Summer 8-8-2017

Pirin Allosterically Modulates The Dynamics And Interactions Of The κB DNA In The NF-κB Supramolecular Complex

Charles Adeniran
cadeniran1@gsu.edu

PIRIN ALLOSTERICALLY MODULATES THE DYNAMICS AND INTERACTIONS OF THE κB DNA IN THE NF-κB SUPRAMOLECULAR COMPLEX

by

CHARLES ADENIRAN

Under the Direction of Donald Hamelberg, PhD

ABSTRACT

The NF-κB family of transcription factors controls a number of essential cellular functions. Pirin is a non-heme iron (Fe) redox specific co-regulator of NF-κB (p65) and has been shown to modulate the affinity between the homodimeric p65 and the DNA. The allosteric effect of the active Fe(III) form of Pirin on the DNA is not known and has not been investigated. We carry out multiple microsecond-long molecular dynamics simulations of the free DNA, p65-DNA complex, and the Pirin-p65-DNA supramolecular complexes in explicit water. We show that, unlike the Fe(II) form of Pirin, the Fe(III) form in the Pirin-p65-DNA supramolecular complex enhances the interactions and affinity between p65 and the DNA, in agreement with experiments. The results further provide atomistic details of the effect of the Fe(III) form of Pirin on the DNA upon binding to p65 to form the supramolecular complex.

INDEX WORDS: NF-κB, PIRIN, DNA, TRANSCRIPTION FACTOR, PRECURSOR PROTEIN, IRON

PIRIN ALLOSTERICALLY MODULATES THE DYNAMICS AND INTERACTIONS OF THE κB DNA IN THE NF-κB SUPRAMOLECULAR COMPLEX

by

CHARLES ADENIRAN

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2017

PIRIN ALLOSTERICALLY MODULATES THE DYNAMICS AND INTERACTIONS OF THE κB DNA IN THE NF-κB SUPRAMOLECULAR COMPLEX

by

CHARLES ADENIRAN

Committee Chair: Donald Hamelberg

Committee: Ivaylo Ivanov

Ming Luo

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2017

ACKNOWLEDGEMENTS

I would like to thank my committee members, Dr. Hamelberg, Dr. Ivanov, and Dr. Luo for helping me through this process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

1 INTRODUCTION

1.1 NF-κB a Transcription Factor

Nuclear factor-κB (NF-κB) family of transcription factors is responsible for the regulation of many sub-cellular processes due to immune and inflammatory responses. These transcription factors are known to take part in a number of cellular processes, ranging from anti-apoptotic response to critical oncogene expression[1]. The mammalian NF-κB transcription factors are structural and functionally related, consisting of RelA (p65), RelB, c-Rel, p50/p105 and p52/p100[1]. They share high sequence homology at the N terminus that is referred to as the rel homology region (RHR) and is responsible for protein dimerization, DNA binding and nuclear translocation through the proteins nuclear localization sequence (NLS).[2, 3] The activities of p65, RelA, RelB and c-Rel are tightly regulated in the cytoplasm by the interaction with an inhibitory protein, IκB[4]. Activation and release of NF-κB is initiated by phosphorylation of IκB.

The IκB proteins form a small Ser/Thr-specific kinase family which includes IκBα, IκBβ, and IκBε that are known as the classical IκBs.[5] There also exists atypical IκB proteins, which include IκBz, Bcl-3, and IκBNS.[6] These atypical family members are not generally expressed in unstimulated cells and therefore are induced upon activation and mediate their effects in the nucleus. [7] The IκB family members are characterized by their C-terminal structural motif that is essential for their function, the ankyrin (ANK) repeat domain (ARD). This repeat domain is typically 6–7 ARDs that each consist of 33 amino acid residues and forms an L-shaped structure having two α-helices connected by a loop. [7] The ARD mediates IκB binding to the NF-κB dimer and has been shown to interfere with the function of the NLS.[6]

Although the IκBs are similar in structure, they each have their own binding preferences and are subject to differential transcriptional regulation by NF-κB family members[8, 9]. For example, the classical p65 homodimer and p65/p50 heterodimer are predominantly regulated by IκBα, the most studied IκB family member[10]. On the other hand, IκBε has been found to regulate the p65 homodimer and c-Rel/p65 heterodimer[8, 11, 12]. The role of IκBβ is less well understood, although it has been shown to bind p65/p50 heterodimers with κB DNA sites. Previous studies have suggested that IκBβ may regulate p65/p50 heterodimer nuclear functions[13, 14]. There are many different stimuli that cause the activation of the NF-κB dimers to induce nuclear activity. Once inside the nucleus, the dimer orchestrates a cascade of signaling responses to external stimuli with the help of a co-regulatory protein, one namely known as Pirin.

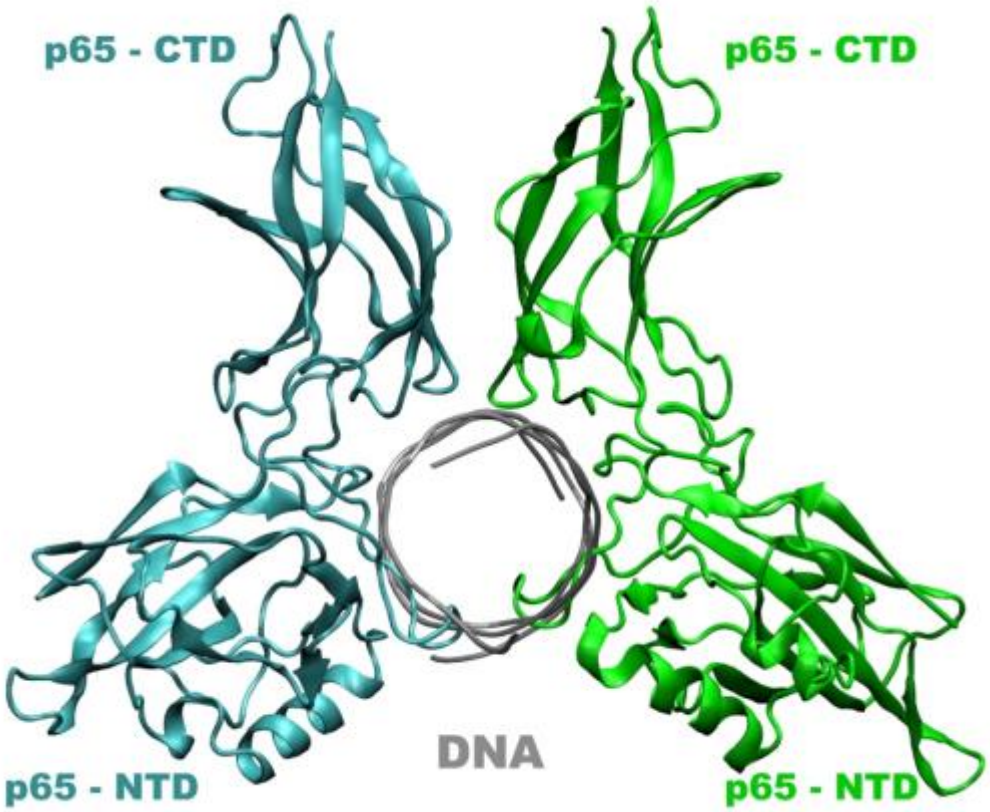

Figure 1.1 Structure of the NF-KB p65 homodimer complex bound to its gene.

1.2 Pirin the Co-regulator

Pirin is a recently discovered nuclear protein, shown in Figure 1.2, and is a sub-family member of the cupin superfamily based on its structure and sequence homology.[15] Pirin is a human protein that is expressed in all human tissues.[16] It is overexpressed in response to oxidative stress.[17-19] It is also up regulated by chronic cigarette smoking and has been linked to a host of other aberrant cellular processes.[17, 20, 21] It is a non-heme iron (Fe) binding protein and has been shown experimentally to modulate the binding of p65 to DNA, Figure 1.1, as a result of a single electron Fe-redox process.[22] Pirin is therefore an iron redox-dependent regulatory protein of p65. Liu et al.[22] have shown using a variety of experimental techniques, including x-ray crystallography, Electron Paramagnetic Resonance (EPR) experiments, fluorescence assays and Surface Plasmon Resonance (SPR), that the Fe center modulates the conformation of a distal surface region of Pirin, a region that is predicted to bind to p65. The Fe(III) form, and not the Fe(II) form, of Pirin was shown to modulate the binding of p65 to the DNA in the homodimeric p65 complex.[22]

Interaction of Pirin in the ferric Fe(III), state not the ferrous Fe(II), state, with p65 increases the affinity of p65 for the κB-gene (DNA) by more than 25-fold in the biomolecular assembly.[22] Using microsecond-long molecular dynamics (MD) simulations, Barman and Hamelberg[23] showed that the single electron redox process could significantly alter the conformational dynamics and electrostatics of Pirin. The results suggest that a restricted conformational space and electrostatic complementarity of the Fe(III) form of Pirin drive the binding of the Fe(III) form of Pirin to p65. Pirin is therefore, suggested to serve as a reversible

functional redox sensor and is believed to modulate transcription of many other genes that are involved in inflammation and stress response.[24-26] However, little is known about the allosteric effect of Pirin on the conformational dynamics of the DNA on an atomic level as it modulates the affinity between p65 and the DNA. Modulation of the conformational dynamics of the κB DNA could allosterically alter gene regulation and other sub-cellular processes. In the cell, these dynamical changes at the p65 binding site on the DNA could modulate subcellular processes[27]. Modulation of the interactions between p65 and the DNA by co-regulators, such as Pirin, could fine-tune the transcriptional level of genes through modulation of the local conformational dynamics of the DNA that could propagate to other regions and protein binding sites on the DNA, for example[28].

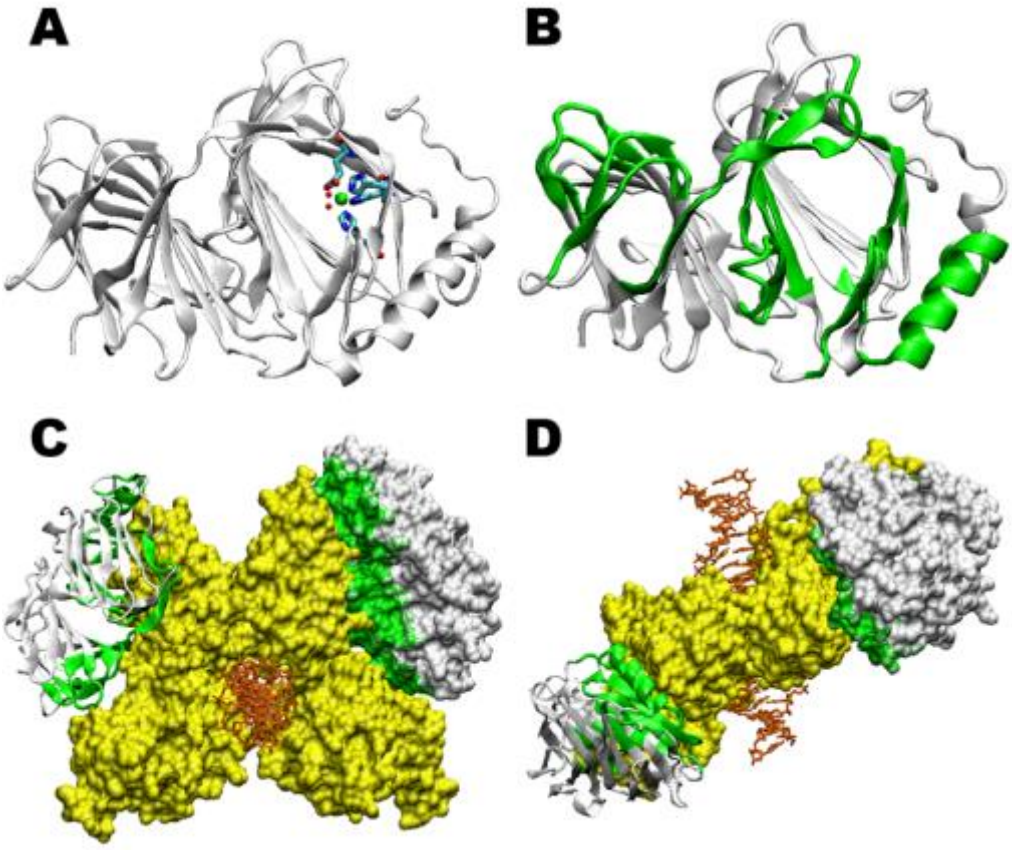

Figure 1.2 Structure of Pirin and the docked Supramolecular complex
(A) Pirin with the iron center shown. The Fe(III) (green) is coordinated to His56, His58, His101, Glu103 and 2 water molecules. (B) The predicted region of Pirin (green) that is suggested to interact with p65 in the Pirin-p65-DNA supramolecular complex. Pirin is shown in two orientations (C and D) as a result of the docking studies. p65 is a homodimer (yellow). The sequence of the DNA is 5'CGGCTGGAAATTTCCAGCCG'3 (brown), and it is the same sequence used in the simulations.

1.3 NF-κB Gene

When NF-κB translocates into the nucleus, it can induce transcription of a number of gene targets, κB DNA. The κB sites on the target DNA sequence vary greatly depending on the homodimer or heterodimer composition. Early experiments discovered that the homodimeric p65 exhibits different DNA target sites than its family members, p50, p65/p50 hetero and homodimers. The experiment during which the p65 crystal structure was resolved, the

homodimeric p65 was found to target a gene sequence, 5'-GGGRNTTTTCC-3'.[1] R denotes a purine and N denotes any nucleotide.[1] The p65 was also found to bind to 5'-GAAATTTCC-3' consensus sequence, which binds with high affinity[1]. The homodimeric p65 was found to not discriminate against the first Guanine.[1]

1.3.1 The Properties of DNA

DNA has a number of special physical and chemical properties that are important to its structure and function. In living organisms, DNA exists as a pair or pairs of molecular strands rather than a single polymer strand. These strands are entwined in the shape of a double helix and this helix is kept stable by hydrogen bonds. The discussion here will focus on the general parameters of DNA which are usually applied to any set of strands.

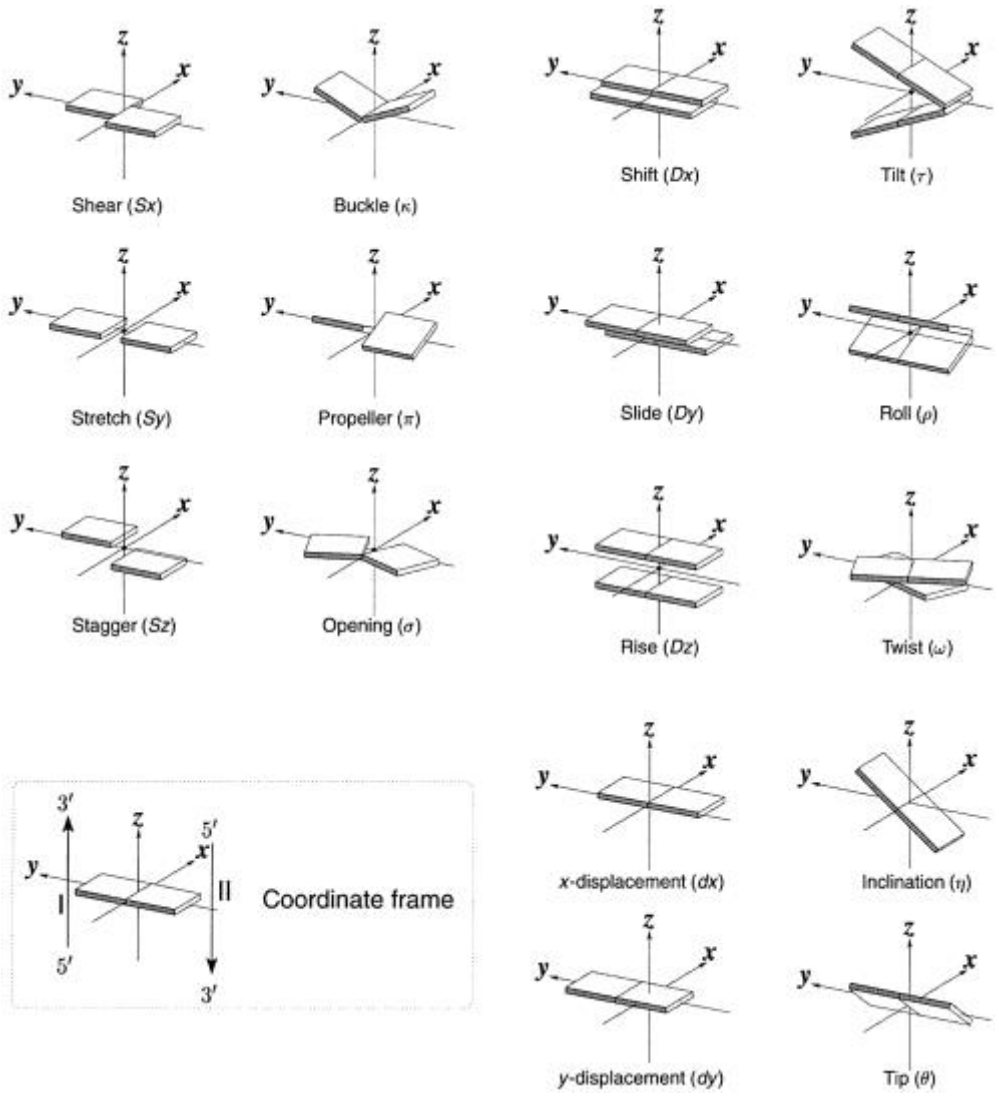

Figure 1.3 DNA Rigid Body Parameters
Visual depictions of DNA rigid-body parameters used to describe the geometry of DNA base pairs and its steps. These images illustrate positive values of the designated parameters. This figure is referenced from 3DNA Software[29] and Research Paper[30].

Base pair parameters are usually calculated using three atomic coordinates per base. For purines these are the C_8, C_6, and N_3 atoms, while for pyrimidines the C_6, C_4, and C_2 atoms are used[31]. The base pair tilt, roll, helical twist and propeller twist are sequence dependent and are based on the influence of stacking interaction energies and the van der Waals constraints imposed by different base pairs. The propeller twist, roll and displacement are extremely

important components in maintaining the stacking interaction of DNA. The axial rise is the distance between adjacent base pairs along the helical axis. The pitch is the distance along the helical axis for one complete helix turn. The pitch also equals the number of nucleotides in one turn multiplied by the axial rise. A turn is 360º, and therefore the helical twist is calculated by taking 360º divided by the number of nucleotides in one turn and is the rotation between neighboring nucleotides. [32]

The base-pair tilt is an angle and calculated when the base pair plane is not exactly perpendicular to the helical axis. The tilt is defined relative to looking at the base pair plane from the 1'-C/N linkage side. Tilting the plane clockwise is a positive tilt while a negative tilt is counterclockwise. There is a linear relationship between the tilt of an individual base with the axial rise per nucleotide. The minor groove is the side of the base pair where the sugars are attached (C1') and the major groove is the opposite side. The width of either groove is the shortest distance between phosphates across the grove minus 5.8 Å, the sum of the van der Waals radii of the two phosphates.

The x-displacement (dx) is the perpendicular distance from the long axis of the base pair to the helix axis. A helix axis is defined by the average symmetry axes of the base pairs. The roll measures the degree of departure of the mean plane of the base pairs from the perpendicular helix axis on the short axis of the base pairs. The helix twist is an angle that defines the orientation of a base pair with respect to the helix axis. That is how big an arc the base pair traces as it measured from one base pair to the next.

The two bases of many base pairs are not perfectly coplanar. Rather, they are arranged like the blades of a propeller. This deviation from the idealized structure, called propeller twisting, enhances the stacking of bases along a strand.[31] Propeller twisting is the angle between the planes of 2 paired bases. It describes the twisting of bases about their long axes within a base pair and is particularly important for DNA structure and flexibility. [31] The buckle is associated with the bases forming a cup with both ends of the bases pointing upwards or downwards relative to the primary strand [31] Shearing is a distance, the tearing apart of a base pair, which unusually occurs with long DNA strands. The opening is associated with an in-base pair plane rotation of the bases such that the major groove sides on both bases rotate away from each other.[31]

1.4 Purpose of the Study

In this study, we characterize the conformational dynamics of the free DNA, p65-DNA complex and Pirin-p65-DNA supramolecular complexes in the Fe(III) and Fe(II) forms using independent microsecond-long atomistic MD simulations in explicit water. Atomistic understanding of the regulatory process is lacking, and our results complement the relatively limited experimental studies on human Pirin and its role in regulating the mechanism of p65. The results provide valuable atomic level understanding and mechanistic details of the allosteric effect of Pirin on the DNA that are necessary for the proper functioning of the NF-κB family of transcription factors in gene expression.

2 EXPERIMENT

2.1 Computational Chemistry

Molecular Dynamics is a commonly used tool applied to bio-molecular investigations such as drug discovery, free energy calculation, protein folding and stability, molecular recognition, and nucleic acid structure to obtain a more refined understanding of chemical reaction at the sub atomic level. It was modeled to suit a variety of chemically related investigations of small molecule binding because the computational expense of sampling large bio-molecular systems can be extremely demanding. There are many different types of methods when discussing the general topic of computational simulations. As it relates to this discussion, the focus here will be put on the molecular dynamics theory used in such simulations, Molecular Mechanics, Quantum Mechanics, and Statistical Mechanics.

2.1.1 Molecular Dynamics

Molecular dynamics, a Molecular Mechanics (MM), also known as Classical Mechanics, based method, is a simulation method based on Newton's second law, the equation of motion. F stands for the force exerted on the particle, m stands for the mass, and a stands for acceleration. The acceleration of each atom in the system can be calculated by knowing the force on each atom.[33] The integration of the equation of motion then produces a trajectory that describes the positions, velocities and accelerations of the particles with respect to time.[33] This trajectory can be used to calculate the average value of properties, which can be acquired, saved and analyzed. The method is deterministic, meaning once the positions and velocities of each atom are calculated, the state of the system can be predicted.[33]

$$F_i = m_i a_i$$

Equation 2.1 Newton's Equation of Motion

In Newton's equation of motion, F_i is the force exerted on particle i, m_i is the mass of particle i and a_i is the acceleration of particle i.

$$F_i = -\Delta_i V$$

Equation 2.2 Gradient of the Potential Energy

The force can also be expressed as the gradient of the potential energy.

$$\frac{dV}{dr_i} = -m_i \frac{d^2 r_i}{dt^2}$$

Equation 2.3 Potential Energy of the System

When combining these two equations, it yields v, which is the velocity of each atom.

$$F = m\, a = m\, \frac{dv}{dt} = m \frac{d^2 r}{dt^2}$$

Equation 2.4 Potential Energy to the Change in Position

Newton's equation of motion can relate the derivative of the velocity to the acceleration.

$$a = \frac{dv}{dt}$$

Equation 2.5 Acceleration/Velocity relation

$$v = at + v_0$$

Equation 2.6 Velocity Definition

$$v = \frac{dr}{dt}$$

Equation 2.7 Velocity Derivative of Position

$$r = vt + r_0$$

Equation 2.8 Atom Position Definition

$$r = \frac{1}{2} a\, t^2 + v_0 t + r_0$$

Equation 2.9 Redefined Atom Position

When combining the equations above with the expression for the velocity, it produces the following relation in Equation 2.9 which gives the value of r at time t as a function of the acceleration a, the initial position, r_0 , and the initial velocity, v_0.

The acceleration is given as the derivative of the potential energy with respect to the position r. In order to calculate a trajectory, only the initial positions of the atoms need to be known, while the initial distribution of velocities and the acceleration is determined by the gradient of the potential energy function. The equations of motion are deterministic, meaning that the positions and the velocities at time zero determine the positions and velocities at any other time t. The initial positions can be obtained from experimental structures, such as an x-ray crystal structure or NMR spectroscopy structure.

$$P = \sum_{i+1}^{N} m_i v_i = 0$$

Equation 2.10 Momentum Definition

The initial distribution of velocities is usually determined from a random distribution with the magnitudes conforming to the required temperature and corrected so there is no overall momentum.

$$prob\,(v_{ix}) = \left(\frac{m_i}{2\pi k_B T}\right)^{\frac{1}{2}} e^{\left[\frac{1 m_i {v_{ix}}^2}{2\ k_B T}\right]}$$

Equation 2.11 Distribution of Velocities

$$T = \frac{1}{(3N)} \sum_{i=1}^{N} \frac{|P_i|}{2m_i}$$

Equation 2.12 Temperature/Velocity Relation

The initial velocity v_i is often chosen randomly from a Maxwell-Boltzmann or Gaussian distribution at a specific temperature, which gives the probability that an atom i has a velocity v_x in the x direction at a temperature T. The temperature can be calculated from the velocities using the relation in Equation 2.12 where N is the number of atoms in the system.

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

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

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

Equation 2.13 Taylor Series Expansion

There exists several common finite difference methods for the solution of Newton's equations of motion with continuous force functions.[33] No single method can be applied generally to provide a solution for any condition. All of the common integration algorithms

assume the positions, velocities and accelerations can be approximated by a Taylor series expansion as shown in Equation 2.13. Different software packages allow the use of other integration techniques, but when choosing an algorithm, it is suggested to consider the following criteria. The algorithm should conserve energy and momentum. As with any calculation, the algorithm should be computationally efficient and function under a reasonable time frame. Lastly, the algorithm should permit a long-time step to allow for integration.

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

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

$$v(t) = \frac{1}{2}\left[v\left(t-\frac{1}{2}\delta t\right) + v\left(t+\frac{1}{2}\delta t\right)\right]$$

Equation 2.14 Taylor Series Expansion for Leap-Frog Algorithm

The Leap-Frog Algorithm as used by default in the AMBER dynamics package calculates the velocities at time $t+\frac{1}{2}\delta t$ as shown in Equation 2.14. These velocities are used to calculate the positions r, at time $t+\delta t$ shown in Equation 2.14. Here the velocities leap over the positions, then the positions leap over the velocities. The advantage of this algorithm is that the velocities are explicitly calculated. However, the disadvantage is that they are not calculated at the same time as the positions. The velocities at time t can be calculated approximately by the relation shown in Equation 2.14.

2.1.1.1 Force Field

The potential function used in this thesis is defined and shown below in Equation 2.15. Molecular mechanics force fields are a key component underlying many investigations of the protein–ligand structure for drug design and other chemical investigations. This prominent force field should work well for biological molecules and the organic molecules that interact with them.

$$V = \sum_{bonds} k_r\left(r - r_{eq}\right)^2 + \sum_{angles} k_\theta\left(\theta - \theta_{eq}\right)^2 + \sum_{dihedrals} \frac{v_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j}\left[\frac{A_{ij}}{{R_{ij}}^{12}} - \frac{B_{ij}}{{R_{ij}}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}\right]$$

Equation 2.15 Energy Equation for Non-bonded Biomolecules

Here, r_{eq} and θ_{eq} are equilibration structural parameters; k_r, k_θ, and v_n are force constants; n is the multiplicity and θ is the phase angle for the torsional angle parameters. The A, B, and q parameters characterize the non-bonded potentials. For the non-bonded portion, van der Waals parameters are incorporated from traditional Amber force fields directly. Partial charges are assigned using the restrained electrostatic potential fit (RESP) model[34, 35] because of its clear and straightforward implementation. For the internal terms bonds, angle and dihedrals, parameterizations were first performed on bond lengths and bond angles that are weakly coupled to other parts in the energy function. Typically, equilibrium bond lengths and bond angles come from experiment and high-level ab initio calculations; the force constants are estimated through an empirical approach and optimized to reproduce experimental and high-level ab initio vibrational frequencies.

Unlike the primary parameters of bond length and bond angle, torsional angle parameters are highly coupled to the non-bonded energy terms. Torsional angle parameters are parameterized last, to cover other effects that cannot be considered in a simple functional form, like polarization, charge transfer, and many body effects. In practice, torsional angle parameters are derived to reproduce the energy differences of two conformations and rotational profiles, based on experimental or high-level ab initio data.

2.1.1.2 Cut Off

Generally the non-bonded potentials, such as Van-der-Waals which are represented by the Lennard-Jones function, or electrostatic interactions are assumed spherically symmetric.[36] Therefore, it is possible to modify the original non-bonded potential in such a way that it would decay faster at large distances r and converge to 0 at some finite r.[36] Three modification techniques are commonly used, which are based on truncation, switch, and shift functions.[37] All three must satisfy several requirements.

The potential should remain minimally perturbed by modifying function at small distances. The modified potential must remain a smooth function. This requirement is crucial for molecular dynamics, Langevin dynamics, or minimization procedures. Violation of this requirement may result in severe instability of the integration of equation of motions due to sudden variation of forces. While energy is conserved in standard molecular dynamics simulations such as the NVE ensemble, modification of potentials should not lead to noticeable energy drift.

$$V_{ij}(r_{ij}) = S(r_{ij})\left[\frac{A_{ij}}{{R_{ij}}^{12}} - \frac{B_{ij}}{{R_{ij}}^{6}} + \frac{q_i q_j}{r_{ij}}\right]$$

Equation 2.16 Cut off Method Non Bonded Potential

The general implementation of cut-off methods is shown in Equation 2.17, where $S(r)$ is a function which modifies the pairwise non-bonded potential V_{ij} for a pair of atoms i and j.

2.1.1.3 Short Ranged Interaction

Short ranged interactions are defined by using a Van-der-Waals potential for point-to-point interactions. The attractive Van-der-Waals pair potential between point particles is proportional to $\frac{1}{r^6}$, here r is the distance between the point particles.[38] The widely used semi-empirical potential is used to describe Van-der-Waals interactions is the Lennard-Jones potential, referred to as the 6-12 potential because of its $\left(\frac{1}{r}\right)^6$ and $\left(\frac{1}{r}\right)^{12}$ distance r dependence of the attractive interaction and repulsive component, respectively.[38] While the 6-potential is derived from point particle dipole-dipole interaction, the 12-potential is based on pure empiricism.

$$\phi(r) = -\frac{C_{vdw}}{r^6} + \frac{C_{rep}}{r^{12}} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

$$\sigma = \left(\frac{C_{rep}}{C_{vdw}}\right)^{\frac{1}{6}}$$

$$\varepsilon = \frac{{C_{vdw}}^2}{4C_{rep}}$$

Equation 2.17 Lennard Jones Potential Function

The Lennard-Jones potential is provided in the following two equivalent forms as function of the particle-particle distance r. C_{vdw} and C_{rep} are characteristic constants. $C = C_{vdw}$ is called the Van-der-Waals interaction parameter. The empirical constant ε represents the characteristic energy of interaction between the molecules. The symbol, σ is a characteristic diameter of the molecule known as the collision diameter and is the distance between two molecules for $\phi(r) = 0$.

2.1.1.4 Periodic Boundary Conditions

In most cases the purpose of simulations is to study the properties of an infinite molecular system at a given concentration, temperature, and various other parameters. The computational power of technology today allows for a study to simulate systems of up to about a million degrees of freedom. Without using special tricks, the simulation system would be confined to a finite volume. A simulation with hard walls surrounding the system or appearing as a bubble of atoms in vacuo, in theory, would not yield realistic results because it is not appropriately accounting for its surroundings. In general, the fraction of atoms on the surface relative to the total number of atoms scales as $N^{-\frac{1}{3}}$.[33]

The walls confining the system lead to a number of so called finite size effects that distort bulk properties. A common method way to avoid finite size effects is to consider periodic boundary conditions (PBC).[33] There are two important rules that must be followed when using PBC for biomolecular or organic systems. The first is to prevent self-interaction of a structure; a

sufficient number of water layers must be used.[33] The number of water molecules between the structure and the unit cell boundary must be the distance of approximately 10 Å. Following this guideline should create a minimum water buffer of about 20 Å between each structural image.[33]

The second is that the size of the unit cell L must be more than twice the cut-off distance r_c.[33] This condition eliminates correlated fluctuations that atoms may experience due to simultaneous interactions with two images of a given particle. The most common geometric container for the unit cell is cubic, but there also exists complicated containers, such as a box (non-cubic), truncated octahedron, hexagonal prism, and etc. The larger the number of planes or dimensions and the closer the shape is to a spherical container; the more efficient salvation of a structure becomes. The sphere itself cannot be used for a unit cell. It is also important to be aware that PBC introduce spurious correlated fluctuations with the wavelength of the order of L.[33] The PBC lead to anisotropic radial distributions of densities, which are manifested in the radial correlation functions $g(r)$.[33]

$$E_{tot} = \frac{1}{2}\sum_{i,j,\vec{n}} \left(\left|\vec{r}_{ij} + \vec{n}L\right|\right)$$

Equation 2.18 Total Non-Bonded Energy

The total non-bonded energy is given by Equation 2.18 where the sum is taken over all pairs of atoms i and j as well as all images. Each image, denoted as $\vec{r}$ is specified by the vector $\vec{n}$. In practice, the cut-offs in $\vec{r}$ non-bonded interactions limit the sum over n to the nearest images.

2.1.1.5 Electrostatic Interaction

Electrostatic interactions exist between cations and anions, or atoms with partial charges. These interactions can be either attractive or repulsive, depending on the sign of the charges on those ions. When a single cation and a single anion are in close proximity, the interactions are considered to be non-covalent electrostatic interactions. Non-covalent electrostatic interactions can be strong, and act as long range electrostatics. Although, electrostatic interactions are very strong, they weaken gradually with the distance $\frac{1}{r}$, where r is the distance between the ions.[39]

Electrostatic interactions are the primary stabilizing interaction between phosphate oxygens of RNA and magnesium ions for example. However, these interactions are increasingly dampened by water. Thus, there is introduced an inherent limitation of molecular dynamic simulations. A common technique that is used to recreate the effects of these interactions, is the use of periodic boundary condition as described in Section 2.1.14.[40] This technique is artificial and sometimes the use of these boundary conditions, introduce adverse effects on the equilibrium properties of the liquid or solvent. There exist several approximations used to address long-range interactions, and the best solution used within AMBER dynamics package tends to be the Ewald summation method.[39, 40]

Ewald method provides the opportunity to compute electrostatic interactions without using cut-off distances and still avoiding explicit enumeration of all atom pairs i and j. Let's consider a system with N point charges, and a net charge of 0 with periodic boundary conditions.

$$E_{ELEC} = \frac{1}{2}\sum_{i} z_i \emptyset(\vec{r}_i)$$

Equation 2.19 Electrostatic Energy

$$\emptyset\vec{r}_i = \sum_{j,\vec{n}} \frac{z_j}{|r_{ij} + \vec{n}L|}$$

Equation 2.20 Electrostatic Infinite Series

The electrostatic energy shown in Equation 2.19, where the electrostatic potential created by the charges j distributes on all unit cell images excludes the case where $i = j$. It is commonly accepted that such infinite series in Equation 2.20 are poorly converging.[39] The method solves this problem by considering two sets of spherically symmetric charge clouds.[39] The screening charge clouds have the signs opposite to the charges z_i and are therefore centered at the position of the same charges z_i.[39] The compensating charge clouds have exactly the same charge distribution as the screening ones, but these charges are assigned the opposite sign. The screening clouds are partially compensating for the point charges at the point r, to ensure fast decay of the total electrostatic potential.

$$E_{ELEC} = E_{ELEC}^{F} - E_{ELEC}^{self} + E_{ELEC}^{R}$$

Equation 2.21 Total Electrostatic Potential

$$E_{ELEC}^{F} = \frac{1}{2}\sum_{\vec{k}\neq 0} \frac{4\pi V}{k^2} \left|\rho(\vec{k})\right|^2 e^{-\frac{k^2}{4a}}$$

$$p|\vec{k}| = \frac{1}{V}\sum_{i=1}^{N} z_i e^{i\vec{k}\vec{r}_i}$$

Equation 2.22 Electrostatic Potential based on Fourier transform

$$E_{ELEC}^{self} = \sqrt{\frac{\alpha}{\pi}} \sum_{i=1}^{N} z_i^2$$

Equation 2.23 Electrostatic Potential of point charges

$$E_{ELEC}^{R} = \frac{1}{2} \sum_{i \neq j} \sum_{\vec{n}} z_i z_j \frac{erfc\left(\sqrt{\alpha} \left| \vec{r}_{ij} + \vec{n}L \right| \right)}{\left| \vec{r}_{ij} + \vec{n}L \right|}$$

Equation 2.24 Electrostatic Potential of screening charge clouds

The idea of the Ewald method is to compute the electrostatic energy as the sum of three components shown in Equation 2.21. The first is the interaction of point charges with the compensating charge clouds shown in Equation 2.22.[39] The distribution of compensating charge clouds is periodic, therefore this term is calculated using Fourier transform in k space.[39] The second term is associated with the interaction of point charges with their own compensating charge clouds shown in Equation 2.23. The third term is the interactions of point charges with the other point charges partially screened by the screening charge clouds shown in Equation 2.24. The second and the third terms are computed in real space (no Fourier transform involved). The introduced charge clouds are narrow, which leads to all the terms converging.

In Equation 2.23 to 2.25, α determines the width of Gaussian distribution of screening and compensating charge clouds, V is the volume of the unit cell. The value of α also determines the convergence of sums over n and k in Equation 2.22. The larger α becomes, the faster the real-space term converges only for large α where only $n = 0$ components survive. However, α has the opposite effect on Equation 2.23. Large α causes slower convergence in the sum over k.

The accuracy of third term computation is determined by α. Also, the trigonometric functions in Equation 2.24 are usually computed using fast interpolation methods, which require the evaluation of the function on the grid points within the interval L. The total number of which should be equal to the product of small integer numbers.

2.1.1.6 Thermostat

It is usually always preferred to study a system in a more experimentally relevant canonical isothermal-isobaric NPT ensemble, where the temperature (T) is constant. While the temperature is fixed, the total system energy is allowed to vary. There exist a large variety of approaches for introducing such a thermostat, which can be roughly classified into two categories: deterministic and stochastic, depending on whether the Newton's equation of motion contain a random component.[41] Among the deterministic approach, the Nose-Hoover Langevin thermostat appears to be the most popular technique. The reason being is because it generates an accurate canonical ensemble of the system phase space.[41]

Thermostat variables are coupled and control only global system quantities such as kinetic energy, these thermostats rely on a very efficient energy transfer within the system to achieve equipartition within the canonical distribution. The average energy of each degree of freedom inside the system should be equal to $k_B T$.[41] This use to impose problems for systems with slow degrees of freedom, as different parts of a system will experience different temperatures. With improved techniques, such as Hoover's alternate formulation of Newton's equations of motion, these issues no longer exist.

$$H = \frac{1}{2}\sum m|p_i|^2 + U(r^N) + \frac{\xi^2 Q}{2} + 3Nk_B T \ln s$$

$$\frac{dr_i}{dt} = v_i$$

$$\frac{dv_i}{dt} = -\frac{1}{m_i}\frac{\partial U(r^N)}{\partial r_i} - \xi v_i$$

$$\frac{d\xi}{dt} = \frac{(\sum m_i|v_i|^2 - 3Nk_B T)}{Q}$$

$$\frac{d\ln s}{dt} = \xi$$

Equation 2.25 Hoover's Revised Equation of Motion

In Equation 2.25, the symbol ξ is the friction coefficient. It no longer changes in time, as it did with Newton's equation, when the instantaneous kinetic energy was equal to $\frac{3}{2}Nk_B T$. The time-evolution of the particle positions and momenta is defined by the other equations listed in Equation 2.25. The Hoover's velocity update of a particle resembles Newton's equations with an additional force component that is proportional to the velocity.

2.1.1.7 Barostat

It is more desirable to maintain a simulated system at a constant pressure of 1 bar, known as the isothermal-isobaric (NPT) ensemble, for a run production simulation.[42] Run production dynamics are often saved under constant temperature and pressure conditions because this more

closely resembles laboratory conditions. However, at low temperatures, the system model has shown to be very inaccurate. Using constant pressure periodic boundaries for equilibration system setup and equilibration can lead to problems. Using constant pressure with restraints can also cause problems, so it is best practice to initially equilibrate and heat a model system at constant volume. Once equilibrated, the restraints are turned off and the barostat settings are changed to constant pressure before generating production files.

In the AMBER MD package, simulations are usually generated with isotropic position scaling, a Berendsen barostat.[42] This method assumes that the pressure is weakly coupled to a pressure bath and that the volume is periodically rescaled. Coupling to a pressure bath can be calculated by first adding an extra term to Newton's equation of motion.

$$\left(\frac{dP}{dt}\right)_{bath} = \frac{P_0 - P}{\tau_P}$$

Equation 2.26 Derivative of Pressure Equation

$$P = \frac{2}{3V}(E_k - \Xi)$$

Equation 2.27 Barostat Pressure Equation

$$\Xi = -\frac{1}{2}\sum_{i<j} r_{ij}F_{ij}$$

$$r_{ij} = r_i - r_j$$

Equation 2.28 Internal Viral for Pair-Addictive Potential

The pressure is calculated and shown in Equation 2.27, where Ξ is the internal viral for pair-addictive potential. F_{ij} is the force on particle i due to particle j.

$$\frac{dP}{dt} = -\frac{1}{\beta V}\frac{dV}{dt} = -\frac{3\alpha}{\beta}$$

Equation 2.29 Pressure change

The Pressure change in Equation 2.29 is related to the isothermal compressibility β. The equations can then be modified to define anisotropic systems, shown below in Equation 2.30.

$$P = \frac{1}{V}\left\{\sum_i m_i v_i v_i^T + \sum_{i<j} r_{ij} F_{ij}^T\right\}$$

Equation 2.30 Pressure for Anisotropic Systems

2.1.1.8 Solvation

In molecular dynamics, a water model is used to simulate and thermodynamically calculate water clusters, liquid water, or aqueous solutions. Several theoretical methods exist but the most prominent methods are implicit, explicit, vacuo, and hybrid solvation models. Under the explicit solvent category, the most commonly used water models are SPC/E, TIP3P, TIP4P, and TIP5P.[43] The TIP3P water model appears to be the most popular, while the SPC model appears very similar and shares TIP3P's minimalist form. Explicit solvent simulations are significantly more computationally expensive to use than the implicit or vacuo solvent simulations, therefore it is essential to reduce the computational complexity as much as possible. One way to do this is to use a triangulated water model, where the angle between the hydrogens is fixed.[43] This rigid

model is considered the simplest water models which relies on non-bonded interactions.[43] The bonding interactions are treated implicitly by a holonomic constraint.[43] The electrostatic interactions are modeled using Coulomb's law, while the dispersion and repulsion forces are defined using the Lennard-Jones potential as discussed in Section 2.1.1.3, they only incorporate polarization effects in an average sense.[43]

$$E_{ab} = \sum_{i}^{on\ a} \sum_{j}^{on\ b} \frac{k_C q_i q_j}{r_{ij}} + \frac{A}{r_{OO}^{12}} - \frac{B}{r_{OO}^{6}}$$

Equation 2.31 Energy of Solvation

In Equation 2.31 k_c is the electrostatic constant, q_i and q_j are the partial charges relative to the charge of the electron. r_{ij} is the distance between two atoms or charged sites. Lastly, A and B are the Lennard-Jones parameters, and r_{OO} is the radial distribution function.

2.1.2 Quantum Mechanics/Molecular Mechanics

In order to model an electronic rearrangement of ions during a chemical reaction of a simulation, a quantum mechanical (QM) description is required for the parts of a system that are explicitly involved in the reaction. For the remainder of the system however, a simple molecular mechanics (MM) force field model will describe the interactions with a classical approach. The interactions in the system are computed with a technique called a hybrid Quantum Mechanic/Molecular Mechanic (QM/MM) framework.[44] Biochemical systems are too large to be described at any level of ab-initio theory. At the same time, the available molecular mechanics force fields are not sufficiently flexible to model processes where chemical bonds are broken and formed.

In order to overcome these limitations, a full quantum mechanical description and a full molecular mechanics treatment are utilized in unison. Such methods have been developed that treat a small part of the system at the level of quantum chemistry (QM), while retaining the computationally cheaper force field (MM) for the larger. The justification for dividing a system into regions that are described at different levels of theory is the local character of most chemical reactions in condensed phases. A distinction can be made between a reaction center with atoms that are directly involved in the reaction and a spectator region, where atoms do not directly participate in the reaction.[44] For example, a reaction in solution can involve the reactants and the first few solvation shells. The bulk of the solvent hardly affects the reaction, but can influence the reaction via long-range interactions. The same is true for most enzymes, in which the catalytic process is restricted to an active site located somewhere inside the protein. The rest of the protein provides an electrostatic background that may or may not facilitate the reaction.

The QM/MM method provides both potential energies and forces.[44] With these forces, it is possible to perform a molecular dynamics simulation. However, because of the great computational costs required to perform ab initio calculations, the timescales that can be reached in QM/MM simulations is rather limited. At the ab-initio or DFT level, the limit is in the order of few hundreds of picoseconds. With semi-empirical methods (e.g., AM1[45, 46], PM3[47], or DFTB[48]) for the QM calculation, the limit is roughly 100 times longer. Therefore, unless the chemical process under consideration is at least an order of magnitude faster than the timescale that can be reached, an unrestrained MD simulation is not the method of choice to investigate that process.

Density-functional theory (DFT) is one of the most popular and successful quantum mechanical approaches. It is routinely applied to calculate the binding energy of molecules and the band structure of solids in physics.[44] DFT broadly spans use in studying superconductivity, atoms in the focus of strong laser pulses, relativistic effects in heavy elements and in atomic nuclei, classical liquids, and magnetic properties of alloys.[44] DFT owes this versatility to the generality of its fundamental concepts and the flexibility one has in implementing them. In spite of this flexibility and generality, DFT is based on a rigid conceptual framework.[44] In quantum mechanics we learn that all information we can possibly have about a given system is contained in the system's wave function, Ψ.[44] The nuclear degrees of freedom appear only in the form of a potential $v(r)$ acting on the electrons, so that the wave function depends only on the electronic coordinates.

$$\left[-\frac{\hbar^2\nabla^2}{2m} + v(r)\right]\Psi(r) = \epsilon\Psi(r)$$

Equation 2.32 Schrodinger's equation

This wave function is calculated from Schrodinger's equation for a single electron moving in a potential $v(r)$, shown in Equation 2.32.

$$\left[\sum_i^N \left(-\frac{\hbar^2\nabla_i^2}{2m} + v(r_i)\right) + \sum_{i<j} U(r_i, r_j)\right]\Psi(r_1, r_2 \ldots, r_N) = E\Psi(r_1, r_2 \ldots, r_N)$$

Equation 2.33 Schrodinger's Equation for Many Body System

If there is more than one electron, the Schrodinger's equation is calculated as shown in Equation 2.33. Here, N is the number of electrons, while $U(r_i, r_j)$ is the electron-electron interaction.

$$\hat{U} = \sum_{i<j} U(r_i, r_j) = \sum_{i<j} \frac{q^2}{|r_i - r_j|}$$

Equation 2.34 Operator for Coulomb Interactions

The operator for any system of particles interacting via the Coulomb interaction, is shown in Equation 2.34.

$$\hat{T} = -\frac{\hbar^2}{2m} \sum_i \nabla_i^2$$

Equation 2.35 Kinetic Energy Operator for Nonrelativistic Systems

The kinetic energy operator is the same for any nonrelativistic system. Whether the system is an atom, molecule, or solid depends only on the potential $v(r_i)$.

$$\hat{V} = \sum_i v(r_i) = \sum_i \frac{Qq}{|r_i - R|}$$

Equation 2.36 Relativistic Velocity of an Atom

For an atom, the Relativistic Velocity is defined in the Equation 2.36. Here Q is the nuclear charge and R is the nuclear position. When dealing with a single atom, R is usually taken to be the start of the coordinate system.

$$\hat{V} = \sum_i v(r_i) = \sum_{ik} \frac{Q_k q}{|r_i - R_k|}$$

Equation 2.37 Relativistic Velocity for a Molecule

For a molecule or a solid, the Relativistic Velocity is shown in Equation 2.37, where the sum on k extends over all nuclei in the system, each with charge $Q_k = Z_k e$ and position R_k. It is only the spatial arrangement, R_k, that distinguishes an atom from a solid structure. Similarly, it is only through the term U that the single-body quantum mechanics of equation (2.32) differs from the extremely complex many-body problem shown in equation (2.33). These properties are built into DFTB in a very fundamental process.

$$v(r) \overset{SE}{\Rightarrow} \Psi(r_1 r_2 \dots r_N) \overset{\langle\Psi|\dots|\Psi\rangle}{\Longrightarrow} observable$$

Equation 2.38 Summarized Schrodinger's equation

The usual quantum-mechanical approach to Schrodinger's Equation (SE) can be summarized and is shown in Equation 2.38. Specify the system by choosing $v(r)$, then plug it into Schrodinger's equation and solve the equation for the wave function Ψ, which then calculates observables by taking the expectation values of operators with this wave function.

$$n(r) = N \int d^3 r_2 \int d^3 r_3 \int d^3 r_N \Psi^*(r, r_2 \dots, r_N)\Psi(r, r_2 \dots r_N)$$

$$n(r) \Rightarrow \Psi(r_1 \dots r_N) \Rightarrow v(r)$$

Equation 2.39 Density-functional Equation

Among the observables that are calculated in this process, is the particle density. The density-functional approach can be summarized by the sequence in Equation 2.39, knowledge of $n(r)$ implies knowledge of the wave function and the potential, and hence of all other observables.

2.1.3 Statistical Mechanics

Statistical mechanics is the study of the system at the molecular level and then applying the observed characteristics and properties to the macroscopic behavior of systems.[49] In order to connect two different sizable states, an investigation starts by analyzing an ensemble. The ensemble is a collection of all the possible confirmations, different microscopic states but have identical thermodynamic properties. Then the states, macroscopic and microscopic systems, are connected by time independent statistical averages introduced to solve the problem.[49]

In statistical mechanics, average values are defined as ensemble averages. There are different types of ensembles with different characteristics. Microcanonical ensemble (NVE) is the thermodynamic state characterized by a fixed number of atoms N, a fixed volume V, and a fixed energy E.[49] This relates to an isolated system. Canonical Ensemble (NVT) is a collection of all the systems whose thermodynamic state is characterized by a fixed number of atoms N, a fixed volume V, and a fixed temperature T.[49] Isobaric-Isothermal Ensemble (NPT) is characterized by a fixed number of atoms N, a fixed pressure P, and a fixed temperature, T.[49] Grand canonical Ensemble (mVT) is characterized by a fixed chemical potential m, a fixed volume V, and a fixed temperature T.[49]

$$\langle A \rangle = \int dp^N dr^N A(p^N, r^N) \rho(p^N, r^N)$$

Equation 2.40 Ensemble Average

$$A(p^N, r^N)$$

Equation 2.41 Ensemble Observable of Interest

The ensemble average, as it refers to statistical mechanics is defined in Equation 2.40. In Equation 2.41 is the observable of interest expressed as a function of the momenta p, and the positions r of the system. The integration is executed over all the variables r and p.

$$p(p^N, r^N) = \frac{1}{Q} e^{\left[\frac{-H(p^N, r^N)}{k_B T}\right]}$$

Equation 2.42 Probability Density of the Ensemble

The probability density of the ensemble is shown in Equation 2.42 where H is the Hamiltonian, T is the temperature, k_B is Boltzmann's constant and Q is the partition function.

$$Q = \int \int dp^N dr^N e^{\left[\frac{-H(p^N, r^N)}{k_B T}\right]}$$

Equation 2.43 Partition Function

The integral in Equation 2.43 is generally difficult to calculate because it requires calculation of all the possible states of the system. In a molecular dynamics simulation, the points in the ensemble are calculated sequentially in time, so to calculate an ensemble average,

the molecular dynamics simulations must pass through all possible states corresponding to the particular thermodynamic constraints.

$$\langle A \rangle_{TIME} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A\big(p^N(t), r^N(t)\big) dt \approx \frac{1}{M} \sum_{t=1}^{M} A(p^N, r^N)$$

Equation 2.44 Alternative Ensemble Average

Another method to determine the time average A, of a molecular dynamics simulation is shown in Equation 2.44. Here, t is the simulation time, M is the number of time steps in the simulation, and $A(p^N, r^N)$ is the instantaneous value of A. The only apparent drawback appears to be that it is possible to calculate time averages utilizing a molecular dynamics simulation, but the experimental observables are assumed to be ensemble averages. This leads to one of the most fundamental axioms of statistical mechanics, the ergodic hypothesis. This theory states that the time average equals the ensemble average.[49] If the system is allowed to evolve indefinitely in time, that system will eventually sample all the possible states.[49] The objective of conducting molecular dynamics simulation is to run long enough to generate enough representative conformations so that this equality is satisfied. If this is the case, experimentally relevant information concerning structural, dynamic and thermodynamic properties can then be calculated.

$$E_{PE} = \langle E_{PE} \rangle = \frac{1}{M} \sum_{i=1}^{M} V_i$$

Equation 2.45 Average Potential Energy

$$E_{KE} = \langle E_{KE} \rangle = \frac{1}{M} \sum_{j=1}^{M} \left\{ \sum_{i=1}^{N} \frac{m_i}{2} v_i v_i \right\}_j$$

Equation 2.46 Average Kinetic Energy

The Equation 2.45 shows how to calculate the Average Potential Energy. here M is the number of configurations in the molecular dynamics trajectory and V_i is the potential energy of each configuration. The Equation 2.46 shows how to calculate the Average Kinetic Energy. Here, M is the number of configurations in the simulation, N is the number of atoms in the system, m_i is the mass of the particle i, and v_i is the velocity of particle i.

2.2 Methods

2.2.1 RMSD

The Root Mean Squared Deviation (RMSD) is the most common method of comparing the differences between two or more molecular structures. It uses a least-squares fit procedure by mathematically assessing the average distance between two atoms. Coupled with the uses of MD analysis, an ensemble of structures is gathered by superimposing a set of snapshots and calculating the mean distances from the sum of each of the Cartesian coordinates of the atom divided by the number of structures, as shown in Equation 2.47.

$$RMSD = \sqrt{\frac{1}{N} \sum \left(r_{ij1} - r_{ij2} \right)^2}$$

Equation 2.47 RMSD Equation

In this equation, N is the number of snapshots in the trajectory, while r is the distance between the atom i and a reference structure of the N equivalent atom.

2.2.2 MMPBSA

MMPBSA stands for Molecular Mechanics/Poisson-Boltzmann Surface Area and it is a method used to calculate the binding free energy of biomolecules, protein-ligand interactions from molecular dynamic simulations.[50] This method is a reasonable calculation with good accuracy while being less computationally expensive, compared to other techniques like full-scale molecular dynamics Free Energy Perturbation (FEP)/Thermodynamic Integration (TI) calculations.[51] With MMPBSA, the free energy of a molecule is estimated as the sum of its gas-phase energy, an implicit solvent model.[52, 53] The gas-phase energy is approximately calculated by the molecular mechanics energy of the molecule. It is determined from a force field with defined terms for bond, angle, torsion energy, Van-der-Waals and electrostatic interactions.

The binding free energy can be calculated as the difference between the free energy of a complex and the sum of the free energies of its components.[54]

$$\Delta G_{bind} = G_{complex} - G_{protein} - G_{ligand}$$

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

Equation 2.48 Binding Free Energy

In Equation 2.48, $G_{complex}$, $G_{protein}$, and G_{ligand} refer to the free energies of the complex, protein, and ligand, respectively, and ΔE_{MM} is the change in the molecular mechanical

energy, ΔG_{PB} is the change in the solvation free energy determined with the Poisson-Boltzmann model, $\Delta G_{nonpolar}$ is the change in the nonpolar free energy based on a surface area calculation, and $T\Delta S$ is the temperature of the simulation multiplied by the change in entropy, which is determined based on snapshots from the MD simulations. When utilizing a conformational ensemble, from a MD simulation, each energy component is determined by an average over the respective energy contributions from all conformations of the ensemble.

2.2.3 CURVES+

CURVES+ is specialized software written to analyze conformations of nucleic acid structures. It helps the study of comparing helical parameters of regular and irregular structures in detail. This software also has the ability to generate visual models of the helical axis and groove geometry. CURVES+ first creates a reference frame by initially identifying each base pair, pyrimidine or purine, of a given PDB structure. There are distinguished vector points used when characterizing a pyrimidine and purine. The software then attempts to calculate parameters, intra-base pair, inter-base pair, helical axis, base pair-axis, helical rise and twist, backbone and groove geometry. Within each set of parameters, there are different techniques to calculate the necessary data. CURVES+ is not limited to the classical double stranded antiparallel nucleic acid structure with standard base pairing. There is also support for single to four-stranded DNA with any number of orientations, or with any atypical lengths and gaps.

2.2.4 Principle Component Analysis (PCA)

Principal Component Analysis (PCA) is a statistical technique, which is widely used to reduce the dimensionality of a data set. It retains the most variance, by finding patterns within the data set. PCA searches for linear combinations with the largest variances, and divides them into Principal Components (PC) where the largest variance is captured by the highest component in order to extract the most important information. At some point, the plot begins to lose its structure and becomes noise or a distribution of Poisson-Bolztman equation

PCA[55] was performed on all the heavy atoms relating to the KB DNA gene and also the backbone of the Pirin. The covariance matrix of the x, y, and z coordinates of the KB DNA heavy atoms were obtained from each snapshot of the combined trajectories of the free DNA, DNA-p65 complex, and the Supramolecular complex were calculated. The same was done separately for the simulations involving Pirin, DNA-p65 and Supramolecular complex. The covariance matrix was further diagonalized to produce orthonormal eigenvectors and their corresponding eigenvalues, ranked on the basis of their corresponding variances. The first five eigenvectors were calculated, the first three principal components that contributed the majority of all the atomic fluctuations, were used to project the conformational space onto them, plotted along two dimensions.

2.2.5 Experimental Details

All simulations were carried out using the CUDA version of pmemd module in the Amber 14[56] suite of programs and the ff14SB[57] modified version of the Cornel et al.[58] force field. Each simulation was run on a single NVIDIA GeForce GTX 980 GPU. Each system was

solvated in a periodic octahedron box of explicit TIP3P[59] water model. The simulations were run at a constant pressure of 1 bar and a constant temperature of 300 K. The SHAKE algorithm[60] was used to constrain all bonds involving hydrogen. The Langevin thermostat was used to regulate the temperature[61] of the system at 300 K with a collision frequency of 1.0 ps^{-1}. All short-range non-bonded interactions were calculated within a cutoff of 9 Å, and all long-range electrostatic interactions were calculated using the particle mesh Ewald summation method.[62] A time step of 2 fs was used to integrate the Langevin equation of motion. The VMD[63] software was used for all graphical representations.

The initial coordinates of the homodimeric p65-DNA complex came from a 2.4 Å resolution x-ray crystal structure with Protein Data Bank (PDB) ID 2RAM.[1] The initial coordinates of Pirin were also obtained from a x-ray crystal structure of the ferric form with PDB ID 4EWA.[22] The tleap module in Amber 14 was used to add the missing hydrogen atoms. The proposed Pirin-p65-DNA complex was constructed by carrying out rigid docking studies of the Fe(III) form of Pirin (PDB ID 4EWA) to the p65-DNA complex (PDB ID 1RAM)[1] using the ZDOCK webserver.[64] The binding interface was restricted to the region containing R23, E32, and K34 on the Pirin and E282, R273 and E234 on p65, as was also previously described by Liu et al.[22] and Barman and Hamelberg.[23] Single point R23E, E32V or K34V mutation of Pirin bu Liu et al. reduced the binding affinity of p65 for the DNA by more than half, when compared to wild type Pirin, suggesting that this region on Pirin was the interacting region for p65.

The free DNA, p65-DNA complex, and Pirin-p65-DNA complexes were solvated in a periodic truncated octahedron water box with the edges of the box at least 10 Å away from any

part of the system. Each system was neutralized with Na^+ counter ions to bring the charge of the entire system to zero. The systems were equilibrated with a series of minimization and molecular dynamics simulations. Applying a harmonic constraint only on the solute with a force constant of 300, 200, 100, and 50 kcal/mol/Å^2 was used to carry out a series of minimization steps. A final minimization step was carried out without applying any harmonic constraint. The entire system was heated from 0 - 300 K using molecular dynamics simulation by applying a harmonic constraint on the solute with a force constant of 100 kcal/mol/Å^2. Three additional molecular dynamics simulations with harmonic constraint and force constants of 75, 50, and 0 kcal/mol/Å^2 were carried out. All of the systems were equilibrated for at least 2 ns. Each system was then simulated for an additional 1.3 μs. The first 300 ns of snapshots were discarded and considered as additional equilibration. The snapshots of the trajectories were saved after every 500 steps or 1 ps of integration time.

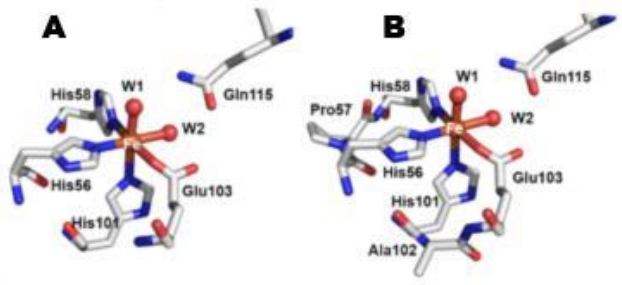

Figure 2.1 Coordinated Iron QM region
Residues included in the QM region and the RESP charge fitting procedure. The rest of the protein (not shown) was treated with MM. The hydrogen atoms are not shown but are included in the calculations. (A) Residues treated in the QM region during QM/MM minimization. (B) Residues included in the QM region during a single point energy calculation to obtain the electrostatic potential and to calculate the partial charges of the atoms using the RESP charge fitting procedure.

The force field parameters for the iron center of the Fe(II) and Fe(III) forms of Pirin were obtained as was previously described.[23] Specifically, the Fe(III) form of Pirin was optimized

using the Quantum Mechanics/Molecular Mechanics (QM/MM) method implemented in the AMBER14 package[56] and the Gaussian03[65] as an external program for the quantum mechanical calculations. The QM region consisted of His56, His58, His101, Glu103, Gln115, two water molecules, and the Fe atom (Figure 2.1). The histidine residues coordinated to the Fe atom were mono-protonated at the δ-nitrogen. The QM/MM minimization was performed with the QM region treated at the B3LYP level of theory and using the LANL2DZ Hay−Wadt effective core potential as the basis set.[66-69] This was followed by calculation of the electrostatic potential and derivation of the partial charges using the RESP partial charge fitting procedure.[35] The minimized geometry of the iron, the coordinated residues and two additional residues Pro57 and Ala102 were included in calculating the electrostatic potential, preceding the RESP charge fitting procedure. The electrostatic potential was calculated at the M06 level of theory and all electron 6-31G(d) basis set. The partial charges were used in the molecular dynamics simulations. The spin states of Fe(III) and Fe(II) were set as high spin and low spin, respectively, based on suggestion from experiments.[22] Table 1.1 shows the partial charges obtained for the coordinating residues and iron in the Fe(II) and Fe(III) forms of Pirin. The rest of the protein was treated with the AMBER14 ff14sb force-field parameters.

Table 1 Partial charges of the atoms in QM region
Partial charges of the atoms in the residues at the iron center of the Fe(II) and Fe(III) forms of Pirin.

	Fe(II)			**Fe(III)**		
	His56	**His58**	**His101**	**His56**	**His58**	**His101**

N	0.2824	-0.7333	-0.0965	0.0814	-0.6056	0.0412
H	0.0970	0.3954	-0.0846	-0.0252	0.3355	-0.1036
CA	0.5477	0.0810	0.1301	-0.3582	-0.2237	-0.0235
HA	-0.0379	0.0784	0.0978	0.2476	0.1578	0.1191
CB	-0.5482	-0.3323	-0.4762	-0.5614	-0.1528	-0.4717
HB2	0.1571	0.1165	0.1781	0.2238	0.0697	0.2141
HB3	0.1571	0.1165	0.1781	0.2238	0.0697	0.2141
CG	0.3847	0.3728	0.2258	0.2374	0.3986	0.0443
ND1	-0.3467	-0.5428	-0.3489	-0.0604	-0.4977	-0.1582
HD1	0.3410	0.4290	0.3686	0.2875	0.4278	0.3494
CE1	0.0177	0.0705	0.0261	-0.1367	0.0787	-0.1053
HE1	0.1823	0.1433	0.1837	0.2225	0.1563	0.1963
NE2	-0.1666	0.0997	-0.2988	-0.1578	-0.0656	-0.1772
CD2	-0.3555	-0.4448	-0.0859	-0.2467	-0.3720	0.0084
HD2	0.2045	0.2655	0.1181	0.1842	0.2692	0.0898
C	-0.3648	0.6331	0.2613	0.3444	0.7851	0.3804
O	-0.1110	-0.4809	-0.3354	-0.3758	-0.4930	-0.4288
	Pro57			**Pro57**		
N	0.4385			0.0659		
CD	-0.0119			0.0804		
HD2	0.0539			0.0353		
HD3	0.0539			0.0353		
CG	-0.3127			-0.2138		
HG2	0.1237			0.1059		
HG3	0.1237			0.1059		
CB	0.0759			0.0388		
HB2	0.0584			0.0421		
HB3	0.0584			0.0421		
CA	-0.5379			-0.2968		
HA	0.1279			0.1156		
C	0.7397			0.6765		
O	-0.4995			-0.5394		
	Ala102			**Ala102**		

N	-0.5091	-0.4409
H	0.3007	0.3005
CA	0.3273	0.1850
HA	-0.0366	0.0260
CB	-0.4494	-0.4148
HB1	0.1271	0.1306
HB2	0.1271	0.1306
HB3	0.1271	0.1306
C	0.7539	0.8220
O	-0.5609	-0.5731
	Glu103	**Glu103**
N	-0.8500	-0.8863
H	0.4221	0.4376
CA	-0.1017	-0.4297
HA	0.1673	0.2893
CB	-0.0466	-0.0040
HB2	0.0354	0.0452
HB3	0.0354	0.0452
CG	-0.1095	-0.0235
HG2	0.0502	0.0449
HG3	0.0502	0.0449
CD	0.5381	0.5159
OE1	-0.5466	-0.4714
OE2	-0.5271	-0.4566
C	0.6640	0.8759
O	-0.5146	-0.5644
	Wat1	**Wat2**
O	-0.8791	-0.8533
H1	0.4823	0.4872
H2	0.4823	0.4872
	Wat2	**Wat2**
O	-0.9599	-0.9097
H1	0.4882	0.5009
H2	0.4882	0.5009
	Gln115	**Gln115**
N	-0.2853	-0.2038
H	0.2082	0.2201
CA	-0.2083	-0.4373
HA	0.2117	0.2710

CB	0.0836	-0.2229
HB2	0.0025	0.1166
HB3	0.0025	0.1166
CG	-0.2601	-0.1657
HG2	0.0607	0.0687
HG3	0.0607	0.0687
CD	0.8261	0.6725
OE1	-0.6308	-0.4808
NE2	-1.0033	-0.9378
HE21	0.4494	0.3896
HE22	0.4494	0.3896
C	0.5755	0.7134
O	-0.4681	-0.5092
	Fe(II)	**Fe(III)**
Fe	0.6724	1.007

The trajectories were mainly analyzed using the cpptraj module[70] in Amber 14. The Cartesian Principal Component Analysis[55] was carried out using cpptraj by combining the trajectories of the DNA from all the systems, so that a common set of eigenvectors could be used to describe the motions in all of the systems. The average groove widths and the helical parameters of the DNA in the different states were calculated using the CURVES+ program[71]. Statistical analysis was carried out using the 1-way analysis of the variance (ANOVA) and the multiple comparison test using Tukey's method. Also, we split each data set into four equal blocks and take the average of each block to estimate the errors and check for convergence.

3 RESULTS

Liu et al. have used Surface Plasmon Resonance (SPR) experiments to show that the Fe(III) form of Pirin increases the affinity of p65 for the DNA by more than 25 fold.[22] The effect of Pirin was studied by measuring the level of binding of p65 to the DNA in the presence and

absence of Pirin. It is proposed that Pirin, by itself, does not bind to the DNA. They also carried out mutational studies on the surface of Pirin to narrow down the region that interacts wih p65. Single point R23E, E32V or K34V mutation on Pirin reduces the binding affinity between p65 and the DNA by more than half, when compared to wild type Pirin, suggesting that this region interacts with p65. The suggested binding region on Pirin has been shown to undergo conformational changes upon going from the inactive (ferrous) form to the active (ferric) form in x-ray crystallographic studies.[1, 22] This suggested binding region corresponds to the same area that undergoes major conformational changes in the recent comparative molecular dynamics studies between the Fe(II) form of Pirin and the Fe(III) form.[23]

As a result, we carried out docking studies of the Fe(III) form (ferric) of Pirin (PDB ID 4EWA) to the p65-DNA complex (PDB ID 1RAM)[1] using ZDOCK,[64] restricting the binding region of Pirin to the surface containing R23, E32 and K34. Figure 1.2B shows the predicted binding region of Pirin and the proposed Pirin-p65-DNA supramolecular complex. Pirin binds complementarily to the space between the two domains on the monomeric p65 (Figure 1.2C). R23, E32, and K34 on Pirin are in close proximity to E282, R273 and E234 on p65, respectively. The docked complex provided the initial starting coordinates for the simulations of the supramolecular complex, since there are no x-ray crystallographic or NMR structures of Pirin in a ternary complex with the p65-DNA assembly. The Pirin-p65-DNA supramolecular complex in explicit water was approximately 180,000 atoms, and the Fe(III) and Fe(II) forms of the complex were each simulated for 1.3 μs. The p65-DNA complex (~120,000 atoms) without Pirin, and the free DNA (~40,000 atoms) were also simulated in explicit water, each for 1.3 μs. The root-mean-square deviations of the different systems are shown in Figures 3.1.1 and 3.1.2. All of the

systems equilibrated within the first 300 ns of simulation time, so all of the analyses were carried out on the last 1.0 μs of the simulations.

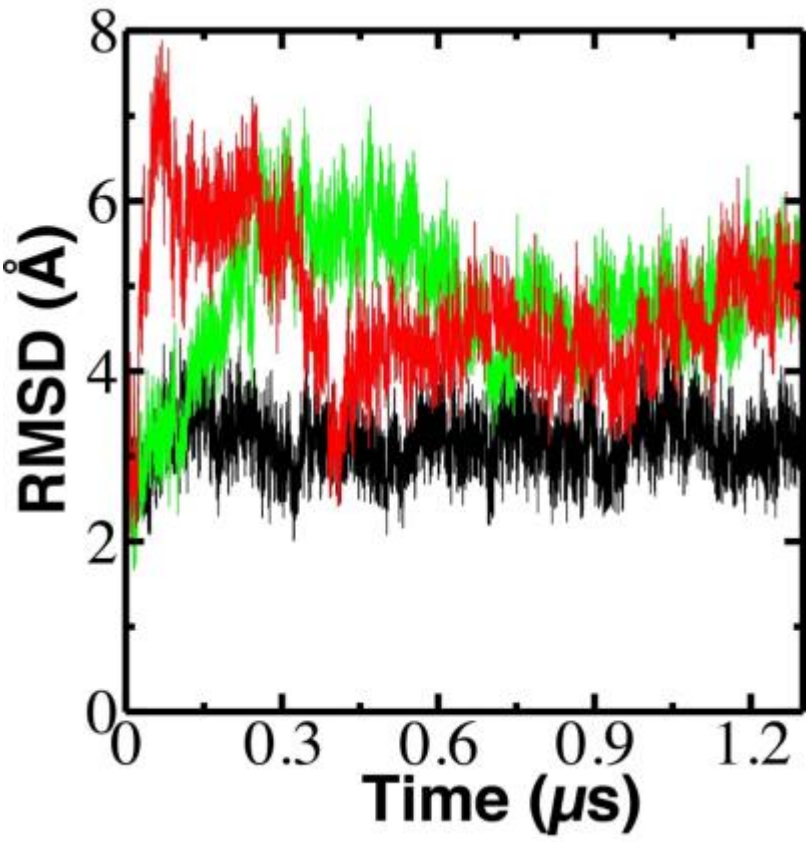

Figure 3.1 Root mean square deviation of the Cα atoms
Root mean square deviation of the Cα atoms of the NF-κB of the p65-DNA complex (black), the Fe(III) form of the Pirin-p65-DNA supramolecular complex (green), and the Fe(II) form of the Pirin-p65-DNA supramolecular (red) during the entire 1.3 μs MD simulation.

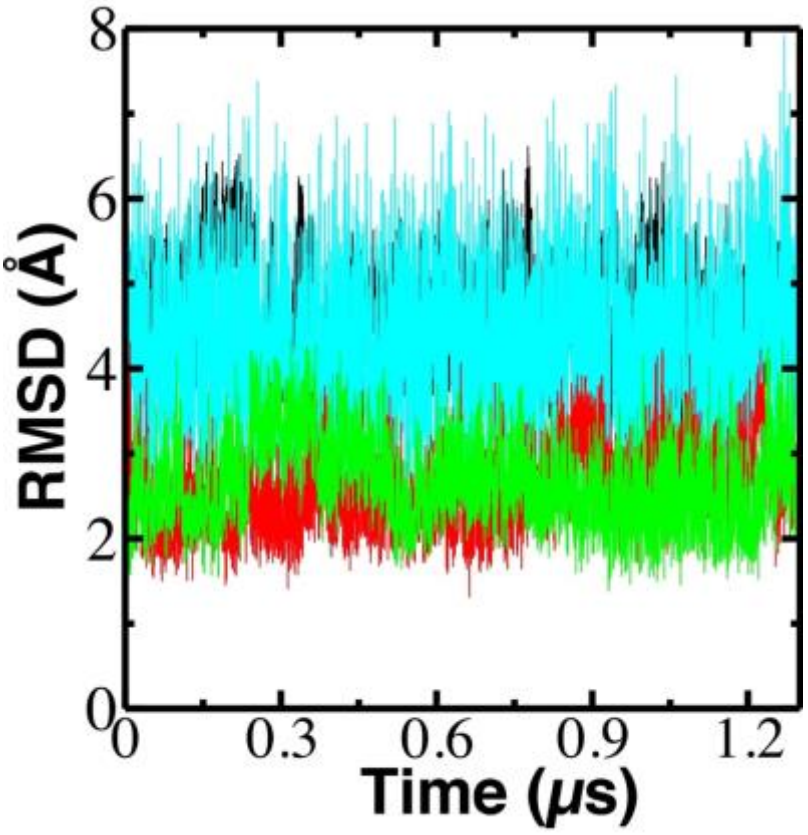

Figure 3.2 Root-mean-square deviation (RMSD) of the heavy atoms of the DNA
Root-mean-square deviation (RMSD) of the heavy atoms of the free DNA (cyan), DNA in the p65-DNA complex (black), DNA in the Fe(III) form of the Pirin-p65-DNA supramolecular complex (green), and DNA in the Fe(II) form of the Pirin-p65-DNA supramolecular complex (red) during the entire 1.3 μs MD simulation.

3.1 Fe(III) form modulates interactions between NF-κB and DNA

Inherently, the regulation of p65 by a co-regulator, such as Pirin, would be expected to modulate the interactions between p65 and the DNA. We therefore investigate the effect of Pirin on the interactions between p65 and the DNA by analyzing the propensity and dynamics of the residue-residue contacts between p65 and the DNA. We considered a residue-residue contact between the p65 and the DNA to be formed if any two inter-residue heavy atoms are within 4.5 Å. We calculate the contact probability between any two residues in p65 and the DNA. The

majority of these contacts are never formed in the p65-DNA complex, and a small percentage of contacts at the interfacial region are either always formed or are dynamic contacts, meaning they form and break during the simulations. A contact is considered to be a dynamic contact if it is formed more than 10% and less than 90% of the total simulation time. We calculate the difference in probabilities between the dynamic contacts of the p65-DNA complex and the Fe(III) form of the Pirin-p65-DNA supramolecular complex as shown in Figure 3.3. The probabilities of forming the dynamic contacts between the p65 and the DNA in the p65-DNA complex are subtracted from those of the Fe(III) form of Pirin-p65-DNA supramolecular complex; therefore, contacts that are more formed upon binding the Fe(III) form of Pirin are positive (blue) and contacts that are less formed upon binding the Fe(III) form of Pirin are negative (red). The width of the cylinders in Figure 2A is proportional to the magnitude of the difference. Some contacts between NF-κB and the DNA are more formed and some contacts are less formed upon binding the Fe(III) form of Pirin, as shown in Figure 2A. In general, the magnitudes of the differences in the contacts that are more formed are greater than those that are less formed, according to the width of the cylinders (Figure 2A). The results suggest that stronger contacts are more formed between NF-κB and the DNA upon binding the Fe(III) form of Pirin to form the supramolecular complex through modulation of the dynamics of NF-κB. Eleven residue-residue contacts are more formed with a probability difference above 50% (Figure 2B), as compared to only six residue-residue contacts that are less formed with a similar probability difference. Notably, Arg 28 and Arg 35 on p65 make tighter contacts with the DNA and were further inserted into the major groove (Figure 2B) upon binding the Fe(III) form of Pirin. The results suggest that stronger contacts are more formed between the p65 and the DNA upon binding the Fe(III) form of Pirin to form the supramolecular complex through modulation

of the dynamics of p65. Interestingly, formation of the supramolecular complex upon binding the Fe(II) form of Pirin behaves similarly to the p65-DNA complex and does not significantly change the contact dynamics between p65 and the DNA significantly as shown in Figure 3.3B.

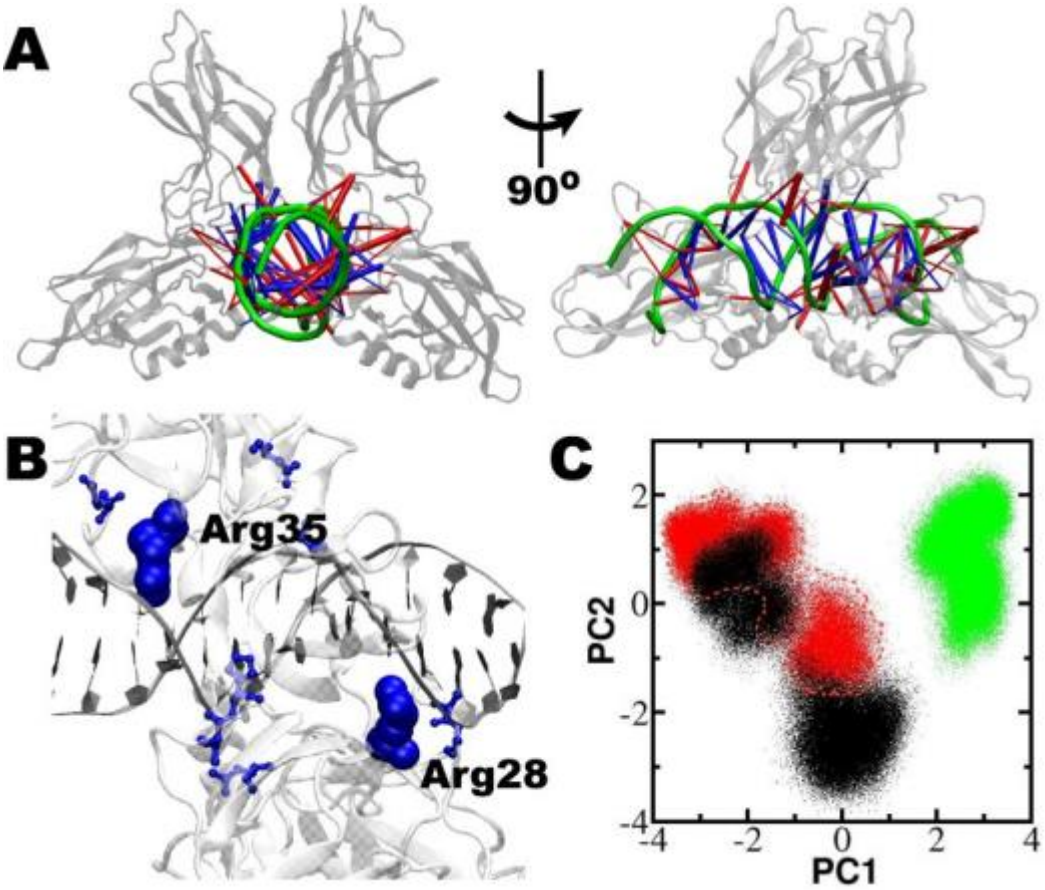

Figure 3.3 Residue-residue contact dynamics
Residue-residue contact dynamics between p65 and the DNA in the p65-DNA and Pirin-p65-DNA supramolecular complexes. (A) Residue-residue contact that are more formed (blue) and less formed (red) upon binding of the Fe(III) form of Pirin to form the Pirin-p65-DNA supramolecular complex. The width of the cylinder represents the magnitude of the difference in probabilities, and only differences above 20% are shown. (B) Residues on p65 that form tighter contacts with the DNA with a difference of above 50%. Arg 28 and Arg 35 are further inserted into the major groove of the DNA upon binding the Fe(III) form of Pirin to form the Pirin-p65- DNA supramolecular complex. (C) The two dominant Principal Components of the dynamic contacts between p65 and the DNA of the p65-DNA complex (black), the Fe(III) form of Pirin- p65-DNA supramolecular complex (green), and the Fe(II) form of Pirin-p65-DNA

supramolecular complex (red). Each dot represents a conformation of the respective complex.

We further probe the interfacial dynamics between p65 and the DNA by generating trajectories of the dynamic contacts between p65 and the DNA and performing Principal Component Analysis (PCA) on the contact trajectories of the p65-DNA complex, the Fe(II) form of Pirin of the supramolecular complex, and the Fe(III) form of the supramolecular complex. Details of the approach have been previously described.[72, 73] More specifically, a dynamic contact is given a '1' when formed and a '0' when not formed. At each time point of the simulation, a contact trajectory is made up of a binary representation of the dynamic contacts, reducing the dimensionality of the description of the interactions between p65 and the DNA.

Figure 3.3B shows the distributions of the interfacial residue-residue contact dynamics in contact space projected on the top two Principal Components (PCs). Each dot represents a confirmation of the respective complex. Three things can be gleaned from the analysis. Firstly, binding of the Fe(III) form of Pirin dramatically changes the dynamics of the interfacial contacts between p65 and the DNA. The distributions of the contact dynamics of the p65-DNA complex and the Fe(III) form of Pirin-p65-DNA supramolecular complex in contact space are quite different and distinct, suggesting that binding of the Fe(III) form of Pirin alters the interactions between p65 and the DNA. Secondly, the contacts are more localized and well formed in the Fe(III) form of the Pirin-p65-DNA supramolecular complex (green cluster in Figure 3.3B) than those of the p65-DNA complex (black cluster in Figure 3.3B). In the p65-DNA complex, the contacts can easily form and break, suggesting a higher stochastic probability of p65-DNA to be in the dissociated state when compared to the Fe(III) form of the Pirin-p65-DNA supramolecular

complex. Lastly, the distribution of contacts in the Fe(II) form of the Pirin-p65-DNA supramolecular complex (red cluster in Figure 3.3B) is similar to that of the p65-DNA complex (Figure 3.3C), suggesting that p65 has a similar probability of dissociating from the DNA in the Fe(II) form of the Pirin-p65-DNA supramolecular complex as in the p65-DNA complex. Figure 3.3C shows that the distributions of residue-residue interfacial contacts in the p65-DNA complex and the Fe(II) form of Pirin-p65-DNA supramolecular complex are similar and overlap. The results suggest that the formation of the Fe(III) form of Pirin-p65-DNA supramolecular complex, and not the Fe(II) form, localizes p65 on the DNA and enhances the interactions between p65 and the DNA in the supramolecular complex. These results therefore suggest that the Fe(II) form of Pirin does not localize p65 on the DNA due to its much looser interfacial dynamic contacts.

The Fe(II) form of Pirin was shown to weakly bind to p65 due partly to the lack of electrostatic complementarity between the Fe(II) form of Pirin and the p65.[23] The results of the present studies also suggest that the Fe(III) form of Pirin binds more strongly to p65, inducing tighter interactions between p65 and the DNA. Figure 3.4 shows that the movement of the Fe(II) form of Pirin on the p65 samples more conformational space than the Fe(III) form of Pirin. We carried out Principal Component Analysis (PCA) on the Cartesian coordinates of the backbone of each Pirin molecule on the Fe(II) and Fe(III) forms of the supramolecular complexes after aligning the corresponding p65 monomer. The results captures both the backbone dynamics, translational, and rotational dynamics of the Pirin molecules in the supramolecular complexes. The Fe(III) form of Pirin is more tightly bound to p65 than the Fe(II) form as shown in Figure 3.4. The results suggest looser interactions between the Fe(II) form of Pirin and p65, leading to

little or no effect on the interactions between p65 and the DNA in the Fe(II) form of the supramolecular complex.

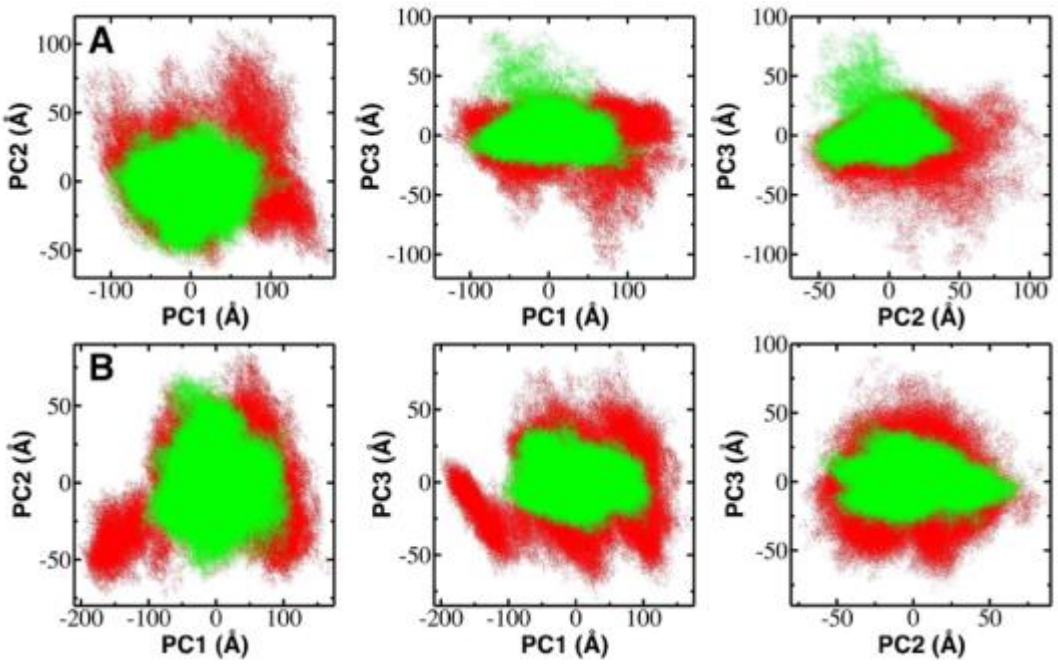

Figure 3.4 Principal Component Analysis of the motions of Pirin
Principal component Analysis of the motions of Pirin in the supramolecular complexes. Projection of the top three principal components of the Fe(III) form of Pirin (green) and the Fe(II) form of Pirin (red) in the Pirin-p65-DNA supramolecular complexes on either side of the side of the supramolecular complex (A) and (B). Each dot represents a conformation.

3.2 Modulation of the interactions lead to higher affinity

We hypothesize that changes in the residue-residue contacts could alter the binding affinity between p65 and the DNA. In determining the consequences these changes upon binding the Fe(III) form of Pirin, we calculate the distributions of binding free energies between p65 and the DNA in the p65-DNA complex and the Fe(II) and Fe(III) forms of the Pirin-p65-DNA supramolecular complexes. The affinity between p65 and the DNA is estimated by calculating the binding free energy for every 200 snapshots in the microsecond trajectories using the

Molecular Mechanics/Poisson Boltzmann Surface Area (MM/PBSA) method.[74] Each snapshot provides a binding free energy for that particular conformation; therefore, each trajectory provides a distribution of binding free energies that is shown in Figure 3.5 for the p65-DNA and the Pirin-p65-DNA supramolecular complexes. Figure 3.5 shows that - upon binding the Fe(III) form of Pirin to p65 - the interactions between the p65 and the DNA are not only altered, as shown in Figure 3.3, the distribution of binding free energies are also shifted to lower values. On average, the binding free energy between p65 and the DNA in the p65-DNA complex goes from approximately -22.1±5.3 kcal/mol to approximately -32±3.2 kcal/mol in the Fe(III) form of the supramolecular complex, suggesting that binding of the Fe(III) form of Pirin enhances the binding of p65 to the DNA – in line with experiments.[22] On the other hand, the Fe(II) form of Pirin does not induce a significant change in the binding affinity between p65 and the DNA with an average binding free energy of approximately -20.5±1.5 kcal/mol, similar to that of p65-DNA complex. The distribution of the binding energies of the p65-DNA complex is very similar to that of the Fe(II) form of the Pirin-p65-DNA supramolecular complex as shown in Figure 3.5.

Splitting each data set into four equal blocks and taking the average of each block was used to estimate the errors. Also, the significance of the differences in the distributions in Figure 3 was analyzed using a 1-way analysis of the variance (ANOVA). The p-value was $<10^{-16}$, showing that at least one of the distributions is significantly different from the others. A multiple comparison test using Tukey's method shows that the average binding free energy of the Fe(III) form of the Pirin-p65-DNA complex is significantly different from that of the p65-DNA complex and the Fe(II) form of the Pirin-p65-DNA complex with p-values of $<10^{-7}$. The average binding free energy of the Fe(II) form of the Pirin-p65-DNA complex is more similar to that of the p65-DNA complex with a p-value of approximately 10^{-3}. Estimating the binding affinity between p65

and the DNA to assess whether the effect of Pirin is in agreement with experiments is also a test of the model of the supramolecular complex, and the results provide a reassurance of the validity of the model. In general, the results suggest that binding of the Fe(III) form of Pirin to p65 changes the dynamical contacts between p65 and the DNA by strengthening several of the residue-residue contacts, leading to a decrease in the binding free energy between p65 and the DNA. The enhanced binding of p65 to the DNA could in turn affect the conformational dynamics of the DNA. In the cell, modulation of the interactions between p65 and the DNA by a co-regulator could lead to fine-tuning of biological function, such as the transcriptional level of genes, through modulation of the local conformational dynamics of the DNA that could propagate to other regions of the DNA.

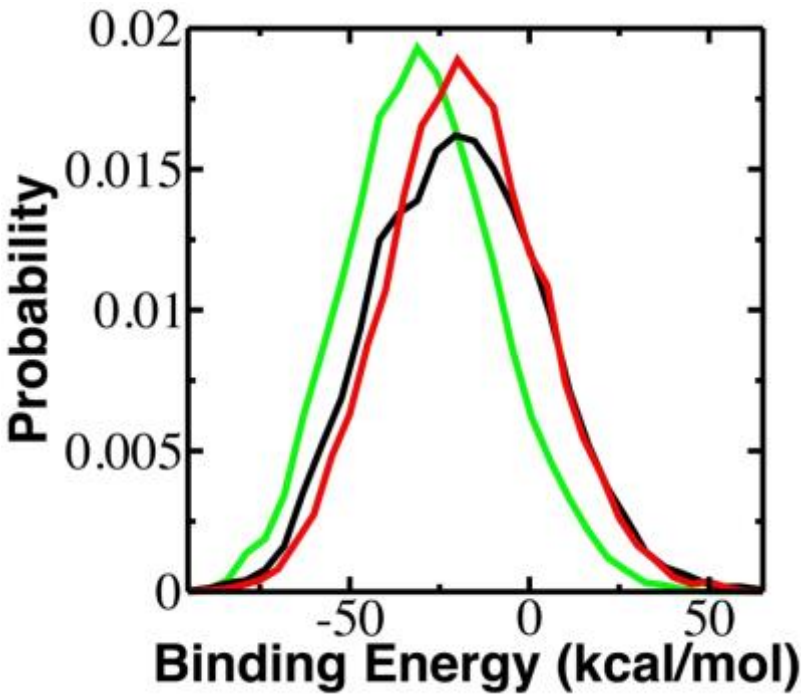

Figure 3.5 Binding Free Energies between p65 and the DNA
Distributions of the binding free energies between p65 and the DNA in the p65-DNA complex (black), the Fe(II) form of the Pirin-p65-DNA supramolecular complex (red), and the Fe(III) form of the Pirin-p65-DNA supramolecular complex (green).

3.3 Fe(III) form alters dynamics and conformations of the DNA

The binding of proteins to DNA often causes local changes in the conformations of the DNA[75, 76], affecting the groove widths and helical parameters.[75] NF-κB binds to the major groove of the DNA and narrowing the minor groove as shown in Figures 3.6A and 3.7. On average, the minor groove width goes from ~5.0 Å to ~4.0 Å around the central binding region of p65 binding site on the DNA. A further narrowing of the minor groove of the DNA at the binding site of p65 to ~2.5 Å is observed up binding of the Fe(III) form of Pirin to p65 in the supramolecular complex (Figures 3.6A and 3.7). Interestingly, the binding of the Fe(II) form of Pirin does not have any significant effect on the minor groove of the DNA at the binding site of p65, with the minor groove width similar to that of the p65-DNA complex, also shown in Figures 3.6A and 3.7. The narrowing of the minor groove of DNA is usually accompanied by changes in helical parameters, such as the propeller twist.[77] The narrower the minor groove the greater the magnitude of the propeller twist (Figure 3.6B). The propeller twist of the free DNA is ~ -15^{o} and is ~ -22^{o} upon interacting with p65. Upon binding the Fe(III) form of Pirin to p65, the propeller twist changes further to ~ -28^{o} at the binding site of p65. Again, binding of the Fe(II) form of Pirin to p65 has little or no effect on the propeller twist at the binding site as shown in Figure 3.6B.

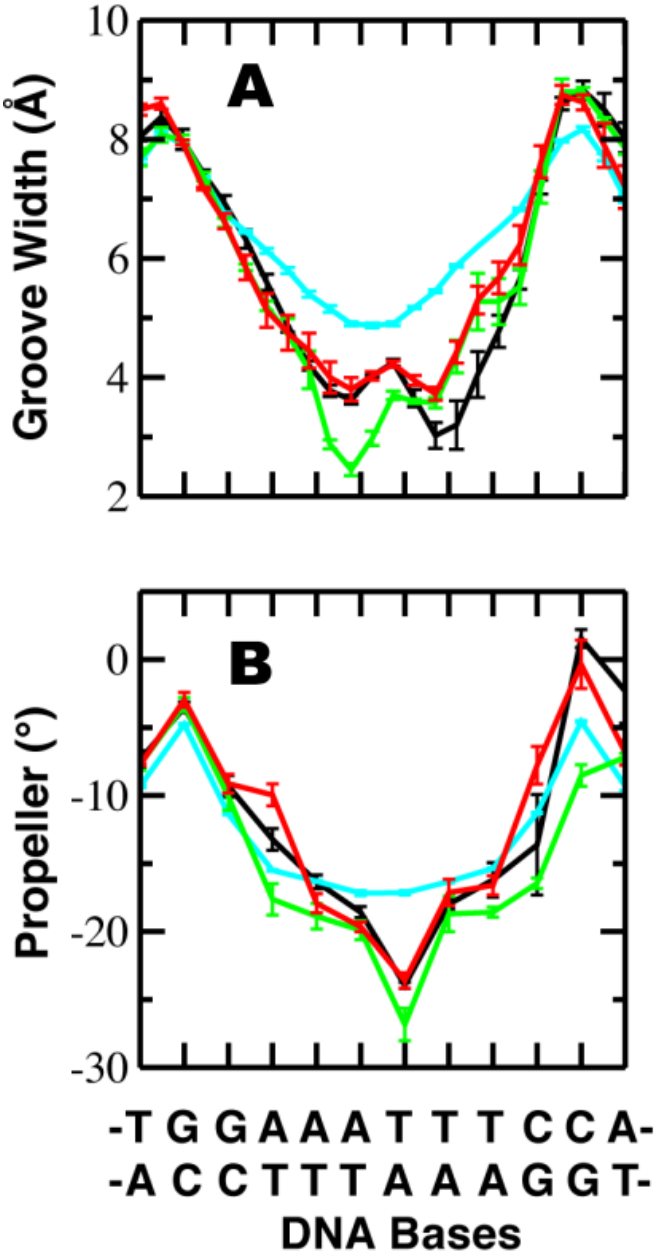

Figure 3.6 Average helical parameters of the DNA
Average helical parameters of the DNA around the binding region of p65. (A) Minor groove width of and (B) propeller twist of the free DNA (cyan), the DNA in the p65-DNA complex (black), the DNA in the Fe(III) form of the Pirin-p65-DNA supramolecular complex (green), and the DNA in the Fe(II) form of the Pirin-p65-DNA supramolecular complex (red). The error bars were calculated by splitting the simulation trajectories into four equal parts and computing the average helical parameters for each block.

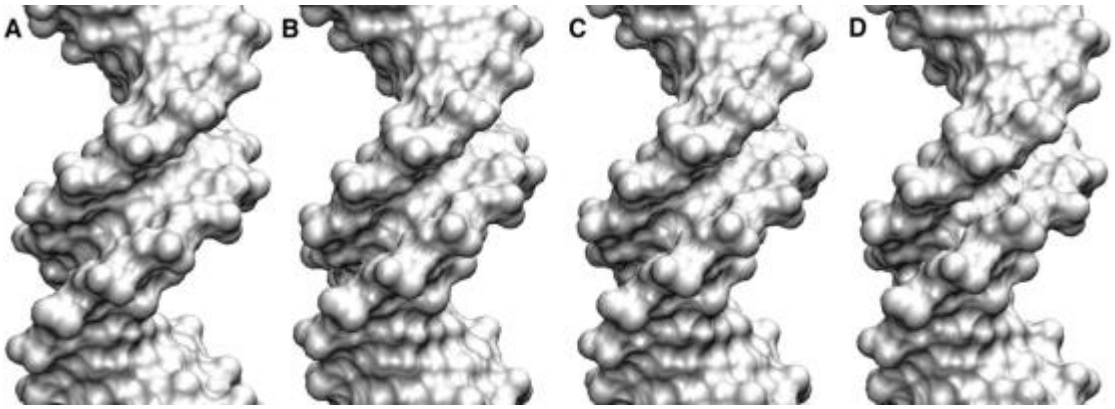

Figure 3.7 Average structure and minor groove of the DNA
Average structure and minor groove of the DNA around the binding region of p65 of the (A) free DNA, (B) the DNA in the p65-DNA complex, (C) the DNA in the Fe(II) form of the Pirin-p65-DNA supramolecular complex, and (D) the DNA in the Fe(III) form of the Pirin-p65-DNA supramolecular complex.

The local perturbation of a DNA double helix could be propagated to other parts of the helix. Clearly, the results suggest that the Fe(III) form of Pirin modulates the local conformation of the p65 binding site on the DNA. The local perturbation could potentially affect the groove width distal to the binding site of p65. Changes of the conformations of the DNA away from the p65 binding site could affect the binding site of other proteins to the DNA and in turn affects their interactions with the DNA. The shape of the DNA is known to affect the affinity and energetics of protein-DNA interactions.[75, 78] The modulation of these long range or adjacent interactions could alter the function of other proteins and fine-tune gene expression levels. The local conformational changes induced by the binding of the Fe(III) form of Pirin causes strain in the conformation of the DNA that is overcompensated for by the enhanced interaction between p65 and the DNA. Consequently, another way of assessing the effect of the binding of the Fe(III)

form of Pirin on the conformation of the DNA is by calculating the total molecular mechanics potential energy of the free DNA, the DNA in the p65-DNA complex, and the DNA in the supramolecular complexes. The components of the energy consist of the bond, angle, dihedral, electrostatics and van der Waals of only the DNA. The distributions of the energies are shown in Figure 3.8. The average total potential energies of the free DNA and the DNA in the p65-DNA complex are similar (Figure 3.8) with an average value of approximately of 5730±5 kcal/mol and 5725±11 kcal/mol respectively. The multiple-comparison statistical analysis using Tukey's method shows that the average potential energy of the free DNA and that of the DNA in the p65-DNA complex are similar with a p-value of 10^{-2} compared to a p-value of $<10^{-7}$ between the potential energy of the DNA in the Fe(III) form of the pirin-p65-DNA complex and the others. Even though p65 slightly perturbs the conformation and helical parameters of the DNA, it does not induce tremendous energetic strain to the DNA. However, unlike the binding of the Fe(II) form of Pirin, the binding of the Fe(III) form of Pirin clearly shifts the distribution of energies of the different conformations of the DNA to higher potential energies with an average value of approximately 5805±9 kcal/mol (Figure 3.8), approximately 80 kcal/mol increase in the potential energy and strain on the DNA. The increased binding interactions between p65 and the DNA due to the binding of the Fe(III) form of Pirin to p65 induce the local strain, which could propagat to other parts of the DNA.

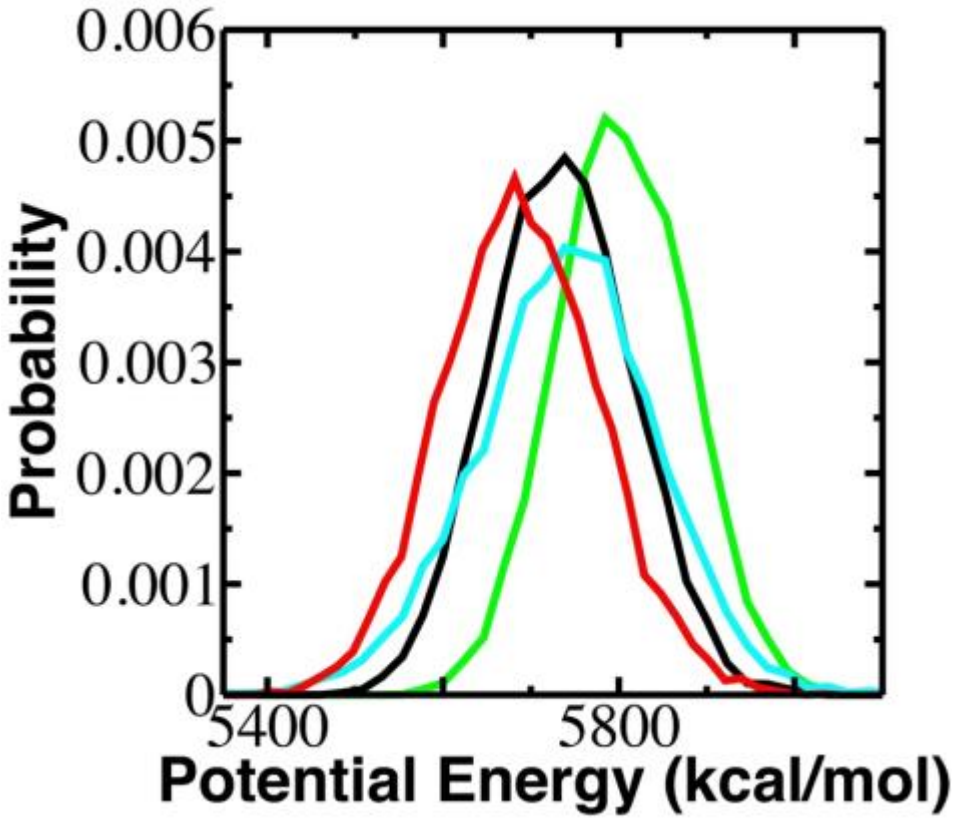

Figure 3.8 Distributions of the Potential Energies of the DNA
Distributions of the potential energies of the free DNA (cyan), the DNA in the p65-DNA complex (black), the DNA in the Fe(II) form of the Pirin-p65-DNA supramolecular complex (red), and the DNA in the Fe(III) form of the Pirin-p65-DNA supramolecular complex (green).

In addition to altering the conformation of the DNA, the binding of the Fe(III) form of Pirin to p65 also tremendously alters the dynamics of the DNA as shown in Figure 3.9. We carried out Principal Component Analysis (PCA) on the Cartesian coordinates of all of the heavy atoms of the DNA in the free DNA, the p65-DNA, and Pirin-p65-DNA supramolecular complexes to calculate the eigenvectors corresponding to the dominant motions of the DNA. The top three eigenvectors that describe the slowest modes were projected on to the trajectories to calculate the corresponding principal components (Figure 3.8). Two-dimensional projections of the conformational dynamics of the DNA on any of two of the principal components are shown

in Figure 3.8. The free DNA (cyan) clearly samples a lot more conformational space than the DNA in the p65-DNA complex (Figure 3.8). The binding of the Fe(III) form of Pirin to form the Pirin-p65-DNA supramolecular complex further reduces the dynamics of the DNA, limiting the conformational space along the dominant motions. The binding of the Fe(II) form of Pirin does not have the same effect as that of the Fe(III) form of Pirin, remaining more similar to that of the p65-DNA complex. Interestingly, the conformational space of the DNA sampled in the Fe(III) form of Pirin-p65-DNA supramolecular complex is a subset of that of the p65-DNA complex, which is also in turn a subset of the conformational space of the free DNA. These results show that the general motions of the DNA in the p65-DNA and Pirin-p65-DNA supramolecular complexes are already present in the free DNA. The results suggest that the binding of the Fe(III) form of Pirin imposes a local conformational strain and reduces the dynamics of the DNA around the p65 binding site. This reduced dynamics and conformational perturbation result in changes in the groove widths and helical parameters around the binding site of p65. These dynamical changes could dictate functional changes in subcellular processes due to long-range propagation.

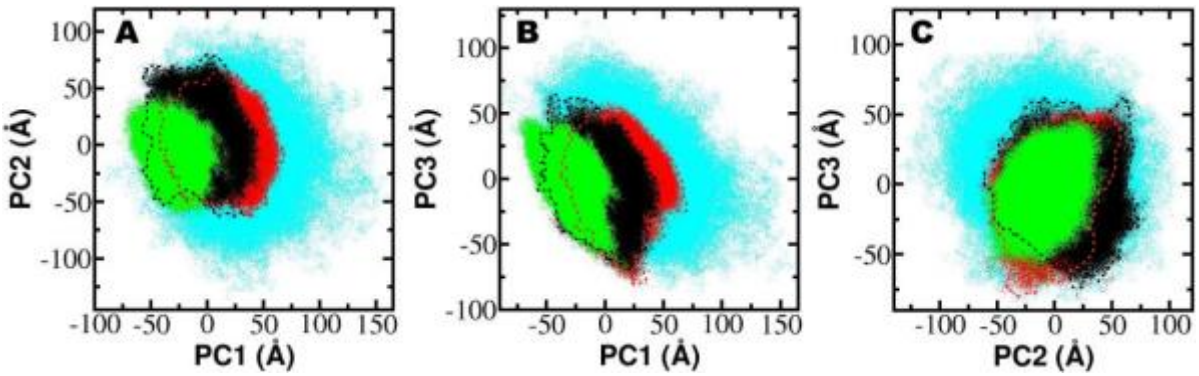

Figure 3.9 Principal Component Analysis of the Cartesian coordinate of DNA
Principal Component Analysis of the Cartesian coordinate of the heavy atoms of the free DNA(cyan), the DNA in the p65-DNA complex (black), the DNA in the Fe(II) form of the

Pirin-p65-DNA supramolecular complex (red), and the DNA in the Fe(III) form of the Pirin-p65-DNA supramolecular complex (green). (A) PC1 and PC2, (B) PC1 and PC3, and (C) PC2 and PC3. The black and red dotted outlines show the range of the data points of the conformational sampling of the DNA in the p65-DNA complex (black) and the DNA in the Fe(II) form of the Pirin-p65-DNA supramolecular complex (red), respectively.

4 CONCLUSIONS

In this study, we have studied the allosteric effect of the Fe(III) form of Pirin on the κB DNA in the Pirin-p65-DNA supramolecular complex. We constructed the Pirin-p65-DNA using knowledge from previous experimental results and protein-protein docking. We carried out extensive microsecond-long molecular dynamics simulations on the free DNA, p65-DNA complex, and the Fe(II) and (III) forms of the Pirin-p65-DNA complexes in explicit water. We show that the Fe(III) form of Pirin forms a tighter complex with p65 in the Pirin-p65-DNA supramolecular complex and, in turn, modulates the interactions between p65 and the DNA, resulting in some of the residues on p65 forming tighter contacts with DNA and further inserting into the major groove. Unlike the Fe(II) form of Pirin, the Fe(III) form of Pirin increases the binding affinity between p65 and the DNA in the Fe(III) form of the supramolecular complex. The enhanced interaction between p65 and the DNA in the Fe(III) form of the supramolecular complex results in changes in the conformational dynamics of the DNA. Interestingly, the Fe(II) form of Pirin has little or no effect on the interaction between p65 and the DNA and on the conformational dynamics of the DNA in the Fe(II) form of the Pirin-p65-DNA supramolecular complex. Our computational results, along with previous computational and experimental results, establish a hypothetical regulatory mechanism of the active Fe(III) form of Pirin in promoting gene expression that is summarized in Figure 8. In the absence of the active Fe(III) form of Pirin, NF-κB weakly binds to the DNA. The resting, inactive Fe(II) form of Pirin only

weakly binds to NF-κB and does not significantly alter the interaction between NF-κB and the DNA. On the other hand, under oxidative stress, the oxidized, active Fe(III) form of Pirin binds more strongly to NF-κB and not only increases the affinity between NF -κB and the DNA but also allosterically alters the conformational dynamics of the DNA. The conformational changes at the NF-κB binding site on the DNA could propagate to other regions of the DNA, activating gene expression and possibly modulating the affinity of other proteins on the DNA. The results provide atomic level understanding of the iron redox specific modulation of the DNA in p65-DNA complexes, details that are difficult to obtain using current experimental techniques. It is interesting how such a subtle Fe(II)/Fe(III) redox process could have a significant structural and dynamical effects in controlling sub-cellular biological processes. The results therefore complement experiments in providing a more detailed picture of the regulatory role of Pirin as a co-regulator of p65 in many sub-cellular processes.

REFERENCES

[1] Chen, Y.-Q., Ghosh, S., and Ghosh, G. (1998) A novel DNA recognition mode by the NF-[kappa]B p65 homodimer, *Nat Struct Mol Biol 5*, 67-73.

[2] May, M. J., and Ghosh, S. (1997) Rel/NF-kappa B and I kappa B proteins: an overview, *Semin. Cancer Biol. 8*, 63-73.

[3] Karin, M. (1999) How NF-kappaB is activated: the role of the IkappaB kinase (IKK) complex, *Oncogene 18*, 6867-6874.

[4] Zheng, C., Yin, Q., and Wu, H. (2011) Structural studies of NF-[kappa]B signaling, *Cell Res 21*, 183-195.

[5] Vallabhapurapu, S., and Karin, M. (2009) Regulation and function of NF-kappaB transcription factors in the immune system, *Annual review of immunology 27*, 693-733.

[6] Manavalan, B., Govindaraj, R., Lee, G., and Choi, S. (2011) Molecular modeling-based evaluation of dual function of IκBζ ankyrin repeat domain in toll-like receptor signaling, *Journal of Molecular Recognition 24*, 597-607.

[7] Hayden, M. S., and Ghosh, S. (2008) Shared principles in NF-kappaB signaling, *Cell 132*, 344-362.

[8] Whiteside, S. T., and Israel, A. (1997) I kappa B proteins: structure, function and regulation, *Semin Cancer Biol 8*, 75-82.

[9] Weil, R., Whiteside, S. T., and IsraÈl, A. (1997) Control of NF-κB Activity by the IκBβ Inhibitor, *Immunobiology 198*, 14-23.

[10] Karin, M., and Ben-Neriah, Y. (2000) Phosphorylation Meets Ubiquitination: The Control of NF-κB Activity, *Annual review of immunology 18*, 621-663.

[11] Whiteside, S. T., Epinat, J. C., Rice, N. R., and Israël, A. (1997) I kappa B epsilon, a novel member of the I kappa B family, controls RelA and cRel NF-kappa B activity, *The EMBO Journal 16*, 1413-1426.

[12] Li, Z., and Nabel, G. J. (1997) A new member of the I kappaB protein family, I kappaB epsilon, inhibits RelA (p65)-mediated NF-kappaB transcription, *Molecular and Cellular Biology 17*, 6184-6190.

[13] Suyang, H., Phillips, R., Douglas, I., and Ghosh, S. (1996) Role of unphosphorylated, newly synthesized I kappa B beta in persistent activation of NF-kappa B, *Molecular and Cellular Biology 16*, 5444-5449.

[14] Thompson, J. E., Phillips, R. J., Erdjument-Bromage, H., Tempst, P., and Ghosh, S. (1995) IκB-β regulates the persistent response in a biphasic activation of NF-κB, *Cell 80*, 573-582.

[15] Pang, H., Bartlam, M., Zeng, Q., Miyatake, H., Hisano, T., Miki, K., Wong, L.-L., Gao, G. F., and Rao, Z. (2004) Crystal Structure of Human Pirin: AN IRON-BINDING NUCLEAR PROTEIN AND TRANSCRIPTION COFACTOR, *Journal of Biological Chemistry 279*, 1491-1498.

[16] Wendler, W. M. F., Kremmer, E., Förster, R., and Winnacker, E.-L. (1997) Identification of Pirin, a Novel Highly Conserved Nuclear Protein, *Journal of Biological Chemistry 272*, 8482-8489.

[17] Brzóska, K., Stępkowski, T. M., and Kruszewski, M. (2011) Putative proto-oncogene Pir expression is significantly up-regulated in the spleen and kidney of cytosolic superoxide dismutase-deficient mice, *Redox Report 16*, 129-133.

[18] Hubner, R. H., Schwartz, J. D., De Bishnu, P., Ferris, B., Omberg, L., Mezey, J. G., Hackett, N. R., and Crystal, R. G. (2009) Coordinate control of expression of Nrf2-modulated genes in the human small airway epithelium is highly responsive to cigarette smoking, *Mol Med 15*, 203-219.

[19] Mercer, B. A., Lemaitre, V., Powell, C. A., and D'Armiento, J. (2006) The Epithelial Cell in Lung Health and Emphysema Pathogenesis, *Current respiratory medicine reviews 2*, 101-142.

[20] Licciulli, S., Luise, C., Scafetta, G., Capra, M., Giardina, G., Nuciforo, P., Bosari, S., Viale, G., Mazzarol, G., Tonelli, C., Lanfrancone, L., and Alcalay, M. (2011) Pirin Inhibits Cellular Senescence in Melanocytic Cells, *The American Journal of Pathology 178*, 2397-2406.

[21] Miyazaki, I., Simizu, S., Okumura, H., Takagi, S., and Osada, H. (2010) A small-molecule inhibitor shows that pirin regulates migration of melanoma cells, *Nature chemical biology 6*, 667-673.

[22] Liu, F., Rehmani, I., Esaki, S., Fu, R., Chen, L., de Serrano, V., and Liu, A. (2013) Pirin is an iron-dependent redox regulator of NF-κB, *Proceedings of the National Academy of Sciences 110*, 9722-9727.

[23] Barman, A., and Hamelberg, D. (2016) Fe(II)/Fe(III) Redox Process Can Significantly Modulate the Conformational Dynamics and Electrostatics of Pirin in NF-κB Regulation, *ACS Omega 1*, 837-842.

[24] Romashkova, J. A., and Makarov, S. S. (1999) NF-kappaB is a target of AKT in anti-apoptotic PDGF signalling, *Nature 401*, 86-90.

[25] Karin, M., and Lin, A. (2002) NF-kappaB at the crossroads of life and death, *Nature immunology 3*, 221-227.

[26] Li, Q., and Verma, I. M. (2002) NF-kappaB regulation in the immune system, *Nature reviews. Immunology 2*, 725-734.

[27] Nussinov, R., Tsai, C. J., and Ma, B. Y. (2013) The Underappreciated Role of Allostery in the Cellular Network, *Annu Rev Biophys 42*, 169-189.

[28] Lefstin, J. A., and Yamamoto, K. R. (1998) Allosteric effects of DNA on transcriptional regulators, *Nature 392*, 885-888.

[29] Lu, X. J., and Olson, W. K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures, *Nucleic Acids Research 31*, 5108-5121.

[30] Lu, X.-J., and Olson, W. K. (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures, *Nat. Protocols 3*, 1213-1227.

[31] Peguero-Tejada, A., and van der Vaart, A. (2017) Biasing Simulations of DNA Base Pair Parameters with Application to Propellor Twisting in AT/AT, AA/TT, and AC/GT Steps and Their Uracil Analogs, *Journal of chemical information and modeling 57*, 85-92.

[32] Dickerson, R. E. (1989) Definitions and nomenclature of nucleic acid structure components, *Nucleic Acids Research 17*, 1797-1803.

[33] Uline, M., and Corti, D. (2013) Molecular Dynamics at Constant Pressure: Allowing the System to Control Volume Fluctuations via a "Shell" Particle, *Entropy 15*, 3941.

[34] Cornell, W. D., Cieplak, P., Bayly, C. I., and Kollmann, P. A. (1993) Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation, *Journal of the American Chemical Society 115*, 9620-9631.

[35] Bayly, C. I., Cieplak, P., Cornell, W., and Kollman, P. A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model, *The Journal of Physical Chemistry 97*, 10269-10280.

[36] Toxvaerd, S., and Dyre, J. C. (2011) Communication: Shifted forces in molecular dynamics, *The Journal of Chemical Physics 134*, 081102.

[37] Wu, X.-W., and Sung, S.-S. (1998) Constraint dynamics algorithm for simulation of semiflexible macromolecules, *Journal of Computational Chemistry 19*, 1555-1566.

[38] Köster, A. M. (1997) Intermolekulare Kräfte: The Theory of Intermolecular Forces. Von A. J. Stone. Oxford University Press, Oxford, 1996. 264 S., geb., 49,50 £. ISBN 0-19-855884-8, *Nachrichten aus Chemie, Technik und Laboratorium 45*, 917-918.

[39] Grzybowski, A., Gwóźdź, E., and Bródka, A. (2000) Ewald summation of electrostatic interactions in molecular dynamics of a three-dimensional system with periodicity in two directions, *Physical Review B 61*, 6706-6712.

[40] III, C. L. B., Pettitt, B. M., and Karplus, M. (1985) Structural and energetic effects of truncating long ranged interactions in ionic and polar fluids, *The Journal of Chemical Physics 83*, 5897-5908.

[41] Davidchack, R. L., Handel, R., and Tretyakov, M. V. (2009) Langevin thermostat for rigid body dynamics, *The Journal of Chemical Physics 130*, 234101.

[42] Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F. v., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath, *The Journal of Chemical Physics 81*, 3684-3690.

[43] Rossky, P. J., and Karplus, M. (1979) Solvation. A molecular dynamics study of a dipeptide in water, *Journal of the American Chemical Society 101*, 1913-1937.

[44] Senn, H. M., and Thiel, W. (2009) QM/MM methods for biomolecular systems, *Angewandte Chemie (International ed. in English) 48*, 1198-1229.

[45] Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., and Stewart, J. J. P. (1985) Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model, *Journal of the American Chemical Society 107*, 3902-3909.

[46] Dewar, M. J. S., and Storch, D. M. (1985) Development and use of quantum molecular models. 75. Comparative tests of theoretical procedures for studying chemical reactions, *Journal of the American Chemical Society 107*, 3898-3902.

[47] Stewart, J. J. P. (1989) Optimization of parameters for semiempirical methods I. Method, *Journal of Computational Chemistry 10*, 209-220.

[48] Porezag, D., Frauenheim, T., Kohler, T., Seifert, G., and Kaschner, R. (1995) Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon, *Physical review. B, Condensed matter 51*, 12947-12957.

[49] Aranow, R. H. (1963) THE STATISTICAL MECHANICS OF MICELLES1, *The Journal of Physical Chemistry 67*, 556-562.

[50] Archontis, G., Simonson, T., and Karplus, M. (2001) Binding free energies and free energy components from molecular dynamics and Poisson-Boltzmann calculations. Application to amino acid recognition by aspartyl-tRNA synthetase, *Journal of molecular biology 306*, 307-327.

[51] Michel , J., Foloppe, N., and Essex, J. W. (2010) Rigorous Free Energy Calculations in Structure-Based Drug Design, *Molecular Informatics 29*, 570-578.
[52] Wang, J., Hou, T., and Xu, X. (2006) Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models, *Current Computer - Aided Drug Design 2*, 287-306.
[53] Gilson, M. K., and Honig, B. (1988) Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis, *Proteins 4*, 7-18.
[54] Huang, N., Kalyanaraman, C., Bernacki, K., and Jacobson, M. P. (2006) Molecular mechanics methods for predicting protein-ligand binding, *Physical chemistry chemical physics : PCCP 8*, 5166-5177.
[55] Stein, S. A. M., Loccisano, A. E., Firestine, S. M., and Evanseck, J. D. (2006) Chapter 13 Principal Components Analysis: A Review of its Application on Molecular Dynamics Data, In *Annual Reports in Computational Chemistry* (David, C. S., Ed.), pp 233-261, Elsevier.
[56] Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs, *Journal of Computational Chemistry 26*, 1668-1688.
[57] Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins 65*, 712-725.
[58] Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, *J Chem Theory Comput 11*, 3696-3713.
[59] Jorgensen, W. L. (1982) Revised TIPS for simulations of liquid water and aqueous solutions, *The Journal of Chemical Physics 77*, 4156-4163.
[60] Ryckaert, J., Cicotti, G., and Berendsen, H. (1977) Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes, *Journal of Computational Physics 23*, 327-341.
[61] Izaguirre, J. A., Catarello, D. P., Wozniak, J. M., and Skeel, R. D. (2001) Langevin stabilization of molecular dynamics, *The Journal of Chemical Physics 114*, 2090-2098.
[62] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method, *The Journal of Chemical Physics 103*, 8577-8593.
[63] Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics, *Journal of molecular graphics 14*, 33-38, 27-38.
[64] Pierce, B. G., Wiehe, K., Hwang, H., Kim, B. H., Vreven, T., and Weng, Z. (2014) ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers, *Bioinformatics (Oxford, England) 30*, 1771-1773.
[65] Frisch, M., Trucks, G., Schlegel, H., Scuseria, G., Robb, M., Cheeseman, J., Montgomery Jr, J., Vreven, T., Kudin, K., and Burant, J. (2004) Gaussian 03. Gaussian, *Inc.: Wallingford, CT*.
[66] Becke, A. D. (1988) Density-functional exchange-energy approximation with correct asymptotic behavior, *Physical Review A 38*, 3098-3100.
[67] Becke, A. D. (1993) Density-functional thermochemistry. III. The role of exact exchange, *The Journal of Chemical Physics 98*, 5648-5652.

[68] Lee, C., Yang, W., and Parr, R. G. (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, *Physical Review B 37*, 785-789.

[69] Hay, P. J., and Wadt, W. R. (1985) Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg, *The Journal of Chemical Physics 82*, 270-283.

[70] Roe, D. R., and Cheatham, T. E. (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data, *J Chem Theory Comput 9*, 3084-3095.

[71] Blanchet, C., Pasi, M., Zakrzewska, K., and Lavery, R. (2011) CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures, *Nucleic Acids Research*.

[72] Doshi, U., Holliday, M. J., Eisenmesser, E. Z., and Hamelberg, D. (2016) Dynamical network of residue–residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation, *Proceedings of the National Academy of Sciences 113*, 4735-4740.

[73] Barman, A., and Hamelberg, D. (2016) Coupled Dynamics and Entropic Contribution to the Allosteric Mechanism of Pin1, *The Journal of Physical Chemistry B 120*, 8405-8415.

[74] Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., and Cheatham, T. E. (2000) Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models, *Accounts of Chemical Research 33*, 889-897.

[75] Rohs, R., Jin, X. S., West, S. M., Joshi, R., Honig, B., and Mann, R. S. (2010) Origins of Specificity in Protein-DNA Recognition, In *Annual Review of Biochemistry, Vol 79* (Kornberg, R. D., Raetz, C. R. H., Rothman, J. E., and Thorner, J. W., Eds.), pp 233-269, Annual Reviews, Palo Alto.

[76] Brennan, R. G., Roderick, S. L., Takeda, Y., and Matthews, B. W. (1990) Protein-DNA conformational changes in the crystal structure of a lambda Cro-operator complex, *Proceedings of the National Academy of Sciences of the United States of America 87*, 8165-8169.

[77] Calladine, C. R. (1982) Mechanics of sequence-dependent stacking of bases in B-DNA, *Journal of molecular biology 161*, 343-352.

[78] Rohs, R., West, S. M., Sosinsky, A., Liu, P., Mann, R. S., and Honig, B. (2009) The role of DNA shape in protein-DNA recognition, *Nature 461*, 1248-U1281.