MOLECULAR MECHANISM, BINDING FREE ENERGY, AND SEQUENCE SELECTIVITY OF INTERCALATION OF DOXORUBICIN AND DNA

A DISSERTATION IN
Physics
and
Cell Biology and Biophysics

Presented to the Faculty of the University of
Missouri-Kansas City in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

by
BAHAA JAWAD

M.Sc., Al-Nahrain University, Baghdad, Iraq, 2012

Kansas City, Missouri
2021

MOLECULAR MECHANISM, BINDING FREE ENERGY, AND SEQUENCE SELECTIVITY OF INTERCALATION OF DOXORUBICIN AND DNA

Bahaa Jawad, Candidate for the Doctor of Philosophy Degree

University of Missouri-Kansas City, 2021

ABSTRACT

Many drugs interact noncovalently with DNA either by groove binding or intercalation. Intercalation is a key process in drug discovery and biosensor development. Doxorubicin (DOX) is an intercalator drug that treats a wide range of cancers. However, its binding process with DNA is still a highly debatable topic on both the experimental and theoretical sides with many unanswered questions. Particularly, what is the key physical factor(s) that drives the complex formation at both conformational change and insertion binding stages? What are the DOX sequence-selectivity and the role of physical factors in determining this selectivity? What is the best model to describe the relationship between binding affinity and selectivity of an intercalator drug? How do the aqueous environment and ionic concentration impact the intercalation process?

A comprehensive microsecond time-scale molecular dynamics study in an explicit aqueous solvent has been performed to address the above-raising questions. In this study, DOX interacts with different dsDNA sequences of various lengths (hexamer or tetradecamer). The molecular mechanics Poisson-Boltzmann or generalized-Born surface area (MM-PB(GB)SA) method is adapted to quantify and partition the binding free energy (BFE) into its

thermodynamic components, for a variety of different solution conditions and different DNA sequences. Our results show that the compulsory DNA conformational changes to form the intercalation cavity, the loss of translational and rotational mobility upon complex formation, and the overall electrostatic interactions are all unfavorable for the DOX-DNA complexation process. However, they are counteracted by the favorable contributions from the attractive van der Waals interaction, the non-polar solvation interaction, the vibrational entropic contribution, and the polyelectrolyte free energy at lower ionic strength. The van der Waals interaction provides the largest contribution to the BFE at each stage of binding. The sequence selectivity depends mainly on the base pairs located downstream from the DOX intercalation site, with a preference for $(AT)_2$ or $(TA)_2$ driven by the favorable electrostatic and/or van der Waals interactions. Invoking the quartet sequence model proved to be most successful to predict the sequence selectivity. Our findings indicate that the aqueous bathing solution (i.e. water and ions) opposes the formation of the DOX-DNA complex at every binding stage, thus implying that this process preferably occurs at low ionic strength and is crucially dependent on solvent effects.

APPROVAL PAGE

The faculty listed below, appointed by the Dean of the School of Graduate Studies have examined a dissertation titled "Molecular Mechanism, binding free energy, and sequence selectivity of intercalation of Doxorubicin and DNA," presented by Bahaa Jawad, candidate for Doctor of Philosophy Degree, and certify that in their opinion it is worthy of acceptance.

Supervisory Committee

Wai-Yim Ching, Ph.D., Committee Chair Department of

Physics and Astronomy

Paul M. Rulis, Ph.D.
Department of Physics and Astronomy

Michelle M. Paquette, Ph.D.
Department of Physics and Astronomy

Ryan D. Mohan, Ph.D.
Department of Cell Biology and Biophysics

Nathan A. Oyler, Ph. D.
Department of Chemistry

CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and deep appreciation to my supervisor, Professor Wai-Yim Ching, for all his patience, time, advice, motivation, wisdom, and support throughout my research. Without his guidance, this dissertation would not have come to fruition. I will forever be grateful to him.

I am also extremely grateful to my committee members, Professor Ryan D. Mohan, Professor Paul M. Rulis, Professor Michelle M. Paquette, and Professor Nathan A. Oyler, for being so flexible during this whole process, for their valuable suggestions and feedback, and for devoting their time and expertise to evaluating my research.

Special thanks to the Higher Committee for Education Development (HCED) in Iraq for sponsoring my PhD program for more than five years. I also would like to thank the University of Technology in Iraq for granting me study leave.

I would also like to thank all members of the Electronic Structure Group (ESG) and Computational Physics Group (CPG) for their kind assistance and friendship.

I sincerely appreciate the resources of the University of Missouri-Kansas City, Department of Physics and Astronomy, and Research Computing Support Services (RCSS) of the University of Missouri System.

Finally, my greatest thank goes to my family, especially my parents, for their continuous love, prayer, and sacrifice. My wife, Aya Alazzawi, deserves special thanks for her enduring love, invaluable help, patience, and understanding, and, most significantly, for taking care of my two daughters (Misk and Shems).

DEDICATION

To my lovely family: my father, mother, wife, daughters (Misk and Shems).

CHAPTER 1. INTRODUCTION

Despite indisputable progress in cancer therapy, cancer is still one of the deadliest diseases in the United States and around the world [1], [2]. According to the National Cancer Institute of the United States in 2020, over 1.8 million new cases of cancer were diagnosed in the US alone and 606,520 deaths were attributed to cancer and related complications [3]. Globally, new cancer cases per year are predicted to reach 29.5 million by 2040, with 16.4 million cancer-related deaths [3]. Consequently, the design of new cancer drugs and/or improvement of the existing ones with desired features - including lower toxic side effects, and higher selectivity and efficacy - are standardly listed among the major clinical demands. These desired therapeutic goals are, however, hampered by many challenges in cancer treatment research, among which the fundamental understanding of how an anticancer drug can recognize and bind to its correct molecular target is one of the prime topics [4]. DNA is the main molecular target for many clinical cancer drugs, since targeting DNA can disrupt the specific gene expression for the treatment of pathogenic diseases at the genetic level, especially in the case of cancers and viral diseases [5], [6]. Most cancer drugs bind to DNA by intercalation or by the groove binding process [7] (**Figure 1.1**). Unlike the groove binding drug, an intercalator drug can alter the local double-stranded DNA (dsDNA) conformation by inserting its planar aromatic chromophore portion between sequential base pairs (BPs), thus modifying the base-base stacking separation of the dsDNA helix, followed first by a local unwinding, and then by a conformational alteration of the sugar-phosphate backbone [8]. This physical distortion of DNA has become the hallmark of the intercalation process. The intercalation process occurs at least in two steps: the formation of the intercalation BP site

within the dsDNA, corresponding to the 'conformational change step', followed by the insertion of the drug into the locally deformed dsDNA, corresponding to the 'insertion step' [9]. In the first step, the dsDNA is undergoing a conformational transition from the B-DNA conformation to the intercalated dsDNA conformation. Apart from the rational drug design, the great scientific interest in the intercalation process is connected also with the essential role it plays in many other biomedical applications, such as in developing DNA biosensors [8], [10]. A fundamental understanding of the intercalation mechanisms and the various energetic components that regulate each stage of the intercalation binding process is therefore important to gain insight into its fundamental properties as well as control its practical consequences.

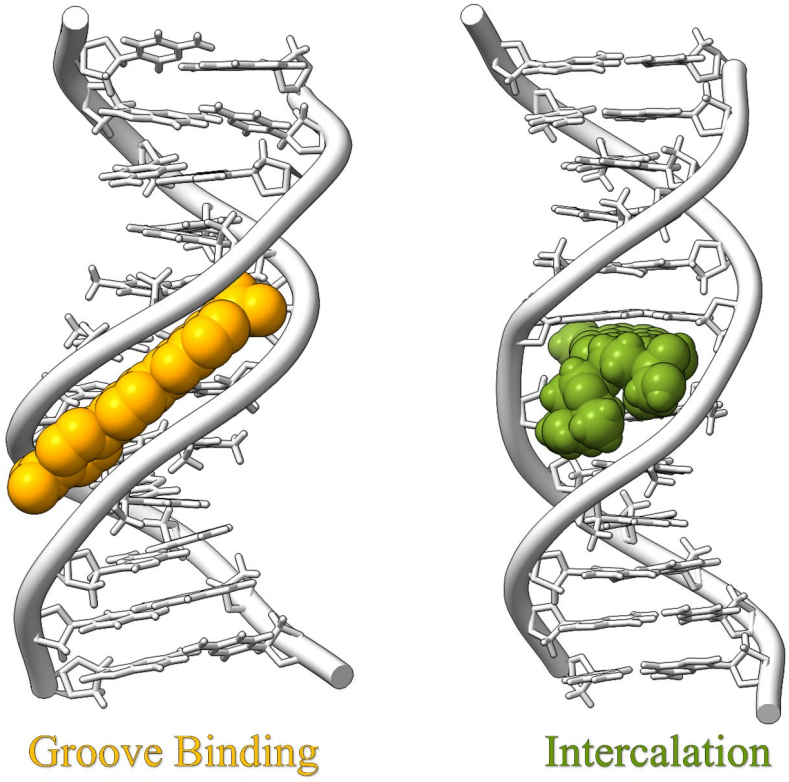


Figure 1.1. Non-covalent binding modes of small drug and DNA. In the left hand is the groove binding of DB921 and d(CGCGAATTCGCG)$_2$ complex (PDB 2B0K [11]) and the right hand is the intercalation of DOX and d(CGCGCGCGCGCG)$_2$ complex (design in this study).

Doxorubicin (trade name Adriamycin, abbreviated DOX) is an intercalator cancer drug that has been widely applied in clinical settings of chemotherapy [12]–[14]. DOX belongs to the anthracycline anticancer group, effective in killing cancer cells in both solid and liquid tumors [15]–[17]. It has three functional domains, as shown in **Figure 1.2**.

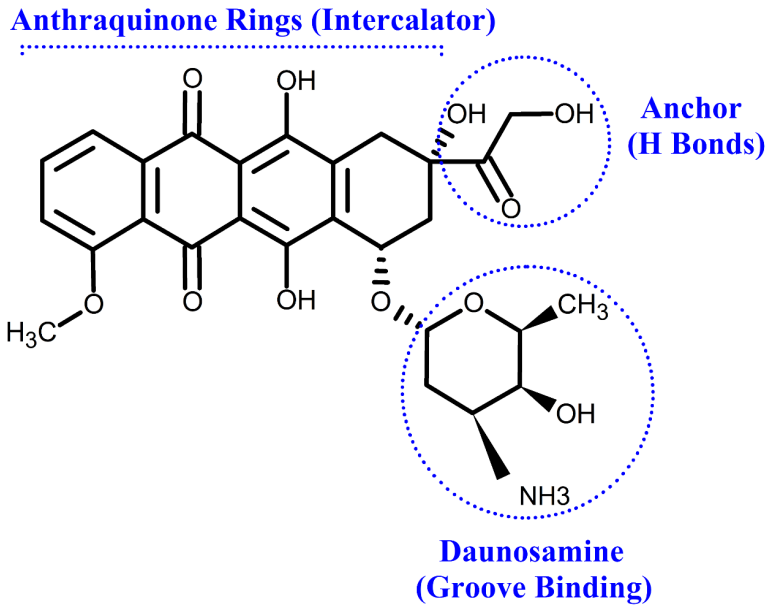


Figure 1.2. Schematic chemical structure and three functional domains of Doxorubicin.

The anthraquinone ring that intercalates between two BPs of dsDNA, a subsection in the anthraquinone ring that stabilizes the DOX-DNA complex by forming hydrogen bonds (HBs) with DNA bases, and the daunosamine domain with an amino sugar group, that acts as a minor groove binding agent [18]. The intercalation of DOX into DNA deforms the polynucleotide structure, resulting in an inhibition of the macromolecular biosynthesis, due to its interference with the enzyme topoisomerase II, and generation of enzyme-mediated DNA [19]–[23]. Despite the considerable success of DOX in clinical applications as a

chemotherapeutic anticancer agent, the underlying modus operandi is still not entirely clear and remains a subject of intense debate. There are many proposed models including topoisomerase II poisoning, DNA adduct formation, oxidative stress, as well as ceramide overproduction to explain the DOX-mediated cell death [24]–[26]. Hence, a detailed analysis of the DOX-DNA interactions can provide in-depth knowledge of the molecular mechanism of intercalator drug action, as well as a better understanding of the general principles for guiding the new DNA-intercalator design.

Apart from DOX's general mechanism, its action is in addition sequence-dependent, with details that are also not entirely clear and remain doubtful, resulting in further unresolved questions in basic and applied research. First, there is still debate about the nature of the preferential binding of DOX to dsDNA [27], [28]. Some studies have suggested that the preferred intercalation site for DOX is $(CG)_2$ and the most energetically favored sequences are 5'TCG and 5'ACG [29]–[35], while other studies have indicated that the 5'TCA sequence is the one connected with the preferred binding [36]–[38]. Second, the model describing the relationship between the DNA sequence and the DOX binding affinity is also not entirely resolved, being modeled by either a triplet sequence model [31]–[33], or equivalently by a quartet sequence model [28], [39], [40], with no emerging comprehensive consensus. Third, the picture of how the nearest neighbor BPs affect the DOX intercalation site is also not clear. The physical structure of DOX's daunosamine sugar domain allows DOX to extend to the nearest BP, vicinal to the intercalation site, thus extending the role of this amino sugar group in the DOX binding process. Fourth, the physical factors that drive the DOX-DNA complex formation and their role in its selectivity have also not been adequately studied. Fifth, the

calculated values of DNA deformation penalty and entropy cost of intercalated DOX are experimentally difficult to obtain and subject to some debate [41]. Finally, there is limited information also on the non-specific polyelectrolyte effects of the DOX-DNA binding mechanism, which are expected to be non-negligible.

In this dissertation, the main aims are to address the questions raised above. We undertook a comprehensive and systematic study focusing on the understanding of the molecular mechanisms and sequence-specificity implied in the DOX intercalation process. The current study is based on molecular dynamics (MD) simulations that have been performed at different solution conditions for many DOX-DNA complexes having various DNA sequences with two distinct lengths (hexamer or 6 BPs and tetradecamer or 14 BPs), two DOX configurations (1 DOX or 2 DOXs), and different DOX sites (in the middle of DNA or at its end terminal). We first verified the intercalation binding process between DOX and DNA. Simultaneously, many molecular factors involved in the formation of a stable DOX-DNA complex have been inspected and quantified [1]. Then, we extended our research to investigate the DOX sequence-specificity at a deeper level and for a larger number of complex models with hopes to put an end to many controversies and unanswered questions about this selectivity [2]. Our research provides useful information that could be used in drug design and intercalator biosensors.

1.1 Outline of Dissertation

The outline of this dissertation is as follows. In chapter 2, we describe the computational theory of molecular dynamics (MD) simulations, solvation models, the end-point binding free energy method (BFE), and the analysis tools used in this dissertation.

Chapter 3 consists of the study of the intercalation binding process between DOX and different hexamer DNA sequences with two complex configurations (1:1 and 2:1 complexes) [1] and chapter 4 consists of the study of DOX sequence-specificity [2]. In chapter 3, we present the computational modeling and BFE calculations of the DOX-DNA complex with two sequences: d(CG**AT**CG)$_2$ or DNA1 and d(CG**TA**CG)$_2$ or DNA2. Several molecular interactions are probed and quantified to identify the key factor responsible for driving the stable intercalated DOX-DNA complex. Chapter 3 [1] (Bahaa Jawad, Lokendra Poudel, Rudolf Podgornik, Nicole F. Steinmetz, and Wai-Yim Ching, Molecular mechanism and binding free energy of doxorubicin intercalation in DNA. Physical Chemistry Chemical Physics, 2019, 21(7).) is reproduced by permission of the PCCP Owner Societies. In chapter 4, we focus primarily on the DOX sequence-dependent and attempt to describe a comprehensive model to explain the relationship between DOX sequence and binding affinity. In this chapter, we describe in more detail our procedures to design sixteen DOX-DNA complexes with tetradecamer sequences of form d(CGCG**XYISXY**CGCG)$_2$. Chapter 4 [2] is reprinted with permission from (Bahaa Jawad, Lokendra Poudel, Rudolf Podgornik, and Wai-Yim Ching, Thermodynamic Dissection of the Intercalation Binding Process of Doxorubicin to dsDNA with Implications of Ionic and Solvent Effects. The Journal of Physical Chemistry B, 2020, 124(36).). Copyright (2020) American Chemical Society. Each chapter has its result and conclusion sections. We conclude our work with some final remarks and future work in chapter 5.

CHAPTER 2. THEORY AND METHODOLOGY

This chapter describes the general concepts of molecular dynamics (MD) simulation, solvation models, the common methods for calculating the binding free energy (BFE), and the analysis tools that are adopted in this dissertation. Details of the specific simulation setups and the BFE protocols are given in chapters 3 and 4.

2.1 Molecular Dynamics Simulation

Molecular dynamics (MD) simulation is a computational technique that uses classical Newtonian dynamics to simulate the time evolution of all interacting atoms in biological macromolecules like protein and nucleic acid [42]–[44]. It can capture a wide variety of key biomolecular dynamics and processes, including conformational changes in protein or nucleic acids [45]–[47], thermodynamics of ligand-receptor [48]–[52], protein-protein [53]–[55], DNA-protein [56], [57] interactions, protein folding [58], [59], etc. Since protein or nucleic acids are large systems and their processes occur at longer timescales (nanoseconds or longer), ideally describing their interactions based on *Ab Initio* quantum mechanic methodologies is still a challenging task [60]. Therefore, it is necessary to resort to a classical MD, which is computationally more efficient for simulating all-atoms of the system on a long timescale. However, it has many simplifications and approximations as follows. First, it only calculates the energy of a system as a function of the nuclear positions which is justified based on the Born-Oppenheimer approximation [42]. Second, it uses an empirical force field to describe the interaction between atoms.

Many popular MD simulation packages exist, including AMBER [60], [61], CHARMM [63], NAMD [64], Gromacs [65], etc. In this dissertation, AMBER (Assisted Model Building with Energy Refinement) has been used. The basic concept behind an MD simulation is to describe the time-dependent behavior of systems by numerically integrating Newton's second law of motion [44]:

$$\boldsymbol{F}_i(t) = m_i \boldsymbol{a}_i(t) = -\frac{\partial}{\partial \boldsymbol{r}_i} V(\boldsymbol{r}_i) \quad (1)$$

where m_i is a mass of the *i*th atom and $\boldsymbol{F}_i(t)$ is the net force acting on it at a given time t, deriving usually from a potential energy $V(\boldsymbol{r}_i)$ as a function of all atomic coordinates $\boldsymbol{r}_i(t)$ of the instantaneous configuration of the system. The corresponding acceleration $\boldsymbol{a}_i(t)$ can be determined from this force, leading to a change in the velocity and position of the *i*th atom within a discrete time-step δt, usually in the range of 1-2 femtoseconds (fs) to capture the fastest motions in the system. Many algorithms, such as SHAKE [66], [67] and LINCS [68], have been developed to constrain the motion of hydrogen-containing bonds, allowing the MD simulation to be performed with timestep lengths greater than 1 fs. Other numerical algorithms, such as Verlet [69] and leap-frog [61] algorithms, use these time-steps to simplify the integration of Newton's equation of motion, especially for large biomolecule systems. In AMBER MD simulation, the leap-frog algorithm is used [61], which is a modified version of the Verlet scheme. In this algorithm, the positions are defined at times t, $(t + \delta t)$, $(t + 2\delta t)$, … according to the velocity at times $(t - \delta t/2)$, $(t + \delta t/2)$, $(t + 3\delta t/2)$, ... respectively. For example, to calculate the position at the time $(t + \delta t)$, the velocity at the time $(t + \delta t/2)$ is calculated first [70]:

$$\boldsymbol{v}\left(t+\frac{1}{2}\delta t\right)=\boldsymbol{v}\left(t-\frac{1}{2}\delta t\right)+\boldsymbol{a}(t)\delta t \tag{2}$$

$$\boldsymbol{r}(t+\delta t)=\boldsymbol{r}(t)+\boldsymbol{v}\left(t+\frac{1}{2}\delta t\right)\delta t \tag{3}$$

In this way, the velocities leap over the positions, then the positions leap over the velocities.

Simply, MD simulation is an iterative process consisting of two steps: calculating the instantaneous forces on each atom and then using those forces to update the motion of the system [42], [43]. This process is typically repeated millions of times to provide insight details about the dynamic and structure aspects of a certain biomolecular process under investigation. As a result, the trajectory of atomic-level motion is obtained, which is essentially a three-dimensional movie describing the system configuration at each point during simulation [43].

2.2 Empirical Force Fields

The force in Eq.(1) is calculated as a derivative of potential energy $V(\boldsymbol{r}_i)$ using a model known as a molecular mechanics (MM) force field for describing the intermolecular and intramolecular interactions of the system. Because the force field (FF) governs all interactions and dynamic behaviors of the system, selecting an accurate FF is crucial in an MD simulation. Various FFs have been developed for biomolecule systems, with the most commonly used being various versions of AMBER [71]–[74], CHARMM [75]–[78], and OPLS [79]–[81]. Notably, the AMBER and CHARMM MD packages should not be confused with the AMBER and CHARMM force fields. Again, the AMBER MD package with appropriate AMBER FFs for DNA and DOX has been used in this dissertation (chapters 3 and 4). Generally, FF contains the functional forms of potential energy, the definition of atom types that differ by their atomic

number and chemical environment, and a set of parameters. The functional form of the AMBER FF is [82]:

$$V(\boldsymbol{r}) = \sum_{bonds} K_b\left(l - l_{eq}\right)^2 + \sum_{angles} K_\theta\left(\theta - \theta_{eq}\right)^2 + \sum_{dihedrals} \sum_{n} \frac{V_n}{2}\left[1 + \cos(n\phi - \gamma)\right] + \sum_{i}^{N} \sum_{i>j}^{N-1} \left\{ \epsilon_{ij} \left[\left(\frac{r_{ij}^0}{r_{ij}}\right)^{12} - 2\left(\frac{r_{ij}^0}{r_{ij}}\right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right\} \quad (4)$$

In Eq.(4), the first three terms represent bonded or intramolecular interactions arising from changes in bond stretching (1-2 interactions), bending (1-3 interactions), and torsions (1-4 interactions) as shown in **Figure 2.1**. Bond stretching and bending interactions are modeled by simple harmonic potentials, which require only two parameters for each potential: the equilibrium values l_{eq} and θ_{eq} and force constants K_b and K_θ, respectively. Torsion interactions are typically represented by a cosine series such as the one used in Eq.(4), where V_n is the torsional barrier corresponding to the nth barrier of a given torsional angle ϕ with phase γ, and n is the multiplicity defining the number of minima or maxima between 0 and 2π. The last two terms of Eq.(4) are the non-bonded or intermolecular interactions associated with van der Waals (E_{vdW}) and electrostatic (E_{ele}) energies. E_{vdW} is modeled using the 6-12 Lennard-Jones (LJ) potential [83], where ϵ_{ij} and r_{ij}^0 define the depth and position of the potential minimum, respectively. E_{ele} is represented by the Coulomb potential, where q_i and q_j are the partial charges on the respective atoms and ε_0 is the dielectric constant.

To reproduce the actual behavior of the biomolecule system, the fixed parametrizations of the FF are determined by fitting to experimental data and quantum mechanical calculations [49]. For an example of the nucleic acid bases, the l_{eq} and θ_{eq} values are taken from X-ray

structural data, the K_b and K_θ values are obtained from vibration spectroscopies, and the fixed partial charges are assigned relied on the HF/6-31G* QM calculations with the RESP (Restrained Electrostatic Potential) fitting approach [82]. For more information on how this FF describes interactions and sets parameterizations, see Ref. [82].

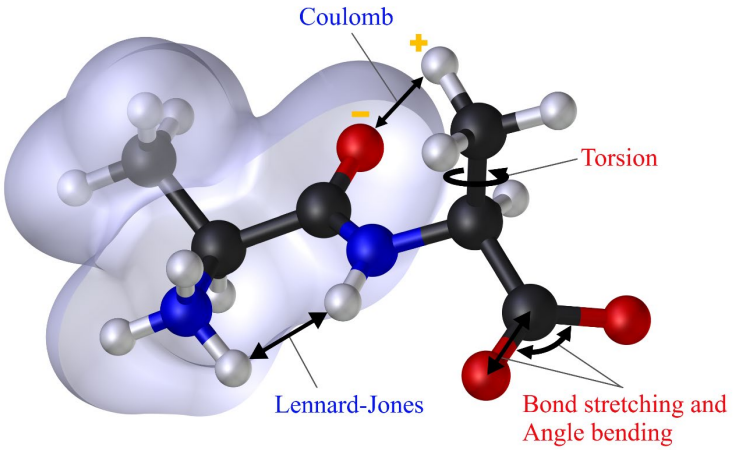


Figure 2.1. Bonded and non-bonded interactions between atoms in a force field [84].

Although FF has many advantages such as simplicity, transferability, additivity of energy components, allowing extended MD simulations of large biomolecular systems, and speeding up calculations, it has many limitations. The major limitations are the fixed partial charges and the inability to describe forming or breaking of the covalent bonding between atoms during the chemical reaction [42]–[44]. To overcome these limitations, polarizable or reactive FFs have been developed [85]–[88], but they are not widely used yet.

2.3 Treating Solvent Effects

Aqueous solvents (water and ions) play a crucial role in governing the structure, stability, dynamics, thermodynamics, and functions of biological molecules [89], [90]. Water

is a vital molecule for all living organisms, yet it has only recently been quantitatively considered as an integral part of biomolecular systems [89]. It influences the biomolecules through participating in hydrogen networks, screening effect of charge-charge interactions, hydrophobic attractions, and solvation charged molecules (e.g. ligand, charged amino acids, base pair of nucleic acids, etc.) [89]–[91]. For these reasons, the inclusion of solvents is a necessary step to obtain a realistic and accurate simulation. Computationally, solvents can be represented using either explicit or implicit models as described below.

2.3.1 Explicit Solvent Models (ESM)

Individual solvent molecules are treated explicitly in explicit solvent models (ESM). ESM is the most comprehensive, accurate, and realistic model for describing solvent effects. However, it is computationally demanding. Many MD programs include ESM such as transferable intermolecular potential with 3 points (TIP3P) [92], 4 points (TIP4P) [92], 5 points (TIP5P) [93], optimal-point charged (OPC) [94], single-point charged (SPC) [95], or modified versions of them. The electrostatic representation (number of charge sites and polarizability), and the flexibility or rigidity of the water structure (internal geometry) are the main distinctions amongst these ESMs [91]. AMBER package supports all these ESMs. TIP3P has been used in this work, which is a rigid and 3-sites of fixed charge model. The dimerization energy for two water molecules, m and n, is calculated using Coulombic electrostatic interactions of all intermolecular pairs and a single Lennard-Jones (LJ) term between oxygens (Eq.5) [92].

$$E_{mn} = \sum_{i}^{on\ m} \sum_{j}^{on\ n} K \frac{q_i q_j}{r_{ij}} + \frac{A}{r_{OO}^{12}} - \frac{C}{r_{OO}^{6}} \tag{5}$$

where K is Coulomb's constant and equal to 332.1 (kcal. Å/mol·e²), r_{ij} are distances between the charged sites with charges q_i, q_j, A and C are the LJ parameters, and r_{OO} is oxygen–oxygen distance. In TIP3P, the equilibrium OH bond length is 0.9572 Å, the equilibrium HOH angle is 104.52°, the partial charges of oxygen and hydrogen are -0.834e and 0.417e, the LJ parameters of A and C are 0.582 (kcal. $Å^{12}$/mol) and 0.595 (kcal. $Å^6$/mol). All these parameters are typically adjusted for best fit against a chosen set of experimental water properties, such as density, radial distribution functions, enthalpy of vaporization, heat capacity, diffusion coefficient, and dielectric constant. TIP3P reproduces well the physico-chemical properties of bulk water [96]. However, many-body effects like electronic polarizability and quantum effects like charge transfer are not explicitly taken into account in Eq.(5) [97]. TIP3P also assumes that the geometry of water molecules remains fixed, ignoring the explicit consideration of internal interactions. It also underestimates the height of the second (tetrahedral) peak in the O-O radial distribution function and overestimates the diffusion constant [96]. Many ESMs have been developed to address these critical limitations [92], [93], [95]. Likewise, ions are explicitly treated using the Joung-Cheatham ion parameters for TIP3P water [98].

The common protocol for performing an MD simulation begins with solvating the solute (protein, DNA, or ligand-DNA complex) in a periodic box using a large number of explicit solvents molecules. Two main points should be mentioned here. First, because simply placing the solute in a solvated box is insufficient to reproduce actual bulk properties, the periodic boundary condition (PBC) is employed to mimic an infinite system with finite repeating samples of the basic simulation box [44]. PBC is also used to overcome unphysical edge effects caused by the finite size. Briefly, by using PBC, the model system is placed in a

unit cell that is replicated in all three dimensions to form an infinite lattice of image atoms (**Figure 2.2**) [99].

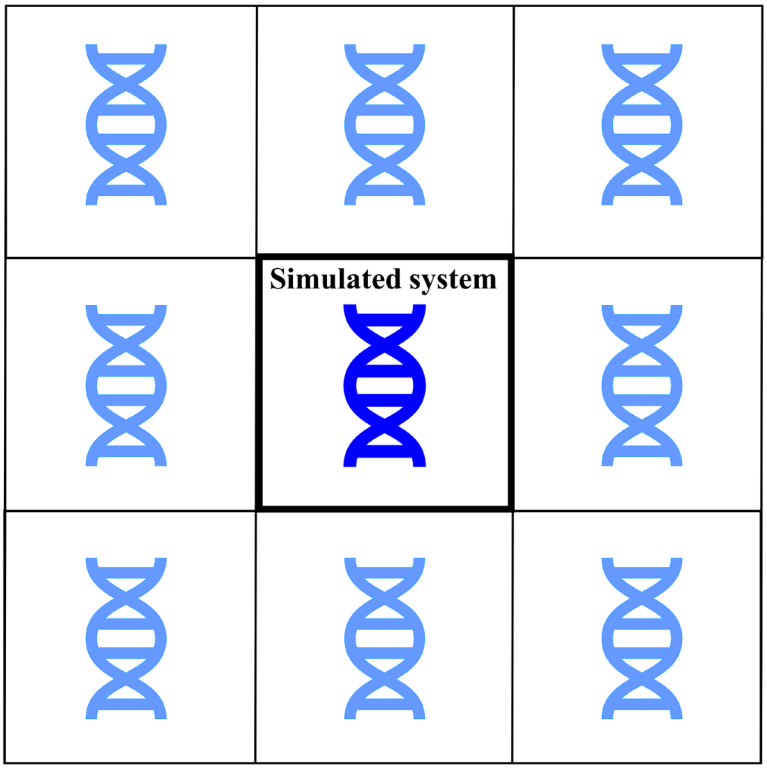


Figure 2.2. 2D schematic of periodic boundary conditions in molecular dynamics for a simple DNA system. Note that the DNA simulated system is a sub-ensemble within an infinite system of identical, small ensembles.

During the simulation, every periodic image in the neighboring boxes moves in the same way as a particle moves in the primary simulation box [100]. Coordinates and velocities are only stored and propagated for the simulation box, but the interactions are reproduced across all the periodic images. Each particle interacts with the other particles and their images in neighboring boxes. As a particle leaves from one side of the box, one of its images enters from the opposite side. As a result, the total number of particles in the central simulation box

is conserved [70]. AMBER includes two PBC geometries: rectangular parallelepiped and truncated octahedron (i.e. box with corners chopped off) [61]. In this study, we used a rectangular PBC.

Second, special attention must be paid when calculating long-range interactions. PBC is commonly combined with Ewald summation methods like the particle mesh Ewald method (PME) [101] to split the pairwise interactions into short-range and long-range contributions. The short-range contributions are sum rapidly in real space using a cutoff scheme in the range of 8-12 Å for both van der Waals and electrostatic interactions. The cutoff distance should be equal to or larger than half the diameter of the primary simulation box. While the slowly decaying long-range electrostatic interactions are treated in reciprocal space using three-dimensional (3D) fast Fourier transforms (3DFFT) with the charge density discretized on a grid [44], [62], whereas the long-range van der Waals interactions are estimated by a continuum model [62].

2.3.2 Implicit Solvent Models (ISM)

Since an MD simulation in ESM involves many discrete explicit solvent molecules, the computational cost is high and slow convergence. These limitations arise from massive calculations of solvent-solvent and solvent-solute interactions during each MD integration step, as well as the use of a large number of samples to estimate thermodynamic properties across many degrees of solvent freedom [102], [103]. To handle these problems, the alternative and simpler strategy is the implicit (or continuum) solvent model (ISM), which treats the solvent effects implicitly as a continuous medium, commonly assumed homogeneous and isotropic, with dielectric and nonpolar properties of water [97], [104]–[106]. In ISM, the

solvent is represented as a high-dielectric continuum that interacts with the solute charges embedded in a lower-dielectric medium like a vacuum [107]. The dielectric boundary between solute and solvent is taken as the molecular surface defined by a 1.4 Å probe sphere of solvent and by spheres centered on each atom with van der Waals radii [108].

Besides the less computational cost and faster convergence of ISM over ESM, it provides a quick and easy way to compute solvation-free energy. Despite these advantages, ISM contains some approximations whose effects are frequently difficult to address. For instance, the solute-solvent or/and solvent–solvent hydrogen bonds, and ion distribution, are no longer explicitly described. However, these interactions are implicit and contribute to the overall solvation energy in the mean-field approximation via a linear dielectric response [62]. ISM is widely applied for elucidating the electrostatic properties of many biological phenomena in solution, particularly the binding process, because it offers a trade-off between efficiency and accuracy [106], [109]–[111].

When modeling the ligand-receptor binding process using MD simulation, the binding free energy (BFE) of the molecule in the presence of solvent is the most important quantity to compute. As will see later, BFE can partition in terms of the nature of the interaction in a vacuum and/or solvent [2]. Solvation free energy (ΔG_{sol}) is the energy required to transfer a molecule from vacuum to solvent, and its thermodynamic cycle is illustrated in **Figure 2.3**. By using ISM, ΔG_{sol} is divided into an electrostatic or a polar contribution (ΔG_p) and a nonelectrostatic or nonpolar contribution (ΔG_{np})

$$\Delta G_{sol} = \Delta G_p + \Delta G_{np} \tag{6}$$

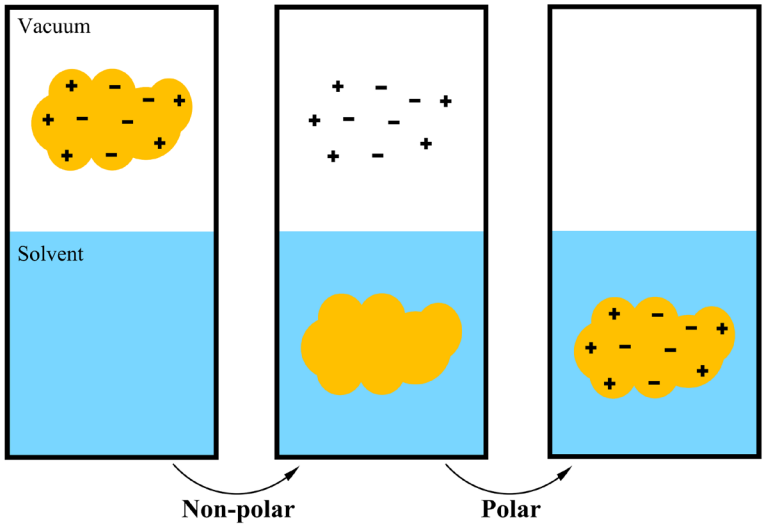


Figure 2.3. Thermodynamic cycle for computing the solvation energy using implicit solvent model (ISM). In ISM, the solvation energy is decomposed into electrostatic (polar) and nonpolar components.

According to **Figure 2.3**, ΔG_{np} is the free energy of solvating a molecule from which all charges of their atoms have been removed, and ΔG_p is the free energy of first removing all charges in the vacuum and then adding them back in the presence of a continuum solvent environment [62]. Generally, ΔG_{np} is composed of two interactions: the unfavorable cost of creating a cavity of the solute in the solvent, and the favorable van der Waals attraction between the solute and solvent. It is usually modeled as a linear function of the solvent-accessible surface area (SASA) of the solute

$$\Delta G_{np} = \gamma . SASA + b \tag{7}$$

where γ is a surface tension parameter derived from experimental solvation energies of small nonpolar molecules, and b is an offset correction term or fitting parameter [62], [109]. SASA is defined as the locus of the center of a solvent-sphere probe (1.4 Å for water) rolling over the solute van der Waals surface as shown in **Figure 2.4(a)** [112]. Based on this approach,

AMBER uses a fast LCPO (linear combination of pairwise overlaps) algorithm to compute an analytical approximation to the SASA of the solute [113].

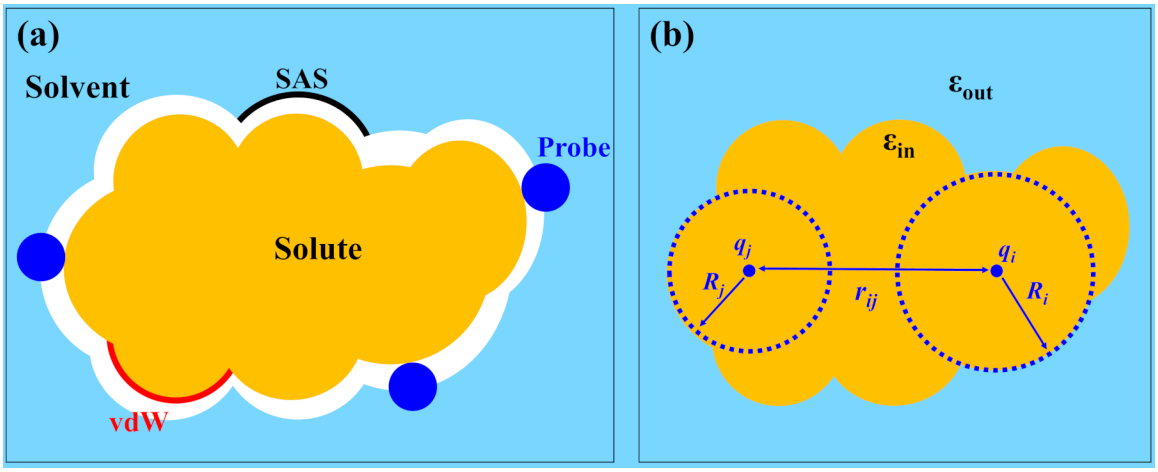


Figure 2.4. (a) Solvent-accessible surface area (SASA) model. The van der Waals (vdW) surface as given by the atomic radii is shown in gold, while the SAS is shown in white and is created by tracing the center of the probe sphere (in blue) as it rolls along the vdW surface. (b) The effective Born radius of an atom reflects the degree of its burial inside the low dielectric region defined by the solvent boundary.

Unlike ΔG_{np}, the computation of ΔG_p to total ΔG_{sol} is the most time-consuming part because the forces involved are long-ranged and the solvent screening effect is a complex phenomenon [62]. ΔG_p can be computed numerically using the finite-difference Poisson-Boltzmann (PB) model or analytically using a pairwise generalized Born (GB) model as described below.

I. Poisson–Boltzmann model

In classical electrostatic theory, within the continuum model framework, the electrostatic solvation component (ΔG_p, here denotes as ΔG_{PB}) is calculated by numerically solving Poisson−Boltzmann (PB) equation using a finite-difference (FD) method.

In the absence of mobile ions, PB equation for a biomolecular system is given by [109]

$$\nabla.\varepsilon(\boldsymbol{r})\nabla\varphi(\boldsymbol{r}) = -4\pi\rho(\boldsymbol{r}) \tag{8}$$

Here, $\varepsilon(\boldsymbol{r})$ is the position-dependent dielectric constant, which equals the bulk water dielectric constant far away from the solute and is expected to decrease fairly rapidly across the solute-solvent boundary. $\varphi(\boldsymbol{r})$ is the electrostatic potential distribution function, and $\rho(\boldsymbol{r})$ is the fixed atomic charged density of the solute. When mobile ions are present, the charge density includes the solute charge distribution inside the cavity and the charge density generated by the ion atmosphere outside the cavity (ρ_{ext}) in solutions ($\rho = \rho(\boldsymbol{r}) + \rho_{ext}(\boldsymbol{r})$. At equilibrium, $\rho_{ext}(\boldsymbol{r})$ can be approximated by a Boltzmann distribution, which is written as [114]

$$\rho_{ext}(\boldsymbol{r}) = -\varepsilon_w\, \kappa^2\, sinh\, \varphi(\boldsymbol{r}) \tag{9}$$

where ε_w is the solvent dielectric constant (e.g. water) and $\kappa^2 = \frac{8\pi e^2 I}{\varepsilon_w\, k_B T}$ represents the Debye–Huckel screening parameter that is used to model the electrostatic screening effects of (monovalent) salt, where e is the charge of the electron, I is the ionic strength of the bulk solution, k_B *is* Boltzmann's constant, and T is the temperature. The PB equation can be expressed in two forms based on the ionic strength. First, for weak ionic strength, the $sinh\, \varphi(\boldsymbol{r})$ function in Eq.(9) is approximated by including only the first term of a Fourier expansion, $sinh\, \varphi(\boldsymbol{r}) \approx \varphi(\boldsymbol{r})$, yielding a linear PB (LPB) as in Eq.(10). Second, for ionic strengths requiring a general nonlinear expression of PB (NLPB) as in Eq.(11), it uses especially for highly charged systems [109], [114]

$$\nabla.\varepsilon(\boldsymbol{r})\nabla\varphi(\boldsymbol{r}) - \varepsilon_w\, \kappa^2\varphi(\boldsymbol{r}) = -4\pi\rho(\boldsymbol{r}) \tag{10}$$

$$\nabla.\varepsilon(\boldsymbol{r})\nabla\varphi(\boldsymbol{r}) + 4\pi\lambda(\boldsymbol{r})\sum_{i}^{N} z_i c_i \, exp\left[\frac{-z_i\varphi(\boldsymbol{r})}{k_B T}\right] = -4\pi\rho(\boldsymbol{r}) \quad (11)$$

where $\lambda(\boldsymbol{r})$ describes the accessibility to ions, it is a predefined ion-exclusion function with a value of 0 within the Stern layer and the molecular interior (i.e., the areas are inaccessible to ions) and a value of 1 outside the Stern layer (accessible areas for ions), z_i is the charge and c_i is the bulk number density of ion type *i* far from the solute at a given temperature T.

Unfortunately, analytical solutions to PB equations are only possible for systems with simple and highly symmetric dielectric boundary geometries, such as spheres [115]. Therefore, they are solved numerically using various numerical methods for realistic biomolecule shapes [106]. One of the most popular and rigorous methods is the finite-difference (FD) method, which relies on the superimposition of a regular rectangular Cartesian mesh over the system [116]. An FD involves the following steps to solve the PB equations: mapping atomic charges to the FD grid points, specifying non-periodic/periodic boundary conditions, and using a molecular dielectric model to define the boundary between high-dielectric (water) and low-dielectric (molecular interior) regions and mapping them to the FD grid edges [62]. The DF scheme is described in more detail elsewhere [106], [116], [117]. Many solvers based on FD schemes are available for solving the PB, including PBSA (an AMBER module) [118], DelPhi [119], ZAP [120], etc.

II. Generalized Born model

Because the numerical solutions of the PB equation are computationally expensive, an alternative dielectric continuum model has been developed, which is a Generalized Born (GB) pairwise approximation of the PB equation. The analytic GB method has gained popularity for

MD applications due to its simplicity and computational efficiency compared to the numerical solution of the PB model [121], [122]. Within the AMBER GB model, each atom in a biomolecule is described as a charged sphere with a radius R_i and a charge q_i at its center, and a lower interior dielectric constant than the surrounding environment [123], [124]. Because the local environment, which has a low dielectric constant, determines the screening that each atom experiences, the more an atom is surrounded by other atoms, the less its electrostatics will be screened. This propriety is known as descreening of one atom by another. The Born radius of each atom is calculated using descreening, and the Born radius of an atom thus describes the degree of descreening. A large Born radius of atom in vacuum indicates small screening (strong electric field) while a small Born radius of atom in water represents large screening (weak electric field) [109].

In the absence of mobile ions, the standard form of the GB model is expressed as [109]

$$\Delta G_{GB} = -\frac{1}{2}\left(\frac{1}{\varepsilon_i} - \frac{1}{\varepsilon_w}\right)\sum_{i,j}\frac{q_i q_j}{f_{GB}} \tag{12}$$

In which

$$f_{GB} = \left[r_{ij}^2 + R_i R_j exp\left(-r_{ij}^2/4R_i R_j\right)\right]^{1/2} \tag{13}$$

where ε_i is the interior dielectric constant of the biomolecule, r_{ij} is the distance between atoms i and j, $R_i \; and \; R_j$ are the effective Born radii reflecting the degree of their burial inside the molecule, and f_{GB} is a certain smooth function of its arguments (**Figure 2.4(b)**) [62].

In the presence of ions, the GB modify model is written as [109]

$$\Delta G_{GB} = -\frac{1}{2}\left(\frac{1}{\varepsilon_i} - \frac{exp(-\kappa f_{GB})}{\varepsilon_w}\right)\sum_{i,j}\frac{q_i q_j}{f_{GB}} \tag{14}$$

Here, κ is the Debye–Huckel screening parameter. The polar solvation energy is strongly dependent on the effective Born radii, as shown by GB equations (12 and 14). These Born radii must be calculated at each time step in an MD simulation because they are dependent on protein conformation. Various approximations have been used to estimate the effective Born radii based on the molecule configuration and intrinsic atomic radii [97]. In AMBER, many GB models have been developed for this purpose such as GB^{HCT} (igb = 1) [125], [126], GB^{OBC}(igb = 2 or 5) [122], [127], and GBn (igb = 7 or 8) [128].

For an isolated ion, R_i is equal to its vdW radius a, and Eq.(12) takes the form of the well-known Born formula for the solvation energy of a single spherical ion [62]:

$$\Delta G_{Born} = -\frac{q^2}{2a}\left(\frac{1}{\varepsilon_i} - \frac{1}{\varepsilon_w}\right) \tag{15}$$

2.4 Binding Free Energy (BFE)

The most important concept in physical chemistry and biophysics is binding free energy (BFE) [129]. BFE is a thermodynamic potential of the system that can be used to do work, and it describes thermodynamic process tendencies as well as the probability of the system remaining in a given state [109]. BFE is an essential quantity for providing physical insight into details difficult to access in experiments such as the molecular driving force, guiding the design of the best drug candidate with desirable pharmacological properties, and speeding the discovery of new medications and biological probes [130]. Because the BFE governs all biomolecular processes, like protein folding, molecular recognition, chemical reaction, etc., an accurate and reliable determination of the BFE is a major goal of computational drug discovery or any biomolecular studies [109]. Many computational

approaches are available to predict BFE including free energy perturbation (FEP) [131] and thermodynamic integration (TI) [132]. These rigorous calculations are computationally expensive as the system size increases [133], mainly due to the explicit treatment of solvent using the explicit solvent model (ESM) [134], and the need for calculations to be performed for many intermediate states along the transition pathway can be obtained.

The alternative and simpler strategy is to treat the solvent effects implicitly using the continuum solvent method, also called the implicit solvent model (ISM) (section 2.3). The computational cost can be significantly reduced by considering only the free energy difference between the end-points [1]. The solvent effect is still partially accounted for in ISM, so it achieves a good balance between computational efficiency and accuracy [1].

The most well-known end-point and implicit solvent methods are the Molecular Mechanics Poisson–Boltzmann Surface Area (MM-PBSA) and the Molecular Mechanics Generalized-Born Surface Area (MM-GBSA) approaches [135]. The MM-PB(GB)SA method has been successfully used in a range of settings including protein design [136], and protein-protein [137]–[139], protein-ligand [140], [141], protein–DNA [142], and drug–DNA interactions [143]–[145]. In particular, the MM-PB(GB)SA method has been carried out to study the binding of DOX to various macromolecules [146], [147]. These two methodologies have also been used to study the interaction mechanism of DOX with carbon nanotube [148], [149]. Most recently, they have been demonstrated to be valuable in SARS-CoV-2 research [150]–[156].

2.4.1 MM-PB(GB)SA Methodology

The common experimental methods that are used to measure the drug−DNA BFE are calorimetric [157]. However, reliable decomposition of BFE into either the binding enthalpies and entropies or various energetic contributions or per-residue normalization has also not been fully implemented. Consequently, there is insufficient information to integrate these energetic components into overall process characterization.

Here, we use the most popular end-point free energy method [107], [133], [149], the MM-PBSA or MM-GBSA methods, to calculate the BFE, ΔG, and the relative ΔG, $\Delta\Delta G$ [1], [2]. It combines the molecular mechanics (MM) energies with the PBSA or GBSA continuum solvation approach to predict the BFE of small ligands to biological macromolecule receptors [159]. The main difference between the MM-PBSA and the MM-GBSA approach is the methodology of calculating the solvation free energy [1]. The MM-PB(GB)SA methods we adopted have many advantages: it is more accurate than most empirical scoring approaches of molecular docking and less computationally demanding than alchemical free energy methods [109], it can be modified to analyze the BFE concerning the binding stages in the intercalation process of drugs, and it can break down the BFE to their thermodynamic components and can be further decomposed into per-residue energetic contributions [160], which make it possible to capture the local dominant interactions between DOX and dsDNA including contributions from individual base pairs (BPs) that are located near the DOX intercalation site [2].

In the MM-PBSA or MM-GBSA method, the BFE is determined based on independent MD simulations as the difference between the free energies of the bound state of the

DOX−DNA complex ($G_{(COM,sol)}$) and the unbound state of free B-DNA ($G_{(DNA,sol)}$) and free DOX ($G_{(DOX,sol)}$) [2], [109], [160]

$$\Delta G_{bind} = G_{(COM,sol)} - G_{(DNA,sol)} - G_{(DOX,sol)} \quad (16)$$

Each free energy term of the species (DOX, DNA, and their complex) in Eq.(16) can be computed from contributions of different interactions and expressed as [2], [109]

$$G = E_{MM} + G_{sol} - TS \quad (17)$$

Thus Eq.(16) can be written as

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{sol} - T\Delta S = \Delta G_{vac} + \Delta G_{sol} \quad (18)$$

where ΔE_{MM} is the changes in the gas phase MM energy, which is calculated by averaging the configurational ensemble along the MD simulation trajectory based on the force field interactions represented by Eq.(4). ΔE_{MM} can be simply written as the sum of the changes in the bonded energy ΔE_{int}, the nonbonded electrostatic energy ΔE_{ele}, and the van der Waals energy ΔE_{vdW}

$$\Delta E_{MM} = \Delta E_{int} + \Delta E_{ele} + \Delta E_{vdW} \quad (19)$$

Here, ΔE_{int} is the sum of energy changes associated with bond stretching (ΔE_{bond}), angle bending (ΔE_{angle}), and dihedral ($\Delta E_{dihedral}$) energies. ΔG_{sol} in Eq.(18) is divided into an electrostatic or polar solvation energy part (ΔG_p), and a non-electrostatic or nonpolar part (ΔG_{np}) between the solute and the continuum solvent as defined in Eq.(6). Again, ΔG_{np} is proportional to SASA as showed in Eq.(7), while ΔG_p is typically computed using either the PB model (equation 8, 10, or 11) or the GB model (equation 12 or 14). The free energy of the total electrostatic contribution is the sum of ΔE_{ele} and ΔG_p ($\Delta G_{ele} = \Delta E_{ele} + \Delta G_p$) [2]. The

change in conformational entropy ($-T\Delta S$) is the sum of the change in translational, rotational, and vibrational entropic contributions. The translational and rotational entropic contributions can be determined using the standard statistical mechanical formulas, while the vibrational entropic contribution can be approximated either through a normal-mode analysis of the vibrational frequencies as adopted in our analysis or through a quasiharmonic approximation [1], [2], [160]. Lastly, ΔG_{vac} is a contribution stemming from the solute.

The relative BFE ($\Delta\Delta G$) for binding DOX to various DNA sequences are directly calculated as the difference between two corresponding BFEs [2]

$$\Delta\Delta G_{ba} = \Delta G_b - \Delta G_a = (G_{(COM,sol)} - G_{(DNA,sol)})_b - (G_{(COM,sol)} - G_{(DNA,sol)})_a \qquad (20)$$

where a and b are the reference and changed sequences respectively. Since the free DOX system remains the same in all models, its contributions are canceled.

The common protocol for performing BFE calculations using MM-PB(GB)SA begins with running MD simulations with an explicit solvent to generate the conformational ensemble. In this step, the gas-phase energies (MM) are calculated from the interactions consistent with the force field and based on the structures that are extracted from the generated MD trajectories. Subsequently, explicit water molecules and counterions are removed, and the implicit PBSA or GBSA solvent model is utilized to compute the solvation energy at different ionic concentrations. The solute conformational entropy change is estimated using a normal-mode analysis [109], [150]. Finally, the BFE is then obtained by summing these individual energy components. **Figure 2.5** shows the flowchart of these steps with two protocols of generating the necessary configuration ensembles for the bound and unbound states in the MM-PB(GB)SA method: single and multiple trajectories protocols (STP and MTP).

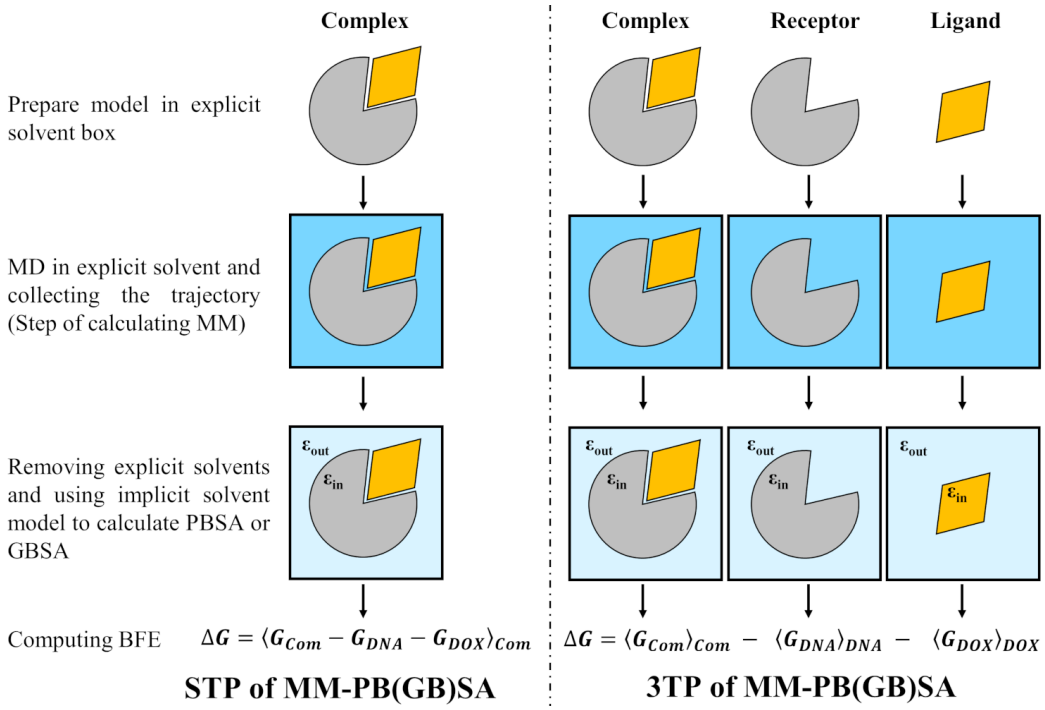


Figure 2.5. The illustration of the flowchart of the MM-PB(GB)SA method protocols for binding free energy calculations with the single-trajectory protocol (1TP) and the three-trajectory protocol (3TP).

In the STP, only one MD simulation of the bound DOX-DNA complex is performed, and all average ensembles of unbound DNA and DOX can be extracted from this single simulation. In MTP, independent MD simulations of the bound DOX-DNA complex, unbound DNA, and free DOX are run, this is known as the three-trajectories protocol (3TP). The two-trajectories protocol (2TP) can also be used when only two independent MD simulations of the complex with only receptor or ligand are conducted. Each protocol has advantages and disadvantages. STP is computationally more efficient than MTP because only one trajectory is used to create all three ensembles (DOX-DNA complex, DNA, and DOX) [1], [160]. In addition, ΔE_{int} cancel out exactly because the configurations in the bound and unbound ensembles are the same, leading to lower fluctuations and standard errors and easier

convergence for BFE [135], [160]. However, STP ignores the change in the structure of the ligand and the receptor upon ligand binding, which may be significant factors, especially when this binding process is associated with large conformational changes.

2.4.2 MM-PB(GB)SA Modification for Studying Intercalation Process

The binding process of each intercalator molecule proceeds in two steps: conformational change of dsDNA to form the drug intercalation site and the insertion proper of the drug into that site [9]. The first step is an endothermic process, which is a high energy barrier process, accompanied by strongly unfavorable interactions that play a crucial role in the kinetics and thermodynamics of the intercalation process. The second step is an exothermic process that must be larger in magnitude than the initial process if intercalation is to occur. Therefore, a detailed investigation of the energy differences involved in various steps in the intercalation process is of great scientific interest [1]. Both steps take place in an aqueous medium, and **Figure 2.6** illustrates the thermodynamic cycle of these two steps. The energy contributions of these two steps are implicitly included in Eq.(16) and can be written as [2]

$$\Delta G_{bind} = \Delta G_{Conf} + \Delta G_{Ins} \tag{21}$$

ΔG_{Conf} is the energy contribution of the conformational changes in DNA and DOX, which can be estimated in the solution medium, as shown in Eq.(22) below, as the difference between bound and unbound energies separate for DNA or DOX. The bound energy of DNA itself or DOX itself is computed from the DOX−DNA complex models by the STP of the MM-PB(GB)SA method. The energy of unbound systems is taken directly from the MD simulations of free unbound DNA and free unbound DOX models.

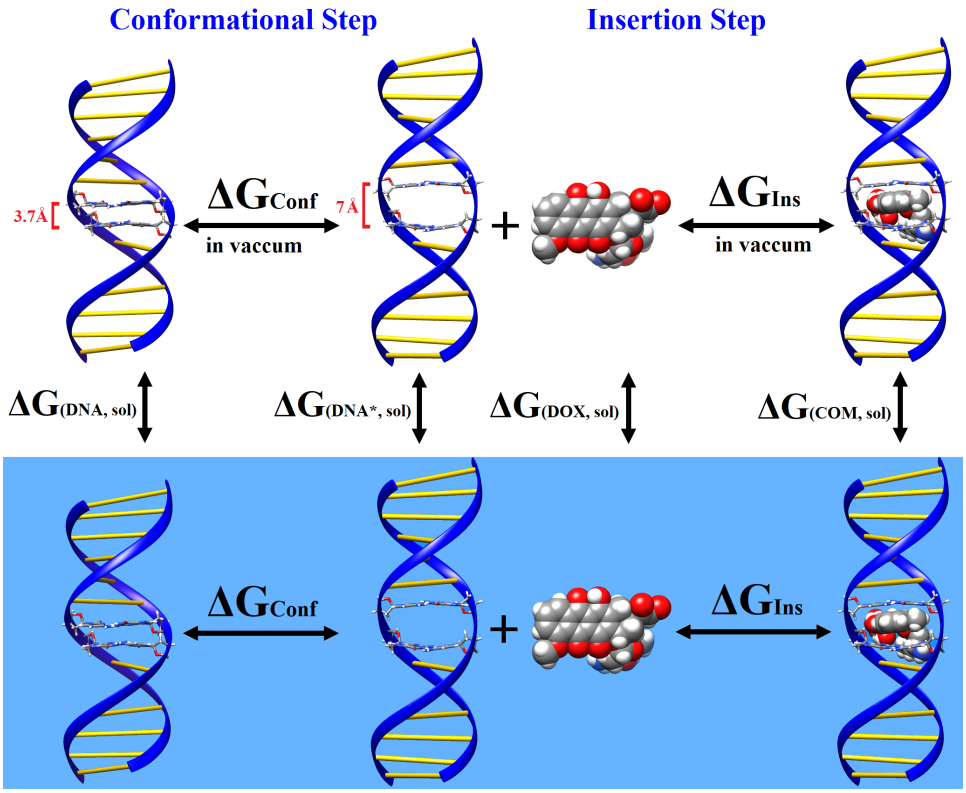


Figure 2.6. Illustration of the thermodynamic cycle for a DNA intercalator binding mode of DOX-DNA complex at the conformational changes and insertion binding stages. Blue color represents the water medium.

These conformational energies are calculated as [2]

$$\Delta G_{Conf} = \Delta G_{Conf(DNA,sol)} + \Delta G_{Conf(DOX,sol)}$$

$$\Delta G_{Conf(DNA,sol)} = \Delta G_{(DNA^*,sol)_{COM}} - \Delta G_{(DNA,sol)} \quad (22)$$

$$\Delta G_{Conf(DOX,sol)} = \Delta G_{(DOX,sol)_{COM}} - \Delta G_{(DOX,sol)}$$

where DNA^* is the intercalated DNA or locally unwound DNA. The DNA structure is usually stable B-form conformation before binding with DOX. After the binding, this B-form conformation can deform due to the generation of the intercalation site. Unlike the undisturbed DNA (B-form of DNA), the creation of an intercalation site causes a doubling of the base–

base distance. Therefore, the DNA deformation penalty is a significant term and indispensable in the calculating of the intercalation binding process [1]. $\Delta G_{(DNA^*,sol)_{COM}}$ and $\Delta G_{(DOX,sol)_{COM}}$ are defined as the energy contributions of bound DNA and DOX on complex formation while $\Delta G_{(DNA,sol)}$ and $\Delta G_{(DOX,sol)}$ are defined in the same way as in Eq.(16). By combining Eqs. 16, 20, and 21, the binding energy of the insertion step (ΔG_{Ins}) assumes the form [2]

$$\Delta G_{Ins} = \Delta G_{(COM,sol)} - \Delta G_{(DNA^*,sol)_{COM}} - \Delta G_{(DOX,sol)_{COM}} \quad (23)$$

The absolute free energy terms in Eqs.(22 and 23) are all of the forms of Eq.(17), so they can be further decomposed into physical interaction components to be discussed subsequently. Here, ΔG_{Ins} can be obtained by using the STP of the MM-PB(GB)SA method.

2.4.3 MM-PB(GB)SA Corrections for Studying Polyelectrolyte and Solvent Effects

For comparing our present results of the intercalation BFE with the experimental results of DOX-DNA intercalating complexes, two necessary corrections of the BFE are described here. First, the standard free energy of the DOX concentration effect should be taken into account. In the dilute concentration limit, the expression of the standard free energy to form an AB particle from the binding of particle A to particle B is [161]

$$\Delta G_{AB}^{0} = -RT\ ln\left(\frac{C^0 C_{AB}}{C_A C_B}\right) = -RT\ ln(K_{AB}) \quad (24)$$

where R is the gas constant, T is the temperature in Kelvin (310 K in our case), K_{AB} is the binding constant, C^0 is the standard concentration (1M), and the C_i's are the concentration of respective species. Here A denotes the drug (DOX) and B denotes the DNA-binding site which is equal to the number of base pairs divided by the binding site size (base pairs per bound drug),

which is 3 or 4 base pairs for DOX [162], [163]. Thus, C_A and C_{AB} cancel each other and a net difference in binding free energy seems from C_B and C^0 concentrations.

The second correction comes from the dependence of the intercalation BFE on the ionic concentration of the solution. Since the DNA is an anionic highly charged polyelectrolyte and DOX is a cationic drug, we expect to find a polyelectrolyte contribution to the BFE for the DOX intercalation process at each stage of the binding. The observed Gibbs free energy (ΔG_{obs}) can be experimentally decomposed into a polyelectrolyte contribution (ΔG_{pe}) and a non-polyelectrolyte contribution ΔG_t as [18]

$$\Delta G_{obs} = \Delta G_{pe} + \Delta G_t = -RT\, ln(K_{obs}) \tag{25}$$

where K_{obs} is the binding constant. ΔG_{pe} is a polyelectrolyte contribution to the ΔG_{obs} at any given salt concentration, whose origin is the release of counterions upon the charged ligand binding to dsDNA. It can be quantified by using the following expression according to the theory of Record and co-workers [18], [164].

$$\Delta G_{pe} = (SK)RT\, ln[Na^+] \tag{26}$$

where SK is the salt dependence of the binding constant. It is obtained by plotting $log(K_{obs})$ vs. $log[Na^+]$, with SK being the linear slope, $SK = \frac{\delta log(K_{obs})}{\delta log[Na^+]}$. The ΔG_t is a term of the BFE that is independent of the salt concentration, resulting from other interactions such as van der Waals interactions, hydrophobic interactions, etc. It is referred to as the standard state of BFE at 1 M monovalent salt [165] with the corresponding ΔG_{pe} equal to zero and salt independent. By using the polyelectrolyte theory, we can also theoretically partition the ΔG_{bind} into terms analogous to the two polyelectrolyte contributions, as in Eq.(25). By using MD simulations

with the MM-PB(GB)SA method, we have quantified the effect of the ionic strength on the BFE at ten different salt concentrations, ranging from 0.1 to 1 M NaCl. Once the SK slope is determined, the ΔG_{pe} can be estimated by using Eq.(26). Alternatively, the ΔG_{pe} can be simply evaluated by taking the difference between the BFE at a specific ionic concentration and the ΔG_t, which is a BFE at 1 M [1]. In our study, the true binding free energy is denoted as ΔG_t^{sim} to distinguish it from ΔG_{bind} without any corrections or ΔG_t from experiment. We also use ΔG_{calc} to represent the subtraction of ΔG_{con} from ΔG_{bind}.

2.5 Analysis Tools

Analysis tools are chosen based on our interest in the model. The following sections describe the tools that have been used for most of the MD simulations in this work.

2.5.1 Root Mean Square Deviation (RMSD)

RMSD is used to analyze the structural stability of our models and the convergence of MD simulation. It is a quantitative measure of the structural difference between two coordinates: a target and a reference, and it is defined as

$$RMSD = \sqrt{\frac{1}{N}\sum_{i}^{N}(X_{j,i} - Y_{R,i})^2} \tag{27}$$

where N is the number of atoms, $X_{j,i}$ is the coordinate vector for target atom i at each trajectory frame j, and $Y_{R,i}$ is the coordinate vector for reference atom i at a reference structure R usually the initial structure for the MD production simulation. In the current study, RMSD was computed for each model from MD simulation trajectories using the CPPTRAJ program [166]

in AMBER by taking the starting point for the MD production simulations (after the minimization and equilibration procedures) as reference. CPPTRAJ employs a version of Kabsch's algorithm [167] for calculating the best-fit RMSD of a structure to a reference structure, which means each structure is rotated and translated to minimize the RMSD to the reference structure [157]. RMSD is typically plotted vs. time of MD simulation.

2.5.2 Root Mean Square Fluctuation (RMSF)

RMSF is a measure for thermal motion that is highly correlated with the experimental B-factor. It is an indicator of individual residue flexibility, or how much a specific residue fluctuates during an MD simulation. RMSF per residue is typically plotted vs. residue number and can show structurally which amino acids in a protein or base pair in DNA contribute the most to molecular motion. RMSF for each residue i is given by

$$RMSF = \sqrt{\frac{1}{T}\sum_{i}^{T}(x_i(t_j) - \bar{x_i})^2} \tag{28}$$

where T is the total number of simulation frames, $x_i(t_j)$ is the coordinates of the residue x_i at time t_j, and $\bar{x_i}$ is the average coordinate of the residue x_i over all frames.

2.5.3 Principal Component Analysis (PCA)

Because RMSD calculations can't capture all structural conformations in phase space perfectly, multidimensional analyses of the sampling of conformational space as a function of time (e.g., by PCA) are much more rigorous. PCA is a standard mathematical tool that can be used to extract large-scale motions occurring in the MD trajectory, providing a brief picture of the underlying structure of atomic fluctuations by applying the dimensionality reduction

method [1]. This technique is based on the determination of a new set of collective coordinates called the principal components (PCs) or "modes" through a linear transformation of the atomic coordinates. The PCs are described as the eigenvectors of the atomic displacement covariance matrix. They represent a correlated motion of many atoms in a 3-dimensional space, and the corresponding eigenvalues indicate the extent of the total motion occurring in each direction. Usually, the first few principal components (PCs) are sufficient to describe the most important slow modes of the total motion observed during the dynamic, which are related to the functional motions of a biomolecular system [168]. PCA has been widely used to study the intrinsic motions of various biomolecular systems such as nucleic acids and proteins [169], [170].

In the present work, PCA was performed with R software using the bio3D package [171]. We applied PCA to the heavy atoms of the complex, DNA and DOX in order to investigate and compare the functional motions of DNA-free and DNA-bound complexes. Our PC analysis is based on Cartesian coordinates.

2.5.4 Block Analysis

Block analysis is an efficient and simple approach for estimating the statistical error of the time average of the quantity in the MD simulation [172]. In this work, this approach was used to calculate the standard mean error (SEM) of the binding free energy (ΔG) or the absolute free energy (G) for each model. This approach works by dividing the time steps of the G or ΔG into large consecutive blocks of a specific size. Then, the average of each block is calculated, and SEM is determined from the variance of these block averages.

2.5.5 Hydrogen Bonds (HBs)

For identifying the global structural features, we focus only on the hydrogen bonds (HBs) between the DOX and DNA which are identified by using the HBonds plugin in VMD [164]. In general, there would be an HB between an electronegative atom (the donor, D) and another electronegative atom (the acceptor, A). In this work, the HBs are counted based on geometric criteria using a cutoff acceptor and donor (A...D) distance < 3.5 Å and 25° angle cutoff among the acceptor, hydrogen, and donor atoms. The percentage of HB occupancy gives an account of how many times a certain HB forms during the MD simulations. The higher the frequency, the higher the stability of this HB [150]. For HBs within DNA base pairings, 3DNA program can be used to achieve the Watson-Crick (WC) base pairs which are three HBs for each C-G base pairing and two for each A-T base pairing [174].

2.5.6 Geometrical Parameters for Characterizing dsDNA Conformation

Many parameters are used to define the helical geometry of the dsDNA. They are schematically illustrated in **Figure 2.7**. These include six local base-pair parameters, six local base pair-step parameters, and four local base pair-axis parameters in a Cartesian coordinate system (X: short axis of the paired base plane, Y: long axis of the paired base plane, and Z: the DNA helix direction). The six base-pair parameters are shear, stretch, stagger, buckle, propeller, and opening which defines one base's deviation from its paired one. Shear, stretch, and stagger parameters describe the translational deviations from ideal base-pair geometry, while the other three (buckle, propeller, and opening) characterize the rotational deviations with respect to X, Y and Z axis respectively.

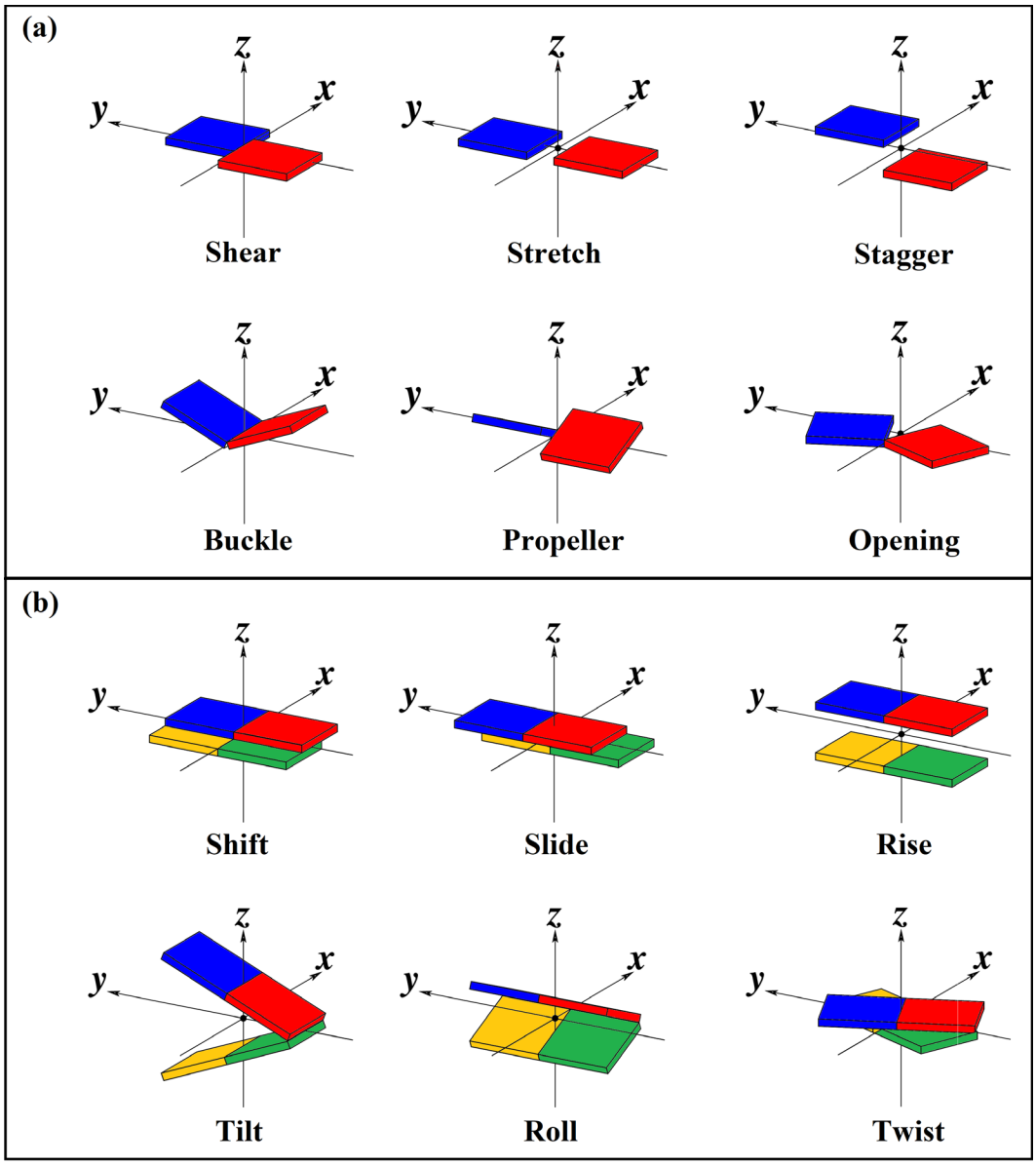


Figure 2.7. DNA Geometric parameters. (a) Base-pair parameters and (b) Sequential base pair-step parameters [174]. In each figure, the first three parameters are translational, while the other three are rotational. DNA base pair is shown by rectangular solids (blue and red for BP and gold and green for adjacent BP).

The base pair-step parameters refer to the deviation of two consecutive base pairs from their ideal geometries relative to each other. These include translation deviations (Shift, Slide, and Rise) and rotational deviations (Tilt, Roll, and Twist). The four local base pair-axis

parameters (X-displacement, Y-displacement, inclination, and tip) represent a base pair's position and orientation relative to the helical axis, here identified by a two-base pair repetition (not shown in **Figure 2.7**).

Our special interests lie in the DOX intercalation site between DNA base pairs because there are specific conformation changes in the DNA associated with the formation of this site of DOX. These geometric parameters were thoroughly analyzed using the 3DNA program [174]. Here, we used the unbound DNA models as a reference to check conformation changes when DOX inserted between their base pairs. These analyses are based on the average coordinate structures extracted from the entire MD simulation of bound and unbound DNA. Sugar-phosphate backbone and glycosyl torsion angles, and sugar conformation were also tested using the 3DNA program. These angles are shown in **Figure 2.8** [174], [175].

The backbone torsion angles of DNA in 3DNA program are defined as

Alpha (α):	O3'(i-1)-P-O5'-C5'
Beta (β):	P-O5'-C5'-C4'
Gamma (γ):	O5'-C5'-C4'-C3'
Delta (δ):	C5'-C4'-C3'-O3'
Epsilon (ε):	C4'-C3'-O3'-P(i+1)
Zeta (ζ):	C3'-O3'-P(i+1)-O5'(i+1)
Chi (χ) for pyrimidines:	O4'-C1'-N1-C2
Chi (χ) for purines:	O4'-C1'-N9-C4.

P is the phase angle of pseudorotation of the sugar ring.

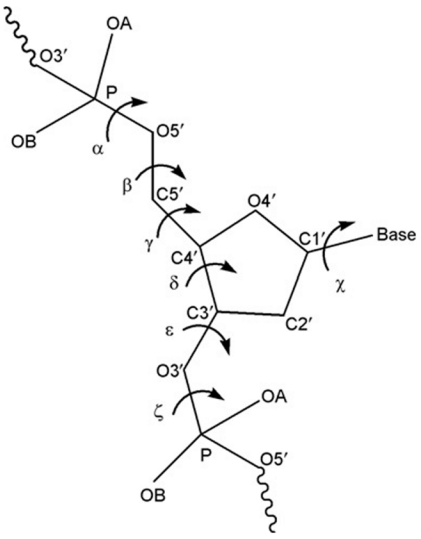


Figure 2.8. The backbone and glyosidic torsion angles in a unit nucleotide.

CHAPTER 3. MOLECULAR MECHANISM AND BINDING FREE ENERGY OF DOXORUBICIN INTERCALATION IN DNA

3.1 Introduction

Many drugs, and in particular chemotherapies, interact with DNA either by groove binding or intercalating between the DNA bases and therefore interfere with cell critical functions such as replication or protein expression [176]. The intercalator drugs directly interact with DNA through the insertion of its planar aromatic ring systems between sequential base pairs of DNA. DOX is a DNA intercalator widely applied in the clinical setting to treat a wide range of malignancies [177]. The detailed molecular mechanism of DOX activity is still not completely elucidated and remains a subject of considerable debate. Many molecular mechanism models have been proposed including DNA intercalation and inhibition of DNA biosynthesis [178], interference with topoisomerase II [22], [179], [180] and induction of DNA double-strand breaks [20] and interference with DNA unwinding [181], [182]. The interaction between DOX and DNA participates in all these proposed models. Therefore, a comprehensive analysis of DOX-DNA interactions can provide insight into the molecular mechanism of intercalator drug action, as well as a better knowledge of the basic principles for developing a new intercalator drug.

Within the last few decades, several experimental and theoretical studies have been performed to understand the relationship between the biological functions and structural characteristics of DOX. Experimentally, there are numerous high-resolution X-ray [34], [35], [162], [183] and NMR analyses [184] of DNA-anthracycline complexes, and detailed structural explanations of DOX as a DNA intercalator have been proposed. Chaires et al.

suggested that the drug-DNA binding process occurs consecutively in three-step: "outside" binding, drug intercalation, and "reshuffling" of the drug at the intercalation site [185]. Rizzo et al. proposed a five-step kinetic model with two additional sub-steps carrying out at the first and third steps of the three-step model [186]. Recently, different methodologies have been applied to study the interaction of DOX with the synthetic polynucleotides of the duplex [27], [187], [188] or single-stranded DNA (ssDNA) [189], native DNA such as calf thymus DNA (ctDNA) [190], [191] and salmon DNA [192], the DNA hairpin [193], [194], G-quadruplex DNA [195], [196] as well as with RNA [197] and tRNA [198]. However, experimental studies alone cannot clearly pinpoint the details of the complicated dynamic behavior of the complex formation. In particular, the energetic contributions of non-bonding interactions that are crucial in determining the intercalation process are still not firmly established.

On the other hand, several theoretical calculations were performed on these drug-DNA complexes. Nakata et al. performed molecular modeling for the conformational analysis of DOX-DNA complex [199], [200]. Barone et al. [201] and Zhu et al. [202] carried out the density functional theory (DFT) to study the anthracycline–DNA intercalation adducts to determine the most stable geometry in vacuo to characterize the structure of DOX [202]. Zhu and co-worker concluded that (i) the intramolecular hydrogen bonds between the quinone and hydroquinone in anthracycline drug plays a vital role to stabilize the complex, (ii) the side chain is perpendicularly aligned to the anthracycline system, and (iii) the sugar residue is the most flexible part of the DOX drug. Poudel et al. studied the electronic structure and partial charge distribution of DOX in three different molecular environments: isolated, solvated, and intercalated in a DNA complex [203], by using first-principles DFT-based methods [204]–

[206]. Barthwal et al. studied self-aggregated drugs in aqueous solution and drug-DNA complexes using restrained molecular dynamics (RMD) with constrained H-H distances [207]–[209].

In the present work, we use the MM-PBSA and MM-GBSA methods to calculate the BFE of DOX-DNA intercalating complexes in two sequences DNA1 and DNA2 of dsDNA. DNA1 and DNA2 sequences are d(CG**AT**CG)$_2$ and d(CG**TA**CG)$_2$, respectively, with the only difference being the reversed position of the base pair AT in the sequences. These two sequences contain two 5'-C|G**A**-3' or 5'-C|G**T**-3' binding sites which allow us to simulate two DOX molecules with a single DNA (2:1 complex) as well as a single DOX molecule with a single DNA (1:1 complex). Here the vertical line | represents the DOX intercalation site. First, we analyze the energetic contributions of DOX bound to DNA sequences using MD simulations implemented in AMBER package. To the best of our knowledge, this may be the first work to compare the BFE of DOX in two different sequences due to the changes in the position of the DOX ($NH3^+$) daunosamine sugar group in the minor groove binding. Another goal of this work is to determine the DNA deformation energy by calculating the energy difference between bound DNA in various DOX-DNA complexes and an unbound (free) B-form DNA. Additionally, this work focusses on the origins of cooperative interactions that occur when the second molecule of DOX is bound to the 1:1 complex to form the 2:1 complex. Although both MM-PBSA and MM-GBSA approaches have been shown to give estimated free energy in good agreement with the experiment, it is worth mentioning that none of them reproduces accurately the absolute experimental BFE values. In this work, we modified these methods to include the DNA deformation energy at the intercalation site and the impact of

drug and ionic solution concentrations. We have also studied the effect of salt concentrations on the BFE, which allows us to predict a more realistic BFE. Finally, we also address the solute entropic contributions.

3.2 Molecular Models Setup

An important goal of the present work is to understand the molecular mechanism and predict the BFE for the intercalated DOX-DNA complex. To this end, we constructed two free DNA models with different DNA sequences (DNA1 and DNA2) and four intercalated DOX-DNA models, that have different configurations of one DOX molecule (1:1 complex) or two DOX molecules (2:1 complex) with DNA1 and DNA2. DNA1 has a sequence of d(CG**AT**CG)$_2$ obtained from the structure resolved by X-ray diffraction (PDB ID:1D12) [34], while DNA2 has a sequence of d(CG**TA**CG)$_2$ selected from another structure resolved by X-ray diffraction (PDB ID:1D11) [35]. Since the (1D11) structure source contains two daunorubicin (DNR) drugs with the DNA2 sequence, we modified it by substituting DNR with DOX to create the DOX-DNA complex for DNA2. From these initial structures of 1D12 and 1D11, we build three different models for each DNA sequence: the free DNA, the DOX-DNA complex with 1:1, and the 2:1 complex. Therefore, in total, we have built six models that are summarized in **Table 3.1** and shown in **Figure 3.1**. The first two models are only DNA models from unbounded DNAs with different sequences of dsDNA obtained separately from two crystal structures (1D12) and (1D11), respectively. These two models were used as a reference model for DNA1 and DNA2. The second two models are the 1:1 complex constructed from only one DOX molecule that intercalated into the DNA of each sequence. The final two models are the 2:1 complex created from two DOX molecules with DNA1 or DNA2.

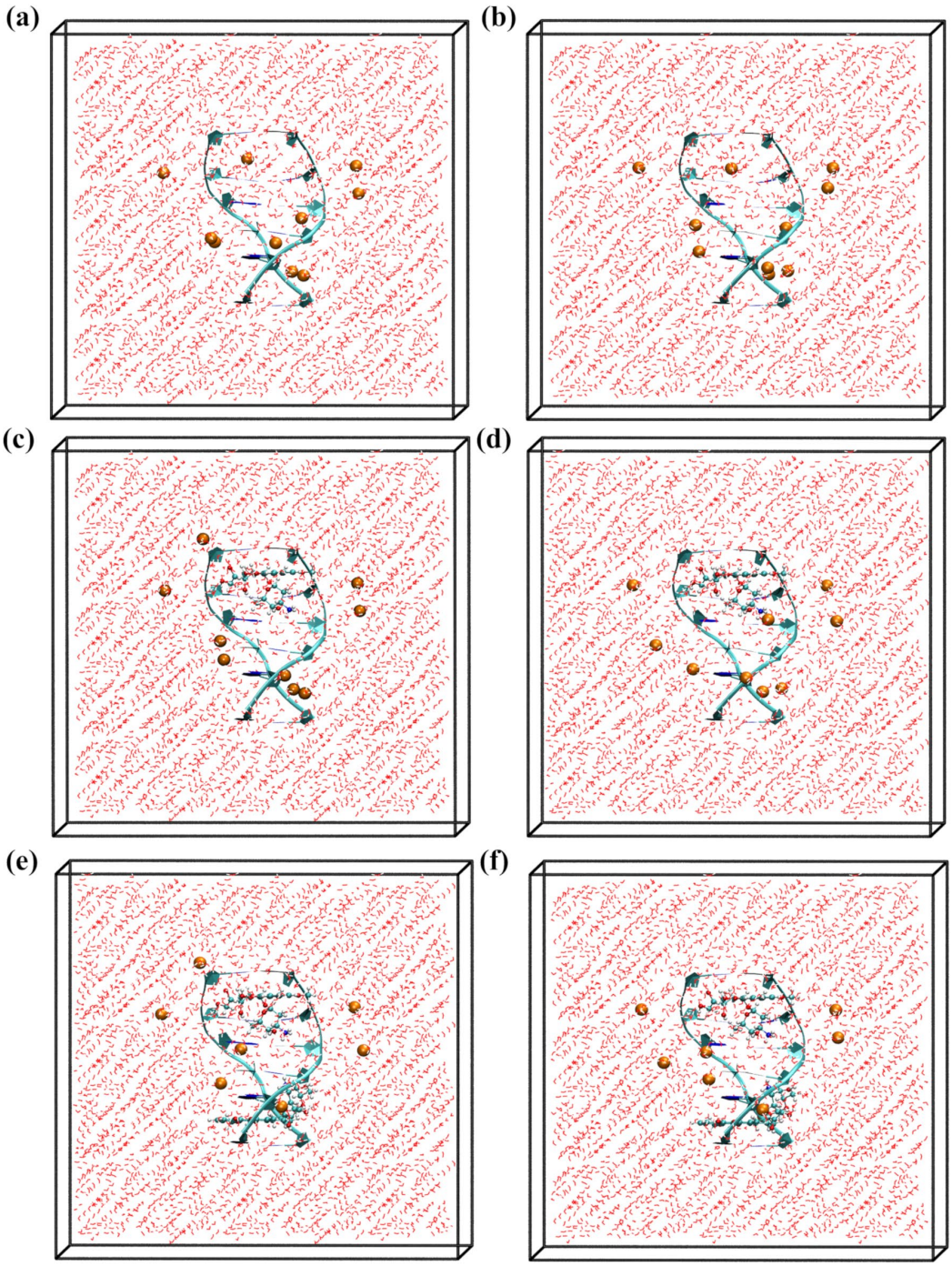


Figure 3.1. Molecular structure modeling of MD simulations. (a) and (b) for solvated free DNA models (M1 and M2 respectively); (c) and (d) for 1:1 solvated complex models (M3 and M4); and (e) and (f) for 2:1 solvated complex models (M5 and M6). The orange balls represent Na ions, the cartoon represents the DNA, the balls and sticks represent the DOX molecules and the red wires represents water molecules.

Table 3.1. Summary of MD simulation models performed in this study.

Models	Atom number				
	DOX	DNA	Water	Na^+ ions	Total
M1	---	378	9321	10	9709
M2	---	378	9252	10	9640
M3	69	378	9300	9	9756
M4	69	378	9498	9	9954
M5	138	378	9510	8	10034
M6	138	378	9360	8	9884

For all models, nucleotides on strand 1 are labeled Cl to G6 in the 5' to 3' direction and C7 to G12 in the 5' to 3'direction on symmetry-related strand 2. The DOX molecule is numbered DOX1 in 1:1 complex models, and DOX1 and DOX2 in 2:1 complex models as shown in **Figure 3.2**.

The specific purposes of using these six models are: first, to identify the driving force that plays a significant role in forming the stable intercalation complex of DOX-DNA. Second, to probe the impact of the dsDNA sequence on the DOX-DNA binding process. Third, to investigate the conformational penalty when one or two DOX molecules are intercalated within DNA. Finally, to study the cooperative effect when the second DOX molecules are bound to 1:1 complex to generate 2:1 complex.

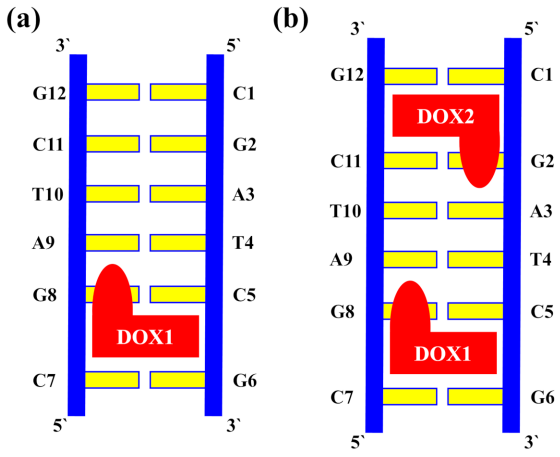


Figure 3.2. The nucleotide complementary bases of DNA d(CGATCG)$_2$ or DNA1 and the intercalated sites of DOX drug.

The molecular mechanics (MM) parameters of a DOX drug are obtained following the standard AMBER 11 protocol [210]. In particular, the electrostatic potential of DOX was obtained after geometry optimization using Gaussian 09 at the HF/6-31G* level [211]. The partial charges are obtained by fitting the electrostatic potential using the RESP (Restrained Electrostatic Potential) method [212] in the R.E.D. server [213], [214]. They are listed in **Table S1 in Appendix**. Other parameters of DOX were taken from the AMBER GAFF/GAFF2 parameter set [74]. The AMBER force field parmbsc0 [73] is used to represent the parameterizations of the intermolecular and intramolecular interactions of dsDNA. Each of these six structural models (M1, M2, M3, M4, M5, and M6) are appropriately solvated with water molecules and described below.

3.2.1 Solvated Free DNA Models (M1 and M2)

M1 and M2 are created by extracting only the isolated double-stranded DNA1 and DNA2 structures from PDB ID: 1D12 and 1D11, respectively, and removing the DOX or DNR

molecule as well as all other bathing solution molecules. To solvate the DNAs, we have inserted the isolated dsDNA into the water box of 10 Å in each direction with periodic boundary conditions by using the TIP3P model [92] in the AMBER 11 package [210]. The water box is maintained to be cubic which having dimensions of 52× 52×52 $Å^3$ by specifying a list of numbers to the solvateBox command through the LEaP module from AMBER 11 [210]. The dsDNA fragment has six base pairs with a total charge of -10e, and to neutralize the system, 10 Na+ ions are added as counter ions.

3.2.2 Solvated 1:1 Complex Models (M3 and M4)

The intercalated structure of the DOX-DNA1 complex is taken from PDB ID 1D12 [34]. The 1:1 complex is built to contain only dsDNA molecule with one protonated DOX molecule by removing all the other molecules from the 1D12 structure. A molecular structure of protonated DOX is $C_{27}H_{30}NO_{11}$ with 69 atoms. For the solvation of the 1:1 DOX-DNA complex, we have used the same approach as in free DNA models. We have also added 9 Na^+ ions as a charge compensation and this model is named as M3. Similarly, we have started from source data 1D11 to generate a 1:1 complex with a DNA2 sequence [35]. We should mention here that the 1D11 structure contains DNR instead of DOX. In this case, we replaced the DNR molecule with just one protonated DOX molecule to create a 1:1 DOX-DNA2 complex. After that, all the other bathing solution molecules were removed, and the complex has been solvated using the same procedure as in previous models. The same number of Na^+ counter ions (9 Na^+ ions) as in M3 was used to neutralize the system, and this model is called M4.

3.2.3 Solvated 2:1 Complex Models (M5 and M6)

The DOX-DNA complex with two DOX molecules (2:1 complexes) is created using the same way as the 1:1 complex, except that two DOX molecules are inserted into the DNA rather than one. Because of these two protonated DOX molecules, only 8 Na^+ ions are required to neutralize the systems. These models are named M5 and M6 for 2:1 DOX-DNA1 and DOX-DNA2 complexes respectively.

3.3 Computational Details

3.3.1 Molecular Dynamics (MD) Simulation

All MD simulations for the six models have been performed using the AMBER 11 simulation package in an explicit solvent with periodic boundaries and the Particle Mesh Ewald (PME) potential function [101]. Before MD simulations, two minimization stages have been used to remove possible steric clashes and adapt the systems to the chosen force field. The first stage kept the solute (DOX-DNA complex or DNA) fixed with a force constant of 500 kcal/mol-$Å^2$ and only optimize the positions of the water and ions, necessary to get the water and ions properly randomized. In this stage, each one of the six models was first minimized for 5000 steps with the steepest descent, followed by 5000 steps of conjugate gradient while keeping the positions of solute fixed. In the second stage of minimization, the entire system is minimized without the restraints for additional 10000 cycles (5000 steps for steepest descent and 5000 steps for conjugate gradient). After minimization, the next stage is to gradually heat our system from 0 K to 310 K for 310 picoseconds (ps) using the NVT ensemble with a 10 kcal/mol-$Å^2$ weak restraint on the solute (complex or DNA). Then, 0.5 ns without the restraint of constant pressure of 1 bar and temperature 310 K (NPT) to allow the system to reach the

proper density. The systems were then equilibrated over 3 ns using NPT ensemble through six independent simulations, the length of each one is 0.5 ns. The following settings were activated in all of the equilibration MD simulations: Langevin dynamics for temperature scaling, 2 ps as the pressure relaxation time, long-range electrostatic interactions were calculated with the PME, both the direct space PME and Lennard-Jones cutoffs were set at 10 Å, the SHAKE algorithm [66] was used to constraint bond length of hydrogen atoms to avoid high-frequency motions involving hydrogen atoms, and 1 fs time step. All the minimizations and equilibration steps were conducted using the SANDER module of AMBER 11. Finally, 30 ns NPT production run with 30 independent MD runs was performed at constant pressure (1 bar) and temperature (310 K). The length of each independent MD production run is 1 ns. During the production run, the atomic coordinates from trajectories were saved every 2 ps for subsequent MM-PB(GB)SA analyses. All settings that were activated in the equilibration MD simulations were kept during the production run except the 2 fs time step is used instead of 1 fs time step. The PMEMD program in AMBER11 was used for production MD simulations.

3.3.2 BFE Calculation Using MM-PB(GB)SA Methods

Briefly, the entire binding profiles of the DOX-DNA complex with 1:1 and 2:1 configurations have been computed using the most popular BFE method, MM-PB(GB)SA. Although both methods show good overall agreement with experimental values, none of them can give sufficiently accurate values. This motivated us to modify these methods by considering two basic corrections from the standard free energy of the DOX concentration effect and the ionic concentration dependency of the intercalation BFE. The background of the methods, including the theory of corrections, was fully described in section 2.4 of chapter 2.

After MD simulations, snapshots were taken for every 10 ps from 0 to 30 ns, so 1500 snapshots were extracted for post-process binding free energies using MM-PB(GB)SA methods. BFE calculations by MM-PBSA or MM-GBSA methods were performed using the MMPBSA.py module of AMBER 11[160]. They are described succinctly as follows.

In MM-PBSA approach, the following settings are used to calculate the BFE of DOX-DNA intercalating complexes: the value of an exterior dielectric constant was set to 80 while for solute dielectric constant was set to 1 [215], the MM-PBSA surface tension (γ) and the non-polar free energy correction term (β) were set to 0.00542 kcal/mol-$Å^2$ and 0.92 kcal/mol, respectively according to the PARSE and mbondi2 sets [120], [205].

In the MM-GBSA method, the model developed by Onufriev et al. (GB^{OBC} with igb = 2) [127] was used as the GB model in this study. By the antechamber program in AMBER 11, the mbondi2 radii set was prepared. The default setting of MM-GBSA surface tension (γ = 0.005 kcal / mol Å2) and the non-polar free energy correction term (β = 0) were applied. The value of an exterior dielectric constant was set to 78.3 and for solute dielectric constants was used 1 as in the MM-PBSA method.

The normal mode analysis was used to evaluate the solute entropic contribution of DOX-DNA complexes. In the normal-mode analysis, the conformational entropy changes ($-T\Delta S$) upon binding DOX to DNA were estimated using the nmode program through the MMPBSA.py module of AMBER 11 [160]. Due to the limitation of the computationally expensive normal-mode analysis, we only considered the residues within a 12 Å sphere centered at the ligand, and these residues were retrieved from an MD snapshot for each DOX-DNA complex. Then, each structure was fully minimized for 10000 steps using a distance-

dependent dielectric of $4r_{ij}$ (r_{ij} is the distance between two atoms) to mimic the solvent dielectric change from the solute to solvent. To reduce the computational demand, 150 snapshots were taken from 0 to 30 ns to estimate the contribution of the entropy to the binding. The final conformational entropy was obtained from the average over the snapshots [1].

By using MD simulations with the MM-PB(GB)SA method, we have quantified the effect of the ionic strength on the BFE at ten different salt concentrations, ranging from 0.1 to 1 M NaCl. The electrostatic screening effects of monovalent salt are implicitly accounted by using the Debye-Huckel screening of charges [2]. As shown in chapter 2, the Debye–Huckel screening parameter is proportional to the square root of the ionic strength ($\kappa^2 = 8\pi e^2 I/\varepsilon_w\, k_B T$). Since this parameter is included in a modified GB or PB theory based on the Debye-Huckel limiting law for ion screening of interactions, one can set a specific salt concentration in molarity (M) of 1-1 mobile counterions in solution [2]. This task can be accomplished by setting a certain ionic concentration as input in the MM-PBSA.py module of AMBER. Again, it should mention that all explicitly charged ions used to neutralize the system are removed before estimating BFE by using MM-PB(GB)SA method. Then the ionic effects are treated implicitly by using a modified GB or PB theory based on the Debye-Huckel screening model [2].

3.4 Results and Discussion

3.4.1 DNA Conformations

The DNA conformation parameters such as sugar-phosphate backbone and glycosidic torsion angles are very important to elucidate the binding mechanism in drug-DNA complexes. They schematically illustrated in **Figure 2.8**. The conformation parameters for the six models

are determined using the 3DNA program [174] and listed in **Tables S2 to S7 in Appendix**, as well as compared with experimental values for B-DNA [216].

In M1 and M2, the phosphodiester linkage in DNA represented by α and ζ torsion angles along the DNA backbone in terms of g (gauche) or t (trans) conformations goes from 5' to 3' direction adopted the gg conformations in all nucleotide residues are in full agreement with B-DNA conformation (**Tables S2** and **S3**). These conformations are slightly changed in the DOX-DNA complexes, especially at the DOX intercalation site(s). We have tabulated $gg\_gg\_gg\_gg\_gg\_tg$ conformation changes for models M3 and M4 (1:1 complexes) and $gg\_tg\_gg\_gg\_gg\_tg$ for models M5 and M6 (2:1 complexes) (**Tables S4 to S7**). It shows that the conformation parameters are close to a B-DNA form of free DNA (M1 and M2) but deviate from it in DOX-DNA complexes. It can be therefore concluded from the value of α and ζ torsion angles that cytosine (C) is influenced more than guanine (G) at the intercalation site(s), as the phosphodiester linkage of these nucleotides (C residues) faces the amino sugar side of DOX. The increased values of the phosphodiester linkage in model M6 are slightly larger than in the case of the M5 model. In addition, there is a reduction in β angle in all complex models at the central residue (A or T residue) of the intercalation site. Furthermore, there is an additional change due to ε angle in residues C5 and C7 in M3 and M4, as well as in all C residues in models M5 and M6. These observations of torsional angle changes are consistent with the experimental results [34], [35]. The glycosidic torsion angles χ between the sugar and the base in the CpG step(s) for intercalated DOX, which is adopted in an anti-conformation, is also slightly changed from the unbound models (M1 and M2). The same behavior of χ angle is found at the intercalation site in both 1:1 and 2:1 complex, however, we

notice an abnormal situation for χ angle at C1 residue in 1:1 complex (see the values of χ angles in **Tables S4 and S5**), which may influence the formation HBs in these models. A possible reason behind this abnormal situation could be the large fluctuations of the terminal base-pairs during MD simulations.

We also calculated the helical rise parameters of DNA1 and DNA2 for three different situations (free, 1:1 complex, and 2:1 complex) and we separately plot them in top and bottom of **Figure 3.3(a)**, respectively. The rise paratemers are almost identical for different sequences of DNA when DOX(s) are intercalated into CpG step(s). There is an increase in the separation between CpG step(s) to intercalate DOX in DNA of about 4.15 Å, consistent with the changing of glycosidic torsion angle explained in the previous paragraph. Hence, the formation of the intercalation sites can be achieved by coupling the change in both χ and ε torsion angles to the rotation of the phosphodiester linkage from a normal gauche (gg conformation as in free DNAs) to a trans conformation (tg as in intercalated complexes). This agrees also with experimental studies [34], [35] reporting an increase in the rise parameter by 4.1 Å per intercalation site.

We have also estimated the helical twist parameters of DNAs in three different situations. They are plotted in **Figure 3.3(b)**. The behavior of twist parameters is quite different when varying the DNA sequences. In M5, the helical twist is reduced from 36° in M1 to about 32° at G2pA3 or G8pA9 steps, and about 30° at central step A3pT4. This behavior appears only on one side in M3, where the DOX is intercalated into DNA (**Figure 3.3(b)**). In M6, the helical twist is first reduced from 36° in M2 to about 30° at G2pT3 or G8pT9 steps and then increases to about 38° at the T3pA4 step. In M4, one observes an identical behavior only on

one side where the DOX is inserted into DNA. Therefore, the unwinding of DNA at the intercalation site of the intercalated complex is up to 4° in DNA1 and 6° in DNA2. This corresponds quite well with 6° of unwinding angle of DNA by these types of drugs (DOX or DNR) using viscosity techniques by Chaires [163]. Our calculated values of twisting angles also lie within the range of X-ray structure studies [34], [35].

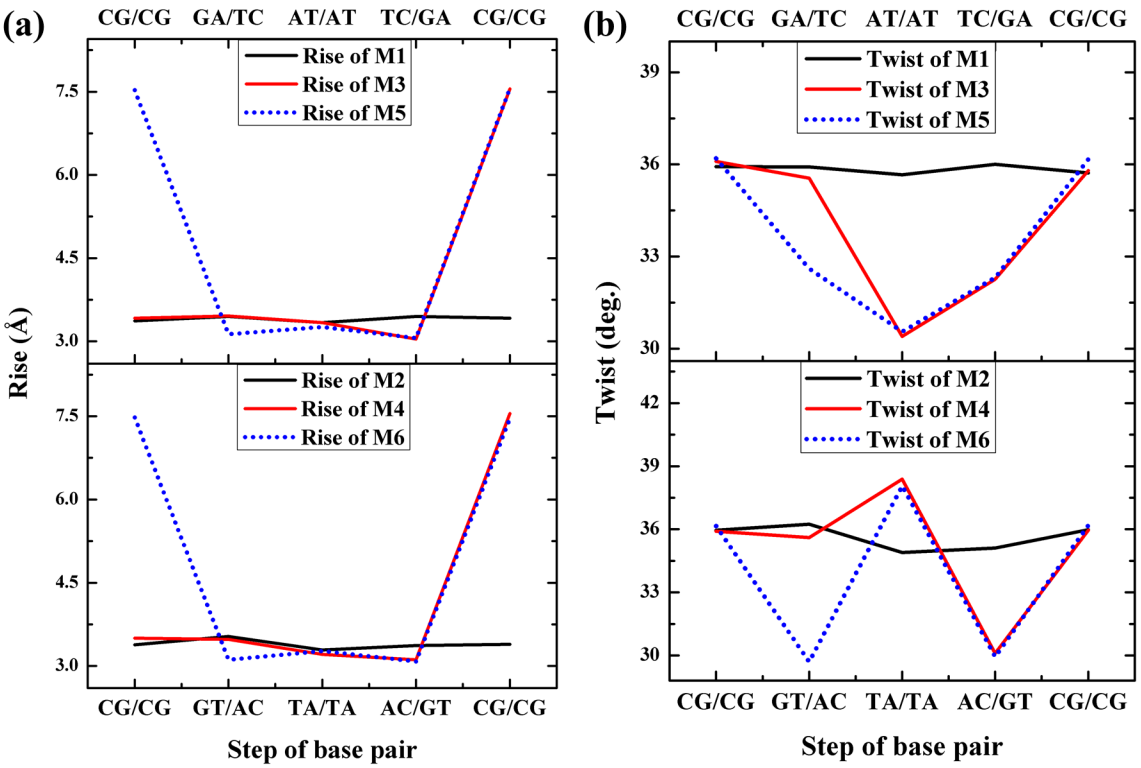


Figure 3.3. Rise and twist helical step parameters of the simulated models. The top and bottom part of figure displayed for DNA1 and DNA2 respectively with three different environments (free DNA, 1:1 complex and 2:1 complex) for (a) helical rise and, (b) helical twist.

3.4.2 Hydrogen Bonds Analysis

Hydrogen bonds (HBs) play a salient role in all biological systems. It holds the key to understanding biomolecular system stability and binding. We have calculated the total number

of HBs between DNA base pairs for all models using the 3DNA program. Moreover, the HBs between specific atoms of DOX and DNA nucleotides in the DOX-DNA complexes, are obtained using HBonds Plugin in VMD [173], based on the following criteria: the donor-acceptor distance is less than 3.5 Å and the donor-acceptor-hydrogen angle is less than 25°.

The previous X-ray structure studies [34], [35], [162], [183] suggested that the DOX is anchored by two direct HBs from its O9 hydroxyl to the N3 and N2 atoms of an adjacent guanine base in both DNA1 and DNA2. The HBs between N3' of the amino sugar of DOX and DNA remain a subject of some debate. Frederick et al. mentioned that the N3' of the amino sugar of DOX could form three HBs with DNA1 but not with DNA2 [34]. However, Wang et al. [35] and Nunn et al. [162] suggested that the N3' of the amino sugar of DOX could form two HBs with DNA2 as well as three HBs with DNA1. These HBs between the DOX and DNA revealed by these previous X-ray studies are now investigated in more detail using our MD study. In this respect, five HBs between the DOX and both sequences of DNA are examined in our study. They are O12(DOX)-N3(G8), O12(DOX)-N2(G8), N3'(DOX)-O4'(C5), N3'(DOX)-O2(T4), and N3'(DOX)-O2(C5).

First, there is a total of 16 HBs between DNA base pairings in all modeled systems except for M3 and M4 where there are only 15. The missing HB is located at the terminal base pair (C1-G12) on the one side of DNA with no DOX intercalated. This may be attributed to the abnormal χ angle at C1 residue alluded earlier for M3 and M4.

Second, the occupancy of the HBs between DOX and DNA as a function of MD simulation time for M3 and M4 are estimated and presented in **Figure 3.4(a)** and **(b)** respectively.

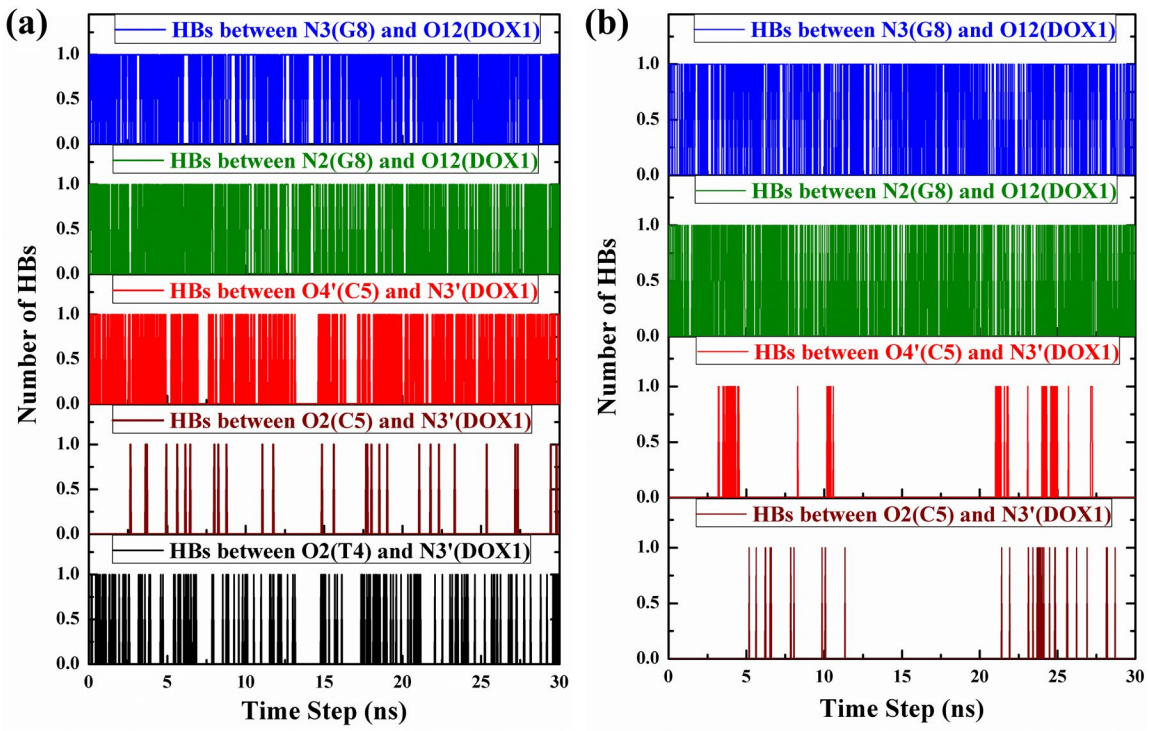


Figure 3.4. Evaluation of the total number of H-bonds between the specific atoms of DOX and DNA residue through whole MD simulation in 1:1 complex of (a) DOX-DNA1, and (b) DOX-DNA2.

This shows that the two HBs between the O12 hydroxyl of DOX and N3 and N2 atom of guanine base are highly conserved throughout the simulation. The percentage of HB occupancy between O12(DOX1)-N3(G8) is 72% which is higher than that between O12(DOX1)-N2(G8) (52% occupancy) in both M3 and M4. In M3, the N3' of amino sugar in DOX forms three HBs with DNA1, but with lesser occupancy. For instance, the HB occupancy between N3'(DOX1)-O4'(C5) is only 30%, N3'(DOX1)-O2(T4) is 18% and N3'(DOX1)-O2(C5) is merely 3%. The HBs distances from RMD analysis of Jain et al. [208] are 3.65 Å, 3.96 Å, and 4.32 Å for N3'(DOX)-O2(C5), N3'(DOX)-O4'(C5), and N3'(DOX)-O2(T4), respectively. So, the lower values of HB occupancy in our MD study may be due to the lower

cutoff distance and angle that are used (the distance cut off less than 3.5 Å and the angle cut off less than 25°).

Similarly, in M4, the N3' of amino sugar form one HB with O4'(C5) with 4% of HB occupancy and another one with O2(C5) has less HB occupancy of 2%, and there are no HBs formed between DOX and adenine or thymine residue. The different behavior of HBs between N3' atom of amino sugar with both atoms in cytosine residue (e.g., O2 and O4') in M4 is contradicted with X-ray structural analysis by Frederick et al. [34] but agrees with Wang et al. [35] and Nunn et al. [162]. The reason behind this difference of these HBs in our study with Frederick's study may be explained by the higher flexibility of the amino sugar group of DOX at certain steps during MD simulation. However, Frederick's study is based on the static position of the DOX-DNA complex to predict the HBs. Hence, we can conclude that the amino sugar group that lies in the minor groove forms a relatively tighter complex with DNA1 than with DNA2. For the 2:1 complex, the same behavior of HBs is seen, but they occur at both intercalation sites.

3.4.3 Principal Component Analysis (PCA)

Principal component analysis (PCA) was performed with R software using the bio3D package [171]. To investigate and compare the functional motions of DNA-free and DNA-bound complex, PCA is applied to the heavy atoms (all atoms except H atoms) of the DNA and DOX. The trajectory snapshots were extracted from MD production runs for the whole 30 ns, so we have 15000 snapshots for each model. The total number of heavy atoms included in PCA analysis is various in three different situations of each DNA sequence (free DNA, 1:1 complex, and 2:1 complex). A total of 240 atoms were included in the analysis of free DNA in

both sequences, resulting in 720 PCs. In the 1:1 complex case for both sequences, a total of 279 atoms were adopted in the analysis, resulting in 837 PCs. While in the 2:1 complex case, 318 atoms were included in the analysis, resulting in 954 PCs. The proportion of variance against its eigenvalue rank which represents the eigenvalue spectrum is shown in **Figure S1 in Appendix** for three different environments for both DNA sequences. For free DNA models, the first three components together make up 49.2% and 47.9% of the variance for M1 and M2 respectively. The top 20 PCs account for 89% of the total variance accumulated during the trajectories. For 1:1 complex models, the first three principal components contribute to making up 44.1% and 50% of the variance for M3 and M4 respectively. The top 20 PCs could capture 83.2% and 89.1% of total variance during the MD trajectories of M3 and M4 respectively. Similarly, the first three PCs contributions into the 2:1 complex make up 34.9% and 46.7% for M5 and M6 respectively and the top 20 PCs account for 77.6% and 83.7% of total variance during the MD trajectories.

The MD trajectories of the three different situations (free DNA, 1:1, and 2:1 complexes) corresponding to the same DNA sequence were then projected onto their respective first, second, and third principal component modes as shown in **Figure 3.5**. It shows that the 2:1 complex in both DNA sequences has smaller fluctuations than free DNA and the 1:1 complex. By comparing the PCs of DNA1 and DNA2 sequences corresponding to each one of the three cases, the model with the DNA1 sequence has lower fluctuations during MD. There are highly overlapping between three situations corresponding to the DNA1 and DNA2 sequence as can be seen in **Figure 3.5(a) and (b)**, respectively.

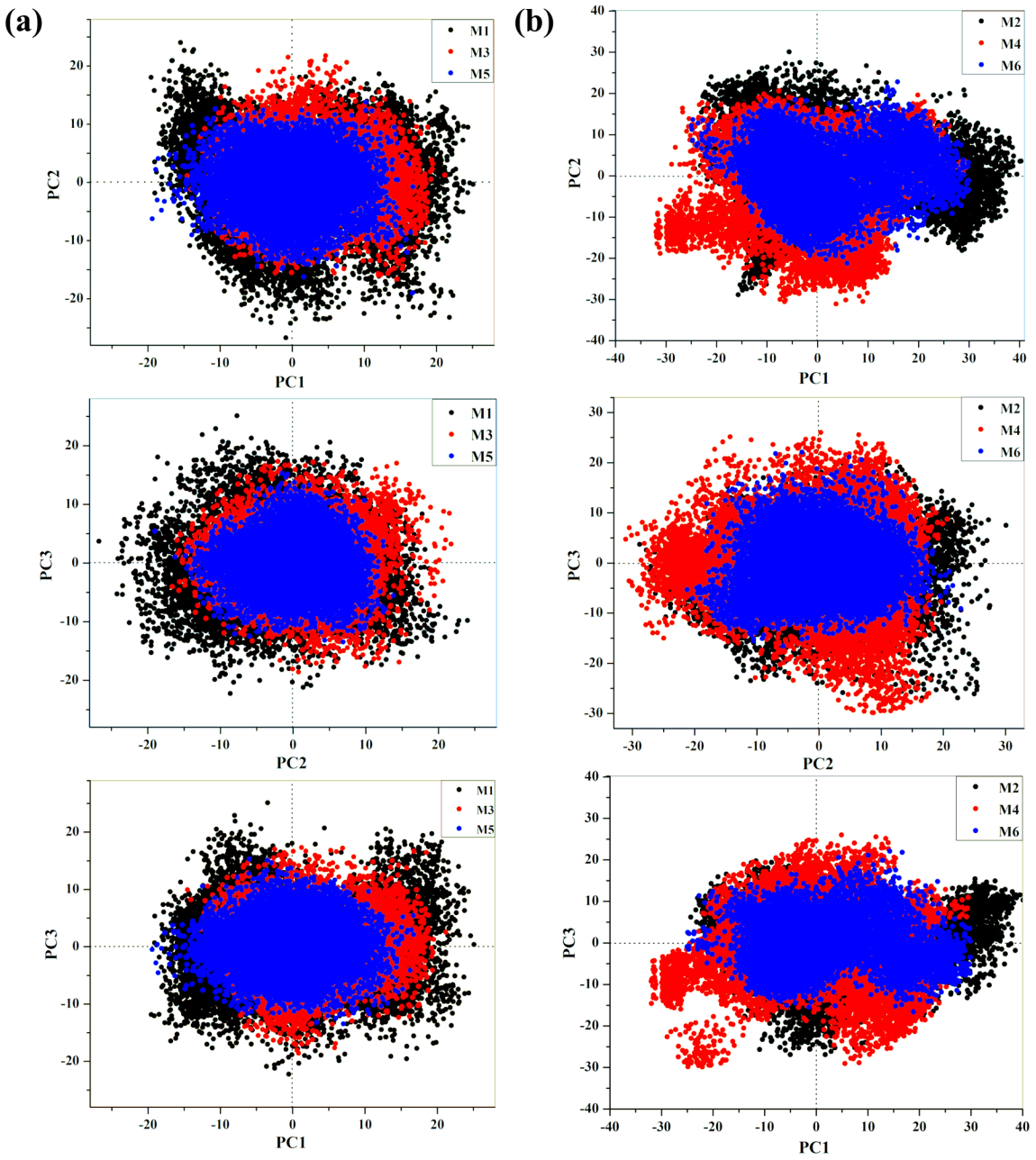


Figure 3.5. Projection of MD simulations with three different environments (free DNA, 1:1 complex and 2:1 complex which are represents by black, red, and blue circles, respectively) onto the corresponding first, second and third PC modes from the principal component analysis (PCA) of (a) the three situations of DNA1 sequence (M1 is free DNA, M3 is 1:1 complex and M5 is 2:1 complex), and (b) the three situations of DNA2 sequence (M2 is free DNA, M4 is 1:1 complex and M6 is 2:1 complex).

For more investigations, the contribution of each residue to the first three PCs is analyzed for three situations for each DNA sequence as shown in **Figure 3.6**. It shows that the DNA terminal base pairs are more flexible in both DNA sequences. The contributions of the DNA terminal base pairs in M1 and M2 of the free DNA along the three PCs are greater than those in the 1:1 complex or the 2:1 complex. This observation could explain why free DNA models have higher PC distributions than intercalated complexes (**Figure 3.5**). However, there is one exception to the 1:1 complex of M4, particularly through the second and third PCs (**Figure 3.6(b)**) which could explain the higher fluctuation of M4 (**Figure 3.5(b)**).

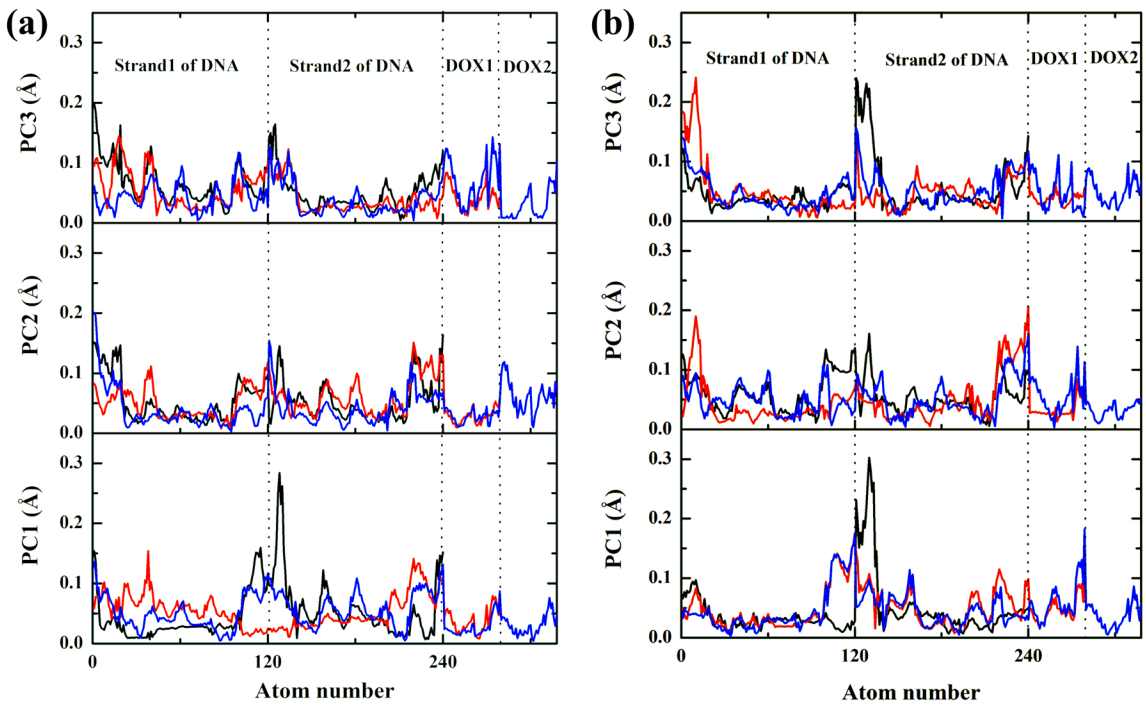


Figure 3.6. Residue-wise loadings for the first three principal components of three different environments (free DNA which represents by black line, 1:1 complex which represents by red line, and 2:1 complex which represents by blue line) for (a) the three situations of DNA1 sequence (M1 is free DNA, M3 is 1:1 complex and M5 is 2:1 complex), (b) the three situations of DNA2 sequence (M2 is free DNA, M4 is 1:1 complex and M6 is 2:1 complex). The dashed lines are used to separate DNA's and DOX's atom numbers.

3.4.4 Energetic Penalty for DNA Deformation (First Step)

The DNA deformation energy is indispensable in the calculation of BFE for the intercalation process. However, the MM-PB(GB)SA methods do not explicitly take the DNA deformation energy into account when calculating the BFE of DOX-DNA intercalation, as indicated in Eq.(16). So, the MM-PB(GB)SA methods have been manipulated to estimate $\Delta G_{Conf(DNA,sol)}$ as expressed in Eq.(22). Since the $\Delta G_{Conf(DOX,sol)}$ is relatively small, it is not considered in this chapter. Briefly, the $\Delta G_{Conf(DNA,sol)}$ can be computed as the energy difference between the intercalated DNA (bound state) and B-form DNA (unbound state). The energy of the bound DNA is calculated from DOX-DNA complex models M3 to M6 using STP since we can extract the ensembles from a single MD trajectory. The energy of an unbound DNA is extracted from reference models M1 and M2 using two trajectories protocol (2TP) of MM-PB(GB)SA methods. For this, we used 2TP, one MD simulation for the complex and another for the free DNA.

The calculated DNA deformation energies for all intercalated DOX-DNA complexes are listed in **Table 3.2**. It shows that the 1:1 complex has energy costs of about 18 kcal/mol but the 2:1 complex has an energy cost of about 28 kcal/mol (i.e., 14 kcal/mol per DOX). This implies that the deformation energy of DNA is less when the second drug (DOX2) attaches to the first one (DOX1) in the 2:1 complex, indicating that there is a cooperative effect. The penalty of DNA deformation in DNA1 is less than DNA2 in both 1:1 and 2:1 complexes. Thus, the binding of DOX to DNA1 could be more favorable than that of DNA2.

Table 3.2. DNA deformation energy that calculated as the difference between the energy of bound DNA that extracted from complexed models (M3 to M6) without other species (DOX or complex) by single trajectory MD (STP) subtracting from energy of unbound free B-DNA that extracted from multiple trajectory MD (2TP) of reference models (M1 and M2).

Difference	Method	DNA Deformation Energy (kcal/mol) at different ionic concentrations (M)					
		0.1	0.15	0.2	0.25	0.5	1
M3-M1	MM-GBSA	17.42	17.44	17.46	17.47	17.56	17.62
M3-M1	MM-PBSA	17.46	17.49	17.51	17.54	17.64	17.72
M4-M2	MM-GBSA	18.53	18.56	18.59	18.61	18.68	18.76
M4-M2	MM-PBSA	18.64	18.68	18.73	18.75	18.80	18.87
M5-M1	MM-GBSA	26.94	26.98	27.01	27.06	27.14	27.26
M5-M1	MM-PBSA	27.05	27.07	27.09	27.13	27.20	27.30
M6-M2	MM-GBSA	28.51	28.55	28.60	28.65	28.73	28.79
M6-M2	MM-PBSA	28.60	28.64	28.67	28.70	28.76	28.83

Our study also shows that the DNA deformation energy in forming an intercalation site for DOX is only slightly affected by the salt concentration in the 1:1 and 2:1 complex. The calculated deformation energy lies within the range of previous studies [217]–[220], but is slightly higher than those of DOX or DNR (10.5 kcal/mol) [31], [32]. We believe that our calculated values are more accurate since we considered explicitly the solvation and entropic effects. Moreover, they provide the overall changes in energy due to conformational changes in DNA, not just the energy necessary for the unstacking of the base pairs at the intercalation site. This analysis shows that the DNA deformation energy is only slightly dependent on the

sequence of DNA. The precise dependence is important to attain minimal DNA deformation energy for higher binding affinity.

Table 3.3 shows the important components that are contributed to $\Delta G_{Conf(DOX,sol)}$, to gain further insight into the origin of the driving force that forms the intercalation site. The intramolecular electrostatic energy is favorably contributing to the DNA deformation. When DOX is intercalated into base pairs of DNA, it increases the average spacing between the charges, leading to the reduction of electrostatic stress between the phosphate groups. Furthermore, it seems that the solvation energy has larger unfavorable interaction.

Table 3.3. Difference energy components that contribute into DNA deformation energy at 0.2 M theoretical ionic concentration using MM-GBSA method as an example.

Difference	ΔE_{int}	ΔE_{vdW}	ΔE_{ele}	$\Delta G_{p(GB)}$	ΔG_{sol}	$T\Delta S$	$\Delta G_{Conf(DNA,sol)}$
M3-M1	2.69	16.99	-15.43	9.16	9.81	-3.44	17.46
M4-M2	2.73	18.37	-24.17	19.31	20.05	-1.61	18.59
M5-M1	8.25	34.36	-31.71	10.93	12.17	-3.93	27.01
M6-M2	8.27	36.28	-43.33	24.18	25.42	-1.97	28.59

On the other hand, the contribution from the intramolecular electrostatic energy can overcome the electrostatic polar energy, leading to the favorable interaction for the total electrostatic interaction. Thus, the unfavorable interaction of the solvation free energy stems only from the non-polar interaction. The large positive values of the solvation free energy can be attributed to two factors. One is related to the increase in the distance of separation when one or two DOX intercalate into DNA. This leads to the decrease in the total charge density of

DNA and consequently reduces the interaction between DNA and the surrounding aqueous environment. The other factor arises from the burial of the SASA due to the intercalation of one or two DOX, leading to the unfavorable contribution from the non-polar solvation energy in the range of 0.7 kcal/mol of the 1:1 complexes. This range is almost double in the 2:1 intercalated complex (1.25 kcal/mol). Despite the cationic nature of the DOX structure, the van der Waals energy and not electrostatics dominate the driving force to create binding of DOX to DNA. The van der Waals energy predominantly contributes to DNA deformation energy, which also agrees with a previous study from David et al. [220]. Similarly, the contribution from non-polar solvation energy and internal energies play the role of an unfavorable interaction for the DNA deformation energy. However, the vibrational entropy favorably contributes to the $\Delta G_{Conf(DNA,sol)}$ because the native B-DNA in M1 and M2 having more freedom for vibrational motion but they are more restricted when DOX is intercalated between them.

We have compared the deformation energy $\Delta G_{Conf(DNA,sol)}$ of M3 with M4 or M5 with M6 to verify the best DNA sequence with the least defamation penalty. The difference in penalty due to van der Waals interaction energy in DNA1 is more favorable in DNA2 by 1.38 kcal/mol (18.37-16.99) in 1:1 complex and 1.92 kcal/mol (36.24-34.36) in 2:1 complex (**Table 3.3**). The net electrostatic interactions ($\Delta G_{ele} = \Delta E_{ele} + \Delta G_p$) for M3 is -6.27 kcal/mol (-15.43+9.16). Similarly, for M4 these numbers is -4.86 kcal/mol. Therefore, the difference in total electrostatic interaction between M3 and M4 is -1.41 kcal/mol (-6.27+4.86). The preferred total electrostatic interactions in M3 over M4 play a significant role in reducing the higher repulsion between the amino sugar of DOX and the guanine amino group. The vibrational

entropic is also more favorable by -1.83 kcal/mol (-3.44+1.61) in M3 over M4. The same can be said for the preferred behavior of M5 over M6 for the 2:1 complex (-1.96 kcal/mol). Besides these energetic investigations on the preferred interactions in the first step of the intercalating process, the unwinding of DNA at the intercalation site, of up to 4° in DNA1 and 6° in DNA2 sequence, is another index that supports the dependence of DNA deformation energy on the sequence of DNA.

3.4.5 Binding Energy of the Insertion Step (Second Step)

As mentioned in chapter 2, the binding energy of the insertion step (ΔG_{Ins}) can be computed directly using the STP of the MM-PB(GB)SA method. To check the stability of calculated ΔG_{Ins}, the existing trajectory data extracted from the 30 ns MD simulations is divided into three equal segments with 10 ns of each one. These segments are defined as the first, second, and third 10 ns of each MD simulation. **Figure 3.7** shows the ΔG_{Ins} at 0.2 M salt concentration as a function of simulations time of different time segments for the 1:1 complex, along with their histogram distributions. The left-hand side of **Figure 3.7(a)** shows that the ΔG_{Ins} for M3 has achieved stability. As dynamic simulation proceeds, the fluctuation of the ΔG_{Ins} decreases around the mean value of ΔG_{Ins} with a total standard deviation (SD) of 4.39 kcal/mol and a total standard error of the mean (SEM) of 0.113 kcal/mol. The right-hand side of **Figure 3.7(a)** shows that around 95% of the histogram distributions of ΔG_{Ins} is confined between -35 to -25 kcal/mol. The same behavior of ΔG_{Ins} in M4 can be noticed from **Figure 3.7(b)** with a slight difference at the beginning of the dynamic simulation. ΔG_{Ins} takes approximately 1 ns to stabilize and become less fluctuation around the mean value with SD of 4.55 kcal/mol and SEM of 0.117 kcal/mol.

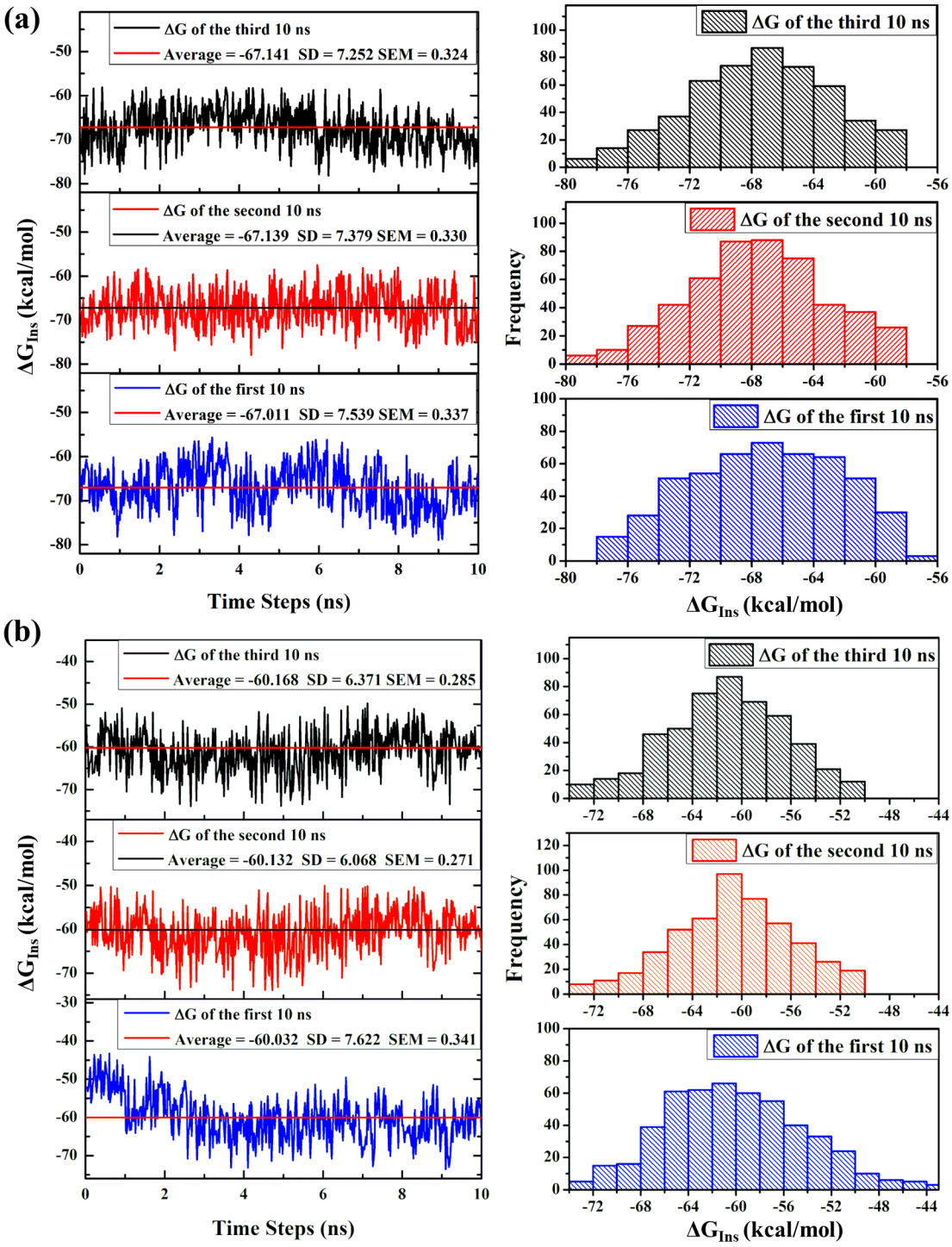


Figure 3.7. The insertion binding free energy against the first, second and third 10 ns segments of whole 30 ns time simulation and histogram distribution of ΔG_{Ins} in 1:1 complex at 0.2 M salt concentration using MM-GBSA method for (a) M3 and (b) M4.

The histogram distributions of **Figure 3.7(b)** show that about 94% of ΔG_{Ins} are confined between -33 to -21 kcal/mol. The same trend can be observed for the 2:1 complex (**Figure 3.8**).

The different energy components that contribute to ΔG_{Ins} for all intercalated models (M3 to M6) have been calculated and tabulated in **Table 3.4** and displayed as a histogram in **Figure 3.9**. In general, the total contribution to ΔE_{int} is nearly zero due to the use of STP of the MM-PB(GB)SA methods. In the gas phase, ΔE_{vdW} and ΔE_{ele} are both favorably contributed to form stable DOX-DNA complexes in the insertion stage, where ΔE_{ele} predominates interaction. Such a highly favored ΔE_{ele} can be explained by the strong electrostatic attractions between the negatively charged phosphate backbone of DNA and the positively charged amino group of DOX. Thus, strong interaction at the insertion step between DOX and dsDNA is electrostatic in origin. However, ΔE_{ele} is completely overshadowed by ΔG_p, leading to an overall unfavorable contribution of the total ΔG_{ele}. Upon the insertion occurred, the solvent-solvent and solvent-solute interactions are reorganized as a result of burying the polar or charged atoms of DOX within dsDNA and thus hidden away from the bathing solution, leading to a largely unfavorable ΔG_p and favorable ΔG_{np}. Another reason for favorable ΔG_{np} is that the complex has less solvent-exposed molecular surface than its unbound species (free DOX and DNA). Finally, the total entropic contribution ($-T\Delta S$) at 310 K is unfavorably significant, stemming from DOX being trapped with little mobility when inserted between BPs of dsDNA.

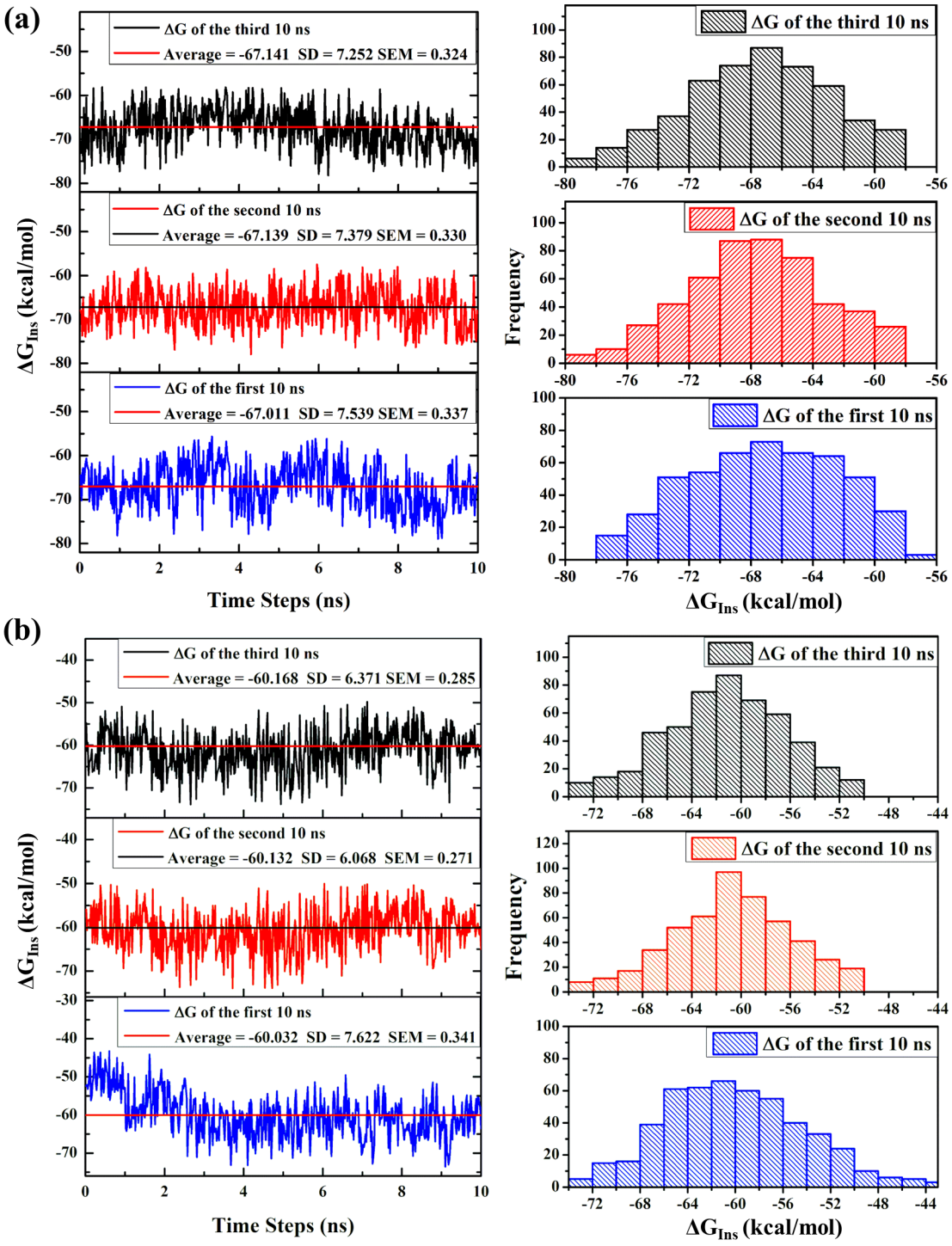


Figure 3.8. The insertion binding free energy against the first, second and third 10 ns segments of whole 30 ns time simulation and histogram distribution of ΔG_{Ins} in 2:1 complex at 0.2 M salt concentration using MM-GBSA method for: (a) M5 and (b) M6.

Table 3.4. Energy contributions to insertion binding of DOX-DNA complex models at 0.2 M (kcal/mol). The last two columns is the differences in energy between corresponding models.

Energy	ΔG of M3	ΔG of M4	ΔG of M5	ΔG of M6	$\Delta\Delta G_{(M3-M4)}$	$\Delta\Delta G_{(M5-M6)}$
ΔE_{int}	0.23	0.21	0.47	0.38	0.02	0.09
ΔE_{vdW}	-64.15	-61.09	-128.69	-122.89	-3.06	-5.81
ΔE_{ele}	-345.16	-327.32	-691.76	-683.66	-17.84	-8.10
ΔE_{MM}	-409.08	-388.21	-819.99	-806.17	-20.87	-13.82
$\Delta G_{p(GB)}$	362.30	344.78	733.71	727.00	17.52	6.71
$\Delta G_{np(GB)}$	-2.86	-2.75	-5.64	-5.59	-0.11	-0.04
$\Delta G_{sol(GB)}$	359.44	342.04	728.07	721.40	17.41	6.67
$\Delta G_{ele(GB)}$	17.14	17.46	41.95	43.33	-0.32	-1.38
$\Delta G_{p(PB)}$	361.46	344.12	730.62	724.48	17.34	6.14
$\Delta G_{np(PB)}$	-5.70	-5.43	-10.39	-10.26	-0.26	-0.13
$\Delta G_{sol(PB)}$	355.76	338.68	720.23	714.22	17.08	6.01
$\Delta G_{ele(PB)}$	16.30	16.79	38.86	40.82	-0.5	-1.96
$T\Delta S$	-19.65	-19.06	-24.82	-24.65	-0.59	-0.17
$\Delta G_{Ins(GB)}$	-29.99	-27.11	-67.09	-60.11	-2.87	-6.98
$\Delta G_{Ins(PB)}$	-33.67	-30.47	-74.93	-67.30	-3.20	-7.64

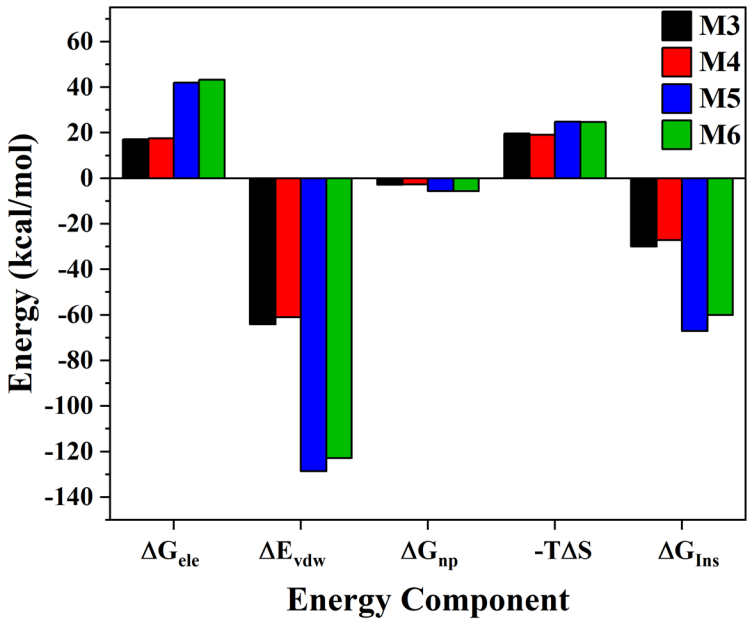


Figure 3.9. Histogram display of the calculated different energy components of insertion binding energy (ΔG_{Ins}) of DOX-DNA intercalated complexes for M3, M4, M5, and M6 at 0.2 M theoretical salt concentration using MM-GBSA method. Note $\Delta G_{ele} = \Delta E_{ele} + \Delta G_p$.

We can conclude from the above analyses that the long-range Coulombic electrostatic interaction (ΔE_{ele}) is responsible for initiating the insertion process between the DOX and DNA, whereas the attractive component of the van der Waals interaction (ΔE_{vdW}) is the major driving force for stabilizing the complex.

We have also compared the difference of ΔG_{Ins} and their energy component with respect to DNA1 and DNA2 in the 1:1 complex (denoted by $\Delta\Delta G_{Ins}$ in **Table 3.4**). In the 1:1 complex, the difference between DOX-DNA1 and DOX-DNA2 in $\Delta\Delta E_{ele}$ is -17.84 kcal/mole. The difference in electrostatic polar energy with PB and GB methods ($\Delta\Delta G_{p(PB/GB)}$) is equal to 17.34 and 17.52 kcal/mole respectively. Hence, the $\Delta\Delta G_{ele}$ is equal to -0.5 and -0.32 kcal/mole, respectively. Furthermore, the $\Delta\Delta E_{vdW}$ between them is -3.06 kcal/mole, and the

$\Delta\Delta G_{np}$ is -0.26 and -0.11 kcal/mole from PBSA and GBSA, respectively. Because both models contain the same drug molecule, there are no differences in translational and rotational entropies (M3 and M4), but there are minor differences in vibrational entropy of -0.6 kcal/mole. Finally, based on these comparisons for the DOX insertion into two different sequences of DNA, we conclude that the $\Delta\Delta E_{vdW}$ of -3.06 kcal/mole plays a crucial role in the preferential sequence of DOX with DNA1 rather than DNA2, which agrees with the experimental studies [34], [162].

In brief, the same conclusions and behaviors for the different energy components can be drawn in the 2:1 complex. On the other hand, by comparing the 2:1 vs 1:1 complex, the binding of DOX to DNA entails a considerably greater entropic penalty for the 1:1 complex (19 kcal/mol) than the 2:1 complex (24 kcal/mol), a reduction of 12 kcal/mol for each DOX. This reveals that the inserting of the first DOX within dsDNA has a relatively greater entropic cost than the inserting of the second DOX.

3.4.6 BFE Corrections due the Effects of DOX and Ionic Concentrations

The first correction of the BFE using MM-PB(GB)SA methods originates from the standard free energy of the DOX concentration effect as given by Eq.(24). Here, it should mention that we used the symbol ΔG_{con} for the standard concentration dependent free energy of DOX rather than ΔG^0_{AB}. We estimate ΔG_{con} to be -2.31 kcal/mole at T=310 K in both 1:1 and 2:1 complexes. This value should be subtracted from our calculated BFE using MM-PB(GB)SA methods to compare with the experiment [221]. **Table 3.5** lists the BFEs of the DOX intercalation process for all models at different salt concentrations, including ΔG_{con}.

Table 3.5. The calculated BFE of DOX-DNA complexes at different salt concentrations.

Model	Method	ΔG_{calc} (kcal/mol) at different salt concentrations (M)					
		0.1	0.15	0.2	0.25	0.5	1
M3	MM-GBSA	-10.72	-10.45	-10.22	-10.08	-9.60	-9.10
	MM-PBSA	-14.35	-14.08	-13.85	-13.66	-13.23	-12.74
M4	MM-GBSA	-6.69	-6.42	-6.21	-6.05	-5.58	-5.10
	MM-PBSA	-9.94	-9.66	-9.43	-9.29	-8.83	-8.35
M5	MM-GBSA	-38.75	-38.21	-37.78	-37.51	-36.56	-35.61
	MM-PBSA	-46.52	-45.96	-45.54	-45.26	-44.34	-43.40
M6	MM-GBSA	-30.16	-29.62	-29.21	-28.87	-27.95	-27.03
	MM-PBSA	-37.29	-36.72	-36.32	-36.01	-35.05	-34.15

The second correction of BFE arises from the dependence of the intercalation BFE on the ionic concentration of the solution. Since ΔG_{Ins} and $\Delta G_{Conf(DNA,sol)}$ have been computed over a wide range of ionic concentrations from 0.1 to 1 M, it can plot the ΔG_{bind} against these theoretical concentrations as shown in **Figure 3.10** to assess the ionic effects on the BFE of DOX-DNA complexes as well as to quantify the polyelectrolyte free energy (ΔG_{pe}) and the non-polyelectrolyte free energy (ΔG_t). **Figure 3.10** shows that the values of ΔG_{bind} becomes less exothermic with increasing ionic concentration in both methods, which is consistent with the relevant experimental findings [163], [222].

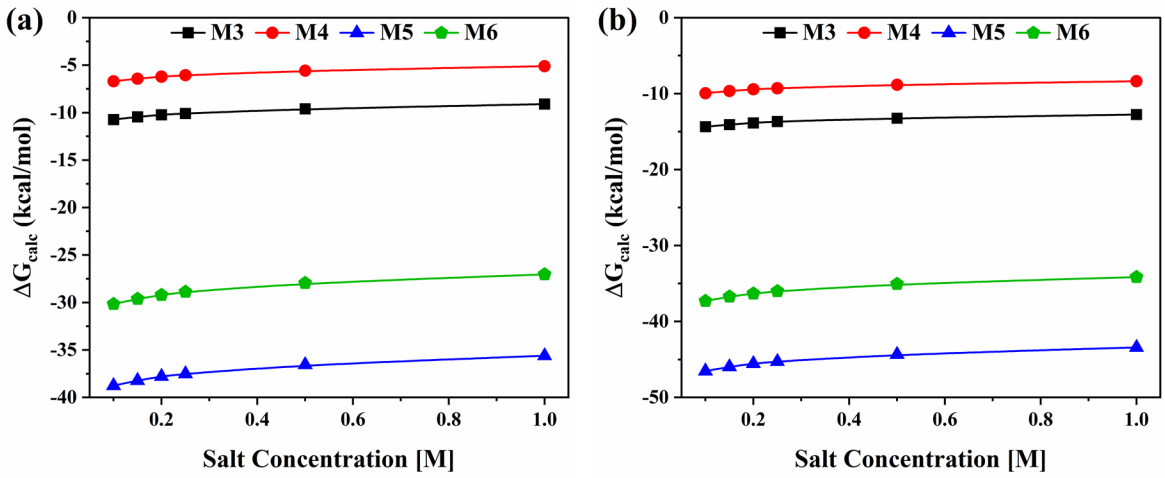


Figure 3.10. The impact of ionic concentrations on the binding free energy of DOX-DNA complexes using modified methods of (a) MM-GBSA and (b) MM-PBSA.

ΔG_{pe} can be computed in two ways: first as the difference between the BFE at a specific ionic concentration and the BFE at 1 M (BFE at 1 M is ΔG_t^{sim}); second by applying Eq.(26) after determining the salt dependence of the binding constants (SK). By using the first approach, the ΔG_{bind}, for example in M3, is -10.72 and -9.1 kcal/mol at 0.1 M and 1 M salt concentrations, respectively, from the MM-GBSA method, leading to ΔG_{pe} is equal to -1.62 kcal/mol, which is close to the experimental value from Chaires et al. (ΔG_{pe} = -2.3 kcal/mol at 0.016 M and T = 293.15 K) [18]. The difference between our calculated of ΔG_{pe} and the experimental result (-2.3 kcal/mol) arises mainly from different solution conditions and differences in the number of base pairs per site (3.4 in the experiment [165] and 3 in our work). We will see later in Chapter 4 that the number of base pairs per site should be greater than three. The calculated value of ΔG_{pe} at 0.2 M is -1.12 kcal/mol, which is quite close to the experimental value of -1 kcal/mol at the same salt concentration and T = 293.15 K [165]. Interestingly, the same results of ΔG_{pe} are obtained using the MM-PBSA method. This ΔG_{pe}

trend is also observed in other intercalated DOX-DNA complexes using MM-PB(GB)SA methods.

Alternatively, the computed slope SK from the binding constant (K_{sim}), with corresponding values of ΔG_{pe} at 0.1, 0.15, 0.2, 0.25, and 0.5 M and T = 310 K, are listed in **Table 3.6** for all models with both MM-GBSA and MM-PBSA methods. SK is about -1.12 for the 1:1 complex, which agrees quite well with the theoretical estimate of -1.24 on Ethidium-DNA complex [223]. The experimental value of SK of DOX is -0.97 [18]. Wilson and Lopp suggested that there are approximately 0.2-0.3 sodium ions released in the intercalation conformation change, regardless of the intercalator charge [224].

Table 3.6. The salt dependence of the binding constants (SK) of DOX-DNA complexes with the corresponding ΔG_{pe} at different theoretical ionic concentration and 310 K temperature.

Model	Method	SK	ΔG_{pe} (kcal/mol) at different salt concentrations (M)				
			0.1	0.15	0.2	0.25	0.5
M3	MM-GBSA	-1.140	-1.62	-1.33	-1.13	-0.98	-0.49
	MM-PBSA	-1.133	-1.61	-1.33	-1.13	-0.97	-0.48
M4	MM-GBSA	-1.119	-1.59	-1.31	-1.11	-0.96	-0.48
	MM-PBSA	-1.119	-1.59	-1.31	-1.11	-0.96	-0.48
M5	MM-GBSA	-2.210	-3.14	-2.59	-2.19	-1.89	-0.95
	MM-PBSA	-2.196	-3.12	-2.57	-2.18	-1.88	-0.94
M6	MM-GBSA	-2.203	-3.13	-2.58	-2.19	-1.88	-0.94
	MM-PBSA	-2.210	-3.14	-2.59	-2.19	-1.89	-0.95

Thus, our value of SK for the 1:1 complex of -1.12 should be lower than the experimental value. Similarly, the SK for 2:1 complexes is -2.2, which is double the value of 1:1 complexes. This value also agrees well with other theoretical studies on similar types of drugs (quinacrine diction) [224]. The values of slope and corresponding values of ΔG_{pe} are independent of the sequence of DNA in the intercalating complexes. The decrease in ΔG_{pe} favorability upon increasing the ionic concentrations is due to the electrostatic repulsion between the positively charged amino sugar of the DOX-DNA complexes and the Na^+ in the solvent medium. When DOX binds to DNA at low ionic concentrations, the positive charge of DOX causes the release of one bound ion, since the positively charged amino sugar in the DOX may be used to neutralize the DNA rather than the Na^+ ion. There is a thermodynamic linkage between the ions in the solvent and the charged drug. The binding of an ion will influence the binding of the drug and vice versa [165].

The formation of DOX-DNA intercalating complexes is predicted to occur at lower solution conditions due to the favorable ΔG_{pe}, which agrees with experimental studies [163], [222], [225]. Interestingly for all DOX-DNA complexes shown in **Figure 3.10**, the calculated values of ΔG_{bind} at different concentrations are all negative. This is consistent with the idea that these DOX-DNA complexes are stabilized predominantly by hydrogen bonds and van der Waals interactions. This is even more pronounced at 1 M ionic concentration since ΔG_{pe} is zero, while the BFEs have negative values for all models. This finding supports our conclusion that van der Waals interactions and the HBs are the primary driving forces in the formation of a stable intercalated complex.

3.4.7 Total BFE of the Intercalation Complex

The true total BFE (ΔG_t^{sim}) represents the portion of the binding free energy that is independent of ionic concentrations. Theoretically, ΔG_t^{sim} has the same value of ΔG_{bind} at 1 M ionic concentration because ΔG_{pe} is zero at 1 M. Alternatively, it can be computed by applying Eq.(25), $\Delta G_t^{sim} = \Delta G_{bind} - \Delta G_{pe}$.

ΔG_t^{sim} for M3 are -12.74 and -9.1 kcal/mol from modified MM-PBSA and MM-GBSA methods, respectively, might be in reasonable agreement with the experimental value of -7.7 ± 0.3 kcal/mol [18]. Our predicted value of M3 (-9.1 kcal/mol) using MM-GBSA is demonstrated to be closer to the experimental value (-7.7 kcal/mol) than that computed using the HINT program of DOX-d(CGAT)$_2$ complex (-9.6 kcal/mol) [28]. For M4, the ΔG_t^{sim} of -5.1 kcal/mol obtained with MM-GBSA varies markedly from the experimental value of -7.7 kcal/mol, however, the ΔG_t^{sim} of -8.35 kcal/mol using MM-PBSA may be in a good match with experimental values [18], [226]. Interestingly, our value of -5.1 kcal/mol using MM-GBSA is in excellent agreement with the recent experimental value of -4.99 kcal/mol [190], [198]. For M5, the ΔG_t^{sim} are -43.4 and -35.61 kcal/mol from MM-PBSA and MM-GBSA methods respectively. Similarly, they are -34.15 and -27.03 kcal/mol for M6. ΔG_t^{sim} in the 2:1 complex is more than twice the corresponding values from the 1:1 complex, implying that the binding of two DOXs to DNA is more favorable than one DOX.

Finally, the calculated value of the entropic contribution of intercalated DOX is experimentally subject to some debate as mentioned in Chaires's studies [41], [157]. In this study, the total entropic contribution is explained in a more detailed manner especially for the 1:1 complex. We aim to obtain an accurate value for total entropy to provide the missing part

for estimating the entropic contribution by experimentalists. From **Table 3.4**, the calculated values of the entropic contributions in M3 and M4 are 19.65 and 19.06 kcal/mol respectively. There are also favorable entropic contributions in forming the intercalation site of DOX at the base pair of DNA with deformation energy of -3.44 and -1.61 kcal/mol, respectively, as shown in **Table 3.3**. Therefore, the total entropic contribution for forming 1:1 complex is 16.8 ± 0.6 kcal/mol, in excellent agreement with the experimental value of 15.5 kcal/mol [41], [157], [227].

3.5 Summary and Conclusion

Extensive and careful investigation of the key parameters that are vital in forming a stable DOX-DNA complex is important for practical and rational drug design. There is also a fundamental interest in predicting accurately the binding free energy including the role of each of the solution components. The extensive MD simulations and an accurate evaluation of the total binding free energy reported here provide a deep molecular-level understanding of the intercalated DOX-DNA complexes. Our study aims also to address the missing part of the estimation of both entropic and DNA deformation energy contributions, faced by experimentalists. We have investigated different energy components that contributed to the TBFE of DOX-DNA binding and our study provides the following solid conclusions:

(i) There are at least three unfavorable interactions for DOX binding to DNA. They are the deformation energy, the energy cost stemming from the translational and rotational entropic contributions, and the total electrostatic interactions of the sum of the intramolecular electrostatic and electrostatic polar solvation interactions.

(ii) We quantified five favorable interactions. They are van der Waals interactions energy, the non-polar solvation free energy, the vibrational entropic contributions, the standard free energy related to DOX solution concentration (ΔG_{con}), and the standard free energy due to ionic concentration (ΔG_{pe}). These favorable interactions overcome all unfavorable interactions.

(iii) The key factor of driving force for intercalated DOX into DNA interaction is the van der Waals interactions.

(iv) DOX prefers to bind with DNA1 more than the DNA2 sequence. This difference is due to three main factors: (1) favorable van der Waals interactions, (2) smaller DNA deformation energy penalty of DOX, and (3) more HBs between the N3' atom of the amino sugar group of DOX and the DNA residues.

(v) Based on the values of DNA deformation penalty and entropy cost, as well as the free energy, the formation of the 2:1 complex, is more favorable and stable than the 1:1 complex.

(vi) The final calculated values of the true BFE, ΔG_t, agree reasonably with experimental values, if the details of the solution conditions, number of base pairs per site, the change of DNA sequence, and the number of DOX are all considered concurrently.

(vii) The binding of DOX to DNA is preferred at lower ionic concentrations.

CHAPTER 4. THERMODYNAMIC DISSECTION AND SEQUENCE SELECTIVITY OF DOXORUBICIN-DNA INTERACTION

4.1 Introduction

DNAs perform their biological function by interacting with other molecules. In the readout of genetic information and the regulation of gene expression, their interactions with proteins or small ligands are critical [228]. On the other hand, the binding of the drug to DNA can induce various biological consequences such as transcription and replication inhibition which are the necessary preconditions for cell division [229], [230]. Particularly, DNA is the main intracellular target for many clinical cancer drugs [5], [6]. Thus, a complete understanding of the nature of the drug-DNA interaction is key to provide insights knowledge into new therapeutic discoveries, identify new DNA recognition motifs, and explore the molecular mechanisms of achieving cellular selectivity.

DNA sequence selectivity plays an important role in a wide range of molecular recognition processes including protein-DNA [231], [232], DNA-DNA [233], and ligand-DNA [230], [234] interactions. Since many drugs bind to DNA in a sequence-dependent fashion [235], the identification and characterization of preferential drug binding sites within DNA sequences are crucial for a deep understanding of the molecular basis of drug action and for designing better drugs with desirable features [29]. Therefore, intensive efforts have been put into finding more selective drugs, and there is considerable excitement that the identification of cancer-specific molecular targets will yield a new generation of less toxic therapeutic [6]. However, the design of sequence-specific DNA-binding drugs is still a highly challenging task for the following general reasons. First, the methods for precisely determining

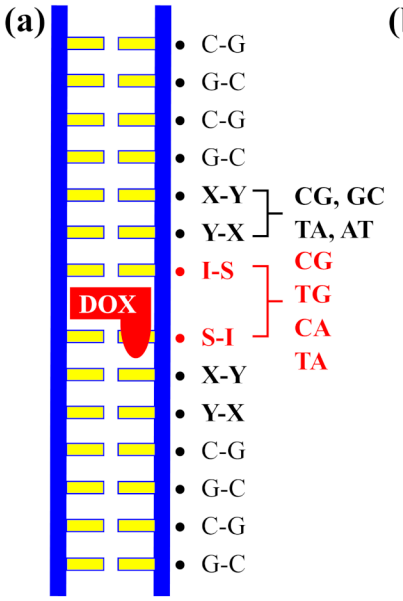
selective drugs are limited [235], [236]. Second, the drug does not have a uniform binding mode because it can bind DNA noncovalently via two common binding modes: grooving binding and intercalation. Third, the major driving forces that play a significant role in the sequence specificity are unique for each drug and are still being argued [237]–[239]. These limitations motivate us to utilize various methodologies based on computational methods in investigating the specificity of anthracycline intercalator drugs such as DOX to provide valuable information on the nature of this preference with high accuracy at the molecular, sequence, and atomic levels.

Intensive experimental and theoretical studies have been conducted to identify the DNA sequence specificity of DOX [27]–[40]. Experimentally, the binding affinity of anthracycline anticancer drugs toward alternating pyrimidine-purine sequence seems to be stronger than non-alternating sequence [31]–[34], [240], and a lower binding predilection toward the G-C base pair over the A-T base pair [241]. While identifying the exact sequence-dependent of DOX remains an open question as mentioned earlier. Additionally, the experimental studies alone are unable to determine how DOX recognizes its specific site in the BP sequence of DNA. On the other hand, theoretical and computational studies for the DNA sequence selectivity of DOX binding are still very limited [28], [31]–[33], [39], [40].

There are numerous debates and unanswered questions regarding the identification of the sequence-specific of DOX as mentioned in chapter 1. For example, does DOX have sequence-specificity, and if so, what is it? What is the best model and DOX site size to describe this specificity? and so on.

The aim of the present work is much more ambitious. To address the specific issues raised above in more detail, we undertook a comprehensive MD study focusing on the understanding of the molecular mechanism and sequence-specificity implied in the DOX intercalation process. The first goal of our endeavor is to identify the specific sequence of DOX and to probe the physical factors that determine its preferential binding and control the intercalation process of the DOX-DNA complex at each of the binding stages. This includes physical variables such as the energetic components of bonded and non-bonded interactions, the conformational energetic cost to form the intercalation cavity, the DOX insertion free energy, the solvation free energy change, the non-specific polyelectrolyte contribution to DOX-DNA binding, as well as the solute entropic contributions to the free energy change. The second goal is to investigate the best model reflecting the relationship between the DNA sequence and the preferential intercalation of DOX into DNA. The third objective is to examine the impact of surrounding nearest neighbor BPs, vicinal to the intercalation site, and to clarify the biological role of the DOX's daunosamine sugar group in the binding process. The last goal is to probe the impact of the solvent and ionic strength effects on the complexation mechanism. Our investigation is based on accurately computing the BFE (ΔG) and the related relative binding energy ($\Delta\Delta G$) for 16 different DOX-DNA complexes using MD simulations at the microsecond time scale, with about 10 microseconds length in total. All MD simulations are implemented in the AMBER package within the context of an explicit solvent TIP3P water model. Here, we used only the MM-GBSA method to predict the BFE of binding DOX with 16 unique tetradecamer sequences of dsDNA. For each of the dsDNA unique sequences, we built two different models, the free standard B-DNA model (without DOX) and the intercalated

DOX-DNA complex model with one configuration of DOX (1:1 DOX-DNA complex). **Figure 4.1** shows the diagram of the DOX-DNA complex as well as the 16 proposed dsDNA sequences. For the analysis of polyelectrolyte effects, each calculation of the BFE involves ten different uni-valent NaCl salt concentrations from 0.1 to 1 M.



(b)

Model

#	Name	Group	DNA Sequence
1	A1	A	d(CGCG**GC**CG**GC**CGCG)$_2$
2	A2	A	d(CGCG**CG**CG**CG**CGCG)$_2$
3	A3	A	d(CGCG**TA**CG**TA**CGCG)$_2$
4	A4	A	d(CGCG**AT**CG**AT**CGCG)$_2$
5	B1	B	d(CGCG**GC**TG**GC**CGCG)$_2$
6	B2	B	d(CGCG**CG**TG**CG**CGCG)$_2$
7	B3	B	d(CGCG**TA**TG**TA**CGCG)$_2$
8	B4	B	d(CGCG**AT**TG**AT**CGCG)$_2$
9	C1	C	d(CGCG**GC**CA**GC**CGCG)$_2$
10	C2	C	d(CGCG**CG**CA**CG**CGCG)$_2$
11	C3	C	d(CGCG**TA**CA**TA**CGCG)$_2$
12	C4	C	d(CGCG**AT**CA**AT**CGCG)$_2$
13	D1	D	d(CGCG**GC**TA**GC**CGCG)$_2$
14	D2	D	d(CGCG**CG**TA**CG**CGCG)$_2$
15	D3	D	d(CGCG**TA**TA**TA**CGCG)$_2$
16	D4	D	d(CGCG**AT**TA**AT**CGCG)$_2$

Figure 4.1. Proposed models to understand the intercalation process and to identify DOX selectivity. (a) Schematic diagram of a DNA intercalator binding mode of DOX-DNA complex. A ladder represents the fragments of dsDNA helix and DOX molecule is shown in red. XY refers to the nearest BPs located directly to the intercalation site at both sides and can take $(GC)_2$ and its reverse $(CG)_2$, and $(TA)_2$ and $(AT)_2$. IS stands for the intercalation site of DOX between BP7 and BP8 and can take any BPs of our proposed groups: A $(CG)_2$, B $(TG)_2$, C $(CA)_2$, and D $(TA)_2$.; (b) The models of DOX-DNA complex with different DNA sequences.

4.2 Construction of Molecular Models

To identify the sequence specificity of DOX and to understand the intercalation process of the DOX-DNA complex at the molecular, sequence, and base pair levels, we explicitly targeted two different models for each unique sequence of dsDNA. There are altogether 16 models of free B-DNA (unbound or free regular B-DNA without DOX) and another 16 models

of the DOX-DNA complex with one configuration of DOX, (1:1 DOX-DNA complex). For all models, we chose the dsDNA tetradecamer sequences of the form d(CGCG**XY**I|S**XY**CGCG)$_2$, where **I|S** stands for the DOX intercalation site that can take any BP among four proposed groups: A **(C|G)$_2$**, B **(T|G)$_2$**, C **(C|A)$_2$**, and D **(T|A)$_2$**. The **XY** denotes the nearest neighbor BP located directly at both sides of the intercalation site and can be either **(GC)$_2$** and its reverse **(CG)$_2$**, or **(TA)$_2$** and its reverse **(AT)$_2$**. These 16 DOX-DNA complexes are divided into four different groups according to the DOX intercalation site (see **Figure 4.1(b)**). For each group, 4 different sequences are selected, the only difference between them being the BP located next to the intercalation site on both sides. The first and second sequences have **(GC)$_2$** and its reverse **(CG)$_2$** at the neighboring intercalation site, whereas the third and fourth sequences have **(TA)$_2$** and **(AT)$_2$**, respectively. The choice of four different groups enables us to examine the preferential DOX intercalation site, while the aim of selecting four different sequences in each group is to understand how the binding is influenced by the position of the daunosamine chain of DOX in the minor groove and to investigate the impact of the nearest neighbor BPs on the DOX binding. The alternating pyrimidine-purine sequences were chosen as intercalation sites since they are known to be energetically preferred [31]–[34].

Two important points should mention here. First, the complex models of DOX and long tetradecamer sequences of DNA accurately mimic the DOX environment, in which one DOX molecule is intercalated at the middle of this sequence (**Figure 4.1**). In these models, the effect of surrounding nearest neighbor PBs on DOX binding is considered. Second, these models are more realistic than the ones we created in chapter 3 based on X-ray structures where the DOX molecules are intercalated at the terminal BP steps of the DNA hexamer sequence. In X-ray

structures, there are end effects in which the end terminal BPs are known to fray and breath quite differently than internal base pairs as well as they could have peculiar electrostatic properties. Surely this would strongly influence the energy required to deform the DNA to make the interaction cavity and ignore the impact from the nearest neighbor PBs.

We develop the following procedure for building the models. First, the free B-DNA models with the desired sequence of 14 BPs are designed, based on the parameters adopted from the B-DNA fiber-diffraction studies by using the 'fiber' module of the 3DNA program [174]. As a result, we build 16 free B-DNA structures with different tetradecamer sequences. In this step, we also use the 'analyze' module of the 3DNA program to create the parameter file for each sequence of these B-DNA structures. This file contains information on the dsDNA helical geometry parameters, including six base pairs and six base pairs step parameters. This file will be used when generating intercalated DNA models in the next steps. Second, the DOX intercalation site is created by modifying the helical geometry parameters, at the local BPs that form this site at BP7 and BP8 in the current study, for the free B-DNA structure built in the first step. These modifying parameters of the intercalation BPs are obtained from the Protein Data Bank (PDB) with ID: 1D12 of the X-ray structure of the DOX-DNA complex [34]. It should be mentioned here that the parameters of all other BPs of B-DNA remain unchanged. Here, the parameter file that is generated in the first step is also modified, but this time it contains the intercalated BP parameters from PDB ID: 1D12 at the local site of DOX intercalation site. Third, the intercalated DNA structure with a DOX intercalation site is generated by using the modifying file parameter from the previous step. The 'rebuilding' module of the 3DNA program is utilized to build this intercalated DNA with the same 16

sequences that were created in the first step. Fourth, the resulting intercalated DNA structures are aligned according to DNA PDB ID: 1D12 using a structure comparison tool in the UCSF Chimera software [242]. We can easily place one DOX molecule at the same intercalation site of all the dsDNAs to form a 1:1 DOX-DNA complex. The DOX molecule is obtained from the crystal structure of 1D12. The first intercalated DNA structure that we created was for model A4 because it can easily align with a sequence of PDB ID: 1D12. Fifth, all structures (16 free B-DNAs and 16 DOX-DNA complexes) are solvated using the TIP3P explicit water model [92] implemented in AMBER, after adding H atoms using the LeaP module [61]. In this way 4000 water molecules are added to each model. The box of water was adjusted to be rectangular, with dimensions of 52× 52×73 $Å^3$, by using specific commands through the LeaP. Finally, since the dsDNA fragment has 14 BPs with a total charge of 26 e^-, 26 Na^+ ions are added as counter ions to neutralize the system. Since the net charge of the free DOX molecule is 1 e^+, a Cl^- ion is added to maintain the overall charge neutrality. These ions were placed by using the program LeaP. In total, we have built thirty-two models, 16 models for the 1:1 complex and 16 models for free dsDNA without DOX. They are summarized in **Table S8 in Appendix**. In addition, we have one model for free DOX in the solvent. The most recently recommended AMBER force field OL15 is used for the DNA parametrization [72], while the parameters and partial charges for free DOX were obtained as described in section 3.2 of chapter 3.

4.3 MD Simulation and BFE Protocols

In brief, we have adopted approaches similar to those used in section 3.3 of chapter 3 with some differences summarized below. First, all MD simulations have been performed in

the explicit TIP3P water model throughout 300 ns by using PMEMD.CUDA module implemented in AMBER 18 simulation package [243]–[245]. Second, each system underwent six stages of energy minimization rather than two stages. In the first five stages, different restraint force constants of 500, 250, 100, 10, 1 kcal/mol-$Å^2$ were applied to hold the solute (DOX-DNA complex, DNA, or DOX) fixed in its position and allow only the positions of the water molecules and ions to relax. The last stage of minimization is the same as the one described in section 3.3, in which the entire system is minimized without any constraints. Third, the heating, equilibration, and production protocols, as well as their setting parameters, are the same as those used in section 3.3, with the exception that each NPT production MD run was initiated for the full 300 ns. Fourth, once the MD simulations were finished, the snapshots were taken for every 10 ps over whole 300 ns, so in total 15000 snapshots were extracted for the BFE post-process analysis at ten different uni-valent NaCl salt concentrations (0.1–1 M). This analysis has been carried out by adopting the 3TP approach of the MM-GBSA method through the MMPBSA.py module of AMBER [160]. In 3TP, it is necessary to simulate three separate systems (DOX-DNA complex, free or unbound DNA, and free or unbound DOX) for each model. The free or unbound DOX system is identically designed for all models, so it is simulated only twice without having to repeat its MDs. The setting parameters for performing the MM-GBSA approach are similar to those in section 3.3. Finally, in addition to the SEM that is estimated by the MMPBSA.py module of AMBER, the block analysis approach [172] is also followed to check the convergence and numerical accuracy of the mean values in the BFE analysis.

4.4 Results and Discussion

4.4.1 Stability of the MD Simulations

RMSD is used to analyze the structural stability of our models. In the current study, RMSD was computed for each model from MD simulation trajectories using the CPPTRAJ program [166] in AMBER by taking the starting point for the MD production simulations (after the minimization and equilibration procedures) as reference. Here we showed only the RMSDs of group A as an example to discuss the structural stability of MD simulation. The RMSDs and their relative frequency (RF) distributions are shown in **Figures 4.2** and **4.3** of heavy atoms of DOX-DNA complex and free dsDNA.

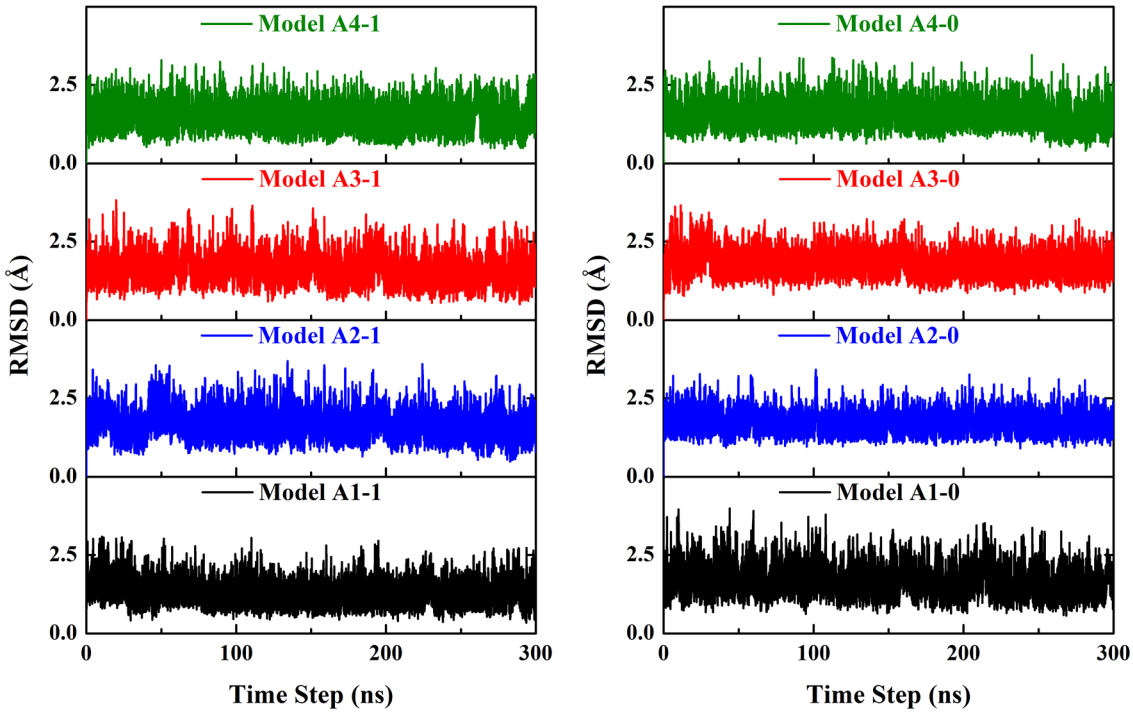


Figure 4.2. The root mean square deviation (RMSD) of the heavy atoms of DNA (right panel) and DOX-DNA complex (left panel) for models in group A as a function of simulation time throughout 300 ns.

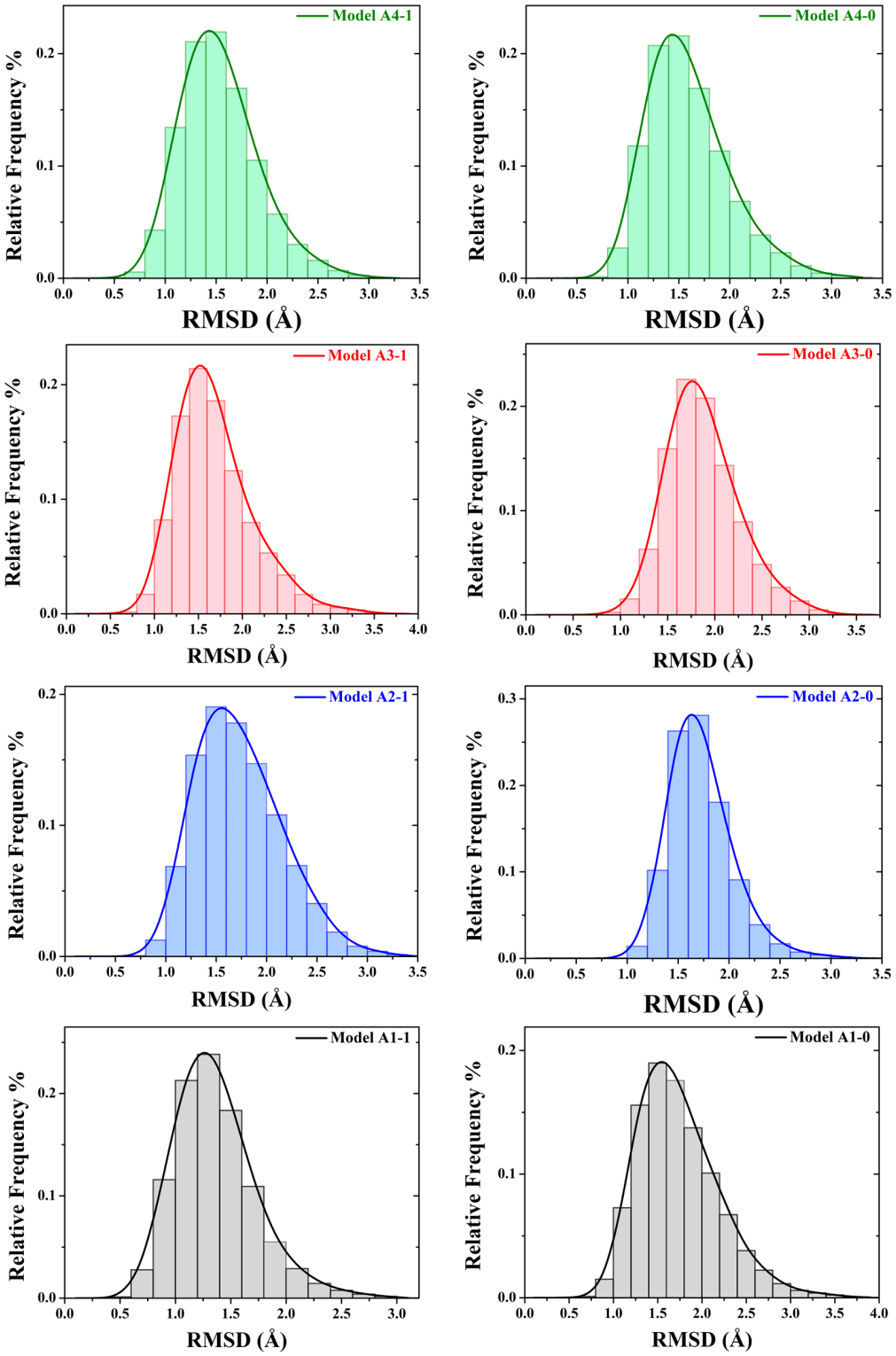


Figure 4.3. Relative frequency (RF) distribution histogram of RMSD of the heavy atoms of DNA (right panels) and DOX-DNA complex (left panels) for models in group A.

As seen in **Figure 4.2**, the RMSD of all models is stable as shown by flat and small RMSD fluctuations, and the MD simulations have attained a steady state. The mean RMSD of free DNA models ranges from 1.6 to 1.8 Å, whereas it ranges from 1.3 to 1.7 Å for the DOX-DNA complex. The mean RMSD of free DNA is slightly larger than that of the DOX-DNA complex. **Figure 4.3** for the RMSD histogram distributions show there is a range from 82 to 94% of the RMSD confined between 1 to 2.2 Å.

To check the convergence of the calculated binding free energy (ΔG_{calc}) at 0.2 M salt concentration as a function of the total simulation time of 300 ns, we used the block analysis [172]. For each MD simulation, the time series of ΔG is divided into 100 blocks corresponding to 3 ns of frames per block (**Figure 4.4**). From **Figure 4.4**, the uncertainties of the ΔG quantities are less than 0.35 kcal/mol in all models.

4.4.2 DOX-DNA Intercalation and Sequence-Dependent Binding

We performed MD simulations in the explicit TIP3P water model throughout 0.3 μs for DOX with 16 separate tetradecamer sequences of dsDNA. We adopted the 3TP approach of the MM-GBSA method to calculate the BFE at 310 K (37 °C), neutral pH, and ten different uni-valent NaCl salt concentrations (0.1–1 M).

Table 4.1 lists the BFE (ΔG_{calc}) at 0.2 M salt concentration of the 16 models, as well as with all the BFEs plotted in **Figure 4.5(a)**. The effect of the standard concentration of DOX on BFE is also taken into account in these BFE values. We estimate the standard concentration dependent free energy (ΔG_{con}) of DOX to be -2.18 kcal/mol using Eq.(24). To be more specific, the concentration of DOX (C_A) is calculated to be 0.0084 M (one molecular of DOX in a rectangular box of 52× 52×73 Å^3) and the value of $C_B = (14/4) \times C_A$ = 0.0294 M.

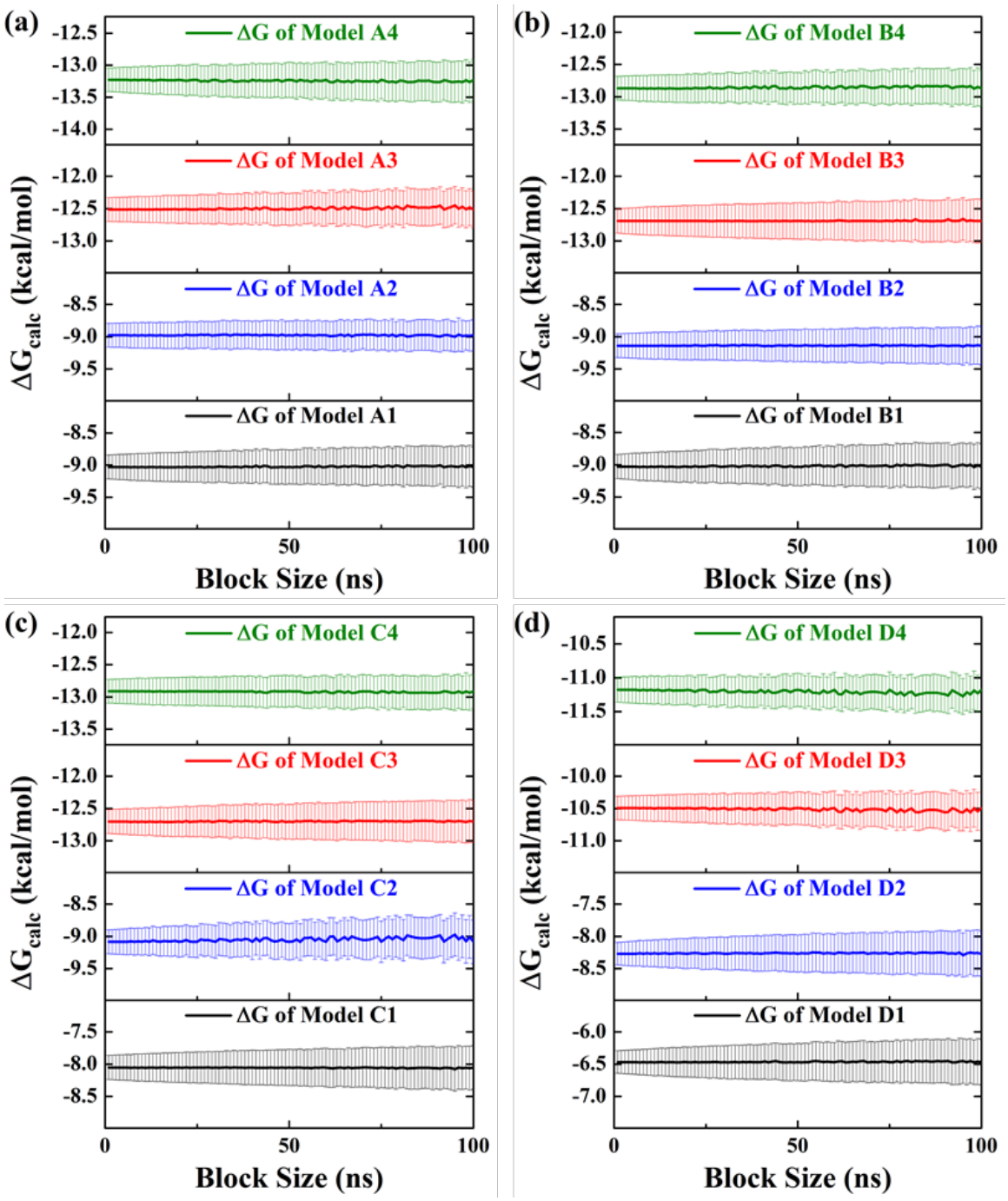


Figure 4.4. Convergence plot of the calculated binding free energy (ΔG_{calc})at 0.2 M salt concentration as a function of the total simulation time of 300 ns. Each block in x-axis is equal to 3 ns (i.e. 3 ns x 100 = 300 ns). (a) For models in group A; (b) For models in group B; (c) For models in group C; and (d) For models in group D.

Table 4.1. Binding free energy (ΔG) and their decompositions in terms of the two-step process of the DOX intercalation process at salt concentration of 0.2 M and different DNA sequences. Their relative BFEs (ΔΔG) is evaluated with reference to Model A4. All energy in kcal/mol unit. Note ΔG_{calc} is the ΔG_{bind} after substratcing ΔG_{con}.

Models	Conformation			ΔG_{Ins}	ΔG_{bind}	ΔG_{calc}	$\Delta\Delta G$
	DOX	DNA	overall				
A1	1.88	15.65	17.53	-28.73	-11.20 ±0.3	-9.02	4.20
A2	1.74	14.65	16.39	-27.55	-11.16 ±0.3	-8.98	4.25
A3	1.52	14.06	15.57	-30.26	-14.69 ±0.3	-12.51	0.71
A4	1.36	14.41	15.77	-31.18	-15.41 ±0.3	-13.23	0
B1	1.87	15.11	16.98	-28.19	-11.21 ±0.3	-9.03	4.20
B2	1.77	14.28	16.05	-27.37	-11.32 ±0.3	-9.14	4.09
B3	1.52	13.58	15.10	-29.97	-14.87 ±0.3	-12.69	0.54
B4	1.37	14.01	15.38	-30.43	-15.05 ±0.3	-12.87	0.36
C1	1.57	15.45	17.03	-27.26	-10.23 ±0.3	-8.05	5.17
C2	1.54	14.50	16.04	-27.30	-11.26 ±0.3	-9.08	4.14
C3	1.39	13.59	14.98	-29.86	-14.88 ± 0.3	-12.70	0.53
C4	1.51	14.01	15.52	-30.61	-15.09 ±0.3	-12.91	0.32
D1	1.37	14.94	16.31	-24.96	-8.65 ±0.3	-6.47	6.76
D2	1.48	14.48	15.96	-26.41	-10.45 ±0.3	-8.27	4.96
D3	1.52	12.91	14.43	-27.10	-12.67 ±0.3	-10.49	2.73
D4	1.48	13.46	14.95	-28.31	-13.36 ±0.3	-11.18	2.05

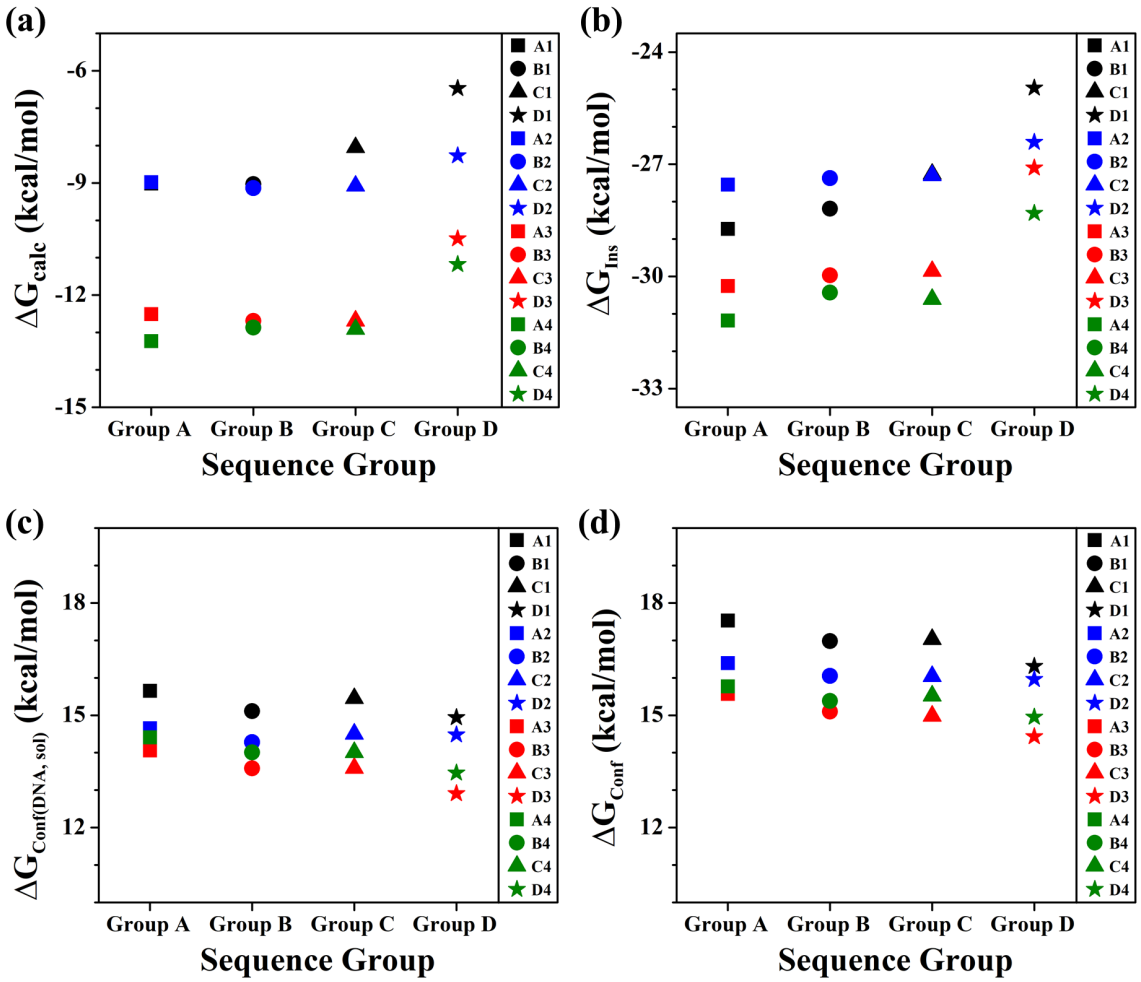


Figure 4.5. The binding free energy (BFE) analysis in terms of the two-stage intercalation process for all models of DOX-DNA complex vs sequence groups at at 0.2 M salt concentration. (a) The calculated binding free energy (ΔG_{calc}); (b) The energetic of the insertion step (ΔG_{Ins}); (c) The DNA defamation energetic cost $\Delta G_{Conf(DNA,sol)}$; and (d) The overall energetic penalty of the conformational change step (ΔG_{Conf}).

Again, ΔG_{con} should be subtracted from our calculated values of BFE computed by the MM-GBSA method. The ranking for preferred DOX binding from high to less favorable is s4 ≥ s3 > s2 ≥ s1, where s_i (i = 1-4) labels a sequence in its group. This order also coincides with the other BFEs at different ionic concentrations. The average total BFE (ΔG_{calc}) of 16 models is -10.41 ± 0.53 kcal/mol which is consistent with the experimental values of DOX binding

with ctDNA of -11.2 ± 1.2 kcal/mol at the same solution conditions (i.e. $\Delta G = \Delta H - T\Delta S$ and this -11.2 = -9.7 – 1.5) [157], [246]. Model A4, which has a d(CGCG**AT**C|G**AT**CGCG)$_2$ DNA sequence, has the best ΔG_{calc} of -13.23 ± 0.29 kcal/mol which falls within the relevant theoretical ranges of daunorubicin (DAU) and DNA complex [221], [247]. Both DOX and DAU are anthracycline drugs with a difference in hydroxyl group vs hydrogen.

By comparing the four sequences in each group, **Table 4.1** and **Figure 4.5(a)** show that the DOX daunosamine side chain, which serves as the minor groove binding agent, prefers binding with the sequences that have **(AT)$_2$** or **(TA)$_2$** BPs adjacent to the intercalation site, which is also consistent with experimental [29], [30] and other theoretical studies [31], [32]. It shows the reversal of **(TA)$_2$** to **(AT)$_2$** BP is slightly important, specifically in models of groups A and D, and confirms the previous findings [34], [162]. Our data also confirms that DOX has no tendency to choose **(CG)$_2$** over the **(GC)$_2$** in groups A and B, and is slightly preferred in groups C and D. Both are unfavorable compared to (TA)$_2$ and (AT)$_2$.

Comparing the four groups, our results indicate that there is a slight difference between the BFEs of groups A **(CG)$_2$**, B **(TG)$_2$**, and C **(CA)$_2$**. The mean ΔG_{calc} values for four sequences in groups A, B, and C are -10.94 ± 1.13, -10.93 ± 1.07, and -10.68 ± 1.24 kcal/mol, respectively. The sequences of group D are relatively less binding for DOX when compared to the corresponding sequences of groups A, B, and C (e.g., comparing D1 with A1, B1, or C1 and other sequences). However, the ΔG_{calc} of model D3, which has a d(CGCG**TA**T**A**T**A**CGCG)$_2$ DNA sequence, is slightly favorable compared to model A2, which has a d(CGCG**CG**C**G**C**G**CGCG)$_2$ DNA sequence. The models D3 and A2 could be considered as poly(dA-dT)$_2$ and poly(dG-dC)$_2$ respectively. The computed values of ΔG_{calc}

are -10.49 ± 0.29 kcal/mol and -8.98 ± 0.29 kcal/mol for D3 and A2, being also close to the experimental values of -9.5 and -8.7 kcal/mol, respectively [248]. It should be mentioned that our values are obtained for T = 310 K, neutral pH, and 0.2 M salt concentration, whereas the experimental values were measured at T = 298.15 K (25 °C) for DAU with regular alternating sequences, the other solution conditions being the same. Since the BFE is largely dependent on temperature and ionic strength [163], [222], a slight discrepancy is understandable. The mean ΔG_{calc} values for four sequences in groups B and C are well matched to the experimental value of -10.78 ± 0.3 kcal/mol at T = 298.15 K and 0.2 M ionic strength for poly(dA-dC).poly(dG-dT) [249]. Since experimental data for BFE of DOX with different DNA sequences are very limited, we chose the existing data for the DAU-DNA complex to compare our results.

Additionally, our calculated values agree well with experimental evidence on sequence selectivity [30]. The experimental observations show that seven of the ten best binding sites are at the end of the triplet sequences 5'(A/T)CG or 5'(A/T)GC, and the remaining three sites are at 5'(A/T)C(A/T). All the six strongest binding complexes (ΔG < -12.51 kcal/mol) in our study have quartet sequences of 5'CG(AT/TA), 5'TG(AT/TA), or 5'CA(AT/TA). The (AT/TA) means either $(AT)_2$ or $(TA)_2$ that are flanking the DOX intercalation site. In our study, the DOX intercalation site is designed to be in the middle of the tetradecamer sequences of dsDNA between BP7 and BP8 (see **Figure 4.1**).

The SEM values that are calculated using the MM-GBSA method (**Table 4.1**) are within the same range of 0.35 kcal/mol estimated by Block analysis (**Figure 4.4**). The high precision and similarity of SEM values in all models are attributed to the fact that the mean

value of ΔG or G for each species is obtained from 15000 independent energy values of the running MD simulations with high stability, while all models are built and simulated in the same manner. Both errors estimated by the MM-PBSA.py and block analysis, together with the high stability of the MDs, enable us to compute the BFEs and their relative comparisons with full confidence.

4.4.3 Importance of BPs Adjacent to DOX Intercalation Site

To investigate the significance of BPs that are located directly next to the DOX intercalation site, we estimated the relative BFE ($\Delta\Delta G$) with respect to the best BFE (model A4 as reported in **Table 4.1**) using the direct method according to Eq.(20). It is obvious from the last column of **Table 4.1** that A3, B3, B4, C3, C4 have slightly less stable BFEs than A4, falling in the range of 0.32 to 0.71 kcal/mol. They are the second strongest binding sequence of DOX. Our results reveal that A4 is more favorable than D3 and D4 by 2.73 and 2.05 kcal/mol, respectively. All these models have sequences with a **$(TA)_2$** or **$(AT)_2$** adjacent to the DOX intercalation site on both sides. A more striking observation regarding the $\Delta\Delta G$ values is that models having the **$(GC)_2$** or **$(CG)_2$** next to the DOX intercalation site, display, in general, a large unfavorable relative energy, ranging from 4.09 to 6.76 kcal/mol when compared to A4. The C1 and D1 are disfavoured by 5.17 and 6.76 kcal/mol, respectively, while both have **$(GC)_2$** BPs next to the DOX intercalation site. Within the context of $\Delta\Delta G$, one can infer that DOX exhibits sequence selectivity toward **$(TA)_2$** or **$(AT)_2$** in the vicinal location to the DOX intercalation site. This selectivity is directly compatible with the sequence preference of topoisomerase II-stimulated DNA cleavage [250], [251].

At this stage, it is appropriate to raise five key questions: At what stage of the intercalation process (conformation change and DOX insertion stages) DOX selectivity occurs? What is the key driving force responsible for the DOX-DNA intercalation process? What are the characteristics of this driving force that play a major role in determining a preferred sequence? What is the best model to describe the relationship between binding affinity and selectivity of an intercalator drug? How does the aqueous environment impact the intercalation process and how does ionic concentration affect the binding mechanism? To address these questions, we have broken down the computed BFE (ΔG) into its different energy components, which can be achieved by four different protocols. First, ΔG of the DOX intercalation process will be analyzed in detail in terms of the two-step process as described in Eq.(21). Second, ΔG can be parsed in terms of the interaction medium (see Eq.(18)), vacuum (i.e. the solute interactions), or with solvent (i.e. solvent interactions). Or ΔG can be further decomposed into physically interpretable terms. Third, the BFE can be further decomposed to the contribution of each residue within the DOX-DNA complex. And finally, ΔG can be decomposed into polyelectrolyte and non-polyelectrolyte contributions, as described in Eq.(25). These developments will be discussed in the next sections.

4.4.4 Origin of Sequence Preference of DOX binding to DNA

To understand in detail the source of DOX sequence selectivity, we have dissected the BFE into its energetic components as mentioned in the previous section. In the five subsections that follow, we will separately discuss the four different energy contributions $\Delta G_{Conf(DOX,sol)}$, $\Delta G_{Conf(DNA,sol)}$, ΔG_{Conf} and ΔG_{Ins}. At the same time, we will describe the relationship between the DNA sequence and the corresponding energetic terms. In the end subsection, we

will summarize this relationship. The BFEs associated with each step of the DOX intercalation process for all models at 0.2 M salt concentration are listed in **Table 4.1**.

I. Energetic cost for DOX conformational change

DOX conformational changes ($\Delta G_{Conf(DOX,sol)}$) are unfavorable (i.e., correspond to positive energy) and quantitatively similar in all models with an average energy of 1.56 kcal/mol. Consequently, it appears that DOX can be classified as an iso-energetic conformer in all DNA sequences. A more detailed investigation shows that these conformational changes are not arising from the structural changes of DOX per se (**Figure 4.6(a)**). ΔE_{int} contribution to $\Delta G_{Conf(DOX,sol)}$ is favorable and mostly similar in all models with an average of -1.46 ± 0.07 kcal/mol, driven by the favorable dihedral energy components. Our understanding of this favored ΔE_{int} is that the bound DOX in the complex is stabilized by unique HBs [34], making its internal dihedral angles more restrictive than those of the free DOX in the solvent. On the other hand, ΔE_{vdW} is 1.27 ± 0.04 kcal/mol on average, which is an energetically disfavored contribution. ΔE_{ele} is consistently favorable with an average value of -0.82 ± 0.09. In contrast to ΔE_{ele}, the polar contribution (ΔG_{GB}) is largely unfavorable with an average of 2.48 ± 0.12 kcal/mol. Therefore, ΔE_{ele} is completely overshadowed by ΔG_{GB}, leading to an overall unfavorable contribution of the total electrostatic component (ΔG_{ele}). ΔG_{SA} in all the models is only -0.05 kcal/mol. The loss of configuration entropy ($-T\Delta S$) at 310 K is quite similar in all models with a small average value of 0.14 kcal/mol. Our findings thus imply that ΔG_{ele} is the main driving force behind the conformational changes in DOX, and these changes are the result of a decreased interaction between DOX and the surrounding solvent molecules when DOX intercalates between the BPs of DNA.

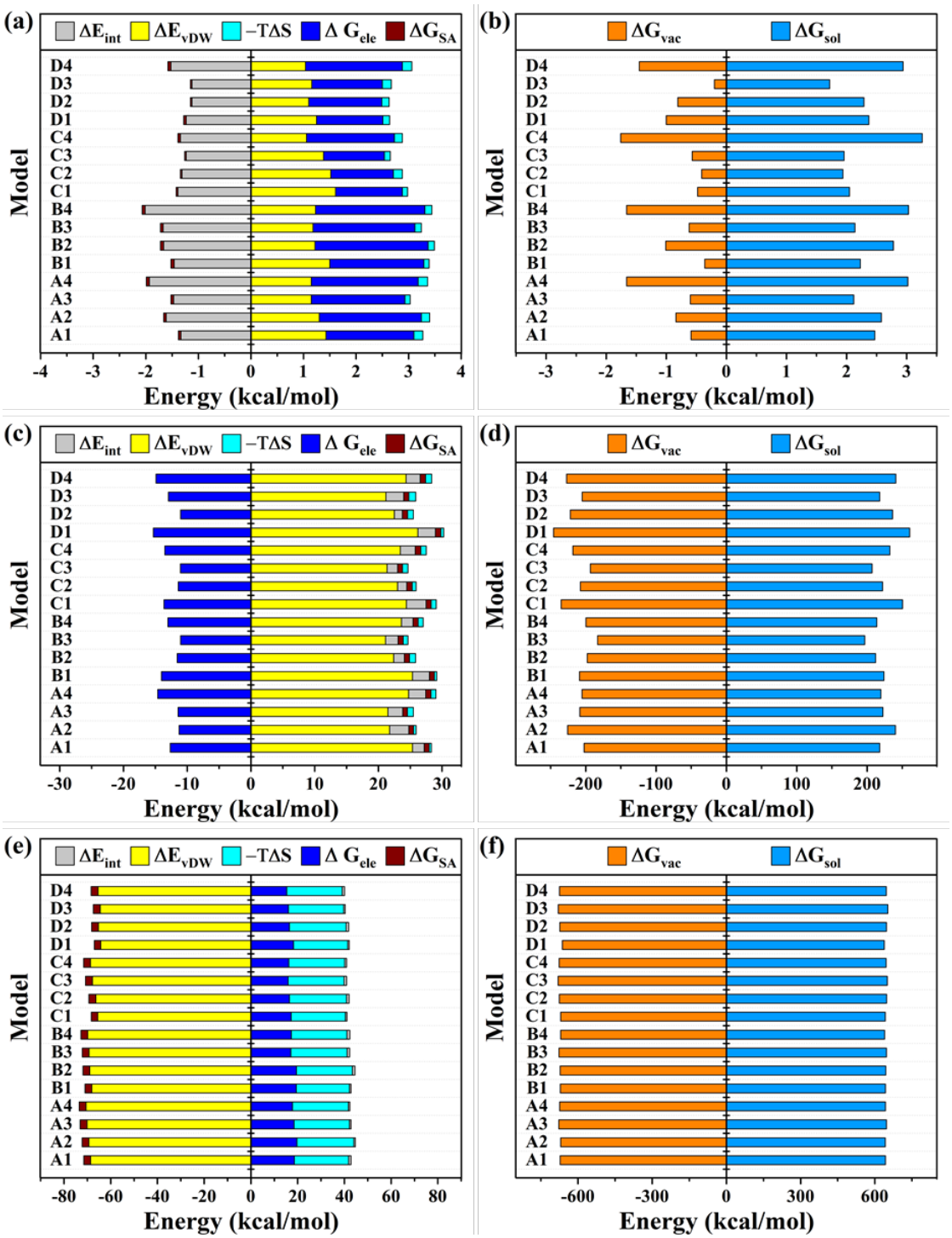


Figure 4.6. The histogram displays BFE analysis of the two-stage DOX binding process decomposed in terms of the physical factors (left panels) or the nature of interactions in a vacuum or solvent (right panels). (a) and (b) for DOX energetic cost; (c) and (d) for DNA deformation cost in the first step; (e) and (f) for insertion energetic binding in the second step.

This behavior becomes even more pronounced when the $\Delta G_{Conf(DOX,sol)}$ is further decomposed in terms of the nature of the interaction in a vacuum and/or solvent, as shown in **Figure 4.6(b)**, yielding DOX conformation ΔG_{vac} and ΔG_{sol} values of -0.88 ± 0.13 and 2.43 ± 0.11 kcal/mol, respectively.

II. Energetic Penalty for DNA Deformation

The energetic penalty for the DNA deformation change, $\Delta G_{Conf(DNA,sol)}$, is significant and amounts to 14.32 ± 0.18 kcal/mol on average. Additionally, the average value of four sequences in each group is 14.69 ± 0.47, 14.25 ± 0.4, 14.39 ± 0.41 and 13.95 ± 0.5 kcal/mol for A, B, C, and D group, respectively. Therefore, the energy penalty order from the highest to the lowest is A $(CG)_2$ > B $(TG)_2$ ≈ C $(CA)_2$ > D $(TA)_2$. Moreover, the DNA deformation cost in each group exhibits a systematic order of s1 > s2 > s4 > s3 (**Table 4.1** or **Figure 4.5(c)**). Our analysis also shows that the binding deformation energy of DNA is moderately dependent on the DNA sequence, which is necessary to have a minimal DNA deformation energy for higher binding affinity. Model D3 appears to have the lowest energy cost of DNA deformation (12.91 kcal/mol), while Model A1 appears to have the highest (15.65 kcal/mol). D4 also has a low deformation cost of 13.46 kcal/mol. The reason behind this could be related to the rigidity and molecular structure of D3 and D4 when compared to other models (see **Figure 4.1(b)**). The calculated DNA energetic penalty for forming the intercalation cavity for DOX in our study is within the range of previously obtained values [217]–[220]. However, we believe that our values are more accurate since we take into account the solvent effect, the configurational entropy changes, the sequence effect and the overall DNA conformational energy changes.

We have also computed the various energy components contributing to the $\Delta G_{Conf(DNA,sol)}$ which are all plotted as a histogram in **Figure 4.6(c)**. Unlike the favorable contribution of ΔE_{int} to $\Delta G_{Conf(DOX,sol)}$, it is energetically unfavorable to $\Delta G_{Conf(DNA,sol)}$ with an average value of 2.22 ± 0.14 kcal/mol. Second, the contribution of ΔE_{vdW} to the DNA deformation is 23.27 ± 0.42 kcal/mol on average, which is mainly unfavorable (positive) as a result of the increase in the distance between the BPs of dsDNA to form the DOX intercalation cavity. Our estimated value is well matched with previously reported 22.2 ± 1.4 kcal/mol of 1:1 DAU-DNA complex [247]. Our results also predict a systematic penalty progression for ΔE_{vdW} with a sequence s1 > s4 > s2 ≥ s3. In addition, D1 has a large ΔE_{vdW} penalty of 26.22 kcal/mol, whereas the lowest penalty is found in almost all s3 models. Third, ΔG_{ele} is negative with an averaged value of -12.71 ± 0.37 kcal/mol and is the result of the interplay of two large contributions of opposite signs, the ΔG_{GB} with a positive value of 225.13 ± 4.14 kcal/mol that is completely balanced by a negative value of -237.84 ± 4.34 kcal/mol from the Coulomb interaction (ΔE_{ele}). The intercalation of DOX between the local BPs of DNA causes an increase in the DNA helix rise parameter of about 4 Å, which leads to a favored ΔE_{ele} due to the reducing electrostatic repulsions between the local charges of the DNA phosphate groups. However, this elongation is also associated with unfavorable ΔG_{GB} due to a decrease in the total charge density of unwinding DNA, leading furthermore to a weakening in the interaction between this DNA and the surrounding aqueous environment. In general, models with sequence s1 or s4 have more favored ΔG_{ele} than s2 and s3. Fourth, a value of 0.81 ± 0.1 kcal/mol of ΔG_{SA} is obtained for DNA unwinding deformation that has a larger SASA than the B-form DNA due to its elongation, leading to an unfavourable contribution due to the

additional molecular surface solvent exposure. This value is also well consistent with previous studies of the 1:1 complex [247]. Fifth, the loss of configuration entropy ($-T\Delta S$) is calculated to be 0.73 ± 0.06 kcal/mol, and depends slightly on the DNA sequence, an effect too small to drive the overall DNA deformation. Our value of ($-T\Delta S$) is also in good agreement with the previously reported result of 0.48 kcal/mol for 1:1 DAU-DNA complex [247]. The main conclusion that can be drawn from these analyses is that the penalty of ΔE_{vdW} is the main factor that drives the unwinding conformational transition from the B-form DNA to form the intercalation cavity, a conclusion consistent also with a previous study [210].

Figure 4.6(d) shows the analysis of $\Delta G_{Conf(DNA,sol)}$ in terms of the interaction nature of the total ΔG_{vac} and ΔG_{sol}. The averaged ΔG_{vac} across all models is -211.62 ± 4.07 kcal/mol, while the desolvation free energy ΔG_{sol} is 225.94 ± 4.15 kcal/mol. The ΔG_{vac} is completely cancelled by ΔG_{sol}. Furthermore, this trend shows that the aqueous solution (i.e., water and ions) opposes the unwinding DNA and consequently we expect that the DNA conformational transition step is salt-dependent (this effect will be discussed in detail later).

III. Energetic penalty of the overall conformational change (First step)

The first thermodynamically unfavorable step of the DOX intercalation process, described by Eq.(22), requires on the average ΔG_{Conf} of 15.87 ± 0.2 kcal/mol to form the DOX binding site. The energy penalty dependence on the DNA sequence is similar to the one found for $\Delta G_{Conf(DNA,sol)}$ (**Figure 4.5(d)**). The decomposition of ΔG_{Conf} in terms of physical components amounts to ΔE_{int} = 0.75, ΔE_{vdW} = 24.54, ΔG_{ele} = -11.06, ΔG_{SA} = 0.76 and $-T\Delta S$ = 0.87 kcal/mol, while its decomposition in terms of the nature of interaction is ΔG_{vac} = -212.5 and ΔG_{sol} = 228.37 kcal/mol.

IV. Binding energy of the insertion step (Second step)

The unfavourable conformational energy barrier ΔG_{Conf} is overcome by the favorable binding energy (ΔG_{Ins}) of the insertion step, corresponding to an average energy of -28.47 ± 0.43 kcal/mol. The insertion step depends crucially on the DNA sequence. ΔG_{Ins} with respect to four sequences in each group is A = -29.43, B = -28.99, C = -28.76, and D = -26.69 kcal/mol. The preferential binding at this step is A > B ≥ C > D and, interestingly, the preferential ordering shown earlier, s4 ≥ s3 > s2 ≥ s1, is slightly different at the insertion step of the DOX intercalation process, amounting to s4 > s3 > s1 ≥ s2 (**Figure 4.5(b)**), with one exception that D2 > D1. In addition, A4 model also has the strongest preferential binding at this step with ΔG_{Ins} of -31.18 kcal/mol. It can be therefore concluded that the DOX preferential binding usually occurs at the insertion binding stage.

Further decomposition of the total ΔG_{Ins} into its energetic components leads to the following conclusions (**Figure 4.6(e)**). ΔE_{int} contribution to ΔG_{Ins} is unfavorable (positive) with an average value of 0.88 ± 0.05 kcal/mol. The contribution of ΔE_{vdW} is largely favorable (negative) in all DOX-DNA complexes and dominates the other components with an average value of -67.67 ± 0.53 kcal/mol. To be effective at the insertion stage, DOX should have a high ability to fit well between the BPs of the DNA, which, of course, requires many intermolecular interactions between DOX and the stacked BPs of dsDNA. This value is well matched by -66.7 kcal/mol reported for 1:1 DAU-DNA complex and calculated by adapting STP of MM-GBSA method [247]. It is also in the same range of values as those reported elsewhere for 1:1 complex of DOX-hexamer sequences [252]. A3 and A4 display the highest attractive ΔE_{vdW} of -70.16 and -70.6 kcal/mol, respectively, while the lowest values are for D1 and D3,

amounting to -64.23 and -64.45 kcal/mol, respectively. Generally, we notice that in all groups the ranking of the most attractive ΔE_{vdW} component is s4 > s3 > s2 > s1, with one exception: D2 > D3. More surprisingly, this preferred order is almost identical to that of the ΔG_{calc}, and we thus expect that the attractive component of ΔE_{vdW} plays an important role in the sequence selectivity (vide infra). In sharp contrast to the ΔE_{vdW} term, the overall electrostatic contributions (ΔG_{ele}) are unfavorable (positive) in all models at 0.2 M salt concentration with an average value of 17.48 ± 0.34 kcal/mol. This value is the result of the two major contributions with opposite signs, ΔE_{ele} = -630.48 ± 1.19 vs. ΔG_{GB} = 647.96 ± 0.96 kcal/mol. The negative ΔE_{ele} in all models is arising from the strong electrostatic atractions between the negatively charged phosphate backbone of DNA and the positively charged amino group of DOX. As ΔG_{GB} contribution is largely unfavorable (positive), the negative ΔE_{ele} gets cancelled by ΔG_{GB}, so the remaining ΔG_{ele} makes an unfavorable contribution to ΔG_{Ins}. The reported values of ΔG_{ele} are again consistent with previous study [252]. We also find that the ΔG_{ele} is an important factor in the sequence-dependent preferential binding (vide infra). The average value of ΔG_{SA} is -2.92 ± 0.05 kcal/mol, in agreement with the reported value of -3.3 kcal/mol of 1:1 DAU-DNA complex [247]. It obviously contributes favorably to ΔG_{Ins}. Importantly, we observe that the ΔE_{vdW}, ΔG_{ele} and ΔG_{SA} at the insertion stage have signs opposite to the same quantities at the conformational change stage (see **Figure 4.6(e) vs 4.6(c)**). Finally, the total entropic contribution ($-T\Delta S$) at 310 K and 0.2 M ionic concentration is 23.76 ± 0.12 kcal/mol, which reflects the important entropic cost stemming from DOX being trapped with little mobility when bound to DNA. This value is the same range as found in the studies of 1:1 complex [247]. From the above analysis it can be concluded that the attractive

component ΔE_{vdW} is the major factor that drives DOX to be inserted between the BPs of dsDNA.

ΔG_{Ins} can also be partitioned into ΔG_{vac} and ΔG_{sol} as shown in **Figure 4.6(f)**. ΔG_{vac} is -673.51 ± 1.1 kcal/mol on average, which is large and dominates over ΔG_{sol} of 645.04 ± 0.95 kcal/mol, and thus leads to a favorable overall ΔG_{Ins}.

V. Energetic contributions to DOX sequence selectivity

Table 4.2 shows the decomposition of ΔG_{bind} into its overall energetic components obtained by summing of two-stage binding energies in all 16 models at 0.2 M salt concentration, plotted in **Figure 4.7**. Briefly, the patterns of the signs of the energetic components of ΔG_{bind} are following those of the insertion step, ΔG_{Ins}, already discussed in the previous section (see **Figure 4.6(e) and Figure 4.7**).

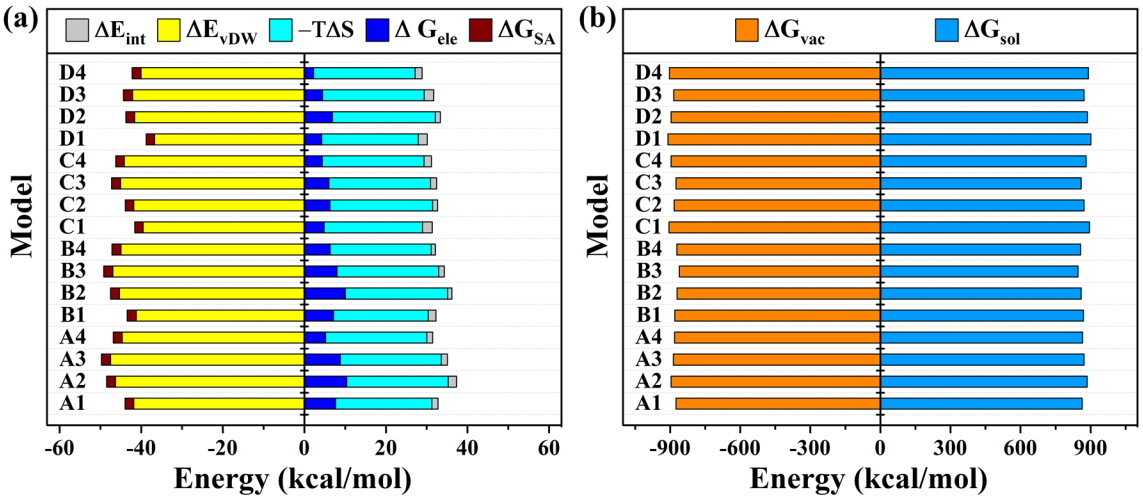


Figure 4.7. The histogram displays the total BFE analysis of the overall two-stage DOX binding process. (a) The total BFE is decomposed in terms of the physical factors; and (b) It is decomposed in terms of interaction nature in a vacuum or solvent.

Table 4.2. Physical component analysis of total BFE (ΔG_{bind}) at the salt concentration of 0.2 M. All energy values are given in kcal/mol. G_{ele} is the sum of ΔE_{ele} and ΔG_{GB}.

Models	ΔE_{int}	ΔE_{vdW}	ΔE_{ele}	$-T\Delta S_{Tot}$	ΔG_{GB}	ΔG_{SA}	ΔG_{ele}	ΔG_{bind}
A1	1.51	-41.83	-858.20	23.64	865.83	-2.15	7.63	-11.20
A2	2.05	-46.28	-877.37	24.89	887.73	-2.19	10.37	-11.16
A3	1.43	-47.50	-865.45	24.78	874.26	-2.21	8.81	-14.69
A4	1.40	-44.70	-862.41	24.83	867.61	-2.14	5.20	-15.41
B1	1.88	-41.24	-864.34	23.22	871.49	-2.23	7.14	-11.21
B2	1.10	-45.35	-851.41	25.07	861.43	-2.15	10.02	-11.32
B3	1.36	-46.95	-839.96	24.95	847.95	-2.23	7.99	-14.87
B4	1.06	-44.99	-852.34	24.76	858.65	-2.18	6.31	-15.05
C1	2.36	-39.53	-892.18	24.12	897.03	-2.04	4.85	-10.23
C2	1.20	-41.81	-867.93	25.22	874.17	-2.12	6.25	-11.26
C3	1.46	-45.06	-855.92	24.94	861.92	-2.22	6.00	-14.88
C4	1.84	-44.17	-878.80	24.91	883.20	-2.07	4.40	-15.09
D1	2.17	-36.76	-898.75	23.75	902.95	-2.02	4.21	-8.65
D2	1.21	-41.65	-881.06	25.23	887.95	-2.12	6.89	-10.45
D3	2.33	-42.12	-870.00	24.90	874.45	-2.24	4.45	-12.67
D4	1.73	-40.06	-890.17	24.88	892.42	-2.16	2.25	-13.36
Ave.	1.63	-43.13	-869.14	24.63	875.57	-2.15	6.42	-12.59

Table 4.3 lists the relative energies $(\Delta\Delta G)$ and their energetic components with respect to A4.

Table 4.3. The analysis of the relative BFE (ΔΔG) in terms of its associated physical decomposition at 0.2 M. ΔΔG is evaluated with respect to Model A4. All values in (kcal/mol).

Models	$\Delta\Delta E_{int}$	$\Delta\Delta E_{vdW}$	$\Delta\Delta G_{ele}$	$\Delta\Delta G_{SA}$	$-T\Delta\Delta S$	$\Delta\Delta G$
A1	0.11	2.86	2.43	-0.01	-1.20	4.20
A2	0.65	-1.59	5.17	-0.05	0.06	4.25
A3	0.03	-2.80	3.61	-0.08	-0.05	0.71
A4	0	0	0	0	0	0
B1	0.48	3.46	1.96	-0.09	-1.61	4.20
B2	-0.30	-0.65	4.82	-0.02	0.24	4.09
B3	-0.04	-2.25	2.79	-0.09	0.12	0.54
B4	-0.34	-0.30	1.11	-0.04	-0.07	0.36
C1	0.96	5.17	-0.34	0.10	-0.71	5.17
C2	-0.20	2.88	1.05	0.02	0.39	4.14
C3	0.06	-0.36	0.81	-0.09	0.10	0.53
C4	0.44	0.53	-0.80	0.07	0.07	0.32
D1	0.77	7.94	-0.99	0.12	-1.08	6.76
D2	-0.19	3.04	1.69	0.01	0.40	4.96
D3	0.93	2.58	-0.74	-0.10	0.07	2.73
D4	0.34	4.64	-2.94	-0.03	0.05	2.05

Among the energetic components of ΔG_{bind}, ΔG_{ele} contributes in an essential manner to the DNA sequence selectivity of the DOX-DNA interaction. The overall dependence of ΔE_{vdW} on the DNA sequence displays more variance, but also a systematic preferential ordering across all groups of s3 > s2 > s4 > s1, with the only exception being C4 > C2 (**Table 4.2**). Our results also show that ΔG_{SA} and the entropic loss of the translational $(-T\Delta S_{tran})$ and rotational $(-T\Delta S_{rot})$ degrees of freedom are not involved in the selectivity. The computed $(-T\Delta S_{tran})$ and $(-T\Delta S_{rot})$ at 310 K in all models are 13.82 and 11.73 kcal/mol, respectively. The s1 in all groups emerges as the more favored vibrational entropic contribution $(-T\Delta S_{vib})$. We conclude that the ΔG_{ele} is the main physical factor reflecting the DNA sequence selectivity and that ΔE_{vdW} plays a secondary role. The other contributions are marginal.

4.4.5 Reliable Model for DOX Selectivity and DOX Binding Site-Size

To propose the best model describing the relationship between BFE and the DNA sequence, as well as to determine the size of the DOX binding site, we performed per-residue BFE decomposition on all models of the DOX-DNA complex. The energetic contribution can be obtained by summing the interactions across all BPs involved in the DOX-DNA complex. It is worth mentioning that the entropic contributions are not included in this decomposition analysis [253], [254], as we have already shown that the sequence selectivity is independent of the $(-T\Delta S_{tran})$ and $(-T\Delta S_{rot})$, and only slightly affected by $(-T\Delta S_{vib})$.

The per-residue interaction spectrum of the DOX-DNA complex is shown as a histogram chart in **Figure 4.8** for all groups at 0.2 M salt concentration.

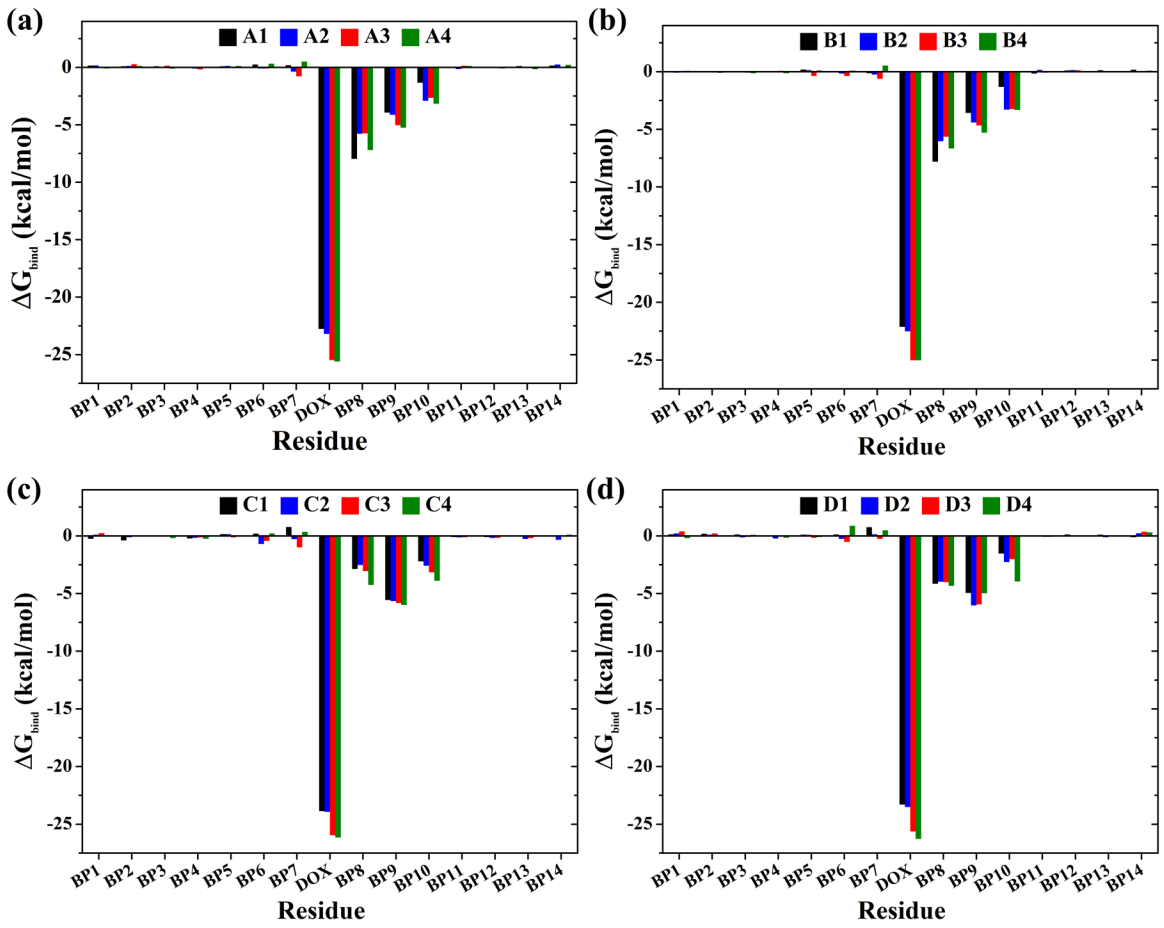


Figure 4.8. Decomposition of the BFE on a per-residue basis of DOX with all BPs of dsDNA at 0.2 M salt concentration. (a) For models in group A; (b) For models in group B; (c) For models in group C; and (d) For models in group D.

From **Figure 4.8**, our results essentially show clear evidence that the quartet sequence model is the best model to characterize the relationship between sequence selectivity and binding affinity. DOX thus requires at least four BPs to develop its selectivity. This important conclusion is also consistent with Kellogg et al. proposal of a quartet sequence model [40]. It is necessary to point out that we reach this conclusion using a rather different methodology based on MD simulation, and we determine the actual values for interaction energy of DOX

with all the BPs. Additionally, the binding size of 4 BPs is also in full accord with previous experimental studies [163], [240].

The main critical findings stemming from the DOX-BP interaction spectrum can be summarized as follows. The vicinal BP below the intercalation site (BP8) contributes much more to the interaction energy than the vicinal BP above the intercalation site (BP7). BPs located downstream of the intercalation site (BP9 and BP10) interact significantly with DOX, while the upstream BPs are largely not involved (BP5 and BP6). This finding clearly reveals the importance of the DOX daunosamine side chain in the DOX intercalation process. In addition, the BP8 contributes more than BP9 in groups A and B while the reverse occurs in groups C and D (**Figures 4.8**). BP8 and BP9 in groups A and B contribute to total BFE, without entropic contribution, -6.6 ± 0.34 and -4.53 ± 0.22 kcal/mol on average, whereas they contribute -3.64 ± 0.25 and -5.6 ± 0.15 kcal/mol in groups C and D, respectively. A close examination of the contributions of BP8 and PB9 in terms of their energetic component analysis shows that the BP8 displays an attractive component, ΔE_{vdW}, smaller than BP9, while its ΔG_{ele} is larger than PB9. Therefore, these different behaviors are primarily a consequence of ΔE_{vdW} and ΔG_{ele} in groups A and B vs. groups C and D.

Finally, the contribution of DOX to the total BFE of the DOX-DNA complex is about 66%, while the whole DNA contributes about 34%, mostly from BP8, BP9, and BP10. More interestingly, this contribution is sequence-dependent, and DOX interacts preferentially with s3 and s4 as opposed to s1 and s2 in all groups (**Figures 4.8**). To our best knowledge, this may be the first time that this contribution has been estimated explicitly. Based on this strong evidence one can assert that DOX recognizes its selective sequence. **Figure 4.9** shows the

decomposition from only DOX contribution in terms of the energetic components. Concisely, the attractive component ΔE_{vdW} is large and dominates other components, while ΔG_{ele} is a major source in the selectivity of s3 and s4. The DOX contribution in s3 is approximately equivalent to that in s4, and its contribution in s1 is like that in s2. We observe a preferred difference of relative energy between the DOX contributions in s3 or s4 and those in s1 or s2. This difference is about -2.5 kcal/mol, which is strong enough to allow DOX to preferentially select its site between the BPs along with DNA.

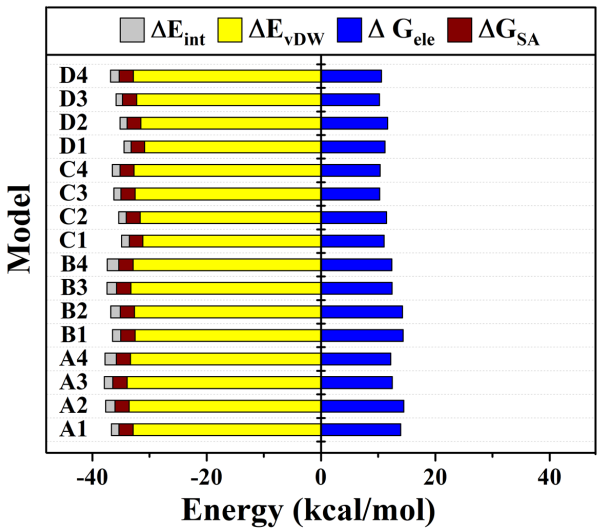


Figure 4.9. Decomposition of the BFE on a per-residue basis into the contribution from the physical factors of the interaction DOX drug only in the intercalation process for all models at 0.2 M.

4.4.6 Polyelectrolyte Effect on DOX Intercalation Binding to DNA

We anticipate finding a polyelectrolyte effect to the BFE for the DOX intercalation into DNA at each stage of the binding process because DNA is a highly charged polyanion and DOX is a cationic drug. Theoretically, the values of BFE (ΔG_{calc}) and its interaction decomposition of (ΔG_{Conf}) and (ΔG_{Ins}) are computed over a wide range of salt concentrations

from 0.1 to 1 M to assess the ionic effects on the BFE of DOX-DNA complexes and to quantify ΔG_{pe} and ΔG_t. **Figure 4.10** shows the ΔG_{calc} and its decomposition for models in all groups against the entire range of ionic concentrations. **Figure 4.10** shows that ΔG_{calc} becomes less exothermic with increasing salt concentration, in excellent agreement with the relevant experimental findings [163], [222]. In addition, our results indicate that salt dependence resides in both stages of binding, especially at low salt concentrations. For instance, ΔG_{Conf} of A4 is 15.4 kcal/mol at 0.1 M salt concentration and 16.41 kcal/mol at 1 M, whereas ΔG_{Ins} is -31.42 kcal/mol at 0.1 M and -30.63 kcal/mol at 1 M. The same trend is true for other models. At higher salt conditions, ΔG_{Conf} becomes more endothermic and ΔG_{Ins} less exothermic. The sensitivity of ΔG_{Conf} to salt effects results purely from the DNA deformation ($\Delta G_{Conf(DNA,sol)}$) during the formation of the DOX intercalation cavity. $\Delta G_{Conf(DOX,sol)}$ does not participate in this sensitivity, for example in A4, it is 1.36 kcal/mol at 0.1 M and 1.37 kcal/mol at 1 M.

A very interesting observation is that the entropic contribution becomes more favored as salt concentrations increase. For example in A4, $-T\Delta S$ is 24.92 at 0.1 M and 24.22 kcal/mol at 1 M, implying that there is a 0.7 kcal/mol favored entropic change when the salt is increased from 0.1 to 1 M. This trend is also observed for other models. Additionally, the response of DOX-DNA complexes to the ionic strength is almost identical in all models examined here and independent of the DNA sequence (**Figures 4.10**). This is not surprising since the Na^+ ions interact with the anionic DNA phosphate backbone only with non-specific electrostatic interactions, independent of the nature of the BP sequence. The analysis of ionic effects on the BFE shows that the stable complex is preferred at lower ionic strength solution environments.

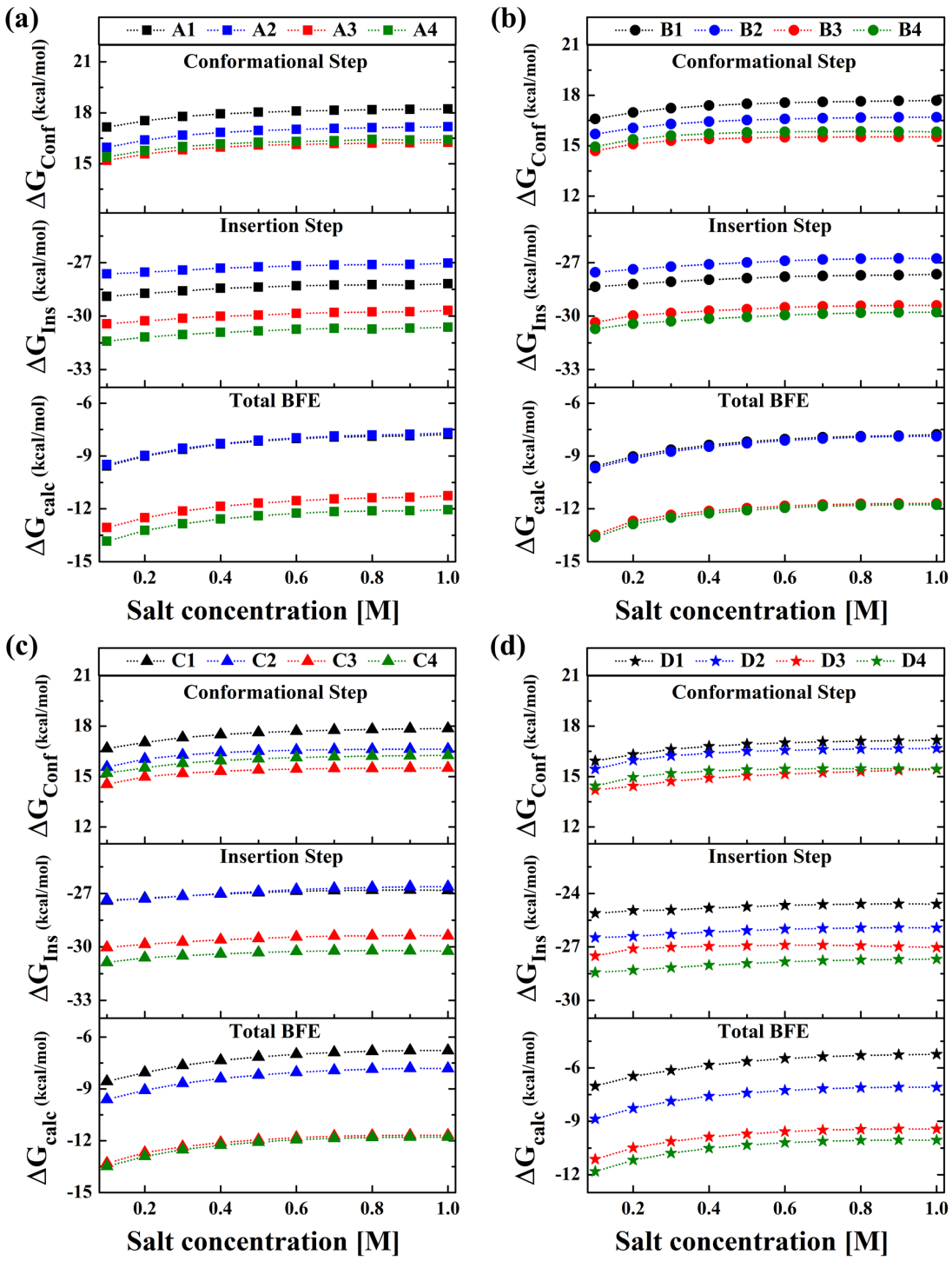


Figure 4.10. Dependence of the binding free energy of DOX–DNA complex on the salt concentration at each stage of binding process for models in group A (a); group B (b); group C (c); and group D (d). The upper panel is for conformational changes stage, the middle panel is for insertion binding stage and the bottom panel is for the total BFE.

Table 4.4 summarizes the values of ΔG_{calc} determined at 0.1, 0.2, and 1 M and temperature 310 K along with its partition into ΔG_{pe} and ΔG_t. ΔG_{calc} at 1 M is equal to ΔG_t because ΔG_{pe} is zero at this concentration. The salt dependence (SK) of the binding constants of DOX in all models is listed in **Table 4.4** and is estimated from a linear fit of a double logarithmic plot of $log(K_{calc})$ vs. $log[Na^+]$ for a set of salt conditions from 0.1 to 1 M. ΔG_{pe} can be quantified as a difference between ΔG_{calc} at a specific ionic concentration and ΔG_t or it can be predicted by using SK value,. The determined slope SK of -1.29 on average is in excellent agreement with Friedman and Manning predicted value of -1.24 for binding a monovalent intercalator drug to DNA as well as with the experimental value of -1.24 ± 0.15 for DUN-ctDNA and -1.2 for ethidium-ctDNA complex [165], [223], [224]. The value is slightly higher than the relevant experimental values of -1.06 or -0.99 for DOX-ctDNA complex [41], [255], where the small discrepancy may be explained by the fact that the effects of ionic concentration on the DNA conformational transition stage are included in our predicted values of SK. Analysis by Wilson and Lopp suggested that about 0.2 to 0.3 sodium ions should be released during the DNA conformation change stage regardless of the intercalator charge [224], [234]. Correcting our value of -1.29 for the DNA transition stage ions release yields a rescaled slope of -1.

ΔG_{pe} for DOX-DNA complexes at 0.1 M is -1.8 kcal/mol (**Table 4.4**), which is close to the experimental value of -2.3 ± 0.2 kcal/mol at 0.016 M and temperature 293.15 K for the DOX-ctDNA complex [18]. At 0.2 M the value of -1.18 kcal/mol is in extremely good agreement with the relevant experimental value of -1 kcal/mol at the same salt condition and T = 293.15 K [165].

Table 4.4. Decomposition of BFE into polyelectrolyte (ΔG_{pe}) and non-polyelectrolyte (ΔG_t) contributions with salt dependence of the binding constants (SK) of DOX–DNA complexes.

Models	SK	Energy (kcal/mol) at different salt concentrations (M)				
		ΔG_{calc}		ΔG_t^{sim}	ΔG_{pe}	
		0.1	0.2	1	0.1	0.2
A1	-1.29	-9.57	-9.02	-7.78	-1.79	-1.25
A2	-1.30	-9.49	-8.97	-7.69	-1.81	-1.28
A3	-1.29	-13.06	-12.51	-11.26	-1.81	-1.26
A4	-1.29	-13.84	-13.23	-12.05	-1.79	-1.18
B1	-1.30	-9.58	-9.03	-7.78	-1.80	-1.25
B2	-1.33	-9.68	-9.14	-7.88	-1.80	-1.26
B3	-1.25	-13.48	-12.69	-11.67	-1.78	-1.02
B4	-1.31	-13.61	-12.87	-11.77	-1.83	-1.10
C1	-1.34	-8.57	-8.05	-6.78	-1.80	-1.28
C2	-1.34	-9.61	-9.08	-7.82	-1.79	-1.27
C3	-1.20	-13.31	-12.70	-11.61	-1.70	-1.04
C4	-1.25	-13.49	-12.91	-11.79	-1.70	-1.12
D1	-1.31	-7.02	-6.47	-5.23	-1.78	-1.24
D2	-1.31	-8.87	-8.27	-7.08	-1.79	-1.19
D3	-1.23	-11.13	-10.49	-9.43	-1.70	-1.06
D4	-1.28	-11.82	-11.18	-10.06	-1.76	-1.13
Ave.	-1.29	-11.01	-10.41	-9.23	-1.78	-1.18
SEM	0.01	0.55	0.53	0.56	0.01	0.02

As the salt concentration increases, ΔG_{pe} becomes less favored due to an increase in electrostatic repulsions between the positive charge of DOX amino sugar and Na^+ ions in the surrounding bathing solution. Since the positive charge of DOX may act to neutralize DNA as the counterion during the binding process, the bound Na^+ cation may be subsequently released [165], implying a thermodynamic link between the binding of DOX and release the Na^+ ion. The binding of DOX to DNA will thus affect the binding of sodium ions and vice versa [18]. The predominant factor supporting our conclusion that a stable complex is preferably formed at lower ionic strength conditions is that ΔG_{pe} is more favored. Since the errors in the average of SK and ΔG_{pe} are not significant across all models, these two quantities are also independent of the DNA sequences (**Table 4.4**). The average value of ΔG_t is -9.23 ± 0.56 kcal/mol, which appears to be in reasonable accord with the experimental value of -7.7 ± 0.3 kcal/mol for DOX-ctDNA complex [18], [165]. The slight difference is primarily arising from the different temperatures (T = 293.15 K vs. 310 K in our study) and differences in the number of base pairs per site (3.4 vs. 4 in our study). This value is also consistent with the average value of -9.6 ± 1.1 for the DOX-DNA complex of 32 DNA quartet sequences obtained by using the HINT program [28], [40]. More importantly, the higher negative value of ΔG_t is another key indicator to reinforce our conclusion that van der Waals interactions are the predominant molecular factor driving force for the formation of a stable DOX-DNA complex. This conclusion is in excellent agreement with the X-ray structure studies [34], [162].

4.5 Summary and Conclusion

We have carried out a systematic and comprehensive analysis of the intercalation binding process of doxorubicin (DOX) to dsDNA in terms of the two-stage binding process,

the sequence selectivity, and the physical components of molecular interactions in different environments at a range of ionic concentrations. The binding free energy (BFE) of DOX with sixteen different tetradecamer DNA sequences has been quantified and decomposed into energetic components by using MD simulation with the modified MM-GBSA method at a microsecond time range. The complete energetic profile of this complexation mechanism has been characterized at each binding stage, containing the total BFE (ΔG_{calc}) and its energetic components, based on different decomposition analyses. Additionally, the per-residue BFE decomposition probes the local dominant interactions between DOX and all BPs of dsDNA to detect the best model describing the relationship between the binding affinity and selectivity, as well as to identify the actual size of the DOX binding site. Our study draws the following solid conclusions:

i) The DOX selective sequence is strongly dependent on the base pairs located downstream from the site of drug intercalation. The primary source of this selectivity is the electrostatic interactions between the positively charged amino group of DOX and the charges of these base pairs. The van der Waals interaction also contributes to this selectivity. Both **$(TA)_2$** and **$(AT)_2$** downstream base pairs are preferred.

ii) The per-residue BFE decomposition implies that the quartet sequence model is the best and most accurate model to reflect the relationship between DOX selectivity and its binding affinity.

iii) The size of the DOX binding site is 4 base pairs. Our analysis introduces the specific energy values of the local interactions between DOX and BPs of dsDNA and has not been previously reported. The BP below the DOX intercalation site (BP8) and the nearest BPs (BP9 and BP10)

downstream from the DOX site contribute approximately 34% of the total BFE and the interaction with the DOX molecule contributes about 66%. This local interaction of DOX is sequence-dependent. There is a -2.5 kcal/mol preferred energy difference between the DOX contributions in s3 or s4 compared to those in s1 or s2.

iv) Parsing total BFE in terms of a two-stage binding process implies the energetic conformational penalty of $\Delta G_{Conf} = 15.87$ kcal/mol on average, necessary to form a DOX binding site in DNA, that needs to be overcome by the energetically favorable exothermic process of the insertion binding stage of $\Delta G_{Ins} =$ -28.47 kcal/mol on average. Our results predict that the preferred binding of DOX occurs during the insertion.

v) There is a competition between the ΔG_{vac} and ΔG_{sol} at both stages of the binding process, where ΔG_{vac} tries to stabilize the process, but it requires being large enough to overcome the ΔG_{sol} which attempts to oppose the process. Our findings point out that this binding process prefers to occur at a lower ionic solution.

CHAPTER 5. FINAL REMARKS AND FUTURE WORK

We performed extensive and systematic MD simulations in combination with MM-PB(GB)SA binding free energy methods to identify and investigate (1) the complete binding profile including the main driving force for the intercalation process between DOX and DNA at both binding stages (conformational changes and insertion binding stages), (2) the DOX sequence selectivity, the comprehensive model to define DOX selectivity and to determine its binding site-size, and the impact of the nearest base-pair on the DOX selectivity, and (3) the impact of the solvent and ionic strength effects on this binding process.

A new procedure has been introduced to design a longer DNA sequence in which DOX can be inserted as desired between the base pairs of the DNA. This approach is very useful for generating a large number of DOX-DNA complexes with new DNA sequences that have never been seen before in X-ray data. Additionally, it can design a more accurate molecular model without the end-effect of fraying and breathing in the terminal base pairs of DNA that happens when the molecular model is constructed based on available X-ray structures.

Our deep-level computationally based analysis of the DOX intercalation process offers an alternative route to the pure experimental approach for the elucidation of molecular mechanisms involved in the drug-DNA binding process as well as allowing to detection of their hidden states. This study provides the missing link to answering many conflicting questions regarding the molecular mechanism and selectivity of DOX that have been bothering experimental and computational scientists for decades. Additionally, the analysis of the BFE and its energetic components is unprecedented because it introduces valuable information

about the balance of driving forces in this biomolecular process that could be used in the development of new sequence-specific anticancer therapeutic drugs.

Much work remains to be done especially in these three directions. First, it may be possible to expand this study for designing doxorubicin analogs. The goal here would be to add structural features on DOX that would either make new sequence-selectivity or enhance its binding affinity or change its binding mode which may result in a new potential agent with less toxicity and side-effect. Here, the daunosamine domain of DOX should deserve special interest. Second, the binding between DOX and other potential targets such as the G-quadruplex, triple helices or nucleic acid junctions has not been explored yet and this will have to be assessed in the future. Third, we have successfully used the methodology and the ensuing analysis of biomolecular interaction presented in this study to study the interaction between spike protein of SARS-CoV-2 and ACE2 receptor for better understanding the virus infection proliferation [150]. We can expand this study to design novel drugs such as small ligands, peptides, and small proteins to target spike protein SARS-CoV-2 and block viral entry.

APPENDIX

SUPPORTING TABLES AND FIGURES

Table S1. Atom types and RESP atomic partial charge (PC) of DOX using R.E.D server.

Atom	PC	Atom	PC	Atom	PC	Atom	PC
C1	-0.0976	O13	-0.5056	C5'	0.1345	H212	0.0938
C2	-0.1718	C14	0.0789	O5'	-0.3166	H213	0.0938
C3	-0.1318	O14	-0.6771	C6'	-0.1304	H1'	0.161
C4	0.2316	C15	-0.0140	H1	0.1515	H2'	0.0776
O4	-0.2450	C16	0.0080	H2	0.1854	H2"	0.0776
C5	-0.1020	C17	0.3361	H3	0.136	H3'	0.1255
C6	0.5709	O17	-0.5944	H8	0.4155	HN31	0.3485
O6	-0.5788	C18	-0.3139	H10	0.1321	HN32	0.3485
C7	-0.1253	C19	0.6228	H112	0.0789	HN33	0.3485
C8	0.0807	O19	-0.5217	H113	0.0789	H4'	0.1058
O8	-0.4440	C20	-0.125	H12	0.499	HO4'	0.4671
C9	-0.0180	C21	-0.0755	H142	0.070	H5'	0.0798
C10	0.0403	C1'	0.0867	H143	0.070	H6'1	0.0634
O10	-0.3547	C2'	-0.0591	HO14	0.4496	H6'2	0.0634
C11	-0.1272	C3'	0.0644	H152	0.0811	H6'3	0.0634
C12	0.1239	N3'	-0.4467	H153	0.0811		
O12	-0.7507	C4'	0.0355	H17	0.4733		
C13	0.6270	O4'	-0.6283	H211	0.0938		

Table S2. Backbone and glycosyl torsion angles and sugar conformations (deg.) for free dsDNA with d(CG**AT**CG)$_2$ sequence (model M1)

Residue	α	β	γ	δ	ε	ζ	χ	P	Puckering
C1	---	---	63	121	-156	-92	-119	133	C1'-exo
G2	-80	166	51	134	-142	-141	-103	146	C2'-endo
A3	-86	161	47	137	-175	-91	-112	165	C2'-endo
T4	-68	167	58	110	-175	-86	-131	122	C1'-exo
C5	-67	170	61	120	-170	-94	-118	129	C1'-exo
G6	-73	172	55	121	---	---	-114	129	C1'-exo
C7	---	---	63	115	-156	-92	-126	123	C1'-exo
G8	-85	170	54	134	-158	-131	-107	146	C2'-endo
A9	-75	166	56	126	-178	-91	-124	128	C1'-exo
T10	-65	170	59	107	-173	-89	-128	110	C1'-exo
C11	-69	171	58	122	-167	-93	-121	130	C1'-exo
G12	-72	170	54	130	---	---	-107	142	C1'-exo
Aver.	-74	168	57	123	-165	-100	-118	134	
B-DNA	-63	171	54	123	-169	-108	-117		

Table S3. Backbone and glycosyl torsion angles and sugar conformations (deg.) for free dsDNA with d(CG**TA**CG)$_2$ sequence (model M2)

Residue	α	β	γ	δ	ε	ζ	χ	P	Puckering
C1	---	---	62	127	-155	-81	-133	137	C1'-exo
G2	-85	163	53	133	-139	-131	-101	138	C1'-exo
T3	-94	161	59	122	-169	-100	-119	133	C1'-exo
A4	-69	174	58	139	-177	-99	-102	168	C2'-endo
C5	-68	166	56	106	-159	-106	-120	103	O4'-endo
G6	-77	169	55	131	---	---	-107	146	C2'-endo
C7	---	---	58	126	-162	-81	-128	137	C1'-exo
G8	-84	173	50	138	-136	-143	-100	146	C2'-endo
T9	-84	160	54	132	-163	-96	-119	147	C2'-endo
A10	-78	169	49	139	-165	-111	-98	144	C1'-exo
C11	-78	161	51	126	-141	-97	-122	136	C1'-exo
G12	-88	162	52	126	---	---	-103	138	C1'-exo
Aver.	-74	168	57	123	-165	-100	-118	134	
B-DNA	-63	171	54	123	-169	-108	-117		

Table S4. Backbone and glycosyl torsion angles and sugar conformations (deg.) for 1:1 DOX-DNA complex with d(CG**AT**CG)$_2$ sequence (model M3)

Residue	α	β	γ	δ	ε	ζ	χ	P	Puckering
C1	---	---	63	133	-167	-78	-45	153	C2'-endo
G2	-72	169	56	137	-145	-169	-101	153	C2'-endo
A3	-80	163	50	135	-177	-95	-113	160	C2'-endo
T4	-66	170	58	114	-162	-82	-124	118	C1'-exo
C5	-76	169	53	139	-97	142	-84	149	C2'-endo
G6	-59	179	32	141	---	---	-93	170	C2'-endo
C7	---	---	35	125	-105	-70	-148	134	C1'-exo
G8	-76	162	49	127	-139	103	-86	132	C1'-exo
A9	-94	153	53	136	-175	-94	-115	167	C2'-endo
T10	-69	168	61	108	-168	-87	-128	109	C1'-exo
C11	-70	173	58	137	-175	-85	-118	153	C2'-endo
G12	-67	169	56	130	---	---	-106	132	C1'-exo
Aver.	-74	168	57	123	-165	-100	-118	134	
B-DNA	-63	171	54	123	-169	-108	-117		

Table S5. Backbone and glycosyl torsion angles and sugar conformations (deg.) for 1:1 DOX-DNA complex with d(CG**TA**CG)$_2$ sequence (model M4)

Residue	α	β	γ	δ	ε	ζ	χ	P	Puckering
C1	---	---	61	119	-151	-65	-41	143	C1'-exo
G2	-113	167	60	135	-176	-103	-100	170	C2'-endo
T3	-71	167	58	117	-172	-94	-121	104	O4'-endo
A4	-70	174	57	127	-170	-88	-108	148	C2'-endo
C5	-72	166	57	116	-101	163	-86	121	C1'-exo
G6	-65	166	41	132	---	---	-94	135	C1'-exo
C7	---	---	43	127	-109	-156	-143	131	C1'-exo
G8	-91	163	57	134	-140	120	-91	149	C2'-endo
T9	-81	151	67	129	-163	-92	-125	147	C2'-endo
A10	-78	170	57	128	-147	-133	-105	141	C1'-exo
C11	-68	162	56	124	-165	-87	-118	134	C1'-exo
G12	-75	169	56	124	---	---	-107	132	C1'-exo
Aver.	-78	166	56	126	-149	-51	-103	138	
B-DNA	-63	171	54	123	-169	-108	-117		

Table S6. Backbone and glycosyl torsion angles and sugar conformations (deg.) for 2:1 DOX-DNA complex with d(CG**AT**CG)$_2$ sequence (model M5)

Residue	α	β	γ	δ	ε	ζ	χ	P	Puckering
C1	---	---	40	128	-101	-70	-149	138	C1'-exo
G2	-75	169	48	130	-144	121	-86	135	C1'-exo
A3	-96	156	52	139	-176	-94	-116	171	C2'-endo
T4	-69	172	59	120	-163	-83	-123	128	C1'-exo
C5	-77	169	53	138	-96	164	-81	148	C2'-endo
G6	-71	178	42	139	---	---	-96	186	C3'-exo
C7	---	---	50	124	-109	-72	-146	134	C1'-exo
G8	-76	164	50	129	-140	135	-88	134	C1'-exo
A9	-94	155	53	138	-175	-94	-116	168	C2'-endo
T10	-69	172	60	119	-163	-84	-121	127	C1'-exo
C11	-76	168	54	138	-103	148	-83	148	C2'-endo
G12	-72	179	43	138	---	---	-97	180	C3'-exo
Aver.	-77	168	50	132	-137	7	-109	150	
B-DNA	-63	171	54	123	-169	-108	-117		

Table S7. Backbone and glycosyl torsion angles and sugar conformations (deg.) for 2:1 DOX-DNA complex with d(CG**TA**CG)$_2$ sequence (model M6)

Residue	α	β	γ	δ	ε	ζ	χ	P	Puckering
C1	---	---	37	130	-96	-69	-149	139	C1'-exo
G2	-74	164	46	133	-139	127	-86	138	C1'-exo
T3	-92	151	56	130	-172	-96	-119	147	C2'-endo
A4	-73	176	58	132	-166	-90	-105	153	C2'-endo
C5	-74	166	52	129	-107	180	-85	137	C1'-exo
G6	-61	177	32	139	---	---	-96	180	C3'-exo
C7	---	---	46	130	-100	-76	-146	139	C1'-exo
G8	-76	161	48	131	-143	141	-88	138	C1'-exo
T9	-87	150	59	126	-168	-101	-116	139	C1'-exo
A10	-71	173	57	133	-167	-88	-120	152	C2'-endo
C11	-74	167	53	131	-107	151	-83	138	C1'-exo
G12	-67	179	39	135	---	---	-97	178	C2'-endo
Aver.	-75	166	49	131	-137	8	-107	148	
B-DNA	-63	171	54	123	-169	-108	-117		

Table S8. Summary of Thirty-two models performed in this study. The zero or one after the dash symbol (i.e., A1-0 and A1-1) means no DOX (as in free DNA) or one DOX respectively (as in 1:1 DOX-DNA complex). The bold black letters represent the nearest BPs and the bold red letters are the BPs that form the DOX intercalation site.

Model	atoms # in DNA	atoms # in DOX	# of Water molecules	# Cl^- ion	# Na^+ ions	Total # of atoms
A1-0	880	—	4000	—	26	12906
A1-1	880	69	4000	1	26	12976
A2-0	880	—	4000	—	26	12906
A2-1	880	69	4000	1	26	12976
A3-0	884	—	4000	—	26	12910
A3-1	884	69	4000	1	26	12980
A4-0	884	—	4000	—	26	12910
A4-1	884	69	4000	1	26	12980
B1-0	881	—	4000	—	26	12907
B1-1	881	69	4000	1	26	12977
B2-0	881	—	4000	—	26	12907
B2-1	881	69	4000	1	26	12977
B3-0	885	—	4000	—	26	12911
B3-1	885	69	4000	1	26	12981
B4-0	885	—	4000	—	26	12911
B4-1	885	69	4000	1	26	12981

Table S8. Continued.

Model	atoms # in DNA	atoms # in DOX	# of Water molecules	# Cl^- ion	# Na^+ ions	Total # of atoms
C1-0	881	—	4000	—	26	12907
C1-1	881	69	4000	1	26	12977
C2-0	881	—	4000	—	26	12907
C2-1	881	69	4000	1	26	12977
C3-0	885	—	4000	—	26	12911
C3-1	885	69	4000	1	26	12981
C4-0	885	—	4000	—	26	12911
C4-1	885	69	4000	1	26	12981
D1-0	882	—	4000	—	26	12908
D1-1	882	69	4000	1	26	12978
D2-0	882	—	4000	—	26	12908
D2-1	882	69	4000	1	26	12978
D3-0	886	—	4000	—	26	12912
D3-1	886	69	4000	1	26	12982
D4-0	886	—	4000	—	26	12912
D4-1	886	69	4000	1	26	12982

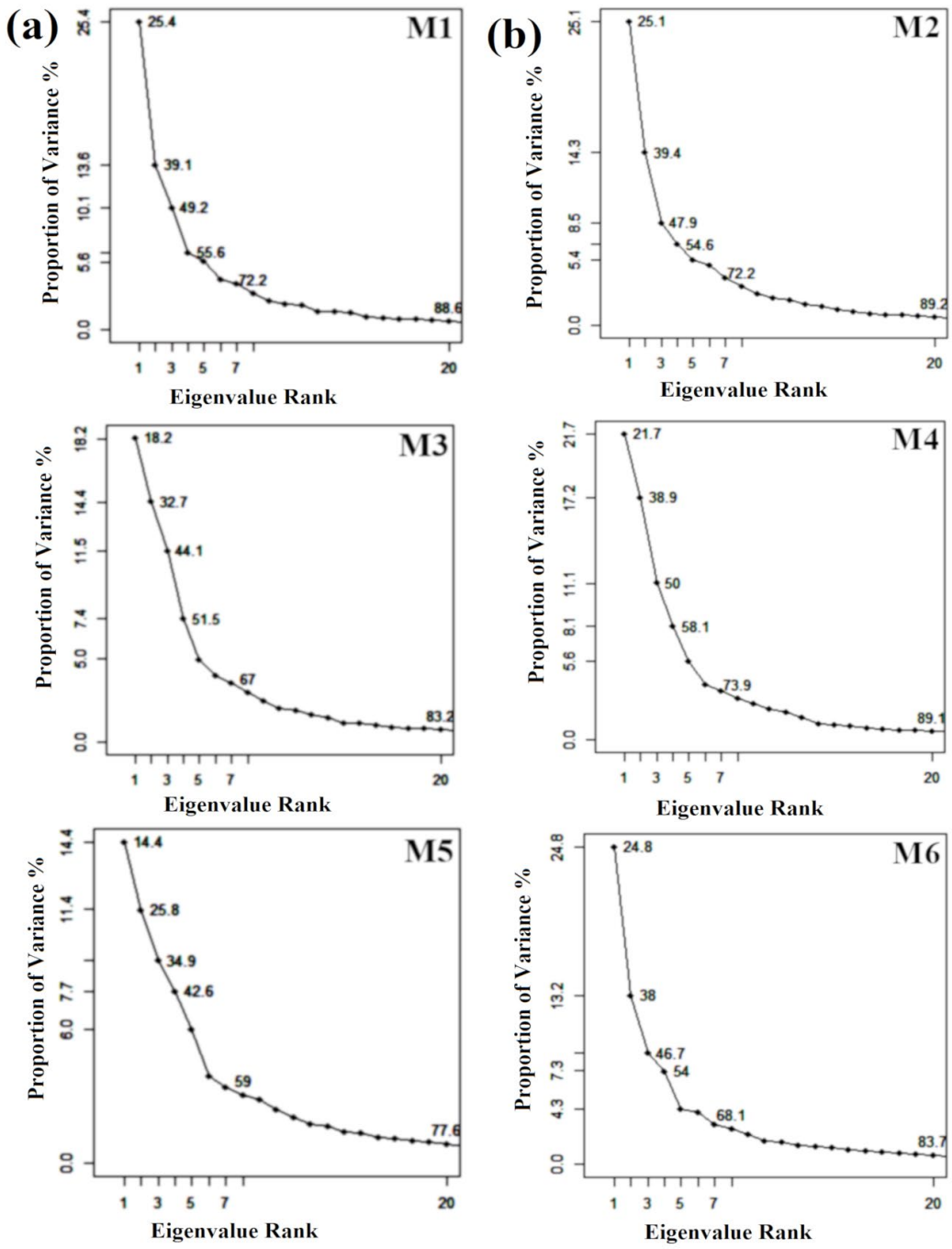


Figure S1. Scree plot for principal component analysis on the MD coordinate data of three different environments (free DNA, 1:1 complex, 2:1 complex) of (a) three situations of DNA1 sequence and (b) three situations of DNA2 sequence. The magnitude of each eigenvalue is expressed as the proportion of the total variance (mean-square fluctuation) captured by the corresponding eigenvector. Labels on each point indicate the cumulative sum of variance accounted for by a particular eigenvector and its preceding eigenvectors.

REFERENCES

[1] B. Jawad, L. Poudel, R. Podgornik, N. F. Steinmetz, and W. Y. Ching, "Molecular mechanism and binding free energy of doxorubicin intercalation in DNA," *Physical Chemistry Chemical Physics*, vol. 21, no. 7, pp. 3877–3893, 2019, doi: 10.1039/c8cp06776g.

[2] B. Jawad, L. Poudel, R. Podgornik, and W. Y. Ching, "Thermodynamic dissection of the intercalation binding process of doxorubicin to dsDNA with implications of ionic and solvent effects," *Journal of Physical Chemistry B*, vol. 124, no. 36, pp. 7803–7818, Sep. 2020, doi: 10.1021/acs.jpcb.0c05840.

[3] "Cancer Statistics - National Cancer Institute." https://www.cancer.gov/about-cancer/understanding/statistics (accessed Oct. 19, 2021).

[4] S. Hoelder, P. A. Clarke, and P. Workman, "Discovery of small molecule cancer drugs: Successes, challenges and opportunities," *Molecular Oncology*, vol. 6, no. 2. 2012. doi: 10.1016/j.molonc.2012.02.004.

[5] J. Sheng, J. Gan, and Z. Huang, "Structure-based DNA-targeting strategies with small molecule ligands for drug discovery," *Medicinal Research Reviews*, vol. 33, no. 5, 2013, doi: 10.1002/med.21278.

[6] L. H. Hurley, "DNA and its associated processes as targets for cancer therapy," *Nature Reviews Cancer*, vol. 2, no. 3. 2002. doi: 10.1038/nrc749.

[7] J. B. Chaires, "Calorimetry and thermodynamics in drug design," *Annual Review of Biophysics*, vol. 37. 2008. doi: 10.1146/annurev.biophys.36.040306.132812.

[8] A. S. Backer, A. S. Biebricher, G. A. King, G. J. L. Wuite, I. Heller, and E. J. G. Peterman, "Single-molecule polarization microscopy of DNA intercalators sheds light on the structure of S-DNA," *Science Advances*, vol. 5, no. 3, 2019, doi: 10.1126/sciadv.aav1083.

[9] M. Baginski, F. Fogolari, and J. M. Briggs, "Electrostatic and non-electrostatic contributions to the binding free energies of anthracycline antibiotics to DNA," *Journal of Molecular Biology*, vol. 274, no. 2, 1997, doi: 10.1006/jmbi.1997.1399.

[10] A. S. Biebricher, I. Heller, R. F. H. Roijmans, T. P. Hoekstra, E. J. G. Peterman, and G. J. L. Wuite, "The impact of DNA intercalators on DNA and DNA-processing enzymes elucidated through force-dependent binding kinetics," *Nature Communications*, vol. 6, 2015, doi: 10.1038/ncomms8304.

[11] Y. Miao *et al.*, "Out-of-shape DNA minor groove binders: Induced fit interactions of heterocyclic dications with the DNA minor groove," *Biochemistry*, vol. 44, no. 45, 2005, doi: 10.1021/bi051791q.

[12] C. Y. Zhang, X. Dong, J. Gao, W. Lin, Z. Liu, and Z. Wang, "Nanoparticle-induced neutrophil apoptosis increases survival in sepsis and alleviates neurological damage in stroke," *Science Advances*, vol. 5, no. 11, 2019, doi: 10.1126/sciadv.aax7964.

[13] K. Sasaki *et al.*, "Engineered collagen-binding serum albumin as a drug conjugate carrier for cancer therapy," *Science Advances*, vol. 5, no. 8, 2019, doi: 10.1126/sciadv.aaw6081.

[14] Q. Hu, H. Li, L. Wang, H. Gu, and C. Fan, "DNA nanotechnology-enabled drug delivery systems," *Chemical reviews*, vol. 119, no. 10. 2019. doi: 10.1021/acs.chemrev.7b00663.

[15] D. Amgalan *et al.*, "A small-molecule allosteric inhibitor of BAX protects against doxorubicin-induced cardiomyopathy," *Nature Cancer*, vol. 1, no. 3, 2020, doi: 10.1038/s43018-020-0039-1.

[16] Q. He, F. Wang, T. D. Ryan, M. Chalasani, and A. N. Redington, "Repeated remote ischemic conditioning reduces doxorubicin-induced cardiotoxicity," *JACC: CardioOncology*, vol. 2, no. 1, 2020, doi: 10.1016/j.jaccao.2020.01.005.

[17] Y. Liu *et al.*, "Visnagin protects against doxorubicin-induced cardiomyopathy through modulation of mitochondrial malate dehydrogenase," *Science Translational Medicine*, vol. 6, no. 266, 2014, doi: 10.1126/scitranslmed.3010189.

[18] J. B. Chaires, S. Satyanarayana, D. Suh, I. Fokt, T. Przewloka, and W. Priebe, "Parsing the free energy of anthracycline antibiotic binding to DNA," *Biochemistry*, vol. 35, no. 7, 1996, doi: 10.1021/bi952812r.

[19] J. Shu, D. Cui, Y. Ma, X. Xiong, Y. Sun, and Y. Zhao, "SCFβ-TrCP-mediated degradation of TOP2β promotes cancer cell survival in response to chemotherapeutic drugs targeting topoisomerase II," *Oncogenesis*, vol. 9, no. 2, 2020, doi: 10.1038/s41389-020-0196-1.

[20] F. Yang, S. S. Teves, C. J. Kemp, and S. Henikoff, "Doxorubicin, DNA torsion, and chromatin dynamics," *Biochimica et Biophysica Acta - Reviews on Cancer*, vol. 1845, no. 1. 2014. doi: 10.1016/j.bbcan.2013.12.002.

[21] J. L. Nitiss, "Targeting DNA topoisomerase II in cancer chemotherapy," *Nature Reviews Cancer*, vol. 9, no. 5. 2009. doi: 10.1038/nrc2607.

[22] K. M. Tewey, T. C. Rowe, L. Yang, B. D. Halligan, and L. F. Liu, "Adriamycin-induced DNA damage mediated by mammalian DNA topoisomerase II," *Science*, vol. 226, no. 4673, 1984, doi: 10.1126/science.6093249.

[23] V. G. S. Box, "The intercalation of DNA double helices with doxorubicin and nagalomycin," *Journal of Molecular Graphics and Modelling*, vol. 26, no. 1, 2007, doi: 10.1016/j.jmgm.2006.09.005.

[24] G. Minotti, P. Menna, E. Salvatorelli, G. Cairo, and L. Gianni, "Anthracyclines: Molecular advances and pharmacologie developments in antitumor activity and cardiotoxicity," *Pharmacological Reviews*, vol. 56, no. 2. 2004. doi: 10.1124/pr.56.2.6.

[25] A. Senchenkov, D. A. Litvak, and M. C. Cabot, "Targeting ceramide metabolism - A strategy for overcoming drug resistance," *Journal of the National Cancer Institute*, vol. 93, no. 5. 2001. doi: 10.1093/jnci/93.5.347.

[26] C. F. Thorn *et al.*, "Doxorubicin pathways: Pharmacodynamics and adverse effects," *Pharmacogenetics and Genomics*, vol. 21, no. 7, 2011, doi: 10.1097/FPC.0b013e32833ffb56.

[27] M. Airoldi, G. Barone, G. Gennaro, A. M. Giuliani, and M. Giustini, "Interaction of doxorubicin with polynucleotides. a spectroscopic study," *Biochemistry*, vol. 53, no. 13, 2014, doi: 10.1021/bi401687v.

[28] D. J. Cashman and G. E. Kellogg, “A computational model for anthracycline binding to DNA: Tuning groove-binding intercalators for specific sequences,” *Journal of Medicinal Chemistry*, vol. 47, no. 6, 2004, doi: 10.1021/jm030529h.

[29] J. B. Chaires, K. R. Fox, J. E. Herrera, M. Britt, and M. J. Waring, “Site and sequence specificity of the daunomycin-DNA interaction,” *Biochemistry*, vol. 26, no. 25, 1987, doi: 10.1021/bi00399a031.

[30] J. B. Chaires, J. E. Herrera, and M. J. Waring, “Preferential binding of daunomycin to 5′TACG and 5′TAGC sequences revealed by footprinting titration experiments,” *Biochemistry*, vol. 29, no. 26, 1990, doi: 10.1021/bi00478a006.

[31] K. X. Chen, N. Gresh, and B. Pullman, “A theoretical investigation on the sequence selective binding of adriamycin to double-stranded polynucleotides,” *Nucleic Acids Research*, vol. 14, no. 5, 1986, doi: 10.1093/nar/14.5.2251.

[32] K. X. Chen, N. Gresh, and B. Pullman, “A theoretical investigation on the sequence selective binding of daunomycin to double-stranded polynucleotides,” *Journal of Biomolecular Structure and Dynamics*, vol. 3, no. 3, 1985, doi: 10.1080/07391102.1985.10508434.

[33] K. X. Chen, N. Gresh, and B. Pullman, “A theoretical study of the comparative binding affinities of daunomycin derivatives to a double-stranded oligomeric DNA. Proposal for new high affinity derivatives,” *Molecular Pharmacology*, vol. 30, no. 3, 1986.

[34] C. A. Frederick *et al.*, “Structural comparison of anticancer drug-DNA complexes: Adriamycin and daunomycin,” *Biochemistry*, vol. 29, no. 10, 1990, doi: 10.1021/bi00462a016.

[35] A. H. Wang, G. Ughetto, G. J. Quigley, and A. Rich, “Interactions between an anthracycline antibiotic and DNA: Molecular structure of daunomycin complexed to d(CpGpTpApCpG) at 1.2-Å resolution,” *Biochemistry*, vol. 26, no. 4, 1987, doi: 10.1021/bi00378a025.

[36] A. Skorobogaty, R. J. White, D. R. Phillips, and J. A. Reiss, “Elucidation of the DNA sequence preferences of daunomycin,” *Drug Design and Delivery*, vol. 3, no. 2 SPEC. ISS., 1988.

[37] A. Skorobogaty, R. J. White, D. R. Phillips, and J. A. Reiss, “The 5′-CA DNA-sequence preference of daunomycin,” *FEBS Letters*, vol. 227, no. 2, 1988, doi: 10.1016/0014-5793(88)80877-9.

[38] D. R. Phillips, C. Cullinane, H. Trist, and R. J. White, “In vitro transcription analysis of the sequence specificity of reversible and irreversible complexes of adriamycin with DNA,” 1990. doi: 10.1007/978-94-011-3728-7_10.

[39] D. J. Cashman, J. N. Scarsdale, and G. E. Kellogg, “Hydropathic analysis of the free energy differences in anthracycline antibiotic binding to DNA,” *Nucleic Acids Research*, vol. 31, no. 15, 2003, doi: 10.1093/nar/gkg645.

[40] G. E. Kellogg, J. N. Scarsdale, and F. A. Fornari, “Identification and hydropathic characterization of structural features affecting sequence specificity for doxorubicin intercalation into DNA double-stranded polynucleotides,” *Nucleic Acids Research*, vol. 26, no. 20, 1998, doi: 10.1093/nar/26.20.4721.

[41] J. B. Chaires, “Energetics of drug-DNA interactions,” *Biopolymers*, vol. 44, no. 3, 1997, doi: 10.1002/(SICI)1097-0282(1997)44:3<201::AID-BIP2>3.0.CO;2-Z.

[42] S. A. Adcock and J. A. McCammon, “Molecular dynamics: Survey of methods for simulating the activity of proteins,” *Chemical Reviews*, vol. 106, no. 5. 2006. doi: 10.1021/cr040426m.

[43] S. A. Hollingsworth and R. O. Dror, “Molecular dynamics simulation for all,” *Neuron*, vol. 99, no. 6. 2018. doi: 10.1016/j.neuron.2018.08.011.

[44] M. de Vivo, M. Masetti, G. Bottegoni, and A. Cavalli, “Role of molecular dynamics and related methods in drug discovery,” *Journal of Medicinal Chemistry*, vol. 59, no. 9. 2016. doi: 10.1021/acs.jmedchem.5b01684.

[45] D. L. Mobley and M. K. Gilson, “Predicting binding free energies: Frontiers and benchmarks,” *Annual Review of Biophysics*, vol. 46. 2017. doi: 10.1146/annurev-biophys-070816-033654.

[46] R. Anandakrishnan, A. Drozdetski, R. C. Walker, and A. v. Onufriev, “Speed of conformational change: Comparing explicit and implicit solvent molecular dynamics simulations,” *Biophysical Journal*, vol. 108, no. 5, 2015, doi: 10.1016/j.bpj.2014.12.047.

[47] S. Kirmizialtin, V. Nguyen, K. A. Johnson, and R. Elber, “How conformational dynamics of DNA polymerase select correct substrates: Experiments and simulations,” *Structure*, vol. 20, no. 4, 2012, doi: 10.1016/j.str.2012.02.018.

[48] H. Guterres and W. Im, “Improving protein-ligand docking results with high-throughput molecular dynamics simulations,” *Journal of Chemical Information and Modeling*, vol. 60, no. 4, 2020, doi: 10.1021/acs.jcim.0c00057.

[49] X. Liu, D. Shi, S. Zhou, H. Liu, H. Liu, and X. Yao, “Molecular dynamics simulations and novel drug discovery,” *Expert Opinion on Drug Discovery*, vol. 13, no. 1. 2018. doi: 10.1080/17460441.2018.1403419.

[50] J. Chen, X. Wang, L. Pang, J. Z. H. Zhang, and T. Zhu, “Effect of mutations on binding of ligands to guanine riboswitch probed by free energy perturbation and molecular dynamics simulations,” *Nucleic Acids Research*, vol. 47, no. 13, 2019, doi: 10.1093/nar/gkz499.

[51] X. Lin, X. Li, and X. Lin, “A review on applications of computational methods in drug screening and design,” *Molecules*, vol. 25, no. 6. 2020. doi: 10.3390/molecules25061375.

[52] M. Araki *et al.*, “Exploring ligand binding pathways on proteins using hypersound-accelerated molecular dynamics,” *Nature Communications*, vol. 12, no. 1, 2021, doi: 10.1038/s41467-021-23157-1.

[53] P. Mokaberi, F. Babayan-Mashhadi, Z. Amiri Tehrani Zadeh, M. R. Saberi, and J. Chamani, “Analysis of the interaction behavior between Nano-Curcumin and two human serum proteins: combining spectroscopy and molecular stimulation to understand protein-protein interaction,” *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 9, 2021, doi: 10.1080/07391102.2020.1766570.

[54] A. Jost Lopez, P. K. Quoika, M. Linke, G. Hummer, and J. Köfinger, "Quantifying protein-protein interactions in molecular simulations," *Journal of Physical Chemistry B*, vol. 124, no. 23, 2020, doi: 10.1021/acs.jpcb.9b11802.

[55] A. C. Pan, D. Jacobson, K. Yatsenko, D. Sritharan, T. M. Weinreich, and D. E. Shaw, "Atomic-level characterization of protein–protein association," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 10, 2019, doi: 10.1073/pnas.1815431116.

[56] L. Dai, Y. Xu, Z. Du, X. D. Su, and J. Yu, "Revealing atomic-scale molecular diffusion of a plant-transcription factor WRKY domain protein along DNA," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 118, no. 23, 2021, doi: 10.1073/pnas.2102621118.

[57] A. Pérez, F. J. Luque, and M. Orozco, "Frontiers in molecular dynamics simulations of DNA.," *Accounts of chemical research*, vol. 45, no. 2. 2012. doi: 10.1021/ar2001217.

[58] S. Bottaro and K. Lindorff-Larsen, "Biophysical experiments and biomolecular simulations: A perfect match?," *Science*, vol. 361, no. 6400. 2018. doi: 10.1126/science.aat4010.

[59] S. Piana, K. Lindorff-Larsen, and D. E. Shaw, "How robust are protein folding simulations with respect to force field parameterization?," *Biophysical Journal*, vol. 100, no. 9, 2011, doi: 10.1016/j.bpj.2011.03.051.

[60] L. E. Ratcliff, S. Mohr, G. Huhs, T. Deutsch, M. Masella, and L. Genovese, "Challenges in large scale quantum mechanical calculations," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 7, no. 1. 2017. doi: 10.1002/wcms.1290.

[61] D. A. Pearlman *et al.*, "AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules," *Computer Physics Communications*, vol. 91, no. 1–3, 1995, doi: 10.1016/0010-4655(95)00041-D.

[62] D. A. Case *et al.*, "The Amber biomolecular simulation programs," *Journal of Computational Chemistry*, vol. 26, no. 16. 2005. doi: 10.1002/jcc.20290.

[63] B. R. Brooks *et al.*, "CHARMM: The biomolecular simulation program," *Journal of Computational Chemistry*, vol. 30, no. 10, 2009, doi: 10.1002/jcc.21287.

[64] L. Kalé *et al.*, "NAMD2: Greater scalability for parallel molecular dynamics," *Journal of Computational Physics*, vol. 151, no. 1, 1999, doi: 10.1006/jcph.1999.6201.

[65] M. J. Abraham *et al.*, "Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, vol. 1–2, 2015, doi: 10.1016/j.softx.2015.06.001.

[66] S. Miyamoto and P. A. Kollman, "Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models," *Journal of Computational Chemistry*, vol. 13, no. 8, 1992, doi: 10.1002/jcc.540130805.

[67] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-

alkanes," *Journal of Computational Physics*, vol. 23, no. 3, 1977, doi: 10.1016/0021-9991(77)90098-5.

[68] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, "LINCS: A linear constraint solver for molecular simulations," *Journal of Computational Chemistry*, vol. 18, no. 12, 1997, doi: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

[69] L. Verlet, "Computer 'experiments' on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules," *Physical Review*, vol. 159, no. 1, 1967, doi: 10.1103/PhysRev.159.98.

[70] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids: Second edition*. 2017. doi: 10.1093/oso/9780198803195.001.0001.

[71] C. Tian *et al.*, "Ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution," *Journal of Chemical Theory and Computation*, vol. 16, no. 1, 2020, doi: 10.1021/acs.jctc.9b00591.

[72] R. Galindo-Murillo *et al.*, "Assessing the current state of AMBER force field modifications for DNA," *Journal of Chemical Theory and Computation*, vol. 12, no. 8, 2016, doi: 10.1021/acs.jctc.6b00186.

[73] A. Pérez *et al.*, "Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers," *Biophysical Journal*, vol. 92, no. 11, 2007, doi: 10.1529/biophysj.106.097782.

[74] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general Amber force field," *Journal of Computational Chemistry*, vol. 25, no. 9, 2004, doi: 10.1002/jcc.20035.

[75] R. B. Best *et al.*, "Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ1 and χ2 Dihedral Angles," *Journal of Chemical Theory and Computation*, vol. 8, no. 9, 2012, doi: 10.1021/ct300400x.

[76] K. Vanommeslaeghe *et al.*, "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields," *Journal of Computational Chemistry*, vol. 31, no. 4, 2010, doi: 10.1002/jcc.21367.

[77] A. D. MacKerell *et al.*, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," *Journal of Physical Chemistry B*, vol. 102, no. 18, 1998, doi: 10.1021/jp973084f.

[78] K. Hart, N. Foloppe, C. M. Baker, E. J. Denning, L. Nilsson, and A. D. MacKerell, "Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium," *Journal of Chemical Theory and Computation*, vol. 8, no. 1, 2012, doi: 10.1021/ct200723y.

[79] M. J. Robertson, J. Tirado-Rives, and W. L. Jorgensen, "Improved peptide and protein torsional energetics with the OPLS-AA force field," *Journal of Chemical Theory and Computation*, vol. 11, no. 7, 2015, doi: 10.1021/acs.jctc.5b00356.

[80] E. Harder *et al.*, "OPLS3: A force field providing broad coverage of drug-like small molecules and proteins," *Journal of Chemical Theory and Computation*, vol. 12, no. 1, 2016, doi: 10.1021/acs.jctc.5b00864.

[81] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids," *Journal of the American Chemical Society*, vol. 118, no. 45, 1996, doi: 10.1021/ja9621760.

[82] C. I. Bayly *et al.*, "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules," *Journal of the American Chemical Society*, vol. 117, no. 19, 1995, doi: 10.1021/ja00124a002.

[83] J. E. Lennard-Jones, "On the determination of molecular fields. —II. From the equation of state of a gas," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 106, no. 738, 1924, doi: 10.1098/rspa.1924.0082.

[84] F. E. Boas and P. B. Harbury, "Potential energy functions for protein design," *Current Opinion in Structural Biology*, vol. 17, no. 2. 2007. doi: 10.1016/j.sbi.2007.03.006.

[85] P. Ren and J. W. Ponder, "Polarizable atomic multipole water model for molecular mechanics simulation," *Journal of Physical Chemistry B*, vol. 107, no. 24, 2003, doi: 10.1021/jp027815+.

[86] J. W. Ponder *et al.*, "Current status of the AMOEBA polarizable force field," *Journal of Physical Chemistry B*, vol. 114, no. 8, 2010, doi: 10.1021/jp910674d.

[87] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, "ReaxFF: A reactive force field for hydrocarbons," *Journal of Physical Chemistry A*, vol. 105, no. 41, 2001, doi: 10.1021/jp004368u.

[88] K. Chenoweth, A. C. T. van Duin, and W. A. Goddard, "ReaxFF reactive force field for molecular dynamics simulations of hydrocarbon oxidation," *Journal of Physical Chemistry A*, vol. 112, no. 5, 2008, doi: 10.1021/jp709896w.

[89] Y. Levy and J. N. Onuchic, "Water and proteins: A love-hate relationship," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 10. 2004. doi: 10.1073/pnas.0400157101.

[90] M. C. Bellissent-Funel *et al.*, "Water determines the structure and dynamics of proteins," *Chemical Reviews*, vol. 116, no. 13. 2016. doi: 10.1021/acs.chemrev.5b00664.

[91] P. Ren *et al.*, "Biomolecular electrostatics and solvation: A computational perspective," *Quarterly Reviews of Biophysics*, vol. 45, no. 4. 2012. doi: 10.1017/S003358351200011X.

[92] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *The Journal of Chemical Physics*, vol. 79, no. 2, 1983, doi: 10.1063/1.445869.

[93] M. W. Mahoney and W. L. Jorgensen, "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions," *Journal of Chemical Physics*, vol. 112, no. 20, 2000, doi: 10.1063/1.481505.

[94] S. Izadi, R. Anandakrishnan, and A. v. Onufriev, "Building water models: A different approach," *Journal of Physical Chemistry Letters*, vol. 5, no. 21, 2014, doi: 10.1021/jz501780a.

[95] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, “The missing term in effective pair potentials,” *Journal of Physical Chemistry*, vol. 91, no. 24, 1987, doi: 10.1021/j100308a038.

[96] P. Florová, P. Sklenovský, P. Banáš, and M. Otyepka, “Explicit water models affect the specific solvation and dynamics of unfolded peptides while the conformational behavior and flexibility of folded peptides remain intact,” *Journal of Chemical Theory and Computation*, vol. 6, no. 11, 2010, doi: 10.1021/ct1003687.

[97] A. v. Onufriev and S. Izadi, “Water models for biomolecular simulations,” *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 8, no. 2. 2018. doi: 10.1002/wcms.1347.

[98] I. S. Joung and T. E. Cheatham, “Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations,” *Journal of Physical Chemistry B*, vol. 112, no. 30, 2008, doi: 10.1021/jp8001614.

[99] E. Braun *et al.*, “Best practices for foundations in molecular simulations [Article v1.0],” *Living Journal of Computational Molecular Science*, vol. 1, no. 1, 2019, doi: 10.33011/livecoms.1.1.5957.

[100] D. Frenkel and B. Smit, *Understanding molecular simulation: From algorithms to applications*. 1996. doi: 10.1063/1.881812.

[101] T. Darden, D. York, and L. Pedersen, “Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems,” *The Journal of Chemical Physics*, vol. 98, no. 12, pp. 10089–10092, 1993, doi: 10.1063/1.464397.

[102] M. Feig and C. L. Brooks, “Recent advances in the development and application of implicit solvent models in biomolecule simulations,” *Current Opinion in Structural Biology*, vol. 14, no. 2. 2004. doi: 10.1016/j.sbi.2004.03.009.

[103] J. Zhang, H. Zhang, T. Wu, Q. Wang, and D. van der Spoel, “Comparison of implicit and explicit solvent models for the calculation of solvation free energy in organic solvents,” *Journal of Chemical Theory and Computation*, vol. 13, no. 3, 2017, doi: 10.1021/acs.jctc.7b00169.

[104] J. Chen, C. L. Brooks, and J. Khandogin, “Recent advances in implicit solvent-based methods for biomolecular simulations,” *Current Opinion in Structural Biology*, vol. 18, no. 2. 2008. doi: 10.1016/j.sbi.2008.01.003.

[105] G. A. Cisneros *et al.*, “Modeling molecular interactions in water: From pairwise to many-body potential energy functions,” *Chemical Reviews*, vol. 116, no. 13. 2016. doi: 10.1021/acs.chemrev.5b00644.

[106] N. A. Baker, “Poisson-Boltzmann methods for biomolecular electrostatics,” *Methods in Enzymology*, vol. 383, 2004, doi: 10.1016/S0076-6879(04)83005-2.

[107] J. Srinivasan, M. W. Trevathan, P. Beroza, and D. A. Case, “Application of a pairwise generalized Born model to proteins and nucleic acids: Inclusion of salt effects,” *Theoretical Chemistry Accounts*, vol. 101, no. 6, 1999, doi: 10.1007/s002140050460.

[108] J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman, and D. A. Case, “Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices,” *Journal of the American Chemical Society*, vol. 120, no. 37, 1998, doi: 10.1021/ja981844+.

[109] E. Wang *et al.*, "End-point binding free energy calculation with MM/PBSA and MM/GBSA: Strategies and applications in drug design," *Chemical Reviews*, vol. 119, no. 16. 2019. doi: 10.1021/acs.chemrev.9b00055.
[110] H. X. Zhou and X. Pang, "Electrostatic interactions in protein structure, folding, binding, and condensation," *Chemical Reviews*, vol. 118, no. 4. 2018. doi: 10.1021/acs.chemrev.7b00305.
[111] S. Decherchi, M. Masetti, I. Vyalov, and W. Rocchia, "Implicit solvent methods for free energy estimation," *European Journal of Medicinal Chemistry*, vol. 91. 2015. doi: 10.1016/j.ejmech.2014.08.064.
[112] M. L. Connolly, "Solvent-accessible surfaces of proteins and nucleic acids," *Science*, vol. 221, no. 4612. 1983. doi: 10.1126/science.6879170.
[113] J. Weiser, P. S. Shenkin, and W. C. Still, "Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO)," *Journal of Computational Chemistry*, vol. 20, no. 2, 1999, doi: 10.1002/(SICI)1096-987X(19990130)20:2<217::AID-JCC4>3.0.CO;2-A.
[114] M. Orozco and F. J. Luque, "Theoretical methods for the description of the solvent effect in biomolecular systems," *Chemical Reviews*, vol. 100, no. 11, 2000, doi: 10.1021/cr990052a.
[115] J. G. Kirkwood, "Theory of solutions of molecules containing widely separated charges with special application to zwitterions," *The Journal of Chemical Physics*, vol. 2, no. 7, 1934, doi: 10.1063/1.1749489.
[116] C. Li, L. Li, M. Petukh, and E. Alexov, "Progress in developing Poisson-Boltzmann equation solvers," *Computational and Mathematical Biophysics*, vol. 1, 2013, doi: 10.2478/mlbmb-2013-0002.
[117] T. Simonson, "Electrostatics and dynamics of proteins," *Reports on Progress in Physics*, vol. 66, no. 5, 2003, doi: 10.1088/0034-4885/66/5/202.
[118] J. Wang, C. Tan, Y. H. Tan, Q. Lu, and R. Luo, "Poisson-Boltzmann solvents in molecular dynamics simulations," *Communications in Computational Physics*, vol. 3, no. 5. 2008.
[119] L. Li *et al.*, "DelPhi: a comprehensive suite for DelPhi software and associated resources," *BMC Biophysics*, vol. 5, no. 1, 2012, doi: 10.1186/2046-1682-5-9.
[120] N. v. Prabhu, M. Panda, Q. Yang, and K. A. Sharp, "Explicit ion, implicit water solvation for molecular dynamics of nucleic acids and highly charged molecules," *Journal of Computational Chemistry*, vol. 29, no. 7, 2008, doi: 10.1002/jcc.20874.
[121] A. v. Onufriev and D. A. Case, "Generalized Born implicit solvent models for biomolecules," *Annual Review of Biophysics*, vol. 48. 2019. doi: 10.1146/annurev-biophys-052118-115325.
[122] A. Onufriev, D. Bashford, and D. A. Case, "Exploring protein native states and large-scale conformational changes with a modified Generalized Born model," *Proteins: Structure, Function and Genetics*, vol. 55, no. 2, 2004, doi: 10.1002/prot.20033.
[123] D. A. Case Ross C Walker and T. E. Darden Junmei Wang Robert Duke, "Amber 2019 reference manual principal contributors to the current codes." [Online]. Available: http://ambermd.org/contributors.html

[124] W. Im, M. S. Lee, and C. L. Brooks, "Generalized Born model with a simple smoothing function," *Journal of Computational Chemistry*, vol. 24, no. 14, 2003, doi: 10.1002/jcc.10321.
[125] G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, "Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium," *Journal of Physical Chemistry*, vol. 100, no. 51, 1996, doi: 10.1021/jp961710n.
[126] V. Tsui and D. A. Case, "Theory and applications of the Generalized Born solvation model in macromolecular simulations," *Biopolymers*, vol. 56, no. 4, 2000, doi: 10.1002/1097-0282(2000)56:4<275::AID-BIP10024>3.0.CO;2-E.
[127] A. Onufriev, D. Bashford, and D. A. Case, "Modification of the generalized born model suitable for macromolecules," *Journal of Physical Chemistry B*, vol. 104, no. 15, 2000, doi: 10.1021/jp994072s.
[128] J. Mongan, C. Simmerling, J. A. McCammon, D. A. Case, and A. Onufriev, "Generalized born model with a simple, robust molecular volume correction," *Journal of Chemical Theory and Computation*, vol. 3, no. 1, 2007, doi: 10.1021/ct600085e.
[129] P. Kollman, "Free energy calculations: Applications to chemical and biochemical phenomena," *Chemical Reviews*, vol. 93, no. 7, 1993, doi: 10.1021/cr00023a004.
[130] S. Decherchi and A. Cavalli, "Thermodynamics and Kinetics of Drug-Target Binding by Molecular Simulation," *Chemical Reviews*, vol. 120, no. 23. 2020. doi: 10.1021/acs.chemrev.0c00534.
[131] A. J. Clark *et al.*, "Free energy perturbation calculation of relative binding free energy between broadly neutralizing antibodies and the gp120 glycoprotein of HIV-1," *Journal of Molecular Biology*, vol. 429, no. 7, 2017, doi: 10.1016/j.jmb.2016.11.021.
[132] M. Lawrenz, R. Baron, and J. A. McCammon, "Independent-trajectories thermodynamic-integration free-energy changes for biomolecular systems: determinants of H5N1 avian influenza virus neuraminidase inhibition by peramivir," *Journal of Chemical Theory and Computation*, vol. 5, no. 4, 2009, doi: 10.1021/ct800559d.
[133] J. Michel, N. Foloppe, and J. W. Essex, "Rigorous free energy calculations in structure-based drug design," *Molecular Informatics*, vol. 29, no. 8–9, 2010, doi: 10.1002/minf.201000051.
[134] R. E. Skyner, J. L. McDonagh, C. R. Groom, T. van Mourik, and J. B. O. Mitchell, "A review of methods for the calculation of solution free energies and the modelling of systems in solution," *Physical Chemistry Chemical Physics*, vol. 17, no. 9. 2015. doi: 10.1039/c5cp00288e.
[135] N. Homeyer and H. Gohlke, "Free energy calculations by the molecular mechanics Poisson-Boltzmann surface area method," *Molecular Informatics*, vol. 31, no. 2, 2012, doi: 10.1002/minf.201100135.
[136] S. Sirin *et al.*, "A computational approach to enzyme design: Predicting W-Aminotransferase catalytic activity using docking and MM-GBSA scoring," *Journal of Chemical Information and Modeling*, vol. 54, no. 8, 2014, doi: 10.1021/ci5002185.

[137] F. Chen *et al.*, "Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein-protein binding free energies and re-rank binding poses generated by protein-protein docking," *Physical Chemistry Chemical Physics*, vol. 18, no. 32, 2016, doi: 10.1039/c6cp03670h.

[138] G. Weng, E. Wang, F. Chen, H. Sun, Z. Wang, and T. Hou, "Assessing the performance of MM/PBSA and MM/GBSA methods. 9. Prediction reliability of binding affinities and binding poses for protein-peptide complexes," *Physical Chemistry Chemical Physics*, vol. 21, no. 19, 2019, doi: 10.1039/c9cp01674k.

[139] G. Weng *et al.*, "HawkDock: a web server to predict and analyze the protein-protein complex based on computational docking and MM/GBSA," *Nucleic Acids Research*, vol. 47, no. W1, 2019, doi: 10.1093/nar/gkz397.

[140] K. Huang, S. Luo, Y. Cong, S. Zhong, J. Z. H. Zhang, and L. Duan, "An accurate free energy estimator: Based on MM/PBSA combined with interaction entropy for protein-ligand binding affinity," *Nanoscale*, vol. 12, no. 19, 2020, doi: 10.1039/c9nr10638c.

[141] G. Poli, C. Granchi, F. Rizzolio, and T. Tuccinardi, "Application of MM-PBSA methods in virtual screening," *Molecules*, vol. 25, no. 8. 2020. doi: 10.3390/molecules25081971.

[142] X. Wang, H. Zhang, and W. Li, "DNA-binding mechanisms of human and mouse cGAS: A comparative MD and MM/GBSA study," *Physical Chemistry Chemical Physics*, vol. 22, no. 45, 2020, doi: 10.1039/d0cp04162a.

[143] A. Soni, P. Khurana, T. Singh, and B. Jayaram, "A DNA intercalation methodology for an efficient prediction of ligand binding pose and energetics," *Bioinformatics*, vol. 33, no. 10, 2017, doi: 10.1093/bioinformatics/btx006.

[144] A. Kamal *et al.*, "Synthesis, DNA-binding ability and anticancer activity of benzothiazole/benzoxazole-pyrrolo[2,1-c][1,4]benzodiazepine conjugates," *Bioorganic and Medicinal Chemistry*, vol. 18, no. 13, 2010, doi: 10.1016/j.bmc.2010.05.007.

[145] W. Treesuwan *et al.*, "A detailed binding free energy study of 2: 1 ligand-DNA complex formation by experiment and simulation," *Physical Chemistry Chemical Physics*, vol. 11, no. 45, 2009, doi: 10.1039/b910574c.

[146] A. v. Vargiu, P. Ruggerone, T. J. Opperman, S. T. Nguyen, and H. Nikaido, "Molecular mechanism of MBX2319 inhibition of Escherichia coli AcrB multidrug efflux pump and comparison with other inhibitors," *Antimicrobial Agents and Chemotherapy*, vol. 58, no. 10, 2014, doi: 10.1128/AAC.03283-14.

[147] S. Panneerselvam, D. Yesudhas, P. Durai, M. A. Anwar, V. Gosu, and S. Choi, "A combined molecular docking/dynamics approach to probe the binding mode of cancer drugs with cytochrome P450 3A4," *Molecules*, vol. 20, no. 8, 2015, doi: 10.3390/molecules200814915.

[148] M. L. Contreras, C. Torres, I. Villarroel, and R. Rozas, "Molecular dynamics assessment of doxorubicin–carbon nanotubes molecular interactions for the design of drug delivery systems," *Structural Chemistry*, vol. 30, no. 1, 2019, doi: 10.1007/s11224-018-1210-5.

[149] K. R. Karnati and Y. Wang, “Understanding the co-loading and releasing of doxorubicin and paclitaxel using chitosan functionalized single-walled carbon nanotubes by molecular dynamics simulations,” *Physical Chemistry Chemical Physics*, vol. 20, no. 14, 2018, doi: 10.1039/c8cp00124c.

[150] B. Jawad, P. Adhikari, R. Podgornik, and W.-Y. Ching, “Key interacting residues between RBD of SARS-CoV-2 and ACE2 receptor: Combination of molecular dynamics simulation and density functional calculation,” *Journal of Chemical Information and Modeling*, vol. 61, no. 9, pp. 4425–4441, Sep. 2021, doi: 10.1021/acs.jcim.1c00560.

[151] E. Laurini, D. Marson, S. Aulic, M. Fermeglia, and S. Pricl, “Computational alanine scanning and structural analysis of the SARS-CoV-2 spike protein/angiotensin-converting enzyme 2 complex,” *ACS Nano*, vol. 14, no. 9, 2020, doi: 10.1021/acsnano.0c04674.

[152] P. K. Panda *et al.*, “Structure-based drug designing and immunoinformatics approach for SARS-CoV-2,” *Science Advances*, vol. 6, no. 28, 2020, doi: 10.1126/sciadv.abb8097.

[153] Z. Li *et al.*, “Identify potent SARS-CoV-2 main protease inhibitors via accelerated free energy perturbation-based virtual screening of existing drugs,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 44, 2020, doi: 10.1073/pnas.2010470117.

[154] J. Wang, “Fast identification of possible drug treatment of coronavirus disease-19 (COVID-19) through computational drug repurposing study,” *Journal of Chemical Information and Modeling*, vol. 60, no. 6, 2020, doi: 10.1021/acs.jcim.0c00179.

[155] P. Sang, S. H. Tian, Z. H. Meng, and L. Q. Yang, “Anti-HIV drug repurposing against SARS-CoV-2,” *RSC Advances*, vol. 10, no. 27, 2020, doi: 10.1039/d0ra01899f.

[156] A. Spinello, A. Saltalamacchia, and A. Magistrato, “Is the rigidity of SARS-CoV-2 spike receptor-binding motif the hallmark for its enhanced infectivity? Insights from all-atom simulations,” *Journal of Physical Chemistry Letters*, vol. 11, no. 12, 2020, doi: 10.1021/acs.jpclett.0c01148.

[157] J. Ren, T. C. Jenkins, and J. B. Chaires, “Energetics of DNA intercalation reactions,” *Biochemistry*, vol. 39, no. 29, 2000, doi: 10.1021/bi000474a.

[158] E. Wang *et al.*, “Development and evaluation of MM/GBSA based on a variable dielectric GB model for predicting protein-ligand binding affinities,” *Journal of Chemical Information and Modeling*, vol. 60, no. 11, 2020, doi: 10.1021/acs.jcim.0c00024.

[159] S. Genheden and U. Ryde, “The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities,” *Expert Opinion on Drug Discovery*, vol. 10, no. 5. 2015. doi: 10.1517/17460441.2015.1032936.

[160] B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke, and A. E. Roitberg, “MMPBSA.py: An efficient program for end-state free energy calculations,” *Journal of Chemical Theory and Computation*, vol. 8, no. 9, 2012, doi: 10.1021/ct300418h.

[161] M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon, "The statistical-thermodynamic basis for computation of binding affinities: A critical review," *Biophysical Journal*, vol. 72, no. 3, 1997, doi: 10.1016/S0006-3495(97)78756-3.

[162] C. M. Nunn, L. van Meervelt, S. Zhang, M. H. Moore, and O. Kennard, "DNA-drug interactions. The crystal structures of d(TGTACA) and d(TGATCA) complexed with daunomycin," *Journal of Molecular Biology*, vol. 222, no. 2, 1991, doi: 10.1016/0022-2836(91)90203-I.

[163] J. B. Chaires, "Thermodynamics of the daunomycin–DNA interaction: Ionic strength dependence of the enthalpy and entropy," *Biopolymers*, vol. 24, no. 2, 1985, doi: 10.1002/bip.360240208.

[164] M. T. Record, C. F. Anderson, and T. M. Lohman, "Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: The roles of ion association or release, screening, and ion effects on water activity," *Quarterly Reviews of Biophysics*, vol. 11, no. 2, 1978, doi: 10.1017/S003358350000202X.

[165] J. B. Chaires *et al.*, "Dissection of the free energy of anthracycline antibiotic binding to DNA: Electrostatic contributions," *Journal of the American Chemical Society*, vol. 115, no. 13, 1993, doi: 10.1021/ja00066a002.

[166] D. R. Roe and T. E. Cheatham, "PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data," *Journal of Chemical Theory and Computation*, vol. 9, no. 7, 2013, doi: 10.1021/ct400341p.

[167] W. Kabsch, "A discussion of the solution for the best rotation to relate two sets of vectors," *Acta Crystallographica Section A*, vol. 34, no. 5, 1978, doi: 10.1107/S0567739478001680.

[168] H. Wan, J. P. Hu, X. H. Tian, and S. Chang, "Molecular dynamics simulations of wild type and mutants of human complement receptor 2 complexed with C3d," *Physical Chemistry Chemical Physics*, vol. 15, no. 4, 2013, doi: 10.1039/c2cp41388d.

[169] S. Haider, G. N. Parkinson, and S. Neidle, "Molecular dynamics and principal components analysis of human telomeric quadruplex multimers," *Biophysical Journal*, vol. 95, no. 1, 2008, doi: 10.1529/biophysj.107.120501.

[170] C. C. David and D. J. Jacobs, "Principal component analysis: A method for determining the essential dynamics of proteins," *Methods in Molecular Biology*, vol. 1084, 2014, doi: 10.1007/978-1-62703-658-0_11.

[171] B. J. Grant, A. P. C. Rodrigues, K. M. ElSawy, J. A. McCammon, and L. S. D. Caves, "Bio3d: An R package for the comparative analysis of protein structures," *Bioinformatics*, vol. 22, no. 21, 2006, doi: 10.1093/bioinformatics/btl461.

[172] H. Flyvbjerg and H. G. Petersen, "Error estimates on averages of correlated data," *The Journal of Chemical Physics*, vol. 91, no. 1, 1989, doi: 10.1063/1.457480.

[173] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," *Journal of Molecular Graphics*, vol. 14, no. 1, 1996, doi: 10.1016/0263-7855(96)00018-5.

[174] X. J. Lu and W. K. Olson, "3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures," *Nucleic Acids Research*, vol. 31, no. 17, 2003, doi: 10.1093/nar/gkg680.
[175] S. Neidle, *Principles of nucleic acid structure*. 2008. doi: 10.1016/B978-0-12-369507-9.X5001-8.
[176] P. E. Nielsen, "Sequence-selective DNA recognition by synthetic ligands," *Bioconjugate Chemistry*, vol. 2, no. 1, 1991, doi: 10.1021/bc00007a001.
[177] P. Mohan and N. Rapoport, "Doxorubicin as a molecular nanotheranostic agent: Effect of doxorubicin encapsulation in micelles or nanoemulsions on the ultrasound-mediated intracellular delivery and nuclear trafficking," *Molecular Pharmaceutics*, vol. 7, no. 6, 2010, doi: 10.1021/mp100269f.
[178] A. M. Meredith and C. R. Dass, "Increasing role of the cancer chemotherapeutic doxorubicin in cellular metabolism," *Journal of Pharmacy and Pharmacology*, vol. 68, no. 6. 2016. doi: 10.1111/jphp.12539.
[179] E. Schneider, Y. H. Hsiang, and L. F. Liu, "DNA topoisomerases as anticancer drug targets," *Advances in Pharmacology*, vol. 21, no. C, 1990, doi: 10.1016/S1054-3589(08)60342-7.
[180] K. Buzun, A. Bielawska, K. Bielawski, and A. Gornowicz, "DNA topoisomerases as molecular targets for anticancer drugs," *Journal of Enzyme Inhibition and Medicinal Chemistry*, vol. 35, no. 1. 2020. doi: 10.1080/14756366.2020.1821676.
[181] F. A. Fornari, J. K. Randolph, J. C. Yalowich, M. K. Ritke, and D. A. Gewirtz, "Interference by doxorubicin with DNA unwinding in MCF-7 breast tumor cells," *Molecular Pharmacology*, vol. 45, no. 4, 1994.
[182] N. Pilco-Ferreto and G. M. Calaf, "Influence of doxorubicin on apoptosis and oxidative stress in breast cancer cell lines," *International Journal of Oncology*, vol. 49, no. 2, 2016, doi: 10.3892/ijo.2016.3558.
[183] G. J. Quigley, A. H. J. Wang, G. Ughetto, G. van der Marel, J. H. van Boom, and A. Rich, "Molecular structure of an anticancer drug-DNA complex: Daunomycin plus d(CpGpTpApCpG)," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 77, no. 12 II, 1980, doi: 10.1073/pnas.77.12.7204.
[184] E. Ragg, R. Mondelli, C. Battistini, A. Garbesi, and F. P. Colonna, "31P NMR study of daunorubicin-d(CGTACG) complex in solution Evidence of the intercalation sites," *FEBS Letters*, vol. 236, no. 1, 1988, doi: 10.1016/0014-5793(88)80320-X.
[185] J. B. Chaires, N. Dattagupta, and D. M. Crothers, "Kinetics of the daunomycin-DNA interaction," *Biochemistry*, vol. 24, no. 2, 1985, doi: 10.1021/bi00323a004.
[186] V. Rizzo, N. Sacchi, and M. Menozzi, "Kinetic studies of anthracycline-DNA interaction by fluorescence stopped flow confirm a complex association Mechanism," *Biochemistry*, vol. 28, no. 1, 1989, doi: 10.1021/bi00427a038.
[187] C. Pérez-Arnaiz, N. Busto, J. M. Leal, and B. García, "New insights into the mechanism of the DNA/doxorubicin interaction," *Journal of Physical Chemistry B*, vol. 118, no. 5, 2014, doi: 10.1021/jp411429g.
[188] R. M. Kenney, K. E. Buxton, and S. Glazier, "Investigating the impacts of DNA binding mode and sequence on thermodynamic quantities and water exchange values

for two small molecule drugs," *Biophysical Chemistry*, vol. 216, 2016, doi: 10.1016/j.bpc.2016.05.002.

[189] X. Chen, L. Zhou, J. Wang, G. Jiang, H. Cheng, and R. Pei, "The study of the interaction between doxorubicin and single-stranded DNA," *ChemistrySelect*, vol. 1, no. 13, 2016, doi: 10.1002/slct.201600473.

[190] D. Agudelo, P. Bourassa, G. Bérubé, and H. A. Tajmir-Riahi, "Intercalation of antitumor drug doxorubicin and its analogue by DNA duplex: Structural features and biological implications," *International Journal of Biological Macromolecules*, vol. 66, 2014, doi: 10.1016/j.ijbiomac.2014.02.028.

[191] X. Zhang, A. Poniewierski, K. Sozański, Y. Zhou, A. Brzozowska-Elliott, and R. Holyst, "Fluorescence correlation spectroscopy for multiple-site equilibrium binding: A case of doxorubicin-DNA interaction," *Physical Chemistry Chemical Physics*, vol. 21, no. 3, 2019, doi: 10.1039/c8cp06752j.

[192] B. Gnapareddy *et al.*, "Chemical and physical characteristics of doxorubicin hydrochloride drug-doped salmon DNA thin films," *Scientific Reports*, vol. 5, 2015, doi: 10.1038/srep12722.

[193] C. H. Stuart, D. A. Horita, M. J. Thomas, F. R. Salsbury, M. O. Lively, and W. H. Gmeiner, "Site-specific DNA-doxorubicin conjugates display enhanced cytotoxicity to breast cancer cells," *Bioconjugate Chemistry*, vol. 25, no. 2, 2014, doi: 10.1021/bc4005427.

[194] F. Yao *et al.*, "Nanopore single-molecule analysis of DNA-doxorubicin interactions," *Analytical Chemistry*, vol. 87, no. 1, 2015, doi: 10.1021/ac503926g.

[195] L. Scaglioni, R. Mondelli, R. Artali, F. R. Sirtori, and S. Mazzini, "Nemorubicin and doxorubicin bind the G-quadruplex sequences of the human telomeres and of the c-MYC promoter element Pu22," *Biochimica et Biophysica Acta - General Subjects*, vol. 1860, no. 6, 2016, doi: 10.1016/j.bbagen.2016.02.011.

[196] I. Manet *et al.*, "Complexes of the antitumoral drugs doxorubicin and sabarubicin with telomeric G-quadruplex in basket conformation: Ground and excited state properties," *Photochemical and Photobiological Sciences*, vol. 10, no. 8, 2011, doi: 10.1039/c1pp05065f.

[197] A. R. Rubio, N. Busto, J. M. Leal, and B. García, "Doxorubicin binds to duplex RNA with higher affinity than ctDNA and favours the isothermal denaturation of triplex RNA," *RSC Advances*, vol. 6, no. 103, 2016, doi: 10.1039/c6ra21387a.

[198] D. Agudelo, P. Bourassa, G. Bérubé, and H. A. Tajmir-Riahi, "Review on the binding of anticancer drug doxorubicin with DNA and tRNA: Structural models and antitumor activity," *Journal of Photochemistry and Photobiology B: Biology*, vol. 158. 2016. doi: 10.1016/j.jphotobiol.2016.02.032.

[199] Y. Nakata and A. J. Hopfinger, "An extended conformational analysis of doxorubicin," *FEBS Letters*, vol. 117, no. 1–2, 1980, doi: 10.1016/0014-5793(80)80957-4.

[200] Y. Nakata and A. J. Hopfinger, "Predicted mode of intercalation of doxorubicin with dinucleotide dimers," *Topics in Catalysis*, vol. 95, no. 2, 1980, doi: 10.1016/0006-291X(80)90824-4.

[201] G. Barone *et al.*, "Intercalation of daunomycin into stacked dna base pairs. Dft study of an anticancer drug," *Journal of Biomolecular Structure and Dynamics*, vol. 26, no. 1, 2008, doi: 10.1080/07391102.2008.10507229.
[202] S. Zhu, L. Yan, X. Ji, and W. Lu, "Conformational diversity of anthracycline anticancer antibiotics: A density functional theory calculation," *Journal of Molecular Structure: THEOCHEM*, vol. 951, no. 1–3, 2010, doi: 10.1016/j.theochem.2010.04.008.
[203] L. Poudel *et al.*, "Electronic structure and partial charge distribution of doxorubicin in different molecular environments," *ChemPhysChem*, vol. 16, no. 7, 2015, doi: 10.1002/cphc.201402893.
[204] W. Y. Ching and P. Rulis, *Electronic structure methods for complex materials: The orthogonalized linear combination of atomic orbitals*, vol. 9780199575800. 2012. doi: 10.1093/acprof:oso/9780199575800.001.0001.
[205] G. Kresse and J. Furthmüller, "Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set," *Computational Materials Science*, vol. 6, no. 1, 1996, doi: 10.1016/0927-0256(96)00008-0.
[206] G. Kresse and J. Furthmüller, "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set," *Physical Review B - Condensed Matter and Materials Physics*, vol. 54, no. 16, 1996, doi: 10.1103/PhysRevB.54.11169.
[207] P. Agrawal, S. K. Barthwal, and R. Barthwal, "Studies on self-aggregation of anthracycline drugs by restrained molecular dynamics approach using nuclear magnetic resonance spectroscopy supported by absorption, fluorescence, diffusion ordered spectroscopy and mass spectrometry," *European Journal of Medicinal Chemistry*, vol. 44, no. 4, 2009, doi: 10.1016/j.ejmech.2008.09.037.
[208] M. Jain, S. K. Barthwal, R. Barthwal, and G. Govil, "Restrained molecular dynamics studies on complex of adriamycin with DNA hexamer sequence d-CGATCG," *Archives of Biochemistry and Biophysics*, vol. 439, no. 1, 2005, doi: 10.1016/j.abb.2005.05.009.
[209] R. Barthwal, P. Agrawal, A. N. Tripathi, U. Sharma, N. R. Jagannathan, and G. Govil, "Structural elucidation of 4′-epiadriamycin by nuclear magnetic resonance spectroscopy and comparison with adriamycin and daunomycin using quantum mechanical and restrained molecular dynamics approach," *Archives of Biochemistry and Biophysics*, vol. 474, no. 1, 2008, doi: 10.1016/j.abb.2008.02.039.
[210] D. A. Case *et al.*, "Amber 11 users' manual principal contributors to the current codes." [Online]. Available: http://ambermd.org/contributors
[211] M. J. Frisch *et al.*, "Gaussian 09, Rev. D.01," *Gaussian Inc., Wallingford, CT*. 2016.
[212] C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model," *Journal of Physical Chemistry*, vol. 97, no. 40, 1993, doi: 10.1021/j100142a004.
[213] F. Y. Dupradeau *et al.*, "The R.E.D. tools: Advances in RESP and ESP charge derivation and force field library building," *Physical Chemistry Chemical Physics*, vol. 12, no. 28, 2010, doi: 10.1039/c0cp00111b.

[214] E. Vanquelef *et al.*, "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 Research*, vol. 39, no. SUPPL. 2, 2011, doi: 10.1093/nar/gkr288.

[215] D. Sitkoff, K. A. Sharp, and B. Honig, "Accurate calculation of hydration free energies using macroscopic solvent models," *Journal of Physical Chemistry*, vol. 98, no. 7, 1994, doi: 10.1021/j100058a043.

[216] R. E. Dickerson and H. R. Drew, "Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure," *Journal of Molecular Biology*, vol. 149, no. 4, 1981, doi: 10.1016/0022-2836(81)90357-0.

[217] M. E. Nuss, F. J. Marsh, and P. A. Kollman, "Theoretical studies of drug-dinucleotide interactions. empirical energy function calculations on the interaction of ethidium, 9-aminoacridine, and proflavin cations with the base-paired dinucleotides GpC and CpG," *Journal of the American Chemical Society*, vol. 101, no. 4, 1979, doi: 10.1021/ja00498a008.

[218] R. L. Ornstein and R. Rein, "Energetics of intercalation specificity. I. Backbone unwinding," *Biopolymers*, vol. 18, no. 5, 1979, doi: 10.1002/bip.1979.360180517.

[219] R. L. Ornstein and R. Rein, "Energetic and structural aspects of ethidium cation intercalation into DNA minihelices," *Biopolymers*, vol. 18, no. 11, 1979, doi: 10.1002/bip.1979.360181112.

[220] D. Řeha *et al.*, "Intercalators. 1. Nature of stacking interactions between intercalators (ethidium, daunomycin, ellipticine, and 4′,6-diaminide-2-phenylindole) and DNA base pairs. Ab initio quantum chemical, density functional theory, and empirical potential study," *Journal of the American Chemical Society*, vol. 124, no. 13, 2002, doi: 10.1021/ja011490d.

[221] A. Mukherjee, R. Lavery, B. Bagchi, and J. T. Hynes, "On the molecular mechanism of drug intercalation into DNA: A simulation study of the intercalation pathway, free energy, and DNA structural changes," *Journal of the American Chemical Society*, vol. 130, no. 30, 2008, doi: 10.1021/ja8001666.

[222] F. Barcelo, J. Martorell, F. Gavilanes, and J. M. Gonzalez-Ros, "Equilibrium binding of daunomycin and adriamycin to calf thymus DNA. Temperature and ionic strength dependence of thermodynamic parameters," *Biochemical Pharmacology*, vol. 37, no. 11, 1988, doi: 10.1016/0006-2952(88)90571-0.

[223] R. A. G. Friedman and G. S. Manning, "Polyelectrolyte effects on site-binding equilibria with application to the intercalation of drugs into DNA," *Biopolymers*, vol. 23, no. 12, 1984, doi: 10.1002/bip.360231202.

[224] W. D. Wilson, C. R. Krishnamoorthy, Y. -H Wang, and J. C. Smith, "Mechanism of intercalation: Ion effects on the equilibrium and kinetic constants for the interaction of propidium and ethidium with DNA," *Biopolymers*, vol. 24, no. 10, 1985, doi: 10.1002/bip.360241008.

[225] D. E. Graves and T. R. Krugh, "Adriamycin and daunorubicin bind in a cooperative manner to deoxyribonucleic acid," *Biochemistry*, vol. 22, no. 16, 1983, doi: 10.1021/bi00285a033.

[226] V. Rizzo *et al.*, "Association of anthracyclines and synthetic hexanucleotides. Structural factors influencing sequence specificity," *Journal of Molecular Recognition*, vol. 2, no. 3, 1989, doi: 10.1002/jmr.300020306.
[227] R. S. Spolar and M. T. Record, "Coupling of local folding to site-specific binding of proteins to DNA," *Science*, vol. 263, no. 5148, 1994, doi: 10.1126/science.8303294.
[228] L. A. Marky, K. Alessi, and D. Rentzeperis, "Calorimetric studies of drug-DNA interactions," *Advances in DNA Sequence-Specific Agents*, vol. 2, no. C, 1996, doi: 10.1016/S1067-568X(96)80004-1.
[229] L. H. Hurley and F. L. Boyd, "DNA as a target for drug action," *Trends in Pharmacological Sciences*, vol. 9, no. 11. 1988. doi: 10.1016/0165-6147(88)90067-3.
[230] M. Sirajuddin, S. Ali, and A. Badshah, "Drug-DNA interactions and their study by UV-Visible, fluorescence spectroscopies and cyclic voltametry," *Journal of Photochemistry and Photobiology B: Biology*, vol. 124. 2013. doi: 10.1016/j.jphotobiol.2013.03.013.
[231] G. D. Stormo and Y. Zhao, "Determining the specificity of protein-DNA interactions," *Nature Reviews Genetics*, vol. 11, no. 11. 2010. doi: 10.1038/nrg2845.
[232] R. Rohs, X. Jin, S. M. West, R. Joshi, B. Honig, and R. S. Mann, "Origins of specificity in protein-DNA recognition," *Annual Review of Biochemistry*, vol. 79. 2010. doi: 10.1146/annurev-biochem-060408-091030.
[233] A. Srivastava, R. Timsina, S. Heo, S. W. Dewage, S. Kirmizialtin, and X. Qiu, "Structure-guided DNA–DNA attraction mediated by divalent cations," *Nucleic Acids Research*, vol. 48, no. 13, 2020, doi: 10.1093/nar/gkaa499.
[234] L. Strekowski and B. Wilson, "Noncovalent interactions with DNA: An overview," *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, vol. 623, no. 1–2, 2007, doi: 10.1016/j.mrfmmm.2007.03.008.
[235] M. Manosas, J. Camunas-Soler, V. Croquette, and F. Ritort, "Single molecule high-throughput footprinting of small and large DNA ligands," *Nature Communications*, vol. 8, no. 1, 2017, doi: 10.1038/s41467-017-00379-w.
[236] J. Ren and J. B. Chaires, "Sequence and structural selectivity of nucleic acid binding ligands," *Biochemistry*, vol. 38, no. 49, 1999, doi: 10.1021/bi992070s.
[237] W. C. Tse and D. L. Boger, "Sequence-selective DNA recognition: Natural products and nature's lessons," *Chemistry and Biology*, vol. 11, no. 12. 2004. doi: 10.1016/j.chembiol.2003.08.012.
[238] T. Laskowski, W. Andrałojć, J. Grynda, P. Gwarda, J. Mazerski, and Z. Gdaniec, "A strong preference for the TA/TA dinucleotide step discovered for an acridine-based, potent antitumor dsDNA intercalator, C-1305: NMR-driven structural and sequence-specificity studies," *Scientific Reports*, vol. 10, no. 1, 2020, doi: 10.1038/s41598-020-68609-8.
[239] A. Kellett, Z. Molphy, C. Slator, V. McKee, and N. P. Farrell, "Molecular methods for assessment of non-covalent metallodrug-DNA interactions," *Chemical Society Reviews*, vol. 48, no. 4, 2019, doi: 10.1039/c8cs00157j.
[240] J. B. Chaires, N. Dattagupta, and D. M. Crothers, "Binding of daunomycin to calf thymus nucleosomes," *Biochemistry*, vol. 22, no. 2, 1983, doi: 10.1021/bi00271a009.

[241] J. B. Chaires, N. Dattagupta, and D. M. Crothers, “Studies on interaction of anthracycline antibiotics and deoxyribonucleic acid: Equilibrium binding studies on interaction of daunomycin with deoxyribonucleic acid,” *Biochemistry*, vol. 21, no. 17, 1982, doi: 10.1021/bi00260a005.

[242] E. C. Meng, E. F. Pettersen, G. S. Couch, C. C. Huang, and T. E. Ferrin, “Tools for integrated sequence-structure analysis with UCSF Chimera,” *BMC Bioinformatics*, vol. 7, 2006, doi: 10.1186/1471-2105-7-339.

[243] S. le Grand, A. W. Götz, and R. C. Walker, “SPFP: Speed without compromise - A mixed precision model for GPU accelerated molecular dynamics simulations,” *Computer Physics Communications*, vol. 184, no. 2, 2013, doi: 10.1016/j.cpc.2012.09.022.

[244] R. Salomon-Ferrer, A. W. Götz, D. Poole, S. le Grand, and R. C. Walker, “Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald,” *Journal of Chemical Theory and Computation*, vol. 9, no. 9, 2013, doi: 10.1021/ct400314y.

[245] A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. le Grand, and R. C. Walker, “Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. generalized born,” *Journal of Chemical Theory and Computation*, vol. 8, no. 5, 2012, doi: 10.1021/ct200909j.

[246] J. B. Chaires, “A thermodynamic signature for drug-DNA binding mode,” *Archives of Biochemistry and Biophysics*, vol. 453, no. 1, 2006, doi: 10.1016/j.abb.2006.03.027.

[247] M. Trieb, C. Rauch, F. R. Wibowo, B. Wellenzohn, and K. R. Liedl, “Cooperative effects on the formation of intercalation sites,” *Nucleic Acids Research*, vol. 32, no. 15, 2004, doi: 10.1093/nar/gkh788.

[248] L. E. Xodo, G. Manzini, J. Ruggiero, and F. Quadrifoglio, “On the interaction of daunomycin with synthetic alternating DNAs: Sequence specificity and polyelectrolyte effects on the intercalation equilibrium,” *Biopolymers*, vol. 27, no. 11, 1988, doi: 10.1002/bip.360271112.

[249] D. P. Remeta, K. J. Breslauer, C. P. Mudd, and R. L. Berger, “Thermodynamic characterization of daunomycin-DNA interactions: Comparison of complete Binding profiles for a series of DNA host duplexes,” *Biochemistry*, vol. 32, no. 19, 1993, doi: 10.1021/bi00070a014.

[250] G. Capranico, E. Butelli, and F. Zunino, “Change of the sequence specificity of daunorubicin-stimulated topoisomerase II DNA cleavage by epimerization of the amino group of the sugar moiety,” *Cancer Research*, vol. 55, no. 2, 1995.

[251] G. Capranico, M. Binaschi, M. E. Borgnetto, F. Zunino, and M. Palumbo, “A protein-mediated mechanism for the DNA sequence-specific action of topoisomerase II poisons,” *Trends in Pharmacological Sciences*, vol. 18, no. 9. 1997. doi: 10.1016/S0165-6147(97)01095-X.

[252] H. Lei, X. Wang, and C. Wu, “Early stage intercalation of doxorubicin to DNA fragments observed in molecular dynamics binding simulations,” *Journal of Molecular Graphics and Modelling*, vol. 38, 2012, doi: 10.1016/j.jmgm.2012.05.006.

[253] K. R. Karnati, Y. Wang, and Y. Du, “Exploring the binding mode and thermodynamics of inverse agonists against estrogen-related receptor alpha,” *RSC Advances*, vol. 10, no. 28, 2020, doi: 10.1039/c9ra10697a.
[254] H. Gohlke, C. Kiel, and D. A. Case, “Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes.,” *Journal of molecular biology*, vol. 330, no. 4, 2003, doi: 10.1016/S0022-2836(03)00610-7.
[255] C. Cera, G. Palù, S. M. Magno, and M. Palumbo, “Interaction between second generation anthracyclines and DNA in the nucleosomal structure,” *Nucleic Acids Research*, vol. 19, no. 9, 1991, doi: 10.1093/nar/19.9.2309.

VITA

Bahaa Abdalhassan Jawad is a doctoral candidate in Physics and Cell Biology and Biophysics at the University of Missouri-Kansas City (UMKC). His research advisor is Professor Wai-Yim Ching. He was born on January 28, 1983, in Najaf, Iraq. He finished his schooling at the local public schools. In 2006, Mr. Jawad completed his bachelor's degree in Science (B.Sc.) from the University of Technology in Baghdad, Iraq. He graduated as class valedictorian and was hired as a physicist assistant at the same university immediately. He became a teaching assistant in 2012, after receiving his master's degree in science (M.Sc.) from Al-Nahrain University in Baghdad.

In 2015, Mr. Jawad came to the USA as a graduate student in the Department of Physics and Astronomy at UMKC, after getting a scholarship from the Higher Committee for Education Development (HCED) in Iraq. In Fall 2015, he started his interdisciplinary PhD program. Starting from September 2020, he works as a Graduate Research Assistant (GRA) in the Electronic Structure Group (ESG) in the Department of Physics and Astronomy at UMKC. After graduation, Mr. Jawad will return to Iraq to continue his career as an Assistance Professor in the Department of Applied Sciences at the University of Technology and to pursue research interests.

Mr. Jawad is a member of the Biophysical Society (BPS). He has authored or co-authored six papers in peer-reviewed journals and has presented two conference posters and four university symposium posters.

PUBLICATIONS

1. Khagendra Baral, Puja Adhikari, **Bahaa Jawad**, Rudolf Podgornik, and Wai-Yim Ching. (2021). Solvent Effect on the Structure and Properties of RGD Peptide (1FUV) at Body Temperature (310 K) Using Ab Initio Molecular Dynamics. Polymers, 13(19), 3434.
2. Puja Adhikari, Rudolf Podgornik, **Bahaa Jawad**, and Wai-Yim Ching. (2021). First-Principles Simulation of Dielectric Function in Biomolecules. Materials, 14(19), 5774.
3. **Bahaa Jawad**, Puja Adhikari, Rudolf Podgornik, and Wai-Yim Ching. (2021). Key Interacting Residues between RBD of SARS-CoV-2 and ACE2 Receptor: Combination of Molecular Dynamics Simulation and Density Functional Calculation. Journal of Chemical Information and Modeling, 61(9), 4425–4441.
4. Wai-Yim Ching, Puja Adhikari, **Bahaa Jawad**, and Rudolf Podgornik. (2021). Ultra-large-scale ab initio quantum chemical computation of bio-molecular systems: The case of spike protein of SARS-CoV-2 virus. Computational and Structural Biotechnology Journal, 19, 1288-1301.
5. **Bahaa Jawad**, Lokendra Poudel, Rudolf Podgornik, and Wai-Yim Ching. (2020). Thermodynamic dissection of the intercalation binding process of doxorubicin to dsDNA with implications of ionic and solvent effects. The Journal of Physical Chemistry B, 124(36), 7803-7818.
6. **Bahaa Jawad**, Lokendra Poudel, Rudolf Podgornik, Nicole F. Steinmetz, and Wai-Yim Ching. (2019). Molecular mechanism and binding free energy of doxorubicin intercalation in DNA. Physical Chemistry Chemical Physics, 21(7), 3877-3893.

CONFERENCES

1. **Bahaa Jawad**, Puja Adhikari, and Wai-Yim Ching. “Potential therapeutics for targeting the SARS-CoV-2 spike protein: ACE2 interface using molecular docking, MD simulation and density functional theory approaches”., Poster Presentation in Cold Spring Harbor Asia COVID19/SARS CoV2 Rapid Research Reports (Virtual), Hosted by Cold Spring Harbor Asia, Suzhou, China, November 2-4, 2020.
2. **Bahaa Jawad**, Lokendra Poudel, and Wai-Yim Ching. "Molecular Mechanism and Binding Free Energy of Doxorubicin Intercalation in DNA”. Poster Presentation in Annual UMKC Faculty Research Symposium, UMKC, USA, April 24, 2019.
3. **Bahaa Jawad**, Lokendra Poudel, and Wai-Yim Ching. "Molecular Dynamic and Free Energy Analysis of Doxorubicin and DNA Complex. ", Poster Presentation in the

Biophysical Society 62nd Annual Meeting, February 17-21, 2018, San Francisco, CA, USA.

4. **Bahaa Jawad**, Lokendra Poudel, and Wai-Yim Ching. “Binding Free Energy of Doxorubicin-DNA Intercalation using Molecular Dynamics Simulation.”, Poster Presentation in Annual UMKC Faculty Research Symposium, UMKC, USA, December 7, 2018.
5. **Bahaa Jawad**, Lokendra Poudel, and Wai Yim Ching. “Molecular mechanism and binding free energy of doxorubicin intercalation in DNA”. Poster Presentation in Annual UMKC Faculty Research Symposium, UMKC, USA, December 8, 2017.
6. **Bahaa Jawad**, Lokendra Poudel, and Wai-Yim Ching. "Study of the interaction energies in Doxorubicin: A Combined DFT and NAMD Methods”. Poster presentation in Annual UMKC Faculty Research Symposium, UMKC, USA, December 9, 2016.