STUDY OF DRUG-DNA INTERACTION USING QUANTUM MECHANICAL TOOLS

THESIS SUBMITTED FOR THE AWARD OF THE DEGREE OF

Doctor of Philosophy

in

Applied Physics

By

Ruchi Mishra

Enrolment No.: 266/13

Under the Supervision of

Dr. Ramesh Chandra

BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY
LUCKNOW
प्रज्ञा शील करुणा
ESTABLISHED 1996

Department of Applied Physics
School for Physical Sciences
Babasaheb Bhimrao Ambedkar University, Lucknow-226025
U.P. (India)

2018

THIS THESIS IS DEDICATED
TO
DEVESH SIR
AND
MY FAMILY

DECLARATION

I declare that the thesis entitled "**Study of Drug-DNA Interaction using Quantum Mechanical Tools**" has been prepared by me under the supervision of **Dr. Ramesh Chandra,** Assistant Professor, Department of Applied Physics, School for Physical Sciences, Babasaheb Bhimrao Ambedkar University, Lucknow. No part of this thesis has formed the basis for the award of any degree, diploma or fellowship previously. I, declare that the material embodied in the present work is based on original research work and the indebteness to others has been duly acknowledged at relevant places. This is also declared that the thesis is essentially free from any kinds of plagiarism.

Ruchi Mishra

(Ruchi Mishra)
Department of Applied Physics
School for Physical Sciences
Babasaheb Bhimrao Ambedkar University, Lucknow

Date: 24-9-18
Place: Lucknow

CERTIFICATE

This is to certify that the thesis titled "**Study of Drug-DNA Interaction using Quantum Mechanical Tools**" submitted by **Ms. Ruchi Mishra** is an original research work and has not been previously submitted in part or full for the award of any other degree or diploma to this or any other university or institutions.

The thesis submitted to the Babasaheb Bhimrao Ambedkar University, Lucknow, satisfies all the requirements as stipulated in the *Doctor of Philosophy (Ph.D.) regulations -1999 as amended in 2010* and it is fit for the submission and evaluation for the award of Doctor of Philosophy of the University.

Date: 24-9-2018

24-9-2018
Supervisor

24/9/18
Head of the Department

ACKNOWLEDGEMENT

It gives me immense pleasure that I have an opportunity to place on record, the contributions of several people, who have been influential in crystallizing this thesis. I would like to thank the almighty for the blessings. I have been benefited from numerous people during the course of this work and I attempt to list these people here, and I fervently hope I have not missed anyone. First and foremost, I would like to express my sincere gratitude to my supervisor **Dr. Ramesh Chandra**, for his continuous support in my Ph.D. work and related research, for his patience, motivation, and immense knowledge. Out of many qualities the one that I really tried to imbibe from him is to stay calm in adverse situations, which is quite essential in carrying out good research. I feel extremely privileged to get an opportunity to work with **Prof. Devesh Kumar**, Head, Department of Applied Physics who has been my inspiration and has provided me with constant support, encouragement and a strong guidance. Apart from this, he patiently taught me that there is a solution to every computational challenge. Without his valued support, the outcome of my theoretical research would have not been the same. I would also like to thank the rest of faculties of my department: **Prof. B. C. Yadav**, **Dr. A. K. Yadav**, **Dr. Devendra Singh** and **Dr. K. B. Thapa** for their encouragement, insightful comments, and hard questions during the DRC meetings. Their critical review and valuable comments on my presentations always helped me to convey the main message easily to the audiences during the oral presentations. I would like to express my heartfelt reverences towards **Dr. G. Narahari Sastry** from Indian Institute of Chemical

Technology, Hyderabad for the scientific discussion on various topics. During my Ph.D., I visited many times to his laboratory.

I am extremely grateful to my friend **Miss Anamika Singh Gaur** for her help to begin my work and constant affectionate support. I have learnt a lot from her, through both personal and scholarly interactions. I extend my thanks to my research group (Dr. Jitendra Kumar, Dr. Pranav Upadhaya, Dr. Suresh Kumar, Mr. Asheesh Kumar, Miss Yash Kaur Singh, Mr. Ratindra Gautam, Mr. Vivek Kumar Nautiyal, Mr. Deep Kumar, Mr. Dharmveer Singh, Mr. Narinder Kumar, Mr. Rajkamal Shastri, Mr. Surya Pratap Goutam and Mr. Ankur Trivedi) and to my juniors (Miss Anamika Shukla, Miss Rolly Yadav, Miss Nidhi Awasthi, and Miss Shivani Chaudhary) who provided a friendly and cooperative environment in the lab during my Ph. D. period. At this point I want to thank one and all who directly or indirectly helped me during my Ph.D.

Next, I would like to thank my parents whose love, guidance and blessings were always with me. Whatever I have achieved till date is due to their sacrifice in their life to fulfill my demands. A warm thanks to my younger brothers **Anurag Mishra** and **Ashutosh Mishra** for making my hectic and stressful Ph.D. time cheerful with all their love, care and sense of humor.

Finally I would like to appreciate the support and facilities provided by the Staff of Library and the Computer Center of the Department as well as the university for necessary literature survey and other requirements.

Ruchi Mishra

(Ruchi Mishra)

ABSTRACT

The fundamental problems in drug discovery are based on the process of molecular recognition by small molecules. DNA has been known as cellular target for many anticancer and antitumor agents. Understanding the interaction of drug molecules with DNA has become an active area of research in the recent past. The binding specificity of DNA-small molecule is identified mainly by studying the hydrogen bonding and polar interactions. Majority of the DNA binding small molecules and their mechanism of action at the molecular level are not well studied. As these small molecules can act as effective therapeutic agents against many diseases, there is a need to have the detailed mechanistic insight on how they interact with DNA. Molecular docking and other molecular modelling techniques are well parameterized for protein-ligand and also for the protein-protein interaction; but there is a lack of systematic parameterization involving nucleic acid ligand binding.

In the present study the mechanism of molecular interaction of different types of DNA binding small molecules (intercalators and minor groove binders) and stability of the complexes using available molecular modelling methods such as Molecular Docking, Molecular Dynamics Simulation, QM/MM etc. has been investigated. Molecular Docking predicts the mode of binding of drug molecule to the DNA. On the other hand Molecular Dynamics simulations assure the stability of the drug-DNA complex. Quantum-mechanics/molecular-mechanics (QM/MM) approach is used for the modelling of reaction mechanism in biomolecular systems which are not understood so far. This study confirms that the intercalators and minor groove binders bind to DNA through different

types of non-covalent interactions. Intercalators are stabilized by π-π stacking between the drug and DNA bases while, DNA minor groove binders form hydrogen bonds between the drug and DNA bases end terminal. DNA recognition by drug molecules does not directly depend on the bases sequence but also depends on the sequence dependent conformation or DNA modifications and distortions. This study provides full understanding of drug-DNA interactions, sequence selectivity and sequence specificity.

PREFACE

Deoxyribonucleic acid (DNA) has a great biological significance. Transcription and replication are the vital processes essential for the survival of the living system. Transcription is responsible for the construction of proteins, while replication produces copies of DNA. Most of the anticancer therapies are based on the interaction of drugs with DNA. To design effective chemotherapeutic agents and better anticancer drugs, it is essential to explore the interaction of drug with DNA. The intercalation and groove binding are the two important modes of binding of drug with DNA. Electrostatic interactions, surface binding to their minor or major grooves, intercalation between adjacent base pairs, covalent attachments to the double helix, or electrostatic binding. Both covalent as well as non-covalent interactions were found between small molecules and DNA. Here we develop such a theoretical model based on a molecular level, which would eventually assist and complement the rational drug- design attempts. The work carried out for the Ph.D. thesis has been divided into 7 chapters.

Chapter 1 discusses about the structure of nucleic acid, different types of nucleic acids interaction between drugs and the different types of experimental and theoretical methods which are used to study the interactions between nucleic acid and drug. This chapter also presents the literature survey of work done so for on drug-DNA interaction which focuses on experimental and theoretical studies.

Chapter 2 discusses about the different types of molecular modelling techniques used to study the drug-DNA interaction.

Chapter 3 presents the comparative studies of DNA minor groove binders using AMBER and GROMACS MD simulation and binding energy between DNA and ligands were calculated.

Chapter 4 illustrates the molecular interaction of some DNA intercalators from acridines, anthracyline and anthraquinone group of family using molecular docking, molecular dynamics simulation (MD) and quantum mechanics/molecular mechanics (QM/MM).

Chapter 5 illustrates the binding mechanism of DNA minor groove binders using different molecular modelling techniques. This study describes the role of different types of non-covalent interaction during the formation of DNA and minor groove binder complex formation.

Chapter 6 explains the role of DNA sequence to predict the exact mode of binding of any drug with the DNA. This chapter also reveals that the binding mode of any ligand cannot be exactly predicted on the basis of molecular docking. MD simulation is also required to predict the stability of the complexes.

Chapter 7 presents the general conclusion of the entire work presented in the thesis.

LIST OF ABBREVIATIONS

A	Adenine
AAC	N-[2-(dimethylamino)-ethyl]-9-aminoacridine-4-carboxamide
ADM	Adriamycin
AFM	Atomic Force Microscopy
AMBER	Assisted Model Building with Energy Refinement
APBS	Adaptive Poisson-Boltzmann Solver
ATP	Adenosine triphosphate
BEs	Binding Energies
C	Cytosine
CC	Coupled Cluster Theory
CD	Circular Dichroism
CHARMM	Chemistry at HARward Macromolecular Mechanics
CI	Configuration Interaction
cNDI 1	Cyclic Napthalene diimide Derivatives
cNDI 2	Non-Cyclic Naphthalene diimide Derivatives
Ct-DNA	Calf Thymus DNA
DACA	N-[2-(dimethylamino)-ethyl]-acridine-4-carboxamide
DAPI	6-amidine-2-(4-amidino-phenyl)indole
DFT	Density Functional Theory
DIT	Ditercalinium
DMA	N, N-dimethylaniline

DMAADD	9-N,N-dimethylaniline decahydroacridinedione
DNA	De-oxy ribonucleic acid
DNM	Daunomycin
dsDNA	Double strand DNA
G	Guanine
GA	Genetic Algorithms
GAFF	Generalized Amber Force Field
GLIDE	Grid-based Ligand Docking with Energetics
GOLD	Genetic Optimisation for Ligand Docking
GTOs	Gaussian Type Orbitals
HF	Hartree Fock
IBB	5-(2-imidazolinyl)-2-[2-(4-hydroxyphenyl)-5-benzimidazolyl]benzimidazole
IR	Infrared
MC	Monte Carlo
MCSCF	Multi configurational self-consistent Field
MD	Molecular Dynamics
MGBs	Minor Groove Binders
MM	Molecular Mechanics
MMGBSA	Molecular Mechanics Generalized-Born Surface Area
MMPBSA	Molecular Mechanics Poisson-Boltzmann Surface Area
MPn	Moller-Plesset perturbation Theory
NDB	Nucleic acid Data Bank

NDIs	Napthalene diimide Derivatives
NMR	Nuclear Magnetic Resonance
PBC	Periodic Boundary Conditions
PDB	Protein Data Bank
PME	Particle Mesh Ewald
PNAs	Peptide nucleic Acids
QM	Quantum Mechanics
QM/MM	Quantum Mechanics/Molecular Mechanics
QSAR	Quantitative structure-activity relationship
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuations
RNA	Ribonucleic Acid
STOs	Slater Type Orbitals
T	Thymine
TFOs	Triplex-forming oligonucleotides
U	Uracil
VS	Virtual Screening

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

Chapter 1

Introduction

Chapter 1

Introduction

The work presented in this thesis is mainly focused on *in silico* studies of Drug-DNA interaction. Molecular interaction between the drugs and deoxyribonucleic acid (DNA) is a field of current research and also plays an important to its biological activity. Many anticancer therapies depends upon the interaction of drug molecule with DNA. These interactions may cause damage of DNA in cancerous cells by inhibiting the process of replication or transcription, which inhibits the growth of cancer cells.[1-3] To design efficient chemotherapeutic agents and better anticancer drugs, it is important to inspect the interaction of drug with DNA. The number of known DNA based drug targets are very limited in comparison to the protein based drug targets and also the number of available structures of DNA-drug complexes is small relative to protein-drug complexes deposited in the PDB.[4]

Molecular modeling investigation depends on effective potentials which has great significance in describing the structure, dynamics and energetics of DNA in free State [5, 6] and in complex with small molecules.[7, 8] There are many force fields and protocols which are developed for large bimolecular systems (including DNA/protein complexes) for which simulation times can reach up to the 100 ns time scale inspite of this there are several difficulties in the fields of DNA-modeling which appears due to the use of effective potentials. Because there are many phenomena which depends upon the

electronic structure in such a manner that they cannot be modeled via effective potential. Other reasons are the bond-breaking and bond-forming processes, which arises when the drugs binds with DNA covalently.

Another method for the investigation of such systems is Quantum Mechanics/ Molecular Mechanics (QM/MM) approaches, in which the active site is treated at the quantum mechanical level and rest of the bimolecular system which includes DNA, solvent and the counter ions, is treated with molecular mechanics. The solvent surroundings have been important for the study of electronic properties in case of DNA.[9-12]

1.1 De-oxy ribonucleic Acid

1.1.1 History

DNA was first discovered by Friedrich Miescher, when he was working with white blood cells obtained from pus drained out of surgical bandages and determined that the DNA was rich in phosphorous and acidic. However the role of DNA to store heredity information was not reported before 1940s until Avery and co-workers published that the nucleic acids are the genetic information carriers and not proteins.[13] In 1950 Chargaff recognized that the composition of DNA is unique for each and every species. Chargaff also found that when DNA is broken into its components, the amount of guanine fluctuated from one organism to another but always equal to the cytosine and the amount of cytosine is equal to the amount of thymine.[14] Rosalind Frankllin elucidates basic helical structure of DNA on the basis of X-ray crystallography technique. In 1953, Watson and Crick scooped Franklin's and Chargaff's information and cracked the code of DNA structure.[15-17] They recognized that the relationship between the nitrogenous bases suggested by Chargaff may be due to the complementary base pairing between

adenine-thymine and guanine-cytosine and due to this type of base pairing they discovered the hydrogen bonding between these bases, which is currently known as Watson-Crick hydrogen bonding. With this information they modeled a right-handed double helical structure of DNA in which phosphate backbone lied outside the helix and the bases are held together by hydrogen bonding pointed towards the center of helix.

In 1979, a first crystal structure of left-handed double helical DNA $d(CGCGCG)_2$ at atomic resolution was published, known as Z-DNA.[18] After a year, the single-crystal structure analysis of right-handed B-DNA, with the self-complementary dodecamer sequence $d(CGCGAATTCGCG)_2$ was discovered by Dickerson and co-workers, this dodecamer is one of the most studied DNA fragments.[19] These discoveries reveals how the genetic information passes from one to next generation.

1.1.2 Structure of Nucleic Acid

DNA is the biomolecule in which all the hereditary information of the organism is stored. The total DNA content of any cell is known as DNA genome which is unique for every single organism. The elementary building block of the nucleic acid is nucleotide which is composed of a 5-membered, non-planar, sugar ring (deoxyribose in DNA and ribose in RNA), a phosphate group and a purine or pyrimidne nucleic bases. The nucleobases can be divided into the purines: adenine (A) and guanine (G) and pyrimidines: cytosine (C) and thymine (T) in DNA and thymine (T) is replaced by uracil (U) in the case of RNA (Fig. 1.1). Nucleosides, the units formed from the sugar and nucleic base only, are linked together via a phosphate group, forming a polynucleotide chain. DNA has two complementary helical strands running in anti-parallel direction (Fig. 1.2). The nucleotides are linked together via phosphodiester linkage to constitute a strand. The two

helical strands are held together via Watson Crick hydrogen bonds in which adenine forms two hydrogen bonds with thymine and cytosine forms three hydrogen bonds guanine (Fig 1.3).[20]

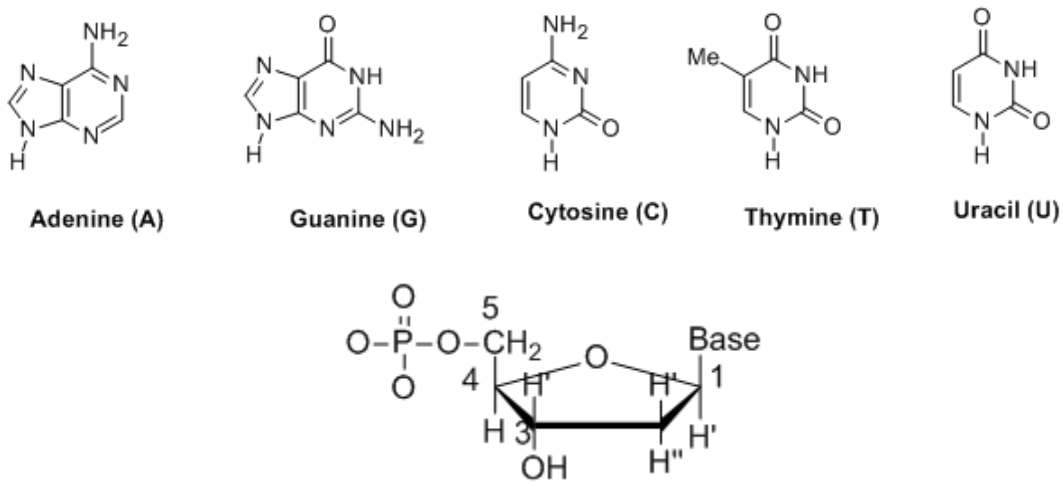


Fig. 1.1: Schematic structure of the nucleic acid bases and of deoxyribose with monophosphate.

The most common conformations of DNA are B-, A-, Z-DNA and B-form DNA is the most common occurring conformation. In this type of DNA, the base pairs are perpendicular to the helix axis and twisted by 36^0 with respect to each other. A single turn in the double helix consists of 10 base pairs. The two strands of the double helix are separated by two different grooves i.e. minor groove and major groove. The major and minor groove of DNA differs from each other not only in size, but also in polarity and chemistry (Table 1.1). The chemistries available in the grooves are specific to the base pairs which can lead to sequence specific binding in the grooves.[21,22] Specific recognition of DNA sequences by small molecules is achieved by the combination of hydrogen bond acceptor/donor sites available on the major groove or minor groove of DNA.[12, 13] DNA is a stiff molecule with a persistence length of about 50nm.[23]

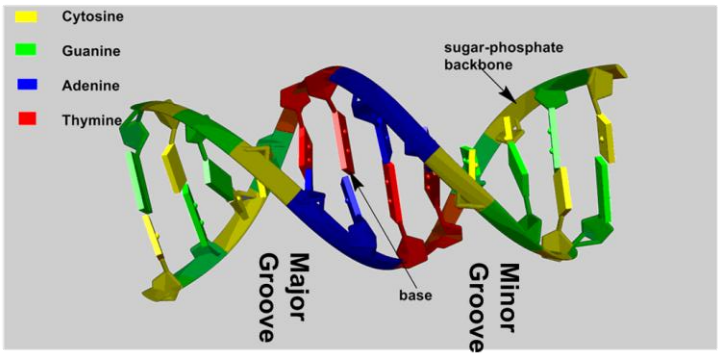


Fig. 1.2: Nucleic Acid Structure.

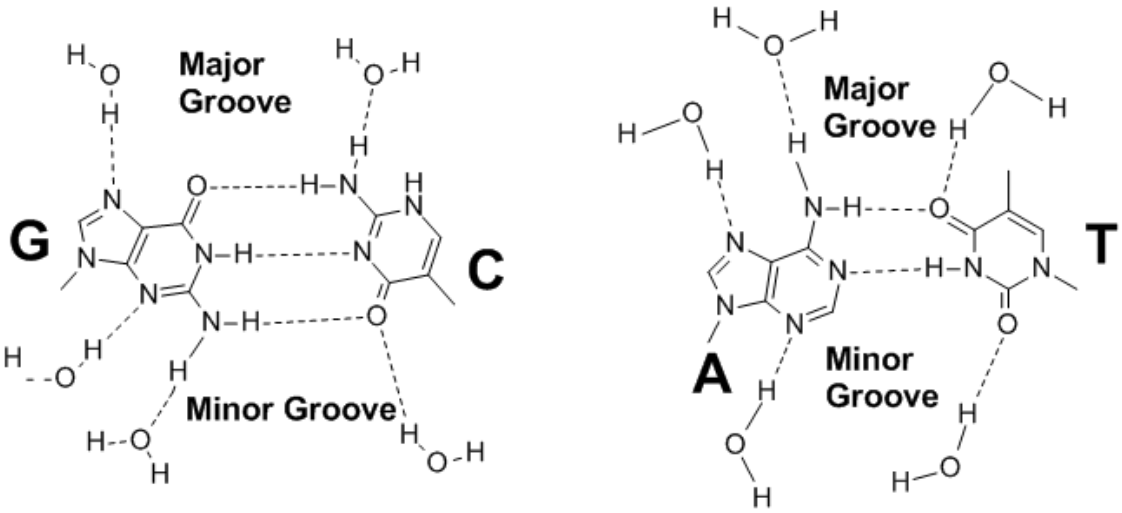


Fig. 1.3: Chemical structure of base pairs extending into major and minor grooves.

Table 1.1: Structural properties of A-, B-, Z-form of DNA.

Conformation	Helix Sense	Twist/bp(Å)	Rise/bp (Å)	Residues/turn	Sugar pucker	Groove Width (Å)		Groove Depth (Å)	
						Minor	Major	Minor	Major
A-DNA	right	32.7	2.56	11	C3'-endo	11.0	2.7	2.8	13.5
B-DNA	right	36	3.4	10	C2'-endo	5.7	11.7	7.5	8.5
Z-DNA	left	-9,-51	3.8	12	C3'-endo (Syn)		8.8	3.7	3.7

1.2 Drug-DNA Interaction

DNA is the pharmacological target of many anticancer drugs which are currently under clinical trials. Transcription and replication are the vital processes essential for the survival of the living system. In transcription, information is fetched from DNA to RNA and has recourse to synthesize protein in the body. In replication, DNA yield self-replication process and reconstruct two identical strands (Fig. 1.4). DNA starts these processes only after receiving the signal which is usually in the form of regulatory protein to a specific region of DNA. If this regulatory protein is mimicked by a drug molecule (mainly heterocyclic aromatic molecule), then the functions of DNA can be artificially modulated, inhibited or activated by this small molecule to cure or control a disease. DNA involved in vital processes such as replication, transcription etc. are of particular interest as target for wide range of anticancer and antibiotic drugs.[24-29]

DNA plays an extremely important role for drug action, for many biological activities (anti-tumor, antiviral and antimicrobial) because many compounds have ability to bind with DNA sequence specifically and interfere with DNA topoisomerases or with transcription factor. [30] The activity of some drugs for the treatment of cancer, genetic disorders, and other viral diseases is depend upon their binding mode with DNA or modification of DNA activity. The study of interaction of drug with DNA is very interesting and efficient not only to learn the mechanism of interactions but also for designing new drugs. Although the mechanism of drug- DNA interaction is not known in detail. It is necessary to explore more simple methods for investigating their interaction mechanism. The current research effort focuses on the computational methods which are used to investigate the molecular basis for drug-DNA recognition, the role of water in

mediating weak interaction. Thus the investigation of drug-DNA interaction could be helpful to discover new and advanced drug candidates.[31, 32] Some of the important Web Tools for the drug-DNA modelling are given in Table 1.2. Number of DNA-drug complexes and DNA-protein complexes deposited in Nucleic Acid Database (NDB) are given in Table 1.3.

Table 1.2: Some important Web Tools for the modelling of DNA and Drug-DNA Complexes.

Tool	**Description**	**References**
Curves	To Analyze DNA structure	[33,34]
Nuparm	Analyze sequence-dependent variations in nucleic acid (DNA and RNA) double helices	[34-40]
3DNA	Analyze, rebuild and visualize 3D nucleic acid structures	[41,42]
NucCGEN	Generate a curved or non-uniform helix	[33]
AMBER	Generate canonical A-and B-duplex geometries of nucleic acids	[43-44]
DNA sequence to structure	Generates canonical A and B DNA and molecular dynamics-averaged DNA structure	[45]

Table 1.3: Number of 3D structures of DNA, DNA-drug and DNA-protein complexes deposited in NDB [46].

Types of Complex	**Total number of NDB entries**
DNA	6393
Drug-DNA complexes	429
DNA-minor Groove Binders	116
DNA-major Groove Binders	24
DNA intercalators	139
DNA-protein Complexes	4479
DNA-covalent binding complexes	15

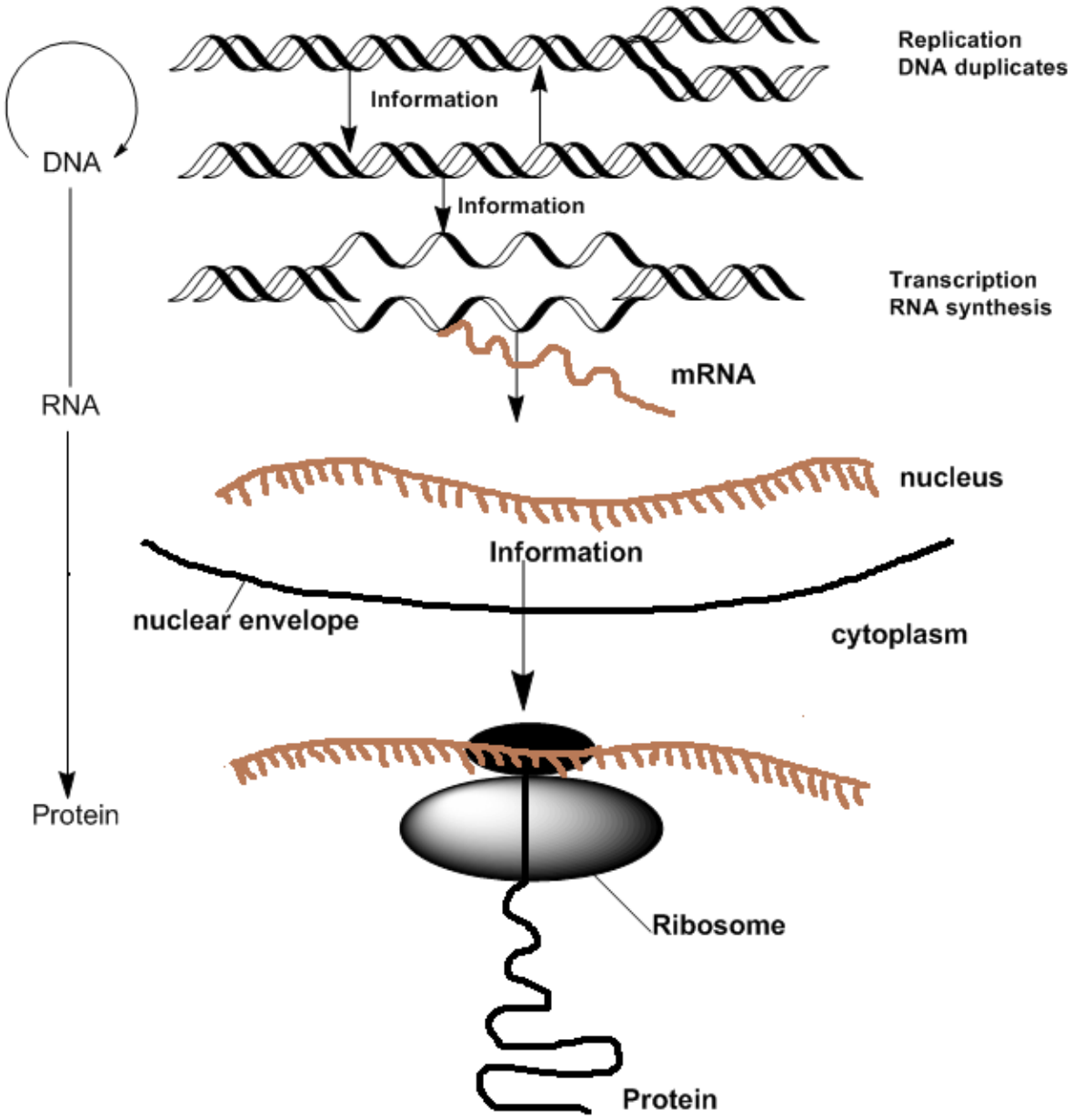


Fig. 1.4: The "Central Dogma of Molecular Biology" depicting key cellular processes of DNA replication and translation.

1.3 Different Modes of DNA binding with Drug

There are two key modes in which drug molecules can interact with DNA such as surface binding to their minor or major grooves, intercalation between adjacent base pairs. Thus both covalent as well as non-covalent interactions were found between drug molecules and DNA.[47] DNA interacting drug molecules are given in Table 1.4.

1.3.1 Covalent Binding

Many chemotherapeutic drug molecules which are in clinical use binds with DNA not only non-covalently but also by covalent binding. Covalent binding in DNA is irretrievable and regularly points to complete inhibition of DNA processes and subsequent cell death. Drug molecule covalently binds with DNA via inter- and intrastrand cross linking or alkylation. Covalent binders of DNA are the high binding strength.[48, 49] The covalent binders are also known as alkylating agents because they can attach an alkyl group to DNA and are also used in the treatment of Cancer. Alkylating agents are the important class of anticancer drugs, they play crucial role in the cure of several types of cancers. Alkylating agents have methyl or other alkyl groups (C_nH_{2n+1}) attached to the molecules. Chemical structure of some important alkylating agents is shown in Fig. 1.5. Alkylating agents are involved in reaction with the preferential N-7 position of guanine and N-3 of adenine in DNA. Thus the base pairing of the DNA could be inhibited and this leads to miscoding of DNA. Alkylating agents can interact with DNA via three mechanisms. In the first mechanism, an alkylating agent attaches alkyl group to the nucleic acid bases, this results in the DNA being fragmented by repair enzymes in their attempts to replace the alkylated bases. In second mechanism alkylating agent is responsible for the DNA damage due to formation of cross-links and

bonds between atoms in the DNA. In this process, two bases are linked together by alkylating agents that has two DNA-binding sites. Cross-linking prevents DNA from being separated for synthesis or transcription. In third type of mechanism, alkylating agents causes the mispairing of the nucleotides leading to mutations.[50] The nitrogen mustards were the first alkylating agent used medically, as well as the first modern cancer chemotherapies.[51] Cis-platin is one of the anticancer antibiotics, that covalently bind to DNA, which makes an intra/interstrand cross-link with nitrogens on the DNA bases and is used in the treatment of testicular, ovarian, head and neck cancers. Most alkylating drugs are monofunctional methylating agents (e.g. temozolomide[TMZ], N-methyl-N'-nitro-N-nitrosoguanidine [MNNG], and dacarbazine), bifunctional alkylating agents such as nitrogen mustards (e.g. chlorambucil and cyclophosphamide), or chloroethylating agents (e.g. nimustine [ACNU], carmustine [BCNU], lomustine [CCNU], and fotemustine) [52-57].

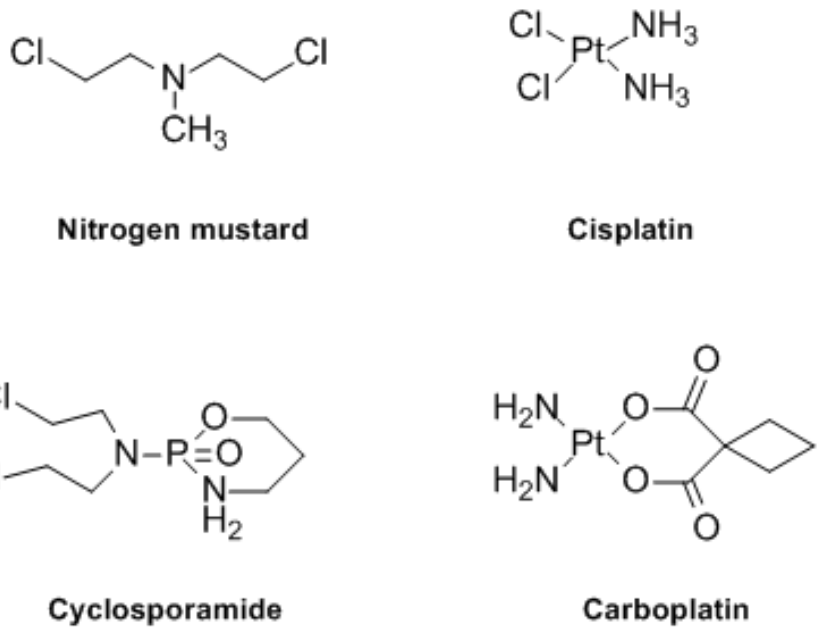


Fig. 1.5: DNA Alkylating Agents.

Table 1.4: DNA interacting drug molecules.

Non-covalent Binding drug molecules			
Groove Binding		**Intercalators**	**Covalent binding drug molecules**
Minor groove Binders	**Major Groove Binders**		
Berenil Netropsin Hoechst 33258 Distamycin A Gunayl Bisuframidine SN6999 SN7176 Pentamidine Mithramycin Pilcamycin Chromomycin A3 Diamidine-2-phenylindole Bisbenzimadoles Bleomycin Mitomycin FR66979 Duocarmycins CC-1065 Yatakemycin Neocarzinostatin Calicheamicins Retrorsine Anthramycins Saframycins Ecteinascidin 743 Isochrysohermidin	Chloroquine Netamycine Cis-{$Pt(NH_3)_2$(pyridine)}$_2$+ Aminoglycoside (NB33) Chlorambucil Nimustein Pluramycins Aflatoxins Azinomycins Leinamycin Ditercalinium	Daunomycin Nogalamycin Ethidium bromide Proflavine Ellipticine Diplamine Chlorpheniramine Bis-napthalimide Doxorubcin Aminoacridines Arylaminoalcohals Coumarines Cystodytin Diplamine YO and YOYO-1 Quinolines Quinoxalines Echnomycin Methapyrilene Tamoxifen M-AMSA Indoles Aclarubicin Idarubicin Epirubcin Pirarubcin Valrubicin Amrubicin Actinomycin D Camptothecin Topotecan Irinotecan Rebeccamycin Podophyllotoxin Etoposide Teniposide Elsamicin Dynemicin Triostin	Nitrogen mustard PBDs CC1065 Cis-platinum Menogril Clomesone Cyclodisone

1.3.2 Non-covalent Binding

Non-covalent binding drug may change DNA torsional tension, interrupt protein-DNA interactions, and potentially lead to DNA strand breaks. All of these can have substantial effects on gene expression. Non-covalent interactions, in particular hydrogen bonding and stacking interactions, determine the structure of biomolecules (such as nucleic acids and proteins). While it is understood that hydrogen bonding is essential for the specificity of base pairing, π-π stacking interactions between planar aromatic rings of nucleobases are equally important contributions to the final stability of nucleic acid structures. Although individually weak, the additive power of these interactions has large cooperative stabilizing effects. Non-covalently binding of drug with DNA is mainly classified into two category viz. groove binders and intercalators. Groove binders are of two types: minor groove binders and major groove binders. Groove binders are highly sequence-specific.[58] The two types of grooves in nucleic acid differ in hydrogen-bonding, electrostatic potential and in degree of hydration. Mainly the large protein molecules binds to the major groove of DNA while small molecules generally bind to the minor groove of DNA which are long elongnated structures with a curvature that accompaniment the shape of the minor groove.[59, 60]

1.3.2.1 Minor Groove Binders

Minor groove binders usually consist of aromatic rings covalently linked by sigma bonds. Small molecules can form hydrogen to the nucleic bases, generally N3 of adenine and O2 of thymine. Minor groove binders generally bind with A-T rich region of the DNA. This preference in addition to the designed propensity for the electro negative pockets of AT sequences is probably due to better vander Waals contacts between the ligand and groove

regions and also because of the steric hinderance in the latter, presented by the C2 amino group of the guanine base.[61-63] Sequence specific DNA-binding proteins commonly binds with the major groove because of numerous possibilities for hydrogen bonds with donors and acceptors on the nucleic bases, which provides complex stability and sequence specificity. Proteins and small molecule binding to the minor groove of DNA; depends upon the hydration properties of minor grooves, the latter binding to that AT-rich regions in which water ordering is most prevalent. Thus the minor groove binding is normally driven by the very large entropy of releasing the ordered water, despite an unfavorable enthalpy.[59, 64-66] Minor-groove binding usually involves greater binding affinity and higher sequence specificity than that of intercalator binding. Minor-groove has been demonstrated for neutral, mono-charged and multicharged ligands.[67] Generally, minor-groove binders show AT-rich region selectivity, several factors are responsible for this preference. The electrostatic potential of AT-rich region is greater than that of GC-rich region. On the other hand, the dimensions of the minor groove at AT sites are narrower and deeper than at the GC sites. This difference is due to the differences in the ionic interactions in the two types of base pairs.[68] The cationic minor-groove binders include the lexitropsins and their conjugates, analogues of Hoechst 33258, DAPI and diarylamidines, Berenil, SN series, and pentamidines.[69-79] Some important minor groove binders are shown in Fig. 1.6.

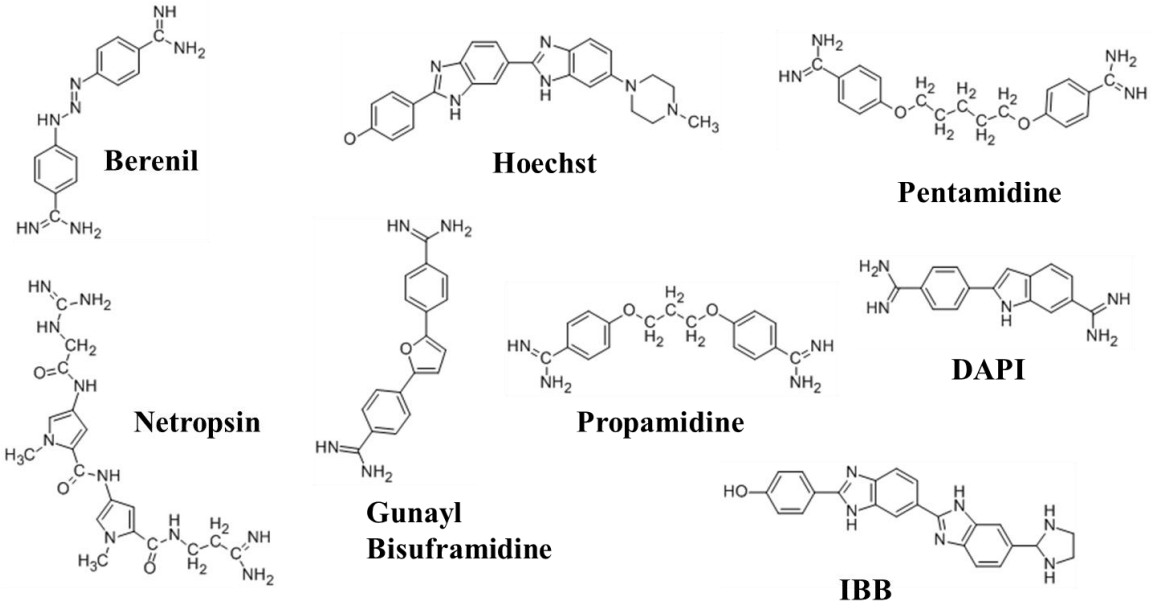


Fig 1.6: DNA Minor Groove Binders.

1.3.2.2 Intercalators

Another type of non-covalently binding drugs is intercalators which are generally planar heterocyclic molecule which stacks between the two adjacent nucleic acid base pairs. The complex remains stabilized because of π-π stacking between the drug molecule and DNA bases. Intercalation, first explained by Lerman, in which the drug molecule is held rigidly perpendicular to the DNA backbone without breaking up the hydrogen bonding between the nucleic bases. This may cause distorting the sugar phosphate backbone and also decreased the pitch.[77] The driving forces for the stability of DNA-intercalator complex are vanderWaals, hydrophobic, stacking or Charge transfer forces and hydrogen bonding and electrostatic forces also become important for the stabilization of complexes.[78] DNA intercalation results conformational changes in DNA structure, causing lengthening, stiffing and unwinding of DNA helix. Intercalation needs changes in the torsional angles of sugar-phosphate backbone to adjust the incoming aromatic compound,

which causes separation between the base pairs with a lengthening of the DNA approximately 3.4 Å and decrease in helical twist, unwinding the DNA in the vicinity of the binding site to less than 36° base pair.[79] Intercalation preferentially occurs at GC-rich sequences because these sequences get unstacked easily. Intercalators generally cause more significant distortion to the conformation of DNA. Echinomycin, noglamycin, triostin A, acridine, cis-Platin, adriamycin, ethidium, propidium, actinomycn D, adriamycin are some examples of the DNA intercalating agents shown in Fig. 1.7.[80-83]

There are few major binding modes for reversible binding of molecule to the DNA: (i) electrostatic interactions with the anionic sugar phosphate backbone of DNA (ii) interaction with DNA minor groove (iii) interaction with DNA major groove (iv) intercalation between DNA base pairs via DNA minor groove (v) intercalation between DNA base pairs via DNA major groove and (vi) threading intercalation mode. After the intercalation of a structure, the access of another intercalator to binding site next to neighboring intercalation pocket is hindered. This phenomenon is referred as the "neighbor exclusion principle" and could be explained considering that due to intercalation the significant structural changes in DNA with deep alterations in the nucleotide secondary structure.[58, 84-87] Intercalating compounds without bulky substituents can intercalate without having significant part of molecule in either minor or major groove, this type of molecules (DACA, proflavin etc.) are called classical intercalators.[79] Some of the intercalating molecules having bulky substituents, and these bulky substituents are placed in the major or minor groove along with intercalating moiety. These types of intercalators are called threading intercalators.[80] In threading

intercalation complexes, an aromatic system is inserted between the base pairs, while bulky substituent binds strongly with both major and minor groove.[84, 85] Other important intercalating drugs are the anthracyclin, daunorubicin, Adriamycin, quinacrine and actinomycin have bulky substituents that must be in one groove or the other after the planar aromatic ring of the drugs is bound by intercalation.[87-100]

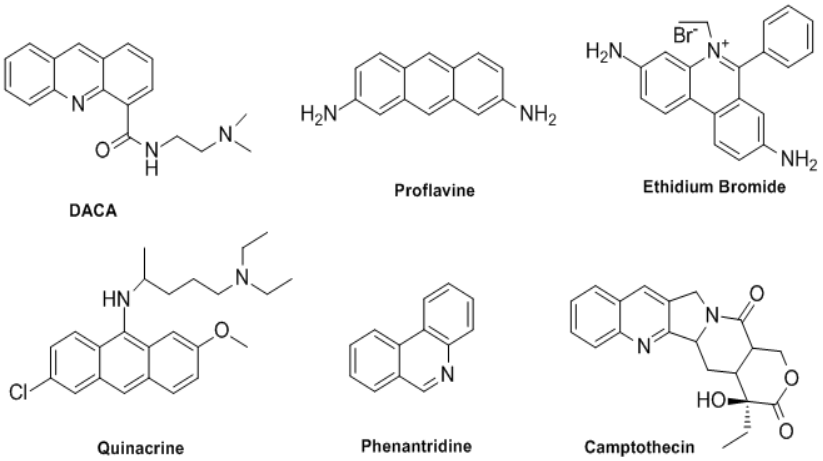


Fig. 1.7: DNA Intercalators

1.3.2.3 Major Groove Binders

The major groove is wider than the minor groove, the groove width values for B-form of DNA are 11.6 and 6.0 Å respectively. Due to this difference in dimension, the major groove is the target for many DNA-interacting proteins. Many biological macromolecules such as proteins interact by the variety of hydrogen bond acceptor and donor supplied in the major groove.[101-103] It is important for a major groove binding molecule that it could block access to proteins that recognize the same groove. This can be achieved by sequence affinity and sequence selectivity.[104] DNA duplexes which are made up of polypurine–polypyrimidine sequences can be read by oligomers and bind to the major groove and form hydrogen bond with nucleic bases of the purine strand. These are called

triplex-forming oligonucleotides (TFOs). Another form of major-groove recognition could be achieved by peptide nucleic acids (PNAs).[105, 106] Chemical structure of some important major groove binders are shown in Fig. 1.8. Fig. 1.9 shows different mode of drug-DNA binding.

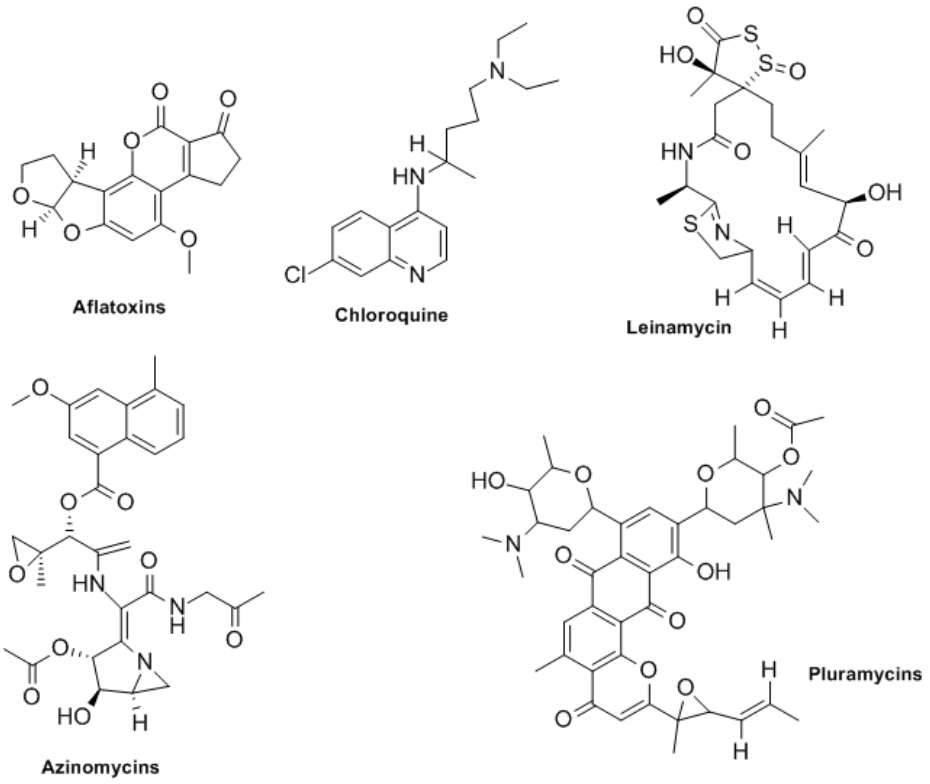


Fig. 1.8: DNA Major Groove Binders.

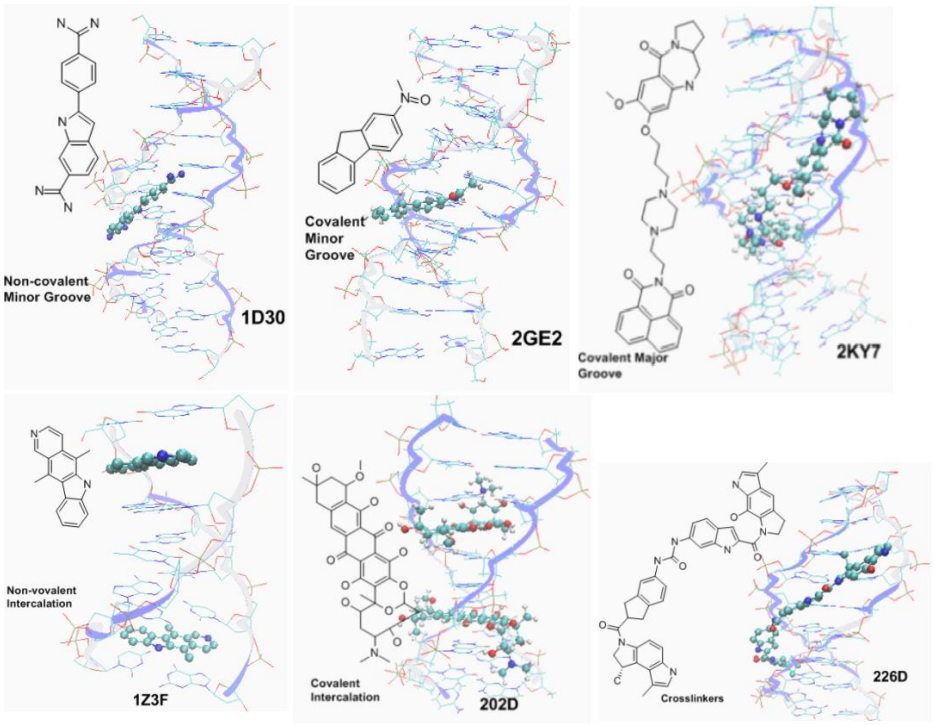


Fig. 1.9: Different modes of Drug-DNA Binding.

1.3.3 DNA interacting organometallic compounds

Many coordination complexes possess an intercalating ligand in their coordination sphere; the study of such complexes reveals the preferential geometry of the metal center, the nature of the intercalating ligand and the nature number and the position of the substituents over the intercalating ligand in the capacity and selectivity of the coordination complexes to intercalate with DNA. These coordination compounds bind to DNA via two interaction modes: irreversible (covalent or coordination binding) and reversible (intermolecular association). The later binding mode can be further classified into electrostatic interactions, groove binding and intercalation. However, these coordination complexes may exhibit a preference for a particular binding mode or a nucleotide sequence depending upon the size and the shape of the molecule.[107-110]

All mononuclear platinum complexes could form intrastrand and interstrand adducts with DNA. When interstrand lesion is formed, massive distortions of the B-DNA are observed. Similarly, intrastrand lesion, while it forms more readily than the interstrand lesion, it induces mutational events via the distortion of its nucleic acid target. Binuclear platinum (II) complexes where designed and synthesized and their interactions were studied with calf thymus DNA and a small 49 base pair oligodeoxyribonucleotide. Owing to the presence of the pyridyl ligands, this compound induces a much higher degree of DNA unwinding than that seen with the either of the ammonia bound complexes, as well as the mononuclear trans-[$PtCl_2$-$(py)_2$]. Similarly, to the mononuclear compound [(Pt(trans)-$(py)_2Cl_2$-μ-(diaminobutane)]$^{2+}$. These alterations likely involve [(Pt(trans)-$(py)_2Cl_2$-μ-(diaminobutane)]$^{2+}$ to undergo π-stacking interactions upon DNA association which in turn, disfavors the Z-DNA conformations. Importantly, interstrand-cross links formation is very efficient for all three complexes. The directionality is dependent upon the nature of the cross link. Interestingly, this is a unique example of anti-cancer drugs behaving in this manner. Molecules normally reach DNA through one of the grooves and react to either the backbone or the nucleobases. Electrostatic binding occurs due to the interaction between cations with the negatively charged phosphate backbone at the exterior surface of the DNA helix. Chemical structure of some organometallic compounds are shown in Fig 1.10.[111-112]

The use of transition metal complexes gives a strong tool to the drug chemists to develop and study molecules capable of obtaining specific DNA-drug interactions considering the multiple options of d-block metals from the periodic table. Transition metals are dynamic

in geometry, electron affinity and reactivity, making them excellent choices to feed the ongoing field of antineoplastic drug discovery.[113-118]

The fundamental factors of these interactions still possess greatest gaps in as much as the results provided by experimental designs that do not involve expensive protocols and equipment are extremely poor to identify the specific interactions due the lower energetic changes involved, making clear the use of methodologies such as computational chemistry to help solve these problems.[118-122]

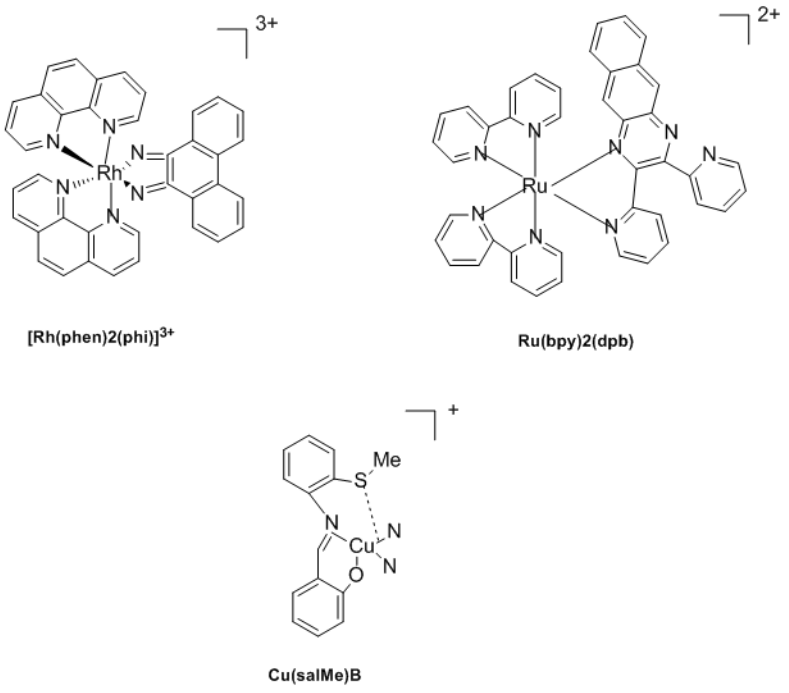


Fig. 1.10: Some organometallic compounds.

1.4 Experimental studies used in drug-DNA interaction

Experimental Studies plays crucial role to explore the drug-DNA interaction. Thermodynamic studies provide the necessary information of free energy, enthalpy, entropy, heat capacity and also about the binding constant changes during complex formation. Various experimental techniques which are used to understand the interaction of drug molecule with nucleic acid holds Infrared (IR), Raman, Circular dichorism, UV-visible, Nuclear magnetic resonance (NMR) spectroscopies, Atomic force Microscopy (AFM), Electrophoresis, Mass spectrometry, Viscosity measurements, Thermal denutration studies, Cyclic square wave and Differential pulse voltammetry, etc. The above techniques have been used as a primary tool to characterize the behavior of drug-DNA binding and the consequences of such type interaction on the structure of nucleic acid. The commonly used experimental techniques are UV-visible, fluorescence spectroscopies and cyclic voltammetry.[123] Chaires also provided the change in experimental values of thermodynamic properties during the drug-DNA complex formation.[124] This information may helpful for the theoretical prediction and structural analysis. In 2006, Freyer et al. during the studies on nucleic acid and therefore also in the binding of drug molecule with these bimolecular system, the solvent water plays important role. In the case of DNA-focused drug approaches have need to understand how water take part in the reorganization.[125-129]

Circular dichroism (CD) spectroscopy is one of the important techniques for drug-DNA interaction. This technique measures the difference in absorption of left and right circularly polarized light by optically active chiral compounds. This technique is able to detect whether a drug is bind with DNA by intercalation or minor groove. Binding site

and also other binding parameters can also be calculated by using CD. CD spectroscopy techniques are used to explore different types of non-covalent interactions between drug-DNA, these effects affect the electronic structure of the molecules and also modify their electronic spectroscopic behavior.[130]

Thermal stability is used to study the stabilizing or destabilizing effects of a drug on the DNA transition which can be calculated by monitoring the 260 nm UV-absorption profile as a function of temperature.[131]

Viscometric Titration is used to examine conformational changes in DNA as a result of drug-DNA interactions.

Electric linear dichroism predicts the information about the binding mode and also about DNA base preferences in drug binding to DNA.[131-134]

X-ray crystallography required a good quality of single crystal of the complex.[135] It is difficult to get good quality of crystal of high molecular weight compounds such as drug-DNA complex. The main disadvantage of this method is the selection of particular conformational forms or structures due to the influence of crystal packing forces and local ionic strengths.

Nuclear Magnetic Resonance (NMR) is very useful for the characterization of ligand-DNA interactions at molecular level. The atomic nuclei which are available for the study of DNA (^{1}H, ^{13}C, ^{15}N and ^{31}P), where ^{1}H is the most common, but NMR of ^{31}P is useful for studying the effects of ligand binding on the phosphate groups of DNA. DNA minor groove binding drugs generally prefer AT-rich, rather than GC-rich regions of DNA. These drugs are usually planar with crescent in shapes. Most of antitumour drugs bind to

the major groove, and they usually do it covalently through N-7 of guanine but their modes of interaction have been studied with techniques different from NMR.[135]

1.5 Molecular modelling studies involved in drug-DNA interaction

Molecular modelling techniques are used to study the 3D structure, dynamics and properties of the molecule. To better comprehend the basis of the action of any molecule with biological activity, it is necessary to know that how this molecule interact with its site of action, particularly its conformational properties and orientation of interaction. Prior to any experimental screening of different compounds uses molecular modeling techniques lessens labor and is financially savvy. Molecular modeling methods can be able to identify and define key details of molecular interaction using high quality molecular graphics tools. Computer modeling has risen as a powerful tool for experimental and theoretical investigations. Visualization of experimental data in a 3D, atomic-scale model cannot just clarify startling outcomes yet regularly brings up new issues, which influences future research. Molecular Docking is one of the important molecular modeling technique which plays key role in setting up the interaction between different molecules and aids in structure based drug discovery. Molecular Docking likewise verifies the experimental results and helps in understanding the mechanism of interaction.[136-138] Molecular modeling methods regularly utilized in deciding the binding mode of various drugs where intercalators and groove binders are easily distinguished.[136, 139-143] Some studies on sequence selectivity of minor groove binding ligands demonstrated that the most solid outcomes for AT-rich DNA sequences are required when MD are performed in explicit solvent, when this information is handled utilizing the MMPBSA approach and when normal mode analysis is utilized to

gauge configurational entropy changes.[144-147] Analogue based approaches such as QSAR, [148-150] pharmacophore modeling, and structure based approaches such as virtual screening (VS) [151-153] and molecular docking[154], free energy of binding to understand binding affinity (MMPBSA and MMGBSA), [155, 156] are the typical molecular methods used for the investigation of ligand drug receptor interaction, establishing the biological activity of the ligand or a compound. Different types of experimental and theoretical methods which are used to study the interaction between drug and DNA are shown in Fig. 1.11.

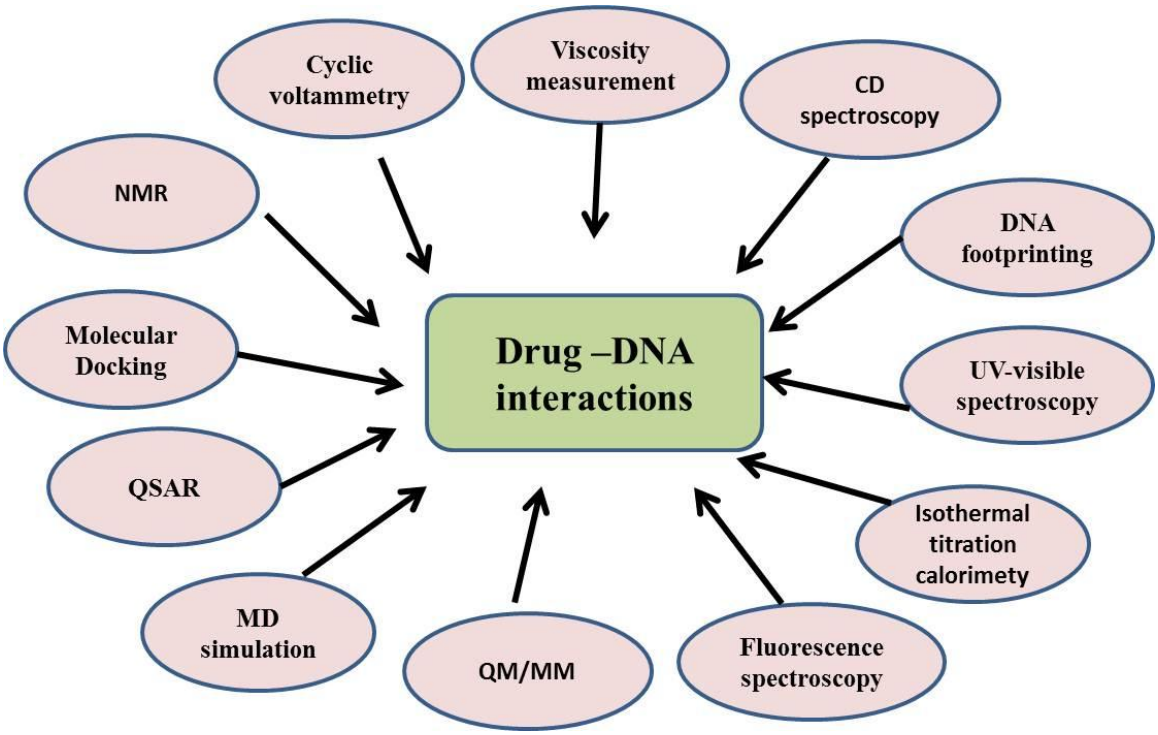


Fig. 1.11: Different types of experimental and theoretical methods used to study drug-DNA interaction.

1.6 Forces involved in Drug-DNA interaction

The binding of drug to DNA contains different energy contributions which includes energy in the form of conformational changes, from the entropic contribution for the

biomolecular complex formation, from the hydrophobic transfer process, from the formation of non-covalent molecular interactions within the complex. Some of them play a key role in stabilizing of drug in drug-DNA complex.[147, 157-160]

1.6.1 Hydrogen bonding

In drug-DNA interaction hydrogen bond formation takes place between the substituents on drug and the bases of the nucleic acid, they both are able to form hydrogen bonds to water. All the hydrogen bonds are linear and have similar free energies, so the contribution to the favourable change in free energy is very less when the drug and DNA interact in solution. In contrast to the formation of poorly aligned hydrogen bonds or absence of some of them on the complex formation carries a 4 KJ mol^{-1} penalty. Therefore hydrogen bonds are one of the most important means of making sequence specific interaction of nucleic acid with drug molecule.[147]

1.6.2 Electrostatic forces: Bridges

Salt bridges are electrostatic interactions that are formed between the ionized phosphates of nucleic acid and the positively charged group of drug molecule in drug-DNA complexes. They provide about 40 KJ mol^{-1} of stabilization per salt bridge. The strength of salt bridge decreases as the concentration of the salt in the drug-DNA complex increases. These forces are much stronger when there is no water molecule between the two ionized groups because water has high dielectric constant. They are relatively long range forces.[161]

1.6.3 Entropic forces: The hydrophobic effect

The hydrophobic effect is due to the presence of water at the surface of complex. Any molecule in water creates a sharp curved surface of ordered water molecule around itself.

When the molecules are accumulated, the ordered water molecules at the interface become the part of disordered bulk water. Thus the system can be stabilized by increasing the entropy. Water molecules which are left at the interface between drug and DNA decrease the entropy of system. Thus the surface of non-planar aromatic chromophore of drug molecule tends to be exactly complementary so that there are no necessary water remains where complex formation takes place.[162]

1.6.4 Base stacking: Dispersion forces

Base stacking is obtained due to two types of interaction mainly: the hydrophobic effect and the dispersion forces. Molecules which have no net dipole moment can attract each other by a transient dipole-induced dipole interaction. Thus the dispersion forces decreases with the inverse of the sixth power of the distance separating the two dipoles, as are exceptionally sensitive to the thermal motion of the molecules involved. In spite of their extreme distance dependence, dispersion force plays a key role in maintaining the double standard nucleic acids structure because they help to cause base stacking. They also allow aromatic ring of the drug to intercalate between bases and stabilize it by base stacking.[163-167]

1.7 Previous research work done

In 1961, Lerman proposed that the acridines and its derivatives interact with nucleic acid by intercalation between the base pairs and it was shown that the bound acridines is held rigidly perpendicular to the helix axis. It was cleared that during intercalation the hydrogen bonding remains undisturbed while the regular helical structure was destroyed due to the unwinding of the DNA at the binding site. [168] The actinomycin and its related compounds bind to DNA via intercalation between the GC base pairs but binding

at a given site creates distortion in the helix. The distortation in helix is due to the pair of hydrogen bonds formed between the deoxyribose oxygen and the –CONH- groups of the actinomycin.[169] In 1977, C.W. Mosher et al. investigates 2- Deaminoactinomycin D, which is an antibiotic drug with antitumor activity and binds with the DNA by intercalation. Experiments show that this drug binds with calf thymus DNA and the stability of complex increased with ΔT_m.[170] Chen et al. presents the theoretical investigation of binding of newly synthesized intercalating anthraquinone antitumor drug mitoxantrone to the six different tetramer sequences. The interaction energy is better when the 5'-purine is a guanine rather than an adenine and the intercalative binding of drug in the minor groove is less favorable than the major grove and also less sequence selective. The only hydrogen bonding involves O1' of S3 and S3' sugars of the intercalation site with the side chains of mitoxantrone.[171,172] The crystal structure of minor groove binding drug berenil was determined and various molecular modelling studies show that this drug has ability to bind in the minor groove of B-DNA with specificity for A-T rich sequences. The specific hydrogen bonding was obtained between the amide groups of the drug and O2 of thymine. [173] Later, in 1988 the binding of two anthracyclines Adriamycin (ADM) and daunomycin (DNM) to the calf thymus DNA with absorbance and fluorescence quenching was studied. The study shows that affinity of ADM is always higher than that of DNM and the binding constant for both the drugs exhibits a strong salt and temperature dependence. The exclusion parameter equals to 3.1±0.4 and 3.3±0.4 base pairs for the ADM and DNM respectively.[174] The molecular model was presented for the previous crystallographic studies on the d (GATACGATAC)

– proflavine complex and the significant changes were observed in conformational, roll, twist and tilt features.[175]

In 1993, Kennard investigated 30 DNA-drug complexes with deoxyolignucleotides of length between 6 and 12 base pairs. These drugs were divided into three categories: first the drug molecules which bind in the minor groove, second which both intercalate and bind to the minor groove and the last bifunctional intercalators. In this study, by comparing the different structures obtained by changing both drug and DNA molecules on complex formation, which type of forces are involved in the interactions is investigated. Interactions between drug and DNA involve hydrogen bonds between the two components and these hydrogen bonds may be direct hydrogen bonds or mediated by water molecules or by counter ions. Van der Waals forces contribute significantly to the stabilization of the complexes while there is no direct electrostatic interactions obtained.[176] Gallego et al. investigated the behavior of complexes of echinomycin with the DNA tetramers $d(ACGT)_2$ and $d(TCGA)_2$ in which the AT base pairs are either a Hoogsteen or a Watson-Crick conformation was explored by molecular dynamics simulation. The Hoogsteen conformation favours the DNA-echnomycin complex, and the theoretically calculated interaction energy is in good agreement with the experimental results. On the other hand the unfavourable dipolar interactions in the $d(TCGA)_2$ complex.[177] In 1994, Singh et al. reported the binding affinities of distamycin and its derivatives to d(CGCAAGTTGGC).d(GCCAACTTGCG) calculated using molecular dynamics studies are in good agreement with the experimental data.[178] Crenshaw et al. predicted the binding energies of acridines antitumor agents with DNA using experimental techniques.[179] Cholody et al. synthesized a new class of antitumor

agents known as bismidazoacridones. Experimental and modelling data indicates that these molecules interact with nucleic acid, but inspite of their structure, they may not be bis-intercalators.[180] In 1995, Berger et al. reported the single crystal X-ray diffraction studies of the halogenated anticancer drug iododoxorubicin with the hexanucleotide duplex sequences d(TGTACA) and d(CGATCG). The iodine substituent does not alter the geometry of intercalation as compared to previously solved anthracyline complexes. [181] In 1998, Shahla et al. performed the spectroscopic and molecular modelling studies on bis(argginyl) conjugate of a tricationic porphyrin (BAP) complexes with a series of dodecanucleotides having common $d(CG)_2$ intercalation site. Spectroscopic studies shows that the binding of BAP to the major groove of DNA.[182]

In 1999, Clare et al. used the fluorescence titration measurements to understand the interaction of analogues of bis-benzimidazole with the decamer duplex $d(GCAAATTTGC)_2$. The experimental values of binding constant, ΔG values and ΔTm values of four analogues were presented and these analogues give the same pattern of hydrogen bonding with the floor of the minor groove.[183,184] A new class of DNA intercalators of pyridazino[1',6':1,2]pyrido[4,3-b]indol-5-inium family were investigated using molecular modelling studies and these studies show a preferred orientation of the intercalating chromophore within a CG intercalation site and this helps in the rational design of novel bis-intercalators based on these chromophores.[185] Mazur et al. reported that furan derivatives interact to the minor groove of DNA.[186] Benzidine has a biphenyl ring, experimental and molecular studies clearly indicate that benzidine weakly interacts with DNA duplex as an intercalator.[187] Two tris-benzimidazole derivatives are designed and synthesized and these molecules bind strongly to the DNA sequences

where the continuity of AT stretch is not interrupted by a GC pair.[188, 189] The molecular modelling studies on benzo[f]azino[2,1-a]phthalazinium cations suggested the preferred orientation for the intercalating chromophores within the CG or TG interclalation site.[190]

In 2006, Rajendran et al. presented the dual binding behavior of 9-N,N-dimethylaniline decahydroacridinedione (DMAADD) with calf thymus DNA by means of traditional experimental and theoretical techniques. DMAADD has been found to be partial intercalator and unusual decrease in melting temperature of CT-DNA by the addition of dye molecule. Molecular modelling gives the various binding mode, the dye prefers the major groove binding to the sites GC-rich residues and to the sites AT-rich it prefers intercalation binding mode either through major or minor groove with the inclusion of the N, N-dimethylaniline (DMA) group inside the double helix which has been stacked in between the bases.[191] Kamal et al. synthesized 10-substituted 2-(4-piperidyl/phenyl)-5,5-dioxo[1,2,4]triazolo[1,5-b][1,2,4]benzothiadiazine derivatives, C2-fluoro substituted pyrrolo[2,1-c][1,4]benzodiazepines, bis-1,2,3-triazolo-bridged unsymmetrical pyrrolobenzodiazepine trimmers and performed experimental and molecular modelling studies with calf thymus DNA and molecular modelling studies confirmed the same binding mode and binding energies as predicted by experimental studies.[192-195] Biophysical studies on a family of asymmetric guanidinium-based diaromatic derivatives have been carried out to evaluate binding constants, stoichiometry and mode of binding and the results shows strong and favourable interaction.[196,197] Pandya et al. performed experimental and computational studies on indole derivatives and these studies suggested that these molecules bind in the minor groove of DNA.[198]

Later in 2011, Srivastava et al. performed a comparative study on DNA minor groove binders using available computational docking tools and the study shows that the GOLD and GLIDE docking programs are more reliable for the molecular modelling of nucleic acid. This study also reveals that the interaction energies calculated from the MMPBSA method are in good agreement with the experimental ΔT_m and the GOLD docking scores.[150] Yan et al. showed that the interaction of amino phosphine ester derivatives with DNA by experimental and molecular modelling methods. The results obtained from the studies confirms that these derivatives bind as intercalator that can slide into the G–C rich region of CT-DNA.[199] A review study on heterocyclic diamidine minor groove binders shows that a new class of DNA minor groove binders can be developed which may incorporate a bound water molecule into their DNA minor groove complex, form a very strong ternary complex, and the compounds can form cooperative stacked dimers to recognize GC and mixed AT/GC base pair sequences.[200] In 2012, Wilhelm et al. analyzed the intercalation pathway for daunomycin using two different processes: metadynamics and umbrella sampling. Daunomycin binds as minor groove binders initially, after an activated step drug is rotated which creates deformation in DNA, creates a gap between DNA base pairs, bends DNA towards major groove of DNA and forms a metastable state. Lastly, after crossing a small free energy barrier leads to further rotation in drug thereafter daunomycin fully intercalate between the DNA flanking base pairs [201] shown in Fig. 1.12.

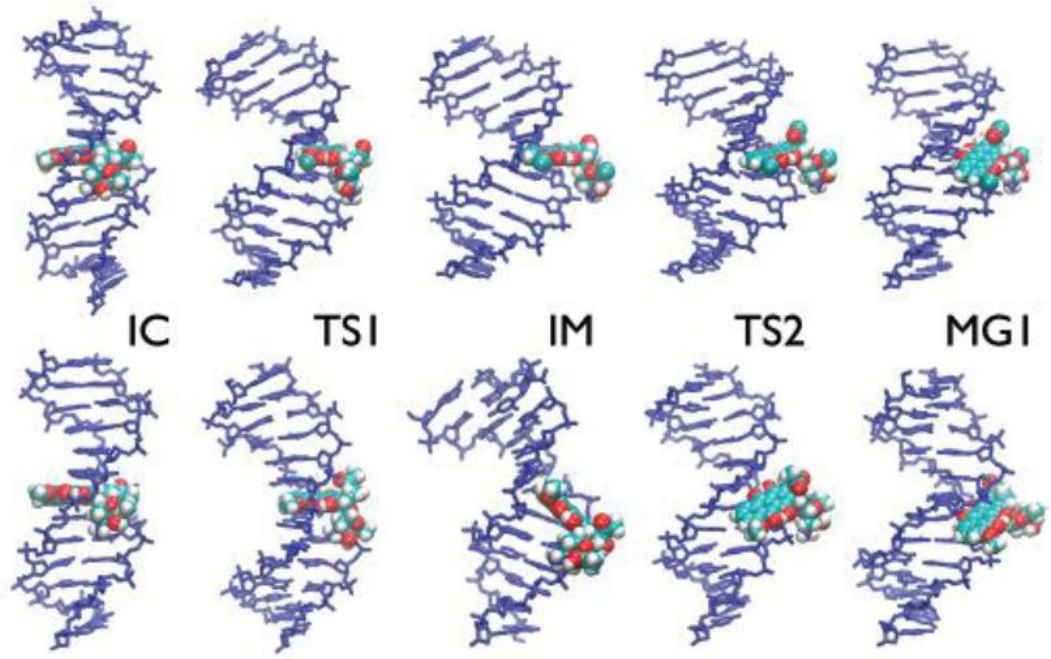


Fig. 1.12: Representative structures leading from the intercalation site to the minor-groove-bound site for the metadynamics (above) and umbrella-sampling (below) pathways. DNA is shown as a blue Dreiding model and daunomycin as a CPK model with standard chemical coloring.[201]

Vijayalakshmi et al. performed a computational study of 130 minor groove binders with Staphylococcus DNA and uses QM/MM force field for docking which predict perfect interactions with the predicting binding sites and MD simulations supports these results.[202]

Srivastava et al. calculated Molecular mechanics with Generalized Born surface area (MMGBSA) based binding energies (BEs) obtained from the MD simulation of 30 aromatic furan derivatives and performed DFT based QSAR studies to calculate the BEs without performing MD simulations and MMGBSA calculations. These studies shows good agreement to the experimental ones with statistically significant values (R^2 = 0.93, R^2_{cv} = 0.83).[155]

In 2013, Gilad et al. analyzed ligand binding sites analysis of 63 DNA-intercalator complexes available in the PDB. Molecular docking calculation was performed on these complexes with preformed intercalation site as in PDB. Agreeable outcomes were obtained while docking ligands into DNA structures with *apriori* known and unknown binding sites. Thus, AUTODOCK could produce a conformation with a RMSD < 2.00 Å in 84% of the cases with a known binding site and in 78% of the cases with an unknown binding site. This study provides potential use of AUTODOCK in the computational investigation of DNA interacting molecules.[203] Synder et al. examined the interaction of 1350 launched drugs with 10 different sequences using two molecular methods: Autodock and Surflex. These drugs were also docked with the crystallographic ATP-binding site of human topoisomerase II. These underlying investigations give impressive proof that DNA intercalation might be an important, to a great extent overlooked, source of drug-induced genotoxicity and further suggest involvement of topoisomerase in that genotoxicity.[204]

In 2014, Barbara et al., studied the dynamics of minor groove binding drug and intercalating drug with the DNA dodecamer by MD and analyzed that the minor groove binders locks the breathing movement of the minor groove instigating a lessening of the configurational entropy of the helix while intercalator enhance flexibility of the double helix.[205] Fu et al. performed experimental and molecular modelling calculations on water soluble DNA minor groove binders (MGBs), the experimental studies such as Spectroscopic titration, viscosity and electrophoresis measurements revealed that these complexes bind with DNA via an outside groove binding mode and the modeling studies

used to investigate the mechanism and the mode of action of these molecules which verifies the experimental results.[206]

In 2015, Mariya et al. checked the accuracy of DNA docking methods by two docking programs Flex and Autodock and this study revealed that the accuracy of DNA docking is directly linked with the choice of the appropriate DNA sequence. If the nature of ligand is known then the DNA template is selected accordingly but the problem arises when the nature of ligand is not known with DNA. Then the molecular docking method is unable to predict the binding mode accurately, as a result other molecular modelling methods such as MD and thermodynamics integration is used to further resolve such problems.[207] Naphthalene diimide derivatives (NDIs) have been explored as potential anti-cancer agents against pancreatic cancer and the selectivity of NDIs towards nucleic acids has been broadly contemplated.[208-211] Interactions of two NDIs derivatives one is cyclic (cNDI 1) and the other is non-Cyclic (cNDI 2) with DNA duplex is presented and it is observed that the mode of binding cyclic compound is bis-intercalator while non-cyclic compound bind as intercalator. Both ligands exhibited appreciable binding affinity to tetraplex DNA structures.[212] Thermodynamics and kinetic studies of derivatives of NDIs was performed with the calf thymus DNA. UV–Vis analysis showed high binding affinity of cNDI 1 to dsDNA in the range of 6.0×10^{5} - 5.3×10^{6} M^{-1} (approximately 10 times higher than that of cNDI 2), with bis-intercalation of four base pairs per ligand molecule. Thermodynamic studies show that entropy-dependent hydrophobic interactions play a major role in their interaction with dsDNA. cNDI 1 showed more entropically favorable interactions with dsDNA than cNDI 2.[213]

Benzimidazoles, are the important biological molecules which are used as anticancer, antiviral, antimicrobial and antifungal contains a nitrogen in its aromatic ring.[214-216] 2-anthryl substituted benzimidazole derivatives were synthesized with hydrogen, carboxyl and benzoyl substituents at the 5^{th} position and the DNA cleavage activities were performed. The variation of DNA cleavage activity was obtained and the carboxyl substituted derivatives show greater DNA cleavage than benzoyl substituted derivative while hydrogen substituted derivative did not show any DNA cleavage.[217-220] Interaction studies on triazine-benzimidazole with calf thymus DNA were explored using spectroscopic and viscometric techniques and these studies revealed that these compounds bind to the minor groove of DNA.[216]

Hoechst 33258 is positively charged minor groove binder which strongly interacts to the electronegative floor of minor groove of AT-rich region which is more negative than GC-rich region.[222-224] In 2013, Fornander et al. examine the binding mechanism of Hoechst 33258 to the three oligonucleotide duplexes containing AT regions of different lengths: $[d(CGCGAATTCGCG)]_2$ (A2T2), $[d(CGCAAATTTGCG]_2$ (A3T3), and $[d(CGAAAATTTTCG)]_2$ (A4T4) and observed that this molecule binds in the middle of the AT-sequence with high binding affinity. When the ratio of Hoechst 33258 to oligonucleotide is increased then the second molecule enters to the minor groove A4T4 and the first one slightly moves away to give space and these two molecules are now both sitting symmetrically in the groove with a small distance apart.[225]

Table 1.5. Experimental data of DNA binding ligands collected from literature with their binding mode.

Ligand	**DNA Sequence**	**Exp. Data**	**Binding Mode**	**Ref.**
DB244	CT-DNA	ΔG= -9.9 kcal/mole	m	[226]
DB75	CT-DNA	ΔG= -9.0 kcal/mole	m	[226]
DB226	CT-DNA	ΔG= -8.5 kcal/mole	m	[226]
Benzidine	CT-DNA	K= $6.2\pm0.4\times10^{3}$ M^{-1}	I	[187, 227]
HT32258	A3T3 CT-DNA	ΔG= -7.7 kcal/mole ΔG= -7.9 kcal/mole	m	[140]
Netropsin	A3T3 CT-DNA	ΔG= -7.7 kcal/mole ΔG= -8.8 kcal/mole	m	[140]
Propamidine	A3T3 CT-DNA	ΔG= -7.0 kcal/mole ΔG= -8.2 kcal/mole	m	[140]
Berenil	A3T3 CT-DNA	ΔG= -8.0 kcal/mole ΔG= -8.6 kcal/mole	m	[140]
Distamycin	A3T3	ΔG= -10.5 kcal/mole	m	[140]
Captopril	CT-DNA	K= $6.2\pm0.4\times10^{3}$ M^{-1}	m	[228]
7-amino ACTD	(AT)2 (AT)3 (AT)4 (AA)(AT)2 (AA)2 (AA)3 (AA)4	ΔG= -7.47 kcal/mole ΔG= -7.24 kcal/mole ΔG= -7.19 kcal/mole ΔG= -7.05 kcal/mole ΔG= -6.78 kcal/mole ΔG= -6.70 kcal/mole ΔG= -6.68 kcal/mole	I	[229]
HT32258	Poly{d(AT)}2 A3T3	ΔG= -11.8 kcal/mole ΔG= -11.7 kcal/mole	m	[230]
DB293	CT-DNA	ΔG= -9.6 kcal/mole	m	[231]
Doxorubicin	CT-DNA	ΔG= -8.9 kcal/mole	I	[232]
Propidium	CT-DNA	ΔG= -7.5 kcal/mole	I	[232]
Daunorubicin	CT-DNA	ΔG= -7.9 kcal/mole	I	[232]
Ethidium	CT-DNA	ΔG= -6.7 kcal/mole	I	[232]
Chartreusim	CT-DNA	ΔG= -7.4 kcal/mole	I	[233]
NB506	CT-DNA	ΔG= -5.9kcal/mole	I	[234]
WP631	CT-DNA	ΔG= -15.3 kcal/mole	B	[235]
WP762	CT-DNA	ΔG= -16.3 kcal/mole	B	[236]

table 1.5 contd…

Echinomycin	CT-DNA	ΔG= -7.5 kcal/mole	B	[237]
DAPI		$K= 5.5\times10^{8}\ M^{-1}$	G	[238]
Sperimine		$K= 2.1\times10^{5}\ M^{-1}$	G	[239]
Proflavin		$K= 2.5\times10^{6}\ M^{-1}$	I	[240]
DAPI	AATT	$K= 1.2\times10^{5}\ M^{-1}$	I	[241]
Ditercalinium	CGCG	$K= 5.0\times10^{8}\ M^{-1}$	B	[242]
Anthracene-9-carbonyl-N1-Sperimine	Poly[d(GC)d(GC)] Poly[d(AT)d(AT)]	$K= 2.2\times10^{7}\ M^{-1}$ $K= 6.6\times10^{5}\ M^{-1}$	G	[243]
2-Imidazolidinethione	CT-DNA	$K= 1.4\times10^{3}\ M^{-1}$	m	[244]
Mesalamine	CT-DNA	ΔG= -4.2 kcal/mole	I	[245]
Doxorubicin	1R2L DNA	$K= 0.17\times10^{5}\ M^{-1}$	I	[246]
Epirubicin	1R2L DNA	$K= 0.14\times10^{5}\ M^{-1}$	G	[246]
Daunorubicin	1R2L DNA	$K= 1.80\times10^{5}\ M^{-1}$	I	[246]
Cisplatin	1R2L DNA	$K= 0.41\times10^{5}\ M^{-1}$	I	[246]
Flourouracil	1R2L DNA	$K= 8.7\times10^{5}\ M^{-1}$	G	[246]
Carboplatin	1R2L DNA	$K= 2.41\times10^{5}\ M^{-1}$		[246]
Etopside	1R2L DNA	$K= 0.14\times10^{3}\ M^{-1}$	G	[246]
Cyclophosphamide	1R2L DNA	$K= 4.21\times10^{5}\ M^{-1}$		[246]
Dactinomycin	1R2L DNA	$K= 3.61\times10^{5}M^{-1}$	I	[246]
Mitoxantrone	1R2L DNA	$K= 61.91\times10^{5}M^{-1}$	I	[246]
Chloridazon	CT-DNA	ΔG= -22.63 kcal/mole	I	[247]
DB921	AATT	$K= 2.0\times10^{8}M^{-1}$	G	[200]
DB911	AATT	$K= 2.0\times10^{7}M^{-1}$	G	[200]
Pirenzepine	CT-DNA	$K= 1.57\times10^{3}M^{-1}$	G	[248]
Distamycin	Dickerson Poly[d(AT)]	ΔG= -44.77 kj/mole ΔG= -36.77 kj/mole	m	[249]
NAX001	Dickerson Poly[d(AT)]	ΔG= -35.70 kj/mole ΔG= -31.25 kj/mole	G	[249]
NAX002	Dickerson Poly[d(AT)]	ΔG= -35.70 kj/mole ΔG= -31.25 kj/mole	G	[249]
3'-azido-3-deamino daunoxrubicin	CT-DNA	ΔG= -31.25 kj/mole	I	[250]
Chlorambucil	CT-DNA	$K= 2.0\times10^{7}M^{-1}$	M	[251]
Neodymium-Naproxen (Nd-NAP)	CT-DNA	$K= 2.90\times10^{4}M^{-1}$	G	[252]
Emodine	CT-DNA	$K= 5.59\times10^{3}M^{-1}$	I	[253]

table 1.5 contd…

Mitoxantrone	CT-DNA	$K= 3.38\times10^{5}M^{-1}$	I	[254]
Ciprofloxacine	Human teomeric DNA	$K= 9.62\times10^{4}M^{-1}$	G	[255]
Daphnetin	CT-DNA	$\Delta G= -25.19$ kj/mol	I	[256]
6-Mercaptopurine	CT-DNA	$K= 7.48\times10^{3}M^{-1}$	G	[257]
Thiabendazole	CT-DNA	$\Delta G= -21.68$ kj/mol	I	[258]
PyTA		$K= 4.75\times10^{4}M^{-1}$	G	[201]
PzTA		$K= 6.04\times10^{3}M^{-1}$	G	[201]
Aryltetrasachcharide	h-TCCT h-ACAT	$\Delta G= -7.6$ kcal/mol $\Delta G= -6.7$ kcal/mol	I	[259]
3,4-Dihydropyrano[c] chromene	CT-DNA	$K= 2.37\pm0.001\times10^{3}M^{-1}$	G	[260]
Troxerutin	CT-DNA	$K= 1.75\times10^{3}M^{-1}$	m	[261]
Pregabalin	CT-DNA	$K= 5.6\times10^{3}M^{-1}$	I	[262]
Nimustein	CT-DNA	$K= 3.27\times10^{3}M^{-1}$	I	[263]
4-chloro-3-ferrocenylaniline	CT-DNA	$K= 2.39\times10^{3}M^{-1}$	G	[264]
Querceitn	Ds-DNA	$K= 3.56\times10^{3}M^{-1}$	I	[265]
Flavones	CT-DNA	$\Delta G= -29.9\pm1.2$ kj/mol	I	[266]
3-Hydroxyflavone	CT-DNA	$\Delta G= -26.9\pm0.9$ kj/mol	I	[267]
5-Hydroxyflavone	CT-DNA	$\Delta G= -27.5\pm1.0$ kj/mol	I	[268]
6-Hydroxyflavone	CT-DNA	$\Delta G= -28.3\pm0.6$ kj/mol	I	[268]
7-Hydroxyflavone	CT-DNA	$\Delta G= -28.0\pm0.1$ kj/mol	I	[268]
Tomoxifen	CT-DNA	$K= 8.15\times10^{4}M^{-1}$	I	[269]
Nitrofurantion	CT-DNA	$K= 8.22\pm0.05\times10^{6}M^{-1}$	I	[270]
Farrerol	CT-DNA	$\Delta G= -26.16\pm0.81$ kj/mol	I	[271]
Levetiracetam	CT-DNA	$K= 4.9\pm0.2\times10^{3}M^{-1}$	m	[272]
Idarubicin	ds-DNA	$K= 2.1\times10^{4}M^{-1}$	I	[273]
Eu(III)(9-ACA)2(NO3)2	ds-DNA	$K= 9.1\times10^{4}M^{-1}$	I	[274]
Ticlopidine	CT-DNA	$K= 4.91\times10^{3}M^{-1}$	G	[275]
Citral	CT-DNA	$K= 4.96\times10^{3}M^{-1}$	I	[276]
Chelerythrine	Poly(dG).poly(dC) Poly(dG–dC).poly(dG–dC) Poly(dA).poly(dT) Poly(dA–dT).poly(dA–dT)	$K= 5.25\times10^{-6}M^{-1}$ $K= 2.17\times10^{-6}M^{-1}$ $K= 1.23\times10^{-6}M^{-1}$ K= 1.60×10-6M-1	G	[277]
6 Mercaptopurine	CT-DNA	$K= 7.48\times10^{3}M^{-1}$	G	[278-282]
Resistomycin	CT-DNA	$K= 1.095\times10^{4}\ M^{-1}$	G/I	[283]

table 1.5 contd…

Epirubicin	Fish sperm DNA	$K= 3.8\times10^{5}\ M^{-1}$	I	[284]
Sodium Benzoate	CT-DNA	$\Delta G= -28.34\pm0.40$ kJ/mole	I	[285]
Carboplatin	CT-DNA			[286]
Sinafloxacin	CT-DNA	$K= 2.33\times10^{2}\ M^{-1}$	G	[287]
Idarubicin	Ds-DNA	$K= 5.14\times10^{5}\ M^{-1}$	I	[288]
Elsculetin	CT-DNA	$K= 1.84\times10^{4}\ M^{-1}$	m	[289]
Methotrexate	Ds-DNA	$K= 1.0\times10^{3}\ M^{-1}$	G	[290]
Olanzapine	CT-DNA	$K= 2.0\times10^{3}\ M^{-1}$	m	[291]
N-phenylbenzohydroxamic acid	CT-DNA	$K= 2.01\times10^{4}\ M^{-1}$	I	[292]
Carbaryl	CT-DNA	$K= 8.68\times10^{3}\ M^{-1}$	I	[293]
Tau-Fluvalinate	CT-DNA	$\Delta G= -24.56$ kJ/mole	m	[294]
Flumethrin	CT-DNA	$\Delta G= -24.80$ kJ/mole	m	[294]
Methyldopa	CT-DNA	$\Delta G= -23.57$ kJ/mole	m	[295]
Coumarin	CT-DNA	$K= 1.69\times10^{11}\ M^{-1}$	m	[296]
Aspirin	CT-DNA	$\Delta T_m=6.4^{0}C$	I	[297]
Diflunisal	CT-DNA	$\Delta T_m=2.3^{0}C$	G	[298]
Ibuprofen	CT-DNA	$\Delta G= -6.96$ kcal/mole	I	[299]
Eugenol	Salmon sperm DNA	$K= 7.28\times10^{3}\ M^{-1}$	I	[300]

Intercalator **(I)**

Bisintercalator **(B)**

Groove Binder **(G)**

Major Groove Binder **(M)**

Minor Groove Binder **(m)**

References

[1] Hemmert, C., Pitie, M., Renz, M., Gornitzka, H., Soulet, S., Meunier, B., *J. Bio. Inorg. Chem.*, **6**, 14 (2001).

[2] Li, V.S., Choi, D., Wang, Z., Jimenez, L.S., Tang, M., Kohn, H., *J. Am. Chem. Soc.*, **118**, 2326 (1996).

[3] Zuber, G., Quada, J.C., Hecht, S.M., *J. Am. Chem. Soc.*, **120**, 9368 (1998).

[4] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N, Bourne, P.E., *Nucleic Acid Res.*, **28**, 235 (2000).

[5] Beveridge, D.L., McConnell, K.J., *Curr. Opin. Struct. Biol.*, **10**, 182 (2000).

[6] Cheatham, T.E. III, Kollman P.A., *J. Mol. Biol.*, **259**, 434 (1996).

[7] Reyes, C.M., Kollman, P.A., *J. Mol. Biol.*, **297**, 1145 (2000).

[8] Rovira, C., *Methods Mol. Biol.*, **305**, 517 (2005).

[9] Kamp, M.W., Mulholland, A.J., *Biochem.*, **52**, 2708 (2013).

[10] Senn, H.M., Theil, W., *Angew. Chem. Int. Ed.*, **48**, 1198 (2009).

[11] Duarte, F., Amrein, B.A., Blaha-Nelson, D., Kamerlin, S.C.L., *Biochimica Acta*, **1850**, 954 (2015).

[12] Patricia, S., Deiman, R., Qiang, C., *J. Phys. Chem.*, **123**, 014905 (2005).

[13] Avery, O.T., Maclend, C., McCarty, M., *J. Exp. Med.*, **79**,137 (1994).

[14] Chargaff, E., *J. Cell. Comp. Physiol.*, **38**, 41 (1951).

[15] Chargaff, E., *Experimentia*, **6**, 201 (1950).

[16] Watson, J.D., Crick, F.H.C., *Nature*, **171**, 737 (1953).

[17] Watson, J.D., Crick, F.H.C., *Nature*, **171**, 964 (1953).

[18] Wang, A.H.J et al., *Nature*, **282**, 680 (1979).

[19] Wing, R., Drew, H., Takano, T., Broka, C., Tanka, S., Itakura, K., Dickeson, R.E., *Nature*, **287**, 755 (1980).

[20] Sinden, R.R., *DNA Structure and Function*, Academic Press, San Diego (1994).

[21] Gottesfeld, J.M., Neely, L., Trauger, J.W., Baird, E.E., Dervan, P.B., *Nature*, **387**, 202 (1997).

[22] W.D. Wilson, *Reversible interactions of small molecules with nucleic acids, in: M. Gait, M. Blackburn (Eds.), The Chemistry and Biology of Nucleic Acids*, Oxford University Press, Oxford, England (1990).

[23] Cluzel, P., Lebrun, A., Heller, C., Lavery, R., Viovy, J.-L., Chatenay, D., Caron, F., *Science*, **271**, 792 (1996).

[24] Chaires, J.B., *Curr. Opin. Struct. Biol.*, **9**, 314 (1998).

[25] Chaires, J.B, *Biopolymers*, **44**, 201 (1997).

[26] Chaires, J.B., *Curr. Opin. Struct. Biol.*, **8**, 314 (1998).

[27] Chaires, J.B, *Biopolymers*, **44**, 201 (1998).

[28] Charies, J.B, *Annu. Rev. Biophys.*, **37**, 135 (2008).

[29] Hurley, L.H., *Nat. Rev. Cancer*, **2**, 188 (2002).

[30] Singh M.P., Joseph, T., Kumar S., Bathini, Y., Lown, J.W., *Chem. Res. Toxicol.*, **5**, 597 (1992).

[31] Pullman, B., *Molecular basis of specificity in nucleic acid-drug interactions, in: The Jerusalem Symposia on Quantum Chemistry and Biochemistry*, Kluwer Academic Publishers, **23** (1990).

[32] Sharma, S., Doherty, K.M., Brosh Jr. R.M., *Curr. Med. Chem. Anticancer Agents*, **5,** 183 (2005).

[33] Lavery, R., Sklenar, H., *J. Biomol. Struct. Dyn.,* **6**, 63 (1988).

[34] Lavery, R., Sklenar, H., *J. Biomol. Struct. Dyn.,* **6**, 655 (1989).

[35] Bhattacharya, D., Bansal, M., *J. Biomol. Struct. Dyn.,* **6**, 93 (1988).

[36] Bhattacharya, D., Bansal, M., *J. Biomol. Struct. Dyn.,* **6**, 635 (1989).

[37] Bhattacharya, D., Bansal, M., *J. Biomol. Struct. Dyn.,* **8**, 539 (1990).

[38] Bhattacharya, D., Bansal, M., *J. Biomol. Struct. Dyn.,* **10**, 213 (1992).

[39] Bansal, M., Bhattacharya, D., Ravi B., *Comput. Appl. Biosci.*, **11**, 281 (1995).

[40] Mukherjee, S., Bansal, M., Bhattacharyya, D., *J. Comp. Aided Mol. Des.,* **20**, 629 (2006).

[41] Lu, X. J., Olson, W. K., *Nucl. Acids Res.,* **31**, 5108 (2003).

[42] Lu, X. J. Olson, W. K., *Nat. Protoc*., **3**, 1213 (2008).

[43] Case, D. A., Cheatham, III, T.E., Darden, T., Gohlke, H., Luo, R., Merz, Jr., K.M., Onufriev, A., Simmerling, C., Wang, B., Woods, R., *J. Computat. Chem.,* **26**, 1668 (2005).

[44] Ponder, J. W., Case, D. A., Adv. Prot. Chem., **66**, 27 (2003).

[45] Arnott, S., Campbell-Smith, P.J., Chandrasekaran, R., In Fasman, Ed. *Handbook of Biochemistry and Molecular Biology, 3rd ed. Nucleic Acids*, G.P.. Cleveland: CRC Press, 2, 411(1976).

[46] Buvaneswari, C.N., Westbrook, J., Ghosh, S., Anton, I.P., Blake, S., Craig, L. Z., Neocles, B.L., Berman, H.M., *Nucleic Acid Res.*, **42**, D114 (2014).

[47] Pindur, U., Jansen, M., Lemster, T., *Curr. Med. Chem.,* **12**, 2805 (2005).

[48] Paul, A., Bhattacharya, S., *Curr. Sci.*, **102**, 212 (2002).

[49] Liu, H., Sadler, P.J., *Acc. Chem. Res.*, **44**, 349 (2011).

[50] Silvestri, C., Brodbelt, J.S., *Mass Specttrom. Rev.*, **32**, 247 (2013).

[51] Bauer, G.B., Povirk, L.F., *Nucleic Acids Res*., **25**, 1211(1997).

[52] Kondo, N., Takahashi, A., Ono, K., Ohnishi, T., *J. Nucleic Acids,* 2010, Article ID 543531, (2010).

[53] Rajski, S. R., Williams, R.M., *Chem. Rev*., **98**, 2723 (1998).

[54] Park, H.J, Hurley, L. H., *J. Am. Chem. Soc.*, **119**, 629 (1997).

[55] Denny, W.A., *Curr. Med. Chem.*, **8**, 533 (2001).

[56] Smaill, J.B., Fan, J.Y., Denny, W.A., *Anticancer Drug Des.,* **13**, 857 (1998).

[57] Majid, A.M., Smythe, G., Denny, W.A., Wakelin, L.P., *Mol. Pharmacol.*, **71**, 1165 (2007).

[58] Wilson, W.D., Tanious, F.A., Ding, D., Kumar, A., Boykin, W.D., Colson, P., Claude, H., Bailly, C., *J. Am. Chem. Soc.*,**120**, 10310 (1998).

[59] Lucjan, S., Beth W., *Mutat. Res.*, **623**, 3 (2007).

[60] Burridge, J.M., Quarendon, P., Reynolds, C.A., Goodford, P.J., *J. Mol. Graph*., **5**, 165 (1987).

[61] Nelson, S.M., Ferguson, L.R., Denny, W.A., *Mutat. Res.*, **623**, 24 (2007).

[62] Khan, G.S., Shah, A., Zia-ur-Rehman, Barker, D., *J. Photochem. PhotoBiol. B: Biology*, **115**, 105 (2012).

[63] Privalov, P.L., Dragan, A. I., Crane-Robinson, C., Breslauer, K. J., Rameta, D. P., Minetti, A. S. A. , *J. Mol. Biol*., **365**, 1 (2007).

[64] Wemmer, D.E., Dervan, P.B., *Curr. Opin. Struct. Biol.*, **7**, 355 (1997).

[65] Kopka, M.L., Yoon, C., Goodsell, D., Pjura, P., Dickerson, R.E., Proc. *Natl. Acad. Sci. U.S.A.*, **82**, 1376 (1985).

[66] Gilbert, D.E., Feigon, J., *Curr. Opin. Struct. Biol.*, **1**, 439 (1991).

[67] Bailly, C., Chaires, J. B., *Bioconjug. Chem.*, **9**, 513 (1998).

[68] Hamelberg, D., Williams, L. D., Wilson, W. D., *J. Am. Chem. Soc.*, **123**, 7745 (2001).

[69] Bhattacharya, S., Thomas, M., *Tetrahedron Lett*., **41**, 5571 (2000).

[70] Eriksson, S., Kim, S. K., Kubista, M., Norden, B., *Biochem.*, **32**, 2987 (1993).

[71] Brown, D. G. et al., *EMBO* , **9**, 1329 (1990).

[72] Wilson, W.D., Tanious, F.A., Barton, H.J., Jones, R.L., Fox, K., Wydra, R.L., Strekowski, L., *Biochem.*, **29**, 8452 (1990).

[73] Tabernero, L., Verdaguer, N., Coll, M., Fita, I., van der Marel, G.A., Boom, J.H., Rich, A., Aymami, J., *Biochem.*, **32**, 8403 (1993).

[74] Breusegem, S.Y., Clegg, R.M., Loontiens, F.G., *J. Mol. Biol*., **315**, 1049 (2002).

[75] Stephanie M. N., Lynnette R. F., William A. D., *Mutat. Res*., **623**, 24 (2007).

[76] Lah, J., Vesnaver, G., *J. Mol. Biol*., **342**, 73 (2004).

[77] Remers, W. A., *The Chemistry of Antitumor Antibiotics*, Wiley, New York, 1979, p. 1.

[78] Williams, L. D., Egli, M. and Gao, Q., *Proc. Natl. Acad. Sci. USA*, **87**, 2225 (1990).

[79] Pigram, W.J., Fuller, W., Hamilton, L.D., *Nature New Biology*, **235**, 17 (1972).

[80] Wang, A.H.J, *Curr. Opin. Struct. Biol.*, **2**, 361(1992).

[81] Neto, B.A.D, Lapis, A.A.M., *Molecules*, **14**, 1725 (2009).

[82] Yen, Shau-Fong, Gabbay, E.J., Wilson, W.D., *Biochem.*, **21**, 2070 (1982).

[83] Tanious, F.A.,Yen, Shau-Fong, Wilson, W.D., *Biochem.*, **30**, 1813 (1991).

[84] Rao, S.N.R, Kollman, P.A., *Proc. Natl. Acad. Sci*., **84**, 5735 (1987).

[85] Bond, P.J, Langridge, R., Jennette, K.W., Lippard, S.J, *Proc. Natl. Acad. Sci.*, **72**, 4825 (1975).

[86] Wilson, W.D. et al., *J. Am. Chem. Soc.*, **120**, 10310 (1998).

[87] Martinez, R., Chacon-Garcia, L., *Curr. Med. Chem.*, **12**, 127 (2005).

[88] Denny, W.A., *Curr. Med. Chem*., **9**, 1655 (2002).

[89] Nakamoto, K., Tsuboi, M., Strahan, G.D, *Drug–DNA Interactions: Structures and Spectra.*

[90] Wheate, N.J, et al., *Med. Chem.*, **7**, 627 (2007).

[91] Baraldi, P.G et al., *IL Farmaco*, **54**, 15 (1994).

[92] Reddy, B.S.P., Sondhi, S.M., Lown, J.W., *Pharmacology and Therapeutics*, **84**, 1(1999).

[93] Ferguson, L.R., Denny, W.A., *Mut. Res.*, **258**, 123 (1991).

[94] Delbarre, A., Delepierre, M., Garbay, C., Igolen, J., Le Pecq, J.-B., Roques, B.P., Proc. *Natl. Acad.Sci. U.S.A.*, **84**, 2155 (1987).

[95] Lawrence, B. H., Virendra, B. M., Edwin D. B. Jr., Douglas, E., *Mutat. Res.*, **623**, 53 (2007).

[96] Long, E.C., Barton, J.K., *Acc. Chem. Res*., **23**, 271 (1990).

[97] Wilson, W.D., in: E., Kool, (Ed.), *DNA and RNA Intercalators; DNA and aspects of molecular biology*, in: D. Barton, K. Nakanishi, (Eds.), Comprehensive Natural Products Chemistry, **7**, (1998).

[98] Simonsson, S., Samuelsson, T., Elias, P., *J. Biol. Chem.*, **273**, 24633 (1998).

[99] Singh, N. N., Lambowitz, A. M., *J. Mol. Biol*., **309**, 361 (2001).

[100] Mamoon, N. M., Song, Y., Wellman, S. E., *Biochem.*, **41**, 9222 (2002).

[101] Schleif, R., *Science*, **241**, 1182 (1998).

[102] Thuong, N. T., Hélène, C., *Angew. Chem., Int. Ed. Engl.,* **32**, 666 (1993).

[103] Jain, A. K., Bhattacharya, S., *Bioconjugate Chem.,* **21**, 1389 (2010).

[104] Nielsen, P. E., *Curr. Opin. Struct. Biol.*, **9**, 353 (1999).

[105] Ganesh, K. N., Kumar, V. A., *Acc. Chem. Res.,* **38**, 404 (2005).

[106] Seog, K.K., Bengt, N., *FEBS Lett.,* **315**, 61 (1993).

[107] Han. D., Wang, H., Ren, N., *J. Mol. Model.,* **10**, 216 (2004).

[108] Hazarika, P., Bezbaruah, B., Deka, R.P., Deka, J., Barman, T.K., Medhi, O.K., Medhi, C., *The Clarion*, **1**, 24 (2012).

[109] Juan, C.G., Rodrigo, G., Fernando, C., Lena, R., *J. Mex. Chem. Soc.,* **57**, 245 (2013).

[110] Konstantinos, G., Shaun, T.M., James, A.P., *RSC Adv***., 3**, 4066 (2013).

[111] Decatris, M.P., Sundar, S., O'Byrne, K.J., *Cancer Treat. Rev*., **30**, 53 (2004).

[112] Ge, R., Sun, H., *Acc. Chem. Res*., **40**, 267 (2007).

[113] Wang, D., Lippard, S.J., *Nature Rev. Drug Discov.,* **4**, 307 (2005).

[114] Umezawa, H., *Prog. Biochem. Pharmacol.,* **11**, 18 (1976).

[115] Wong, E., Giandomenico, C.M., *Chem. Rev.*, **99**, 2451(1999).

[116] Bonnett, R., *Chem. Soc. Rev.,* **24**, 19 (1995).

[117] Zhang, O.O., Zhang, F., Wang, W.G., Wang, X.L., *J. Inorg. Biochem.*, **100**, 1344 (2006).

[118] Baruah, H., Barry, C.G., Bierbach, U., *Curr. Top. Med. Chem.,* **4**, 1537 (2004).

[119] Jiang, G.B., Xie, Y.Y., Lin, G.J., Huang, H.L., Liang, Z.H., Liu, Y.J., *J. Photochem. Photobiol. B,* **5**, 48 (2013).

[120] Christian, G., Ivano, T., Ursula, R., *J. Am. Chem. Soc.*, **130**, 10921 (2008).

[121] Shaun, T.M., James, A.P., *J. Phys. Chem. A.*, **115**, 11293 (2011).

[122] Mozhgan, K., Meissam, N., Sommaieh, K., Spectrochim. Acta A, **78**, 389 (2011).

[123] Sirajuddin, M., Ali, S., Badshah, A., *J. Photochem. Photobiol. B*: *Biology*, **124**, 1 (2013).

[124] Garbett, N.C., Chaires, J.B., *Expert opin. Drug Discov.*,**7**, 299 (2012).

[125] Cheatham III, T.E., Miller, J.L., Fox, T., Darden, T.A., Kollman, P.A., *J. Am. Chem. Soc.*, **117**, 4193 (1995).

[126] Bellissent-Funel, M., Hassanali, A., Havenith, M., Henchman, R., Pohl, P., Sterpone, F., van der Spoel, D., Xu, Y., Garcia, A.E., *Chem. Rev.*, **116**, 7673 (2016).

[127] Yu, H., Ren, J., Chaires, J.B., Qu, X., *J. Med. Chem.*, **51**, 5909 (2008).

[128] Chalikian, T.V., Breslauer, K.J., *Curr. Opin. Struct. Biol.*, **8**, 657 (1998).

[129] Zimmer, C., Luck, G., *Use of circular dichroism to probe structure and drug binding to DNA*. In B. J. Graham (Ed.), *Advances in DNA Sequence Specific Agents*, Greenwich: JAI Press, **1**, 51 (1992).

[130] Singh, M. P., Lown, J. W., *Lexitropsins: design and development of sequence-selective DNA minor groove binding agents as new chemotherapeutics*. In M. I. Choudhary (Ed.), *Progress in Medicinal Chemistry* (pp. 49–171). Amsterdam: Harwood Academic Publishers, **49** (1996).

[131] Nordeni, B., Tjernald, F., *Biophys Chem.*, **4**, 191 (1976).

[132] Bailly, C., Colson, P., Houssier, C., Hamy, F., *Nucleic Acids Res.*, **24**, 1460 (1996).

[133] Colson, P., Bailly, C., Houssier, C., *Biophys. Chem*., **58**, 125 (1996).

[134] Kennard, O., Hunter, W. N., *Angew Chem Int Ed Engl*., **30**, 1254 (1991).

[135] Gonzalez-Ruiz V., Olives A.I., Martin M.A., Ribelles P., Ramos M.T. and Menendez J. C., *An Overview of Analytical Techniques Employed to Evidence Drug-DNA Interactions, Applications to the Design of Genosensors*, In: M. A. Komorowska and S. OlsztynskaJanus, Eds., *Biomedical Engineering, Trends, Research and Technologies*, InTech, Castle Rock, 65(2011).

[136] Husain, M.A., Yaseen, Z., Rehman, S.U., Sarwar, T., Tabish, M., *FEBS J.*, **280**, 6569 (2013).

[137] Nejedly, K., Chladkova, J., Vorlickova, M., Hrabcova, I., Kypr, J., *Nucleic Acids Res.*, **33**, 1 (2005).

[138] Saito, S.T., Silva, G., Pungartnik, C., Brendel, M., *J. Photochem. Photobiol. B,* **111**, 59 (2012).

[139] Husain, M.A., Sarwar, T., Rehman, S.U., Ishqi, H.M., Tabish, M., *Phys. Chem. Chem.Phys.*, **17**, 13837 (2015).

[140] Haq, Arch. *Biochem. Biophys.*, **403**, 1 (2002).

[141] Rehman, S.U., Yaseen, Z., Husain, M.A., Sarwar, T., Ishqi, H.M., Tabish, M., *PLoS ONE*, **9**, e93913 (2014).

[142] Rehman, S.U., Sarwar, T., Ishqi, H.M., Husain, M.A., Hasan, Z., Tabish, M., Arch. *Biochem. Biophys.*, **566**, 7 (2015).

[143] Ni, Y., Lin, D., Kokot, S., *Anal. Biochem.*, **352**, 231(2006).

[144] Wang, J., Kang, X., Kuntz, I.D., Kollman, P.A., *J. Med. Chem.*, **48**, 2432 (2005).

[145] Wang, H., Laughton, C.A., *Phys. Chem. Chem. Phys.*, **11**, 10722 (2009).

[146] Wang, H., Laughton, C.A., *Molecular modeling methods to quantitative drug-DNA interactions*, In: Drug-DNA interaction protocols, *Methods Mol. Biol.*, Humana Press, Germany, 613, 19 (2010).

[147] Shaikh, S.A., Ahmed, S.R., Jayaram, B., *Arch. Biochem.* Biophys., **429**, 81(2004).

[148] Bohari, M. H., Srivastava, H. K., Sastry, G. N., *Org. Med. Chem. Lett.*, **1**, 1 (2011).

[149] Srivastava, H.K., Bohari, M. H.,Sastry, G. N., *Curr. Comput-Aid. Drug*, **8**, 224 (2012).

[150] Srivastava, H.K, Chourasia, M., Kumar, D., Sastry, G.N., J. *Chem. Inf. Model.*, **51**, 558 (2011).

[151] Reddy, A. S., Pati, S. P., Kumar, P. P., Pradeep, H. N., Sastry, G. N., *Curr. Protein Pept. Sc.*, **8**, 329 (2007).

[152] Badrinarayan, P., Sastry, G. N., J. *Mol. Graph. Model.*, **34**, 89 (2012).

[153] Badrinarayan, P., Sastry, G. N, *Comb. Chem. High. T Scr*, **14**, 840 (2011).

[154] Bohari, M. H., Sastry, G. N., *J. Mol. Model*., **18**, 4263 (2012).

[155] Srivastava, H. K., Sastry, G. N., *J. Biomol. Struct. Dyn.*, **31**, 522 (2013).

[156] Srivastava, H. K., Sastry, G. N., *J. Chem. Inf. Model.*, **52**, 3088 (2012).

[157] Seibel, G.L., Singh, U.C., Kollman, P.A., *Proc. Natl. Acad. Sci. USA*, **82**, 6537 (1985).

[158] Sharp, K.A., Friedman, R.A., Misra, V., Hecht, J., Honig, B., *Biopolymers*, **36**, 245 (1995).

[159] Misra, V.K., Honig, B., *Proc. Natl. Acad. Sci. USA*, **92**, 4691 (1995).

[160] Singh, S.B., Ajay, Wemmer, D.E., Kollman, P.A., *Proc. Natl. Acad. Sci. USA*, **91**,

7673 (1994).

[161] Mikheikin, A.L., Zhuze, A.L., Zasedatelev, A.S., *J. Biomol. Struct. Dyn.*, 19, 175 (2001).

[162] Sergey K. F., Cestmir, K., Pavla, K., Larisa, S., Milena, S., Petr, S., *Langmuir*, **26**, 4999 (2010).

[163] Ray, L., Hillary, S. R. G., Michael, J. P., Michael, K. G., *Biophysical*, 80, 140 (2011).

[164] Chi, H. M., *J. Phys. Chem. B*, **120**, 6010 (2016).

[165] Gellman, S. H., Haque, T. S., Newcomb, L. F., *Biophys. J*., **71**, 3523 (1996).

[166] Peter, Y., Ekaterina, P., Maxim D., Frank-Kamenetskii, *Nucleic Acids Res*., **34**, 564 (2006).

[167] Friedman, R. A., Honig, B., *Biophys. J.*, **69**, 1528 (1995).

[168] Lerman, L.S., *Biochem.*, **49**, 94 (1963).

[169] Muller, W., *J. Mol. Biol*., **35**, 251 (1968).

[170] Carol W. M., Karl F. K., Dennis G. K., David W. H., *J. Med. Chem.*, **20**, 1055 (1977).

[171] Kai-Xian, C., Nohad, G., Bernard, P., *Nucleic Acid Res.*, **14**, 3799 (1986).

[172] Laurence, H.P., Jane, V.S., Brian, D.H., Stephen, N., *Nucleic Acid Res.*, **15**, 3469 (1987).

[173] Kuntebomanahalli, T., Apoorva, G. U., Elvis, F. M., Mushtaque, S. S., Evans, C. C., Mayur, C. Y., *Nucleosides, Nucleotides and Nucleic Acids*, **34**, 309 (2015).

[174] Francisca, B., Jordi, M., Francisco, G., Jose M. G., *Biochem. Pharmacol*., **37**, 2133 (1988).

[175] Neidle, S., Pearl, P.H, Herzyk, P., Helen, M.B., *Nucleic Acid Res.*, **16**, 8999 (1988).

[176] Kennard, O., *Pure & Appl. Chem.*, **65**, 1213 (1993).

[177] Jose, G., Angel, R. O., Federico, G., *J. Med. Chem.*, **36**, 1548 (1993).

[178] Suresh, B. S., Ajay, David, E. W., Peter, A. K., *Proc. Natl. Acad. Sci. USA*, **91**, 7673 (1994).

[179] James, M. C., David, E. G., William, A. D., *Biochem.*, **34**, 13682 (1995).

[180] Wieslaw, M. C., Lidia, H., Lawrence, H., Dominic, A. S., Draginaja, B. D., Christopher, J. M., *J. Med. Chem.*, **38**, 3043 (1995).

[181] Berger, I., Su, Li, Jeffrey, R.S., ChulHee, K., Thomas, G.B., Alexander, R., *Nucleic Acid Res.*, **23**, 4488 (1995).

[182] Shahla, M., Martine, P.F, Gresh, N., Hillairet, K., Taillandier, E., *Biochem.*, **37**, 6165 (1998).

[183] Clare, E.B., Searle, M.S., *Nucleic Acid Res.*, **27**, 1619 (1999).

[184] Chaudhari, P., Ganguly, B., Bhattacharya, S., *J. Org. Chem.*, **72**, 1912 (2007).

[185] Molina, A., et al., *J. Org. Chem.*, **64**, 3907 (1999).

[186] Mazur, S., Tanious, F.A., Ding, D., Kumar, A., Boykin, D.W., Simpson, I.J., Neidle, S., Wilson, W.D., *J. Mol. Biol*., **300**, 321 (2000).

[187] Amutha, R., Subramanian, V., Nair, B.U., *Chem. Phys. Lett.*, **344**, 40 (2001).

[188] Ji, Y., *Biorg. Med. Chem.*, **9**, 2905 (2001).

[189] Satz, A.L., Brucie, T.C., *J. Am. Chem. Soc.*, **123**, 2469 (2001).

[190] Martinez, V.,et al., *J. Med. Chem.*, **47**, 1136 (2004).

[191] Rajendra, A., Nair, B.U., *Biochem. Biophys. Acta*., **1760**, 1794 (2006).

[192] Kamal, A., et al., *Biorg. Med. Chem. Lett.*, **17**, 5400 (2007).

[193] Kamal, A., et al., *Biorg. Med. Chem. Lett.*, **17**, 1557(2009).

[194] Kamal, A., et al.,, *Tetrahedron*, **66**, 5498 (2010).

[195] Kamal A, et al., *Biorg. Med. Chem. Lett.*, **18**, 4747 (2010).

[196] Sankaraiah, N. et al., *Biorg. Chem.*, **59**, 130 (2015).

[197] Nagle, P.S., *Org. Biomol. Chem.*, **8**, 5558 (2010).

[198] Pandya, P., Islam, M.M., Kumar, G.S., Jayaram, B., Kumar S., *J. Chem. Sci.*, **122**, 247 (2010).

[199] Yan, L., Gongke, W., Wen, T., Xiaoxiao, H., Meihua, X., Xiang, L., *Spectrochim. Acta A*, **82**, 247 (2011).

[200] Rupesh, N., Wilson, W.D., *Binding to the DNA Minor Groove by Heterocyclic Dications: From AT Specific Monomers to GC Recognition with Dimers*, *Curr. Protoc. Nucleic Acid Chem*., CHAPTER: Unit8.8.

[201] Matthieu, W., Arnab M., Benjamin, B., Krystyna, Z., James, T. H. , Richard L.,*J. Am. Chem. Soc.*,**134**, 8588 (2012).

[202] Periyasamy, V., Chandrabose, S., Singh, S.K. , Jaganathan, N. , Kandasamy, S. , Pitchai D., *J. Biomol. Struct. Dyn.*, **31**, 561 (2013).

[203] Yocheved, G., Hanoch S., *J. Chem. Inf. Model.*, **54**, 96 (2014).

[204] Snyder, R.D., Patrick, A. H., Maguire, J.M., Trent, J.O., *Environ. Mol. Mutagen*, **54**, 668 (2013).

[205] Barbara, F., Remacle, F., *Phys. Chem. Chem. Phys.*, **16**, 14070 (2014).

[206] Fu, X., Liu, D., Yuan L., Wei H., Mao, Z., Le, X., *Dalton Trans.*, **43**, 8721 (2014).

[207] Mariya, R., Sana, A., *RSC Adv.*, **5**, 72394 (2015).

[208] Cuneca, F., Greciano, O., Gunaratnam, M., Haider, S., Munnur, D., Nanjunda, R., Wilson, W.D., Neidle, S., *Bioorg. Med. Chem. Lett.*, **18**, 1668 (2008).

[209] Sissi, C., Lucatello, L., Krapcho, A.P., Maloney, D.J., Boxer, M.B., Camarasa, M.V., Pezzoni, G., Menta, E., Palumbo, M., *Bioorg. Med. Chem. Lett.*, **15,** 555, (2007).

[210] Nadai, M., Doria, F., Di Antonio, M., Sattin, G., Germani, L., Percivalle, C., Palumbo, M., Richter, S.N., Freccero, M., *Biochimie*, **93**, 1328 (2011).

[211] Hampel, S.M., Sidibe, A., Gunaratnam, M., Riou, J.F., Neidle, S. *Biorg. Med. Chem. Lett.,* **20**, 6459 (2010).

[212] Izabella, C., Shinobu, S., Bernard, J., Shigeori, T., *Biorg. Med. Chem.,* **22**, 2593 (2014).

[213] Islam, M.M, Satoshi F., Shinobu S., Tatsuo O., Shigeori T., *Biorg. Med. Chem.*, **23**, 4769 (2015).

[214] Bansal, Y., Silakari, O., *Biorg. Med. Chem.,* **20**, 6208 (2012).

[215] Anthony, R.P., Rodrigo, V. D., Louis, S.K., John, C.D., Leory, B.T., *J. Med. Chem.*, **41**, 1252 (1998).

[216] Chen, J., Wang, Z., Li, C.M., Lu, Y.P., Vaddady, K., Meibohm, B., Dalton, J.T., Miller, D.D., Li, W.J., *J. Med. Chem.*, **53**, 7414 (2010).

[217] Vyankat, A.S., Anup, N.K., Sougata, G., Piyush, M., Rajesh, G., Navanath, M.K., Anupa, A.K., Balu, A. C., Vaishali, S.S., New J. Chem., **39**, 4882 (2015).

[218] Gellis, A., Kovaic, H., Boufatah, N., Vanelle, P., Euro. *J. Med. Chem.,* **43**, 1858 (2008).

[219] Guven, O.O., Erdogan, T., Goker, H., Yildiz, S., *Biorg. Med. Chem. Lett.*, **17**, 2233 (2007).

[220] Kopanska, K., Najda, A., Zebrowska, J., *Biorg. Med. Chem.*, **12**, 2617 (2004).

[221] Prinka, S., Vijay, L., Kamaldeep, P., *RSC Adv.*, **6**, 14741 (2016).

[222] Zimmer, C., Wahnert, U., *Prog. Biophys. Mol. Biol*., **47**, 31 (1986).

[223] Baraldi, P.G., Bovero, A., Fruttarolo, F., Preti, D., Tabrizi, M.A., Pavani, M.G., Romagnoli, R., Med. Res. Rev., **24**, 475 (2004).

[224] Fede, A., Billeter, M., Leupin, W., Wuethrich, K., *Structure (Cambridge, MA, U.S.)*, **1**, 177 (1993).

[225] Fornander, L.H., Lisha, W., Martin, B., Per, L., Bengt, N., *J. Phys. Chem. B*., **117**, 5820 (2013).

[226] Suzann, M., Farial, A.T., Daoyuan, D., Arvind, K., David, W.B., Ian, J.S., Stephen, N., Wilson, W.D., *J. Mol. Biol*., **300**, 321 (2000).

[227] Reis, L.A., Ramos, E.B., Rocha, M.S., *J. Phys. Chem. B*, **117**, 14345 (2013).

[228] Mukherjee, A., Singh, B., *J. Luminescence*, **190**, 319 (2017).

[229] Xiaogang, Q., Jinsong, R., Peter, V.R., Albert, S.B., Chaires, J.B., *Biochem.*, **42**, 11960 (2003).

[230] Ihtshamul, H., John, E.L., Babur, Z.C., Terence, C.J., Chaires, J.B., *J. Mol. Biol*., **271**, 244 (1997).

[231] Frank, G.L., Peter, R., Annelies, Z., Lieve D., Robert, M.C., *Biochem.*, **29**, 9029 (1990).

[232] Jinsong, R., Terence, C.J., Chaires, J.B., *Biochem.*, **39**, 8439, (2000).

[233] Francisca, B., Damiana, C., Jose, P., *Nucleic Acids Res.*, **30**, 4567 (2002).

[234] Carrasco, C., Veizn, H., Wilson, W.D., Ren, J., Chaires, J.B., Jonathan, B., *Anticancer Drug Dis.*, **16**, 99 (2001).

[235] Fenfei, L., Waldemar, P., Chaires, J.B., *Biochem.*, **37**, 1743 (1998).

[236] Jose, P., Derek, J.C., John, O.T., Neus, F., Teresa, P., Izabela, F., Waldemar, P., Chaires, J.B., *J. Med. Chem.*, **48**, 8209 (2005).

[237] Fenfei, L., Chaires, J.B., Waring, M.J., *Nucleic Acids Res.*, **31**, 6191 (2003).

[238] Sophia, Y.B., Robert, M.C., Frank, G.L., *J. Mol. Biol*., **315**, 1049 (2002).

[239] Lucjan, S., Donald, B.H., Roman, L.W., Kent, D.S., Wilson, W.D., *J. Mol. Recognit.*, **2**, 158 (1989).

[240] Lynnette, R.F., William, A.D., *Mutat. Res.*, **258**, 123 (1991).

[241] Wilson, W.D., Farial, A.T., Henryk, J.B., Lucjan, S., Boykin, W.D., *J. Am. Chem. Soc.*, **111**, 5008 (1989).

[242] Alain, D., Muriel, D., Christiane, G., Jean, I., Jean-Bernard, L.P., Bernard, P.R., *Proc. Nati. Acad. Sci.*, **84**, 2155 (1987).

[243] Lucjan, S., Beth, W., *Mutat. Res.*, **623**, 3 (2007).

[244] Ahmadi, F., Alizadeh, A.A., Bakhshandeh-Saraskanrood, F., Jafari, B., Khodadadian, M., *Food Chem. Toxiocol.*, **48**, 29 (2010).

[245] Nahid, S., Soraya, M.F., Fahimeh, K., *J. Photochem. Photobiol. B*: *Biology*, **128**, 20 (2013).

[246] Fouzia, P., Rumana, Q., Farzana, L.A., Saima, K., Safeer, A., *J. Mol. Struc*., **1004**, 67 (2011).

[247] Ahmadia, F., Jamalia, N., Jahangard-Yektaa, S., Jafari, B., Nourib, S., Najafic, F., Rahimi-Nasrabadid, M., *Spectrochim. Acta A*, **79**, 1004 (2011).

[248] Yusra, R., Shumaila, A., Mohammed, A.H., Tarique, S. Abad, A., Shamsuzzaman, Mohammad, T., *Arch. Biochem. Biophys.*, **625**, 1 (2017).

[249] Gaetano, M., Giambattista, G., Alessio, L., Chiara, F., Maria, G.M., Paolo, L., Maria, P.C., *Amino Acids*, **42**, 641 (2012).

[250] Fengling, C., Guangquan, H., Xiaoying, J., Guisheng, Z., *Int. J. Biol. Maccromol.*, **51**, 406 (2012).

[251] Sonika, C., Manish, S., Gunjan, T., Ranjana, M., *Int. J. Biol. Maccromol.*, **51**, 406 (2012).

[252] Ruina, H., Guiqing, X., Xiaoying, J., Yao, Ge, Zaikun, X., Fengling, C., *J. Biochem. Molecular Toxicology*, **26**, 193 (2012).

[253] Samuel, T.S., Givaldo, S., Cristina, P., Martin, B., *J. Photochem. Photobiol. B: Biology*, **111**, 59 (2012).

[254] Agrawal, S., Jangir, D.K., Mehrotra, R., *J. Photochem. Photobiol. B: Biology*, **120**, 177 (2013).

[255] Huihui, L., Xiaoyang, B., Jia, L., Chongzheng, X., Xianlong, W., Xiaodi, Y., *Spectrochim. Acta A: Mol. Biomol. Spect.*, **107**, 227 (2013).

[256] Xiaoyue, Z., Guowen, Z., Junhi, P., *Int. J. Biol. Maccromol.*, **74**, 185 (2015).

[257] Rabindranath, B., Bijaya, K.S., Kalyan, S.G., Swagata, D., *Int. J. Biol. Maccromol.*, **42**, 14 (2008).

[258] Fahimeh, J., Parisa, S.D., *Arab. J. Chem.*, **10**, S3947 (2017).

[259] Claudia, S., Stefano, M., Donald, M.C., *Biopolymers*, **103**, 449 (2014).

[260] Mahvash, F.D., Gholamreza, D., Majid, M., Mohammad, A.H.F., *Adv. Pharm. Bull.*, **5**, 477 (2015).

[261] Subastria, A., Ramamurthya, C.H., Suyavarana, A., Mareeswarana, R., Lokeswara Raoa, P., Harikrishnab, M., Kumar, M.S., Sujathae, V., Thirunavukkarasua, C., *Int. J. Biol. Maccromol.,* **78**, 122 (2015).

[262] Sahabadi, N., Amiri, S., *Spectrochim. Acta A: Mol. Biomol. Spect.*, **138**, 840 (2015).

[263] Chadha, D., Agarwal, S., Mehrotra, R., *MAPAN- J. Metrol. Soc. I*, **28**, 273 (2016).

[264] Faiza, A., Amin, B., Ian, S.B., Saira, T., Bhajan, L., Muhammad, N.T., *Inorganica Chimica Acta*, **442**, 46 (2016).

[265] Mahvash, F. D., Gholamreza, D., Majid, M., Mohammad, A. H. F., *Journal of Reports in Pharmaceutical Sciences*, **5**, 80, 2016.

[266] Joana, V., Sottomayor, M.J., *J. Mol. Struc*., **975**, 292 (2010).

[267] Barnali, J., Sudipta, S., Debanjana, G., Debosreeta, B., Nitin, C., *J. Phys. Chem. B*, **116**, 639 (2012).

[268] Naveed, K.J., Asima, S., Azra, Y., Sana, S., Rumana, Q., Sayed ul H., *Spectrochim. Acta A*, **74**, 1135 (2009).

[269] Changqun, C., Xiaoming, C., Fei, G., *Spectrochim. Acta A* ,**76**, 202 (2010).

[270] Mehmet, L.Y., Nuran, O., *J. Electro. Chem*., **653**, 56 (2011).

[271] Guowen, Z., Peng, F., Lin, W., Mingming, H., *J. Agric. Food Chem.,* **59**, 8944 (2011).

[272] Nahid, S., Saba, H., *Spectrochim. Acta A: Mol. Biomol. Spect.*, **96**, 278 (2012).

[273] Sonika, C., Ranjana, M., *Int. J. Biol. Maccromol.,* **60**, 213 (2013).

[274] Belal, H.M.H, Mostafa, A.G., Abdullah, I.E., *Int. J. Electrochem. Sci.,* **12**, 9488 (2017).

[275] Shumaila, A., Yusra, R., Tarique, S. Mohammed, A.H., Abad, A., Shamsuzzaman, Mohammad, T., *Spectrochim. Acta A: Mol. Biomol. Spect.*, **186**, 66 (2017).

[276] Md. F.A., Supriya, V. Masood, A.K., Amaj, A.L., Hina, Y., *Bio. Mac.*, **113**, 300 (2018).

[277] Pritha, B., Gopinatha, S.K., *J. Photochem. Photobiol. B: Biology*, **138**, 282 (2014).

[278] Chan, S.L., Lee, M.C., Tan, K.O., Yang, L.K., Lee, A.S., Flotow, H., Fu, N.Y, Butler, M.S., Soejatro,D.D., Buss, A.D., Yu, V.C., J. *Biol. Chem.*, **278**, 20453 (2003).

[279] Basu, P., Bhowmik, D., Kumar, G.S., *J. Photochem. Photobiol. B*, **129**, 57(2013).

[280] Kim, H.L., Kim, H.N., Lee, E.J., *Genomics & Informatics*, **4**, 16 (2006).

[281] Urbanova, J., Lubal, P., Slaninova, I., Taborska, E., Taborsky, P., *Anal. Bioanal. Chem.*, **394**, 997 (2009).

[282] Vijayabharathi, R., Sathyadevi, P., Krishnamoorthy, P., Senthilaraja, D., Brunthadevi, P., Sathyabama, S., Priyadarisini, V.B., *Spectrochim. Acta A*, **89**, 294 (2012).

[283] Haijian, R., Ekhlasi, E., Daneshwar, R., *E-J. Chem*., **9**, 1587 (2012).

[284] Josephine, S.S., Daniel, G.M., Balis, M.E, *Experimental .and Molecular Pathology*, **6**, 199 (1967).

[285] Guowen, Z., Yadi, Ma., *Food Chem.*, **141**, 41 (2013).

[286] Deepak, K.J., Gunjan, T., Ranjana, M., Suma, K., *J. Mol. Struc*., **969**, 126 (2010).

[287] Yan, F., Guocai, L., Gurong, F., Yutian, W., *Analytical Sciences*, **25**, 1333

(2009).

[288] Can, O., Hayriye, E.S.K., *J. Photochem. Photobiol. B*: *Biology*, **138**, 36 (2014).

[289] Tarique, S., Mohammed, A.H., Sayeed Ur R., Hassan, M.I., Mohammad, T., Mol. Biosyst. DOI: 10.1039/c4mb00636d.

[290] Haijan, R., Tavakol, M., *E-J. Chem.*, **9**, 471 (2012).

[291] Nahid, S., Somaych, B., *Spectrochim. Acta A: Mol. Biomol. Spect.*, **136**, 1454(2015).

[292] Deepesh, K., Rama, *P. Der Pharma Chemica*, **4**, 66 (2012).

[293] Guowen, Z., Xing, Hu., Peng, Fu, ., *J. Photochem. Photobiol. B*: *Biology*, **105**, 53 (2012).

[294] Mo, T., Guowen, Z., Junhui, P., Chunhong, X., *Spectrochim. Acta A: Mol. Biomol. Spect.*, **155**, 28 (2016).

[295] Nahid, S., Maryam, M., *Mol. BioSyst.*, **10**, 338 (2014).

[296] Tarique, S., Sayeed, Ur R., Mohammed, A.H., Hassan, M.I., Mohammad, T., *Int. J. Biol. Maccromol.*, **73**, 9 (2015).

[297] Mohammed, A.H., Sayeed, ur R., Hassan. M.I., Tarique, S., Mohammad, T., *RSC Adv.*, **5**, 64335 (2015).

[298] Mohammed, A.H., Tarique, S. Sayeed, Ur R., Hassan, M.I., Mohammad, T., *Phys. Chem. Chem. Phys.*, **17**, 13837 (2015).

[299] Shuyun, B., Lili, Y., Yu, W., Bong, P., Tianjiao, W., *J. Luminescence*, **132**, 2355 (2012).

[300] Mrksich, M., Dervan, P.B., *J. Am. Chem., Soc.*, **115**, 2572 (1993).

Chapter 2

Methodology

Chapter 2

Methodology

2.1 Quantum Mechanics

Quantum mechanical methods are capable of theoretically determining the molecular, electronic structures and other properties of the molecules for many electronic systems, after solving the Schrödinger equation. *Ab intio* means from the beginning. The basic aim is to calculate the molecular properties of molecule without considering any experimental parameters.

Quantum mechanics is one of the oldest mathematical formalisms of theoretical chemistry. According to quantum mechanics (QM), all possible information on a molecular system can be obtained from wave function, Ψ which is obtained by solving the Schrödinger wave equation.[1,2] The Schrödinger equation is the fundamental equation in QM and provides a complete electronic description of the molecule and can be written as:

$$H\Psi = E\Psi \qquad (2.1)$$

This is the time independent Schrödinger wave equation, where H is Hamiltonian Operator of the system and E is the energy eigenvalue of operator H. Ψ is known as wave function and is a well behaved mathematical function whose square represents the probability density.[3]

The Hamiltonian operator for a molecular system can be written as the kinetic and potential energies of the nuclei and electrons:

$$\hat{H} = \hat{T}_n + \hat{T}_e + \hat{V}_{ne} + \hat{V}_{ee} + \hat{V}_{nm} \tag{2.2}$$

where, $\hat{T}$ is the kinetic energy operator, $\hat{V}_{ne}$ is the external potential produced by the nuclei acting on the electrons, and the rest two terms, $\hat{V}_{ee}$ and $\hat{V}_{nm}$ are the electron-electron and nuclei-nuclei potentials, respectively.

The Schrödinger equation can be solved for only one electron systems but it is unsolvable for many electron systems. Therefore various approximations must be introduced in order to extend the utility of the method for polyatomic systems.

One of the primary approximations to solve the wave function according to the Born Oppenheimer approximation which allows the wave function of a molecule to be broken into its electronic and nuclear (vibrational, rotational) components.[4-6] Using this approximation solutions to the Schrödinger equation lead to value of effective electronic energy, which are dependent on relative nuclear coordinates. Electrons are lighter than the nucleons, so they move much faster and relax rapidly to the ground state configuration given by nuclear positions. So the wave function for electrons moving in the potential field of nuclei is calculated and treated as fixed point charges:

$$\Psi(r_l, R_\alpha) = \Psi(r_l, R_\alpha)\phi(R_\alpha) \tag{2.3}$$

Thus the complete Hamiltonian is given by

$$\hat{H}_{elec} = \sum_{i=1}^{N}(-\frac{1}{2}\nabla^2{}_l) + \sum_{i=1}^{N} v(r_l) + \sum_{i<j}^{N}\frac{1}{r_{ij}} = \hat{T}_e + \hat{V}_{ne} + \hat{V}_{ee} \tag{2.4}$$

in which

$$v(r_l) = -\sum_{\alpha} \frac{Z_\alpha}{r_{1\alpha}} \tag{2.5}$$

The electronic wave function $\Psi(r_l, R_\alpha)$ depends upon the electronic coordinates, while the nuclear coordinates enter only parametrically.

Quantum chemical methods are classified into two main classes: semi-empirical methods (such as CNDO, MNDO etc.) and non-empirical *(ab initio,* DFT etc.) methods. Semi-empirical methods use some parameters derived from experimental data to simplify the calculations and it is less demanding than *ab initio* methods while *ab initio* calculation uses the correct Hamiltonian and does not use the experimental data other than the values of some physical constants such as speed of light, masses, charges of the electrons and nuclei, Plank's constant etc.

2.1.1 Hartree Self -Consistent Field Method

The solution of the electronic Schrödinger wave equation of molecule, which is formed by writing the Slater determinant, consists of the nuclear–nuclear interaction energy, which has constant value for a given geometry, the nuclear-electron attraction, which is dependent on one electron coordinate and the electron–electron repulsion, which depends on two electron coordinates.[7,8]

The Hamiltonian is

$$H_e = \sum_{p}^{N} h_p + \sum_{i=1}^{N} \sum_{j>i}^{N} g_{ij} + V_{nn} \tag{2.6}$$

where,

$$h_p = -\frac{1}{2}\nabla^2 - \sum_{a} \frac{Z_a}{|R_a - r_p|} \tag{2.7}$$

and

$$g_{ij} = \frac{1}{|r_i - r_j|} \tag{2.8}$$

Where one electron operator h_i describes the motion of i[th] electron in field of all nuclei, g_{ij} is two electron operator giving the repulsion between two electrons while V_{nn} is the nuclear-nuclear interaction energy. The energy can be expressed as

$$E = \sum_{i}^{N} \langle \varphi_i | h_i | \varphi_i \rangle + \frac{1}{2} \sum_{ij}^{N} \left(\langle \varphi_j | J_i | \varphi_j \rangle - \langle \varphi_j | K_i | \varphi_j \rangle \right) + V_{nn} \tag{2.9}$$

$$J_{12} = \langle \varphi_1^{(1)} \varphi_2^{(2)} | g_{12} | \varphi_1^{(1)} \varphi_2^{(2)} \rangle \tag{2.10}$$

where, J operator represents the classical repulsion between the two charge distributions described by $\varphi_{12}(1)$ and $\varphi_{22}(2)$

$$K_{12} = \langle \varphi_1^{(2)} \varphi_2^{(2)} | g_{12} | \varphi_2^{(1)} \varphi_1^{(2)} \rangle$$

The K operator represents the exchange integral that has no classical analogue.[9] The Hartree-Fock method is known as mean field approximation in which the average electron-electron repulsion is taken into account.[10]

2.1.2 Density Functional Theory

Density functional theory (DFT) in contrast to wave function QM describes the energy as a functional of the electron density. The use of DFT theory is known for its applicability in medium and large molecular system. Kohn et al. derived a set of one-electron equations that enables one to calculate the electron density and consequently the total energy of the system. His methodology has been proven very successful recently as calculations using this method are found to be computationally inexpensive while at the same time they have reasonable agreement with experimental values for relatively large chemical systems.[11,12] Thus, similar to a Hartree-Fock calculation, as shown above,

the total energy (E_{el}) of a DFT calculation is split into a kinetic energy term, a term representing the electron-nucleus attractions, a term for the Coulomb interactions between the electrons and an exchange-correlation term (E_{xc}), respectively.[13] The first three terms resemble the Hartree-Fock Hamiltonian shown in Eq. 2.6, 2.7 and 2.8 above as a function of the nuclear coordinates (R), and the coordinates of the electrons (r).

$$E_{el} = -\frac{1}{2}\sum_i \int \varphi_i(r_1)\nabla^2\varphi_i(r_1)dr_1 + \sum_A \int \frac{Z_A}{|R_A - r_1|}\rho(r_1)dr_1 + \frac{1}{2}\frac{\rho(r_1)\rho(r_2)}{|r_1 - r_2|}dr_1 dr_2 + E_{XC} \tag{2.11}$$

The exchange-correlation functional is unknown and therefore approximate equations have been set-up to estimate its contribution. Generally, E_{xc} is split into an exchange functional (E_X) and a correlation functional (E_C). The exchange functional essentially represents the interactions of two ferromagnetic spins in different orbitals, whereas the correlation is the pairing energy of electrons in the same orbital.[14,15]

$$E_X^{Slater} = -\frac{9}{4\alpha_{ex}}\left(\frac{3}{4\pi}\right)^{1/3}\sum_\gamma \int \left[\rho_1^\gamma(r_1)\right]^{4/3} dr_1 \tag{2.12}$$

In this equation α_{ex} is an exchange scale factor, which has the value ⅔ for an electron gas. Commonly used correlation energy functional E_C^{VWN} is due to Vosko, Wilk and Nusair [14] and represents the correlation energy per electron in a gas $\varepsilon_C\left[\rho_1^\alpha \rho_1^\beta\right]$ with spin densities ρ_1^α and ρ_1^β.

$$E_C^{VWN} = \int \rho_1(r_1)\varepsilon_c\left[\rho_1^\alpha(r_1), \rho_1^\beta(r_1)\right]dr_1 \tag{2.13}$$

Combination of Slater exchange and Vosko-Wilk-Nusair correlation, which both are directly derived from the homogenous electron gas equations, gives the Local Density Approximation. However, to correct the non-local terms other (better) exchange and

correlation functional have been developed. Two popular approaches are due to Lee, Yang and Parr (LYP correlation functional)[15] and Perdew and Wang (PW91 correlation functional)[16] but there are many more available functional, each with their own qualities and/or drawbacks.

Thus, Becke benchmarked DFT methods against a test set of experimentally known ionization energies, electron affinities and proton affinities with high accuracy.[17] He came up with a three parameter (hybrid) density functional method to estimate the contributions of the exchange and correlation functions and optimized the values of these three fit parameters (A, B, and C) against the experimental data in the test set. There are many possibilities of combining the various exchange and correlation functional, but throughout the years, the most popular one has become the B3LYP method, although it should be realized that it is not necessarily the most accurate one.[17] Essentially the hybrid density functional method B3LYP has the following form:

$$E_{XC}^{B3LYP} = AE_X^{Slater} + (1-A)E_X^{HF} + B\Delta E_X^{Becke} + E_C^{VWN} + C\Delta E_C^{LYP} \qquad (2.14)$$

Thus, it takes the local density approximation functions of Slater and Vosko-Wilk-Nusair (Eqs.2.11 and 2.12), the Hartree-Fock exchange, a correction term to the exchange due to Becke and Lee-Yang-Parr correction for non-local correlation factors. The coefficients A, B, and C are essentially fit-parameters obtained through fitting the energies of B3LYP/6–31G* calculations against experimentally obtained ionization energies and electron affinities. Consequently, the B3LYP method in essence is not an *ab initio* method. In strict sense as the term *ab initio* means starting from scratch without prior knowledge of experiment. Nevertheless, these fit-parameters have created an accurate and low-cost computational method and as a result B3LYP has become one of the widest used

techniques in science over the past decade. The B3LYP method, as well as other hybrid and non-hybrid DFT methods, have been shown to be extremely accurate and versatile for computational chemists. Although in general the accuracy of DFT is not as good as high level ab initio methods, such as coupled cluster methods, their speed in combination with reasonable accuracy makes them a very popular and useful methodology.[15,18]

The commonly used B3LYP method fails to predict dispersion energy. The application of DFT is limited, and used for the systems where the dispersion part is the dominant part. In that case the calculated interaction energy values are always underestimated. A breakthrough in DFT and computational chemistry in general appeared when Becke developed the hybrid density functional procedures.[17-18]

2.1.3 Basis Set

A basis set is a set of mathematical functions which are used to describe the shape of orbitals in an atom, and also used to approximate theoretical calculation or modelling. There are two types of basis sets that dominate the area of *ab initio* calculations. Those are Slater Type Orbitals (STO) and Gaussian Type Orbitals (GTO).[19,20] Slater type orbitals are simple exponentials that mimic the exact eigen fucntions of the hydrogen atom and have the functional form:

$$S(r) = N_s e^{-\xi r^2} \tag{2.15}$$

where r is radial distance from the nucleus, N_s, is normalization constant, and ζ is a constant known as the orbital exponent, which governs with size of the orbital.

Slater functions are good approximations to atomic wave function. The introduction of GTOs to replace STOs in calculations made the evaluation of three and four center

integrals more rapid. A Gaussian Type Orbital for an s-type atomic orbital with the same orbital exponent as STO has the form

$$g(r) = N_g e^{-\zeta r^2} \tag{2.16}$$

where N_g is the normalization constant .

The STO basis set has an advantage that they have direct physical interpretation and thus are naturally good basis for molecular orbitals. The STOs have the shortcoming that most of the required integrals needed for the SCF procedure must be calculated numerically; hence it is computationally expensive whereas wave functions with GTO basis set are much easier to compute.

The minimal basis set is a smallest basis set which represent one basis function for each type of occupied orbital in the separated atoms.[21] At least three Gaussian functions are required to represent STO closely. STO-3G basis set is the most commonly used minimal basis set and the notation represents that basis set approximates shape of STO by single contraction three GTO orbitals. In general STO-nG basis set is minimal basis set in which n Gaussian functions are used to represent each orbital (in general n=2-6).

The Pople basis set [22, 23] are another family of basis sets written as 6-31G. This indicates that each core orbital is described by a single contraction of six GTO primitives which describes each core orbital and two contractions, of which one with three primitives and another with one primitive describe each valence shell orbital. These types of basis set are very popular for organic molecules. Sometimes it may be denoted as 6-31G* or 6-31G**. A single asterisk (*) indicates that a set of d primitives has been added to atoms other than hydrogen while two asterisks (**) mean that a set of p primitives has been added to hydrogen as well. These are called polarization functions because they give

the wave function more flexibility to change shape. One or two plus signs can also be added, such as 6-31+G* or 6-31++G*. A single plus sign indicates that diffuse functions have been added to atoms other than hydrogen. The second plus sign indicates that diffuse functions are being used for all atoms.

2.1.4 Electron Correlation

Since electrons repel each other, they try to avoid each other. This type of effect is called electron correlation. Thus the motion of each electron is correlated with the motion of the other electrons in the system. When the molecule forms, there is also the possibility for the angular correlation around the bond direction. Within the HF formalism the antisymmetric wavefunction is approximated by a single Slater determinant, which does not include this Coulomb correlation. Therefore the energy calculated with HF theory is different from the exact energy of the system. The difference between these two is called the correlation energy

$$E_{corr} = E_{exact} - E_{HF} \tag{2.17}$$

The development of methods to determine the correlation contributions accurately and efficiently leads to a broad categories of approaches called post HF methods.

Many post-HF calculations which include configuration interaction (CI), multi configurational self-consistent field (MCSCF), Moller-Plesset perturbation theory (MPn) and coupled cluster theory (CC) considers the electron correlation.

2.2 Molecular Mechanics

In case of large molecules or biomolecules where the semiemprical methods are less effective, it is still possible to model the behavior of molecule by totally avoiding quantum mechanics. This method is referred as molecular mechanics (MM), which set up

a simple algebraic expression for the total energy of compound and there is no need to compute a wave function or total energy or total electron density. Molecular mechanics is based on the simple classical model where the atoms are treated as hard spheres and bonds as spring. In MM the energy of molecule is expressed as the energy due to the geometry of a molecule, from a few specific interactions within a molecule. These interactions include bonded terms such as bond stretching, bending and twisting and non-bonded terms such as van der Waals attractions or repulsions of atoms that come close together, and the electrostatic interactions between partial charges in a molecule due to polar bonds.[24-26] The total potential energy can be written as the sum of energies of these two types of interactions:

$$E = E_{bonded} + E_{non-bonded} \tag{2.18}$$

$$E_{bonded} = E_{streching} + E_{bending} + E_{improper} + E_{dihedral} \tag{2.19}$$

$$E_{non-bonded} = E_{elec} + E_{vdw} \tag{2.20}$$

The approximations adopted in MM calculations make the computation less expensive. Due to the speed of the MM, the large biomolecular system such as protein and DNA are treated using this method. For the system where the database for parameterization is available, it is possible to predict the accurate geometries and corresponding energies in a short time. The shortcoming of this method is that it ignores electrons and there are many chemical properties which are not defined within this method such as electronic excited states, shape and energies of molecular orbitals. This method is for the limited class of molecules for which force field is parameterized.

2.2.1 Force Field

Force field refer to a mathematical function with a set of parameters (obtained experimentally as well as theoretically from computer intensive quantum calculations) to represent the potential energy of a molecular system. All the bonded and non-bonded terms in equation 2.14 are represented as a force field to calculate the potential energy of the system at a given conformation. The Simple MM energy expression will be

$$E_{MM} = \sum_{bonds} k_b (d - d_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} k_\phi [1 + \cos(n\phi + \delta)] + \sum_{\substack{non-bonded \\ pairsAB}} \left\{ \varepsilon_{AB} \left[\left(\frac{\sigma_{AB}}{r_{AB}} \right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}} \right)^{6} \right] + \frac{1}{4\Pi\varepsilon_0} \frac{q_A q_B}{r_{AB}} \right\} \quad (2.21)$$

Where d, θ, and ϕ are the bond distances, angles and torsions respectively. d_0 and θ_0 are equilibrium values; and η and δ are the torsion multiplicity and phase, respectively. K_b, K_θ and K_ϕ are bonded force constant. r_{AB} is the non-bonded distance, ε_{AB} and σ_{AB} are the van der Waals parameters between atoms A and B. q_A, q_B are atomic partial charges and ε_0 is the vacuum permittivity.

There are various types of force fields depending upon the level of accuracy. Prior to any molecular modeling calculation it is important to first choose the required force field according to the calculation. The various force fields have been developed for different reasons and with different applications. AMBER [27, 28] CHARMM [29], GROMOS [30,31] are the force fields which are widely used for the simulation of biological macromolecules (proteins and DNA). OPLS, MM1-4, MMFF, TRIPOS and CVFF are the other existing force field. A comprehensive comparison of force fields found that some of force fields worked well at general level, however many of them encountered

problems with some molecular systems, so it is important to carefully test a force field when it is applied to molecules for which it was not originally developed.[32]

2.2.2 Popular Force Fields

AMBER (Assisted Model Building with Energy Refinement) was first developed by Kollman et al. [http://ambermd.org/] has a historical significance on the parameter transferability and use of charges derived from electrostatic potential fitting.[27,28] The results obtained with this force field is good for proteins and nucleic acids but not for other systems. So the force field should be chosen carefully.

CHARMM (Chemistry at HARward Macromolecular Mechanics) developed by Karplus et al. [http://www.charmm.org] was originally derived for proteins and nucleic acids, [33] and now used for large number of macromolecules for molecular dynamics, solvation, crystal packing, vibrational analysis and also for the quantum mechanics/molecular mechanics (QM/MM) studies. This force field uses five valence terms, one of which is electrostatic and is a basis for other force fields.

GROMOS (Groningen Molecular Simulation) developed at the University of Groningen and the ETH (Eidgenössische Technische Hochschule) of Zurich [http://www.igc.ethz.ch/GROMOS/index] is quite popular for predicting the dynamical motion of molecules and bulk liquids, also being used for modelling biomolecules. It uses five valence terms, one of which is electrostatic.[34] Its parameters are currently being updated.[31]

2.3 Molecular Docking

Molecular docking method is used to predict the structure (or structures) of the intermolecular complex formed between two or more molecules. The docking program generates large number of possible structures, and so they require ranking according to

their score to identify those of most interest.[35-40] The three important components of docking are:

(1) Representation of the system

(2) Conformational space search via a search algorithm

(3) Ranking of potential solutions using the scoring function.

The main application of docking process is to computationally simulate the molecular identification process and accomplish an optimized conformation so that the free energy of the overall complex is minimized. The docking method involves many degrees of freedom. With six degrees of translational and rotational freedom as well as conformational degrees of freedom of each molecule there are large number of possible binding modes between two molecules. Computationally it would be too expensive to generate all the possible conformations. Various algorithms have been developed to tackle the docking problem. These algorithms can be characterized according to the number of degrees of freedom that they ignore.[41-46]

The simplest algorithms treat the two molecules as rigid bodies and explore only six degrees of translational and rotational freedom. This approximation is used in the earliest algorithms for docking of small ligands to the binding sites of proteins and DNA. Kuntz and co-workers used this algorithm in the docking program DOCK.[47]

To perform conformationally flexible docking the conformational degrees of freedom need to be accounted. In this method only the conformational degrees of freedom is included that only considers the conformational space of ligand: the receptor is invariably assumed to be rigid. All of the common methods for searching conformational space have been incorporated at some stage into various docking algorithms.[48]

Monte Carlo (MC) methods have been used to perform molecular docking, often in conjunction with simulated annealing. At each iteration of the MC procedure the internal conformation of ligand is changed (by rotating about a bond) or the entire molecule is randomly translated or rotated. The conformation obtained by this transformation is tested with an energy-based selection criterion. If it passes the criterion, it will be saved and further modified to generate conformations. The iterations will proceed until the predefined quantity of conformations is collected.[48-50]

Genetic algorithms (GA) can also be used to perform molecular docking.[51-53] The idea of the GA stems from Darwin's theory of evolution. Degrees of freedom of the ligand are encoded as binary strings called genes. Genes make-up the 'chromosome' which represents the pose of the ligand. GA involves two types of genetic operator's mutation and crossover. Mutation makes random changes to the genes and crossover exchanges genes between two chromosomes. Each chromosomes codes not only for the internal conformation of the ligand but also for the orientation of the ligand within the receptor site. Both the orientation and the internal conformation will thus vary as the populations evolve. The score of each docked structure within the site act as the fitness function used to select the individuals for the next iteration.[53-55] Genetic algorithms have been used in Autodock,[54] GOLD,[55] DIVALI[56] and DARWIN.[57]

2.3.1 Scoring Functions

Docking algorithms are able to generate a large number of potential solutions. Some of these can be rejected because of high energy clashes with the receptor and remainder must be assessed using some scoring function. Scoring functions involve in estimating, rather than calculating binding affinity for the ligand binding to the receptor. Scoring

functions can be divided in force-field based, empirical and knowledge-based scoring functions.[58]

Classical force field based scoring function [59-63] computes the binding energy by calculating the sum of non-bonded (electrostatics and van der Waals) interactions. The electrostatic terms are calculated by a coulombic formulation while the van der Waals terms are handled using Lennard –Jones potential function. Hydrogen bonding was taken into account with an additional term in the electrostatic energy term. In case of force-field scoring function, the problem of slow computational speed arises, so the cut-off distance is used to analyze the non-bonded interaction. For example, early versions of DOCK,[64,65] GOLD [66] and Autodock [54] employed energy functions based on the AMBER force field as their internal scoring engine.

The empirical scoring functions also known as energy component methods is based on the physical interaction terms such as Van-der-Waals interactions or hydrogen bonding.[67] This type of scoring function is used in PLP,[68] ChemScore,[69,70] X-score,[71] GOLD[66] and GLIDE.[72,73] In this method, the binding energy decomposes into several energy components, such as hydrogen bond, ionic interaction, hydrophobic effect and binding entropy. All these components are multiplied by a coefficient and then summed up to give a final binding energy. The coefficients are optimized to give a good fit to a training set of molecules. For different softwares, each term in empirical scoring functions may be treated in a different manner.[74-77]

Knowledge-based scoring function [78-82] uses statistical analysis for the observed interatomic contact frequencies and/ or distances between the ligand and the biomolecule. It can be assumed that only those molecular interactions which are close to maximum

frequency of the interactions in the database favor the binding process and this results increase in the overall binding affinity. The observed frequency distributions are further converted into pair-wise atom type potentials or knowledge based potentials. Examples for this type of scoring function, ranked by their first publication date, include Muegge's PMF,[83-85] DrugScore,[86,87] IT-Score[88-91] and KECSA.[92]

As compared to other physics based methods, knowledge–based potentials are conceptual and computational simple due to their pairwise characteristics. Knowledge-based potentials include all the energetic factors between the receptor-ligand interactions implicitly with pairwise potential. They are derived from statistical analysis of pure structural information without the help of experimental binding data. Knowledge based scoring functions are applied to reproduce protein-ligand binding pose rather than binding energies but some recent studies correlates these potential to the binding affinity data which shows better correlation than other types of scoring functions.[91,92]

2.3.2 Popular docking programs

Molecular docking is the most important and useful areas of bioinformatics whose purpose is to determine the binding geometry of a biomolecule onto a second molecule generally a small ligand and is used by the pharmaceutical industry for identifying drug candidate compounds. Docking algorithms are computationally demanding since they consist of generating and evaluating a large amount of different molecule conformations and placements. There are large number of docking programs and search algorithm that are available for carrying out docking calculations.

Morris et al. developed the program **AutoDock** which uses a grid based scheme for energies of individual atoms, allows quick computation of the interaction energy of the

receptor-ligand complex. The strength of AutoDock can be attributed to the Monte Carlo simulated annealing, evolutionary, genetic and Lamarckian genetic algorithm methods. The bound conformation of enzyme-inhibitor complexes, peptide-antibody complexes, nucleic acid-drug complexes, protein-protein complexes etc. predicted from AutoDock shows great success.[54]

GLIDE (Grid Based Ligand Docking with Energetics) software also uses a grid-based scheme to represent the shape and properties of the receptor and approximate a close and complete systematic search for conformational, orientational and positional space of the docked ligand using a OPLS-AA force field.[93]

SURFLEX is a fully automatic flexible molecular docking algorithm that presents results evaluated for reliability and accuracy in comparison to the crystallographic experimental results on 81 protein/ligand pairs of substantial structural diversity.[94]

GOLD performs flexible docking for ligands into receptor binding site which uses genetic algorithm for the conformational search that produces powerful tool identification and screening of novel lead compounds. Gold gives high accuracy and reliability and its genetic algorithm parameters are optimized for vast range of virtual screening applications.[66,72,73,95,96]

FlexX classify the docked ligand conformers using a pose-clustering, where the placement of rigid core fragment is based on interaction geometry between fragment and receptor groups. Before docking, FlexX cuts the ligand into pieces at the rotatable bond and puts an important fragment into the active site, and incrementally builds up the ligand again, using the other pieces. For a protein which has known 3D structure and a drug

molecule, FlexX modeled the geometry of the protein-ligand complex and estimates the binding affinity.[97,98]

In **DOCK** the geometries of ligands and binding sites are described by sets of spheres, which attempts to fit each compound from a selected database into the active site and the spheres could be overlapped by means of an approximate clique-detection procedure. In the recent version of DOCK, steric matching-scores with electrostatic and molecular mechanics interaction energies are considered for the ligand receptor complex. In scoring the docked orientations atomic hydrophobicity descriptors are being considered.[99]

2.4 Molecular Dynamics Simulations

The internal motions of proteins and consequences of their inherent dynamic nature are important for their functional aspects. The dynamic properties such as simple conformational changes, bending modes, ligand binding, protein folding etc. are different at different time scales. Biological activity of a molecule is the result of its time dependent interactions with other molecules or its environment. Molecular dynamics (MD) simulations which are a computer simulation of physical movements of atoms and molecules as a function of time. MD simulations provide the atomistic view of dynamic properties of biomolecules at different time scales which is often not possible or difficult to calculate using experimental techniques.[100]

In MD, the motion of any bimolecular system is studied under the effect of a "force" is simulated according to time by solving the Newton's equation of motion (second law), which for simple atomic system may be written as

$$F_i = m_i.a = m_i.\frac{dv_i}{dt} = m_i.\frac{d^2r_i}{dt^2} \qquad (2.22)$$

This describes the motion of particle of mass m_i along the coordinate r_i with force F_i on m_i in same direction. The forces on every atom are calculated by the force-fields which are the complicated equations but easy to calculate. The force-field calculates the forces acting on the molecular system, and classical Newton's law of motion is used to calculate accelerations and velocities on each particle. The force can also be expressed as the gradient of the potential energy function

$$F_i = -\nabla_i V \tag{2.23}$$

Combining eqn. 2.22 and 2.23 two equations yields

$$m_i . \frac{d^2 r_i}{dt^2} = -\frac{\partial V}{\partial r_i} \tag{2.24}$$

The above equation relates the derivative of the potential energy to the changes of the atomic coordinates as a function of time. As the potential energy is a function of the atomic positions (3N) of all the atoms in the system. Due to the complicated nature of this function, this equation can only be solved numerically with some approximations. After calculating the force on each particle, Newton's laws of motions are integrated to generate new positions and velocities for specified time-steps. Various numerical algorithms have been developed for integrating the equation of motion (Table 2.1) [101,102]. At each time step a set of coordinates is calculated which can be expressed in the form of trajectories and a single coordinate set as frame. MD trajectories calculate the micro and macroscopic properties and these calculations are based on the principles of statistical mechanics.[100] MD calculates the microscopic properties of the system such as position and velocities of each individual atom of the system. However, the

macroscopic properties such as number of particles (N), volume (V), energy (E), temperature (T), pressure (P), chemical potential of particles (μ).[103] These macroscopic properties are used to calculate the thermodynamic properties with time.

A microscopic state of the system is defined by the positions and momenta of all the particles of a system. The state of all the particles in any system can be defined by its position and momentum coordinates of 6N dimensional space which is known as phase space. Thus at any given time, the system corresponds to a point of the multidimensional space.The total energy of the system can be distributed by different ways among the N particles of the system. An ensemble is the collection of all possible systems which have different microscopic states but have the same macroscopic or thermodynamic state. On the basis of set of constant macroscopic properties the different types of ensembles are canonical (*NVT*), the grand canonical (μVT), microcanonical (*NVE*) and the isothermal-isobaric (*NPT*) ensemble. Properties and partition function for different types of ensembles are shown in Table 2.2.[104]

Table 2.1: Different types of integration algorithms with associated mathematical equation used to generate velocity and position at each time step.[102]

Algorithms	**Equation to update new position**	**Equation to update new velocity**
Verlet	$x(t+\Delta t) = 2x(t) - x(t-\Delta t) + \frac{f(t)}{m}(\Delta t)^2$	$v(t) = \frac{x(t+\Delta T) - x(t-\Delta t)}{2x(t)} + \mathrm{O}(\Delta t)^2$
Velocity Verlet	$x(t+\Delta t) = x(t) + v(t) + \frac{f(t)}{m}(\Delta t)^2$	$v(t+\Delta t) = v(t) + \frac{f(t+\Delta t) + f(t)}{2m}\Delta t$
Leap-Frog	$x(t+\Delta t) = x(t) - \Delta t v\left(t + \frac{\Delta t}{2}\right)$	$v\left(t + \frac{\Delta t}{2}\right) = v\left(t - \frac{\Delta t}{2}\right) + \Delta t \frac{f(t)}{m}$
Velocity corrected Verlet	$x(t+\Delta t) = x(t) + v(t)\Delta t + \frac{4f(t) - f(t-\Delta t)}{6m}\Delta t^2$	$v(t+\Delta t) = v(t) + \frac{2f(t+\Delta t) + 5f(t) - f(t-\Delta t)}{6m}\Delta t$

Table 2.2: Different types of statistical ensembles used in MD simulation.[100]

Ensemble	**Properties**	**Partition Function**
Microcanonical	Number of particles(N), volume (V) and energy (E) are constant	Ω(E) where Ω(E) is the number of micro-states corresponding the system's energy E
Canonical	Number of particles(N), volume (V) and Temperature (T) are constant	$Z_{NVT} = \sum_i e^{-\beta E_i}$ where $\beta = \frac{1}{KT}$ and E_i is the total energy of the system in the respective microstate
Grand canonical	Chemical potential (μ), volume (V) and Temperature (T) are constant	$Z_{\mu VT} = \sum_i e^{\frac{(N_i\mu - E_i)}{KT}}$ where $\beta = \frac{1}{KT}$, Ni is total number of particles and E_i is the total energy of the system in the respective microstate

2.4.1 Steps in MD simulation

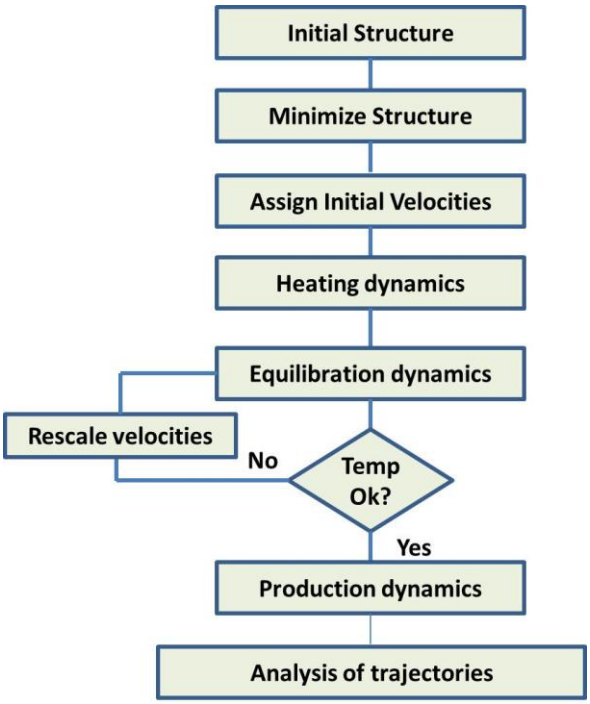


Fig. 2.1: Steps of MD Simulation.

An initial set of coordinates or configuration of the simulated molecule is required before starting the MD simulation (Fig. 2.1). Generally, the structures are obtained from protein data bank,[104] these structures are retrieved from X-ray scattering or NMR experiments. These crystal structures have high energy interaction like Pauli repulsions which leads to the local distortion that results in an unstable simulation. So these systems required energy minimization before starting the calculation and minimization results a structure with lowest possible energy. Thus, this state corresponds to a temperature near 0K, where no motion can be seen.[105] There is no motion or no forces on the atoms of minimized structure. For the minimum energy the gradient of potential energy vector equals to zero this means that there is no force acting on any atom in any direction and therefore no motion will appear inside the system.

To raise the temperature from 0K to the desired value, it is required to assign initial velocities to the atoms corresponding to a Gaussian distribution for a certain temperature to provide energy to the system.[106] Now, the Newton's equations of motion are integrated to propagate the system according to the time. Fig. 2.2 shows how some properties of the system vary with time during a heating process. The kinetic energy and the temperature show exactly the same behavior. Total energy remains constant during equilibration while kinetic and potential energy behaves irregularly, showing that the total energy distributes between them in a vibrational kind of way.

The structure may be unstable after heating the system so quickly and the temperature may drop.[107] So it is needed to equilibrate the system properly before running the real dynamics simulation or the production run. Equilibration is the process where the kinetic energy and the potential energy evenly distribute themselves throughout the system. For the constant period of time, the velocities are rescaled to the values for the desired temperature. This process is done until the simulation becomes stable with respect to time which means till thermodynamic terms like temperature and energy are retained in a certain, small interval for a sufficiently long time.

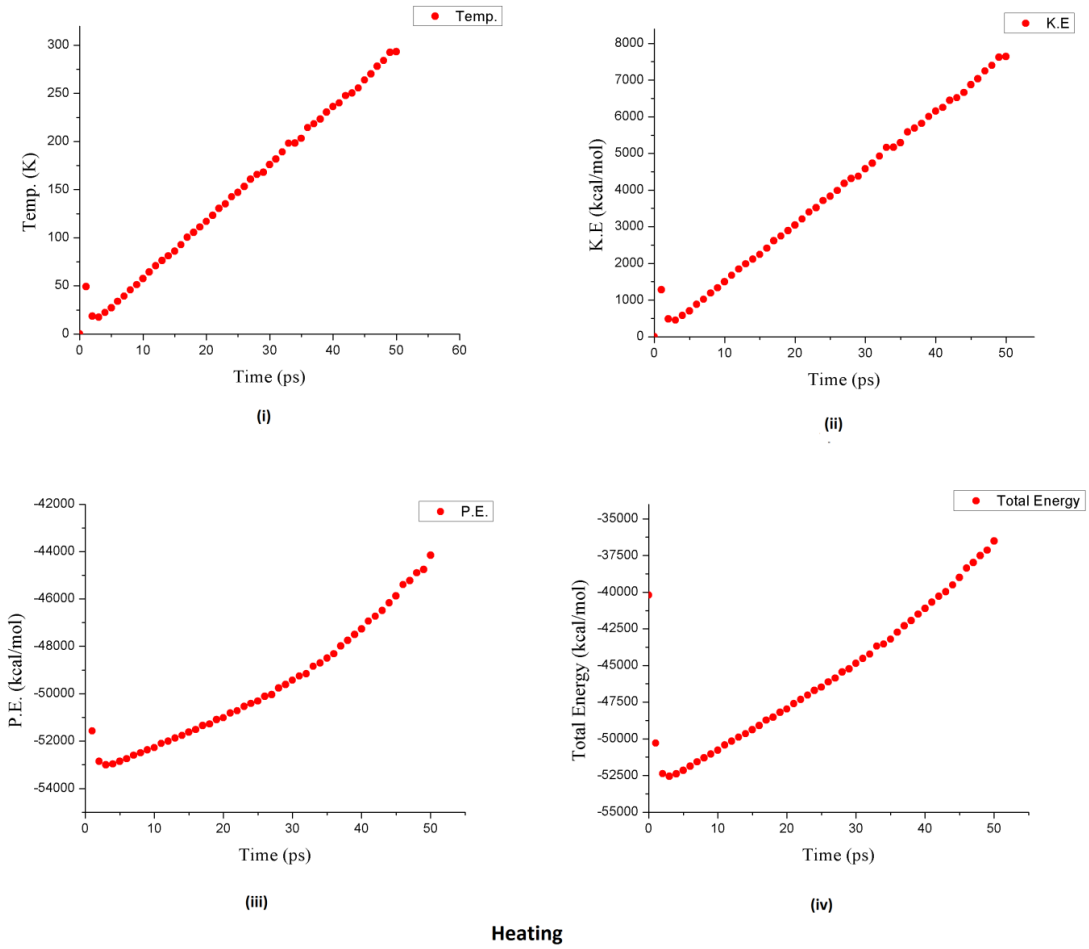


Fig. 2.2: (i) Temperature vs. Time (ii) Kinetic Energy vs. Time (iii) Potential Energy vs. Time (iv) Total Energy vs. Time.

After all these steps, the actual molecular dynamics simulation can be started by integrating Newton's equations of motion of the system for the desired period of time. This can be from several hundred picoseconds up to some nanoseconds. All the coordinates, velocities, accelerations and momenta generated during the production run are saved and used for analysis.

The important time dependent properties such as different energy terms, angles and dihedral angles or distances between atoms or whole selection of atoms can be analyzed and interpreted. Other important time dependent property of dynamic system is the root

mean square deviation (RMSD). RMSD indicates how much two structures vary in terms of differences between the coordinates of the structures and is calculated with

$$RMSD(t) = \left[\frac{1}{M} \sum_{i=1}^{N} m_i \left| r_i(t) - r_i^{ref} \right|^2 \right]^2 \tag{2.25}$$

where $M = \sum_i m_i$ and $r_i(t)$ is the position of atom i at time t after least square fitting the structure to the reference structure.

A special case of RMSD is the root mean square fluctuation (RMSF) where the reference structure is an average structure over the whole trajectory. RMSF calculation can also give better information on how structures fluctuate during a simulation, while RMSD is more appropriate to show that simulations are performed close to experimental structures.

$$RMSF_i = \left[\frac{1}{T} \sum_{t_j=1}^{T} m_i \left| r_i(t_j) - r_i^{ref} \right|^2 \right]^2 \tag{2.26}$$

where t is the time over which one wants to average and r_i^{ref} is the reference position of particle i.

MD simulations of nucleic acid are performed in the presence of explicit solvent water molecules by using periodic boundary conditions (PBC). Explicit water models used in biomolecular simulations include TIP3P, TIP4P, SPC, extended SPC/E, and F3C models among these TIP3P is the most commonly used model.[108-110]

In mid 1990s several groups performed successful MD simulations of DNA and RNA using the AMBER, CHARMM nucleic acid, or GROMOS force field. Various force fields used in the nucleic acid simulation includes, CHARMM, AMBER, GROMOS, OPLS, ENCAD and BMS26. AMBER,[111] GROMACS,[112] CHARMM [113] and NAMD [114] are the popular software packages for the simulation of nucleic acid ligand

complexes. The three currently used force-fields (AMBER-99, CHARMM-27 and BMS) provide accurate representations of standard DNA and RNA structures.[115-121] The root-mean square deviation (RMSD) and the average helical parameters of the nucleic acid are approximately equal to the experimental results and also the dihedral distributions are correct. CHARMM and AMBER are the most popular force fields.[123-128]

2.5 MMPBSA/MMGBSA Method

The molecular mechanics energies combined with the Poisson-Boltzmann or generalized Born and surface area continuum solvation (MMPBSA/MMGBSA) methods are popular approaches to evaluate the free energy difference between the bound and unbound state of two solvated molecules.[129-133] Snapshots obtained from MD simulation are used for the calculation, to get an average of the energies. The free energy of binding is calculated by the equations mentioned below-

$$\Delta G_{bind} = \Delta H - T\Delta S \tag{2.27}$$

$$\Delta G_{bind} = (\Delta E_{MM} + \Delta G_{SOL}) - T\Delta S \tag{2.28}$$

where, $\Delta E_{MM} = (E_{MM}{}^{complex} - E_{MM}{}^{receptor} - E_{MM}{}^{ligand})$ (2.29)

$$\Delta G_{SOL} = (\Delta G_{SOL}{}^{complex} - \Delta G_{SOL}{}^{receptor} - \Delta G_{SOl}{}^{ligand}) \tag{2.30}$$

$$\Delta S = (S^{complex} - S^{receptor} - S^{ligand}) \tag{2.31}$$

where ΔH, ΔE_{MM}, ΔG_{SOL}, T and ΔS is the enthalpy contribution to binding energy, the average difference in molecular mechanics energy, the solvation free energy (including

both polar and non-polar component), the temperature and change in entropy respectively.

Thus the final binding free energy of complex is equal to the sum of an intermolecular energy, a solvation free energy and an entropic term. Polar solvation free energies are calculated either by solving the linear Poisson Boltzmann (PB) equation or by Generalized Born (GB) model. The non-polar component of energy is evaluated from a linear relation to the solvent accessible surface area (SASA). Entropy contributions are calculated either by solving quasi-harmonic analysis or by using normal mode analysis.[134,135]

The MMPBSA calculations are based on single minimized structures and save computational time by ignoring dynamical effects. By performing the minimizations using MMGBSA method more time can be saved. The MMGBSA results varies as the simulation length increases, but there is no benefit for the simulation longer than 4ns.[132] H.K. Srivastava *et al.* proved that the interaction energies calculated using MMPBSA from the MD Simulations are in good agreement with the experimental results for the DNA-ligand system.[136,137] The MMPBSA method was first developed for the AMBER but nowadays some scripts have also been developed for the NAMD, GROMACS and APBS freely available softwares. A study reveals that the energies calculated using g_mmpbsa (GROMACS) and the AMBER MMPBSA package is approximately similar and the difference of 1-3 kcal/mol has been observed due to the difference in ΔG_{polar}.[138]

2.6 Hybrid QM/MM method

The QM/MM concept was first presented by Warshel and Levitt in 1976. They presented this method to study the chemical reaction in lysozyme.[139] Combined QM/MM theory has developed as an important approach for modeling local electronic events in large biomolecular systems. The basic concept of this method is to demonstrate the chemically active site by QM with accuracy, while the effect of the bimolecular environment is described by MM. The accuracy of QM and speed of MM combined QM/MM methods enable the modelling of reactive bimolecular systems at reasonable computational cost with the necessary accuracy. Due to the potential use of this method, this field gained the Nobel Prize in chemistry in 2013. The total energy of the system is calculated using the two available schemes, the additive and the subtractive scheme.[140-142] Regarding this boundary scheme, the labeling convention is given in **Fig. 2.3**

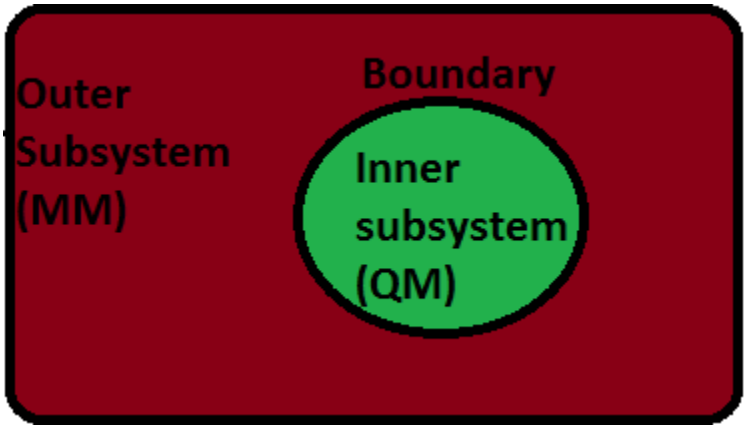


Fig. 2.3: QM/MM partitioning Scheme.

Subtractive schemes:

$$E_{QM/MM}(system) = E_{MM}(system) + E_{QM}(QM) - E_{MM}(QM) \qquad (2.32)$$

where $E_{QM/MM}$(system), E_{MM}(system), E_{QM}(QM), E_{MM}(QM) are the total energy, the MM energy of the entire system, the QM energy of the QM region and the energy of the QM region respectively.

The drawback of this scheme is that the interactions between QM and MM region is considered at MM level only, which is unreliable and this scheme requires the MM parameters for the QM region also. These parameters are not generally accessible for these systems which are present in excited electronic states or contain transition metals.

Additive schemes:

$$E_{QM/MM}(system) = E_{MM}(system) + E_{QM}(QM) - E_{QM-MM}(QM, MM) \qquad (2.33)$$

In this scheme, the total energy $E_{QM/MM}$ (system) contained only three components, E_{MM} (MM) the MM energy of the MM region only, E_{QM} (QM) the QM energy of the QM region and the E_{QM-MM} (QM, MM) this term is due to the inclusion of bonded and non-bonded interaction between QM/MM The bonded term elucidate the bond stretching, bending and torsion and the non-bonded elucidate the van-der Waals and electrostatic interactions.

The main objective of these QM/MM methods is that they considered the pairing between the electric field from the surrounding and the QM Hamiltonian from the active-site region and these methods also need that the boundary between QM and MM region is treated accurately.

The most important part of QM/MM is partitioning of the system. The basic considerations for QM/MM partitioning are:

(a) The QM region should be chosen very carefully according to the chemical problem; the QM region should be small which can then be enlarged to check the affectability of the QM/MM results with respect to such an extension.

(b) If a QM/MM division through covalent bonds cannot be avoided, cut only unconjugated single bonds, preferably without electronically demanding substituent's (e.g., cut unpolar C–C bonds).

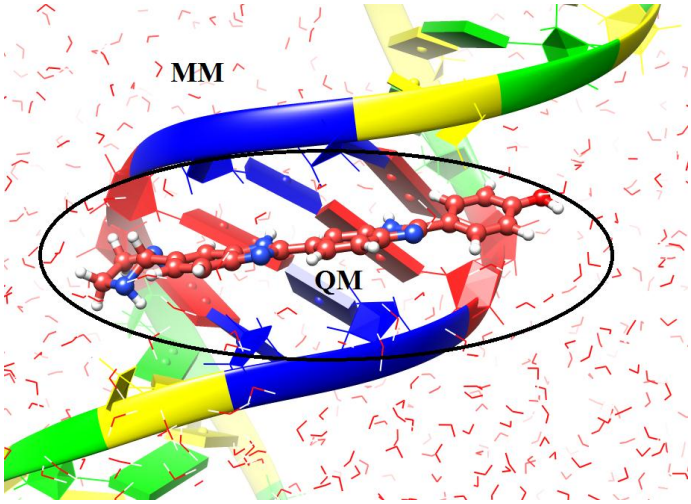


Fig. 2.4: QM/MM partitioning: drug and associated bases (QM region), remaining bases and water molecules and ions (MM region).

A common procedure for QM/MM calculations is to take a crystal structure from protein data bank as a starting point for the calculations and add hydrogen, missing atoms etc. and water molecules (soak the complete system). Thereafter, the system is equilibrated using MM followed by molecular dynamics production run and low-energy snap-shots are studied. These snapshots will eventually be taken for QM/MM calculations. These structures contain the bimolecular system in a droplet of water (20000-30000 atoms) and this setup requires a lot of prior work to avoid errors and wrong choices for the actual

QM/MM calculations. Further study of the reaction mechanism is similar to the typical methods for the study of gas-phase reaction.[143,144]

Morokuma and co-workers developed the computational chemistry techniques, especially the hybrid methods (ONIOM) for large biomolecular systems. [145-147] The two layered ONIOM scheme is shown in Fig. 2.4. In many studies, the ONIOM method is used for the study of DNA binding drugs.[148-136]

References

[1] Shiff, L. I., *Quantum Mechanics*, McGraw- Hill, NewYork, Fifth Ed. (1968).

[2] Schroedinger, E., *Ann. Physik*, **79**, 361 (1926).

[3] Pauling, L., Wilson, E. B., *Introduction to Quantum Mechanics*, McGraw-Hill, New York (1935).

[4] Born, M., Oppenheimer, J. R., *Ann. Physik.*, **84**, 457 (1927).

[5] Kolos, W., Wolniewicz, L., *J. Chem. Phys.*, **41**, 3663 (1964).

[6] Sutsliffe, B. T., *Adv. Quantum. Chem.*, **28**, 65 (1997).

[7] Levine, I. N., *Quantum Chemistry*; Chapter-11, "*The Hartree Fock Self- Consistent Method*", Pearson, Fifth Ed. (2000).

[8] Levine, I. N., *Quantum Chemistry.*, Chapter-8, "*Perturbation Theory*", Pearson, Fifth Ed. (2000).

[9] Hohenberg, P., Khon, W., *Phys. Rev.*, **136**, B864 (1964).

[10] Edimiston, C., Ruedenberg, K., *Rev. Mol. Phys.*, **34**, 457 (1963).

[11] Ziegler, T., *Chem. Rev*., **91**, 651 (1991).

[12] Khon, W., Sham, L., *J. Phys. Rev.*, **140**, 1133 (1965).

[13] Hohenberg, P., Khon, W., *Phys. Rev.*, **136**, B864 (1964).

[14] Vosko, S. H., Wilk, L., Nusair, M., *Can. J. Phys*., **58**, 1200 (1980).

[15] Lee, C., Yang, W., Parr, R. G., *Phys. Rev*., **B37**, 785 (1988).

[16] Perdew, J. P., Wang, Y., *Phys. Rev*., **B45**, 13244 (1992).

[17] Becke, A. D., *J. Chem. Phys*., **98**, 5648 (1993).

[18] Parr, R. G., Yang, W., *Density Functional Theory*, Oxford University Press (1989).

[19] Helgaker, T., Taylor, P. R., "*Modern Electronic Structure Theory*", Part II (D.

Yarkony ed.), World Scientific, pp. 727 (1995).

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

[21] Hehre, W. J., Stewart, R. F., Pople, J. A., *J. Chem. Phys.*, **51**, 2657 (1969).

[22] Binkley, Pople, J. A., *J. Am. Chem. Soc.*, **102**, 939 (1980).

[23] Frisch, M. J., J. A., Pople, Binkley, *J. Chem. Phys.*, **80**, 3265 (1984).

[24] Field, M. J., Moleculaire, L. D., Grenoble, C. *Practical introduction to the simulation of molecular systems.* Second edn. Cambridge University Press (2007).

[25] Ponder, J. W., Case, D. A., *Adv. Prot. Chem.*, **66**, 27 (2003).

[26] Tieleman, D. P., *Clin. Exp. Pharmacol. Physiol.*, **33**, 893 (2006).

[27] Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Fergurson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., Kollman, P.A., *J. Am. Chem. Soc.,* **117**, 5179 (1995).

[28] Brooks, B.R., Brooks, C.L. 3rd, Mackerell, A.D. Jr, et al., *J. Comput. Chem.*, **30**, 1545 (2009).

[29] Weiner, S.J, Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S., Weiner, P., J. *Am. Chem. Soc.*, **106**, 765 (1984).

[30] Daura, X., Mark, A.E., van Gunsteren, W.F.J., *Comp. Chem.*, **19**, 535 (1998).

[31] Horta, B.A.C., Fuchs, P.F.J., van Gunsteren, Wilfred, F.H., Philippe, H., *J. Chem. Theory Comput.*, **7**, 1016 (2011).

[32] Halgren, T. A., J .*Comput. Chem.*, **20**, 730 (1999).

[33] Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M., *J. Comput. Chem.*, **4**, 187 (1983).

[34] Gunsteren van, W.F., Berendsen, H.J.C., *GROMOS 86: Groningen Molecular*

Simulation Program Package; University of Groningen: Groningen, The Netherlands (1986).

[35] Blaney, J.M., Dixon, J.S., *Pers. Drug Discov. Desig.*, **1**, 301 (1993).

[36] Abagyan, R.T, otrov, M., *Curr. Opin. Chem. Biol.*, **5**, 375 (2001).

[37] Kuntz I., *Science*, **257**, 1078 (1992).

[38] Lengauer, T. and Rarey, M., *Curr. Opin. Struct. Biol.*, **6**, 402 (1996).

[39] Joseph-McCarthy, D., *J. Mol. Biol.*, **267**, 727 (1997).

[40] Gane, P.J., Dean, P.M., *Curr. Opin. Struct. Biol.*, **10**, 401 (2000).

[41] Schneider, G. and Böhm, H.-J., *Drug Discov. Today*, **7**, 64 (2002).

[42] Waszkowycz, B., *Curr. Opin. Drug Discov.*, **5**, 407 (2002).

[43] Toledo, S., L.M. and Chen, D., *Curr. Opin. Drug Discov. Dev.*, **5**, 414 (2002).

[44] Verkhivker, G.M., Bouzida, D., Gehlhaar, D.K., Rejto, P.A., Arthurs, S., Colson, A.B., Freer, S.T., Larson, V., Luty, B.A., Marrone,T. and Rose, P.W., *J. Comput. Aided Mol. Des.*, **14**,731 (2002).

[45] Stahl, M. and Rarey, M., *J. Med. Chem.*, **44**, 1035 (2001).

[46] Doman, T.N., McGovern, S.L., Witherbee, B.J., Kasten, T.P., Kurumbail, R., Stallings, W.C., Connolly, D.T., Shoichet, B.K. ,*J. Med. Chem.*, **45**, 2213 (2002).

[47] Kuntz, I.D., Blaney, J.M., Oately, S.J., Langridge, R. and Ferrin, T.E., *J. Mol. Biol.*, **161**, 269 (1982).

[48] Goodsell, D.S., Olson, A.J., *Protein Struc. Funct. Genet.*, **8**,195 (1990).

[49] Goodsell, D.S., Lauble, H., Stout, C.D., Olson, A.J., *Proteins*, **17**, 1 (1993).

[50] Hart, T.N., Read, R.J., *Proteins*, **13**, 206 (1992).

[51] Judson, R.S., Jaeger, E.P. and Treasurywala, A.M., *J. Mol. Struct. Theochem*, **114**,

191 (1994).

[52] Jones, G., Willet, P. and Glen, R.C., *J. Mol. Biol.,* **245**, 43 (1995b).

[53] Oshiro, C.M., Kuntz, I.D. and Dixon, J.S., *J. Comput. Aided Mol. Des.*, **9**, 113 (1995).

[54] Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., Olson, A.J., *J. Com. Chem.* **19**, 1639 (1998).

[55] Jones, G., Willett, P., Glen, R.C., Leach, A.R., Taylor, R., *J. Mol. Biol.,* **267**, 727 (1997).

[56] Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W., Taylor, R.D., *Proteins*, **52**, 609 (2003).

[57] Clark, K.P., Ajay, *J. Comput. Chem.,* **16**, 1210 (1995).

[58] Taylor, J.S., Burnett, R.M., *Proteins*, **41**, 173 (2000).

[59] Kitchen, D.B., Decornez H., Furr J.R., Bajorath J., *Nat. Rev. Drug Discov.,* **3**, 935 (2004).

[60] Gohlke, H. and Klebe, G., *Curr. Opin. Struct. Biol*., **11**, 231 (2001).

[61] Kollman, P.A., *Chem. Rev.,* **93**, 2395 (1993).

[62] Aqvist, J., Luzhkov, V.B., Brandsdal, B.O., *Acc. Chem. Res.,* **35**, 358 (2002).

[63] Carlson, H.A., Jorgensen, W.L., *J Phys Chem.,* **99**, 10667 (1995).

[64] Ewing, T.J., Makino, S., Skillman, A.G., Kuntz, I.D., *J. Comput. Aided. Mol. Des.* , **15**, 411 (2001).

[65] Welch W., Ruppert J., Jain A.N., *Chem Biol.***, 3**, 449 (1996).

[66] Accelrys Inc., San Diego, CA, USA.

[67] Bohm H.J., *J. Comput. Aided. Mol. Des.*, **12**, 309 (1998).

[68] Verkhivker, G., Appelt, K., Freer, S. T., Villafranca, J. E., *Protein Eng.*, **8**, 677 (1995).

[69] Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., Mee, R. P. , J. *Comput. Aided. Mol. Des.*, **11**, 425 (1997).

[70] Murray, C. W., Auton, T. R., Eldridge, M. D., *J. Comput. Aided. Mol. Des.*, **12**, 503 (1998).

[71] Wang, R., Lai, L., Wang, S., *J. Comput. Aided. Mol. Des.*, **16**, 11 (2002).

[72] Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., Shenkin, P. S., *J. Med. Chem.,* **47**, 1739 (2004).

[73] Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., Mainz, D. T., *J. Med. Chem.,* **49**, 6177 (2006).

[74] Gehlhaar, D.K., Verkhivker, G.M., Rejto, P.A., Sherman, C.J., Fogel, D.B., Fogel, L.J., Freer, S.T., *Chem Biol*., **2**, 317 (1995).

[75] Verkhivker, G.M., Bouzida, D., Gehlhaar, D.K., Rejto, P.A., Arthurs, S., Colson, A.B., Freer, S.T., Larson, V., Luty, B.A., Marrone, T., Rose, P.W., *J. Comput. Aided. Mol. Des.***, 14**, 731 (2000).

[76] Jain, A.N., *J. Comput. Aided. Mol. Des.,* **10**, 427 (1996).

[77] Head, R.D., Smythe, M.L., Oprea, T.I., Waller, C.L., Green, S.M., Marshall, G.R., *J. Am. Chem. Soc.* ,**118**, 3959 (1996).

[78] Mengzhu, X., Mingyue, Z., Bing, X., Yanlian, L., Hualiang, J., Jingkang, S., *J. Chem. Inf. Model.,* **50**, 1378 (2010).

[79] Mitchell, J.B.O., Laskowski, R.A., Alex, A., Thornton, J.M., *J. Comput. Chem.*, **20**,

1165 (1999).

[80] Ishchenko, A.V., Shakhnovich, E.I., *J. Med. Chem.*, **45**, 2770 (2002).

[81] Feher, M., Deretey, E., Roy, S., *J. Chem. Inf. Comput. Sci.*, **43**, 1316 (2003).

[82] Mingyue, Z., Bing, X., Cheng, L., Shanshan, L., Xian, L., Qianchen, S., Jing, L., Weiliang, Z., Xiaomin, L., Hualiang, J., *J. Chem. Inf. Model.*, **51**, 2994 (2011).

[83] Muegge, I., Martin, Y. C., *J. Med. Chem.*, **42**, 791 (1999).

[84] Muegge, I., *Perspectives in Drug Discovery & Design.*, **20**, 99 (2000).

[85] Muegge, I., *J. Comput. Chem.*, **22**, 418 (2001).

[86] Gohlke, H., Hendlich, M., Klebe, G., *J. Mol. Biol.*, **295**, 337 (2000).

[87] Velec, H. F. G., Gohlke, H., Klebe, G., *J. Med. Chem.*, **48**, 6296 (2005).

[88] Neudert, G., Klebe, G., *J. Chem. Inf. Model.*, **51**, 2731 (2011).

[89] Huang, S.-Y., Zou, X, *J. Comput. Chem.*, **27**, 1865 (2006).

[90] Huang, S. Y., Zou, X., *J. Comput. Chem.*, **27**, 1876 (2006).

[91] Huang, S. Y., Zou, X., *J. Chem. Inf. Model.*, **50**, 262 (2010).

[92] Zheng, Z., Merz, K. M., *J. Chem. Inf. Model.* **53**, 1073 (2013).

[93] Kirkpatrick, P., *Nat. Rev. Drug Discov.*, **3**, 299 (2004).

[94] Jain, A. N., *J. Med. Chem.*, **46**, 499 (2003).

[95] Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray C. W., Taylor, R. D., *Proteins*, **52**, 609 (2003).

[96] Aparna, V., Rambabu, G., Panigrahi, S. K., Sarma, J. A. R. P., Desiraju, G. R., *J. Chem. Inf. Model.*, **45**, 725 (2005).

[97] Rarey, M., Kramer, B., Lengauer, T., Klebe, G., *J. Mol. Biol.*, **261**, 470 (1996).

[98] Schellhammer, I.,Rarey, M., *Bioinformatics*, **57**, 504 (2004).

[99] Knegtel, R. M., Kuntz, I. D., Oshiro, C. M., *J. Mol. Biol.*, **266**, 424 (1997).

[100] Frenkel D, Smit B.,*Understanding molecular simulations: from algorithms to applications, 2nd edn. vol 1, Computational Science Series Academic Press, San Diego* (2002).

[101] Allen, M.P., Tildesley, D.J., *Computer simulations of liquids. Clarendon Press, Oxford* (1987).

[102] Zeigler, B., Praehofer, H., Kim, T. (eds) *Theory of modeling and simulation: integrating discrete event and continuous complex dynamic systems*, 2nd edn. Academic Press, NewYork (2000).

[103] Rapaport, D.C., *The art of MD simulation*, 2nd edn. Cambridge University Press, NewYork (2004).

[104] Hinchliffe, A., *Molecular Modelling for Beginners*, John Wiley & Sons Ltd, England (2003).

[105] Bhm, H.J., Klebe, G., Kubinyi, H., Wirkstoffdesign, Spektrum Heidelberg, D, (1996).

[106] Leach, A.R., *Molecular Modelling*, Pearson Prentice Hall, Harlow, GB, (2001).

[107] Brooks, B.R., Bruccoleri, R.E., Olafson, B. D., States, D. J., Swaminathan, S. and Karplus, M., *J. Comp. Chem.*, **4**, 187 (1983).

[108] http://www.ch.embnet.org/MD tutorial/index.html.

[109] Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L., *J. Chem. Phys*., **79**, 926 (1983).

[110] Berendsen, H.J.C., Grigera, J.R., Straatsma T.P., *J. Phys. Chem.*, **91**, 6269 (1987).

[111] Levitt, M., Hirshberg, M., Sharon, R., Laidig, K.E., Daggett, V., *J. Phys. Chem. B*,

101, 5051 (1997).

[112] Case, D.A., Berryman, J.T., Betz, R.M., Cerutti, D.S., Cheatham, III T.E., Darden, T.A., Duke, R.E., Giese, T.J., Gohlke, H., Goetz, A.W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T.S., LeGrand, S., Li, P., Luchko, T., Luo, R., Madej, B., Merz, K.M., Monard, G., Needham, P., Nguyen, H., Nguyen, H.T., Omelyan, I., Onufriev, A., Roe, D.R., Roitberg, A., Salomon-Ferrer, R., Simmerling, C.L., Smith, W., Swails, J., Walker, R.C., Wang, J., Wolf, R.M., Wu, X., York, D.M., Kollman, P.A., AMBER 2015, University of California, San Francisco (2015)..

[113] Hess, B., Kutzner, C., van der Spoel D., Lindahl, E., *J. Chem. Theory Comput.*, **4**, 435 (2008).

[114] Sapay, N., Tieleman, D.P., *J. Comput. Chem.*, **32**, 1400 (2011).

[115] Nelson, M.T., Humphrey, W., Gursoy, A., Dalke, A., Kale, L.V., Skeel, R.D., Schulten, K., *Int. J. Supercomput. Appl. High Perform Comput.* **10**, 251 (1996).

[116] York, D.M., Yang, W., Lee, H., Darden, T., Pedersen, L.G., *J. Am. Chem. Soc.*, **117**, 5001 (1995).

[117] Cheatham, III T.E., Miller, J.L., Fox, T., Darden, T.A., Kollman, P.A., *J. Am. Chem. Soc.*, **117**, 4193 (1995).

[118] Weerasinghe, S., Smith, P.E., Mohan, V., Cheng, Y.K., Pettitt, B.M., *J. Am. Chem. Soc.*, **117**, 2147 (1995).

[119] Weiner, P.K., Kollman, P.A., *J. Comput. Chem.*, **2**, 287 (1981).

[120] Nilsson, L., Karplus, M., *J. Comput. Chem.*, **7**, 591 (1986).

[121] Erik, L., Berk, H., David, S., *J. Mol Model*, **7**, 306 (2001).

[122] Cheatham, T.E., Young, M.A., *Biopolymers*, **5**, 232 (2001).

[123] Cheatham, T.E., Cieplak, P., Kollman, P.A., *J. Biomol. Struct. Dyn.*, **16**, 845 (1999).

[124] Weiner, S.J., Kollman, P.A., Nguyen, D.T., Case, D.A., *J. Comput. Chem.*, 7, 230 (1986).

[125] Foloppe, N., Mackerell, A.D., *J. Comput. Chem.*, **21**, 86 (2000).

[126] MacKerell, A.D., Jr., Banavali, N.K., *J. Comp. Chem.*, **21**, 105 (2000).

[127] Langley, D.R., *J. Biomol. Struct. Dyn.*, **16**, 487 (1998).

[128] Modesto, O., Alberto, P., Agnes, N., Luque, F.J., *Chem. Soc. Rev.*, **32**, 350 (2003).

[129] Gohlke H., Klebe, G., *Angew. Chem. Int. Ed.*, **41**, 2644 (2002).

[130] Kollman, P.A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Lillian, C., Matthew, L., Taisung, L., Yong, D., Wei, W., Oreola, D., Piotr, C., Jaysharee, S., Case, D.A., CheathamIII, T.E., *Acc. Chem. Res.*, **33**, 889 (2000).

[131] Srinivasan, J., Cheatham, T.E., Cieplak, P., Kollman, P.A., Case, D.A., *J. Am. Chem. Soc.*, **120**, 9401(1998).

[132] Hou, T., Wang, J., Li, Y.Y., Wang, W., *J. Comp. Chem.*, **32**, 866(2011).

[133] Homeyer, N., Gohlke, H., *Mol. Inf.*, **31**, 114 (2012).

[134] Schwarzl, S.M., Tschopp, T.B., Smith, J.C., Fischer, S., *J. Comput. Chem.*, **23**, 1143 (2002).

[135] Rastelli, G., Del Rio, A., Degliesposti, G., Sgobba, M., *J. Comput. Chem.*, **31**, 797 (2010).

[136] Srivastava, H.K., Chourasia, M., Kumar, D., Sastry, G.N., *J. Chem. Inf. Model.*, **51**, 558 (2011).

[137] Spackova, N., Cheatham, T.E., Ryjacek, F., Lankas, F., Meervelt, L., Hobza, P., Sponer, J., *J. Am. Chem. Soc.*, **125**, 1759 (2003).

[138] Kumari, R., Kumar, R., Lynn, A., *J. Chem. Inf. Model.*, **54**, 1951 (2014).

[139] Warshel, A., Levitt, M., *J. Mol. Biol.*, **103**, 227 (1976).

[140] Sherwood, P., Brooks, B.R., Sansom, M.S., *Curr. Opin. Struct. Biol.*, **18** 630 (2008).

[141] Sherwood, P., de Vries, A.H., Guest, M.F. et al., *J. Mol. Struct. Theochem*, **632**, 1 (2003).

[142] Senn, H.M., Thiel, W., *Angew.Chem. Int. Ed Engl.*, **48**, 198 (2009).

[143] Senn, H.M., Theil, W. *QM/MM methods for biological systems in Topics in Current Chemistry*. M. Reiher (Ed.), Springer, Berlin, **268**, 173 (2007).

[144] Friesner, R.A., Guallar, V., *Annu. Rev. Phys. Chem.* **56**, 389 (2005).

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

[146] Morokuma, K., *Korean Chem. Soc.*, **24**, 797 (2003).

[147] Morokuma, K., Musaev, D.G., Verena, T., Basch, H., Torrent, M., Khoroshun, D.V., *IBM J. Res. Dev.* **45**, 367 (2001).

[148] Rebeca, R., Begoña, G., Giuseppe, R., Arturo, S., Giampaolo, B., *J. Mole. Struct.:* Theochem, **915**, 86 (2009).

[149] Ahmadia, F., Jamalia, N., Jahangard-Yektaa, S., Jafari, B., Nourib, B., Najafic, F., Rahimi-Nasrabadid, M., *Spectrochimica Acta Part A*, **79**, 1004 (2011).

[150] Robertazzi, A., Platts, J.A., *Chem. Eur. J.*, **12**, 5747 (2006).

Chapter 3

Molecular Docking and Molecular Dynamics Study of DNA Minor Groove Binders

Chapter 3

Molecular Docking and Molecular Dynamics Study of DNA Minor Groove Binders

3.1 Introduction

Deoxyribonucleic acid (DNA) is the bio molecule which has two complementary helical strands running in anti-parallel directions, which carries genetic information from parents to offspring.[1] DNA plays an important role in cellular processes, including cell division (DNA replication) and protein synthesis (Transcription and Translation). Most of the anticancer therapies are involved in the interaction of drugs with DNA. The intercalation and groove binding are the two important modes of binding of drug with DNA. Both covalent and non-covalent types of interactions are possible in these two binding modes. Small molecules that can bind between nucleic acid base pairs are categorized as intercalators. These molecules contain planar heterocyclic groups which stack between adjacent DNA base pairs, as a result there is a decrease in the DNA helical twisting and lengthening of the DNA. On the other hand, groove binding does not induce large conformational changes in DNA and may be considered similar to standard lock and key models for ligand-macromolecular binding. Such molecules bind to both major and minor groove of nucleic acid. Minor groove binders are crescent in shape and they complement the shape of minor groove.[2-5] The binding mechanism of drug with the DNA minor groove can be described in mainly two steps. In the first step, the transfer of ligand to the

DNA minor groove by the electrostatic and hydrophobic interaction. In the second step, various types of non-covalent interactions occur between the ligand and the functional groups of DNA base pairs. These interactions usually include hydrogen bonds, hydrophobic and van dar Waals contacts, and electrostatic interactions. Most of the minor groove binding drugs bind to A/T rich region.[6-9]

In the present study, two major DNA minor groove binders classes, polyamides and diarylamidines are under taken. Different interaction models are available to explain protein-ligand binding but these models are not reasonable for DNA-ligand systems because there is no prescribed active site in the DNA, unlike protein/enzymes. Our study examines the available popular molecular modelling approaches for the estimation of DNA- drug binding free energies and compared with the experimental results. Molecular modeling methods are a powerful tool to investigate various types of non-covalent interactions, which exist between the receptor and ligand. In the number of studies for the minor groove binders, these computational methods have shown good agreement with the experimental results.[10-14]

The special approaches like molecular dynamics simulation is required to understand the complex systems like nucleic acids. The force field used in various studies to simulate nucleic acid simulation includes, CHARMM, AMBER, GROMOS, OPLS, ENCAD and BMS26. There have been many studies for the comparison of force fields for the nucleic acids but still there is a need to analyze them critically at the molecular level. AMBER 03 and CHARMM force fields were chosen for our study to simulate the DNA with small molecules. Here we have used the popular molecular mechanics energies combined with the Generalized-Born surface area (MMGBSA)

and Poisson−Boltzmann Surface Area (MMPBSA) methods to estimate the binding free energy of the binding of small molecules to DNA. These methods have been applied to wide range of molecules to estimate their ligand binding affinities.

3.2 Materials and Methods

3.2.1 Dataset

The crystal data of the B-DNA (1D30, 195D, 1D86 and 102D) were downloaded from the Protein Data Bank [15] and their experimental binding energies were collected from literature.[16] The water molecules and the ligands were removed from the 1D30,195D, 1D86 and 102D. The drug molecules extracted from these complexes were subjected to geometry optimization using Gaussian 09 at B3LYP/6-31G* level (Fig.3.1).[17]

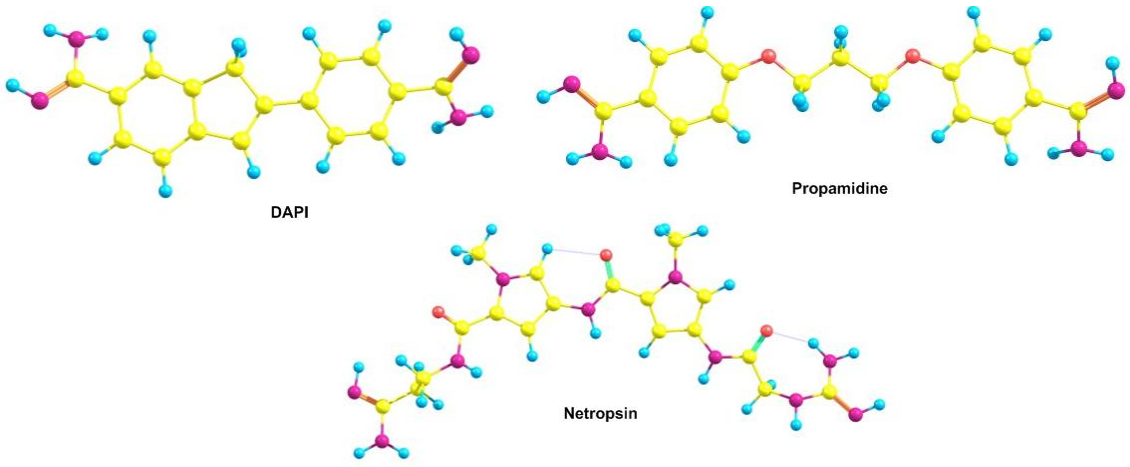


Fig. 3.1: Optimized structures of DNA minor groove binders using Gaussian 09 at B3LYP/6-31G* level.

3.2.2 Molecular Docking

Molecular docking studies were performed using Autodock 4.2 program.[18] The Gasteiger charges were added to the complex by Autodock Tools (ADT) before performing docking calculations. The binding site was centered on the macromolecule and a grid box was created with $55 \times 55 \times 60$ points and a 0.375 Å grid spacing which almost covered the entire DNA involved. Docking simulations were performed using the classical Lamarckian genetic algorithm (LGA). The 20 LGA runs with maximum of 25,00000 energy evaluations were performed. In addition, the other parameters were set to default. The pose with lowest energy of binding or binding affinity was extracted and aligned with receptor structure for further analysis.

3.2.3 Simulations

Molecular Dynamics simulation of 5 ns was carried out using the AMBER 15 [19-22] and GROMACS 4.5.6 packages.[23] AMBER 15 program was used for MD simulations of the selected docked poses. The 'leaprc.gaff" (generalized amber force field) was used to prepare the ligands, while the 'leaprc.ff03' was used for the DNA. The 'add ions' command implemented in 'tleap' of AMBER15 was used to add the Na+ ions explicitly to neutralize the system. Each system was placed in a box of TIP3P water by using 'solvateOct' command with the minimum distance between any solute atom and the boundary of the box is set to 8Å. Energy minimization was then performed to achieve the nearest stable low energy conformations (500 steps of each steepest descent and conjugate gradient method), 50 ps of heating and 50ps of density equilibration with weak restraints on the complex followed by 500 ps of

constant pressure equilibration at 300K. Cut off size of 12 Å was used for MD simulations. All long –range electrostatics were included by means of a Particle mesh Ewald (PME) method. All hydrogen and heavy atom bonds were constrained by the Shake Method, and simulations were performed with a 2fs time step and langvenin dynamics was used for temperature control. The same conditions for the final phase of equilibration were used for the production run and the coordinates were recorded at every 10 ps. Periodic boundary conditions were used for the final production run. Five hundred snapshots of the complex are obtained at every 10 ps from the MD trajectories, and all the water molecules and ions were removed before MMPBSA/MMGBSA [24] calculations using the "extract_coords.mmpbsa" script and the $\Delta G_{bind} - PB/GB$ values were calculated using the "binding_energy.mmpbsa" script.

In GROMACS MD simulation the topology and co-ordinate files for the DNA were generated by pdb2gmx program of the GROMACS package taking parameters from CHARMM [25] all atom force field and for the ligand using SwissParam [26] Web server. The coordinate and topology files of DNA and ligand were merged to obtain the final starting structure and topology file for each complex. The drug-DNA complex was placed in the center of dodecahedron periodic box. The system was then solvated in TIP3P water molecules. The total charge on the system was then neutralized by adding counter ions. The energy was minimized using steepest descent algorithm. Then the system was heated to 300K during 50 ps of constant volume simulation with 2 fs time step. The pressure was equilibrated to 1 atm during 50 ps NPT simulation with 2 fs time step. In both the simulations a restrained with force

constant of 1000 kJ/(mol/nm^2) was used. Both temperature and pressure were regulated using Berendsen algorithm. Production simulations were performed 5 ns with a 2 fs time step. The temperature and pressure were maintained at 300 K and 1 atm using the v-rescale temperature and Parninello-Rahman pressure coupling method. The binding energy was calculated using g_mmpbsa.[27] Binding energy of each snapshot was calculated for each complex. The entropy contribution was not included in the binding energy.

3.3 Results and Discussions

3.3.1 Molecular Docking Studies

Molecular docking calculations shows that all the ligands bind to AT-rich region of DNA with good docking fitness score (Table 3.1). Comparison between experimental and calculated binding energies obtained from docking studies show that the complex having lowest experimental binding energy also has the lowest calculated binding energy and vice-versa. This shows that the ability of Autodock in predicting the correct binding modes for drug DNA complex. The PDB Id **1D30** has the lowest binding energy, this shows that the interaction between DNA duplex of sequence d (CGCGAATTCGCG) $_2$ and DAPI (Fig. 3.2) is more stable in comparison to other complexes. The further studies were performed on the best pose with lowest binding energy for all complexes.

Table 3.1: List of PDB Ids considered in the present work with their DNA sequences, calculated binding energies (kcal/mol) and experimental binding energies obtained from literature.

PDB Id	DNA sequence	Ligand	$\Delta G_{calc.}$ [a]	$\Delta G_{exp.}$ [b]
1D30	5'-CGCGAATTCGCG-3'	DAPI	-9.07	-8.8
1D86	5'-CGCGAATTCGCG-3'	Netropsin	-5.42	-8.7
195D	5'-CGCGTTAACGCG-3'	Netropsin	-4.35	-8.0
102D	5'-CTTTTGCAAAAG-3'	Propamidine	-4.61	-8.2

[a]Calculated binding Energy in kcal/mol.

[b]Experimental binding energy kcal/mol.

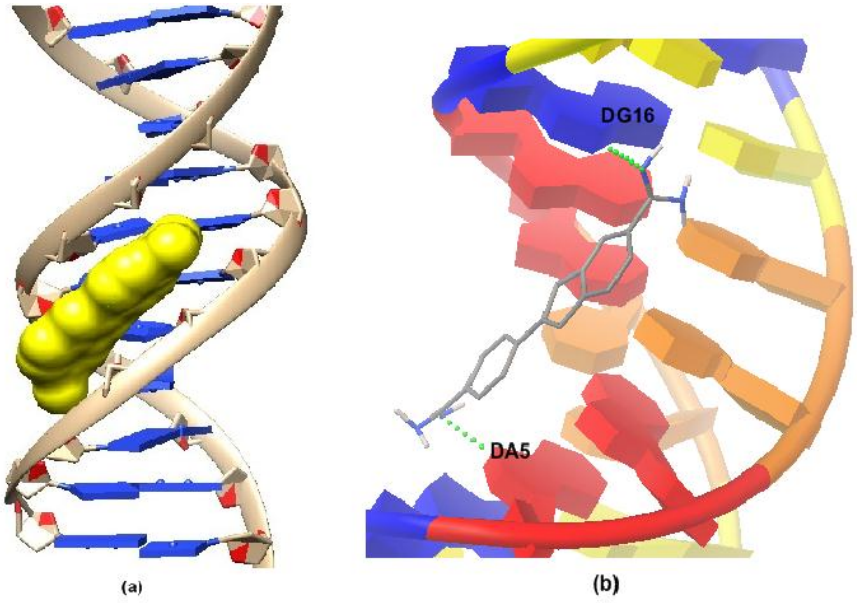


Fig. 3.2: A molecular docked model for DAPI with DNA duplex of sequence d $(CGCGAATTCGCG)_2$ (PDB ID: 1D30). (a) the full view of docking between DAPI and 1D30; (b) the binding mode between DAPI and 1D30 and the green dashed line showing hydrogen bond interactions between them.

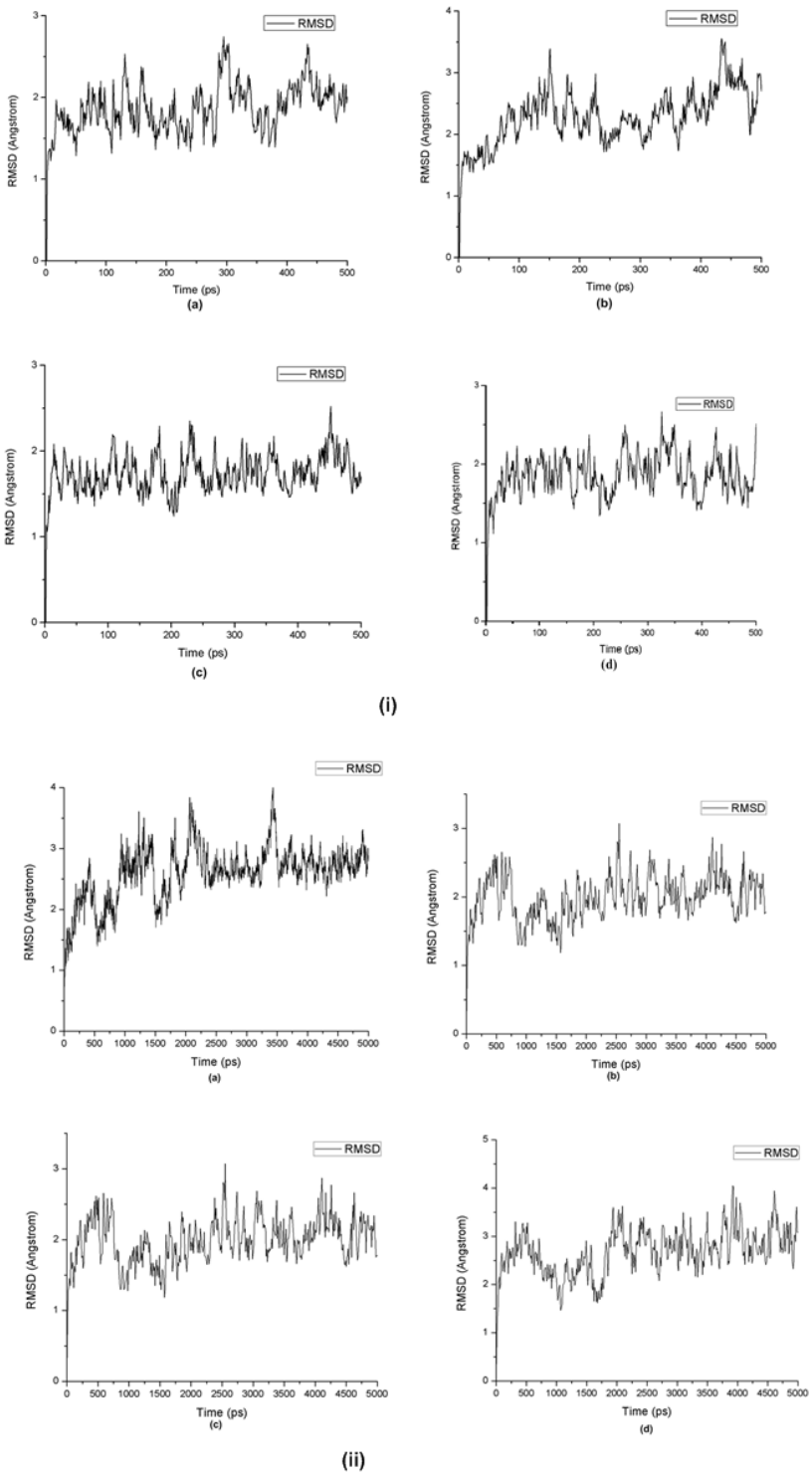


Fig. 3.3: Plots of RMSD vs. time of their trajectory for PDB Id (a) 1D30 (b) 1D86 (c) 102D (d) 195D obtained from (i) AMBER (ii) GROMACS.

3.3.2 Molecular Dynamics Studies

MD simulation from AMBER 15 and GROMACS 4.5.6 were performed for the best poses selected from the docking studies. RMSD as a function of time is plotted for all the complexes in Fig. 3.3 to obtain the systematic deviation of complexes. The rmsd profile shows that the ligands remain bound to the DNA near the preferential binding site. It has been observed that all the four complexes show almost same RMSD variation from both types of simulations. The range of RMSD is 1-3 Å for AMBER and 1-3.5 Å for GROMACS throughout the simulation (Fig. 3.3). Convergence of RMSD values shows the stability of complex. The comparison of rmsd analysis between AMBER and GROMACS simulation shows that the complexes are more stable in AMBER molecular dynamics simulation. This difference in stability is due to the different force fields used for the study. In order to identify the conformational changes during 5 ns of MD simulation, the snapshots were extracted from the MD trajectory at every 10 ps for both AMBER and GROMACS (Fig. 3.4). It has been observed from snapshots that the ligand forms interact with AT-rich region of DNA duplex and remains bounded in minor groove up to the end of simulation.

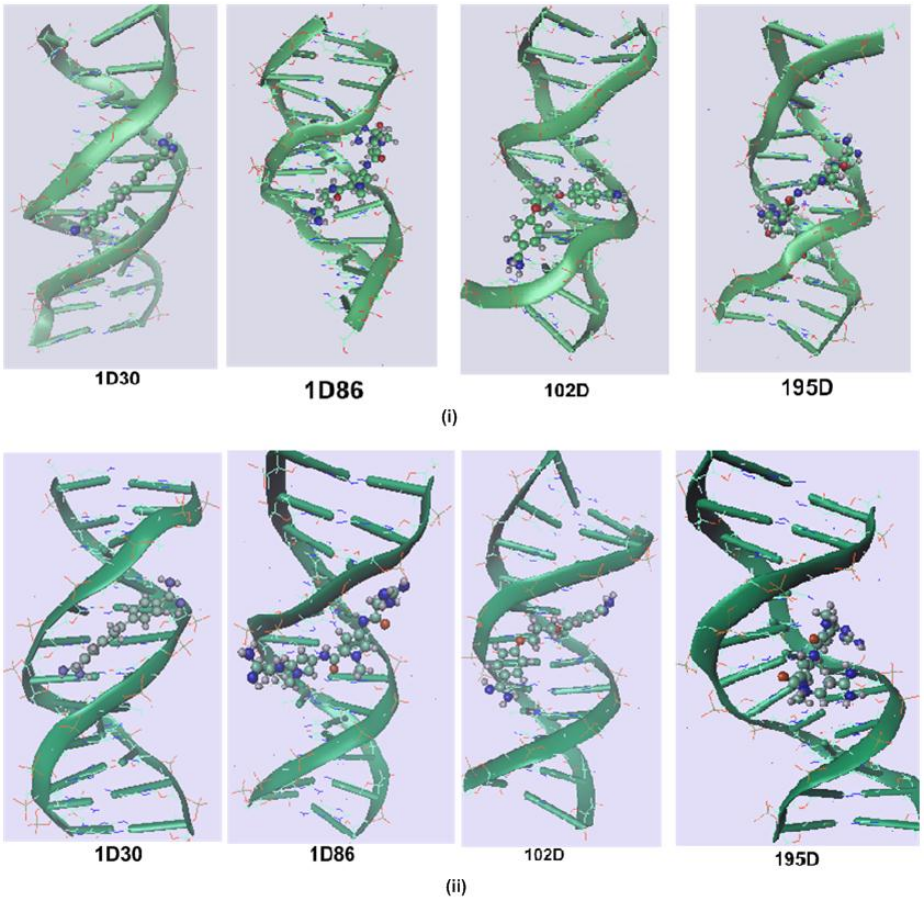


Fig. 3.4: Snapshots of different DNA-ligand complexes at 5 ns MD simulation for PDB Id (a) 1D30 (b) 1D86 (c) 102D (d) 195D obtained from (i) AMBER (ii) GROMACS.

Further, MMPBSA/MMGBSA calculations were performed from AMBER 15 by using MD trajectories to obtain binding energy values. Binding energy values predict the strength of ligand with their respective receptors. Calculated binding free energy of all four drug-DNA complexes with the contribution of van der Waals, electrostatic, solvation energy etc. is shown in Fig. 3.5 and it was found that the total binding free energy of complex with PDB Id 1D30 has lowest binding energy (-25.52 kcal/mol), so this complex is slightly more stable than others.

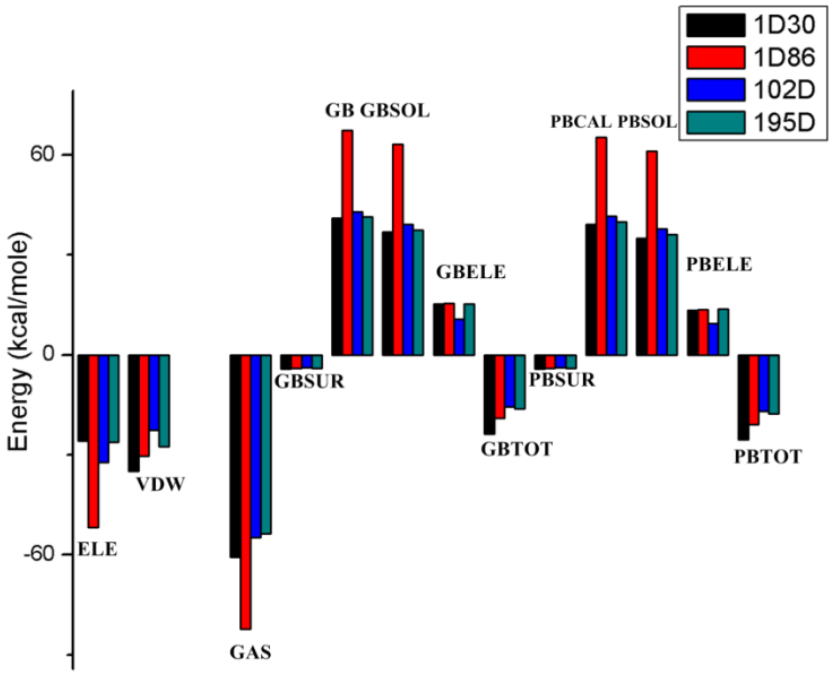


Fig. 3.5: Histogram depicting view of the contribution of various energy components to the binding free energy.

ELE: electrostatic energy; **VDW**: van der Waals energy; **INT**: internal energy; **GAS**: total gas-phase energy; **PBSUR/GBSUR**: nonpolar contribution to the solvation free energy; **PBCAL/GB**: electrostatic contribution to the solvation free energy calculated by PB or GB, respectively; **PBSOL/GBSOL**: sum of non-polar and polar contribution to solvation; **PBELE/GBELE**: sum of the electrostatic solvation free energy and MM electrostatic energy; **PBTOT/GBTOT**: final estimated binding free energy.

The binding free energy has also been recalculated for all the four complexes using the g_mmpbsa for the MD trajectories obtained from GROMACS MD simulation. It has been observed that binding energy obtained from both types of programs that the PDB Id 1D30 have lowest binding energy. To solve the PB equation g_mmpbsa uses the APBS package whereas mm_pbsa.pl uses the PBSA program of the AMBER suite. It has been observed that the energies calculated using g_mmpbsa and the AMBER MM-PBSA package is approximately similar and the difference of 1-3 kcal/mol has been observed due to the difference in ΔG_{polar} (Table 3.2). The

difference in ΔG_{polar} observed because of different algorithms, implemented in APBS and PBSA.

The binding energy calculation shows that DNA-DAPI complex is more stable than others. So the MD structures of complexes were selected to study the hydrogen bonds which retained throughout the MD simulation. The results show that only few hydrogen bonds that were present in the original complex (energy minimized) were retained during the simulations. In the DNA-DAPI complex two hydrogen bonds between LIG25:H4-DT19:O2 and LIG25:H5-DT20-O4' were retained throughout the simulation using AMBER and GROMACS. It has also been observed that hydrogen bonds are mainly formed between minor groove binders and the functional group of the bases exposed in the grooves via their end groups and also through their amide and linker group.

Table 3.2: Comparison of binding energy components obtained from AMBER MM-PBSA and g_mmpbsa.

PDB Id	program	ΔE_{elec} [a]	ΔE_{vdw} [b]	ΔG_{polar} [c]	$\Delta G_{nonpolar}$ [d]	$\Delta G_{binding}$ [e]
1D30	mm_pbsa.pl	-25.80±6.54	-34.83±3.39	39.27±6.40	-4.15±0.22	-25.52±3.64
	g_mmpbsa	-12.01±5.18	-16.42±7.24	17.84±7.12	-1.71±0.58	-25.95±10.11
1D86	mm_pbsa.pl	-51.80±13.59	-30.43±4.80	65.35±12.44	-4.10±0.33	-20.98±6.58
	g_mmpbsa	-17.60±12.47	-27.36±12.61	25.41±14.16	-2.92±1.16	-22.46±13.18
102D	mm_pbsa.pl	-32.25±10.30	-22.56±3.63	41.63±8.77	-3.76±0.20	-16.94±4.06
	g_mmpbsa	-11.91±4.10	-26.51±5.13	20.65±5.75	-2.89±0.41	-20.66±7.33
195D	mm_pbsa.pl	-26.25±14.81	-27.46±3.69	39.99±14.48	-3.93±0.39	-17.65±4.36
	g_mmpbsa	-26.26±4.31	-34.99±3.61	37.72±6.65	-3.75±0.27	-27.30±4.17

[a]Electrostatic component to the binding energy in kcal/mol.

[b]van der Waals component to the binding energy in kcal/mol.

[c]Polar solvation energy.

[d]Non-polar solvation energy.

[e]Binding energy.

3.4 Conclusions

The computational studies were performed to evaluate and analyze the binding energies of DNA minor groove binders. The focus of the study is to provide a detailed perspective on drug-DNA interactions at the molecular level. Thus, our study attempts to give detail insight on the complexity in binding modes of small molecules with DNA. Molecular dynamics results on comparison with the molecular docking results, showed a similar binding mode. This lends credit to the reliability of the active conformations obtained from the Autodock. The binding affinities obtained from both the methods are also comparable and shows the inhibitory capability of minor groove binders with the

DNA. Later, the analysis from the root mean square deviation for the stability of complex shows that the AMBER force fields are better for the drug-DNA complexes up to 5ns of molecular dynamics simulations. However, longer simulations may give a more detailed view of the difference in the force fields. These analyses can be helpful for the improvement of existing minor groove binders, and also in designing novel chemical entities which can act as a good DNA inhibitor. This study may also provide a theoretical protocol for complementing experimental techniques, generation of database for structure-energy relationship in drug-DNA complexes.

References:

[1] Watson, J.D., Crick, F.H.C., *Nature*, **171**, 737 (1953).

[2] Song, Y.M., Wu, Q., Yang, P.J., Luan, N.N., Wang, L.F., Liu, Y.M., *J. Inorg. Biochem.*, **100**, 1685 (2006).

[3] Dervan, P.B., *Bioorg. Med. Chem.*, **9**, 2215 (2001).

[4] Dervan, P.B., Edelson, B.S., *Curr. Opin. Struct. Biol.*, **13**, 284 (2003).

[5] Masta, A., Gray, P.J., Phillips, D.R., *Nucleic Acids Res*., **23**, 3508 (1995).

[6] Tse, W.C., Boger, D.L., *Chemistry & Biology*, **11**, 1607 (2004).

[7] Nunn, M.C., Neidle, S., *J. Med. Chem.*, **38**, 2317 (1995).

[8] Kennard, O., *Pure & Appl. Chem.*, **65**, 1213 (1993).

[9] Srivastava, H.K., Chourasaia, M., Kumar, D, Sastry, G.N., *J. Chem. Inf. Model.*, **51**, 558 (2011).

[10] Kamal, A., Shetti, R.V., Rmaiah, M.J., Swapna, P., Reddy, K.S., Mallareddy, A., Rao, M.P.N., Chourasia, M., Sastry, G.N., Juvekar, A., Zingde, S., Sarma, P., Pushpavalli, S.N., Bhadra, M.P., *Med. Chem. Comm*., **2**, 780 (2011).

[11] Kamal, A., Shankaraiah, N., Reddy, Ch. R., Prabhakar, S., Markandeya, N., Srivastava, H.K., Sastry, G.N., *Tetrahedron*, **66**, 5498 (2010).

[12] Kamal, A., Reddy, K.S., Khan, M.N.A., Shetti, R.V.C.R.N.C., Ramaiah, M.J., Pushpavalli, S.N.C.V.L., Srinivas, C., Pal-Bhadra, M., Chourasia, M., Sastry, G.N., Juvekar, A., Zingde, S., Barkume, M., *Bioorg. Med. Chem.*, **18**, 4747 (2010).

[13] Kamal, A., Bharathi, E.V., Ramaiah, M.J., Dastagiri, D., Reddy, J.S., Viswanath, A., Sultana, F., Pushpavalli, S.N.C.V.L., Pal-Bhadra, M., Srivastava, H.K., Sastry, G.N., Juvekar, A., Sen, S., Zingde, S., *Bioorg. Med. Chem.*, **18**, 526 (2010).

[14] Kamal, A., Rajender, Reddy D.R., Reddy, M.K., Balakishan, G., Shaik, T.B., Chourasia, M., Sastry, G.N., *Bioorg. Med. Chem.*, **17**, 1557 (2009).

[15] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., *Nucleic Acids Res.*, **28**, 235 (2000).

[16] Shaikh, S.A., Jayaram, B., *J. Med. Chem.*, **50**, 2240 (2007).

[17] Gaussian 09, Revision A.02, Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G.A., Nakatsuji, H., Caricato, M., Li X., Hratchian, H.P., Izmaylov, A.F., Bloino, J., Zheng, G., Sonnenberg, J.L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J.A., Jr., Peralta, J.E., Ogliaro, F., Bearpark, M., Heyd, J.J., Brothers, E., Kudin, K.N., Staroverov, V.N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J.C., Iyengar, S.S., Tomasi, J., Cossi, M., Rega, N., Millam, J.M., Klene, M., Knox, J.E., Cross, J.B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli, C., Ochterski, J.W., Martin, R.L., Morokuma, K., Zakrzewski, V.G., Voth, G.A., Salvador, P., Dannenberg, J.J., Dapprich, S., Daniels, A.D., Farkas, O., Foresman, J.B., Ortiz, J.V., Cioslowski, J., Fox D.J., Gaussian, Inc., Wallingford CT, 2009.

[18] Morris, G., Goodsell, D., Halliday, R., Huey, R., Hart, W., Belew, R., Olson, A.J., *J. Comput. Chem*., **19**, 1639 (1998).

[19] Case, D.A., Berryman, J.T., Betz, R.M., Cerutti, D.S., Cheatham, III T.E., Darden, T.A., Duke, R.E., Giese, T.J., Gohlke, H., Goetz, A.W., Homeyer, N., Izadi, S.,

Janowski, P., Kaus, J., Kovalenko, A., Lee, T.S., LeGrand, S., Li, P., Luchko, T., Luo, R., Madej, B., Merz, K.M., Monard, G., Needham, P., Nguyen, H., Nguyen, H.T., Omelyan, I., Onufriev, A., Roe, D.R., Roitberg, A., Salomon-Ferrer, R., Simmerling, C.L., Smith, W., Swails, J., Walker, R.C., Wang, J., Wolf, R.M., Wu, X., York, D.M., Kollman, P.A., AMBER 2015, University of California, San Francisco (2015).

[20] Salomon-Ferrer, R., Case, D.A., Walker, R.C., *WIREs Comput. Mol. Sci.*, **3**, 198 (2013).

[21] Case, D.A., Cheatham, T.E. III, Darden, T., Gohlke, H., Luo, R., Merz, K.M., Jr., Onufriev, A., Simmerling, C.,Wang, B., Woods, R., *J. Computat. Chem*., **26**, 1668 (2005).

[22] Cheatham, T.E., III, Case, D.A., *Biopolymers*, **99**, 969 (2013).

[23] Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., van der Spoel, D., Hes, B., Lindahl, E., *Bioinformatics*, **29**, 845 (2013).

[24] Thompsan, J.J., Lill, M.A., *J. Chem. Inf. and Model*., **51**, 2680 (2011).

[25] Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., uvench, O.G., Lopes, P., Vorobyov, I., MacKerell, Jr. A.D., *J. Comput. Chem.*, **31**, 671 (2010).

[26] Zoete, V., Cuendet, M.A., Grosdidier, A., Michielin, O., *J. Comput. Chem*., **32**, 2359 (2011).

[27] Kumari et al *J. Chem. Inf. Model.*, **54**, 1951(2014).

Chapter 4

Illustrating Binding Mechanism of DNA intercalators using Computational Approaches

Chapter 4

Illustrating Binding Mechanism of DNA intercalators using Computational Approaches

4.1 Introduction

DNA is a biomolecule that acts as an information hub for the living beings which carries all the genetic information of the forthcoming generation. DNA controls various cellular processes such as, protein and enzyme synthesis and also an important target for the drug delivery. Thus, it is a target for many of the drug molecules which are used in chemotherapy.[1-4] In order to improve the clinical efficiency of available drugs and to design novel and potent drugs, it is important to study the molecular basis of drug-DNA interactions along with their thermodynamic, kinetic and structural details.[5-8] Ligands can bind with the nucleic acid via two modes of binding; namely, groove binding and intercalation. Both covalent and non-covalent types of interactions are possible for these binding modes. In nucleic acid structure, two strands of DNA winds together to form two types of grooves namely, minor groove and major groove which differs in their size and binding elements.[9-11] Proteins and large molecules generally bind to the major groove while small molecules can bind to both types of grooves.[12, 13] However, small molecules generally prefer to bind with the minor groove of the DNA. Minor groove binders are usually crescent in shape and symphonize the shape of the groove. Minor groove binders usually bind to the AT-rich region of the nucleic acid.[14-18] Another

common binding mode of ligands with nucleic acid is intercalation, which involves the insertion of these molecules between the nucleic acid base pairs. This type of insertion requires an opening gap between the nucleic acid base pairs. Intercalation, first explained by Lerman, in which the ligand is held rigidly perpendicular to the DNA backbone without breaking up the hydrogen bonding between the nucleic bases.[19-21] This causes the distortion of the sugar phosphate backbone and also decreases the helical pitch in DNA. Intercalators usually bind to the GC intercalation site of DNA, and are generally aromatic, planar and polycyclic in nature.[22-24] Due to the presence of aromatic ring in planar molecule, π-π interactions are observed between the two flanking nucleic acid bases. These types of interactions play an important role in the intercalation binding mode, while other interactions such as van der Waals, hydrogen bonding, electrostatic and hydrophobic also contributes to the formation of DNA-intercalators complexes. The intercalated DNA becomes geometrically distorted that results in decoding of genetic information and structural changes in DNA which can affect the cellular processes such as inhibition of transcription, and translation.[25-27] Echinomycin, nogalamycin, triostin A, acridine, cis-Platin, adriamycin, ethidium bromide, propidium, actinomycin D, are some examples of the DNA intercalating agents.[28-36] Acridines are an important class of antitumor agents of threading type of intercalators. In threading type of intercalators, ligand interacts with both major and minor grooves. Two important acridine-based compounds N-[2-(dimethylamino)-ethyl]-acridine-4-carboxamide (DACA) and N-[2-(dimethylamino)-ethyl]-9-aminoacridine-4-carboxamide (AAC) are used in the treatment of cancer. In spite of structural similarity these two drug molecules, DACA and AAC belong to different classes of drugs. Since there is a difference between the structures of

these two compounds, due to the presence of amino substituent at the C9 position, compelling changes in the electronic properties can be raised.[37-39] Mitoxantrone is another drug intercalator which structurally belongs to the anthracycline antibiotic group. This drug molecule is generally used for the treatment of different types of cancers including metastatic breast cancer, acute myeloid leukemia, non-small cell lung cancer, non-Hodgkin's lymphoma and prostate cancer. In the context of many studies, it is yet not admitted that which type of conformational changes arise in the DNA upon the interaction of mitoxantrone.[40-44] Therefore, it is important to investigate the drug-DNA interaction in the view of biological significance. The emodin (1, 3, 8-trihydroxy-6-methylanthracene-9, 10-Dione) is an anthraquinone which was derived from a traditional Chinese medicinal plant 'rhizome of rhurab'; used for its antitumor, immunosuppressive, anti-inflammatory and antibacterial properties.[45,46] Ordinarily aromatic molecules having three or four fused rings are the possible choices for the DNA intercalators. The experimental results clearly indicate that the classical mode of binding is ruled out in the case of benzidine. Benzidine has a biphenyl ring that weakly interacts with DNA duplex as a partial intercalator.[47] In the present work, the interaction studies on DNA intercalators, DACA, AAC, mitoxantrone, emodin and benzidine has been performed with the nucleic acid using various molecular modelling approaches. The experimental binding constant of these compounds with DNA has been collected from the literature for comparison with computational results and provide insight on how the various interactions affects binding of these compounds.[38, 43, 46, and 47] Our study examines the available popular molecular modeling approaches for the estimation of DNA-ligand binding free energies. Molecular modeling methods are powerful tools to investigate

various types of non-covalent interactions and binding energies between the receptor and ligand. In a number of studies for the minor groove binders and intercalators, computational methods have shown good agreement with the experimental results.[48-51] Molecular dynamics simulation, used to investigate the conformations of biological macromolecules with time. Various force fields used in the nucleic acid simulation includes, CHARMM, AMBER, GROMOS, OPLS, ENCAD and BMS26. In the present work, Molecular Mechanics/Generalized Born Surface Area (MMGBSA) and Molecular Mechanics/Poisson−Boltzmann Surface Area (MMPBSA) methods are used for the free energy estimation using MD trajectories. The analysis of drug-DNA interactions, sequence selectivity and specificity, and the mechanism of drug-DNA interaction have been investigated using Quantum Mechanics/Molecular Mechanics method (QM/MM). Other computational techniques such as Docking and Molecular Dynamics (MD) have been used for predicting binding of the drug to DNA and analyze conformational stability of bound inhibitor respectively. Nowadays, combined QM/MM approaches have become the method of choice for the modeling of reaction mechanism in biomolecular systems at a reasonable computational cost with accuracy. With this method, several snapshots of the reaction can be generated which help to get an insight into the mechanisms which were not understood so far. The detailed investigation of interactions between DNA-intercalators and their role in binding may provide more successful drug candidates that can target specific disease.

4.2 Materials and Methods

The experimental binding constant K of five DNA intercalators DACA, AAC, mitoxantrone, emodin and Benzedine were collected from the literature (Table 4.1).

Further, the binding constants obtained from the literature were correlated with the binding free energy obtained from computational methods. Basic equation for the calculation of binding constant and Gibbs free energy is

$$\Delta G = -RT \ln K$$

where ΔG is free energy, R is a gas constant (1.98 cal /mol K), T is the temperature at which the experiment was performed (300K) and K is binding constant between drug intercalators and free DNA.

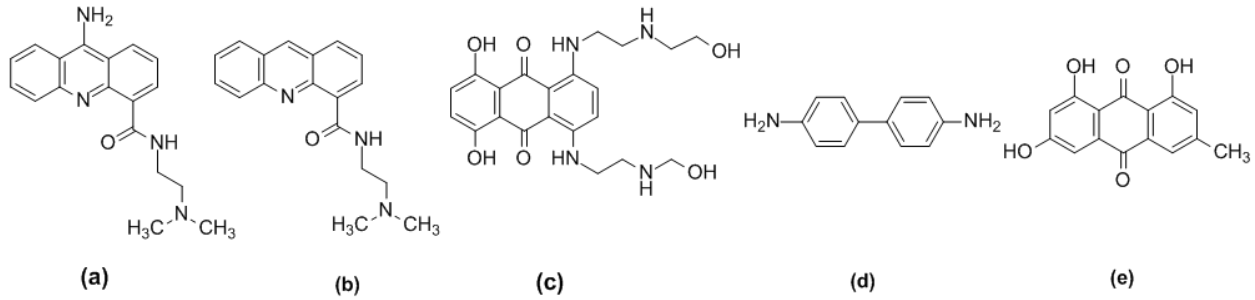


Fig. 4.1: Chemical structure of ligands (a) AAC (b) DACA (c) Mitoxantrone (d) Benzidine (e) Emodin.

Table 4.1: List of intercalating drugs used for the study with their chemical structure, number of heavy atoms in drug, experimental binding const. K, and experimental binding energy ΔGbind collected from the literature with references.

S. No.	Drug	Exp. Binding Const. K (M^{-1})	Exp. ΔG_{bind} (kcal/mol)
1	Mitoxantrone [47]	3.88×10^{5}	-7.64
2	AAC [38]	4.1×105	-7.68
3	DACA [38]	7.4×10^{5}	-8.03
4	Emodin [43]	5.59×10^{3}	-5.12
5	Benzidine [46]	1.3×10^{3}	-4.26

The present investigation was carried out using biological databases such as Protein Data Bank,[52] Nucleic Acid Database [53,54] and various software's such as Gaussian 09, [55] Chimera,[56] Visual Molecular Dynamics (VMD),[57] Autodock 1.5.6,[58] Discovery Studio,[59] GROMACS 4.5.6,[60] AMBER 15, [61] and g_mmpbsa program.[62]

All the structures of DNA intercalating drugs (Fig. 4.1) were optimized at B3LYP/6-31G* level using Gaussian 09 software. The optimized ligands were further used for docking studies.

4.2.1 Molecular Docking Studies

MGL tools 1.5.6 with AutoGrid4 and AutoDock4 were used to perform blind docking calculations. For the docking studies, the DNA fragments bounded with intercalators were searched from protein databank and the DNA from the crystal structure having PDB Id,1DSC was selected for further investigation.[63] The original ligand was removed and the DNA (5'-GAAGCTTC-3') was used as a receptor for docking studies. Initially, all the heteroatom's including water molecules were deleted from the DNA structure. Receptor (DNA) and ligand files were prepared and the Gasteiger charges were added to the complex by Autodock Tools before performing docking calculations. The DNA was enclosed in a box with the grid points 70×70×70 in x × y × z directions with grid spacing of 0.375Å. Lamarckian genetic algorithms were used for the docking calculations and all other parameters were set to default. The 20 LGA runs with maximum of 25,00000 energy evaluations were performed. For each docked case, the lowest energy conformation, according to the autodock scoring function, was considered as the best

binding mode. The DNA-ligand complexes obtained from Autodock were analyzed using Discovery Studio visualizer.

4.2.2 Molecular Dynamics Simulation

Molecular Dynamics Simulation (MD) of 5ns was carried out using GROMACS 4.5.6 package and AMBER 15 for the selected docked poses. In GROMACS, the topology and co-ordinate files for the DNA were generated by pdb2gmx program of the GROMACS package taking the parameters from the Amber 03 force field while for ligand these files were generated using 'acpype.py' python script. The coordinate and topology files of DNA and ligands were merged to obtain the final structure and topology file for each complex. The simulation system was constructed by immersing the macromolecules in a dodecahedron of explicit TIP3P water molecules. The simulation system was neutralized by the addition of excess positively charged counter-ions. The system was then made to acquire the aqueous ionic environment by applying 50,000 cycles of steepest descent algorithm. Then the system was heated to 300K during 50 ps of constant volume simulation with 2 fs time step. The pressure was equilibrated to 1 atm during 50 ps NPT simulation with 2 fs time step. In both the simulations, a restrained of force constant (of 1000kJ/ (mol/nm^2) was applied. Both temperature and pressure were regulated using Berendsen algorithm. To compute MD trajectories, unrestrained – production level MD simulations were performed for 5ns with a time step of 2 fs. Temperature and pressure were sustained at 300 K and 1 atm using the v-rescale temperature and Parninello-Rahman Algorithm. Long range electrostatics effects were calculated using Particle-Mesh-Ewald algorithm (PME). The integration time step applied was 2.0 fs and the co-

ordinates were saved at every 2.0 ps. Trajectories were visualized by means of VMD. All the graphs were plotted using GNUPLOT.

Further, MD simulations of the selected docked poses were also performed using AMBER 15. The 'leaprc.gaff" (generalized amber force field) was used to prepare the ligands, while the 'leaprc.ff03' was used for the preparation of DNA. The 'add ions' command implemented in 'tleap' of AMBER15 was used to add the Na+ ions explicitly to neutralize the system. Each system was placed in a box of TIP3P water by using 'solvateOct' command with the minimum distance between any solute atom and the boundary of the box was set to 8Å. Energy minimization was then performed to achieve the nearest stable low energy conformations (500 steps of each steepest descent and conjugate gradient method), 50 ps of heating and 50 ps of density equilibration with weak restraints on the complex followed by 500 ps of constant pressure equilibration at 300K. Cut off size of 12 Å was used for MD simulations. All long –range electrostatics were included by means of a Particle Mesh Ewald (PME) method. All hydrogen and heavy atom bonds were constrained by the shake method and simulations were performed with a 2 fs time step and langvenin dynamics was used for temperature control. The same conditions for the final phase of equilibration were used for the production run and the coordinates were recorded at every 10 ps. Periodic boundary conditions were used for the final production run.

4.2.3 Binding Energy Calculations

In order to estimate the relative binding energy between drug and DNA, g_mmpbsa method in case of GROMACS and mm_pbsa.pl script in case of AMBER were used. In case of GROMACS, the average binding energies were obtained from 5000 snapshots

extracted at every 2 ps from the MD trajectories between 0 ns to 5 ns. The contribution of non-polar solvation energy was analyzed based on SASA model. This program can calculate molecular mechanics potential energy (electrostatic and van der Waals energies) and free energy of solvation (polar and non-polar solvation energies). The entropy contribution was not included to the final binding energy.

In case of AMBER, 5,000 snapshots of the complex are obtained at every 10 ps from the MD trajectories, and all the water molecules and ions were removed before MMPBSA/MMGBSA calculations using the "extract_coords.mmpbsa" script. The $\Delta G_{bind} - PB/GB$ values were calculated using the "binding_energy.mmpbsa" script.

The binding energy is calculated by MMPBSA method only in case of MD trajectories obtained from GROMACS, while for AMBER the binding energies between DNA and ligands are calculated from both MMPBSA and MMGBSA method.

4.2.4 QM/MM calculations

Snapshots obtained at each nanosecond from the MD simulation for both GROMACS and AMBER, were selected as starting geometries for the QM/MM calculations. These structures contain the bimolecular system in a droplet of water (20000-30000 atoms) and this setup requires a lot of prior work to avoid errors and wrong choices for the actual QM/MM calculations. The QM/MM calculations were performed with the two-level ONIOM method within the Gaussian 09 program suite. The two layered ONIOM scheme is shown in Fig. 4.2. B3LYP method along with basis set 6-31+G is used for the high layer and AMBER force field for the low layer and all the water molecules and Na^+ ions are frozen. Geometry optimization of the above system was performed and the energy of the high or QM region is calculated. Now, the co-ordinate of QM region from the

optimized system is extracted, which is further subjected to geometry optimization and the energy of QM region in the gas phase is calculated by the same method. The interaction energy between the DNA and ligand is calculated with the help of formula-

$$\Delta E_{I.E} = E_{QM_{(g,p)}} - E_{QM_{(p,p)}}$$

where, $\Delta E_{I.E}$ is the interaction energy between drug and DNA. $E_{QM\ (p,\ p)}$ is the energy of QM region in nucleic acid phase and $E_{QM\ (g,\ p)}$ is the energy of QM region in the gas phase.

The single point energy of the optimized geometry was calculated using the basis set with dispersion correction M062X for the QM region and AMBER force field for the MM region. The single point energy of QM region in the gas phase with the same basis set was also calculated. The interaction energy between the DNA and ligand was calculated using above mentioned formula.

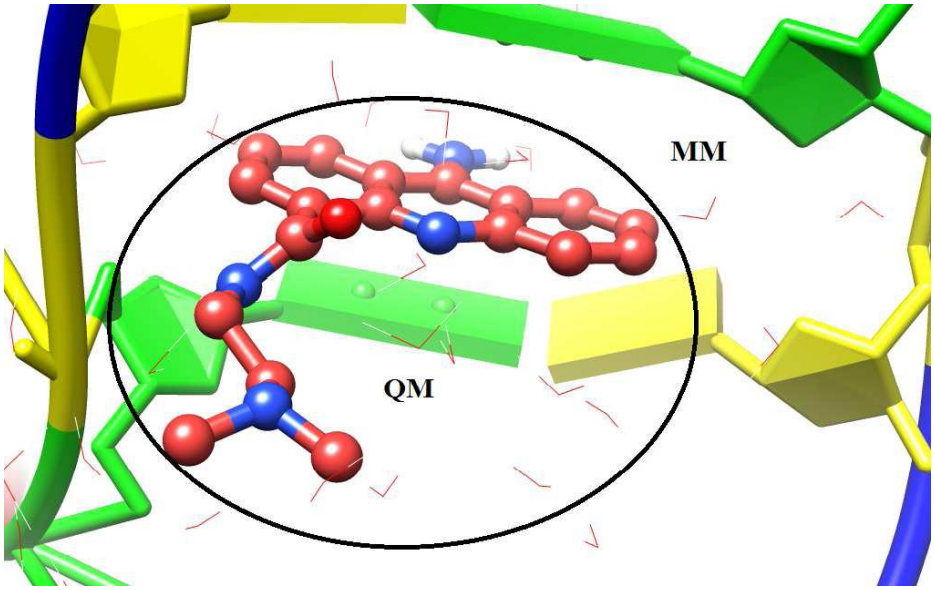


Fig. 4.2: Two Layered ONIOM Scheme.

4.3 Result and Discussion

4.3.1 Molecular Docking Results

The conformation with minimum energy is taken into consideration among the 20 energy conformers. From the docking studies, it was found that the drug molecules inserted between the nucleic acid bases or intercalation site of nucleic acid had van der Waal forces with the DNA duplex. All the drug molecules intercalate between the G-C pairs of DNA. The docked poses are stabilized by π–π stacking and by the formation of hydrogen-bonds between the DNA base pairs and functional group (Fig 4.3). The calculated binding free energies (ΔG) obtained from docking studies for the all DNA-inhibitor complexes in this study are listed in Table 4.2.

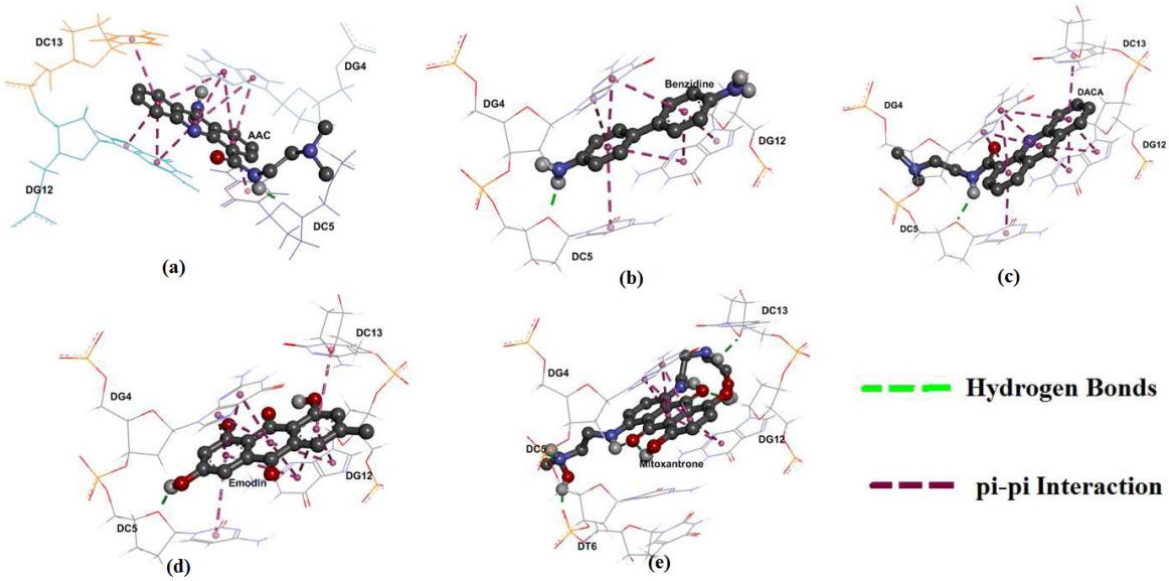


Fig. 4.3: Different types of non-covalent Interactions obtained from Molecular Docking. Hydrogen bonds are indicated by dotted lines and π–π interactions by purple dotted lines for the ligand (a) AAC (b) Benzidine (c) DACA (d) Emodin (e) Mitoxantrone.

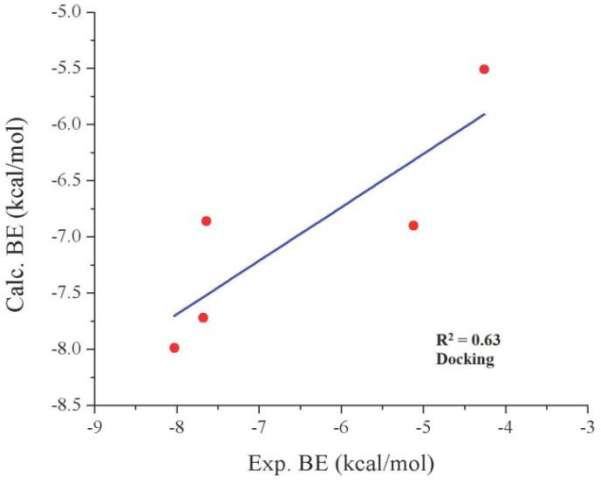


Fig. 4.4: Variation of calculated binding energy with experimental binding energy (R^2 = 0.63).

The graph (Fig. 4.4) indicates that there is a linear correlation (R^2=0.63) between the calculated binding free energy (ΔG_{bind}) and experimental binding free energy obtained from molecular docking calculations. The docking result seems therefore in agreement with the experimental results. The DNA-benzidine complex has the highest experimental ($\Delta G_{bindvalue}$) and calculated binding energy ($\Delta G_{bindvalue}$) because chemical structure of benzidine has an unfused biphenyl ring and the complex with mitoxantrone has lowest binding energy, thus, this complex is more stable than other complexes under study.

4.3.2 Molecular Dynamics Results

To elucidate the conformational stability of DNA-inhibitor complexes, the most stable intercalative binding mode obtained from molecular docking for each system were taken as starting points for molecular dynamics simulation. All the molecular dynamics simulations were carried out with GROMACS and AMBER as discussed in the method section. The analysis of RMSD, comparing DNA without ligand, free ligand and the intercalated bound complexes are shown in Fig. 4.5 and Fig. 4.6. The ligands bound to

the intercalation site leads to structural fluctuations in the beginning of the simulation. This behavior indicates that the intercalated ligand opens up the gap between nucleic acid bases because of π-π interactions. The π-π interaction was obtained between the DNA residues with Id DG4, DC5, DG12 and DC13. The intercalating ligands AAC, DACA, benzidine, emodin and mitoxantrone found hydrogen bonds with Guanine (DG 12), Cytosine and Guanine (DC5, DG5), Cytosine (DC13, DC5), Cytosine (DC5) and Cytosine and Guanine (DG4, DC15) base pairs of DNA respectively and hydrogen bonds also formed with the solvent. These interaction show that intercalators form non-covalent interactions with CG base pairs of DNA.

In case of MD simulation from GROMACS, the range of RMSD is 0.1-0.4 nm for the entire DNA-ligand complexes and free DNA, while the RMSD variation of free ligands for AAC, DACA, and mitoxantrone shows higher RMSD variation (0.2 nm) and another two ligands emodin and benzidine shows RMSD variation less than 0.1 nm. It is clear from Fig. 4.5 and Fig. 4.6 that lowest experimental binding energy complexes show lowest deviation in RMSD. The RMSD of free ligand is always lower than the RMSD of free DNA and DNA-ligand complexes. In the case of AMBER MD simulation, the range of RMSD variation is 0.1-0.45 nm for all the DNA-Ligand complexes, while the RMSD variation of free ligand is always lower than that of DNA and DNA-Ligand complexes. The variation in RMSD obtained from AMBER and GROMACS MD simulation are almost identical. The Fig. 4.7 and Fig. 4.8 shows the snapshots of drug-DNA complexes at 5 ns, which shows that the ligand remains, binds at the intercalation site of DNA up to the end of simulation from AMBER and GROMACS.

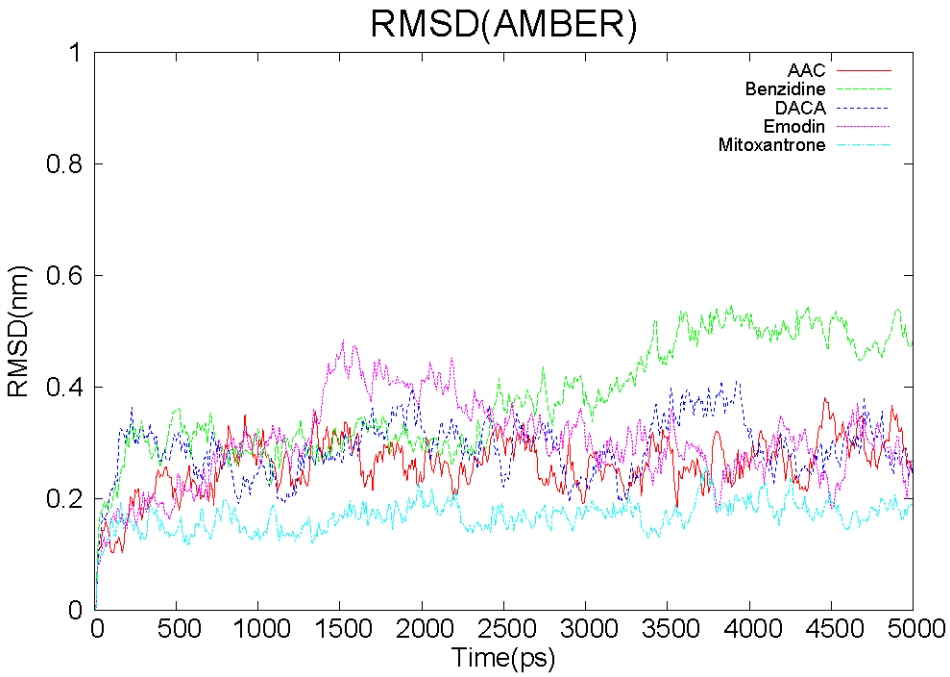


Fig. 4.5: RMSD plots for DNA-ligand complexes (AMBER).

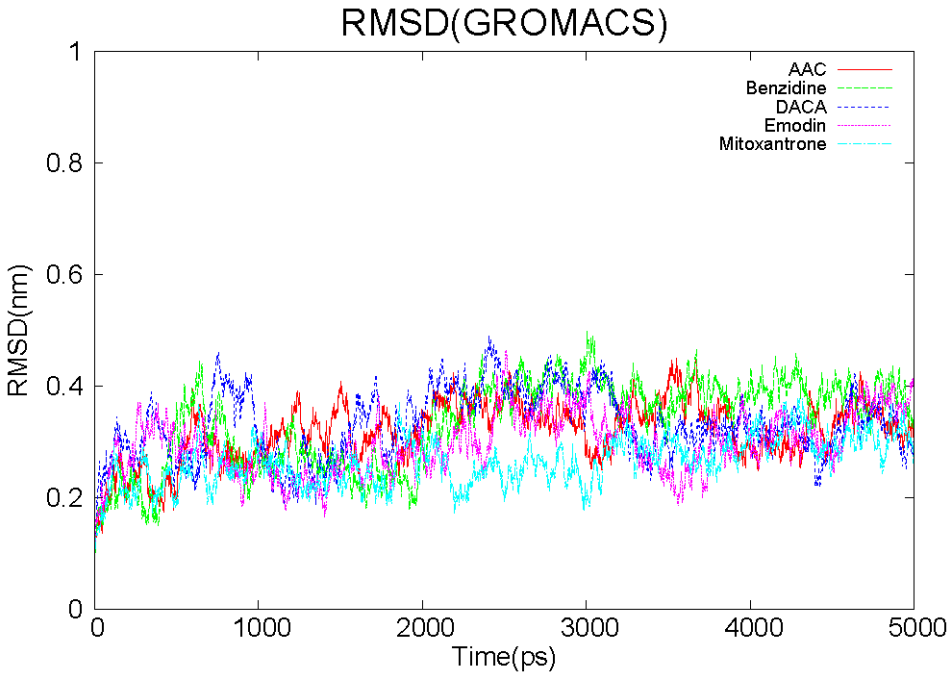


Fig. 4.6: RMSD plots for complex, DNA and ligand (GROMACS).

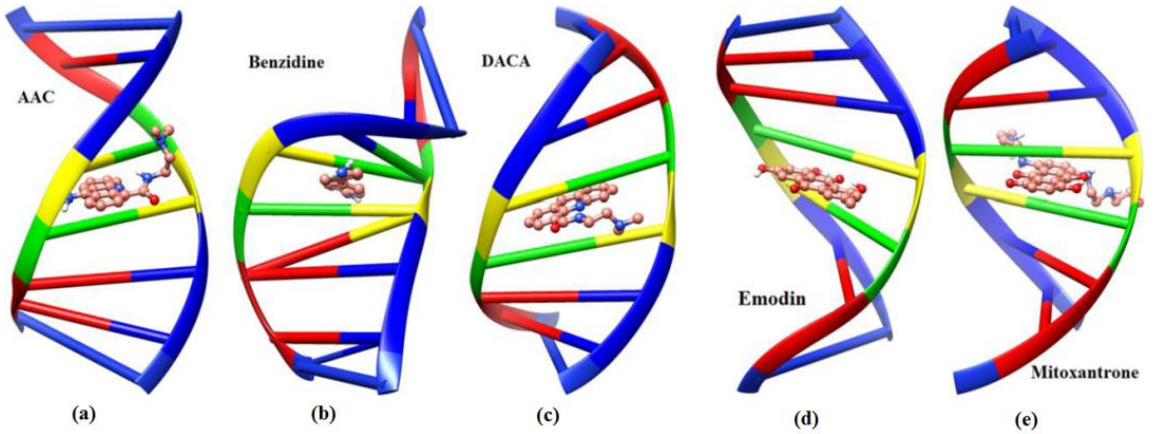


Fig. 4.7: Snapshots of all DNA-ligand complexes at 5ns MD simulation (AMBER) for ligands (a) AAC (b) Benzidine (c) DACA (d) Emodin (e) Mitoxantrone.

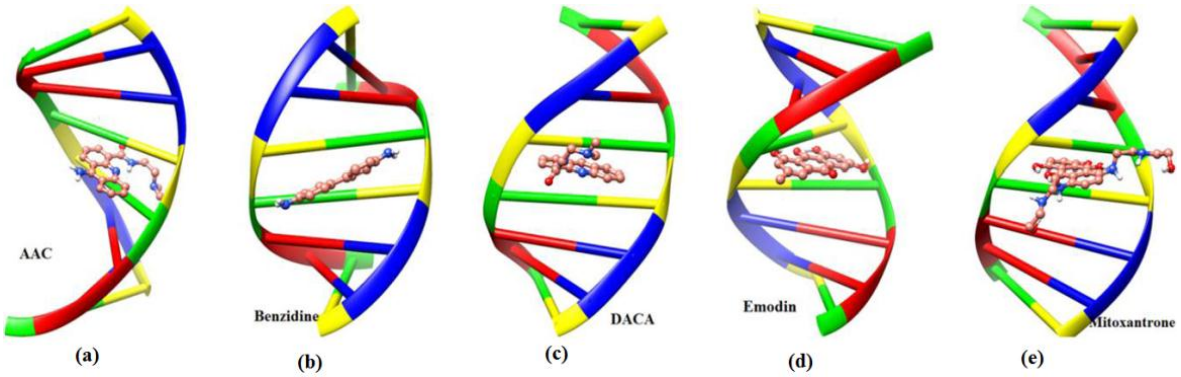


Fig. 4.8: Snapshots of all DNA-ligand complexes at 5ns MD simulation (GROMACS) for ligands (a) AAC (b) Benzidine (c)DACA (d) Emodin (e) Mitoxantrone.

Table 4.2: Comparison of experimental and calculated ΔG_{bind} obtained from molecular docking studies and MMPBSA/MMGBSA calculations using AMBER and GROMACS.

Ligand	No. of heavy Atoms in Ligand	Exp. ΔG_{bind}	Calc. ΔG_{bind} (Docking)	Calc. ΔG_{bind} (MD)		
				GROMACS	AMBER	
				MMPBSA	MMPBSA	MMGBSA
Mitoxantrone	32	-7.64	-6.86	-46.66±4.36	-32.39±2.86	-27.25±2.30
AAC	23	-7.68	-7.72	-38.67±3.16	-29.07±2.55	-32.73±3.23
DACA	22	-8.03	-7.99	-35.80±3.63	-30.54±2.64	-31.55±2.59
Emodin	20	-5.12	-6.90	-35.18±2.55	-25.85±2.99	-24.35±2.36
Benzidine	14	-4.26	-5.51	-19.12±4.96	-16.51±2.61	-17.61±2.94

All the energies are in kcal/mol.

4.3.3 Binding Energy calculations

The MMPBSA/MMGBSA calculations were performed by using MD trajectories to obtain Binding Energy (BE) values, the strength of interaction between DNA and ligand can be measured from BE values, which are presented in Table 4.2. In the view of binding energy calculations obtained from MMPBSA method (GROMACS), it was analyzed that lower experimental binding energy value has lower calculated binding energy value and vice versa and the plot between experimental binding energy and calculated binding energy values shows linear correlation (R^2 = 0.24). The plot between the number of heavy atoms versus calculated binding energy also has a linear correlation (R^2 = 0.81), and indicates that with the increase in the size of the drug, the binding energy of drug-DNA complexes also increases. From the above discussion and Fig 4.9, it is clear that the BE values obtained from MMPBSA calculations linearly correlates with the

number of heavy atoms in each drug molecule better than the experimental binding energy values.

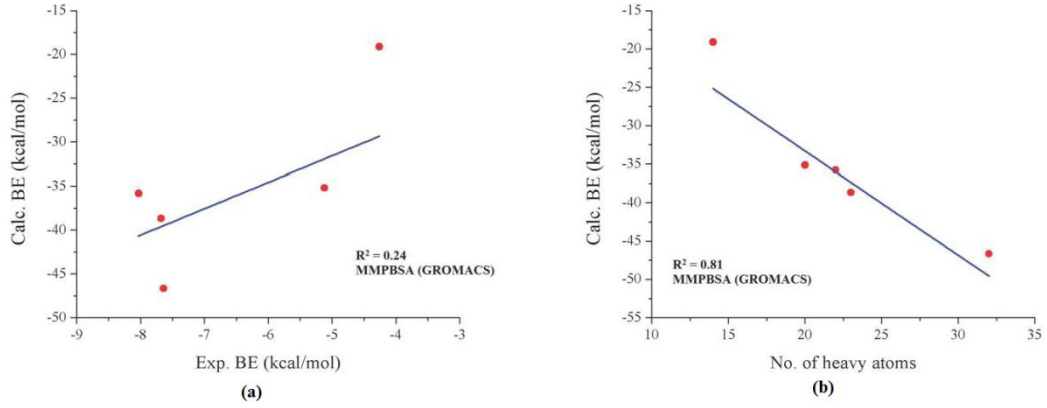


Fig. 4.9: Variation of calculated binding energy obtained from MMPBSA with respect to (a) experimental binding energy (b) No. of heavy atoms in each drug molecule (GROMACS).

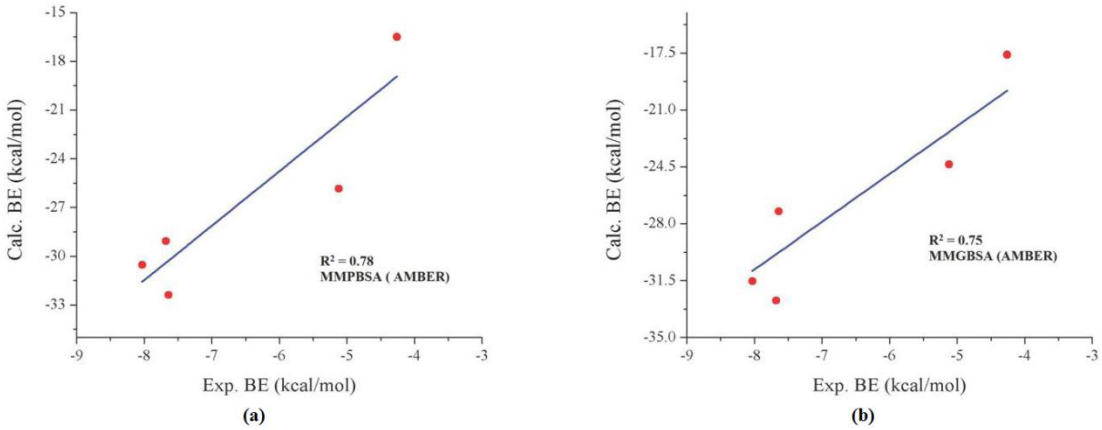
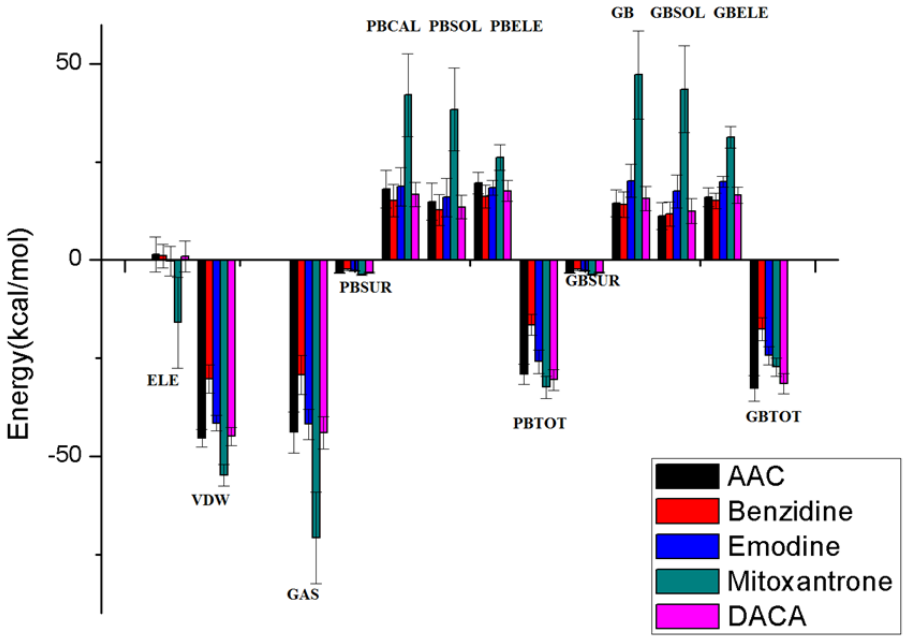


Fig. 4.10: Variation of calculated binding energy obtained from (a) MMPBSA (b) MMGBSA calculations using AMBER with experimental binding energy.

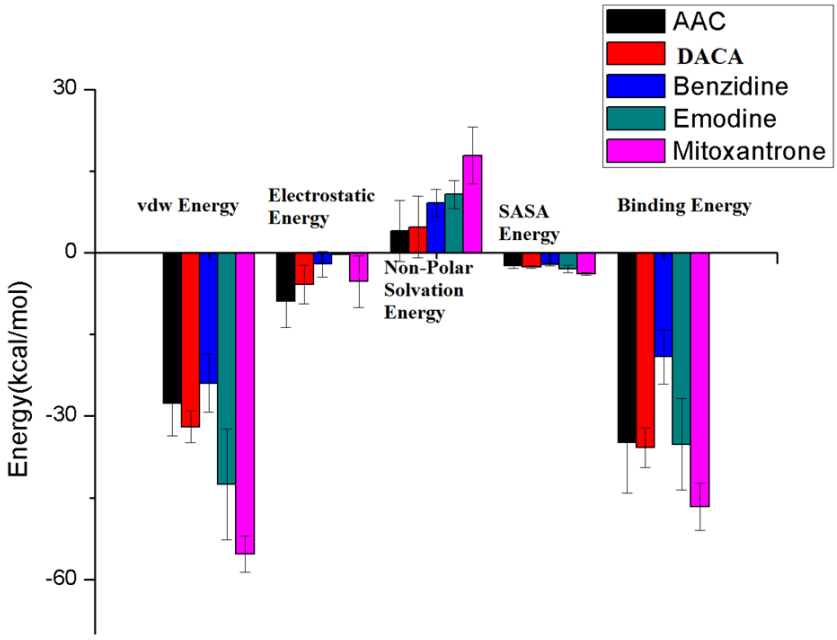
In the case of AMBER, the binding energy values were calculated using MMPBSA and MMGBSA method which shows good correlation with experimental values with R^2 = 0.78 for MMPBSA and R^2 = 0.75 for MMGBSA methods as shown in Fig. 4.10. The

binding energy values obtained from AMBER MD simulation better correlates with experimental values than GROMACS, this is due to the difference in ΔG_{polar} observed which is because of different algorithms, implemented in APBS and PBSA.

Fig 4.11 shows the contribution of each component of energy to the final binding energy, which clearly indicates that the better van der Waals interaction leads to the better binding of the drug to DNA and the electrostatic energy does not contribute much to the final binding energy. The binding energy of the complex with drug benzidine is higher because benzidine has unfused biphenyl rings while other intercalators have fused rings.



(a)



(b)

Fig. 4.11: Histogram depicting view of the contribution of various energy components to the binding free energy (a) GROMACS (b) AMBER.

4.3.4 QM/MM calculations

Six snapshots at each nanosecond (0ns, 1ns, 2ns, 3ns, 4ns, and 5ns) from the MD simulation trajectories from GROMACS were taken and BE values were calculated using earlier discussed ONIOM scheme by employing two different methods namely B3LYP and M062X for QM region, which are presented in Table 4.3 and Table 4.4 for all the studied DNA intercalators. The optimized structure of drug mitoxantrone is shown in Fig. 4.12. The average of these binding energy values is compared with the experimental binding energy values. The experimental binding energy values shows good correlation with calculated binding energy values for both method B3LYP (R^2 = 0.95) and M062X (R^2 = 0.96) in Fig. 4.13. This plot also shows that the lowest experimental binding energy has lowest calculated binding energy and vice versa. Further, Snapshots at each

nanosecond were obtained from the MD trajectories using AMBER MD simulation. The correlation plot between calculated binding energy and experimental binding energy shows good linear correlation with correlation coefficient R^2 = 0.72 for the B3LYP and R^2 = 0.96 for M062X shown in Fig. 4.14.

From the above discussion and figures, it is clear that the binding energy values obtained from the QM/MM calculations complements more with experimental data than that of the molecular docking calculations and MMPBSA calculations.

Table 4.3: Calculated binding energy from QM/MM method using basis set B3LYP for QM region with the help of snapshots obtained from GROMACS MD simulation trajectories at each nanosecond.

Snapshots	AAC		DACA		Benzidine		Emodin		Mitoxantrone	
	B3LYP	M062X	B3LYP	M062X	B3LYP	M062X	B3LYP	M062X	B3LYP	M062X
0ns	-15.15	-12.87	-13.31	-12.10	-0.88	0.004	-0.28	-0.10	-8.53	-7.64
1ns	-3.02	-3.17	-12.66	-12.22	-0.32	-0.19	-0.28	-0.56	-3.90	-5.10
2ns	-2.20	-2.92	-1.40	-9.45	-0.31	-0.08	-0.52	-0.86	-10.45	-14.80
3ns	-5.18	-6.85	-1.67	-1.52	-0.42	-0.15	-0.31	-0.55	-7.61	-8.81
4ns	-4.64	-7.45	-6.59	-7.06	-0.31	-0.04	-0.69	-1.48	-2.20	-3.76
5ns	-5.18	-7.16	-5.78	-5.78	-1.11	-0.28	-0.92	-0.92	-12.29	-14.07

All the energies are in kcal/mol.

Table 4.4: Calculated binding energy from QM/MM method using basis set B3LYP for QM region with the help of snapshots obtained from AMBER MD simulation trajectories at each nanosecond.

Snapshots	AAC		DACA		Benzidine		Emodin		Mitoxantrone	
	B3LYP	M062X	B3LYP	M062X	B3LYP	M062X	B3LYP	M062X	B3LYP	M062X
0ns	-4.32	-3.19	-13.09	-11.09	-0.93	-0.24	-0.27	-0.79	-1.54	-2.22
1ns	-1.70	-1.41	-0.96	-1.43	-1.02	-0.25	-0.50	-1.02	-2.90	-4.12
2ns	-1.61	-2.46	-1.05	-0.63	-0.10	-0.06	-0.17	-0.58	-2.05	-2.12
3ns	-0.28	0.33	-0.25	0.19	-0.21	0.15	-0.11	-0.41	-2.69	-1.70
4ns	-0.20	0.15	-2.02	-0.39	-1.12	-0.30	-0.22	-0.52	-2.68	-1.70
5ns	-2.30	-3.83	-0.72	-0.87	-0.60	-0.04	-0.07	-0.46	-1.30	-2.30

All the energies are in kcal/mol.

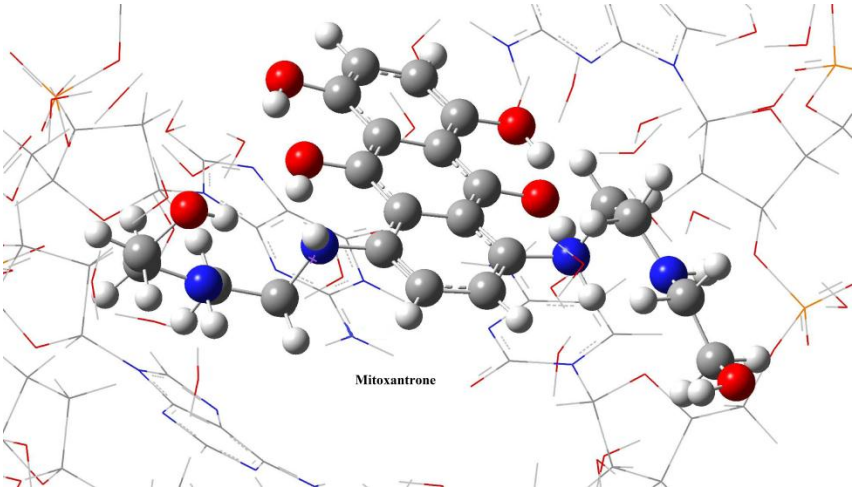


Fig. 4.12: Optimized structure of ligand mitoxantrone using ONIOM method.

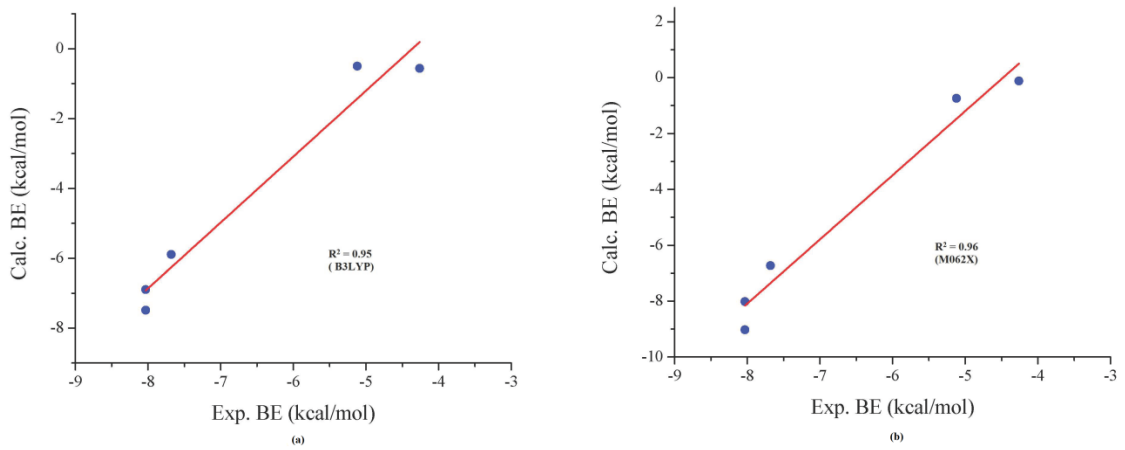


Fig. 4.13: Variation of calculated binding energy with experimental binding energy from QM/MM calculations using (a) B3LYP (R^2 = 0.95) (b) M062X (R^2 = 0.96) as a basis set for the QM region (GROMACS).

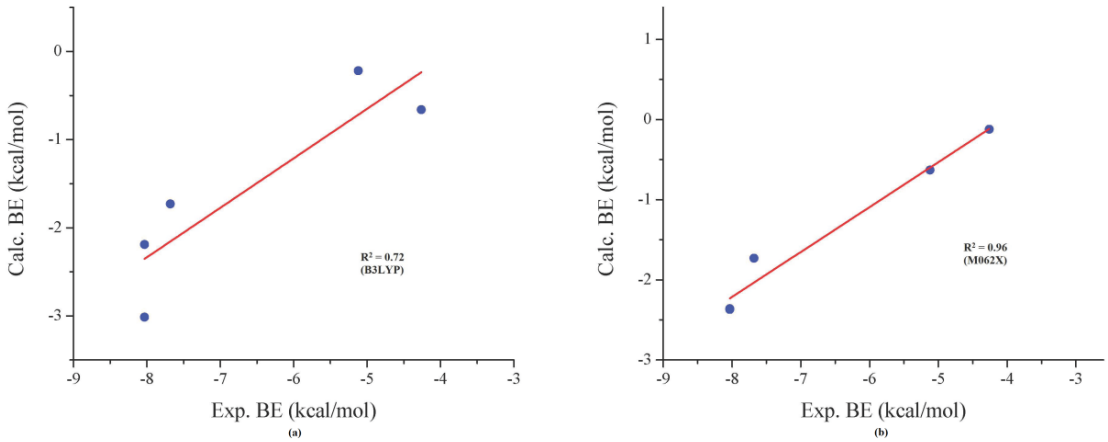


Fig. 4.14: Variation of calculated binding energy with experimental binding energy from QM/MM calculations using (a) B3LYP (R^2 = 0.72) (b) M062X (R^2 = 0.96) as a basis set for the QM region (AMBER).

4.4 Conclusion

The theoretical studies between DNA and ligand via intercalation mode show that DNA intercalators generally bind between the CG regions of nucleic acid via π-π interaction.

The molecular docking results linearly correlate with the experimental results. Both molecular dynamics results and molecular docking results validate the similar binding mode. This lends credit to the reliability of the active conformations obtained from the Autodock. In the case of GROMACS MD simulation, the binding energies obtained from MMPBSA calculations shows that the contribution of van der Waals interaction is greater than that of electrostatic to the binding energy and are not in good correlation with experimental binding energy values. While, in the case of AMBER MD simulation, the binding energy values calculated from MMPBSA/MMGBSA method are in good correlation with experimental binding energy. QM/MM calculations also confirm the binding of a ligand with DNA through the same intercalation site as predicted from docking and MD simulations. The binding energies obtained from the QM/MM calculations are in good correlation with experimental results and the QM/MM calculations provide much better results than that of other molecular modeling methods such as molecular docking and molecular dynamics results. Therefore, QM/MM study provides better insight on the complexity in binding modes of small molecules to DNA. The use of MD simulations approach with QM/MM calculations allowed us to provide a theoretical protocol for complementing experimental techniques, and intercalation disrupts the structure of the DNA helix which in turn inhibits the cellular functions performed by the DNA and lead to the death of cancerous cells. Thus, these molecular modeling studies performed at electronic structure level can give a great insight on designing inhibitors for cancer and microbial infections.

References

[1] Watson, J.D., Crick, F.H.C, *Nature*, **171**, 737 (1953).

[2] Dervan, P.B., *Bioorg. Med. Chem.*, **9**, 2215 (2001).

[3] Boer, D.R., Canals, A., mColl, M., *Dalton Trans.*, **3**, 399 (2009).

[4] Gurova, K., *Future Oncol.*, **5**, 1685 (2009).

[5] Reyes, C.M., Kollman, P.A, *J. Mol. Biol.*, **297**, 1145 (2000).

[6] Shaikh, S.A., Ahmed, S.R., Jayaram B., *Arch. Biochem. Biophys.*, **429**, 81 (2004).

[7] Charies, J.B., *Arch. Biochem. Biophy.*, **453**, 26 (2006).

[8] Paul, A., Bhattacharya, S., *Curr. Sci.*, **102**, 212 (2012).

[9] Charies, J.B., *Curr. Opin. Struc. Biol.*, **8**, 314 (1998).

[10] Gibson, D., *Pharmacogenomics J.*, **2**, 275 (2002).

[11] Tse, W.C., Boger, D.L., *Chemistry and Biology*, **11**, 1607 (2004).

[12] Pindur, U., Jansen, M., Lemester, T., *Curr. Med. Chem.*, **12**, 2805 (2005).

[13] Hamilton, P.L., Arya, D.P., *Nat. Prod. Rep.*, **29**, 134 (2012).

[14] Khan, G.S., Shah, A., Zia-ur-Rehman, Barker, D., *J. Photochem. & Photobiol. B: Biology*, **115**, 105 (2012).

[15] Nanjunda, R., Wilson, W.D., *Curr. Protoc. Nucleic Acid Chem.*, **51**, 1 (2012)**.**

[16] Kumar, S., Pandya, P., Pandav K., Gupta S.P., Chopra A., *J. Biosci.*, **37**, 553 (2012).

[17] Srivastava, H.K., Sastry, G.N., *J. Biomol. Struc. Dyn.*, **31**, 522 (2013).

[18] Charies, J.B., *Biopolymers*, **103**, 473 (2015).

[19] Lerman, L.S., *Biochem.*, **49**, 95 (1963).

[20] Neto, B.A.D., Lapis A.A.M, *Molecules*, **14,** 1725 (2009).

[21] Beiebricher, A.S., Heller, I., Roijmans, R.F.H, Hoekstra, T.P., Petermen, E.J.G, Wuite, G.J.L., *Nature Comm.*, **6**, 7304(2015).

[22] Hendry, L.B., Mahesh, V.B., Bransome Jr., E.D., Ewing, D.E., *Mutat. Res.*, **623**, 53 (2007).

[23] Ferguson, L.R., Denny, W.A., *Mutat. Res.*, **623**, 14 (2007).

[24] Lohani, N., Singh, H.N., Moganty, R.R., *Inter. J. Biol. Macr.*, **87**, 433 (2016).

[25] Charies, J.B., *Biopolymers*, **44**, 201 (1997).

[26] Marco, A., Mejia, L., Rocha, A.L., *J. Mol. Graph. Modell.*, **27**, 9000 (2009).

[27] Rehman, S., Sarwar, T., Hussain, M.A., Ishqi, H.M., *Arch. Biochem. Biophy.*, **576**, 49 (2015).

[28] Muller, W., Crothers, D.M, *J. Mol. Biol.*, **35**, 251 (1968).

[29] Mosher, C.W., Kuhlmann, K.F., Kleid, D.G., Henry, W.H., *J. Med. Chem.*, **20**, 1055 (1977).

[30] Tong, G.L., Cory, M., Lee, W.W., Henry, D.W., *J. Med. Chem.*, **21**, 732 (1978).

[31] Qu, X., Ren, J., Riccelli, P.V., Benight, A.S., *Biochemistry*, **42**, 11960 (2003).

[32] Martinez, V., Burgos, C., Builla, J.A., Fernandez, G., Domingo, A., Nieto, R.G., Gago, F., Manzanares, I., Cuevas, C., Vaquero, J.J., *J. Med. Chem.*, **47**, 1136 (2004).

[33] Rajendran, A., Nair, B.U., *Biochimica et Biophysica Acta*, 1760, 1794 (2006).

[34] Liu, H., Sadler, P.J., *Acc. Chem. Res.*, **44**, 349 (2011).

[35] Lu, Y., Wang, G., Tang, W., Hao, X., Xu, M., Li, X., *Spectrochimi. Acta Part A*, **82**, 247 (2011).

[36] Islam, M., Fuji, S., Sato, S., Okauchi, T., Takenaka, S., *Bioorg. Med. Chem.*, **23**, 4769 (2015).

[37] Coulombeau, C., Gresh, N., *Nucleic Acids Res.*, **18**, 711 (1990).

[38] Crenshaw, J.M., Graves, D.E., Denny, W.A., *Biochem.*, **34**, 13682 (1995).

[39] Todd, A.K., Admas, A., Thorpe, J.H., Denny, W.A., Wakelin, L.P.G, Cardin, C.J, *J. Med. Chem.*, **42**, 536 (1999).

[40] Chen, K., Gresh, N., Pullman, B., *Nucleic Acid Res.*, **14**, 3799 (1986).

[41] Con, P., Phillips, D.R., *Nucleic Acid Res.*, **22**, 1342 (1994).

[42] Varadwaj, P., Misra, K., Sharma, A., Kumar, R., *Elec. J. Biol.*, **6**, 36 (2010).

[43] Agarwal, S., Jangir, D.K., Mehrotra, R., *J. Photochem. Photobiol. B: Biology*, **120**, 177 (2013).

[44] Thimmaiah, K., Ugarkar, A.G., Martis, E.F., Shaikh, M.S., Coutinho, E.C., Yergeri, M.C., *Nucleosides, Nucleotides and Nucleic Acids*, **34**, 309 (2015).

[45] Yim, H., Lee, Y.H., Lee, C.H., Lee, S.K., *Planta Medica*, **65**, 9 (1995).

[46] Saito, S.T., Silva, G., Pungartnik, C., Brendel, M., *J. Photochem. Photobiol. B: Biology*, **111**, 59 (2012).

[47] Amutha, R., Subramanian, V., Nair, B.U., *Chem. Phys. Lett.*, **344**, 40 (2001).

[48] Kamal, A., Shetti, R.V., Rmaiah, M.J., Swapna, P., Reddy, K.S., Mallareddy, A., Rao, M.P.N., Chourasia, M., Sastry, G.N., Juvekar, A., Zingde, S., Sarma, P., Pushpavalli, S.N., Bhadra, M.P., *Med. Chem. Comm.*, **2**, 780 (2011).

[49] Kamal, A., Shankaraiah, N., Reddy, Ch. R., Prabhakar, S., Markandeya, N., Srivastava, H.K., Sastry, G.N., *Tetrahedron*, **66**, 5498 (2010).

[50] Kamal, A., Reddy, K.S., Khan, M.N.A., Shetti, R.V.C.R.N.C., Ramaiah, M.J., Pushpavalli, S.N.C.V.L., Srinivas, C., Pal-Bhadra, M., Chourasia, M., Sastry, G.N., Juvekar, A., Zingde, S., Barkume, M., *Bioorg. Med. Chem.*, **18**, 4747 (2010).

[51] Kamal, A., Bharathi, E.V., Ramaiah, M.J., Dastagiri, D., Reddy, J.S., Viswanath, A., Sultana, F., Pushpavalli, S.N.C.V.L., Pal-Bhadra, M., Srivastava, H.K., Sastry, G.N., Juvekar, A., Sen, S., Zingde, S., *Bioorg. Med. Chem.*, **18**, 526 (2010).

[52] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., *Nucleic Acids Res*., **28**, 235 (2000).

[53] Berman, H.M., Olson, W.K., Beveridge, D.L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh S., Srinivasan, A.R., Schneider, B., *Biophysical Journal*, **63**, 751 (1992).

[54] Narayanan, B.C., Westbrook, J., Ghosh, S., Petrov, A.I., Sweeney, B., Zirbel, C.L., Neocles, B.L., Bermen, H.M., *Nucleic Acid Res.*, **42**, 114 (2014).

[55] Gaussian 09, Revision A.02, Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G.A., Nakatsuji, H., Caricato, M., Li X., Hratchian, H.P., Izmaylov, A.F., Bloino, J., Zheng, G., Sonnenberg, J.L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J.A., Jr., Peralta, J.E., Ogliaro, F., Bearpark, M., Heyd, J.J., Brothers, E., Kudin, K.N., Staroverov, V.N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J.C., Iyengar, S.S., Tomasi, J., Cossi, M., Rega, N., Millam, J.M., Klene, M., Knox, J.E., Cross, J.B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli, C., Ochterski, J.W., Martin, R.L., Morokuma, K., Zakrzewski, V.G., Voth,

G.A., Salvador, P., Dannenberg, J.J., Dapprich, S., Daniels, A.D., Farkas, O., Foresman, J.B., Ortiz, J.V., Cioslowski, J., Fox D.J., Gaussian, Inc., Wallingford CT, 2009.

[56] Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meing, E.C., Ferrine, T.E., *J. Comput. Chem.*, **25**, 1605 (2004).

[57] Humphrey, W., Dalke, A., Schulten, K., *J. Mol. Graph.*, **14,** 33 (1996).

[58] Morris, G.M., Huey, R., Lindstorm, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J., *J. Compt. Chem*., **30**, 2785 (2009).

[59] Diego, S., Accelrys Software Inc., **2007**.

[60] Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., Spoel, D., Hes, B., Lindahl, E., *Bioinformatics*, **29**, 845 (2013).

[61] Case, D.A., Berryman, J.T., Betz, R.M., Cerutti, D.S., Cheatham, III T.E., Darden, T.A., Duke, R.E., Giese, T.J., Gohlke, H., Goetz, A.W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T.S., LeGrand, S., Li, P., Luchko, T., Luo, R., Madej, B., Merz, K.M., Monard, G., Needham, P., Nguyen, H., Nguyen, H.T., Omelyan, I., Onufriev, A., Roe, D.R., Roitberg, A., Salomon-Ferrer, R., Simmerling, C.L., Smith, W., Swails, J., Walker, R.C., Wang, J., Wolf, R.M., Wu, X., York, D.M., Kollman, P.A., AMBER 2015, University of California, San Francisco (2015).

[62] Kumari, R., Kumar, R., Lynn, A., *J. Chem. Inf. Model.*, **54**, 1951 (2014).

[63] Lian, C., Robinson, H., Wang, A.H.J., *J. Am. Chem. Soc.*, **118**, 8791 (1996).

Chapter 5

The role of Quantum Mechanics/Molecular Mechanics calculations in understanding the binding of DNA minor groove binders

Chapter 5

The role of Quantum Mechanics/Molecular Mechanics calculations in understanding the binding of DNA minor groove binders

5.1 Introduction

The nucleic acid is the target for many anticancer, antiviral, antitumor, antiprotozoal and antiparasitic drugs which are currently in clinical use or in clinical trials. Most of the DNA interactive drugs are aromatic compounds with low molecular weight. DNA groove binding and DNA intercalation are the two leading DNA-binding modes of small molecules/drugs. Intercalation needs conformational changes in DNA due to the formation of intercalation gap between the DNA base pairs, in contrast to minor groove binders which does not require any conformational changes to the DNA. The structural arrangement of base pairs gives rise to the architecture of major and minor grooves and the environment of each groove is different at molecular prescriptive. The major groove has a width of 11.7 Å and depth of 8.7 Å and have multiple sites of interaction which results in comparatively strong binding with ligands. The major groove provides the easy entry of large molecules. In spite of that, the minor groove is smaller in size, with the width of 5.7 Å and depth of 7.5 Å and has less binding sites. DNA minor groove binders are commonly flexible, crescent in shape contains rotatable bonds and are able to direct itself towards the minor groove of DNA. Generally, these molecules form non-covalent interaction with DNA and also some of these molecules cause cleavages in DNA backbone. Typically, minor groove binders

bind to B-DNA with the higher binding affinity towards AT- rich sequences.[1-7] Intercalators are the planar molecules with the incorporation of several fused rings, which insert between the adjacent base pairs of DNA duplex by their planar moiety. These interactions are quite strong although the fact that energy is consumed due to the unwinding of the helix and unstacking of the base pairs during complex formation. Contributing to the stability of intercalated complexes are hydrophobic interactions.[8-10]

DNA minor groove is the target for many anticancer and antitumor drugs. Hoechst 33258 (2'-(4-hydroxyphenyl)-5-(4-methyl-1-piperazinyl)-2, 5'-bi-benzimidazole) and IBB (5-(2-imidiazolinyl)-2-[2-(4-hydroxyphenyl)-5-benzimidazolyl] benzimidazole) fall into the category of bis-benzimidazole.[11-15] Hoechst 33258 is the experimental antitumor agent and also exhibits fluorescence when bound to DNA which may vary according to their binding sequence with DNA.[21] IBB is the analogue of Hoechst 33258, in which the piperazine has been replaced by an imidazoline group, due to which higher binding affinity is shown by this analogue with DNA compared with Hoechst itself. [16-20, 21] Netropsin is known experimental antitumor drug which have been isolated from Streptomyces netropsis which contains two different terminal cationic centers: one an amidine and other is guanidine and interact to the minor groove of DNA to the AT sites.[21-24] Berenil is a synthetic diarylamidinetrypanocidal drug used for the treatment of trypanosmal infections in animals in Africa. This drug exhibits the characteristics of both minor groove binder and intercalator but shows strong binding towards minor groove at AT sites and berenil can also bind to both RNA and DNA.[25-29] 2, 5-bis (4-guanylphenyl) furan (furamidine) is an analogue of berenil but differs only with the central triazene unit is replaced with a furan moiety, effective against Pneumocystis *carinii* pathogen.[27]

The synthetic DAPI (4',6-diamidino-2-phenylindole) is diamidine used for its antitrypanosomal activity and is also one of the important DNA binding drug because of its application as antiviral, antibiotic and anticancer drug. It forms a fluorescent complex by interacting to the minor groove of AT-rich DNA sequences and also shows weak binding interaction as an intercalator at GC sites.[30-35] Pentamidine (1,5-bis(4-amidinophenoxy)pentane), is a minor groove binder which has been significant in clinical use. Pentamidine is considered as a derivative of bisbezamidine in which bezamidine moieties are linked by a flexible pentyldioxy chain. Pentamidine is used against primary stage African trypanosomiasis, antimony-resistant leishmaniasis and AIDs associated Pneumocystis carinii.[36-40] Propamidine (1,3-bis(4-amidinophenoxy)propane) is a short chain analogue of pentamidine active against the *Pneumocystis carinii* pathogen.[41-42] The forces that involve in DNA-minor groove binding are electrostatic, van der Waals, hydrophobic, and hydrogen bonding. Sequence specificity is important key tool for the drug-DNA interaction and another important factor for the interaction is that the small molecule has a crescent shape which is complementary to the minor groove of DNA. Frequently, minor groove shows selectivity towards AT region, several factors are responsible for this, First is that the electrostatic potential of AT-rich region is higher than that of GC-filled ones and another one is that the AT-rich grooves are narrower and deeper than GC ones. DNA minor groove with alternating A and T, allows favorable van der Waals contacts between the drug and DNA instead of GC-rich region where the geometry of groove is altered by bulky amino groups of guanine bases.

In this study, different minor groove binders classes such as bis-benzimidazole (Hoechst and IBB), polyamides, diarylamidines, and furamidines are undertaken. The detailed interaction view of minor groove recognition by small molecules such as

Netropsin, hoechst, berenil, DAPI, pentamidine, propamidine, furamidine via computational chemical techniques is investigated. Molecular modelling techniques such as molecular docking, molecular dynamics simulation, MMPBSA calculation and Quantum Mechanics/Molecular Mechanics method (QM/MM) has been used for the estimation of binding free energies, which was compared to the experimental results. In the number of studies for the minor groove binders, these molecular modelling methods show good agreement with the experimental results.[43-47]

Molecular Docking is simplest computational representation of DNA-ligand interaction and is usually unable to predict the correct binding affinity. Many docking protocols are available for the protein ligand complex structure, these docking methods are not reasonable for the DNA-ligand complexes because there is no well-defined active site for DNA. Srivastava et al. demonstrated that GOLD and GLIDE docking protocols are reliable in modelling of DNA-ligand complexes in comparison of CDOCKER and AUTODOCK.[43] So, it is important to include the dynamics information to accurately estimate the binding free energy, Hence the weakness of scoring function used in docking protocols is ruled out. The force field proposed for the MD simulation studies of nucleic acid are CHARMM, AMBER, GROMOS, OPLS, ENCAD and BMS. There are many studies for the comparison of force fields for the nucleic acids but AMBER force fields seems to be good for nucleic acid simulation due to the presence of specific topologies for the terminal nucleotides. MD simulation followed by MMPBSA can also contribute conformational fluctuations and entropic contribution to the binding free energy. The MMPBSA method contributes three energetic terms to the binding free energy. The first term represents change in the potential energy, which includes both bonded (bond, angle and torsion energies) and non-bonded terms (van der waals and electrostatic term). MD

simulation has been used to generate ensemble of different binding conformations in the presence of explicit water. The MMPBSA method was used to calculate the free binding energy and to calculate the relative stabilities of bio molecular system.

The QM/MM approach may be taken as reliable technique for the large biomolecules to investigate the structural, electronic and chemical properties than the MM computational technique. The QM/MM approach combines computational methods of different accuracy and efficiency, resulting in an extremely powerful tool for the study of biological systems. The atoms which were in the chemically active region are included in the QM region, while the rest of the region is treated at the MM level. The QM/MM approach was first applied by Warshel & Levitto on the enzyme system.[48] Now a day this approach has widely spread and successfully describe and predict chemical processes in complex environment. Due to the potential application in various fields of this method, this field gained the Nobel Prize in chemistry in 2013.

This study used advance computational techniques for a detailed mechanistic understanding of the process of drug-DNA interactions at the molecular level using QM/MM approach other molecular modelling techniques such as molecular docking and molecular dynamics (MD) are used to predict binding mode and to check the stability of bounded ligand respectively, which will provide a theoretical protocol for experimental techniques, generation of database for structure activity/property relationship in drug-DNA complexes.

5.2 Material and Methods

5.2.1 Dataset

The crystal data of the drug-DNA complexes were downloaded from the Protein Data Bank [49] having PDB Id 127D, 2D64, 1D30, 1D64, 1D63, 121D, 102D, 109D,

195D, 1D86, 2DBE, 1PRP, 227D and their experimental binding energies were collected from literature.[19,23-27, 36, 41-42, 50-53] The water molecules and the ligands were removed and the crystal structure of B-DNA was extracted from their respective drug-DNA complex crystal structure. The drug molecules extracted from these complexes were subjected to geometry optimization using Gaussian 09[54] at B3LYP/6-31G* level. The chemical structure of minor groove binders which were used for the further study is shown in Fig. 5.1.

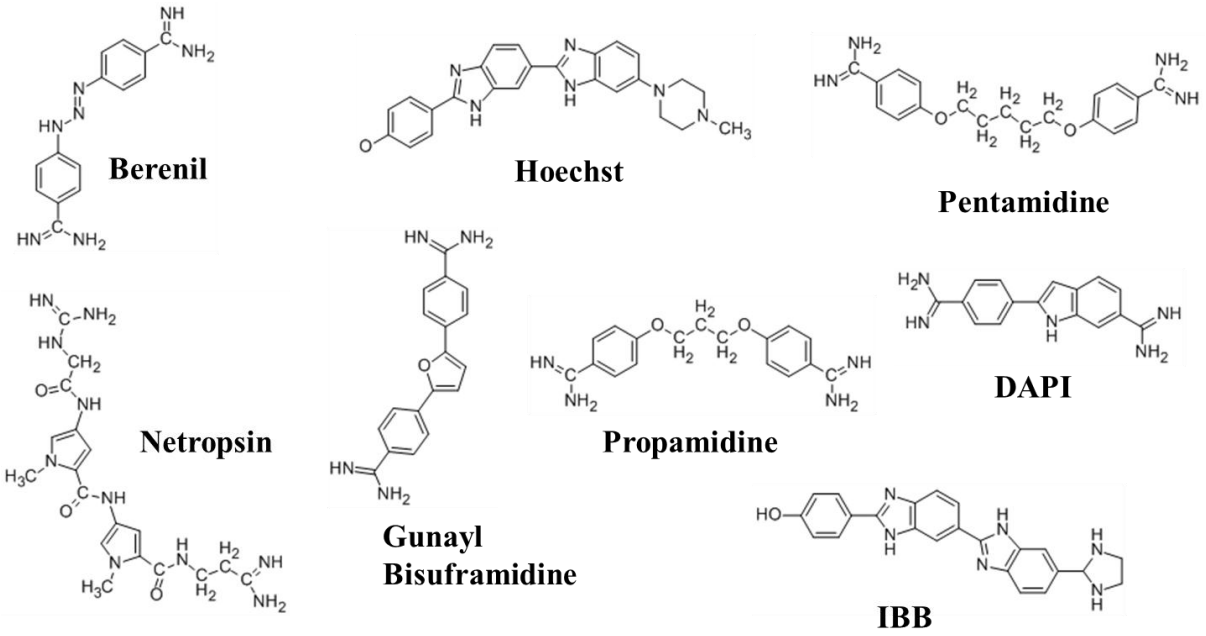


Fig. 5.1: Chemical Structure of Minor Groove Binders.

5.2.2 Molecular Docking

Blind molecular docking calculations between drug and DNA sequence were performed using MGL tools 1.5.6 with AutoGrid4 and AutoDock4.[55] DNA was extracted from the respective complexes and all the heteroatoms and water molecules were deleted. A grid box of grid points 70×70×70 was created with grid spacing of 0.375Å in order to include the entire DNA fragment. After the creation of grid map, grid potential map was calculated using Auto Grid 4.0. Docking calculation were

performed using Lamarckian genetic algorithms, as accomplished in Autodock. The 20 LGA runs with maximum of 25,00000 energy evaluations were performed. All other parameters were used with default settings. According to the scoring function, the lowest energy docked conformation was selected as best docked pose for each docking case, this best docked pose was used for the further study.

5.2.3 Molecular Dynamics Simulation

Molecular Dynamics simulation (MD) of 5ns for each selected best poses was carried out using GROMACS 4.5.6. The pdb2gmx program of the GROMACS [56] package was used to generate the parameters and topology files for the DNA using AMBER94 force field while for the ligands these files were processed via the 'acpype.py' python script using generalized amber force field (GAFF), which "translated" them into GROMACS topology file formats. The coordinate and topology files of DNA and ligands were merged to obtain the final starting structure and topology file for each complex. The simulation system was constructed by immersing the macromolecules in a dodecahedron of explicit TIP3P water molecules. The drug-DNA complex was placed in the center of dodecahedron periodic box of size 1.2Å. The system was then solvated in TIP3P water molecules. The total charge on the system was then neutralized by adding counter ions. The energy was minimized using steepest descent algorithm. Then the system was heated to 300K during 50 ps of constant volume simulation with 2 fs time step. The pressure was equilibrated to 1 atm during 50 ps NPT simulation with 2 fs time step. In both the simulations a restrained with force constant of 1000kJ/(mol/nm^2). Both temperature and pressure were regulated using Berendsen algorithm. Production simulations were performed for 5 ns with a 2fs time step. The temperature and pressure were maintained at 300 K and 1 atm using the v-rescale temperature and Parninello-Rahman pressure coupling method. Root mean

square deviation (RMSD), root mean square fluctuations (RMSF), radius of gyration (Rg) and hydrogen bond distribution for each system was determined by the analysis of MD trajectories produced during the 5ns of the production run at every 2ps. Trajectories were visualized by means of VMD. All the graphs were plotted using GNUPLOT.

5.2.4 Binding Energy calculations

The binding energy of each DNA-ligand complex was determined using the g_mmpbsa [57] tool. From 5,000 snapshots structure extracted from the 5ns time period, the binding free energy (ΔG_{bind}) was determined using the equations below-

$$\Delta G_{bind} = G_{complex} - (G_{receptor} + G_{ligand}) \quad (5.1)$$

$$\Delta G_{bind} = E_{gas} + G_{sol} - T\Delta S \quad (5.2)$$

$$E_{gas} = E_{\text{int}} + E_{vdw} + E_{ele} \quad (5.3)$$

$$G_{sol} = G_{pol} + G_{SA} \quad (5.4)$$

$$G_{SA} = SASA + b \quad (5.5)$$

where, $G_{complex}$, $G_{receptor}$ and G_{ligand} are the absolute free energies of the DNA-ligand complex, a DNA and ligands respectively. The ΔG_{bind} was decomposed to its individual contributions (equations 5.1 to 5.5); gas-phase energy (E_{gas}) which is a sum of bonded (E_{int}) and non-bonded terms (E_{vdw} and E_{ele}); the solvation free energy (G_{sol}) polar (G_{pol}) and nonpolar (GSA) solvation energy components, and an entropy terms ($T\Delta S$).Polar solvation energies were determined by solving the Poisson-Boltzmann linear equation while nonpolar solvation through the solvent accessible surface area with an offset value (b) of 3.84928 $kJ.mol^{-1}$ and surface tension proportionality (γ) set

at 0.0226778 kJ.mol^{-1}.Å^{-2}. The individual contributions of DNA residues to the three energetic components were determined through per-residue decomposition.

5.2.5 QM/MM Studies

Snapshots obtained at each nanosecond from the MD simulation from GROMACS, were used as starting point for the QM/MM modeling. These structures contain the bimolecular system in a droplet of water (20000-30000 atoms) and this setup requires a lot of prior work to avoid errors and wrong choices for the actual QM/MM calculations. The QM/MM calculations were performed with the two-level ONIOM method within the Gaussian 09 program suite using the B3LYP functional and AMBER force field. The two layered ONIOM scheme is shown in Fig. 5.2. B3LYP method along with basis set 6-31+G is used for the high layer and AMBER force field for the low layer and all the water molecules and Na^+ ions are frozen. The mechanical-embedding version of ONIOM was used for geometry optimization and the energy of the high or QM region is calculated. Now, the co-ordinate of QM region from the optimized system is extracted, which is further subjected to geometry optimization and the energy of QM region in the gas phase is calculated by the same method. The interaction energy between the DNA and drug molecule is calculated with the help of formula-

$$\Delta E_{I.E} = E_{QM_{(g,p)}} - E_{QM_{(p,p)}} \qquad (5.6)$$

where, $\Delta E_{I.E}$ is the interaction energy between drug and DNA. $E_{QM\,(p,\,p)}$ is the energy of QM region in nucleic acid phase and $E_{QM\,(g,\,p)}$ is the energy of QM region in the gas phase.

The electronic-embedding model of ONIOM (ONIOM-EE) was used for single point calculation by using the method B3LYP with basis set 6-31G** (d,p) and M062X for the QM region and AMBER force field for the MM region. The single point energy of

QM region in the gas phase with the same method was also calculated. The interaction energy between the DNA and drug was calculated using above mentioned formula.

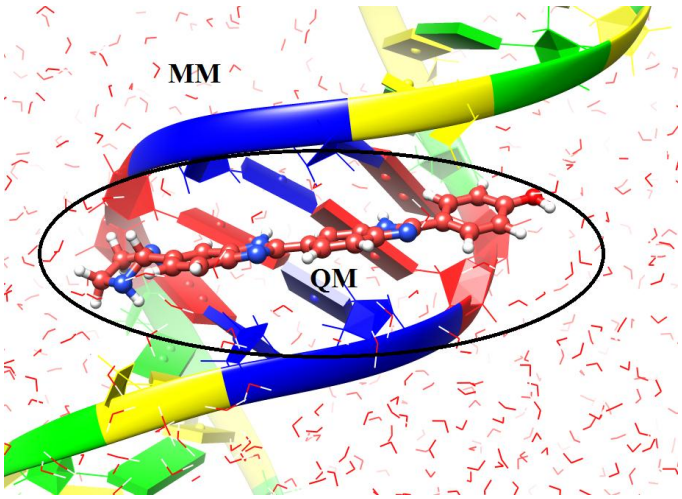


Fig. 5.2: Two layered ONIOM scheme.

5.3 Result and Discussions

5.3.1 Molecular Docking

The DNA minor groove binding drugs were successively docked with their respective DNA sequence in order to predict binding site along with the preferred orientation of ligand to the DNA minor groove. From the docking calculations, the most energetically favorable conformer among the 20 runs is picked up as initial conformer for the further study. The docking studies reveal that the ligands are stacked in the minor groove of DNA duplex via van der Waal and hydrophobic interactions. Additionally, the minor groove binding drugs exhibit strong hydrogen bonding to the nucleic acid bases. From the results, this is concluded that the minor groove binders generally bind to the A and T DNA residues. All the docked complexes are stabilized by hydrogen bonding and hydrophobic interactions between ligand and nucleic acid bases. The hydrogen bond formation occurs between minor groove binding ligands and the functional groups on the bases are exposed in the grooves via their end groups

and also their amide or other linker groups. The different type of interactions takes place in top three most stable complexes shown in Fig. 5.3. The experimental binding energy linearly correlates to the calculated binding free energy with R^2 = 0.87 shown in Fig. 5.4. Molecular docking calculations show that all the ligands bind to AT-rich region of DNA with good docking fitness score (Table 5.1). The complex having PDB Id 109D has lowest binding energy so this complex is more stable than others.

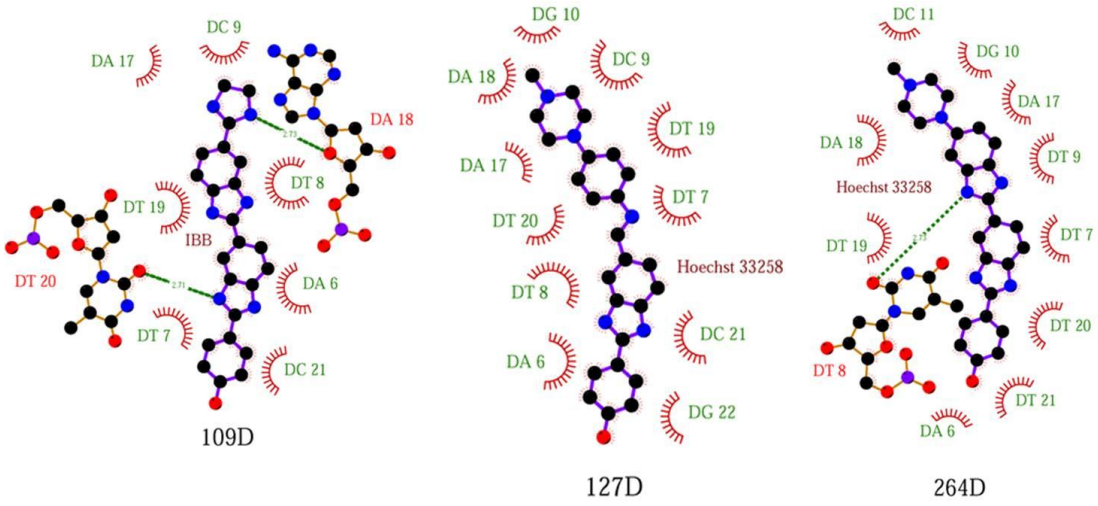


Fig. 5.3: Different types of interactions obtained from molecular docking studies for top three stable complexes.

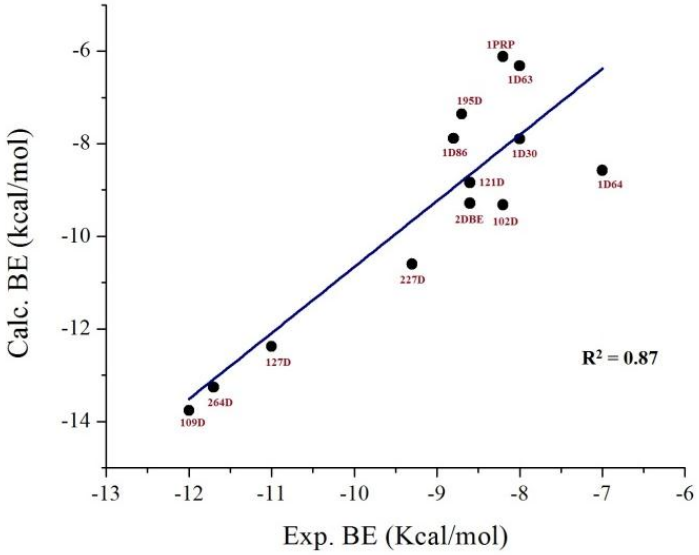


Fig. 5.4: Variation of calculated binding energy with experimental binding energy (R^2 = 0.87).

Table 5.1: List of PDB Ids taken for the study with their DNA sequence, resolution, No. of heavy atoms, calculated binding energies (kcal/Mol) from molecular Docking and experimental binding energies (kcal/mol) obtained from literature.

PDB Id	Drug	DNA Sequence	No. of Heavy Atoms	Exp. ΔG_{bind}	Calc. ΔG_{bind} (Docking)	Residue Ids of nucleic acid bases involved in hydrogen bond formation
127D	Hoechst33258	5'-CGCGAATTCGCG-3'	32	-11.0	-12.38	-
264D	Hoechst33258	5'-CGCAAATTTGCG-3'	32	-11.7	-13.26	DT8
1D86	Netropsin	5'-CGCGAATTCGCG-3'	31	-8.8	-7.89	DT20,DT19,DA18,DA6
121D	Netropsin	5'-CGCAAATTTGCG-3'	31	-8.6	-8.84	DT7, DT9,DA18,DT19
195D	Netropsin	5'-CGCGTTAACGCG-3'	31	-8.7	-7.36	DA20,DG10,DA7,DT6
109D	IBB	5'-CGCGAATTCGCG-3'	30	-12.0	-13.76	DA18, DT20
1D64	Pentamidine	5'-CGCGAATTCGCG-3'	25	-7.0	-8.58	DA7,DG10
227D	Guanyl Bisfuramidine	5'-CGCGAATTCGCG-3'	23	-9.3	-10.60	DC9,DA18
1PRP	Propamidine	5'-CGCGAATTCGCG-3'	23	-8.2	-6.12	-
1D30	DAPI	5'-CGCGAATTCGCG-3'	21	-8.0	-7.90	DA6,DT8
1D63	Berenil	5'-CGCAAATTTGCG-3'	21	-8.0	-6.32	DT20,DT21
2DBE	Berenil	5'-CGCGAATTCGCG-3'	21	-8.6	-9.29	DA6,DA18,DT8
102D	Propamidine	5'-CGCAAATTTGCG-3'	23	-8.2	-9.32	DT9,DA6, DT7

All the energies are in kcal/mol.

5.3.2 Molecular Dynamics Results

MD simulation from GROMACS 4.5.6 was performed for the best docked poses selected from the docking studies. To investigate the structural behavior of the drug-DNA complexes and to test the stability of simulations, the root mean square deviation (RMSD) of the atoms have been calculated from the starting structures. The average RMSD values for the DNA-Ligand complexes and free DNA is 0.2-0.3 nm and ligands within the last 3ns of the simulations are 0.2-0.3 nm and 0.03-0.1 nm

respectively, while the average RMSD of free ligands are 0.03-0.1 nm except for PDB Id 195D. The RMSD profile of all the systems is shown in Fig. 5.5 indicates the result of dynamic behavior of DNA-Ligand complexes from the initial position to final position throughout the simulation time. The root-mean square fluctuations of all the DNA-ligand complexes were calculated within the 5ns simulation. The value RMSF is higher at the end terminal of nucleic acid and for the ligand shown in Fig. 5.6. Fig. 5.7 depicts the hydrogen bonding pattern of the most top three ligands to the minor groove. The complex with **PDB Id 127D has** lowest binding energy -65.83±3.42 kcal/mol shows three hydrogen bonds with Thymine (T) base pairs of DNA (one Hydrogen bond each with DT-7, DT-19 and DT-20) at the end of simulation and six hydrogen bonds with solvent i.e. water. For the **PDB Id 109D** three hydrogen bonds with Adenine (A) and Thymine (T) base pairs of DNA (one hydrogen bond each with DA-6, DT19 and DT20), while hydrogen bond with water solvent and for **PDB Id 264D** four hydrogen bonds were obtained with base pairs DT7, DT 8, DA18 and DT 19 of nucleic acid and five hydrogen bond with solvent. The snapshots at 5ns was obtained shown in Fig. 5.8 depicts that the ligand remain binds at the AT region of DNA up to the end of simulation.

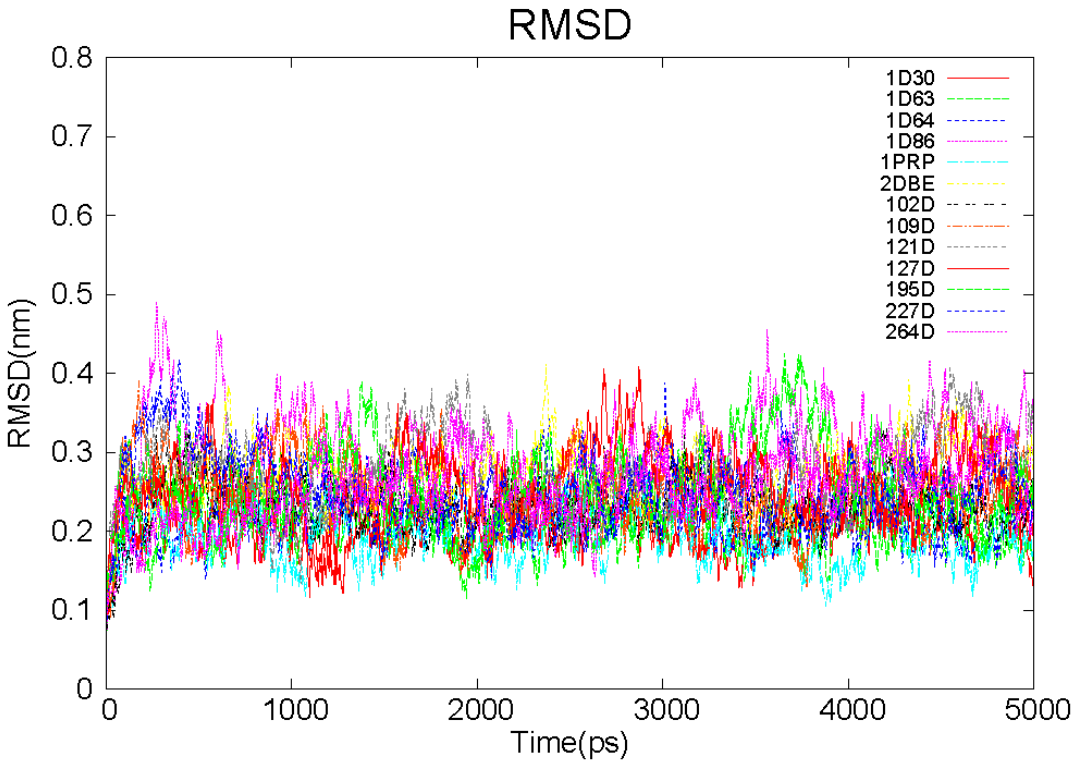


Fig. 5.5: RMSD profile for DNA-Ligand complexes.

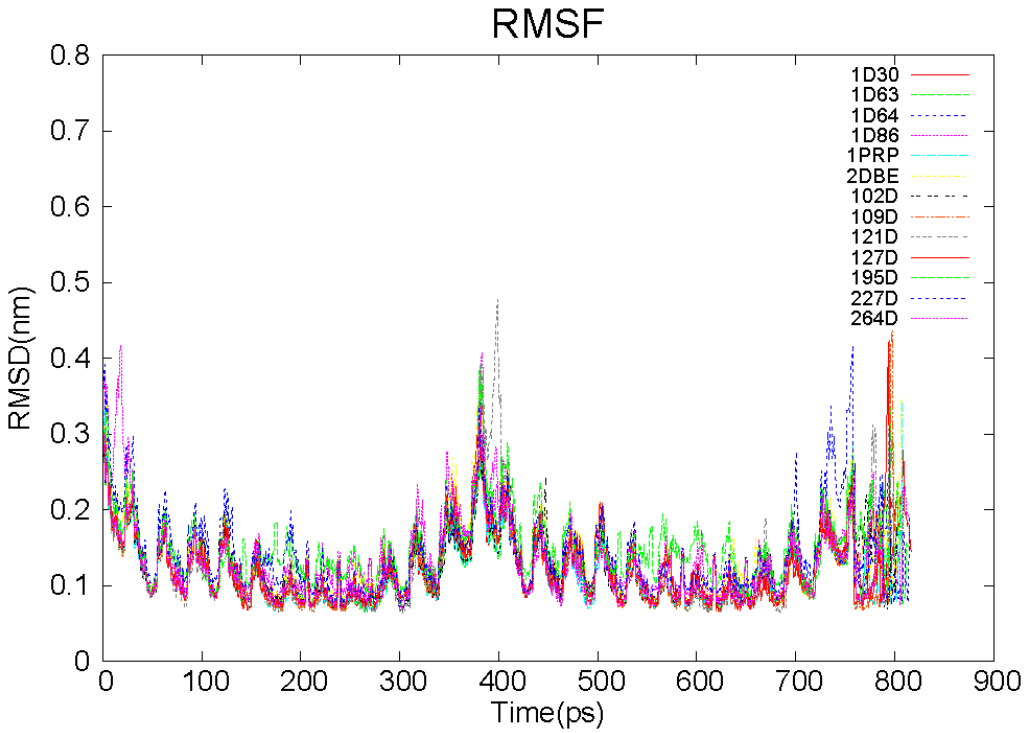


Fig. 5.6: RMSF profile for DNA-Ligand complexes.

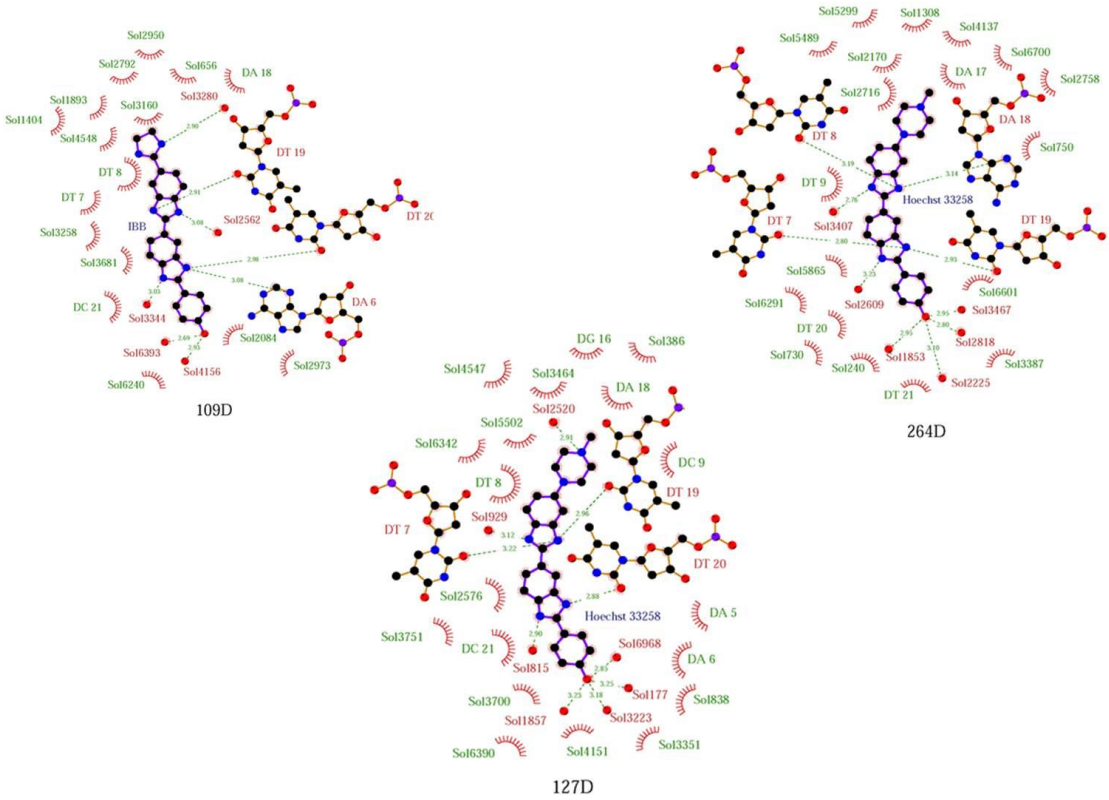


Fig. 5.7: Hydrogen bonding between the most potent top three minor groove binders and DNA bases.

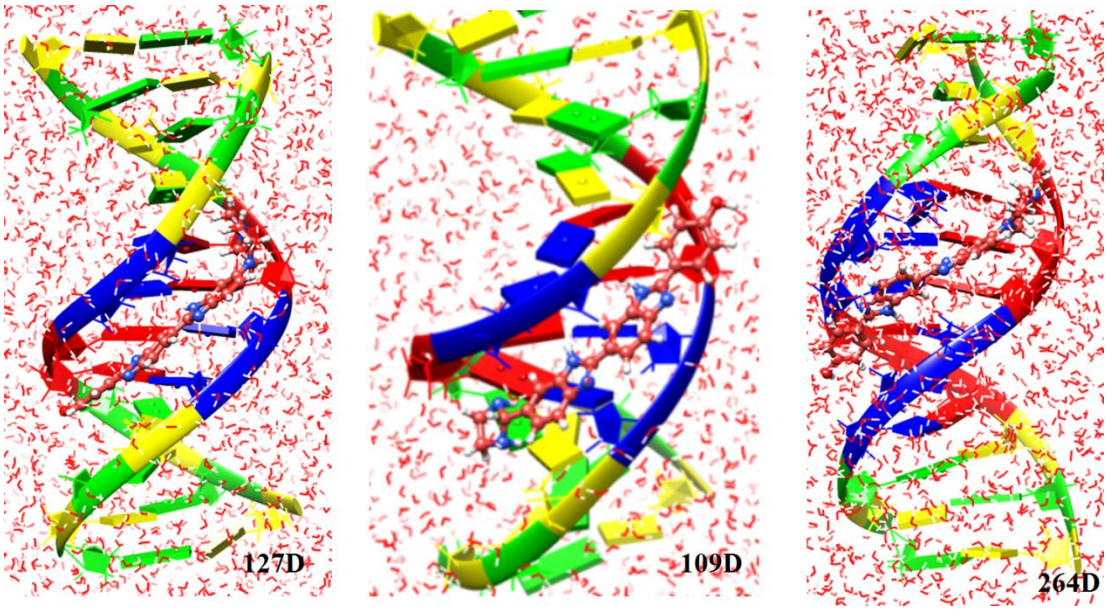


Fig. 5.8: Snapshots of top three DNA-ligand complexes at 5ns MD Simulation.

5.3.3 Binding Energy Calculations

The MMPBSA calculations were performed to calculate the strength of interaction between ligand and DNA, which can be measured by the binding energy values presented in Table 5.2. Table 5.2 clearly indicates that all the ligands show consistent interaction until the 5ns period without much variation in the IE values. Most of the DNA-ligand complexes have lowest IE values between the 1-3ns of the time scale except for the 195D, 1D64, and 1D30, 102D. To understand the DNA-ligand interaction at molecular level, it is important to analyze the DNA residues that interact with the ligand. The decomposition of total binding energy into the contribution made by each residue enables the comparison of the relative contribution of residues to the overall binding energy. The energy per residue contribution was calculated for all the 13 complexes and it can be seen that the energy contribution profiles are similar for all the complexes. The residues which strongly interact with ligands are DT7, DT8, DC9, DA17, DA18, DT19 and DT20 that reveals that the minor groove binders binds to AT-rich region**.** Fig. 5.9 shows the binding energy contributions of residues for all DNA-ligand complexes. Fig. 5.10 shows the contribution of each component of energy to the final binding energy and it was analyzed that the contribution of van der Waals interaction is greater than that of electrostatic energy to the final binding energy. The moderate correlation was observed between experimental BE values and calculated IE values with $R^2 = 0.77$, $R^2 = 0.78$, $R^2 = 0.79$, $R^2 = 0.76$ and $R^2 = 0.71$ for the 0-1ns, 1-2ns, 2-3ns, 3-4ns and 4-5ns time scale respectively. The plot between the experimental binding energy and final calculated binding energy value shows linear correlation with $R^2 = 0.77$ shown in Fig. 5.11**.**

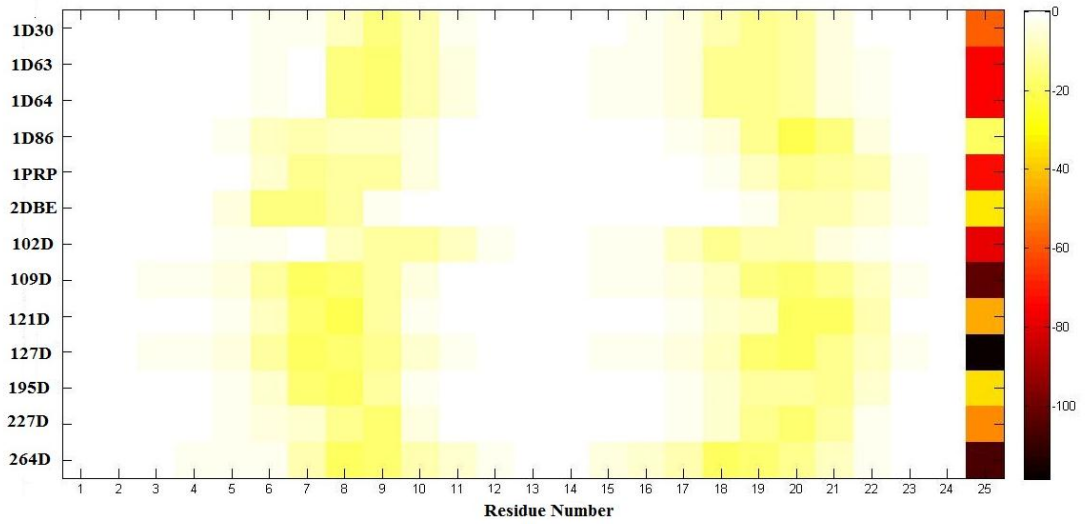


Fig. 5.9: Energetic contribution of DNA residues in the binding in kilojoules/mol.

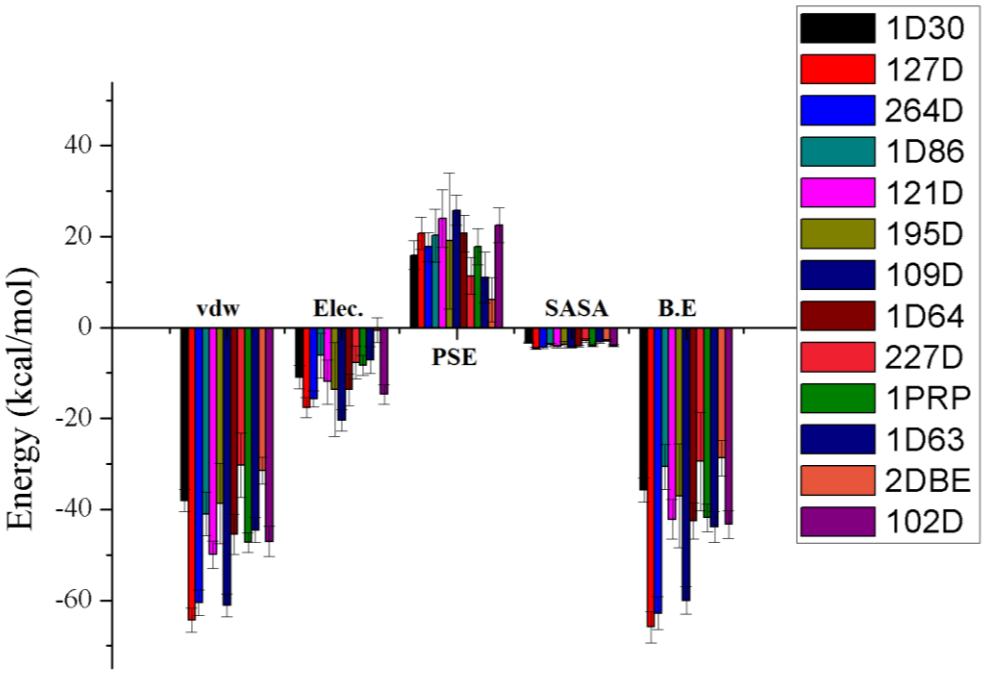


Fig. 5.10: Histogram depicting view of the contribution of each and every component of energy to the final binding energy.

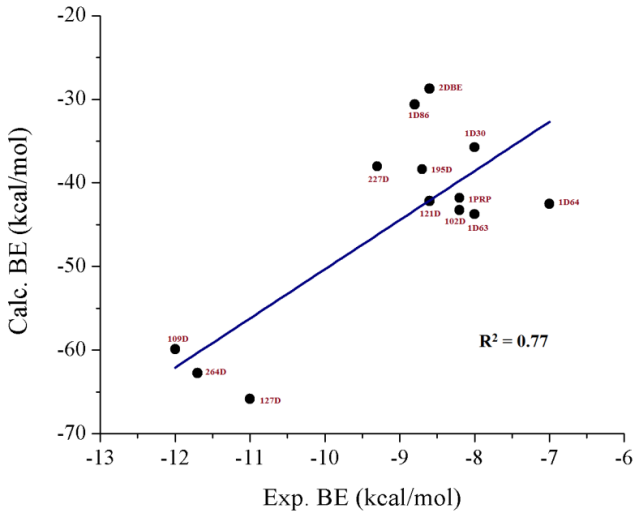


Fig. 5.11: Variation of calculated binding energy with the experimental binding energy obtained from MMPBSA calculation using GROMACS.

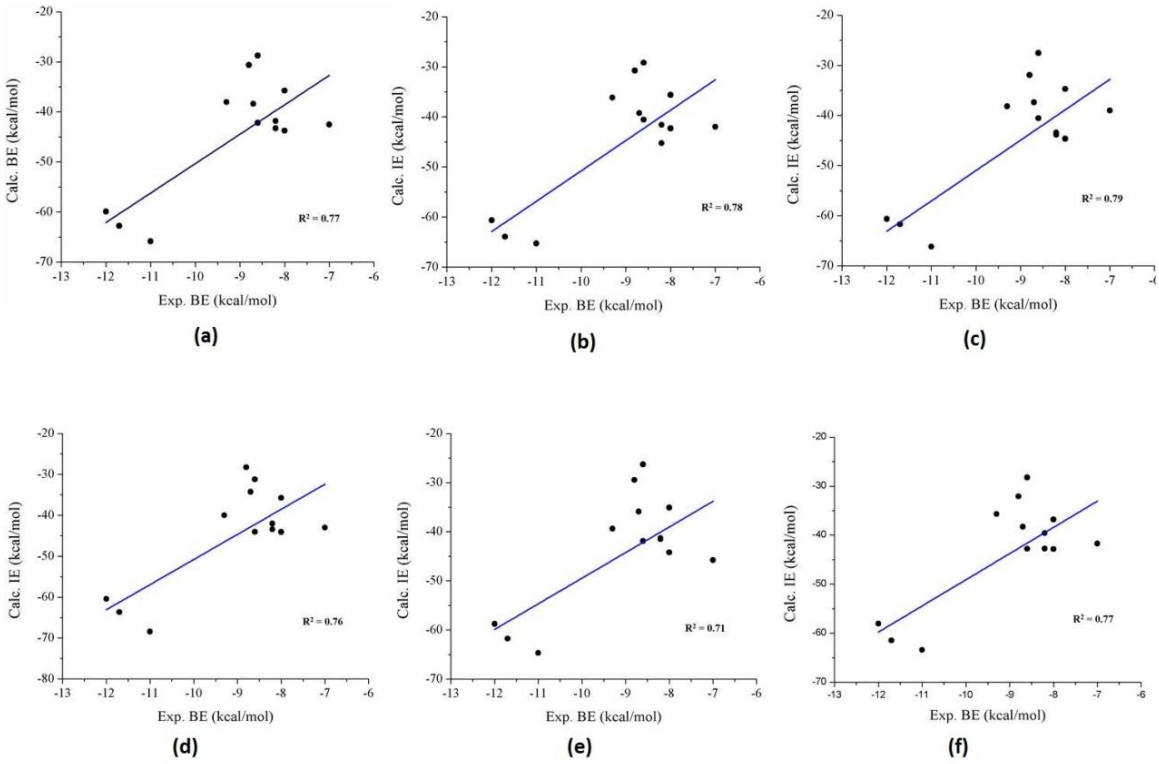


Fig. 5.12: Variation of calculated binding energy with the experimental binding energy obtained from MMPBSA calculation using GROMACS at (a) 0ns (b) 1ns (c) 2ns (d) 3ns (e) 4ns (f) 5ns.

Table 5.2: Interaction energies of DNA-Ligand complexes at various time scales and comparison with experimental binding energy (ΔG_{bind}).

PDB Id	Exp. BE	IE					
		0-1ns	1-2ns	2-3ns	3-4ns	4-5ns	0-5ns
127D	-11.0	-63.41±3.22	-65.25±2.93	-66.16±2.83	-68.43±2.95	-64.65±2.87	-65.83±3.42
264D	-11.7	-61.48±3.47	-63.93±3.31	-61.70±3.12	-63.66±3.58	-61.76±3.51	-62.75±3.58
1D86	-8.8	-32.10±6.16	-30.74±3.58	-31.95±4.74	-28.25±4.24	-29.46±4.15	-30.62±4.90
121D	-8.6	-42.79±3.45	-40.55±3.67	-40.55±4.04	-44.05±4.20	-41.88±5.01	-42.17±4.35
195D	-8.7	-38.27±4.33	-39.24±5.65	-37.38±6.42	-34.27±4.24	-35.90±4.82	-37.01±5.45
109D	-12.0	-58.04±3.12	-60.64±2.73	-60.65±2.78	-60.45±2.46	-58.75±2.67	-59.90±2.98
1D64	-7.0	-41.73±3.13	-41.96±3.65	-39.01±3.59	-43.02±3.63	-45.80±2.76	-42.51±4.04
227D	-9.3	-35.68±3.34	-36.16±4.68	-38.16±3.11	-39.99±2.99	-39.36±3.24	-37.87±3.91
1PRP	-8.2	-39.58±2.78	-41.57±2.71	-43.41±2.67	-43.41±2.67	-41.28±2.81	-41.80±3.05
1D30	-8.0	-36.82±2.38	-35.60±3.01	-34.71±2.24	-35.75±2.50	-35.10±2.44	-35.76±2.64
1D63	-8.0	-42.85±3.32	-42.29±2.87	-44.64±3.00	-44.10±4.02	-44.21±2.96	-43.75±3.39
2DBE	-8.6	-28.22±3.43	-29.17±3.57	-27.55±5.38	-31.20±4.07	-26.30±3.26	-28.74±3.98
102D	-8.2	-42.75±2.85	-45.24±2.69	-43.85±2.64	-41.99±2.70	-41.52±2.62	-43.28±3.03

All the energy values are in kcal/mol.

5.3.4 QM/MM Results

Interaction energies at each nanosecond were calculated using earlier discussed ONIOM scheme by employing two different methods B3LYP and M062X. The geometry optimization was performed using the basis set 6-31G* for the QM region and Amber force field for MM region. The optimized geometry of complex having PDB Id 109D, 127D is shown in Fig. 5.12. The single point calculation was performed using two method B3LYP with basis set 6-31G**(d,p) and M062X method. The interaction energies were calculated using the equation 5.6**.** The calculated interaction energies from QM/MM calculation shows good correlation with experimental binding energies at 1ns snapshots from both B3LYP and M062X method for QM after that the stability of DNA-ligand complex start to decrease

shown in Fig. 5.13 and Fig. 5.14. The calculated interaction energies from ONIOM scheme are listed in Table 5.3 and Table 5.4. The calculated interaction energy of DNA-ligand complexes increases as the experimental binding energy increases. It was observed that the interaction energy of DNA-Ligand complexes does not only depend on the chemical structure of ligand but also depends on the DNA sequence and specificity. The different interaction energies were observed in the interaction of Hoechst with two different DNA sequences, this is due to difference in the DNA sequences and similarly for the case of Netropsin, Berenil and propamidine. It was observed that for the same ligand the interaction energy of ligand with AT-rich DNA is more stable than that of CG-rich DNA.

Table 5.3: Calculated Interaction energies of complexes at each nanosecond obtained from MD trajectories using B3LYP method for QM region.

PDB Id	Exp. BE	0ns	1ns	2ns	3ns	4ns	5ns
127D	-11.0	-47.07	-71.54	-54.20	-66.93	4.86	-72.06
264D	-11.7	-59.37	-93.36	-34.70	-63.64	-61.28	-64.80
1D86	-8.8	-65.82	-49.34	-28.95	-55.57	-40.79	-63.87
121D	-8.6	-95.46	-58.55	-65.75	-64.06	-75.30	-89.37
195D	-8.7	-32.60	-31.95	-76.12	-52.13	-99.44	-55.74
109D	-12.0	-51.98	-75.94	-75.87	-72.31	-60.89	-52.50
1D64	-7.0	-50.76	-33.67	-51.95	-56.61	-55.99	-31.68
227D	-9.3	-33.81	-42.15	-38.81	-23.84	-47.98	-44.56
1PRP	-8.2	-44.31	-41.30	-46.67	2.63	-38.09	-24.70
1D30	-8.0	-29.48	-29.28	-53.31	-59.61	-43.65	-31.94
1D63	-8.0	-30.47	-42.96	-66.86	-30.52	-38.14	-46.83
2DBE	-8.6	-26.46	-16.58	-14.43	-26.43	-53.36	-47.74
102D	-8.2	-50.84	-43.50	-56.13	-67.57	-58.78	-44.55

All energies in kcal/mol.

Table 5.4: Calculated Interaction energies of complexes at each nanosecond obtained from MD trajectories using B3LYP method for QM region.

PDB Id	Exp. BE	0ns	1ns	2ns	3ns	4ns	5ns
127D	-11.0	-31.94	-52.30	-39.59	-50.93	8.39	-54.08
264D	-11.7	-43.35	-70.32	-25.53	-48.41	-42.62	-46.06
1D86	-8.8	-41.20	-28.19	-5.83	-31.13	-18.67	-34.66
121D	-8.6	-43.04	-15.73	-52.45	-30.99	-66.00	-30.83
195D	-8.7	-56.48	-33.57	-38.75	-33.45	7.22	-52.47
109D	-12.0	-37.76	-59.77	-59.77	-53.78	-47.10	-37.80
1D64	-7.0	-3.12	-21.52	-37.92	-40.62	-36.49	-31.70
227D	-9.3	-23.15	-29.87	-25.41	-11.30	-31.44	-30.81
1PRP	-8.2	-32.32	-24.86	-33.15	3.80	-23.91	-15.22
1D30	-8.0	-18.43	-20.57	-37.04	-40.04	-28.72	-21.08
1D63	-8.0	-20.76	-28.12	-47.01	-22.73	-27.64	-32.86
2DBE	-8.6	-17.06	-5.68	-5.14	-14.21	-33.45	-32.73
102D	-8.2	-35.77	-28.47	-38.30	-47.20	-43.31	-33.28

All energies in kcal/mol.

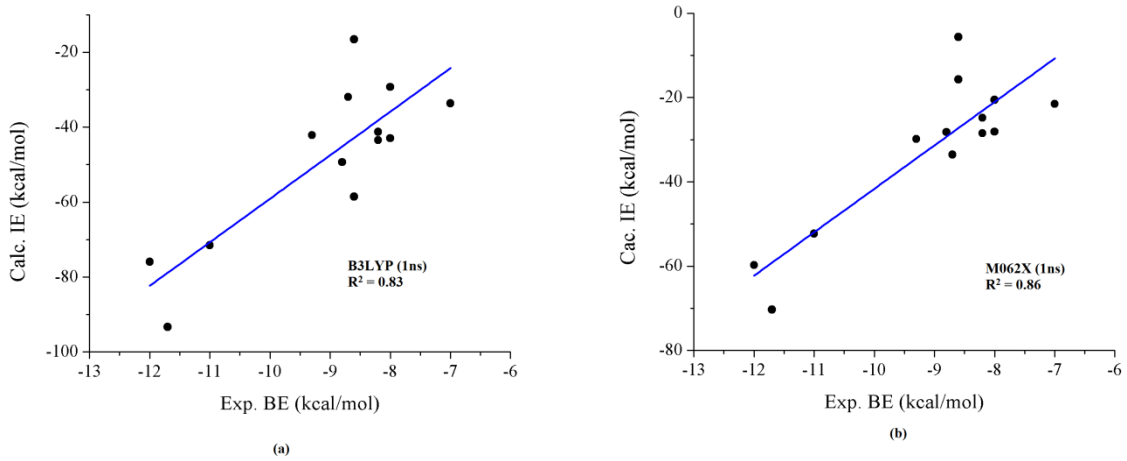


Fig. 5.13: Variation of interaction energy with experimental binding energy from QM/MM calculation using (a) B3LYP (b) M062X method for QM region.

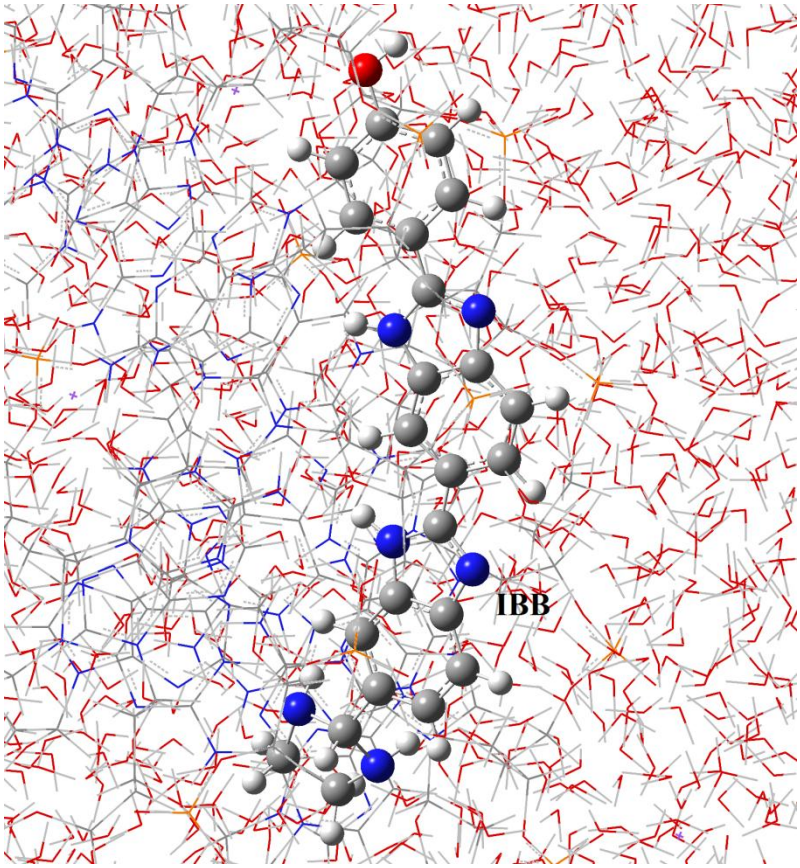


Fig. 5.14: ONIOM optimized structure of complex having PDB Id 109D.

5.4 Conclusions

Minor groove binders are of high pharmaceutical potential application because of their mode of action towards anticancer, antiviral, antibacterial activities are directly related to their binding with nucleic acid. A computational approach that combines with quantum mechanics, molecular docking, and molecular dynamics simulation predict the energetic pattern of the DNA-Ligand binding. Molecular Docking helps able to predict the perfect binding pockets and MD simulation results highly supports the predicted binding site. RMSD analysis also indicates that none of the compounds are detached from the system throughout the simulation. This study reveals the role of different types of non-covalent interactions such as van der Waals, electrostatic, hydrophobic, hydrogen bonding interactions etc. and it was analyzed that van der

Waals interactions are strongly linked to the drug-DNA binding. QM/MM calculations also predict the similar binding mode predicted from Docking and Molecular dynamics. Therefore this study, confirm the importance of DNA sequence and specificity in directing the complex formation and might provide novel insight into the mechanism of Drug-DNA interactions at molecular level.

References

[1] Neidle, S., Nunn, C.M., *Nat. Prod. Rep.*, **2**, 1 (1998).

[2] Neidle, S., *Nat. Prod. Rep.*, **18**, 291(2001).

[3] Wemmer, D.E., *Biomol. Struct*., **29**, 439 (2000).

[4] Ren, J., Chaires, J.B., *Biochem.*, **38**, 16067 (1999).

[5] Bailly, C., Chaires, J.B., *Bioconjugate Chem.*, **9**, 513 (1998).

[6] Watson, J.D., Crick, F.H.C., *Nature*, **171**, 737 (1953).

[7] Dervan, P.B., *Bioorg. Med. Chem.*, **9**, 2215 (2001).

[8] Gibson, D., *Pharmacogenomics J.*, **2**, 275 (2002).

[9] Pindur, U., Jansen, M., Lemster, T., *Curr. Med. Chem.*, **12**, 2805 (2005).

[10] Lerman, L.S., *Biochem.*, **49**, 95 (1963).

[11] Wilson, W.D., Tanious, F.A., Mathis, A., Tevis, D., Hall, J.E., Boykin, D.W., *BioChimie.*, **90**, 999 (2008).

[12] Demeunynck, M., Bailly, C., Wilson, W.D., *Small Molecule DNA and RNA Binders: From Synthesis to Nucleic Acid Complexes*, Volume 2, 3-527-30595-5.

[13] Boykin, D.W., *J. Braz. Chem. Soc.*, **13**, 763 (2002).

[14] Cai, X., Jr., P.J. G., Hoff, D.D.V., *Cancer Treat. Rev.*, **35**, 437 (2009).

[15] Wilson, W.D., Nguyen, B., Tanious, F.A., Mathis, A., Hall, J.E, Stephens, C.E., Boykin, D.W., *Current Medicinal Chemistry-Anti-Cancer Agents*, **5**, 389 (2005).

[16] Karisson, H.J., Eriksson, M., Perzon, E., Akerman, B., Lincoln, P., Westman, G., *Nucleic Acid Res.*, **31**, 6227 (2003).

[17] Bostock-Smith, C.E., Searle, M.S., *Nucleic Acid Res.*, **27**, 1619 (1999).

[18] Alpan, A.S., Yilmaz, O., Kilickaya, O., Kara, P., Gunes, H.S., Ozsoz, M.S., *Marmara Pharmaceutical Journal*, **16**, 48 (2012).

[19] Sriram, M., Marel, G.A.V.D., Roelen, H.L.P.F, Boom, J.H.V.,Wang, A.H.J., *The EMBO J.*, **11**, 225 (1992).

[20] Wood, A.A., Nunn, C.M., Czarny, A., Boykin, D.W., Neidle, S., *Nucleic Acid Res.*, **23**, 3678 (1995).

[21] Zimmer, C., Wahnert, U., *Prog. Biophys. Molec. Biol.*, **47**, 31 (1986).

[22] Tabernero, L., Verdaguer, N., Coll, M., Fita, I., Marel, G.A.V., Boom, J.H.V., Rich, A., Aymamf, J., *Biochem.*, **32**, 8403 (1993).

[23] Sriram, M., Marel, G.A.V.D., Roelen, H.L., Boom, J.H.V., Wang, A.H.J., *Biochem.*, **31**, 11823 (1992).

[24] Balendiran, K., Rao, S.T., Sekharudu, C.Y., Zon, G., Sundralingam, M., *Acta Cryst.*, **D51**, 190 (1995).

[25] Brown, D.G., Sanderson, M.R., Garman, E., Neidle, S., *J. Mol. Biol.*, **226**, 481 (1992).

[26] Brown, D.G., Sanderson, M.R., Skelly, J.V., Jenkins, T.C., Brown, T., Garman, E., Stuart, D.I., Neidle, S., *EMBO J.*, **9**, 1329 (1990).

[27] Laughton, C.A., Tanious, F., Nunn, C.M., Boykin, D.W., Wilson, W.D., Neidle, S., *Biochem.*, **35**, 5655 (1996).

[28] Greenhill, J.V., *Prog. Med. Chem.*, **30**, 206 (1993).

[29] Pilch, D.S., Kirolos, M.A., Breslauer, K.J., *Biochem.*, **34**, 16107 (1995).

[30] Kapuscinski, J., *Biotechnic and Histochemistry*, **70**, 220 (1995).

[31] Estandarte, A.K., Botchway, S., Lynch, C., Yusuf, M., Robinson, I., *Scientific Reports*, **6**, 1 (2016).

[32] Larsen, T.A., Goodsell, D.S., Cascio, D., Grzeskowiak, K., Dickerson, R.E., *J. Biomol. Struc. Dyn.*, **7**, 477 (1989).

[33] Kapuscinski, J., Yanagi, K., *Nucleic Acid Res.*, **6**, 3535 (1979).

[34] Wilson, W.D., Tanious, F.A., Barton, H.J., Jones, R.L., Fox, K., Wydra, R.L., Strekowski, L., *Biochem.*, **29**, 8452 (1990).

[35] Vlieghe, D., Sponer, J., Meervelt, L., *Biochem.*, **38**, 16443 (1999).

[36] Edwards, K.J., Jenkins, T.C., Neidle, S., *Biochem.*, **31**, 7104 (1992).

[37] Boykin, D.W., *J. Braz. Chem. Soc.*, **6**, 763 (2002).

[38] Walzer, P.D., Kim, C.K., Foy, J., Linke, M.J., Cushion, M.T., *Antimicrob. Agents Chemother.*, **32**, 896 (1998).

[39] Poeta, M.D., Schell, W.A., Dykstra, C.C., Jones, S., Tidwell, R.R., Czarny, A., Bajic, M., Bajic, M., Kumar, A., Boykin, D., Perfect, J.R., *Antimicrob. Agents Chemother.*, **42**, 2495 (1998).

[40] Cushion, M.T., Walzer, P.D., Collins, M.S., Rebholz, S., Eynde, J.J.V., Mayence, A., Huang, T.L., *Antimicrob. Agents Chemother.*, **48**, 4209 (2004).

[41] Nunn, C.M., Jenkins, T.C., Neidle, Stephen, *Biochem.*, **32**, 13838 (1993).

[42] Nunn, C.M., Neidle, S., *J. Med. Chem.*, **38**, 2317 (1995).

[43] Srivastava, H.K., Chourasaia, H., Kumar D., Sastry G.N., J. *Chem. Inf. Model.*, **51**, 558 (2011).

[44] Kamal, A., Khan, M.N.A., Reddy, K.S., Rohini, K., Sastry, G.N., Sateesh, B., Sridhar, B., *Bioorg. Med. Chem. Lett.*, **17**, 5400 (2007).

[45] Kamal, A., Rajender, Reddy, D.R., Reddy, M.K., Balakishan, G., Shaik, T.B., Chourasia, M., Sastry, G.N., *Bioorg. Med. Chem.*, **17**, 1557 (2009).

[46] Kamal, A, Shankaraiah, N., Reddy, Ch. R., Prabhakar, S., Markandeya, N.; Srivastava, H.K.; Sastry, G.N., *Tetrahedron*, **66**, 5498 (2010).

[47] Kamal, A., Reddy, K.S., Khan, M.N.A., Shetti, R.V.C.R.N.C., Ramaiah, M.J., Pushpavalli, S.N.C.V.L., Srinivas, C., Pal-Bhadra, M., Chourasia, M., Sastry,

G.N., Juvekar, A., Zingde, S., Barkume, M., *Bioorg. Med. Chem.*, **18**, 4747 (2010).

[48] Warshel, A., Levitt, M., *J. Mol. Biol.*, **103**, 227 (1976).

[49] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., *Nucleic Acid Res.*, **28**, 235 (2000).

[50] Vega, M.C., Saez, I.G., Aymami, J., Eritja, R., Marel, G.A.V., Boom, J.H., Rich, A., Coll, M., *Eur. J. Biochem.*, **222**, 721 (1994).

[51] Clark, G.R., Squire, C.J., Gray, E.J., Leupin, W., Neidle, S., *Nucleic Acid Res.*, **24**, 4882 (1996).

[52] Nunn, C.M., Jenkins, T.C., Neidle, S., *Eur. J. Biochem.*, **226**, 953 (1994).

[53] Tabernero, L., Verdaguer, Coll, M., Fita, I., Marel, G.A.V, Boom, J.H., Rich, A., Aymami, J., *Biochem.*, **32**, 8403 (1993).

[54] Gaussian 09, Revision A.02, Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G.A., Nakatsuji, H., Caricato, M., Li X., Hratchian, H.P., Izmaylov, A.F., Bloino, J., Zheng, G., Sonnenberg, J.L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J.A., Jr., Peralta, J.E., Ogliaro, F., Bearpark, M., Heyd, J.J., Brothers, E., Kudin, K.N., Staroverov, V.N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J.C., Iyengar, S.S., Tomasi, J., Cossi, M., Rega, N., Millam, J.M., Klene, M., Knox, J.E., Cross, J.B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli, C., Ochterski, J.W., Martin, R.L., Morokuma, K., Zakrzewski, V.G., Voth, G.A., Salvador, P., Dannenberg,

J.J., Dapprich, S., Daniels, A.D., Farkas, O., Foresman, J.B., Ortiz, J.V., Cioslowski, J., Fox D.J., Gaussian, Inc., Wallingford CT, 2009.

[55] Morris, G.M., Huey, R., Lindstorm, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J., *J. Compt. Chem.*, **30**, 2785 (2009).

[56] Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., Spoel, D., Hes, B., Lindahl, E., *Bioinformatics*, **29**, 845 (2013).

[57] Kumari, R., Kumar, R., Lynn, A., *J. Chem. Inf. Model.*, **54**, 1951 (2014).

Chapter 6

Cross Molecular Docking and Molecular Dynamics Studies Of DNA Binding Ligands

Chapter 6

Cross Molecular Docking and Molecular Dynamics Studies Of DNA Binding Ligands

6.1 Introduction

Large number of biological structures in databases is increasing rapidly, molecular docking is a popular approach to deal with conformation or even to explain the interaction between potential drug molecules and their bio molecular targets.[1,2] Many molecular methods are available but only some of them are able to correctly reproduce the binding mode of co-crystallized ligand with their respective targets but none of them can be applicable universally.[3] Groove binding and intercalation are the two important mode of binding of small molecules to the DNA.[4] Groove binding molecules are flexible and contain rotatable bonds as they can easily occupy the shape of groove (major or minor), thereby inhibiting its regular function.[5,6] However, major groove of DNA having more H-bonding donor and acceptor sites provide shallow binding space for the incoming drug molecule than the minor groove. In spite of this almost all the small molecules interact with the minor groove while proteins and peptide binds to the major groove of DNA.[7, 8] The DNA intercalators are small and rigid aromatic molecules that stack-in between the DNA base pairs causing structural alterations to DNA to create an intercalation gap between two consecutive DNA base pairs.[9-12]

There is large number of docking studies reported for protein-ligand as compared to the nucleic acid-ligand complexes; though DNA is the major target for many anticancer,

antiviral, antibiotics, antitumor and antifungal drugs.[13-18] Mostly all the scoring function are parameterized for protein-ligand complexes and docking programs are validated for protein-ligand docking.[3, 4] The structural features of nucleic acid is unique such as high density charge and helix chiral geometry and nucleic acid do not have any predefined binding site which is present in proteins.[3] This leads to a question whether the docking programs which are applicable for the protein-ligand docking can also be able to reproduce results for nucleic acid-ligand or not. This matter has been taken care by Holt et al. in 2008, who predicted that Autodock and Surflex docking programs can reproduce the crystal structure of minor groove binders and intercalators with in the resolution of 2Å.[19] Self-dockings (i.e. original crystallographic target) are able to provide the information about docking accuracy in real drug discovery.[3] The critical issue arises when the binding mode of ligand is unknown to the DNA and there is no previous experimental data available, and then it is difficult to select which oligomer conformation should be used as target?

To address the above mentioned issue Ricci and Netz performed self and cross docking studies using Autdock 4.0 software. The two ligands one is known minor groove binder (Netropsin) and other is known intercalator (ellipcitne) docked with four different DNA sequences: 1) crystallographic DNA obtained from the PDB Id 1DNE (DNA without intercalation gap); 2) crystallographic DNA obtained from the PDB Id 1Z3F; 3) canonical B-DNA; and 4) modified B-DNA without intercalation gap. These docking studies reveal that the current limitations of docking methods can be overcome by a proper choice of the target conformation. It is observed that if the target DNA has an artificial intercalation gap, then the docking is able to predict the correct binding mode which is energetically

most favorable; which suggests that the AutoDock score function is efficient to evaluate ligand-DNA interactions at least in a qualitative way.[20] Later, in 2015, Mariya et. al. concluded that if the nature of ligand with DNA is not known, the exact mode of binding of ligand to DNA cannot be predicted on the basis of molecular docking as a result other molecular modelling techniques such as molecular dynamics simulation and thermodynamics integration will be required to further resolve the problem.[21]

In the present study some of the DNA binding ligands which are known intercalator or minor groove binders are used and the two DNA sequences are taken, 1) with intercalation gap (PDB Id 2DES) ; 2) without intercalation gap (PDB Id 1BNA). Molecular modelling studies such as molecular docking, MD simulation, MMPBSA and MMGBSA calculations and QM/MM studies etc. were performed to predict the binding mode, to check the stability of the complex and to calculate the interaction energy.

6.2 Material and Methods

The chemical structure of known minor groove binders and intercalators which were used for the study are shown in Fig. 6.1. The experimental values for the ligands were obtained from literature shown in Table 6.1. These ligands were optimized using Gaussian 09 [22] at B3LYP/6-31G* level. The two DNA sequences with intercalation gap (PDB Id 1DSC) and without intercalation gap (PDB Id 1BNA) were obtained from the Protein Data Bank.

PDB Id	**DNA Sequence**
1BNA	5’-CGCGAATTCGCG-3’
1DSC	5’-GAAGCTTC-3’(I.G)

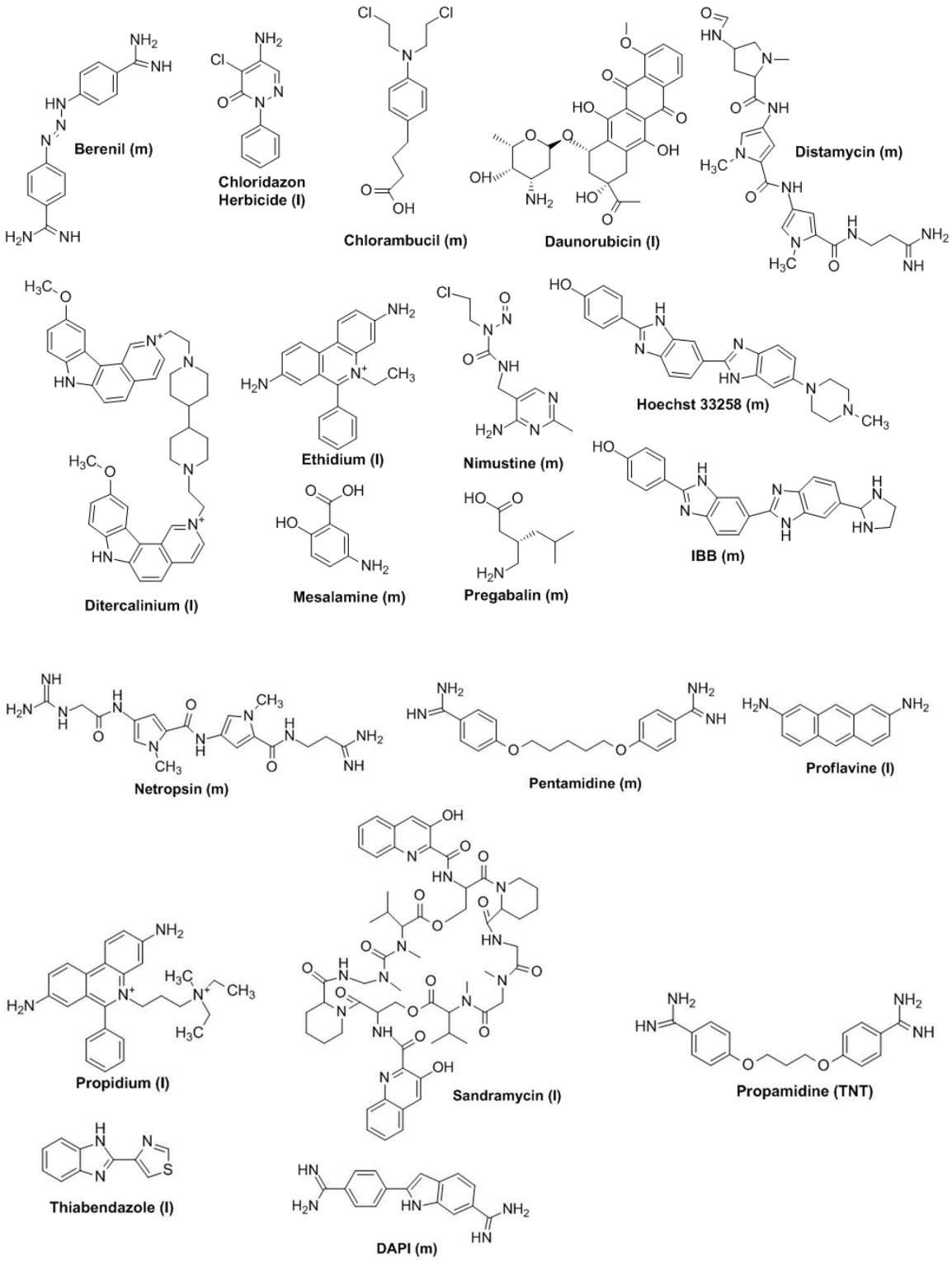


Fig. 6.1: Chemical Structure of DNA binding Ligands with their binding mode.

Table 6.1: Comparison of experimental and calculated ΔG_{bind} obtained from molecular docking studies and MMPBSA calculations using GROMACS.

Ligand	Exp. B.E.	Docking		MMPBSA		References
		6mer	12mer	6mer	12mer	
Berenil (m)	-8.0	-6.82 (m)	-8.57 (m)	-28.33±3.84	-35.55±3.11	[23]
Chloridazon Herbicide (I)	-5.91	-6.08 (m)	-7.14 (m)	-24.86±3.09	-25.37±3.03	[24]
Cholrambucil (m)	-4.26	-3.67 (m)	-6.04 (m)	-25.66±4.51	-23.28±5.58	[25]
Daunorubicin (I)	-7.9	-9.26 (I)	-11.07 (m)	-55.28±4.8	-35.42±10.92	[26]
Ditercalinium (I)	-11.9	-11.68 (I)	-13.21 (m)	-338.52±7.17	-469.15±7.03	[27]
Distamycin (m)	-10.5	-5.82 (m)	-8.4 (m)	-24.17±10.37	-24.25±8.11	[23]
Ethidium (I)	-6.7	-7.28 (I)	-8.23 (m)	-190.82±3.58	-242.81±3.57	[26]
Hoechst (m/I)	-11.7	-9.34 (m)	-13.93 (m)	-16.41±9.62	-43.51±7.71	[28]
IBB (m)	-12.0	-10.04 (m)	-14.59 (m)	-50.34±7.05	-54.20±9.86	[29]
Mesalamine (m)	-4.2	-3.77 (m)	-5.4 (m)	-18.59±5.61	-21.17±2.44	[30]
Nimustein (m)	-4.8	-5.8 (m)	-7.39 (m)	-24.14±3.76	-19.73±7.03	[31]
Netropsin (m)	-8.7	-6.85 (m)	-8.35 (m)	-31.91±4.92	-36.92±7.18	[23]
Pentamidine (m)	-7.0	-5.84 (m)	-8.23 (m)	-22.91±6.10	-41.31±3.76	[32]
Pregabalin (m)	-6.5	-4.76 (m)	-5.22 (m)	-5.80±10.29	-14.13±3.15	[33]
Proflavin (I)	-8.7	-6.31 (I)	-8.22 (m)	-182.71±3.21	-241.90±2.63	[34]
Propidium (I)	-7.5	-7.29 (I)	-7.72 (m)	-30.66±5.40	-29.01±6.57	[26]
Sandramycin (I)	-10.3	-9.34 (I)	-6.52 (m)	-43.30±3.2	-40.05±8.09	[35]
Thiabendazole (I)	-5.2	-5.4 (m)	-7.61 (m)	-4.46±9.00	-28.29±7.78	[36]
Propamidine (m)	-7.0	-6.5 (m)	-9.22 (m)	-35.08±5.93	-46.10±3.84	[23]
DAPI(m)	-8.0	-6.86 (m)	-9.55 (m)	-29.59±5.19	-39.83±2.23	[37,38]

6.2.1 Molecular Docking Studies

Ligands were docked with the two types of DNA: with intercalation gap (PDB Id 2DES) and without intercalation gap (PDB Id 1BNA). Blind molecular docking calculations between drug and DNA sequence were performed using MGL tools 1.5.6 with AutoGrid4 and AutoDock4.[23] DNA was extracted from the above PDB Ids and all the heteroatoms and water molecules were deleted. A grid box of grid points 70×70×70 was created with grid spacing of 0.375Å in order to include the entire DNA fragment. After the creation of grid map, grid potential map was calculated using Auto Grid 4.0. Docking calculation were performed using Lamarckian genetic algorithms, as accomplished in Autodock. The 20 LGA runs with maximum of 25,00000 energy evaluations were performed. All other parameters were used with default settings. According to the scoring function, the lowest

energy docked conformation was selected as best docked pose for each docking case, this best docked pose were used for further study.

6.2.2 Molecular Dynamics Simulation

Molecular Dynamics simulation (MD) of 5ns for the each selected best poses was carried out using GROMACS 4.5.6. The pdb2gmx program of the GROMACS [24] package was used to generate the parameters and topology files for the DNA using AMBER94 force field while for the ligands these files were processed via the 'acpype.py' python script using generalized amber force field (GAFF), which "translated" them into GROMACS topology file formats. The coordinate and topology files of DNA and ligands were merged to obtain the final starting structure and topology file for each complex. The simulation system was constructed by immersing the macromolecules in a dodecahedron of explicit TIP3P water molecules. The drug-DNA complex was placed in the center of dodecahedron periodic box of size 1.2Å. The system was then solvated in TIP3P water molecules. The total charge on the system was then neutralized by adding counter ions. The energy was minimized using steepest descent algorithm. Then the system was heated to 300K during 50 ps of constant volume simulation with 2 fs time step. The pressure was equilibrated to 1 atm during 50 ps NPT simulation with 2 fs time step. In both the simulations a restrained with force constant of 1000kJ/(mol/nm^2) was used. Both temperature and pressure were regulated using Berendsen algorithm. Production simulations were performed 5ns with a 2 fs time step. The temperature and pressure were maintained at 300 K and 1 atm using the v-rescale temperature and Parninello-Rahman pressure coupling method. Root mean square deviation (RMSD), root mean square fluctuations (RMSF), radius of gyration (Rg) and hydrogen bond distribution for each

system was determined by the analysis of MD trajectories produced during the 5 ns of the production run at every 2 ps. Trajectories were visualized by means of VMD. All the graphs were plotted using GNUPLOT.

6.2.3 MMPBSA/MMGBSA Method

The binding energy of each DNA-ligand complex was determined using the g_mmpbsa [25] tool. From 5,000 snapshots structure extracted from the 5 ns time period. The binding free energy (ΔG_{bind}) was determined using the equations below-

$$\Delta G_{bind} = G_{complex} - (G_{receptor} + G_{ligand}) \quad (6.1)$$

$$\Delta G_{bind} = E_{gas} + G_{sol} - T\Delta S \quad (6.2)$$

$$E_{gas} = E_{\text{int}} + E_{vdw} + E_{ele} \quad (6.3)$$

$$G_{sol} = G_{pol} + G_{SA} \quad (6.4)$$

$$G_{SA} = SASA + b \quad (6.5)$$

where, $G_{complex}$, $G_{receptor}$ and G_{ligand} are the absolute free energies of the DNA-ligand complex, DNA and ligands respectively. The ΔG_{bind} was decomposed to its individual contributions (equations 6.1 to 6.5); gas-phase energy (E_{gas}) which is a sum of bonded (E_{int}) and nonbonded terms (E_{vdw} and E_{ele}); the solvation free energy (G_{sol}), polar (G_{pol}) and non-polar (GSA) solvation energy components, and an entropy terms ($T\Delta S$). Polar solvation energies were determined by solving the Poisson-Boltzmann linear equation while nonpolar solvation through the solvent accessible surface area with an offset value (b) of 3.84928 $kJ.mol^{-1}$ and surface tension proportionality (γ) set at 0.0226778 $kJ.mol^{-1}Å^{-2}$. The individual contributions of DNA residues to the three energetic components were determined through per-residue decomposition.

6.3 Result and Discussions

6.3.1 Molecular Docking Results

Molecular docking results show that the binding of any ligand to the DNA does not depend on chemical structure of ligand only but also on the DNA sequence and the conformation of DNA. An intercalator intercalates between the base pairs of DNA, if and only if, the DNA has a required intercalation gap between DNA base pairs, otherwise, the ligand will bind to DNA groove (major or minor). If DNA has intercalation gap then minor groove binder can also be able to bind as an intercalator. It can be further observed from Fig. 6.2 that binding energies calculated from docking followed the same trend as set up by the experimental binding energy values. This provides a good agreement of the theoretical calculations with the experimental results. The binding energy of ligands for the DNA without intercalation gap is always lower than that of the DNA with intercalation gap shown in Table 6.1. Molecular Docking studies reveal that all the minor groove binders are stabilized by hydrogen bond and hydrophobic interaction, while intercalators are stabilized by π-π interaction between nucleic acid bases and intercalating ligand which were shown in Fig. 6.3 for the ligand Berenil (minor groove binder) and DIT (intercalators). Intercalators intercalate between the DG4, DC5, DG12 and DC13 DNA residues of the DNA with intercalation gap. The interaction analyses between ligands reveal that the intercalators bind to CG-rich region while minor groove binders bind to the AT-rich region of DNA.

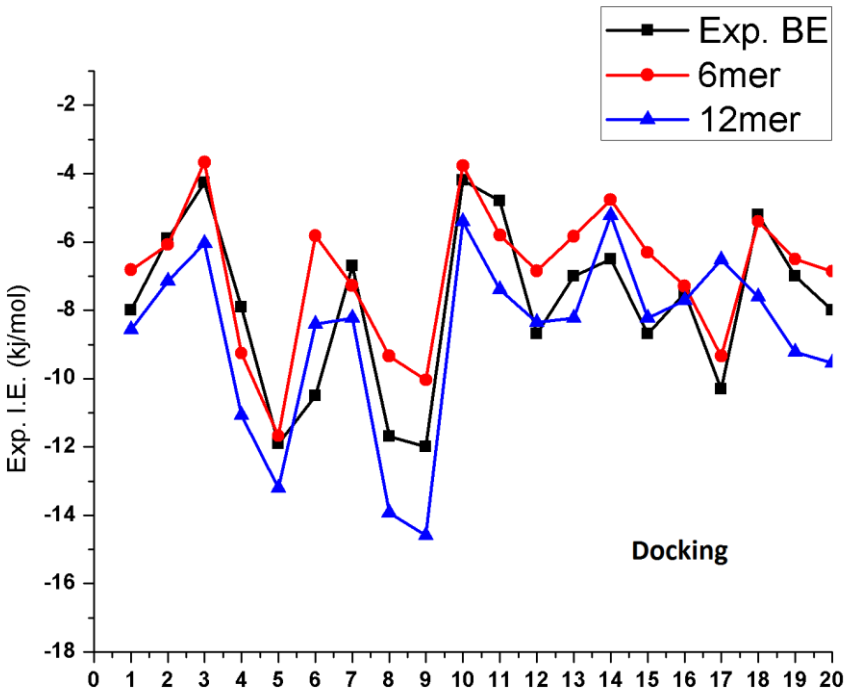


Fig. 6.2: Binding Energy Trend.

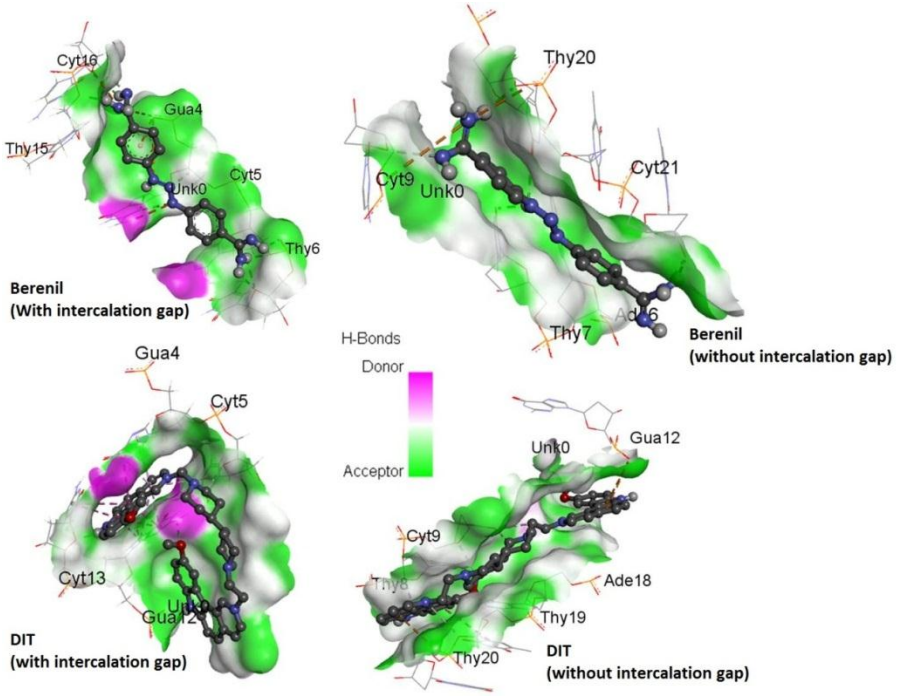
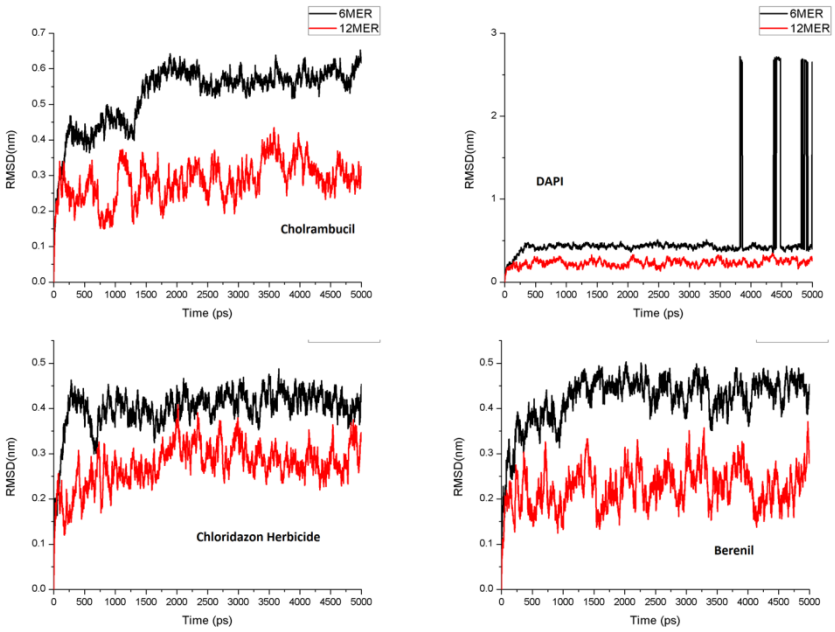
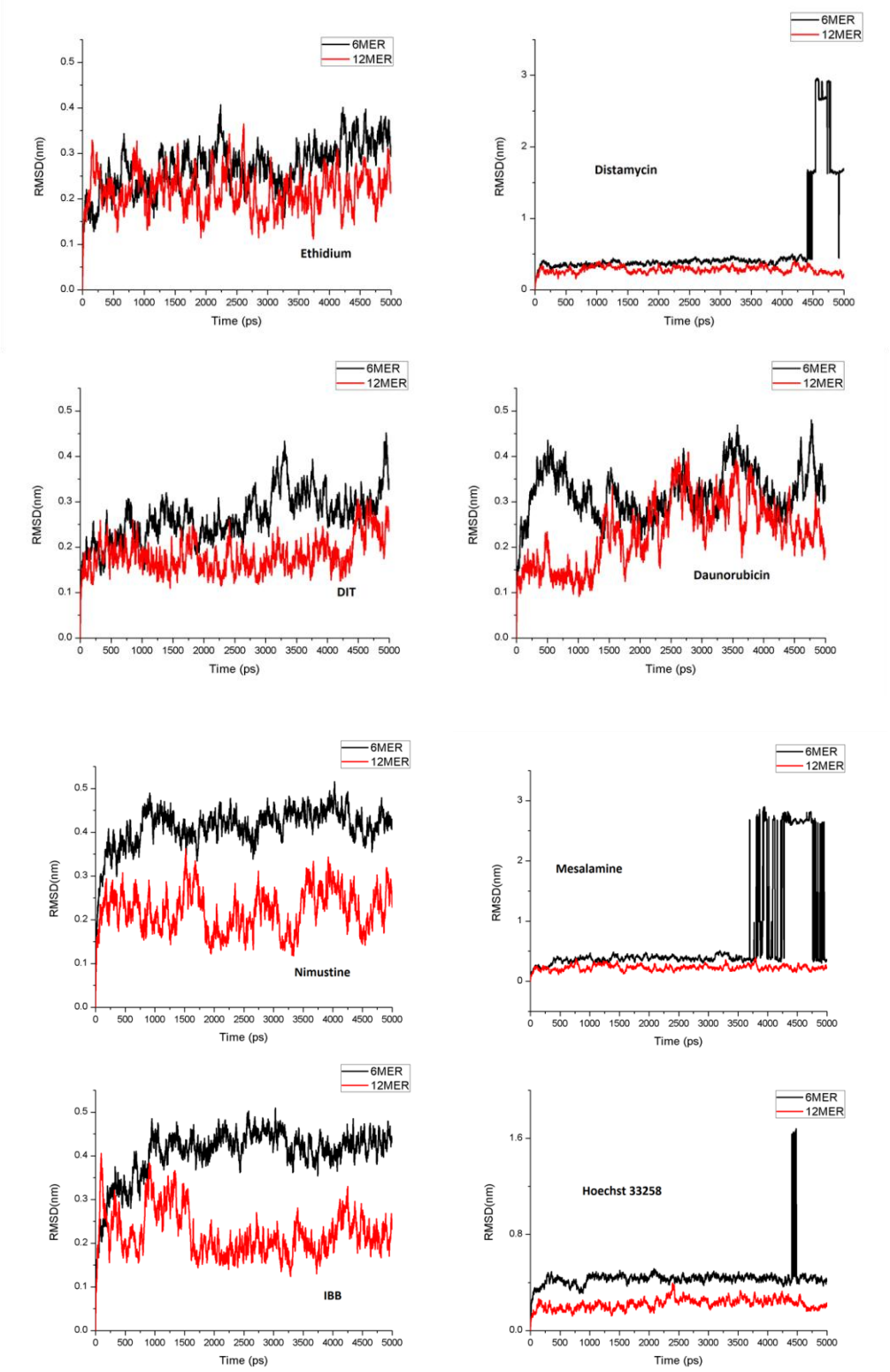


Fig. 6.3: Interaction of ligand DIT (intercalators) and Berenil (minor groove binder) with both types of DNA.

6.3.2 Molecular Dynamics Simulation Results

MD simulations of best docked poses obtained from molecular docking were performed by GROMACS using AMBER force field. The root mean square deviations (RMSD) have been calculated from starting structure to check the stability of the complex during MD simulation. The average value of RMSD is 0.2-0.3 nm for the DNA without intercalation gap, while for the DNA with intercalation gap the average value of RMSD is much higher. Convergence of RMSD shows the stability of the complex. The average rmsd value of complex of ligand DAPI, distamycin, mesalamine, hoeschst 33258, pregabalin to the DNA with intercalation gap are not converged throughout the simulation and the hydrogen bonding between the nucleic acid base pairs of two helix breakdown, while for the ligand thiabendazole which is an intercalator when formed complex with the DNA without intercalation gap, then the complex is not stable during MD simulation shown in Fig. 6.4.




6MER
12MER
RMSD(nm)
Time (ps)
Ethidium
Distamycin
DIT
Daunorubicin
Nimustine
Mesalamine
IBB
Hoechst 33258

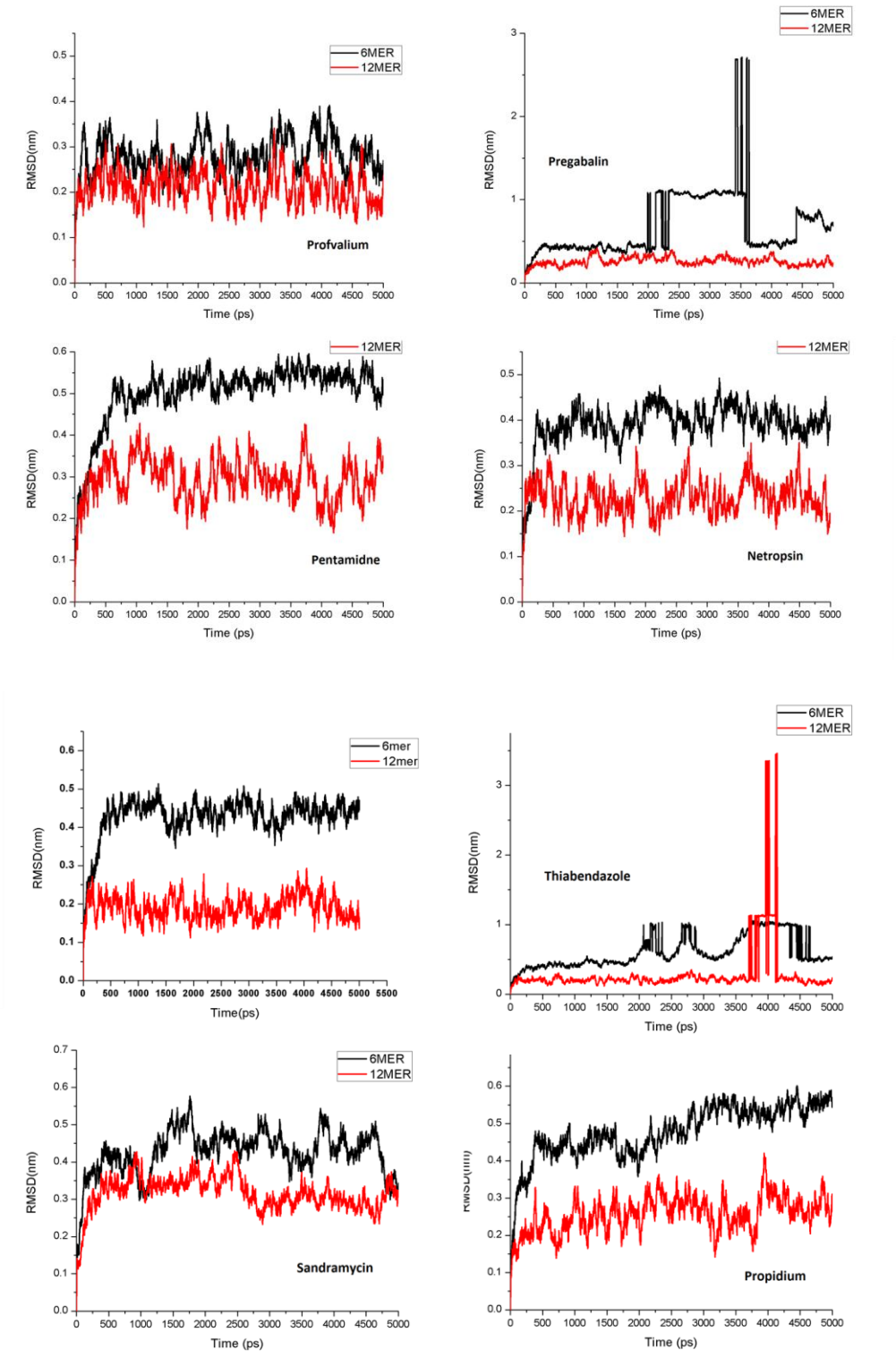


Fig. 6.4: RMSD Plots.

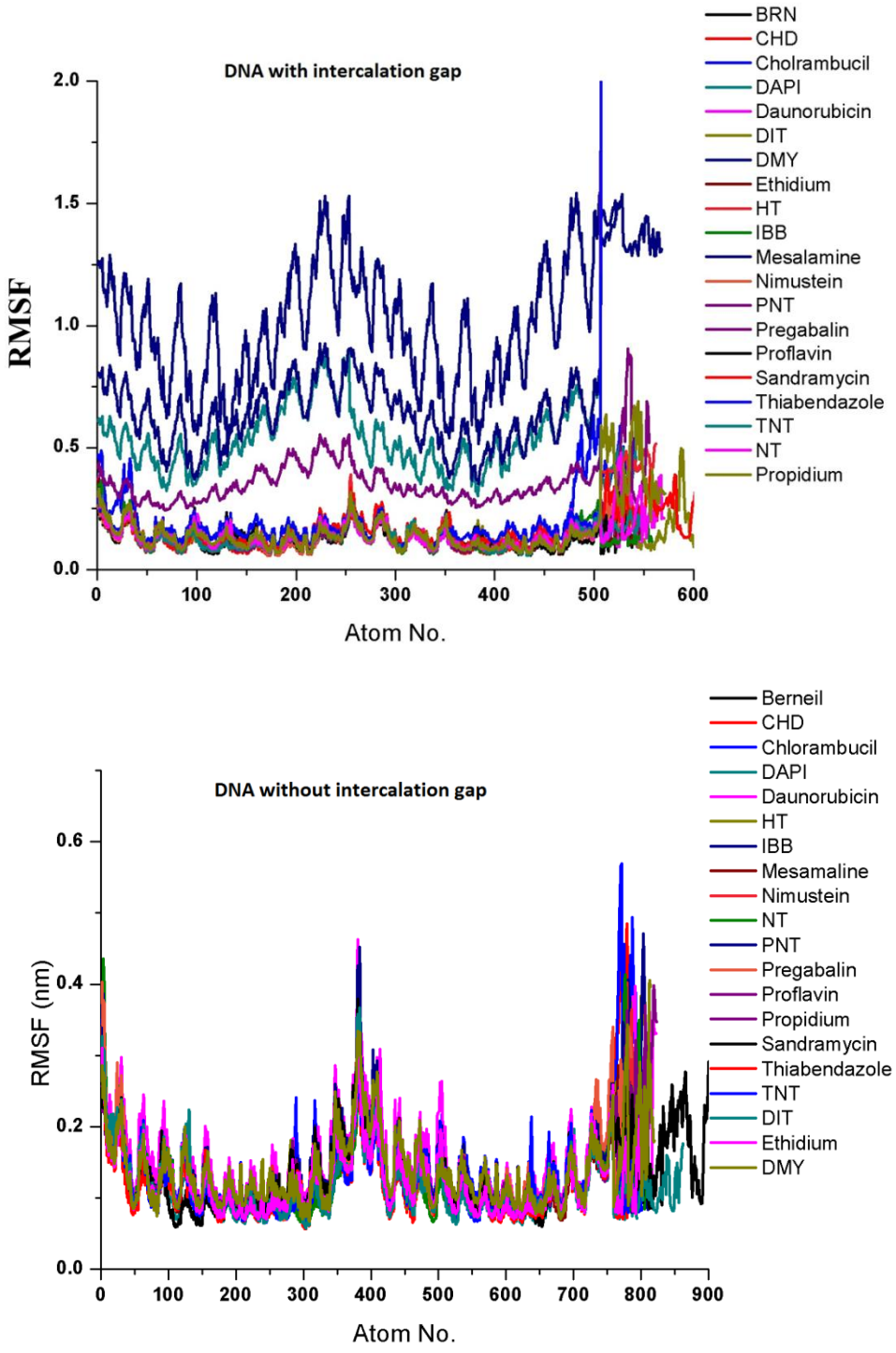


Fig. 6.5: RMSF plots.

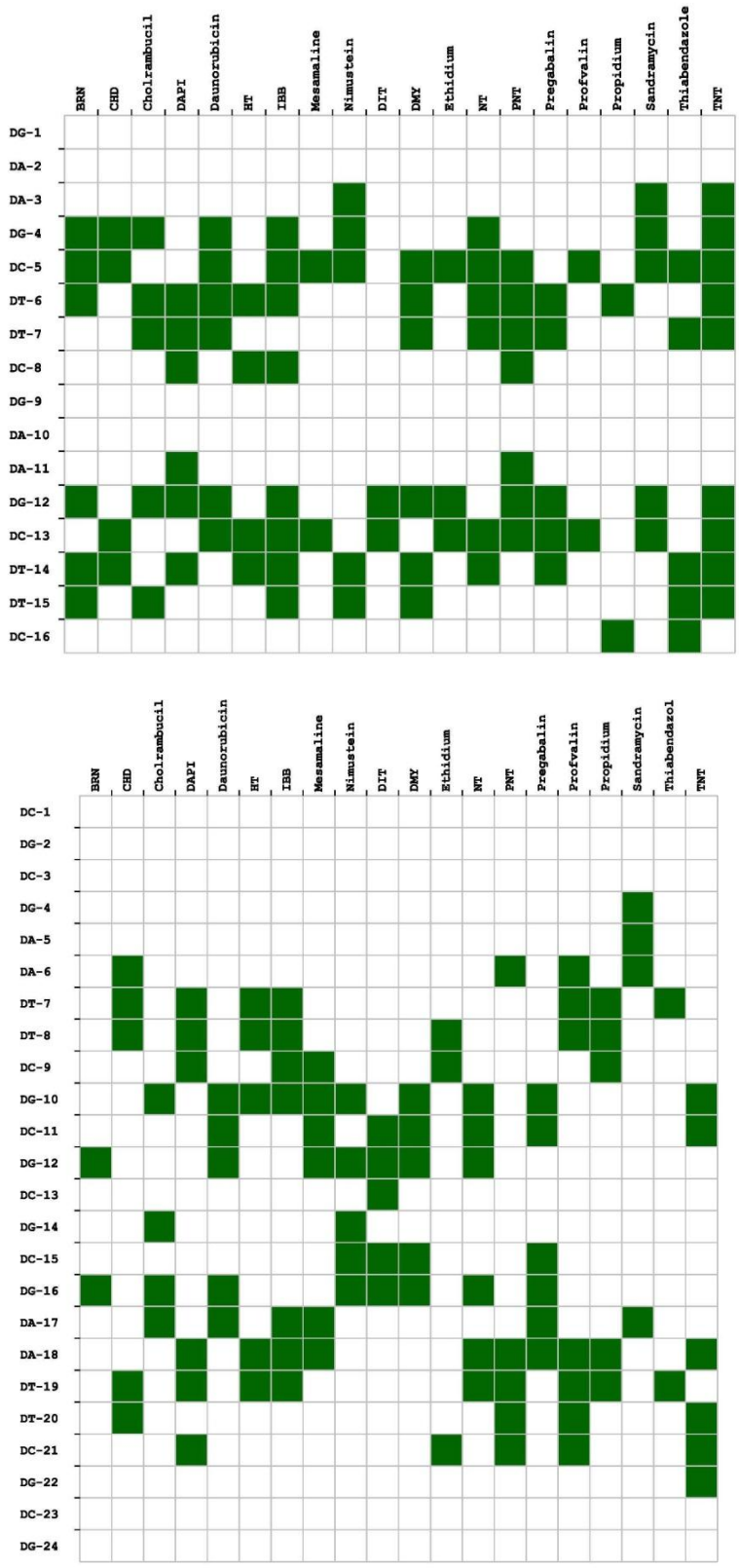


Fig. 6.6: Hydrogen bond analysis.

The root mean square fluctuation (RMSF) values are higher at the end terminal of nucleic acid and for the ligand shown in Fig. 6.5. The hydrogen bond analysis were performed and the analysis reveal that in the case of MD simulation with the DNA with intercalation gap the DNA residues involves in hydrogen bond interaction are DG4, DC5, DT6 and DT7, while for the DNA without intercalation gap the residues are DA6, DT7, DT8, DG10, DC11, DG12, DG16, DA17, DA18 and DT19 shown in Fig. 6.6.

6.3.3 MMPBSA Calculations

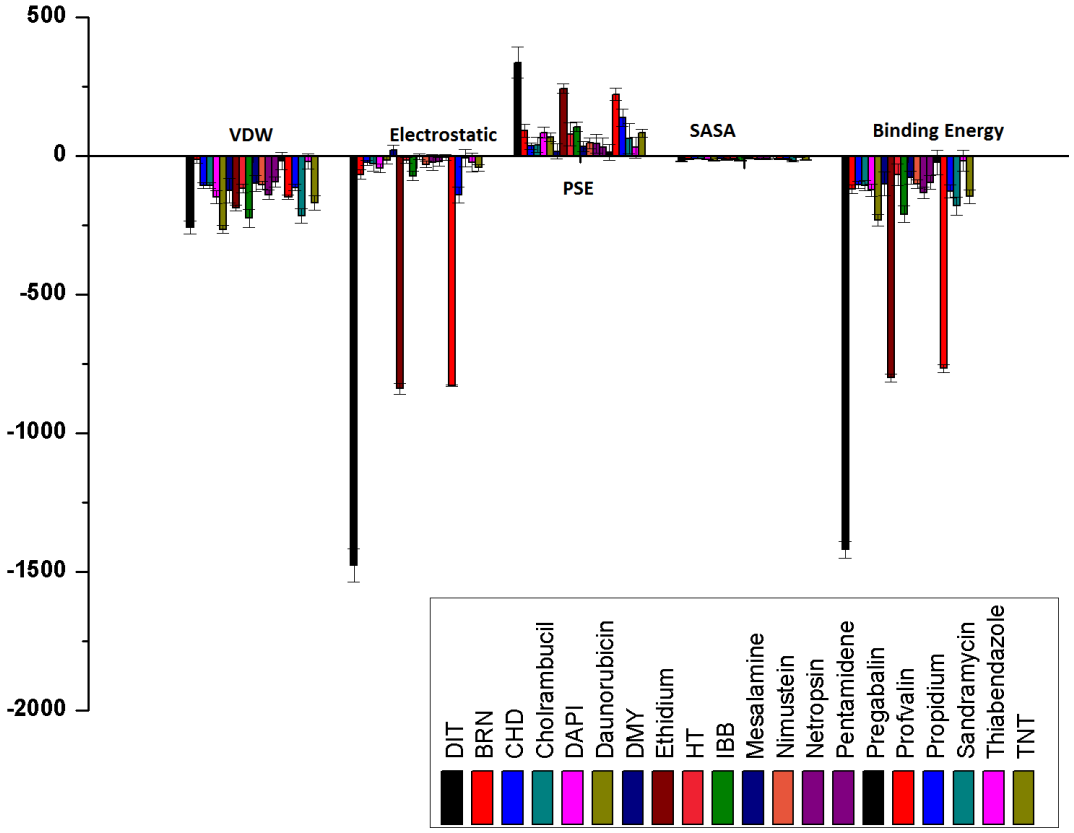


Fig. 6.7: Histogram depicting the view of contribution of each and every component of energy to the final binding energy.

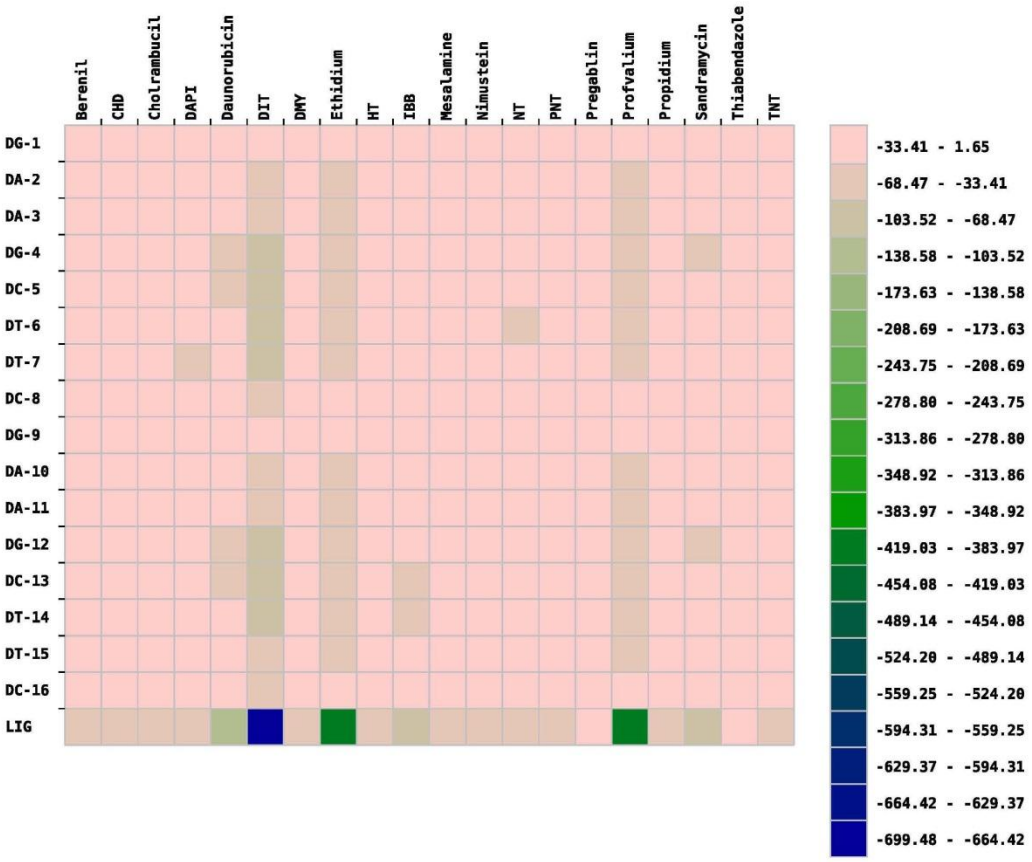


Fig. 6.8: Energetic contribution of DNA residues in the binding in kilojoules/mol for the DNA with intercalation gap.

The MMPBSA calculation shows that the electorstatic contribution of the energy for the ligand DIT, Ethidium and profavalin is much higher than that of the vdw contribution, so the interaction energy for these complexes is higher than other complexes. For all other remaining complexes the vdw energy contributes much to the final binding energy. Residue wise contribution to the final interaction energy was calculated. For the DNA with intercalation gap the DNA residues DG4, DC5, DT6, DT7, DA11, DG 12, DC13 and DT14 strongly interact with the ligand (Fig. 6.8), while in case of DNA without intercalation gap the DNA residues DG4, DC5, DT6, DT7, DG12 and DC13 involves the

ligand (Fig. 6.9). The interaction energy calculated from the MMPBSA method shows the similar trend with the experimental data.

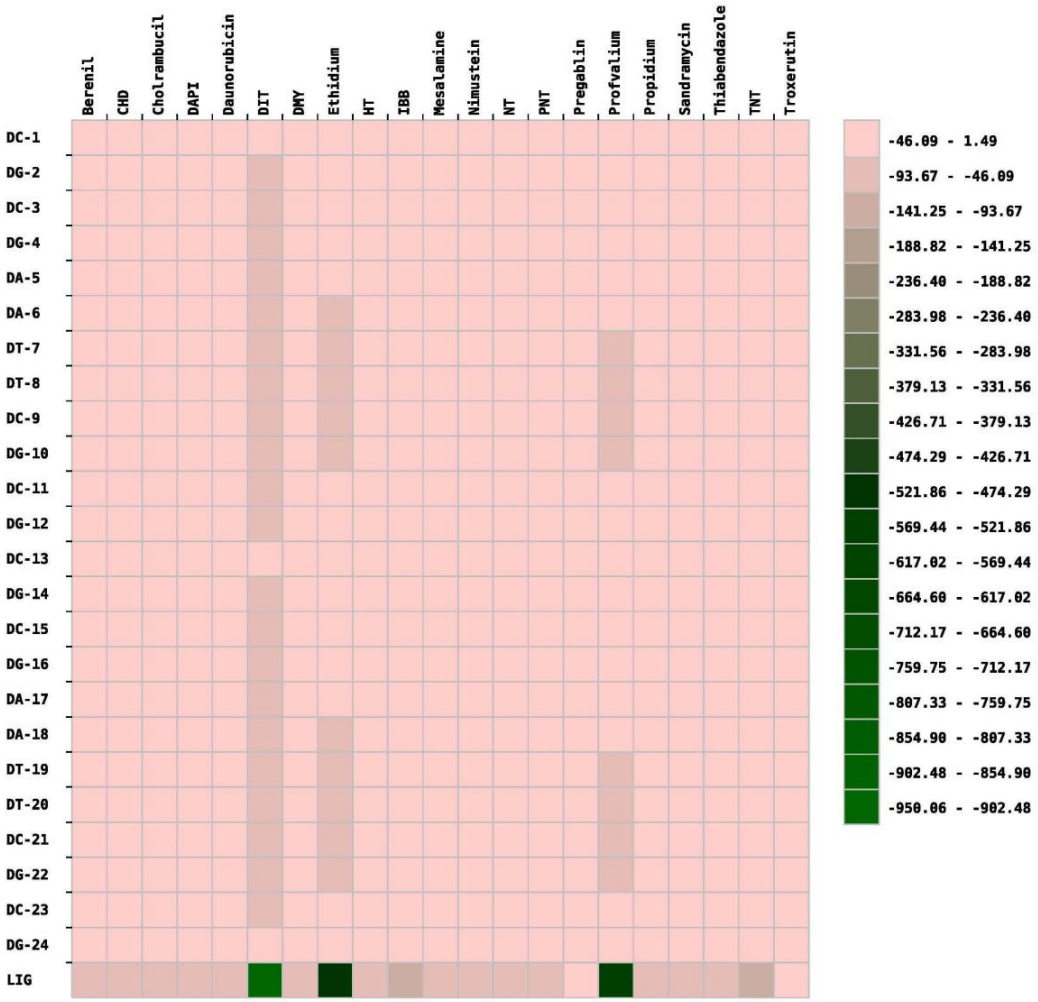


Fig. 6.9: Energetic contribution of DNA residues in the binding in kilojoules/mole for the DNA without intercalation gap.

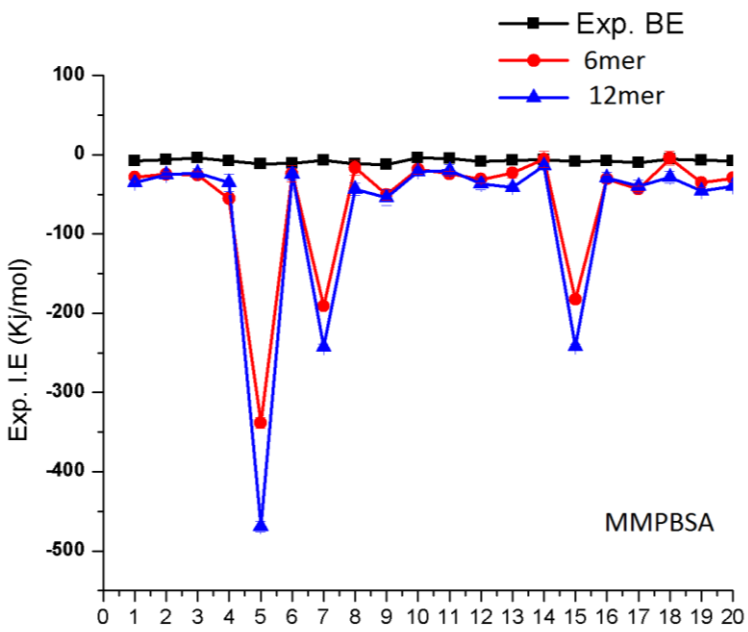


Fig. 6.10: Interaction Energy Trend (MMPBSA).

6.4 Conclusions

This study revealed that the interaction of any ligand with the DNA not only depends upon the chemical structure of ligand but also on the DNA sequence. An intercalator intercalates between the base pairs of DNA, if and only if, the DNA has a required intercalation gap between DNA base pairs, otherwise, the ligand will bind to DNA groove (major or minor). Molecular Docking from AUTODOCK shows that if the DNA has intercalation site then minor groove binders are also able to bind as intercalators. MD simulations were performed to check the stability of docked complexes and RMSD analysis shows that the minor groove binder that bind as an intercalator with DNA having intercalation gap are not stable up to the end of simulation. From the RMSD analysis, the average RMSD variation of minor groove binders that bind to the minor groove of DNA having intercalation site, is always greater than that of the DNA without intercalation site as minor groove binder. So, it is clear from above discussion that on the basis of

Molecular Docking only, the exact mode of binding of ligand to DNA cannot be predicted and MD simulation is required to check the stability of docked complexes.

References:

[1] Kitchen, D. B., Decornez, H., Furr, J. R., Bajorath, J., *Nat. ReV. Drug DiscoVery*, **3**, 935 (2004).

[2] Reddy, A.S., Pati, S.P., Kumar, P.P., Pradeep, H.N., Sastry, G.N., *Curr. Protein Pept. Sci.*, **8**, 329 (2007).

[3] Moitessier, N., Englebienne, P., Lee, D., Lawandi, J., Corbeil, C. R., *Br. J. Pharmacol.*, **153**, S7 (2008).

[4] Sternberg, M. J. E., Gabb, H. A., Jackson, R. M., *Curr. Opin. Struct. Biol.*, **8**, 250 (1998).

[5] P.R. Turner, W.A. Denny., *Mutat. Res.*, **355**, 141 (1996).

[6] Baraldi, P.G., Bovero, A., Fruttarolo, F., Preti, D., Tabrizi, M.A., Pavani, M.G., Romagnoli, R., *Med. Res. Rev.*, **24**, 475 (2004).

[7] Hannon, M. J., *Chem. Soc. Rev.*, **36**, 280 (2007).

[8] Tse, W. C., Boger, D. L., *Chem. Biol.*, **11**, 1607(2004).

[9] Waring, M. J., Bailly, C., *J. Mol. Recognit.*, **7**, 109 (1994).

[10] Brana, M.F., Cacho, M., Gradillas, A., de Pascual-Teresa, B., Ramos. A., *Curr. Pharm. Des.*, **7**, 1745 (2001).

[11] Wheate, N.J., Brodie, C.R., Collins, J.G., Kemp, S., Aldrich-Wright J.R., *Mini Rev. Med. Chem.*, **7**, 627 (2007).

[12] Martínez, R., Chacón-García. L., *Curr. Med. Chem.*, **12**, 127 (2005).

[13]Grootenhuis, P. D. J., Roe, D. C., Kollman, P. A., Kuntz, I. D., *J. Comput. Aided Mol. Des.*, **8**, 731 (1994).

[14] Sobhani, A. M., Amini, S. R., Tyndall, J. D. A., Azizi, E., Daneshtalab, M., Khalaj, A., *J. Mol. Graphics Model.,* **25**, 459 (2006).

[15] Evans, D. A.; Neidle, S., *J. Med. Chem.,* **39,** 4232 (2006).

[16] Navarrete, J. T., Casado, J., Ram, F. J., *J. Mol. Struct.*, **834**, 176 (2007).

[17] Yan, Z., Sikri, S., Beveridge, D. L., Baranger, A. M., *J. Med. Chem.,* **50**, 4096 (2007).

[18] Hurley, L. H., *Nat. ReV. Cancer,* **2**, 188 (2002).

[19] Holt, P. A., Chaires, J. B., Trent, J. O., *J. Chem. Inf. Model.,* **48**, 1602 (2008).

[20] Ricci, C.G., Netz, P.A., *J. Chem. Inf. Model.*, **49,** 1925 (2009).

[21] Mariya, al-Rashidaa, Ahsen, S., *RSC Advances,* **5,** 72394 (2015).

[22] Gaussian 09, Revision A.02, Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G.A., Nakatsuji, H., Caricato, M., Li X., Hratchian, H.P., Izmaylov, A.F., Bloino, J., Zheng, G., Sonnenberg, J.L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J.A., Jr., Peralta, J.E., Ogliaro, F., Bearpark, M., Heyd, J.J., Brothers, E., Kudin, K.N., Staroverov, V.N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J.C., Iyengar, S.S., Tomasi, J., Cossi, M., Rega, N., Millam, J.M., Klene, M., Knox, J.E., Cross, J.B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli, C., Ochterski, J.W., Martin, R.L., Morokuma, K., Zakrzewski, V.G., Voth, G.A., Salvador, P., Dannenberg, J.J., Dapprich, S., Daniels, A.D., Farkas, O., Foresman, J.B., Ortiz, J.V., Cioslowski, J., Fox D.J., Gaussian, Inc., Wallingford CT, 2009.

[23] Morris, G.M., Huey, R., Lindstorm, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J., *J. Compt. Chem.*, **30**, 2785(2009).

[24] Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., Spoel, D., Hes, B., Lindahl, E., *Bioinformatics*, **29**, 845 (2013).

[25] Kumari, R., Kumar, R., Lynn, Andrew, *J. Chem. Inf. Model.*, **54**, 1951(2014).

Chapter 7

General Conclusions

Chapter 7

General Conclusions

The work done in this thesis highlights following general conclusions:

A. The focus of this study is to confirm the importance of DNA sequence and specificity in directing the complex formation at molecular level. Thus, our study attempts to give detailed insight on the complexity in binding modes of small molecules to DNA.

B. In this study, the drug molecules which targeted DNA have been discussed. The array of available computational approaches and molecular modelling methods are being used for complementing the experimental efforts to improve the existing drugs and also in designing novel drug candidates which can act as good DNA inhibitors.

C. The computational studies between minor groove binders and intercalators with DNA show that DNA intercalators generally bind between the CG regions of nucleic acid via π-π interaction between ligand and nucleic acid bases, while minor groove binders form hydrogen bonds between ligands and the functional groups on the bases are exposed in the grooves via their end groups and also their amide or other linker groups in AT-rich region of DNA.

D. The interaction of any ligand to the DNA depends not only on the chemical structure of ligand but also on the DNA sequence. Intercalation requires conformational changes in DNA and a minor groove binder can also as an

intercalator if DNA has required intercalation gap and an intercalator bind as minor groove binder if DNA has no intercalation gap.

E. QM/MM calculations provide much better results than that of other molecular modelling methods such as molecular docking and molecular dynamics results. So, QM/MM study provides better insight on the complexity in binding modes of small molecules to DNA. The use of MD simulations approach with QM/MM calculations provides a theoretical protocol for complementing experimental techniques.

List of Publications

Published Papers:

1. Molecular Docking and Molecular Dynamics Study of DNA Minor Groove Binders

   Ruchi Mishra, Anamika Singh Gaur, Ramesh Chandra, Devesh Kumar

   International Journal of Pharmaceutical Chemistry and Analysis, 2015, 2, 161-169.

2. Assessing Therapeutic Potential of Molecules: Molecular Property Diagnostic Suite for Tuberculosis

   Anamika Singh Gaur, Anshu Bhardwaj, Arun Sharma, Lijo John, M Ram Vivek, Neha Tripathi, Prasad V Bharatam, Rakesh Kumar, Sridhara Janardhan, Abhaysinh Mori, Anirban Banerji, Andrew M Lynne, Anmol J Hemrom, Anurag Passi, Aparna Singh, Asheesh Kumar, Charuvaka Muvva, Chinmai Madhuri, Chinmayee Choudhury, D Arun Kumar, Deepak Pandit, Deepak R. Bharti, Devesh Kumar, Er Azhagiya Singam, Gajendra Ps Raghava, Hari Sailaja, Harish Jangra, Kaamini Raithatha, Karunakar Tanneeru, Kumardeep Chaudhary, M Karthikeyan, M Prasanthi, Nandan Kumar, N Yedukondalu, Neeraj K Rajput, P Sri Saranya, Pankaj Narang, Prasun Dutta, R Venkata Krishnan, Reetu Sharma, R Srinithi, **Ruchi Mishra**, S Hemasri, Sandeep Singh, Subramanian Venkatesan, Suresh Kumar, Uca Jaleel, Vijay Khedkar, Yogesh Joshi and G Narahari Sastry

   J. Chem. Sci. 2017, 129, 515-531.

3. A review on theoretical studies of various types of Drug-DNA Interaction

   Ruchi Mishra, Asheesh Kumar, Ramesh Chandra, Devesh Kumar

   International Journal of Science, Technology and Society, 2017, 3, 11-27

4. A review on QM/MM studies of nucleic bases interactions with graphene and carbon nanotubes

   Asheesh Kumar, **Ruchi Mishra**, Deep Kumar, Devesh Kumar

   International Journal of Science, Technology and Society, 2017, 3, 1-10

Communicated Papers:

5. DNA binding study of Dicationic Carbazoles and its Analogs

   Anwesh Pandey, **Ruchi Mishra**, Anil Kumar Yadav, Devesh Kumar

6. Illustrating Binding Mechanism of DNA intercalators using Computational Approaches

   Ruchi Mishra, Anamika Singh Gaur, Asheesh Kumar, Ramesh Chandra, Devesh Kumar, G.N. Sastry

7. The role of Quantum Mechanics/Molecular Mechanics calculations in understanding the binding of DNA minor groove binders

   Ruchi Mishra, Asheesh Kumar, Ramesh Chandra, Devesh Kumar

8. Cross Molecular Docking and Molecular Dynamics Studies Of DNA Binding Ligands

   Ruchi Mishra, Asheesh Kumar, Ramesh Chandra, Devesh Kumar

J. Chem. Sci. Vol. 129, No. 5, May 2017, pp. 515–531. © Indian Academy of Sciences.
DOI 10.1007/s12039-017-1268-4

REGULAR ARTICLE

Assessing therapeutic potential of molecules: molecular property diagnostic suite for tuberculosis (MPDS^{TB})

ANAMIKA SINGH GAUR[a], ANSHU BHARDWAJ[b], ARUN SHARMA[b], LIJO JOHN[a], M RAM VIVEK[a], NEHA TRIPATHI[c], PRASAD V BHARATAM[c], RAKESH KUMAR[b], SRIDHARA JANARDHAN[a], ABHAYSINH MORI[c], ANIRBAN BANERJI[a,†], ANDREW M LYNN[e], ANMOL J HEMROM[e], ANURAG PASSI[b], APARNA SINGH[a], ASHEESH KUMAR[g], CHARUVAKA MUVVA[d], CHINMAI MADHURI[f], CHINMAYEE CHOUDHURY[a], D ARUN KUMAR[a], DEEPAK PANDIT[f], DEEPAK R. BHARTI[e], DEVESH KUMAR[g], ER AZHAGIYA SINGAM[d], GAJENDRA PS RAGHAVA[b], HARI SAILAJA[h], HARISH JANGRA[c], KAAMINI RAITHATHA[h], KARUNAKAR TANNEERU[a], KUMARDEEP CHAUDHARY[b], M KARTHIKEYAN[f], M PRASANTHI[a], NANDAN KUMAR[a], N YEDUKONDALU[a], NEERAJ K RAJPUT[b], P SRI SARANYA[a], PANKAJ NARANG[e,†], PRASUN DUTTA[h], R VENKATA KRISHNAN[c], REETU SHARMA[a], R SRINITHI[a], RUCHI MISHRA[g], S HEMASRI[a], SANDEEP SINGH[b], SUBRAMANIAN VENKATESAN[d], SURESH KUMAR[g], UCA JALEEL[h], VIJAY KHEDKAR[f], YOGESH JOSHI[f] and G NARAHARI SASTRY[a,*]

[a]Centre for Molecular Modeling, CSIR-Indian Institute of Chemical Technology, Tarnaka, Hyderabad 500 007, India
[b]Bioinformatics Centre, CSIR-Institute of Microbial Technology, Chandigarh 160 036, India
[c]Department of Medicinal Chemistry, National Institute of Pharmaceutical Education and Research (NIPER), Mohali 160 062, India
[d]Chemical Laboratory, CSIR-Central Leather Research Institute, Chennai 600 020, India
[e]School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi 110 067, India
[f]Chemical Engineering and Process Development, CSIR-National Chemical Laboratory, Pune 411 008, India
[g]Department of Applied Physics, Babasaheb Bhimrao Ambedkar University, Lucknow 226 025, India
[h]Open Source Drug Discovery Consortium, New Delhi, India
E-mail: gnsastry@gmail.com

MS received 5 March 2017; revised 20 March 2017; accepted 22 March 2017

Abstract. Molecular Property Diagnostic Suite (MPDS^{TB}) is a web tool (http://mpds.osdd.net) designed to assist the in silico drug discovery attempts towards Mycobacterium tuberculosis (Mtb). MPDS^{TB} tool has nine modules which are classified into data library (1–3), data processing (4–5) and data analysis (6–9). Module 1 is a repository of literature and related information available on the Mtb. Module 2 deals with the protein target analysis of the chosen disease area. Module 3 is the compound library consisting of 110.31 million unique molecules generated from public domain databases and custom designed search tools. Module 4 contains tools for chemical file format conversions and 2D to 3D coordinate conversions. Module 5 helps in calculating the molecular descriptors. Module 6 specifically handles QSAR model development tools using descriptors generated in the Module 5. Module 7 integrates the AutoDock Vina algorithm for docking, while module 8 provides screening filters. Module 9 provides the necessary visualization tools for both small and large molecules. The workflow-based open source web portal, MPDS^{TB} 1.0.1 can be a potential enabler for scientists engaged in drug discovery in general and in anti-TB research in particular.

Keywords. Tuberculosis; chemoinformatics; open science; neglected diseases; drug discovery portal; web-based technology.

*For correspondence
†Deceased: ANIRBAN BANERJI and PANKAJ NARANG.

1. Introduction

Data and knowledge generated in drug discovery have been escalating exponentially in recent years owing to the demanding nature of pharmaceutical industry to deliver affordable and safer drugs for existing and emerging diseases.[1–10] How to make the knowledge thus generated available to the practicing scientists is an issue of great significance as it reduces the redundancy, enables research activity and focuses on the grand challenges in the healthcare sector.[11–17] Open science and open innovation are extremely important in the drug discovery approaches in general and those directed towards neglected and orphan diseases in particular.[18–27] How the existing knowledge helps medicinal chemists can be addressed by answering the following two questions: a) What is the value or relevance of molecules that were synthesized? and (b) which of those molecules are the most promising? Because currently chemist's ability to synthesize complex molecules has increased tremendously and more often than not, the question is which molecule to synthesize rather than how to synthesize.

Tuberculosis (TB) has become a global threat killing nearly 1.4 million people with 10.4 million new cases in 2015.[28] The disease-causing bacteria Mtb is a rather challenging microorganism that takes over six months of treatment with multiple drugs to curb its infection.[29] The therapeutic interventions further get confounded due to the fact that most of the time Mtb remains in latent phenotype, which is not well understood.[30] The emergence of multi-drug resistant, extensively drug-resistant and totally drug-resistant forms have resulted in long duration therapies with various side effects and toxicity issues with a risk of non-compliance.[31–38] Therefore, the need of new therapy to combat this dreadful disease is inevitable and demands exploration of new chemical space. Computational approaches to obtain and optimize anti-tubercular leads have been extensively employed in this area.[39–44] There are several public databases having diverse chemical classes of the compounds including PubChem, ZINC, KEGG, DrugBank, ASINEX, ChEMBL and NCI.[45–62] The computational tools and databases are required to facilitate rational prioritization and analysis of compounds from the available large chemical space. Further, the development of new algorithms/scripts is required to classify the chemical space for searching or extracting the useful information for each molecule. The tools are needed to predict the physical, chemical and biological properties of small molecules in order to improve search capabilities.

Before synthesizing any molecule, the medicinal chemist should have prior knowledge to optimize various physicochemical properties, structural alerts, in addition to the understanding of protein-ligand interactions that help in improving drug-likeness and avoiding toxicity issues. There are a number of open source scripts and algorithms developed by various developers for solving drug discovery issues.[63,64] Ideally, developing a disease-specific web portal which integrates the publicly available tools could provide a right platform to conduct drug discovery research in TB.

A variety of chemoinformatics analysis tools have been previously implemented in workflow systems. Steinbeck *et al.,* have implemented the chemoinformatics library of Chemistry Development Kit (CDK)[65,66] in the Taverna workflow suite.[67] Steinbeck *et al.*[68] have also implemented CDK in Konstanz information miner (KNIME), which is an open source workflow platform. It contains functions like format conversion, signatures, fingerprints and molecular properties generation. The Galaxy platform[69–71] is another workflow management system that provides easy to use interfaces of tools to the users and allows easy connection of the tools as well. Various instances of Galaxy have already been established.[72] For example, Ballaxy[73] is a Galaxy instance for structural bioinformatics wherein functions like protein preparation, ligand and protein checker, docking and many other tools have been implemented.

The current manuscript presents a web-based MPDS^{TB} Galaxy tool, which provides an open source platform for the chemoinformaticians, bioinformaticians, medicinal chemists, computational biologists, pharmacologists and others scientists to work on the design of anti-tuberculosis (anti-TB) drugs. The Galaxy based web tool is conveniently designed to integrate with any other software or script and can be used by designing user-defined workflows, a feature conveniently exploited by the users of Galaxy in many cases. The MPDS^{TB} tool provides three class of modules: a) Data Library (modules: 1. Literature, 2. Target library, 3. Compound library); b) Data Processing (4. File format conversion, 5. Descriptor calculation); and c) Data Analysis

Principal investigator: G NARAHARI SASTRY
Co-principal investigators: P ANSHU BHARDWAJ, PRASAD V BHARATAM, ANDREW M LYNN, DEVESH KUMAR, GAJENDRA P S RAGHAVA, M KARTHIKEYAN, SUBRAMANIAN VENKATESAN
Core developers: ANAMIKA SINGH GAUR, ANSHU BHARDWAJ, ARUN SHARMA, LIJO JOHN, M RAM VIVEK, NEHA TRIPATHI, PRASAD V BHARATAM, RAKESH KUMAR, SRIDHARA JANARDHAN, G NARAHARI SASTRY
Co-developers: ABHAYSINH MORI, ANIRBAN BANERJI, ANMOL J HEMROM, ANURAG PASSI, APARNA SINGH, ASHEESH KUMAR, CHARUVAKA MUVVA, CHINMAI MADHURI, CHINMAYEE CHOUDHURY, D ARUN KUMAR, DEEPAK PANDIT, DEEPAK R BHARTI, ER AZHAGIYA SINGAM, HARI SAILAJA, HARISH JANGRA, KAAMINI RAITHATHA, KARUNAKAR TANNEERU, KUMARDEEP CHAUDHARY, M PRASANTHI, NANDAN KUMAR, N YEDUKONDALU, NEERAJ K RAJPUT, P SRI SARANYA, PANKAJ NARANG, PRASUN DUTTA, R VENKATA KRISHNAN, REETU SHARMA, R SRINITHI, RUCHI MISHRA, S HEMASRI, SANDEEP SINGH, SURESH KUMAR, UCA JALEEL, VIJAY KHEDKAR, YOGESH JOSHI.

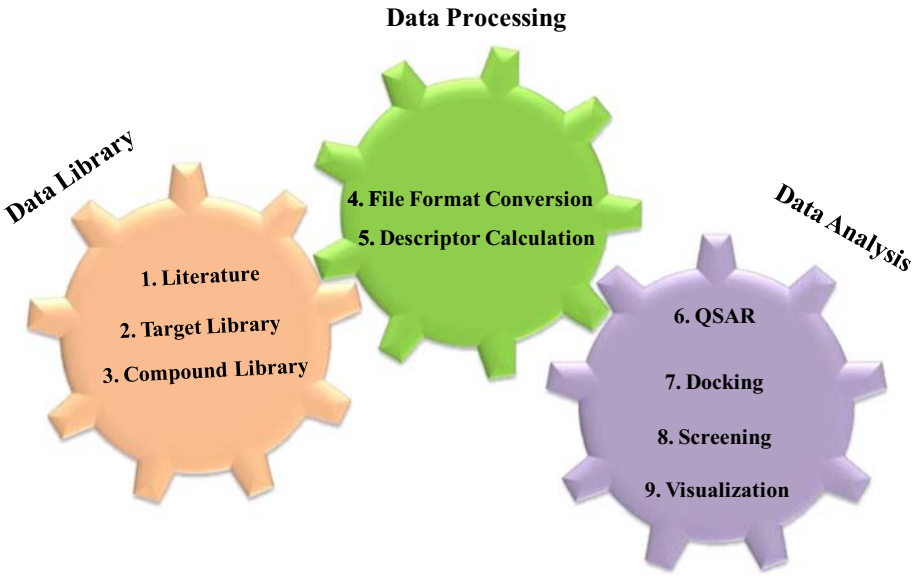


Figure 1. MPDS is structured into data library (literature, target library, compound library), data processing (file format conversion, descriptor calculation) and data analysis (QSAR, docking, screening and visualization).

(6. QSAR, 7. Docking, 8. Screening, 9. Visualization) (Figure 1). While modules 1, 2, 7 and 8 are specific to a particular disease (in this case TB), it is quite possible to make at least three out of the four modules, 2, 7, and 8 as generic. Efforts are underway to achieve this in the near future. However, in MPDS^{TB}1.0.1, only five modules Compound library, QSAR, Docking, Screening, Visualization are generic in nature and can be used 'as is' in drug discovery platforms directed towards other diseases. As the main focus of the current endeavor is to store all the data that is generated for each of the molecules, we decided to generate a unique MPDS ID, which is akin to Aadhar number in India, or social security number in the USA where all the information pertaining to a molecule is stored. A structure-based classification tool has been employed which facilitates the navigation through the chemical space.

The current work has the following major objectives: a) Quantitatively evaluating the multifarious aspects of drug-likeness of a given molecule, in order to diagnose its potential application as a drug; b) Calculate various drug-like physico-chemical properties for prioritization of compounds; c) Provide necessary framework to employ virtual screening of large dataset of compounds; d) Help synthetic medicinal chemists for the design of novel compounds; e) Coordinate the strategic development and integration of chemoinformatics efforts; f) Develop new multi-disciplinary collaborative projects. The current work focuses on the anti-tubercular lead discovery, design and optimization. Expectedly, a suitably altered protocol can be generated for other disease specific Galaxy web MPDS portals by customizing some of the modules. Thus, MPDS in the long run, can emerge as a general purpose open source drug discovery web portal.

2. Methods and modules

Galaxy (http://galaxyproject.org/) is a web-based workflow management system implemented in Python programming language, which is widely used for making data libraries, data integration, data processing and data analysis. In the current work, MPDS^{TB} is developed using Linux (CentOS 6.4) operating system having the python version 2.7. It provides a graphical user interface (GUI) to many computational tools that helps in computational chemistry, drug design, image analysis, climate modeling, linguistics, and biomedical research. Basically, it was developed to analyze the genomic data including gene expression, proteomics, transcriptomics, and gene assembly. Galaxy can be used directly on the web or can be installed in local machines which gives freedom to the users to integrate their own tools. It has flexibility in using diverse biological, chemical data formats and it allows the integration of tools that is written in any programming language or script for which a command line invocation can be constructed. Once the piece of code is written, a tool definition file should be written in XML that describes the working of the tool and its input/output parameters (Table 1).

For each module, the XML code and its complete path should be incorporated into the main configuration file of the Galaxy. The new tool implemented gets displayed in the

Table 1. Description of XML file as implemented in Galaxy MPDS[TB].

Tool ID	Description
<tool id>	Gives a unique name to the tool whose description is mentioned in the XML file.
<name>	It has the name of the tool that will be displayed as hyperlink in Galaxy.
<description>	This is displayed just after the hyperlinked name.
<command>	It describes how the tool (which compiler) will be executed and its input and output parameters.
<inputs>	Defines the input parameters.
<outputs>	Defines the output parameters.
<help>	Describes what the tool does.

tool panel of the Galaxy home page. The Galaxy workspace mainly consists of four areas, the first one is the navigation bar which provide links to Galaxy's major components, analysis workspace, workflows, data libraries, and user repositories (histories, shared data, workflows, pages), the second one is the tool panelist containing the analysis tools and data sources available to the user, the third one is detailed panel display interfaces for tools selected by the user and the fourth one is history panel that shows data and the results of analyses performed by the user, as well as automatically tracked metadata and user-generated annotations.

3. Structure of MPDS[TB]

MPDS[TB] is structured into data library (literature, target library, compound library), data processing (file format conversion, descriptor calculation), and data analysis (QSAR, docking, screening, and visualization) (Table 2). Each of these modules is customized for TB drug discovery and will be described in the following sections.

3.1 *Data library*

3.1a *Module 1: Literature*: Module 1 provides information of druggable protein targets/gene information, FDA-approved drugs, and polypharmacology for Mtb. The genetic information provided includes RvID, gene name, gene product, class of protein, structural details from PDB, active site, function, metabolic pathway, localization, method of validation, drug/inhibitor information, druggability index, and mechanism of action. The literature module provides the list of available FDA approved drugs along with their identification, pharmacology, potential targets and corresponding references. The polypharmacology covers the structure of Mtb cell wall, biosynthetic, metabolic pathways (cell wall, chorismate, amino acids, lipids, carbohydrate, cofactor, nitrogen/sulphur, DNA, protein), bibliography and hyperlinks were given to various servers related to TB.

Table 2. Description of various modules in MPDS[TB] 1.0.1.

Category	Modules	Description
Data Library	Module 1: Literature	Contains Mtb proteins and its genetic information; FDA approved drug information and polypharmacological information.
	Module 2: Target Library	Contains crystal structures and homology models for Mtb proteins.
	Module 3: Compound Library	Contains a single window interface for searching a compound in MPDS compound database.
Data Processing	Module 4: File Format Conversion	Conversions of files from one chemical format to another chemical format, 2D to 3D file conversion using Open Babel.
	Module 5: Descriptor Calculation	Calculation of descriptors and fingerprints using PaDEL and CDK tools.
Data Analysis	Module 6: QSAR	Generation of QSAR models using the data mining tools, McQSAR and SVMlight.
	Module 7: Docking	Ligand Optimization; Conformer Generation and Protein-Ligand docking.
	Module 8: Screening	Prioritization of compounds for drug-like features using DruLiTo tool; Biopharmaceutical Classification System (BCS); Identification of toxicophoric groups in a compound.
	Module 9: Visualization	Visualizing protein-ligand interactions using Jmol and Ligplot.

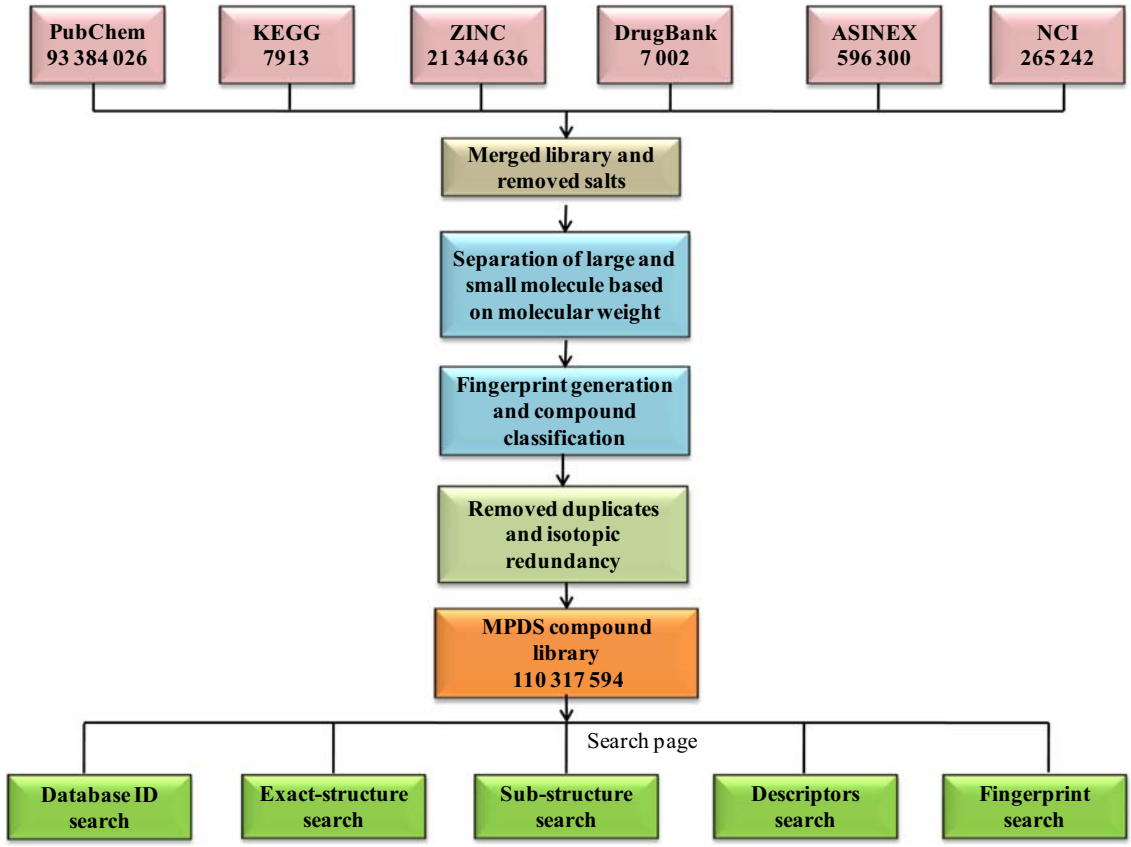


Figure 2. Schema for the generation of MPDS compound library from various data sources and representation of compound search engine.

3.1b *Module 2: Target library*: The target library consists of crystal structures of 140 Mtb proteins which are reported targets and also the ones prioritized through systems level analysis of Mtb interactome. Each protein is also annotated for key residues around the active site. Most of these target structures were collected from protein data bank (PDB) and some of them were homology models. If the structural information is not available, then mutagenesis study results were collected from literature. Multiple sequence alignment of the same family of protein was employed for identification of active site residues for the protein. One of the principal objectives of the target library is to provide a list of prepared proteins in Mtb suitable for molecular docking, and give hyperlink of the data source, if available. The collected PDB structures were prepared by using standard protocols such as assigning bond orders, adding hydrogens, and minimizing protein complexes.

The targets were selected from Mtb H37Rv genome family and they majorly belong to various enzyme classes, such as oxidoreductase, transferase, hydrolase, lyase, isomerase, and ligase. These targets are involved in various biological functions that include signal transduction, peptidoglycan and cell wall synthesis, amino acid synthesis, drug metabolism, DNA precursor synthesis, post-translational modifications and nitrogen metabolism, *etc.* The protein structures in MPDS^{TB} target library can be potentially exploited in structure-based drug design approaches.

3.1c *Module 3: Compound library*: The compound library is generated with the objective of establishing a single window interface to search compounds available across different public domain databases. An efficient small molecule search, implemented using multiple search strategies, facilitates the comprehensive analysis of the available chemical space that may be utilized for identification of novel anti-TB compounds.

Preparation: For the preparation of MPDS^{TB} compound library, existing small molecule databases such as PubChem, Drugbank, KEGG, NCI, ZINC and ASINEX were downloaded. To ensure that globally acceptable standards are followed to store and search this data, each compound in the library is converted into SMILES, InchI and InchIKey using OpenBabel 2.3.2.[74] Indexing of the compound library is done using InchIKey (Figure 2).

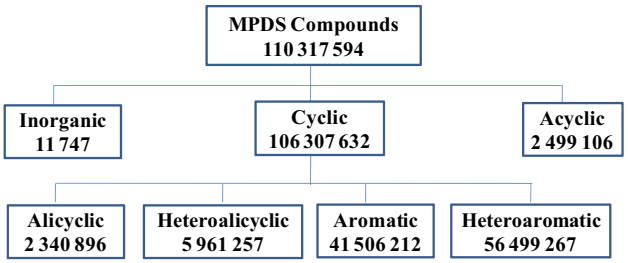


Figure 3. A structure-based chemical classification system of MPDS compound library.

Storage and search: As mentioned above, each compound in the library is stored using InchI standards along with SMILES string and the IDs from where the compound is sourced. Multiple search options are provided through MPDS^{TB} web interface which includes structure-based search, property-based search, and fingerprint-based search (Figure 3). Compound property data is mostly calculated using PaDEL. The pKa values and IUPAC names are calculated using ChemAxon command line tools. The substructure search is implemented using RDkit.[75] A novel fingerprinting algorithm is developed to classify compounds at various levels of structural features. This binary fingerprinting algorithm is developed and employed for the classification and clustering of all the molecules included in the MPDS^{TB} compound library database. Currently, a 30-bit qualitative and partially quantitative fingerprinting scheme is adopted to cluster similar compounds together and reduce the search space. Furthermore, two open source tools, 'molecule cloud' and 'open molecule generator' are also implemented in compound library module to create a molecular cloud of scaffolds and to generate molecule library respectively.[76,77]

3.2 *Data processing*

3.2a *Module 4: File format conversion*: A core requirement of any workflow management system is the connecting links between different analysis tools. Most often these links are based on the file formats that are read by these tools. As there are various molecular file formats to represent chemical structures, an open source file format converter is implemented to facilitate the creation of workflows over the web. Different tools require specific input file formats and will produce output in another specific format. In order to maintain a smooth flow of data between different tools in a workflow, the file format converter module utilizes one of the tools from the Galaxy toolshed to convert one file format to another based on user requirements. As of now, a few input formats like mol2, mol, sdf, SMILES, *etc.,* and output formats, mol2, sdf, mol, pdb are incorporated. This module also contains a tool that generates 3D coordinate from 2D structural file or SMILES which utilizes OpenBabel 2.3.2. The 3D structure generated follows geometrical rules based on hybridization of atoms. Once the structure is generated, the stereochemistry of the structure is taken care of and the lowest free energy conformer is generated by MMFF94 force field using weighted rotor search.

3.2b *Module 5: Descriptor calculation*: In order to computationally assess the properties of the compounds present in MPDS^{TB} compound library or those provided by end users, two descriptor calculation tools, namely, PaDEL and CDK, are incorporated in MPDS^{TB}. These tools may be used to calculate different compound properties. The input format for both of these descriptor tools is sdf and provides the output in CSV format. The descriptor module may read the output from compound library search or user uploaded sdf. The output may be used to build machine-learning models for target specific filters, predicting anti-TB properties, drug-like properties or toxicity of the compounds.

3.3 *Data analysis*

3.3a *Module 6: QSAR*: Two methods of data mining, quantitative structure activity relationships (QSAR) and support vector machine (SVM), are incorporated in MPDS^{TB}. A Multi-conformational Quantitative Structure-Activity Relationship (McQSAR) using Genetic algorithms is implemented for developing QSAR models.[78] The 'Build_QSAR_Model' tool of the QSAR module takes the descriptor file of the compounds with known activity and prompts the user to enter the name of the column whose value needs to be predicted (activity, in this case). In order to remove the redundancy and unwarranted features, the user has been given six options. The feature selection options are as follows: exclude correlated descriptors, exclude identical conformers, exclude inactive compounds, exclude sparse conformers, exclude sparse descriptors and exclude descriptors with zeros. The user also has the flexibility to set the number of times user wants cross-validation step to be repeated. The user can perform four kinds of cross-validation using the percentage of bins 3, 5, 7, 10 folds to divide the compounds. McQSAR can be used to predict the activity of new compounds using the model created by the previously mentioned tool. This accepts two input files: one that contains the descriptors of the compounds whose activity needs to be predicted and the second is a model file created by the 'Build_QSAR_Model' tool (Figure 4).

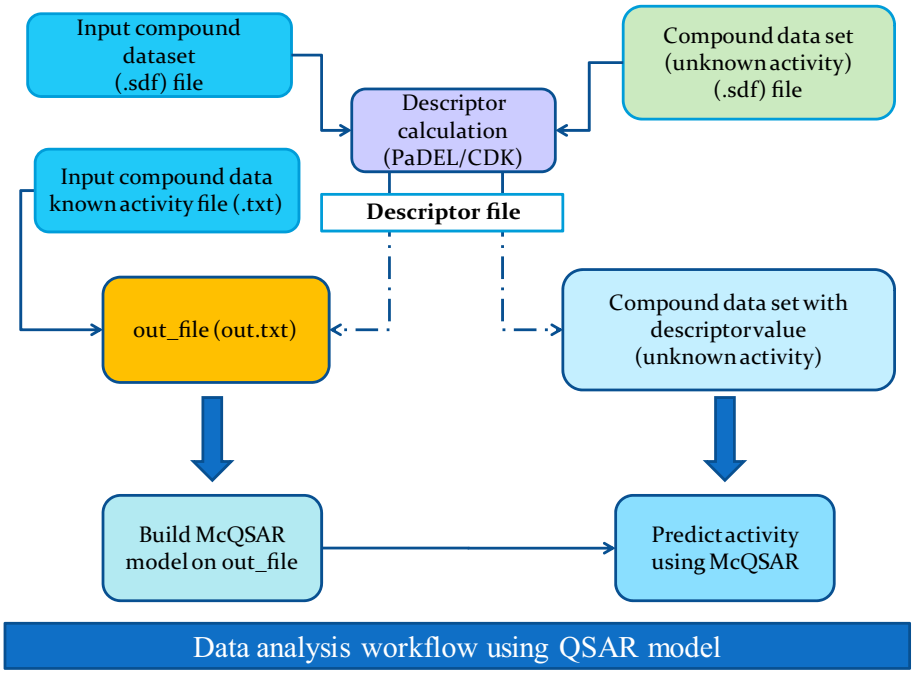


Figure 4. Workflow for QSAR model generation.

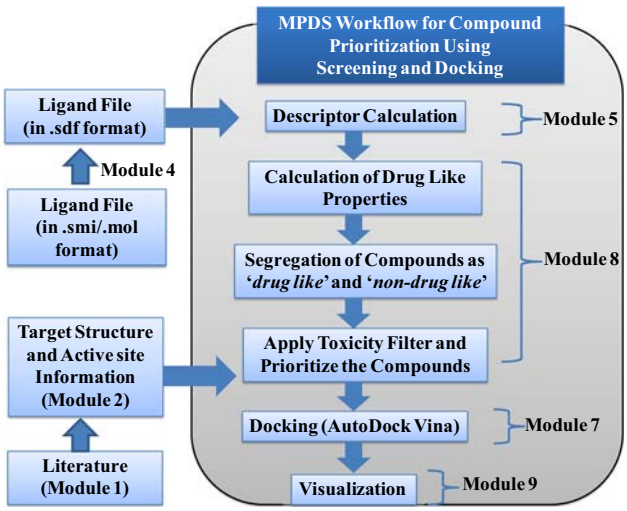


Figure 5. Workflow for compound prioritization using screening and docking modules.

Along with McQSAR, SVMlight[79] has also been incorporated in MPDS[TB] which helps in classifying the compound data into actives and inactives. The 'Build QSAR Model: SVMlight' builds a QSAR model based on descriptor files of active and inactive molecules. Then 'Classify Data: SVMlight' tool takes the model file and the descriptor file of compounds with unknown activity, and classifies them into actives and inactives.

3.3b *Module 7: Docking*: The docking protocol involves multiple steps including energy minimization (chemical structure optimization), conformer generation, docking and its analysis, and visualization. In MPDS[TB], the docking module contains four tools to carry out these steps. The ligand optimization tool incorporates the Phenix electronic Ligand Builder and Optimisation Workbench (eLBOW).[80] The tool uses semi-empirical quantum chemistry based calculations using AM1 (Austin Model 1).[81] Hydrogen atoms are automatically added to the compound by eLBOW. The conformer generation tool utilizes the OpenBabel 2.3.2. genetic algorithm approach to generate diverse conformers based on RMSD. AutoDock Vina[82] has been implemented as the docking software in module 7. Thus, the use of docking module allows the user to carry out efficient and robust docking calculations (Figure 5).

The proteins can be uploaded as a PDB file, or can directly be accessed from target library available in the MPDS[TB] shared library. The ligands can be obtained from the compound library of MPDS[TB] package or a user can upload their choice of ligand in sdf or pdb format with 3D coordinates. The docking calculation can be started using the default parameters and with some user defined parameters. Users can also download the docking results as a zip file containing the complex files in PDB format and the Vina log file containing the ranked binding free energy scores. The docked complexes can be visualized through the visualization module for examining the protein-ligand interactions.

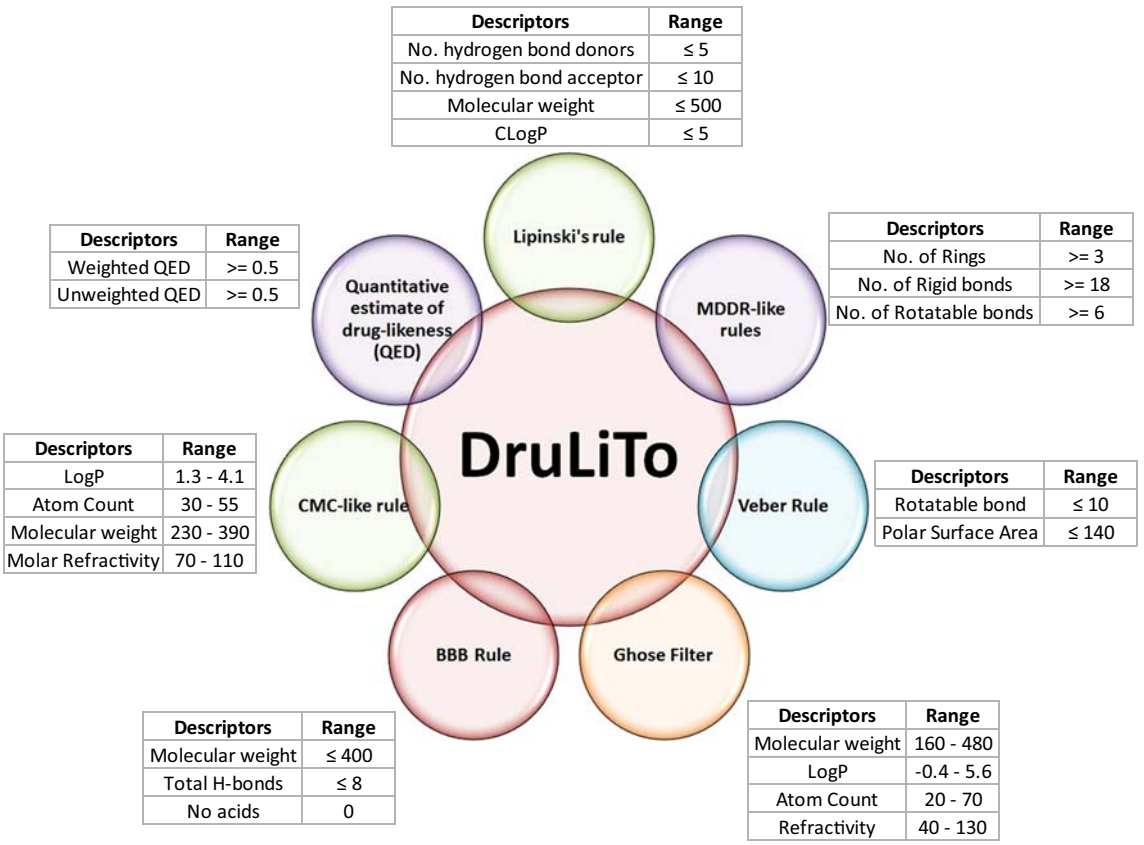


Figure 6. Drug-Likeness prediction tool (DruLiTo) for screening chemical compounds, databases or libraries.

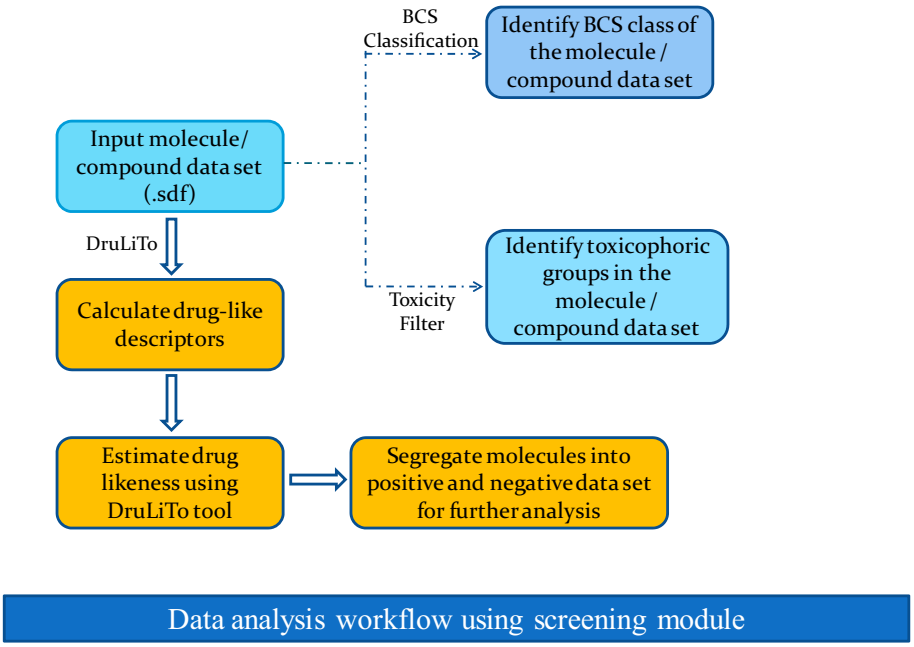


Figure 7. Workflow for compound screening using DruLiTo and filters.

3.3c *Module 8: Screening*: To prioritize compounds for their drug-like features, DruLiTo[83] is integrated into MPDS[TB]. This tool provides the assessment of drug-like features based on eight different published algorithms.[84–88] To calculate drug-likeness of a compound, DruLiTo needs descriptor data, which may be calculated using CDK. DruLiTo uses these descriptors to determine the drug-likeness of the molecule,

based on the different rules as defined by the end user. The user can then segregate the input dataset into passed and failed groups based on one or more rules selected (Figure 6, 7). In addition to this, the biopharmaceutics classification system (BCS) is implemented using in-house scripts to evaluate the solubility and the permeability of compounds. These predictions can subsequently be used to process the positive ligand set or negative ligand set or the original input dataset, all of which are in sdf format, to determine the BCS class (I, II, III, IV) into which the molecules fall based on their predicted permeability and intrinsic solubility values.

Another feature provided in MPDS[TB] is to perform toxicity analysis. A comprehensive list of structural alerts including data from FAF-Drugs2 (204 substructures), OChem and literature are generated. The toxicity filter performs a substructure search on the query molecule and highlights the matched toxicophoric groups from this data and displays it in an image format. It also produces the mol format of the molecules with toxicophoric groups that can be saved. Along with this, a text file listing the molecule ID, the total number of toxicophoric matches found and the SMARTS pattern of the matched groups can also be generated with this module.

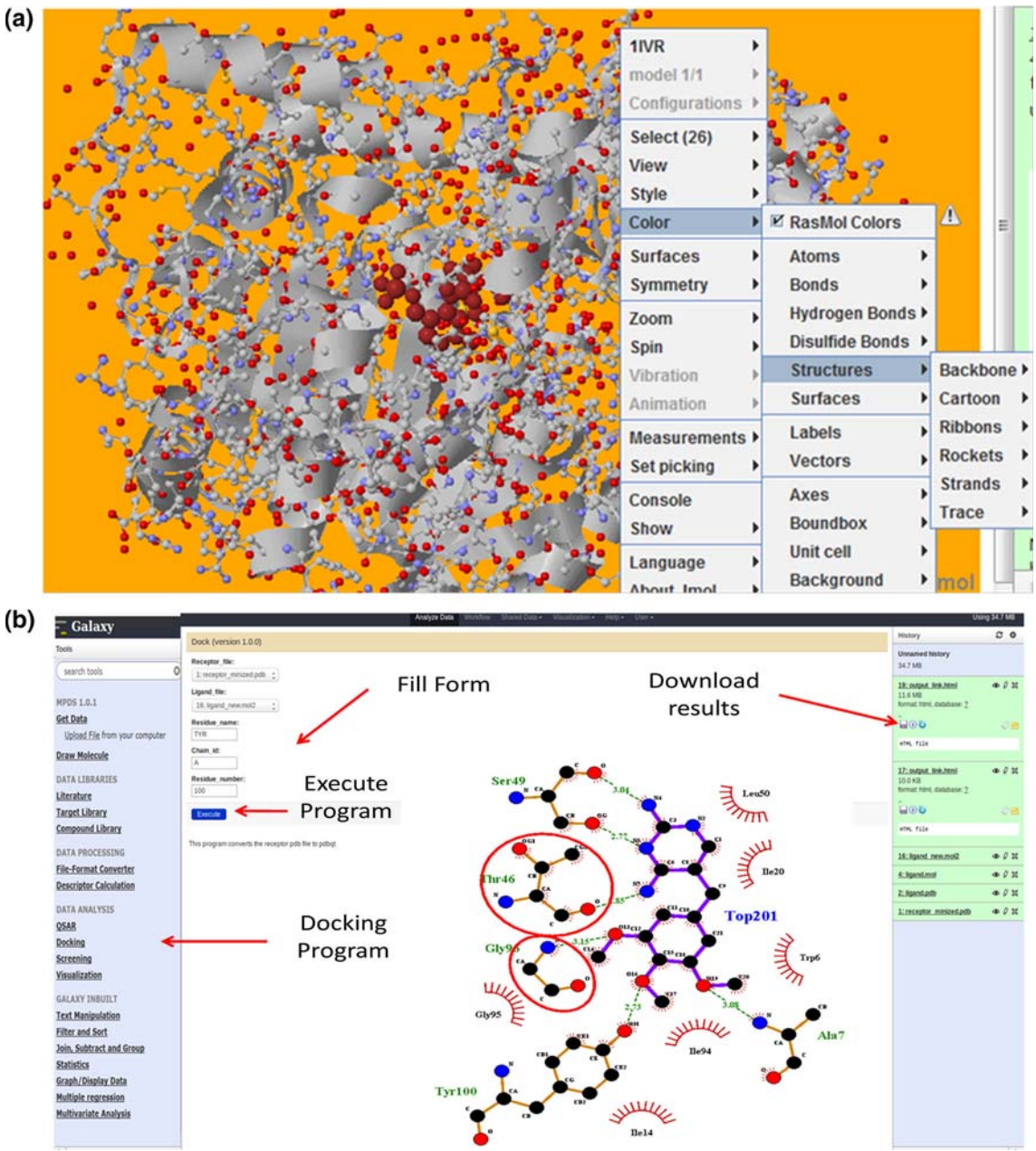


Figure 8. (**a**) Visualization of protein-ligand complex by various structural representations, (**b**) LigPlot for visualizing protein-ligand interactions (hydrogen bonding and hydrophobic contacts).

3.3d *Module 9: Visualization*: Protein-ligand complex can be viewed in LigPlot^{+} tool available in this module in order to examine the protein-ligand interactions. The key interactions of the ligand with active site amino acids are displayed by hydrogen bonding (dashed lines between the atoms) and hydrophobic contacts (an arc with spokes radiating towards the ligand atoms). The protein-ligand complex can be visualized in Jmol by various structural representations (Figure 8).

4. Results and Discussion

4.1 *The compound library and search page*

MPDS^{TB} compound library is a compiled and curated database containing 110.31 million unique compounds (as on 31st December 2016) from various publicly available databases including PubChem, KEGG, ZINC, DrugBank, ASINEX and NCI. Before making unique compounds, salt containing compounds were removed and further, the compounds were separated on the basis of molecular weight taking 750 Da as cut-off. Fingerprints were generated and compounds were classified into classes and clusters. Chemically well-defined classes obtained from fingerprinting algorithm were subjected to redundancy as well as removal of isotopes based on the InchIKey and truncated InChI respectively (Table 3). A class is a chemically well-defined group, whereas cluster is a set of compounds with a size limit of 0.25 million. After classification, the compound library majorly consists of 106.30 million cyclic compounds (2.34 million alicyclic, 5.96 million heteroalicyclic, 41.51 million aromatic and 56.49 million heteroaromatic). For aromatic compounds, class 12 (aromatic compounds contain two rings) has a maximum number of compounds (8.97 million) having 42 clusters, whereas class 20 (aromatic compounds contains more than or equal to four rings) has 0.91 million compounds in clusters. Highest number of compounds is present in the heteroaromatic class 22 (10 million), which are stored in 46 clusters (Figure 9). MPDS^{TB} fingerprinting algorithm and search engine offer a number of advantages over the available fingerprinting algorithms. The various available algorithms match a set of SMARTS patterns against each molecule to calculate each bit of the fingerprint that leads to reduced search time.[89] MPDS-Database uses SMILES notation of the molecules for storage and fingerprinting which makes it faster than the other approaches utilizing SMARTS notation for pattern search and matching. The available fingerprinting algorithms are not suitable for the large data size (110.31 million) as the speed is compromised due to matching a large set of SMARTS patterns against each molecule. The limitation of the current version of MPDS-fingerprinting is that it is not at the stage of being molecule specific. The purpose of designing this fingerprinting algorithm was to reduce the search space to a level where the 2D structural comparison can be performed without compromising the computational power and time utilization. Therefore, more efforts are required to make this fingerprinting robust and molecule specific. Fingerprinting of inorganic molecules is an issue with most of the available fingerprinting algorithms. However, MPDS^{TB} fingerprinting algorithm adopts the structure-based fingerprint classifying the inorganic compounds efficiently. The program for the generation of 30-bit fingerprinting is written in Java (using NetBeans IDE 7.2.1).[90] Further,

Table 3. Distribution of number of compounds in structural classes of MPDS^{TB} compound library.

Chemical Classes	Class‡	No. of clusters$^{£}$	No. of compounds
Acyclic	1	8	1,790,302
	2	4	708,804
Alicyclic	3	7	1,436,757
	4	3	483,877
	5	1	127,301
	6	2	292,961
Heteroalicyclic	7	12	2562,119
	8	11	2,289,075
	9	4	772,643
	10	2	337,420
Aromatic	11	31	6,842,221
	12	42	8,978,649
	13	32	6,898,540
	14	13	2,649,637
	15	31	6,276,919
	16	10	2,071,127
	17	5	1,024,251
	18	15	3,089,744
	19	14	2,771,697
	20	5	903,427
Heteroaromatic	21	16	3,461,762
	22	46	10,086,573
	23	25	5,484,547
	24	32	6,580,855
	25	45	9,268,612
	26	13	2,588,133
	27	12	2,228,509
	28	39	7,829,460
	29	45	8,970,816
Large Molecules	30	7	1,499,109
Inorganic	31	1	11747

‡Class is a chemically well-defined group, $^{£}$Cluster is a set of compounds with a size limit of 0.25 million.

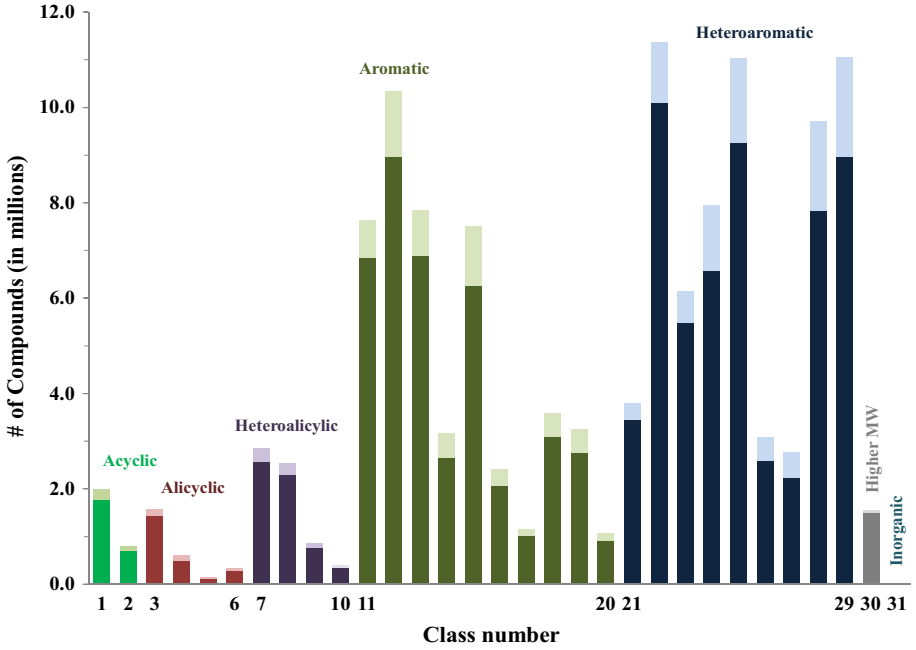


Figure 9. Distribution of unique (dark color) and duplicate (light color) compounds in different chemically well-defined classes.

the choice of SMILES notation for the assignment of fingerprinting and structure search was attributed to the easy and fast handling of the SMILES patterns rather than SMARTS patterns (Table 4).

4.2 *The search page*

MPDS^{TB} provides an interface for searching molecule based on their diverse properties through compound library search tool designed on the Galaxy platform. The user can search the compound library by database ID, exact structure, substructure, descriptor and fingerprint. The result of search page shows a list of molecules fulfilling the criteria, which are further linked to MPDS ID card, displaying the basic information of molecule. The in-house program will calculate molecular fingerprints and searches the query molecules in the database for generating MPDS ID card. For assisting the molecular drawing in MPDS-Galaxy portal, JS draw has been incorporated.

4.3 *Workflow system*

This is one of the most important features of the current platform, which paves the way to generate custom design workflows. The Galaxy workflow system can be used by a graphical drag-and-drop interface in which tools are configured and interconnected between the appropriate input and output points. The user can also extract/download the workflow from the history pane. MPDS^{TB} presents a platform where a data library, data processing and data analysis are configured and interconnected to assist *in silico* drug discovery. The user can utilize the inbuilt workflows or design a new one to carry out chemoinformatics analysis for any given molecule. The MPDS^{TB} workflow system can be interconnected between different modules for a specific task.

4.4 *MPDS ID card*

The output from MPDS^{TB} portal is MPDS ID card (which is akin to Aadhar card in India, or social security number in the USA) that represents a molecular profile report specific to a given molecule. This card reports the vital physico-chemical properties of a molecule essential to estimate the drug-likeness and activity of the molecule in the early stages of the drug discovery and development pipeline. MPDS ID which provides all pertinent information can have several pages. First page is standard and essentially aids in registering the molecule in the database, besides providing vital physico-chemical parameters. However, if more pertinent data are available on the molecule, such as their biological activity, spectroscopic data, toxicity, PK/PD data, *etc*., subsequent pages will be created. The first page of the molecule can be viewed publicly from the portal; second and subsequent pages are stored in the external hard disks at the development site, due to apparent disk storage limitations (Figures 10 and 11).

Table 4. 30-bit fingerprinting scheme and its various levels employed for the classification of molecules in MPDS^{TB} 1.0.1.

Fingerprint Level	Fingerprint bits	Fingerprint bits and Molecular features
Level 1 (Skeleton)	1 2 3 4 5 6	1st bit: Acyclic/Cyclic **Acyclic:** 2nd and 3rd bits: Saturation & Conjugation 4th bit: Straight/Branched 5th bit: Homo/Hetero atomic **Cyclic:** 2nd and 3rd bits: Number of rings 4th and 5th bits: Alicyclic/Aromatic 6th bit: Large/Small
Level 2 (Atom type)	7 8 9 10 11 12	**Acyclic and Cyclic:** 7th bit: Geometrical isomerism 8th bit: Hydrogen Bond Donor 9th bit: Hydrogen Bond Acceptor 10th bit: Halogens 11th bit: Heteroatom in side chain/backbone **Acyclic:** 12th bit: Presence of metal ion **Cyclic:** 12th bit: Presence of fused/unfused rings
Level 3 (Functional group)	13 14 15 16 17 18 19 20	**Acyclic and Cyclic:** 13th bit: Carbonyl group 14th bit: Phosphate containing group 15th bit: Cyanide group 16th bit: Nitro group 17th bit: Sulphur group 18th bit: Amino group 19th bit: Alcohol/Ether group 20th bit: Boron
Level 4 (Size of molecule)	21 22 23 24 25 26	**Acyclic and Cyclic:** 21st, 22nd, 23rd and 24th bits: Number of heavy atoms 25th bit: Chirality 26th bit: Connected/Disconnected structure
Level 5 (Heteroatom type in ring)	27 28 29 30	**Acyclic:** 27th, 28th, 29th and 30th bits: Even/Odd number of carbons **Cyclic:** 27th bit: Nitrogen in ring 28th bit: Oxygen in ring 29th bit: Sulphur in ring 30th bit: Phosphorus in ring
Level 6	–	Number of heteroatom containing rings

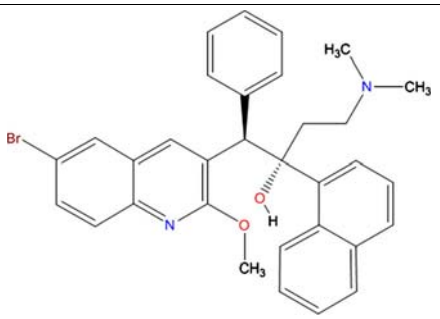
5. Conclusions

MPDS^{TB} 1.0.1 is a comprehensive open source Galaxy-based web tool, which provides a platform to integrate the data collection, processing, and analysis customized towards anti-TB drug discovery and design. One of the main features of this program is its ability to assess and estimate the activity of a given molecule using chemoinformatics, bioinformatics tools, and the existing knowledge in the area. The major attainment of MPDS^{TB} is the creation of a unique compound library (110.31 million) by removing duplicates, isotopic redundancy and salts. A structure-based classification program was developed, which classifies the compound library into 31 classes and 533 clusters.

We believe that the web-based chemoinformatics and modeling tools are great enablers to tackle grand challenges in healthcare in general and drug discovery in particular. One of the major bottlenecks for carrying out research in drug discovery in the academia can be

Molecular Property Diagnostic Suite

MPDS ID:27-09-137939



Molecular Formula:

$C_{32}H_{31}BrN_2O_2$

IUPAC Name:

(1R,2S)-1-(6-Bromo-2-methoxy-3-quinolinyl)-4-(dimethylamino)-2-(1-naphthyl)-1-phenyl-2-butanol

Remarks:

Name/Synonyms: Bedaquilina, Bedaquilinum, TMC207, Sirturo, TMC-207, R207910, TMC207, R207910

Molecular Properties:

Mol. Wt	555.5	LogP	6.37
HBD	1	LogS	-6.5
HBA	4	pKa	pKa1: 11.64; pKa2: 13.47; pKa3: 7.67; pKa4: 2.67
Molar refractivity	154.02	Polar surface area	45.59
Heavy atom count	37	Aromatic rings count	5
Rotatable bonds	8	Polarizability	57.29

Figure 10. Molecular Property Diagnostic Suite ID card (MPDS ID Card).

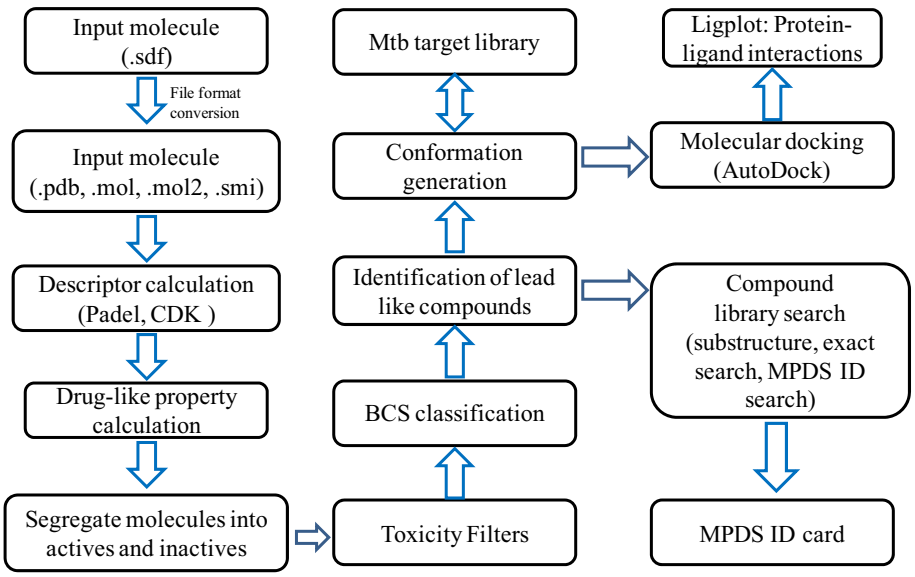


Figure 11. A workflow system for the generation of MPDS ID card by integration of data library, data processing and data analysis modules.

traced to the lack of proper software, access to the comprehensive information on the disease, and also what the contemporary challenges are in developing therapeutics in a given area. Developed countries are not significantly affected by TB and thus the onus of coordinating and conducting frontline research on neglected diseases will be on the developing world. It is very important that India, possessing a tremendous pool of talent and human resource, takes up the leadership role in undertaking research on TB and other neglected diseases. The

current web-based tool, MPDS^{TB} is expected to provide exactly such a platform to drive the research in this area.

Abbreviations

AM1: Austin Model 1; **BCS**: Biopharmaceutics Classification System; **CDK**: Chemistry Development Kit; **CSV**: Comma-Separated Values; **DruLiTo**: Drug Likeness Tool; **eLBOW**: electronic Ligand Builder and Optimisation Workbench; **FAF-Drugs3**: Free ADME-Tox Filtering Tool; **FDA**: Food and Drug Administration; **GUI**: Graphical User Interface; **ID**: Identification number; **InChI**: International Chemical Identifier; **InchIKey**: International Chemical Identifier key; **IUPAC**: International Union of Pure and Applied Chemistry; **KEGG**: Kyoto Enclyclopedia of Genes and Genomes; **KNIME**: Konstanz Information Miner; **McQSAR**: Multi-conformational Quantitative Structure-Activity Relationship; **MMFF94**: Merck Molecular Force Field 94; **MPDS**: Molecular Property Diagnostic Suite; **Mtb**: Mycobacterium tuberculosis; **NCI** National Cancer Institute; **PaDEL**: Pharmaceutical Data Exploration Laboratory; **PDB**: Protein Data Bank; **PK/PD**: PharmacoKinetic/Pharmaco Dynamic; **QSAR**: Quantitative Structure Activity Relationship; **RMSD**: Root Mean Square Deviation; **SDF**: Structure Data File; **SMARTS**: SMiles ARbitrary Target Specification; **SMILES**: Simplified Molecular-Input Line-Entry System; **SVM**: Support Vector Machine; **TB**: Tuberculosis; **XML**: Extensible Markup Language

Acknowledgements

We are thankful to OSDD, CSIR and Sir Dorabji TATA trust for providing TCOF fellowships to some of the authors in the study. CSIR 12th five year program GENESIS (BSC 0121), Department of Science and Technology (New Delhi) and Department of Biotechnology (New Delhi) are also thanked for funding. Code development has taken about 5 years of time starting from 2012 and has witnessed 5 Workshops in IICT, IMTECH, OSDD centre, Bangalore, and NCL. Besides there were several exchange of students between various institutes. We thank CSIR OSDD consortium, NIPER, JNU, and BBAU for providing support. GNS thank J C Bose fellowship of DST. This manuscript is dedicated to the memory of Dr. Anirban Banerji and Dr. Pankaj Narang who have provided a lot of energy and enthusiasm during the kick-start stages of the MPDS teamwork.

References

1. Searls D B 2005 Data integration: Challenges for drug discovery *Nat. Rev. Drug Discovery* **4** 45
2. Nwaka S, Ramirez B, Brun R, Maes L, Douglas F and Ridley R 2009 Advancing drug innovation for neglected diseases-criteria for lead progression *PLoS Negl. Trop. Dis.* **3** e440
3. Sachs J D 2001 A new global commitment to disease control in Africa *Nat. Med.* **7** 521
4. Jagarlapudi S A and Kishan K V 2009 Database systems for knowledge-based discovery *Methods Mol. Biol.* **575** 159
5. Winter M J, Owen S F, Murray-Smith R, Panter G H, Hetheridge M J and Kinter L B 2010 Using data from drug discovery and development to aid the aquatic environmental risk assessment of human pharmaceuticals: Concepts, considerations, and challenges *Integr. Environ. Assess Manage.* **6** 38
6. Lushington G H, Dong Y and Theertham B 2013 Chemical informatics and the drug discovery knowledge pyramid *Comb. Chem. High Throughput Screening* **16** 764
7. Bajorath J 2017 Compound Data Mining for Drug Discovery *Methods Mol. Biol.* **1526** 247
8. Boran A D and Iyengar R 2010 Systems approaches to polypharmacology and drug discovery *Curr. Opin. Drug Discovery Dev.* **13** 297
9. Badrinarayan P and Sastry G N 2011 Virtual high throughput screening in new lead identification *Comb. Chem. High Throughput Screening* **14** 840
10. Reddy A S, Pati S P, Kumar P P, Pradeep H N and Sastry G N 2007 Virtual screening in drug discovery – a computational perspective *Curr. Protein Pept. Sci.* **8** 329
11. Collins P Y, Patel V, Joestl S S, March D, Insel T R, Daar A S; Scientific Advisory Board and the Executive Committee of the Grand Challenges on Global Mental Health, Anderson W, Dhansay M A, Phillips A, Shurin S, Walport M, Ewart W, Savill S J, Bordin I A, Costello E J, Durkin M, Fairburn C, Glass R I, Hall W, Huang Y, Hyman S E, Jamison K, Kaaya S, Kapur S, Kleinman A, Ogunniyi A, Otero-Ojeda A, Poo M M, Ravindranath V, Sahakian B J, Saxena S, Singer P A and Stein D J 2011 Grand challenges in global mental health *Nature* **475** 27
12. Varmus H, Klausner R, Zerhouni E, Acharya T, Daar A S and Singer P A 2003 Public health. Grand Challenges in Global Health *Science* **302** 398
13. Paul S M, Mytelka D S, Dunwiddie C T, Persinger C C, Munos B H, Lindborg S R and Schacht A L 2010 How to improve R&D productivity: The pharmaceutical industry's grand challenge *Nat. Rev. Drug Discov.* **9** 203
14. Dubois D J 2010 Grand Challenges in Pharmacoeconomics and Health Outcomes *Front. Pharmacol.* **1** 7
15. Yildirim O, Gottwald M, Schüler P and Michel MC M 2016 Opportunities and Challenges for Drug Development: Public–Private Partnerships, Adaptive Designs and Big Data *Front. Pharmacol.* **7** 461
16. Gostin L O and Mok E A 2009 Grand challenges in global health governance *Br. Med. Bull.* **90** 78
17. Pai M, Daftary A and Satyanarayana S 2016 TB control: Challenges and opportunities for India *Trans. R. Soc. Trop. Med. Hyg.* **110** 158
18. Wells T N, Willis P, Burrows J N and Hooft V H R 2016 Open data in drug discovery and development: Lessons from malaria *Nat. Rev. Drug Discov.* **15** 661

19. Van Voorhis W C, Adams J H, Adelfio R, Ahyong V, Akabas M H, Alano P, Alday A, Alemán Resto Y, Alsibaee A, Alzualde A, Andrews K T, Avery S V, Avery V M, Ayong L, Baker M, Baker S, Ben Mamoun C, Bhatia S, Bickle Q, Bounaadja L, Bowling T, Bosch J, Boucher L E, Boyom F F, Brea J, Brennan M, Burton A, Caffrey C R, Camarda G, Carrasquilla M, Carter D, Belen Cassera M, Chih-Chien Cheng K, Chindaudomsate W, Chubb A, Colon B L, Colón-López D D, Corbett Y, Crowther G J, Cowan N, D'Alessandro S, Le Dang N, Delves M, DeRisi J L, Du A Y, Duffy S, Abd El-Salam El-Sayed S, Ferdig M T, Fernández Robledo J A, Fidock D A, Florent I, Fokou P V, Galstian A, Gamo F J, Gokool S, Gold B, Golub T, Goldgof G M, Guha R, Guiguemde W A, Gural N, Guy R K, Hansen M A, Hanson K K, Hemphill A, Hooft van Huijsduijnen R, Horii T, Horrocks P, Hughes T B, Huston C, Igarashi I, Ingram-Sieber K, Itoe M A, Jadhav A, Naranuntarat Jensen A, Jensen L T, Jiang R H, Kaiser A, Keiser J, Ketas T, Kicka S, Kim S, Kirk K, Kumar V P, Kyle D E, Lafuente M J, Landfear S, Lee N, Lee S, Lehane A M, Li F, Little D, Liu L, Llinás M, Loza M I, Lubar A, Lucantoni L, Lucet I, Maes L, Mancama D, Mansour N R, March S, McGowan S, Medina Vera I, Meister S, Mercer L, Mestres J, Mfopa A N, Misra R N, Moon S, Moore J P, Morais Rodrigues da Costa F, Müller J, Muriana A, Nakazawa Hewitt S, Nare B, Nathan C, Narraidoo N, Nawaratna S, Ojo K K, Ortiz D, Panic G, Papadatos G, Parapini S, Patra K, Pham N, Prats S, Plouffe D M, Poulsen S A, Pradhan A, Quevedo C, Quinn R J, Rice C A, Abdo Rizk M, Ruecker A, St Onge R, Salgado Ferreira R, Samra J, Robinett N G, Schlecht U, Schmitt M, Silva Villela F, Silvestrini F, Sinden R, Smith D A, Soldati T, Spitzmüller A, Stamm S M, Sullivan D J, Sullivan W, Suresh S, Suzuki B M, Suzuki Y, Swamidass S J, Taramelli D, Tchokouaha L R, Theron A, Thomas D, Tonissen K F, Townson S, Tripathi A K, Trofimov V, Udenze K O, Ullah I, Vallieres C, Vigil E, Vinetz J M, Voong Vinh P, Vu H, Watanabe N A, Weatherby K, White P M, Wilks A F, Winzeler E A, Wojcik E, Wree M, Wu W, Yokoyama N, Zollo P H, Abla N, Blasco B, Burrows J, Laleu B, Leroy D, Spangenberg T, Wells T and Willis P A 2016 Open Source Drug Discovery with the Malaria Box Compound Collection for Neglected Diseases and Beyond *PLoS Pathog.* **28** e1005763
20. Williamson A E, Ylioja P M, Robertson M N, Antonova-Koch Y, Avery V, Baell J B, Batchu H, Batra S, Burrows J N, Bhattacharyya S, Calderon F, Charman S A, Clark J, Crespo B, Dean M, Debbert S L, Delves M, Dennis A S, Deroose F, Duffy S, Fletcher S, Giaever G, Hallyburton I, Gamo F J, Gebbia M, Guy R K, Hungerford Z, Kirk K, Lafuente-Monasterio M J, Lee A, Meister S, Nislow C, Overington J P, Papadatos G, Patiny L, Pham J, Ralph S A, Ruecker A, Ryan E, Southan C, Srivastava K, Swain C, Tarnowski M J, Thomson P, Turner P, Wallace I M, Wells T N, White K, White L, Willis P, Winzeler E A, Wittlin S and Todd M H 2016 Open Source Drug Discovery: Highly Potent Antimalarial Compounds Derived from the Tres Cantos Arylpyrroles *ACS Cent. Sci.* **2** 687
21. Rottmann M, McNamara C, Yeung B K, Lee MC, Zou B, Russell B, Seitz P, Plouffe D M, Dharia N V, Tan J, Cohen S B, Spencer K R, González-Páez GE, Lakshminarayana S B, Goh A, Suwanarusk R, Jegla T, Schmitt E K, Beck H P, Brun R, Nosten F, Renia L, Dartois V, Keller T H, Fidock D A, Winzeler E A and Diagana T T 2010 Spiroindolones, a potent compound class for the treatment of malaria *Science* **329** 1175
22. Meister S, Plouffe D M, Kuhen K L, Bonamy G M, Wu T, Barnes S W, Bopp S E, Borboa R, Bright A T, Che J, Cohen S, Dharia N V, Gagaring K, Gettayacamin M, Gordon P, Groessl T, Kato N, Lee M C, McNamara C W, Fidock D A, Nagle A, Nam T G, Richmond W, Roland J, Rottmann M, Zhou B, Froissard P, Glynne R J, Mazier D, Sattabongkot J, Schultz P G, Tuntland T, Walker J R, Zhou Y, Chatterjee A, Diagana T T and Winzeler E A 2011 Imaging of Plasmodium liver stages to drive next-generation antimalarial drug discovery *Science* **334** 1372
23. Gamo F J, Sanz L M, Vidal J, de Cozar C, Alvarez E, Lavandera J L, Vanderwall D E, Green D V, Kumar V, Hasan S, Brown J R, Peishoff C E, Cardon L R and Garcia-Bustos J F 2010 Thousands of chemical starting points for antimalarial lead identification *Nature* **465** 305
24. Guiguemde W A, Shelat A A, Bouck D, Duffy S, Crowther G J, Davis P H, Smithson D C, Connelly M, Clark J, Zhu F, Jiménez-Díaz M B, Martinez M S, Wilson E B, Tripathi A K, Gut J, Sharlow E R, Bathurst I, El Mazouni F, Fowble J W, Forquer I, McGinley P L, Castro S, Angulo-Barturen I, Ferrer S, Rosenthal P J, Derisi J L, Sullivan D J, Lazo J S, Roos D S, Riscoe M K, Phillips M A, Rathod P K, Van Voorhis W C, Avery V M and Guy R K 2010 Chemical genetics of Plasmodium falciparum *Nature* **465** 311
25. Wells T N 2010 Microbiology. Is the tide turning for new malaria medicines? *Science* **329** 1153
26. Rees S 2015 The promise of open innovation in drug discovery: An industry perspective *Future Med. Chem.* **7** 1835
27. Allarakhia M 2014 The successes and challenges of open-source biopharmaceutical innovation *Expert Opin. Drug Discovery* **9** 459
28. *Global Tuberculosis report*http://apps.who.int/iris/bitstream/10665/250441/1/9789241565394-eng.pdf?ua=1 (accessed on 31st January 2017)
29. *Guidelines for treatment of tuberculosis, fourth edition*http://apps.who.int/iris/bitstream/10665/44165/1/9789241547833_eng.pdf?ua=1&ua=1 (accessed on 31st December 2016)
30. Esmail H, Barry C E, Young D B and Wilkinson R J 2014 The ongoing challenge of latent tuberculosis*Philos. Trans. R. Soc. London, Ser. B* **369** 20130437
31. Davis C E, Carpenter J L, McAllister C K, Matthews J, Bush B A and Ognibene A J 1985 Tuberculosis. Cause of death in antibiotic era *Chest* **88** 726
32. Frieden T R, Sterling T R, Munsiff S S, Watt C J and Dye C 2003 Tuberculosis *Lancet* **362** 887
33. Dye C, Scheele S, Dolin P, Pathania V and Raviglione M C 1999 Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project *JAMA* **282** 677
34. Norton B L and Holland D P 2012 Current management options for latent tuberculosis: a review *Infect. Drug Resist.* **5** 163

35. Johnson R, Streicher E M, Louw G E, Warren R M, van Helden P D and Victor T C 2006 Drug resistance in Mycobacterium tuberculosis *Curr. Issues Mol. Biol.* **8** 97
36. Kremer L S and Besra G S 2002 Current status and future development of antitubercular chemotherapy *Expert Opin. Invest. Drugs* **11** 1033
37. Chan E D and Iseman M D 2008 Multidrug-resistant and extensively drug-resistant tuberculosis: a review *Curr. Opin. Infect. Diseases* **21** 587
38. Daley C L and Caminero J A 2013 Management of multidrug resistant tuberculosis *Semin. Respir. Crit. Care Med.* **34** 44
39. Choudhury C, Priyakumar U D and Sastry G N 2014 Molecular dynamics investigation of the active site dynamics of mycobacterial cyclopropane synthase during various stages of the cyclopropanation process *J. Struct. Biol.* **187** 38
40. Choudhury C, Priyakumar U D and Sastry G N 2015 Dynamics based pharmacophore models for screening potential inhibitors of mycobacterial cyclopropane synthase *J. Chem. Inf. Model.* **55** 848
41. Choudhury C, Priyakumar U D and Sastry G N 2016 Dynamic ligand-based pharmacophore modeling and virtual screening to identify mycobacterial cyclopropane synthase inhibitors *J. Chem. Sci.* **128** 719
42. Janardhan S, Ram Vivek M and Sastry G N 2016 Modeling the permeability of drug-like molecules through the cell wall of Mycobacterium tuberculosis: an analogue based approach *Mol. Biosyst.* **12** 3377
43. Reddy A S, Amarnath H S, Bapi R S, Sastry G M and Sastry G N 2008 Protein ligand interaction database (PLID) *Comput. Biol. Chem.* **32** 387
44. Srivastava H K, Choudhury C and Sastry G N 2012 The efficacy of conceptual DFT descriptors and docking scores on the QSAR models of HIV protease inhibitors *Med. Chem.* **8** 811
45. Dobson C M 2004 Chemical space and biology *Nature* **432** 824
46. Lipinski C and Hopkins A 2004 Navigating chemical space for biology and medicine *Nature* **432** 855
47. Barker A, Kettle J G, Nowak T and Pease J E 2013 Expanding medicinal chemistry space *Drug Discovery Today* **18** 298
48. Reymond J L and Awale M 2012 Exploring Chemical Space for Drug Discovery Using the Chemical Universe Database *ACS Chem. Neurosci.* **3** 649
49. Oprea T I and Gottfries J 2001 Chemography: The art of navigating in chemical space *J. Com. Chem.* **3** 157
50. Xu J and Stevenson J 2000 Drug-like index: A new approach to measure drug-like compounds and their diversity *J. Chem. Inf. Comput. Sci.* **40** 1177
51. Irwin J J and Shoichet B K 2005 ZINC-a free database of commercially available compounds for virtual screening *J. Chem. Inf. Model.* **45** 177
52. Bolton E E, Wang Y, Thiessen P A and Bryant S H 2008 PubChem: Integrated platform of small molecules and biological activities *Annu. Rep. Comput. Chem.* **4** 217
53. Wang Y, Xiao J, Suzek T O, Zhang J, Wang J, Zhou Z, Han L, Karapetyan K, Dracheva S and Shoemaker B A 2012 PubChem's BioAssay database *Nucleic Acids Res.* **40** D400
54. Vasilevich N I, Kombarov R V, Genis D V and Kirpichenok M A 2012 Lessons from natural products chemistry can offer novel approaches for synthetic chemistry in drug discovery *J. Med. Chem.* **55** 7003
55. Milne G W and Miller J 1986 The NCI drug information system. 1. System overview *J. Chem. Inf. Comput. Sci.* **26** 154
56. Wishart D S, Knox C, Guo A C, Shrivastava S, Hassanali M, Stothard P, Chang Z and Woolsey J 2006 DrugBank: A comprehensive resource for in silico drug discovery and exploration *Nucleic Acids Res.* **34** D668
57. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M and Tanabe M 2014 Data, information, knowledge and principle: Back to metabolism in KEGG *Nucleic Acids Res.* **42** D199
58. Pence H E and Williams A 2010 ChemSpider: An online chemical information resource *J. Chem. Educ.* **87** 1123
59. Chen C Y 2011 TCM Database@ Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico *PLoS One* **6** e15939
60. Kiss R, Sandor M and Szalai F A 2012 http://Mcule.com: A public web service for drug discovery *J. Cheminf.* **4** P17
61. Olah M, Rad R, Ostopovici L, Bora A, Hadaruga N, Hadaruga D, Moldovan R, Fulias A, Mractc M and Oprea T I 2008 In *Small Molecules to Systems Biology and Drug Design -WOMBAT and WOMBAT-PK: Bioactivity Databases for Lead and Drug Discovery Chemical Biology* S L Schreiber, T M Kapoor and G Wess (Eds.) (Weinheim: Wiley-VCH Verlag GmbH) Vol. **1–3** p. 760
62. Anna G, Louisa J B, Bento A P and Jon C 2012 ChEMBL: A large-scale bioactivity database for drug discovery *Nucleic Acids Res.* **40** D1100
63. Jiang C, Jin X, Dong Y and Chen M 2016 Kekule.js: An Open Source JavaScript Chemoinformatics Toolkit *J. Chem. Inf. Model.* **56** 1132
64. Wojcikowski M, Zielenkiewicz P and Siedlecki P 2015 Open Drug Discovery Toolkit (ODDT): a new open-source player in the drug discovery field *J. Cheminf.* **7** 26
65. Kuhn T, Willighagen E L, Zielesny A and Steinbeck C 2010 CDK-Taverna: An open workflow environment for chemoinformatics *BMCBioinformatics* **11** 159
66. Steinbeck C, Han Y, Kuhn S, Horlacher O, Luttmann E and Willighagen E 2003 The Chemistry Development Kit (CDK): An open-source Java library for Chemo- and Bioinformatics *J. Chem. Inf. Comput. Sci.* **43** 493
67. Wolstencroft K, Haines R, Fellows D, Williams A, Withers D, Owen S, Soiland-Reyes S, Dunlop I, Nenadic A, Fisher P, Bhagat J, Belhajjame K, Bacall F, Hardisty A, Nieva H A, Balcazar V M P, Sufi S and Goble C 2013 The Taverna workflow suite: Designing and executing workflows of Web Services on the desktop, web or in the cloud *Nucleic Acids Res.* **41** W557
68. Beisken S, Meinl T, Wiswedel B, de Figueiredo L F, Berthold M and Steinbeck C 2013 KNIME-CDK: Workflow-driven cheminformatics *BMC Bioinf.* **14** 257
69. Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A and Taylor J 2010 Galaxy: a web-based genome analysis tool for

experimentalists *Curr. Protoc. Mol. Biol.* Chapter 19 Unit 19.10.1-21
70. Afgan E, Baker D, Beek M V D, Blankenberg D, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Eberhard C, Gruning B, Guerler A, Jackson J H, Kuster G V, Rasche E, Soranzo N, Turaga N, Taylor J, Nekrutenko A and Goecks J 2016 The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update *Nucleic Acids. Res.* **44** W3
71. Blankenberg D, Kuster G V, Bouvier E, Baker D, Afgan E, Stoler N, Galaxy Team, Taylor J and Nekrutenko A 2014 Dissemination of scientific software with Galaxy ToolShed *Genome Biol.* **15** 403
72. *Publicly Accessible Galaxy Servers* https://wiki.galaxyproject.org/PublicGalaxyServers (accessed on 31st December 2016)
73. Hildebrandt A K, Stockel D, Fischer N M, de la Garza L, Kruger J, Nickels S, Rottig M, Scharfe C, Schumann M, Thiel P, Lenhof H P, Kohlbacher O and Hildebrandt A 2015 ballaxy: web services for structural bioinformatics *Bioinformatics* **31** 121
74. O'Boyle N M, Banck M, James C A, Morley C, Vandermeersch T and Hutchison G R 2011 Open Babel: An open chemical toolbox *J. Cheminf.* **3** 33
75. *Landrum G RDKit: Open-Source Cheminformatics* http://www.rdkit.org (accessed on 31st December 2016)
76. Ertl P and Rohde B 2012 The Molecule Cloud - compact visualization of large collections of molecules *J. Cheminf.* **4** 12
77. Peironcely J E, Cherto M R, Fichera D, Reijmers T, Coulier L, Faulon J L and Hankemeier T 2012 OMG: Open Molecule Generator *J. Cheminf.* **4** 21
78. Vainio M J and Johnson M S 2005 McQSAR: a multiconformational quantitative structure-activity relationship engine driven by genetic algorithms *J. Chem. Inf. Model.* **45** 1953
79. Joachims T 1999 *Advances in Kernel Methods- Making Large-Scale SVM Learning Practical* B Scholkopf, C Burges and A Smola (Eds.) (Cambridge: MIT-Press) p. 169
80. Moriarty N W, Grosse-Kunstleve R W and Adams P D 2009 electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation *Acta Crystallogr., D: Biol. Crystallogr.* **65** 1074
81. Dewar M J S, Zoebisch E G, Healy E F and Stewart J J P 1985 Development and use of quantum mechanical molecular models. 76. AM1: A new general purpose quantum mechanical molecular model *J. Am. Chem. Soc.* **107** 3902
82. Trott O and Olson A J 2010 AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading *J. Comput. Chem.* **31** 455
83. *Drug Likeness Tool (DruLiTo)* http://www.niper.ac.in/pi_dev_tools/DruLiToWeb/DruLiTo_index.html (accessed on 31st December 2016)
84. Lipinski C A, Lombardo F, Dominy B W and Feeney P J 2001 Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings *Adv. Drug Delivery Rev.* **46** 3
85. Oprea T I 2000 Property distribution of drug-related chemical databases *J. Comput. -Aided. Mol. Des.* **14** 251
86. Ghose A K, Viswanadhan V N and Wendoloski J J 1999 A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases *J. Comb. Chem.* **1** 55
87. Bickerton G R, Paolini G V, Besnard J, Muresan S and Hopkins A L 2012 Quantifying the chemical beauty of drugs *Nat. Chem.* **4** 90
88. Veber D F, Johnson S R, Cheng H Y, Smith B R, Ward K W and Kopple K D 2002 Molecular properties that influence the oral bioavailability of drug candidates *J. Med. Chem.* **45** 2615
89. Yap C W 2011 PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints *J. Comput. Chem.* **32** 1466
90. Jensen C and Scacchi W 2005 *Collaboration, leadership, control, and conflict negotiation and the netbeans.org open source software development community IEEE* 196b

Molecular Docking and Molecular Dynamics Study of DNA Minor Groove Binders

Ruchi Mishra[1], Anamika Singh Gaur[2], Ramesh Chandra[3,*], Devesh Kumar[4,*]

[1,3,4]Department of Applied Physics, School for Physical Sciences, Babasaheb Bhimrao Ambedkar University, Vidya Vihar, Rae Bareilly Road, Lucknow, [2]Centre for Molecular Modeling, CSIR-Indian Institute of Chemical Technology, Tarnaka, Hyderabad

***Corresponding Author:**
Email: dkclcre@yahoo.com, ramesh.luphy@gamil.com

ABSTRACT

The fundamental problems in drug discovery are based on the process of molecular recognition by small molecules. The binding specificity of DNA-small molecule is identified mainly by studying the hydrogen bonding and polar interactions. Majority of the minor groove binders and their mechanism of action at the molecular level are not well studied. As these small molecules can act as effective therapeutic agents against many diseases, there is a need to have the detailed mechanistic insights on how they interact with DNA. In this study we have investigated the binding mechanism and stability of the complexes using molecular modeling methods. The molecular docking studies were performed to explore the exact binding sites and affinity inside the DNA minor grove. A 5ns molecular dynamics (MD) simulation for the DNA minor groove binders has been performed using AMBER and GROMACS program. Further, to study the systematic deviation of docked complexes during MD simulation, RMSD as a function of time have been analyzed and it has been found that RMSD variation obtained using AMBER and GROMACS MD simulation are approximately same. The binding free energies between the DNA and minor groove binders were calculated and decomposed by molecular mechanics/generalized born surface area (MM-GBSA) and Molecular Mechanics/Poisson−Boltzmann Surface Area (MM-PBSA) methods. The comparative and systematic analysis presented in this study can provide guidance for the choice of MD methods and the designs of new potent inhibitors targeting DNA.

Keywords: *Minor groove binders, Molecular docking, Molecular dynamics (MD), MM-GBSA method, MM-PBSA method.*

INTRODUCTION

Deoxyribonucleic acid (DNA) is the bio molecule has two complementary helical strands running in anti-parallel directions, which carries genetic information from parents to offspring[1]. DNA plays an important role in cellular processes, including cell division (DNA replication) and protein synthesis (Transcription and translation). Most of the anticancer therapies are involved in the interaction of drugs with DNA. The intercalation and groove binding are the two important modes of binding of drug with DNA. Both covalent and non-covalent types of interactions are possible in these two binding modes. Small molecules that can bind between nucleic acid base pairs are categorized as intercalators. These molecules contain planar heterocyclic groups which stack between adjacent DNA base pairs, which results decrease in the DNA helical twisting and lengthening of the DNA. On the other hand, groove binding does not induce large conformational changes in DNA and may be considered similar to standard lock and key models for ligand-macromolecular binding. Such molecules bind to both major and minor groove of nucleic acid. Minor groove binders are crescent in shape and they complement the shape of minor groove[2-5]. The binding mechanism of drug with the DNA minor groove can be described in mainly two steps. In the first step, the transfer of ligand to the DNA minor groove by the electrostatic and hydrophobic interaction. In the second step, various types of non-covalent interactions occur between the ligand and the functional groups of DNA base pairs. These interactions usually include hydrogen bonds, hydrophobic and van der Waals contacts, and electrostatic interactions. Most of the minor groove binding drugs bind to A/T rich region[6-9].

In the present study, two major DNA minor groove binders classes, polyamides and diary lamidines is undertaken. Different interaction models are available to explain protein-ligand binding but these models is not reasonable for DNA-ligand systems because there is no prescribed active site in the DNA, unlike protein/enzymes. Our study examines the available popular molecular modeling approaches for the

estimation of DNA- drug binding free energies and compares with the experimental results. Molecular modeling methods are a powerful tool to investigate various types of non-covalent interactions, which exist between the receptor and ligand. In the number of studies for the minor groove binders, these computational methods have shown good agreement with the experimental results.[10-14]

The special approaches like molecular dynamics simulation is required to understand the complex systems like nucleic acids. The force field used in various studies to simulate nucleic acid simulation includes, CHARMM, AMBER, GROMOS, OPLS, ENCAD and BMS26. There have been many studies for the comparison of force fields for the nucleic acids but still there is need to analyze them critically at the molecular level. AMBER 03 and CHARMM force fields were chosen for our study to simulate the DNA with small molecules. In this paper, we have used the popular molecular mechanics energies combined with the Generalized-Born surface area (MM-GBSA) and Poisson−Boltzmann Surface Area (MM-PBSA) methods to estimate the binding free energy of the binding of small molecules to DNA. These methods have been applied to wide range of molecules to estimate their ligand binding affinities

MATERIALS AND METHODS

Dataset

The crystal data of the B-DNA (1D30, 195D, 1D86 and 102D) were downloaded from the Protein Data Bank[10] and their experimental binding energies were collected from literature[11]. The water molecules and the ligands were removed from the 1D30, 195D, 1D86 and 102D. The drug molecules extracted from these complexes were subjected to geometry optimization using Gaussian 09 at B3LYP/6-31G* level[12].

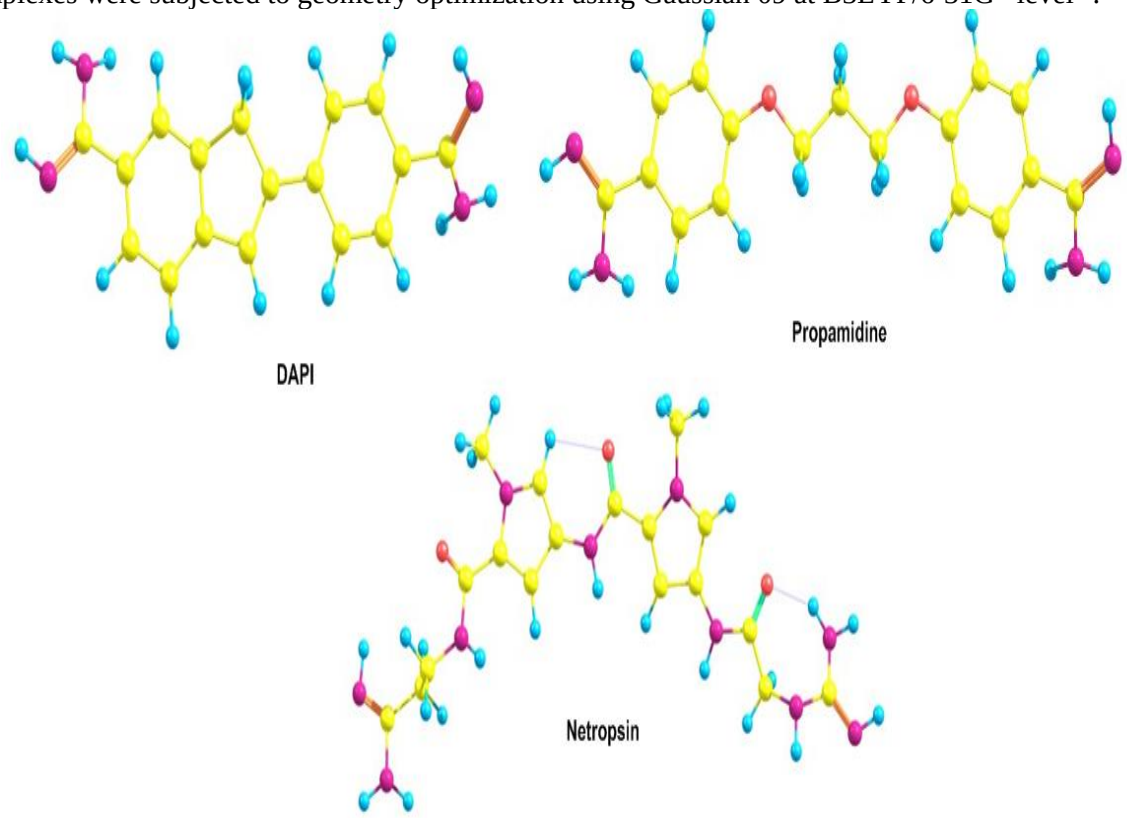


Fig. 1: Optimized structures of DNA minor groove Binders using Gaussian 09 at B3LYP/6-31G* level

Molecular Docking

Molecular docking studies were performed using Autodock 4.2 program[13]. The Gasteiger charges were added to the complex by Autodock Tools (ADT) before performing docking calculations. The binding site was centered on the macromolecule and a grid box was created with 55 × 55 × 60 points and a 0.375 Å grid spacing in which almost covered the entire DNA were involved. Docking simulations were performed using

the classical Lamarckian genetic algorithm (LGA). The 20 LGA runs with maximum of 2500000 energy evaluations performed. In addition, the other parameters were set to default. The pose with lowest energy of binding or binding affinity was extracted and aligned with receptor structure for further analysis.

Simulations

Molecular Dynamics simulation of 5 ns was carried out using the AMBER 15[14-17] and GROMACS 4.5.6 packages[18]. AMBER15 program was used for MD simulations of the selected docked poses. The 'leaprc.gaff" (generalized amber force field) was used to prepare the ligands, while the 'leaprc.ff03' was used for the DNA. The 'add ions' command implemented in 'tleap' of AMBER15 was used to add the Na+ ions explicitly to neutralize the system. Each system was placed in a box of TIP3P water by using 'solvateOct' command with the minimum distance between any solute atom and the boundary of the box is set to 8Å. Energy minimization was then performed to achieve the nearest stable low energy conformations (500 steps of each steepest descent and conjugate gradient method), 50 ps of heating and 50ps of density equilibration with weak restraints on the complex followed by 500ps of constant pressure equilibration at 300K. Cut off size of 12 Å was used for MD simulations. All long –range electrostatics were included by means of a Particle mesh Ewald (PME) method. All hydrogen and heavy atom bonds were constrained by the shake method, and simulations were performed with a 2fs time step and langvenin dynamics was used for temperature control. The same conditions for the final phase of equilibration were used for the production run and the coordinates were recorded at every 10 ps. Periodic boundary conditions were used for the final production run. Five hundred snapshots of the complex are obtained at every 10 ps from the MD trajectories, and all the water molecules and ions were removed before MMPBSA/MMGBSA[19] calculations using the "extract_coords.mmpbsa" script and the $\Delta G_{bind} - PB/GB$ values were calculated using the "binding_energy.mmpbsa" script.

In GROMACS MD simulation the topology and co-ordinate files for the DNA were generated by pdb2gmx program of the GROMACS package taking parameters from CHARMM[20] all atom force field and for the ligand using SwissParam[21] Web server. The coordinate and topology files of DNA and ligand were merged to obtain the final starting structure and topology file for each complex. The drug-DNA complex was placed in the center of dodecahedron periodic box. The system was then solvated in TIP3P water Molecules. The total charge on the system was then neutralized by adding counter ions. The energy was minimized using steepest descent algorithm. Then the system was heated to 300K during 50ps of constant volume simulation with 2fs time step. The pressure was equilibrated to 1 atm during 50ps NPT simulation with 2 fs time step. In both the simulations a restrained with force constant of 1000kJ/ (mol/nm^2). Both temperature and pressure were regulated using Berendsen algorithm. Production simulations were performed 5ns with a 2fs time step. The temperature and pressure were maintained at 300 K and 1 atm using the v-rescale temperature and Parninello-Rahman pressure coupling method. The binding energy was calculated using g_mmpbsa[22]. Binding energy of each snapshot was calculated for each complex. The entropy contribution was not included in the binding energy.

RESULTS AND DISCUSSIONS

Molecular docking studies

Molecular docking calculations shows that all the ligands bind to AT-rich region of DNA with good docking fitness score (Table 1). Comparison between experimental and calculated binding energies obtained from docking studies shows that the complex having lowest experimental binding energy also has the lowest calculated binding energy and vice-versa. This shows the ability of Autodock in predicting the correct binding modes for drug DNA complex. The PDB ID **1D30** have lowest binding energy, this shows that the interaction between DNA duplex of sequence d (CGCGAATTCGCG) $_2$ and DAPI (fig. 2) is more stable in comparison to other complexes. The further studies were performed on the best pose with lowest binding energy for all complexes.

Table 1: List of PDB IDs taken for the study with their DNA sequence, calculated binding energies (kcal/Mol) and experimental binding energies obtained from literature

PDB ID	DNA sequence	Ligand	$\Delta G_{calc.}$[a]	$\Delta G_{exp.}$[b]
1D30	5'-CGCGAATTCGCG-3'	DAPI	-9.07	-8.8
1D86	5'-CGCGAATTCGCG-3'	Netropsin	-5.42	-8.7
195D	5'-CGCGTTAACGCG-3'	Netropsin	-4.35	-8.0
102D	5'-CTTTTGCAAAAG-3'	Propamidine	-4.61	-8.2

[a] Calculated binding Energy in kcal/mol.
[b] Experimental binding energy kcal/mol.

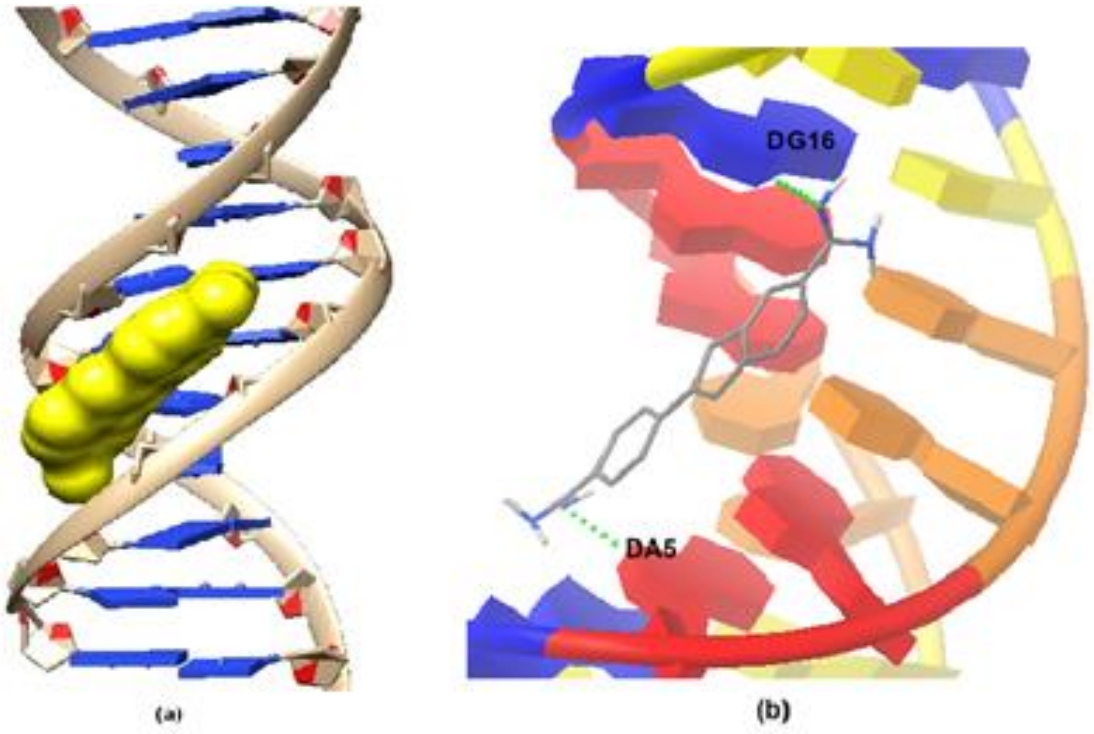


Fig. 2: A molecular docked model for DAPI with DNA duplex of sequence d (CGCGAATTCGCG) $_2$ (PDB ID: 1D30). (a)The full view of docking between DAPI and 1D30; (b) The binding mode between DAPI and 1D30 and the green dashed line showing hydrogen bond interactions between them

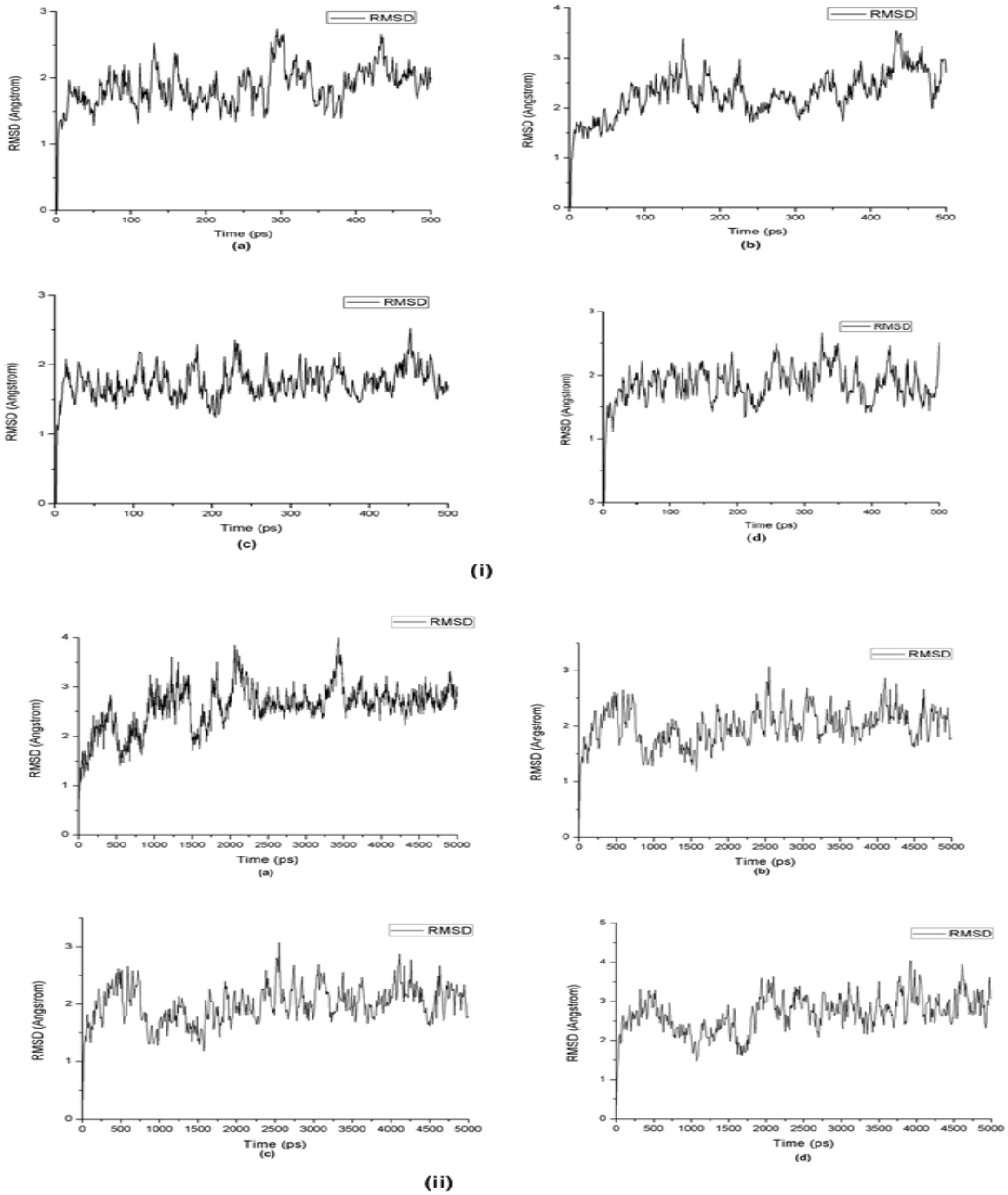


Fig. 3: Plots of RMSD vs. time of their trajectory for PDB ID (a) 1D30 (b) 1D86 (c) 102D (d) 195D obtained from (i) AMBER (ii) GROMACS

Molecular dynamics studies

MD simulation from AMBER15 and GROMACS 4.5.6 were performed for the best poses selected from the docking studies. RMSD as a function of time is plotted for all the complexes in fig 3 to obtain the systematic deviation of complexes. The RMSD profile shows that the ligands remain bound to the DNA near the preferential binding site. It has been observed that the all the four complexes show RMSD variation from both types of simulations. The range of RMSD from fig. 3 is between 1Å-3Å from AMBER and 1Å-3.5Å for GROMACS during the course of simulation. Convergence of RMSD values shows the stability of complex. In order to identify the conformational changes during 5ns of MD simulation, the snapshots were extracted from the MD trajectory at every 10 ps for both AMBER and GROMACS (fig 4). It has been

observed from snapshots that ligand forms interactions with AT-rich region of DNA duplex and bounded in minor groove up to the end of simulation.

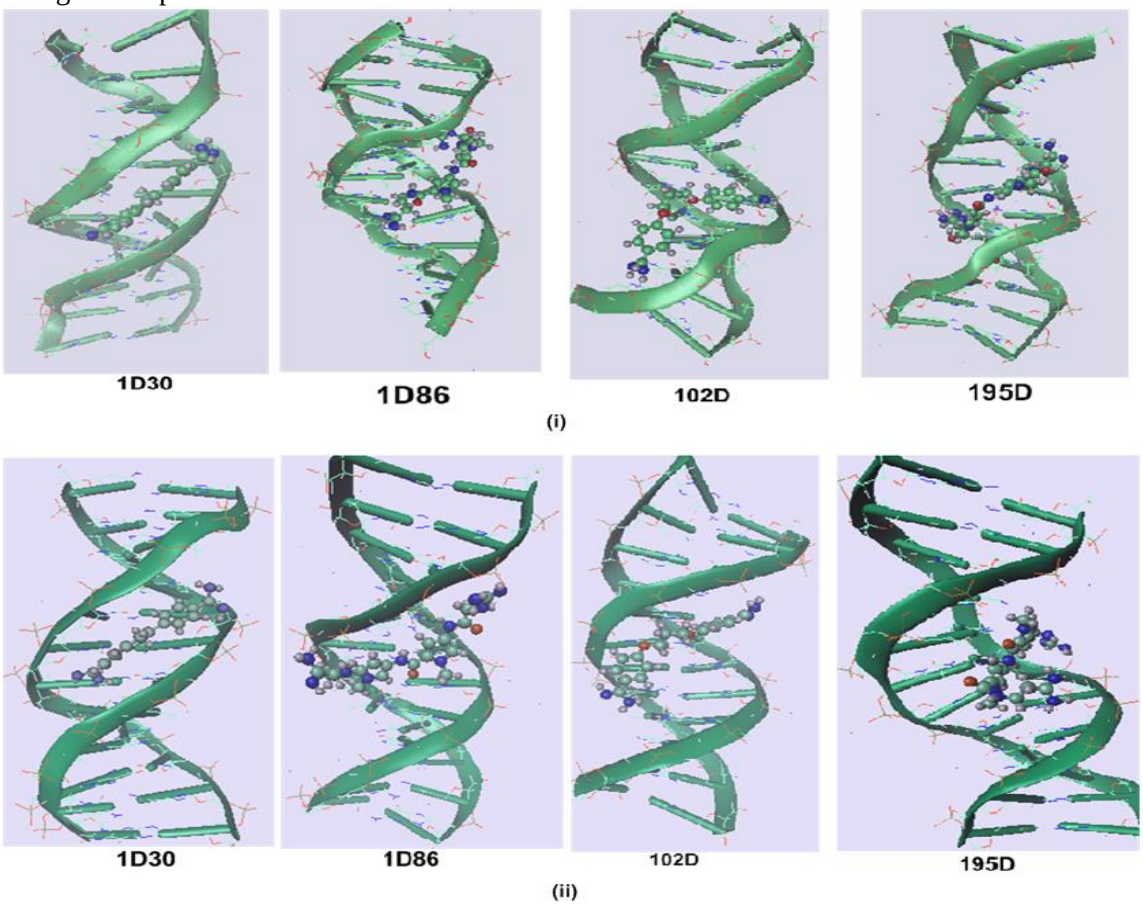


Fig. 4: Snapshots of different DNA-ligand complexes at 5 ns MD simulation for PDB ID (a) 1D30 (b) 1D86 (c) 102D (d) 195D obtained from (i) AMBER (ii) GROMACS

Further, MMPBSA/MMGBSA calculations were performed from AMBER 15 by using MD trajectories to obtain binding energy values. Binding energy values predict the strength of ligand with their respective receptors. Calculated binding free energy of all four drug-DNA complexes with the contribution of van der Waals, electrostatic, salvation energy etc. is shown in fig. 5 and it was found that the total Binding free energy of complex with PDB ID 1D30 is lowest binding energy (-25.52 kcal/mole), so this complex is slightly more stable than others.

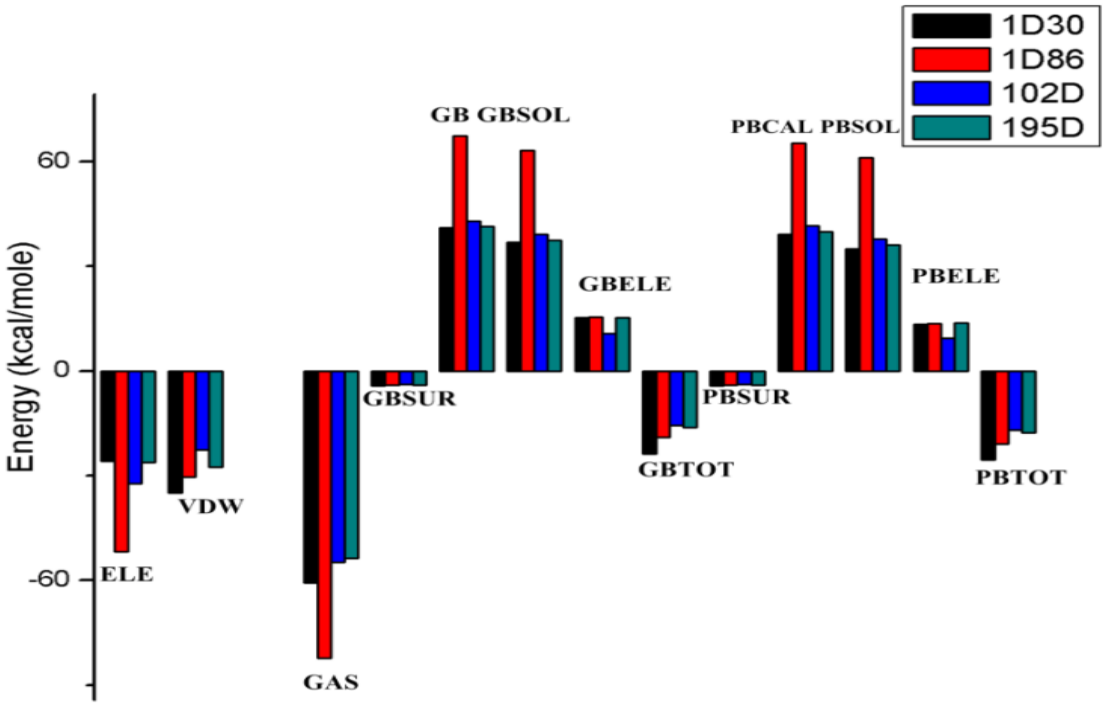


Fig. 5: Histogram depicting view of the contribution of various energy components to the binding free energy

ELE: electrostatic energy; **VDW**: van der Waals energy; INT: internal energy; GAS: total gas-phase energy; PBSUR/GBSUR: nonpolar contribution to the solvation free energy; PBCAL/GB: electrostatic contribution to the solvation free energy calculated by PB or GB, respectively; PBSOL/GBSOL: sum of non-polar and polar contribution to solvation; PBELE/GBELE: sum of the electrostatic solvation free energy and MM electrostatic energy; PBTOT/GBTOT: final estimated binding free energy.
The binding free energy has been also recalculated for all the four complexes using the g_mmpbsa for the MD trajectories obtained from GROMACS MD simulation. It has been observed that binding energy obtained from both types of programs that the PDB ID 1D30 having lowest binding energy. To solve the PB equation g_mmpbsa uses the APBS package whereas mm_pbsa.pl uses the PBSA program of the AMBER suite. It has been observed that the energies calculated using g_mmpbsa and the AMBER MM-PBSA package is approximately similar and the difference of 1-3 kcal/mol has been observed due to the difference ΔG_{polar} (Table 3). The difference in ΔG_{polar} observed because of different algorithms, implemented in APBS and PBSA.

The binding energy calculation shows that DNA-DAPI complex is more stable than others. So the MD structures of above complex were selected to study the hydrogen bonds which retained throughout the MD simulation. The results show that only few hydrogen bonds that were present in the original complex (energy minimized) were retained during the simulations. In the DNA-DAPI complex two hydrogen bonds between LIG25:H4-DT19:O2 and LIG25:H5-DT20-O4' were retained throughout the simulation using AMBER and GROMACS. It has been also observed that hydrogen bonds, mainly forms between minor groove binders and the functional groups of the bases exposed in the grooves via their end groups and also through their amide and linker group.

Table 2: Comparison of binding energy components obtained from AMBER MM-PBSA and g_mmpbsa

PDB Id	program	ΔE_{elec} [a]	ΔE_{vdw} [b]	ΔG_{polar} [c]	$\Delta G_{nonpolar}$ [d]	$\Delta G_{binding}$ [e]
1D30	mm_pbsa.pl	-25.80±6.54	-34.83±3.39	39.27±6.40	-4.15±0.22	-25.52±3.64
	g_mmpbsa	-12.01±5.18	-16.42±7.24	17.84±7.12	-1.71±0.58	-25.95±10.11
1D86	mm_pbsa.pl	-51.80±13.59	-30.43±4.80	65.35±12.44	-4.10±0.33	-20.98±6.58
	g_mmpbsa	-17.60±12.47	-27.36±12.61	25.41±14.16	-2.92±1.16	-22.46±13.18
102D	mm_pbsa.pl	-32.25±10.30	-22.56±3.63	41.63±8.77	-3.76±0.20	-16.94±4.06
	g_mmpbsa	-11.91±4.10	-26.51±5.13	20.65±5.75	-2.89±0.41	-20.66±7.33
195D	mm_pbsa.pl	-26.25±14.81	-27.46±3.69	39.99±14.48	-3.93±0.39	-17.65±4.36
	g_mmpbsa	-26.26±4.31	-34.99±3.61	37.72±6.65	-3.75±0.27	-27.30±4.17

[a] Electrostatic component to the binding energy in kcal/mol.
[b] van der Waals component to the binding energy in kcal/mol.
[c] Polar solvation energy.
[d] Non-polar solvation energy.
[e] Binding energy.

CONCLUSIONS

The computational studies were performed to evaluate and analyze the binding energies of DNA minor groove binders. The focus of the study is to provide a detailed perspective on drug-DNA interactions at the molecular level. Thus, our study attempts to give detail insight on the complexity in binding modes of small molecules to DNA. Theses analysis will be helpful for the improvement of existing minor groove binders, and also in designing novel chemical entities which can act as good DNA inhibitor. This will provide a theoretical protocol for complementing experimental techniques, generation of database for structure-energy relationship in drug-DNA complexes.

ACKNOWLEDGMENTS

RM likes to acknowledge UGC for the financial support. DK is DST-Ramanujan Fellow (SR/S2/RJN-11/2008). The authors are thankful to Dr G. Narahari Sastry for his support and discussions.

REFERENCES

1. Watson JD, Crick FHC, Nature, 1953,171:737–738.
2. Song YM, Wu Q, Yang PJ, Luan NN, Wang LF, Liu YM, J. Inorg. Biochem., 2006,100:1685–1691.
3. Dervan PB, Bioorg. Med. Chem., 2001,9:2215–2235.
4. Dervan PB, Edelson BS, Curr. Opin. Struct. Biol., 2003,13:284–299.
5. Masta A, Gray PJ, Phillips DR, Nucleic Acids Res., 1995, 23:3508–3515.
6. Tse WC, Boger DL, Chemistry&Biology, 2004,11:1607-1617.
7. Nunn MC, Neidle S, J. Med. Chem., 1995,38:2317-2325.
8. Kennard O, Pure &Appl. Chem., 1993,65:1213-1222.
9. Srivastava HK, Chourasaia H, Kumar D, Sastry GN, Chem. Inf. and Model, 2011,51:558-571.

10. Kamal A, Shetti RV, Rmaiah MJ, Swapna P, Reddy KS, Mallareddy A, Rao MPN, Chourasia M, Sastry GN, Juvekar A, Zingde S, Sarma P, Pushpavalli SN, Bhadra MP, Med. Chem. Comm., 2011,2:780-788.
11. Kamal A, Shankaraiah N., Reddy Ch. R, Prabhakar S, Markandeya N, Srivastava HK, Sastry GN, Tetrahedro*n*, 2010,66:5498-5506.
12. Kamal A, Reddy KS, Khan MNA, Shetti RVCRNC, Ramaiah MJ, Pushpavalli SNC VL, Srinivas, C, Pal-Bhadra, M, Chourasia M, Sastry GN, Juvekar A, Zingde S, Barkume M, Bioorg. Med. Chem., 2010,18:4747-4761.
13. Kamal A, Bharathi EV, Ramaiah MJ, Dastagiri D, Reddy JS, Viswanath A, Sultana F, Pushpavalli SNCVL, Pal-Bhadra M, Srivastava HK, Sastry GN, Juvekar A, Sen S, Zingde S. Bioorg. Med. Chem., 2010**,**18:526-542.
14. Kamal A, Rajender, Reddy DR, Reddy MK, Balakishan G, Shaik, TB, Chourasia M, Sastry GN, Bioorg. Med. Chem., 2009,17:1557-1572.
15. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE ."The Protein Data Bank". Nucleic Acids Res. 2000,28:235–242.
16. Shaikh SA, Jayaram B, J. Med. Chem., 2007,50:2240–2244.
17. Gaussian 09, Revision E.01, Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, .Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA, Jr., Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas O, Foresman JB, Ortiz JV, Cioslowski J , Fox DJ, Gaussian, Inc., Wallingford CT, 2009.
18. Morris G, Goodsell D, Halliday R, Huey R, Hart W, Belew R, Olson AJ, J. Comput. Chem., 1998,19:1639–1662.
19. Case D. et al. AMBER 15, University of California, San Francisco, 2015.
20. Salomon-Ferrer R, Case DA, Walker RC, WIREs Comput. Mol. Sci., 2013,3:198-210.
21. Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM, Jr., Onufriev A, Simmerling C, Wang B and Woods R. J. Computat. Chem., 2005,26:1668-1688.
22. Cheatham TE, III and Case DA. Biopolymers, 2013,99:969-977.
23. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hes B, Lindahl E, Bioinformatics, 2013,29:845−854.
24. Thompsan JJ, Lill MA, Journal of Chemical Information and Modeling, 2011,51:2680–2689.
25. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, uvench OG, Lopes P, Vorobyov I, MacKerell Jr. AD, Journal of Computational Chemistry, 2010,31:671-690.
26. Zoete V, Cuendet MA, Grosdidier A, Michielin O, J. Comput. Chem., 2011,32:2359-68.
27. Kumari et al J. Chem. Inf. Model., 2014,54:1951-1962.

REVIEW ARTICLE

International Journal of Science, Technology and Society
Vol 3, No. 2 (2017) : 11-27

A review on theoretical studies of various types of Drug-DNA Interaction

Ruchi Mishra, Asheesh Kumar, Ramesh Chandra, Devesh Kumar*

Department of Physics, School of Physical and Decision Sciences, Babasaheb Bhimrao Ambedkar University, VidyaVihar, Rae Bareilly Road, Lucknow-226 025, India.

Publication Info

Article history:
Received : 17.10.2017
Accepted : 20.12.2017
DOI: https://doi.org/10.18091/ijsts.v3i02.11408

Key words:
Minor Groove Binders, Intercalators, Major Groove Binders, Molecular Docking, Molecular Dynamics, MMPBSA/ MMGBSA, QM/MM.

Corresponding author:
Devesh Kumar

Email:
*dkclcre@yahoo.com

ABSTRACT

A large number of the currently used chemotherapeutic anticancer agents fall into the category of DNA-binding drugs. Study of interactions of various drugs with DNA plays a key role in pharmacology. Due to the potential application of such drugs to cancer and beyond, further discovery and characterization of such compounds are of considerable interest. The combinations of distinct binding modes will improve the stability of recognition and enhance target specificity with respect to both DNA structure as well as sequence. The current review gives an overview of the recently used computational chemical techniques to understand mechanism of drug-DNA interaction. The discussions will provide a theoretical protocol for complementing experimental techniques, generation of database for structure activity/property relationship in drug-DNA complexes. This will be helpful for the improvement of existing drugs, design of new drugs etc.

INTRODUCTION

Deoxyribonucleic acid (DNA) was first discovered by Friedrich Miescher, when he was working with white blood cells obtained from the pus drained out from surgical bandages and determined that the DNA was rich in phosphorous and acidic in nature. However the role of DNA to store heredity information was not reported before 1940s until Avery and co-workers published that the nucleic acids are the genetic information carriers and not proteins (Avery *et al.*, 1944). In 1950 Chargaff recognized that the composition of DNA is unique for each and every species. Chargaff also found that when DNA is broken into its components, the amount of guanine fluctuated from one organism to another is always equal to the cytosine and the amount of cytosine is equal to the amount of thymine (Chargaff, 1951). Rosalind Frankllin elucidates basic helical structure of DNA on the basis of X-ray crystallography technique. In 1953, Watson and Crick scooped Franklin's and Chargaff's information and cracked the code of DNA structure (Chargaff, 1950; Watson, 1953a; Watson, 1953b). They recognized that the relationship between the nitrogenous bases suggested by Chargaff may be due to the complementary base pairing between adenine-thymine and guanine-cytosine and due to this type of base pairing they discovered the hydrogen bonding between these bases, which is currently known as Watson-Crick hydrogen bonding. With this information they modeled a right-handed double helical structure of DNA in which phosphate backbone lied outside the helix and the bases are held together by hydrogen bonding pointed towards the center of helix.

In 1979, a first crystal structure of left-handed double helical DNA $d(CGCGCG)_2$ at atomic resolution was reported, known as Z-DNA (Wang, 1979). After a year, the single-crystal structure analysis of right-handed B-DNA, with the self-complementary dodecamer sequence $d(CGCGAATTCGCG)_2$ was discovered by Dickerson and co-workers. This dodecamer is one of the most studied DNA fragments (Wing, 1980). These discoveries revealed that how the genetic information passes from one to next generation. The most common conformations of DNA are B-, A-, Z-DNA and B-form DNA is the most common occurring conformation. In this type of DNA, the base pairs are perpendicular to the helix axis and twisted by 36^0 with respect to each other. A single turn in the double helix consists of 10 base pairs (Table 1). The two strands of the

Table 1. Structural properties of A-, B-, Z-form DNA.

Conformation	Helix Sense	Twist/bp (Å)	Rise/bp (Å)	Residues /turn	Sugar pucker	Groove Width (Å)		Groove Depth (Å)	
						Minor	Major	Minor	Major
A-DNA	right	32.7	2.56	11	C3'-endo	11	2.7	2.8	13.5
B-DNA	right	36	3.4	10	C2'-endo	5.7	11.7	7.5	8.8
Z-DNA	left	-9,-51	3.8	12	C3'-endo (Syn)	-	8.8	3.7	3.7

double helix are separated by two different grooves minor groove and major groove. Specific recognition of DNA sequences by small molecules is achieved by the combination of hydrogen bond acceptor/donor sites available on the major groove or minor groove of DNA.

DNA is the pharmacological target of many anticancer drugs which are currently under clinical trials. Transcription and replication are the vital processes essential for the survival of the living system. In transcription, information is fetched from DNA to RNA and has recourse to synthesize protein in the body. In replication, DNA yield self-replication process and reconstruct two identical strands. DNA starts these processes only after receiving the signal which is usually in the form of regulatory protein to a specific region of DNA. If this regulatory protein is mimicked by a drug molecule (mainly heterocyclic aromatic molecule), then the functions of DNA can be artificially modulated, inhibited or activated by this small molecule to cure or control a disease. DNA involved in vital processes such as replication, transcription etc. are of particular interest as target for wide range of anticancer and antibiotic drugs (Chaires, 1998; Chaires, 1997; Chaires, 2008; Hurley, 2002).

Molecular interaction between drugs and DNA is a field of current research and also plays an important role in its biological activity. Many anticancer therapies depend on the interaction of drug molecule with DNA. These interactions may cause damage of DNA in cancerous cells by inhibiting the process of replication or transcription, which inhibits the growth of cancer cells. To design efficient chemotherapeutic agents and better anticancer drugs, it is important to inspect the interaction of drug with DNA. The number of known DNA-based drug targets is very limited in comparison to the protein based drug targets and also the number of available structures of DNA-drug complexes is also small relative to protein-drug complexes deposited in the PDB (Berman *et al.,* 2000).

Different Modes of DNA binding with Drug

There are many binding modes in which drug molecule can interact with DNA such as surface binding to their minor or major grooves, intercalation between adjacent base pairs, covalent attachments to the double helix, or electrostatic binding. Thus both covalent as well as non-covalent interactions were found between drug molecules and DNA. DNA interacting drug molecules are given in Table 1.

Covalent Binding

Many chemotherapeutic drug molecules which are in clinical use bind with DNA not only non-covalently but also by covalent binding. Covalent binding in DNA is irretrievable and regularly points to complete inhibition of DNA processes and subsequent cell death. Drug molecule covalently binds with DNA via inter- and intra-strand cross linking or alkylation. Covalent binders of DNA are the high binding strength (Paul and Bhattacharya, 2012; Liu and Sadler, 2011). The covalent binders are also known as alkylating agents because they can attach an alkyl group to DNA and are also used in the treatment of Cancer. Alkylating agents are the important class of anticancer drugs, they play crucial role in the cure of several types of cancers. Alkylating agents have methyl or other alkyl groups (C_nH_{2n+1}) onto molecules. Chemical structure of some important alkylating agents is shown in Figure 1. Alkylating agents are involved in reaction with the preferential N-7 position of guanine and N-3 of adenine in DNA. Thus the base pairing of the DNA could be inhibited and this leads to miscoding of DNA. Alkylating agents can interact to DNA via three mechanisms. In first mechanism an alkylating agent attaches alkyl group to the nucleic acid bases, this results in the DNA being fragmented by repair enzymes in their attempts to replace the alkylated bases. In second mechanism alkylating agent leads to DNA damage due to formation of cross-links and bonds between atoms in the DNA. In this process, two bases are linked together by alkylating agents that has two DNA-binding sites. Cross-linking prevents DNA from being separated for synthesis or transcription. In third type of mechanism, alkylating agents causes the mispairing of the nucleotides leading to mutations (Silvestri and Brodbelt, 2013; Kondo *et al.,* 2010). The nitrogen

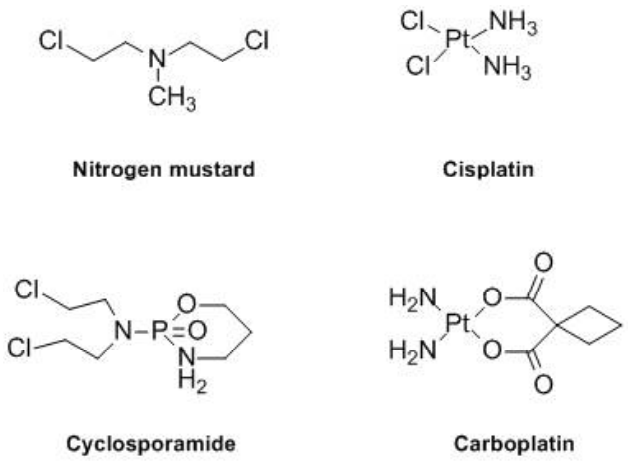


Fig. 1. Chemical structure of some DNA alkylating agents.

mustards were the first alkylating agent used medically, as well as the first modern cancer chemotherapies (Bauer and Povrik, 1997). Cis-platin is one of the anticancer antibiotics, which covalently binds to DNA, which makes an intra/ interstrand cross-link with nitrogens on the DNA bases and is used in the treatment of testicular, ovarian, head and neck cancers. Most alkylating drugs are mono functional methylating agents (e.g. temozolomide[TMZ], N-methyl-N'-nitro-N-nitrosoguanidine [MNNG], and dacarbazine), bifunctional alkylating agents such as nitrogen mustards (e.g. chlorambucil and cyclophosphamide), or chloroethylating agents (e.g. nimustine [ACNU], carmustine [BCNU], lomustine [CCNU], and fotemustine) (Kondo *et al.,* 2010; Rajski and Williams, 1998; Park and Hurley, 1997).

Non-covalent Binding

Non-covalent binding drug may change DNA torsional tension, interrupt protein-DNA interactions, and potentially lead to the breaking of DNA strands. All of these can have substantial effects on gene expression. Non-covalent interactions, in particular hydrogen bonding and stacking interactions, determine the structure of biomolecules (such as nucleic acids and proteins). While it is understood that hydrogen bonding is essential for the specificity of base pairing, ð-ð stacking interactions between planar aromatic rings of nucleobases are equally important contributions to the final stability of nucleic acid structures. Although individually weak, the additive power of these interactions has large cooperative stabilizing effects. Non-covalently binding of drug with DNA is mainly classified into two category viz. groove binders and intercalators. Groove binders are of two types: minor groove binders and major groove binders. Groove binders are highly sequence-specific. The two types of grooves in nucleic acid differ in hydrogen-bonding, electrostatic potential and in degree of hydration. Mainly the large protein molecules binds to the major groove of DNA while small molecules generally bind to the minor groove of DNA which are long elongated structures with a curvature that acquires the shape of the minor groove.

Minor Groove Binders

Minor groove binders usually consist of aromatic rings covalently linked by sigma bonds. Small molecules can form hydrogen bond to the nucleic bases, generally N3 of adenine and O2 of thymine. Minor groove binders generally bind with A-T rich region of the DNA. This preference in addition to the designed propensity for the electro negative pockets of AT sequences is probably due to better vander Waals contacts between the ligand and groove regions and also because of the steric hindrance in the latter, presented by the C2 amino group of the guanine base (Nelson *et al.,* 2007; Khan *et al.,* 2012; Privalov *et. al,* 2007). Sequence specific DNA-binding proteins commonly binds with the major groove because of numerous possibilities for hydrogen bonds with donors and acceptors on the nucleic bases, which provides complex stability and sequence specificity. Proteins and small molecule binding to the minor groove of DNA; depends upon the hydration properties of minor grooves, the latter binding to that AT-rich regions in which water ordering is most prevalent. Thus the minor groove binding is normally driven by the very large entropy of releasing the ordered water, despite an unfavorable enthalpy (Wemmer and Dervan, 1997; Sterkowski and Wilson, 2007; Gilbert and Feigon, 1991). Minor-groove binding usually involves greater binding affinity and higher sequence specificity than that of intercalator binding. Minor-groove has been demonstrated for neutral, mono-charged and multicharged ligands (Baily and Chaires, 1998). Generally, minor-groove binders show AT-rich region selectivity, several factors are responsible for this preference. The electrostatic potential of AT-rich region is greater than that of GC-rich region. On the other hand, the dimensions of the minor groove at AT sites are narrower and deeper than at the GC sites. This difference is due to the differences in the ionic interactions in the two types of base pairs (Hamelberg *et al.,* 2001). The cationic minor-groove binders include the lexitropsins and their conjugates, analogues of Hoechst 33258, DAPI and diarylamidines, Berenil, SN series, and pentamidines (Bhattacharya and Thomas, 2000; Erikson *et al.,* 1993; Brown *et al.* 1979).

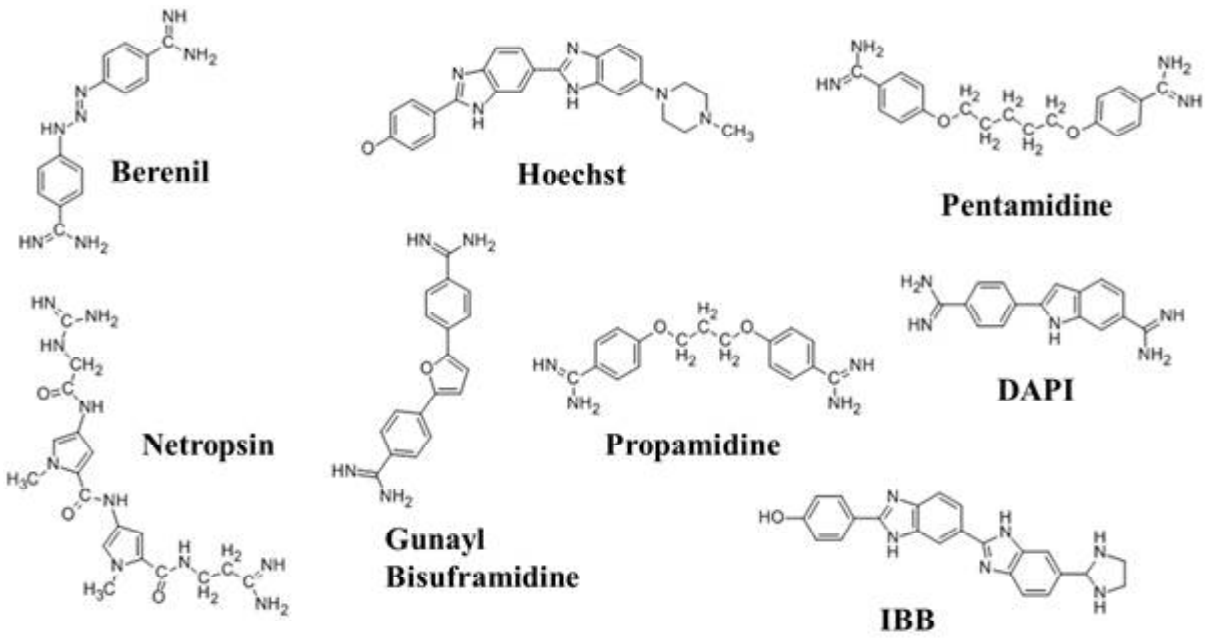


Fig. 2. Chemical Structure of some DNA minor groove binders.

Intercalators

Another type of non-covalently binding drugs is intercalators which are generally planar heterocyclic molecule which stacks between the two adjacent nucleic acid base pairs. The complex remains stabilized because of ð-ð stacking between the drug molecule and DNA bases. Intercalation was first explained by Lerman, in which the drug molecule held rigidly perpendicular to the DNA backbone without breaking up the hydrogen bonding between the nucleic bases (Lerman, 1963). This may be distorting the sugar phosphate backbone and also leads to decrease in the pitch (Williams *et al.,* 1990). The driving forces for the stability of DNA-intercalator complex are vanderWaals, hydrophobic, stacking or charge transfer forces and hydrogen bonding and electrostatic forces also become important for the stabilization of complexes (Wang, 1992). DNA intercalation results in conformational changes in DNA structure, causing lengthening, stiffing and unwinding of DNA helix. Intercalation needs changes in the torsional angles of sugar-phosphate backbone to adjust the incoming aromatic compound, which causes separation between the base pairs with a lengthening of the DNA approximately 3.4Å and decrease in helical twist, unwinding the DNA in the vicinity of the binding site to less than 36° base pair (Neto and Lapis, 2009). Intercalation preferentially occurs at GC-rich sequences because these sequences get unstacked easily. Intercalators generally cause more significant distortion to the conformation of DNA. Echinomycin, noglamycin, triostin A, acridine, cis-Platin, adriamycin, ethidium, propidium, actinomycn D, adriamycin are some examples of the DNA intercalating agents (Pigram *et al.,* 1992).

There are few major binding modes for reversible binding of molecule to the DNA: (i) electrostatic interactions with the anionic sugar phosphate backbone of DNA (ii) interaction with DNA minor groove (iii) interaction with DNA major groove (iv) intercalation between DNA base pairs via DNA minor groove (v) intercalation between DNA base pairs via DNA major groove and (vi) threading intercalation mode. After the intercalation of a structure, the access of another intercalator to binding site next to neighboring intercalation pocket is hindered. This phenomenon is referred as the "neighbor exclusion principle" and could be explained considering that due to intercalation the significant structural changes in DNA with deep alterations in the nucleotide secondary structure (Neto and Lapis, 2009; Yen *et al.,* 1982; Tanious *et al.,* 1991). Intercalating compounds without bulky substituents can intercalate without having significant part of molecule in either minor or major groove, this type of molecules (DACA, proflavin etc.) are called classical intercalators (Pigram *et al.,* 1972). Some of the intercalating molecules have bulky substituents, and these bulky substituents are placed in the major or minor groove along with intercalating moiety. These types

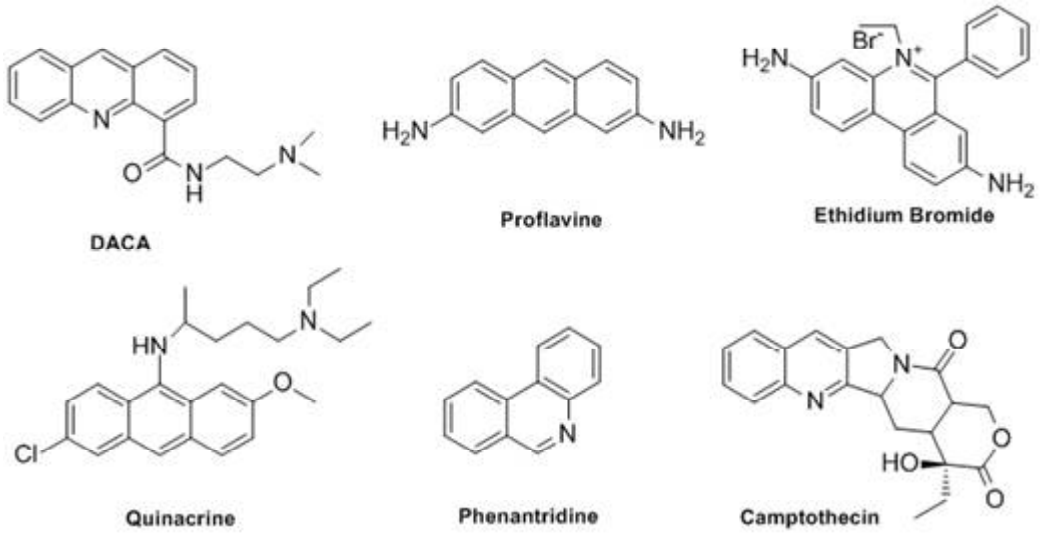


Fig. 3. Chemical Structure of some DNA intercalators.

of intercalators are called threading intercalators (Martinez and Chacon-Garica, 2005). In threading intercalation complexes, an aromatic system inserted between base pairs, while bulky substituent binds strongly with both major and minor groove (Neto and Lapis, 2009; Yen *et al.,* 1982). Other important intercalating drugs are the anthracyclin, daunorubicin, adriamycin, quinacrine and actinomycin have bulky substituents that must be in one groove or the other after the planar aromatic ring of the drugs is bound by intercalation (Tanious *et al.,* 1991; Rao and Kollman, 1987; Bond *et al.,* 1975; Wilson *et al.,* 1998; Martinez and Chacon-Garica, 2005; Denny, 2002; Nakamoto *et al.,* 2008; Wheate *et al.,* 2007; Baraldi *et al.,* 1999; Reddy *et al.,* 1999).

Major Groove Binders

The major groove is wider than the minor groove, the groove width values for B-form of DNA are 11.6 Å and 6.0 Å respectively. Due to this difference in dimension, the major groove is the target for many DNA-interacting proteins. Many biological macromolecules such as proteins interact by the variety of hydrogen bond acceptor and donor supplied in the major groove (Simonsson *et al.,* 1998; Singh and Lambowitz, 2001; Mamoon *et al.,* 2002). It is important for a major groove binding molecule that it could block access to proteins that recognize the same groove. This can be achieved by sequence affinity and sequence selectivity (Scheilf, 1988). DNA duplexes which are made up of polypurine–polypyrimidine sequences can be read by oligomers and bind to the major groove and form hydrogen bond with nucleic bases of the purine strand. These are called triplex-forming oligonucleotides (TFOs) (Thoung and Helene, 1993; Jain and Bhattacharya, 2010). Another form of major-groove recognition could be achieved by peptide nucleic acids (PNAs) (Neilsen, 1999; Ganesh and Kumar, 2005).

DNA interacting organometallic compounds

Many coordination complexes possess an intercalating ligand in their coordination sphere; the study of such complexes reveals the preferential geometry of the metal center, the nature of the intercalating ligand and the number and the position of the substituents over the intercalating ligand in the capacity and selectivity of the coordination complexes to intercalate with DNA. These coordination compounds bind to DNA via two interaction modes: irreversible (covalent or coordination binding) and reversible (intermolecular association). The later binding mode can be further classified into electrostatic interactions, groove binding and intercalation. However, these coordination complexes may exhibit a preference for a particular binding mode or a nucleotide sequence depending upon the size and the shape of the molecule (Han *et al.,* 2004; Hazarika *et al.,* 2012; Juan *et al.,* 2013; Rodrigo 2015).

All mononuclear platinum complexes could form intrastrand and interstrand adducts with DNA. When interstrand lesion is formed, massive distortions of the B-DNA are observed. Similarly, intrastrand lesion, while it forms much more readily than the interstrand lesion, it induces mutational events via the distortion of its nucleic acid target. Binuclear platinum (II) complexes were designed and synthesized and their interactions were studied with calf thymus DNA and a small 49 base pair

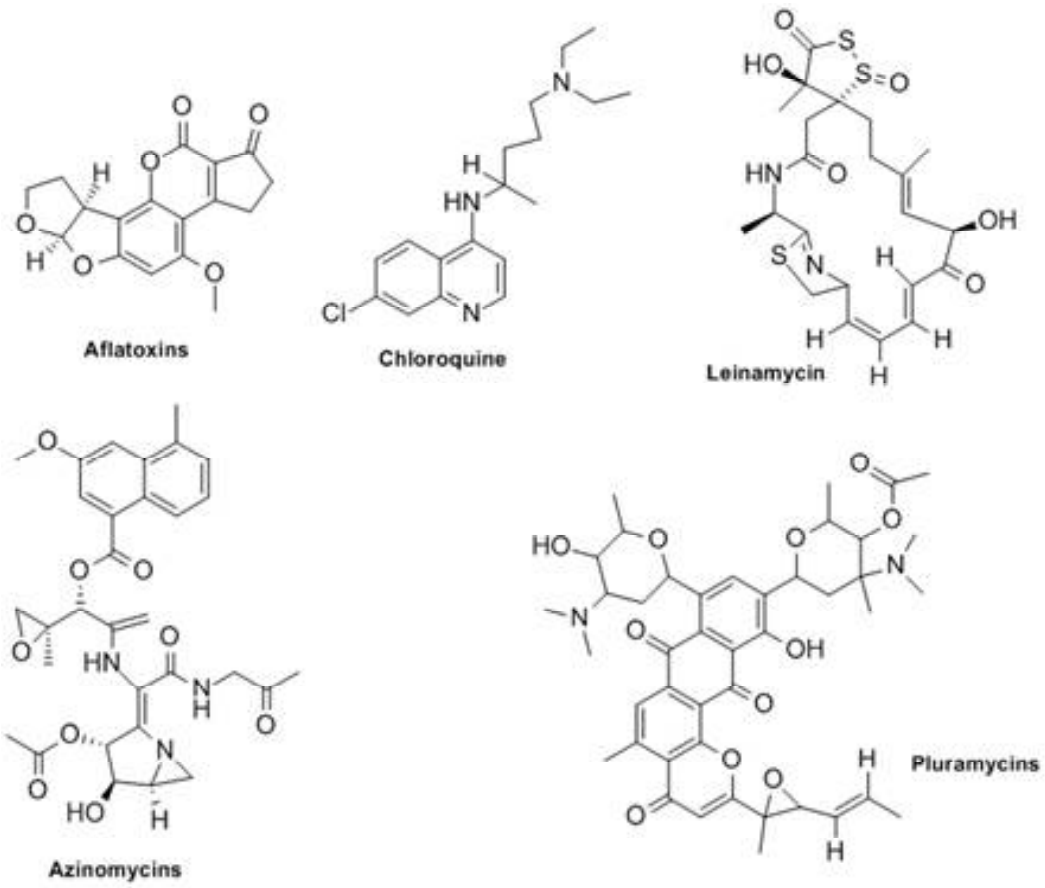


Fig. 4. Chemical Structure of some Major Groove Binders.

oligodeoxyribonucleotide. Owing to the presence of the pyridyl ligands, this compound induces a much higher degree of DNA unwinding than that seen with the either of the ammonia bound complexes, as well as the mononuclear trans-[$PtCl_2$-$(py)_2$]. Similarly, to the mononuclear compound [(Pt(trans)-$(py)_2Cl_2$-µ-(diaminobutane)]$^{2+}$. These alterations likely involve [(Pt(trans)-$(py)_2Cl_2$-µ-(diaminobutane)]$^{2+}$ to undergo ð-stacking interactions upon DNA association which in turn, disfavors the Z-DNA conformations. Importantly, interstrand-cross links formation is very efficient for all three complexes. The directionality is dependent upon the nature of the cross link. Interestingly, this is a unique example of anti-cancer drugs behaving in this manner. Molecules normally reach DNA through one of the grooves and react to either the backbone or the nucleobases. Electrostatic binding occurs due to the interaction between cations with the negatively charged phosphate backbone at the exterior surface of the DNA helix (Konstantinos *et al.,* 2013; Decatris *et al.,* 2004).

The use of transition metal complexes gives a strong tool to the drug chemists to develop and study molecules capable of obtaining specific DNA-drug interactions considering the multiple options of d-block metals from the periodic table. Transition metals are dynamic in geometry, electron affinity and reactivity, making them excellent choices to feed the ongoing field of antineoplastic drug discovery.

The fundamental factors of these interactions still possess greatest gaps in as much as the results provided by experimental designs that do not involve expensive protocols and equipment that are extremely poor to identify the specific interactions due the lower energetic changes involved, making clear the use of methodologies such as computational chemistry to help solve these problems (Ge and Sun, 2007; Wang and Lippard, 2005; Umezawa, 1976; Wong and Giandomenico, 1999; Bonnet, 1995).

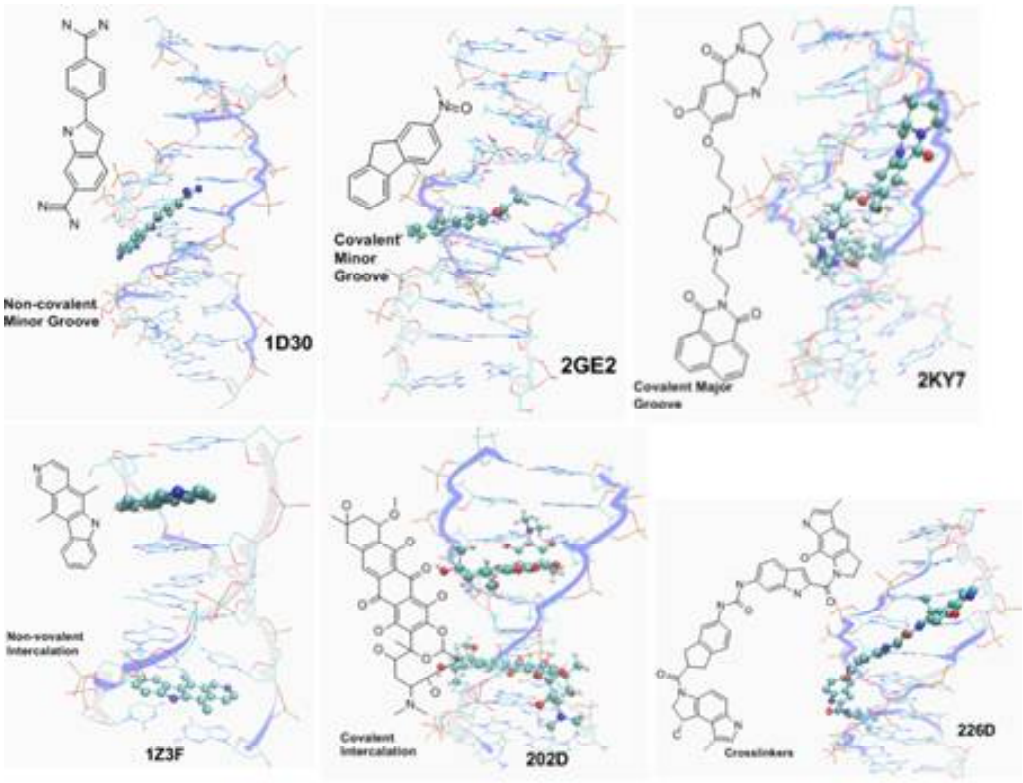


Fig. 5. Different modes of Drug-DNA Binding.

Experimental studies used in drug-DNA interaction

Experimental Studies play crucial role to explore the drug-DNA interaction. Thermodynamic studies provide the necessary information of free energy, enthalpy, entropy, heat capacity and also about the binding constant changes during complex formation. Various experimental techniques which are used to understand the interaction of drug molecule with nucleic acid holds Infrared (IR), Raman, Circular dichorism, UV-visible, Nuclear magnetic resonance (NMR) spectropies, Atomic Force Microscopy (AFM), Electrophoresis, Mass spectrometry, Viscosity measurements, Thermal denutration studies, Cyclic square wave and Differential pulse voltammetry, etc. The above techniques have been used as a primary tool to characterize the behavior of drug-DNA binding and the consequences of such type interaction on the structure of nucleic acid. The commonly used experimental techniques are UV-visible, fluorescence spectroscopies and cyclic voltammetry (Sirajuddin *et al,* 2013). Chaires also provided the change in experimental values of thermodynamic properties during the drug-DNA complex formation (Garbett and Chaires, 2012). This information may be helpful for the theoretical prediction and structural analysis. In the binding of drug molecule with bimolecular system, the solvent water plays important role. In the case of DNA-focused drug approaches there is a need to understand how water take part in the reorganization (CheathamIII *et al.,* 1995; Bellissent-Funel *et al.,* 2016; Yu, 2008; Chalikian and Breslauer, 1998).

Molecular Modelling Studies for drug-DNA interaction

Molecular Docking

Molecular docking method is used to predict the structure (or structures) of the intermolecular complex formed between two or more molecules. The docking program generates large number of possible structures, and so it is required to rank them according to their score to identify those of most interest (Blaney and Dixon, 1995; Abagyan and Totrov, 2001; Kuntz, 1992; Lengauer and Rarey, 1996). The three important components of docking are:

(1) Representation of the system.

(2) Conformational space search via a search algorithm.

(3) Ranking of potential solutions using the scoring function.

Table 2. DNA interacting drug molecules.

Non-covalent Binding drug molecules			Covalent binding drug molecules
Groove Binding		Intercalators	
Minor groove Binders	Major Groove Binders		
Berenil	Chloroquine	Daunomycin	Nitrogen mustard
Netropsin	Netamycine	Nogalamycin	PBDs
Hoechst 33258	Cis-{$Pt(NH_3)_2$(pyridine)}$_2$+	Ethidium bromide	CC1065
Distamycin A	Aminoglycoside (NB33)	Proflavine	Cis-platinum
GunaylBisuframidine	Chlorambucil	Ellipticine	Menogril
SN6999	Nimustein	Diplamine	Clomesone
SN7176	Pluramycins	Chlorpheniramine	Cyclodisone
Pentamidine	Aflatoxins	Bis-napthalimide	
MithramycinPilcamycin	Azinomycins	Doxorubcin	
Chromomycin A3	Leinamycin	Aminoacridines	
Diamidine-2-phenylindole	Ditercalinium	Arylaminoalcohals	
Bisbenzimadoles		Coumarines	
Bleomycin		Cystodytin	
Mitomycin		Diplamine	
FR66979		YO and YOYO-1	
Duocarmycins		QuinolinesQuinoxalines	
CC-1065		Echnomycin	
Yatakemycin		Methapyrilene	
Neocarzinostatin		Tamoxifen	
Calicheamicins		M-AMSA	
Retrorsine		Indoles	
Anthramycins		Aclarubicin	
Saframycins		Idarubicin	
Ecteinascidin 743		Epirubcin	
Isochrysohermidin		Pirarubcin	
		Valrubicin	
		Amrubicin	
		Actinomycin D	
		Camptothecin	
		Topotecan	
		Irinotecan	
		Rebeccamycin	
		Podophyllotoxin	
		Etoposide	
		Teniposide	
		Elsamicin	
		Dynemicin	
		Triostin A	
		Luzopeptins	
		Sandramycin	

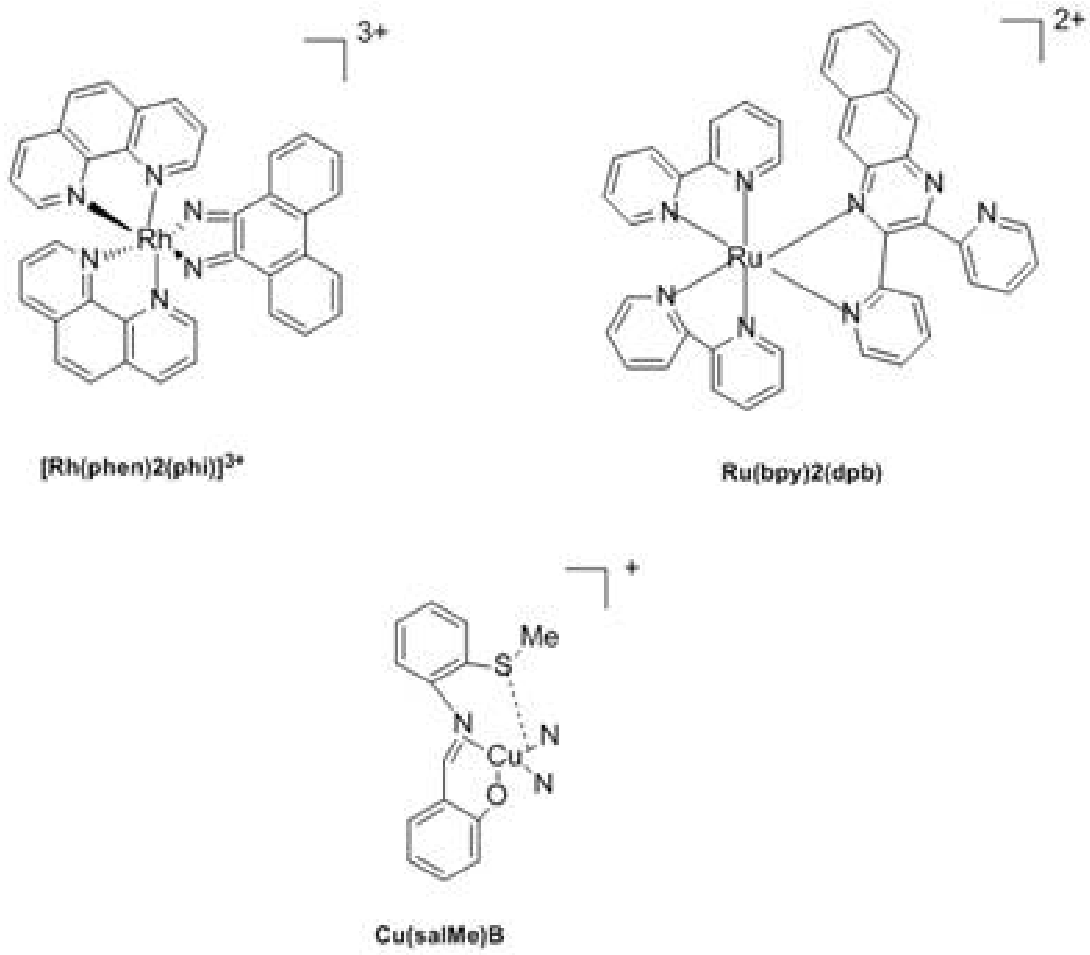


Fig. 6. Chemical Structure of some organometallic compounds.

The aim of docking process is to computationally simulate the molecular identification process and accomplish an optimized conformation so that the free energy of the overall complex is minimized. The docking method involves many degrees of freedom. With six degrees of translational and rotational freedom along with the conformational degrees of freedom of each molecule results large number of possible binding modes between the two molecules. Computationally it would be too expensive to generate all the possible conformations. Various algorithms have been developed to tackle the docking problem. These algorithms can be characterized according to the number of degrees of freedom that they ignore. The search algorithm generates number of conformations for a particular ligand, and scoring functions are then applied in order to identify the energetically most favorable pose (Gane and Dean, 2000; Schneider and Bohm, 2002; Sthal and Rarey, 2001; Koremer, 2003; Gohlke and Klebe, 2002).

Surflex (Surflex, 2007), Autodock (Morris *et al.,* 1998), DOCK (Rarey*et al.,* 1996), GOLD (Jones *et al.,* 1997), Glide, Flex (Rarey *et al.,* 1996), CDOCKER (Wu *et al.,* 2003) are the docking programs which are used for the Molecular Docking of drug and DNA. A study shows that the GOLD and GLIDE docking protocols seem to be very reliable in modelling of nucleic acid ligand complexes (Srivastava *et al.,* 2011). Molecular docking studies show that the intercalators generally bind to the CG –rich region of DNA and minor groove binders to the AT-rich region of DNA (Srivastava *et al.,* 2011; Rashidaa and Ahsen, 2015). ð-ð interaction dominates in case of intercalators. In case of minor groove binders, hydrogen bonds are mainly formed between minor groove binders and the functional groups on the bases are exposed in the grooves via their end groups and also their amide or other linker groups. Other studies shows that if the nature of ligand with DNA is not known, the exact mode of binding of ligand to DNA cannot be predicted on the

basis of molecular docking as a result other molecular modelling techniques such as molecular dynamics simulation and thermodynamics integration will be required to further resolve the problem (Mariya and Ahsen, 2015).

Molecular Dynamics Simulation

Molecular Dynamics (MD) is the most important computational approach for the study of flexible nucleic acids. In MD, the motion of a biomolecular system under the effect of a "force" (i.e a specified force field) is simulated by following its molecular configurations in time, according to Newton's equation of motion (second law). A MD calculation starts with a set of initial co-ordinates and velocities. The force-field calculates the potential energy and the forces acting on the system, and Newton's second law is used to determine the accelerations on each particle. Numerical integration of these accelerations provides a set of new velocities and positions, which are used to build up a trajectory. MD protocols include algorithms to fix the temperature and the pressure, allowing the simulation of nucleic acids under conditions close the physiological ones. MD simulations of nucleic acid are performed using explicit solvent representations including thousands of water molecules and periodic boundary conditions (PBC). Explicit water models used in bio molecular simulations include TIP3P, TIP4P, SPC, extended SPC/E, and F3C models among these TIP3P is the most commonly used model (Jorgensen *et al.,* 1983; Berendsen *et al.,* 1987; Levitt *et al.,* 1997). Ions (generally Na^+ and Cl^-) are introduced to neutralize the system and simulate the given ionic strength. The evolution of trajectories shows the movement from one stable state to a stable one.

In mid 1990s several groups performed successful MD simulations of DNA and RNA with an explicit representation of solvent using the AMBER, CHARMM nucleic acid, or GROMOS force field. Various force fields used in the nucleic acid simulation includes, CHARMM, AMBER, GROMOS, OPLS, ENCAD and BMS26. AMBER (Case *et al.,* 2012), GROMACS (Hess *et al.,* 2008), CHARMM (Brooks *et al.,* 2009) and NAMD (Nelson *et al.,* 1996) are the popular software packages for the simulation of nucleic acid ligand complexes. The three latest force-fields (AMBER-99, CHARMM-27 and BMS) provide accurate representations of standard DNA and RNA structures (York *et al.,* 1995; Cheatham *et al.,* 1995; Weerasinghe *et al.,* 1995; Weiner and Kollman, 1981; Nilsson and Karplus, 1986; Gunsteren and Berendsen, 1986). The root-mean square deviation (RMSD) of the simulated nucleic acids with respect to experimental structures is small, the dihedral distributions are correct, and the average helical parameters are also reasonably close to the accepted experimental values. Among the variety of available force fields, CHARMM and AMBER are the most popular force fields (Cheatham and Young, 2001; Cheatham *et al.,* 1999; Cornell *et al.,* 1995; Foloppe and Mackerell, 2000; Mackerell and Banavali, 2000; Langley, 1998; Modesto *et al.,* 2003). In a number of studies for the minor groove binders and intercalators, computational methods have shown good agreement with the experimental results (Kamal *et al.,* 2007, 2009, 2010a, 2010b).

MMPBSA/MMGBSA Method

The molecular mechanics energies combined with the Poisson-Boltzmann or generalized Born and surface area continuum solvation (MMPBSA/MMGBSA) methods are popular approaches to calculate the free energy difference between two states, generally the bound and unbound state of two solvated molecules, or ultimately to compare free energy of two different solvated conformations of the same molecule (Gohlke and Klebe, 2002; Kolman *et al.,* 2000; Srinivasan *et al.,* 1998; Hou *et al.,* 2011; Homeyer and Gohlke, 2012). Snapshots obtained from MD simulation are used for the calculation, yielding an average of the energies. The free energy of binding is calculated by the equations mentioned below-

$$\Delta G_{bind} = \Delta H - T\Delta S$$

$$\Delta G_{bind} = (\Delta E_{MM} + \Delta G_{SOL}) - T\Delta S$$

Where,

$$\Delta E_{MM} = (E_{MM}{}^{complex} - E_{MM}{}^{receptor} - E_{MM}{}^{ligand})$$

$$\Delta G_{SOL} = (\Delta G_{SOL}{}^{complex} - \Delta G_{SOL}{}^{receptor} - \Delta G_{SOl}{}^{ligand})$$

$$\Delta S = (S^{complex} - S^{receptor} - S^{ligand})$$

where ΔH is the enthalpic contribution to binding energy, ΔE_{MM} is the average difference in molecular mechanics energy, while ΔG_{SOL} term accounts for the solvation free energy(including both polar and non-polar component); T is the temperature and ΔS is a change in entropy.

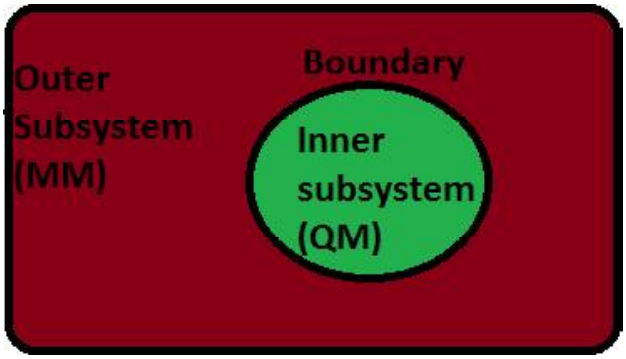
Thus the net binding free energy of complex system is equal to the sum of an intermolecular energy (calculated using MM force field), a solvation free energy term and an

entropic term. Polar solvation free energies is calculated either solving the linear Poisson Boltzmann (PB) equation or more approximate, or computationally effective Generalized Born (GB) model, whereas the non-polar contribution is estimated from a linear relation to the solvent accessible surface area (SASA). Entropy contributions to the free energy are estimated either by quasi-harmonic analysis or by using normal mode analysis (Schwarzl *et al.,* 2002; Rastelli *et al.,* 2010).

The MMPBSA calculations are based on single minimized structures, rather than on a large number of MD snapshots. This method will save much computational time but ignores dynamical effects. It has been observed that the more time can be saved by performing the minimizations in a GB continuum model. Hou *et al.* reveals that the MMGBSA results varied with the length of simulation, but there is no gain of using simulation longer than 4ns (Hou *et al.,* 2011). H.K. Srivastava *et. al.* proves that the MM-PBSA based interaction energies calculated from the MD Simulations are in good agreement with the experimental values for the DNA-ligand complexes (Srivastava *et al.,* 2011; Spackova *et al.,* 2003). The MMPBSA approach was originally developed for the AMBER software but recently some automatic scripts have also been presented for the freely available GROMACS, NAMD and APBS software. A study reveals that the energies calculated using g_mmpbsa (GROMACS) and the AMBER MM-PBSA package is approximately similar and the difference of 1-3 kcal/mol has been observed due to the difference in ΔG_{polar} (Kumari *et al.,* 2014). The difference in ΔG_{polar} is observed because of different algorithms, implemented in APBS and PBSA (Mishra *et al.,* 2015).

Quantum Mechanics/Molecular Mechanics (QM/MM) method

The QM/MM concept was introduced, as early as 1976, by Warshel and Levitt, who presented a semi empirical QM/ MM treatment for a chemical reaction in lysozyme (Warshel and Levitt, 1976). Combined QM/MM theory has emerged as the method of choice for modeling local electronic events in large bimolecular systems.The basic idea is to describe the active site (where chemical reactions or electronic excitations occur) by quantum mechanics, as accurately as needed, while capturing the effects of the bimolecular environment by molecular mechanics, i.e., at the classical force field level. The accuracy of QM and speed of MM, combined QM/MM methods enable the modelling of reactive bimolecular systems at reasonable computational cost with the necessary accuracy. Due to the potential uses of this method, this field gained the Nobel Prize in chemistry in 2013. There are two schemes to calculate the total energy of the system the additive and the subtractive scheme (Sherwood *et al.,* 2008, Senn and Theil, 2009, Sherwood *et al.,* 2003). Regarding this boundary schemes, the labeling conventions given in **Figure** that apply to covalent bonds across the QM-MM boundary.



Subtractive schemes:

$$E_{QM/MM}(system) = E_{MM}(system) + E_{QM}(QM) - E_{MM}(QM)$$

Where $E_{QM/MM}$(system) is the total energy, E_{MM}(system) is the MM energy of the entire system, E_{QM}(QM) the QM energy of the QM region and E_{MM}(QM) the energy of the QM region.

The scheme encounters shortcomings due to the treatment of interactions between QM and MM region only at MM level which is inaccurate. This scheme needs the MM parameters for the QM region. Parameters are not usually available for these systems which are present in excited electronic states or contain transition metals.

Additive schemes:

$$E_{QM/MM}(system) = E_{MM}(system) + E_{QM}(QM) - E_{QM-MM}(QM, MM)$$

In this scheme, the total energy of the system $E_{QM/MM}$ (system) comprises of only three components viz., E_{MM} (MM) theMM energy of the MM region only, E_{QM} (QM) the QM energy of the QM region andthe $E_{QM^{\circ}MM}$(QM, MM) a term which interfaces between the QM/MM through the inclusion of bonded and non-bonded interactions. The bonded interactions account for bond stretching, bending and torsion while the non-bonded account for the vander Waals and electrostatic interactions.

The key to such QM/MM methods is the coupling between the electric field from the surrounding and the QM Hamiltonian in the active-site region. This requires careful treatment of the boundary between the QM and MM

regions.

The most important part of QM/MM is partitioning of the system. The basic considerations for QM/MM partitioning are:

(a) The choice of the QM region is usually made by consideration of the chemical problem; chemical arguments normally suggest a minimum-size QM region which can then be enlarged to check the sensitivity of the QM/MM results with regard to such an extension.

(b) If a QM/MM division through covalent bonds cannot be avoided, cut only unconjugated single bonds, preferably without electronically demanding substituent's (e.g., cut unpolar C–C bonds).

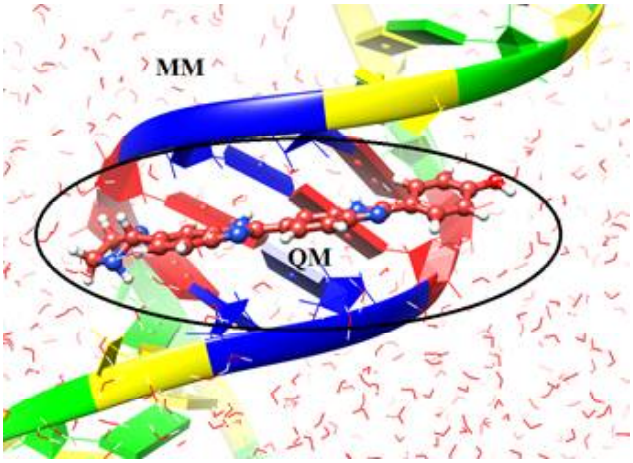


Fig. 7. QM/MM partitioning: *drug* and associated bases (QM region), remaining bases and water molecules and ions (MM region).

A common procedure for QM/MM calculations is to take a crystal structure from protein data bank as a starting point for the calculations and add hydrogen, missing atoms etc. and water molecules (soak the complete system). Thereafter, the system is equilibrated using MM followed by molecular dynamics production run and low-energy snap-shots are studied. These snapshots will eventually be taken for QM/MM calculations.These structures contain the bimolecular system in a droplet of water (20000-30000 atoms) and this setup requires a lot of prior work to avoid errors and wrong choices for the actual QM/MM calculations. Further study of the reaction mechanism is similar to the typical methods for the study of gas-phase reaction (Senn and Theil, 2007, Frieesner and Guallar, 2005).

The computational chemistry techniques, especially the hybrid methods (ONIOM) based on combination of several theoretical approaches, have been developed by Morokuma and co-workers for large biomolecular systems (Svensson *et al.,* 1996, Morokuma *et al.,* 2001, Morokuma, 2003, Kuno *et al.,* 2003). In many studies, the ONIOM method is used for the study of DNA binding drugs (Rebeca *et al.,* 2009, Ahmadia *et al.,* 2011, Robertazzi and Platts, 2006).

CONCLUSION

In this review the different types of small organic molecules which targeted DNA have been discussed. The array of available computational approaches and molecular modelling methods are being used for complementing the experimental efforts to improve the existing drugs and also in designing novel drug candidates which can act as good DNA inhibitor. Advances in computational resources over the last years have made screening of large chemical libraries and application of molecular dynamics and quantum chemical calculations feasible. This review seeks to highlight some recent molecular modelling studies performed at electronic structure level to study the mechanism of drug interaction to DNA which can give an insight in designing inhibitors for the treatment of cancer and AIDS.

ACKNOWLEDGEMENTS

AK like to acknowledge the UGC for financial support. The authors are thankful to Dr G. Narahari Sastry for his support and discussions.

REFERENCES

Abagyan R, Totrov M (2001). High-throughput docking for lead generation. Curr. Opin. Chem. Biol., 5: 375-382.

Ahmadia, F, Jamalia N, Jahangard-Yektaa S, Jafari B, Nourib B, Najafic F, Rahimi-Nasrabadid M (2011). The experimental and theoretical QM/MM study of interaction of chloridazon herbicide with ds-DNA, Spectrochimica Acta Part A, 79: 1004– 1012.

Avery OT, Maclend C, Mc Carty M (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. The Journal of Experimental Medicine, 79: 137-158.

BaillyC ,Chaires JB (1998). Sequence-specific DNA minor groove binders. Design and synthesis of netropsin and Distamycin analogs. Bioconjugate Chem., 9: 513–538.

Baraldi PG, Cacciari B, Guiotto A, Romagnoli R, Zaid AN, Spalluto G (1999). DNA minor-groove binders: results and design of new antitumor agents. IL Farmaco, 54: 15-25.

Bauer GB, Povirk LF (1997). Specificity and kinetics of interstrand and intrastrandbifunctional alkylation by nitrogen mustards

at a G-G-C sequence. Nucleic Acids Research, 25: 1211-1218.

Bellissent-Funel M, Hassanali A, Havenith M, Henchman R, Pohl P, Sterpone F, Spoel D, Xu Y, Garcia AE (2016). Water Determines the Structure and Dynamics of Proteins. Chem. Rev., 116: 7673-7697.

Berendsen HJC, Grigera JR, Straatsma TP (1987). The missing term in effective pair potentials J PhysChem, 91: 6269-6271.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhatt TN, Weissig H, Shindyalov IN, Bourne PE (2000). The protein Data bank. Nucleic Acid Research, 28: 235-242.

Bhattacharya S, Thomas M (2000). Facile synthesis of oligopeptidedistamycin analogs devoid of hydrogen-bond donors or acceptors at the N-terminus: sequence-specific duplex DNA binding as a function of peptide chain length. Tetrahedron Lett., 41: 5571–5575.

Blaney JM, Dixon JS (1993). A good ligand is hard to find: Automated docking methods Perspectives in Drug Discovery and design, 1: 301-319.

Bond PJ, Langridge R, Jennette KW, Lippard SJ (1975). X-ray fiber diffraction evidence for neighbor exclusiona binding of a platinum metallointercalation reagent to DNA. Proc. Natl. Acad. Sci., 72: 4825-4829.

Bonnett R (1995). Photosensitizers of the porphyrin and phthalocyanine series for photodynamic therapy. Chem. Soc. Rev., 24: 19–33.

Brooks BR, Brooks CL 3rd, Mackerell AD Jr, et al. (2009). CHARMM: the biomolecular simulation program. J Comput Chem., 30: 1545–1614.

Brown DG, Sanderson1 MR, Skelly JV, Terence CJ, Brown2 T, Garman3 E, Stuart3 DI, Neidle1 S (1990). Crystal structure of a berenil–dodecanucleotide complex: the role of water in sequence-specific ligand binding. EMBO J., 9: 1329–1334.

Case DA, Berryman JT, Betz RM, Cerutti DS, Cheatham III TE, Darden TA, Duke RE, Giese TJ, Gohlke H, Goetz AW, Homeyer N, Izadi S, Janowski P, Kaus J, Kovalenko A, Lee TS, LeGrand S, Li P, Luchko T, Luo R, Madej B, Merz KM, Monard G, Needham P, Nguyen H, Nguyen HT, Omelyan I, Onufriev A, Roe DR, Roitberg A, Salomon-Ferrer R, Simmerling CL, Smith W, Swails J, Walker RC, Wang J, Wolf RM, Wu X, York DM, Kollman PA (2015). AMBER 2015, University of California, San Francisco.

Chaires JB (1997). Energetics of drug-DNA interactions. Biopoly., 44: 201-215.

Chaires JB (1998). Drug—DNA interactions. Curr. Opin. Struc. Biol., 8: 314-320.

Chaires JB (2008). Calorimetry and thermodynamics in drug design. Annu. Rev. Biophys., 37:135-151.

Chalikian TV, Breslauer KJ (1998). Thermodynamic analysis of biomolecules: a volumetric approach. Current Opinion in Structural Biology, 8: 657-664.

Chargaff E (1950). Chemical Specificity of Nucleic Acids and Mechanism of their Enzymatic Degradation. Experimentia, 6: 201-209.

Chargaff E (1951). Some recent studies on the composition and structure of nucleic acids. J Cell Physiol. Suppl., 38: 41-59.

Cheatham III TE, Miller JL, Fox T, Darden TA, Kollman, PA(1995). Molecular Dynamics Simulations on Solvated Biomolecular Systems: The Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins. J Am ChemSoc, 117: 4193-4194.

Cheatham TE ,Cieplak P, Kollman PA (1999). A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. J. Biomol. Struct. Dyn., 16: 845-862.

Cheatham TE, Young MA (2001). Molecular Dynamics Simulation of Nucleic Acids: Sucesses, Limitations, and Promise. Biopolymers, 5: 232-256.

Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Fergurson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. J. Am. Chem. Soc., 117: 5179-5197.

Decatris, MP, Sundar S, O'Byrne KJ (2004). Platinum-based chemotherapy in metastatic breast cancer: current status. Cancer Treat. Rev., 30: 53–81.

Denny WA (2002). Acridine Derivatives as Chemotherapeutic Agents. Current Medicinal Chemistry, 9: 1655-1665.

Eriksson S, Kim SK, Kubista M, Norden B (1993). Binding of 42 6-diamino-2-phenylindole (DAPI).to AT regions of DNA: evidence for an allosteric conformational change. Biochemistry, 32: 2987–2998.

Foloppe N, Mackerell AD (2000). All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. J. Comput. Chem., 21: 86-104.

Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004). Glide: A New approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J. Med. Chem. 2004, 47: 1739-1749.

Friesner RA, Guallar V (2005). Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics methods for studying enzymatic catalysis. Annu. Rev. Phys. Chem. 56: 389-427.

Gane PJ, Dean PM (2000). Recent advances in structure-based rational drug design.Curr. Opin. Struct. Biol., 10: 401-404.

Ganesh KN, Kumar VA (2005). Conformationally constrained PNA analogs: Structural evolution towards DNA/RNA binding selectivity. Acc. Chem. Res., 38: 404–412.

Garbett NC, Chaires JB (2012). Thermodynamic studies for drug design and screening. Expert opin. Drug Discov., 7: 299-314.

Gareth J, Peter W, Robert CG, Andrew RL, Robin T (1997). Development and Validation of a Genetic Algorithm for Flexible Docking. J. Mol. Biol., 267: 727-748.

Ge R, Sun H (2007). Bioinorganic chemistry of bismuth and antimony: target sites of metallodrugs. Acc. Chem. Res., 40: 267-274.

Gilbert DE, Feigon J (1991). Structural analysis of drug-DNA interactions. Current Opinion in Structural Biology, 1: 439-445.

Gohlke H, Klebe, G (2002). Approaches to the Description and Prediction of the Binding Affinity of Small-Molecule Ligands to Macromolecular Receptors. Angew. Chem. Int. Ed., 41: 2644-2676.

Gunsteren van, WF, Berendsen HJC (1986). GROMOS 86: Groningen Molecular Simulation Program Package; University of Groningen: Groningen, The Netherlands.

Hamelberg D, Williams LD, Wilson WD (2001). Influence of the dynamic positions of cations on the structure of the DNA minor groove: sequence-dependent effects. J. Am. Chem. Soc., 123: 7745-7755.

Han D, Wang H, Ren N (2004). Molecular modeling of B-DNA site recognition by Ru intercalators: molecular shape selection. J Mol Model, 10: 216-222.

Hazarika P, Bezbaruah B, Deka RP, Deka J, Barman TK, Medhi OK, Medhi C (2012). The DNA binding features of ruthenium complexes compared with cisplatin : docking, force field and QM/MM studies. The Clarion, 1: 24-32.

Hess B, Kutzner C, van der Spoel D, Lindahl E (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput., 4: 435-447.

Homeyer N, Gohlke H (2012). Free energy calculations by the molecular mechanics Poisson-Boltzmann surface area method. Mol Inf, 31:114-122.

Hou T, Wang J, Li YY, Wang W (2011). Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. J Comp Chem, 32: 866-877.

Hurley LH (2002). DNA and its associated processes as targets for cancer therapy. Nat. Rev. Cancer, 2: 188-200.

Jain AK, Bhattacharya S (2010). Groove binding ligands for the interaction with parallel-stranded ps-duplex DNA and triplex DNA. Bioconjugate Chem., 21: 1389-1403.

Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997). Development and validation of a genetic algorithm for flexible docking. J. Mol. Biol., 267: 727-748.

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983). Comparison of simple potential functions for simulating liquid water. J ChemPhys , 79: 926-935.

Juan CG, Rodrigo G, Fernando C, Lena R (2013). Metal-Based Drug-DNA Interactions. J. Mex. Chem. Soc., 57: 245-259.

Kamal A, Rajender, Reddy DR, Reddy MK, Balakishan G, Shaik TB, Chourasia M, Sastry GN, (2009).Remarkable enhancement in the DNA-binding ability of C2-fluoro substituted pyrrolo[2,1-c][1,4]benzodiazepines and their anticancer potential. Bioorg. Med. Chem., 17: 1557-1572.

Kamal A, Khan MNA, Reddy KS, Rohini K, Sastry GN, Sateesh B, Sridhar B (2007). Synthesis, Structure Analysis and Antibacterial Activity of Some Novel 10-Substituted 2-(4-Piperidyl/Phenyl) -5, 5-Dioxo [1,2,4] triazolo [1,5b] [1,2,4] Benzothiadiazine Derivatives. Bioorganic and Medicinal Chemistry Letters, 17: 5400-5405.

Kamal A, Reddy KS, Khan MNA, Shetti RVCRNC, Ramaiah MJ, Pushpavalli SNCVL, Srinivas C, Pal-Bhadra M, Chourasia M, Sastry GN, Juvekar A, Zingde S, Barkume M (2010). Synthesis, Synthesis, DNA-binding ability and anticancer activity of benzothiazole/benzoxazole–pyrrolo[2,1-c][1,4]benzodiazepine conjugates. Bioorg. Med. Chem., 18: 4747-4761.

Kamal A, Shankaraiah N, Reddy C R, Prabhakar S, Markandeya N, Srivastava HK, Sastry GN (2010). Synthesis of bis-1,2,3-triazolo-bridged unsymmetrical pyrrolobenzodiazepinetrimers via 'click' chemistry and their DNA-binding studies. Tetrahedron, 66: 5498-5506.

Khan GS, Shah A, Zia-ur-Rehman, Barker D(2012). Chemistry of DNA minor groove binding agents. Journal of photochemistry and photobiology B: Biology, 115: 105-118.

Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Lillian C, Matthew L, Taisung L, Yong D, Wei W, Oreola D, Piotr C, Jaysharee S, Case DA, CheathamIII TE (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. AccChem Res, 33: 889-97.

Kondo N, Takahashi A, Ono K, Ohnishi T (2010). DNA Damage Induced by Alkylating Agents and Repair Pathways. Journal of Nucleic Acids, 54351: 1-7.

Konstantinos G, Shaun TM, James AP (2013). QM/MM description of platinum–DNA interactions: comparison of binding and DNA distortion of five drugs. RSC Adv., 3: 4066-4073.

Koremer RT (2003). Molecular modelling probes: docking and scoring. Biochemical Society Transactions, 31: 980-984.

Kumari R, Kumar R, Lynn A (2014). G_mmpbsa - a GROMACS tool for high throughput MM-PBSA calculations. J ChemInf Model, 54: 1951-1962.

Kuntz I (1992). Structure-based strategies for drug design and discovery. Science, 257: 1078-1082.

Langley DR (1998). Molecular dynamic simulations of environment and sequence dependent DNA conformations: the development of the BMS nucleic acid force field and comparison with experimental results. J. Biomol. Struct. Dyn., 16: 487-509.

Lengauer T, Rarey M (1996). Computational methods for biomoleculardocking.Curr. Opin. Struct. Biol., 6: 402-406.

Lerman LS (1963).The structure of the DNA-acridinecomplex. Biochemistry, 49: 95-101.

Levitt M, Hirshberg M, Sharon R, Laidig KE, Daggett V (1997). Calibration and Testing of a Water Model for Simulation of the Molecular Dynamics of Proteins and Nucleic Acids in Solution. J PhysChem B, 101, 5051-5061.

Liu H, Sadler PJ (2011). Metal complexes as DNA intercalators. Accounts of Chemical Research, 44: 349-359.

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

MacKerell AD, Jr., Banavali NK (2000). All-atom empirical force field for nucleic acids: 2). Application to solution MD simulations of DNA. J. Comp. Chem. 21: 105–120.

Mamoon NM, Song Y, Wellman SE (2002). Histone h1(0).and its carboxyl-terminal domain bind in the major groove of DNA. Biochemistry, 41: 9222-9228.

Mariya al-Rashidaa, Ahsen, S (2015). In search of a docking protocol to distinguish between DNA intercalators and groove binders: Genetic algorithm Vs shape-complementarity based docking methods. RSC Advances, 1-27.

Martinez R, Chacon-Garcia L (2005). The Search of DNA-Intercalators as Antitumoral Drugs: What it worked and what did not Work. Current Medicinal Chemistry, 12: 127-151.

Mishra R, Gaur AS, Chandra R, Kumar D (2015). Molecular Docking and Molecular Dynamics Study of DNA Minor Groove Binders. International Journal of Pharmaceutical Chemistry and Analysis, 2: 161-169.

Modesto O, Alberto P, Agnes N, Luque FJ (2003). Theoretical methods for the simulation of nucleic acids. Chem. Soc. Rev, 32: 350–336.

Morokuma K (2003). ONIOM and Its Applications to Material Chemistry and Catalysis Korean Chem. Soc. 24: 797-801.

Morokuma K, Musaev DG, Verena T, Basch H, Torrent M, Khoroshun DV (2001). Model studies of the structures, reactivities, and reaction mechanisms of metalloenzymes, IBM J. Res. Dev. 45: 367-375.

Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J. Comp. Chem. 19: 1639-1662.

Nakamoto K, Tsuboi M, Strahan GD (2008). Drug-DNA Interactions: Structures and Spectra. John Wiley & Sons, Inc, 119-208.

Nelson MT, Humphrey W, Gursoy A, Dalke A, Kale LV, Skeel RD, Schulten K (1996). NAMD: a parallel, object oriented molecular dynamics program. Int J SupercomputAppl High Perform Comput. 10 : 251-268.

Nelson SM, Ferguson LR, Denny WA (2007). Non-covalent ligand/DNA interactions: Minor groove binding agents. Mutation Research, 623: 24-40.

Neto BAD, Lapis AAM (2009). Recent Developments in the Chemistry of Deoxyribonucleic Acid (DNA). Intercalators: Principles, Design, Synthesis, Applications and Trends. Molecules, 14: 1725-1746.

Nielsen PE (1999). Peptide nucleic acids as therapeutic agents. Curr. Opin. Struct. Biol., 9: 353–357.

Nilsson L, Karplus M (1986). Empirical energy functions for energy minimization and dynamics of nucleic acids. J Comput Chem, 7: 591-616.

Park HJ, Hurley LH, (1997). Covalent Modification of N3 of Guanine by (+).-CC-1065 Results in Protonation of the Cross-Strand Cytosine. J. Am. Chem. Soc., 119: 629-630.

Paul A, Bhattacharya S (2012). Chemistry and biology of DNA-binding small molecules. Current Science, 102: 212-231.

Pigram WJ, Fuller W, Hamilton LD (1972). Stereochemistry of Intercalation: Interaction of Daunomycin with DNA. Nature New Biology, 235: 17-19.

Privalov PL, Dragon AI, Colyn C, Breslauer KJ, Remeta DP, Minetti CSA (2007). What Drives Proteins into the Major or Minor Grooves of DNA? J. Mol. Biol., 365: 1-9.

Rajski SR, Williams RM (1998). DNA Cross-Linking Agents as Antitumor Drugs.Chem. Rev., 98: 2723-2795.

Rao SNR, Kollman PA (1987). Molecular mechanical simulations on double intercalation of 9-amino acridine into d(CGCGCGC).d(GCGCGCG).: Analysis of the physical basis for the neighbor-exclusion principle. Proc. Natl. Acad. Sci., 84: 5735-5739.

Rarey, M. et al. (1996). A fast flexible docking method using an incremental construction algorithm. J. Mol. Biol. 261, 470–489.

Rastelli G, Del Rio A, Degliesposti G, Sgobba M (2010). Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. J Comput Chem., 31: 797-810.

Rebeca R, Begoña G, Giuseppe R, Arturo S, Giampaolo B, (2009). Computational study of the interaction of proflavine with d(ATATATATAT).2 and d(GCGCGCGCGC).2. Journal of Molecular Structure: THEOCHEM , 915: 86–92.

Reddy BSP, Sondhi SM, Lown JW (1999). Synthetic DNA minor groove-binding drugs. Pharmacology and Therapeutics, 84: 1-111.

Remers WA (1979). The Chemistry of Antitumor Antibiotics, Wiley, New York, 1-290.

Robertazzi A, Platts JA (2006). A QM/MM Study of Cisplatin–DNA Oligonucleotides: From Simple Models to Realistic Systems. Chem. Eur. J. 12: 5747 - 5756.

Schleif R (1988). DNA binding by proteins. Science, 241: 1182-1187.

Schneider G, Bohm H (2002). Virtual screening and fast automated docking methods. Combinational chemistry: reviews, 7: 64-70.

Schwarzl SM, Tschopp TB, Smith JC, Fischer S (2002). Can the calculation of ligand binding free energy be improved with continuum solvent electrostatics and an ideal-gas entropy correction. J ComputChem, 23: 1143-1149.

Senn HM, Thiel W (2009). QM/MM methods for biomolecular systems. AngewChemInt Ed Engl 48:198-229.

Senn, HM, Theil W (2007). QM/MM methods for biological systems in Topics in Current Chemistry. M. Reiher (Ed.)., Springer, Berlin, 268: 173-290.

Sherwood P, Brooks BR, Sansom MS (2008). Multiscale methods for macromolecular simulations.CurrOpinStructBiol 18: 630-640.

Sherwood P, de Vries AH, Guest MF et al (2003). QUASI: a general purpose implementation of the QM/MM approach and its application to problems in catalysis. J Mol StructTheochem, 632:1-28.

Silvestri C, Brodbelt JS (2013). Tandem Mass Spectrometry for Characterization of Covalent Adducts of DNA with Anti-cancer Therapeutics. Mass Specttrom Rev., 32: 247-266.

Simonsson S, Samuelsson T, Elias P (1998). The Herpes Simplex Virus Type 1 Origin Binding Protein specific recognition of phosphates and methyl groups defines the interacting surface for a monomeric dna binding domain in the major groove of DNA. J. Biol. Chem., 273: 24633-24639.

Singh NN, LambowitzAM (2001). Interaction of a group II intron ribonucleoprotein endonuclease with its DNA target site investigated by DNA footprinting and modification interference. J. Mol. Biol., 309: 361-386.

Sirajuddin M, Ali S, Badshah A (2013). Drug-DNA interactions and their study by UV-Visible, fluorescence spectroscopies and cyclic voltammetry. Journal of Photochemistry and Photobiology B: Biology, 124: 1-19.

Spackova N Cheatham TE, Ryjacek F, Lankas F, Meervelt L, Hobza P, Sponer J (2003). Molecular dynamics simulations and thermodynamics analysis of DNA—drug complexes. Minor groove binding between 4',6-diamidino-2-phenylidole and DNA duplexes in solution. J Am Chem Soc., 125: 1759-1769.

Srinivasan J, Cheatham TE, Cieplak P, Kollman PA, Case DA (1998). Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. J Am ChemSoc, 120: 9401-4409.

Srivastava HK, Chourasia M, Kumar D, Sastry GN (2011). Comparison of Computational Methods to Model DNA Minor Groove Binders. J. Chem. Inf. Model., 51: 558-571.

Stahl M, Rarey M (2001). Detailed Analysis of Scoring Functions for Virtual Screening. J. Med. Chem., 44: 1035-1042.

Sterkowski L, Wilson B (2007). Noncovalent interactions with DNA: An overview. Mutation Research, 623: 3-13.

Surflex, version 2.11; Tripos, Inc.: St. Louis, MO, 2007.

Svensson MJ, Humbel S, Froese RDJ, Matsubara T, Sieber S, Morokuma K (1996). ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alder Reactions and Pt (P(t-Bu).3).2 + H2 Oxidative Addition. J. Phys. Chem., 100: 19357-19363.

Tanious FA, Yen S, Wilson WD (1991). Kinetic and Equilibrium Analysis of a Threading Intercalation Mode: DNA Sequence and Ion Effects, Biochemistry, 30: 1813-1819.

Thuong NT, Hélène C (1993). Stereospecific detection and modification of double helix DNA by oligonucleotides. Angew. Chem., Int. Ed. Engl., 32: 666-690.

Umezawa H (1976). Structure and action of bleomycin. Prog. Biochem. Pharmacol., 11: 18-27.

Wang AHJ (1992). Intercalative drug binding to DNA. Current Opinion in Structural Biology, 2: 361-368.

Wang AHJ, Quigley GJ, Kolpak FJ, Crawford JL, Boom JH, Marel G, Rich A (1979). Molecular structure of a left-handed double helical DNA fragment at atomic resolution. Nature, 282: 680-686.

Wang D, Lippard, SJ (2005). Cellular processing of platinum anticancer drugs. Nature Rev. Drug Discov., 4: 307–320.

Wang J, Hou T, Xu X (2006). Recent advances in free energy calculations with a combination of molecular mechanics and continuum models. CurrComput-Aided Drug Design, 2: 95-103.

Warshel A, Levitt M (1976). Theoretical studies of enzymic

reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. J. Mol. Biol., 103: 227-249.

Watson JD, Crick FHC (1953a). Molecular Structure of Nucleic Acids. Nature, 171: 737-738.

Watson JD, Crick FHC (1953b). Genetical implications of the structure of De-oxy ribonucleic Acid. Nature, 171: 964-967.

Weerasinghe S, Smith PE, Mohan V, Cheng YK, Pettitt BM (1995). Nanosecond Dynamics and Structure of a Model DNA Triple Helix in Saltwater Solution. J Am ChemSoc, 117: 2147-2158.

Weiner PK, Kollman PA (1981). AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. J ComputChem , 2: 287-303.

Wemmer DE, Dervan PB (1997). Targeting the minor grove of DNA. Current Opinion in Structural Biology, 7: 355-361.

Wheate NJ, Brodie CR, Collins JG, Kemp S, Janice R, Aldrich-Wright (2007). DNA Intercalators in Cancer Therapy: Organic and Inorganic Drugs and Their Spectroscopic Tools of Analysis. Medicinal Chemistry, 7: 627.

Williams LD, Egli M, Gao Q (1990). Structure of nogalamycine bound to a DNA hexamer. Proc. Natl. Acad. Sci. USA, 87: 2225-2229.

Wilson WD, Tanious FA, Ding D, Kumar A, Boykin DW, Colson P, Houssier C, Bailly C (1998). Nucleic Acid Interactions of Unfused Aromatic Cations: Evaluation of Proposed Minor-Groove, Major-Groove, and Intercalation Binding Modes. J. Am. Chem. Soc., 120: 10310-10321.

Wing R, Drew H, Takano CB, Tanaka S, Itakura K, Dickerson RE (1980). Crystal structure analysis of a complete B-DNA. Nature, 287: 755-758.

Wong E, Giandomenico CM (1999). Current status of platinum based antitumour drugs. Chem. Rev., 99: 2451-2466.

Wu G, Roberston, DH, Brooks CLIII, Vieth M (2003). Detailed analysis of grid-based molecular docking: A case study of CDOCKER A CHARMM-based MD docking algorithm. J. Comput. Chem., 24: 549-562.

Yen S, Gabbay EJ, Wilson WD (1982). Interaction of aromatic Imides with Deoxyribonucleic Acid. Spectrophotometric and Viscometric Studies, Biochemistry, 21: 2070-2076.

York DM, Yang W, Lee H, Darden T, Pedersen LG (1995). Toward the accurate modeling of DNA: the importance of long-range electrostatics. J. Am. Chem. Soc., 117: 5001-5002.

Yu H, Ren J, Chaires JB, Qu X (2008). Hydration of Drug-DNA Complexes: Greater Water Uptake for Adriamycin Compared to Daunomycin. J. Med. Chem., 51: 5909-5911.

REVIEW ARTICLE

International Journal of Science, Technology and Society
Vol 3, No. 2 (2017) : 01-10

A review on QM/MM studies of nucleic bases interactions with graphene and carbon nanotubes

Asheesh Kumar, Ruchi Mishra, Deep Kumar, Devesh Kumar*

Department of Physics, School of Physical and Decision Sciences, Babasaheb Bhimrao Ambedkar University, Vidya Vihar, Rae Bareilly Road, Lucknow-226 025, India.

Publication Info

Article history:
Received : 10.12.2017
Accepted : 22.12.2017
DOI: https://doi.org/10.18091/ijsts.v3i02.11404

Key words:
Nucleic Acid (NA), Carbon nanotube (CNT), Graphene, Binding Energy, Initial Configurations (IC), QM (Quantum Mechanics) / MM (Molecular Mechanics).

Corresponding author:
Devesh Kumar

Email:
*dkclcre@yahoo.com

ABSTRACT

Nucleic bases interaction with carbonaceous materials finds significant attention due to their application in various fields such as DNA sequencing, DNA sensing and drug delivery. Nucleic bases, building blocks of nucleic acids interact with carbon nanotube and contribute significantly to the stability of the nucleic bases, carbon nanotube hybrids and their properties. In the present work, a thorough review of previous studies on the binding of nucleic bases with graphene and CNT is presented, with a focus on the simulation works that attempted to evaluate the structure and strength of binding. Dissimilitude among these works is noticed and factors that might contribute to such discrepancies are discussed in detail.

INTRODUCTION

Graphene and carbon-nanotubes have different applications for many reasons, but the differences can be ultimately attributed to the difference between one-dimensional materials and two dimensional materials. For example, a single walled carbon nanotube can be regarded as a single crystal with a high length–diameter ratio. However, the current synthesis and assembly technology cannot prepare the carbon nanotube crystals on a macroscopic scale, which limits their applications. While, graphene may be considered as a two-dimensional crystal structure, and its strength, conductivity and thermal conductivity are seen to be the best in two-dimensional crystal materials. It has a broad range of applications because of its ability to have a large area of continuous growth.

Graphene, is a planar form of carbon atoms designed in a two-dimensional hexagonal lattice fashion. It has emerged as the most dominating allotropes of carbon during the last few years. Its extended honeycomb network is the basic building block of other important allotropes such as 3D graphite formed by the stacking of several layers of graphene; 1D nanotube, obtained by rolling the graphene and the 0D fullerene prepared by wrapped graphenes (J. Allen Matthew *et al.,* 2010). Graphene is being used in the designing of new nanomaterials for energy storage devices, fuel cells and biosensors owing to its high stability, elasticity and electromechanical modulation (Stoller Meryl D *et al.*, 2008, Si Yongchao and T Samulski Edward, 2008, Pumera Martin *et al.,* 2010). Also, graphene exhibits extraordinary electronic properties in comparison to many of the conventional materials; the highly conductive graphene becomes an insulator after hydrogenation. This hydrogenation of graphene is highly reversible; the intrinsic conductivity as well as the structure of graphene can be restored on annealing (Chen Liang *et al., 2007;* Denis Pablo A *et al.,* 2009; Rubes Miroslav *et al.,* 2009). Graphene is also an important material in nanoscale electronics due to its compatibility with industry standard lithographic processing. The electron mobilities is up to 150 times greater than Si, and the thermal conductivity is approximately twice that of diamond (Ritter Kyle A *et al.,* 2009). Thus one can say that graphene has revolutionalized the technology.

Graphene sensors have emerged as another area of recent interest. The chemical and physical properties of graphene make it a promising candidate that can be used as a sensor to detect different gases such as H_2, NO_2, and NH_3. Schedin *et* al., in their experimental results, illustrated that graphene based sensors allow the sensitivity levels such that the adsorption of individual gas molecules could be detected accurately (Rao, C. N. R. *et al.,* 2009; Schedin F *et al.,* 2007). Graphene-polyaniline nanocomposite is found to be a good sensor for H_2 gas while nitrogen doped graphene find its application in electrochemical biosensing (Al-Mashat Laith *et al.,* 2010). It is also shown that, through functionalization, properties of graphene can be modified. The functionalization of graphene with hydrogen, oxygen, or other chemical groups is of prime importance as a way to engineer the different properties of graphene. A recent study reveals that with controlled epoxide functionalization, graphene can be used as a starting material for diverse chemical functionalization by chemical modification of the epoxide group. The functionalization of graphene and single-walled carbon nanotubes with individual 3d transition metal atoms were also modeled using density functional theory calculations.(Lee Geunsik *et al.,* 2009; Hubert Valencia *et al.,* 2010; Ghaderi Nahid *et al.,* 2010;Park Sungjin *et al.,*2008; Wang Donghai *et al.,2010;* Quintana Mildred *et al.*, 2010; Tachikawa Hirotoet *al.,* 2010; Al-Aqtash Nabil *et al.,* 2009).

The capability to detect single bio-molecules with high accuracy and efficiency is of prime importance in many areas of environmental science, biology, and chemistry (Lim Sung H. *et al.,* 2009; Bano Fouzia *et al.,* 2009; Zwolak Michael *et al.,* 2008; Fredlake P. Christopher *et al.,* 2006; Jonkheijm Pascal *et al.,* 2008; Patolsky Fernando *et al.,* 2006; Vidic Jasmina *et al.,* 2006). Efficient bio-sensors are expected to contribute to the improvement of medicine and medical treatment (Zwolak Michael, 2008). It is, quite uncertain whether traditional chemical techniques can be simultaneously fast and inexpensive which is another very important aspect that needs to be taken care of (Fredlake P. Christopher *et al.,* 2006). Nano-materials, due to their extreme sensitivity of the electron-transport properties in confined materials to external perturbations, form an excellent technological platform for single-molecule recognition (Zhang Guangyu *et al.,* 2006; Meyer Jannik C *et al.,* 2007; Shapir Errez *et al.,* 2008; N Kang *et al.,* 2007). Recently, graphene nano-ribbon (GNR) has emerged as a suitable candidate for making sensors for single small molecules, such as H_2, H_2O, and NO. This concept is based on measuring a variation in the source-drain current of a GNR based field-effect transistor originating from the covalent bond formed between the molecule to be detected and a defect (or an edge) of GNR. However, few reports exist on the use of GNRs as bio-sensors. One of the main reasons is that the biomolecules do not usually bind GNR via covalent bonds as a result of which the electrical perturbation induced by a biomolecule on a GNR is too weak to be detected. However the GNR is proposed to be used for the DNA sequence via π-π stacking in many reports.

Relevance of DNA bases and graphene/carbon nanotube interaction

The interaction of the biomolecules such as nucleic bases on the surface of GNR and CNT has attracted many researchers. In particular, the DNA-CNT interaction has cast its spell in the research community due to its application in various fields such as DNA sensing, DNA sequencing, and drug delivery (Zhao Xiongce 2011; Paul Ambarish 2010; Liu Zhuang *et al.,* 2011; Yarotski Dzmitry A *et al.,* 2009). It has also been found that the determination of a patient's DNA sequence can even reveal his risk of falling ill with particular diseases and it also helps to design "personalized medicine", and it is therefore the DNA sequencing that appears to be one of the most potential applications for the carbon nanostructures (Sanchez Jimenez Gerardo *et al.,* 2001; Nelson Tammie *et al.,* 2010; Prasongkit Jariyanee *et al.,* 2011). Sensors for amplified detection methods based on CNT-biomolecule composites is an area of recent interest, and such sensors can be efficiently used to detect various carbon nanostructures as well as different biomaterials such as DNA, protein, and so on (Barone Paul W *et al.*, 2005). Also, DNA-functionalized carbon nanotubes form the basis for not only a new class of chemical sensors but also for the molecular electronic devices. An ultrasensitive graphene-embedded nano channel device which effectively controls the motion of nucleobases via π-π interaction was also reported (Min Seung Kyu *et al.,* 2011).Weizmann *et al.,* 2011, recently reported that DNA-CNT nanowire networks can be used for DNA detection and Zheng Y*et al.,* 2009, constructed a carbon nanotube-based DNA biosensor for sensing the phenolic pollutants. Beside biomedical applications, comprehending the DNA-CNT interaction can also be used in the separation of carbon nanotubes as it has been shown that single-stranded DNA can be effectively used for the dispersion and separation of single-walled carbon nanotubes (Zheng M *et al.,* 2003). Some of the important conclusions can also be drawn from the studies (Wang X et al., (2011); Lu G et al.,(2009) ; Li J et al.,(2003)). Many research groups have focused on

determining the DNA-CNT interaction and tried to explore the strength of binding of different nucleosides, nucleobases, and nucleobases pairs on the carbon nanotubes and graphene, in both experimental and computational studies (Chen Robert J. *et al.,* 2003; Stepanian S.G. *et al.,* 2008; Shtogun Yaroslav V. *et al.,* 2007; Wang Hongming *et al.,* 2009; Wang Po *et al.,* 2011). The different binding energy orders for different studies are found in many experimental studies and it is understood that this may be due to the different experimental conditions applied. For most cases, in computational studies, the order is G > A > T > C > U, and in some cases, were found to be in the order as G~A~T~C > U. There are a number of theoretical and experimental studies on the nucleobases interaction with carbon nanotube and graphene surfaces as shown below:

Table 1. BE (kJ/mol) of nucleobases with SWCNT and graphene in theoretical and experimental studies.

Nucleic bases interaction with different Carbon nanomaterials			
Type	**Order**	**Method**	**Ref.**
SWCNT	T A ~C	Exp.	100
CNT(5,0)	G A T C U	Comp.	23
CNT(7,0)	G A ~T ~C U	Comp.	65
CNT (5,5), CNT(10,0)	G A T C	Comp.	86
GNR	G A ~T ~C U	Comp.	22

The binding energy of all the considered complexes illustrates that the binding energy increases as the curvature of SWCNT (single walled carbon nanotube) decreases and reaches the maximum for graphene. In general, the nucleobases tend to have π-π stacking (Zheng Y *et al.,* 2009) type of interaction with the carbon nanostructures. Hence, as the curvature of SWCNT decreases, there will be more efficient stacking between the carbon nanotube and the nucleobases surface, resulting in an increase in the binding energy of the complexes. The effect of size and curvature generally plays an important role in the non-bonded interactions (Zheng Ming *et al.,* 2003; Chen Robert J. *et al.,* 2003).

METHODS

First-Principles Methodology

In the QM methods, mostly DFT has been used in the computational chemistry and quantum physics due to their relatively low computational cost compared to high level *ab initio* methods and high accuracy in comparison to the semi- empirical methods.

The dispersion forces in the dispersion interaction are the most important interactions in molecular systems that are not addressed well in several DFT approaches. Efforts were made by the several research groups (Rutledge *et al.*, 2009; Rutledge and Wetmore, 2010; Johnson *et al.*, 2004, 2009; Dion *et al.*, 2004a; Zhao and Truhlar, 2005, 2011; Meijer and Sprik, 1996; Tkatchenko and Scheffler, 2009; Grimme, 2004, 2006; Grimme *et al.*, 2010, etc) to precisely incorporate dispersion in the correlation term of DFT. There were several studies of interaction of nucleic bases with CNT or graphene which however did not consider the dispersion interaction into account. Some early works based on LDA scheme of DFT, also lack dispersion correction. However the recent studies adopted either classical FF (force field) or dispersion corrected DFTs to consider the dispersion factor. It is believed that the π-π stacking plays a key role in the binding of nucleobase to graphene or CNT, dispersion therefore can play a significant role in determining the binding structure and BE (Binding Energy). So, one can precisely say that different approaches lead to different results. Therefore the past studies can be broadly classified into two categories: those performed with methods that consider dispersion, and those which do not consider dispersion-corrected methods.

Among the QM studies, there are also various methods with different levels of complexity and accuracy, including *ab initio* methods (HF, MP2 and CCSD (T)), DFT and semi-empirical methods. HF, originally named SCF method, is the first *ab initio* method and forms the basis of post- HF methods. Despite of having the correct description for the exchange energy, HF does not address the electron correlation precisely. Post-HF methods include MP2 (Møller and Plesset, 1934; Head- Gordon *et al.*, 1988), CI and CCSD (T) that were proposed to properly describe the correlation energy. These *ab initio* methods are usually employed for very small atomic systems due to their high computational cost. Semi-empirical QM methods are based on *ab initio* methods but include empirical parameters to speed up the calculations, examples include AM1 (Dewar *et al.*, 1985), PM3 (Stewart, 1989a,b, 1991) and PM6 (Stewart, 2007). In computational quantum chemistry and physics, DFT has been widely used, due to its relatively low computational cost compared with high level *ab initio* methods and high accuracy compared with semi-empirical methods. In 1964, Kohn and Hohenberg published the first paper on DFT in which they substituted the many electron wavefunction with the electron density and reduced the number of

variables. One year later, Kohn and Sham in 1965 improved the Hohenberg and Kohn's theory by introducing effective potential that included external potential, exchange and correlation interactions.

Methods lacking the dispersion correction

Gowtham *et al*., studied the adsorption of nucleobases (A, C, G, T and U) on graphene using MP2 and LDA (Gowtham *et al.*, 2007). Nucleobases in their work were attached to a methyl group. Plane wave basis set was used in the LDA calculations (Supercell approach), while in the MP2 calculations, 6-311++G(d,p) basis set was used with the graphene containing 28 carbon atoms terminated by hydrogen atoms at the edges. For each configuration of the nucleobase on the graphene, they initially performed a force relaxation to determine the preferred orientation and kept the bases at optimum separation distance. This was followed by a scan of the potential energy surface (PES) where the nucleobases were kept parallel to the graphene surface at a fixed distance. For each configuration, single point energy calculations were also performed and the minimum potential energy was determined. This configuration was subjected to a further optimization step in which all atoms were free to move and the final optimized structure was identified. Thereafter, BE was then calculated for the optimized structure using both MP2 and LDA. Table 2 shows the values of the obtained BEs. Among these two, MP2 predicted BE values that were almost doubled the LDA values. The BEs with respect to the different nucleobases almost remained in the same order: it was G>A=T=C>U using LDA and G>A>T>C>U using MP2. The final optimized nucleobases were found to be parallel to the graphene sheet with the separation distance being 3.5 Å.

Table 2. BE (kJ/mol) between nucleobases and graphene [Gowtham *et al.*, (2007)].

Nucleobase	LDA	MP2
G	58.86	103.24
A	47.28	90.70
T	47.28	80.08
C	47.28	77.19
U	42.25	71.40

In a later work, Gowtham *et al.,* also studied the adsorption of the same nucleobases on a (5,0) CNT (Gowtham *et al.,* 2008), using the same approach except that the BE calculation was only done with LDA only, and not with MP2. The order of the BE was found to be the same, i.e., G>A>T>C>U with the values being 47.28, 37.63, 32.81, 27.98 and 27.02 kJ/mol, respectively. Their results confirmed that the BEs for CNT were much smaller than those for graphene, that was attributed to the larger curvature of the CNT and resulting smaller area of contact.

Meng *et al.,* (Meng *et al.*, 2007a) first optimized the structures using CHARMM FF which includes an empirical description of dispersion interaction, but this dispersion was neglected again during the re-optimization step using LDA.

Meng *et al*., used a different approach (time-dependent LDA method) to study the binding between DNA nucleosides and a CNT (10,0) (Meng *et al.*, 2007b). From these simulations, the optical absorbance spectrum for DNA nucleosides were obtained, which were used to determine the preferred orientation of the nucleosides on the CNT. Optimized binding structures were also obtained using MM (CHARMM) calculations, and were found in good agreement between the MM results and LDA results. According to MM calculations, the order of the BE for the most stable structures was G>A>T>C with the BE values of 82.01, 78.15, 74.29 and 67.54 kJ/mol, respectively.

The dependence of BE on CNT chirality was studied by Wang and Ceulemans (Wang *et al.*, 2009) using LDA. They considered two connected adenosine-monophosphates with the phosphate groups terminated by H atoms. The resulting molecule was neutral and was taken to interact with different CNTs, including five (m,0) zigzag tubes with m = 7,8,9,10,17 and four (n,n) armchair tubes with n= 4,5,6,7. Periodic boundary condition using supercell approach and the linear combination of numerical atomic orbitals (LCAO) basis set with double-zeta polarizations were used.

In another work, Wang considered all four DNA nucleobases interacting with two types of CNTs: (5,5) and (10,0) (Wang, 2008). Same as his first work (Wang *et al.*, 2007), for each type of CNT, only a small part (C24H12) was made to interact with the nucleobases. Both DFT and MP2 methods were adopted in the simulations. The geometry optimization was carried out at MPWB1K/cc-pVDZ level where carbon and hydrogen atoms were kept frozen in the C24H12 fragments. The optimized structures were then subjected to a single point energy calculation at MP2/6-311++G(d,p) level. The BSSE-corrected BE for the C(5,5) CNT hybrid in vacuum was 46.46 kJ/mol which is quite different from Wang's former study (Wang *et al.*, 2007) in which the BE for the same system was determined to be 32.76 kJ/mol. The order of the BE between nucleobase and

CNT in the gas phase was found to be G>A>T>C for both CNTs. This is in agreement with the DFT studies of Gowtham *et al.,* on the interaction of nucleobases with graphene and (5,0) CNT (Gowtham *et al.*, 2007, 2008), and also with the MM results of Meng *et al.*, for the interaction of nucleosides with a (10,0) CNT.

The simulation works reviewed above are all based on methods that lack correction for dispersion interaction. With the pace of time and advancement in computational chemistry, more accurate dispersion corrected methods have been introduced.

Methods with dispersion-corrected methods

Recent works using dispersion-corrected DFT also gave rise to different results, possibly due to the difference in ways of incorporating dispersion interaction in these methods. The choice of basis sets can affect the BE evaluation, even with the same method (Shukla *et al.*, 2009). In addition, it is also found that BSSE can be large and has to be taken into account (Tournus *et al.*, 2005). Performance of simulation methods and basis set are still being widely evaluated in the computational chemistry community.

Though a large number of dispersion-corrected methods exist in literature however benchmarking has been performed by some of them (Johnson *et al.*, 2004; Dion *et al.*, 2004a; Hohenstein *et al*., 2008; Zhao *et al.*, 2008; Johnson *et al.*, 2009; Rutledge *et al.*, 2010; Zhao *et al.*, 2011; Grimme, 2011; Ehrlich *et al.*, 2013). Among these methods, Minnesota density functional developed by Truhlar's group, e.g., M05, M05-2X, M06, M06-L, M06-2X and M06-HF, are based on meta-GGA approximations (Zhao *et al.*, 2005, 2006; Zhao *et al.*, 2008, 2006 a,b). The exchange-correlation term in all Minnesota functionals depend on kinetic energy.

In the M06 family, M06-2X has shown good performance in several studies where vdW interaction played an important role (Rutledge *et al.*, 2010). Panigrahi *et al.*, employed dispersion-corrected DFT using wB97XD functional to study nucleobase-graphene binding (Panigrahi *et al.*, 2012). Nucleobases in their work were attached to a methyl group, similar to the study by Gowtham *et al.,* (Gowtham *et al.*, 2007). Each nucleobase was placed above a square graphene sheet with eight carbon rings in each direction and H atoms at the edges. The IC (initial configuration) of the base plane was parallel to the graphene surface with a separation distance of 4 Å, which was subjected to a full optimization at wB97XD/6-31G(d,p) level. The separation distance in the optimized structures was found to be around 3.5 Å. BSSE corrected BE was calculated at the same level and found to be 94.16, 85.03, 79.30, 77.04 and 68.41 kJ/mol respectively for G, A, C, T and U, i.e., G>A>C>T>U. Such order is identical to what was observed by Gowtham *et al.*, (Gowtham *et al.*, 2007) on the same system using LDA optimization accompanied by MP2 energy calculation. The BE values are also close to the MP2 results (Gowtham *et al.*, 2007) but almost double to those obtained using LDA alone.

Swathi and Chandra Shekar (with wB97XD functional) examined physisorption of nucleobases on coronene (C24H12) as a model of graphene (Chandra Shekar *et al.*, 2014). Different ICs were considered while the separation distance was considered to 3Å in all ICs. Geometry optimization was carried out at wB97XD/6-31G(d,p) level followed by a single point energy calculation at wB97XD/6-311+G (d,p). The order of the BSSE corrected BEs was determined to be G>T>A>C>U with the values of 75.73, 66.53, 65.27, 64.43 and 56.48 kJ/mol, respectively. BE values in this work were less than the ones obtained by Panigrahi *et al*., (Panigrahi *et al.*, 2012), which may be attributed to the smaller size of graphene in this study compared to that in Panigrahi *et al*. The separation distance in the optimized structures was found to be 3.24, 3.25, 3.30, 3.22 and 3.20 Å, respectively for G, T, A, C and U. These separation distances were also smaller than the ones obtained by Panigrahi *et al.,* (Panigrahi *et al.*, 2012).

Antony and Grimme studied the interaction of nucleobases with graphene in which four different sizes of graphene were considered, with 24 (C24H12), 54 (C54H18), 96 (C96H24) and 150 (C150H30) carbon atoms respectively (Antony *et al.*, 2010). Hybrids were fully optimized at B97-D/TZV(d,p) level. A three dimensional PES scan was also performed for the interaction of nucleobases with the C96H24 fragment, and no other minima was found except the one obtained from optimization. Nucleobases were attached to a methyl group and PBC was applied in their study. Full geometry optimization for the hybrid structures was also performed but no detailed explanations were given for the ICs.

Vovusha *et al.,* studied the interaction of nucleobases with graphene using M05-2X and M06-2X functional. Vovusha *et al.*, 2013 in their study of graphene model included 54 carbons with 18 hydrogen atoms capping the edge carbons. Geometry optimizations were all performed at M05-2X/6-31G(d) level and BEs were evaluated using both M05-2X and M06-2X methods with 6-31+G(d,p) and 6311++G(d,p) basis sets. The separation distance between nucleobases and graphene in the optimized structures was determined to be 3.2-3.5 Å that is close to previously reported results. Results obtained using M06-2X were considerably

larger than the ones obtained using M05-2X method. The order of the BE using M05- 2X was determined to be G>C=T>A>U and G>C>T>A>U respectively with 6-31+G(d,p) and 6-311++G(d,p) basis sets. When M06-2X was used for the BE calculation, the order was changed to G>T>A>C>U and G>T>C>A>U respectively using 6-31+G(d,p) and 6-311++G(d,p) basis sets. This demonstrates the great effect of method and basis set on the value and order of the BE. In most of the previous results on the BE between nucleobases and graphene, BE of A was only second to G, while this was not obtained by Vovusha *et al.*

Studies on semi-empirical and force-field methods

The studies of the binding of nucleobases with graphene or CNT at lower level methods involve classical MM or semi-empirical QM approaches. AM1 (Dewar *et al.*, 1985), PM3 (Stewart, 1989 a,b, 1991) and PM6 (Stewart, 2007) are the widely used semi-empirical methods. Non-bonded interactions including electrostatic and vdW forces that are implemented in classical FFs such as Amber (Cornell *et al.*, 1995) and CHARMM (MacKerel Jr. *et al.*, 1998). It has been shown that Amber FF can even be more accurate than some of the semi-empirical QM methods when evaluating the BE for biological systems (Rutledge *et al.*, 2009; Rutledge *et al.*, 2010).

Optimization process also plays a key role in BE calculation for these weakly bound systems where PES is expected to be near local minima. Direct optimization may lead system to nearby local minima, but not near the global minima. So, optimization is very sensitive to the IC chosen. The different IC result in different BE values and can even change the order of BE for different nucleobases (NB).

Umadevi *et al.*, brought to light the dependence of the curvature by considering the binding of nucleobases with graphene and a series of armchair (n,n) CNTs where n=3, 4 and 5 (Umadevi *et al.*, 2011). The graphene and CNTs were made using the Gaussian software package with H atoms at the edges was used to saturate the dangling bonds at the boundaries. Each system was optimized using ONIOM method at the (M06-2X/6-31G(d):AM1) level. Atoms of the nucleobases and the "reacting atoms" of CNTs were modeled as the high layer using M06-2X/6-31G(d). The atoms in CNT were considered as the low layer using semi-empirical AM1. Single point energy calculations were performed for the optimized structures using the dispersion-corrected B3LYP method (B3LYP-D) with the 6-31G(d) basis set. The BE was found to be graphene>CNT(5,5)> CNT(4,4)> CNT(3,3) for all nucleobases except T, for which the order was graphene>CNT(5,5)> CNT(3,3)>CNT(4,4). The BSSE-corrected BE was 30-51 kJ/mol for CNTs and 50-73 kJ/mol for graphene. The order of the BE with respect to different nucleobases was determined to be G>T>A>C>U for the CNTs and G>A>T>C>U for the graphene. In another work, using the M06-2X/6-311G**, Umadevi *et al*., found the order of binding for nucleobases with graphane in the order G > A > C > T > U (Umadevi *et al.*, 2015).

DISCUSSION AND FUTURE PERSPECTIVES

This paper presents a comprehensive review of past computational work, where three categories of methods have been used: (1) first-principles studies based on methods lacking dispersion correction, (2) first-principles studies based on dispersion corrected methods and (3) studies based on semi-empirical and FF methods. In nearly all studies reviewed above, the nucleobases were found to be parallel to the graphene or CNT with the separation distance being around 3 Å, which confirms the π-π stacking nature of the interaction. On the other hand, drastically different results have been reported for the BE.

Previous QM calculations on BE already illustrates some effects of CNT chirality (Akdim *et al.,* 2012), however it is not yet clear whether such effects are correlated with the electronic structure of the CNT. Finally, it can be noted that BE has been used as the main parameter for comparisons made in this review. Other properties such as charge transfer and density of states have only been reported in some

Table 3. BE (kJ/mol) between nucleobases and graphene [Vovusha *et al.*, (2013)].

	DFT level			
	M05-2X		M06-2X	
Nucleobase	6-31+G(d,p)	6-311++G(d,p)	6-31+G(d,p)	6-311++G(d,p
G	37.62	27.23	65.08	57.46
A	27.01	16.70	52.19	44.23
T	27.98	19.64	52.93	46.23
C	27.98	20.50	51.02	45.10
U	22.19	13.93	46.36	35.00

(<50%) of the cited works and hence are not suitable for systematic comparison. Also, the calculation of charge transfer does not only depend on the QM method but also on the charge distribution scheme (e.g., Mulliken, ESP, RESP, etc.). This makes the comparison among different studies more complicated.

CONCLUSION

In the present work, a thorough review on the theoretical studies, mainly at the QM level, on the binding of nucleobases (and in a few cases, nucleosides or nucleotides) with graphene or CNT has been performed. BE, as an indicator for the stability of the binding, is used to compare different studies. Due to the different simulated systems and procedure considered for the study, a large range of binding energy values were reported, and considerable discrepancies exist among the past investigations. So, the importance of using dispersion-corrected method and proper design of the optimization procedure plays a crucial role in understanding the interaction of the nucleobases with the CNTs or graphene.

ACKNOWLEDGEMENT

AK would like to acknowledge the UGC for the financial support.

REFERENCES

Akdim B, Pachter R, Day PN, Kim SS and Naik RR (2012). On modelling biomolecular-surface non-bonded interactions: application to nucleobase adsorption on single-wall carbon nanotube surfaces. Nanotechnology, 23: 165703(1-6).

Al-Aqtash N and Vasiliev I (2009). Ab Initio Study of Carboxylated Graphene. J. Phys. Chem. C, 113: 1290-12975.

Allen MJ, Tung VC, and Kaner RB (2010). Honeycomb Carbon: A Review of Graphene. Chem. Rev., 110: 132-145.

Al-MashatLaith, Shin K, Kalantar-zadeh K, Plessis JD, Han SH, Kojima RW, KanerRB, Li D, Gou X, Ippolito SJ, and Wlodarski W (2010). Graphene/Polyaniline Nanocomposite for Hydrogen Sensing. J. Phys. Chem. C, 114: 16168-16173.

Bano F, Fruk L, Sanavio B, Glettenberg M, Casalis L, Niemeyer CM, and Scoles G (2009). Toward Multiprotein Nanoarrays Using Nanografting and DNA Directed Immobilization of Proteins. Nano Letters, 9: 2614-2618.

Barone PW, Baik S, Heller DA and Strano MS (2005). Near-Infraed optical sensors based on single-walled carbon nanotubes. Nature Materials, 4: 86-92.

Becke AD (1993). Density functional thermochemistry. III. The role of exact exchange. J. Chem. Phys. 98: 5648-5652.

Chandra SS and Swathi RS (2014). Stability of Nucleobases and Base Pairs Adsorbed on Graphyne and Graphdiyne. J. Phys. Chem. C, 118: 4516-4528.

Chen L, Cooper AC, Pez GP, and Cheng H (2007). Mechanistic Study on Hydrogen Spillover onto Graphitic Carbon Materials. J. Phys. Chem. C, 111: 18995-19000.

Chen RJ, Bangsaruntlp S, Drouvalakls KA, Kam NWS, Shim M, Li Y, Kim W, Utz PJ, Dai H (2003). Noncovalent functionalization of carbon nanotubes for highly specific electronic biosensors. PNAS, 100: 4984-4989.

Chung C, Gautier C, Campidelli S, Filoramo A. (2010) Hierarchical Functionalization of Single-Wall Carbon Nanotubes with DNA through Positively Charged Pyrene. Chem. Commun, 46: 6539-6541.

Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, and Kollman PA (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. J. Am. Chem. Soc., 117: 5179-5197.

Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, and Kollman PA (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. Journal of the American Chemical Society, 117(19):5179-5197.

Denis PA (2011).Theoretical investigation of the stacking interactions between curved conjugated systems and their interaction with fullerenes. Chem. Phys. Lett., 516: 82-87.

Denis PA, Iribarne F (2009). On the hydrogen addition to graphene. Journal of Molecular Structure: Theochem, 907: 93-103.

Dewar MJS, Zoebisch EG, Healy EF, and Stewart JJP (1984). AM1: A New General Purpose Quantum Mechanical Molecular Model. J. Am. Chem. Soc., 107: 3902-3909.

Dion M, Rydberg H, Schrçder E, Langreth DC, and Lundqvist BI (2004a). Van derWaals Density Functional for General Geometries. Physical Review Letters, 92: 246401 (1-4).

Dion M, Rydberg H, Schrçder E, Langreth DC, and Lundqvist BI (2004b). Van derWaals Density Functional for General Geometries. Physical Review Letters, 92: 246401(1-4).

Ehrlich S, Moellmann J, and Grimme S (2013). Dispersion-Corrected Density Functional Theory for Aromatic Interactions in Complex Systems. Acc. Chem. Res., 46: 916-926.

Fredlake CP, Hert DG, Mardis ER, Barron AE (2006) What is the future of electrophoresis in large- scale genomic sequencing? Electrophoresis, 27: 3689-3702.

Ghaderi N and Peressi M (2010). First-Principle Study of Hydroxyl Functional Groups on Pristine, Defected Graphene, and Graphene Epoxide. J. Phys. Chem. C, 114: 21625-21630.

Gowtham S, Scheicher RH, Ahuja R, Pandey R and Karna SP (2007). Physisorption of nucleobases on graphene: Density-functional calculations. Physical Review B 76: 033401(1-4).

Gowtham S, Scheicher RH, Pandey R, Karna SP and Ahuja R (2008). First-principles study of physisorption of nucleic acid bases on small-diameter carbon nanotubes. Nanotechnology, 19: 125701 (1-6).

Grimme S (2004). Accurate Description of van der Waals Complexes by Density Functional Theory Including Empirical Corrections. Journal of Computational Chemistry, 25: 1463-1473.

Grimme S (2006). Semi empirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. Journal of Computational Chemistry, 27: 1787-1799.

Grimme S, Antony J, Ehrlich S, and Krieg H (2010). A consistent and accurate ab initio parametrization of density functional dispersion correction DFT-D for the 94 elements H-Pu. The J. Chem. Phys., 132: 154104(1-19).

Grimme S, Ehrlich S, Goerigk L (2010). Effect of the Damping Function in Dispersion Corrected Density Functional Theory. Journal of Computational Chemistry, 32: 1456-1465.

H-G Martin, Pople JA (1988). MP2 Energy Evaluation By Direct Methods. Chemical Physics Letters, 153: 503-506.

Hohenstein EG, Chill ST and Sherrill CD (2008). Assessment of the Performance of the M05-2X and M06-2X Exchange-Correlation Functionals for NoncovalentInteractions in Biomolecules. J. Chem. Theory Comput., 4: 1996-2000.

Hubert V, Gil A, and Frapper G (2010). Trends in the Adsorption of 3d Transition Metal Atoms onto Graphene and Nanotube Surfaces: A DFT Study and Molecular Orbital Analysis. J. Phys. Chem. C, 114: 14141-14153.

Johnson ER, Wolkow RA , DiLabio GA (2004). Application of 25 density functionals to dispersion-bound homomolecular dimmers. Chemical Physics Letters, 394: 334-338.

Johnsona ER, Mackieb ID and DiLabio GA (2009). Dispersion interactions in density-functional theory. J. Phys. Org. Chem., 22: 1127-1135.

Jonkheijm P, Weinrich D, Schrçder H, Niemeyer CM., and Waldmann H (2008). Chemical Strategies for Generating Protein Biochips. Angew. Chem. Int. Ed., 47: 9618-9647.

Kang N, Erbe A and Scheer E (2008). Electrical characterization of DNA in mechanically controlled break-junctions. New Journal of Physics, 10: 023030 (1-9).

Lee C, Yang W, and Parr RG (1988). Development of the Colic-Salvetti correlation-energy formula into a functional of the electron density. Physical Review B, 37: 785-789.

Lee G, Lee B, Kim J, and Cho K (2009). Ozone Adsorption on Graphene: Ab Initio Study and Experimental Validation. J. Phys. Chem. C, 113: 14225-14229.

Li J, Lu Y, Ye Q, Cinke M, Han J and Meyyappan M (2003). Carbon nanotube sensors for gas and organic vapor detection. Nano Lett., 3: 929-933.

Lim SH., Feng L, Kemling. JW, Musto. CJ, and Suslick KS. (2009). An optoelectronic nose for the detection of toxic gases. nature chemistry, 1: 562-567.

Liu Z, Yang K, and Lee ST (2011). Single-walled carbon nanotubes in biomedical imaging. J. Mater. Chem., 21: 586-598.

Lu G., Ocola, L. E. and Chen, J (2009). Gas detection using low-temperature reduced graphene oxide sheets. Appl. Phys. Lett., 94, 083111-083115.

Meijer EJ and Sprik M (1996). A density functional study of the intermolecular interactions of benzene. J. Chem. Phys. 105: 8684-8689.

Meng S, Maragakis P, Papaloukas C, and Kaxiras E (2007a). DNA Nucleoside Interaction and Identification with Carbon Nanotubes. Nano Lett., 7: 45-50.

Meng S, Wang WL, Maragakis P, and Kaxiras E (2007b). Determination of DNA-Base Orientation on Carbon Nanotubes through Directional Optical Absorbance. Nano Letters, 7: 2312-2316.

Meyer JC, Geim AK, Katsnelson MI, Novoselov KS, Booth TJ & Roth S (2007). The structure of suspended graphene sheets. Nature Letters, 446: 60-63.

Min SK, Kim WY, Cho Y and Kim KS. (2011). Fast DNA sequencing with a graphene-based nanochannel device. nature nanotechnology, 6: 162-165.

Moller C and Plesset MS (1934). Note on an Approximation Treatment for Many-Electron Systems. Physical Review, 46: 618-622.

Nelson T, Zhang B, Prezhdo OV (2010). Detection of Nucleic Acids with Graphene Nanopores: Ab Initio Characterization of a Novel Sequencing Device. Nano Letters, 10: 3237-3242.

Panigrahi S, Bhattacharya A, Banerjee S and Bhattacharyya D (2012). Interaction of Nucleobases with Wrinkled Graphene Surface: Dispersion Corrected DFT and AFM Studies. J. Phys. Chem. C, 116: 4374-4379.

Park S, Lee K-S, Bozoklu G, Cai W, Nguyen ST, and Ruoff RS (2008). Graphene Oxide Papers Modified by Divalent Ions-Enhancing Mechanical Properties via Chemical Cross-Linking. Acs Nano, 3: 572-578.

Patolsky F, Zheng G and Lieber CM (2006). Fabrication of silicon nanowire devices for ultrasensitive, label-free, real-time detection of biological and chemical species. Nature protocols, 1: 1711-1724.

Paul A and Bhattacharya B (2010). DNA Functionalized Carbon Nanotubes for Nonbiological Applications. Materials and Manufacturing Processess, 25: 891-908.

Prasongkit J, Grigoriev A, Pathak B, Ahuja R, Scheicher RH (2011). Transverse Conductance of DNA Nucleotides in a Graphene nano gap from first principles. Nano Letters, 11: 1941-1945.

Priyakumar UD and Sastry GN (2003). Cation-pi interactions of curved polycyclic systems: M+ (M = Li and Na) ion

complexation with buckybowls. Tetrahedron Lett. , 44: 6043-6046.

Pumera M, Ambrosi A, Bonanni A, Chng ELK, Poh HL (2010). Graphene for electrochemical sensing and biosensing. Trends in Analytical Chemistry, 29: 954-965.

Quintana Mildred, SpyrouKonstantinos, GrzelczakMarek, Browne Wesley R, Rudolf Petra, and Prato Maurizio (2010). Functionalization of Graphene via 1,3- Dipolar Cycloaddition. Acs Nano, 4: 3527-3533.

Rao CNR, Sood AK, Subrahmanyam KS, and Govindaraj A (2009). Graphene: The New Two-Dimensional Nanomaterial. Angewandte Chemie, 48: 7752-7777.

Ritter KA and Lyding JW (2009). The influence of edge structure on the electronic properties of graphene quantum dots and nanoribbons. Nature materials, 8: 235-242.

Rubes M and Bludsky O (2009). DFT/CCSD(T) Investigation of the Interaction of Molecular Hydrogen with Carbon Nanostructures. ChemPhysChem, 10: 1868-1873.

Rutledge LR and Wetmore SD (2010). The assessment of density functionals for DNA protein stacked and T-shaped complexes. Canadian Journal of Chemistry, 88: 815-830.

Rutledge LR, Durst HF, and Wetmore SD (2009). Evidence for Stabilization of DNA/RNA-Protein Complexes Arising from Nucleobase-Amino Acid Stacking and T-Shaped Interactions. J. Chem. Theory Comput., 5: 1400-1410.

Sanchez-Jimenez G, Childs B and Valle D (2001). Human disease genes. nature analysis, 409: 853-855.

Schedin F, Geim AK, Morozov SV, Hill EW, Blake P, Katsnelson MI and Novoselov KS (2007). Detection of individual gas molecules adsorbed on graphene. Nature Materials Lett., 6: 652-655.

Shapir E, Cohen H, Calzolari A, Cavazzoni C, Ryndyk DA., Cuniberti G, Kotlyar A, Felice RDF and Porath D (2008). Electronic structure of single DNA molecules resolved by transverse scanning tunnelling spectroscopy. Nature Materials, 7: 68-74.

Shtogun YV, Woods LM, Dovbeshko GI (2007). Adsorption of Adenine and Thymine and their Radicals on Single- Walled Carbon Nanotubes. J. Phys. Chem. C, 111:18174-18181.

Shukla MK, Dubey M, Zakar E, Namburu R, Czyznikowsk Z, Leszczynski J (2009). Interaction of nucleic acid bases with single-walled carbon nanotube. Chemical Physics Letters 480: 269-272.

Si Y and Samulski ET (2008). Exfoliated Graphene Separated by Platinum Nanoparticles. Chem. Mater., 20: 6792-6797.

Song B, Elstner M, and Cuniberti G (2008). Anomalous Conductance Response of DNA Wires under Stretching. Nano Letters, 8: 3217-3220.

Stepanian SG ,Karachevtsev MV, Glamazda AY , Karachevtsev VA, Adamowicz L (2008). Stacking interaction of cytosine with carbon nanotubes: MP2, DFT and Raman. Chemical Physics Letters, 459: 153-158.

Stewart JJP (1988). Optimization of Parameters for Semi empirical Methods II. Applications. Journal of Computational Chemistry, 10: 221-264.

Stewart JJP (1989). Optimization of Parameters for Semi empirical Methods I. Method. Journal of Computational Chemistry, 10: 209-220.

Stewart JJP (1990). Optimization of Parameters for Semi empirical Methods III. Extension of PM3 to Be, Mg, Zn, Ga, Ge, As, Se, Cd, In, Sn, Sb, Te, Hg, T1, Pb, and Bi. Journal of Computational Chemistry, 12: 320-341.

Stewart JJP (2007). Optimization of parameters for semi empirical methods V: Modification of NDDO approximations and application to 70 elements. J Mol Model, 13: 1173–1213.

Stoller MD, Park S, Zhu Y, An J, and Ruoff RS (2008). Graphene-Based Ultracapacitors. Nano letters, 8: 3498-3502.

Tachikawa H and Iyama T (2010). Density Functional Theory Method for Study of the Mechanism of C–H Bond Formation on Finite-Sized Graphene Surface. Japanese Journal of Applied Physics, 49: 06GJ12 (1-4).

Tang S and Cao Z (2011). Adsorption of nitrogen oxides on graphene and graphene oxides: Insights from density functional calculations. The journal of chemical physics, 134: 044710(1-4).

Tkatchenko A and Scheffler M (2009). Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. PRL, 102: 073005(1-4).

Tournus F and Charlier JC (2005). Ab initio study of benzene adsorption on carbon nanotubes. Physical Review B 71:165421 (1-8).

Umadevi D and Sastry GN (2011). Quantum Mechanical Study of Physisorption of Nucleobases on Carbon Materials: Graphene versus Carbon Nanotubes. J. Phys. Chem. Lett., 2:1572-1576.

Umadevi D, Sastry GN (2015). Graphane versus graphene: a computational investigation of the interaction of nucleobases, aminoacids, heterocycles, small molecules (CO2, H2O, NH3, CH4, H2), metal ions and onium ions. Phys. Chem. Chem. Phys., 17: 30260-30269.

Vidic JM, Grosclaude J, Persuy M-A, Aioun J, Salessea R and Pajot-Augy E (2006). Quantitative assessment of olfactory receptors activity in immobilized nanosomes: a novel concept for bioelectronic nose. Lab chip, 6: 1026-1032.

Vovusha H, Sanyal S and Sanyal B (2013). Interaction of Nucleobases and Aromatic Amino Acids with Graphene Oxide and Graphene Flakes. J. Phys. Chem. Lett., 4: 3710-3718.

Wang D, Kou R, Choi D, Yang Z, Nie Z, Li J, Saraf LV, Hu D, Zhang J, Graff GL, Liu J, Pope MA, and Aksay IA. (2010). Ternary Self-Assembly of Ordered Metal Oxide Graphene Nanocomposites for Electrochemical Energy Storage. Acs

Nano, 4: 1587-1595.

Wang H and Ceulemans A (2009). Physisorption of adenine DNA nucleosides on zigzag and armchair single-walled carbon nanotubes: A first-principles study. Physical Review B, 79:195419(1-6).

Wang P, Wu H, Dai Z, Zou X (2011). Simultaneous detection of guanine, adenine, thymine and cytosine at choline monolayer supported multiwalled carbon nanotubes film. Biosensors and Bioelectronics, 26: 3339-3345.

Wang X and Liew KM (2011). Silicon carbide nanotubes serving as a highly sensitive gas chemical sensor for formaldehyde. J. Phys. Chem. C, 115: 10388-10393.

Wang Y (2008). Theoretical Evidence for the Stronger Ability of Thymine to Disperse SWCNT than Cytosine and Adenine: self-stacking of DNA bases vs their cross-stacking with SWCNT. J Phys Chem C Nanomater Interfaces, 112: 14297-14305.

Wang Y, Shao Y, Matson DW, Li J, and Lin Y (2010). Nitrogen-Doped Graphene and Its Application in Electrochemical Biosensing. Acs Nano, 4: 1790-1798.

Wang Yi and Bu Y (2007). Noncovalent Interactions between Cytosine and SWCNT: Curvature Dependence of Complexes via π-π Stacking and Cooperative CH-pi/NH-pi. J. Phys. Chem. B, 111: 6520-6526.

Weizmann Y, Chenoweth DM, and Swager TM (2011). DNA-CNT Nanowire Networks for DNA Detection. J. Am. Chem. Soc., 133: 3238-3241.

Yan Z and Truhlar DG. (2005). Benchmark Databases for Nonbonded Interactions and their use to test Density Functional Theory. J. Chem. Theory Comput., 1: 415-432.

Yarotski DA, Kilina SV, Talin AA, Tretiak S, Prezhdo OV, Balatsky AV, and Taylor AJ (2009). Scanning Tunneling Microscopy of DNA-Wrapped Carbon Nanotubes. Nano Letters, 9: 12-17.

Zhang G, Qi P, Wang X, Lu Y, Li Xiaolin, Tu R, Bangsaruntip S, Mann D, Zhang L, Dai H (2006). Selective Etching of Metallic Carbon Nanotubes by Gas-Phase Reaction. Science, 314: 974-977.

Zhao X (2011). Self-Assembly of DNA Segments on Graphene and Carbon Nanotube Arrays in Aqueous Solution: A Molecular Simulation Study. J. Phys. Chem. C, 115: 6181-6189.

Zhao Y and Truhlar DG (2005). Benchmark Databases for Nonbonded interactions and their use to test Density Functional Theory. J. Chem. Theory Comput., 1: 415-432.

Zhao Y and Truhlar DG (2006a). A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. J. Chem. Phys., 125: 194101 (1-18).

Zhao Y and Truhlar DG (2006b). Density Functional for Spectroscopy: No Long-Range Self-Interaction Error, Good Performance for Rydberg and Charge-Transfer States, and Better Performance on Average than B3LYP for Ground States. J. Phys. Chem. A, 110: 13126-13129.

Zhao Y and Truhlar DG (2008). The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functional. TheorChem Account, 120: 215-241.

Zhao Y and Truhlar DG. (2008). Density Functionals with Broad Applicability in Chemistry. Acc. Chem. Res, 41: 157-167.

Zhao Y and Truhlar DG. (2011). Applications and validations of the Minnesota density functional. Chemical Physics Letters, 502: 1-13.

Zheng M, Jagota A, Semke ED, Diner BA, Mclean Robert S, Lustig Steve R, Raymond Richardson E. And Tassi Nancy G. (2003). DNA-assisted dispersion and separation of carbon nanotubes. nature materials, 2: 238-242.

Zheng Y, Yang C, Pu W, Zhang J (2009). Carbon Nanotube-based DNA Biosensor for Monitoring Phenolic Pollutants. Microchim. Acta, 166: 21-36.

Zwolak Michael and Ventra M Di (2008). Colloquium: Physical approaches to DNA sequencing and detection. Rev. Mod. Phys., 80: 141-165.