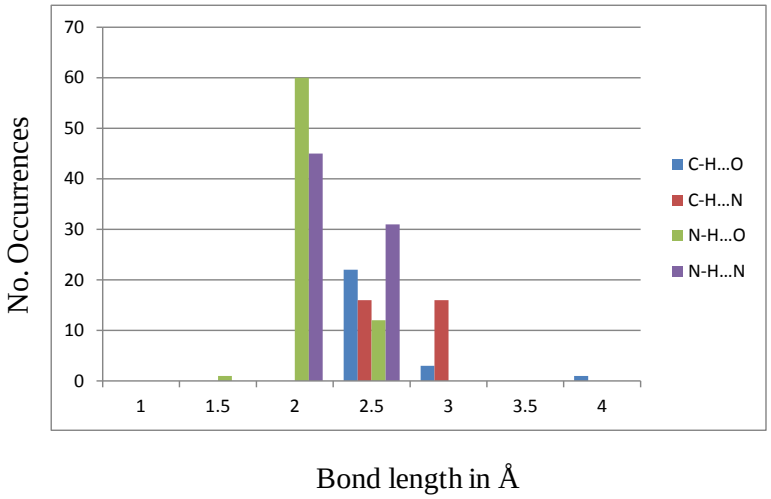


Figure 3.3.6. Histogram plot of distribution of bond lengths for all the 126 optimized basepairs

Frequent basepairs, with N-H…O/N types H-bonds, are found to be very stable in terms of their interaction energy strengths (Panigrahi et al., 2011b). Interaction energies of the polar and non-polar basepairs are given in table 3.3.1. The average interaction energy of the polar mediated basepairs is calculated as -11.93 kcal/mol with standard deviation value as 3.68. From the Table 3.3.1, I observe that interaction energy of the G:G W:W T model basepair is found to be -21.77 kcal/mol (Table 3.3.1) which is comparable to interaction energy of that of the most frequent G:C W:W C basepair. Its optimized structure has two most strong N-H…O bonds. The non-polar basepairs have

average interaction energy as -6.81 kcal/mol with standard deviation value as 1.86. I observe that some of the C-H…O mediated model basepairs are very stable in terms of their interaction energy value (Table 3.3.1), for example the G:U w:h T basepair has -11.98 kcal/mol interaction energy, which is comparable to the polar basepair, this basepair is found to be stable with formation of two stable N1-H1…O4 and C5-H5…O6 H-bonds, the basepair parameters values also signifies stable geometry. Eight basepairs stabilized by two C-H…O/N mediated hydrogen bonds, have interaction energy much lower than the other two types of hydrogen bonded basepairs as discussed above as these bonds are very weak in nature with large bond lengths values as shown in Figure 3.3.6. Among the modeled C-H…O mediated basepairs, A:U h:h T, A:U w:h T, G:U h:h T optimized basepairs are associated with large open angle values. This indicated that the basepairs were not stable in the initial orientation and some other orientation was preferred. Their interaction energy lies within the range of -0.72 to 5.6 kcal/mol but the H-bonds retain in the optimized structures.

In my previous section (section 3.2), I have discussed about 5 frequently observed sugar mediated basepairs, where in all the cases O2'-H2' of the ribose sugar act as donor. The 2' hydroxyl group attached to the C2' of the sugar moiety can act as both H-bond donor and acceptor, and plays important role in stabilizing the tertiary structures of RNA. Many such examples were observed of the model basepairs, where O2' can act as an acceptor also. The interaction energy varies in wide range for the sugar mediated basepairs (Table 3.3.1). Greater values of interaction energy are obtained for the G:U W:S C (E_{int}= -12.96 kcal/mol), A:U H:S C (E_{int}= -12.96 kcal/mol), G:C S:W C (E_{int}= -10.91 kcal/mol) and G:U W:S T (E_{int}= -15.48 kcal/mol) basepairs. Most of these

basepairs are associated with large amount of buckle and open angle values. It may be noted here that formation of these basepairs require specific sugar puckering and glycosidic torsion angle. It was earlier studied that some of these are between successive residues and are held together by sugar-phosphate backbone (Sharma et al., 2010a). Recent quantum chemical calculations by Sponer and co-workers have proved the importance of backbone conformations in enhancing the stability of the GPU dinucleotide platform, which is mediated by O2'…O2P H-bond, in addition to H-bond between the bases (Mládek et al., 2012). As these models were always using typical C2'-endo sugar conformation and the optimization process could not convert the sugar pucker to the required form. This is presumably the reason behind non-planar structures of these basepairs with highly deformed geometry.

Distributions of the basepair parameters from the crystal data base show that buckle, open angle, propeller, stagger and shear show normal distributions with mean zero, while stretch values shows normal distribution with mean value around 3.0 for the standard canonical basepairs such as G:C W:W C and A:U W:W C. The crystal analysis of the G:U wobble basepair shows distribution of shear value around -2, unlike that of the usual canonical basepairs, which is in agreement with the value obtained from the optimized geometry. Among the non-WC basepairs, A:U H:W T, A:U W:W T, and G:G W:W T shows normal distribution of their basepair parameter values. Most of the sugar mediated basepairs show mean stretch value close to 4. Optimization of most of the sugar mediated basepairs, either they have taken from the crystal structures or modeled basepairs results in unusual geometry.

I have also analyzed all the types of base triplets found with each base, involving its different edges. The numbers of triplets associated with each of basepair are given in the database. I observed the canonical basepairs G:C W:W C and A:U W:W C have 13 such base triplet types, which signifies their greatest involvement in base triplets formation. Several non-canonical basepairs such as G:U W:W C, A:G H:S T, A:U H:W T have also associated with large number of triplet types. I observed significant frequencies of base triplets formed by A:A H:W T (frequency = 32) , A:U H:W C (frequency = 14) and A:A W:W T (frequency = 39) basepairs as given in Table 3.3.1. I observed that some of the rare basepairs (frequencies in basepaired state given in parenthes) (Table 3.3.1) are also observed as part of base triplets. Among the protonated basepairs, only G:C H:+ T is observed as part of triplets. I have also classified the total number triplets associated with each doublet in to different types depending on the basepairing edges of the third base, and the first base. Our database gives information of number of such variants and also we provide an initial representative of best representative in terms of E-value of each of variant type, along with its PDB and residue information.

3.3.4. Discussion

We have prepared a database of basepairs, including protonated basepairs, base triplets from crystal structure. The number of base triplets formation are found to be very large. From the database it is evident that the structures of base triplets can be very diverse and be used to build distinct RNA structural elements. Potential applications of triplets are however not properly understood yet. Optimization of the observed base

triplets may open some new insights in understanding the functions of riboswitch, ribozyme etc. and may help in designing new therapeutic applications.

Chapter 4

Quantum Chemical Analysis of Citrate Capped Gold Nanoparticle-Quercetin Complex

4.1. Introduction

Metallic nanoparticles have attracted fundamental interest in recent years because of their unique chemical and electronic properties and have practical interest for wide varieties of potential applications. As property of a material depends on its atomic arrangement and in nano meter size metal, local disorder in atomic arrangements results in unique electronic properties in the material (Petkov et al., 2005). In recent years, considerable attentions have been paid to the synthesis and characterization of nano particle-mediated therapeutic treatments. Gold nanoparticles (GNP) have received importance in wide areas like catalysis, biolabelling, nonlinear optical devices etc. These have applications in drug delivery process also, due to their unique physical and chemical properties (Daniel and Astruc, 2004). Gold, in the normal state is one of the most inert materials, as it does not cause any serious side effects in biological systems. Type I diabetes, commonly known as insulin dependent diabetes mellitus (IDDM) can be produced in animals by the action of chemicals or by immune attack on pancreatic beta cells. In both the cases, oxygen free radicals are produced and play crucial role in beta cell killing (Oberley, 1988). GNP can act as an anti-glycating agent. This anti-glycating properties of the gold nano particle has been proved in eye protein known as α-crystallin (Singha et al., 2009). Glycation in diabetic patient also results in damage of physiologically important proteins like albumin, collagen etc. So GNP-quercetin complex can serve as a therapeutic agent in nanomedicine, where GNP acts both as a carrier and anti-glycating agent. GNP has been reported to act as antioxidant (BarathManiKanth et al., 2010). Gold nanoparticles can be used to assay antioxidant activity of several

phenolic acids and several food samples (Wang et al., 2007). In several studies gold nanoparticles have been used for drug delivery (Wieder et al., 2006, El-Sayed et al., 2006, Huang et al., 2008), and delivery of proteins, peptides, and oligonucleotides (Paciotti et al., 2006, Ghosh et al., 2008, Han et al., 2007a).

Flavonoids are present in many plant based food items and beverages, and have been shown to possess important therapeutic activities. Quercetin is the most abundantly consumed bio-flavonoid and is present in high concentrations in tea, apple, and onion. The flavonoid exhibits antioxidant (Anjaneyulu and Chopra, 2004) and anti-inflammatory properties (Guardia et al., 2001), which helps protecting against heart diseases and cancer. They scavenge damaging particles in the body known as free radicals, which damage cell membranes, interacts with DNA, and sometimes cause cell death. It is also known to maintain the blood glucose and insulin level in diabetic condition (Oberley, 1988, Vessal et al., 2003, Shetty et al., 2004). Wu and Yen (Wu and Yen, 2005) have reported the antiglycating property of the flavonoid. Therefore, it is expected that GNP-quercetin complex can serve as a better therapeutic agent in nanomedicine.

There have been several theoretical studies on the interaction of gold nano clusters with different nucleobases (Shukla et al., 2009, Kryachko and Remacle, 2005, Sharma et al., 2007). Electronic properties of gold nano cluster have been studied extensively using density functional theory (DFT) by different groups (Deka and Deka, 2008, Wang et al., 2002, Sankaran and Viswanathan, 2006). Different physical and cemical methods are available for preparation of gold nanoparticles. It has been well established that due to decrease in the size of the nanoparticles, two additional effects,

namely (i) quantum confinement and (ii) increase in the fraction of surface atoms relative to the total number of atoms in the nanoparticles, become important factors. Due to the above properties, the naked nanoparticles tend to nucleate and aggregate, so they are often capped with molecular species to prevent such aggregation (Murray et al., 1993). The nature of capping mostly depends on the method of synthesis of GNP. The reduction procedures for preparation of GNP generally give rise to capped molecules. They are usually capped with thiol, citrate groups, aminoacids etc. (Jadzinsky et al., 2007, Pal and Chakraborti, 2010, Selvakannan et al., 2003, Ding et al., 2008). However, the proper modelling of the capping of GNPs have not been done to reproduce the actual experimental findings. Besides these methods, other relatively cheap and environmentally friendly methods are also available for preparation of GNP, such as arc discharge and spark discharge methods (Lung et al., 2007, Tseng et al., 2008). Both the methods require water medium for production of GNP, where water molecules or hydroxyl groups possibly act as capping agents. Although GNP micelles are formed in water medium in both these methods, proper molecular structure of the complex has not been proposed.

It was shown that formation of stable quercetin-gold nanoparticle complex (GNPQ) is possible by sodium citrate reduction method (Pal and Chakraborti, 2010). This finding has prompted us to investigate the interaction of GNP with quercetin in more detail. To mimic the experimental findings, a miniature model of gold nanoparticle along with the citrate capping has been taken as the initial geometry to study its interaction with the quercetin molecule. To my knowledge, this is the first trial to theoretically represent

the proper geometry of the gold-citrate capped complex. Density functional theory (DFT) has been employed to understand the properties of GNP and GNPQ.

4.2. Methods

Synthesis and characterization of the gold nano particle quercetin complex (GNPQ) has been carried out and in vivo analysis of the GNPQ has been also done by my experimental collaborators (Pal et al.).

4.2.1. In Silico Studies

Modeling of the molecular systems has been done with the help of MOLDEN (Schaftenaar and Noordik, 2000) and Accelrys Discovery studio 2.5 software (Studio, 2007) maintaining the standard bond lengths and angles between the atoms involved. The initial geometry optimizations were carried out by density functional theory (DFT) using the plane-wave pseudo-potential approach within the generalized gradient approximation (GGA) (Perdew, 1986, Perdew and Yue, 1986, Perdew and Wang, 1992) as implemented in the Amsterdam Density Functional package (ADF) (Te Velde et al., 2001, Fonseca Guerra et al., 1998). All the quantum chemical calculations were carried out with Perdew−Wang exchange and correlation functional (PW91) (Perdew et al., 1996, Perdew et al., 1992). Valence triple ξ Slater basis set with polarization functions (TZP) has been used for all atoms. As gold has 79 protons, the electron around the heavy nucleus moves very fast with a velocity close to the speed of the light. This causes increase in its mass as elaborated in section 2.15. Hence, we have used DFT with relativistic corrections. The same GGA/PW91 approach has been applied for the relativistic approach also, but this

time with valence triple ξ and doubly polarized basis sets (TZ2P), along with the zeroth order regular approximation Hamiltonian (ZORA) (van Lenthe et al., 1993, van Lenthe et al., 1994), which is an approximate method for solving Dirac equation (Dirac, 1928) as on increasing the size of atom, relativistic effect becomes more important.

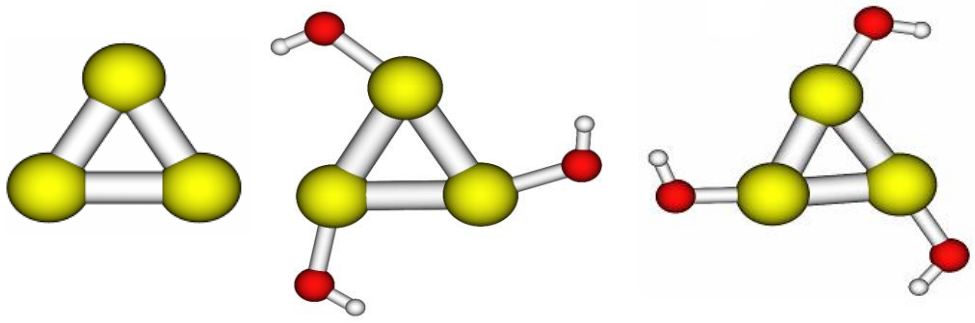
To correlate the results obtained from of the ADF software, Gaussian g09 package (Frisch et al., 2009) has been used for all the electronic structure calculation. Full geometry optimization of the 3gold-3citrate capped system (without any constrain) has been carried out with the most widely accepted B3LYP (Becke, 1993, Lee et al., 1988) (Becke three parameters hybrid functional with Lee-Yang-Parr) functional using 6-31G** basis set along with effective core potential LANL2DZ (Ehlers et al., 1993) for the gold atoms to take in account of the relativistic effect. The DFT calculation along with the LANL2DZ basis set is a commonly used method for gold cluster (Sankaran and Viswanathan, 2006). The harmonic frequencies were calculated using the same level of theory used for the optimization process.

4.3. Results and Discussions

4.3.1. Structural Analysis

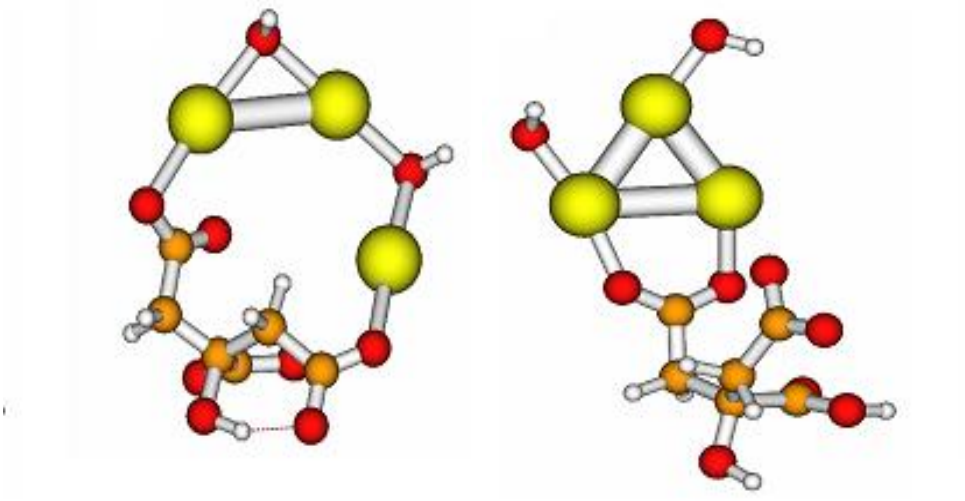
As quantum chemical calculations for many electron atoms are possible for small molecules only, in this present investigation, a triangular gold cluster (Au_3) has been considered as starting geometry for quantum chemical calculations. Such approximation could be made as it has been observed by various groups that the gold reactivity increases with the presence of highly non or low co-ordinate gold atoms (Wells, 2004, Wells Jr et al., 2002, Tang et al., 1998). In the initial geometry the Au-Au bond lengths are assumed

to be 2.70 Å and Au-Au-Au bond angles are kept at 60^{o} (Jadzinsky et al., 2007). Optimization of the triangular Au_3 has been carried out using GGA:PW91/TZ2P method as it contains only the heavy gold atoms, the optimization leads to similar stable geometry with the bond lengths reduced to 2.66 Å (Figure 4.1a) (Jadzinsky et al., 2007). However, the process of preparation of gold nano-particle by citrate reduction indicates presence of capping radicals associated with Au atoms. It may be mentioned here that the arc discharge process also is likely to give some capping (possibly –OH groups) to gold nano-particles as the dangling bonds would readily be neutralized by the environment. Thus, I have tried to study gold-nano particle with different capping groups. Taking in account of the previous theory, preliminary Au_3 structure has been capped with -OH groups (Figure 4.1b,c) and bond lengths and energy of the final optimized geometry are presented in Table 4.1.



(a) (b) (c)

Figure 4.1. Optimized Structure of (a) triangular Au_3 (Relativistic) (b) Au_3 capped with hydroxyl (-OH) groups (Non-relativistic) (c) Au_3 capped with hydroxyl (-OH) groups (Relativistic)



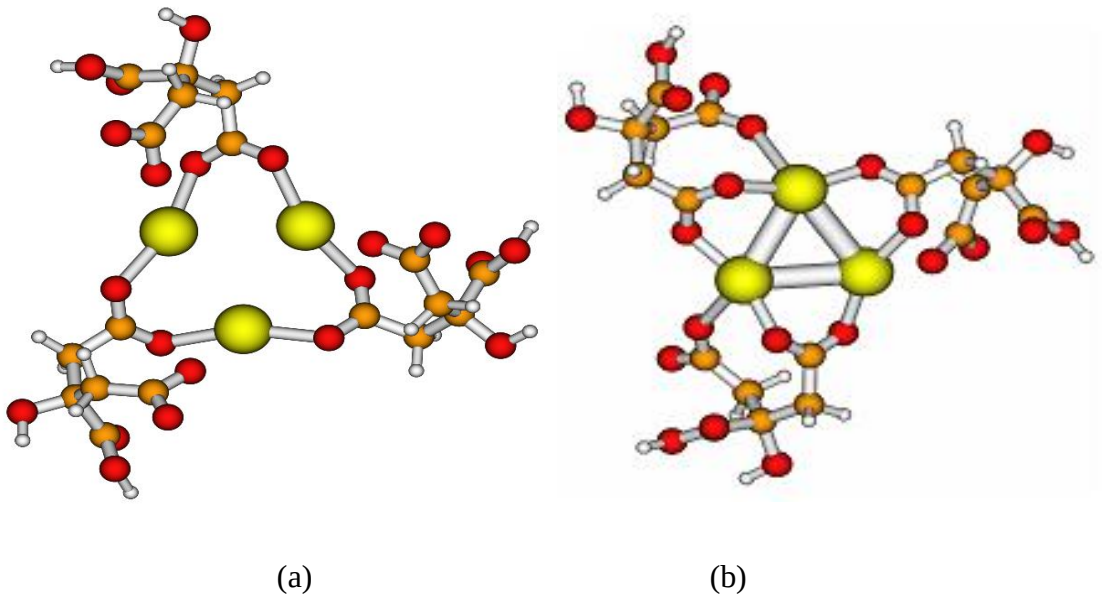
(a) (b)

Figure 4.2. Optimized geometry of the Au_3 capped with one citrate group by (a) non-relativistic approach (b) relativistic approach

I have further attempted to study gold clusters capped with two hydroxyl –OH groups and one citrate radical, assuming the experimental procedure of preparation of gold nano particles may keep some citrate groups. This structure almost dissociates during optimization by non-relativistic approach (Figure 4.2a). However optimization through relativistic approach using GGA:PW91/TZ2P method leads to stable geometry with intact Au-Au bonds, although with larger bond lengths (Figure 4.2b).

This was attributed to the limitations of the DFT method in approximating electron i.e. configuration of the heavy elements. ZORA has been incorporated with the ADF package especially for heavy atoms, where relativistic approach plays a crucial role in stabilizing the systems. As in case of relativistic approach of an atom, the effective

Bohr radius decreases for the inner electrons, the core electrons move with large average speed. This result in energy stabilization and relative radial contraction of the s and p orbital those lie closer to the nucleus, while d and f orbital (Chemical configuration of gold is [Xe] $4f^{14}$ $5d^{10}$ $6s^{1}$) are destabilized and tend to expand radially and outwardly. However the higher s-shells also suffer similar contraction, as they are orthogonal to the lower ones. So according to this theory the outer 6s orbitals of gold also suffer similar contraction, and even larger than the 1s orbital. This contraction gives a stabilization of the 6s shell in gold atom, which leads to an increase in the affinity of the gold atom (Pyykko, 1988).



(a) (b)

Figure 4.3. Optimized geometry of the Au_3 capped with three citrate groups by (a) non-relativistic (b) relativistic approach

This signifies that symmetry plays a very crucial role in stabilizing the systems. So I approached to design the system with a better symmetry and have put three-citrate caps near the three gold atoms of the triangular cluster maintaining proper symmetry. The optimized geometries obtained through non-relativistic and relativistic approach are shown in Figure 4.3.

Table 4.1. Bond lengths and energy of the optimized geometry

System no	System Type	Methods	Bond lengths (Au-Au) in Å	Energy in eV
1	Au_3	Relativistic	2.66	-4.07
		Non-relativistic	2.70	-3.92
2	Au_3-3OH	Relativistic	2.58	-35.66
		Non-relativistic	2.92	-32.20
3	Au_3-2OH-1citrate cap	Relativistic	2.58 2.73 3.00	-148.11
		Non-relativistic	3.21 5.28 3.61	-152.00
4	Au_3-3citrate cap	Relativistic	3.12 3.08 3.10	-373.99
		Non-relativistic	3.51 3.52 3.51	-370.98
5	Au_3-3citrate cap+ quercetin	Relativistic	3.09 3.05 3.05	-593.59

Optimization using relativistic effect leads to symmetrical geometry (Figure 4.3b), with oxygen of citrates attached to the gold atoms in a similar geometrical fashion in terms of their bond lengths (Table 4.1). All the energy and the bond-lengths of the optimized geometries of all the systems studied and as obtained by relativistic and non-relativistic approaches are given in Table 4.1.

Following Variational principle of quantum mechanics, the relativistic Hamiltonian always gives more stable energy than that of the non-relativistic Hamiltonian. The bond lengths with ZORA are smaller, indicating higher binding energy between the Au atoms. I have also carried out Mullikan charge analysis of the 3gold-3citrate capped system. A representative Au_3 cluster with one citrate group along with the nomenclature for the constituent's carbon, oxygen, and hydrogen atoms of the citrate group has been given in a schematic diagram (Figure 4.4).

The average of the Mulliken charges of the similar kinds of atoms for all the three citrate groups (Following the nomenclature of Figure 4.4) has been calculated and presented in Table 4.2. The standard deviation values of the charges in all the cases are found to be close to zero. From the table 4.2, I observed a symmetrical distribution of charges in the citrate groups, which lies towards the surface of the system. I observed that accumulation of electropositive charges is more prominent in case of H6 and H3 of OH6-H6 and O3-H3 bonds, which are the hydrogen bond donor groups. There are also numbers of hydrogen bond acceptors, those are present in the citrate groups, such as O6, O11, O12, O51, O52, O6 atoms and I observed they acquired more amounts of electronegative charges.

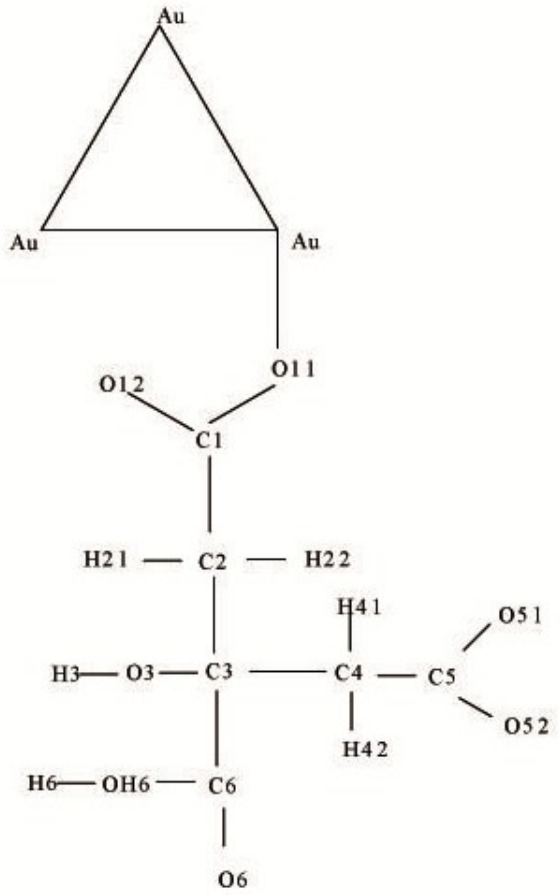


Figure 4.4. Schematic diagram of Au_3-1citrate complex along with the atom type

The oxygen atoms forming bonds with Au, namely O11, O12, and O51 always are more negatively charged. These hydrogen bond donor and acceptor groups of the citrate molecule may interact with quercetin by forming strong hydrogen bonds. The oxygen atoms, which lie outwardly interact with quercetin leading to decrease in surface potential, which is evident from the zeta potential analysis as explained by Pal

and Chakraborti (Pal and Chakraborti, 2010), leading to formation of stable gold-quercetin complex.

Table 4.2. Geometry and the Mullikan charges of the 3gold-3citrate capped system

Distance (Au…Au) in Å	Distance (Au…O) in Å	Angle <O-C-O> in (^{o})	Average Mullikan charges on the carbon atoms of citrate group	Average Mullikan charges on the oxygen atoms of citrate group	Average Mullikan charges on the hydrogen atoms of citrate group
3.07 3.06 3.05	2.07 2.08 2.08 2.07 2.07 2.08	122.627 122.614 122.627	C1 (0.71) C2 (-0.28) C3 (0.23) C4 (-0.29) C5 (0.63) C6 (0.54)	O11 (-0.46) O12 (-0.46) O3 (-0.54) O51(-0.45) O52 (-0.34) O6 (-0.39) OH6 (-0.48)	H21(0.20) H22(0.21) H41(0.15) H42(0.18) H6(0.36) H3(0.34)

4.3.2. Observed Interactions between the GNP – Quercetin Complex

As the three gold and three citrate capped complex system appears to be a reasonable miniature model for the theoretical calculation, I have attempted to study the effect of quercetin on the gold clusters. The final optimized geometry of the 3Au-3citrate capped structure (obtained by relativistic approach) has been taken and a single quercetin molecule was modeled near one of the citrate group, in such a way that H-bonding

interactions may occur and which may lead to a stabilized geometry. Optimization of this complex geometry has been carried out with GGA:PW91/TZ2P approach with ZORA Hamiltonian. The optimized structure shows a very stable geometry with formation of two stable O-H…O hydrogen bonds of bond lengths 1.57 Å and 1.76 Å between the citrate group and the quercetin molecule (Figure 4.5). The Au-Au bond lengths are obtained as around 3.05 Å, which is comparable with the bond lengths obtained from the crystal structure (Jadzinsky et al., 2007). These bond lengths are also found to be slightly smaller than that of the Au-Au bond lengths of symmetric Au_3-3citrate cap complex (Table 4.1), indicating that a single quercetin does not destabilize the capped cluster and perhaps make it stronger. This is also supported by the experimental finding that GNPQ is more stable than the GNP as found by TGA analysis (Pal et al.).

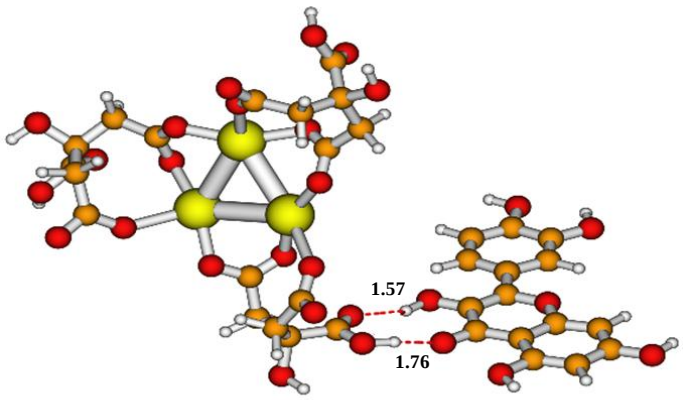


Figure 4.5. Optimized geometry of the citrate capped gold nano-particle along with quercetin with the hydrogen bond lengths are given in Å

With increase in the size of the gold clusters, the interaction of the quercetin with the citrate caps will become more effective and the number of hydrogen bond formation will also increase and the system will become more stable. The schematic diagram in the Figure 4.6 mimics the above type of complex system. The considered miniature model of GNPQ has only one quercetin molecule, but in nanometer size systems, there can be additional interactions between a quercetin molecule with multiple citrate groups. Moreover, the quercetin molecules also can stack with each other or form hydrogen bonds between themselves. This could give additional stability to the GNPQ, as found by the TGA analysis by my experimental collaborators.

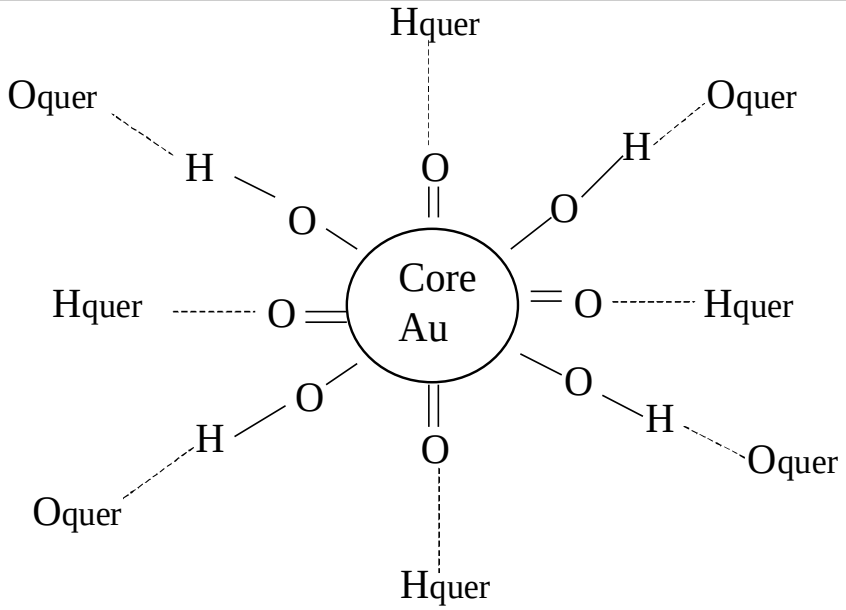


Figure 4.6. Schematic diagram showing the citrate capped gold nanoparticle interacting with the quercetin through hydrogen bonded interaction; the gold-nano particles are represented as the core Au

Among all the types of interactions observed in nature, hydrogen bonding interaction is dependent on the hydrogen ion concentration of the system, in other words on the pH value of the system. When the hydrogen-bonded system is kept at low pH, more number of hydrogen ions (H^+) is available in the system. This results in protonation of some of the hydrogen bond acceptor sites, so restricting the formation of some of the required hydrogen bond. Whereas if the system is kept at high pH, then the desired interactions will be restricted by relatively less concentration of the hydrogen, and in this case some of the important protonated group may become deprotonated. So in both the cases, i.e at both at high and low pH, the formation of desirable hydrogen bonds gets hindered. Experimental analysis also suggests that the interaction of the gold-citrate capped complex with quercetin is pH dependent, with optimum at around pH 6.2, which, in turn, proves that GNPQ formation is mediated through mostly hydrogen bonding interactions, whereas electrostatic and van der Waals interactions rarely play any role.

4.3.3. Elemental Analysis of the Complex Systems

Elemental analysis of the gold–citrate capped complex (GNP) and gold-citrate with quercetin complex (GNPQ) models have been carried out separately and presented in Table 4.3. In the gold-citrate capped complex, the percentage of mass of carbon is 27.8%, and that of hydrogen, gold and oxygen are calculated as 2.3%, 28.9%, 41 % respectively out of the total mass. However when it forms complex with quercetin, the percentage of mass of the carbon, hydrogen, gold and oxygen change to 35.2%, 2.6%,

21.0% and 41.2%. The increase in the mass of carbons and oxygen are due to binding of the GNP with the quercetin. Present theoretical calculation also correlates with the EDX spectra analysis performed by my experimental collaborators as given in Table 4.3 (Pal et al.).

Moreover theoretical study has several limitations (i) the theoretical system contains only three gold atoms, however in experiment, there are thousands of gold atoms form the nano particles and (ii) Experimentally EDX analysis scans the surface of the system; therefore it will not encounter gold atoms, as they always appear as capped with citrate groups, and therefore the gold percentage is found to be almost negligible in the GNPQ complex experimentally. So my theoretical results may not exactly match with the experimental data, however trends are similar as obtained experimentally.

Table 4.3. Elemental analysis of the GNP and GNPQ complex obtained by both experimental and theoretical calculation

Elements	Experimental EDAX Analysis (Wt %)		Theoretical Analysis (Mass %)	
	GNP	GNPQ	GNP	GNPQ
C	9.69	66.02	27.8	35.17
H			2.32	2.58
Au	75.85	0.25	28.9	21.05
O	7.14	25.74	40.98	41.21

4.3.4. Analysis of theoretical Infra-red (IR) spectrum

The theoretically calculated IR spectra of the quercetin and 3gold-3citrate capped system has been analyzed and is shown in Figure 4.7. The DetectMode program (Sen et al., 2006), which has been developed by our group was used to characterize the vibration patterns of both the systems. The frequency of the GNPQ complex could not be analyzed, since it required memory and scratch files of size larger than our computers capacity.

The peaks at frequency 3828 cm^{-1}, 3820 cm^{-1}, 3807 cm^{-1}, 3786 cm^{-1}, 3454 cm^{-1} correspond to –OH bond stretching vibration of the quercetin (Figure 4.7a). Among these, higher intensity (181.53 KM/mol) has been observed for the frequency 3454 cm^{-1}, which corresponds to coupled stretching vibrations of the some of the –OH groups of the quercetin. The band at around 1689 cm^{-1} represents the aromatic carbonyl C=O stretching, which corresponds to the highest peak with intensity 331.88 KM/mol. The band at around 1328 cm^{-1} is mainly due to C-O stretching of the phenolic –OH group of the quercetin. At lower frequency region the peaks observed are due to the coupled vibrations of the C=C stretching of the aromatic rings, associated with angle bending of the C-C-C, C-C-O, C-C-H bonds.

I have also analyzed the IR spectra of the gold citrate capped complex (GNP) (Figure 4.7b). The peaks around 3753 cm^{-1}, 3752 cm^{-1}, 3784 cm^{-1} correspond to –OH stretching of the citrate groups. Intensity is found to be very high (213.87 KM/mol) for the 3753 cm^{-1} peak, which corresponds to simultaneous stretching of the –OH bonds of

the citrate groups. From frequency region 3140 cm^{-1} to 3077 cm^{-1}, I observed C-H vibrations with low intensities. The carbonyl groups present in the gold-citrate complex can be categorized into three types, (i) those bind directly to the gold atoms (ii) C=O group with free oxygen (closer to gold atoms but do not form any bond with gold atoms) (iii) those which lies towards the outer surface of the system (can interact with the quercetin). The symmetrical and asymmetrical stretching of the carbonyl groups connected to Au atoms is observed at 1451 cm^{-1}, 1550 cm^{-1} and 1551 cm^{-1} respectively. The carbonyl stretching for the free oxygens are observed at 1703 cm^{-1} and 1704 cm^{-1}. Stretching and bending of the carbonyl groups of the citrate groups those lies towards the outer surface is observed at 1858 cm^{-1} and 1859 cm^{-1}. The highest peak is observed for the frequency around 1161 cm^{-1} with intensity 1003 KM/mol which corresponds to C-O stretching of the phenolic –OH group of the citrate molecules. At low frequencies 795 cm^{-1}, 621 cm^{-1} out of plane torsions, angle bending were observed, which are characterized by coupled movements of all the atoms in the molecules. The out of plane torsions also show wagging of hydrogen atoms of the $–CH_2$ groups. The IR spectra of the isolated citrate group has also been analyzed, and I observed that C=O stretching vibration of the carbonyl group is observed at 1902 cm^{-1}, 1866 cm^{-1} and 1861 cm^{-1}, which got reduced upon complexation with triangular gold clusters.

The theoretical findings follow the similar trend as observed by experimental FT-IR spectra (Pal et al.) . I observed systematic increase in the frequency values obtained by theoretical method than those by experimental method. Since it is a general practice to scale the theoretical values by some factor, while comparison with experimental data is

required (Pliego Jr et al., 1999). However in my case I could not compare the systems fully with the experiments due to difference in size of the systems. However I can correlate certain intense peaks which corresponds to vibration due to stretching, angle bending, out of plane torsion due to specific groups. I have not observed any peak corresponding to the Au-Au vibrations, which is perhaps expected due to larger mass of Au. So on the basis of the present theoretical investigation, it can be concluded that gold atoms are not present as isolated atom or bare clusters; they always come with citrate cap, therefore vibration of the associated atoms of the citrate groups are well observed by theoretical IR spectra analysis. The vibration of the –OH stretching of the quercetin is found to decrease in the experimental FT-IR spectra, this is due to involvement of the –OH groups in forming hydrogen bonds with the citrate groups of the GNP complex. I could expect similar trend in case of theoretical approach also (Figure 4.5), where quercetin form stable hydrogen bonds with the GNP complex.

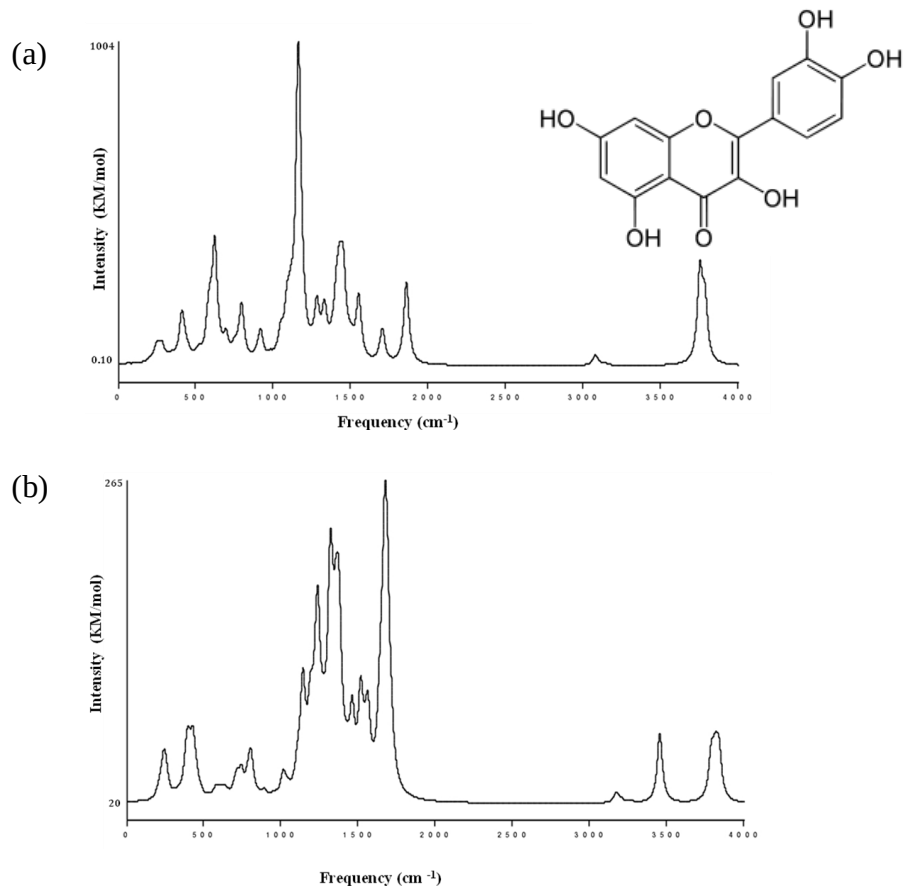


Figure 4.7. IR spectra of (a) Quercetin (Inset: structure of quercetin)
(b) GNP (Citrate capped gold nano particle)

4.4. Discussion

Present theoretical findings suggest stable formulation of gold nanoparticle-quercetin complex, which is confirmed by experimental techniques carried out by my collaborators. The triangular nanocluster considered for the theoretical analysis is considered as the best model for theoretical calculations (Mohan et al., 2006). There are

several reports on computational studies of the gold nano particle with nucleic acids like DNA, RNA (Sharma et al., 2007, Mohan et al., 2006), where they have neglected the inclusion of capping in the gold complexes. However, from the present theoretical investigation, using very systematic approaches, I can propose that capping is necessary for stabilizing the gold nano particles, which in turn favor its interaction with other molecules, such as ligands, drugs etc. The theoretical IR frequency of gold citrate complex does not show any characteristic peak corresponding to the Au-Au vibrations, which gives us a clear idea of gold atoms not being in the free state any longer, rather being capped with citrate. Recent study by Ray and groups (Lu et al., 2010a) proved the importance of capping in nanoparticle based drug formulation. Taking all these into accounts, we can presume that the GNP (triangular nano cluster with the citrate capping) complex could be considered as the best theoretical model, and can be used to study its interaction with other molecules, such as drugs, ligands etc. Quercetin is tagged with citrate capped gold nanoparticle through hydrogen bonding. Size and other associated properties of GNPQ are compatible with penetration through the cell membranes, and will be worth trying to study its therapeutic efficiency in biological systems.

Chapter 5

Wetting Property of the Edges of Monoatomic Step on Graphite: ab initio Quantum Chemical Studies

5.1. Introduction

Rapid development of nanotechnology opens up many potential applications in nano electronics, composite materials, energy search and biomedicine. Graphene represents a new class of novel nano materials, which are one atom thick, made of carbon atoms in sp^2 hybridization, and arranged in a hexagonal lattice form. These graphene sheets can be transformed into various dimensional carbon materials by self-assembly which depends on their growth conditions. The 2D graphene sheets can be periodically stacked to form 3D graphite, can be rolled to form 1D nanotubes and can be wrapped into 0D fullerenes (buckyballs), nanocones etc. Graphene with its unique physical and chemical properties has attracted tremendous attention in different areas. Graphene possesses unique electronic properties, can act as field effect transistors (FET) (Novoselov et al., 2005). It is an excellent conductor, while its conducting property can be changed under certain conditions (Banerjee et al., 2005, Banerjee et al., 2006). It has been observed that single layer graphene is found to be more reactive than the higher layers, with the edges of the graphene showing more reactivity than its surface. Graphene possesses two distinct edges, arm-chair (Cis edge) and zig-zag edge (Trans edge), as evident from the schematic diagram shown in Figure 5.1, if one of the edges is arm-chair, (A), then the other 30^o bend cut will be zig-zag (Z) or vice versa (Kobayashi et al., 2005, Banerjee et al., 2005, Banerjee et al., 2006). A partial zig-zag- arm-chair (Z,A) edge can be obtained if the cut bends to 60^o angle (Kobayashi et al., 2006). Intuitively one may expect similar electron accumulation at both the edges. However, it was reported earlier that there exist distinct electronic properties at these edges (Fujita et al., 1996a, Nakada et

al., 1996, Wakabayashi et al., 1999, Stein and Brown, 1987, Hosoya et al., 1993, Tanaka et al., 1987, Fujita et al., 1996b, Ryu and Hatsugai, 2004, Wakabayashi, 2001, Elstner et al., 1998, Kobayashi et al., 2005, Dedkov et al., 2001, Niimi et al., 2005, Geim and Novoselov, 2007). It has been observed that specific electronic states are localized only at the zig-zag edge of the graphene, which is not found in the armchair edge. It was noted that the zig-zag edge of graphene layer, which can be part of the top most layer on the pyrolytic graphite, accumulates more electron density as compared to the arm-chair edge (Sarkar et al., 2010, Banerjee et al., 2005, Banerjee et al., 2006). Previous theoretical calculations based on density functional theory also demonstrated zig-zag edge to be more electron rich (Banerjee and Bhattacharyya, 2008) as compared to that of the arm-chair edge. However, differential wetting properties associated with different edges of graphene have never been studied before.

Since understanding the wetting property of solid surfaces has been of fundamental interest for long time from the point of basic science and numerous technological applications (Herminghaus et al., 2008, Bonn et al., 2009, Sarkar et al., 2010), in this present chapter, I have tried to explain the wetting property of zig-zag (trans) and arm-chair (cis) edges of the graphene sheet,

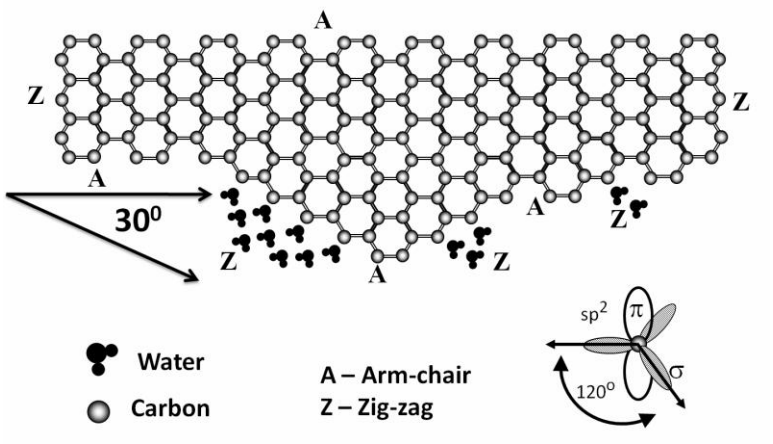


Figure 5.1. Schematic diagram of an arbitrary cut graphene

Macromolecular recognitions in cellular environment are mostly governed by hydrophobic and hydrogen bonding (hydrophilic) interactions. These are the major driving forces which stabilize native structures of proteins, nucleic acids as well as other biomolecules (Jiang and Lai, 2002, Scheiner et al., 2001). The origin of the hydrophobic interaction is known to arise from entropy gained by the release of bound water molecules. Additional contribution to this stability often comes from van der Waals interaction, which is non-directional in nature. The hydrogen bond, on the other hand, is sensitive to geometry and type of atoms involved in such interaction. Traditionally polar groups were known to form hydrogen bond, where an electronegative atom acts as acceptor (A) while another electronegative atom acts as the proton donor (D) with A...H-D in a linear, or near linear arrangement, where A and D may be nitrogen, oxygen or other electronegative atoms. It is often stated that the hydrogen bond interaction is mostly

electrostatic in nature. In addition, it is also described as overlapping of orbitals of donor, acceptor and hydrogen atoms and the charge transfer through the hydrogen bond. Numerous studies (Desiraju, 2011) also indicate that in some cases carbon atoms, which are otherwise known as non-polar type, can also act as the proton donor and/or proton acceptor (Desiraju and Steiner, 2001, Cho et al., 2006). It has been shown that such non-polar hydrogen bonds are often important in DNA and RNA structures (Ghosh and Bansal, 1999, Roy et al., 2008) and protein-protein interface (Jiang and Lai, 2002, Scheiner et al., 2001). Strength and stability of the non-polar hydrogen bonds also depend on the acidity of the C-H group involved, but its direct quantification is not available in the literature. Hydrogen bond donor capacity of carbon atoms in C-H group depends on the state of hybridization of bonding orbitals on carbon. An increase in the s-character of the bonding orbital makes the carbon atom more acidic with accumulation of more negative charge on it. Such carbon atoms behave as better proton donors than those having less acidity and lower negative charges.

As discussed earlier (Banerjee and Bhattacharyya, 2008) carbon atoms at different edges of graphene molecule accumulate different amount of electronic charge, which probably indicates differential hydrogen bonding capacity of these associated C-H bonds. Water molecule, often involved in hydrogen bond interaction, can act both as donor and acceptor. In this case I have chosen oxygen atom of water molecule as an acceptor species to test the strength of hydrogen bonds exerted by C-H groups from different graphene edges. I have carried out detailed quantum chemical calculations to characterize hydrogen-bonding interaction between water and the different edges of graphene.

Supporting experimental studies carried out by my collaborators using frictional force microscopy on a pyrolytic graphite surface also indicate differential interaction between water and the two edges of graphene, showing the possibility of the existence of weak attractive C-H...O interactions for some of the C-H groups. Interaction between graphene, as an extension of benzene with water using quantum chemical methods has been reported recently (Rubeš et al ., 2009), however, they considered very small systems, inadequate to provide terminal properties. The real graphene structures are sometimes characterized by the existence of the dangling electrons, which are highly reactive (Koskinen et al., 2008). Hence I assume that the model systems taken for calculations in this study mimic at least approximately the real stable structures of graphene at stationary state. Recently it was experimentally shown (Banerjee et al., 2005, Banerjee et al., 2006) that these edges of monoatomic step on graphite have two distinct electronic states corresponding to two different edge structures (zig-zag and arm-chair). Later, theoretically I have also shown that similar electronic property is observed in nano graphene (Banerjee and Bhattacharyya, 2008).

My experimental collaborators observed presence of two different frictional properties along the edges of a graphene sheet. They observed that with increase in the relative humidity, the water molecules condense at one particular edge (Panigrahi et al., 2011a) of the graphene. To complement the experimental findings, I have carried out quantum chemical analysis on model graphene-water complexes to explain the differential wetting properties of the edges.

5.2. Theories and Computation

A representative nano-graphene associated with both the arm-chair and zig-zag edges has been modeled by MOLDEN (Schaftenaar and Noordik, 2000) software maintaining the standard C-C bond lengths as 1.421 Å. The edges were terminated with hydrogen atoms so as to neutralize the valencies of the carbon atoms. Although there are several reports of theoretical studies of graphene without such capping, I feel the molecules with dangling electrons would behave like radicals and soon be reduced by ambient proton (Koskinen et al., 2008). The C-H bond lengths are maintained at 1.009 Å and all the angles are fixed to 120^{o}. Three graphene-water complex systems have been modeled maintaining the above geometry criteria and a single water molecule was placed near the graphene (i) at the arm-chair edge (ii) at the zig-zag edge and (iii) at the top of the graphene. In the initial models of the arm-chair and zig-zag edge systems, the oxygen of the water molecule points towards one of the hydrogen of the terminal C-H bond of graphene edge. In the top-graphene system one of the hydrogen atoms of water molecule was oriented towards the center of a hexagonal ring of graphene. All the three systems have been geometry optimized with GAMESS-US (Schmidt et al., 1993) employing Dunning Correlation Consistent basis set (Dunning Jr, 1989, Kendall et al., 1992, Peterson et al., 1993) (HF/cc-pVDZ) constraining covalent bonds of graphene by IFREEZ option.

I observed two hydrogen bond like interactions between C-H of graphene and oxygen atom of water molecule in both arm-chair and zig-zag edges after optimization

(Figure 5.2). In both the cases, the water molecule remains in the plane of the graphene, the optimized geometries have the C-H…O hydrogen bonding distances (H…O) of around 2.4 Å and angles subtended by donor-H-acceptor <D-H-A> are also close to 180^{o}. In the case of arm-chair-graphene, one of those two H-bond distances is comparatively larger while in case of zig-zag-graphene, I found two C-H…O types of hydrogen bonds of similar magnitudes (Table 5.1). From these data one can presume that graphene can form weak hydrogen bond interactions with water oxygen atoms through C-H…O contacts (Cho et al., 2006, Rubeš et al ., 2009). Apart from the arm-chair- and zig-zag-graphene systems, top-graphene complex was also optimized. The initial model had the water molecule placed at top of graphene, about 2.88 Å angstroms vertically above the center of the graphene. One of the two H-atoms of water faced towards the center of the hexagonal ring. The oxygen atom was placed in such an orientation that the ring-center…H–O angle is close to 180^{o}, suitable for an O-H…π type of hydrogen bonding (Grabowski, 2007). After optimization the free water molecule moves from the center of graphene towards the zig-zag edge but remain vertically above the graphene plane, forming slightly elongated H-bond. Here C…H–O type of interaction is observed with H...C distance of about 2.6 Å while O-H...C angle is smaller than 150^{o} (Table 5.1). Since the H…C distance is close to the corresponding sum of van der Waals radii one can expect that the O-H…C interaction could not be classified as hydrogen bonding but rather as van der Waals type. Considering recent understanding (Jain et al., 2009) that lone pair of the oxygen can also interact attractively with π-electron clouds, I have also optimized a graphene-water complex from an initial geometry where one of lone-pair

electrons of the water oxygen was facing towards the central hexagonal ring of the graphene. The optimized geometry shows feature similar to the other top-graphene system – the water molecule moves in such a way that the hydrogen atoms come closest to carbons near the zig-zag edge. These two observations indicate that a free water molecule prefers to bind to graphene through its zig-zag edge as compared to all other edges. Moreover the hydrophobic surface of graphene cannot trap any water molecule to form a hydrated layer (Yang et al., 2007, Biswas and Drzal, 2008, Leenaerts et al., 2009).

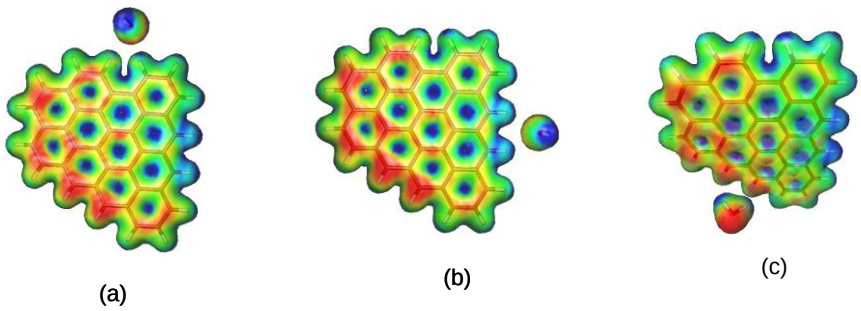


Figure 5.2. Structures of the optimized graphene water complexes in three orientations for water interacting through (a) arm-chair edge (b) zig-zag edge (c) top, along with electrostatic potential. The electrostatic potential values are color coded with red indicating less than -0.10 H, yellow indicating –0.05 H, green indicating 0.00, light blue indicating +0.05 H and dark blue indicating greater than 0.10 H.

Structures of the optimized complexes are shown in Figure 5.2. along with their electrostatic potential. These electrostatic potentials were calculated by MOLDEN using multipole derived method from the orbital electron density of the optimized structures. Contour value of 0.02 is used for these calculations. These figures indicate that most of the C-H bonds near arm-chair edge are neutral (green in color) while the C-H bonds at the zig-zag edge have large dipole moment. The near triangular graphene considered in

the theoretical study (Figure 5.2) has arm-chair edge at one side and zig-zag edges at the two other sides with arm-chair edges also at the corners. The nature of potential of the two zig-zag edges are both polar but different in nature – (i) in one edge the hydrogen atoms are more electropositive (blue) and (ii) in the other edge the carbons are more electronegative (red), both leading to large dipole moments. These electrostatic potentials qualitatively indicate that the zig-zag edge have stronger dipoles, as compared to the arm-chair edges (neutral, green), and may interact strongly with water.

Table 5.1. Hydrogen bond length, angle and interaction energy of the optimized structures obtained through different method

System	Optimization Method	H-Bond	Bond Length (H...A) in Å	Bond Angle <D-H...A>	Energy[a] (kcal/mol)	Energy[b] (kcal/mol)
Arm-chair	HF/cc-pVDZ	Cs1-Hs1…O Cs4-Hs4…O	2.62 2.40	179.2° 169.2°	-2.41	-2.59
Zig-zag	HF/cc-pVDZ	Ct2-Ht2…O Ct4-Ht4…O	2.45 2.45	150.7° 150.2°	-4.47	-3.33
Top-edge	HF/cc-pVDZ	CT2…H1-O CT4…H2-O	2.67 2.88	144.3° 124.0°	--------	--------

[a] Interaction energy calculated by HF/cc-pVDZ method and

[b] Interaction energy calculated by MP2/6-31G** method

Atoms in molecules (AIM) approach is evolving as a very important theoretical technique for characterization of hydrogen bonds (Bader, 1991), which is very elaborately discussed in chapter 2. I have carried out the AIM calculations for all three

types of graphene systems discussed above using AIM2000 program (Biegler-König, 1990).

I have calculated the ρ_c, $\nabla^2\rho_c$, H_c, G_c, V_c for all interactions, mainly hydrogen bonds, observed in the arm-chair, zig-zag and top-graphene systems (Table 5.2). One can see that the values of $\nabla^2\rho_c$ and H_c are positive in all types of graphene systems signifying formation of weak hydrogen bonds. Since there exist a correlation between hydrogen bond length and $\nabla^2\rho_c$, I observe two C-H…O interactions for arm-chair-graphene, both are weak but they differ in the strength (Table 5.1). The $\nabla^2\rho_c$ values also show the similar trend. However in case of zig-zag-graphene system I notice two C-H…O hydrogen bonds of similar strength as suggested by similar values of Laplacian and charge density. In case of the top-graphene system I observe very small values of $\nabla^2\rho_c$ at the BCP, which signifies very weak and almost negligible interaction between the water and graphene, as it was mentioned before it may be classified as weak van der Waals interaction. This is also in agreement with previous experiments (Yang et al., 2007, Leenaerts et al., 2009). So I have not considered the top-graphene system in further analysis.

Table 5.2. Topological parameters of BCPs (in a.u), corresponding to C-H…O distance for all the three types of graphene system

System		ρ_c	$\nabla^2\rho_c$	G_C	V_C	H_C
Arm-chair	Cs1-Hs1…O	0.007	0.024	0.006	-0.005	0.0003
	Cs4-Hs4…O	0.010	0.036	0.008	-0.008	0.0003
Zig-zag	Ct2-Ht2…O	0.009	0.032	0.008	-0.007	0.0004
	Ct4-Ht4…O	0.010	0.032	0.008	-0.007	0.0003
Top	O-H....C	0.006	0.020	0.004	-0.003	0.0007
	O-H.....C	0.006	0.020	0.005	-0.004	0.0005

Interaction energy and basis set superposition error were calculated from the optimized structures with HF/cc-pVDZ level and also with MP2/631G** (Møller and Plesset, 1934), a higher level of approximation method. Both these methods include electron correlation and electron exchange functions in their Hamiltonian, justify pursuing these methods to study orbital overlapping and electron exchange due to the hydrogen bond formation in water-graphene complexes. I have employed Morokuma method (Kitaura and Morokuma, 1976) of GAMESS-US and Boys-Bernardi function counterpoise method (Boys and Bernardi, 1970) of Gaussian03 (Frisch et al., 2003) for the calculation of BSSE, which has been discussed in chapter 2. The BSSE corrected interaction energy of the systems is calculated as follows

$$E_{int} = E(\text{graphene+water})_{opt} - E(\text{graphne alone})_{opt} - E\ (\text{water alone})_{opt} + \text{BSSE} \quad \ldots[5.1]$$

The interaction energies are calculated for the geometry-optimized arm-chair- and zig-zag-graphene water–complexes (Table 5.1). The interaction energy obtained from

HF/cc-pVDZ in zig-zag-graphene system indicates it to be stronger by around 2 kcal/mol as compared to that of the arm-chair-graphene. Similar trend is found for the interaction energy calculated by more accurate MP2/6-31G** method. The general conclusion is thus, water prefers to bind to the zig-zag edge more strongly as compared to the arm-chair edge of graphene but the energy differs by few k_BT at room temperature (Table 5.1).

Charge transfer due to hydrogen bond formation for the optimized structures of the arm-chair- and zig-zag- systems were also analyzed by natural bond orbital (NBO) (Carpenter and Weinhold, 1988, Reed et al., 1985, Reed et al., 1988) approach using Gaussian03, which is discussed in chapter 2. Electronic properties, such as occupancy of the natural orbitals, stabilization energy, natural charges for the relevant atoms and bonds of graphene, which are closest to the water molecule, are reported in Table 5.3.

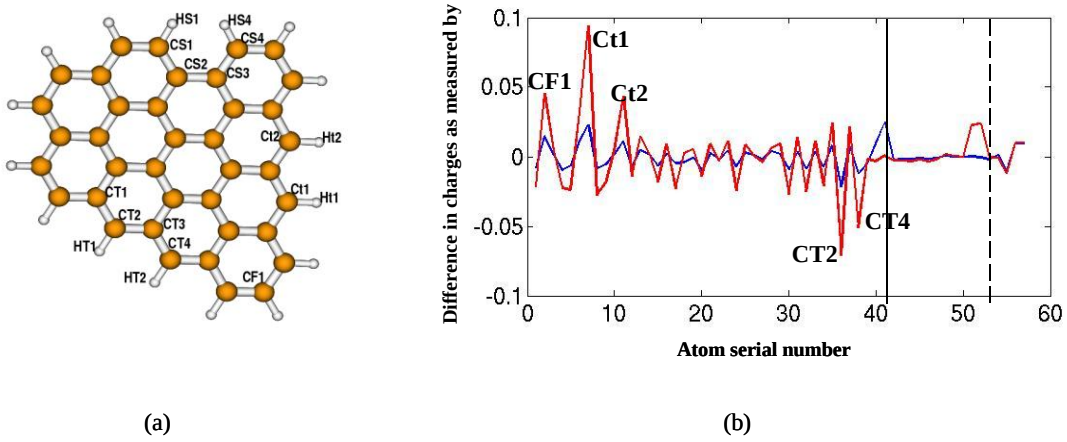
It indicates that charge transfer takes place, in both arm-chair- and zig-zag-graphene, from the lone pair of oxygen to the anti-bonding orbital of the closest C-H groups of the graphene ring. The significant change in natural charges of the carbon, hydrogen and oxygen of the complex and components in zig-zag-graphene, particularly for O, Ct1 and Ct2 atoms (Figure 5.3a) implies that zig-zag-graphene has more polarity than that of the arm-chair-graphene. I have also compared alterations in the NBO charges of all the atoms of graphene and water upon complex formation (Figure 5.3b). This clearly indicates that larger charge modifications take place for few carbon atoms when water comes close to the zig-zag edge.

Table 5.3. Results from Natural bond orbital analysis characterizing hydrogen bond formation between water oxygen atom and C-H moiety of graphene molecule in different orientations

System	Donor NBO (LP/BD)	Occupancy	Acceptor NBO C-H (BD*)	Occupancy	NσE(N-σ*) (kcal/mol)	qO	qH	qC
Arm-chair	O(LP1) O(LP2)	1.996 (1.997) 1.989 (1.996)	Cs1-Hs1 Cs4-Hs4	1.9793 (1.9799) 1.9788 (1.9800)	LP1…Cs4-Hs4:0.56 LP2…Cs1-Hs1:1.46 LP2…Cs4-Hs4:2.72	-0.918 (-0.908) -0.918 (-0.908)	0.225 (0.216) 0.224 (0.199)	-0.165 (-0.164) -0.286 (-0.290)
Zig-zag	O(LP1) O(LP2)	1.995 (1.997) 1.991 (1.996)	Ct1-Ht1 Ct2-Ht2	1.9827 (1.9831) 1.9813 (1.9820)	LP1…Ct1-Ht1:0.38 LP1…Ct2-Ht2:0.87 LP2…Ct1-Ht1:2.04 LP2…Ct2-Ht2:1.59	-0.921 (-0.909) -0.921 (-0.909)	0.240 (0.217) 0.234 (0.211)	-0.153 (-0.247) -0.299 (-0.342)

It may also be noted that large charge-transfer takes place for some carbon atoms situated very far from the water molecule, particularly in zig-zag edge. Largest differences are seen for Ct1 and Ct2 (they are closest to the water molecule) and also for CF1, CT2 and CT4, which are far away from the binding site. The large change in NBO charges for CT2 and CT4, in zig-zag edge complex, is presumably due to electron delocalization within graphene, the transferred charge is pushed away. Similar trend, but of very small magnitude, of alteration of NBO charges on these sites are also observed in arm-chair edge complex. This presumably indicates that arm-chair edges are more

hydrophobic and the water molecule close to the arm-chair edge does not perturb the graphene molecule. The zig-zag edge, on the other hand, is polar and attracts water molecules strongly. Furthermore binding of a water molecule at the zig-zag edge enhances further water binding at the zig-zag edges.



(a) (b)

Figure 5.3. NBO charge donors and acceptors at the arm-chair and zig-zag edges involved in C-H…O interaction with the water molecules are shown in (a) and (b) denotes the change in Natural charges in the atoms of graphene and water before and after the complex formation in arm-chair edge (blue lines) and zig-zag edge (red lines) orientations (As obtained from NBO analysis), the black vertical line distinguishes the carbon and hydrogen atoms, whereas the black vertical broken line distinguishes the hydrogen and the water molecules

So from this analysis it is confirmed that electron delocalization in the graphene system is more significant due to the presence of the water molecule at the zig-zag edge than that of the arm-chair edge. This in turn proves that water molecule is bound to the zig-zag edge more strongly than that of the arm-chair-edge.

5.3. Discussion

I observed two distinct electronic properties at the two edges of the graphene, upon interacting with water molecule. My present theoretical calculations enabled me to find out exact mechanism of interaction of water with the two different types of edges of the graphene. The interaction energies and NBO charge transfer analysis suggest that the zig-zag edge of graphene has stronger binding capacity with water molecule. The NBO analysis also clearly reveals that water binding enhances further water-binding capacity to the other zig-zag edges of the nano-graphene sheet through a cooperative mode. I have also observed that water does not remain on graphite surface and hence exhibits hydrophobic nature of graphene/graphite surface. My theoretically analysis well correlates the experimental findings carried out by my collaborators.

Chapter 6

Quantum Chemical Studies of Interaction of Nucleobases and Nucleosides with Graphene and Carbon Nanotube

Section 6.1

Binding of Nucleobases with Wrinkled Graphene Surface: Analyzed through Dispersion Corrected DFT Approach

6.1.1. Introduction

Graphene, one of the allotropes of carbon, has inspired many theoretical and experimental applications due to its unique chemical and electrical properties (Rosas et al., 2011, Novoselov et al., 2006). it is a mono-layered sp^2-bonded carbon sheet, possesses excellent thermal conductivity and mechanical strengths (Balandin et al., 2008, Lee et al., 2008). Similar to the surface of the graphite, graphene can adsorb various small molecules or atoms on its surface, thus can be used in biosensors (Yang et al., 2010b, Shao et al., 2010) and chemical sensors (Fowler et al., 2009). It can also be used as transparent conductors (Wassei and Kaner, 2010), and has many biological applications (Sun et al., 2008). A numbers of reports have shown the potential of graphene for different biomedical applications (Yang et al., 2010a, Feng and Liu, 2011). Along with its unique physical, chemical, and electronic properties, its interesting shape and size make it a promising nano material in many biological sensing. Graphene surface with delocalized π electrons can be utilized for effective loading of aromatic anticancer drugs. Graphene can also be used as a drug delivery carrier (Liu et al., 2008, Liu et al., 2011). Recently, *in vivo* cancer treatment with graphene has been realized in animal experiments with graphene sheet used as a vehicle for carrying drugs (Yang et al., 2010a). As nano graphene has limited hydrophobicity, the edges being slightly polar and the faces are highly hydrophobic (Yang et al., 2007, Panigrahi et al., 2011a, Banerjee et al., 2005, Banerjee and Bhattacharyya, 2008), it has the capability to pass through biological membranes. Graphene-nucleobase complexes are even more amphiphilic as the nucleobases can favorably interact with the lipid head-groups while the graphene components would tend to be stabilized towards the central hydrophobic part of the

bilayer membrane. These interactions together can permit the complex to pass through the cell membrane. The above characteristics allow graphene to act as a drug-delivery vehicle, which has been discussed by various groups (Sun et al., 2008, Liu et al., 2008, Yang et al., 2010a).

Among the several interactions of nature, non-covalent interactions have crucial importance in chemical, catalytic, in biological systems and nanoengineering. The main contributions to the non-covalent interactions come from electrostatics, hydrogen bonding, stacking, and van der Waals interactions. Among these contributions, dispersion interaction, a component of van der Waals interaction, acts as a major attractive interaction between nonpolar molecules. Dispersion interaction is an attractive interaction arises due to electrostatic interactions of the fluctuating distributions, instantaneous dipoles and higher multipoles. Interactions of graphene with different host molecules such as many organic small molecules, nucleobases, drugs etc., are also guided by non-covalent interactions. However, dispersion interaction including π-π stacking plays crucial role in stabilizing such kind of systems. It has been observed that dispersion interaction between stacked amino acids and stacked DNA or RNA basepairs can be sufficiently large and sometimes comparable to the strengths of the hydrogen bonding interactions also (Černý and Hobza, 2007). Many DNA sensors have been developed taking in advantage of the π-π stacking of single strand DNA on graphene (Brett and Chiorcea, 2003, Akca et al., 2011). Different aromatic drugs, which are used in cancer therapy, can be loaded on the graphene sheet through π-π stacking for intracellular drug delivery. Recently, additional studies in drug loading and delivery via graphene have been reported by several groups (Yang et al., 2008, Yang et al., 2010c). A very important

characteristic of graphene sheet is its high surface area, where maximum number of atoms can get exposed to the surface, thereby making it a very important therapeutic agent for drug delivery (Geim and Novoselov, 2007). Many studies by different groups have been carried out to analyze the interactions between the graphene sheet with different biomolecules like amino acids, nucleobases, and different organic compounds (Rajesh et al., 2009, Varghese et al., 2009, Antony and Grimme, 2008, Gowtham et al., 2007, Umadevi and Sastry, 2011). However, different groups adopt different assumptions to optimize these structures. Gowtham *et. al* (Gowtham et al., 2007) used density functional theory (DFT) formalism with plane-wave pseudopotential considering periodic lattice of graphene-nucleobase complexes. Varghese *et. al* (Varghese et al., 2009) optimized structures of graphene–nucleobase complexes using Hartree-Fock method. Sastry and coworkers (Umadevi and Sastry, 2011) recently studied interaction of graphene and nanotubes of different radius with the nucleobases using dispersion-corrected DFT based approach and observed that the binding energy depends on curvature of the molecule. However, they did not study any effect of non-planarity of graphene in their hybrid quantum mechanical-molecular mechanics (QM/MM) approach.

All the above studies assume perfectly planar graphene sheets and mostly of infinite dimensions in X and Y directions. However, it is well known that two-dimensional crystals are never stable (Hohenberg, 1967, Mermin and Wagner, 1966) and have a tendency to curl towards the edges (Pereira et al., 2010, Gil et al., 2010). Also many reports by different groups confirmed that bending in the graphene can be due to topological lattice defects, impurities, tensions etc. (Banerjee and Bhattacharyya, 2008, Thompson-Flagg et al., 2009, Guinea et al., 2009). Recent electron microscopy and

atomic force microscopy studies also show evidences of such curvature at the edges, but the wavelengths measured by different groups vary significantly (Meyer et al., 2007, Bao et al., 2009).

I have carried out *ab initio* quantum chemical calculations using density functional theory with different DFT functionals on free hydrogen terminated nano graphene as well as graphene-nucleobases complexes to understand the structure and interactions of graphene. Supportive experiments have been also carried out by my collaborators to understand its natural curvature (Panigrahi et al., 2012). All of these indicate that significantly larger graphene sheet is curved in general and these curvatures may give additional stability to nucleobase binding.

6.1.2. Modeling and Computation

I have considered square graphene sheets of different dimensions, by varying the number of benzene rings such as 4-ring (4×4), 5-ring (5×5), 6-ring (6×6), 7-ring (7×7) and 8-ring (8×8) for theoretical quantum chemical calculations. All the graphene sheets considered here are terminated with hydrogen atoms in their cis (arm-chair) and trans (zig-zag) edges. The molecular models were constructed with the help of MOLDEN software (Schaftenaar and Noordik, 2000) using standard C-C and C-H bond lengths as 1.421 Å and 1.009 Å, respectively and all bond angles were fixed to 120^{o}. I have optimized all the graphene sheets of different dimensions by the standard B3LYP/6-31G** and dispersion corrected density functional theory (DFT-D) based approach with ωB97XD/6-31G** (Chai and Head-Gordon, 2008) basis set using Gaussian 09 (Frisch et

al., 2009). Considering their overall properties and to correlate theoretical data with the experimental findings, I have considered (8×8) graphene sheet as a standard miniature model for carrying out further calculations.

I have modeled the constituent bases of DNA and RNA, namely adenine, guanine, cytosine, thymine and uracil using their standard bond lengths and angles. The N9 position of the purines (adenine and guanine) and the N1 position of pyrimidines (cytosine, thymine, and uracil) have been terminated with methyl groups instead of hydrogen. This removes the artificial dipole moment due to the N-H bonds and also gives some steric effect, though smaller than the sugar moiety of nucleic acids. Initially the bases were kept parallel to the 8-ring graphene sheet at around 4.0 Å vertically above the center of the hexagonal carbon rings of the planar graphene sheet (modeling is done with the help of Discovery Studio software (Studio, 2007)). During the optimization process, all the atoms of the graphene sheet and nucleobases were allowed to relax freely without any constrain. Equation of plane was fitted to the coordinates of the carbon atoms of the optimized graphene by least squares fit method and root mean square deviation (rmsd) of the carbon atomic positions from the best-fit plane was calculated.

Interaction energy between graphene and the nucleobases were calculated considering Basis Set Superposition Error (BSSE) and deformation correction using the following equation.

$$E_{int} = E(complex) - E_{Xo}(isolated\ graphene) - E_{Yo}(isolated\ nucleobase) + BSSE \quad \ldots. [6.1.1]$$

Boys-Bernardi function counterpoise method (Boys and Bernardi, 1970) has been employed for calculation of the BSSE contribution. I have also analyzed the pyramidalization of the amino groups of the nucleobases following earlier publications

(Mukherjee et al., 2005). As electronic structure of the graphene sheet is highly sensitive to the environment, I have carried out Natural Bond Orbital (NBO) (Carpenter and Weinhold, 1988, Reed et al., 1985, Reed et al., 1988) analysis to evaluate the charge transfer and charge distribution of the bases and the graphene sheet before and after complex formation, which in turn affects it's conducting properties also.

6.1.3. Results

6.1.3.1. Structure of Nano Graphene

Optimized structures of most of the graphene models of various dimensions show near planar geometry. The root mean square deviations (rmsd) of the carbon atomic positions from the best-fit planes were calculated to estimate planarity of the optimized models and are tabulated for the B3LYP/6-31G** and ωB97XD/6-31G** methods in Table 6.1.1. The smaller models show similar and small rmsd values by both standard DFT and DFT-D quantum chemical methods. The model having 8-rings optimized by DFT-D method shows a significant curvature as reflected by the high rmsd value (Figure 6.1.1). The central region of the graphene sheet remains invariant, whereas the corners tend to role, which is in agreement with the study by Meyer *et. al*. (Meyer et al., 2007). The difference in DFT and DFT-D methods arises from inclusion of long-range dispersion interaction in the DFT-D functional, which plays significant role in making a larger graphene sheet non-planar. I have also optimized the curved model with 8-rings, as obtained from DFT-D optimization, by the standard B3LYP/6-31G** method and observed that absence of dispersion correction makes the structure planar (rmsd value drops from 3.7 Å to 0.06 Å as denoted by DFT-D_B3LYP in Table 6.1.1). This clearly

indicates that the curvature in the 8-ring graphene (2.4 nm × 2.4 nm dimensions) by DFT-D is not an artifact and it arises due to dispersion interaction.

Table 6.1.1. RMSD values for different dimensions of graphene sheets obtained from plane fit method

System	B3LYP_RMSD	ωB97XD_RMSD
4-ring	0.0019 Å	0.0052 Å
5-ring	0.0023 Å	0.0049 Å
6-ring	0.0053 Å	0.0076 Å
7-ring	0.0062 Å	0.0027 Å
8-ring	0.04 Å	3.70 Å
8-ring (DFT-D_B3LYP)	0.06 Å	

On the basis of these results, I can hypothesize that larger graphene sheets would have stronger dispersion interactions and would be more curved. It has been discussed by Bao *et. al*. (Bao et al., 2009) that the free standing graphene sheet may have much smaller wavelength ripples and as the thickness of graphene sheet increases (few layers of graphene bound together), the wavelength will increase. As dimension of the models considered for my theoretical calculations are an order of magnitude smaller than the measured wavelengths, it is difficult to estimate the degree of bending to compare accurately with experiments. My experimental collaborators also observed distinct electronic properties of the carbons at the bends (Panigrahi et al., 2012). Using the conducting tip atomic force microscopy technique around the folded region of the graphene underlying on the highly oriented pyrolytic graphite surface, they proposed that the appearance of curvature is presumably due to adoption of partial sp^3 character by some of the carbon atoms present towards the edges, owing to unusual edge property.

6.1.3.2. Interaction of Graphene with Nucleobases

Each of the complex systems (graphene sheet with nucleobase), upon optimization by ωB97XD/6-31G** method results in stacked geometry with the (8×8) dimension graphene sheet as shown in Figure 6.1.2. The bases remain parallel to the graphene sheet in AB stacking configuration, with distances around 3.5 Å vertically above the central hexagonal carbon rings, which are the characteristics of π-π stacking interaction.

Figure 6.1.1. Isolated graphene optimized by (a) DFT-D formalism with ωB97XD/6-31G** (b) B3LYP/6-31G** functional

However, guanine moved slightly from the center towards the trans edge unlike the other bases. An analysis of dipole moment of individual bases reveals that guanine has maximum dipole moment (6.96 D) as compared to the other bases (2.48, 6.13, 4.44 and 4.65 D for adenine, cytosine, thymine and uracil, respectively). As trans edge is more polar than the cis edge, (Panigrahi et al., 2011a, Banerjee and Bhattacharyya, 2008) guanine shows a tendency to move towards it. The associated graphene sheets of the graphene-nucleobase complexes also show similar amount of wrinkles as observed in case of free graphene sheet (Figure 6.1.1).

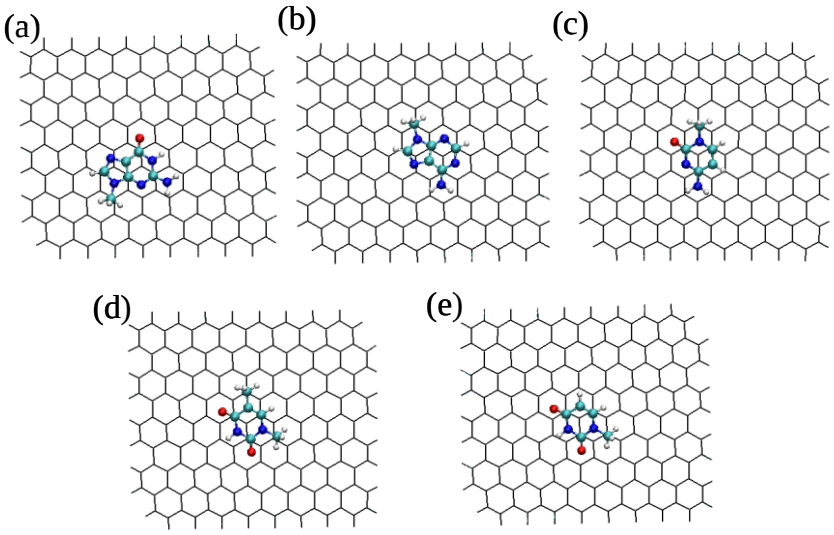


Figure 6.1.2. Optimized geometry of stacked complexes of graphene with (a) Guanine (b) Adenine (c) Cytosine (d) Thymine (e) Uracil

The rmsd values of the curved graphene sheets in the complexes lie in between 3.5 Å to 4.5 Å (Figure 6.1.3), indicating nucleobase binding does not further enhance the non-planarity of the graphene sheets. The values of interaction energy, as given in Table 6.1.2, indicate its trend follows the order of G > A > C > T > U. These results are well correlated with the theoretical and experimental observations by various groups (Gowtham et al., 2007, Varghese et al., 2009, Umadevi and Sastry, 2011). Guanine has more stabilization energy presumably due to presence of both the amino ($-NH_2$) and carbonyl ($=O_6$) groups, which can interact with the π electron cloud of the graphene sheet. Such groups are known to interact with π-center to form a weak hydrogen bond like interactions in two ways: N-H…π as well as lone-pair electron…π types of interactions (Jeffrey, 1997). The cytosine base also possesses amino and carbonyl groups, but its interaction energy is significantly smaller than that of guanine. This difference possibly arises due to stacking component that depends on the overlapping area and

number of atoms involved. It may be mentioned here that the guanine, adenine, and cytosine bases have amino groups, which can form N-H…π hydrogen bond while the uracil and thymine bases possess only carbonyl oxygen atoms, capable to form lone-pair…π interactions. Among these, the N-H...π is perhaps stronger than the lone-pair…π interaction, which is reflected in the above interaction energy pattern. The strong interaction of N-H…π over the weak C=O… π interaction in different aromatic compounds have been studied both experimentally and theoretically by various groups (Ottiger et al., 2009, Egli and Sarkhel, 2007, Jain et al., 2009). Thus, in addition to hydrophobic stacking interaction, there are also some electrostatic interactions present in the complexes.

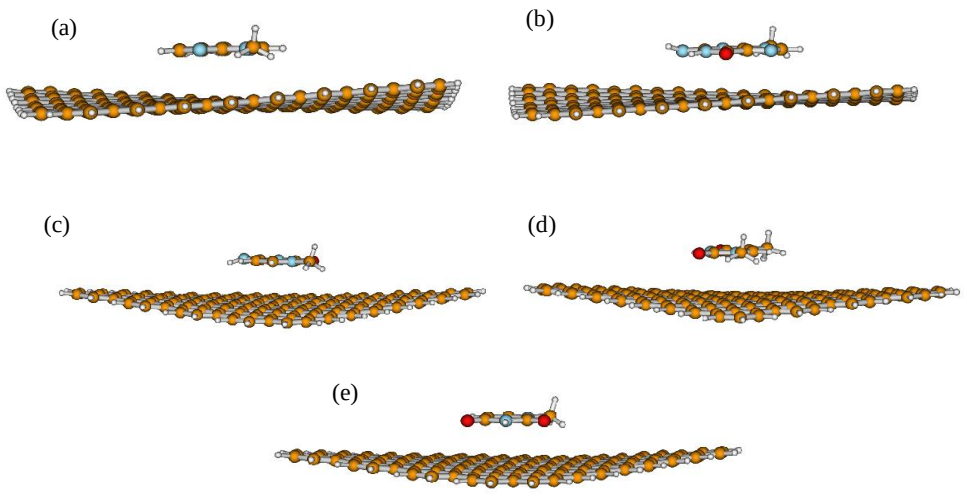


Figure 6.1.3. Optimized structure of the graphene from complex with (a) Adenine (b) Guanine (c) Cytosine (d) Thymine (e) Uracil

It has been observed by various groups that amino groups have the tendency to be pyramidal and the value of pyramidalisation changes when it is involved in any kind of

interactions, mostly when it is involved in hydrogen bonding. I observed that guanine shows maximum pyramidalization (37.77^{o}) in the isolated environment (Table 6.1.2), and its value decreases by 10^{o} when it forms complex with graphene sheet, while in case of adenine and cytosine, value of pyramidalization increases upon complex formation. This indicates a strong possibility of guanine interacting through N-H…π type of hydrogen bonding.

Table 6.1.2. Interaction energy and amino group geometry of the stacked graphene-nucleobases

System	Interaction Energy (kcal/mol)	Pyramidalization of amino groups in Complex	Pyramidalization of amino groups in Component
Graphene+ Guanine	-22.49	27.02 ^{o}	37.77 ^{o}
Graphene+ Adenine	-20.31	23.10 ^{o}	5.2 ^{o}
Graphene+ Cytosine	-18.94	26.21 ^{o}	17.68 ^{o}
Graphene+ Thymine	-18.40		
Graphene+ Uracil	-16.34		

As the graphene surfaces are hydrophobic in nature, partial charges on the inner carbon atoms are close to zero in isolation. Their values can change in presence of charged groups, such as amino or carbonyl groups, which can be best characterized by NBO analysis. I observed noticeable change in the natural charges of the methyl groups in the complex and in the isolated geometry. The differences in natural charges of the amino groups and carbonyl groups of the nucleobases before and after complex formation are given in Table 6.1.3, all the other atoms show negligible amount of charge transfer.

It is observed that NBO charges of HN21, HN22 and O6 of guanine change significantly after complex formation as compared to functional groups of the other bases, which is probably indicative of weak hydrogen bond formation. I have also picked the carbon atoms of the graphene sheet, those lie within 4.0 Å to the amino groups and carbonyl groups of the nucleobases, and compared their charges as average and standard deviation (Table 6.1.4).

Table 6.1.3. NBO analysis of nucleobases of the graphene-nucleobases complexes

System	Atoms	Natural charge of complex (Qc)	Natural charge of component (Qo)	Difference (Qc-Qo)
Graphene+guanine	O6	-0.617	-0.604	-0.013
	HN21	0.445	0.437	0.008
	HN22	0.426	0.419	0.007
Graphene+ adenine	HN61	0.433	0.438	-0.005
	HN62	0.437	0.435	0.002
Graphene+ cytosine	O2	-0.655	-0.644	-0.011
	HN41	0.436	0.437	-0.002
	HN42	0.424	0.421	0.002
Graphene+ thymine	O2	-0.646	-0.642	-0.004
	O4	-0.618	-0.610	-0.008
Graphene + uracil	O2	-0.639	-0.637	-0.003
	O4	-0.613	-0.602	-0.011

In case of isolated graphene, the average NBO charges are generally zero with very small variations. The average charges of the carbon atoms present close to amino hydrogen atoms (having weak hydrogen bonding distance) become negative with wide fluctuations as observed from its large standard deviation values.

Table 6.1.4. NBO charge analysis of carbon atoms of the graphene-nucleobase complexes along with closest approaching distance from the nucleobases.

System	Atoms of the Nucleobases	Distance D-H…A	No. of atoms of the graphene sheet close to the atoms indicated in column 2	Average NBO charge and Std. in parenthesis
Gua+graphene	HN21	3.22 Å	7	-0.006 (0.021)
	HN22	3.44 Å	5	0.000 (0.019)
	O6	3.30 Å	6	0.009 (0.011)
Ade+graphene	HN61	2.98 Å	9	-0.002(0.022)
	HN62	3.03 Å	9	0.000(0.017)
Cyt+graphene	HN41	3.01 Å	9	0.001 (0.017)
	HN42	2.94 Å	9	-0.008 (0.016)
	O2	3.36 Å	6	0.011 (0.015)
Thy+graphene	O2	3.37 Å	6	0.005 (0.016)
	O4	3.34 Å	6	0.006 (0.014)
Ura+graphene	O2	3.43 Å	6	0.004 (0.016)
	O4	3.32 Å	6	0.007 (0.014)
Isolated graphene	Single ring		6	0.000 (0.004)
Isolated graphene	Fused ring		10	0.003 (0.008)

Similar wide fluctuations are also observed for the NBO charges of the carbon atoms close to the carbonyl oxygen atoms. This can well explain why the graphene-guanine complex has more attractive interaction energy as compared to others.

6.1.4. Discussion

From the computational approach, I observed that ripple formation is the inherent nature of the larger graphene sheet, which provides structural stability to graphene sheet. Interaction of graphene with the nucleobases proved that electrostatic interaction is not the major driving forces for stabilizing such systems and greatest contributions for the binding energy often come from the dispersion interaction, which is isotropic in nature.

Among the nucleobases, guanine shows maximum charge transfer with the graphene sheet, associated with maximum interaction energy and pyramidalization values, which in turn demonstrate that N-H…π interaction in addition to lone pair…π contacts play crucial role in stabilizing such kind of systems. My theoretical observations also prove the reliability of dispersion corrected density functional theory (DFT-D) for optimizing the stacked complex systems. My theoretical studies were well supported by the experimental analysis carried out by my collaborators, where ripple formation in graphene is associated with unique electronic properties of the carbons at the bends.

This investigation opens up a new area of research to understand the mode of interaction between graphene with other biomolecules. One can expect that amphipathic compounds like small interfering RNA, which are generally unstructured, have their hydrophobic base faces as well as hydrophilic sugar-phosphate groups exposed to solvent environment. Such molecules may bind to graphene more strongly through π–π stacking as well as hydrogen bonding interaction involving their 2′-hydroxyl groups, which are discussed in the next section.

Section 6.2

siRNA Unzipping on Graphene : Quantum Chemical Approach

6.2.1. Introduction

Exploring of new classes of RNA, popularly known as non-coding RNA (nc-RNA) reveals several fascinating features of small RNA (Carrington and Ambros, 2003). These RNAs in general control gene expression by repressing the translation of the target genes, by binding to the 3' untranslated region of the m-RNA targets. Such small RNAs in animal, plant and fungi are collectively termed as RNA interference (RNAi). RNA interface refers to inhibition of gene expressions by small double stranded RNA molecule, typically known as siRNA (Couzin, 2002). The siRNA is emerging as widely accepted tools for down regulating gene expressions. Producing endogenous siRNA is regulated by three enzymes, (i) RNA dependent RNA polymerase which converts single stranded RNA into double RNA, (ii) DNA dependent RNA polymerase produce siRNA by transcribing inverted DNA repeats and (iii) dicer that cut the large RNA fragments into smaller siRNA. RNA induced silencing complex (RISC) is formed when argonaute protein binds to the siRNA, which promoted gene silencing. The base-pairing of siRNA with messenger RNA (mRNA) sequence silences the encoded protein. The mechanism of RNAi interference involves RISC (RNA-induced silencing complex) loading complex comprising of Dicer, Argonaute2 and siRNA binding protein that induces unzipping of siRNA into two single strand RNAs (Zamore et al., 2000, Tomari et al., 2004). One of these two strands acts as a guiding strand to specifically base-pair with mRNA. The siRNA controls the gene expression in two ways, it may bind to the messenger RNA, and inhibits the translation process, or it may direct binds to the chromatin (Plasterk, 2002) (Zamore, 2002) and modify the chromatin structures. SiRNA can be synthetic, they can be designed specifically to target specific genes, thus have important therapeutic

applications. RNAi technology can be used in the treatment of cancer, HIV-AIDS, viral infections (Li et al., 2002) . Efforts are being made for the efficient and safe siRNA delivery systems to achieve the desired RNAi effect. Dendrimers (Tsubouchi et al., 2002, Huang et al., 2003) and carbon nanotubes (Liu et al., 2007, Liu et al., 2009), cell penetrating peptides (Morris et al., 1997, Simeoni et al., 2003) etc. are good carriers of siRNA into disease infected cell. Graphene are molecular structures having exceptional electronic, thermal, and mechanical properties (Han et al., 2007b, Lee et al., 2008). These properties along with applications oriented characteristics such as high surface area, diverse capabilities of chemical modifications and functionalization, made them a promising material in wide varieties of applications. Numbers of reports have exploded the potential for graphene for different biomedicine applications (Feng and Liu, 2011, Yang et al., 2008). In recent years graphene has been envisaged to be very efficient delivery materials for siRNA as well as various oligonucelotides (Zhang et al., 2011). Different groups have attempted to study the properties of nucleic acid interaction with graphene (Varghese et al., 2009, Umadevi and Sastry, 2011, Mohanty and Berry, 2008). However the interaction between graphene and siRNA/DNA has not been studied well.

In this section, I have tried to understand the structure and thermodynamics of the siRNA-graphene complex by dispersion corrected density functional theory (DFT-D) based approach, since dispersion force including $\pi-\pi$ stacking serves as the major attractive interactions between the non-polar molecules (Brett and Chiorcea, 2003). My quantum chemical results are well supported by the molecular dynamics studied carried out by my collaborators (Santosh et al.). It was found from molecular dynamics simulation of graphene-siRNA and graphene-dsDNA complexes, for same sequence, that

the siRNA has a tendency to unzip and binding to graphene more strongly. On the other hand the dsDNA remains in stable double helical form and probably has weaker interactions with graphene.

6.2.2. Modeling and Computation

I have carried out quantum chemical analysis to understand the interaction of the planar nano graphene with the nucleosides of RNA and DNA. For the present investigation, I have taken a graphene sheet of (6×6) dimension with alternate arm chair edge and zig-zag edges with C-C bond lengths as 1.42 Å, C-H length as 1.09 Å while all the angles are kept at 120^{o}. The edges are terminated with hydrogen atoms to avoid any unwanted terminal effect (Koskinen et al., 2008). I have modeled adenine, guanine, cytosine, thymine and uracil with pentose sugar connected to the respective bases by the β-glycosidic bond. Since the structural features that distinguishes RNA from DNA are present of uracil base and 2' hydroxyl –OH group in the constituted ribose sugar in RNA. So I have modeled thymine with the deoxy-ribose sugar, and all other bases are modeled with ribose sugar. Now these two structures act as proper miniature models of RNA and DNA to understand the interaction with the planar graphene molecule by quantum chemical calculations. Initially the nucleosides are placed vertically above, around 4 Å from the center of the planar hexagonal ring of the graphene molecule parallel to the graphene sheet. All the modeling was done with the help of Accelrys Discovery studio (Studio, 2007) and Molden (Schaftenaar and Noordik, 2000) software.

Dispersion force serves as the major attractive interactions between the non-polar molecules (Paton and Goodman, 2009). Dispersion forces in stacked bases are

sometimes very large and contribute a significant amount of the stabilization energy, sometimes comparable to hydrogen bond energy also. Graphene has amphiphilic nature, the surface is hydrophobic, while the terminal edges carry some amount of hydrophilic property with varying strengths (Panigrahi et al., 2011a). These differential nature of nano graphene make it a promising carrier in nanomedicine (Sun et al., 2008). The surface of nucleobases is also hydrophobic in nature. So these nucleobases may interact with the graphene with the non-covalent dispersion interaction. I have optimized all the complex systems (graphene + nucleosides) to get the energy minimized configuration, as different physical and chemical properties of a system can be best studied in its minimized configuration only. For all the quantum chemical calculations, I have employed dispersion corrected density functional approach using ωb97xd/6-31G** (Chai and Head-Gordon, 2008) method, which has been incorporated in Gaussian g09 (Frisch et al., 2009) to take in account of the dispersion interaction specially. Equation of plane was fitted to the coordinates of the carbon atoms of the optimized graphene by least squares fit method and root mean square deviation (rmsd) of the carbon atomic positions from the best-fit plane was calculated for all the systems.

The total interaction energy of each system have been calculated using the following equation

E_{int} = E(complex)- E_{xo}(isolated graphene)- E_{Yo}(isolated nucleobase) + BSSE …… [6.2.1]

The Basis set superposition error (BSSE) has been calculated using Boys and Bernardi function counter poise method (Boys and Bernardi, 1970). I have also carried out frequency calculation on the optimized geometry, using the same method and basis set that was used for optimizing the systems, to calculate different thermodynamic

parameters such as entropy, enthalpy, and free energy of the systems. To analyze the degree of binding, I have also carried out the charge transfer analysis of the nucleosides and graphene sheet using the Natural Bond Orbital (NBO) approach (Carpenter and Weinhold, 1988, Reed et al., 1988)

The enthalpy of the reaction is calculated as

$$\Delta H = \sum(\varepsilon_0 + H_{corr})_{products} - \sum(\varepsilon_0 + H_{corr})_{reacants} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad [6.2.2]$$

and Gibbs free energy of reaction is calculated as

$$\Delta G = \sum(\varepsilon_0 + G_{corr})_{products} - \sum(\varepsilon_0 + G_{corr})_{reacants} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [6.2.3]$$

ε_0 is the electronic energy of the system, E_{total} and S_{total} denote total internal energy and total entropy of the system, having contributions from translational, rotational, vibrational and electronic motion given as

$$E_{total} = E_t + E_r + E_v + E_e \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [6.2.4]$$

$$S_{total} = S_t + S_r + S_v + S_e \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [6.2.5]$$

All of the term includes zero-point energy.

H_{corr} is the corrections to the enthalpy due to internal energy defined as

$$H_{corr} = E_{total} + K_B T \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [6.2.6]$$

G_{corr} is the corrections to the Gibb's free energy due to internal energy defined as

$$G_{corr} = H_{corr} - TS_{total} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [6.2.7]$$

6.2.3. Results

Upon optimization, the nucleosides are found to be stacked on the planar nano graphene sheet (Figure 6.2.1). I have calculated the rmsd for the optimized isolated

graphene as well as graphene nucleoside complexes (Table 6.2.1). I observed that the isolated (6×6) graphene remain in the planar conformation after optimization through DFT-D method with rmsd value 0.008 Å, however I found deviations from the planar structure of graphene in all the optimized structures of the graphene nucleoside complexes, as noted from its rmsd values, which range from 1.63 Å to 3.35 Å. The deviation from the planar graphene follow the order A > G > U > C > T. I focus mainly on the thymine and uracil complex systems since these are the principal bases, which can differentiate between DNA and RNA. On analyzing the optimized geometry of the complex systems graphene + uridine (Graphene + Uracil sugar) and Graphene + Thymidine) (graphene + thymine sugar), I find that in both the cases O3'-H3' of the constituents sugar points towards the planar graphene sheet with close approach forming O-H...π contacts.

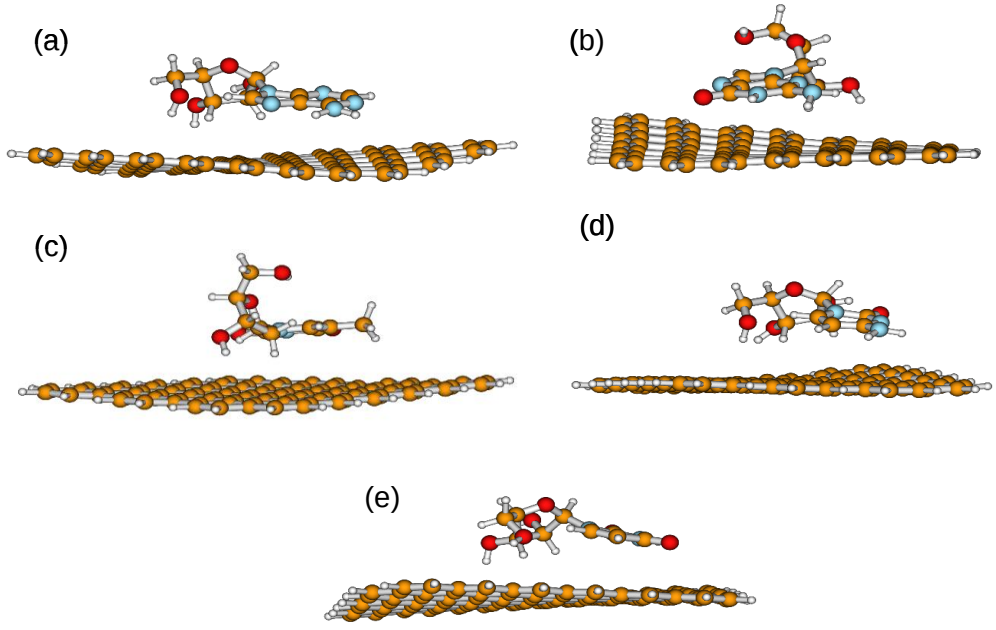


Figure 6.2.1. Optimized geometry of stacked complexes of graphene with (a) Adenosine (b) Guanosine (c) Thymidine (d) Cytidine (e) Uridine

In case of thymine nucleoside, the closest O3'-H3'…C is found to be around 2.54 Å, and the angle <O3'-H3'…C> is obtained as 129.35^{o}. These distances and angles corresponds to weak O-H… π types of hydrogen bonds. While in case of uracil nucleoside, the O3'-H3'…C bond distances are found to be 2.34 Å, and the <O3'-H3'…C> be obtained as 165.32^{o}, indicating strong hydrogen bonds between the graphene and uracil sugar, comparable to that of thymine sugar complex (Panigrahi et al., 2011b). The O2 and O4 groups of the thymine and uracil also interact with the graphene sheet, but the magnitudes of interaction seems to be very low as comparable to that of O-H… π types of contact.

The BSSE corrected interaction energy has been calculated for all the complex systems and they follows the trends G > A > U > T > C. Interaction energy of the graphene-uracil nucleoside is found to be -22.26 kcal/mol, while that of graphene-thymine nucleoside is found to be -20.30 kcal/mol. One can get the conclusion about the stability of the system from the interaction energy data. So I can infer that uracil nucleoside interacts with the graphene more strongly than that of the thymine nucleoside. These interaction energy values also well correlate the O-H…π type H-bond lengths and angles values obtained in both the cases.

Once a stationary point has been reached by geometry optimization, it is necessary to carry out the frequency calculation to check whether this point is a minimum/maximum or a hilltop. The nature of stationary point can be checked by calculating the frequency and observing how many imaginary frequencies are present. A stationary point is said to be minimum if it does not have any imaginary frequency (Jensen, 2007).

Table 6.2.1. Interaction energy, rmsd and thermochemistry of the graphene-basesugar complexes

System	RMSD plane fit in Å	ΔE (kcal/mol) BSSE corrected	ΔH (kcal/mol)	ΔG (kcal/mol)	ΔS (Cal/Mol-Kelvin)
Graphene+ Adenosine	3.35	-26.84	-30.84	-15.92	-50.13
Graphene+ Guanosine	2.82	-28.85	-33.41	-16.66	-56.15
Graphene+ Thymidine	1.63	-21.66	-23.80	-10.22	-45.54
Graphene+ Cytidine	2.03	-15.94	-20.13	-5.92	-48.00
Graphene+ Uridine	2.73	-22.26	-26.39	-11.95	-48.44
Graphene isolated (6 x 6)	0.008				

Frequency studies of the entire complex as well as isolated systems give no imaginary frequency indicating all have attained their minima. Thermodynamic parameters for binding can be calculated from these frequencies. All the calculations are carried out at 298.15 K and 1 atm. pressure. I have calculated the change in enthalpy and free energy of all the complex systems (Table 6.2.1). The ΔH of Graphene + Uridine is found to be -26.39 kcal/mol, where as that of ΔH of Graphene + Thymidine is calculated to be -23.80 kcal/mol. Similarly the ΔG of graphene+ uridine and graphene + thymidine are found to be -11.95 kcal/mol and -10.22 kcal/mol respectively. As it is well known that the system is more favorable with increase in the negative value of ΔG. The ΔG, ΔH, ΔS for all the complex graphene nucleoside complexes follow the trend G > A > U > T > U. All these values indicated formation of stable graphene-uracil nucleoside complex and

this may affect subsequent unzipping of the si-RNA structure as observed by my counterpart MD simulation studies conducted by my collaborators (Santosh et al.).

I have also calculated the NBO charges of the adenosine, guanosine, thymidine, cytidine, and uridine in complex systems and compared them with that of the isolated nucleosides. The differences in NBO charges of the major hydrogen bond donor atoms of the nucleosides which interact with the graphene sheet are given in Table 6.2.2. In case of adenosine, the O3'-H3' and hydrogens of the $-NH_2$ groups come closer to the graphene sheet, with distances of closest approach as 2.30 Å, 2.84 Å and 2.95 Å respectively. The difference in NBO charges are calculated as 0.009, 0.002, and 0.002 in case of the H3', hydrogens of the $-NH_2$ group respectively after complexation with the graphene sheet. In case of guanosine, the closest interaction is observed in between the O2'-H2' (2.73 Å), hydrogens of $-NH_2$ group (3.21 Å and 2.93 Å) and O6 (3.08 Å) groups and graphene sheet. The differences in NBO charges are obtained as -0.025 for O6, 0.010 for hydrogen of O2'-H2' and 0.003 and 0.005 for the $-NH_2$ group. These can be well correlated with the interaction energy of the graphene nucleosides complexes. I observed that charge transfer is more significant for the O3'-H3' and O4 atoms for uridine molecule with graphene, since they interacts strongly with the graphene sheet (Table 6.2.2). The O2 of thymine shows negligible amount of charge transfer with the graphene sheet. So from this analysis, I can clearly observed that uridine interacts with the graphene sheet more strongly than that of the thymidine, which is well correlated with the hydrogen bond strengths, interaction energy, and thermochemical analysis of the systems.

Table 6.2.2. Difference in NBO charges of major hydrogen bond donors of Nucleosides with graphene sheet

Atom No.	Graphene + Thymidine	Graphene + Uridine
	Difference in NBO charge	Difference in NBO charge
O3'-H3'	0.005	0.007
O2	0.002	0.000
O4	-0.007	-0.014

6.2.4. Discussion

From the present investigation using dispersion corrected density functional theory, I demonstrate that uracil residues make strongest contacts with the graphene sheet through van der Waals and specific H-bonding interaction involving 3'-OH group of the ribose sugar. The 2'-OH also has some contribution, but its contribution is comparatively less. These interactions can be the dragging force for the double helical siRNA to unzip, which are stabilized subsequently by several such O-H...π interaction. Subsequent molecular dynamics study by my collaborators also shows complete siRNA unzipping on graphene substrate where the equivalent double stranded DNA remains stable throughout the simulation time. This may be due to absence of –OH group in DNA and presence bulky methyl group in thymine, leading to less interaction energy of the graphene+thymidine complex. I have also observed deformation on the plane of the graphene upon complexation with nucleosides. This signifies that inclusion of nucleosides also effect the structural features of the graphene. These unusual phenomena of these findings may have very important implications while designing graphene based delivery platform for siRNA delivery.

Section 6.3

Interaction of Small Interfering RNA (siRNA) with Single Walled Carbon Nanotube: Quatum Chemical Approach

6.3.1. Introduction

The discovery of reagents to inhibit RNA synthesis, translation has been attracted researchers to study the biological processes. These tools include small interfering RNA (si-RNA), small hairpin loops and ribozymes. Small interfering RNA is also commonly known as silencing RNA, typically 21 to 23 nucleotides in length are actively being studied due to their potential influence on cell functionality and applications in medicine (Fire et al., 1998, Novina et al., 2002, Medarova et al., 2007). The potential role of the siRNA has already been discussed more elaborately in section 6.2. A number of groups have shown that single-walled carbon nanotubes (SWNTs) and multiwalled carbon nanotubes (MWNTs) can be used to deliver therapeutic and diagnostic small molecules and macromolecules to cells (Kostarelos et al., 2007, Heller et al., 2005). CNTs can act as suitable candidates for loading biomolecules and drugs on its surface for drug delivery and medical imaging (Liu et al., 2009) due to their high surface area, which can potentially act as a template for loading molecules such as peptides, proteins, nucleic acids, and drugs through $\pi-\pi$ stacking interaction. The cellular uptake of free peptides and oligonucleotides is very poor; therefore these molecules are conjugated onto CNT for improvements in the delivery of such biologically important molecules (Kostarelos et al., 2007, Pantarotto et al., 2003). Chemical modifications or proper functionalization of the surface of the CNT can improve the biocompatibility and reduce toxicity to cells. CNTs that are functionalized with polymers such as Polyethylene glycol (PEG), CONH-(CH2)(6)-NH3+Cl- or single stranded DNA (ssDNA) are efficient transporters of siRNA into human T cells and primary cells (Liu et al., 2007, Liu et al., 2009, Kam et al., 2006). Recently, CNTs have also been conjugated with siRNA that help in siRNA-mediated

gene silencing. However, the mechanism underlying this effect has not been clearly elucidated. The delivery of siRNA by CNTs will certainly become more widespread and it is expected that therapeutic applications will also diversify. With this knowledge, we have studied the mode of binding of the siRNA with nucleobases and nucleosides, and try to analyze several other properties such as charge transfer, hydrogen bonding interaction through *ab initio* quantum chemical approach. My computational findings are well supported by the all atom molecular dynamics (MD) simulations of siRNA-CNT complex carried out by my collaborator (Santosh et al., 2012).

6.3.2. Methods

I have carried out quantum chemical analysis to understand the interactions of the CNT with the DNA and RNA. The valencies of the Carbon atoms at the ends of the (6, 6) CNT were satisfied by adding necessary hydrogen atoms. The structural features that distinguish RNA from DNA are the presence of uracil base and 2'-OH groups of the ribose sugars. So I have modeled four miniature systems, which carry the features of the bigger system. My considered systems are as follows (i) CNT with one uracil nucleobase, (ii) CNT with one thymine nucleobase, (iii) CNT with one uridine nucleoside (uracil attached with C3'-endo ribose sugar) and (iv) CNT with thymidine nucleoside (thymine nucleobase attached with C2'-endo deoxyribose sugar) as shown in Figure 6.3.1. These initial structures were built using MOLDEN (Schaftenaar and Noordik, 2000) software. Initially the bases were placed parallel to the CNT. Free geometry optimization of all the four systems discussed above, were carried out without any constraints using density functional theory with ωB97XD/6-31G(d,p) basis set (Chai and Head-Gordon, 2008),

which includes dispersion correction to density functional theory, giving rise to energy E_{XY}, by Gaussian09 (Frisch et al., 2009). I have also optimized the isolated CNT, giving rise to energies E_{X0}, the two nucleobases and the two nucleosides in un-complexed isolated from having energy E_{Y0}. Basis set superposition error (BSSE) has been corrected by Boys-Bernardi function counterpoise method (Boys and Bernardi, 1970). The BSSE corrected interaction energies (E_{int}) of each system were calculated using the following equation,

$$E_{int.} = E_{XY} - E_{X0} - E_{Y0} + BSSE \quad \ldots\ldots\ldots\ldots\ldots\ldots [6.3.1]$$

where E_{XY} represents potential energy of the whole system (CNT + nucleobase or nucleoside), and E_{X0} and E_{Y0} represent energy of the isolated CNT and optimized nucleoside/nucleobase residues. The deformation energy which signifies the energy required to deform the isolated monomers in forming the complex system is calculated as follows.

$$E(\ Deformation) = (E_X - E_{X0}) + (E_Y - E_{Y0}) \quad \ldots\ldots\ldots\ldots\ldots\ldots [6.3.2]$$

The deformation energies and the corrected interaction energies obtained from all the four systems are compared. I have also carried out the Mullikan charge analysis of the optimized systems which involved the sugar ring in its configurations as presence of hydroxyl -OH group of the sugar may have some effect in the CNT system.

All molecular dynamics simulations were carried out by my experimental collaborators (Santosh et al., 2012).

6.3.3. Results

I observed that upon optimization all the nucleobases and nucleosides remain stacked on the CNT wall through π-π stacking interaction (Figure 6.3.1). I find interaction energy of the CNT-uracil nucleobase complex is -9.64 kcal/mol whereas that of the CNT-thymine nucleobase complex is about -12.25 kcal/mol. Interaction energy between thymine and CNT was calculated earlier using Hartree-Fock and related methods giving significant attraction between the two (Varghese et al., 2009).

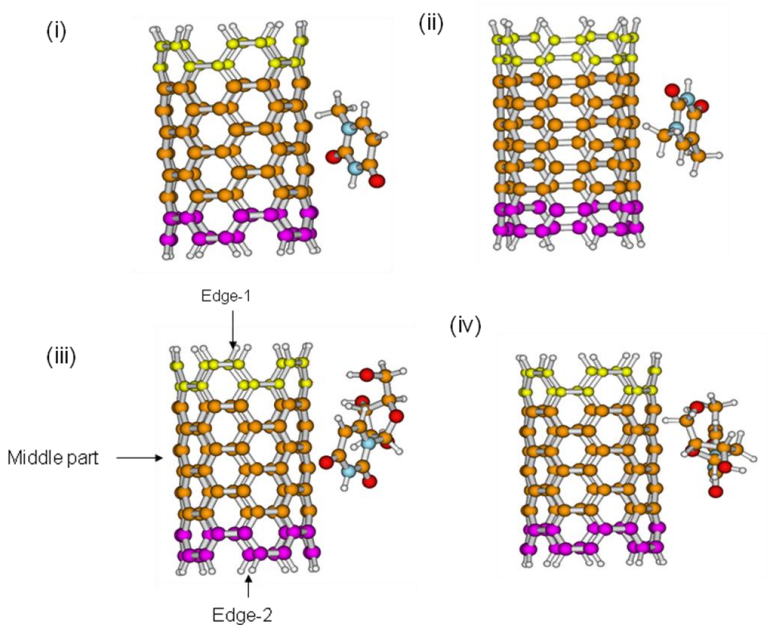


Figure 6.3.1. Optimized geometries of the four complex systems studied: CNT with (i) uracil (ii) thymine (iii) uracil-sugar and (iv) thymine-sugar. The yellow colored atoms represent the terminal carbon atoms of the CNT those are in close approach to the base/nucleoside (Edge-1), whereas the violet colored atoms represent the terminal carbon atoms, those remain far away from the base/nucleotide (Edge-2)

Although I have optimized the systems and calculated the energies using more robust density functional theory which includes dispersion correction (DFT-D), the value

of interaction energy of thymine obtained in this method is quite similar to that of the previous estimate (-11.3 kcal/mol), without dispersion and electron correlation effects. The difference in energy between thymine and uracil bases obtained in this method may arise due to the stronger non-polar interaction between the methyl groups of thymine and the carbons of CNT. Since sugar backbone plays a crucial role in maintaining the structure and stability of RNA or DNA, I have also optimized the CNT-uridine and CNT-thymidine complexes including sugars attached to the nucleobases. This now represents the true model of CNT-siRNA complex in miniature form. The interaction energy values shown in Table 6.3.1 indicate that CNT with uridine has stronger binding (-18.72 kcal/mol), than that of the CNT with thymidine (-16.25 kcal/mol). This is due to possibility of weak hydrogen bond formation between the three -OH groups of uridine molecule with carbon atoms of CNT.

Table 6.3.1. Interaction energy values of the optimized systems

System	System name	BSSE (kcal/mol)	Deformation energy (kcal/mol)	$E_{Int.}$ (kcal/mol)
Figure 6.3.1(i)	CNT+Uracil	3.27	1.49	-9.64
Figure 6.3.1(ii)	CNT+Thymine	3.68	0.77	-12.65
Figure 6.3.1 (iii)	CNT+Uridine	5.36	1.79	-18.72
Figure 6.3.1 (iv)	CNT+Thymidine	5.94	1.00	-16.25

The deformation energy data of the former system also follow the same trend, with CNT with uracil-sugar is found to be comparatively more deformed. This signifies that CNT with uracil-sugar complex is found to be more stable than that of thymine-sugar complex, in other words the CNT is found to be more reactive towards the uracil-sugar complex, which is in agreement with MD simulation data (Santosh et al., 2012). From the

interaction energy values, I can also conclude that presence of ribose sugar added a good amount of stability to the CNT. I believe this favors the siRNA unzipping and subsequent wrapping on CNT whereas only adsorption of dsDNA on CNT is observed. The uridine molecule forms two hydrogen bonds involving O5'-H5' of ribose sugar with two carbon atoms at edge-1 (Figure 6.3.1) of CNT with hydrogen bond distances of 2.62 Å and 2.52 Å and associated O-H...C angles of 150.81^{o} and 166.16^{o}, respectively. Another hydrogen bond is found between the O3'-H3' of the ribose sugar ring with one of the carbon atom that lies towards the middle of the CNT with hydrogen bond distance and angle of 2.50 Å and 154.39^{o}, respectively. So uridine is able to form three good hydrogen bonds with the CNT and stabilize the system provided the carbon atoms have sufficient negative charges.

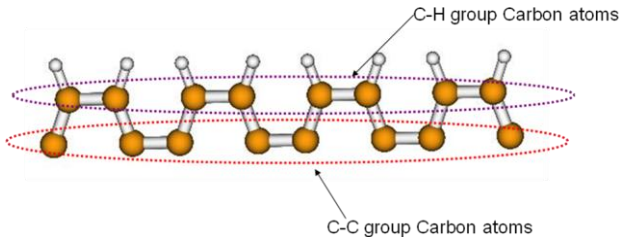


Figure 6.3.2. Schematic diagram of terminal Carbon atoms, C-C bond Carbon atoms are circled in red dotted lines, whereas the C-H bond Carbon atoms are circled in violet dotted line

Generally hydrogen bonds take place between polar moieties and the carbons of CNT are supposed to be rather neutral, questioning its participation as hydrogen bond acceptor. I have therefore analyzed partial charge of all the atoms of CNT calculated by Mulliken population analysis. As expected the terminal C-H groups are slightly polar, the carbons not bonded to hydrogen are neutral (Carbons of C-C group, those lies inwardly at the terminal edges, Figure 6.3.2) and the central Carbons present at the middle region of

the CNT have nearly zero charge. Moreover all the charges at the two edges are symmetrically distributed with zero standard deviations when it is not complexed with any nucleobase/nucleoside. In presence of thymine or uracil bases, the properties of the CNT remain nearly unchanged. Presence of nucleoside residues, particularly the uridine nucleoside, breaks the symmetry of the CNT significantly (shown in Table 6.3.2), as detected from the larger standard deviations of the charges of different groups of atoms. I have classified the CNT atoms into the following types (Figure 6.3.2) (i) carbon atoms at the edges which are not bonded to any hydrogen C-C carbons, (ii) carbon atoms of C-H group and iii) carbon atoms which lie in the central region of CNT. The terminal atoms can be further classified into edge-1 and edge-2, depending on proximity to the binding nucleoside. When uridine binds to CNT, partial charges of the C-C Carbon atoms at the edge-1 change significantly. In this case, the average charge of the carbon atom decreases and the standard deviation increases, which signifies delocalization of the charges. In presence of the polar uridine, the electrons of the CNT move significantly through the extended conjugation and accumulate near the uridine. The thymidine also alters charges of these Carbon atoms in CNT but to a lesser extent. I notice that the atoms, which are far away from the nucleoside (those of the edge-2), do not undergo any noticeable changes in both the cases. The partial charges of the central Carbon atoms also alter significantly, particularly in case of uridine, which is reflected in larger standard deviations (partial charge of Carbon changes from 0.002 to –0.060 for acting as hydrogen bond acceptor). The other carbon atoms of CNT that are forming hydrogen bonds with the O-H groups of uridine acquire Mulliken charges of -0.063 and -0.154 from 0.020 and -0.149, respectively. This signifies that the uridine can polarize CNT and has strong binding

affinity with CNT than that of the thymidine, which is correlated with its interaction energy data also.

Table 6.3.2. Average values of Mulliken charges of the seven types of atoms presented in CNT. Values given in parentheses are standard deviations.

Systems	Edge-1 of C of C-C group	Edge-1 C of C-H group	edge-1 H	edge-2 C of C-C group	edge-2 C of C-H group	edge-2 H	Middle carbons
Isolated CNT	0.020 (0.000)	-0.150 (0.000)	0.124 (0.000)	0.020 (0.000)	-0.150 (0.000)	0.124 (0.000)	0.002 (0.003)
CNT + uridine	0.012 (0.027)	-0.149 (0.003)	0.128 (0.002)	0.020 (0.002)	-0.149 (0.001)	0.126 (0.001)	0.0005 (0.024)
CNT + thymidine	0.018 (0.011)	-0.150 (0.002)	0.125 (0.00)	0.021 (0.005)	-0.150 (0.000)	0.124 (0.001)	0.003 (0.008)

Supportive MD simulation studies also confirm that siRNA undergoes unzipping and wrapping around the CNT, which I do not observe in case of dsDNA. This can be explained due to the relatively strong A-T base-pair interaction energy compared A-U base-pair interaction energy (Sponer et al., 2004) that is present in siRNA. The van der Waals attraction, binding energy and the number of close contacts of dsDNA to CNT remain less compared to siRNA.

6.3.4. Discussion

I have demonstrated the interactions of the uracil and thymine nucleobases and nucleosides with the CNT through dispersion corrected density functional (DFT-D) approach. I observed that uracil nucleoside form more stable complex with the graphene, than that of the thymine nucleoside. The Mullikan charge analysis also confirms that charge transfer taking place between the CNT and uracil nucleoside is significant. From

this present theoretical investigation, I can conclude that siRNA has strong binding capacity with the CNT, than that of the dsDNA. The subsequent molecular dynamics simulation carried out by my collaborators also confirms QM finding (Santosh et al., 2012). The adsorbed siRNA can be delivered to virus infected cell via endocytosis to reduce the expression of specific unwanted genes.

Further functionalization of the CNT may increase the efficiency of the drug delivery capacity, may open up several new research on CNT bases drug delivery system.

Chapter 7

Summary and Future Prospective

7.1. Summary

Throughout my research work, I have provided detail descriptions of different types of basepairing patterns observed and possible in RNA. I have studied the structures, stability and dynamics of these basepairs using Density Functional Theory (DFT), with popular hybrid functionals. Taking all these into accounts, I have also prepared a database, named as RNA non-canonical basepair database, which provides a complete description of the possible RNA basepairs, including protonated basepairs and model basepairs. The database provides complete qualitative and quantitative analysis of the basepairs, which includes geometry of the optimized basepairs, interaction energy, isosteric parameters, basepair orientation parameters. Since the basepairs can be part of the triplets also, so I have also provided the information about the triplets associated with each basepair type. These study and classification is expected to be helpful in understanding tertiary structure of RNA, RNA folding etc.

I have also provided an in-depth description of understanding some of the specific properties of the nanomaterials such as gold nano cluster, graphene, and carbon nanotube through quantum chemical studies, and I have also enough evidences from experimental and molecular dynamics study to correlate the quantum chemical data. I have studied proper mechanism of interaction of capping agent such as citrate group with gold nano cluster. My study confirmed importance of capping agent in stabilizing the gold nano cluster and how this capping enhances its interaction with quercetin, which is one of the most widely used antioxidant. I have analyzed the wetting properties associated with the graphene. From my quantum chemical study, I observed that zig-zag edge of the graphene is found to be more hydrophilic to water than that of the arm-chair edge, while

surface of the graphene is found to be hydrophobic in nature. I have also analyzed that the wrinkling of the larger graphene sheet as its inherent property through dispersion corrected DFT theory. The interaction of nucleobases with the graphene sheet have been studied, and the binding energy is found to be follow the trends G > A > C > T > U. My collaborators through their molecular dynamics study observed that siRNA unzips on graphene and carbon nanotube, while dsDNA does not. I have studied these properties quantum mechanically. My subsequent charge analysis, hydrogen bonding interactions also prove that uridine has stronger interaction with graphene and CNT than that of the thymidine. These may further enhance unzipping of siRNA.

7.2. Future Prospective

In such a dynamic scientific world, where in each moment we are exposed to new scientific thoughts, it is really difficult to conclude any research work. I always remember the lines from my thesis supervisor, which says that finding of one paper or one research topic should open up several doors to carry out new research. In my thesis, I have presented the detail analysis of the RNA basepairs, however triplets are emerging as new variants of RNA. Analysis of the triplets in more detail will be helpful in understanding its importance in various biological processes. Throughout my work, I have used quantum chemical calculations to study the biological molecules, however in reality water and environment are known to play crucial role in the functions of proteins. So the systems can be studied in more realistic approach using Car-Parrinello molecular dynamics (CPMD) (Car and Parrinello, 1985), since the biological systems are intrinsically dynamic in nature.

Through experimental and theoretical approach, we proved the importance of citrate capping in stabilizing the gold nano particles. However, there are several other capping agents are there such as thio group, peptides etc. Study of detail mechanism of these capping agents with gold nano cluster will be helpful.

I have studied graphene and CNT without any functionalization, but functionalization are important to manipulate the physical and chemical properties and to study its interactions with environments. Functionalization can be included by oxidation or hydrogenation process. They can be included through covalent or non-covalent interactions. Functionalization of graphene and CNT can be useful to study its interactions with different biological macromolecules, and identifications of new therapeutic targets in foreseeable futures.

BIBLIOGRAPHY

Akca, S., Foroughi, A., Frochtzwajg, D. & Postma, H. W. C. (2011) Competing Interactions in DNA Assembly on Graphene. *PloS one,* 6**,** e18442.

Allerhand, A. & Von Rague Schleyer, P. (1963) A survey of C-H groups as proton donors in hydrogen bonding. *J. Am. Chem. Soc.,* 85**,** 1715.

Almakarem, A. S. A., Petrov, A. I., Stombaugh, J., Zirbel, C. L. & Leontis, N. B. (2012) Comprehensive survey and geometric classification of base triples in RNA structures. *Nucleic Acids Res.,* 40**,** 1407.

Altona, C. & Sundaralingam, M. (1972) Conformational analysis of the sugar ring in nucleosides and nucleotides. New description using the concept of pseudorotation. *J. Am. Chem. Soc.,* 94**,** 8205.

Anjaneyulu, M. & Chopra, K. (2004) Quercetin, an anti-oxidant bioflavonoid, attenuates diabetic nephropathy in rats. *Clin. Exp. Pharmacol. Physiol.,* 31**,** 244.

Antony, J. & Grimme, S. (2008) Structures and interaction energies of stacked graphene–nucleobase complexes. *Phys. Chem. Chem. Phys.,* 10**,** 2722.

Babcock, M. S. & Olson, W. K. (1994) The Effect of Mathematics and Coordinate System on Comparability and "dependencies" of nucleic acid structure parameters,. *J. Mol. Biol.,* 237**,** 98.

Bader, R. F. W. (1991) A quantum theory of molecular structure and its applications. *Chem. Rev.,* 91**,** 893.

Balandin, A. A., Ghosh, S., Bao, W., Calizo, I., Teweldebrhan, D., Miao, F. & Lau, C. N. (2008) Superior thermal conductivity of single-layer graphene. *Nano Lett.,* 8**,** 902.

Bandyopadhyay, D. & Bhattacharyya, D. (2006) Estimation of strength in different extra Watson–Crick hydrogen bonds in DNA double helices through quantum chemical studies. *Biopolymers,* 83**,** 313.

Banerjee, S. & Bhattacharyya, D. (2008) Electronic properties of nano-graphene sheets calculated using quantum chemical DFT. *Comput. Mat. Sci.,* 44**,** 41.

Banerjee, S., Sardar, M., Gayathri, N., Tyagi, A. & Raj, B. (2005) Conductivity landscape of highly oriented pyrolytic graphite surfaces containing ribbons and edges. *Phys. Rev. B,* 72**,** 075418.

Banerjee, S., Sardar, M., Gayathri, N., Tyagi, A. & Raj, B. (2006) Enhanced conductivity in graphene layers and at their edges. *Appl. Phys. Lett.,* 88**,** 062111.

Bhattacharyya, D. & Bansal, M. (1990) Local variability and base sequence effect in DNA crystal structures. *J. Biomol. Struct. Dyn.,* 8**,** 539.

Bansal, M., Bhattacharyya, D. & Ravi, B. (1995) NUPARM and NUCGEN: software for analysis and generation of sequence dependent nucleic acid structures. *CABIOS,* 11**,** 281.

Bao, W., Miao, F., Chen, Z., Zhang, H., Jang, W., Dames, C. & Lau, C. N. (2009) Controlled ripple texturing of suspended graphene and ultrathin graphite membranes. *Nat. Nanotechnol.,* 4**,** 562.

BarathManiKanth, S., Kalishwaralal, K., Sriram, M., Pandian, S. B. R. K., Youn, H., Eom, S. H. & Gurunathan, S. (2010) Anti-oxidant effect of gold nanoparticles restrains hyperglycemic conditions in diabetic mice. *J Nanobiotechnology,* 8:16.

Basu, S., Majumdar, R., Das, G. K. & Bhattacharyya, D. (2005) Energy barriers and rates of tautomeric transitions in DNA bases: ab initio quantum chemical study. *Indian J. Biochem. Biophys.,* 42**,** 378.

Becke, A. D. (1993) Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys,* 98**,** 5648.

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S. H., Srinivasan, A. & Schneider, B. (1992) The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.,* 63**,** 751.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000) The protein data bank. *Nucleic Acids Res.,* 28**,** 235.

Betts, L., Josey, J. A., Veal, J. M. & Jordan, S. R. (1995) A nucleic acid triple helix formed by a peptide nucleic acid-DNA complex. *Science,* 270**,** 1838.

Bhattacharyya, D., Koripella, S. C., Mitra, A., Rajendran, V. B. & Sinha, B. (2007) Theoretical analysis of noncanonical base pairing interactions in RNA molecules. *J. Biosci.,* 32**,** 809.

Biegler-König, F. (1990) AIM2000, version 1.0; University of Applied Science: Bielefeld, Germany.(b) Bader, RFW Atoms in Molecules: A Quantum Theory. Oxford University Press: Oxford.

Biswas, S. & Drzal, L. T. (2008) A novel approach to create a highly ordered monolayer film of graphene nanosheets at the liquid− liquid interface. *Nano Lett.,* 9**,** 167.

Blackburn, G. M. (2006) *Nucleic acids in chemistry and biology,* Royal Society of Chemistry.

Bloomfield, V. A., Crothers, D. M. & Tinoco, I. (2000) *Nucleic acids: structures, properties, and functions,* Univ Science Books.

Boggs, J. M. (1986) Effect of lipid structural modifications on their intermolecular hydrogen bonding interactions and membrane functions. *Biochem. Cell Biol.,* 64**,** 50.

Boggs, J. M. (1987) Lipid intermolecular hydrogen bonding: influence on structural organization and membrane function. *Biochim. Biophys. Acta,* 906**,** 353.

Bonn, D., Eggers, J., Indekeu, J., Meunier, J. & Rolley, E. (2009) Wetting and spreading. *Rev. Mod. Phys.,* 81**,** 739.

Boys, S. & Bernardi, F. (1970) The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.,* 19**,** 553.

Brameld, K., Dasgupta, S. & Goddard III, W. A. (1997) Distance dependent hydrogen bond potentials for nucleic acid base pairs from ab initio quantum mechanical calculations (LMP2/cc-pVTZ). *J. Phys. Chem. B,* 101**,** 4851.

Brett, A. M. O. & Chiorcea, A. M. (2003) Atomic force microscopy of DNA immobilized onto a highly oriented pyrolytic graphite electrode surface. *Langmuir,* 19**,** 3830.

Broitman, S. L., Im, D. D. & Fresco, J. R. (1987) Formation of the triple-stranded polynucleotide helix, poly (AAU). *Proc. Natl. Acad. Sci. USA,* 84**,** 5120.

Broker, T. R. (1973) An electron microscopic analysis of pathways for bacteriophage T4 DNA recombination. *J. Mol. Biol.,* 81**,** 1.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. comput. Chem,* 4**,** 187.

Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.,* 34**,** 5402.

Cao, X. & Dolg, M. (2010) Relativistic pseudopotentials in: Relativistic methods for chemists. *Ed. by M. Barysz, and Y. Ishikawa, Springer UK***,** 215.

Capra, J. A., Paeschke, K., Singh, M. & Zakian, V. A. (2010) G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae. *PLoS Comp. Biol.,* 6**,** e1000861.

Car, R. & Parrinello, M. (1985) Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.,* 55**,** 2471.

Carpenter, J. E. & Weinhold, F. (1988) Analysis of the geometry of the hydroxymethyl radical by the "different hybrids for different spins" natural bond orbital procedure. *J. Mol. Struct. (Theochem),* 169**,** 41.

Carrington, J. C. & Ambros, V. (2003) Role of microRNAs in plant and animal development. *Science,* 301**,** 336.

Case, D. A., Darden, T. A., Cheatham III., T.E. , Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Merz, K. M., Wang, B., Pearlman, D. A., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C., Caldwell, J. W., Ross, W. S. & P.A., K. (2004) AMBER 8, University of California, San Francisco

Černý, J. & Hobza, P. (2007) Non-covalent interactions in biomacromolecules. *Phys. Chem. Chem. Phys.,* 9**,** 5291.

Chai, J. D. & Head-Gordon, M. (2008) Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.,* 128**,** 084106.

Chastain, M. & Tinoco, I. (1991) Structural elements in RNA. *Prog. Nucleic Acid Res. Mol. Biol.,* 41**,** 131.

Chawla, M., Sharma, P., Halder, S., Bhattacharyya, D. & Mitra, A. (2011) Protonation of Base Pairs in RNA: Context Analysis and Quantum Chemical Investigations of Their Geometries and Stabilities. *J. Phys. Chem. B,* 115**,** 1469.

Cheong, C. & Moore, P. B. (1992) Solution structure of an unusually stable RNA tetraplex containing G-and U-quartet structures. *Biochemistry (Mosc).* 31**,** 8406.

Cho, K., Lee, K. Y., Lee, K. H., Kim, D. & Lee, D. (2006) Specificity of molecular interactions in transient protein–protein interaction interfaces. *Protein: Struct. Funct. Bioinf.,* 65**,** 593.

Clay, E. H. & Gould, I. R. (2005) A combined QM and MM investigation into guanine quadruplexes. *J. Mol. Graphics Model.,* 24**,** 138.

Clowney, L., Jain, S. C., Srinivasan, A., Westbrook, J., Olson, W. K. & Berman, H. M. (1996) Geometric parameters in nucleic acids: nitrogenous bases. *J. Am. Chem. Soc.,* 118**,** 509.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.,* 117**,** 5179.

Couzin, J. (2002) Small RNAs make big splash. *Science,* 298**,** 2296.

Cramer, C. J. (2004) *Essentials of computational chemistry: theories and models*, John Wiley & Sons Inc.

Cybulski, S., Bledson, T. & Toczyłowski, R. (2002) Comment on "Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory treatment"[J. Chem. Phys.114, 5149 (2001)]. *J. Chem. Phys.,* 116**,** 11039.

Daniel, M. C. & Astruc, D. (2004) Gold nanoparticles: assembly, supramolecular chemistry, quantum-size-related properties, and applications toward biology, catalysis, and nanotechnology. *Chem. Rev.,* 104**,** 293.

Danilov, V. I. & Anisimov, V. M. (2005) Post Hartree-Fock studies of the canonical Watson-Crick DNA base pairs: molecular structure and the nature of stability. *J. Biomol. Struct. Dyn.,* 22**,** 471.

Danilov, V. I., Anisimov, V. M., Kurita, N. & Hovorun, D. (2005) MP2 and DFT studies of the DNA rare base pairs: the molecular mechanism of the spontaneous substitution mutations conditioned by tautomerism of bases. *Chem. Phys. Lett.,* 412**,** 285.

Das, G. & Lyngdoh, R. H. D. (2012) Role of wobble base pair geometry for codon degeneracy: purine-type bases at the anticodon wobble position. *J. Mol. Model.,* DOI: 10.1007/s00894-012-1385-4**,** 1.

Das, J., Mukherjee, S., Mitra, A. & Bhattacharyya, D. (2006) Non-canonical base pairs and higher order structures in nucleic acids: crystal structure database analysis. *J. Biomol. Struct. Dyn.,* 24**,** 149.

Dashnau, J. L., Sharp, K. A. & Vanderkooi, J. M. (2005) Carbohydrate intramolecular hydrogen bonding cooperativity and its effect on water structure. *J. Phys. Chem. B,* 109**,** 24152.

De La Paz, M. L., Ellis, G., Penadés, S. & Vicent, C. (1997) Conformational restriction by intramolecular hydrogen bonding. Carbohydrate-carbohydrate self-assembly. *Tetrahedron Lett.,* 38**,** 1659.

Dedkov, Y. S., Shikin, A., Adamchuk, V., Molodtsov, S., Laubschat, C., Bauer, A. & Kaindl, G. (2001) Intercalation of copper underneath a monolayer of graphite on Ni (111). *Phys. Rev. B,* 64**,** 035405.

Deka, A. & Deka, R. C. (2008) Structural and electronic properties of stable Au_n (n= 2-13) clusters: A density functional study. *J. Mol. Struct. (Theochem.),* 870**,** 83.

Derewenda, Z., Derewenda, U. & Kobos, P. (1994) (His)C epsilon-H...O=C hydrogen bond in the active sites of serine hydrolases. *J. Mol. Biol.,* 241, 83.

Derewenda, Z. S., Lee, L. & Derewenda, U. (1995) The Occurence of C-H...O Hydrogen Bonds in Proteins. *J. Mol. Biol.,* 252**,** 248.

Desiraju, G. R. (2011) A Bond by Any Other Name. *Angew. Chem. Int. Ed.,* 50**,** 52.

Desiraju, G. R. & Steiner, T. (2001) *The weak hydrogen bond: in structural chemistry and biology*, Oxford University Press, USA.

Dewar, M. J. S., Zoebisch, E. G., Healy, E. F. & Stewart, J. J. P. (1985) AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.,* 107**,** 3902.

Dickerson, R. (1989) Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res.,* 17**,** 1797.

Dickerson, R., Bansal, M. & Calladine, C. R. (1989) Definitions and nomenclature of nucleic acid structure parameters. *The EMBO Journal,* 8**,** 1.

Dickerson, R. E. & Chiu, T. K. (1997) Helix bending as a factor in protein/DNA recognition. *Biopolymers,* 44**,** 361.

Ding, Y., Chen, Z., Xie, J. & Guo, R. (2008) Comparative studies on adsorption behavior of thionine on gold nanoparticles with different sizes. *J. Colloid Interface Sci.,* 327**,** 243.

Dirac, P. A. M. (1928) The quantum theory of the electron. *Proc. R. Soc. Lond. A,* 117**,** 610.

Dunning Jr, T. H. (1989) Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys,* 90**,** 1007.

Egli, M. & Sarkhel, S. (2007) Lone pair-aromatic interactions: To stabilize or not to stabilize. *Acc. Chem. Res.,* 40**,** 197.

Ehlers, A., Bohme, M., Dapprich, S., Gobbi, A., Hollwarth, A., Jonas, V., Kohler, K., Stegmann, R., Veldkamp, A. & Frenking, G. (1993) A set of f-polarization functions for pseudo-potential basis sets of the transition metals Sc---Cu, Y---Ag and La---Au. *Chem. Phys. Lett.,* 208**,** 111.

El-Sayed, I. H., Huang, X. & El-Sayed, M. A. (2006) Selective laser photo-thermal therapy of epithelial carcinoma using anti-EGFR antibody conjugated gold nanoparticles. *Cancer Lett.,* 239**,** 129.

El Hassan, M. & Calladine, C. (1995) The assessment of the geometry of dinucleotide steps in double-helical DNA; a new local calculation scheme. *J. Mol. Biol.,* 251**,** 648.

Elstner, M., Hobza, P., Frauenheim, T., Suhai, S. & Kaxiras, E. (2001) Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory based treatment. *J. Chem. Phys.,* 114**,** 5149.

Elstner, M., Porezag, D., Jungnickel, G., Elsner, J., Haugk, M., Frauenheim, T., Suhai, S. & Seifert, G. (1998) Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B,* 58**,** 7260.

Felsenfeld, G., Davies, D. R. & Rich, A. (1957) Formation of a three-stranded polynucleotide molecule. *J. Am. Chem. Soc.,* 79**,** 2023.

Feng, L. & Liu, Z. (2011) Graphene in biomedicine: opportunities and challenges. *Nanomedicine,* 6**,** 317.

Fire, A., Xu, S. Q., Montgomery, M. K., Kostas, S. A., Driver, S. E. & Mello, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature,* 391**,** 806.

Fonseca Guerra, C., Snijders, J., Te Velde, G. & Baerends, E. (1998) Towards an order-N DFT method. *J. Theor. Chem. Acc.,* 99**,** 391.

Fowler, J. D., Allen, M. J., Tung, V. C., Yang, Y., Kaner, R. B. & Weiller, B. H. (2009) Practical chemical sensors from chemically derived graphene. *ACS nano,* 3**,** 301.

Frank-Kamenetskii, M. D. & Mirkin, S. M. (1995) Triplex DNA structures. *Annu. Rev. Biochem.,* 64**,** 65.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery, J., 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. W., Ayala, P. Y., Morokuma, K., Voth, G. A., Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels, A. D., Strain, M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Ortiz, J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Gonzalez, C. & Pople, J. A. (2003), Gaussian, Inc., Wallingford CT.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J., J. A., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, N. J., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, Ö., Foresman, J. B., Ortiz, J. V., Cioslowski, J. & Fox, D. J. (2009) Gaussian G09. Gaussian, Inc. Wallingford, CT.

Fujita, M., Wakabayashi, K., Nakada, K. & Kusakabe, K. (1996a) Peculiar localized state at zigzag graphite edge. *J. Phys. Soc. Jpn.,* 65**,** 1920.

Fujita, M., Yoshida, M. & Nakada, K. (1996b) Polymorphism of Extended Fullerene Networks: Geometrical Parameters and Electronic Structures. *Fullerene Sci. Technol.,* 4**,** 565.

Geim, A. K. & Novoselov, K. S. (2007) The rise of graphene. *Nat. Mater.,* 6**,** 183.

Gesteland, R. F., Cech, T. & Atkins, J. F. (2006) *The RNA world: the nature of modern RNA suggests a prebiotic RNA world*, Cold Spring Harbor Laboratory Press.

Ghosh, A. & Bansal, M. (1999) Three-centre C-HO hydrogen bonds in the DNA minor groove: analysis of oligonucleotide crystal structures. *Acta Crystallogr. Sect. D. Biol. Crystallogr.,* 55**,** 2005.

Ghosh, A. & Bansal, M. (2003) A glossary of DNA structures from A to Z. *Acta Crystallogr. Sect. D. Biol. Crystallogr.,* 59**,** 620.

Ghosh, P., Han, G., De, M., Kim, C. K. & Rotello, V. M. (2008) Gold nanoparticles in delivery applications. *Adv. Drug Del. Rev.,* 60**,** 1307.

Gil, A., Adhikari, S., Scarpa, F. & Bonet, J. (2010) The formation of wrinkles in single-layer graphene sheets under nanoindentation. *J. Phys.: Condens. Matter,* 22**,** 145302.

Gilbert, D. E. & Feigon, J. (1999) Multistranded DNA structures. *Curr. Opin. Struct. Biol.,* 9**,** 305.

Gill, P. M. W., Johnson, B. G., Pople, J. A. & Frisch, M. J. (1992) The performance of the Becke--Lee--Yang--Parr (B--LYP) density functional theory with various basis sets. *Chem. Phys. Lett.,* 197**,** 499.

Giudice, E., Varnai, P. & Lavery, R. (2003) Base pair opening within B-DNA: free energy pathways for GC and AT pairs from umbrella sampling simulations. *Nucleic Acids Res.,* 31**,** 1434.

Gould, I. R. & Kollman, P. A. (1994) Theoretical investigation of the hydrogen bond strengths in guanine-cytosine and adenine-thymine base pairs. *J. Am. Chem. Soc.,* 116**,** 2493.

Gowtham, S., Scheicher, R. H., Ahuja, R., Pandey, R. & Karna, S. P. (2007) Physisorption of nucleobases on graphene: Density-functional calculations. *Phys. Rev. B,* 76**,** 033401.

Grabowski, S. J. (2006) Theoretical studies of strong hydrogen bonds. *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.,* 102**,** 131.

Grabowski, S. J. (2007) pi-H...O Hydrogen Bonds: Multicenter Covalent pi-H Interaction Acts as the Proton-Donating System. *J. Phys. Chem. A,* 111**,** 13537.

Guardia, T., Rotelli, A. E., Juarez, A. O. & Pelzer, L. E. (2001) Anti-inflammatory properties of plant flavonoids. Effects of rutin, quercetin and hesperidin on adjuvant arthritis in rat. *Il farmaco,* 56**,** 683.

Guinea, F., Horovitz, B. & Le Doussal, P. (2009) Gauge fields, ripples and wrinkles in graphene layers. *Solid State Commun.,* 149**,** 1140.

Halder, S., Bansal, M. & Bhattacharyya, D. *(Manuscript under preparion)*.

Halder, S. & Bhattacharyya, D. (2010) Structural Stability of Tandemly Occurring Noncanonical Basepairs within Double Helical Fragments: Molecular Dynamics Studies of Functional RNA. *J. Phys. Chem. B,* 114**,** 14028.

Han, G., Ghosh, P. & Rotello, V. M. (2007a) Functionalized gold nanoparticles for drug delivery. *Nanomedicine,* 2**,** 113.

Han, M. Y., Özyilmaz, B., Zhang, Y. & Kim, P. (2007b) Energy band-gap engineering of graphene nanoribbons. *Phys. Rev. Lett.,* 98**,** 206805.

Harkins, W. D., Mattoon, R. W. & Corrin, M. L. (1946) Structure of soap micelles as indicated by X-rays and interpreted by the theory of molecular orientation* 1:: II. The solubilization of hydrocarbons and other oils in aqueous soap solutions. *J. Colloid Sci.,* 1**,** 105.

Hay, P. J. & Wadt, W. R. (1985) Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg. *J. Chem. Phys.,* 82**,** 270.

Heller, D. A., Baik, S., Eurell, T. E. & Strano, M. S. (2005) Single-Walled Carbon Nanotube Spectroscopy in Live Cells: Towards Long-Term Labels and Optical Sensors. *Adv. Mater.,* 17**,** 2793.

Hendrix, D. K., Brenner, S. E. & Holbrook, S. R. (2005) RNA structural motifs: building blocks of a modular biomolecule. *Q. Rev. Biophys.,* 38**,** 221.

Hermann, T. & Westhof, E. (1999) Non-Watson-Crick base pairs in RNA-protein recognition. *Chem Biol.,* 6**,** 335.

Herminghaus, S., Brinkmann, M. & Seemann, R. (2008) Wetting and dewetting of complex surface geometries. *Annu. Rev. Mater. Res.,* 38**,** 101.

Hobza, P., Kabeláč, M., Sponer, J., Mejzlík, P. & Vondrášek, J. (1997) Performance of empirical potentials (AMBER, CFF95, CVFF, CHARMM, OPLS, POLTEV), semiempirical quantum chemical methods (AM1, MNDO/M, PM3), and ab initio

Hartree–Fock method for interaction of DNA bases: Comparison with nonempirical beyond Hartree–Fock results. *J. Comput. Chem.,* 18**,** 1136.

Hobza, P. & Sponer, J. (1999) Structure, energetics, and dynamics of the nucleic acid base pairs: nonempirical ab initio calculations. *Chem. Rev.,* 99**,** 3247.

Hohenberg, P. (1967) Existence of long-range order in one and two dimensions. *Phys. Rev.,* 158**,** 383.

Hohenberg, P. & Kohn, W. (1964) Inhomogeneous electron gas. *Phys. Rev.,* 136**,** B864.

Holbrook, S. R. (2005) RNA structure: the long and the short of it. *Curr. Opin. Struct. Biol.,* 15**,** 302.

Holliday, R. (1964) A mechanism for gene conversion in fungi. *Genet. Res,* 5**,** 282.

Hosoya, H., Gao, Y., Nakada, K., Ohuchi, M. & Tsuruta, T. (1993) in New Functionality Materials edited by Tsuruta, C.T.; Doyama, M. and Seno, M., Elsevier, New York. 27.

Huang, X., Jain, P. K., El-Sayed, I. H. & El-Sayed, M. A. (2008) Plasmonic photothermal therapy (PPTT) using gold nanoparticles. *Lasers Med Sci.,* 23**,** 217.

Huang, Y. Z., Zang, M., Xiong, W. C., Luo, Z. & Mei, L. (2003) Erbin suppresses the MAP kinase pathway. *J. Biol. Chem.,* 278**,** 1108.

Hunenberger, P. H., Borjesson, U. & Lins, R. D. (2001) Electrostatic interactions in biomolecular systems. *CHIMIA* 55**,** 861.

Jadzinsky, P. D., Calero, G., Ackerson, C. J., Bushnell, D. A. & Kornberg, R. D. (2007) Structure of a thiol monolayer-protected gold nanoparticle at 1.1 Å resolution. *Science,* 318**,** 430.

Jain, A., Ramanathan, V. & Sankararamakrishnan, R. (2009) Lone pair$\cdots\pi$ interactions between water oxygens and aromatic residues: Quantum chemical studies based on high-resolution protein structures and model compounds. *Protein Sci.,* 18**,** 595.

Jain, A., Wang, G. & Vasquez, K. M. (2008) DNA triple helices: biological consequences and therapeutic potential. *Biochimie,* 90**,** 1117.

Jayaram, B., Latha, N., Jain, T., Sharma, P., Gandhimathi, A. & Pandey, V. S. (2006) Sanjeevini: A comprehensive active site directed lead design software. *Indian J. of Chem. Sect. A: Inorg, Bio-inorg, Phys, Theor, Anal Chem,* 45**,** 21.

Jeffrey, G. A. (1997) *An introduction to hydrogen bonding,* Oxford University Press New York.

Jensen, F. (2007) *Introduction to computational chemistry,* Wiley.

Jiang, L. & Lai, L. (2002) CH$\cdots$ O hydrogen bonds at protein-protein interfaces. *J. Biol. Chem.,* 277**,** 37732.

Joyce, G. F. (1989) RNA evolution and the origins of life. *Nature,* 338**,** 217.

Jucker, F. M. & Pardi, A. (1995) GNRA tetraloops make a U-turn. *RNA,* 1**,** 219.

Jurečka, P., Sponer, J., Černý, J. & Hobza, P. (2006) Benchmark database of accurate (MP2 and CCSD (T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.,* 8**,** 1985.

Kabelác, M., Kratochvíl, M., Sponer, J. & Hobza, P. (2000) Structure, energetics, vibrational frequencies and charge transfer of base pairs, nucleoside pairs, nucleotide pairs and B-DNA pairs of trinucleotides: ab initio HF/MINI-1 and empirical force field study. *J. Biomol. Struct. Dyn.,* 17**,** 1077.

Kam, N. W. S., Liu, Z. & Dai, H. (2006) Carbon nanotubes as intracellular transporters for proteins and DNA: an investigation of the uptake mechanism and pathway. *Angew. Chem.,* 118**,** 591.

Karow, J. K., Constantinou, A., Li, J. L., West, S. C. & Hickson, I. D. (2000) The Bloom's syndrome gene product promotes branch migration of Holliday junctions. *Proc. Natl. Acad. Sci. USA,* 97**,** 6504.

Kendall, R. A., Dunning Jr, T. H. & Harrison, R. J. (1992) Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.,* 96**,** 6796.

Keniry, M. A. (2000) Quadruplex structures in nucleic acids. *Biopolymers,* 56**,** 123.

Kitaura, K. & Morokuma, K. (1976) A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation. *Int. J. Quantum Chem,* 10**,** 325.

Klosterman, P. S., Tamura, M., Holbrook, S. R. & Brenner, S. E. (2002) SCOR: a structural classification of RNA database. *Nucleic Acids Res.,* 30**,** 392.

Knauert, M. P. & Glazer, P. M. (2001) Triplex forming oligonucleotides: sequence-specific tools for gene targeting. *Hum. Mol. Genet.,* 10**,** 2243.

Kobayashi, Y., Fukui, K., Enoki, T. & Kusakabe, K. (2006) Edge state on hydrogen-terminated graphite edges investigated by scanning tunneling microscopy. *Phys. Rev. B,* 73**,** 125415.

Kobayashi, Y., Fukui, K., Enoki, T., Kusakabe, K. & Kaburagi, Y. (2005) Observation of zigzag and armchair edges of graphite using scanning tunneling microscopy and spectroscopy. *Phys. Rev. B,* 71**,** 193406.

Kohn, W. (1999) Nobel Lecture: Electronic structure of matter-wave functions and density functionals. *Rev. Mod. Phys. ,* 71**,** 1253.

Kohn, W. & Sham, L. J. (1965) Self-consistent equations including exchange and correlation effects. *Phys. Rev,* 140**,** A1133.

Koskinen, P., Malola, S. & Häkkinen, H. (2008) Self-passivating edge reconstructions of graphene. *Phys. Rev. Lett.,* 101**,** 115502.

Kostarelos, K., Lacerda, L., Pastorin, G., Wu, W., Wieckowski, S., Luangsivilay, J., Godefroy, S., Pantarotto, D., Briand, J. P. & Muller, S. (2007) Cellular uptake of functionalized carbon nanotubes is independent of functional group and cell type. *Nat. Nanotechnol.,* 2**,** 108.

Kristyán, S. & Pulay, P. (1994) Can (semi) local density functional theory account for the London dispersion forces? *Chem. Phys. Lett.,* 229**,** 175.

Kryachko, E. & Remacle, F. (2005) Complexes of DNA bases and Watson-Crick base pairs with small neutral gold clusters. *J. Phys. Chem. B,* 109**,** 22746.

Kurita, N., Danilov, V. I. & Anisimov, V. M. (2005) The structure of Watson-Crick DNA base pairs obtained by MP2 optimization. *Chem. Phys. Lett.,* 404**,** 164.

Lavery, R. & Sklenar, H. (1988) The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.,* 6**,** 63.

Le Doan, T., Perrouault, L., Praseuth, D., Habhoub, N., Decout, J. L., Thuong, N. T., Lhomme, J. & Héène, C. (1987) Sequence-specific recognition, photocrosslinking and cleavage of the DNA double helix by an oligo-(α]-thymidylate covalently linked to an azidoproflavine derivative. *Nucleic Acids Res.,* 15**,** 7749.

Leach, A. R. (1996) *Molecular modelling,* Longman Singapore.

Lee, C., Wei, X., Kysar, J. W. & Hone, J. (2008) Measurement of the elastic properties and intrinsic strength of monolayer graphene. *Science,* 321**,** 385.
Lee, C., Yang, W. & Parr, R. G. (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B,* 37**,** 785.
Leenaerts, O., Partoens, B. & Peeters, F. (2009) Water on graphene: Hydrophobicity and dipole moment using density functional theory. *Phys. Rev. B,* 79**,** 235440.
Lemieux, S. & Major, F. (2002) RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire. *Nucleic Acids Res.,* 30**,** 4250.
Leontis, N. B., Stombaugh, J. & Westhof, E. (2002) The non-Watson–Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.,* 30**,** 3497.
Leontis, N. B. & Westhof, E. (1998) Conserved geometrical base-pairing patterns in RNA. *Q. Rev. Biophys.,* 31**,** 399.
Leontis, N. B. & Westhof, E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA,* 7**,** 499.
Leontis, N. B. & Westhof, E. (2003) Analysis of RNA motifs. *Curr. Opin. Struct. Biol.,* 13**,** 300.
Lewars, E. G. (2011) *Computational chemistry: introduction to the theory and applications of molecular and quantum mechanics*, Springer Verlag.
Li, H., Li, W. X. & Ding, S. W. (2002) Induction and suppression of RNA silencing by an animal virus. *Science,* 296**,** 1319.
Lipps, H. J. & Rhodes, D. (2009) G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.,* 19**,** 414.
Liu, Z., Robinson, J. T., Sun, X. & Dai, H. (2008) PEGylated nanographene oxide for delivery of water-insoluble cancer drugs. *J. Am. Chem. Soc.,* 130**,** 10876.
Liu, Z., Robinson, J. T., Tabakman, S. M., Yang, K. & Dai, H. (2011) Carbon materials for drug delivery & cancer therapy. *Mater. Today,* 14**,** 316.
Liu, Z., Tabakman, S., Welsher, K. & Dai, H. (2009) Carbon nanotubes in biology and medicine: in vitro and in vivo detection, imaging and drug delivery. *Nano Res.,* 2**,** 85.
Liu, Z., Winters, M., Holodniy, M. & Dai, H. (2007) siRNA Delivery into Human T Cells and Primary Cells with Carbon-Nanotube Transporters. *Angew. Chem. Int. Ed.,* 46**,** 2023.
Lu, W., Singh, A. K., Khan, S. A., Senapati, D., Yu, H. & Ray, P. C. (2010a) Gold nano-popcorn-based targeted diagnosis, nanotherapy treatment, and in situ monitoring of photothermal therapy response of prostate cancer cells using surface-enhanced Raman spectroscopy. *J. Am. Chem. Soc.,* 132**,** 18103.
Lu, X. J., El Hassan, M. & Hunter, C. (1997) Structure and conformation of helical nucleic acids: analysis program (SCHNAaP) *J. Mol. Biol.,* 273**,** 668.
Lu, X. J. & Olson, W. K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.,* 31**,** 5108.
Lu, X. J. & Olson, W. K. (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nature protocols,* 3**,** 1213.

Lu, X. J., Olson, W. K. & Bussemaker, H. J. (2010b) The RNA backbone plays a crucial role in mediating the intrinsic stability of the GpU dinucleotide platform and the GpUpA/GpA miniduplex. *Nucleic Acids Res.,* 38**,** 4868.

Lung, J. K., Huang, J. C., Tien, D. C., Liao, C. Y., Tseng, K. H., Tsung, T. T., Kao, W. S., Tsai, T. H., Jwo, C. S. & Lin, H. M. (2007) Preparation of gold nanoparticles by arc discharge in water. *J. Alloys Compd.,* 434**,** 655.

Mallajosyula, S. S., Datta, A. & Pati, S. K. (2005) Structure and electronic properties of the Watson-Crick base pairs: Role of hydrogen bonding. *Synth. Met.,* 155**,** 398.

Mandal, M. & Breaker, R. R. (2004) Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.,* 5**,** 451.

Mangang, S. U. & Lyngdoh, R. (2001) Wobble base-pairing in codon-anticodon interactions: a theoretical modelling study. *Indian J. Biochem. Biophys.,* 38**,** 115.

Maréchal, Y. (2007) *The hydrogen bond and the water molecule: the physics and chemistry of water, aqueous and bio media*, Elsevier Science.

Medarova, Z., Pham, W., Farrar, C., Petkova, V. & Moore, A. (2007) In vivo imaging of siRNA delivery and silencing in tumors. *Nat. Med.,* 13**,** 372.

Mermin, N. D. & Wagner, H. (1966) Absence of ferromagnetism or antiferromagnetism in one-or two-dimensional isotropic Heisenberg models. *Phys. Rev. Lett.,* 17**,** 1133.

Meyer, J. C., Geim, A., Katsnelson, M., Novoselov, K., Booth, T. & Roth, S. (2007) The structure of suspended graphene sheets. *Nature,* 446**,** 60.

Meyer, M., Steinke, T., Brandl, M. & Sühnel, J. (2001) Density functional study of guanine and uracil quartets and of guanine quartet/metal ion complexes. *J. Comput. Chem.,* 22**,** 109.

Mládek, A., Sharma, P., Mitra, A., Bhattacharyya, D., Sponer, J. & Sponer, J. E. (2009) Trans Hoogsteen/Sugar Edge Base Pairing in RNA. Structures, Energies, and Stabilities from Quantum Chemical Calculations. *J. Phys. Chem. B,* 113**,** 1743.

Mládek, A., Sponer, J. E., Kulhánek, P., Lu, X. J., Olson, W. K. & Sponer, J. (2012) Understanding the Sequence Preference of Recurrent RNA Building Blocks using Quantum Chemistry: The Intrastrand RNA Dinucleotide Platform. *J. Chem. Theory Comput.,* 8**,** 335.

Mo, Y. (2006) Probing the nature of hydrogen bonds in DNA base pairs. *J. Mol. Model.,* 12**,** 665.

Mohan, P., Datta, A., Mallajosyula, S. S. & Pati, S. K. (2006) Structures of nucleobases trapped within Au triangles and its effects on hydrogen bonding in base pairs of DNA. *J. Phys. Chem. B,* 110**,** 18661.

Mohanty, N. & Berry, V. (2008) Graphene-based single-bacterium resolution biodevice and DNA transistor: interfacing graphene derivatives with nanoscale and microscale biocomponents. *Nano Lett.,* 8**,** 4469.

Mokdad, A., Krasovska, M. V., Sponer, J. & Leontis, N. B. (2006) Structural and evolutionary classification of G/U wobble basepairs in the ribosome. *Nucleic Acids Res.,* 34**,** 1326.

Møller, C. & Plesset, M. S. (1934) Note on an approximation treatment for many-electron systems. *Phys. Rev.,* 46**,** 618.

Morris, M., Vidal, P., Chaloin, L., Heitz, F. & Divita, G. (1997) A new peptide vector for efficient delivery of oligonucleotides into mammalian cells. *Nucleic Acids Res.,* 25**,** 2730.

Mukherjee, S., Bansal, M. & Bhattacharyya, D. (2006) Conformational specificity of non-canonical base pairs and higher order structures in nucleic acids: crystal structure database analysis. *J. Comput.-Aided Mol. Des.,* 20**,** 629.

Mukherjee, S., Majumdar, S. & Bhattacharyya, D. (2005) Role of hydrogen bonds in protein-DNA recognition: Effect of nonplanar amino groups. *J. Phys. Chem. B,* 109**,** 10484.

Müller-Dethlefs, K. & Hobza, P. (2000) Noncovalent interactions: A challenge for experiment and theory. *Chem. Rev.,* 100**,** 143.

Murray, C., Norris, D. & Bawendi, M. G. (1993) Synthesis and characterization of nearly monodisperse CdE (E= sulfur, selenium, tellurium) semiconductor nanocrystallites. *J. Am. Chem. Soc.,* 115**,** 8706.

Murthy, V. L. & Rose, G. D. (2003) RNABase: an annotated database of RNA structures. *Nucleic Acids Res.,* 31**,** 502.

Nagaswamy, U., Larios-Sanz, M., Hury, J., Collins, S., Zhang, Z., Zhao, Q. & Fox, G. E. (2002) NCIR: a database of non-canonical interactions in known RNA structures. *Nucleic Acids Res.,* 30**,** 395.

Nakada, K., Fujita, M., Dresselhaus, G. & Dresselhaus, M. S. (1996) Edge state in graphene ribbons: Nanometer size effect and edge shape dependence. *Phys. Rev. B,* 54**,** 17954.

Neidle, S. (2007) *Principles of nucleic acid structure,* Academic Press.

Niimi, Y., Matsui, T., Kambara, H., Tagami, K., Tsukada, M. & Fukuyama, H. (2005) Scanning tunneling microscopy and spectroscopy studies of graphite edges. *Appl. Surf. Sci.,* 241**,** 43.

Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B. & Steitz, T. A. (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl. Acad. Sci. USA,* 98**,** 4899.

Novina, C. D., Murray, M. F., Dykxhoorn, D. M., Beresford, P. J., Riess, J., Lee, S. K., Collman, R. G., Lieberman, J., Shankar, P. & Sharp, P. A. (2002) siRNA-directed inhibition of HIV-1 infection. *Nat. Med.,* 8**,** 681.

Novoselov, K., Geim, A., Morozov, S., Jiang, D., Grigorieva, M. I. K. I. V., Dubonos, S. & Firsov, A. (2005) Two-dimensional gas of massless Dirac fermions in graphene. *Nature,* 438**,** 197.

Novoselov, K., Mccann, E., Morozov, S., Fal'ko, V. I., Katsnelson, M., Zeitler, U., Jiang, D., Schedin, F. & Geim, A. (2006) Unconventional quantum Hall effect and Berry's phase of 2π in bilayer graphene. *Nat. Phys.,* 2**,** 177.

Oberley, L. W. (1988) Free radicals and diabetes. *Free Radical Biol. Med.,* 5**,** 113.

Ogawa, T., Kurita, N., Sekino, H., Kitao, O. & Tanaka, S. (2003) Hydrogen bonding of DNA base pairs by consistent charge equilibration method combined with universal force field. *Chem. Phys. Lett.,* 374**,** 271.

Olson, W. K., Bansal, M., Burley, S. K., Dickerson, R. E., Gerstein, M., Harvey, S. C., Heinemann, U., Lu, X. J., Neidle, S. & Shakked, Z. (2001) A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.,* 313**,** 229.

Ortiz-Lombardía, M., González, A., Eritja, R., Aymamí, J., Azorín, F. & Coll, M. (1999) Crystal structure of a DNA Holliday junction. *Nat. Struct. Biol.,* 6**,** 913.

Ottiger, P., Pfaffen, C., Leist, R., Leutwyler, S., Bachorz, R. A. & Klopper, W. (2009) Strong N− H··· π Hydrogen Bonding in Amide− Benzene Interactions. *J. Phys. Chem. B,* 113**,** 2937.

Paciotti, G. F., Kingston, D. G. I. & Tamarkin, L. (2006) Colloidal gold nanoparticles: a novel nanoparticle platform for developing multifunctional tumor-targeted drug delivery vectors. *Drug Dev. Res.,* 67**,** 47.

Pal, R. & Chakraborti, A. S. (2010) Preparation Of Gold Nanoparticle Quercetin Complexes By Citrate Reduction Method. *ICANN-2009, AIP Conference Proceedings, 1276, 283.*

Pal, R., Panigrahi, S., Bhattacharyya, D. & Chakraborti, A. S. Characterization of citrate capped gold nanoparticle-quercetin complex: experimental and quantum chemical approach. *(Manuscript under preparion)*.

Panigrahi, S., Banerjee, S., Bhattacharya, A. & Bhattacharyya, D. (2012) Interaction of Nucleobases with Wrinkled Graphene Surface: Dispersion Corrected DFT and AFM Studies. *J. Phys. Chem. C,* 116**,** 4374.

Panigrahi, S., Bhattacharya, A., Bandyopadhyay, D., Grabowski, S. J., Bhattacharyya, D. & Banerjee, S. (2011a) Wetting Property of the Edges of Monoatomic Step on Graphite: Frictional-Force Microscopy and ab initio Quantum Chemical Studies. *J. Phys. Chem. C,* 115**,** 14819.

Panigrahi, S., Pal, R. & Bhattacharyya, D. (2011b) Structure and energy of non-canonical basepairs: comparison of various computational chemistry methods with crystallographic ensembles. *J. Biomol. Struct. Dyn.,* 29**,** 541.

Pantarotto, D., Partidos, C. D., Hoebeke, J., Brown, F., Kramer, E., Briand, J. P., Muller, S., Prato, M. & Bianco, A. (2003) Immunization with peptide-functionalized carbon nanotubes enhances virus-specific neutralizing antibody responses. *Chem. Biol.,* 10**,** 961.

Paton, R. S. & Goodman, J. M. (2009) Hydrogen Bonding and π-Stacking: How Reliable are Force Fields? A Critical Evaluation of Force Field Descriptions of Nonbonded Interactions. *J. Chem. Inf. Model. ,* 49**,** 944.

Perdew, J. P. (1986) Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B,* 33**,** 8822.

Perdew, J. P., Burke, K. & Ernzerhof, M. (1996) Generalized gradient approximation made simple. *Phys. Rev. Lett.,* 77**,** 3865.

Perdew, J. P., Chevary, J., Vosko, S., Jackson, K. A., Pederson, M. R., Singh, D. & Fiolhais, C. (1992) Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B,* 46**,** 6671.

Perdew, J. P. & Wang, Y. (1992) Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B,* 45**,** 13244.

Perdew, J. P. & Yue, W. (1986) Accurate and simple density functional for the electronic exchange energy: Generalized gradient approximation. *Phys. Rev. B,* 33**,** 8800.

Pereira, V. M., Castro Neto, A., Liang, H. & Mahadevan, L. (2010) Geometry, Mechanics, and Electronics of Singular Structures and Wrinkles in Graphene. *Phys. Rev. Lett.,* 105**,** 156603.

Peterson, K. A., Kendall, R. A. & Dunning Jr, T. H. (1993) Benchmark calculations with correlated molecular wave functions. II. Configuration interaction calculations on first row diatomic hydrides. *J. Chem. Phys.,* 99**,** 1930.

Petkov, V., Peng, Y., Williams, G., Huang, B., Tomalia, D. & Ren, Y. (2005) Structure of gold nanoparticles suspended in water studied by x-ray diffraction and computer simulations. *Phys. Rev. B,* 72**,** 195402.

Pimentel, G. C. & Mcclellan, A. L. (1960) The hydrogen bond. *WH Freeman San Francisco , 475*.

Plasterk, R. H. A. (2002) RNA silencing: the genome's immune system. *Science,* 296**,** 1263.

Pliego Jr, J. R., De Almeida, W. B., Celebi, S., Zhu, Z. & Platz, M. S. (1999) Singlet-Triplet Gap, and the Electronic and Vibrational Spectra of Chlorophenylcarbene: A Combined Theoretical and Experimental Study. *J. Phys. Chem. A,* 103**,** 7481.

Praseuth, D., Guieysse, A. & Helene, C. (1999) Triple helix formation and the antigene strategy for sequence-specific control of gene expression. *Biochim Biophys Acta.,* 1489**,** 181.

Pyykko, P. (1988) Relativistic effects in structural chemistry. *Chem. Rev.,* 88**,** 563.

Pyykko, P. (1997) Strong closed-shell interactions in inorganic chemistry. *Chem. Rev.,* 97**,** 597.

Rajesh, C., Majumder, C., Mizuseki, H. & Kawazoe, Y. (2009) A theoretical study on the interaction of aromatic amino acids with graphene and single walled carbon nanotube. *J. Chem. Phys.,* 130**,** 124911.

Ramstein, J. & Lavery, R. (1988) Energetic coupling between DNA bending and base pair opening. *Proc. Natl. Acad. Sci. USA,* 85**,** 7231.

Ravishanker, G., Swaminathan, S., Beveridge, D., Lavery, R. & Sklenar, H. (1989) Conformational and helicoidal analysis of 30 PS of molecular dynamics on the d (CGCGAATTCGCG) double helix:" curves", dials and windows. *J. Biomol. Struct. Dyn.,* 6**,** 669.

Ray, S. S., Halder, S. & Bhattacharyya, D. (2009) HD-RNAS: Hierarchical Database of RNA Structures. *Proceedings of International Conference on Frontiers of Interface Between Statistics and Sciences***,** 724.

Ray, S. S., Halder, S., Kaypee, S. & Bhattacharyya, D. (2012) HD-RNAS: an automated hierarchical database of RNA structures. *Front Genet.,* 3.

Read, M. A., Wood, A. A., Harrison, J. R., Gowan, S. M., Kelland, L. R., Dosanjh, H. S. & Neidle, S. (1999) Molecular modeling studies on G-quadruplex complexes of telomerase inhibitors: structure-activity relationships. *J. Med. Chem.,* 42**,** 4538.

Reed, A. E., Curtiss, L. A. & Weinhold, F. (1988) Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chem. Rev.,* 88**,** 899.

Reed, A. E., Weinstock, R. B. & Weinhold, F. (1985) Natural population analysis. *J. Chem. Phys.,* 83**,** 735.

Rhee, S., Han, Z., Liu, K., Miles, H. T. & Davies, D. R. (1999) Structure of a triple helical DNA with a triplex-duplex junction. *Biochemistry (Mosc).* 38**,** 16810.

Riley, K. E., Pitoňák, M., Černý, J. & Hobza, P. (2009) On the structure and geometry of biomolecular binding motifs (hydrogen-bonding, stacking, X− H··· π): WFT and DFT calculations. *J. Chem. Theory Comput.,* 6**,** 66.

Roberts, R. W. & Crothers, D. M. (1992) Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. *Science,* 258**,** 1463.

Rosas, J. J. H., Gutiérrez, R. E. R., Escobedo-Morales, A. & Anota, E. C. (2011) First principles calculations of the electronic and chemical properties of graphene, graphane, and graphene oxide. *J Mol Model,* 17**,** 1133.

Roy, A., Panigrahi, S., Bhattacharyya, M. & Bhattacharyya, D. (2008) Structure, stability, and dynamics of canonical and noncanonical base pairs: quantum chemical studies. *J. Phys. Chem. B,* 112**,** 3786.

Rozas, I., Alkorta, I. & Elguero, J. (2000) Behavior of ylides containing N, O, and C atoms as hydrogen bond acceptors. *J. Am. Chem. Soc.,* 122**,** 11154.

Rubeš, M., Nachtigall, P., Vondrásek, J. & Bludský, O. (2009) Structure and Stability of the Water− Graphite Complexes. *J. Phys. Chem. C,* 113**,** 8412.

Ryu, S. & Hatsugai, Y. (2004) Zero-energy edge states and chiral symmetry breaking at edges of graphite sheets. *Physica E (Amsterdam),* 22**,** 679.

Saenger, W. (1984) *Principles of nucleic acid structure*, Springer-Verlag New York.

Sankaran, M. & Viswanathan, B. (2006) A DFT study of the electronic property of gold nanoclusters (Au_x, x = 1-12 atoms) *Bulletin of the Catalysis Society of India,* 5**,** 26.

Santamaria, R., Charro, E., Zacarias, A. & Castro, M. (1999) Vibrational spectra of nucleic acid bases and their Watson–Crick pair complexes. *J. Comput. Chem.,* 20**,** 511.

Santosh, M., Panigrahi, S., Bhattacharyya, D., Sood, A. & Maiti, P. (2012) Unzipping and binding of siRNA with single walled Carbon Nanotube: a platform for siRNA delivery. *J. Chem. Phys.,* 136**,** 065106.

Santosh, M., Panigrahi, S., Sood, A. K., Bhattacharyya, D. & Maiti, P. K. siRNA Unzipping on Graphene. *(Submitted)*.

Sarkar, S., Patra, S., Gayathri, N. & Banerjee, S. (2010) Effect of self-affine fractal characteristics of surfaces on wetting. *Appl. Phys. Lett.,* 96**,** 063112.

Sarver, M., Zirbel, C. L., Stombaugh, J., Mokdad, A. & Leontis, N. B. (2008) FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *J. Math. Biol.,* 56**,** 215.

Scaria, P. V., Will, S., Levenson, C. & Shafer, R. H. (1995) Physicochemical Studies of the d (G3T4G3)* d (G3A4G3)• d (C3T4C3) Triple Helix. *J. Biol. Chem.,* 270**,** 7295.

Schaftenaar, G. & Noordik, J. H. (2000) Molden: a pre-and post-processing program for molecular and electronic structures*. *J. Comput-Aided Mol. Des.,* 14**,** 123.

Scheiner, S., Kar, T. & Gu, Y. (2001) Strength of the $C^{\alpha}H\cdot\cdot$ O hydrogen bond of amino acid residues. *J. Biol. Chem.,* 276**,** 9832.

Schmidt, M. W., Baldridge, K. K., Boatz, J. A., Elbert, S. T., Gordon, M. S., Jensen, J. H., Koseki, S., Matsunaga, N., Nguyen, K. A. & Su, S. (1993) General atomic and molecular electronic structure system. *J. Comput. Chem.,* 14**,** 1347.

Seibert, E., Ross, J. & Osman, R. (2003) Contribution of opening and bending dynamics to specific recognition of DNA damage. *J. Mol. Biol.,* 330**,** 687.

Selvakannan, P., Mandal, S., Phadtare, S., Pasricha, R. & Sastry, M. (2003) Capping of gold nanoparticles by the amino acid lysine renders them water-dispersible. *Langmuir,* 19**,** 3545.

Sen, D. & Gilbert, W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature,* 334**,** 364

Sen, K., Basu, S. & Bhattacharyya, D. (2006) Ab initio studies on excited state intramolecular electron transfer in 4-amino-N-methylphthalimide and 3-amino-N-methylphthalimide. *Int. J. Quantum Chem,* 106**,** 913.

Shao, Y., Wang, J., Wu, H., Liu, J., Aksay, I. A. & Lin, Y. (2010) Graphene based electrochemical sensors and biosensors: a review. *Electroanalysis,* 22**,** 1027.

Sharma, P., Chawla, M., Sharma, S. & Mitra, A. (2010a) On the role of Hoogsteen: Hoogsteen interactions in RNA: Ab initio investigations of structures and energies. *RNA,* 16**,** 942.

Sharma, P., Mitra, A., Sharma, S., Singh, H. & Bhattacharyya, D. (2008) Quantum chemical studies of structures and binding in noncanonical RNA base pairs: the trans Watson-Crick: Watson-Crick family. *J. Biomol. Struct. Dyn.,* 25**,** 709.

Sharma, P., Singh, H. & Sharma, S. (2007) Binding of Gold Nanoclusters with Size-Expanded DNA Bases: A Computational Study of Structural and Electronic Properties. *J. Chem. Theory Comput. ,* 3**,** 2301.

Sharma, P., Sponer, J. E., Sponer, J., Sharma, S., Bhattacharyya, D. & Mitra, A. (2010b) On the Role of the cis Hoogsteen: Sugar-Edge Family of Base Pairs in Platforms and Triplets-Quantum Chemical Insights into RNA Structural Biology. *J. Phys. Chem. B,* 114**,** 3307.

Sheina, G., Stepanian, S., Radchenko, E. & Blagoi, Y. P. (1987) IR spectra of guanine and hypoxanthine isolated molecules. *J. Mol. Struct.,* 158**,** 275.

Shetty, A., Rashmi, R., Rajan, M., Sambaiah, K. & Salimath, P. (2004) Antidiabetic influence of quercetin in streptozotocin-induced diabetic rats. *Nutrition Research,* 24**,** 373.

Shukla, M. K., Dubey, M., Zakar, E. & Leszczynski, J. (2009) DFT Investigation of the Interaction of Gold Nanoclusters with Nucleic Acid Base Guanine and the Watson- Crick Guanine-Cytosine Base Pair. *J. Phys. Chem. C,* 113**,** 3960.

Simeoni, F., Morris, M. C., Heitz, F. & Divita, G. (2003) Insight into the mechanism of the peptide-based gene delivery system MPG: implications for delivery of siRNA into mammalian cells. *Nucleic Acids Res.,* 31**,** 2717.

Singha, S., Bhattacharya, J., Datta, H. & Dasgupta, A. K. (2009) Anti-glycation activity of gold nanoparticles. *Nanomed. Nanotechnol. Biol. Med.,* 5**,** 21.

Southall, N. T., Dill, K. A. & Haymet, A. (2002) A view of the hydrophobic effect. *J. Phys. Chem. B,* 106**,** 521.

Sponer, J. (2006) *Computational Studies of RNA and DNA*, Springer Verlag.

Sponer, J., Florián, J., Hobza, P. & Leszczynski, J. (1996) Nonplanar DNA base pairs. *J. Biomol. Struct. Dyn.,* 13**,** 827.

Sponer, J. & Hobza, P. (2003) Molecular interactions of nucleic acid bases. A review of quantum-chemical studies. *Collect. Czech. Chem. Commun.,* 68**,** 2231.

Sponer, J., Jurečka, P. & Hobza, P. (2004) Accurate interaction energies of hydrogen-bonded nucleic acid base pairs. *J. Am. Chem. Soc.,* 126**,** 10142.

Sponer, J., Mokdad, A., Sponer, J. E., Spackova, N., Leszczynski, J. & Leontis, N. B. (2003) Unique tertiary and neighbor interactions determine conservation patterns of cis Watson-Crick A/G base-pairs. *J. Mol. Biol.,* 330**,** 967.

Sponer, J., Sponer, J. E., Petrov, A. I. & Leontis, N. B. (2010) Quantum Chemical Studies of Nucleic Acids: Can We Construct a Bridge to the RNA Structural Biology and Bioinformatics Communities? *J. Phys. Chem. B,* 114**,** 15723.

Sponer, J. E., Leszczynski, J., Sychrovský, V. & Sponer, J. (2005a) Sugar edge/sugar edge base pairs in RNA: stabilities and structures from quantum chemical calculations. *J. Phys. Chem. B,* 109**,** 18680.

Sponer, J. E., Spackova, N., Kulhánek, P., Leszczynski, J. & Sponer, J. (2005b) Non-Watson-Crick base pairing in RNA. quantum chemical analysis of the cis Watson-Crick/sugar edge base pair family. *J. Phys. Chem. A,* 109**,** 2292.

Sponer, J. E., Spackova, N., Leszczynski, J. & Sponer, J. (2005c) Principles of RNA base pairing: structures and energies of the trans Watson-Crick/sugar edge base pairs. *J. Phys. Chem. B,* 109**,** 11399.

Sponer, J. E., Réblová, K., Mokdad, A., Sychrovský, V., Leszczynski, J. & Sponer, J. (2007) Leading RNA tertiary interactions: structures, energies, and water insertion of A-minor and P-interactions. A quantum chemical view. *J. Phys. Chem. B,* 111**,** 9153.

Starikov, E. B. & Steiner, T. (1997) Computational support for the suggested contribution of C-H…O=C interactions to the stability of nucleic acid base pairs. *Acta Crystallogr. Sect. D. Biol. Crystallogr.,* 53**,** 345.

Stein, S. E. & Brown, R. (1987). pi.-Electron properties of large condensed polyaromatic hydrocarbons. *J. Am. Chem. Soc.,* 109**,** 3721.

Steitz, T. A. & Moore, P. B. (2003) RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem. Sci.,* 28**,** 411.

Stepanian, S., Sheina, G., Radchenko, E. & Blagoi, Y. P. (1985) Theoretical and experimental studies of adenine, purine and pyrimidine isolated molecule structure. *J. Mol. Struct.,* 131**,** 333.

Stewart, J. J. P. (1989) Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.,* 10**,** 209.

Stojanovi , S. Đ. & Zari , S. D. (2009) Hydrogen Bonds and Hydrophobic Interactions of Porphyrins in Porphyrin-Containing Proteins. *Open Struct Biol J,* 3**,** 34.

Stombaugh, J., Zirbel, C. L., Westhof, E. & Leontis, N. B. (2009) Frequency and isostericity of RNA base pairs. *Nucleic Acids Res.,* 37**,** 2294.

Studio, D. (2007) version 2.0. *Accelrys Inc.: San Diego, CA.*

Sun, X., Liu, Z., Welsher, K., Robinson, J. T., Goodwin, A., Zaric, S. & Dai, H. (2008) Nano-graphene oxide for cellular imaging and drug delivery. *Nano Res.,* 1**,** 203.

Sutor, D. J. (1963) Evidence for the existence of C–H…O hydrogen bonds in crystals. *J. Chem. Soc.***,** 1105.

Tanaka, K., Yamashita, S., Yamabe, H. & Yamabe, T. (1987) Electronic properties of one-dimensional graphite family. *Synth. Met.,* 17**,** 143.

Tanford, C. (1962) Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J. Am. Chem. Soc.,* 84**,** 4240.

Tanford, C. (1987) Organizational consequences of the hydrophobic interaction. *Proc. Indian Acad. Sci. (Chem. Sci.),* 98**,** 343.

Tang, Z., Litvinchuk, A., Lee, H. G. & Guloy, A. M. (1998) Crystal structure and vibrational spectra of a new viologen gold (I) iodide. *Inorg. Chem.,* 37**,** 4752.

Te Velde, G., Bickelhaupt, F. M., Baerends, E. J., Fonseca Guerra, C., Van Gisbergen, S. J. A., Snijders, J. G. & Ziegler, T. (2001) Chemistry with ADF. *J. Comput. Chem.,* 22**,** 931.

Thompson-Flagg, R. C., Moura, M. J. B. & Marder, M. (2009) Rippling of graphene. *EPL (Europhysics Letters),* 85**,** 46002.

Tomari, Y., Matranga, C., Haley, B., Martinez, N. & Zamore, P. D. (2004) A protein sensor for siRNA asymmetry. *Science,* 306**,** 1377.

Tseng, K. H., Liao, C. Y., Huang, J. C., Tien, D. C. & Tsung, T. T. (2008) Characterization of gold nanoparticles in organic or inorganic medium (ethanol/water) fabricated by spark discharge method. *Mater. Lett.,* 62**,** 3341.

Tsubouchi, A., Sakakura, J., Yagi, R., Mazaki, Y., Schaefer, E., Yano, H. & Sabe, H. (2002) Localized suppression of RhoA activity by Tyr31/118-phosphorylated paxillin in cell adhesion and migration. *J. Cell. Biol.,* 159**,** 673.

Umadevi, D. & Sastry, G. N. (2011) Quantum Mechanical Study of Physisorption of Nucleobases on Carbon Materials: Graphene versus Carbon Nanotubes. *J. Phys. Chem. Lett.,* 2**,** 1572.

van Holde, K. E., Johnson, W. C. & Ho, P. S. (2006) Principles of physical biochemistry.

van Lenthe, E., Baerends, E. J. & Snijders, J. G. (1993) Relativistic regular two-component Hamiltonians. *J. Chem. Phys.,* 99**,** 4597.

van Lenthe, E., Baerends, E. J. & Snijders, J. G. (1994) Relativistic total energy using regular approximations. *J. Chem. Phys.,* 101**,** 9783.

van Mourik, T. & Gdanitz, R. J. (2002) A critical note on density functional theory studies on rare-gas dimers. *J. Chem. Phys.,* 116**,** 9620.

van Mourik, T. & van Duijneveldt, F. (1995) Ab initio calculations on the CH... O hydrogen-bonded systems CH4-H2O, CH3NH2-H2O and CH3NH3^{+}-H2O. *J. Mol. Struct. (Theochem.),* 341**,** 63.

Varani, G. & Mcclain, W. H. (2000) The G· U wobble base pair. *EMBO reports,* 1**,** 18.

Vargas, R., Garza, J., Dixon, D. A. & Hay, B. P. (2000) How Strong Is the Cα−H···O=C Hydrogen Bond? *J. Am. Chem. Soc.,* 122**,** 4750.

Varghese, N., Mogera, U., Govindaraj, A., Das, A., Maiti, P. K., Sood, A. K. & Rao, C. (2009) Binding of DNA nucleobases and nucleosides with graphene. *ChemPhysChem,* 10**,** 206.

Varnai, P. & Lavery, R. (2002) Base flipping in DNA: pathways and energetics studied with molecular dynamic simulations. *J. Am. Chem. Soc.,* 124**,** 7272.

Vessal, M., Hemmati, M. & Vasei, M. (2003) Antidiabetic effects of quercetin in streptozocin-induced diabetic rats. *Comp Biochem Physiol C Toxicol Pharmacol.,* 135**,** 357.

Voet, D. & Voet, J. G. (1994) Biochemistry; John Wiley&Sons. *Inc.: New York.*

Wakabayashi, K. (2001) Electronic transport properties of nanographite ribbon junctions. *Phys. Rev. B,* 64**,** 125428.

Wakabayashi, K., Fujita, M., Ajiki, H. & Sigrist, M. (1999) Electronic and magnetic properties of nanographite ribbons. *Phys. Rev. B,* 59**,** 8271.

Wang, J., Wang, G. & Zhao, J. (2002) Density-functional study of Au_n(n= 2–20) clusters: Lowest-energy structures and electronic properties. *Phys. Rev. B,* 66**,** 035418.

Wang, J., Zhou, N., Zhu, Z., Huang, J. & Li, G. (2007) Detection of flavonoids and assay for their antioxidant activity based on enlargement of gold nanoparticles. *Anal. Bioanal. Chem.,* 388**,** 1199.

Wassei, J. K. & Kaner, R. B. (2010) Graphene, a promising transparent conductor. *Mater. Today,* 13**,** 52.

Watson, J., Hays, F. A. & Ho, P. S. (2004) Definitions and analysis of DNA Holliday junction geometry. *Nucleic Acids Res.,* 32**,** 3017.

Wells, D. H. (2004) Formation of hydrogen peroxide from H2 and O2 over a neutral gold trimer: a DFT study. *J. Catal.,* 225**,** 69.

Wells Jr, D. H., Delgass, W. N. & Thomson, K. T. (2002) Density functional theory investigation of gold cluster geometry and gas-phase reactivity with O_2. *J. Chem. Phys.,* 117**,** 10597.

Wieder, M. E., Hone, D. C., Cook, M. J., Handsley, M. M., Gavrilovic, J. & Russell, D. A. (2006) Intracellular photodynamic therapy with photosensitizer-nanoparticle conjugates: cancer therapy using a 'Trojan horse'. *Photochem. Photobiol. Sci.,* 5**,** 727.

Williamson, J. R. (1994) G-quartet structures in telomeric DNA. *Annu. Rev. Biophys. Biomol. Struct.,* 23**,** 703.

Wu, C. H. & Yen, G. C. (2005) Inhibitory effect of naturally occurring flavonoids on the formation of advanced glycation endproducts. *J. Agric. Food Chem.,* 53**,** 3167.

Xin, Y. & Olson, W. K. (2009) BPS: a database of RNA base-pair structures. *Nucleic Acids Res.,* 37**,** D83.

Xu, H. & Dill, K. A. (2005) Water's hydrogen bonds in the hydrophobic effect: A simple model. *J. Phys. Chem. B,* 109**,** 23611.

Yang, H., Fung, S. Y., Pritzker, M. & Chen, P. (2007) Modification of hydrophilic and hydrophobic surfaces using an ionic-complementary peptide. *PloS one,* 2**,** e1325.

Yang, H., Jossinet, F., Leontis, N., Chen, L., Westbrook, J., Berman, H. & Westhof, E. (2003) Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.,* 31**,** 3450.

Yang, K., Zhang, S., Zhang, G., Sun, X., Lee, S. T. & Liu, Z. (2010a) Graphene in mice: ultrahigh in vivo tumor uptake and efficient photothermal therapy. *Nano Lett.,* 10**,** 3318−3323.

Yang, W., Ratinac, K. R., Ringer, S. P., Thordarson, P., Gooding, J. J. & Braet, F. (2010b) Carbon nanomaterials in biosensors: should you use nanotubes or graphene? *Angew. Chem. Int. Ed.,* 49**,** 2114.

Yang, X., Wang, Y., Huang, X., Ma, Y., Huang, Y., Yang, R., Duan, H. & Chen, Y. (2010c) Multi-functionalized graphene oxide based anticancer drug-carrier with dual-targeting function and pH-sensitivity. *J. Mater. Chem.,* 21**,** 3448.

Yang, X., Zhang, X., Liu, Z., Ma, Y., Huang, Y. & Chen, Y. (2008) High-efficiency loading and controlled release of doxorubicin hydrochloride on graphene oxide. *J. Phys. Chem. C,* 112**,** 17554.

Zamore, P. D. (2002) Ancient pathways programmed by small RNAs. *Science,* 296**,** 1265.

Zamore, P. D., Tuschl, T., Sharp, P. A. & Bartel, D. P. (2000) RNAi:: Double-Stranded RNA Directs the ATP-Dependent Cleavage of mRNA at 21 to 23 Nucleotide Intervals. *Cell,* 101**,** 25.

Zhang, L., Lu, Z., Zhao, Q., Huang, J., Shen, H. & Zhang, Z. (2011) Enhanced Chemotherapy Efficacy by Sequential Delivery of siRNA and Anticancer Drugs Using PEI-Grafted Graphene Oxide. *Small,* 7**,** 460.

Zhao, Y. & Truhlar, D. G. (2006) A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys.*, 125**,** 194101.