NEW PROTOCOL FOR SCANNING MINOR AND MAJOR GROOVES ALONG DNA TO FIND BEST INTERACTION MODE WITH DRUG MOLECULES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

AYBÜKE GÜLKAYA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
CHEMISTRY

JANUARY 2022

Approval of the thesis:

NEW PROTOCOL FOR SCANNING MINOR AND MAJOR GROOVES ALONG DNA TO FIND BEST INTERACTION MODE WITH DRUG MOLECULES

submitted by **AYBÜKE GÜLKAYA** in partial fulfillment of the requirements for the degree of **Master of Science** i**n Chemistry Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Özdemir Doğan
Head of the Department, **Chemistry** _____

Assist. Prof. Dr. Antoine Marion
Supervisor, **Chemistry, METU** _____

Examining Committee Members:

Assoc. Prof. Dr. Özgül Persil Çetinkol
Chemistry Dept., METU _____

Assist. Prof. Dr. Antoine Marion
Chemistry Dept., METU _____

Assist. Prof. Dr. Erol Yıldırım
Chemistry Dept., METU _____

Assoc. Prof. Dr. Mehmet Fatih Danışman
Chemistry Dept., METU _____

Prof. Dr. Fethiye Aylin Sungur
Computational Science and Engineering Division, ITU Informatics Institute _____

Date: 18.01.2022

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name Last name : Aybüke Gülkaya

Signature :

ABSTRACT

NEW PROTOCOL FOR SCANNING MINOR AND MAJOR GROOVES ALONG DNA TO FIND BEST INTERACTION MODE WITH DRUG MOLECULES

Gülkaya, Aybüke
M.S., Chemistry Department
Supervisor : Assist. Prof. Dr. Antoine Marion

January 2022, 126 pages

Major and minor groove binder drugs could function as possible treatment agents in genetic disorders or cancer. DNA is recognized by different proteins during replication, translation, and transcription. Basically, by binding to the grooves of DNA, small drug molecules can change the helical structure of DNA, interfere with protein binding and prevent replication, translation or transcription. To alter and inhibit DNA - protein interactions, especially sequence specific ligands should be designed. Accordingly, we investigated the interactions of different anticancer drugs binding to specific DNA sequences via minor or major grooves by a novel, automatizated protocol as a combination of docking and molecular mechanics refinement. With this methodology, we obtained best interaction modes of the drug molecules with given DNA sequences and most preferred sequences.

Keywords: Molecular Docking, Molecular Mechanics, Binding Free Energy, Molecular Modelling, Drug Discovery

ÖZ

DNA MOLEKÜLÜNÜN BÜYÜK VE KÜÇÜK OLUKLARININ TAMAMINI TARAYARAK EN İYI LİGAND VE DNA ETKİLEŞİMİNİ BULMAK ÜZERE METOT GELİŞTİRİLMESİ

Gülkaya, Aybüke
Yüksek Lisans, Kimya Bölümü
Tez Yöneticisi: Dr. Öğretim Üyesi Antoine Marion

Ocak 2022, 126 sayfa

DNA'nın büyük ve küçük oluklarına bağlanan ilaçlar, kanser veya genetik hastalıkların tedavisinde kullanılmaktadır. Replikasyon, translasyon ve transkripsiyon sırasında, DNA farklı proteinler tarafından tanınır. Temel olarak, ilaç molekülleri DNA'nın oluklarına bağlanarak DNA'nın sarmal yapısını değiştirir ve replikasyon mekanizmasını engeller. Farklı DNA ve protein etkileşimlerini inhibe edecek yeni ilaçlar ve yöntemler geliştirmek için, farklı sekanslarla etkileşim gösteren spesifik moleküller geliştirilmelidir. Bu nedenle yeni geliştirdiğimiz hızlı protokol ile DNA moleküllerine ilaç yerleştirme ve moleküler mekanik ile farklı dizilerin küçük ve büyük oluklarının, farklı ligandlarla etkileşimleri incelenmiştir. Bu method ile, seçilen ilaçların DNA ile en iyi etkileşime sahip olduğu bağlanma konformasyonları ve seçici oldukları DNA sekansları elde edilmiştir.

Anahtar Kelimeler: Moleküler Kenetleme, Moleküler Mekanik, Bağlanma Serbest Enerjisi, Moleküler Modelleme , İlaç Keşfi

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisor Assist. Prof. Dr. Antoine Marion for his guidance, continuous support, sincerity, motivation, vision, encouragements and insight throughout the research. His counseling and brilliant way of thinking contributed me to gain critical thinking ability and scientific perspective. Also, I would like to thank Dr. İlke Uğur Marion for her valuable suggestions in the docking part.

I would like to express my thankfulness to my thesis committee members Assoc. Prof. Dr. Özgül Persil Çetinkol, Assist. Prof. Dr. Antoine Marion, Assist. Prof. Dr. Erol Yıldırım, Assoc. Prof. Dr. Mehmet Fatih Danışman and Prof. Dr. Fethiye Aylin Sungur for their kind interest to participate in my thesis defense and their insightful comments and feedbacks.

I would like to give my special thanks to my previous lab members Assoc. Prof. Dr. Özgül Persil Çetinkol and Dr. Mehrdad Forough. They have lighted the fire of science inside me with their enthusiasm, motivation, sincerity and knowledge and contributed me both academically and socially since I have met them.

I also want to thank my family since they have supported me in every step of my life with love.

I am so lucky to have my dear friends Gamze Ağacayak, Cem Kaya, Ecrin Çalışan, Gülbihan Bestelci, İrem Beril Karaçalık, Elif Büke, Buse Çankaya, Rojda Kalkan, Gizem Kars, Ekinsu Özmen, Koray Beyaz, Muhammet Durmaz, Elif Gür, Kübra Ferah and Edanur Metin. They make the world funny, enjoyable and better place to live for me with their presence, support and fun.

I want to also thank MML team Dr. Georgios Antipas and Dr. Nikolaos Ntallis who encouraged and supported me while I am writing my thesis.

I would like to express my endless love and deepest thanks to my heroes Batuğhan Çavuş and Yoda the huge dog who covered every site of my heart. They have given so much of their time, love, energy and endless support. I always felt their warmth around me. They have always stood by me at every challenge that I faced. I feel so blessed and lucky to have them.

Calculations are performed by Turkish National e-science e-Infrastructure (TRUBA) of TUBITAK ULAKBIM High Performance and Grid Computing Center.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS

E :	Energy
DNA :	Deoxyribonucleic acid
3D :	Three dimensional
5-hmC :	5-hydroxymethyl cytosine
5mC :	5-methylcytosine
A :	Adenine
AMBER :	Assisted Model Building with Energy Refinement
bZIP :	Basic leucine zipper domains
˚C :	Celcius degree
C :	Cytosine
CADD :	Computer-Aided Drug Design
CDK :	Cyclin dependent kinase
DBD :	DNA binding domains
DNMT :	DNA methyltransferase
ds :	Double stranded
ds-DNA :	Double stranded deoxyribonucleic acid
G :	Guanine
G_0 :	Quiescence phase
G_1 :	Growth phase 1 (First Gap)
G_2 :	Growth phase 2 (Second Gap)
GB :	Generalized Born model
H :	Hydrogen
HF :	Hartree-Fock
K :	Kelvin
LBDD :	Ligand-based drug discovery
M :	M phase
MD :	Molecular dynamics
MM	Molecular mechanics
MM/GBSA :	Molecular Mechanics/ Generalized Born Surface Area
MM-refinement :	Molecular mechanics-based refinement
N :	Nitrogen
O :	Oxygen
ps :	Picoseconds
Pi :	Phosphate atom
QM :	Quantum mechanics

QSAR :	Quantitative structure-activity relationship
RNA :	Ribonucleic acid
RNAi :	RNA interference
S :	Synthesis phase
SASA :	Solvent-accessible surface area
SBDD :	Structure-based drug discovery
ssDNA :	Single stranded deoxyribonucleic acid
T :	Thymine
TBP :	TATA-box binding proteins
THT :	Helix-turn-helix domains
U :	Uracil

LIST OF SYMBOLS

SYMBOLS

Å	angstrom
ΔH	enthalpy change
K_θ	force constant for angles
ΔS	entropy change of the system
ΔG_{sol}	solvation free energy
K_b	force constant of covalent bonds
L	ligand
LR	complex
q_i	atomic charge of the i^{th} atom
R	receptor
R_i	effective Born radius of atom *I*
r_{ij}	the inter-particle distance
T	absolute temperature
V_n	force constant of dihedral angle
$\emptyset$	dihedral angle
ϵ	dielectric constant
ΔG_{bind}	Gibbs binding free energy
ΔG_{solv}	Gibbs solvation free energy
ΔG_{GB}	electrostatic solvation energy of polar contribution
ΔG_{SA}	electrostatic solvation energy of nonpolar contribution
ΔE_{MM}	change in the gas phase molecular mechanics (MM) energy
ΔE_{int}	internal energy change
ΔE_{ele}	electrostatic energy change
ΔE_{vdW}	van der Waals energies

CHAPTER 1

INTRODUCTION

Major and minor groove binder drugs could function as possible treatment agents in genetic disorders or cancer. DNA is recognized by different proteins during replication, translation, and transcription. Basically, by binding to the grooves of DNA, small drug molecules can change the helical structure of DNA, interfere with protein binding and prevent replication, translation or transcription(Khan et al., 2012).

In the current literature, it has been shown that anticancer drugs or antibiotics prefer specific sequences in different grooves of DNA for binding(Rahman et al., 2019). However, the interaction patterns and mechanisms between drug molecules and DNA remain unclear for many of them. Understanding the mechanisms of interactions between DNA and drug molecules would provide significant help in designing novel drugs that would bind specifically to the target sites.

To alter and inhibit DNA - protein interactions, especially sequence specific ligands should be designed. Groove binder drugs bind to DNA by non-covalent interactions with different binding strengths depending on the sequence specificity of the drug, and thus might suppress different transcription factors or oncogenes. Within this thesis, we investigate the interactions of different major and minor groove binders with the minor and major grooves of different DNA sequences by a new and automatized docking method as combination of docking, molecular mechanics refinement and molecular dynamics protocol that we developed.

In this study, four major groove binder drugs, namely altromycin B, pluramycin A, hedamycin, nogalamycin and six minor groove binder drugs, berenil, DAPI, distamycin, hoechst33258, netropsin and pentamidine were selected as ligands to investigate the best interacting binding site by scanning along the whole major and

minor grooves of AAAAAAAAAA, GGGGGGGGGG, ATCGCGCGAT, CGTATATACG and GGCCAATTGG with this new automatization protocol.

1.1 DNA Structure

DNA (deoxyribonucleic acid) is the genetic code that carries molecular information and transferred throughout generations by translation, transcription, and replication. A DNA chain consists of nucleotides of four major bases: i.e., adenine (A), guanine (G), cytosine (C) and thymine (T) and a five-carbon deoxyribose sugar linked by phosphodiester bonds. The horizontally aligned base pairs are stacked in the interior and sugar-phosphate backbone surrounds the stacked base pairs. At physiological pH, each phosphate has a negative charge and, therefore, the genomic DNA is a highly negatively charged double stranded polymer built up by two intertwined single strandsthat are held together by hydrogen bonds among the bases of two opposite helices. Orientations of the nucleotides in a single strand are indicated by 5' and 3' directions, which refer to 5^{th} and 3^{rd} carbon atom of the deoxyribose sugars, respectively. Phosphodiester bonds links the adjacent nucleotides in the backbone via the 5' phosphate of one nucleotide to hydroxyl of 3' end of the other nucleotide. There are three main known conformations of DNA: B-DNA, A-DNA and Z-DNA(Travers & Muskhelishvili, 2015). B-DNA is right-handed and is the most common structure at physiological pH as shown in figure 1.

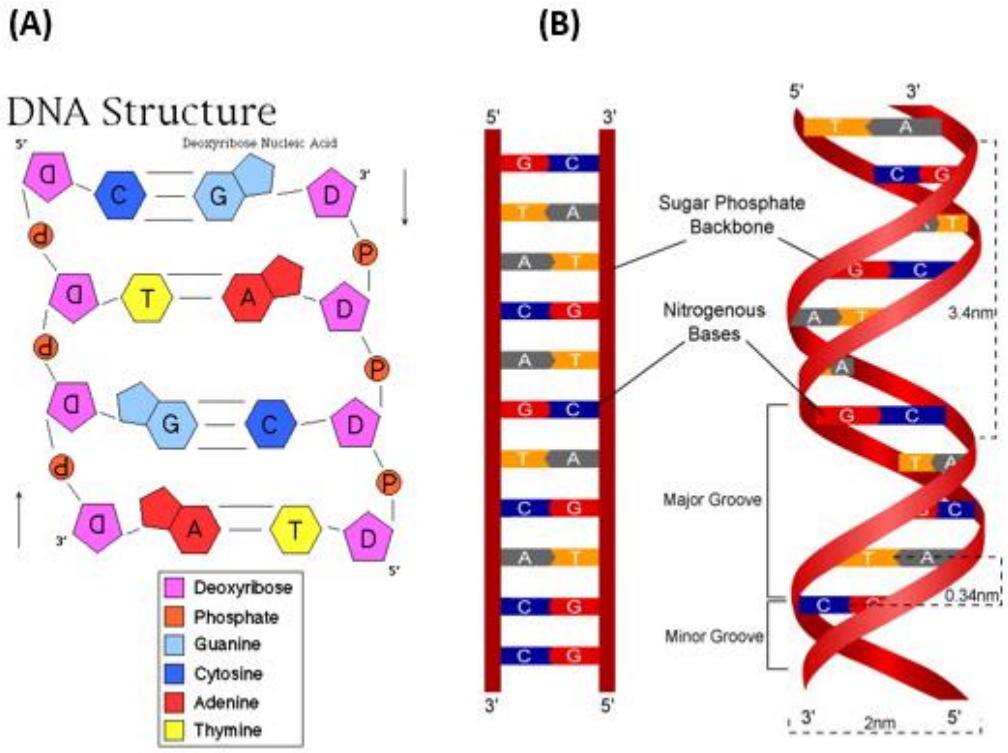


Figure 1. DNA Structure (A). There are three hydrogen bonds among the cytosine and guanine bases and two hydrogen bonds among the adenine and thymine bases. Backbone consists of covalently attached alternating phosphate and deoxyribose sugars. In (B), B-DNA is shown with major and minor grooves. (Northern Arizone University, n.d.)

The negative charges on phosphate groups in the backbone of each opposite strand create repulsion among them. However, the hydrogen bonds of paired bases and hydrophobic interactions and π-π stacked bases in this specific stair wise structure help to balance this negatively charged backbone repulsion and stabilize the double stranded helix. Additionally, the solvation and ions around the ds-DNA is also crucial for the stability.

DNA is so important that it has been studied a lot over the past decades and a lot of mysteries have been resolved as shortly described above. However, we still need to know more about the energetics, thermodynamics, behaviour, and machinery of DNA interactions. This would help us to rationally target DNA via drugs aiming specific groove and/or sequences. Moreover, solving out the specific interactions

between drugs and specific sites of DNA would open up new doors in design of better drugs with respect to their interaction patterns of either covalent interactions, intercalation, or weak interactions.

1.2 Anticancer and Antibiotic Activity of Small Molecules

Cancer and bacterial infections lead quite high death rates around the whole world. It has been stated (Soo et al., 2016) that cancer cells/tumors and bacterial infections have similar features due to similar activity like spreading quickly in host, similar metabolic pathways, quick resistance promotion (chemotherapy resistance and antibiotic resistance) and aggressive proliferation during disease progression. Therefore some anticancer drugs are used also as antibacterial agents(Soo et al., 2016). In clinical applications, new antibiotics and anticancer drugs are needed more and more due to increasing antibiotic and chemotherapy resistance. Furthermore, the pursuit of new cancer drug candidates aims at limiting toxicity and malignant side effects on host. On the other hand, pharmaceutical companies have not released a new brand of antibiotic in over a decade, which leads to a serious problem that is, clinical applications are ineffective due to drug resistance. In this regard, developing new methods to design new drugs by raveling out the interaction patterns and mechanism of those medicines without toxicity and side effects holds potential to speed up the discovery of new therapeutics (Soo et al., 2016). Also, small molecules have potential of therapeutic effect due to their higher cell permeability opposite of nucleic acid-based approaches (antisense, decoy and triple helix-forming oligonucleotides, ribozymes and RNAi), retroviral or adenovirus vectors, liposomes, and designed zinc finger peptides which cannot enter inside the cell easily due to low permeability (Melander et al., 2019).

Cancer can be defined as uncontrolled and rapid cell proliferation and spreading. Anticancer drugs and antibiotics basically impair the cancer cell cycles in different phases and ways. There are three major anticancer drug interactions that are: interaction with DNA-binding protein, RNA-ssDNA hybridization along the

replication forks, and ligand-DNA binding (Hamad, F. D. , 2017). How the ligand and DNA interact changes upon the ligand structure, interaction energetics, patterns formed by hydrogen bond donor and acceptor groups, depending on the composition of bases, and compatibility of the size of the drug with size of the binding pockets on DNA. Since the way in which a ligand interacts with DNA depends on many parameters and patterns, it is crucial to resolve interaction mechanics in order to propose new treatments and modifications on drugs to decrease their cytotoxicity (Hamilton & Arya, 2012). In this study, the focus has been concentrated on the type of drugs that are directly interacting with DNA and that prevent either RNA transcription, DNA replication, or translation cycles, which should lead to apoptosis.

1.3 Gene Expression Regulation

Genes carry the required information to make proteins, and cell can function properly if proteins are synthesized at the correct time and in the right amount. Herein, gene expression can be defined as producing RNA and proteins by using genes. Transcription is the first step of gene expression where the DNA sequence is used to synthesize mRNA molecule, which will later be proccessed to synthesize the desired protein. Since transcription is the very initial point of gene expression, it is a relevant target to regulate the gene expression, to select which genes will be transcribed into the primary RNA transcript and hence proteins (Latchman, 1993). For every different cell type, the genes that are expressed to make proteins are different. These different genes transcribed in different cells are activated and regulated by proteins called transcription factors. This gene regulation allows to form different organ and tissue cells that have specific structure and function depending on the which genes are actively transcribed and translated (Hillenbrand et al., 2016).

1.4 Transcription Factors

Transcription factors are DNA-binding proteins that control the activity of transcription by recognizing the regulatory region of DNA close to the initial point of transcription. This DNA regulatory region recognition is extremely selective and is formed by non-covalent interactions and hydrogen bonding patterns of the surfaces of DNA bases in major and minor grooves and DNA binding domains (DBD) of the transcription factor. As the number of complementary contacts between transcription factor and DNA interface increases, specificity and strength of DNA-protein interlocking increases, and sequence is recognized distinguishably (Alberts B, 2002). Major classes of DNA binding domains that recognize the major groove are helix-turn-helix domains (THT), basic leucine zipper domains (bZIP) and zinc finger domains (Struhl, 1989). Major classes of DNA binding domains that recognize the minor groove are TATA-box binding proteins (TBP), integration host factors (IHF), and HMG-Box domains (Bewley et al., 1998). TATA-box is one of the most important core promoters recognized by TBP from minor groove to initiate transcription that contains TATA sequences (Tora & Timmers, 2010).

1.5 Cancer Cell Behavior

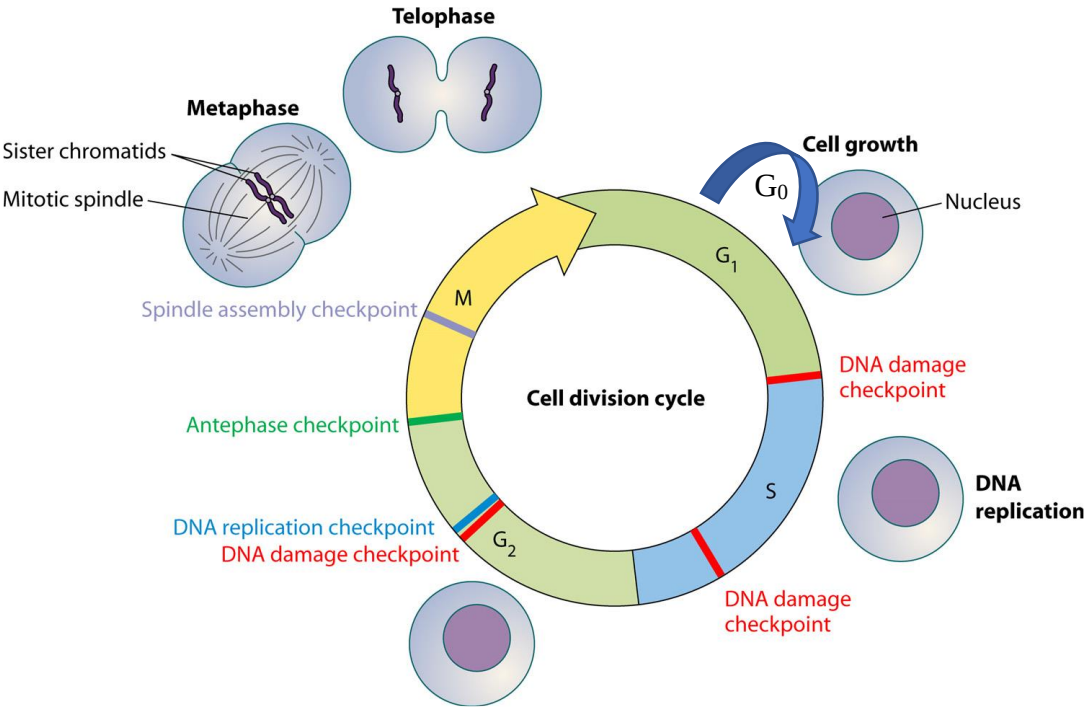


Figure 2. Cell cycle checkpoints(Chin & Yeong, 2015)

The cell cycle is critical to comprehend the behavior of cancer cells. It is a continuous process that is shown in figure 2. Normal cell is at rest at quiescence phase (G_0) and enters the cell cycle. Growth phase (G_1) is the first phase of the cell cycle where organelles of the cell, such as mitochondria, are duplicated. The cell enters the synthesis phase (S) when the DNA is replicated. Afterward, it enters the growth phase (G_2) where the cell grows again and prepares itself for the mitosis (M phase). In mitosis, cell divides itself to two identical cells. Those new cells either can enter the cell cycle or go back to the G_0 phase. In the cell cycle there are checkpoints to verify the problems occurred during the cell cycle. The first checkpoint (G_1 checkpoint) is at the G_1 phase. It assures the lack of any abnormality in the cell and in the DNA. The second checkpoint is G_2 checkpoint at the G_2 phase that checks whether there is no problem before the mitosis or not. The next checkpoint is found at the mitosis phase. When the cell enters the cell cycle, cyclins and CDKs (cyclin dependent kinases) start to be produced to allow the cell to enter and progress into

this continuous cell cycle. Those proteins can be regarded as the drivers of the cell cycle. For example, when the cell enters the cell cycle, it produces CDK4/6 and cyclin D in the early G_1 phase. When the cyclin D binds to CDK4/6, E2F detaches from the retinoblastoma protein and when E2F is released, it acts like a transcription factor which is allowing this cell to synthesis (S) phase. However, at the end of the G_1 phase, there is another CDK and cyclin which are CDK2 and cyclin E. Once at the S phase, the cell will produce another CDK and cyclin which are CDK1/2 and cyclin A. At the G_2 phase, CDK1 and cyclin B are produced that allow the cell to go through the cell cycle. So, the CDK's and cyclins are the drivers of the cell cycle. These proteins can hold cells in G_0 (resting) or even kill the cell if there is an abnormality to prevent tumor development. Alteration in one of the cyclin regulation pathways or mutations may hamper the control of the cell division. For example, if the amounts of those proteins are low, the cell does not progress in the cell cycle and if the amount of those proteins is too high, these cells continuously enter the cell cycle. As a result, cell growth becomes uncontrolled which is one of the mechanisms of cancer development (Hydbring et al., 2016).

A point mutation is a type of mutation where a single change of nucleotide occurs in a gene coding for a specific protein. An oncogenic mutation of one of the genes involved in checkpoints may cause increase in CDK and cyclin proteins within the cell which leads to propagation of the cell to S phase without the detection of appropriate extracellular growth factor. DNA amplification is another reason for the overexpression of cyclin in cancer cells that occurs when a certain gene gets amplified so many times such that the overexpressed cyclins function as oncoproteins. The other factor leading to overexpression of cyclins is called chromosomal rearrangement where the chromosome basically attached to one another where it should not. Another factor is epigenetic modifications such as methylation and acetylation of genes. Methylation and acetylation of genes provide silencing or activation of certain genes. With one of these alterations, a cell becomes a cancer cell in which the checkpoints of the cell cycle have been bypassed and grows in uncontrolled way. This uncontrolled cell growing is essentially caused by two

major changes occurring in the cancer cell which are activation of oncogenes like RAS and MYC genes and inactivation of tumor suppressor genes like p53, APC and BRCA1/2. When these genetic changes that are inactivation of tumor suppressor gene and activation of oncogene occur, the abnormal cell can enter the cell cycle and bypass the checkpoints and can continuously grow (Kastan & Bartek, 2004).

1.5.1 Oncogene Activation

Overexpression or mutation of oncogenes may result in dysregulated cell signaling, proliferation and tumorigenesis. For instance, RAS gene switches signaling pathways to control the differentiation or proliferation of cells and conveys signals from cell membrane to the nucleus. RAS gene produces RAS proteins that are mainly found below the plasma membrane. RAS proteins are activated once the growth factor receptor is stimulated by the cell growth. Activated RAS proteins cause a cascade of intracellular phosphorylation of specific tyrosine residues, which eventually activate transcription factors. Activated transcription factors bind to DNA and read the genes to produce CDK's and cyclin proteins particularly to allow this cell to go from G_1 phase to S phase for cell growth. When mutations occur in the RAS-coding gene, RAS proteins remain in the active state. Therefore, the cascade of phosphorylation and activation of transcription factors take place continuously. It leads to overproduction of proteins that allow cell growth like cyclins and CDKs (Adjei, 2001; Simanshu, D. K., Nissley, D. V. , McCormick, 2017).

MYC gene normally produce proteins that are responsible for cell growth, survival and activity. When a mutation occurs in the MYC-coding gene, cell growth and activity increases (Dang, 1999). Thus, activation of these oncogenes will allow the cell to bypass the checkpoints of the cell cycle, resulting in the uncontrolled cell growth.

1.5.2 Tumor Suppressor Gene Deactivation

Normally, healthy cells are able to stop any problematic cell to enter the cell cycle. For example, when a cell gets held up at G_2 phase due to an abnormal or damaged DNA, the cell will not progress in the cell cycle. When the cell has damaged DNA, it produces p53 proteins via the tumor suppressor gene p53 which can act like a transcription factor. Those p53 proteins will interact with DNA to produce cell arrest proteins such as p21. P21 protein inhibits the drivers of the cell cycle which are CDK and cyclin proteins and halt the progression of the cell cycle. P53 gene also produces proteins which are required for cell repair. So, when the cell is arrested, it can repair itself and its DNA. P53 protein also makes the proteins important for the apoptosis of the abnormal cell. If a mutation leads to inactivation of the tumor suppressor gene p53, cell arrest proteins, cell repair proteins, and proteins for apoptosis cannot be produced. Therefore, the abnormal cell can continue to the cell cycle and bypasses the checkpoints. This will lead to continuous growth and proliferation (Chen, 2016).

1.5.3 DNA Methylation

Even though every cell in the body contains the same DNA sequence, different cells express different genes. This process is regulated also partly by chemical modifications of nucleotides, which, without altering the DNA sequence, are responsible for gene silencing or activation. Such modifications form a second level of our genomes which are considered as one of the main epigenetic modifications. DNA methylation resulting in the modification of a cytosine to 5-methylcytosine (5mC) is one of those epigenetic modifications that is responsible for gene silencing or activation. DNA methylation is effective on cellular differentiation, X-chromosome inactivation, gene imprinting, physiological conditions, and repression of retrotransposons. Cytosine methylation is highly prone to mutation where occurrence of spontaneous deamination of 5-methylcytosine to thymine leads approximately 20% decrease in the methylation of CpG islands.

DNA methylation is important for maintaining cell types and it is carried out by DNA methyltransferases (DNMT), e.g., DNMT1, DNMT3a and DNMT3b. DNA is wrapped around histone octamers and forms structures known as nucleosomes. DNMT3a and DNMT3b are responsible for *de novo* methylation allowing embryonic cells to differentiate into a cell type such as skin cells. DNA methylation profiles of corresponding cell types are conserved and inherited during cell division. DNMT1 is responsible for the maintenance of DNA methylation following differentiation and it is active during cell division. Thereafter, the methylation patterns of each cell type are different, and this reflects the gene expression pattern of the different cell types. Methylated cytosines are found primarily in CpG-poor regions at repetitive elements. On the other hand, the expected CpG content is found in CpG-dense regions (CpG islands) where cytosines are non-methylated state. Cytosine guanine sites (CPG islands) are mostly methylated in adult cells except that the promoter CpG-rich regions are typically nonmethylated. Promoter regions contain regulatory elements that control transcription of genes. DNA methylation blocks the interaction of some transcriptional activator or enables binding of repressive factors to prevent transcription. DNMT3A and DNMT3B are responsible for DNA methylation in early development. DNMT obtains the methyl group from a molecule called SAM. Then, methyl group is added to cytosine to form 5-methyl cytosine. Another important enzyme is TET that has an important role in regulating DNA methylation patterns which adds a hydroxyl group initially to 5-methylcytosine to form 5-hydroxymethyl cytosine (5-hmC) or form cytosine back from 5-hydroxymethyl cytosine through several pathways. Therefore, TET enzyme is considered as responsible for DNA demethylation. In a normal cell, the two counter processes methylation and demethylation are strictly regulated during cell development. In cancer cells, the methylation pattern of DNA is altered, where the balance between the methylation and demethylation processes and balance between the TET and DNMT proteins are disturbed. Normally, TET proteins have higher concentrations where transcription starts in CpG island promoter regions. However, in tumors TET mutations cause loss of function, low 5-hmC levels and abnormal

DNA methylation levels. Additionally, decreased levels of DNMT leads to genome-wide hypomethylation. Consequently, DNA methylation patterns change because of the mutations of DNMTD and TET proteins. Typically, in cancer cells, promoter CpG regions are hypermethylated which is associated with tumor suppressor gene inactivation. In contrast to the focal regions of hyper methylation, DNA in cancer cells also undergoes widespread hypo methylation across the entire genome. This deregulation of epigenetic landscape is found in every type of human tumors (Gujar et al., 2019).

1.6 Drug-DNA Binding Types

Drug molecules can be designed to bind to the transcription regulation regions. The direct interaction between ligand and DNA can be classified into two main groups that are covalent and non-covalent interactions (B.R., 1958). Covalent interactions occur, when ligands are attached to DNA by covalent bonds and irreversibly halt DNA action by causing permanent damage (Sechi et al., 2009). Alkylation is one of the covalent interactions, where alkylating agents can attach the methyl or other alkyl groups they have in their initial structure to DNA with three different mechanism that are cross-bridging, base alkylation and mispairing (Silvestri & Brodbelt, 2013). Cross-bridging is a type of covalent interaction, where bifunctional alkylating agents of cross-link DNA by forming inter-strand cross linking or intra-strand bridge as shown in figure 3. Since cross-linking changes the normal DNA structure, DNA cannot be used anymore as a template for replication, transcription or translation. As a result, cell goes to apoptosis (Mishra et al., 2017; Silvestri & Brodbelt, 2013).

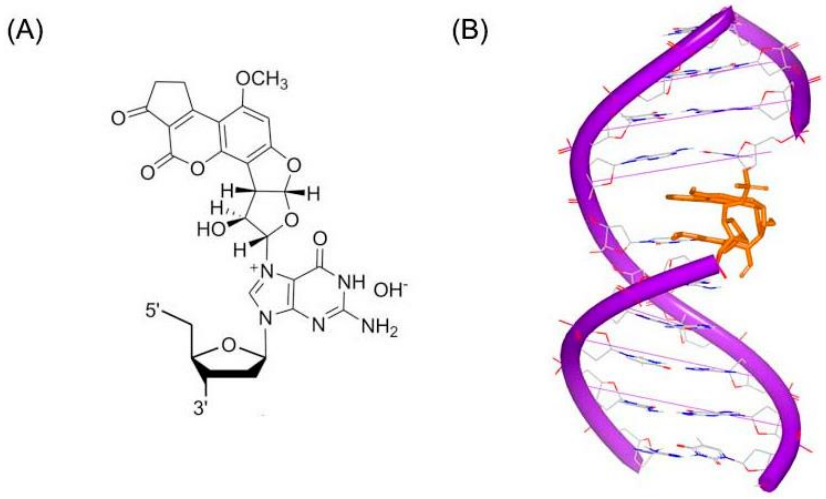


Figure 3. (A) Structure of Aflotoxin B1. (B) Crosslinked complex of DNA backbone which is colored purple and Aflotoxin B1 which is colored orange (M. Wang et al., 2016)

Base alkylation is also classified as covalent interaction since the pairing of the bases are impaired by alkylation of mainly the N-7 position of guanine and N-3 of adenine. DNA becomes miscoded and hence it is started to being fragmented to be repaired to replace the alkylated bases by enzymes (Mishra et al., 2017). Mispairing is another covalent interaction type, where this type alkylating agents cause mutations due to mispairing (Bauer & Povirk, 1997).

1.6.1 Non-Covalent Interaction

Non-covalent binders interact with the DNA molecule temporarily by weak interactions and therefore do not damage the DNA molecule even though they cause temporary deformation in DNA conformation and backbone torsions (Sirajuddin et al., 2013).

1.6.1.1 Intercalation

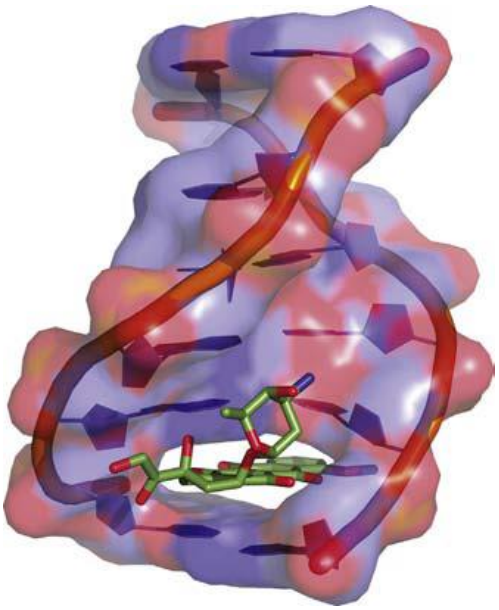


Figure 4. Doxorubicin-DNA intercalation complex (Yang et al., 2013).

Polycyclic planar aromatic ring structures can stack themselves between subsequent base-pairs of DNA by compensating the energy cost arising from the deformation of the DNA for unwinding and elongation of DNA by established π-π stacking interactions between the sandwiched base pairs and dipole-dipole interactions (fig. 4). Steps are opening of the DNA sequence where intercalation will occur, insertion horizontally to the created space and establishing interaction modes with the intercalation environment (Chaires, 2006; Lei et al., 2012). However, average to poor selectivity toward a specific binding site can be shown as a disadvantage of intercalator type binders (Ren & Chaires, 1999). It is difficult to select specifically the pathogenic genes or transcription starter genes since the deformation energy of interaction is mostly compensated by hydrophobic interactions of aromatic rings (Chaires, 1998). Due to low to average selectivity of intercalator drugs, intercalator type ligands would be intercalated at many sites in the very long genome, it would

block whole DNA instead of only targeted sites and consequently, they damage the whole DNA which may be identified as toxic (Sheng et al., 2013).

1.6.1.2 Minor Groove Binding

Minor groove is shallower and narrower than the major groove. Therefore, it is a good binding pocket for smaller molecules. In comparison with the intercalation, minor groove binding does not require large deformation energy cost (Neidle, 2001). Therefore, the binding site is determined by specific hydrogen bonds and van der Waals interactions between the ligand and minor groove sequence. These specific non-covalent interactions allow the binding of ligand to specific sites and therefore are more advantageous with respect to the non-selective intercalation.

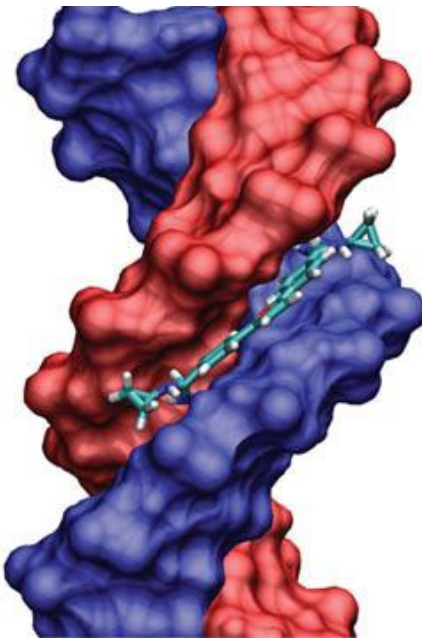


Figure 5. Space filling model of a minor groove binder (a cyclopropyl substituted diamidine) bound to DNA shown with tube model. (Nanjunda & Wilson, 2012).

Minor groove binders have common typical characteristics that are: high flexibility, torsional freedom of the sigma bonds among the curvature of the ring system and

possession of neutral or positive charge to easily interact with the negatively charged phosphate backbone. The flexibility enables the ligand to interact with the minor groove without any unfavorable free energy barrier and deformation of the minor groove (Chaires, 1998) as shown in figure 5.

1.6.1.3 Major Groove Binding

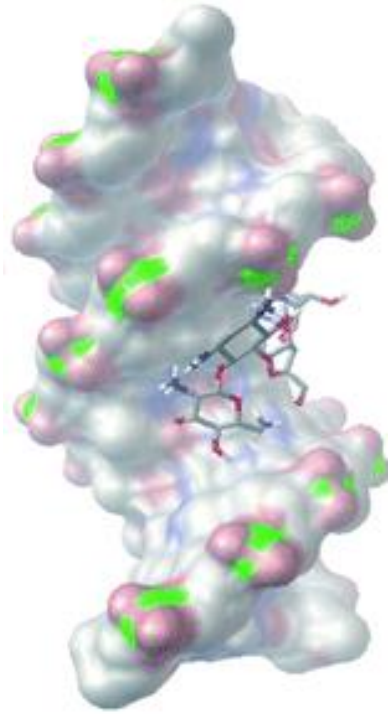


Figure 6. Major groove binding ligand complexed with DNA (Hamilton & Arya, 2012).

Most of the proteins that are playing a role in DNA processing are quite large and their size is compatible with interactions in the major groove of DNA (fig. 6). In addition to this size compatibility, those proteins bind according to molecular interactions sequence specifically and read the corresponding information. Even though the number of known major groove binders is rather small, it has been found

that bulky and large carbohydrate compounds prefer to bind to major grooves of DNA (Hamilton & Arya, 2012).

1.7 Sequence Specificity Based on Electrostatic Potentials and Functional Groups of Base Pairs

Sequence selective groove binders with high affinity to DNA and so with high therapeutic potential can be designed to prevent cancer and bacterial infections based on sequence specific electrostatic interactions. The working principle behind this can be explained by the modulation of the gene expression through binding of the ligand to specific genes sequence selectively. To understand those interactions, a closer look to the topology of DNA, helix conformation, electrostatic potentials and hydrogen bonding groups of bases directing to major and minor grooves is insightful.

1.7.1 Electrostatic Potential Patterns

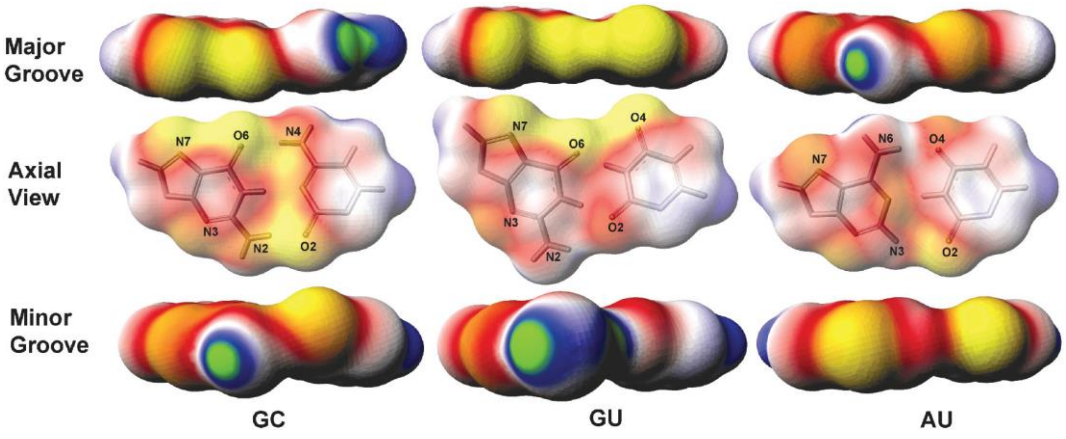
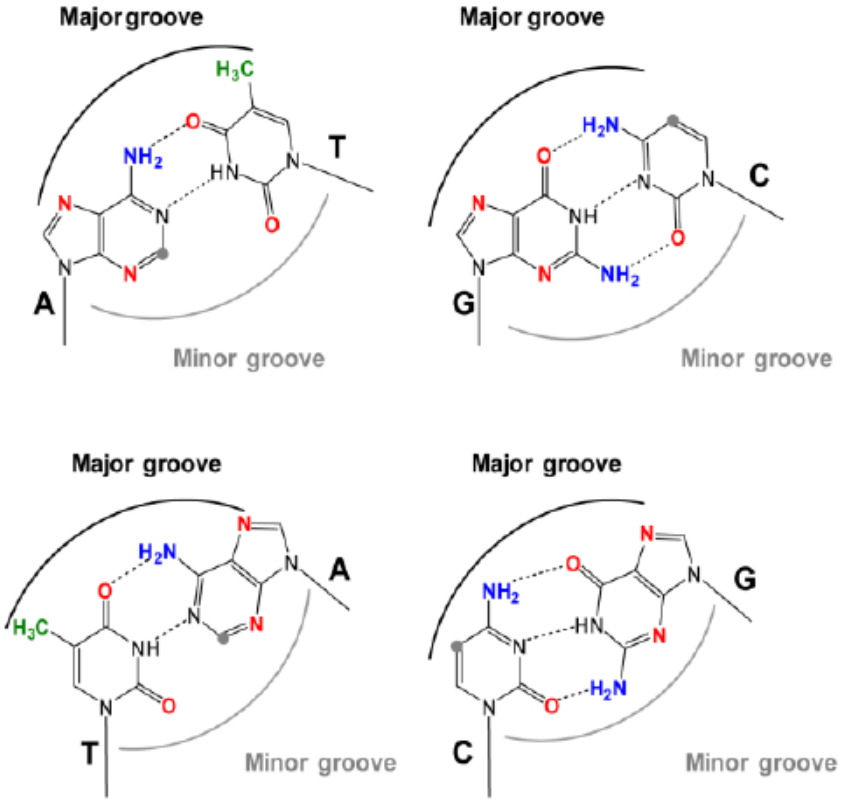


Figure 7. Surface electrostatic potentials of isolated G.C, G.U and A.U base pairs. The color scheme is as follows: yellow is the most negative, red is negative, white is neutral, blue is positive, and green is the most positive electrostatic potential, respectively (Xu et al., 2007).

In figure 7, it can be seen that in the major groove, the G.C base pair is significantly more electronegative than the A.U base pair. On the other hand, the A.U base pair has higher electronegativity in minor grove than the G.C base pair. The total electronegativity of sequences along major and minor groove can be adjusted, and different electrostatic patterns can form depending on the sequence. For example, GGGGGGGGGG has the most negative electrostatic value whereas AAAAAAAAAA has the most positive electrostatic value in the major groove. On the other hand, ds(CACACACACA) has an overall electrostatic potential value close to neutral since the neighbor alternating positive and negative potentials neutralize each other. By using this information, we can make a rough interpretation about the electrostatic potentials of AAAAAAAAAA, GGGGGGGGGG, ATCGCGCGAT, CGTATATACG and GGCCAATTGG of minor and major grooves. The electronegativity of major grooves of the studied sequences can be ordered as follows: GGGGGGGGGG (most electronegative major groove) > ATCGCGCGAT > GGCCAATTGG > CGTATATACG > AAAAAAAAAA (least electronegative major groove). Similarly, the electronegativity of minor grooves of studied sequences can be ordered as: AAAAAAAAAA (most electronegative minor groove) > CGTATATACG > GGCCAATTGG > ATCGCGCGAT > GGGGGGGGGG. Beside the total electronegativity on grooves, ordering and location of base pairs change the electroactivity and electrostatic wall pattern. Herein, electronegativity is a global descriptor. For example, even though the CACACACACA is neutral overall, it has a very specific electrostatic pattern, where positive and negative potentials are present locally.

1.7.2 Functional Group Patterns



H-bond acceptor	
H-bond donor	
Methyl group	
Ring Hydrogen	

Figure 8. Functional groups on the edges of cytosine (C)-guanine(G) and adenine (A)-thymine(T) base pairs directing both minor and major grooves. The colors assigned for the given functional groups are shown in the table (Harteis & Schneider, 2014).

In figure 8, hydrogen bond donor and acceptor groups, methyl groups and a ring hydrogen are shown. In this study, AAAAAAAAAA, GGGGGGGGGG, ATCGCGCGAT, CGTATATACG and GGCCAATTGG sequences were used as

receptors. Docking studies were performed along both major and minor grooves of those sequences for all ligands. Functional group pattern representations along each groove of each sequence were shown in this section to emphasize the receptor characteristics. This information is used to create functional group patters of major and minor grooves of studied sequences in figure 10 and 11.

1.7.3 Minor Groove

In current literature, it has been stated that minor groove binder drugs show high affinity to bind A-T rich sequences which are narrow and deep. Since minor groove of A-T rich sequences is narrower whereas G-C rich sequences are wider, van der Waals interactions and contacts between the binding pocket of A-T rich minor groove and ligand increase (Nanjunda & Wilson, 2012).

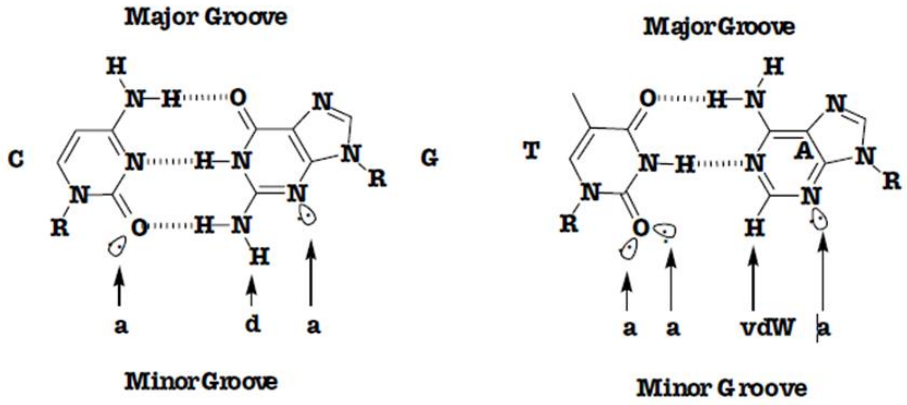


Figure 9. Hydrogen bond donor and acceptor groups on the edges of cytosine (C)-guanine(G) and adenine (A)-thymine(T) base pairs directing minor groove (Melander et al., 2019).

Even though the binding pattern information in the minor groove is less rich than that of the major groove, it can be understood from figure 8 that the direction of the exocyclic (2)-amino group provides discrimination of G-C from C-G base pair. Additionally, thymine O(2) and H(2) are placed asymmetrically which create extra discrimination pattern to increase the information in minor groove for better selectivity. Figure 9 shows the lone pairs of C-G and T-A base pairs in minor groove,

where the effect of direction of hydrogens and lone pairs enhances topological information in minor groove to discriminate G-C from C-G and T-A from A-T. Discrimination between A-T and T-A base pairs is that is O (2) of thymine has two lone pair (two hydrogen bond acceptor in very close distance) whereas the N(3) of adenine only has one (Melander et al., 2019). Therefore, binding patterns of minor grooves of studied sequences given in figure 10 might be a rough picture of bonding patterns. However, the binding patterns might be useful as qualitative descriptors to see the bonding pattern characteristics of the sequences used.

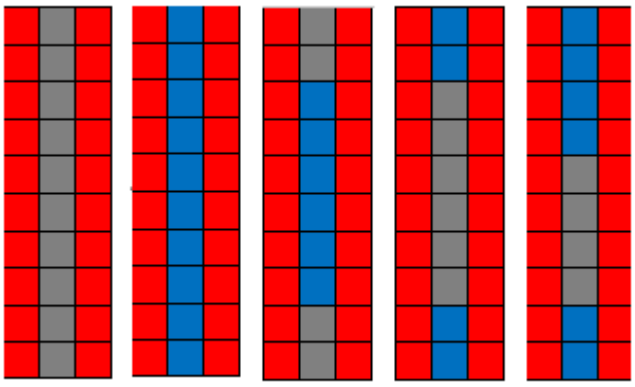


Figure 10. Functional group bonding patterns of minor groove of AAAAAAAAAA, GGGGGGGGGG, ATCGCGCGAT, CGTATATACG and GGCCAATTGG sequences respectively. Colors are assigned based on figure 8.

1.7.4 Major Groove

Previously, we have stated that larger molecules which cannot fit into minor groove, may bind to the major grooves. From figure 7, we noticed that the number of accessible functional groups in the major groove is greater than the minor groove. Moreover, bidentate hydrogen bonds also increase the information for selectivity. Therefore, the encoded pattern is greater and the information in major groove is richer (Harteis & Schneider, 2014). Figure 11 shows the interaction patterns of selected sequences of major grooves used within this thesis. AAAAAAAAAA, GGGGGGGGGG, ATCGCGCGAT, CGTATATACG and GGCCAATTGG

sequences were selected to be able to see the effect of various sequence and neighbors on binding interactions.

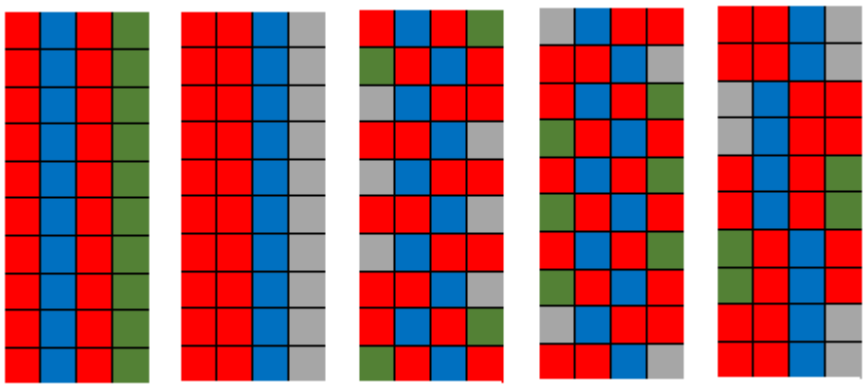


Figure 11. Functional group bonding patterns of major groove of AAAAAAAAAA, GGGGGGGGGG, ATCGCGCGAT, CGTATATACG and GGCCAATTGG sequences, respectively. Colors are assigned based on figure 8.

1.8 Aim

Discovery of a drug starts with the definition of the target and binding free energy. Then, drug is subjected to pre-clinical *in vitro* and *in vivo* assays for optimization until it reaches clinical trials and, ultimately, the market. Computational methods in drug development shorten time and cost of drug development. Computer-Aided Drug Design (CADD) is the name of the field defined for biomolecular modelling of drug-target binding which can be divided into two subgroups that are ligand-based drug discovery (LBDD) and structure-based drug discovery (SBDD). When identification of the target structure is hard or target structure is not known experimentally, ligand-based drug discovery methods are used. Some LBDD approaches are molecular similarity approaches, quantitative structure-activity relationship (QSAR) and pharmacophore modelling. As a disadvantage, ligand-based approaches are either based on similarity between structural features of known drugs and developed drug, or relation between the known drug and its corresponding biological activity on the target site. When the 3D structure of the target is known

and solved either by NMR or X-ray crystallography, structure-based drug development methods such as structure-based virtual screening and direct docking can be employed. Large molecular structure datasets can be scanned to estimate the best binding probability and conformation of each molecule candidate via a scoring function to bind to the target by docking in SBDD approach. Different binding pockets (binding sites) might be available in the target (receptor). In this case, geometry-based binding site identification algorithms can be utilized to predict possible locations that ligand can bind. Moreover, new drugs can be developed by *de novo* ligand design by adding/changing the functional groups on the main structure without scanning large molecule libraries. Affinity of the possible drug molecules are correlated with the experimental results. As a result of this correlation, drug molecules are optimized, examined and eliminated until best drug molecules are obtained.

However, there are some serious problems in structure based virtual screening processes. One of them is, most of the time flexibility of the receptor (target) is ignored (Aminpour et al., 2019). To overcome this issue, we generated multiple receptor conformations by extracting 10 snapshots from a molecular dynamics simulation of each to account receptor flexibility. Additionally, Autodock Vina (docking program used in this research) results do not always lead reliable binding scores. Therefore, we merged additional MM/GBSA (Molecular Mechanics/ Generalized Born Surface Area) methods after docking to calculate binding free energies. The ligand and receptor complex conformation found by Autodock Vina is usually not the most ideal complex structure from a molecular mechanics point of view. This is mostly due to the fact that hydrogen atoms are ignored during the docking process and placed arbitrarily at the end (Oleg, T., Olson, 2010). Torsions should also be adjusted to match the MM potential. Therefore, we minimized all complex structures after docking to relax the torsions. Another limitation of docking is incomplete sampling. To draw a conclusion from a docking study, many docking confirmations and results are required. Therefore, in this study we performed multiple docking runs, where we accounted for all docking poses of each docking

site of each ligand conformation. We extracted binding probability distribution plots of studied ligand-binding pocket interactions versus corresponding binding free energy values of investigated ligand-receptor pairs.

Computational binding free energy prediction would be a very cost and time efficient tool for drug design and discovery. A quick and trustable binding free energy approach would yield results faster than expensive experimental studies to test possible interaction modes of novel drugs with the target site. Herein, we created a fairly time-efficient computational binding free energy method, which is automatization of combined docking, molecular mechanics (MM), molecular dynamics (MD) by bash/python programming. We selected set of the ligands whose binding site preferences were already known and DNA sequences as receptor molecules. We compared the obtained results with the literature data to check the correlation and we identified best binding sites for each drug-DNA pair. After this point, selectivity of any possible drug candidate toward any sequence can be predicted and these automatization codes can be used for drug discovery.

CHAPTER 2

METHODOLOGY

In this study, sequence specificity of a series of major and minor groove binder molecules toward 5 different DNA sequences was investigated via computational chemistry methods involving docking, molecular mechanics (MM), molecular dynamics (MD) and bash/python programming for automatization.

2.1. Molecular Docking

Docking can be defined as a computational study that tries to predict possible binding modes and conformations of a small molecule to a cavity of macromolecule. Relative binding free energies are calculated by using scoring functions. Docking has a significant importance in drug discovery and design. In this work, Autodock Vina was used as a docking program. It searches the binding mode of the ligand in given central coordinates and dimension sizes of the grid box starting with a random orientation without visual interface. Therefore, it enables automatization with its high speed. Also, Autodock Vina enables to sample various bound ligand conformations. However, it makes a lot of approximations to work fast and this leads to a decrease in the accuracy and requires improvement (Cosconati et al., 2011). Autodock Vina scoring function works based on van der Waals-like potential including non-directional hydrogen bond term, a hydrophobic term and a conformational entropy penalty where electrostatics and solvation are not taken into consideration (Eberhardt et al., 2021).

The next step after input structures required for docking (pdbqt files for both ligands (drugs) and receptors (DNAs)) was to decide the grid box sizes that encapsulates

each of the drugs. Smallest cubic box sizes were selected as big as to contain all atoms of the ligands. Another step for docking was to decide the docking sites where the grid box containing the ligands are placed. Since the purpose of this study is to scan binding energy along whole major and minor grooves of selected DNA molecules sequentially, a specific strategy was applied to place those grid boxes to the center of masses of phosphate atoms (Pi) of opposite strands. Basically, center of mass of each Pi of 5' strand and Pi+3 of 3' strand was calculated for minor groove. For major groove, center of mass of each Pi of 3' strand and Pi+4 of 5' strand was calculated. Mainly, 8 binding sites are presented for major groove and 9 binding sites are presented for minor groove of 10 base pair long DNA. 5 DNA sequences were selected as receptors which are AAAAAAAAAA, GGGGGGGGGG, ATCGCGCGAT, CGTATATACG and GGCCAATTGG with same length. For each DNA, 10 different conformations extracted from different nanoseconds of production were used in the docking studies to get unbiased docking results. Autodock Vina has been set to generate 9 docking poses for each docking. For one DNA conformation, there are in total 17 docking centers as sum of docking centers in minor and major grooves. Since 10 conformations of each sequence were used, 170 docking studies were performed to scan one sequence and one drug interaction. Since we have scanned 5 different DNA sequences, we needed to perform 850 docking studies. We investigated interactions of 10 ligands, therefore we needed to perform 8500 docking studies and MM-refinement of those dockings to calculate interaction energies of each drug to find sequence specificity. Unfortunately, calculation of center of mass for each docking study was impossible by hand. Therefore, we wrote an automatized python script to create Autodock Vina inputs. Then box size was set according to the ligand size mentioned in the code as 26 Å for altromycin B, 21 Å for pluramycin A, 21 Å for hedamycin, 21 Å for nogalamycin, 18 Å for berenil, 17 Å for DAPI, 21 Å for distamycin, 23 Å for hoechst33258, 20 Å for netropsin and 23 Å for pentamidine, respectively. Auto Vina then calculated the center of mass of each docking pose with respect to corresponding conformation of DNA. All codes of automatization are available upon reasonable request.

2.1 Molecular Mechanics (MM)

With molecular mechanics, reliable binding free energies can be obtained. That requires parametrization based on quantum mechanics (QM). Descriptors such as atomic charges, electronic structure, dipole moments, orbital energies, molecular electrostatic potentials, conformational and thermodynamic parameters of a ligand molecule can be obtained by quantum chemical calculations. Those descriptors are critical for ligand receptor interactions since non-bonded interactions are directly calculated by using derived atomic charges (Sizochenko et al., 2017). Electronic structure of a polyelectronic molecular or atomic system could be obtained if time-dependent Schrödinger equation could be solved analytically. Therefore, approximations are essential to solve Schrödinger equation. Hartree-Fock (HF) is one of those approximations which ignores correlation of electrons which results wave function of a system turns out product of independent one particle wave functions. In HF theory, each electron experiences the Coulombic repulsion with averaged electrostatic field of rest of the electrons instead of experiencing each Coulombic repulsion of one-to-one electron interaction. Because of this reason, the Schrödinger equation of HF approximation contains the kinetic energy of each individual electron, each electron-nuclei interaction and sum of the Coulombic interaction of each electron with the mean electrostatic field created by rest of the electrons. HF approximation is relatively cheap and used for ligand ground state geometry optimization in this work (Veryazov, 2016).

In molecular mechanics, bonds are treated as classical springs with a force constant and atoms are treated as the masses on the opposite edges of this spring. Energy is calculated by simple harmonic potential and similar approximations are used for bond angles and dihedral angles. Parameters used in the force field are derived from quantum mechanical calculations such as partial charges of atoms or experiments such as van der Waals radii, atomic mass, bond lengths, angles, force constants (Marion, 2014). In this research, AMBER (Assisted Model Building with Energy Refinement) force field has been used for MM-refinement as shown below, in which

the first three terms stand for bonding interactions and last term stands for non-bonding interactions that are van der Waals and Coulomb interactions:

$$E = \sum_{bonds} K_b(b - b_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\emptyset - \gamma)] + \sum_i \sum_{j>i}[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}}]$$

(1)

where K_b is a force constant of covalent bonds, b_0 is the center of the harmonic potential of covalent bonds, K_θ is the force constant for angles, θ_0 is the center of the harmonic potential of angles, V_n is the force constant of dihedral angle $\emptyset$ and phase angle γ, ϵ is the dielectric constant, q_i is the atomic charge of the i^{th} atom (Cornell et al., 1995; Marion, 2014). In this study, energy of each complex structure produced by docking was minimized with AMBER force field by using bsc1, ff14SB and GAFF parameters.

2.1.1 Binding Free Energy (ΔG_{bind}) and Solvation Free Energy (ΔG_{solv})

Binding free energy is a measure which describes the binding strength of potential drug candidate to the target, in this case to a macromolecular biomolecule (Genheden & Ryde, 2015). To create a medium similar to physiological conditions, water has been used as the solvent in calculations. Hydrophilic parts of the solvated molecules in water are able to form hydrogen bonds with the solvent molecule and so might be stabilized by these hydrogen bonds. In the process of forming the complex, there is a cost for separating polar groups from water molecules where the drug molecule is transferred from water to the binding pocket. On the other hand, hydrophobic parts of the solvated molecules would interact easily with the hydrophobic parts of the receptor in the process of formation of complex structure. If the new electrostatic interactions in the formed complex overcome the loss by matching of hydrophobic residues in the binding pocket, binding affinity increases. Therefore, solvation free energy is a key point for the calculation of binding energy (Decherchi, S., Masetti,

M., Vyalov, I., Rocchia, 2015). Binding free energy is calculated as in the following equation:

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

(2)

where ΔH is the enthalpy change, T is the absolute temperature (K), and ΔS is the entropy change of the system (Forouzesh, N., Mishra, 2021). In this work, implicit solvent model was used to represent water, which has dielectric properties of water due to low computational cost. However, in this research entropy contribution is neglected.

2.1.2 Molecular Mechanics Generalized Born Surface Area (MM/GBSA) Method

In MM/GBSA method to calculate binding free energies, classical force fields are employed with implicit solvation models. Electrostatic contribution was obtained by Generalized Born (GB) model. For the binding of ligand (L) to receptor (R) to form complex (LR) in a solution where L+R↔LR:

$$\Delta G_{bind} = G_{LR} - G_L - G_R$$

(3)

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{sol} - T\Delta S$$

(4)

$$\Delta E_{MM} = \Delta E_{int} + \Delta E_{ele} + \Delta E_{vdW}$$

(5)

$$\Delta G_{bind} = \Delta G_{GB} + \Delta G_{SA}$$

(6)

$$\Delta G_{SA} = \gamma . SASA + b$$

(7)

where ΔG_{sol} is the solvation free energy that is the sum of electrostatic solvation energy of polar contribution (ΔG_{GB}) and nonpolar contribution (ΔG_{SA}) calculated by solvent-accessible surface area (SASA) between the solute and continuum solvent, $-T\Delta S$ is the conformational entropy upon ligand binding mostly neglected due to high computational cost, ΔE_{MM} is the change in the gas phase molecular mechanics (MM) energy containing the internal energy changes (ΔE_{int}) that are bond, angle and dihedral energies, electrostatic energies (ΔE_{ele}), and the van der Waals energies (ΔE_{vdW}) (E. Wang et al., 2019).

In this work, solvation free energy was obtained by Generalized Born (GB) model as shown in equation 8 below:

$$\Delta G_{GB} = -\frac{1}{2}\left(\frac{1}{\varepsilon_{in}} - \frac{1}{\varepsilon_{out}}\right)\sum_{i,j}\frac{q_i.q_j}{f_{ij}^{GB}(r_{ij})} \quad (8)$$

$$f_{ij}^{GB} = \left[r_{ij}^2 + R_{ij}R_{ij}\exp\left(\frac{-r_{ij}^2}{4R_iR_j}\right)\right]^{1/2} \quad (9)$$

Where ε is the dielectric constant of bulk solvent (dielectric constant of water for our study), q_i is the partial atomic charge of atom i, r_{ij} is the inter-particle distance and R_i is the effective Born radius of atom *I* (Nguyen et al., 2013).

2.2 Minimization

The docking poses created by Autodock Vina are not optimized where the scoring function is based on distances between atoms only instead of actual energy calculation(Oleg, T., Olson, A. J., 2010). But in MM, it is crucial to optimize the structure prior to evaluating the binding energy because all pairwise interactions are associated with a potential. Also, this is a technicality, but it is important to mention it: since Vina does not use the hydrogen atoms explicitly, it orients them randomly in groups such as -OH. This can cause issues in MM if not optimized. Complex

structures that are docked ligand to the binding pockets and receptor should be refined, and their energy values should be minimized with molecular mechanics to be able to be more realistic where the complex is stable without any clashes. Atomic charges, electronic structure, dipole moments, orbital energies, molecular electrostatic potentials, conformational and thermodynamic parameters of a ligand molecule can be used in energy minimization to have better atomic descriptions and parameters to calculate the interaction energies more realistically.

2.3 Molecular Dynamics (MD)

Molecular dynamics simulation study has significant importance on drug discovery by helping to understand the dynamical behaviors of biomolecules of interest like in ligand binding, conformational change, motion over a simulation duration. Working principle of MD simulation is that Newton's law of motion is used to predict the spatial position of each atom as a function of time by calculating the forces exerted on each atom and then, calculating and updating the new positions and velocities of atoms in every timestep to create a trajectory. Resultant trajectory is a movie that is formed by a visualization of three-dimensional motions and velocities over the time steps (Hollingsworth & Dror, 2018). In this study, MD was used to equilibrate the DNA sequences and we obtained different conformations of these DNA sequences from MD. The reason of this is to consider receptor flexibility in docking to prevent any bias or possible errors. Additionally, MD would be useful even after molecular mechanics study since it enables us to see whether the formed complex structures are stable over time in periodic conditions or not.

2.4 Pipeline

Mainly, 8 binding sites were assigned for major groove and 9 binding sites were assigned for minor groove for 10 base pair long DNA structures. 5 DNA sequences were selected as receptor which are AAAAAAAAAA, GGGGGGGGGG,

ATCGCGCGAT, CGTATATACG and GGCCAATTGG with the same length from 5' to 3' direction. For each DNA structure, 10 different conformations extracted from different nanoseconds of production were used in the docking studies to get unbiased docking results. Autodock Vina was set to generate 9 docking poses for each docking. For one DNA conformation, there are totally 17 docking centers as sum of docking centers in minor and major grooves. For each docking center, Autodock Vina produced 9 docked poses. Therefore, for only one conformation of DNA sequence, there were 153 docking and MM-refinement studies to be performed. Since 10 conformations of each sequence were used, 1530 docking and MM-refinement studies were performed to scan one sequence and one drug interaction. Since we scanned 5 different DNA sequences, we needed to perform 7650 docking and MM-refinement studies to see sequence selectivity of only one drug. We investigated interactions of 10 ligands, therefore we performed 76500 docking and MM-refinement of those dockings to calculate interaction energies of each drug to find sequence specificity. Such process should be automatized because, it is not possible to calculate one by one. All steps of this research were performed by the combination of python codes and bash/shell scripting.

2.4.1 Ligand Preparation

Selected major and minor groove binder drugs were optimized with Hartree-Fock method (HF) by using 6-31G* basis set by using Orca 4.2.1 in their protonation states at 7.4 pH, and the corresponding electrostatic potentials (ESP) were calculated. Atomic point charges were calculated by restrained electrostatic potential atomic charges (RESP) by combination of Merz-Kollman grid points resulted from the orientation of molecules. General Amber Force Field (GAFF) was used to set ligand parameters (mol2 files) by antechamber program of AMBER. To check the missing bond parameters and produce frcmod files containing all parameters required, parmchck2 of Ambertools19 was used. Parameter and coordinates files were prepared by tleap program of AMBER. After that, short minimization was performed

to relax the molecule with 1000 cycles with 999999.99 Angstroms cutoff. Resultant rst files were converted to pdb format by ambpdb tool of AMBER for further use. Then, pdbqt files required for docking were generated from pdb files by running the prepare_ligand4.py code obtained from MGLTools/Utilizes/24. Autodock Vina was used as docking tool to perform all docking studies.

2.4.2 DNA Preparation

5 different B-DNA sequences containing 10 base pairs were selected which are AAAAAAAAAA, GGGGGGGGGG, ATCGCGCGAT, CGTATATACG and GGCCAATTGG Those sequences were prepared with AMBER NAB (nucleic acid builder) to create initial DNA structures in the pdb format. By using tleap facility of AMBER and pdb files obtained from AMBER NAB tool, parameter (prmtop) and coordinate (inpcrd) files of each DNA sequences were prepared and ff14SB and bsc1 force fields were used for this process. To relax the system prior to MD (molecular dynamics), two step minimization algorithm was applied to all DNA sequences by AMBER. In the first step of minimization, constant volume 10000 steps of gas phase minimization without pressure control were performed with 12 Å non-bonded cut-off distance. Atoms in DNA residues were restrained with 3.0 kcal/mol force constant. In the second step of minimization, same periodic boundary conditions, method and same number of cycles of minimizations were applied but restraint was removed. Then, all DNA systems were heated up gradually in 7 steps and allowed to equilibrate to prevent any possible blow up and trap in any local minima. In the first step of heat-up, all DNA systems were heated up from 0 K to 10 K by constant volume (NVT) ensemble. 500 number of MD steps were used for MD runs of 10 picosecons (ps) with 0.002 picosecons (ps) timestep. The seed for the pseudo-random number generator to assign initial velocities was set as to get velocities differently in each start with reproducibility (ig = -1) and Langevin Thermostat was used (ntt=3). Cut-off distance was set to 12 Å. Coordinates were read from the minimization outputs (rst files). Shake algorithm was set only to restrain the bonds

involving the hydrogens (ntc = 2, ntf = 2). In the second step, DNA molecules were heated from 10 K to 50 K with same periodic boundary conditions. In the third step, restrains were removed and systems were heated from 50 K to 100 K with same conditions. In the fourth step of heatup, ensemble was switched from NVT to NPT ensemble at 100 K. In the fifth step, systems were heated from 100 K to 200 K with NPT ensemble. In the sixth step, systems were heated from 200 K to 300 K with NPT ensemble. In the seventh step, systems were equilibrated at 300 K with NPT ensemble.

3000 ns long production of 5 selected DNA sequences were performed by pmemd.cuda implementation of AMBER19. Isothermic-isobaric (NPT) ensemble was applied at 300 K with isotropic position scaling (ntb=2, ntp = 1), with 12 Å cutoff distance. Langevin thermostat was used. The seed for the pseudo-random number generator to assign initial velocities was set as to get velocities differently in each start with reproducibility (ig = -1) and Langevin Thermostat was used (ntt=3). Cut-off distance was set to 12 Å and timestep was 0.002 ps. Shake algorithm was set only to restrain the bonds involving the hydrogens (ntc=2, ntf=2).

By using cpptraj of AMBER, each rst file of each nanosecond which were generated by 3000 ns production were converted to pdb files. Then, pdbqt files required for docking were generated from pdb files by running the prepare_receptor4.py code obtained from MGLTools/Utilizes/24. 10 pdb file for each DNA sequence have been used for docking studies as receptor.

2.4.3 Automatization

Automatization consists of three major codes after drug and DNA molecules were optimized and parametrized. Inputs for Autodock Vina were created with the python script that basically calculates the center of mass of the distance of each phosphate (Pi) in 5’ direction starting from the 1^{st} base and $(Pi+4)^{th}$ phosphate in the opposite strand for major groove as shown in the figure 12 below and calculates the center of

mass of the distance of each phosphate (Pi) of 3' strand starting from 20th base and (Pi+3)th phosphate in the opposite strand for minor groove as shown in the figure 13 below. For each DNA, docking was performed 10 times at each of the docking centers shown in figure 12.

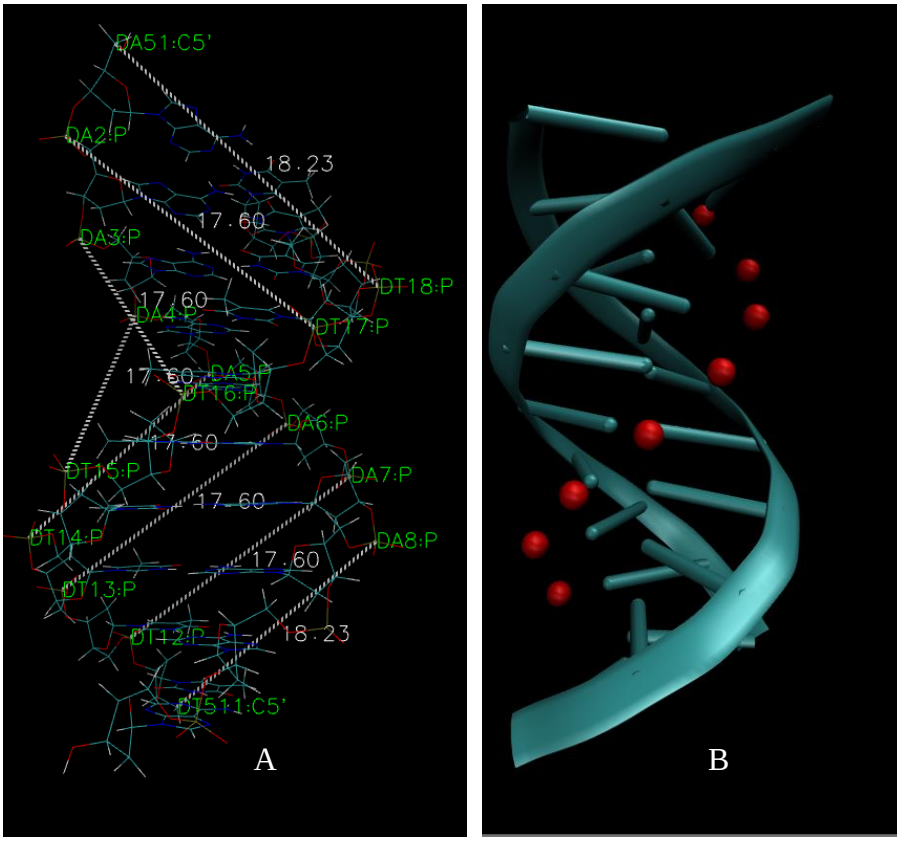


Figure 12. (A) Calculation of the center of mass of the distance of each phosphate (Pi) in 5' direction starting from the 1st base and (Pi+4)th phosphate in the opposite strand for major groove. (B) Major groove docking centers along AAAAAAAAAA represented with red dots.

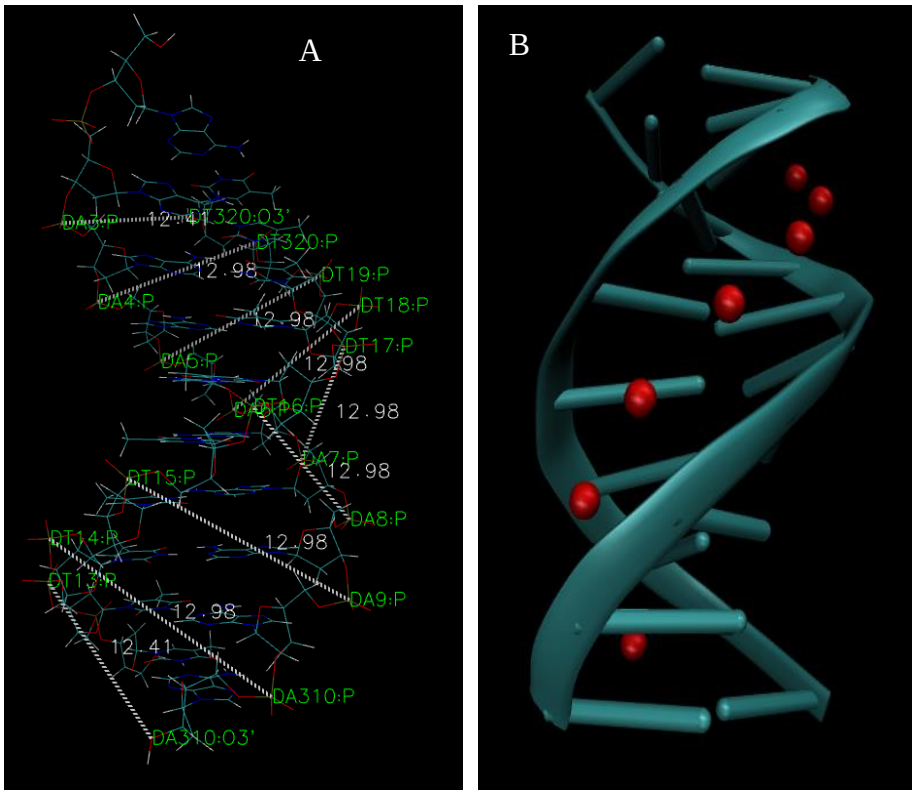


Figure 13. (A) Calculation of the center of mass of the distance of each phosphate (Pi) of 3' strand starting from 20th base and (Pi+3)th phosphate in the opposite strand for minor groove. (B) Minor groove docking centers along AAAAAAAAAA represented with red dots

Specific atom names and residue numbers that are going to be used for center of mass calculation is written in the code as regular expressions with respect to each DNA sequence. Each regular expression is searched in corresponding nanosecond of every DNA pdb files to find the x, y and z coordinates and calculate the center of mass to define box center coordinates to be written in each Autodock Vina input. Also, box size is written by python code to the created Autodock Vina input with respect to smallest cubic dimensions to contain each drug molecule separately that are 26 Å for altromycin B, 21 Å for pluramycin A, 21 Å for hedamycin, 21 Å for nogalamycin, 18 Å for berenil, 17 Å for DAPI, 21 Å for distamycin, 23 Å for hoechst33258, 20 Å for netropsin and 23 Å for pentamidine respectively. Number of binding modes was selected as 9 and exhaustiveness was set as 25 for all inputs.

Figure 14. First binding mode of altromycin B at each docking center of along AAAAAAAAAA/minor groove.

In figure 14, the first binding mode of altromycin B at each docking center along AAAAAAAAAA/minor groove is shown. Both major and minor grooves of all DNA for all drug molecules were scanned by this systematic docking method.

MM-refinement is MM-minimization and calculation of binding free energies. Docking and MM-refinement are automatized in this step together with bash scripting. There are 3 main scripts for major and minor grooves separately. With given regular expressions describing the docking site naming, all the created Autodock Vina inputs created in the previous step were used for docking. Formed pdbqt files contain all 9 docked poses together. Therefore, those binding poses of

drugs were turned into pdb files and then, splitted into sperate pdb files in the next step. Next, all vina output txt fileswere moved to another folder to use the Vina docking scores during the binding energy calculation step. After that, corresponding DNA and drug binding pose were combined to create parameter and coordinate files of the complex structure (docked drug pose + DNA). To prepare Amber files of complex structure with tleap program of AMBER19, implicit solvation model with mbondi3 was used. For DNA, bsc1 force field was employed. For ligands, GAFF force field was utilized. Next, complex structures were minimized in two steps. In first step, all hydrogen atoms were minimized while others were constrained with a 3.0 kcal/mol force constant for 1000 cycles without cutoff and igb=8 Generalized Born (GB) solvation model was utilized. In the second step, the minimization of the complex structure was performed for 5000 steps, with restraints set only on phosphorus atoms.

After minimization of complex structure, DNA structure and ligand structure were extracted by cpptraj utility of AMBER19 separately. Parameter and coordinate files for only drug, only DNA and complex together were created via cpptraj of AMBER. Then, all energy values were calculated for complex structure (ligand+receptor), ligand and receptor separately by using ig=8 GB solvation model and ntslim=-1 sander facility of AMBER19.

Then, corresponding Autodock Vina outputs were copied to every subfolder where energies were calculated. The purpose of this step is to include Vina scores in the output files for better comparison with MM results. All these processes were performed via bash scripts in created subdirectories automatically. For each DNA sequence, bash codes were written separately with respect to corresponding name of the DNA sequence, regular expressions and nanoseconds of the DNA sequences.

After MM-Refinement completed, binding free energy of each docked pose was automatically calculated by another python code and outputs were written in corresponding csv files. Basically, the code finds total energy of the complex, drug and DNA sequence of each docked conformation from the corresponding MM-

refinement outputs in corresponding subdirectory. The binding energy of each structure was calculated by subtracting the total energy of ligand and total energy of DNA from the total energy of complex structure, including solvation energy. Then, calculated binding free energy values for major and minor grooves were written to 2 csv files respectively for each drug and DNA interactions. The same calculation for the solvation energy was applied and and the results are written in next column. Last column contains corresponding Autodock Vina scores. Codes and obtained data are available upon a reasonable request.

2.5 Studied Drug Molecules

For this study, we selected a series of major and minor groove binder drugs that are experimentally confirmed in the literature. The selection was based on binding specificity toward either major or minor groove of the drugs and extent of the availability of the information in the literature for comparison with the results obtained in this study. In this section, we describe these molecules and review the binding characteristics with respect to literature information.

2.5.1 Major Groove Binders

We selected 4 major groove binder molecules that are pluramycin A, altromycin B, hedamycin and nogalamycin.

2.5.1.1 Pluramycin A ([4-(dimethylamino)-6-[8-[4-(dimethylamino)-5-hydroxy-6-methyloxan-2-yl]-11-hydroxy-5-methyl-2-[2-methyl-3-[(*E*)-prop-1-enyl]oxiran-2-yl]-4,7,12-trioxonaphtho[2,3-h]chromen-10-yl]-2,4-dimethyloxan-3-yl] acetate)

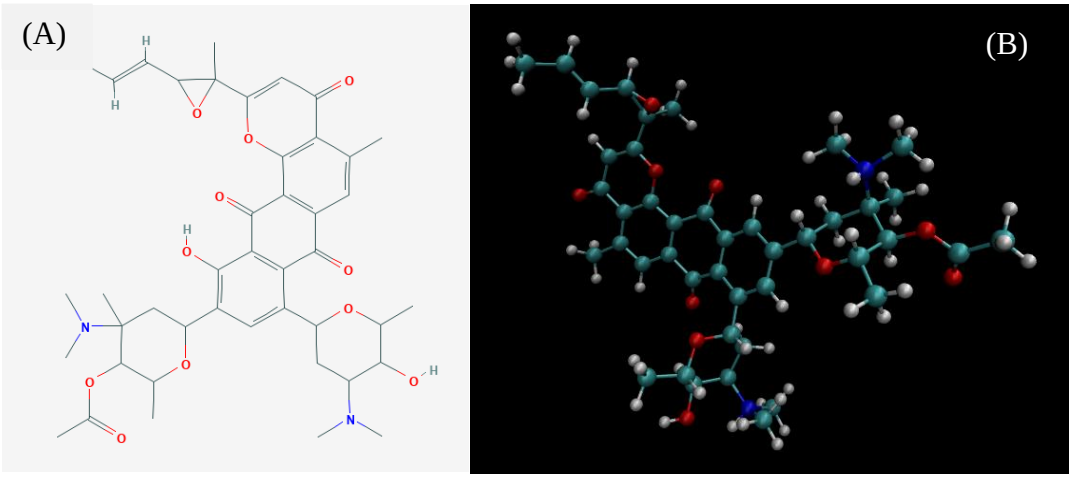


Figure 15. (A) Lewis structure of pluramycin A (PubChem CID: 5906990) (NIH, n.d.). (B) Three-dimensional (3D) conformation of pluramycin A. Color representations are as follows: cyan stands for carbon, grey stands for hydrogen, red stands for oxygen, blue stands for nitrogen.

It has been proposed that pluramycin A selectively binds to TATA sequence by intercalation. Its flanking carbohydrate groups are in major groove and epoxy group located towards minor groove when it intercalates. TATA sequence is called as TATA box where TATA-binding protein (TBP) binds to DNA from minor groove side in RNA transcription. When pluramycin A intercalates to TATA sequence as described, it alkylates N7s of guanines (Sun & Hurley, 1995). In one article, it has been stated pluramycin A shows higher selectivity towards 5'(C/T)G sequence (Hansen & Hurley, 1995). In another study, selectivity pluramycin A and hedamycin were ordered as 5' -CG* > 5'-TG* >> 5' -AG* = 5' -GG* (Okamato, 1998). On the other hand, pluramycins are grouped as major groove binders by Mishra et. al. (Mishra et al., 2017). Compared to small intercalating agents, opposite sided bulky flanking groups around the planar ring system might make the intercalation process

more difficult. The results of the present work in combination with intercalation and dynamics could help determining the mode of action of pluramycin A.

2.5.1.2 Altromycin B (Methyl2-(3,5-dihydroxy-4-methoxy-6-methyloxan-2-yl)-2-[10-[4-(dimethylamino)-5-(5-hydroxy-4-methoxy-6-methyloxan-2-yl)oxy-4,6-dimethyloxan-2-yl]-2-(2,3-dimethyloxiran-2-yl)-11-hydroxy-4,7,12-trioxonaphtho[2,3-h]chromen-5-yl]-2-hydroxyacetate)

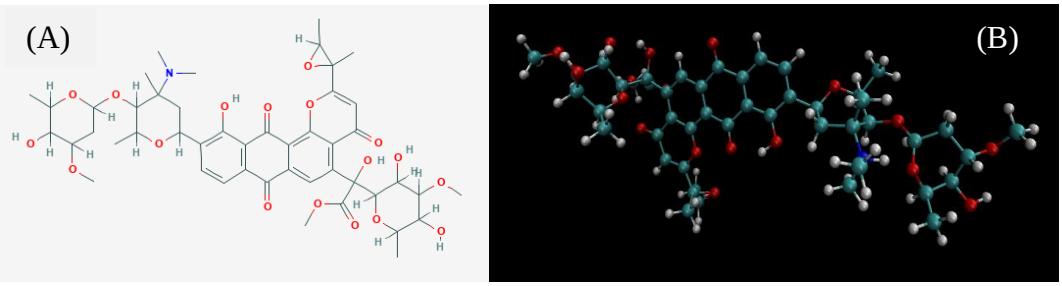


Figure 16. (A) Lewis structure of altromycin B (PubChem CID:148125) (NIH, n.d.). (B) Three-dimensional (3D) conformation of altromycin B. Color representations are as follows: cyan stands for carbon, grey stands for hydrogen, red stands for oxygen, blue stands for nitrogen.

Altromycin B is a DNA alkylating antitumor antibiotics that works in a manner that the epoxy group of altromycin B performs nucleophilic attack to N7 of guanine and prefers to bind to 5' -AG* (Hansen & Hurley, 1995; Okamato, 1998). It intercalates by stacking to the 5’ side of guanine by placing the disaccharide into minor groove and monosaccharide into the major groove. Glycosidic attachments at the C5, C8 and C10 chromophore positions can adjust the selectivity and reactivity in both major and minor grooves as much as attachments of electrophilic groups to the C2 position (Hansen & Hurley, 1995). In NMR studies, it has been detected that altromycin B interacts with major grooves (Hamilton & Arya, 2012). On the other hand, altromycin B is reported as a major groove binder by Mishra et. al.(Mishra et al., 2017). However, since there was no crystal structure reported in the literature, this study could help to understand how altromycin B interacts with DNA.

2.5.1.3 Hedamycin (10-[4-(dimethylamino)-5-hydroxy-4,6-dimethyloxan-2-yl]-8-[4-(dimethylamino)-5-hydroxy-6-methyloxan-2-yl]-11-hydroxy-5-methyl-2-[2-methyl-3-(3-methyloxiran-2-yl)oxiran-2-yl]naphtho[2,3-h]chromene-4,7,12-trione)

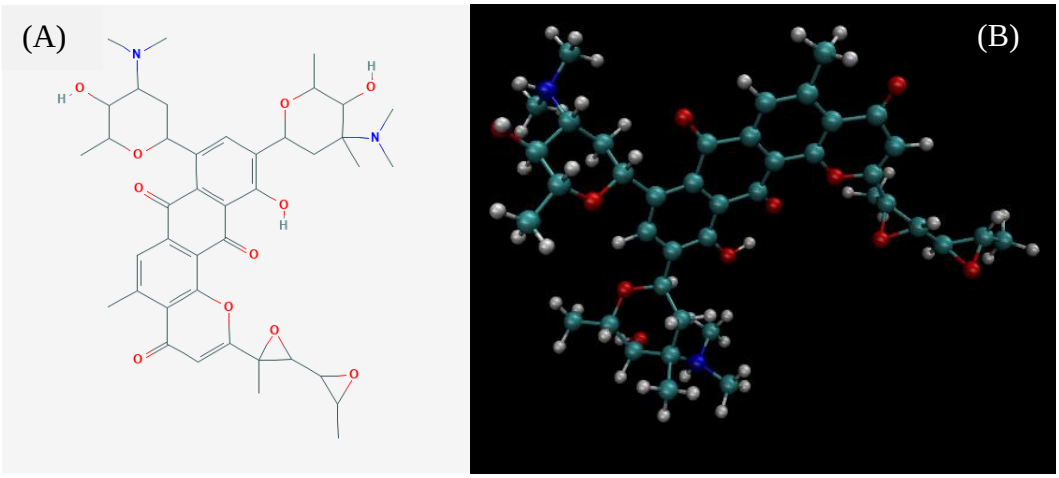


Figure 17. (A) Lewis structure of hedamycin (PubChem CID: 98033) (NIH, n.d.). (B) Three-dimensional (3D) conformation of hedamycin. Color representations are as follows: cyan stands for carbon, grey stands for hydrogen, red stands for oxygen, blue stands for nitrogen.

Hedamycin is a member of pluramycin family which are antitumor and antrimicrobial agents and it is more selective for 5'(C/T)G sequence (Hansen & Hurley, 1995). Another study shows that hedamycin binds to GC rich sequences and inhibits the transcription of survivin protein which is an apoptosis inhibitor protein leading to cancer cell death (Wu et al., 2005). It has been found that hedamycin shows higher preference for 5'-CGT-3' over 5'-CGC-3' based on location of N,N-dimethylglucosamine saccharide group facing the minor groove (Owen et al., 2002). Hedamycin has two different sequential binding modes that are; intercalation to first 5' C*G of $d(CGTACG)_2$ that is identified as reversible binding mode, and alkylation of N7 of guanine from 3' GC that is identified as irreversible covalent binding. Also, they observed that second hedamycin binds symmetrically to second 5'C*G (Pavlopoulos et al., 1999). On the other hand, hedamycin is grouped as major groove binder (Hamilton & Arya, 2012; Mishra et al., 2017).

2.5.1.4 Nogalamycin (methyl(1*R*,10*S*,12*S*,13*R*,21*R*,22*S*,23*R*,24*R*)-23-(dimethylamino)-4,8,12,22,24-pentahydroxy-1,12-dimethyl-6,17-dioxo-10-[(2*R*,3*R*,4*R*,5*S*,6*S*)-3,4,5-trimethoxy-4,6-dimethyloxan-2-yl]oxy-20,25-dioxahexacyclo[19.3.1.$0^{2,19}$.$0^{5,18}$.$0^{7,16}$.$0^{9,14}$]pentacosa-2,4,7(16),8,14,18-hexaene-13-carboxylate)

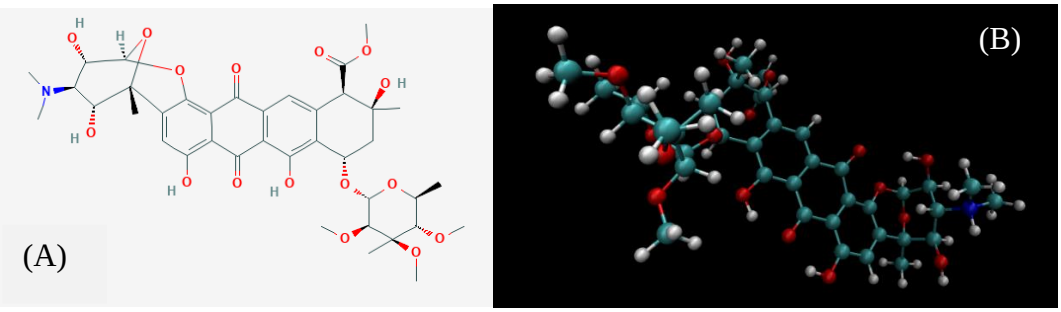


Figure 18. (A) Lewis structure of nogalamycin (PubChem CID: 5289019) (NIH, n.d.). (B) Three-dimensional (3D) conformation of nogalamycin. Color representations are as follows: green stands for carbon, grey stands for hydrogen, red stands for oxygen, blue stands for nitrogen.

Nogalamycin is classified as a major groove targeting binder antibiotic/antitumor due to large size of carbohydrates interacting with the major groove (M. Wang et al., 2016). Controversially, it has been suggested that nogalamycin is an intercalator which shows sequence selectivity towards 3' guanine residue where alternating pyrimidine-purine sequences are present via hydrogen boding between N2 and N7 atoms of guanine (Smith et al., 1996). Orientation of threaded intercalation is suggested as where nogalose sugar placed in minor groove and hydrogen bond formed between carbonyl oxygen of methyl ester and 2-NH_2 of guanine at intercalated site and positively charged bicyclic amino group placed in major groove of 5'-CpG of the center of $d(GACGTC)_2$ via interactions of 2'-OH and 4'-OH of bicyclo sugar (Seakle & Bicknell, 1992). It has been suggested in another study that nogalamycin intercalates to GC sequences sandwiched with AT pairs when DNA opened up for penetration of bulky groups (Liaw et al., 1989).

2.5.2 Minor Groove Binders

We selected 6 minor groove binders that are netropsin, distamycin, DAPI, berenil, pentamidine and hoechst33258.

2.5.2.1 Netropsin (*N*-[5-[(3-amino-3-iminopropyl)carbamoyl]-1-methylpyrrol-3-yl]-4-[[2-(diaminomethylideneamino)acetyl]amino]-1-methylpyrrole-2-carboxamide)

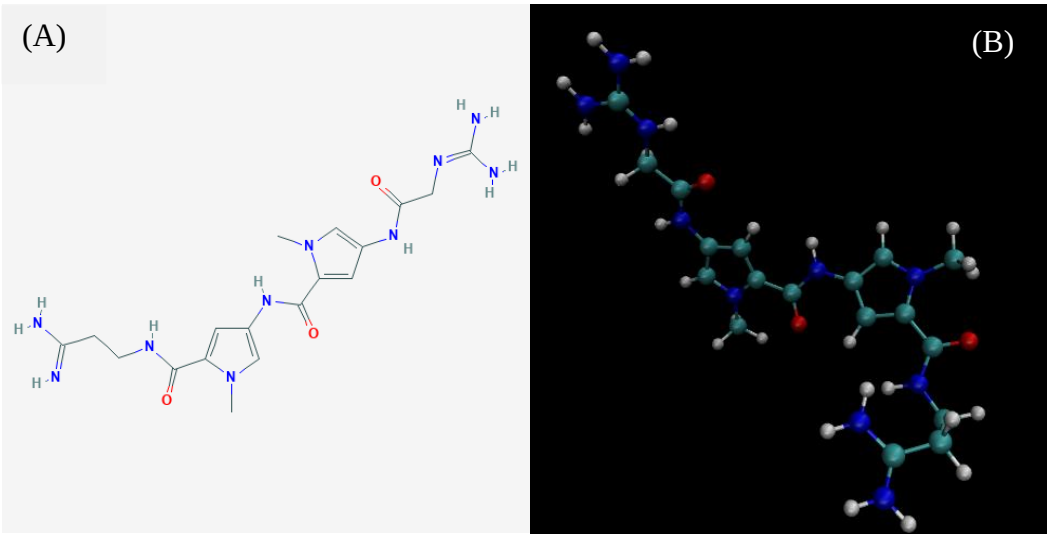


Figure 19. (A) Lewis structure of netropsin (PubChem CID: 4461) (NIH, n.d.). (B) Three-dimensional (3D) conformation of netropsin. Color representations are as follows: green stands for carbon, grey stands for hydrogen, red stands for oxygen, blue stands for nitrogen.

Netropsin is a natural oligopeptide antibiotic which is selective for $(AT)_4$ (Hartley et al., 2009). In another study, it has been shown that netropsin shows highest affinity for AATT sequence with the binding order AAAA = AATT > ATTA/TAAT = TTAA = ATAT = TATA (Abu-daya et al., 1995). It has been shown by agreement of experimental and theoretical studies that minor groove width depends on the orientation of netropsin molecule where the guanidium terminus binds to narrower part and amidinium terminus binds to the wider part. Netropsin has four degrees of freedom before binding and degrees of freedoms of netropsin are limited by its

binding to DNA (Fang, Y., Morris V. R., Lingani, G. M., Long, E., Southrland, 2010).

2.5.2.2 Distamycin (*N*-[5-[[5-[[5-[[5-[(3-amino-3-iminopropyl)carbamoyl]-1-methylpyrrol-3-yl]carbamoyl]-1-methylpyrrol-3-yl]carbamoyl]-1-methylpyrrol-3-yl]carbamoyl]-1-methylpyrrol-3-yl]-4-formamido-1-methylpyrrole-2-carboxamide)

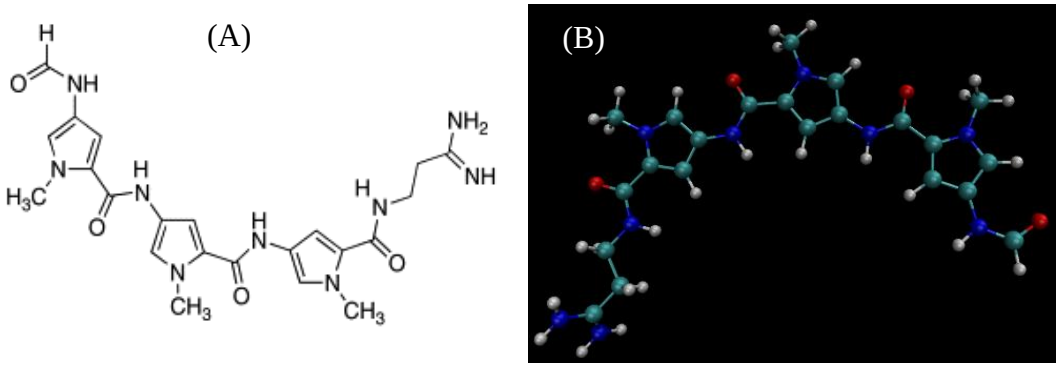


Figure 20. (A) Lewis structure of distamycin (Madkour, 2019) . (B) Three-dimensional (3D) conformation of distamycin. Color representations are as follows: green stands for carbon, grey stands for hydrogen, red stands for oxygen, blue stands for nitrogen.

Distamycin is a natural oligopeptide antibiotic which is selective for $(AT)_5$ (Hartley et al., 2009). Distamycin shows similar affinity order with netropsin (AAAA = AATT > ATTA/TAAT > TTAA=TATA > ATAT) but the affinity for best and worst binding sites for distamycin is lower (Abu-daya et al., 1995). It has been stated that the sequence specificity of distamycin is a result of the hydrogen bonding between NHs of amide of distamycin and O(2) of thymine and N(3) of adenine of 5-base-pair binding site (Schultz & Dervan, 1984).

2.5.2.3 DAPI (2-(4-carbamimidoylphenyl)-1*H*-indole-6-carboximidamide)

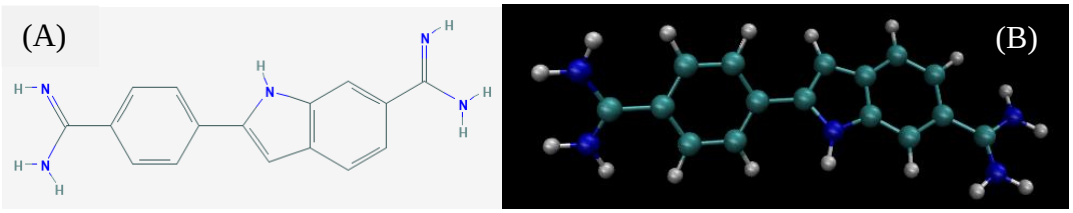


Figure 21. (A) Lewis structure of DAPI (PubChem CID: 2954 (NIH, n.d.)). (B) Three-dimensional (3D) conformation of DAPI. Color representations are as follows: green stands for carbon, grey stands for hydrogen, blue stands for nitrogen.

DAPI is a fluorescent stain that is commonly used in fluorescence microscopy to quantify DNA. Its affinity is in order of AATT >>TAA ≈ATAT > TATA ≈TTAA. Minor contribution in stabilization of DAPI in the minor groove is coming from the hydrogen bonding of two benzimidazole imino NHs to thymine O-2 and adenine N-3 (Breusegem et al., 2002). In a study it has been shown that amino group of guanine base does not prevent sterically the minor groove binding as commonly believed, but it even allows the formation of strong hydrogen bonds with cytosine base. The reason of better binding toward A/T rich sequences is described as electrostatic repulsions of G/C rich sequences that prevents the penetration of DAPI to the center of the G/C rich minor groove (Mohan et al., 1994). DAPI has strong side effects when it is used as an anti-microbial (Cai et al., 2009).

2.5.2.4 Berenil (4-[2-(4-carbamimidoylphenyl)iminohydrazinyl]benzenecarboximidamide)

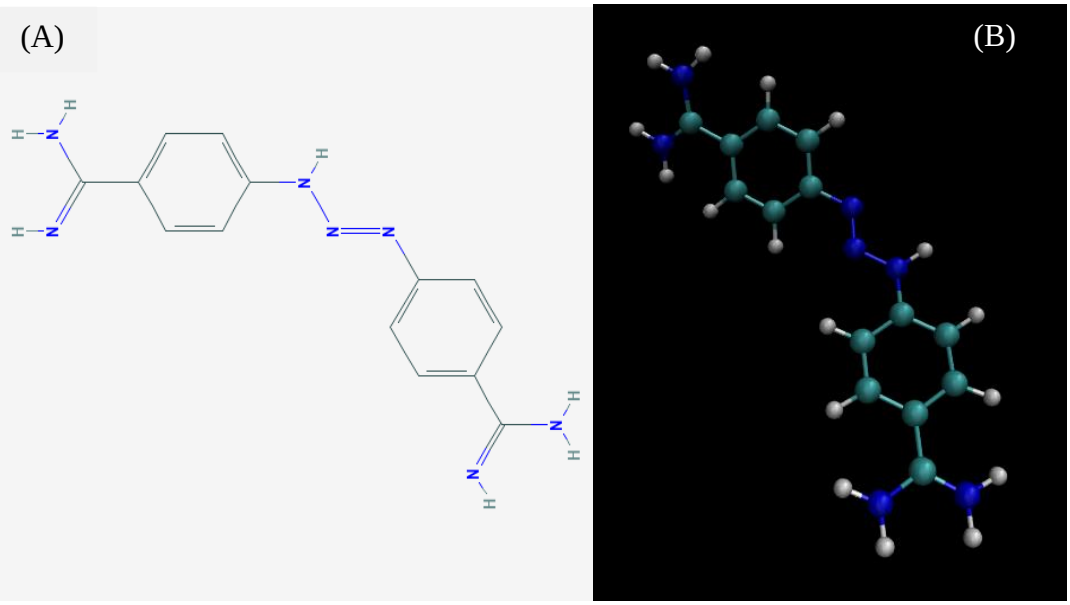


Figure 22. (A) Lewis structure of berenil (PubChem CID: 2354) (NIH, n.d.). (B) Three-dimensional (3D) conformation of berenil. Color representations are as follows: green stands for carbon, grey stands for hydrogen, blue stands for nitrogen.

Binding affinity of berenil is ordered as:

AAAA=ATAT=AATT=TTAA>ATTA/TAAT>TATA in minor groove (Abu-daya et al., 1995). Berenil shows also high affinity for three-base pair ATT sequence and is used as a medicine for trypanosomaisis in veterinary beside its anticancer activity (Cai et al., 2009). It has been suggested that berenil which can bind both DNA and RNA might also be an intercalator, as evidenced by unwinding of negative supercoils in the pBR322 plasmid (Pilch et al., 1995). By Circular Dichroism (CD) experiments and Molecular Mechanics (MM) calculations, it has been proven that berenil not only binds to A/T rich sequences of minor grooves, but it also intercalates through major groove of G-C rich sequences where one of its phenyl groups stacks in one strand, the half of the other aminophenyl group is not stacked and interacting with the minor groove via the hydrogen bonds provided by one of the amino groups and the O4' of G6 which leads to the better distribution of charges (Barceló et al., 2001) .

2.5.2.5 Pentamidine (4-[5-(4-carbamimidoylphenoxy)pentoxy]benzenecarboxim

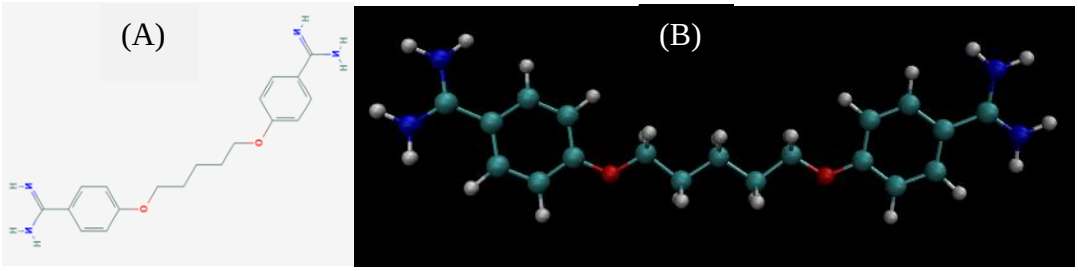


Figure 23. (A) Lewis structure of pentamidine (PubChem CID:4735) (NIH, n.d.). (B) Three-dimensional (3D) conformation of pentamidine. Color representations are as follows: green stands for carbon, grey stands for hydrogen, red stands for oxygen, blue stands for nitrogen.

Pentamidine is effectively used in AIDS treatment with an unknown mechanism and has A/T rich sequence specificity which bound to 5'-AATT sequence in d(CGCGAATTCGCG)$_2$ (Montanari et al., 1998). Also, pentamidine is one of the well-known anti-microbial DNA groove binders used against protozoa *Pneumocystis jirovecii* pneumonia with known side effects like toxicity, nephrotoxicity, cardiotoxicity, and hepatotoxicity. Additionally, is known that pentamidine inhibits the oncogenic PRL phosphatases in treatment of pancreatic cancer (Cai et al., 2009)

2.5.2.6 Hoechst 33258 (4-[6-[6-(4-methylpiperazin-1-yl)-1H-benzimidazol-2-yl]-1H-benzimidazol-2-yl]phenol)

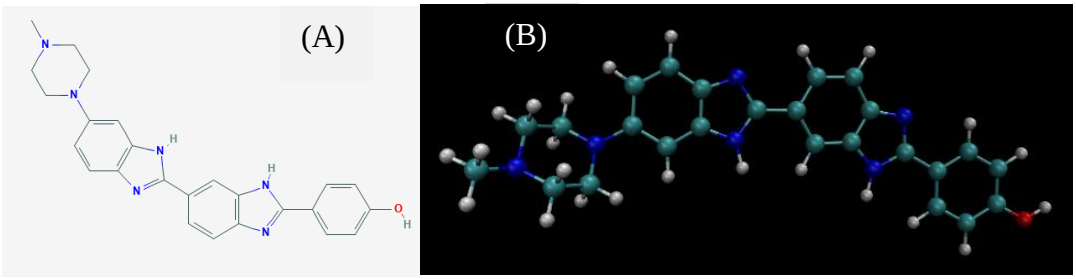


Figure 24. (A) Lewis structure of hoechst 33258 (PubChem CID:2392) (NIH, n.d.). (B) Three-dimensional (3D) conformation of Hoechst 33258. Color representations

are as follows: green stands for carbon, grey stands for hydrogen, red stands for oxygen, blue stands for nitrogen.

Hoechst 33258 shows similar trend with netropsin in affinity order AATT > AAAA > ATTA/TAAT = TTAA = ATAT > TATA, but affinity for AATT of hoechst 33258 is much higher than the affinity for AATT of netropsin and affinity for rest of the sequences(TAAT, TTAA, ATAT, TATA) of hoechst 33258 is significantly lower with respect to netropsin (Abu-daya et al., 1995). Hoechst 33258 shows same affinity order with DAPI but with higher sensitivity due to its larger size providing higher number of contacts which is AATT >>TAAT ≈ATAT > TATA ≈TTAA (Breusegem et al., 2002).

CHAPTER 3

RESULTS AND DISCUSSIONS

3.1 Visualization of the Results

Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values were plotted by using probability density function (PDF) of bound ligands along each DNA sequence. These distribution plots are shown in figure 26. PDFs were calculated by scipy kernel density estimation function using Gaussian kernels. Each plot represents the distribution of docking attempts. The x-axis indicates the score or MM-refinement free energy of the pose, while the y-axis represents the projection of the center of mass coordinate of each docked ligand pose to the DNA sequence onto the principal axis of the DNA sequence, i.e., the location of the drug in the sequence. The z-axis is the probability density with color scaling.

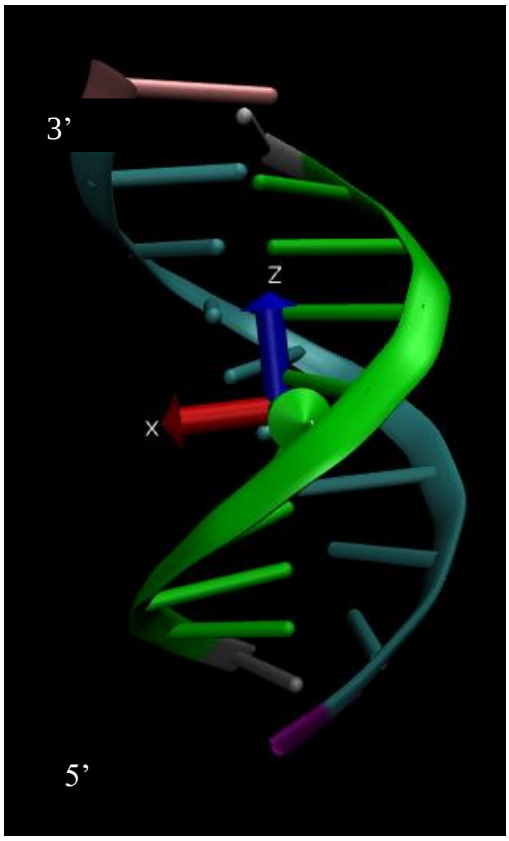


Figure 25. DNA alignment on z-axis and centered to origin

DNA molecules in the complex structures were aligned on their principal axis (z-axis in Figure 25) with same directionality. Then, center of mass coordinate of each docked ligand is projected onto the principal axis of DNA. A value of +15 on the y-axis of the plots represents the upper-most base location of 3’ end, 0 represents the center of the sequence, and -15 represents the bottom-most base location of 5’ end. The color scale on the PDFs shows the probability density. Dark red represents highest probability density values and dark blue shows lowest density probability values. At this point it should be noted that probability density values indicate the amount of times a given pose with a given score was generated by AutoDock Vina. It is not the probability of occurrence for a given binding mode. The plots for berenil in interaction with GGCCAATTGG are presented in figure 26 and discussed below, while corresponding figures for all other complexes can be found in appendices A.

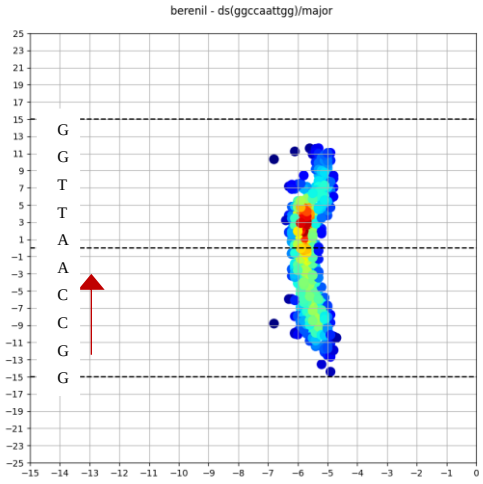
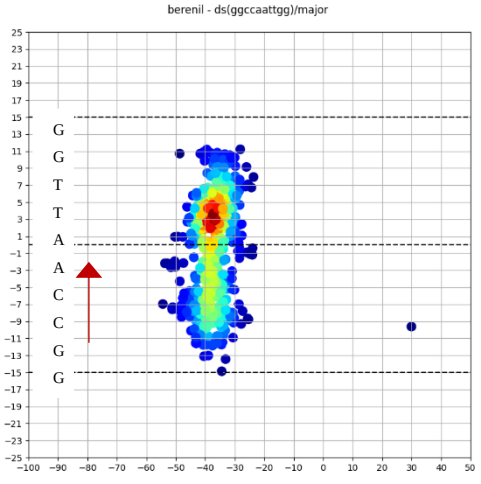
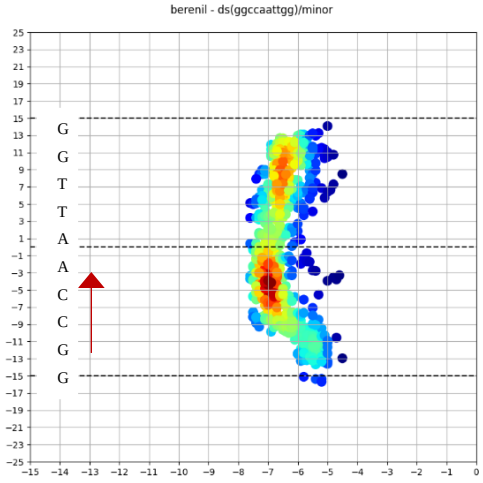
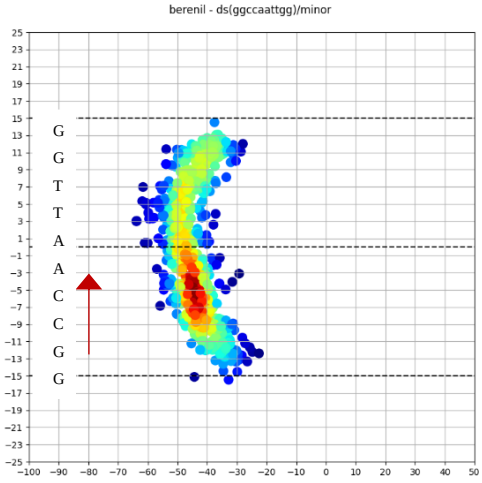
Autodock Vina Score Distribution	MM-Refinement Distribution
A 	B 
C 	D 

Figure 26. Three-dimensional (3D) distribution graphs of Autodock Vina scores (shown in first column) and MM-refinement binding free energy values (shown in second column) of berenil-5'-GGCCAATTGG -3'.

In figure 26, 3D binding free energy distribution plots of Autodock Vina docking scores and MM-refinement binding energy distribution plots along the DNA were shown. Here each point belongs to one docked pose. More negative scores along x

axis represent more favorable and realistic poses. Assessed binding scores of Autodock Vina for corresponding binding poses only vary within 0.1 kcal/mol precision and are all confined in a rather narrow range (i.e., -4.0 to -8.0 kcal/mol; figures 26.A and 26.C). Such a narrow range and small difference in binding score makes it difficult to discriminate between good and bad poses. That is one issue of the scoring function of Autodock Vina that requires further refinement. The energy values obtained after MM-refinement range between -70.00 kcal/mol and -20.00 kcal/mol (figures 26.B and 26.D), with one outlier at +30 kcal/mol. In comparison with the Autodock Vina score distribution plots, the new energy distribution plots obtained after MM-refinement show a much wider distribution of energy and a few points that move out of the distribution centers. The significant difference in free energy score in the latter case allows to better discriminate between good and bad poses, and to identify the most promising ones.

When Autodock Vina binding free energy scores in major groove (fig. 26.A) and in minor groove (fig.26.C) are compared, one can see that binding scores in minor groove are more favorable rather than major groove. Autodock Vina is already capable of identify groove binding preference, as expected from experimental observations. However, Autodock Vina distribution plot in the minor groove (fig. 26.C) cannot show the best binding site in the sequence since it predicts quite similar binding scores for any site along GGCCAATTGG In comparison with Autodock Vina score distribution plot in minor groove (fig. 26.C), the energy distribution pattern is significantly different after MM-refinement (fig. 26.D). Minor groove MM-binding free energy distribution plot of berenil-GGCCAATTGG complex shows a very distinctive binding trend, where the distribution is bent. Here, as the drug moves towards the edges of the GGCCAATTGG sequence, the absolute value of binding energy decreases. It indicates that binding is less favorable toward the guanine or cytosine rich sequences in the minor groove. Reversely, the binding at AATT is predicted to be significantly greater by the MM-score. Therefore, MM-energy distribution plot in minor groove (table 1.D) shows sequence selectivity of berenil toward AATT region.

3.2 Autodock Vina Score vs. MM Binding Free Energy

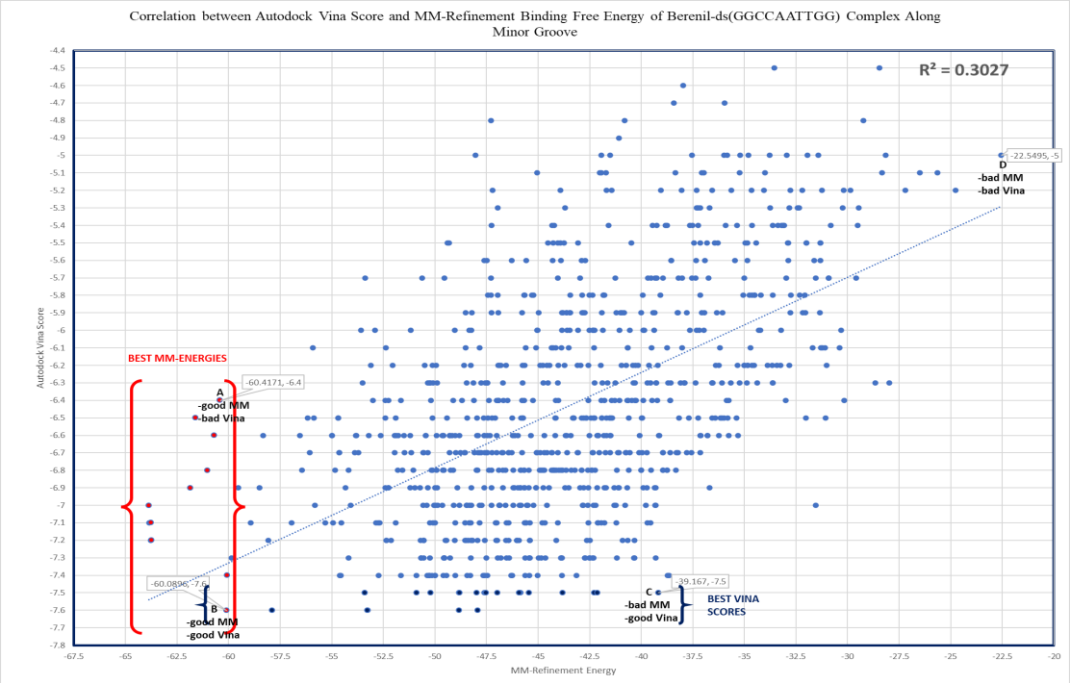


Figure 27. Correlation between Autodock Vina score and MM-Refinement binding free energy

Figure 27 shows the correlation between Autodock Vina score and MM-Refinement binding free energy. Better Autodock Vina Scores can be found in the lower part of the y-axis and better MM-refinement binding free energy values can be found on the left-hand side of the x-axis. From this figure it can be understood that correlation between Autodock Vina results and MM-refinement results is very poor, where no linearity is present between two data sets. The problem is that Vina predicts some poses as very good, while they shouldn't score that high. At the very bottom of this correlation plot, best Autodock Vina scores yielded various MM-refinement binding free energy values where most of them very bad. Only a few of the best Vina poses yielded good MM-refinement binding free energy values. Also, one can appreciate that for the same Autodock Vina score of -7.5 or -7.6, MM-binding free energy values vary roughly from -60.00 to -30.00 kcal/mol. This outlines that the scoring function is not sensitive enough to distinguish some poses from one another. On the

other hand, most negative MM-refinement binding free energy values are not necessarily the poses with the best Autodock Vina scores. It indicates that using filtering only the best Autodock Vina poses and leaving out the others would result in missing of the best poses. Yet, Autodock Vina yields a very good sampling of relevant poses, which must be refined to yield reliable scoring.

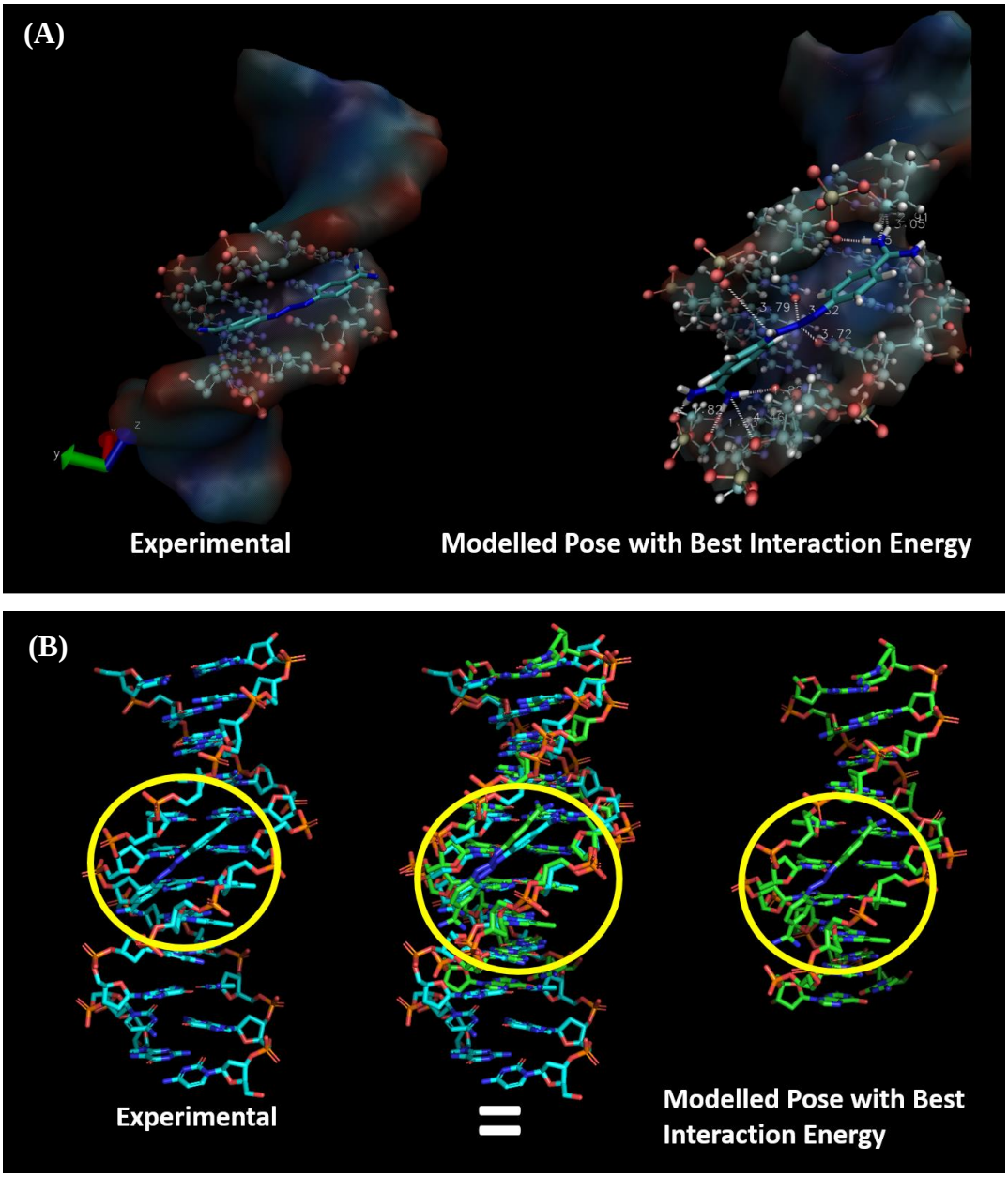


Figure 28. (A) Comparison of experimental crystal structure (2GVR) and best modelled pose (B) Overlap of experimental and modelled pose

Figure 26.D shows that best binding free energies are obtained at AATT region in the middle of the minor groove of the sequence where a lot of poses obtained with negative binding free energies. In the left-hand side of the figure 28, crystal structure of berenil-ds(CGCGAATTCGCG) complex (Brown et al., 1990) with PDB ID 2GVR shows that berenil prefers to bind 5'-AATT sequence of minor groove. The most favorable MM-refined modelled pose is taken from figure 27 which is the red dot at the very left of the plot and illustrated in the right-hand side of figure 28. MM-refinement yielded very accurate result where the best modelled pose is in a very good agreement with the experimental observations.

3.3 Interesting Poses for Berenil

Four points were selected from figure 27 to investigate what the Autodock Vina identifies correctly and what Autodock Vina cannot identify correctly during its scoring. At point A, Autodock Vina score is not so good, but MM-refinement energy is very good. Therefore, point A in figure 29 is a pose that could not be scored well by Autodock Vina. At point B, both of Autodock Vina score and MM-refinement energy is good and matching. Therefore, point B in figure 30 is a pose that could be scored well by Autodock Vina. At point C, Autodock Vina score is good, but MM-refinement energy is not favorable. Therefore, point C in figure 31 is not a pose that could be scored well by Autodock Vina. At point D, neither Autodock Vina score nor MM-refinement energy is not good. At this point, result is matching. Therefore, point D in figure 32 is a pose that could be scored well by Autodock Vina.

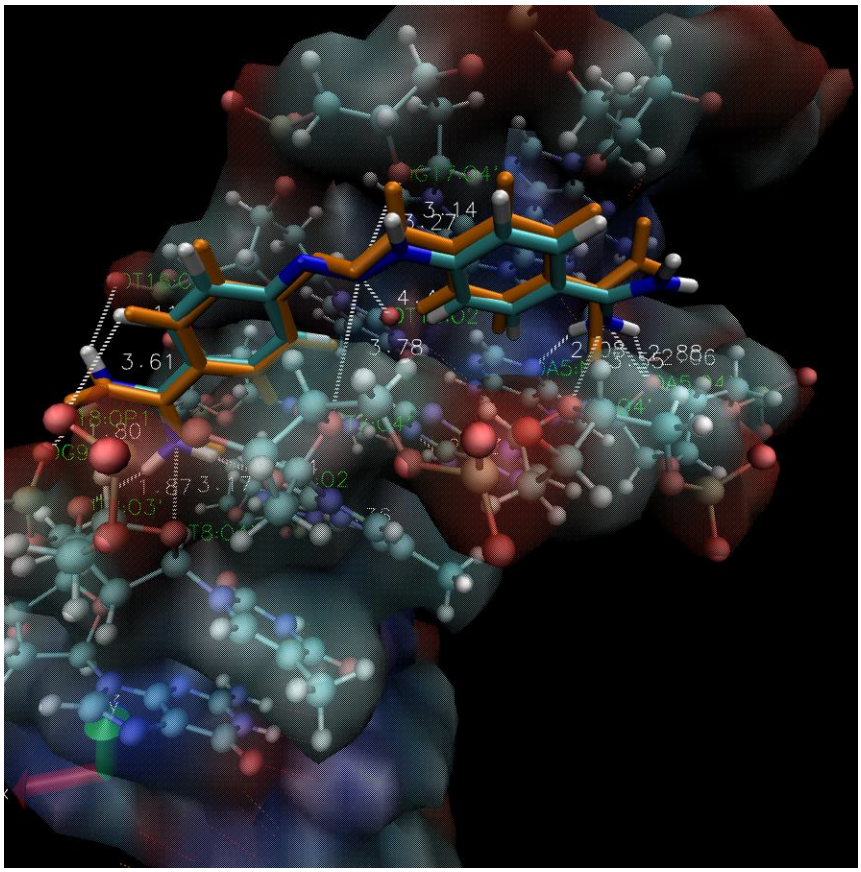


Figure 29. Binding pose at point A where MM-energy is very good and Vina score is not so good. Orange: Vina docking pose & blue: MM-refinement pose

In figure 29, Autodock Vina pose at point A and MM-Refinement pose at point A were shown respectively. At point A, Autodock Vina score is -6.4 which is not so good, whereas MM-score is good -60.42 kcal/mol. After MM-refinement, functional groups at the end, which are positively charged guanidiums, establish hydrogen bonds and stabilized better by electrostatic interactions.

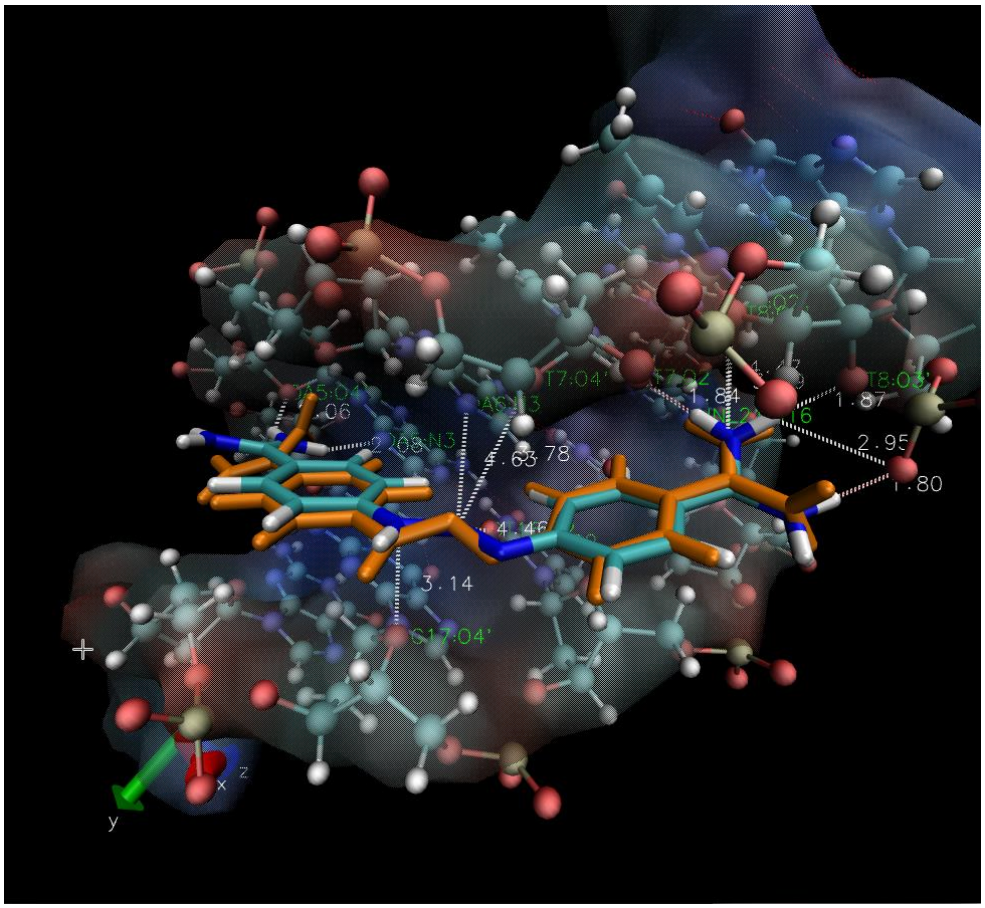


Figure 30. Binding pose at point B where MM-energy is good and Vina score is good as well. Orange : Vina docking pose& blue :MM-refinement pose

In figure 30, Autodock Vina pose at point B and MM-Refinement pose at point B were shown respectively. At point B, Autodock Vina score is -7.6 which can be counted good and MM-score shows that berenil binds to AATT region with favorable value -60.09 kcal/mol. In Autodock Vina pose, there no hydrogen bonds are observed. Here it can be seen that, the structure of the drug after MM-refinement did not change much and therefore interactions are remained the same.

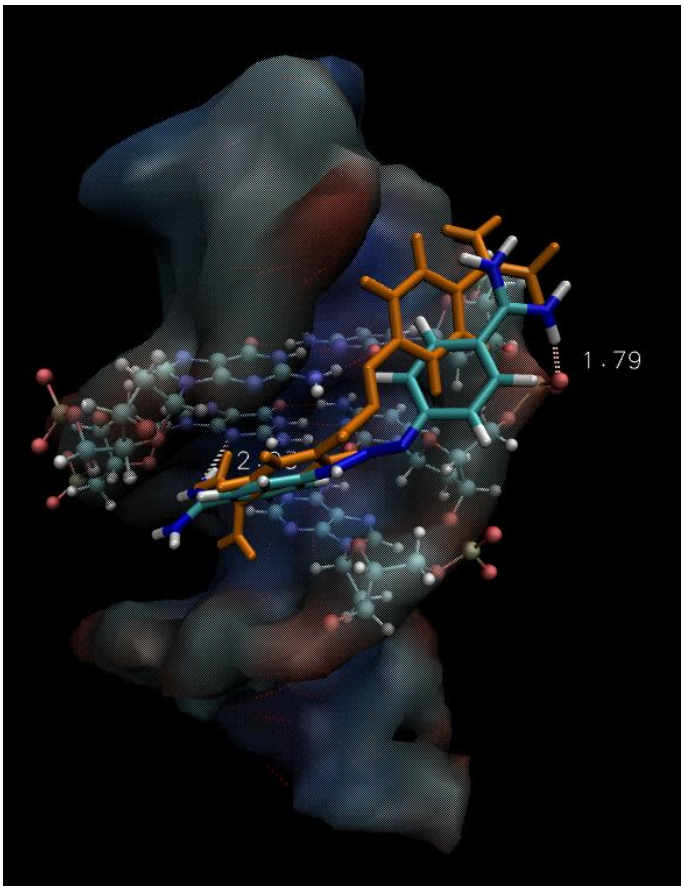


Figure 31. Binding pose at point C where MM-energy is bad and Vina score is quite good. Orange : Vina docking pose & blue: MM-refinement pose

In figure 31, Autodock Vina pose at point C and MM-Refinement pose at point C were shown respectively. At point C, Autodock Vina score is -7.5 which can be counted good, whereas MM-score is quite bad with a value of -39.167. In Autodock Vina pose, drug molecule is in the DNA's groove. But after MM-refinement, the molecule has drifted away from the groove, indicating that the Vina pose was not well-stabilized.

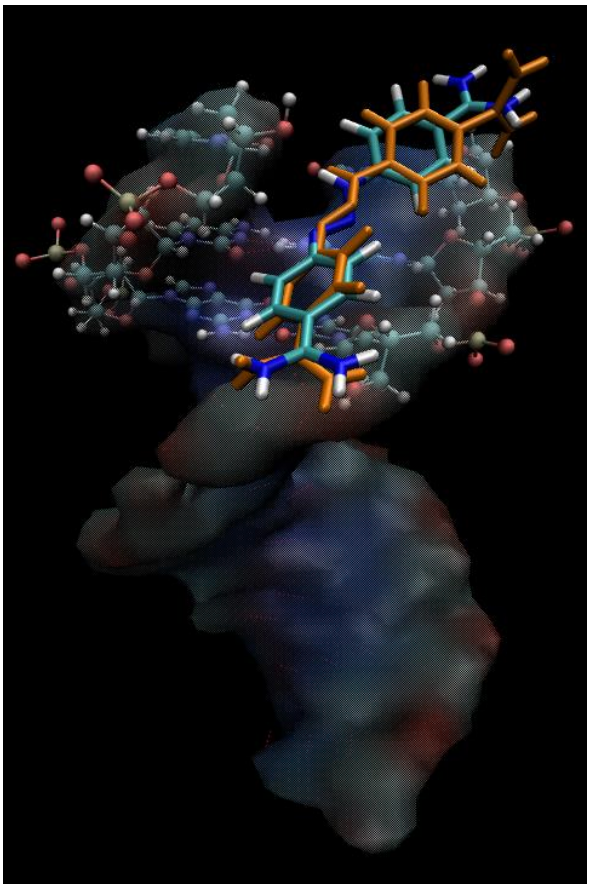


Figure 32. Binding pose at point D where MM-energy is bad and Vina score is quite bad as well. Orange : Vina docking pose & blue :MM-refinement pose

In figure 32, Autodock Vina pose at point D and MM-Refinement pose at point D were shown respectively. At point D, Autodock Vina score is -5 which can be counted very bad and MM-score is also quite bad with value -22.55 kcal/mol. The pose is predicted outside of the groove and is similar in both methods. Here both methods identified that neither electrostatic interactions nor hydrogen bonding stabilized the molecule at all. At this worst binding point, Autodock Vina predicted correct. By looking all of those figures, one can state that Autodock Vina yields mixed combination of good and bad accuracy. Therefore, to be able to better identify electrostatic and hydrogen bonding contributions of docked poses, MM-refinement seems mandatory.

3.4 The Importance of Choice of MM Parameters

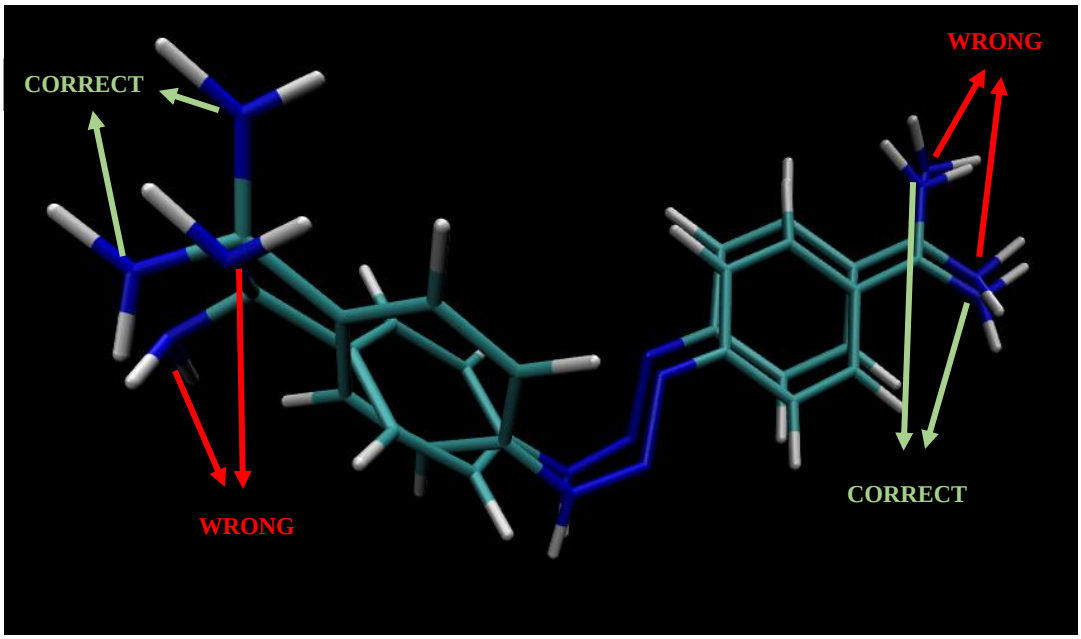


Figure 33. Wrong assignment of the atom types in the guanidinium group in berenil by ANTECHAMBER (red arrows) and corrected version of these atom types (green arrows).

Figure 33 illustrates the geometry of amidinium ends in berenil molecule resulted by wrong atom type assignment by ANTECAMBER by red arrows and green arrows show the corrected geometry of amidinium by correct atom type assignment. Automatic assignment of the atom types of the amidinium groups by ANTECHAMBER made the C-C bond rotatable in the amidinium, although it should not be. This disturbed planarity of amidinium groups and deviation from the actual binding free energy values. This is one of the difficulties of MM that makes it harder to generalize compared to standard docking (e.g., Autodock Vina). This error was spotted for berenil and the results corrected accordingly. For other molecules in our set containing an amidinium gorup, molecular mechnics steps will be repeated to correct the results.

3.5 Other Complexes

Table 1. Experimental and computed selectivity for selected drug molecules

LIGAND	EXPERIMENTAL KNOWLEDGE	CALCULATED SEQUENCE	PREFERRED GROOVE	BINDING FREE ENERGY (KCAL/MOL)
Hedamycin	-Intercalation to GC rich sequence -Major groove binder	ds(ATCGCGCGAT)	Major	-92.08
Hoechst 33258	-Minor groove binder AATT >>TAAT ≈ATAT > TATA ≈TTAA	ds(GGCCAATTGG)	Minor	-61.62
Altromycin B	-Intercalation to CG -Major groove binder	ds(CGTATATACG) ds(GGGGGGGGGG)	Minor Minor	-89.35 -87.49
Nogalamycin	-Intercalates to GC sequences sandwiched with AT	ds(GGGGGGGGGG) ds(CGTATATACG) ds(GGCCAATTGG) ds(AAAAAAAAAA)	Major Minor Minor Minor	-67.73 -66.77 -66.54 -65.69
Pluramycin A	-Intercalation to TATA -Major groove binder -Intercalation selectivity towards 5'(C/T)G	ds(GGGGGGGGGG) ds(ATCGCGCGAT)	Major Major	-85.24 -67.00

A detailed analysis of the other complexes in our set can be made similar to the case of berenil. Here we mention the most relevant findings from this analysis and all plots are presented in Appendix A. As shown in table 1, altromycin B prefers to bind CG parts of ds(CGTATATACG)/minor groove with -89.35 kcal/mol and ds(GGGGGGGGGG)/minor groove with -87.49 kcal/mol instead of major grooves of the sequences as expected from experiements. Hedamycin has the best binding free energy -92.08 kcal/mol in the CG region of ds(ATCGCGCGAT)/mojor groove, which is compatible with the literature given in the section 2.5. Hoechst 33258 binds to AATT region of ds(GGCCAATTGG)/minor groove with best -61.62 kcal/mol, which also fits to the literaure information in section 2.5, ds(AAAAAAAAAA)/minor with -59.91 kcal/mol. Nogalamycin binds best to ds(GGGGGGGGGG)/major groove with -67.73 kcal/mol, unexpectedly to the TATA region of ds(CGTATATACG)/minor groove with -66.77 kcal/mol, to the AATT region of ds(GGCCAATTGG)/minor groove with -66.54 kcal/mol and to ds(AAAAAAAAAA)/minor groove with -65.69 kcal/mol. Pluramycin A interacts best with ds(GGGGGGGGGG)/major groove with best energy value,that is -85.24 kcal/mol compared to all sequneces. After that, best interaction energy -67.00

kcal/mol of pluramycin A is given by the interaction with GC region of ds(ATCGCGCGAT)/major groove as given in the literature section 2.5.

CHAPTER 4

CONCLUSIONS AND FUTURE PLANS

In this study, we devised new protocol for prediction of binding mode of any drug molecule and DNA complex. We applied molecular docking with Autodock Vina, and molecular mechanics based refinement. We trained our software with total of 50 DNA/drug complexes. We plotted distribution of energies of each drug/DNA complex along both major and minor grooves. These distribution plots give hint for sequence selectivity of drug molecules which can be counted a very promising progression for DNA-binding drug discovery. New drug candidates can be tested to identify sequence selectivity with our software.

With our current protocole, many pose in the ensemble generated by Autodock Vina are identical or nearly identical. Therefore, clustering should be added to the pipeline in order to eliminate identical binding poses in the future.

Additionally, molecular dynamics should be applied after molecular mechanics based refinement. Molecular mechanics geometry optimization consider the complex at 0 K and ignores the temperature contribution to the dynamics of the system. Hovewer, the body temperature is 37 °C which contributes to kinetic energy and therefore motion of the molecules. Therefore, stability of the formed complex structures should be assesed at body temperature over time.

In this study, we disregarded intercalation and focused only groove binding. To have complete non-covalent interaction scan over DNA molecule, intercalation type of interaction also be included.

REFERENCES

Abu-daya, A., Brown, P. M., & Fox, K. R. (1995). DNA sequence preferences of several AT-selective minor groove binding ligands. *Nucleic Acid Research.23*(17), 3385–3392.

Adjei, A. A. (2001). Blocking oncogenic ras signaling in cancer. *REVIEW Journal of the National Cancer Institute*, *93*(14), 1062–1074.

Alberts B, Johnson A, Lewis J. (2002). DNA binding motifs in gene regulatory proteins. *Molecular Biology of the Cell.* 4th ed. New York.

Aminpour, M., Montemagno, C., & Tuszynski, J. A. (2019). An overview of molecular modeling for drug discovery with specific illustrative examples of applications. *Molecules*, *24*(9). https://doi.org/10.3390/molecules24091693

B.R. (1958). Chemistry and biology of mucopolysaccharides. *American Heart Journal*, *56*(5), 796–797. https://doi.org/10.1016/0002-8703(58)90224-2

Barceló, F., Ortiz-Lombardía, M., & Portugal, J. (2001). Heterogeneous DNA binding modes of berenil. *Biochimica et Biophysica Acta - Gene Structure and Expression*, *1519*(3), 175–184. https://doi.org/10.1016/S0167-4781(01)00233-0

Bauer, G. B., & Povirk, L. F. (1997). Specificity and kinetics of interstrand and intrastrand bifunctional alkylation by nitrogen mustards at a G-G-C sequence. *Nucleic Acid Research. 25*(6), 1211–1218.

Breusegem, S. Y., Clegg, R. M., & Loontiens, F. G. (2002). Base-sequence specificity of Hoechst 33258 and DAPI binding to five (A/T)4 DNA sites with kinetic evidence for more than one high-affinity Hoechst 33258-AATT complex. *Journal of Molecular Biology*, *315*(5), 1049–1061. https://doi.org/10.1006/jmbi.2001.5301

Brown, D. G., Sanderson, M. R., Skelly, J. V., Jenkins, T. C., Brown, T., Garman, E., Stuart, D. I., & Neidle, S. (1990). Crystal structure of a berenil - dodecanucleotide complex: The role of water in sequence-specific ligand

binding. *EMBO Journal*, *9*(4), 1329–1334. https://doi.org/10.1002/j.1460-2075.1990.tb08242.x

Cai, X., Gray, P. J., & Von Hoff, D. D. (2009). DNA minor groove binders: Back in the groove. *Cancer Treatment Reviews*, *35*(5), 437–450. https://doi.org/10.1016/j.ctrv.2009.02.004

Chaires, J. B. (1997). Energetics of Drug – DNA.*Biopolymers*.44(3),201-205. https://doi.org/10.1002/(SICI)1097-0282(1997)44:3<201::AID-BIP2>3.0.CO;2-Z

Chaires, J. B. (2006). A thermodynamic signature for drug – DNA binding mode. *Archives of Biochemistry and Biophysics*.*453*, 26–31. https://doi.org/10.1016/j.abb.2006.03.027

Chen, J. (2016). The cell-cycle arrest and apoptotic and progression. *Cold Spring Harbor Perspectives in Biology*, 1–16.

Chin, C. F., & Yeong, F. M. (2010). Safeguarding Entry into Mitosis : the Antephase Checkpoint MINIREVIEW Safeguarding Entry into Mitosis : the Antephase Checkpoint .*Molecular and Cellular Biology.* 30(1), 22-32.https://doi.org/10.1128/MCB.00687-09

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., & Kollman, P. A. (1995). A Second Generation Force Field for the Simulation of Proteins , Nucleic Acids , and Organic Molecules. *J. Am. Chem. Soc.* 117, 5179–5197.

Dang, C. V. (1999). c-Myc Target Genes Involved in Cell Growth, Apoptosis, and Metabolism. *Molecular and Cellular Biology*, *19*(1), 1–11. https://doi.org/10.1128/mcb.19.1.1

Decherchi, S., Masetti, M., Vyalov, I., Rocchia, W. (2015). Implicit solvent methods for free energy estimation. *Eur J Med Chem*, *91*, 27–42. https://doi.org/10.1016/j.ejmech.2014.08.064.Implicit

Eberhardt, J., Santos-Martins, D., Tillack, A. F., & Forli, S. (2021). AutoDock Vina

1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *Journal of Chemical Information and Modeling*, *61*(8), 3891–3898.

Fang, Y., Morris V. R., Lingani, G. M., Long, E., Southrland, W. M. (2010). Genome-Targeted Drug Design: Understanding the Netropsin- DNA Interaction. *The Open Conference Proceedings Journal*, *23*(1), 157–163. https://doi.org/https://doi.org/10.2174/22102892010010100157

Forouzesh, N., Mishra, N. (2021). An Effective MM / GBSA Protocol for Absolute Binding Free. *Molecules*, *26*(2383). https://doi.org/https://doi.org/10.3390/molecules26082383 Academic

Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*, *10*(5), 449–461. https://doi.org/10.1517/17460441.2015.1032936

Gujar, H., Weisenberger, D. J., & Liang, G. (2019). The roles of human DNA methyltransferases and their isoforms in shaping the epigenome. *Genes*, *10*(2). https://doi.org/10.3390/genes10020172

Hamad, D. F. (Western M. U. (2017). Novel Spectroscopic Tools to Differentiate Drug-DNA Binding Interactions. In *Master's Theses. 1130.* https://scholarworks.wmich.edu/masters_theses/1130/?utm_source=scholarworks.wmich.edu%2Fmasters_theses%2F1130&utm_medium=PDF&utm_campaign=PDFCoverPages

Hamilton, P. L., & Arya, D. P. (2012). Natural product DNA major groove binders. *Natural Product Reports*, *29*(2), 134–143. https://doi.org/10.1039/c1np00054c

Hansen, M., & Hurley, L. (1995). Altromycin B Threads the DNA Helix Interacting with Both the Major and the Minor Grooves To Position Itself for Site-Directed Alkylation of Guanine N7. *Journal of the American Chemical Society*, *117*(9), 2421–2429. https://doi.org/10.1021/ja00114a006

Harteis, S., & Schneider, S. (2014). Making the Bend : DNA Tertiary Structure and Protein-DNA Interactions. *Int. J. Mol. Sci*,15, 12335–12363. https://doi.org/10.3390/ijms150712335

Hartley, J. A., Lown, J. W., Mattes, W. B., Kohn, K. W., Lown, J. W., Mattes, W. B., & Sequence, K. W. K. D. (2009). Dna Sequence Specificity of Antitumor Agents : Oncogenes as possible targets for cancer therapy Oncogenes as possible targets for cancer therapy. *Acta Oncologica*,27(5),503-510. https://doi.org/10.3109/02841868809093578

Hollingsworth, S. A., & Dror, R. O. (2018). Molecular Dynamics Simulation for All. *Neuron*, *99*(6), 1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011

Hydbring, P., Malumbres, M., & Sicinski, P. (2016). Non-canonical functions of cell cycle cyclins and cyclin-dependent kinases. *Nature Reviews Molecular Cell Biology*, *17*(5), 280–292. https://doi.org/10.1038/nrm.2016.27

Kastan, M. B., & Bartek, J. (2004). Cell-cycle checkpoints and cancer. *Nature*, *432*(7015), 316–323. https://doi.org/10.1038/nature03097

Khan, G. S., Shah, A., Zia-Ur-Rehman, & Barker, D. (2012). Chemistry of DNA minor groove binding agents. *Journal of Photochemistry and Photobiology B: Biology*, 115, 105–118. https://doi.org/10.1016/j.jphotobiol.2012.07.003

Lei, H., Wang, X., & Wu, C. (2012). Early stage intercalation of doxorubicin to DNA fragments observed in molecular dynamics binding simulations. *Journal of Molecular Graphics and Modelling*, *38*, 279–289. https://doi.org/10.1016/j.jmgm.2012.05.006

Liaw, Y., Gao, Y., Robinson, H., Marel, G. A. Van Der, Boom, J. H. Van, & Wang, A. H. (1989). Antitumor Drug Nogalamycin Binds DNA in Both Grooves Simultaneously : Molecular Structure of Nogalamycin-DNA Complex. *Biochemistry*,28,9913-9918.

Madkour, L. H. (2019). Classifications of DNA binding molecules—Drug interactions. In *Nucleic Acids as Gene Anticancer Drug Delivery Therapy* (pp. 87–101). https://doi.org/10.1016/b978-0-12-819777-6.00007-x

Marion, A. (2014). *Molecular dynamics using a semiempirical quantum force field: development and applications to systems of biological interest*. à l'Université de Lorraine.

Melander, C., Burnett, R., & Gottesfeld, J. M. (2019). Regulation of gene expression with pyrrole – imidazole polyamides. *Journal of Biotchnology,112,195-220*. https://doi.org/10.1016/j.jbiotec.2004.03.018

Mishra, R., Kumar, A., Chandra, R., & Kumar, D. (2017). A review on theoretical studies of various types of Drug-DNA Interaction. *International Journal of Science, Technology and Society*, *3*, 11–27.

Mohan, S., Yathindra, N., & Yathindra, N. (1994). A study of the interaction of dapi with dna containing at and non-at sequences - molecular specificity of minor groove binding drugs. *Journal of Biomolecular Structure and Dynamics*, *11*(4), 849–867. https://doi.org/10.1080/07391102.1994.10508037

Montanari, C. A., Trent, J. O., & Jenkins, T. C. (1998). Molecular recognition of B-DNA minor-groove binders: The rigid analogue approach to synthesise antileishmaniasis compounds. A molecular modeling study. *Journal of the Brazilian Chemical Society*, *9*(2), 175–180. https://doi.org/10.1590/S0103-50531998000200010

Nanjunda, R., & Wilson, W. D. (2012). Binding to the DNA minor groove by heterocyclic dications: From AT-specific monomers to GC recognition with dimers. *Current Protocols in Nucleic Acid Chemistry*, *SUPPL.51*, 1–27. https://doi.org/10.1002/0471142700.nc0808s51

Neidle, S. (2001). DNA minor-groove recognition by small molecules. *Nat. Prod. Rep*,18, 291–309. https://doi.org/10.1039/a705982e

Nguyen, H., Roe, D. R., & Simmerling, C. (2013). Improved Generalized Born Solvent Model Parameters for Protein Simulations. *J. Chem. Theory Comput.*, *9*, 2020−2034. https://doi.org/2034 dx.doi.org/10.1021/ct3010485

NIH.(n.d).*Pubchem*. https://pubchem.ncbi.nlm.nih.gov/compound/

Northern Arizona State University.(n.d). DNA structure. https://www2.nau.edu/lrm22/lessons/dna_notes/dna_notes.html

Okamato, A. (1998). Design, Synthesis and Evaluation of Novel DNA Alkylating

Agents Based on the Chemistry of Antibiotic Kapurimycin A3 [Kyoto University]. https://doi.org/10.11501/3135524

Oleg, T., Olson, J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comput Chem*, *31*(2), 455–461. https://doi.org/10.1002/jcc.21334.AutoDock

Owen, E. A., Keniry, M. A., Burley, G. A., Carver, J. A., & Wickham, G. (2002). Structural investigation of the hedamycin: D(ACCGGT)$_2$complex by NMR and restrained molecular dynamics. *Biochemical and Biophysical Research Communications*, *290*(5), 1602–1608. https://doi.org/10.1006/bbrc.2002.6369

Pavlopoulos, S., Bicknell, W., Wickham, G., & Craik, D. J. (1999). Characterization of the sequential non-covalent and covalent interactions of the antitumour antibiotic hedamycin with double stranded DNA by NMR spectroscopy. *Journal of Molecular Recognition*, *12*(6), 346–354. https://doi.org/10.1002/(SICI)1099-1352(199911/12)12:6<346::AID-JMR476>3.0.CO;2-L

Pilch, D. S., Kirolos, M. A., Liu, X., Plum, G. E., & Breslauer, K. J. (1995). Berenil [1,3-Bis(4'-amidinophenyl)triazene] Binding to DNA Duplexes and to a RNA Duplex: Evidence for Both Intercalative and Minor Groove Binding Properties. *Biochemistry*, *34*(31), 9962–9976. https://doi.org/10.1021/bi00031a019

Rahman,A.,O'Sullivan,P.,Rozas,I.(2018). Recent developments in compounds acting in the DNA minor groove. *Med. Chem. Commun*., 2019, 10,26-40. DOI: 10.1039/c8md00425k

Ren, J., & Chaires, J. B. (1999). Sequence and Structural Selectivity of Nucleic Acid Binding Ligands. *Biochemistry*, 38(49), 16067–16075.

Schultz, P. G., & Dervan, P. B. (1984). Distamycin and penta-n-methylpyrrolecarboxamide binding sites on native dna a comparison of methidiumpropyi-edta-fe(Ii) footprinting and dna affinity cleaving. *Journal of Biomolecular Structure and Dynamics*, *1*(5), 1133–1147. https://doi.org/10.1080/07391102.1984.10507508

Seakle, M. S., & Bicknell, W. (1992). Interaction of the anthracycline antibiotic nogalamycin with the hexamer duplex d(5'-GACGTC)2. *Eur. J. Biochem*, *205*, 45–58.

Sechi, M., Derudas, M., Dallocchio, R., Dessì, A., & Cosseddu, A. (2009). DNA Binders : 1 . Evaluation of DNA-Interactive Ability , Design , and Synthesis of Novel Intercalating Agents. *Letters in Drug Design & Discovery*, 6(1), 56–62.

Sheng, J., Gan, J., & Huang, Z. (2013). Structure-based DNA-targeting strategies with small molecule ligands for drug discovery. *Medicinal Research Reviews*, *33*(5), 1119–1173. https://doi.org/10.1002/med.21278

Silvestri, C., & Brodbelt, J. S. (2013). Tandem mass spectrometry for characterization of covalent adducts of dna with anticancer therapeutics. *Mass Spectrometry Reviews*, 32, 247–266. https://doi.org/10.1002/mas

Simanshu, D. K., Nissley, D. V. , McCormick, F. (2017). RAS Proteins and Their Regulators in Human Disease. *Cell*, *170*(1), 17–33. https://doi.org/10.1016/j.cell.2017.06.009.RAS

Sirajuddin, M., Ali, S., & Badshah, A. (2013). Journal of Photochemistry and Photobiology B : Biology Drug – DNA interactions and their study by UV – Visible , fluorescence spectroscopies and cyclic voltametry. *Journal of Photochemistry & Photobiology, B: Biology*, *124*, 1–19. https://doi.org/10.1016/j.jphotobiol.2013.03.013

Sizochenko, N., Majumdar, D., & Roszak, S. (2017). Application of Quantum Mechanics and Molecular Mechanics in Chemoinformatics. In M. K. S. Jerzy Leszczynski, Anna Kaczmarek-Kedziera, Tomasz Puzyn, Manthos G. Papadopoulos, Heribert Reis (Ed.), *Handbook of Computational Chemistry* (2nd Editio, pp. 2041–2063). Springer International Publishing. https://doi.org/10.1007/978-3-319-27282-5

Smith, C. K., Brannigan, J. A., & Moore, M. H. (1996). Factors affecting DNA sequence selectivity of nogalamycin intercalation: The crystal structure of d(TGTACA)2-nogalamycin2. *Journal of Molecular Biology*, *263*(2), 237–258.

https://doi.org/10.1006/jmbi.1996.0572

Soo, V., Kwan, B., Quezada, H., Castillo-Juárez, I., Pérez-Eretza, B., García-Contreras, S., Martínez-Vázquez, M., Wood, T., & García-Contreras, R. (2016). Repurposing of Anticancer Drugs for the Treatment of Bacterial Infections. *Current Topics in Medicinal Chemistry*, *17*(10), 1157–1176. https://doi.org/10.2174/1568026616666160930131737

Sun, D., & Hurley, L. H. (1995). TBP binding to the TATA box induces a specific downstream unwinding site that is targeted by pluramycin. *Chemistry and Biology*, *2*(7), 457–469. https://doi.org/10.1016/1074-5521(95)90263-5

Tanchuk, V. Y., Tanin, V. O., Vovk, A. I., & Poda, G. (2016). A New, Improved Hybrid Scoring Function for Molecular Docking and Scoring Based on AutoDock and AutoDock Vina. *Chemical Biology and Drug Design*, 87(4), 618–625. https://doi.org/10.1111/cbdd.12697

Travers,A.,Muskhelishvili,G.(2015).DNA Structure and Function.*The FEBS Journal*. 282, 2279–2295. doi:10.1111/febs.13307

Veryazov, V. (2016). Hartree-Fock Theory. In and P.-O. W. Björn O. Roos, Roland Lindh, Per Åke Malmqvist, Valera Veryazov (Ed.), Multiconfigurational Quantum Chemistry (First Edit, pp. 44–58). John Wiley & Sons, Inc.

Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z. H., & Hou, T. (2019). End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design [Review-article]. *Chemical Reviews*, *119*(16), 9478–9508. https://doi.org/10.1021/acs.chemrev.9b00055

Wang, M., Yu, Y., Liang, C., Lu, A., & Zhang, G. (2016). Recent advances in developing small molecules targeting nucleic acid. *International Journal of Molecular Sciences*, *17*(6). https://doi.org/10.3390/ijms17060779

Wu, J., Ling, X., Pan, D., Apontes, P., Song, L., Liang, P., Altieri, D. C., Beerman, T., & Li, F. (2005). Molecular mechanism of inhibition of survivin transcription by the GC-rich sequence-selective DNA binding antitumor agent, hedamycin: Evidence of survivin down-regulation associated with drug sensitivity. *Journal*

of Biological Chemistry, *280*(10), 9745–9751. https://doi.org/10.1074/jbc.M409350200

Xu, D., Landon, T., Greenbaum, N. L., & Fenley, M. O. (2007). The electrostatic characteristics of G . U wobble base pairs. *Nucleic Acid Research. 35*(11), 3836–3847. https://doi.org/10.1093/nar/gkm274

Yang, F., Teves, S. S., Kemp, C. J., & Henikoff, S. (2013). Doxorubicin, DNA torsion, and chromatin dynamics. *BBA - Reviews on Cancer*, 1–6. https://doi.org/10.1016/j.bbcan.2013.12.002

APPENDICES

A. Energy Score Distribution Plot of Each Drug/DNA Pair

Table 2. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of altromycin B-AAAAAAAAAA.

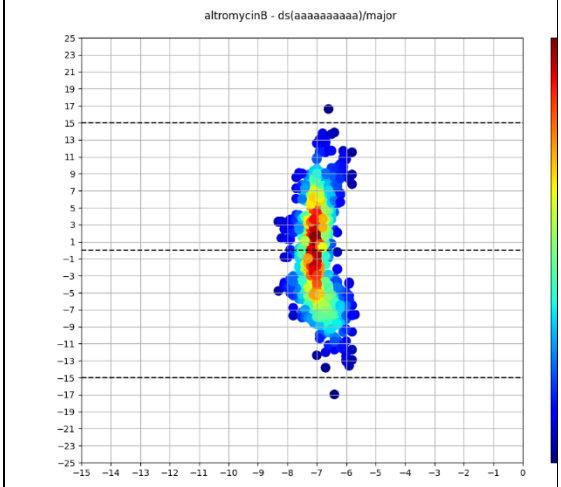
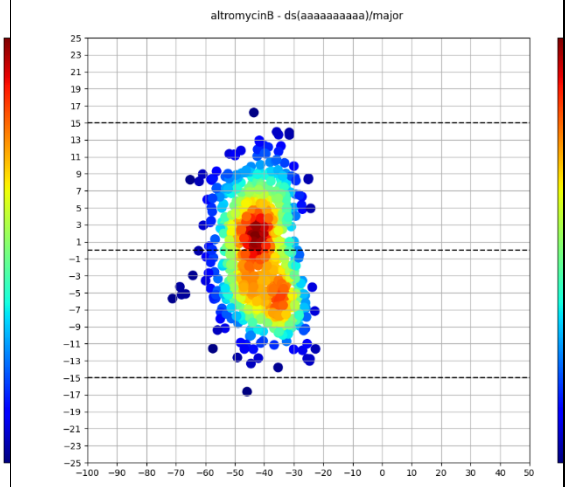
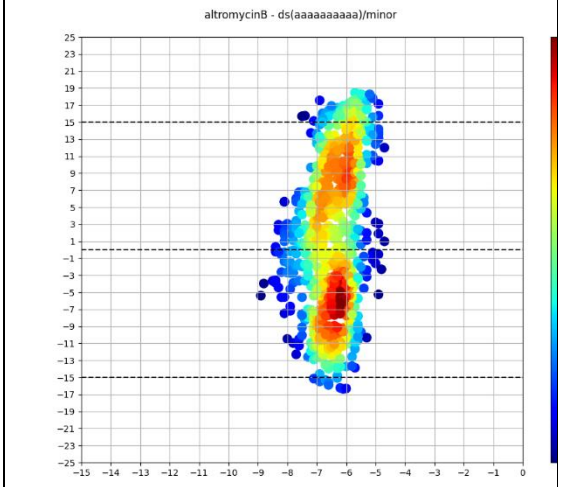
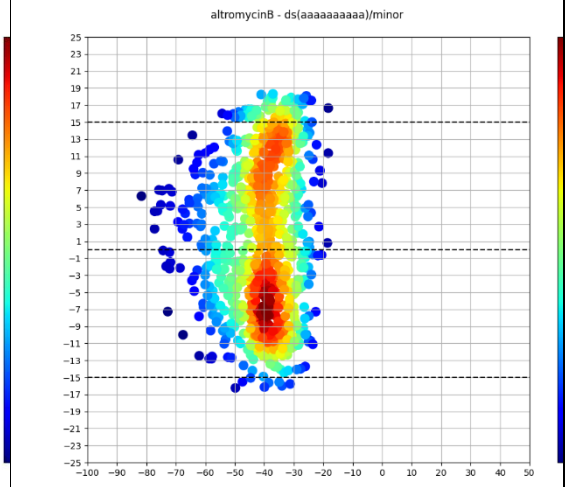
Autodock Vina Score Distribution	MM-Refinement Distribution
	
	

Table 3. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of altromycin B-ATCGCGCGAT.

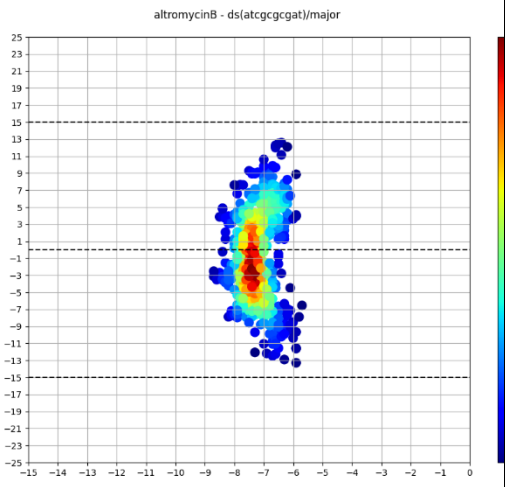
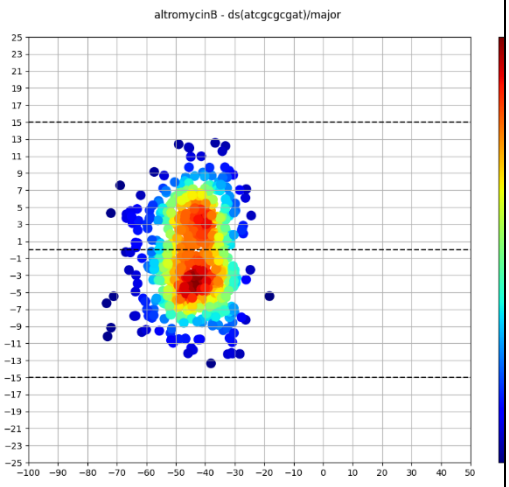
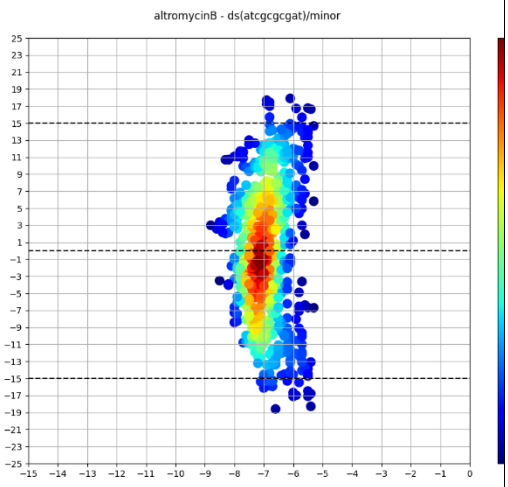
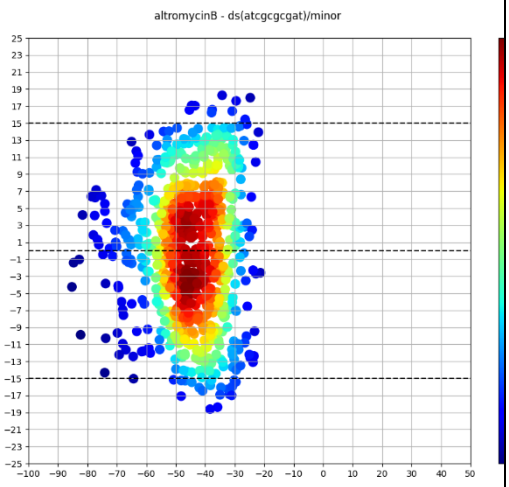
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 4. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of altromycin B-CGTATATACG .

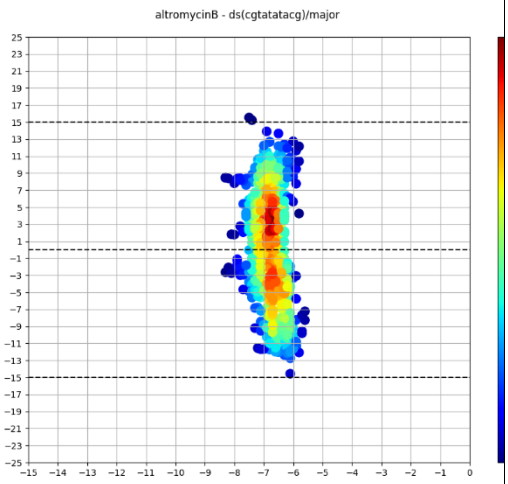
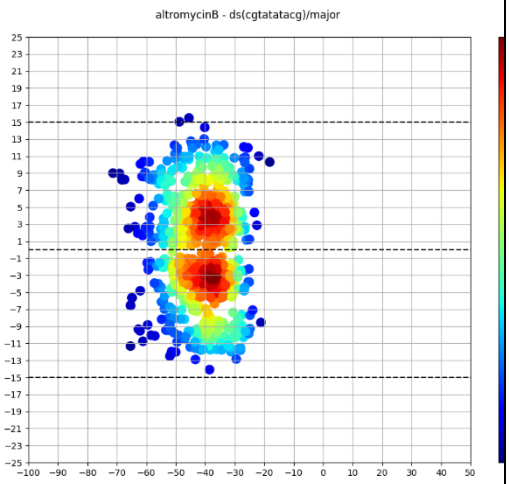
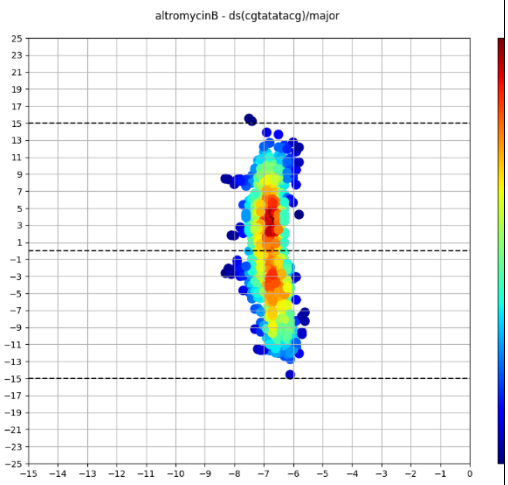
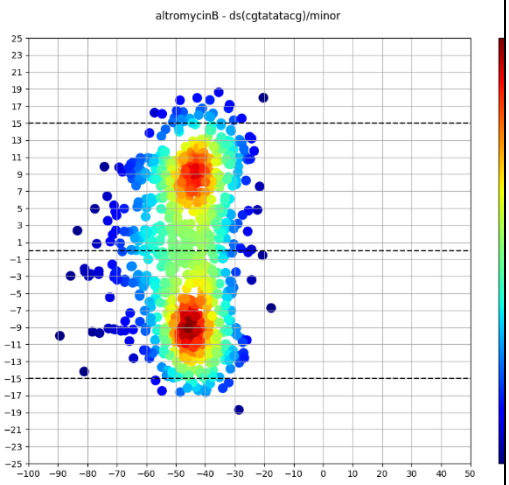
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 5. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of altromycin B-GGCCAATTGG .

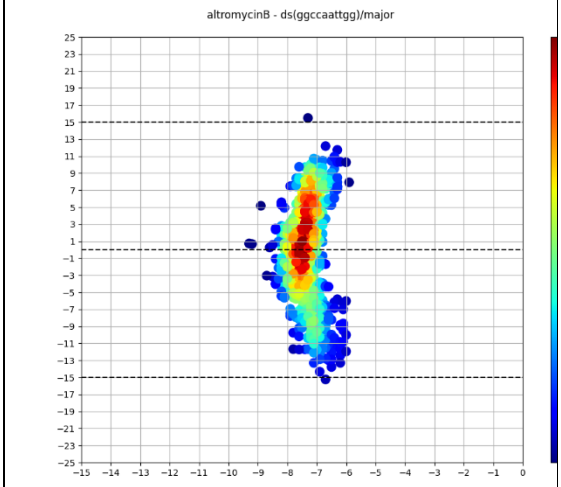
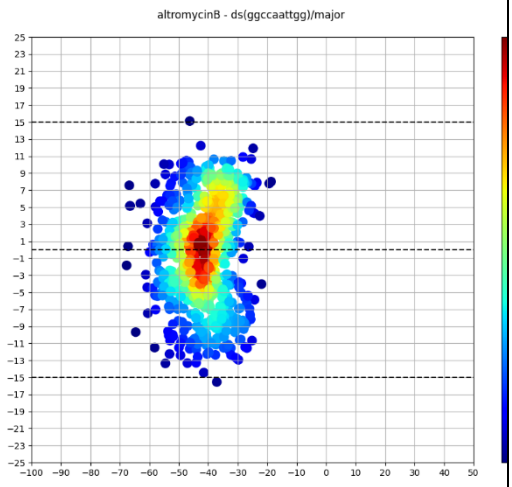
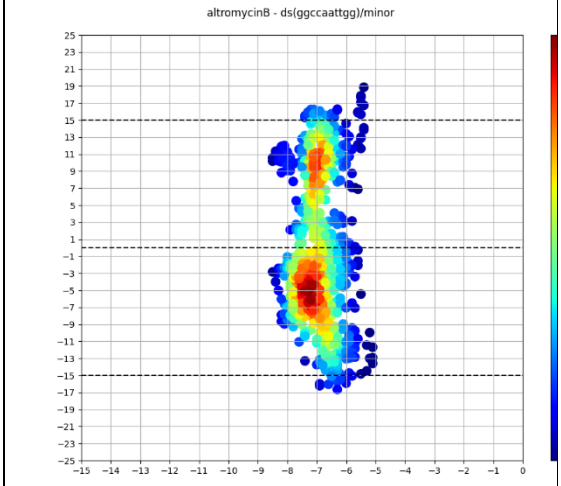
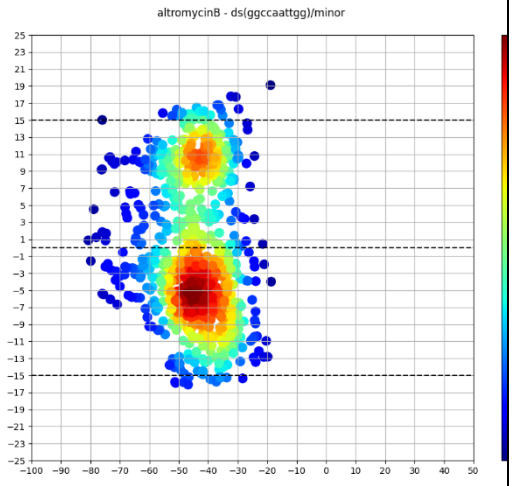
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 6. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of altromycin B-GGGGGGGGGG.

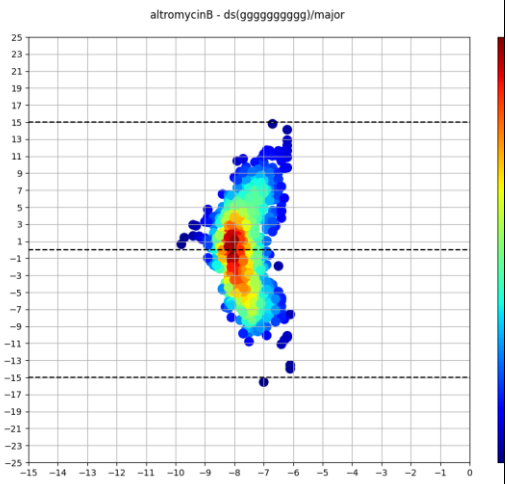
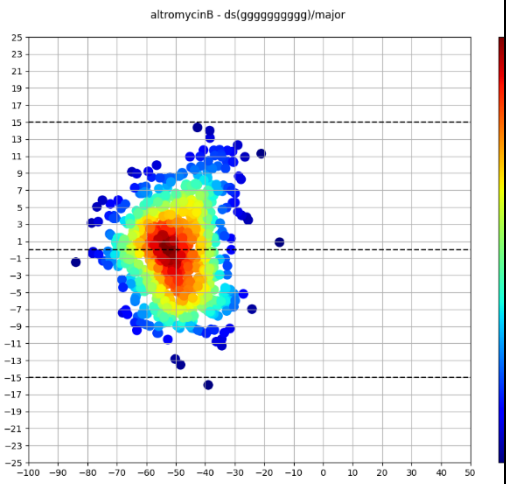
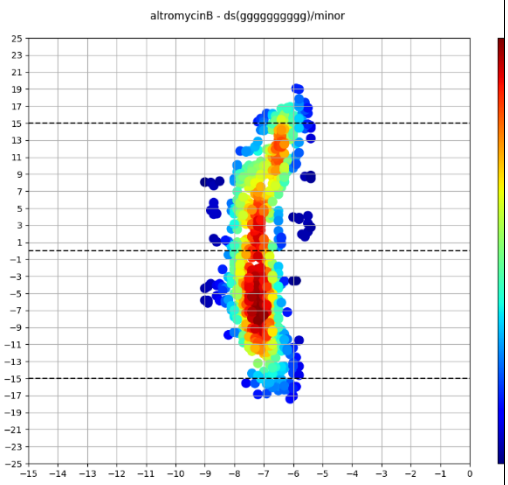
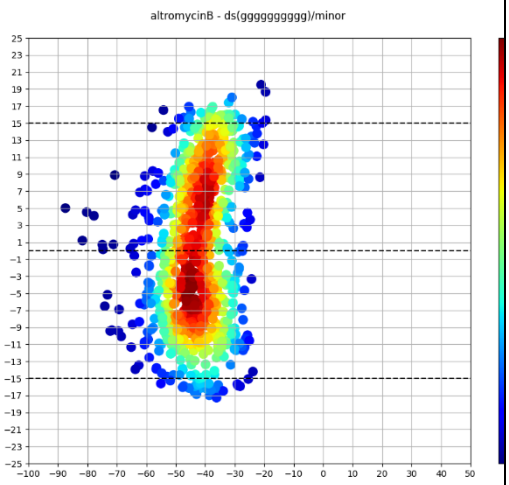
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 7. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of berenil-AAAAAAAAAA.

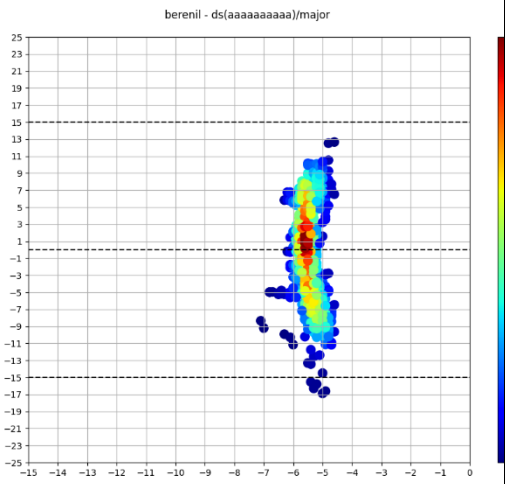
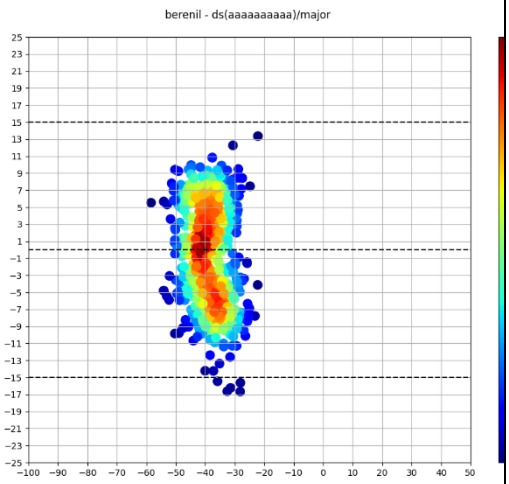
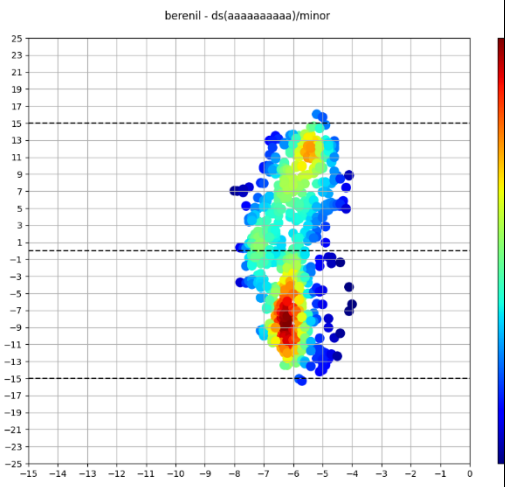
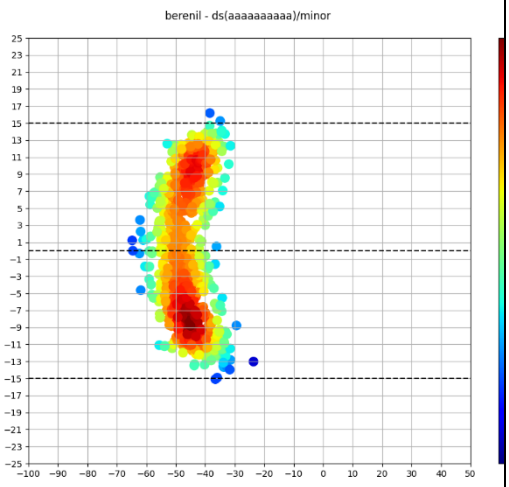
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 8. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of berenil-ATCGCGCGAT.

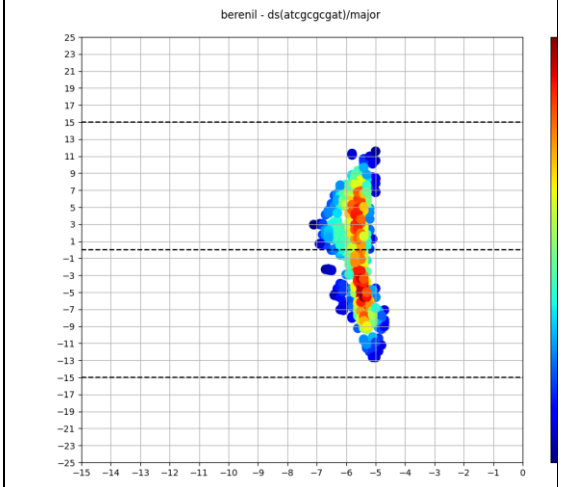
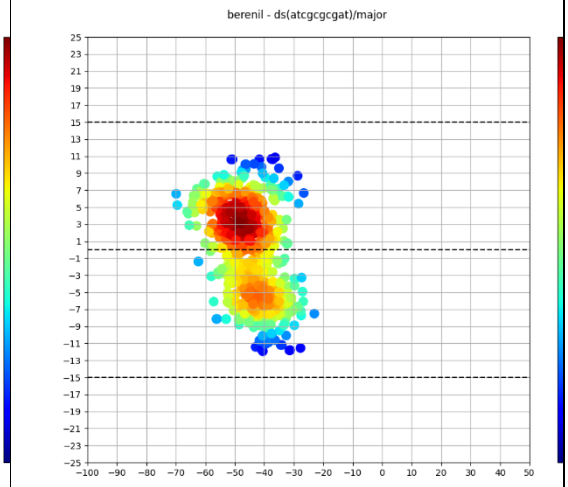
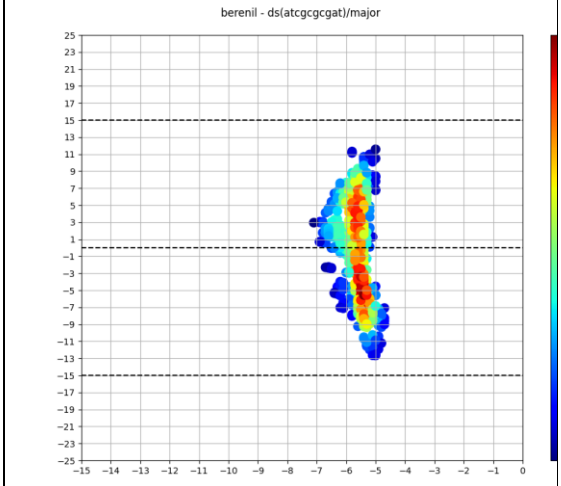
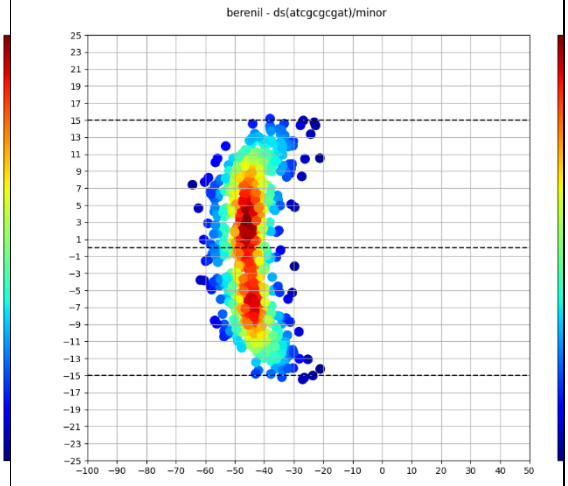
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 9. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of berenil-CGTATATACG .

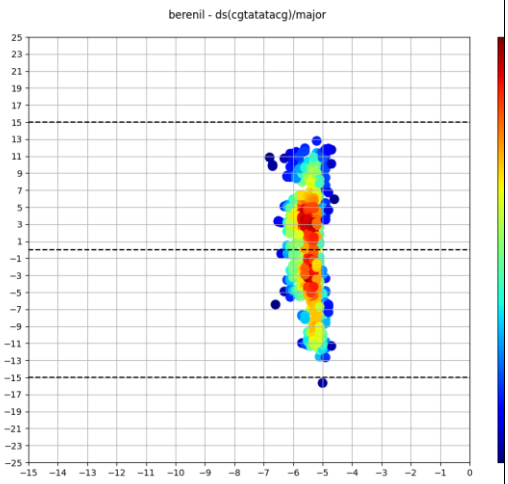
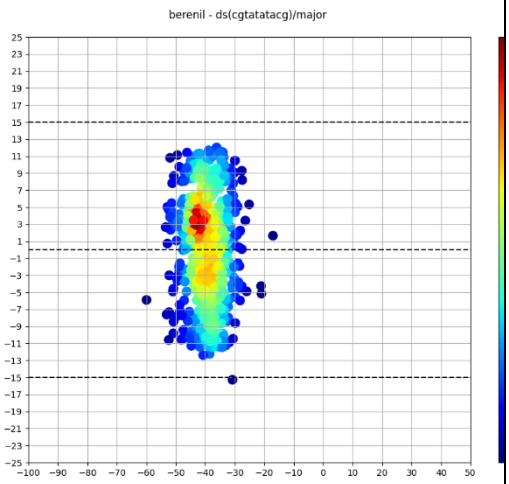
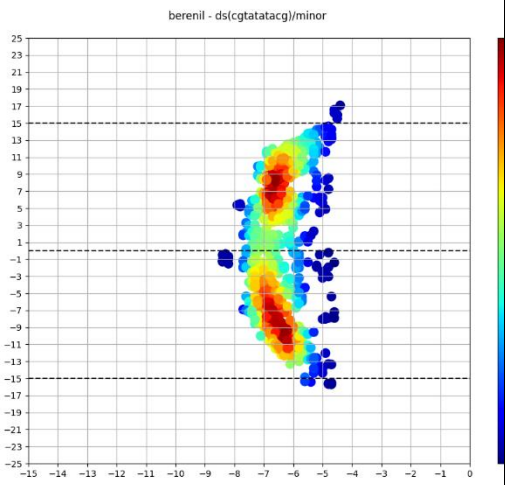
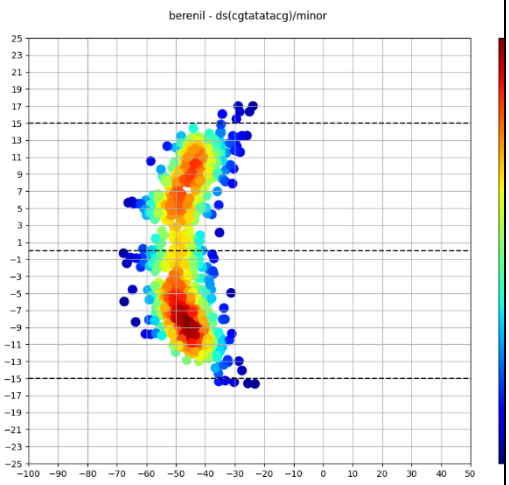
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 10. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of berenil-GGCCAATTGG .

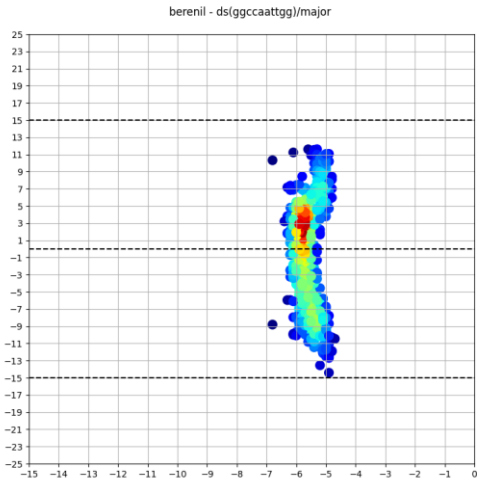
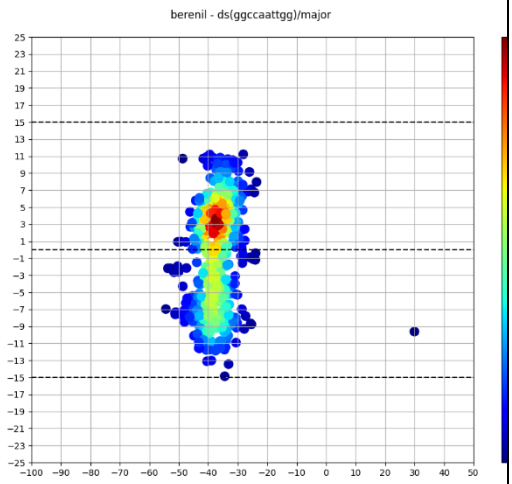
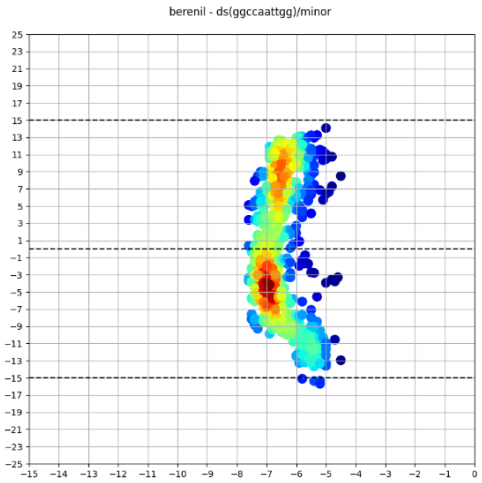
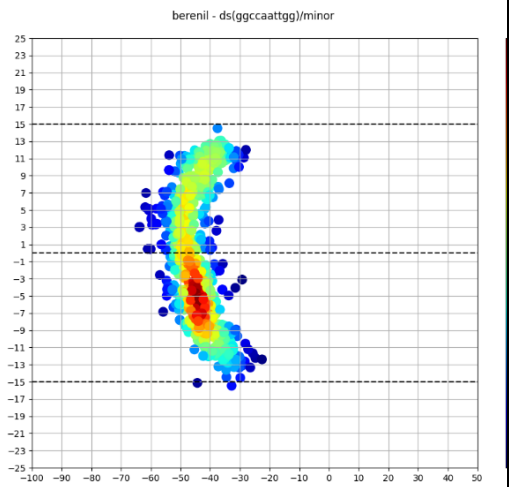
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 11. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of berenil-GGGGGGGGGG.

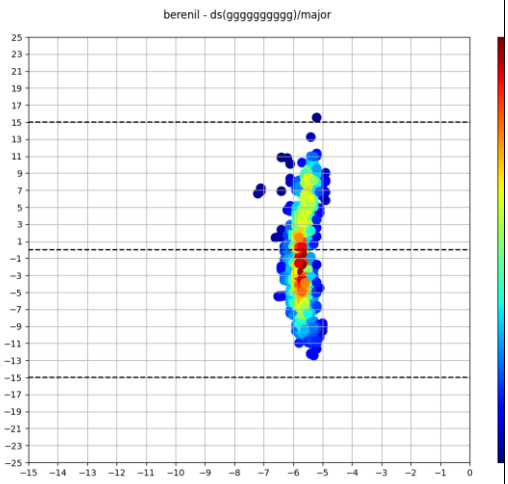
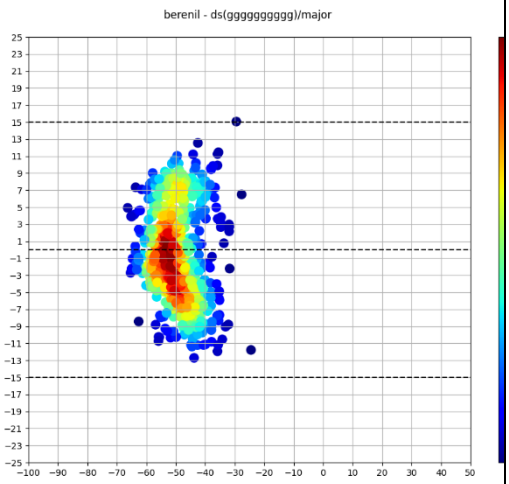
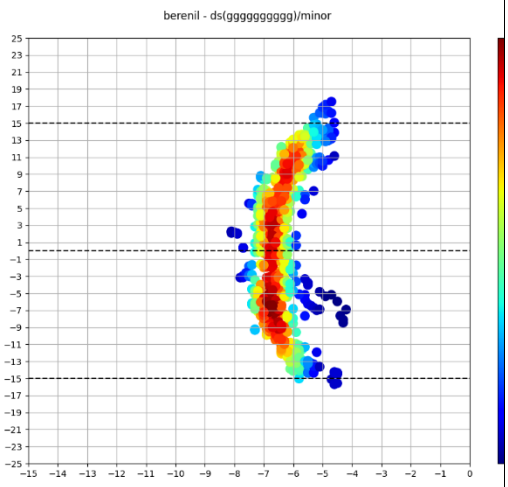
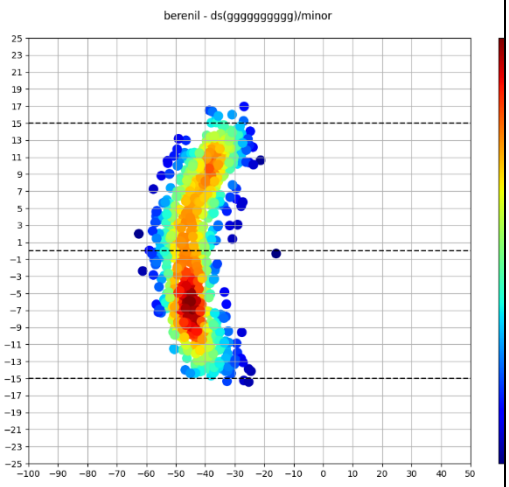
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 12. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of DAPI-AAAAAAAAAA.

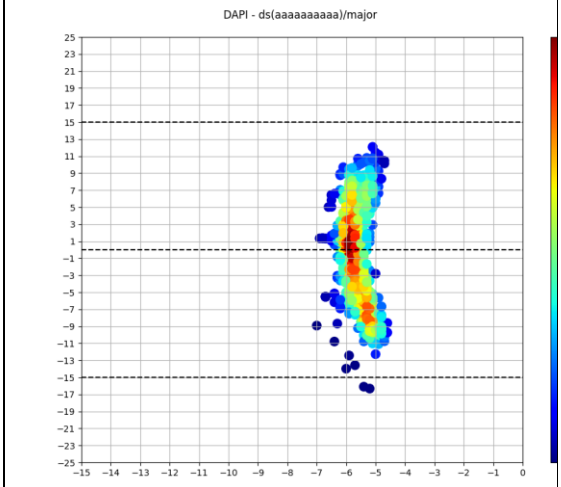
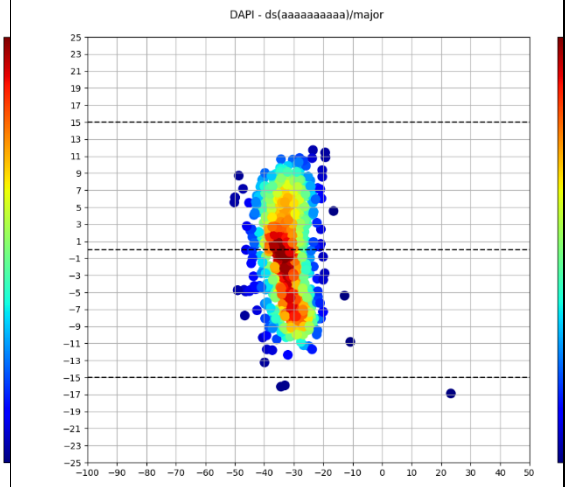
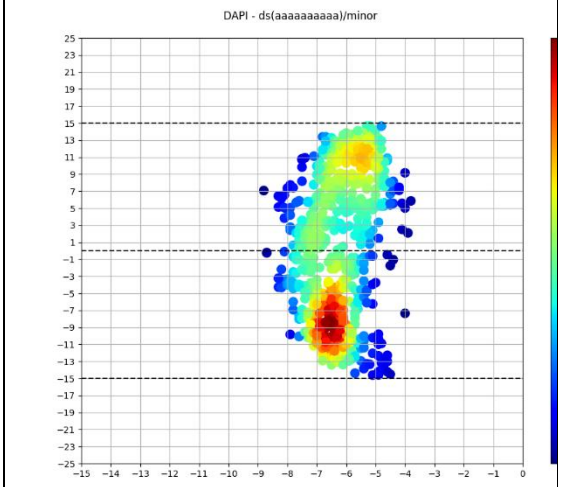
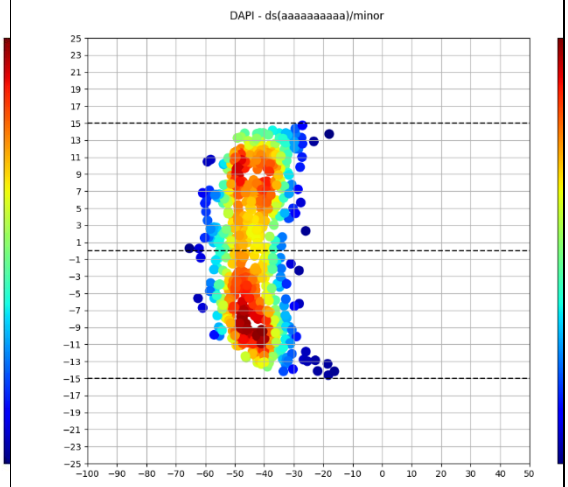
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 13. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of DAPI-ATCGCGCGAT.

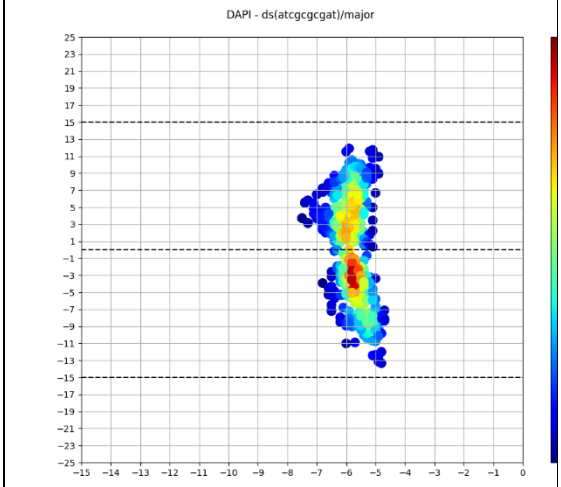
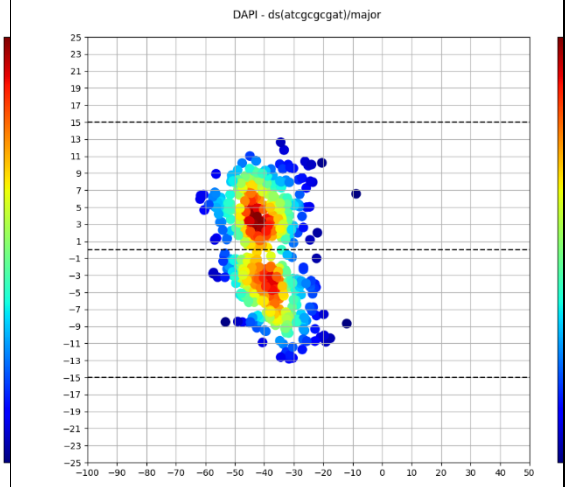
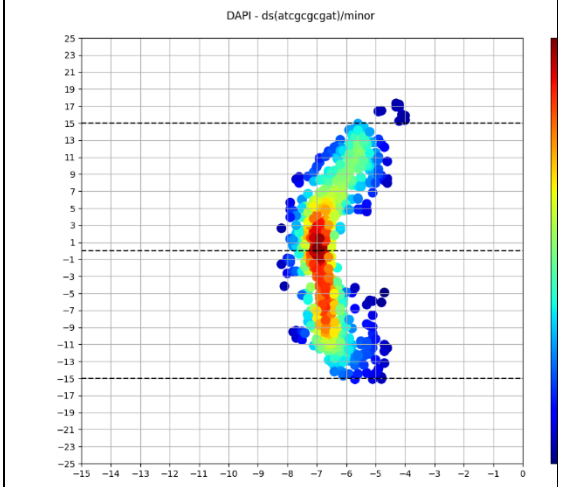
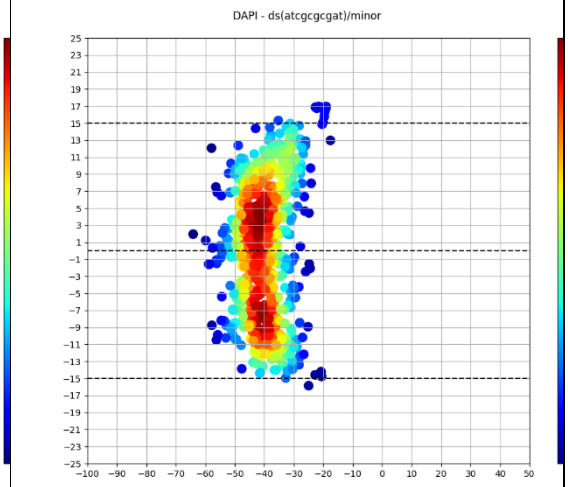
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 14. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of DAPI-CGTATATACG .

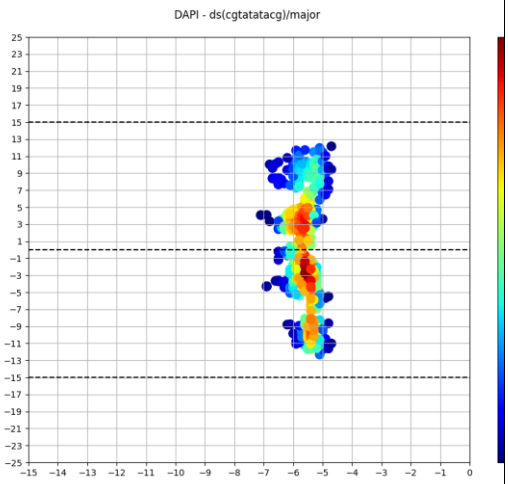
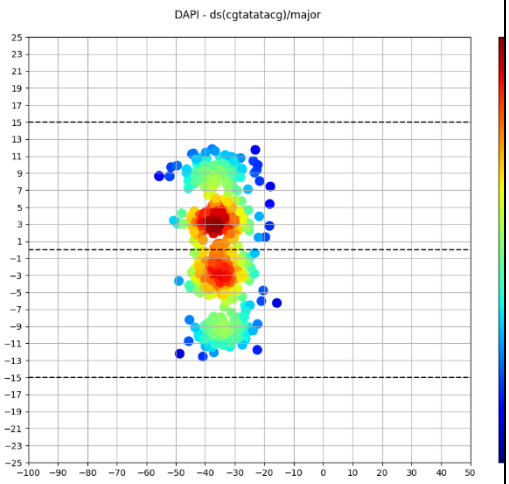
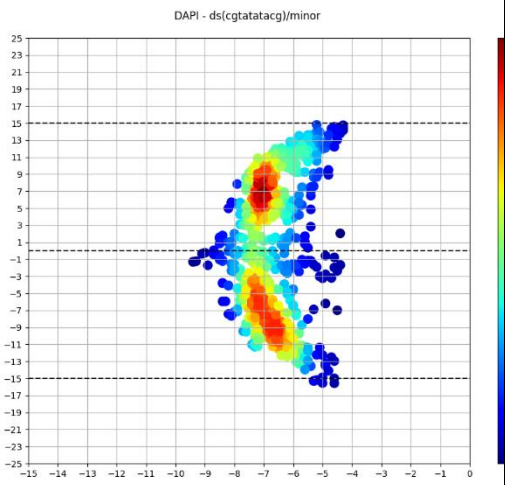
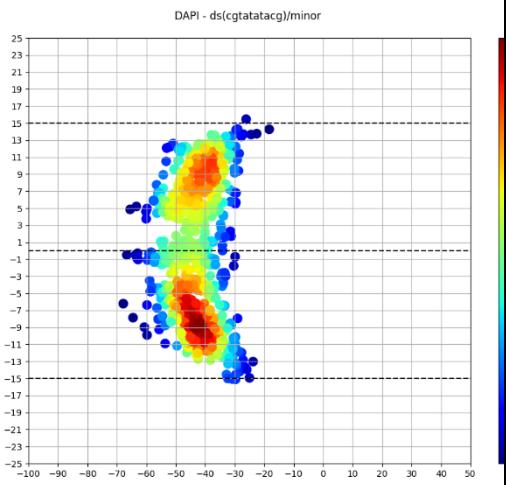
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 15. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of DAPI-GGCCAATTGG .

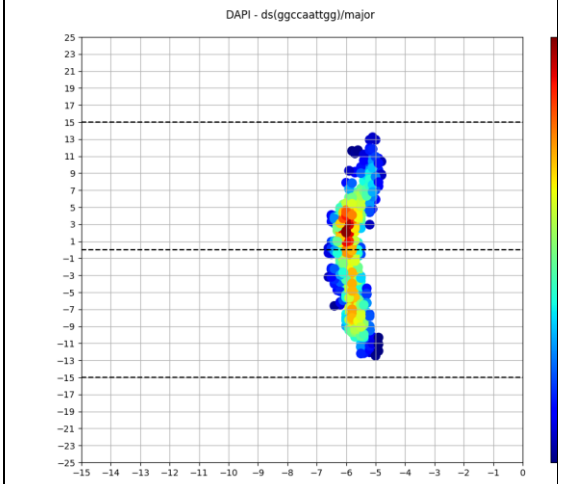
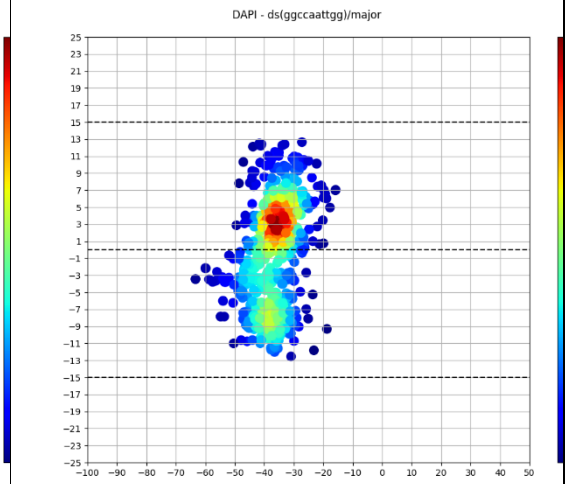
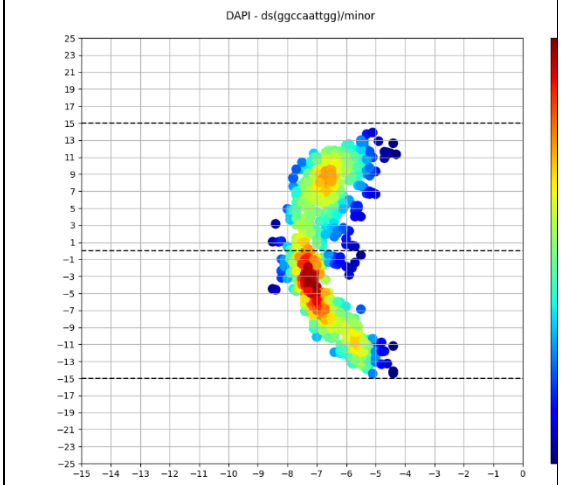
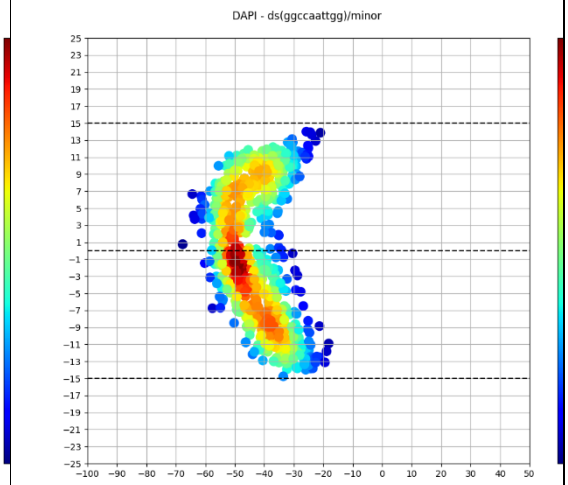
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 16. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of DAPI-GGGGGGGGGG.

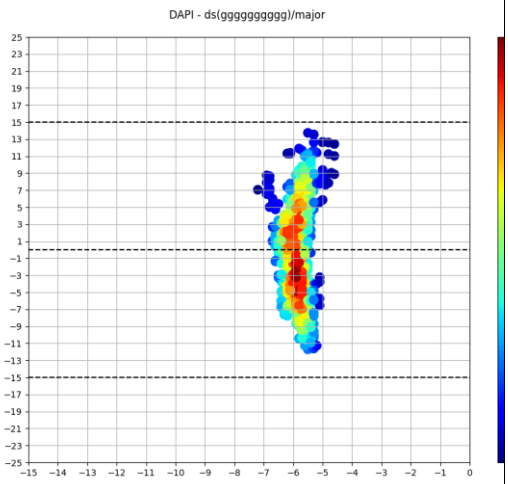
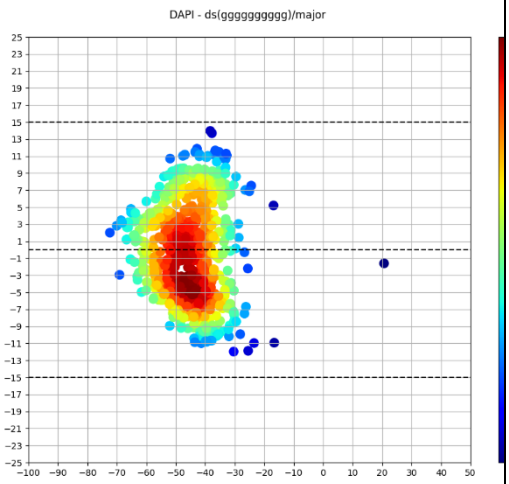
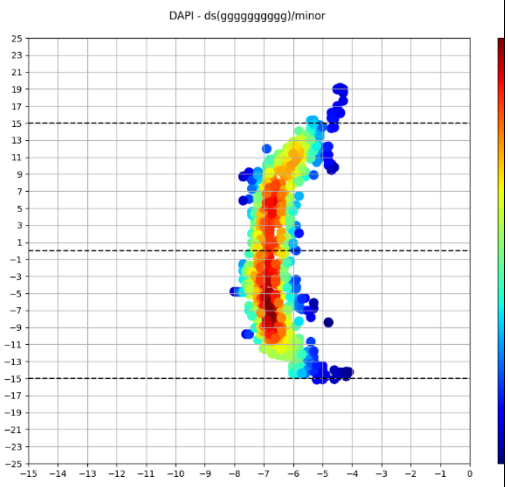
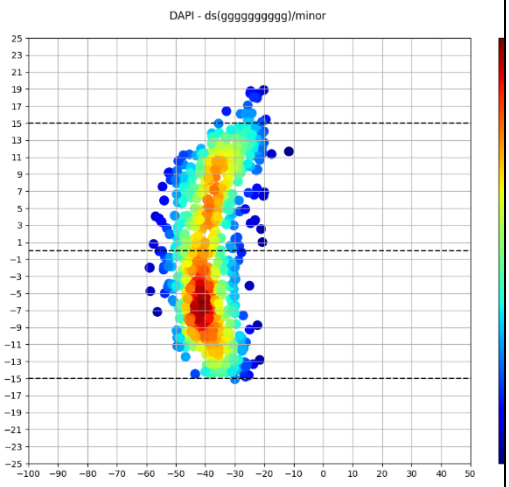
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 17. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of distamycin-AAAAAAAAAA.

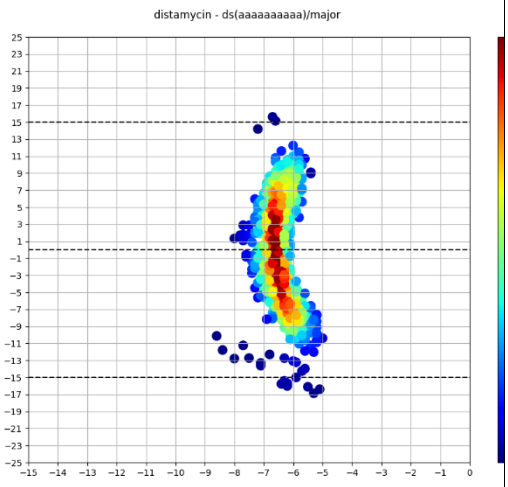
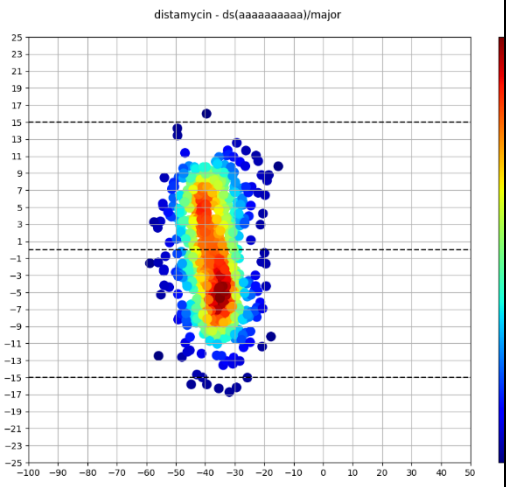
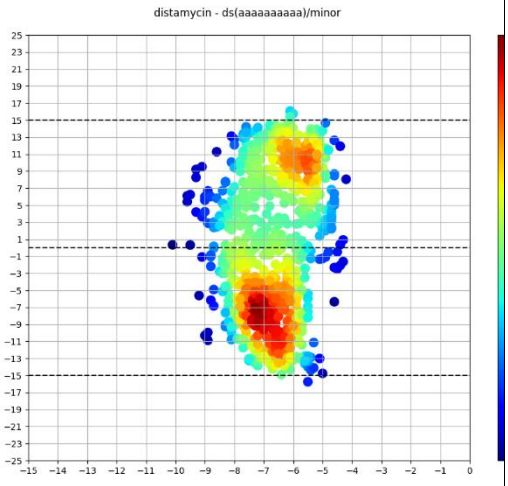
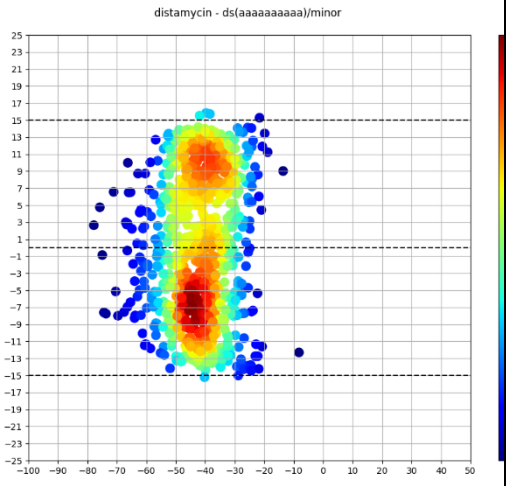
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 18. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of distamycin-ATCGCGCGAT.

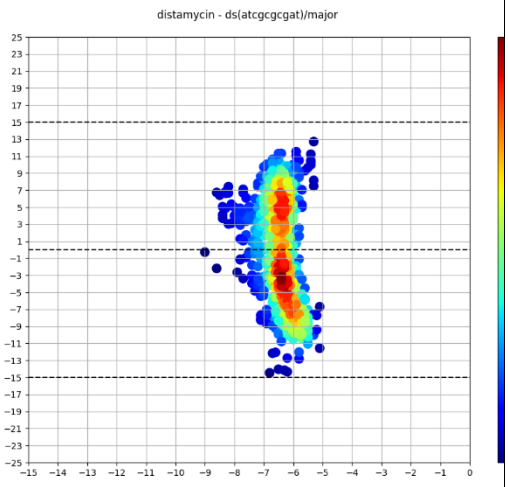
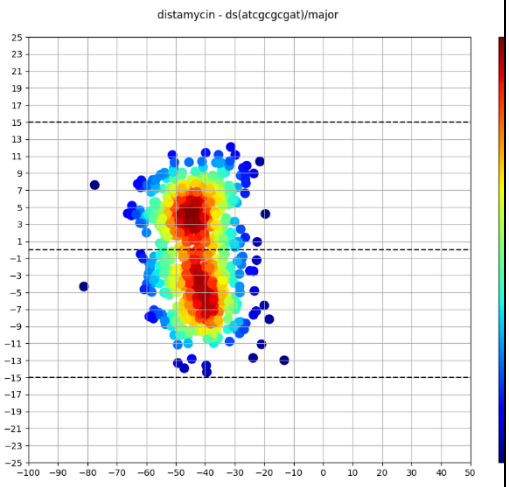
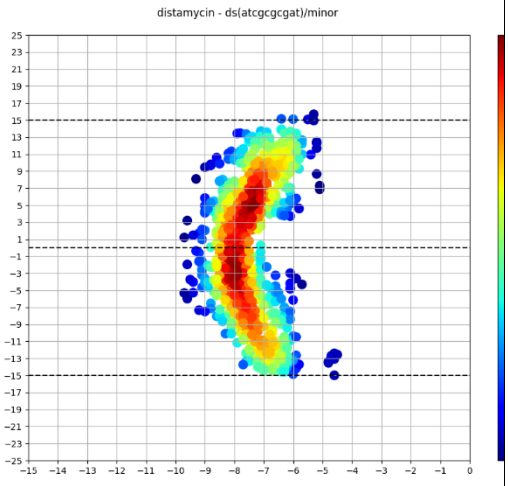
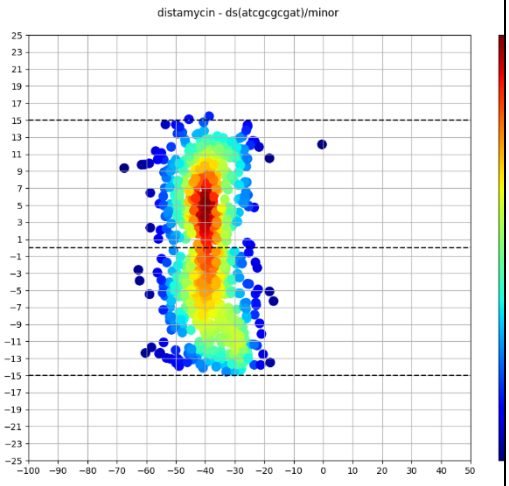
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 19. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of distamycin-CGTATATACG .

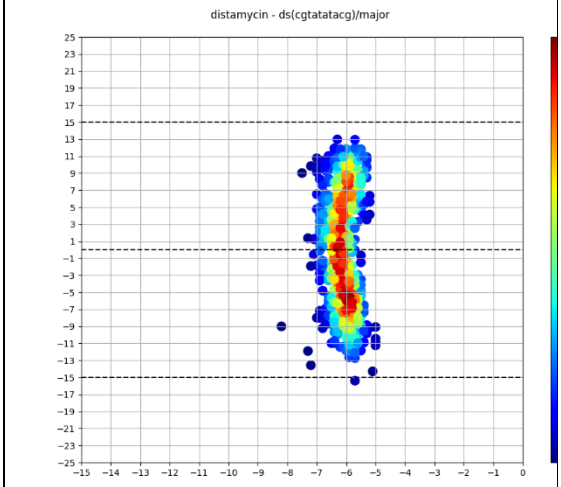
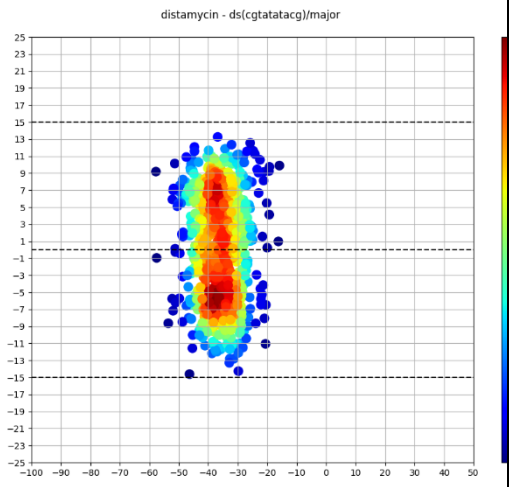
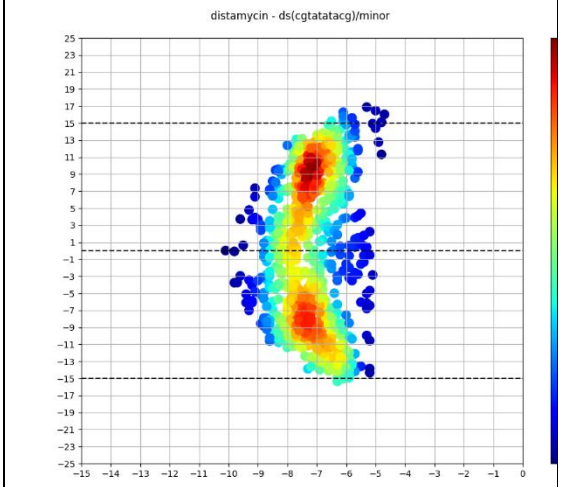
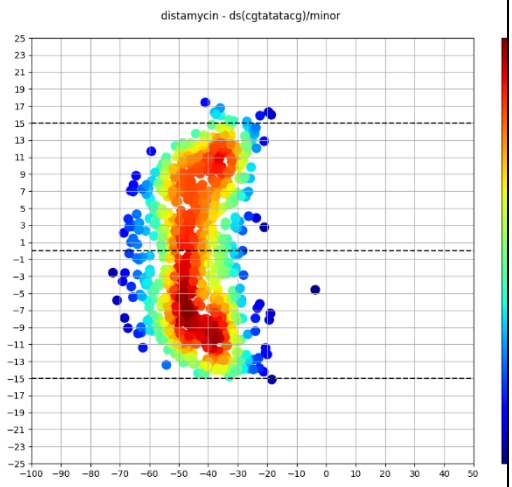
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 20. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of distamycin-GGCCAATTGG .

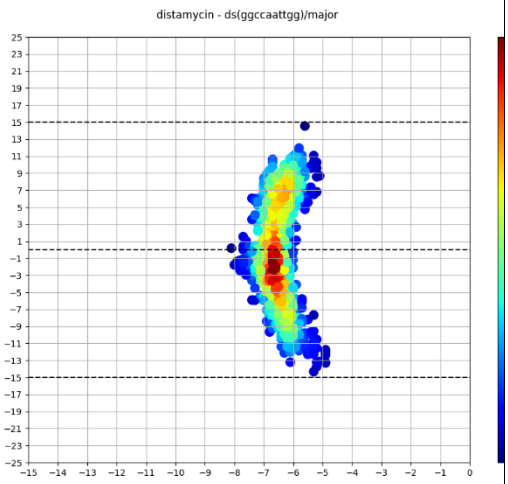
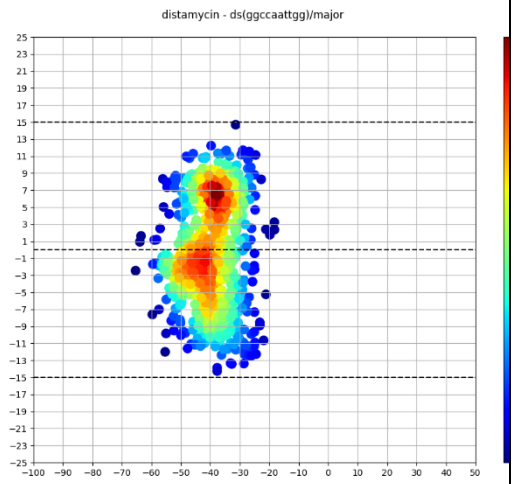
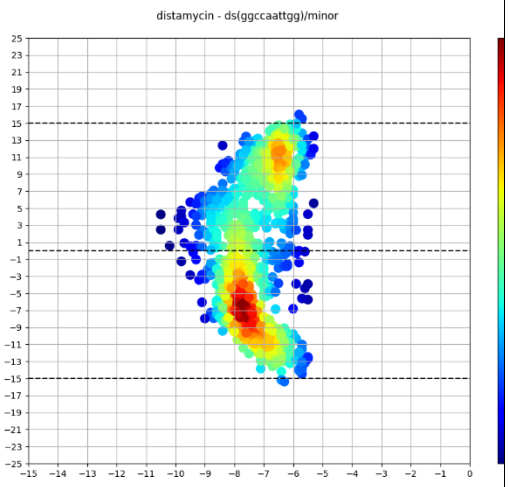
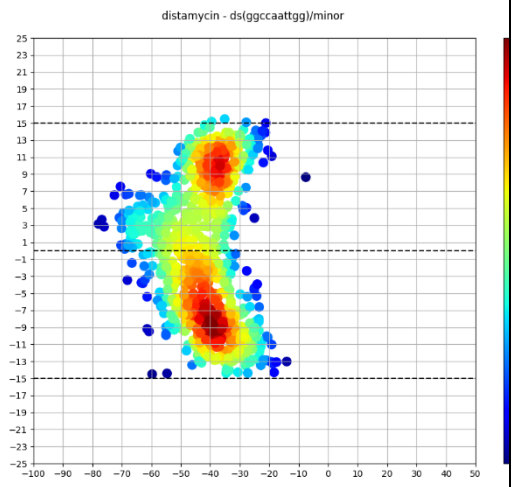
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 21. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of distamycin-GGGGGGGGGG.

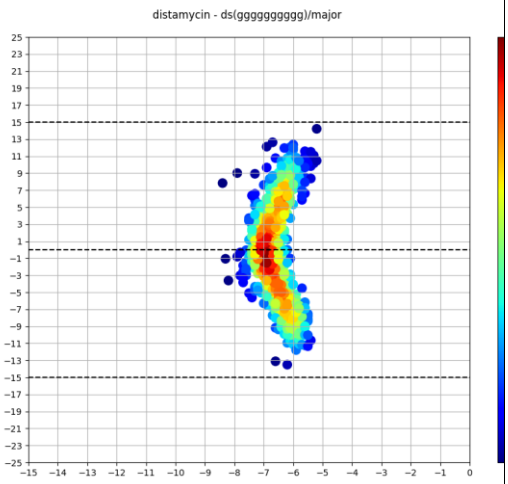
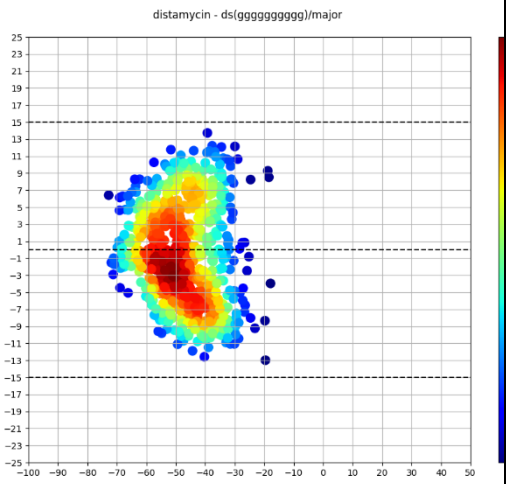
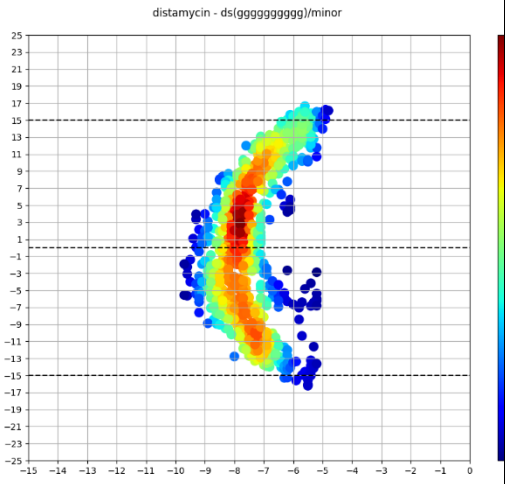
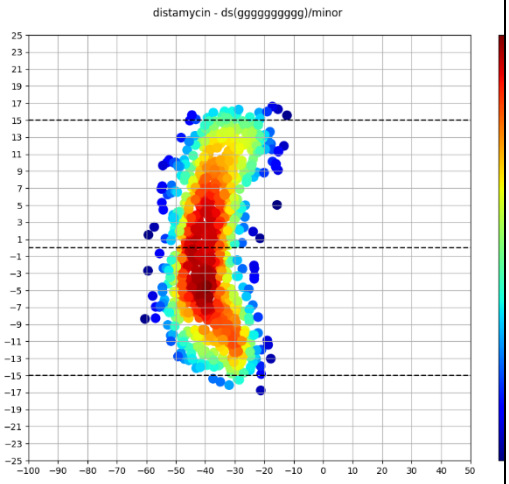
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 22. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of hedamycin-AAAAAAAAAA.

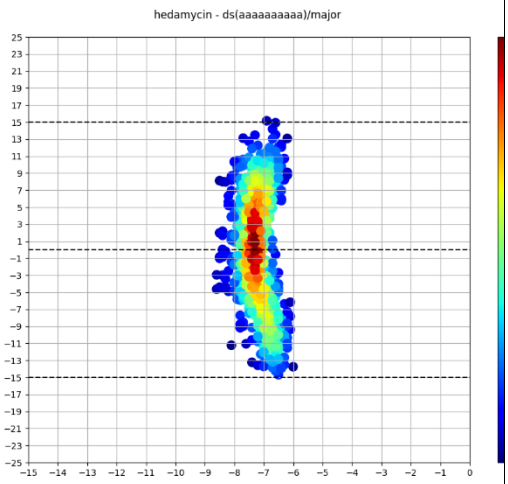
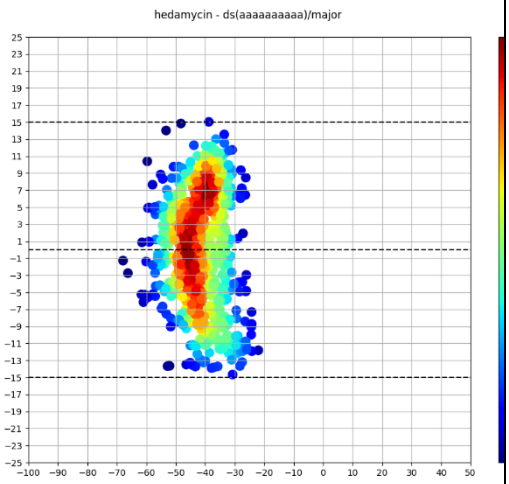
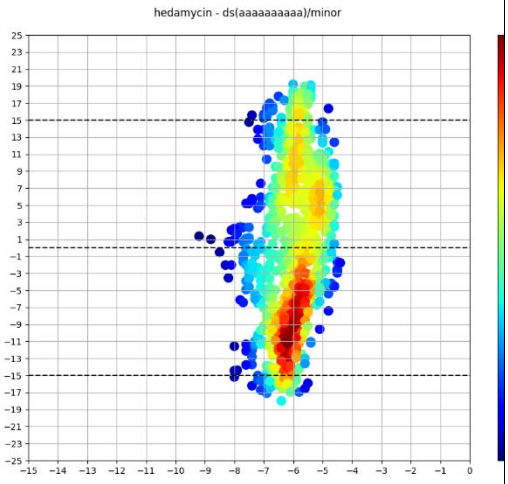
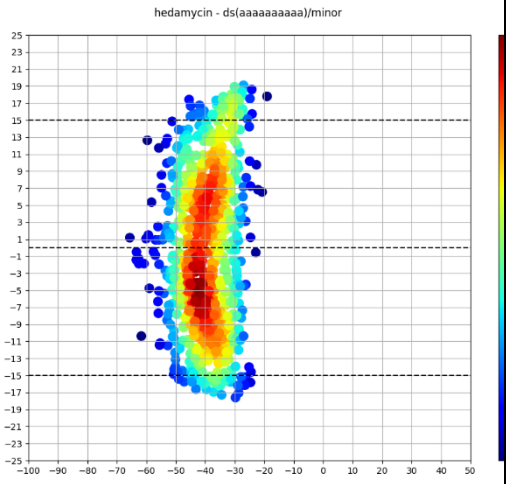
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 23. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of hedamycin-ATCGCGCGAT.

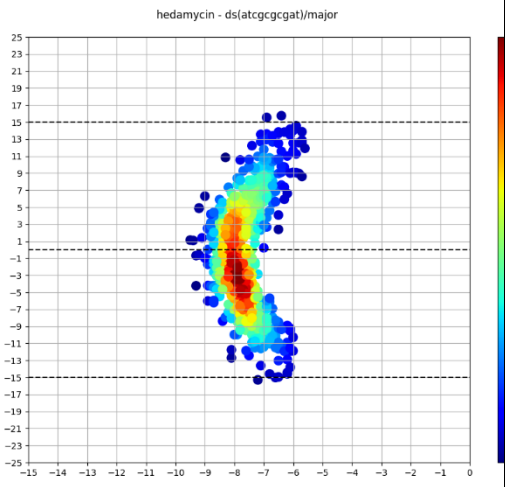
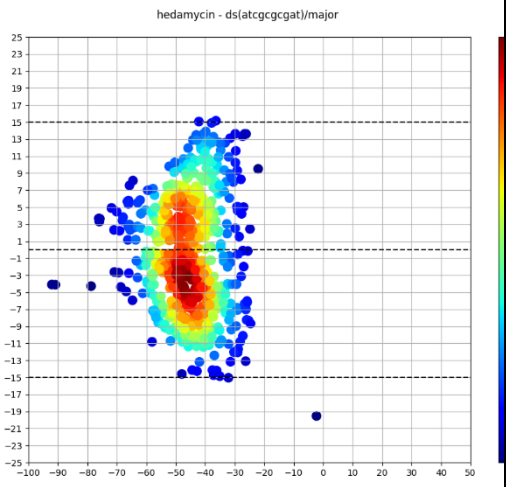
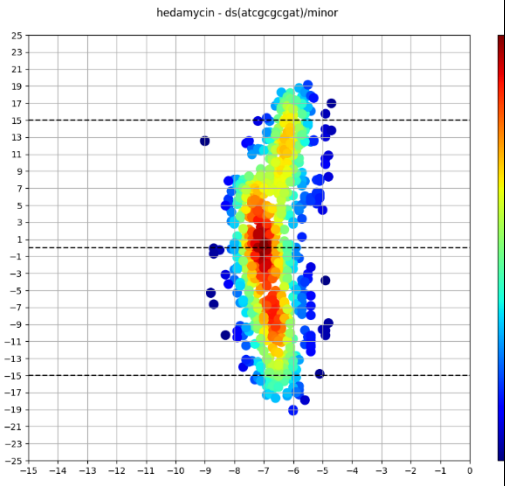
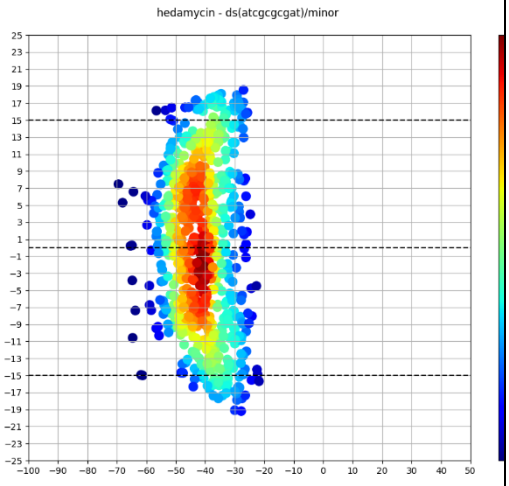
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 24. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of hedamycin-CGTATATACG .

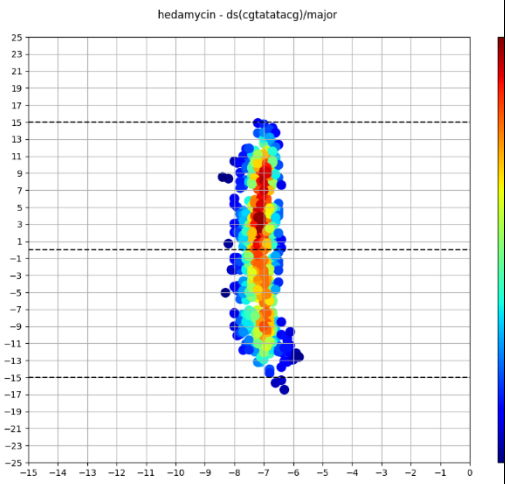
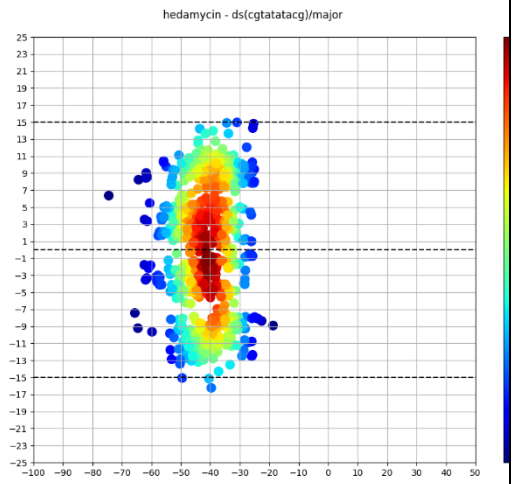
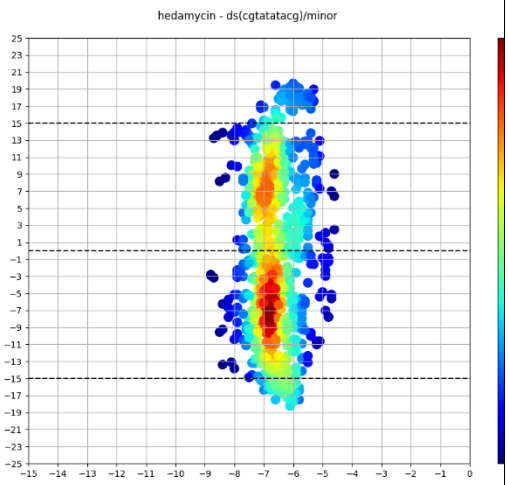
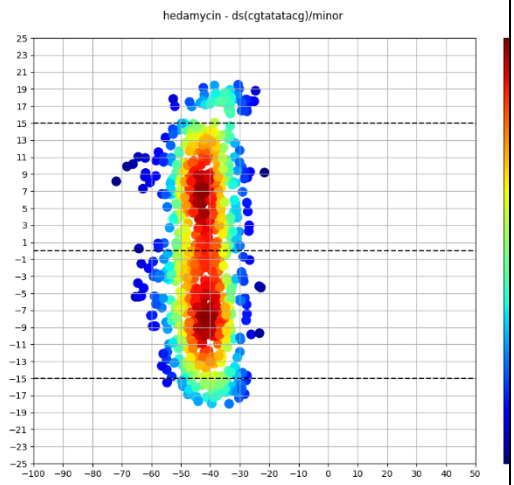
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 25. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of hedamycin-GGCCAATTGG .

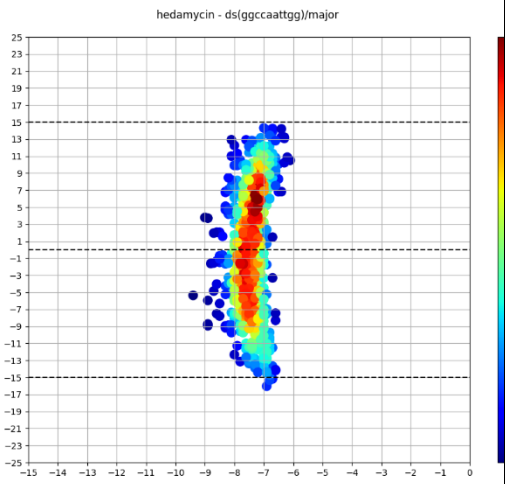
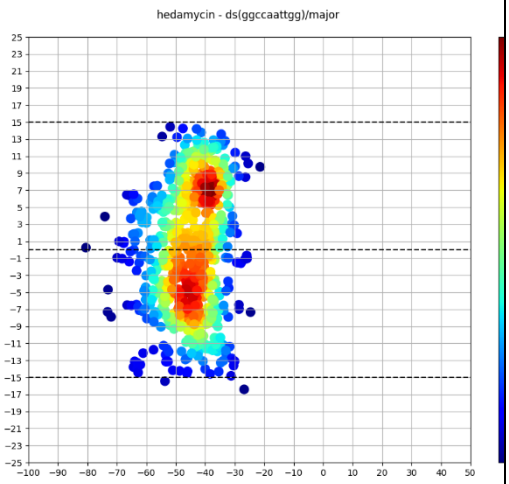
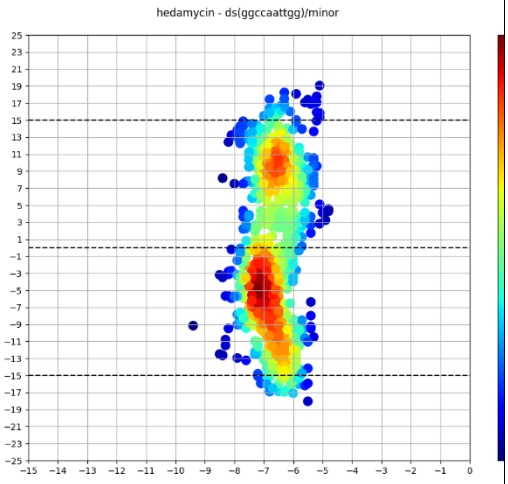
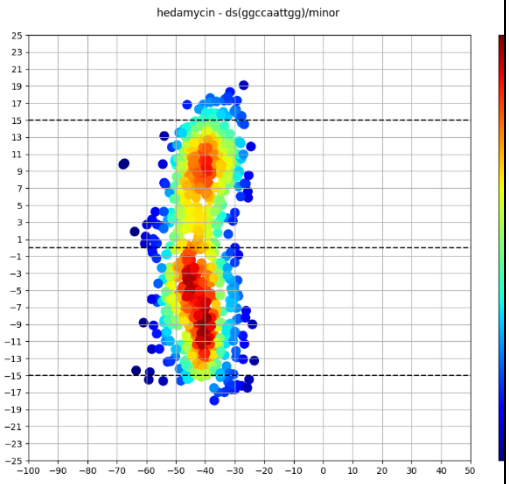
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 26. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of hedamycin-GGGGGGGGGG.

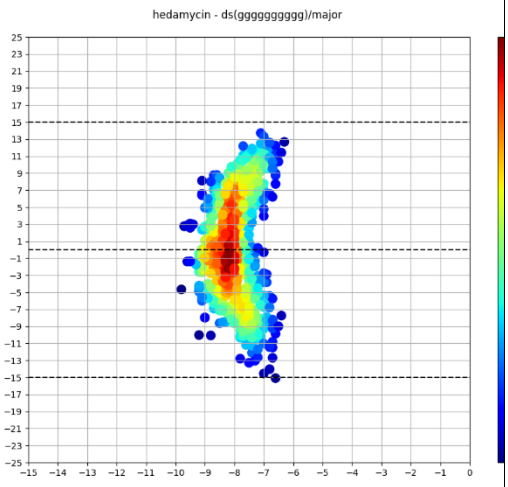
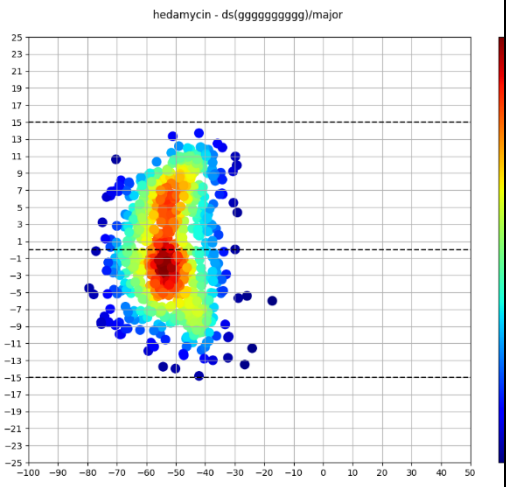
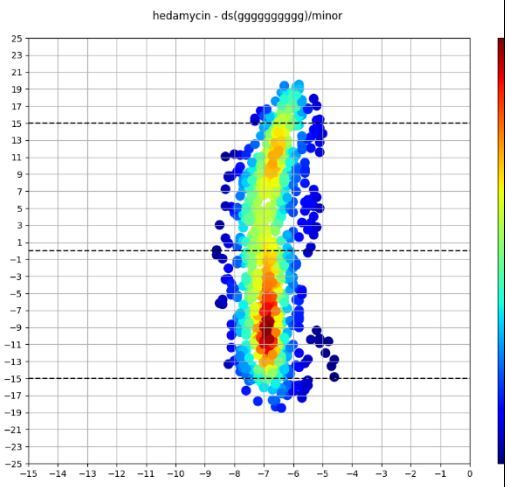
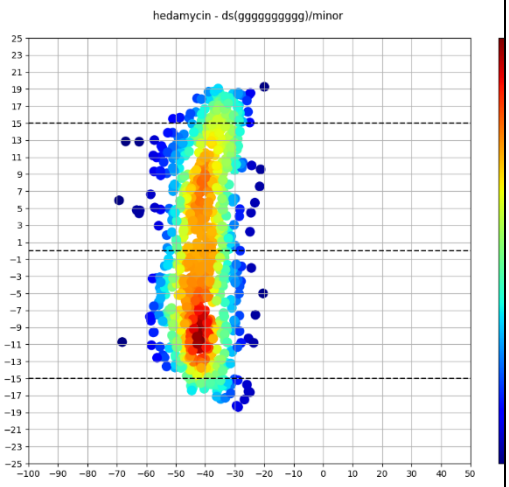
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 27. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of hoechst33258-AAAAAAAAAA.

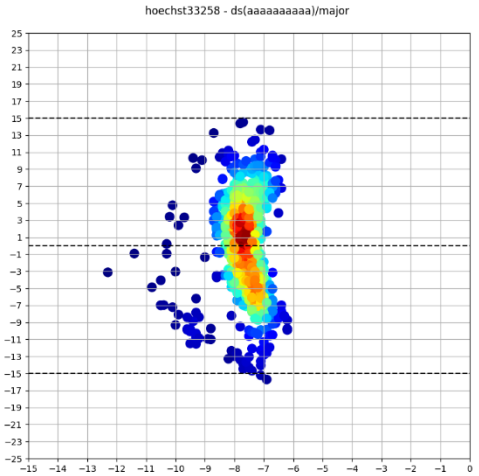
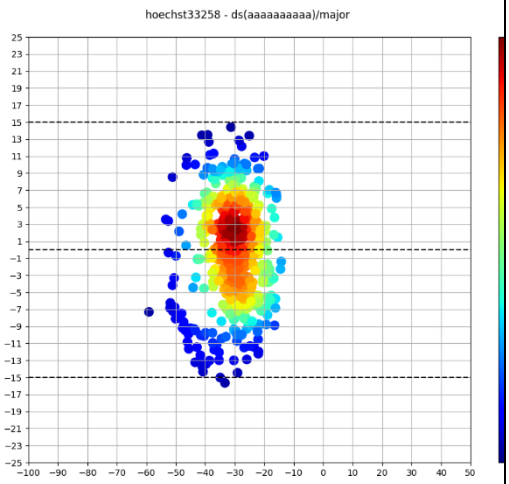
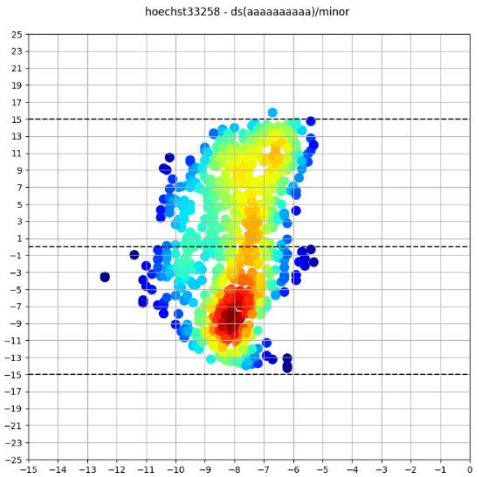
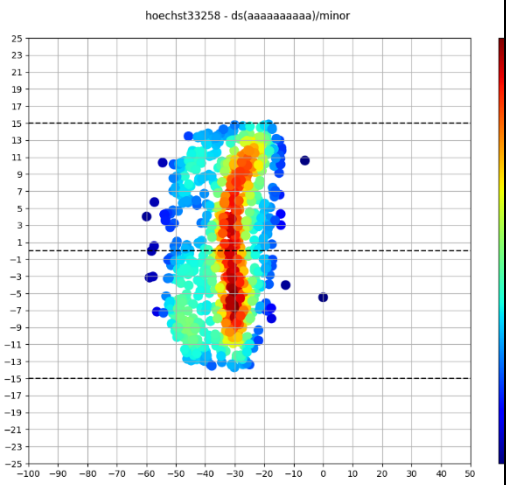
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 28. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of hoechst33258-ATCGCGCGAT.

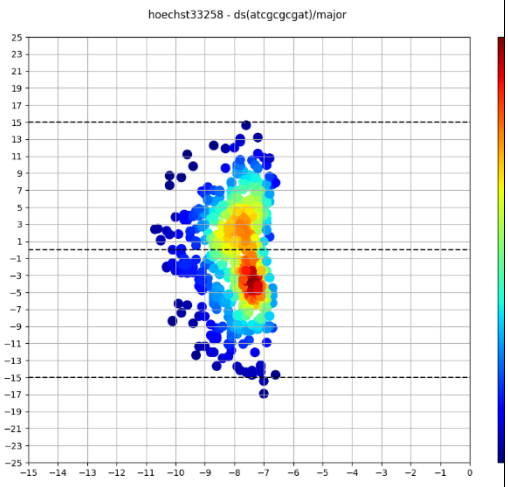
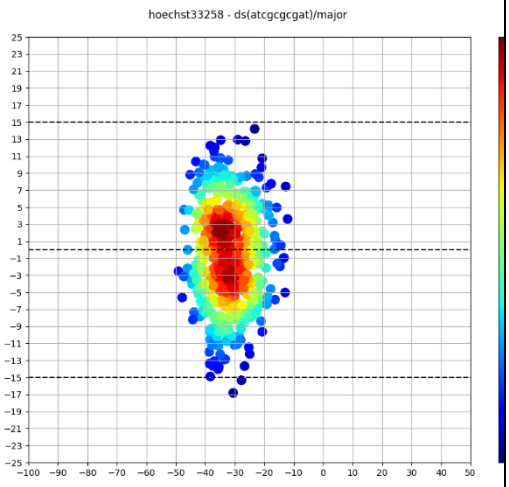
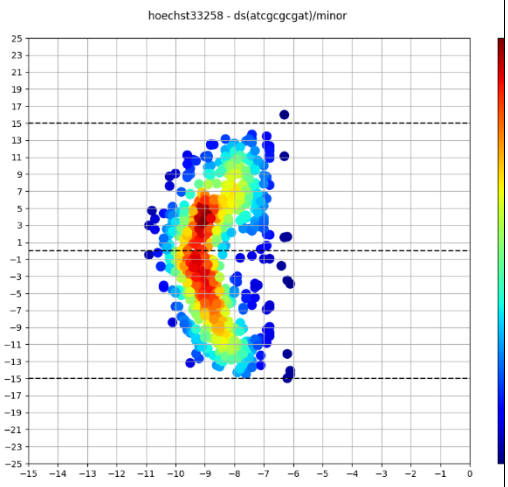
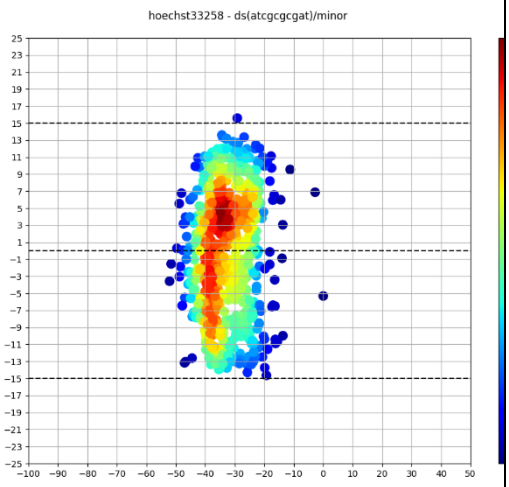
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 29. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of hoechst33258-CGTATATACG .

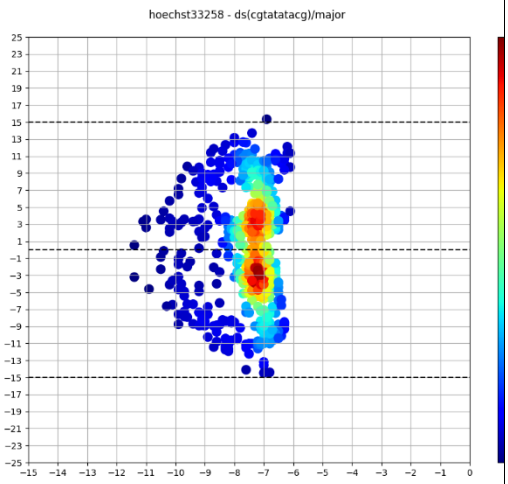
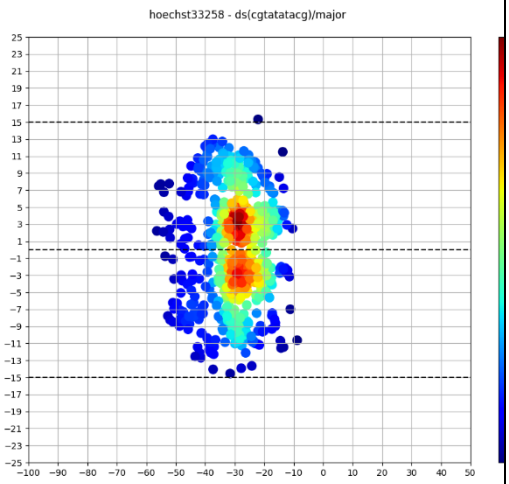
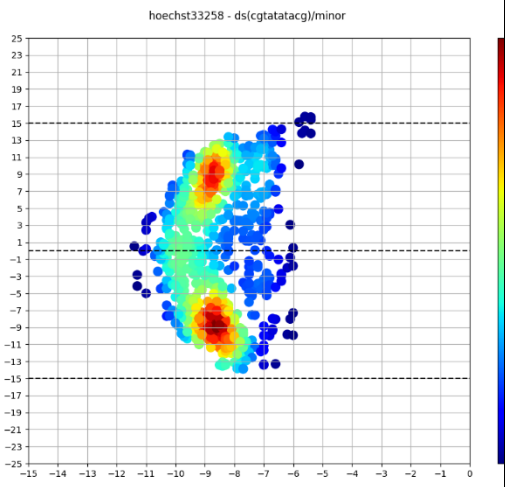
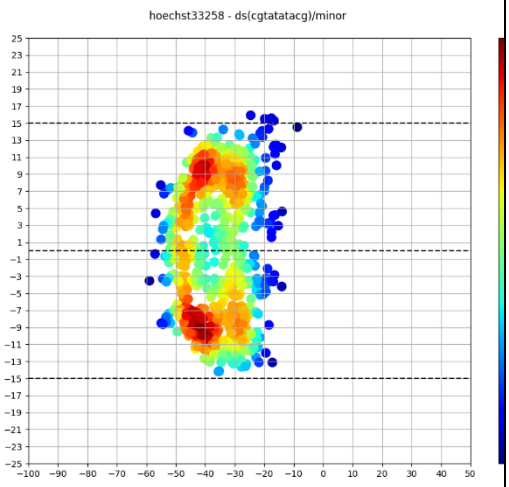
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 30. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of hoechst33258-GGCCAATTGG .

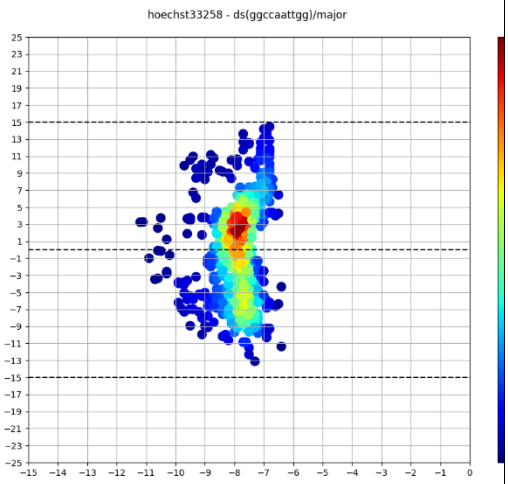
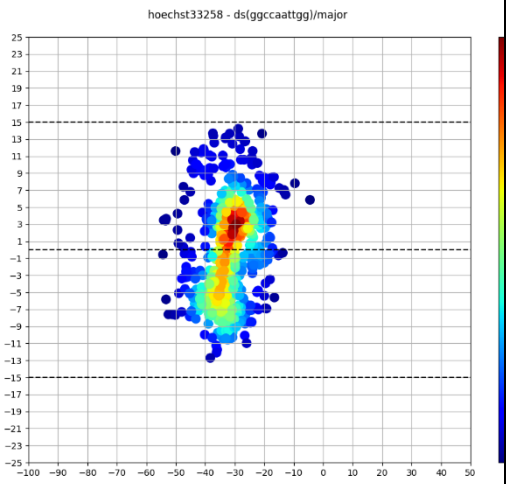
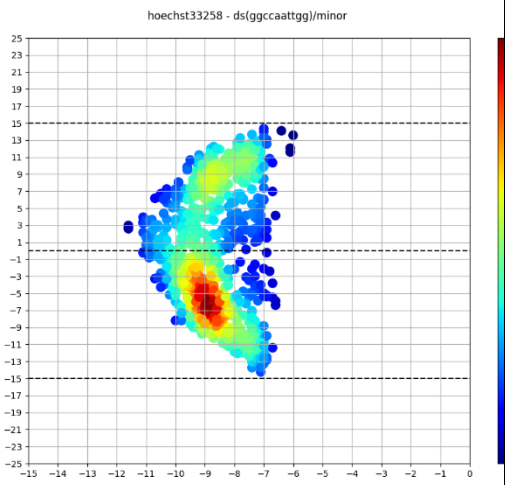
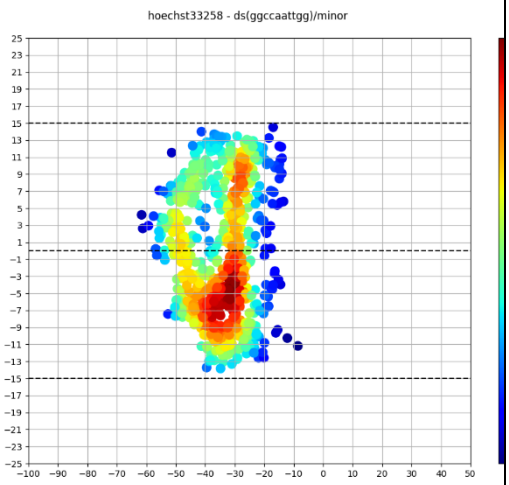
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 31. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of hoechst33258-GGGGGGGGGG.

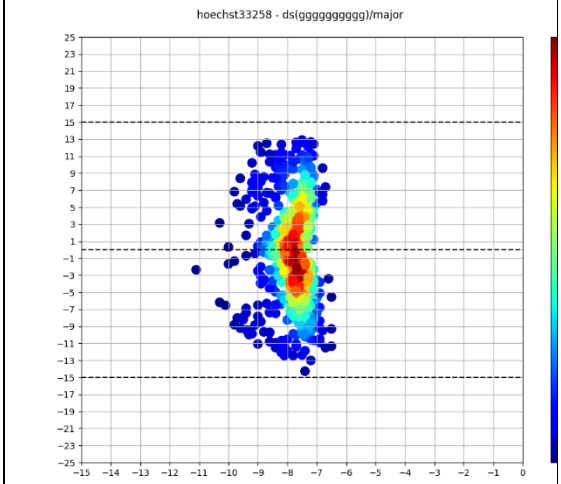
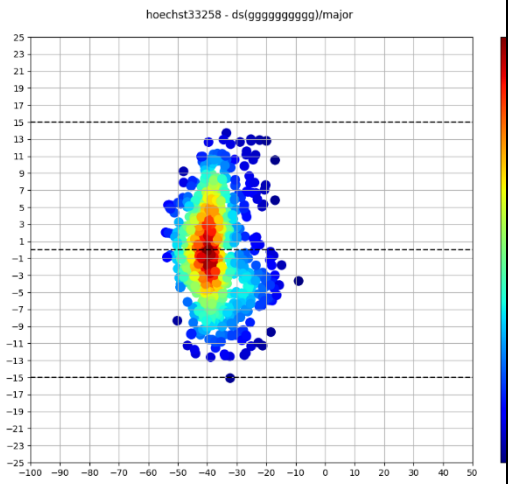
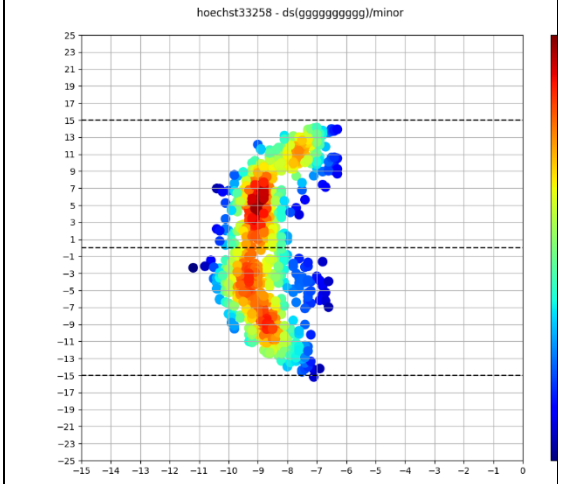
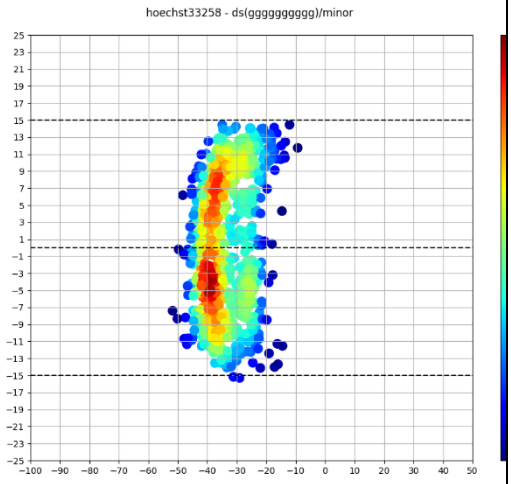
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 32. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of netropsin-AAAAAAAAAA.

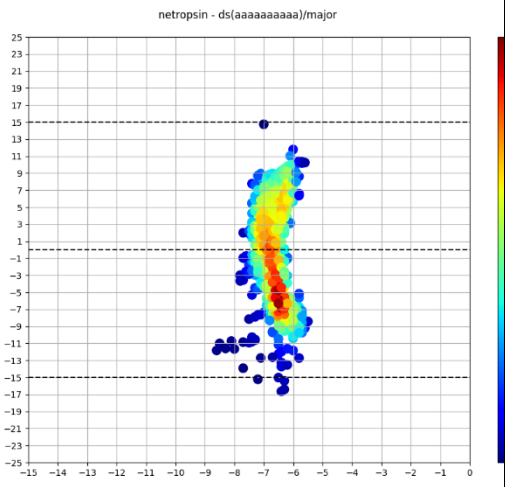
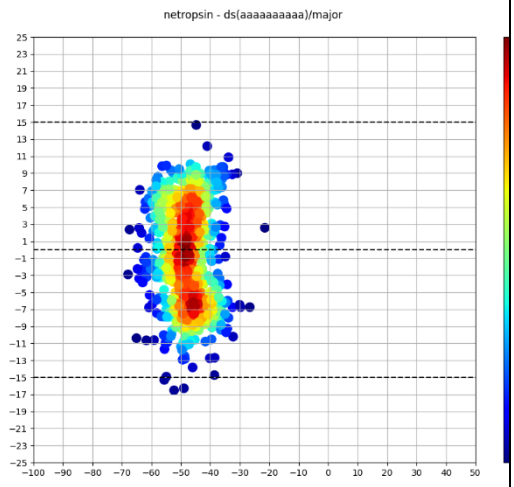
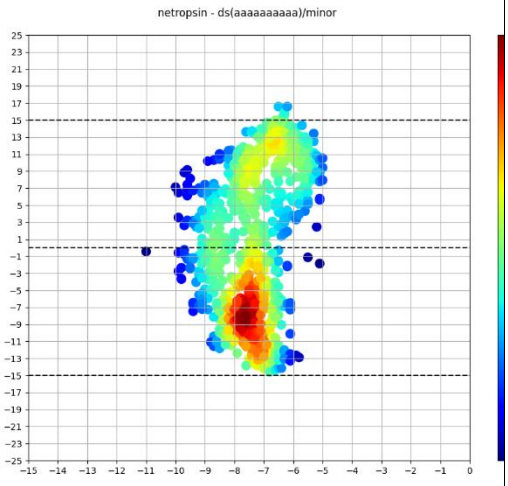
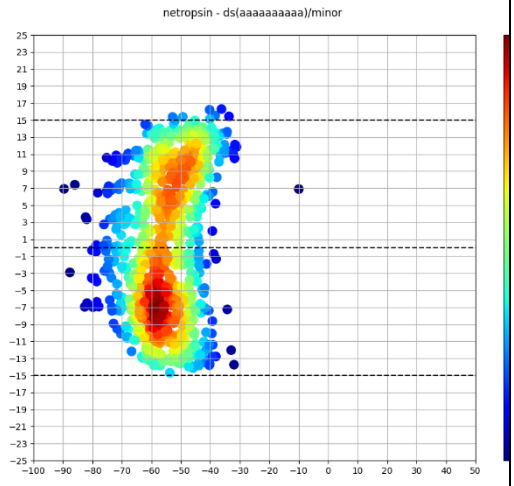
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 33. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of netropsin-ATCGCGCGAT.

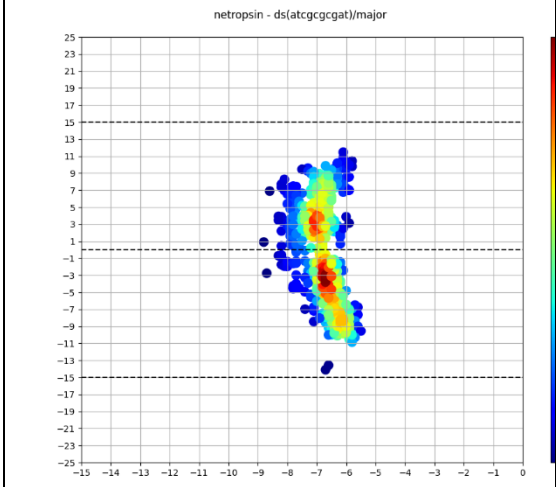
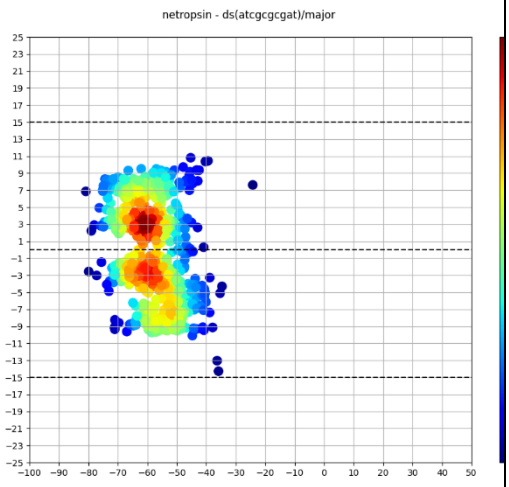
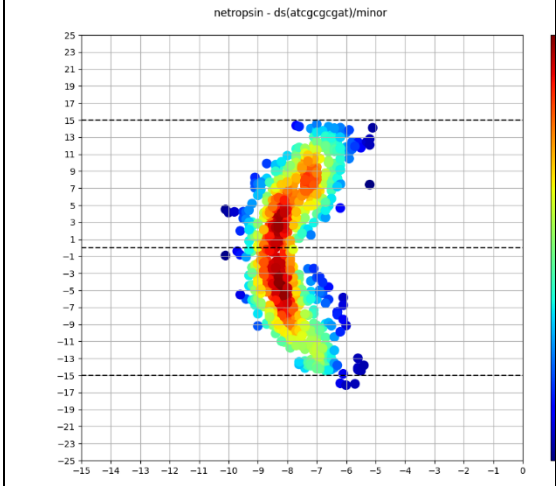
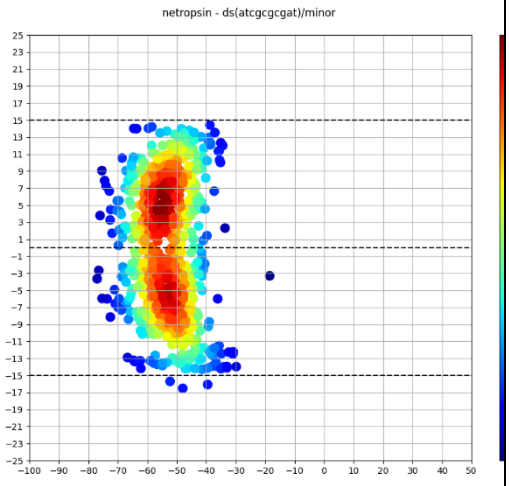
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 34. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of netropsin-CGTATATACG.

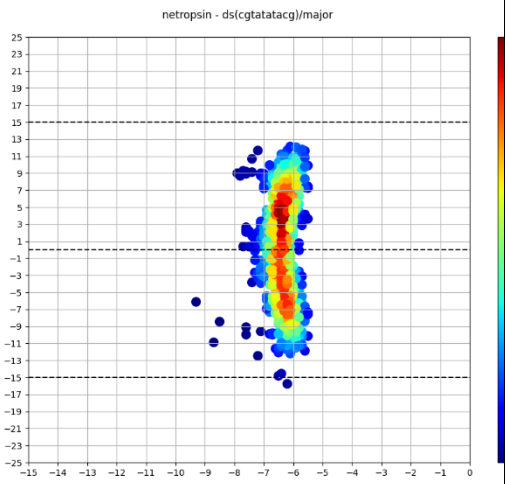
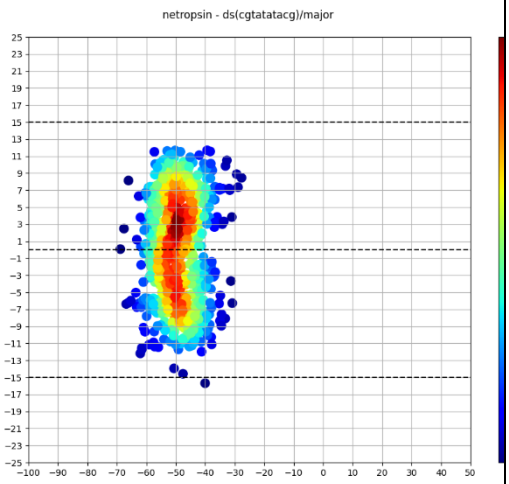
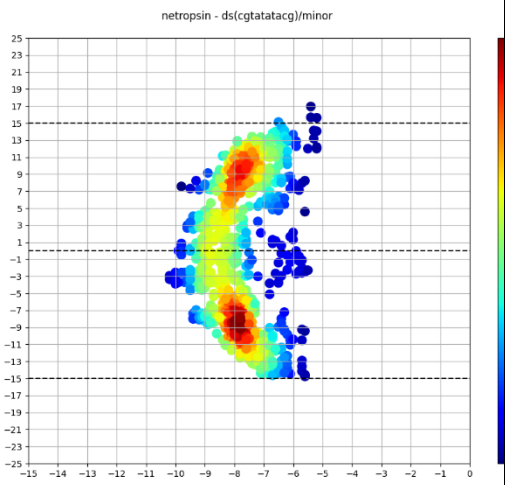
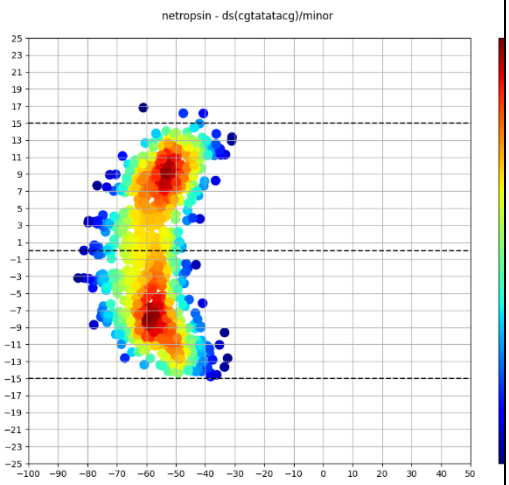
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 35. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of netropsin-GGCCAATTGG .

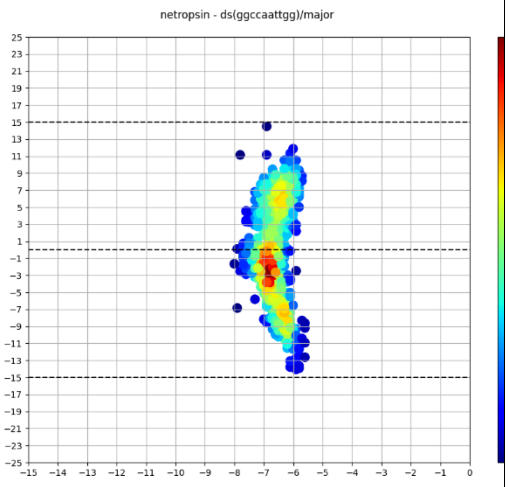
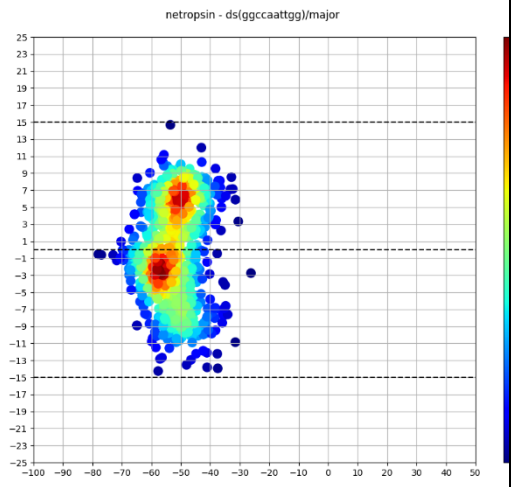
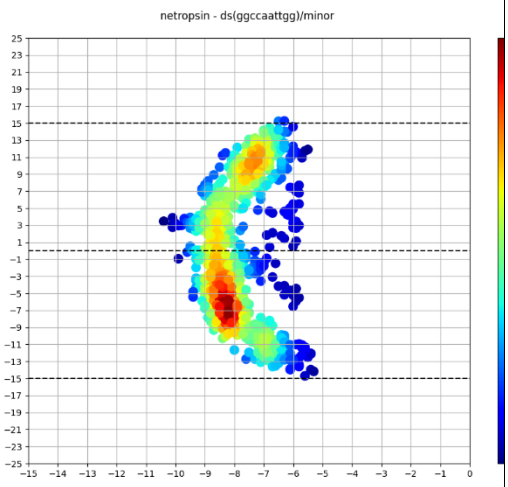
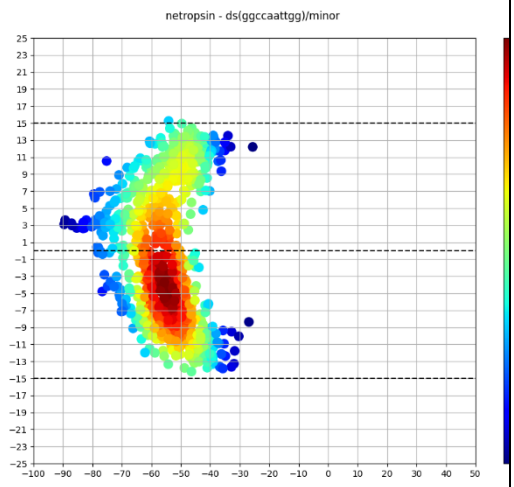
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 36. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of netropsin-GGGGGGGGGG.

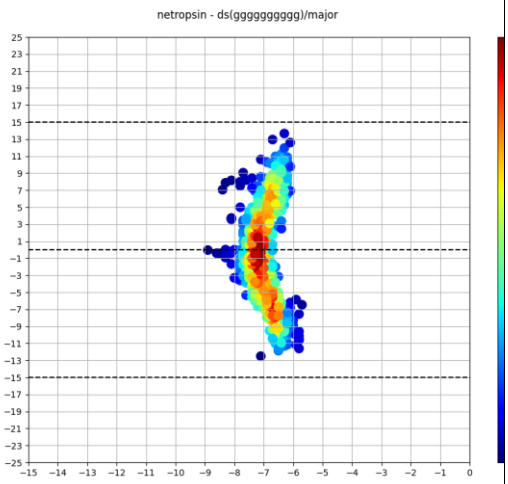
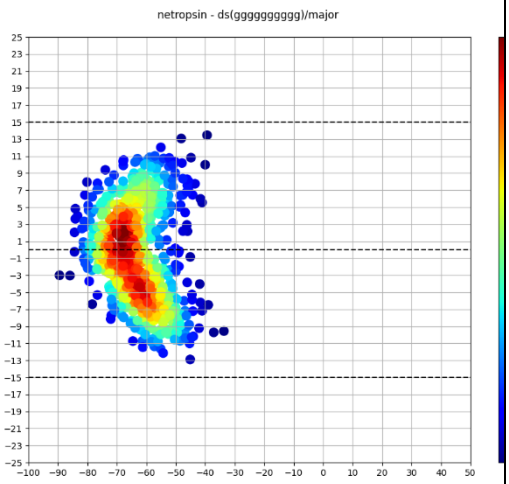
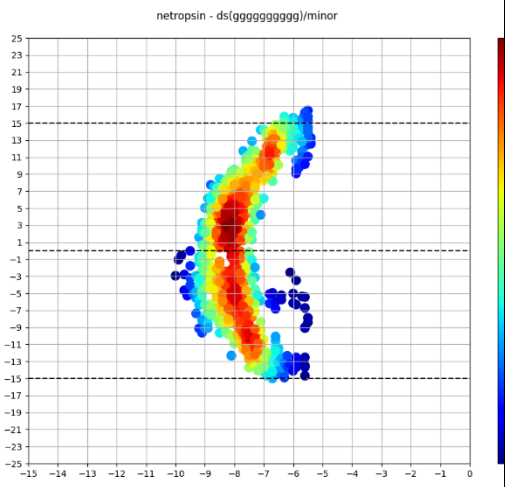
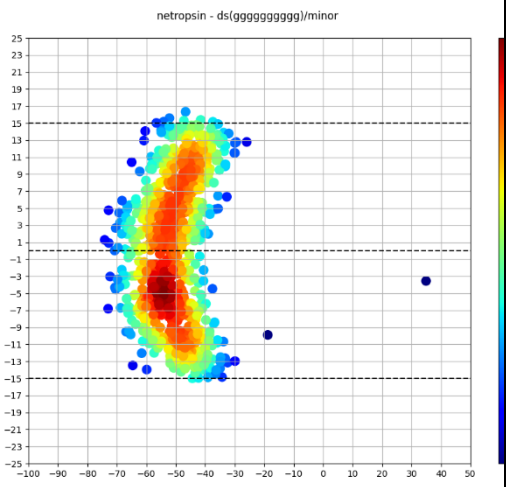
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 37. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of nogalamycin-AAAAAAAAAA.

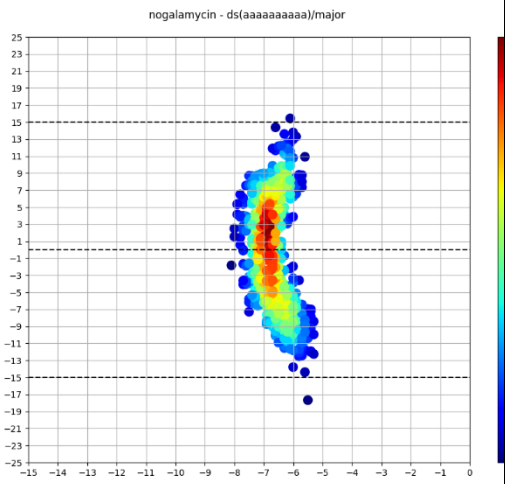
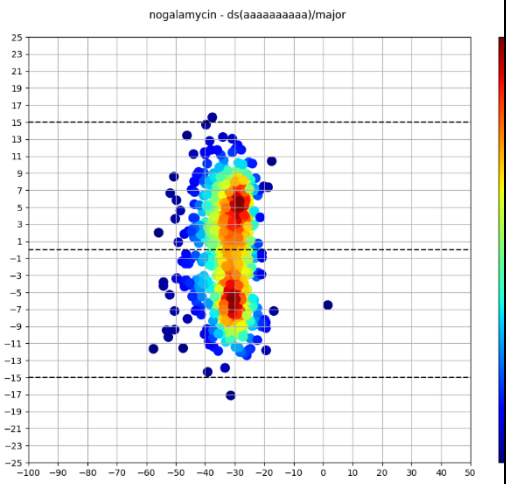
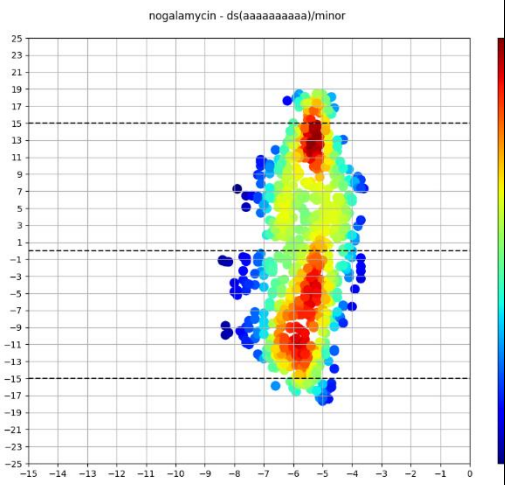
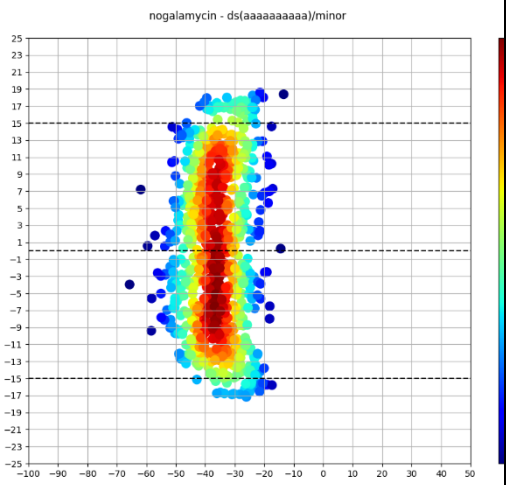
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 38. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of nogalamycin-ATCGCGCGAT.

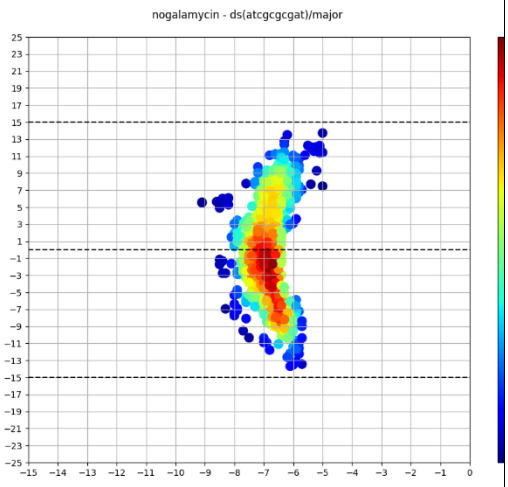
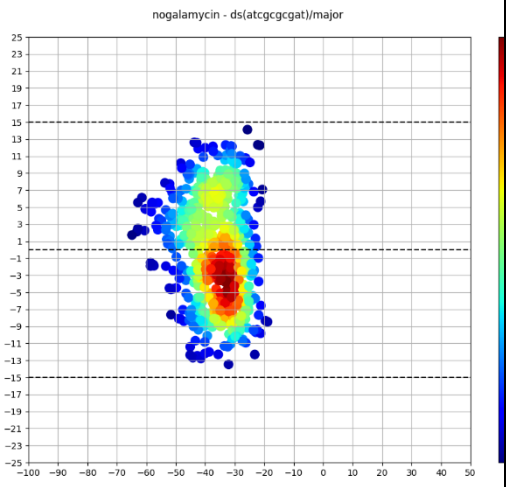
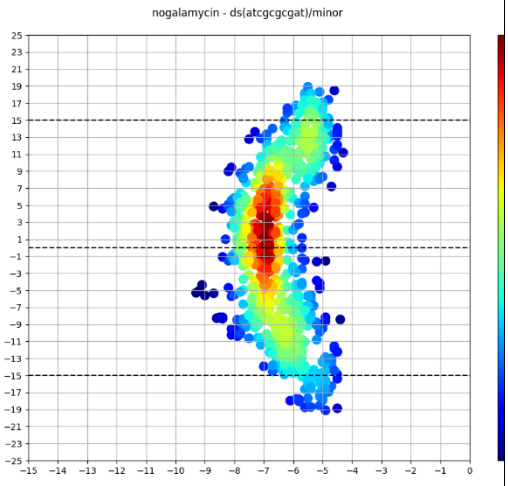
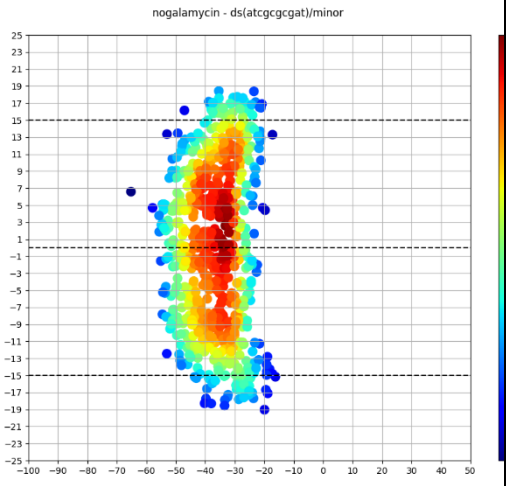
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 39. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of nogalamycin-CGTATATACG .

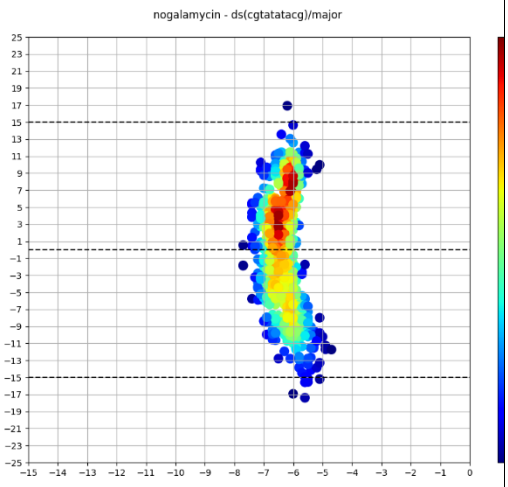
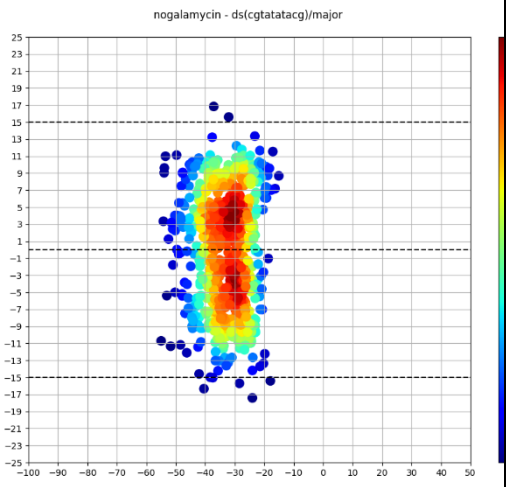
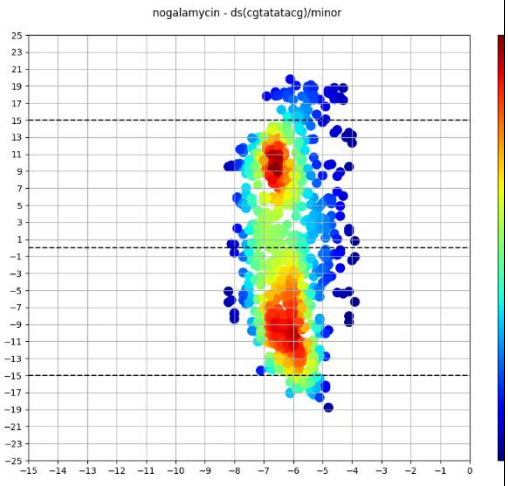
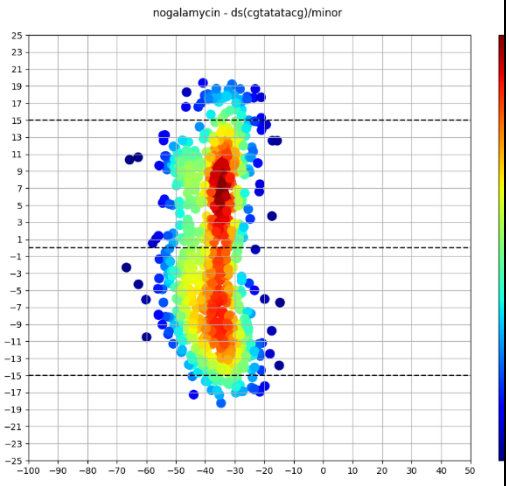
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 40. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of nogalamycin-GGCCAATTGG .

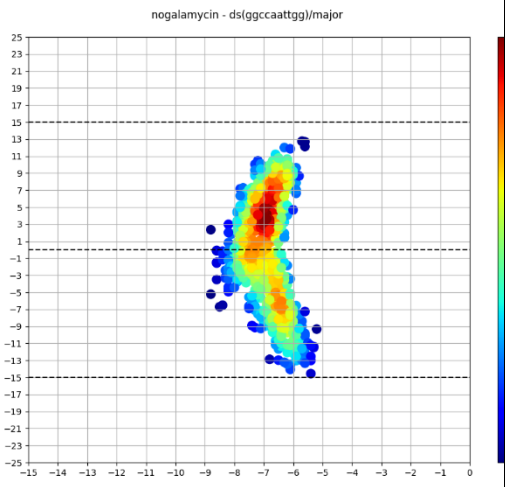
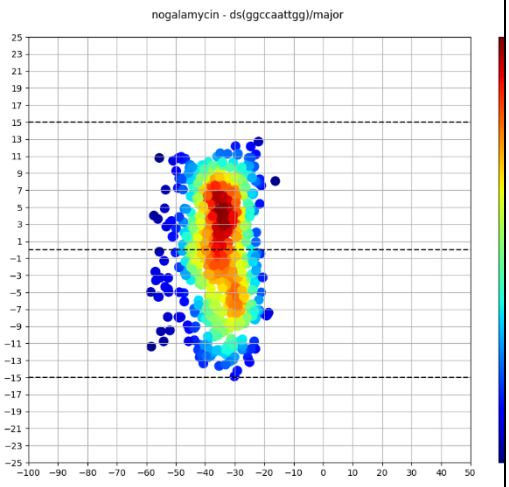
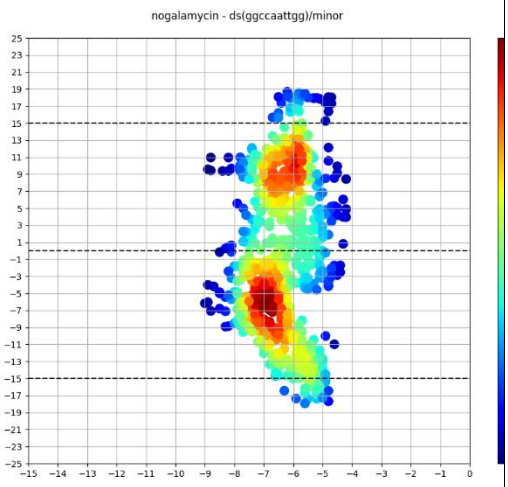
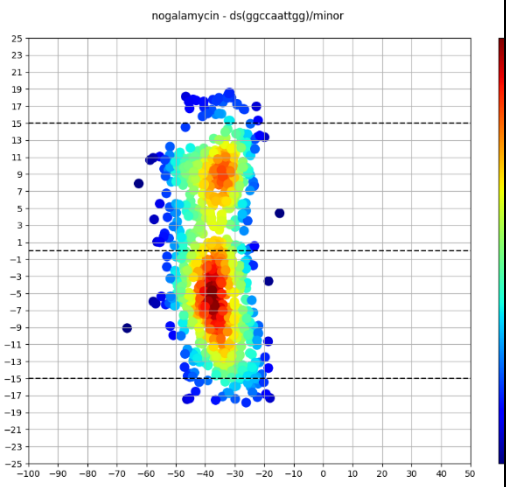
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 41. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of nogalamycin-GGGGGGGGGG.

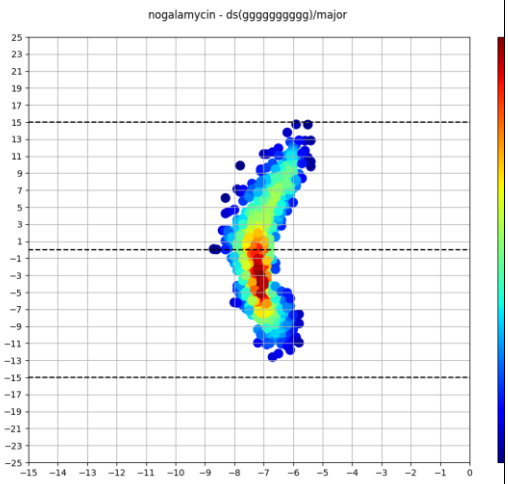
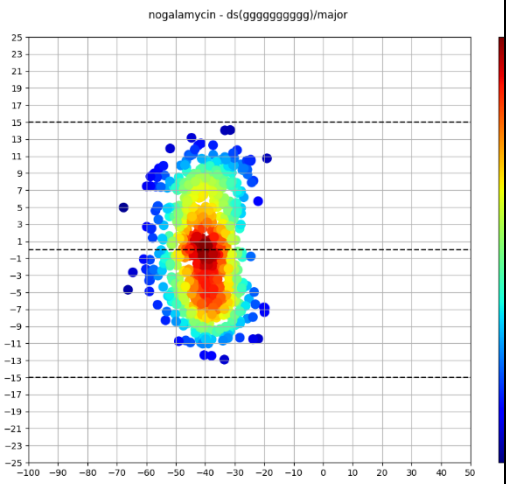
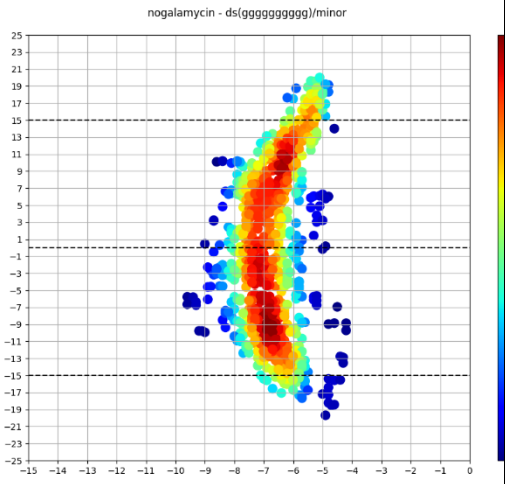
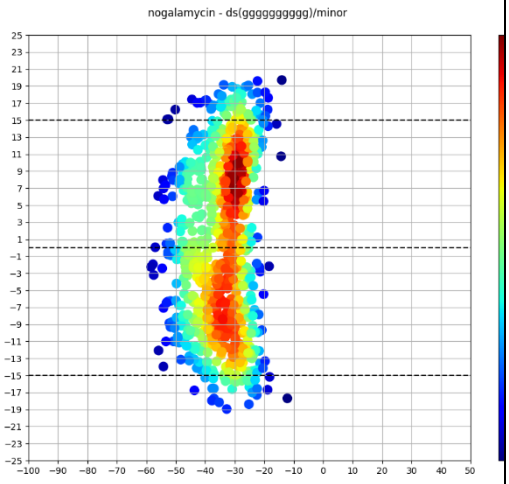
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 42. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of pentamidine-AAAAAAAAAA.

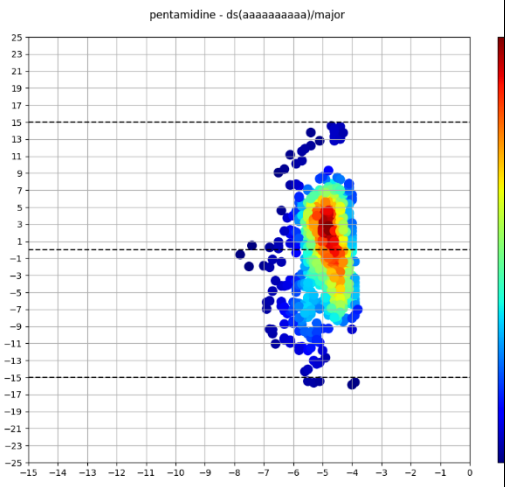
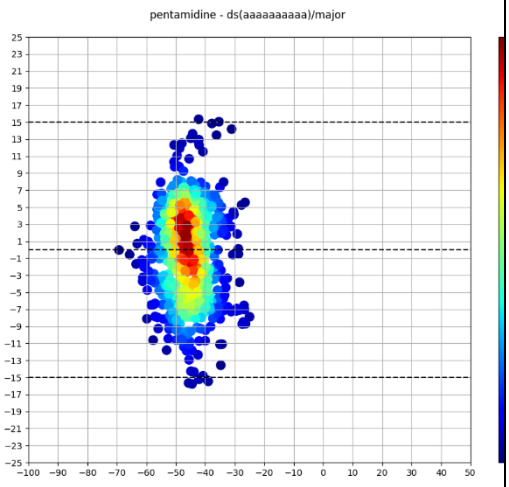
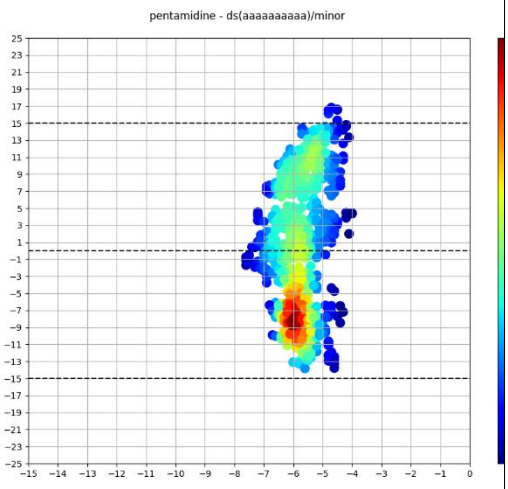
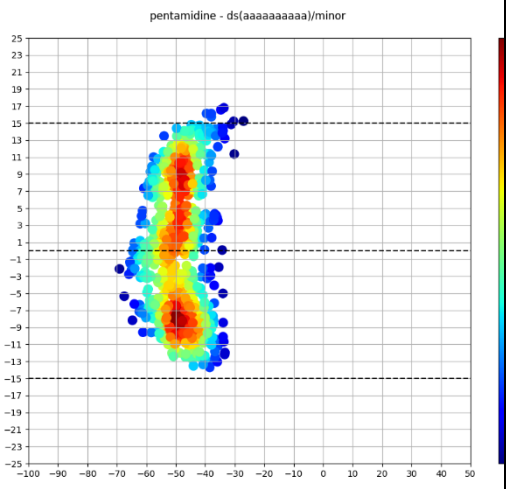
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 43. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of pentamidine-ATCGCGCGAT.

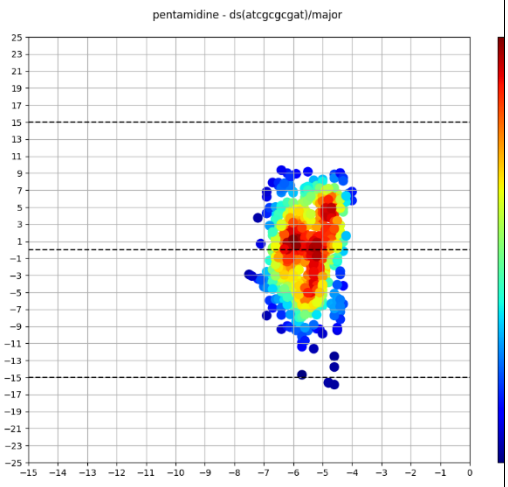
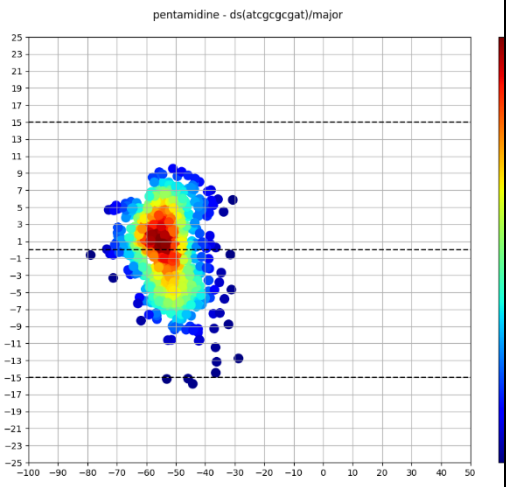
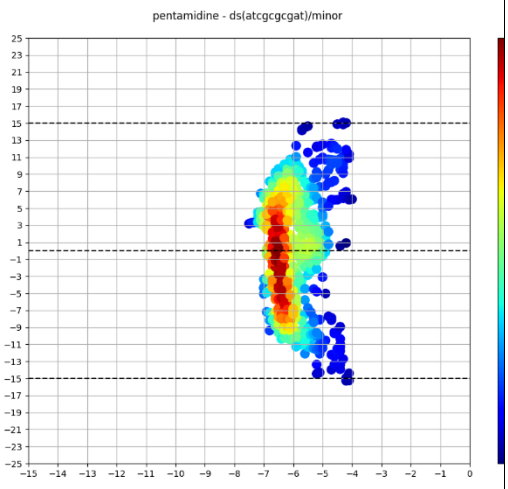
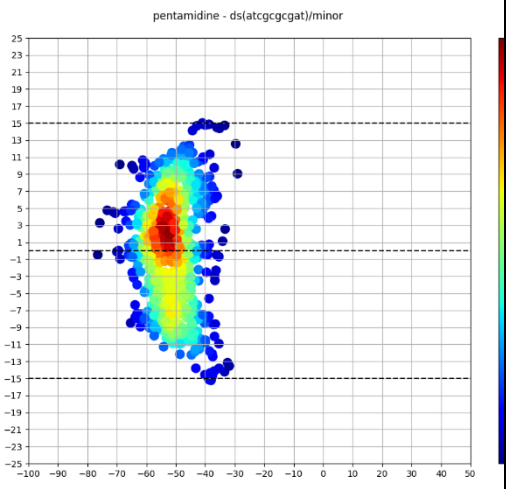
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 44. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of pentamidine-CGTATATACG .

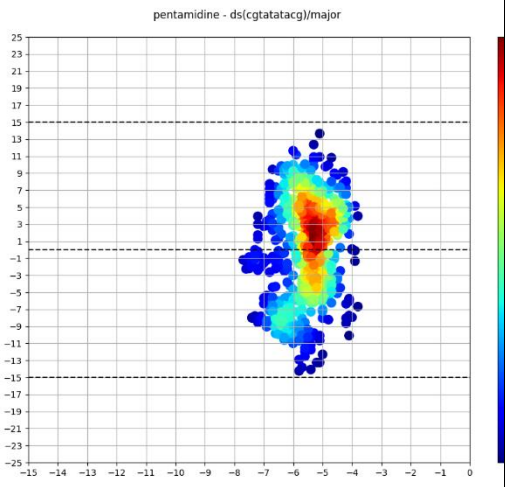
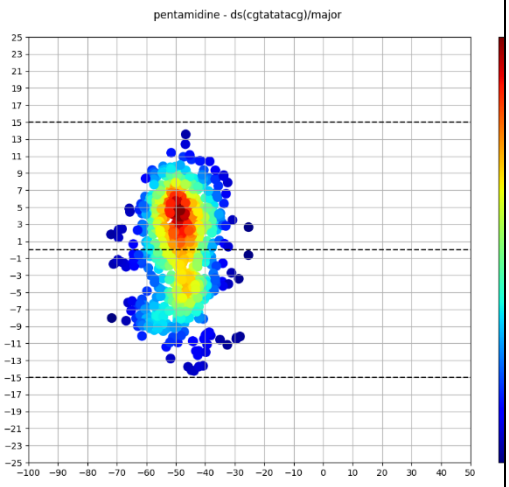
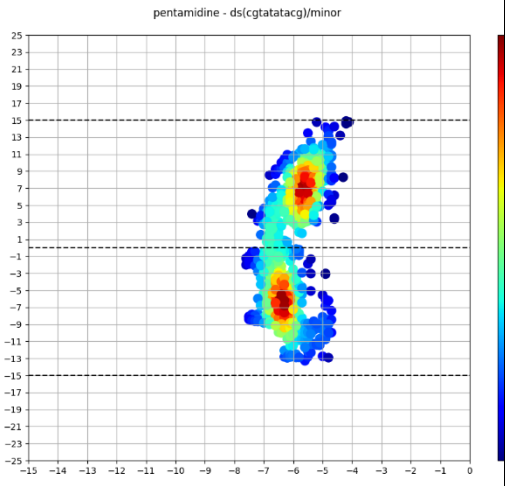
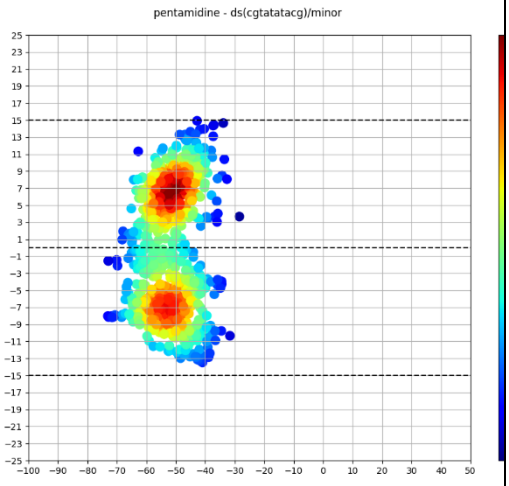
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 45. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of pentamidine-GGCCAATTGG .

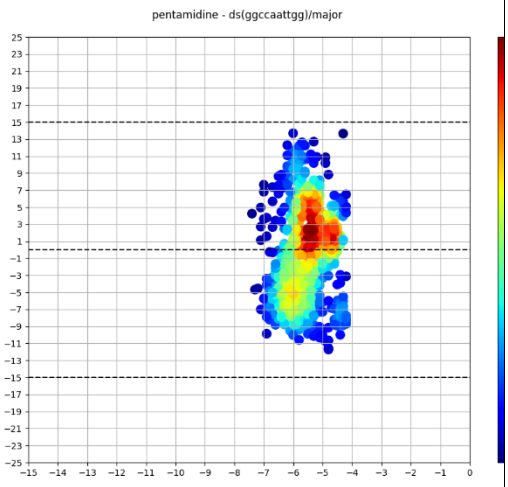
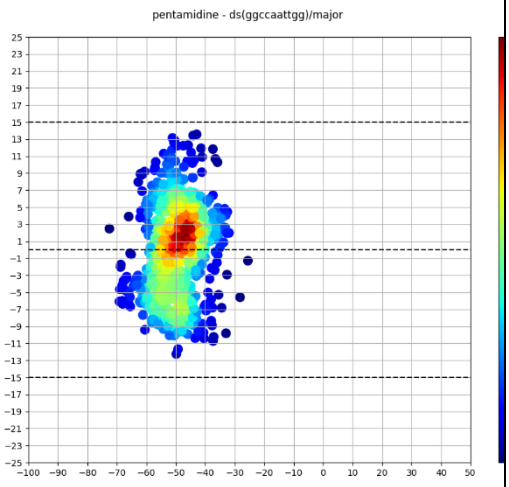
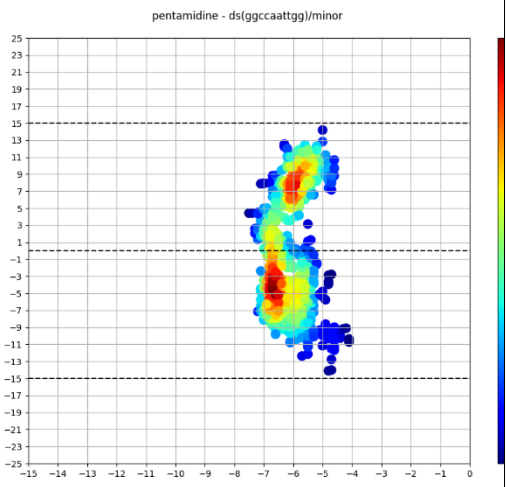
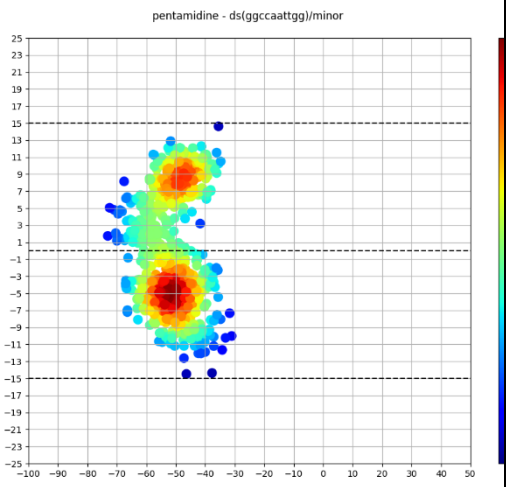
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 46. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of pentamidine-GGGGGGGGGG.

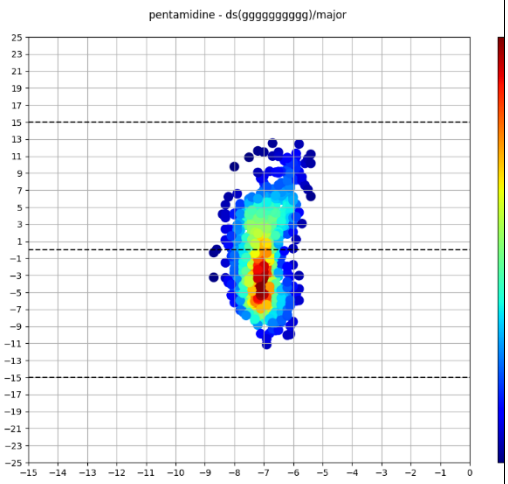
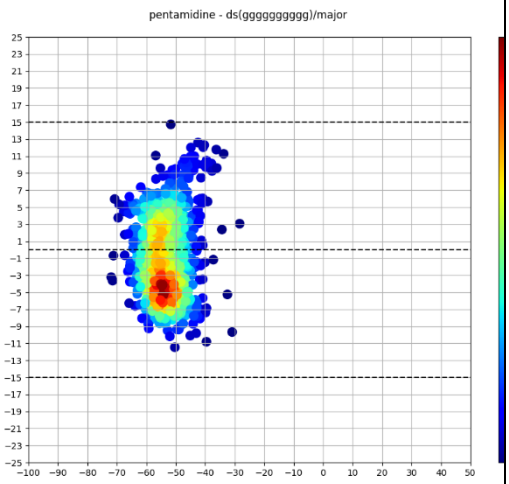
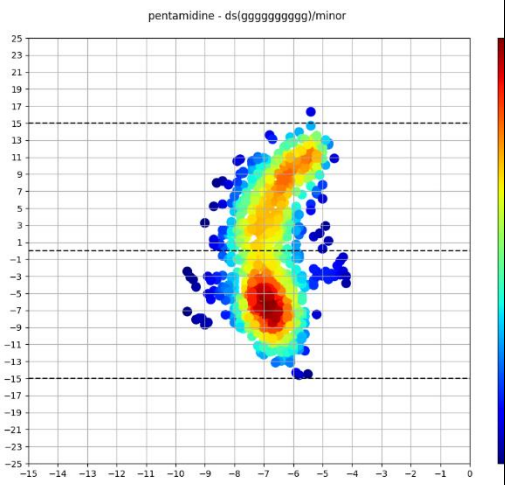
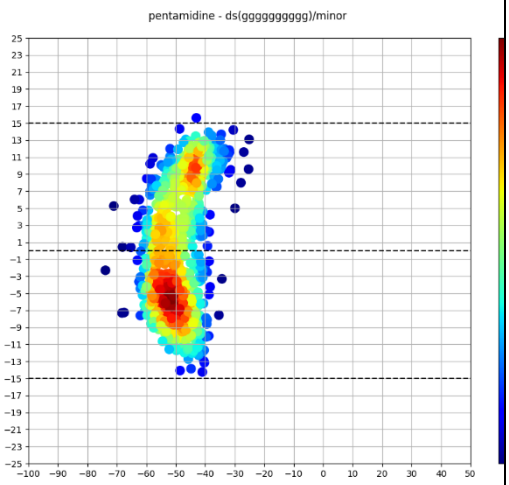
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 47. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of pluramycin A-AAAAAAAAAA.

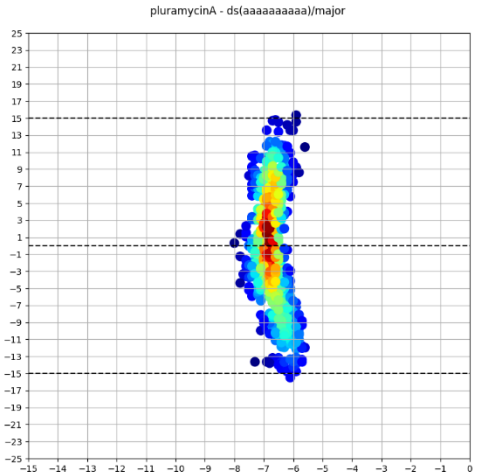
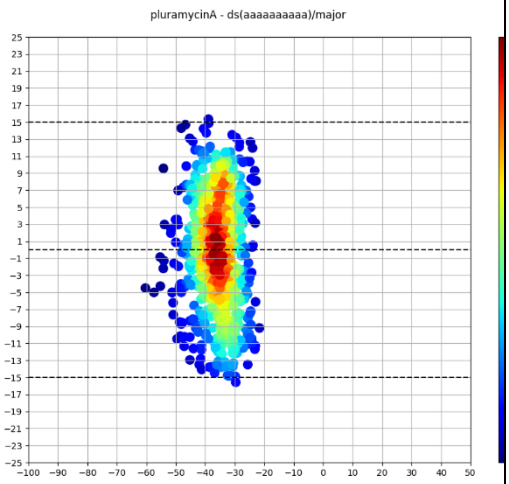
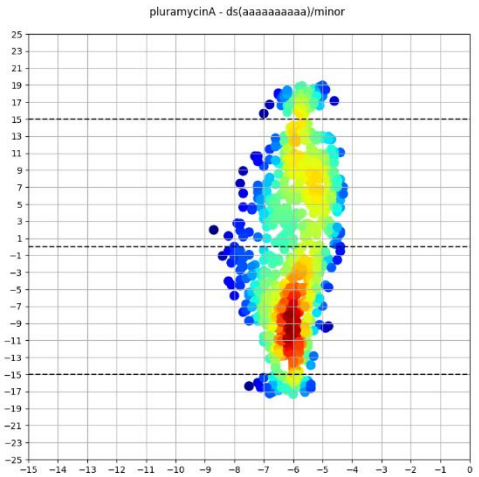
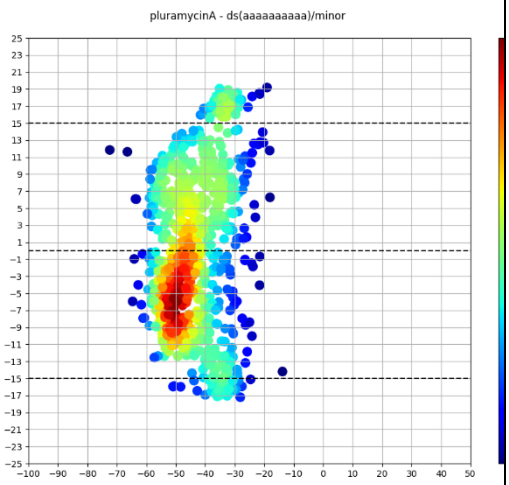
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 48. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of pluramycin A-ATCGCGCGAT.

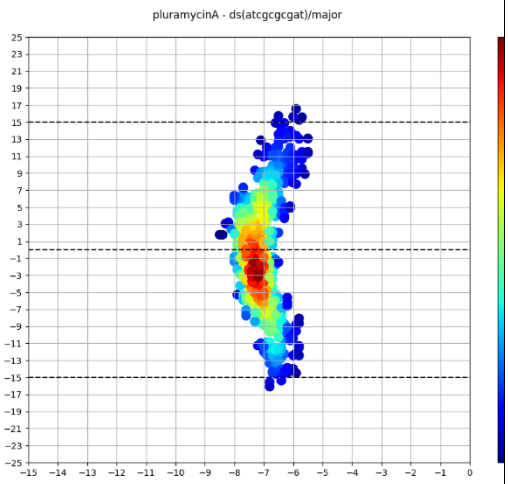
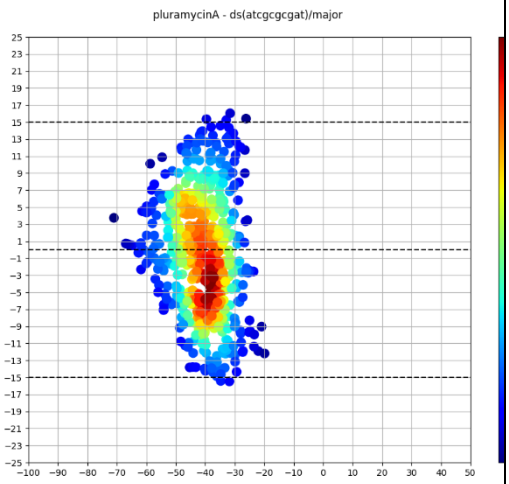
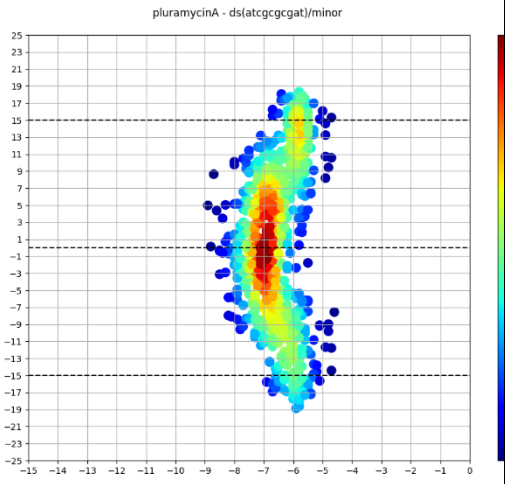
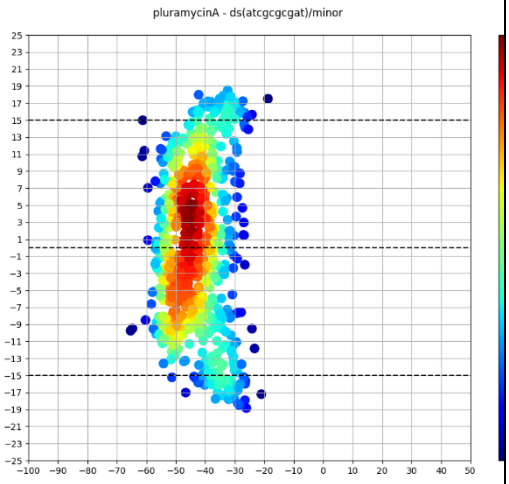
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 49. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of pluramycin A-CGTATATACG .

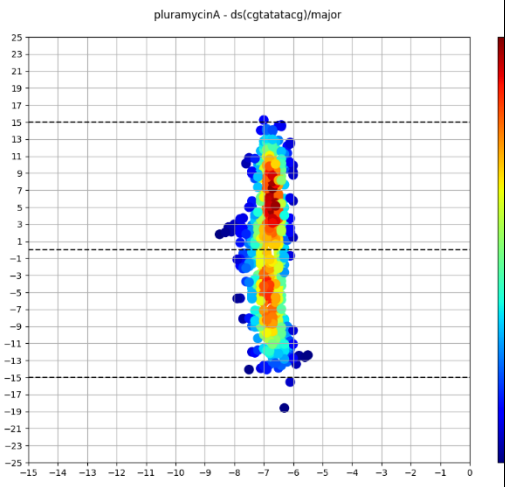
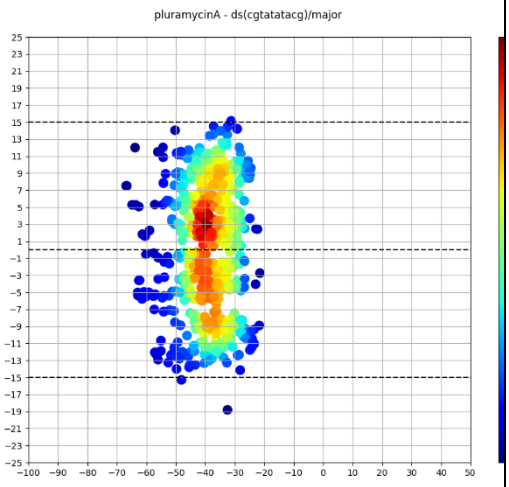
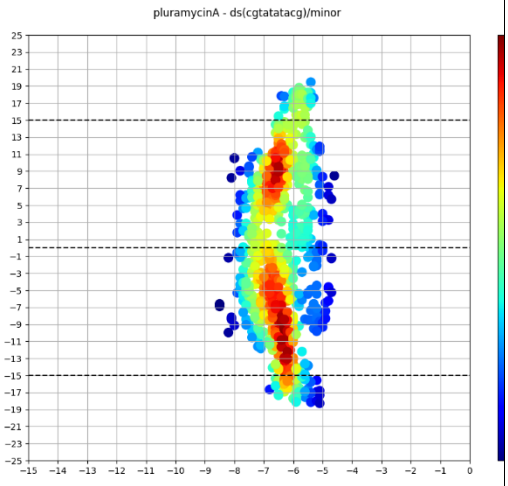
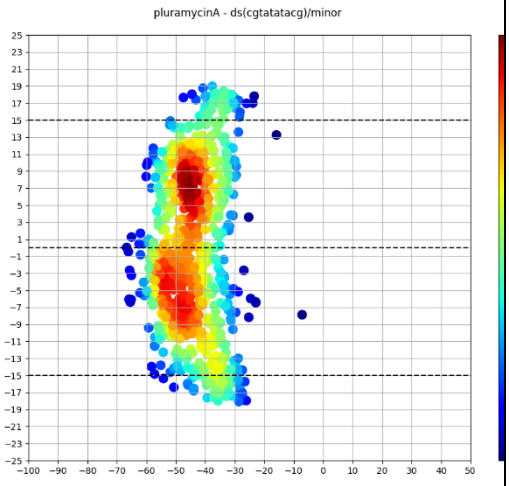
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 50. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of pluramycin A-GGCCAATTGG .

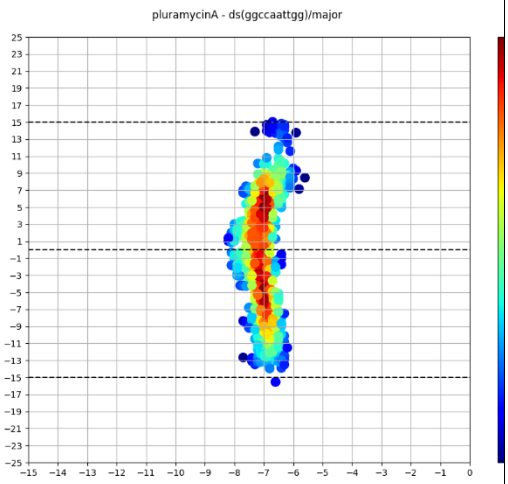
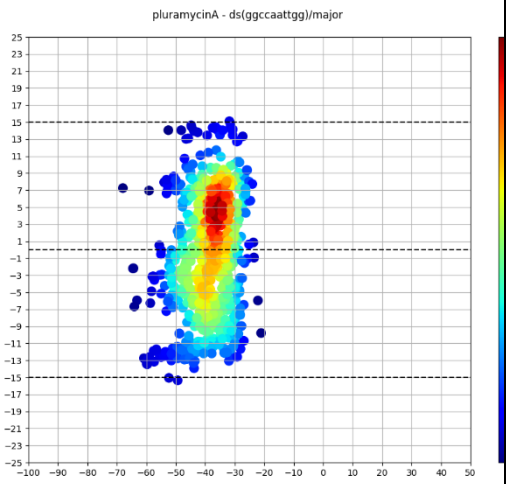
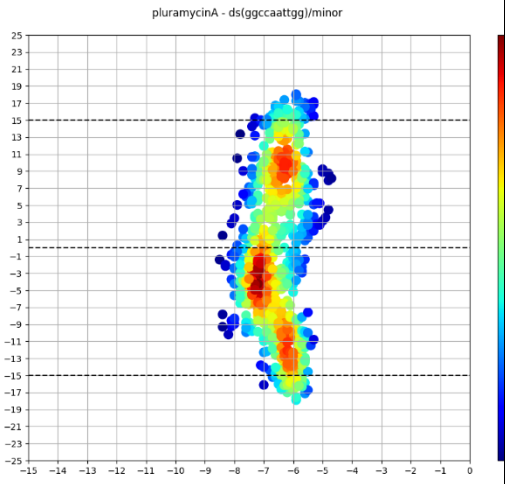
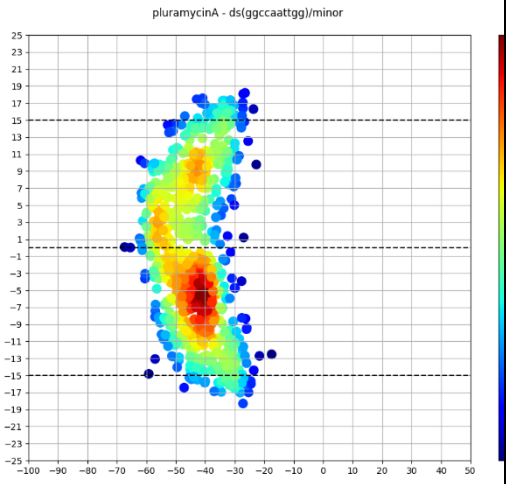
Autodock Vina Score Distribution	**MM-Refinement Distribution**
	
	

Table 51. Three-dimensional (3D) distribution graphs of Autodock Vina scores and MM-refinement binding free energy values of pluramycin A-GGGGGGGGGG.

Autodock Vina Score Distribution	**MM-Refinement Distribution**