2012

31P NMR of Backbone Conformation and Dynamics in DNA at Cre Binding Site in Terms of Sequence Context

Kelly A. Garton
Scripps College

^{31}P NMR of Backbone Conformation and Dynamics in DNA at Cre Binding Site in Terms of Sequence Context

A Thesis Presented

by

Kelly Garton

To the W.M. Keck Science Department

Of Claremont McKenna, Pitzer, and Scripps Colleges

In Partial Fulfillment of

The Degree of Bachelor of Arts

Senior Thesis in Biochemistry

April 2012

Table of Contents

Abstract

The Cre sequence (ACGT) is a site responsible for the binding of specific transcription factors that determine the activation of genes. Due to its major role in gene transcription, it has become a subject of immense research. The binding of transcription factors to the Cre binding site has been determined to be dependent on DNA conformation. In this study, the effects of flanking sequence around the Cre binding site on the conformation and the dynamics of DNA were investigated. The Cre binding site was studied in its native form with differing flanking sequences to determine the BI/BII profile (conformation) and the magnitude of the energy transition barrier (dynamics) between the BI and BII conformations of each phosphate step of the following three dodecamer sequences: CreACAG, CreGGAG, and CreTATA. In order to obtain the BI/BII profile of each phosphate step, 2D ^{31}P-NMR NOESY and HSQC experiments at various temperatures were utilized. Based of the basic principles of kinetics, the lower the energy barrier between the two conformations, the easier the transition between the BI and BII conformation. Therefore, it was hypothesized that *low* and *high* %BII character lead to a large energy barrier (high $\Delta G^{\ddagger}$ values), whereas *average* %BII character leads to a small energy barrier (low $\Delta G^{\ddagger}$ values). The results of the 2D ^{31}P-NMR experiments of the three dodecamer sequences confirmed this relationship between the %BII character and the magnitude of the energy barrier ($\Delta G^{\ddagger}$). However, further conformation and dynamics studies must be conducted to further understand the correlation.

Introduction

The Cre Sequence (ACGT)

Due to the complexity of multicellular organisms, activities of specialized cells tend to be controlled by chemical messengers. Cyclic AMP (cAMP), one such chemical messenger, is a product of the reaction between cytosolic ATP and the enzyme adenyl cyclase. Some of the cellular events regulated by cAMP include glycogen degradation, fatty acid production, heart rate and blood pressure, water reabsorption, and bone resorption. The cAMP molecule targets an enzyme known as cAMP-dependent kinase, or protein kinase A (PKA).[1]

An effective approach for controlling the activity of transcription factors is the addition of phosphate groups, a process known as protein phosphorylation. PKA phosphorylates a wide variety of cellular proteins, which includes the transcription factor called CREB.

The CREB protein is responsible for binding to specific DNA sequences known as cAMP response elements (CRE), which are located adjacent to genes whose transcription is induced by cAMP. The phosphorylation of the CREB protein activates CBP (CREB-binding protein), a transcriptional coactivator. CBP catalyzes histone acetylation. The acetylation loosens the packing of tightly wound nucleosomes and interacts with RNA polymerase to assist transcription machinery at surrounding gene promoters.[1] **Figure 1** illustrates the activation of gene transcription by cAMP.

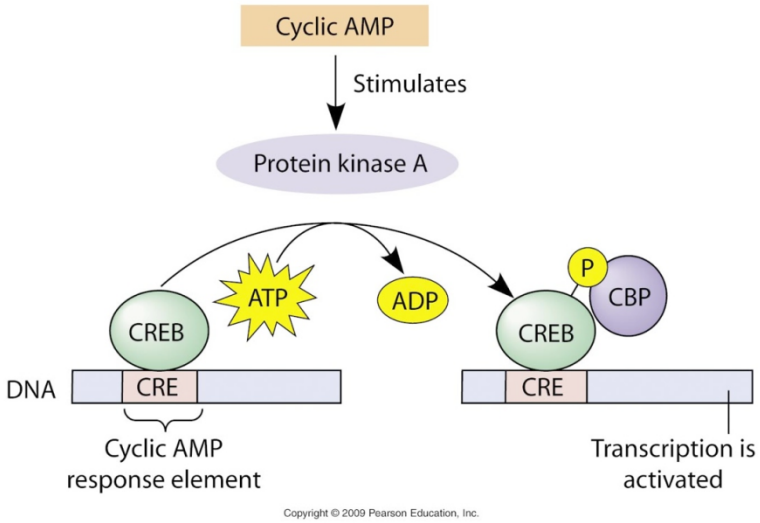

Figure 1. The role of cyclic AMP on gene transcription. Adapted from Becker et al.

Oncogenes are specific DNA sequences whose presence can trigger the development of cancer. Several oncogenes code for components of growth signal pathways. They can produce mutant forms or excessive quantities of various transcription factors, including CREB, which is present in a variety of tumor cells.[1]

The seemingly simple, but monumental, four nucleotide series—adenine-cytosine-guanine- thymine—of the Cre sequence has been extensively researched. The binding of CREB to the Cre sequence is thought to be greatly affected by the conformation of the DNA. Therefore, the Hatcher-Skeers research group investigates the effect of the flanking sequence on the backbone conformation of the Cre sequence.

Protein-DNA Recognition

The binding of protein to DNA depends on sequentially and structurally specific mechanisms, also known as direct recognition or direct readout. The interaction between amino acid residues and specific bases in DNA sequences are direct readout mechanisms. Several observations report that sequence dependence alone does not explain the phenomena of protein-DNA binding.[2,3,4] Studies of base mutations lacking direct contact with the protein, such as flanking sequences, greatly affect binding affinity.[5]

In addition to sequence-dependent interactions, there is increasing evidence that structural properties of DNA affects protein-DNA binding.[5] Some of these structural properties include flexibility, elasticity, bending and kinking, major and minor groove widths, as well as hydration.[5] Ultimately the protein-DNA process is based on DNA geometry.[6]

Regulatory proteins are known to recognize particular DNA sequences through atomic contact between the protein and DNA or indirectly through the conformational characteristics of DNA.[7] Using a sequence dependent stiffness scale, Grohima and colleagues have computed the average DNA stiffness parameters for numerous protein-DNA complexes.[8] The importance of DNA stiffness was determined by experimental free energy change due to protein-DNA binding in the tested complexes. The results showed a 0.65 to 0.97 binding correlation between the stiffness factor and the free energy of protein-DNA binding. Furthermore, they noted that the binding correlation varied based on the protein complex.[8] In a later study, statistical analysis of base-amino acid interaction within protein-DNA complexes in addition to computer simulations of base-amino acid interactions was

performed.[7] The results provided binding specificities of different complexes and reconfirmed that protein–DNA recognition correlates with the stiffness of the DNA.

DNA Conformations

As previously mentioned, conformational behavior is important to determine local flexibility in terms of DNA nucleotide sequence. Studies have shown that some backbone states cause dramatic helical changes. Therefore, sequence dependent DNA backbone flexibilities affect the attraction and binding of proteins.[9]

Two backbone conformations were discovered from a series of NMR studies measuring specific spin-spin coupling in B-form double-stranded DNA. These geometries were characterized by two backbone states: BI and BII.[9] The conformations differ in their torsion angles ε (epsilon) and ζ (zeta). In BI ε is *trans* and ζ is *gauche* with ε-ζ centered at -90 °. In contrast, in BII ε is *gauche* and ζ is *trans* with ε-ζ centered at +90 °. With respect to both the minor and major grooves, the more common BI conformation is symmetric. The phosphate group moves towards the minor groove in the BII conformation.[10] **Figure 2** illustrates the B-form DNA conformations.

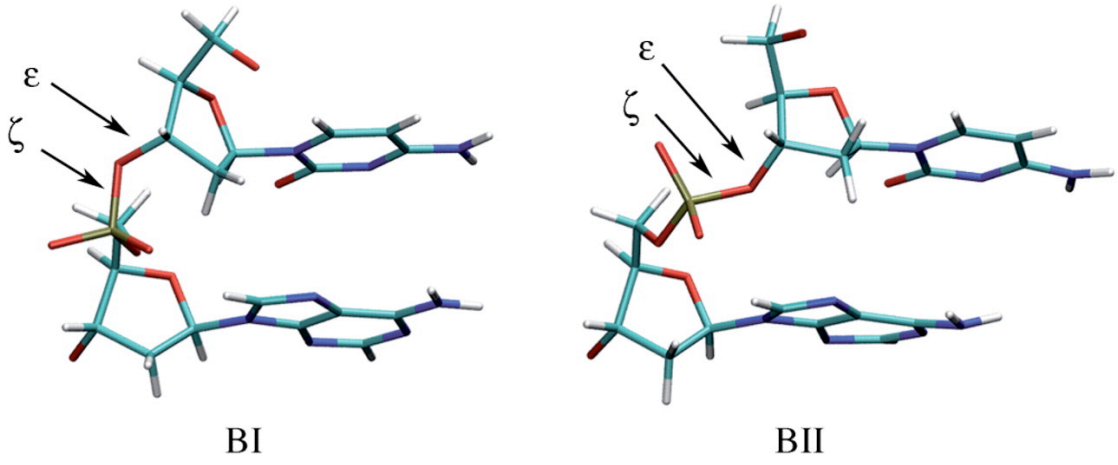

Figure 2. B-form conformations of DNA phosphodiester backbone. Adapted from Heddi et al.

DNA backbones experience conformational exchange between BI and BII states. Since geometry derived protein-DNA binding is central to our research, it is important to focus on the BI/BII conformations, which include three combinations: BI.BI, BI.BII, or BII.BII.[11] The notation BI.BI denotes that both nucleotides involved in the dinucleotide binding step are in the BI conformation and therefore have minimal backbone flexibility. BI.BII dinucleotide steps have the maximal flexibility.[10]

DNA Dynamics

The BI and BII conformations exist in a dynamic equilibrium according to the following equation:[12]

$$B_I \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} B_{II}$$

The switch between the B conformations occurs when the phosphate group switches from the major to minor grove or vice versa. The ability to change from one conformation to the other is an example of backbone dynamics. For most dinucleotide steps, the BI conformation is lower in energy and therefore is said to be more stable. An energy barrier must be overcome to make the transition between the two conformations. Typically, the value of the Gibbs free energy of this transition lies between 4.6 and 15.0 kcal/mol.[13] Based on simple kinetics, the smaller the Gibbs free energy barrier, the more flexible the DNA. This elasticity results in a more rapid transition between BI and BII states. **Figure 3** effectively demonstrates the assumed two-state model, to describe the transition using a potential energy diagram. The BI conformation is referred to as the zero energy state, while ΔG (free energy difference between the two conformers) and $\Delta G^{\ddagger}$ (free energy of activation) are defined relative to the zero-energy BI state.[12]

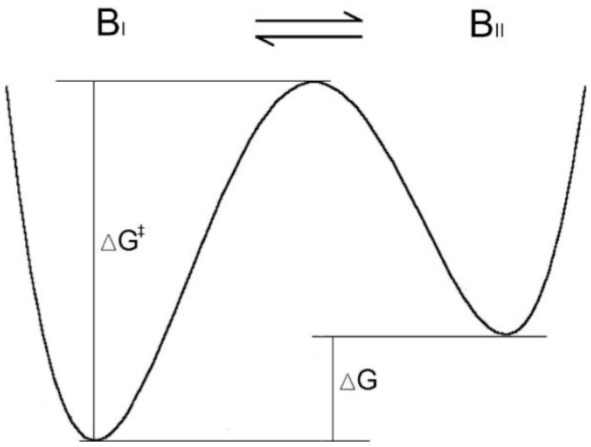

Figure 3. Potential energy diagram representing the two-state conformational transition. Adapted from Tian et al.

Two-dimensional nuclear magnetic resonance spectroscopy provides a practical method for determining both the free energy of activation and percent BII conformation of B-form DNA.

Analysis Technique: Nuclear Magnetic Resonance Spectroscopy (NMR)

Introduction to General Spectroscopy and Nuclear Magnetic Resonance (NMR):

In order to understand the concepts and theory behind Nuclear Magnetic Resonance (NMR), it is necessary to first understand and review its broader family, the field of spectroscopy. Spectroscopy is defined as the interaction between both energy and matter.[14] In general, the energy involved in spectroscopy is in the form of electromagnetic radiation (EMR). This type of radiation includes the exclusion of time dependent variables and the existence of atoms that dwell only in well-defined nuclear and electronic states with fixed energies. Depending on the system, either atomic or molecular, there exist several quantized states that can change. In order to move from one quantized state to another (also known as the transition from one state to the next), the energy difference between the quantized states must be overcome. In many types of spectroscopy, such as NMR, the energy necessary for this transition is absorbed as a **photon**.[14] While photons present particle-like qualities, for the purposes of understanding spectroscopy, it is necessary to focus on their wave-like nature. The equation below encapsulates the relationship between the energy (E) of a single photon of EMR and its frequency (v).[14]

$$h\nu = \Delta E \qquad (1)$$

In spectroscopic equations the most efficient absorption of EMR occurs when EMR of the exact frequency (v) required to **excite** the transition from one state to the next is absorbed.[14] This efficient absorption is also known as **resonance**.[14] Many experts approach spectroscopy with its semi-classical definition. The assumption of this approach is that EMR is described mathematically as a wave. In addition, the atoms and the molecules that interact

with EMR must be thought of in terms of quantized energy levels and frequencies. Overall, the quantized energy states of the atoms and molecules are viewed as quantum oscillators.[14]

Most importantly, it is possible to express these standing waves as mathematical wave functions. The exact equation of the wave for quantum wave functions need not be known. One of the requirements is that the wave function must yield observables of that state of the system. In other terms, the wave function, according to the fundamentals of quantum chemistry, must follow the Schrödinger equation (**Equation 2**).[14]

$$\mathbf{E\psi=\psi\hat{H}} \qquad \mathbf{(2)}$$

Where ψ is an **eigenfunction** of the Hamiltonian operator with **eigenvalue,** E.[14]

Quantum mechanical transitions provide insight to a system's relaxation from a higher energy state to a low energy state (ground) in spectroscopic experiments. When there are multiple quantum oscillators present in a molecule, the response of one oscillator applied to EMR may depend on the state of the other oscillators present in the same molecule. This phenomenon is known as **coupling**.[14]

The Basic Theory Behind NMR:

One area of particular interest to this thesis is the concept known as **Zeeman splitting**.[14] Zeeman splitting explores the division of energetically equivalent or degenerate nuclear spin states (or electronic) into energetically nonequivalent or nondegenerate state upon exposure to a magnetic field. Nuclei that respond by a change of spin states are targets for spectroscopists. The nuclear spin of a particle has a definitive quantum number, usually with value ½. A quadrapole moment occurs when the quantum number of the nucleus is greater than ½. [14]

It is vital to quickly review the quantum aspects of nuclear spin. **I** defines the number of stationary spin states that the nucleus may occupy in an imposed magnetic field, which is represented by **Equation 3**:[14]

$$\textbf{Number of levels} = (2I+1) \qquad (3)$$

When **I** is equal to ½, which is true for ^{1}H, ^{3}H, ^{13}C, ^{15}N, ^{19}F, ^{29}Si and ^{31}P, the nucleus can have two spin states. In addition to **I** (the spin quantum number), each individual state also has an orientation given by the quantum number **m**. The quantum number **m** can have values shown in the equation:[14]

$$m = -I, -I+1, ..I-1, I \qquad (4)$$

Therefore, according to equation 4 the allowable spin states for **I** = ½ are **m** = -½, +½. All spin states of the nucleus are degenerate until a magnetic field is present. There is also a "spinning top" model of nuclear spin, but it relies heavily on classical models rather than quantum theory.[14]

There are three general important pieces of information that are available from NMR: nuclear shielding and chemical shift, scalar coupling and dynamics.

Nuclear shielding and chemical shift:

The local environment of the reporting nucleus can be identified using NMR. The magnetic environment of the nuclear spin is affected by a variety of sources, such as valence shell orbital hybridization, charge and bond polarity, and electronegativity of nearby atoms.[14] **Chemical shifts** are caused by changes in the environment that induce small changes in the resonant frequency of the nuclear spin.[14] The origins of a chemical shift begin with an

applied magnetic field denoted by $\vec{B}_0$, which creates a magnetization $\vec{M}$. The induced magnetization is proportional to the applied magnetic field, $\vec{M}=\rho_0 \chi_0 \vec{B}_0$, where χ is known as the **magnetic susceptibility**.[14] The Greek letter ρ_0 indicates the concentration of induced dipole per unit volume. When χ has a negative value, the induced dipoles that arise as a result of the introduced magnetic field directionally oppose the applied field. In addition, the vector sum of all of the individual induced dipoles summed over its volume, $\vec{M}$ will be antiparallel to that of $\vec{B}_0$. In other words, the induced magnetization partially counters the applied field. When χ has a positive value there is an increase in the field and the induced magnetization adds to the applied field. The sign of χ describes whether or not the experimental substance is **diamagnetic** or **paramagnetic**.[14] For the purposes of NMR, the magnetic susceptibility is relevant only in that distortions to the magnetic field occur at boundaries between phases with different susceptibilities. This discontinuity affects the NMR line shape.

Again, chemical shifts depend on the resonance frequency of a nucleus resulting from its local electronic environment. When a nucleus is located in a region where the induced $\vec{B}_{loc}$ is in the same direction as the applied field, $\vec{B}_0$, the effective field that is detected by the nucleus, is larger than the applied field. In this situation the nucleus is said to be **deshielded** and will resonate at a higher frequency than expected from the applied $\vec{B}_0$.[14] In contrast, a nucleus that is located in a region where the induced $\vec{B}_{loc}$ is in the opposite direction as the applied field, $\vec{B}_0$ is smaller than the applied field. In this particular situation the nucleus is said to be **shielded** and resonates at a lower frequency than expected.[14] It is important to note that in the early history of NMR, in order to reach resonance for a deshieled nucleus, the magnetic field was reduced and was classified as 'downfield.' Similarly, by

increasing the applied field, resonances of shielded nuclei were reached and were designated to be 'upfield.'[14] The degree of shielding and deshielding that a nucleus experiences is based on the magnitude of the applied field. There exists a shielding constant, σ_i, for a nucleus in a particular environment that describes the proportionality between the chemical shift and applied field. **Equation 5** shows the mathematical relationship between the observed resonance frequency of the i^{th} nucleus and the applied field:[14]

$$\omega_i = \gamma B_0(1-\sigma_i) \quad \textbf{(5)}$$

It is best to keep in mind that a chemical shift is normally measured relative to a standard reference. In order to calculate the chemical shift, δ_i (usually recorded as ppm), a dimensionless ratio, must be determined using the following relationship between the shift of the signal of interest and that of the standard reference:[14]

$$\delta_i = \frac{\omega_i - \omega_{ref}}{\omega_{ref}} \quad \textbf{(6)}$$

Scalar Coupling:

Scalar coupling separates NMR from other types of spectroscopy, also known as through-bonding or **J coupling**.[14] This type of coupling is transmitted by the interactions of nuclear spin with the spins of bonding electrons. Take a moment to picture two nuclei, A and B. Nucleus A is spin **I** = ½ and therefore can occupy both **m** = -½ and +½. The electrons located within bonding orbitals overlapping with nuclear spin A will be affected by the spin state of A. Similarly, the electron spin states will change as a result of the spin of the nucleus. If nucleus B overlaps with the affected orbitals, the electronic spin states can also affect nucleus B. Due to the perturbation of all of the spins, there is a slight change in the resonance frequency of nucleus B, depending on the **m** value of nucleus A. In this case

nuclei A and B are J-coupled. The **coupling constant** (J_{AB}) is defined as the distance between the resonance frequency of spin B when A is **m** = -½ versus **m** = +½ and vice versa.[14]

Furthermore, two distinct nuclei, A and C, are coupled and the coupling constant (J_{AC}) is significantly smaller than the difference between the resonance frequencies of A and C. This condition results in weak coupling and first-order spectra. A **doublet** (present in first-order spectra) is two split lines.[14] Each line corresponds to the transition frequencies of the resonant spin when the coupling partner is either in lower or upper spin state. The coupling constant is equal to the exact splitting between the two lines. Refer to **Figure 4** for spectral simulations of coupled spins.

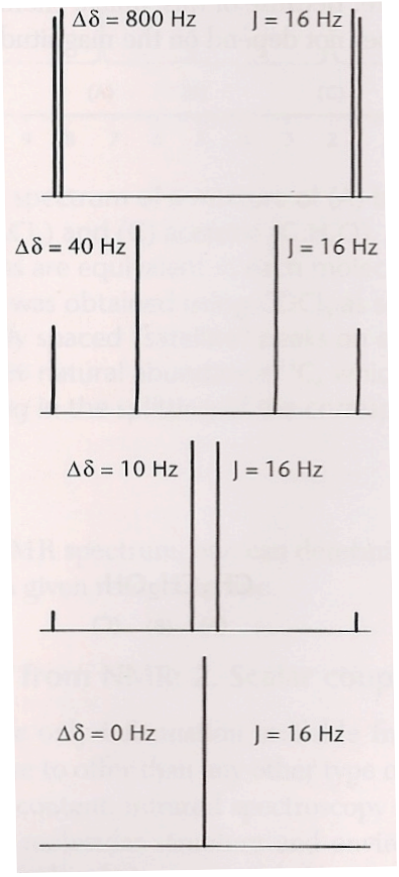

Figure 4. Spectral simulations of a two-spin system with a constant coupling J=16 Hz showing field effects on line position and intensities of coupled spins. Adapted from Pochapsky.

Dynamics:

Most importantly to this thesis, NMR provides a plethora of information regarding molecular dynamics on a variety of time scales.[14] This includes measuring rates of processes, such as conformational equilibria, tautomerization, and weak complex formation. The time range can vary from 10^{-12} seconds, to hours, or even days.[14] A common example of a dynamic NMR is the analysis of chemical shifts in a situation where chemical exchange is taking place. If an NMR-active nucleus samples two distinct environments (A and B), each gives rise to resolved resonances. If the exchange is slow, the two different resonances will be observed at two distinct frequencies.[14] According to the NMR, two different nuclei are being observed. When the exchange rate increases, the lifetime of a spin state may become shortened and the line may become broader; this incident is unsurprisingly referred to as **exchange broadening**.[14]

Line width is found to be inversely proportional to fast exchange rates but directly proportional to slow exchange rates. As the exchange rate increases, the probability that exchange between the sites will occur during the observation also increases. If the exchange becomes fast enough, it competes with relaxation for limiting the lifetime of individual excited states. Near the coalescence point, new sets of states, both ground and excited, begin to form. These new states represent a weighted time-average of the original two sets of ground and excited states.[14]

To summarize, when the exchange rate is slow there are two distinct nuclei environments present that are represented by two discrete lines in the NMR spectrum. As the

exchange rate increases, the more indistinct the two nuclei environments become. In the fast-exchange regime, the spins switch environments in the course of one precession creating a single line, which give the average precessional frequency.[14]

In addition to DNA conformational exchange, there are countless dynamic processes in macromolecules that are of interest for both chemical and biological process, such as particular ligand binding. Let $X_1 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} X_2$ represent a two-site exchange, which is almost identical to the two-state conformational change of DNA.[14] This equation could represent an isomerization, a ligand substrate binding, or a protein-protein interaction. When the exchange is extremely slow on the chemical-shift time-scale, there are different resonances for the same spin in different conformers. However, as mentioned in more general terms above, the exchange rate is usually fast enough that the result is a single resonance showing the weighted average of the chemical shifts of the individual conformers.[14]

The BI/BII conformational change falls into the fast exchange regime. Therefore, the exchange rate is larger than the difference in chemical shifts between the two sites. The theory for this exchange is developed in a study by Tian et al. The observed magnetization in the NMR experiments was written as the sum of unitary magnetization precession, relaxation, and exchange between the BI and BII. For the purposes of the study, the T_2 was assumed to be equal for both the BI and BII conformer. Through the eigenvector method and a combined matrix, the eigenvalues were determined. An effective line width was determined and an averaged chemical shift was also calculated ($<\omega>$). To better comprehend the fast exchange between conformations, the chemical shift values and equilibrium constant allow for the fit of phosphorus line widths in the fast exchange limit.[12]

Multi-Dimensional NMR:

Multi-dimensional NMR allows spectroscopists to make several correlations in a single experiment. Multi-dimensional NMR provides information that makes it possible to characterize large molecules such as proteins, polymers and even DNA.[14]

For a 2D NMR there are four successive time periods: *preparation, evolution, mixing,* and *detection*. During the *preparation* time, a delay time when thermal equilibrium must be obtained. In addition the different spins must have the same phases relative to one another. The *evolution* time involves a specified state of coherence that is dependent on the elapsed time and the Hamiltonian operator. However, for NMR and spectroscopic transitions in general, small time-dependent **perturbations** of the Hamiltonian operator are added to the original time-independent Schrödinger equation.[14] Next, the *mixing* period is when the frequency pairs are determined. Lastly, the *detection* period is when the system becomes further evolved under the elapsed time and Hamiltonian operator. In order to determine the coordinates of the signal peaks in a spectra, the precessional frequencies during the time period of *evolution* and *detection* must be obtained.[15] **Figure 5** illustrates the four successive time periods of 2D NMR.

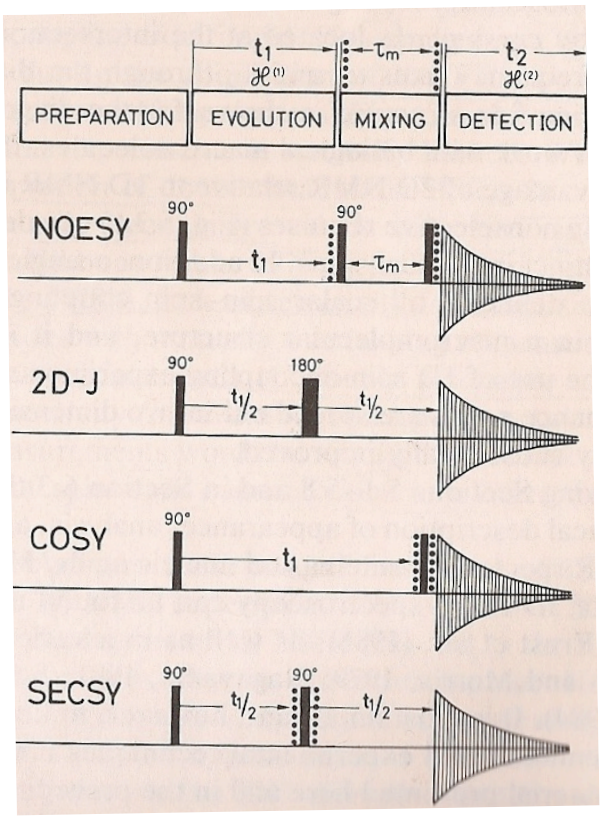

Figure 5. General schematic of time periods of 2D NMR followed by four specific schematics of homonuclear ^{1}H 2D NMR experiments (NOESY, 2D-J, COSY, and SECSY). Adapted from Wüthrich.

Two different two-dimensional NMR experiments are used to analyze the conformation and dynamics of DNA: Nuclear Overhauser Enchantment and Exchange Spectroscopy (NOESY) and Heteronuclear Single-Quantum Coherence (HSQC). Nuclear Overhauser enhancement and exchange spectroscopy (NOESY) is a particularly attractive NMR method for biological macromolecules, such as DNA.[15] NOESY experiments use the dipolar interaction in the form of cross relaxation to correlate spins that are close in distance.[16] Heteronuclear single-quantum coherence (HSQC) focuses on obtaining the heteronuclear correlations via one-bond scalar coupling, by observing the proton (^{1}H) signals. The proton signals are utilized since ^{1}H is more sensitive relative to heteronuclei. For the DNA HSQC experiments the heteronuclei are ^{31}P.[16]

Sequences and Hypothesis

This study utilizes BI/BII conformation and DNA dynamics data obtained from ^{31}P-NMR studies to determine how flanking sequence of the Cre binding site affect the conformation and the dynamics of the DNA sequence. The current study involves the investigation of the relationship between the BI/BII profile and the Gibbs free energy of each phosphate step. For two of the studied sequences CreACAG (ACAG ACGT CTGT) and CreGGAG (GGAG ACGT CTCC) BI/BII profiles were previously collected (**Figure 6** and **Figure 7**).[17] These previously collected profiles were used to look at the effect of methylation of the BII profiles. However, for this study, only the native or non-methylated sequence was studied as shown in blue. Each phosphate step indicated on the x-axis of **Figure 6** and **Figure 7** is represented by the following scheme: (1) nucleotide abbreviation (2) nucleotide number in the sequence (3) "p" for phosphate (4) the following nucleotide abbreviation (5) nucleotide number in the sequence (ex.: **C2pA3** for A**CA**G ACGT CTGT).

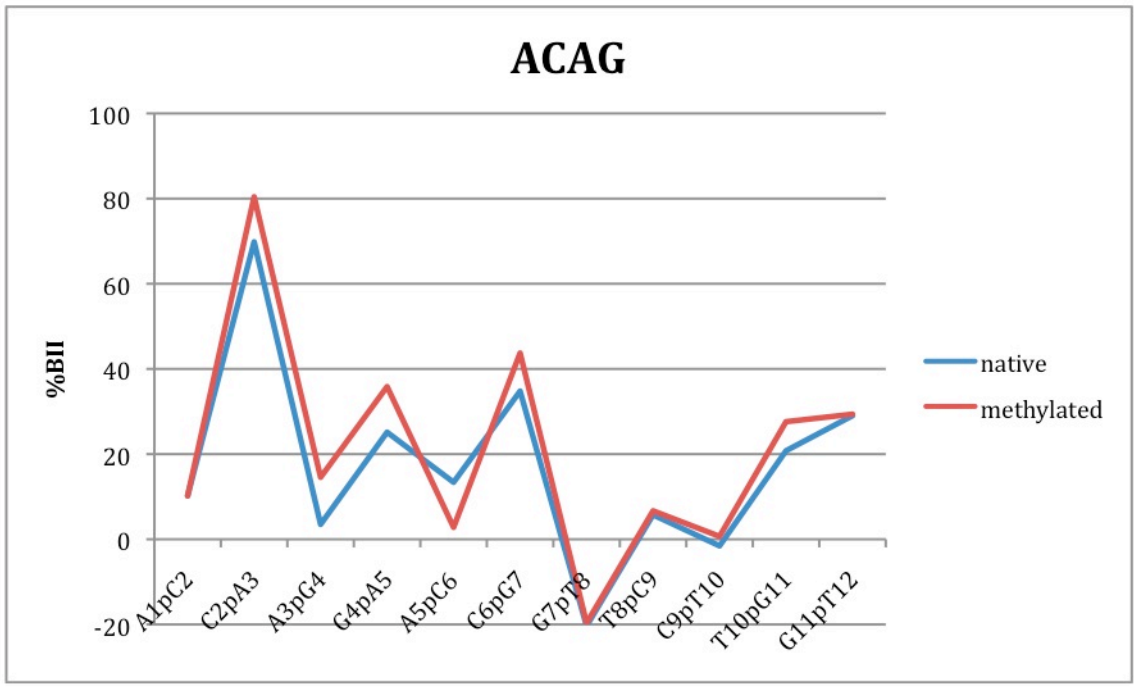

Figure 6. %BII for each phosphate step in the ACAGACGTCTGT (native) and ACAGA-meCGTCTGT (methylated) sequences at 296.64 K.[17]

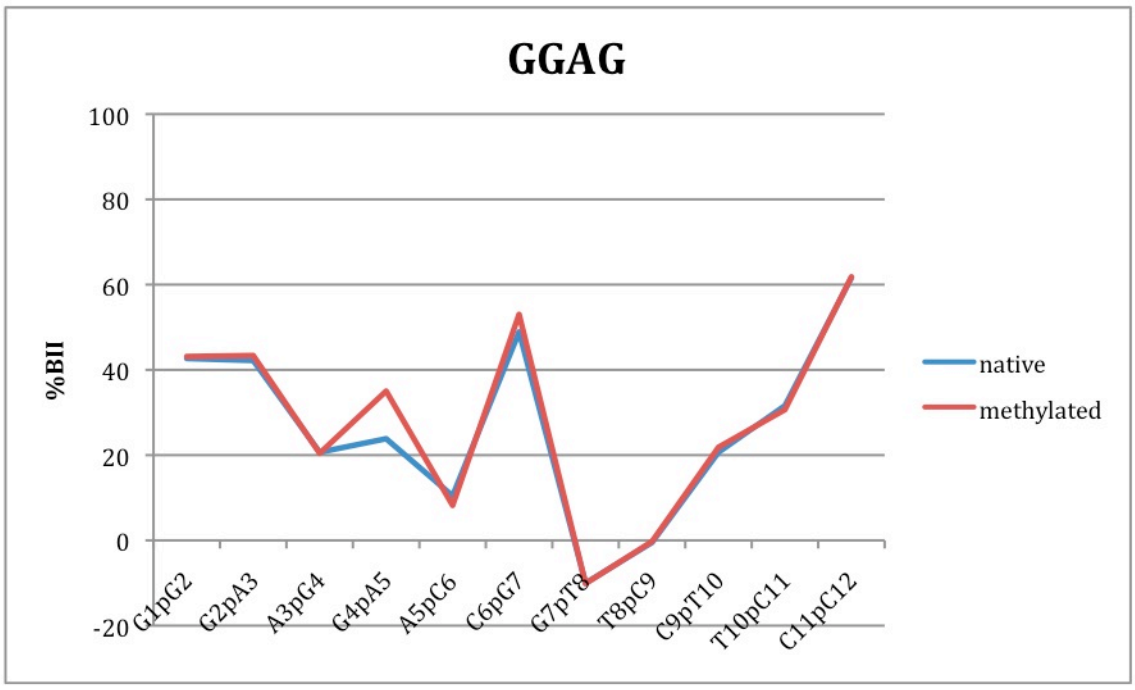

Figure 7. %BII for each phosphate step in the GGAGACGTCTCC (native) and GGAGA-meCGTCTCC (methylated) sequences at 296.64 K.[17]

The resulting BII profiles indicated that the differing flanking sequences caused significant changes in the %BII. This study attempts to correlate the percent BII to the Gibbs free energy ($\Delta G^{\ddagger}$) or the energy barrier that must be overcome to switch between the two conformations, BI and BII. Therefore, it is expected that phosphate steps demonstrating low and high %BII will have high-energy barriers (large $\Delta G^{\ddagger}$ value), whereas phosphate steps exhibiting average %BII will have low energy barriers (small $\Delta G^{\ddagger}$ value). This predicted relationship between %BII and $\Delta G^{\ddagger}$ values was validated through a study involving a single sequence EcoCGC (CGCAATTCGCG) and the effects of the presence of cations, such as aluminum (Al^{3+}) and sodium (Na^{+}), on the BI/BII profile and Gibbs free energy of each phosphate step.[18]

Methods

Sample Preparation

DNA dodecamers of the native sequences $d(TATAACGTTATA)_2$, $d(ACAGACGTCTGT)_2$, and $(GGAGACGTCTCC)_2$ were obtained from Alpha DNA (Quebec, Canada). **Table 1** provides the abbreviations as well as the nucleotide sequences of the dodecamers. The samples were desalted by temperature-controlled centrifugation in Y300 Centricons (Millipore) for a minimum of two repeats. After the desalting process, the samples were lyophilized overnight and stored at 4ºC until quantification and analysis.[19]

Table 1. The dodecamer sequences including Cre binding sequence analyzed for NMR experiments

Abbreviated Sequence Name	**Nucleotide Sequence**
CreTATA	TATA ACGT TATA
CreACAG	ACAG ACGT CTGT
CreGGAG	GGAG ACGT CTCC

NMR Experiments

1D ^{1}H-NMR and ^{31}P-NMR and 2D NOESY and HSQC experiments for the samples were performed using a Bruker Avance spectrometer (^{1}H frequency 500MHz) with a 5 mm gradient probe. HSQC spectra were recorded at increments of 2 ˚C for a range of 22 ˚C and 44 ˚C using an FTS System AIRJET temperature preconditioner (FTS, Stone Ridge, NY). All temperatures were provided by a calibrated thermocouple.[19]

The data from 2D ^{1}H Nuclear Overhauser Effect (NOESY) experiments were analyzed using Sparky (T. D. Goddard and D. G. Kneller, CA). The peaks were identified using the sequential walk method in the 1'H and 2'H-2"H regions of the NOESY spectrum.[20] The 2'H-2"H chemical shifts were used to determine chemical shifts in the 3'H and 4'H regions. **Figure 8** shows the layout of a typical NOESY NMR spectrum. Using these

chemical shifts, the 3'H and 4'H regions, the proton shifts of the HSQC spectra were assigned. The known chemical shifts were utilized to determine the step and chemical shift of each phosphorus peak using TopSpin (Bruker, MA).[19]

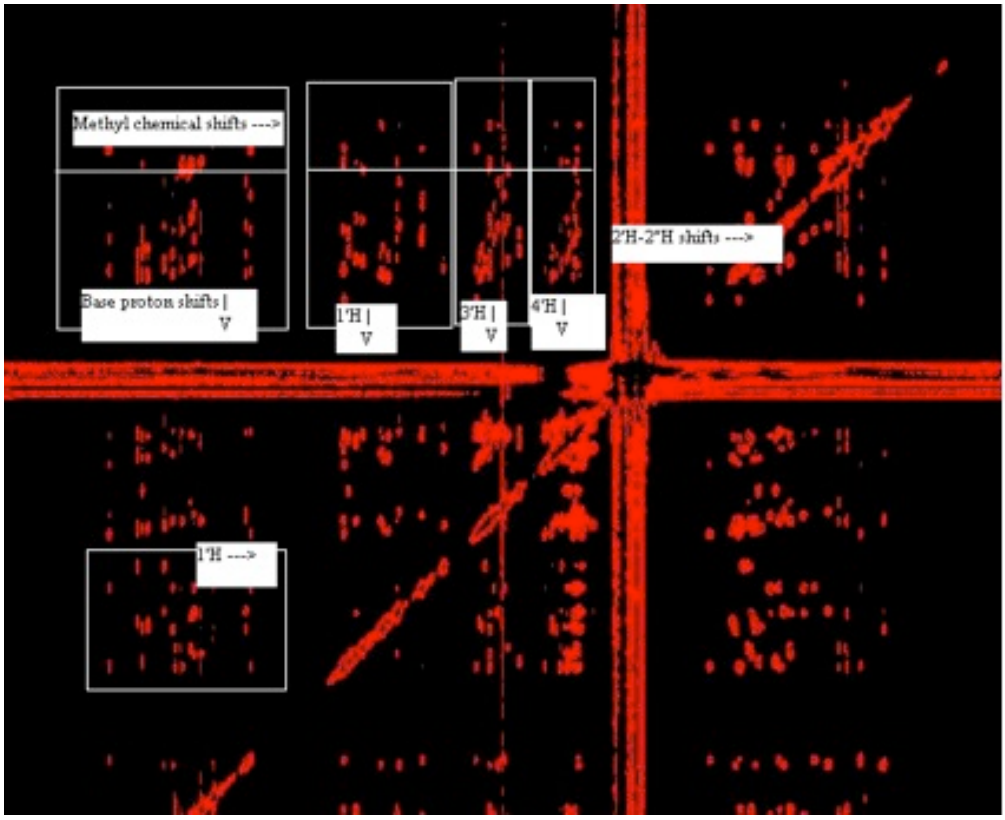

Figure 8. Layout of spectra and proton shifts of a NOESY NMR experiment. Adapted from Yonis.

Data Analysis

1D ^{31}P traces were extracted from the indirect dimensions of the HSQC spectra for each step of the DNA sequence and fit to a Lorenztian curve and a simple baseline offset in Mathematica™ (Wolfram, Champaign, IL). The temperature dependent rate constants were obtained from the Lorentzian fits of widths. The chemical shifts along with the Lorentzian half widths were then used to determine the %BII character and the Gibbs free energy for each phosphate step.[13]

Results

Mathematica™ Data Analysis

Thorough analysis of the NOESY and HSQC spectra led to the assignment of all chemical shifts of each phosphate step at each temperature for CreTATA, CreACAG, and CreGGAG. These chemical shift values were then incorporated into **Equation 7**(Adapted by Tian et al.) to calculate the %BII of each phosphate step for all three sequences. The calculated %BII is based on the torsional angle of the backbone of the phosphate as well as the time the phosphate resides in a specific conformation.

$$\%B_{II} = \frac{100}{2.368 - 0.005110T}\delta P\,(ppm) + \frac{134.5 - 0.2031T}{2.368 - 0.005110T} \qquad \textbf{(7)}$$

To obtain %BII values within reasonable parameters of the data collected **Equation 8** and **10** were utilized for CreACAG and CreGGAG. Similarly, **Equation 9** and **10** were used for CreTATA. These equations were responsible for determining the temperature-dependent limiting shifts for the BI and BII conformations (Adapted from Tian et al).

$$\delta I = -1.345 + 0.002031\ K - 0.15 \qquad \textbf{(8)}$$

$$\delta I = -1.345 + 0.002031\ K - 0.20 \qquad \textbf{(9)}$$

$$\delta II = 1.023 - 0.003079\ K \qquad \textbf{(10)}$$

To determine the Gibbs free energy of activation ($(\Delta G^{\ddagger}_{max})$ **Equation 11** was employed (Adapted from Tian et al).

$$r=\frac{1}{\pi T_2}+\frac{K^2(\omega_{BII}-\omega_{BI})^2}{(1+K)^3}\frac{h}{\pi K_b T}e^{\Delta G^{\ddagger}/RT} \quad \textbf{(11)}$$

The variables were defined as follows: **r** = Lorentzian half-widths; **T2**= spin-spin relaxation time; **ω**= chemical shift; **K**= ($<\omega>-\omega_{B1}$)/($\omega_{BII}-<\omega>$); **h**=Planck's constant; $\mathbf{K_b}$=Boltzmann's constant; **R**=gas constant.

The CreTATA Sequence

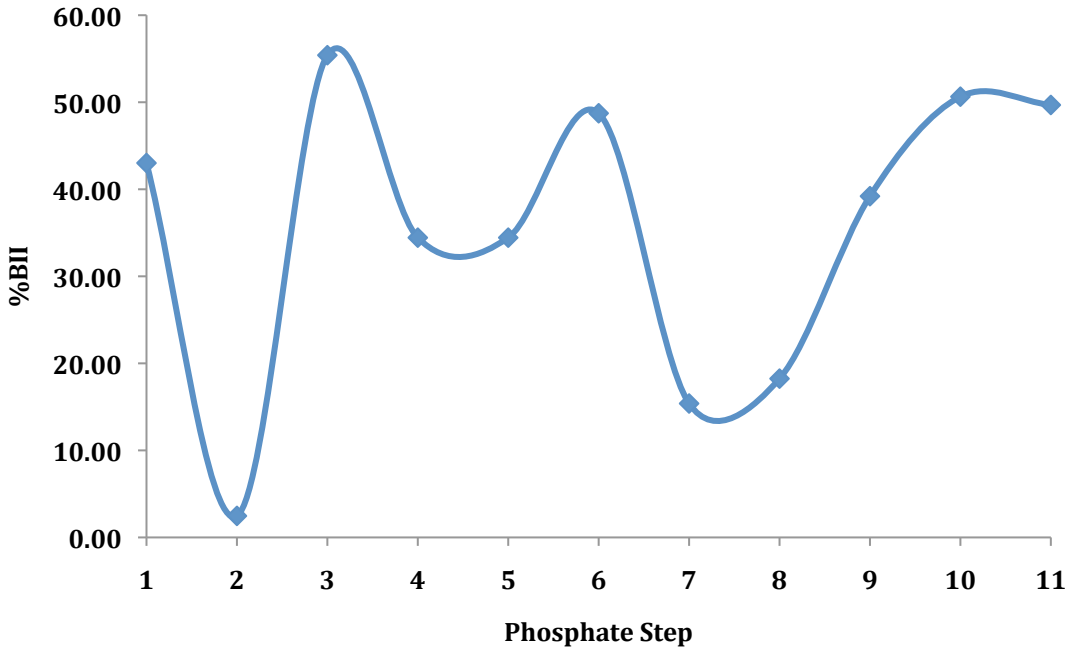

Figure 9. %BII of each phosphate step of the CreTATA sequence at 297.15 K (24°C). Please refer to Supplemental Information **Table 5**.

The CreTATA sequence showed a wide range of BII character, varying from nearly 0% to almost 60%. Interestingly, the A2pT3 step demonstrated notably less BII character, while the A3pA4, C6pG7, A10pT11, and T11pA12 steps showed a %BII character around 50%.

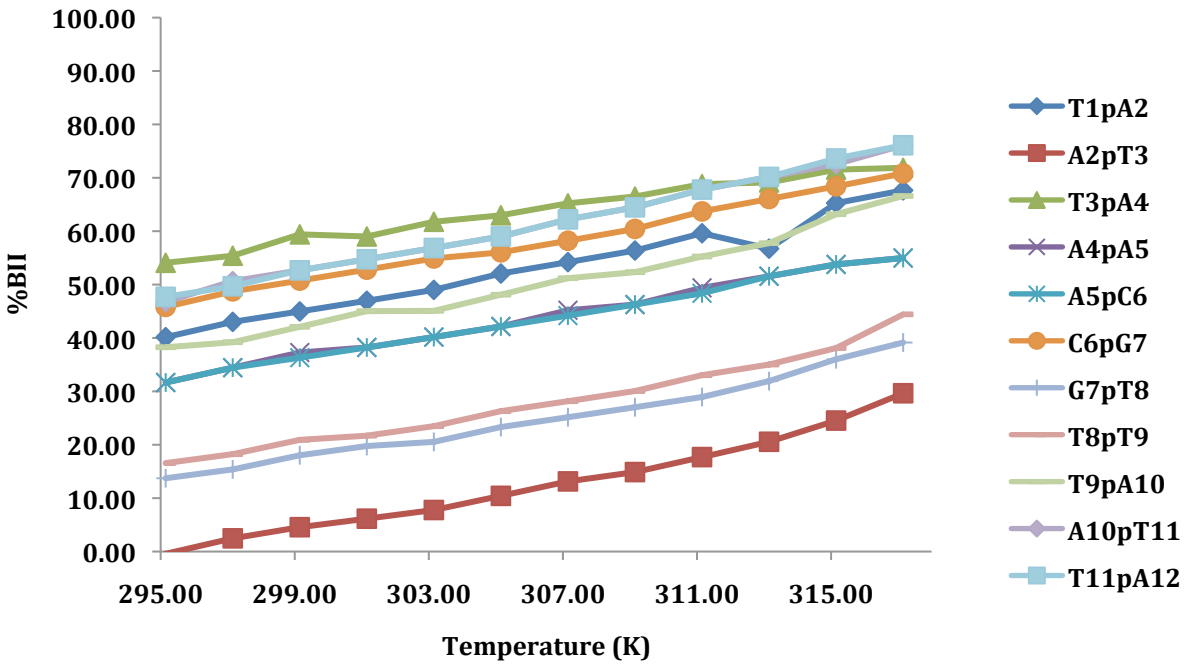

Figure 10. %BII of each phosphate step of the CreTATA sequence at varying temperatures sranging from 295.15 K-317.15 K at 2.00 K intervals. Refer to Supplemental Information **Table 5**.

In the case of the CreTATA sequence, the A2pT3 step showed limited BII character that did not rise above 30%. In contrast, the T11pA12 showed the most BII character reaching above 70%. The general trend indicates that in relation to increasing temperature, there is also a consistent increase in %BII character for each phosphate step.

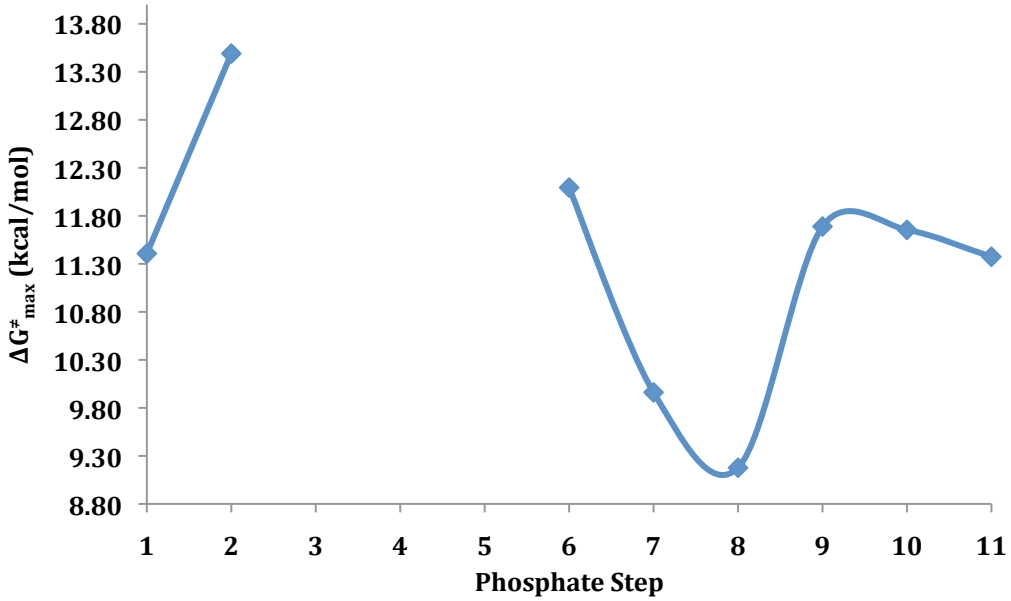

Figure 11. The maximum Gibbs free energy ($\Delta G^{\ddagger}_{max}$) between the BI and BII conformations of the CreTATA sequence for each phosphate step. Refer to Supplemental Information **Table 6**.

The CreTATA sequence exhibited the largest Gibbs free energy at step 2 (A2pT3 and the smallest Gibbs free energy at step 8, T8pC9. In general, $\Delta G^{\ddagger}_{max}$ appeared to stay within the range of 11-12 kcal/mol. Due to discrepancies in the data, steps 3, 4, and 5 were not included in the visual representation of the calculated $\Delta G^{\ddagger}_{max}$ values.

The CreACAG Sequence

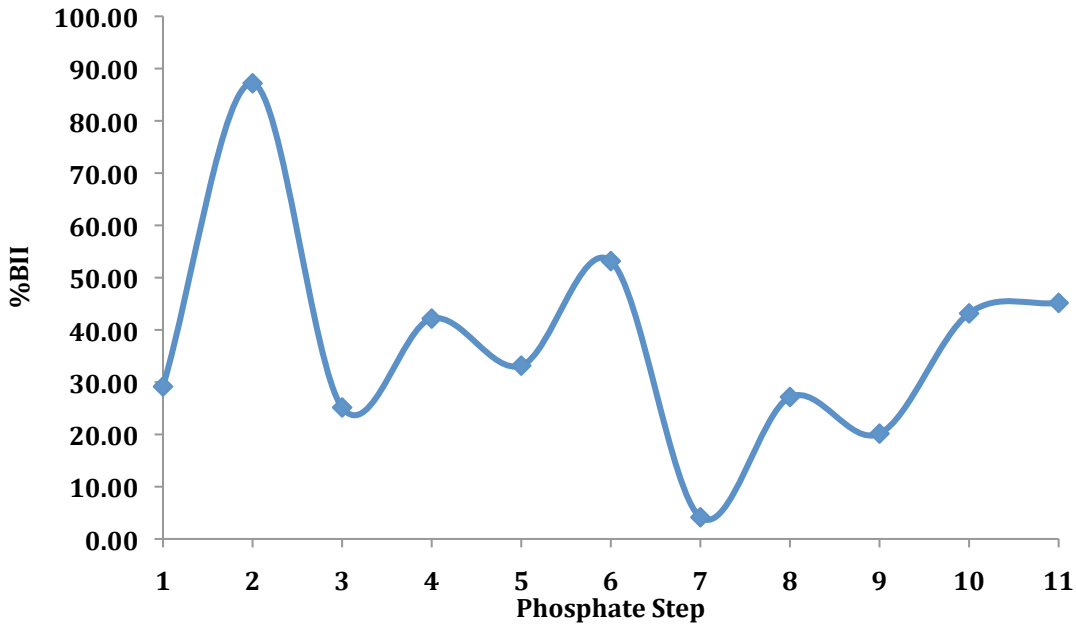

Figure 12. %BII of each phosphate step of the CreACAG sequence at 297.15 K (24°C). Refer to Supplemental Information **Table 9**.

Compared to the %BII profile of the CreTATA sequence, the CreACAG sequence showed a wider range of %BII of under 10% to almost 90%. The C2pA3 step was seen to have the highest %BII, whereas the G7pT8 marked the lowest %BII.

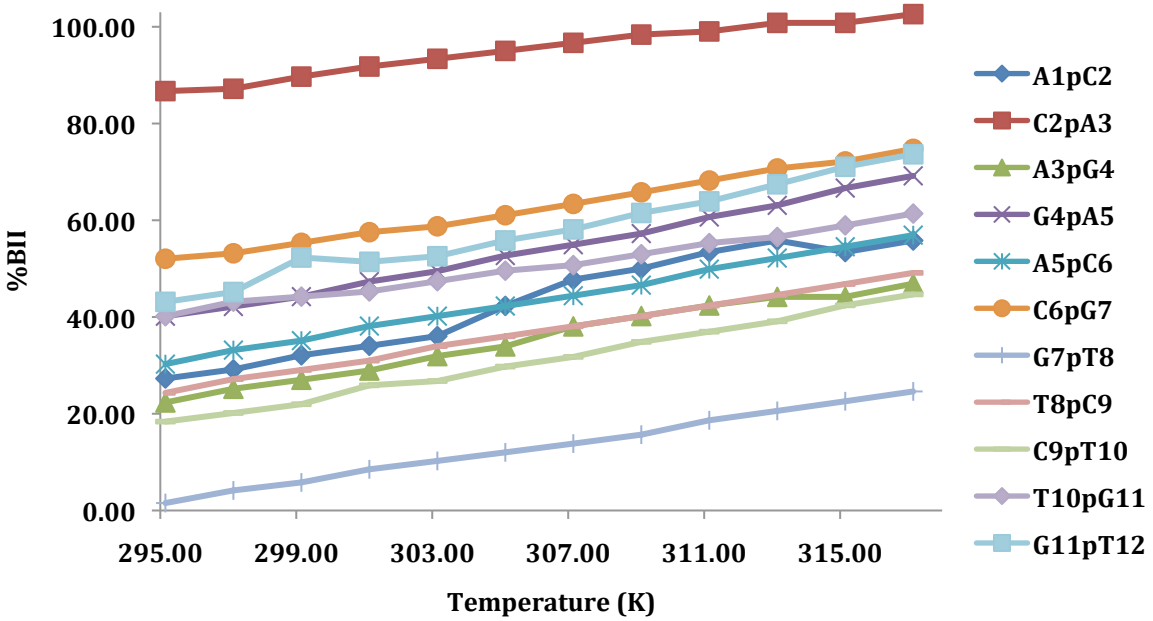

Figure 13. %BII of each phosphate step of the CreACAG sequence at varying temperatures ranging from 295.15 K-317.15 K at 2.00 K intervals. Refer to Supplemental Information **Table 9.**

The CreACAG sequence shows an evident variance in %BII character. The C2pA3 step shows high %BII character, whereas the G7pT8 indicates low %BII character. The other nine phosphate step hover around 20% to 75%. Again, the general trend indicates that in relation to increasing temperature there is also a consistent increase in %BII character for each phosphate step.

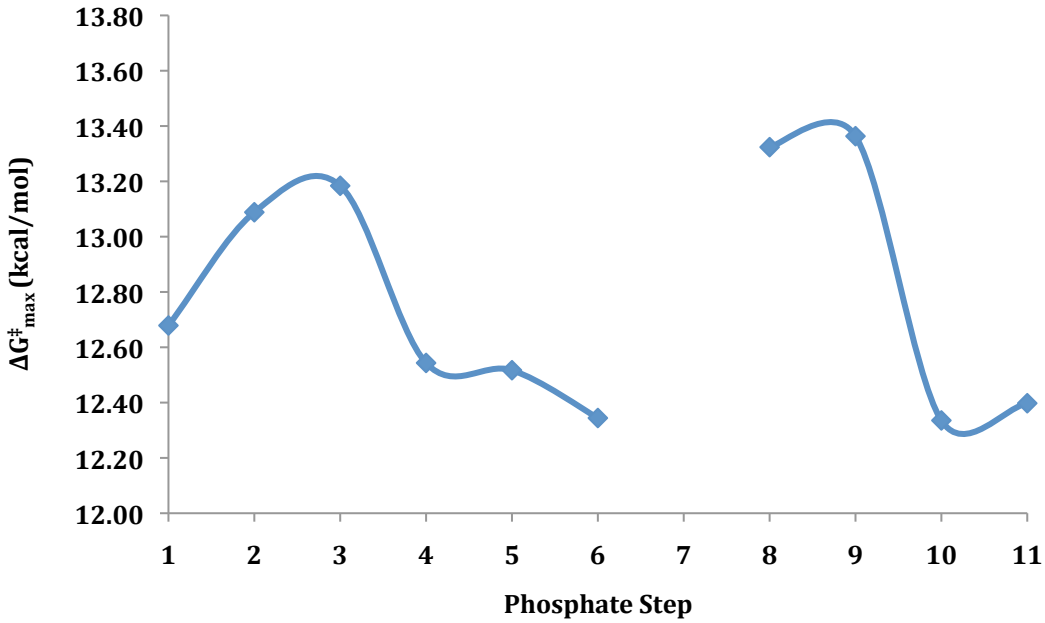

Figure 14. The maximun Gibbs free energy ($\Delta G^{\ddagger}_{max}$) between the BI and BII conformations of the CreACAG sequence for each phosphate step. Refer to Supplemental Information **Table 10.**

A large $\Delta G^{\ddagger}_{max}$ value was seen for step 9, C9pT10, and a small $\Delta G^{\ddagger}_{max}$ was viewed for step 6, C6pG7, for the CreACAG sequence. Most phosphate steps varied within a range of 1 kcal/mol. Due to discrepancies in the data, step 7 and step 8 were eliminated from the figure above.

The CreGGAG Sequence

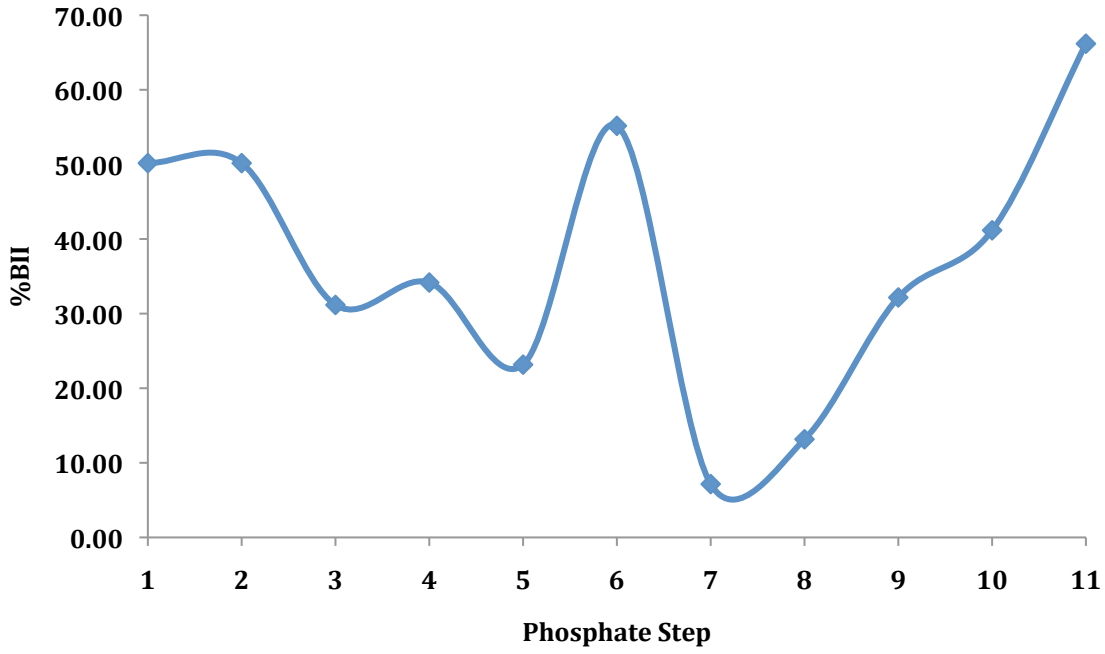

Figure 15.%BII of each phosphate step of the CreGGAG sequence at 297.15 K (24°C). Refer to Supplemental Information **Table 13.**

The CreGGAG sequence exhibited the highest %BII phosphate step at the C11pC12 and the lowest %BII phosphate step at the G7pT8 step. The G1pG2, G2pA3, and the C6pG7 steps showed average values of %BII character.

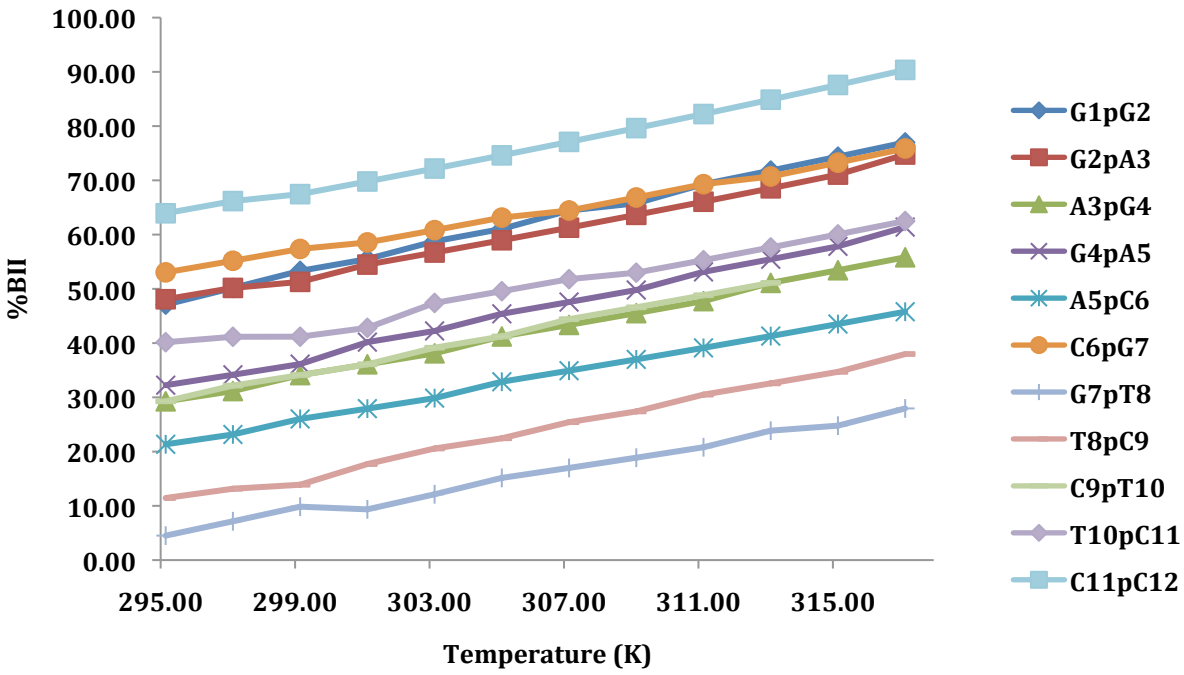

Figure 16. %BII of each phosphate step of the CreGGAG sequence at varying temperatures ranging from 295.15 K-317.15 K at 2.00 K intervals. Refer to Supplemental Information **Table 13.**

Compared to CreACAG, the CreGGAG sequence demonstrated a more dispersed range of %BII for the eleven phosphate steps. The G7pT8 step showed the lowest %BII character, while the C11pC12 gave rise to the highest %BII character.

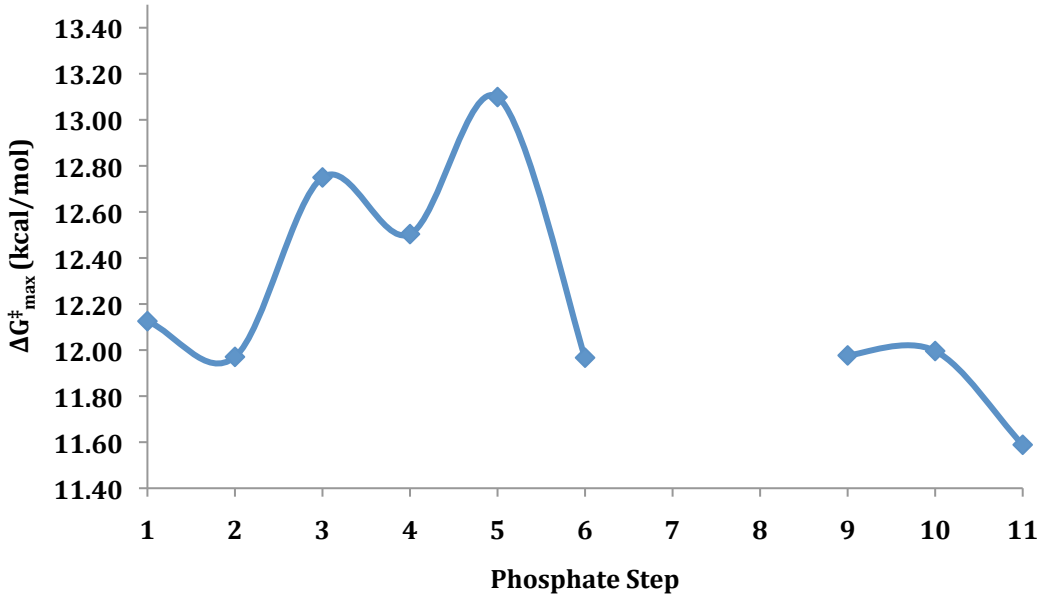

Figure 17. The maximun Gibbs free energy ($\Delta G^{\ddagger}_{max}$) between the BI and BII conformations of the CreGGAG sequence for each phosphate step. Refer to Supplemental Information **Table 14.**

The largest Gibbs free energy of activation was determined to be at step 6, A5pC6. In contrast, the smallest Gibbs free energy of activation was found to occur at the final step, C11pC12. Due to seemingly unreliable data, step 7 and 8 were excluded from **Figure 17.**

Discussion

The Relationship between DNA Conformation and DNA Dynamics

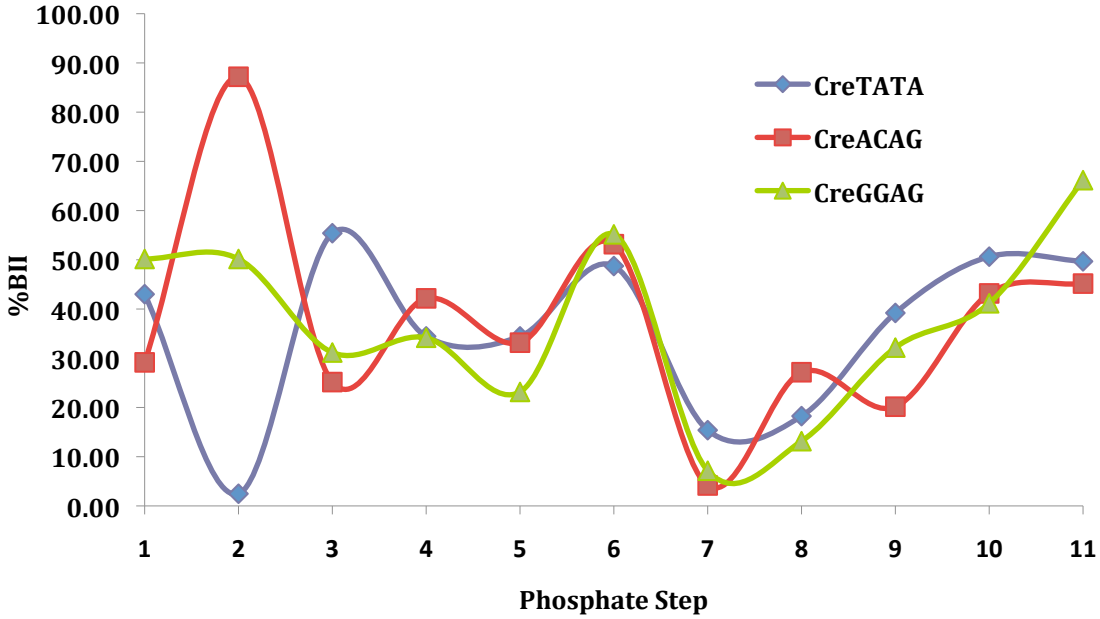

Figure 18. %BII of each phosphate step of the CreTATA, CreACAG, and CreGGAG sequences at 297.15 K (24°C)

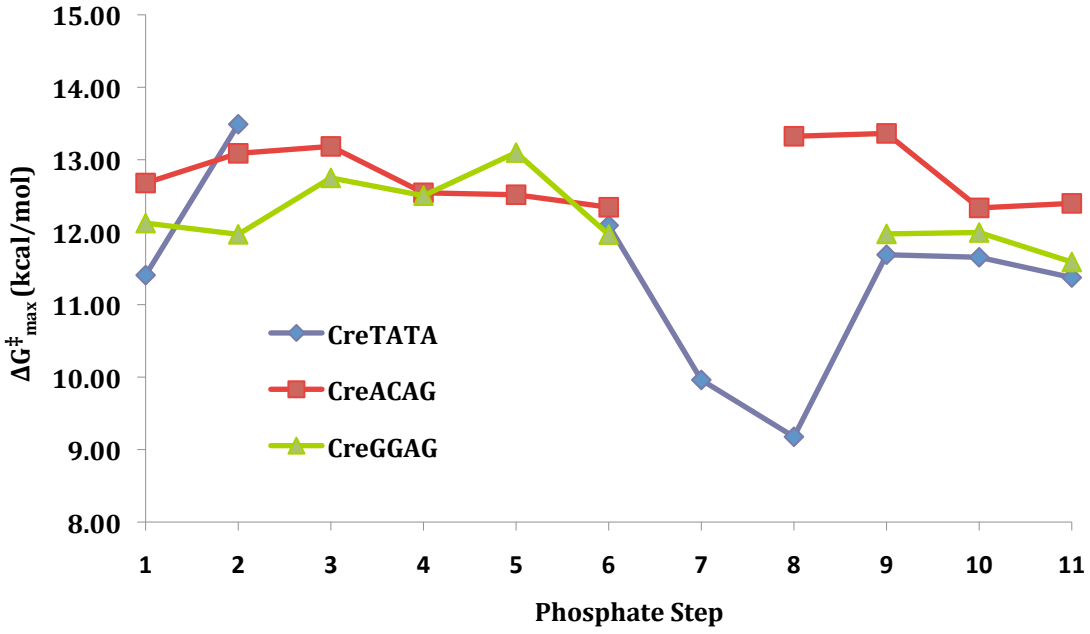

Figure 19. The maximum Gibbs free energy transition barrier ($\Delta G^{\ddagger}_{max}$) between the two B-form DNA conformations of each phosphate step of the CreTATA, CreACAG, and CreGGAG sequences

As demonstrated by **Figure 18,** the differing flanking sequences resulted in varying %BII for the phosphate steps of the CreTATA, CreACAG, and CreGGAG sequences. In general, CreTATA appears to have a dramatic BI step with several phosphate steps that exhibited close to 50% BII character. Comparatively, CreGGAG demonstrated average %BII character for many of the phosphate steps. These average %BII steps for CreGGAG were lower than the later steps of CreTATA. Furthermore, unlike the two other sequences, CreGGAG was defined by average %BII character for step 2. The CreACAG sequence showed an astoundingly high %BII character for step 2. The other phosphate steps of CreACAG appeared to follow a similar trend to the other two sequences. It is evident that the dodecamer sequences with their differing flanking sequences led to different %BII character for not only the flanking phosphate steps, but for the central Cre phosphate steps (steps 5-8) as well.

When comparing the %BII profile to Gibbs free energy in **Figure 19** of each phosphate step we see that our original hypothesis is confirmed. Low and high %BII lead to large energy barriers ($\Delta G^{\ddagger}_{max}$), whereas average %BII results in low energy barriers ($\Delta G^{\ddagger}_{max}$). Step 2 is an exemplary phosphate step to demonstrate the relationship between conformation profile and dynamics. As shown in **Figure 18,** step 2 for CreTATA indicates low %BII character, while step 2 for CreGGAG exhibited high %BII character. Based on conclusions of past studies including Heddi et al., we would have expected to have seen a high energy barrier for step 2 of CreTATA and a low energy barrier for step 2 for CreGGAG. However, our results indicated that step 2 for CreTATA *and* CreACAG were faced with a high energy barrier compared to step 2 of CreGGAG, which was characterized to have average %BII (50.17%). CreGGAG indicated significantly lower %BII compared to CreACAG (87.19%).

This trend indicates a *paradigm shift* in the literature. Several previous studies have concluded that high %BII leads to low energy barriers and therefore an increase in flexibility. [7,8,10] However, this study presents us with an alternative hypothesis. Average %BII (50%) may be an indicator that the phosphate is participating in an exchange between the two conformations and spending equivalent amounts of time in each conformation. This fast exchange between the two backbone conformations may be related to increased flexibility of the backbone.[17]

Lorentzian Fit of 1D ^{31}P Traces

Unfortunately, multiple data points were excluded from **Figures 11**, **14**, **17** and **19** because of the calculation of negative $\Delta G^{\ddagger}_{max}$ values. Due to the definition of the BI conformation as the more stable or lower in energy conformation and the concept that in order to reach the BII conformation an energy barrier defined by $\Delta G^{\ddagger}$ must be overcome, negative values are impossible (**Figure 3**). Therefore, the equation responsible for the ultimate calculation of $\Delta G^{\ddagger}_{max}$ must be modified to ensure reasonable values for the transition.

One possible source for the unreasonable $\Delta G^{\ddagger}_{max}$ values is some of the 1D ^{31}P traces extracted from the assigned phosphate steps of the HSQC spectra lead to undesirable Lorentzian curve fits. Improperly fit Lorentzian curves resulted in incorrect line widths. **Figure 20** shows the Mathematica™ screen view of two Lorentzian fit outcomes.

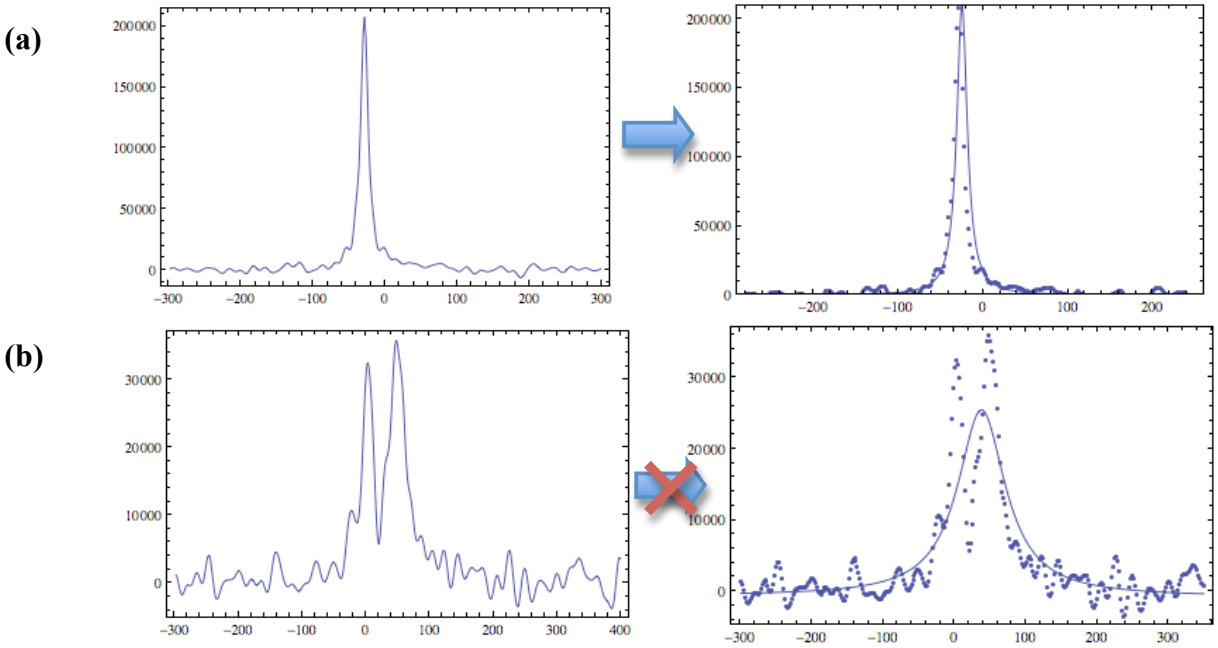

Figure 20. **(a)** 1D ^{31}P trace of T1pA2 of CreTATA at 22ºC trace accurately fit to a Lorentzian curve **(b)** 1D ^{31}P trace of T3pA4 of CreTATA at 44ºC trace improperly modeled by the Lorentzian curve fit

Conclusions and Future Research

Through the analysis of 2D ^{31}P NOESY and HSQC NMR experiments the flanking sequences of CreTATA, CreACAG, and CreGGAG were reconfirmed to result in different %BII profiles. Furthermore, a general trend between %BII (conformation) and $\Delta G^{\ddagger}_{max}$ (dynamics) was determined. Low **and** high %BII phosphate steps caused large $\Delta G^{\ddagger}_{max}$ values; average %BII phosphate steps resulted in small $\Delta G^{\ddagger}_{max}$ values .

These findings indicate a *paradigm shift*, since the average %BII **not** high %BII character leads to a low Gibbs free energy transition barrier ($\Delta G^{\ddagger}_{max}$). Average %BII may be correlated to increased flexibility of the backbone due to the quick exchange between these two conformations. This flexibility of the backbone can be traced back to the ability of proteins to bind to DNA. Binding studies of these three sequences CreTATA, CreACAG, and CreGGAG were conducted in the Hatcher-Skeers Research group using fluorescence spectroscopy to quantify the dissociation constants of a specific drug, 7-AMD.[17,21] The dissociation constants (K_D) collected are as shown in **Table 2**.

Table 2. Dissociation constants (K_D) measured in μM for each sequence including the standard deviation at 25 °C. The program Kaleidagraph was applied to obtain a Michaelis-Menten fit.[17,21]

Sequence	K_D (μM)
CreTATA	5.70 ± 0.45
CreACAG	2.88 ± 0.91
CreGGAG	8.55 ± 1.60

The CreGGAG sequence appeared to contain the most phosphate steps with average %BII and also the lowest Gibbs free energy transition barrier. In conjunction with the reported

dissociation constants, 7-AMD appears to bind the worst to the CreGGAG sequence, due to its high reported K_D value. Therefore, according to these set of data it seems that 7-AMD disfavors the most dynamic sequence and may favor a backbone conformation that is more consistent or rigid. However, there are numerous factors that affect DNA binding and in order to make this conclusion confidently, more sequences must be analyzed through 2D ^{31}P-NMR NOESY and HSQC NMR experiments with accurate line widths. Additional conformation and dynamic studies are necessary to further support the development of average %BII character correlating with low Gibbs free energy.

References

1. Becker W, Kleinsmith L, Hardin J, and Berton G. *The World of the Cell*. 398, 741-775. Pearson Education, Inc.: San Francisco, **2009**.

2. Abi-Ghanem, J.; Heddi, B.; Foloppe, N.; Hartmann, B. *Nucleic Acids Research* **2010**. 38(3):18.

3. Hartmann, B.; Piazzola, D.; Lavery, R. *Nucleic Acids Resesarch* **1993**, *21*, 561-568.

4. Oguey C, Foloppe N, Hartmann B.*PLoS ONE* **2010**; 5(12).

5. Aeling, K., Steffen, N., Johnson, M., Hatfield W., Lathrop, R., and Denear, D. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **2007**, 4 (1), 117-125.

6. Fuxreiter, M.; Simon, I.; Bondos, S. *Trends in Biochemical Sciences* **2011**, 36 (8), 415-423.

7. Gromiha, M.; Siebers, J.; Selvaraj, S.; Kono, H.; Sarai, A. *Gene* **2005**, *364*, 108-113.

8. Gromiha, M. M. *Journal of Biotechnology* **2005**, *117*, 137-145.

9. Djuranovic, D., and Hartmann, B. *Biopolymers* **2004**, *73*, 356-368.

10. Heddi, B.; Oguey, C.; Lavelle, C.; Foloppe, N., and Hartmann, B. *Nucleic Acids Resesarch* **2010**, *38*, 1034-1047.

11. Hartmann, B.; Piazzola, D.; Lavery, R. *Nucleic Acids Resesarch* **1993**, *21*, 561-568.

12. Tian, Y.; Kayatta, M.; Shultis, K.; Gonzalez, A.; Mueller, L.; Hatcher, M. *J. Phys. Chem. B,* **2008**, *113*, 2596-2603.

13. Gorenstein, D. G. *Chemical Reviews* **1994**, *94*, 1315-1338.

14. Pochapsky, Thomas C. and Susan S. *NMR for Physical and Biological Scientists.*1-46. Taylor & Francis Group, LLC: New York, **2007**.

15. Wüthrich, Kurt. *NMR of Proteins and Nucleic Acids.* 44-92. John Wiley & Sons, Inc.: New York, **1986**.

16. Teng, Quincy. *Structural Biology: Practical NMR Application.* Springer Science+Business Media, Inc.: New York, **2005**.

17. Pincus, A. *Sequence Context and Methylation Dependence of the Conformation and Binding Affinity of the Cre Sequence*. B.A. Thesis, W.M. Keck Science Department of the Claremont McKenna, Pitzer, and Scripps Colleges, Claremont, CA, **2011**.

18. Munoz, E. *The Effects of Al^{3+} on DNA: An NMR Study*. B.A. Thesis, W.M. Keck Science Department of the Claremont McKenna, Pitzer, and Scripps Colleges, Claremont, CA, **2011**.

19. Yonis, N. *Methylation induced changes in phosphate conformation of the Cre binding site*. B.A. Thesis, W.M. Keck Science Department of the Claremont McKenna, Pitzer, and Scripps Colleges, Claremont, CA, **2010**.

20. Buckett, D. *Spectroscopic Binding Studies on the Structure and Binding Affinity of the Cre Sequence*. B.A. Thesis, W.M. Keck Science Department of the Claremont McKenna, Pitzer, and Scripps Colleges, Claremont, CA, **2009**.

21. Ullman, S. *Sequence Context Effects on Binding of 7-AMD of the Cre Binding Sequence*. B.A. Thesis, W.M. Keck Science Department of the Claremont McKenna, Pitzer, and Scripps Colleges, Claremont, CA, **2012**.

Acknowledgements

First, I would like to thank Professor Mary Hatcher-Skeers for her continual guidance and support not only as a thesis advisor, but also as an academic advisor and professor throughout the past four years. Further, I am grateful for the opportunities that I have had to collaborate with Professor Katie Purvis-Roberts, Len Mueller, Alexandra Pincus, Evan Munoz, Samuel Ullman, Esmeralda Trejo, Kiley Lawrence, Zach Panzer, Claire Mazahery, and Rebecca Dutta. Also, I would like to thank the entire faculty at the W.M. Keck Science Department for fueling my passion for science during my academic career at Scripps College. Lastly, I would like to acknowledge my friends, in particular Joshua Buss, Tom Boerigter, and Laura Wyatt, and my father and my mother, Gary and Ilona Garton, for always believing in me. Without the contributions of all of these people, none of this would have been possible.

Supplemental Information

The CreTATA Sequence

Table 3. Calibrated chemical shift values of each phosphate step for the CreTATA sequence for various temperatures ranging from 295.15K-317.15K at 2.00K intervals

	295.15	297.15	299.15	301.15	303.15	305.15	307.15	309.15	311.15	313.15	315.15	317.15
T1pA2	-0.52	-0.49	-0.47	-0.45	-0.43	-0.40	-0.38	-0.36	-0.33	-0.36	-0.28	-0.26
A2pT3	-0.95	-0.92	-0.89	-0.87	-0.85	-0.82	-0.79	-0.77	-0.74	-0.71	-0.67	-0.62
T3pA4	-0.37	-0.36	-0.32	-0.32	-0.30	-0.29	-0.27	-0.26	-0.24	-0.24	-0.22	-0.22
A4pA5	-0.61	-0.58	-0.55	-0.54	-0.52	-0.50	-0.47	-0.46	-0.43	-0.41	-0.39	-0.38
A5pC6	-0.61	-0.58	-0.56	-0.54	-0.52	-0.50	-0.48	-0.46	-0.44	-0.41	-0.39	-0.38
C6pG7	-0.46	-0.43	-0.41	-0.39	-0.37	-0.36	-0.34	-0.32	-0.29	-0.27	-0.25	-0.23
G7pT8	-0.80	-0.78	-0.75	-0.73	-0.72	-0.69	-0.67	-0.65	-0.63	-0.60	-0.56	-0.53
T8pT9	-0.77	-0.75	-0.72	-0.71	-0.69	-0.66	-0.64	-0.62	-0.59	-0.57	-0.54	-0.48
T9pA10	-0.54	-0.53	-0.50	-0.47	-0.47	-0.44	-0.41	-0.40	-0.37	-0.35	-0.30	-0.27
A10pT11	-0.45	-0.41	-0.39	-0.37	-0.35	-0.33	-0.30	-0.28	-0.25	-0.23	-0.21	-0.18
T11pA12	-0.44	-0.42	-0.39	-0.37	-0.35	-0.33	-0.30	-0.28	-0.25	-0.23	-0.20	-0.18

Table 4. Extracted line widths of each phosphate step for the CreTATA sequence for various temperatures ranging from 295.15 K-317.15 K at 2.00 K intervals

	295.15	297.15	299.15	301.15	303.15	305.15	307.15	309.15	311.15	313.15	315.15	317.15
T1pA2	16.57	17.35	17.52	15.67	14.22	13.39	12.38	13.52	13.67	13.32	16.15	16.31
A2pT3	17.09	32.86	16.38	14.78	18.84	14.42	16.21	16.20	14.65	12.17	23.02	56.77*
T3pA4	22.20	18.16	19.20	20.61	15.01	14.52	15.89	17.04	21.87	33.50	96.69	81.31*
A4pA5	17.79	32.64	25.88	29.37	16.85	17.54	12.34	14.46	28.51	32.05	40.09*	14.53
A5pC6	20.30	17.66	18.56	17.25	17.17	15.26	14.54	15,11	12.75	15.23	14.47	29.79
C6pG7	23.85	24.54	24.59	28.83	19.97	19.35	20.13	17.20	16.39	16.75	16.53	15.72
G7pT8	18.22	15.78	17.71	18.30	17.09	10.62	15.73	14.60	21.34	32.65	20.43	22.08
T8pT9	18.10	19.31	17.68	16.02	15.63	15.45	15.70	14.04	22.99	22.65	161.65*	151.48*
T9pA10	44.66*	23.90	25.72	24.19	20.08	16.11	14.14	19.67	34.13	11.52	38.33*	37.02*
A10pT11	14.98	16.2	15.37	12.03	11.40	12.47	11.18	11.22	12.16	10.25	10.96	11.01
T11pA12	11.55	19.04	29.43	17.00	16.00	10.98	12.90	13.71	16.09	15.14	17.78	15.17

Table 5. %BII of each phosphate step for the CreTATA sequence for various temperatures ranging from 295.15 K-317.15 K at 2.00 K intervals

	295.15	297.15	299.15	301.15	303.15	305.15	307.15	309.15	311.15	313.15	315.15	317.15
T1pA2	40.15	43.02	44.97	46.97	49.00	52.07	54.20	56.37	59.62	56.73	65.26	67.65
A2pT3	-0.42*	2.47	4.56	6.16	7.78	10.43	13.14	14.89	17.69	20.56	24.53	29.65
T3pA4	54.08	55.40	59.41	59.01	61.76	62.98	65.22	66.49	68.82	69.12	71.53	71.87
A4pA5	31.66	34.44	37.28	38.22	40.17	42.16	45.19	46.26	49.39	51.56	53.77	54.98
A5pC6	31.66	34.44	36.31	38.22	40.17	42.16	44.19	46.26	48.37	51.56	53.77	54.98
C6pG7	45.82	48.73	50.75	52.80	54.89	56.04	58.21	60.42	63.71	66.03	68.39	70.81
G7pT8	13.73	15.39	18.03	19.76	20.54	23.32	25.16	27.03	28.94	31.93	36.02	39.15
T8pT9	16.56	18.24	20.92	21.70	23.49	26.30	28.16	30.07	33.03	35.03	38.11	44.43
T9pA10	38.27	39.21	42.09	45.03	45.08	48.11	51.20	52.33	55.26	57.76	63.17	66.59
A10pT11	46.76	50.64	52.67	54.74	56.86	59.01	62.21	64.47	67.80	70.16	72.57	76.09
T11pA12	47.70	49.69	52.67	54.74	56.86	59.01	62.21	64.47	67.80	70.16	73.62	76.09

Table 6. $\Delta G^{\ddagger}_{max}$ (kcal/mol) of each phosphate step for the CreTATA

Phosphate Step	**$\Delta G^{\ddagger}_{max}$ (kcal/mol)**
T1pA2	11.41
A2pT3	13.49
T3pA4	-2.95*
A4pA5	2.48*
A5pC6	-3.06*
C6pG7	12.09
G7pT8	9.96
T8pT9	9.18
T9pA10	11.69
A10pT11	11.65
T11pA12	11.37

The CreACAG Sequence

Table 7. Calibrated chemical shift values of each phosphate step for the CreACAG sequence for various temperatures ranging from 295.15K-317.15K at 2.00K intervals

	295.15	297.15	299.15	301.15	303.15	305.15	307.15	309.15	311.15	313.15	315.15	317.15
A1pC2	-0.62	-0.60	-0.57	-0.55	-0.53	-0.50	-0.47	-0.45	-0.42	-0.40	-0.37	-0.35
C2pA3	-0.02	-0.02	0.00	0.01	0.01	0.02	0.03	0.04	0.05	0.05	0.06	0.07
A3pG4	-0.67	-0.64	-0.62	-0.60	-0.57	-0.55	-0.53	-0.51	-0.49	-0.47	-0.45	-0.43
G4pA5	-0.49	-0.47	-0.45	-0.42	-0.40	-0.37	-0.35	-0.33	-0.30	-0.28	-0.25	-0.23
A5pC6	-0.59	-0.56	-0.54	-0.51	-0.49	-0.47	-0.45	-0.43	-0.40	-0.38	-0.36	-0.34
C6pG7	-0.37	-0.36	-0.34	-0.32	-0.31	-0.29	-0.27	-0.25	-0.23	-0.21	-0.20	-0.18
G7pT8	-0.88	-0.85	-0.83	-0.80	-0.78	-0.76	-0.74	-0.72	-0.69	-0.67	-0.65	-0.63
T8pC9	-0.65	-0.62	-0.60	-0.58	-0.55	-0.53	-0.51	-0.49	-0.47	-0.45	-0.43	-0.41
C9pT10	-0.71	-0.69	-0.67	-0.63	-0.62	-0.59	-0.57	-0.54	-0.52	-0.50	-0.47	-0.45
T10pG11	-0.49	-0.46	-0.45	-0.44	-0.42	-0.40	-0.39	-0.37	-0.35	-0.34	-0.32	-0.30
G11pT12	-0.46	-0.44	-0.37	-0.38	-0.37	-0.34	-0.32	-0.29	-0.27	-0.24	-0.21	-0.19

Table 8. Extracted line widths of each phosphate step for the CreACAG sequence for various temperatures ranging from 295.15 K-317.15 K at 2.00 K intervals.

	295.15	297.15	299.15	301.15	303.15	305.15	307.15	309.15	311.15	313.15	315.15	317.15
A1pC2	20.75	17.62	16.43	18.03	14.07	13.29	12.34	12.10	11.99	11.72	10.41	10.56
C2pA3	22.54	18.74	18.49	22.12	17.58	17.50	15.18	15.23	14.01	13.51	12.64	12.16
A3pG4	22.47	37.61	27.15	17.31	21.79	30.00	17.90	17.71	17.52	16.95	14.00	13.76
G4pA5	32.21	23.77	23.94	22.87	18.07	17.27	15.81	14.37	15.91	13.52	13.10	13.08
A5pC6	21.51	16.97	18.64	21.54	15.26	15.18	14.86	13.39	12.93	12.83	11.61	11.21
C6pG7	33.23	23.34	24.05	21.49	19.77	17.27	19.85	16.30	16.35	15.84	14.67	13.58
G7pT8	15.30	20.46	19.26	21.11	22.47	17.99	15.76	16.42	14.28	14.38	13.40	12.34
T8pC9	85.95*	24.14	43.55	26.40	22.11	19.04	23.26	22.91	18.74	20.29	17.90	15.60
C9pT10	96.26*	27.04	18.47	20.32	11.57	10.78	11.05	13.05	13.11	12.38	12.94	15.07
T10pG11	22.38	22.80	23.49	23.75	19.57	15.70	16.51	15.05	13.80	13.98	13.08	12.89
G11pT12	25.54	22.05	23.02	26.23	23.23	20.35	17.58	13.74	17.35	12.73	10.40	10.37

Table 9. %BII of each phosphate step for the CreACAG sequence for various temperatures ranging from 295.15 K-317.15 K at 2.00 K intervals

	295.15	297.15	299.15	301.15	303.15	305.15	307.15	309.15	311.15	313.15	315.15	317.15
A1pC2	27.29	29.16	32.08	34.05	36.05	42.30	47.74	50.01	53.43	55.82	53.43	55.82
C2pA3	86.71	87.19	89.70	91.78	93.38	95.01	96.68	98.39	99.04	100.81	100.81	102.62
A3pG4	22.34	25.16	27.03	28.94	31.92	33.93	38.06	40.20	42.38	44.16	44.16	46.90
G4pA5	40.16	42.17	44.21	47.32	49.47	52.70	54.95	57.25	60.67	63.08	66.65	69.19
A5pC6	30.26	33.16	35.12	38.13	40.18	42.27	44.41	46.59	49.90	52.19	54.53	56.92
C6pG7	52.05	53.17	55.33	57.54	58.76	61.05	63.38	65.77	68.22	70.71	72.16	74.76
G7pT8	1.54	4.15	5.80	8.51	10.25	12.02	13.83	15.68	18.65	20.59	22.58	24.61
T8pC9	24.32	27.16	29.05	30.98	33.99	36.01	38.08	40.19	42.35	44.56	46.82	49.13
C9pT10	18.38	20.16	21.98	25.88	26.76	29.75	31.75	34.86	36.97	39.11	42.41	44.67
T10pG11	40.16	43.17	44.21	45.28	47.40	49.57	50.73	52.98	55.28	56.55	58.94	61.39
G11pT12	43.13	45.17	52.30	51.41	52.56	55.83	58.11	61.51	63.91	67.44	71.06	73.65

Table 10. $\Delta G^{\ddagger}_{max}$ (kcal/mol) of each phosphate step for the CreACAG sequence

Phosphate Step	$\Delta G^{\ddagger}_{max}$ (kcal/mol)
A1pC2	12.68
C2pA3	13.09
A3pG4	13.18
G4pA5	12.54
A5pC6	12.52
C6pG7	12.34
G7pT8	-9.64*
T8pC9	13.32
C9pT10	13.36
T10pG11	12.34
G11pT12	12.40

The CreGGAG Sequence

Table 11. Calibrated chemical shift values of each phosphate step for the CreGGAG sequence for various temperatures ranging from 295.15K-317.15K at 2.00K intervals

	295.15	297.15	299.15	301.15	303.15	305.15	307.15	309.15	311.15	313.15	315.15	317.15
G1pG2	-0.42	-0.39	-0.36	-0.34	-0.31	-0.29	-0.26	-0.25	-0.22	-0.20	-0.18	-0.16
G2pA3	-0.41	-0.39	-0.38	-0.35	-0.33	-0.31	-0.29	-0.27	-0.25	-0.23	-0.21	-0.18
A3pG4	-0.60	-0.58	-0.55	-0.53	-0.51	-0.48	-0.46	-0.44	-0.42	-0.39	-0.37	-0.35
G4pA5	-0.57	-0.19	-0.53	-0.49	-0.47	-0.44	-0.42	-0.40	-0.37	-0.35	-0.33	-0.30
A5pC6	-0.68	-0.66	-0.63	-0.61	-0.59	-0.56	-0.54	-0.52	-0.50	-0.48	-0.46	-0.44
C6pG7	-0.36	-0.34	-0.32	-0.31	-0.29	-0.27	-0.26	-0.24	-0.22	-0.21	-0.19	-0.17
G7pT8	-0.85	-0.82	-0.79	-0.79	-0.76	-0.73	-0.71	-0.69	-0.67	-0.64	-0.63	-0.60
T8pC9	-0.78	-0.76	-0.75	-0.71	-0.68	-0.66	-0.63	-0.61	-0.58	-0.56	-0.54	-0.51
C9pT10	-0.60	-0.57	-0.55	-0.53	-0.50	-0.48	-0.45	-0.43	-0.41	-0.60	-0.37	-0.35
T10pC11	-0.49	-0.48	-0.48	-0.44	-0.83	-0.40	-0.38	-0.37	-0.35	-0.33	-0.31	-0.29
C11pC12	-0.25	-0.23	-0.22	-0.48	-0.18	-0.16	-0.14	-0.12	-0.10	-0.08	-0.68	-0.04

Table 12. Extracted line widths of each phosphate step for the CreGGAG sequence for various temperatures ranging from 295.15 K-317.15 K at 2.00 K intervals

	295.15	297.15	299.15	301.15	303.15	305.15	307.15	309.15	311.15	313.15	315.15	317.15
G1pG2	26.14	25.08	22.84	21.53	20.01	19.26	17.18	16.49	16.25	15.56	16.85	15.89
G2pA3	24.09	22.77	16.19	14.56	17.52	15.94	16.26	15.64	14.22	13.99	15.37	13.97
A3pG4	33.33	23.75	21.25	14.81	21.16	18.48	16.10	16.51	18.31	16.71	16.18	11.39
G4pA5	25.72	23.44	23.68	22.38	13.31	17.34	18.41	16.60	13.64	15.32	16.73	15.54
A5pC6	102.32*	11.36	30.46*	20.86	12.91	11.48	19.12	13.75	14.11	12.34	13.41	11.41
C6pG7	29.39	19.59	14.90	24.38	18.78	17.70	18.63	16.27	16.24	17.33	17.58	17.36
G7pT8	43.10	26.07	17.42	19.70	20.20	16.40	15.07	15.69	14.80	13.53	13.03	13.31
T8pC9	21.03	15.92	20.57	16.90	18.67	17.34	14.69	15.03	11.95	10.69	11.72	23.04
C9pT10	20.75	23.28	15.00	12.52	12.67	14.35	11.31	12.46	11.31	14.51	12.96	23.04
T10pC11	19.24	16.66	104.20*	12.06	65.08	49.02	24.25	30.36	22.45	23.86	28.33	37.39*
C11pC12	7.39	15.40	24.19	15.96	13.71	14.04	13.73	13.42	11.60	11.15	11.68	10.86

Table 13. %BII of each phosphate step for the CreGGAG sequence for various temperatures ranging from 295.15 K-317.15 K at 2.00 K intervals

	295.15	297.15	299.15	301.15	303.15	305.15	307.15	309.15	311.15	313.15	315.15	317.15
G1pG2	47.09	50.17	53.31	55.50	58.76	61.05	64.44	65.77	69.29	71.80	74.37	76.99
G2pA3	48.08	50.17	51.29	54.47	56.69	58.96	61.28	63.64	66.06	68.53	71.06	74.76
A3pG4	29.27	31.16	34.11	36.09	38.12	41.23	43.35	45.52	47.74	51.10	53.43	55.82
G4pA5	32.24	34.16	36.13	40.18	42.24	45.40	47.57	49.79	53.13	55.46	57.84	61.39
A5pC6	21.35	23.16	26.02	27.92	29.86	32.88	34.92	37.00	39.12	41.29	43.51	45.79
C6pG7	53.04	55.17	57.35	58.56	60.82	63.13	64.44	66.84	69.29	70.71	73.27	75.87
G7pT8	4.51	7.15	9.85	9.36	12.13	15.15	16.99	18.88	20.80	23.86	24.78	27.96
T8pC9	11.44	13.15	13.89	17.71	20.57	22.45	25.43	27.40	30.50	32.58	34.70	37.99
C9pT10	29.27	32.16	34.11	36.09	39.15	41.23	44.41	46.59	48.82	51.10	53.43	55.82
T10pC11	40.16	41.17	41.18	42.82	47.40	49.57	51.79	52.98	55.28	57.64	60.04	62.50
C11pC12	63.93	66.18	67.46	69.79	72.17	74.61	77.09	79.63	82.22	84.88	87.59	90.36

Table 14. $\Delta G^{\ddagger}_{max}$ (kcal/mol) of each phosphate step for the CreGGAG sequence

Phosphate Step	$\Delta G^{\ddagger}_{max}$ **(kcal/mol)**
G1pG2	12.13
G2pA3	11.97
A3pG4	12.75
G4pA5	12.50
A5pC6	13.10
C6pG7	11.97
G7pT8	-12.70*
T8pC9	-9.62*
C9pC10	11.98
C10pT11	12.00
T11pC12	11.59