12-13-2019

The binding modes of diminazene aceturate with c-MYC G-quadruplexes

Jerrano Bowleg

The binding modes of diminazene aceturate with c-MYC G-quadruplexes

By

Jerrano L. Bowleg

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Chemistry
in the Department of Chemistry

Mississippi State, Mississippi

December 2019

The binding modes of diminazene aceturate with c-MYC G-quadruplexes

By

Jerrano L. Bowleg

Approved:

_____

Steven R. Gwaltney
(Major Professor)

_____

Debra Ann Mlsna
(Committee Member)

_____

Joseph P. Emerson
(Committee Member and Graduate Coordinator)

_____

Rick Travis
Dean
College of Arts & Sciences

Name: Jerrano L. Bowleg

Date of Degree: December 13, 2019

Institution: Mississippi State University

Major Field: Chemistry

Major Professor: Steven R. Gwaltney

Title of Study: The binding modes of diminazene aceturate with c-MYC G-quadruplexes

Pages in Study: 56

Candidate for Degree of Master of Science

Interactions between DNA and ligands are important in the rational design of drugs and in research into DNA function. In particular, the interaction of DMZ with DNA structures named "G-quadruplexes" was considered. G-quadruplexes are structures present in telomeres and several oncogenes.

The main purpose of this project was to provide a computational tool to study DNA-ligand interactions using a variety of molecular modeling techniques that include molecular docking, molecular dynamics simulations (MD) and MM/PBSA (Molecular Mechanics/Poisson Boltzmann Surface Area). We investigated the binding modes and binding affinities of DMZ with c-MYC G-quadruplexes (G4s). We found that the conformation and structural design of the quadruplex can dramatically influence the binding profiles of the ligand. The binding free energies for each site were estimated by the MM/PBSA method. The binding of small molecules to DNA can result in the disruption of oncogene transcription, making it an effective anticancer strategy.

DEDICATION

I would like to dedicate this project to my beloved brother, Garrison Pyfrom, who passed away suddenly and to my mother, sister, stepfather, and grandfather who have provided with me the support.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my advisor, Dr. Steven R. Gwaltney, for affording me an opportunity to pursue my degree under his supervision. Under his supervision, I have grown as a scientist and have gained important skills to carry out scientific research.

I want to thank Dr. Edwin Lewis for affording me this opportunity to partake in the project of DMZ-G4 binding along with Dr. Clinton Mikek.

I want to thank Dr. Debra Mlsna, for serving as a teaching mentor and a committee member and offering me a chance to teach recitation for general chemistry.

I want to thank Dr. Joseph Emerson and Dr. Stephen Foster who served as the graduate coordinators during my time here at Mississippi State University. Their continued advice provided me the confidence to do my best.

I want to thank the Ministry of Education in the Bahamas and the Lyford Cay Foundation for their financial support for my master's studies.

I want to thank my lab mates Eyram Asempa, Mawuli Deegbey, and Hao Chang and my friend Teresa Brown for their continuous support and encouragement.

Finally, I want to thank my family for the love and unyielding support.

TABLE OF CONTENTS

DEDICATION iii

ACKNOWLEDGEMENTS iv

LIST OF TABLES vii

LIST OF FIGURES viii

CHAPTER

I. INTRODUCTION 1

1.1 Background 1
1.2 DNA 1
1.3 G-quadruplex DNA 3
1.4 G-quadruplex binders 7

II. COMPUTATIONAL METHODS 9

2.1 Molecular docking 10
2.2 Molecular dynamics 11
2.2.1 Force fields 12
2.2.2 Energy minimization 13
2.2.3 Integration algorithms 14
2.2.4 Timestep 14
2.2.5 Temperature coupling 15
2.2.6 Pressure Coupling 15
2.2.7 Periodic boundary conditions 15
2.2.8 Ensemble averages 17
2.3 Molecular mechanics Poisson/Boltzmann and Generalized Born 18

III. RESULTS AND DISCUSSION 22

3.1 Simulation protocols 22
3.2 MM/PBSA 26
3.3 Results and discussion 27
3.3.1 Molecular docking 28
3.3.2 Three drug binding modes were observed. 28
3.3.2.2 End Stacking mode 31

3.3.2.3 Groove Binding 32
3.3.2.4 Loop insertion 32
3.3.3 Overall structural stability 34
3.3.4 Important noncovalent interactions: hydrogen-bonding and π-π stacking 45
3.3.5 Binding free energies 46
3.4 Conclusion 47

REFERENCES 49

LIST OF TABLES

Table 3.1 Binding free energies for each complex (kcal/mol) 33

LIST OF FIGURES

Figure 1.1 The structure of the four nitrogenous bases.[7] 2

Figure 1.2 Crystal structure of a dodecanucleotide showing DNA base pairs in double-stranded DNA. Adapted from Brown *et al*.[8] 3

Figure 1.3 Cartoon of a G-tetrad and typical G-quadruplex motifs. Adapted from Mikek.[10] 4

Figure 1.4 Cartoon image showing the path of c-MYC transcription silencing by G-quadruplex stabilizing compounds. Adapted from Brooks *et al*.[21] 6

Figure 1.5 Crystal Structure of the Quadruplex DNA-Drug Complex 8

Figure 2.1 Under periodic boundary conditions, when a particle drifts out of a box, then its image will replace it in the box 16

Figure 3.1 Structure of PDB ID: 2LBY c-MYC G-quadruplex 22

Figure 3.2 Structure of c-MYC 1-2-1 G-quadruplex 23

Figure 3.3 Structure of c-MYC 1-6-1 G-quadruplex 23

Figure 3.4 2D Structure of DMZ 24

Figure 3.5 Best representative structure from MD simulations (100 ns) of DMZ with c-MYC 1-6-1. 29

Figure 3.6 Best representative structure from MD simulations (100 ns) of DMZ with c-MYC 1-2-1. 30

Figure 3.7 Best representative structure from MD simulations (100 ns) of DMZ with c-MYC 2LBY. 30

Figure 3.8 RMSD of DMZ Aa with c-MYC 1-6-1 G4. 35

Figure 3.9 RMSD of DMZ Ab with c-MYC 1-6-1 G4. 35

Figure 3.10 RMSD of DMZ Ac with c-MYC 1-6-1 G4. 36

Figure 3.11 RMSD of DMZ Ad with c-MYC 1-6-1 G4. 36

Figure 3.12 RMSD of DMZ Ae with c-MYC 1-6-1 G4. 37

Figure 3.13 RMSD of DMZ Af with c-MYC 1-6-1 G4. 37

Figure 3.14 RMSD of DMZ Ba with c-MYC 1-2-1 G4. 38

Figure 3.15 RMSD of DMZ Bb with c-MYC 1-2-1 G4. 38

Figure 3.16 RMSD of c-MYC DMZ Bc with 1-2-1 G4 39

Figure 3.17 RMSD of c-MYC DMZ Bd with 1-2-1 G4 39

Figure 3.18 RMSD of c-MYC DMZ Be with 1-2-1 G4 40

Figure 3.19 RMSD of c-MYC DMZ Bf with 1-2-1 G4. 40

Figure 3.20 RMSD of DMZ Ca with c-MYC 2LBY G4. 41

Figure 3.21 RMSD of c-MYC DMZ Cb with c-MYC 2LBY G4. 41

Figure 3.22 RMSD of c-MYC DMZ Cc with c-MYC 2LBY G4. 42

Figure 3.23 RMSD of c-MYC DMZ Cd with c-MYC 2LBY G4. 42

Figure 3.24 RMSD of c-MYC DMZ Ce with c-MYC 2LBY G4. 43

Figure 3.25 RMSD of c-MYC DMZ Cf with c-MYC 2LBY G4. 43

Figure 3.26 RMSF of each c-MYC 1-6-1 nucleotide in complexes and apo as a function of time. 44

Figure 3.27 RMSF of each c-MYC 1-2-1 nucleotide in complexes and apo as a function of time. 44

Figure 3.28 RMSF of each c-MYC 2lBY nucleotide in complexes and apo as a function of time. 45

CHAPTER I

INTRODUCTION

1.1 Background

Cancer is a major public health issue in the world and is the second leading cause of mortality in the United States. In 2019, there will be an estimated 1,762,450 new cancer cases diagnosed and 606,880 cancer deaths in the United States.[1] DNA damage is known to increase the propensity for cancer. Proposed mechanisms for cancer development include oxidation,[2] carcinogen exposure,[3] and ultraviolet-induced strand breakage.[4] Therefore, DNA serves as an attractive anticancer target for therapeutic intervention. Zhou *et al*. revealed that Berenil (diminazene aceturate or DMZ), which has been shown to bind to the minor groove of AT-rich DNA with a micromolar dissociation constant, binds to G-quadruplexes with a nanomolar dissociation constant, *i.e.* with a three orders of magnitude stronger affinity for G-quadruplexes than for AT-rich duplexes.[5] DMZ is used to treat animal trypanosomiasis and hence its toxicological profile is already known. Thus, the DMZ scaffold is a good starting point to develop potent G-quadruplex ligands.

1.2 DNA

Deoxyribonucleic acid (DNA) is a complex molecule that contains all the genetic information necessary to construct and maintain an organism. DNA is a linear polymer that is comprised of a sequence of nucleotides, with each monomer containing a phosphate group, a cyclic furanoside sugar (deoxyribose) and a nitrogenous base. The sugar is phosphorylated at the 5′ position, and a purine or pyrimidine base is attached to deoxyribose at the 1′ carbon. The most

common structure of DNA is the double helix proposed by Watson and Crick in 1953.[6] This duplex structure is comprised of two strands in a double helix formation. The two DNA strands are oriented antiparallel to each other. The DNA strands are read from the 5′ to the 3′ end, where the 5′ end terminates in a phosphate group and the 3′ end terminates in a sugar molecule. The bases are shown in Figure 1.1.

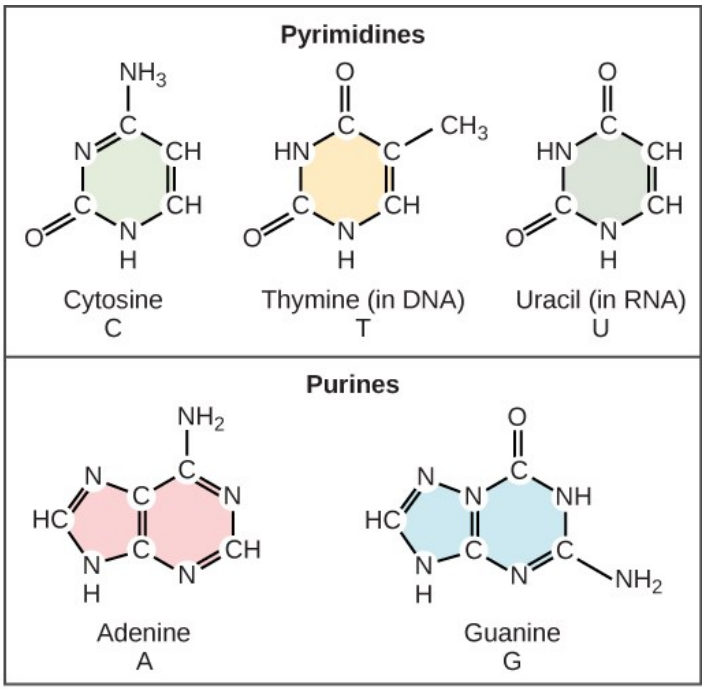


Figure 1.1 The structure of the four nitrogenous bases.[7]

According to Chargaff's rules, the two strands will engage in hydrogen bonds matching the purine bases adenine (A) and guanine (G) with pyrimidine bases thymine (T) and cytosine (C), respectively. The A-T combination possesses two hydrogen bonds, while the G-C interaction is of three hydrogen bonds, as depicted in Figure 1.2.

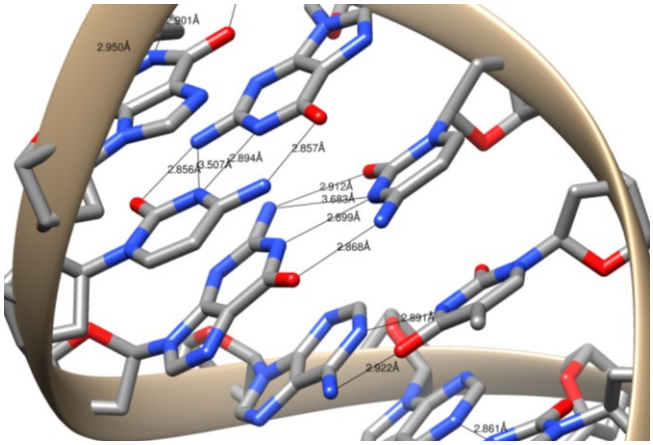


Figure 1.2 Crystal structure of a dodecanucleotide showing DNA base pairs in double-stranded DNA. Adapted from Brown *et al.*[8]

The three-dimensional structure of the dodecanucleotide d(CGCGAATTCGCG). Hydrogen bonds are shown in black with their respective distances. Guanine (top left) hydrogen bound to cysteine (top right).[8]

1.3 G-quadruplex DNA

G-quadruplexes (G4) are found at specific guanine-rich regions of genes and telomeres. Recent attention has been given to these structures because of their potential to serve as a therapeutic target.[9] Guanine has both hydrogen bond donors and acceptors, which allow four guanines bases from a common DNA strand to align in a pseudo-plane through Hoogsteen hydrogen bonds, resulting in a (G-G-G-G) tetrad. In a tetrad, eight hydrogen bonds stabilize the quartet. Central cations and π orbital interactions between stacked aromatic bases help hold the tetrad together. All G4 structures have three distinct structural features: the quadruplex stem, formed from a set of stacked tetrads, the phosphodiester backbone, which forms the grooves, and the unpaired bases that links the quartets and forms loops. A G4 structure can be antiparallel or parallel depending on the position of the connecting loop, as illustrated in Figure 1.3.

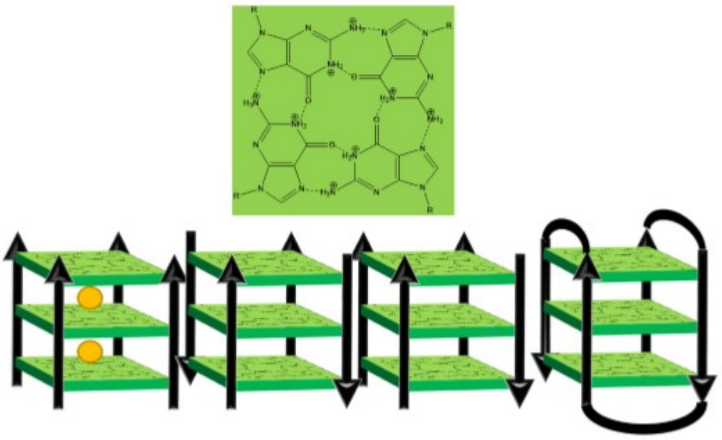


Figure 1.3 Cartoon of a G-tetrad and typical G-quadruplex motifs. Adapted from Mikek.[10]

G-tetrad (top) highlighted in green. Cartoon G-quadruplexes showing the several possible conformations including (from left to right) tetramolecular parallel, tetramolecular antiparallel, tetramolecular mixed parallel/antiparallel, and unimolecular antiparallel. Metal cations are represented by gold spheres only in the tetramolecular parallel G4 cartoon but are present in each G4.[10]

G-quadruplexes can be regarded as being highly polymorphic in nature. G4s are in the genome, which contains between 350,000 and 700,000 distinct putative quadruplex-forming sequences, of which the telomeric and those present in the oncogenic promoter sequences have been extensively studied.[11–13] Quadruplex formations decrease the replicative DNA at each cell cycle, which can lead to cell apoptosis. Telomeres are ensembles of proteins and noncoding DNA that provides protection for chromosomes' terminal ends from unwanted events such as degradation, and recombination. Hence, this ability to control the mitotic clock, particularly in cancer cells, makes G4s potential therapeutic targets for cancer treatment.[14,15] G-rich strands that form G-quadruplexes are free of the complementary C-rich strand, which allows for a myriad of folding topologies when G-quadruplexes are formed. In humans, the single-stranded 3′ telomere ends can extend up to 200 nucleotides (7-33 hexanucleotide repeats), which allows for many

self-association and specific single-stranded binding to proteins. Telomerase is a reverse transcriptase that utilizes its own RNA template to hybridize to the DNA 3′ end and form additional d(TTAGGG) repeats that maintain the 3′ overhangs. In over 85 % of cancers, telomerase was identified to be upregulated.[16] With this discovery, research moved toward disrupting telomerase inhibition to alter cell growth and affect telomere maintenance.[17] Moreover, apart from the telomeric region, other parts of the genome can form G-quadruplexes. Many studies have been done on the promoter regions of the c-MYC oncogenes. The human c-MYC gene is responsible for regulating the expression of proteins involved in cellular growth and proliferation, and overexpression of c-MYC has been linked to cancer.[18,19] The nuclease hypersensitive element III_1 (NHE), which is found to be present at the promoter regions of c-MYC, is responsible for up to 90% of transcription.[20] The c-MYC process is outlined in Figure 1.4.

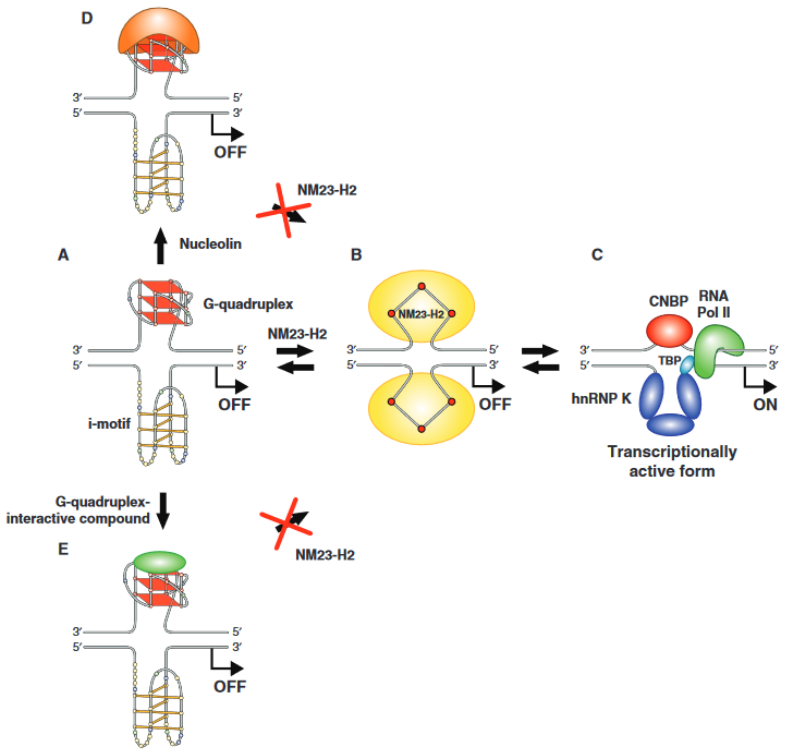


Figure 1.4 Cartoon image showing the path of c-MYC transcription silencing by G-quadruplex stabilizing compounds. Adapted from Brooks *et al*.[21]

A shows the G-quadruplex/i-motif form of the nuclear hypersensitivity element (NHE III1), which is the silencer element. (A) to (C) via (B) illustrates the remodeling of the G-quadruplex/i-motif complex by NM23-H2, in which a stepwise unfolding of the secondary DNA structure is proposed to take place. Binding of nucleolin (A,D) or a G-quadruplex-interactive compound (A,E) to the silencer element prevents conversion by NM23-H2 to the transcriptionally active form of the NHE III1(C)[21]

The promoter region of the c-MYC oncogene was the first to be studied and is of great interest because 1) In several human cancers, c-MYC plays an essential role; 2) the instability of the MYC gene product makes it a good drug target; 3) Small molecules have been reported to stabilize the c-MYC G4 structure and down-regulate gene expression.[22–24]

1.4 G-quadruplex binders

A major focus is to design small molecules with enhanced affinity and selectivity to the DNA substrate of telomerase and c-MYC oncogenes by stabilizing the G-quadruplex formation. Several drug-like molecules were developed to target quadruplex formation specifically in c-MYC oncogenes. Small molecules have been reported to stabilize the c-MYC G4 structure and promote the down-regulation of gene expression. Such ligands include quarfloxin,[25] TmPyP4,[26] quindoline,[27,28] and metal complexes.[29] To stabilize a G-quadruplex structure, the binder (ligand) should possess an aromatic system, allowing for stacking interactions with the tetrad, and should possess an overall positive charge. The positive charges on the ligand would allow for increased water solubility and increase electrostatic interactions with the negatively charged phosphates. G-rich sequences with the propensity to form G4s are present in the promoter regions of several oncogenes such as c-MYC,[30] KRAS,[31] and BCL2.[32] An example of a ligand bound to a quadruplex structure is shown in Figure 1.5.

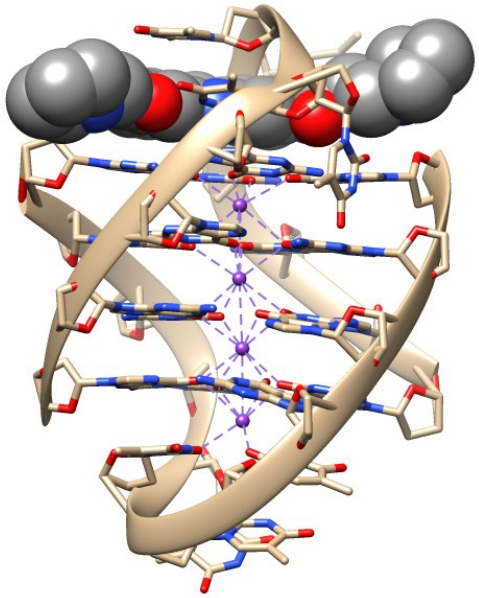


Figure 1.5 Crystal Structure of the Quadruplex DNA-Drug Complex

Adapted from Haider *et al*.[33]

Diminazene aceturate has been reported as a tight binder to G-quadruplexes.[5] In this study, attempts have been made to characterize the binding modes of DMZ with c-MYC G4s. Recently reported thermodynamic studies of DMZ binding to G4 suggested that DMZ does bind to G-quadruplex. However there lacks both structural and energetic understanding of the specific binding events.[34] In this work, we carried out molecular dynamics simulations (MD) simulations together with employing the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA)[35] approach to explore the molecular basis of the DMZ binding to c-MYC variants based on experimental observations of DMZ binding to c-MYC G4.[34]

CHAPTER II

COMPUTATIONAL METHODS

Normally, the modeling of DNA-ligand interactions uses two methods, which are used to predict the binding mode and to estimate the relevant binding affinity. Molecular docking locates the favorable binding position of a ligand, while MM/PBSA determines the binding affinity. The binding free energy is a very useful quantity as it measures how strongly the ligand can bind to the DNA. Calculated binding affinities can reduce the necessity to measure the ligand binding constants experimentally, which can be very time consuming and particularly expensive. As mentioned earlier, molecular docking is used to predict the binding modes of small molecules to G4 DNA. However, the scoring functions that are used in docking programs do not provide accurate binding affinities. The empirical scoring function used in molecular docking is a very simplified method for screening many compounds. The empirical scoring functions do not include factors like solvation, entropic effects, or sampling effects. Hence, more accurate methods are needed to validate binding sites and identify potentially active compounds.

A type of free energy calculation method that is used with molecular docking is the Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) method. The binding energies are calculated using a molecular mechanical force field, the entropies are calculated by normal mode analysis or the quasi-harmonic method, and the solvation free energy calculated by the Poisson-Boltzmann or Generalized Born implicit solvent models. The MM/PBSA approaches have proven successful in many research areas and are hence used in this thesis. [36] In the present

thesis work, the host-guest interactions described above have been investigated by isothermal calorimetry and circular dichroism spectroscopy, which provide indirect evidence of the interaction between ligands and DNA. Unfortunately, there lacks a descriptive mechanism of the atomic scale process of ligand binding to c-MYC G4s.

2.1 Molecular docking

Docking is a computational method that is used to predict the binding of a ligand to its receptors.[37] The binding of a small molecule to a G-quadruplex serves as an attractive strategy for cancer treatment. Molecular docking plays an essential role in structure-based drug design and in the research of DNA/RNA structure and function. Numerous docking programs exist, and much research is geared towards optimizing these methods. In this thesis, we used AutoDock 4.2.[38] The process of molecular docking occurs in two stages. The first stage involves the use of sampling algorithms to predict the orientation and position of the ligand with a defined binding site. Autodock 4.2 uses a rapid grid-based energy method in conjunction with an efficient search algorithm for torsions. Many studies involve a blind docking, which requires a grid to be placed around the entire molecule to allow docking to occur at many different sites on the receptor. To evaluate the energies in an efficient way, the potentials for each atom type is precalculated in AutoDock. Each grid point stores the affinity potentials for all the atom types of the ligand with the DNA. In an AutoDock simulation, the interaction energy of a ligand conformation with the DNA is calculated from the grids. The search for ligand conformations is connected to the torsional degrees of freedom by the various search methods implemented in Autodock.[38] The most common search algorithm used is the Lamarckian genetic algorithm (LGA),[39] which uses the genetic algorithm and a local search algorithm to achieve efficient global phase space convergence and local search optimization. The second stage elucidates the chemical interactions

that exist between the ligand and its receptor, and from there it estimates binding affinities. AutoDock uses a semi-empirical scoring function to estimate the DNA-ligand binding free energies.[38] The scoring function combines classical force fields with empirical parameters and can rank the binding poses quickly. The enthalpic contributions are calculated with a molecular mechanics approach, and the solvation free energy and conformational entropy are calculated by empirical approaches. The DNA and ligand molecules start in unbound conformations and form a bound complex after docking.

2.2 Molecular dynamics

The use of MD simulations in conjunction with automated docking is the most prevalent method to study quadruplex-ligand interactions. Molecular dynamics allows for one to understand the dynamic nature of the ligand binding to quadruplexes. It also provides atomistic detail to aid in interpreting experimental findings. In an MD simulation, the classical Newtonian equation of motion is solved. The force is calculated by taking the gradient of the potential energy (V). The potential energy is determined using an appropriate forcefield.

$$F = -\vec{\nabla}V \tag{2.1}$$

By combining Newton's second law of motion and the above equation we obtain

$$-\frac{dV}{d\vec{r}} = m\frac{d^2\vec{r}}{dt^2} \tag{2.2}$$

The potential energy of the system is calculated by considering the initial coordinates of the atoms, which can be obtained from crystal structures or homology models. The initial velocities of the atoms are assigned randomly via the Maxwell-Boltzmann distribution at a given temperature:

$$P(v) = \sqrt{\frac{m_i}{2\pi k_B T}} \exp\left(\frac{-1}{2}\frac{mv^2}{k_B T}\right) \tag{2.3}$$

where m = mass, v = velocity, k_B = the Boltzmann constant, and T = temperature.

2.2.1 Force fields

In molecular dynamics, the function of the force field is to model the potential energy of an atomic system. The force an atom feels receives contributions from covalent bonding parameters that include bond length, bond angle, torsions, and out-of-plane distortions, and the non-bonded terms that include van der Waals interactions, electrostatic interactions, and hydrogen bonding.

$$V_{net} = V_{BL} + V_{BA} + V_{torsion} + V_{non-bonded} \tag{2.4}$$

$$V_{net} = \sum_{bonds} \frac{k}{2}(l - l_i)^2 + \sum_{angles} \frac{k}{2}(\theta - \theta_i)^2 + \sum_{dihedral} \frac{V_n}{2}(1 + \cos(n\omega - \gamma)) +$$

$$\sum_{non-bonded} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\epsilon_o r_{ij}} \tag{2.5}$$

In equation 2.4, the first term (V_{BL}) is computed from the square of the difference between the bond distance and its reference value. This accounts for bond stretching. The second term (V_{BA}) is calculated by squaring the difference between the bond angle and its equilibrium value, which accounts for the harmonic bond angle potential. The torsional energy ($V_{torsion}$), with where ω is being periodic in the angle, (V_n) is the height of the rotational boundary, n is the periodicity, and γ is the stage factor, which decides where the torsional point goes through its base value. The van der Waals (vdW) potential is modeled using a Lennard-Jones potential function which incorporates repulsive (r^{-12}) and attractive (r^{-6}) terms. The coulombic term calculates the electrostatic interactions between two charged species separated r_{ij}. Most force fields, for example, AMBER,[40] CHARMM,[41] NAMD,[42] and GROMACS[43] normally contain parameters for proteins, nucleic acids, and other biological polymers, as well as common small

molecules such as ATP, ADP, NADP. The force field for a ligand is usually absent; therefore, additional forcefield parameters must be specified. Two force fields used to generate ligand parameters are the CHARMM general force field (CGENFF)[44] and the general AMBER force field(GAFF).[45]

2.2.2 Energy minimization

The initial coordinates of a system, in this case, a biological system, may not be at a local minimum. Therefore, the system must be subjected to minimization before equilibration or sampling in molecular dynamics. The most commonly used optimization techniques include the steepest descent,[46] conjugate gradient,[47] and the Newton-Raphson method.[48] In this thesis, we have employed the use of steepest descent and conjugate gradient. Steepest descent is an efficient algorithm for most minimization processes due to its robustness. The steepest descent method involves taking some initial coordinates represented by a vector x_i. At every step of the calculation, we obtain the gradient (g_k). The step size of each iteration may be adjusted using the *Line search* method. Hence, we can locate the minima on the potential energy surface (PES). This method is very effective if the system is located far away from the minima. However, the convergence criterion becomes slow as the system approaches a minimum. Therefore, in the Amber software, the initial iterations (user-defined) will be performed using the steepest descent method followed by a conjugate gradient to quickly locate the minimum. The conjugate gradient method makes use of the gradient history to decide a better direction for the next step. The systems move in the direction of V_k from a point X_k. The direction is calculated using the gradient at that point and the direction of the previous point:

$$V_k = -g_k + \gamma_k V_{k-1} \tag{2.6}$$

The γ_k is a constant. The first step in this method is like that of the steepest descent algorithm. A line search or arbitrary step method can be used to reach the minima.

2.2.3 Integration algorithms

In the MD simulations, the physical quantities, for example, positions, speeds, increasing speed and so forth, are approximated utilizing Taylor series of expansions as given by

$$r(t+\delta t) = r(t) + \delta t v(t) + \frac{1}{2}\delta t^2 a(t) + \cdots \quad (2.7)$$

$$v(t+\delta t) = v(t) + \delta t a(t) + \frac{1}{2}\delta t^2 b(t) + \cdots \quad (2.8)$$

$$a(t+\delta t) = a(t) + \delta t b(t) + \cdots \quad (2.9)$$

where r = position, v = velocity, and a = acceleration. These quantities are computed utilizing the Verlet,[49] Leap Frog,[50] or Velocity Verlet algorithm.[51] In this investigation, we have utilized Velocity Verlet.

2.2.4 Timestep

There is no standard principle for picking a time step for an MD simulation. The time step δt should be chosen to not influence the general elements of a framework or system while investigating the phase space. One of the general rules is that the time step should be around one-tenth of the briefest time of movement of a system. The bond stretches will have the highest frequencies. Particularly, the period for a C-H bond stretch is around 10 fs. Therefore, the most commonly used time step in MD is 1 fs. A constraint algorithm is a method to constrain bodies in Newton's equation of motion. The use of constraint algorithms is often used in MD simulations to omit some parts of the trajectory calculations. The most commonly used algorithm is the SHAKE algorithm.[52] The SHAKE algorithm has become the standard approach for doing

molecular dynamics with fixed bond lengths. It can also be used to hold angles fixed, but this is less common.

2.2.5 Temperature coupling

In certain MD simulations, different properties of a biological system are assessed at a steady temperature. Therefore, a theoretical regulator is utilized to keep a consistent temperature all through the simulation. The temperature of a framework relies upon the average kinetic energy.

$$< K >_{NVT} = \frac{3}{2} N k_B T \tag{2.10}$$

2.2.6 Pressure Coupling

Likewise, with the temperature coupling, we can use a "pressure bath" or barostat to control the pressure of the system during a simulation. There exist many algorithms such as Berendsen,[53] which scales the box vectors and coordinates of the atoms at each of the time steps in order to maintain a constant pressure.

2.2.7 Periodic boundary conditions

Most biological experiments are conducted in the solution phase. Therefore, simulations are required to be immersed in an explicit solvent environment with solvent molecules and counterions added to maintain the neutrality of the system. The box, however, must obey Periodic Boundary Conditions (PBC). By applying Periodic Boundary Conditions and using a minimum number of particles, one can efficiently sample the phase space and calculate various properties of the system. Under the PBC, the box will be replicated in all 3 directions. The 2D arrangement of this replication is shown in Figure 2.1.

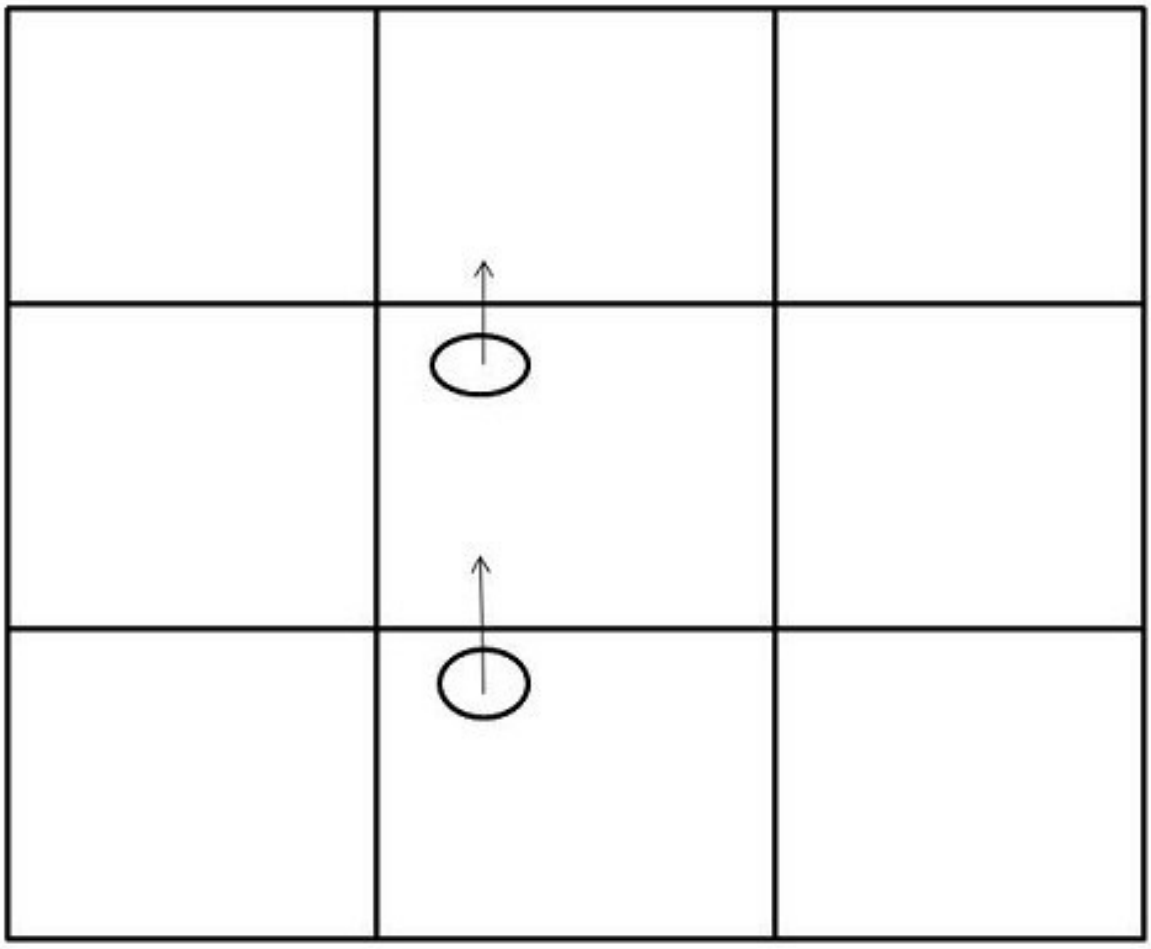


Figure 2.1 Under periodic boundary conditions, when a particle drifts out of a box, then its image will replace it in the box

As a simulation progresses, a particle that may drift out of the box. If it does, its image will immediately replace it in the box, as illustrated in Figure 2.1. This benefits the simulation, as it ensures that the number of particles in the simulation is kept constant.

The calculation of non-bonded interactions is the most expensive step in performing MD simulations. Practically, it is very difficult to calculate the sum of the nonbonded interactions, as the number of non-bonded interactions is proportional to the square of the number of particles in

the system. However, in most biomolecular simulations, a non-bonded cutoff is implemented to decrease the time of the calculation. When a non-bonded cutoff is used between pairs of atoms, if the atoms are farther apart than the cutoff distance, the interaction is set to zero. For an accurate simulation, all the contributions to the energy of the system should be considered overall. However, it is impractical to attempt to sum up all the nonbonded interactions in a standard MD simulation. Hence, one must introduce spherical cutoff distances in order to dissipate long-range electrostatic interactions outside of the specified distance without hurting the quality of the results.

The highly charged DNA backbone, cations, and charged ligands have a profound effect on quadruplexes. Therefore, if the long-range interactions are poorly treated, the simulation can become unstable. A useful method to truncate the long-range electrostatics is the particle-mesh Ewald (PME).[54,55]

$$V = \frac{1}{2}\sum'_{|n|=0}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{4\pi\varepsilon_o|r_{ij}+n|} \tag{2.11}$$

It eliminates that fundamental problem and helps produce a stable simulation. The size of this cutoff depends upon the case, but normally in biomolecular simulations, a suitable cutoff is usually between 8-10 Å.

2.2.8 Ensemble averages

An ensemble must also be chosen to adequately reflect the experimental conditions. An ensemble is a collection of phase space which corresponds to a thermodynamic (macroscopic) state of the system.[56] In MD simulations of biological molecules, different points are generated in the phase space over time in an ensemble. MD simulations can be performed in a variety of ensembles, e.g. canonical (NVT), microcanonical (NVE), and isobaric-isothermal (NPT), where

the number of atoms (N), pressure (P), volume (V) and temperature (T), and/or energy (E) is kept constant. In order to determine macroscopic properties and experimental values, the positions (r) and momenta (p) of a system must be known. Any measured value in the system is measured as an average over a duration of time. Hence, for an N particle system, the average value for a particular property over a period is given by the equation:[56]

$$A_{avg} = \int \int dp^N dr^N A\,(p^N, r^N) \tag{2.12}$$

where A = an observable property of the system. In order to computationally calculate the property of a system, an average value is replaced by an ensemble average. Wherein, numerous replicas of a system are simultaneously considered, and the property of the system can be represented as

$$< A >= \iint dp^N dr^N A(p^N, r^N)\rho(p^N, r^N) \tag{2.13}$$

where ρ = The probability density of the system. Under the *Ergodic* hypothesis,[56] the ensemble average is equal to the time average according to the equation:

$$A_{avg} =< A > \tag{2.14}$$

2.3 Molecular mechanics Poisson/Boltzmann and Generalized Born

Solvent effects are important to consider when studying biomolecules such as DNA, as they often exist in water. In an MD simulation water is normally included in the simulation and is represented explicitly by various water models such as TIP3P,[57] SPC,[57] or SPC/E.[58] While it is useful to add in a solvent to a simulation, this also presents an increase in computational time, and it is also very difficult to obtain free energy values from simulations in explicit solvent in addition as the increase in the degrees of freedom from the solvent molecules can influence the calculation. To address this shortcoming, implicit solvation can be used to represent the solvent

effect. Essentially, the explicit water environment is replaced by a continuum which has similar dielectric properties as water. The usefulness of implicit solvation provides more ease in calculating the free energies as the solvent effect is taken out. It also reduces the computational requirement because many water molecules are represented implicitly, and the sampling is enhanced because the conformation space due to solvent viscosity is reduced.

The prediction of ligand binding free energy to a quadruplex remains a challenging problem to this day. Although several methods exist, for this thesis, we employed the use of the implicit solvent methods. The free energy of binding for a receptor-ligand system can be expressed as

$$\Delta G_{bind} = \Delta E_{forcefield} + \Delta G_{sol} - T\Delta S \tag{2.15}$$

where ΔE is the total energy of a system in the gas phase which is calculated by a defined forcefield

$$\Delta E = \Delta E_{bonding} + \Delta E_{non-bonded} \tag{2.16}$$

Similarly, ΔG_{sol} is calculated using the following relationship

$$\Delta G_{sol} = \Delta G_{elec} + \Delta G_{vdw} + \Delta G_{cavity} \tag{2.17}$$

The ΔG_{elec} is a polar contribution i.e. associations between charged solutes and polarized solvent. The last two terms are considered as non-polar contributions. The solvent molecules must reorganize and create a cavity when a solute is added to a pure solvent. In this way, the entropy of the solvent declines, and the term ΔG_{cavity} will be positive. The ΔG_{sol} is the solvation free energy. Different techniques have been proposed for ascertaining ΔG_{elec}. The molecular mechanics energies joined with the Poisson– Boltzmann[59] or Generalized Born[60] and surface region continuum solvation (MM/PBSA and MM/GBSA) strategies are well-known ways to estimate the free energy of the binding of small ligands to biological macromolecules. The

electrostatic contribution can be obtained using the Poisson-Boltzmann (PB) or Generalized Born (GB) model.

In the 1920s, Born developed a method to calculate the free energy of solvation when a charged particle is transferred from *in vacuo* to a solvent. This in turn, can be correlated proportionally to ΔG_{elec} and is represented as

$$\Delta G_{elec} = \frac{-q^2}{2a}\left(1 - \frac{1}{\varepsilon}\right) \tag{2.18}$$

where q = charge of the ion, a = radius of the cavity, and ε is the dielectric constant of the medium. The ionic radius is assigned as the radius of the cavity. We used the Generalize Born equation for calculating electrostatic interactions within the scheme of study.[61] Each particle was represented by a sphere of radius (r_i) and charge (q_i). The interactions between every two atoms are calculated, and the net electrostatic interactions of the system can be written as equation 2.19.

$$\Delta G_{elec} \approx \Delta G_{gb} = \frac{-1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{f(r_{ij}, a_{ij})} \tag{2.19}$$

The function f depends on r_{ij} and the Born radii a_{ij} and is given by equation 2.20,

$$f(r_{ij}, a_{ij}) = \sqrt{(r^2{}_{ij} + a^2{}_{ij}\, e^{-D})} \tag{2.20}$$

where $a_{ij} = (a_i a_j)^{1/2}$ and $D = r^2_{ij}/(2a_{ij})^2$

In this method, the spherical particle is assumed to be filled with uniform matter and a dielectric constant ranging from 1-4. The surrounding solvent medium is usually water, containing a dielectric constant of 80 (F/m). In the Poisson-Boltzmann strategy, the solute and solvent are assigned dielectric constants of 1-4 and 80 separately. A summed-up Poisson condition is characterized as

$$\nabla^2\phi(r) = \frac{-4\pi\rho(r)}{\varepsilon} \tag{2.21}$$

where $\phi(r)$ is the electrostatic potential, ρ = charge thickness, and ε = dielectric consistent of the medium. When we have in excess of one dielectric component in a framework, the Poisson condition should be modified. The modified Poisson equation is known as the Poisson-Boltzmann equation[62] and is given by equation 2.23,

$$\nabla[\varepsilon(r)\nabla\phi(r)] - \kappa'^2 sinh[\phi(r)] = -4\pi\rho(r) \tag{2.22}$$

where κ'^2 can be calculated using the Debye-Hückel inverse length, and κ is given by equation 2.23,

$$\kappa^2 = \frac{\kappa'^2}{\varepsilon} = \frac{8\pi N_A e^2 I}{1000\varepsilon k_B T} \tag{2.23}$$

where e = charge of a proton, I = ionic strength of the solution, and N_A = Avogadro's number.

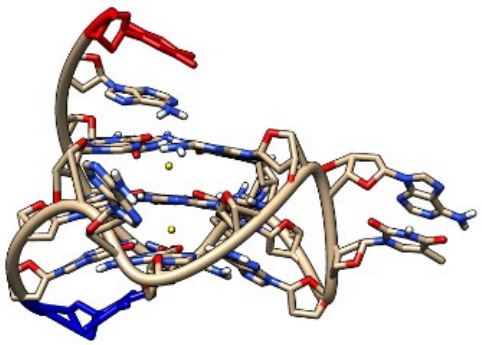
In an MD simulation, the electrostatic interactions utilizing this strategy are ascertained utilizing framework focuses. Every framework point is assigned a dielectric constant. The solute and solvent framework focuses can be recognized by evaluating the solvent available zone.

CHAPTER III

RESULTS AND DISCUSSION

3.1 Simulation protocols

We constructed the DNA-ligand systems using the c-MYC G-quadruplex (PBD ID: 2LBY) as a template.[30] The ligand molecule present in the PDB structure was removed to allow for docking of the DMZ. From the PDB, two variants of c-MYC were modeled, the c-MYC 1-2-1 and c-MYC-1-6-1 based on the sequences shown in Figures 3.1 to 3.2.

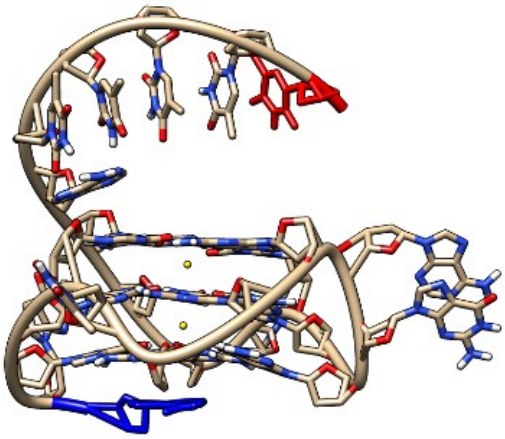


1 **10** **19**

5′ -d(TAGGGAGGGTAGGGAGGGT)-3′

Figure 3.1 Structure of PDB ID: 2LBY c-MYC G-quadruplex

Structure of human parallel c-MYC G-quadruplex DNA used in this study. (A), parallel c-MYC G4 DNA (PDB ID: 2LBY) with K^+ ions. 5′ and 3′ of the DNA chain are indicated by a red and blue, respectively. K^+ ions are indicated by yellow balls.

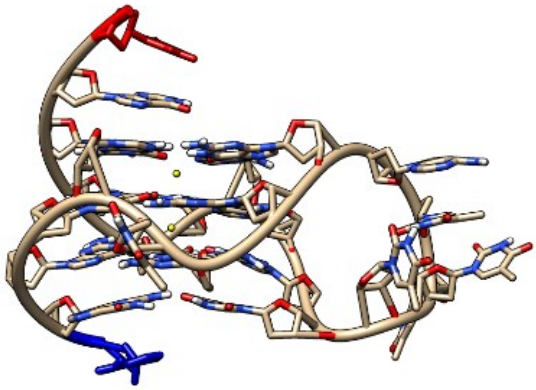


1 **10** **24**

5′-d(TTTTTAGGGTGGGGAGGGTGGGGA)-3′

Figure 3.2 Structure of c-MYC 1-2-1 G-quadruplex

Parallel c-MYC-1-2-1 G4 DNA with K^+ ions. 5′ and 3′ of the DNA chain are indicated by a red and blue, respectively. K^+ ions are indicated by yellow balls.



1 **10** **24**

5′-d(TGGGGAGGGTTTTTAGGGTGGGGA)-3′

Figure 3.3 Structure of c-MYC 1-6-1 G-quadruplex

Parallel c-MYC-1-6-1 g-quadruplex with K^+. 5′ and 3′ of the DNA chain are indicated by a red and blue, respectively. K^+ ions are indicated by yellow balls.

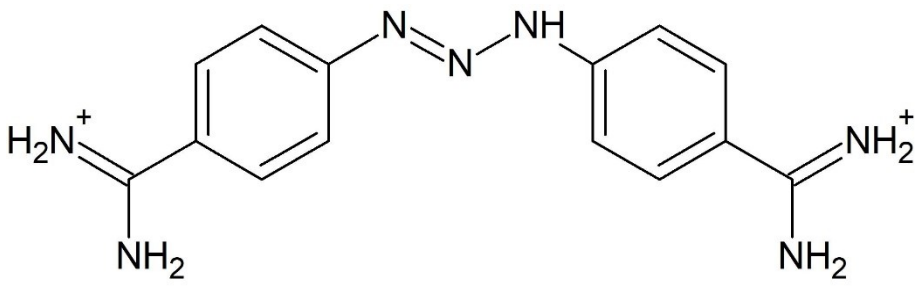


Figure 3.4 2D Structure of DMZ

The models were constructed by taking the crystal structure of a G4 core from the protein data bank and modifying the loops to match the sequence of the oligonucleotides used for the experiments. Two potassium ions were manually added to the cavity between consecutive G-quartets of the parallel structures as the K^+ ions are needed to stabilize the tetrad. A short MD simulation was performed to acquire the starting structure for docking using Autodock 4.2.[37] We docked DMZ to the different c-MYC G4s, while AutodockTools 1.5.6 was employed to create the AutoGrid points of the DMZ-c MYC G-quadruplexes. The grid maps were established by centering the grid box on the entire molecule to locate all possible binding sites. The grid consisted of 126 x 126 x 126 grid points with a 0.375 Å spacing, which allowed the ligand to explore the whole conformational space. The Lamarckian genetic algorithm was used for docking with a maximum number of 2.5 x 10^6 energy evaluations with an initial population of 500 randomly placed individuals having a mutation rate of 0.02 along with a maximum number of 27000 generations. A crossover rate of 0.8 and 300 iterations of local search were used. All other docking parameters were left at default values. For the ligand, 150 independent docking runs were carried out. The output of the docking conformations was clustered based on RMSD between the cartesian coordinates of the ligand atoms (cutoff = 2.0 Å) and was ranked based on the scoring function. The best-docked structures were selected based on the lowest binding

energies and preferential binding sites based on experimental work done by Mikek *et al*.[34] The chemical structure of the ligand used in this study is shown in Fig. 3.4. The DMZ was optimized at the HF/6-31G* level using Spartan 14.[63] Atom types and bond types were assigned using the antechamber program of AMBER 14.[64] The conformer with the lowest energy was used for the molecular docking. The partial charges and electrostatic potential of the DMZ molecule were obtained at the HF/6-31G* level after geometry optimization at the same level; the electrostatic potential using the RESP (Restrained Electrostatic Potential)[65,66] method determined the partial charges. Other force field parameters were taken from the AMBER GAFF force field.[45] The DNA-ligand systems were solvated in a truncated octahedron water box with 10 Å of water molecules between the system and the edges of the box added. Additionally, K^+ ions were added as counter ions to neutralize the system. To simulate a salt concentration of 0.15 M, we added additional K^+ and Cl^- ions. A refined version of the ff144SB force field with the parmbsc0 nucleic acid parameters[67] was applied to represent the DNA fragments, along with the TIP3P model[57] for the water molecules. The system was minimized by 500 steepest descent energy minimizations and then 500 steps of conjugate gradient minimization with constraints of 50 kcal mol^{-1} Å and 500 kcal mol^{-1} Å on the central K^+ ions and the G4-DNA, respectively. The entire system was then subjected to 3000 steps of steepest descent followed by 500 steps of conjugate gradient minimization. The system was heated for 100 ps from 10 K to 300 K under NVT conditions with a restraint of 5.0 kcal/mol/$Å^2$ on the G4 DNA. Next, it was equilibrated for 900 ps at 300 K under NPT conditions. During the equilibration and sampling process, the hydrogens were constrained using the SHAKE[68] algorithm, and a constant temperature was maintained using Langevin dynamics with a collision frequency of 1.0 ps^{-1} The electrostatic interactions were handled using a Particle Mesh Ewald (PME) summation[55] under periodic boundary

conditions with a 10 Å cutoff. All the dynamics were performed with a time step of 2 fs. Under NPT conditions, the unrestrained MD simulations were performed for 100 ns using the CUDA version of PMEMD.[69] The same MD simulation protocol was followed for the ligand-free G4 DNA. The output trajectory files were saved every 1 ps for further analysis, with the last 50 ns of the trajectory chosen for the MM-PBSA calculation.

3.2 MM/PBSA

The binding free energies were calculated using the MM-PBSA approach. [35,70,71] The dielectric constants were set to 1 for the solute and 80 for the surrounding solvent molecules. The K^+ radius was set to 1.33 Å.[72] A total of 400 snapshots were taken from the last 50 ns using the single trajectory approach. For each snapshot, the free energy was calculated for each species (complex, ligand, and quadruplex) using equation 3.1,

$$\Delta G_{bind} = G_{complex} - G_{DNA} - G_{ligand} \tag{3.1}$$

where $G_{complex}$, G_{DNA} and G_{ligand} are complex, G4, and ligand free energies, respectively. Each was calculated using

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{solv} - T\Delta S \tag{3.2}$$

The enthalpy change ΔE_{MM} in the gas phase upon ligand binding, is obtained by the sum of the bonded terms (bond energy, angle energy, and torsional energy) and non-bonded terms (van der Waals energy and electrostatic energy), ΔG_{solv} is the solvation free energy, and $-T\Delta S$ is the entropic contribution at temperature T, which T was 300 K. The molecular mechanics free energy was calculated from equation 3.3.

$$E_{MM} = \Delta E_{bond} + \Delta E_{angle} + \Delta E_{torsion} + \Delta E_{vdw} + \Delta E_{EEL} \tag{3.3}$$

The solvation free energy was calculated using equation 3.4 where ΔG_{PB} was the polar contribution to solvation obtained by solving the Poisson-Boltzmann equation for the MM-PBSA method.

$$\Delta G_{solv} = \Delta G_{PB} + \Delta G_{np} \quad (3.4)$$

ΔG_{np} is the nonpolar contributions to solvation, which is computed by equation 3.5 where γ is the surface tension that was set to 0.0072 kcal/ (mol- Å^2) and β was a constant = 0.00 kcal/(mol- Å^2).

$$\Delta G_{np} = \gamma * \Delta SASA + \beta \quad (3.5)$$

SASA is the solvent accessible surface area (Å^2) that is used to derive the nonpolar free energy *(ΔG_{np})* as per equation 3.5 and estimated using the MOLSURF algorithm. The solvent probe radius was set to 1.4 Å to define the electric boundary around the molecular surface. Note that since the solute conformational entropy is not included in our analysis, the binding energies by MMGPSA generally over-estimate the true binding free energy *(i.e.* the binding affinity). However, when the solute conformational entropies in different binding poses are comparable, the relative binding free energy can be estimated from the relative MMGBSA binding energies.[73]

3.3 Results and discussion

In the present study, DMZ was docked into different c-MYC G-quadruplex variants. It was revealed that DMZ, which has shown to bind to the minor groove of AT-rich DNA with a micromolar dissociation constant, binds to G-quadruplexes with a nanomolar dissociation constant. Therefore it serves as an attractive scaffold for developing effective G-quadruplex ligands.[5,74] Each DMZ was docked with G4s, and docking conformations with the lowest energy of the best cluster were taken for further analysis. To explore other possible binding modes, we selected binding poses that included groove binding, top stacking, bottom stacking, and loop

binding to explore the possibilities of other distinct binding modes. However, it is important to note that the scoring function in the docking program does not provide accurate binding free energies. Therefore, it is common to validate binding poses with a more accurate method, the MM/PBSA method.

3.3.1 Molecular docking

Molecular docking employing Autodock 4.2[38] provided the most suitable poses for ligand interactions with the c-MYC G4 DNA. For each ligand, 150 docked poses were generated, and the maximum number of poses and energy showed that the binding site of DMZ can occur at the loops, grooves, and both the top and the bottom of the tetrads. Consequently, 100 ns of unrestrained MD simulations were carried out on six G4–ligand complexes for each quadruplex with DMZ bound at different positions. To compare the ligand-induced conformational changes in the c-MYC G4 DNA, MD simulations of 2LBY were carried out for 100 ns in the absence and presence of the ligand.

3.3.2 Three drug binding modes were observed.

Starting from the docked conformations, 100 ns MD simulation were carried out on each binding modes. The convergence of the simulations was confirmed through the RMSD and binding free energies. Cluster analyses of the MD trajectories were carried out using USCF Chimera[75] and the CPPTRAJ module in AmberTools 14. All solvent and solute atoms except the G4 channel ions were removed from each of the MD trajectories. The identification of major binding modes was achieved via a clustering method, where similar structures are grouped into the same cluster. There are quite a few metrics that can be used for such clustering. However, we used an RMSD metric, which is essentially a measure of the similarity between two structures or

binding modes. The best representative conformers from the six clusters were extracted to show the non-covalent interactions. The confirmed binding modes for each of the G4s are shown in Figure 3.5, 3.6, and 3.7.

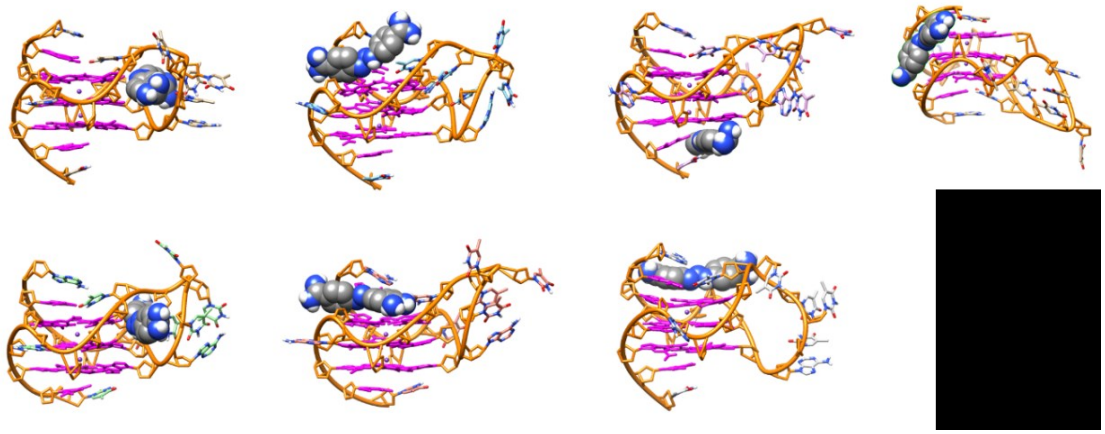


Figure 3.5 Best representative structure from MD simulations (100 ns) of DMZ with c-MYC 1-6-1.

The backbone of the DNA is shown as a cartoon, the atoms are in stick representation, and the ligands are sphere representation and colored by element with carbon being grey, nitrogen being blue and hydrogen being white. K^+ ions are shown as non-bonded purple spheres.

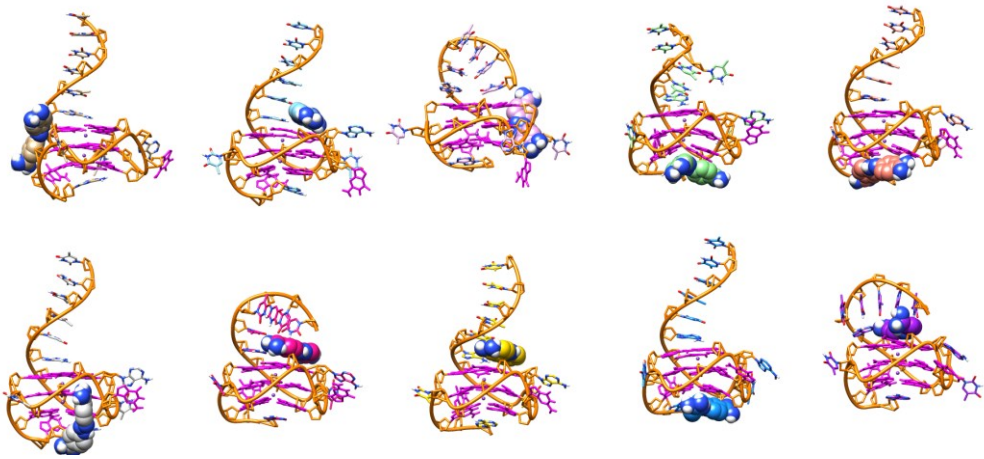


Figure 3.6 Best representative structure from MD simulations (100 ns) of DMZ with c-MYC 1-2-1.

The backbone of the DNA is shown as a cartoon, the atoms are in stick representation, and the ligands are sphere representation and colored by element with carbon being grey, nitrogen being blue and hydrogen being white. K^+ ions are shown as non-bonded purple spheres.

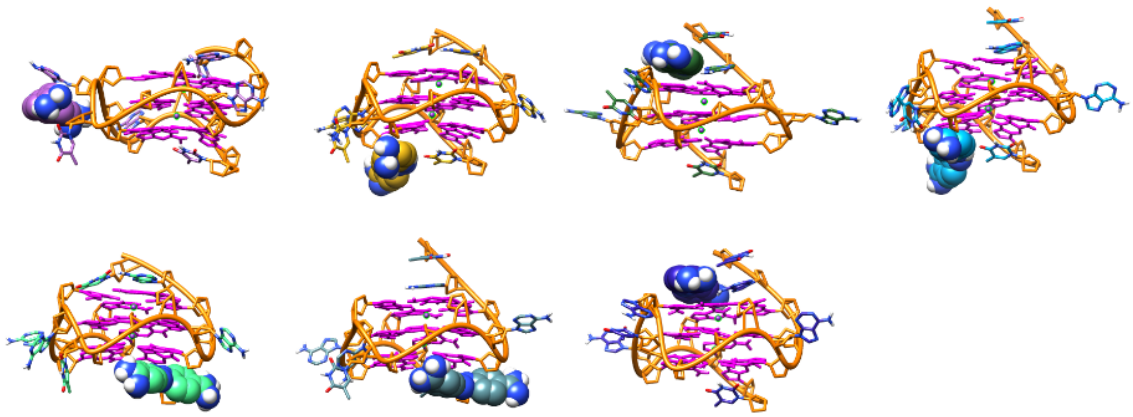


Figure 3.7 Best representative structure from MD simulations (100 ns) of DMZ with c-MYC 2LBY.

The backbone of the DNA is shown as a cartoon, the atoms are in stick representation, and the ligands are sphere representation and colored by element with carbon being grey, nitrogen being blue and hydrogen being white. K^+ ions are shown as non-bonded purple spheres.

3.3.2.2 End Stacking mode

End stacking refers to the binding of DMZ to the terminal G-quartets. This binding mode is shown in all three of the G-quadruplexes. This binding mode is shown to be favorable for all complexes, as illustrated in Table 3.1. For the c-MYC 1-6-1, the bottom binding mode exhibits larger fluctuations, which suggest that DMZ binding to the bottom of the G-quadruplex is not as favorable as the top binding mode. However, both binding modes are accessible based on the MMPBSA calculations in which the binding energy for this mode varied from -40 to -48 kcal/mol. The same trend is observed for the c-MYC 1-2-1 where energy values varied from -33 to -40 kcal/mol. The crystal structure showed these two binding modes predominantly with binding energy values that varied from -26 to -35 kcal/mol. The end stacking mode seems the most favorable for the c-MYC 1-2-1 and the 2LBY based on the binding energies and structural stability suggested from the RMSD graphs. Although there exists no crystal structure of c-MYC binding with DMZ available, a crystal structure of c-MYC G4 with quindoline (PDB ID: 2L7V) showed the end stacking binding mode with quindoline and a computational study of thioflavin T (ThT).[76] The quindoline and ThT has a similar shape to that DMZ, which suggests that two DMZs can stack side by side on the terminal G-quartets, which is consistent with our observations. The end stacking position is most favorable for the c-MYC 1-2-1 and 2LBY. For the 2LBY, this is no surprise as it is not capped by terminal bases as the c-MYC 1-2-1 is or possesses edgewise loops as the c-MYC 1-6-1 does. Interestingly, for the c-MYC 1-2-1, the longer 5` tail can undergo a conformational change when DMZ is bounded by folding over the G-tetrad in the presence of the ligand.

3.3.2.3 Groove Binding

In the groove binding mode, DMZ inserts into the groove formed by lateral loops on the G4, similarly to those seen with B-DNA. This binding mode was observed in the c-MYC 1-6-1 and c-MYC 1-2-1 but was not observed in the 2LBY. The binding energy for the c-MYC 1-6-1 and 1-2-1 were -34 and -36 kcal/mol respectively, which was similar between the two complexes. However, with the 2LBY, this binding mode was observed from the docking, but the DMZ moves from the groove and into an end-stacking binding mode during the simulation, suggesting that DMZ binding in the groove is not as favorable binding site for the c-MYC G4 when there is no terminal bases or large enough edgewise loops.

3.3.2.4 Loop insertion

This binding mode was observed in the c-MYC 1-6-1 and 2LBY but was not a favorable binding mode in the c-MYC 1-2-1. The c-MYC 1-2-1 docked structure of the DMZ bounded included a DMZ docked in the loop of the c-MYC 1-2-1, but it was not favorable based on the MMPBSA calculations and its large fluctuations. For the 2LBY, the binding energy was -33 kcal/mol. While this mode was observed in the c-MYC 1-2-1 it was not favorable based on the MMPBSA calculation, so it was excluded. Perhaps the size of the loop in the c-MYC 1-2-1 will not allow for loop insertion. For the c-MYC 1-6-1, this binding mode showed large fluctuations, but its MMPBSA energy (-53 kcal/mol) was the most favorable out of all binding modes observed in the quadruplexes, making this binding pose the most favorable of the three for the c-MYC 1-6-1. This suggests that the conformers with large loops are most accessible for binding interactions with the side chains of the ligands, and therefore, they can be harnessed for the design of novel ligands using DMZ as a template.[77]

Table 3.1 Binding free energies for each complex (kcal/mol)

c-*MYC* 1-6-1 G4 DNA								
Ligand	**Site**	ΔE_{ELE}	ΔE_{vdw}	ΔE_{PBSUR}	ΔE_{PBCAL}	ΔE_{PBSOL}	ΔE_{PBELE}	ΔE_{PBTOT}
DMZ Aa	**Loop**	**-1219.5 ± 14.7**	**-37.9 ± 2.9**	**-3.6 ± 0.0**	**1207.2 ± 11.5**	**1203.5 ± 11.5**	**-12.3 ± 5.2**	**-53.3 ± 4.2**
DMZ Ab	**Bottom 3'**	**-1117.1 ± 27.3**	**-33.1 ± 4.1**	**-3.2 ± 0.0**	**1104.9 ± 21.2**	**1101.6 ± 21.1**	**-12.1 ± 7.9**	**-48.0 ± 5.5**
DMZ Ac	**Groove**	**-966.2 ± 151.8**	**-27.9 ± 6.4**	**-2.7 ± 0.5**	**961.9 ± 151.6**	**959.2 ± 151.0**	**4.3 ± 3.4**	**-34.4 ± 7.4**
DMZ Ad	**Top 5'**	**-992.6 ± 21.5**	**-31.9 ± 2.7**	**-2.8 ± 0.1**	**986.7 ± 20.2**	**983.9 ± 20.1**	**-5.9 ± 3.4**	**-40.1 ± 3.4**
DMZ Ae	**Loop**	**-856.0 ± 437.1**	**-25.1 ± 17.8**	**-2.4 ± 1.7**	**853.4 ± 437.6**	**850.9 ± 435.9**	**-2.6 ± 3.6**	**-29.7 ± 19.0**
DMZ Af	**Bottom 3'**	**-1145.1 ± 33.8**	**-26.6 ± 3.5**	**-2.9 ± 0.1**	**1132.0 ± 28.5**	**1129.0 ± 28.4**	**-13.0 ± 7.9**	**-42.1 ± 7.0**
c-*MYC* 1-2-1 G4 DNA								
Ligand	**Site**	ΔE_{ELE}	ΔE_{vdw}	ΔE_{PBSUR}	ΔE_{PBCAL}	ΔE_{PBSOL}	ΔE_{PBELE}	ΔE_{PBTOT}
DMZ Ba	**Groove**	**-1022.8 ± 11.5**	**-27.9 ± 2.5**	**-2.9 ± 0.0**	**1017.2 ± 10.2**	**1014.3 ± 10.23**	**-5.6 ± 3.0**	**-36.4 ± 2.7**
DMZ Bb	**Top 5'**	**-1064.9 ± 54.6**	**-31.5 ± 3.4**	**-2.7 ± 0.2**	**1059.8 ± 53.2**	**1057.0 ± 53.1**	**-5.1 ± 5.1**	**-39.4 ± 4.7**
DMZ Bc	**Bottom 3'**	**-972.9 ± 15.5**	**-25.7 ± 2.8**	**-2.4 ± 0.0**	**966.8 ± 13.3**	**964.3 ± 13.31**	**-6.1 ± 4.4**	**-34.3 ± 3.8**
DMZ Bd	**Groove**	**-889.6 ± 20.3**	**-14.5 ± 2.9**	**-1.8 ± 0.2**	**875.8 ± 21.8**	**874.0 ± 21.68**	**-13.7 ± 4.2**	**-30.0 ± 2.6**
DMZ Be	**Bottom 3'**	**-962.4 ± 13.2**	**-32.6 ± 2.5**	**-2.6 ± 0.0**	**964.3 ± 13.1**	**961.7 ± 13.08**	**1.9 ± 3.0**	**-33.3 ± 3.4**
DMZ Bf	**Top 5'**	**-1123.1 ± 13.9**	**-32.7 ± 2.6**	**-3.51 ± 0.1**	**1122.8 ± 14.9**	**1119.3 ± 14.92**	**-0.2 ± 5.2**	**-36.5 ± 4.9**
c-*MYC* 2LBY G4 DNA								
Ligand	**Site**	ΔE_{ELE}	ΔE_{vdw}	ΔE_{PBSUR}	ΔE_{PBCAL}	ΔE_{PBSOL}	ΔE_{PBELE}	ΔE_{PBTOT}
DMZ Ca	**Loop**	**-851.5 ± 14.4**	**-33.4 ± 2.83**	**-3.1 ± 0.0**	**854.6 ± 13.0**	**851.5 ± 13**	**3.1 ± 4.8**	**-33.4 ± 3.8**
DMZ Cb	**Bottom 3'**	**-810.4 ± 72.7**	**-22.4 ± 3.38**	**-2.2 ± 0.2**	**805.0 ± 72.7**	**802.7 ± 72.4**	**-5.3 ± 3.0**	**-30.0 ± 4.0**
DMZ Cc	**Top 5'**	**-801.3 ± 77.8**	**-32.5 ± 4.80**	**-2.7 ± 0.3**	**801.7 ± 78.3**	**798.9 ± 78.0**	**0.3 ± 4.6**	**-34.9 ± 5.1**
DMZ Cd	**Bottom 3'**	**-802.0 ± 14.2**	**-23.6 ± 2.20**	**-2.4 ± 0.1**	**800.3 ± 13.4**	**797.9 ± 13.4**	**-1.7 ± 4.6**	**-27.8 ± 4.3**
DMZ Ce	**Bottom 3'**	**-829.6 ± 46.7**	**-24.4 ± 3.05**	**-2.5 ± 0.1**	**823.2 ± 46.0**	**820.7 ± 45.8**	**-6.4 ± 3.7**	**-33.4 ± 3.4**
DMZ Cf	**Top 5'**	**-782.7 ± 12.7**	**-27.1 ± 1.54**	**-2.2 ± 0.0**	**785.9 ± 13.2**	**783.7 ± 13.1**	**3.2 ± 2.4**	**-26.1 ± 2.1**

The binding free energy components of DMZ binding to c-MYC G4s in (kcal/mol). Four sites are viable for DMZ binding: loop binding, groove binding, and end stacking at the 5' and 3' ends, as demonstrated by the negative value of the total binding free energy. The loop and groove binding modes showed similar binding strength to all the G4s examined.

3.3.3 Overall structural stability

The convergence of the MD simulations was monitored by the root mean square deviations (RMSDs) of the structures of the complex and its components with respect to the first frame of the simulation. The comparison of RMSD graphs for the G-quartets in comparison with the ligand-free G-quartet showed that DMZ binding stabilizes the G4 efficiently as shown in Figures 3.8 to 3.25. The backbone atoms converged towards the last 40-50 ns of the simulations. The RMSD graphs show that DMZ is more flexible in the groove and loop binding modes but more rigid in the end stacking mode when complexed with the G4s. To probe the dynamics of the nucleotides in the G4, root-mean-square fluctuations (RMSF) were calculated as illustrated in Figures 3.26 to 2.28 For the c-MYC 1-6-1, the RMSF graph suggests that the loops and flanking nucleotides fluctuated less in the complex with DMZ than in the ligand-free G4. For c-MYC 1-2-1, the 5' end nucleotides exhibited less fluctuation in comparison to the ligated G4. For the c-MYC crystal structure, there was more fluctuation in the 5' end in contrast to the 3' end which shows little fluctuation when complexed with DMZ. From the RMSF values, we can see that the structure fluctuations originate primarily from the loop and flanking nucleotides, while the G-tetrads are quite stable in the simulations.

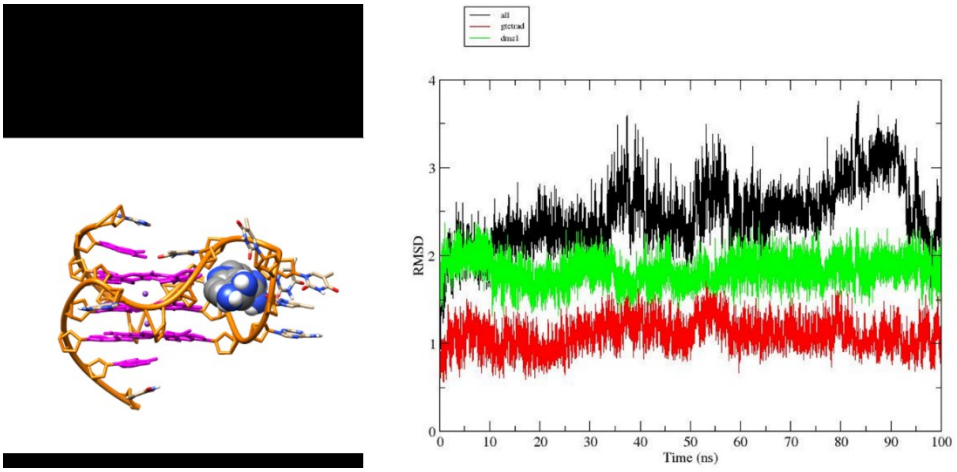


Figure 3.8 RMSD of DMZ Aa with c-MYC 1-6-1 G4.

RMSD graphs of the c-MYC 1-6-1 G4 DNA and DMZ Aa complex during the 100 ns of MD simulations. The black, red and green lines indicate the RMSDs for the DNA backbone, G-quartet, and ligand, respectively.

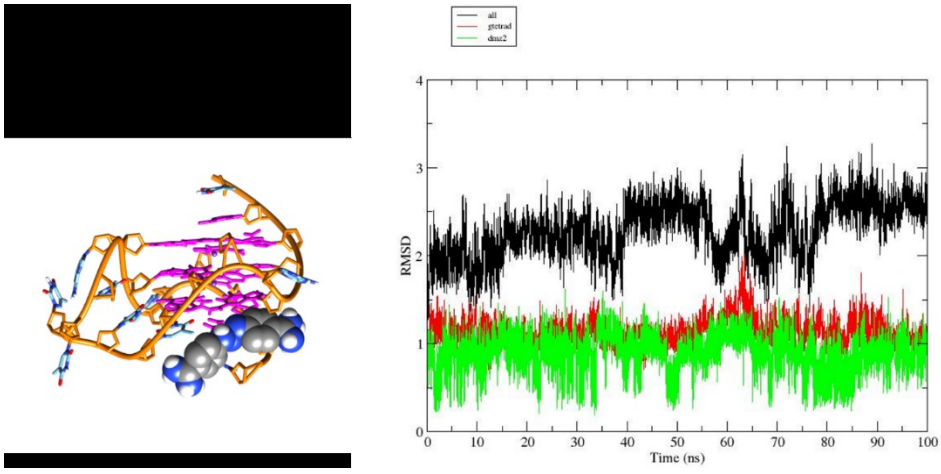


Figure 3.9 RMSD of DMZ Ab with c-MYC 1-6-1 G4.

RMSD graphs of the c-MYC 1-6-1 G4 DNA and DMZ Ab complex during the 100 ns of MD simulations. The black, red and green lines indicate the RMSDs for the DNA backbone, G-quartet, and ligand respectively.

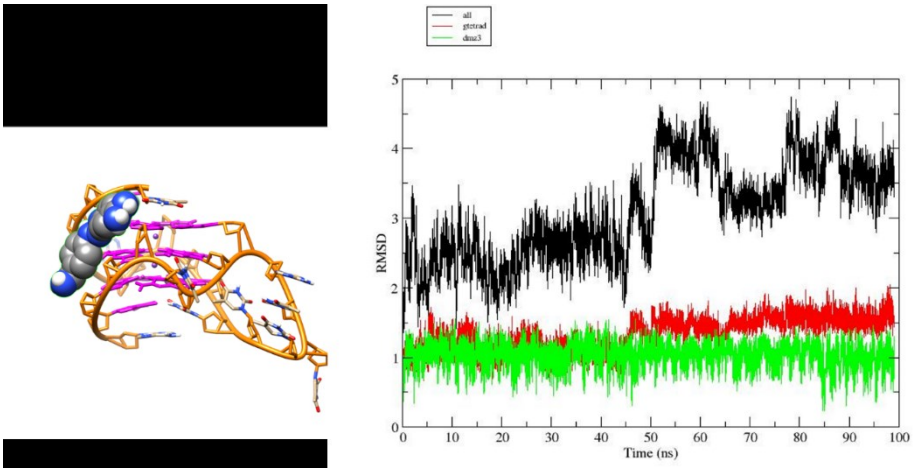


Figure 3.10 RMSD of DMZ Ac with c-MYC 1-6-1 G4.

RMSD graphs of the c-MYC 1-6-1 G4 DNA and DMZ Ac complex during the 100 ns of MD simulations. The black, red and green lines indicate the RMSDs for the DNA backbone, G-quartet, and ligand respectively.

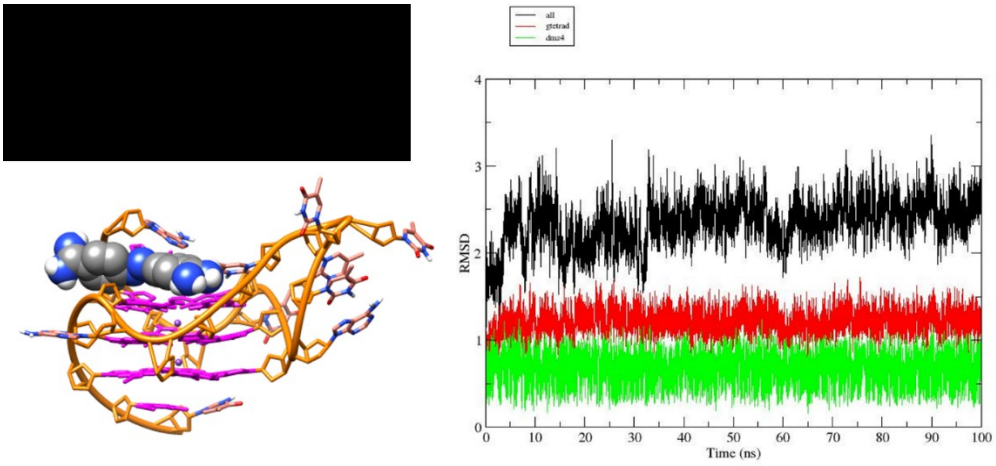


Figure 3.11 RMSD of DMZ Ad with c-MYC 1-6-1 G4.

RMSD graphs of the c-MYC 1-6-1 G4 DNA and DMZ Ad complex during the 100 ns of MD simulations. The black, red and green lines indicate the RMSDs for the DNA backbone, G-quartet, and ligand respectively.

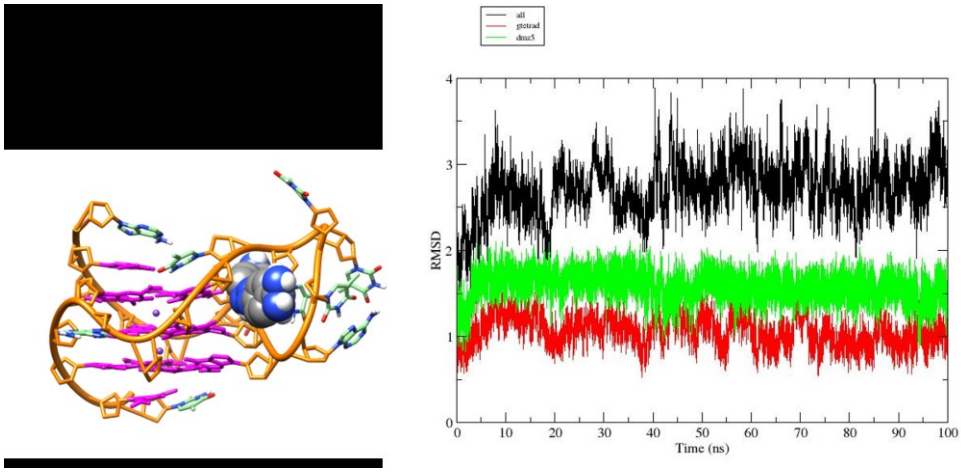


Figure 3.12 RMSD of DMZ Ae with c-MYC 1-6-1 G4.

RMSD graphs of the c-MYC 1-6-1 G4 DNA and DMZ Ae complex during the 100 ns of MD simulations. The black, red and green lines indicate the RMSDs for the DNA backbone, G-quartet, and ligand respectively.

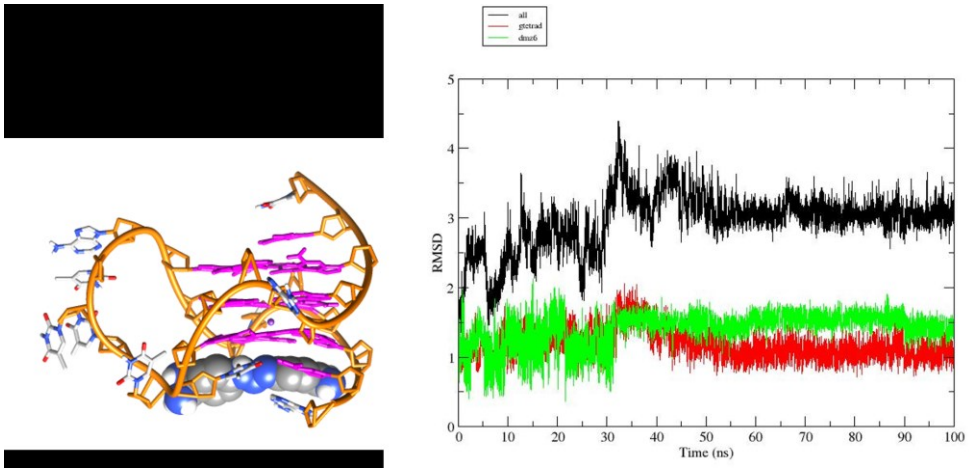


Figure 3.13 RMSD of DMZ Af with c-MYC 1-6-1 G4.

RMSD graphs of the c-MYC 1-6-1 G4 DNA and DMZ Af complex during the 100 ns of MD simulations. The black, red and green lines indicate the RMSDs for the DNA backbone, G-quartet, and ligand respectively.

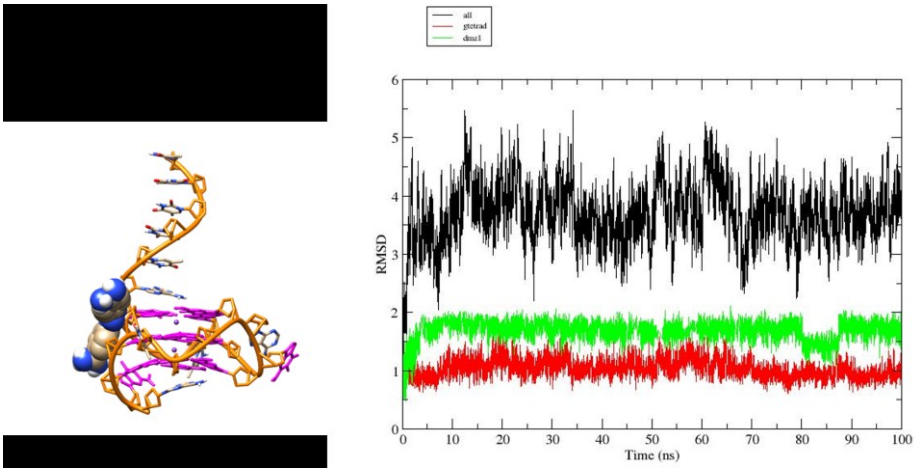


Figure 3.14 RMSD of DMZ Ba with c-MYC 1-2-1 G4.

RMSD graphs of the c-MYC 1-2-1 G4 DNA and DMZ Ba complex during the 100 ns of MD simulations. The black, red and green lines indicate the RMSDs for the DNA backbone, G-quartet, and ligand respectively.

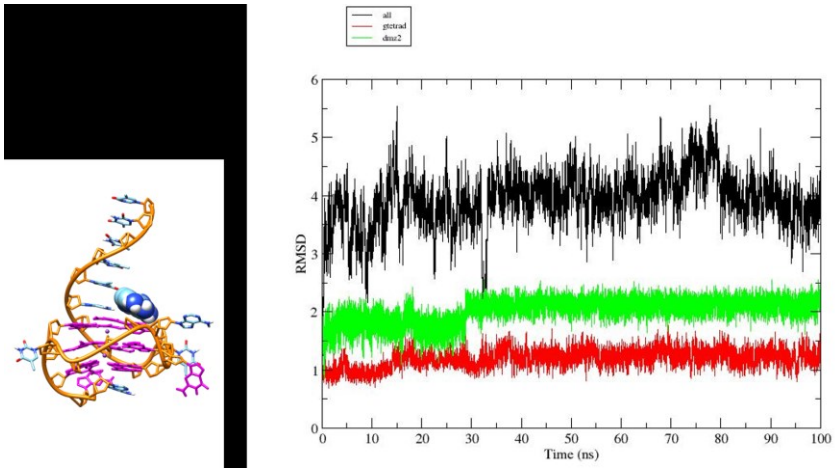


Figure 3.15 RMSD of DMZ Bb with c-MYC 1-2-1 G4.

RMSD graphs of the c-MYC 1-2-1 G4 DNA and DMZ Bb complex during the 100 ns of MD simulations. The black, red and green lines indicate the RMSDs for the DNA backbone, G-quartet, and ligand respectively.

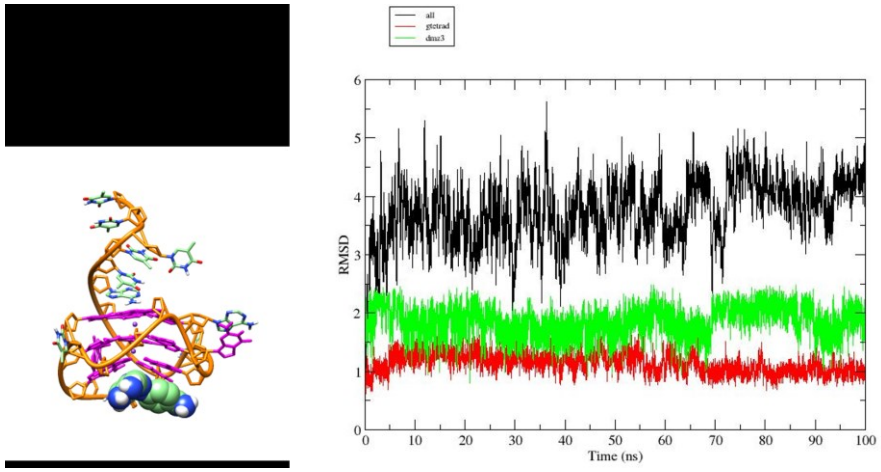


Figure 3.16 RMSD of c-MYC DMZ Bc with 1-2-1 G4

RMSD graphs of the c-MYC 1-2-1 G4 DNA and DMZ Bc complex during the 100 ns of MD simulations. The black, red and green lines indicate the RMSDs for the DNA backbone, G-quartet, and ligand respectively.

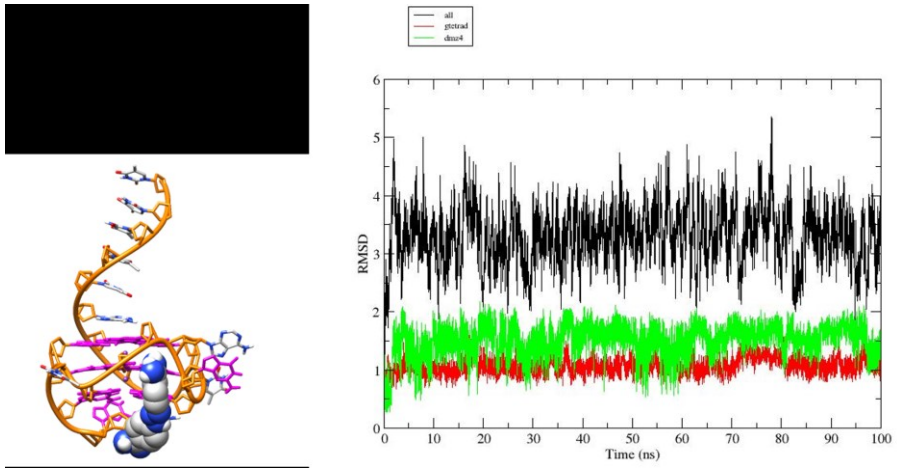


Figure 3.17 RMSD of c-MYC DMZ Bd with 1-2-1 G4

RMSD graphs of the c-MYC 1-2-1 G4 DNA and DMZ Bd complex during the 100 ns of MD simulations. The black, red and green lines indicate the RMSDs for the DNA backbone, G-quartet, and ligand respectively.

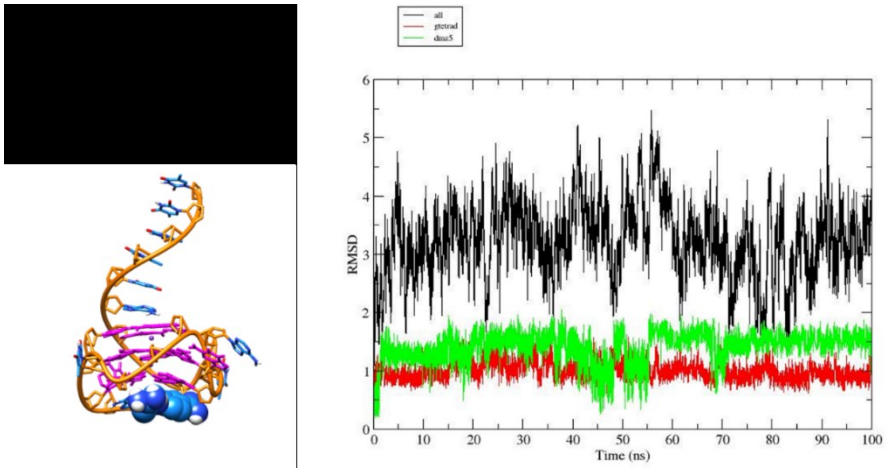


Figure 3.18 RMSD of c-MYC DMZ Be with 1-2-1 G4

RMSD graphs of the c-MYC 1-2-1 G4 DNA and DMZ Be complex during the 100 ns of MD simulations. The black, red and green lines indicate the RMSDs for the DNA backbone, G-quartet, and ligand respectively.

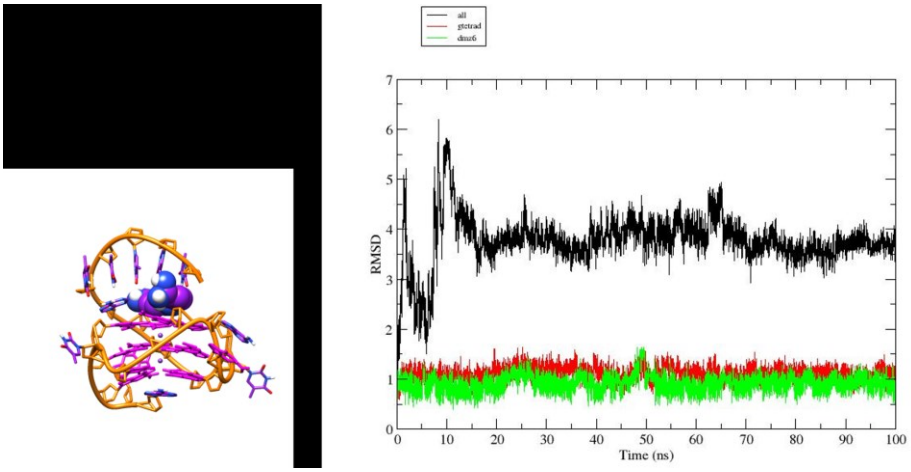


Figure 3.19 RMSD of c-MYC DMZ Bf with 1-2-1 G4.

RMSD graphs of the c-MYC 1-2-1 G4 DNA and DMZ Bf complex during the 100 ns of MD simulations. The black, red and green lines indicate the RMSDs for the DNA backbone, G-quartet, and ligand respectively.

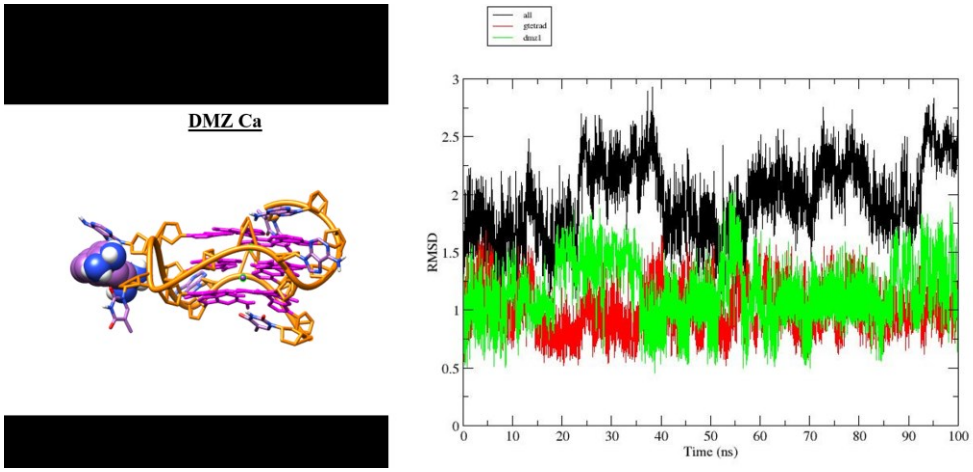


Figure 3.20 RMSD of DMZ Ca with c-MYC 2LBY G4.

RMSD graphs of the c-MYC 2LBY G4 DNA and DMZ Ca complex during the 100 ns of MD simulations. The black, red and green lines indicate the RMSDs for the DNA backbone, G-quartet, and ligand respectively.

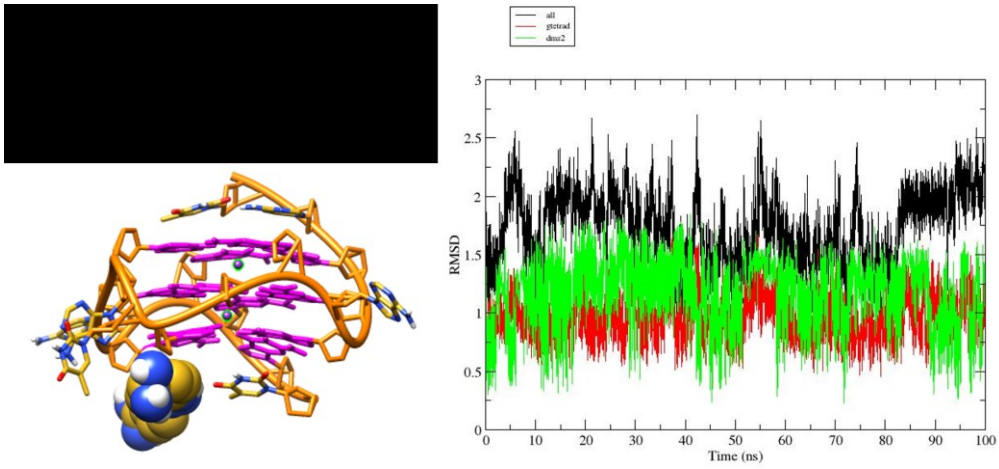


Figure 3.21 RMSD of c-MYC DMZ Cb with c-MYC 2LBY G4.

RMSD graphs of the c-MYC 2LBY G4 DNA and DMZ Cb complex during the 100 ns of MD simulations. The black, red and green lines indicate the RMSDs for the DNA backbone, G-quartet, and ligand respectively.

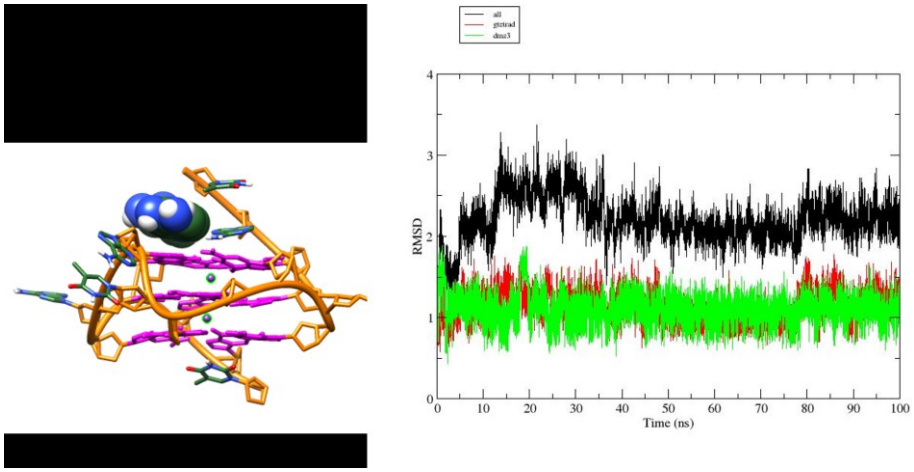


Figure 3.22 RMSD of c-MYC DMZ Cc with c-MYC 2LBY G4.

RMSD graphs of the c-MYC 2LBY G4 DNA and DMZ Cc complex during the 100 ns of MD simulations. The black, red and green lines indicate the RMSDs for the DNA backbone, G-quartet, and ligand respectively.

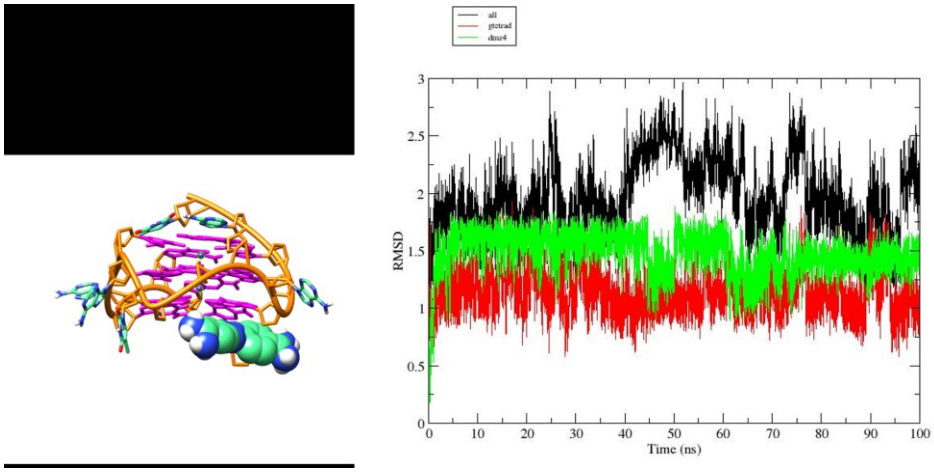


Figure 3.23 RMSD of c-MYC DMZ Cd with c-MYC 2LBY G4.

RMSD graphs of the c-MYC 2LBY G4 DNA and DMZ Cd complex during the 100 ns of MD simulations. The black, red and green lines indicate the RMSDs for the DNA backbone, G-quartet, and ligand respectively.

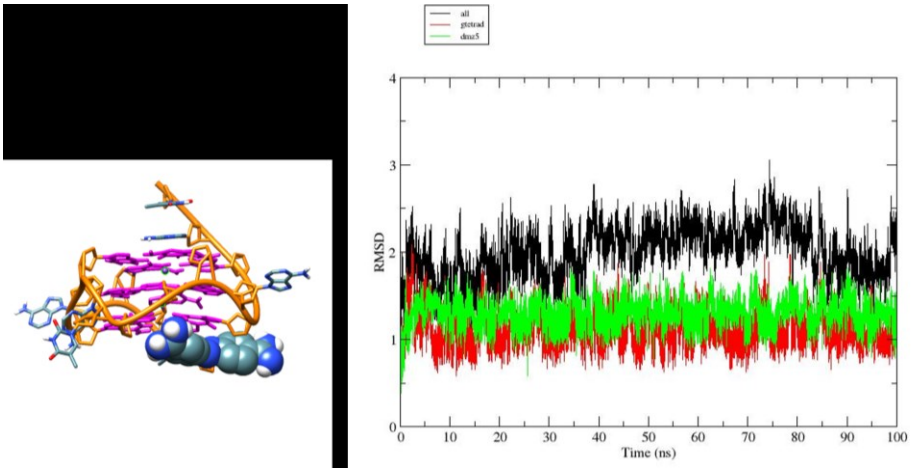


Figure 3.24 RMSD of c-MYC DMZ Ce with c-MYC 2LBY G4.

RMSD graphs of the c-MYC 2LBY G4 DNA and DMZ Ce complex during the 100 ns of MD simulations. The black, red and green lines indicate the RMSDs for the DNA backbone, G-quartet, and ligand respectively.

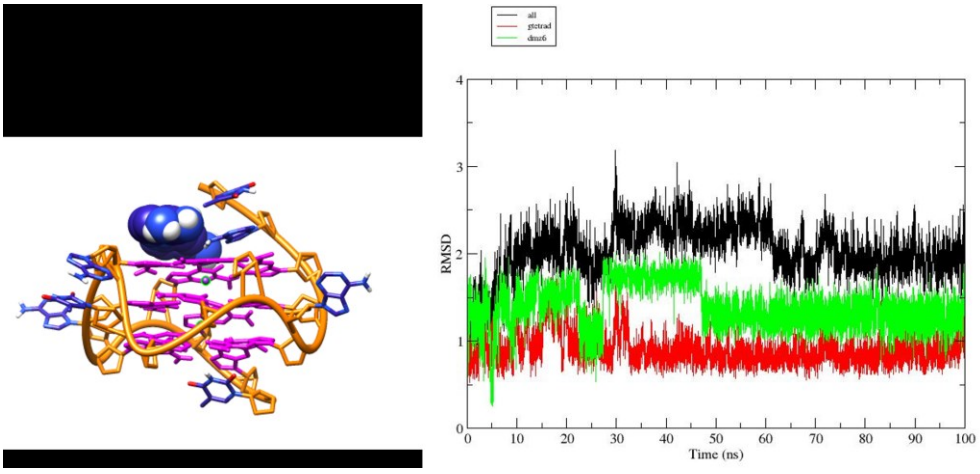


Figure 3.25 RMSD of c-MYC DMZ Cf with c-MYC 2LBY G4.

RMSD graphs of the c-MYC 2LBY G4 DNA and DMZ Cf complex during the 100 ns of MD simulations. The black, red and green lines indicate the RMSDs for the DNA backbone, G-quartet, and ligand respectively.

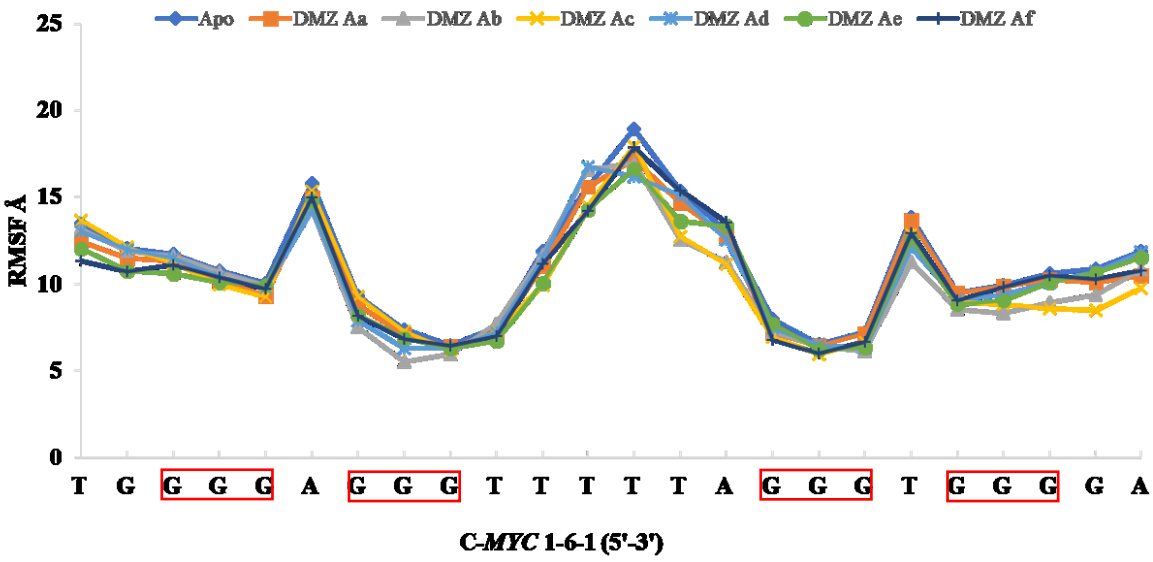


Figure 3.26 RMSF of each c-MYC 1-6-1 nucleotide in complexes and apo as a function of time.

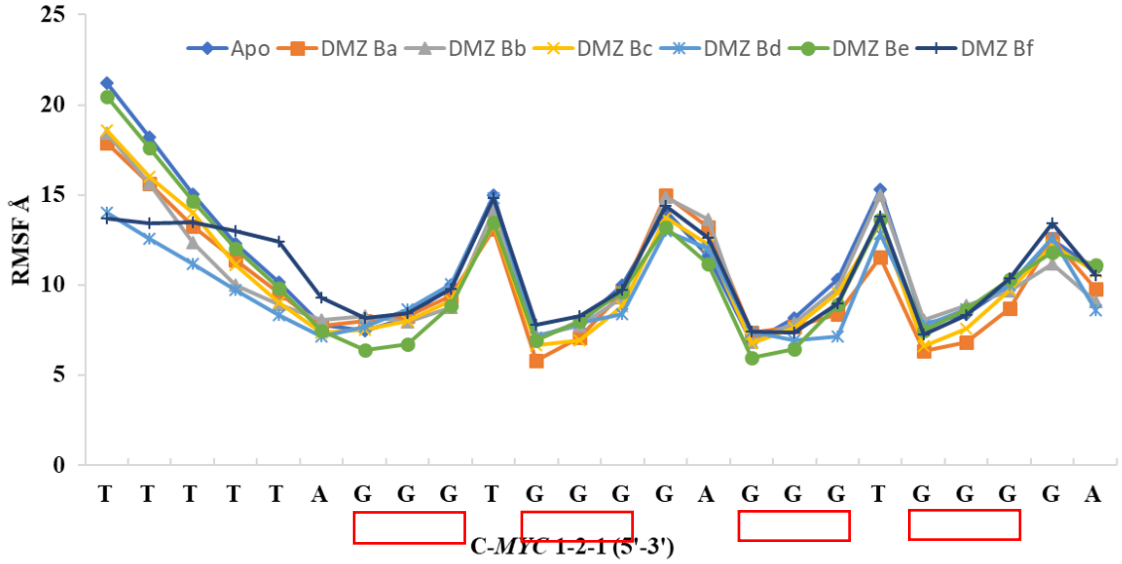


Figure 3.27 RMSF of each c-MYC 1-2-1 nucleotide in complexes and apo as a function of time.

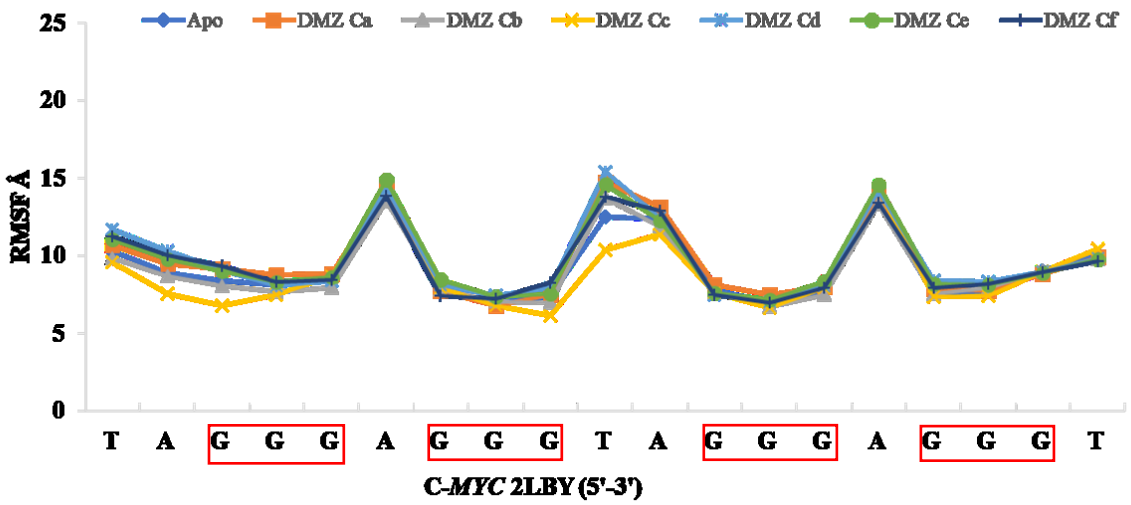


Figure 3.28 RMSF of each c-MYC 2lBY nucleotide in complexes and apo as a function of time.

These observations suggest that, depending on the binding mode of DMZ, it can stabilize and cause conformational changes in loops and flanking nucleotides of the G4 structures as well as minimize the dynamic behavior.

3.3.4 Important noncovalent interactions: hydrogen-bonding and π-π stacking

It is important to note that DNA possesses a negative charge, so ligands with positive charges such as DMZ will bind to them but without selectivity. The stability of the G4 is defined by the hydrogen bonding and the inherent π-π stacking. In a G4, each G-quartet is held together by eight Hoogsten hydrogen bonds (N2-H---N7 and N1-H---O6), which are the main bonds keeping the quartet stable. To study the effects of DMZ binding on G4s, the hydrogen bond occupancies were calculated along all the MD trajectories. For the H-bond analysis, we only considered H-bonds in the G-tetrad with occupation rates greater than 90% for the last 50 ns of the simulation. The analysis indicates that the H-bonds provide stability of the apo and

complexed G-quadruplexes. However, upon DMZ binding, hydrogen bonds appear without as large a frequency, which is concurrent with experimental observations.[5] Also, the π-π stacking interactions contribute significantly to the stability of the complexes. G4s are stable structures in solution, so having a ligand that can promote a conformational change in the G4 may serve as an attractive strategy. From this study, the loops provide a suitable site for DMZ binding, especially those loops rich in thymine bases, as they are more prone to forming hydrogen bonds with ligands such as DMZ.[78]

3.3.5 Binding free energies

The binding free energies were computed by means of the MM-PBSA. The results are listed in Table 3.1. The binding free energies ranged from -26 to -53. kcal/mol. The calculated energies are substantially overestimated, which may be because of the highly charged species involved in this study. Based on the results, the four sites that are viable for DMZ binding are loop binding, groove binding, and end stacking at the 5' and 3' ends, as indicated by the negative value of the total binding free energy. The electrostatic energies (ΔE_{ELE}), which arise from the positive charges on the amidines of the DMZ and negative charges on the phosphate groups of the quadruplexes, are shown to be favorable for all the complexes. Also, the electrostatic energy contribution is strong enough to compensate for the contribution from the polar solvation (ΔPB_{SOLV}). The van der Waals energy (ΔE_{vdw}), which stems from the π-interactions of DMZ with the G-tetrad are shown to provide a favorable contribution. A slightly favorable contribution arose from the nonpolar contribution to the solvation energy (ΔE_{PBSUR}), but the electrostatic term to the solvation energy (ΔE_{PBCAL}) was highly unfavorable. Of all the G4 structures examined in this work, the largest binding free energy observed was from the loop insertion binding mode in the large-looped c-MYC 1-6-1 structure. The DMZ ligand shows the importance of positively

charged ligands having aromatic units for binding and stabilization. The binding entropy can be computed through quasiharmonic analysis or normal mode analysis. However, the binding entropy was not considered here. Our aim is this work was to identify the possible binding modes of DMZ and key interactions that play in role in the stabilizing the DMZ-G-quadruplex structure. It should be noted that a normal mode analysis, disregards the anharmonic contributions, which can lead to systematic errors in calculating the vibrational entropy term, while the quasiharmonic method may fail to converge when estimating the entropy.[59]

3.4 Conclusion

Motivated by recent research using the DMZ scaffold to bind to G-quadruplexes, we have performed unbiased molecular dynamics simulations to study the binding. Based on snapshots from the simulations, the effective binding free energy between DMZ and G-quadruplexes that was computed by the MM-PBSA approach. It is shown in this study that DMZ, which is a potent duplex binder, can in fact bind to G-quadruplexes with high affinity and can do so by three possible binding modes that include end stacking, loop, and groove binding. Our results are consistent with the experimental observations[34] and indicate that ligands such as DMZ can induce a conformational change in the loop and flanking nucleotides. The large flexibility of the loops and flanking nucleotides can be reduced, which serves as an attractive method for drug design. Drugs that interact with loops have recently been proposed to be the most accessible for binding.[79] Our results show that DMZ can bind to different variants of c-MYC depending on its 3-dimensional structure, but it should be noted that it is not specific towards any c-MYC. Research into G4 loops has proposed that ligand binding to the G4s can have a significant degree of loop conformation variability.[80] Of all the G4 structures examined in this work, the largest binding free energy observed was from the loop insertion binding mode in the large-looped c-

MYC 1-6-1 structure. This feature of loops is likely to be common to all quadruplexes with loops, whether they contain single-nucleotide or much longer loops. From the experimental work and theoretical studies, DMZ can serve as a great template for loop specific ligands that can be harnessed for drug design. Also, the study underlines the use of MD simulations in combination molecular docking and MM-PBSA free energy calculations to provide a detailed description of the energetics at the atomistic level for DMZ binding.

REFERENCES

(1) Jemal, A.; Siegel, R.; Ward, E.; Murray, T.; Xu, J.; Thun, M. J. Cancer Statistics, 2007. *CA. Cancer J. Clin.* **2007**, *57* (1), 43–66. https://doi.org/10.3322/canjclin.57.1.43.

(2) Eckstein, N. Platinum Resistance in Breast and Ovarian Cancer Cell Lines. *Journal of Experimental and Clinical Cancer Research*. 2011. https://doi.org/10.1186/1756-9966-30-91.

(3) Hartinger, C. G.; Jakupec, M. A.; Zorbas-Seifried, S.; Groessl, M.; Egger, A.; Berger, W.; Zorbas, H.; Dyson, P. J.; Keppler, B. K. KP1019, a New Redox-Active Anticancer Agent - Preclinical Development and Results of a Clinical Phase I Study in Tumor Patients. *Chemistry and Biodiversity*. 2008, pp 2140–2155. https://doi.org/10.1002/cbdv.200890195.

(4) Dickerson, M.; Howerton, B.; Bae, Y.; C. Glazer, E. Light-Sensitive Ruthenium Complex-Loaded Cross-Linked Polymeric Nanoassemblies for the Treatment of Cancer. *J. Mater. Chem. B* **2016**, *4* (3), 394–408. https://doi.org/10.1039/c5tb01613d.

(5) Zhou, J.; Le, V.; Kalia, D.; Nakayama, S.; Mikek, C.; Lewis, E. A.; Sintim, H. O. Diminazene or Berenil, a Classic Duplex Minor Groove Binder, Binds to G-Quadruplexes with Low Nanomolar Dissociation Constants and the Amidine Groups Are Also Critical for G-Quadruplex Binding. *Mol. Biosyst.* **2014**, *10* (10), 2724–2734. https://doi.org/10.1039/c4mb00359d.

(6) Wilkins, M. H. F.; Stokes, A. R.; Wilson, H. R. Molecular Structure of Deoxypentose Nucleic Acids. In *50 Years of DNA*; Nature Publishing Group, 2016; Vol. 171, pp 84–86. https://doi.org/10.1038/nature01396.

(7) Smithson S. DNA Structure and Sequencing. In *Biology*; 2015; pp 366–373.

(8) Brown, D. G.; Sanderson, M. R.; Skelly, J. V.; Jenkins, T. C.; Brown, T.; Garman, E.; Stuart, D. I.; Neidle, S. Crystal Structure of a Berenil-Dodecanucleotide Complex: The Role of Water in Sequence-Specific Ligand Binding. *EMBO J.* **1990**, *9* (4), 1329–1334. https://doi.org/10.1002/j.1460-2075.1990.tb08242.x.

(9) Düchler, M. G-Quadruplexes: Targets and Tools in Anticancer Drug Design. *Journal of Drug Targeting*. 2012, pp 389–400. https://doi.org/10.3109/1061186X.2012.669384.

(10) Mikek, C. Establishing the Structure Function Relationship of Polypyridyl Ruthenium and Berenil-Type Compounds in the Formation of Complexes with B-DNA and / or G-Quadruplex DNA, 2017, Vol. 0.

(11) Chambers, V. S.; Marsico, G.; Boutell, J. M.; Di Antonio, M.; Smith, G. P.; Balasubramanian, S. High-Throughput Sequencing of DNA G-Quadruplex Structures in the Human Genome. *Nat. Biotechnol.* **2015**, *33* (8), 877–881. https://doi.org/10.1038/nbt.3295.

(12) Huppert, J. L.; Balasubramanian, S. Prevalence of Quadruplexes in the Human Genome. *Nucleic Acids Res.* **2005**, *33* (9), 2908–2916. https://doi.org/10.1093/nar/gki609.

(13) Todd, A. K.; Johnston, M.; Neidle, S. Highly Prevalent Putative Quadruplex Sequence Motifs in Human DNA. *Nucleic Acids Res.* **2005**, *33* (9), 2901–2907. https://doi.org/10.1093/nar/gki553.

(14) Han, H.; Hurley, L. H. G-Quadruplex DNA: A Potential Target for Anti-Cancer Drug Design. *Trends in Pharmacological Sciences*. Elsevier Current Trends April 1, 2000, pp 136–142. https://doi.org/10.1016/S0165-6147(00)01457-7.

(15) Jing, N.; Sha, W.; Li, Y.; Tweardy, W. X. and D. J. Rational Drug Design of G-Quartet DNA as Anti-Cancer Agents. *Current Pharmaceutical Design*. 2005, pp 2841–2854. https://doi.org/http://dx.doi.org/10.2174/1381612054546761.

(16) Kim, N. W.; Piatyszek, M. A.; Prowse, K. R.; Harley, C. B.; West, M. D.; Ho, P. L. C.; Coviello, G. M.; Wright, W. E.; Weinrich, S. L.; Shay, J. W. Specific Association of Human Telomerase Activity with Immortal Cells and Cancer. *Science (80-.).* **1994**, *266* (5193), 2011–2015. https://doi.org/10.1126/science.7605428.

(17) Sun, D.; Thompson, B.; Cathers, B. E.; Salazar, M.; Kerwin, S. M.; Trent, J. O.; Jenkins, T. C.; Neidle, S.; Hurley, L. H. Inhibition of Human Telomerase by a G-Quadruplex-Interactive Compound. *Journal of Medicinal Chemistry*. 1997, pp 2113–2116. https://doi.org/10.1021/jm970199z.

(18) González, V.; Hurley, L. H. The C- *MYC* NHE III $_1$: Function and Regulation. *Annu. Rev. Pharmacol. Toxicol.* **2010**, *50* (1), 111–129. https://doi.org/10.1146/annurev.pharmtox.48.113006.094649.

(19) Pelengaris, S.; Rudolph, B.; Littlewood, T. Action of Myc in Vivo - Proliferation and Apoptosis. *Current Opinion in Genetics and Development*. 2000, pp 100–105. https://doi.org/10.1016/S0959-437X(99)00046-5.

(20) Sun, D.; Liu, W.-J.; Guo, K.; Rusche, J. J.; Ebbinghaus, S.; Gokhale, V.; Hurley, L. H. The Proximal Promoter Region of the Human Vascular Endothelial Growth Factor Gene Has a G-Quadruplex Structure That Can Be Targeted by G-Quadruplex-Interactive Agents. *Mol. Cancer Ther.* **2008**, *7* (4), 880–889. https://doi.org/10.1158/1535-7163.mct-07-2119.

(21) Brooks, T. A.; Kendrick, S.; Hurley, L. Making Sense of G-Quadruplex and i-Motif Functions in Oncogene Promoters. *FEBS Journal*. 2010, pp 3459–3469. https://doi.org/10.1111/j.1742-4658.2010.07759.x.

(22) Balasubramanian, S.; Neidle, S. G-Quadruplex Nucleic Acids as Therapeutic Targets. *Current Opinion in Chemical Biology*. 2009, pp 345–353. https://doi.org/10.1016/j.cbpa.2009.04.637.

(23) Brooks, T. A.; Hurley, L. H. Targeting MYC Expression through G-Quadruplexes. *Genes Cancer* **2010**, *1* (6), 641–649. https://doi.org/10.1177/1947601910377493.

(24) Neidle, S. *Therapeutic Applications of Quadruplex Nucleic Acids*; 2012. https://doi.org/10.1016/C2009-0-30510-3.

(25) Drygin, D.; Siddiqui-Jain, A.; O'Brien, S.; Schwaebe, M.; Lin, A.; Bliesath, J.; Ho, C. B.; Proffitt, C.; Trent, K.; Whitten, J. P.; et al. Anticancer Activity of CX-3543: A Direct Inhibitor of RRNA Biogenesis. *Cancer Res.* **2009**, *69* (19), 7653–7661. https://doi.org/10.1158/0008-5472.CAN-09-1304.

(26) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. Direct Evidence for a G-Quadruplex in a Promoter Region and Its Targeting with a Small Molecule to Repress c-MYC Transcription. *Proc. Natl. Acad. Sci.* **2002**, *99* (18), 11593–11598. https://doi.org/10.1073/pnas.182256799.

(27) Ou, T. M.; Lu, Y. J.; Zhang, C.; Huang, Z. S.; Wang, X. D.; Tan, J. H.; Chen, Y.; Ma, D. L.; Wong, K. Y.; Tang, J. C. O.; et al. Stabilization of G-Quadruplex DNA and down-Regulation of Oncogene c-Myc by Quindoline Derivatives. *J. Med. Chem.* **2007**, *50* (7), 1465–1474. https://doi.org/10.1021/jm0610088.

(28) Rangan, A.; Fedoroff, O. Y.; Hurley, L. H. Induction of Duplex to G-Quadruplex Transition in the c-Myc Promoter Region by a Small Molecule. *J. Biol. Chem.* **2001**, *276* (7), 4640–4646. https://doi.org/10.1074/jbc.M005962200.

(29) Trajkovski, M.; Morel, E.; Hamon, F.; Bombard, S.; Teulade-Fichou, M. P.; Plavec, J. Interactions of Pt-Ttpy with G-Quadruplexes Originating from Promoter Region of the c-Myc Gene Deciphered by NMR and Gel Electrophoresis Analysis. *Chem. - A Eur. J.* **2015**, *21* (21), 7798–7807. https://doi.org/10.1002/chem.201500347.

(30) Mathad, R. I.; Hatzakis, E.; Dai, J.; Yang, D. C-MYC Promoter G-Quadruplex Formed at the 5′-End of NHE III 1 Element: Insights into Biological Relevance and Parallel-Stranded G-Quadruplex Stability. *Nucleic Acids Res.* **2011**, *39* (20), 9023–9033. https://doi.org/10.1093/nar/gkr612.

(31) Cogoi, S.; Xodo, L. E. G-Quadruplex Formation within the Promoter of the KRAS Proto-Oncogene and Its Effect on Transcription. *Nucleic Acids Res.* **2006**, *34* (9), 2536–2549. https://doi.org/10.1093/nar/gkl286.

(32) Dai, J.; Dexheimer, T. S.; Chen, D.; Carver, M.; Ambrus, A.; Jones, R. A.; Yang, D. An Intramolecular G-Quadruplex Structure with Mixed Parallel/Antiparallel G-Strands Formed in the Human BCL-2 Promoter Region in Solution. *J. Am. Chem. Soc.* **2006**, *128* (4), 1096–1098. https://doi.org/10.1021/ja055636a.

(33) Haider, S. M.; Parkinson, G. N.; Neidle, S. Structure of a G-Quadruplex-Ligand Complex. *J.Mol.Biol.* **2003**, *326*, 117–125. https://doi.org/10.2210/PDB1L1H/PDB.

(34) Mikek, C. G.; West, S. J.; Gwin, J. C.; Dayal, N.; Sintim, H. O.; Lewis, E. A. Berenil Binds Tightly to Parallel and Mixed Parallel/Antiparallel G-Quadruplex Motifs with Varied Thermodynamic Signatures. *ACS Omega* **2018**, *3* (9), 11582–11591. https://doi.org/10.1021/acsomega.8b01621.

(35) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discov.* **2015**, *10* (5), 449–461. https://doi.org/10.1517/17460441.2015.1032936.

(36) Wang, C.; Greene, D.; Xiao, L.; Qi, R.; Luo, R. Recent Developments and Applications of the MMPBSA Method. *Front. Mol. Biosci.* **2018**, *4* (JAN). https://doi.org/10.3389/fmolb.2017.00087.

(37) Morris G.M.; Dallakyan S. AutoDock — AutoDock. *02-27* **2013**.

(38) Version, A.; Morris, G. M.; Goodsell, D. S.; Pique, M. E.; Lindstrom, W. L.; Huey, R.; Hart, W. E.; Halliday, S.; Belew, R.; Olson, A. J. AutoDock Version 4.2 User Guide. *Citeseer* **2012**, 1–66. https://doi.org/10.1101/354621.

(39) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19* (14), 1639–1662. https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B.

(40) Weiner, S. J.; Kollman, P. A.; Singh, U. C.; Case, D. A.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106* (3), 765–784. https://doi.org/10.1021/ja00315a051.

(41) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4* (2), 187–217. https://doi.org/10.1002/jcc.540040211.

(42) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *Journal of Computational Chemistry*. 2005, pp 1781–1802. https://doi.org/10.1002/jcc.20289.

(43) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *Journal of Computational Chemistry*. 2005, pp 1701–1718. https://doi.org/10.1002/jcc.20291.

(44) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31* (4), 671–690. https://doi.org/10.1002/jcc.21367.

(45) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174. https://doi.org/10.1002/jcc.20035.

(46) Gonzalez, C.; Bernhard Schlegel, H. An Improved Algorithm for Reaction Path Following. *J. Chem. Phys.* **1989**, *90* (4), 2154–2161. https://doi.org/10.1063/1.456010.

(47) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B - Condens. Matter Mater. Phys.* **1996**, *54* (16), 11169–11186. https://doi.org/10.1103/PhysRevB.54.11169.

(48) Garavelli, M.; Celani, P.; Bernardi, F.; Robb, M. A.; Olivucci, M. The C5H6NH2+ Protonated Shiff Base: An Ab Initio Minimal Model for Retinal Photoisomerization. *J. Am. Chem. Soc.* **1997**, *119* (29), 6891–6901. https://doi.org/10.1021/ja9610895.

(49) Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159* (1), 98–103. https://doi.org/10.1103/PhysRev.159.98.

(50) Hockney, R. W.; Goel, S. P.; Eastwood, J. W. Quiet High-Resolution Computer Models of a Plasma. *J. Comput. Phys.* **1974**, *14* (2), 148–158. https://doi.org/10.1016/0021-9991(74)90010-2.

(51) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. *J. Chem. Phys.* **1982**, *76* (1), 637–649. https://doi.org/10.1063/1.442716.

(52) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327–341. https://doi.org/10.1016/0021-9991(77)90098-5.

(53) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690. https://doi.org/10.1063/1.448118.

(54) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092. https://doi.org/10.1063/1.464397.

(55) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593. https://doi.org/10.1063/1.470117.

(56) pripciples and aplications. Molecular Modelling: Principles and Applications, Second Edition. *Second Ed.* **2001**, 784.

(57) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935. https://doi.org/10.1063/1.445869.

(58) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105* (43), 9954–9960. https://doi.org/10.1021/jp003020w.

(59) Homeyer, N.; Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method. *Mol. Inform.* **2012**, *31* (2), 114–122. https://doi.org/10.1002/minf.201100135.

(60) Shaikh, S. A.; Ahmed, S. R.; Jayaram, B. A Molecular Thermodynamic View of DNA-Drug Interactions: A Case Study of 25 Minor-Groove Binders. *Arch. Biochem. Biophys.* **2004**, *429* (1), 81–99. https://doi.org/10.1016/j.abb.2004.05.019.

(61) Clark Still, W.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112* (16), 6127–6129. https://doi.org/10.1021/ja00172a038.

(62) Fogolari, F.; Brigo, A.; Molinari, H. The Poisson-Boltzmann Equation for Biomolecular Electrostatics: A Tool for Structural Biology. *Journal of Molecular Recognition*. 2002, pp 377–392. https://doi.org/10.1002/jmr.577.

(63) PRODUCTS | Wavefunction, Inc. https://www.wavefun.com/products (accessed Jul 11, 2019).

(64) Case, D. A.; Babin, V.; Berryman, J.; Betz, R.; Cai, Q.; Cerutti, D.; Cheatham, T. E.; Darden, T. A.; Duke, R.; Gohlke, H.; et al. Amber 14. *Univ. California, San Fr.* **2014**. https://doi.org/citeulike-article-id:2734527.

(65) Dupradeau, F. Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. Tools: Advances in RESP and ESP Charge Derivation and Force Field Library Building. *Phys. Chem. Chem. Phys.* **2010**, *12* (28), 7821–7839. https://doi.org/10.1039/c0cp00111b.

(66) Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F. Y. R.E.D. Server: A Web Service for Deriving RESP and ESP Charges and Building Force Field Libraries for New Molecules and Molecular Fragments. *Nucleic Acids Res.* **2011**, *39* (SUPPL. 2). https://doi.org/10.1093/nar/gkr288.

(67) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **2007**, *92* (11), 3817–3829. https://doi.org/10.1529/biophysj.106.097782.

(68) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327–341. https://doi.org/10.1016/0021-9991(77)90098-5.

(69) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9* (9), 3878–3888. https://doi.org/10.1021/ct400314y.

(70) Massova, I.; Kollman, P. A. Combined Molecular Mechanical and Continuum Solvent Approach (MM- PBSA/GBSA) to Predict Ligand Binding. *Perspectives in Drug Discovery and Design*. 2000, pp 113–135. https://doi.org/10.1023/A:1008763014207.

(71) Thompson, D. C.; Humblet, C.; Joseph-McCarthy, D. Investigation of MM-PBSA Rescoring of Docking Poses. *J. Chem. Inf. Model.* **2008**, *48* (5), 1081–1091. https://doi.org/10.1021/ci700470c.

(72) Pauling, L. The Sizes of Ions and the Structure of Ionic Crystals. *J. Am. Chem. Soc.* **1927**, *49* (3), 765–790. https://doi.org/10.1021/ja01402a019.

(73) Kongsted, J.; Söderhjelm, P.; Ryde, U. How Accurate Are Continuum Solvation Models for Drug-like Molecules? *J. Comput. Aided. Mol. Des.* **2009**, *23* (7), 395–409. https://doi.org/10.1007/s10822-009-9271-6.

(74) Wang, C.; Carter-Cooper, B.; Du, Y.; Zhou, J.; Saeed, M. A.; Liu, J.; Guo, M.; Roembke, B.; Mikek, C.; Lewis, E. A.; et al. Alkyne-Substituted Diminazene as G-Quadruplex Binders with Anticancer Activities. *Eur. J. Med. Chem.* **2016**, *118*, 266–275. https://doi.org/10.1016/j.ejmech.2016.04.030.

(75) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612. https://doi.org/10.1002/jcc.20084.

(76) Luo, D.; Mu, Y. All-Atomic Simulations on Human Telomeric G-Quadruplex DNA Binding with Thioflavin T. *J. Phys. Chem. B* **2015**, *119* (15), 4955–4967. https://doi.org/10.1021/acs.jpcb.5b01107.

(77) Marchand, A.; Granzhan, A.; Iida, K.; Tsushima, Y.; Ma, Y.; Nagasawa, K.; Teulade-Fichou, M. P.; Gabelica, V. Ligand-Induced Conformational Changes with Cation Ejection upon Binding to Human Telomeric DNA G-Quadruplexes. *J. Am. Chem. Soc.* **2015**, *137* (2), 750–756. https://doi.org/10.1021/ja5099403.

(78) Hou, J. Q.; Chen, S. Bin; Tan, J. H.; Ou, T. M.; Luo, H. Bin; Li, D.; Xu, J.; Gu, L. Q.; Huang, Z. S. New Insights into the Structures of Ligand-Quadruplex Complexes from Molecular Dynamics Simulations. *J. Phys. Chem. B* **2010**, *114* (46), 15301–15310. https://doi.org/10.1021/jp106683n.

(79) Harikrishna, S.; Kotaru, S.; Pradeepkumar, P. I. Ligand-Induced Conformational Preorganization of Loops of c-MYC G-Quadruplex DNA and Its Implications in Structure-Specific Drug Design. *Mol. Biosyst.* **2017**, *13* (8), 1458–1468. https://doi.org/10.1039/c7mb00175d.

(80) Collie, G. W.; Campbell, N. H.; Neidle, S. Loop Flexibility in Human Telomeric Quadruplex Small-Molecule Complexes. *Nucleic Acids Research*. 2015, pp 4785–4799. https://doi.org/10.1093/nar/gkv427.