Published in final edited form as:

J Biomol NMR. 2018 April ; 70(4): 229–244. doi:10.1007/s10858-018-0177-2.

Atomic Structures of Excited State A-T Hoogsteen Base Pairs in Duplex DNA by Combining NMR Relaxation Dispersion, Mutagenesis, and Chemical Shift Calculations

Honglue Shi[1], **Mary C. Clay**[2], **Atul Rangadurai**[2], **Bharathwaj Sathyamoorthy**[1,2], **David A. Case**[3,*], and **Hashim M. Al-Hashimi**[1,2,*]

[1]Department of Chemistry, Duke University, Durham, NC 27710, USA

[2]Department of Biochemistry, Duke University School of Medicine, Durham, NC 27710, USA

[3]Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854, USA

Abstract

NMR relaxation dispersion studies indicate that in canonical duplex DNA, Watson-Crick base pairs (bps) exist in dynamic equilibrium with short-lived low abundance excited state Hoogsteen bps. *N1*-methylated adenine (m^1A) and guanine (m^1G) are naturally occurring forms of damage that stabilize Hoogsteen bps in duplex DNA. NMR dynamic ensembles of DNA duplexes with m^1A-T Hoogsteen bps reveal significant changes in sugar pucker and backbone angles in and around the Hoogsteen bp, as well as kinking of the duplex towards the major groove. Whether these structural changes also occur upon forming excited state Hoogsteen bps in unmodified duplexes remains to be established because prior relaxation dispersion probes provided limited information regarding the sugar-backbone conformation. Here, we demonstrate measurements of C3′ and C4′ spin relaxation in the rotating frame ($R1\rho$) in uniformly $^{13}C/^{15}N$ labeled DNA as sensitive probes of the sugar-backbone conformation in DNA excited states. The chemical shifts, combined with structure-based predictions using an Automated Fragmentation Quantum Mechanics/Molecular Mechanics (AFQM/MM) method, show that the dynamic ensemble of DNA duplexes containing m^1A-T Hoogsteen bps accurately model the excited state Hoogsteen conformation in two different sequence contexts. Formation of excited state A-T Hoogsteen bps is accompanied by changes in sugar-backbone conformation that allow the flipped *syn* adenine to form hydrogen-bonds with its partner thymine and this in turn results in overall kinking of the DNA toward the major groove. Results support the assignment of Hoogsteen bps as the excited state observed in canonical duplex DNA, provide an atomic view of DNA dynamics linked to formation of Hoogsteen bps, and lay the groundwork for a potentially general strategy for solving structures of nucleic acid excited states.

*To whom correspondence should be addressed. Telephone: (919) 660-1113, hashim.al.hashimi@duke.edu or david.case@rutgers.edu. Present address: Bharathwaj Sathyamoorthy, Department of Chemistry, Indian Institute of Science Education and Research Bhopal, Bhopal 462066, India

Keywords

DNA dynamics; rotating frame spin relaxation; nucleic acids; sugar pucker; major groove kinking; DNA bending

Introduction

Techniques based on spin relaxation in the rotating frame ($R1\rho$)[1–3] and chemical exchange saturation transfer (CEST)[4–6] are providing rare insights into short-lived (μs-ms) low-populated (<5%) excited states (ESs)[7] in nucleic acids that play important roles in gene expression and regulation[8,9]. For example, canonical Watson-Crick A–T and G–C base pairs (bps) in duplex DNA have been shown to exist in dynamic equilibrium with non-canonical Hoogsteen bps[10] (Figure 1A). The Hoogsteen bps are proposed to play important roles in DNA recognition[11], damage repair[12], and replication[9,13]. RD studies have also uncovered an inverse process, whereby non-canonical G–T/U wobble mismatches morph into canonical Watson-Crick-like tautomeric and anionic mismatches[14]. These Watson-Crick-like mismatches can evade fidelity checkpoints and give rise to replication and translational errors[15]. ESs have also been reported in a wide variety of RNAs where they are proposed to play roles as regulatory switches[16,17], folding intermediates[8], and in the gene regulation by riboswitches[18–20].

Determining the atomic resolution structure and dynamic properties of nucleic acid ESs is key for elucidating their functional roles as well as for potentially advancing new therapeutic approaches targeting these alternative conformational states. This however presents a considerable challenge to conventional structure determination methods, given that ES are short-lived and exist in low abundance. A variety of approaches have been developed to address this problem in proteins[21,22]. The ES chemical shifts have been used to guide determination of conformational ensembles of protein folding intermediates. Methods have also been introduced to directly obtain structural constraints on the ES including orientational constraints from measurements of residual dipolar couplings (RDCs)[23–25], residual chemical shift anisotropies (RCSAs)[26,27], and distance-based constraints from paramagnetic relaxation enhancements (PRE)[28,29]. Methods have also been developed to gain insights into the dynamic properties of protein ESs that rely on the analysis of chemical shifts[30] and H/D exchange[31]. Recently, CEST-based methods have also been introduced to measure RDCs in RNA ESs paving the way for atomic resolution structure characterization[32].

An alternative strategy relies on mutations or chemical modifications to stabilize a given ES so that its structural and dynamic properties can be characterized by conventional methods[8,16,30,33,34]. This has proven to be an attractive approach for nucleic acids because their ESs often feature non-canonical mismatches that can be stabilized by point substitution mutations or single atom substitutions[10,16,35]. The mutants also provide an opportunity to explore the function of the ESs *in vitro* and *in vivo*[36,37]. The approach does however suffer from one major drawback; the mutants may not perfectly recapitulate the ES. This is

particularly problematic when the RD data is not sufficient to fully characterize all aspects of the structure around a given modification.

Mutations have successfully been used to stabilize ES Hoogsteen bps in duplex DNA. In particular, we previously showed that *N1*-methylated adenine (m^1A) and *N1*-methylated guanine (m^1G), which are naturally occurring forms of DNA damage that are repaired by damage repair enzymes[38–40], stabilize A-T and G-C^+ Hoogsteen bps respectively in a variety of DNA duplex contexts[10,12,41–44] (Figure 1B). The methyl group eliminates one Watson-Crick hydrogen-bond as well as sterically collides with the partner pyrimidine effectively destabilizing the Watson-Crick bp without significantly impacting the Hoogsteen conformation (Figure 1B). The chemical shifts of the m^1A^+-T and m^1G-C^+ Hoogsteen bps recapitulate those measured for the ES using NMR R1ρ in unmodified duplexes, indicating that they are valid models for the ES Hoogsteen bps[10,41,45,46].

Recently, we reported high-resolution NMR structures and dynamic ensembles for two duplexes (A_2-DNA and A_6-DNA, Figure 1C) containing m^1A^+-T Hoogsteen bps using RDCs, NOEs, and molecular dynamics (MD) simulations[42]. In these structural ensembles, m^1A forms the expected hydrogen bonded m^1A^+-T Hoogsteen bp. However, the m^1A^+-T Hoogsteen bp also induced significant perturbations to the sugar-backbone conformation, including kinking of the DNA toward the major groove (~ 9°), a bias toward the BI backbone conformation, and sugar repuckering toward O4′-*endo* in and around the Hoogsteeen bp[42]. Whether similar changes in sugar-backbone conformation take place when forming ES Hoogsteen bps in unmodified DNA duplexes remains to be established. RD studies[35] have so far relied on base (C2, C5, C6, C8, N1/3) and sugar (C1′) probes that provide limited information regarding sugar-backbone conformation. The C1′ chemical shift is insensitive to phosphodiester torsion angles, and while it is sensitive to sugar pucker, this can be obscured by an even greater sensitivity to the glycosidic X angle, which can vary significantly when forming non-canonical bps[47]. The *N1*-methyl group introduces a positive charge on the adenine base that can change its stacking properties and also lead to steric contacts that don't occur in the unmodified duplex[12,42].

Here, we address this 'blind spot' in RD studies by demonstrating measurements of sugar C3′ and C4′ RD in uniformly $^{13}C/^{15}N$ labeled DNA samples as new probes of sugar-backbone conformation in DNA ESs. By further combining these RD measurements with structure-based predictions of chemical shifts, we evaluated how well the structure and dynamic ensembles of the m^1A Hoogsteen stabilizing mutants recapitulate the corresponding ES Hoogsteen bp in the unmodified duplex. Results support the assignment of Hoogsteen bps to ESs observed in canonical duplex DNA, provide an atomic view of sugar-backbone DNA dynamics linked to formation of Hoogsteen bps, as well as lays the groundwork for a potentially general strategy for solving structures of nucleic acid ESs.

Materials and methods

Survey of DNA sugar chemical shifts

To gain insights into the deoxyribose ^{13}C chemical shifts, we analyzed all nine entries deposited in the Biological Magnetic Resonance Data Bank (BMRB)[48] (Supplementary

Table 1). The ^{13}C chemical shifts for entry BMRB ID: 17409 was through addition of 2.7 p.p.m. to account for differences in referencing (TMS versus DSS)[49,50]. We then analyzed the sugar carbon chemical shift distribution for 'Helical' nucleotides defined as unmodified Watson-Crick bps that are surrounded by Watson-Crick bps (Supplementary Table 1).

Sample preparation

NMR buffer—All DNA and RNA samples were buffered exchanged with centrifugal concentrators (Amicon Untra-15 3-kDa cutoff EMD Milipore) into a NMR buffer containing 25 mM sodium chloride, 15 mM sodium phosphate, 0.1 mM ethylenediaminetetraacetic acid (EDTA) at pH 5.4 or pH 6.8. Samples were then flash frozen, lyophilized overnight and resuspendend into 100% D_2O. The majority of the NMR experiments were conducted in 100% D_2O (EMD Milipore).

Unmodified DNA samples—Unmodified A_2-DNA and A_6-DNA oligonucleotide samples were purchased from Integrated DNA Technologies (IDT) with standard desalting purification. The DNA samples were annealed at equimolar ratio with unlabeled complementary strands purchased from IDT and exchanged into NMR buffer to a final concentration of 1–2 mM.

m^1A- and m^1G- containing DNA samples—m^1A and m^1G modified DNA oligonucleotide samples were purchased from Midland DNA Technologies. The m^1A DNA samples were synthesized following ultra-mild deprotection to minimize Dimoroth rearrangement[51]. The m^1A DNA samples were then purified by reverse-phase HPLC while the m^1G DNA samples were purified by gel filtration. The m^1A DNA samples were heated at 65 °C to avoid Dimoroth rearrangement, then annealed at equimolar ratio with its unmodified complementary strand from IDT and exchanged into NMR buffer to a final concentration of 1–2 mM.

Unmodified and dA-, dG- containing RNA samples—RNA single strands with or without incorporated dNTPs were synthesized using an in-house MerMade 6 Oligo Synthesizer on 1 μmol standard synthesis columns (1,000 Å) from BioAutomation and using 2′-TBDMS acetyl protected RNA phosphoramidites and benzoyl/isobutyl protected DNA adenine/guanine phosphoramidites. The final 4,4′-dimethoxytrityl (DMT) group was used in cartridge purification. The oligonucleotides were cleaved from each 1-μmol column with ~1 mL ammonia methylamine (1:1 ratio of 30% ammonium hydroxide and 30% methyl amine) and then incubated for 2 hrs at room temperature to allow base deprotection. The solution was then placed under airflow until evaporation was complete, leaving the desired product oligonucleotide as dried crystals. The crystals were dissolved in 115 μL DMSO, mixed with 60 μL Triethylamine (TEA) and 75 μL TEA-3HF, and incubated at 65 °C for 2.5 hrs for 2′-deprotection. The reaction was quenched with Glen-Pak RNA quenching buffer and loaded onto Glen-Pak RNA cartridges (Glen Research Corporation) for purification according to the standard protocol (http://www.glenresearch.com/Technical/GlenPak_UserGuide.pdf). The single strands were then ethanol-precipitated, annealed at equimolar ratio with complementary strands, and exchanged into NMR buffer to a final concentration of 1–2 mM.

$^{13}C,^{15}N$-labeled labeled DNA samples—Uniformly or residue-type (C, A or G, T) ^{13}C, ^{15}N labeled samples were prepared using the Zimmer and Crothers method[52] employing a chemically synthesized DNA template (IDT) and Klenow fragment DNA polymerase (New England Biolab). To improve spectral resolution (Supplementary Figure 2A), residue-type labeled (C, A or G, T) strands were prepared for A_2-DNA using the appropriately labeled $^{13}C,^{15}N$-dNTPs (Sigma-Aldrich) whereas uniform labeling was used for A_6-DNA (Supplementary Figure 2B). The oligonucleotides were purified by 20% 29:1 polyacrylamide denaturing gel with 8 M urea, 20 mM Tris borate and 1 mM EDTA followed by electro-elution (Whatmann, GE Healthcare) in 40 mM Tris Acetate and 1 mM EDTA. The oligonucleotides were then ethanol precipitated, annealed at equimolar ratio with unlabeled complementary strands (IDT) to generate two labeled samples for each duplex. The samples were exchanged into NMR buffer to a final concentration of 1–2 mM.

NMR experiments

Resonance assignment—NMR experiments were carried out at 25 °C on a Bruker Avance III 600-MHz NMR spectrometer or Bruker Avance III 700-MHz spectrometer equipped with a 5mm triple-resonance cryogenic probe. The assignments for C1′-H1′ and C4′-H4′ of A_2- and A_6-DNA (± m^1A) were previously reported[10,42]. Complete sugar assignments were obtained using 2D [^{13}C, ^1H] Heteronuclear Single Quantum Coherence (HSQC), 2D [^1H, ^1H] Total Correlation Spectroscopy (TOCSY), 2D [^1H, ^1H] WATERGATE Nuclear Overhauser Effect Spectroscopy (NOESY) experiments, 2D [^{13}C, ^1H] Constant Time (CT) HSQC and 3D HCCH-TOCSY experiments.

Off-resonance R1ρ relaxation dispersion—^{13}C RD experiments were carried out on Bruker Avance III 700-MHz spectrometer equipped with a 5mm triple-resonance cryogenic cryogenic probe at 25 °C as described previously[47]. Experiments were carried out using selective Hartmann-Hahn polarization transfers to excite specific spin of interest and collect data in a 1D manner[1,53]. The lowest spin-lock power attainable is predominately limited by approximately 3x the largest homonuclear scalar coupling which corresponds to 3 x $^1J_{C4'C5'}$ ~ 130 Hz[53]. Unlike in RNA[47], the deoxyribose carbon chemical shifts are all separated by more than 17 p.p.m relative to their directly bonded carbon atoms, thus minimizing the possibility of unwanted Hartman-Hahn transfers. Measurements across a broad range of spin locks and offsets showed no evidence for deviations from mono-exponential behavior. On- and off-resonance C3′ and C4′ R1ρ RD profiles as well as for C1′ and base C8 and C6 were recorded using the spin lock offsets and power levels listed in Supplementary Table 2 for each spin of interest using six to nine delay times.

Analysis of R1ρ data—Fitted peak intensities determined using NMRpipe[54] at each delay time were fitted to a mono-exponential decay to obtain the R1ρ values and the uncertainty in fitted parameters estimated using Monte Carlo simulations as described previously[14]. Bloch-McConnell (BM) numerical simulations were used to fit R1ρ values to two-state or three-state exchange (for T9 C3′ and C15 C3′) with star-like topology[14].

Automated fragmentation quantum mechanics/molecular mechanics calculations

Chemical shift calculations followed the procedure described earlier[55]. We used an ensemble of six structures for A_2-DNA, ten structures for A_6-DNA, ten structures of A_2-DNA^{m1A16}, and fifteen structures of A_6-DNA^{m1A1642}. For each of the modified structures, we also computed shifts where the methyl group was replaced with a hydrogen. For each of these 66 duplex structures, 24 quantum fragments were constructed (one for each nucleotide), containing the "central" and three to five neighboring nucleotides, as described in previous study[55]. The effects of atoms outside the quantum region, and of water and ions in the solvent, were represented by point charges uniformly distributed on the molecular surface of the quantum region, fit to Poisson-Boltzmann calculations using the "solinprot" module of MEAD[56]. The quantum region was assigned a local dielectric of ε=1 (vacuum), the remaining nucleic acid region had ε=4, and the solvent region ε=80. For each quantum fragment, GIAO chemical shift calculations were carried out using version 4.3.6 of the demon-2k program[57], using the OLYP functional[58] with the TZVP basis set and the GEN-A2* fitting set. Further details are available elsewhere[55,59].

Hoogsteen accommodation analysis

Analysis of Hoogsteen bp translation—Each DNA structure from both modified and unmodified A_2- and A_6-DNA ensembles was superimposed onto an idealized B-form DNA helix constructed using 3DNA[60], using all the heavy atoms of the two bps neighboring the Hoogsteen bp (G10-C15, T8-A17). For a given Watson-Crick bp and its Hoogsteen counterpart, the displacement of the purine base when transitioning from Watson-Crick to Hoogsteen was defined as the projection of the vector joining the C1′ atoms of the purine of the Hoogsteen and Watson-Crick bp onto the C1′-C1′ vector of the Watson-Crick bp. The above calculation was repeated for all possible combinations of Hoogsteen and Watson-Crick structures, and the obtained distances were then averaged. An analogous procedure was performed to compute the displacement of the pyrimidine.

Analysis of C15 sugar and backbone torsion angle—An idealized B-form helix model of the 5′-CA-3′ dinucleotide step was generated using 3DNA[60]. The torsion angles (α, β, γ, δ, ε, ζ, and X) and sugar pucker of both residues were adjusted to match the corresponding average values in the A16-C15 step in the dynamic ensemble of A_6-DNA (with size N=2400) reported previously[42]. First, the torsion angles and sugar pucker of the 3′-dA were changed to match the average values obtained for A16 in the dynamic ensemble of A_6-DNA^{m1A16} (N=2400)[42]. Second, the backbone (β, γ, ε, ζ, X) and sugar (δ and sugar pucker) torsion angles of C15 were adjusted in steps to match the average values obtained for C15 in the dynamic ensemble of A_6-DNA^{m1A16} (N=2400)[42]. This transition was accompanied by an increase in surface area of overlap/stacking between the bases of C15 and A16 as quantified using the ‘analyze utility’ in the 3DNA suite[61].

Results

Deoxyribose C3′ and C4′ chemical shift as probes of DNA sugar-backbone conformation

We recently reported measurements of C4′ R1ρ in uniformly $^{13}C/^{15}N$ labeled RNA as new probes of sugar-backbone conformation in RNA ESs[47]. A similar approach can in principle

be applied to uniformly $^{13}C/^{15}N$ labeled DNA and extended to include both C3′ and C4′ given the improved C3′ spectral dispersion in DNA. In contrast to C1′, the C3′ and C4′ chemical shifts are less sensitive to glycosidic angle (X) in both DNA[62,63] and RNA[62–64]. Rather, prior DFT[63,65,66], solution-state[66], and solid-state NMR[66,67] studies have shown that the deoxyribose C3′ chemical shift primarily depends on sugar pucker (upfield shift ~7 p.p.m. from C2′-*endo* to C3′-*endo*), and also on the phosphodiester torsion angle γ (upfield shift ~2–3 p.p.m. from *gauche* to *trans*). The deoxyribose C4′ chemical shift is strongly dependent on sugar pucker (upfield shift ~1–4 p.p.m. from C2′-*endo* to C3′-*endo*) but less so on the backbone angle γ.

We recently verified the sugar chemical shift structure relationships in the RNA ribose moiety using the database of available chemical shifts[47]. Unfortunately, a similar approach cannot be applied to DNA given the much smaller number of entries in the Biological Magnetic Resonance Data Bank (BMRB)[48] all of which fall in the canonical C2′-*endo* conformation (Supplementary Figure 1 and Supplementary Table 1). We therefore verified computational predictions and prior solid-state NMR studies by incorporating dA or dG into an A_6-RNA duplex (Figure 2A). This allowed us to trap the deoxyribose sugar in the C3′-*endo* rather than C2′-*endo* conformation, as confirmed through measurements of the sums of the coupling constants for the H1′ and H2″ (Figure 2D)[42,68]. To our knowledge, this is the first solution state measurement of the sugar chemical shifts for a DNA in the C3′-*endo* conformation. Indeed, we find that the C3′ and C4′ chemical shifts are upfield shifted by ~5.6 p.p.m. and ~2 p.p.m., respectively relative to unmodified duplex DNA, in very good agreement with theoretical predictions[63,65,66] and solid-state NMR studies on A-form DNA[67] and nucleotide analogues[66,67] containing C3′-*endo* sugars (~ 7 p.p.m. and 1–4 p.p.m. respectively) (Figure 2B and 2C).

Using C3′ and C4′ chemical shifts to probe the sugar-backbone conformation of m^1A^+-T and m^1G-C^+ Hoogsteen bps

We previously reported[42] NMR ensembles for two DNA duplexes containing the Hoogsteen stabilizing m^1A modification (A_2-DNA^{m1A16} and A_6-DNA^{m1A16}). Comparison of these ensembles to ensembles determined for their unmodified counterparts (A_2-DNA and A_6-DNA) revealed that the m^1A^+-T Hoogsteen bps induce significant changes in the sugar backbone conformation[42]. These include changes in the sugar pucker at m^1A16 (from C2′-*endo* to O4′-*endo*), C15 (more C2′-*endo*), and T9 (more C3′-*endo*); a bias toward the BI backbone conformation at C15pm^{1}A16; and kinking of the DNA toward the major groove[42]. We hypothesized that such changes in sugar-backbone conformation should give rise to significant changes in C3′ and C4′ chemical shifts in the DNA duplexes containing the Hoogsteen stabilizing m^1A modification as compared to their unmodified counterparts. These chemical shift probes could in turn provide RD probes for examining the sugar-backbone in the ES Hoogsteen.

Comparison of the spectra obtained for modified (Hoogsteen) versus unmodified (Watson-Crick) duplexes revealed significant (> 0.5 p.p.m) C3′ and C4′ chemical shift perturbations in both A_2-DNA and A_6-DNA (Figure 3A and 3B and Supplementary Figure 3A and 3B). These include upfield shifts in C3′, (2.1–3.5 p.p.m.) and C4′ (1.2–1.5 p.p.m.) at the *syn*

purine nucleotide as well as for the C3′ (1.2–1.7 p.p.m.) of the pyrimidine base partner (Figure 3C and 3D). These perturbations are consistent with a change in sugar pucker from C2′-*endo* towards C3′-*endo*, as observed in the NMR ensembles of A_2-DNA^{m1A16} and A_6-DNA^{m1A16} [42]. Here, A16 and T9 adopt a predominantly O4′-*endo* sugar pucker compared to C2′-*endo* and C1′-*exo* respectively, in the unmodified duplexes. The C3′ and C4′ chemical shifts of both m^1A16 and m^1G10 and the C3′ chemical shifts of their complementary T9 and C15 fall out of the range observed for canonical Watson-Crick bps (Supplementary Figure 1) based on analysis of nine entries in the BMRB[48], indicating a significant change in sugar-backbone conformation relative to the canonical C2′-*endo* conformation.

Interestingly, significant chemical shift perturbations were also observed at nucleotides neighboring the Hoogsteen bp (Figure 3C and 3D). We will refer to these as “remote” perturbations to emphasize that they involve sites distal to the Hoogsteen bp. In particular, the C3′ chemical shifts of nucleotides 5′ to the *syn* purine are downfield shifted by ~ 2.0–2.5 p.p.m (Figure 3C and 3D) which is consistent with NMR ensembles of A_2-DNA^{m1A16} and A_6-DNA^{m1A16} in which the C15 sugar pucker is shifted toward C1′-*exo*[42]. Additionally, the C3′ and C4′ chemical shifts of these 5′-neighboring residues of both m^1A and m^1G fall near the upper boundary (more downfield shifted) of the canonical Watson-Crick chemical shift region deduced from analysis of BMRB, suggesting a small bias toward the C2′-*endo* conformation (Supplementary Figure 1). Smaller perturbations (< 1.0 p.p.m.) in sugar and base carbon chemical shifts are also observed at other neighboring residues (G10-C3′, C15-C4′ for A_2-DNA^{m1A16} and A_6-DNA^{m1A16} and T9-C4′, G11-C3′, C14-C3′ for A_2- and A_6-DNA^{m1G10}) (Figure 3C and 3D).

Using C3′ and C4′ R1ρ to probe sugar-backbone conformation in excited state Hoogsteen bps

The structural ensembles of modified duplexes combined with the sugar chemical shift analysis presented here strongly suggest that m^1A^+-T and m^1G-C^+ Hoogsteen bps induce significant changes in sugar-backbone conformation. To examine whether similar conformational changes occur when forming ES A-T and G-C^+ Hoogsteen bps in unmodified duplexes, we carried out C3′ and C4′ R1ρ on unmodified A_2-DNA and A_6-DNA. Based on comparisons of chemical shifts for modified (m^1A^+ and m^1G) and unmodified A_2-DNA and A_6-DNA, several active (A16-C3′, A16-C4′, T9-C3′, G10-C3′, G10-C4′ and C15-C3′) and remote (G11-C3′ and C14-C3′) sites are predicted to have sizeable differences in chemical shifts ($\Delta\omega > 0.5$ p.p.m.), which should give rise to detectable RD. Fitting of these RD profiles should in turn yield chemical shifts that match those of the modified duplexes and exchange parameters (population and rates of exchange) that match those measured for ES Hoogsteen bp[10]. Likewise, several sites have small perturbations ($\Delta\omega < 0.5$ p.p.m.) and are predicted to have flat C3′ and/or C4′ profiles. Some sites are predicted to simultaneously sense active and remote RD. For example, the T9-C3’ is predicted to experience active RD due to formation of A16-T9 Hoogsteen bp and remote RD due to formation of a neighboring G10-C^+15 Hoogsteen bp. Since the active and remote chemical shift perturbations are opposite (Figure 3C and 3D), and since Hoogsteen transitions at different sites occur independently[10,46], this site is predicted to experience 3-

state exchange in an RD experiment. While other sites such as A17-C1′, A17-C8, and G10-C3′ are also predicted to experience active and remote RD, the perturbations have similar signs, and the contribution from active RD is predicted to dominate (see Supplementary Discussion). Indeed, all of these predictions were borne out experimentally as detailed below.

We first measured C3′ and C4′ RD on A_6-DNA at pH 6.8 so as to reduce the abundance of any ES G-C^+ Hoogsteen bp below the RD detection limit[10,69]. We observed the predicted active RD at A16-C3′, A16-C4′, and T9-C3′ (Figure 4A). Fitting of the RD data to a two-state model yielded ES populations (p_B) and exchange rates ($k_{ex} = k_1+k_{-1}$) that are in excellent agreement with ES Hoogsteen bp as deduced based on A16-C1′ and A16-C8 RD data[10]. Despite working under conditions that suppress RD contributions due to transient G-C^+ Hoogsteen bps, we observed RD at C15-C3′ and G10-C3′. This was exactly as predicted due to remote RD and transient formation of a neighboring A16-T9 Hoogsteen bp (Figure 4A). A global fit of these RD data to a two-state model yielded $p_B \sim 0.45$ % and $k_{ex} = 2885\ s^{-1}$ in excellent agreement with exchange parameters ($p_B \sim 0.39$ % and $k_{ex} = 3680\ s^{-1}$) obtained from fitting the active A16-C1′ and A16-C8 RD data under similar conditions[10] (Supplementary Figure 4 and Supplementary Table 3). Furthermore, all of the sites that are predicted to have insignificant (< 0.5 p.p.m.) chemical shift perturbations on Hoogsteen bp formation had flat RD profiles (Figure 4A and Supplementary Figure 5).

Next, we measured C3′ and C4′ RD linked to formation of G-C^+ Hoogsteen bps by lowering the pH to 5.4 for A_6-DNA[10,69]. Much more significant RD was observed for G10-C3′ and G10-C4′ that are similar to those observed for A16-C3′ and A16-C4′, consistent with the onset of G10-C^+15 Hoogsteen bp. Strikingly, in accord with predictions, T9-C3′ and C15-C3′ exhibit clear signs of three-state exchange due to additional active and remote RD, respectively. This is evidenced by two rather than one oppositely shifted humps in the RD profile reflecting two distinct ES chemical shifts with opposite Δω signs (Figure 4B). A three-state fit to the RD data assuming a star-like topology[3,14] yielded one set of exchange parameters (p_B and k_{exB}) that are consistent with active formation of transient G10-C^+15 Hoogsteen bp and another set (p_C and k_{exC}) that are consistent with the remote formation of a neighboring transient A16-T9 Hoogsteen bp (Supplementary Figure 4). Once again, all of the sites that were predicted to have insignificant (< 0.5 p.p.m.) chemical shift perturbations on G-C^+ Hoogsteen bp formation had flat RD profiles (Figure 4B).

Finally, we also carried out C3′ and C4′ RD measurements on C, A and G, T residue type labeled A_2-DNA to aid spectral resolution (Supplementary Figure 2A). Similar results and levels of agreement were again observed for G10-C15 and A16-T9 at pH 5.4 (Supplementary Figure 4, 6 and Supplementary Table 4).

The ^{13}C chemical shifts of the ES match those of the m^1A and m^1G mutants

Next, we compared the difference between the ES and GS chemical shifts ($\Delta\omega = \omega_{ES} - \omega_{GS}$) measured by RD with the corresponding difference in chemical shifts obtained when comparing the chemical shifts of the Hoogsteen stabilizing mutants (m^1A16 or m^1G10) and the unmodified duplexes ($\Delta\omega = \omega_{m1A16} - \omega_{A16}$ and $\Delta\omega = \omega_{m1G10} - \omega_{G10}$) (Figure 5). Excellent agreement was obtained for both A_2-DNA and A_6-DNA with RMSD = 0.31–0.81

p.p.m. and R = 0.91–0.98 (Figure 5). Excellent agreement was observed for sites that experience active or remote RD as well as sites that experience both process and that were fit to 3-state exchange. The only exception is A16-C8, which as predicted based on DFT calculations[10], is significantly more downfield shifted in m^1A due to protonation of N1. These results indicate that the Hoogsteen stabilizing mutants accurately and robustly model key conformational signatures of A-T and G-C^+ Hoogsteen ESs in unmodified duplexes in two different sequence contexts.

Validating structures and ensembles of Hoogsteen stabilizing mutants using structure-based chemical shift calculations

Next we asked how well do the previously reported NMR structural ensembles of A_2-DNA^{m1A16} and A_6-DNA^{m1A16} recapitulate key conformational signatures of the ES Hoogsteen bps in the unmodified duplexes by using an AF-QM/MM approach[55] to predict the chemical shifts for these ensembles[42]. Specifically, we computed the Δω values ($\Delta\omega_{calc} = \omega_{m1A16} - \omega_{A16}$) and compared them to values measured using RD ($\Delta\omega_{meas}$). Chemical shifts were predicted for each conformer in the ensemble and values averaged to account for motional averaging. Predicting differences between chemical shifts (Δω) rather than the absolute chemical shifts potentially helps to minimize systematic or referencing errors in the chemical shift predictions.

As shown in Figure 6A, AFQM/MM predicts Δω with RMSD ~ 0.66 p.p.m. and R = 0.88 for the A_6-DNA ensemble. The agreement is slightly worse for A_2-DNA (RMSD ~ 0.87 p.p.m. and R = 0.85) for which there was greater uncertainty in the NMR structure due to a lower number of constraints[42]. This level of agreement is consistent with the agreement seen when comparing RD $\Delta\omega_{meas}$ with values measured by taking the differences between the modified and unmodified duplexes (Figure 5). The differences between calculated and observed values for Δω are much less than those seen for QM/MM estimates of ground-state ^{13}C shifts in DNA and RNA (*ca.* 3.5 ppm, see[55]), presumably due to cancellation of systematic errors in taking the difference between ES and GS shifts (see below). The trends in Figure 6A clearly support the notion that our structural models for the ES states in DNA are fully consistent with the observed chemical shift changes from the ground state.

The agreement deteriorated when using only the lowest energy structure (RMSD = 0.80–1.05 p.p.m., and R = 0.77–0.79) (Figure 6B and Supplementary Figure 7), when computing Δω by taking the difference between chemical shift predicted for a randomly chosen conformer in modified and unmodified ensembles (R = 0.18–0.87; RMSD = 0.78–1.46 p.p.m.) (Figure 6B), or the difference between every conformer in the modified ensemble and an unmodified idealized B-form helix (A_6-DNA^{synA16}: R = 0.66, RMSD = 0.99 p.p.m.; A_2-DNA^{synA16}: R = 0.68, RMSD = 1.09 p.p.m.) (Supplementary Figure 7). These controls reinforce the importance of having an accurate ensemble representation to account for motional averaging when computing chemical shifts in DNA. Poorer agreement was also observed when comparing the measured versus predicted chemical shifts rather than their difference for all four duplexes (RMSD = 3.37–3.81 p.p.m. and R = 0.98–0.99, Supplementary Figure 8). This can be attributed to systematic deviations in slope and intercept for different residue and carbon sites, which largely cancel one another when

calculating the differences (Supplementary Figure 9A). Indeed, it was feasible to apply a linear correction to the predicted chemical shifts by multiplying the predicted value by the slope followed by addition of the intercept for each type of nucleus based on the comparison with experimental values measured for unmodified A_2-DNA and A_6-DNA (Supplementary Figure 9B). Applying this correction leads to far better predictions of the chemical shifts (RMSD = 0.82–0.84 p.p.m. and R = 1.00) (Supplementary Figure 9C and 9D), and may point the way to future semi-empirical models that combine quantum calculations with information from databases of observed shifts[70].

Accommodation of excited state Hoogsteen bps

Our results strongly support the validity of the previously reported dynamic ensembles of A_2-DNA^{m1A16} and A_6-DNA^{m1A16} as mimics of the ES A-T Hoogsteen bps. A closer examination of these structures now provides an atomic view of what happens when Watson-Crick bps morph into Hoogsteen bps as well as helps to rationalize the observed differences in chemical shifts.

Since their discovery[71], it was recognized that relative to Watson-Crick bps, Hoogsteen bps require that the two base partners come into closer proximity by ~2 Å in order to form hydrogen-bonds (Figure 1A). In principle, the purine, pyrimidine, or both bases could translate to form the hydrogen bonds. Analysis of the structure and dynamic ensembles of A_2-DNA^{m1A16} and A_6-DNA^{m1A16} reveals that it is in fact the purine adenine that primarily translates to engage in hydrogen-bonding with its partner thymine (Figure 7A and Supplementary Figure 10A). This explains the change in sugar pucker at the adenine nucleotide from C2′-*endo* to O4′-*endo* as well as the bias in backbone angles ε and ζ toward the BI conformation (Figure 7C). This change in sugar pucker at the adenine-C1′ counteracts the ~6–8 p.p.m downfield shift predicted purely due to changes in the glycosidic X angle when flipping the base[10]. This change in sugar pucker in turn leads to major groove kinking of the duplex at the Hoogsteen bp (Figure 7C and 7D). This helps to rationalize the observed correlation between pyrimidine-purine distance and the magnitude of major groove kinking[72].

In both solution structures and dynamic ensembles of A_2-DNA^{m1A16} and A_6-DNA^{m1A16}, the sugar pucker of C15, which is the 5′-neighbor of the flipping purine base, changes from C1′-*exo* toward C2′-*endo* and the BI conformation is enriched at the step between C15 and A16. This conformational change is supported by large differences in RDCs (19–27 Hz) measured previously for C15 C4′-H4′ in A_2-DNA^{m1A16} and A_6-DNA^{m1A16} as compared to the unmodified DNA[42]. Molecular modeling (see Methods) shows that this change in sugar pucker helps to restore stacking between C15 and A16 which is otherwise lost as A16 flips and translates to form Hoogsteen H-bonds (Figure 7B). Therefore, the changes in sugar pucker and backbone angles in C15 most likely serve to maintain stacking interactions with the flipping and sliding purine involved in the Hoogsteen bp.

Discussion

There are now multiple sources of evidence that support the assignment of Hoogsteen bps as the ES observed in duplex DNA. These include the ^{13}C[10,45 and 15]N chemical shifts[46], the

enhanced populations seen with inosine substitutions[73], diminished populations observed with 7-deazopurine substitutions[46], and the pH dependence of the exchange process[69]. The strong agreement observed here between the Δω values measured for the base and sugar moieties of the transient A-T and G-C^+ Hoogsteen bps as well as their neighbors with values measured in the Hoogsteen stabilizing mutants in two different duplexes, as well as the strong agreement between the measured Δω values and those predicted using AFQM/MM leave little doubt that the ESs measured in duplex DNA correspond to Hoogsteen bps. In this regard, it is also satisfying to note that ES Hoogsteen bps are not observed in RNA duplexes due the energetic instability of the Hoogsteen bp in the A-form helix[41], and accordingly, no sugar RD was recently reported in the sugar moieties of A-form RNA duplexes[47].

Advances in NMR methods for accessing structural information on ESs are making it possible to solve their structure at atomic resolution[22]. Our study demonstrates that a mutational approach can also be valid in the case of nucleic acids. Indeed, we previously showed that point substitution mutations as well as single-atom substitutions can be used to stabilize RNA ESs[8,16,17,74]. In addition, various forms of mutagenesis have successfully been used to stabilize ESs of proteins[30,33,34,75,76]. Here, it is critically important to measure a comprehensive set of RD data that probe various aspects of conformation. With the measurement of sugar C3′ and C4′ RD demonstrated here, it is now feasible to characterize most structural aspects of nucleic acid ESs, providing a more robust means by which to assess ES-stabilizing mutants. This combined with recent developments involving measurements of RDCs in nucleic acid ESs using CEST[32] will make it increasingly possible to access the structures of nucleic acid ESs at atomic resolution.

Our results also highlight the feasibility of using quantum chemical methods for computing chemical shifts in studies of DNA structure and dynamics. However, limitations of the current approach are also seen in our results. In particular, they highlight systematic errors in the predicting chemical shifts for various sites as well as the importance of using a dynamic representation of the nucleic acid in order to account for motional averaging of the chemical shifts. While our approach relied on computing differences between the GS and ES in order to minimize systematic errors in structure-based chemical shift predictions, it was also feasible to apply a linear correction to the predicted chemical shifts (Supplementary Figure 9). Further studies are needed to evaluate the robustness and general validity of the chemical shift corrections.

Based on the structures and dynamic ensembles of A_2-DNA^{m1A16} and A_6-DNA^{m1A16}, a dynamic picture of duplex DNA emerges in which formation of transient Hoogsteen bps leads to major groove kinking of the duplex accompanied by sugar repuckering and changes in backbone torsion angles which allow the purine base to optimally hydrogen-bond with its partner dT as well as form stacking interactions with its neighbors (Figure 7B). The RDC ensembles further indicate that the kinked Hoogsteen conformation is globally more flexible, undergoing larger amplitude kinking motions relative to the Watson-Crick conformation (Figure 7D and Supplementary Figure 10D). An important step for the future will be to develop methods that can also probe these dynamic properties of nucleic acid ESs, as has now been done for proteins[30,76]. This will make it possible to assess to what extent ES-stabilizing mutants accurately capture the dynamic properties of the ESs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank all Al-Hashimi lab members for critical comments on the manuscript. We acknowledge the technical support and resources from the Duke Magnetic Resonance Spectroscopy Center. This work was supported by US National Institute for General Medical Sciences (5P50GM103297 to H.M.A. and D.A.C and R01GM089846 to H.M.A.). The authors declare no conflict of interest.

References

1. Hansen AL, Nikolova EN, Casiano-Negroni A, Al-Hashimi HM. 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. 2009; 131:3818–9. [PubMed: 19243182]
2. Palmer AG 3rd. NMR characterization of the dynamics of biomacromolecules. Chem Rev. 2004; 104:3623–40. [PubMed: 15303831]
3. Palmer AG 3rd, Massi F. Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. Chem Rev. 2006; 106:1700–19. [PubMed: 16683750]
4. Zhao B, Hansen AL, Zhang Q. Characterizing slow chemical exchange in nucleic acids by carbon CEST and low spin-lock field R(1rho) NMR spectroscopy. J Am Chem Soc. 2014; 136:20–3. [PubMed: 24299272]
5. Vallurupalli P, Bouvignies G, Kay LE. Studying "invisible" excited protein states in slow exchange with a major state conformation. J Am Chem Soc. 2012; 134:8148–61. [PubMed: 22554188]
6. Bouvignies G, Kay LE. A 2D C-13-CEST experiment for studying slowly exchanging protein systems using methyl probes: an application to protein folding. J Biomol NMR. 2012; 53:303–310. [PubMed: 22689067]
7. Mulder FA, Mittermaier A, Hon B, Dahlquist FW, Kay LE. Studying excited states of proteins by NMR spectroscopy. Nat Struct Biol. 2001; 8:932–5. [PubMed: 11685237]
8. Xue Y, Gracia B, Herschlag D, Russell R, Al-Hashimi HM. Visualizing the formation of an RNA folding intermediate through a fast highly modular secondary structure switch. Nat Commun. 2016; 7
9. Nikolova EN, et al. A historical account of hoogsteen base-pairs in duplex DNA. Biopolymers. 2013; 99:955–68. [PubMed: 23818176]
10. Nikolova EN, et al. Transient Hoogsteen base pairs in canonical duplex DNA. Nature. 2011; 470:498–502. [PubMed: 21270796]
11. Kitayner M, et al. Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs. Nat Struct Mol Biol. 2010; 17:423–9. [PubMed: 20364130]
12. 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. 2010; 38:4415–25. [PubMed: 20223766]
13. Nair DT, Johnson RE, Prakash L, Prakash S, Aggarwal AK. Human DNA polymerase iota incorporates dCTP opposite template G via a G.C + Hoogsteen base pair. Structure. 2005; 13:1569–77. [PubMed: 16216587]
14. Kimsey IJ, Petzold K, Sathyamoorthy B, Stein ZW, Al-Hashimi HM. Visualizing transient Watson-Crick-like mispairs in DNA and RNA duplexes. Nature. 2015; 519:315–20. [PubMed: 25762137]
15. Topal MD, Fresco JR. Complementary base pairing and the origin of substitution mutations. Nature. 1976; 263:285–9. [PubMed: 958482]
16. Dethoff EA, Petzold K, Chugh J, Casiano-Negroni A, Al-Hashimi HM. Visualizing transient low-populated structures of RNA. Nature. 2012; 491:724–8. [PubMed: 23041928]
17. Lee J, Dethoff EA, Al-Hashimi HM. Invisible RNA state dynamically couples distant motifs. Proc Natl Acad Sci U S A. 2014; 111:9485–90. [PubMed: 24979799]

18. Zhao B, Guffy SL, Williams B, Zhang Q. An excited state underlies gene regulation of a transcriptional riboswitch. Nat Chem Biol. 2017; 13:968–974. [PubMed: 28719589]
19. Chen B, LeBlanc R, Dayie TK. SAM-II Riboswitch Samples at least Two Conformations in Solution in the Absence of Ligand: Implications for Recognition. Angew Chem Int Ed Engl. 2016; 55:2724–7. [PubMed: 26800479]
20. Kladwang W, VanLang CC, Cordero P, Das R. A two-dimensional mutate-and-map strategy for non-coding RNA structure. Nat Chem. 2011; 3:954–62. [PubMed: 22109276]
21. Kay LE. New Views of Functionally Dynamic Proteins by Solution NMR Spectroscopy. J Mol Biol. 2016; 428:323–331. [PubMed: 26707200]
22. Sekhar A, Kay LE. NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. Proc Natl Acad Sci U S A. 2013; 110:12867–74. [PubMed: 23868852]
23. Igumenova TI, Brath U, Akke M, Palmer AG. Characterization of chemical exchange using residual dipolar coupling. J Am Chem Soc. 2007; 129:13396. [PubMed: 17929930]
24. Vallurupalli P, Hansen DF, Stollar E, Meirovitch E, Kay LE. Measurement of bond vector orientations in invisible excited states of proteins. Proc Natl Acad Sci U S A. 2007; 104:18473–7. [PubMed: 18006656]
25. Hansen DF, Vallurupalli P, Kay LE. Quantifying two-bond 1HN-13CO and one-bond 1H(alpha)-13C(alpha) dipolar couplings of invisible protein states by spin-state selective relaxation dispersion NMR spectroscopy. J Am Chem Soc. 2008; 130:8397–405. [PubMed: 18528998]
26. Vallurupalli P, Hansen DF, Kay LE. Probing structure in invisible protein states with anisotropic NMR chemical shifts. J Am Chem Soc. 2008; 130:2734. [PubMed: 18257570]
27. Vallurupalli P, Hansen DF, Kay LE. Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. Proc Natl Acad Sci U S A. 2008; 105:11766–71. [PubMed: 18701719]
28. Sekhar A, Rosenzweig R, Bouvignies G, Kay LE. Hsp70 biases the folding pathways of client proteins. Proc Natl Acad Sci U S A. 2016; 113:E2794–801. [PubMed: 27140645]
29. Clore GM, Tang C, Iwahara J. Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement. Current Opinion in Structural Biology. 2007; 17:603–616. [PubMed: 17913493]
30. Korzhnev DM, Religa TL, Banachewicz W, Fersht AR, Kay LE. A Transient and Low-Populated Protein-Folding Intermediate at Atomic Resolution. Science. 2010; 329:1312–1316. [PubMed: 20829478]
31. Long D, Bouvignies G, Kay LE. Measuring hydrogen exchange rates in invisible protein excited states. Proc Natl Acad Sci U S A. 2014; 111:8820–5. [PubMed: 24889628]
32. Zhao B, Zhang Q. Measuring Residual Dipolar Couplings in Excited Conformational States of Nucleic Acids by CEST NMR Spectroscopy. J Am Chem Soc. 2015; 137:13480–3. [PubMed: 26462068]
33. Korzhnev DM, et al. Nonnative interactions in the FF domain folding pathway from an atomic resolution structure of a sparsely populated intermediate: an NMR relaxation dispersion study. J Am Chem Soc. 2011; 133:10974–82. [PubMed: 21639149]
34. Lorieau JL, Louis JM, Bax A. The complete influenza hemagglutinin fusion domain adopts a tight helical hairpin arrangement at the lipid:water interface. Proc Natl Acad Sci U S A. 2010; 107:11341–11346. [PubMed: 20534508]
35. Xue Y, et al. Characterizing RNA Excited States Using NMR Relaxation Dispersion. Methods Enzymol. 2015; 558:39–73. [PubMed: 26068737]
36. Bhabha G, et al. A Dynamic Knockout Reveals That Conformational Fluctuations Influence the Chemical Step of Enzyme Catalysis. Science. 2011; 332:234–238. [PubMed: 21474759]
37. Whittier SK, Hengge AC, Loria JP. Conformational motions regulate phosphoryl transfer in related protein tyrosine phosphatases. Science. 2013; 341:899–903. [PubMed: 23970698]
38. Yang CG, Garcia K, He C. Damage detection and base flipping in direct DNA alkylation repair. Chembiochem. 2009; 10:417–23. [PubMed: 19145606]

39. Trewick SC, Henshaw TF, Hausinger RP, Lindahl T, Sedgwick B. Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. Nature. 2002; 419:174–8. [PubMed: 12226667]
40. Yang CG, et al. Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. Nature. 2008; 452:961–5. [PubMed: 18432238]
41. Zhou H, et al. m1A and m1G disrupt A-RNA structure through the intrinsic instability of Hoogsteen base pairs. Nat Struct Mol Biol. 2016; 23:803–810. [PubMed: 27478929]
42. Sathyamoorthy B, et al. Insights into Watson-Crick/Hoogsteen breathing dynamics and damage repair from the solution structure and dynamic ensemble of DNA duplexes containing m1A. Nucleic Acids Res. 2017
43. Yang H, Zhan Y, Fenn D, Chi LM, Lam SL. Effect of 1-methyladenine on double-helical DNA structures. FEBS Lett. 2008; 582:1629–33. [PubMed: 18435925]
44. Yang H, Lam SL. Effect of 1-methyladenine on thermodynamic stabilities of double-helical DNA structures. FEBS Lett. 2009; 583:1548–53. [PubMed: 19376116]
45. Alvey HS, Gottardo FL, Nikolova EN, Al-Hashimi HM. Widespread transient Hoogsteen base pairs in canonical duplex DNA with variable energetics. Nat Commun. 2014; 5:4786. [PubMed: 25185517]
46. Nikolova EN, Gottardo FL, Al-Hashimi HM. Probing transient Hoogsteen hydrogen bonds in canonical duplex DNA using NMR relaxation dispersion and single-atom substitution. J Am Chem Soc. 2012; 134:3667–70. [PubMed: 22309937]
47. Clay MC, Ganser LR, Merriman DK, Al-Hashimi HM. Resolving sugar puckers in RNA excited states exposes slow modes of repuckering dynamics. Nucleic Acids Res. 2017
48. Ulrich EL, et al. BioMagResBank. Nucleic Acids Res. 2008; 36:D402–8. [PubMed: 17984079]
49. Aeschbacher T, Schubert M, Allain FH. A procedure to validate and correct the 13C chemical shift calibration of RNA datasets. J Biomol NMR. 2012; 52:179–90. [PubMed: 22252483]
50. Dallmann A, et al. Structure and dynamics of triazole-linked DNA: biocompatibility explained. Chemistry. 2011; 17:14714–7. [PubMed: 22131102]
51. Macon JB, Wolfenden R. 1-Methyladenosine. Dimroth rearrangement and reversible reduction. Biochemistry. 1968; 7:3453–8. [PubMed: 5681457]
52. Zimmer DP, Crothers DM. NMR of enzymatically synthesized uniformly 13C15N-labeled DNA oligonucleotides. Proc Natl Acad Sci U S A. 1995; 92:3091–3095. [PubMed: 7724521]
53. Korzhnev DM, Orekhov VY, Kay LE. Off-resonance R1(p) NMR studies of exchange dynamics in proteins with low spin-lock fields: An application to a fyn SH3 domain. J Am Chem Soc. 2005; 127:713–721. [PubMed: 15643897]
54. Delaglio F, et al. Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. J Biomol NMR. 1995; 6:277–293. [PubMed: 8520220]
55. Swails J, Zhu T, He X, Case DA. AFNMR: automated fragmentation quantum mechanical calculation of NMR chemical shifts for biomolecules. J Biomol NMR. 2015; 63:125–39. [PubMed: 26232926]
56. Richardson WH, Peng C, Bashford D, Noodleman L, Case DA. Incorporating solvation effects into density functional theory: Calculation of absolute acidities. International Journal of Quantum Chemistry. 1997; 61:207–217.
57. Koster AMGG, , Calaminici P, , Casida ME, , Dominguez VD, , Flores-Moreno R, , Gamboa GU, , Goursot A, , Heine T, , Ipatov A, , Janetzko F, , del Campo JM, , Reveles JU, , Vela A, , Zuniga-Gutierrez B, , Salahub DR. deMon2k, Version 3 The deMon developers; Cinvestav, Mexico City: 2011
58. Handy NC, Cohen AJ. Left-right correlation energy. Molecular Physics. 2001; 99:403–412.
59. Case DA. Chemical shifts in biomolecules. Curr Opin Struct Biol. 2013; 23:172–6. [PubMed: 23422068]
60. Lu XJ, Olson WK. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic Acids Res. 2003; 31:5108–5121. [PubMed: 12930962]

61. Lu XJ, Olson WK. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. Nat Protoc. 2008; 3:1213–27. [PubMed: 18600227]
62. Xu XP, Au-Yeung SCF. Investigation of chemical shift and structure relationships in nucleic acids using NMR and density functional theory methods. J Phys Chem B. 2000; 104:5641–5650.
63. Fonville JM, et al. Chemical Shifts in Nucleic Acids Studied by Density Functional Theory Calculations and Comparison with Experiment. Chemistry-a European Journal. 2012; 18:12372–12387.
64. Rossi P, Harbison GS. Calculation of 13C chemical shifts in rna nucleosides: structure-13C chemical shift relationships. J Magn Reson. 2001; 151:1–8. [PubMed: 11444931]
65. Dejaegere AP, Case DA. Density functional study of ribose and deoxyribose chemical shifts. J Phys Chem A. 1998; 102:5280–5289.
66. Xu XP, Chiu WLAK, Au-Yeung SCF. Chemical shift and structure relationship in nucleic acids: Correlation of backbone torsion angles gamma and alpha with C-13 chemical shifts. J Am Chem Soc. 1998; 120:4230–4231.
67. Santos RA, Tang P, Harbison GS. Determination of the DNA Sugar Pucker Using C-13 Nmr-Spectroscopy. Biochemistry. 1989; 28:9372–9378. [PubMed: 2611236]
68. van Wijk J, Huckriede BD, Ippel JH, Altona C. Furanose sugar conformations in DNA from NMR coupling constants. Methods Enzymol. 1992; 211:286–306. [PubMed: 1406311]
69. Nikolova EN, Goh GB, Brooks CL 3rd, Al-Hashimi HM. Characterizing the protonation state of cytosine in transient G.C Hoogsteen base pairs in duplex DNA. J Am Chem Soc. 2013; 135:6766–9. [PubMed: 23506098]
70. Brown JD, Summers MF, Johnson BA. Prediction of hydrogen and carbon chemical shifts from RNA using database mining and support vector regression. J Biomol NMR. 2015; 63:39–52. [PubMed: 26141454]
71. Hoogsteen K. Crystal and Molecular Structure of a Hydrogen-Bonded Complex between 1-Methylthymine and 9-Methyladenine. Acta Crystallographica. 1963; 16:907.
72. Zhou H, et al. New insights into Hoogsteen base pairs in DNA duplexes from a structure-based survey. Nucleic Acids Res. 2015; 43:3420–33. [PubMed: 25813047]
73. Nikolova EN, Stull F, Al-Hashimi HM. Guanine to inosine substitution leads to large increases in the population of a transient G.C Hoogsteen base pair. Biochemistry. 2014; 53:7145–7. [PubMed: 25339065]
74. Merriman DK, et al. Shortening the HIV-1 TAR RNA Bulge by a Single Nucleotide Preserves Motional Modes Over a Broad Range of Timescales. Biochemistry. 2016
75. Barette J, Velyvis A, Religa TL, Korzhnev DM, Kay LE. Cross-validation of the structure of a transiently formed and low populated FF domain folding intermediate determined by relaxation dispersion NMR and CS-Rosetta. J Phys Chem B. 2012; 116:6637–44. [PubMed: 22148426]
76. Bouvignies G, et al. Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. Nature. 2011; 477:111–4. [PubMed: 21857680]
77. Cavanagh J, , Fairbrother WJAGP, , III, Rance M, , Skelton NJ. Protein NMR Spectroscopy 2. Elsevier Inc; 2007

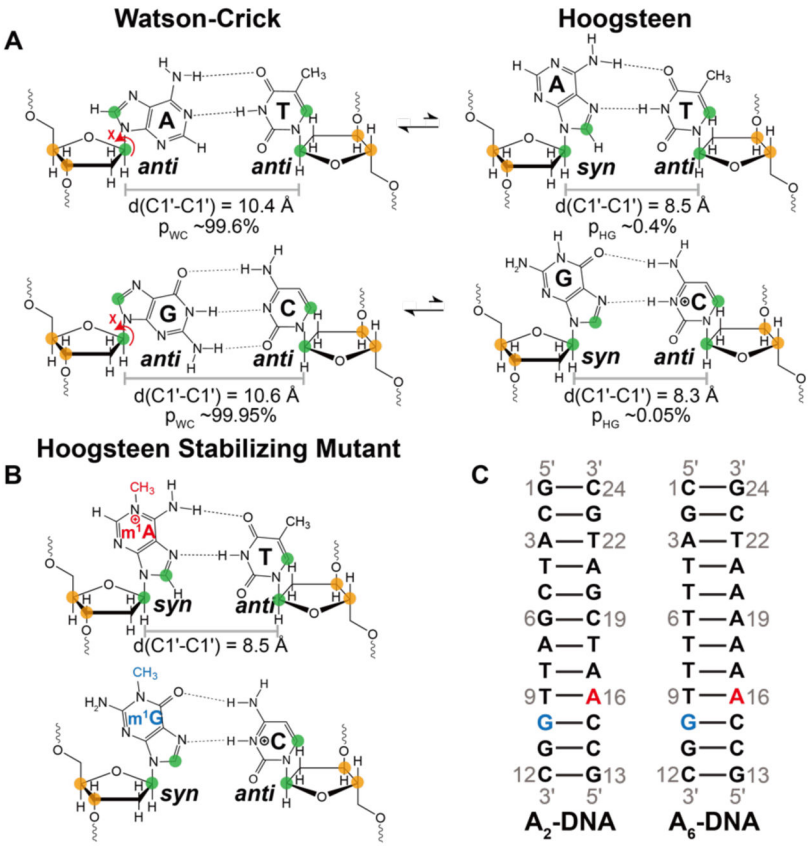

Figure 1. m^1A and m^1G as A•T and G•C^+ Hoogsteen stabilizing mutants

(A) In canonical duplex DNA, Watson-Crick bps exist in dynamic equilibrium with their Hoogsteen counterparts. Note that the population of the G-C^+ Hoogsteen bp is strongly dependent on pH; values shown correspond to pH 6.8[10]. (B) Stabilizing Hoogsteen bps using m^1A and m^1G. (A, B) Nuclei targeted for RD measurements in this work are colored green (C1′ and base) and orange (C3′ and C4′). (C) Duplexes used in this study. A16 or G10 were substituted by m^1A16 (red) or m^1G10 (blue), to stabilize A-T and G-C^+ Hoogsteen bps, respectively.

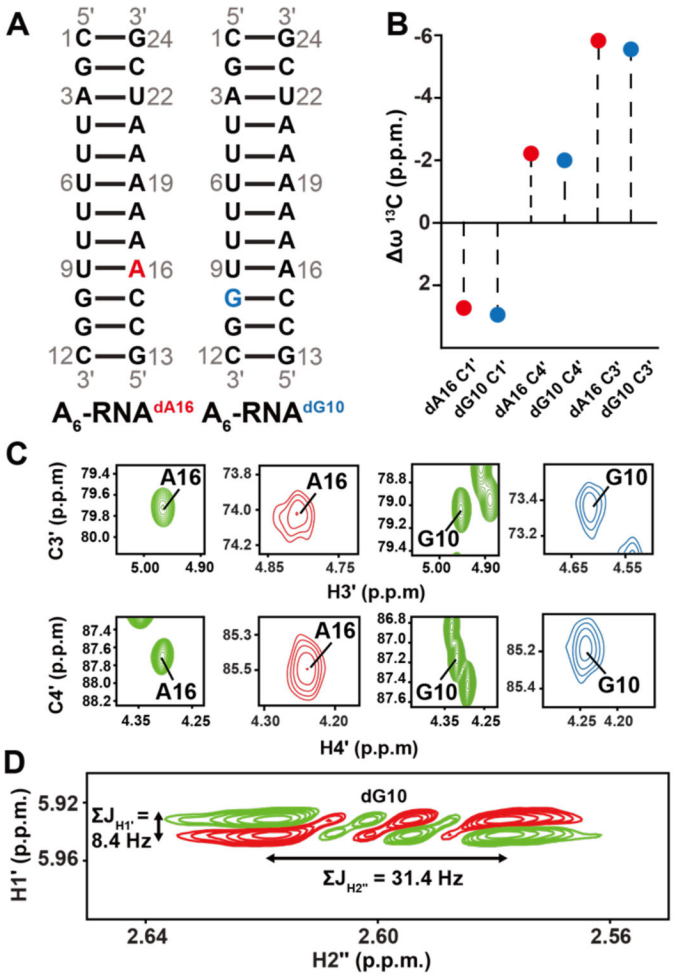

Figure 2.

^{13}C chemical shifts in C3′-*endo* deoxyribose. (A) A_6-RNA duplex highlighting the dA16 and dG10 dexoyribose substitutions. (B) Differences ($\Delta\omega = \omega_{A6RNA/dNTP} - \omega_{A6DNA}$) between sugar ^{13}C chemical shifts at A16 and G10 when comparing A_6-DNA duplex versus dA16 and dG10 substituted A_6-RNA. (C) 2D C3′-H3′ and C4′-H4′ HSQC spectra of the corresponding A_6-DNA duplex (green) and dA16 (red) and dG10 (blue) substituted A_6-RNA. (D) The sum of the scalar coupling constants $\Sigma J_{H1'}$ and $\Sigma J_{H2''}$ in DQF-COSY[77] indicate that the sugar pucker at dA16 and dG10 in substituted A_6-RNA is C3′-*endo*[42,68].

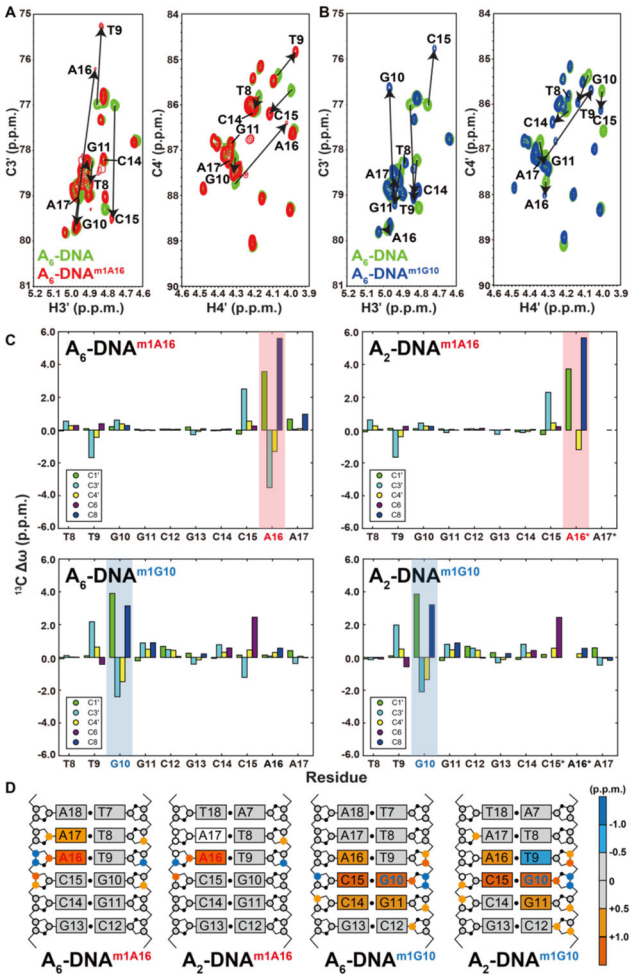

Figure 3. C3′ and C4′ chemical shift perturbations induced by m^1A and m^1G in duplex DNA

Overlays of 2D C3′-H3′ and C4′-H4′ HSQC spectra for A_6-DNA (green) with (A) A_6-DNA^{m1A16} (red) and (B) A_6-DNA^{m1G10} (blue). Corresponding overlays for A_2-DNA are shown in Supplementary Figure 3. (C) ^{13}C chemical shift perturbations induced by m^1A and m^1G (shade in red and blue, respectively) relative to the unmodified duplex. Star denotes sites that are broadened out of detection. (D) ^{13}C chemical shift perturbations induced by m^1A16 or m^1G10 highlighted on the structures of A_2-DNA and A_6-DNA (residues 1 to 6 which experience minor perturbations are not shown for clarity). Colored circles and rectangles represent sites with significant carbon chemical shift perturbations (> 0.5 p.p.m.) on deoxyribose ^{13}C and aromatic ^{13}C, respectively. Open circles on sugars and rectangles on base denote sites that are broadened out of detection.

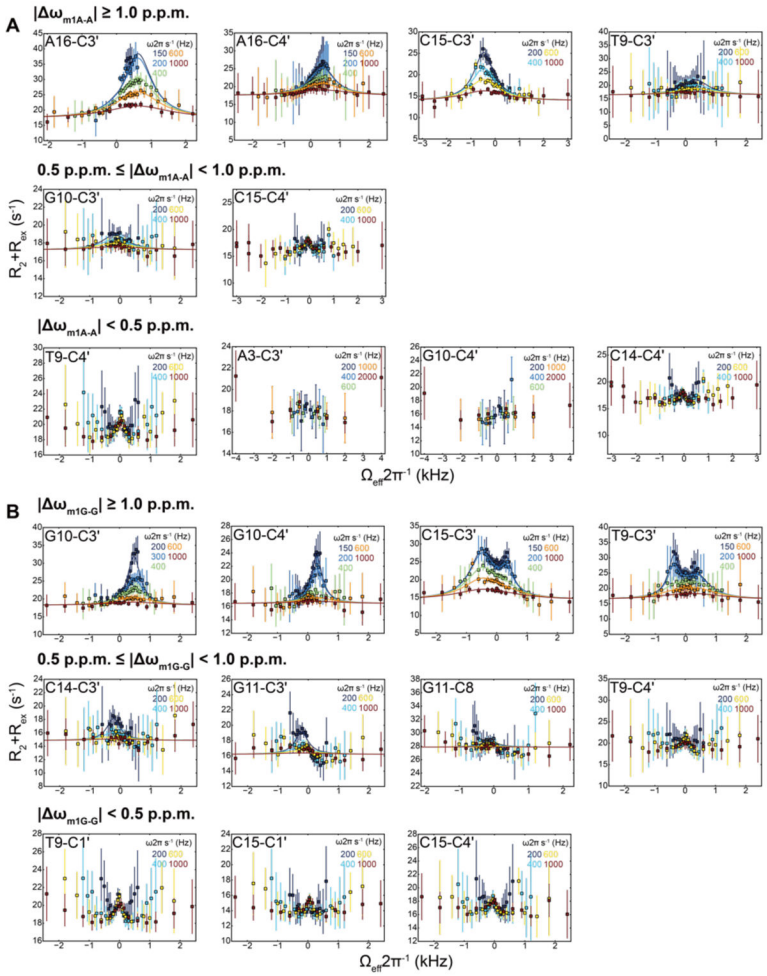

Figure 4. Off-resonance ^{13}C R1ρ RD profiles measured in A_6-DNA

Shown are representative (all other flat RD profiles shown in Supplementary Figure 5) ^{13}C off-resonance R1ρ RD profiles with best global fit of A_6-DNA measured at (A) pH 6.8 highlighting A16-T9 Hoogsteen bp exchange and (B) pH 5.4 highlighting both A16-T9 and G10-C$^+$15 Hoogsteen bp exchange at 25 °C. Error bars represent experimental uncertainty (one s.d., see Methods). Also shown are the predicted differences in chemical shifts Δω between Watson-Crick and Hoogsteen based on comparison of chemical shifts measured in the m^{1}A and m^{1}G Hoogsteen stabilizing mutants with their unmodified counterparts. RD is not readily detectable for |Δω| < 0.5 p.p.m whereas it should be observed for |Δω| > 1.0 p.p.m. |Δω| values between 0.5 to 1.0 p.p.m. may or may not lead to detectable RD depending the quality of the data and the exchange parameters.

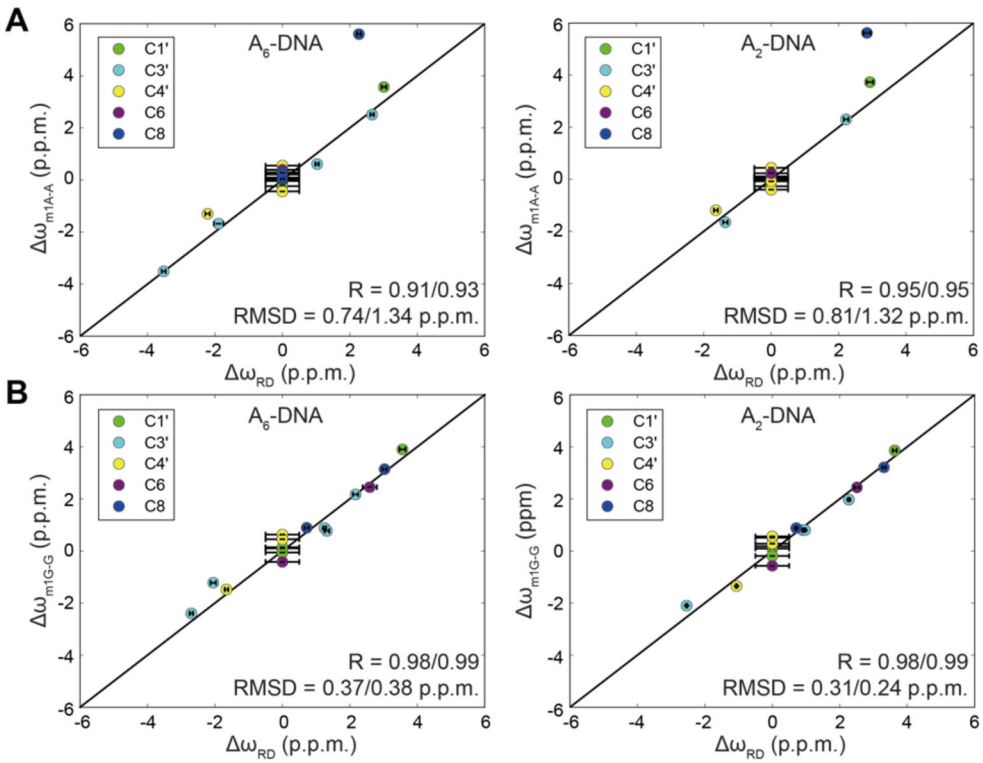

Figure 5. Comparison of chemical shifts for ES Hoogsteen bps and m^1A or m^1G stabilizing mutants

(A) Comparison of $\Delta\omega_{RD} = \omega_{ES} - \omega_{GS}$ measured by RD with values obtained from the differences in chemical shifts observed in the modified and unmodified duplexes (A) $\Delta\omega_{m1A\text{-}A} = \omega_{m1A} - \omega_A$ and (B) $\Delta\omega_{m1G\text{-}G} = \omega_{m1G} - \omega_G$ in A_2-DNA and A_6-DNA. RMSD and R values are shown for all data and when excluding data points with flat RD for which $\Delta\omega_{RD}$ was assumed to be 0.

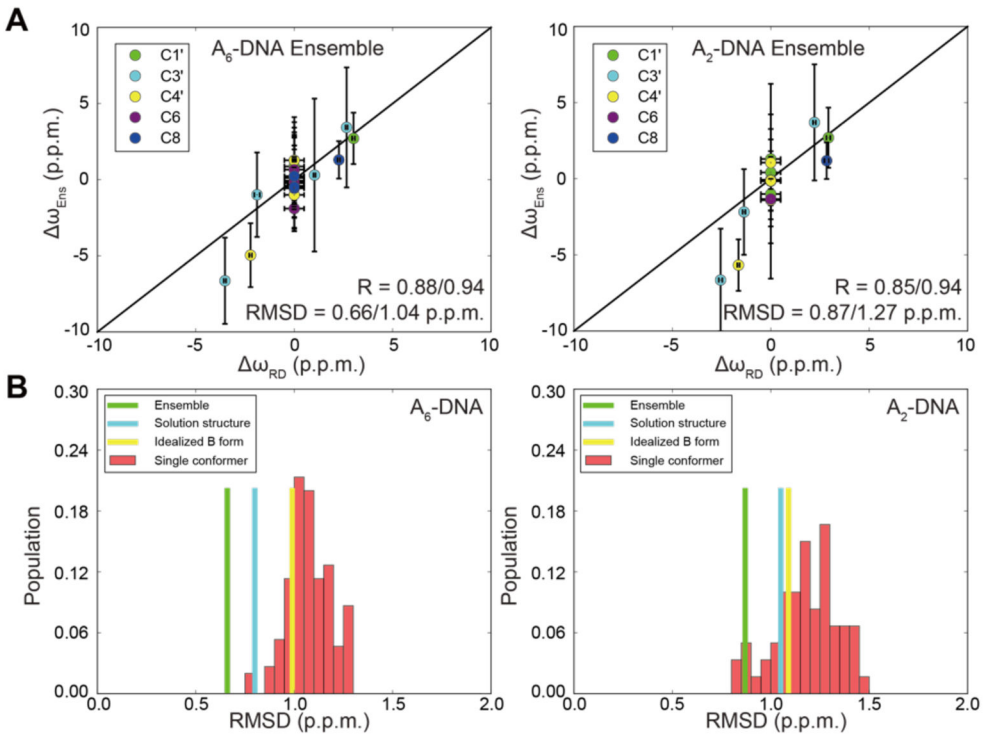

Figure 6. Testing validity of structures and ensembles of Hoogsteen stabilizing mutant as mimics of ES Hoogsteen using structure-based predictions of chemical shifts

(A) Comparison of $\Delta\omega_{RD} = \omega_{ES} - \omega_{GS}$ measured by RD with values obtained from the differences in calculated chemical shifts predicted based on the ensemble of structures of the modified and unmodified duplexes i.e. $\Delta\omega_{Ens} = \omega_{calc\text{-}m1A} - \omega_{calc\text{-}A}$ in A_2-DNA and A_6-DNA. RMSD and R values are shown for all data (before the slash), and when excluding data points with flat RD for which $\Delta\omega_{RD}$ was assumed to be 0 (after the slash). Error bar represents the standard deviation in the calculated chemical shifts across all conformers in the ensemble. (B) Histogram showing the root-mean-square-deviation (RMSD) between $\Delta\omega_{meas}$ and $\Delta\omega_{calc}$ when using different input NOE structures and RDC ensembles to compute $\Delta\omega_{calc}$: dynamic ensembles (green), solution structures (cyan), idealized B-form DNA (yellow), and taking the difference between the chemical shifts predicted for every conformer in the ES- and GS-ensemble (red).

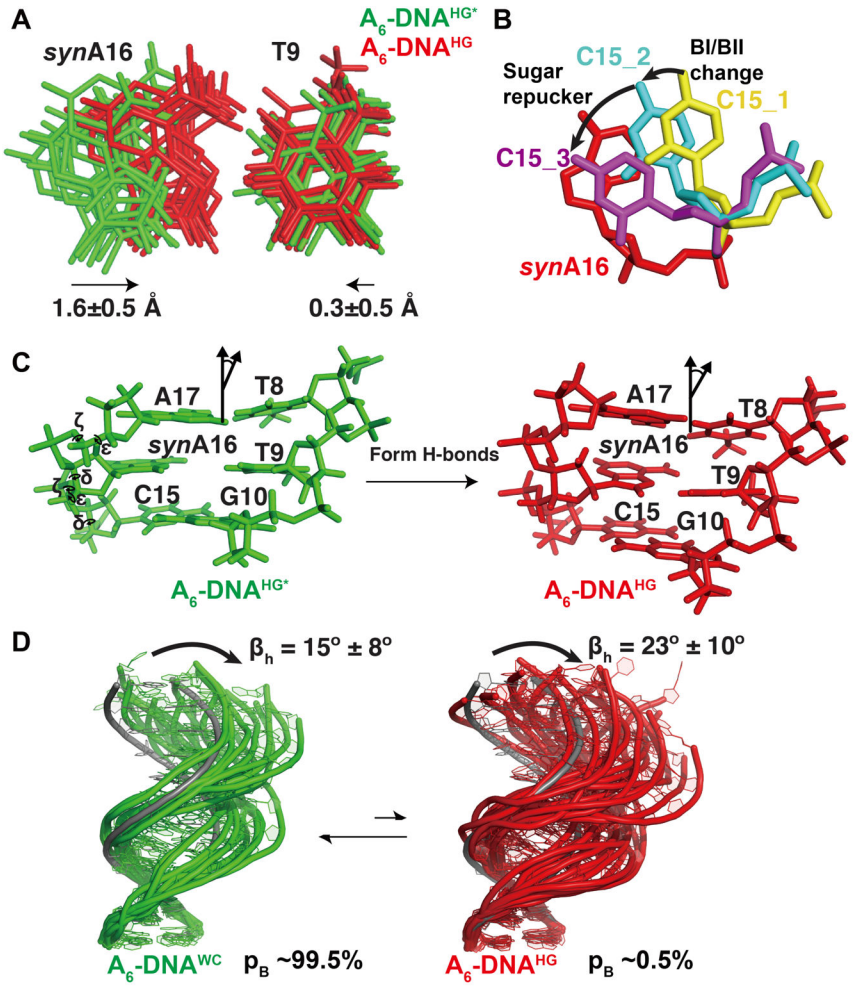

Figure 7. Structure and dynamic ensembles of GS Watson-Crick and ES Hoogsteen bps in duplex DNA

(A) Comparison of GS A_6-DNA (green) and Hoogsteen ES-mimic A_6-DNA^{m1A16} (red) ensembles showing translation of *syn*A16 towards the complementary strand. The A16 in A_6-DNA was flipped into a *syn* conformation to aid comparison. (B) Translation of the *syn* purine to form Hoogsteen H-bonds with the partner thymine requires changes in sugar pucker and backbone torsion angles at C15. (C) Changes in the local torsion angles (ε, ζ, sugar phase angle/δ) at C15 increase overlap area with the flipping and sliding *syn*A16. (D) Ensembles of the GS A_6-DNA (green) and Hoogsteen ES-mimic A_6-DNA^{m1A16} (red) showing global changes in structure and dynamics accompanying formation of Hoogsteen bps. An idealized B-form DNA helix (in gray) is overlaid for reference. The ensembles were taken from a previous study[42].