The Double Helix in Motion: New Insights into Sequence-specific, Functional DNA Dynamics Using NMR Spectroscopy

by
Evgenia Nikolaevna Nikolova

A dissertation submitted in partial fulfillment
Of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in The University of Michigan
2011

Doctoral Committee:
      Professor Hashim M. Al-Hashimi, Chair
      Professor Carol Fierke
      Professor Nils G. Walter
      Assistant Professor Patrick J. O'Brien

2011

To my family and close ones.

For their endless love and support during this truly metamorphic experience that is graduate research.

Acknowledgements

Table of Contents

Chapter 4

Characterizing the dependence of transient Hoogsteen base pairs on pH, cations, and base modifications, and their cooperativity of formation

Chapter 5

Conclusions and future perspectives...153

Appendix 1

Summary of $R_{1\rho}$ relaxation dispersion data and best-fit parameters for DNA

dodecamers decays and DFT ^{13}C chemical shift calculations

List of Figures

List of Tables

Abstract

DNA is a highly flexible molecule that undergoes a variety of structural transitions in response to cellular cues. Sequence-directed variations in the canonical double helix structure that retain Watson-Crick base-pairing play important roles in DNA recognition, topology, and nucleosome positioning. Here, we use NMR relaxation methods to study sequence-directed dynamics occurring at picosecond to millisecond timescales in variable size DNA duplexes. Traditionally, atomic-level spin relaxation studies of DNA dynamics have been limited to short duplexes, in which sensitivity to biologically relevant nanosecond fluctuations is often inadequate. We introduce a method for preparing residue-specific $^{13}C/^{15}N$-labeled elongated DNA along with a strategy for establishing resonance assignments and apply it towards probing fast inter-helical bending motions induced by an adenine tract. Our results suggest the presence of elevated A-tract independent end-fraying and/or bending internal nanosecond motions, which evade detection in short constructs and that penetrate deep within the helix and gradually fade away towards its interior. By studying picosecond-nanosecond dynamics in short DNA dodecamers with variable length A-tracts, we discover that A-tracts are relatively rigid and can modulate the flexibility of their junctions in a length-dependent manner. We identify the presence of large-amplitude deoxyribose internal motions in CA/TG and CG steps placed in different sequences that likely represent rapid sugar repuckering. Moreover, by using NMR relaxation dispersion in concert with steered molecular dynamics simulations, we observe transient sequence-specific excursions away from Watson-Crick base-pairing at CA/TG and TA steps inside DNA dodecamers towards low-populated and short-lived A•T and G•C Hoogsteen base pairs. We show that their populations and lifetimes can be modulated by environmental factors like acidity, monovalent and divalent ions as well as intrinsic sequence and chemical modifications. The observation of Hoogsteen base pairs in duplexes specifically bound to transcription

factors and in damaged sites implies that the DNA double helix intrinsically codes for excited state Hoogsteen base pairs as a means of expanding its structural complexity beyond Watson-Crick base-pairing. The methods presented here provide a new route for characterizing transient nucleic acid structures, which we predict will be abundant in the genome and constitute a second transient layer of the genetic code.

Chapter 1

Introduction

1.1. NMR studies of DNA dynamics

1.1.1 Role of DNA flexibility in biology

       The DNA double helix carries within its nucleotide sequence multiple codes. The best-known of all is the nearly universal "genetic code", which uses triplets of DNA bases or "codons" to specify the sequence of amino acids that make up proteins, in all known forms of life. But DNA is not simply a repository of genetic information used for the production of proteins. It was realized early that the assignment of codons to specific amino acids is not completely random, but instead has been adaptively selected to minimize translational misread errors[1,2] and frameshift mutations[3], suggesting that there are other layers of information residing within the DNA sequence. Coexisting with the genetic message are "parallel codes" that serve to direct DNA packaging into nucleosomes[4,5], that guide the binding of regulatory and structural proteins to DNA, that carry signals for RNA splicing[6] and secondary structure[7]. All of these codes are conveyed by specific nucleotide sequences. More intriguingly, many have been linked to the flexibility of the DNA double helix that is encoded by its sequence.

       DNA sequence and intrinsic flexibility play essential roles in the positioning and organization of nucleosomes, the fundamental building blocks of chromatin comprised of ~ 160 - 240 base pairs of DNA wrapped around a core of eight histone proteins. Based on patterns of distribution and abundance of short DNA motifs in collections of chicken and yeast nucleosomes[4,5], Widom and coworkers were able to construct a DNA-histone interaction model where AA/TT/TA dinucleotides are favored at the interface of the DNA backbone with histones, while GC dinucleotides are preferentially placed on the outward face, and used this model to predict nucleosome positioning *in vivo* (Figure 1.1).

This means that oppositely directed local bends, presumed to be positive roll and negative roll for GC and AA steps respectively, would add in-phase to produce greater overall curvature. Their team further observed that homopolymeric stretches of adenines, so-called A-tracts, that are believed to adopt a non-generic B-DNA conformation *in vitro*, tend to resist incorporation into nucleosomes, possibly due to their bending stiffness, thereby facilitating the access to nearby origins of replication and promoters by their cognate regulatory proteins[8]. In addition, $(GC)_n$ repeats prone to Z-DNA formation have
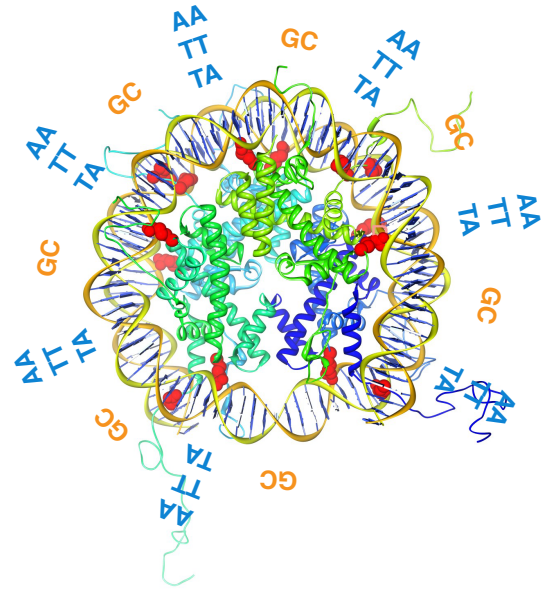
Figure 1.1: Preferential positioning of DNA dinucleotides inside a nucleosome core particle. Red spheres indicate the location of Arginine residues that recognize the narrow DNA minor groove (PDB ID: 1KX3).

been primarily localized to the DNA linking nucleosomes and found to block nucleosome assembly, producing transcriptionally favorable sites for active promoters[9]. The biased positioning of these DNA sequences is largely attributed to their conformational preferences and degree of bending and torsional flexibility[10] that could either increase or decrease the affinity for nucleosome formation. In short, their specific location can modulate DNA accessibility and exert transcriptional control. As a result, functional DNA sites have lower than average occupancy, DNA exons found in between coding sequences are more highly inhabited, while the transcriptionally silent centromeric DNA, the most tightly packed region of the chromosome, contains the highest stability nucleosomes.

Another remarkable feature of DNA is that it can regulate its own readout. The genome is flagged with DNA motifs that architectural and regulatory proteins, such as transcription factors, can single out and specifically bind in order to carry out their functions. The determination of 3-D structures of numerous DNA-protein complexes has provided a detailed picture of complexation. The ever-expanding structural database has revealed a large conformational diversity in protein folds that use a wide repertoire of

interactions to recognize the DNA double helix. It has become evident that a protein-DNA "recognition code" cannot be constructed purely based on a static structural framework because DNA sequences are observed to adaptively change conformation, often in different ways, on binding to distinct protein targets.

We can distinguish two general mechanisms by which proteins recognize DNA sequences – "direct" and "indirect" readout[11,12]. The first involves formation of hydrogen bonds and Van der Waals' contacts between individual base pairs and amino acid side chains, it exploits the unique chemical signature that every base pair has in the major groove, and partially in the minor groove[13]. Protein ligands that utilize this chemical complementarity are expected to be primarily major-groove binders, and this prediction is evident in a wealth of high-resolution structures of DNA-protein complexes[12], exemplified by major DNA binding domains (DBDs) such as the helix-turn-helix motif from λ repressor[14] and the zinc-finger motif from Zif268[15].

More intriguing is the second mechanism, which relies on the readout of the DNA shape, and specifically, the ability of the DNA sequence to assume non-canonical conformations that optimize its interaction with the protein. In this mode of recognition, the DNA helix does not usually make direct contact with the protein but rather has a distinct propensity to deform – kink, bend or twist – in response to protein binding in a way that is complementary to the protein surface. This is observed as local dynamic anisotropy that lowers the energy for formation of sharp kinks and bends seen in the consensus CA/TG DNA dinucleotide upon binding of catabolic activator protein (CAP)[16] (Figure 1.2). Or it takes the form of a larger conformational change such as the severe TATA box DNA bending and major groove

Figure 1.2: Indirect readout of DNA sequences by CAP. Black wedges indicate the primary kink sites in the consensus DNA sequence for CAP binding (PDB ID: 1O3Q).

3

compression used to conform to the concave surface of TBP[17]. And it is seen as a minor groove narrowing in AT-rich sequences, which enhances their negative electrostatic potential and, consequently, interaction with positively charged amino acids (Arginine and Lysine) in the prokaryotic integration host factor (IHF)[18], *Drosophila* Hox protein sex comb reduced (SCR)[19], and many other proteins including histone folds[20] (Figure 1.1) and tumor suppressor p53[21]. On a larger scale, mesoscopic DNA anisotropy could facilitate formation of DNA loops and curvature over hundreds of base pairs. These large-scale architectures are essential in assisting the communication of remotely anchored regulatory proteins, such as two bacterial LAC[18] or two GAL[22] repressors, and the compaction of the bacterial genome[23].

Given the complexity of these DNA-protein interactions, there is no single DNA conformation within a given nucleotide sequence that can satisfy all of them. Therefore, DNA must be found in a dynamic ensemble of interconverting structures, and we find traces of these conformations in the large dataset of crystallographic data for complexes between DNA and proteins. The relationship between DNA flexibility and protein affinity has been examined quantitatively by several research groups[24-28]. By analyzing correlated fluctuations in stereochemical features in an array of DNA-protein crystal structures, Olson *et al.*[25] were able to construct sequence-dependent empirical energy functions that describe local base pair movements in dinucleotide steps, operating in sequence-specific recognition and protein-induced DNA deformations. The "rules" that describe DNA nearest-neighbor dynamics, later expanded to tetramers, hexamers and octamers[29,30], have identified pyrimidine-purine steps (TA and CA/TG) as most flexible and purine-purine steps (AA) as least flexible. They highlight a clear empirical correlation between reduced base pair flexibility and high levels of propeller twisting that can stereochemically lock AA dinucleotides into rigid,
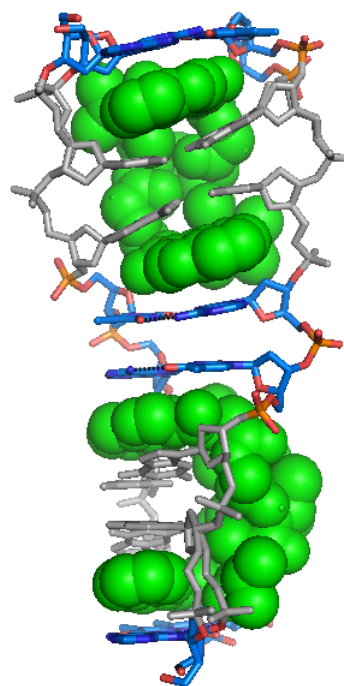
Figure 1.3: Binding of an antibiotic intercalator drug traps non-canonical Hoogsteen base pairs in DNA (PDB ID: 1XVN).

4

inflexible frameworks and inhibit their interaction with proteins. Because of this, stretches of adenines or A-tracts are presumed to be structurally rigid, while the bending observed when A-tracts are phased periodically with the helical repeat, is presumed to originate at other sites[31]. In addition, these studies have discovered other couplings in helical parameters such as bending and base pair displacement that can regulate its superhelical handedness in the context of tightly wound or looped DNA located in nucleosomes or transcription initiation complexes.

Similar recognition strategies that take advantage of sequence-specific DNA flexibility are also used by DNA-binding drugs[32]. DNA is emerging as a prominent drug target, and there is growing consensus that targeting DNA dynamic switches requires a dynamics- rather than structure-based framework. Needless to say, effective tools for analysis of DNA-drug recognition can facilitate the design and development of drugs that discriminate specific DNA sequences over others or can modify the conformational state of particular DNA sites (Figure 1.3), but require extensive knowledge of sequence-specific flexibility of the native DNA state. Thus, the prospect of engineering gene regulation by protein- or drug-DNA interactions places a tremendous importance on how well we understand and can manipulate sequence-dependent DNA dynamics.

Until now, we have focused on DNA flexibility within the general B-DNA framework. However, it is well known that DNA is highly polymorphic in character. Despite its preference for a B-form double helix under physiological conditions, DNA can adopt a variety of non-canonical conformations including cruciforms, left-handed Z-DNA, or G-quadruplexes that are also dependent on its sequence and that can perform physiological functions not achievable by the B-form helix. These structural transitions serve to regulate access to DNA, direct the binding of proteins and drugs, dissipate cellular forces arising from DNA supercoiling, or allow adaptive recognition of diverse protein targets. Such large-scale non-canonical transformations further demonstrate the ability of DNA to take different shapes, encoded in its sequence, in order to fulfill a vast number of functions.

As we have seen, defining the dynamic landscape of the native DNA double helix is of paramount importance to understanding many cellular processes. X-ray structures of native DNA, despite their large numbers, generally highlight smaller-amplitude local

fluctuations around the ground state that paint an incomplete picture of the dynamic movements of bases and sugars and contain no information about the dynamic timescales. A further limitation of the current data is that anisotropic structure factors are infrequently reported in crystal structures. Crystal packing forces can also bias the average structure from a broader conformational ensemble and stabilize higher energy conformations that might not be substantially populated in solution. Moreover, many large and heterogeneous DNA and RNA systems that interconvert between dramatically different conformations could be extremely challenging, if not impossible, to crystallize. Needless to say, the "real-time" flexibility of DNA in any given complex with proteins or small molecules cannot be determined directly from the co-crystal data. One cannot use only the ensemble-averaged parameters, provided by crystallographic data, to learn about the intrinsic motions of the double helix, especially when some DNA states are so low-populated and short-lived that their fractions would have no reflection on the structural ensemble.

Solution state NMR spectroscopy is a powerful experimental tool for characterizing dynamic transitions in nucleic acids on biologically relevant timescales that span picoseconds to seconds, and beyond. It can provide unprecedented insights into the kinetics and thermodynamics of biomolecular motions as well as structural characteristics for low populated states that cannot be observed otherwise in conformational ensembles. Over the years, much greater emphasis has been placed on developing techniques optimized for the study of protein dynamics[33-35]. Recent studies in our lab have demonstrated a successful strategy that combines helix elongation and RDCs to visualize spatially correlated dynamics between two RNA helices that are linked by a functionally important trinucleotide bulge over timescales extending up to milliseconds[36]. NMR relaxation methods geared towards the study of nucleic acids have recently emerged[37,38] and are motivating the study of sequence-specific motions in DNA that are the focus of this thesis. Below I will provide an overview of NMR developments and a detailed perspective on DNA dynamics studies by NMR relaxation experiments that have provided significant insights into global and sequence-specific DNA flexibility.

1.1.2 A brief history of NMR spectroscopy geared towards biomolecular studies

The concept of a nuclear magnetic spin dates back almost a century ago. Pauli was the first to propose the notion of a spin for elementary particles in the early 1920s and later developed an in-depth mathematical theory to describe it. It was not until 1938 that the phenomenon of nuclear magnetic resonance was successfully demonstrated by Rabi[39] in an experiment where he measured the nuclear magnetic moment using a molecular beam traversing a magnetic field. Soon after, in 1946, Block[40] and Purcell[41] revolutionized the field of NMR by extending its application to liquids (H_2O) and solids (paraffin), for which they were awarded the Nobel Prize in 1952. NMR was at its infancy in the 1950s that marked the discovery of the NMR chemical shift[42-45] and the development of many important theories and experiments including the Bloch equations for NMR relaxation[40,46], the effect of chemical exchange on NMR spectra[47,48], the Redfield relaxation in the rotating frame[49], and the Carr-Prucell spin echo method for measurement of transverse relaxation times[50]. Magic angle spinning of solid samples[51] was found to yield high resolution spectra and, combined with pulsed methods for line narrowing and methods for polarization transfer, lay the foundation for studying materials by solid-state NMR.

In the 1960s, the introduction of the fourier transform (FT) technique by Ernst and Anderson[52,53] presented the next powerful invention in NMR. They performed a pulsed FT experiment that enhanced dramatically NMR spectra by replacing the problematic slow sweep of radio frequencies with short pulses of high intensity. Soon after, commercial FT spectrometers that afforded dramatic increase in sensitivity, especially useful at that time for detecting ^{13}C nuclei at natural abundance, became available. This revolutionary step triggered a wave of innovations in NMR technology and techniques that allowed for 2D NMR spectra to be implemented[54,55] and, consequently, for the first complete resonance assignment[56] and 3D conformations[57,58] of small proteins to be obtained with the use of ^{1}H Nuclear Overhauser Effect SpectroscopY (NOESY) and distance restraints, one of the most powerful NMR techniques for structural studies thus far. The late 1980s and early 1990s saw the development of novel 3D and 4D pulse sequences[59-62] designed to target protein and nucleic acid samples enriched with ^{13}C and ^{15}N isotopes, effectively surmounting the resonance overlap problems for larger size biomolecules that ^{1}H spectroscopy was facing.

The pinnacle of high-resolution structure determination for proteins and nucleic acids was the development of biomolecular alignment methods to measure residual dipolar couplings (RDCs) in the solution state. Dipolar couplings, normally averaged to zero due to random Brownian rotational diffusion, could be resurrected in the form of RDCs by orienting a small fraction of the molecules with an alignment media or the magnetic field itself[63-65]. In addition, the last two decades have seen the development and improvement of methods for measurement and interpretation of dynamics on the picosecond to greater than seconds timescales for ^{1}H, ^{13}C and ^{15}N nuclei in proteins and nucleic acids, including spin relaxation (ps-ns) and relaxation dispersion (μs-ms), residual anisotropic interactions (<ms), and time-resolved spectroscopy combined with fast acquisition methods (>s)[66-70]. Furthermore, the development of TROSY-based techniques[71], that exploit the favorable cross-correlation between the dipole and chemical shift anisotropy (CSA) to detect slowly relaxing components (narrow linewidths), has opened up the window for multidimensional observation of large biomolecules (> 100 kDa), previously impossible to study by NMR. Armored with clever innovations and great advances made in the last several decades, NMR spectroscopy is becoming unrivaled in the quantity and quality of dynamic and structural information that it can provide for biomolecules of vastly variable sizes, at atomic resolution.

1.1.3 Historical view of DNA dynamics by NMR

In 1953, Watson and Crick proposed the spiral staircase structure of DNA, a right-handed double helix comprised of two strings of nucleotides held together by hydrogen bonds between opposite nucleobases, later coined as "Watson-Crick" base pairing and B-form DNA (Figure 1.4)[72]. At the time, there were many unknowns about the structure and function of nucleic acids. As it turned out, B-DNA was only the major state of DNA under particular environmental conditions and with particular nucleotide sequences, while it could easily transform into left-handed and triple- or quadruple-stranded structures with distinct biological functions[73].

The atomic-detail exploration of the DNA double helix and its biological functions has benefited enormously from advances in NMR methods for the study of biomolecular structure and dynamics. Prior to the 1970s, very few NMR experiments on

DNA, restriction fragments or homopolymers containing hundreds of base pairs, had been done with ^{1}H being the primary diagnostic nucleus given its high isotopic abundance and ubiquity in nucleic acids. Despite the somewhat unpromising beginning, nucleic acid NMR enjoyed a renaissance in the 1970s with the introduction of high-field FT spectrometers and design of new NMR pulse sequences that gave valuable insights into DNA dynamics.

The early DNA experiments relied on the detection of basic properties such as chemical shifts, coupling constants, resonance intensities and line broadening, and their variation with environmental conditions. Facilitated by ring current calculations of chemical shifts for different nuclei, they were used to extract information about the number and type of bases, hydrogen bonding schemes, the lifetime of base pairs or specific structural features, and the local DNA conformation[74]. For example, Crothers and Patel monitored changes in the chemical shift and linewidth of exchangeable imino/amino protons of short DNA duplexes with temperature and pH to learn about the thermodynamics of transient opening of the double helix[75-77], termed "fraying" or "helix-to-coil" transition. Crothers *et al.*[75] initially hypothesized, based on a two-step mechanism for imino ^{1}H exchange with solvent, that the exchange was limited by helix dissociation. However, Patel *et al.*[76] discovered that the fraying occurred rapidly on the NMR timescale and appeared to be "decoupled" between individual base pairs under certain conditions, being more pronounced for terminal versus internal base pairs, which provided initial clues that complete helix dissociation was not required for exchange to occur. In another study at the time, altered chemical shifts and linewidth profiles for AT
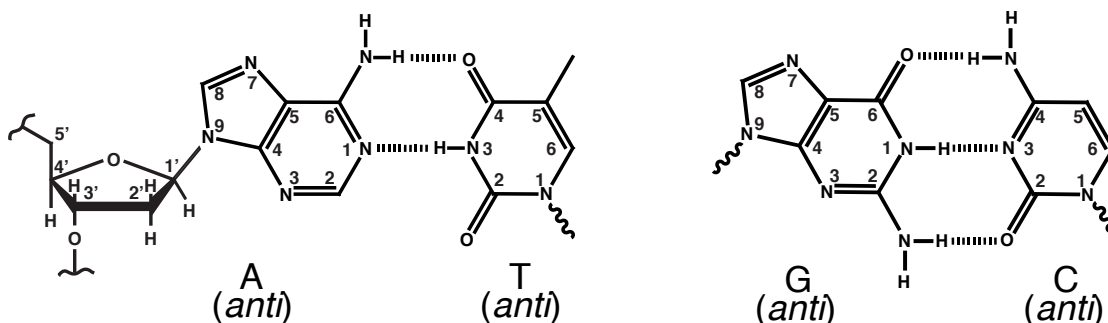
Figure 1.4: Watson-Crick A•T and G•C base pairs.

imino protons at the interface of two homopolymeric blocks, A_{15}/T_{15} and G_{15}/C_{15}, suggested that the conformation and dynamics of the junctional base pairs was perturbed to accommodate the coupling of the different B-form structures[78]. All physical techniques previously used to investigate DNA structure, such as UV and CD spectroscopy, suffered from the fact that they could not draw a direct correlation between observed global properties and specific structural perturbations[79], whereas NMR promised to directly report on the conformation of specific DNA sites and at the level of individual atoms.

In the early 1970s, the time-resolved decay in the fluorescence polarization anisotropy (FPA) of an ethidium dye intercalated between base pairs revealed that DNA is a flexible polymer in solution that undergoes both twisting and bending internal motions[80]. This finding contradicted the previously assumed rigidity of the DNA chain and motivated later studies of internal DNA motions in order to characterize the dynamic behavior of DNA as well as to gain insights into rapid helical dynamics pertaining to sequence-specific DNA recognition by proteins and small molecules. At the time, the development of NMR techniques to measure longitudinal (R_1) and transverse (R_2) spin relaxation rate constants and heteronuclear nuclear Overhauser effect (NOE) for ^{1}H, ^{31}P and ^{13}C at natural abundance stimulated continued interest in characterizing DNA dynamics and improving theoretical models for the analysis of molecular motions. The techniques used were basic saturation/inversion recovery (R_1) and spin-echo (R_2) type experiments. NMR spin relaxation rates depend on global molecular motions as well as internal motions that are faster than overall rotational diffusion, typically less than μs. This makes them sensitive to dynamics on the ps-ns timescale, which could be extended to μs in superlarge molecular weight DNAs like plasmids. Initial solution relaxation studies primarily probed long DNA constructs (> 150 bp) that were assumed to reorient isotropically in solution with rotational correlation times of tens and hundreds of ns to μs, and, unlike fast-tumbling (< 10 ns) short DNA fragments, could provide unique insights into slow bending and torsional mobility in DNA. ^{31}P and ^{13}C nuclei (at natural abundance) were preferentially targeted for relaxation since the extensive network of coupled protons in DNA bases and sugars, as well as the large fluctuation of interproton distances, complicates the interpretation of ^{1}H relaxation data. By measuring ^{31}P, Crothers *et al.* obtained similar ns internal motions for backbone phosphates in dsDNA

when naked or packaged into nucleosomes (0.2 - 0.4 ns), but their reductionist theory was somewhat flawed in that it ignored anisotropic ^{31}P relaxation mechanisms[81]. More in-depth investigations performed by Bolton et al.[82,83] analyzed the relaxation behavior of both ^{31}P and ^{13}C in linear, circular, and supercoiled DNA, including relaxation contributions from chemical shift anisotropy (CSA) with an axially symmetric tensor in ^{31}P and modeling internal motions as unrestricted rotations. Bolton's data could be justified by two types of internal motions: similar large-amplitude nanosecond (~ 1 - 7 ns) mobility for backbone sugar and sugar-phosphate bonds superimposed onto microsecond long-range bending fluctuations, with DNA bases being relatively immobilized. Clearly, motions in the sugar-phosphate backbone were less restricted and large dynamic variability was present among the sugar carbons, ranking them as 5' > 2' > 1', 3', and 4' in decreasing mobility. Subsequent studies by the same group[84,85] revealed that these slower bending motions increased in frequency from linear to circular nicked to supercoiled DNA that could be explained by the excess of conformational free energy and coupling to higher frequency torsional motions in the closed circular forms. These studies generated similar results and established that the DNA helix was a flexible polymer with internal nanosecond torsional motions and microsecond bending motions. Yet, the detailed nature and extent of internal fluctuations in DNA was poorly defined by the existing motional models and the use of simplistic theories.

In parallel, Hogan and Jardetzky were able to accumulate a large set of relaxation data on ^{31}P, ^{13}C and ^{1}H nuclei in variable size long DNA duplexes (140 to 600 bps) under different conditions (temperature, viscosity, and field strength)[86,87]. They recognized that dynamics on a local scale are not "unrestricted" and implemented the formalism of King et al. for a "two-site jump" model, isomerization between two discrete states, inside a cylinder (DNA helix) to analyze contributions from internal motions[86,87]. Their results revealed that the DNA double helix was far from rigid; large-scale fluctuations between nearly isoenergetic conformations were present in all components of DNA – base planes (~ 15 - 20°), deoxyribose rings (~ 20 - 33°), and sugar-backbone phosphates (~ 27°) – with similar internal correlation times on the order of 1 - 2 ns, which was consistent with the idea that they may be coupled. Although simplistic, the two-state isomerization model used by Hogan et al. seemed to yield similar results as a more complex model for

oscillations inside a harmonic potential. For example, Lipari and Szabo were able to fit the Hogan-Jardetzky data to a variety of models for internal motions including two-site jumps, fluctuations of the azimuthal angle of internuclear vectors (twisting), and wobbling of internuclear vectors, which all implied the existence of large-amplitude ns internal motions[88,89]. In an attempt to provide a meaningful, unified picture of internal dynamics in DNA, they formulated and tested the "model-free formalism"[88,89], which has set the standard for analysis of solution NMR relaxation data, although other more elegant and sophisticated motional models have found broad application in solid-state NMR (ssNMR), electron paramagnetic resonance (EPR), and fluorescence depolarization techniques for interpreting fast DNA dynamics. In the model-free approach, internal motions are conceptualized as being partitioned into two variables – a generalized order parameter (S^2) that defines the amplitude of motion and an effective internal correlation time (τ_e) that defines the timescale of the motion – those variables could be easily expanded to suit far more sophisticated models than originally proposed (see Section 1.2.2). Despite the higher sensitivity of ^{1}H and ^{31}P to detect local dynamics, Lipari and Szabo came to the conclusion that ^{13}C spin probes, having simple relaxation properties at natural abundance (the only accessible data set at the time), would be the best candidate to build a coherent physical view of internal motions in nucleic acids since relaxation for ^{1}H and ^{31}P could be easily obscured by uncertain spin interactions and far more challenging to analyze.

Extensive measurements of ^{13}C relaxation at natural abundance and at different fields soon followed made possible by recent instrumental improvements and gain in sensitivity. CPMG pulse sequences for R_2 measurements were customarily used without complications that we see nowadays for ^{13}C-labeled samples, since ^{13}C-^{1}H bonds can be treated as isolated spin systems at natural abundance. ^{13}C relaxation data collected by Levy et al.[90-92] for large DNAs (120 – 160 bp) and at multiple fields (proton 150 MHz to 500 MHz) indicated that both the base and sugar undergo rapid uniform internal motion (20 - 25º) with a correlation time of ~ 1 ns, arguing against independent flexibility of the DNA backbone. It also provided a basis set for careful evaluation of current models for internal and overall motions, and other experimental insights into ^{13}C CSA contribution to relaxation. For internal motions, Levy et al. found that "diffusion-in-a-cone" was the

most realistic model to represent fast local fluctuations. For overall motions, they rejected the rigid-rod model, especially for DNAs longer than the persistence length (147 bp), in favor of the "flexible rod" model – much more disordered, axially symmetric overall tumbling that involves considerable, high frequency helical bending. At the basis of these conclusions were substantially smaller linewidths and larger R_1 and NOE values for long DNA fragments that could not be reconciled with predicted correlation times for a rigid rotor. Unlike other sugar ring carbons, the data fits for sugar C2' sites could be improved considerably by incorporation of two concurrent internal motions that could justify the disproportionately large dynamic amplitudes seen in prior studies. Levy and colleagues also made another interesting discovery: they found that DNA underwent a spontaneous transition to an ordered phase at some critical high concentration, inversely dependent on DNA length, where C3', C4' and especially C5' motions became progressively frozen, while rapid C2' motions persisted. This led them to believe that molecular crowding and the phase transition to an ordered state dramatically reduced collective bending fluctuations, whereas the "decoupled" local sugar dynamics felt largely by C2' spins were still retained.

The power of NMR was also employed to elucidate of the unusual and somewhat controversial conformation and dynamics of DNA homopolymers, poly(dA-dT) and poly(dG-dC); the first proposed to adopt a heteronomous (or Hoogsteen paired) double helix with mixed sugar conformations in solution[93,94]; the second characterized by extreme helical stiffness[95] and prone to dramatic B-to-Z DNA transitions. Poly(dA-dT), also found by fluorescence anisotropy to be torsionally stiffer than random DNA[95], was especially interesting because of the recent discovery that short runs of adenines (A_{3-6}) or A-tracts cause anomalous electrophoretic behavior in kinetoplast DNA[96-98] and macroscopic curvature when periodically phased by ~ 10 bp[99,100]. Based on proton 1D and 2D (NOE) relaxation studies that provide information on interproton distances, Kearns and colleagues resolved the controversy around the poly(dA-dT) structure by showing it adopts B-form with WC base pairs and C2'-endo sugar puckers[101-103]. They utilized the model-free formalism to fit the data for base protons to progressively more complex models, where simple torsional motions alone could not satisfy the experimental data. Though complex and perhaps over-interpreted, their best fit included fast local

torsional base motions (0.2 - 2 ns, ± 20°) coupled with slower longitudinal base motions along the helix (10 - 100 ns, ± 25°) that were superimposed onto fast collective torsional motions of the helix (~ 6 ns) (or alternatively a three-state jump with a single correlation time of ~ 3 ns). Similar results were obtained by Poly(dG-dC) in B-form[104], while ^{1}H relaxation data for the non-canonical left-handed Z-form DNA suggested slower or lower amplitude backbone motions.

The NMR-based view of DNA motions was consistent with concurrent studies of random DNA sequences[105], poly(dA-dT)[106,107], and poly(dG-dC)[108] using another powerful technique, EPR. EPR can probe torsional and flexural motions as well as the stiffness of the double helix to twisting with the use of stable free radicals ("spin labels") coupled to the DNA. By installing nitroxide spin labels covalently bound to pyrimidine bases in poly(dA-dT) and poly(dG-dC) respectively, Bobst *et al.*[106] and Strobel *et al.*[108] were able to detect base movements with correlation times of several ns (~1 - 4 ns) in B-DNA polymers but ~ 2-fold slower base dynamics in Z-DNA, arguing for increased helix rigidity in the non-canonical form. These EPR studies use a continuous elastic model, where DNA is treated as a flexible rod that exhibits individual base pair motions and correlated bending and twisting deformations, as described by Barkley and Zimm[109]. Another notable model for interpretation of fast internal motions from EPR, ssNMR and fluorescence polarization anisotropy (FPA) studies is the weakly bending rod model by Schurr *et al.*[110]. There, DNA is modeled as spherical beads in a chain coupled to nearest neighbors by a harmonic potential characterized by a torsion, R, and a bending, κ, elastic constant. This elegant and more sophisticated formalism has found broad applications in the analysis of dynamic linewidths from EPR and ssNMR data, FPA where the model-free approach can fail to describe the motional behavior.

In the 1980s, ssNMR emerged as a powerful partner in the study of internal DNA dynamics and conformation, especially for large molecular weight species[111]. The typical probes were backbone ^{31}P and ^{2}H incorporated at purine C8 or thymine methyls. Rotational motions in solids are generally restricted in amplitude, angular dispersion or rate compared to those in liquids. However, relatively fast internal motions are sufficient to reduce or even average CSA and dipolar nuclear interactions in ^{31}P. The ^{2}H nucleus, though existing at low abundance in DNA, is highly favored because of its simple

relaxation mechanism from the dominant quadrupolar interaction and its large dynamic range spanning ps to μs by employing relaxation and lineshape analysis, and even longer by other techniques[111]. The ^{2}H quadrupolar coupling and spin-lattice relaxation can provide information about both the amplitude and timescales of motions. In early studies, Shindo and coworkers used the CSA of ^{31}P to study the conformation and dynamics of highly oriented Li-DNA fibers in the A-, B- and C- DNA forms[112]. In parallel, Opella and coworkers[113,114] investigated internal motions in large molecular-weight calf thymus DNA by employing both ^{31}P and ^{2}H relaxation measurements. Opella *et al.* observed non-axial averaging of ^{31}P CSA tensor on timescales faster than 100 μs, while no averaging of the ^{2}H powder pattern of purine C8-labeled DNA indicated large internal flexibility of the phosphate backbone and absence of large-amplitude ns motions in the nucleobases[113,114]. These findings were consistent with short ^{2}H spin-lattice relaxation times in poly(I)•poly(C) measured by Bendel *et al.*[115], from which they estimated amplitudes of < 5° for base components using the diffusion-in-a-cone model. In another study, Mai and coworkers[116] were able to associate the transition from A-DNA to B-DNA with ± 30° backbone fluctuation by monitoring ^{31}P lineshapes, relaxation and ^{31}P-^{1}H NOE with increasing levels of hydration.

The first attempts to quantitatively model the ^{2}H DNA dynamics was reported in a series of papers by the Kearns and Vold groups[117-121]. The most complete set of ^{2}H spin-relaxation data for C8-deuterated Li-DNA polymer in the B-form was measured by Brandes *et al.*[117], where the decrease in the quadrupolar splitting was interpreted using (i) diffusion-in-cone to yield high order parameters for the base $S^2 \sim 0.97$ ($\theta \sim 12^o$) and $S^2 \sim 0.94$ ($\theta \sim 16^o$) at 66% and 84% hydration or (ii) restricted biaxial diffusion in two planes, parallel (ϕ) and perpendicular (θ) to the base plane, to yield angles $\phi \sim 8^o$ and $\theta \sim 12^o$ at 66% hydration. Their deuterium data clearly suggested that the nucleobases move into a new motional regime as they become more hydrated, while the parallel change in ^{31}P and ^{2}H relaxation rates with hydration pointed to couplings between base and backbone motions. The results by Brandes *et al.* were reproduced by subsequent studies of B-DNA base dynamic anisotropy, showing that the amplitude of base pair tilting motions increases with the degree of hydration from $\theta \sim 5^o$ to $\theta \sim 15^{o}$[119].

While a compelling picture for DNA dynamics had started to emerge for large systems, little was known about fast DNA dynamics at the level of individual residues in specific sequences. Most NMR studies in liquids and solids had focused on long DNA fragments with more complex relaxation mechanisms and where individual residues were intractable. Due to substantial spectral overlap and unfavorable relaxation properties (large linewidths) in high MW duplexes, the NMR field shifted more and more towards the investigation of short synthetic DNA fragments. However, a large drawback of using short DNA constructs is that the dynamic information about ns motions occurring at the same frequency as rotational diffusion would be lost. The first comprehensive relaxation study of short DNA fragments was done by Kearns and colleagues[122-124] in the early 80s, where they measured R_1 and R_2 relaxation rates of individual imino protons using inversion recovery and Hahn spin-echo pulse sequence with long selective pulses. Based on analysis of ^{1}H relaxation rate constants dominated by dipolar interactions with their directly bonded nitrogen and closest ^{1}H neighbors in a rigid rotor overall diffusion frame, internal motions were found to have relatively low contributions.

Due to uncertainties in DNA structure, as discussed above, ^{1}H spins are not ideal probes for to quantifying DNA motions. Because of the low abundance and sensitivity of ^{13}C, ample efforts were directed towards the isotopic ^{13}C incorporation in DNA nucleotides to facilitate resonance assignment, structure determination and, last by not least, dynamics studies that would make the unfavorable ^{1}H spins obsolete in studies of rapid ps-ns DNA motions. For example, by labeling the thymine base with ^{13}C at position C6 in a synthetic short DNA hairpin, Williamson and Boxer[125,126] were able to examine, site-specifically, the conformational flexibility of the DNA loop region as well as the relative contributions of the two major relaxation mechanisms, dipole-dipole and CSA interactions, to ^{13}C relaxation rates. Using the model-free formalism with isotropic tumbling and one internal correlation time (considering the small DNA size), they found very large amplitude base fluctuations in the T-loop with $S^2 \sim 0.5 - 0.6$ ($\pm 31 - 36$ °) with internal correlation times of $\sim 10 - 50$ ps. By comparison, Borer et al.[127] reported much lower flexibility ($S^2 \sim 0.8$) for base-paired purines and pyrimidines in short, palindromic duplexes (4 - 8 bp) compared to Williamson et al., but relatively larger flexibility of sugar carbons ($S^2 \sim 0.6$). They resorted to studying very short DNAs in order to measure

natural abundance ^{13}C relaxation (R_1 and NOE) at site-specific resolution. Apart from characterizing the dynamic signatures of internal residues, one of their major findings was that, due to partial strand melting, terminal residues were extremely disordered ($S^2 \sim$ 0.2-0.6). In another study, Boxer and colleagues[128] paired NMR ^{13}C relaxation with depolarized dynamic light scattering (DDLS) relaxation, which is sensitive to only end-over-end rotational correlation times and can easily distinguish between DNA shapes and sizes, for example, duplex and hairpin conformation that their DNA fragments partitioned between. They demonstrated that the two techniques complement each other very well to resolve uncertainties in the interpretation of overall motions from NMR, and that DDLS was a viable and non-invasive alternative to EPR methods for studying overall dynamics[128].

The success of subsequent solution NMR studies to accurately characterize sequence-specific motions in short DNA duplexes replied primarily on simple, efficient methods for high-yield incorporation of uniformly of fractionally ^{13}C/^{15}N-labeled nucleotides in DNA fragments by *in vitro* enzymatic reaction of bacterial systems that emerged in the 1990s[129-133]. Because the traditional CPMG relaxation experiments suffer significantly from homonuclear ^{13}C-^{13}C interactions in uniformly ^{13}C-labeled samples, $R_{1\rho}$ relaxation pulse sequences were implemented in the measurement of transverse relaxation rates (see Section 1.2.2), originally developed by Kay and coworkers for the study protein dynamics[68,134,135]. Solution NMR ^{13}C relaxation studies focused on characterizing rigorously site-specific motions in biologically relevant DNA sequences, a few prominent examples being the AT-rich *EcoR*I endonuclease consensus sequence (Drew-Dickerson dodecamer)[136], GC-rich *Hha*I methyltransferase target DNA sequence[137-139] as well as damaged or drug-bound DNA[140,141], using the model-free formalism. Studies on damaged DNA, still performed at ^{13}C natural abundance, identified a significant increase in ps-ns backbone dynamic disorder around the site of psoralen damage in DNA[141], or increase in order at sugar moieties around the site of drug intercalation[140], relative to unmodified DNA. Spielmann and coworkers further utilized their ^{13}C relaxation parameters for unmodified DNA to construct correlations, for the first time, between DNA dynamics and structure using high-resolution structures[142]. Their most prominent findings were (i) dynamic couplings between deoxyribose conformation

for at dinucleotide steps and across the strand and (ii) profound effects of inter base pair spacing (rise) on collective internal motions, also supported by MD simulation analysis of rise fluctuations[143].

More general dynamics trends across canonical DNA sequences revealed that non-terminal DNA nucleobases were relatively rigid ($S^2 > 0.9$), consistent with prior studies on long heterogeneous and homopolymeric DNA fragments. A much greater dynamic and sequence-specific variability was observed at the sugar backbone (S^2 as low as ~ 0.5). One prominent trend across a number of dynamic studies was the enhanced backbone motions (low S^2) in cytosine deoxyribose moieties, especially those found at YR or YY steps (CG, CA, CT) shown to be conformationally anisotropic based on crystallographic studies (above)[136,141,144,145]. By using a combined NMR and MD approach, Duchardt *et al.* were able to link low NMR order parameters at cytosine sugars to rapid sugar repuckering interconversions between the predominant C2'-endo and the less common C3'-endo state[136], found in A-form helices. Drobny, Varani and colleagues[138,139,146,147], who employed a cojoint analysis of solution and solid state NMR data using ^{13}C and ^{2}H as dynamic probes, hypothesized that the extensive large-amplitude fluctuations in the cytosine targeted for methylation by *Hha*I methyltransferase (5' GCGC) relative to other cytosines can be used for specific DNA recognition. Additionally, they found that C5-methylation at the same cytosine quenched significantly the amplitude of motion, suggesting how backbone dynamics may assist the endonuclease in discerning between methylated and unmethylated DNA.

Considerable attention in the last two decades has been also concentrated on elucidating motions occurring on NMR timescales much slower than overall diffusion, base pair opening dynamics and chemical exchange processes occurring on μs-ms timescales. Base pair opening dynamics are typically examined by NMR techniques that probe the transfer of imino/amino protons to solvent upon spontaneous opening of individual base pairs. These protons are relatively well protected from solvent due to hydrogen bonding interactions inside base pairs and have sufficiently slow exchange rates to allow detection of their NMR signals. Depending on the exchange times, two general experimental scheme are used – magnetization transfer from water (or hydrogen exchange) experiments for faster exchange times ranging between $\sim$ 10 s to 5 ms and

tims-resolved hydrogen-deuterium exchange for slower exchange times greater than 100 s. Because base pairs in regular DNA helices are usually shorter lived than the detection limits of hydrogen-deuterium exchange, magnetization transfer from water experiments are routinely used, where the magnetization of water protons is initially selectively perturbed, inverted or zeroed in "inversion" and "saturation" recovery methods respectively, and this perturbation is allowed to transmit to exchangeable DNA protons via the exchange process for a variable mixing delay.

Even though the experimental schemes for studying proton exchange date back to the 1970s and were employed in the 1980s to characterize helix opening[124] in short DNAs, the accurate quantitation of base pair lifetimes did not come until Gueron and colleagues implemented a robust method of measuring imino proton exchange rates as a function of base catalyst concentration, which clearly established that base pairs open one at a time[148]. By collecting data with increasing amounts of base catalyst, one can extract individual opening/closing rate constants and the base pair dissociation constant that permits determination of the opening and activation free energies. Since then, the largest contributions to the field of DNA base pair opening dynamics have been made by the groups of Russu[149-157], Gueron and Leroy[158-165]. Together these studies reveal that base pair lifetimes in heterogeneous DNA helices typically range within ~ 1 - 25 ms for A•T and ~ 5 - 50 ms for G•C (at room temperatures), while base pair open states occur very transiently with typical lifetimes of ns (~ 100 - 300 ns). A•T base pairs that belong to A-tract sequences are an exception to this trend – longer base pair lifetimes were found inside longer A-tracts, progressively increasing towards the A-tract interior to ~ 125 ms[160] and evidence for a secondary base pair opening mode was put forward with longer lived open states approaching 1 μs[166]. In striking contrast to the behavior of A-tracts, GC tracts are characterized by unusually rapid base pair opening kinetics and short base pair lifetimes < 5 ms[167], which was proposed to play a role in the singling out of dynamic cytosine nucleobases by *Hha*I methyltransferases (see above). Similarly, G•C base pairs duplexes containing an internal CACA/TGTG motif were found to open at 3-8X higher frequencies than if the second G•C base pairs were inverted to a C•G[151] and provided clues for the intrinsically lower base pair stability at consecutive CA/TG steps. These investigations have highlighted that the opening of the double helix, much like rapid

internal motions, can be strongly influenced by sequence-specific features of the DNA helix. A continued motivation behind base pair opening studies of canonical and non-canonical/damaged DNA duplexes has been that spontaneous base excursions to extrahelical states can play a critical role in the recognition of proteins, particularly those that capture and modify bases or excise damaged nucleotides. In fact, recent reports by the group of Stivers strongly suggest that an enzyme that removes uracil from DNA (uracil DNA glycosylase, UNG) uses thermally induced base pair opening and not enzyme-induced base flipping to discriminate between U and T in A•U and A•T base pairs[168].

While NMR proton exchange experiments can monitor the ms breaking of base pairs held together by hydrogen bonds, measurements of resonance linewidths and chemical shifts for non-exchangeable protons can tell a more detailed story about the movements of individual bases and sugars at the µs-ms timescales. Peak linewidths are described by the transverse relaxation times (R_2) that can be significantly influenced by µs-ms chemical exchange contributions (see Section 1.2.2). The exchange between chemically distinct conformations is also associated with differences in chemical shifts between the two (or more) exchanging species. The simplest approach to identification of chemical exchange is to observe a non-monotonic variation of the "excess" linewidth and chemical shift with temperature, and this analysis can provide some estimation of chemical exchange parameters. Very few reports of site-specific conformational transitions have been published thus far, for instance, at the base H2 and H8 sites of TA step in the context of Pribnow box DNA promoter sequence[169]. To obtain a quantitative description of chemical exchange at various TA steps, subsequent studies employed ^{1}H rotating-frame ($R_{1\rho}$) relaxation dispersion experiments allowing them to estimate exchange rates on the order of ~ 10 - 150 µs[170] (see Section 1.2.2). ^{1}H and ^{13}C relaxation dispersion targeting CA/TG steps[171] and G•G mismatch base pairs[172] have similarly identified conformational flexibility with exchange times of ~ 20 - 300 µs and ~ 70 µs respectively. These transitions have been attributed to *anti-syn* base rotation in the G•G pair, where structural studies support the existence of these states, while the origin behind CA/TG steps dynamics have not been elucidated. These early experiments were prone to significant errors for ^{1}H spins and relied on 1D-resolved resonances for unlabeled DNA

samples, and in general provided limited information beyond the timescales of the conformational exchange. The development of sensitive methods suitable for ^{13}C-labeled nucleic acids have lagged behind those for proteins, where experiments were continuously improved and where elaborate specific-residue/atom ^{13}C-labeling strategies practically eliminated complications from CPMG relaxation dispersion methods[68,173]. A significant obstacle to the application of protein-based CPMG relaxation dispersion schemes to uniformly ^{13}C-labeled nucleic acids are the extensive network of coupled carbons in the aromatic and sugar rings. In this thesis, we will present a recent development in carbon $R_{1\rho}$ relaxation dispersion experiment that is suitable for ^{13}C-labaled DNA samples and utilizes low spinlock fields to access millisecond timescales.

In summary, we have provided a detailed overview of the application of NMR relaxation and relaxation dispersion experiments to study dynamics in unlabeled and labeled, primarily canonical DNA systems with vastly variable sizes. After decades of investigation, a view of the DNA helix as a highly dynamic biopolymer is slowly taking shape. Currently, there is an urgent need for more extensive characterization of sequence-specific DNA dynamics that can aid in the construction of a dynamics-based set, assisted by other techniques such as simulations of molecular dynamics, for analysis and prediction of protein and small molecule binding. This would require a hand-in-hand development of experimental and computational tools.

1.2 NMR Theory and Methods

1.2.1 Chemical Shifts and Resonance Assignments

Every NMR active atom in a DNA molecule (e.g. ^{1}H, ^{13}C, ^{15}N) has a unique property, the NMR chemical shift, which can discriminate between nuclei of the same type and allow for the atomic-resolution study of large biomolecules. In general, each type of nucleus resonates at a characteristic frequency, known as the Larmor frequency, when placed in an external magnetic field. The Larmor frequency depends on the separation between energy levels for a given nuclear spin and the strength of the magnetic field, so one would naturally predict that all ^{1}H, for example, would resonate at the same frequency. However, the total magnetic field experienced by a nucleus in the context of a molecule is also influenced by local magnetic fields induced by electron
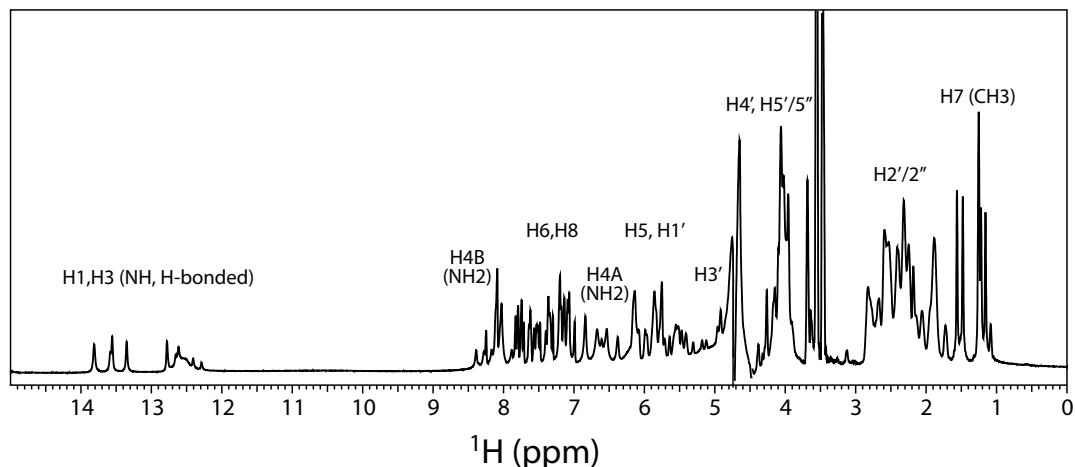
Figure 1.5: ^{1}H NMR chemical shifts in a DNA duplex.

currents in the molecular orbitals. These perturbations in the local electron "cloud" or density distribution of each nucleus are extremely small, yet they cause deviations in its resonance frequency significant enough to be resolved by modern NMR spectrometers operating at high magnetic fields.

The chemical shift is defined as the variation of the nuclear magnetic resonance frequency for the same kind of nucleus due to variations in its electronic environment. Atoms that have higher induced fields or higher electron density are referred to as *shielded* and have upfield (lower) chemical shifts, while those that have smaller induced fields and are deprived of electrons, are called *deshielded* and have downfield (higher) chemical shifts. The degree of shielding is determined by electron density, electronegativity of neighboring groups (electron-donating or withdrawing) and anisotropic induced magnetic field effects. Thus, the chemical shift constitutes a particularly useful probe of the local geometry and bonding at each nucleus – bonding partners, bond lengths and angles, torsion angles (sugar pucker, *syn* vs. *anti* base configuration) – and can sense structural perturbations and formation of new interactions such as H-bonding and stacking upon ligand binding and chemical modification.

The first essential step in an NMR study is to assign the chemical shift of each resonance peak in an NMR spectrum to a given atom (Figure 1.5). NMR methodologies for resonance assignments in nucleic acids have been well established and extensively reviewed[174,175]. In general, these experiments establish connections between nearby

nuclei via through-bond and through-space correlations. Subsequently, a network of sequential connectivities can be constructed for a known sequence of presumed geometry (i.e. B-form or A-form in helical regions) and can aid in the determination of secondary structure and base pairing where it is not known (i.e. non-canonical regions, bulges and loops). Original methods used primarily 2D homonuclear or heteronuclear ^{1}H/^{31}P correlation spectroscopy to establish assignments for abundant ^{1}H and ^{31}P nuclei (~ 100 %), and sometimes protonated ^{13}C nuclei at natural abundance (~1%), but provide only partial assignment and limited structural insights. More recently, elaborate strategies have been designed for the full assignment (^{1}H, ^{13}C, ^{15}N and ^{31}P nuclei) of small to moderately sized DNA/RNA constructs that require ^{13}C/^{15}N isotopic enrichment and long NMR acquisition times. Recent developments of *in vitro* ^{13}C/^{15}N-labeling of DNA sequences [129,133] at affordable cost has made it possible to devise heteronuclear through-bond methods that can successfully target almost any atom in the molecule, enhancing the prospect for high-resolution structural studies. However, those methods would not be the subject of this study and are reviewed elsewhere[174]. Here, we will outline only conventional techniques for assignment of ^{1}H, ^{13}C and ^{15}N in unlabeled DNA samples, which were sufficient to obtain the necessary peak assignments in our study.

Prescriptions for partial assignment of short unlabeled DNA constructs (< 20 bp) involve a combination of through-space and through-bond NMR methods. Through-space ^{1}H,^{1}H correlation experiments rely on the nuclear Overhauser effect (NOE) that occurs between protons in close proximity to one another (< 5 Å). NOE is the transfer of nuclear spin polarization from one nuclear spin population to another via dipolar cross-relaxation mechanisms, where the transfer efficiency is highly dependent on the distance between the interacting spins. Because of that, NOE SpectroscopY (NOESY) is by far one of the most powerful tools for structural investigation of biomolecules in the solution state. In a basic experimental scheme, an initial RF pulse creates transverse magnetization, which precesses during the evolution time and encodes the chemical shift in the second dimension; a second pulse returns the magnetization of all spins to the Z-axis, allowing for "mixing" or cross-relaxation to occur; finally, a third pulse converts the mixed magnetization to the transverse plain that is observed during acquisition. The result is a 2D ^{1}H,^{1}H spectrum where proton resonance peaks occupy the diagonal and the cross-

peaks due to dipolar mixing appear as off-diagonal elements labeled by the frequencies of the pair of interacting protons in each dimension.

The simple ^{1}H,^{1}H NOESY scheme provides a wealth of information to build correlations within a single nucleotide or between neighboring nucleotides. First, one can establish the base pairing pattern (or secondary structure) by cross peaks between imino and amino protons (i.e. G H1 to C H6a/b; T H3 to A H6a/b) and "walk" sequentially from the imino/amino proton of one base pair to the next to determine which residue they belong to (Figure 1.6). The identity of the non-exchangeable base H2, H6, H8, H7M and sugar H1' protons can also be resolved via cross-peaks with the exchangeable imino/amino protons in the same or

Figure 1.6: Imino ^{1}H NOE walk and connectivities to C H4a/b, A H2, and T CH3.

adjacent nucleotide, especially for nearby A H2 and T H3/G H1 protons in the B-form conformation. However, non-exchangeable H6/H8/H5M/H1'/H2',2" sites are commonly assigned via a different NOE pathway that connects the base H6/H8/H5M to the sugar H1'/H2',2" of the same nucleotide and the sugar H1'/H2',2" of the 3'-adjacent nucleotide (Figure 1.7). Thus, one can assign consecutive residues by moving from the 5' to 3' direction along the DNA strand. Further correlations can be identified between pyrimidine H5 and H6 on the same residue, sugar protons (H1', H2'/2", H3', H4' and H5'), H6/H8 of adjacent residues, etc. Proton assignments could be easily translated to 2D ^{1}H/^{13}C and ^{1}H/^{15}N heteronuclear correlation spectra (HSQC or HMQC) acquired at

natural abundance or for equivalent $^{13}C/^{15}N$ labeled DNA samples. NOE walks provide a robust approach for assignment of helical B-form (or A-form) regions with predictable range of inter-proton distances, but may not apply to non-canonical conformations and base pairing patterns. Unique NOE cross-peaks could be critical in differentiating between Watson-Crick (WC) and other base pairing configurations such as Hoogsteen (HG) base pairs discussed in Chapter 3.

Additional assignment methods are based on though-bond coherence transfer due to appreciable scalar couplings (J coupling) between nuclei, yielding similar 2D correlation spectra. Correlation Spectroscopy (COSY) experiments can connect directly coupled protons with the sugar spin system (H1' to H2'/2", H3' and/or H4') or base H6 to H5/H7M for pyrimidines C and T. Total Correlation Spectroscopy (TOCSY) can take it a step further by correlating nuclei that are connected by a chain of couplings. Because proton J couplings in the deoxyribose moiety vary with the sugar ring conformation, one can obtain further insights into the sugar pucker type (S or N) and backbone torsion angles. For example, transfer from H1' to H4' can be achieved for S-type pucker or mixture of S and N-type (B-form), but not for N-type, which allows an easy way of identifying sugar puckering in DNA and RNA. An equivalent of the $^{1}H/^{1}H$ COSY experiment is the heternuclear $^{1}H/^{31}P$ correlation (HETCOR) that relates sugar H3'/H4'/H5'/5" to backbone P, and can complement information from homonuclear methods.
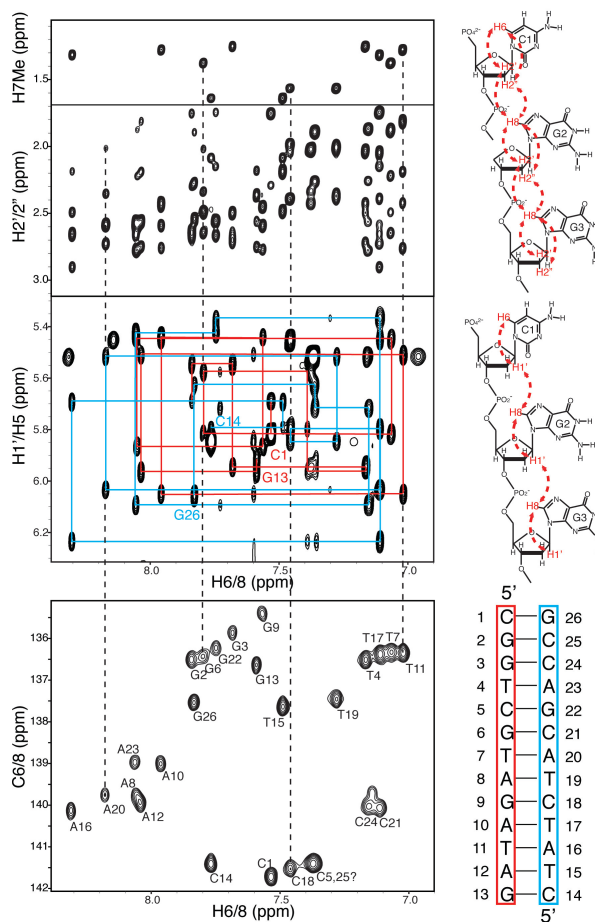
Figure 1.7: Base H6/H8 to sugar H1' NOE walk and connectivities to H2'.2"/CH3.

These assignment strategies can be implemented successfully for short DNA constructs, but have limited applications for large DNA molecules where spectral crowding becomes a significant issue. Even if analogous experiments with ^{13}C and/or ^{15}N-editing are used to resolve overlapped resonances in the 3rd and 4th dimension in ^{13}C/^{15}N labeled DNA constructs, the overlap in ^{1}H/^{13}C and ^{1}H/^{15}N 2D experiments can preclude conventional dynamic measurements. Moreover, unfavorably long overall rotational correlation times yield large resonance linewidths and present further challenge for dynamic characterization of large DNAs. It is the aim of this thesis to develop new approaches for partial DNA labeling that can solve the overlap problem, as well as devise suitable assignment strategies, so that dynamics in large DNA systems could be investigated at a residue-specific and atomic-level resolution.

1.2.2. Spin relaxation: theory and methods

1.2.2.1 Introduction

When placed in a strong external magnetic field, nuclear spins behave like tiny bar magnets that are partially polarized, aligning parallel (α state) and antiparallel (β state) to the magnetic field. As a result, nuclei with positive gyromagnetic ratio acquire a small net magnetization in the direction of the magnetic field at thermal equilibrium. During an NMR experiment, strong RF pulses are applied at a particular frequency causing spin-flips that convert the nuclear spin from parallel to antiparallel, higher energy configurations. When the RF pulse is switched off, the nuclear spin magnetization has a natural tendency to return to thermal equilibrium through interaction with the thermal environment. This process is referred to as nuclear spin relaxation.

Nuclear relaxation depends on the rate at which local magnetic fields fluctuate (in magnitude and direction) with time as the molecule reorients in solution with respect to the external magnetic field. These fluctuations are sensitive to overall molecular motion and internal motions that occur on timescales faster than the correlation time for global rotation, usually on the tenths of ns for DNA/RNA systems. Slower motions generally have smaller effect since the molecule reorients multiple times before the motion takes place. Thus, spin relaxation measurements are a valuable reporter of molecular dynamics

on ps-ns timescales and the most widely used approach for characterizing biomolecular dynamics by NMR spectroscopy.

The decay of the magnetization for an isolated spin ½ nuclei (^{1}H, ^{13}C, ^{15}N or ^{31}P) can be generally expressed by two relaxation constants – the longitudinal (spin-lattice) relaxation rate constant, R_1, and the transverse (spin-spin) relaxation rate constant, R_2 (Figure 1.8). The first describes the recovery of the magnetization component parallel to the magnetic field (or populations) back to the equilibrium Boltzmann distribution due to interaction with the complex magnetic field of the surroundings, the *lattice*. The second describes the dephasing of magnetization perpendicular to the magnetic field (single quantum coherences) due to interaction with neighboring nuclei that have different magnetic quantum states. For more complex systems of coupled spins, multiple spin operators such as zero and double quantum coherence have differing relaxation rate constants. In either case, the prevalent relaxation mechanisms are dipole-dipole (DD) and chemical shift anisotropy (CSA) interactions that occur between coupled nuclei. DD relaxation originates from variations in the through-space dipolar coupling between two nearby spins, while CSA relaxation arises from variations in the anisotropic chemical shielding of the nucleus, as the molecule rotates relative to the static magnetic field. In addition, R_2 relaxation can be affected by chemical exchange (R_{ex}) occurring on microsecond-to-millisecond timescales.

1.2.2.2 Spin relaxation probes in DNA

As discussed in the introduction, early dynamic relaxation studies of unlabeled DNA systems commonly probed the most abundant ^{1}H and ^{31}P nuclei. For an isolated N-H or C-H bond vector, ^{1}H spin relaxation is generally dominated by dipolar interaction with its directly bonded nitrogen or carbon that depends on the inverse sixth distance between the two nuclei. However, the extensive network of protons in close proximity (< 5 Å) to a targeted ^{1}H nucleus, in both nucleobase and sugar units, can contribute significantly to the observed relaxation rate constants. To properly analyze ^{1}H spin relaxation in a DNA molecule that tumbles anisotropically, one needs to have an accurate knowledge of the average inter-proton distances and the angular orientation of the proton-proton vectors relative to the principal axis of the diffusion tensor. The principal diffusion

axis could theoretically be approximated by the long helix axis for DNA constructs comprised of a single helix, but its direction could deviate substantially in DNA systems with statically/anisotropically bent or more complex multi-helical architectures. Accurate distance information, however, is extremely difficult to acquire for ^{1}H nuclei, since their position can deviate significantly from predicted values based on ideal B-form and is not well determined even in high-resolution X-ray structures. One way to avoid these challenges is perform relaxation measurements on DNA that is partially deuterated at sites in close proximity to the target protons; unfortunately, such deuterated nucleotides are very costly and not commercially available with ^{15}N/^{13}C-isotopic labeling. On the other hand, ^{31}P, despite its ~ 100 % natural abundance and its unique position to sense large-amplitude backbone motions, shows small chemical shift dispersion in solution NMR spectra and is not beneficial for site-specific dynamic studies.

The other two spin ½ nuclei in DNA, ^{13}C and ^{15}N, have < 1% natural abundance and lower magnetic sensitivity, making them historically non-ideal probes for dynamics. Developments in *in vitro* and *in vivo* methods for ^{13}C/^{15}N-labeled DNA synthesis over the last two decades, however, have made it possible to install ^{13}C and ^{15}N probes in DNA. In principle, the simplest nucleus to target in NMR spin relaxation measurements is the protonated imino nitrogen (N1 in G; N3 in T/U) on the heterocyclic base. Protonated imino ^{15}N nuclei can be treated as isolated two-spin (N-H) systems that enjoy the benefits of simple relaxation mechanisms dominated by N-H dipolar and ^{15}N CSA couplings. Other ^{15}N dipolar and scalar couplings with neighboring C, N or H are either negligible or can be easily suppressed by using selective RF pulses and continuous waves, or effectively eliminated in ^{15}N-labeled DNA samples. Considerable disadvantages of using imino ^{15}N as probes for dynamics are i) their scarcity in DNA/RNA helices, only one per base-pair (in G and T/U), which provide insufficient dynamic information for individual nucleobase and sugar backbone moieties, and ii) their unfavorable relaxation properties or even lack of NMR signal in non-canonical regions where they tend to exchange rapidly with solvent. This has made it necessary to rely on protonated carbons, whose value for studying nucleic acid dynamics was demonstrated by Hansen *et al*[37] as described above. ^{13}C can provide rich dynamic information about multiple base and sugar backbone sites, yet such measurements face unique challenges posed by the inherent

systems of coupled carbons comprising the building blocks of DNA and RNA. The common strategies to eliminate unwanted ^{13}C-^{13}C interactions are to use site-specific or fractional carbon labeling schemes, to carry out experiments at natural carbon abundance, or to devise appropriate NMR pulse sequences that can remove the effects of such couplings. Unfortunately, producing fractionally labeled nucleic acid samples still comes at a considerable cost, while natural abundance experiments are relatively insensitive and extremely time-consuming, making it intractable to study. However, advances in NMR experiments that deal with these issues have come forward.

1.2.2.3 Spin relaxation analysis

NMR relaxation for a given spin system can be treated theoretically by considering the transition probabilities between different energy levels or eigenstates, which depend on the fluctuations of the relaxation-inducing Hamiltonian, especially on those frequency components that match the transition frequency. The Hamiltonian that describes the spin system can be separated into two terms, spin operator functions and spatial functions, where the latter contains all the temporal fluctuations. The average spatial terms of the fluctuating field are correlation functions that rigorously describe the rotational motion of the nuclear interaction. How rapidly a local field B fluctuates can be expressed by the autocorrelation function of the field, defined as $C(t) = \langle B(t)B(t+\tau) \rangle$, where τ is an arbitrary time interval that separates any time point t from a later point $t + \tau$. The autocorrelation function tends to be large for small values of τ and approaches zero for large values of τ. The magnitude of the fluctuating field can be expressed as the mean square fluctuating field $\langle B^2(t) \rangle$. In the case of a single characteristic timescale, an isotropic rotational diffusion with no internal motions, one assumes a simple exponential behavior of the correlation function $C(t) = \langle B^2 \rangle e^{-|\tau|/\tau_m}$, where τ_m is the correlation time for the fluctuation, or overall diffusion.

The autocorrelation function defines another important parameter used in relaxation theory, the spectral density function $J(\omega)$:

$$J(\omega) = 2\int_0^\infty C(\tau)e^{-i\omega\tau}\,d\tau = 2\left\langle B^2 \right\rangle \mathbf{A}(\omega;0,\tau_m^{-1}) = 2\left\langle B^2 \right\rangle g(\omega) \tag{1.1}$$

where A is the real part of the complex Lorentzian (absorption line) and $g(\omega)$ is the normalized spectral density function. One can see that the spectral density function is the half-sided fourier transform of the autocorrelation function and can be explicitly written as:

$$J(\omega) = 2\left\langle B^2 \right\rangle \frac{\tau_m}{1+(\omega\tau_m)^2} \tag{1.2}$$

In practice, the idealized formulation of the spectral density above can be extended to reflect internal molecular motions (one or more) and account for anisotropic global diffusion, which is only briefly discussed here.

The most common approach to treat simultaneously overall and internal motions that occur on sufficiently different times, according to the model-free formalism developed by Lipari and Szabo[88], is to assume that they are independent components of the autocorrelation function:

$$f(t) = C(t) = C_0(t)C_I(t) \tag{1.3}$$

The correlation function for an isotropic overall diffusion can be rigorously expressed as:

$$C_0(t) = \frac{1}{5}e^{-t/\tau_m} \tag{1.4}$$

In general, the internal correlation function can be approximated by a series of exponential terms. For a single internal motion, it is given by:

$$C_I(t) = S^2 + (1-S^2)e^{-t/\tau_e} \tag{1.5}$$

where S is the generalized order parameter that defines the degree of motion and τ_e is an effective correlation time that depends on the local diffusive rates and the spatial properties of the motion (Figure 1.8). The squared order parameter, S^2, is a measure of the spatial restriction of the internal motion and ranges from zero to unity for unrestricted (isotropic) orientational distribution to a fixed orientation (lack of motion). These two parameters, S^2 and τ_e, are sufficient to specify the dynamic information about ps-ns internal motions and are model-independent. The fourier transform of the correlation function for isotropic overall motion is now given by:

$$g(\omega) = \frac{2}{5}\left(\frac{S^2}{1+(\omega\tau_m)^2} + \frac{(1-S^2)\tau}{1+(\omega\tau)^2}\right)$$ (1.6)

where

$$\tau^{-1} = \tau_m^{-1} + \tau_e^{-1}.$$ (1.7)

If the overall motion is anisotropic, the internal correlation and spectral density functions take a more complicated form. However, by fitting of experimental data to various motional models, Lipari and Szabo[88,89] showed that the formalism above is exact to explain experimental data as long as $(\omega\tau_e)^2 \ll 1$ and $\tau_e \ll \tau_m$, that is, the fast internal motions do not modulate overall tumbling and are effectively decoupled from it. There are numerous scenarios where these assumptions will not be correct, for example, in the case of long modular nucleic acid helices with different types of internal motions or helical fluctuations approaching the timescales of overall diffusion.

For the case when a single internal motion fails to describe NMR relaxation data, Clore and coworkers[176] proposed an extended model-free formalism including two internal correlation times, τ_f and τ_s:

$$C_I(t) = S^2 + (1-S_f^2)e^{-t/\tau_f} + (S_f^2 - S^2)e^{-t/\tau_s}$$ (1.8)

where $S^2 = S_f^2 S_s^2$, and S_f and S_s are order parameters (and correlation times) for the fast and slow internal dynamics, respectively. This relationship is valid under the assumption that τ_f and τ_s are separated by at least one order of magnitude.

Although proteins can generally be modeled as isotropic rotors, modular nucleic acids shorter than the persistence length ($\ll 150$ bp) do not tumble isotropically and can be more accurately represented by flexible rods with an axially symmetric diffusion tensor. The corresponding spectral density function developed by Spiess[177] can be, therefore, reformulated as:

$$g_2^\lambda(\omega) = \sum_{i=0}^{2} c_i^\lambda \left(\frac{S_f^2 S_s^2 \tau_i}{1+(\omega\tau_i)^2} + \frac{(1-S_f^2)\tau_{i,f}}{1+(\omega\tau_{i,f})^2} + \frac{(S_f^2 - S_f^2 S_s^2)\tau_{i,s}}{1+(\omega\tau_{i,s})^2}\right)$$ (1.9)

where $\tau_{i,n}^{-1} = \tau_i^{-1} + \tau_n^{-1}$ ($i = 0, 1, 2$ and $n = f$ or s), in which $\tau_0^{-1} = 6D_s - 2D_a$, $\tau_1^{-1} = 6D_s - D_a$, $\tau_2^{-1} = 6D_s + 2D_a$. $D_a = D_{zz} - 1/2(D_{xx} + D_{yy})$, $D_s = 1/3(D_{xx} + D_{yy} + D_{zz})$, and $\tau_m = (6D_s)^{-1}$, and D For an axially symmetric diffusion, the coefficients c_i^λ are

dependent on the orientation of the interaction frame relative to the long (unique) axis of the principle frame of the diffusion tensor, and can take into account the asymmetry of the CSA (i.e. in ^{13}C nuclei):

$$c_0^\lambda = \left(\frac{3\cos^2 \beta^\lambda - 1}{2} - \frac{\eta^\lambda}{2}\cos(2\alpha^\lambda)\sin^2 \beta^\lambda \right)^2$$

$$c_1^\lambda = \frac{1}{3}\left[\left(\frac{\eta^\lambda}{2}\cos(2\alpha^\lambda)\sin(2\beta^\lambda) + 3\cos\beta^\lambda \sin\beta^\lambda \right)^2 + \eta^\lambda \sin(2\alpha^\lambda)\sin\beta^\lambda \right]$$

$$c_2^\lambda = \frac{1}{3}\left[\left(\frac{\eta^\lambda}{4}\cos(2\alpha^\lambda)\left(3 + \cos(2\beta^\lambda)\right) - \frac{3}{2}\sin^2 \beta^\lambda \right)^2 + \left(\eta^\lambda \sin(2\alpha^\lambda)\sin\beta^\lambda \right)^2 \right] \qquad (1.10)$$

in which η^λ is the asymmetry of the interaction (λ = DD or CSA), where $\eta^{DD} = 0$ and $\eta^{CSA} = \left(\sigma_{xx} - \sigma_{yy} \right)/\sigma_{zz} \le 1$. The angles α^λ and β^λ are the polar angle defining the orientation of the axially symmetric dipolar ($\alpha^{DD} = 0$) and asymmetric CSA interaction tensors with respect to the principal axis of diffusion (D_{zz}).

For the dipolar and CSA interactions, the relaxation rate constants are linear combinations of the spectral density functions evaluated at different frequencies. The auto-relaxation rate constants R_1 and R_2 and cross-relaxation rate constant σ_{CX} used to compute the heteronuclear ^{13}C-^{1}H NOE, in the absence of chemical exchange, can be expressed as:

$$R_1 = R_1^{DD} + R_1^{CSA} = \frac{1}{10}D_{CX}^2\left(6g_2^{DD}(\omega_C + \omega_X) + g_2^{DD}(\omega_C - \omega_X) + 3g_2^{DD}(\omega_C) \right)$$
$$+ \frac{1}{10}C_C^2\left(3g_2^{CSA}(\omega_C) \right) \qquad (1.11)$$

$$R_2 = R_2^{DD} + R_2^{CSA} = \frac{1}{20}D_{CX}^2\left(\begin{array}{l} 4g_2^{DD}(0) + 3g_2^{DD}(\omega_C) + 6g_2^{DD}(\omega_X) + 6g_2^{DD}(\omega_C + \omega_X) \\ +g_2^{DD}(\omega_C - \omega_X) \end{array} \right)$$
$$+ \frac{1}{20}C_C^2\left(4g_2^{DD}(0) + 3g_2^{CSA}(\omega_C) \right) \qquad (1.12)$$

$$\sigma_{CX} = \frac{1}{10}D_{CX}^2\left(6g_2^{DD}(\omega_C + \omega_X) - g_2^{DD}(\omega_C - \omega_X) \right); \ NOE = 1 + \left(\frac{\gamma_X}{\gamma_C R_1} \right)\sigma_{CX} \qquad (1.13)$$

where X = H or C, $D_{CX} = \dfrac{\mu_0 h \gamma_C \gamma_H}{8\pi^2} \dfrac{1}{\langle r_{CX}^3 \rangle}$ is the dipolar coupling and $C_C = (\omega_C \sigma_{zz})$ is the

CSA constant, μ_0 is the permeability of free space, h is Plank's constant, γ_H and γ_C are the gyromagnetic ratios of ^{1}H and ^{13}C, r_{CX} the ^{1}H-^{1}H or ^{13}C-^{1}H bond vector where angular brackets indicate time average. The Lipari-Szabo model-free parameters are obtained from the experimental relaxation data typically by nonlinear least squares minimization of a χ^2 variable[178] or by using a Bayesian statistical approach[179,180] for various dynamics models.

This thesis takes advantage of ^{13}C as a ubiquitous probe for characterizing internal local and collective domain motions in base and sugar moieties in short and elongated DNA constructs. As mentioned before, ^{13}C relaxation, in analogy to ^{1}H relaxation, is complicated by extensive carbon-carbon dipolar and scalar interactions in uniformly labeled nucleotides. Moreover, data interpretation is challenging for short DNA/RNA constructs where the "decoupling approximation" under the model free formalism, the separation of internal and overall motions, can be violated. Hansen *et al*[37] addressed these challenges by developing an NMR pulse sequence that effectively suppresses undesired ^{13}C-^{13}C relaxation mechanisms in uniformly ^{13}C/^{15}N-labeled RNA via the use of selective excitation, selective off-resonance continuous waves (spin-lock), and gradient dephasing techniques. They show that the quantitative analysis of ^{13}C relaxation data to obtain order parameters and timescales of internal, both local and collective, dynamics can be simplified by using large elongated RNAs, where internal and overall motions are, to a
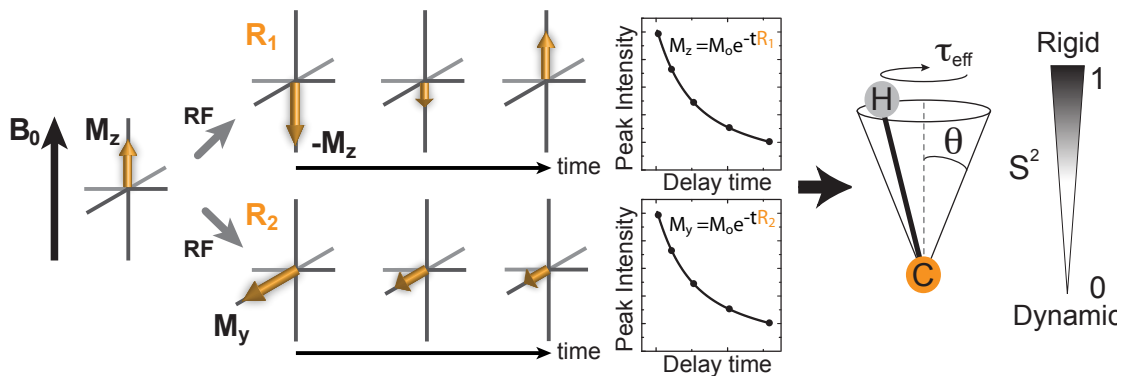
Figure 1.8: Charactering ps-ns motions using ^{13}C spin relaxation. Model free analysis of the relaxation data for longitudinal (R_1) and transverse (R_2) magnetization components yields an order parameter (S^2) describing the amplitude of motion and a constant (τ_{eff}) describing its timescale.

good approximation, decoupled. As a result, the prevalent relaxation contributions to nucleobase C2, C5, C6 and C8 and sugar C1' sites can be reduced to a combination of ^{13}C-^{1}H and ^{13}C-^{13}C dipolar, and ^{13}C CSA interactions. In addition, Hansen *et al.* were able to account for the non-collinearity of the C-H/C-C dipolar and CSA interactions by implementing the general spectra density function ($g_2(\omega)$) in their relaxation data analysis using the extended model free formalism with axially symmetric overall diffusion.

1.2.2.4 The generalized order parameter (S^2)

The generalized order parameter S^2 obtained from a model free analysis could be conceptualized in several different ways[181]. Generally, S^2 describes the equilibrium distribution of accessible orientations of the bond vector or tensor principal axis. If the distribution is for a small-amplitude axially symmetric motion, then $S^2 = 1 - 3\langle\theta^2\rangle$ (1.14), where θ is the polar angle. Alternatively, the motion could be modeled as a restricted diffusion in a cone with amplitude $S^2 = \left[\cos\theta(1 + \cos\theta)/2\right]^2$ (1.15) and semi-angle θ, proposed originally by Lipari and Szabo [89]. An even more complex interpretation is the Gaussian Axial Fluctuation (GAF), where the vector stochastically moves within a parabolic potential on the surface of a cone[182,183]:

$$S^2 = 1 - 3\sin^2\theta\left(\cos^2\theta(1 - e^{\left(-\sigma_f^2\right)} + 0.25\sin^2\theta(1 - e^{\left(-4\sigma_f^2\right)})\right)$$ (1.16)

where θ is the angle between the vector and principal axis for the motion and σ_f is the standard deviation in the fluctuation in the azimuthal angle. Changes in S^2 represent changes in the ps-ns flexibility and thus are related to the local conformational entropy[181].

The measurement of ^{13}C-^{1}H *NOE* in nucleic acids with current experimental schemes still carries significant errors for C5, C6 and C1' spin due to contributions from homonuclear ^{13}C-^{13}C relaxation[37]. An alternative method to model-free to examine ^{13}C relaxation rates (R_1 and R_2) in the absence of hetero-nuclear ^{13}C-^{1}H *NOE,* is to calculate a relative order parameter (S_{rel}^2) that is, in some sense, analogous to the generalized order

parameters S^2. S^2_{rel} values are computed from the value of $(2R_2 - R_1)$ normalized to the highest $(2R_2 - R_1)$ value for the same type of nucleus located in a well-structured helical region, assuming no chemical exchange contributions $(0 \le S^2_{rel} \le 1)$. To a good approximation, S^2_{rel} is proportional to $S^2 g_2(0)$, and is independent of internal motional timescales, the number of contributing relaxation mechanisms as long as all mechanisms experience similar motional amplitudes, and mostly insensitive to orientational differences of interaction tensor frames[37,184]. Thus, S^2_{rel} provides an estimate for the amplitude of internal motions occurring on ps-ns timescales and would be implemented here to compare the amplitude of motions in short ^{13}C/ ^{15}N-labeled DNA samples.

1.2.2.5 ^{13}C spin relaxation experiments

Experimental methods for measuring ^{13}C R_1 and R_2 relaxation are based on conventional HSQC pulse sequences or modified to use the TROSY technique for enhanced sensitivity in larger systems. The basic heteronuclear R_1 and R_2 relaxation NMR pulse sequence can be divided into five building blocks: (i) preparation of (^{13}C) magnetization, (ii) relaxation, (iii) (^{13}C) frequency labeling, (iv) mixing and (v) acquisition (Figure 1.9). The first period uses a series of pulses and delays called Insensitive Nucleus Enhanced Polarization Transfer (INEPT) or refocused INEPT element to initially excite ^{1}H, transfer the magnetization to the directly bonded ^{13}C, and prepare the proper ^{13}C magnetization, S_z for R_1 and $S_{x/y}$ for R_2 measurements. The magnetization is allowed to decay for a variable delay during the adjustable relaxation block, and incorporates ^{1}H decoupling to suppress unwanted relaxation mechanisms. In the following period, chemical shift is encoded in the indirect (^{13}C) dimension to generate the 2D spectrum, during which sizeable J_{CC} couplings can be removed by a constant-time evolution or application of selective pulses. The magnetization is returned to the ^{1}H dimension for observation during the mixing period via a refocused INEPT, a Spin State Selective Coherence Transfer (S$_3$CT) [185] or other schemes. Finally, the proton magnetization is recorded and frequency labeled during the acquisition period.

The ^{13}C relaxation block is different between R_1 and R_2 measurements. The inversion recovery technique[186] is used to record longitudinal relaxation. At the beginning

of the relaxation block, the ^{13}C magnetization is aligned along the z-axis and then allowed to relax for a variable delay, while a z-phase alternation is used to suppress unwanted artifacts. ^{1}H decoupling is accomplished by a train of 180° pulses at short intervals that suppresses ^{1}H-^{15}N dipolar cross-relaxation and ^{1}H-^{15}N DD/^{15}N CSA relaxation interference. The resonance intensity extracted from separate experiments with increasing time delays behaves monoexponentially with a rate constant equal to R_1.

The transverse relaxation rate R_2 can be measured using two types of experimental schemes, the Carr-Purcell-Meiboom-Gill (CPMG) or rotating-frame ($R_{1\rho}$) relaxation. In the CPMG pulse sequence, a spin-echo element is applied during the relaxation block that consists of a train of 180° pulses on ^{13}C/^{15}N channel used to refocus the transverse magnetization in the xy plane. This experiment is routinely used for the measurement of ^{15}N relaxation in protein amide and nucleic acid imino nitrogens, but does not effectively eliminate the extensive homonuclear interactions between nucleobase or sugar carbons, making it less suitable for measurement of ^{13}C relaxation. When applied to imino ^{15}N spins, softer (lower power) refocusing CPMG pulses are necessary in order to prevent J_{NN} coupling ($J_{NN} \sim 7$ Hz) evolution during relaxation and a particular phase cycle is applied
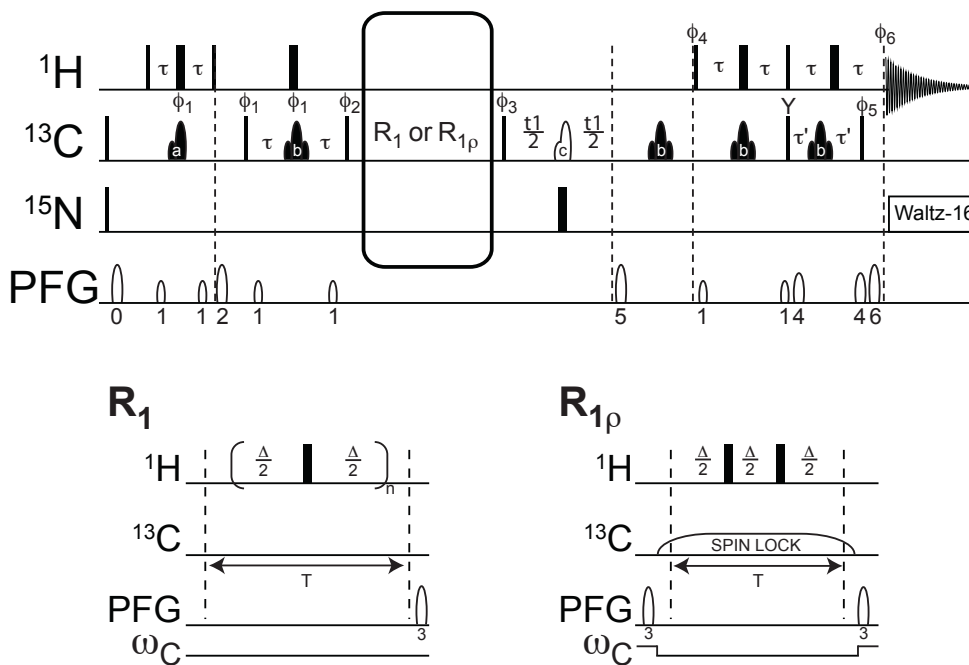
Figure 1.9: Sample pulse sequence for the measurement of off-resonance $R_{1\rho}$ or R_1 ^{13}C relaxation. (courtesy of Anette Casiano). Thin and thick rectangles represent hard 90° and 180° pulses respectively, *a* and *b* shapes represent shaped pulses and open shapes represent pulse field gradients (PFG).

to suppress off-resonance effects. A more suitable choice for nucleic acid carbons is the $R_{1\rho}$ relaxation scheme, which utilizes a continuous RF field during the relaxation block, a "spin-lock", to lock the magnetization at a particular tilt angle in the laboratory frame. The effective tilt angle is related to the spin-lock field strength/power (ω_1) and the spin-lock frequency offset (Ω) from the resonance of interest by the relationship, $\theta = \arctan(\omega_1 / \Omega)$. R_1 relaxation in the rotating frame has components both from longitudinal and transverse relaxation:

$$R_{1\rho} = R_1 \cos^2\theta + R_2 \sin^2\theta \tag{1.17}$$

$R_{1\rho}$ experiments at various relaxation delays produce a monoexponential intensity profile, from which $R_{1\rho}$ is determined and R_2 rates can be subsequently calculated. In a 2D experiment, adiabatic passages are used to properly align the magnetization of all resonances into their effective tilt angle[187] to avoid loss of signal. Accurate calibration of the spin-lock power can be accomplished by collecting intensities with and without adiabatic passages to estimate the relationship between the tilt angle and the actual power/offset experienced by each nucleus[66]. ^{1}H decoupling is accomplished by two proton 180_degree pulses that suppress the effects of ^{1}H-^{13}C dipole/^{13}C CSA cross‑correlated relaxation and J_{CH} evolution. Homonuclear Hartmann-Hahn (H-H) matching conditions between two carbons (1 and 2) can be avoided (<< 1%) by selecting values for the spin-lock power and offset that minimize the following function reflecting the efficiency of H-H interactions:

$$A_{HAHA} = \left(1 + \left(\frac{\omega_{eff,1} + \omega_{eff,2}}{J_{12}\left(1 + \cos\left(\theta_1 - \theta_2\right)\right)/2} \right) \right)^{-1} \tag{1.18}$$

where $\omega_{eff,j} = \left(\omega_1^2 + \Omega_j^2\right)^{1/2}$ is the effective power for carbon j = 1 or 2, J_{12} is the scalar coupling between the two carbons. In addition, ^{13}C $R_{1\rho}$ relaxation experiments typically collected at high spin-lock powers ($\omega_1 \sim 3.5 - 5.0$ kHz) in order to suppress contribution from chemical exchange occurring at the microsecond-to-millisecond timescales.

1.2.2.6 Chemical exchange (µs-ms)

In most general terms, chemical exchange refers to the process whereby a nuclear spin exchanges between different chemical environments or local magnetic fields, and can be inter- or intramolecular in nature. Biomolecules are characterized by complex energy landscapes and can dynamically exchange between different conformations on a variety of timescales that NMR spectroscopy is sensitive to. Flexibility on the ps-ns timescale includes local diffusive motions that occur via shallow energy barriers around the most energetically stable and populated "ground state", and can be probed by spin relaxation techniques (described above). Zooming out of the global energy minimum, one can find local minima that lie higher in free energy (by several kcal/mol) and are dynamically accessible from the ground state via transitions over much larger energy barriers. Such higher-energy conformational sub-states are referred to as "excited states", and are found in dynamic equilibrium with the ground state.

The equilibrium between a ground (A) and an excited (B) state can be represented by simple two-state kinetics:

$$A \underset{k_B}{\overset{k_A}{\rightleftharpoons}} B \tag{1.19}$$

where k_A and k_B are the forward and reverse rate constants. The kinetic rate laws for the system can be written in the matrix form:

$$\frac{d}{dt}\begin{pmatrix} [A](t) \\ [B](t) \end{pmatrix} = \begin{pmatrix} -k_A & k_B \\ k_A & -k_B \end{pmatrix}\begin{pmatrix} [A](t) \\ [B](t) \end{pmatrix} \tag{1.20}$$

The solution to these coupled differential equations is of the familiar form:

$$\begin{pmatrix} [A](t) \\ [B](t) \end{pmatrix} = \begin{pmatrix} \dfrac{k_A e^{-(k_A+k_B)t} + k_B}{k_A + k_B} & -\dfrac{k_B(e^{-(k_A+k_B)t} - 1)}{k_A + k_B} \\ -\dfrac{k_A(e^{-(k_A+k_B)t} - 1)}{k_A + k_B} & \dfrac{k_B e^{-(k_A+k_B)t} + k_A}{k_A + k_B} \end{pmatrix}\begin{pmatrix} [A]_0 \\ [B]_0 \end{pmatrix} \tag{1.21}$$

where the observed rate constant $k_{ex} = k_A + k_B$. This can be generalized to exchange between N different sites using the following relationship:

$$\frac{dA(t)}{dt} = KA(t) \tag{1.22}$$

$$A(t) = e^{Kt}A_0 = U^{-1}e^{Dt}UA_0$$

38

where $A(t)$ and A_0 are the Nx1 vectors that contain the time dependence of the initial concentrations of the exchanging species, U is unitary transformation matrix, D is a diagonal matrix and $D = URU^{-1}$, and K is an NxN kinetic rate matrix for which the following rules apply:

$$K_{ij} = k_{ji} \text{ for } i \neq j,\ K_{ii} = \sum_{\substack{j=1 \\ j \neq i}}^{N} k_{ij} \tag{1.23}$$

When applying this kinetics formalism to analysis of NMR signals, we assume the concentrations of a given species to be proportional to magnetization and integrate the evolution of magnetization into the Block-McConnell equations[48]. In the absence of radiofrequency fields, the longitudinal and transverse relaxation behave as:

$$\frac{dM_z(t)}{dt} = (-R + K)(M_z(t) - M_0)$$

$$\frac{dM^+(t)}{dt} = (i\Omega - R + K)M^+(t) \tag{1.24}$$

where M_0 is the equilibrium longitudinal magnetization, R represents the intrinsic relaxation rates with $R_{ij} = \delta_{ij}R_j$ and Ω represents the chemical shift offsets with $\Omega_{ij} = \delta_{ij}\Omega_j$ (δ is the Kronecker delta function).

Here, we will consider exchange processes that occur on the ms-µs timescales that tend to modulate the transverse relaxation rates and resonance lineshapes. This modulation manifests differently depending on the frequency of exchange, the populations and chemical shift difference between the exchanging states. Changes between different chemical environments often have sizeable effects on the NMR chemical shifts. If the two states A and B exchange at a rate constant ($k_{ex} = k_A + k_B$) that is comparable to the chemical shift difference between them ($\Delta\omega = \omega_B - \omega_A$) and B is appreciably populated, the stochastic fluctuation between the two frequencies of nuclei located on different DNA molecules during an NMR experiment will cause the transverse magnetization to "dephase" and relax with a faster rate constant than the intrinsic R_2. This leads to an observable exchange contribution, R_{ex}, to the apparent R_2 and additional line broadening of the observed NMR resonance signal. If $k_{ex} > \Delta\omega$, a broader resonance is

observed at a population weighted chemical shift, while the exchange contributions vanish for extremely fast $k_{ex} \gg \Delta\omega$; if $k_{ex} \ll \Delta\omega$, then separate resonances for A and B are observed and their linewidths and weight modulated by the exchange rate and populations. Even though the signal for B is usually not observed due to minute population (p_B), excessive line broadening, and/or relatively fast exchange rate, the transverse relaxation properties of the observable signal are modulated by the conformational exchange.

1.2.2.7 Relaxation dispersion experiments and analysis

If the exchange rate is on the ms-µs timescale, one can take advantage of spin relaxation techniques to monitor the chemical exchange[66,68]. When measuring R_2 spin relaxation by CPMG or $R_{1\rho}$ methods, the contribution from chemical exchange is intentionally suppressed by either i) tighter spacing by 180 degree pulses in CPMG experiments that refocus the transverse magnetization at a rate faster than the exchange rate or ii) higher spin-lock powers that minimize the dephasing of transverse magnetization. However, if the spacing between CPMG pulses or the spin-lock power is incrementally decreased in a series of measurements, the R_{ex} contribution can be revived and quantified from the so-called as relaxation dispersion profiles. The upper and lower bounds are determined by the frequency at which 180 degree pulses can be applied or the amplitude of continuous-wave spinlock without significant sample heating or NMR probe damaging effects and or the frequency necessary to suppresses scalar coupling evolution and off-resonance effects, etc. These specific requirements make CPMG experiments sensitive to slower millisecond motions, while traditional $R_{1\rho}$ experiments are more sensitive to faster, microsecond, motions.

Practically, CPMG experiments are not favorable for ^{13}C spins in nucleic acids that suffer from extensive ^{13}C-^{13}C interaction. As a result, $R_{1\rho}$ experiments are more applicable to nucleic acids for reasons discussed above. Traditionally for $R_{1\rho}$ measurements, the poor timescale sensitivity into millisecond motions has been determined by the high lower limit of usable powers (~ 1000 Hz) in 2D experiments that suppresses exchange processes only slower than ~ 300 µs. Below those powers,

conventional decoupling techniques for 2D experiments would fail to eliminate scalar coupling evolution and off-resonance effects, and would lead to oscillation superimposed onto the exponential decay. However, this decade has seen the implementation of off-resonance $R_{1\rho}$ experiments that can successfully use spin-lock fields as low as ~ 25 Hz without causing detrimental oscillatory effects on the relaxation profile[188]. These experiments have been used as a basis for development of suitable ^{13}C relaxation dispersion experiments by our laboratory to measure chemical exchange by our lab, which are partially outlined in this thesis.

Ideally, by mapping out the $R_{1\rho}$ relaxation dispersion as a function of spin-lock power and offset, one can extract the rate constant of exchange (k_{ex}), the populations (p) and lifetimes of the exchanging states, and the chemical shift difference between them ($\Delta\omega$). Theoretical expressions to analyze experimental $R_{1\rho}$ relaxation rates can be derived by several approaches that make particular assumptions depending on the rate and populations of exchange. For all exchange regimes, one can approximate $R_{1\rho}$ from the Block-McConnell equations (above) or Stochastic Liouville Equations, as described by Trot $et\ al.$[68,189].

When exchange is sufficiently fast or when the populations are skewed, the equations for $R_{1\rho}$ can be reduced to a much simpler expressions[68]. For examples, the expression for the fast exchange limit for a system of two exchanging states (A and B), is of the form:

$$R_{1\rho} = R_1 \cos^2\theta + (R_2 + R_{ex})\sin^2\theta = R_1 \cos^2\theta + \sin^2\theta\left(R_2 + \frac{\Phi_{ex}k_{ex}}{\omega_1^2 + k_{ex}^2}\right); \quad \Phi_{ex} = p_A p_B \Delta\omega_{AB}^2$$

(1.25)

where R_1 and R_2 are the longitudinal and transverse relaxation rates respectively, Ω is the resonance offset from the spinlock carrier, ω_1 is the spinlock strength; θ = arctan(ω_1/Ω_{ave}), $\Delta\omega_{AB} = \Omega_B - \Omega_A$, $\Omega_{ave} = p_A\Omega_A + p_B\Omega_B$, where p_A (p_B) is the major (minor) state fractional population ($p_A + p_B = 1$). Under this scenario, $k_{ex} \gg \Delta\omega_{AB}$ and one cannot discriminate the individual populations and chemical shift difference that are condensed in a single term Φ_{ex}. More often, chemical exchange occurs between a major (A) and

minor (B) species whose population is only a small fraction of A ($p_A \gg p_B$). Strictly, under that scenario and without assumptions about the exchange timescales, one can use the so-called asymmetric population expression that allows for quantification of all exchange parameters:

$$R_{1\rho} = R_1 \cos^2\theta + R_2 \sin^2\theta + \sin^2\theta \frac{p_A p_B \Delta\omega_{AB}^2 k_{ex}}{\Omega_B^2 + \omega_1^2 + k_{ex}^2} \tag{1.26}$$

where all parameters are as described above. By extracting k_{ex}, p_A, and p_B separately, one can quantify the forward and reverse rate constants given by $k_A = p_B k_{ex}$ and $k_B = p_A k_{ex}$ respectively ($k_{ex} = k_A + k_B$). In that case, the effect of chemical exchange is maximum when the spinlock is on resonance with the minor state (Ω_B). Finally, a general expression for a two-state equilibrium that gives accurate results uses the Laguerre approximation for polynomial root fitting, derived by Miloushev and Palmer[190]. Additional expressions have also been derived that improve the accuracy for the special cases of equal populations ($p_A = p_B$) or placing the spinlock at the average resonance frequency of the two states ($(\Omega_A + \Omega_B)/2$). In summary, these relaxation dispersion techniques provide the rare opportunity to be able to characterize the simultaneously the population, lifetime, and even conformation of low-populated (~ 1%) and transient (<1 ms) excited states in DNA.

1.3 References

1.	Woese, C.R. On the evolution of the genetic code. *Proc Natl Acad Sci U S A* **54**, 1546-52 (1965).
2.	Freeland, S.J. & Hurst, L.D. The genetic code is one in a million. *J Mol Evol* **47**, 238-48 (1998).
3.	Itzkovitz, S. & Alon, U. The genetic code is nearly optimal for allowing additional information within protein-coding sequences. *Genome Res* **17**, 405-12 (2007).
4.	Satchwell, S.C., Drew, H.R. & Travers, A.A. Sequence periodicities in chicken nucleosome core DNA. *J Mol Biol* **191**, 659-75 (1986).
5.	Segal, E. et al. A genomic code for nucleosome positioning. *Nature* **442**, 772-8 (2006).
6.	Cartegni, L., Chew, S.L. & Krainer, A.R. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* **3**, 285-98 (2002).

7. Katz, L. & Burge, C.B. Widespread selection for local RNA secondary structure in coding regions of bacterial genes. *Genome Res* **13**, 2042-51 (2003).

8. Segal, E. & Widom, J. Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr Opin Struct Biol* **19**, 65-71 (2009).

9. Wong, B., Chen, S., Kwon, J.A. & Rich, A. Characterization of Z-DNA as a nucleosome-boundary element in yeast Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **104**, 2229-34 (2007).

10. Widom, J. Role of DNA sequence in nucleosome stability and dynamics. *Q Rev Biophys* **34**, 269-324 (2001).

11. Otwinowski, Z. et al. Crystal structure of trp repressor/operator complex at atomic resolution. *Nature* **335**, 321-9 (1988).

12. Garvie, C.W. & Wolberger, C. Recognition of specific DNA sequences. *Mol Cell* **8**, 937-46 (2001).

13. Seeman, N.C., Rosenberg, J.M. & Rich, A. Sequence-specific recognition of double helical nucleic acids by proteins. *Proc Natl Acad Sci U S A* **73**, 804-8 (1976).

14. Beamer, L.J. & Pabo, C.O. Refined 1.8 A crystal structure of the lambda repressor-operator complex. *J Mol Biol* **227**, 177-96 (1992).

15. Pavletich, N.P. & Pabo, C.O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science* **252**, 809-17 (1991).

16. Chen, S., Vojtechovsky, J., Parkinson, G.N., Ebright, R.H. & Berman, H.M. Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: DNA binding specificity based on energetics of DNA kinking. *J Mol Biol* **314**, 63-74 (2001).

17. Kim, Y., Geiger, J.H., Hahn, S. & Sigler, P.B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512-20 (1993).

18. Rice, P.A., Yang, S., Mizuuchi, K. & Nash, H.A. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* **87**, 1295-306 (1996).

19. Joshi, R. et al. Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* **131**, 530-43 (2007).

20. Rohs, R. et al. The role of DNA shape in protein-DNA recognition. *Nature* **461**, 1248-53 (2009).

21. Kitayner, M. et al. Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs. *Nat Struct Mol Biol* **17**, 423-9 (2010).

22. Geanacopoulos, M. et al. GalR mutants defective in repressosome formation. *Genes Dev* **13**, 1251-62 (1999).

23. Tolstorukov, M.Y., Virnik, K.M., Adhya, S. & Zhurkin, V.B. A-tract clusters may facilitate DNA packaging in bacterial nucleoid. *Nucleic Acids Res* **33**, 3907-18 (2005).

24. Koudelka, G.B. & Carlson, P. DNA twisting and the effects of non-contacted bases on affinity of 434 operator for 434 repressor. *Nature* **355**, 89-91 (1992).

25. Olson, W.K., Gorin, A.A., Lu, X.J., Hock, L.M. & Zhurkin, V.B. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc Natl Acad Sci U S A* **95**, 11163-8 (1998).

26. Packer, M.J., Dauncey, M.P. & Hunter, C.A. Sequence-dependent DNA structure: dinucleotide conformational maps. *J Mol Biol* **295**, 71-83 (2000).

27. Zhang, Y., Xi, Z., Hegde, R.S., Shakked, Z. & Crothers, D.M. Predicting indirect readout effects in protein-DNA interactions. *Proc Natl Acad Sci U S A* **101**, 8337-41 (2004).

28. Svozil, D., Kalina, J., Omelka, M. & Schneider, B. DNA conformations and their sequence preferences. *Nucleic Acids Res* **36**, 3690-706 (2008).

29. Packer, M.J., Dauncey, M.P. & Hunter, C.A. Sequence-dependent DNA structure: tetranucleotide conformational maps. *J Mol Biol* **295**, 85-103 (2000).

30. Gardiner, E.J., Hunter, C.A., Packer, M.J., Palmer, D.S. & Willett, P. Sequence-dependent DNA structure: a database of octamer structural parameters. *J Mol Biol* **332**, 1025-35 (2003).

31. Haran, T.E. & Mohanty, U. The unique structure of A-tracts and intrinsic DNA bending. *Q Rev Biophys* **42**, 41-81 (2009).

32. Arauzo-Bravo, M.J. & Sarai, A. Indirect readout in drug-DNA recognition: role of sequence-dependent DNA conformation. *Nucleic Acids Res* **36**, 376-86 (2008).

33. Palmer, A.G., 3rd. NMR characterization of the dynamics of biomacromolecules. *Chem Rev* **104**, 3623-40 (2004).

34. Korzhnev, D.M. & Kay, L.E. Probing invisible, low-populated States of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding. *Acc Chem Res* **41**, 442-51 (2008).

35. Loria, J.P., Berlow, R.B. & Watt, E.D. Characterization of enzyme motions by solution NMR relaxation dispersion. *Acc Chem Res* **41**, 214-21 (2008).

36. Zhang, Q., Stelzer, A.C., Fisher, C.K. & Al-Hashimi, H.M. Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* **450**, 1263-7 (2007).

37. Hansen, A.L. & Al-Hashimi, H.M. Dynamics of large elongated RNA by NMR carbon relaxation. *J Am Chem Soc* **129**, 16072-82 (2007).

38. Hansen, A.L., Nikolova, E.N., Casiano-Negroni, A. & Al-Hashimi, H.M. Extending the range of microsecond-to-millisecond chemical exchange detected in labeled and unlabeled nucleic acids by selective carbon R(1rho) NMR spectroscopy. *J Am Chem Soc* **131**, 3818-9 (2009).

39. Rabi, II, Zacharias, J.R., Millman, S. & Kusch, P. A new method of measuring nuclear magnetic moment. *Physical review* **53**, 318-318 (1938).

40. Bloch, F., Hansen, W.W. & Packard, M. Nuclear induction. *Physical review* **69**, 127-127 (1946).

41. Purcell, E.M., Torrey, H.C. & Pound, R.V. Resonance absorption by nuclear magnetic moments in a solid. *Physical review* **69**, 37-38 (1946).

42. Knight, W.D. Nuclear magnetic resonance shift in metals. *Physical review* **76**, 1259-1260 (1949).

43. Proctor, W.G. & Yu, F.C. The dependence of a nuclear magnetic resonance frequency upon chemical compound. *Physical review* **77**, 717-717 (1950).

44. Dickinson, W.C. Dependence of the F-19 nuclear resonance position on chemical compound. *Physical review* **77**, 736-737 (1950).

45. Arnold, J.T., Dharmatti, S.S. & Packard, M.E. Chemical effects on nuclear induction signals from organic compounds. *Journal of Chemical Physics* **19**, 507-507 (1951).

46. Bloch, F. Line-narrowing by macroscopic motion. *Physical review* **94**, 496-497 (1954).

47. Gutowsky, H.S., McCall, D.W. & Slichter, C.P. Nuclear magnetic resonance multiplets in liquids. *Journal of Chemical Physics* **21**, 279-292 (1953).

48. McConnell, H.M. Reaction rates by nuclear magnetic resonance. *Journal of Chemical Physics* **28**, 430-431 (1958).

49. Redfield, A.G. Nuclear magnetic resonance saturation and rotary saturation in solids. *Physical review* **98**, 1787-1809 (1955).

50. Carr, H.Y. & Purcell, E.M. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical review* **94**, 630-638 (1954).

51. Andrew, E.R., Bradbury, A. & Eades, R.G. Nuclear magnetic resonance spectra from a crystal rotated at high speed. *Nature* **182**, 1659-1659 (1958).

52. Ernst, R.R. Senstivity enhancement in magnetic resonance .I. Analysis of method of time averaging. *Review of Scientific Instruments* **36**, 1689-& (1965).

53. Ernst, R.R. & Anderson, W.A. Application of fourier transform spectroscopy to magnetic resonance. *Review of Scientific Instruments* **37**, 93-+ (1966).

54. Jeener, J. *Ampere Summer School, Basko Polje* (1971).

55. Aue, W.P., Bartholdi, E. & Ernst, R.R. 2-dimensional spectroscopy - application to nuclear magnetic resonance. *Journal of Chemical Physics* **64**, 2229-2246 (1976).

56. Wagner, G. & Wuthrich, K. Sequential resonance assignments in protein 1H nuclear magnetic resonance spectra. Basic pancreatic trypsin inhibitor. *J Mol Biol* **155**, 347-66 (1982).

57. Braun, W., Wider, G., Lee, K.H. & Wuthrich, K. Conformation of glucagon in a lipid-water interphase by 1H nuclear magnetic resonance. *J Mol Biol* **169**, 921-48 (1983).

58. Williamson, M.P., Havel, T.F. & Wuthrich, K. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. *J Mol Biol* **182**, 295-315 (1985).

59. Torchia, D.A., Sparks, S.W. & Bax, A. NMR signal assignments of amide protons in the alpha-helical domains of staphylococcal nuclease. *Biochemistry* **27**, 5135-41 (1988).

60. Marion, D. et al. Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. *Biochemistry* **28**, 6150-6 (1989).

61. Kay, L.E., Clore, G.M., Bax, A. & Gronenborn, A.M. Four-dimensional heteronuclear triple-resonance NMR spectroscopy of interleukin-1 beta in solution. *Science* **249**, 411-4 (1990).

62. Clore, G.M. & Gronenborn, A.M. Two-, three-, and four-dimensional NMR methods for obtaining larger and more precise three-dimensional structures of proteins in solution. *Annu Rev Biophys Biophys Chem* **20**, 29-63 (1991).

63. Tolman, J.R., Flanagan, J.M., Kennedy, M.A. & Prestegard, J.H. Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc Natl Acad Sci U S A* **92**, 9279-83 (1995).

64. Tjandra, N., Wingfield, P., Stahl, S. & Bax, A. Anisotropic rotational diffusion of perdeuterated HIV protease from 15N NMR relaxation measurements at two magnetic fields. *J Biomol NMR* **8**, 273-84 (1996).

65. Tjandra, N. & Bax, A. Measurement of dipolar contributions to 1JCH splittings from magnetic-field dependence of J modulation in two-dimensional NMR spectra. *J Magn Reson* **124**, 512-5 (1997).

66. Palmer, A.G., 3rd, Kroenke, C.D. & Loria, J.P. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol* **339**, 204-38 (2001).

67. Prestegard, J.H., Bougault, C.M. & Kishore, A.I. Residual dipolar couplings in structure determination of biomolecules. *Chem Rev* **104**, 3519-40 (2004).

68. Palmer, A.G., 3rd & Massi, F. Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chem Rev* **106**, 1700-19 (2006).

69. Jarymowycz, V.A. & Stone, M.J. Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem Rev* **106**, 1624-71 (2006).

70. Furtig, B. et al. Time-resolved NMR studies of RNA folding. *Biopolymers* **86**, 360-83 (2007).

71. Pervushin, K., Riek, R., Wider, G. & Wuthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A* **94**, 12366-71 (1997).

72. Watson, J.D. & Crick, F.H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-8 (1953).

73. Rich, A. DNA comes in many forms. *Gene* **135**, 99-109 (1993).

74. Kearns, D.R. High-resolution nuclear magnetic resonance studies of double helical polynucleotides. *Annu Rev Biophys Bioeng* **6**, 477-523 (1977).

75. Crothers, D.M., Hilbers, C.W. & Shulman, R.G. Nuclear magnetic resonance study of hydrogen-bonded ring protons in Watson-Crick base pairs. *Proc Natl Acad Sci U S A* **70**, 2899-901 (1973).

76. Hilbers, C.W. & Patel, D.J. Proton nuclear magnetic resonance investigations of the nucleation and propagation reactions associated with the helix-coil transition of d-ApTpGpCpApT in H2O solution. *Biochemistry* **14**, 2656-60 (1975).

77. Patel, D.J. Proton and phosphorus NMR studies of d-CpG(pCpG)n duplexes in solution. Helix-coil transition and complex formation with actinomycin-D. *Biopolymers* **15**, 533-58 (1976).

78. Early, T.A., Kearns, D.R., Burd, J.F., Larson, J.E. & Wells, R.D. High resolution proton nuclear magnetic resonance investigation of the structural and dynamic properties of d(C15A15)-d(T15G15). *Biochemistry* **16**, 541-51 (1977).

79. Wells, R.D., Larson, J.E., Grant, R.C., Shortle, B.E. & Cantor, C.R. Physicochemical studies on polydeoxyribonucleotides containing defined repeating nucleotide sequences. *J Mol Biol* **54**, 465-97 (1970).

80. Wahl, P., Paoletti, J. & Le Pecq, J.B. Decay of fluorescence emission anisotropy of the ethidium bromide DNA complex evidence for an internal motion in DNA.

Proceedings of the National Academy of Sciences of the United States of America **65**, 417-421 (1970).

81. Klevan, L., Armitage, I.M. & Crothers, D.M. 31P NMR studies of the solution structure and dynamics of nucleosomes and DNA. *Nucleic Acids Res* **6**, 1607-16 (1979).

82. Bolton, P.H. & James, T.L. Molecular motions in RNA and DNA investigated by P-31 and C-13 NMR relaxation. *Journal of Physical Chemistry* **83**, 3359-3366 (1979).

83. Bolton, P.H. & James, T.L. Conformational mobility of deoxyribonucleic acid, transfer ribonucleic acid, and poly(adenylic acid) as monitored by carbon-13 nuclear magnetic resonance relaxation. *Biochemistry* **19**, 1388-92 (1980).

84. Bendel, P., Laub, O. & James, T.L. Molecular motions of supercoiled and circular DNA - a P-31 nuclear magnetic-resonance study. *Journal of the American Chemical Society* **104**, 6748-6754 (1982).

85. Bendel, P. & James, T.L. Structural and dynamic differences between supercoiled and linear DNA from proton NMR. *Proc Natl Acad Sci U S A* **80**, 3284-6 (1983).

86. Hogan, M.E. & Jardetzky, O. Internal motions in DNA. *Proc Natl Acad Sci U S A* **76**, 6341-5 (1979).

87. Hogan, M.E. & Jardetzky, O. Internal motions in deoxyribonucleic acid II. *Biochemistry* **19**, 3460-8 (1980).

88. Lipari, G. & Szabo, A. Nuclear magnetic resonance relaxation in nucleic acid fragments: models for internal motion. *Biochemistry* **20**, 6250-6 (1981).

89. Szabo, A. & Lipari, G. Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 1. Theory and range of validity. *Journal of the American Chemical Society* **104**, 4546-4559 (1982).

90. Levy, G.C., Hilliard, P.R., Jr., Levy, L.F., Rill, R.L. & Inners, R. Carbon 13 spin-lattice relaxation, linewidth, and nuclear Overhauser enhancement measurements of nucleosome length DNA. *J Biol Chem* **256**, 9986-9 (1981).

91. Rill, R.L., Hilliard, P.R., Jr. & Levy, G.C. Spontaneous ordering of DNA. Effects of intermolecular interactions on DNA motional dynamics monitored by 13C and 31P nuclear magnetic resonance spectroscopy. *J Biol Chem* **258**, 250-6 (1983).

92. Levy, G.C., Craik, D.J., Kumar, A. & London, R.E. A critical evaluation of models for complex molecular dynamics: application of NMR studies of double- and single-stranded DNA. *Biopolymers* **22**, 2703-26 (1983).

93. Arnott, S. & Selsing, E. Structures for the polynucleotide complexes poly(dA) with poly (dT) and poly(dT) with poly(dA) with poly (dT). *J Mol Biol* **88**, 509-21 (1974).

94. Arnott, S., Chandrasekaran, R., Hall, I.H. & Puigjaner, L.C. Heteronomous DNA. *Nucleic Acids Res* **11**, 4141-55 (1983).

95. Hogan, M., LeGrange, J. & Austin, B. Dependence of DNA helix flexibility on base composition. *Nature* **304**, 752-4 (1983).

96. Simpson, L. Isolation of maxicircle component of kinetoplast DNA from hemoflagellate protozoa. *Proc Natl Acad Sci U S A* **76**, 1585-8 (1979).

97. Marini, J.C. et al. Physical characterization of a kinetoplast DNA fragment with unusual properties. *J Biol Chem* **259**, 8974-9 (1984).

98. Marini, J.C., Levene, S.D., Crothers, D.M. & Englund, P.T. Bent helical structure in kinetoplast DNA. *Proc Natl Acad Sci U S A* **79**, 7664-7668 (1982).

99. Wu, H.M. & Crothers, D.M. The locus of sequence-directed and protein-induced DNA bending. *Nature* **308**, 509-13 (1984).

100. Hagerman, P.J. Sequence dependence of the curvature of DNA: a test of the phasing hypothesis. *Biochemistry* **24**, 7033-7 (1985).

101. Assa-Munt, N., Granot, J., Behling, R.W. & Kearns, D.R. 1H NMR relaxation studies of the hydrogen-bonded imino protons of poly(dA-dT). *Biochemistry* **23**, 944-55 (1984).

102. Assa-Munt, N. & Kearns, D.R. Poly(dA-dT) has a right-handed B conformation in solution: a two-dimensional NMR study. *Biochemistry* **23**, 791-6 (1984).

103. Behling, R.W. & Kearns, D.R. 1H two-dimensional nuclear Overhauser effect and relaxation studies of poly(dA).poly(dT). *Biochemistry* **25**, 3335-46 (1986).

104. Mirau, P.A., Behling, R.W. & Kearns, D.R. Internal motions in B- and Z-form poly(dG-dC).poly(dG-dC): 1H NMR relaxation studies. *Biochemistry* **24**, 6200-11 (1985).

105. Robinson, B.H. et al. Analysis of double-helix motions with spin-labeled probes: binding geometry and the limit of torsional elasticity. *J Mol Biol* **139**, 19-44 (1980).

106. Bobst, A.M., Kao, S.C., Toppin, R.C., Ireland, J.C. & Thomas, I.E. Dipsticking the major groove of DNA with enzymatically incorporated spin-labeled deoxyuridines by electron spin resonance spectroscopy. *J Mol Biol* **173**, 63-74 (1984).

107. Kao, S.C. & Bobst, A.M. Local base dynamics and local structural features in RNA and DNA duplexes. *Biochemistry* **24**, 5465-9 (1985).

108. Strobel, O.K., Keyes, R.S. & Bobst, A.M. Base dynamics of local Z-DNA conformations as detected by electron paramagnetic resonance with spin-labeled deoxycytidine analogues. *Biochemistry* **29**, 8522-8 (1990).

109. Barkley, M.D. & Zimm, B.H. THEORY OF TWISTING AND BENDING OF CHAIN MACROMOLECULES - ANALYSIS OF THE FLUORESCENCE DEPOLARIZATION OF DNA. *Journal of Chemical Physics* **70**, 2991-3007 (1979).

110. Schurr, J.M. Rotational diffusion of deformable macromolecules with mean local cylindrical symmetry. *Chemical Physics* **84**, 71-96 (1984).

111. Alam, T.M., Orban, J. & Drobny, G.P. Deuterium NMR investigation of backbone dynamics in the synthetic oligonucleotide [d(CGCGAATTCGCG)]2. *Biochemistry* **30**, 9229-37 (1991).

112. Shindo, H., Wooten, J.B., Pheiffer, B.H. & Zimmerman, S.B. Nonuniform backbone conformation of deoxyribonucleic acid indicated by phosphorus-31 nuclear magnetic resonance chemical shift anisotropy. *Biochemistry* **19**, 518-26 (1980).

113. Opella, S.J., Wise, W.B. & DiVerdi, J.A. Deoxyribonucleic acid dynamics from phosphorus-31 nuclear magnetic resonance. *Biochemistry* **20**, 284-90 (1981).

114. DiVerdi, J.A. & Opella, S.J. Dynamics of B-DNA in the solid state. *J Mol Biol* **149**, 307-11 (1981).

115. Bendel, P., Murphy-Boesch, J. & James, T.L. Deuterium NMR in the solid-state and in solution of the molecular motion of the bases in poly(I) and poly(I) . poly(C). *Biochim Biophys Acta* **759**, 205-13 (1983).

116. Mai, M.T., Wemmer, D.E. & Jardetzky, O. Effects of hydration on the dynamics of deoxyribonucleic acids. *Journal of the American Chemical Society* **105**, 7149-7152 (1983).

117. Brandes, R., Vold, R.R., Vold, R.L. & Kearns, D.R. Effects of hydration on purine motion in solid DNA. *Biochemistry* **25**, 7744-51 (1986).

118. Brandes, R. & Ehrenberg, A. Kinetics of the proton-deuteron exchange at position H8 of adenine and guanine in DNA. *Nucleic Acids Res* **14**, 9491-508 (1986).

119. Brandes, R., Vold, R.R., Kearns, D.R. & Rupprecht, A. Static disorder and librational motions of the purine bases in films of oriented Li-DNA. *J Mol Biol* **202**, 321-32 (1988).

120. Brandes, R., Kearns, D.R. & Rupprecht, A. A 2H-NMR study of the DNA hydration water in solid Li-DNA assembles. *Biopolymers* **27**, 717-32 (1988).

121. Brandes, R., Vold, R.R., Kearns, D.R. & Rupprecht, A. Deuterium relaxation and internal motion in solid Li-DNA. *Biochemistry* **29**, 1717-21 (1990).

122. Early, T.A., Kearns, D.R., Hillen, W. & Wells, R.D. A 300-MHz proton nuclear magnetic resonance investigation of deoxyribonucleic acid restriction fragments: dynamic properties. *Biochemistry* **20**, 3764-9 (1981).

123. Early, T.A., Kearns, D.R., Hillen, W. & Wells, R.D. A 300- and 600-MHz proton nuclear magnetic resonance investigation of a 12 base pair deoxyribonucleic acid restriction fragment: relaxation behavior of the low-field resonances in water. *Biochemistry* **20**, 3756-64 (1981).

124. Feigon, J., Denny, W.A., Leupin, W. & Kearns, D.R. Proton nuclear magnetic resonance investigation of the conformation and dynamics in the synthetic deoxyribonucleic acid decamers d(ATATCGATAT) and d(ATATGCATAT). *Biochemistry* **22**, 5930-42 (1983).

125. Williamson, J.R. & Boxer, S.G. Synthesis of a thymidine phosphoramidite labelled with 13C at C6: relaxation studies of the loop region in a 13C labelled DNA hairpin. *Nucleic Acids Res* **16**, 1529-40 (1988).

126. Williamson, J.R. & Boxer, S.G. Multinuclear NMR studies of DNA hairpins. 1. Structure and dynamics of d(CGCGTTGTTCGCG). *Biochemistry* **28**, 2819-31 (1989).

127. Borer, P.N. et al. 13C-NMR relaxation in three DNA oligonucleotide duplexes: model-free analysis of internal and overall motion. *Biochemistry* **33**, 2441-50 (1994).

128. Eimer, W., Williamson, J.R., Boxer, S.G. & Pecora, R. Characterization of the overall and internal dynamics of short oligonucleotides by depolarized dynamic light scattering and NMR relaxation measurements. *Biochemistry* **29**, 799-811 (1990).

129. Zimmer, D.P. & Crothers, D.M. NMR of enzymatically synthesized uniformly 13C15N-labeled DNA oligonucleotides. *Proc Natl Acad Sci U S A* **92**, 3091-5 (1995).

130. Chen, X. et al. A PCR-based method for uniform 13C/15N labeling of long DNA oligomers. *FEBS Lett* **436**, 372-6 (1998).

131. Kojima, C., Ono, A., Kainosho, M. & James, T.L. DNA duplex dynamics: NMR relaxation studies of a decamer with uniformly 13C-labeled purine nucleotides. *J Magn Reson* **135**, 310-33 (1998).

132. Louis, J.M., Martin, R.G., Clore, G.M. & Gronenborn, A.M. Preparation of uniformly isotope-labeled DNA oligonucleotides for NMR spectroscopy. *J Biol Chem* **273**, 2374-8 (1998).

133. Masse, J.E., Bortmann, P., Dieckmann, T. & Feigon, J. Simple, efficient protocol for enzymatic synthesis of uniformly 13C, 15N-labeled DNA for heteronuclear NMR studies. *Nucleic Acids Res* **26**, 2618-24 (1998).

134. Yamazaki, T., Muhandiram, R. & Kay, L.E. NMR Experiments for the Measurement of Carbon Relaxation Properties in Highly Enriched, Uniformly 13C,15N-Labeled Proteins: Application to 13C.alpha. Carbons. *Journal of the American Chemical Society* **116**, 8266-8278 (1994).

135. Korzhnev, D.M., Skrynnikov, N.R., Millet, O., Torchia, D.A. & Kay, L.E. An NMR experiment for the accurate measurement of heteronuclear spin-lock relaxation rates. *J Am Chem Soc* **124**, 10743-53 (2002).

136. Duchardt, E., Nilsson, L. & Schleucher, J. Cytosine ribose flexibility in DNA: a combined NMR 13C spin relaxation and molecular dynamics simulation study. *Nucleic Acids Res* **36**, 4211-9 (2008).

137. Pederson, K., Meints, G.A., Shajani, Z., Miller, P.A. & Drobny, G.P. Backbone dynamics in the DNA HhaI protein binding site. *J Am Chem Soc* **130**, 9072-9 (2008).

138. Echodu, D. et al. Furanose dynamics in the HhaI methyltransferase target DNA studied by solution and solid-state NMR relaxation. *J Phys Chem B* **112**, 13934-44 (2008).

139. Shajani, Z. & Varani, G. 13C relaxation studies of the DNA target sequence for hhai methyltransferase reveal unique motional properties. *Biochemistry* **47**, 7617-25 (2008).

140. Spielmann, H.P. Dynamics of a bis-intercalator DNA complex by 1H-detected natural abundance 13C NMR spectroscopy. *Biochemistry* **37**, 16863-76 (1998).

141. Spielmann, H.P. Dynamics in psoralen-damaged DNA by 1H-detected natural abundance 13C NMR spectroscopy. *Biochemistry* **37**, 5426-38 (1998).

142. Isaacs, R.J. & Spielmann, H.P. Relationship of DNA structure to internal dynamics: correlation of helical parameters from NOE-based NMR solution structures of d(GCGTACGC)(2) and d(CGCTAGCG)(2) with (13)C order parameters implies conformational coupling in dinucleotide units. *J Mol Biol* **307**, 525-40 (2001).

143. Flatters, D. & Lavery, R. Sequence-dependent dynamics of TATA-Box binding sites. *Biophys J* **75**, 372-81 (1998).

144. Paquet, F., Gaudin, F. & Lancelot, G. Selectively 13C-enriched DNA: evidence from 13C1' relaxation rate measurements of an internal dynamics sequence effect in the lac operator. *J Biomol NMR* **8**, 252-60 (1996).

145. Isaacs, R.J. & Spielmann, H.P. NMR evidence for mechanical coupling of phosphate B(I)-B(II) transitions with deoxyribose conformational exchange in DNA. *J Mol Biol* **311**, 149-60 (2001).

146.   Meints, G.A., Miller, P.A., Pederson, K., Shajani, Z. & Drobny, G. Solid-state nuclear magnetic resonance spectroscopy studies of furanose ring dynamics in the DNA Hhal binding site. *Journal of the American Chemical Society* **130**, 7305-7314 (2008).

147.   Pederson, K., Meints, G.A., Shajani, Z., Miller, P.A. & Drobny, G.P. Backbone dynamics in the DNA Hhal protein binding site. *Journal of the American Chemical Society* **130**, 9072-9079 (2008).

148.   Gueron, M., Kochoyan, M. & Leroy, J.L. A single mode of DNA base-pair opening drives imino proton exchange. *Nature* **328**, 89-92 (1987).

149.   Moe, J.G. & Russu, I.M. Proton exchange and base-pair opening kinetics in 5'-d(CGCGAATTCGCG)-3' and related dodecamers. *Nucleic Acids Res* **18**, 821-7 (1990).

150.   Moe, J.G. & Russu, I.M. Kinetics and energetics of base-pair opening in 5'-d(CGCGAATTCGCG)-3' and a substituted dodecamer containing G.T mismatches. *Biochemistry* **31**, 8421-8 (1992).

151.   Folta-Stogniew, E. & Russu, I.M. Sequence dependence of base-pair opening in a DNA dodecamer containing the CACA/GTGT sequence motif. *Biochemistry* **33**, 11016-24 (1994).

152.   Chen, C. & Russu, I.M. Sequence-dependence of the energetics of opening of AT basepairs in DNA. *Biophys J* **87**, 2545-51 (2004).

153.   Coman, D. & Russu, I.M. Site-resolved stabilization of a DNA triple helix by magnesium ions. *Nucleic Acids Res* **32**, 878-83 (2004).

154.   Coman, D. & Russu, I.M. Base pair opening in three DNA-unwinding elements. *J Biol Chem* **280**, 20216-21 (2005).

155.   Coman, D. & Russu, I.M. A nuclear magnetic resonance investigation of the energetics of basepair opening pathways in DNA. *Biophys J* **89**, 3285-92 (2005).

156.   Every, A.E. & Russu, I.M. Probing the role of hydrogen bonds in the stability of base pairs in double-helical DNA. *Biopolymers* **87**, 165-73 (2007).

157.   Every, A.E. & Russu, I.M. Influence of magnesium ions on spontaneous opening of DNA base pairs. *J Phys Chem B* **112**, 7689-95 (2008).

158.   Kochoyan, M., Leroy, J.L. & Gueron, M. Proton exchange and base-pair lifetimes in a deoxy-duplex containing a purine-pyrimidine step and in the duplex of inverse sequence. *J Mol Biol* **196**, 599-609 (1987).

159.   Kochoyan, M., Lancelot, G. & Leroy, J.L. Study of structure, base-pair opening kinetics and proton exchange mechanism of the d-(AATTGCAATT) self-complementary oligodeoxynucleotide in solution. *Nucleic Acids Res* **16**, 7685-702 (1988).

160.   Leroy, J.L., Charretier, E., Kochoyan, M. & Gueron, M. Evidence from base-pair kinetics for two types of adenine tract structures in solution: their relation to DNA curvature. *Biochemistry* **27**, 8894-8 (1988).

161.   Leroy, J.L., Kochoyan, M., Huynh-Dinh, T. & Gueron, M. Characterization of base-pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton. *J Mol Biol* **200**, 223-38 (1988).

162.   Kochoyan, M., Leroy, J.L. & Gueron, M. Processes of base-pair opening and proton exchange in Z-DNA. *Biochemistry* **29**, 4799-805 (1990).

163. Nonin, S., Leroy, J.L. & Gueron, M. Terminal base pairs of oligodeoxynucleotides: imino proton exchange and fraying. *Biochemistry* **34**, 10652-9 (1995).

164. Nonin, S., Leroy, J.L. & Gueron, M. Acid-induced exchange of the imino proton in G.C pairs. *Nucleic Acids Res* **24**, 586-95 (1996).

165. Varnai, P., Canalia, M. & Leroy, J.L. Opening mechanism of G.T/U pairs in DNA and RNA duplexes: a combined study of imino proton exchange and molecular dynamics simulation. *J Am Chem Soc* **126**, 14659-67 (2004).

166. Warmlander, S., Sen, A. & Leijon, M. Imino proton exchange in DNA catalyzed by ammonia and trimethylamine: evidence for a secondary long-lived open state of the base pair. *Biochemistry* **39**, 607-15 (2000).

167. Dornberger, U., Leijon, M. & Fritzsche, H. High base pair opening rates in tracts of GC base pairs. *J Biol Chem* **274**, 6957-62 (1999).

168. Parker, J.B. et al. Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature* **449**, 433-7 (2007).

169. Lefevre, J.F., Lane, A.N. & Jardetzky, O. A temperature dependent transition in the Pribnow box of the trp promoter. *FEBS Lett* **190**, 37-40 (1985).

170. Kennedy, M.A., Nuutero, S.T., Davis, J.T., Drobny, G.P. & Reid, B.R. Mobility at the TpA cleavage site in the T3A3-containing AhaIII and PmeI restriction sequences. *Biochemistry* **32**, 8022-35 (1993).

171. Kojima, C., Ulyanov, N.B., Kainosho, M. & James, T.L. Slow motion in the CAA*TTG sequence of a DNA decamer duplex studied by NMR. *Biochemistry* **40**, 7239-46 (2001).

172. Lane, A.N. & Peck, B. Conformational flexibility in DNA duplexes containing single G.G mismatches. *Eur J Biochem* **230**, 1073-87 (1995).

173. Palmer, A.G., 3rd. Probing molecular motion by NMR. *Curr Opin Struct Biol* **7**, 732-7 (1997).

174. Wijmenga, S.S. & van Buuren, B.N.M. The use of NMR methods for conformational studies of nucleic acids. *Progress in Nuclear Magnetic Resonance Spectroscopy* **32**, 287-387 (1998).

175. Cromsigt, J., van Buuren, B., Schleucher, J. & Wijmenga, S. Resonance assignment and structure determination for RNA. in *Nuclear Magnetic Resonance of Biologica Macromolecules, Pt A*, Vol. 338 371-399 (Academic Press Inc, San Diego, 2001).

176. Clore, G.M. et al. Deviation from the simple 2-parameter model-free approach to the interpretation of N-15 nuclear magnetic-relaxation of proteins. *Journal of the American Chemical Society* **112**, 4989-4991 (1990).

177. Hentschel, R., Schlitter, J., Sillescu, H. & Spiess, H.W. Orientational distributions in partially ordered solids as determined from NMR and ESR line-shapes. *Journal of Chemical Physics* **68**, 56-66 (1978).

178. Palmer, A.G., Rance, M. & Wright, P.E. Intramolecular Motions of a Zinc Finger DNA-Binding Domain from Xfin Characterized by Proton-Detected Natural Abundance C-12 Heteronuclear Nmr-Spectroscopy. *Journal of the American Chemical Society* **113**, 4371-4380 (1991).

179. McMahon, M.T. & Oldfield, E. Determination of order parameters and correlation times in proteins: a comparison between Bayesian, Monte Carlo and simple graphical methods. *J Biomol NMR* **13**, 133-7 (1999).
180. Andrec, M., Montelione, G.T. & Levy, R.M. Estimation of dynamic parameters from NMR relaxation data using the Lipari-Szabo model-free approach and Bayesian statistical methods. *J Magn Reson* **139**, 408-21 (1999).
181. Palmer, A.G., 3rd. Nmr probes of molecular dynamics: overview and comparison with other techniques. *Annu Rev Biophys Biomol Struct* **30**, 129-55 (2001).
182. Bruschweiler, R. & Wright, P.E. Nmr Order Parameters of Biomolecules - a New Analytical Representation and Application to the Gaussian Axial Fluctuation Model. *Journal of the American Chemical Society* **116**, 8426-8427 (1994).
183. Bremi, T., Bruschweiler, R. & Ernst, R.R. A protocol for the interpretation of side-chain dynamics based on NMR relaxation: Application to phenylalanines in antamanide. *Journal of the American Chemical Society* **119**, 4272-4284 (1997).
184. Tjandra, N., Cowburn, D. & Fushman, D. An approach to direct determination of protein dynamics from N-15 NMR relaxation at multiple fields, independent of variable N-15 chemical shift anisotropy and chemical exchange contributions. *Journal of the American Chemical Society* **121**, 8577-8582 (1999).
185. Sorensen, M.D., Meissner, A. & Sorensen, O.W. Spin-state-selective coherence transfer via intermediate states of two-spin coherence in IS spin systems: Application to E.COSY-type measurement of J coupling constants. *Journal of Biomolecular Nmr* **10**, 181-186 (1997).
186. Vold, R.L., Waugh, J.S., Klein, M.P. & Phelps, D.E. Measurement of spin relaxation in complex systems. *Journal of Chemical Physics* **48**, 3831-& (1968).
187. Mulder, F.A.A., de Graaf, R.A., Kaptein, R. & Boelens, R. An Off-resonance Rotating Frame Relaxation Experiment for the Investigation of Macromolecular Dynamics Using Adiabatic Rotations. *J Magn Reson* **131**, 351-7 (1998).
188. Korzhnev, D.M., Orekhov, V.Y. & Kay, L.E. Off-resonance R(1rho) NMR studies of exchange dynamics in proteins with low spin-lock fields: an application to a Fyn SH3 domain. *J Am Chem Soc* **127**, 713-21 (2005).
189. Trott, O. & Palmer, A.G., 3rd. R1rho relaxation outside of the fast-exchange limit. *J Magn Reson* **154**, 157-60 (2002).
190. Miloushev, V.Z. & Palmer, A.G., 3rd. R(1rho) relaxation for two-site chemical exchange: general approximations and some exact solutions. *J Magn Reson* **177**, 221-7 (2005).

Chapter 2

Probing pico- to nanosecond dynamics in short and elongated DNA helices and their modulation by variable size A-tracts

2.1 Introduction

Sequence-specific DNA flexibility plays essential roles in a variety of cellular processes that are key for gene packaging, expression and regulation[1-3]. For example, intrinsic sequence-specific DNA flexibility is believed to play an important function in directing adaptive changes in DNA conformation that occur upon protein and ligand recognition[1,4,5]. Local deformability at the dinucleotide level and unusual elements such as adenine tracts have been implicated in nucleosome positioning and DNA accessibility by the transcriptional machinery, potentially providing a new layer of genetic regulation[6-9]. Sequence-specific modifications such as damage can also modulate DNA flexibility, which is believed to actively participate in the efficient recruitment of repair enzymes[10-12].

Solution NMR techniques based on measurements of spin relaxation data[4,13-15] and RDCs[16,17] provide a unique opportunity to probe DNA flexibility at the site-specific level over ps to ms timescales[18-21]. More recent examples of ^{13}C relaxation studies have revealed multi-scale dynamics in a methyltransferase enzyme DNA target sequence[22-24] and high-amplitude motions in the deoxyribose moieties of cytosine residues[25,26], while RDCs have allowed the detailed construction of conformational ensembles of the Dickerson dodecamer[26,27]. Despite advances in the application of spin relaxation and RDC techniques in studies of DNA dynamics illustrated above and in Chapter 1, methodological challenges abound. As shown for RNA, extended nucleic acids can undergo large-scale fluctuations in global structure, such as bending, that can result in

coupled changes in overall motions[28-32]. This, in turn, can lead to a breakdown in the "decoupling approximation", which is at the crux of formalisms currently used to quantitatively interpret NMR spin relaxation data and RDCs[33-35]. In addition, motions occurring at ns timescales remain difficult to address by solution NMR because they fall outside the detection limit of both spin relaxation and relaxation dispersion techniques. While RDCs are in principle sensitive to such fluctuations, their broad timescale sensitivity makes it complicated to tease out individual contributions from motions occurring at specific rates.

Our laboratory has developed an NMR-invisible domain elongation approach for decoupling internal from overall motions in RNA[30,32] that not only serves to resolve internal from global motions, but also predefines the overall diffusion and alignment tensor to be axially symmetric with the principal axis oriented along the elongated helix, thus further simplifying analysis of spin relaxation[30,32] and RDC data[31]. By reducing the rate of overall tumbling, the elongation extends the timescale sensitivity of relaxation data deeper into the ns regime[30]. When combined with measurements of ^{15}N and ^{13}C spin relaxation, and RDCs, this approach allowed for the characterization of a network of local and collective motions in RNA spanning ps to ms timescales that play key roles in adaptive recognition[30-32,36].

Here, we extend the elongation methodology and tailor it towards the atomic characterization of dynamical features in elongated DNA (E-DNA) and their modulation by A-tracts[i] (see Chapter 1). Globally, A-tracts cause microscopic bending when periodically phased, while, locally, A-tracts tend to assume a non-canonical, so-called B'-DNA conformation with distinct structural features such as minor groove compression and high base pair propeller twist (see Haran et al.[37]). Despite decades of investigation, it remains unclear whether A-tracts induce static or dynamic bends. Although there are conflicting views, most studies suggest that A-tract sequences themselves are "straight" and locally stiff but that A-tract junctions with neighboring GC-rich B-DNA sequences are flexible, and possibly, the source of deformation and inter-helical bending, reviewed in Haran et al.[37,38] and Beveridge et al.[37,38]. This highlights one of the biggest challenges

---

[i] Nikolova, E.N. & Al-Hashimi, H.M. Preparation, resonance assignment, and preliminary dynamics characterization of residue specific 13C/15N-labeled elongated DNA for the study of sequence-directed dynamics by NMR. *J Biomol NMR* **45**, 9-16 (2009).

in the field of DNA dynamics, which is to experimentally distinguish "static" versus "dynamic" deformations in a B-form duplex structure. Here, we investigate rapid ns bending and torsional motions in A-tract containing E-DNA using ^{15}N and ^{13}C spin relaxation combined with qualitative intensity analysis. Using the same relaxation techniques and additional tricks to resolve global from internal motions in short DNA duplexes, we further tap into the local ps-ns dynamics of A-tract interior and junctions and their dependence on A-tract length to look for rare sequence-dependent dynamic features.

2.2 Materials and Methods

2.2.1 Synthesis and preparation of $^{13}C/^{15}N$-labeled and unlabeled DNA

All unlabeled DNA oligonucleotides were purchased from Integrated DNA Technologies, Inc. (Coralville, IA). E-DNA was synthesized using the Polymerase chain reaction (PCR) method. Site-selectively labeled E-DNA duplexes were prepared in a series of enzymatic reactions involving the target DNA, a complementary primer pair, *Taq* (or *Vent*) DNA polymerase (NEB, Inc.), and one type of uniformly $^{13}C/^{15}N$-labeled dNTPs (Isotec, Sigma-Aldrich). Reaction conditions such as buffer and primer annealing temperatures were optimized prior to each large-scale synthesis. A sample 100 µl reaction for E-DNA^G included: 20 ng/µl template DNA, 2 µM primer, 60 µM each dNTP, 1U *Taq* DNA Polymerase and 1X ThermoPol buffer (NEB, Inc.). The PCR cycle performed was: 1) 2 min at 94 ^{o}C, 2) 15 sec at 94 ^{o}C, 15 sec at 51 ^{o}C, 30 sec at 72 ^{o}C (35 cycles), 3) 5 min at 72 ^{o}C. DNA samples were purified by polyacrylamide gel electrophoresis (PAGE) and electroelution, followed by ethanol precipitation, and resuspended in 15 mM sodium phosphate, 25 mM sodium chloride, 0.1 mM EDTA, pH 6.8 NMR buffer. The final yields were ~ 0.1 - 0.2 mM duplex in 300 µl NMR buffer starting with ~ 250 - 300 nmoles of primers. To improve reaction yields and obtain more NMR-friendly concentrations, one could use higher initial primer quantities and/or more expensive DNA polymerases that exhibit higher thermal stability and processivity. We prepared the three constructs shown in Figure 2.2 and described in the main text, which contain three isotopically labeled nucleotides (G or T) per domain placed in a unique sequence environment.

Short $^{13}C/^{15}N$-labeled labeled DNA dodecamers (Figure 2.4) were synthesized *in vitro* by the method of Zimmer *et al.*[39], using a template hairpin DNA (IDT, Inc.), Klenow fragment DNA polymerase (NEB, Inc.), and uniformly $^{13}C/^{15}N$-labeled dNTPs (Isotec, Sigma-Aldrich). Single-stranded DNA products were purified by a 20% denaturing PAGE, isolated by passive elution from crushed gels and desalted on a C18 reverse-phase column (Sep-pak, Waters). Oligonucleotides were further lyophilized and complementary strands were resuspended separately in NMR buffer (15 mM sodium phosphate (pH 6.8), 25 mM sodium chloride, 0.1 mM EDTA) supplied with 10% D_2O. Sample annealing was monitored by quick $^{13}C,^{15}N$ HSQCs until single strand signal were not longer observed, with typically duplex concentrations of 0.5 – 1.0 mM for NMR studies. Unlabeled DNA constructs were prepared directly from oligos purchased from the manufacturer. Oligos were resuspended in NMR buffer at ~ 200 μM concentration and their exact concentration measured by UV absorbance at 260 nm using extinction coefficients provided by the manufacturer. DNA duplexes were annealed by mixing an equal molar ratio of the complementary DNA strands, heating for 2 min at 95°C and gradual cooling (~ 30 min) at room temperature. DNA preparations were further washed 3X in resuspension buffer by micro-centrifugation using an Amicon Ultra-4 centrifugal filter (3 kDa cutoff), concentrated to a volume of ~ 250 μl (~ 2-4 mM) for NMR studies and supplied with 10% D2O.

2.2.2 Resonance assignments and intensities

Unlabeled DNA duplexes were assigned using conventional 2D $^1H,^1H$ NOESY (mixing time ~ 150 - 200 ms) in 10% D_2O at 5°C and/or 26°C. Proton assignments were transferred to 2D $^1H,^{13}C$ or $^1H,^{15}N$ HSQC spectra, allowing convenient assignment of base C2H2, C6H6, C8H8, N1H1, N3H3 and sugar C1'H1' in unlabeled DNA and their equivalent short $^{13}C/^{15}N$-labeled duplexes or corresponding fragments from $^{13}C/^{15}N$-labeled E-DNA. Resonance intensities were obtained from $^1H,^{13}C$ or $^1H,^{15}N$ HSQC correlation spectra and normalized for each type of bond vector (C2H2, C6H6, C8H8, N1H1, N3H3 and C1'H1') to the intensity of a helical residue showing an average value that was set to 0.1. All NMR experiments were performed on a Bruker Avance 600 MHz NMR spectrometer equipped with a 5mm triple-resonance cryogenic probe.

2.2.3 Measurement and analysis of ^{13}C and ^{15}N relaxation data

^{13}C and ^{15}N $R_{1\rho}$ relaxation measurements for E-DNA were conducted at 37 $^{\circ}$C using 1D on-resonance $R_{1\rho}$ relaxation experiments that employ selective Hartmann-Hahn excitation of the carbon or nitrogen nuclei of interest, as described[40,41]. Data was recorded using high spinlock field strengths of 3.5 kHz (^{13}C) and 1.0 kHz (^{15}N) for the deoxyribose C1' and base imino N1 spins respectively and four-point relaxation profiles as a function of spin-lock time. In this experimental scheme, each decay is recorded individually, which allows for optimization of signal-to-noise (S/N) and number and values of relaxation delays. For C1', data was collected with 1024 scans (~ 2 hrs) and relaxation delays {0, 6, 12, 20 ms} for G53 (ctr), G51 (ctr), and G50 (ctr) (domain II) and with 2048 or 4096 (~ 4 or 8 hrs) scans and relaxation delays {0, 2, 4, 7 ms} or {0, 5, 10, 15 ms} for G34 (ctr), G80 (ctr), G83 (ctr) (domain I) (see Figure 2.2). For N1, data was acquired with 1024 (G53 (ctr)), 2048 (G50 (ctr)) and 4096 (G83 (ctr)) scans and the following relaxation delays: G53, {0, 10, 20, 30 ms}; G53 (ctr), G50 (ctr), and G83, {0, 8, 16, 24 ms}; G83 (ctr), {0, 5, 10, 15 ms} (Figure 2.3). Relaxation profiles were processed with nmrPipe[42] and relaxation rate constants determined by fitting the resonance intensities to monoexponential decays using Mathematica 6.0.

For ^{13}C/^{15}N-labeled DNA dodecamers, ^{13}C relaxation rate constants R_1 and R_2 were measured using a 2D $R_{1\rho}$ relaxation experiment[32,43,44] for base C2, C6, and C8, and sugar C1' spins using a 3.5 kHz spinlock field strength and a spinlock carrier centered at C6 (for C2/C6/C8) or C1' resonances. Spinlock powers were sufficiently high to suppress undesired chemical exchange contributions and ensure Hartmann-Hahn contributions of < 1 % for J_{CC} ~ 10 Hz and < 0.1 % for J_{CC} ~ 1 Hz. Relaxation data were collected with 8 scans (~ 6 - 7 hrs) and delay series {20, 100, 250, 450 (X3) ms} for R_1 and {4, 16, 32, 48 (X3) ms} for $R_{1\rho}$ with triplicate measurements for error estimation. Relaxation profiles were processed and analyzed as described above. R_2 relaxation rates were computed from R_1 and $R_{1\rho}$ using the established relationship $R_2 = (R_{1\rho} - R_1 \cos^2 \theta) / \sin^2 \theta$ [14]. Relative order parameters (S^2_{rel}) represent estimates of ps-ns motional amplitudes and are largely

independent of motional timescales, were computed as normalized $2R_2 - R_1$ values, as described in Chapter 1[32,45].

Table 2.1: ^{13}C and ^{15}N R_2 relaxation parameters for E-DNA constructs.

Residue	E-DNA^G		E-DNA^{GCtr}	
	^{13}C R_2 (Hz)	^{15}N R_2 (Hz)	^{13}C R_2 (Hz)	^{15}N R_2 (Hz)
Domain I				
G83	89.8 ± 6.5	48.9 ± 1.2	90.1 ± 9.5	44.0 ± 4.8
G80	95.9 ± 9.0	n/a	86.6 ± 5.4	n/a
G34	106.9 ± 1.9	n/a	102.2 ± 2.7	n/a
Domain II				
G50	56.8 ± 2.4	31.7 ± 2.0	59.2 ± 1.7	30.8 ± 2.8
G51	49.8 ± 1.1	n/a	53.9 ± 3.6	n/a
G53	48.1 ± 0.5	27.5 ± 1.2	46.1 ± 0.6	28.9 ± 2.1

n/a indicates data not collected due to resonance overlap.

Table 2.2: R_1 and R_2 ^{13}C relaxation rates for ^{13}C/^{15}N-labeled DNA dodecamers.

Residue	Spin	A_6-DNA		A_4-DNA		A_2-DNA	
		R_1 (Hz)	R_2 (Hz)	R_1 (Hz)	R_2 (Hz)	R_1 (Hz)	R_2 (Hz)
1	C2	2.42±0.04	25.26±0.43				
	C6/C8	2.62±0.02	23.83±0.47	2.69±0.01	22.70±0.04		
	C1'	1.52±0.01	7.82±0.07	1.56±0.01	8.37±0.11		
2	C2						26.03±0.42
	C6/C8	2.28±0.02	21.18±0.27	2.35±0.01	20.30±0.23	2.76±0.03	
	C1'	1.96±0.02	13.46±0.06	1.97±0.01	13.81±0.28	1.88±0.01	11.97±0.20
3	C2	2.42±0.04				2.43±0.03	
	C6/C8	2.31±0.02	22.24±0.20	2.91±0.02	26.58±0.31	2.34±0.02	22.40±0.26
	C1'	1.83±0.01	14.68±0.13	1.84±0.02	11.80±0.35	1.92±0.01	14.48±0.22
4	C2		25.80±0.12				
	C6/C8	2.77±0.01					25.78±0.37
	C1'	1.86±0.02	13.14±0.11	1.99±0.02	14.53±0.17	1.82±0.01	14.51±0.12
5	C2			2.54±0.01	25.36±0.46		
	C6/C8			2.40±0.01	21.52±0.20	2.89±0.02	27.56±0.34
	C1'	1.87±0.02	15.81±0.11	1.92±0.01	14.35±0.28	1.94±0.01	11.18±0.08
6	C2						
	C6/C8			2.81±0.01	25.90±0.17		
	C1'	1.88±0.02	15.44±0.16	1.88±0.01	14.66±0.50	1.95±0.01	14.72±0.16
7	C2					2.49±0.03	26.03±0.42
	C6/C8						
	C1'	1.86±0.03	15.54±0.07	1.89±0.01	15.35±0.22		
8	C2						
	C6/C8					2.86±0.04	26.67±0.44
	C1'	1.84±0.03	15.80±0.10	1.91±0.03	14.83±0.41	1.84±0.01	14.34±0.24
9	C2						

	C6/C8	2.77±0.01	26.33±0.41	2.89±0.01	25.33±0.29	2.80±0.02	25.87±0.27
	C1'	1.80±0.02	14.90±0.10	1.92±0.01	13.10±0.25	1.86±0.01	12.93±0.38
10	C2						
	C6/C8	2.33±0.03	21.33±0.29	2.43±0.01	20.58±0.28	2.37±0.03	20.93±0.27
	C1'	1.88±0.02	14.18±0.14	1.93±0.02	14.20±0.24	1.92±0.01	14.94±0.22
11	C2						
	C6/C8	2.30±0.02	21.19±0.39	2.39±0.01	20.06±0.21	2.36±0.03	20.60±0.24
	C1'	1.88±0.02	14.60±0.09	1.96±0.01	12.32±0.35		
12	C2						
	C6/C8	2.75±0.01	26.74±0.25	2.83±0.03	24.35±0.35	2.82±0.03	25.93±0.39
	C1'	1.88±0.02	11.32±0.11	1.74±0.01	11.64±0.10	1.72±0.01	11.90±0.16
13	C2						
	C6/C8	2.23±0.03	20.13±0.26	2.29±0.01	18.72±0.34		
	C1'	1.76±0.01	8.31±0.09	1.71±0.01	8.09±0.21		
14	C2						
	C6/C8	2.78±0.02	27.51±0.32	2.88±0.02	26.56±0.27		
	C1'	1.82±0.02	12.32±0.13	1.85±0.02	12.48±0.35	1.85±0.01	12.62±0.13
15	C2						
	C6/C8	2.79±0.02	27.72±0.48	2.91±0.02	27.99±0.71		
	C1'	1.86±0.01	11.28±0.09	1.92±0.01	11.02±0.07	1.92±0.01	11.56±0.25
16	C2	2.40±0.03	26.14±0.27	2.51±0.01	25.55±0.27	2.45±0.03	25.34±0.22
	C6/C8	2.28±0.02	22.26±0.17	2.34±0.01	21.52±0.09	2.31±0.02	22.20±0.23
	C1'	1.83±0.04	14.55±0.11	1.90±0.03	14.38±0.32		
17	C2	2.40±0.02	26.34±0.32	2.52±0.02	25.49±0.34	2.46±0.04	25.81±0.33
	C6/C8	2.28±0.02	22.12±0.14	2.38±0.01	21.40±0.34		
	C1'	1.84±0.02	14.55±0.11	1.91±0.01	13.97±0.24	1.85±0.01	14.60±0.15
18	C2			2.56±0.01	25.61±0.39		
	C6/C8			2.38±0.01	21.83±0.37	2.78±0.03	26.17±0.25
	C1'	1.82±0.04	15.21±0.08	1.88±0.03	14.32±0.14	1.82±0.01	14.01±0.26
19	C2			2.50±0.02	26.08±0.40		
	C6/C8	2.32±0.02	22.55±0.25	2.39±0.02	22.31±0.24	2.89±0.02	28.60±0.31
	C1'	1.82±0.03	14.77±0.10	1.93±0.02	14.64±0.40	1.91±0.01	11.87±0.14
20	C2	2.49±0.02	26.96±0.23				
	C6/C8	2.34±0.02	22.83±0.37	2.84±0.01	25.19±0.50		
	C1'	1.86±0.01	14.62±0.08	1.84±0.01	13.91±0.09	1.93±0.02	14.36±0.20
21	C2	2.21±0.03	26.39±0.31			2.48±0.03	25.34±0.38
	C6/C8			2.98±0.01	27.64±0.36		
	C1'	1.93±0.03	12.63±0.10				
22	C2						
	C6/C8	2.79±0.02	26.55±0.23			2.77±0.02	25.50±0.30
	C1'	1.84±0.03	14.00±0.08	1.93±0.02	14.46±0.26	1.86±0.01	12.87±0.25
23	C2						
	C6/C8	2.83±0.03	27.68±0.37	2.93±0.01	26.51±0.30		
	C1'	1.85±0.02	10.62±0.07				
24	C2						
	C6/C8	2.28±0.02	19.60±0.30	2.28±0.01	19.19±0.30		
	C1'	1.76±0.02	11.50±0.05	1.79±0.01	11.44±0.23	1.73±0.01	11.70±0.09

3D models for E-DNA with a straight or a bent A-tract were constructed in Insight II (Accelrys, San Diego) by assembling B-form helices corresponding to domain I and domain II on to a B-form A_6-tract or an known NMR structure of an A_6-tract with an overall 19 degree bend (PDB ID: 1FZX)[46]. The overall diffusion tensor was predicted using HydroNMR[47] at 37 °C with terminal base pairs excluded. Relaxation rates were computed using in-house software[32] assuming $S^2 = 0.85$ and a internal correlation time $\tau_f = 100$ ps.

2.3 Results and Discussion

2.3.1 Design of site-specifically ^{13}C/^{15}N-labeled E-DNA

Here, we extend the elongation methodology and tailor it towards the atomic characterization of dynamical features in DNA. Specifically, we present a strategy for preparation and assignment of long DNA duplexes (E-DNA) that are ^{13}C/^{15}N-labeled in a residue-selective manner to probe inter-helical bending motions induced by an A-tract sequence[48,49]. In the new DNA elongation strategy, a sequence of interest, such as an A-tract, is flanked by short and long B-form helical domains, which act as reporters of any inter-helical motions occurring across the target sequence (Figure 2.1). The longer helix dominates overall diffusion and alignment whereas the short helix reports on local as well as collective dynamics occurring across the DNA target sequence. The sequence is devised in such a way that enables the site-specific incorporation of labeled nucleotides at particular positions in each helical domain. The design minimizes resonance overlap, which is key for dynamics measurements, and yet provides adequate domain-specific information. The elongated DNAs provide a unique opportunity to resolve
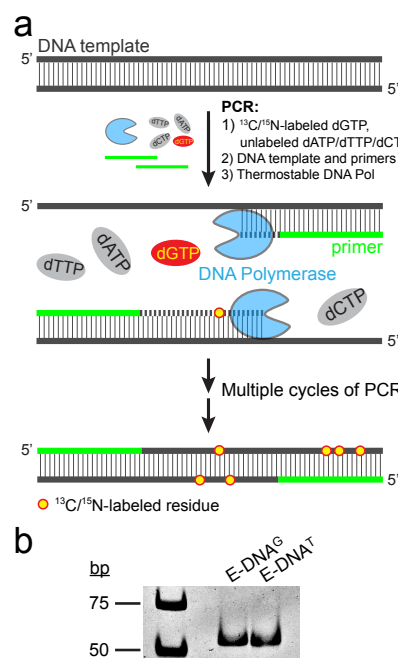
Figure 2.1: PCR scheme for the synthesis of site-selective ^{13}C/^{15}N-labeled E-DNA. (a) An example with labeled G nucleotides, depicted as yellow-red circles. (b) 15% native PAGE of 56 base pair long E-DNA^G (left) and E-DNA^T (right) relative to a DNA standard.

"static" versus "dynamic" bends using NMR spin relaxation measurements, as previously described for RNA[30]. For example, static directional bends are expected to reduce the angle between the nucleobase C-H (or N-H) bonds in the short helix and the slowly tumbling long axis of the DNA, thus resulting in an increase in the observed correlation times relative to perpendicular bonds in the elongated helix. Conversely, dynamic bends at timescales faster than overall tumbling are expected to reduce the observed correlation times for C-H (or N-H) bonds in the short relative to the elongated helix.

The elongated DNA (E-DNA) duplexes shown in Figure 2.2 consist of 56 base pairs (~ 35 kDa) and are optimized to probe collective helical bending motions activated by an A-tract element. The E-DNA duplexes are prepared in a series of enzymatic reactions involving the target DNA, an unlabeled primer pair, a thermostable DNA polymerase and only one type of uniformly ^{13}C/^{15}N-labeled nucleotides (G or T). Various efficient PCR and primer extension methods for the synthesis of large quantities of isotopically labeled oligonucleotides with applications to NMR studies have been previously presented [39,50-53] and are suitable for preparation of short or non-site-specifically labeled DNA. Here, a traditional PCR-based method combined with an appropriate sequence design and labeled nucleotide selection was implemented for the *in vitro* preparation of E-DNA. Although the synthesis requires multiple reaction runs, current advances in large-scale PCR can facilitate the rapid production of E-DNA samples in NMR-friendly quantities. We prepared the three constructs shown in Figure 2.2a, in which the identity of the labeled nucleotides and partially the sequence around them are altered. E-DNA^G and E-DNA^T contain G and T ^{13}C/^{15}N-labeled residues respectively and share the same primer-defined regions, including the A-tract element and the short helical domain II. E-DNA^{GCtr} is used as a control and is identical to E-DNA^G except that the A$_6$-tract is replaced with an (AC)$_3$ element. The final DNA constructs contain three isotopically labeled nucleotides per domain placed in a unique sequence environment in order to reduce spectral crowding (Figure 2.2).

2.3.2 "Divide and conquer" assignment strategy for E-DNA

The specific labeling strategy allowed us to collect well-resolved, excellent quality 2D spectra for the 56 base pair E-DNA. The sparsely labeled, elongated DNA
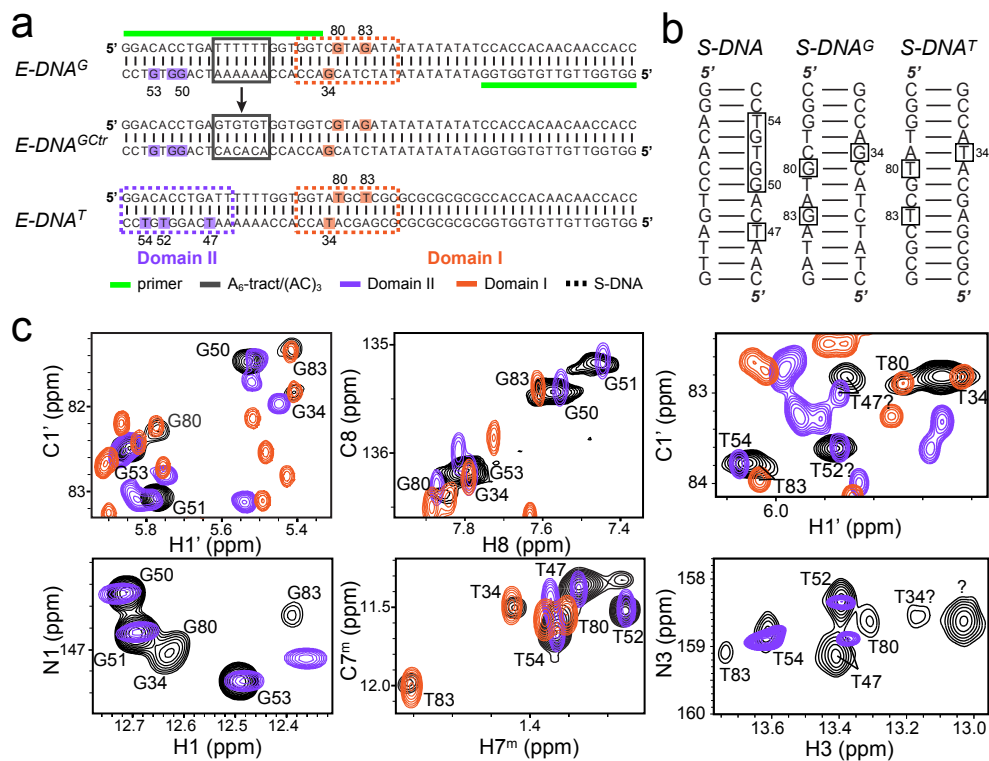
Figure 2.2: Preparation and resonance assignment of E-DNA. (a) E-DNA constructs (56 base pairs) showing the location of the A_6-tract or $(AC)_3$ (solid black), the PCR primers (solid green), and the overlapping sequence with short S-DNA duplexes (dashed purple/orange) used for resonance assignments of E-DNAs. $^{13}C/^{15}N$-labeled residues are boxed and numbered (b) S-DNA duplex sequences. (c) 2D HSQC (S^3CT) and HMQC spectral overlays of E-DNA^G/E-DNA^T (black) with S-DNA (purple) and S-DNA^G/S-DNA^T (orange) showing excellent chemical shift correspondence.

requires a non-traditional approach for resonance assignment given its large molecular size and mostly non-sequential positioning of the labeled nucleotides. To this end, we have implemented a "divide and conquer" approach for assigning resonances in E-DNA (Figure 2.2). Here, assignments are first determined using conventional methods for unlabeled short duplexes (S-DNA) corresponding to the labeled fragments from the individual domains in E-DNA (Figure 2.2a). The S-DNA constructs (Figure 2.2b) preserve the sequence identity of nearest neighbors and are expected to have chemical shifts comparable to those of labeled probes in E-DNA. The E-DNA resonances were then assigned by overlaying natural abundance 2D correlation maps (combined with 1D imino spectra) of S-DNA with those of E-DNA. As shown in Figure 2.2, we observe excellent chemical shift correspondence between short and long DNA constructs allowing us to unambiguously assign the far majority of sugar and nucleobase resonances

in the elongated duplexes. Since chemical shifts are sensitive reporters of nucleic acid structure, this observation supports local structural similarity between the short and long DNA molecules with the exception of the A-tract junction. Decomposition of the E-DNA sequence into small fragments can also be employed in the reverse manner; inexpensive unlabeled DNA sequences can be used to predict optimal spectral quality and avoid potential overlap prior to the synthesis of large constructs.

2.3.3 Position-dependent fast dynamics in A-tract and control E-DNA

Despite the agreement in chemical shift, 2D spectra of E-DNA^G and E-DNA^T recorded at 37 ^{o}C featured marked differences in resonances intensities that were not apparent in corresponding short DNA duplexes. For example, simple examination of the spectra of Figure 2.2 clearly shows that residue G53 in the short helix has elevated intensities compared to G83 in the elongated helix. Measurements of sugar backbone (C1'H1') and base (C7^{m}H7^m, C8H8, N1H1, N3H3) resonance intensities revealed an interesting pattern in which values for the short helix are 2 - 3 fold higher than those belonging to the long helix (Figure 2.3a). This data suggest that residues in the short helix experience a greater degree of picosecond-to-nanosecond internal motions compared to residues in the elongated helix and/or that residues in the long helix experience a greater degree of linewidth broadening due to chemical exchange at the μs-ms timescale. A similar pattern of resonance intensities is observed in the E-DNA^{GCtr} control sample, indicating that the observed motional process is not induced by the A-tract (Figure 2.3a).

To rule out the possibility of chemical exchange on the μs-ms timescale and gain further insights into the motions in E-DNA, we measured ^{15}N and ^{13}C transverse relaxation rate constants for the sugar C1' and imino N1 nuclei in E-DNA^G and E-DNA^{GCtr} (Figure 2.3b). Such quantitative measurements presented a challenge given the large size and relatively low concentration (0.1-0.2 mM) of the E-DNA samples. To overcome this problem, we used $R_{1\rho}$ experiments that employ selective Hartmann-Hahn polarization transfers to excite specific spins of interest and collect data in a 1D manner[40,41,54]. The 1D acquisition scheme affords shorter data collection times and thus improved sensitivity over corresponding 2D schemes. The experiment interrogates

individual peaks in a single run and was thus perfectly suited for probing the small number of well-resolved resonances in E-DNA.

The on-resonance $R_{1\rho}$ relaxation data paint a picture analogous to that obtained from examination of resonance intensities. Similar R_2 values were measured in E-DNA^G and E-DNA^{GCtr} (Figure 2.3b), indicating that the A-tract does not induce fluctuations detectable by spin relaxation measurements under our experimental conditions. For both E-DNA^G and E-DNA^{GCtr}, the R_2 data measured in the elongated domain is in excellent agreement with expectations based on a hydrodynamically estimated overall tumbling rate of ~ 24 ns (37 ^{o}C) and restricted internal motions (S^2 ~ 0.85) (Figure 2.3b). In contrast, for both E-DNA^G and E-DNA^{GCtr}, the R_2 values measured for the shorter helix II were significantly smaller. The differences in R_2 measured in the long and short helices cannot be explained simply based on differences in the orientation of spin interactions relative to the elongated axis. First, as described in applications to E-RNA[30,32], for the nucleobases, the carbon and nitrogen interactions in the long helix are already positioned optimally perpendicular to the long axis and thus correspond to the lower limit of R_2 rates
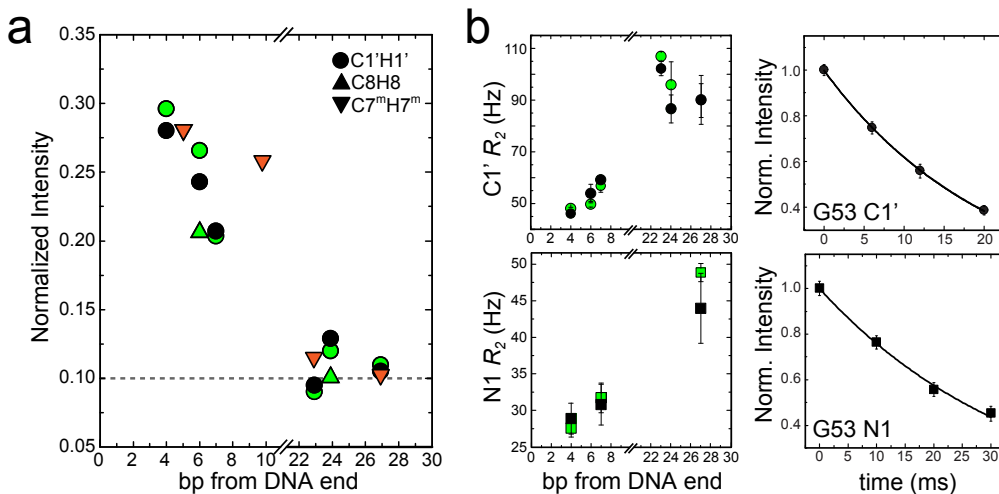
Figure 2.3: Dynamics characterization of E-DNA. (a) Resonance intensities as a function of distance (in base pair (bp)) from the DNA ends showing elevated motions in helical domain II (based on spectra in Figure 2.2). Different constructs are color-coded with green (E-DNA^G), black (E-DNA^{GCtr}), and red (E-DNA^T). (b) ^{13}C (C1', top) and ^{15}N (N1, bottom) R_2 relaxation rates (left panels) supporting fast dynamics in domain II. Data was collected using selective 1D $R_{1\rho}$ relaxation experiments with the spinlock carrier placed on-resonance and spinlock field strength of 3.5 kHz for ^{13}C and 1.0 kHz for ^{15}N. Right panels show sample monoexponential decays for G53 in E-DNA^G

that can be measured in the absence of motions due to anisotropic tumbling. Second,

hydrodynamic computations on an E-DNA model incorporating a bent A-tract verses a straight A-tract suggest that the curvature induces minor differences in R_2 in the opposite direction from what is observed in relaxation measurements (data not shown). Thus, the relaxation data exclude differential chemical exchange broadening as the source of large variations in resonance intensities and suggest the presence of internal motions at ps-ns timescales.

Interestingly, the internal motions are not observed in short (13 base pair) DNA constructs containing identical sequences. However, upon introduction of glycerol (~ 20% w/v) to a short DNA duplex (14 base pair, $\tau_m \sim 6$ ns) to increase solvent viscosity and decrease the rate of overall tumbling, we observed a relative increase in the intensities of resonances belonging to terminal as well as internal residues (Figure 2.6). This suggests that the local fluctuations resolved in E-DNA occur at timescales comparable to overall tumbling of short DNAs (typically < 6 ns) and thus evade detection, as has been observed in a number of different RNA contexts[30,55,56].

While the current preliminary analysis cannot allow for a detailed characterization of the nature of the motions, the resonance intensities and the relaxation data measured in the E-DNA indicate that the motions are gradually attenuated as a function of distance from the terminal end and are completely diminished at centrally positioned residues (Figure 2.3). It is possible that end-fraying motions occurring at ns timescales reverberate deep (seven base-pairs) within the helix and gradually fade away as one penetrates further into the helix interior. This observation is supported by recent theoretical predictions for a weakly deformable DNA with internal flexibility[57]. Moreover, the fast transitions could be accompanied by torsional or flexural backbone deformations potentially coupled to these internal modes near the DNA ends. Although the positional effect seems to dominate the trend, we cannot exclude contributions from sequence-specific local dynamics. Additional constructs, which contain domain-swapped residues, could be used in the future to resolve this ambiguity.

Finally, we note that our results do not rule out the presence of bending motions that occur at timescales much slower than can be probed by spin relaxation. Preliminary results on a short duplex containing the equivalent A_6-tract provide evidence for microsecond exchange dynamics near the A-tract junctions, which could activate slower

bending dynamics (manuscript in preparation). Measurements of additional relaxation data as well as RDCs in E-DNA should provide insights into the nature of both local and bending motions over a wider range (ps-ms) of timescales. A-tract induced bending has also been shown to be strongly dependent on temperature and presence/absence of divalent ions such as Mg^{2+}[58-68]. The high temperature and lack of Mg^{2+} ion conditions used in our experiments are believed to suppress A-tract induced bending. Thus, it will be important to also perform NMR dynamics studies at variable ambient conditions to comprehensively explore the landscape of A-tract induced dynamics.

2.3.4 A-tract dependent conformational variations in short $^{13}C/^{15}N$-labeled DNA dodecamers

We prepared three uniformly $^{13}C/^{15}N$-labeled DNA dodecamers that incorporate two, four, and six consecutive adenines, A-tracts, capped by GC-rich helices (Figure 2.4a). The A_6-tract (A_6-DNA) appears in a context (5' CA_6T) commonly encountered in the kinetoplast sequence originally found to exhibit microscopic bending when phased periodically with the helical repeat[48]. The A_4-tract and A_2-tract (A_4-DNA and A_2-DNA) are designed as shorter versions of the above but with identical 5' and 3' junctional base pairs (Figure 2.4a). An isolated 5' CA/TG step was additionally included in A_2-DNA as a control for A-tract dependent dynamics at the 5' CA/TG boundary. Complete resonance assignments were obtained for nucleobase C2H2, C6H6, and C8H8 and deoxyribose C1'H1' in the three DNA constructs (Figure 2.4b). $^{13}C,^{1}H$ HSQC and 2D NOESY spectra of A_6-DNA and A_4-DNA displayed chemical shifts (CS) and NOE connectivities characteristic of asymmetric A-tracts, which have been ascribed by prior solution NMR studies and CS predictions to structural departure from canonical B-form[46,69-71]. For example, we observed strong inter-strand NOE cross-peaks between the base H2 of A and the sugar H1' or imino H3 of the 3'-neighboring T on the complementary strand previously correlated with minor groove compression and large propeller twist[69,70]. In addition, purine A H8/H1' and pyrimidine T H3/H1' protons displayed upfield and downfield chemical shifts respectively characteristic of A-tracts (Figure 2.4b). Unusual upfield proton CS perturbations of sugar moieties at the 5' CA/TG junction, C15 and A16

H1', also reproduced the unique A-tract spectral signature seen in previous reports[69,70] and attributed to deformations at the boundaries.

The A-tract-specific spectral characteristics gradually diminished as the adenine run was shortened from six to two base pairs, with a more dramatic change between A_4-DNA and A_2-DNA. Upon A-tract shortening, internal A H8/H1' and T H3/H1' protons shifted to lower or higher fields respectively relative to residues found in the heterogeneous part of the sequence (Figure 2.4b). Base and sugar protons at the common 5' CA/TG junction, C15 H6/H1' and A16 H8/H1', experienced a small downfield shift between the A_6-tract and A_4-tract, while the effect was more pronounced between A_4-DNA and A_2-DNA. Such significant perturbations would not be typically predicted for changes in non-immediate neighbors in canonical B-DNA and suggest the existence of sequence-specific conformational variability between different length A-tracts. Smaller CS deviations with A-tract length were observed near the 3' junction and could reflect effects from proximity to the DNA terminus. Overall, the isotropic chemical shift analysis provides evidence for the presence of size-dependent A-tract conformational differences around the 5' and 3' junctions that could potentially correlate with A-tract bending.
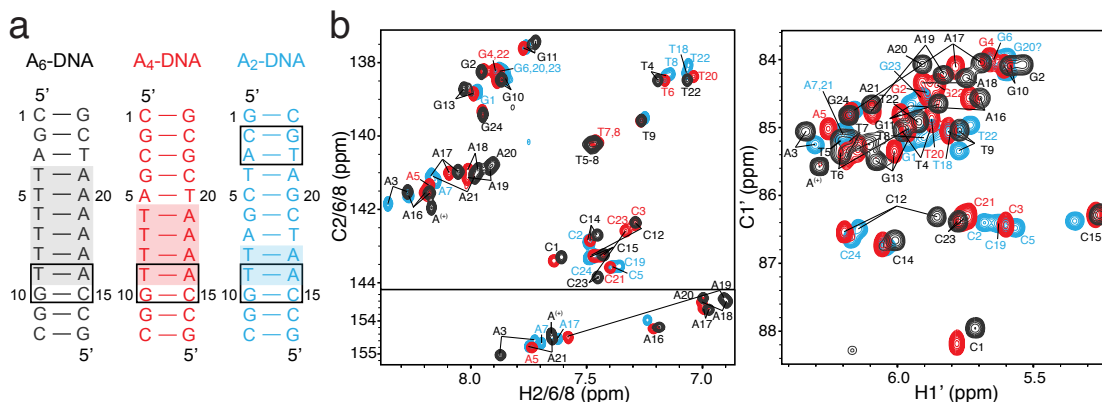
Figure 2.4: DNA dodecamer constructs with CA steps and varying A-tract length and resonance assignments. (a) DNA sequences for A_6-DNA (black), A_4-DNA (red), and A_2-DNA (blue). (b) 2D ^{1}H,^{13}C HSQC spectra showing base and sugar correlations at 26 °C and pH 6.8. Residue labels are color-coded except for common residues between the three DNA constructs (labeled in black).

68

2.3.5 Ps-ns motional trends from resonance intensity and ^{13}C relaxation analysis

We have examined the ps-ns motional signature of short $^{13}C/^{15}N$ isotopically enriched samples, containing variable size A-tracts. The sequence for the A_6-DNA is identical to the central A_6-tract in E-DNA used to study fast ps-ns motions induced in the two satellite helices. We measured fast dynamics using longitudinal (R_1) and transverse (R_2) ^{13}C spin relaxation for nucleobase C2, C6, C8 and deoxyribose C1' nuclei combined with resonance intensities for the same C-H bond vectors in variable conditions. Figure 2.5 shows the comparative analysis of the normalized intensities and ^{13}C relaxation data expressed as relative order parameters (S_{rel}^2) for the different bond vectors (see Section 1.2.2)[32,45,72].

The intensity profiles provide a qualitative insight into biomolecular dynamics over a broad range of timescales spanning ps to ms, and can be analyzed as a first quick-and-easy step towards dynamics characterization. NMR linewidths are directly related to the transverse relaxation rate constant R_2 and can be influenced by a variety of motions. The phenomenon of line narrowing is a result of rapid conformational averaging of the dipole interactions cause slower relaxation on timescales faster than overall rotational diffusion, and strongly indicate faster ps-ns motions. The rotational correlation time for the DNA dodecamers here is estimated to be ~ 4.5 ns based on hydrodynamic predictions, meaning that resonance line shapes would be sensitive to motions faster than this diffusion limit. Conversely, line broadening arises from a rapid dephasing of the transverse magnetization for a given nucleus due to exchange between distinct chemical environments on timescales slower than diffusion, typically in the μs-ms window.

In general, we observed fairly uniform intensities for base and sugar C-H bonds of internal residues and across the three dodecamers. Terminal and next-to-terminal residues demonstrated significant degree of ps-ns flexibility as expected from reduced stacking and base pair fraying effects at the DNA ends. This observation was in contrast to E-DNA, where intensities were still 2-fold above the norm seven base pairs away from the duplex end. The most consistent trend in short DNAs was found for C6H6 resonances corresponding to Ts, which appeared to be more intense than other pyrimidines (C) and purines (A and G). In addition, A16 C2H2 and C8H8 located at 5' A-tract junctions was characterized by broader linewidths, pointing to a slow chemical exchange process that

persisted from A_6-DNA to A_2-DNA. Contrary to the nucleobases, intensities for sugar C1'H1' resonances of internal residues were more variable with lowest values seen for Ts inside A_6- and A_4-tracts, and highest values seen for Cs.

Further insights into dynamics can be obtained by a temperature dependence of the resonances intensities of the short DNA constructs. In general, higher temperatures would induce greater disorder and faster local fluctuations that would translate into overall higher resonance intensities. However, higher temperatures simultaneously would decrease solvent viscosity and speed up global diffusion, enhancing the motional coupling between ns internal and overall motions and limiting the window for observable internal contributions, which would oppose/reverse the increase in resonance intensities and mask ns local dynamics. When temperature was varied in the range between 10 $^{\circ}$C and 40 $^{\circ}$C, the normalized resonance intensities normalized showed negligible variations for most base and sugar probes, implying that the change in thermal environment had a uniform effect on the timescale/amplitude of internal mobility. However, a number of residues also showed consistent reduction in intensity with decrease in temperature that could arise from a combination of relatively slower timescales and smaller amplitudes of internal motions. This trend was especially prominent for sugar C1'H1' of Ts and As that are part of longer A-tracts and for the 3' A-tract TA junction, but also extended to the base C8H8 sites for several Gs (G2, G10 and G11). The reverse trend, higher intensity values with lower temperature, was characteristic of cytosine sugar C1'H1' and signified the presence of ns internal motions that became progressively resolved as global tumbling was retarded in colder conditions.

Due to the broad timescale sensitivity of resonance intensities, simultaneous fast and slow motions could have opposing effects on the intensity and mask each other's contribution. To tap specifically into the ps-ns motions in a quantitative fashion, we analyzed ^{13}C relaxation data for the same base and sugar carbons above in terms of relative motional amplitudes (S_{rel}^2). S_{rel}^2 values range from 1 to 0 for a fully rigid to spatially unrestricted internal motion (see Section 1.2.2). Several trends emerge while examining the relaxation properties of the three DNA dodecamers. First, the nucleobases in DNA appeared relatively immobilized with $S_{rel}^2 > 0.9$ for non-terminal residues,

independent of sequence context. In particular, internal purines (A and G) as well as internal Cs were found to be relatively rigid with $S_{rel}^2 > 0.95$, while internal Ts, particularly those located at the A-tract junctions, were more flexible with $S_{rel}^2 \sim 0.9$. Both the C2 and C8 probes in longer A-tracts conveyed local rigidity at the adenine bases that was also comparable to other purines across the sequences. These data were consistent with trends in previously reported order parameters implying lower-amplitude base motions inside Watson-Crick B-DNA helices (see Introduction). Relative order parameters ranging from 0.9 to 1.0 could be translated to motional fluctuations on the order of $15 - 20^o$ using a nominal $S^2 \sim 0.9$ (based on estimations from preliminary model-free analysis of A_6-DNA) and the "diffusion-in-a-cone" motional model (see Section 1.2.2).

Contrary to bases, backbone sugars showed largely variable ps-ns flexibility ($0.65 < S_{rel}^2 < 1$) that was high for some internal nucleotides in a sequence-specific manner. Out of all residues, Ts exhibited the lowest-amplitude sugar motions, regardless of neighboring sequence, and were characterized by fairly uniform S_{rel}^2 values approaching unity (Figure 2.5c). This observation was in line with resonance intensity profiles and their variation with temperature. An important exception were Ts found in the context of CA/TG steps, T22 in A_2-DNA and T9, which stood out with substantially higher amplitude motions ($S_{rel}^2 \sim 0.85$). The motional profile of T9 located at the 5' A-tract end was evidently unique to its position in a CA/TG step and not dependent on the A-tract length. In contrast to most Ts, the sugars of non-terminal Cs demonstrated high level of mobility ($S_{rel}^2 < 0.8$), especially Cs that belonged to CG or CA dinucleotides ($S_{rel}^2 \sim 0.65 - 0.75$). The relative order parameters for sugar C1' of A-tract As ranged between 0.9 and unity, implying elevated motions and greater dynamic variability than for their complementary Ts. In fact, the number of consecutively stacked As seemed to exert an effect on their local dynamics. For example, A18 sugar dynamics were slightly sensitive to the A-tract length or its specific position along the A-tract, and were greater in A_4-DNA versus A_6-DNA. The A-tract length had the inverse effect on A17, whose sugar motions gradually became more restricted as the A-tract became shorter, with S_{rel}^2 of 0.87

71

in A_6-DNA and 1.0 in A_2-DNA. A more significant trend was observed for the sugars of G and T located at the 5' (G10) and 3' (T22/T20/T18) junction of A-tracts, respectively. Their motions increased consistently in amplitude, S_{rel}^2 varying from ~1.0 to 0.89 for G and ~0.94 to 0.88 for T, as the adjoining A-stretch was elongated.

To summarize the general findings from ^{13}C relaxation measurements, we discovered the following trends regarding ps-ns dynamics across the three DNA dodecamers: (i) high degree of stiffness for sugars of internal A-tract Ts, (ii) high degree of stiffness for bases of internal A-tract As, (iii) A-tract length dependence of sugar motions in internal As and junctions (5'G and 3'T), generally decreasing in amplitude with shorter A-tracts, (iv) generally a direct correlation between the relative mobility of base and sugar units for residues of the same type (i.e. more mobile A sugars correspond to more mobile A bases), (v) generally an inverse correlation between the base and sugar mobility for T and G in CA/TG steps as the A-tract length changes (i.e. G base becomes more rigid, while the G sugar becomes more mobile with longer A-tracts), (vi) a large gap of flexibility between C base (low flexibility) and sugar (high flexibility) at pyrimidines in pyrimidine-purine (YR) CA/TG and CG/CG steps. By and large, the relaxation data portrayed A-tracts as relatively stiff DNA elements, where base rigidification at As was correlated with slightly higher sugar flexibility, while the bases of Ts seemed to move more freely "at the expense" of less flexible sugars (based on few data points for T). This inverse correlation was intriguing and reminiscent of an enthalpy-entropy compensation phenomenon for correlated base-sugar motions. Moreover, the A-tract length and likely variable conformation between short and long adenine stretches seemed to affect the motional properties of residues found at the interface with B-DNA. Thus, we have discovered that the length of A-tracts can modulate their own flexibility and the flexibility of adjacent base pairs, which could have important consequences for their interaction with cellular factors or inclusion into nucleosomes.

The structural rigidity of A-tracts is not a new concept. In fact, AA steps are ranked as the most rigid dinucleotides out of the ten unique ones[73-75]. By examination of crystallized oligomers, it has been shown that AA steps display the largest propeller twist and most limited slide mobility, which seems to have a severe stereochemical "locking" effect and keeps the nucleotides essentially "in place". Though the high propeller twist

introduces a mechanical strain, its effect might be offset by improved π -π base-stacking interactions[76], potential bifurcated hydrogen bonds[73], and the formation of an ordered water network in the resultant narrow minor groove[77], which could all serve as a source of overall stabilization. The preferred conformation of AA steps also explains the very low structural variability across a large database of crystal structures and sequence contexts[75].

Most interestingly, extensive sugar dynamics appeared to be a general feature of T and, particularly, C nucleotides in YR steps (CA/TG and CG). These findings are supported by recent reports that implement ^{13}C NMR relaxation and a combined solution NMR/Molecular Dynamics (MD) approach to investigate fast motions in DNA[19,25,78,79]. There, the nature of a cytosine-specific motion (in CA/TG and CG steps) is linked to conformational heterogeneity involving rapid deoxyribose interconversions, which correspond to changes in sugar pucker from the predominant C2'-endo (S-type) to the less common C3'-endo (N-type) state. Although deoxynucleotides prefer to occupy the S conformational space in B-DNA, C sugar moieties consistently exhibit a higher tendency towards the N conformation, as observed in both solution and computational studies[25,80,81]. These sugar fluctuations from MD simulations show a strong correlation with both C1' order parameters and the BI/BII (epsilon-zeta) equilibrium of backbone phosphates[79], and are accompanied by changes in the glycosidic torsion angle that dictates the relative orientation of the sugar and base moieties. In fact, analysis of DNA crystal structures suggests that CA/TG steps, together with TA steps, are the most flexible of all unique dinucleotides[75] and favor the BII conformer (associated with N-form) more than any other dinucleotide[82]. We note, however, that sugar repuckering could also be facilitated by the frequent removal of the same or nearby base from the helical stack[83]. Enhanced sugar re-puckering with sizeable N-type populations or other sugar-backbone torsional fluctuations could easily transmit to C1' nuclei and explain the high C1' flexibility at YR steps and at certain A-tract residues.
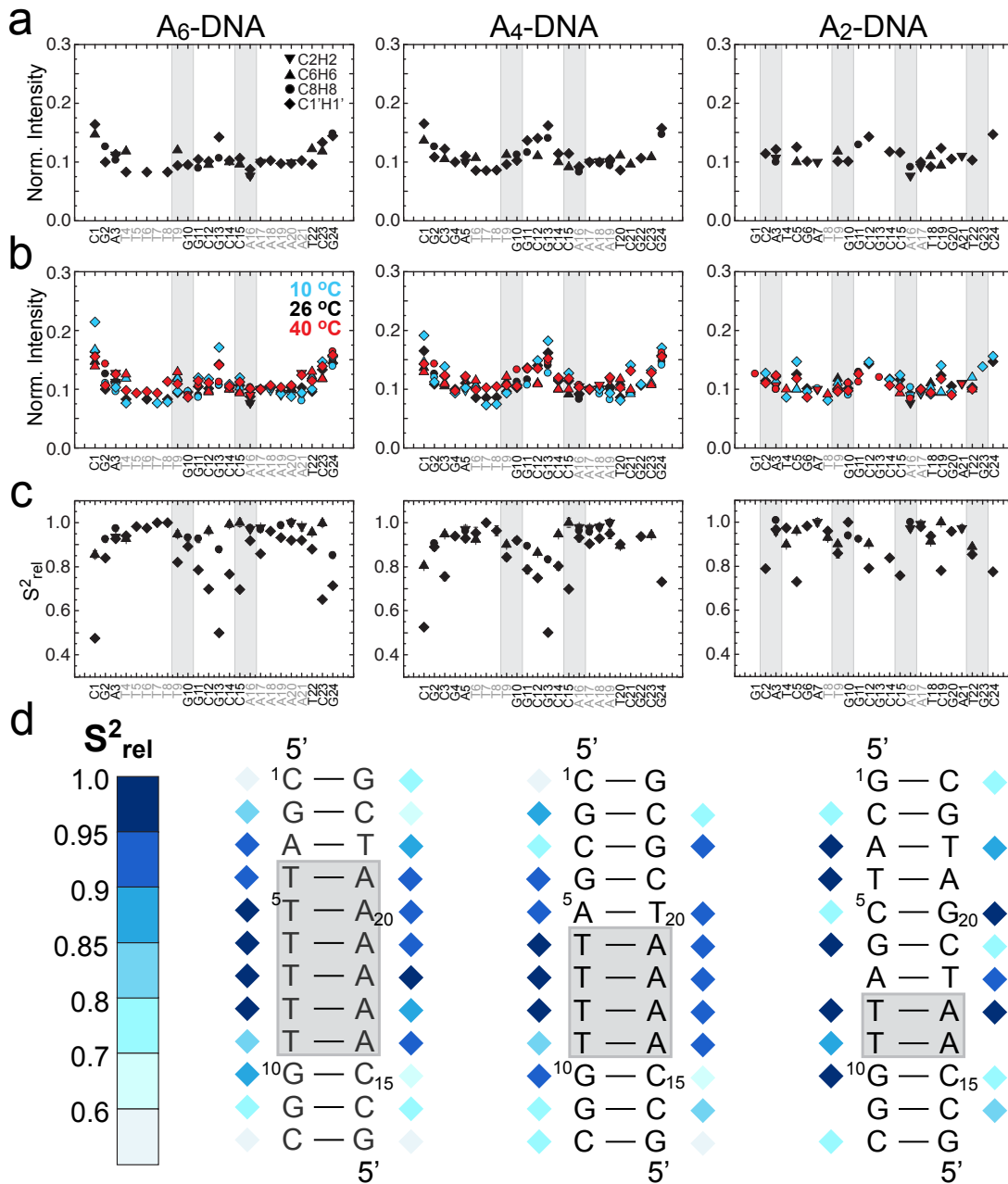
Figure 2.5: Fast dynamics in DNA dodecamers. (a) Normalized resonances intensities for C2H2 (inverted triangle), C6H6 (triangle), C8H8 (circle), C1'H1' (diamond) in A_6-DNA, A_4-DNA, and A_2-DNA collected at 26^oC. The grey boxes highlight the position of CA/TG steps. (b) Intensities as a function of temperature at 10^oC (blue), 26^oC (black), and 40^oC (red). (c) Corresponding relative order parameters obtained from ^{13}C spin relaxation data at $26\ ^oC$. (d) A depiction of S^2_{rel} values for C1'H1' at $26\ ^oC$ showing high variability within and across DNA sequences.

2.3.6 Extending the "NMR-visible" nanosecond window in short DNA

NMR spin relaxation data is most commonly analyzed and interpreted using the model free formalism, where one can extract both the amplitude and timescale of internal motion(s) based on various motional models (see Section 1.2.2). Here, however, we focused on comparing the motional amplitudes rather than timescales, which, on their own, offer valuable information about the dynamic anisotropy at each site. We further recognized that fast-tumbling short DNA dodecamers are not ideal for quantifying the timescales of ns motions due to potential coupling between internal and overall dynamics, which could render the former "invisible". In order to obtain further insights into what residues exhibit elevated ns flexibility, we devised a quick strategy to selectively slow down overall molecular rotation and thereby expose any ns internal dynamics that could have been previously concealed. Our goal is to achieve conditions where internal motions start to decouple from overall motions by simply adjusting the buffer content and without the use of multiple isotopically labeled samples that are required by the domain elongation approach[30]. In our approach, we collect resonance intensities in the absence and presence of glycerol, which increases the viscosity of solvent (water) and retards the overall rotational diffusion in a predictable manner. Here, we used resonance intensities as a rapid diagnostic tool for unusual ns dynamics, although spin relaxation data could be further acquired to quantify these motions.

As a benchmark for this method, we used a 27-nt HIV-1 TAR construct containing a mutant UUCG tetraloop (mTAR), whose ns dynamics have been extensively characterized by $^{15}N^{30}$ and $^{13}C^{32}$ spin relaxation using a helical elongation technique devised in our lab[30]. Upon addition of ~ 25% Glycerol that increases the rotational correlation time of mTAR from ~ 6 ns to ~ 11 ns, we observed line narrowing for several nucleotides that were among those showing ns internal mobility[30] (data not shown).

To probe for unusual ns motions in DNA, resonance intensities were collected for the sugar C1'H1' of unlabeled A_6-DNA, A_4-DNA, and A_2-DNA in the absence and presence of ~ 20% glycerol that increases the viscosity of solvent and retards the overall rotational diffusion of the DNA (~ 1.6X at 25 $^{\circ}C$) (Figure 2.6a). The sugar C1'H1' spectrum was specifically targeted since C1' sites showed noticeable increase in resonance intensities with decrease in temperature, contrary to expectations, and also

unusually high degree of flexibility as judged by S_{rel}^2 parameters. Glycerol did not seem to interact specifically with the DNA, as judged by nearly perfect spectral overlay with and without the retarding agent (Figure 2.6b). Overall, terminal residues and internal Cs as well as G11 showed noticeable increase in intensity upon addition of glycerol, which correlates with their dynamic behavior at low temperature. The data for the 5'CA/TG A-tract junction seemed to be fairly uniform across all three DNA constructs. Thus, by using non-invasive methods to manipulate selectively global molecular tumbling, we were able to deconvolute, to some extent, ns internal motions from overall diffusion and demonstrate their presence in cytosine ribose moities and terminal residues for unlabeled DNA duplexes.

Glycerol concentration could, in principle, be increased to further resolve any invisible ns dynamics. Unfortunately, upon doubling the glycerol from ~20% to 40% (wt/vol) for A_6-DNA, we encountered significant line broadening and spectral overlap as well as undesired artifacts from residual glycerol signals that interfered with DNA resonances. Instead, we decided to use a combination of minimal DNA elongation and glycerol addition to look for further increase in intensity. We repeated the experiments with a 14 base pair DNA construct (A_6-DNA(14)), built by extending A_6-DNA with an extra base pair at each terminus ($\tau_m \sim 6$ ns), and 20% glycerol. We also examined a $(CA)_3$-DNA(14) construct that replaces the A-tract with a run of CA/TG steps, and contains the same sequence as the control E-DNA^{CtrG}. It is interesting to note that the terminal ends showed different profiles even in the absence of glycerol – the reduced normalized intensity for $(CA)_3$-DNA(14) and A_6-DNA(14) could be interpreted as either less terminal-end flexibility in the former or more rigidity in the interior of the latter. These explanations seem physically plausible, and could signify i) larger scale or faster oscillations at the satellite helices of A_6-tracts versus CA-repeats as a result of bending motions and/or lower thermodynamic stability with lower GC content in A_6-DNA(14), ii) higher stiffness or slower oscillations of A_6-tracts versus a CA-repeats, or iii) both. Considering results for E-DNA constructs that show negligible differences in fast dynamics at the satellite helices (Domain I and II in Figure 2.2) between A_6-tract and

CA$_3$ containing E-DNAs[84] we are more inclined towards ii. Further support for A-tract stiffness that is provided by S^2_{rel} data for A$_6$-DNA, discussed above.

With ~ 20 % glycerol, we obtained up to 1.5X larger fractional increase in the intensity of both terminal nucleotides and some internal Cs (C15) for A$_6$-DNA(14) as compared to A$_6$-DNA (Figure 2.6). Also, the fractional increase by adding glycerol to the 14 base pair samples was comparable for terminal residues between A$_6$-DNA(14) and (CA)$_3$-DNA(14), yet C15 next to the A$_6$-tract experienced much larger glycerol-dependent linewidth narrowing than C15 adjacent to the CA-repeat (Figure 2.6b). This is the first evidence the A$_6$-tract induces faster and/or higher-amplitude ns motions, which
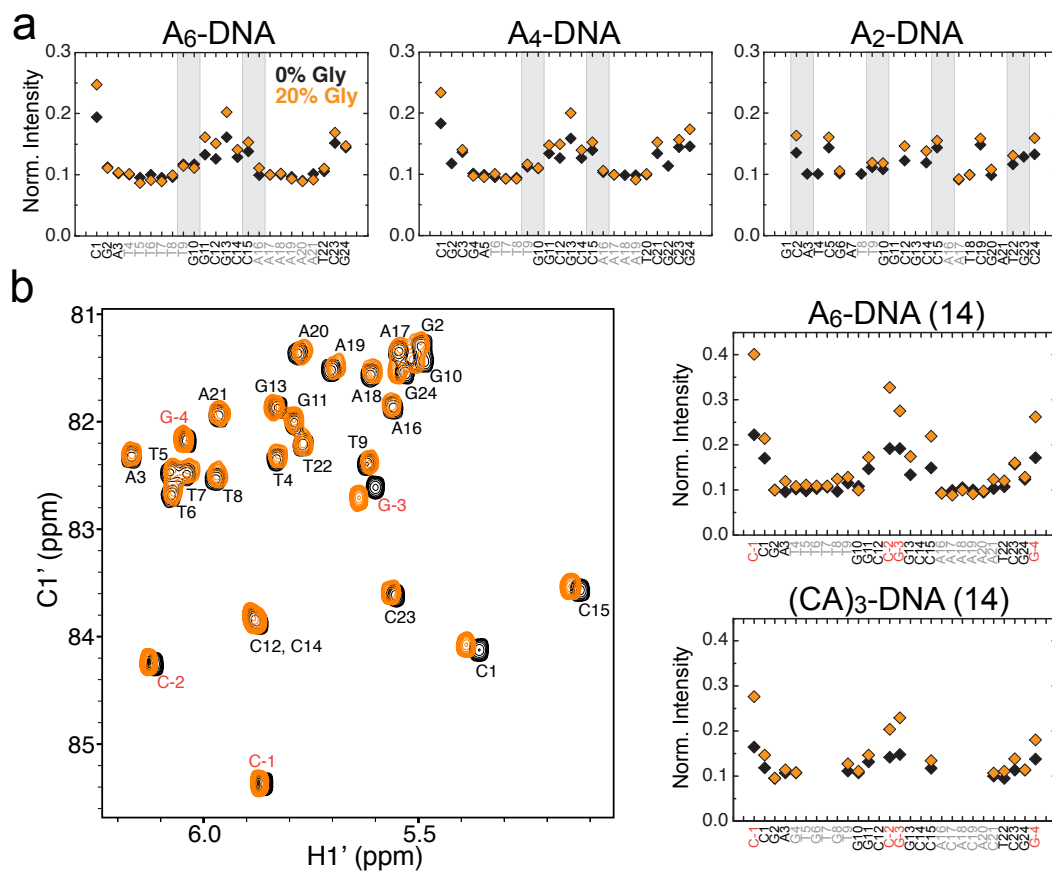
Figure 2.6: Decoupling **ns** motions in short DNA by glycerol. (a) Normalized resonances intensities for C1'H1' in A$_6$-DNA, A$_4$-DNA, and A$_2$-DNA collected at 26°C in the presence (orange) and absence (black) of glycerol. (b) Normalized resonance intensities for A$_6$-DNA(14) and CA$_3$-DNA(14) with and without glycerol, and a C1'H1' HSQC spectral overlay of A$_6$-DNA(14), showing negligible chemical shift perturbation upon addition of glycerol (residues highlighted in red are the two added base pairs to A$_6$-DNA).

went undetected in our intensity as well as ^{13}C relaxation analysis of short DNA dodecamers in the absence of glycerol. Therefore, we speculate that the A-tract size could modulate, in particular cases, both the timescales and amplitudes of pre-existing rapid motions at 5' CA/TG steps and possibly other junction types, which would be the subject of follow-up investigations.

2.4 Conclusion

In conclusion, we have implemented an approach for preparation, assignment, and analysis of dynamics in elongated DNA using NMR. Our results establish the ability to quantitatively study the dynamics of DNA, which are significantly larger and more anisotropic than customarily studied by NMR relaxation methods. We have detected rapid ns motions that are manifested and enhanced when the residues approach the terminal ends and that likely can be explained due to a combination of internal position- and sequence-specific motions and possibly coupled backbone deformations. While we do not observe any concrete evidence for ns bending dynamics, our results uncover deep penetrating ns motions, which seem to be independent of the A-tract at physiological temperatures. These effects are not observed in non-elongated DNA samples (~ 14 base pair long), indicating that the motions detected here occur at timescales greater than ~ 6 ns. However, we were able to resolve similar pre-existing ns motions near terminal ends of short DNA duplexes by adding glycerol that selectively slows down overall tumbling just enough to make contributions from internal motions manifest into spectral linewidths. We do not see these motions transmit further than three base pairs into the helix, perhaps because i) we have not fully decoupled global and internal motions, ii) the overall flexibility of the short duplex is already higher and masks the extent of internal motions and iii) the rapid bending modes are dependent on duplex length and sustained better by longer DNA duplexes. In general, they comprise an interesting finding that could have implications, for instance, in the sensing and trapping of thermally-driven transient fluctuations by DNA unwinding helicases at replication forks or other DNA end-interacting proteins[85]. The observation of such gradual deep-penetrating end effects, which are thought to be localized at the first two base pairs, highlights the kinds of new dynamical insights that can be obtained from the new NMR domain elongation.

Our examination of dynamics inside and at the junctions of variable A-tracts revealed several interesting motions trends that might correlate with the unusual conformation that A-tracts tend to adopt. Most notably, we discovered higher degree of stiffness in the interior of longer A-tracts, where a subtle but noticeable inverse correlation between base and sugar internal mobility was evident. These findings comply well with the commonly accepted notion that AA steps, together with AT steps, are the most rigid and structurally invariable[73-75,86]. AA steps can achieve high propeller twisting that enables them to stack quite optimally and allows for a hydration spine to form in the compressed minor groove, that ultimately confines their ability to wobble around. Moreover, we discover that A-tract length can modulate the flexibility of internal base pairs and bracketing sequences that could stem from variable conformation, hydration and stability of longer A-tracts. Last but not least, we find pronounced ps-ns disorder at the pyrimidine (C and T) sugar but not nucleobase of CA/TG and CG dinucleotide steps that indicate high torsional backbone fluctuations. A similar observation in the cytosine only of CG, CA and CT steps have been made by other groups[19,25,78,87,88] and could be associated with sugar pucker inter-conversions by correlation of MD simulations and NMR order parameters[25]. Indeed, the events that underlie these molecular transitions are difficult to probe solely by NMR, which the development of motional models can benefit tremendously from NMR-informed state-of-the-art computations of molecular dynamics to elucidate the nature of DNA flexibility, as has been demonstrated for canonical[25,89], non-canonical[90] and damaged DNA[91].

In a biological context, fast backbone dynamics can ultimately facilitate recognition of specific DNA sequences by their target ligands and protein binders such as gene regulators, restriction endonucleases or methylases. Such studies by the groups and Schleucher[25], Drobny[92] and Varani[23,24] have made significant progress in understanding the sugar-backbone dynamics of AT-rich *EcoR*I and CG-rich *Hha*I methyltransferase target sites, and the impact of methylation on backbone flexibility, which might play a role in protein recognition. CA steps are also targeted specifically by many biological factors, such the ubiquitous and gene-regulating CAP[93,94] and p53[95] proteins that are known to induce large deformations or trap non-canonical base-pairing conformations, that can sense the dynamic flexibility and preferentially localize at those sites.

2.5 References

1. Travers, A.A. The structural basis of DNA flexibility. *Philos Transact A Math Phys Eng Sci* **362**, 1423-38 (2004).
2. Garcia, H.G. et al. Biological consequences of tightly bent DNA: the other life of a macromolecular celebrity. *Biopolymers* **85**, 115-30 (2007).
3. Gimenes, F., Gouveia Fde, S., Fiorini, A. & Fernandez, M.A. Intrinsic bent DNA sites in the chromosomal replication origin of Xylella fastidiosa 9a5c. *Braz J Med Biol Res* **41**, 295-304 (2008).
4. Shajani, Z. & Varani, G. NMR studies of dynamics in RNA and DNA by 13C relaxation. *Biopolymers* **86**, 348-59 (2007).
5. Fujii, S., Kono, H., Takenaka, S., Go, N. & Sarai, A. Sequence-dependent DNA deformability studied using molecular dynamics simulations. *Nucleic Acids Res* **35**, 6063-74 (2007).
6. Segal, E. et al. A genomic code for nucleosome positioning. *Nature* **442**, 772-8 (2006).
7. Fernandez, A.G. & Anderson, J.N. Nucleosome positioning determinants. *J Mol Biol* **371**, 649-68 (2007).
8. Kaplan, N. et al. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**, 362-6 (2009).
9. Segal, E. & Widom, J. Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr Opin Struct Biol* **19**, 65-71 (2009).
10. Fuxreiter, M., Luo, N., Jedlovszky, P., Simon, I. & Osman, R. Role of base flipping in specific recognition of damaged DNA by repair enzymes. *J Mol Biol* **323**, 823-34 (2002).
11. Isaacs, R.J. & Spielmann, H.P. A model for initial DNA lesion recognition by NER and MMR based on local conformational flexibility. *DNA Repair (Amst)* **3**, 455-64 (2004).
12. Stivers, J.T. Extrahelical damaged base recognition by DNA glycosylase enzymes. *Chemistry* **14**, 786-93 (2008).
13. Palmer, A.G., 3rd, Kroenke, C.D. & Loria, J.P. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol* **339**, 204-38 (2001).
14. Palmer, A.G., 3rd & Massi, F. Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chem Rev* **106**, 1700-19 (2006).
15. Mittermaier, A. & Kay, L.E. New tools provide new insights in NMR studies of protein dynamics. *Science* **312**, 224-8 (2006).
16. Prestegard, J.H., Bougault, C.M. & Kishore, A.I. Residual dipolar couplings in structure determination of biomolecules. *Chem Rev* **104**, 3519-40 (2004).
17. Bax, A. & Grishaev, A. Weak alignment NMR: a hawk-eyed view of biomolecular structure. *Curr Opin Struct Biol* **15**, 563-70 (2005).
18. Kojima, C., Ono, A., Kainosho, M. & James, T.L. DNA duplex dynamics: NMR relaxation studies of a decamer with uniformly 13C-labeled purine nucleotides. *J Magn Reson* **135**, 310-33 (1998).

19. Spielmann, H.P. Dynamics in psoralen-damaged DNA by 1H-detected natural abundance 13C NMR spectroscopy. *Biochemistry* **37**, 5426-38 (1998).

20. Isaacs, R.J. & Spielmann, H.P. Relationship of DNA structure to internal dynamics: correlation of helical parameters from NOE-based NMR solution structures of d(GCGTACGC)(2) and d(CGCTAGCG)(2) with (13)C order parameters implies conformational coupling in dinucleotide units. *J Mol Biol* **307**, 525-40 (2001).

21. Ying, J., Grishaev, A., Bryce, D.L. & Bax, A. Chemical shift tensors of protonated base carbons in helical RNA and DNA from NMR relaxation and liquid crystal measurements. *J Am Chem Soc* **128**, 11443-54 (2006).

22. Shajani, Z. & Varani, G. 13C NMR relaxation studies of RNA base and ribose nuclei reveal a complex pattern of motions in the RNA binding site for human U1A protein. *J Mol Biol* **349**, 699-715 (2005).

23. Echodu, D. et al. Furanose dynamics in the HhaI methyltransferase target DNA studied by solution and solid-state NMR relaxation. *J Phys Chem B* **112**, 13934-44 (2008).

24. Shajani, Z. & Varani, G. 13C relaxation studies of the DNA target sequence for hhai methyltransferase reveal unique motional properties. *Biochemistry* **47**, 7617-25 (2008).

25. Duchardt, E., Nilsson, L. & Schleucher, J. Cytosine ribose flexibility in DNA: a combined NMR 13C spin relaxation and molecular dynamics simulation study. *Nucleic Acids Res* **36**, 4211-9 (2008).

26. Wu, Z., Delaglio, F., Tjandra, N., Zhurkin, V.B. & Bax, A. Overall structure and sugar dynamics of a DNA dodecamer from homo- and heteronuclear dipolar couplings and 31P chemical shift anisotropy. *J Biomol NMR* **26**, 297-315 (2003).

27. Schwieters, C.D. & Clore, G.M. A physical picture of atomic motions within the Dickerson DNA dodecamer in solution derived from joint ensemble refinement against NMR and large-angle X-ray scattering data. *Biochemistry* **46**, 1152-66 (2007).

28. Showalter, S.A., Baker, N.A., Tang, C. & Hall, K.B. Iron responsive element RNA flexibility described by NMR and isotropic reorientational eigenmode dynamics. *J Biomol NMR* **32**, 179-93 (2005).

29. Zhang, Q., Throolin, R., Pitt, S.W., Serganov, A. & Al-Hashimi, H.M. Probing motions between equivalent RNA domains using magnetic field induced residual dipolar couplings: accounting for correlations between motions and alignment. *J Am Chem Soc* **125**, 10530-1 (2003).

30. Zhang, Q., Sun, X., Watt, E.D. & Al-Hashimi, H.M. Resolving the motional modes that code for RNA adaptation. *Science* **311**, 653-6 (2006).

31. Zhang, Q., Stelzer, A.C., Fisher, C.K. & Al-Hashimi, H.M. Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* **450**, 1263-7 (2007).

32. Hansen, A.L. & Al-Hashimi, H.M. Dynamics of large elongated RNA by NMR carbon relaxation. *J Am Chem Soc* **129**, 16072-82 (2007).

33. Lipari, G. & Szabo, A. Nuclear magnetic resonance relaxation in nucleic acid fragments: models for internal motion. *Biochemistry* **20**, 6250-6 (1981).

34. Tolman, J.R., Flanagan, J.M., Kennedy, M.A. & Prestegard, J.H. NMR evidence for slow collective motions in cyanometmyoglobin. *Nat Struct Biol* **4**, 292-7 (1997).

35. Tolman, J.R., Al-Hashimi, H.M., Kay, L.E. & Prestegard, J.H. Structural and dynamic analysis of residual dipolar coupling data for proteins. *J Am Chem Soc* **123**, 1416-24 (2001).

36. Zhang, Q. & Al-Hashimi, H.M. Extending the NMR spatial resolution limit for RNA by motional couplings. *Nat Methods* **5**, 243-5 (2008).

37. Haran, T.E. & Mohanty, U. The unique structure of A-tracts and intrinsic DNA bending. *Q Rev Biophys* **42**, 41-81 (2009).

38. Beveridge, D.L., Dixit, S.B., Barreiro, G. & Thayer, K.M. Molecular dynamics simulations of DNA curvature and flexibility: helix phasing and premelting. *Biopolymers* **73**, 380-403 (2004).

39. Zimmer, D.P. & Crothers, D.M. NMR of enzymatically synthesized uniformly 13C15N-labeled DNA oligonucleotides. *Proc Natl Acad Sci U S A* **92**, 3091-5 (1995).

40. Korzhnev, D.M., Orekhov, V.Y. & Kay, L.E. Off-resonance R(1rho) NMR studies of exchange dynamics in proteins with low spin-lock fields: an application to a Fyn SH3 domain. *J Am Chem Soc* **127**, 713-21 (2005).

41. Hansen, A.L., Nikolova, E.N., Casiano-Negroni, A. & Al-Hashimi, H.M. Extending the range of microsecond-to-millisecond chemical exchange detected in labeled and unlabeled nucleic acids by selective carbon R(1rho) NMR spectroscopy. *J Am Chem Soc* **131**, 3818-9 (2009).

42. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-93 (1995).

43. Mulder, F.A.A., de Graaf, R.A., Kaptein, R. & Boelens, R. An Off-resonance Rotating Frame Relaxation Experiment for the Investigation of Macromolecular Dynamics Using Adiabatic Rotations. *J Magn Reson* **131**, 351-7 (1998).

44. Korzhnev, D.M., Skrynnikov, N.R., Millet, O., Torchia, D.A. & Kay, L.E. An NMR experiment for the accurate measurement of heteronuclear spin-lock relaxation rates. *J Am Chem Soc* **124**, 10743-53 (2002).

45. Dethoff, E.A. et al. Characterizing complex dynamics in the transactivation response element apical loop and motional correlations with the bulge by NMR, molecular dynamics, and mutagenesis. *Biophys J* **95**, 3906-15 (2008).

46. MacDonald, D., Herbert, K., Zhang, X., Pologruto, T. & Lu, P. Solution structure of an A-tract DNA bend. *J Mol Biol* **306**, 1081-98 (2001).

47. Garcia de la Torre, J., Huertas, M.L. & Carrasco, B. HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J Magn Reson* **147**, 138-46 (2000).

48. Marini, J.C., Levene, S.D., Crothers, D.M. & Englund, P.T. Bent helical structure in kinetoplast DNA. *Proc Natl Acad Sci U S A* **79**, 7664-7668 (1982).

49. Wu, H.M. & Crothers, D.M. The locus of sequence-directed and protein-induced DNA bending. *Nature* **308**, 509-13 (1984).

50. Chen, X. et al. A PCR-based method for uniform 13C/15N labeling of long DNA oligomers. *FEBS Lett* **436**, 372-6 (1998).

51. Louis, J.M., Martin, R.G., Clore, G.M. & Gronenborn, A.M. Preparation of uniformly isotope-labeled DNA oligonucleotides for NMR spectroscopy. *J Biol Chem* **273**, 2374-8 (1998).

52. Yan, J. & Bushweller, J.H. An optimized PCR-based procedure for production of 13C/15N-labeled DNA. *Biochem Biophys Res Commun* **284**, 295-300 (2001).

53. Masse, J.E., Bortmann, P., Dieckmann, T. & Feigon, J. Simple, efficient protocol for enzymatic synthesis of uniformly 13C, 15N-labeled DNA for heteronuclear NMR studies. *Nucleic Acids Res* **26**, 2618-24 (1998).

54. Pelupessy, P., Chiarparin, E. & Bodenhausen, G. Excitation of selected proton signals in NMR of isotopically labeled macromolecules. *J Magn Reson* **138**, 178-81 (1999).

55. Getz, M.M., Andrews, A.J., Fierke, C.A. & Al-Hashimi, H.M. Structural plasticity and Mg2+ binding properties of RNase P P4 from combined analysis of NMR residual dipolar couplings and motionally decoupled spin relaxation. *RNA* **13**, 251-66 (2007).

56. Sun, X., Zhang, Q. & Al-Hashimi, H.M. Resolving fast and slow motions in the internal loop containing stem-loop 1 of HIV-1 that are modulated by Mg2+ binding: role in the kissing-duplex structural transition. *Nucleic Acids Res* **35**, 1698-713 (2007).

57. Smith, A.L. et al. Theory for spin-lattice relaxation of spin probes on weakly deformable DNA. *J Phys Chem B* **112**, 9219-36 (2008).

58. Marini, J.C. et al. Physical characterization of a kinetoplast DNA fragment with unusual properties. *J Biol Chem* **259**, 8974-9 (1984).

59. Koo, H.S., Wu, H.M. & Crothers, D.M. DNA bending at adenine . thymine tracts. *Nature* **320**, 501-6 (1986).

60. Levene, S.D., Wu, H.M. & Crothers, D.M. Bending and flexibility of kinetoplast DNA. *Biochemistry* **25**, 3988-95 (1986).

61. Diekmann, S. Temperature and salt dependence of the gel migration anomaly of curved DNA fragments. *Nucleic Acids Res* **15**, 247-65 (1987).

62. Herrera, J.E. & Chaires, J.B. A premelting conformational transition in poly(dA)-Poly(dT) coupled to daunomycin binding. *Biochemistry* **28**, 1993-2000 (1989).

63. Chan, S.S. et al. Physical studies of DNA premelting equilibria in duplexes with and without homo dA.dT tracts: correlations with DNA bending. *Biochemistry* **29**, 6161-71 (1990).

64. Toth, K., Sauermann, V. & Langowski, J. DNA curvature in solution measured by fluorescence resonance energy transfer. *Biochemistry* **37**, 8173-9 (1998).

65. Stellwagen, N.C., Magnusdottir, S., Gelfi, C. & Righetti, P.G. Preferential counterion binding to A-tract DNA oligomers. *J Mol Biol* **305**, 1025-33 (2001).

66. Hud, N.V. & Feigon, J. Characterization of divalent cation localization in the minor groove of the A(n)T(n) and T(n)A(n) DNA sequence elements by (1)H NMR spectroscopy and manganese(II). *Biochemistry* **41**, 9900-10 (2002).

67. Jerkovic, B. & Bolton, P.H. Magnesium increases the curvature of duplex DNA that contains dA tracts. *Biochemistry* **40**, 9406-11 (2001).

68. Tchernaenko, V., Halvorson, H.R. & Lutter, L.C. Topological measurement of an A-tract bend angle: effect of magnesium. *J Mol Biol* **341**, 55-63 (2004).

69. Kintanar, A., Klevit, R.E. & Reid, B.R. Two-dimensional NMR investigation of a bent DNA fragment: assignment of the proton resonances and preliminary structure analysis. *Nucleic Acids Res* **15**, 5845-62 (1987).

70. Katahira, M. et al. One- and two-dimensional NMR studies on the conformation of DNA containing the oligo(dA)oligo(dT) tract. *Nucleic Acids Res* **16**, 8619-32 (1988).

71. Barbic, A., Zimmer, D.P. & Crothers, D.M. Structural origins of adenine-tract bending. *Proc Natl Acad Sci U S A* **100**, 2369-73 (2003).

72. Tjandra, N., Cowburn, D. & Fushman, D. An approach to direct determination of protein dynamics from N-15 NMR relaxation at multiple fields, independent of variable N-15 chemical shift anisotropy and chemical exchange contributions. *Journal of the American Chemical Society* **121**, 8577-8582 (1999).

73. Nelson, H.C., Finch, J.T., Luisi, B.F. & Klug, A. The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature* **330**, 221-6 (1987).

74. Hunter, C.A. Sequence-dependent DNA structure. The role of base stacking interactions. *J Mol Biol* **230**, 1025-54 (1993).

75. El Hassan, M.A. & Calladine, C.R. Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA. *J Mol Biol* **259**, 95-103 (1996).

76. Levitt, M. How many base-pairs per turn does DNA have in solution and in chromatin? Some theoretical calculations. *Proc Natl Acad Sci U S A* **75**, 640-4 (1978).

77. Drew, H.R. & Dickerson, R.E. Structure of a B-DNA dodecamer. III. Geometry of hydration. *J Mol Biol* **151**, 535-56 (1981).

78. Paquet, F., Gaudin, F. & Lancelot, G. Selectively 13C-enriched DNA: evidence from 13C1' relaxation rate measurements of an internal dynamics sequence effect in the lac operator. *J Biomol NMR* **8**, 252-60 (1996).

79. Isaacs, R.J. & Spielmann, H.P. NMR evidence for mechanical coupling of phosphate B(I)-B(II) transitions with deoxyribose conformational exchange in DNA. *J Mol Biol* **311**, 149-60 (2001).

80. LaPlante, S.R., Zanatta, N., Hakkinen, A., Wang, A.H. & Borer, P.N. 13C-NMR of the deoxyribose sugars in four DNA oligonucleotide duplexes: assignment and structural features. *Biochemistry* **33**, 2430-40 (1994).

81. Foloppe, N. & MacKerell, A.D., Jr. Intrinsic conformational properties of deoxyribonucleosides: implicated role for cytosine in the equilibrium among the A, B, and Z forms of DNA. *Biophys J* **76**, 3206-18 (1999).

82. Madhumalar, A. & Bansal, M. Sequence preference for BI/BII conformations in DNA: MD and crystal structure data analysis. *J Biomol Struct Dyn* **23**, 13-27 (2005).

83. Giudice, E., Varnai, P. & Lavery, R. Base pair opening within B-DNA: free energy pathways for GC and AT pairs from umbrella sampling simulations. *Nucleic Acids Res* **31**, 1434-43 (2003).

84. Nikolova, E.N. & Al-Hashimi, H.M. Preparation, resonance assignment, and preliminary dynamics characterization of residue specific 13C/15N-labeled elongated DNA for the study of sequence-directed dynamics by NMR. *J Biomol NMR* **45**, 9-16 (2009).

85. Delagoutte, E. & von Hippel, P.H. Helicase mechanisms and the coupling of helicases within macromolecular machines. Part I: Structures and properties of isolated helicases. *Q Rev Biophys* **35**, 431-78 (2002).

86. Packer, M.J., Dauncey, M.P. & Hunter, C.A. Sequence-dependent DNA structure: dinucleotide conformational maps. *J Mol Biol* **295**, 71-83 (2000).

87. Borer, P.N. et al. 13C-NMR relaxation in three DNA oligonucleotide duplexes: model-free analysis of internal and overall motion. *Biochemistry* **33**, 2441-50 (1994).

88. Isaacs, R.J., Rayens, W.S. & Spielmann, H.P. Structural differences in the NOE-derived structure of G-T mismatched DNA relative to normal DNA are correlated with differences in (13)C relaxation-based internal dynamics. *J Mol Biol* **319**, 191-207 (2002).

89. Nikolova, E.N. et al. Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **470**, 498-502 (2011).

90. Isaacs, R.J. & Spielmann, H.P. Insight into G[bond]T mismatch recognition using molecular dynamics with time-averaged restraints derived from NMR spectroscopy. *J Am Chem Soc* **126**, 583-90 (2004).

91. Chen, J., Dupradeau, F.Y., Case, D.A., Turner, C.J. & Stubbe, J. DNA oligonucleotides with A, T, G or C opposite an abasic site: structure and dynamics. *Nucleic Acids Res* **36**, 253-62 (2008).

92. Pederson, K., Meints, G.A., Shajani, Z., Miller, P.A. & Drobny, G.P. Backbone dynamics in the DNA HhaI protein binding site. *J Am Chem Soc* **130**, 9072-9 (2008).

93. Parkinson, G. et al. Structure of the CAP-DNA complex at 2.5 angstroms resolution: a complete picture of the protein-DNA interface. *J Mol Biol* **260**, 395-408 (1996).

94. Chen, S., Vojtechovsky, J., Parkinson, G.N., Ebright, R.H. & Berman, H.M. Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: DNA binding specificity based on energetics of DNA kinking. *J Mol Biol* **314**, 63-74 (2001).

95. Kitayner, M. et al. Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs. *Nat Struct Mol Biol* **17**, 423-9 (2010).

Chapter 3

Probing micro- to millisecond motions in canonical duplex DNA:
Transient Hoogsteen base pairs

3.1 Introduction

Soon after its discovery[1], it was recognized that the DNA double helix could accommodate a range of conformations that retain Watson-Crick (WC) base pairing[3]. Sequence-directed variations in duplex DNA structure, shape, and flexibility have since been shown to play fundamental roles in biology including in the indirect readout of DNA sequences by recognition factors[4,5], nucleosome positioning[5,6], and formation of loops and other large-scale architectures[7] involved in DNA packaging, replication, transcription, and recombination. DNA duplexes resiliently maintain WC base pairing even when supercoiled and wrapped around histone octamers in nucleosomes[8] and when adopting left-handed double-helical conformations known as Z-DNA[9]. Deviations from the WC base pairing have so far only been observed in duplex DNA bound to proteins[10,11] and small molecule ligands[12,13] and in the context of damaged DNA[14,15], but never within naked canonical B-DNA duplexes.

Thus far, atomic resolution structural studies of the iconic DNA double helix have exclusively focused on its dominant, experimentally accessible, ground state conformation. Far less is known about other low-energy DNA conformations that may be sampled only transiently in solution. NMR relaxation dispersion experiments[16,17] provide a rare opportunity to detect and characterize such short-lived (<5 ms) and low-populated species (>0.1%), often referred to as "excited states". This methodology has been widely used to characterize protein excited states that have been implicated in folding[16,17], recognition[18], and catalysis[19], culminating in the recent structure determination of a transient protein-folding intermediate[20]. Recent advances in carbon-based relaxation dispersion experiments combined with selective labeling schemes have addressed

limitations that have hindered application of this methodology to nucleic acids, allowing detection of excited states in both RNA[21,22] and DNA[22,23]. However, the structures of nucleic acid excited states remain elusive.

Here, we continue our investigations of sequence-dependent dynamics in DNA duplex dodecamers (Chapter 2) in search for transient states that form on µs-ms timescales. Such chemical exchange dynamics occurring on timescales slower than overall tumbling tend to modulate transverse relaxation rates and resonance linewidths, and, under particular conditions, manifest themselves as reduced resonance intensities. The lower intensity profile for adenines in 5' CA/TG A-tract junctions had suggested the presence of chemical exchange in DNA duplexes. Previously, slow mobility has been observed for TA[24-26] and CA[27] steps and a G•G mismatch base pair[28] embedded in short duplexes using line broadening and $^1H/^{13}C$ relaxation dispersion on fractionally labeled DNA. However, those experiments cannot afford accurate quantitation and timescale sensitivity to characterize thoroughly these motions estimated to occur with exchange times of hundreds of microseconds. In order to probe for chemical exchange and how altering the A-tract length influences it, we develop suitable NMR ^{13}C relaxation dispersion experiments[22] and apply them towards characterization of excited states in DNA[29i]. In the process, we devise a new strategy for determining the conformations of excited states that relies on (i) matching the experimentally deduced kinetic-thermodynamic profiles for the transition with those obtained from computational methods, (ii) trapping the proposed ground and excited states by chemical modification, nucleotide mutations or ligand binding, (iii) comparing directly the chemical shifts deduced by relaxation dispersion with theoretical predictions of proposed excited states.

3.2 Materials and Methods

3.2.1 Preparation and resonance assignment of unlabeled and $^{13}C/^{15}N$-labeled DNA

Unlabeled DNA oligonucleotides were purchased from IDT, Inc. (A_2-DNA, A_4-DNA, A_6-DNA, and E-DNA), Exiqon A/S (A_6-DNA^{A16LNA} and A_2-DNA^{A16LNA}) and

---

[i] Hansen, A.L., Nikolova, E.N., Casiano-Negroni, A. & Al-Hashimi, H.M. Extending the range of microsecond-to-millisecond chemical exchange detected in labeled and unlabeled nucleic acids by selective carbon R(1rho) NMR spectroscopy. *J Am Chem Soc* **131**, 3818-9 (2009); Nikolova, E.N. et al. Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **470**, 498-502 (2011).

Midland Certified, Inc. (A$_6$-DNA^{1mA16}, A$_6$-DNA^{1mG10}, A$_6$-DNA^{7dNA16}, and A$_6$-DNA^{7dNG10}). Unlabeled and ^{13}C/^{15}N-labeled DNA oligos and duplexes were prepared and assigned as described in Section 2.2.1 using sample-specific buffer conditions. DNA samples were exchanged into the desired buffer by using Amicon Ultra-4 centrifugal filters (3 kDa cutoff) 3-5X. The desired pH was adjusted by buffer exchange as above or by directly titrating a dilute HCl/NaOH solution, and in either case the pH was carefully checked using a pH meter (Thermo Scientific). Unlabeled DNA samples were annealed at 2.0-4.0 mM concentrations. Labeled DNA samples were prepared by resuspending/exchanging the single strands into the desired buffer, titrating one strand directly into the other and monitoring the disappearance of single stranded DNA using fast HSQCs. The 2:1 complex between echinomycin (Selleck Chemicals) and E-DNA was prepared as previously described[30]. Briefly, echinomycin was resuspended in 100% methanol, added at a 2:1 ratio to E-DNA in NMR buffer, and the excess methanol evaporated under a stream of N$_2$, ensuring effective binding of echinomycin to DNA. All NMR experiments were performed on a Bruker Avance 600 MHz NMR spectrometer equipped with a 5mm triple-resonance cryogenic probe.

3.2.2. Selective ^{13}C $R_{1\rho}$ relaxation dispersion

Rotating frame ($R_{1\rho}$) relaxation dispersion profiles were measured at a single field (14.1 T) using a selective carbon experiment with a 1D acquisition scheme[22] that extends the sensitivity to chemical exchange into millisecond timescales relative to conventional 2D relaxation dispersion methods. On-resonance data were recorded at variable (100 to 3500 Hz) effective spin-lock field strength (ω_{eff}) (Appendix 1) for various sites in ^{13}C/^{15}N-labeled and unlabeled DNA constructs. For ^{13}C/^{15}N-labeled DNA samples and E-DNA octamer, off-resonance dispersion data were collected in a temperature dependent manner at various spinlock offset frequencies (Ω_{eff}) and at three to four different spinlock powers (ω)(Appendix 1). In each case, the following relaxation delays were used: {0, 4, 8, 12 (2X), 16, 20, 26, 32 (2X) ms} for C2/C6/C8 and {0, 4, 8, 12 (2X), 18, 26, 34, 42 (2X) ms} for C1' in ^{13}C/^{15}N-labeled DNA constructs; {0, 40(2X) ms} for C8 and {0, 48(2X) ms} for C1' at 17 °C, {0, 42(2X) ms} for C8 (17.0 °C) and {0, 60(2X) ms} for C1' at 26 °C in E-DNA; {0, 30(2X) ms} for C8 (17.0 °C) in A$_6$-

DNA^{A16LNA}, {0, 32(2X) ms} for C8 and {0, 42(2X) ms} for C1' at 26 °C in A$_6$-DNA^{7dNA16} and A$_6$-DNA^{7dNG10}. Data points corresponding to Hartmann-Hahn matching conditions were omitted from the data fits as previously described[22]. Data were processed using nmrPipe[31] and $R_{1\rho}$ relaxation rate constants were computed by fitting the resonance intensities as a function of spinlock time to monoexponential decays using Mathematica 6.0 (Wolfram Research, Inc., Champaign, IL). Errors in $R_{1\rho}$ were estimated from duplicate measurements and errors in R_{2eff} ($R_{2eff} = R_2 + R_{ex}$) were calculated by error propagation using Eq. 3.1 and assuming 1 % error in spinlock power.
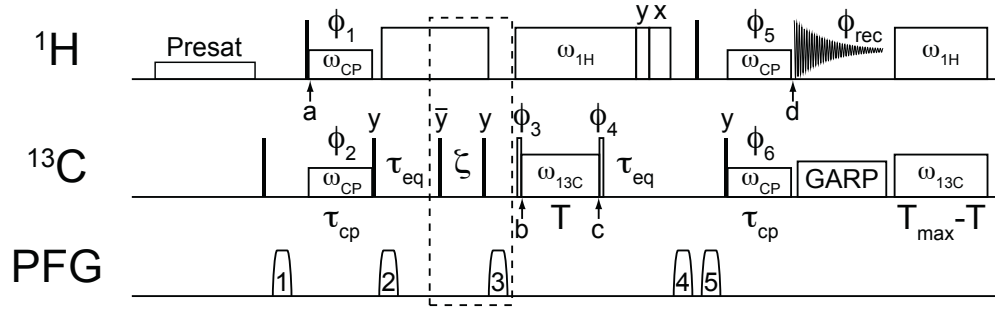
Figure 3.1: Off-resonance $R_{1\rho}$ ^{13}C pulse sequence for quantifying μs‑ms chemical exchange in nucleic acids. Solid narrow bars represent 90° pulses, open narrow bars represent 13C pulses applied at a tip angle of $\theta = \mathrm{arccot}(\Omega / \omega)$, where ω and Ω are the spinlock power and resonance offset from the spinlock respectively. Open rectangles depict blocks of continuous wave irradiation including ^{1}H and ^{13}C cross-polarization (ω_{CP}), ^{13}C $R_{1\rho}$ spinlock period (ω_{13C}), ^{1}H decoupling period for suppression of ^{13}C-^{1}H DD/CSA cross-correlated relaxation, and water presaturation (ca. 10 Hz). Additional purge pulses (small open rectangles) are used at the end of the 1H decoupling for water suppression. A 3.5 kHz GARP and 1.0 kHz GARP sequence is used for decoupling during acquision on ^{13}C and ^{15}N channels respectively. Initially, the ^{1}H magnetization is selectively transferred to the directly bonded ^{13}C spin using matching ω_{CP} for a period of 1/J$_{CH}$, followed by a ^{13}C relaxation period with variable spinlock power and offset, and finally the magnetization is returned back using the same strategy for detection.
ζ

Measured relaxation dispersion profiles with on- and off-resonance data were fit by Eq.1 (below) that assumes a two-state equilibrium ($A \underset{kB}{\overset{kA}{\rightleftharpoons}} B$) with asymmetric population distributions ($p_A >> p_B$)[32] using Mathematica 6.0 (Wolfram Research, Inc., Champaign, IL)

$$R_{1\rho} = R_1 \cos^2\theta + R_2 \sin^2\theta + \sin^2\theta \frac{p_A p_B \Delta\omega_{AB}^2 k_{ex}}{\Omega_B^2 + \omega_1^2 + k_{ex}^2} \tag{3.1}$$

where R_1 and R_2 are the intrinsic longitudinal and transverse relaxation rates respectively (assumed to be identical for A and B species), Ω is the resonance offset from the spinlock carrier, ω_1 is the spinlock strength; $\theta = \arctan(\omega_1/\Omega_{ave})$, $\Delta\omega_{AB} = \Omega_B - \Omega_A$, $\Omega_{ave} = p_A\Omega_A +$

$p_B\Omega_B$, where p_A (p_B) is the major (minor) state fractional population ($p_A + p_B = 1$); $k_{ex} = k_A + k_B$ is the exchange rate constant for a two-state equilibrium, where $k_A = p_B k_{ex}$ and $k_B = p_A k_{ex}$ are the forward and reverse rate constants respectively. Similar results were obtained when fitting relaxation dispersion profiles against more complex two-state exchange models including the Laguerre approximation[33] that do not assume a skewed population distribution (data not shown) and statistical analysis implied that application of the more complex model to extract exchange parameters here was not justified. Temperature-dependent data for base/sugar resonances within the same nucleotide or base pair were fit individually and globally with shared parameters (k_{ex} and p_B for each temperature) (Appendix 1). The best-fit model was assessed using F-test (data not shown), which uses chi-square (χ^2) and F-distribution analysis to determine the feasibility of a complicated model (i.e. individual fits) versus a more simple model (i.e. shared-parameter fits) nested inside the first model. The chemical shift difference $\Delta\omega_{AB}$ was assumed to be invariant over the narrow temperature range investigated. On-resonance $R_{1\rho}$ profiles were fit by a simplified two-state fast exchange expression[32] ($k_{ex} \gg \Delta\omega_{AB}$) with parameters are as above:

$$R_{1\rho} = R_2 + R_{ex} = R_2 + \frac{\Phi_{ex} k_{ex}}{\omega_1^2 + k_{ex}^2}; \Phi_{ex} = p_A p_B \Delta\omega_{AB}^2 \qquad (3.2)$$
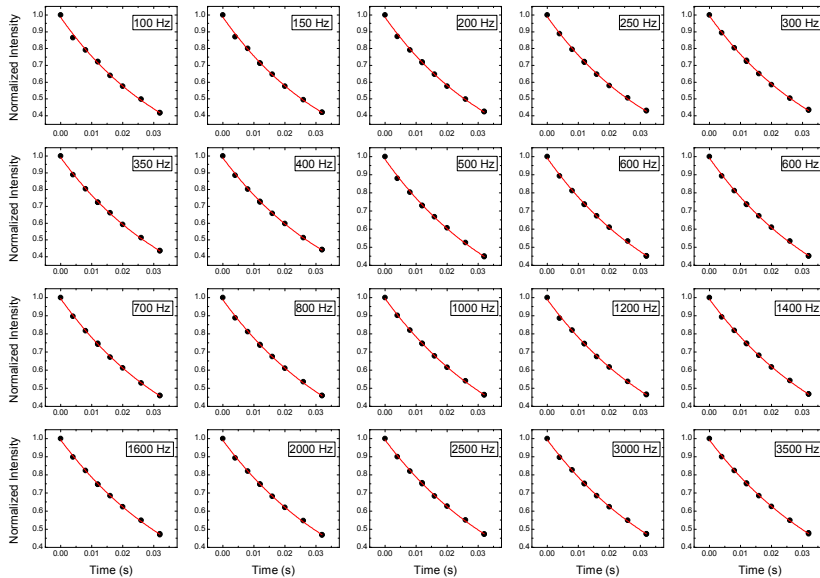
Figure 3.2: $R_{1\rho}$ relaxation decays. Representative on-resonance relaxation intensity profiles for G10 C8 in A_6-DNA at various spinlock powers (inset) and best fits to a monoexponential decay.

3.2.3. Thermodynamic Analysis

The observed temperature dependence of the forward (k_A) and reverse (k_B) rate constants was fit by a modified van't Hoff equation that accounts for statistical compensation effects and assumes a smooth energy surface[2]:

$$\ln\left(\frac{k_i(T)}{T}\right) = \ln\left(\frac{k_B \kappa}{h}\right) - \frac{\Delta G_i^T}{RT_{hm}} - \frac{\Delta H_i^T}{R}\left(\frac{1}{T} - \frac{1}{T_{hm}}\right)$$

(3.3)

where k_i (i = A, B) is the rate constant, ΔG_i^T and ΔH_i^T are the free energy and enthalpy of activation respectively, R is the universal gas constant, T is temperature, and T_{hm} is the harmonic mean of the experimental temperatures computed as $T_{hm} = n / \sum_{i=1}^{n} 1/T_i$; k_B is Boltzmann's constant, h is Plank's constant, κ is the transmission coefficient (assumed to be 1) in the pre-exponential factor of Eyring's theory. The entropy of activation (ΔS_i^T) was calculated as $T\Delta S_i^T = \Delta H_i^T - \Delta G_i^T (T_{hm})$ (3.4). Semi-logarithmic plots are included in Figure 3.4 and best-fit thermodynamic parameters are reported in Table 3.1.

Table 3.1: Thermodynamic parameters for WC-to-HG transition. Data was obtained from best fits of experimentally derived rate constants (k_A, k_B) for a two-site chemical exchange to Eq. 3.1 (TST) or Eq. 3.5 (FL). The entropy terms for TST were calculated from Eq. 3.4.

Residue (spin)	Construct	Thermodynamic Parameters (kcal/mol)		
Transition State Theory (TST)		ΔG_A^T	ΔG_B^T	ΔG_{AB}
G10 (C8/C1')[a]	A$_6$-DNA	16.49±0.01	13.56±0.01	2.93±0.01
A16 (C8/C1')[a]	A$_6$-DNA	15.95±0.01	12.78±0.01	3.16±0.01
A16 (C8/C1')[b]	A$_6$-DNA	16.01 ±0.04	12.71±0.04	3.30±0.05
A16 (C1')[b]	A$_4$-DNA	16.08±0.04	12.76±0.04	3.33±0.06
A16 (C8)[b]*	A$_2$-DNA	*16.35±0.12*	*12.67±0.09*	*3.68±0.15*
A3 (C8/C1')[b]	A$_2$-DNA	15.66±0.06	12.41±0.03	3.25±0.06
		ΔH_A^T	ΔH_B^T	ΔH_{AB}
G10 (C8/C1')[a]	A$_6$-DNA	26.83±0.72	25.17±0.79	1.66±0.98
A16 (C8/C1')[a]	A$_6$-DNA	25.23±0.36	23.34±0.17	1.89±0.40
A16 (C8/C1')[b]	A$_6$-DNA	23.35±2.82	24.23±1.55	-0.88±3.22
A16 (C1')[b]	A$_4$-DNA	23.02±3.33	22.33±3.11	0.68±4.55
A16 (C8)[b]*	A$_2$-DNA	-	-	-
A3 (C8/C1')[b]	A$_2$-DNA	12.46±3.17	16.57±1.30	-4.11±3.42

		$T\Delta S_A^T$	$T\Delta S_B^T$	$T\Delta S_{AB}$
G10 (C8/C1')[a]	A_6-DNA	10.34±0.72	11.61±0.66	-1.27±0.98
A16 (C8/C1')[a]	A_6-DNA	9.29±0.36	10.56±0.17	-1.27±0.40
A16 (C8/C1')[b]	A_6-DNA	7.34±2.82	11.51±1.55	-4.17±3.22
A16 (C1')[b]	A_4-DNA	6.93±3.33	9.58±3.11	-2.65±4.55
A16 (C8)[b]*	A_2-DNA	-	-	-
A3 (C8/C1')[b]	A_2-DNA	-3.20±3.17	4.16±1.30	-7.36±3.42

Ferry Law (FL)		$< H_A^2 >^{1/2}$	$< H_B^2 >^{1/2}$	
G10 (C8/C1')[a]	A_6-DNA	2.86±0.01	2.67±0.11	
A16 (C8/C1')[a]	A_6-DNA	2.76±0.03	2.66±0.01	
A16 (C8/C1')[b]	A_6-DNA	2.65±0.15	2.70±0.08	
A16 (C1')[b]	A_4-DNA	2.64±0.18	2.60±0.17	
A16 (C8)[b]*	A_2-DNA	-	-	
A3 (C8/C1')[b]	A_2-DNA	1.96±0.23	2.24±0.08	

* *van't Hoff plots for A16 C8 were non-linear and only average ΔG parameters (italicized) are reported.*
[a] *Data reported for measurements at pH 5.4.*
[b] *Data reported for measurements at pH 6.8.*

An alternative interpretation of the thermodynamic parameters is given by the phenomenological Ferry law[34,35], which incorporates a lower energy barrier with a rough enthalpic surface:

$$\ln\left(k_i\right) = \ln C - \frac{\Delta H_i^T}{RT} - \frac{\left\langle H_i^2\right\rangle}{(RT)^2}$$

(3.5)

where C is a constant, where $\left\langle H_i^2\right\rangle^{1/2}$ represents the enthalpy due to ruggedness of the

barrier. The maximum $\left\langle H_i^2\right\rangle^{1/2}$ values were calculated by taking the smooth Arrhenius-

like enthalpic barrier to be vanishing ($\Delta H_i^T = 0$) and are reported in Table 3.1.

3.2.4 Molecular dynamics simulations of WC-to-HG base pair transition pathways

(collaboration with Dr. Eunae Kim and Prof. Ioan Andricioaei from the University of California, Irvine)

An initial duplex DNA in standard B-form with Watson-Crick (WC) base pairing corresponding to A_6-DNA was generated using 3DNA[36]. Hoogsteen (HG) base pairs

were generated at A16•T9, A3•T22, and G10•C15 positions using known X-ray coordinates[37,38], where the purine adopts a *syn* conformation, while the complementary pyrimidine retains an *anti* conformation. Initially, all conformers with WC or HG base pairing were equilibrated through a series of energy minimizations using the Adopted Basis Newton-Raphson (ABNR) algorithm. Canonical (NVT) ensemble MD were performed with the CHARMM27 all-atom force field[39] and the generalized Born molecular volume (GBMV) implicit solvation model (GBMV II)[40-42]. The velocity-Verlet algorithm was used with a time step of 2 fs. A temperature of 300 K was kept constant with a Nose-Hoover thermostat[43,44]. The cutoff for non-bonded list generation was set to 21 Å, the cutoff for non-bonded interactions was set to 18 Å, and the onset of switching for non-bonded interactions was set to 16 Å. The SHAKE algorithm was used to constrain vibrations of covalent bonds to hydrogen atoms involved. To conserve the duplex DNA structure during pre-equilibration, flat-bottom distance restraints were applied, which prevented the hydrogen-bond donor from moving more than 2.0 Å. The simulations ran for a total of 6.0 ns.

For the collection and analysis of equilibrium data, the initial 1 ns simulation data were discarded. We constructed a 2D free-energy map using the following two reaction coordinates: the DNA backbone root-mean-square deviation (rmsd) and the potential energy value; the minimization of this approximate free energy surface permitted us to choose the most populated structure in the canonical ensemble. [$F=-k_BT \ln P(X,Y)$, X and Y are reaction coordinates, k_B is Boltzmann's constant, T is the absolute temperature]. The DNA backbone rmsd is defined by selecting the P-O5'-C5'-C4'-C3'-O3' atoms. The reference structure was a standard B-form DNA (Twist angle $\Omega = 36.0°$, Rise per base pair along helix axis = 3.3 Å) without minimization. In order to probe the transition pathway between a WC and a HG base pair conformer, we used the conjugate peak refinement (CPR) method[45] applicable to the study of complex isomerization reactions including allosteric transitions in proteins and more general conformational changes in macromolecules. The resulting paths follow the adiabatic energy surface without applying any constraints and path-points between saddle-points ensure the continuity of the path, not necessarily constrained to find the absolute bottom of the energy valley. The initial WC-to-HG pathways were generated using a targeted molecular dynamics method

that applied a holonomic constraint, which decreased gradually the rmsd to the final target structure. In each CPR cycle, a heuristic procedure was used to modify the path by improving, removing, or inserting one path-point, so that the new path avoids the maximum energy peak. Finally, to refine the CPR path further, we used a synchronous chain minimization method of all path-points, under the constraint that the points move within hyper-planes orthogonal to the path. The most populated structures with the WC base pair or with the HG base pair from the normal MD simulations corresponded to "reactant" and "product" wells, respectively, on the energy surface. In order to sample a wider range of possible transition pathways, additional putative intermediates that differed in either the flip-over or flip-out were added by modifying, respectively, the glycosidic angle, χ (O4'-C1'-N9-C4), of the purine base and a center-of-mass pseudo-dihedral angle, θ^{46}, which describes the extent to base opening.

3.2.5 Density functional theory (DFT) calculations of carbon chemical shifts

DFT chemical shift calculations were conducted on a high-performance computing cluster using Gaussian 03[47]. DNA conformations of the target A•T and G•C base pair that represent a range of (χ, θ) pairs were selected from each simulated WC-to-HG transition pathway and capped by 3'OH/5'OH (UCSF Chimera[48]) for DFT calculations (Appendix 1) without further geometry optimization. NMR ^{13}C chemical shift calculations were conducted using the GIAO method with the B3LYP/6-311+G(2d,p) basis set. The isotropic carbon chemical shifts (σ_{ISO}) were referenced to TMS (σ_{TMS} = 182.759) using the relationship $\delta_{ISO} = \sigma_{TMS} - \sigma_{ISO}$, where the structure of TMS was optimized at the same level of theory. Computed carbon chemical shifts were referenced to the most stable WC conformer in a given transition pathway and matched with NMR excited-state chemical shift differences ($\Delta\omega_{AB}$). For benchmarking, similar calculations were performed on single guanosine nucleotides with *anti* and *syn* glycosidic conformations from crystal structures of *Oxytricha Nova* telomeric G-quadruplex (5' (G)$_4$(T)$_4$(G)$_4$; PDB ID: IJPQ, 1JRN, 2GWQ, 2GWE, and 2NPR) with added hydrogen atoms (UCSF Chimera[48]) and no further geometry optimization, and compared to observed NMR C8 chemical shifts for the same G-quadruplex (courtesy of Dr. Michelle Gill and Dr. Patrick Loria) (Appendix 1). DFT chemical shift calculations performed for

unmodified and N1-methyladenine modified HG base pairs used existing either crystal structures (PDB ID: 1K61, 3IGK, 3IGL, 3KZ8, 3H8O) or MD-generated HG base pairs that have been manually modified (UCSF Chimera[58]) by adding hydrogens and without further geometry optimization.

3.3 Results and Discussion

3.3.1 Developing a selective ^{13}C $R_{1\rho}$ relaxation dispersion experiment

(collaboration with Dr. A.L. Hansen and Dr. A. Casiano-Negroni, experimental details are only briefly reviewed here and fully outlined in Dr. Hansen's dissertation)

We developed an off-resonance ^{13}C $R_{1\rho}$ relaxation dispersion NMR experiment that targets protonated carbons in nucleic acid base and sugar units[22] (Figure 3.1). This experiment has been adapted from a similar off-resonance ^{15}N $R_{1\rho}$ relaxation dispersion experiment reported by Korzhnev *et al.*[49] designed to measure chemical exchange in protein backbone amide nitrogens. The adapted method utilizes selective Hartmann-Hahn cross-polarization transfers to inspect individual ^{13}C sites at a time coupled with a 1D acquisition scheme that saves significant experimental time as compared to routinely used 2D experiments (Figure 3.2). Because of the dramatically short acquisition times, the experiment allows one to thoroughly examine the dependence of $R_{1\rho}$ relaxation on spinlock power and offset and accurately determine all chemical exchange parameters – the exchange rate constant (k_{ex}) and the individual rate constants for the forward and reverse reactions (k_A and k_B), the populations of the two exchanging species (p_A and p_B), and the chemical shift difference between them ($\Delta\omega_{AB}$).

The experiment can also accommodate the use of low on/off-resonance spinlock powers ($\omega_{min} \sim 100$ Hz) without complications from unwanted interactions, which extends significantly the timescale sensitivity to chemical exchange relative to conventional 2D $R_{1\rho}$ relaxation dispersion methods ($\omega_{min} \sim 1000$ Hz)[50,51]. In general, the major difficulty in the use of weak (< 1000 Hz) spinlock fields in these traditional 2D experiments is attributed to the train of hard ^{1}H 180° pulses or ^{1}H broadband decoupling schemes used to suppress the effects of cross-correlation between dipole-dipole (DD) and chemical shift anisotropy (CSA) relaxation mechanisms. These methods can result in

incomplete decoupling of scalar coupling interactions and lead to systematic errors in measured $R_{1\rho}$ relaxation rates, especially at low spinlock fields[52,53]. An important advantage of the newly developed pulse sequence is the ability to implement a strong on-resonance continuous wave ^{1}H decoupling during the relaxation period that very effectively suppress [13]C-^{1}H one-bond scalar coupling (J_{CH}) evolution and cross correlated relaxation between [13]C-^{13}C DD and ^{13}C CSA during the relaxation period. Thus, utilizing low power ^{13}C spinlocks is no longer presents an issue. Moreover, homonuclear [13]C-^{13}C interactions with neighboring protonated and quaternary carbons are effectively controlled by the selective excitation of the carbon nucleus of interest. These homonuclear interactions present challenges for measuring accurate transverse relaxation rates for uniformly ^{13}C/^{15}N-labeled systems by 2D CPMG relaxation dispersion experiments that can extend to much slower timescales.

Although 1D acquisition is not preferred when multiple peaks are targeted, it is to our advantage that μs-ms chemical exchange in nucleic acids is typically limited to a smaller number of residues localized around non-canonical regions. Here, we tested the ^{13}C relaxation dispersion experiment on a ^{13}C/^{15}N-labeled ribosomal A-site RNA construct containing a dynamic internal loop and an unlabeled DNA construct containing a damaged nucleotide. We were able to detect and characterize μs chemical exchange dynamics in both systems that could have functional importance in ribosomal decoding and damage recognition, respectively. The details of this study focusing on damaged DNA have been discussed elsewhere[54].

3.3.2 Characterizing excited states at CA steps and their sequence dependence

We used the selective ^{13}C relaxation dispersion experiment (above) that allows detection of **μs-ms** chemical exchange to probe for the existence of excited states in the three DNA duplexes studied in Chapter 2 (Figure 3.3a). On the basis of selective line broadening at the adenine (A16), we had inferred that a chemical exchange process was present at the CA/TG step located at the 5' A-tract junction. Indeed, we uncovered pronounced μs dynamic transitions towards excited states occurring specifically at A•T and G•C base pairs in CA/TG steps (Figure 3.3b). Chemical exchange of such magnitude is observed only for CA/TG steps that, together with TA and CG steps, show the most

structural variety and anisotropic flexibility at the ps-ns timescale. In particular, we observed significant $R_{1\rho}$ relaxation dispersion at the base C8 and sugar C1' carbons of the A and G residues, and for the C6 carbon of the C residue (Figure 3.3b). No relaxation dispersion was observed for the adenine base C2, the thymine residue, or the cytosine sugar C1' sites (Figure 3.3b). None or very small chemical exchange contributions were detected for other residues (Appendix 1).

A two-state equilibrium analysis ($A \underset{k_B}{\overset{k_A}{\rightleftharpoons}} B$) of the off-resonance $R_{1\rho}$ relaxation dispersion data (Figure 3.3b,c) revealed a single base pair exchange process that occurred in an seemingly uncorrelated manner between the neighboring A•T and G•C. Specifically, the exchange rate constants and excited state populations were statistically the same for C8 and C1' sites within the same purine nucleotides (A16 and G10). Moreover, similar chemical exchange parameters between C6 of C15 and C8/C1' of G10 strongly argued for a concerted base pair motion. In A_6-DNA, the exchange was slightly faster for A•T versus G•C and was directed towards minutely populated ($p_B \sim 0.64$ % and ~ 0.47 % for G•C and A•T, 26 $^{\circ}$C) excited states that have exceptionally short lifetimes ($\tau_B = 1/k_B \sim 1.5$ ms and ~ 0.3 ms for G•C and A•T, 26 $^{\circ}$C), and that have downfield-shifted ^{13}C chemical shifts ($\Delta\omega_{AB(C8)} \sim 2.7$ - 3.2 ppm, $\Delta\omega_{AB(C1')} \sim 3.7$ ppm, and $\Delta\omega_{AB(C6)} \sim 2.2$ ppm, Appendix 1).

Initial analysis revealed that the exchange process was modulated by adjacent 3' neighboring A-tracts and by positional context. While the best-fit exchange parameters were indistinguishable between the A_6- and A_4-tract, the excited state populations and exchange rates for A16 were ~ 2-fold higher for A_6-DNA ($p_B \sim 0.4$ %, $k_{ex} \sim 3700$ Hz) in comparison to A_2-DNA ($p_B \sim 0.2$ %, $k_{ex} \sim 2100$ Hz) at 26 $^{\circ}$C. Although the exact nature of these differences cannot be determined, we offer possible scenarios for how relative changes in energetics of the populated species could produce these effects. Specifically, the kinetic and thermodynamic effects induced at A•T by the longer 3' A-tract could stem from (i) relatively lower stability for both the WC base pair and the excited state (ES) or (ii) relatively lower stability for the WC base pair and higher stability for the transition state (TS). The G•C base pair (G10•C15) experienced a noticeable kinetic, but not thermodynamic, impact from the adjacent A-tract length. We observed a combination of gradual decrease in k_{ex} values and no sizeable change in p_B values that could signify a

relative TS destabilization at G•C next to longer A-tracts. For the second CA/TG steps in A_2-DNA, which is found among different neighbors and in closer proximity to the terminal end, we observed faster k_{ex} values (~ 1.5 to 3 fold) at the adenine (A3), yet very similar stability of the excited and ground states (p_B) as compared to other CA/TG steps. These results could be interpreted in terms of destabilization of both WC and ES with respect to the high-energy TS and can be equally attributed to either sequence or end-fraying effects.

With this information at hand, we cannot differentiate between the various molecular origins for the altered chemical exchange dynamics at CA/TG that are correlated to changes in sequence and position. On one hand, longer A-tracts seem to relatively destabilize the TS for G•C, making it harder to achieve the transition between the two states. On the other hand, longer A-tracts seem to relatively stabilize the TS for A•T, making it easier to achieve the transition. This could be a result of progressively larger static deformations or bends at the A-tract interface with B-DNA that affect inversely the inter-conversion between the two states in G•C and A•T. Our hypothesis provides only one explanation for the observed differences in microsecond dynamics, though alternative scenarios could be at play. Nevertheless, we can acknowledge that sequence-specific effects, though not dramatic in magnitude, are present when it comes to
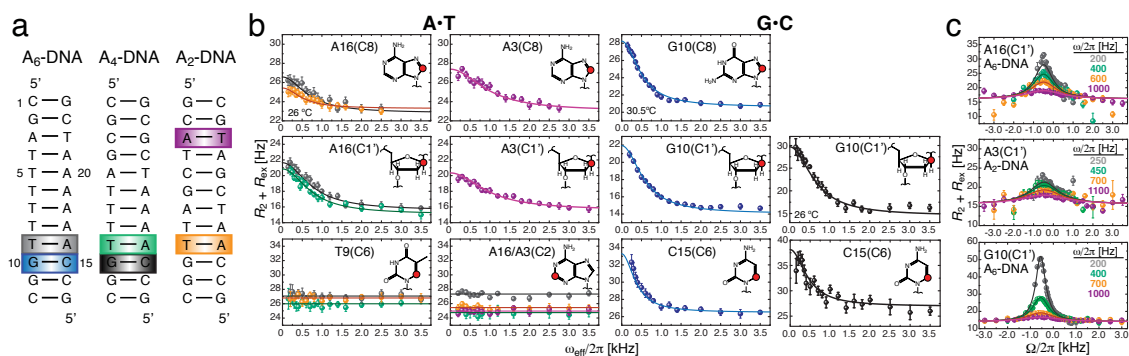
Figure 3.3: Detection of base pair specific excited states in CA/TG steps of duplex DNA. (a) DNA constructs containing varying length A-tracts with color-coded A•T and G•C base pairs at CA/TG steps that show carbon chemical exchange. (b) On-resonance ^{13}C $R_{1\rho}$ relaxation dispersion profiles for A•T (26°C) and G•C showing CA/TG specific chemical exchange at purine base C8 and sugar C1' and at cytosine base C6. Data shown for G10 in A_6-DNA is at 30.5 °C/pH 5.4 and G10 in A_4-DNA at 26 °C/pH5.2. Shown are the best base pair global fits (solid line) to a two-state asymmetric exchange model. (c) Representative off-resonance relaxation dispersion profiles for corresponding C1' sites and best global fits as in (b). (all data included in Appendix 1)

the slow conformational dynamics observed here. Ultimately, this provides further support for our hypothesis that conformational preferences encoded in specific DNA sequences can govern the dynamics of the DNA double helix.

3.3.3 Chemical shift and kinetic-thermodynamic assignment of excited state Hoogsteen (HG) base pairs

What is the excited state encoded by CA/TG steps and detected by NMR relaxation dispersion? NMR imino proton exchange measurements[55,56] and computer simulations[57] have previously shown that WC base pairs can spontaneously break open and access extrahelical conformations. While the forward exchange rates (k_A) measured by relaxation dispersion (~ 4 - 20 s^{-1} at 26 ^{o}C) are within an order of magnitude of rates reported previously for WC base pair opening (~ 40 - 400 s^{-1} at 25 ^{o}C)[2] inside canonical duplexes, the excited states detected here are at least three orders of magnitude more populated as judged from a large body of literature[2,56,58-63]. Although "open" states for canonical DNA base pairs show thermodynamic stabilities relative to "closed" states that span several kcal/mol (~ 6 - 10 kcal/mol), the corresponding fractional populations (p_{op}) never seem to exceed ~ 0.001%. The only notable exception where p_{op} approaches p_B for the excited state are non-canonical base pairs such as G•T wobble[64] (p_{op} ~ 0.1 - 1%). Even though their relative stabilities are comparable, the lifetime of the G•T wobble open state (< 4 μs at 0°C) is estimated to be much shorter than the lifetime of the excited state (hundreds of μs to ms at 0°C), and the lower energy gap is mainly due to diminished stability of the "closed" wobble pair. This presents strong evidence that the excited state is a far more energetically favorable species than base pair open states.

To obtain more insights into the excited state, we measured carbon $R_{1\rho}$ relaxation dispersion as a function of temperature (Figure 3.4a, Appendix 1). We then used transition-state theory and van't Hoff analysis to extract a complete thermodynamic-kinetic description of the two-state equilibria (Table 3.1). Semi-logarithmic van't Hoff plots revealed a linear dependence characteristic of a two-state process for A•T and G•C base pairs (Figure 3.4b). The analysis yielded activation free energies (~ 16 kcal/mol) and enthalpies (~ 12 - 26 kcal/mol) for the forward transition that are comparable to values reported previously for base pair opening (~ 14 - 25 kcal/mol and ~ 8 - 29 kcal/mol

respectively)[2,46]. Thus, the transition to the excited state entails disruption of stacking and hydrogen-bonding interactions in the WC base pair. However, for both A•T and G•C, this loss in enthalpy is nearly entirely restored when the exited state is formed. In fact, the excited state is in part destabilized relative to the WC ground state by less favorable entropy, implying a more "rigid" excited state conformation and counter to what would be expected for a flexible looped-out or base pair open state. Interestingly, we observed strong enthalpy-entropy compensation between activation and overall energies that leads to much narrower differences in the total free energies for various base pairs. This phenomenon has similarly been observed for base pair opening in both DNA[2] and RNA[65] double helices.

Taken together, our data point to an excited state conformation whose creation requires complete disruption of WC base pairs, but whose thermodynamic stability is comparable to that of a WC base pair. One possibility is that the excited state represents an alternative base pair. Here, the correlated nature of the exchange at purine (A and G)
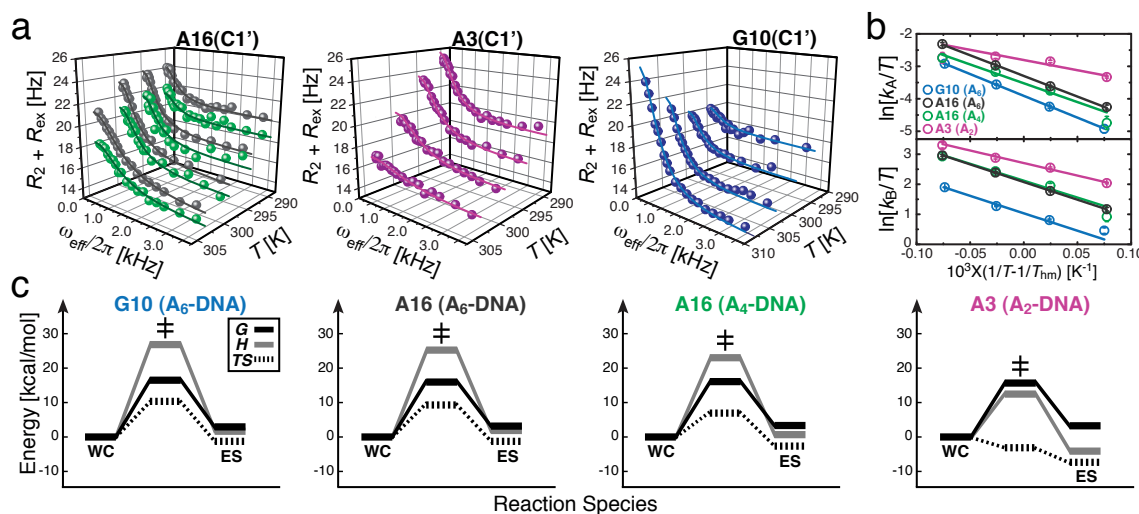
Figure 3.4: Kinetic-thermodynamic analysis of ground-to-excited state transitions. (a) Representative on-resonance ^{13}C $R_{1\rho}$ relaxation dispersion profiles as a function of temperature for A16 (A_6-DNA and A_4-DNA), A3 (A_2-DNA), and G10 (A_6-DNA) C1' (all plots included in Appendix 1). The analysis was performed at pH 5.4 for A16/G10 A_6-DNA (see Table 3.1 for A16 at pH 6.8) and at pH 6.8 for A16/A3 in A_4-DNA and A_2-DNA. (b) Modified van't Hoff plots showing temperature dependence of the forward (k_A) and reverse (k_B) rate constants for the two-site exchange in A•T and G•C base pairs highlighted in Figure 3.3. (c) Corresponding kinetic-thermodynamic profiles for exchange between the Watson-Crick (WC) ground state and the excited state (ES) via a transition state (‡), showing activation and net free energy (G), enthalpy (H), and entropy (TS) changes (referenced to 0).

C8 and C1' nuclei and the large downfield ^{13}C chemical shifts provides important clues for the structural transition. In particular, the magnitude and direction of $\Delta\omega_{AB(C8)}$ and $\Delta\omega_{AB(C1')}$ are strongly indicative of rotation about the glycosidic bond that links the base and the sugar moieties, also known as an *anti*-to-*syn* transition, as deduced from a survey of ^{13}C chemical shifts in G-tetrads[66] and density functional theory (DFT) calculations[67]. For canonical DNA WC base pairs, glycosidic angles occupy a narrow range of 20° - 30° inside the *anti* conformational space. Thus, an *anti*-to-*syn* purine transition would require a dramatic base reorientation of $\sim 180^o$ concurrent with complete loss of base pairing and inter-base interactions. Thus, it would be natural for a purine constrained inside the framework of the double helix to require energies for base flipping on the order of what we observe. Without the constraints of the duplex, the energetic barrier to cross from an *anti* to a *syn* orientation is much smaller, estimated by quantum mechanical methods to be only $\sim$ 3 - 6 kcal/mol in purine mononucleotides[68]. Foloppe *et al.* also computed a separation of at least 3 kcal/mol between the energy minima of *anti* and *syn* conformers[68], though NMR studies suggest that single nucleoside/nucleotides, which have greater rotational freedom, exhibit higher proportions of the *syn* conformer (>10 %)[69] that is coupled with downfield-shifted chemical shifts relative to duplex nucleotides. Therefore, we suspect that poorly stacked or non-hydrogen bonded residues in non-canonical regions or terminal ends would have a greater natural tendency to adopt the *syn* conformer, unless adverse interactions hinder this process.

Remarkably, an *anti*-to-*syn* rotation of the adenine or guanine base results in creation of a Hoogsteen (HG) base pair optimally stabilized by two hydrogen bonds (Figure 3.5). The HG G•C base pair would require protonation of cytosine N3 (G•C$^+$, pK$_a$(N3) ~4.2[70]), which can in turn explain the relaxation dispersion observed at cytosine C6 at acidic pH, and the downfield shifted excited state C6 chemical shift, as well as the dramatic pH dependence of dispersion measured at carbon sites in the G•C but not in the A•T base pair (see Chapter 4). HG base pairs have widely been observed in non-canonical DNA structures, such as DNA triplexes and quadruplexes, where they are specifically recognized by proteins[71]. In a few cases, HG base pairs have also been observed in duplex DNA containing alternating AT repeats[72] or in complex with ligands[12,13] and proteins, including the active site of DNA polymerase-ι[73] and a complex

between TATA binding protein (TBP) and a mutant TATA-box DNA[10], where HG G•C$^+$ base pairs are formed at pH ~ 6. Indeed, HG base pairs could form more easily in the context in an all-*anti* DNA triplex, where they show similar stability to WC base pairs in duplex DNA[74]. Moreover, computational simulations of WC-HG hybrid duplexes suggest that B-DNA can easily tolerate incorporation of single HG base pairs without significant structural, dynamic and energetic perturbations[75].

An excited-state HG base pair can explain the lower enthalpy and lower entropy of the exited state relative to a base pair open state. Its creation requires a ~ 180-degree base rotation around the glycosidic linkage and disruption of the WC base pair, consistent with our measured transition state barriers. These excited state HG base pairs may have evaded detection by prior solvent exchange measurements due to hydrogen-bond protection of the thymine imino proton in (A•T) and/or inaccessibility of the guanine imino proton in (G•C) in addition to slower exchange rates. For example, the observation of a non-hydrogen-bonded G imino H1 resonance in a G(*syn*)•A$^+$(*anti*) mismatch base pair at 20 °C inside a DNA duplex[76] supports lower exchange rates with solvent, possibly due to interaction with the backbone and/or steric effects. Yet another reason for the slower *syn*-G H1 exchange rates with solvent when in an HG base pair could be the inclusion of G N7 into a hydrogen bond and protection from solvent, the reason being that G N7 protonation is believed to accelerate H1 transfer to water (lowers the pK_a of N1)[77]. In the context of a WC G•C base pair, C N3 acts as a powerful intrinsic catalyst on N7-protonated G to facilitate proton transfer, the prevalent exchange mechanism under acidic conditions, while this mechanism would not be possible in the context of an HG G•C$^+$ base pair[78].

To summarize, on the basis of chemical shifts and kinetic-thermodynamic parameters computed from NMR relaxation dispersion data, we have assigned the excited states at A•T and G•C in CA/TG steps as Hoogsteen base pairs that form transiently following an *anti*-to-*syn* purine rotation and, in the case of G•C, protonation of C N3 to optimize H-bonding. Although highly uncommon and energetically unfavorable inside a naked B-DNA duplex, the HG geometry has been observed in DNA when forming a continuous HG double helix or when complexed with proteins and drugs, factors that can alter the energy landscape of DNA and/or capture excited state conformations. The

potential for HG base pair formation suggests that it can already exist as an accessible excited state in B-DNA.

3.3.4 NMR-informed simulations of the WC-to-HG transition

(collaboration with Dr. Eunae Kim and Prof. Ioan Andricioaei from the University of California, Irvine)

To assess the energetic and stereochemical feasibility of the proposed WC-to-HG transition, as well as obtain insights into the transition pathway, we used conjugate peak refinement (CPR)[45] methods to simulate multiple transition pathways between WC and excited state HG base pairs that sample various glycosidic (χ) and base opening (θ) angles (Figure 3.5). CPR simulations of the WC-to-HG transition were performed for three base pairs in A_6-DNA – A16•T9, G10•C15, and A3•T22 – where the hypothetical HG base pair was created from crystallographic data. The first two base pairs located at a CA step displayed chemical exchange, while the third one located at an AT step showed no detectable exchange contribution and was used as a control. For all purines, the *anti*-to-*syn* transition was achieved via bi-directional rotation (clockwise and counterclockwise) around the glycosidic bond (χ) and via various degrees of base flipping ($\theta < 70^{\circ}$) through the major groove (meaningful simulations could not be obtained with opening into the minor groove). In general, CPR trajectories showed smooth WC-to-HG transitions accompanied by little variation in the sugar pucker (C2'-endo for A and G), and minor adjustments in the complementary pyrimidine residue and neighboring WC base pairs. Optimal pathways featured purine base-flips at low base pair opening angles (θ) into the major groove and transition states, which require complete disruption of WC base pairing – this is also observed experimentally by relaxation dispersion – in which the purine base is near-orthogonal to its pyrimidine partner. This unusual geometry is well accommodated by the rather large inter-base pair spacing in the B-form helix (i.e. base pair rise ~3.3 Å) and its ability to mold without loss of other WC base pairing. Even though rise shows very little variation across DNA structural databases[79,80], it has been observed that some CA/TG and CG steps exhibit higher rise exceeding 4 Å that could open up more space for base rotation[81].

We should note that the CPR protocol generated unique trajectories with variable χ-θ correlations and extent of base opening for each base pair. For example, A16 experienced a simultaneous change in χ and θ coordinates, while A3 sampled a path where base rotation around χ was preceded by base opening (θ), reminiscent of a two-step process. Despite that, comparable transition energies were simulated for A3 versus A16 at similar θ values and the overall pseudo-free energy change was on average within 3 kcal/mol for A3 as compared to A16. Therefore, the CPR profiles suggest that the
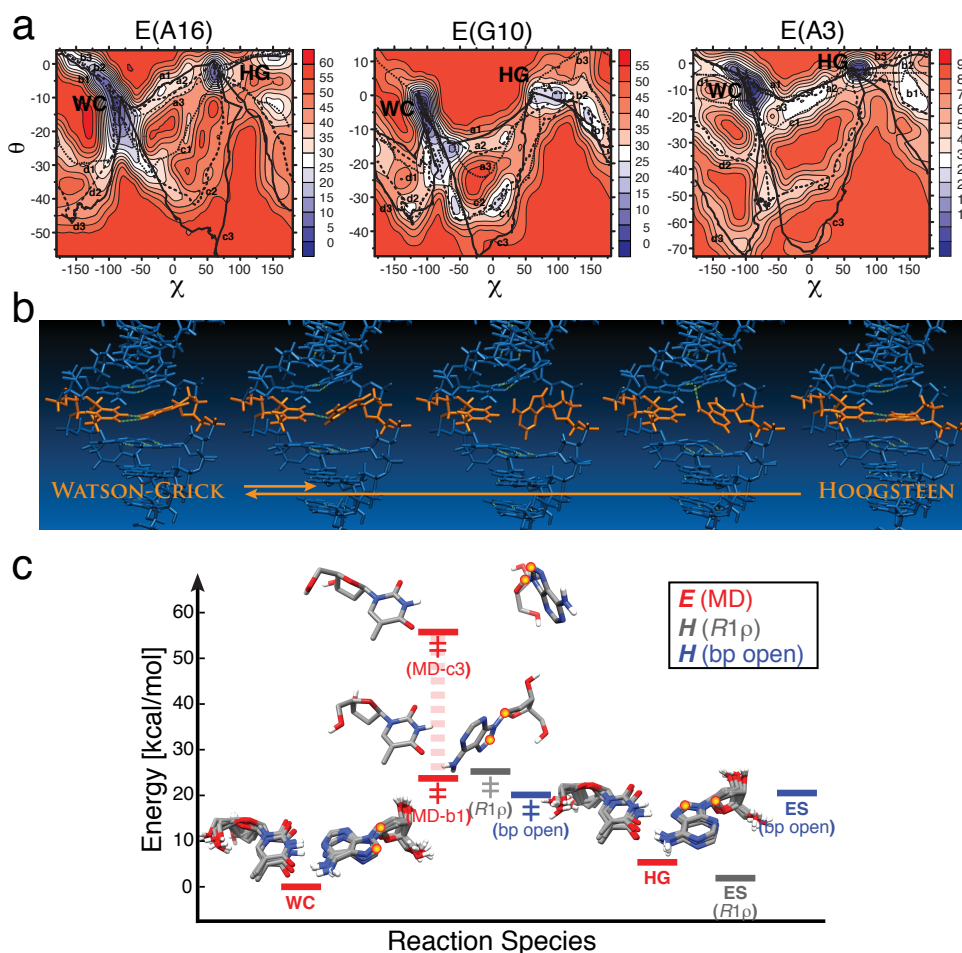
Figure 3.5: Watson-Crick to Hoogsteen base pair transition simulations. (a) Pseudo-free-energy (E, kcal/mol) contour plots as a function of (χ, θ) pairs for A16•T9, G10•C15, and A3•T22 in A_6-DNA obtained from multiple CPR trajectories (*a-d*). (b) Snapshots from a representative CPR transition pathway (*a1*) for A16•T9. (c) Initial WC and final exited-state (ES) HG structures and representative lowest-energy (*b1*) and highest-energy (*c3*) transition state (‡) structures of A16•T9 illustrating the span of CPR transition barriers, and their relative potential energies (averaged for WC and HG) compared with enthalpies (*H*) derived from NMR relaxation dispersion for chemical exchange or NMR imino proton exchange for A•T opening[2].

A3•T22 base pair can also access the HG configuration, but that exchange likely escapes detection by NMR relaxation dispersion due to small kinetic and/or thermodynamic differences.

Although a large spread of transition barriers were sampled, the minimal CPR energy barriers and net energy changes were within 2 kcal/mol of NMR-derived ensemble enthalpy terms (Figure 3.5). The stability of the final HG states relative to their WC counterparts was, on average, much lower for G•C than for A•T. This likely results from the inability of G•C to achieve optimal HG geometry and hydrogen bonding in the absence of protonated C N3. The final HG state for G10•C15 is observed to adopt variable *syn*-guanine geometry, with either one or no optimal hydrogen bonds to the unprotonated C, but on occasion forming an intraresidue hydrogen bond from amine groups to backbone oxygens. Unfortunately, our MD simulations were not extended to DNA constructs with a protonated C since CHARMM force fields are not currently available to support protonated C systems and because of our inability to determine precisely the most favorable geometry for protonation during the CPR simulation.

Overall, the CPR simulations demonstrate that WC-to-HG transitions are structurally possible, and via distinct motional trajectories, in the framework of a B-DNA duplex. The relative CPR energies between various pathways suggest, almost counterintuitively, that preferred pathways could involve purine rotation inside the helix without significant base flipping. Here, we assume that CPR energies can be directly compared to NMR-derived enthalpies, yet we have no prediction of the free energy and entropy components that these dramatic transformations entail. Thus, we should also not discard pathways that lead to more extensive base opening, since the gain in conformational entropy for a base that has "escaped" the helix could make these pathways competitive, if not dominant, mechanisms in reality. In fact, the excellent agreement between activation energies from WC to TS states in base pair opening processes measured by imino proton exchange and in excited state formation measured by relaxation dispersion provides support for a common, or at least energetically similar, TS between the two dynamic events.

3.3.5 Matching of chemical shifts between ES and HG base pairs generated by computational methods or "trapped" by chemical modification and drug binding

3.3.5.1 DFT chemical shift calculations

The CPR MD simulations were essential for generating a conformational pool of WC, HG base pairs and intermediate structures that we in turn used to explore the structure-chemical shift space for a range of χ-θ pairs. Comprehensive DFT carbon chemical shift calculations were performed using selected conformers along the MD CPR trajectories for A16•T9 and G10•C15. Chemical shifts computed for the WC and lowest-energy HG geometries exhibited consistently the characteristic downfield-shifted $\Delta\omega_{AB(C8)}$ and $\Delta\omega_{AB(C1')}$ observed experimentally by relaxation dispersion (Figure 3.6a). Predicted values that gave the best simultaneous match to NMR values ($\Delta\omega_{AB,C8}$ and $\Delta\omega_{AB,C1'}$) were within ~ 0.1 - 1.0 ppm of the experimental data

Figure 3.6: Matching of DFT-predicted and excited state ^{13}C chemical shift differences for WC and HG base pairs. (a) Comparison between excited state chemical shift changes from relaxation dispersion and best matching DFT predictions (or average for T9 C6 and C15 C1'). (b) $R_{1\rho}$ relaxation dispersion profiles for T9 C6 and C15 C1' showing no detectable chemical exchange consistent with DFT results.

(Table 3.6). Moreover, our trends of C8/C1' chemical shift variations with χ for small base-opening angles (θ) reproduced well previously published DFT predictions for a χ dependence[30]. Calculations for C C6 in the final HG state, without N3 protonation, showed small and random chemical shift differences from the WC state (Appendix 1). By contrast, we saw large downfield perturbations for C6 when we manually protonated C N3 in the context of a G•C$^+$ HG base pair (Figure 3.6a), which agreed with the presence of chemical exchange at C15 C6 and strongly supported a protonated excited state. In addition, little variation was observed for A C2, T C6, and C C1', consistent with the lack of observable relaxation dispersion at those sites (Figure 3.6).

The comprehensive DFT calculations (Appendix 1) yielded a small number of geometries, other than HG base pairs, that match the measured excited-state chemical shifts, as a result of the intrinsic C8/C1' isotropic chemical shift dependencies on the glycosidic angle. However, those alternative conformations involved barrierless transitions and/or represented high-energy structures that disagreed with experimentally derived enthalpies, and could be readily excluded as candidates for the excited state. In our view, HG base pairs offered the best match to the excited state in terms of both chemical shift and enthalpic stability.

We recognize that DFT calculations performed on hypothetical, MD-generated structures could be associated with large bias and uncertainty due to lack of knowledge of the actual conformation for each species as well as unoptimized geometries at a high theoretical level. Despite that, we believe that this is the most adequate approach we could have taken in terms of allowing for structural variability of both ground and excited states, thus sampling a much wider conformational space, and in terms of time saving. In a way, the availability of an ensemble of structures for the DFT calculations provided a measure for the "structural noise" given by differences in local geometries or CA/TG dinucleotide conformations. For example, the spread of NMR-observed C8 chemical shift differences between *anti* and *syn* guanines inside a single G-quadruplex had a standard deviation of ~ 1.2 ppm, which is on the order of the differences that we observe for conformations with similar χ, while the mean value was ~ 3.0 ppm and matches well with $\Delta\omega_{AB(C8)}$ ~ 3.2 ppm from relaxation dispersion. Finally, the only currently attainable high-resolution structure/ensemble would be that of the ground WC state, and the benefits of knowing the exact ground state structure without knowledge of the excited state structure are not apparent.

3.3.5.2 Trapping HG base pairs by chemical modification

We further used chemical modifications to trap the excited state HG base pair within duplex DNA, which allowed us to directly compare the carbon chemical shift signatures of the trapped HG base pair with those measured for the excited state using relaxation dispersion. By installing an *N*1-methylated adenine (1mA), which is a common DNA lesion known to sterically impair WC base pairs and favor HG base pairs[14,82] at the

CA step, we trapped the HG A•T base pair in the A_6-DNA duplex (A_6-DNA^{1mA16}) with all surrounding base pairs in WC geometry, as confirmed by analysis of proton NOE connectivities, chemical shifts, and intensities (Figure 3.7). Specifically, we observed an upfield shift in T9 imino H3 and a downfield shift in 1mA16 amino H6 (H-bonded) that are characteristic for the alternate HG hydrogen-bonding geometry[14,82], in addition to connectivities from H3 and H6 to the neighboring WC base pairs, indicating that the HG base pair was well accommodated inside the B-DNA duplex. We also detected other structural features that are unique to HG base pairs with *syn* purine geometries, the most notable being (i) a cross-peak between 1mA16 H8 and T9 H3 that are found in close proximity in HG base pairs but not WC base pairs, (ii) a very intense cross-peak between 1mA16 H8 and H1' that are brought within a shorter distance in the *syn* versus the *anti* conformation, and (iii) connectivities between C15 H1'/H5 and 1mA16 H2 instead of 1mA16 H8 normally seen in B-DNA, since H2 has taken the place of H8 in the new *syn* geometry.

Though never observed before, we also successfully trapped an HG G•C$^+$ base pair using instead the structural analog to 1mA – an *N*1-methylated guanine (1mG) – at the same CA step (A_6-DNA^{1mG10}) (Figure 3.7b). We used acidic conditions to ensure that the HG base pair is largely protonated at C N3. G•C resonances were either significantly broadened (C16 C6H6) or not observable (1mG10 C8H8 and C15 C1'H1') at pH 6.8 due to enhanced chemical exchange, while these resonances became well resolved at the lower pH 5.2 that stabilizes the HG base pair. We found some of the same HG base pair indicators by inspecting the proton NOESY spectra of A_6-DNA^{1mG10}. Those included a downfield-shifted C15 amino H4 and cross-peaks to T9 H3 that supports the alternate H-bonding scheme and the intrahelical base pair stacking, and an intense 1mG10 H8-H1' cross-peak that supports the *syn* geometry (Figure 3.7b). Unfortunately, the C15 imino H3 resonance and any corresponding cross-peaks that would provide further evidence for the HG base pair were not observable in the NOESY, possibly due to rapid exchange of H3 with solvent. However, we assumed that the C imino N3 was mostly protonated because of the large C6 downfield chemical shift that is predicted to occur with protonation.

Once we had confirmed the presence of the modified A•T and G•C$^+$ HG base pairs, we studied 2D ^{1}H,^{13}C HSQC spectra at natural abundance for the behavior of base and sugar carbons. The chemical shifts for 1mA16 C8 and C1' exhibited large perturbations in the same downfield direction (Figure 3.7a), and with the same magnitude for C1' (within 0.3 ppm), as the NMR-derived excited state A16 chemical shifts (Figure 3.9b). Remarkably, we observed an excellent chemical shift correspondence between 1mG10 C8 and C1' and C15 C6 and relaxation dispersion data for unmodified G•C (Figure 3.7b). The noticeably downfield-shifted C8 in the trapped *syn*-1mA16 versus the unmodified excited state HG base pair (~ 2.9 ppm) could be attributed to changes in the electronic environment arising from introduction of the methyl group and positive charge on the heterocyclic base, or possible differences in the HG geometry. This was supported by direct comparison of computed C8 chemical shifts for an unmodified A and an *N*1-methylated-A in the *syn* conformation, based on crystal structures or MD-generated structures, which showed a several ppm downfield shift (~ 5 - 6 ppm) for the modified adenine. By contrast, the chemical shift perturbation induced by *N*1-methylation of G in the HG base pair, where no positive charge is acquired, was negligible (~ 0.2 ppm) and consistent with the excellent match between 1mG10 and excited state G10 C8 chemical shifts. Finally, carbon chemical shifts of T9 C6 and A16 C2 (A$_6$-DNA^{1mA16}), C15 C6 (A$_6$-DNA^{1mG10}), and base pairs neighboring the damaged sites were not significantly altered (< 1.0 ppm), in agreement with the absence of pronounced chemical exchange at those sites in unmodified A$_6$-DNA. Yet, the dynamics at some adjacent base pairs were clearly perturbed, judging from their broad lineshapes, and signified diminished base pair stability that would be subject to further investigation.

The ability to independently trap HG A•T or G•C$^+$ base pairs is consistent with the relaxation dispersion data showing that transitions to excited state A•T and G•C HG base pairs within CA steps are semi-independent of one another. The differences in carbon chemical shift measured between modified and unmodified DNA were, where anticipated, in excellent agreement with the chemical shift differences measured between the ground and excited state by relaxation dispersion (Figure 3.9). Conversely, resonances that showed small differences in carbon chemical shift between modified and unmodified constructs (Appendix 1) exhibited little to no carbon chemical exchange
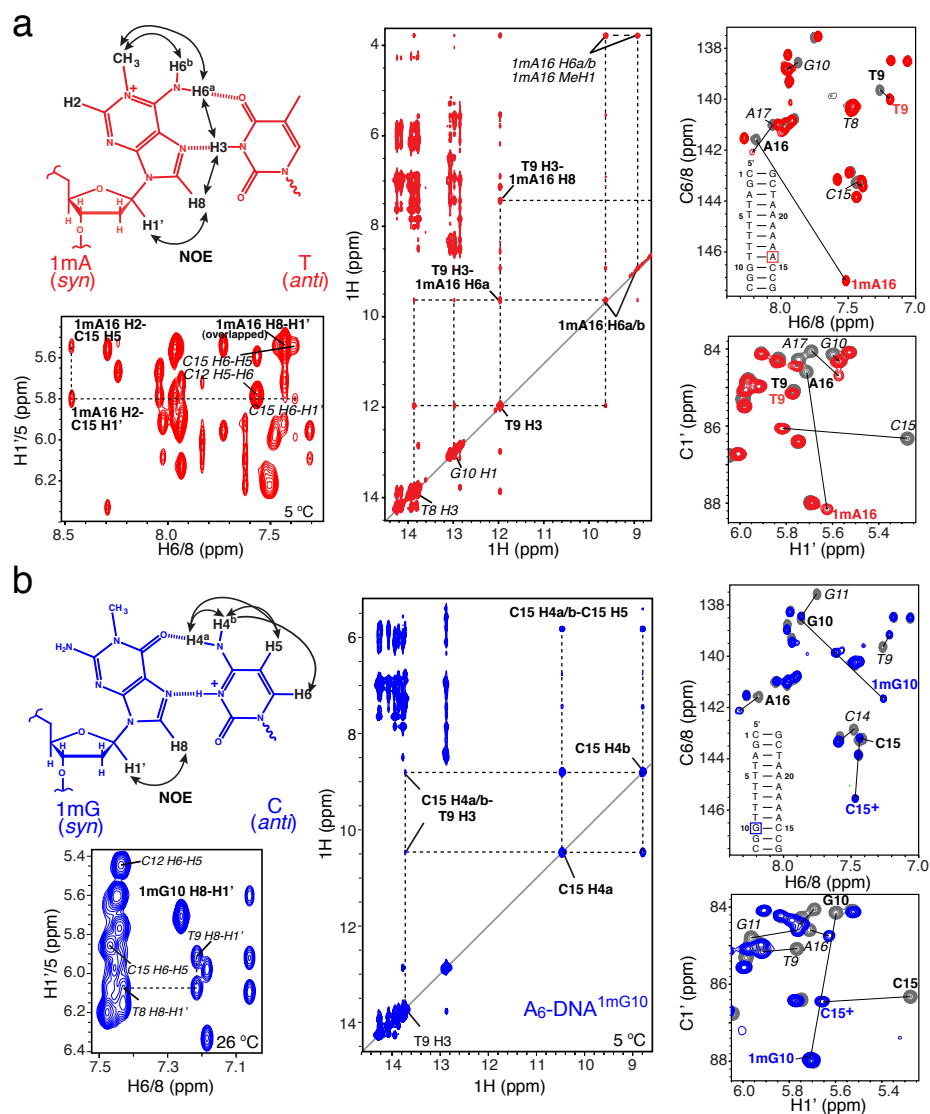
Figure 3.7: Trapping of HG base pairs in damaged DNA. Schematic of the 1mA•T (a, red) and the 1mG•C[+] (b, blue) HG base pair and corresponding fragments of ^{1}H,^{1}H NOESY correlation spectra for A$_6$-DNA^{1mA16} and A$_6$-DNA^{1mG10} showing NOE connectivities, cross-peak intensities and proton chemical shifts characteristic for HG base pairs (see text for details). Shown on the right are ^{1}H, ^{13}C HSQC correlation spectra depicting chemical shift changes in the CA/TG step from unmodified A$_6$-DNA.

(Figure 3.7, Figure 3.3). Our data rule out other non-canonical base pairing modes involving pyrimidine base rotation, repuckered sugar conformation from C2'-endo to C3'-endo, or base nitrogen protonation. Although the C3'-endo sugar causes a very similar downfield shift for C1' and does the *syn* conformation, it also yields an upfield rather than a downfield nucleobase carbon chemical shift and does not match the excited state (Figure 3.9). As for protonation, gaining a proton by any of the base nitrogen would

result in minimal C1' chemical shift perturbation and in large upfield shifts for C2 in case the nitrogen is N1 or N3, as judged by DFT chemical shift predictions for protonated purines at N1, N3, or N7. This result generally contradicts the relaxation dispersion data where we observe large C1' downfield shifts and no exchange at A C2. Thus, comparison of known HG ground state chemical shifts with the excited state chemical shifts provides strong support for its assignment as an HG base pair.

3.3.5.3 Capturing pre-existing HG base pairs in a drug-DNA complex

As an inverse experiment, we asked whether TA steps, which have also been observed to form HG base pairs in duplex DNA bound to small molecule ligands[12,13], exhibit the characteristic HG exited state. We measured ^{13}C relaxation dispersion data for a palindromic DNA sequence, which has previously been shown to form tandem A•T HG base pairs in solution[30] when in complex with the bis-intercalating antibiotic echinomycin (Figure 3.8). Strikingly, we observed the same chemical exchange pattern, correlated dynamics at A C8 and C1', in the TA step of a DNA octamer (E-DNA) as in the CA/TG step of DNA dodecamers (Figure 3.8d). Furthermore, we found that carbon chemical shift differences between ground and exited state in TA steps that are in excellent agreement with the differences between the free (WC) DNA and drug-bound (HG) DNA (Figure 3.8, 3.9).

First, we assigned E-DNA in the free and drug-bound state, and verified that the same binding mode was present in the drug-DNA complex by comparing our NOESY spectra with previously published NOE data. For example, we observed the unique intra-HG base pair cross-peak between T4 imino H3 an A5 H8, also seen in the damaged HG base pairs above (Figure 3.8b). In addition, we identified an echinomycin-DNA cross-peak that firmly distinguishes between the *syn* and *anti* A5 geometries. The present NOE connectivity links Val CH3 to A5 H8 (~ 4.2 Å) and not to A5 H2 (~ 9.9 Å) that is consistent with the *syn* conformation (Figure 3.8b), while this connectivity would be reversed if A5 assumed the *anti* conformation in a regular WC base pair. In fact, both scenarios have been observed before and – echinomycin tends to trap HG base pairs in 5'--CG**TA**CG-- sequences[30] and WC base pairs in 5'--CG**AT**CG-- sequences[83] that, in

each case, are favored because they yield the optimal stacking between the DNA adenine bases and drug quinoxaline rings.
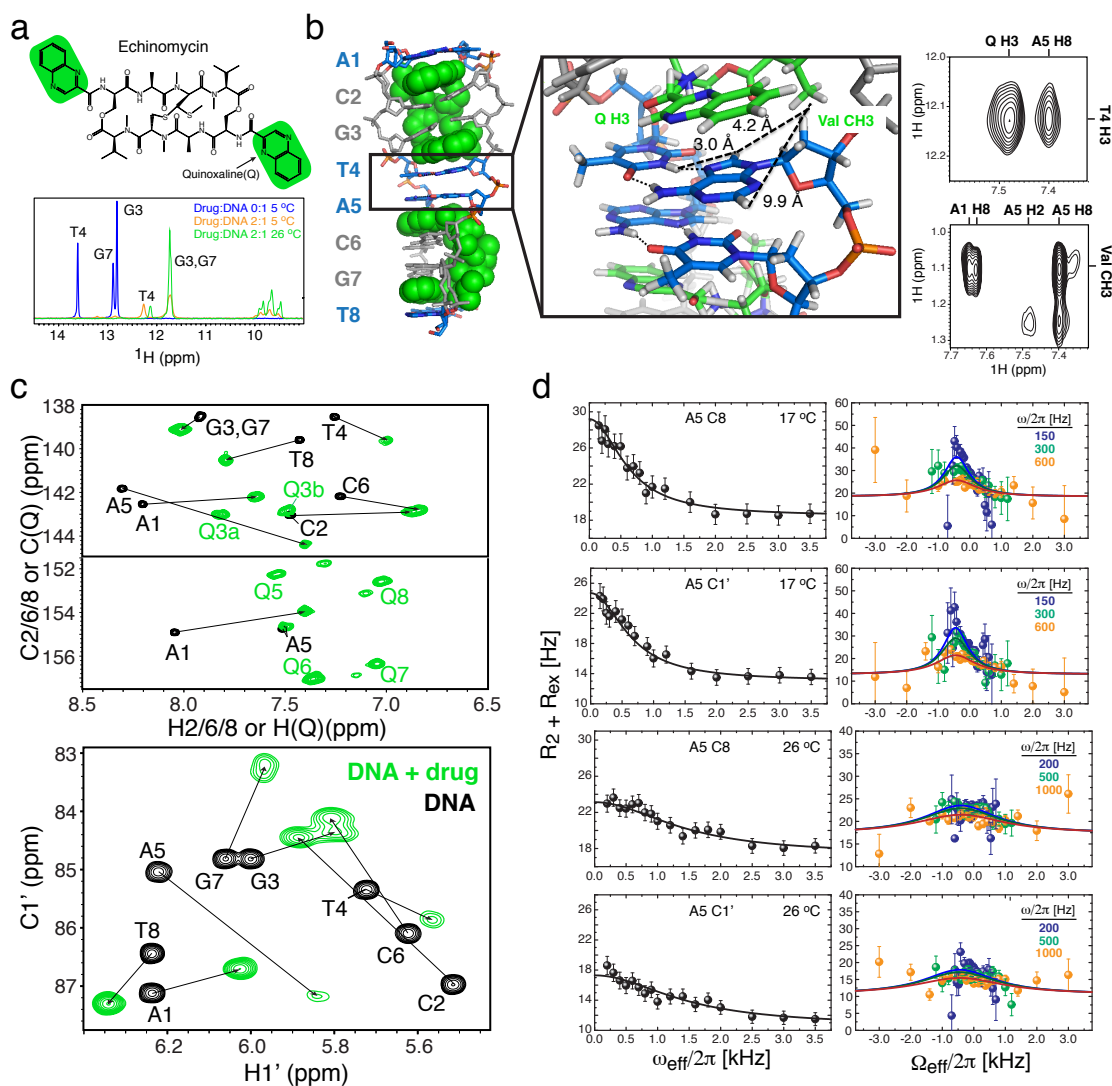
Figure 3.8: Drug binding traps pre-existing HG base pairs in a DNA duplex. (a) Structure of the peptide-based quinoxaline (green) antibiotic echinomycin and 1D ^{1}H spectral overlay of the imino/amino region of free and echinomycin-bound E-DNA (NMR buffer, pH 6.8), showing upfield-shifted T4, G3, and G7 H3 resonances upon drug binding and HG base pair formation at A5•T4[38] (b) Echinomycin intercalation mode into E-DNA (PDB ID: 3H8O) with HG A•T base pairs depicted in blue, a zoomed-in view of the interaction between the drug and HG base pairs highlighting NOE distances characteristic for the HG geometry, which is confirmed by observation of the respective NOE cross-peaks in E-DNA at 26 ^{o}C. (c) ^{1}H,^{13}C HSQC spectral overlays and assignments of free (black) and echinomycin-bound (green) E-DNA at 26 ^{o}C . (d) $R_{1\rho}$ on- and off-resonance relaxation dispersion profiles (powers in inset) for A5 C8 and C1' sites at 17°C and 26 ^{o}C showing best fits to the asymmetric two-site exchange (eq. 3.1). (see text for details)

Thus, excited state HG base pairs are not restricted to CA/TG steps but also include TA steps, and they can be conformationally captured by recognition factors. Moreover, the weaker binding affinity of TBP to an HG-containing mutant TATA box, where a WC base pair would not be tolerated, versus the wild-type sequence with all-WC base pairing ($\Delta\Delta G$ ~3 kcal/mol)[84] could be attributed to conformational selection of a low-populated HG base pair and could be further correlated with transcriptional regulation of gene repression[85].

3.3.6 Suppressing chemical exchange at CA/TG by destabilizing HG base pairs

As a final step towards assigning the excited states at CA/TG steps as HG base pairs, we performed a "negative control" experiment where we suppressed the potential for HG base pair formation at the appreciable populations or at the timescales (~ 1 ms to ~ 5 μs) required for detection by NMR relaxation dispersion. We installed a chemically modified 7-deaza nucleotide that substitutes a protonated carbon C7 for the purine N7 and is isosteric to A or G, in place of A16 or G10 in unlabeled A_6-DNA (7-deazaadenine in A_6-DNA[7dNA16] or 7-deazaguanine in A_6-DNA[7dNG10]). This single-atom substitution eliminates the ability for one of the HG base pair
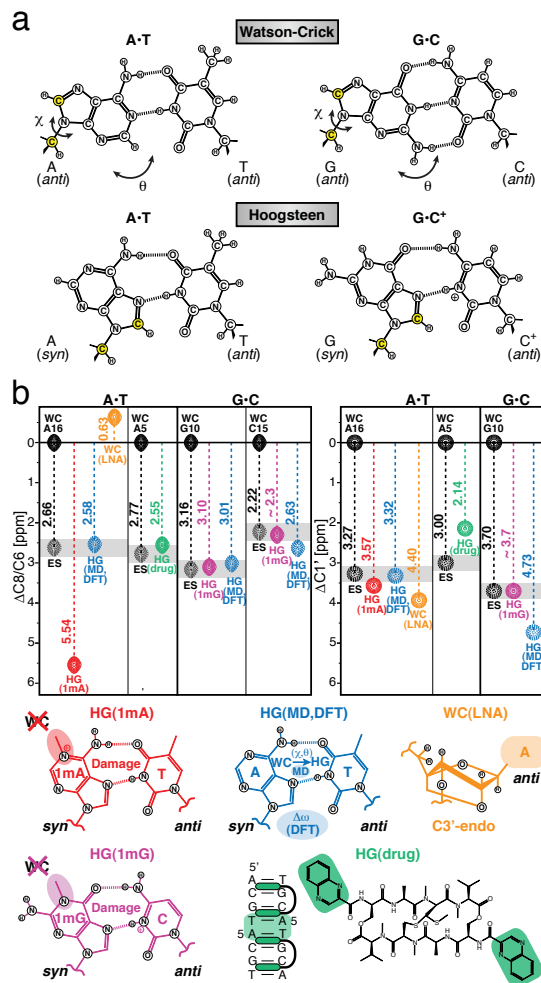
Figure 3.9: Summary of chemical shift matching for assignment of the excited state HG base pairs. (a) Chemical structures for WC and HG A•T and G•C+ base pairs showing C8 and C1' (yellow). (b) Depiction of C8 and C1' chemical shifts relative to WC for the ES (ES, grey); an *N*1-methyladenine modified A_6-DNA (HG(1mA), red), an *N*1-methylguanine modified A_6-DNA (HG(1mG), violet), and an echinomycin-bound DNA (HG(drug), green) forming HG base pairs; a simulated A16•T9 HG base pair in A_6-DNA (HG(MD), blue); a C3'-endo locked A_6-DNA (WC(LNA), orange); and representative cartoons.

113

hydrogen bonds, A/G N7---H3N3 T/C, to form, which could cost up to several kcal/mol of stability relative to the optimal HG geometry. For the *syn*-purine conformation, this could be compensated to a certain degree by creation of a staggered geometry with two less favorable hydrogen bonds, for example between G O6---H4a/bN4 C for G•C and T O4---H6a/bN6 A for A•T. As a result, we anticipate that excursions from WC to HG base pairs would be highly unfavorable and we would not be able to detect chemical exchange at carbon sites that otherwise show pronounced μs dynamics in A_6-DNA (Figure 3.3).

A spectral overlay of unmodified and modified A_6-DNA (Figure 3.10) showed the anticipated upfield chemical shift perturbations at the 7dNA C8H8 (~ 18 ppm) and C2H2 (~ 1.5 ppm), and 7dNG C8H8 (~ 18 ppm) as a result of the single-atom substitution (N7 → C7H7) that alters the electronic environment of the heterocyclic ring. Surprisingly, changes in base and backbone chemical shifts extended to the opposite T9 or C15 and the two surrounding base pairs (G10•C15 and A17•T8 for A_6-DNA^{7dNA16}; G11•C14 and A16•T9 for A_6-DNA^{7dNG10}). For A_6-DNA^{7dNA16}, the most noticeable neighboring shifts were for the sugar C1'H1' of C15 on the 5' side and the base C8H8 of A17 on the 3' side of 7dNA16. A_6-DNA^{7dNG10} was characterized by even larger perturbations to the backbone at the modified 7dNG10 and T9 and transmitting at least two base pairs away in each direction, to A17•T8 and the near terminus (G11, C13 C1'H1'), the latter being more susceptible to changes in conformation due to diminished stability. Although we cannot compare directly the resonance intensities C8H8 of 7dNA16 and A16, those of the deoxyribose C1'H1' site were slightly higher than for neighboring adenines, indicating the presence of increased sugar disorder.

For consistency with previous data on damaged A_6-DNA (above), we collected on-resonance $R_{1\rho}$ relaxation dispersion at pH 6.8 for A_6-DNA^{7dNA16} and at lower pH 5.2 for A_6-DNA^{7dNG10}, where G10 exhibits chemical exchange. Figure 3.10 shows that, as expected, μs dynamics were suppressed simultaneously at C8 and C1' sites of 7dNA16 (A_6-DNA^{7dNA16}), C8 site of 7dNG10 and C6 site of C15 (A_6-DNA^{7dNG10}). At the same time, the presence of chemical exchange was not altered in the unmodified A16 C8 neighboring the 7dNG10•C15 base pair in A_6-DNA^{7dNG10} (Figure 3.10b), which served as an internal control and provided an estimate for the current sensitivity in unlabeled DNA samples.

Our "negative control" was successful to demonstrate that DNA excited states do not form at appreciable populations or fast/slow enough exchange rates for NMR detection when we selectively disrupt the potential for A•T or G•C$^+$ HG base pairs to form. We "surgically" replaced a single atom, N7 → C7, which would most significantly impair HG base pairs and, as far as we can tell, should not disfavor the formation of base
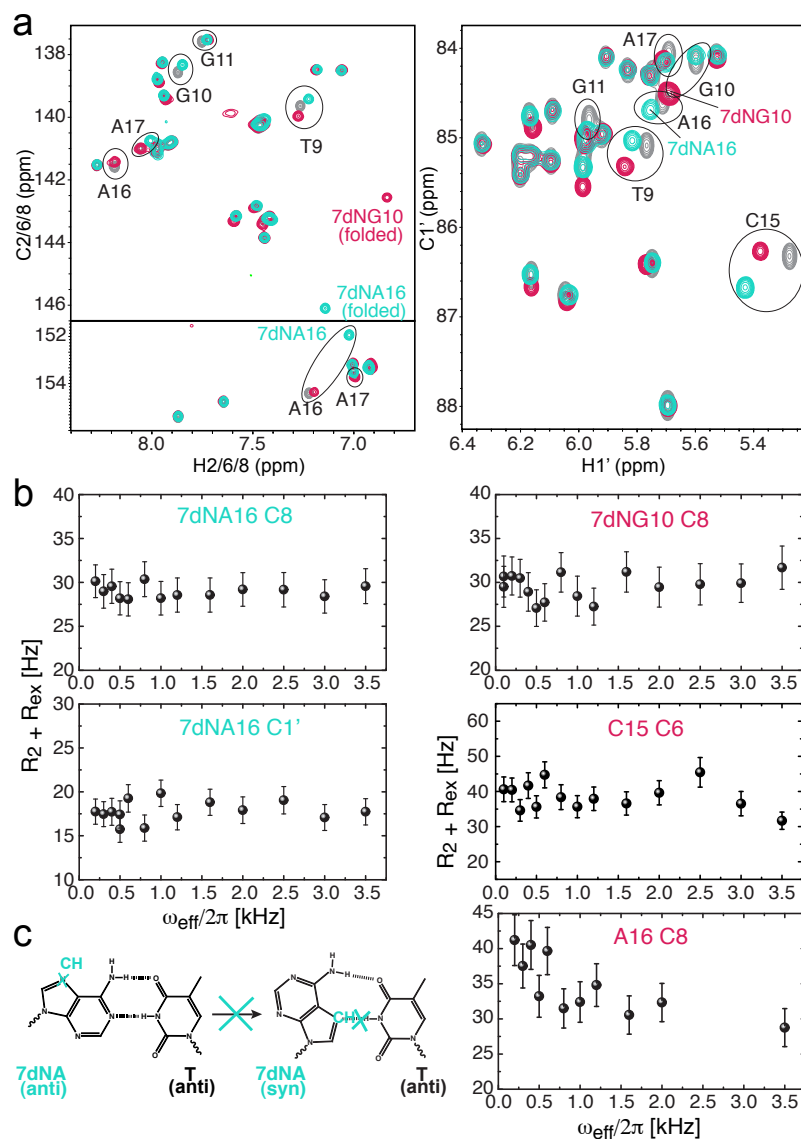
Figure 3.10: Suppressing excited state HG base pairs by chemical modification. (a) 2D ^{1}H,^{13}C HSQC spectral overlay of unlabeled A$_6$-DNA (grey), A$_6$-DNA^{7dNA16} (cyan), and A$_6$-DNA^{7dNG10} (magenta). (b) $R_{1\rho}$ on-resonance relaxation dispersion profiles for 7dNA16 C8/C1' (A$_6$-DNA^{7dNA16}), 7dNG10 C8 and C15 (A$_6$-DNA^{7dNG10}) showing no chemical exchange, and the control A16 (A$_6$-DNA^{7dNG10}) showing dispersion. (c) N7 to C7H7 substitution that disrupts HG base pairs formation and correlates with suppressed chemical exchange.

pair open states – the other significant candidate for the ES conformation. However, the single-atom substitution could have multiple complex effects on WC and open base pairs that we cannot fully comprehend. 7-deazapurines can destabilize WC base pairs and DNA duplexes by preventing the interaction of water molecules with the H-bond acceptor N7 and by disrupting the π-electron system of the heterocyclic ring that affects both base pairing strength and vertical stacking. Based on these potential effects, we expect the WC base pair to be destabilized to a certain degree. At the same time, the WC base pair could open less readily because the opening mechanism involves protonation (or other interactions) at N7. At present, we suspect that these complex effects are small compared to the extent of HG destabilization by 7-deazapurines, which is why we are able to effectively suppress chemical exchange.

3.4 Conclusion

By using a newly developed ^{13}C relaxation dispersion experiment, we have observed transient sequence-specific excursions away from Watson-Crick base pairing at CA and TA steps inside canonical duplex DNA towards low-populated and short-lived A•T and G•C excited state. The formation of these excited states at CA/TG steps is modulated by increasing the length of a 3' neighboring A-tract in a way that is consistent with transition state destabilization for G•C and stabilization for A•T (faster transition for G•C and slower transition for A•T), which might be explained by more pronounced structural deformations at the 5' junction of longer A-tracts seen in high-resolution structures[86]. Moreover, the pronounced sugar dynamics at CA/TG steps may play an important role in the initiation of base flipping events that occur on slower timescales. To determine the nature of ES at atomic resolution, we ambitiously came up with a strategy that included (i) thorough kinetic-thermodynamic and chemical shift characterization of the chemical exchange process combined with (ii) steered molecular dynamic for a proposed structural transition and pathway, (iii) matching of NMR-derived ES chemical shifts with chemical shifts for the proposed ES from DFT predictions and from chemically modified/drug-bound DNAs that trap the proposed ES as their ground state, and finally (iv) verifying suppression of the proposed ES formation by selective destabilization.

116

All of our data, based on NMR and computation, strongly suggests that these excited states are in fact transient A•T and G•C$^+$ Hoogsteen base pairs that can be well accommodated inside a B-DNA duplex. Kinetic-thermodynamic profiles point to a species with similar stability to WC base pairs but whose formation requires complete disruption of WC interactions, consistent with transition to an ES HG base pair by a 180-degree purine rotation and not favoring transition to an ES base open state. NMR chemical shifts for damaged and drug-bound DNA that trap A•T and G•C$^+$ HG base pairs as well as computed DFT chemical shifts for MD-generated HG base pairs present almost perfect agreement to the relaxation dispersion data. Last but not least, chemically modified DNAs with suppressed ability to form HG base pairs do not exhibit the chemical exchange pattern for ES.

Our data strongly argue that the DNA double helix codes for a pre-existing WC-to-HG equilibrium, with HG base pairs representing an accessible and energetically competent alternative to WC base pairing that present very distinct electrostatic and hydrophobic signatures. This makes it possible to trap HG base pairs by interactions with cellular triggers, thereby expanding the structural and functional diversity of the double helix beyond that which can be achieved based on an alphabet of only WC base pairing. There are several examples of transcription factors including TBP[10] and p53 tumor suppressor[11] that specifically recognize HG base pairs embedded in different WC contexts, where the modulation in binding affinity, conceivably, by an HG base pair could even be correlated with an essential biological function[85]. In addition, HG base pairs are often trapped by oxidative and alkylation lesions[15], providing unique recognition signals for repair enzymes in search of damage sites in a sea of undamaged DNA. Transient formation of HG base pairs inside B-DNA may also serve to promote non-canonical structures such as contiguous HG motifs, especially in tandem CA and TA repeats, or more dramatic transformations to Z-DNA[87], and may well exist in much greater abundance for native genomic DNA, which is under torsional stress in the cellular environment. The methods presented here provide a general strategy for detecting and characterizing excited states in DNA and RNA, which we predict will be abundant in the genome and constitute another transient layer of the genetic code.

3.4 References

1. Watson, J.D. & Crick, F.H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-8 (1953).
2. Coman, D. & Russu, I.M. A nuclear magnetic resonance investigation of the energetics of basepair opening pathways in DNA. *Biophys J* **89**, 3285-92 (2005).
3. Record, M.T., Jr. et al. Double helical DNA: conformations, physical properties, and interactions with ligands. *Annu Rev Biochem* **50**, 997-1024 (1981).
4. Koudelka, G.B., Mauro, S.A. & Ciubotaru, M. Indirect readout of DNA sequence by proteins: the roles of DNA sequence-dependent intrinsic and extrinsic forces. *Prog Nucleic Acid Res Mol Biol* **81**, 143-77 (2006).
5. Rohs, R. et al. The role of DNA shape in protein-DNA recognition. *Nature* **461**, 1248-53 (2009).
6. Segal, E. et al. A genomic code for nucleosome positioning. *Nature* **442**, 772-8 (2006).
7. Saiz, L. & Vilar, J.M. DNA looping: the consequences and its control. *Curr Opin Struct Biol* **16**, 344-50 (2006).
8. Richmond, T.J. & Davey, C.A. The structure of DNA in the nucleosome core. *Nature* **423**, 145-50 (2003).
9. Wang, A.H. et al. Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282**, 680-6 (1979).
10. Patikoglou, G.A. et al. TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev* **13**, 3217-30 (1999).
11. Kitayner, M. et al. Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs. *Nat Struct Mol Biol* **17**, 423-9 (2010).
12. Ughetto, G. et al. A comparison of the structure of echinomycin and triostin A complexed to a DNA fragment. *Nucleic Acids Res* **13**, 2305-23 (1985).
13. Seaman, F.C. & Hurley, L. Interstrand cross-linking by bizelesin produces a Watson-Crick to Hoogsteen base pairing transition region in d(CGTAATTACG)2. *Biochemistry* **32**, 12577-85 (1993).
14. Yang, H., Zhan, Y., Fenn, D., Chi, L.M. & Lam, S.L. Effect of 1-methyladenine on double-helical DNA structures. *FEBS Lett* **582**, 1629-33 (2008).
15. Shanmugam, G., Kozekov, I.D., Guengerich, F.P., Rizzo, C.J. & Stone, M.P. Structure of the 1,N2-ethenodeoxyguanosine adduct opposite cytosine in duplex DNA: Hoogsteen base pairing at pH 5.2. *Chem Res Toxicol* **21**, 1795-805 (2008).
16. Palmer, A.G., 3rd. NMR characterization of the dynamics of biomacromolecules. *Chem Rev* **104**, 3623-40 (2004).
17. Korzhnev, D.M. & Kay, L.E. Probing invisible, low-populated States of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding. *Acc Chem Res* **41**, 442-51 (2008).
18. Boehr, D.D., Nussinov, R. & Wright, P.E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* **5**, 789-96 (2009).
19. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964-72 (2007).

20. Korzhnev, D.M., Religa, T.L., Banachewicz, W., Fersht, A.R. & Kay, L.E. A transient and low-populated protein-folding intermediate at atomic resolution. *Science* **329**, 1312-6 (2010).

21. Johnson, J.E., Jr. & Hoogstraten, C.G. Extensive backbone dynamics in the GCAA RNA tetraloop analyzed using 13C NMR spin relaxation and specific isotope labeling. *J Am Chem Soc* **130**, 16757-69 (2008).

22. Hansen, A.L., Nikolova, E.N., Casiano-Negroni, A. & Al-Hashimi, H.M. Extending the range of microsecond-to-millisecond chemical exchange detected in labeled and unlabeled nucleic acids by selective carbon R(1rho) NMR spectroscopy. *J Am Chem Soc* **131**, 3818-9 (2009).

23. Shajani, Z. & Varani, G. 13C relaxation studies of the DNA target sequence for hhai methyltransferase reveal unique motional properties. *Biochemistry* **47**, 7617-25 (2008).

24. Kennedy, M.A., Nuutero, S.T., Davis, J.T., Drobny, G.P. & Reid, B.R. Mobility at the TpA cleavage site in the T3A3-containing AhaIII and PmeI restriction sequences. *Biochemistry* **32**, 8022-35 (1993).

25. McAteer, K., Ellis, P.D. & Kennedy, M.A. The effects of sequence context on base dynamics at TpA steps in DNA studied by NMR. *Nucleic Acids Res* **23**, 3962-6 (1995).

26. McAteer, K. & Kennedy, M.A. NMR evidence for base dynamics at all TpA steps in DNA. *J Biomol Struct Dyn* **17**, 1001-9 (2000).

27. Kojima, C., Ulyanov, N.B., Kainosho, M. & James, T.L. Slow motion in the CAA*TTG sequence of a DNA decamer duplex studied by NMR. *Biochemistry* **40**, 7239-46 (2001).

28. Lane, A.N. & Peck, B. Conformational flexibility in DNA duplexes containing single G.G mismatches. *Eur J Biochem* **230**, 1073-87 (1995).

29. Nikolova, E.N. et al. Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **470**, 498-502 (2011).

30. Gilbert, D.E., van der Marel, G.A., van Boom, J.H. & Feigon, J. Unstable Hoogsteen base pairs adjacent to echinomycin binding sites within a DNA duplex. *Proc Natl Acad Sci U S A* **86**, 3006-10 (1989).

31. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-93 (1995).

32. Palmer, A.G., 3rd & Massi, F. Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chem Rev* **106**, 1700-19 (2006).

33. Miloushev, V.Z. & Palmer, A.G., 3rd. R(1rho) relaxation for two-site chemical exchange: general approximations and some exact solutions. *J Magn Reson* **177**, 221-7 (2005).

34. Ferry, J., Grandine, L. & Fitzgerald, E. The Relaxation Distribution Function of Polyisobutylene in the Transition from Rubberlike to Glasslike Behavior and its Dependence on Temperature. *Physical review* **91**, 217-217 (1953).

35. Denisov, V.P., Peters, J., Horlein, H.D. & Halle, B. Using buried water molecules to explore the energy landscape of proteins. *Nat Struct Biol* **3**, 505-9 (1996).

36. Olson, W.K. et al. A Standard Reference Frame for the Description of Nucleic Acid Base pair Geometry. *J. Mol. Biol.* **131**, 9 (2001).

37. Abrescia, N.G., Gonzalez, C., Gouyette, C. & Subirana, J.A. X-ray and NMR studies of the DNA oligomer d(ATATAT): Hoogsteen base pairing in duplex DNA. *Biochemistry* **43**, 4092-100 (2004).

38. Aishima, J. et al. A Hoogsteen base pair embedded in undistorted B-DNA. *Nucleic Acids Res* **30**, 5244-52 (2002).

39. MacKerell, A.D., Jr., Banavali, N. & Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **56**, 257-65 (2000).

40. Chocholousova, J. & Feig, M. Implicit solvent simulations of DNA and DNA-protein complexes: agreement with explicit solvent vs experiment. *J Phys Chem B* **110**, 17240-51 (2006).

41. Feig, M. et al. Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J Comput Chem* **25**, 265-84 (2004).

42. Lee, M.S., Feig, M., Salsbury, F.R., Jr. & Brooks, C.L., 3rd. New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations. *J Comput Chem* **24**, 1348-56 (2003).

43. Nose, S. A Unified Formulation of the Constant Temperature Molecular-Dynamics Methods. *Journal of Chemical Physics* **81**, 511-519 (1984).

44. Hoover, W.G. Canonical Dynamics - Equilibrium Phase-Space Distributions. *Physical Review A* **31**, 1695-1697 (1985).

45. Fischer, S. & Karplus, M. Conjugate Peak Refinement - an Algorithm for Finding Reaction Paths and Accurate Transition-States in Systems with Many Degrees of Freedom. *Chem. Phys. Lett.* **194**, 252-261 (1992).

46. Song, K. et al. An Improved Reaction Coordinate for Nucleic Acid Base Flipping Studies. *J. Chem. Theory Comput.* **5**, 3105-3113 (2009).

47. M. J. Frisch, G.W.T., H. B. Schlegel, G. E. Scuseria, et al. Gaussian 03, Revision C.02. (Gaussian, Inc., Wallingford CT, 2004).

48. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-12 (2004).

49. Korzhnev, D.M., Orekhov, V.Y. & Kay, L.E. Off-resonance R(1rho) NMR studies of exchange dynamics in proteins with low spin-lock fields: an application to a Fyn SH3 domain. *J Am Chem Soc* **127**, 713-21 (2005).

50. Hansen, A.L. & Al-Hashimi, H.M. Dynamics of large elongated RNA by NMR carbon relaxation. *J Am Chem Soc* **129**, 16072-82 (2007).

51. Dethoff, E.A. et al. Characterizing complex dynamics in the transactivation response element apical loop and motional correlations with the bulge by NMR, molecular dynamics, and mutagenesis. *Biophys J* **95**, 3906-15 (2008).

52. Korzhnev, D.M., Skrynnikov, N.R., Millet, O., Torchia, D.A. & Kay, L.E. An NMR experiment for the accurate measurement of heteronuclear spin-lock relaxation rates. *J Am Chem Soc* **124**, 10743-53 (2002).

53. Massi, F., Johnson, E., Wang, C., Rance, M. & Palmer, A.G., 3rd. NMR R1 rho rotating-frame relaxation with weak radio frequency fields. *J Am Chem Soc* **126**, 2247-56 (2004).

54. Hansen, A.L. Development of 13C Nuclear Magnetic Resonance Methods for Studying the Structural Dynamics of Nucleic Acids in Solution. (2009).

55. Pardi, A., Morden, K.M., Patel, D.J. & Tinoco, I., Jr. Kinetics for exchange of imino protons in the d(C-G-C-G-A-A-T-T-C-G-C-G) double helix and in two similar

helices that contain a G . T base pair, d(C-G-T-G-A-A-T-T-C-G-C-G), and an extra adenine, d(C-G-C-A-G-A-A-T-T-C-G-C-G). *Biochemistry* **21**, 6567-74 (1982).

56. Gueron, M., Kochoyan, M. & Leroy, J.L. A single mode of DNA base pair opening drives imino proton exchange. *Nature* **328**, 89-92 (1987).

57. Perez, A., Luque, F.J. & Orozco, M. Dynamics of B-DNA on the microsecond time scale. *J Am Chem Soc* **129**, 14739-45 (2007).

58. Kochoyan, M., Lancelot, G. & Leroy, J.L. Study of structure, base pair opening kinetics and proton exchange mechanism of the d-(AATTGCAATT) self-complementary oligodeoxynucleotide in solution. *Nucleic Acids Res* **16**, 7685-702 (1988).

59. Leroy, J.L., Kochoyan, M., Huynh-Dinh, T. & Gueron, M. Characterization of base pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton. *J Mol Biol* **200**, 223-38 (1988).

60. Folta-Stogniew, E. & Russu, I.M. Sequence dependence of base pair opening in a DNA dodecamer containing the CACA/GTGT sequence motif. *Biochemistry* **33**, 11016-24 (1994).

61. Chen, C. & Russu, I.M. Sequence-dependence of the energetics of opening of AT basepairs in DNA. *Biophys J* **87**, 2545-51 (2004).

62. Coman, D. & Russu, I.M. Base pair opening in three DNA-unwinding elements. *J Biol Chem* **280**, 20216-21 (2005).

63. Every, A.E. & Russu, I.M. Influence of magnesium ions on spontaneous opening of DNA base pairs. *J Phys Chem B* **112**, 7689-95 (2008).

64. Varnai, P., Canalia, M. & Leroy, J.L. Opening mechanism of G.T/U pairs in DNA and RNA duplexes: a combined study of imino proton exchange and molecular dynamics simulation. *J Am Chem Soc* **126**, 14659-67 (2004).

65. Rinnenthal, J., Klinkert, B., Narberhaus, F. & Schwalbe, H. Direct observation of the temperature-induced melting process of the Salmonella fourU RNA thermometer at base pair resolution. *Nucleic Acids Res* (2010).

66. Greene, K.L., Wang, Y. & Live, D. Influence of the glycosidic torsion angle on 13C and 15N shifts in guanosine nucleotides: investigations of G-tetrad models with alternating syn and anti bases. *J Biomol NMR* **5**, 333-8 (1995).

67. Xu, X. & Au-Yeung, S. Investigation of chemical shift and structure relationships in nucleic acids using NMR and density functional theory methods. *J. Phys. Chem. B* **104**, 5641-5650 (2000).

68. Foloppe, N., Hartmann, B., Nilsson, L. & MacKerell, A.D., Jr. Intrinsic conformational energetics associated with the glycosyl torsion in DNA: a quantum mechanical study. *Biophys J* **82**, 1554-69 (2002).

69. Stolarski, R., Dudycz, L. & Shugar, D. NMR studies in the syn-anti dynamic equilibrium in purine nucleosides and nucleotides. *Eur J Biochem* **108**, 111-21 (1980).

70. Izatt, R.M., Christensen, J.J. & Rytting, J.H. Sites and thermodynamic quantities associated with proton and metal ion interaction with ribonucleic acid, deoxyribonucleic acid, and their constituent bases, nucleosides, and nucleotides. *Chem Rev* **71**, 439-81 (1971).

71. Ghosal, G. & Muniyappa, K. Hoogsteen base pairing revisited: resolving a role in normal biological processes and human diseases. *Biochem Biophys Res Commun* **343**, 1-7 (2006).

72. Abrescia, N.G., Thompson, A., Huynh-Dinh, T. & Subirana, J.A. Crystal structure of an antiparallel DNA fragment with Hoogsteen base pairing. *Proc Natl Acad Sci U S A* **99**, 2806-11 (2002).

73. Nair, D.T., Johnson, R.E., Prakash, L., Prakash, S. & Aggarwal, A.K. Human DNA polymerase iota incorporates dCTP opposite template G via a G.C + Hoogsteen base pair. *Structure* **13**, 1569-77 (2005).

74. Powell, S.W., Jiang, L. & Russu, I.M. Proton exchange and base pair opening in a DNA triple helix. *Biochemistry* **40**, 11065-72 (2001).

75. Cubero, E., Luque, F.J. & Orozco, M. Theoretical study of the Hoogsteen-Watson-Crick junctions in DNA. *Biophys J* **90**, 1000-8 (2006).

76. Sau, A.K. et al. Evidence for A+(anti)-G(syn) mismatched base pairing in d-GGTAAGCGTACC. *FEBS Lett* **377**, 301-5 (1995).

77. McConnell, B. & Politowski, D. Buffer catalysis of amino proton-exchange in compounds of adenosine, cytidine and their endocyclic N-methylated derivatives. *Biophysical Chemistry* **20**, 135-148 (1984).

78. Nonin, S., Leroy, J.L. & Gueron, M. Acid-induced exchange of the imino proton in G.C pairs. *Nucleic Acids Res* **24**, 586-95 (1996).

79. Packer, M.J., Dauncey, M.P. & Hunter, C.A. Sequence-dependent DNA structure: dinucleotide conformational maps. *J Mol Biol* **295**, 71-83 (2000).

80. Packer, M.J. & Hunter, C.A. Sequence-dependent DNA structure: the role of the sugar-phosphate backbone. *J Mol Biol* **280**, 407-20 (1998).

81. Yanagi, K., Prive, G.G. & Dickerson, R.E. Analysis of local helix geometry in three B-DNA decamers and eight dodecamers. *J Mol Biol* **217**, 201-14 (1991).

82. Lu, L., Yi, C., Jian, X., Zheng, G. & He, C. Structure determination of DNA methylation lesions N1-meA and N3-meC in duplex DNA using a cross-linked protein-DNA system. *Nucleic Acids Res* **38**, 4415-25 (2010).

83. Gilbert, D.E. & Feigon, J. The DNA sequence at echinomycin binding sites determines the structural changes induced by drug binding: NMR studies of echinomycin binding to [d(ACGTACGT)]2 and [d(TCGATCGA)]2. *Biochemistry* **30**, 2483-94 (1991).

84. Hoopes, B.C., LeBlanc, J.F. & Hawley, D.K. Contributions of the TATA box sequence to rate-limiting steps in transcription initiation by RNA polymerase II. *J Mol Biol* **277**, 1015-31 (1998).

85. Meyer, T., Carlstedt-Duke, J. & Starr, D.B. A weak TATA box is a prerequisite for glucocorticoid-dependent repression of the osteocalcin gene. *J Biol Chem* **272**, 30709-14 (1997).

86. MacDonald, D., Herbert, K., Zhang, X., Pologruto, T. & Lu, P. Solution structure of an A-tract DNA bend. *J Mol Biol* **306**, 1081-98 (2001).

87. Segers-Nolten, G.M., Sijtsema, N.M. & Otto, C. Evidence for Hoogsteen GC base pairs in the proton-induced transition from right-handed to left-handed poly(dG-dC).poly(dG-dC). *Biochemistry* **36**, 13241-7 (1997).

Chapter 4

Characterizing the dependence of transient Hoogsteen base pairs on pH, cations, and base modifications, and their cooperativity of formation

4.1 Introduction

Earlier, we discovered the existence of extensive chemical exchange at CA/TG and TA steps of duplex DNA and identified the process to be transient formation of A•T and G•C$^+$ Hoogsteen base pair (Chapter 3). In addition, having observed ground state HG base pairs in DNA bound to proteins and small molecules in various sequences, we hypothesized that they might exist *in vivo* in populations appreciable enough to present an alternative to WC base pairs for selective protein and small molecule recognition. This other "face" of DNA has the potential to emerge at any point, but how often that happens can be modulated by intrinsic DNA modifications, conformational switches induced by superhelical stress, various external factors such as salts, small molecule and protein ligands that are in constant flux in the cellular milieu.

To better understand how HG base pairs form and react to their environment, it is essential that we characterize their intrinsic properties and response to changes in conditions such as pH, monovalent and divalent cations. Small variations in pH are expected to modulate significantly the occupancies of G•C$^+$ HG base pairs that require a protonation to achieve optimal stability. The condensation of counterions (protons and other cations) on the DNA surface, driven by favorable electrostatic interactions with backbone phosphate groups, can modify the stability of the duplex, local groove charge potentials, and affect the intrinsic basicity/acidity of nucleobase atoms that have different geometries and electrostatics in the WC and HG state[1]. Ultimately, this can alter the equilibrium and kinetics for transition to and from excited state HG geometry. Even more pronounced effects could be caused by divalent Mg^{2+} ions that are known to form long-lived complexes with DNA bases and sugars through direct or water-mediated coordination[2-5]. The proton/cation affinity of nucleobase sites is not only modulated by

external factors, but also by internal base modifications that can be endogenous to the cell. One such modification that occurs *in vivo* at extremely high abundance (~ 1 %) in the mammalian genome is 5-methylcytosine[6,7] (5mC). Cytosine C5-methylation is maintained through multiple cell divisions and plays a critical role in the epigenetic control of chromatin condensation and gene expression, while deregulation of methylation is a hallmark of cancer[8]. C5-methylcytosine commonly occurs at CG steps in DNA, but has been observed as the predominant marker in CA steps of embryonic stem cells[9]. Concurrent with studies of pH and ion dependence, we probe the effect of 5mC on WC-to-HG equilibria in the following investigation.

Another eminent question is whether HG base pairs can form cooperatively and at higher proportions, especially in repetitive DNA segments comprised of CA/TG and TA base pair steps. If so, this can transform dramatically the phase of DNA in poly-CA sequences that belong to microsatellite DNA and constitute approximately 0.25 % of the genome! CA-repeats affect the frequencies of intraplasmid bacteria recombination[10], homologous recombination in yeast[11] and are hotspots for human genetic recombination[12]. The increased possibility of oligonucleotide invasion within linear poly-CA duplexes[13] has been associated with their conformational plasticity[14,15] and ability to assume non-canonical conformations such as Z-DNA[16]. All of these could be linked to the greater potential of consecutive CA/TG steps to form non-canonical HG states. Moreover, based on MD simulations, we anticipate for HG base pairs at TA steps, analogous to CA steps, to achieve better stacking than HG-WC junctions[17]. To study the cooperativity of HG base pairs, we examine how CA-repeat sequences and the trapping of HG base pairs or reducing their occurrence at CA/TG steps by nucleobase modifications can influence the populations of excited HG states and the rates at which they form and disappear.

4.2 Materials and Methods
4.2.1 Preparation and resonance assignment of unlabeled and ^{13}C/^{15}N-labeled DNA

Unlabeled and ^{13}C/^{15}N-labeled DNA oligos and duplexes were prepared and assigned as described in Section 2.2.1 using sample-specific buffer conditions. DNA samples were exchanged into the desired buffer by using Amicon Ultra-4 centrifugal

filters (3 kDa cutoff) 3-5X. $^{13}C/^{15}N$-labeled DNA. The desired pH was adjusted by buffer exchange as above or by directly titrating a dilute HCl/NaOH solution in the DNA sample, in either case the pH was carefully checked using a pH meter (Thermo Scientific). Unlabeled DNA samples were annealed at 2.0-4.0 mM concentrations. Semi-labeled DNA samples were prepared by titrating the unlabeled strand directly into the NMR tube containing the $^{13}C/^{15}N$-labeled strand and monitoring the disappearance of ssDNA using quick HSQCs. All NMR experiments were performed on a Bruker Avance 600 MHz NMR spectrometer equipped with a 5mm triple-resonance cryogenic probe.

4.2.2. Selective ^{13}C $R_{1\rho}$ relaxation dispersion

$R_{1\rho}$ relaxation dispersion profiles were measured as described in Section 3.2.

4.2.3. Analysis of pH dependence of $R_{1\rho}$ relaxation dispersion

Carbon $R_{1\rho}$ relaxation dispersion data for A16 and G10 C8 was collected as a function of pH. The apparent pH dependence for the G•C excited state population was analyzed based on equation 4.1 below derived for an effective two-state equilibrium from the Henderson–Hasselbalch equation, which subsumes the proposed three-state and four-state equilibria in Scheme 2 and 3:

Scheme 1:

$$WC + H^+ \xrightleftharpoons{K_{obs}} HG^+$$
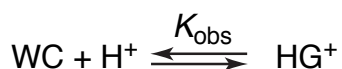
$$p(HG^+) = p_B = \frac{10^{pK_{obs}-pH}}{1+10^{pK_{obs}-pH}} \tag{4.1}$$

where $p(HG^+)$ is the excited state population (p_B) measured by $R_{1\rho}$ relaxation dispersion, and K_{obs} is the observed equilibrium constant that subsumes the protonation and conformational change. We can also introduce a Hill coefficient, n, as a multiple of the exponent ($n(pK_{obs} - pH)$) to assess the extent of cooperativity of the reaction (n < 1

anticooperative, n = 1 non-cooperative, n >1 positively cooperative). More precisely, the transition from WC to HG$^+$ base pair can be decomposed into three- or four-state models, linked equilibria, which involve protonation events coupled to conformational changes:

Scheme 2: 3-state

$$WC \; \underset{}{\overset{K_1}{\rightleftharpoons}} \; HG + H^+ \; \underset{}{\overset{K_2}{\rightleftharpoons}} \; HG^+$$
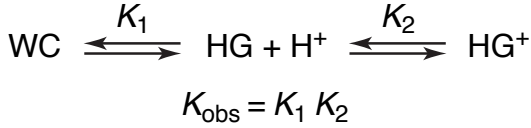
$$K_{obs} = K_1 \, K_2$$

$$p(HG^+) = p_B = \frac{K_1 10^{pK_2 - pH}}{1 + K_1 + K_1 10^{pK_2 - pH}} \tag{4.2}$$

Scheme 3: 4-state

$$
\begin{array}{ccc}
WC & \overset{K_1}{\rightleftharpoons} & HG \\
+ H^+ & \overset{K_{obs}}{\searrow} & + H^+ \\
K_4 \updownarrow & & \updownarrow K_2 \\
WC^+ & \underset{K_3}{\rightleftharpoons} & HG^+
\end{array}
$$
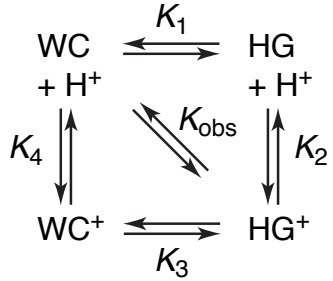
$$K_{obs} = K_1 \, K_2$$

$$p(HG^+) = p_B = \frac{K_1 10^{pK_2 - pH}}{1 + K_1 + 10^{pK_4 - pH} + K_1 10^{pK_2 - pH}} \tag{4.3}$$

The following definitions were used to derive the equations:

$$K_1 = \frac{p_{HG}}{p_{WC}}, \; K_2 = \frac{p_{HG+}}{p_{HG}H^+}, \; K_3 = \frac{p_{HG+}}{p_{WC+}}, \; K_4 = \frac{p_{WC+}}{p_{WC}H^+}, \; K_1 K_2 = K_3 K_4$$

$$p_{WC} + p_{WC+} + p_{HG} + p_{HG+} = 1$$

$$pH = -\log[H^+]$$
$$pK_i = \log[K_i]$$

The above two models were analyzed with the simplification that $K_1 \sim 10^{-3}$ as the highest value and $K_4 \ll 10^{-3}$ since protonation of the WC base pair at C N3 is highly unfavorable in the pH range investigated (4.0 - 7.6), making the term $10^{pK_4-pH} \ll 1$ and negligible.

<u>4.2.4. DFT chemical shift predictions of protonated bases</u>

DFT chemical shift calculations were performed as described in Section 3.2.6. The singly protonated species for nucleobase A, G, T, and C were generated manually by addition a hydrogen at each available nitrogen using UCSF Chimera[18].

4.3 Results and Discussion

<u>4.3.1 pH dependence of WC-to-HG transition</u>

In Chapter 3, the high pH dependence of relaxation dispersion for G•C base pairs, absent in A•T base pairs for the 5.4 to 7.6 pH range, suggested the involvement of a protonatable group and formation of G•C$^+$ HG base pairs. Specifically, the excited state populations became larger as the proton concentration was increased (lower pH). The most likely group to be protonated in G•C HG geometry based on pK_a values for free nucletides was C N3 ($pK_a \sim 4.2 - 4.4$) and not G N7 ($pK_a \sim 2.9 - 3.3$)[19]. Protonation of C N3 would result in the favorable formation of a second H-bond between G and C that stabilizes the HG base pair, and would affect chemical shifts at cytosine carbons, specifically C6, based on DFT calculations (Chapter 3, Appendix 2), that would lead to relaxation dispersion. C N3 protonation accompanies the third-strand binding to form HG interactions in the major grove of DNA triplexes[20] and formation of CC duplexes. HG base pair formation in A•T is not generally associated with a change in the protonation state of T N3, and therefore we do not expect to observe relaxation dispersion at T C6. If instead the relaxation dispersion data were to report only on a protonation event and not a WC-to-HG transition, than G N7 in G•C and A N7 in A•T[21] would be the best proton acceptors based on pK_a predictions[19,21]. If G N7 was protonated in the context of a WC base pair, then we do not expect to see a significant impact on the chemical shift of the opposite C C6 carbon, or on the G C1' as judged from DFT predictions, translating to lack of chemical exchange. If A N7 is protonated in the WC configuration, then we

would also not predict sizeable chemical exchange at T C6 or A C1' based on DFT calculations for the latter.

Let us briefly review our findings from Chapter 3. Several lines of evidence come together to support protonation of C N3 in the context of an HG base pair. First, we observe a direct correlation between chemical exchange at C C6 and G C8/C1' that increases simultaneously with increasing acidity. Second, a protonated C N3 could most reasonably explain the large downfield shift for the excited state chemical shift in C C6 that has previously been observed in triplex DNA[20] and predicted in Chapter 2 by DFT calculations. Lastly, the stability of the trapped 1mG•C$^+$ HG base pair was dramatically increased by lowering pH and yielded an excellent chemical shift agreement with relaxation dispersion data for C C6 in unmodified G•C. An alternative explanation for chemical exchange at C C6 could be a cytosine rotation around the glycosidic bond from *anti* to *syn* geometry. However, cytosine rotation by itself that achieves an interhelical *syn*-geometry would not allow for any productive base pairing between C and G. Cytosine rotation to form an unstacked extrahelical conformation was discarded base on the high stability, low enthalpy and low entropy of the excited state relative to energetic parameters previously observed for base pair open states (see Chapter 3). On the other hand, we observed the presence of chemical exchange at A C8/C1' and absence at T C6, which together with thermodynamic energies for the excited state would be most consistent with an intrahelical species and an A•T HG base pair.

Moreover, it was intriguing to discover in Chapter 3 that the 1mG•C$^+$ HG base pair was relatively stable at pH 5.2, which indicated that the pK_a of C N3 inside an HG base pair is higher than that of free cytosine ($pK_a \sim$ 4.2 - 4.4). This was not entirely surprising since, for example, an increase of several pH units in the pK_a of A N1 or C N3 from that of free adenine ($pK_a \sim$ 3.5) or free cytosine has been observed in non-canonical A$^+$•C wobble base pairs in DNA duplexes[22] and for a C$^+$-G•C base triplet in a DNA triplex[20]. The driving force behind the pK_a shift for internal base pairs has been proposed to be favorable enthalpy from interactions of the cationic species with the DNA backbone. However, 1mG•C$^+$ showed pronounced chemical exchange at pH 6.8 indicative of reduced stability, while the trapped 1mA•T base pair was stable at pH 6.8 (see Chapter 3). Therefore, we attempted to characterize in more detail the pH

dependence of chemical exchange in G•C and A•T base pairs as part of gathering further evidence for the nature of the excited states and insights into the pK_a of transient HG base pairs nested in a B-DNA duplex.

We measured the pH dependence of carbon $R_{1\rho}$ relaxation dispersion for G•C and A•T base pairs in the range of pH 4.0 to 7.6 at 26 °C by using G10 and A16 (A_6-DNA) base C8 carbons as probes. Since there was a correlation between G C8 and C C6 chemical exchange parameters, it is safe to assume that G C8 directly monitors formation of the protonated HG$^+$ base pair. The dispersion profiles at each pH were analyzed separately and globally, where we assumed that the ground-to-excited state chemical shift difference ($\Delta\omega_{AB}$) was invariable with pH (we observed a small and random < 0.2 ppm $\Delta\omega_{AB}$ variations from individual fits). Specifically, the G•C base pair experienced a steep increase in excited state population (p_B) from undetectable (< 0.05 %) at pH 7.6 to ~ 8.5
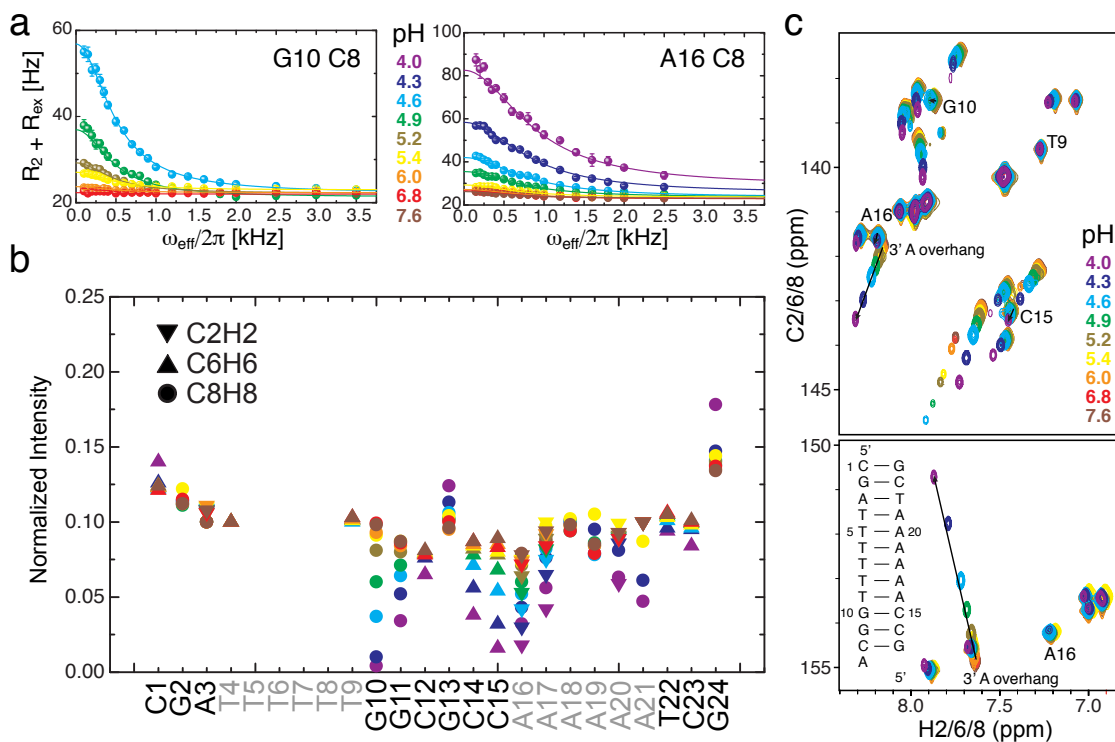
Figure 4.1: pH dependence of HG base pair formation for CA/TG. (a) On-resonance $R_{1\rho}$ relaxation dispersion profiles for G10 C8 and A16 C8 (A_6-DNA) as a function of pH (color-coded). (b) 2D ^{1}H,^{13}C HSQC spectra of the aromatic region of A_6-DNA during the pH titration showing little chemical shift perturbations for CA/TG residues (labeled) and large shifts for a 3'A overhang due to N1 protonation. (c) Corresponding normalized resonance intensities extracted from the 2D spectra.

% at pH 4.3 (Figure 4.2). Excessive line broadening or lack of chemical exchange for G10 C8 and C1' prevented us from measuring data at more acidic (< 4.3) and basic (> 6.8) pHs, respectively, for better definition of the pH profile. The increase in p_B for G10 C8 was correlated with an increase in the forward rate constant k_A, while the reverse rate constant k_B showed a very shallow increase (Figure 4.2). This interplay between p_B, k_A and k_B with lower pH can be interpreted, relatively, as either a decrease in the stability of the WC state or simultaneous increase in the stability of TS and HG states for a simplified two-state equilibrium. By using this analysis and fitting the G•C pH profile to Eq. 4.1, we obtained a $pK_{obs} \sim 3.25$ and a Hill coefficient of unity (n ~ 1.01), indicating no cooperativity. In the context of a WC-to-HG conversion, pK_{obs} is a complex parameter that contains information about both C N3 protonation and G *anti*-to-*syn* rotation.

To exhaust all options, let us consider that a single protonation event in the context of the WC base pair causes the pH dependent chemical exchange. As discussed above, G N7 is the only available atom for protonation and its pK_a in free guanine is ~ 2.9 – 3.3, which is similar to the pK_{obs} value that we obtained. However, DFT chemical shift calculations performed on N7-protonated G predict a downfield shift for G C8 but negligible perturbation for G C1' and, therefore, do not support G N7 protonation as the only reaction that takes place. The C C6 chemical shift could potentially be explained by transient abstraction of the N7-protonated G N1 proton by C N3 that occurs at low pH, but we expect this to be very unfavorable. If it was a single-step protonation mechanism, than we should observe a linear dependence of k_{ex} with proton concentration $[H^+]$, $k_{ex} = k_A + k_B = k_A^*[H^+] + k_B$, where k_A^* (the intrinsic rate constant for protonation) would be the slope and k_B the y-intercept. Under that

Figure 4.2: pH dependence of chemical exchange parameters for G10 and A16 C8 in CA/TG. Excited-state population (p_B) and best fits to Eq. 4.1 (solid) and Eq 4.3 (dashed), chemical exchange rate constant (k_{ex}), and computed forward (k_A) and reverse (k_B) rate constants for the two-state process plotted as a function of pH.
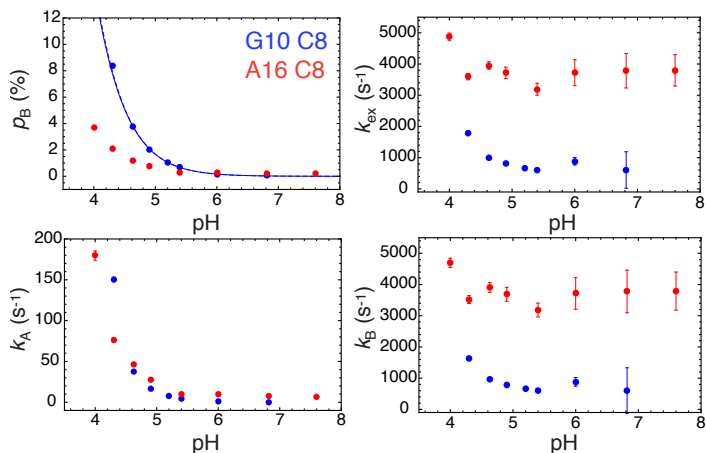
scenario, k_B should not be dependent on pH, which contradicts our results (Figure 4.2). Moreover, the lack of exchange of other G C8 sites at pH 5.4 – 6.8 observed in Chapter 3 (Appendix 1) further suggests that the chemical exchange is not a simple N7 protonation event.

It is clear that a simple protonation reaction cannot encompass the complexity of this process, and that the mechanism involves minimally a conformational change linked to the protonation. There are several complex reactions besides a WC-to-HG transition that we can invoke to explain our data, but they involve base pair opening and/or large sugar repuckering that are unlikely for reasons discussed in Chapter 3. From a physical standpoint, we regard a WC-to-HG transition as a minimally two-step process, where the WC base pair (i) first flips over and then gains a proton or (ii) first gains a proton and then flips over (Scheme 2). Alternatively, we can build a thermodynamic box that couples the two equilibria and four species together (Scheme 3). The process could be even more complicated if there is a flipped out intermediate, which we will not regard here. In order to estimate the pK_a of C N3 in the HG state, we could assume the three-state (ii) or four-state equilibria above. Analysis of the pH profile by using directly Eq. 2 and Eq. 3 and dissection of all equilibrium constants turned out to be extremely challenging – we could not obtain the desired convergence with either, likely due to insufficient data and sampling of the pH range. In order to estimate a reasonable lower bound for pK_a (C N3), we can make the tolerant assumption that the equilibrium constant for transition from WC to unprotonated HG (K_1) is $< 10^{-3}$. Thus, can estimate a pK_2 for the HG protonation reaction of at least ~ 6.23 using the three-state equation (Eq 4.2). For the four-state equation to yield convergence in the data fit, we would have to make a further assumption that the protonation of C N3 in the WC configuration is extremely unfavorable and that $10^{pK_4 - pH} \rightarrow 0$, which reduces Eq 4.3 to Eq 4.2. A similar value of ~ 6.25 is obtained if we use the relationship $pK_{obs} = pK_1 + pK_2$ (or $K_{obs} = K_1 K_2$) and $K_1 < 10^{-3}$, and derive pK_2 from pK_{obs} ~ 3.25 (n = 1.01) obtained from the two-state fit. This "minimum" pK_a ~ 6.2 that we calculate after a number of assumptions does fall in the range of pK_a values reported previously for the well-characterized A$^+$•C wobble base pair (~ 6.2 – 8.0) in DNA and RNA duplexes[23,24]. Much like the pK_a of A N1 in the wobble base pair, the pK_a of C N3 in the HG base pair is likely to be shifted by at least two units

(or two orders of magnitude in K) from that of free cytosine (~ 4.2). The common trends suggest that the shift in C N3 proton affinity would be primarily driven by enthalpic stabilization from favorable electrostatic interactions relative to the unfolded state (free cytosine).

We tried to assess, independently, the shift in pK_a (C N3) in the excited state HG base pair by monitoring directly the chemical shift of C C6 in a trapped HG 1mG•C$^+$ base pair (Section 3.3.5) as a function of pH (Figure 4.3). As discussed in Chapter 3, we could not observe the exchangeable C H3 imino proton directly, which is likely due to rapid exchange with solvent at pH close to the pK_a of C N3, superimposed onto a chemical exchange process whereby HG 1mG•C$^+$ base pair forms (we did observe the exchangeable T H3 in the trapped HG 1mA•T base pair since pK_a (T N3) ~ 9.2). On the other hand, the C H6 proton does not exchange with solvent and its linewidth is dominated by the pH-dependent chemical exchange process.

Following chemical shift changes with pH at a single nucleus is a commonly used NMR technique for monitoring protonation of atoms and for quantifying pK_a values in a site-specific manner. The unprotonated and protonated states should be ideally in fast exchange on the NMR timescale so that (i) there is a gradual change in chemical shift as the populations invert and (ii) the resonance peak remains observable throughout the pH titration. As discussed in Chapter 3, the chemical shifts for C C6 and G C8/C1' in the trapped HG 1mG•C$^+$ base pair at pH 5.2 matched perfectly our excited state chemical shifts for the same nuclei. Therefore, at low pH 5.2 we expect that the chemical shift of C C6 would correspond to > 90 % protonated C N3 fraction, while the chemical shift C C6 for the WC state would correspond to ~ 99 % unprotonated C N3 fraction. The
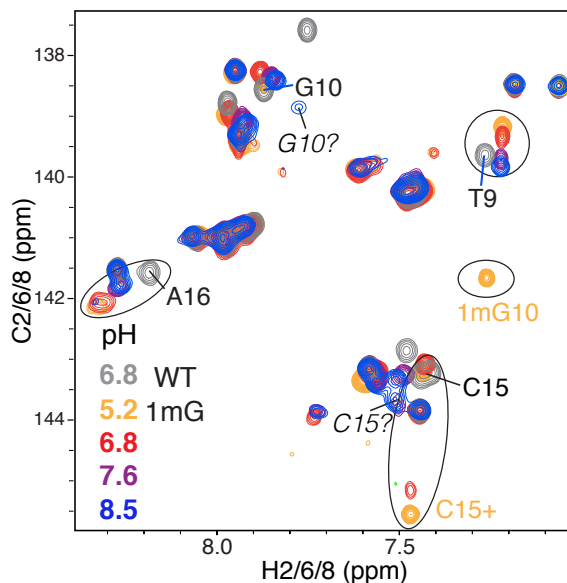
Figure 4.3: Insights into cytosine N3 pK_a in a trapped G•C$^+$ HG base pair. pH dependent change in the chemical shift of C C6 in A$_6$-DNA^{1mG10} (1mG) versus the unprotonated C C6 in A$_6$-DNA (WT).

only caviat to that assumption would be that the true chemical shift at C C6 in the unprotonated HG state might not coincide with that in the WC state due to different local geometries. Unfortunately, the chemical exchange at the trapped C C6 was on the intermediate timescale and we could not clearly observe the resonance above pH 6.8, although a resonance peak clearly appeared in the unprotonated C6 region at pH 8.5 ("C15?"). Figure 4.3 shows that the change in C6 chemical shift from pH 5.2 to 6.8 is small (~ 0.4 ppm) relative to the change from C6 in HG$^+$ to C6 in WC (~ 2.3 ppm). Moreover, the new cytosine C6 peak near the unprotonated WC C6 at pH 8.5 can be tentatively ascribed to C15 and would represent a mostly unprotonated state. Based on these observations and the large gap in carbon chemical shifts at C15, as well as the neighboring A16 and T9 that are sensitive to the transition, for pH 6.8 and pH 8.5, which pH values seem to approach the two endpoints of the titration, we can estimate the pK_a for C N3 to be somewhere between those values. This is consistent with our minimum pK_a estimation (~ 6.2) from the analysis of unmodified G•C chemical exchange (above).

Although the WC-to-HG transition for A•T base pairs does not involve a change in the nucleotides' protonation state, we observe appreciable increase in the HG fraction with more acidic pH (Figure 4.2). The value of p_B was steady ~ 0.3 % in the range of pH 7.6 and 5.4, but increased dramatically up to ~ 3.4 % at pH 4.0. In comparison, p_B for G10 at pH 4.3 was approximately 4-fold greater than p_B for A16. At this point, we are not certain what the origin of this pH dependence is, but it might be rooted in the different proton affinities of solvent exposed adenine nitrogens. Theoretically, the intrinsic basicities of the three nitrogens in free adenosine are arranged as N1 > N7 > N3, with micro acidity constants (in the absence of protonation at other Ns) given by pK_{N1} ~ 3.63, pK_{N7} ~ 2.15, and pK_{N3} ~ 1.5 from a recent experimental study[21]. One hypothesis is that the higher proton affinity of the solvent exposed N1 in the HG major groove than for the N7 in the WC major groove could create more proton and cation density at N1 and favorable electrostatic base-phosphate interactions at considerably low pH, biasing the equilibrium towards the HG base pair. At the same time, the likelihood for protonation of the H-bonded N1 and disruption of WC interactions becomes more significant than the likelihood for protonation of the H-bonded A N7 and disruption of HG interactions, which could affect the kinetics and thermodynamics. Also, we cannot exclude that

protonation at other nitrogen sites (N3) as well as neighboring residues at low pH, deduced from visible chemical shift changes during the pH titration, could affect the relative stability of the inter-converting WC and HG states. Yet another, perhaps less likely, possibility is that the increased frequency of the neighboring G•C HG base pair in the CA/TG step could contribute in a cooperative manner to the formation of the A•T HG base pair. Further studies including the use of acid catalysts or the presence of trapped neighboring HG base pairs will be critical in elucidating the source of these pH variations in the WC-to-HG balance of A•T base pairs and could help understand the transition mechanism.

Because we observed a trend of significant increase in HG base pair populations at low pH for G•C and A•T, we probed the pH dependence of other base pairs in A_6-DNA in order to investigate whether acidic conditions would bring into detection excited HG states that were hidden at high pH. Those could have remained undetected as a result of slower kinetics and/or sufficiently low populations that place them beyond the detection limit. Indeed, we observed a global "activation" of transient HG base pairs across the entire DNA sequence based on on-resonance $R_{1\rho}$ relaxation dispersion profiles at pH 4.6 versus pH 5.4 (Figure 4.4). Because we see the same trend of correlated exchange at purine C8 and C1' carbons (i.e. A17), we assume that these exchange processes represent excursions to HG base pairs as for the CA/TG step. Judging from the extent of R_{ex} contribution, the effect was most striking around the CA/TG step and clearly diminished towards the interior of the A-tract that coincides with the interior of the duplex. This was concurrent with
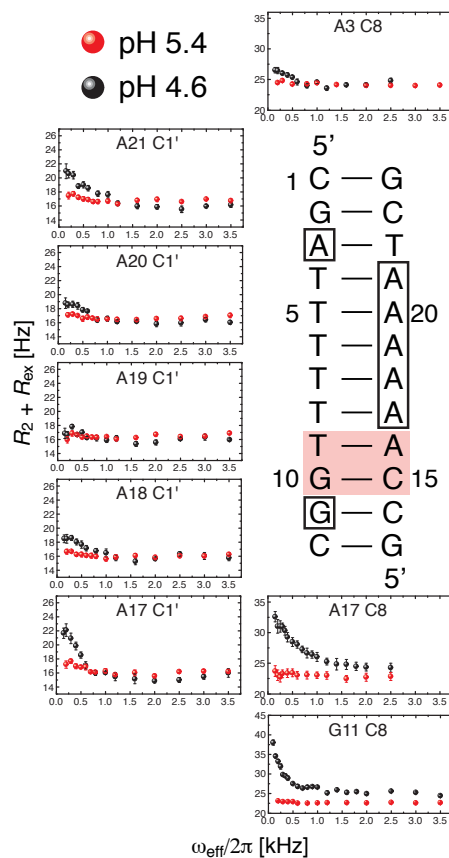
Figure 4.4: Low-pH induced of chemical exchange in sequences other than CA/TG steps. Shown are on-resonance $R_{1\rho}$ relaxation dispersion profiles at pH 5.4 and 4.6 for a number of C8 or C1' sites (boxed) in A_6-DNA (profiles for G10 and A16 in included in Figure 4.1). CA/TG is highlighted in red).

enhanced line broadening as seen in pH dependent resonance intensity profiles (Figure 4.1b). The HG base pair formation seems to generally correlate with sequence-specific anisotropic flexibility based on crystallographic studies and ps-ns dynamics presented here (Chapter 2). The most notable example is the highly deformable and dynamic CA/TG step in contrast to the structurally invariable and rigid AA/TT and AT step[14]. Therefore, we believe that this trend is dictated mostly by increasing A_6-tract rigidity and AA-stacking towards the interior rather than the overall higher stability of the duplex core. Our conclusion is strengthened by the observation of progressively longer base pair lifetimes in longer A-tracts made by Leroy $et\ al.$[25]. It is possible that the well-stacked and propeller-twisted AA step conformation, that tends to have an average base-pair rise but smaller than some CA/TG steps[26], inhibits the WC-to-HG transition. Any positive cooperativity in the creation of neighboring HG base pairs could potentially be a contributing factor as well. Needless to say, we would need further off-resonance relaxation dispersion studies as well as data as a function of sequence and A-tract length to be able to reach more definite conclusions about the process and its sequence-specificity. Most importantly, by selecting conditions that we know favor HG base pairs, we have been able to access HG states that seem to elude detection at neutral pH and have demonstrated that they are encoded by various DNA sequences.

Prior observations of upfield-shifted imino proton resonances at acidic pH (pH < 5) have suggested that appreciable fractions of HG base pairs in B-DNA could contribute to those NMR signals apart from single strands and frayed duplexes[27], but no concrete evidence had been presented thus far at the molecular level. Our studies suggest that the low-pH induction of large HG base pair fractions could alter appreciably the state of DNA and chromatin under conditions of cellular stress and oxygen starvation that upset the pH balance, and affect its reactivity and interaction with proteins and chemicals in the body, or stimulate DNA damage.

4.3.2. Effect of monovalent and divalent cations on WC-to-HG transitions

To obtain insights into the cation binding and salt dependence of WC-to-HG transition, we examined spectral features and carbon $R_{1\rho}$ relaxation dispersion of A•T and G•C base pairs at CA/TG step upon addition of monovalent (Na^+) and divalent (Mg^{2+})

cations. Initially, comparison of ^{1}H/^{13}C HSQC spectra for A$_6$-DNA in 25 mM and 125 mM added Na$^+$ at pH 5.4 and pH 6.8 revealed no significant chemical shift changes for base and sugar sites with increasing monocation concentrations (data not shown). On the contrary, addition of 10 mM Mg^{2+} caused large chemical shift perturbations for selected internal residues, specifically, the four base pairs around the 5' A-tract junction (5'CCAA/TTGG) containing the CA/TG step (Figure 4.5b). Addition of more Mg^{2+} up to 25 mM did not cause further spectral changes for internal residues, implying that Mg^{2+} binding and structural transition was already nearly saturated with 10 mM Mg^{2+} (Figure 4.5b). The Mg^{2+}-induced perturbations were evident for both base and sugar sites and signified specific cation binding and local conformational distortions in the duplex. Conversely, only minor changes were observed at the 3' A-tract junction (AT step).

Divalent cations (M^{2+}) are known to bring about thermal stabilization of DNA structure. Crystallographic data and MD simulations have shown that divalent metal ions (Mg^{2+}, Ca^{2+}, Zn^{2+}, etc.) can bind the double helix sequence- and site-specifically in both the major and minor grooves[2-5], and cause helical bending by base-roll compression when bound to the major groove, especially at GG and AG steps (M^{2+} binds to G N7/O6 or A N7). Moreover, the macroscopic curvature of DNA containing phased A-tracts can be augmented by such groove-specific binding of Mg^{2+} to intervening GC-rich sequences[28]. Here, the binding at the TTGG sequence was consistent with the high affinity of Mg^{2+} for GG sequences. Therefore, the structural perturbation we detect could involve helical bending around the A-tract boundaries. On the other hand, the lack of binding to the A-tract interior is consistent with the unfavorable accommodation of hydrated Mg^{2+} in the narrow minor groove of A-tracts.

Relaxation dispersion data for A16 and G10 (CA/TG step) at variable cation concentrations showed complex and opposing effects for HG base pair formation at A•T and G•C (Figure 4.5a). For the G•C base pair, increasing the Na$^+$ content from 25 mM to 125 mM (pH 5.4) reduced the excited state population (p_B) almost 2-fold and led to concurrent smaller decrease in k_A and increase in k_B. The same was not true for the A•T base pair, where a several fold increase in Na$^+$ concentration did not have any impact on chemical exchange at A•T (pH 6.8). On the other hand, the Mg^{2+} titration (10 and 25 mM) had a gradual effect on the relaxation dispersion at both base pairs. For G•C, adding

Mg^{2+} continued the trend observed with Na^+, namely suppression of the HG population and modulation of the rate constants, until exchange was no longer detectable at 25 mM Mg^{2+}. On the contrary, the presence of divalent ions led to a small but persistent increase in p_B for the A•T base pair. The Mg^{2+}-induced modulation of the transition rate constants was most complex – the lower Mg^{2+} concentration caused a simultaneous decrease, while the higher Mg^{2+} concentration restored their values (k_B) or brought them higher (k_A) relative to their values without Mg^{2+}. These observations also highlight the independence of excited state HG base pair formation at G•C and A•T.

Let us focus on the G•C base pair first. With Na^+ interacting non-specifically, suppression of the HG base pair by higher amounts of monovalent salts could be a consequence of condensed cation electrostatic effects that oppose the protonation at C N3. Positively charged counterions interacting favorably with backbone phosphates would compete with the uptake of similarly charged protons, offsetting the proton affinity of C N3. Such anticooperative effects between monovalent ions and protons have been reported for a salt dependent stability of A•C$^+$ wobble base pairs[22]. We can provide the same explanation for the large, Mg^{2+}-induced depletion of HG base pairs. The specific
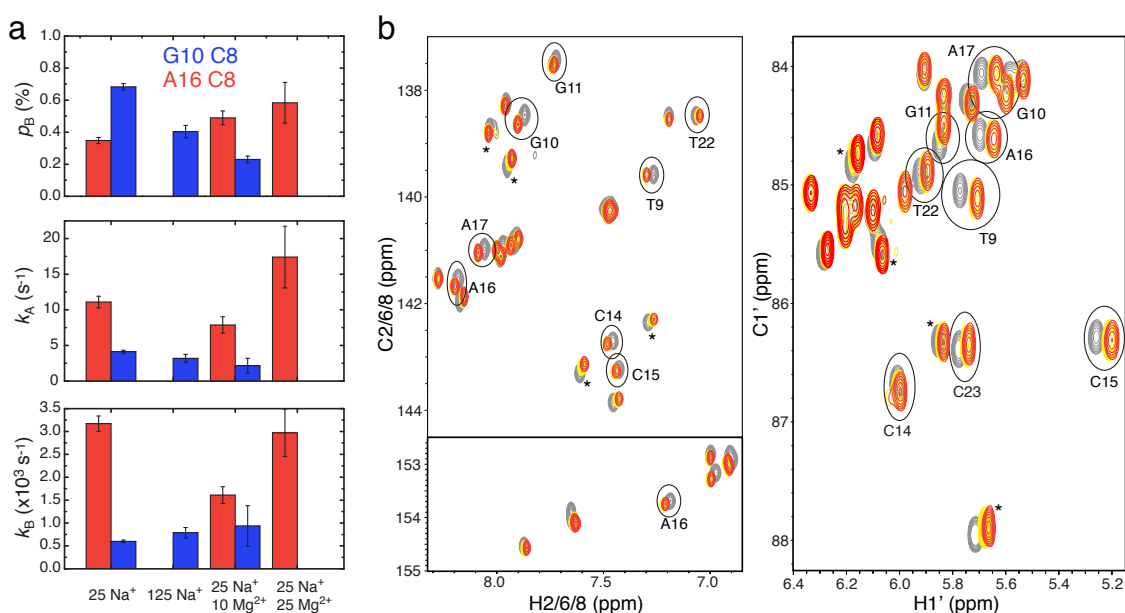
Figure 4.5: Salt dependence of chemical exchange at CA/TG. (a) Variation in excited-state population (p_B), and computed forward (k_A) and reverse (k_B) rate constants as a function of Na^+ and Mg^{2+} concentration (mM). (b) 2D $^1H,^{13}C$ HSQC spectra overlay of A_6-DNA in absence (grey) and presence of 10 mM Mg^{2+} (yellow) and 25 mM Mg^{2+} (red), showing significant chemical shift perturbations at CA/TG sites.

binding of Mg^{2+} or hydrated Mg^{2+} ions at the CA/TG step, which increases its effective local concentrations relative to Na^+, would produce the more dramatic anticooperative effect that we observe. We see that pattern when we compare results from 10 mM Mg^{2+}/25 mM Na^+ and 125 mM Na^+, where the ionic strength of the second mixture is actually higher. Ultimately, this reduction in the population of HG base pairs with higher salt concentrations would be due to a decrease in the cytosine N3 pK_a. Another explanation, complementing the first, for the reduced population and forward kinetics and population trend in the formation of G•C HG state with Mg^{2+} could be the actual major groove binding of Mg^{2+} to the GG step (N7 and O6). Such binding can induce groove compression[4] and, on one hand, obstruct the *anti-syn* rotation of the bulky guanine base via the major groove and, on another hand, stabilize selectively the WC state since these unique contacts/geometry would not be feasible in the HG state where N7 has been replaced with the (semi-)protonated N1.

For the for A•T base pair, increasing the cation concentration did not result in any change in exchange parameters in the case of Na^+ or induced a higher HG population in the case of Mg^{2+}, opposite to what was observed for G•C. Unlike G•C, the A•T WC base pair should not change its protonation state when it switches to the HG geometry, so mechanisms operating on A•T could be intrinsically different. Because monovalent cations do not produce any noticeable effects for A•T, we speculate that the specific water-mediated or direct contacts that divalent ions make with the CCAA/GGTT stretch play a role. First, since Mg^{2+} can bind in the major groove (A N7) of DNA[4,5], a similar argument can be made for metal-ion affinities of A N1 versus A N7 as for proton affinities in the previous section. Specifically, in the case of hydrated Mg^{2+} coordination to an adenine nitrogen in the major groove, the higher pK_a of N1 could allow for greater cation localization in HG and stabilize more the non-canonical HG base pair. Although we favor this argument, it is not clear whether Mg^{2+} binding at that location or, alternatively, at the GG step is what causes the Mg^{2+}-induced chemical shift changes at the A•T base pair (Figure 4.5). If it is the case that a hydrated Mg^{2+} coordinates to the neighboring GG step and changes in the groove width, such deformations could also produce a more favorable transition to HG for the A•T base pair. A careful examination of Mg^{2+} binding could be achieved by following chemical shifts of G N7 and A N7 via

138

long-range HSQC NMR experiments to try to resolve some ambiguities about specific cation binding.

4.3.3. Effect of 5-methylcytosine on G•C WC-to-HG transition

Our pH dependent study of transient HG base pairs motivated us to look into how manipulating the pK_a(N3) of cytosine will affect the proportion or lifetime of transient HG base pairs. Addition of various substituents at the C5 position has been predicted by density functional studies[29] to affect the pK_a of N3. One such cytosine modification that is believed to increase the proton affinity of N3 is C5-methylation (5mC). This modification adds an electron-donating group and changes the charge distribution on the cytosine base. C5-methylation is commonly known as the "fifth natural base"[7]. As discussed in the introduction, C5-methylation occurs primarily at CG steps as a way to regulate the epigenetic state of the mammalian genome, but it has been detected prevalently in CA (and CT) steps of embryonic stem cells[9].

To test whether C5-methylation shifts the cytosine pK_a(N3) and increases the potential for G•C HG formation at neutral pH, we prepared a 5mC modified DNA construct and collected carbon $R_{1\rho}$ relaxation dispersion for G base paired with 5mC (G•5mC) (Figure 4.6). The chemically altered base was incorporated in place of C15 in an unlabeled strand annealed to a $^{13}C/^{15}N$-labeled strand to form a modified semi-labeled A_6-DNA duplex (A_6-DNA^{5mC15}). The 5-methyl group induced very little perturbation at G10•5mC15 and no detectable effect on the conformation of neighboring and other parts of the duplex, as judged by chemical shift overlay of $^1H/^{13}C$ HSQC spectra for base and sugar nuclei. Rather surprisingly, the relaxation dispersion profiles for G10 C8 at three different pHs (5.2 - 6.8) revealed little change in the population of the excited state HG base pair as compared to unmodified G•C (Figure 4.6a). Thus, 5mC does not disturb the thermodynamic equilibrium between WC and HG base pair and, likely, does not shift the cytosine pK_a(N3) substantially based on the three-point comparison.

While the WC-HG balance was carefully maintained, we observed a coordinated increase in the forward and reverse rate constants (k_A and k_B) relative to unmodified G•C, reaching 2-3 fold at pH 5.2, where the relaxation dispersion profiles are most accurate. This signified a reduction of the lifetimes for both WC and HG base pairs with 5mC

modification, making it easier for the transition to occur. Thus, the perturbation either destabilizes both the WC and HG state and/or stabilizes the transition state. It is highly unlikely that the 5-methyl modification destabilizes the DNA duplex, either WC or HG states. Prior studies report that 5mC brings about a strong thermal stabilizing effect in duplex and triplex DNA[30,31] (containing a Hoogsteen third strand), which is believed to be predominantly entropic in origin from the release of ordered water molecules in the major groove to free space for the deeply-inserted methyl group. The same is valid for methyl-purine modifications that place the methyl group in the major groove[32]. 5mC incorporation also dampens the exchange rate and increases the barrier for imino proton transfer ($\sim$ 4 kcal/mol)[33], and arrests fast motions in cytosine sugar moieties of CG steps[34,35]. On the contrary, we observe a decrease in transition barriers between the two exchanging conformations ($\sim$ 2-3 fold), indicative of relative stabilization of the TS, which could involve simultaneous stabilization of WC and HG, and more significantly, the TS conformations relative to unmodified DNA. The discrepancy between TS behavior in relaxation dispersion and imino proton exchange provide another evidence
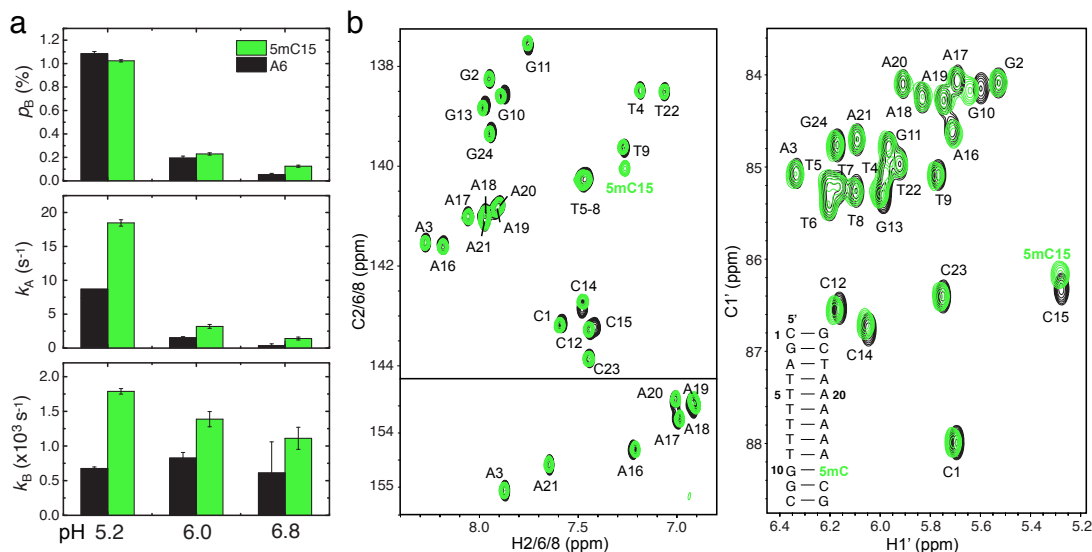
Figure 4.6: Effect of 5-methylcytosine on G•C[+] HG base pair formation. (a) Comparison of excited-state population (p_B), and computed forward (k_A) and reverse (k_B) rate constants for G10 C8 in unmodified A_6-DNA (black) and semi-labeled methylated A_6-DNA[5mC15] (green). (b) 2D ^{1}H,^{13}C HSQC spectra overlay of unlabeled A_6-DNA (grey) and A_6-DNA[5mC15] (green), showing only minimal perturbations around the site of methylation (5mC C6 is shifted due to direct base modification, resembling a T C6). Unlabeled and semi-labeled constructs displayed perfect chemical shift overlay for the labeled strand (data not shown).

140

that the excited-state is not a base-pair open state. The effect of the methyl group alone on overall duplex stability could be explored in a control experiment by replacing another cytosine with 5mC.

Although C5-methylation does not alter the local DNA conformation in any significant way, the methyl group causes a profound effect on the local charge environment of cytosine and the major groove and, as expected, interacts better with positively charged species while the H5 proton interacts better with negatively charged species. It is possible that a modified charge density and/or hydration pattern lowers the energetic cost for the transition and accelerates HG base pair formation. Additional insights into the energetics of the process can be obtained from a temperature-dependent study of the WC-to-HG transition and comparison of the enthalpy and entropy with the unmodified G•C.

From a biological standpoint, methylated HG base pairs have not been observed in any functional context so far because of their elusive nature, but may very well participate in various cellular functions. Certainly, the higher frequency of transformation between WC and HG base pairs relative to unmodified G•C could allow for rapidly diffusing proteins and small molecules to sense HG sites and, perhaps, facilitate HG-based intermediates in the transition pathway towards Z-DNA in poly-CA repeat sequences.

4.3.4. Exploring the cooperativity of HG base pair formation in CA/TG steps that trap or abolish one HG base pair, or that neighbor multiple CA steps

4.3.4.1 Effect of neighboring trapped HG base pairs

So, do HG base pairs prefer to form cooperatively or not at CA/TG steps? MD simulations suggest that an antiparallel all-HG double helix is nearly as stable as an all-WC double helix, being slightly disfavored by intramolecular entropic considerations[36]. Moreover, simulations of WC-HG junctions find that, while the stacking energy for AT steps is the same with a WC or HG H-bonding scheme, the stacking energy for TA steps is ~ 3.8 kcal/mol more favorable for the HG H-bonding scheme[17]. We expect a similar behavior for CA steps. We attempted to address this question by introducing a trapped 1mA•T base pair at the CA/TG step of semi-labeled A_6-DNA and probing the WC-to-HG

transition at the adjacent G•C base pair (we will refer to that as G•C*). The DNA construct contains an unlabeled modified strand and a $^{13}C/^{15}N$ enriched unmodified strand, and is identical to the A_6-DNA^{1mA16} studied in Chapter 3, where we detected formation and accomodation of a single 1mA•T HG base pair inside the otherwise B-DNA double helix (Figure 3.7). Perfect chemical shift overlay was observed between unlabeled and semi-labeled A_6-DNA^{1mA16} for observable resonances (data not shown).

When comparing ^{13}C $R_{1\rho}$ relaxation dispersion profiles on G10 and G10* C8/C1', we observed a small, almost negligible difference in the excited state populations but a large > 5-fold simultaneous increase in the forward and reverse rate constants (k_A and k_B) (Figure 4.7a). Even though this effect is analogous (and much greater) to the effect of the 5-methylcytosine substitution in the previous section, the basis for these kinetic differences are probably not the same. Unlike 5mC that stabilized the DNA helix, a single modified 1mA•T HG base pair reduces significantly the thermodynamic stability of DNA duplexes[37], and is consistent with our observation of line broadening at the modified site (1mA16•T9) and, more so, its immediate neighbors (A17•T8) (Figure 4.7b, Figure 3.7). A potential source of entropic and enthalpic stability for the trapped HG base pair could be placement of the adenine methyl group in the major groove and favorable interaction of the positive charge with the sugar-backbone. Naturally, we would think that the reduction in the transition barriers at G•C* originates from, coincidentally, similar destabilization of the WC and HG states from the neighboring protonated 1mA•T HG pair. It is easier to picture how the WC base pair is weakened, but the reason for the HG base pair could be a diminished proton affinity due to charge repulsion with the unnatural, positively charged adenine nextdoor. Even though this seems far-fetched, we believe this is plausible and it is partially supported findings of Wu et al.[38] that adjacent placement of C$^+$-G•C triplets inside a DNA triplex (forming a third-strand HG) lowers simultaneously the stability of WC and HG base pairs. If that is the reason, than our study is not truly representative of the coupling between HG geometries, and to resolve that ambiguity we should study the reverse situation where we look at dynamics of the A•T base pair that does not require protonation placed next to a 1mA•T and 1mG•C, which we plan to do in the future. Of course, we should also be mindful of the alternative explanation, stabilization of TS for G•C* for multiple reasons, one of which could be that

the weakened and distorted 1mA•T, on average, opens up more space to allow for an intrahelical base rotation.

To analyze this process in more depth, we also have to consider what the new excited state with two consecutive HG base pairs looks like. To this end, we prepared an unlabeled, doubly *N1*-methylated duplex at G10 and A16 nucleotides (A_6-DNA$^{1mG10/mA16}$) that was expected to mimic the G•C* excited state by forming two HG base pairs side by side, and this is indeed what we observed from ^{1}H/^{13}C HSQC at pH 5.2 (Figure 4.7b). A comparison of A_6-DNA$^{1mG10/mA16}$ with A_6-DNA^{1mG10} and A_6-DNA^{1mA16}, and all of these to unmodified A_6-DNA, revealed more extensive line broadening and upfield proton and carbon chemical shifts of HG base pair resonances in the doubly-methylated duplex. This implies that the two HG base pairs are undergoing more chemical exchange when together and that their average conformations are less stacked as compared to a single HG base pair confined inside B-DNA. The other state(s) with which they exchange could not be WC base pairs, whose formation is hindered by the methyl groups, but some partially unstacked or base pair open species. Proportionally higher chemical exchange is also seen at the WC base pairs at the junction between the WC and HG regions. G•C* WC base pair shows signs of instability inside A_6-DNA^{mA16} (ground state) relative to unmodified A_6-DNA, and so does the trapped G•C* HG base pair inside A_6-DNA$^{1mG10/mA16}$ (excited state-like). From this qualitative analysis, we cannot assess whether their stabilities have changed to a similar degree, as we see in the relaxation dispersion data, and how the enthalpy-entropy balance has been shifted. Despite that, we are able to present evidence for base pair weakening for both the ground and excited state that can allow for faster inter-conversion between WC and HG conformations. If the TS was stabilized, that would also support the more rapid two-state kinetics.

The reduced base pair stability of one or more *N1*-methylpurines could certainly have implications in the recognition of one or more alkylated, damaged bases by repair enzymes, which use a base-flipping mechanism to capture the damaged residue and excise it[39]. Whether the incidence of coexisting unmodified HG base pairs is sufficient to induce faster, more synchronized HG transitions still remains unclear. Currently, the two junctions between the WC helix and HG base pairs are abrupt with the HG base pairs

absorbing some of the stress that can offset their optimal positioning/stacking relative to an ideal HG helix. The two base-pair HG mini-helix formed here is potentially not long enough to counter the adverse effects of being trapped inside a B-DNA duplex, like a "nucleation penalty", while extending the HG helix might provide the sufficient energy for a more coherent transition. As mentioned before, HG base pairs in YR-type steps (TA) tend to interact better together than with surrounding WC base pairs[17], which could be the driving force for more enhanced transitions in YR steps. Conversely, RY (AT) steps do not discriminate between the two stacking modes and, therefore, YR-repeats might not have any advantage over single steps, based solely on stacking, to act in collaborative manner in the conversion to HG helices. If anything, the driving force could be the intrinsic instability of stacked YR steps in B-form helices that manifest in increased dynamic disorder in crystal structures and NMR order parameters, discussed in Chapter 2. Either one could have a contributing effect to more rapid dynamic transitions to non-canonical base pairs in DNA sequences comprised of repeating CA or TA units. We present preliminary results for CA_3-repeat below.
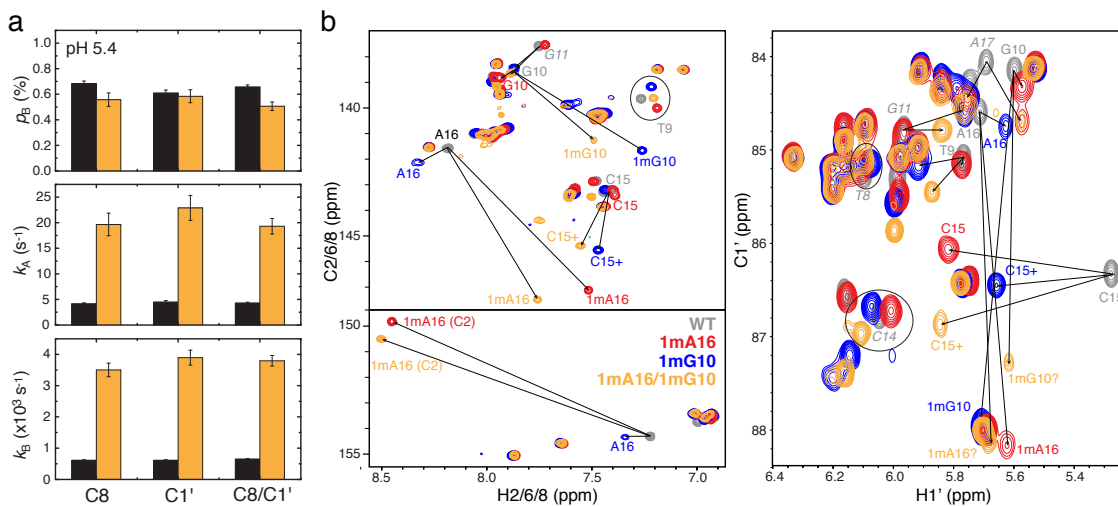
Figure 4.7: Effect of trapped HG base pair on WC-to-HG equilibrium at a neighboring site. (a) Comparison of excited-state population (p_B), and computed forward (k_A) and reverse (k_B) rate constants for G10 C8, C1' or C8/C1' fit globally in unmodified A_6-DNA (black) and semi-labeled *N1*-methylated A_6-DNA^{1mA16} (orange) containing a trapped 1mA•T HG base pair. (b) 2D ^{1}H,^{13}C HSQC spectra overlay of unlabeled A_6-DNA (grey), A_6-DNA^{1mA16} (red) with 1mA16•T9 HG base pair, A_6-DNA^{1mA16} (blue) with 1mG10•C15$^+$ HG base pair, A_6-DNA$^{1mG10/mA16}$ (orange) showing formation of both HG base pairs in the last but also significant upfield chemical shifts for corresponding resonances.

4.3.4.2 Effect of reduced HG potential in neighboring base pairs

To look at the question of HG base pair cooperativity from the opposite perspective, we tested how the removal of chemical exchange at one of the base pairs in CA/TG would correlate with less frequent or less populated HG states. To probe this effect, we used a semi-labeled construct containing a 7-deazaadenine (7dNA) modification at the A•T base pair (A_6-DNA^{7dNA16}) that we showed in Chapter 3 to make undetectable the formation of the excited state HG base pair, by lowering the population of slowing/speeding up the exchange rate, to an appreciable extent most likely due to the reduced H-bonding potential. A spectral overlay of unmodified and modified A_6-DNA (Chapter 3, Figure 3.10) showed the anticipated upfield chemical shift perturbations at the 7dNA C8H8 (~ 18 ppm) and C2H2 (~ 1.5 ppm) as a result of the single-atom substitution (N7 $\rightarrow$ C7H7) that alters the electronic environment of the heterocyclic ring. Surprisingly, changes in base and backbone chemical shifts extended to the opposite T9 and the two surrounding base pairs (G10•C15 and A17•T8), the most noticeable shift being for the sugar C1'H1' of C15 on the 5' side of 7dNA16. Although we cannot compare directly the resonance intensities C8H8 of 7dNA16 and A16, those of the deoxyribose C1'H1' site were slightly higher than for neighboring adenines, indicating the presence of increased sugar disorder.

^{13}C $R_{1\rho}$ relaxation dispersion profiles for G10 C8 on the labeled strand reveal an unanticipated ~ 2.8-fold increase in the excited state HG population, accompanied by a ~ 7.3-fold increase in k_A and ~ 2.6-fold increase in k_B (Figure 4.8). Evidently, the removal of N7 in the neighboring base had much more profound effects than we had anticipated. It would not be an accurate reporter of HG base pair cooperativity considering all the different structural and energetic "side effects" from the substitution. A feasible explanation for these exchange statistics is a parallel decrease in stability of the WC and, less so, of the HG conformation relative to the TS conformation, as compared to unmodified A_6-DNA. In general, the thermal stability of the modified duplex is compromised as a result of less favorable gain of entropy of folding (from fewer expelled waters) associated with 7-deazaadenine relative to adenine in AT-rich duplexes[40]. We believe this would, more or less, equally affect the WC and HG duplex states and would not discriminate between *anti* and *syn* conformers for G10. In addition, at the molecular

level, the substitution of N7 would abolish water coordination at that H-bond acceptor site, and the enthalpic loss might outweigh the entropic gain of water liberation for 7dNA16. A disrupted major groove hydration pattern could easily disturb water contacts at the neighboring WC G•C base pair with N7, but might not affect as dramatically the HG state that has placed N1H1 in the major groove, which would not coordinate to waters in its protonated state. Again, a decomposition of enthalpy and entropy terms would be helpful in the analysis of this process.

Evidently, the removal of N7 in the neighboring base had much more profound effects than we had anticipated. Thus, it would not be an accurate reporter of HG base pair cooperativity considering all the different structural and energetic "side effects" on the ground state.

4.3.4.3 Effect of CA-repeats

In order to test whether repeats of CA/TG steps will have an enhancement of HG base pair formation, we prepared a site-specifically labeled A_6-DNA dodecamer where the A_6-tract has been replaced with a $(CA)_3$ repeat (CA_3-DNA). To reduce spectral overlap in the highly homogeneous sequence, only A/C on the $(CA)_3$ side and G on the $(TG)_3$ side were $^{13}C/^{15}N$-labeled. We have obtained preliminary relaxation dispersion profiles for most of the non-overlapped resonances in CA/TG steps including A C8/C1', G C1' and C C6 that are sensitive to the slow dynamics. First
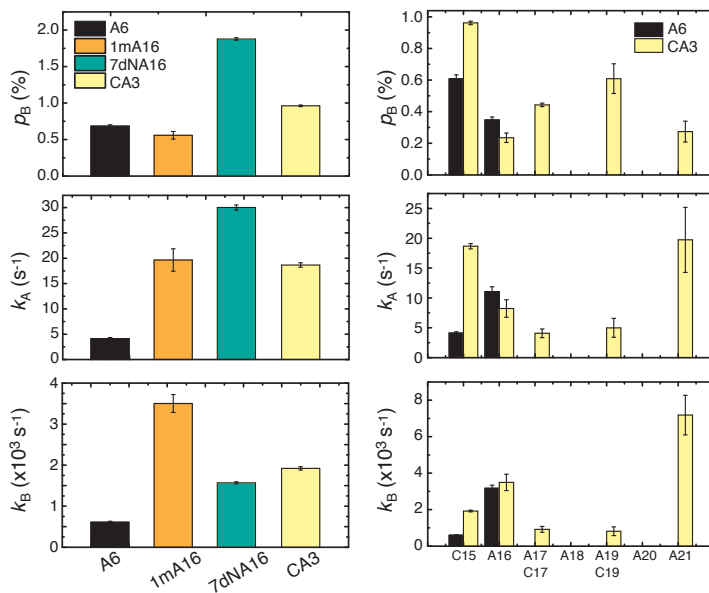
Figure 4.8: Probing cooperativity of HG base pair formation by modification or expansion of CA/TG steps. (a) Comparison of excited-state population (p_B), and computed forward (k_A) and reverse (k_B) rate constants for G10 C8 or C1' in unmodified A_6-DNA (black), *N1*-methyladenine A_6-DNA^{1mA16} (orange), 7-deazaadenine A_6-DNA^{7dNA16} (aqua), and in CA_3-DNA (yellow). (b) Comparison between chemical exchange parameters (as defined in (a)) for A_6-DNA and CA_3-DNA as a function of residue number.

of all, we see that all non-overlapped sites exhibit relaxation dispersion that confirms the universality of excited state transitions at various CA steps. If we compare the single "conserved" CA/TG step, the chemical exchange at A16•T9 remains largely unchanged (within error) while the chemical exchange at G10•C15 is modified between CA_3-DNA and A_6-DNA. Specifically, the excited state population, the forward and reverse rate constants simultaneously increase by ~ 1.5-fold, ~ 4.5-fold and ~ 3-fold, respectively. To interpret these results in any reasonable manner, we can assume that both the WC and, less so, the HG base pair are destabilized relative to TS when placed in the new sequence context of a CA_3-repeat. As we have found from our temperature-dependent study of WC-HG equilibria in Chapter 3, excited state HG base pairs are characterized by lower entropies than ground state WC base pairs, which is consistent with theoretical studies of antiparallel HG helices[36]. To analyze the contributions from enthalpy and entropy and elucidate the mechanism by which HG base pairs become more favored in the CA_3-repeat versus A_6-tract, we would need to perform a similar temperature study of the relaxation dispersion. And whether this effect comes from the CA_3-repeat or lack of A_6-tract, we can assess by probing the chemical exchange of G10 in the shorter A-tract, A_2-DNA, from Chapter 3.

The two new G•C base pairs in the CA_3 tract showed similar chemical exchange parameters that were comparable to G10•C15 in A_6-DNA (Figure 4.8). Therefore, WC ground states at these steps are more stable relative to WC at the 5' end CA/TG. Clearly, we need at least two more control sequences to be able to interpret this data. One such control is a permutation of the CA/TG step that places it more towards the duplex interior without a CA_3-repeat, the other is A_2-DNA, for which we have not to collected yet relaxation dispersion under these conditions.

A more surprising discovery is the enhanced relaxation dispersion at A19 found at the 3' end of the CA_3-repeat and in an A_2-tract. A16 and A21 are both four base pairs removed from the duplex end. In comparison to A16, A21 exhibits similar p_B values but ~ 2-fold more rapid exchange kinetics in both the forward and backward direction that are indicative of (i) simultaneous WC and HG destabilization of (ii) TS stabilization. To understand the dynamic behavior of A21 in CA_3-DNA, we would have to compare it to A21 in A_6-DNA, A19 in A_6-DNA, and A17 in A_6-DNA (from Chapter 3) that are found

in the same immediate neighbor context but vary in their position along the helix or in the length of neighboring A-tract. A17 would be the most analogous to A21 in CA_3-DNA, however, both C8 and C1' resonances were overlapped and could not be investigated. The overlap could only be combated with selective isotopic labeling as we have done for CA_3-DNA. We had collected on-resonance profiles of A21 and A19 that show a smaller exchange contribution (Appendix 1), but we had not thoroughly quantified chemical exchange parameters from off-resonance relaxation dispersion (Chapter 3) to allow for accurate comparison.

At this point, the most prominent finding regarding the dynamics of CA_3-DNA is enhanced chemical exchange at the 5' end G•C and 3' end A•T base pair relative to control sequences. We hypothesize that both are due to destabilization of WC and HG base pairs simultaneously, which is likely induced by the CA_3-repeat. Certainly, the dynamic instability of CA_3-repeats can facilitate more and more frequent formation of HG base pairs relative to rigid sequences like long A-tracts. Additional experiments are necessary to establish the origin for these motional differences in various DNA samples

4.4 Conclusion

In summary, by probing the pH dependence of relaxation dispersion in A_6-DNA, we have shown that transient HG base pairs in A•T and, more so, G•C become more populated under acidic conditions. Acidic conditions in the cell under stress or cellular cues that affect the cytosine pK_a could have profound effects on the structure of DNA and its interaction with architectural and regulatory proteins. From a quantitative analysis of the pH profile and chemical shifts in a trapped 1mG•C HG base pair, we have also shown that the cytosine pK_a(N3) in the G•C$^+$ HG state is heightened by at least two units from that of free cytosine, which is likely a result of enthalpy gain from electrostatics and intrahelical H-bonding relative to the unfolded state, similarly observed for an A$^+$•C wobble base pair. Moreover, we have demonstrated that HG base pairs have the ability to form transiently and sequence-specifically in other DNA sequences, besides CA/TG and TA steps, but that their populations are inhibited and elude NMR detection at neutral pH.

The thermodynamics and kinetics of transient HG base pairs can be affected by monovalent/divalent ions, chemical modifications that favor or disfavor HG formation at

neighboring sites, and CA-repeats. Mono- and divalent ions can form interactions with DNA grooves and sugar-backbones that modulate the dynamic landscape of WC-to-HG transitions in a way that suppresses the G•C and activates the A•T HG base pairs selectively in the CA/TG step. The abundant genomic C5-methylation was observed to accelerate the kinetics of exchange without affecting the relative probabilities of WC and HG states, possibly due to favorable electrostatics that stabilize the transition state. Our attempt to investigate the cooperative behavior of HG base pairs with the use of chemical modifications that induce or remove neighboring HG configurations was met with many challenges and complex side effects of the modifications on both WC and HG configurations. Introduction of a trapped *N1*-methylated HG base pair or an 7-deaza WC base pair, or replacement of the A-tract by a CA_3-repeat all accelerated the kinetics of WC-to-HG inter-conversion at a 5' neighboring G•C, and in the case of 7-deaza substitution and CA_3-repeat increased the fraction of transient HG base pairs. These effects are likely to arise from parallel WC and HG destabilization in the modified duplexes due to reduced stacking and altered hydration, but cannot be necessarily attributed to cooperativity in HG base pair formation inside B-form helices. Additional experiments are needed to gain further insights into the complex underlying phenomena that affect WC-HG equilibria.

4.5 References

1. Manning, G.S. Electrostatic free energy of the DNA double helix in counterion condensation theory. *Biophysical Chemistry* **101**, 461-473 (2002).
2. Young, M.A. et al. Structure determination and analysis of local bending in an A-tract DNA duplex: comparison of results from crystallography, nuclear magnetic resonance, and molecular dynamics simulation on d(CGCAAAAATGCG). *Methods Enzymol* **261**, 121-44 (1995).
3. Young, M.A. & Beveridge, D.L. Molecular dynamics simulations of an oligonucleotide duplex with adenine tracts phased by a full helix turn. *J Mol Biol* **281**, 675-87 (1998).

4. Chiu, T.K. & Dickerson, R.E. 1 A crystal structures of B-DNA reveal sequence-specific binding and groove-specific bending of DNA by magnesium and calcium. *J Mol Biol* **301**, 915-45 (2000).

5. Hud, N.V. & Polak, M. DNA-cation interactions: The major and minor grooves are flexible ionophores. *Curr Opin Struct Biol* **11**, 293-301 (2001).

6. Gal-Yam, E.N., Saito, Y., Egger, G. & Jones, P.A. Cancer epigenetics: modifications, screening, and therapy. *Annu Rev Med* **59**, 267-80 (2008).

7. Kim, J.K., Samaranayake, M. & Pradhan, S. Epigenetic mechanisms in mammals. *Cell Mol Life Sci* **66**, 596-612 (2009).

8. Esteller, M. Aberrant DNA methylation as a cancer-inducing mechanism. *Annu Rev Pharmacol Toxicol* **45**, 629-56 (2005).

9. Ramsahoye, B.H. et al. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc Natl Acad Sci U S A* **97**, 5237-42 (2000).

10. Murphy, K.E. & Stringer, J.R. RecA independent recombination of poly[d(GT)-d(CA)] in pBR322. *Nucleic Acids Res* **14**, 7325-40 (1986).

11. Gendrel, C.G., Boulet, A. & Dutreix, M. (CA/GT)(n) microsatellites affect homologous recombination during yeast meiosis. *Genes Dev* **14**, 1261-8 (2000).

12. Majewski, J. & Ott, J. GT repeats are associated with recombination on human chromosome 22. *Genome Res* **10**, 1108-14 (2000).

13. Gasanova, V.K. et al. [Complementary oligonucleotide invasion into (CA/TG)31-repetitive region of linear and circular DNA duplexes]. *Mol Biol (Mosk)* **44**, 520-8.

14. El Hassan, M.A. & Calladine, C.R. Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA. *J Mol Biol* **259**, 95-103 (1996).

15. Packer, M.J., Dauncey, M.P. & Hunter, C.A. Sequence-dependent DNA structure: dinucleotide conformational maps. *J Mol Biol* **295**, 71-83 (2000).

16. Ho, P.S. The non-B-DNA structure of d(CA/TG)n does not differ from that of Z-DNA. *Proc Natl Acad Sci U S A* **91**, 9549-53 (1994).

17. Cubero, E., Luque, F.J. & Orozco, M. Theoretical study of the Hoogsteen-Watson-Crick junctions in DNA. *Biophys J* **90**, 1000-8 (2006).

18. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-12 (2004).

19. Izatt, R.M., Christensen, J.J. & Rytting, J.H. Sites and thermodynamic quantities associated with proton and metal ion interaction with ribonucleic acid, deoxyribonucleic acid, and their constituent bases, nucleosides, and nucleotides. *Chem Rev* **71**, 439-81 (1971).

20. Asensio, J.L., Brown, T. & Lane, A.N. Comparison of the solution structures of intramolecular DNA triple helices containing adjacent and non-adjacent CG.C+ triplets. *Nucleic Acids Res* **26**, 3677-86 (1998).

21. Kapinos, L.E., Operschall, B.P., Larsen, E. & Sigel, H. Understanding the Acid-base properties of adenosine: the intrinsic basicities of n1, n3 and n7. *Chemistry* **17**, 8156-64.

22. Chen, G., Kennedy, S.D. & Turner, D.H. A CA(+) Pair Adjacent to a Sheared GA or AA Pair Stabilizes Size-Symmetric RNA Internal Loops. *Biochemistry* **48**, 5738-5752 (2009).

23. Siegfried, N.A., O'Hare, B. & Bevilacqua, P.C. Driving forces for nucleic acid pK(a) shifting in an A(+).C wobble: effects of helix position, temperature, and ionic strength. *Biochemistry* **49**, 3225-36.

24. Moody, E.M., Brown, T.S. & Bevilacqua, P.C. Simple method for determining nucleobase pK(a) values by indirect labeling and demonstration of a pK(a) of neutrality in dsDNA. *J Am Chem Soc* **126**, 10200-1 (2004).

25. Leroy, J.L., Charretier, E., Kochoyan, M. & Gueron, M. Evidence from base-pair kinetics for two types of adenine tract structures in solution: their relation to DNA curvature. *Biochemistry* **27**, 8894-8 (1988).

26. Yanagi, K., Prive, G.G. & Dickerson, R.E. Analysis of local helix geometry in three B-DNA decamers and eight dodecamers. *J Mol Biol* **217**, 201-14 (1991).

27. Nonin, S., Leroy, J.L. & Gueron, M. Acid-induced exchange of the imino proton in G.C pairs. *Nucleic Acids Res* **24**, 586-95 (1996).

28. Tchernaenko, V., Halvorson, H.R. & Lutter, L.C. Topological measurement of an A-tract bend angle: effect of magnesium. *J Mol Biol* **341**, 55-63 (2004).

29. Moser, A., Guza, R., Tretyakova, N. & York, D.M. Density Functional Study of the Influence of C5 Cytosine Substitution in Base Pairs with Guanine. *Theor Chem Acc* **122**, 179-188 (2009).

30. Xodo, L.E., Manzini, G., Quadrifoglio, F., van der Marel, G.A. & van Boom, J.H. Effect of 5-methylcytosine on the stability of triple-stranded DNA--a thermodynamic study. *Nucleic Acids Res* **19**, 5625-31 (1991).

31. Xodo, L.E., Alunnifabbroni, M. & Manzini, G. EFFECT OF 5-METHYLCYTOSINE ON THE STRUCTURE AND STABILITY OF DNA - FORMATION OF TRIPLE-STRANDED CONCATENAMERS BY OVERLAPPING OLIGONUCLEOTIDES. *Journal of Biomolecular Structure & Dynamics* **11**, 703-720 (1994).

32. Seela, F., Ramzaeva, N. & Chen, Y.M. Oligonucleotide duplex stability controlled by the 7-substituents of 7-deazaguanine bases. *Bioorganic & Medicinal Chemistry Letters* **5**, 3049-3052 (1995).

33. Mirau, P.A. & Kearns, D.R. Effect of environment, conformation, sequence and base substituents on the imino proton exchange rates in guanine and inosine-containing DNA, RNA, and DNA-RNA duplexes. *J Mol Biol* **177**, 207-27 (1984).

34. Meints, G.A., Miller, P.A., Pederson, K., Shajani, Z. & Drobny, G. Solid-state nuclear magnetic resonance spectroscopy studies of furanose ring dynamics in the DNA HhaI binding site. *Journal of the American Chemical Society* **130**, 7305-7314 (2008).

35. Pederson, K., Meints, G.A., Shajani, Z., Miller, P.A. & Drobny, G.P. Backbone dynamics in the DNA HhaI protein binding site. *Journal of the American Chemical Society* **130**, 9072-9079 (2008).

36. Cubero, E., Abrescia, N.G., Subirana, J.A., Luque, F.J. & Orozco, M. Theoretical study of a new DNA structure: the antiparallel Hoogsteen duplex. *J Am Chem Soc* **125**, 14603-12 (2003).

37. Yang, H. & Lam, S.L. Effect of 1-methyladenine on thermodynamic stabilities of double-helical DNA structures. *FEBS Lett* **583**, 1548-53 (2009).

38. Wu, P., Kawamoto, Y., Hara, H. & Sugimoto, N. Effect of divalent cations and cytosine protonation on thermodynamic properties of intermolecular DNA double and triple helices. *J Inorg Biochem* **91**, 277-85 (2002).
39. Leiros, I. et al. Structural basis for enzymatic excision of N1-methyladenine and N3-methylcytosine from DNA. *EMBO J* **26**, 2206-17 (2007).
40. Seela, F. & Chen, Y. Oligonucleotides containing fluorescent 2'-deoxyisoinosine: solid-phase synthesis and duplex stability. *Nucleic Acids Res* **23**, 2499-505 (1995).

Chapter 5

Conclusion and future perspectives

The DNA double helix is held firmly together by specific base-pairing and base-stacking interactions that make the structure of DNA very resilient to change and chemical modification, which contributes to genomic integrity and reliable reproduction. This poses the naïve question – how can such a rigid framework composed of only four nucleotide varieties discriminate among so many diverse proteins targets? Crystallographic studies have shown that DNA can wind around histone octamers to assemble into nucleosomes[1] or can bend and twist dramatically upon the binding of transcription factors[2], without disruption of Watson-Crick base pairing. Clearly, the local structural adjustments that DNA can tolerate – kinking, groove expansion and compression, local displacements of bases and sugars, etc. – arise from its intrinsic ability to deform and access a range of conformations. It has been proposed that the energies required for DNA-accessible fluctuations are largely determined by its base composition and specific sequence. This sequence-specific flexibility, beyond direct interactions, can modulate the binding of proteins and small molecules by altering the DNA shape and regional physicochemical properties. Numerous studies of DNA structure, reviewed in the introduction, have been able to tackle questions concerning the inherent flexibility of particular sequences. Much fewer studies of the sequence-directed dynamic events that ultimately give rise to observed variations in the average structure exist, despite a huge body of literature on DNA dynamics. In this thesis, I have used NMR spectroscopy to probe atomic-level dynamic fluctuations spanning picoseconds to milliseconds timescales across various sequences in order to gather more in-depth view of sequence-specific motional modes that exist in the canonical DNA double helix.

5.1 Sequence-specific rapid internal motions in DNA helices

Atomic level spin relaxation NMR studies of DNA dynamics have been limited to short duplexes in which sensitivity to biologically relevant fluctuations occurring at ns timescales is often inadequate. In Chapter 1, we developed and implemented an approach to quantitatively study site-, position- and sequence-specific motions of larger and more anisotropic $^{13}C/^{15}N$-labeled DNA systems (56 base pairs). Our results on elongated DNA reveal rapid ns motions (< 20 ns) that are enhanced towards the terminal ends and penetrate deep (7 base pairs) within the DNA helix, and that are not influenced by the insertion of a rigid A-tract motif nearby. These dynamic modes, not observed in short DNA helices, could represent enhanced fraying near terminal ends and/or large-scale helical bending fluctuations that persist in long DNA helices; those could be superimposed onto sequence-specific local dynamics that evade detection in short DNA due to couplings with overall tumbling. Even though NMR, EPR and FAD studies have presented plenty of evidence for large-scale ns motions in much longer DNA systems that no longer behave like flexible rods (see Chapter 1), few have been designed to detect directly, site-specifically and without base/backbone modifications such fast end-breathing or bending dynamics.

A strategy to elucidate the molecular origins for this complex motional behavior would involve several approaches: (i) permuting the current labeled residue probes to assess the end-position dependence, (ii) changing the sequence around the labeled probes to access the sequence dependence, (iii) changing the position of the entire labeled segment or swapping the two domains to characterize end-fraying contributions, (iv) varying the length of the double helix to approximate the timescales of these motions and test whether they arise from length-dependent bending modes that see enhanced flexing of the DNA ends, (v) changing the ambient conditions such as divalent salts to study whether A-tract induced bending at GC-rich regions nearby would amplify these rapid oscillations. In future experiments, we could also resort to RDCs to probe potentially slower bending motions with timescale sensitivity extending up to milliseconds. We anticipate it would be challenging and time-consuming to prepare these DNA constructs and separate contributions from

each of the proposed mechanisms in order to provide a coherent description for the internal dynamics.

Such experiments would benefit tremendously from the development of faster and less expensive strategies for sample preparation as well as suitable sensitive NMR methods for the study of larger DNA systems with unfavorable relaxation properties. Here, we demonstrated that the use of HMQC TROSY experiments that exploit slowly relaxing coherences in rapidly-rotating DNA methyl groups[3] could provide the required sensitivity to study sequence-specific nucleobase dynamics by using thymine as a probe. In the future, these experiments could also be applied to other chemically methylated nucleotides as model systems for damage dynamics or damage induction in long linear and supercoiled DNAs, or to examine damage recognition and capture by DNA repair proteins. The development of efficient isotopic labeling methods for synthetic nucleic acids, i.e. double methyl deuteration ($^{13}CD_2H$) or selective ^{13}C labeling at base (C8/C6) and sugar (C1'/C3')[4] and deuteration of unwanted protons in the base and sugar components, would be instrumental in quantitative TROSY-based relaxation measurements in nucleic acids. We can envision many applications of these methods towards the study of functional dynamics in large DNAs and DNA-protein complexes, hundreds of kDa in size, such as plasmids or nucleosome core particles (~ 300 kDa).

In Chapter 2, we also used short DNA helices for in-depth investigation of sequence specific ps-ns dynamics (< 4.5 ns) inside A-tract motives of variable length, their GC-rich junctions and other sequences. While non-terminal nucleobases exhibited similar constrained motions, we found large sequence-specific variations in the sugar-backbone (torsional) dynamics that correlated with rankings of dinucleotide conformational flexibility: A-tracts composed of rigid AA-steps were dynamically stiff as compared to flexible YR steps (CA and CG) that displayed large dynamic disorder. Though our study encompassed a range of sequences, a more comprehensive database of order parameters/motional timescales as a function of DNA sequence and environmental conditions (salts, hydration, pH) needs to be compiled so that we can understand the dynamic and conformational couplings over longer stretches of DNA, not simply dinucleotide steps. For example, our findings

that the A-tract length, and presumably distinct conformation, can modulate the dynamics of flanking sequences provides clues for the type of long-range communication that exists and is purely dictated by the DNA sequence, in the absence of external factors. To investigate the origins for sequence-specific backbone flexibility, NMR dynamics data should be supplemented by or fed into MD simulations to generate plausible 3D motional trajectories for the corresponding molecular transitions, as has been demonstrated for canonical[5,6], non-canonical[7] and damaged DNA[8]. To aid in these studies, one can incorporate mutations and chemical modifications into DNA base/backbone that bias the conformation in a predictable way to monitor changes in dynamics and select for best-matching motional models. Ultimately, the prospect of engineering gene regulation by protein-DNA or drug-DNA interactions relies on how well we understand and can manipulate sequence-dependent DNA dynamics.

5.2 Transient Hoogsteen base pairs and other excited states in DNA helices

In the course of these dynamics studies, we have shown that the information encoded in the B-DNA double helix is even richer than previously expected. In Chapter 3, we described the identification of transient Hoogsteen A•T and G•C base pairs in regular DNA duplexes that occur at ~ 0.1 - 1 % populations and ~ 0.1 - 2 ms lifetimes under physiologically relevant conditions. We have further demonstrated in Chapter 4 that the populations and lifetimes of these low-populated non-canonical states can depend significantly on environmental factors, chemical modifications, and DNA sequence. In fact, acidic conditions can induce large fractional population of HG states that could compete with WC B-DNA in the cell when the pH is not properly maintained and/or is acidic as a result of cellular stress. Moreover, alkylated damaged nucleotides that preclude formation of WC geometry can trap HG base pairs as their predominant states, which could play a role in damage recognition by DNA repair enzymes.

Such low-populated and short-lived states of DNA would be impossible to detect with conventional crystallographic techniques that capture conformational "snapshots" around the most stable, ground state. We also demonstrated the power of

NMR relaxation dispersion experiments, combined with steered computational tools and quantum-mechanical calculations, to characterize kinetics and thermodynamics of equilibria between ground and excited states as well as to resolve excited-state structures based on NMR carbon chemical shifts. The success of these experiments relies on a newly developed ^{13}C NMR pulse sequence with increased timescale sensitivity and improved scheme for suppressing undesired cross-relaxation and scalar coupling interactions, including those between like carbon spins in uniformly ^{13}C-labeled samples[9]. Needless to say, the design of efficient methods for selective or fractional ^{13}C labeling of DNA, some of which have already been implemented towards DNA[10] and RNA[4,11,12] by using isotopic mixtures or manipulating bacterial biosynthetic pathways, would remove interactions between spin probes that presently complicate $R_{1\rho}$ and CPMG relaxation dispersion experiments. Partial deuteration of nucleic acids[13,14], especially at the sugar ring, can lead to slower relaxation rates and higher sensitivity in the measurements of dynamics by relaxation dispersion.

In addition, the development of a suite of relaxation dispersion experiments that can target various nuclei susceptible to exchange – protons, protonated nitrogens or unprotonated nucleobase carbons/nitrogens – is currently underway in our laboratory. Such experiments would provide us with extra dynamics probes that (i) could be more sensitive to particular conformational changes which protonated carbons alone miss out on because of small/limited changes in their chemical shifts or (ii) would contribute additional exited state chemical shifts that might help constrain the conformation of the excited state. We have already developed and tested an imino ^{15}N $R_{1\rho}$ relaxation dispersion experiment on CA/TG steps that exhibit WC-to-HG transitions. Our preliminary data suggests that the imino nitrogen of guanines are also sensitive to the conformational switch that involves changes in the H-bonding pattern and H-bond distances/angles. Moreover, experiments that measure RDCs of the excited states, which exist for proteins[15], need to be tailored to the study of nucleic acids in order to obtain atomic resolution *structures* of low-populated states. Currently, the characterization of excited state structures relies heavily on knowledge of how the NMR chemical shift depends on conformation and the existence of feasible models of the conformational transitions. To aid with chemical-shift

informed prediction of conformations for "invisible" low-populated states, a comprehensive survey of DNA chemical shifts as a function of structure and sequence needs to be constructed. The integration of such chemical shift information can even help in *de novo* DNA structure prediction without the use of NOE and RDC constraints. Long MD simulations of DNA can be utilized to generate a pool of accessible conformations on longer (µs-ms) biologically relevant timescales and could narrow down the ensemble of feasible excited state structures. Chemical shift predictions could then select out the conformations that match the experimental chemical exchange parameters, much like the strategy we have used in the identification of Hoogsteen base pairs.

We are hopeful that our exciting discovery of transient Hoogsteen base pairs and novel strategy for identification of excited state structures will prompt many such investigations into the dynamics of canonical and non-canonical DNA motifs and architectures. It is essential that such studies be gradually translated to systems that mimic better the native, supercoiled state of DNA in the cell, such as nucleosome particles and supercoiled minicircles, and eventually performed *in vivo*, in the cellular milieu. Only then can we establish a more clear connection between transient DNA states and their hidden biological functions.

5.3 References

1. Richmond, T.J. & Davey, C.A. The structure of DNA in the nucleosome core. *Nature* **423**, 145-50 (2003).
2. Kim, Y., Geiger, J.H., Hahn, S. & Sigler, P.B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512-20 (1993).
3. Korzhnev, D.M., Kloiber, K., Kanelis, V., Tugarinov, V. & Kay, L.E. Probing slow dynamics in high molecular weight proteins by methyl-TROSY NMR spectroscopy: application to a 723-residue enzyme. *J Am Chem Soc* **126**, 3964-73 (2004).
4. Johnson, J.E., Jr. & Hoogstraten, C.G. Extensive backbone dynamics in the GCAA RNA tetraloop analyzed using 13C NMR spin relaxation and specific isotope labeling. *J Am Chem Soc* **130**, 16757-69 (2008).

5. Duchardt, E., Nilsson, L. & Schleucher, J. Cytosine ribose flexibility in DNA: a combined NMR 13C spin relaxation and molecular dynamics simulation study. *Nucleic Acids Res* **36**, 4211-9 (2008).

6. Nikolova, E.N. et al. Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **470**, 498-502 (2011).

7. Isaacs, R.J. & Spielmann, H.P. Insight into G[bond]T mismatch recognition using molecular dynamics with time-averaged restraints derived from NMR spectroscopy. *J Am Chem Soc* **126**, 583-90 (2004).

8. Chen, J., Dupradeau, F.Y., Case, D.A., Turner, C.J. & Stubbe, J. DNA oligonucleotides with A, T, G or C opposite an abasic site: structure and dynamics. *Nucleic Acids Res* **36**, 253-62 (2008).

9. Hansen, A.L., Nikolova, E.N., Casiano-Negroni, A. & Al-Hashimi, H.M. Extending the range of microsecond-to-millisecond chemical exchange detected in labeled and unlabeled nucleic acids by selective carbon R(1rho) NMR spectroscopy. *J Am Chem Soc* **131**, 3818-9 (2009).

10. Kojima, C., Ono, A., Kainosho, M. & James, T.L. DNA duplex dynamics: NMR relaxation studies of a decamer with uniformly 13C-labeled purine nucleotides. *J Magn Reson* **135**, 310-33 (1998).

11. Thakur, C.S., Sama, J.N., Jackson, M.E., Chen, B. & Dayie, T.K. Selective 13C labeling of nucleotides for large RNA NMR spectroscopy using an E. coli strain disabled in the TCA cycle. *J Biomol NMR* **48**, 179-92.

12. Dayie, T.K. & Thakur, C.S. Site-specific labeling of nucleotides for making RNA for high resolution NMR studies using an E. coli strain disabled in the oxidative pentose phosphate pathway. *J Biomol NMR* **47**, 19-31.

13. Cromsigt, J., Schleucher, J., Gustafsson, T., Kihlberg, J. & Wijmenga, S. Preparation of partially 2H/13C-labelled RNA for NMR studies. Stereo-specific deuteration of the H5" in nucleotides. *Nucleic Acids Res* **30**, 1639-45 (2002).

14. Lukin, M. & de los Santos, C. Stereoselective nucleoside deuteration for NMR studies of DNA. *Nucleosides Nucleotides Nucleic Acids* **29**, 562-73.

15. Korzhnev, D.M., Religa, T.L., Banachewicz, W., Fersht, A.R. & Kay, L.E. A transient and low-populated protein-folding intermediate at atomic resolution. *Science* **329**, 1312-6 (2010).

Appendix 1

A$_6$-DNA (A16)
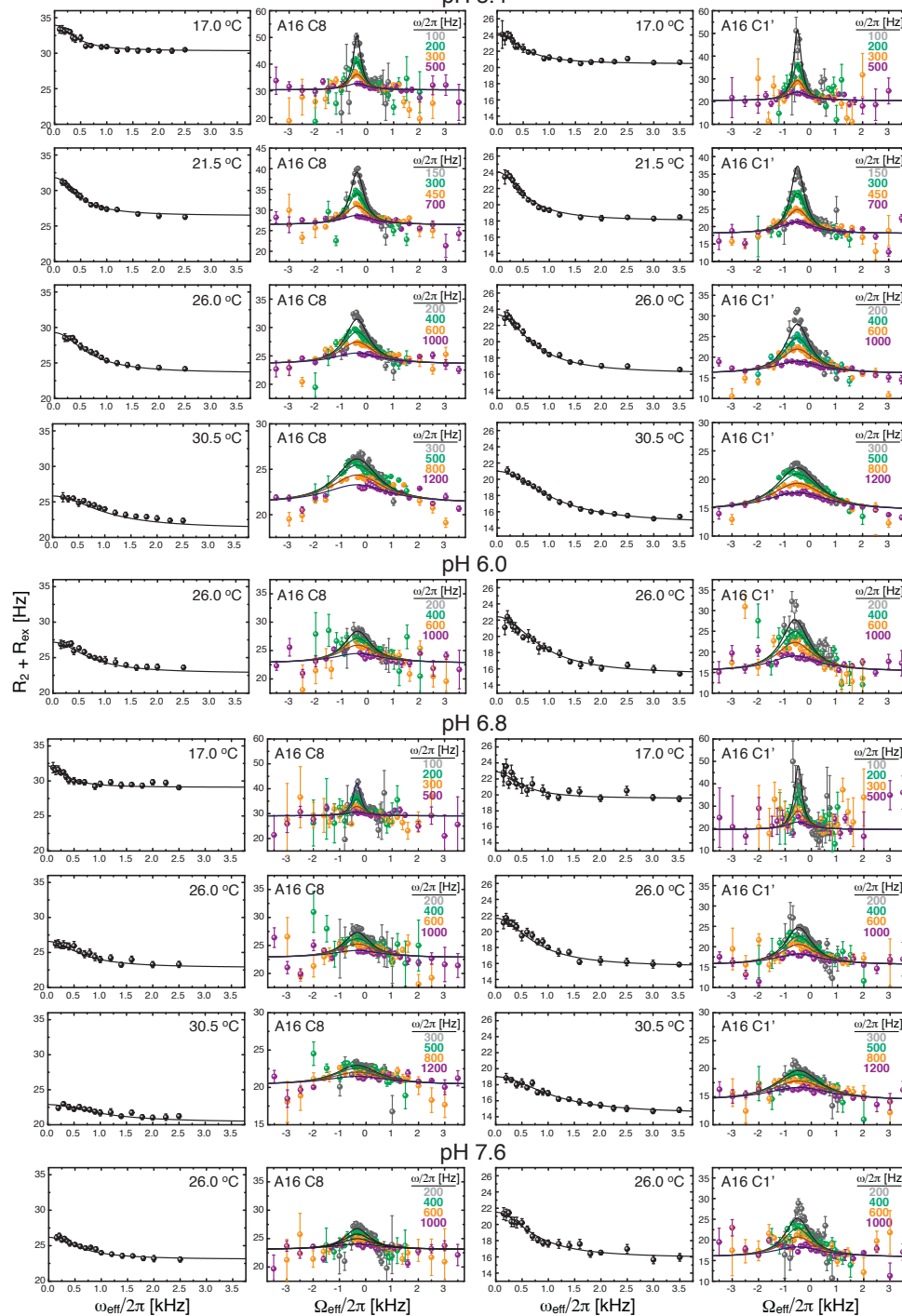
Figure A1.1: Sample carbon $R_{1\rho}$ relaxation dispersion profiles for A16 C8 and C1' in A$_6$-DNA at variable temperature and pH. Shown are best fits to the Eq.3.1.
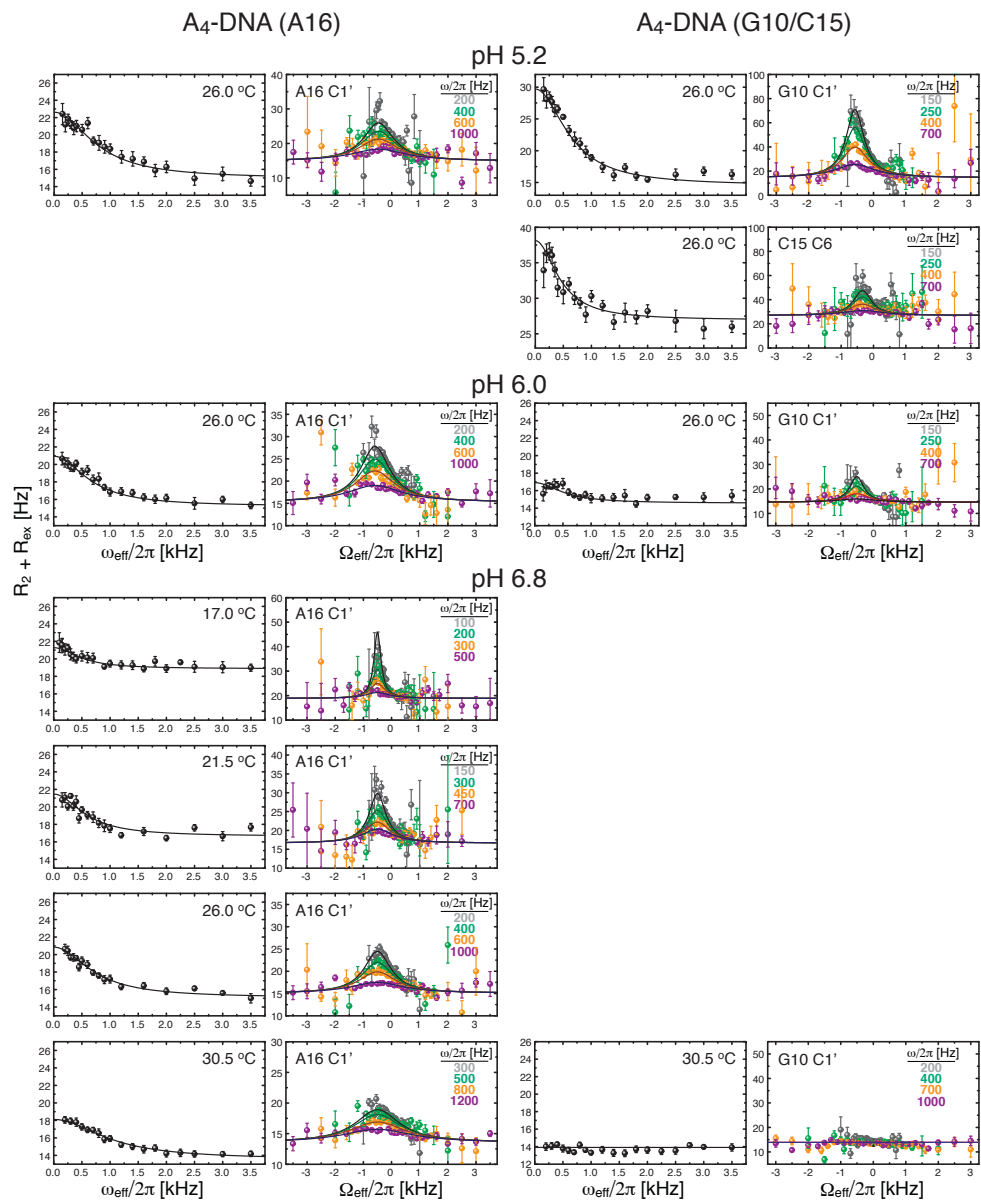
Figure A1.2: Carbon $R_{1\rho}$ relaxation dispersion profiles for A16 C1', G10 C1' and C15 C1' in A$_4$-DNA at variable temperature and pH. Shown are best fits to the Eq.3.1
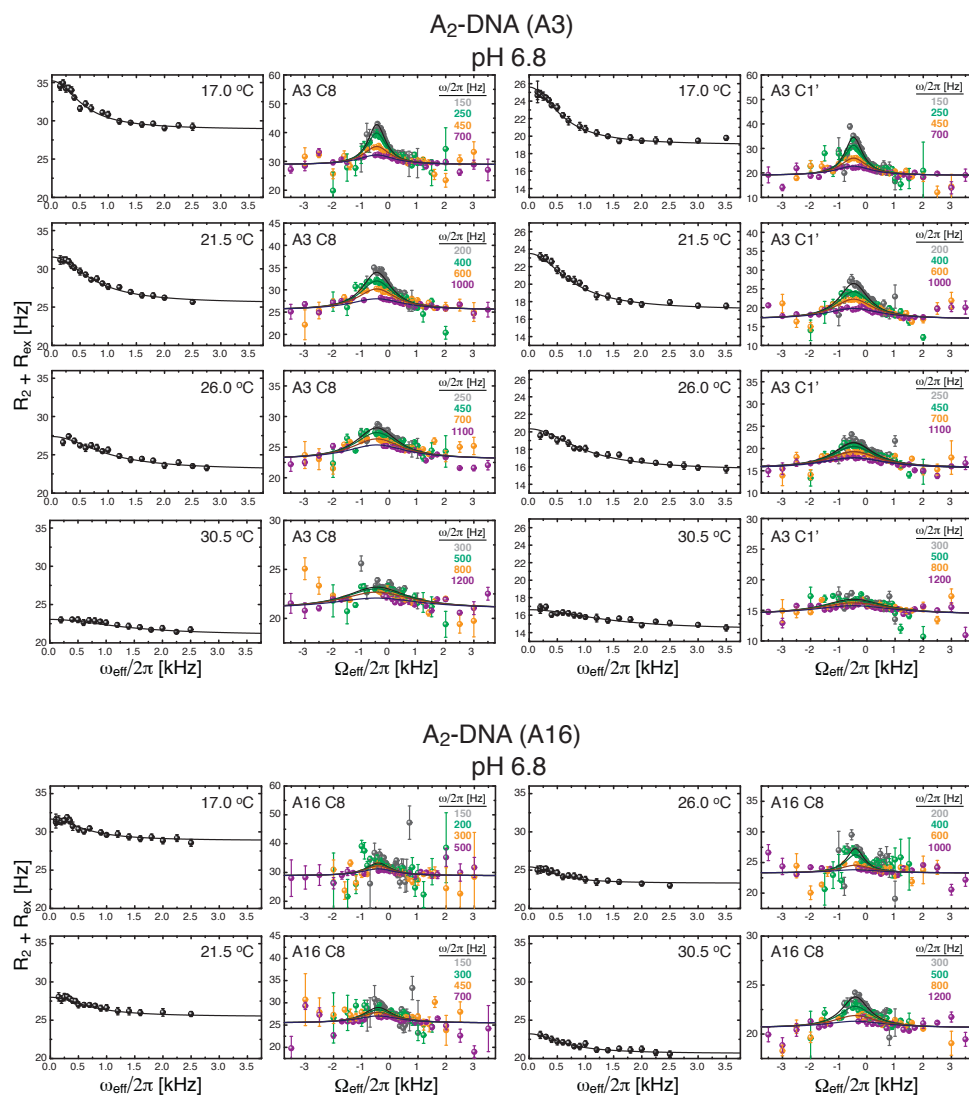
Figure A1.3: Carbon $R_{1\rho}$ relaxation dispersion profiles for A16 C8 and A3 C8 and C1' in A_2-DNA at variable temperature. Shown are best fits to the Eq. 3.1.
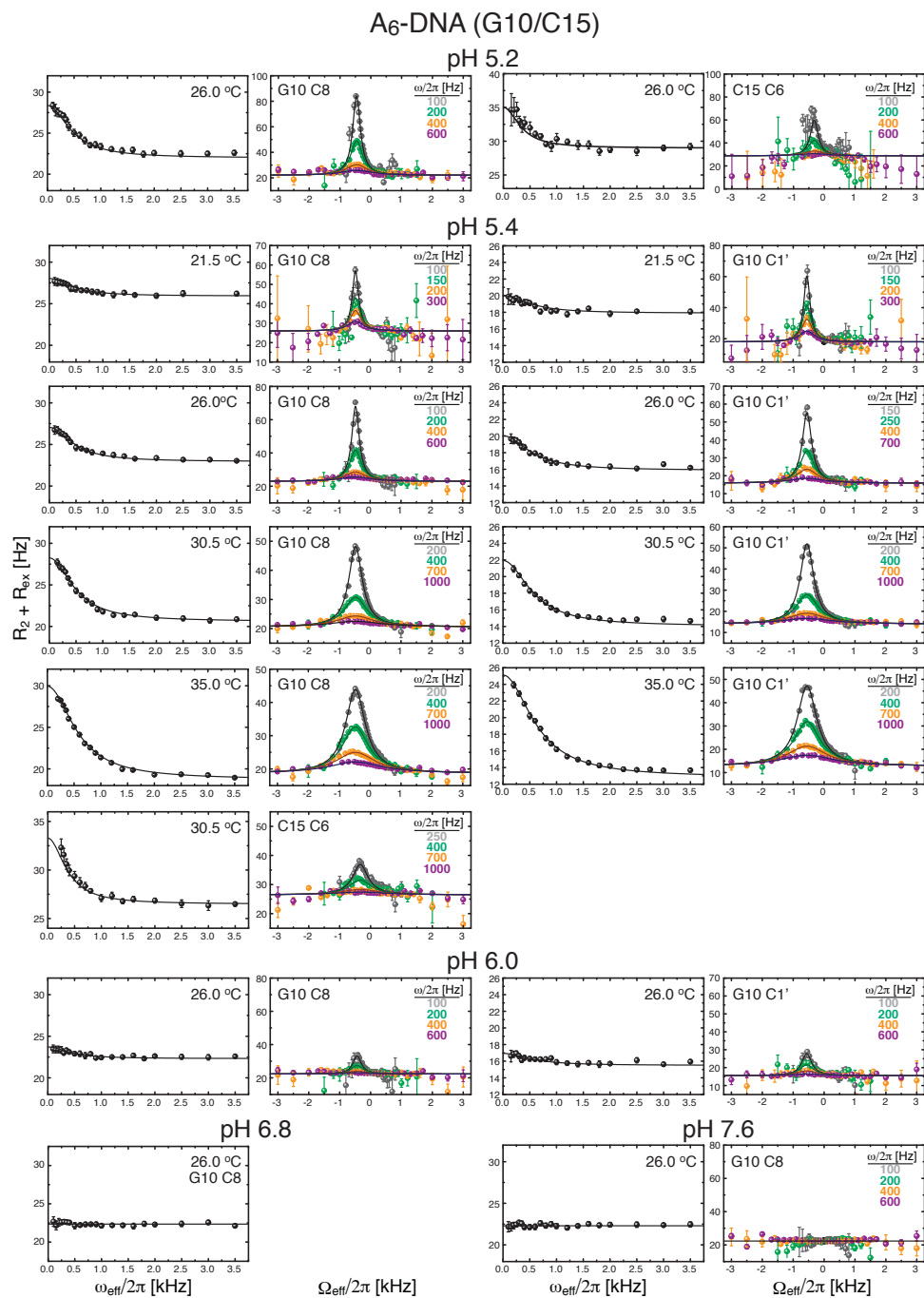
Figure A1.4: Sample carbon $R_{1\rho}$ relaxation dispersion profiles for G10 C8 and C1' in A_6-DNA at variable temperature and pH. Shown are best fits to the Eq.3.1.
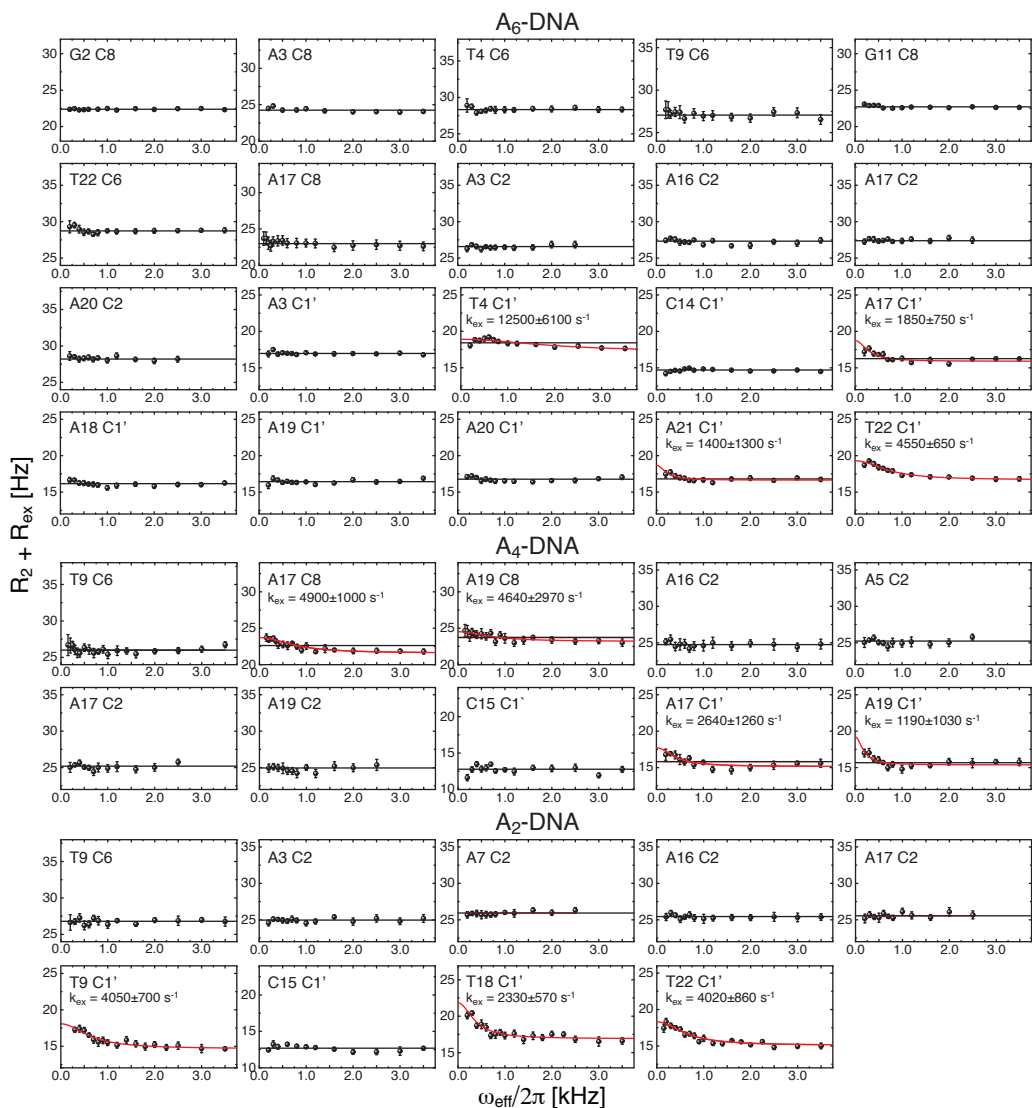
Figure A1.5: Carbon $R_{1\rho}$ relaxation dispersion profiles showing $R_2 + R_{ex}$ as a function of on-resonance spinlock power ($\omega_{eff}/2\pi$) for different sites in A_6-DNA, A_4-DNA, and A_2-DNA at 26 ^{o}C and pH 5.4. Solid lines indicate best fits to the simplified fast exchange model (Eq.3.2) in the absence (black) or presence (red) of chemical exchange. Small R_{ex} contributions approaching the detection limit were observed for i) adenine C8 and C1' sites outside of CA/TG steps, which could also reflect WC-to-HG transitions, and for ii) thymine C1' sites at and outside of CA/TG steps that could be due to concomitant backbone rearrangements.
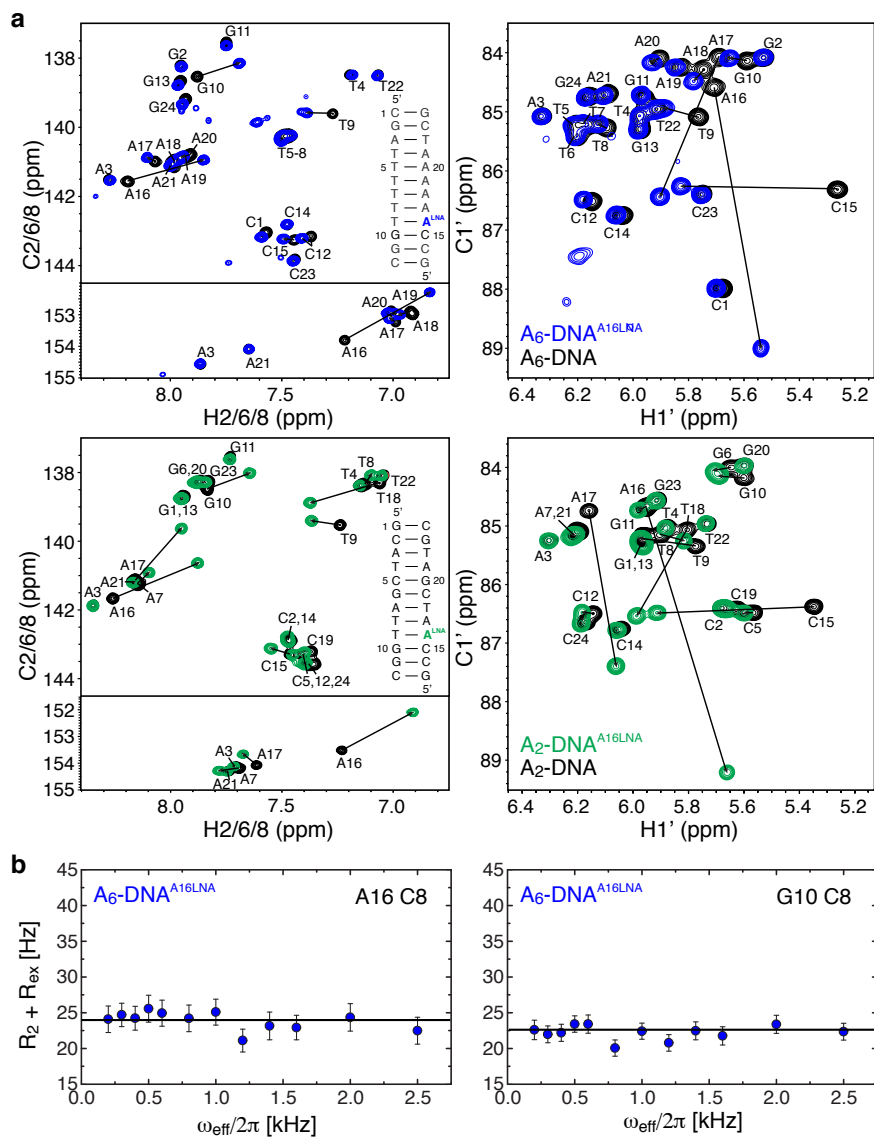
Figure A1.6: Effect of C3'- endo locked nucleic acid (LNA) on chemical shifts and chemical exchange at CA/TG steps. (a) 2D ^{1}H-^{13}C HSQC spectra and assignments for LNA-modified unlabeled A_6-DNA (A_6-DNA^{A16LNA} in blue) and A_2-DNA (A_2-DNA^{A16LNA} in green) showing base/sugar chemical shift perturbations at the CA/TG step relative to A_6-DNA or A_2-DNA (black) at 26 ^{o}C and pH 6.8. (b) On-resonance ^{13}C $R_{1\rho}$ relaxation dispersion profiles at 30.5 ^{o}C for A16^{LNA} C8 (left) and G10 C8 (right) in A_6-DNA^{A16LNA} showing lack of detectable chemical exchange contrary to A16 C8 sites in unmodified A_6-DNA.
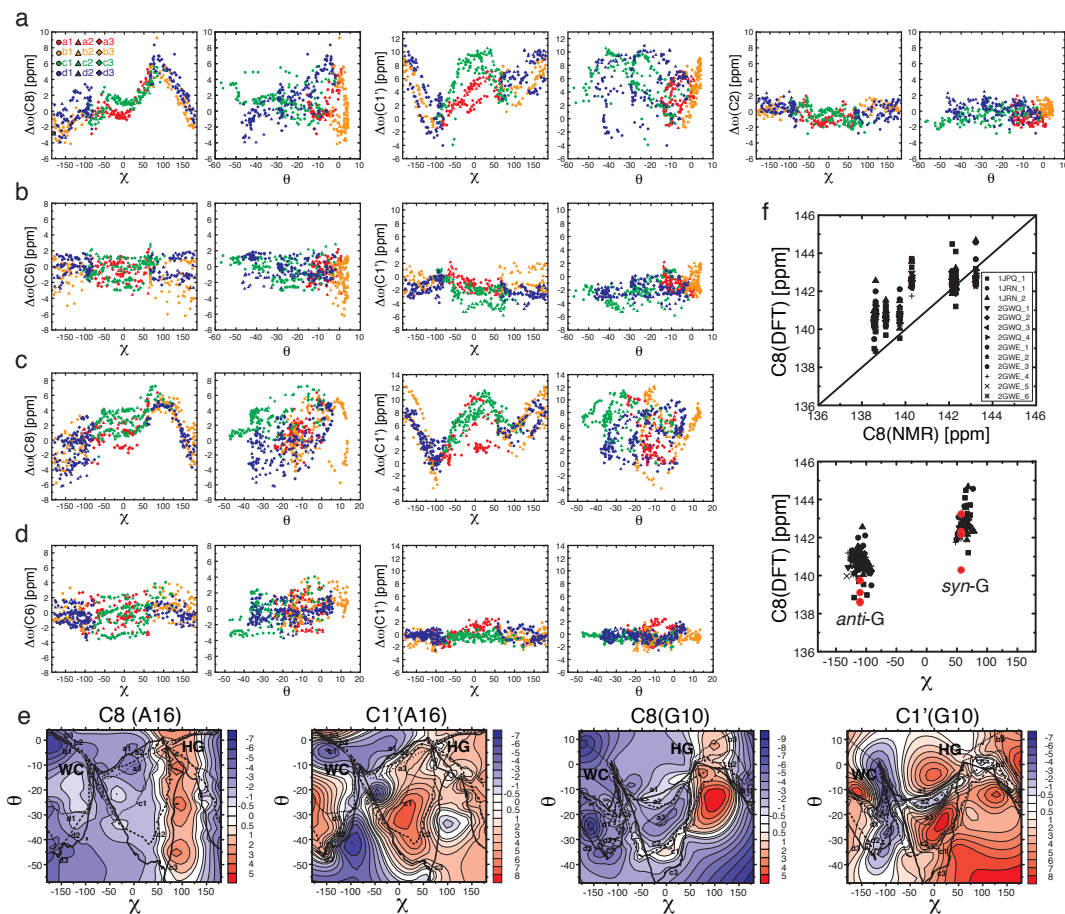
Figure A1.7: DFT chemical shift calculations for the CA/TG step in A$_6$-DNA. a-d, DFT predictions for (a) A16 C8, C1', and C2, (b) T9 C6 and C1', c G10 C8 and C1', (d) C15 C6 and C1' chemical shift changes ($\Delta\omega_{AB}$) as a function of χ and θ and relative to the minimal-energy MD-generated WC conformation for each WC-to-HG transition pathway color-coded as in Fig. S7. (e) Contour plots of Score C8 and Score C1' versus (χ, θ) coordinate pairs for A16 and G10 constructed by combining all data sets in a and c respectively, where the Score value was calculated as $\Delta\omega_{AB,DFT}$ -$\Delta\omega_{AB,NMR}$ (e.g. white color is the best match to $\Delta\omega_{AB,NMR}$). (f) Correlation between NMR C8 chemical shifts for *anti* and *syn* guanines in a G-quadruplex and corresponding DFT predictions for multiple crystal structures (top, see inset for PDB ID) and the spread of measured (red) and predicted (black) chemical shifts versus χ (bottom) for *anti* and *syn* guanines.

166

Table A1.1: Parameter sets for off-resonance $R_{1\rho}$ relaxation dispersion profiles of A16 C8/C1' (A$_6$-DNA, A$_4$-DNA, A$_2$-DNA), A3 C8/C1' (A$_2$-DNA), and G10 C8/C1' & C15 C6 (A$_6$-DNA, A$_4$-DNA) measured at 14.1 T and variable temperature and pH.

Residue (spin)	T (°C)	On-resonance spinlock power (ω_1)/ Off-resonance spinlock power (ω_1) & ±{offset (Ω)}
A16 (C8/C1')	17.0 (pH 5.4 & 6.8)	100, 150, 200, 250, 300, 350, 400, 500, 600, 700, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2250, 2500, 3000, 3500 Hz/ 100 Hz & ±{30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 360, 420, 480, 540, 600, 700, 800, 1000 Hz} 200 Hz & ±{50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1200, 1500, 2000 Hz} 300 Hz & ±{75, 150, 225, 300, 375, 450, 550, 650, 800, 1000, 1200, 1400, 1600, 2000, 2500, 3000, 4000, 5000 Hz} 500 Hz & ±{100, 200, 300, 400, 500, 600, 700, 900, 1100, 1300, 1600, 2000, 2500, 3000, 3500, 4000, 5000, 7500 Hz}
	21.5 (pH 5.4 & 6.8)	100, 150, 200, 250, 300, 350, 400, 500, 600, 700, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2250, 2500, 3000, 3500 Hz/ 150 Hz & ±{30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 360, 420, 480, 540, 600, 700, 800, 1000 Hz} 300 Hz & ±{50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1200, 1500, 2000 Hz} 450 Hz & ±{75, 150, 225, 300, 375, 450, 550, 650, 800, 1000, 1200, 1400, 1600, 2000, 2500, 3000, 4000, 5000 Hz} 700 Hz & ±{100, 200, 300, 400, 500, 600, 700, 900, 1100, 1300, 1600, 2000, 2500, 3000, 3500, 4000, 5000, 7500 Hz}
	26.0 (pH 5.2-7.6)	150, 200, 250, 300, 350, 400, 500, 600, 700, 900, 1000, 1200, 1400, 1600, 2000, 2500, 3000, 3500 Hz/ 200 Hz & ±{30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 360, 420, 480, 540, 600, 700, 800, 1000 Hz} 400 Hz & ±{50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1200, 1500, 2000 Hz} 600 Hz & ± {75, 150, 225, 300, 375, 450, 550, 650, 800, 1000, 1200, 1400, 1600, 2000, 2500, 3000, 4000, 5000 Hz} 1000 Hz & ± {100, 200, 300, 400, 500, 600, 700, 900, 1100, 1300, 1600, 2000, 2500, 3000, 3500, 4000, 5000, 7500 Hz}
	30.5 (pH 5.4 & 6.8)	200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2250, 2500, 3000, 3500 Hz/ 300 Hz ±{30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 360, 420, 480, 540, 600, 700, 800, 1000 Hz} 500 Hz ±{50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1200, 1500, 2000 Hz} 800 Hz ± {75, 150, 225, 300, 375, 450, 550, 650, 800, 1000, 1200, 1400, 1600, 2000, 2500, 3000, 4000, 5000 Hz} 1200 Hz ± {100, 200, 300, 400, 500, 600, 700, 900, 1100, 1300, 1600, 2000, 2500, 3000, 3500, 4000, 5000, 7500 Hz}
	17.0 (pH 6.8)	150, 200, 250, 300, 350, 400, 500, 600, 700, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2250, 2500, 3000, 3500 Hz/ 150 Hz & ±{30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 360, 420, 480, 540, 600, 700, 800, 1000 Hz} 250 Hz & ±{50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1200, 1500, 2000 Hz} 450 Hz & ±{75, 150, 225, 300, 375, 450, 550, 650, 800, 1000, 1200, 1400, 1600, 2000, 2500, 3000, 4000, 5000 Hz} 700 Hz & ±{100, 200, 300, 400, 500, 600, 700, 900, 1100, 1300, 1500, 1700, 2000, 2500, 3000, 3500, 4000, 5000, 7500 Hz}
		200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2500, 3000, 3500 Hz

A3 (C8/C1')	21.5 (pH 6.8)	200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2500, 3000, 3500 Hz 200 Hz & ±{30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 360, 420, 480, 540, 600, 700, 800, 1000 Hz} 400 Hz & ±{50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1200, 1500, 2000 Hz} 600 Hz & ±{75, 150, 225, 300, 375, 450, 550, 650, 800, 1000, 1200, 1400, 1600, 2000, 2500, 3000, 4000, 5000 Hz} 1000 Hz & ±{100, 200, 300, 400, 500, 600, 700, 900, 1100, 1300, 1500, 1700, 2000, 2500, 3000, 3500, 4000, 5000, 7500 Hz}
	26.0 (pH 6.8)	200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2250, 2500, 2750, 3000, 3500 Hz 250 Hz & ±{30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 360, 420, 480, 540, 600, 700, 800, 1000 Hz} 450 Hz & ±{50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1200, 1500, 2000 Hz} 700 Hz & ±{75, 150, 225, 300, 375, 450, 550, 650, 800, 1000, 1200, 1400, 1600, 2000, 2500, 3000, 4000, 5000 Hz} 1100 Hz & ±{100, 200, 300, 400, 500, 600, 700, 900, 1100, 1300, 1500, 1700, 2000, 2500, 3000, 3500, 4000, 5000, 7500 Hz}
	30.5 (pH 6.8)	200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2250, 2500, 3000, 3500 Hz 300 Hz & ±{30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 360, 420, 480, 540, 600, 700, 800, 1000 Hz} 500 Hz & ±{50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1200, 1500, 2000 Hz} 800 Hz & ±{75, 150, 225, 300, 375, 450, 550, 650, 800, 1000, 1200, 1400, 1600, 2000, 2500, 3000, 4000, 5000 Hz} 1200 Hz & ±{100, 200, 300, 400, 500, 600, 700, 900, 1100, 1300, 1500, 1700, 2000, 2500, 3000, 3500, 4000, 5000, 7500 Hz}
G10 (C8/C1') & C15 (C6)	21.5 (pH 5.4)	100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 2000, 2500, 3500 Hz 100 Hz & ±{30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 330, 360, 390, 420, 480, 540, 600, 700, 800 Hz} 150 Hz & ±{50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 700, 800, 900, 1000, 1200, 1500 Hz} 200 Hz & ±{75, 150, 225, 300, 375, 450, 500, 650, 800, 1000, 1200, 1400, 1600, 2000, 2500, 3000, 3500 Hz} 300 Hz & ±{100, 200, 300, 400, 500, 700, 900, 1100, 1300, 1500, 1700, 2000, 2500, 3000, 3500, 4000, 5000 Hz}
	26.0 (pH 5.2-7.6)	100, 150, 200, 250, 300, 350, 400, 500, 600, 700, 800, 1000, 1200, 1400, 1600, 2000, 2500, 3000, 3500 Hz 100 or 150 Hz & ±{30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 330, 360, 420, 480, 540, 600, 700, 800 Hz} 200 or 250 Hz & ±{50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 700, 800, 900, 1000, 1200, 1500 Hz} 400 Hz & ±{75, 150, 225, 300, 375, 450, 500, 650, 800, 1000, 1200, 1400, 1600, 2000, 2500, 3000, 3500 Hz} 600 or 700 Hz & ±{100, 200, 300, 400, 500, 600, 700, 900, 1100, 1300, 1500, 1700, 2000, 2500, 3000, 3500, 5000 Hz}
	30.5 (pH 5.4)	150, 200, 250, 300, 350, 400, 500, 600, 700, 800, 1000, 1200, 1400, 1600, 2000, 2500, 3000, 3500 Hz 200 (250 for C6) Hz & ±{30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 360, 420, 480, 540, 600, 700, 800, 1000 Hz} 400 Hz & ± {50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1200, 1500, 2000 Hz} 700 Hz & ± {75, 150, 225, 300, 375, 450, 550, 650, 800, 1000, 1200, 1400, 1600, 2000, 2500, 3000, 4000, 5000 Hz} 1000 Hz & ± {100, 200, 300, 400, 500, 600, 700, 900, 1100, 1300, 1600, 2000, 2500, 3000, 3500, 4000, 5000, 7500 Hz}
	35.0	200, 250, 300, 350, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 2000, 2500, 3000, 3500 Hz 200 Hz & ± {30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 360, 420, 480, 540, 600, 700, 800, 1000 Hz}

(pH 5.4)   400 Hz & ± {50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1200, 1500, 2000 Hz}

700 Hz & ± {75, 150, 225, 300, 375, 450, 550, 650, 800, 1000, 1200, 1400, 1600, 2000, 2500, 3000, 4000, 5000 Hz}

1000 Hz & ± {100, 200, 300, 400, 500, 600, 700, 900, 1100, 1300, 1600, 2000, 2500, 3000, 3500, 4000, 5000, 7500 Hz}

Table A1.2: Chemical exchange parameters and computed forward and backward exchange rates (k_A, k_B) representing best global and corresponding individual fits of off-resonance $R_{1\rho}$ relaxation dispersion data to Eq.3.1 collected at 14.1 T ([a]pH 5.4 and [b]pH 6.8)

Residue	Parameter	17.0 (°C)	21.5 (°C)	26.0 (°C)	30.5 (°C)	17.0 (°C)	21.5 (°C)	26.0 (°C)	30.5 (°C)
		Global Fit				Individual Fit			
	$p_{B,C8}$ (%)	0.439±0.012	0.453±0.005	0.468±0.009	0.507±0.017	0.464±0.016	0.453±0.011	0.346±0.020	0.561±0.089
	$p_{B,C1'}$ (%)					0.394±0.017	0.452±0.007	0.491±0.016	0.451±0.025
	$k_{ex,C8}$ (s⁻¹)	937±50	1750±50	3270±90	5890±180	916±62	1660±70	3180±170	5660±420
	$k_{ex,C1'}$ (s⁻¹)					1060±80	1850±60	3270±100	5910±190
	$k_{A,C8}$ (s⁻¹)	4.11±0.25	7.91±0.23	15.3±0.5	29.9±1.4	4.25±0.32	7.52±0.37	11.0±0.9	31.8±5.6
	$k_{A,C1'}$ (s⁻¹)					4.18±0.38	8.33±0.32	16.04±0.73	26.7±1.7
A16	$k_{B,C8}$ (s⁻¹)	933±50	1740±50	3260±90	5860±180	912±62	1650±70	3170±170	5630±420
(C8/C1')	$k_{B,C1'}$ (s⁻¹)					1060±80	1840±60	3250±100	5880±190
A₆-DNAᵃ	$R_{1,C8}$ (Hz)	2.00±0.02	2.17±0.02	2.36±0.02	2.66±0.02	1.99±0.02	2.16±0.02	2.33±0.02	2.65±0.02
	$R_{2,C8}$ (Hz)	30.3±0.1	26.5±0.1	23.6±0.1	21.2±0.1	30.2±0.1	26.5±0.1	23.6±0.1	21.4±0.1
	$R_{1,C1'}$ (Hz)	1.79±0.03	1.83±0.02	2.06±0.02	2.18±0.02	1.83±0.03	1.85±0.02	2.06±0.02	2.18±0.02
	$R_{2,C1'}$ (Hz)	20.4±0.1	18.0±0.1	16.1±0.1	14.5±0.1	20.5±0.1	18.0±0.1	16.0±0.1	14.5±0.1
	$\Delta\omega_{AB,C8}$ (ppm)	2.66±0.02				2.67±0.04	2.71±0.05	3.07±0.10	2.42±0.20
	$\Delta\omega_{AB,C1'}$ (ppm)	3.27±0.02				3.22±0.06	3.24±0.03	3.21±0.06	3.52±0.09
	$p_{B,C1'}$ (%)	0.347±0.040	0.333±0.021	0.368±0.015	0.343±0.026	0.343±0.040	0.383±0.021	0.351±0.018	0.293±0.031
	$k_{ex,C1'}$ (s⁻¹)	728±117	2070±160	3410±180	5690±390	735±118	2090±150	3390±180	5710±400
A16	$k_{A,C1'}$ (s⁻¹)	2.52±0.50	6.91±0.60	12.5±0.8	19.5±2.0	2.52±0.50	6.91±0.60	12.5±0.8	19.5±2.0
(C1')	$k_{B,C1'}$ (s⁻¹)	725±117	2070±160	3390±180	5670±390	733±118	2080±150	3380±180	5700±390
A₄-DNAᵇ	$R_{1,C1'}$ (Hz)	1.87±0.03	1.90±0.04	1.96±0.03	2.29±0.03	1.87±0.03	1.93±0.04	1.95±0.03	2.26±0.03
	$R_{2,C1'}$ (Hz)	18.9±0.2	16.6±0.2	15.1±0.1	13.5±0.2	18.9±0.1	16.6±0.2	15.1±0.1	13.5±0.2
	$\Delta\omega_{AB,C1'}$ (ppm)	3.32±0.06				3.37±0.10	2.93±0.12	3.43±0.10	3.66±0.19
	$p_{B,C8}$ (%)	0.224±0.020	0.211±0.031	0.169±0.011	0.205±0.019	0.282±0.032	0.272±0.062	0.140±0.012	0.167±0.020
	$k_{ex,C8}$ (s⁻¹)	2680±290	3820±530	2090±260	3260±380	2470±260	3870±520	2330±300	3390±400
	$k_{A,C8}$ (s⁻¹)	5.99±0.83	8.06±1.62	3.53±0.50	6.68±0.98	6.94±1.07	10.52±2.77	3.26±0.50	5.65±0.98

A16 (C8) A$_2$-DNA[b]								
$k_{B,C8}$ (s^{-1})	2670±290	3810±530	2080±260	3260±380	2460±260	3860±520	2330±300	3380±400
$R_{1,C8}$ (Hz)	2.01±0.02	2.18±0.02	2.39±0.02	2.73±0.03	2.02±0.02	2.19±0.02	2.38±0.02	2.72±0.03
$R_{2,C8}$ (Hz)	28.9±0.2	25.5±0.1	23.3±0.1	20.6±0.1	28.9±0.2	25.4±0.1	23.2±0.1	20.6±0.1
$\Delta\omega_{AB,C8}$ (ppm)	2.73±0.11				2.29±0.15	2.36±0.27	3.28±0.25	3.15±0.25
A3 (C8/C1') A$_2$-DNA[b]								
$p_{B,C8}$ (%)	0.464±0.007	0.449±0.012	0.376±0.017	0.239±0.032	0.473±0.015	0.404±0.021	0.381±0.049	0.227±0.075
$p_{B,C1'}$ (%)					0.477±0.011	0.465±0.024	0.289±0.075	0.127±0.029
$k_{ex,C8}$ (s^{-1})	2240±60	3790±100	5340±210	8220±760	2200±80	3733±157	6320±400	9980±1500
$k_{ex,C1'}$ (s^{-1})					2210±70	3869±133	4800±260	7940±990
$k_{A,C8}$ (s^{-1})	10.4±0.3	17.0±0.6	20.1±1.2	19.7±3.2	10.4±0.5	15.1±1.0	24.1±3.5	22.7±8.2
$k_{A,C1'}$ (s^{-1})					10.5±0.4	18.0±1.1	13.8±1.3	10.1±2.6
$k_{B,C8}$ (s^{-1})	2220±60	3770±101	5320±210	8200±750	2190±80	3718±157	6300±400	9960±1490
$k_{B,C1'}$ (s^{-1})					2200±70	3851±132	4780±260	7930±990
$R_{1,C8}$ (Hz)	2.05±0.02	2.21±0.02	2.35±0.02	2.64±0.02	2.06±0.02	2.20±0.02	2.35±0.02	2.66±0.03
$R_{2,C8}$ (Hz)	28.8±0.1	25.5±0.1	23.0±0.1	21.0±0.1	29.0±0.1	25.5±0.1	22.7±0.2	20.5±0.3
$R_{1,C1'}$ (Hz)	1.87±0.02	1.88±0.02	2.07±0.02	2.15±0.02	1.87±0.02	1.89±0.02	2.04±0.03	2.13±0.02
$R_{2,C1'}$ (Hz)	19.0±0.1	17.0±0.1	15.6±0.1	14.3±0.2	19.0±0.1	17.0±0.1	15.8±0.1	14.5±0.2
$\Delta\omega_{AB,C8}$ (ppm)	2.97±0.03				2.87±0.06	3.20±0.10	3.21±0.21	3.72±0.59
$\Delta\omega_{AB,C1'}$ (ppm)	3.14±0.04				3.09±0.05	3.07±0.08	3.54±0.15	4.17±0.41
A16 (C8/C1') A$_6$-DNA[b]								
$p_{B,C8}$ (%)	0.378±0.018		0.386±0.019	0.359±0.018	0.384±0.021		0.393±0.072	0.859±0.491
$p_{B,C1'}$ (%)					0.345±0.036		0.418±0.032	0.353±0.048
$k_{ex,C8}$ (s^{-1})	944±78		3680±200	6080±390	936±98		3290±390	5930±420
$k_{ex,C1'}$ (s^{-1})					1020±160		3780±240	6170±430
$k_{A,C8}$ (s^{-1})	3.57±0.34		14.2±1.03	21.8±2.3	3.60±0.42		12.9±2.9	50.9±30.3
$k_{A,C1'}$ (s^{-1})					3.53±0.67		15.8±1.6	21.8±3.33
$k_{B,C8}$ (s^{-1})	941±78		3670±200	6060±390	936±98		3280±390	5880±930
$k_{B,C1'}$ (s^{-1})					1020±160		3760±230	6150±430
$R_{1,C8}$ (Hz)	1.70±0.02		2.41±0.04	2.67±0.04	1.70±0.02		2.41±0.04	2.67±0.04
$R_{2,C8}$ (Hz)	29.1±0.1		23.0±0.2	20.4±0.3	29.1±0.1		23.0±0.2	20.4±0.3
$R_{1,C1'}$ (Hz)	1.64±0.05		1.97±0.04	1.98±0.04	1.64±0.05		1.99±0.04	1.98±0.04
$R_{2,C1'}$ (Hz)	19.5±0.2		15.6±0.2	14.3±0.1	19.5±0.2		15.5±0.2	14.5±0.1
$\Delta\omega_{AB,C8}$ (ppm)	2.31±0.05				2.30±0.07		2.21±0.22	1.45±0.45
$\Delta\omega_{AB,C1'}$ (ppm)	3.35±0.08				3.58±0.15		3.20±0.15	3.41±0.22

	21.5 (°C)	26.0 (°C)	30.5 (°C)	35.0 (°C)	21.5 (°C)	26.0 (°C)	30.5 (°C)	35.0 (°C)
	Global Fit				Individual Fit			
$p_{B,C8}$ (%)					0.551±0.078	0.683±0.023	0.806±0.010	0.798±0.006
$p_{B,C1'}$ (%)	0.452±0.033	0.637±0.016	0.788±0.008	0.799±0.003	0.432±0.039	0.608±0.024	0.749±0.013	0.785±0.007
$p_{B,C6}$ (%)							0.682±0.055	
$k_{ex,C8}$ (s⁻¹)					367±62	603±28	1110±20	2110±30
$k_{ex,C1'}$ (s⁻¹)	473±44	673±24	1090±20	2070±20	522±61	734±40	1120±30	2060±30
$k_{ex,C6}$ (s⁻¹)							1090±130	
$k_{A,C8}$ (s⁻¹)					2.02±0.45	4.12±0.24	8.96±0.23	16.85±0.24
$k_{A,C1'}$ (s⁻¹)	2.14±0.26	4.29±0.19	8.60±0.17	16.6±0.2	2.26±0.33	4.46±0.30	8.37±0.27	16.19±0.28
$k_{A,C6}$ (s⁻¹)							7.43±1.09	
$k_{B,C8}$ (s⁻¹)					365±62	599±28	1100±20	2090±30
$k_{B,C1'}$ (s⁻¹)	471±44	669±24	1080±20	2060±20	520±61	729±40	1110±30	2050±30
$k_{B,C6}$ (s⁻¹)							5320±210	
$R_{1,C8}$ (Hz)	2.16±0.01	2.44±0.01	2.65±0.01	2.97±0.02	2.17±0.01	2.44±0.01	2.65±0.01	2.96±0.02
$R_{2,C8}$ (Hz)	25.9±0.2	22.9±0.0	20.6±0.0	18.7±0.0	26.0±0.1	23.1±0.1	20.5±0.0	18.6±0.0
$R_{1,C1'}$ (Hz)	1.66±0.02	1.86±0.01	2.04±0.02	2.19±0.1	1.66±0.02	1.86±0.02	2.05±0.02	2.19±0.01
$R_{2,C1'}$ (Hz)	17.9±0.1	15.9±0.0	14.0±0.0	12.8±0.0	17.9±0.1	15.7±0.1	14.1±0.0	12.9±0.1
$R_{1,C6}$ (Hz)			3.09±0.03				3.06±0.03	
$R_{2,C6}$ (Hz)			26.5±0.1				26.0±0.1	
$\Delta\omega_{AB,C8}$ (ppm)	3.16±0.01				3.02±0.03	3.09±0.02	3.16±0.01	3.19±0.02
$\Delta\omega_{AB,C1'}$ (ppm)	3.70±0.01				3.67±0.04	3.82±0.03	3.74±0.02	3.72±0.02
$\Delta\omega_{AB,C6}$ (ppm)	2.22±0.03						2.42±0.07	

Row group label: G10 (C8/C1') & C15(C6) A₆-DNA[a]

Table A1.3: Parameter sets for experimental off-resonance $R_{1\rho}$ relaxation dispersion profiles of A5 C8/C1' in E-DNA collected at 14.1 T and pH 6.8, and best global and corresponding individual data fits to Eq. 3.1.

Parameter	17.0 °C		26.0 °C	
On-resonance spinlock power (ω_1)	200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2500, 3000, 3500 Hz		200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2500, 3000, 3500 Hz	
Off-resonance spinlock power (ω_1) & ±{offset (Ω)}	150 Hz & ±{30, 60, 90, 120, 180, 240, 300, 420, 480, 540, 600, 700 Hz} 300 Hz & ±{50, 100, 150, 200, 250, 300, 400, 500, 600, 800, 1000, 1200 Hz} 600 Hz & ±{75, 150, 225, 300, 375, 450, 600, 800, 1000, 1400, 1600, 2000, 2500, 3000 Hz}		200 Hz & ±{30, 60, 90, 120, 180, 240, 300, 420, 480, 540, 600, 700 Hz} 500 Hz & ±{50, 100, 150, 200, 250, 300, 400, 500, 600, 800, 1000, 1200 Hz} 1000 Hz & ±{75, 150, 225, 300, 375, 450, 600, 800, 1000, 1400, 1600, 2000, 2500, 3000 Hz}	
	Global Fit	*Individual Fit*	*Global Fit*	*Individual Fit*
$p_{B,C8}$ (%)	0.822±0.059	0.806±0.087	0.800±0.142	2.00±1.90
$p_{B,C1'}$ (%)		0.815±0.075		0.756±0.670
$k_{ex,C8}$ (s⁻¹)	2900±190	2870±290	8670±940	9200±1380
$k_{ex,C1'}$ (s⁻¹)		2920±270		7360±1120
$k_{A,C8}$ (s⁻¹)	23.8±2.4	23.0±3.5	69.4±14.4	185±174
$k_{A,C1'}$ (s⁻¹)		23.9±3.2		54.9±28.3
$k_{B,C8}$ (s⁻¹)	2870±200	2850±290	8600±930	9010±1360
$k_{B,C1'}$ (s⁻¹)		2900±270		7310±1110
$R_{1,C8}$ (Hz)	3.48±0.13	3.47±0.14	4.40±0.12	4.39±0.11
$R_{2,C8}$ (Hz)	18.4±0.3	18.5±0.4	17.3±0.5	17.1±0.7
$R_{1,C1'}$ (Hz)	2.39±0.14	2.39±0.14	2.87±0.13	2.87±0.13
$R_{2,C1'}$ (Hz)	13.0±0.4	13.0±0.4	10.6±0.5	11.2±0.6
$\Delta\omega_{AB,C8}$ (ppm)	2.77±0.15	2.82±0.22	2.77±0.15	1.80±0.83
$\Delta\omega_{AB,C1'}$ (ppm)	3.00±0.16	3.02±0.21	3.00±0.16	2.79±0.70

*Global fits were performed by sharing k_{ex} and p_B for each nucleotide and assuming a temperature invariant $\Delta\omega_{AB}$ for each site.

Table A1.4: Comparison of NMR-derived carbon chemical shift differences ($\Delta\omega_{AB}$) between the ground and excited state and chemical shifts in corresponding modified DNA constructs or drug-bound DNA that show formation of either HG or WC ground-state base pairs (experimental data is for pH 6.8 unless otherwise indicated).

Residue/ DNA	$\Delta\omega_{AB}$ (ppm)											
	C6/C8						C1'					
	NMR	m1A16 (HG)[a]	m1G10 (HG)[a,d]	MD/DFT (HG)[c]	LNA (WC)[a]	Drug (HG)[b]	NMR	m1A16 (HG)[a]	m1G10 (HG)[a,d]	MD/DFT (HG)[c]	LNA (WC)[a]	Drug (HG)[b]
A16/A$_6$-DNA	2.24±0.09 2.66±0.02[e]	5.54	0.55	2.58 4.29±0.98	-0.63	-	3.16±0.11 3.27±0.02[e]	3.57	0.15	3.32 4.65±1.69	4.40	-
A16/A$_4$-DNA	-	-	-	-	-	-	3.32±0.06 3.17±0.02[e]	-	-	-	-	-
A16/A$_2$-DNA	2.73±0.11	-	-	-	-1.03	-	-	-	-	-	4.54	-
A3/A$_2$-DNA	2.97±0.03	-	-	-	-	-	3.14±0.04	-	-	-	-	-
T9/A$_6$-DNA	n/d	0.37	-0.47	-	-0.05	-	n/d	0.07	0.07	-	-0.16	-
T9/A$_2$-DNA	n/d	-	-	-	0.88	-	n/d	-	-	-	-0.10	-
T22/A$_2$-DNA	n/d	-	-	-	-	-	n/d	-	-	-	-	-
G10/A$_6$-DNA	3.09±0.02[d] 3.16±0.01[e]	0.26	3.10	3.01 4.01±1.99 3.74±1.97	-0.42	-	3.70±0.01[e]	0.17	3.70	4.73 5.17±1.44 5.14±1.40	-0.06	-
G10/A$_4$-DNA	-	-	-	-	-	-	3.80±0.07[d]	-	-	-	-	-
G10/A$_2$-DNA	-	-	-	-	-0.24	-	-	-	-	-	-0.24	-
C15/A$_6$-DNA	2.19±0.06[d] 2.22±0.03[e]	0.18	2.30	2.63 5.62±1.99 -0.63±1.73	-0.02	-	n/d	-0.26	0.13	0.61 0.08±0.96 -0.52±0.95	-0.06	-
C15/A$_4$-DNA	2.30±0.07[d]	-	-	-	-	-	-	-	-	-	-	-
C15/A$_2$-DNA	-	-	-	-	-0.17	-	n/d	-	-	-	0.11	-
A5/E-DNA	2.77±0.15	-	-	-	-	2.55	3.00±0.16	-	-	-	-	2.14
T4/E-DNA	-	-	-	-	-	1.07	-	-	-	-	-	0.52

'-' data not measured; 'n/d' chemical exchange not detectable. Italicized data refers to perturbation observed at neighboring sites to the modified one.

[a] NMR chemical shift differences between unmodified and modified A_6-DNA or A_2-DNA

[b] NMR chemical shift differences between WC and HG base pairs at a TA step in free and echinomycin-bound E-DNA

[c] DFT predictions for chemical shift differences between for WC and HG base showing best matching values to the excited state $\Delta\omega_{AB}$ followed by the average chemical shift difference ± one s.d. Calculations for G10•C15 base pair in A_6-DNA reflect a protonated C15 N3 except for the bottom values.

[d] Data reported for measurements at pH 5.2

[e] Data reported for measurements at pH 5.4

Table A1.5: Chemical exchange parameters from best individual or global fits to to Eq.3.1 of pH dependent off-resonance $R_{1\rho}$ relaxation dispersion data for G10 C8/C1' and A16 C8 in A_6-DNA, collected at 14.1 T and 26 ^{o}C.

Parameter	pH 4.0	pH 4.3	pH 4.6	pH 4.9	pH 5.2	pH 5.4	pH 6.0	pH 6.8	pH 7.6
G10 - Individual fits									
p_B (%)	-	8.38±0.13	3.71±0.03	2.05±0.03	1.08±0.02	0.68±0.02	0.18±0.01	0.05±0.03	n/d
k_{ex} (s^{-1})	-	1795±29	1010±14	804±14	682±22	603±24	827±81	610±450	n/d
$R_{1,C8}$ (Hz)	-		2.35±0.04	2.47±0.04	2.32±0.02	2.44±0.01	2.36±0.01	2.29±0.02	n/d
$R_{2,C8}$ (Hz)	-		22.2±0.1	21.3±0.1	22.2±0.1	23.1±0.1	22.3±0.1	22.2±0.1	n/d
$R_{1,C1'}$ (Hz)	-	1.76±0.43	-	-	-		-	-	n/d
$R_{2,C1'}$ (Hz)	-	18.1±1.6	-	-	-		-	-	n/d
$\Delta\omega_{AB,C8}$ (ppm)	-		3.22±0.01	3.10±0.02	3.09±0.02	3.11±0.02	2.95±0.06	3.19±0.33	n/d
$\Delta\omega_{AB,C1'}$ (ppm)	-	3.94±0.06	-	-	-		-	-	n/d
G10 - Global fit (shared $\Delta\omega_{AB,C8}$)									
p_B (%)	-	8.38±0.07	3.77±0.04	2.05±0.04	1.08±0.03	0.68±0.03	0.17±0.01	0.05±0.04	n/d
k_{ex} (s^{-1})	-	1795±29	1010±19	804±24	675±30	606±33	885±119	604±584	n/d
$R_{1,C8}$ (Hz)	-		2.39±0.05	2.46±0.04	2.32±0.02	2.44±0.01	2.36±0.02	2.29±0.02	n/d
$R_{2,C8}$ (Hz)	-		22.3±0.1	21.2±0.1	22.2±0.1	23.0±0.1	22.3±0.0	22.2±0.1	n/d
$R_{1,C1'}$ (Hz)	-	1.63±0.14	-	-	-		-	-	n/d
$R_{2,C1'}$ (Hz)	-	17.2±0.5	-	-	-		-	-	n/d
$\Delta\omega_{AB,C8}$ (ppm)	-				3.14±0.01				n/d
$\Delta\omega_{AB,C1'}$ (ppm)	-	3.78±0.02	-	-	-	-	-	-	n/d
A16 - Individual fits									
p_B (%)	4.13±0.15	1.79±0.04	1.21±0.04	0.86±0.04	-	0.35±0.02	0.30±0.04	0.39±0.08	0.37±0.07
k_{ex} (s^{-1})	4700±110	3770±100	3920±110	3560±170	-	3180±20	3570±360	3290±400	3310±360
$R_{1,C8}$ (Hz)	2.96±0.08	2.43±0.08	2.33±0.05	2.37±0.05	-	2.33±0.02	2.21±0.04	2.41±0.04	2.39±0.04
$R_{2,C8}$ (Hz)	28.8±0.3	25.5±0.4	22.2±0.1	23.6±0.2	-	23.8±0.1	23.0±0.2	23.0±0.2	23.1±0.2
$\Delta\omega_{AB,C8}$ (ppm)	3.12±0.07	3.89±0.06	3.32±0.07	3.02±0.12	-	3.08±0.10	2.97±0.23	2.21±0.23	2.10±0.24
A16 - Global fit (shared $\Delta\omega_{AB,C8}$)									
p_B (%)	3.68±0.08	2.12±0.05	1.18±0.03	0.74±0.02	-	0.30±0.01	0.25±0.02	0.21±0.02	0.19±0.02
k_{ex} (s^{-1})	4870±110	3600±100	3950±120	3710±180	-	3190±190	3720±410	3790±550	3800±500
$R_{1,C8}$ (Hz)	2.96±0.09	2.59±0.08	2.32±0.05	2.35±0.05	-	2.32±0.02	2.19±0.04	2.38±0.04	2.36±0.04
$R_{2,C8}$ (Hz)	28.1±0.4	26.2±0.4	23.5±0.3	23.4±0.3	-	23.8±0.1	23.0±0.2	22.8±0.2	23.0±0.3
$\Delta\omega_{AB,C8}$ (ppm)					3.39±0.04				