Charles University in Prague

Faculty of Science

Ph.D study program: Modeling of Chemical Properties of Nano- and Biostructures

M.Sc. Susanta Haldar

Theoretical Study of Non-covalent Interaction from Small Molecules to Biomolecules

Dissertation Thesis

Supervisor:

Prof. Pavel Hobza, DrSc., FRSC, Dr. h. c.

Academy of Sciences of the Czech Republic

Institute of Organic Chemistry and Biochemistry, v.v.i.

Center of Excellence: “Controlling structure and function of biomolecules at the molecular level: theory meets experiment”

Prague, 2015

Univerzita Karlova v Praze

Přírodovědecká fakulta

Doktorský studijní program: Modelování chemických vlastností nano- a biostruktur

Mgr. Susanta Haldar

Teoretická studie nekovalentních interakcí, od malých molekul k biomolekulám

Disertační práce

Školitel:

Prof. Ing. Pavel Hobza, DrSc., FRSC, Dr. h. c.

Akademie věd, Česká Republika

Ústav organické chemie a biochemie

Centrum Excelence: "Řizení struktury a funkce biomolekul na molekulové úrovni: souhra teorie a experimentu"

Praha, 2015

Declaration of Authorship

I, SUSANTA HALDAR, hereby declare that the matter embodied in this dissertation "**Theoretical Study of Non-covalent Interaction from Small Molecules to Biomolecules**" is the result of the work carried out by me. All the literature is properly cited, and it has not been submitted for the award of any degree or diploma or associateship of any university or institute.

Prague, October 2015 (Susanta Haldar)

Candidate

CERTIFIED

Prof. Ing. Pavel Hobza

(Research Supervisor)

Dedicated To

MY BELOVED PARENTS

Acknowledgements

I would like to greet my sincere thanks to my research supervisor Prof. Pavel Hobza, who constantly encouraged me to pursue my research with utmost enthusiasm. His patience, guidance and support have pushed me to do work with great devotion. The regular scientific discussions and meeting have created a disciplined and promising platform for us to learn many things as a part of research training. The frequently birthday celebrations of my lab colleagues and their subsequent treats have created a homely environment to rejoice. It was indeed a very good learning opportunity in his laboratory.

I express my sincere gratitude to all my lab colleagues Dr. Martin Lepšík, Dr. Jindřich Fanfrlík, Dr. Jan Řezáč, Dr. Ramachandran Gnanasekaran, Dr. Róbert Sedlák, Adam Pecina, Dr. Rabindranath Lo, Michaela Nekardová, Jiří Hostaš, Haresh Ajani, Cemal Köprülüoğlu and Saltuk M. Eyrilmez. At last I would also like to pursue special thank to Dr. Michal H. Kolář for his kind support. Learning Molecular Dynamics stuffs with Michal was always fun.

I thank my parents and all family members, friends and well-wishers for their constant support throughout my Ph. D study and their sacrifices and blessings have always encouraged me to take my further steps.

Finally, the Sunday Lunch with Harish Bhai, Rabin Da and Naresh Bhai is appreciated forever.

Table of contents

List of Tables

List of Figures

List of Abbreviations

QM	**Q**uantum **M**echanical
HF	**H**atree **F**ock
WFT	**W**ave **F**unction **T**heory
CCSD(T)	**C**oupled **C**luster theory with **S**ingles, **D**oubles and **P**erturbative **T**riples excitations
MP2	**M**øller-**P**lesset second order perturbation theory
SCS-MP2	**S**pin **C**omponent **S**caled **MP2**
SCS(MI)-MP2	**S**pin **C**omponent **S**caled **M**olecular **I**nteraction **MP2**
MP3	**M**øller-**P**lesset third order perturbation theory
MP2.5	**M**øller-**P**lesset second order perturbation theory with the inclusion of third order correction
DFT	**D**ensity **F**unctional **T**heory
DFT-D	**D**ensity **F**unctional **T**heory with empirical **D**ispersion correction
BSSE	**B**asis **S**et **S**uperposition **E**rror
AR	**A**ldose **R**eductase
MD	**M**olecular **D**ynamics
ESH	**E**xplicit σ-hole
EP	**E**xtra **P**oint **C**harge
MetaD	**M**etadynamics
TL	**T**etraloop
WT-MetaD	**W**ell **T**empered **M**etadynamics
RESP	**R**estricted **E**lectrostatic **P**otential
MD	**M**olecular **D**ynamics
CV	**C**ollective **V**ariables
CFT	**C**rooks **F**luctuation **T**heory
SGTI	**S**low **G**rowth **T**hermodynamic **I**ntegration
DTI	**D**iscrete **T**hermodynamic **I**ntegration
FGTI	**F**ast **G**rowth **T**hermodynamic **I**ntegration
ss-RNA	**S**ingle-**S**tranded **R**ibonucleic **A**cid
PDB	**P**rotein **D**ata **B**ank
ESP	**E**lectro**S**tatic **P**otential
HOMO	**H**ighest **O**ccupied **M**olecular **O**rbital
LUMO	**L**owest **U**noccupied **M**olecular **O**rbital
NBO	**N**atural **B**ond **O**rbital

Abstract

The aim of this thesis is to investigate the accurate stabilization energy and binding free energy in various non-covalent complexes spanned from small organic molecules to biomolecules. Non-covalent interactions such as H-bonds, π...π stacking and halogen bonds are mainly responsible for understanding of most biological processes, such as small molecule interactions with surface, protein-ligand binding in the cell machinery, etc.

In the thesis, different non-covalent complexes such as graphene…electron donor-acceptor complexes, DNA base pair interaction with silica surface, etc, were investigated. The reference stabilization energies were calculated at *ab initio* level, e.g., CCSD(T)/CBS method wherever possible. On the other hand, more approximated scaled MP2 method (MP2.5/CBS/6-31G*(0.25)) is taken as reference instead of CCSD(T)/CBS due to the size of the complexes. Further, the DFT and MM energies were also tested towards the reference one. The knowledge of non-covalent interaction is required for rationalizing of any association processes in nature which requires accurate description of the free energy change. The state-of-the-art molecular dynamics simulation in full atomic scale and biased metadynamics free energy method is used for binding free energy calculations. The well tempered metadynamics simulation was adopted for obtaining the conformational free energy change in biomolecules. For alchemical free energy, the non-equilibrium fast growth thermodynamics integration has also been discussed.

Abstrakt

Cílem této práce je prozkoumat přesné stabilizační energie a volné vazebné energie pro různé nekovalentní komplexy počínaje malými organickými molekulami a konče biomolekulami. Nekovalentní interakce např. vodíkové vazby, π…π patrové interakce či halogenové vazby jsou zodpovědné za pochopení většiny biologických procesů, jako jsou interakce malých molekul s povrchem, protein–ligand interakce v buňkách atd.

V práci byly vypočteny stabilizační energie pro různé nekovalentní komplexy, jako elektronové donor-akceptorové komplexy grafenu, páry bazí DNA interagující s povrchem oxidu křemičitého atd. Referenční stabilizační energie, kdekoli to bylo možné, byly výpočteny pomocí metody CCSD(T)/CBS. Vzhledem k velikosti studovaných komplexů byly v některých případech použity jako referenční metody místo CCSD(T)/CBS více aproximativni metody, např. škálované MP2 (MP2.5/CBS/6-31G*(0.25)). Mimo jiné byly stabilizační energie také počítané pomocí metod DFT a MM. Znalost nekovalentních interakcí je nevyhnutelná pro racionalizaci asociačních procesů v přírodě a vyžaduje přesný popis změn volné energie. Nejmodernější molekulově dynamické simulace s plně atomistickým popisem a “biased“ metadynamické simulace byly použity pro výpočet volné vazebné energie. “well“ temperované metadynamické simulace byly použity pro získání změn volné energie v důsledku konformačních změn biomolekul. V práci se také pojednává o alchymických výpočtech změn volné energie pomocí nerovnovážné “fast-growth“ termodynamické integrace.

CHAPTER 1

1 Introduction

1.1 Thesis Overview

This thesis contains accurate calculations of non-covalent interactions between the small organic molecules, interactions of biomolecules with the surfaces and also the folding/un-folding dynamics of small RNA-hairpins. In addition, the standard Molecular Dynamics (MD) simulation has also been used at an atomistic level in explicit solvent to encounter the free energy of binding between the electron donor-acceptor complexes.[1]

A portion of the thesis is based on the tuning of halogenated ligand in the human Aldose Reductase (AR).[2] The halogen bond is mutated to a H-bond to check the tunability of the binding affinity inside the cavity. To do this, we empirically defined the halogen bond with the description of the explicit σ-hole (ESH). The anisotropy of the ESH has been defined with an extra point (EP) charge on the position of the σ-hole which has already been shown in the previous studies. The non-equilibrium fast growth thermodynamic integration free-energy MD simulations have been used to mutate the halogen atom σ-hole to another atom.

Besides the standard MD simulation, we adopted state-of-the-art enhanced sampling technique, namely Metadynamics (MetaD). This technique is based on biasing the simulation on some selected degrees of freedom called Collective Variables (CVs). Although, MetaD simulation has previously been used in many studies such as peptide folding, protein folding and ligand binding to the protein cavity, a proper choice of CVs still remains a challenging task. The advantage of MetaD is that it can easily samples the rare events in a metastable system. The MetaD technique has been used to obtain the association free energy between the DNA base pairs on a hydrophobic silica surface.[3] The MetaD technique is crucial not only for calculating the binding affinity but also for observing the different orientation of binding of the molecules on the surface. It has correctly predicted all the expected modes of association of the base pairs on the surface.

Further, the thesis also focuses on the folding/un-folding dynamics of small RNA-

hairpins tetraloops (TLs).[4] It is an extremely challenging task to fold RNA hairpins with the advanced enhanced dynamics because the choice of proper CVs is difficult. It is known that biomolecules has thousands of conformations, and therefore, the search for the conformation of interest, needs a better definition of CVs. To the best of our knowledge, this is the very first time that a CV based approach is used to unfold or fold small RNA hairpins. The chosen CVs stands out as physics-based CVs that takes into account the process of interest, however, due to the orthogonality characteristic (hidden variable, same as chosen CV) of the CVs, it does not allow for the proper sampling of the requested FES. To perform this, we employed advanced MetaD technique, called well tempered MetaD (WT-MetaD). The advantage of using WT-MetaD is that the deposition rate of the biasing potential is inversely proportional to the time. Therefore, the WT-MetaD does not overfill the free energy surface (FES) which is presumably more effective over the standard MetaD.

1.2 Non-covalent Interactions

Non-covalent interaction represents a dominant type of interaction in biodisciplines and in supramolecular chemistry. Non-covalent interactions are one or two orders of magnitude weaker than the covalent interactions. There are few types of non-covalent interactions such as electrostatic interaction, induction/polarization, dispersion, charge transfer etc. Presently, an intense research interest has been given to the non-covalent interaction because it plays an important role in chemistry, physics and especially in biodisciplines. The non-covalent interactions determine the structure of biomacromolecules such as DNA, RNA and proteins and are responsible for the molecular recognition process. The non-covalent interactions are also seen to contribute vastly in the field of drug design, crystallography or designing and synthesis of new materials[5-10] etc.

Electrostatic interactions such as Hydrogen bonding (H-bond), Halogen bonding (X-bond where X= Cl, Br, I) and in the following sub-sections, those types of non-covalent interactions will be briefly discussed. Besides them also π…π interactions which are basically determined by dispersion interaction will be mentioned.

1.2.1 H-bond Interactions

The X-H...Y H-bond is the bond between proton donor (X) and electron donor (Y)

where hydrogen which is covalently bound to electronegative atom X carries a positive charge. The formation of the H-bond is mainly due to electrostatic interaction between positively charged H and negatively charged Y. Further, the charge transfer energies also contribute to the stabilization the H-bond.[11] Formation of an H-bond is mostly accompanied by elongation of the X-H bond which leads to increase of the dipole of the proton donor thus also for the increase of the dipole-dipole attraction between proton donor and acceptor.

The Quantum Mechanical (QM) calculation revealed the charge transfers from the lone pairs of the proton acceptors to the X-H σ* antibonding orbitals of the proton donor. An increase in the electron density in the antibonding orbitals weakens the X-H covalent bond, which leads to its elongation, accompanied by the lowering of the X-H stretching frequency in the infrared spectroscopy.[11,12] There is another type of H-bond where X-H stretching frequency shift to a higher values. This type of H-bond called as "improper, blue-shifting" H-bonds. The "Blue-shifting" was first observed by Trudeau et al. in 1980. They found that C-H stretching frequency of Fluoroparaffins containing $–CHF_2$ was shifted to higher frequency upon binding with various proton acceptors.[13]

Later, the reason behind "improper, blue-shifting" H-bonds was explained on the basis of the natural bond orbital (NBO) analysis[11] The electron density (ED) in the Benzene...CHF_3 complex was transferred mostly to C–F σ* antibonding orbitals which are not directly involved in the X-H...Y contacts. There is negligible or only small ED transfer to the C–H σ* antibonding orbital. The structural reorganization of the proton-donor leads to the contraction of the X–H bond directly involved in the X–H...Y contact ended up with a concomitant blue shift of its stretch frequency.

H-bonds are of different strength depending upon the character of atoms X and Y. Complexes of Adenine (A)…Thymine (T) and Guanine (G)…Cytosine (C) possess two and three very strong H-bonds, respectively.[14] A schematic representation of the A…T and G…C pair is shown in Figure 1.1 in a ds-DNA duplex. The benchmark study by Jurecka et al. in the S22 dataset predicted the CCSD(T)/CBS interaction energy for A...T and G...C H-bond pairs of −16.86 and −32.06 kcal/mol, respectively.[15]

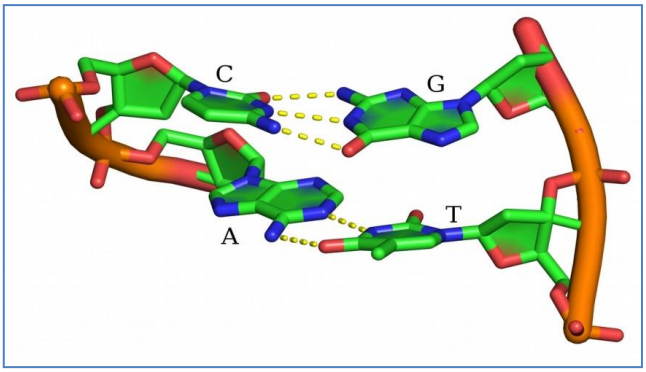

Figure 1.1: Schematic representation of A…T and G…C pair in the ds-DNA duplex.

Weaker H-bonds are those where the H-atom is attached to a carbon moiety. Even weaker are the C-H…π H-bonds such as in the Benzene…Benzene and Benzene…CHF_3 T-shaped complexes. The work in the thesis is mainly related to the first type of H-bonds where the interactions between the base pairs are thoroughly investigated with available QM based and force-field method.[2]

1.2.2 π…π Interactions

The π…π interaction is another type of non-covalent interaction. Its importance can be understood from the stability of ds-DNA upon base stacking.[16] The benchmark study in the S22 data set revealed that the interaction energy for the stacked A-T and G-C pair contribute 73% and 60% to the total H-bond interaction energy and they are −12.3 and −19.02 kcal/mol.[15] The π…π interaction is also known to participate in the stability of the biomolecules such as protein folding, the drug binding to the protein cavity, and also in the material science and in the molecular recognition.[5,6,17-20] The driving force for the π…π interaction comes from the London dispersion forces are also as Induced-dipole-induced-dipole interaction.

For example, π…π stacking interaction stabilizes the parallel displaced (PD) and T-shaped configurations of Benzene dimer.[21-29] The symmetrical stacked arrangements are stabilized by dispersion interactions but destabilized by quardrapole-quardrapole interaction. In the case of parallel displaced and T-shaped configurations, quardrapole-quardrapole interaction is attractive.

1.2.2.1 Surface…π Interactions

Surface…π interaction is mainly driven by the dispersion interaction between the surface and the adsorbed molecule. There are several studies on the Surface…π interaction such as interaction between graphene and electron acceptor molecules[2,30-34] and also graphene-DNA interactions,[35,36-47] graphene-peptide interaction.[48-50] The former type of interaction is popular for discovering fine semiconductor based electronics, since it is known that the graphene has mobile π electrons. The interaction between graphene and DNA is used for sequencing of DNA. DNA is known to possess specific recognition interactions that can interact with graphene through hydrophobic adsorption and π…π stacking.[51,52] There are several studies on the non-covalent functionalization of graphene oxide and reduced graphene oxide with DNA.[53-55]

The computations of π…π stacking interactions are demanding and depend upon the size and shape of the complex. In this work, graphene...organic electron acceptor complexes (*cf.* Figure 1.2) have been investigated with the density functional theory (DFT) and with the force field description. Further, the DNA base pairs interaction with the graphene and hydrophobic silica surface has also been studied. In all cases the dispersion interactions plays a key role.[2]

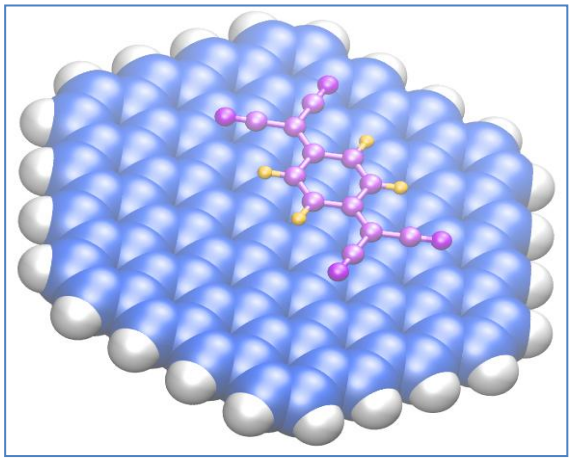

Figure 1.2: Schematic representation of dicircumcoronene tetracyanoquinodimethane π…π stacked complex.

The Wave Function Theories (WFT) such as MP2.5, SCS-MP2 methods were applied to the small complexes of the electron donor-acceptor molecules[1] (see chapter 3). The standard MP2 method overestimates the π…π interaction.[56-59] In case

of DFT, the empirically added dispersion correction almost accurately describes the dispersion effect generated in the π…π stacking interaction.[60-62]

1.2.3 Halogen Bond Interactions

Halogen bond (X-bond) is another type of non-covalent interaction that occurs between a halogen atom, acting as a Lewis acid and an electron donor acting as a Lewis base.[63,64] A typical example of a halogen bond consists of A-X…Y where X (X = Cl, Br, I) is electron acceptor and Y is electron donor.

The interaction of dihalogen (Cl_2, Br_2, and I_2) is known from the middle of the 18th century when Frederik Guthrie gave the first example of chemical adducts between the halogen atoms and an electron donor species.[65,66] Since then several studies have been reported so far on the halogen reactions to the various types of organic donor molecules. In 1950, Mulliken came up with the theory of interaction between the electron donor-acceptor complexes[67-69] and Mulliken's theory has been used to describe the mechanism of Halogen bond formation.[67]

The driving force of the halogen bonds was explained by Tim Clark in 2007, who showed the existence of σ-hole.[70] The calculations predicted that the σ-hole is basically a region of positive electrostatic potential (ESP) at the top of the halogen atom opposite to the A-X covalent bond. A clear picture of σ-hole in halogen bonds is shown in Figure 1.3. Due to the existence of the positive hole, the electron acceptor acts as a Lewis acid.

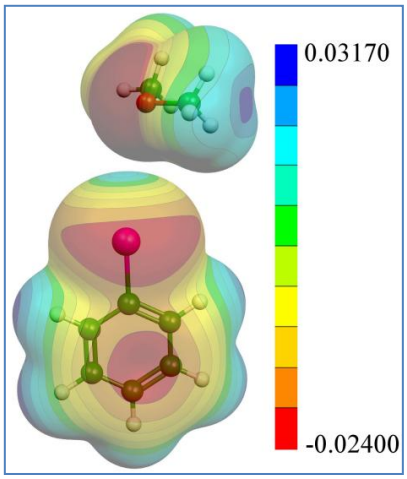

Figure 1.3: Schematic representation of an halogen bond adduct between Bromobenzene and dimethyl ether. The electrostatic potential was calculated on 0.001 au molecular surface.

The "σ-hole" has been basically defined as anisotropic distribution of electron density around a halogen atom. The hole is arrived when an atom of uniform electron density is covalently bonded to another atom; the electron density is polarized towards the bond direction. The important role of halogen bonding so far has been investigated in many natural biological processes and in the medicinal areas, and also in the areas of drug design, mainly in ligand binding to protein cavity etc.

1.2.3.1 Tuning of Halogen Bonds

The halogen bond tuning is rather easy to treat theoretically. There are few factors which determine strength of a halogen bond. The first one is the size, shape and the polarizability factor of a halogen atom.[64] So far, halogen bonds are well documented for Cl, Br and I atom. Fluorine atom is basically excluded from the halogen bond series because of its high electronegativity and low polarizability.[71] However, F atom is capable of forming halogen bonds under some certain condition. The strength of a halogen bond interaction follows the following trend F << Cl < Br < I, with I forming the most stable halogen bonds.[64,72]

The second factor is the incorporation of electron withdrawing group near the halogen atom that makes the σ-hole more positive. For example, it was shown that the CH_3Cl molecules does not contain a positive σ-hole, when however, all the hydrogen atoms are replaced by F atoms one by one, the σ-hole becomes more pronounced and found to be the maximum.[70]

The third factor is most likely the angle between A-X…Y atoms. An ideal halogen bond interaction would be a "head-on" contact between the σ-hole of the halogen atoms to the halogen bond acceptor.[73] The ideal angle has been found to be 180^o in many studies so far. If the angle deviates to 25^o to 30^o, the interaction energy reduces approximately to ~50%.[73] The reason behind the reduction might be due to the repulsion between the partially charged halogen bond acceptor and the negative ESP of the halogen atom, the σ-hole becomes inaccessible to the electron donor. The most typical electron donor molecule has been recognized as the carbonyl moiety.[74] When a strong halogen bond is formed, the corresponding inter-atomic distances are shorter than the sum of the VDW's radii of partner atoms in the complex. For example, average calculated Cl...O, Br...O, and I...O distances are 3.18, 3.10 and 3.06 Å, whereas, the sum of VDW's radii is 3.27, 3.37 and 3.50 Å, respectively.

There are several theoretical studies published so far applying the Coupled Cluster (CCSD(T)) and MP2 methods to estimate the halogen bond strength. Table 1.1 shows a summary of few halogen bonds and their strength taken from the work of Narayan et al.[75]

Complex	ΔE (kcal/mol)	Method	A-X…D Angle(deg)	Source
PhI…BB	−3.39	MP2/TZVPP	175.6	Wilchen et al.[76]
PhBr…BB	−2.15	MP2/TZVPP	177.4	Wilchen et al. [76]
PHCl…BB	−1.34	MP2/TZVPP	171.2	Wilchen et al. [76]
PhBr…OCH_2	−0.71	MP2/aug-cc-pVDZ	168.9	Lu et al.[73]
PhBr…NH_3	−0.66	MP2/aug-cc-pVDZ	179.7	Lu et al. [73]
PhBr…SH_2	−0.51	MP2/aug-cc-pVDZ	165.1	Lu et al. [73]
PhBr…OH_2	−0.44	MP2/aug-cc-pVDZ	178.0	Lu et al. [73]

BB = N-acetyamide

Table 1.1: Summary of strength of few halogen bonds.

1.2.3.2 Importance of Halogen Bonding in Rational Drug Design

The halogen bond is currently receiving increased attention as an important non-covalent interaction acting in the protein-ligand (P-L) complexes.[77-79] Many drugs in the market contain halogens allowing to establish halogen bonds with proteins.

A recent systematic investigation of the selectivity of halogen bonds in P-L complexes by Hardegger et al. showed that the halogen bonds can serve as a powerful tool in increasing the binding affinity.[65] A few examples of halogen bonding impact in the P-L complexes include the following:

1) Incorporation of a halogen atom in the *ortho* position of the 1-phenyl ring of CDPPB [3-cyano-N-(1,3-diphenyl-1H-pyrazol-5-yl)-benzamide] leads to an increase in both the binding and functional activities of mGluR.[80]

2) Iodine in 4-(3,5-dimethylphenoxy)-5-(furan-2-ylmethylsulfanylmethyl)-3-iodo-6-methylpyridin-2 (1H)-one (R221239) increases the binding affinity to wild type HIV-1 reverse transcriptase through the formation of a halogen bond with the carbonyl oxygen of TYR188. This interaction is regarded as one of the important interactions leading to the overall P-L binding affinity.[81,82]

3) Dichloroindolyl enaminonitrile (KH−CB19) is a potent inhibitor CLK1

and CLK4. The Cl···O interaction between a chlorine on the dichloroindolyl enaminonitrile (KH−CB19) and the carbonyl group of Glu242 is recognized as a key mediator of the hinge interaction of KH-CB19.[83]

Recently, in our laboratory, we investigated the selectivity and the binding affinity of CDK2-inhibitors which explicitly includes Bromine atoms. Later the Bromine atoms were substituted by hydrogen and some other alkyl groups to investigate the binding motifs of the drugs. It was shown that the Bromine compounds are more potent than the other selective compounds inside the CDK2 cavity.

1.2.3.3 Force Field Description of σ-hole

The Molecular Mechanics (MM) is unable to describe the halogen properly because it cannot account for the anisotropic distribution of the charge density on the halogen atoms. The idea of treating the anisotropy for halogen bonding empirically was first given by Ibrahim.[84] A mass less (no Lennard Jones (LJ) parameter) extra-point (EP) charge had been introduced on the halogen atoms to mimic the lone pair of electrons. The EP is used to maintain the σ-hole present in the halogen atom. The actual position of the EP was decided upon the correlation between the X-EP and X…Y halogen bond distances. Authors tested the σ-hole approach on the basis of solvation free energy calculations on halobenzenes in implicit solvent. Later, Kolar et al. introduced "aF" (all fit) charge methodology[85] to explicitly account for the EP charge on the halogen atoms which is almost identical to that introduced by Ibrahim.[84] Here the EP is called explicit σ-hole (ESH). Only one empirical parameter was introduced i.e., the distance between the halogen atom center and the ESH. The charges were calculated with the so-called restricted electrostatic potential approach (RESP).[86,87] Here before the RESP calculation, an additional fitting position was placed on top of the halogen atom and optimized the charges in order to obtain the ESP charges on halogen atoms. The results obtained were thoroughly compared with the CCSD(T)/CBS data and found to be excellent in agreement. In this study, we computed the "aF" approach[85] to calculate the solvation free energy of few AR-inhibitors and compared the results with SMD solvation data[3] (see chapter 3).

1.2.3.4 MM Free Energy with σ-hole Approach

Calculating free energy either in small molecules or in biomolecules is always a challenging task in all atom simulation. The free energy cannot be accurately

described unless all the necessary parameters are well defined either from some experimental data or from benchmark theoretical data base. Here we performed a well defined, well parametrized ESH on the AR-inhibitors and mutated in the explicit waters and finally compared the results with available QM data.

Fragment Based Approach: The Fragment Based Approach (FBA) has been applied for the very first time in this work of mutation of several AR-inhibitors inside the explicit waters in all-atom Molecular dynamics (MD) simulation. The FBA approach has been mainly introduced to treat different fragments in the drug candidate contain halogens and ESH. It is already mentioned that the "aF" charge[85] is used for the free energy simulation which is constructed from the all fit delocalized charges throughout the whole molecule. Therefore, if there is a rotatable fragment present during mutation in the free energy simulation contains ESH, it can destabilize the alchemical perturbation. However, the problem can be solved by separating the system into fragments depending upon the number of the rotatable bonds containing ESH present in the system. After fragmentation, each fragment will contain delocalized charges with ESH, so that, during free energy simulations, the alchemical perturbation will be simple and further electrostatic destability will not hamper the simulation. In the recent study of AR-inhibitors, we divided all the inhibitors in two different fragments (fragment 1 possessing net charge of 0 and fragment 2 with the net charge of –1, *cf.* Figure 1.3) in order to minimize the extent of alchemical perturbation what improves the convergence of the resulting free energies.

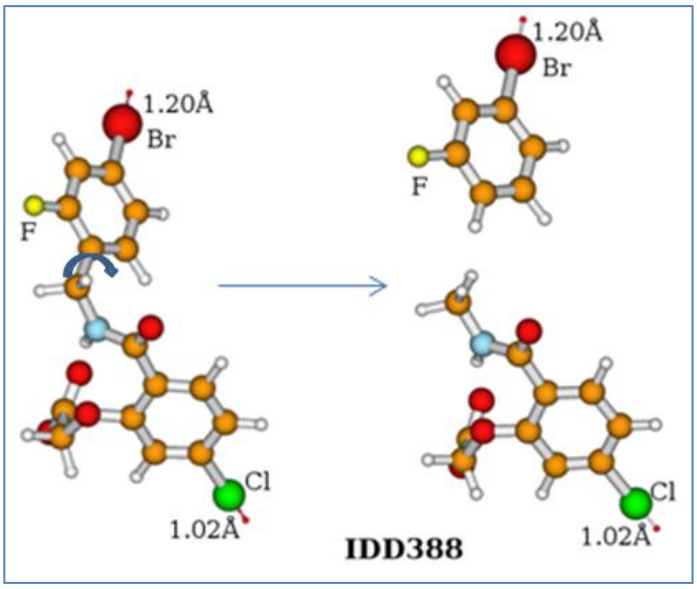

Figure 1.4: Schematic representation of IDD388 AR-inhibitor which possesses two different fragment due to the rotatable bond shown.

1.3 Dynamics of short RNA Hairpin Tetraloops

1.3.1 Overview

The MD simulation represents the most suitable tool to perform molecular recognition, flexibility and dynamic behavior of RNA biomolecules. However, the MD simulation of nucleic acids (NA) did not reach much to the expectation like proteins and other biomolecules. This is due to the lack of available experimental high-resolution structure which would allow validating the force-field and long MD dynamics for nucleic acid structures.[88] Unfortunately, so far, there has been less experimental and simulation data for RNA compared to proteins and the DNA biomolecules which basically brings more complication to explore the dynamical behavior of the corresponding RNA structures. Unlike DNA structures, often found in double helix structures (mostly in the canonical form), RNA biomolecules can frequently feature non-canonical forms such as bulged states, hairpins, etc.

1.3.2 RNA Dynamics with AMBER Force Field

The dynamics of short RNA has been improved progressively in the recent years. Previously, with available force field, the simulations were not able to produce sufficient data of RNA beyond 500 ps timescale. This was due to the lack of experimentally resolved RNA data as well as lack of computational power etc.

Currently, we witness an enormous change of producing RNA structures with increasing the computing power as well as introducing new methodology such as PME treatments for the long range electrostatic interaction in the RNA simulation.

The first RNA simulation was done with the Cornell's AMBER force field. Nowadays, there exists several RNA simulations such as single and double-stranded RNA's, catalytic RNA's, binding of small RNA's to the proteins and also adsorption on the surfaces. Those studies provided nice agreement in terms of structures. The respective simulations were not longer than 100 ns, therefore, the question on reliable RNA structures and convergence of the simulation arrives.

1.3.3 Force Field Inaccuracies

The AMBER force field is one of the most frequently used force filed for RNA, DNA and also for the protein biomolecules. Despite the fact that it produces reliable description of the structures and thermodynamics of RNA systems, however, the force field has some serious problems:

1) Most importantly, it is the proper description of the sugar-phospahate backbone torsions which has multiple degrees of freedom.[89,90] It creates a considerable problem since the torsions are "physics-based" parameters, cannot be directly derived from either experiment or QM data. They cannot also be correctly described by only a single set of partial atomic charges centered on a particular atom.

2) The formation of non-canonical structures of RNA biomolecules, i.e., loops, hairpins, bulged states etc. A further testing and validation of recent parameterization of force field is needed.[91]

3) The force field should also properly describe the RNA ion interaction in the explicit water simulation, since, it is known that the RNA folding/un-folding mechanism also dependent on the concentration of ions in the solution.[92]

However, the combination of increased computational power and more reliable experimental data to refine force field makes MD simulation a promising tool to study structures and dynamics of human RNA biomolecules.

1.3.4 Development in RNA Force Field

The structure and conformational dynamics of molecules in the MD simulations critically depends on the quality and accuracy of the applied empirical force field. At the beginning of 21th century, the MD simulation using AMBER force field (parm94, parm99) have been shown to accurately reproduce the structural and dynamic properties of a large variety of canonical and non-canonical nucleic acids in water.[93-99] Moreover, the simulations were able to describe the complex conformational changes such as A→B transition in duplex and triplex DNAs in the extreme environments. However, all the simulations were performed in a very short time scale (under 10 ns) which was the normal simulation period at that time. Later, in an extended simulation (50ns trajectories), Zakrzewska et al. showed that a massive α/γ transitions to the *gauche*$^+$, *trans* geometry (away from the g^-, g^+ state) introduced severe distortions in DNA simulation.[100] These types of ambiguity in the DNA simulation was realized as a general sequence-independent problem of parm94 or parm99 force field simulation.[101-103]

To overcome the force field inaccuracies, Orozco et al. in 2007, fully reparameterized the α/γ torsion term and derived a new AMBER force field based on the AMBER-parm99 which is named as parmbsc0 force field.[104] This force field was recognized to accurately model the structural dynamics of nucleic acids over the logical MD simulation time reported so far, but also provided very good representations of the nucleic acid structures in simulations 20 times longer. However, in a much longer time scale than the reported one in the simulation of B-DNA revealed that the helical twist remains underestimated, the occasional γ = *trans* flips are still too frequent. Later, Zgarbova et al. parameterize the glycosidic torsion, χ, describing rotation about the bond that links the base to the sugar moiety and determines the relative orientation of the nucleaobase and sugar moieties in DNA and RNA. The improvement of glycosidic torsion is important because it is involved in the equilibrium of A and B forms of DNA as well as the C2'-endo and C3'-endo equilibrium. The current force field by Zgarbova et al. was reparametrized over parmbsc0 and named as ff99bsc0χ_{OL3}.[105] The ff99bsc0χ_{OL3} remove destabilization of the anti region found in the ff99 force field and thus eliminates the formation of the ladder-like structures and spurious artifacts generated by older versions of the force

field.

1.3.5 Folding/un-folding Dynamics of RNA Hairpins

The RNA hairpin, consisting of a double-stranded RNA (ds-RNA) stem and a terminal loop, is the most abundant secondary structure of RNA motif, and therefore, essential to the folding and functioning of RNAs. Thus the RNA hairpin is an ideal model system for understanding the general folding mechanism of the RNA molecules.

Until recently, a number of folding/un-folding studies have been performed both experimentally and theoretically.[106-120] There has been a considerable progress in understanding the mechanism of RNA folding. However, several studies revealed that the folding/un-folding mechanism contains many intermediate states, and therefore, the knowledge of RNA folding is still limited. Many studies further showed that the RNA folded in a two-state model where only a native and a denaturated state were found without any intermediate states. However, other studies suggested that the folding of RNA-hairpins should be more complex due to energetic and topological frustration. For example, Thirumalai et al. performed the complicated behavior of folding of 22-nucleotide RNA hairpin.

In 2011, Kuhrova et al. studied the folding/un-folding mechanism of 5'-UNCG-3' and 5'-GNRA-3' RNA hairpin tetraloops (TLs) using Replica Exchange Molecular Dynamics (REMD) simulation with the recently developed ff99bsc0χ_{OL3} force field.[121] The study clearly revealed that starting from fully un-folded conformations resulted in the folded TL structure within 2 Å all-atoms RMSD of the native structure. The simulation was found to be in better agreement than the previously reported REMD simulation on TLs folding which was only able to reach structures having ~4 Å RMSD from the native state due to force field limitations.

In the present study, we represented a well tempered MetaD (WT-MetaD) simulation and long standard MD in microsecond time scale to determine the stability of GNRA and UNCG TLs. Both the simulations clearly mentioned the same mechanism of un-folding for both the TLs.[4]

1.4 Thesis Organization

At Chapter 2, we present the computational methods used in this thesis which includes WFT methods (Coupled Cluster Singles Doubles with perturbative triples (CCSD(T)), Second Order Møller-Plesset Perturbation (MP2) theory etc). The DFT and MM methods such as standard MD simulation and the CV based biased free energy method to calculate the folding/un-folding dynamics of small RNA-hairpins are also present.

Chapter 3 is divided in five sections, thoroughly discuss the results obtained. Specifically,
Section 3.1 discusses the accurate interaction energy and binding affinity of the graphene...organic electron acceptor complexes.
Section 3.2 shows the interaction and association free energy of the DNA base pairs on the hydrophobic silica surface.
Section 3.3 describes the effect of halogen-to-hydrogen bond substitution on human AR inhibition.
Section 3.4 discusses the accurate dissociation energies for eleven hydrogen-bonded and eleven dispersion-bound complexes.
Section 3.5 describes the CV based bias free energy approach on the folding/un-folding dynamics of RNA-hairpins.

Chapter 4 represents the conclusions of the thesis.

CHAPTER 2

2 Methods

2.1 Interaction Energy Calculations

The interaction energy of a binary complex can be expressed as follows:

$$\Delta E_{int}(A \ldots B) = E(A \ldots B) - E(A) - E(B), \qquad 2.1$$

where E(A...B), E(A) and E(B) denotes the total electronic energy of the complex A...B, and the subsystems A and B, respectively. This so-called supermolecular approach is applicable to all type of binary molecular complexes.

However, there is a drawback of the above-mentioned supermolecular approach which is called the Basis Set Superposition Error (BSSE). The error originates in an unequal description of the dimer and monomers. In 1970's, Boys and Bernardi[122] introduced the Counterpoise-corrected procedure (CP) method to eliminate BSSE error. The typical CP-corrected interaction energy is calculated as follows:

$$\Delta E_{int}^{CP}(AB) = E_{AB}^{AB} - E_{A}^{AB} - E_{B}^{AB}, \qquad 2.2$$

where the dimer and the monomer energies are calculated in the dimer basis set.

The total electronic energy is expressed as the sum of Hatree Fock (HF) energy and the correlation interaction energy.

$$\Delta E_{int} = \Delta E^{HF} + \Delta E^{corr} \qquad 2.3$$

Nowadays, the accurate calculation of ΔE^{corr} is the most challenging part in the total interaction energy calculation from the computational point of view. There are many wave function based methods such as CCSD(T)/CBS, MP2 and also DFT methods which can be used for the calculation of correlation interaction energies. However, it is well known that the different types of non-covalent interaction require different methods for counting proper correlation energies. For example, the MP2 based methods are known to produce accurate interaction energies for H-bonded complexes but overestimates the interaction energy for π…π stacked complexes.[123-126] Reasonable interaction energies for these complexes are obtained by using DFT

methods empirically corrected for the dispersion energy. In the next sections different methods, which are used for the description of non-covalent interactions, will be discussed.

2.1.1 CCSD(T) Method

The CCSD(T)/Complete Basis Set (CBS) method is taken as a reference method for producing reliable binding energies. The CCSD(T)/CBS interaction energy has been used to produce a large body of benchmark data which now is widely used for the validation of the other methods and parameterization of the empirical models.[127] The idea of the Coupled Cluster method was first given by Jiri Cizek[128-130] in the late sixties. Nowadays, CCSD(T)/CBS theory is called as the "*golden standard*" in the field of computational chemistry because of producing highly accurate numbers. The accuracy goes beyond 1 kcal/mol, called as the "chemical accuracy" but usually reaches even the "subchemical accuracy" (0.1 kcal/mol).[131-133] However, the increase of the computational cost with the size of the molecule is enormous.

The CCSD(T) method is based on the extension of the CC theory which relies on the exponential formulation of the wave operator and its expansion into clusters of excitation operator. The most computationally demanding step with the CCSD(T) scales as $N_o^3 N_v^3$, where N_o stands for the number of correlated occupied and N_v, the number of active virtual orbitals. Nowadays, due to massive parallelization of the method,[134-137] system with more than 30 atoms can be routinely treated.

The estimate of the CCSD(T)/CBS interaction energy is composed as:

$$\Delta E\big(CCSD(T)\big) = \Delta E(SCF) + \Delta E(MP2) + \Delta\Delta E\big(CCSD(T)\big), \qquad 2.4$$

where the ΔE(SCF) is the fastest converging component among three. Mostly, in this work, the SCF energy was computed with the aug-cc-pVQZ basis set. The ΔE(MP2) is the MP2 correlation interaction energy which is recognized as the slowest converging term in the CCSD(T) correlation energy expansion. To obtain close estimate of the CBS limit, the ΔE(MP2) is extrapolated from aug-cc-pVDZ and aug-cc-pVTZ basis set using Helgaker's extrapolation scheme. The Helgaker's[138,139] two point extrapolation scheme has been used.

$$E_{corr}^{X} = E_{corr}^{CBS} + B/X^{\beta}, \quad 2.5$$

where $E_{corr}^{X}$is the correlation energy calculated at some specific basis set, E_{corr}^{CBS} is the complete basis set limit of the correlation energy, B is per-power factor, X is the cardinal number of the basis set and β is set to 3.

The ΔΔE(CCSD(T)) is defined as the difference between the interaction energies calculated at the CCSD(T) and MP2 level (ΔE(CCSD(T) − ΔE(MP2)), called as the CCSD(T) correction term. This term is considered as the most computationally demanding, despite the fact that it is usually calculated in a small or medium size basis set. In this work, the ΔΔE(CCSD(T)) term is calculated with aug-cc-pVDZ basis set which is known to produce accurate ΔΔE(CCSD(T)) energies.[140-142]

2.1.2 MP2 Methods

The popularity of the MP2 method, covering a large portion of correlation energy is due to its low computational cost and high performance/cost ratio. The MP2 method also has some of its own disadvantages. The first is its basis set dependence. It was shown recently that for dispersion bound complexes, the MP2/6-31G*(0.25) closely agrees with the CCSD(T) data.[143] This is preferably due to the error compensation. The second disadvantage is that it significantly overestimates the binding energy for the π…π stacked complexes (overestimation by ~80% in the parallel displaced complexes).[144-146] The main reason behind the overestimation is that the dispersion energy is considered at the uncoupled Hatree-Fock (UCHF) level.[125,126]

The above mentioned disadvantages were eliminated by introducing the Spin Component Scaled MP2 (SCS-MP2) methods are based on the separate scaling of the antiparallel (singlet, S) and parallel (triplet, T) spin components of the correlation energy.[147]

$$E_{SCS}^{corr} = pSE_{S}^{corr} + pTE_{T}^{corr} \quad 2.6$$

Both $E_{S}^{corr}$and E_{T}^{corr}represents the singlet and triplet components of the energy. The two parameters pS and pT, were derived from theory and fitted towards the reference calculations, respectively.[148]

2.1.2.1 Scaled MP2 Method

The MP2.5 method provides accurate binding energies for different types of non-covalent complexes. The MP2.5 binding energy is constructed as the average of MP2 and MP3 binding energies. Specifically, the MP2.5 interaction energy is defined as follows:

$$\Delta E^{MP2.5} = \Delta E^{MP2} + \frac{1}{2}E^{(3)} = \Delta E^{MP2} + \frac{1}{2}(\Delta E^{MP3} - \Delta E^{MP2}), \qquad 2.7$$

where $\Delta E^{MP2.5}$ and ΔE^{MP2} are MP2.5 and MP2 interaction energies and $E^{(3)}$ is third-order correction which is defined as follows: $E^{(3)} = \Delta E^{MP3} - \Delta E^{MP2}$. The main advantage of the method is that it employs a single empirical parameter and is thus biased by two rigorously defined, asymptotically correct *ab-initio* methods, MP2 and MP3. The MP2.5 method has been tested over the S22,[15] S66[141] benchmark datasets and shown to be superior over the scaled spin-component MP2 based methods e.g., SCS-MP2, SCS(MI)-MP2 methods.[149]

2.1.3 Density Functional Methods

Density functional theory is one of the most popular and successful QM approaches. Presently, it is routinely applied for calculating the binding energies of various non-covalent molecules, understanding of various complicated reaction mechanism,[150-158] thermodynamic properties of molecules, etc. The main drawback of DFT method is the fact that it does not count for the long range correlation interaction (dispersion interaction). Most of the density functional such as LDA, GGA, meta-GGA are considered as local one. They provide poor description of the binding energy of the non-covalently bounded complexes. The hybrids functional which are non-local in their exchange part are expected to give better description of non-covalent complexes. They contain some portion of the explicit Hatree-Fock exchange but the correlation part is completely local. For example, B3LYP functional which is non-local in its exchange part only whereas its correlation part is completely local. The most straightforward way how to include the missing dispersion energy is by introducing the empirical atom-atom correction (e.g., Grimme's Dispersion,[61-63] Jurecka's Dispersion.[159]

2.2 Molecular Dynamics Simulations

Nowadays, Computer simulation has been one of the most frequently used tool for scientific research, mainly in the field of bimolecular simulation. The main advantage of so-called Molecular Dynamics (MD) is that it can be easily combined with the available experimental data. Another advantage is that it can make calculations at extreme conditions such as systems with very high/low temperature, pressure and density.

Recently improved algorithms for different types of MD simulations such as constraint algorithm, long-range electrostatics, handling proper thermodynamic conditions and also in the field of the solvent (implicit, explicit) modeling were introduced. There exist different potentials for different biomolecules such as proteins, DNAs, RNAs, etc. The most simplified model in the field of MD simulation is the coarse-grained force field (MARTINI).[160] This method is computationally less demanding and have been used to access long time scale runs with the enormous decrease of the molecular degrees of freedom. Many advanced enhanced sampling techniques has also been introduced such as temperature replica exchange, targeted MD, and locally enhanced sampling and high temperature simulations which enable us to configure very high energy barriers wherever there is metastability in the systems.

Some typical applications of MD simulations which will be discussed in the following text include

1) Characterization of the structural dynamics of small molecules such as peptides to the protein
2) Dynamics of small molecules binding to the nano-surfaces
3) Drug design studies
4) The roll of water in the protein cavity
5) Validations and determination of the structures to experiments

2.2.1 Molecular Mechanics Potential

MD is one of the oldest techniques to treat many-body systems. Over the years, there has been an enormous improvement on the methodology in terms of better potential

energy functions that counts accurate description of the physical and thermodynamical properties of the system.

MD simulation is based on calculating the time evolution of the system of interest by integrating the Newton's equation of motion. Provided, the forces on each atom obtained use an appropriate potential energy function. The potential energy function approximates the Quantum Mechanical energy functions. The equation of motion, which is given below, updates the position and velocities in the system with time.

$$F = m.a \tag{2.8}$$

$$\frac{d^2 r_i}{dt^2} = \frac{F r_i}{m_i}, \qquad i = 1,\ldots,\ldots,N \tag{2.9}$$

the equation describes the Force (F) generate on a particle m_i, with the coordinate r_i moving on a particular direction. The total potential energy depends on the positions of all atoms (r) and can be expressed as follow:

$$V(r) = \sum_{bonds} K_r\left(r - r_{eq}\right)^2 + \sum_{angles} K_v(v - v_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}(1 + cos(n\varphi - \gamma_0) + \sum_{i<j}\left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}\right) + \sum_{i<j}\frac{q_i q_j}{r_{ij}} \tag{2.10}$$

The equation describes bonds stretching, angles bending, dihedral angles torsions and also non-bonded interactions represented by Lennard-Jones potential and Coulomb law, respectively.

2.3 Free Energy

According to the classical thermodynamics, any spontaneous change occurring in nature is directly connected with the negative change of the free energy. In the following, an overview on some aspects related to the free energy definition will be given. In a statistical ensemble, the free energy is defined as follows:

$$F = -\frac{1}{\beta}\ln \int e^{-\beta V_{(r)}} dr \tag{2.11}$$

In a system in with the canonical probability distribution, the absolute free energy is given by following equation:

$$F = -\frac{1}{\beta}\ln Q_{(r,p)} \qquad 2.12$$

where $Q_{(r,p)}$ is the canonical partition function of the ensemble and F is the Helmholtz free energy ($F = U - TS$). The F is the thermodynamical potential associated to the canonical ensemble, U is the internal energy and S is the entropy. In case of isothermal-isobaric ensemble, the free energy recovered as Gibbs free energy ($G = F + PV$). The free energy basically inherits from the overall potential energy, $V(r)$ (Eqn. 2.10), the property of being defined up to an immaterial additive constant. This means that the free energy differences (ΔF) rather than the absolute free energy F will be calculated. The calculation of such differences requires the definition of at least two different thermodynamic and/or chemical states of the system e.g., reactants and products of a chemical reaction which can be distinguished by means of few parameters, called reaction coordinate (z). Within the new reaction coordinate (z), the free energy of the ensemble can be written as follows:

$$F(z) = -\frac{1}{\beta}\ln \rho(z) \qquad 2.13$$

where $\rho(z)$ is the estimate of the probability density over the new sampled reaction coordinate (z)

$$\rho(z) = \int \frac{e^{-\beta H\left(p^N, r^N\right)}}{Q_{(r,p)}} \delta_{\xi(r)-z}(dr)dp \qquad 2.14$$

where H is the Hamiltonian of the system as a function of momentum (p) and the coordinate (r). The measure of $\delta_{\xi(r)-z}(dr)$ represents all the degrees of freedom that have to be integrated out from the whole coordinates r to describe the system in terms of the newly configured reaction coordinate z. From the definition of the Eqn. 2.14, the free energy of two different arbitrary states can be defined as follows:

$$\Delta F = F(z_1) - F(z_0) = -\frac{1}{\beta}\, ln\left(\frac{\int e^{-\beta H\left(p^N, r^N\right)}{}_{\delta_{\xi(r)-z_1}(dr)dp}}{\int e^{-\beta H(p^N, r^N)}{}_{\delta_{\xi(r)-z_0}(dr)dp}}\right) \qquad 2.15$$

2.3.1 Sampling of Rare Events

Most natural phenomena exhibit a wide range of characteristics time scale. For example, protein folding,[161] and single-stranded (ss)-RNA folding,[107-121] all are happening in the microseconds to second time scale. The above defined phenomena are occurring at time scale several orders of magnitude larger than the common integration time scale (typically dt ~ 1-2 fs), very far from the accessible simulation time. Therefore, the systems which prone to change in such a long time scale are characterized by rare events. They do exhibit meta-stabilities (in which the thermodynamical basins are separated by very high energy barriers) highly suffer for the time scale separation.

The systems which suffer such situations are non-ergodic in practice. From the Eqn. 2.13, it is clear that the ergodicity is not fulfilled enough because the condition $z(t) = \xi\big(r(t)\big)$ remains trapped somewhere in the free energy basins and therefore a long MD cannot sample the requested thermodynamic properties (ΔF, ΔH, ΔS, etc) to the other positions.

To fully overcome the non-ergodicity, many theoretical approaches have been introduced so far. There are basically two ways of how the Hamiltonian can be expressed to enhance the sampling in the canonical ensemble.

1) The phase space sampling method, where the Hamiltonian is modified with the time-independent terms, basically improves the statistical sampling in the user defined regions. For example, thermodynamic integration (TI)[162,163] and Umbrella sampling[164] where the sampling can be re-weighted in order to recover the unbiased probability distribution.

2) The phase space searching method, where the Hamiltonian is modified under the time-dependent terms, progressively builds the time-dependent biased potential. The biased potential basically neglects the system to return back to its previously sampled configuration. The main disadvantage of the method is that although the sampling is many orders of magnitude faster than the normal MD to find the properties of interest but it suffers the scaling with the dimensionality of the system. The available popular phase space searching methods are metadynamics (MetaD),[165] replica exchange or/parallel tempering.[166]

In the next section, only phase space searching and sampling methods which are used in the thesis will be discussed thoroughly.

2.3.2 Umbrella Sampling

So far, many methodologies have been introduced of how to estimate the conformational free energy change in complete phase space. Umbrella sampling[164] is one of them where the system has to combine with a bias potential to overcome the kinetically trapped free energy minimum to the next local or/global minimum. As it is already mentioned, due to the ruggedness of the free energy landscape, MD simulation suffers from sampling issues.

The free energy along some reaction coordinate (z) is connected with the probability distribution (see Eqn. 2.14) where $\rho(z)$ is the unbiased probability distribution of the reaction coordinate (z). Introducing a bias to the Hamiltonian of the system, one evolves the biased probability distribution $\rho'(z)$, and the overall free energy is defined as follows:

$$F'(z) = -\frac{1}{\beta}\ln \rho'(z) - V_{bias} + C \qquad 2.16$$

where C is an arbitrary constant and V_{bias} is the harmonic bias potential to overcome the barrier height and is define as follows:

$$V_{bias} = k_f(r - r_0)^2 \qquad 2.17$$

The umbrella potential can be obtained in such a way where both states can be sampled in a single run. But if the sampling along the chosen reaction co-ordinate is not improved enough, one has to divide the whole reaction co-ordinate in several small windows and sampled each window with different umbrella potential. The overall free energy can be obtained by summing up the unbiased results of the individual windows. An efficient way to count for the overall free energy is to the equation of weighted histogram analysis method (WHAM).[167]

2.3.3 Metadynamics

Understanding the process of a complex reaction mechanism or sampling the rare events in complex polyatomic systems ranging from chemistry to biology to solid-

state physics, has been an ever-growing interest introduced powerful methodologies. Metadynamics (MetaD)[165,168] is one of them which is a class of method that enhances the sampling of a molecular dynamics simulation and help to reconstruct the free energy profile as a function of few selected degrees of freedom, specially called as collective variables (CVs). Basically MetaD is an algorithm capable of both eliminating the metastability and reconstructing the FES.

The MetaD sampling is favored by the introduction of an additional bias potential that acts on those selected number of CVs. The external history dependent bias potential is added to the Hamiltonian of the system. The potential can be easily written as a sum of the Gaussian deposited along the system trajectory in the CVs space, not to return to the previously visited points in the configurational space.

The history dependent bias potential at a certain time *t* can be written as follows:

$$V(S,t) = \int_0^t dt' \omega \, e^{-\sum_{i=1}^{d} \frac{(S_i(r) - S_i(r(t')))^2}{2\sigma_i^2}} \qquad \text{where } \omega = \frac{W}{\tau_G} \qquad 2.18$$

which is the sum of the Gaussian functions deposited every τ_G time interval with a Gaussian height W and the width σ_i. At last when the simulation converges, the bias potential provides the unbiased free energy which is given as follows:

$$V_G(S, t \to \infty) = -F(S) + C \qquad 2.19$$

where *C* an irrelevant additive constant. The correctness and accuracy of the MetaD method has already been tested with biologically relevant molecules so far.[169] The free energies obtained with MetaD were compared with other free energy methods on the complex systems such as docking. Parrinello et al. showed that MetaD, a flexible method can be successfully adapted to docking where it easily found the docked geometry and the overall binding affinity and the process of binding.[170]

In the thesis, the standard MetaD has been used for searching of the base pair association free energy minima on the silica surface[2] (see chapter 3). The major advantage of MetaD over the MD is the following:

1) MetaD can sample the rare events which are far from the local free energy minima.

2) Not a priori knowledge is required before start to the MetaD simulation. It explores always the local or/ low free energy minimum first.

3) MetaD can easily search the free energy minima of interest depending on the chosen degrees of freedom or CVs.

4) MetaD is easily parallelizable with multiple cores in CPU's and now it is also available with GPU's too.

In contrast with the advantages described above, some of the major drawbacks are as follows:

1) The bias potential does not converge and constantly overfills the FES where it has been sampling for long. The bias potential also pushes the system towards the high energy regions in the CVs space.
2) The time is continuous; it is not known when to stop the simulation.

To stop the overfilling the FES, the MetaD community introduced well tempered MetaD (WT-MetaD),[171] where the bias potential is combined with the simulation time and it is inversely proportional to the deposition rate. Therefore, as the free energy minimum is nearly filled, the algorithm try to escape the already sampled free energy minimum and shift to the other minimum with the lowest free energy path.

The calculated biasing potential in the WT-MetaD simulation is as follows:

$$V_G(S,t) = \int_0^t dt' \omega \tau_G e^{-V_G(S(r(t'),t'))/\Delta T} e^{-\sum_{i=1}^{d} \frac{(S_i(r)-S_i(r(t')))^2}{2\sigma_i^2}} \quad 2.20$$

where ΔT is the user defined parameter related to the dimension of the temperature and the corresponding free energy surface is defined as follows:

$$F(S,t) = \frac{-T+\Delta T}{\Delta T}\big(V(S,t) - C(t)\big) \quad 2.21$$

Therefore, two main features have been introduced here

1) The deposition is proportional to the 1/t which helped system not to overfilled FES for long. Therefore, all the microscopic properties progressively trying to reach as much as closer to the thermodynamic equilibrium of the system.

2) From the Eqn. 2.20, it is clear that when $\Delta T \rightarrow 0$, ordinary MD is recovered, where as for $\Delta T \rightarrow \infty$, which corresponds to standard MetaD. In between, the FES can be tuned with the parameter ΔT.

There are few biased methods which also follow the same methodology but tuned in a better way and showed to be converged much faster than the WT-MetaD simulation, such as MetaD with the well tempered ensemble (WTE)[172] and parallel tempering MetaD (PT-MetaD).[173] Parrinello et al. simulated folding free energy of the 16-residue C-terminal fragment of protein G-B1 in explicit water. It was shown that using PT-MetaD, the simulation converged in a much shorter time relative to standard MetaD. The replica exchange algorithm can easily switch between the high and low temperature replica. It was also shown lately that when a WTE is combined with the PT-MetaD,[173] called WTE-PT-MetaD,[174] the simulation converged in a much faster rate than others.

2.3.4 Choice of Collective Variables

The choice of proper degrees of freedom or the so-called CVs is always a non-trivial task, which usually requires substantial insight into the studied system and also some experience (i.e., several trial- and -error experiments). The proper CV includes all the information about the relevant changes in the system of interest (the slow or the rare events) without taking into account any microscopic details of the system dynamics.

One should consider the following points before choosing any CV for running MetaD.

1) The CVs should be defined in such a way that it must distinguish all the different states (initial, intermediate, and also the final state) of the system.
2) It must include all important slow degrees of freedom (dynamical motions) that contribute to the process of interest of the system.
3) The chosen CVs must not be too many in number.

2.4 Alchemical Free Energy

The Free energy change is one of the most important thermodynamic properties that drive most biochemical processes. Therefore, the accurate finding of the free energies of the corresponding changes is very much essential for quantitative

understanding. For small molecules, the exploration of the complete phase space is within the time limit which brings more accuracy in the calculation of the free energies. But for larger systems like drug binding to the protein, the sampling problem occurs which is due to the insufficient overlap of the phase space densities. Sometimes, the mentioned problem can questioned about the accuracy and convergence of the free energy.

There are few equilibrium methods introduced so far, the free energy perturbation (FEP) is one of them which was first introduced by Zwanzig.[175] Later, these methods have been widely used in calculating the free energies of ligand binding to the protein. However, FEP method suffers severe sampling problem mentioned above which needed excessive sampling, whenever the system perturbation is very high in number.[176-178] Except FEP, there are few equilibrium methods which were developed to calculate free energies accurately in the recent years. The slow growth thermodynamic integration (SGTI) and discrete thermodynamic integration (DTI) are widely used in the field of computational chemistry. In the present work, only the non-equilibrium work (NEW) simulations are used. However, a brief introduction of SGTI and DTI is required to understand more about the NEW simulation.

2.4.1 Thermodynamic Integration

Thermodynamic Integration[162,163] is the typical free energy method which is used to calculate the respective free energy difference between the two given states. In this method, the free energy difference is calculated by defining a suitable thermodynamic path between the considered states of interest and integrating over the ensemble averaged potential energy changes along the path.

Let us consider two states A and B with the potential energies U_A and U_B, respectively. The potential in either system can be evaluated as an ensemble average over the configuration sampled in a MD or Monte Carlo (MC) simulation with the proper Boltzmann weighting. The new potential energy along the reaction path can be defined as follows:

$$U(\lambda) = U_A + \lambda(U_B - U_A) \qquad 2.22$$

where λ is defined as the coupling parameter with a value from 0 to 1, and thus the overall potential energy varies from the energy of the system state *A* (λ =0) to system state *B* (λ =1) or vice versa. The overall changes in the free energy between states *A* and *B* can thus be computed from the integral of the ensemble averaged of the derivatives of potential energy over the range of coupling parameter λ.

$$\Delta F(A \rightarrow B) = \int_0^1 \left\langle \frac{\delta U(\lambda)}{\delta \lambda} \right\rangle_\lambda d\lambda \qquad 2.23$$

2.4.2 Equilibrium SGTI and DTI

The SGTI method is based on the TI method where only one long equilibrium trajectory is used to calculate the free energy. Therefore, in this case only one forward (A→B) and one reverse (B→A) trajectory is available. Due to the long trajectory, a longer integration time is sampled to get a high statistical convergence via the work hysteresis of the forward and the reverse simulation using Jarzynski equality.[179,180]

$$W_\tau = \int_0^1 \frac{\delta H_\lambda}{\delta \lambda} d\lambda \qquad 2.24$$

where W_τ is the work over a switching process of arbitrary length τ which could be very short. In case of DTI, like SGTI, only one trajectory both in the forward (A→B) and reverse (B→A) direction is used. The integration is considered over discrete summation of the ensemble average.

$$\Delta F_{DTI} = \int_0^1 \left\langle \frac{\delta H}{\delta \lambda} \right\rangle_\lambda \approx \sum_i^N \left\langle \frac{\delta H}{\delta \lambda} \right\rangle_{\lambda_i} \Delta\lambda \qquad 2.25$$

The statistical accuracy of ΔF_{DTI} is computed via a Gaussian error propagation of the error contributions from each trajectory, calculated by block averaging method.

2.4.3 Non-equilibrium Fast Growth TI

Among all the established free energy methods, the NEW simulations have been given preferences over the equilibrium work simulations. This is because of the difficulties in decreasing phase space density overlap which requires excessive sampling. In this study, the alchemical free energy simulation has been performed

using Crooks Fluctuation Theory (CFT). According to CFT method,[181] the work distributions for forward $P_f(W)$ and reverse $P_r(-W)$ transitions obey

$$\frac{P_f(W)}{P_r(-W)} = e^{\beta(W-\Delta F)} \quad 2.27$$

It relates the free energy differences ΔF to the work distributions in the forward and reverse ensemble of non-equilibrium transitions.

According to Eqn. 2.27, the ΔF is the work W for which $P_f(W) = P_r(-W)$, i.e, the probable intersection point of the corresponding two mentioned work distributions in the forward and reverse direction. It was shown that the work distribution $P_{f,r}(W)$ can be approximated by a Gaussian distribution.

$$P_{f,r}(W) \approx \frac{1}{\sigma_{f,r}\sqrt{2\pi}} \exp\left[-\frac{(W-W_{f,r})^2}{2\sigma_{f,r}^2}\right] \quad 2.28$$

where $W_{f,r}$ and $\sigma_{f,r}$ are the means and the standard deviations of the work distributions, respectively. The index f denotes the forward ensemble, where λ = 0→1, and r the reverse ensemble, where λ = 1→0, respectively. It should be noted that for $\sigma_f = \sigma_r$ the unique solution for the ΔF is close to $(W_f - W_r)/2$ whereas, for $\sigma_f \neq \sigma_r$ generally two intersection points can be found. The first one is always can be expected in between the W_f and W_r and therefore, very close to the $(W_f - W_r)/2$. However, the second intersection point sometime located far in the extreme tail region and therefore produces artifacts. To estimate statistical accuracy in the ΔF calculations, we followed the Monte Carlo approach introduced by Goette et al.[182]

2.4.4 Free Energy of Mutation

Free Energy of Mutation (FEM) is the alchemical change in the free energy upon mutation of an atom or group to another group or atoms. To compute FEM is always a nontrivial task in biology or chemistry, for example, the FEM of a ligand to its analogue in explicit water or inside the protein cavity exploit nonphysical pathways over thermodynamic cycles involving particle introduction or annihilation. Such alchemical transitions require the modification of the classical non-bonded potential energy terms by applying soft-core potential functions to avoid the singularity

prospective. Moreover, strong repulsive force may appear as a numerical instability upon mutation when the interatomic distances between the atoms decreases and becomes very small. Therefore, to avoid this instability one should modify the integration time step during the alchemical change.

To avoid the singularities and numerical instabilities, Zacharias et al.[183] proposed a method to scale and shift the Lennard-Jones (LJ) potential such that at a very short distance, the LJ repulsion between two atoms converges to a finite value. Beutler et al.[184] proposed the similar shifting of the LJ potential which was also fitted and applied to the Coulomb potential for the electrostatic interactions. All the simulations in the work were performed with Gromacs simulation package[185] where the Beutler et al. soft-core potential is implemented for correcting the hard-core potential. In gromacs, the soft-core potential does scales the r_{ij}^{LJ} and r_{ij}^{Q} for the LJ and Coulomb interactions, respectively as follows:

$$V_{SC}(r) = (1-\lambda)V^{A}(r_A) + \lambda V^{B}(r_B) \qquad 2.29$$

$$r_A = (\alpha\sigma_A^6\lambda^p + r^6)^{\frac{1}{6}} \qquad 2.30$$

$$r_B = (\alpha\sigma_A^6(1-\lambda)^p + r^6)^{\frac{1}{6}} \qquad 2.31$$

where $V^A$and $V^B$are the normal hard-core van der Waals or electrostatic potentials in state A ($\lambda = 0$) and state B ($\lambda = 1$) respectively, α is the soft-core parameter, p is the soft-core λ power, σ is the radius of the interaction.

CHAPTER 3

3 Project

3.1 Adsorption of Organic Electron Acceptors on Graphene-like Molecules: Quantum Chemical and Molecular Mechanical Study

3.1.1 Introduction

In the recent years, an intense research interest has been given to the graphene since it was first experimentally discovered in 2004 as a fine semiconductor based nanoelectronics.[186-188] The reason behind working with graphene is because of its excellent transport property due to its delocalized π-electrons. The extra π-electrons can be used as a donor and upon adsorption of an organic electron acceptor molecule can notably modify the electronic properties of the graphene surface.[189,190] It is well known that adsorption of an electron acceptor molecule ends up with a P-type doping and increase of the electron transportation by many orders of magnitude. So far, there have been a large number of studies found in the literature, where the range of the adsorbate particles spanned from metal ions to biomolecules such as ds-DNA,[191-193] ss-RNA, etc. Therefore, to tune the electronic properties of the graphene and also to see deep insight into the behavior of the adsorbate molecules, a proper study of the corresponding energetics, thermodynamic properties were needed.

The extent of the electron/charge transfer (CT) in the graphene...organic electron acceptor complexes mainly depends upon the quality of the adsorbate and also the surface coverage.[194-196] The charge transfer already contributes a large portion in the stabilization of the donor-acceptor complexes. The adsorbate quality mainly represents how strongly it sucks the electron (strong electron deficiency) from a highly reached electron donor species. To calculate charge transfer is already a non-trivial task. The charge transfer term is covered in the induction energy.

To understand the binding, we need to know how the extent of the finite size of graphene affects the interaction energy upon adsorpotion to a organic electron acceptor molecule. In this study, the graphene surface is modeled by carbon aromatic

molecules of increasing size, starting from coronene (C2), circumcoronene (C3), dicircumcoronene (C4), and ending with the C5 (one extra layer of benzene ring than C4) system. The tetracyanoethylene (TCNE) and tetracyanoquinodimethane (TCNQ) is taken as organic electron acceptor.

Further, we have tested the performance of the Wave Function Theory (WFT), DFT and MM methods of above mentioned set of graphene...organic electron acceptor complexes. A few WFT methods are employed such as MP2.5,[149] SCS-MP2[147] and SCS-MI-MP2[197] methods. From the DFT family, BLYP-D3[198] is used with the Grimme's dispersion correction. The semiempirical DFTB-D[199-203] and MM (AMBER)[204,205] methods are also used for the estimation of the stabilization energies. Besides, stabilization energies, binding free energies were also performed using MM and DFTB-D methods.

Overall, this study aims the following:

1) To evaluate the performance of various computational methods including the DFTB-D and MM ones, for graphene…organic electron acceptor complexes.

2) To elucidate the role of charge transfer on the stabilization of the graphene…electron acceptor complexes.

3) To calculate binding free energies using the DFTB-D and MM methods.

3.1.2 System Preparation

The graphene molecules are modeled as polycyclic aromatic molecules with an increasing number of the aromatic (benzene) ring. The model systems C2, C3, C4 and C5 are abbreviated as CX (X = 2-5), adsorbed systems TCNE and TCNQ, abbreviated as TCNY (Y=E,Q), and the resulting complexes are abbreviated as CXY. The complexes were prepared by stacking the electron acceptor to the center of the carbon aromatic molecule in two different structural orientations (“a” and “b”) (*cf.* Figure 3.1). In the C2 cases, only one orientation was considered.

3.1.3 Calculation Strategies

3.1.3.1 Structure and Geometries

The structures of all the complexes (*cf.* Figure 3.1) were optimized at the BLYP-D3 level. These geometries were used for interaction energy calculation at the WFT

level. Further, the DFT-D3 optimized geometries were reoptimized at the DFTB-D and MM (AMBER) level for the stabilization energy, binding free energy calculation in the corresponding level of theory.

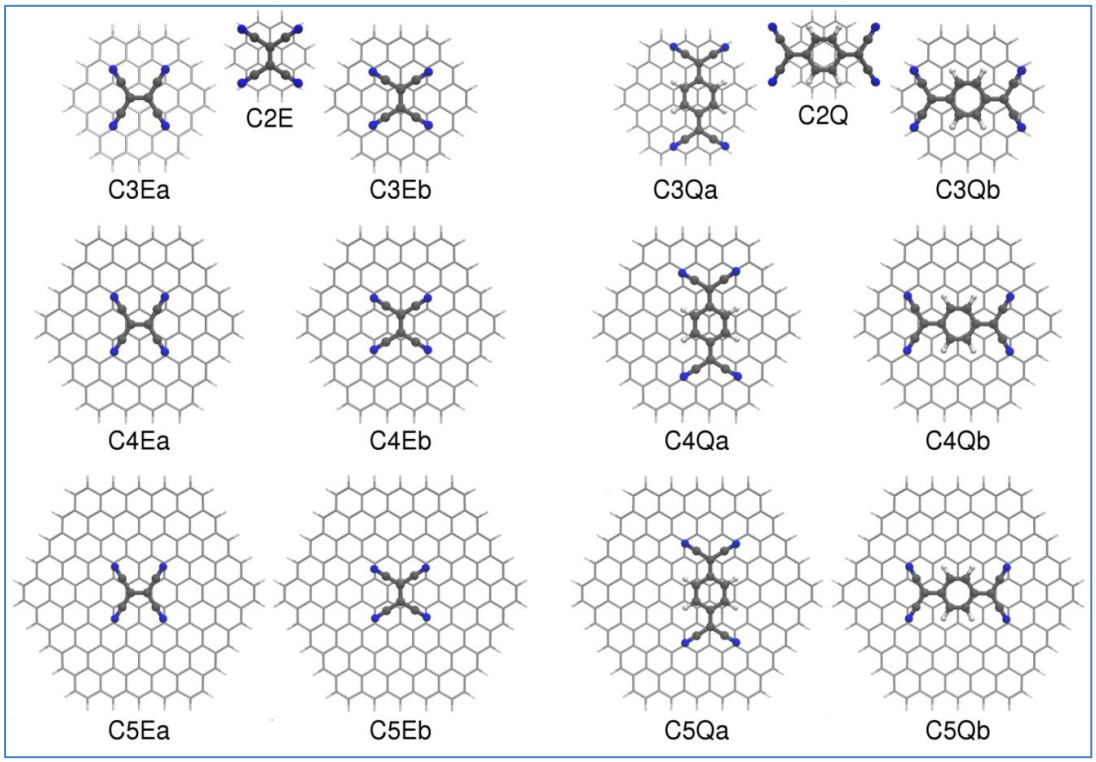

Figure 3.1: Top view of all the investigated C2, C3, C4 and C5 graphene…organic electron acceptor complexes.

3.1.3.2 Interaction Energy

The interaction energies of the complexes studied were determined at the following level of theories

1) Scaled MP2 methods.
2) Dispersion-corrected density functional theory (DFT-D3)
3) Dispersion-corrected density functional tight binding (DFTB-D)[199-203] and
4) Empirical force field.

The interaction energies were calculated according to the supermolecular approach shown in Eqn. 2.1. The interaction energies based on the scaled MP2 methods were corrected for the BSSE using the standard counterpoise procedure.[122] The subsystem deformation energies were not included. For the DFT-D3 calculations, BLYP-D3/TZVPP level of theory were adopted with the Grimme's 3rd generation empirical

dispersion correction.[63] The empirical potential calculations were based on the general AMBER force field (GAFF).[204,205] For the GAFF calculations, the RESP approach[86,87] was used for partial charge calculations which was performed onto a grid of electrostatic potential points calculated at the HF/6-31G* level.

For the C2E and C2Q complexes, the benchmark calculations were performed at the MP2.5/CBS/6-31G*(0.25),[149] SCS-MP2[147] and SCS-MI-MP2[197] level of theory. For MP2.5 method, the MP2/CBS interaction energy was extrapolated from aug-cc-pVDZ and aug-cc-pVTZ basis sets, while the MP2.5 correction term was determined with the 6-31G*(0.25) basis set. The advantage of using SCS-MI-MP2 method over MP2.5 calculation is that it consumes less CPU-time. The interaction energy calculations were performed with Turbomole 6.3 package.

3.1.3.3 Charge Transfer Characteristics

The CT energy contribution was determined by the second-order perturbation theory analysis of the Fock matrix in the NBO basis[206] as follows:

$$E^{CT} \simeq E^{(2)}_{\sigma\sigma^*} = -2\frac{\langle\sigma|F|\sigma^*\rangle^2}{\varepsilon_{\sigma^*}-\varepsilon_{\sigma}} \qquad 3.1$$

where F is the Fock matrix element between the σ and σ^* NBO orbitals, and ε_{σ^*} and ε_{σ} are the energies of the σ and σ^* orbitals, respectively.

3.1.3.4 Free Energy Calculations

RR-HO-IG Approximation: The association free energies were calculated using rigid rotor-harmonic oscillator-ideal gas (RR-HO-IG) approximation at standard room temperature and a pressure of 1 bar as follows:

$$\Delta G = \Delta E + \Delta ZPVE + \Delta H(0 \rightarrow T) - T\Delta S \qquad 3.2$$

Various implicit solvation models were used. For the BLYP-D3 calculations, the COSMO implicit solvation model was employed.[207] For the DFTB-D calculations, the SMD model by Truhlar[208] was applied, whereas, for the MM calculations, the generalized Born (GB) solvation model with parameters by Hawkins et al[209] was used.

Potential of Mean Force: The association free energies calculated using potential of mean force (PMF) method was compared of with results obtained from RR-HO-IG approximation. The advantage of using PMF free energies is that it comes from a statistical average of a sufficiently long equilibrated ensemble.

For all the PMF calculations, graphene...organic acceptor complexes were placed in a cubic box whereas for periodic graphene, a rectangular box was considered. The size of the box varied with the size of the graphene molecule. The box was prepared in such a way that the distance of the solute from the box wall was at least 10 Å. The aqueous environment was modeled either by the GB implicit solvent[209] model or the TIP3P explicit water model.[210] The entire box was well minimized to remove any possible close contacts and gradually heated to 300K during a 20 ps long simulation, keeping the box volume constant. The density of the system was equilibrated by a 1 ns long simulation employing the biasing umbrella potential between the graphene and the CT acceptor of a strength of 2.39 kcal/(mol.Å^2) with the equilibrium Z-distance 5.5 Å. A pressure of 1 bar and a temperature of 300K were maintained by a Berendsen barostat[211] and thermostat,[211] respectively. For solvent and gas phase PMF free energies, a total 45 and 65 sampling windows were taken into account and distance spanned from 2.2 to 11.0 Å and 2.2 to 15.0 Å, respectively. Each umbrella windows were equilibrated up to 3 ns using Nose-Hoover thermostat[212,213] and parrinello-rahman barostat[214] to generate a isothermal-isobaric (NPT) ensemble. All the explicit water simulations used the particle-mesh Ewald (PME) algorithm[215] with a 10.0 Å direct space cutoff for treating electrostatics and the plain 10.0 Å cutoff for the Lennard-Jones (LJ) interactions. Periodic boundary conditions (PBC) were applied for all the PMF simulations. Only the third nanosecond of each umbrella windows was considered for the PMF evaluation while the first, second, and third nanoseconds were used for the block averaging error analysis. All the simulations were performed with gromacs molecular simulation package.[185]

3.1.4 Results and Discussion

3.1.4.1 Interaction Energies

Table 3.1 shows the interaction energies for the C2E and C2Q complexes calculated at different levels and here the MP2.5 represents the benchmark values. Among different methods used, the SCS-MP2 provides the best values and interaction

energies for both structures lie within 1 kcal/mol to the reference values. Surprisingly, the SCS-MI-MP2 energies differ with the reference values more.

Methods	C2E	C2Q	$\Delta E^{(E-Q)a}$
MP2.5/CBS/6-31G*(0.25)	−16.7	−24.3	7.5
SCS-MP2/CBS	−17.1	−25.0	7.8
SCS(MI)-MP2/CBS	−19.8	−29.0	9.2
BLYP-D3	−14.5	−20.6	6.1
DFTB-D	−11.6	−17.4	5.8
AMBER	−13.4	−22.1	8.7

Table 3.1: Interaction energies of the C2E and C2Q complexes (in kcal/mol). [a]the interaction energy difference between C2E and C2Q complexes.

The BLYP-D3, DFTB-D and MM levels tend to underestimate the complex stabilization. Largest difference was found for the DFTB-D method where it underestimates the result by −5.1 and −6.9 kcal/mol for C2E and C2Q complex, respectively relative to benchmark values. The BLYP-D3 method provides reasonable agreement with the benchmark MP2.5 values and thus the BLYP-D3 method can be used for larger complexes as the reference methods.

Complex	BLYP-D	DFTB-D	MM
C3Ea	−20.2	−13.8	−15.7
C3Eb	−19.3	−13.5	−15.7
C4Ea	−19.7	−14.1	−16.5
C4Eb	−19.6	−14.1	−16.5
C5Ea	−22.9	−14.5	−17.0
C5Eb	−22.0	−14.3	−17.1
C3Qa	−26.9	−21.9	−24.6
C3Qb	−28.6	−21.9	−24.7
C4Qa	−31.2	−23.8	−26.6
C4Qb	−29.2	−23.2	−26.8
C5Qa	−32.2	−24.3	−27.7
C5Qb	−32.8	−24.1	−27.8

Table 3.2: Interaction energies of the C2, C3 and C5 complexes with TCNE and TCNQ (in kcal/mol).

Table 3.2 represents the interaction energies of larger C3, C4 and C5 complexes calculated at BLYP-D3, DFTB-D and MM level, respectively. The complex stabilization is seen to be large, reaching about 20 and 30 kcal/mol in the TCNE and TCNQ cases, respectively. Figure 3.2 shows the dependence of the interaction energy on the graphene size. The interaction energy increases as the size of the

graphene molecule increases.

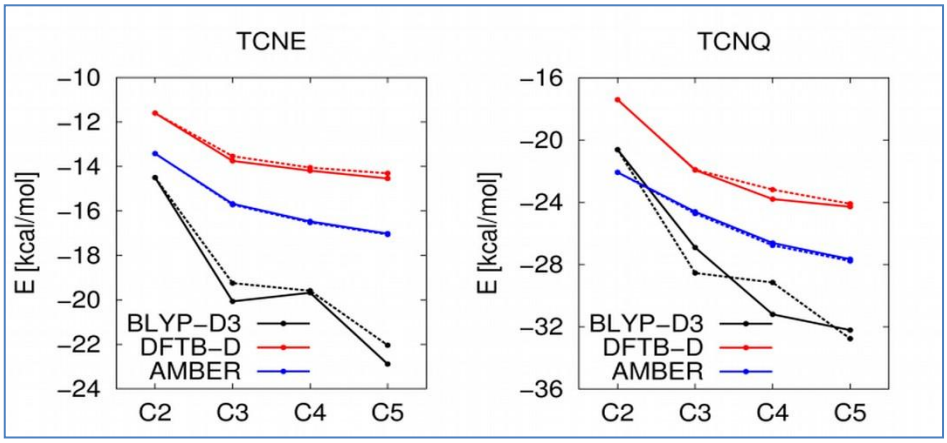

Figure 3.2: The dependence of the interaction energies on the size of the graphene…organic electron acceptor complexes. The full lines stand for the "a" orientations while the dotted lines for the "b" orientations.

The average difference of the interaction energy between the C5Ea and C5Qa complexes is 10.3, 9.8, and 10.7 kcal/mol for the BLYP-D3, DFTB-D, and AMBER methods, respectively. In terms of the benchmark data and the fact that the difference is well converged relative to the graphene size. The value of ~10.5 kcal/mol might represent a good estimate of the true difference of the stabilization of the TCNE and TCNQ complexes with graphene. The interaction energies calculated at DFTB-D and AMBER levels converge slowly and are not too orientation dependent. However, the BLYP-D3 curves are nonmonotonic and more orientation dependent.

3.1.4.2 Charge Transfer

The charge transfer calculations showed that the Mulliken charge transfer value agrees well with the more accurate NBO values. Both Mulliken and NBO charge transfer energies were seen to be nonmonotonic as the interaction energy calculated at the BLYP-D3 level. Therefore, the charge transfer energies are proportional to the interaction energies. The reason behind this nonmonotonicity can be explained on the basis of overlap of the frontier orbitals of electron donor acceptor complexes (shown in Figure 3.3). The HOMO of C3 differs from that of C2 and C4, and this difference is quite large especially at the central benzene ring. Therefore, it can be concluded that the different shape of the HOMO orbital may responsible for bringing unexpected shape of the charge transfer curve and the interaction curve as well.

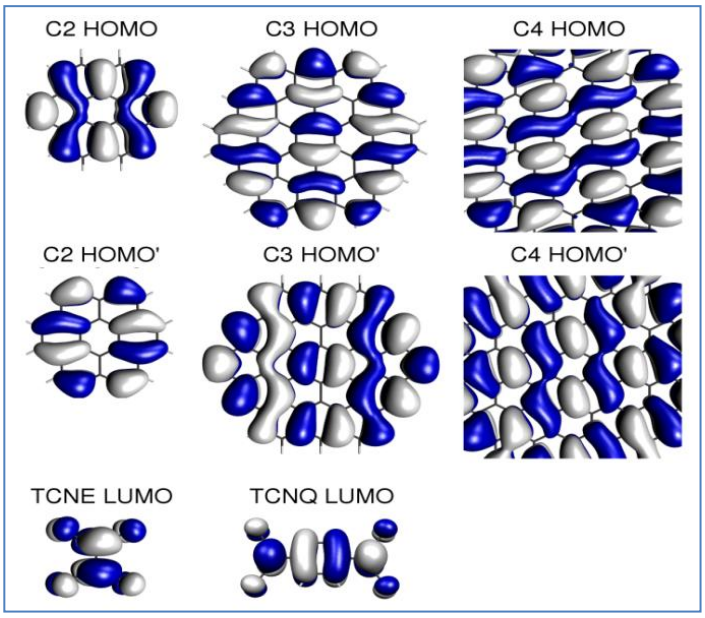

Figure 3.3: Frontier orbital of the molecules investigated. Two energetically degenerated HOMO orbitals are shown for C2, C3 and C4 molecules. The HOMO' orbitals represent HOMO^{-1}.

3.1.4.3 Free Energies

Table 3.3 shows the association free energies calculated at the DFTB-D and MM (AMBER) force field employing RR-HO-IG approximation and at PMF/MM level. The gas phase DFTB-D free energies are less negative than the MM one. The average difference between the interaction energy and association free energy for the DFTB-D and MM level amounts to 3.2 and 2.2 kcal/mol, respectively. This difference is rather small, which indicates that the binding of the complexes studied is governed mainly by the interaction energy, with the other term (such as entropy) not being dominant or compensating for each other. The solvent free energies led to the lower stabilization as compared to the gas phase. The average solvation penalty for CXE complexes at DFTB-D and MM level amount to be 1.4 and 4.4 kcal/mol, whereas for CXQ complexes, it is 2.0 and 5.5 kcal/mol, respectively. Destabilization of complexes studied when passing from the gas phase to solvent is thus at MM level larger than at DFTB-D level. The largest difference was found for C5 complexes where the solvent destabilizes the binding by ~5 kcal/mol.

Table 3.3 also presents the gas and solvent PMF values. The gas phase PMF values are less negative than the respective MM interaction energies by ~5 kcal/mol in average. The PMF Gas as well as Sol values are well converged and the association free energies with periodic graphene amount to −12.3 and −21.1 kcal/mol for E and

Q, respectively. The lower absolute values of association free energies (relative to RR-HO-IG values) are caused by the loss of translational freedom. When two molecules bind, they loss three degrees of freedom, and this loss is captured by the entropy decrease which is also approximately true in the case of RR-HO-IG approximation. However, when a molecule adsorbed onto a surface (e.g., graphene), the number of degrees of freedom which are lost is reduced.

Complex[a]	RR-HO-IG/DFTB-D		RR-HO-IG/MM		PMF/MM	
	Gas	Sol	Gas	Sol	Gas	Sol
C2E	−8.3	−6.2	−9.5	−3.6	−10.5±0.14	−2.7±0.16
C3Ea	−9.6	−8.3	−12.6	−8.6	−11.5±0.28	−4.6±0.25
C3Eb	−11.5	−10.2	−12.6	−8.6	−11.8±0.11	−4.0±0.50
C4Ea	−10.5	−9.3	−13.4	−9.5	−12.2±0.40	−5.2±0.39
C4Eb	−11.0	−9.7	−15.2	−10.3	−12.5±0.16	−5.2±0.17
C5Ea	−10.7	−9.5	−16.8	−12.6	−12.7±0.28	−4.9±0.10
C5Eb	−11.9	−10.7	−16.0	−11.7	−12.7±0.26	−5.0±0.05
PeriodicE	---	---	---	---	−12.3±0.98	−7.1±0.15
C2Q	−12.6	−9.4	−15.6	−10.3	−16.8±0.50	−7.1±0.14
C3Qa	-16.5	−14.2	−19.6	−14.1	−18.9±0.29	−10.2±0.06
C3Qb	−17.3	−14.6	−22.5	−14.1	−19.0±0.23	−9.8±0.37
C4Qa	−18.3	−16.7	−23.3	−17.4	−20.9±0.17	−11.4±0.20
C4Qb	−19.2	−18.0	−22.1	−18.4	−20.8±0.29	−11.7±0.69
C5Qa	−18.9	−17.4	−22.1	−17.3	−21.0±0.27	−12.4±0.14
C5Qb	−20.2	−19.0	−22.2	−17.3	−21.6±0.14	−11.4±0.37
PeriodicQ	---	---	---	---	−21.1±0.77	−13.7±0.69

Table 3.3: Association free energies of all the graphene molecules (C2, C3, C4 and C5) investigated at the DFTB-D and MM level (in kcal/mol), [a]*cf.* Figure 3.1.

The mobility of the organic electron acceptor molecule on the surface was completely revealed from the MD simulation. The visual inspection of the electron acceptor molecules on the surface of graphene shows that it is not fixed and it is more mobile in solvent than in the gas phase. Consequently, it suggests that the actual number of the degrees of freedom lost upon adsorption is not one, as was assumed in the RR-HO-IG approximation, but it is higher. This gives an explanation to the difference between RR-HO-IG and MM values, where the PMF is less negative (more favorable entropy) according to the correct treatment of the E and Q mobility. This fact emphasizes how valuable the data extracted from the MD simulations might be, especially in the cases which are difficult to be tackled by the approximate RR-HO-IG method.

3.2 On the Association of the Base Pairs on the Silica Surface Based on Free Energy Biased Molecular Dynamics Simulation and Quantum Mechanical Calculations

3.2.1 Introduction

The DNA-duplex is considered as one of the most important biomolecules in any living cell because among other functions, it stores and transfer genetic information.[216] In 1953, Watson and Crick[217] discovered the structure and shape of the DNA-duplex and explained its stabilization mostly governed by the inter-molecular H-bonding between the different purine and pyrimidine bases. The guanine (G)…cytosine (C) Watson-Crick (WC) pair has three and the adenine (A)…thymine (T) pair has two hydrogen bonds. However, it was soon realized that besides H-bonding stability, staking (intra- as well as inter-strand) interaction contributes to the stability of the DNA-duplex.[218-220] Further, in our laboratory, it was shown that the methylated analogue of the A…T and G…C pair is most stable in the planner H-bonded form.[221-223] However, passing to the water environment with the several water molecules (microsolvaion) as well as bulk water, the scenario is changed and the stacked structures of mA…mT and mG…mC pairs become the most stable structures.[222]

The question concerning the preferential structure of the base pairs remains when it is absorbed on the solid surface. This question is of immense importance, because it can elucidate the role of the non-covalent interaction in the stabilization of the nucleic acids, formation of large molecules upon polymerization[224-229] and also in the stabilization of the genetic codes. However, it was also found that the genetic materials such as DNA retain its integrity and functionality upon binding with the solid and clay surfaces while maintaining their biological activity.[230]

This work is concerned about the association of the base pairs and their adsorption on the hydrophobic silica surface. The reason behind adopting a hydrophobic silica surface is that it is biologically more relevant clay materials and its interaction with DNA could help us to evaluate the processing of the biomolecules (such as peptide bond formation). To perform the adsorption study, the biased simulation is used, called metadynamics (MetaD). The MetaD[165] method is defined by several collective

variables (CV) and it introduces a finest history dependent biasing potential. The bias potential mainly prevents the system from visiting the regions that it has already explored before. This helps the system escape from the traps of the free energy minima on the free energy surface (FES).

The aim of this work is the following:

1) To investigate the association of the DNA base pairs upon adsorption on the hydrophobic silica surface and its comparison with the previously reported graphene surface.
2) To test the force field based methods over QM methods.

3.2.2 System Preparation

First, a finite 3D hydrophobic silica surface with the methylated (mA…mT and mG…mC) base pairs has been prepared. The silica surface contains 194 atoms and the chemical formula is $Si_{60}O_{10}6H_{28}$ and is taken from the literature.[231] The purine and pyrimidine bases were methylated at the N9 and N1 position, respectively by replacing H-9 and H-1 proton. Figure 3.4 represents a snapshot of the mG…mC base pair sitting on the silica surface taken from MD simulation.

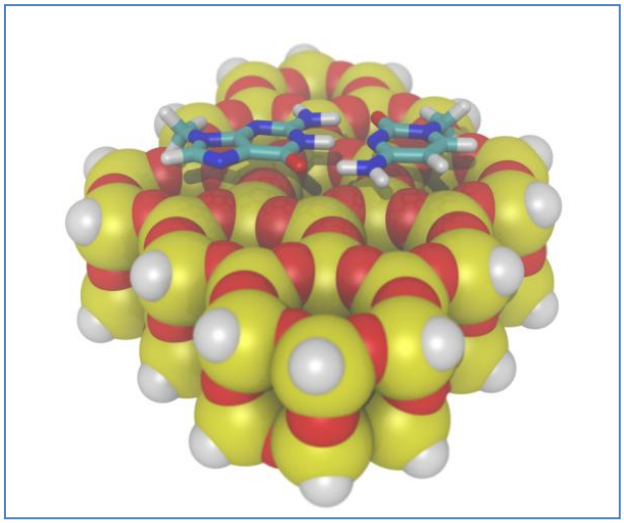

Figure 3.4: Schematic representation of mG…mC base pairs on the silica surface.

3.2.3 Calculation Strategies

3.2.3.1 Structure and geometries

The structures of the base pairs on the surface were obtained from the MetaD simulation. Each structure of surface…mA…mT and surface…mG…mC was taken from the free energy minimum and optimized at the DFT-D and MM level. After optimization the geometrical motifs of all the structures remains the same.

3.2.3.2 Interaction Energy

The gas phase interaction energies were calculated as shown in Eqn. 2.1. The interaction energies of all the complexes were calculated at the following level of theories.

1) Dispersion-corrected density functional theory (DFT-D)[232,233] TPSS[234]/TZVP was used in combination with Jurecka's dispersion.[159]
2) The MM (AMBER) force field.

The empirical potential calculations are based on the general AMBER force field (GAFF). For the GAFF calculations, the RESP approach was used for the partial charge calculations which was performed onto a grid of electrostatic potential points calculated at HF/6-31G* level. Here we introduced DFT (B3LYP/cc-pVTZ)[235] charges also for the gas phase interaction energy calculations. The RESP charges calculated at the DFT level are considered as the more accurate one over HF/6-31G* charges for the interaction energies because the latter lacks electron correlation. The DFT-D calculations were performed in the Turbomole 6.3 package;[236] the RESP charges and MM calculations were performed with the Gaussian09[237] and AMBER packages,[204,205] respectively.

3.2.3.3 Free Energy

Simulation setup: The classical MD as well as the MetaD simulations were carried out with General AMBER Force Field (GAFF).[204,205] For the simulation, the surface including base pairs were placed into a rectangular box of 35x26x24 $Å^3$ along with 661 TIP3P explicit water[210] molecules. The molecules were free to move anywhere in the box, none of them was constrained to its initial coordinated and the box was used solely to confine water molecules. The geometry of the whole system was

optimized to avoid any possible close contacts and gradually heated to 300K during a 50 ps long simulation using an NVT ensemble approach keeping the box volume constant. The initial velocities were chosen randomly from the Maxwell-Boltzmann distribution corresponding to the temperature at 10K. The system was equilibrated to long 50 ns simulation prior start of the MetaD simulation. The equilibration simulation was done with the NPT ensemble employing Nose-Hoover thermostat[212,213] (reference temperature of 300 K, coupling coefficient of τ_T = 0.5 ps) and Parrinello-Rahman barostat[214] (reference pressure 1 bar, a coupling coefficient of τ_P = 0.5 ps). The H-bonds were constrained with the LINCS algorithm.[238] The electrostatics was treated by Particle-Mesh-Ewald (PME)[215] using 10 Å non-bonded cut-off and the same cut-off was used for Lennard-Jones (LJ) interactions too. The MD simulation was performed in Gromacs 4.5.4[185] with a Plumed 1.1.0 plug-in.[239]

Collective variables: The classical MD simulation was followed by a 130 ns long biased MetaD simulation, during which the hills of the biased potential were added at every picosecond (130000 hills in total). Two collective variables were biased during the simulation and they are as follows:

1) The inter-molecular WC H-bonds (S_{wc}), defined in such a way where all the possible inter-molecular H-bonds will be considered wherever possible. The H-bonds CV is defined as coordination number in the following way-

$$S_{wc} = \sum_i^{N_{purine}} \sum_j^{N_{pyrimidine}} \frac{1-(r_{ij}/r_0)^{10}}{1-(r_{ij}/r_0)^{12}} \qquad 3.3$$

2) The inter-molecular stacking (S_{stack}) as the coordination number again and thus it looks like

$$S_{stack} = \sum_i^{N_{purine}} \sum_j^{Npyrimidine} \frac{1-(r_{ij}/r_0)^6}{1-(r_{ij}/r_0)^{12}} \qquad 3.4$$

The S_{wc} acts on the donor/acceptors of WC H-bonding (two pairs of mA…mT and three pairs of mG…mC). S_{stack} was calculated from the inter-molecular distances r_{ij} between the nine ring atoms of purine (i) and the six ring atoms of pyrimidine (j). The reference distance r_0 equal to 0.4 nm (4.0 Å) for S_{stack} and 0.3 nm (3.0 Å) for S_{wc}.

RR-HO-IG approximation: The binding free energies of all the systems were

determined within the applied MM force field using RR-HO-IG approximation. The size of the complexes investigated did not make it possible to use the more accurate QM methods such as WFT methods. The generalized Born (GB) implicit solvent model[209] was used to calculate the solvation free energy. The overall binding free energy of the base pairs on the surface was determined as shown in Eqn. 3.2.

3.2.4 Results and Discussion

3.2.4.1 Geometries

The biased MetaD simulation clearly predicts that in the fully solvated system in presence of silica surface, the base pairs prefer to stay as H-bonded form over π…π stacked form. The chosen two CVs (H-bond and stacked) efficiently describe all the possible H-bonded and silica-purine-pyrimidine π…π stacked assembly. It should mention that when the base pairs are placed on the surface, one translational degrees of freedom of the base pairs are removed (perpendicular to the surface). Therefore, the sampling of the base pair association at the surface becomes much simpler than that in the bulk water. After 130 ns of MetaD simulation, different WC and non-WC base pairs were observed along with the three layer surface...π…π stacked configuration. The FES's obtained after the simulation is shown in Figure 3.5. The different colors at each minimum correspond to the depth of the free energy accumulated by the populations of the base pair association in different configurations. Figure 3.6 depicts the several most probable base pair configurations taken from the FES. A comparison of the base pair associations at the graphene and silica surfaces clearly showed that bases are more flexible upon adsorption on the silica surface. The evidence comes from counting the number of transitions from H-bonded pair to the three layer surface...π…π state. There were total 45 and 40 transitions found for mA…mT and mG…mC pair, respectively, whereas these numbers were smaller for the graphene surface.[240] This also provides evidence of higher flexibility of the silica surface over the graphene one.[240] Mostly two kind of base flipping was seen during the biased simulations and they are WC to the reverse-WC configuration and the silica-purine-pyrimidine to the silica-pyrimidine-purine π…π stacked complex or vice-versa. Most importantly, the base flippings on the silica one also carries the more hydrophobic nature of the surface itself, containing several oxygen atoms, which creates H-bonds with bases during the flipping events.

After long 130 ns MetaD simulation, five and six minima were observed on the FES for mA…mT and mG…mC pair, respectively (*cf.* Figure 3.6).

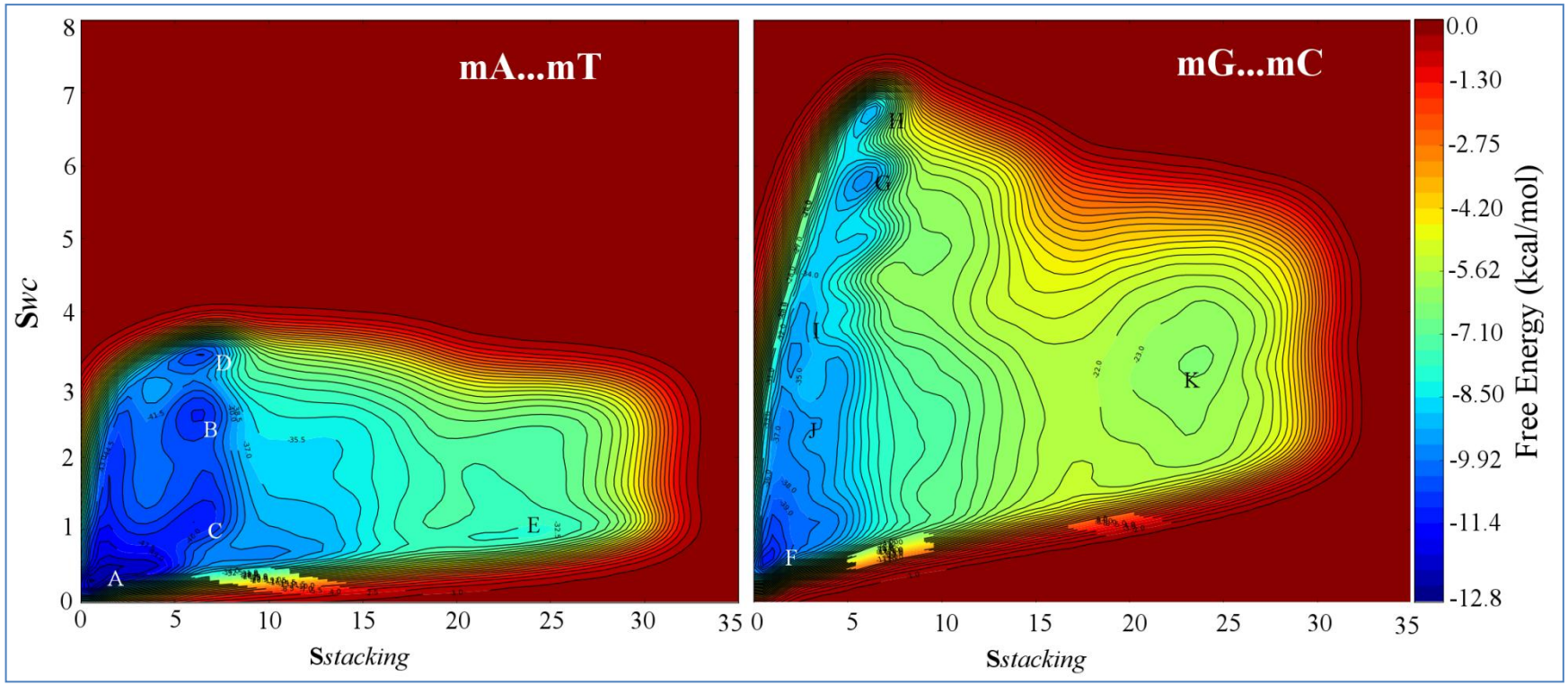

Figure 3.5: The FES of the association of the mA…mT and mG…mC base pairs on the silica surface calculated with the negative of the MetaD biased potential. The axis depicts the collective variables describing base stacking and WC H-bonding. All minima from A–K are highlighted (energies are in kcal/mol).

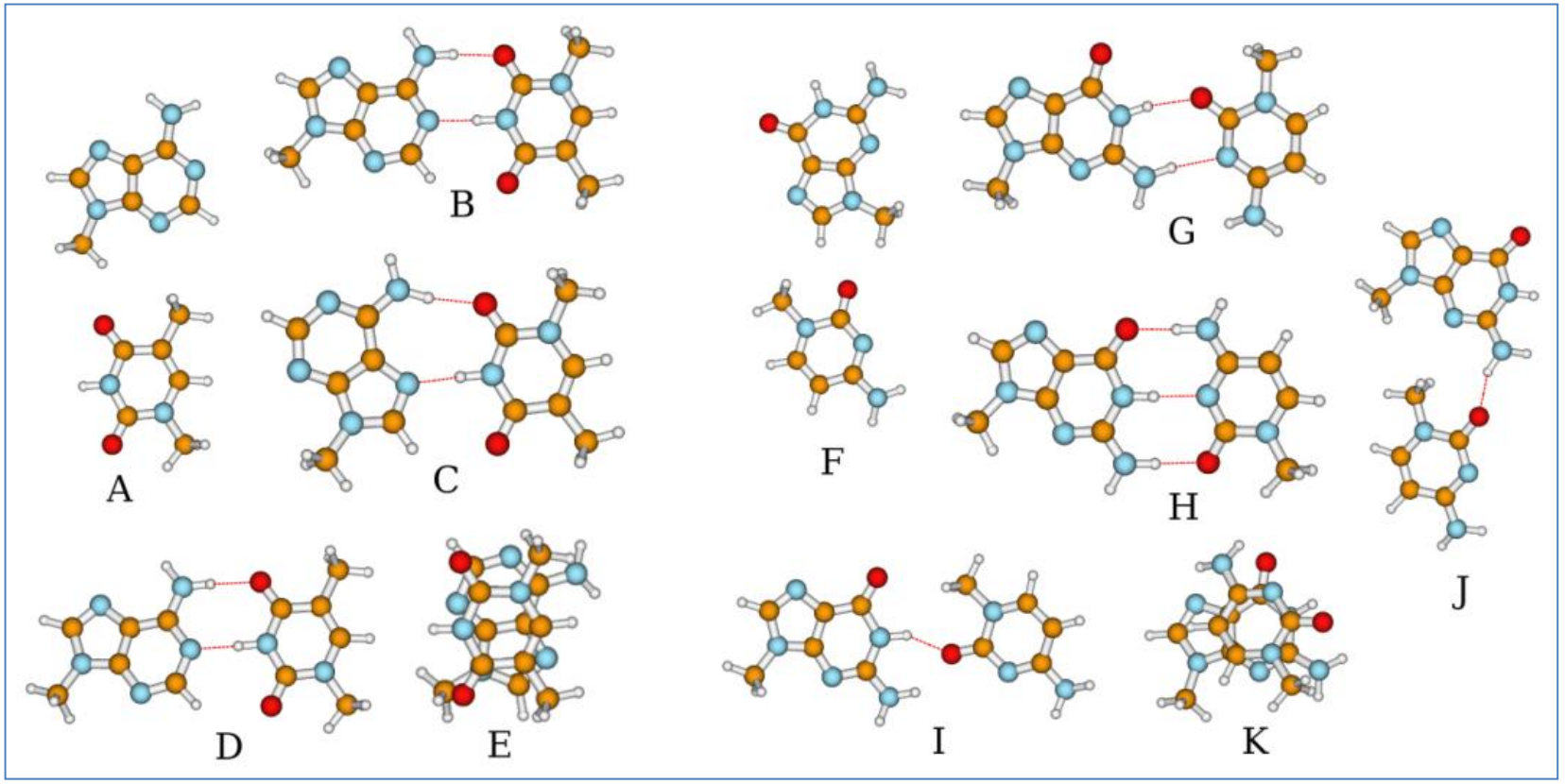

Figure 3.6: All the snapshots (A–K) obtained from MetaD simulation. The minima A–E correspond to the mA...mT base pairs and the minima F–K corresponds to the mG...mC base pairs. The minima D and H represent WC; B is the reverse WC for the mA...mT pair; A and F are the dissociated pairs; E and K depicts the stack pairs for the mA...mT and mG...mC pairs, respectively.

For mA…mT pairs, besides WC base pairing, all the other H-bonded states were found to be more stable than surface...π…π stack states. The WC base pairing (D) is less populated than the reverse WC (B) pair. Another minimum, called C, characterized by an interaction via N7 of the purine base as an acceptor (i.e., Hoogsteen and Hoogsteen-like base pairing) also turned out to be a more effective base pair association than the WC pairing. The minima A, B, and C are all found to be in *cis*-orientation, whereas several stacked pairs (E) were found to be both in the *cis*- and *trans*-orientation. The FES associated with mG…mC pair (*cf.* Figure 3.6) is more complicated than that of the mA…mT pair which is caused by the presence of various H-bonded motifs. For example, the minimum G, is associated with two H-bonds between N1-H of mG and O2 of mC and between N2-H of mG and N3 of mC. The minimum I, is characterized by the presence of two H-bonds between N1-H and N2-H of mG, and O2 of mC. The minimum J, is associated with only one H-bond between N2-H of mG and O2 of mC, which means that the pyrimidine base may flip and adopt different positions. The minima similar to the mA…mT pair are found to be WC H-bond (H) and the surface...π…π stacked conformation.

3.2.4.2 Interaction Energies

Table 3.4 shows the interaction energies of all the clusters calculated at the TPSS-D/TZVP and MM levels.

Systems		TPSS-D	MM
mA...mT	A	–24.3	–32.79
	B	–24.7(–0.4)	–33.45(–0.66)
	C	–23.39(0.91)	–32.94(–0.15)
	D(WC)	–25.27(–0.97)	–33.55(–0.76)
	E	–14.62(9.68)	–19.57(13.22)
mG...mC	F	–23.78	–31.16
	G	–23.67(0.11)	–32.17(–1.01)
	H(WC)	–25.1(–1.32)	–32.42(–1.26)
	I	–24.13(–0.35)	–31.62(–0.46)
	J	–23.75(0.03)	–32.09(–0.93)
	K	–17.11(6.67)	–20.1(11.06)

Table 3.4: Interaction energies of all the snapshots A–K for mA...mT and mG...mC pairs calculated with QM and MM methods (in kcal/mol).

The numbers in parenthesis in column 2 and 3 depict the difference in interaction energy relative to the dissociated state A (*cf.* Figure 3.6) for both the mA…mT and

mG…mC base pairs. The results clearly depict that the average differences in the interaction energies (shown in parenthesis in Table 3.4) calculated for mA…mT and mG…mC pairs with the TPSS-D and MM methods amount to 2.31 and 2.92, and 1.03 and 1.85 kcal/mol, respectively. The difference of the average differences of interaction energies between these two methods for the mA…mT and mG…mC pairs are 0.61 and 0.82 kcal/mol, respectively, which suggests good agreement between the MM and QM methods. Both DFT-D and MM method shows that in case of mA…mT and mG…mC pairs, the WC and non-WC structures, and WC and other H-bonded structures, respectively, are more stable than the dissociated state. This means that both base pairs prefer to stay in the different H-bonded form rather than in the dissociated (stabilized by water-bridging) or stacked forms. At last it has to be mention that the dispersion energies are more pronounced at the graphene surface rather than the silica surface (results are not shown).

3.2.4.3 Free Energy Statistics

Table 3.5 represents free energies of adsorption/binding calculated with RR-HO-IG approximation and MetaD for different configurations of the DNA base pairs on the surface. The empirical potential produces reliable results in comparison with the much more expensive QM results, which means that MM can be successfully used for the description of the base-base interaction on the silica surface. Both RR-HO-IG approximation and MetaD simulation shows that, for mA…mT pair, minimum B (reverse-WC) is more attractive than the WC (D) pair by 0.96 and 0.44 kcal/mol whereas surface...π…π stacking (E) is less attractive by 10.33 and 2.8 kcal/mol, respectively than the WC pair (shown in parenthesis in Table 3.5). In the case of mG…mC pair, the minimum G and J is found to be more attractive by 0.81 and 0.41, and 0.04 and 0.61 kcal/mol, respectively, whereas surface...π…π stacking (K) was less attractive by 8.16 and 2.6 kcal/mol, respectively, than the WC pair (H) for both methods. It is further seen that for mA…mT pair, both calculations prefer non-WC structure over the WC one and in the case of mG…mC pair, minimum G and J is preferred over the WC one. Therefore, the free energy results clearly show that non-WC structures of both base pairs are more stable than the respective WC one. However, the opposite is true when the base pairs are placed on top of a graphene surface.[240]

Systems		MM(RR–HO–IG)$_{sol}$	MetaD
mA...mT	A	–12.88	–12.90
	B	–14.70(–0.96)[–1.82]	–11.07(–0.44)[1.83]
	C	–13.39[–0.51]	–11.36[1.54]
	D(WC)	–13.74[–0.86]	–10.63[2.27]
	E	–3.41(10.33)[9.47]	–7.83(2.8)[5.07]
mG...mC	F	–10.53	–10.22
	G	–12.02(–0.81)[–1.49]	–8.80(–0.41)[1.42]
	H(WC)	–11.21[–0.68]	–8.39[1.83]
	I	–11.20[–0.67]	–8.73[1.49]
	J	–11.25(–0.04)[–0.72]	–9.00(–0.61)[1.22]
	K	–3.05(8.16)[7.48]	–5.79(2.6)[4.43]

Table 3.5: Binding free energies of all snapshots A–K for mA...mT and mG...mC pairs calculated with MM (RR-HO-IG)$_{sol}$ in solvent and the MetaD simulation (in kcal/mol).

The numbers in the squire bracket in Table 3.5 refer to the binding free energy difference relative to the dissociated pair. Analyzing the MM values, it was found that the H-bonded structures were more stable than the dissociated pair, while the opposite is true for stacked structures. Let us add that the same conclusion was drawn in previously calculated base pair adsorption on the graphene surface.[240]

3.3 The effect of Halogen-to-Hydrogen bond Substitution on Human Aldose Reductase Inhibition

3.3.1 Introduction

A halogen bond consists of A-X...Y where X (Cl, Br, I) is an electron acceptor species contain σ-hole and Y is an electron donor (O, N, phenyl ring). The electrostatic potential (ESP) around a halogen atom is strongly anisotropic due to the unequal occupation of the valence orbital. Besides with negative ESP, there is also an area of positive ESP, called the σ-hole.[71,241] In the recent years, an intense research interest has been given to halogen bonds because of its important role in molecular recognition, crystal engineering[242] and interactions of drug with biological molecules.[243-245] The strength of an X-bond depend on the properties of the σ-hole, such as its magnitude and size.[246,247] The magnitude is defined as the value of the most positive ESP of the electron density surface and the size as the special extent of the positive region. The magnitude of the σ-hole increases with increasing the atom size of halogen, i.e. when passing from Cl to I. Furthermore, it has been shown that the σ-hole magnitude correlates with the strength of the X-bond in structurally similar complexes.[246,247]

Apart from the small molecular complexes, the X-bond has been studied also in the protein-ligand (P-L) complexes. It was shown that, a Cl replacement by Br and I has enhanced the X-bond in the cathepsin-inhibitor complex and reduced IC50 from 30 nM to 6.5 and 4.3 nM, respectively.[65] The strength of the X-bond can also be increased by substitutions in the vicinity of the halogens. Electron-withdrawing groups make the σ-hole larger and more positive.[246,247] Therefore, it is possible to tune the inhibition activity by moduling the X-bond strength in a P-L complex. The recent study[248] showed that the IC50 of aldose reductase (AR) inhibitors was reduced from 1900 to 190 nM by fluorination close to the halogen atom that is involved in the X-bond. It should be noted that there has not been any direct correlation between the magnitude of the σ-hole and the inhibition activity, because the fluorination also can change the other properties such as solvation free energy.[248]

The aim of this study is to investigate the effect of the halogen-to-hydrogen bond substitution on the binding energy and biological activity of human aldose reductase

inhibitors. The experimental techniques such as X-ray crystallography and IC50 measurements with the binding free energy calculations have been employed. Here, the X-bond is replaced by an H-bond and it is examined whether this substitution changes the binding mode of the AR inhibitor. For the determination of the binding statistics for the individual AR-inhibitors, a QM based scoring[249] functions is used and the relative free energies obtained was thoroughly compared with the MD based free energy calculations.

3.3.2 System Preparation

To study the effect of halogen-to-hydrogen substitution inside AR, the inhibitors IDD388 and MK315, containing Br and I, respectively involved in the X-bond have been taken into account. Both the inhibitor participates in an X-bond with an oxygen (OG1) atom of the side chain of Thr113 of AR and maintains almost linear arrangement. To introduce an H-bond, one should keep in mind that the replacement must follow the linear arrangement. This was successfully conserved upon introducing a NH_2 group and forms an H-bond with the same acceptor. The respective compound (AK198) with NH_2 substitution in the place of X (Br, I) atom was synthesized and the X-ray structure of the AR-AK198 complex was determined (PDB ID 4QXI) at the ultrahigh resolution of 0.87 Å.

3.3.3 Calculation Strategies

3.3.3.1 Structure and Geometries

The P-L complexes were modeled from the X-ray structure taken from the PDB (PDB code 2IKI).[250] A triple layer QM/SQM/MM scheme[249] has been applied for optimization and binding free energy calculation of the ligand to the AR. The scheme couples the small QM region treated at the DFT-D (TPSS/TZVP//BLYP/SVP) level of theory and the large SQM region treated by PM6-D3H4X[251,252] with an MM description for the rest of the system.

For MD free energy simulation, all three ligands were divided into two fragments (fragment 1 possessing net charge of 0 and fragment 2 with a net charge of –1) (*cf.* Figure 1.4) in order to minimize the extent of alchemical perturbation, and thus to improve the convergence of the resulting free energies. We named this as the

'*Fragment based approach*' (FBA) alchemical free energy calculations which produced relevant results relative to the available QM based results.

3.3.3.2 Molecular Modeling

The binding free energy is approximated by the total score expressed as follows:

$$Score = \Delta E_{int} + \Delta\Delta G_{solv} + \Delta G'^{w}_{conf}(L) \quad 3.5$$

where

$$\Delta\Delta G_{solv} = \Delta\Delta G_{int,solv} + (\Delta G^{low}_{solv}(L) - \Delta G^{high}_{solv}(L)) \quad 3.6$$

Particular terms describe the gas-phase interaction energy (ΔE_{int}), the solvation/desolvation free energy ($\Delta\Delta G_{solv}$), and the change of the 'free' energy of ligand ($\Delta G'^{w}_{conf}(L)$). The gas phase interaction energy (ΔE_{int}) was calculated using the QM/SQM/MM method. The solvation free energy change of the whole complex ($\Delta\Delta G_{solv}$) was determined by the generalized Born solvent[209] model implemented in AMBER 11[204,205] The SMD/HF/6-31G* level[208] was used to calculate the solvation free energy of the ligand in order to increase the accuracy of the $\Delta\Delta G_{solv}$ term, which is the difference between the solvation free energy calculated at high (SMD) and low (GB) level. The $\Delta G'^{w}_{conf}(L)$ term is the "free" energy change between the ligand in the ligand conformation adopted in the P-L complex and its optimal solution structure. For the evaluation of the change in the conformational "free" energy of ligand ($\Delta G'^{w}_{conf}(L)$), the gas-phase DFT-D (TPSS/TZVP//B-LYP/SVP) energy is combined with the SMD solvation free energy.

3.3.3.3 Molecular Dynamics Simulation

The MD simulation has been employed only for the ligand simulation in the explicit water environment. We carried out the mutation of the ligand from the AK198-to-IDD388, AK198-to-MK315, and IDD388-to-MK315 and evaluated the solvation free energy.

Simulation Setup: The general AMBER Force Field (GAFF) in AMBER 11[204,205] was utilized for the ligand parameters. The starting structures of all of the ligands were taken from the respective complex crystal structures (AR-AK198, PDB code 4QXI; AR-IDD388, PDB code 2IKI; and AR-MK315, PDB code 4LB3). The standard

RESP charge set obtained at the HF/6-31G* level was used for the ligand atomic point charges.[86,87] For the I atom, HF with the DGDZVP basis set was used. For IDD388 and MK315 containing Br and I atoms, respectively, we used explicit σ-hole (ESH) with the "aF" (all fit) charges[85] to account for the charge anisotropy of the electrostatic potential on top of the halogen atoms). The ESH charge was calculated to be 0.128e at a distance of 1.2 Å for the Br atom and 0.196e at a distance of 1.32 Å for the I atom (*cf.* Figure 3.7). The Br and I charges were –0.263e and –0.379e, respectively. The ESH was modeled as a mass less virtual site as provided in the Gromacs software package.[185]

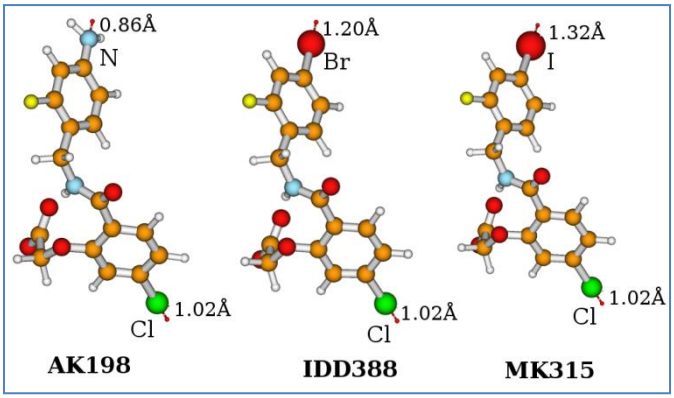

Figure 3.7: Pictorial representation of the AR-inhibitors with the corresponding explicit σ-hole.

For the solvation free energy calculations, 443 TIP3P[210] explicit water molecules were added into a periodic cubic box with dimensions 24.2x24.2x24.2 $Å^3$. The system was neutralized by a single Na^+ ion.[253] The electrostatics were treated by the Particle-Mesh-Ewald (PME) method[215] using a non-bonded cut-off distance of 10 Å. The same cut-off was used for Lennard-Jones (LJ) interactions. For production equilibrium runs, the isothermal-isobaric ensemble was generated employing a Nose-Hoover thermostat[212,213] (a reference temperature of 300 K, a coupling coefficient of τ_T = 0.5 ps) and a Parrinello-Rahman barostat[214] (a reference pressure of 1 bar, a coupling coefficient of τp = 0.5 ps). All the bonds were constrained with the LINCS algorithm.[238] The simulation length was 20 ns with a time step of 1 fs. All the simulations were carried out with the Gromacs (4.5.4).[185]

3.3.3.4 Alchemical Free Energy Simulation

The alchemical mutations were performed as described in ref. 182 and only fragment 1 was mutated (*cf.* Figure 3.8). Fragment 2 was kept identical across all of the

ligands, including atomic charges, bonding and nonbonding parameters. The overall mutation of H-to-X-bond was done in three steps.

1) The initial atomic charges (ensemble A) were turned off to zero charges (ensemble B).

2) The LJ parameters, masses and bonding terms (bonds, angles and dihedrals) were switched from their initial (ensemble B) to their final values (ensemble C) keeping the charges zero.

3) The zero charges (ensemble C) were turned on to their final values (ensemble D). The standard soft-core potentials (Gromacs mdp options sc-power=1, sc-alpha=0.25 and sc-sigma=0.3) were used in all of the steps. The relative solvation free energies were calculated according to a well constructed thermodynamic cycle.

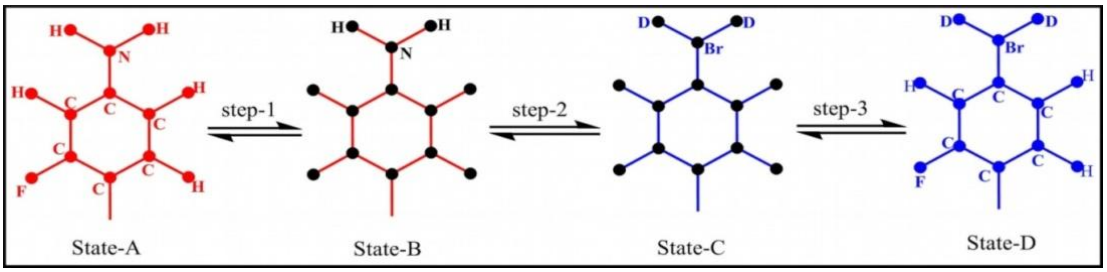

Figure 3.8: Pictorial representation of all the states in a simulation of alchemical free energy of mutation. State-A depicts the complete charged state of a ligand. In the 1st step, the charges are switched to zero to state-B. The black atoms represents the non-charged atoms but within the same bonding and nonbonding environment. To show this the bonds are sketched in red. In the 2nd step, all the bonding and nonbonding parameters are switched to state-C (blue bonds) along with the corresponding masses but the atoms still remains black which means non-charged. Finally in the 3rd step, the atoms are recharged on their final state to state-D. D represents the dummy atoms bonded with bromine. Dummy atoms do not interact with any charge and LJ but contain proper bonding parameters.

The fast-growth thermodynamic integration (FGTI)[162,163] scheme was utilized along with the Crooks Gaussian intersection (CGI)[181] to calculate the free energies. From the equilibrium ensembles A−D (20 ns), we extracted 100 snapshots as starting sates for the non-equilibrium mutations (100 ps each) in solution. For vacuum, each snapshot was further equilibrated to 5 ns prior to the non-equilibrium simulation.

3.3.4 Results and Discussion

3.3.4.1 QM/SQM/MM and MD Geometries

The experimental (X-ray) geometries of all the adopted AR-AK198, AR-IDD388, and AR-MK315 complexes are compared with the geometries obtained upon optimization in the QM/SQM/MM three layer scheme[249] (*cf.* Figure 3.9). The result clearly shows that the QM/SQM/MM structures have accurately reproduced the X-bonded complexes (see Table 3.6). However, the length of the H-bond was about 10% longer in the QM/SQM/MM structure as compared to the X-ray crystal structure. It might be caused by an insufficient basis set used in the gradient optimization. The strength of the H-bond calculated using the larger basis set in single point calculations on the structures with the too long H-bond might thus be underestimated and should be interpreted only as its lower bound.

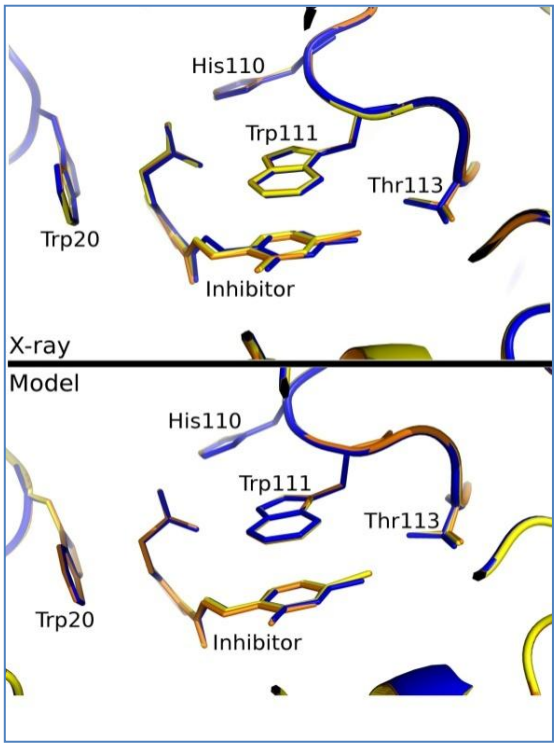

Figure 3.9: Alignment of the AR-inhibitor complexes based on X-ray and QM/SQM/MM optimized structures (coloring code IDD388, yellow; MK315, orange, and AK198, blue).

The $\Delta G'^{w}_{conf}(L)$ term was seen not to be changed much for all three inhibitors and they were close to zero. It is thus considered as the lower bound of the real values of the ligand conformational change as the global minimum cannot be expected to be reached by a single optimization. To support this statement, a long 20ns MD

simulation has been carried out for all three inhibitors in the explicit waters under ambient conditions.

Inhibitor	X	Distance	Angle	
		X...O	C−X...O	X...O−CA
		X-ray		
IDD388	Br	2.90	154.5	132.2
MK315	I	2.95	158.1	130.2
AK198	N	1.92	146.7	124.9
		QM/SQM/MM		
IDD388	Br	2.92	153.1	132.6
MK315	I	2.91	156.6	134.9
AK198	N	3.24	144.4	127.6

Table 3.6: Geometrical properties of the X-bond and H-bond in the studied AR complexes, distances in Å and angles in degrees.

Overall, the molecules were unable to undergo any large conformational change. The radii of gyration (R_g) are comparable and calculated to be 3.91±0.09, 3.96±0.1, and 3.96±0.1 Å for AK198, IDD388 and MK315, respectively. Further, we also calculated the distance between N, Br, and I atoms with the two O atoms of the carboxylate groups. The average distances are calculated to be about 7.5 Å for the N...O, Br...O, and I...O contacts. The present analysis clearly demonstrated that the ligands behave similarly and did not undergo any larger fluctuation.

3.3.4.2 Score

The score value is calculated according to the Eqn. 3.5 and results are summarized in Table 3.7. The ΔE_{int} of AK198 is more negative than that of the halogenated compounds IDD388 and MK315 (by about 4.9 and 1.1 kcal/mol, respectively). In order to see the origin of the change in ΔE_{int}, the interaction energies between the inhibitors and the side chain of the amino acids within the 4 Å of the inhibitor was calculated at the RI-DFT-D/TPSS/TZVP level (results is not shown). Obviously, the H-bond is stronger than the X-bond, and the change in the total ΔE_{int} is caused mainly by weakening the interaction with the residue Thr113. The H-bond with Thr113 was calculated to be approximately 3 kcal/mol stronger than the X-bond. Infact, ΔE_{int} of the X-bond is close to zero (ΔE_{int} with Thr113 of 0.6, –0.4, and –3.1 kcal/mol for IDD388, MK315, and AK198, respectively), which corresponds to the bent C-X...O angle.

Inhibitor	X	ΔE_{int}	$\Delta\Delta G_{solv}$	Score	ΔG_b^o	IC_{50}
AK198	NH_2	−86.9	45.6	−41.3	−8.4	1.3 ± 0.2
IDD388	Br	−82.0	34.3	−47.6	−9.1	0.40 ± 0.02
MK315	I	−85.8	37.8	−48.0	−9.6	0.19 ± 0.09

Table 3.7: Calculated gas-phase interaction energy (ΔE_{int}), interaction desolvation free energy ($\Delta\Delta G_{solv}$), scores, experimental ΔG_b^o (calculated by $\Delta G_b^o = RTln(IC_{50}/2)$, and IC_{50} of the studied AR inhibitors, energetics and IC_{50} in kcal/mol and μM, respectively.

Passing to the solvent, the advantage of the H-bond is offset by the solvation. The SMD solvent model calculation showed that solvation free energy ($\Delta G_{solv}^{high}(L)$) of AK198 is significantly more negative than that of halogenated compounds IDD388 and MK315. They are calculated to be –70.99, –62.65, and –66.28 kcal/mol, for AK198, IDD388 and MK315, respectively. The score value was seen to be less negative for AK198 compared to IDD388 and MK315 by –6.3 and –6.7 kcal/mol, respectively. Whereas, the overall $\Delta\Delta G_{solv}$ term was found to be more positive for AK198 relative to IDD388 and MK315 by 11.3 and 7.8 kcal/mol, respectively. Therefore, the numbers clearly revealed that the lower score value for AK198 come from the desolvation penalty. The score values were seen to be consistent with the IC50 measurements. The IC50 value of AK198 is larger (1.3 μM) than the IC50 of IDD388 and MK315 (0.4 and 0.2 μM, respectively). This trend in the binding affinities is fully supported by the inhibitor occupancy in the X-ray structures. In the case of AR-IDD388 and AR-MK315 complexes, the inhibitors have 100% occupancy, whereas the AK198 inhibitor in the AR-AK198 complex has occupancy of only 56%. The other 44% correspond to the holoenzyme structure, as the B conformation of the Leu300 overlaps with the inhibitor and corresponds to the conformation preceding the inhibitor binding.

3.3.4.3 Alchemical Free Energy

The relative alchemical free energy of mutation ($\Delta\Delta G_{sol}^{mut}$) for the systems AK198-to-IDD388, AK198-to-MK315, and IDD388-to-MK315 were calculated to be 6.90±0.03, 5.85±0.11, and –1.21±0.34 kcal/mol, respectively (see Table 3.8). Methodologically, this represents very well converged values: the thermodynamic enclosure of –0.16 kcal/mol is close to its theoretically correct values 0.0 kcal/mol. These results confirm that the penalty for the ligand desolvation is considerably larger for AK198 than for the two halogenated inhibitors.

AK198-to-IDD388						
Systems	Step-1	Step-2	Step-3	ΔG_{total}	$\Delta\Delta G_{sol}^{mut}$	$\Delta\Delta G_{solv}^{high}(L)$
SOL	46.07±0.06	−1.96±0.03	−9.83±0.05	34.28±0.14		
GAS	37.00±0.05	−1.51±0.06	−8.11±0.06	27.38±0.17	6.90±0.03	8.34
AK198-to-MK315						
SOL	46.07±0.06	−3.15±0.02	−12.52±0.06	30.4±0.14		
GAS	37.00±0.05	−1.90±0.09	−10.55±0.11	24.55±0.25	5.85±0.11	4.71
IDD388-to-MK315						
SOL	9.71±0.32	−1.19±0.07	−12.53±0.04	−4.01±0.43		
GAS	8.11±0.04	−0.44±0.00	−10.47±0.05	−2.8±0.09	−1.21±0.34	-3.63

Table 3.8: The Summary of the alchemical free energy (kcal/mol) change upon mutation of AK198-to-IDD388, AK198-to-MK315 and IDD388-to-MK315, respectively. The $\Delta\Delta G_{sol}^{mut}$ has been calculated according to a well defined thermodynamic cycle.

It is encouraging to see a semiquantitative agreement considering the different approximations in the two approaches. SMD, on one hand boasts for the accurate quantum-based treatment but suffers from the single-conformational approach; on the other hand, FGTI is inherently bound by force-field inaccuracies but has the advantage of a multiconformational treatment. Therefore, both methods provide quantitatively consistent results showing that the amino analog is the best solvent variant, followed by the Br and I analogs. These results also consistent with the $\Delta\Delta G_{solv}$ term of the score values.

3.4 Comparison of *ab initio* Quantum Mechanical and Experimental D_0 Binding Energies for Eleven H-bonded and Eleven Dispersion Bound Complexes

3.4.1 Introduction

In recent years, computational chemistry has made enormous progress in determining the binding energies of non-covalently bound complexes, and the wide application of the CCSD(T)/complete basis set (CBS) method[218,254] has dramatically increased the accuracy of calculated stabilization energies. How accurate are these stabilization energies? Answer to this question is surprisingly not straightforward. It can be relatively easily estimated by making "theory to theory" comparison.[15] For example, the accuracy of about 1–2% was reached for stabilization energies of 66 non-covalently bound complexes from the S66 database[140] by comparing the stabilization energies calculated at two different levels: First, based on the composite scheme of determining CCSD(T)/CBS interaction energies from MP2/CBS[149] interaction energies, extrapolated from aug-cc-pVTZ and aug-cc-pVQZ basis sets, and CCSD(T) correction terms evaluated as a difference between CCSD(T) and MP2 interaction energies, calculated with a smaller basis set (aug-cc-pVDZ), and second, based on the same scheme but using aug-cc-pVTZ and aug-cc-pVQZ basis sets for extrapolation and the aug-cc-pVTZ basis set for the evaluation of the CCSD(T) correction term.[127,255] Such a comparison is straightforward and can be extended for example to higher-electron excitations (Q, P, …) or to the role of relativistic and Born-Oppenheimer corrections.[256] On the other hand, a comparison of theory to experiment[257] is not straightforward even in the simplest case represented by the isolated gas-phase complex at very low temperature. Under such conditions, the binding free energy (ΔG) can be identified with binding energy, which means that entropy, whose calculation is tedious, can be neglected. Nevertheless, the experimental binding energy, D_0, consists of stabilization energy, D_e, and the change of zero-point vibrational energy ($\Delta ZPVE$), which cannot be separated in experimental measurements. The former term can be accurately calculated by e.g. the above-mentioned CCSD(T)/CBS method, whereas the comparably accurate evaluation of the ZPVE requires complex and computationally intensive calculations of vibrational frequencies. The calculations must be performed at a very high level in

order to ensure an accurate description of the potential energy surface (PES). Further, it is inevitable to go beyond the harmonic approximation, which is the main problem. As a result, the determination of the ΔZPVE is the step limiting the accuracy of the calculated D_0.

The aim of the present study is to determine D_0 for 11 H-bonded complexes (having from 7 to 23 atoms) and 11 dispersion-bound complexes (from 12 to 23 atoms) and to compare these values with the experimental ones. The structures of these complexes were determined by gradient optimization at the RI-MP2[258]/cc-pVTZ and B97D3/def2-QZVP[259] levels of theory for the H-bonded and dispersion-bound complexes, respectively. The D_e energies were calculated with the CCSD(T)/CBS method. Finally, the harmonic ΔZPVE term was calculated with the same methods as those used for optimization; on the other hand for the anharmonic ΔZPVE term, the VPT2 theory[260] was applied along with the respective methods used for optimization. The use of the VPT2 method for larger complexes is limited. Therefore, we are trying to find an efficient method for the assessment of the ΔZPVE, which goes beyond the harmonic level, and consider diagonal approximation,[261] where only one-dimensional (1-D) non-coupled anharmonic frequencies are taken into account for the construction of ΔZPVE term.

3.4.2 System Investigated

Altogether 11 H-bonded complexes, depicted in Figure 3.10, was investigated. In seven of them (complexes 1–7) aromatic chromophore containing the polar group (OH, NH) act as a proton donor and water (H_2O), ammonia (NH_3) and methanol (CH_3OH) act as an electron donor. In three complexes (8–10) water act as proton donor and oxygen, fluorine and nitrogen in aromatic chromophore act as electron donor. Finally, in complex (11) double bond in ethylene (C_2H_4) acts as electron donor while H in HCl acts as proton donor. Set of 11 H-bonded complexes was augmented by 11 dispersion-bound stacked complexes which are visualized in Figure 3.11. They all possess aromatic chromophore and an inert gas (Ne, Ar, and Kr). Structures of all the complexes were manually prepared using Molden.[262]

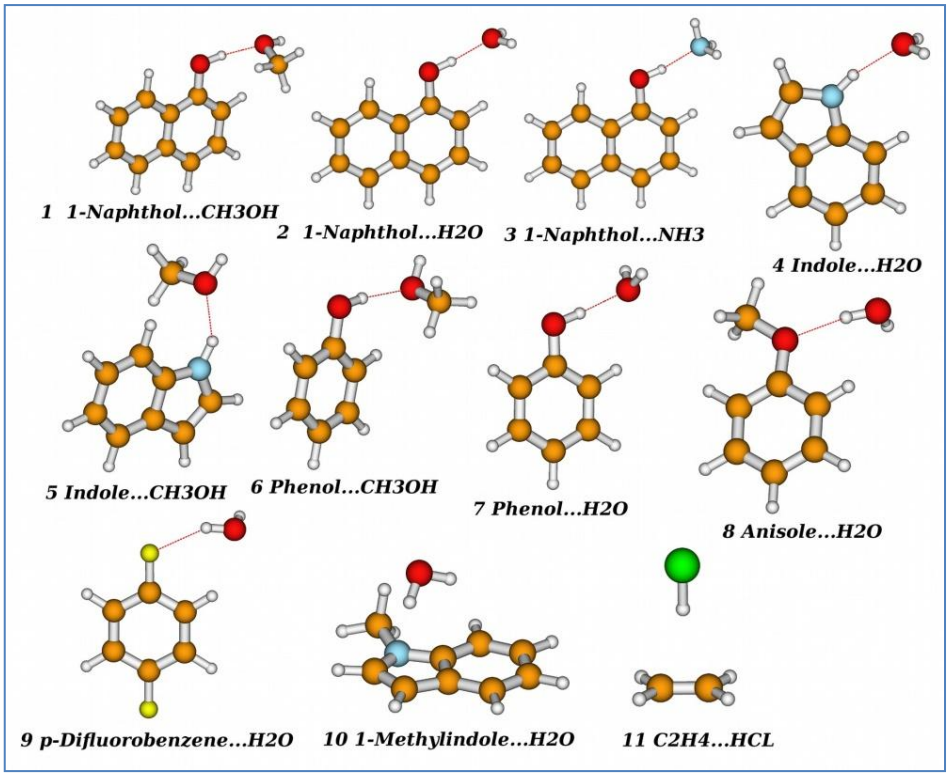

Figure 3.10: Pictorial representation of all the eleven H-bonded complexes investigated.

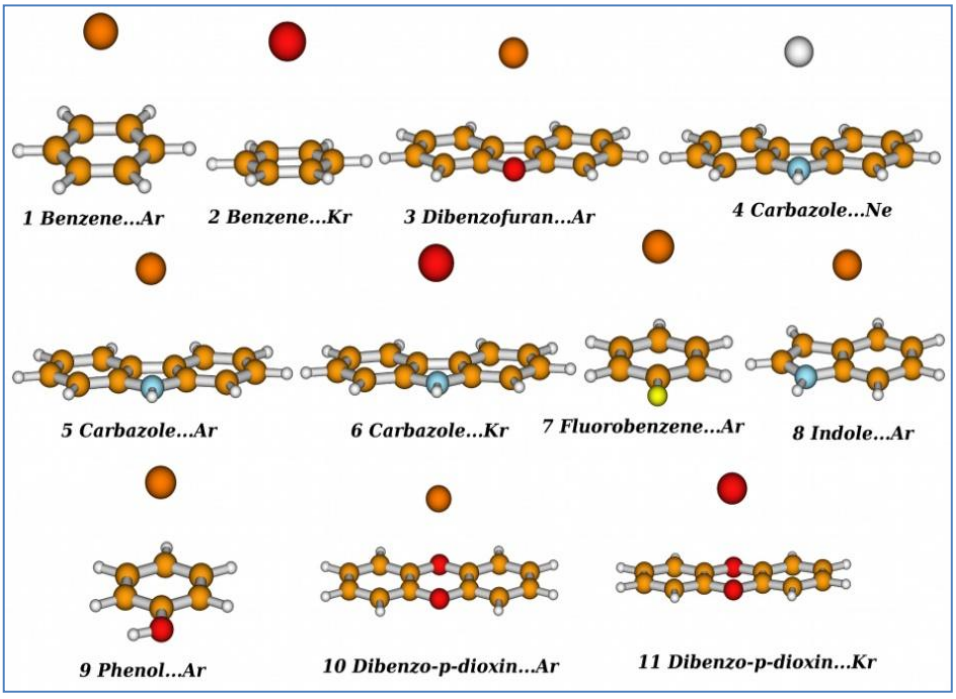

Figure 3.11: Pictorial representation of all the eleven dispersion-bound complexes investigated.

3.4.3 Calculation Strategies

3.4.3.1 Structure and Interaction Energy

Structures and interaction energies of H-bonded and dispersion-bound complexes were obtained by gradient optimization at RI-MP2/cc-pVTZ and B97D3/def2-QZVP levels of theory, respectively. The interaction energies of all the complexes were determined following Eqn. 2.1. Throughout the study the basis set superposition error was eliminated by using the function counterpoise (CP) procedure introduced by Boys and Bernardi.[122] Deformation energy is important for H-bonded complexes but not for dispersion-bound systems and it was thus considered only for the former complexes. Finally, the accurate interaction energies were determined at the CCSD(T)/CBS level of theory shown in Eqn. 2.4. The 2nd term in Eqn. 2.4 ($\Delta E(MP2)$) is determined by extrapolation from aug-cc-pVTZ and aug-cc-pVQZ basis sets using Helgaker et al. Scheme.[138,139] The 3rd term $\Delta\Delta E(CCSD(T))$ was evaluated with aug-cc-pVDZ basis set. All the MP2, DFT and VPT2 calculations were performed in Gaussian09 suite programme[237] with RI approximation (RI-MP2[258]) and CCSD(T) calculations in the Molpro package.[263]

3.4.3.2 Theoretical and Experimental Dissociation Energy

Stabilization energy, (absolute value of interaction energy (eq. 2.1)) which is denoted also as D_e, is not observable. Observable is D_0 energy, which corresponds to stabilization enthalpy, defined as:

$$D_0 = D_e + \Delta ZPVE \qquad 3.7$$

where $\Delta ZPVE$ term denotes the difference in the zero-point vibrational energies between the complex and the subsystems. The RI-MP2/cc-pVTZ and B97D3/def2-QZVP levels has been applied to perform the harmonic $\Delta ZPVE$ calculations for all H-bonded and dispersion-bounded complexes, respectively. We also calculated the anharmonic contribution using second-order perturbation method (VPT2), which approximately covers the anharmonic effects. The use of the VPT2 method especially for larger complexes is limited; here we have tested the use of the diagonal approach, where only the diagonal 1-D anharmonic vibrational frequencies are considered.

3.4.3.3 Diagonal Approach

The 1-D anharmonic vibration frequencies were calculated for all six inter-molecular vibration modes and for the X–H intra-molecular stretching mode, which participates in the H-bonding (X–H...Y). For each mode, we performed a scan using curvilinear valence coordinates and fitted the potential with Morse, sine and cosine functions, for which the vibrational energy levels can be expressed analytically. For more details, see the reference.[261,264,265]

3.4.4 Results and Discussion

3.4.4.1 Stabilization Energies

H-bonded Complexes: Table 3.9 shows the stabilization energies determined at the RI-MP2/cc-pVTZ and the benchmark CCSD(T)/CBS level of theory.

Sr no.	Complexes[a]	D_e(RI-MP2/cc-pVTZ)	D_e(CCSD(T)/CBS)	D_f[b]	D_0((RI-MP2/cc-pVTZ))	D_0(CCSD(T)/CBS)	EXPT[c]
1	1-Naphthol...CH_3OH	8.02	8.58	0.22	6.49	7.05	7.56[266]
2	1-Naphthol...H_2O	6.79	7.07	0.10	5.16	5.44	5.82[266]
3	1-Naphthol...NH_3	9.23	9.57	0.39	7.18	7.51	7.66[266]
4	Indole...H_2O	5.75	5.70	0.06	4.59	4.55	4.66[267]
5	Indole...CH_3OH	6.78	7.02	0.11	5.75	5.99	5.60[268]
6	Phenol...CH_3OH	7.65	8.21	0.18	6.14	6.71	6.11[269]
7	Phenol...H_2O	6.54	6.82	0.09	4.91	5.18	5.43[269]
8	Anisole...H_2O	4.07	4.89	0.06	2.48	3.30	3.69[268]
9	*p*-Difluorobenzene...H_2O	3.04	3.66	0.05	1.91	2.53	2.79[270]
10	C_2H_4...HCl	2.99	3.01	0.01	1.83	1.85	3.18[271]
11	1-metheylindole...H_2O	5.21	5.63	0.22	3.57	3.98	4.10[267]

Table 3.9: Stabilization energies (D_e) and dissociation energies (D_0) (in kcal/mol) of 11 H-bonded complexes calculated at the RI-MP2/cc-pVTZ and CCSD(T)/CBS levels, [a]*cf.* Figure 3.10. [b]the deformation energy, [c]experimental values.

Evidently, deformation energies are not negligible and should be considered and its main contribution comes from X–H bond lengthening upon formation of X–H…Y H-bond. Deformation energy correlates with the strength of H-bond and is largest for the strongest complex 1-Naphthol….NH_3 (0.39 kcal/mol). The CCSD(T)/CBS stabilization energies are mostly larger than the RI-MP2 ones and the differences are not negligible. Largest difference (–0.82 kcal/mol) was found for Anisole...H_2O

complex while the average difference was smaller than one half (–0.37 kcal/mol). The ΔCCSD(T) correction term has been seen to be converged at the aug-cc-pVDZ basis when the results were compared with the aug-cc-pVTZ basis.

Dispersion-Bound Complexes: The stabilization energies of all 11 dispersion-bound complexes are collected in Table 3.10. Present stabilization energies are much smaller than these of previously discussed H-bonded complexes and the Dibenzo-*p*-dioxin...Kr complex was found to be the strongest one (DFT-D3 level) with D_e of 2.32 kcal/mol. At the CCSD(T) level the strongest complex is the Carbazole…Kr one with D_e of 2.15 kcal/mol. Contrary to previous case the CCSD(T)/CBS stabilization energies are smaller than the DFT-D3 values (in average by 0.08 kcal/mol) and the difference is not exceeding 0.24 kcal/mol or 10%. The difference in D_e between lower- and higher-level calculations is not as large as with H-bonded complexes and the use of DFT-D3 level especially for large dispersion-bounds complexes can thus be here recommended.

Sr No.	complexes[a]	D_e(DFT-D3)	D_e(CCSD(T)/CBS)	D_0(DFT-D3)	D_0(CCSD(T)/CBS)	EXPT[b]
1	Benzene...Ar	1.20	1.13	0.99	0.92	0.97[272]
2	Benzene...Kr	1.59	1.42	1.38	1.21	1.15[272]
3	Dibenzofuran...Ar	1.58	1.57	1.47	1.46	1.49[273]
4	Carbazole...Ne	0.58	0.51	0.49	0.42	0.61[274]
5	Carbazole...Ar	1.68	1.68	1.53	1.54	1.52[274]
6	Carbazole...Kr	2.29	2.15	2.19	2.06	1.97[274]
7	Fluorobenzene...Ar	1.19	1.13	0.99	0.93	0.92[275]
8	Indole...Ar	1.48	1.48	1.32	1.33	1.29[276]
9	Phenol..Ar	1.24	1.19	1.01	0.96	1.04[257]
10	Dibenzo-*p*-dioxin...Ar	1.72	1.65	1.51	1.44	1.51[277]
11	Dibenzo-*p*-dioxin...Kr	2.32	2.08	2.09	1.85	1.83[277]

Table 3.10: Stabilization energies (D_e) and Dissociation energies (D_0) (in kcal/mol) of 11 dispersion-bound complexes calculated at the DFT-D3 and CCSD(T)/CBS levels, [a]*cf.* Figure 3.11, [b]experimental values.

3.4.4.2 D_0 Binding Energy

H-bonded Complexes: Harmonic D_0 energies (covering the deformation energies and ΔZPVE) are shown in Table 3.9. The D_0 is based on RI-MP2 and CCSD(T)/CBS levels whereas the ΔZPVE is calculated only at the RI-MP2 level. The average

calculated ΔZPVE is 1.32 (462 cm^{-1}) which forms a non-negligible part of D_0 energies (about 29% of the CCSD(T)/CBS D_0) and, evidently, its neglect will make agreement with experimental D_0 value much worse (see later).

The Mean Relative Error (MRE) and Mean Absolute Error (MAE) for RI-MP2 D_0 energies when compared to the experimental values were 15.1% and 0.63 kcal/mol. Passing to the CCSD(T)/CBS level both errors were reduced to 9.4% and 0.41 kcal/mol, respectively. The best agreement between calculated and experimental D_0 values was found for Phenol...CH_3OH, surprisingly at the RI-MP2/cc-pVTZ level (MRE, 0.5%). Passing to the more reliable CCSD(T)/CBS level, the MRE error slightly increased. At the CCSD(T)/CBS level the best agreement with experiment was found for 1-Naphthol...NH_3 complex (MRE, 2.0%). On the other hand the worst agreement (MRE of about 42% at both theoretical levels) was found for C_2H_4...HCl complex. Very large error clearly indicates that experimental value is not reliable and should be considered with care. When this complex is omitted the agreement between theoretical and experimental D_0 values considerably improves: MRE at RI-MP2 and CCSD(T) level are 12.3% and 6.2%, respectively. This implies the need to include the ΔZPVE term. Without considering the ΔZPVE, both above mentioned errors dramatically increased to 18.0% and 26.8% at RI-MP2 and CCSD(T) levels, respectively. It is thus demonstrated that for approaching experimental D_0 value, it is inevitable to include the ΔZPVE term.

Dispersion-Bound Complexes: Harmonic D_0 energies (covering ΔZPVE at DFT-D3 level) are shown in Table 3.10. The ΔZPVE energies are now considerably smaller than in previous case and the average ΔZPVE is only 0.17 kcal/mol (60 cm^{-1}). This energy forms only about 13% of the D_0 energy what is less than one half of that contribution for H-bonded complexes (see above). The MRE and MAE amount to 7.7% and 0.09 kcal/mol, respectively at DFT-D3 level and to 6.2% and 0.06 kcal/mol, respectively at the CCSD(T)/CBS level. The best agreement between theoretical and experimental D_0 was found for Dibenzo-*p*-dioxin…Ar complex (MRE, 0.3%) at DFT-D3 level; passing to the more reliable CCSD(T)/CBS level both errors slightly increased. At the CCSD(T)/CBS level the best agreement was found for the Dibenzo-*p*-dioxin...Kr complex (MRE, 1.4%).

For dispersion-bound complexes the ΔZPVE plays less important role. If omitted the calculated errors were found to be 18.8% (MRE) and 0.24 kcal/mol (MAE) at DFT-D3 level, and 14.5% (MRE) and 0.17 (MAE) at the CCSD(T)/CBS level what is considerably less than in the case of H-bonded complexes. Nevertheless, also here for reaching an agreement with experiment an inclusion of ΔZPVE term is inevitable.

3.4.4.3 Anharmonicity

H-bonded Complexes: The perturbation VPT2 calculations are CPU-time intensive; therefore, it is performed only for three strong H-bonded complexes, Phenol...H_2O, *p*-Difluorobenzene...H_2O, Phenol...CH_3OH. Table 3.11 presents D_0 energies for these three H-bonded complexes based on RI-MP2 and CCSD(T)/CBS D_e energies. Comparing the harmonic ΔZPVE with the anharmonic one, we can see that the later one is smaller (similarly as for the HF dimer[278]) than the harmonic one, this difference is practically negligible for the first complex. However, contrary to our expectation the relative error for these three complexes did not changed systematically when passing from harmonic to VPT2 anharmonic ΔZPVE term. Specifically, for the first and second complex, it increased from 4.6 to 5.5% and from 9.3 to 18.8%, respectively, and only for the third one it decreased from 9.8 to 6.8%. The fact that harmonic D_0 is for Phenol...H_2O complex closer to the experimental value is evidently due to quality of the ΔZPVE term, since the D_e energy is already converged.

Dispersion-Bound Complexes: Only four dispersion-bound complexes were investigated at the CCSD(T)//VPT2/DFT-D3 level (see Table 3.11). The calculated D_0 based at both, harmonic and anharmonic ΔZPVE terms, agree reasonably well with experiment. When D_0 was constructed from harmonic ΔZPVE term then the mean relative error for complexes was 5%. Passing to anharmonic ΔZPVE term this error slightly increased to 5.6%. Evidently, both approaches for determining the ΔZPVE term provide similar values of theoretical D_0 which agree reasonably well with experimental values.

H-bond				
Complexes	ΔZPVE[a] (RI-MP2/cc-pVTZ)	D_0(RI-MP2/cc-pVTZ)	D_0(CCSD(T)/CBS)	EXPT
Phenol...H_2O	1.59	4.86	5.14	5.43
Phenol...CH_3OH	1.51	5.96	6.52	6.11
p-Difluorobenzene...H_2O	1.35	1.65	2.27	2.79
HF...HF[278]			2.96	3.04
Dispersion				
Benzene...Ar	0.26	----	0.88	0.97
Benzene...Kr	0.21	----	1.21	1.15
Phenol..Ar	0.14	----	1.05	1.04
Fluorobenzene...Ar	0.15	----	0.98	0.92

Table 3.11: The dissociation energy (D_0) (in kcal/mol) calculated at the anharmonic VPT2 level for 4 H-bonded and 4 dispersion-bound complexes. [a]the anharmonic ΔZPVE is calculated at the RI-MP2/cc-pVTZ//VPT2 level for both H-bonded and dispersion-bound complexes.

3.4.4.4 Diagonal Approximation

In the diagonal approximation, the ΔZPVE term is determined using only 1-D (i.e. non-coupled) anharmonic vibrational frequencies. The evaluation of all 3n-6 vibrational frequencies, where n equals the number of atoms of a complex, could be tedious, but it can be simplified by considering only six inter-molecular degrees of freedom. To obtain accurate D_0, we executed diagonal approximation to all the low vibration modes of a particular complex, Phenol...H_2O (see Table 3.12). Here we considered all six inter-molecular vibration frequencies plus the O–H intra-molecular one corresponding to O–H stretching. The *ρ1*, *β1* and *σ(O..O)* are the three hindered translational modes whereas *τ*, *ρ2* and *β2* are the three hindered rotational modes, respectively (*cf.* Figure 3.12). The ΔZPVE term based on these seven 1-D anharmonic vibrational wavenumbers equals 497 cm^{-1} (1.41 kcal/mol), which surprisingly well agrees with the respective VPT2 value (460 cm^{-1}, 1.31 kcal/mol). The close agreement between both values could be, however, caused by the compensation of errors, and further verification of this very promising simple technique for the estimation of the anharmonic ΔZPVE term is required.

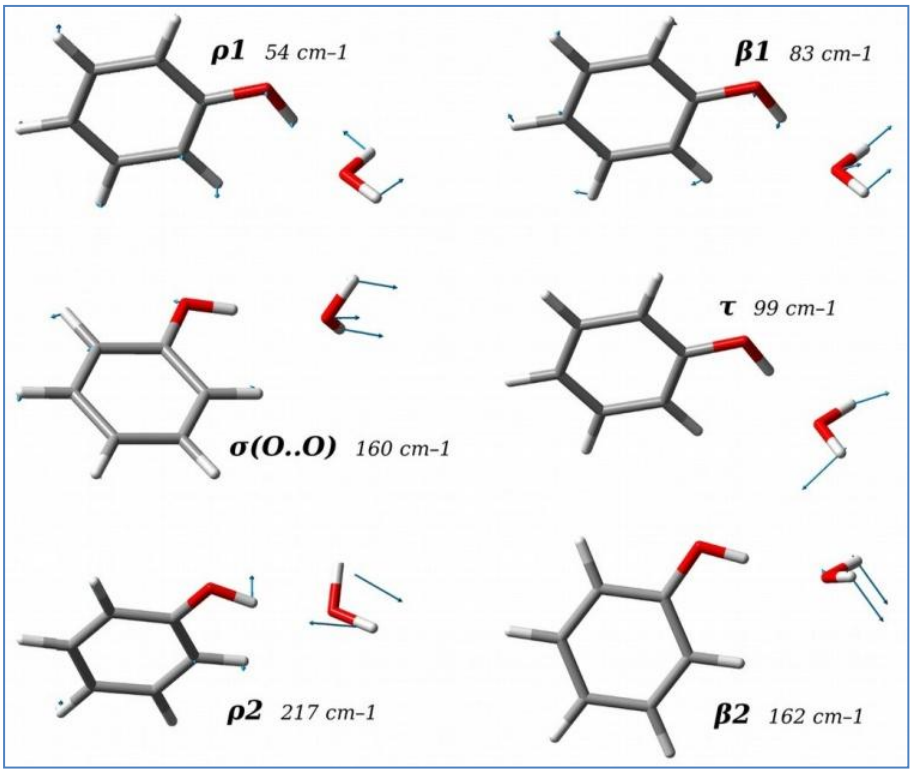

Figure 3.12: Pictorial representation of all the low-vibrational inter-molecular modes of Phenol...H_2O complex.

Phenol...H_2O	Full-dimensional VPT2 calculation	Diagonal approximation
ρ1	54	40
β1	58	83
σ(O...O)	145	160
τ	98	99
ρ2	235	217
β2	162	162
σ(O...H)	3492	3451
free(O...H)	3660	3683

Table 3.12: The anharmonic fundamental wavenumbers (in cm^{-1}) calculated at the standard full-dimensional VPT2 RI-MP2/cc-pVTZ level and applying the 1-D diagonal approximation for the Phenol…H_2O complex.

3.5 Insight into Stability and Folding of GNRA and UNCG Tetraloops Revealed by Microseconds Molecular Dynamics and Well Tempered Metadynamics

3.5.1 Introduction

The RNA hairpins are one of the most common secondary structure motifs that appear almost in every large RNA structure.[279-281] RNA is an unbranched, linear polymer, consists of a primarily Watson-Crick (WC) base paired stem capped with a loop (composed of four nucleotide units, A, C, G, and U, called tetraloop (TL)) of unpaired or non-WC base paired nucleotideds.[282] In this study, we adopted two most important biologically relevant RNA TLs and they are GNRA (N stands for any nucleotide and R for purine bases) and UNCG TL. These TLs enable a sharp bend of RNA backbone and are involved in many biologically important processes, including folding, stabilization of RNA helices and interactions with other biomacromolecules. The highly thermostable UNCG TLs are consider centers where RNA folding can be initiated.[283,284] GNRA TLs are involved in recognition process, which are realized via tertiary RNA-RNA, RNA-protein, and RNA-ligand interactions.[285] X-ray and NMR experiments have shown that GNRA and UNCG TLs have pre*ci*sely defined 3D structures stabilized by a unique combination of rare non-canonical inter-molecular interactions, determining their consensus sequences.[286-290]

This study focused on the folding/un-folding mechanism of the abovementioned TLs because it helps to fold back long RNA motif. Although, many theoretical studies so far has been carried out to dig inside into the mechanism of folding/un-folding such as long classical MD, replica exchange MD (REMD) etc, but, still the understanding is limited.[291,292] Previous MD studies failed to correctly describe the non-canonical but structurally strictly conserved structural features (known as signature interactions) that maintain the GNRA and UNCG TLs folds. Although often not admitted, these simulations resulted in gradual loss of the native structure of both TLs, despite achieving correct base pairing of the stems. The lack of appropriate folding can be masked, for example, by using inadequate folding criteria, such as the RMSD from the target structure.[293] This unsatisfactory simulation behavior was ultimately attributed to limitations of the empirical force fields used in the classical

MD simulations.[282] Later, the reparameterization of the χ torsion term of the Cornell et al. AMBER force field[294] gives almost correct description of the RNA motifs to fold back which has been shown by many studies so far.[105,295,296] These successes may indicate that the new force fields are able to closely represent the native conformation of the TLs, and therefore, could potentially be used to explore folding/un-folding processes and the accompanying free energy changes. We used well tempered metadynamics (WT-MetaD)[171] which is a well-established simulation technique designed to overcome free energy barriers and provide a more robust exploration of the free energy surfaces (FESs). This technique is able to search for different conformations of RNA on a reasonable timescale, which alleviates the enormous computational cost required for reconstruction of FESs of RNA molecules from unbiased MD simulations.

The aim of this work was to study the stability and free energy changes accompanying folding/un-folding of two short RNA hairpin loops (GAGA and UUCG TLs). Another aim was to assess limits of CV-based approaches to study folding of small nucleic acids.

3.5.2 System Preparation

This study concerns two small 10-mer RNA sequences. They are 5'-CGCGAGAGCG-3' (henceforth referred to as the GAGA hairpin) and 5'-CGCUUCGGCG-3' (henceforth referred to as the UUCG hairpin) containing the GAGA and UUCG TL motifs, respectively. They are capping an A-RNA stem comprising three WC base pairs. The structure of the GAGA TL, including one GC base pair adjacent to the TL, was taken from the 1.04 Å resolution X-ray structure of the sarcin-ricin loop from *E. coli* 23S rRNA (PDB ID 1Q9A, residues 2658-2663).[297] It was extended by two additional GC base pairs. The structure of the UUCG TL, including one GC base pair adjacent to the TL, was taken from the high resolution NMR structure,[298] and again, two GC base pairs were added to complete the hairpin loop (*cf.* Figure 3.13).

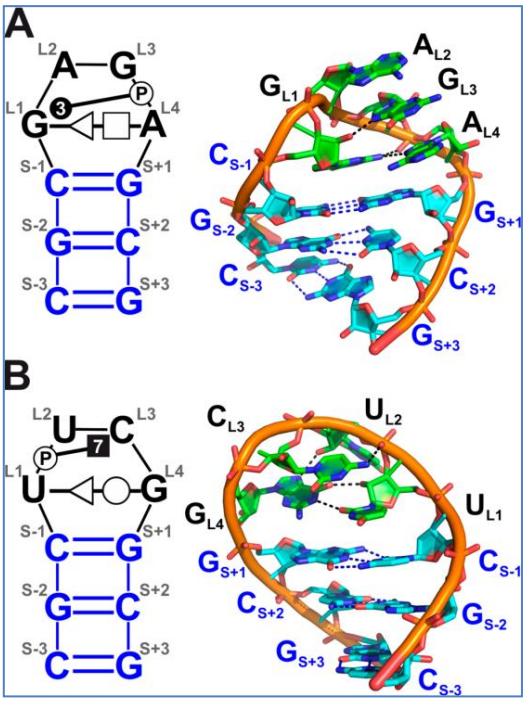

Figure 3.13: Secondary (left) and 3D (right) structures of the studied (A) GNRA and (B) UNCG hairpins, consisting of a short A-RNA stem (blue labels) and the tetraloop (TLs, black labels). The non-canonical base-pairing and base-phosphate interactions in the TL region are classified according to the standard Leontis-Westhof-Zirbel nomenclature; the signature interactions of the TLs are shown in the structures by black dashed lines.

3.5.3 Calculation Strategies

3.5.3.1 Classical MD Setup

To start the dynamics, both TLs were immersed in a rectangular box, then solvated by TIP3P explicit water[210] models and neutralized by adding 9 Na^+ counter ions (r = 1.87 Å and ε = 0.0028 kcal/mol).[253] The simulation were carried out using PBC according to the PME method[215] with a 10 Å non-bonded cutoff. The whole system was first minimized to avoid any close contacts followed by a slow heating of the system starting from 10K to room temperature (300K) keeping the box volume constant. Initial velocities were generated from the Boltzmann distribution. Later the equilibration step is started maintaining NPT conditions through the use of a V-rescale thermostat[299] and isotropic Parrinello-Rahman barostat.[214] A 2.0 fs integration step and temperature of 300K were used and run through the cutting edge timescale of 15 μs for each TL. The simulation were executed using the Gromacs 4.5.1 software package[185] with the all atomic AMBER ff99bsc0χ_{OL3} ff.[105]

3.5.3.2 WT-MetaD Setup

The same system and simulation setup as described for the unbiased MD simulations were used for the WT-MetaD simulations. Prior to initiating the WT-MetaD, 5 ns equilibration MD runs (NPT) were performed with the Nose-Hoover thermostat[212,213] and isotropic Parrinello-Rahman barostat.[214] The equilibration MD runs were followed by a 200-ns-long biased WT-MetaD simulation, performed with 1.0 fs time step at a temperature of 300K. The PLUMED 1.3 plugin[239] was used to carry out WT-MetaD simulations with the Gromacs 4.5.5 code.[185] The deposition rate, ω, and deposition stride, τ_G, of the Gaussian hills was set to 0.478 kcal/mol.ps (2 kJ/mol.ps) and 1 ps, respectively. The bias factor (T+ΔT)/T was set to 15.

Collective variables: Several CVs have been tested in this study on the basis of their performance to enable to construct the FES associated with all the possible conformational changes undergoing in the process of un-folding of small RNA hairpin.

1) H_{core}, which calculate the extent of native H-bonding in three base pairs of the stem, the L1-L4 non-canonical base pairs, and the other signature interactions of the TL. The H_{core} were calculated using the switching functions as follows:

$$H_{core} = \sum_i \frac{1-(r_i/r_0)^n}{1-(r_i/r_0)^m} \qquad 3.8$$

where r_0 was set either to 2.0 Å or to 2.5 Å, henceforth labeled as $H_{core,2.0}$ and $H_{core,2.5}$, respectively. The *n* and *m* parameters were set to 6 and 12, respectively. In case of $H_{core,2.0}$ the index *i* corresponded in both TLs to the eleven H-bonds involved in base pairing of the three GC base pairs of the stem region and additional two H-bonds of the L1-L4 non-canonical base pairs of the TL. The $H_{core,2.5}$ including addition one or two signature H-bonds outside the L1-L4 base pairs. The r_i was the distance between the H-acceptor and the hydrogen atom bound to the H-donor of the abovementioned H-bonds.

2) R_{core} (radius of gyration), which accounted for the global shape of the RNA strand. This CV aimed to describe global bending from straight ss-RNA to the bent hairpin fold. The R_{core} CVs were calculated as follows:

$$R_{core} = \frac{\sum_i^n (r_i - r_{com})^2 m_i}{\sum_i^n m_i} \quad 3.9$$

where the sum extended over either six nitrogens of the stem nucleobases, one from each nucleobases, henceforth $R_{core,base}$, or over all nine phosphorus atoms, henceforth $R_{core,P}$ and the center of mass r_{com} was defined by

$$r_{com} = \frac{\sum_i^n r_i m_i}{\sum_i^n m_i} \quad 3.10$$

3) RMSD of the first, third and fourth nucleotide of the loop (those nucleotides are involved in the signature interactions within the loop) from the corresponding native structure. This CV should sample un-folding behavior of the loop region. The RMSD CV was defined as mass-weighted root-mean-squire-deviation of L1, L3, and L4 necleotides (i.e., G_{L1}, G_{L3} and A_{L4}, and U_{L1}, C_{L3} and G_{L4} nucleotides in case of GAGA and UUCG TLs, respectively).

4) In the case of UUCG TL, also the glycosidic torsion of the G_{L4} nucleotide (G_{L4}-χ) was used as CV. This nucleotide is in *syn*-orientation in the native structure. The G_{L4}-χ CV was defined as the G_{L4}(O4')-G_{L4}(C1')-G_{L4}(N9)-G_{L4}(C4) dihedral angle.

The Gaussian widths were set to 0.2 arbitrary unit, 2 Å, 0.5 Å, and 0.2 radians for H_{core}, R_{core}, RMSD, and G_{L4}-χ, respectively.

3.5.4 Results and Discussion

3.5.4.1 Overview of WT-MetaD Simulation

The choice of useful CV is very much crucial for the applicability of WT-MetaD simulation. The basic technical convergence of the FESs was monitored by calculating differences between two consecutive FESs constructed at times *i*-10 and *i* ns, according to Branduardi et al.[300] This approach uses the former FES as a reference and examines regions in the CV space defined by points lying up to 5 kcal/mol above the global minimum of the reference FES. It should be keep in mind that a progressive convergence between consecutive FESs does not always guaranteed that the WT-MetaD simulation is fully converged. Sometimes, a false convergence can appeared because of sampling to some other local minima due to

hidden CVs (orthogonal to the selected CVs) that does not allow the system to sample on the proper region asked.

So far, in this study, we have carried out a total of eleven independent biased WT-MetaD simulations including few classical MD simulations. The biased simulations were performed with the CVs proposed above for the un-folding of the both GAGA and UUCG TLs.

3.5.4.2 WT-MetaD of GAGA TL

For the GAGA TL, a 200ns long WT-MetaD simulation using two CVs, namely $H_{core,2.0}$ and $R_{core,base}$ have been used. Similar choice of CVs was successfully used, e.g., for folding of small beta-hairpin peptide,[301] G-DNA triplex,[302] and G-DNA quardraplex.[303] In the initial phase of the simulation, the rapid flooding of the native state region ($R_{core,base}$ ~ 4-6, $H_{core,2.0}$ ~ 4-7; the initial X-ray structure corresponds to $R_{core,base}$ ~ 3.7 and $H_{core,2.0}$ ~ 6.5) was observed. However, at later times, the system never returned back to the native state region and sampling only the un-folded and misfolded states with $R_{core,base}$ and $H_{core,2.0}$ in the range 3-14 and 0-3, respectively (*cf.* Figure 3.14). Therefore, the results can be taken as a complement that the selected CVs were not able to accelerate the re-folding events, or the possibility that GAGA TL is inherently not stable under the force field description. Another possibility is that the native state may kinetically stable on microsecond time scale in classical MD simulation. Later, a snapshot corresponding to the misfolded state (R_{core} = 9.0 Å and H_{core} = 1.5 a.u) was taken and run a long classical MD simulation. It was soon clearly realized that GAGA TL however did not maintain the native CV values in space. The system transformed to a straight ss-RNA conformation with random formation of few native H-bonds which were, however, not able to form any inherently stable structures with a longer life time. In that case, it can be concluded that the chosen CVs were not robust enough to accelerate the re-folding events.

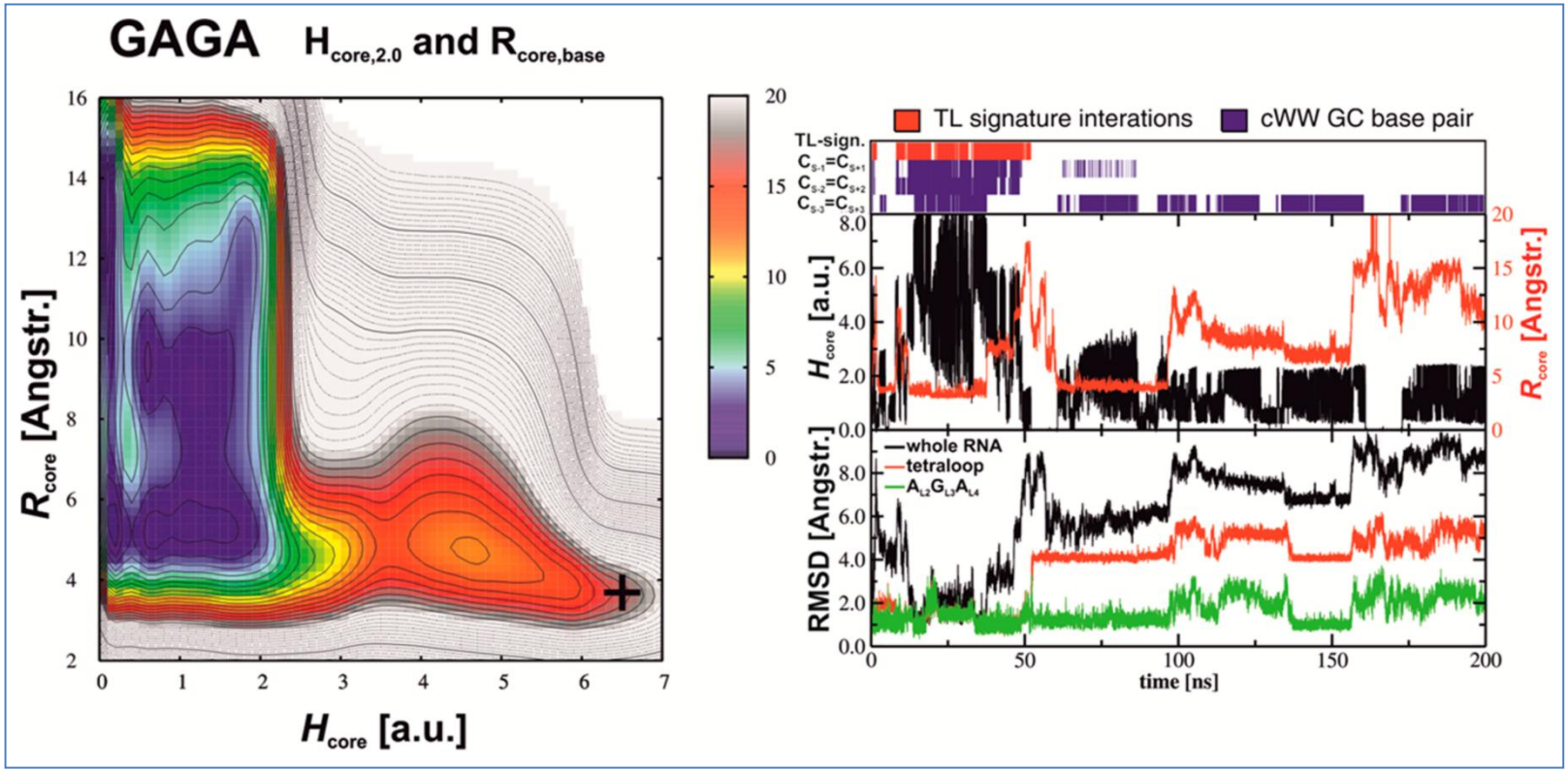

Figure 3.14: FES plot (left) and the development of key structural features (right) obtained from 200-ns-long WT-MetaD simulation of GAGA TL using $H_{core,2.0}$ and $R_{core,base}$ CVs. The FES is contoured by 2 kcal/mol (thin lines) and 10 kcal/mol (thick lines). Energies lower than 20 kcal/mol relevant for structural dynamics are colored. An X-ray structure corresponding to $H_{core,2.0}$/$R_{core,base}$ of 6.5/3.7 Å, respectively, is denoted by black cross. In the top of the right panel, the red and blue stripes indicate the presence of TL signature interactions and GC base pairs of the stem, respectively, both defined on the basis of hydrogen bonding with 4.0 Å cut-off for heavy-atom distance. The middle left panel shows the time evolution of the $H_{core,2.0}$ (black) and $R_{core,base}$ (red) CVs. The lower left panel shows evolution of the root-mean-square deviation (RMSD) of the whole RNA hairpin (black), the TL (red), and the tripurine $A_{L2}G_{L3}A_{L4}$ stack (green)

Further, another robust sampling technique called well tempered ensemble (WTE) parallel tempering MetaD (WTE-PT-MetaD) is used in order to rule out the possibility that the WT-MetaD was not efficient enough to accelerate the sampling of the re-folding events (result is not shown). The advanced WTE-PT-MetaD result predicts that the choice of CV is not well enough to carry out the re-folding events of GAGA TL. However, another testing of the WT-MetaD method was carried out where the new definition of the CVs look like, $H_{core,2.5}$ and $R_{core,P}$. This simulation (400 ns long) was again failed to show the re-folding events of the GAGA TL.

Finally, a few WT-MetaD simulations were again carried out with complete redefining of the CVs which includes the description of the native loop formation using RMSD of the loop region. The following WT-MetaD simulation was extended to same time scale (as it is before) using RMSD and $H_{core,2.5}$ or RMSD, $H_{core,2.5}$ and $R_{core,P}$ CVs, i.e., with two or three CVs, respectively. Similarly to the previous WT-MetaD ($H_{core,2.0}$; $R_{core,base}$) simulation, the rapid loss of the native structure during the initial phase of the simulation was seen for both two and three CVs simulation. The simulation with three CVs never returned to its native state and sampled the transitions between fully un-folded ss-RNA and misfolded states. However, the re-folding events were finally sampled in the simulation with two CVs (RMSD and $H_{core,2.5}$) (*cf.* Figure 3.15). Here the system was seen to be rapidly flooding the native fold basin (state V) during the initial phase and sampled un-folded (state I) and misfolded states with formed stem region but unstructured loop (states VI and VII). However, after that the system formed the $A_{L2}|G_{L3}|A_{L4}$ tripurine stack with the ss-RNA conformation (stat II), finished formation of the loop by establishing *t*HS $A_{L4}G_{L1}$ base pair (state III) followed by zipping the stem region (state IV and V). After this re-folding event, the system lost the native conformation of the loop and sampled the misfolded states (states VI and VII). Therefore, the overall FES revealed several minima corresponding to above mentioned folded (state IV and V), un-folded (state I), and misfolded states. It seems that the folded and misfolded states are similar in free energy, while un-folded state appeared to be the most stable. The simulation was extended to 1000 ns long, after ~700 ns several above-mentioned re-folding events were seen.

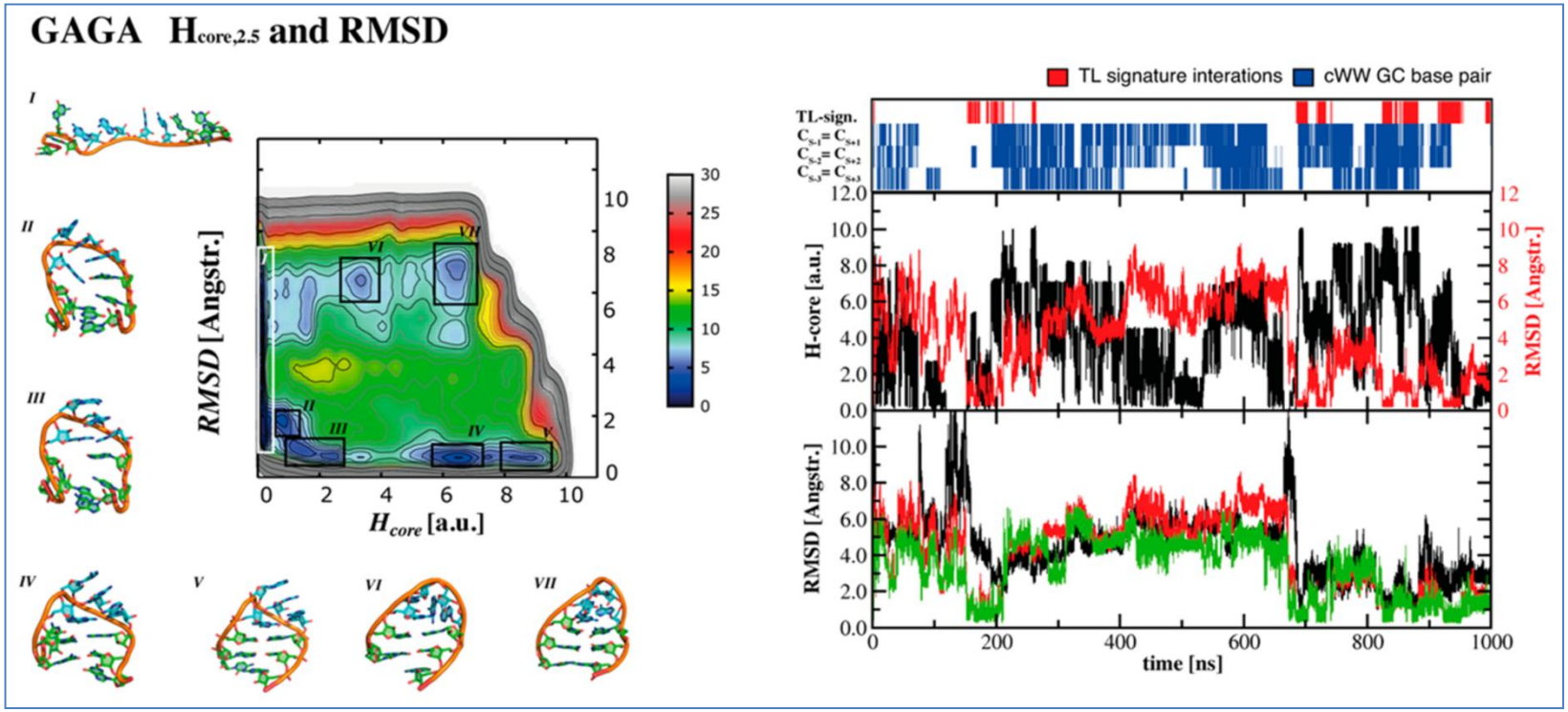

Figure 3.15: FES plot (left) and the development of key structural features (right) obtained from 1000-ns-long WT-MetaD simulations of GAGA WT-MetaD simulations using $H_{core,2.5}$ and RMSD of the loop region from the native conformation as CVs. The colors and contours correspond of the FES as well as the definition of all structural parameters of the right panel corresponds to those of Figure 3.14. The regions depicted on the FES by black boxes correspond to the conformational states populated during the simulation; the corresponding structures are shown in sticks and cartoons.

Thus in this case also, the simulation cannot be considered fully converged and most likely much longer simulation would require achieving a quantitative convergence.

3.5.4.3 WT-MetaD of UUCG TL

Similar to GAGA TL, a 200 ns long WT-MetaD simulation (*cf.* Figure 3.16) was performed for UUCG TL with the combination of $H_{core,2.0}$ and $R_{core,base}$ CVs. The rapid expected native state ($R_{core,base}$ ~ 4–6, $H_{core,2.0}$ ~ 4–7 region was seen to be flooded; the initial NMR structure correspond to $R_{core,base}$ ~ 4.2 and H_{core}, 2.0 ~ 4.6). However, in this WT-MetaD simulation, the TL reversibly fluctuated between un-folded and near-native states (denoted as frayed or flipped-over states, sharing all features with the native folded state except that the terminal base pair was possibly broken).[304] Although, the G_{L4} nucleotide was seen to be retained its *syn*-conformation during the whole 200-ns long WT-MetaD simulation. The *syn*-conformation is an important part of the UNCG signature, so that the molecules remained poised to fold. This means that the used WT-MetaD protocol did not sample conformational states along a genuine un-folding pathway with the *anti*-orientation of G_{L4} nucleobase.

In order to understand the G_{L4} *syn-anti* dynamics, another independent WT-MetaD simulation was performed for 200-ns long with the same CVs (*cf.* Figure 3.17). In this case, the G_{L4} nucleobase was seen to be spontaneously flipped from *syn* to *anti* state during the initial part of the simulation. Therefore, it should be mentioned that the difference between this two independent WT-MetaD simulations was solely due to the coincidental occurrences of a rare event not accelerated by recent CVs used. This simulation only sampled the G_{L4} *anti*-orientation, never came back to its *syn*-state and helped the system to repeatedly visit the completely un-folded single-stranded state as well as bent states with the formed A-RNA states. Those bent states were considered as the near native states because of the highly distorted TL region. Therefore, once the G_{L4} nucleobase had flipped from *syn* to *anti*, our WT-MetaD simulation sampled the un-folded to misfolded transition but not the un-folding/re-folding same as the previous UUCG WT-MetaD simulation.

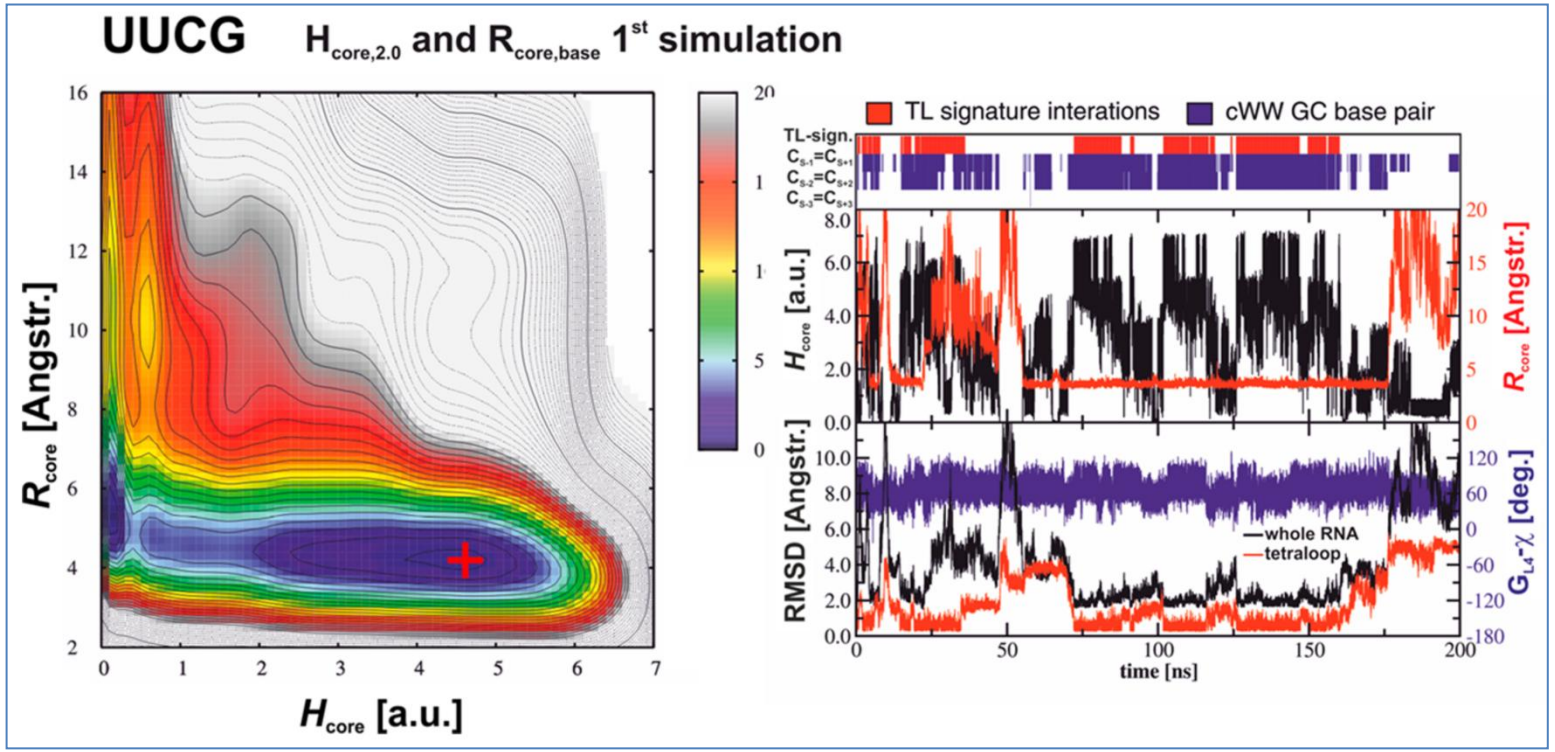

Figure 3.16: The FES plot (left) and the development of key structural features (right) obtained from the first (G_{L4}-*syn* populating) 200-ns-long WT-MetaD simulation of UUCG TL using $H_{core,2.0}$ and $R_{core,base}$ CVs. The contouring and coloring scheme of the FES is same as in Figure 3.14. The MNR structure corresponding to $H_{core,2.0}$/$R_{core,base}$ of 4.6/4.2 Å, respectively, is denoted by red cross. The time development of the TL signature interactions and GC base pairs of the stem (top right) as well as the CVs (middle right) are presented as in Figure 3.14. The lower panel shows evolution of the root-mean-square deviation (RMSD) of the whole RNA hairpin (black), the TL (red), and glycosidic χ torsion of G_{L4} (blue).

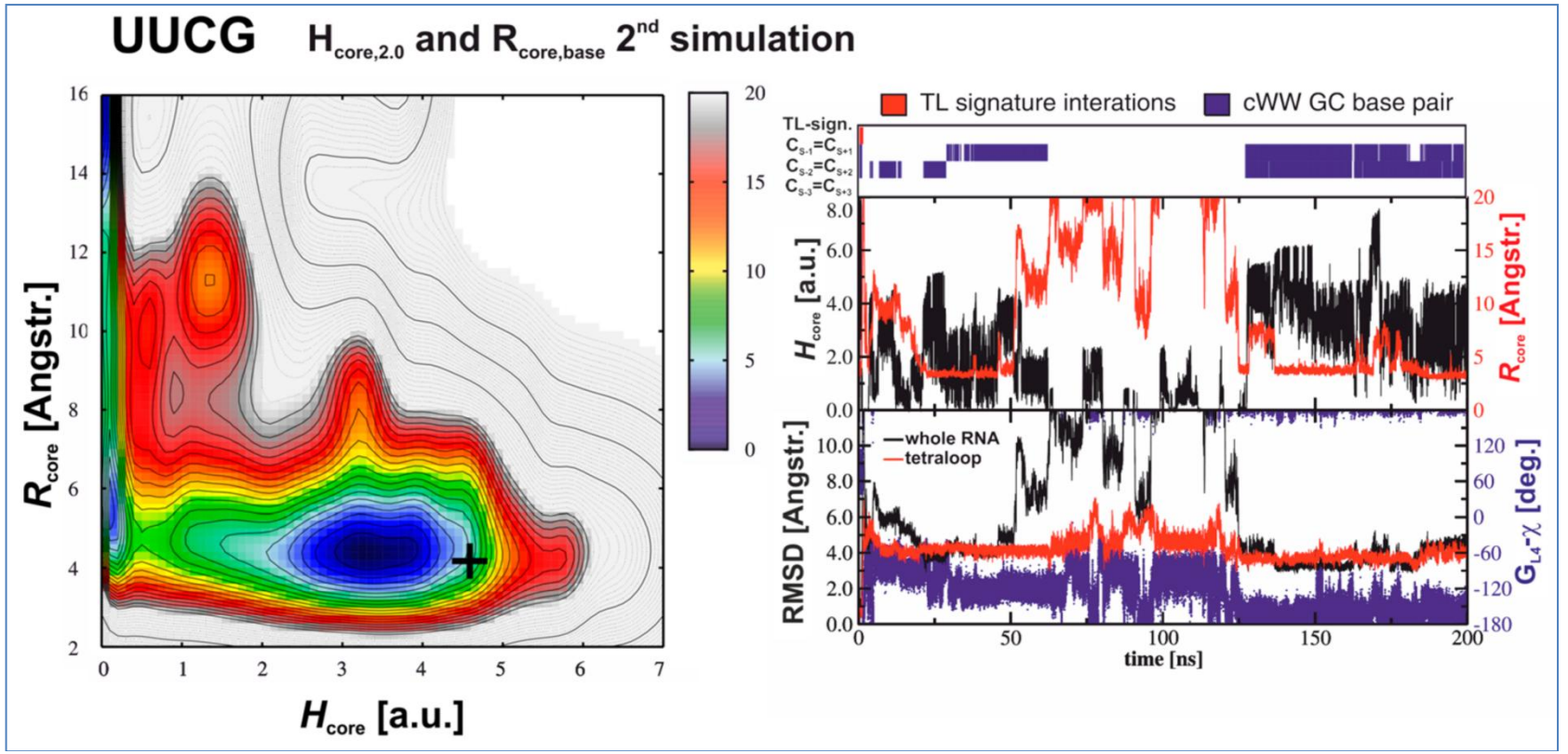

Figure 3.17: The FES plot (left) and the development of key structural features (right) obtained from the second (G_{L4}-*anti* populating) 200-ns-long WT-MetaD simulation of UUCG TL using $H_{core,2.0}$ and $R_{core,base}$ CVs.

Now it is remained a question of how to treat both the G_{L4} *syn-anti* conformation in a single WT-MetaD simulation? Therefore, torsion angle (G_{L4}-χ) was used as a third CV with the other two CV ($H_{core,2.5}$ and $R_{core,base}$) and extended to 400-ns long. However, the results show that WT-MetaD simulation were even not able to sample the re-folding events although the system sampled both G_{L4} *syn-anti* orientation. This might be simply caused by the increased complexity of the CV space. i.e., due to its increased dimensionality. Overall, from all three simulations, it can be concluded that the observed repetitive un-folding/re-folding events in the first G_{L4}-*syn* populating WT-MetaD simulations were allowed because the sugar-phosphate backbone of the loop region retain its structural memory. Finally, the observed re-folding events were facilitated by relatively simple conformational space of the loop within the framework of the G_{L4}-*syn* orientation. In other words, once the sampling over all orientation of G_{L4} was achieved, the conformational space became significantly more complex and the method was no more capable of sample un-folding/re-folding events. So some more sophisticated CVs would require guiding un-folding/re-folding events.

We tested another CV as RMSD of the loop region following the scheme used for GAGA TL. So a 400-ns long WT-MetaD simulation of UUCG TL using RMSD and $H_{core,2.5}$ or RMSD, $H_{core,2.5}$ and G_{L4}-χ CVs, i.e., with two or three CVs respectively. Again in both cases, the native state sampling is failed. Therefore, in particular, the G_{L4}-*anti* state may act as a kinetic trap along the folding pathway of the UUCG TL, which is consistent with experimental kinetic measurements of G_{L4} 8-bronoguanisine[110,305] as well as with the recent REMD simulation.

3.5.4.4 Mechanism of Un-folding

The WT-MetaD simulations were able to capture some attributes of the structural changes along the un-folding and eventual re-folding events.

Un-folding of GAGA TL: The WT-MetaD simulation showed the native un-folding process which was already shown by Sorin. et al.[306] The un-folding process of GAGA TL comprised several steps. It includes opening of the CG pairs via base pair fraying initiated by 5‘-terminal nucleotide, A-RNA stem opening, i.e., rupture of the remaining two pairs of the stem, and, finally breakage of the TL signature

interactions. At the beginning, the system rapidly lost the base pairing and the signature interactions in the TL region, however, the TL retain its bent conformation ($A_{L2}|G_{L3}|A_{L4}$ tripurine stack). After ~15-ns, both base pairing and signature interactions of the TL were reestablished and formed the native conformation (*cf.* Figure 3.14). Note that, this was the only re-folding events in the WT-MetaD simulation which is most likely facilitated by the near native conformation of the loop backbone and finally, the fraying of the 5‘-terminal of the cytosine base occurred. This type of fraying is not surprising because the 5‘-nucleobase are known to exhibit weaker stacking with the adjacent base pairs within the stem than 3‘-end nucleobase. Immediately after unstacking of the 3‘-end from the stem, the second and third pairs ($G_{S-2}=C_{S+2}$, $C_{S-1}=G_{S+1}$) were destablished which resulted in loss of the helical stem. However, the TL region remained temporarily folded for next few ns which was maintaining the *trans*-Hoogsteen sugar edge (*t*HS) A_{L4}-G_{L1} interaction. After destabilization of the *t*HS A_{L4}-G_{L1} pair interaction, the system was seen to be completely un-folded. It is observed that such un-folding and re-folding mechanisms resembled the folding mechanism suggested by the recent REMD simulation.[122]

Un-folding of UUCG TL: In case of UUCG TL, two independent WT-MetaD simulations were employed where most importantly, the G_{L4} nucleaobase made the difference of different folding/un-folding mechanism. Therefore, the un-folding mechanism should be discussed on the basis of G_{L4} *syn-anti* orientation. In contrast to GAGA TL, in the 1st simulation of UUCG TL, a reversible un-folding/re-folding event was observed. Like GAGA, at the beginning, the system rapidly lost the stem interactions but after ~2-ns the interactions were reestablished except the terminal base pair (*cf.* Figure 3.16). Note that, the terminal base pair did not reestablish its canonical *c*WW base pairing through the entire simulation. However, after ~40-ns the 3‘-end guanine flipped over and changed its glycosidic torsion from *anti to syn* and after ~55-ns stacked back to the stem with the opposite stacking face and established a pair with the 5‘-end cytosine in *t*HS GC base pairing family. The hairpin was repeatedly fully un-folded during the simulation; however, the glycosidic torsion of the G_{L4} guanine remained in its *syn*-orientation throughout the whole simulation. In contrast, the 2nd equivalent UUCG WT-MetaD simulation, the G_{L4} glycosidic torsion retains its *anti*-conformation which resists re-folding events (see discussion above). Finally it can be concluded that the *anti*-orientation of G_{L4} makes

a kinetic trap for the folding of UNCG TLs.[110,122,305]

3.5.4.5 Classical MD simulation

Despite enormous sampling of conformational dynamics of the RNA systems, classical unbiased MD simulation is also important for description of the structural flexibility of these systems. Therefore, the structural dynamics of both hairpins using unbiased MD simulations on microseconds timescales (15-μs long) has been employed at 300K with all-atomic AMBER ff99bsc0χ_{OL3} force field. Both TLs fluctuated around their native states for more than 8 μs, suggesting that the native state corresponding to a relatively stable minimum in the utilized force field. During this part of the simulations, the behavior of both TLs was rather uniform, the loop and the central section of the stem were stable and the systems fluctuated around the starting structures. The terminal base pair was the most flexible part of both the hairpins. Overall, the native structure was seen to be more stable in the 15 μs long classical MD simulation.

CHAPTER 4

4 Conclusions

4.1 Graphene...Organic Electron Acceptor Charge Transfer Complexes

The interaction energies and the association free energies of graphene-TCNE and -TCNQ electron acceptor charge transfer complexes have been investigated with various WFT, DFT and MM (AMBER) force field methods. The following results are obtained.

1) The MP2.5/CBS/6-31G*(0.25) method produced accurate results for C2 complexes. The SCS-MP2 method agreed better than SCS(MI)-MP2 with the benchmark interaction energies. All the non-MP2 methods such as DFT-D3, DFTB-D and MM, provided lower stabilization as compared to benchmark values.

2) For larger complexes (beyond C2), only DFT-D3, DFTB-D and MM methods were employed and the DFT-D3 performed better. The simpler DFTB-D and MM provided monotonically increasing stabilization with graphene model size.

3) The association free energies were evaluated at DFTB-D, MM and also using PMF/MM methods. All the methods showed nice agreement of lowering the stabilization of the complex passing from gas to solvent phase. The gas phase free energies agreed well to each other whereas, in the solvent, the PMF/MM value was seen to be ~5 kcal/mol lower than the DFTB-D and MM method. This is probably due the high loss in translational degrees of freedom calculated with DFTB-D and MM method using RR-HO-IG approximation on a single geometry.

4.2 Adsorption of DNA Base Pairs on Silica Surface

The association free energies of the DNA base pairs have been investigated using DFT-D (TPSS-D/TZVP), MM (AMBER) and also with advanced biased MetaD simulation. The following results are obtained.

1) The empirical potential produces reliable results in comparison with the

much more expensive QM results, which means that MM can be successfully applied for the description of base-base interactions on the silica surfaces. All QM and MM calculations as well as biased MetaD simulations demonstrate that the mG…mC and mA…mT base pairs are adsorbed as H-bonded form on the fully solvated silica surface.

2) Biased MetaD simulation and RR-HO-IG approximation convincingly demonstrate that the adsorption of all the H-bonded structures of both base pairs is stronger than that of any π…π stacked structures.

3) The association free energies obtained on the silica surface was compared with the results obtained on the graphene surface. It was realized that on both the surfaces, the H-bonded structures for both the base pairs are more stable than the stacked structures. Further, the WC, non-WC, and Hoogsteen H-bonded structures are adsorbed at solvated graphene surface more strongly than the solvated silica surface. The dispersion energies are more pronounced at the graphene surface than the silica surface.

4.3 Aldose Reductase Inhibition

The human aldose reductase (AR) inhibition has been investigated by tuning an X-bond-to-H-bond in the AR inhibitors. The geometries and binding energies obtained from QM/SQM/MM scoring are thoroughly compared with the geometry obtained from X-ray crystallography and the binding energy with IC50 measurements. The comparison brings the following results.

1) The QM/SQM/MM optimization clearly predicted that the binding mode has been conserved when the X-bond is substituted by H-bond. The three layer QM/QM/MM approach produced reliable geometry relative to X-ray crystallography. However, the distance between N atom of AK198 and the oxygen atom of Thr113 is found to be 3.24 Å, whereas, it is 2.92 Å in the X-ray crystallography. This is expected to be due to insufficient basis set used in the calculations.

2) In the gas phase, the H-bond was found to be significantly stronger than the X-bond and the total binding affinity decreased due to the larger desolvation

penalty of the NH_2 group which is proved by both SMD (single molecule approach) and the MD free energy simulation (average statistical approach).

3) Fragment based approach in the alchemical free energy simulation produced relevant results in comparison to SMD data.

4.4 Benchmark Dissociation Energies for Eleven H-bonded and Eleven Dispersion-Bound Complexes

The benchmark dissociation energies of 11 H-bonded and 11 dispersion-bound complexes have been evaluated with the several WFT (RI-MP2, CCSD(T)/CBS) and DFT-D3 methods. The "theory to theory" and "theory to experiment" comparison brings the following results.

1) The agreement between theoretical D_0, based on harmonic ΔZPVE term, and experimental D_0 for H-bonded complexes is reasonable (the relative error 12.3% and 6.2% at RI-MP2 and CCSD(T) levels, respectively) and, following expectation it is considerably better at the CCSD(T) level. Without considering the harmonic ΔZPVE, the error dramatically increased to 18.0% and 26.8% at RI-MP2 and CCSD(T) levels, respectively. It clearly demonstrates the importance of considering the ΔZPVE term. When passing to anharmonic ΔZPVE term, the agreement did not improved what indicates that the applicability of the perturbation VPT2 technique for H-bonded complexes is limited.

2) Theoretical D_0, based on harmonic ΔZPVE term, agree with experimental D_0 for 11 dispersion-bound complexes comparably well and the relative error at DFT-D3 and CCSD(T) levels amounts to 7.7% and 6.2%, respectively. Also here the inclusion of ΔZPVE term is important (though less than for H-bonded complexes) and omitting the term the error at DFT-D3 and CCSD(T) levels increased to 18.8% and 14.5%, respectively. When instead of harmonic, the anharmonic ΔZPVE term was included, the agreement between theoretical and experimental D_0 slightly deteriorated.

3) Applicability of diagonal approximation for estimation of ΔZPVE term was investigated for Phenol…H_2O complex for which the perturbational VPT2 technique converged. The agreement between ΔZPVE term determined using the

diagonal approximation and VPT2 method has been excellent (within 0.1 kcal/mol).

4.5 Stability and Folding of GNRA and UNCG Tetraloops using Microsecond MD and WT-MetaD Simulation

The un-folding study of two biologically most important RNA (GAGA and UUCG) hairpins have been investigated using biased WT-MetaD and long (15-μs) standard MD simulation. The simulations bring the following conclusions.

1) Despite the insightful and careful choice of CVs, the WT-MetaD simulations achieved only the semiquantitative convergence. Specifically, neither reversible folding/un-folding events with sufficient number of interconversions during the simulations nor full sampling of the un-folded state has been observed.

2) In case of GAGA TL, the WT-MetaD simulations predicted the same mechanism of un-folding as it was predicted by previous REMD simulation. When compared the results obtained with unbiased 15-μs-long MD simulation, the same mechanism of folding/un-folding events was seen. Therefore, both simulations and the force field used in the simulation do properly represent the structural features of the GNRA (GAGA) TLs.

3) In case of UUCG TLs, the crucial role of the orientation of G_{L4} nucleobase around its glycosidic bond in both unbiased MD and WT-MetaD simulations has been confirmed. The simulation data strongly supported the hypothesis that the G_{L4}-*anti* conformation represents a kinetic trap along the UUCG TL folding pathway. The WT-MetaD simulation also predicted that once the G_{L4} is in the *syn* orientation, the force field described the native state frequently, whereas, in the *anti* conformation, the re-folding event is restricted which is probably due to the sugar-phosphate backbone loop region.

Overall, both the WT-MetaD and extensive unbiased MD simulation provided new insights into the conformational behavior of the GAGA and UUCG TLs as well as information about the performance of the used ff99bsc0χ_{OL3} force field.

Bibliography

1. Haldar S.; Kolar M.; Sedlak R.; Hobza P. *J. Phys. Chem. C* **2012**, 116, 25328.
2. Haldar S.; Spiwok V.; Hobza P. *J. Phys. Chem. C* **2013**, 117, 11066.
3. Fanfrlik J.; Ruiz F. X.; Kadlcikova A.; Rezac J.; Cousido-Siah A.; Mitschler A.; Haldar S.; Lepsik M.; Kolar M. H.; Majer P.; Podjarny A. D.; Hobza P. *ACS Chem. Biol.* **2015**, 10, 1637.
4. Haldar S.; Kuhrova P.; Banas P.; Spiwok V.; Sponer J.; Hobza P.; Otyepka M. *J. Chem. Theory Comput.* **2015**, 11, 3866.
5. Wilcken R.; Zimmermann M. O.; Lange A.; Joerger A. C.; Boeckler F. M. *J. Med. Chem.* **2013**, *56*, 1363.
6. Bissantz C.; Kuhn B.; Stahl M. *J. Med. Chem*. **2010**, 53, 5061.
7. Gonnade R. G.; Shashidhar M. S.; Bhadbhade M. M. *J. Indian I. Sci. A-Eng.* **2007**, 87, 149.
8. Burley S. K.; Petsko G. A. *Science* **1985**, 229, 23.
9. Hunter C. A.; Singh J.; Thornton J. M. *J. Mol. Biol*. **1991**, 218, 837.
10. Riley K. E.; Hobza P. *Acc. Chem. Res*. **2013**, 46, 927.
11. Hobza P.; Havlas Z. *Theor. Chem. Acc*. **2002**, 108, 325.
12. Hobza P.; Havlas Z. *Chem. Rev*. **2000**, 100, 4253.
13. Trudeau G. T.; Dumas J. M.; Dupuis P.; Guerin M.; Sandorfy C. *Top. Curr. Chem.* **1980**, 93, 91.
14. Watson J. D.; Crick F. H. *Nature* **1953**, 171, 737.
15. Jurecka P.; Sponer J.; Cerny J.; Hobza P. *Phys. Chem. Chem. Phys*. **2006**, 8, 1985.
16. Hobza P.; Sponer J. *J. Am. Chem. Soc*. **2002**, 124, 11802.
17. Burley S. K.; Petsko G. A. *Adv. Protein Chem*. **1988**, 39, 125.
18. Bhattacharyya R.; Samanta U.; Chakrabarti P. *Protein Eng*. **2002**, 15, 91.
19. Kannan N.; Vishveshwara S. *Protein Eng*. **2000**, 13, 753.

20. McGaughey G. B.; Gagne M.; Rappe A. K. *J. Biol. Chem*. **1998**, 273, 15458.

21. Grimme S.; Antony J.; Schwabe T.; Muck-Lichtenfeld C. *Org. Biomol. Chem*. **2007**, 5, 741.

22. Hunter A. C. *Angew. Chem. Int. Ed*. **1993**, 32, 1584.

23. Gung W. B.; Amicangelo C. J. *J. Org. Chem*. **2006**, 71, 9261.

24. Tsuzuki S.; Honda K.; Uchimaru T.; Mikami M. *J. Chem. Phys*. **2006**, 124, 114304.

25. Ahlrichs R.; Bar M.; Haser M.; Horn H.; Kolmel C. *Chem. Phys. Lett*. **1989**, 162, 165.

26. Hobza P.; Selzle H. L.; Schlag E. W. *Chem. Rev*. **1994**, 94, 1767.

27. Hobza P.; Selzle H. L.; Schlag E. W. *J. Phys. Chem*. **1993**, 97, 3937.

28. Hobza P.; Selzle H. L.; Schlag E. W. *J. Am. Chem. Soc*. **1994**, 116, 3500.

29. Watson D. M.; Fechtenkotter A.; Mullen K. *Chem. Rev*. **2001**, 101, 1267.

30. Zhang Z.; Huang H.; Yang X.; Zang L. *J. Phys. Chem. Lett*. **2011**, 2, 2897.

31. Zhang Z.; Che Y.; Smaldone R. A.; Xu M.; Bunes B. R.; Moore J. S.; Zang, L. *J. Am. Chem. Soc*. **2010**, 132, 14113.

32. Ghosh A.; Rao K. V.; Voggu R.; George S. J. *Chem. Phys. Lett*. **2010**, 488, 198.

33. Lauffer P.; Emtsev K. V.; Graupner R.; Seyller T.; Ley L. *Phys. Status Solidi B* **2008**, 245, 2064.

34. Sim D.; Liu D.; Dong X.; Xiao N.; Li S.; Zhao Y.; Li L. J.; Yan Q.; Hng H. H. *J. Phys. Chem. C* **2011**, 115, 1780.

35. Kabelac M.; Kroutil O.; Predota M.; Lankas F.; Sip M. *Phys. Chem. Chem. Phys*. **2012**, 14, 4217.

36. Saito R.; Dresselhaus G.; Dresselhaus M. S. *Imperial College Press: London*, **1998**.

37. Saenger W. *Springer-Verlag: New York*, **1984**.

38. Staii C.; Chen M.; Gelperin A.; Johnson A. T. *Nano Lett*. **2005**, 5, 1774.

39. Jeng E. S.; Moll A. E.; Roy A. C.; Gastala J. B.; Strano M. S. *Nano Lett*. **2006**, 6, 371.

40. Zheng M.; Jagota A.; Semke E. D.; Diner B. A.; McLean R. S.; Lustig S. R.; Richardson R. E.; Tassi N. G. *Nature Mater*. **2003**, 2, 338.

41. Zheng M.; Jagota A.; Strano M. S.; Santos A. P.; Barone P.; Chou S. G.; Diner B. A.; Dresselhaus M. S.; McLean R. S.; Onoa G. B.; Samsonidze G. G.; Semke E. D.; Usrey M.; Walls D. J. *Science* **2003**, 302, 1545.

42. Meng S.; Maragakis P.; Papaloukas C.; Kaxiras E. *Nano Lett*. **2007**, 7, 45.

43. Campbell J. F.; Tessmer I.; Thorp H. H.; Erie D. A. *J. Am. Chem. Soc*. **2008**, 130, 10648.

44. Jin H.; Jeng E. S.; Heller D. A.; Jena P. V.; Kirmse R.; Langowski J.; Strano M. S. *Macromolecules* **2007**, 40, 6731.

45. Dukovic G.; Balaz M.; Doak P.; Berova N. D.; Zheng M.; McLean R. S.; Brus L. E. *J. Am. Chem. Soc*. **2006**, 128, 9004.

46. Johnson R. R.; Johnson A. T. C.; Klein M. L. *Nano Lett*. **2008**, 8, 69.

47. Manohar S.; Tang T.; Jagota A. *J. Phys. Chem. C* **2007**, 111, 17835.

48. Deighan M.; Pfaendtner J. *Langmuir* **2013**, 29, 7999.

49. Mijajlovic M.; Penna M. J.; Biggs M J. *Langmuir* **2013**, 29, 2313.

50. Camden A. N.; Barr S. A.; Berry R. J. *J. Phys. Chem. B* **2013**, 117, 10691.

51. Antony J.; Grimme S. *Phys. Chem. Chem. Phys*. **2008**, 10, 2722.

52. Panigrahi S.; Bhattacharya A.; Banerjee S.; Bhattacharyya D. *J. Phys. Chem. C* **2012**, 116, 4374.

53. Zhang H.; Gruner G.; Zhao Y. *J. Mater. Chem. B* **2013**, 1, 2542.

54. Sanchez V. C.; Jachak A.; Hurt R. H.; Kane A. B. *Chem. Res. Toxicol*. **2011**, 25, 15.

55. Premkumar T.; Geckeler K. E. *Prog. Polym. Sci*. **2012**, 37, 515.

56. Burg A. B. *Inorg. Chem*. **1964**, 3, 1325.

57. Brown M. P.; Heseltine R. W. *Chem. Commun*. **1968**, 1551.

58. Brown M. P.; Heseltine R. W.; Smith P. A.; Walker P. J. *J. Chem. Soc. A* **1970**, 410.

59. Brown M. P.; Walker P. J. *Spectrochim. Acta* **1974**, 30, 1125.

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

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

62. Grimme S.; Antony J.; Ehrlich S.; Krieg H. *J. Chem. Phys*. **2010**, 132, 154104.

63. Auffinger P.; Hays F.; Westhof E.; Ho P. *Proc. Natl. Acad. Sci. U.S.A* **2004**, 101, 16789.

64. Hardegger L. A.; Kuhn B. S.; Anselm L.; Ecabert R.; Stihle M.; Gsell B.; Thoma R.; Diez J.; Benz J.; Plancher J. M.; Hartmann G.; Banner D. W.; Haap W.; Diederich F. *Angew. Chem. Int. Ed*. **2011**, 50, 314.

65. Guthrie F. *J. Chem. Soc*. **1863**, 16, 239.

66. Remsen I.; Norris J. F. *Am. Chem. J*. **1896**, 18, 90.

67. Mulliken R. S. *J. Am. Chem. Soc.* **1950**, 72, 600.

68. Mulliken R. S. *J. Am. Chem. Soc.* **1952**, 74, 811.

69. Mulliken, R. S. *J. Phys. Chem.* **1952**, 56, 801.

70. Clark T.; Hennemann M.; Murray J. S.; Politzer P. *J. Mol. Model*. **2007**, 13, 291.

71. Pavan M. S.; Prasad K. D.; Row T. N. G. *Chem. Commun*. **2013**, 49, 7558.

72. Politzer P.; Lane P.; Monice C. C.; Ma Y.; Murray J. S. *J. Mol. Model*. **2007**, 13, 305.

73. Lu Y.; Zou J.; Wang Y.; Jiang Y.; Yu Q. *J. Phys. Chem. A* **2007**, 111, 10781.

74. Lu Y.; Wang Y.; Zhu W. *Phys. Chem. Chem. Phys*. **2010**, 12, 4543.

75. Sirimulla S.; Bailey J. B.; Vegesna R.; Narayan M. *J. Chem. Inf. Model*. **2013**, 53, 2781.

76. Wilchen R.; Zimmermann M.; Lange A.; Zahn S.; Boeckler F. *J. Comput.-Aided Mol. Des.* **2012**, 26, 935.

77. Metrangolo P.; Neukirch H.; Pilati T.; Resnati G. *Acc. Chem. Res*. **2005**, 38, 386.

78. Murray J. S.; Riley K. E.; Politzer P.; Clark T. *Aust. J. Chem*. **2010**, 63, 1598.

79. Politzer P.; Murray J. S. *ChemPhysChem* **2013**, 14, 278.

80. De Paulis T.; Hemstapat K.; Chen Y.; Zhang Y.; Saleh S.; Alagille D.; Baldwin R.; Tamagnan G.; Conn P. *J. Med. Chem*. **2006**, 49, 3332.

81. Lu Y.; Shi T.; Wang Y.; Yang H.; Yan X.; Luo X.; Jiang H.; Zhu W. *J. Med. Chem*. **2009**, 52, 2854−2862.

82. Himmel D.; Das K.; Clark A.; Hughes S.; Benjahad A.; Oumouch S.; Guillemont J.; Coupa S.; Poncelet A.; Csoka I.; Meyer C.; Andries K.; Nguyen C.; Grierson D.; Arnold E. *J. Med. Chem.* **2005**, 48, 7582.

83. Fedorov O.; Huber K.; Eisenreich A.; Filippakopoulos P.; King O.; Bullock A. N.; Szklarczyk D.; Jensen L. J.; Fabbro D.; Trappe J.; Rauch U.; Bracher F. *Chem. Biol*. **2011**, 18, 67.

84. Ibrahim M. A. A. *J. Comput. Chem*. **2011**, 32, 2564.

85. Kolar M.; Hobza P. *J. Chem. Theory Comput*. **2012**, 8, 1325.

86. Wang J. M.; Wolf R. M.; Caldwell J. W.; Kollman P. A.; Case D. A. *J. Comput. Chem.* **2004**, 25, 1157.

87. Bayly C. I.; Cieplak P.; Cornell W.; Kollman P. A. *J. Phys. Chem.* **1993**, 97, 10269.

88. Mackerell AD. Jr.; Nilsson L. *Curr. Opin. Struct. Biol*. **2008**, 18, 194.

89. Reynolds C. A.; Essex J. W.; Richards. W. G. *J. Am. Chem. Soc.* **1992**, 114, 9075.

90. Cieplak P.; Cornell W. D.; Bayly C.; Kollman P. A. *J. Comput. Chem*. **1995**, 16, 1357.

91. Orozco M.; Noy A.; Perez A. *Curr. Opin. Struc. Biol*. **2008**, 18, 185.

92. Sorin E. J.; Rhee Y. M.; Pande V. S. *Biophys. J.* **2005**, 88, 2516.

93. Beveridge D. L.; McConnell K. J. *Curr. Opin. Struct. Biol*. **2000**, 10, 182.

94. Cheatham T. E.; Kollman P. A. *Annu. Rev. Struct. Dyn*. **2000**, 51, 435.

95. Cheatham T. E. *Curr. Opin. Struct. Biol.* **2004**, 14, 360.

96. Giudice E.; Lavery R. *Acc. Chem. Res*. **2000**, 35, 350.

97. Orozco M.; Perez A.; Noy A.; Luque F. *J. Chem. Soc. Rev*. **2003**, 32, 350.

98. Perez A.; Blas J. R.; Rueda M.; Lopez-Bes J. M.; de la Cruz X.; Orozco M. *J. Chem. Theory Comput*. **2005**, 1, 790.

99. Cheatham T. E.; Young M. A. *Biopolymers*. **2000**, 56, 232.

100. Varnai P.; Zakrzewska K. *Nucleic Acids Res*. **2004**, 32, 4269.

101. Beveridge D. L.; Barreiro G.; Byun K. S.; Case D. A.; Cheatham T. E.; Dixit S. B.; Giudice E.; Lankas F.; Lavery R.; Maddocks J. H.; Osman R.; Seibert E.; Sklenar H.; Stoll G.; Thayer K. M.; Varnai P.; Young M. A. *Biophys. J.* **2004**, 87, 3799.

102. Dixit S. B.; Beveridge D. L.; Case D. A.; Cheatham T. E.; Giudice E.; Lankas F.; Lavery R.; Maddocks J. H.; Osman R.; Sklenar H.; Thayer K. M.; Varnai P. *Biophys. J.* **2005**, 89, 3721.

103. Dixit S. B.; Beveridge D. L. *Bioinformatics*. **2006**, 22, 1007.

104. Perez A.; Marchan I.; Svozil D.; Sponer J.; Cheatham T. E.; Laughton C. A.; Orozco M. *Biophys. J.* **2007**, 92, 3817.

105. Zgarbova M.; Otyepka M.; Sponer J.; Mladek A.; Banas P.; Cheatham T. E.; Jurecka P. *J. Chem. Theory Comput*. **2011**, 7, 2886.

106. Svoboda P.; Cara A. D. *Cell. Mol. Life Sci.* **2006**, 63, 901.

107. Noller H. F. *Science* **2005**, 309, 1508.

108. Pljevaljcic G.; Millar D. P.; Deniz A. A. *Biophys. J.* **2004**, 87, 457.

109. Ma H. R.; Proctor D. J.; Kierzek E.; Kierzek R.; Bevilacqua P. C.; Gruebele M. *J. Am. Chem. Soc.* **2006**, 128, 1523.

110. Zhang W. B.; Chen S. J. *Biophys. J.* **2006**, 90, 765.

111. Zhang W. B.; Chen S. J. *Biophys. J.* **2006**, 90, 778.

112. Hyeon C.; Thirumalai D. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, 102, 6789.

113. Bundschuh R.; Gerland U. *Phys. Rev. Lett.* **2005**, 95, 208104.

114. Draper D. E. *RNA* **2004**, 10, 335.

115. Thirumalai D.; Hyeon C. *Biochemistry* **2005**, 44, 4957.

116. Bonnet G.; Krichevsky O.; Libchaber A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, 95, 8602.

117. Liphardt J.; Onoa B.; Smith S. B.; Tinoco I. Jr.; Bustamante C. *Science* **2001**, *292*, 733.

118. Garcia A. E.; Paschek D. *J. Am. Chem. Soc.* **2008**, 130, 815.

119. Hyeon C.; Thirumalai D. *J. Am. Chem. Soc.* **2008**, 130, 1538.

120. Bowman G. R.; Huang X.; Yao Y.; Sun J.; Carlsson G.; Guibas L. J.; Pande, V. S. *J. Am. Chem. Soc.* **2008**, 130, 9676.

121. Kuhrova P.; Banas P.; Best R. B.; Sponer J.; Otyepka M. *J. Chem. Theory Comput.* **2013**, 9, 2115.

122. Boys S. F.; Bernardi F. *Mol. Phys.* **1970**, 19, 553.

123. Jaffe R. L.; Smith G. D. *J. Chem. Phys*. **1996**, 105, 2780.

124. Sinnokrot M. O.; Valeev E. F.; Sherrill C. D. *J. Am. Chem. Soc*. **2002**, 124, 10887.

125. Cybulski S. M.; Lytle M. L. *J. Chem. Phys*. **2007**, 127, 141102.

126. Hesselmann A. *J. Chem. Phys*. **2008**, 128, 9.

127. Rezac J.; Hobza P. *J. Chem. Theory Comput.* **2013**, 9, 2151.

128. Cizek J. *J. Chem. Phys.* **1966**, 45, 4256.

129. Paldus J.; Shavitt I.; Cizek, *J. Phys. Rev. A* **1972**, 5, 50.

130. Bartlett R. J.; Musial M. *Rev. Mod. Phys.* **2007**, 79, 291.

131. Podeszwa R.; Szalewicz K. *J. Chem. Phys.* **2007**, 126, 194101.

132. Rode M.; Sadlej J.; Moszynski R.; Wormer P. E. S.; van der Avoird A. *Chem. Phys. Lett*. **1999**, 314, 326.

133. Rybak S.; Jeziorski B.; Szalewicz K. *J. Chem. Phys*. **1991**, 95, 6576.

134. Janowski T.; Pulay P. *J. Chem. Theory Comput.* **2008**, 4, 1585.

135. Lotrich V.; Flocke N.; Ponton M.; Yau A. D.; Perera A.; Deumens E.; Bartlett R. J. *J. Chem. Phys.* **2008**, 128, 194104.

136. Noga J.; Valiron P. *Mol. Phys.* **2005**, 103, 2123.

137. Olson R. M.; Bentz J. L.; Kendall R. A.; Schmidt M. W.; Gordon M. S. *J. Chem. Theory Comput.* **2007**, 3, 1312.

138. Halkier A.; Helgaker T.; Jorgensen P.; Klopper W.; Koch H.; Olsen J.; Wilson A. K. *Chem. Phys. Lett.* **1998**, 286, 243.

139. Helgaker T.; Klopper W.; Koch H.; Noga J. *J. Chem. Phys.* **1997**, 106, 9639.

140. Rezac J.; Riley K. E.; Hobza P. *J. Chem. Theory Comput.* **2011**, 7, 2427.

141. Koch H.; Fernandez B.; Christiansen O. *J. Chem. Phys.* **1998**, 108, 2784.

142. Tsuzuki S.; Honda K.; Uchimaru T.; Mikami M.; Tanabe K. *J. Am. Chem. Soc.* **2002**, 124, 104.

143. Riley K. E.; Hobza P. *J. Phys. Chem. A* **2007**, 111, 8257.

144. Hobza P.; Selzle H. L.; Schlag E. W. *J. Phys. Chem.* **1996**, 100, 18790.

145. Tsuzuki S.; Luthi H. P. *J. Chem. Phys.* **2001**, 114, 3949.

146. Tsuzuki S.; Uchimaru T.; Matsumura K.; Mikami M.; Tanabe K. *Chem. Phys. Lett.* **2000**, 319, 547.

147. Grimme S. *J. Chem. Phys.* **2003**, 118, 9095.

148. Becke A. D.; Johnson E. R. *J. Chem. Phys.* **2005**, 122, 154104.

149. Pitonak M.; Neogrady P.; Cerny J.; Grimme S.; Hobza P. *ChemPhysChem* **2009**, 10, 282.

150. Anand M.; Sunoj R. B.; Schaefer H. F. *J. Am. Chem. Soc*. **2014**, 136, 5535.

151. Jindal G.; Sunoj R. B. *J. Org. Chem.* **2014**, 79, 7600.

152. Sharma A. K.; Roy D.; Sunoj R. B. *Dalton Trans.* **2014**, 43, 10183.

153. Jindal G.; Sunoj R. B. *Angew. Chem. Int. Ed.* **2014**, 53, 4432.

154. Jindal G; Sunoj R. B. *Org. Biomol. Chem.* **2014**, 12, 2745.

155. Verma P; Verma P; Sunoj R. B. *Org. Biomol. Chem.* **2014**, 12, 2176.

156. Kuniyil R.; Sunoj R. B. *Org. Lett.* **2013**, 15, 5040.

157. Parija A.; Sunoj R. B. *Org. Lett.* **2013**, 15, 4066.

158. Rajeev R.; Sunoj R. B. *J. Org. Chem.* **2013**, 78, 7023.

159. Jurecka P.; Cerny J.; Hobza P.; Salahub D. R. *J. Comput. Chem.* **2007**, 28, 555.

160. Marrink S. J.; Risselada H. J.; Yefimov S.; Tieleman D. P.; de Vries A. H., *J. Phys. Chem. B* **2007,** 111, 7812.

161. Dobson C. M.; Sali A.; Karplus M. *Angew. Chem. Int. Ed.* **1998**, 37, 868.

162. Kollman P. *Chem. Rev.* **1993**, 93, 2395.

163. Straatsma T. P.; McCammon J. A. *J. Chem. Phys.* **1991**, 95, 1175.

164. Torrie G. M.; Valleau J. P. *J. Comput. Phys.* **1977**, 23, 187.

165. Liao A.; Parrinello M. *Proc. Natl. Acad. Sci. U.S.A* **2002**, 99, 12562.

166. Sugita Y.; Okamoto Y. *Chem. Phys. Lett.* **1999**, 314, 141.

167. Kumar S.; Bouzida D.; Swendsen R. H.; Kollman P. A.; Rosenberg J. M. *J. Comput. Chem.* **1992**, 13, 1011.

168. Barducci A.; Bonomi M.; Parrinello M. *WIREs Comput. Mol. Sci.* **2011**, 1, 826.

169. Liao A.; Rodriguez-Fortea A.; Gervasio F. L.; Ceccarelli M.; Parrinello M. *J. Phys. Chem. B* **2005**, 109, 6714.

170. Gervasio F. L.; Liao A.; Parrinello M. *J. Am. Chem. Soc*. **2005**, 127, 2600.

171. Barducci A.; Bussi G.; Parrinello M. *Phys. Rev. Lett.* **2008**, 100, 020603.

172. Bonomi M.; Parrinello M. *Phys. Rev. Lett.* **2008**, 104, 190601.

173. Bussi G.; Gervasio F. L.; Liao A.; Parrinello M. *J. Am. Chem. Soc*. **2005**, 128, 13435.

174. Deighan M.; Bonomi M.; Pfaendtner J. *J. Chem. Theory Comput.* **2012**, 8, 2189.

175. Zwanzig R. *J. Chem. Phys.* **1954**, 22, 1420.

176. Lu N.; Kofke D. *J. Chem. Phys.* **2001**, 114, 7303.

177. Lu N.; Kofke D. *J. Chem. Phys.* **2001**, 115, 6866.

178. Lu N.; Singh J.; Kofke D. *J. Chem. Phys.* **2003**, 118, 2977.

179. Jarzynski C. *Phys. Rev. Lett.* **1997**, 78, 2690.

180. Hendrix D.; Jarzynski C. *J. Chem. Phys.* **2001**, 114, 5974.

181. Crooks G. *J. Statis. Phys.* **1998**, 90, 1481.

182. Goette M.; Grubmuller H. *J. Comput. Chem.* **2008**, 30, 447.

183. Zacharias M.; Straatsma T. P.; McCammon J. A. *J. Chem. Phys.* **1994**, 100, 9025.

184. Beutler T. C.; Mark A. E.; van Schaik R. C.; Gerber P. R.; van Gunsteren W. F. *Chem. Phys. Lett.* **1994**, 222, 529.

185. Hess B.; Kutzner C.; van der Spoel D.; Lindahl E. *J. Chem. Theory Comput.* **2008**, 4, 435.

186. Novoselov K. S.; Geim A. K.; Morozov S. V.; Jiang D.; Zhang, Y.; Dubonos S. V.; Grigorieva I. V.; Firsov A. A. *Science* **2004**, 306, 666.

187. Novoselov K. S.; Geim A. K.; Morozov S. V.; Jiang D.; Katsnelson M. I.; Grigorieva I. V.; Dubonos S. V.; Firsov A. A. *Nature* **2005**, 438, 197.

188. Tian X.; Xu j.; Wang X. *J. Phys. Chem. B* **2010**, 114, 11377.

189. Zhang Z.; Huang H.; Yang X.; Zang L. *J. Phys. Chem. Lett*. **2011**, 2, 2897.

190. Ishikawa R.; Bando M.; Morimoto Y.; Sandhu A. *Nanoscale Res. Lett*. **2011**, 6, 111.

191. Lu C. H.; Yang H. H.; Zhu C. L.; Chen X.; Chen G. N. *Angew. Chem. Int. Ed.* **2009**, 48, 4785.

192. Shan C. S.; Yang H. F.; Song J. F.; Han D. X.; Ivaska A.; Niu L. *Anal. Chem.* **2009**, 81, 2378.

193. Baby T. T.; Aravind S. S. J.; Arockiadoss T.; Rakhi R. B.; Ramaprabhu S. *Sens. Actuators, B* **2010**, 145, 71.

194. Chen W.; Chen S.; Qi D. C.; Gao X. Y.; Wee A. T. S. *J. Am. Chem. Soc.* **2007**, 129, 10418.

195. Lu Y. H.; Chen W.; Feng Y. P.; He P. M. *J. Phys. Chem. Lett.* **2009**, 113, 2.

196. Coletti C.; Riedel C.; Lee D. S.; Krauss B.; Patthey L.; von Klitzing K.; Smet J. H.; Starke U. *Phys. Rev. B* **2010**, 81, 235401.

197. Steele R. P.; DiStasio R. A.; Gordon M. H. *J. Chem. Theory Comput.* **2009**, 5, 1560.

198. Grimme S.; Antony J.; Ehrlich S.; Krieg H. *J. Chem. Phys*. **2010**, 132, 154104.

199. Rezac J.; Fanfrlik J.; Salahub D. R.; Hobza P. *J. Chem. Theory Comput.* **2009**, 5, 1749.

200. Korth M.; Pitonak M.; Rezac J.; Hobza P. *J. Chem. Theory Comput.* **2010**, 6, 344.

201. Rezac J.; Hobza P. *J. Chem. Theory Comput.* **2012**, 8, 141.

202. Elstner M.; Poreza G. D.; Jungnickel G.; Elsner J.; Haugk M.; Frauenheim T.; Suhai S.; Seifert G. *Phys. Rev. B* **1998**, 58, 7260.

203. Elstner M.; Hobza P.; Frauenheim T.; Suhai S.; Kaxiras E. *J. Chem. Phys.* **2001**, 114, 5149.

204. Wang J. M.; Wolf R. M.; Caldwell J. W.; Kollman P. A.; Case D. A. *J. Comput. Chem.* **2004**, 25, 1157.

205. Wang J.; Wang W.; Kolmann P. A.; Case D. A. *J. Mol. Graphics Modell.* **2006**, 25, 247.

206. Reed E. A.; Weinstock B. R.; Weinhold F. *J. Chem. Phys.* **1985**, 83, 735.

207. Klamt A.; Schuurmann G. *J. Chem. Soc., Perkin Trans.* **1993**, 2, 799.

208. Marenich A. V.; Cramer C. J.; Truhlar D. G. *J. Phys. Chem. B* **2009**, 113, 6378.

209. Hawkins G. D.; Cramer C. J.; Truhlar D. G. *J. Phys. Chem.* **1996**, 100, 19824.

210. Jorgensen W. L.; Chandrasekhar J.; Madura J. D.; Impey R. W.; Klein M. L. *J. Chem. Phys.* **1983**, 79, 926.

211. Berendsen H. J. C.; Postma J. P. M.; Van Gunsteren W. F.; Dinola A.; Haak J. R. *J. Chem. Phys.* **1984**, 81, 3684.

212. Nose S. *Mol. Phys*. **1984**, 52, 255.

213. Hoover W. G. *Phys. Rev. A* **1985**, 31, 1695.

214. Parrinello M.; Rahman A. *J. Appl. Phys.* **1981**, 52, 7182.

215. Darden T.; York D.; Pedersen L. *J. Chem. Phys.* **1993**, 98, 10089.

216. Chakravarti A. *Science* **2011,** 333, 53.

217. Watson J. D.; Crick F. H. C. *Nature* **1953**, 171, 737.

218. Jurecka P.; Hobza P. *J. Am. Chem. Soc.* **2003**, 125, 15608.

219. Cerny J.; Kabelac M.; Hobza P. *J. Am. Chem. Soc.* **2008**, 130, 16055.

220. Kolar M.; Kubar T.; Hobza P. *J. Phys. Chem. B* **2011**, 115, 8038.

221. Kratochvil M.; Sponer J.; Hobza P. *J. Am. Chem. Soc.* **2000**, 122, 3495.

222. Kabelac M.; Hobza P. *Chem. Eur. J.* **2001**, 7, 2067.

223. Kabelac M.; Zendlova L.; Reha D.; Hobza P. *J. Phys. Chem. B* **2005**, 109, 12206.

224. Bernal, J. D. *The Physical Basis of Life*, (Routledge and Kegan Paul, London, **1951**).

225. Rao M.; Odom D. G.; Oro J. *J. Mol. Evol.* **1980**, 15, 317.

226. Cairns Smith A. G. *Clay Minerals and the Origin of Life*, Cambridge University Press, Cambridge, UK, **1986**.

227. Ferris J. P.; Ertem G.; Agarwal V. K. *Orig. Life Evol. Biosph.* **1989**, 19, 153.

228. Ferris J. P.; Ertem G. *Orig. Life Evol. Biosph.* **1992**, 22, 369.

229. Ferris J. P. *Orig. Life Evol. Biosph.* **1993**, 23, 307.

230. Stotzky J. V.; Gallori E.; Khanna M.; Akkermans A. D. L.; van Elsas Y. D.; Brujin F. J. (Eds.), *Transformation in Soil, Molecular Microbial Ecology Manual*, Kluwer, Dordrecht, **1996**, 1.

231. Tosoni S.; Civalleri B.; Ugliengo P. *J. Phys. Chem. C* **2010**, 114, 19984.

232. Hohenberg P.; Kohn W. *Phys. Rev*. **1964**, 136, 864.

233. Kohn W.; Sham L. J. *Phys. Rev*. **1965**, 140, 1133.

234. Cerny J.; Jurecka P.; Hobza P.; Valdes H. *J. Phys. Chem. A* **2007**, 111, 1146.

235. Duan Y.; Wu C.; Chowdhury S.; Lee M. C.; Xiong G.; Zhang W.; Yang R.; Cieplak P.; Luo R.; Lee T. *J. Comput. Chem.* **2003**, 24, 1999.

236. Ahlrichs A.; Bar M.; Haser M.; Horn H.; Kolmel C. *Chem. Phys. Lett*. **1989**, 162, 165.

237. 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. *Gaussian 09, revision A.1; Gaussian,* Inc.; Wallingford, CT, **2009**.

238. Hess B.; Bekker H.; Berendsen H. J. C.; Fraaije J. G. E. M. *J. Comput. Chem.* **1997**, 18, 1463.

239. Bonomi M.; Branduardi D.; Bussi G.; Camilloni C.; Provasi D.; Raiteri P.; Donadio D.; Marinelli F.; Pietrucci F.; Broglia R. A. *Comput. Phys. Commun.* **2009**, 180, 1961.

240. Spiwok V.; Hobza P.; Rezac J. *J. Phys. Chem. C* **2011**, 115, 19455.

241. Kolař M. H.; Carloni P.; Hobza P. *Phys. Chem. Chem. Phys.* **2014**, 16, 19111.

242. Metrangolo P.; Resnati G. *Chem. Eur. J.* **2001**, 7, 2511.

243. Lu Y.; Liu Y.; Xu Z.; Li H.; Liu H.; Zhu W. *Expert Opin. Drug Discov.* **2012**, 7, 375.

244. Wilcken R.; Zimmermann M. O.; Lange A.; Joerger A. C.; Boeckler F. M. *J. Med. Chem.* **2013**, 56, 1363.

245. Scholfield M. R.; Zanden C. M. V.; Carter M.; Ho P. S. *Protein Sci.* **2013**, 22, 139.

246. Riley K. E.; Murray J. S.; Fanfrlik J.; Řezač J.; Sola R. J.; Concha M. C.; Ramos F. M.; Politzer, P. *J. Mol. Model.* **2011**, 17, 3309.

247. Kolař M.; Hostaš J.; and Hobza P. *Phys, Chem. Chem. Phys*, **2014**, 16, 9987.

248. Fanfrlik J.; Kolař M.; Kamlar M.; Hurny D.; Ruiz F. X.; Cousido-Siah A.; Mitschler A.; Řezač J.; Munusamy E.; Lepšik M. et al. *ACS Chem. Biol.* **2013**, 8, 2484.

249. Fanfrlik J.; Bronowska A. K,; Rezac J.; Prenosil O.; Konvalinka J.; Hobza P. *J. Phys. Chem. B* **2010**, 114, 12666.

250. Steuber H.; Heine A.; Klebe G. *J. Mol. Biol.* **2007**, 368, 618.

251. Rezac J.; Hobza P. *J. Chem. Theory Comput.* **2012**, 8, 141.

252. Rezac J.; Hobza P. *Chem. Phys. Lett.* **2011**, 506, 286.

253. Aaqvist J. *J. Phys. Chem.* **1990**, 94, 8021.

254. Sponer J.; Jurecka p.; Hobza p. *J. Am. Chem. Soc.* **2004**, 126, 10142.

255. Marshall M. S.; Burns L. A.; Sherrill C. D. *J. Chem. Phys.* **2011**, 135, 194102.

256. Rezac J. Hobza P. *J. Chem. Theory Comput.* **2014**, 10, 3066.

257. Cerny j.; Tong X.; Hobza P.; Dethlefs K. M. *J. Chem. Phys.* **2008**, 128, 114319.

258. Weigend F.; Häser M. *Theory. Chem. Acc.* **1997**, 97, 331.

259. Grimme S.; Ehrlich S.; Goerigk L. *J. Comput. Chem*. 2011, **32,** 1456.

260. Barone V. *J. Chem. Phys.* **2005**, 122, 014108.

261. Muchova E.; Spirko V.; Hobza P. Nachtigallova D. *Phys. Chem. Chem. Phys.* **2006**, 8, 4866.

262. Schaftenaar G.; Noordik J. H. *J. Comput. Aided Mol. Design.* **2000**, 14, 123.

263. Werner H. J.; Knowles P. J.; Lindh R.; Schu¨tz M.;, Celani P.; Korona T.; Manby F. R.; Rauhut G.; Amos R. D.; Bernhardsson A.; Berning A.; Cooper D. L.; Deegan M. J. O.; Dobbyn A. J.; Eckert F.; Hampel C.; Hetzer G.; Lloyd A. W.; McNicholas S. J.; Meyer W.; Mura M. E.; Nicklass A.; Palmieri P.; Pitzer R.; Schumann U.; Stoll H.; Stone A. J.; Tarroni R.; Thorsteinsson T. Molpro, version **2002.6**, a package of *ab initio* programs, **2003**.

264. Kabelac M.; Hobza P.; Spirko V. *Phys. Chem. Chem. Phys.* **2009**, 11, 3885.

265. Spirko V. *J. Phys. Chem. A* **2011**, 115, 11313.

266. Burgi T.; Droz T.; Leutwyler S. *Chem. Phys. Lett.* **1995**, 246, 291.

267. Mons M.; Dimicoli I.; Tardivel B.; Piuzzi F.; Brenner V.; Millie P., *J. Phys. Chem. A* **1999**, 103, 9958.

268. Mons M.; Dimicoli I.; Piuzzi F. *Int Rev. Phys. Chem.* **2002**, 21, 101.

269. Courty A.; Mons M.; Domicoli I.; Piuzzi F.; Brenner V.; Millie P. *J. Phys. Chem. A* **1998**, 102, 4890.

270. Brenner V.; Martrenchard S.; Millie P.; Dedonder-Lardeux C.; Jouvet C.; Solgadi D. *J. Phys. Chem.* **1995**, 99, 5848.

271. Walters E. A.; Grover J. R.; White M. G. *Z. Phys. D*, **1986**, 4, 103.

272. Krause H.; H. Neusser H. J. *J. Chem. Phys.* **1993**, 99, 6278.

273. Grebner Th. L.; Stumpf R.; Neusser H. J. *Int. J. Mass Spectrom. Ion Processes* **1997**, 167, 649.

274. Droz T.; Burgi T.; Leutwyler S. *J. Chem. Phys.* **1995**, 103, 4035.

275. Grebner Th. L.; Unold P. V.; Neusser H. J. *J. Phys. Chem. A* **1997**, 101, 158.

276. Braun J. E.; Grebner Th. L.; Neusser H. J. *J. Phys. Chem. A* **1998**, 102, 3273.

277. Grebner Th. L.; Neusser H. J. *Chem. Phys. Lett.* **1995**, 245, 578.

278. Rezac J.; Hobza P. *J. Chem. Theory Comput.* **2014**, 10, 3066.

279. Uhlenbeck O. C. *Nature* **1990**, 346, 613.

280. Jucker F. M.; Heus H. A.; Yip P. F.; Moors E. H.; Pardi, A. *J. Mol. Biol.* **1996**, 264, 968.

281. Woese C. R.; Winker S.; Gutell R. R. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, 87, 8467.

282. Banas P.; Hollas D.; Zgarbova M.; Jurecka P.; Orozco M.; Cheatham III T. E.; Sponer J.; Otyepka M. *J. Chem. Theory Comput.* **2010**, 6, 3836.

283. Fiore J. L.; Nesbitt D. J. *Q. Rev. Biophys.* **2013**, 46, 223.

284. Varani G.; Cheong C.; Tinoco I. Jr. *Biochem.* **1991**, 30, 3280.

285. Jaeger L.; Michel F.; Westhof E. *J. Mol. Biol.* **1994**, *236*, 1271.

286. Hsiao C.; Mohan S.; Hershkovitz E.; Tannenbaum A.; Williams L. D. *Nucleic Acids Res.* **2006,** 34, 1481.

287. Allain F. H.; Varani G.; *J. Mol. Biol.* **1995**, 250, 333.

288. Correll C. C.; Swinger, K. *RNA* **2003**, 9, 355.

289. Heus H. A.; Pardi A. *Science* **1991**, 253, 191.

290. Leontis N. B.; Westhof E. *Curr. Opin. Struct. Biol.* **2003**, 13, 300.

291. Chen A. A.; Garcia A. E. *Proc. Natl. Acad. Sci. USA* **2013**, 110, 16820.

292. Chakraborty D.; Collepardo-Guevara R.; Wales D. J. *J. Am. Chem. Soc.* **2014**, 136, 18052.

293. Bottaro S.; Di Palma F.; Bussi G. *Nucleic Acids Res.* **2014**, 42, 13306.

294. 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. *J. Am. Chem. Soc.* **1996**, 118, 2309.

295. Mlynsky V.; Banas P.; Hollas D.; Reblova K.; Walter N. G.; Sponer J.; Otyepka M. *J. Phys. Chem. B* **2010**, 114, 6642.

296. Sponer J.; Mladek A.; Sponer J. E.; Svozil D.; Zgarbova M.; Banas P.; Jurecka P.; Otyepka M. *Phys. Chem. Chem. Phys.* **2012**, 14, 15257.

297. Correll C. C.; Beneken J.; Plantinga M. J.; Lubbers M.; Chan Y. L. *Nucleic Acids Res.* **2003**, *31*, 6806.

298. Nozinovic S.; Furtig B.; Jonker H. R. A.; Richter C.; Schwalbe H. *Nucleic Acids Res.* **2010**, 38, 683.

299. Bussi G.; Zykova-Timan T.; Parrinello M. *J. Chem. Phys.* **2009**, 130. 074101.

300. Branduardi D.; Bussi G.; Parrinello M. *J. Chem. Theory Comput.* **2012**, 8, 2247.

301. Bussi G.; Gervasio F. L.; Laio A.; Parrinello M. *J. Am. Chem. Soc.* **2006**, 128, 13435.

302. Limongelli V.; De Tito S.; Cerofolini L.; Fragai M.; Pagano B.; Trotta R.; Cosconati S.; Marnelli, L.; Novellino E.; Bertini I.; Randazzo A.; Luchinat C.; Parrinello M. *Angew. Chem. Int. Ed.* **2013**, 52, 2269.

303. Bian Y.; Tan C.; Wang J.; Sheng Y.; Zhang J.; Wang W. *PLoS Comput. Biol.* **2014**, 10, 1003562.

304. Zgarbova M.; Otyepka M.; Sponer J.; Lankas F.; Jurecka P. *J. Chem. Theory Comput.* **2014**, 10, 3177.

305. Proctor D. J.; Ma H. R.; Kierzek E.; Kierzek R.; Gruebele M.; Bevilacqua P. C. *Biochem.* **2004**, *43*, 14004.

306. Sorin E. J.; Engelhardt M. A.; Herschlag D.; Pande S. *J. Mol. Biol.* **2002**, 317, 493.

List of Attached Publications

This thesis is based on the following five articles

Attachment A:
Haldar S.; Kolar M.; Sedlak R.; Hobza P. "Adsorption of Organic Electron Acceptors on Graphene-like Molecules: Quantum Chemical and Molecular Mechanical Study". *J. Phys. Chem. C* **2012**, 116, 25328.

Attachment B:
Haldar S.; Spiwok V.; Hobza P. "On the Association of the base pairs on the Silica Surface Based on Free Energy Biased Molecular Dynamics Simulation and Quantum Mechanical Calculations". *J. Phys. Chem. C* **2013**, 117, 11066.

Attachment C:
Fanfrlik J.; Ruiz F. X.; Kadlcikova A.; Rezac J.; Cousido-Siah A.; Mitschler A.; **Haldar S**.; Lepsik M.; Kolar M. H.; Majer P.; Podjarny A. D.; Hobza P. "The Effect of Halogen-to-Hydrogen Bond Substitution on Human Aldose Reductase Inhibition". *ACS Chem. Biol.* **2015**, 10, 1637.

Attachment D:
Haldar S.; Gnanasekaran R.; Hobza P. "A Comparison of *ab initio* Quantum-Mechanical and Experimental D_0 Binding Energies of Eleven H–bonded and Eleven Dispersion–bound Complexes". *Phys. Chem. Chem. Phys*. **2015**, 17, 26645.

Attachment E:
Haldar S.; Kuhrova P.; Banas P.; Spiwok V.; Sponer J.; Hobza P.; Otyepka M. "Insights into Stability and Folding of GNRA and UNCG Tetraloops Revealed by Microsecond Molecular Dynamics and Well Tempered Metadynamics". *J. Chem. Theory Comput.* **2015**, 11, 3866.

Attached Publications