A Computational Study of Drug-DNA Interactions

THESIS SUBMITTED FOR THE AWARD OF THE DEGREE OF

Doctor of Philosophy

in

Applied Physics

by

Anwesh Pandey

(Enrollment No. - 074/14)

Under the supervision of

Dr. Anil Kumar Yadav

Department of Applied Physics

School for Physical Sciences

Babasaheb Bhimrao Ambedkar University

Lucknow-226025, U.P., India

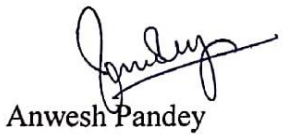
2020

...to you, **Mummy** & **Papa**!!

Thank you for all the motivation, encouragement and never-ending love & support!

DECLARATION

I declare that the thesis entitled "**A Computational Study of Drug-DNA Interactions**" has been prepared by me under the supervision of **Dr. Anil Kumar Yadav**, 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. Further, I declare that the material embodied in the present work is based on original research work and the indebtedness to others has been duly acknowledged at relevant places. This is also declared that the thesis is essentially free from any kind of plagiarism.



Anwesh Pandey
Department of Applied Physics
School for Physical Sciences
Babasaheb Bhimrao Ambedkar University
Vidya Vihar, Raebareli Road
Lucknow-226025,
U.P., India

Date: 20-08-2020
Place: Lucknow

CERTIFICATE

This is to certify that the thesis titled **"A Computational Study of Drug-DNA Interactions"** submitted by **Mr. Anwesh Pandey** 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.

The thesis submitted to Babasaheb Bhimrao Ambedkar University, Lucknow satisfies all the requirements as stipulated in the *Doctor of Philosophy (Ph.D.) regulations - 1999 as amended in 2013* and it is fit for submission and evaluation for the award of Doctor of Philosophy of the University.

Supervisor

Dr. Anil Kumar Yadav
Assistant Professor
Department of Physics
B.B.A.U., LUCKNOW

Date:20-08-2020

Head of the Department

ACKNOWLEDGEMENTS

Thank you, **Lord Shiva** for your blessings and for giving me the will power and determination to pursue my dream. Thank you for everything!

First and foremost, I would like to thank my mother **(Mrs. Neeta Pandey)** and my father **(Mr. Anirudh Pandey)** for their immense blessings and encouragement and love and support and for allowing me to dream and for always standing by my side to see me pursue a carrier that made me happy. No permutation/combination of 24 alphabets can ever convey the overwhelming feelings I have for them. Neither do I wish to do so!

I am very much indebted to my supervisor **Dr. Anil Kumar Yadav** for his generous support, encouragement, guidance and for providing me inspiring research atmosphere during the period of my research work. I am thankful to him for introducing me to the subject. I would like to extend my thanks for his patience with which he dealt me, his critical remarks and valuable suggestions have proven a guiding lamp on the road to success. But most importantly I learnt patience from you sir. I will always remain in-debt for your guidance and support!

I feel myself falling short of words to thank **Prof. Devesh Kumar**, Former Head, Dept. of Physics, for providing me all the necessary softwares and computational facilities and specially for having had faith in my ideas and to have supported me all throughout this journey. I can never repay you sir! I wish you always remain healthy and have a long life, sir!

I would like to extend my gratitude towards **Prof. B.C. Yadav**, Head, Dept. of Physics, **Dr. Ramesh Chandra**, **Dr. Devendra Singh**, **Dr. K.B. Thapa** and **Dr. S. Sikarwar** for constantly encouraging me through their experiences during DRC meetings. Their valuable suggestions and strong criticisms have made this work more significant.

I would also like to mention and thank my senior, **Dr. Ruchi Mishra**, Asst. Prof., Invertis University, Bareilly, who was always there for all the frequent doubts and discussions during the entire course of my research work. I further extend my regards to my seniors **Dr. R.K. Tripathi**, Dr. D.S.K. Post-Doc. Fellow, Dept. of Physics, Babasaheb Bhimrao Ambedkar University, Lucknow, **Dr. Rajkamal Shastry**, Asst. Prof., BSNV PG College, Lucknow and **Dr. Suresh Kumar**, Dr. D.S.K. Post-Doc. Fellow, Dept. of Physics, Banaras Hindu University, Varanasi, for their strong support, criticisms and humble appreciations.

A very warm and special thanks to **Prof. C.V. Shastry**, Dept. of Chemistry, Indian Institute of Technology, Guwahati and **Dr. H.K. Shrivastava**, Associate Professor, Dept. of Medicinal Chemistry, National Institute of Pharmaceutical Education and Research, Guwahati for giving me a chance to visit their labs for the computational training.

Further, I would like to thank the authorities of Gautam Buddha Central Library, Babasaheb Bhimrao Ambedkar University and the computer lab facility of Dept. of Physics which were very useful at every stage of my research work. I would also like to thank the staff of Dept. of Physics for their generous support in the paper formalities.

I also gratefully acknowledge the financial support from the University Grants Commission (UGC), Government of India, New Delhi for providing fellowship during the course of my research work.

Family, teachers, colleagues, friends and everyone else! I would like to name a few special people who were always there with me through every thick and thin. Rolly Yadav, Sachin Yadav, Akash Mishra, Nitesh Jangid and Pawan Rastogi. Thank you so much guys for bearing with me every time and for all the support and delightful conversations between lunch, coffee & tea breaks. Friends like you are worth keeping!

I would also like to mention a few of my colleagues, Shivani Chaudhary, Anamika Shukla, Bhavana Pal, Deeptarka Roy and Nidhi Awasthi. Thank you, guys, for sharing your food during the lunch and for sharing a cheerful and joyful environment in the lab & department by your presence.

A very warm and special thanks to Arpit Verma, Adarsh, Ajay and Harsh for opening their hearts and home at the hour of need during the COVID-19 pandemic. A mere thank you cannot justify the generosity of yours. Always keep in touch!

This journey has been a dream, not only for me but for my family too. I feel extremely joyful in conveying my love to my family members for just being there and seeing me and keep pushing me to bring out the best out of myself. A special name needs to be mentioned here, my younger sister, **Krishna**. I owe her a lot! She was a constant source of joy and happiness for me. I cannot put them to words! Just keep spreading the smiles, like always!

Last but not least, there are indeed a few more people who have been an indispensable part of this journey, from mess manager Raja Shukla to the doctors of

the University Health Center; I would like to thank one and all, who directly or indirectly helped me out during my research work, eventually leading to this thesis. Thank you from the bottom of my heart!

Anwesh Pandey

ABSTRACT

DNA is a well-established biomolecular target for a wide range of drugs that may or may not specifically interact with it and hence affect its biological functions. There is a huge number of anticancer, antibiotic, and antiviral drugs which express their primary biological effects by interacting with nucleic acids, viz., DNA. In addition to it, there is a whole lot number of drugs which are still under various clinical trials and therefore making DNA an utmost import drug target due to its gene expression tendencies. Therefore, it can be said that, significant progress has been gradually made over the past few years in the field of drug-DNA interactions.

Interaction between small molecules and DNA is classified into of two categories, viz., covalent interactions and non-covalent interactions. Three major modes of non-covalent interactions are electrostatic interactions, groove binding and intercalative binding. Numerous elaborative computational techniques have been employed to decipher and provide in-depth explanations for these interactions. To the best of our believes, these techniques provide data that are not only reliable but simple to interpret owing to computational rigorousness.

The research work carried out in this thesis deals with drug–DNA interactions, their types and applications of elaborate computational techniques to study these interactions and provide underlying explanations for binding of molecules that hold some potential pharmaceutical interest. The explanations of the molecular interactions in this research work; between different DNA binding molecules has been done by using benchmarked computational techniques such as, Molecular Docking, Molecular

Dynamics Simulation, Free Energy Calculations and QM/MM calculations etc. Molecular docking is a probabilistic approach to determine the possible drug binding site in the vicinity of the biological macromolecule. Molecular dynamics assures the stability of the drug-DNA complexes with the evolution of time. Free energy calculations offer key insights into the molecular forces that drive and determine the key factors responsible for stable complex formation. Quantum mechanical/molecular mechanical (QM/MM) approach is used for the modelling of reaction mechanism by computationally mimicking the biomolecular systems which are not understood so far.

For the improvement of the clinical efficacy of existing drugs and in the design of new ones, complete understanding of the molecular basis of drug–DNA interactions in structural, thermodynamic, and kinetic detail is a prerequisite. The most fundamental problems related to drug discovery are based upon the progress of recognition by small molecules. Binding specificity of DNA with small molecules is explained mainly by studying the formation of hydrogen bond and polar interactions. The action mechanism of majority of DNA binding molecules at molecular levels is still unknown and therefore due to urgency it becomes and utmost important field of global interest.

This study confirms that the minor groove binders bind to DNA through different types of non-covalent interactions. DNA minor groove binders form hydrogen bonds between the drug and DNA bases. DNA recognition by drug molecules does not directly depend upon the base sequences but also depends on the sequence dependent conformation or DNA modifications and distortions. This study

provides a complete understanding of drug-DNA interactions, sequence selectivity and sequence specificity through usage of advanced computational techniques.

PREFACE

This thesis, entitled **"A Computational Study of Drug-DNA Interactions"** encapsulate the results obtained on theoretical investigations carried out in the Department of Applied Physics, Babasaheb Bhimrao Ambedkar University under the supervision of Dr. Anil Kumar Yadav, Assistant Professor, Department of Applied Physics, Babasaheb Bhimrao Ambedkar University, Lucknow.

Deoxyribonucleic acid DNA is one amongst various other biomolecules having great biomedical significance. Transcription and replication are the two vital processes that are essential for the survival of living organisms. Transcription holds itself responsible for the production of proteins whereas replications is accountable for making copies of the DNA. The knowledge of drug activity for various cancer, anti-microbial and anti-viral drugs depend upon their binding with DNA and hence it makes this a fascinating field of current global interest. Therefore, in order to design effective therapeutic agents targeting DNA, it is essential to explore their interactions with DNA.

Computational modelling studies in such field holds great significance in lieu of experimental setups. Therefore, it is crucial to explore such fields of biomedical interests with benchmarked computational techniques. This thesis will focus upon the interactions taking place between selected antimicrobial molecules and DNA and the factors responsible for stability of the drug-DNA complexes. Specific attention will be paid to the formation of hydrogen bonds and their critical in-depth analysis followed by the implementation of advanced computational techniques.

The work carried out in the present thesis has been divided into seven chapters.

Chapter 1 discusses the structure of DNA, different types of interactions between drugs and DNA, the different methods which are used to study the interactions between drug & DNA. This chapter also reviews the literature work done so far in this field. In **Chapter 2**, different molecular modelling techniques that are used in this thesis are discussed in detail, giving a brief idea of the theoretical protocol being followed in their development and implementation.

Chapter 3 elucidates the comparative molecular docking studies on two classes of antimicrobial agents and discusses the underlying factors for their stability and interactions with DNA.

Chapter 4 describes the interactions between some minor groove binding agents and DNA through the usage of molecular modelling techniques such as molecular docking and molecular dynamics and therefore provides necessary information to complement the docking results.

Chapter 5 presents the binding mechanisms between some antimicrobial drugs and DNA studied through extensive computational techniques viz., molecular docking, molecular dynamics and free energy calculations and thus provides elaborative descriptions for the present interactions.

Chapter 6 uses elaborate and advanced computational technique like, QM/MM to study interactions between selected antimicrobial agents and DNA and thus explains the role of interactions with great sophistications. The general conclusions drawn from the present thesis are summarized in the **Chapter 7**.

Hopefully, a better understanding of interactions occurring between small molecules and DNA will enable more efficient and effective work of computational drug design to be performed, well into the future.

ABBREVIATIONS

A	Adenine
AMBER	Assisted Model Building with Energy Refinement
C	Cytosine
CC	Coupled Cluster Theory
CD	Circular Dichroism
CHARMM	Chemistry at HARward Macromolecular Mechanics
DFT	Density Functional Theory
DNA	De-oxy ribonucleic acid
G	Guanine
GA	Genetic Algorithms
GAFF	Generalized Amber Force Field
GLIDE	Grid-based Ligand Docking with Energetics
GOLD	Genetic Optimization for Ligand Docking
GTOs	Gaussian Type Orbitals
HF	Hartree Fock
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
NMR	Nuclear Magnetic Resonance
PBC	Periodic Boundary Conditions
PDB	Protein Data Bank
PME	Particle Mesh Ewald
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
U	Uracil
VS	Virtual Screening

LIST OF FIGURES

Chapter **Page No.**

Chapter-1

Chapter-2

Chapter-3

Chapter-4

Chapter-5

Chapter 6

LIST OF TABLES

Chapter **Page No.**

Chapter-1

Chapter-2

Chapter-3

Chapter-4

Chapter-5

Chapter 6

CONTENTS

Chapter 1

Introduction

Chapter-1

Introduction

Ever since the dawn of civilization, mankind has been prone to many deadly contagious, non-contagious and life-threatening diseases. Earlier the doctors had devised medicines based on natural products for the cure of simple viral and bacterial infections and various other diseases. But as the generations passed and civilization started getting indulged into industrialization and started depending heavily upon machinery and technology, there came modern advancements in the field of healthcare, medicines and drug discovery followed by gradual advancement in the techniques of drug discovery and their implementations [1-5]. From natural products to the usage of modern computational techniques such as, artificial intelligence, machine learning, data science, etc., the field of drug discovery has evolved itself accordingly for the maximum benefit of mankind [6-11].

The information of accurate biomolecular interactions is the prerequisite for structure based rational and effective therapeutic drug design. In particular, modeling studies on DNA, RNA, enzymes, proteins, lipids, etc. is certainly a huge interdisciplinary enterprise including biology, chemistry, physics, computer science and even mathematics [12,13]. Biologists describe the problems at the cellular levels, chemists fill in it the atomic and molecular designations, physicists formulate the problems at electronic interaction levels and tend to describe the underlying forces beneath, mathematicians formulate the algorithms and develop numerical models

whereas computer scientists support in terms of challenging computational techniques [14-16].

The deoxyribonucleic acid (DNA) is the core bio-molecule, essential for controlling various bio-chemical reactions taking place inside the cell. Interactions of DNA with variety of drugs is amongst one of the most important aspects of bio-medical studies. The prior knowledge of target specificity in rational design of chemo-therapeutics is a significant fundamental factor. Observing the significance of drug-DNA interactions in the field of drug discovery and drug mechanism provides good evidence for the development of new rational design of target-oriented drugs and explanations of their possible mechanism of action. Several similar approaches can also prove themselves to be key-players in providing new insights into target specific drug design and then lead in understanding of their possible interaction mechanism with DNA [17, 18]. DNA is a core biological target for a class of huge number of chemical compounds that act as drugs, because of its gene expression and protein formation tendencies. Many drugs act as inhibitors thereby inhibiting the action of a specific protein. In addition, the DNA replication processes can also be altered significantly, which is important for cell growth and cell division thereby implying DNA to be one of the prominent drug targets in various disease's therapies [19].

In the past several years, research has also shifted its focus towards how the drugs, that target DNA proteins can actually attach themselves to the vicinity of DNA for better implementation of drug action and thus producing lesser side-effects. Eventually, researchers identified three distinct classes of drugs, based upon their binding mechanisms for attaching themselves to DNA, viz., major & minor groove

binders and intercalators [20-22]. Therefore, we can, for sure, say that, drugs will continue to be developed targeting DNA. Our increasing knowledge of DNA and its significant role in human body can hopefully provide more successful treatments that target the disease and eventually help the people who suffer from it's devastating effects [23].

Computational techniques have always been key players in *In-silico* drug design and in modeling the targets for the therapeutic action of drugs to begin [24-27]. The interactions present between drug & DNA helps in controlling and regulating the replication and transcription processes in DNA and thereby inhibiting the progression of the diseased part of the DNA. This thesis encapsulates the robust computational techniques used in solving challenging biological problems; herein, our problem of interest was drug-DNA interactions. The computational techniques, viz., molecular docking calculations [24], molecular dynamics simulations [25] and quantum mechanical/molecular mechanical (QM/MM) calculations [26,27]; used while carrying out the entire research work are benchmarked through years of gradual research.

The problems raised in this thesis mainly focus upon how certain drugs find their target binding site in the vicinity of DNA and then deals with the implementation of various advance computational techniques for the studies and analysis of the interaction and stability of drugs with those binding sites. To name a few techniques that are implemented in this thesis; molecular docking calculations, molecular dynamics simulations, free energy calculations and quantum mechanical (QM) & quantum mechanical/molecular mechanical (QM/MM) calculations, using the two

layered ONIOM scheme. This thesis not only accounts for the analysis of interaction and stability of drug molecules in the vicinity of DNA but also incorporates the roles of various underlying factors responsible for the same and thus fulfills its objective.

1.1 Nucleic Acids

Nucleic acids are the prime contents of cells of all living organisms. Because of their primary occurrence and acidic nature, they are termed as nucleic acids. The unit cell of a nucleic acid is called nucleotide, i.e., nucleic acids are composed of linear polymeric chains of nucleotides. DNA & RNA are the two basic kinds of nucleic acids that hold long thread like macromolecular structures and play essential roles in various hereditary processes. The main component of chromosomes is the DNA, and is associated with a small amount of chloroplast and mitochondria whereas, RNA is mostly found in plant viruses and cytoplasm [28]. **Figure 1.1** shows the constituents of a cell:

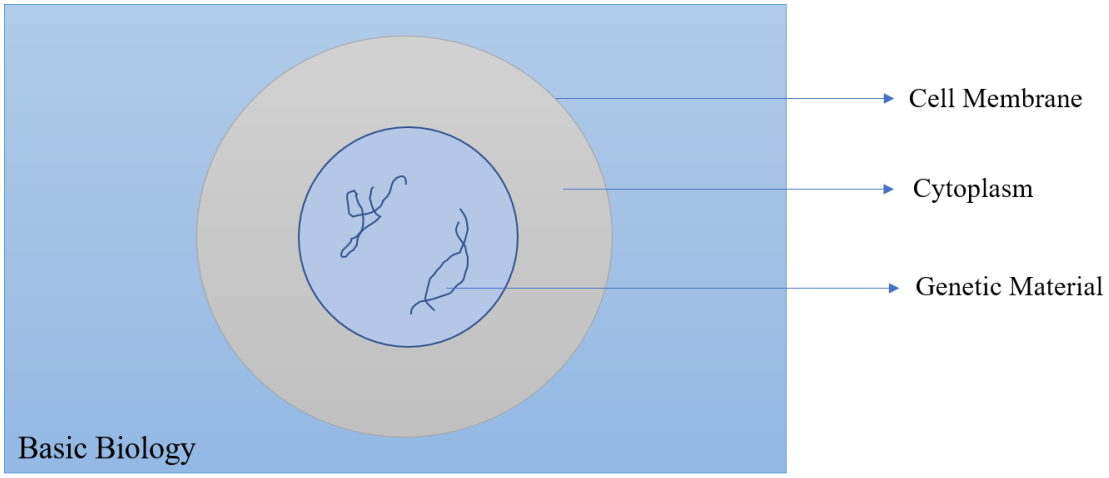


Figure 1.1: Basic constituents of the cell

Free occurrence of DNA & RNA is very rare. Each nucleic acid consists of sugar-phosphate poly-deoxy/oxy ribonucleotide double helical entangled strand.

Ordinarily, nucleic acids are composed of nitrogenous bases, sugars, phosphate groups, each of which are discussed as follows:

1.1.1 Nitrogenous Bases

Nitrogenous bases are nitrogen containing organic rings. Each of the nucleotide in the DNA consists of adenine (A), thymine (T), guanine (G) and cytosine (C), whereas in RNA there is another base called uracil (U) in place of thymine (T). These nitrogenous bases are further classified into two types, viz., purines and pyrimidines. In case of purines, the parent substance purine consists of a six-membered pyrimidine ring joined to a five-membered imidazole ring. Adenine (A) and guanine (G) fall into the category of major purines and have been isolated from nucleic acids. Whereas pyrimidines are six-membered heterocyclic rings and have not yet been isolated from nucleic acids. Cytosine (C), thymine (T) and uracil (U) fall into the category of pyrimidines. **Figure 1.2** represents the chemical structures of purines as well as pyrimidines.

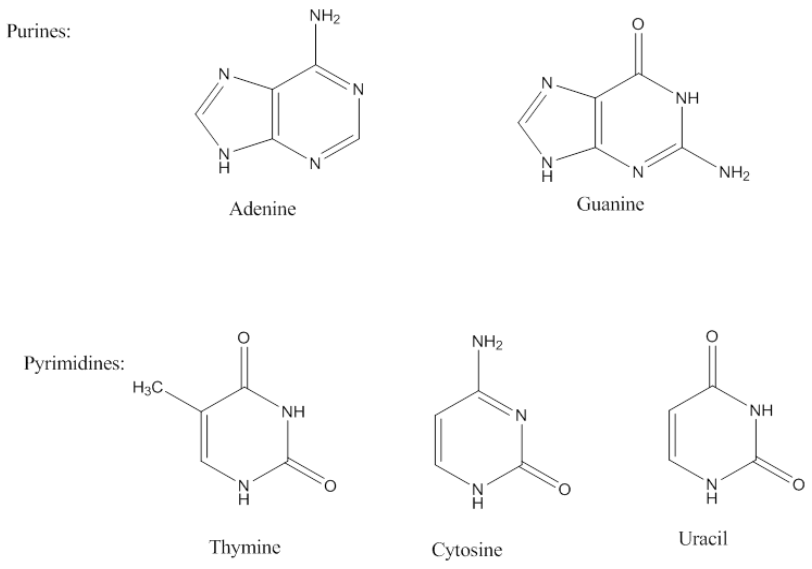


Figure 1.2: Figure depicting purines and pyrimidines nitrogenous bases

1.1.2 Sugars

The sugar components of nucleic acids are ribose and deoxyribose sugars. These sugars are pentoaldoses and occur in the form of furanose rings. Both, DNA as well as RNA contains sugars but each of them consists of different sugar units. DNA is made up of deoxyribose sugar, whereas RNA is made up of ribose sugar. These sugars differ by the fact that the second carbon atom bears a hydroxyl (–OH) group. **Figure 1.3** [29] depicts both deoxyribose sugar as well as ribose sugar.

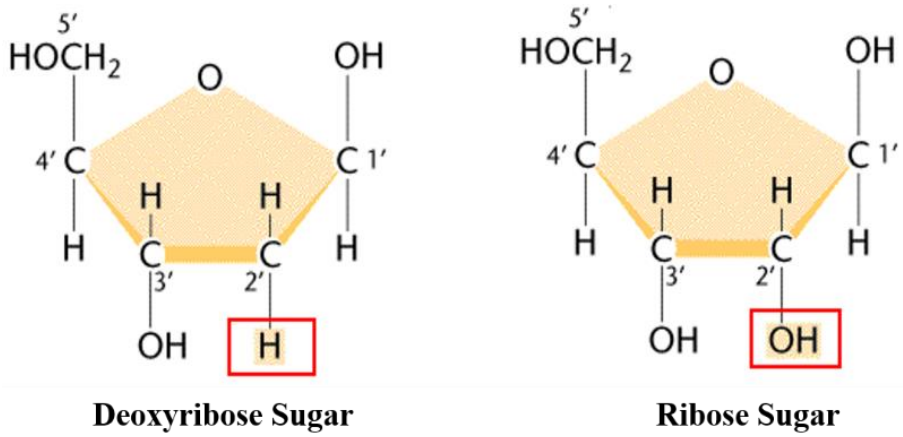


Figure 1.3: Figure depicting deoxyribose and ribose sugars

1.1.3 The Phosphate Group

The phosphate group consists of d-electrons. In a phosphate group, a phosphorous atom is linked to four oxygen atoms via sp^3 hybridized σ bonds having tetrahedral geometry. **Figure 1.4** shown below represents a phosphate group.

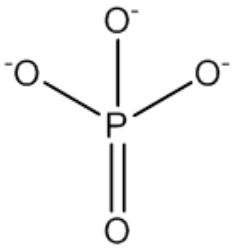


Figure 1.4: Figure representing a sp^3 hybridized phosphate group

1.1.4 Nucleoside and Nucleotide

A deoxyribose or ribose sugar attached to a purine or pyrimidine is termed as a nucleoside. Adenine linked to a ribose or deoxyribose is called adenosine, similarly guanine nucleoside is called guanosine. The pyrimidine nucleosides are called uridine, cytidine and thymidine. Nucleotides are phosphate esters of nucleosides and are strongly acidic in nature. Similarly, oxyribose sugar also form nucleosides and are termed as, adenosine monophosphate (AMP) and similarly GMP, TMP, CMP and UMP respectively. **Figure 1.5** depicts the arrangement of nucleosides and nucleotides along with sugar and phosphate group.

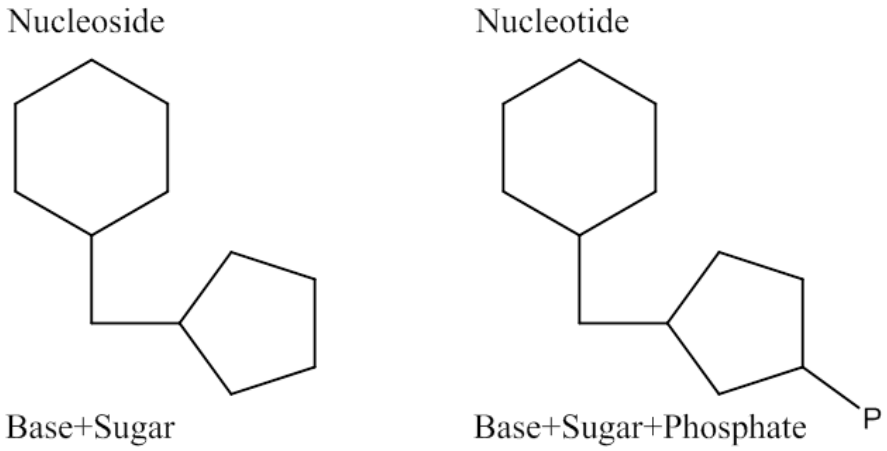


Figure 1.5: Figure depicting a nucleoside and a nucleotide group

1.2 History & Structure of DNA

Life emerges out of a cell (embryo & sperm) and then it gradually replicates (via cell division) and thus forming a human baby. DNA is the key ingredient in such biological process; it regulates and transfers the genetic codes from one generation to another, passing on the important chromosomal contained information. DNA always

existed in all living beings, but was discovered formally by Friedrich Mischer in 1868 [30], while working with white blood cells obtained from pus drained out of surgical bandages.

Ever since then, there was a strong agitation among the scientists all around the globe to get deeper insights into DNA and its conformational structure. To this, Avery and coworkers reported that nucleic acids are the carriers of genetic information and not proteins [31]. In 1950, Chargaff discovered that DNA for each and every species was unique in its own way [32]. Roseland Franklin used experimental technique of X-Ray crystallography to elucidate the helical model of DNA [33] however, it was Watson and Crick who is credited to have modeled the double helical structure of the DNA and for recognizing the relationship between the nitrogenous bases Adenine (A), guanine (G), cytosine (C) and thymine (T), for the first time, ever.

Their double helical model for DNA, significantly suggested that A/T and G/C bases pair with each other, whereas the phosphate backbone lied outside as a supporting structure and the nitrogenous bases were held to each other via hydrogen bonds pointing inwards owing to have modeled a right handed double helical DNA [34,35]. **Figure 1.6** represents the Watson and Crick double helical DNA model.

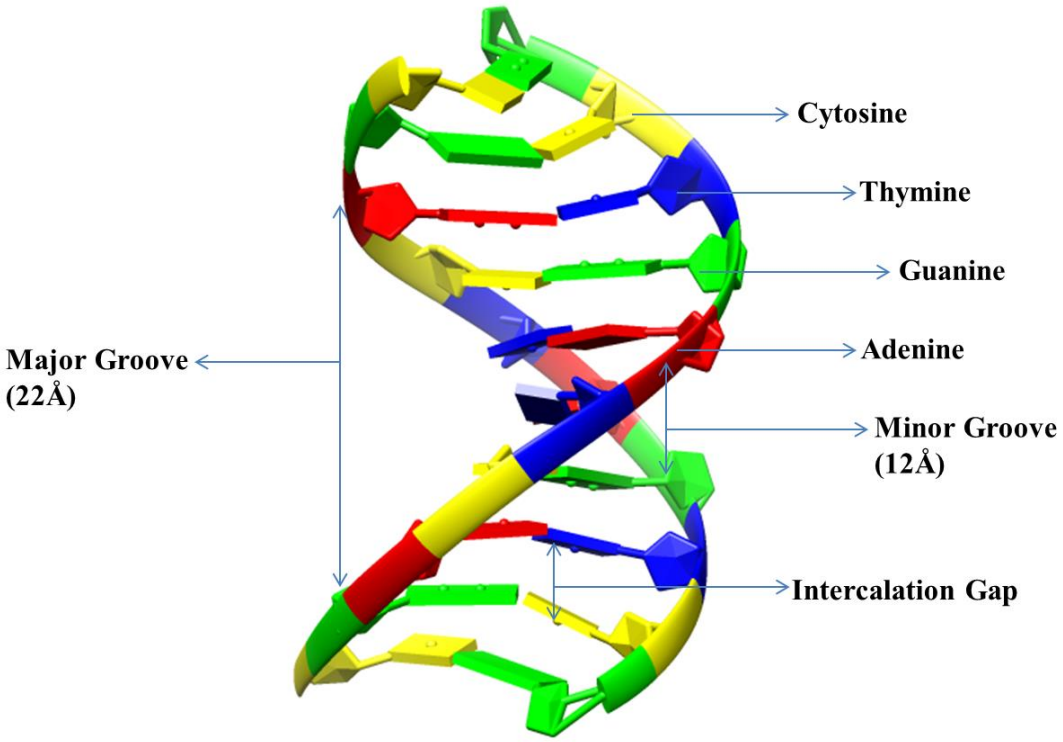


Figure 1.6: Watson & Crick's double helical DNA model

The functions of the DNA, and thus the underlying biological processes, can be artificially controlled, activated or inhibited as per requirement by mimicking the binding specificity of small molecules to DNA. Since then many efforts were frequently made and thus DNA was eventually modeled into various conformations depending upon its topological information obtained. Some of the few common conformations of DNA along with their structural properties are mentioned in [36] **table 1.1**. Whereas **figure 1.7** shows the central dogma [37] of protein synthesis from DNA and RNA.

Table 1.1: Structural properties of different DNA conformations

Property		A-DNA	B-DNA	Z-DNA
Helix Sense		Right	Right	Left
Twist/bp (Å)		32.7	36	-9, -51
Rise/bp (Å)		2.56	3.4	3.8
Residues/bp (Å)		11	10	12
Mean rotation/bp		33.6°	35.9°	-30°
Helix axis location		Through base pairs	Through Major groove	Through Minor groove
Groove width	**Major(Å)**	2.7	11.7	8.8
	Minor(Å)	11	5.7	-
Groove depth	**major(Å)**	13.5	8.8	3.7
	minor(Å)	2.8	7.5	3.7

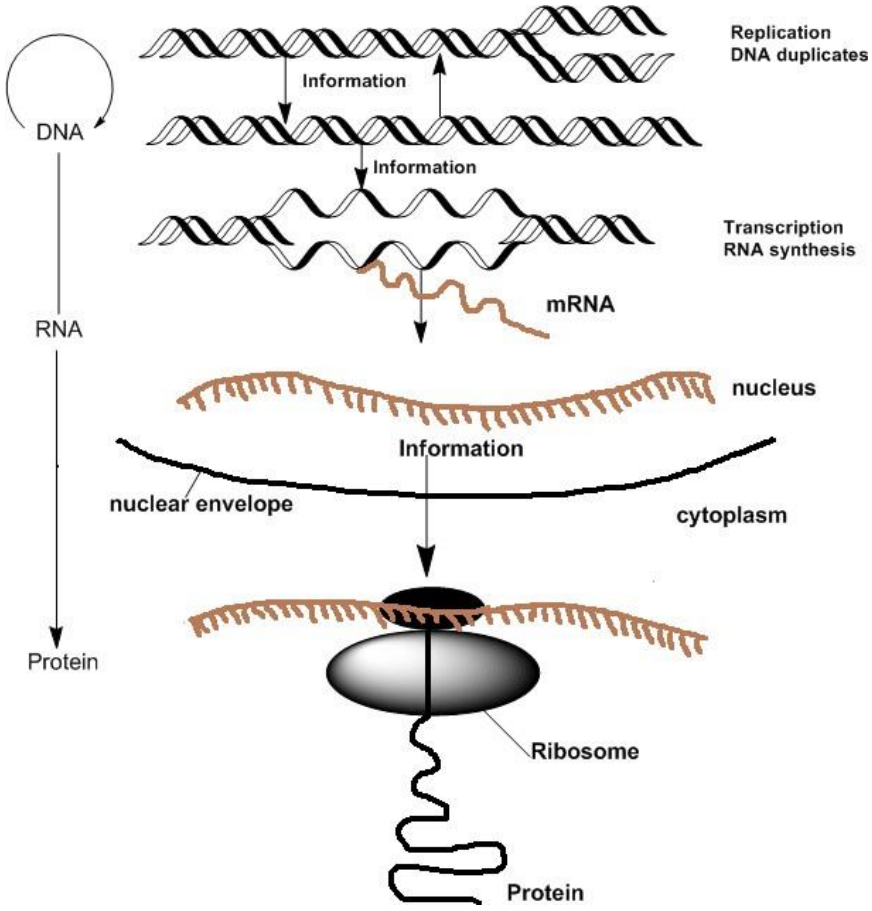


Figure 1.7: Figure representing key cellular processes of DNA replication and translation

1.3 Drug-DNA Interactions

DNA owes itself to be a prime target for many anti-bacterial, anti-fungal, anti-microbial and deadly diseases such as cancer, etc. It is an essential target of many clinical drugs which are still under trials. The vital life processes, that are essential for living organisms involve transcription and replication of DNA [38]. Transcription involves fetching of information from DNA to RNA whereas in replication DNA self-yields two identical DNA strands. These processes are governed by specific form of regulatory proteins. These proteins are generated at particular regions of DNA where certain inhibition in required [39]. Now, if this DNA regulatory mechanism is mimicked by some drug molecule (usually heterocyclic aromatic molecules), then the vital life processes of DNA can be artificially controlled and regulated and thereby inhibiting a particular genetic process or a severe disease. The DNA held vital life processes such as; replication, transcription, etc. are of significant interest as targets of wide range of diseases [40,41].

Molecular interactions occurring between DNA base pairs and drugs is the field of current global interest. It also plays a very significant role in the biological activity of drug controlled vital DNA processes. Many disease inhibiting therapies depend upon the interaction between the drug molecule and DNA. These drug molecules act by interacting with the diseased part of the DNA and thereby inhibiting the growth of the disease because these compounds have the tendencies to bond to DNA at specific sequences and thus interfere with DNA topoisomerases or with DNA transcription factors [42]. To design such effective and efficient therapeutics, it is very important to have a deep insight into drug binding mechanics with DNA.

The mechanism of action of various cancer, anti-microbial and anti-viral drugs depend upon the binding mechanism involved of that drug with DNA or eventually leading to modified DNA activity and thus leading to inhibition of diseases. Although complete information for drug-DNA interactions is still a topic of current global research and has not yet been fully deciphered, but on the basis of literatures and research work conducted we can say that, indeed it is very interesting in learning about new drug action mechanisms as well as in contributing insignificantly towards design of new and potent drugs with minimal side effects [43]. In the current research work, focus has not only been made upon the investigations on the molecular basis underlying the drug-DNA interactions but also a few significant efforts have been made to study the role of hydrogen bonds in such interactions. Thus, these studies can be very fruitful in discovery and design of new and efficient drugs provided they have minimal side effects [44,45].

1.3.1 Forces Involved in Drug-DNA Interactions

Binding of drugs to DNA involves many energy contributions terms of the form of conformational charges, entropy contribution during biomolecular complex formation, hydrophobic charge transfer process, and from some of the non-covalent inter/intra-molecular interactions occurring within the dug-DNA complex [46]. Some of them play a very important role in determining the factors responsible for the stability of drug-DNA complexes [47-52].

1.3.1.1 Van der Waals Interactions

When two neutral molecules are brought into the vicinity of each other, the electrons in their valence shells tend to repel and thus exhibiting a force upon the atoms. During such an interaction a transient dipole is set up in one of the atom/molecules and thus it tends to generate another dipole in the other atom/molecule leading to their mutual interaction. Such interactions are termed to be Van der Waals interactions. These forces are highly dependent upon the shape of the interacting molecules. Van der Waals radii describe the closest proximity to which the interacting species would approach. In the case of drug-DNA interactions, Van der Waals forces hold significant importance in describing the binding and stability of the complexes.

1.3.1.2 Hydrogen Bonding

Formation of hydrogen bonds in drug-DNA complexes is an essential feature and is a key player in determining the involved interaction and in predicting the stability of the system. It usually forms between the functional groups attached on the drug molecule and the bases of DNA. Since all the hydrogen bonds are linear and therefore their contribution to favorable energy change is very less when drug interacts with DNA in solution phase. However, a penalty of 4kJ/mol is offered when a non-linear or poorly aligned hydrogen bond is formed, it is also offered when there is absence of any hydrogen bond. Therefore, the presence of hydrogen bonds signifies the sequence specific interactions between drug & DNA and adds to the sustained stability of the system [47].

1.3.1.3 Electrostatic Forces

Electrostatic forces or salt bridges are those interactions that are formed between ionized phosphates of nucleic acids and the positively charged group of the drug molecule in the drug-DNA complexes. They offer stabilization energy of about ~40kJ/mol per salt bridge. The strength of salt bridge tends to decrease the concentration of salts in drug DNA complexes. These kinds of electrostatic forces are of significant strength when there is absence of water molecule between the two ionized groups reason being water has considerably high dielectric constant. These forces fall into the category of long ranged interactions [52].

1.3.1.4 Hydrophobic Forces: Entropic Forces

This force is caused due to the presence of water molecules at the interaction surfaces between drug & DNA. Any molecule surrounded by water tends to create a sharp cusp like curved surface of ordered molecules around itself. As there is accumulation of water molecules at the interface it gets disordered thereby increasing the entropy of the system. Whereas water molecules which are left at the interface of drug-DNA interaction site tend to decrease the entropy of the system [53]. Thus, it is essential that the chemical structure of the non-aromatic drug chromophore to be exactly complementing in such a manner that there are no unnecessary water molecules thereby hindering the moiety where the drug-DNA complex formation is to take place [54].

1.3.1.5 Dispersive Forces due to Base Stacking

Base pair stacking occurs mainly due to two major kinds of interactions, viz., the hydrophobic effects and dispersive forces. Molecules which do not possess net dipole moment tend to attract each other by transient dipole-induced dipole interactions, this leads to a decrease in the dispersive forces by an inverse of the sixth power of the distance between the two dipoles; the reason behind this is that they are exceptionally sensitive to thermal agitation of the molecules involved [55-57]. Despite of thermal dependence and being so sensitive towards distance dependency, dispersive forces play a very crucial role in maintaining the firmness of the double helix of the DNA strands by providing support in base pair stacking. Another very interesting feature of them is allowing planar insertion of any aromatic chromophore between the base pairs of the DNA, the process to be commonly known as, Intercalation [58,59].

1.4 Modes of Drug-DNA Interactions

While drug interacts with DNA, it binds itself to the vicinity of the DNA macromolecule. This binding of drug molecule to the DNA macromolecule occurs in mainly two ways viz., covalent and non-covalent binding [106]. Non-covalent interactions can be further classified into three categories, viz., minor groove binding, major groove binding and intercalation. Other modes of drug-DNA interactions are also available depending upon the choice of preferential binding site by the DNA and restrictions imposed upon the DNA macromolecule due to the effect of various underlying interaction forces. Following **figure 1.8** represents a flow chart which gives us a brief idea of the different drug-DNA interaction modes.

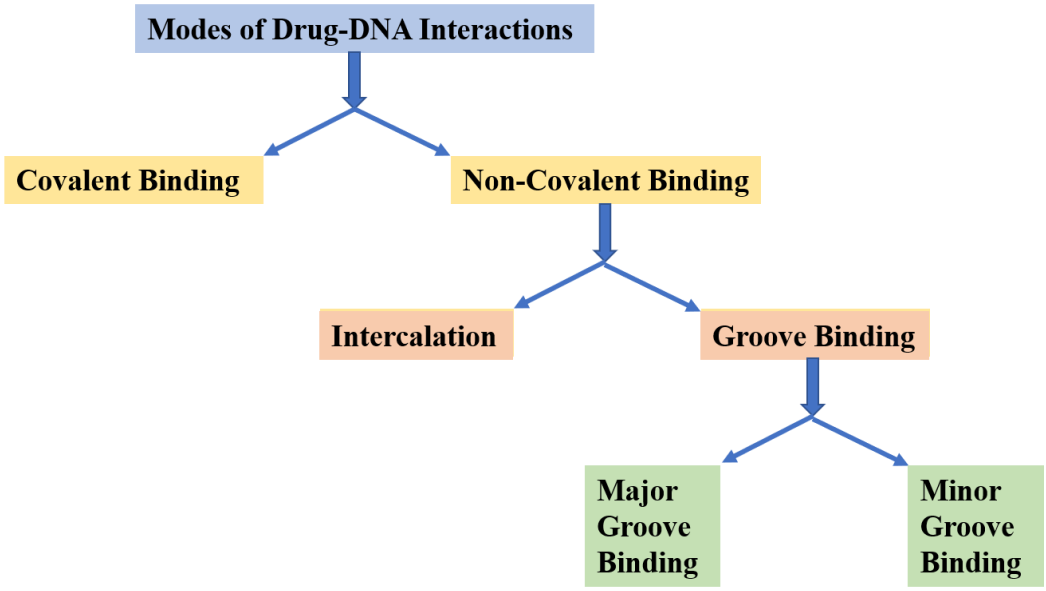


Figure 1.8: Figure representing different modes of drug-DNA interactions

Each of the above-mentioned binding modes are discussed in brief as follows:

1.4.1 Covalent Binding

Covalent binding of drugs with DNA is observed to be irreversible in nature and eventually leads to cell death. Covalent binding of drugs to DNA occurs in two ways, viz., either they bind via inter- or intra-strand cross linking or through DNA alkylation; covalent binding involves high binding strengths [107,108]. DNA covalent binders are often called DNA alkylating agents also, as they may attach an alkyl group to the DNA and thus inhibit the cancerous cell, leading to cell death. DNA alkylation involves attachment of alkyl groups from drug molecule to DNA leading to cure of various types of cancers. Alkylating agents belong to one of the most significant class of anticancer drugs. These alkylating agents have a methyl or alkyl (C_nH_{2n+1}) group attached to the molecules. Some common DNA alkylating agents are: Carmustine (BCNU), Chlorambucil, Cyclophosphamide, cis-Platin, Dacarbazine, Fotemustine,

Lomustine (CCNU), N-methyl-N'-nitro-N-nitrosoguanidiene (MNNG), Nimustine (ACNU), Temozolomide (TMZ) [110-115]. **Figure 1.9** shown below illustrates some of the few DNA alkylating agents. DNA alkylating agents prefer to react to N7 of guanine and N3 of adenine, and in this way base pairing of the DNA gets inhibited and thus leading to miscoding of DNA and consequently inhibiting the disease.

Alkylating agents interact with DNA through three mechanisms. In the first mechanism, the alkylating agent which has to bind to DNA attaches an alkyl group to DNA and therefore rendering the DNA into fragments. This results in release of DNA repairing enzymes it their attempt to replace the alkylated bases of the DNA. In the second mechanism, the alkylating agent forms cross linking with the DNA and thus leading to DNA damage. In this process, the two DNA bases are linked to each other by alkylating agents that has two DNA-binding sites. Cross linking prevents DNA from transcription. And in the third process, alkylating agents lead to mutations due to mis-pairing of the nucleotides [109].

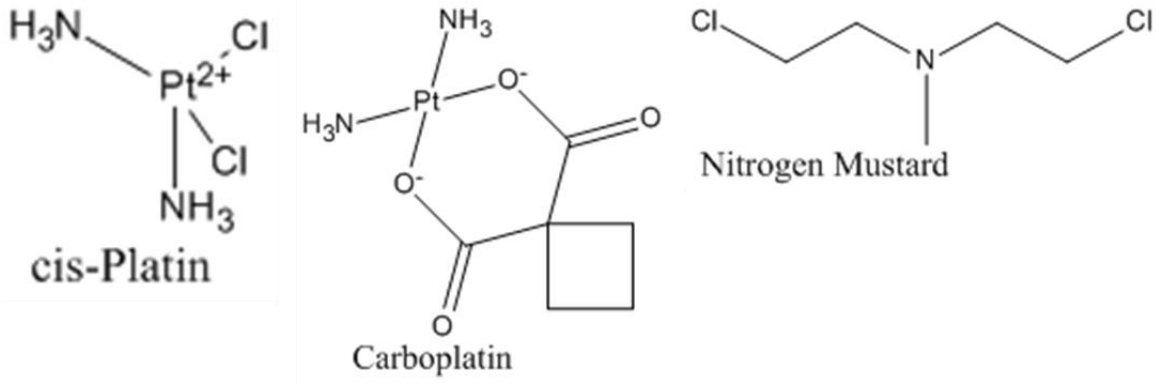


Figure 1.9: Figure showing some common DNA covalent binding agents

1.4.2 Non-Covalent Binding

Non-covalent interactions, viz., hydrogen bonding and base stacking interactions, tend to determine the conformational structure of the biological molecules (nucleic acids and proteins). It is certainly known through years of research that the hydrogen bonding is essential for the specificity of base pairing whereas interactions due to π-π stacking occur between the planar aromatic rings of the nucleobases, and both contribute significantly in the final stability attained by the nucleic acid's structures [117]. Though, individually each of these interacting forces is insignificant but their combined effect produces considerable stabilizing effects. Non-covalent binding in the DNA also tends to rupture the DNA strands, may tend to change the torsional angles and interrupt the protein-DNA interactions. Non-covalent interactions in DNA take place via three modes, viz., Major groove binding, minor groove binding and intercalation. Each of them is discussed as follows:

1.4.2.1 Minor Groove Binding

DNA minor groove binders usually consist of aromatic rings (rings involving benzene) linked via covalent bonds through sigma bonds. Small therapeutic molecules usually bind themselves to the nucleic acid bases via hydrogen bonds with N3 of adenine and O2 of thymine, whereas minor groove binders prefer binding at AT-rich region of the DNA. The reason behind the choice of this preferential binding in addition to the designated propensity for electronegative pockets of numerous AT-rich sites is due to better Van der Waals contact between the drug and the groove regions and also the steric hindrance offered is lower in latter, presented by the C2 amino acid of the guanine base [119-121].

Sequence specific DNA binding proteins prefer binding in the vicinity of major groove of the DNA as it offers numerous positions preferring the formation of hydrogen bond, and thus adding to complex stability and sequence specificity. However, binding of small molecules and proteins to the minor groove of the DNA usually depends upon the hydration properties of the available minor grooves, the former prefers binding to the AT-rich site of the macromolecule due to more prevalent water ordering. Thus, the driving force operating behind minor groove binding is the very large entropy change during the release of ordered water molecules, although the enthalpy being unfavorable [118, 122-124]. Minor groove binding involves higher binding affinity and greater extent of sequence specificity than any other binding modes. Minor groove has been demonstrated as a preferential binding site for neutral, mono-charged and multi-charged ligands [125].

Figure 1.10 shown below represents the both covalent and non-covalent minor groove binding modes of drug with DNA:

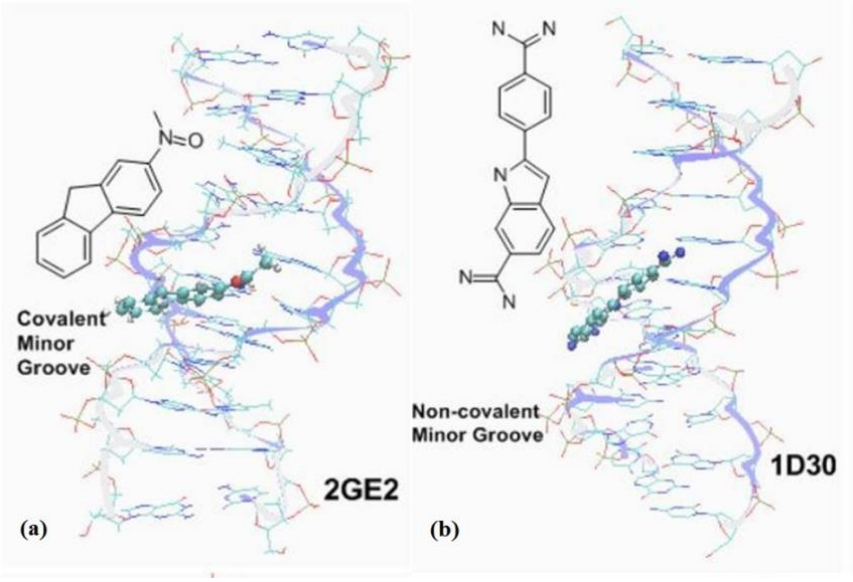


Figure 1.10: Figure depicting the (a) covalent and (b) non-covalent minor groove bindings

There are several factors that govern the selectivity of minor groove binders towards the AT-rich regions of the DNA, viz., firstly the electrostatic potential of the AT-rich region is observed to be higher than that of the GC-rich region and secondly the structural differences between the groove widths, i.e., the AT-rich site is narrower and deeper than GC-rich site, this also favors the minor groove binders to AT-rich sites [126]. Some of the crucial minor groove binders having significant therapeutic importance are shown below in **figure 1.11**.

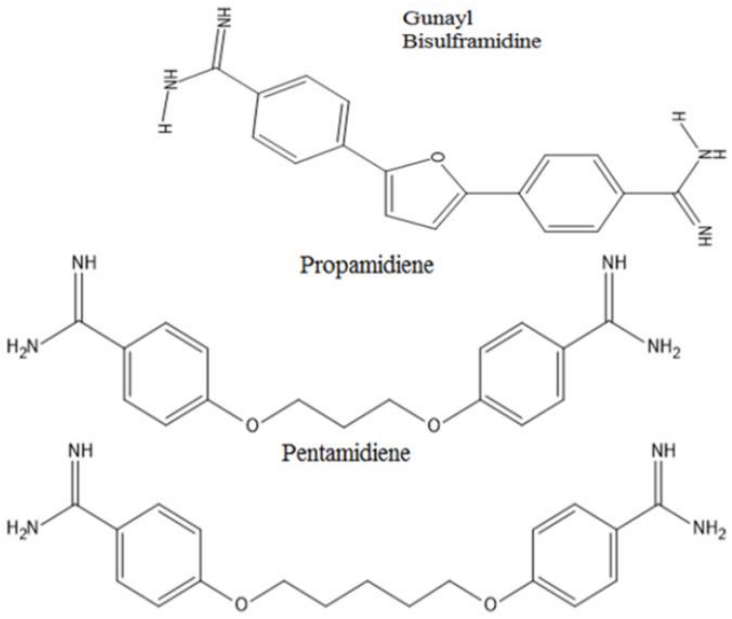


Figure 1.11: Figure showing some common DNA minor groove binders

1.4.2.2 Major Groove Binding

The groove widths for major groove are wider than that of minor groove, as a matter of statistics, the groove widths for B-form of DNA are 11.7 Å and 5.7 Å respectively. This widened groove width attracts DNA-interacting proteins as binding site. Proteins interact by forming a number of hydrogen bonds as donors and acceptors supplied by major groove residues [139-141]. It is the job of major groove binding protein to block the access of other proteins that recognize the similar binding site as

their preferential binding site. **Figure 1.12** shown below depicts the binding of drug in the major groove of the DNA. This is attained through sequence specificity and sequence selectivity [142].

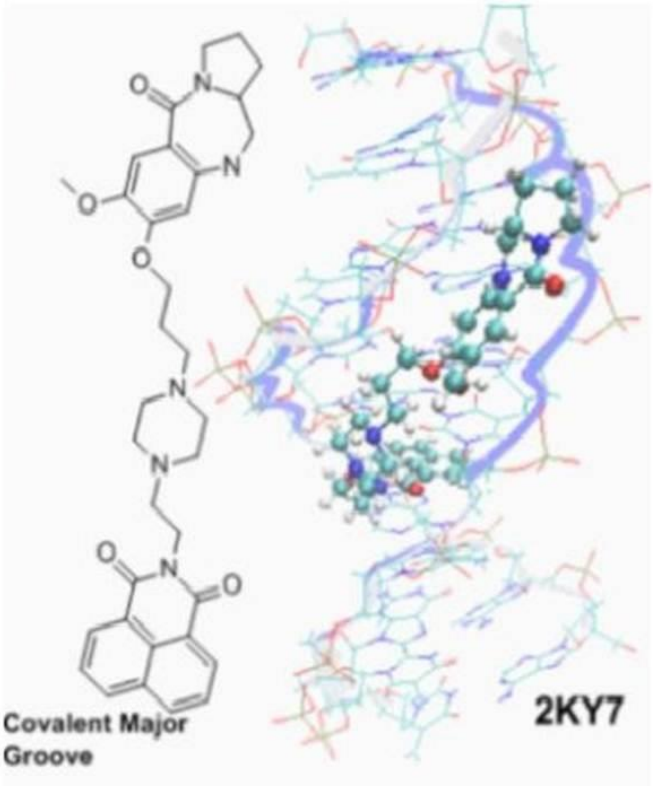


Figure 1.12: Figure showing covalent major groove DNA binding mode

DNA duplexes, which are composed of ring structures (polypurine-polypyrimidine), such sequences are read by oligomers and they tend to form hydrogen bonds with the nucleic acid base pairs of the purine strands. These are termed to be triplex-forming oligonucleotides (TFOs). Another format of major groove binding can be achieved by peptide nucleic acids (PNAs) [143,144]. **Figure 1.13** representing some important major groove binders are shown below.

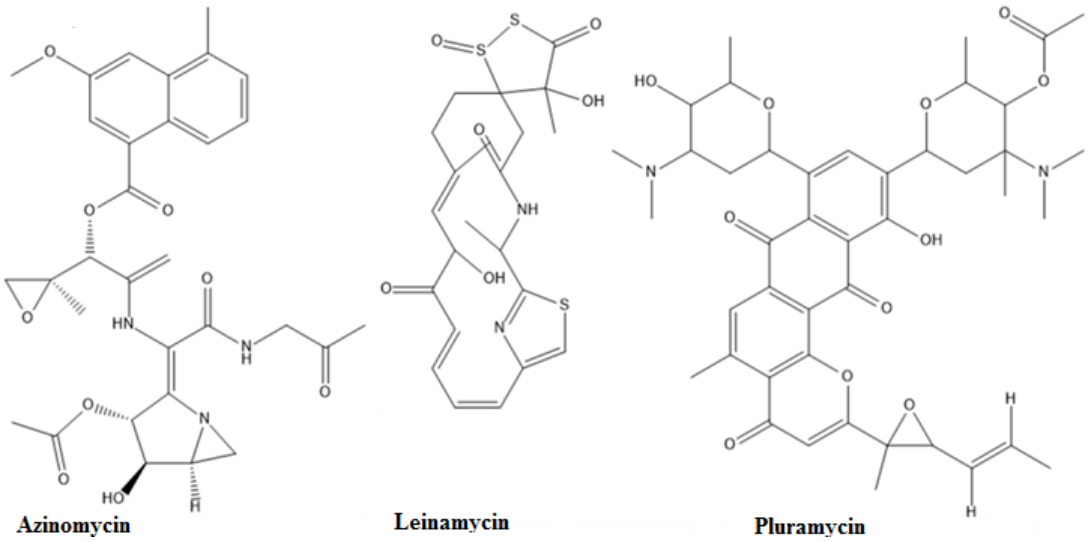


Figure 1.13: Figure showing some common DNA major groove binders

1.4.2.3 Intercalation

Intercalation was first termed by Lerman [127]. He explained that during intercalation the drug molecules are held perpendicularly to the nucleic acids base pairs of the DNA strands. However, this does not render the hydrogen bonds formed between the DNA base pairs and eventually the structural conformation of the DNA remains intact during intercalation. But this causes the distortion in the sugar contained phosphate backbones and hence distorts the pitch of the DNA [128]. During intercalation, the drug molecules bind via stacking non-covalently between two adjacent DNA base pairs. Generally, planar aromatic drug molecules fall into the category of DNA intercalators. Interaction between the drug molecules and DNA base pairs takes place during intercalation via means of π-π stacking. The stability while intercalation is attained via, van Der Waal's force, hydrophobic forces, stacking or charge transfer mechanisms and most important is hydrogen bonding, all these factors tend to stabilize the system, eventually [129].

Intercalation also leads to some structural anomalies in DNA double helix, viz., lengthening, unwinding, or stiffening of the DNA double helix. Data-wise, intercalation leads to change in the DNA torsional angles of the phosphate backbones in a way to create a vacancy for the incoming drug molecule, this tends to separate the two base pairs by lengthening of DNA approximately by 3.4 Å and decreases the twists in the helix by unwinding the DNA near the vicinity of the incoming drug molecule by 36° per base pair [130]. Acridine, Actinomycin D, Adriamycin, cis-Platin, Echinomycin, Ethidium, Noglamycin, Propidium, Trostin A, are some common DNA intercalating drugs and are shown in **figure 1.14**, below [131-134].

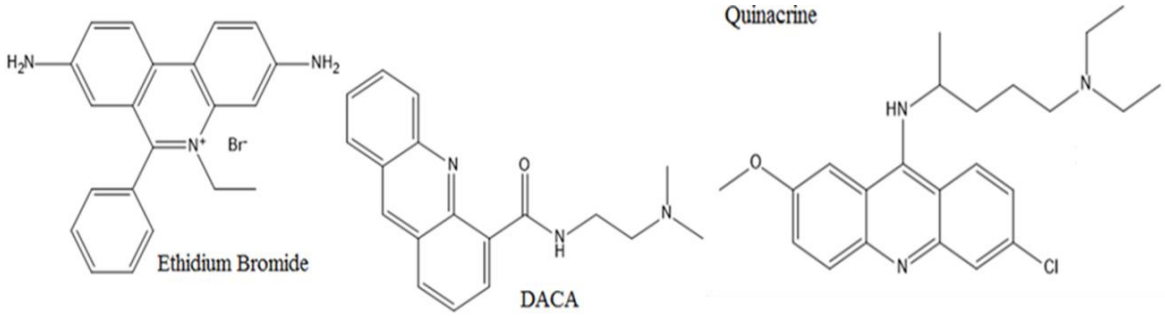


Figure 1.14: Figure showing some common DNA intercalators

Once the intercalation of a drug molecule occurs in the vicinity of DNA, then it hinders the access of another molecule ready to intercalate at the next intercalation site, this phenomenon is termed as "Neighbor Exclusion Principle" and the reason behind this is that due to intercalation, deep significant structural deformations occurs in the nucleotide secondary structure of the DNA [116, 135-138]. There are a few intercalating molecules without bulky groups attached to them, and therefore they can intercalate to the DNA without getting their significant part either into major or minor groove of the DNA, such intercalators are termed as, Classical intercalators [113].

And the ones which get their significant part inserted either into the major or minor groove of the DNA are termed as, threading intercalators [131]. In case of threading intercalation, the bulky group usually contains an aromatic ring and tends to bind itself to minor or major groove of the DNA [135,136]. **Figure 1.15** depicts the two DNA intercalation modes.

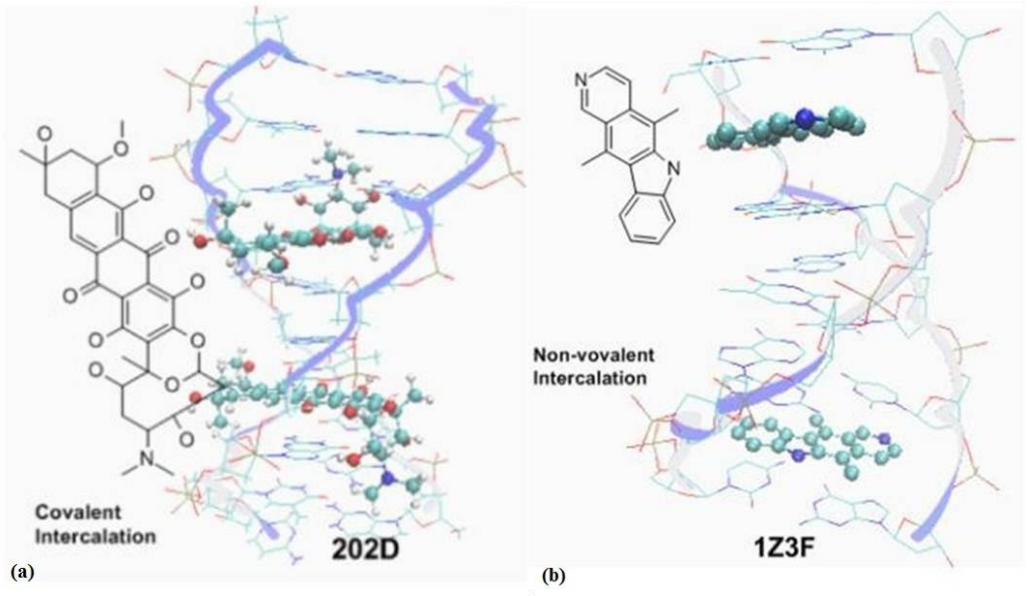


Figure 1.15: Figure showing (a) covalent DNA intercalation and (b) non-covalent DNA intercalation

1.5 Experimental Techniques in Studies of Drug-DNA Interactions

Experimental techniques play a very crucial role in the studies of drug-DNA interactions [195]. Thermodynamic study provides 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 acids (DNA/RNA) are discussed below:

Circular dichroism (CD) spectroscopy holds significant importance in the studies of drug-DNA interactions. This technique is used to detect whether a drug is groove binding or intercalator. Various other binding parameters along with the binding site can also be calculated by using circular dichroism [196].

Thermal stability is used to study the role of stabilizing or destabilizing effects of a drug on the DNA transition. This can be calculated by monitoring the variations in UV-absorption spectra as a function of temperature [197].

Viscometric Titration is used to understand the conformational changes occurring in DNA strands as a result of drug-DNA interactions.

Electric linear dichroism gives significant information regarding the preferential binding mode of the drug with DNA and also informs about base pair preferences of the nucleic acid during the drug binding [197-200].

X-ray crystallography requires a good quality of the single crystal of drug-DNA complex for crystallographic analysis [201]. However, it is very difficult to obtain a good quality crystal of compounds having high molecular weights such as, drug-DNA complexes. The major disadvantage of this technique is the selection of a particular conformational form or structure of the macromolecule due to the influence of crystal packing forces and underlying local ionic strengths.

Nuclear Magnetic Resonance (NMR) is very useful for the characterization of drug-DNA interactions at molecular levels. The atomic nuclei which are available for the study of DNA are ($_{1}$H, $_{13}$C, $_{15}$N and $_{31}$P); among them, $_{1}$H is the very common for NMR studies, but NMR of $_{31}$P is useful for studying the effect of binding of ligands to the phosphate groups of DNAs [201].

Figure 1.16 shows various computational and experimental tools and methodologies used in the studies of drug-DNA interactions.

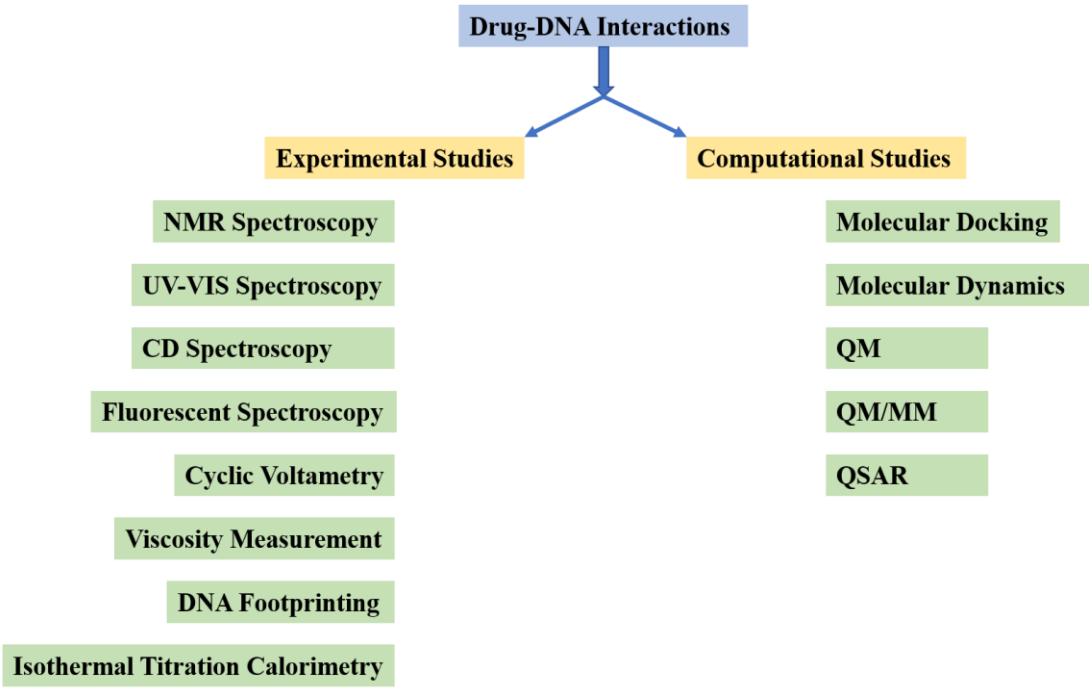


Figure 1.16: Figure representing different methods used to study drug-DNA interactions

1.6 Thermodynamic Data for Drugs as obtained from Previous Researches

Various research groups have been working in the field of drug-DNA interactions via experimental and computational techniques and have reported the obtained thermodynamic data in terms of various parameters, viz., K, ΔG, ΔT_m, etc. [145-194]. The thermodynamic data as obtained from their literature survey, along with their corresponding binding modes with DNA; can be summarized in the **table 1.2** given below:

Table 1.2: Table representing various drug-DNA interaction data as obtained through literature

S. No.	Ligand	DNA Sequence	Exp. Data	Binding Mode
1	Benzidine	CT-DNA	$K=6.2\pm0.4x10^{3}\ M^{-1}$	I
2	Netropsin	A3T3	ΔG=-7.7 kcal/mole	m
		CT-DNA	ΔG=-8.8 kcal/mole	
3	Propamidine	A3T3	ΔG=-7.0 kcal/mole	m
		CT-DNA	ΔG=-8.2 kcal/mole	
4	Berenil	A3T3	ΔG=-8.0 kcal/mole	m
		CT-DNA	ΔG=-8.6 kcal/mole	
5	Distamycin	A3T3	ΔG=-10.5 kcal/mole	m
6	Captopril	CT-DNA	$K=6.2\pm0.4x10^{3}\ M^{-1}$	m
7	Doxorubicin	CT-DNA	ΔG=-8.9 kcal/mole	I
8	Propidium	CT-DNA	ΔG=-7.5 kcal/mole	I
9	Daunorubicin	CT-DNA	ΔG=-7.9 kcal/mole	I
10	Ethidium	CT-DNA	ΔG=-6.7 kcal/mole	I
11	Chartreusim	CT-DNA	ΔG=-7.4 kcal/mole	I
12	Echinomycin	CT-DNA	ΔG=-7.5 kcal/mole	B
13	Dapi	AATT	$K=1.2x10^{5}\ M^{-1}$	I
14	Ditercalinium	CGCG	$K=5.0x10^{8}\ M^{-1}$	B
15	Mesalamine	CT-DNA	ΔG=-4.2 kcal/mole	I
16	Doxorubicin	1R2L DNA	$K=0.17x10^{5}\ M^{-1}$	I
17	Epirubicin	1R2L DNA	$K=0.14x10^{5}\ M^{-1}$	G
18	Daunorubicin	1R2L DNA	$K=1.80x10^{5}\ M^{-1}$	I
19	Cisplatin	1R2L DNA	$K=0.41x10^{5}\ M^{-1}$	I
20	Flourouracil	1R2L DNA	$K=8.7x10^{5}\ M^{-1}$	G
21	Carboplatin	1R2L DNA	$K=2.41x10^{5}\ M^{-1}$	
22	Etopside	1R2L DNA	$K=1.4x10^{3}\ M^{-1}$	G

23	Cyclophosphamide	1R2L DNA	$K=4.21x10^{5}\ M^{-1}$	
24	Dactinomycin	1R2L DNA	$K=3.61x10^{5}\ M^{-1}$	I
25	Mitoxantrone	1R2L DNA	$K=61.91x10^{5}\ M^{-1}$	I
26	Chloridazon	CT-DNA	ΔG=-22.63 kcal/mole	I
27	Pirenzepine	CT-DNA	$K=1.57x10^{3}\ M^{-1}$	G
28	Distamycin	DICKERSON POLY[d(AT)]	ΔG=-44.77 kcal/mole ΔG=-31.25 kcal/mole	m
29	Chlorambucil	CT-DNA	$K=2.0x10^{7}\ M^{-1}$	M
30	Emodine	CT-DNA	$K=5.559x10^{3}\ M^{-1}$	I
31	Mitoxantrone	CT-DNA	$K=3.38x10^{5}\ M^{-1}$	I
32	Ciprofloxacine	HUMAN TELOMERIC DNA	$K=9.62x10^{4}\ M^{-1}$	G
33	Daphnetin	CT-DNA	ΔG=-25.19 kcal/mole	I
34	Thiabendazole	CT-DNA	ΔG=-21.68 kcal/mole	I
35	Troxerutin	CT-DNA	$K=1.75x10^{3}\ M^{-1}$	M
36	Pregabalin	CT-DNA	$K=5.6x10^{3}M^{-1}$	I
37	Nimustein	CT-DNA	$K=3.27x10^{3}\ M^{-1}$	I
38	Querceitn	Ds-DNA	$K=3.56x10^{3}\ M^{-1}$	I
39	Tomoxifen	CT-DNA	$K=8.15x10^{4}\ M^{-1}$	I
40	Nitrofurantion	CT-DNA	$K=8.22\pm0.05x10^{6}\ M^{-1}$	I
41	Farrerol	CT-DNA	ΔG=-26.16±0.81 kcal/mole	I

42	Levetiracetam	CT-DNA	$K=4.9\pm0.2x10^{3}\ M^{-1}$	M
43	Idarubicin	Ds-DNA	$K=2.1x10^{4}\ M^{-1}$	I
44	Ticlopidine	CT-DNA	$K=4.91x10^{3}\ M^{-1}$	G
45	Citral	CT-DNA	$K=4.96x10^{3}\ M^{-1}$	I
46	Chelerythrine	Poly(dG).poly(dC) Poly(dG-dC).poly(dC-dG) Poly(dA).poly(dT) Poly(dA-dT).poly(dA-dT)	$K=5.25x10^{-6}M^{-1}$ $K=2.17x10^{-6}\ M^{-1}$ $K=1.23x10^{-6}\ M^{-1}$ $K=1.60x10^{-6}\ M^{-1}$	G
47	Resistomycin	CT-DNA	$K=1.095x10^{4}\ M^{-1}$	G/I
48	Epirubicin	Fish Sperm DNA	$K=3.8x10^{5}\ M^{-1}$	I
49	Sodium benzoate	CT-DNA	$\Delta G=-28.34\pm0.40$ kcal/mole	I
50	Carboplatin	CT-DNA		
51	Sinafloxacin	CT-DNA	$K=2.33x10^{2}\ M^{-1}$	G
52	Idarubicin	Ds-DNA	$K=5.14x10^{5}\ M^{-1}$	I
53	Elsculetin	CT-DNA	$K=1.84x10^{4}\ M^{-1}$	m
54	Methotrexate	CT-DNA	$K=1.0x10^{3}\ M^{-1}$	G
55	Olanzapine	CT-DNA	$K=2.0x10^{3}\ M^{-1}$	m
56	Carbaryl	CT-DNA	$K=8.68x10^{3}\ M^{-1}$	I
57	Tau-fluvalinate	CT-DNA	$\Delta G=-24.56$ kcal/mole	m
58	Flumethrin	CT-DNA	$\Delta G=-24.80$ kcal/mole	m

59	Methyldopa	CT-DNA	ΔG=-23.57 kcal/mole	m
60	Coumarin	CT-DNA	$K=1.69x10^{11}\ M^{-1}$	m
61	Aspirin	CT-DNA	ΔT_m=6.4˚C	I
62	Diflunisal	CT-DNA	ΔT_m=2.3˚C	G
63	Ibuprofen	CT-DNA	ΔG=-6.96 kcal/mole	I
64	Eugenol	Salmon Sperm DNA	$K=7.28x10^{3}\ M^{-1}$	I

Intercalator (I)

Bisintercalator (B)

Groove Binder (G)

Major Groove Binder (M)

Minor Groove Binder (m)

1.7 Web resources for Drugs and related data

Computational studies of drug-DNA interaction involve a huge number of web-based resources for deposited data. **Table 1.3** represents the list of some important and widely used such web-based resources [60-91].

Table 1.3: List of web-based resources for studying drug-DNA interactions

S. No.	Name	Description	Website
1	BindingDB	Database focusing upon experimental binding affinities between small molecules and macromolecules	http://www.bindingdb.org/bind/index.jsp
2	BioGRID	General repository for interaction dataset for protein, genetic and chemical interaction for humans	https://thebiogrid.org/
3	canSAR	Knowledgebase chemical and pharmacological database	https://cansarblack.icr.ac.uk/

4	CarlsbadOne	Confederated database of compounds associated with protein tarets	http://carlsbad.health.unm.edu/wp/
5	ChEMBL	Manually curate database of drug-like molecules	https://www.ebi.ac.uk/chembl/
6	DGIdb	Database for information about drug-gene interaction	http://www.dgidb.org/
7	DINIES	Helps in predicting potential interaction between drug molecules and target proteins	https://www.genome.jp/tools/dinies/
8	DisGeNET	Contains information about gene-disease associations and variant – disease associations	https://www.disgenet.org/
9	DrugBank	Web resource containing detailed drug data and drug-target information	https://www.drugbank.ca/
10	DrugCentral	Contains information about active ingredients of chemical entities and pharma products, including mode of action of drugs	http://drugcentral.org/
11	DTO	Interactive visualization for drug-target organization	http://drugtargetontology.org/
12	ECOdrug	To identify human drug targets and explore predictions for drug targets	http://www.ecodrug.org/
13	GUILDify	Characterization of genes through biological data	http://aleph.upf.edu/guildify2/
14	HitPick	Helps analysis of chemical screenings and then generates hits for predicting targets	http://mips.helmholtz-muenchen.de/hitpick/cgi-bin/index.cgi?content=targetPrediction.html
15	iDrug-Target	Webserver for predicting drug-target interactions	http://www.jci-bioinfo.cn/iDrug-Target/

16	idTarget	Identifying targets for small molecules	http://idtarget.rcas.sinica.edu.tw/index.php
17	IUPHAR/BPS	One stop shop portal to pharmacological information	http://www.guidetopharmacology.org/
18	OpenTargets	To find new targets for drug discovery	https://www.targetvalidation.org/
19	PASS Online	Predicts biological activity, mechanism of action, pharma effects, adverse effects, metabolism with enzymes, etc. for drugs	http://www.pharmaexpert.ru/passonline/index.php
20	Pharm Mapper	Pharmacophore matching platform for potential target identification	http://lilab-ecust.cn/pharmmapper/
21	Pharos	Database for druggable genome for most common drug-targated protein family	https://pharos.nih.gov/
22	PhID	Provides visual relationship about diseases, drugs, targets, genes, pathways and side-effects	http://phid.ditad.org/
23	PPB	Predicts possible targets for small molecules	http://gdbtools.unibe.ch:8080/PPB/index.html
24	PROMISCOUS	Exhaustive resource for protein-protein & drug-protein interactions	http://bioinformatics.charite.de/promiscuous/
25	SLAP	For target prediction of drugs	http://cheminfov.informatics.indiana.edu:8080/slap/
26	SPiDER	Prediction of drug-equivalence relationship	http://modlabcadd.ethz.ch/software/spider/
27	SuperDRUG2	Resource for approved drugs	http://cheminfo.charite.de/superdrug2/
28	SuperTarget	Information about drug target relations	http://bioinformatics.charite.de/supertarget/

29	SwissTarget Prediction	To predict target for bioactive small molecules	http://www.swisstargetprediction.ch/
30	TargetMine	Identification of suitable targets for characterization	https://targetmine.mizuguchilab.org/
31	TargetNet	For predicting binding of multiple targets for a given molecule	http://targetnet.scbdd.com/home/index/
32	WITHDRAWN	A resource for discontinued drugs	http://cheminfo.charite.de/withdrawn/index.html

1.8 Tools to model Nucleic Acids (DNA)

Some very useful tools [92, 93, 99-104] that are often used to model nucleic acids (DNA) are listed below in **table 1.4**:

Table 1.4: List of tools used to model DNA & drug-DNA interactions

S. No.	Tool	Description
1	3DNA	To analyze, rebuild & visualize 3D structures of nucleic acids
2	AMBER	Generates canonical A- and B-duplex geometries of nucleic acids
3	CURVES	To analyze DNA structure
4	DNA Sequence to Structure	Generates canonical A- and B-duplex geometries of DNA and molecular dynamics averaged DNA structure
5	NucCGEN	Generates a curved or non-uniform helix
6	Nuparm	To analyze sequence dependent variations in nucleic acid double helices

1.9 Database repository entries in NDB

The number of entries of 3D structures of DNA, drug-DNA and DNA-protein complexes as deposited in NDB [105] are listed below in **table 1.5**.

Table 1.5: Number of entries in NDB

S. No.	Type of System	No. of NDB entries
1	DNA	6393
2	Drug-DNA	429
3	DNA minor groove binders	116
4	DNA major groove binders	24
5	DNA intercalators	139
6	DNA-protein complexes	4479
7	DNA covalent binders	15

References:

[1] Nicolaou, K.C., *Angewandte Chemie*., **126,** 9280-9292 (2014).

[2] Gershell, L., Atkins, J., *Nature Reviews Drug Discovery,* **2**, 321–327 (2003).

[3] Gawehn, E., Hiss, J.A., Brown, J.B., Schneider, G., *Expert Opinion on Drug Discovery*, **13**, 579-582 (2018).

[4] Bahuguna, A., Rawat, D.S., *Med. Res. Rev.*, **40**, 263-292 (2020).

[5] Seyhan, A.A., Carini, C., *J. Transl. Med.* **17**,114 (2019).

[6] David, B., Wolfender, J., Dias, D., *Phytochemistry Reviews*, **14** (2015).

[7] Baskin, I.I., Winkler, D., Tetko, I.V., *Expert Opinion on Drug Discovery*, **11**, 785-795 (2016).

[8] Bolis, G., Pace, L.D., Fabrocini, F., *Journal of Computer-Aided Molecular Design*, **5,** 617-328 (1991).

[9] Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., Blaschke, T., *Drug Discovery Today,* **23**, 1241-1250 (2018).

[10] Kononenko, I., *Artificial Intelligence in Medicine*, **23**, 89-109 (2001).

[11] Korotcov, A., Tkachenko, V., Russo, D.P., Ekins, S., *Mol. Pharmaceutics*, **14**, 4462−4475 (2017).

[12] Yuan., H., Ma., Q., Ye., L., Piao., G., *Molecules*, **21**, 559 (2016).

[13] Schlick, T., Molecular Modeling and Simulation: An Interdisciplinary Guide, 2nd edition, *Springer* (2010).

[14] Jacobson, K., Mouritsen, O.G., Anderson, R.G.W., *Nature Cell Biology*, **9,** 7-14 (2007).

[15] Saiz, L., Klein, M.L., *Acc. Chem. Res.*, **35**, 482-489 (2002).

[16] Herrero, M.A., *J. Math. Biol.*, **54**, 887–889 (2007).

[17] Wang JC. Nature reviews. Cellular roles of DNA topoisomerases: A molecular perspective. Molecular and Cellular Biology, **3**, 430-440 (2002)

[18] Tian, T., Chen, Y.Q., Wang, S.R., Zhou, X., *Chem.*, **4**, 1–31, (2018).

[19] Kallenbach, N. R. ed, Chemistry and Physics of DNA-Ligand Interactions, Adenine Press, Schenectady, NY, 1990.

[20] Baraldi, P.G., Bovero, A., Fruttarolo, F., Preti, D., Tabrizi, M.A., Pavani, M.G., Romagnoli, R., *Medicinal Research Reviews*, **24**, 475-528 (2004).

[21] Escudé, C., Sun, J.S., *Top. Curr. Chem.*, **253**, 109–148 (2005).

[22] Martínez, R., Chacón-García, L., *Current Medicinal Chemistry*, **12**, 127-151 (2005).

[23] Fox, K.R., Drug-DNA Interaction Protocols, 2nd Ed., Hanuma Press (2010).

[24] Meng, X.Y., Zhang, H.X., Mezei, M., Cui, M., *Curr Comput Aided Drug Des.*, **7**, 146–157, (2011).

[25] Hospital, A., Goñi, J.R., Orozco, M., Gelpí, J.L., Advances and Applications in Bioinformatics and Chemistry, **8**, 37–47 (2015)

[26] Senn, H.M., Thiel, W., *Angew. Chem. Int. Ed.*, **48**, 1198 – 1229 (2009)

[27] Senn, H.M., Thiel, W., *Current Opinion in Chemical Biology*, **11**, 182–187 (2007)

[28] Lodish, H., Berk, A., Kaiser, C.A., Krieger, M., Scott, M.P., Bretscher, A., Ploegh, H., Matsudaira, P., Molecular Cell Biology, W.H. Freeman and Company, 6th Ed. (2008).

[29] Klug, W.S., Cummings, M.R., Concepts of Genetics, 5th Ed., Pearson Publications (1997).

[30] Dahm, R., *Developmental Biology*, **278**, 274-288 (2005).

[31] Avery,O.T., Maclend, C., Mc Carty, M., *The Journal of Experimental Medicine*, **79**,137-158 (1944).

[32] Chargaff, E., *J Cell Physiol. Suppl.*, **38**, 41-59 (1951).

[33] Franklin, R., and R. G. Gosling, R.G., *Nature*, **171**, 740–741 (1953).

[34] Watson, J.D., Crick, F.H.C., *Nature*, **171**, 737-738 (1953).

[35] Watson, J.D., Crick, F.H.C., *Nature*, **171**, 964-967 (1953).

[36] Dickerson, R.E., Drew, H.R., Conner, B.N., Wing, R.M., Fratini, A.V., Kopka, M.L., *Science*, **216,** 475-85 (1982).

[37] Crick, F.H.C., *Nature*, **227**, 561–563 (1970).

[38] Chaires, J.B., *Biopoly.*, **44**, 201-215 (1997).

[39] Chaires, J.B., *Curr. Opin. Struc. Biol.*, **8**, 314-320 (1998).

[40] Chaires, J.B., *Annu. Rev. Biophys.*, **37**, 135-151 (2008).

[41] Hurley, L.H., *Nat. Rev. Cancer*, **2**, 188-200 (2002).

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

[43] Franklin, T.J., Snow, G.A., Biochemistry and Molecular Biology of Antimicrobial Drug Design, 6th Ed., Springer (2005).

[44] Pullman, B., *The Jerusalam Symposia on Quantum Chemistry and Biochemistry, Kluwer Academic Publishers*, **23** (1990).

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

[46] Mahadevi, A.S., Shastry, G.N., *Chem. Rev.,* **116**, 2775−2825 (2016).

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

[48] Seibel, G.L., Singh, U.C., Kollman, P.A., *Proc. Nat. Acad. Sci., U.S.A.*, **82**, 6537 (1985).

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

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

[51] Singh, S.B., Ajay, Wemmer, D.E., Kollman, P.A., *Proc. Nat. Acad. Sci., USA*, **91**, 7673 (1994).

[52] Mikheikin, A.L., Zhuze, A.L., Zasedatelev, A.S., *J. Biomol. Struc. & Dynam*., **19**, 175 (2001).

[53] Dashnau, J. L., Sharp, K. A., Vanderkooi, J. M., *J. Phys. Chem. B*, **109**, 24152−24159 (2005).

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

[55] Ray, L., Hillary, S.R.G., Michael, J.P., Michael, K.G., *Biophysical*, **80**, 140 (2010).

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

[57] Gellman, S.H., Haque, T.S., Newcomb, L.F., *Biophysics Journel*, **71**, 3523 (1996).

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

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

[60] Liu, T., Lin, Y., Wen, X., Jorissen, R.N., Gilson, M.K., *Nucl. Acids Res*., **35**, 198-201 (2007).

[61] Oughtred, R., Stark, C., Breitkreutz, B.J., Rust, J., Boucher, L., Chang, C., Kolas, N., O'Donnell, L., Leung, G., McAdam, R., Zhang, F., Dolma, S., Willems, A., Coulombe-Huntington, J., Chatr-Aryamontri, A., Dolinski, K., Tyers, M., *Nucl. Acids Res*., **47**, 529-541 (2019).

[62] Tym, J.E., Mitsopoulos, C., Coker, E.A., Razaz, P., Schierz, A.C., Antolin, A.A., Al-Lazikani, B., *Nucl. Acids Res*., **44**, 938-943 (2016).

[63] Mathias, S.L., Hines-Kay, J., Yang, J.J., Zahoransky-Kohalmi, G., Bologa, C.G., Ursu, O., Opera, T.I., *Database*, bat044(2013).

[64] Awale, M., Reymond, J.L., *J Cheminform*, **9,**11(2017).

[65] Griffith, M., Griffith, O. L., Coffman, A. C., Weible, J. V., McMichael, J. F., Spies, N. C.,Wilson, R. K., *Nature Methods*, **10**, 1209–1210 (2013).

[66] Yamanishi, Y., Kotera, M., Moriya, Y., Sawada, R., Kanehisa, M., Goto, S., *Nucl. Acids Res.*, **42**, W39–W45 (2014).

[67] Piñero, J., Bravo, À., Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., Furlong, L. I., *Nucl. Acids Res*., **45,** 833–839 (2016).

[68] Wishart, D. S., *Nucl. Acids Res*., **34,** 668–672 (2006).

[69] Ursu, O., Holmes, J., Knockel, J., Bologa, C. G., Yang, J. J., Mathias, S. L., Oprea, T. I., *Nucl. Acids Res*., **45**, 932–939 (2016).

[70] Lin, Y., Mehta, S., Küçük-McGinty, H., Turner, J. P., Vidovic, D., Forlin, M., Schürer, S. C., *Journal of Biomedical Semantics*, **8,** (2017).

[71] Verbruggen, B., Gunnarsson, L., Kristiansson, E., Österlund, T., Owen, S. F., Snape, J. R., Tyler, C. R., *Nucl. Acids Res*., **46**, D930–D936 (2017).

[72] Guney, E., Garcia-Garcia, J., Oliva, B., *Bioinformatics*, **30**, 1789–1790 (2014).

[73] Liu, X., Vogt, I., Haque, T., Campillos, M., *Bioinformatics*, **29**, 1910–1912 (2013).

[74] Xiao, X., Min, J.L., Lin, W.Z., Liu, Z., Cheng, X., Chou, K.C., *Journal of Biomolecular Structure and Dynamics*, **33**, 2221–2233 (2015).

[75] Wang, J.C., Chu, P.Y., Chen, C.M., Lin, J.H., *Nucl. Acids Res*., **40**, 393–399 (2012).

[76] Pawson, A.J., Sharman, J.L., Benson, H.E., Faccenda, E., Alexander, S.P., Buneman, O.P., Davenport, A.P., McGrath, J.C., Peters, J.A., Southan, C., Spedding, M., Yu, W., Harmar, A.J., NC-IUPHAR. *Nucl. Acids Res*., **42**, 1098–1106 (2013).

[77] Koscielny, G., An, P., Carvalho-Silva, D., Cham, J. A., Fumis, L., Gasparyan, R., Papa, E., *Nucl. Acids Res*., **45**, 985–994 (2016).

[78] Filimonov, D. A., Lagunin, A. A., Gloriozova, T. A., Rudik, A. V., Druzhilovskii, D. S., Pogodin, P. V., & Poroikov, V. V., *Chemistry of Heterocyclic Compounds*, **50**, 444–457 (2014).

[79] Liu, X., Ouyang, S., Yu, B., Liu, Y., Huang, K., Gong, J., Jiang, H., *Nucl. Acids Res*., **38**, 609–614 (2010).

[80] Nguyen, D.T., Mathias, S., Bologa, C., Brunak, S., Fernandez, N., Gaulton, A., Guha, R., *Nucl. Acids Res.*, **45**, 995–1002 (2016).

[81] Deng, Z., Tu, W., Deng, Z., Hu, Q.N., *Journal of Chemical Information and Modeling*, **57**, 2395–2400 (2017).

[82] Awale, M., Reymond, J.L., *Journal of Cheminformatics*, **9**, (2017).

[83] Von Eichborn, J., Murgueitio, M.S., Dunkel, M., Koerner, S., Bourne, P.E., Preissner, R., *Nucl. Acids Res.*, **39**, 1060–1066 (2010).

[84] Chen, B., Ding, Y., & Wild, D.J., *PLoS Computational Biology*, **8**, e1002574, (2012).

[85] Rodrigues, T., *Organic & Biomolecular Chemistry*, **15**, 9275–9282 (2017).

[86] Siramshetty, V.B., Eckert, O.A., Gohlke, B.O., Goede, A., Chen, Q., Devarakonda, P., Preissner, R., *Nucl. Acids Res.*, **46**, 1137–1143 (2017).

[87] Gunther, S., Kuhn, M., Dunkel, M., Campillos, M., Senger, C., Petsalaki, E., Preissner, R., *Nucl. Acids Res.*, **36**, 919–922 (2007).

[88] Byrne, R., Schneider, G., *Systems Chemical Biology*, 273–309 (2018).

[89] Chen, Y.A., Tripathi, L. P., Mizuguchi, K., *PLoS ONE*, **6**, e17844 (2011).

[90] Yao, Z.J., Dong, J., Che, Y.J., Zhu, M.F., Wen, M., Wang, N.N., Cao, D.S., *Journal of Computer-Aided Molecular Design*, **30**, 413–424 (2016).

[91] Siramshetty, V. B., Nickel, J., Omieczynski, C., Gohlke, B.O., Drwal, M. N., Preissner, R., *Nucl. Acids Res.*, **44**, 1080–1086 (2015).

[92] Lavery, R., Sklenar, H., *J. Biomol. Struc. & Dynam.*, **6**, 63 (1988).

[93] Lavery, R., Sklenar, H., *J. Biomol. Struc. & Dynam.*, **6**, 655 (1989).

[94] Bhattacharya, D., Bansal, M., *J. Biomol. Struc. &Dynam.*, **6**, 93 (1988).

[95] Bhattacharya, D., Bansal, M., *J. Biomol. Struc. & Dynam.*, **6**, 635 (1989).

[96] Bhattacharya, D., Bansal, M., *J. Biomol. Struc. & Dynam.*, **8**, 539 (1990).

[97] Bhattacharya, D., Bansal, M., *J. Biomol. Struc. &Dynam.*, **10**, 213 (1992).

[98] Bansal, M., Bhattacharya, D., Ravi, B., *Comput. Appl. Biosci.*, **11**, 8 (1995).

[99] Mukherjee, S., Bansal, M., Bhattacharyya, D., *Journal of Computer-Aided Molecular Design*, **20**, 629 (2006).

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

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

[102] 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).

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

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

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

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

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

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

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

[110] Kondo, N., Takahashi, A., Ono, K., Ohnishi, T.J., *Nucl. Acids Res*., 54353 (2010).

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

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

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

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

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

[116] 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).

[117] Manalo, M.N., Pérez, L.M., LiWang, A., *J. Am. Chem. Soc.*, **129,** 11298–11299 (2007).

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

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

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

[121] 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).

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

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

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

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

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

[127] Lerman, L.S., *Journal of Molecular Biology*, **3,** 18–30 (1961).

[128] Remers, W.A., The Chemistry of Antitumor Antibiotics, *Wiley,New York,*(1979)

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

[130] Pigram, W.J., Fuller, W., Hamilton, L.D., *Nature New Biology*, **5**, 17 (1972).

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

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

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

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

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

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

[137] Wilson,W.D., Tanious, F.A., Barton, H.J., Jones, R.L., Strekowski, L., Boykin D.W., *J. Am. Chem. Soc.*, **120**,10310 (1998).

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

[139] Schleif, R., *Science*, **11**, 82 (1998).

[140] Thuong, N.T., Helene, C., *Angew. Chem*., **32**, 666 (1993).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[166] 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).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[182] Yan, F., Guocai, L., Gurong, F., Yutian, W., *Analytical Sciences*, **25**, 1333 (2009).

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

[184] Tarique, S., Mohammed, A.H., Sayeed Ur, R., Hassan, M.I., Mohammad, T., *Mol. Biosyst.* **11,** 522-531 (2015).

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

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

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

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

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

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

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

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

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

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

[195] Nakamoto, K., Tsuboi, M., Strahan, G.D., Drug–DNA Interactions: Structures and Spectra, John Wiley & Sons, (2008).

[196] Singh, M. P., Lown, J. W., *Progress in Medicinal Chemistry*, **49,** 49–171 (1996).

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

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

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

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

[201] Gonzalez-Ruiz, V., Olives, A.I., Martin, M.A., Ribelles, P., Ramos, M.T., Menendez, J.C., *Biomedical Engineering, Trends, Research and Technologies,* **65,** (2011).

Chapter-2

Methodology

Chapter 2

Methodology

This chapter provides an overview about the different computational techniques that are used to understand the structure, dynamics and other physical and chemical properties of atomic and molecular systems.

2.1 Quantum Mechanics

A molecule is composed of atoms and atoms are made up of electrons, protons & neutrons. Protons & neutrons reside inside the nucleus and there are electrons revolving around the nucleus in various energy levels also called orbits or shells. This is most basic idea of the arrangement of subatomic particles inside an atom. This arrangement of the subatomic particles inside the atom in a definite manner leads them to hold and exhibit unique physical and chemical properties, viz., stable geometry, charge distribution, dipole moment, vibrational frequency, etc. These sets of unique physical and chemical properties can be calculated precisely using quantum mechanical formulations for a given set of atoms or molecules [1].

According to quantum mechanics (QM), all information regarding a molecular system is held by its wave function (ψ) and can be obtained by solving the well-known Schrödinger wave equation [2]:

$$H\Psi = E\Psi \tag{2.1}$$

Equation 2.1 represents the time independent form of the Schrödinger wave equation, where H represents the Hamiltonian Operator and E is the energy eigenvalue

corresponding to the Hamiltonian Operator H. Ψ is a well-behaved mathematical function also called the wave function which is square integrable [3] i.e., whose square represents the probability density.

The wave function of the molecule represents its nuclear as well as its electronic motions together; the electronic wave function is separated from total wave function using Born Oppenheimer approximation [4-6].

When the nuclei of two or more atoms are fixed at a particular distance, then their electronic wave function sufficiently provides all the physical and chemical properties of the molecule. But for a multi electron system, the electronic part of Hamiltonian operator of Schrödinger equation is given by:

$$H_e = -\sum_P \frac{1}{2}\nabla_P^2 - \sum_A \sum_P Z_A r_{AP}^{-1} + \sum_{p<q} r_{pq}^{-1} \qquad (2.2)$$

The first term in **equation 2.2** represents the kinetic energy, second term represents the potential energy whereas the third term represents the raise in potential energy due to inter-electron interactions. Thus, the modified Schrödinger wave equation for 'n' electron system [7] is given by:

$$H_e(1,2,....,n)\Psi_e(1,2,....,n) = E_e\Psi_e(1,2,....,n) \qquad (2.3)$$

Where, Ψ_e is electronic part of the total wave function. Due to the presence of several inter-electronic repulsion terms, the Schrödinger equation for a single atom is inseparable and therefore we obtain the solution Ψ_0 by neglecting these inter-electronic repulsion terms. The wave function Ψ_0 for 'n' electrons is given by the product of 'n' single electronic wave functions. This product wave function is known as Hartree Product of the wave function [8] and is given as:

$$\Psi_0 = \Psi_1(r_1, \theta_1, \varphi_1)\Psi_2(r_2, \theta_2, \varphi_2).....\Psi_n(r_n, \theta_n, \varphi_n) \quad (2.4)$$

The electrons are fermions and thus they obey half spin Fermi-Dirac statistics [9]. One electron wave function are the molecular orbitals which are the products of the spatial orbitals and the spin functions (α or β) [10]. In order to satisfy Pauli's exclusion principle, the wave function must be anti-symmetric in nature [11]. The anti-symmetry wave functions may be derived from Slater Determinants [12]. The spin orbital Slater determinant for 'n' electrons is given by:

$$\Psi = \frac{1}{\sqrt{N}} \begin{vmatrix} \varphi_1^{(1)}\alpha^{(1)} & \varphi_2^{(1)}\beta^{(1)}..... & \varphi_n^{(1)}\beta^{(1)} \\ \varphi_1^{(2)}\alpha^{(2)} & \varphi_2^{(2)}\beta^{(2)}..... & \varphi_n^{(2)}\beta^{(2)} \\ \varphi_1^{(n)}\alpha^{(n)} & \varphi_1^{(n)}\alpha^{(n)}..... & \phi_n^{(n)}\beta^{(2)} \end{vmatrix} \quad (2.5)$$

The solution to eq. 2.5 can be obtained by using perturbation over the solution Ψ_0 and a trial wave function is constructed with the help of Slater determinant by applying the Variational principle. The equation of energy of the system given below is written in form of Dirac notation [13-15] given by:

$$E_e = \frac{\left\langle \Psi'_i \mid H_e \mid \Psi'_i \right\rangle}{\left\langle \Psi'_i \middle| \Psi'_j \right\rangle} \quad (2.6)$$

2.2 Hartree Self -Consistent Field Method

The solution of the 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 [16,17]. The Hamiltonian is given as:

$$H_e = \sum_{p}^{N} h + \sum_{i=1}^{N}\sum_{j>i}^{N} g_{ij} + V_{nn} \quad (2.7)$$

where,

$$h_p = -\frac{1}{2}\nabla^2 - \sum_a \frac{Z_a}{|R_a - r_p|} \tag{2.8}$$

and

$$g_{ij} = \frac{1}{|r_i - r_j|} \tag{2.9}$$

Where one electron operator (h_i) quantifies the motion of the i[th] electron in field of all nuclei, (g_{ij}) is the two-electron operator quantifying the repulsion between two electrons while V_{nn} is to record the nuclear-nuclear interaction energy. Therefore, the energy can be finally expressed using the following sets of equations:

$$E = \sum_i^N \langle \varphi_i \mid h_i \mid \varphi_i \rangle + \frac{1}{2}\sum_{ij}^N \left(\langle \varphi_j \mid J_i \mid \varphi_j \rangle - \langle \varphi_j \mid K_i \mid \varphi_j \rangle \right) + V_{nn} \tag{2.10}$$

$$J_{12} = \left\langle \varphi_1^{(1)} \varphi_2^{(2)} \mid g_{12} \mid \varphi_1^{(1)} \varphi_2^{(2)} \right\rangle \tag{2.11}$$

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

$$K_{12} = \left\langle \varphi_1^2 \varphi_2^{(2)} \mid g_{12} \mid \varphi_2^{(1)} \varphi_1^{(2)} \right\rangle \tag{2.12}$$

The ‘K’ operator represents the exchange integral that has no classical analogue [17]. To determine the set of M.O. (Molecular Orbital) which has minimum energy or is stationary in respect to change of orbitals, variation is carried out in a manner that the MOs should remain orthogonal and normalized. This type of optimization is termed as constrained optimization and it can be achieved by implementing Lagrange’s multipliers [16, 19]. The Lagrange function is stationary with respect to the orbital variations:

$$L = E - \sum_{ij}^{N} \lambda_{ij} \left(\langle \varphi_i \mid \varphi_j \rangle - \delta_{ij} \right) \tag{2.13}$$

'L' is the Lagrange function. Variation of the Lagrange function is written as:

$$\delta L = \delta E - \sum_{ij}^{N} \left(\langle \delta\varphi_i \mid \varphi_i \rangle + \langle \varphi_i \mid \delta\varphi_i \rangle \right) = 0 \tag{2.14}$$

$$F_i = h_i + \sum_{j}^{N} \left(J_j - K_j \right) \tag{2.15}$$

Where, F_i is the Fock Operator; whereas operator J is the electron repulsion term, known as the Coulomb Operator while operator K is termed as the Exchange Operator. Hence, the Hartree Fock equation can be written as below [16, 20]:

$$F_i \phi_i = \varepsilon_i \phi_i \tag{2.16}$$

A set of functions that are the solutions of eq. 2.16 are called Self-Consistent Field Orbitals and are determined through an iterative method. The Hartree-Fock method is known as mean field approximation which considers the average electron-electron repulsion [21].

2.3 Electron Correlation

Electrons avoid confronting themselves by repelling each other. Such effect of electron repulsion (force between electrons) is also called electron correlation. Thus, the motion of electron is correlated and hence affected due the motion of the other electrons in the system. When the molecules are formed, there is a possibility for the existence of angular electron correlation around the bond direction. Within the Hartree-Fock formalism, the anti-symmetric wave function is approximated using a Slater determinant, which does not include this electron correlation (or Coulomb correlation). Therefore, the energy calculated with Hartree-Fock theory is different

from the exact energy of the system. This difference between these two energies is termed as the correlation energy and is given by following equation:

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

There has been a significant development in the fields of computational methods in order to accurately determine the contributions of various correlation terms and thus lead to form a broader category of approach called the post Hartree-Fock method. Many post-HF calculations methods are also available which include configuration interaction (CI), Moller-Plesset perturbation theory (MPn), coupled cluster theory (CC) and multi configurational self-consistent field (MCSCF).

2.4 Density Functional Theory

Density functional theory (DFT), in contrast to wave function formalism in quantum mechanics, describes the energy as a functional of the density of the electron cloud ($\rho(\Psi r)$) [22]. Kohn et al. derived a set of one-electron equations that enables us to calculate the electron density and eventually the total energy of the system [23]. This methodology has proven itself to be very successful in recent years as it is less computationally expensive while the results obtained through it show reasonable agreement with experiments [23]. Thus, similar to Hartree-Fock calculations, the total energy (E_{el}) of a DFT calculation setup is also split into different components, viz., a kinetic energy term, a term representing the electron-nucleus attractions, a term for the Coulomb interactions between the electrons and lastly an exchange-correlation term (E_{xc}), respectively [23,24]. The first three terms resemble the Hartree-Fock Hamiltonian shown in **equations 2.7, 2.8** & **2.9** above as a function of the nuclear coordinates (R), and the coordinates of the electrons (r), and is shown below:

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

(2.18)

Generally, Exchange-Correlation Functional (E_{xc}) is split into an exchange functional (E_X) and a correlation functional (E_C). The exchange functional essentially represents the interaction of spins of the two electrons in different orbitals, whereas the correlation functional represents the pairing energy of the two electrons in the same orbital [25,26]. The exchange energy is often estimated using the Slater exchange function given as:

$$E_X^{Slater} = -\frac{9}{4\alpha_{ex}}\left(\frac{3}{4\pi}\right)^{1/3}\sum_\gamma \int \left[\rho_1^\gamma(r_1)\right]^{4/3} dr_1 \qquad (2.19)$$

In the above **equation 2.19** (α_{ex}) is an exchange scale factor, which has the value equal to ⅔ for an electron gas system. The commonly used correlation energy functional (E_C^{VWN}) was named after Vosko, Wilk and Nusair [25] and represents the correlation energy per electron in an electron gas system $\varepsilon_C\left[\rho_1^\alpha \rho_1^\beta\right]$ with spin densities ρ_1^α & ρ_1^β and is given as:

$$E_C^{VWN} = \int \rho_1(r_1)\varepsilon_c\left[\rho_1^\alpha(r_1), \rho_1^\beta(r_1)\right]dr_1 \qquad (2.20)$$

The Local Density Approximation (LDA) is obtained through the combination of Slater exchange functional and Vosko-Wilk-Nusair correlation functional. However, to correct the non-local terms other (better and accurate) exchange and correlation functionals have now been developed. Two such netter and popular

approaches are named after Lee, Yang & Parr (i.e., LYP correlational functional) [26] and Perdew & Wang (i.e., PW91 correlational functional) [27].

A groundbreaking breakthrough regarding the applications of density functional theory to computational chemistry, in general, appeared when Becke developed the hybrid density functional procedures [28]. Becke benchmarked the DFT methods against a test set of experimentally pre-known ionization energies and electron & proton affinities with higher accuracies. Essentially the hybrid density functional method B3LYP now has the following form:

$$E_{XC}^{B3LYP} = AE_{X}^{Slater} + (1-A)E_{X}^{HF} + B\Delta E_{X}^{Becke} + E_{C}^{VWN} + C\Delta E_{C}^{LYP} \qquad (2.21)$$

In essence, the B3LYP method is not an *ab-initio* method. In strict sense, the term *ab-initio* means starting from scratch without prior knowledge of experiment. The B3LYP method including the other hybrid and non-hybrid DFT methods, have been tested to be extremely accurate and versatile for computational chemists. Although, in general, DFT lags in accuracy when compared to high-level *ab-initio* methods, such as coupled cluster methods, but their calculation speed when combined with reasonable accuracy makes them a very popular choice and a useful methodology.

Comparative calculations when performed to test the accuracy and speed of the three different techniques, i.e. Hartree-Fock theory, MP2 theory and DFT methods (using B3LYP) on complete test-set molecules using showed that B3LYP methodology gives a considerable improvement over Hartree-Fock as well as over MP2 calculations [29]. Therefore, it can be said that, DFT is a low-cost and accurate

result providing alternative than other quantum mechanical procedures and it is as such highly suitable to perform calculations on biochemical systems [30].

2.4.1 Basis Sets

A basis set is a mathematical representation of acceptable functions used to define the shape of the orbitals in an atom. To solve the Schrödinger equation, one has to optimize the wave function with the help of the Hamilton operator to find the corresponding energy eigenvalues. There is no straightforward method to obtain the wave functions for molecules with distinct levels of orbitals. To overcome this problem, an approach known as, the Linear Combination of Atomic Orbitals (LCAO) [31] was used to describe the molecular wave function through molecular orbitals. In this approach, the occupied atomic orbitals (χ) of all the atoms in the molecule are considered to create a set of molecular orbitals (φ_i) through certain fractions (c_{ri}) of all the atomic orbital components. Mathematically, it can be interpreted as following equation:

$$\varphi_i = \sum_r c_{ri}\chi_r \tag{2.22}$$

There are two broad classes of basis sets i.e., Slater Type Orbitals (STO's) [32,33] and Gaussian Type Orbitals (GTO's) [16,34]. A Slater Type Orbital for an s-type atomic orbital has the following form:

$$S(r) = N_s e^{-\xi r^2} \tag{2.23}$$

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

$$g(r) = N_g e^{-\xi r^2} \tag{2.24}$$

Where, N_g is the normalization constant

The smallest basis set, which represents a particular basis function for each type of occupied orbital in the separated atoms, is called a minimal basis set [35]. The Pople basis set [36,37] are another family of basis set and they are represented using the notation 6-31G. This means 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. The Pople basis sets are very popular choice for the simulation of organic molecules. These basis sets can be modified by adding one or two asterisks (*), such as 6-31G* or 6-31G**. A single asterisk means a set of *d*-primitives has been added to atoms other than hydrogen. 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. Similarly, 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. Diffuse functions are used for anions, which have larger electron density distributions. They are also used for describing interactions at long distances, such as, van der Waal's interactions. The effect of adding diffuse functions is usually to change the relative energies of the various

geometries associated with these systems. Basis sets with diffuse functions are called augmented basis sets [36,38].

Due to the constraints implemented in Variational principle, the electronic energy of an atom or a molecule approaches to its exact value closely with an increase in the number of basis functions used in during the calculation. Hence, it is beneficial to choose such a basis sets which has more basis functions than the number of basis functions in minimal basis set. The Double Zeta (DZ) basis sets consist of two minimal basis functions for the separated atoms. Since, the coefficients of hybrid density functional theory were benchmarked against experimental data using calculations through a double-ζ quality basis set, this type of basis set is generally sufficient in DFT calculations for geometry optimizations [39].

2.4.1.1 Popular Basis Sets

Following **table 2.1** provides a concise list of popular basis sets and their description:

Table 2.1: Popular basis sets with their descriptions

S. No.	Basis Set	Description
1	STO-3G	Minimal basis set: used to obtain more qualitative results on very large systems. Implemented when one cannot afford to use 3-21G.
2	3-21G	Split valence: consists of two sets of functions in the valence region to provide a more accurate representation of orbitals. implemented for very large molecules for which 6-31G(d) is computationally too expensive.

3	6-31G(d)	Adds polarization functions to all the heavy atoms present in the molecule. Implemented in the calculations of medium/large sized systems. This basis set uses the 6-component type functions.
4	6-311+G(d, p)	Triple zeta function: adds extra valence functions (three sizes of s- and p-functions)

2.5 Molecular Mechanics

In case of large molecules or biomolecules where the semi-empirical methods are less effective, and modeling the system via quantum mechanics is very difficult; there stands out another method referred to as the molecular mechanics (MM). MM is a set up of simple algebraic expressions to evaluate the total energy of compound and implementation of MM requires no need to compute a wave function or total energy or total electron density [18] for the system. Molecular mechanics (MM) is based on the simple classical model where the atoms are treated as hard spheres and bonds as springs. In MM, the energy of molecule is expressed as the energy due to the orientation (geometry) of the molecule, from a few specific interactions within the molecule. These interactions include the bonded terms such as bond stretching, bending and twisting and the non-bonded terms such as Van der Waals attractions or repulsions between the atoms that come close together, and the electrostatic interactions between partial charges in a molecule due to the presence of polar bonds [24-26]. Therefore, the total potential energy as calculated using MM can be written as the sum of the contributing energies of these two types of interactions, viz.:

$$E = E_{bonded} + E_{non\text{-}bonded} \tag{2.25}$$

$$E_{bonded} = E_{stretching} + E_{bending} + E_{improper} + E_{dihedral} \tag{2.26}$$

$$E_{non\text{-}bonded} = E_{elec} + E_{vdw} \quad (2.27)$$

The approximations adopted in MM calculations make the computation less expensive. Due to the speed of the MM calculations, large biomolecular systems, such as, protein and DNA are treated using this method. The shortcoming of this method is that it ignores various chemical and physical properties associated to electrons as there are many chemical properties which are not defined within the framework of this method such as, electronic excited states, shape and energies of molecular orbitals, etc. This method is valued for the limited class of molecules for which force field is parameterized.

2.6 Molecular Docking Simulation

Molecular docking simulation is a computer simulation method implemented to predict the structure (or structures) of the resulting intermolecular complex formed due to coalescion between two or more molecules/macro biomolecules. The docking program generates a large number of possible docked conformations, and so they require a definite ranking according to their docking scores to identify those conformations that draw the most interest [43-48]. The three important components of docking simulation include:

a. Representation of the biomolecular system
b. Search of the definite conformational space via an inbuilt search algorithm
c. Hierarchical ranking of obtained potential solutions using the inbuilt scoring function

The main application of docking software is to computationally simulate the molecular identification process and accomplish an optimized docked

conformation so that the free energy of the overall complex gets minimized. The docking method involves large numbers of degrees of freedom. With six translational and rotational degrees of freedom as well as certain conformational degrees of freedom available for each molecule, there is a possibility of obtaining a large number of binding modes. It would be computationally too expensive to generate all those possible conformations. Therefore, various docking algorithms have been developed to tackle this problem [49-54].

The simplest amongst the available docking algorithms treats the two molecules as rigid bodies and explores only six translational and rotational degrees of freedom. This approximation was generally used in the earliest algorithms for docking small ligands to the binding sites of proteins and DNA. Kuntz and co-workers used this algorithm in the docking program DOCK [55].

In order to perform a conformationally flexible docking, we need to take into account the conformational degrees of freedom of the molecule. In this method, the receptor (macro biomolecule) is assumed to be rigid and the ligand is supposed to be flexible and hence can orient itself in different poses. Later, all the common methods for searching conformational space are incorporated into various docking algorithms [56].

Monte Carlo (MC) methods are also used to perform molecular docking simulations, in conjunction with simulated annealing techniques. At each iteration during the MC simulation, the internal conformation (geometrical orientation) of the ligand gets changed due to rotation about a bond or the entire molecule gets randomly rotated or translated about a particular bond. The conformations obtained by these

transformations are tested on an energy-based selection criterion. If the criterion is passed, it gets saved and if not, then these iterations will proceed until the predefined quantity of conformations is collected. Later, the final information is further modified to generate conformations [56-58].

Genetic algorithms (GA) can also be used to perform molecular docking [59-61]. The main idea of docking through the usage of genetic algorithm stems out from Darwin's theory of evolution. Each chromosome codes are not only for the internal geometrical conformation of the ligand but also for the orientation of the ligand within the domains of the binding site of the receptor. Both the orientation and the internal conformation thus vary as the population evolves. The score of each docked structure within the site act as the fitness function used to select the individuals for the next iteration [61].

2.6.1 Popular Docking Softwares

Following **table 2.2** gives the list of popular docking softwares and their description:

Table 2.2: Popular Docking Softwares

S. No.	Software	Description	Reference
1	AutoDock	Grid based scheme; quick computation of interaction energies of ligand-receptor complexes; incorporates evolutionary, genetic and Lamarckian Genetic algorithms; affinities predicted using autodock show great success	62
2	DOCK	ligands geometries and binding sites are described by sets of spheres; overlapping of spheres is done	63

		for binding; in recent versions, electrostatic & molecular-mechanics based interaction energies are incorporated	
3	FlexX	Docked ligand conformers are classified using pose-clustering; it models the entire geometry of the protein-ligand complex and then estimates the binding affinity complex	64,65
4	GLIDE	Grid Based Ligand Docking with Energetics; uses a grid-based scheme is used for docking, representation the shape and properties of the receptor; whereas OPLS-AA force field is used to search for conformational, orientational and positional space of the docked ligand	66
5	GOLD	performs flexible docking; uses genetic algorithm for the conformational search; a powerful tool for screening of novel lead compounds; provides high accuracy and reliability	67-71
6	SURFLEX	performs fully automatic and flexible molecular docking; provides results evaluated for reliability and accuracy in comparison to the existing crystallographic experimental results	72

2.7 Molecular Dynamics Simulation

Molecular dynamics simulation is an important computational tool to describe physical movements of atoms and molecules, i.e. how positions, velocities & orientations of molecules changes with respect to each other over the passage of time. MD provides an interface between laboratory experiment and theory, so it is also called **virtual experiment**. MD is generally used when the real-world experiments are too expensive, time consuming & impossible to do.

In MD simulation atoms and molecules are persuaded to interact with each other for an allotted period of time. The trajectories of atoms and molecules are calculated by mathematically solving the Newton's equations of motions whereas the forces and potential energies are calculated using molecular-mechanics based force fields [73]. Long ranged MD simulations are mathematically difficult to simulate because they generate some errors in numerical integration and we can minimize these errors by using some algorithms but not entirely remove them. Molecular dynamics is an important tool to analyze structural and dynamical properties ones at a different scale of time and space. The two main families of molecular dynamics are classical mechanics approach and quantum molecular dynamics. The classical molecular dynamics quantifies the dynamics of the system and the quantum molecular dynamics approach tells about the nature of chemical bonds. Quantum molecular dynamics provides crucial information for a number of biological problems but they require more computational resources being computationally expensive. The choice of ensemble to be used is also a question in a simulation.

Table 2.3 represents various types of integration algorithms and associated mathematical equation with them that are used to generate velocity and position at each time step during the simulation [76]. One can perform simulation at constant NPT and NVT ensemble. This seems that NPT ensemble is widely used as a standard approach for simulating bio membranes etc. [74]. From **table 2.4,** as shown below, one can get a brief idea of different types of statistical ensembles used in MD simulation [77].

Table 2.3: Different types of integration algorithms with associated mathematical equation used to generate velocity and position at each time step

S. No.	Algorithm	Equation for Position	Equation for Velocity
1	Verlet	$x(t+\Delta t) = 2x(t) - x(t-\Delta t) + \frac{f(t)}{m}(\Delta t)^2$	$v(t) = \frac{x(t+\Delta t) - x(t-\Delta t)}{2x(t)} + O(\Delta t)^2$
2	Velocity Verlet	$x(t+\Delta t) = x(t) + v(t) + \frac{f(t)}{m}(\Delta t)^2$	$v(t+\Delta t) = v(t) + \frac{f(t+\Delta t) + f(t)}{2m}(\Delta t)$
3	Leap Frog	$x(t+\Delta t) = x(t) - \Delta t.\, v(t + \frac{\Delta t}{2})$	$v\left(t+\frac{\Delta t}{2}\right) = v\left(t-\frac{\Delta t}{2}\right) + (\Delta t)\frac{f(t)}{m}$
4	Corrected Velocity 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 2.4: Different types of statistical ensembles used in MD simulation

S. No.	Ensemble	Partition Function	Property
1	Microcanonical (E, V, N)	Energy (E), volume (V) and number of particles (N) of the system remain constant during this simulation scheme	$\Omega(E)$ where, $\Omega(E)$ is the number of micro-states corresponding the system's energy E
2	Canonical (T, V, N)	Temperature (T), volume (V) and number of particles(N) of the	$Z_{TVN} = \sum_i e^{-\beta E_i}$ where, $\beta = 1/kT$ and E_i is the

		system remain constant during this simulation scheme	total energy of the system in the respective microstates
3	Grand Canonical (T, V, μ)	Temperature (T), volume (V) and chemical potential (μ) of the system remain constant during this simulation scheme	$Z_{TV\mu} = \sum_i e^{-\beta(N_i\mu - E_i)}$ Where, $\beta = 1/kT$ and $(N_i\mu - E_i)$ is the total energy of the system in the respective microstates

2.7.1 Basic Scheme of Molecular Dynamics Simulation

When the system under consideration contains large number of atoms then the quantum mechanical description is not sufficient. Hence classical mechanics describes a method which is used to approximate the motion executed by heavy atoms of molecule. The general scheme of MD simulation is obtained by numerically integrating Newton's equations of motions for any system of interest, given by:

$$m_i \frac{d^2 r}{dt^2} = -\nabla_i [U(r_1, r_2, r_3 \ldots\ldots\ldots r_n)] \qquad , i = 1, N \tag{2.28}$$

By solving these equations of motion, the position and velocities as a function of time are obtained. Here m_i and r_i are is the mass and position of particle i & U is the potential energy which depends on positions of N particles in the system. In the molecular system we can divide Newton's equation into two parts-

a) evaluation of energies and underlying forces

b) propagations of atomic positions and velocities

The force field equation contains several energy terms to represent each, chemical bonds, angles and improper torsions and also rotations about bonds & pair

wise non-bonded interactions (Van der Waals and Columbic interactions). Sum of all pairs of effective potential [75], as shown in equation below:

$$U(r_{1,}r_{2,\ldots\ldots}r_n) = \sum \frac{1}{2}k_b(b-b_0)^2 + \sum \frac{1}{2}k_\theta(\theta-\theta_0)^2 + \sum k_w(w-w_0)^2 + \sum k_\varphi \left[1 + \cos(nx-\tau)\right] + \sum\{4_{ij}\left[(\frac{\sigma_{ij}}{r})^{12} - \left(\frac{\sigma_{ij}}{r}\right)^6\right] + \frac{q_i q_j}{r}\} \quad (2.29)$$

Here $k_{b,}k_{\theta,}k_w$ and k_φ are force constants.

2.7.2 Classification of Molecular Dynamics

2.7.2.1 ab-initio Molecular Dynamics

Molecular dynamics is performed with standard code like LAMMPS, GROMACS, etc. while ab-initio molecular dynamics is performed with DFT code like VASP. The difference between ab-initio and standard molecular dynamics arises due to the difference between their approaches in calculation of inter atomic forces. However, in both the cases the motions of atoms are computed using Newton's equation of motion, which means by treating them classically. In ab-initio molecular dynamics we solve inter atomic forces at a given instant of solving prescribed potential [78].

In case of systems, in which forces have not been parameterized in terms of force field, ab-initio molecular dynamics enables to calculate dynamics for those systems. The effects like polarization, charge transfer, bonding etc. may also be treated with ab-initio molecular dynamics while in standard MD these effects are imposed artificially.

2.7.2.2 Coarse Grained Molecular Dynamics

Coarse grained molecular dynamics play essential role in computer simulation of biological systems. The MD simulation for any large system for long time scale is extremely expensive for all atom models. Hence, an idea of molecular dynamics reduces the computational cost by simply re-describing the system and at the same time retains the ability of the technique for the prediction of properties of interest. Such methods hold applicability for a wide range viz., protein-protein interactions, folding, lipid-bilayer structure etc. coarse-grained simulation are helpful in larger systems to achieve fast simulation still coarse-grained simulation limits the systems and processes having lack of detailed atomistic information. Hence, mixed all atom coarse grained is an attractive approach providing full atomic details in to coarse grained models. In coarse grained particles are positioned in a way that they describe lowest frequency modes of the molecule and to represent the excluded volume interaction [79]. Steps involved in coarse grained machinery, are-

a) Defining goals
b) Degree of coarsening
c) Mapping atomistic into coarse grained
d) Interaction between coarse grain particles
e) Reproduce atomistic target function
f) Optimize functions

2.7.2.3 Steered Molecular Dynamics

In steered molecular dynamics an external force is applied to one or more atoms and these atoms are known as steered atoms. The steered molecular dynamics

follows the same idea as that of harmonic constraint. In steered molecular dynamics, atoms are constrained only along the constraint direction. By applying steered molecular dynamics, we can investigate molecular mechanics & elastic properties exhibit by proteins subjected to deformations in AFM etc. Steered molecular dynamics is a derivative of molecular dynamics. We can demonstrate steered molecular dynamics as an important tool in the rational drug design. Why steered molecular dynamics?

a) Accelerates processes to simulation time scale (ns).
b) Explanation of bio-polymers
c) Generate and test hypotheses.
d) Finds underlying unbinding potentials

2.7.3 Force Fields

Force field is widely used term in MD simulation & also known as potential energy function. Force field is the functional form and a set of parameters which are used for the evaluation of the potential energy of a system of atoms. Moreover, it is used to describe the bond angles & torsions, the time evolution of bond lengths & also nonbonding Van der Walls and electrostatic interactions between atoms [80].

Our understanding of macromolecular structure is based upon the forces acting to maintain the atoms in their position. With the understanding of these forces we can know how the atoms in structure move about their mean position [81]. Hence, our aim is to understand the force field around each atom in a structure. Empirical potentials are the force field associated with chemical bond, bond angle & bond dihedrals (bonded forces) and Van-der Waals & electrostatic charge (non-bonded). Non bonded

energy formulation is represented as sum over interactions between the particles. The potential function representing non bonded energy is pair potential. In pair potential, total energy is the sum of energies that contribute between pairs of atoms. Lenard-Jones potential is one of the examples of pair potential:

$$U(r) = 4 \in [(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^{6}] \quad (2.30)$$

These are the electrostatic and Lenard Jones potentials (which describe the van-Der Waal's interactions):

$$U(R) = \sum_{bonds} k_i^{bond}(r_i - r_0)^2 + \sum_{angle} k_i^{angle}(\theta_i - \theta_0)^2 + \sum_{dihedral} k_i{}^{dihedrals}[1 + \cos(n_i\emptyset_i + \delta_i)] + \sum_i \sum_{j \neq i} 4\varepsilon ij \left[\left(\frac{q_{ij}}{r_{ij}}\right)^{12} - \left(\frac{a_{ij}}{r_{if}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\varepsilon r_{ij}} \quad (2.31)$$

$U = \sum_{bond} k_b(r - r_0)^2$ and $U = \sum_{angle} k_\theta(\theta - \theta_0)^2$ are the stretching and bending energy equations and are based on the Hook's law. $\sum_{dihedrals} k_i{}^{dihedrals}[(1 + \cos(n_i\emptyset_i + \delta_i)]$ is the torsion energy and is molded by simple periodic function.

2.7.3.1 Popular Force Fields

AMBER (Assisted Model Building with Energy Refinement) was developed by Kollman et. al. [http://ambermd.org/]. It is based upon parameter transferability and usage of charges derived from electrostatic potential fitting [82,83]. The results obtained using this force field holds good for proteins and nucleic acids.

CHARMM (Chemistry at HARward Macromolecular Mechanics) was developed by Karplus et al. [http://www.charmm.org]. It was originally derived for

proteins and nucleic acids, [84] but now a days, it is used for a large number of macromolecules for their molecular dynamics, solvation, crystal packing, vibrational analysis and also for quantum mechanical/molecular mechanical (QM/MM) calculations. This force field uses five valence terms, one of which is electrostatic and is a basis for other force fields.

GROMOS (Groningen Molecular Simulation) was developed jointly by the University of Groningen and ETH (Eidgenössische Technische Hochschule) of Zurich [http://www.igc.ethz. ch/GROMOS/index]. It is quite popular program for predicting the dynamical motion of molecules and bulk liquids [85]. It is also being used for modeling biomolecules. It uses five valence terms, one of which is electrostatic.

2.7.4 Steps in Molecular Dynamics

An initial set of configurations of the simulated molecule is required before starting the MD simulation. A general scheme showing the flowchart to perform MD simulation is shown in **Figure 2.1**. Generally, the structures of nucleic acids are obtained from protein data bank [86], these structures are retrieved experimentally. 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 0 K, where no motion is observed [87].

There is no mutual force acting upon the atoms in their minimized structure. For the minimum energy, the system calculates the gradient of potential energy and puts it equals to zero, this means that no force is acting on any atom along any direction, and therefore no motion will appear to take place inside the system.

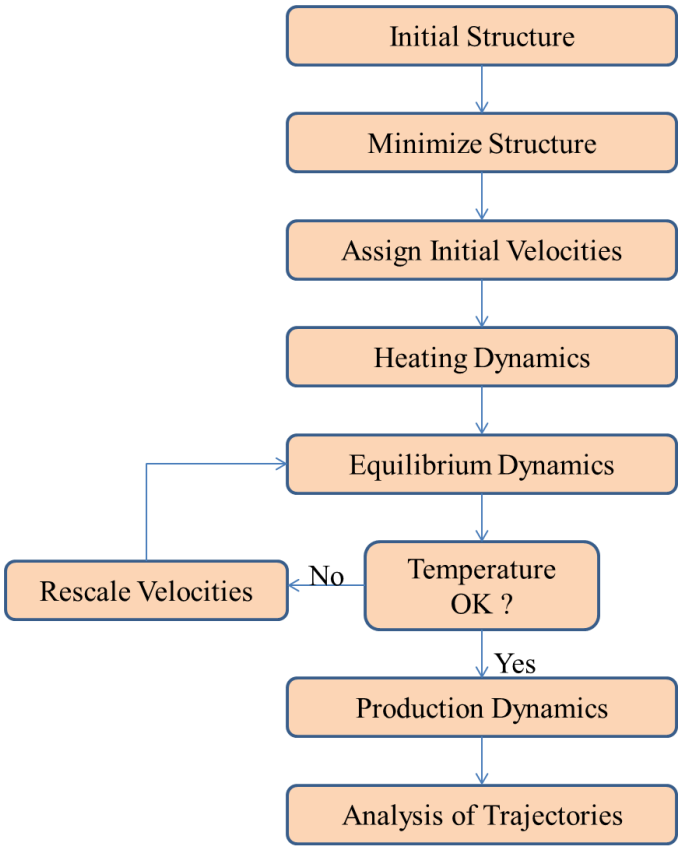


Figure 2.1: Steps involved in MD Simulation

Initial velocities are assigned to the atoms corresponding to a Gaussian distribution in order to raise the temperature from 0K to the desired value, to provide energy to the system [88]. Subsequently, Newton's equations of motion are then integrated to let the system propagate with time. First, the kinetic energy and later the temperature, show the exact same behavior. Since, the total energy remains constant during equilibration process while the kinetic and the potential energies behaves irregularly, this shows that the total energy distribution occurs in a vibrational kind of way between the system components.

The structure may be unstable after heating the system and so the temperature may drop quickly [89]. So, it is needed to equilibrate the system before the production run. Equilibration implies the process in which the kinetic energy and the potential

energy are distributed evenly throughout the system. For the constant period of time, the velocities are rescaled to desired values at the desired temperature. This process is repeated 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.

2.7.5 Popular Molecular Dynamics Softwares

Table 2.5: Popular Molecular Dynamics Softwares

S. No.	Software	Description	Reference
1	AMBER	provides a set of inbuilt modules to apply the AMBER forcefields for simulations of biomolecules viz., LEAP, antechamber, sander, pmemd, cpptraj, mmpbsa, etc.	90
2	DESMOND	computes energies and forces for many standard force fields used in biomolecular simulations; compatible with Drude formalism based polarizable force fields; has facilities for accurate checkpointing and restart.	91
3	GROMACS	excellent CUDA-based GPU acceleration; user-friendly, as topologies and parameter files are written as text format; all programs use a simple interface; keeps updating about the progress of the simulation; it's a free software.	92
4	LAMMPS	classical molecular dynamics program focusing on materials modeling; can simulate solid-state systems, soft matter (biological systems) and coarse-grained or mesoscopic systems	93
5	NAMD	claimed for its parallel simulation efficiency; its interface provides it an edge of access to advanced techniques such as, hybrid QM/MM simulations in an easy-to-use suite	94

2.8 MMPBSA/MMGBSA Calculations

The molecular-mechanics based energies when combined with the Poisson-Boltzmann or Generalized Born Surface Area based continuum solvation models are termed as MMPBSA/MMGBSA methods. This is a very popular approach to evaluate the free energy differences between the bound and unbound states of two solvated molecules [96-100]. Snapshots obtained from the molecular dynamics simulation are used for the calculation to get an average value of the energies. The free energy of binding is calculated by following thermodynamical equations mentioned below:

$$\Delta G_{binding} = \Delta H - T.\,\Delta S \quad (2.32)$$

$$\Delta G_{binding} = (\Delta E_{MM} + \Delta G_{SOL}) - T.\,\Delta S \quad (2.33)$$

Where,

$$\Delta E_{MM} = (E_{MM}^{complex} - E_{MM}^{receptor} - E_{MM}^{ligand}) \quad (2.34)$$

$$\Delta G_{SOL} = (\Delta G_{SOL}^{complex} - \Delta G_{SOL}^{receptor} - \Delta G_{SOL}^{ligand}) \quad (2.35)$$

$$\Delta S = (S^{complex} - S^{receptor} - S^{ligand}) \quad (2.36)$$

Where, ΔH, ΔE_{MM}, ΔG_{SOL}, T and ΔS are the enthalpy contributions to the binding energy, the average difference in molecular mechanical energy, the solvation free energy (including both polar and non-polar components), the absolute temperature and the change in entropy, respectively. Thus, the final binding free energy of the complex equals the sum of intermolecular energy, solvation free energy and the entropic energy contribution terms. Polar solvation free energies are calculated either by solving the linear Poisson Boltzmann (PB) equation or by Generalized Born approximation (GB) models. The non-polar component of energy is evaluated from a linear relation to the solvent accessible surface area (SASA)

model. The entropic contributions are calculated either by solving quasi-harmonic analysis or through normal mode analysis [101,102].

The MMPBSA calculations are based on single minimized structures, and thus are computationally time saviors as they ignore the dynamical effects. By performing the minimizations using MMGBSA method even more time can be saved. The MMGBSA results vary as the simulation length increases, but studies have suggested that there is no benefit for the simulation for longer than 4ns [99]. A study further reveals that the energies calculated using g_mmpbsa (GROMACS) and the AMBER MMPBSA packages are approximately similar and the difference in calculating ΔG_{polar} using both, lies between 1~3 kcal/mol [103].

2.9 Hybrid QM/MM Calculations

The concept of quantum mechanical/molecular mechanical (QM/MM) calculations in biological systems was first proposed and presented by Warshel and Levitt in 1976. They used this method to study the chemical reaction in lysozyme [104]. With due passage of time, the QM/MM theory and related method has developed itself as a crucial computational tool for the modeling of local electronic events in macro biomolecular systems giving precise results. The basic concept behind this method is to chemically demonstrate the active site by QM with accuracy, while the effect of the bimolecular environment is described by MM. Due to the potential use of this method in understanding the biological phenomenon at atomistic scales; this field won, Warshel, Levitt & Karplus, the Nobel Prize in chemistry in 2013. The total energy of the macro biomolecular system is calculated using the two available schemes, viz., the additive and the subtractive schemes, respectively [105-

107]. Regarding this boundary scheme, a diagram showing the convention is given below in **figure 2.2**:

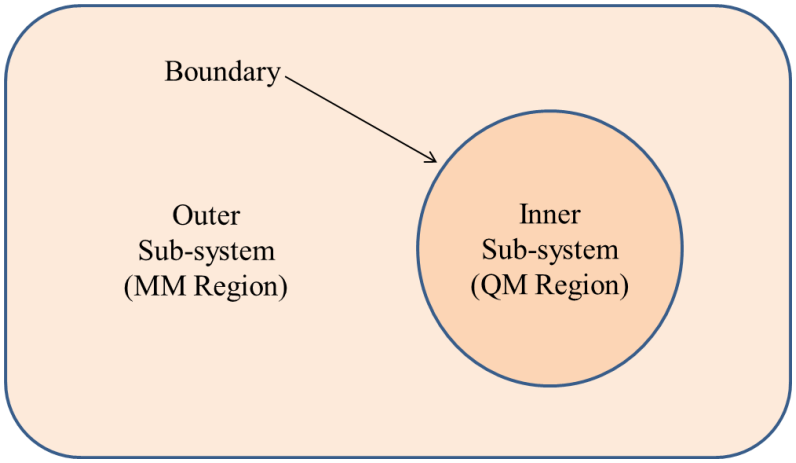


Figure 2.2: Figure representing a typical QM/MM partitioning Scheme

2.9.1 Additive QM/MM Scheme

In this scheme, the total energy ($E_{QM\text{-}MM}$ (system)) contains only three components, the E_{MM} (MM), the molecular mechanical energy of the MM region only, E_{QM} (QM) the quantum mechanical 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 regions. The bonded term elaborates and includes the bond stretching, bending and torsion whereas the non-bonded term elucidates the Van der Waals and electrostatic interactions, respectively. Following **equation 2.37** represents the additive scheme for obtaining QM/MM energy:

$$E_{QM\text{-}MM}\text{ (system)} = E_{MM}\text{ (system)} + E_{QM}\text{ (QM)} - E_{QM\text{-}MM}\text{ (QM-MM)} \quad (2.37)$$

2.9.2 Subtractive QM/MM Scheme

It is given by following **equation 2.38**:

$$E_{QM\text{-}MM}\text{ (system)} = E_{MM}\text{ (system)} + E_{QM}\text{ (QM)} - E_{MM}\text{ (QM)} \quad (2.38)$$

where E_{QM-MM}(system), E_{MM}(system), E_{QM}(QM), E_{MM}(QM) are the total energy, the molecular mechanical energy of the entire system, the quantum mechanical energy of the QM region and the molecular mechanical 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 QM region also.

The main objective of implying QM/MM methods is that they consider the pairing between the electric field from the surrounding and the Hamiltonian based on quantum mechanics from the region of active site and the boundary between QM and MM region is treated accurately. The most important part of QM/MM calculations is the partitioning of the systems. The basic considerations for QM/MM partitioning are:

a. The QM region should be chosen carefully as per the chemical problem; the QM region should be small which can be enlarged later to check the effect of the QM/MM results.
b. If a QM/MM partition is through such that the covalent bonds cannot be avoided, then only unconjugated single bonds are cut and avoided, preferably without electronically demanding substituents (e.g., cut unpolar C–C bonds).

A common procedure for QM/MM calculations starts by taking a crystal structure from protein data bank as a starting point for the calculations. Then missing hydrogens are added, followed by missing atoms etc. Then the system is equilibrated using MM followed by molecular dynamics production run and a few low energy snapshots are generated and studied. These snapshots are eventually taken are starting points for deciding boundaries for QM/MM calculations. These structures contain the

bimolecular system solvated in 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.

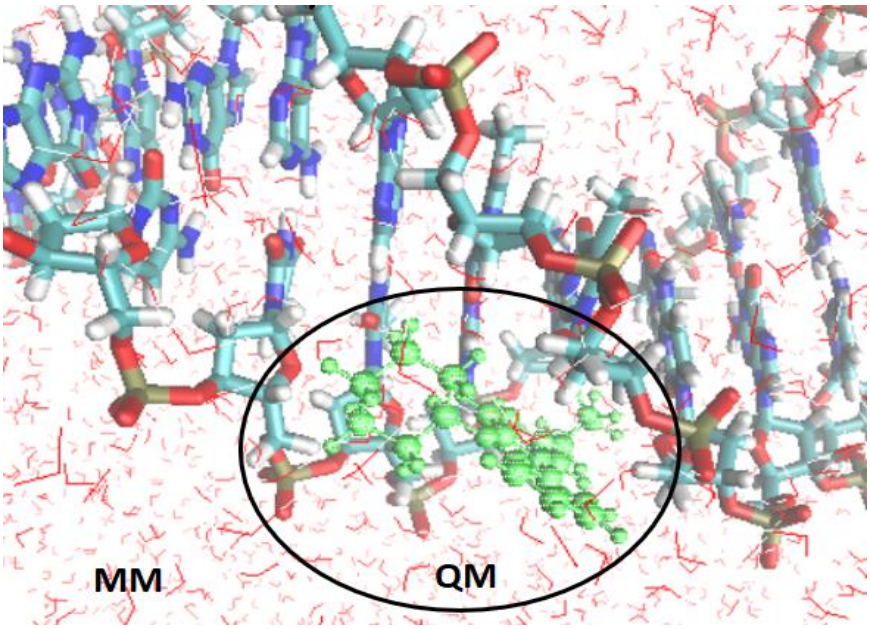


Figure 2.3: Figure representing a two layered QM/MM Scheme

Further, the study of reaction mechanism is similar to the typical methods for the study of gas-phase reactions [108,109]. Morokuma and his co-workers developed the computational chemistry techniques, especially the hybrid QM/MM methods (ONIOM method) for large biomolecular systems [110-112]. The two layered ONIOM scheme is shown above in **figure 2.3**. In many studies, the ONIOM method is used for the study of DNA binding drugs [113, 114].

References:

[1] Shiff, L.I., *Quantum Mechanics,* McGraw- Hill, NewYork., 3rd edition (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] Pople J.A., Beveridge, D.L., *Approximate Molecular Orbital Theory*, McGraw-Hill Book Co., New York (1970).

[8] Hartree, D.R., *Proc. Cambridge Phil. Soc.,* **24,** 426 (1928).

[9] Uhlenbeck, G., Goudsmit, S., *Nature wissenschaften*, **13**, 953 (1925).

[10] Pople, J.A., Beveridge, D.L., Dobosh, P.A., *J. Chem. Phys.*, **47**, 2026 (1967).

[11] Pauli, W., *Z. Physik.*, **31**, 765 (1925).

[12] Slater, J.C., *Phys. Rev.*, **35**, 509 (1930) and **34**, 1239 (1959).

[13] Dirac, P.A.M., *The Principle of Quantum Mechanics*, Oxford University Press, London (1958).

[14] McDonald, J.K.L., *Phys. Rev.*, **43**, 830 (1933).

[15] Young, R.H., *Int. J. Quant. Chem.*, **6**, 596 (1972).

[16] Jenson, F., *Introduction to Computational Chemistry,* 2nd edition, John Wiley & Sons Ltd., Chichester (2007).

[17] Levine, I.N., *Quantum Chemistry*, Pearson, Fifth edition (2000).

[18] Poltev, V., Molecular Mechanics: Method and Applications, Handbook of Computational Chemistry (2012).

[19] Pople, J.A., Beveridge, D.L., *Z. Physik.*, **61**, 126 (1930).

[20] Lennard-Jones, J.E., *Proc. Roy. Soc. (London).*, **A198,** 14 (1949).

[21] Edimiston, C., Ruedenberg, K., *Rev. Mol. Phys.*, **34**, 457 (1963); J. *Chem. Phys.*, **43**, 597 (1965).

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

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

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

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

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

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

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

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

[30] Siegbahn, P.E.M., *Quaet. Rev. Biophys*., **36**, 91 (2003).

[31] Hückel, E., *Z. Physik.*, **70**, 204 (1931).

[32] Feller, D., Davidson, E.R., *Rev. Comp. Chem.*, 1 (1990).

[33] Helgaker, T., Taylor, P.R., "*Modern Electronic Structure Theory*", Part II (D. Yarkony ed.), World Scientific, pp. 727 (1995).

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

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

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

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

[38] Hehre, W.J., Ditchfield, R., Pople, J.A., *Chem. Phys*., **56**, 2257 (1972).

[39] Dunning Jr., T.H., Hay, P.J., *Modern Theoretical Chemistry*, Ed. H. F. Schaefer III, Vol. 3, pp. 1, Plenum, New York (1977).

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

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

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

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

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

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

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

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

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

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

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

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

[52] 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).

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

[54] 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).

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

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

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

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

[59] Judson, R.S., Jaeger, E.P. and Treasurywala, A.M., *J. Mol. Struct. Theochem*, **114**, 191 (1994).

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

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

[62] 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).

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

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

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

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

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

[68] 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).

[69] 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).

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

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

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

[73] Orozco M, Luque FJ., *Chem Rev*. **100**, 4187–4226 (2000)

[74] Martin karplus, J.A., *Nature structural biology*, Nature Publishing group, 2002.

[75] D. Frenkel and B. Smit, "*Understanding Molecular Simulations from Algorithms to Applications*", second edition, Academic Press.

[76] 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).

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

[78] Chaudhuri R, Carrillo O, Laughton CA, Orozco M., *J Chem Theory Comput*., **8**, 2204–2214 (2012).

[79] Mirjalili V, Feig M. *J Chem Theory Comput*., **9**, 1294–1303 (2013)

[80] Lazaridis T, Karplus M., *Proteins*., **35**, 133–152 (1999).

[81] Voelz VA, Bowman GR, Beauchamp K, Pande V.S., *J Am Chem Soc*, **132**, 1526–1528 (2010).

[82] 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).

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

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

[85] Gunsteren van, W.F., Berendsen, H.J.C., *GROMOS 86: Groningen Molecular Simulation Program Package; University of Groningen: Groningen, The Netherlands* (1986).

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

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

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

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

[90] Salomon-Ferrer, R., Case, D.A., Walker, R.C., *WIREs Comput. Mol. Sci.* **3**, 198-210 (2013).

[91] Bowers, K.J., Chow, E., Xu, H., Dror, R.O., Eastwood, M.P., Gregersen, B.A.,

Klepeis, J.L., Kolossvary, I., Moraes, M.A., Sacerdoti, F.D., Salmon, J.K., Shan, Y., Shaw, D.E., Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida, 2006, November 11-17.

[92] Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindahl, E., *SoftwareX,* **1–2**, 2015, 19-25.

[93] Plimpton, S., *J Comp Phys,* **117**, 1-19 (1995).

[94] Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., Schulten, K., *Journal of Computational Chemistry*, **26**,1781-1802, 2005.

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

[96] 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).

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

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

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

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

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

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

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

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

[105] Sherwood, P., de Vries, A.H., Guest, M.F., Schreckenbach, G., Catlow, C.R.A., French, S.A., Sokol, A.A., Bromley, S.T., Thiel, W., Turner, A.J., Billeter, S., Terstegen, F., Thiel, S., Kendrick, J., Rogers, S.C., Casci, J., Watson, M., King, F., Karlsen, E., Sjøvoll, M., Fahmi, A., Schäfer, A., Lennartz, C., *J. Mol. Struct. Theochem*, **632**, 1 (2003).

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

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

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

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

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

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

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

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

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

Chapter-3

Unveiling the Interactions of Some DNA Minor Groove binders through Molecular Docking Calculations

Chapter-3

Unveiling the Interactions of Some DNA Minor Groove Binders through Molecular Docking Calculations

3.1 Introduction

Deoxyribonucleic acid (DNA) is a twisted, double helical strand that serves as the main ingredient by acting as the carrier of all the genetic information [1]. Almost all the anti-cancer therapies involve the interaction of drugs with DNA. The interactions between drug & DNA can be classified broadly into two major categories viz., intercalation and groove binding. Groove binding in DNA takes place via two modes, major groove binding and minor groove binding [2]. In the present chapter, we prefer to analyze the minor groove binding tendency of several ligands with DNA. The ability to predict the conformation as well as the associated energy of binding of ligands with DNA has played a crucial role in bioinformatics and related disciplines. The design of new drugs and test of their potency has been very easy since past few years due to the presence and wide use of various computational tools [3]. Molecular docking technique has proven itself to be of great importance in such a field of research. Since almost all the anti-cancer and anti-microbial drugs preferentially bind itself to the DNA for their therapeutic action over DNA to begin, Minor groove binders are often crescent shaped and therefore their geometry complements the shape of the minor groove of DNA resulting in better binding [4-7].

The binding mechanism of ligands in the vicinity of minor groove of the DNA can be described via following two steps: in the first step, the ligand gets itself transferred to the minor groove of the DNA via hydrophobic and electrostatic interactions whereas in the second step, numerous covalent interactions are formed in between ligand/drug and DNA base pairs [8,9]. These interactions include van-der Waal's contacts, hydrogen bonds, electrostatic interactions and hydrophobic interactions. It has been observed that most of the minor groove binders bind themselves at AT-rich regions of the DNA [10,11]. Various research groups have performed experimental studies in order to explain and understand the binding mechanism of anti-cancer, anti-bacterial and various other diseases. The results obtained from these experiments serve as an excellent database for theoretical scientists to carry out and test the authenticity of their simulations [12-18].

There are many predefined models to analyze and study the protein-DNA interactions but these models are not reliable for the analysis of interactions between drug & DNA. The reason behind this is that unlike in the case proteins and enzymes there is no predetermined active binding site in DNA, whole of the double helical strand of the DNA serves itself as binding site. Therefore, for such cases, various computational molecular modeling techniques have proven themselves to be of great relevance in understanding various non-covalent interactions existing between drug/ligand and DNA base pairs [19]. In this chapter, two different classes of compounds were selected for study, viz., 2,5-bis(4-amidinophenyl) furan & its derivatives (which is labelled as class-1 of compounds) and carbazoles & its analogs (which is labelled as class-2 of compounds) having anti-microbial activity, from

literature [28] and then docked to selected DNA sequences in search of deeper and better insight of their binding modes, binding affinities and to understand the role of hydrogen bonding.

3.2 Computational Methodology

3.2.1 Dataset

The crystal structures of the selected eight DNA sequences (1BNA [20], 1DNE [21], 1QSX [22], 1RMX [23], 195D [24], 2MNB [25], 2MNE [25], 4AH0 [26]) were downloaded from Protein Data Bank (PDB) [27] and are listed below in **table 3.1**. The generalized structure showing the fused rings and the position of the substituent derivatives for two different class of selected compounds, viz., 2,5-bis(4-amidinophenyl) furan & its derivatives (class-1) and carbazole compounds & their analogs (class-2), that were docked with the selected DNA sequences were obtained from literature [28]. Their generalized structures are shown in **figures 3.1** and **3.2**, respectively.

Table 3.1: PDB Id's and sequence of the selected DNAs

S. No.	PDB Id.	DNA Sequence
1	1BNA	5'-CGCGAATTCGCG-3'
2	1DNE	5'-CGCGATATCGCG-3'
3	1QSX	5'- CTTTTGCAAAAG-3'
4	1RMX	5'- CGACTAGTCG-3'
5	195D	5'-CGCGTTAACGCG-3'
6	2MNB	5'- CGACGCGTCG-3'
7	2MNE	5'- CGACTAGTCG-3'
8	4AH0	5'- CGCAAATTTGCG-3'

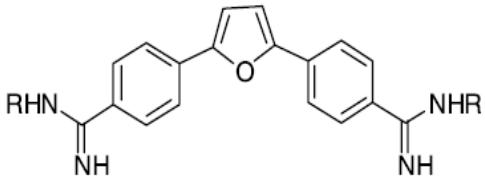


Figure 3.1: Figure showing the generalized structure of the fused ring for class-1 of compounds along with the position where the substituent (R) is to be attached

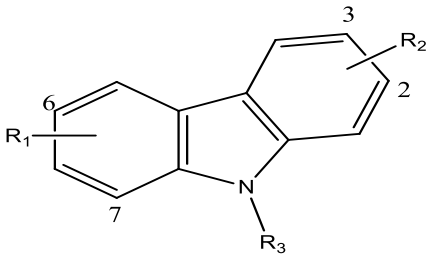


Figure 3.2: Figure showing the generalized structure of the fused ring for class-2 of compounds along with the position where the substituent (R) is to be attached

3.2.2 Geometry Optimization

Water molecules from each of these DNA sequences were removed using Chimera [29]. Chemical structures of all the ligands/drugs were selected from the literature. These chemical structures were then subjected to geometry optimization using Gaussian 09 software package using B3LYP hybrid density functional and 6-31G(d,p) basis set [30].

3.2.3 Molecular Docking

Molecular Docking calculations were performed using Autodock4 software package [31]. The Gasteiger charges were added to the drug-DNA complex using Autodock Tools (ADT) before starting the docking calculations. A grid box, having different dimensions, was prepared for each drug-DNA complex that enclosed the

macromolecule. This helped the drug/ligand in finding the most preferential binding site while the docking calculations were performed [32,33].

Docking calculations were set up using the Lamarckian Genetic Algorithm (LGA). A 20 LGA run having a maximum cycle of 2500000 energy evaluations was performed for each of the drug-DNA complex. The docked pose with the lowest binding affinity was extracted and aligned with the receptor for further analysis [34-38].

3.3 Results and Discussion

The result as obtained by the docking studies has been summarized and discussed as follows:

3.3.1 Optimized Geometries

Optimizing the ligand geometry before docking and other computational calculations is of utmost importance, as it brings the entire system in the lowest possible energy state; having least steric hindrances and charge repulsions. The reason behind this is because the system would then have attained a maximized distance between the bonds, the angles and the dihedrals. Further, the electrostatic charges also tend to get well distributed throughout the system leading to the electron transfer between definite pair of atoms; eventually no distortions in the chemical structures are offered. In the present chapter, the geometry optimization of selected ligands was carried out using Gaussian 09 software using B3LYP hybrid functional at 6-31G** level of theory for their potentials to attain a local minimum. These optimized ligands were then docked to the selected DNA sequence.

3.3.1.1 Class-1 of compounds

Figure 3.3, shown below, represents the optimized structures of the fused ring of the selected ligands obtained after the specific substitutions to the generalized chemical structure belonging to class-1 of compounds.

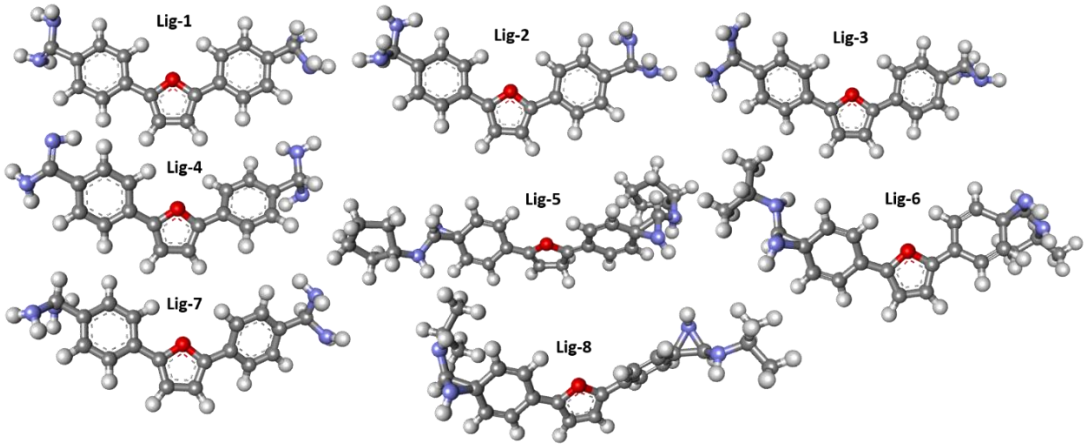


Figure 3.3: Optimized geometrical structures of class-1 of compounds

3.3.1.2 Class-2 of compounds

Figure 3.4, shown below, represents the optimized structures of the fused ring of the selected ligands obtained after the specific substitutions to the generalized chemical structure belonging to class-1 of compounds.

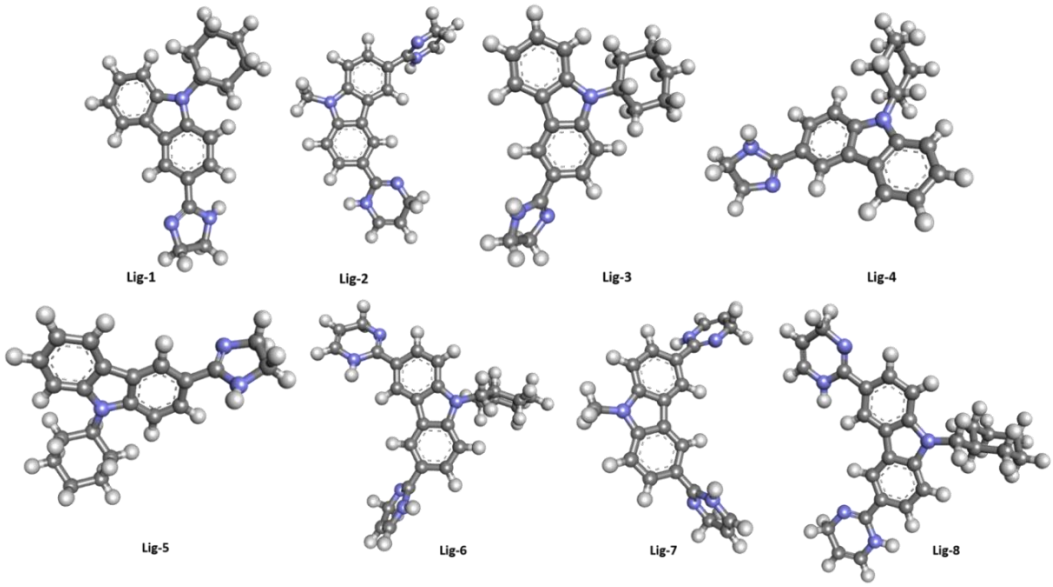


Figure 3.4: Optimized geometrical structures of class-2 of compounds

3.3.2 Molecular Docking Studies

Molecular docking studies were done for the two sets of compounds in search of their preferential binding modes, corresponding binding affinities and for gaining atomistic insights regarding the factors that contribute to the stable complex formation. It was observed that class-1 of compounds were better DNA binders than class-2 of compounds, because of the presence of O-atom in the chemical structure of class-1 compounds whereas class-2 compounds had bulky groups attached to it owing to decreased interactions. The docking results for both, class-1 and class-2 complexes are discussed in detail as follows:

3.3.2.1 Class-1 of compounds

The binding energies obtained by computational docking for the class-1 of compounds are listed below in **table 3.2**, the compounds having the least binding affinities are shown in bold; these represent the formation of stable complexes [2] with corresponding mentioned DNA sequence. **Figure 3.5** represents the best docked

poses corresponding to compounds having least binding affinities, whereas **figure 3.6** represents the interaction profile in 2D obtained using Discovery studio visualizer [39] which reveals various types of interactions taking place between the DNA bases and the selected drugs eventually leading to stabilization or de-stabilization of the complexes thus formed.

Table 3.2: DNA binding affinities of complexes as measured by computational docking along with their change in thermal melting values for class-1 compounds

Substitution R	**Exp.** (ΔTm)	**1BNA** (kcal / mol)	**1DNE** (kcal / mol)	**1QSX** (kcal / mol)	**1RMX** (kcal / mol)	**195D** (kcal / mol)	**2MNB** (kcal / mol)	**2MNE** (kcal / mol)	**4AH0** (kcal / mol)
H	11.7	**-9.89**	**-13.45**	**-10.19**	**-10.69**	-10.88	-9.59	**-10.44**	-10.64
i-Pr	14.4	-8.60	-12.55	-9.71	-10.29	-10.03	**-9.74**	-9.54	**-11.36**
c-Pentyl	15.8	-9.42	-13.40	-10.12	-9.49	**-11.35**	-9.53	-9.83	-10.73
c-Pr	12.4	-7.60	-9.32	-6.29	-8.22	-7.58	-6.50	-6.36	-7.21
2-Pentyl	7.7	-7.60	-10.48	-9.04	-7.50	-9.44	-8.50	-7.38	-10.43

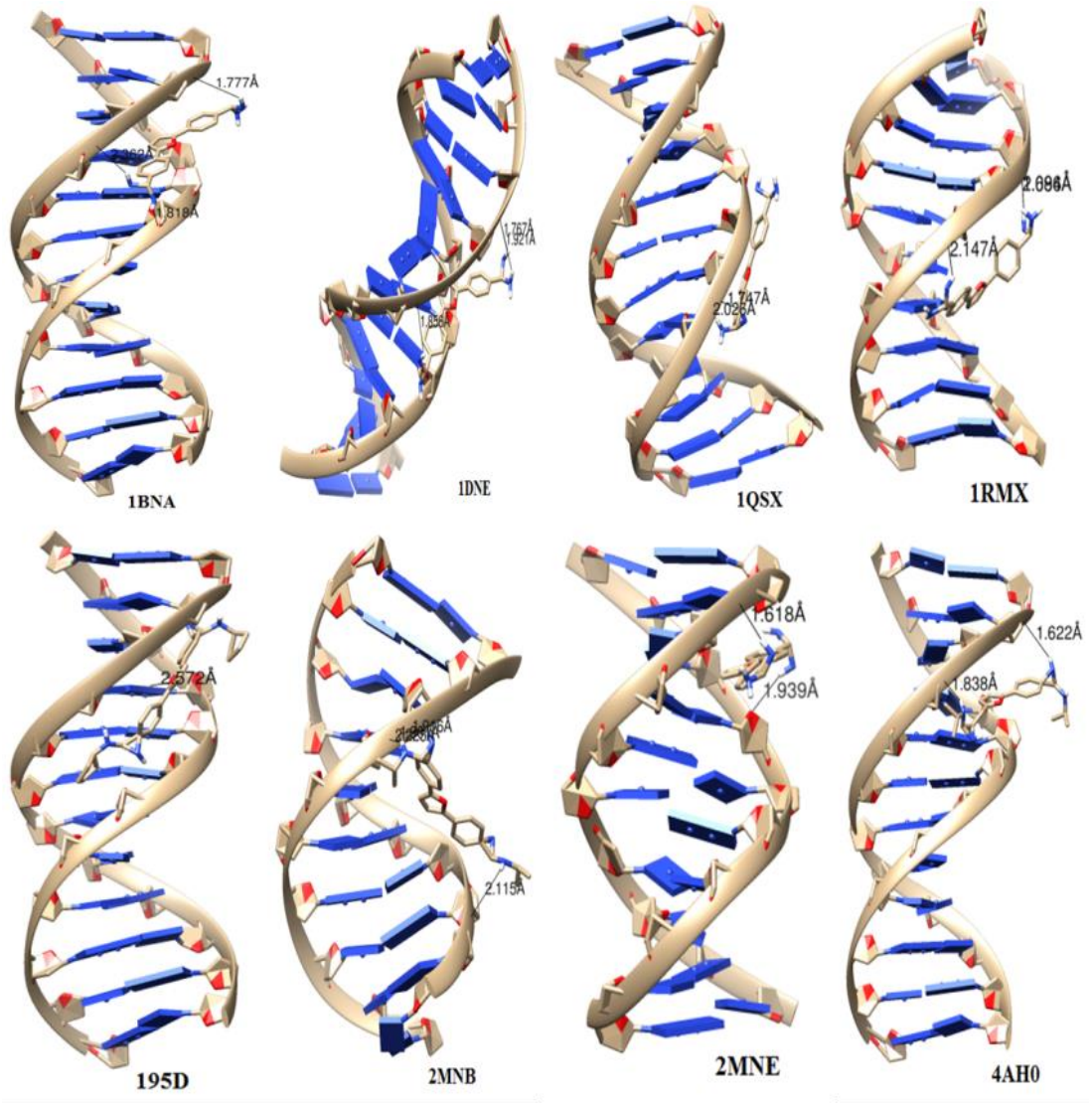


Figure 3.5: Figure representing the best docked posed complexes for class-1 compounds

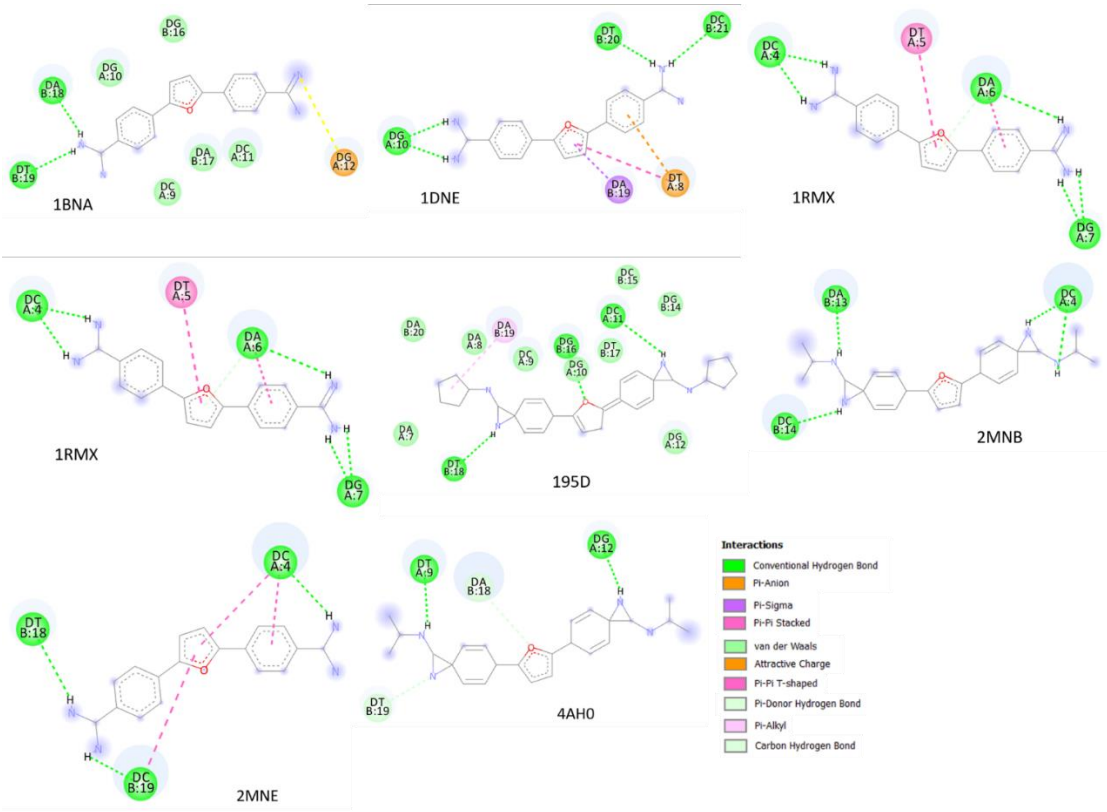


Figure 3.6: Figure representing the interaction profile for the best docked posed complexes for class-1 compounds

3.3.2.2 Class-2 of compounds

The binding energies obtained by computational docking for the class-2 of compounds are listed below in **table 3.3**, the compounds having the least binding affinities are shown in bold; these represent the formation of stable complexes [2] with corresponding mentioned DNA sequence. **Figure 3.7** represents the best docked poses corresponding to compounds having least binding affinities, whereas **figure 3.8** represents the interaction profile in 2D obtained using Discovery studio visualizer [39], which reveals various types of interactions taking place between the DNA bases and the selected drugs eventually leading to stabilization or de-stabilization of the complexes thus formed.

Table 3.3: DNA binding affinities of complexes as measured by computational docking along with their change in thermal melting values for class-2 compounds

Substitutions				Exp. ΔT_m	1BNA	1DNE	1QSX	1RMX	195D	2MNB	2MNE	4AH0
R_1, R_2	**R_1**	**R_2**	**R_3**	**(°C)**	**(kcal/mol)**							
3,6	Am	Am	H	17.2	-7.51	-7.19	-6.81	-6.16	-7.90	-6.02	-6.98	-8.58
3,6	Am	Am	CH_3	19.5	-6.82	-7.24	-7.97	-6.00	-7.06	-6.23	-6.81	-6.63
2,6	IsoAm	IsoAm	H	9.6	-6.49	-7.41	-7.10	-6.86	-7.55	-6.05	-6.37	-9.02
3,6	Im	Im	H	19.5	-8.69	-7.92	-8.73	-5.96	-8.75	-7.35	-7.72	-9.20
3,6	Im	Im	CH_3	24.0	-7.72	-8.07	-8.27	-6.12	-8.54	-7.48	-7.57	-9.31
3,6	Im	H	CHM	16.8	**-9.75**	-8.47	**-9.78**	**-7.56**	**-9.86**	-7.11	-6.74	-9.66
3,6	THP	THP	CH_3	13.6	-8.58	**-8.89**	-8.67	-6.31	-8.92	-7.54	**-7.95**	-8.32
3,6	THP	THP	CHM	7.3	-8.78	-8.72	-6.57	-6.86	-9.78	**-7.72**	-7.54	**-10.24**
2,7	Am	Am	H	19.0	-7.21	-8.25	-7.97	-6.88	-8.37	-6.55	-6.65	-8.06
2,7	Im	Im	H	18.6	-8.25	-7.04	-9.11	-7.05	-9.05	-6.12	-6.04	-8.38
2,7	Im	Im	CH_3	19.1	-7.42	-8.84	-7.76	-6.12	-8.08	-6.50	-6.59	-8.02

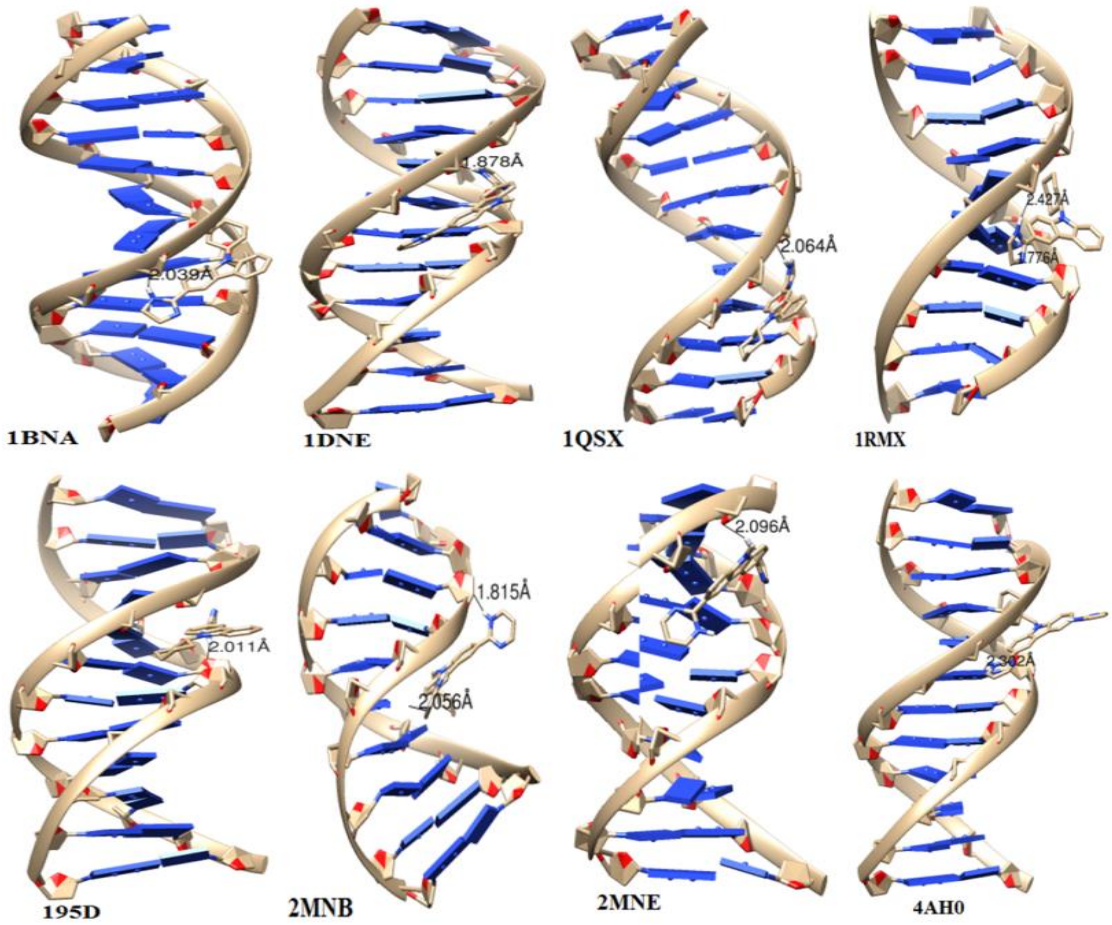


Figure 3.7: Figure representing the best docked posed complexes for class-2 compounds

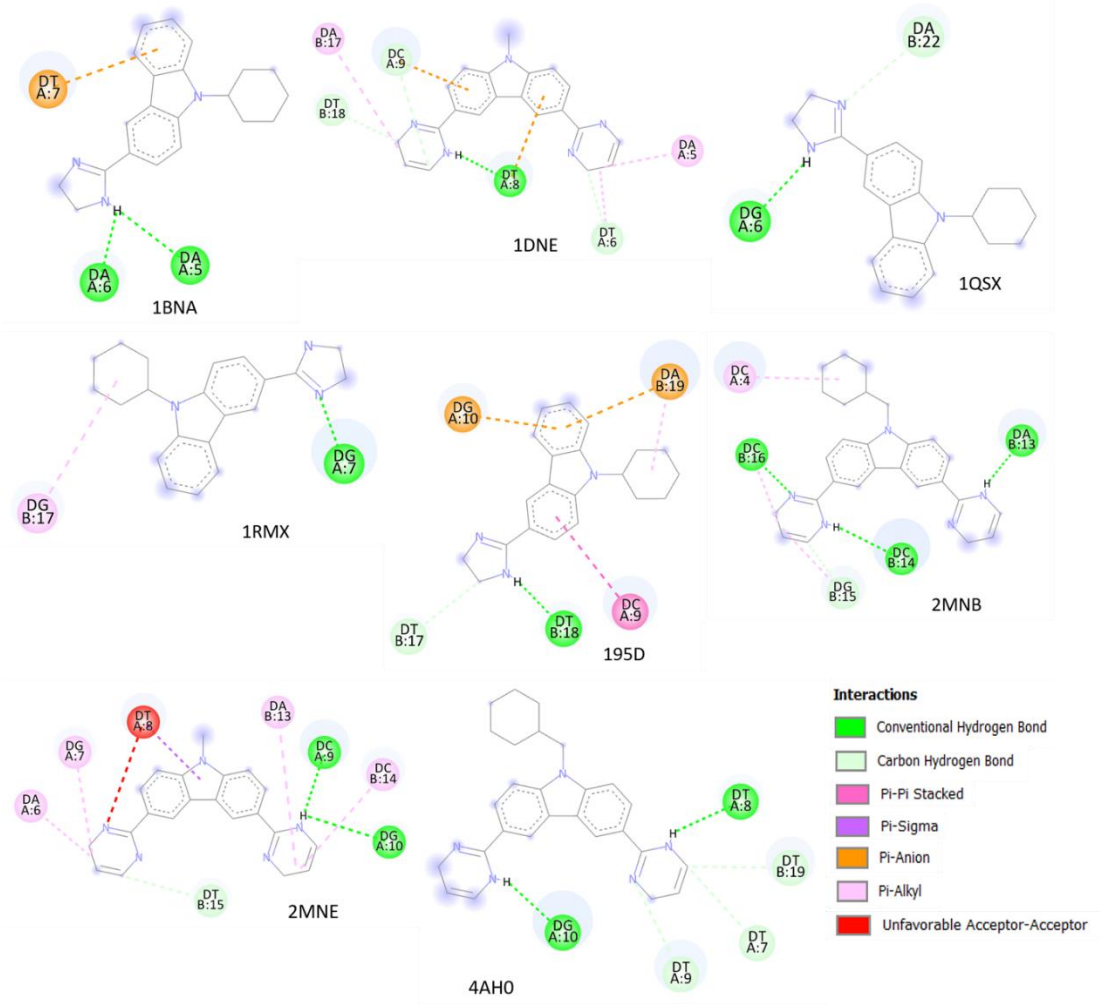


Figure 3.8: Figure representing the interaction profile for the best docked posed complexes for class-2 compounds

3.3.3 Hydrogen Bonding Analysis

Formation of hydrogen bond owes solely to the complex stability [2]. The atoms of the ligands and the DNA residues involved in the formation of hydrogen bonds are visualized in **figures 3.9** & **3.11** respectively, whereas, the acceptor and donor regions during the formation of hydrogen bonds are shown in **figures 3.10** & **3.12.** These assist in demonstrating the electron rich and electron deficient sites in the drug-DNA complexes, and thus helps in the prediction of possible hydrogen binding

sites within the system. These regions also represent the atoms which have the tendency to donate/accept electrons so as to achieve stability [40] during the docking calculations. In this section, the hydrogen bonding interactions between the DNA bases and the ligands are presented as follows:

3.3.3.1 Class-1 of compounds

Figure 3.9 shown below, gives the 3D representation of the hydrogen bonds existing between the DNA bases and the class-1 of compounds corresponding to the best docked posed complexes, whereas **table 3.4** depicts the donor and the acceptor bases of the DNA involved in the formation of hydrogen bond and corresponding length and number of hydrogen bonds formed.

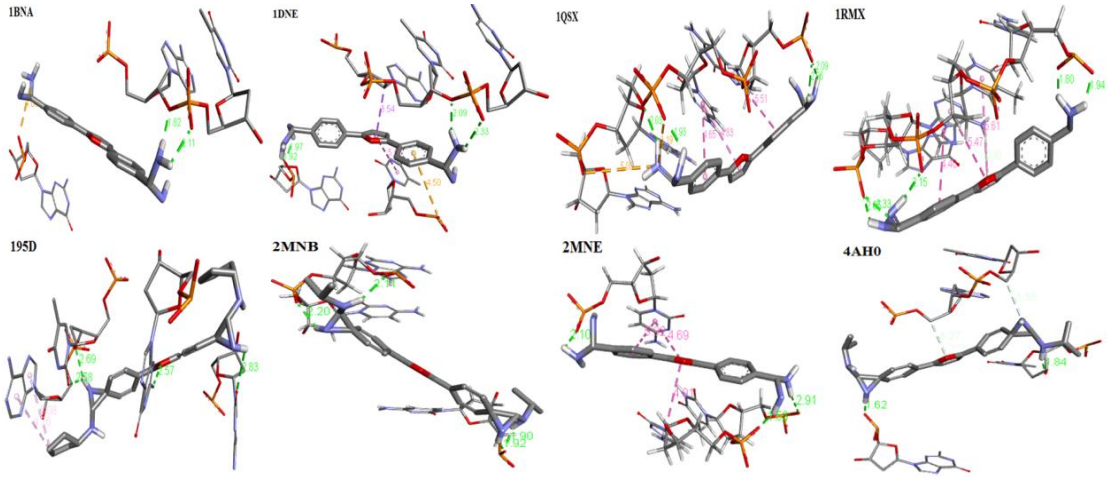


Figure 3.9: Figure representing the H-bonds formed corresponding to the best docked posed complexes for class-1 of compounds in 3D

Table 3.4: Table representing the donor and the acceptor species and H-bond length formed between the DNA and class-1 of compounds

S. No.	Complexes	Number of H-Bonds	Interacting Species	H-Bond length (Å)
1	1bna + lig-6	2	LIG0:H - DT19:OP1	2.105473
			LIG0:H - DA18:O3'	1.818074
2	1dne + lig-7	4	LIG0:H - DG10:OP1	1.920956
			LIG0:H - DG10:OP1	1.970445
			LIG0:H - DT20:O3'	2.092594
			LIG0:H - DC21:OP1	2.326452
3	1qsx + lig-6	4	LIG0:H - DA20:O5'	2.025997
			LIG0:H - DA20:OP2	1.933799
			LIG0:H - DG18:OP2	2.064850
			LIG0:H - DG18:OP2	2.089325
4	1rmx + lig-6	5	LIG0:H - DC4:OP2	1.796149
			LIG0:H - DC4:OP1	1.939713
			LIG0:H - DA6:OP2	2.146880
			LIG0:H - DG7:OP2	2.053564
			LIG0:H - DG7:OP2	2.330354
5	195d + lig-6	4	LIG0:H - DT18:O3'	2.692488
			LIG0:H - DT18:O2	2.579076
			DG16:N2 - LIG0:O	2.571794
			LIG0:H - DC11:O2	2.834610
6	2mnb + lig-8	4	LIG0:H - DC4:OP2	1.916382
			LIG0:H - DC4:OP2	1.901026
			LIG0:H - DC14:OP2	2.196361
			LIG0:H - DA13:OP1	2.114568
7	2mne + lig-7	3	LIG0:H - DT18:OP1	2.912220
			LIG0:H - DC19:OP1	1.534208
			LIG0:H - DC4:OP1	2.100831
8	4ah0 + lig-8	2	LIG0:H - DG12:OP1	1.621622
			LIG0:H - DT9:O3'	1.837660

Following **figure 3.10** represent the H-donor/acceptor clouded regions at the atomistic binding sites for each of the drug-DNA complexes for class-1 of compounds. These specific regions in the biomolecular systems help in quantifying the electron rich and electron deficient sites in the biomolecular complexes, and thus helps in the prediction of possible hydrogen binding sites within the system and the strength of the bond formed.

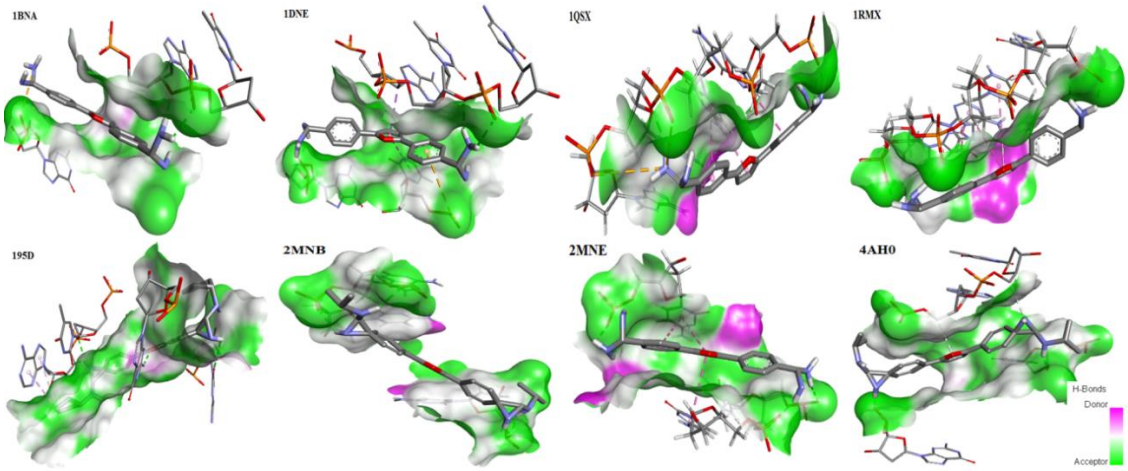


Figure 3.10: Figure showing the donor and acceptor regions for the H-bonds formed corresponding to the best docked posed complexes for class-1 of compounds

3.3.3.2 Class-2 of compounds

Figure 3.11 shown below, gives the 3D representation of the hydrogen bonds existing between the DNA bases and the class-2 of compounds corresponding to the best docked posed complexes, whereas **table 3.5** represents the donor and the acceptor bases of the DNA involved in the formation of hydrogen bond and corresponding length and number of hydrogen bonds formed.

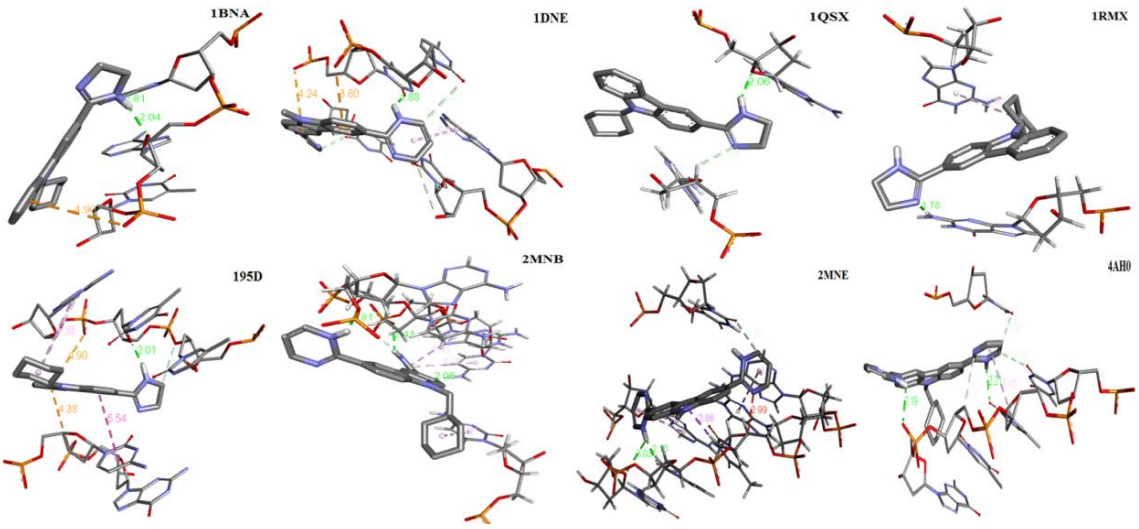


Figure 3.11: Figure representing the H-bonds formed corresponding to the best docked posed complexes for class-2 of compounds in 3D

Table 3.5: Table representing the donor and the acceptor species and H-bond length formed between the DNA and class-2 of compounds

S. No.	Complexes	Number of H-Bonds	Interacting Species	H-Bond length (Å)
1	1bna + lig-1	2	LIG0:H - DA5:N3 LIG0:H - DA6:O4'	2.806182 2.039128
2	1dne + lig-1	1	LIG0:H - DT8:O2	1.877852
3	1qsx + lig-1	1	LIG0:H - DG6:O4'	2.064408
4	1rmx + lig-1	1	DG7:H22 - LIG0:N	1.776228
5	195d + lig-3	1	LIG0:H - DT18:O4'	2.011384
6	2mnb + lig-2	3	LIG0:H - DC14:O3' DC16:H42 - LIG0:N LIG0:H - DA13:O5'	2.410066 2.055858 1.814939
7	2mne + lig-1	2	LIG0:H - DC9:O2 LIG0:H - DG10:O4'	2.095689 3.027310

8	4ah0 + lig-2	2	LIG0:H - DG10:OP1 LIG0:H - DT8:O2	2.188940 2.209946

Following **figure 3.12** represent the H-donor/acceptor clouded regions at the atomistic binding sites for each of the drug-DNA complexes for class-2 of compounds. These concentrated regions in the biomolecular systems help in the demonstration of the electron rich and deficient sites in the biomolecular complexes, and thus helps in the prediction of possible hydrogen bond formation, with their formation sites and also help in the determination of the strength of the hydrogen bond formed.

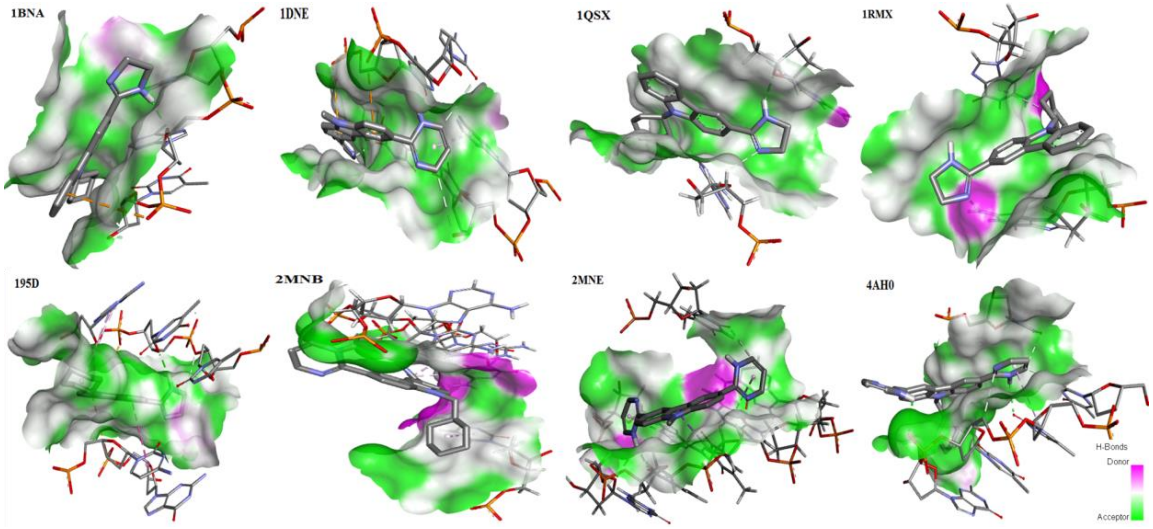


Figure 3.12: Figure showing the donor and acceptor regions for the H-bonds formed corresponding to the best docked posed complexes for class-2 of compounds

3.4 Conclusions

Docking study performed in the current research work was done in order to evaluate the DNA binding affinities for the two class of ligands. The major focus was to get an elaborated perspective of DNA minor groove binders at molecular levels. Thus, our study not only helps in getting a deeper insight regarding the DNA binding mechanism but also evaluates the binding affinities of 2,5-bis(4-amidinophenyl) furans & its derivatives (class-1) and carbazoles & its analogues (class-2) with selected DNA sequences.

Both the class of ligands, class-1 and class-2 were docked to the minor groove of the selected DNA sequences. Their binding site was found to be AT-rich regions of the DNA, as preferred by most of the DNA minor groove binders.

Further, we can conclude that class-1 of the drugs are better DNA binders than class-2 of drugs, as the former has formed a greater number of hydrogen bonds than the later. The reason behind this is the geometry of the two class of compounds, class-1 compounds involved O-atom eventually leading them to find more hydrogen bonded interactions whereas, in the later class of atoms the attached groups were bulky eventually offering more steric hindrances. However, no such steric hindrances were involved in class-1 compounds and thus leading to better binding with DNA.

The results obtained through hydrogen bonding analysis suggested that donor and acceptor clouded regions essential for the formation of hydrogen bonds had greater extent for class-1 compounds, eventually leading the drug-DNA complex to form hydrogen bonds with greater ease than class-2 compounds, this factor also added to the increased interactions and stability of the drug-DNA complexes.

Thus, it may be concluded that this analysis will certainly help in the improvement of existing DNA minor groove binders having antimicrobial potency and would also prove helpful in the design of new and potent drugs. It also adds to enrich the database regarding the structure activity relationship of ligands with DNA through computer simulation studies.

References:

[1] Watson, J.D., Crick, F.H.C., *Nature*, **171**, 737–738 (1953).

[2] Mishra, R., Gaur, A.S., Chandra, R., Kumar, D., *International Journal of Science, Technology and Society*, **3**, 11-27 (2017).

[3] Mehmood, M.A., Sehar, U., Ahmad, N., *J Data Mining Genomics Proteomics*, **5**, 158 (2014).

[4] Song, Y.M., Wu, Q., Yang, P.J., Luan, N.N., Wang, L.F., Liu, Y.M.J., *Inorganic Biochemistry*, **100**, 1685–1691 (2006).

[5] Dervan, P.B., *Bioorganic Medicinal Chemistry*, **9**, 2215–2235 (2001).

[6] Dervan, P.B., Edelson, B.S., *Current Opinions in Structural Biology*, **13**, 284–299 (2003).

[7] Masta, A., Gray, P.J., Phillips, D.R., *Nucleic Acids Research*, **23**, 3508–3515 (1995).

[8] Tse, W.C., Boger, D.L., *Chemistry & Biology*, **11**, 1607-1617 (2004).

[9] Nunn, M.C., Neidle, S. J., *Medicinal Chemistry*, **38**, 2317-2325 (1995).

[10] Kennard, O., *Pure & Applied Chemistry*, **65**, 1213-1222 (1993).

[11] Srivastava, H.K., Chourasaia, H., Kumar, D., Sastry, G.N., *Journal of Chemical Information and Modeling*, **51**, 558-571 (2011).

[12] 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., *Medicinal Chemistry Communications*, **2**, 780-788 (2011).

[13] Kamal, A., Shankaraiah, N., Reddy, C.R., Prabhakar, S., Markandeya, N., Srivastava, H.K., Sastry, G.N., *Tetrahedron*, **66**, 5498-5506 (2010).

[14] 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., *Bioorganic and Medicinal Chemistry*, **18**, 526-542 (2010).

[15] Kamal, A., Rajender, Reddy, D.R., Reddy, M.K., Balakishan, G., Shaik, T.B., Chourasia, M., Sastry, G.N., *Bioorganic and Medicinal Chemistry*, **17**, 1557-1572 (2009).

[16] Kamal, A., Khan, M.N.A., Reddy, K.S., Rohini, K., Sastry, G.N., Sateesh, B., Sridhar, B., *Bioorganic and Medicinal Chemistry Letters*, **17**, 5400-5405 (2007).

[17] 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., *Bioorganic Medicinal Chemistry*, **18**, 4747-4761 (2010).

[18] Ramya, P.V.S., Guntuku, L., Angapelly, S., Karri, S., Digwal, C.S., Babu, B.N., Naidu, V.G.M., Kamal, A., *Bioorganic & Medicinal Chemistry Letters*, **28**, 892–898 (2018).

[19] Mishra, R., Kumar, A., Chandra, R., Kumar, D., *International Journal of Science, Technology and Society*, **3**, 11-27 (2017).

[20] Drew, H.R., Wing, R.M., Takano, T., Broka, C., Tanaka, S., Itakura, K., Dickerson R.E., *Proc. Natl. Acad. Sci. USA*, **78,** 2179-2183 (1981).

[21] Coll, M., Aymami, J., van der Marel, G.A., van Boom, J.H., Rich, A., Wang, A.H., *Biochemistry,* **28,** 310-320 (1989).

[22] Gavathiotis, E., Sharman, G.J., Searle, M.S., *Nucleic Acids Res.,* **28,** 728-735 (2000).

[23] Anthony, N.A., Johnston, B.F., Khalaf, A.I., MacKay, S.P., Parkinson, J.A., Suckling, C.J., Waigh, R.D., *J. Am. Chem. Soc.,* **126,** 11338-49 (2004).

[24] Balendiran, K., Rao, S.T., Sekharudu, C.Y., Zon, G., Sundaralingam, M., *Acta Crystallogr D Biol Crystallogr.*, **51**, 190-8 (1995).

[25] Alniss, H.Y., Salvia, M.V., Sadikov, M., Golovchenko, I., Anthony, N.G., Khalaf, A.I., MacKay, S.P., Suckling, C.J., Parkinson, J.A., *Chembiochem.*, **15**, 1978-1990 (2014).

[26] Munnur, D.G., Mitchell, E.P., Forsyth, V.T., Teixeira, S.C.M., Neidle, S., *RCSB PDB*, http://www.rcsb.org/structure/4AH0, (not yet published).

[27] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., *Nucleic Acids Research*, **28**, 235–242 (2000).

[28] Tidwell, R.R., Boykin, D.W., *Journal of Brazilian Chemical Society*, **13**, 414-460 (2002).

[29] Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., *Journal of Computational Chemistry*, **25**, 1605-1612 (2004).

[30] Gaussian 09, Revision B.01, Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, G. A., Nakatsuji, H., Li, X., Caricato, M., Marenich, A. V., Bloino,

J., Janesko, B. G., Gomperts, R., Mennucci, B., Hratchian, H. P., Ortiz, J. V., Izmaylov, A. F., Sonnenberg, J. L., Williams-Young, D., Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D., Zakrzewski, V. G., Gao, J., Rega, N., Zheng, G., Liang, W., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Throssell, K., Montgomery, J. A., Jr., Peralta, J. E., Ogliaro, F., Bearpark, M. J., Heyd, J. J., Brothers, E. N., Kudin, K. N., Staroverov, V. N., Keith, T. A., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A. P., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Millam, J. M., Klene, M., Adamo, C., Cammi, R., Ochterski, J. W., Martin, R. L., Morokuma, K., Farkas, O., Foresman, J. B., Fox, D. J. Gaussian, Inc., Wallingford CT, 2009.

[31] Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J., *Journal Computational Chemistry*, **16**, 2785-91 (2009).

[32] Baraldi, P.G., Cacciari, B., Guiotto, A., Romagnoli, R., Zaid, A.N., Spalluto, G., *IL Farmaco*, **54**, 15-25 (1999).

[33] Blaney, J.M., Dixon, J.S., *Perspectives in Drug Discovery and Design*, **1**, 301-319 (1993).

[34] Chaires, J.B., *Biopolymers*, **44,** 201-215 (1997).

[35] Chaires, J.B., *Current Opinions in Structural Biology*, **8**, 314-320 (1998).

[36] Chaires, J.B., *Annual Review of Biophysics*, **37**, 135-151 (2008).

[37] Chalikian, T.V., Breslauer, K.J., *Current Opinion in Structural Biology*, **8**, 657-664 (1998).

[38] Gilbert, D.E., Feigon, J., *Current Opinion in Structural Biology*, **1**, 439-445 (1991).

[39] Dassault Systèmes BIOVIA, Discovery Studio Visualizer, 2020, San Diego, Dassault Systèmes, (2020).

[40] Aamir, M., Singh, V.K., Dubey, M.K., Meena, M., Kashyap, S.P., Katari, S.K., Upadhyay, R.S., Umamaheswari, A., Singh, S., *Front. Pharmacol.*, **9,** 1038 (2018).

Chapter-4

Interaction, Dynamics and Stability Analysis of Some Minor Groove Binders with B-DNA Dodecamer 5'-(CGCAAATTTGCG)-3'

Chapter-4

Interaction, Dynamics and Stability Analysis of some Minor Groove Binders with B-DNA Dodecamer 5'-(CGCAAATTTGCG)-3'

4.1 Introduction

Deoxyribonucleic acid (DNA) has been a known cellular target for many antibacterial and anticancer agents due to its gene expression tendencies. The interaction of drugs with nucleic acid is very important feature for drug industry and holds a significant role in understanding the action mechanism of drugs and helps in designing of more efficient drugs with minimal side effects.

Minor groove of the DNA is a proven target for many anticancer and antitumor drugs. The forces that are involved in minor groove binding with DNA are: electrostatic interactions, Van der Waals interactions, hydrophobic interactions, and hydrogen bonding [1]. Sequence specificity is an important tool for studying drug-DNA interactions. Another important factor that governs the interactions of small molecules with DNA is their crescent shape which is complementary to the minor groove and consequently DNA minor groove binding is favored [2]. Furans have been known since long and are often claimed to have possessed various medicinal and therapeutic tendencies.

In the current chapter, we selected four molecules belonging to two different classes viz., 2,4-bis(4-amidinophenyl) furans and reversed diamidino 2,5-diarylfurans, labelled as mol-1, mol-2 mol-3 & mol-4. These molecules were computationally studied for their relative binding strengths and stable complex

formation tendencies with DNA (PDB Id: 4AH0). Molecular docking was performed to predict binding pocket of the drug in the vicinity of DNA and molecular dynamics simulation was performed to study the interaction dynamics in support of the stability of the predicted binding mode.

Frequently, minor groove binders show binding selectivity towards AT-rich regions of DNA; several factors are responsible for this, first one is the electrostatic potential of AT-rich region is higher than that of GC-rich regions and another one is that the AT-rich grooves are deeper and narrower than GC-rich ones [3]. A DNA minor groove with alternating A and T, allows favorable Van der Waals contacts between the drug and DNA unlike the GC-rich region where the geometry of groove is altered by bulky amino groups of guanine bases [4,5].

In the current study, molecular modeling techniques such as geometry optimization, molecular docking & molecular dynamics simulation were implemented for better understanding of the factors responsible for binding and stability of the drug-DNA complexes. The selected ligands were first geometrically optimized in order for them to attain a minimum potential. Further they were docked to the selected DNA in search of the best fit pose. Molecular Docking is simplest computational representation of DNA-ligand interactions [6,7]. Docking also gave information regarding the site binding energy and role of hydrogen bond in obtained docked poses. Various literatures have reported that, the minimum the site binding energy, the better the corresponding docked pose [8]. After geometry optimization and docking, the drug-DNA complexes were put to molecular dynamics simulation [9-11], for a better insight of their dynamical behavior, stable complex formation

tendency and regarding any distortions obtained in the DNA double helix, by means of RMSD & RMSF calculations.

4.2 Computational Methodology

4.2.1 System Selection and Preparation

The lead molecules (mol-1, mol-2, mol-3 & mol-4) were selected from literature [12]. Their chemical structures are shown in **figure 4.1.** The DNA sequence; 4AH0 [13] was obtained from Protein Data Bank [14]. Its structural data including its specific nucleic acid bases sequences and PDB Id. is mentioned in **table 4.1** as obtained from Discovery Studio software package [15]. Water molecules and other residues, from selected DNA sequence were removed using UCSF Chimera before the start of docking calculations [16].

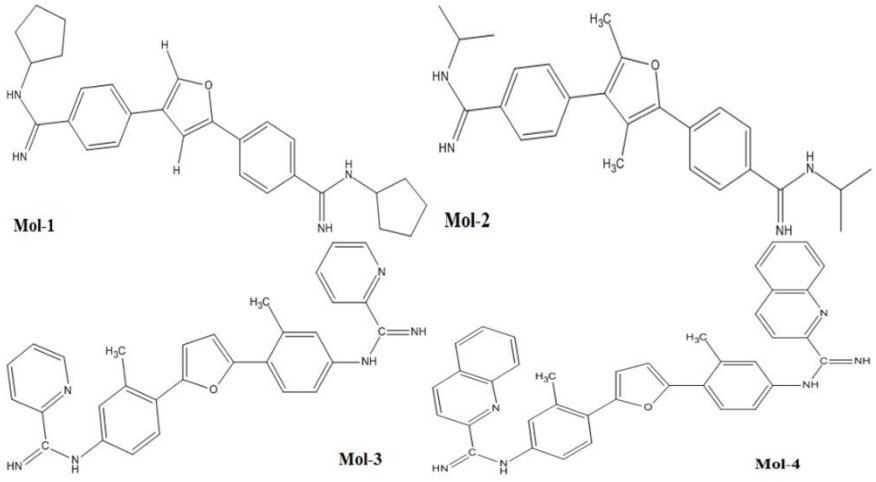


Figure 4.1: Figure showing chemical structures of the selected ligands

Table 4.1: Nucleic Acid Report for 4AH0

S. No.	PDB Id.	Nucleic Acid Sequence from Available Structure	Basic Information
1.	4AH0	B-DNA DODECAMER (5'-D(CGCAAATTTGCG)-3')	• Cell Space Group: 19 (P212121 Origin-1 Choice: 1) • Crystallographic Resolution: 1.20 Å • Molecular Weight of Nucleic Acid Chains: 7268.84 • Number of Nucleic Acids: 24 • Experimental pH: 6.5

4.2.2 Molecular Geometry Optimization

Using optimized ligand geometry before docking and other computational calculations is very important before proceeding for any computational simulations, the reason behind this is that geometry optimization would put the entire system in the lowest possible energy state; having least steric hindrances and charge repulsions; as the system would then have attained a maximized distance between the bonds, the angles and the dihedrals. Further, the electrostatic charges also tend to get well distributed throughout the surface of the system leading to the electron transfer between definite pair of atoms; and eventually no distortions in the chemical structures are offered [17]. Here, the geometry optimization of selected ligands was carried out using Gaussian 09 software [18] using B3LYP hybrid functional at 6-31G** level of theory for their potentials to attain a local minimum. These optimized ligands were then docked to the selected DNA sequence.

4.2.3 Molecular Docking Studies

Molecular Docking is a frequently used computational technique for the prediction of the binding pocket of potent inhibiters in the vicinity of target molecules.

The ability to determine the binding affinity of ligand-receptor complexes is the key objective of docking studies [19]. Here, molecular docking was carried out using Autodock4 software [20]. Docking calculations were set up using Lamarckian Genetic Algorithm (LGA). Gasteiger charges were added to each drug-DNA complex using Autodock Tools (ADT) before starting the docking simulations. A grid box, having various dimensions along the three coordinate axes, was prepared for each drug-DNA complex that enclosed the macromolecule. A 20 LGA run with a maximum cycle of 2500000 energy evaluations was performed for each of the drug-DNA complex. The docked pose with the least binding affinity was extracted and aligned with the receptor (DNA) for further analysis.

4.2.4 Molecular Dynamics Simulation

Molecular Dynamics is a computational technique used for the study the dynamical characteristics of biomolecular systems, viz., protein folding, unwinding and other conformational changes including complex stability, over a specified period of time. Its applications have gained significant importance due to lack of experimental resources [21]. In the present work, GROMACS 5.0.4 software package [22] was used to carry out the molecular dynamics simulations.

A total of four drug-DNA complexes were generated, after docking for molecular dynamics simulations viz., (4AH0+mol-1, 4AH0+mol-2, 4AH0+mol-3 & 4AH0+mol-4). These drug-DNA complexes were put to for 5000ps time scale simulation. There are many studies for the comparison of force fields for the nucleic acids but AMBER force fields [23] seems to be good for nucleic acid simulation due to the presence of specific topologies for the terminal nucleotides. Amber03 force

field already embedded in GROMACS software suite was used to generate the topology for the selected DNA sequence and antechamber module of AMBER program was used to generate topology of selected ligands through a python script: 'acpype.py' [24]. The ligand-DNA complex was then solvated in an octahedral box of using TIP3P water model at 298K [25]. Sodium ions were then added to the solvated box containing the DNA-ligand complex by randomly replacing the water molecules in order to neutralize the system. Particle Mesh Ewald (PME) was used to handle long-ranged electrostatic interactions under periodic boundary conditions [26]. Energy minimization of the whole system was carried out in 25000 steps using Steepest Descent leap- Frog Integration Method followed by NVT ensemble equilibration at a constant temperature of 300K for 50s using Berendsen thermostat [27]. The system was then equilibrated with NPT ensemble at a constant pressure of 1atm in 25000 steps using steepest descent leap-frog integrator [27]. All the bonds involving hydrogen atoms were constrained using the LINCS algorithm [28]. Graphs were plotted using XMgrace software [29].

4.3 Results and Discussion

Present study was aimed to identify new leads, targeting binding affinity and structural & dynamical stability of some antimicrobial ligands with DNA. The results obtained through various computational calculations are discussed and summarized as follows:

4.3.1 Molecular Geometry Optimization

The geometry of the lead molecules needs to be optimized before docking so that the repulsions between the nearest bonded atoms, between their angles and

between their dihedrals could get minimized and the lead molecules would attain a state having least repulsion and hence owing to maximum stability and correspond to minimum energy [30]. The optimized geometries of the selected ligands are shown below in **figure 4.2**.

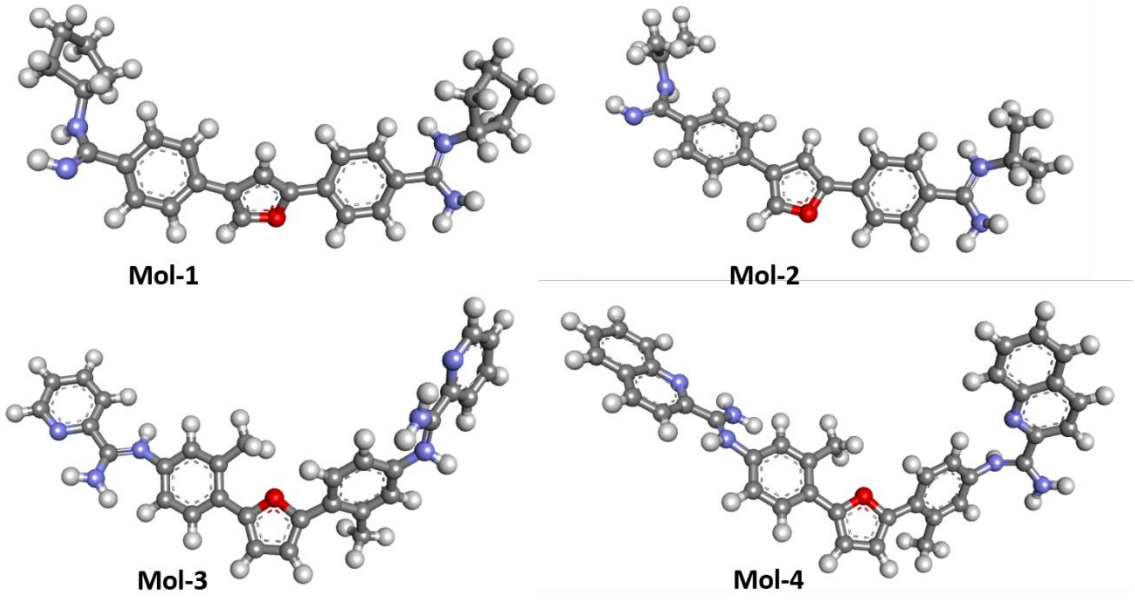


Figure 4.2: Figure showing optimized structures of the selected ligands

4.3.2 Molecular Docking

The selected ligands (mol-1, mol-2, mol-3 & mol-4) which were claimed to possess antimicrobial tendencies in the literature [12] were docked to 4AH0 in search of best docked posed complex leading to stability. The docking result corresponding to the selected DNA sequences are summarized below in **table 4.2**. Docking calculations revealed that 4AH0 with mol-4 had least binding energy of -12.39 kcal/mol. This indicates that since mol-4 had lesser binding energy, it owes to maximum stability of the formed complex. The 3D representation of the docked pose

corresponding to each drug-DNA complex is shown in **figure 4.3.** These figures reveal that both the drugs bonded themselves to the minor groove of each DNA sequence, whereas **figure 4.4** represents the interaction profile in 2D which reveals various types of interactions taking place between the DNA bases and the selected ligands, viz., conventional hydrogen bonds, carbon hydrogen bonds, π-σ bonds, π-alkyl bonds and presence of some attractive charges. These interactions eventually lead to the stabilization or de-stabilization of the drug-DNA complexes formed.

Further, docking calculations revealed that due to presence of more planar aromatic rings in mol-4, its geometry oriented in such a way so as to offer maximum binding affinity with the DNA. Further, from **figures 4.2** & **4.3** and from the binding affinity mentioned in **table 4.2,** it can be seen that as the aromatic rings decreased the binding energy also decreased and similar pattern was observed for inhibition constant also. This means that presence of aromatic rings tends to stabilize the drug-DNA complex owing to higher binding affinity and lower inhibition constants.

Table 4.2: Table representing docking results obtained for 4AH0 DNA sequence

S. No.	Ligand	Binding Energy (kcal/mol)	Inhibition Constant (nM)
1.	Mol-1	-12.03	1.52
2.	Mol-2	-10.94	9.55
3.	Mol-3	-12.06	1.44
4.	Mol-4	-12.39	0.82

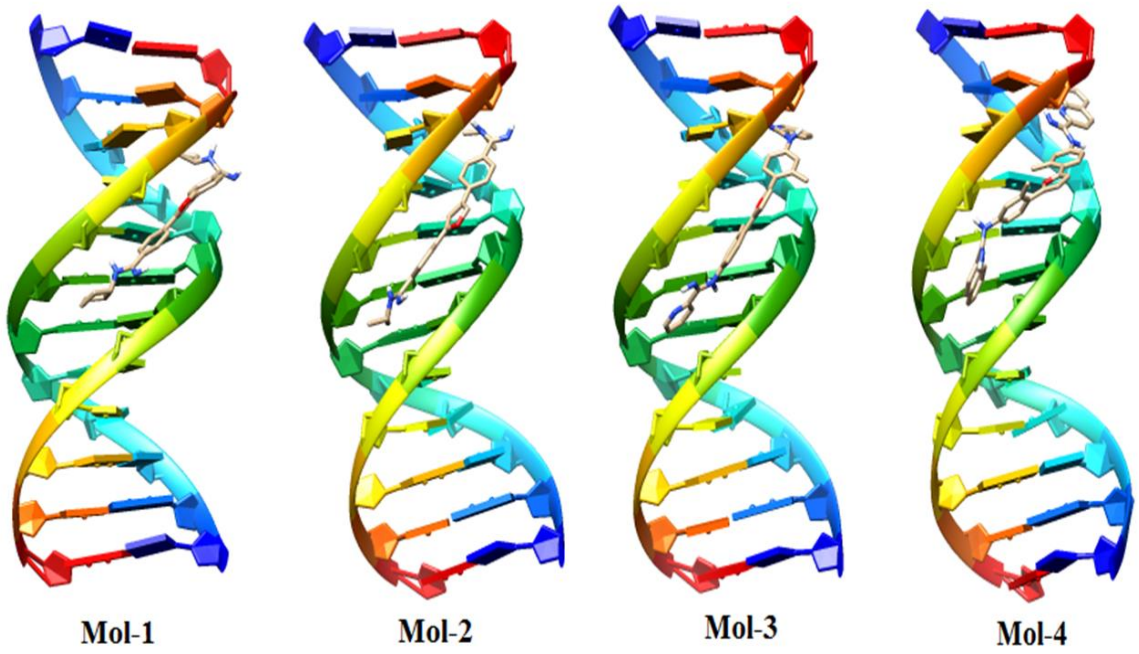


Figure 4.3: Figure showing best docked posed complexes for 4AH0 in 3D

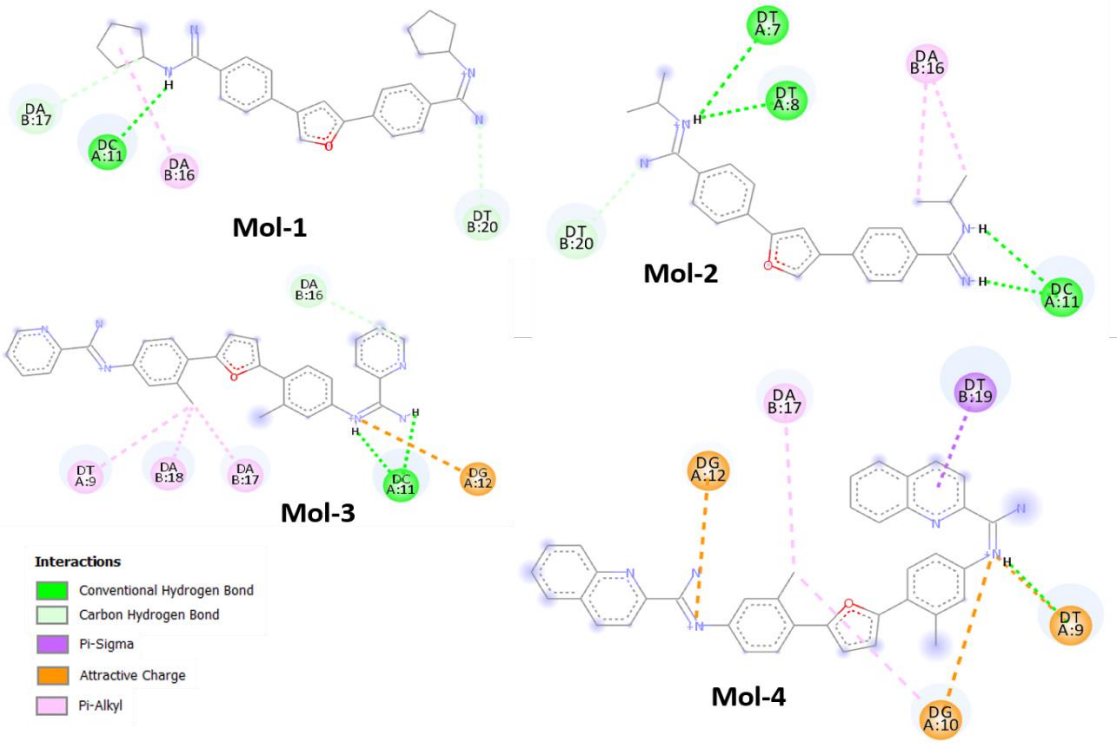


Figure 4.4: Figure representing the interaction profile for the best docked posed complexes in 2D

4.3.2.1 Hydrogen Bonding Analysis

There are few specific regions in the biomolecular systems that help in the quantification of the electron rich and electron deficient sites in the biomolecular complexes, and thus helps in the prediction of possible hydrogen bond formation sites within the system and the strength of the bond formed. **Figure 4.5** represents the atomistic binding site for each of the drug-DNA complex and **figure 4.6** represents the corresponding H-donor/acceptor clouded regions at binding sites. These clouded regions represent the atoms or group of atoms which have the tendency to donate/accept electrons for hydrogen bonding so as to achieve stability [31] during the docking calculations. **Table 4.3** summarizes the donor and the acceptor residues involved in the formation of hydrogen bond between the DNA and ligands along with the length of the hydrogen bond. In this section, the interactions established, specially the hydrogen bonds formed between the DNA bases and the ligands are presented as follows:

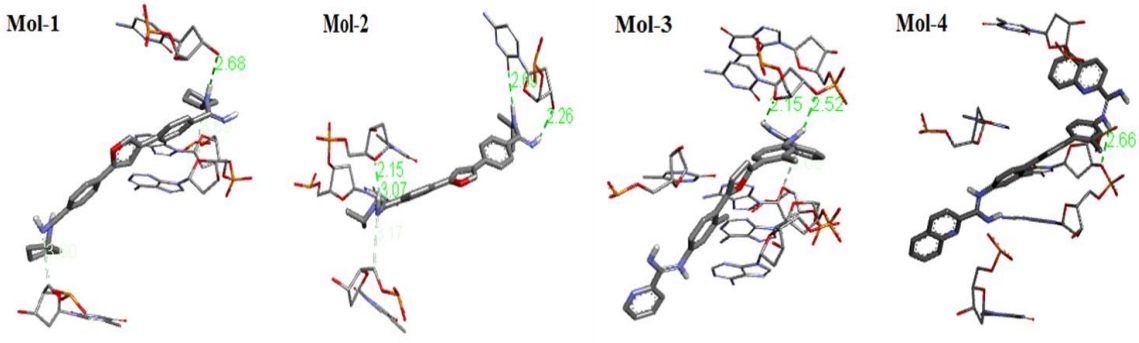


Figure 4.5: Figure showing H-bonds in best docked posed complexes

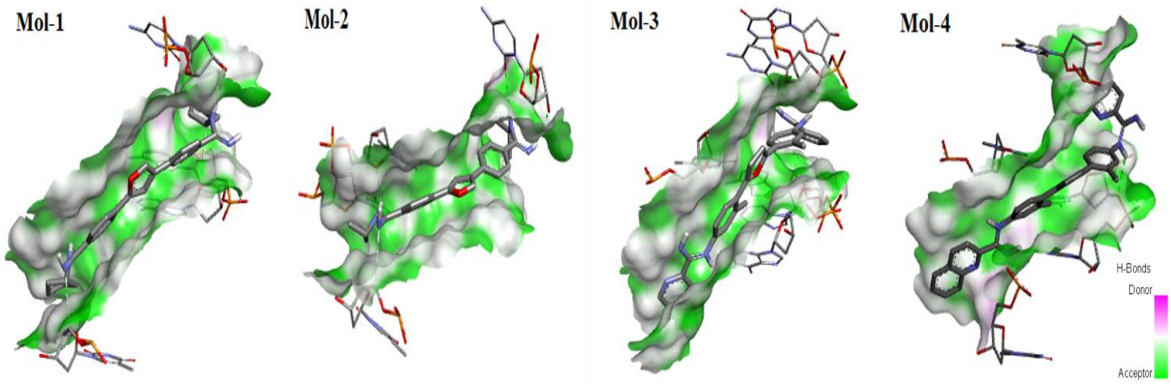


Figure 4.6: Figure representing the H-bond donor and acceptor regions in best docked posed complexes

Table 4.3: Table showing the donor and the acceptor species and H-bond length formed between the DNA and ligand atoms

S. No.	Complex	No. of H-Bonds	Interacting Species	H-Bond Length (Å)
1.	4AH0+Mol-1	1	H – DC11:O3’	2.679714
2.	4AH0+Mol-2	4	H – DT8:O4’	2.148102
			H – DT7:O2	3.067220
			H – DC11:O2	2.689916
			H – DC11:O3’	2.256256
3.	4AH0+Mol-3	2	H – DC11:O4’	2.153036
			H – DC11:O3’	2.520974
4.	4AH0+Mol-4	1	H – DT9:O3’	2.656984

4.3.3 Molecular Dynamics Simulation

Structural stability of biomolecules under dynamical conditions over a specifically mimicked environment for a pre-defined period of time can be studied via molecular dynamics simulations. Such studies hold significant importance in the

structure and dynamics of biomolecules owing to less experimental costs. Various parameters were studied and analyzed for their contribution to stability analysis of the drug-DNA complexes, and are discussed as follows:

4.3.3.1 Variation in Energy

The energy variation obtained from molecular dynamics simulation of the drug-DNA complexes for a time scale of 5000ps are shown below in **figure 4.7**. These variations in energies were obtained through gmx_energy programme inbuilt in GROMACS software suite [25]. The variations in the energy values of all the ligands are comparable and energy values for all lie in between 1500kJ/mol~2100kJ/mol. So, in the lieu of any direct evidence, other factors, viz., number of hydrogen bonds, RMSD, RMSF & variations in radius of gyration, would be decisive regarding the stable complex formation.

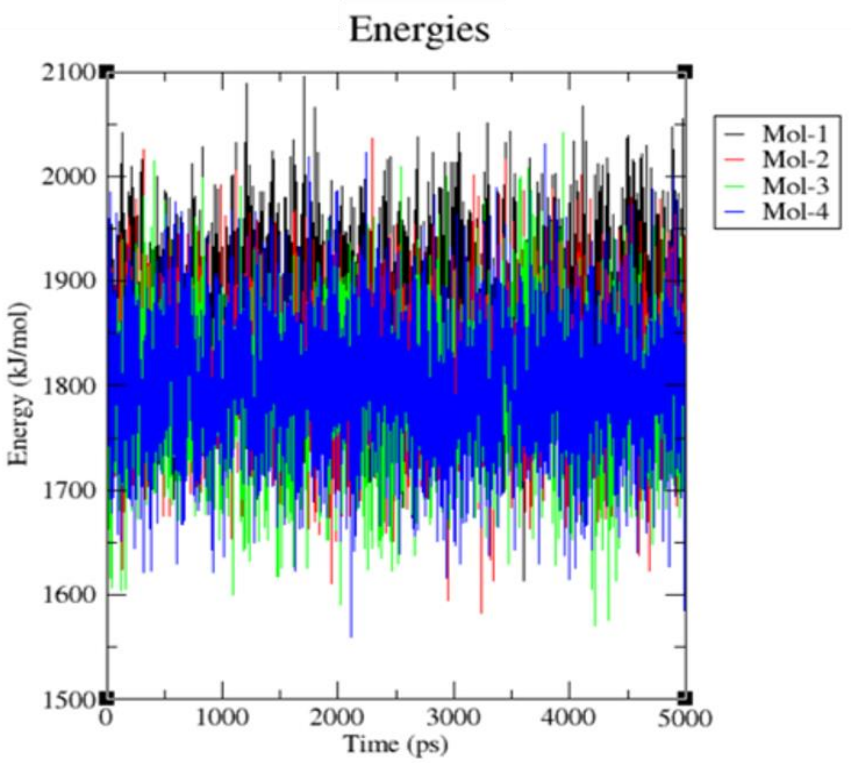


Figure 4.7: Figure representing variations in energy of the drug-DNA complexes

4.3.3.2 Variation in Hydrogen Bonds

The role of hydrogen bonding in stability of biomolecular systems is a well-established fact. These bonds play a crucial role in determining the stability of the drug-DNA complexes. **Figure 4.8** depicts the variations in the number of hydrogen bonds being formed during the molecular dynamics trajectory for 5000ps. The same figure not only reveals the hydrogen bonding forming pattern but also gives information about the average number of bonds formed after simulation; i.e., mol-1 forms 2 H-bonds, mol-2 also forms 2 H-bonds whereas mol-3 & mol 4 form 3 & 1 H-bonds, respectively (all numbers are average after the simulation). Above variations in energies were obtained through gmx_hbond programme inbuilt in GROMACS software suite [25].

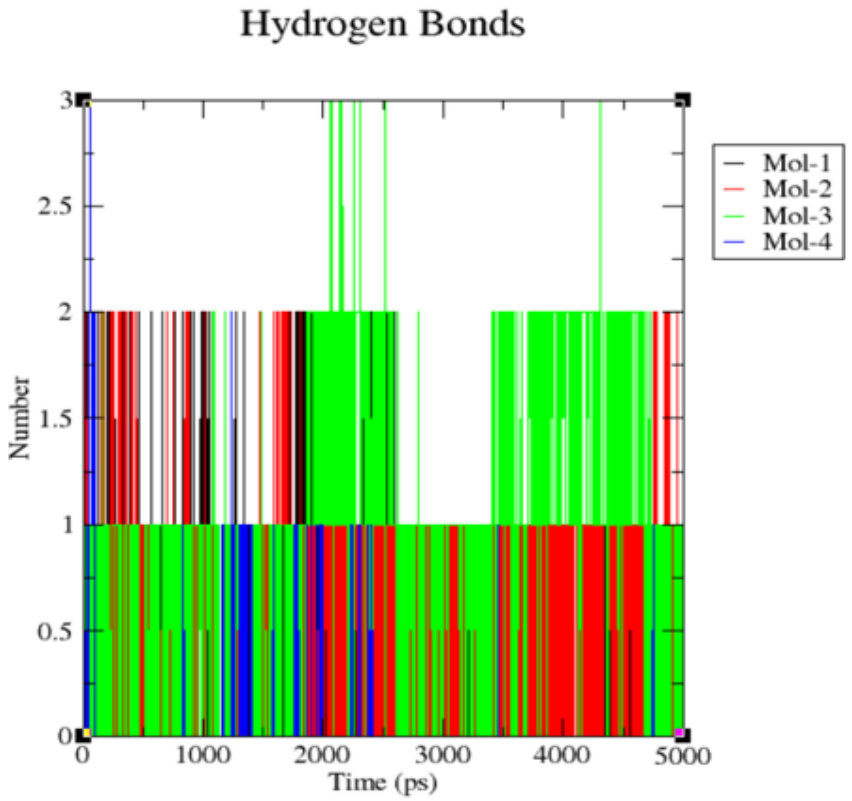


Figure 4.8: Figure representing variations in number of hydrogen bonds for drug-DNA complexes

4.3.3.3 Variation in Radius of Gyration

To understand the compactness of DNA & the dynamical stability of DNA-ligand complexes, radii of gyration values were determined. Clearly the average radius of gyration lies between 1.25nm to 1.50nm resp. The radius of gyration variations for each drug-DNA pair is shown in **figure 4.9**. This data reveals that mol-2, mol-3 & mol-4 with 4AH0 sequence form the most stable complexes whereas for mol-1, the radius remains intact till ~3500ps but after that time limit there are frequent perturbations in the conformations of the complex leading to distorting and decreased stability. These variations in energies were obtained through gmx_gyrate programme inbuilt in GROMACS software suite [25]. The variations in radii of gyration, can be summarized in **table 4.4** shown below:

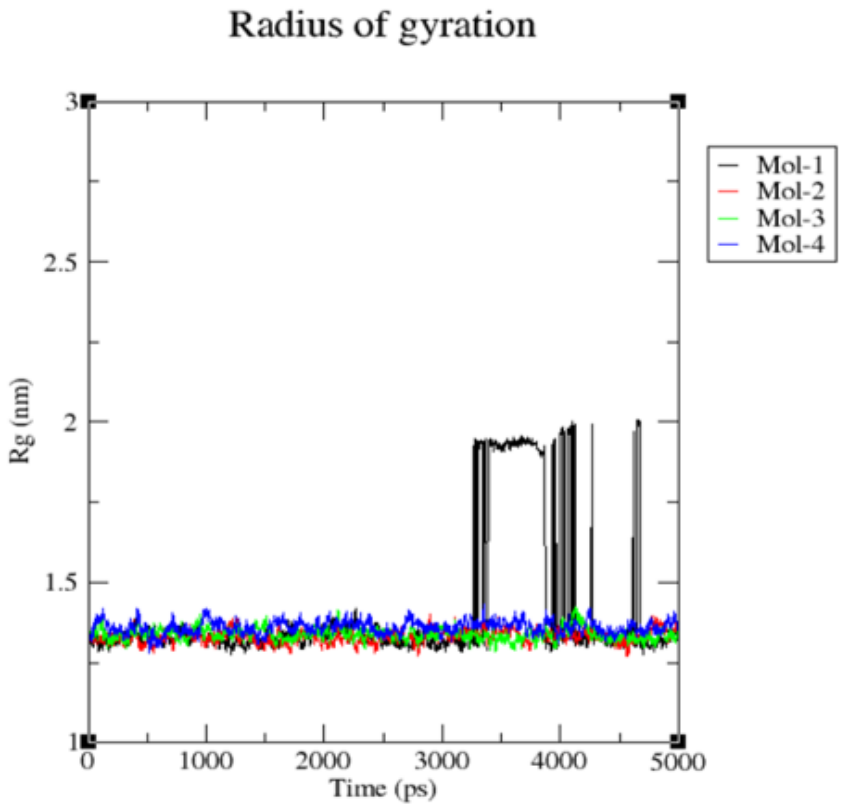


Figure 4.9: Figure representing variations in radius of gyration

Table 4.4: Table summarizing the **v**ariation in radius of gyration for drug-DNA Complexes

S. No.	Drug Molecule	4AH0 (avg. RG)
1.	Mol-1	Compact between 1.25nm to 1.50nm for first 3000ps, But between 3000~3500ps, 4000ps & 4500~5000ps there are several sharp fluctuations in the radii values **Remark:** non compact complex hence unstable
2.	Mol-2	Compact between 1.25nm to 1.50nm for entire 5000ps, **Remark:** Compact drug-DNA complex hence stable
3.	Mol-3	Compact between 1.25nm to 1.50nm for entire 5000ps, **Remark:** Compact drug-DNA complex hence stable
4.	Mol-4	Compact between 1.25nm to 1.50nm for entire 5000ps, **Remark:** Compact drug-DNA complex hence stable

4.3.3.4 Root Mean Square Deviation

The RMSD is the measure of the conformational stability of biological macromolecules. The plots for RMSD of all the drug- DNA complexes are represented in **figure 4.10**. From the variations shown below, the RMSD corresponding to mol-2, mol-3 & mol-4 with 4AH0 almost lie between 0 to 0.5nm, i.e., sharply around 0.25nm representing insignificant deviations. However, complex 4AH0+mol-1, shows deviations between 3000~3500ps, 4000ps & between 4500~5000ps and hence confirms for having had formed least stable complex with considerable deviations in the DNA double helix. The variations in RMSD, can be summarized in **table 4.5** shown below:

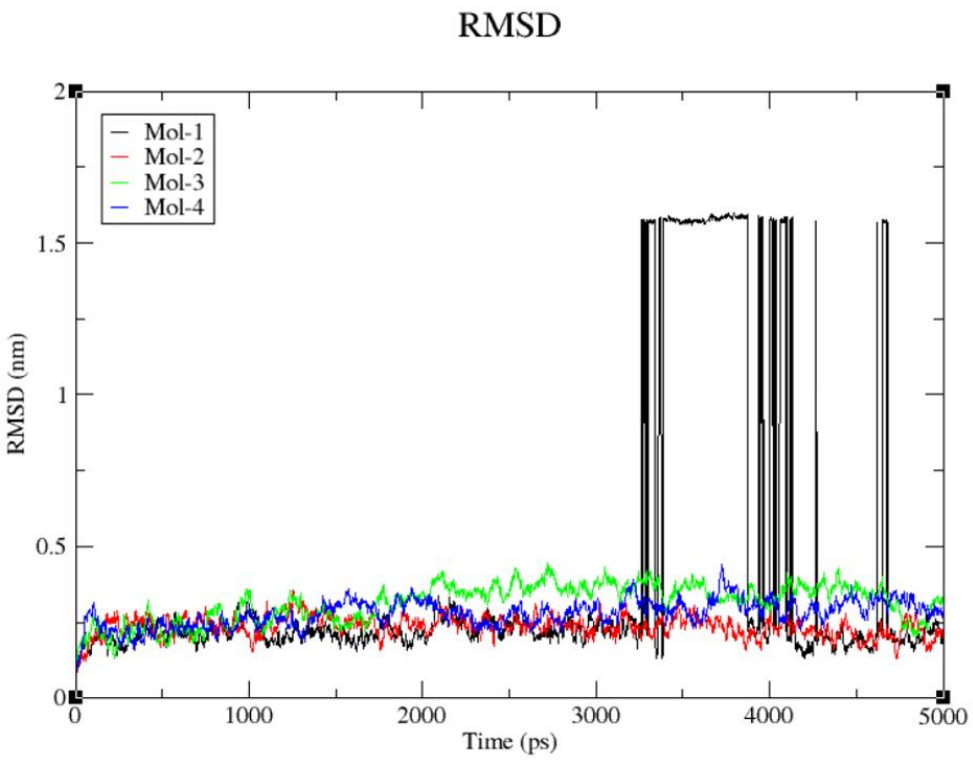


Figure 4.10: Figure representing RMSD for Drug-DNA Complexes

Table 4.5: Table summarizing the RMSD for drug-DNA Complexes

S. No.	Drug Molecule	4AH0 (avg. RG)
1.	Mol-1	Compact between 0 to 0.5nm for first 3000ps, But between 3000~3500ps, 4000ps & 4500~5000ps there are several sharp fluctuations in the RMSD values **Remark:** unstable complex
2.	Mol-2	Compact between 0 to 0.5nm for entire 5000ps, i.e., variations sharply take place around 0.25nm for entire 5000ps **Remark:** Compact drug-DNA structure hence stable
3.	Mol-3	Compact between 0 to 0.5nm for entire 5000ps, i.e., variations sharply take place around 0.25nm for entire 5000ps **Remark:** Compact drug-DNA structure hence stable
4.	Mol-4	Compact between 0 to 0.5nm for entire 5000ps, i.e., variations sharply take place around 0.25nm for entire 5000ps **Remark:** Compact drug-DNA structure hence stable

4.3.3.5 Root Mean Square Fluctuation

RMSF records the fluctuation of each amino acid base pair including the fluctuations in the flexible regions within the nucleic acid during the course of the MD simulation. Well-structured regions and loosely bound regions in DNA strands are distinguished by low and high root mean square fluctuations values, respectively. For small proteins, a fluctuation lying between 1~3 Å is acceptable. The graphs shown below in **figure 4.11** suggest that, 4AH0+mol-2, 4AH0+mol-3 & 4AH0+mol-4 had the least fluctuations, and especially around and beyond residue number 800 among, and hence owe towards the stability of the complexes. Here this range lies between 0~0.5 Å. Whereas,4AH0+mol-1 shows maximum deviations as seen from the corresponding graph.

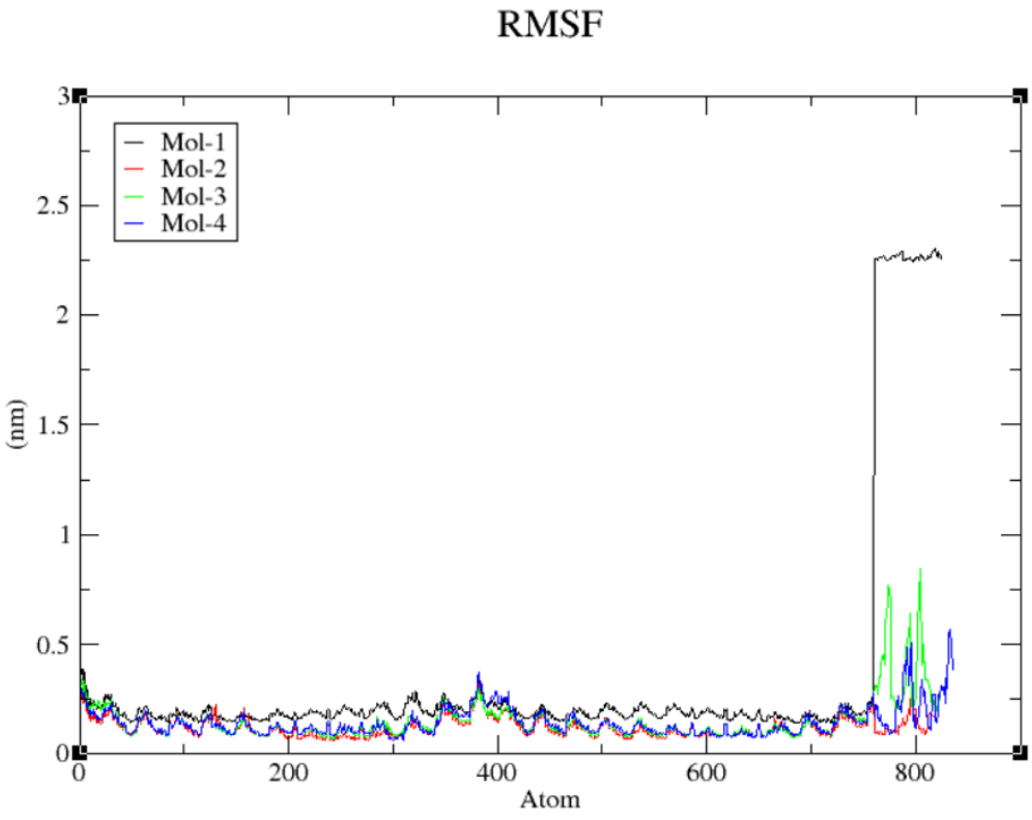


Figure 4.11: Figure representing RMSF for Drug-DNA Complexes

4.4 Conclusions

Current chapter addressed a problem regarding interaction and stability of drugs in the vicinity of DNA, and our findings successfully resolve this claim. The optimized drugs structures when docked to selected DNA generated best docked pose by finding the most stable binding site, these docked complexes when subjected to molecular dynamics for time scale evaluation of their stability yielded similar results that complemented the docking results. Our evaluated parameters viz., hydrogen bonding, radius of gyration, RMSD & RMSF were sufficient to support our claim.

Docking results revealed that all the selected ligands were bound to the minor groove of the DNA. Their binding site was AT-rich region, as preferred by minor groove binders. Docking calculations revealed that 4AH0 with mol-4 had least binding energy of -12.39 kcal/mol. This claims to maximum stability of the formed complex between mol-4 and 4AH0.

The results obtained through hydrogen bonding analysis suggested that donor and acceptor clouded regions essential for the formation of hydrogen bonds had comparable extent for acceptor regions for all the selected molecules. This also represented the DNA bases involved in the hydrogen bond formations.

Results obtained from the molecular dynamics simulation show the time dependence of the stability of selected molecules in the vicinity of DNA. Radius of gyration, RMSD and RMSF analysis were done from the trajectories obtained via MD studies; all these results favor the stable complex formation for mol-2, mol-3 & mol-4 with 4AH0. Radius of gyration & RMSD values reveal that ligands remain bound to the preferred binding positions of the DNA without any considerable deviations in

its minor groove; whereas the RMSF values reveal that the topological structure of DNA remains intact during the entire course of the simulation, inferring the stability of drug-DNA complexes.

Obtained results not only predicted the conformational stability of the DNA with the ligands but also gave a firm affirmation regarding the time dependence of the interaction and stability of drug-DNA complexes. They also add strong affirmations regarding the benchmark of various computational techniques.

References:

[1] Watson, J.D., Crick, F.H.C., *Nature*, **171**, 737-738 (1953).

[2] Pindur, U., Jansen, M., Lemster, T., *Current Medicinal Chemistry*, **12**, 2805-2847 (2005).

[3] Yadava, U., Yadav, S.K., Yadav, R.K., *Journal of Molecular Liquids*, **280**, 135-152 (2019).

[4] Wilson, W.D., Tanious, F.A., Mathis, A., Tevis, D., Hall, J.E., Boykin, D.W., *BioChimie.*, 90, 999-1014 (2008).

[5] Yadava, U., Yadav, S.K., Yadav, R.K., *DNA Repair*, **60**, 9-17 (2017).

[6] Pandey, A., Mishra, R., Shukla, A., Yadav, A.K., Kumar, D., ISAFBM Conference Proceedings (2019) ISBN: 978-93-5351-824-0.

[7] Pandey, A., Yadav, R., Shukla, A., Yadav, A.K., *Advanced Science, Engineering and* Medicine, **11**, 1-5 (2019).

[8] Srivastava, H.K., Chourasaia, H., Kumar, D., Sastry, G.N., *Chem. Inf. and Model.*, **51**, 558-571 (2011).

[9] Shukla, R., Shukla, H., Kalita, P., Sonkar, A., Pandey, T., Singh, D.B., Kumar, A., Tripathi, T., *Journal of Biomolecular Structure and Dynamics*, **36**, 2147-2162, (2018).

[10] Vora, J., Patel S., Sinha S., Sharma S., Srivastava A., Chhabria M., Shrivastava N., *Journal of Biomolecular Structure and Dynamics*, **37**, 131-146, (2018).

[11] Mishra, R., Gaur, A.S., Chandra, R., Kumar, D., *International Journal of Science, Technology and Society*, **3**, 11-27 (2017).

[12] Tidwell, R.R., Boykin, D.W., *Journal of Brazilian Chemical Society*, **13**, 414-460 (2002).

[13] Balendiran, K., Rao, S.T., Sekharudu, C.Y., Zon, G., Sundaralingam, M., *Acta Crystallogr.*, **51**, 190-198 (1998).

[14] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., *Nucleic Acids Research*, **28**, 235–242 (2000).

[15] Dassault Systèmes BIOVIA, Discovery Studio Visualizer, San Diego, Dassault Systèmes, (2020).

[16] Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M, Meng, E.C., Ferrin, T.E., *Journal of Computational Chemistry*, **25**, 1605-1612 (2004).

[17] Dubey, K.D., Chaube, A.K., Parveen, A., Ojha, R.P., *Journal of Biophysics and Structural Biology*, **2**, 47-54 (2010).

[18] Gaussian 09, Revision B.01, Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, G. A., Nakatsuji, H., Li, X., Caricato, M., Marenich, A. V., Bloino, J., Janesko, B. G., Gomperts, R., Mennucci, B., Hratchian, H. P., Ortiz, J. V., Izmaylov, A. F., Sonnenberg, J. L., Williams-Young, D., Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D., Zakrzewski, V. G., Gao, J., Rega, N., Zheng, G., Liang, W., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Throssell, K., Montgomery, J. A., Jr., Peralta, J. E., Ogliaro, F., Bearpark, M. J., Heyd, J. J., Brothers, E. N., Kudin,

K. N., Staroverov, V. N., Keith, T. A., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A. P., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Millam, J. M., Klene, M., Adamo, C., Cammi, R., Ochterski, J. W., Martin, R. L., Morokuma, K., Farkas, O., Foresman, J. B., Fox, D. J. Gaussian, Inc., Wallingford CT, 2009.

[19] Shukla, A., Yadav, R., Pandey, A., Awasthi, N., *Advanced Science, Engineering and Medicine.*, **11**, 1-5 (2019).

[20] Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J., *Journal Computational Chemistry*, **16**, 2785-91 (2009).

[21] Zhao, H., Caflisch, A., *Eur J Med Chem*, **91**, 4-14 (2015).

[22] Abraham, M.J., Murtola, T., Schulz, R., Pall, S., Smith, J.C., Hess, B., Lindahl, E., *SoftwareX*, 1–2, 19–25 (2015).

[23] Ponder, J.W., Case, D.A., *Adv. Prot. Chem.* **66,** 27-85 (2003).

[24] da Silva, A.W. S., Vranken, W.F., *BMC Res. Notes*, **5**, 367 (2012).

[25] Mark, P., Nilsson, L., *Journal Physical Chemistry-A*, **105**, 9954-9960 (2001).

[26] Darden, T., York, D., Pedersen, L., *J Chem Phys*, **98**, 10089-10092 (1993).

[27] Ryckaert, J.P., Ciccotti, G., Berendsen, H.J., *J Comput Phys.*, **23**, 327-341 (1977).

[28] Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M., *Journal of Computational Chemistry*, **18**, 1463-1472 (1997).

[29] Turner, P.J., XMGRACE, Version 5.1. 19. Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR (2005).

[30] Shastri, R., Awasthi, N., Kumar, D., Yadav, A.K., Roy, D., Goutam, S.P., Pandey, A., *Advanced Science, Engineering and Medicine*, **10**, 1-5 (2018).

[31] Aamir, M., Singh, V.K., Dubey, M.K., Meena, M., Kashyap, S.P., Katari, S.K., Upadhyay, R.S., Umamaheswari, A., Singh, S., *Front. Pharmacol.*, **9,**1038 (2018).

Chapter-5

Understanding Interactions of Some DNA Minor Groove Binders through Molecular Dynamics Simulations and MMPBSA Free Energy Calculations

Chapter-5

Understanding Interactions of some DNA Minor Groove Binders through Molecular Dynamics Simulations and MMPBSA Free Energy Calculations

5.1 Introduction

Interaction of DNA with various drugs has been of most significance in the field of drug discovery and the pharmaceutical industry. In the past few years, there has been rapid growth in the interest of studying DNA as the target molecule for the investigation of anticancer, antimicrobial, antibacterial and various other types of diseases. Also, these studies can be helpful in the quantification of such drugs and for the determination of new targets followed by the development and enhancement of existing drugs. DNA owes to be the key ingredient in most biological and genetic processes and therefore is an apt target for such studies [1,2]. Studying the structural properties of DNA, the processes involving mutation of genes, the reasons involving the origin of various deadly diseases and the action and mechanism of various drugs that target DNA for the same is of utmost importance in the current scenario.

Interaction between drug and DNA occurs in many ways, viz., intercalation, covalent binding or cross-linking, DNA cleaving, groove binding, [3] etc. Groove binding in DNA takes place via two modes, viz., minor groove (width being 12Å) binding and major groove (width being 22Å) binding whereas intercalation involves the insertion of the drug ligand in between the DNA base pairs [4,5]. The complex formation between DNA and drug leads to manipulation of its thermodynamic

parameters [6] and thus studies involving complex formation are of great importance and are in current trend as they may lead the researches towards more rational and target-oriented drug design. Apart from drug design and enhancement of its existing properties, computational methods can also be used to model the enzyme catalysis reactions involved in the metabolism of drugs [7-9].

In this chapter, a few di-cationic DNA minor groove binders (2,5-diaryl furans), which are supposed to have antimicrobial drug potency [10], are studied for their interactions, stability, binding affinities and complex forming tendencies with DNA using machine learning techniques.

Molecular docking studies help in finding the best docking site and corresponding docking score followed by corresponding binding site energy. Docking studies performed over such ligands suggested the formation of stable complexes [11]. However, molecular dynamics simulations carried out for drug-receptor complexes helps in determining the structural stability of the complexes over time. Free energy calculations were performed to predict the stable complex formation tendencies of these analogs with DNA with time [12]. Various computational studies have proven computational resources to be of significant importance in this field of computational biology and drug discovery [13, 14].

5.2 Methodology

5.2.1 System Selection & Preparation

The ligands were selected from literature [10]. A generalized chemical structure is shown in **figure 5.1** followed by a table informing about the substitutions

made in **table 5.1**. The DNA sequences {1DNE [15], 195D [16]} were obtained from Protein Data Bank [17]. Their structural data including their specific nucleic acid bases sequences and PDB ID's, is mentioned in **table 5.2**. Geometry optimization of ligands was carried out using Gaussian 09 software [18] by applying B3LYP hybrid density functional at 6-31G** basis set for them to attain a local potential minimum. Water molecules, from downloaded DNA sequences were removed using UCSF Chimera software [19] to prepare them for docking and molecular dynamics simulations.

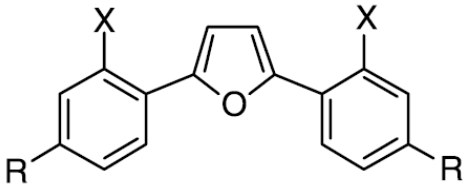


Figure 5.1: Figure showing generalized chemical structures of the ligands with position for attachment of substituents

Table 5.1: Table representing the chemical structures substitutions for the selected ligands

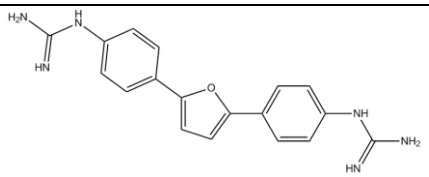
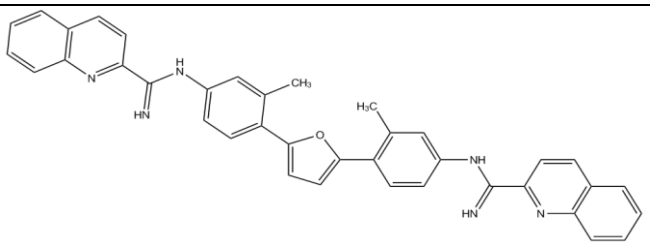
S. No.	Substitution R	Substitution X	Chemical Structure	Identification
1.	$NHC(=NH)NH_2$	H		mol-1
2.	NHC(=NH)-2-Qu	CH_3		mol-2

Table 5.2: PDB ID's of selected DNA sequences

S. No.	PDB Id.	DNA Sequence	Experimental Parameters
1.	1DNE	5'-CGCGATATCGCG-3'	• Total Structure Weight: 7757.25 • Atom Count: 517 • Residue Count: 24 • Unique nucleic acid chains: 1 • Resolution: 2.4 Å
2.	195D	5'-CGCGTTAACGCG-3'	• Total Structure Weight: 7757.25 • Atom Count: 517 • Residue Count: 24 • Unique nucleic acid chains: 1 • Resolution: 2.3 Å

5.2.2 Molecular Docking Studies

Molecular Docking was carried out using Autodock4 software [20]. Gasteiger charges were added to the drug-DNA complex using Autodock Tools (ADT) before beginning the docking calculations computationally. A grid box, having various dimensions along the three coordinate axes, was prepared for each drug-DNA complex which enclosed the macromolecule, here DNA. This helped the ligand (drug)

in searching the most appropriate binding site while the docking calculations were done. Docking calculations were set up using Lamarckian Genetic Algorithm (LGA). A 20 LGA simulation with a maximum cycle of 2500000 energy evaluations was carried out for each of the drug-DNA complex. The docked pose with the least binding affinity was extracted and aligned with the receptor (DNA) for further analysis.

5.2.3 Molecular Dynamics Simulation

Molecular Dynamics Simulation is a cutting-edge computational technology for the precise atomistic simulation of biomolecular systems which mimics the natural environment during the simulation and hence generates accurate results [21]. Applications of molecular dynamics simulation to drug discovery and complex stability followed by atomistic insights into the binding of ligands, unwinding of proteins and conformational changes over time, is expanding and has attained utmost importance due to lack of experimental resources [22,23]. In the current research work, molecular dynamics simulations were carried out using GROMACS 5.0.4 (Groningen Machine for Chemical Simulations) software package [24].

A total of 4 ligand-DNA complexes were created, post docking simulations for molecular dynamic calculations viz., (1DNE: mol-1 & mol-2; 195D: mol-1 & mol-2). All these drug-DNA complexes were put to molecular dynamics for 5000ps time scale simulation. CHARMM27 force field was used to generate the topology for all the selected DNA sequences [25]. However, Swiss PARAM webserver [27] was put into use for generating topology of selected ligands. The Ligand-DNA complex was solvated in a box of varying dimensions having P1 space group using TIP3P water model at 298K [27]. Sodium ions were then added to the box already solvated

containing the DNA-ligand complex by randomly replacing the water molecules to neutralize the charge of system. Particle Mesh Ewald (PME) was used to handle long-range electrostatic interactions in periodic boundary conditions [28]. Energy minimization of the whole system was carried out in 25000 steps using Steepest Descent leap- Frog Integration Method followed by NVT ensemble equilibration at a constant temperature of 300K for 50s using Berendsen thermostat [29]. The system was then equilibrated with NPT ensemble at a constant pressure of 1atm in 25000 steps using steepest descent leap-frog integrator [29]. All the bonds involving hydrogen atoms were constrained using the LINCS algorithm [30].

Various analyses were then done to judge the potency of the antimicrobial agents interaction, binding and stability with DNA. The gmx_energy, gmx_gyrate, gmx_hbond, gmx_rmsd, gmx_rmsf programmers were used to calculate comparative variations in energies, radii of gyration, hydrogen bonds being formed, root mean square deviations and root mean square fluctuations. XMgrace software was used to plot all graphs [31].

5.2.4 Free Energy Calculations

Accurate estimation of binding free energies is of significant importance in structural biology and computer-based drug design [12] MMPBSA method is based upon molecular mechanics and continuum solvent model [32]. In this method the receptor and ligand complex are assumed to split into two components: association in the gaseous phase, generally vacuum and dissolution in the aqueous phase. Classical force fields are implemented in the energy evaluation of single point gaseous phase

whereas implicit solvent models are applied for the solvation energy terms; this is the peculiarity of this method.

In MMPBSA method the free energy of a ligand receptor system is estimated by following equations [32,33]:

$$\Delta G = \Delta H - T\Delta S \quad (5.1)$$

Where,

$$\Delta H = \Delta E_{mm} + \Delta G_{solv} - T.\Delta S \quad (5.2)$$

$$\Delta E_{mm} = \Delta E_{cov} + \Delta E_{elect} + \Delta E_{vdW} \quad (5.3)$$

$$\Delta E_{cov} = \Delta E_{bond} + \Delta E_{angle} + \Delta E_{tor} \quad (5.4)$$

$$\Delta G_{solv} = \Delta G_{pol} + \Delta G_{apol} \quad (5.5)$$

Here, ΔE_{mm} is the MM energy term which includes contribution from bonded interactions viz., covalent (ΔE_{cov}), electrostatic (ΔE_{elect}) and van-der Waal's (ΔE_{vdW}) interactions respectively. Further the covalent bonding interaction free energy (ΔE_{cov}) comprises of energy changes in bond terms (ΔE_{bond}), angle terms (ΔE_{angle}) and torsional terms (ΔE_{tor}) respectively. Whereas ΔG_{solv} comprises of polar (ΔG_{pol}) and non-polar (ΔG_{apol}) solvation free energies respectively. The entropy term is the most difficult to compute amongst all the energy terms, and it is often approximated with a normal mode method. However, in the current work the entropy contribution to free energy was forbidden owing to the expensive computational cost [34]. The Polar solvation free energy (ΔG_{pol}) is obtained by solving the Poisson- Boltzmann (the MMPBSA method) equation or by solving the generalized Born model (the MMGBSA method), whereas the non-polar solvation free energy (ΔG_{apol}) is obtained

from the Solvent accessible surface area (SASA model) [35]. In the current research work, G-MMPBSA method was applied to calculate the binding free energies of all receptor ligand complexes [36].

5.3 Results and Discussion

This study was aimed to identify new leads, targeting binding affinity, structural stability and drug likeness and applicability for antimicrobial ligands with DNA. The results obtained through various computational calculations are summarized and discussed as follows:

5.3.1 Molecular Docking

A total of two ligands which were claimed to possess antimicrobial tendencies, were docked to two DNA sequences in search of best docked posed complex, as shown in **figure 5.2**. The docking results, corresponding to each selected DNA sequences are summarized below in **table 5.3** and **table 5.4** respectively. Docking calculations revealed that 1DNE formed most stable complex with second ligand (mol-2) having binding energy of -13.54 kcal/mol. 195D also formed its best docked posed complex with second ligand (mol-2) having binding energy of -11.72 kcal/mol. Amongst all the obtained best docked posed complexes 1DNE had the maximum binding affinity with mol-2 and thus claiming itself to have formed the most stable drug-DNA complex. The docked pose corresponding to each drug-DNA complex is shown in **figures 5.3 & 5.5** respectively. These figures reveal that both the drugs bonded themselves to the minor groove of each DNA sequence, whereas **figures 5.4 & 5.6** represent the interaction profile in 2D which reveals various types of interactions taking place between the DNA bases and the selected ligands, viz.,

conventional hydrogen bonds, carbon hydrogen bonds, π-σ bonds, π-alkyl bonds and presence of Van der Waals & some attractive charges, etc. These interactions eventually lead to the stabilization or de-stabilization of the drug-DNA complexes formed.

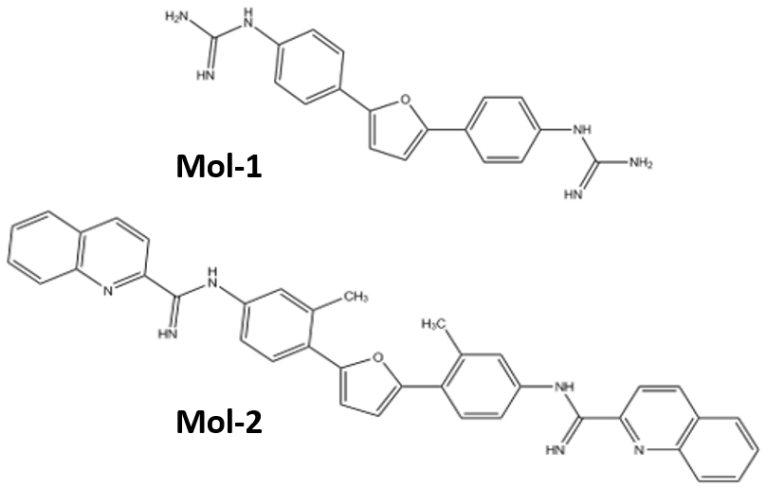


Figure 5.2: Figure showing chemical structures of the selected ligands

Table 5.3: Table representing docking results for 1DNE sequence

S. No.	Molecule	Substitution		Exp. (ΔT_m)	1DNE				
		R	X	(°C)	Binding Free Energy (kcal/mol)	Inhibition Constant (nM)	Docking RMSD (Å)	No. of Conformers	Docking Temp. (K)
1.	mol-1	$NHC(=NH)NH_2$	H	10.8	-11.17	6.49	72.575	3	298.15
2.	mol-2	NHC(=NH)-2-Qu	CH_3	10.8	-13.54	119.86	83.548	1	298.15

*all energies are in kcal/mol

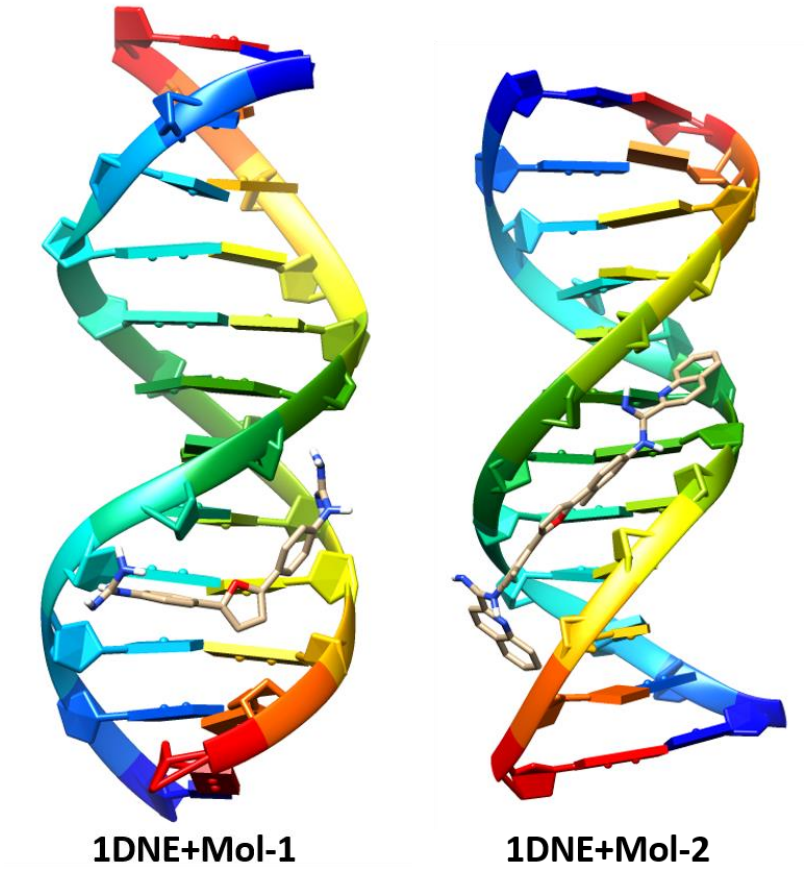


Figure 5.3: Figure showing best docked posed complexes for 1DNE in 3D

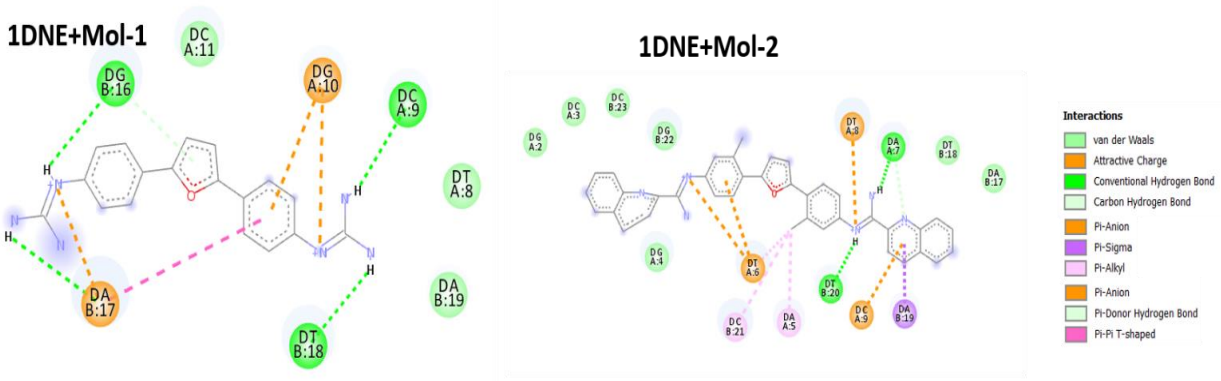


Figure 5.4: Figure representing the interaction profile for the best docked posed complexes in 2D for 1DNE

Table 5.4: Table representing docking results for 195D sequence

S. No.	Molecule	Substitution		Exp. (ΔT_m)	1DNE				
		R	X	(°C)	Binding Free Energy (kcal/mol)	Inhibition Constant (nM)	Docking RMSD (Å)	No. of Conformers	Docking Temp. (K)
1.	mol-1	$NHC(=NH)NH_2$	H	10.8	-11.38	4.52	25.643	2	298.15
2.	mol-2	NHC(=NH)-2-Qu	CH_3	10.8	-11.72	2.56	24.023	1	298.15

*all energies are in kcal/mol

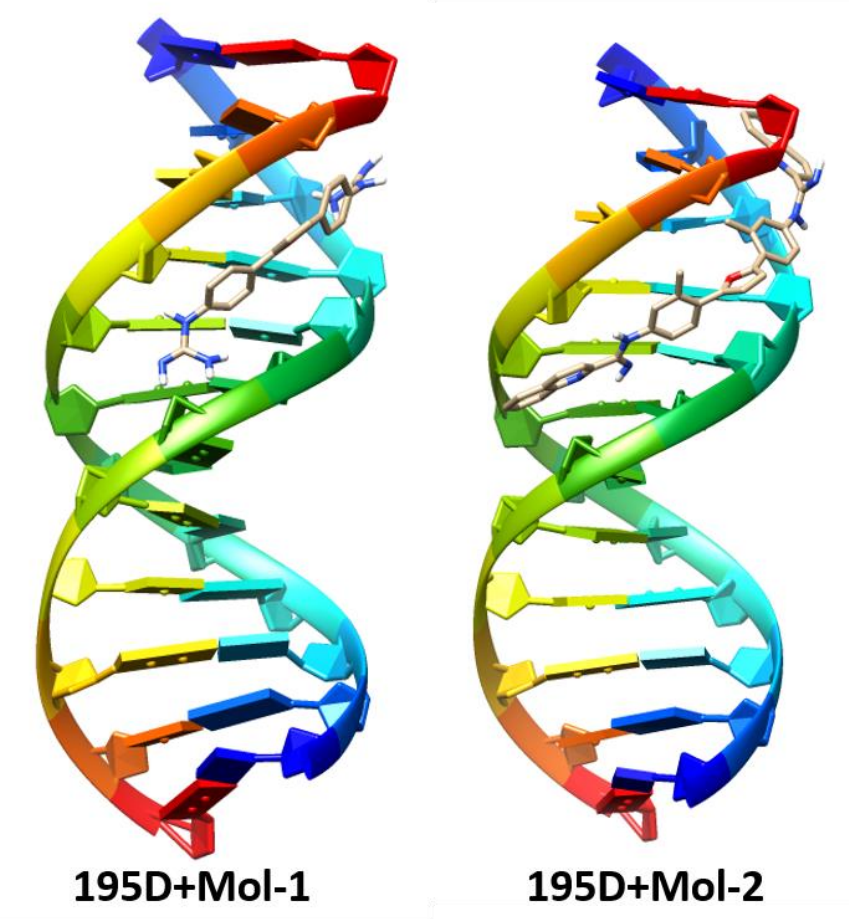


Figure 5.5: Figure showing best docked posed complexes for 195D in 3D

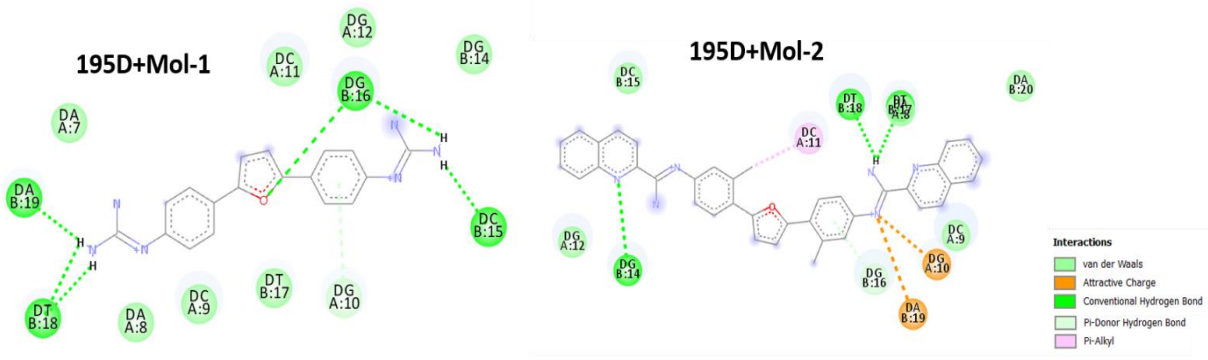


Figure 5.6: Figure representing the interaction profile for the best docked posed complexes in 2D for 195D

5.3.1.1 Hydrogen Bonding Analysis

There are few specific regions in the biomolecular systems that help in the quantification of the electron rich and electron deficient sites in the biomolecular complexes, and thus helps in the prediction of possible hydrogen bond formation sites within the system and the strength of the bond formed. **Figures 5.7** & **5.8** represent the atomistic binding site for each of the drug-DNA complex along with corresponding H-donor/acceptor clouded regions at those binding sites. These clouded regions represent the atoms or group of atoms which have the tendency to donate/accept electrons for hydrogen bonding so as to achieve stability [45] during the docking calculations. **Table 5.5** summarizes the donor and the acceptor residues involved in the formation of hydrogen bond between the DNA and ligands along with the length of the hydrogen bond.

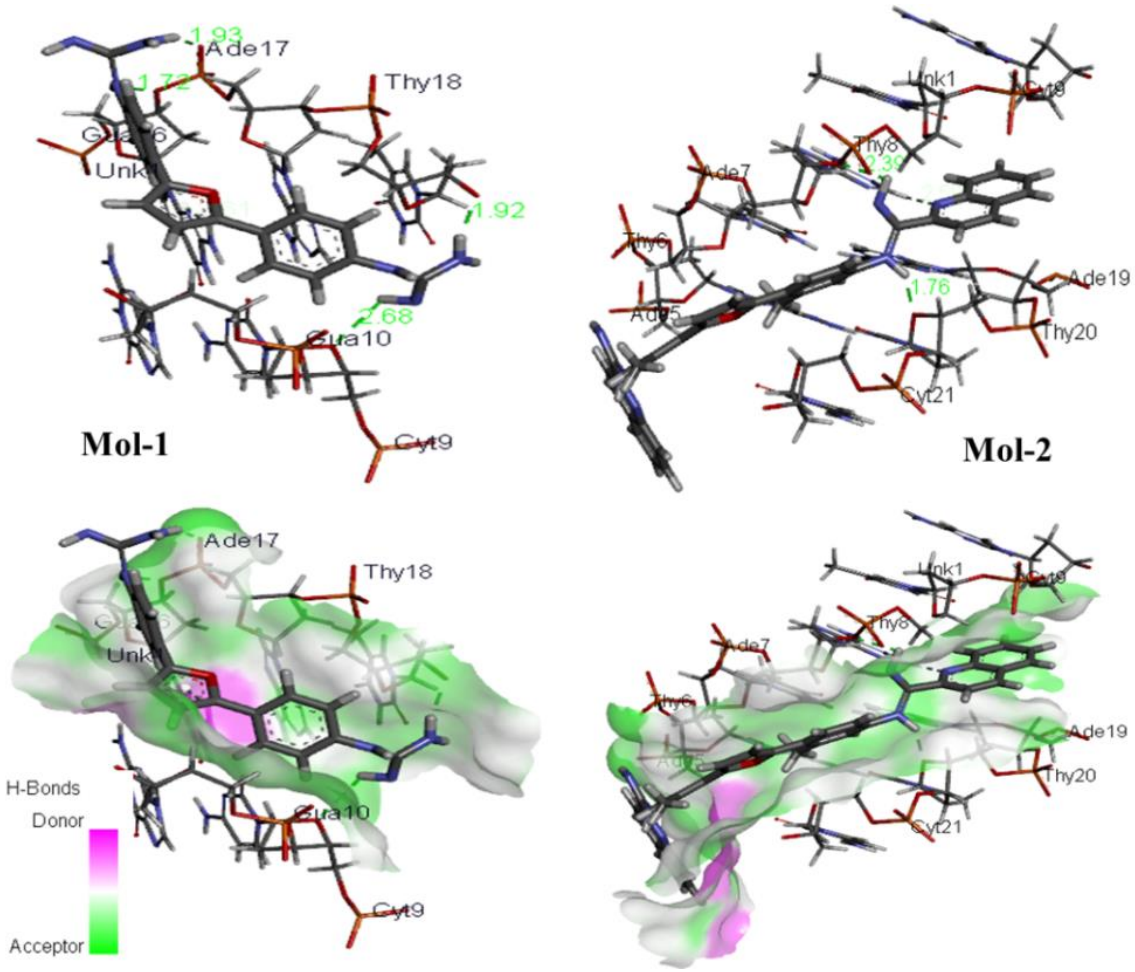


Figure 5.7: Figure representing the H-bond donor and acceptor regions in best docked posed complexes for 1DNE

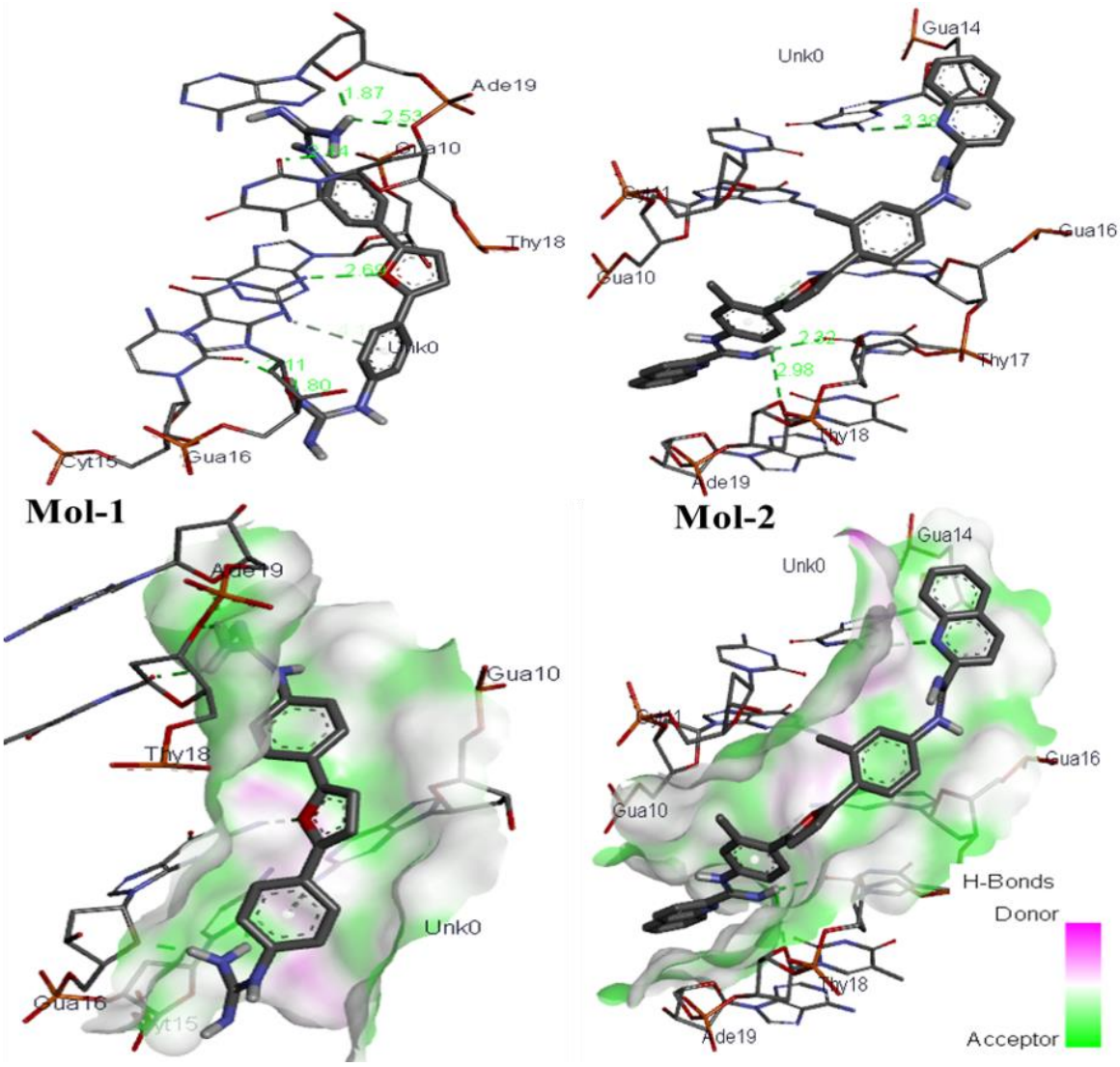


Figure 5.8: Figure representing the H-bond donor and acceptor regions in best docked posed complexes for 195D

Table 5.5: Following table represents the donor and the acceptor residues involved in the formation of hydrogen bond between the DNA and ligand atoms

S. No.	DNA Sequence-Drug	No. of H-Bonds Formed	Interacting Species	Donor Species	Accepting Species	H-Bond Length (Å)
1.	1DNE (Mol-1)	4	LIG1:H16 - DC9:O4'	LIG1:H16	DC9:O4'	2.684142
			LIG1:H19 - DT18:O3'	LIG1:H19	DT18:O3'	1.923581
			LIG1:H32 - DA17:O1P	LIG1:H32	DA17:O1P	1.930623
			LIG1:H27 - DG16:O3'	LIG1:H27	DG16:O3'	1.721654
2.	1DNE (Mol-2)	2	LIG1:H1 - DA7:O3'	LIG1:H1	DA7:O3'	2.387511
			LIG1:H - DT20:O4'	LIG1:H	DT20:O4'	1.758792
3.	195D (Mol-1)	6	LIG0:H - DA19:O4'	LIG0:H	DA19:O4'	1.865318
			LIG0:H - DT18:O3'	LIG0:H	DT18:O3'	2.534521
			LIG0:H - DT18:O2	LIG0:H	DT18:O2	2.139825
			DG16:N2 - LIG0:O	DG16:N2	LIG0:O	2.687842
			LIG0:H - DC15:O2	LIG0:H	DC15:O2	2.109972
			LIG0:H - DG16:O4'	LIG0:H	DG16:O4'	1.800594
4.	195D (Mol-2)	3	LIG0:H - DT18:O4'	LIG0:H	DT18:O4'	2.981288
			LIG0:H - DT17:O2	LIG0:H	DT17:O2	2.323480
			DG14:N2 - LIG0:N	DG14:N2	LIG0:N	3.375634

5.3.2 Molecular Dynamics Simulation

Structural stability of the complexes over a pre-defined period can be elucidated via molecular dynamics simulations followed by certain analysis viz., variations in energy, variations in radius of gyration, variations in number of hydrogen bonds, root mean square deviations and root mean square fluctuations; during the

course of the simulation. Each of the above-mentioned variations are discussed separately as follows:

5.3.2.1 Energy Variations

The GROMACS energy variation obtained from molecular dynamics simulation of 1DNE and 195D for a time scale of 5000ps are shown in **figure 5.9**. Above variations in energies were obtained through gmx_energy programme inbuilt in GROMACS software suite [24]. The contributions of various kinds of energies obtained are tabulated in **table 5.6**. The variations in the energy values of 1DNE are minimum (variations up to a lesser extent) as compared to that of 195D. This indicates the formation of comparatively stable complexes between 1DNE: mol-1 & mol-2 respectively. And hence the activity of selected drugs for 1DNE DNA sequence act as better antimicrobial agents than that of 195D [9]. However, confirmation of this fact requires analysis from various other perspectives and this lets the door opened for further research.

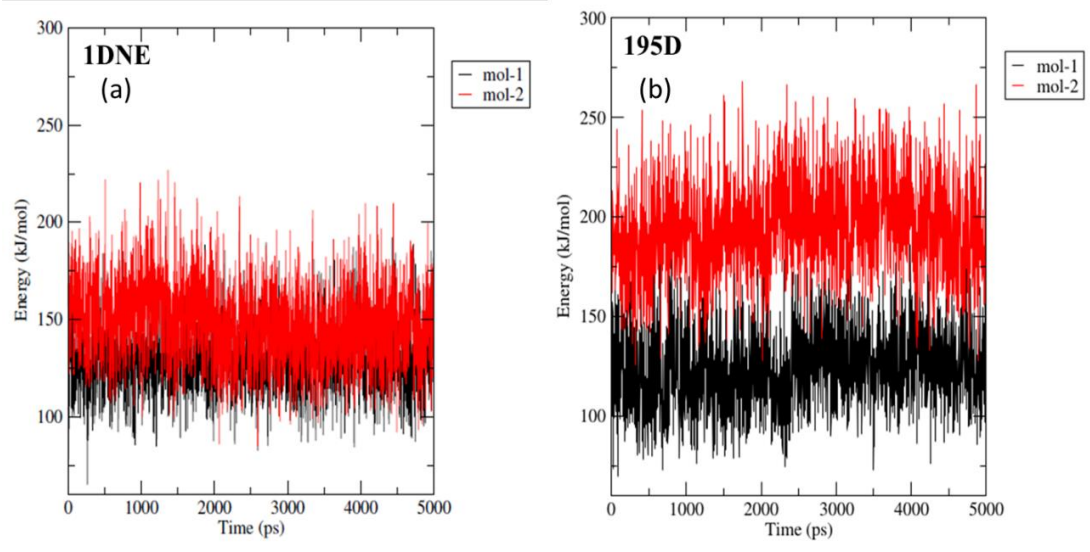


Figure 5.9: Figure representing variations in energy of the drug-DNA complexes

Table 5.6: Various energy contributions during variations in energy

Drug-DNA Complex	Bond (X 10^2)	Angle (X 10)	UB (X 10^3)	Proper Dihedral (X 10^3)	Improper Dihedral (X 10)	LJ-14 (X 10^3)	Coulomb-14 (X 10^4)	Potential (X 10^5)
1DNE (Mol-1)	3.57	9.37	1.59	3.54	2.44	1.81	-1.36	-3.13
1DNE (Mol-2)	3.93	7.68	1.57	3.38	2.79	2.00	-9.89	-3.18
195D (Mol-1)	3.31	7.26	1.31	3.34	1.62	1.75	-1.36	-3.36
195D (Mol-2)	3.66	6.58	1.38	3.43	2.65	1.93	-1.00	-3.34

5.3.2.2 Variation in Radius of Gyration

To understand the compactness and dynamic stability of DNA-ligand complexes, radii of gyration values are determined [37, 38]. Radius of gyration was calculated through gmx_gyrate programme inbuilt in GROMACS software suite [24]. The variation in radius of gyration of DNA- ligand complexes can be seen from **table 5.7**. The avg. radius of gyration for 1DNE and 195D are 1.2nm and 1.3nm~1.4nm, respectively. However, from the above table and figure, it can be concluded that 1DNE shows fewer variations in its radius of gyration than the 195D and therefore this confirms the formation of stable complexes for 1DNE with both mol-1 as well as mol-2 than that of 195D. The collective Radius of Gyration variations are shown in **figure 5.10**. These variation results reveal that 1DNE DNA sequence remains most compact for the whole 5000ps molecular dynamics simulation apart from a few sharp fluctuations at around 2000ps and between 4500ps~4750ps respectively and hence confirms the stability of the complex [35].

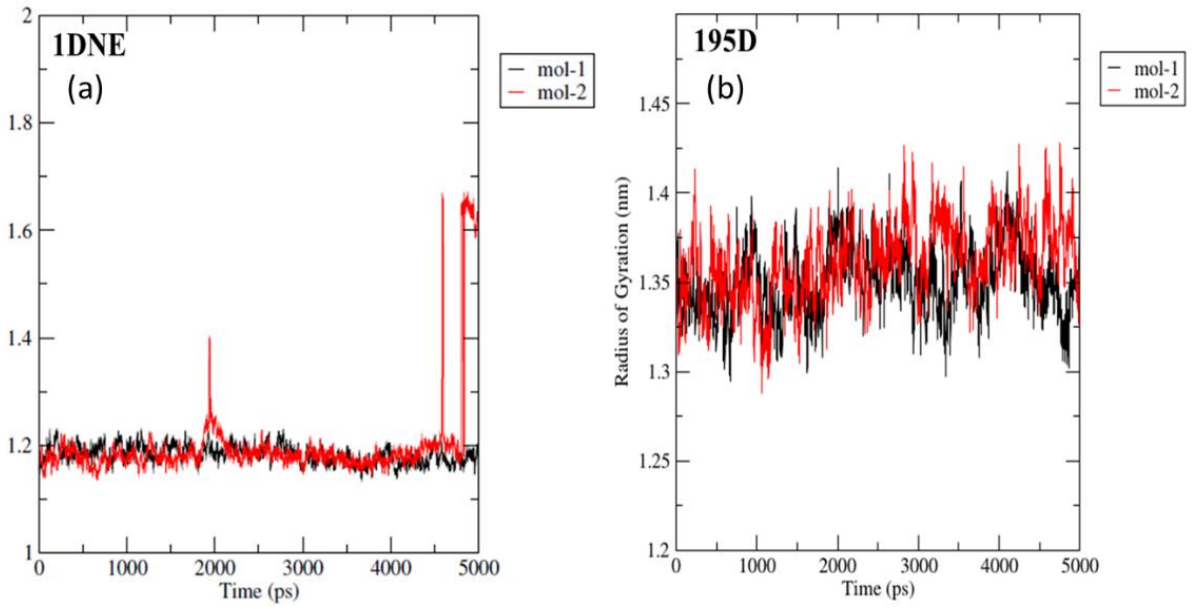


Figure 5.10: Figure representing variations in radius of gyration for drug-DNA complexes

Table 5.7: Variation in radius of gyration for drug-DNA Complexes

S. No.	Drug Molecule	1DNE (avg. RG)	195D (avg. RG)
1.	Mol-1	1.2nm, consistent throughout the simulation **Remark:** Compact DNA structure; Most stable complex	1.3nm~1.4nm, with numerous frequent sharp fluctuations **Remark:** unstable complex
2.	Mol-2	1.2nm, with sharp fluctuation at 2000ns, 4500ns and ~4750ns resp. **Remark:** Compact DNA structure; stable complex within 4500ns	1.275nm~1.425nm with numerous frequent sharp fluctuations **Remark:** unstable complex

5.3.2.3 Variation in Number of Hydrogen Bonds

Hydrogen bonds play a crucial role in determining the binding affinity as well as binding specificity in DNA-ligand interactions. **Figure 5.9** depicts the variations in

the number of hydrogen bonds being formed and broken during the molecular dynamics trajectory for 5000ps. **Figure 5.11** shows that 1DNE forms a maximum of 3 H-bonds with mol-1 and a maximum of 5 H-bonds with mol-2 respectively. However, 195D forms a maximum of 5 H-bonds with mol-1 and a maximum of 4 H-bonds with mol-2 respectively. This suggests the stronger interaction between DNA-ligand complexes of 195D. Thus, the overall results of hydrogen bonds suggest that despite the presence of lesser number of hydrogen bonds ligands attached to DNA sequence 1DNE could act as better antimicrobial agents than that of 195D [37] on account of previous results. However, concerning hydrogen bonding, 195D forms best complexes with its ligands.

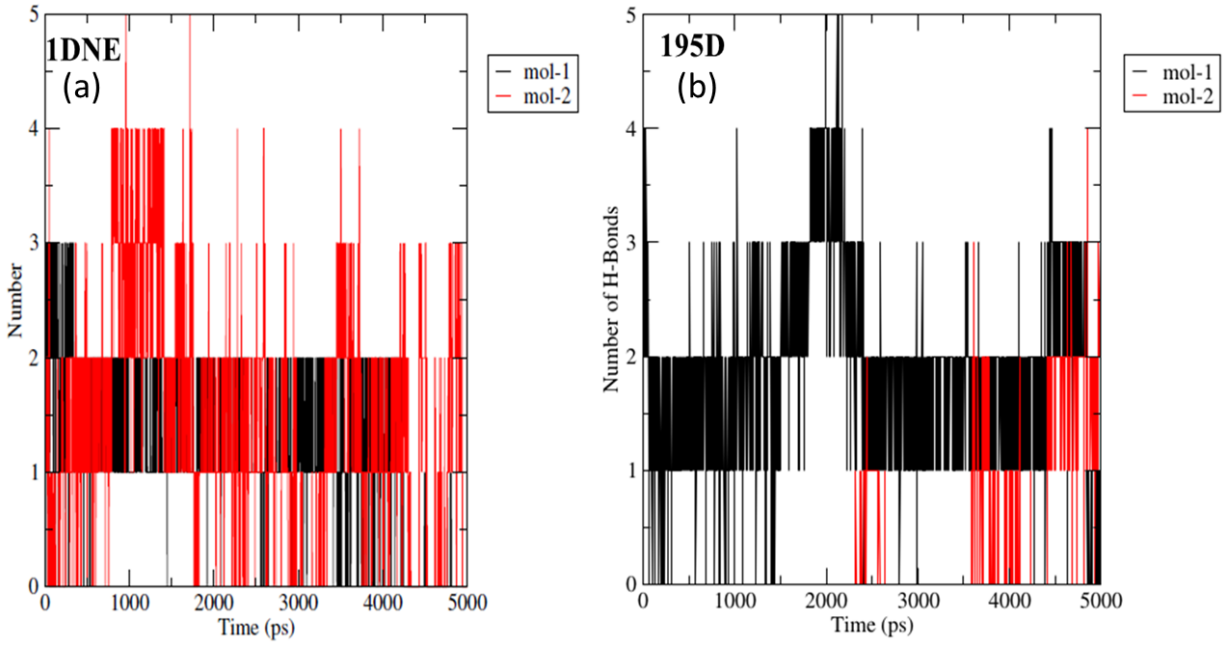


Figure 5.11: Figure representing variations in number of hydrogen bonds for drug-DNA complexes

5.3.2.4 Root Mean Square Deviation

The RMSD can be treated as the measure of the conformational stability of the complex during the course of molecular dynamics simulation [37]. Here, we are interested in examining the conformational stability and dynamical effect of ligand-DNA complexes and to determine the most stable complex. The plots for RMSD of all the drug- DNA complexes are represented in **figure 5.12**. From the variations shown below, the RMSD for all ligand complexes corresponding to DNA sequence 1DNE are almost consistent apart from that for 1DNE: mol-2, which shows a kink between 4500ps and 4750ps and thus destabilizes the complex. However similar pattern was also observed for 195D sequence which too had frequent sharp kinks throughout the 5000ps simulation and thus claiming to the stability of all its complexes. Whereas 1DNE: mol-1, owes to the least deviations and hence confirms for having had formed most stable complex.

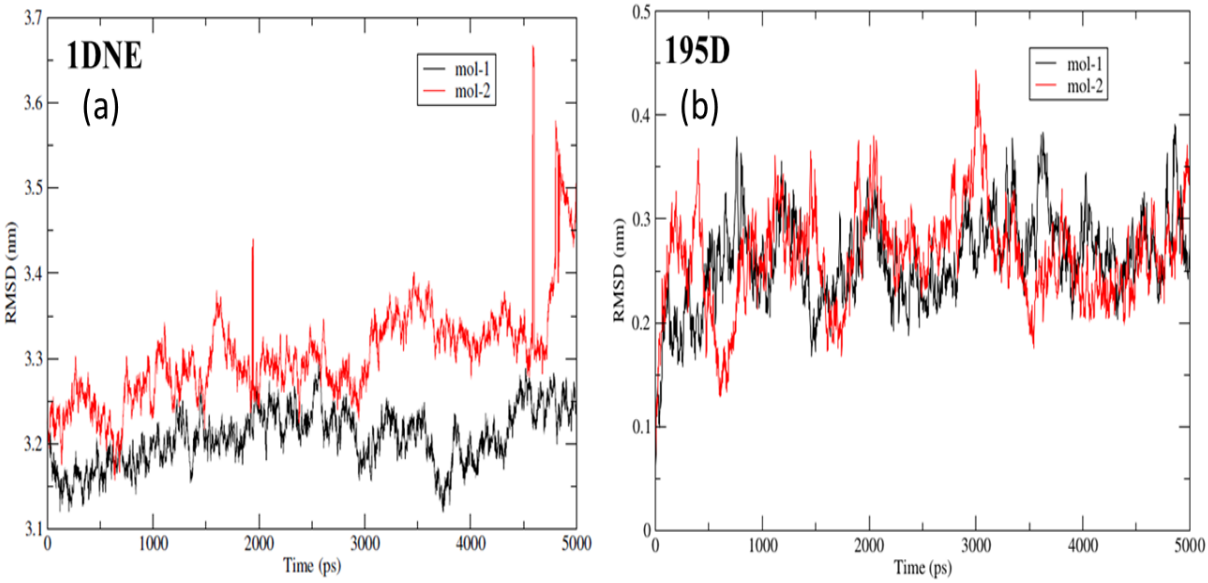


Figure 5.12: Figure representing RMSD for Drug-DNA Complexes

5.3.2.5 Root Mean Square Fluctuation

Well-structured regions and loosely bound regions in DNA strands are distinguished by low and high root mean square fluctuations values respectively [38]. RMSF also records the fluctuation of each amino acid base pair including the fluctuations in the flexible regions within the nucleic acid during the course of the MD simulation. A fluctuation between 1~3 Å is acceptable for biomolecular systems [38]. The graphs shown below in **figure 5.13** suggest that 1DNE: mol-1 had the least fluctuations among and hence forms the most stable complexes among all.

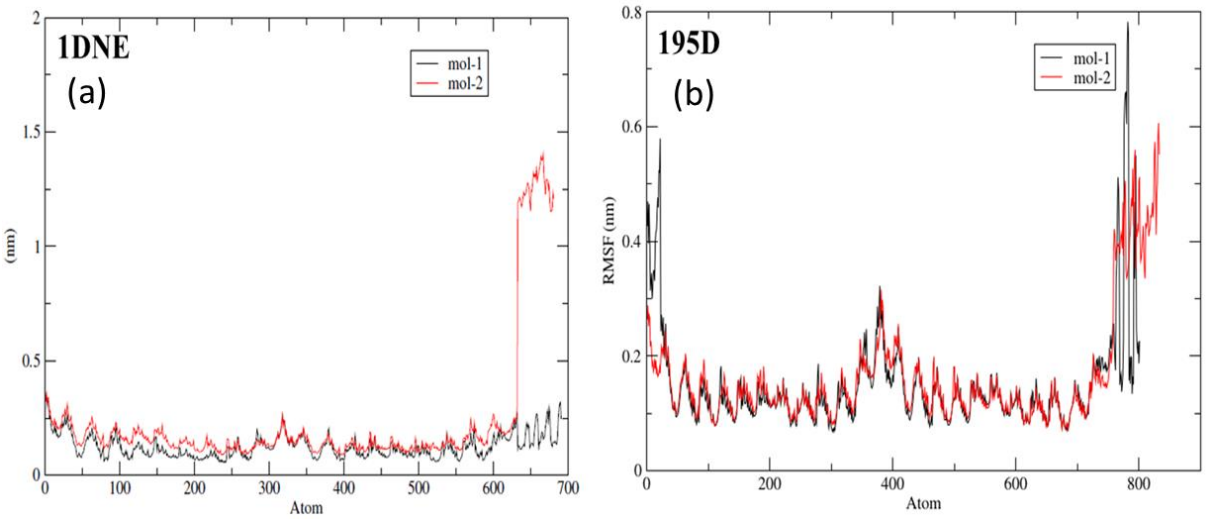


Figure 5.13: Figure representing RMSF for Drug-DNA Complexes

5.3.3 Free Energies

Variations in polar solvation free energies are shown below in **figure 5.14**. The Poisson equation for polar solvation is a fundamental equation of continuum electrostatics [39]. It is a linear, second- order, partial differential equation:

$$-\nabla \bullet (\varepsilon(\vec{r})\nabla\varphi(\vec{r})) = \rho(\vec{r}) \quad for \ \vec{r} \in \Omega \tag{5.6}$$

which expresses the electrostatic potential (φ) terms of a dielectric coefficient (ε) and a charge distribution (ρ) for all points (r) in some domain (Ω). The Poisson-Boltzmann equation for polar solvation is [40]:

$$-\nabla \bullet (\varepsilon(\vec{r})\nabla\varphi(\vec{r})) = \sum_{i}^{N} Q_i \delta(\vec{r} - \vec{r}_i) + \sum_{j}^{m} c_j^b q_j \exp\left[\frac{-q_j\varphi(\vec{r})}{k_B T} - \frac{V_j(\vec{r})}{k_B T}\right] \quad (5.7)$$

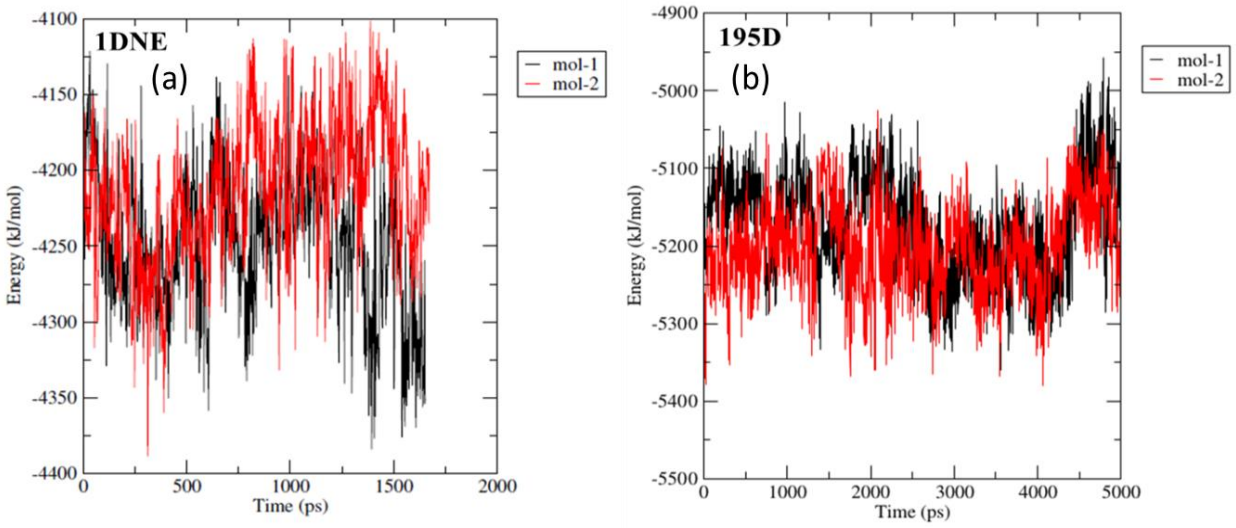


Figure 5.14: Variation in polar solvation free energy for Drug-DNA Complexes

A-polar or non-polar solvation involves the interaction of the system with uncharged solutes. This method of solvation usually includes hydrophobic interaction studies. Several biological phenomena are associated with non-polar solvation of the system, beginning from protein-protein interactions to protein folding and energetics involved [40].

Research in the field of atomistic simulations will continue to enhance our understandings of the role of solvation methods (polar or non-polar) of biomolecular structure and dynamics. Since the major contribution of solvation energies is that

which is required to create a solvation cavity. This cavity generation term explains the amount of work done to accommodate within the solvent and extremely strong solute-solvent interactions [41], etc. **Figure 5.15** shows that the a-polar solvation free energy for ligand complexes of 1DNE and 195D lies between 90~105kJ/mol and 110~130kJ/mol respectively. Therefore, we may predict that 1DNE requires lesser cavity creation energy that 195D. Also, the non-polar solvation process (strong and attractive solvent-solute interactions) cannot be easily decoupled [42-44] and therefore on this basis 1DNE complexes attain more stability over 195D. **Table 5.8** shows the component-wise free energy contribution of each drug-DNA complex. The residue wise total energy as calculated by residue decomposition analysis of the G-MMPBSA module is represented in **figure 5.16**.

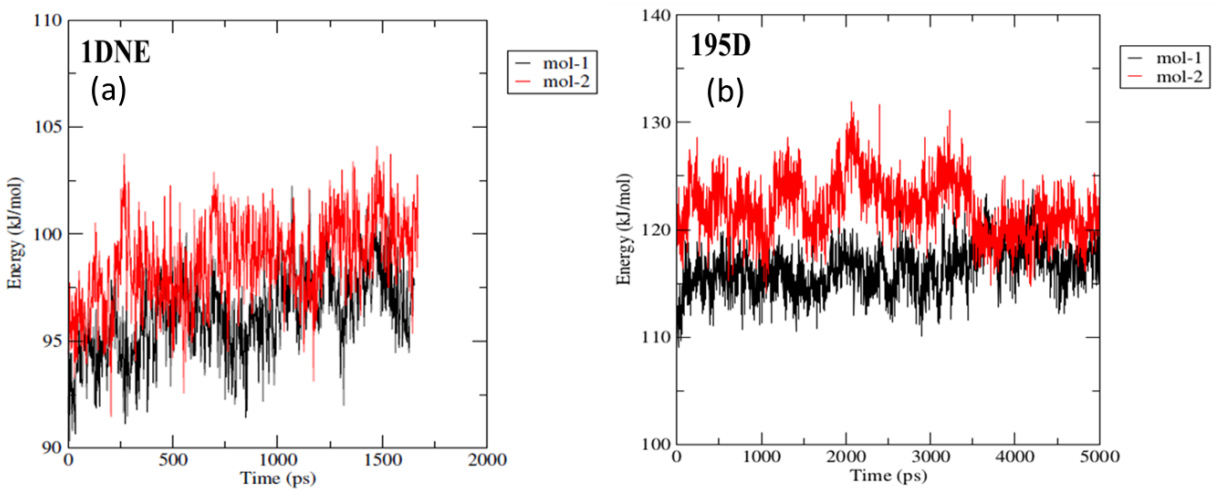


Figure 5.15: Variation in a-polar solvation free energy for Drug-DNA Complexes

Table 5.8: Table representing the component-wise free energy contributions for each drug-DNA pair

S. No.	Energy Term	1DNE		195D	
		Mol-1 (kJ/mol)	**Mol-2 (kJ/mol)**	**Mol-1 (kJ/mol)**	**Mol-2 (kJ/mol)**
1	Van der Waals energy	-97.318±9.428	-185.255±23.278	-66.352±31.113	-121.989±19.503
2	Electrostatic energy	-49.417±12.248	-44.114±17.444	-50.176±34.099	-33.822±25.373
3	Polar solvation energy	51.252±9.816	34.384±33.969	41.507±35.380	17.709±27.876
4	SASA energy	-10.254±0.829	-18.334±2.055	-7-719±2.810	-13.414±2.072
5	Binding energy	-105.736±17.243	-213.319±39.215	-82.740±37.767	-151.517±28.417

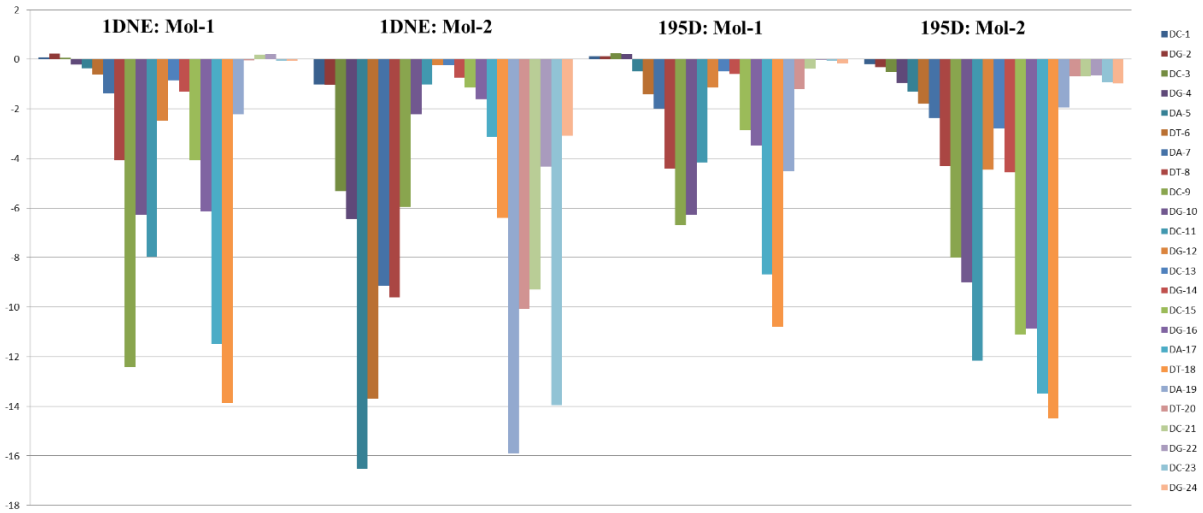


Figure 5.16: The residue interaction map calculated by residue decomposition analysis of G-MMPBSA module

5.4 Conclusions

Computational techniques have been key players in drug designing and pharmacy industries. New and potent drugs are designed and existing drugs are studied for their relative action tendencies using various computational methodologies. In this chapter, interaction, stability and dynamics of drugs in the vicinity of DNA was studied. The optimized chemical structures of ligands when docked to selected DNA generated best docked pose by finding the most stable binding site. These docked complexes were subjected to molecular dynamics for evaluation of their stability. We then performed free energy calculations in order to support the results obtained through molecular dynamics simulations.

Molecular docking results reveal that both the selected molecules corresponding to 1DNE & 195D show stabilizing interactions because of the comparable binding affinities. However, since binding energy obtained for 1DNE+mol-2 is least and also more interactions were observed in complexes of 1DNE; this supports the stability of its complexes. It was also observed from the obtained results that all the ligands were minor groove binders and were docked to AT-rich sited of the DNA, as preferred by minor groove binders. The results obtained through hydrogen bonding analysis suggested that donor region had greater extent in 1DNE and acceptor clouded regions had more extent in 195D. The results also represented the DNA bases involved in the hydrogen bond formations.

Results obtained from the molecular dynamics simulation show the time dependence and the stability of selected molecules in the vicinity of DNA. Radius of gyration, RMSD and RMSF analysis were done from the trajectories obtained via MD

studies; all these results were comparable & thus favor the stable complex formation for both. However, the radius of gyration, RMSD & RMSF analyses reveal a stable complex formation for mol-1 & mol-2 with 1DNE.

Free energy calculations were carried out for better understanding of the results obtained through MD simulations. The results obtained suggest that since binding energies for 195D is less than 1DNE, obtained through various contribution terms, so 195D has the most stable complex formation tendency than 1DNE.

Conclusively, this study describes the properties and dynamics of DNA on the interaction with 2,5-diaryl furan derivatives, taking the account of deformation upon binding which can play significant role in the discovery of new minor groove binder as a regulator of gene expression. Obtained results not only predicted the conformational stability of the DNA with the ligands but also gave a firm affirmation regarding the usage of computational techniques for studies of time dependence of the interaction and stability of drug-DNA complexes.

References:

[1] Chen, Q., Dong, L., Huang, Y., Qing, Z., *Analyst*, **124**, 901–906 (1999).

[2] Navarro, J.A.R., Juan, M.S., Romero, M.A., Vilaplana, R., Gonzalez-vı, F., Lyon, C.B. Restrictive New Cisplatin Analogue., *J. Med. Chem.*, **97**, 332–338 (1998).

[3] Blackburn, G.M., Gait, M.J., Williams, D.M. *Nucleic Acids in Chemistry and Biology*, 3rd Edition, RSC Publishing (2006).

[4] Mishra, R., Kumar, A., Chandra, R., Kumar, D., *International Journal of Science, Technology and Society*, **3**, 11-27 (2017).

[5] Pandey, A., Mishra, R., Shukla, A., Yadav, A.K., Kumar, D., *Conference Proceedings of IASFBM-2019*, 11-18 (2019).

[6] Graves, D.E., Velea, L.M., *Current Organic Chemistry*, **9**, 915–929 (2000).

[7] Kansy, M., Senner, F., Gubernator, K., **7,** 0–3 (1998).

[8] Yadav, R., Pandey, A., Awasthi, N., Shukla, A., *Advanced Science, Engineering and Medicine,* **12(1)**, 83-87 (2019).

[9] Senn, H. M., Thiel, W., *Curr. Opin. Chem. Biol.*, **11,** 182–187 (2007).

[10] Tidwell, R.R., Boykin, D.W., *Journal of Brazilian Chemical Society,* 2, (2003).

[11] Pandey, A., Yadav, R., Shukla, A., Yadav, A.K., *Advanced Science, Engineering and Medicine,* **12(1)**, 40-44 (2019).

[12] Chipot, C., *Computational Molecular Science*, **4**, 71–89 (2013).

[13] Senn, H.M., Thiel, W., *Angew. Chemie Int. Ed.*, **48,** 1198–1229 (2009).

[14] Srivastava, H.K., Chourasia, M.K., Kumar, D., Sastry, G.N., *Journal of Chemical Information and Modeling,* **51**, 558–571 (2011).

[15] Coll, M., Aymami, J., Van Der Marel, G.A., Van Boom, J.H., Rich, A., Wang, A.H.J., *Biochemistry*, **28**, 310–20 (1989).

[16] Balendiran, K., Rao, S.T., Sekharudu, C.Y., Zon, G., Sundaralingam, M., *Acta Crystallogr.,* **Sect.D 51**, 190-198 (1995).

[17] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., *Nucleic Acids Research,* **28**, 235-242 (2000).

[18] Gaussian 09, Revision E.01, 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,

Ö., Foresman, J. B., Ortiz, J. V., Cioslowski, J., Fox, D. J. Gaussian, Inc., Wallingford CT, 2009.

[19] Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.M., Ferrin, T.E., *Journal of Computational Chemistry*, **25**, 1605-1612 (2004).

[20] Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J., *Journal of Computational Chemistry*, **30**, 2785–2791 (2009).

[21] Braun, E., Gilmer, J., Mayes, H.B., Mobley, D.L., Jacob, I., Prasad, S., Zuckerman, D.M., **1**, 1–28 (2019).

[22] Zhao, H., Caflisch, A., *European Journal of Medicinal Chemistry*, **94**, 4-14 (2015).

[23] Durrant, J.D., McCammon, J.A., *BMC Biology*, **9**, 71 (2011).

[24] Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B. Lindahl, E., *SoftwareX*, **1-2**, 19-25 (2015).

[25] Sapay, N., Tieleman, D.P., *Journal of Computational Chemistry*, **32**, 1400-1410 (2011).

[26] Zoete, V., Cuendet, M.A., Grosdidier, A., Michielin, O., *Journal of Computational Chemistry*, **32**, 2359-2368 (2011).

[27] Mark, P., Nilsson, L., *Journal Physical Chemistry-A*, **105,** 9954-9960 (2001).

[28] Darden, T., York, D., Pedersen, L.,*The Journal of Chemical Physics*, **98**, 10089 (1993).

[29] Ryckaert, J.P., Ciccotti, G., Berendsen, H.J.C., *Journal of Computational Physics*, **23**, 327-341 (1977).

[30] Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M., *Journal of Computational Chemistry,* **18**, 1463-1472 (1997).

[31] Turner, P.J., XMGRACE, Version 5.1. 19. Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR (2005).

[32] Kollman, P.A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D.A., Cheatham, T.E., *Accounts of Chemical Research*, **33,** 889-897 (2000).

[33] Srinivasan, J., Cheatham, T.E., Cieplak, P., Kollman, P.A., Case, D.A., *Journal of American Chemical Society,* **120**, 9401-9409 (1998).

[34] Changhao, W., D'Artagnan, G., Li, X., Ruxi, Q., Ray, L., *Frontiers in Molecular Biosciences,* **4**, 87 (2018).

[35] Genheden, S., Ryde, U., *Expert Opinion on Drug Discovery*, **10**, 449-46 (2015).

[36] Kumari, R., Kumar, R., Lynn, A., *Journal of Chemical Information and Modeling,* **54**, 1951-1962 (2014).

[37] Shukla, R., Shukla, H., Kalita, P., Sonkar, A., Pandey, T., Singh, D.B., Kumar, A., Tripathi, T., *Journal of Biomolecular Structure and Dynamics*, **36**, 2147-2162 (2018).

[38] Chaubey, A.K., Dubey, K.D., Ojha, R.P., *Medicinal Chemistry Research*, **24**, 753–763 (2015).

[39] Jackson, J.D. *Classical Electrodynamics*, 3rd, Edition, John Wiley & Sons, Inc. (1975).

[40] Ren, P., Chun, J., Thomas, D., Schnieders, M., Marucho, M., Zhang, J., Baker, N., *Quarterly Reviews of Biophysics*, **45**, 427-491 (2012).

[41] Hummer, G., Garde, S., Garcia, A.E., Pohorillet, A., Pratr, L.R., *Proceedings of National Academy of Science USA*, **93**, 8951-8955 (1996).

[42] Dzubiella, J., Hansen, J., *Journal of Chemical Physics*, **124**, 5514-5530 (2004).

[43] Dzubiella, J., Swanson, J.M.J., McCammon, J.A., *Phys. Rev. Lett.*, **96**, 4 (2006).

[44] Dzubiella, J., Swanson, J.M.J., McCammon, J.A., *Journal of Chemical Physics*, 124 (2006).

[45] Aamir, M., Singh, V.K., Dubey, M.K., Meena, M., Kashyap, S.P., Katari, S.K., Upadhyay, R.S., Umamaheswari, A., Singh, S., *Front. Pharmacol.*, **9**,1038 (2018).

Chapter-6

Exploring Interactions of Some DNA Binding Ligands through Molecular Docking, Molecular Dynamics Simulations and Quantum Mechanical/Molecular Mechanical (QM/MM) Calculations

Chapter-6

Exploring Interactions of Some DNA Binding Ligands through Molecular Docking, Molecular Dynamics Simulations and Quantum Mechanical/Molecular Mechanical (QM/MM) Calculations

6.1 Introduction

Nucleic acids are the most prominent targets for a number of anticancer, antiviral, antitumor, antiprotozoal and antiparasitic drugs. Some of these drugs are either currently in medicinal usages or under clinical trials. Most of these drugs hold aromatic rings within them i.e., they are the derivatives of aromatic class of compounds. Groove binding and intercalation are the two major drug binding modes with which drugs bind themselves with DNA. Intercalation needs conformational changes in DNA due to the formation of intercalation gap between the DNA base pairs whereas, in contrast, minor groove binders which do not require any conformational changes to the DNA [1-4]. Major groove of the DNA has a width of 11.7Å and depth of 8.8Å, and therefore it holds multiple sites for intercalation which result in the formation of comparatively stronger bonds. Whereas, the width and depth for minor groove of the DNA are 5.7Å and 7.5Å respectively and hence there is no availability of binding pockets in the vicinity of DNA. Geometrical flexibility is an important parameter of DNA minor groove binders. They tend to acquire crescent shape while binding them to DNA and hence stabilize the complex [5]. 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. On a statistical note, typically, minor groove binders bind to B-DNA with the higher binding affinity towards AT- rich centers [6,7]. However, the intercalators are usually geometrically planar molecules with several fused rings, and are usually 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. Hydrophobic interactions between ligand and DNA bases contribute towards the stability of intercalated complexes [8-10]. DNA minor groove is the target for many anticancer and antitumor drugs. The forces that contribute in DNA-minor groove binding are electrostatic forces, van der Waal's interactions, hydrophobic interactions, and hydrogen bonding. Sequence specificity is a major clause in drug-DNA interactions. It has been observed that, minor groove binding drugs show selectivity towards AT rich region, several factors are also responsible for this first is electrostatic potential of AT-rich region is higher than that of GC-filled ones and the second one is that the AT-rich grooves are narrower and deeper than GC ones and thus allowing better binding and interactions. Minor groove of the DNA has alternating A and T bases which allows van der Waals contacts between the drug and DNA however no such contact is observed in GC-rich regions because the geometry of groove is altered by bulky amino groups of guanine bases.

In this work, some ligands were taken that were claimed to have antimicrobial activities and were subjected to exhaustive computational studies. Molecular modeling techniques such as molecular docking, molecular dynamics simulation, and Quantum

Mechanical/Molecular Mechanical (QM/MM) calculations have been used for the estimation of binding energies, which was then comparatively studies and evaluations were made to reach a conclusive remark.

Molecular docking was first done using AutoDock4 software for identifying the drug binding site followed by molecular dynamics simulation for both the drug-DNA complexes using GROMACS software. The data obtained from molecular dynamics simulation was analyzed for energy variations, variations in radii of gyrations, variations in number of hydrogen bonds, root mean square deviations and root mean square fluctuations. Further, QM/MM calculations were carried out using the ONIOM scheme of Gaussian09 software by taking several snapshots at time interval of 1ns for the entire molecular dynamics simulations.

The results obtained from these calculations will create opportunities for further *In-vitro* & *In-vivo* studies for the correct determination and applicability of better antimicrobial agent amongst the selected molecules.

6.2 Methodology

6.2.1 System Selection and Preparation

The lead molecules (ligand-1 & ligand-2), for the current study were taken from the literature [11]. Their chemical structures are shown below in **figure 6.1**. The DNA sequence 1BNA [12] was obtained from Protein Data Bank [13]. Its structural information including its specific nucleic acid bases sequences and PDB ID is mentioned in **table 6.1**, as obtained from Discovery studio visualizer [14]. Water molecules, from downloaded DNA sequences were removed using UCSF Chimera software [15] to prepare them for docking and molecular dynamics simulations.

Table 6.1: Nucleic Acid Report for 1BNA

S. No.	PDB Id.	Nucleic Acid Sequence from Available Structure	Basic Information
1.	1BNA	B-DNA DODECAMER (5'-D(CGCGAATTCGCG)-3') Chain A: CGCAAATTTGCG Chain B: CGCAAATTTGCG	• Cell Space Group: 19 (P212121 Origin-1 Choice: 1) • Crystallographic Resolution: 1.20 Å • Molecular Weight of Nucleic Acid Chains: 7268.84 • Number of Nucleic Acids: 24 • Experimental pH: 6.5

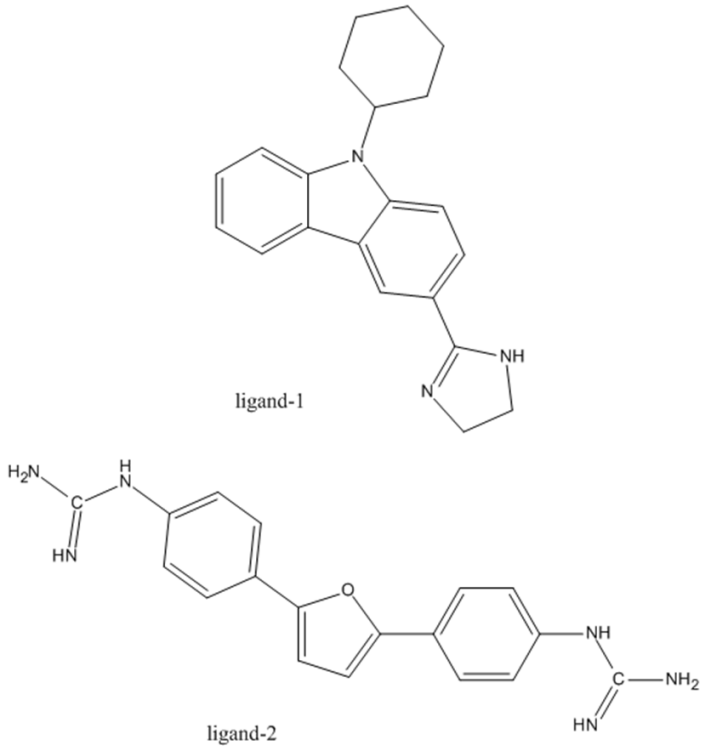


Figure 6.1: Figure representing the selected antimicrobial ligands

6.2.2 Molecular Geometry Optimization

The geometry optimization is a crucial step in molecular modeling studies. The reason that leads to it is that, geometry optimization would put the entire molecular

system to its lowest possible energy state; in which it experiences least steric hindrances and least charge repulsions; as the system would then have attained a maximized distance between the bonds, the angles and corresponding dihedrals [16]. Also, the electrostatic charges tend to get equally distributed throughout the system leading to the electron transfer between definite pair of atoms; and eventually no distortions in the chemical structures are offered [17]. Here, the geometry optimization of selected ligands was carried out using Gaussian 09 software [18] by applying B3LYP hybrid density functional and 6-31G** basis set for their potentials to attain a local minimum. These optimized ligands were then led to docking calculations.

6.2.3 Molecular Docking Studies

Molecular Docking is a widely used computational technique to predict the binding site and a few underlying interactions of potent inhibiters in the vicinity of their corresponding target molecules. Molecular docking plays an important role in the determination of ligand binding site and associated binding affinity [19].

In the current study, docking was carried out using Autodock4 software [20]. Molecular docking calculations were set up using classical Lamarckian Genetic Algorithm (LGA). Then, missing Gasteiger charges were added to each drug-DNA complex using Autodock Tools (ADT) before starting the docking simulations. A grid box that enclosed the macromolecule having different dimensions along all the three coordinate axes was prepared for each of the two drug-DNA complexes. A simulation was carried out with 20 LGA runs having a maximum cycle of 2500000 energy evaluations for each of the drug-DNA complex. The docked pose corresponding to the

least binding affinity was extracted and aligned with the receptor (DNA) for further analysis.

6.2.4 Molecular Dynamics Simulation

Over several years, molecular dynamics has evolved itself to be a very reliable computational technique in the studies of the time evolution of systems. The study of dynamic characteristics of biological systems viz., protein folding, winding & unwinding and other conformational changes including complex stability, over a specified period of time gets easier to depict by studying it through molecular dynamics. Its applications have gained significant importance due to lack of experimental resources [21].

In the current study, GROMACS 5.0.4 software package [22] was used to carry out all the molecular dynamics simulations. A total of two drug-DNA complexes were created keeping using the best docked poses obtained after docking for molecular dynamics simulations viz., (1BNA+ lig-1 and 1BNA+lig-2). A simulation for 5000ps was carried out for the studies of their dynamical behavior. There are many force fields proposed for the MD simulation studies of nucleic acids such as, CHARMM, AMBER, GROMOS, OPLS, ENCAD and BMS. Moreover, many studies have been performed regarding the comparison of force fields for the nucleic acids but AMBER force fields seem to be good for nucleic acid simulation due to the presence of specific topologies for the terminal nucleotides [23]. During the current simulation, Amber03 force field, which was already embedded in GROMACS software suite, was used to generate the topology for the DNA sequence and antechamber module of AMBER program was

used to generate topology of selected ligands through a python script: ‘acpype.py’ [24].

The Ligand-DNA complexes were solvated in a box of different dimensions along the three axes using TIP3P water model at 298K [25]. Sodium ions were then added to the solvated boxes containing the DNA-ligand complexes by randomly replacing the water molecules in order to neutralize the system, as the biological systems needs to be neutral overall before the production run begun. Particle Mesh Ewald (PME) was used to handle long-range electrostatic interactions under periodic boundary conditions [26]. Energy minimization of the whole system was carried out in 25000 steps using Steepest Descent leap- Frog Integration Method followed by NVT ensemble equilibration at a constant temperature of 300K for 50s using Berendsen thermostat [27]. The system was then equilibrated with NPT ensemble at a constant pressure of 1atm in 25000 steps using steepest descent leap-frog integrator [27]. All the bonds involving hydrogen atoms were constrained using the LINCS algorithm [28]. Graphs were plotted using XM-grace software [29]. Variations in energies, 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. The MD trajectories were visualized by means of VMD [30].

6.2.5 QM/MM Calculations

A total of 5 snapshots at 1ns each from the molecular dynamics results as obtained from GROMACS, for each of the two drug-DNA complexes was extracted using VMD and that formed the basis of the starting point of the QM/MM calculations.

These structures contained the bimolecular system surrounded by water droplets of around 20000-30000 atoms in number and this setup requires a lot of prior work to avoid errors and wrong choices regarding the actual QM/MM calculations.

In this chapter, the quantum mechanical/molecular mechanical (QM/MM) calculations were performed using the two-level ONIOM method embedded within the Gaussian 09 program. B3LYP hybrid functional along with 6-31++G basis set was used for the higher layer (QM region of the system) whereas AMBER force field was incorporated for the lower layer (MM region of the system) and all the surrounding water molecules and Na^+ ions were kept frozen. **Figure 6.2**, shown below, depicts the two layered ONIOM scheme of Gaussian09 used in the current study. 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 equation:

$$\Delta E_{IE} = E_{QM\,(gp)} - E_{QM\,(pp)} \qquad (6.1)$$

Where,

ΔE_{IE} is the interaction energy between drug and the DNA,

$E_{QM\,(pp)}$ is the energy of QM region in nucleic acid phase,

$E_{QM\,(gp)}$ is the energy of QM region in the gas phase.

The electronic-embedding model of the two layered ONIOM (ONIOM-EE) method was used for single point energy calculation using the B3LYP hybrid DFT functional along with 6-31++G basis set for the QM region and AMBER force field for the MM region. The single point energy of QM region in the gas phase using the same method was also calculated. The interaction energy between the DNA and drug was calculated using above mentioned formula.

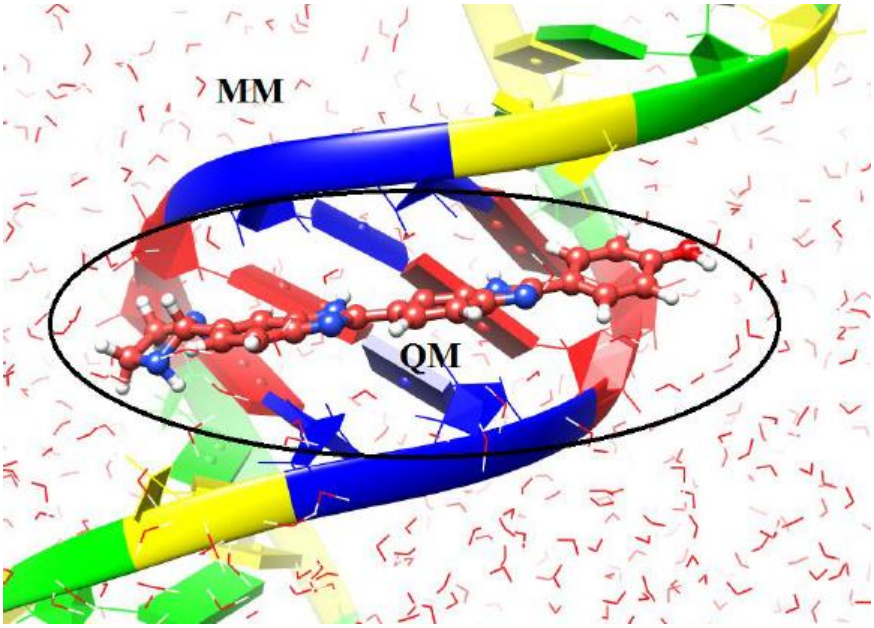


Figure 6.2: Figure depicting the two layered ONIOM Scheme for QM/MM

6.3 Results and Discussion

This study was aimed to identify new leads, targeting binding affinity, structural stability, underlying interactions at atomic levels and applicability for antimicrobial ligands as potent drugs which target DNA. The results obtained through various computational calculations are discussed and summarized as follows:

6.3.1 Molecular Geometry Optimization

It is often suggested to optimize the geometry of the lead molecules before performing the docking simulations because then the repulsions between their nearest bonded atoms, between their angles and between corresponding dihedrals would get minimized and the lead molecules would attain such an energy state having least repulsion and therefore owing to maximum stability and the geometrical orientation of that state would correspond to minimum energy [31]. The optimized geometries of the selected ligands are shown below in **figure 6.3**.

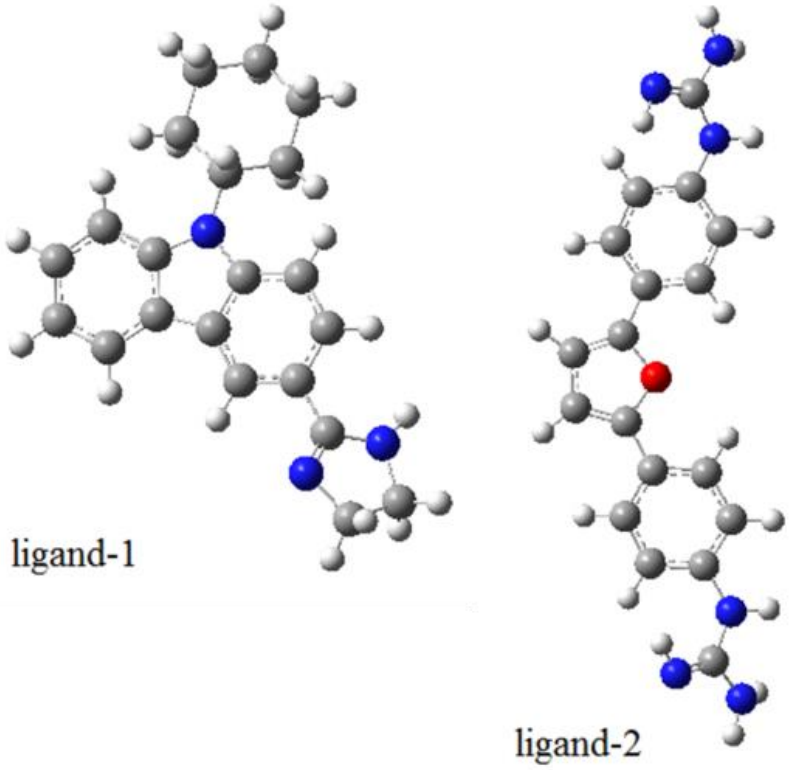


Figure 6.3: Figure showing optimized structures of the selected ligands

6.3.2 Molecular Docking

The lead molecules (lig-1 & lig-2) were docked to the selected sequence calf

thymus DNA sequence (PDB Id: 1BNA) for the evaluation of its interaction tendencies, complex formation tendencies, and stability analysis of the best docked posed complex. The docking results, corresponding to the selected DNA sequences are summarized below in **table 6.2**.

A perusal of **table 6.2** revealed that 1BNA with lig-1had the least binding affinity and therefore has formed more stable complex with the DNA than lig-2. Various other parameters that contributed towards docking binding affinity are mentioned in **table 6.3**. This indicates that since lig-1 had lesser binding energy, it owes to maximum stability of the formed complex. The docked pose corresponding to each drug-DNA complex is shown in **figure 6.4**. These figures reveal that lig-1 bonded itself to the major groove of the DNA whereas lig-2 bonded itself to the minor groove of the DNA.

Further **figure 6.5** represents the interaction profile in 2D which reveals various types of interactions taking place between the DNA bases and the selected ligands, viz., conventional hydrogen bonds, carbon hydrogen bonds, π-σ bonds, π-alkyl bonds and presence of some attractive charges, etc. These interactions eventually lead to the stabilization of the drug-DNA complexes formed.

Molecular docking results also revealed that as bulky groups were attached to base structure of lig-2 as compared to lig-1 and when docked to DNA, there were less number of hydrogen bonding interactions with lig-1 and DNA than that with lig-2; owing to more steric hindrances and consequently the fact that number of hydrogen bonding interactions decrease on attaching bulky groups complement the obtained results. These hydrogen bonding interactions were obtained from Discovery Studio

Visualizer and are discussed below in detail.

Table 6.2: Table representing various docking results for 1BNA DNA sequence

S. No.	Ligand	Binding Energy (kcal/mol)	Docking RMSD (Å)
1.	lig-1	-5.37	12.794
2.	lig-2	-10.69	27.648

Table 6.3: Table representing various contributing terms to binding energy

S. No.	Contributing Factors	Corresponding Energy	
		Lig-1 (kcal/mol)	**Lig-2 (kcal/mol)**
1	vdW + H-bond + desolv. Energy	-5.87	-9.26
2	Electrostatic Energy	-0.09	-2.62
3	Final Total Internal Energy	-0.67	-0.23
4	Torsional Free Energy	+0.60	+1.19
5	Unbound System's Energy	-0.67	-0.23

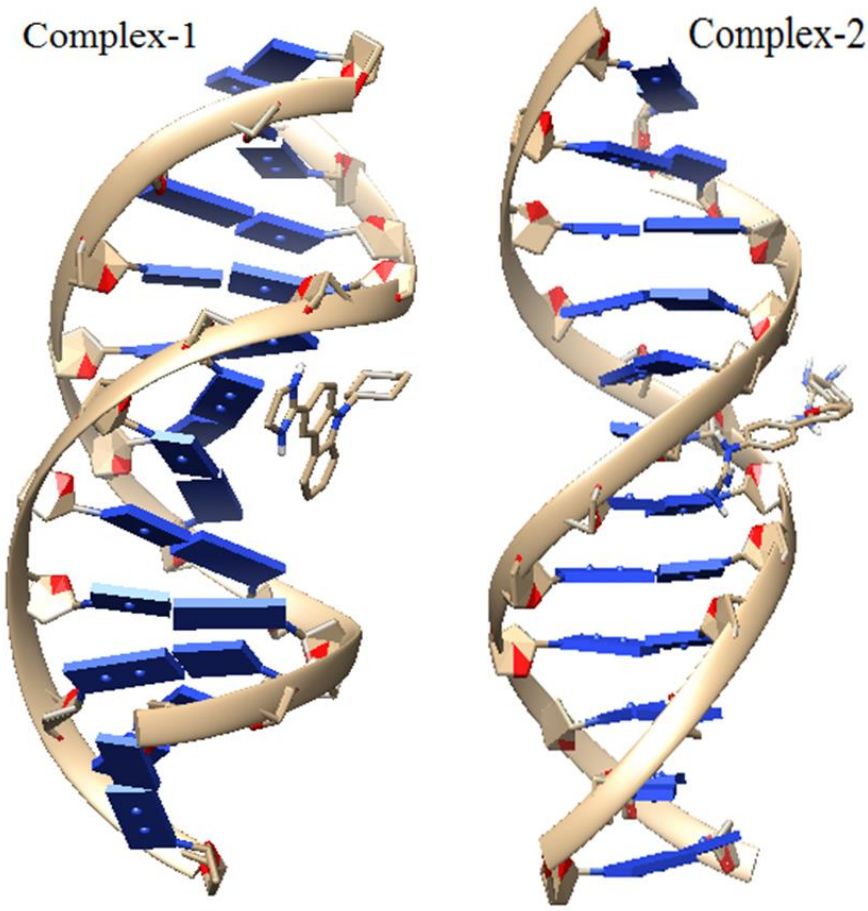


Figure 6.4: Figure showing best docked posed complexes for 1BNA

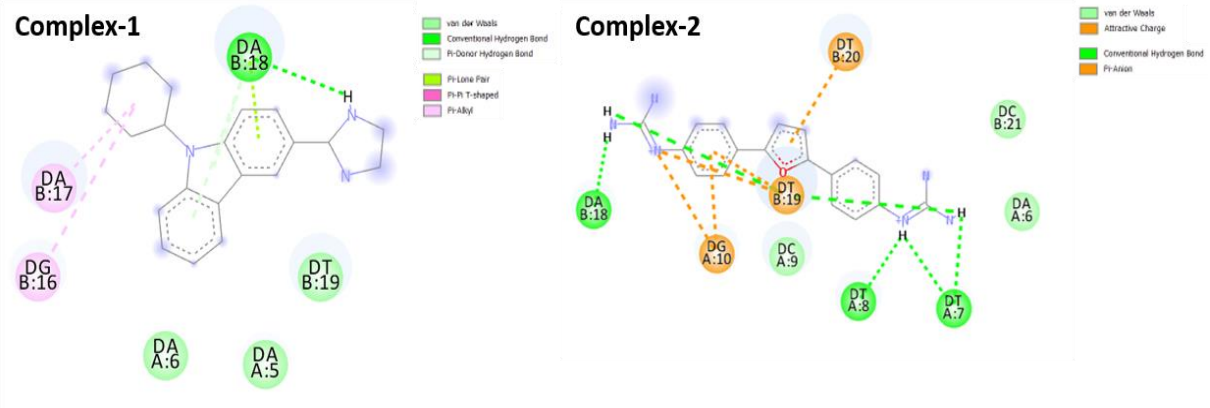


Figure 6.5: Figure representing the interaction profile for the best docked posed complexes in 2D

6.3.2.1 Hydrogen Bonding Analysis

In biomolecular systems there are some electron rich and electron deficient sites and thus helps in the prediction of possible hydrogen bond formation within the system and the strength of the bond formed. **Figures 6.6** & **6.7** represent the atomistic binding site for each of the drug-DNA complex along with corresponding H-donor/acceptor clouded regions at those binding sites. These clouded regions represent the atoms or group of atoms which have the tendency to donate/accept electrons for hydrogen bonding so as to achieve stability [35] during the docking calculations. Whereas **table 6.4** summarizes the residues involved in the formation of hydrogen bond between the DNA and ligands along with the length of the hydrogen bond.

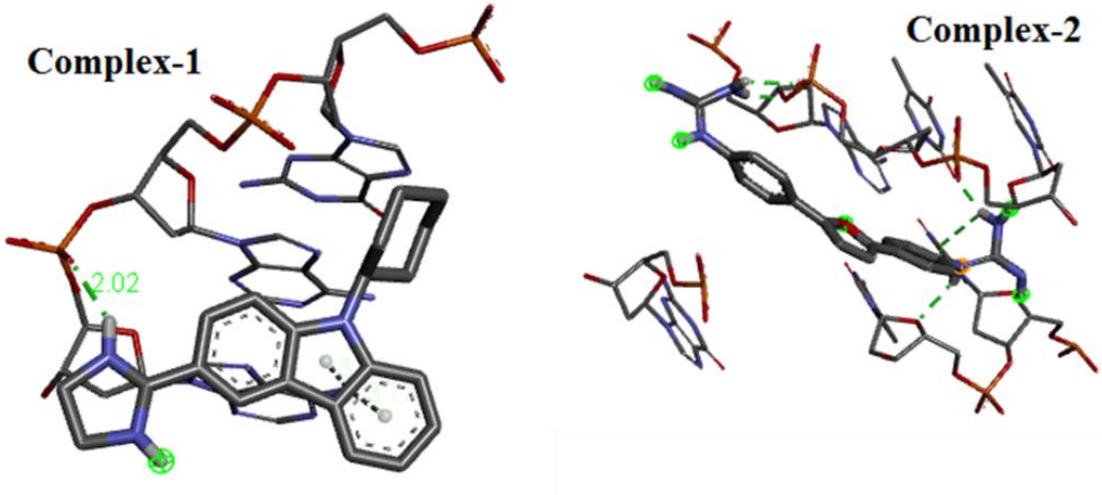


Figure 6.6: Figure showing H-bonds in best docked posed complexes

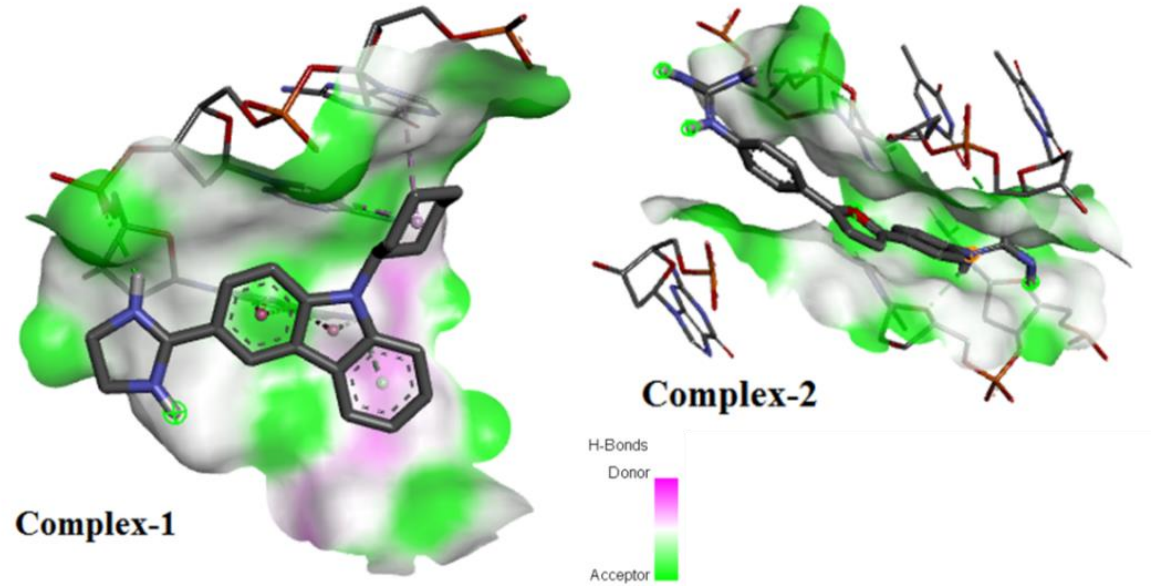


Figure 6.7: Figure representing the H-bond donor and acceptor regions in best docked posed complexes

Table 6.4: Following table represents the donor and the acceptor species and H-bond length formed between the DNA and drug atoms

S. No.	Complex	No. of H-Bonds	Interacting Species	H-Bond Length (Å)
1.	1BNA+lig-1	3	LIG1:H23 - DA18:OP2	2.016971
			DA18:N6 - LIG1	3.767351
			DA18:N6 - LIG1	3.702694
2.	1BNA+lig-2	6	LIG0:H - DT19:OP1	2.044760
			LIG0:H - DA18:O3'	1.938946
			LIG0:H - DT7:O2	2.743386
			LIG0:H - DT19:O2	1.870977
			LIG0:H - DT7:O2	2.491069
			LIG0:H - DT8:O4'	2.166167

6.3.3 Molecular Dynamics Simulation

Structural stability of the drug-DNA complexes can be elucidated, over a pre-defined period of time, via molecular dynamics simulations. The MD simulations were accompanied by certain analysis viz., variations in energy, variations in radius of gyration, variations in number of hydrogen bonds, RMSD and RMSF for better and atomistic insights into the drug-DNA complexes during the course of the simulation. Each of the above-mentioned variation are discussed individually as follows:

6.3.3.1 Energy Variations

The variations in the contributing terms in the energy of the complexes are shown below in **figure 6.8**. These variations were obtained through gmx_energy programme inbuilt in GROMACS software suite. The variations in the energy values for both the complexes are comparable. This indicates the formation of stable complexes for both lig-1 & lig-2 with same DNA sequence. However, confirmation of this fact requires analysis from various other perspectives and this lets the door opened for further research.

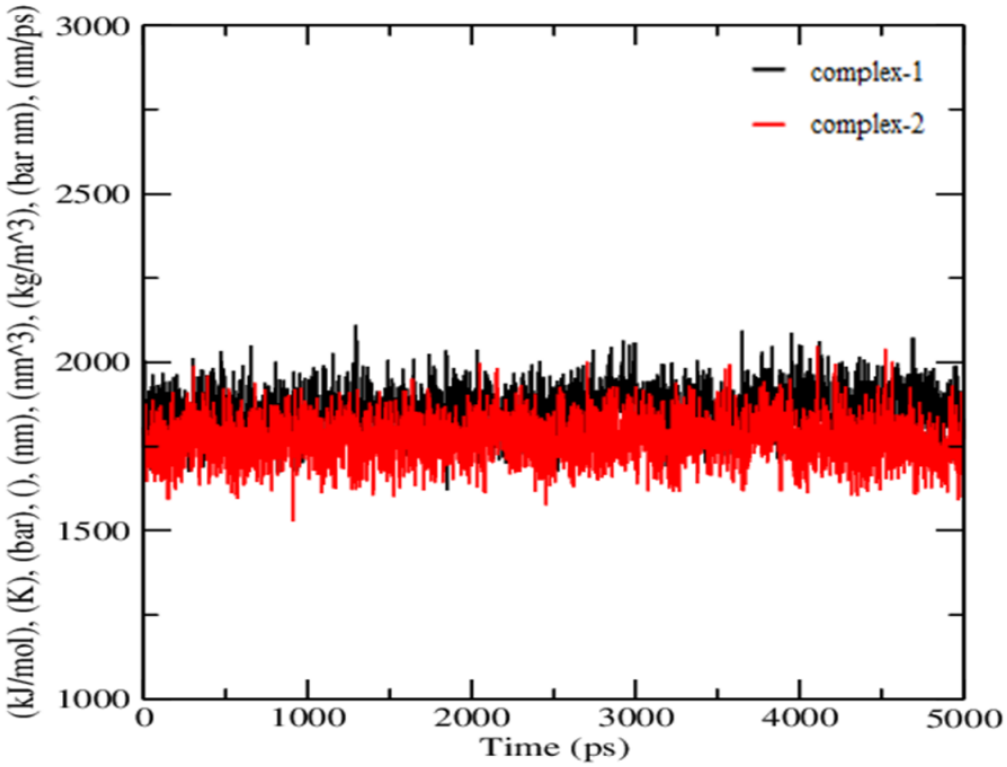


Figure 6.8: Figure representing variations in energy of the drug-DNA Complexes

6.3.3.2 Variation in Radius of Gyration

Radii of gyration values are determined in order to understand the compactness and dynamic stability of drug-DNA complexes [32, 33]. Radius of gyration was calculated through gmx_gyrate programme inbuilt in GROMACS software. The average radius of gyration for both the complexes falls between 1.275nm and 1.475nm, respectively. The collective radius of gyration variations is shown in **figure 6.9**. These variation results reveal that DNA in both the complexes remains compact for the whole 5000ps molecular dynamics simulation and hence claims for the stability of the complex [34].

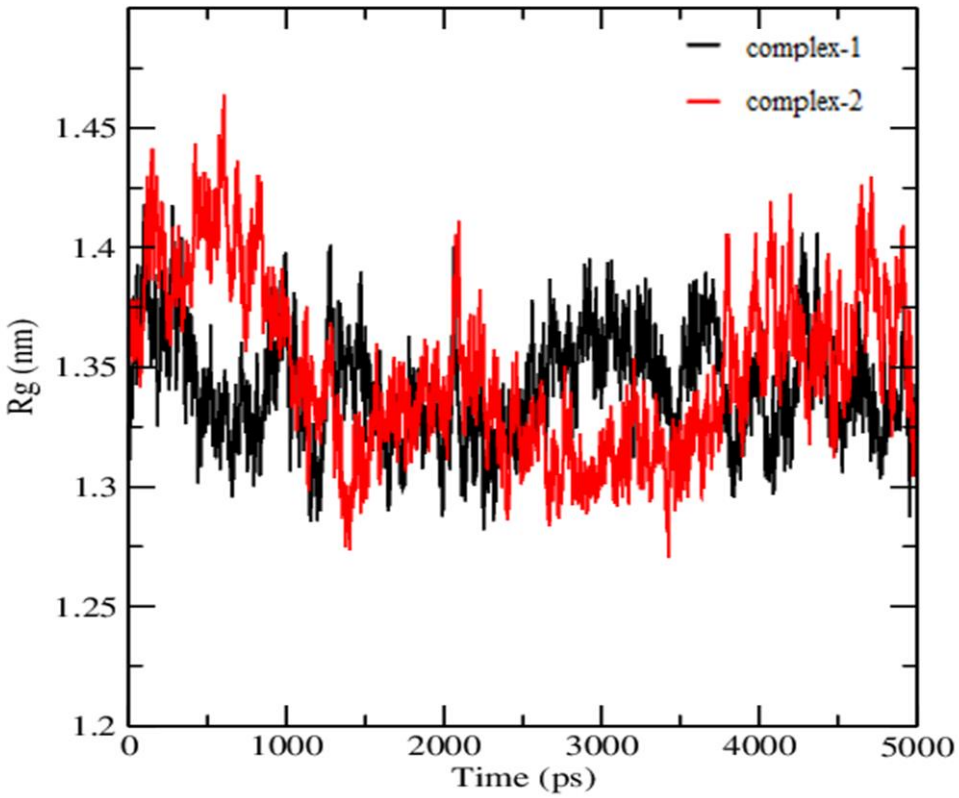


Figure 6.9: Figure representing variations in radius of gyration of the drug-DNA Complexes

6.3.3.3 Variation in Number of Hydrogen Bonds

Hydrogen bonds are important parameters in the determination of specific binding sites in DNA-ligand interactions. **Figure 6.10** depicts the variations in the number of hydrogen bonds being formed and broken during the molecular dynamics trajectory for 5000ps. The same figure also shows that complex-1 forms a maximum of 4 H-bonds whereas complex-2 forms a maximum of 8 H-bonds. This suggests the stronger interaction between DNA and ligand in complex-2. Thus, the overall results of hydrogen bonds suggest that complex-2 could act as better antimicrobial agents [32] on account of previous results.

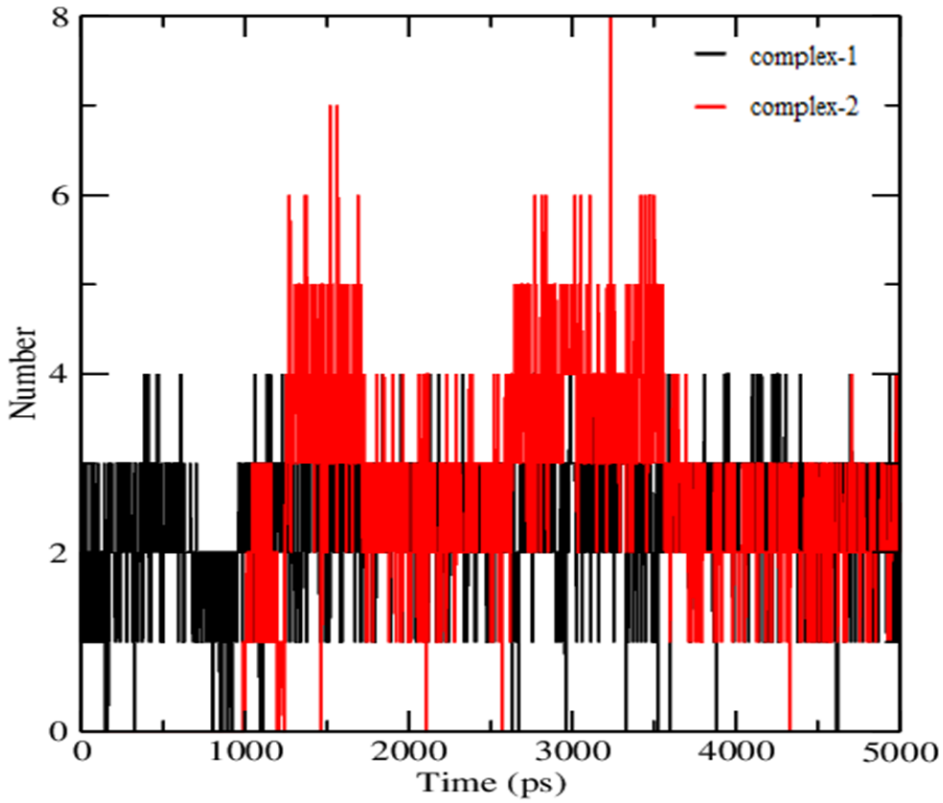


Figure 6.10: Figure representing variations in number of hydrogen bonds for drug-DNA Complexes

6.3.3.4 Root Mean Square Deviation

RMSD is considered as the measure of the conformational stability of the complex during the course of molecular dynamics simulation [32]. Here, our interest lies in examining the conformational stability and dynamical effect of ligand-DNA complexes and hence to determine the most stable complex amongst the two. The plots for RMSD of both drug-DNA complexes are represented in **figure 6.11**. In the variations shown below, the RMSD for both complexes are similar and thus both of them claim towards the stabilization of the complexes. However, complex-2 shows consistency in the variations which is missing in that ofcomplex-1 and this adds to the benefit of the results obtained in favor for complex-2.

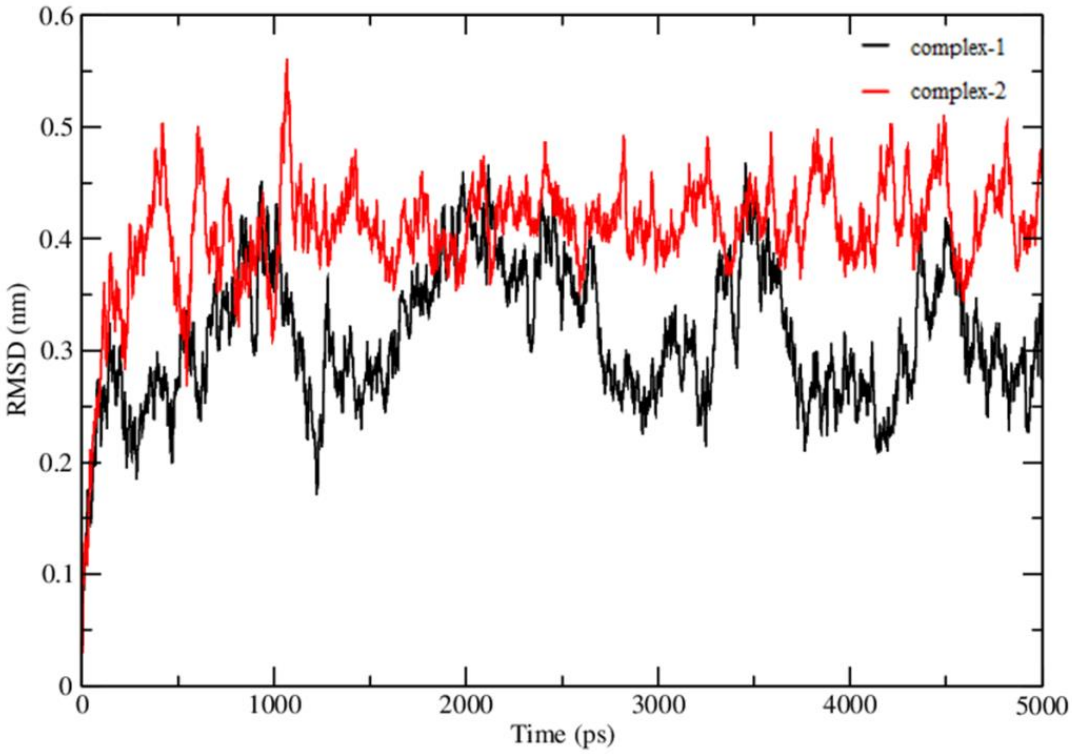


Figure 6.11: Figure representing RMSD for Drug-DNA Complexes

6.3.3.5 Root Mean Square Fluctuation

Well-structured regions and loosely bound regions in proteins and nucleic acid strands are distinguished by low and high root mean square fluctuations values respectively [33]. RMSF also records the fluctuation of each amino acid base pair including the fluctuations in the flexible regions within the nucleic acid during the course of the MD simulation. For small proteins, a fluctuation between 1~3 Å is acceptable [33]. The graph shown below in **figure 6.12** suggests that complex-2 had the least fluctuations in its residues from beginning till the end of the simulations, and hence claims to be the most stable complex.

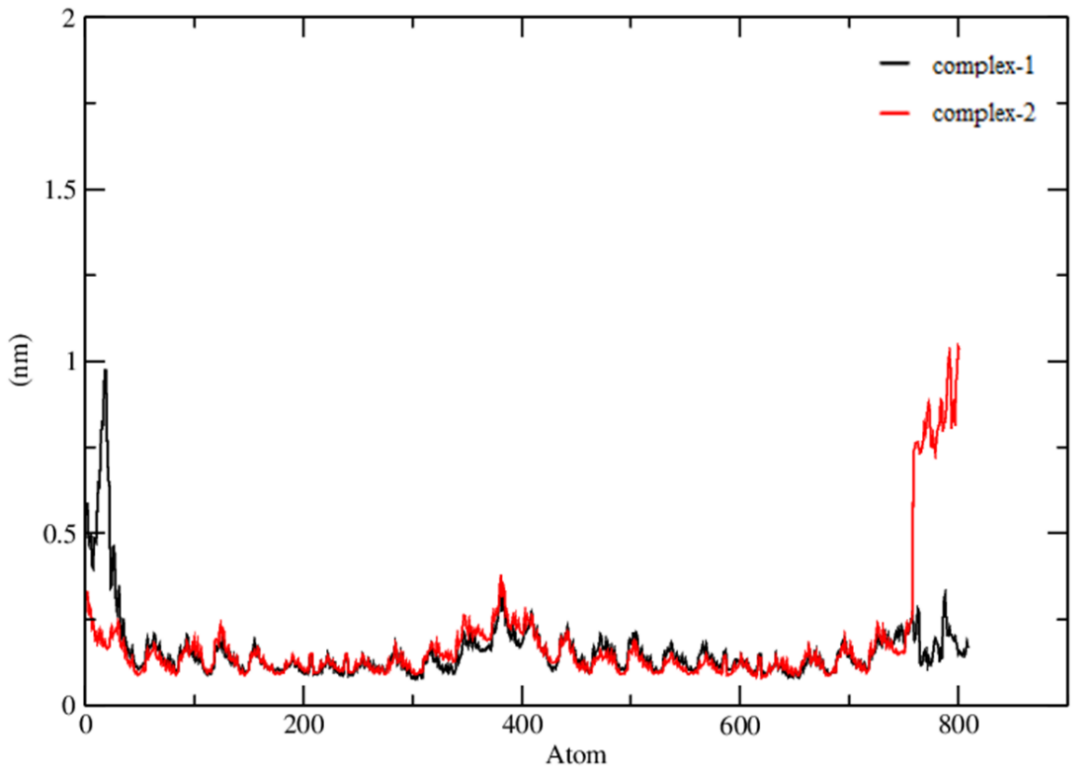


Figure 6.12: Figure representing RMSF for Drug-DNA Complexes

6.3.4 Quantum Mechanical/Molecular Mechanical (QM/MM) Calculations

The interactions energies between drug & DNA were calculated for both the complexes using previously discussed ONIOM method by employing B3LYP hybrid functional. The geometry optimization was performed using the basis set 6-31G** for the quantum mechanical (QM) region (high layer) and amber force field for molecular mechanical (MM) region (low layer). The optimized geometry of complex having the least interaction energy is shown in **figure 6.13**. All the interaction energies were calculated using equation 1**.** The calculated interaction energies from QM/MM calculation using ONIOM scheme are mentioned below in **table 6.5.** 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.

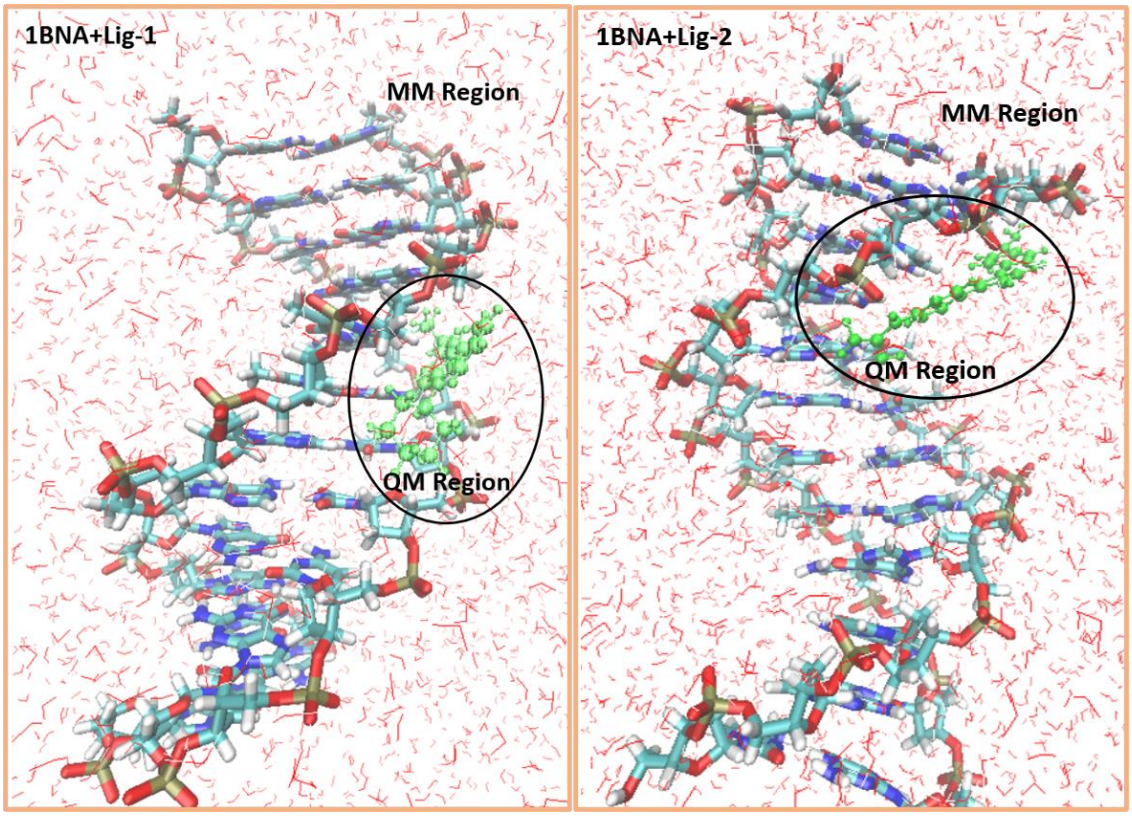


Figure 6.13: Figure representing one of the snapshots of drug-DNA complexes

Table 6.5: Table showing the obtained energies (ΔE_{IE}) through QM/MM calculations

S. No.	Time Scale	Complex-1	Complex-2
		ΔE_{IE} (kcal/mol)	ΔE_{IE} (kcal/mol)
1	**1ns**	-1.733	-4.367
2	**2ns**	+2.902	-7.282
3	**3ns**	-1.714	-3.334
4	**4ns**	-1.300	-4.578
5	**5ns**	-2.125	-1.776

6.4 Conclusions

DNA minor groove binding drugs are of high pharmaceutical potential because of their mode of action towards anticancer, antiviral, antibacterial activities are directly related to their binding with nucleic acids. A computational approach that combines with quantum mechanics, molecular docking, and molecular dynamics simulation predict the energetic pattern of the DNA-ligand binding was presented.

Docking results revealed that lig-1 was bound to major groove of the DNA whereas lig-2 was bound to minor groove of DNA. They were strictly bound to AT-rich regions of the DNA. Docking calculations revealed that with mol-2 had least binding energy of -10.69 kcal/mol. This claims to maximum stability of the formed complex between mol-2 and 1BNA.

The results obtained through hydrogen bonding analysis suggested that complex-1 had both donor and acceptor clouded regions essential for the formation of hydrogen bonds but due to presence of bulky groups greater interactions could not be achieved. However, complex-2 had greater extent of donor clouded regions but due to presence of non-bulky substituents, hydrogen bonding was preferred. This analysis also represented the DNA bases involved in the hydrogen bond formations.

Molecular dynamics simulation shows the time evolution of the stability of ligands in the vicinity of DNA. Radius of gyration, hydrogen bonding, RMSD and RMSF analysis were done from the trajectories obtained via MD studies; all these results favor the stable complex formation for lig-2 with 1BNA. Radius of gyration & RMSD values reveal that ligands remain bound to the preferred binding positions of the DNA without any considerable deviations; whereas the hydrogen bonding data

reveals that complex-2 has greater number of hydrogen interactions, as predicted by docking calculations. RMSF analysis revealed that the topological structure of DNA remains intact during the entire course of the simulation, inferring the stability of drug-DNA complexes.

QM/MM calculations were then carried out for better understanding of the results obtained through MD simulations. The results obtained suggest that since binding energies for lig-2 is less than lig-1, so complex-2 has the most stable complex. Because lig-2 did not offer any steric hindrances during the DFT studies and hence the repulsions between their nearest bonded atoms, between the angles and between corresponding dihedrals was minimized and the complex-2 attained an energy state having least bonded repulsions and therefore maximum stability was achieved and that state would correspond to minimum energy.

Molecular docking predicted the binding pockets and MD simulation results highly support the predicted binding site. RMSD and RMSF analysis also indicate that none of the compounds detached from the system throughout the simulation. Advanced computational techniques such as, QM/MM calculations also predicted the similar results as obtained via molecular docking and molecular dynamics simulation, owing to incorporation of DFT. The results obtained through computational calculations not only predicted the conformational stability of the DNA with the ligands but also gave a firm affirmation regarding the time dependence of the interaction and stability of drug-DNA complexes. The extension of such studies involving computational techniques is not only limited to drug design and molecular modelling but also to drug metabolism.

References:

[1] Neidle, S., Nunn, C.M., *Natural Product Reports*, **2**, 1-15 (1998).

[2] Neidle, S., *Natural Product Reports*, **18**, 291-309 (2001).

[3] Wemmer, D.E., *Biomol. Struct.*, **29**, 439-461 (2000).

[4] Ren, J., Chaires, J.B., *Biochemistry*, **38**, 16067-16075 (1999).

[5] Bailly, C., Chaires, J.B., Bioconjugate *Chem*., **9**, 513-538 (1998).

[6] Watson, J.D., Crick, F.H.C., *Nature*, **171**, 737-738 (1953).

[7] Dervan, P.B., *Bioorganic and Medicinal Chemistry*, **9**, 2215-2235 (2001).

[8] Gibson, D., *The Pharmacogenomics Journal*, **2**, 275-276 (2002).

[9] Pindur, U., Jansen, M., Lemster, T., *Current Medicinal Chemistry*, **12**, 2805-2847 (2005).

[10] Lerman, L.S., *Biochemistry*, **49**, 95-101 (1963).

[11] R.R. Tidwell, D.W. Boykin, *Journal of Brazilian Chemical Society*, **13**, 414-460 (2002).

[12] Drew, H.R., Wing, R.M., Takano, T., Broka, C., Tanaka, S., Itakura, K., Dickerson, R.E., *Proc Natl Acad Sci U. S. A.*, **78**, 2179-2183 (1981).

[13] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., *Nucleic Acids Research*, **28**, 235–242 (2000).

[14] Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, San Diego (2020).

[15] Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., *Journal of Computational* Chemistry, **25**, 1605-1612 (2004).

[16] Shastri, R., Yadav, A.K., Kumar, D., *Eur. Phys. J. Plus*, **132**, 313 (2017)

[17] Dubey, K.D., Chaube, A.K., Parveen, A., Ojha, R.P., *Journal of Biophysics and Structural Biology*, **2**, 47-54 (2010).

[18] Gaussian 09, Revision B.01, Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, G. A., Nakatsuji, H., Li, X., Caricato, M., Marenich, A. V., Bloino, J., Janesko, B. G., Gomperts, R., Mennucci, B., Hratchian, H. P., Ortiz, J. V., Izmaylov, A. F., Sonnenberg, J. L., Williams-Young, D., Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D., Zakrzewski, V. G., Gao, J., Rega, N., Zheng, G., Liang, W., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Throssell, K., Montgomery, J. A., Jr., Peralta, J. E., Ogliaro, F., Bearpark, M. J., Heyd, J. J., Brothers, E. N., Kudin, K. N., Staroverov, V. N., Keith, T. A., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A. P., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Millam, J. M., Klene, M., Adamo, C., Cammi, R., Ochterski, J. W., Martin, R. L., Morokuma, K., Farkas, O., Foresman, J. B., Fox, D. J. Gaussian, Inc., Wallingford CT, 2009.

[19] Mishra, R., Gaur, A.S., Chandra, R., Kumar, D., *International Journal of Science, Technology and Society*, **3**, 11-27 (2017).

[20] Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J., *Journal Computational Chemistry*, **16**, 2785-91 (2009).

[21] Zhao, H., Caflisch, A., *Eur J Med Chem*., **91**, 4-14 (2015).

[22] Abraham, M.J., Murtola, T., Schulz, R., Pall, S., Smith, J.C., Hess, B., Lindahl, E., *SoftwareX*, **1–2**, 19–25 (2015).

[23] Ponder, J.W., Case, D.A., *Adv. Prot. Chem.*, **66**, 27-85 (2003).

[24] da Silva, A.W. S., Vranken, W.F., *BMC Res. Notes*, **5**, 367 (2012).

[25] Mark, P., Nilsson, L., *Journal Physical Chemistry-A*, **105**, 9954-9960 (2001).

[26] Darden, T., York, D., Pedersen, L., *J Chem Phys*., **98**, 10089-10092 (1993).

[27] Ryckaert, J.P., Ciccotti, G., Berendsen., H.J., *J Comput Phys.*, **23**, 327-341 (1977).

[28] Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M., *Journal of Computational Chemistry*, **18**, 1463-1472 (1997).

[29] Turner, P.J., XMGRACE, Version 5.1. 19. Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR (2005).

[30] Humphrey, W., Dalke, A. and Schulten, K*J. Molec. Graphics*, **14**, 33-38 (1996).

[31] Shastri, R., Awasthi, N., Kumar, D., Yadav, A.K., Roy, D., Goutam, S.P., Pandey, A., *Advanced Science, Engineering and Medicine*, **10**, 1-5 (2018).

[32] Shukla, R., Shukla, H., Kalita, P., Sonkar, A., Pandey, T., Singh, D.B., Kumar, A., Tripathi, T., *Journal of Biomolecular Structure and Dynamics*, **36**, 2147-2162, (2018).

[33] Chaubey, A.K., Dubey, K.D., Ojha, R.P., *Medicinal Chemistry Research*, **24**, 753–763 (2015).

[34] Genheden, S., Ryde. U., *Expert Opinion on Drug Discovery*, **10**, 449-46, (2015).

[35] Aamir, M., Singh, V.K., Dubey, M.K., Meena, M., Kashyap, S.P., Katari, S.K., Upadhyay, R.S., Umamaheswari, A., Singh, S., *Front. Pharmacol*., **9**, 1038 (2018).

Chapter-7

Conclusions

Chapter-7

Conclusions

This chapter focuses upon the conclusions drawn from the investigations carried out on the computational studies of drug-DNA interactions in this thesis. It also provides the strategies for further research work in the field of drug discovery via computational advances. The concluding remarks made and the recommendations suggested for future works are as follows:

1. The focus of the study is to confirm the importance of DNA sequence and specificity in directing the complex formation at molecular levels. And our study attempts to give detail insight on the complexity in binding modes of small molecules to DNA.
2. Docking results were well explained through various series of analysis and thus were found to be in good co-relation with the results reported in the literature.
3. Dynamical behavior of the system was well studied and critical analysis of the in-depth information regarding the stability of the drug-DNA complex with evolving time was done by evaluating various parameters such as, variations in energy, variations in radii of gyrations, variations in number of hydrogen bonds, RMSD and RMSF.
4. The free energy calculations performed also added to the computational rigrousity of the calculations by incorporating the electrostatic effects and their exhaustiveness added to more detailed results, viz., energy contributions for

each component (van der Waals energy, electrostatic energy, polar solvation energy, binding energy and SASA energy) and also per-residue contribution was also obtained and hence critical analysis of the obtained results was achieved.

5. QM/MM calculations provide much better results than that of other molecular modeling methods such as molecular docking and molecular dynamics results owing to incorporation of DFT.
6. The use of MD simulations approach with QM/MM calculations allowed us to provide a theoretical protocol for complementing experimental techniques.

These collaborative computational modelling studies performed at electronic structure levels can give a great insight and add huge valuable information in designing novel inhibitors in close relationship with experimental works.

APPENDIX

A. List of Publications

B. List of Communicated Manuscripts

C. Developed Scripts

A. List of Publications

1. **Anwesh Pandey**, Ruchi Mishra, Anil Kumar Yadav; **Understanding interactions of DNA minor groove binders using advanced computational techniques**, *The International journal of analytical and experimental modal analysis*, 2020, 12, 1300-1315.
2. **Anwesh Pandey,** Rolly Yadav, Anamika Shukla, and Anil Kumar Yadav; **Unveiling the Antimicrobial Activities of Dicationic Carbazoles and Related Analogs Through Computational Docking**, *Adv. Sc. Engg. Med*., 2020, 1, 40-44.
3. **Anwesh Pandey**, Ruchi Mishra, Anamika Shukla, Anil Kumar Yadav, Devesh Kumar; **In- silico docking studies of 2,5-bis(4-amidinophenyl) furan and its derivatives**, published in conference proceedings of *ISAFBM-2019.*
4. Rolly Yadav, **Anwesh Pandey**, Nidhi Awasthi, and Anamika Shukla; **Molecular Docking Studies of Enzyme Binding Drugs on Family of Cytochrome P450 Enzymes**, *Adv. Sc. Engg. Med*., 2020, 1, 83-87.
5. Anamika Shukla, Ruchi Mishra, **Anwesh Pandey**, A.K. Dwivedi, Devesh Kumar; **Interactions of Flavonols with DNA: Molecular Docking Studies**, published in conference proceedings of *ISAFBM-2019*.
6. Rajkamal Shastri, Nidhi Awasthi, Devesh Kumar, Anil Kumar Yadav, Diptarka Roy, SP Goutam, **Anwesh Pandey**; **A Density Functional Theory Study on Structural Stability and Electronic Properties of Co_xO_y (x+y=4-12) Nanoclusters**; *Adv. Sc. Engg. Med*., 2018, 10, 814-818.

B. List of Communicated Manuscripts

1. **Anwesh Pandey**, Ruchi Mishra, Anil Kumar Yadav, Devesh Kumar; ***DNA binding study of Dicationic Carbazoles and its Analogs***.
2. **Anwesh Pandey**, Rolly Yadav, Anamika Shukla, Anil Kumar Yadav, Devesh Kumar; ***In- silico docking studies of 2,4-bis(4-amidinophenyl) furan and its derivatives***.
3. **Anwesh Pandey**, Anil Kumar Yadav, Devesh Kumar; ***Exploring interactions of some DNA binding ligands through Molecular Docking, Molecular Dynamics Simulations and Quantum Mechanical/Molecular Mechanical (QM/MM) Calculations***.
4. **Anwesh Pandey**, Anil Kumar Yadav, Devesh Kumar; ***Interaction, Dynamics & Stability analysis of some minor groove binders with B-DNA dodecamer 5'-(CGCAAATTTGCG)- 3'.***
5. **Anwesh Pandey**, Anil Kumar Yadav, Devesh Kumar; ***Molecular Dynamics insights into interaction and stability of ligands in the vicinity of DNA***.
6. Rolly Yadav, **Anwesh Pandey**, Nidhi Awasthi, Devesh Kumar; ***Interaction of Carbazoles and their analogs with human Cytochrome***.
7. Anamika Shukla, Ruchi Mishra, **Anwesh Pandey**, Rolly Yadav, Devesh Kumar; ***Interaction of Flavones with DNA: Molecular Docking Studies.***
8. Anamika Shukla, Ruchi Mishra, **Anwesh Pandey**, Asheesh Kumar, Devesh Kumar; ***Evaluation of binding properties of some common drugs with DNA as intercalator and groove binder***.

C. Developed Scripts

---

1. Bash (Shell) script for geometry optimization using Gaussian09

2. Python script for extracting data from Gaussian09 log file

3. Bash (Shell) script for automated docking using Autodock4

4. Python script for extracting data from Autodock4 log file

5. Bash (Shell) script for molecular dynamics simulation using GROMACS with amber03 force field

6. Bash (Shell) script for molecular dynamics simulation using GROMACS with charmm27 force field

7. Bash (Shell) script for molecular dynamics data analysis for amber03 & charmm27 force fields

8. Bash (Shell) script for ligand preparation for molecular dynamics simulation using AMBER software

1. Bash (Shell) script for geometry optimization using Gaussian09

```
#!/bin/bash
for i in 1 2 3 4
do
cd $i
g09 $i.com
cd ..
done
echo "*****optimization done*****"
```

2. Python script for extracting data from Gaussian09 log file

```
import glob
import sys
sys.stdout=open("data.txt", "w")
for fn in glob.glob('*.log'):
f=open(fn, 'r')
atom=[]; basis=[]; charge=[]; scf=[]; homo=[];
lumo=[]; freq=[]; meth=[]; multi=[]
while 1:
  line=f.readline()
  if not line: break
       if line[0:8]==' NAtoms=':
       at1=line[11:13]
       atom.append(at1)
       if line[0:16]==' Standard basis:':
       bs1=line[17:25]
       basis.append(bs1)
       if line[0:9]==' Charge =':
       ch1=line[11:12]
       charge.append(ch1)
       if line[0:10]==' SCF Done:':
       scf1=line[25:39]
       scf.append(scf1)
       if line[0:12]==' Frequencies':
       fq1=line[21:73]
       freq.append(fq1)
       if line[0:27]==' Alpha   occ. eigenvalues --':
       homo1=line[28:38]
       homo.append(homo1)
       if line[0:27]==' Alpha virt. eigenvalues --':
       lumo1=line[28:38]
```

```
        lumo.append(lumo1)
        if line[0:3]==' #p':
        me1=line[13:19]
        meth.append(me1)
        if line[0:9]==' Charge =' and line[11:12]=='0':
        mul1=line[28:29]
        multi.append(mul1)
  print fn, '\t', atom[1], '\t', basis[1], '\t',
charge[1], '\t', scf[-1], '\t', freq[0], '\t', homo[-
1], '\t', lumo[1], '\t', meth[0], '\t', multi[0]
sys.stdout.close()
```

3. Bash (Shell) script for automated docking using Autodock4

```
#!/bin/bash
for i in lig1 lig2 lig3 lig4
do
cd $i
prepare_ligand4=/home/anwesh/softwares/mgltools_x86_6
4Linux2_ 1.5.6/bin/pythonsh prepare_ligand4.py -l
$i.pdb
prepare_receptor4=/home/anwesh/softwares/mgltools_x86
_64Linux 2_1.5.6/bin/pythonsh prepare_receptor4.py -r
rec.pdb
prepare_gpf4=/home/anwesh/softwares/mgltools_x86_64Li
nux2_1.5.6/bin/pythonsh prepare_gpf4.py -l $i.pdbqt -
r rec.pdbqt -o grid.gpf
autogrid4 -p grid.gpf -l grid.glg
prepare_dpf42=/home/anwesh/softwares/mgltools_x86_64L
inux2_1. 5.6/bin/pythonsh prepare_dpf42.py -l
$i.pdbqt-r rec.pdbqt -o dock.dpf
autodock4 -p dock.dpf -l dock.dlg
cd ..
done
echo "*****docking done*****"
```

4. Python script for extracting data from Autodock4 log file

```
import glob
import sys
sys.stdout=open("data.txt", "w")
for fn in glob.glob('*.dlg'):
f=open(fn, 'r')
be=[]
while 1:
  line=f.readline()
  if not line: break
       if line[0:4]=='     1':
       be1=line[11:17]
       be.append(be1)
  print(fn, be)
sys.stdout.close()
```

5. Bash (Shell) script for molecular dynamics simulation using GROMACS with amber03 force field

Step 1:

```
#!/bin/bash
for i in 1 2 3 4
do
cd $i
gmx pdb2gmx -ignh -f dna.pdb -o conf.gro -water tip3p
cd ..
done
echo "**********"
```

Step 2:

```
#!/bin/bash
for i in 1 2 3 4 do
cd $i
gmx editconf -f conf.gro -o newbox.gro -c -bt
dodecahedron -d 1.2
gmx solvate -cs -cp newbox.gro -p topol.top -o
solv.gro gmx grompp -f em.mdp -c solv.gro -p
topol.top -o ions.tpr
gmx genion -s ions.tpr -o solv_ions.gro -p topol.top
-pname NA -nname CL -np 20
cd ..
done
echo "**********"
```

Step 3:

```
#!/bin/bash
for i in 1 2 3 4
do
cd $i
gmx grompp -f em_real.mdp -c solv_ions.gro -p
topol.top -o em.tpr
gmx mdrun -v -nt 8 -deffnm em
cd ..
done
echo "**********"
```

Step 4:

```
#!/bin/bash
for i in 1 2 3 4
do
cd $i
gmx genrestr -f lig_GMX.gro -o posre_lig.itp -fc 1000
1000 1000
cd ..
done
echo "**********"
```

Step 5:

```
#!/bin/bash
for i in 1 2 3 4
do
cd $i
gmx make_ndx -f em.gro -o index.ndx
cd ..
done
echo "**********"
```

Step 6:

```
#!/bin/bash
for i in 1 2 3 4
do
cd $i
gmx grompp -f nvt.mdp -c em.gro -p topol.top -n
index.ndx -o nvt.tpr
gmx mdrun -v -nt 8 -deffnm nvt
gmx grompp -f npt.mdp -c nvt.gro -t nvt.cpt -p
topol.top -n index.ndx -o npt.tpr
gmx mdrun -v -nt 8 -deffnm npt
gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p
topol.top -n index.ndx -o md_0_1.tpr
gmx mdrun -v -nt 8 -deffnm md_0_1
cd ..
done
echo "**********"
```

6. Bash (Shell) script for molecular dynamics simulation using GROMACS with charmm27 force field

Step 1:

```
#!/bin/bash
for i in 1 2 3 4
do
cd $i
gmx pdb2gmx -f dna.pdb -ff conf.pdb -nochargegrp
charmm27 -water tip3p -ignh -o
cd ..
done
echo "**********"
```

Step 2:

```
#!/bin/bash
for i in 1 2 3 4
do
cd $i
gmx editconf -f  conf.pdb -o boxed.pdb -c -d 1.2    -
bt octahedron
gmx solvate -cs -cp boxed.pdb -o solvated.pdb -p
topol.top gmx grompp -f em.mdp -c solvated.pdb -p
topol.top
trjconv-s topol.tpr -f     solvated.pdb -o
compact.pdb -ur compact -pbc mol
cd ..
done
echo "**********"
```

Step 3:

```
#!/bin/bash
for i in 1 2 3 4
do
cd $i
gmx editconf -f conf.pdb -o boxed.pdb -bt octahedron
-d 1.0 gmx solvate -cp boxed.pdb -o solvated.pdb -p
topol.top
gmx grompp -f em.mdp -c solvated.pdb -p topol.top -o
ions.tpr
gmx genion -s ions.tpr -o solvated.pdb -p topol.top -
pname NA -np 22
cd ..
done
echo "**********"
```

Step 4:

```
#!/bin/bash
for i in 1 2 3 4
do
cd $i
gmx grompp -f em_real.mdp em.tpr -c solvated.pdb -p
topol.top -o
gmx mdrun -v -nt 6 -deffnm em
cd ..
done
echo "**********"
```

Step 5:

```
#!/bin/bash
for i in 1 2 3 4
do
cd $i
gmx genrestr -f  ligand.pdb -o posre_lig.itp -fc 1000
1000 1000
cd ..
done
echo "**********"
```

Step 6:

```
#!/bin/bash
for i in 1 2 3 4
do
cd $i
gmx make_ndx -f em.gro -o index.ndx
cd ..
done
echo "**********"
```

Step 7:

```
#!/bin/bash
for i in 1 2 3 4
do
cd $i
gmx grompp -f nvt.mdp -c em.gro -p topol.top -n
index.ndx -o nvt.tpr
gmx mdrun -v -nt 6 -deffnm nvt
gmx grompp -f npt.mdp -c nvt.gro -t nvt.cpt -p
topol.top -n index.ndx -o npt.tpr
gmx mdrun -v -nt 6 -deffnm npt
```

```
gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p
topol.top -n index.ndx -o md_0_1.tpr
gmx mdrun -v -nt 6 -deffnm md_0_1
cd ..
done
echo "**********"
```

7. Bash (Shell) script for molecular dynamics data analysis for amber03 & charmm27 force fields

```
#!/bin/bash
for i in 1 2 3 4
do
cd $i
trjconv -s md_0_1.tpr -f md_0_1.xtc -o
md_0_1_noPBC.xtc -pbc mol -ur compact
trjconv -s md_0_1.tpr -f md_0_1_noPBC.xtc -n
index.ndx -dt 1 -b 1 -o last.pdb -fit rot+trans
g_rms -s md_0_1.tpr -f md_0_1_noPBC.xtc -n index.ndx
-o rmsd.xvg -tu ps
g_rmsf -s md_0_1.tpr -f md_0_1_noPBC.xtc -n index.ndx
-o rmsf.xvg
g_hbond -s md_0_1.tpr -f md_0_1_noPBC.xtc -num
hbond.xvg
g_gyrate -s md_0_1.tpr -f md_0_1_noPBC.xtc -n
index.ndx -o gyrate.xvg
g_energy -f md_0_1.edr -o ener.xvg
cd ..
done
echo "**********"
```

8. Bash (Shell) script for ligand preparation for molecular dynamics simulation using AMBER software

```
#!/bin/bash
for i in lig1 lig2 lig3 lig4
do
cd $i
antechamber -i $i.pdb -fi pdb -o $i.mol2 -fo mol2 -c
bcc -s 0
parmchk2 -i $i.mol2 -f mol2 -o $i.frcmod
cd ..
done
echo "**********"
```