Using NMR and molecular dynamics to link structure and dynamics effects of the universal base 8-aza, 7-deaza, N8 linked adenosine analog

Alexander M. Spring-Connell[1], Marina G. Evich[1], Harald Debelak[2], Frank Seela[2,3] and Markus W. Germann[1,*]

[1]Department of Chemistry, Georgia State University, Atlanta, GA 30303, USA, [2]Laboratorium für Organische und Bioorganische Chemie, Institut für Chemie neuer Materialien, Universität Osnabrück, Barbarastraße 7, 49069 Osnabrück, Germany and [3]Laboratory of Bioorganic Chemistry and Chemical Biology, Center for Nanotechnology, Heisenbergstraße 11, 48149 Münster, Germany

ABSTRACT

A truly universal nucleobase enables a host of novel applications such as simplified templates for PCR primers, randomized sequencing and DNA based devices. A universal base must pair indiscriminately to each of the canonical bases with little or preferably no destabilization of the overall duplex. In reality, many candidates either destabilize the duplex or do not base pair indiscriminatingly. The novel base 8-aza-7-deazaadenine (pyrazolo[3,4-d]pyrimidin- 4-amine) N^8-(2′deoxyribonucleoside), a deoxyadenosine analog (UB), pairs with each of the natural DNA bases with little sequence preference. We have utilized NMR complemented with molecular dynamic calculations to characterize the structure and dynamics of a UB incorporated into a DNA duplex. The UB participates in base stacking with little to no perturbation of the local structure yet forms an unusual base pair that samples multiple conformations. These local dynamics result in the complete disappearance of a single UB proton resonance under native conditions. Accommodation of the UB is additionally stabilized via heightened backbone conformational sampling. NMR combined with various computational techniques has allowed for a comprehensive characterization of both structural and dynamic effects of the UB in a DNA duplex and underlines that the UB as a strong candidate for universal base applications.

INTRODUCTION

The genetic code relies on the precise base pairing of the four natural bases. However, an objective in synthetic bi-

ology is the expansion of the 'genetic alphabet' with the development of unnatural bases. These efforts have shown promise in a variety of roles including polymerase chain reaction (PCR) amplification, RNA transcription and DNA primers and probes (1–3). In contrast to novel synthetic base pairs that feature specific pairing, a universal base would be indiscriminate in its base pairing and ideally pair with all four natural bases uniformly. In the biotechnology field, a truly universal base could be implemented in probes and primers and would be highly useful (4–11).

Ideally, a universal base should be incorporated into a DNA duplex environment with little perturbation to the duplex structure. Natural DNA base pairs are formed via canonical base stacking and Watson–Crick hydrogen bonding patterns generating a regular B-type DNA. Thus, regular base stacking, hydrogen bonding and steric interactions are important design features for the stable incorporation of any synthetic nucleobase analog within a duplex environment. Examples of universal base designs range from single and fused aromatic ring systems to aliphatic and even abasic designs; each of these schemes undermine either structure and/or stability of the DNA duplex (12–15). Recent examples of a prominent family of universal bases include the non-hydrogen bonding 3-nitropyrrol and 5-nitrolindole analogs. NMR and thermostability studies of theses analogs reveal a unique intercalating base stacking scheme, which aids in its ambiguous base pairing scheme while maintaining an overall B-type conformation with a reduced thermostability ($\Delta T_m \sim 6°C$ per synthetic nucleobase in a 16er oligonucleotide sequence) (16). Although novel, this intercalation presents an altered local conformation as compared to canonical B-type DNA, which may or may not be desirable.

The Seela lab has developed a series of N^8 linked 2′ deoxyadenosine analogs that exhibit universal base properties (17–22). The 8-aza-7-deazaadenine (pyrazolo[3,4-

---

*To whom correspondence should be addressed. Tel: +1 404 413 5561; Fax: +1 404 413 5505; Email: mwg@gsu.edu
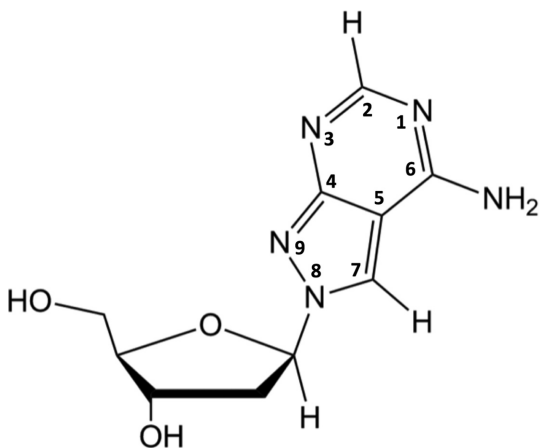
Figure 1. UB schematic. Schematic of 8-aza-7-deazaadenine (pyrazolo[3,4-d]pyrimidin-4-amine) N^8-(2′deoxyribonucleoside) (UB) in the "anti" orientation.

d]pyrimidin-4- amine) N^8-(2′deoxyribonucleoside) de-oxyadenosine analog (UB) base pairs indiscriminately with all four natural base pairs despite the unique N^8 glycosidic linkage (Figure 1) (18,20–22). In dodecamer duplexes where UB is paired with all four natural bases, the T_m values are within 2°C of one another. Compared to an A:T base pair, the UB-(A,C,G,T) base pair results in a drop in thermal stability with ΔT_m of −4 to −6°C. Additionally, it exhibits less fluctuations than mismatches (mismatch ΔT_m values of −2 to −12°C), and is more stable than abasic sites (abasic ΔT_m values of −10 to −17°C) (18).

Understanding how the UB is incorporated into a duplex environment will aid in a refined design of universal bases and shed light on its interaction with DNA servicing proteins and enzymes (16). Solution state NMR was utilized to determine a high-resolution structure of a DNA duplex containing a UB opposite a dT; the duplex was also probed for regions of unusual dynamics and chemical exchange. NMR data complemented with extended molecular dynamics, molecular dynamics with time averaged restraints and chemical shift calculations reveal that the UB is well accommodated in the duplex with very little impact to global and local geometry. Nevertheless, UB is also the source of unusual dynamics on the high μs to low ms timescale, which results in the disappearance of the UB base H2 proton in all NMR spectra under native conditions but with little impact on other resonances.

MATERIALS AND METHODS

Oligonucleotides

UB containing oligonucleotides were synthesized and purified as previously described (18,19). All other oligonucleotides were purchased from University of Calgary Core DNA Services and purified via ion exchange (Resource 15Q column) and size exclusion liquid chromatography. Purity was assessed via denaturing polyacrylamide gels. Concentrations were calculated using extinction coefficients from the sum of mononucleotides (absorbance at 260 nm, 80°C) (23). The extinction coefficient for the UB is 6600 $M^{-1}cm^{-1}$

(18). Sequences used for NMR are as follows (X = UB or dA):

```
     1     5     9
  5′-ATGGXGCTC-3′
  3′-TACCTCGAG-5′
    18        10
```
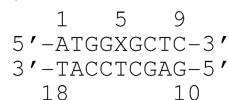
Melting temperature studies

DNA duplexes for T_m measurements are detailed in Table 1. Samples for T_m measurements were in 100 mM NaCl, 10 mM $MgCl_2$ and 10 mM sodium cacodylate with a pH of 7.0. Total DNA strand concentration was 10 μM. UV absorbance was measured on a Cary-1/1E UV/VIS spectrophotometer. Thermodynamic data were calculated using the MeltWin 3.0 package. T_m measurements in 100 mM NaCl with 10 mM sodium phosphate at a pH of 6.8 were also completed on the NMR control sample (A:T) and UB sample (UB:T) and analyzed as described previously (23).

NMR sample conditions

NMR samples were completed in a buffer containing 10 mM sodium phosphate, 100 mM NaCl, with 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) added as an internal reference. The pH* for D_2O samples was 6.36; the pH for H_2O samples was 6.79. NMR samples for imino proton spectra collected at supercooled conditions were transferred to 1 mm borate glass capillaries with the ends flame sealed; capillaries were combined in a standard 8 mm NMR tube (24,25). Temperatures below −5°C were reduced at a rate of 1°C per hour (26).

NMR experiments and structure calculations

Detailed methods for NMR experiments and structure calculations are provided in Supplementary Data. Briefly, NMR experiments were conducted on a Bruker Avance 600 MHz NMR equipped with a 5 mm or 8 mm QXI probe, ^1H {^{13}C, ^{15}N, ^{31}P}; HSQC and ^{13}C T_1 experiments were conducted on a 500 MHz Avance systems equipped with a 5 mm TXI cryoprobe ^1H{^{13}C, ^{15}N}. Standard protocols were utilized for the collection of 1D ^1H and ^{31}P and 2D COSY, TOCSY, CTNOESY, HPCOR, HSQC, and NOESY spectra as previously described (27–30). D_2O NOESY spectra were collected with mixing times of 75, 150 and 250 ms and an 8 s interscan delay. RDC values were determined for NMR signals from the sugar and base regions using f2 coupled HSQC spectra collected in the presence and absence of pf1 phage (ASLA) (^2H splitting of 18.0 and 32.2 Hz giving concentrations of 20.3 and 36.3 mg/ml for the UB and control samples respectively) (31). Structure calculation utilized an iterative RANDMARDI process with MARDI-GRAS, CORMA and AMBER 9.0 cycles as described previously (27,28,32–39). The B-type backbone torsion angles (except for the experimentally determined ϵ torsion angles) and Watson–Crick hydrogen bond restraints were not included for the UB5 residue.

The AMBER parm99 force-field was appended to include specific parameters for the UB (40). (Protocols for the AMBER force-field paramaterization can be found in

Supplementary Data). Starting structures were canonical B-type DNA with the incorporation of UB at position 5. Fully restrained molecular dynamics were conducted in explicit water (TIP3P) and charges were neutralized with the addition of 16 Na+ counterions; simulation times were 10 ns. The final bundle of structures was selected from the last 100 ps at a rate of 1 structure per 10 ps. Structural and helical parameters were determined using the Curves+ algorithm (41). Heavy atom RMSD calculations were completed in VMD 1.9 (42). Molecular visualization was completed in VMD 1.9 and UCSF Chimera (42–44).

Computational methods

For a detailed description of the molecular dynamics simulations, see Supplementary Data. Briefly, molecular dynamics simulations using time-averaged restraints (MDtar) and AMBER 9.0 with a $\tau = 20$ ps were completed as described previously (45). The PDQPRO algorithm was used in conjunction with CORMA to select high probability structures from MDtar trajectory of structures (46). Unconstrained extended molecular dynamic simulations were completed on the UB and control oligonucleotides with additional GC base pairs at the terminal ends to stabilize the ends and avoid artifacts (47). The sequence used for unconstrained extended molecular dynamics simulations is:

```
      1     5     9
   5'-GCATGGXGCTCGC-3'
   3'-CGTACCTCGAGCG-5'
     18          10
```
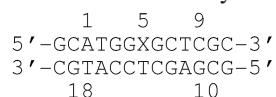
where X is either the UB or dA. NMR chemical shift calculations were completed with Gaussian 03 (revision C.02) using a DFT B3LYP/6-31g(d,p) basis set and the GIAO method in vacuum.

RESULTS

Thermal stability

The thermal stability for the four UB containing oligonucleotides are similar with T_m and ΔG values all within 1°C and 0.5 kcal/mol of one another with the UB opposite dT having a slightly higher T_m than the others (Table 1, section A). Although the UB thermal stabilities are compromised compared to the canonical A:T base pairing (a reduction from 42°C for the A:T duplex to 33.2–34°C for the UB oligonucleotides), all four canonical bases opposite the UB exhibit a similar stability underlining its universal base pairing capabilities (Table 1). Differences in ΔG as compared to the control are ∼2.2-2.7 kcal/mol lower than the control potentially indicating different stacking or hydrogen bonding pattern for the UB containing duplexes. Additionally, thermal stability studies were conducted on the UB:T and control samples using NMR buffer conditions (Supplementary Figure S17). As anticipated, the thermal stabilities for the UB and control sequence were lower in the absence of Mg^{2+} with the difference in T_m and ΔG between the two samples being higher as compared to the Mg^{2+} containing samples (UB:dT and control T_m 31.4 ± 0.4°C and 41.7 ± 0.4°C, respectively; $\Delta\Delta G$ 2.6 kcal/mol) (Table 1, section B). Based on these results and the UB being an adenosine ana-

Table 1. (A) Thermodynamic properties of UB containing oligonucleotides in the same sequence context but with varying base pair partners for the UB. Also included for comparison are the same sequences with an A:T base pair and a A:C mismatch. The X represents the UB. Sample conditions are as follows: 10 μM total strand concentration, 100 mM NaCl, 10 mM MgCl$_2$ and 10 mM sodium cacodylate with a pH of 7.0. [a]Melting curve is not fitable. (B) Thermodynamic properties of the UB containing oligonucleotide and control sequence using NMR buffer conditions (100 mM NaCl, 10 mM sodium phosphate with a pH of 6.8). The X represents the UB. For both sections, ΔG was calculated at 21°C using $\Delta G = \Delta H - T\Delta S$

Duplex	T_m (°C)	ΔH (kcal mol^{-1})	ΔS (cal mol^{-1} K^{-1})	$\Delta G_{21°C}$ (kcal mol^{-1})
A				
5' A-T-G-G-**A**-G-C-T-C 3' T-A-C-C-**T**-C-G-A-G	42.8	-77.0	-218.1	-12.9
5' A-T-G-G-**A**-G-C-T-C 3' T-A-C-C-**C**-C-G-A-G	< 25[a]	-	-	-
5' A-T-G-G-**X**-G-C-T-C 3' T-A-C-C-**T**-C-G-A-G	34.0	-67.1	-192.8	-10.4
5' A-T-G-G-**X**-G-C-T-C 3' T-A-C-C-**C**-C-G-A-G	33.2	-79.1	-232.7	-10.7
5' A-T-G-G-**X**-G-C-T-C 3' T-A-C-C-**A**-C-G-A-G	33.5	-64.4	-184.3	-10.2
5' A-T-G-G-**X**-G-C-T-C 3' T-A-C-C-**G**-C-G-A-G	33.9	-62.0	-176.3	-10.2
B				
5' A-T-G-G-**A**-G-C-T-C 3' T-A-C-C-**T**-C-G-A-G	41.6 ± 0.4	-65.6 ± 5.5	-194.6 ± 0.3	-11.4 ± 0.1
5' A-T-G-G-**X**-G-C-T-C 3' T-A-C-C-**T**-C-G-A-G	31.3 ± 0.4	-54.5 ± 4.5	-158.6 ± 0.3	-8.8 ± 0.1

log, NMR and dynamics studies were limited to the slightly more stable UB opposite the dT.

UB glycosidic bond orientation

The unique UB glycosidic bond orientation will be defined using the same naming convention as originally reported where χ = O4'-C1'-N8-C7 and the minimum distance between H7 and H1' is defined as 'anti' (18). The 'syn' and 'anti' conformations yield quite different H7-H1' distances ('syn' 3.8 Å and 'anti' 2.5 Å) and was used to characterize the glycosidic bond torsion angle (Figure 1). A NOESY spectra with a 75 ms mixing time was used to estimate the distance between the UB5 H7 and H1' protons at 2.3 Å. This establishes the base orientation as 'anti' with the H7 proton pointing into the minor groove and the H2 proton in the major groove. These results were verified with H7 to H2', H2″ and H3' distances. The residues surrounding the UB are all in canonical anti conformations.

NMR structure data

All anticipated NOE base to H1' contacts were identified, with the only unusual features being a reduction in intensity between G4 H8 and UB5 H1' and the previously described high intensity UB5 H7 to UB5 H1'. All base to base, H2', H2″ and H3' pathways are accounted for with the exception of the base to base and base to H3' pathways between G4 to UB5 (Figure 4A). In addition, several unique crosspeaks
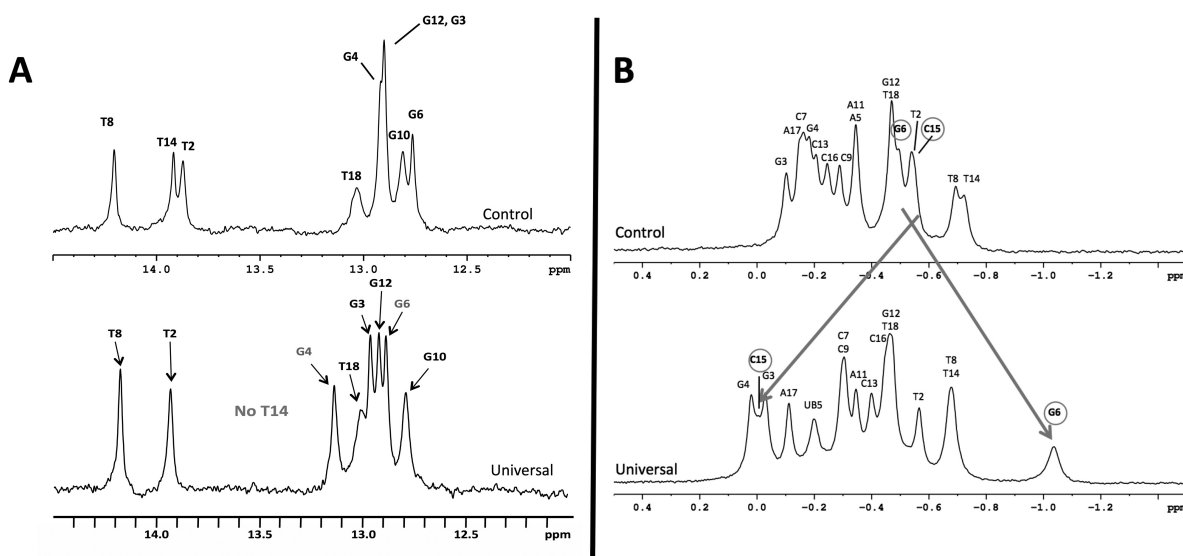
Figure 2. Imino proton and ^{31}P NMR spectra. (**A**) ^{1}H imino proton spectrum of the control and UB sequences at 7°C (top and bottom, respectively). Imino protons for residues G4 and G6 of the UB sequence (which flank UB5) are labeled in grey. (**B**) ^{31}P spectra for the control and UB sequences at 22°C. Arrows and circles indicate the G6 and C15 phosphorous resonances, which exhibit significant perturbations in chemical shifts as compared to the control.

were observed: G6 H1′ to UB5 H7 and G6 H3′ to UB5 H7 as well as UB5 H1′ to G6 H3′ and UB5 H1′ to G6 H4′ (Figure 4A). Collectively, these unique contacts along with the G4 H8 to UB5 H1′ NOESY crosspeaks, confirm an intrahelical orientation of the UB base (Supplementary Figure S2B). The measured NOEs and intensities for the rest of the oligonucleotide are consistent with an overall B type helix.

1D and 2D imino proton spectra were used to detect and assign base pairs. All imino proton resonance except the T14 imino proton were observed under normal conditions (7°C) and exhibit chemical shifts indicative of normal A:T and G:C Watson–Crick base pairing. Of particular note, the bases flanking the UB5 (G4 and G6) exhibit sharp line shapes indicating stable base pair formation. All imino protons resonances were detected for the control sequence (Figure 2A). NOESY spectra collected in H$_2$O were used to identify the imino/amino pathways. The UB5 amino group could not be identified in NOESY spectra collected in H$_2$O. Finally, a weak NOESY crosspeak between G4's imino proton and UB5 H7 was observed, which further confirms the UB5 intrahelical orientation. Interstrand NOESY crosspeaks involving UB5 or T14 were not observed.

Absolute values of differences in chemical shifts (|UB – control| = Δ ppm) for sugar protons and phosphorus resonances were calculated. As shown in Supplementary Figure S2A, the largest chemical shift perturbations for the sugar protons were localized near the UB5:T14 base pair with the T14 sugar protons containing the largest overall chemical shift changes and the largest changes in individual proton resonances. Base proton chemical shift changes were mainly localized to C15 (Supplementary Tables S2 and S3). As compared to the control sequence, ^{31}P chemical shifts for UB5-P-G6 and T14-P-C15 are shifted +0.52 and −0.56 ppm from their control values, respectively (Figure 2B, Supplementary Tables S2 and S3).

All measured three bond sugar coupling values for UB5 were consistent with a 2′ endo / 3′ endo equilibrium and place the UB5 sugar in an S conformation with a pseudorotation value of 125–170° and fraction south of 80%. With the exception of C7 (50% f$_s$), all other residues are predominately in an S conformation (Supplementary Table S4). For the control sequence, all residues measured were found to be in S conformation. Coupling values for control C7 could not be measured due to signal overlap, however, examination of the COSY coupling patterns suggest a similar fraction south range as the UB sequence (50% fs). Experimentally determined epsilon (ϵ) torsion angles did not reveal a unique perturbation for any of the residues including UB5 and flanking bases (ϵ torsion angle range of 163° to 171°) (Supplementary Table S5). Interestingly, the ^{1}H – ^{31}P HETCOR experiments did show a weak intensity for the H3′ to ^{31}P crosspeak in the UB5-P-G6 sequence (Figure 4D).

Structure calculation

The starting conformation for both duplexes was a B-type geometry with the UB5:T14 base pair in a reverse Watson–Crick-like base pair. For both sequences, the rMD trajectory was analyzed for variations in structure conformation through the monitoring of key torsion angles and base pair distances (torsion angles presented in Figure 5; base pair distance data not shown). Little variation was observed allowing for sampling from the final 100 ps of the trajectory. The UB containing duplex is highly restrained containing a total of 480 restraints giving an average 27 restraints per residue (Table 2). The average heavy atom R.M.S.D. for the final bundle of structures is 0.42 Å, and the structure with the overall lowest CORMA R^X and AMBER restraint violations was chosen as the final representative structure (Supplementary Figure S3). The structure shows excellent agreement with NOESY data with total CORMA R^X values ranging from 4.4 to 5.5 and AMBER distance and RDC

Table 2. Summary of NMR restraints used for the structure development of the UB and control structures. Qualitative distance restraints involving exchangeable protons were obtained from a H_2O NOESY spectrum with a mixing time of 150 ms. The structure rMD calculations were run for 10 ns. Charges were neutralized with 16 Na+ ions and simulations were conducted in explicit TIP3P water. Final structures underwent fully restrained energy minimization (2000 steps steepest decent). Heavy atom R.M.S.D. values were calculated in VMD version 1.9. Of note, residue A1 for both the control and UB sequences exhibit a syn/anti equilibrium with <25% syn population, which was evident in the integrated NOESY crosspeaks (Supplementary Figure S4). Because this observation was present and similar for both sequences and obviously not a result of the UB incorporation, the A1 residue was constrained to an anti conformation for both sequences.

Parameter	UB	Control	Force Constant [a]
Distance Restraints			
Non-exchangeable (RANDMARDI) distance restraints	210	144	
intra residue	139	78	20 or 30 [b]
inter residue (sequential)	71	66	30
Average well width (Å)	0.70	0.80	
Qualitative exchangeable distance restraints	34	0	30
Qualitative A1 anti glycosidic angle restraints	2	2	30
Watson–Crick Base Pair Restraints			
Watson-Crick distatnce restraints	21	23	25
Watson-Crick flat angle restraints	21	23	10
Backbone Torsion Angle Restraints			
Broad angle DNA restraints ($\alpha, \beta, \gamma, \zeta$)	61	66	50
ε (CTNOESY) torsion angle restraints	16	16	50
Endocyclic torsion angle restraints	75	90	50
Total Restraints	**440**	**364**	
Average Restraints / Residue	**24.4**	**20.2**	
RDC Restraints	**40**	**34**	1.0 (dwt)
Total Restraints with RDC	**480**	**398**	
Average Restraints / Residue	**26.7**	**22.1**	
Average Heavy Atom R.M.S.D. (Å)	0.42	0.53	
Final AMBER Penalties			
AMBER distance penalty (kcal/mol)	36.6	14.3	
Average deviation (Å)	0.03	0.02	
Standard deviation (Å)	0.06	0.05	
Maximum deviation (Å)	0.33	0.29	
AMBER angle penalty (kcal/mol)	0.4	0.2	
Average deviation (degrees)	1.5	1.0	
Standard deviation (degrees)	2.1	1.6	
Maximum deviation (degrees)	6.7	5.6	
AMBER Torsion angle penalty (kcal/mol)	4.0	11.2	
Average deviation (degrees)	0.9	1.8	
Standard deviation (degrees)	3.4	8.0	
Maximum deviation (degrees)	19.1	69.3	
Total RDC alignment constraint (kcal/mol)	12.9	3.0	

[a] (kcal/mol * unit of violation)
[b] 20 kcal/mol for intraresidue sugar restraints, 30 kcal/mol for all others

CORMA R^X

UB Structure				Control Structure			
T_m (ms)	intra	inter	total	T_m (ms)	intra	inter	total
75 ms	5.14	6.05	5.49	75 ms	4.72	6.13	5.38
150 ms	4.88	4.92	4.89	150 ms	4.43	6.01	5.10
250 ms	4.17	4.66	4.35	250 ms	3.34	4.51	3.89

violations of 36.6 and 12.9 kcal/mol respectively (R^X values multiplied by 100). The control structure has fewer restraints than the UB structure due to signal overlap (~22 restraints per residue); it too exhibits excellent agreement with NOESY data with total CORMA R^X values ranging from 3.9 to 5.4 and AMBER distance and RDC violations of 14.3 and 3.0 kcal/mol for all restraints (Table 2). Both the UB5 and the A5 residues are well restrained in each of the structures with the UB5 H7 proton having a total of 7 distance restraints and the A5 H8 proton having 8 restraints (Supplementary Figure S2B).

Structure analysis

All bases including both the UB5 and T14 bases are intrahelical and the overall global geometry is B-type (Supplementary Figure S3). The UB5:T14 base pair does not form a hydrogen bond involving the T14 imino proton, which is consistent with the absence of the lack of the T14 imino proton. Instead, the amino group of UB5 is hydrogen bonded to the T14 O4 carbonyl (Figure 3A). This unusual base pairing results in heightened stretch and opening (0.75 Å and −23°, respectively) manifesting in an increase of 1 Å in the UB5:T14 anomeric carbon distance as compared to the control structure (Figure 3A, Supplementary Figure S3). As a consequence of the unique N8 UB base connectivity and orientation, the local stacking of G4-UB5-G6 is slightly disrupted with somewhat G4 poorly stacked on top of UB5. Interestingly, the perturbation is confined to the UB5; G4 and G6 remain in a similar conformation as the control sequence (Figure 3B). This unusual orientation was persistent with repeated structure calculations indicating that the energetically unfavorable base pair configuration is represented in all of the UB5:T14 base pair restraints. The overall solved UB and control structures are quite comparable with a total heavy atom R.M.S.D. of 0.98 Å (due to different connectivity and atom placement, UB5/A5 were excluded from the calculation).

Supercooled aqueous NMR yields a T14 imino proton signal

Previous thermal stability studies with UB containing oligonucleotides and other UB-like analogs predicted a reverse Watson–Crick like UB-T base pair and would exhibit a hydrogen bonding pattern with the involvement of the thymine imino proton [17,18] (Figure 3C). Due to the quite unusual UB5-T14 hydrogen bonding pattern found in the solved structure, further probing of the imino proton spectra was conducted via supercooled aqueous NMR utilizing 1 mm capillaries. At temperatures >4°C, no T14 imino proton signal is observed. When temperatures are dropped to −8°C, a shoulder begins to emerge at 13.6 ppm suggestive of the formation of an additional A:T–like base pair (Supplementary Figure S8). A NOESY spectrum collected at −12°C was used to confirm the identity of the new resonance as T14 through the appearance of a cross peak between the new imino proton resonance and a methyl resonance similar to T14's methyl chemical shift at 21°C (Figure 4B, Supplementary Table S2). All observed imino proton resonances shifted as anticipated with decreasing temperatures indicating that an overall global conformation change in the oligonucleotide did not occur (Supplementary Figure S9).

Base opening rates

Base pair lifetimes were calculated for non-overlapped imino proton peaks of the UB and control sequences through the addition of NH_4OH as a catalyst (Supplementary Figure S18). For the control sequence, the A5:T14 base pair exhibits a lifetime of 3 to 5 ms; because no T14 imino proton signal is observed under standard conditions for the UB sequence, a corresponding base opening for the
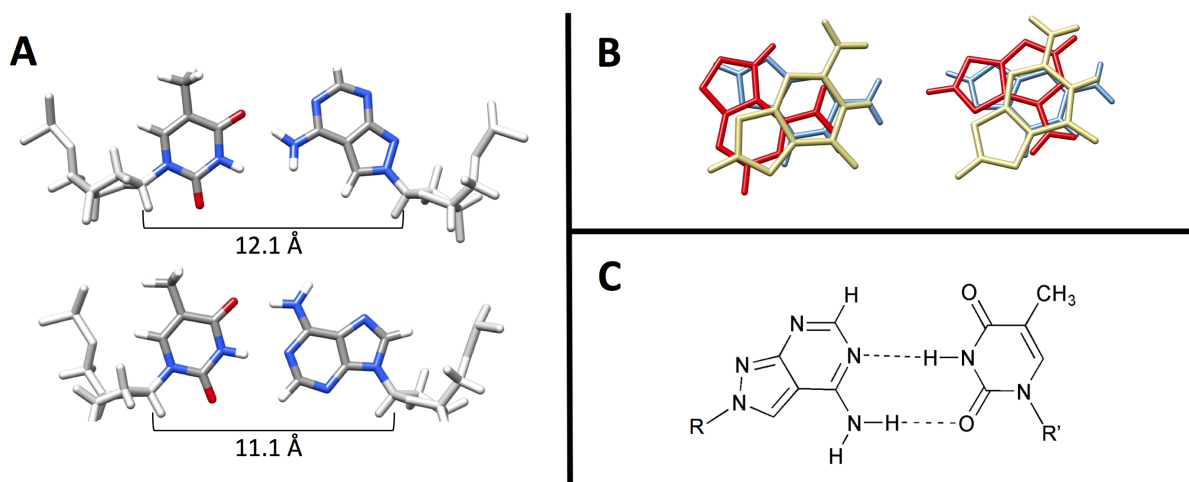
Figure 3. UB5:T14 and A5:T14 base conformations. (**A**) Base pair for UB5:T14 (top) and the control A5:T14 (bottom) (both are from the solved rMD structures). Anomeric carbon distances are referenced. Sugar and backbone are in grey; base colors: blue = nitrogen, grey = carbon, red = oxygen, white = hydrogen. (**B**) Base stacking for G4 (yellow), UB5/A5 (red) and G6 (blue). UB sequence on left, control on right (both are from the solved rMD structures). (**C**) Predicted UB-T base pair hydrogen bonding pattern determined from previous thermal stability studies (Seela F. & Debelak, H., NAR, 2000).

UB5:T14 base pair could not determined. Significant overlap of the control sequence limited quantitative analysis to G6:C13 which was found to have a base pair lifetime of 40 to 50 ms. Based on findings from other studies, these results are consistent with canonical A:T and G:C Watson–Crick base pair lifetimes (48–50). Similar to the control sequence, the base pair lifetimes for many of the UB sequence residues could not be quantitatively determined due to signal overlap. G4-C15 exhibits a lifetime of ∼3 ms. The G6 and G12 imino proton signals are overlapped and lifetimes could not be determined individually. Therefore, a qualitative approach was taken with some overlapped peaks where the imino proton resonance's recovery was compared to other peaks in the same sequence. Under high concentrations of catalyst, stable base pairs recover more slowly than short lived ones. Inspection of the UB sequence's G6/G12 overlapped peaks reveal that the signal recovers more slowly than the G4 peak suggesting that G6 and G12 have a longer base pair lifetime (Supplementary Figure S19). Additionally, although the control G3, G4 and G12 imino proton peaks are overlapped, the relative position of the G4 resonance within the overlapped peak is known to be on the downfield side (Figure 2A). With this, G6 appears to recover comparably with G4 suggesting that G4 and G6 have similar base pair lifetimes (Supplementary Figure S20). For both sequences, the imino proton resonance for T2:A17 and T8:A11 disappear upon the first addition of NH_4OH (∼3 mM NH_3).

'The case of the disappearing H2 proton'

The UB5-H2 proton has a similar chemical environment as compared to an adenosine H2 and is anticipated to have similar NMR and chemical traits. Under denaturing conditions, all 4 H2 protons in the UB sequence are readily identified by both 1D ^1H NMR and ^1H–^{13}C HSQC spectra. A spectrum recorded for the single strand identified the UB5 H2 and C2 resonances at 7.92 and 156 ppm, respectively. Unexpectedly, when annealed, the UB5-H2 proton vanishes

from all spectra; only the three adenosine H2 protons can be identified in 1D ^1H, HSQC and NOESY spectra (Figure 4C, Supplementary Figure S1). The UB H2 proton reappears upon subsequent denaturing of the duplex; this process is readily reversible. Careful examination and integration of the base region ruled out resonance overlap. This anomaly is restricted to the UB5 H2 proton; the UB5 H7 proton, UB5 sugar protons as well as protons of other neighboring residues including the T14 protons were readily identified and exhibit sharp peaks in 1D spectra.

Reduced crosspeak intensities

A number of 2D NMR crosspeaks in the HSQC, low-flip angle COSY and ^1H–^{31}P HETCOR spectra have lower intensities than anticipated. These include the HSQC T14 H1′/C1′, G6 H1′/C1′ and UB5 base H7/C7 crosspeaks, COSY UB5 and T14 H1′–H2′ and H1′–H2″ crosspeaks, and the UB5 H3′–G6 P HETCOR crosspeak (Figure 4D, Supplementary Figures S5, S6 and S7). In all instances, the underlying coupling values responsible for the crosspeaks were identified and measured. Measured $^1J_{C-H}$ values were within anticipated ranges with the UB5 $^1J_{C7-H7}$ at 197 Hz. COSY $^3J_{H-H}$ couplings for UB5 and T14 sugar protons exhibit values anticipated for a predominant f_s conformation, and the $^3J_{H3'-P}$ for UB5 H3′–G6 P was measured at 3.85 Hz from ctNOESY experiments (Supplementary Tables S4 and S5). Taken together, these peak anomalies are suggestive of unusual dynamics.

^{13}C T_1 measurements

To gain insight into motions on the ps–ns time frame, natural abundance ^{13}C T_1 relaxation rates were measured for the UB sequence (Supplementary Figure S21). For this size molecule, lower T_1 values correlate with increased mobility. To avoid artifacts from terminal residues, the analysis was limited to the immediate bases surrounding the UB residue. At 19°C, the anomeric carbons for G4, UB5, G6 and T14
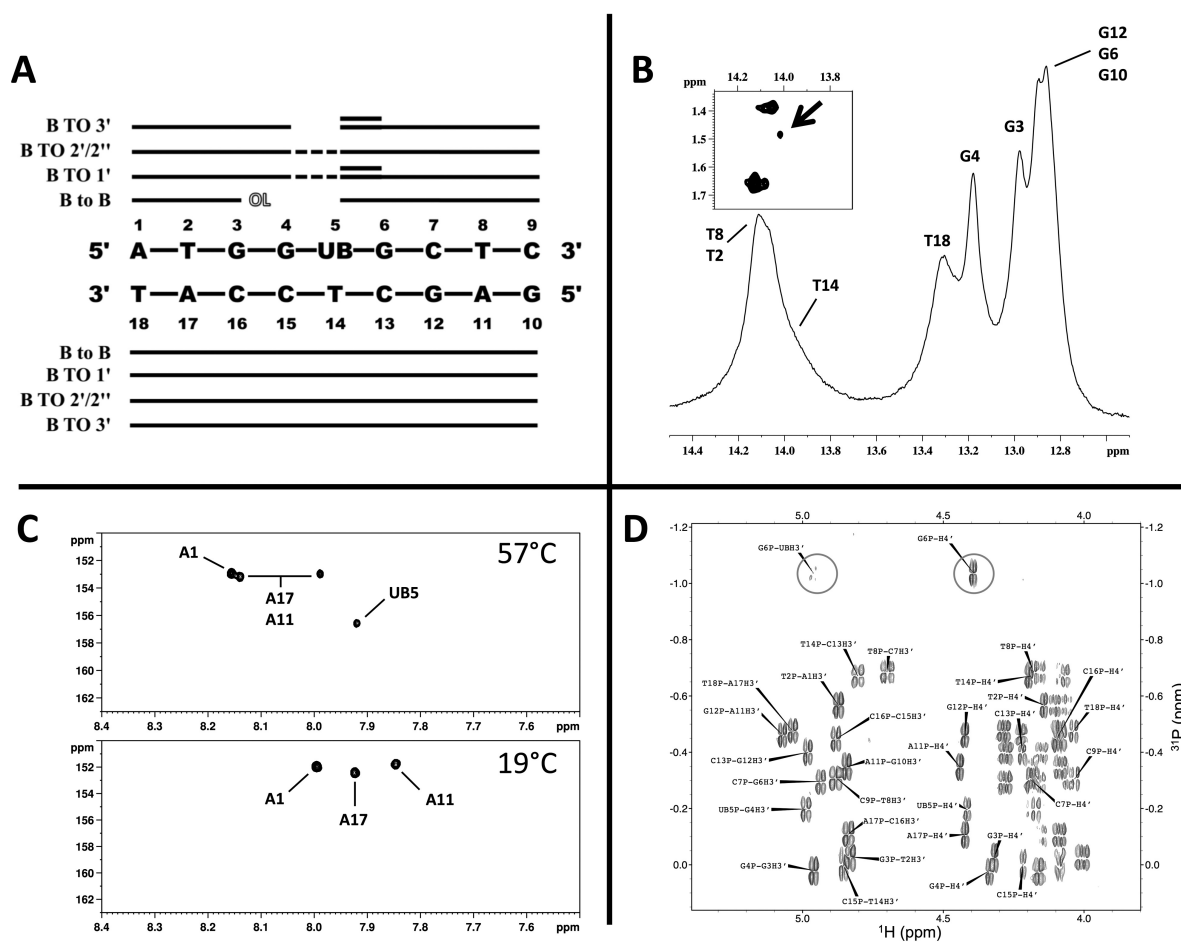
Figure 4. NOE connectivity and NMR spectra indicating unusual dynamics. (**A**) NOE pathways. Solid line signifies normal NOE contacts; dotted line = weak NOE, double line = unanticipated NOE contacts; OL = signal overlap. See text for details of unanticipated NOE contacts. (**B**) Supercooled aqueous NMR spectra of the imino proton region at −12°C showing the appearance of a 'shoulder' peak at 14 ppm. Inset: NOESY spectrum collected at −12°C showing crosspeaks between the imino proton region and methyl region. Arrow indicates the appearance of a new crosspeak. (**C**) ^{1}H-^{13}C HSQC spectra focused on the adenosine C2–H2 region; top spectrum is collected at 57°C, bottom spectrum is at 19°C. (**D**) ^{1}H–^{31}P HETCOR spectrum showing crosspeaks for H3′–P (4.6–5.1 ppm, ^{1}H) and P–H4′ (3.9–4.5 ppm, ^{1}H) (22°C). Circled are the weakened intensity for the UB5 H3′– G6 P and normal intensity G6 P–G6 H4′ crosspeaks.

are comparable with only the G4 showing a slightly reduced value (0.34 ± .03 s compared to a range of 0.38 to 0.41 s) (C13 and C15 not determined due to signal overlap). However, at 9°C, G4's anomeric T_1 value is markedly lower than the other values (0.38 ± 0.05 s compared to the others, range of 0.47 to 0.52 s). Data for the G4 and G6 base carbons could not be determined, however, C13 and C15 base carbons were evaluated revealing that C15 exhibits a somewhat lower T_1 relaxation compared to C13 (C13 = 0.35 and 0.42 s, C15 = 0.25 and 0.21 s for C5 and C6 carbons, respectively). This is consistent with the reduced anomeric T_1 value observed for G4 (C15's base pair) indicating a slight increase in mobility on the 5′ side of UB5. Of note, UB5's C7 relaxation was measured at 0.33 ± 0.03 s, which is comparable to A11's C8 (0.31 ± 0.03 s).

Molecular dynamics with time averaged restraints

In order to evaluate possible conformational sampling while still using the NMR restraints, molecular dynamics calculations using time-averaged restraints (MDtar) were per-

formed where distance restraints are enforced over a period of time as opposed to constant enforcement (51,52). This treatment allows for the evolution of dynamics when they are not readily apparent in the NMR data (implementation detail are in Supplementary Data Materials and Methods). For the control structure, the MDtar simulations do not yield any significant changes in conformations or torsion angles for A5 or the surrounding residues. In contrast, the UB MDtar simulation reveals dynamic flexibility of the backbone torsion angles at the UB5 residue with no significant changes in base orientations (Supplementary Figure S10A and B). As shown in Figure 5A, the UB5 α/γ torsion angles transition from a canonical g−/g+ (gauche−/gauche+) through a g−/t (gauche−/ trans) and ultimately to a g+/t (gauche+/ trans) conformation (labeled as populations A, B and C on Figure 5A, respectively). Concurrent with these transitions, the β torsion angle transitions from trans to gauche+ back to trans over the course of the simulation. The ε and ζ torsion angles were used to analyze B_I/B_{II} transitions. For the G4 residue, pop-
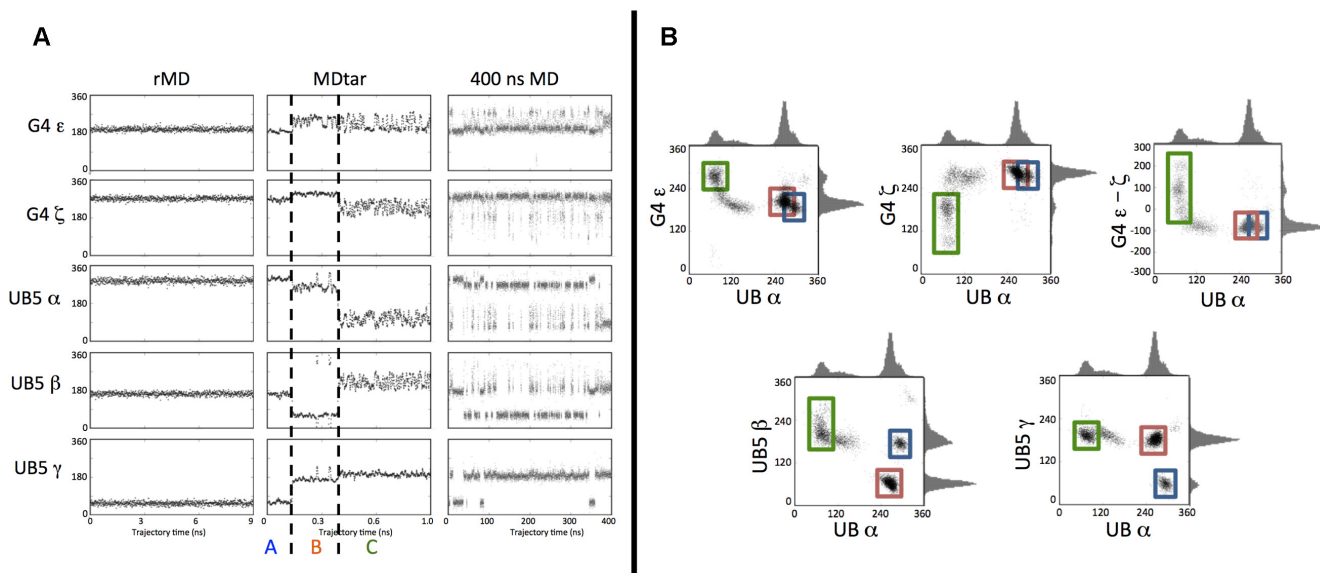
Figure 5. (**A**) G4 and UB5 backbone torsion angles versus MD trajectory time for rMD, MDtar and long term unconstrained MD simulations for the UB sequence. From top to bottom: G4 epsilon, G4 zeta, UB5 alpha, UB5 beta and UB5 gamma backbone torsion angles. From left to right: 9 ns of restrained MD, 1 ns of MD using time averaged restraints and 400 ns of unconstrained MD. Occurrences of populations A, B and C are labeled and separated with dashed lines in the MDtar graphs. For all graphs: X-axis is the trajectory time in ns and the Y-axis is torsion angle in degrees. Gauche− (g−) = −60° (i.e. 300°), gauche+ (g+) = +60° and trans (t) = 180°. (**B**) Scatter plots of torsion angles versus UB5 alpha torsion angle from the unconstrained MD trajectory. For all plots, the X axis is UB5 alpha torsion angle. Side projections are histograms for the respective torsion angle. Populations are outlined as follows: blue box = population A, orange box = population B, green box = population C.

ulations A and B are in a B_I like conformation. However, once the α/γ torsion angles adopt a g+/t geometry (population C), G4 samples the entire B_I and B_{II} conformational space. All other torsion angles surrounding the UB residue including its base pair T14 exhibit stable, canonical torsion angles. As anticipated, the AMBER restraint penalties over the course of the MDtar trajectory drop to nearly half (∼18 kcal/mol and 8 kcal/mol for the UB and control trajectories, respectively).

In order to obtain representative structures, the PDQPRO algorithm was used to compare the MDtar ensemble of structures with the NMR NOESY data (45,46). The analysis yields 5 high probability structures (individual structure probabilities are 11%, 10%, 20%, 13% and 14%) for the UB sequence; one each from clusters A and B, and 3 from cluster C. All PDQPRO selected structures have the UB5:T14 base pair in the same configuration as the solved structure with little variation in orientation (Supplementary Figure S10A and B). For the three PDQPRO structures from cluster C, none were in full canonical B_I or B_{II} conformations, but 2 exhibit B_{II} type characteristics ($\epsilon − \zeta$: 39.5° and 26.6°), and one with B_I characteristics ($\epsilon − \zeta$: −40.7°). The two other structures representing clusters A and B were in the canonical B_I range (−83.0° and −86.9°). Despite the time-averaged enforcement of distance restraints, the five structures overlay well with most conformational changes occurring in the backbone geometries (Supplementary Figure S10A, B and C). These results indicate that the enhanced backbone flexibility is unique to the UB5 residue, is not exhibited by the control A5 residue, and is supported by the NMR experimental data.

Long-term unconstrained Molecular Dynamics

Following MDtar analysis, long-term unconstrained dynamics simulations were run to predict whether the backbone conformational changes are consistently concerted, if they are long lived, and if there are other changes. The heavy atom RMSD values reveal a bimodal distribution of populations for the UB5-T14 base pair only; this is not observed for other base pairs in the UB duplex or the control duplex (Supplementary Figure S11). At the onset of the unconstrained molecular dynamics simulation, the UB5–T14 base pair moves to the predicted reverse Watson–Crick like base pair and does not sample another configuration (Figure 3C, Supplementary Figure S14B). Other than this initial shift in the UB5–T14 hydrogen bonding pattern, no other prominent changes were evident in the base orientation of the surrounding residues. Remarkably, the shift in base pair hydrogen bonding patterns only affects the conformational sampling of the backbone geometry at the UB5 residue as observed in the MDtar calculations.

The backbone at UB5 exhibits 3 distinct populations similar to those observed in the MDtar trajectory where the UB5 α, β and γ torsion angles exhibit concerted transitions (Figure 5A, Supplementary Figure S12). Correlating the UB5 α torsion angle versus other torsion angles confirms concerted backbone transitions and allows the populations to be culled into three separate populations with torsion angles similar to the MDtar results (i.e. A, B and C) (Figure 5A and B, Supplementary Figure S22, Supplementary Table S1). Interestingly, only the UB5 β torsion angle exhibits a bimodal distribution with heavy atom RMSD for the UB5:T14 base pair (Supplementary Figures S11C and S12). The α, β and γ torsion angles for the residues flanking

the UB as well as T14 and flanking residues remain in essentially static canonical torsion angles (Supplementary Figure S12). Population B is the most frequently sampled conformation in the MD simulation representing ∼50% of the total trajectory (Supplementary Table S1). Unconstrained MD simulations with the control sequence do not reveal any significant changes in conformation at the A5 residue emphasizing that the observed results for the UB sequence are unique to the UB5 (Supplementary Figure S12).

Analysis of the UB sequence G4's ϵ and ζ torsion angles reveals that populations A and B are in a B_I conformation, while C samples both B_I and B_{II} space with B_{II} sampling occurring 14.1% of the trajectory. Within the C population, lower values of UB5 α ($\alpha < 100°$) tend favor a B_{II} like conformation while higher values favors a B_I like conformation (Figure 5B, Supplementary Figure S13). Analysis of T14's ϵ and ζ torsion angles reveal that it is in a B_{II} conformation 8.6% of the trajectory. However, unlike G4, the sampling appears to be independent from the UB5 torsion angle changes (Supplementary Figure S13). Surprisingly, despite the heightened G4 ϵ and ζ and UB5 α, β and γ flexibility, the UB5 ϵ and ζ torsion angles are quite static and remain in the B_I conformation (UB5 = 0% B_{II}) (Supplementary Figures S12 and S13).

DISCUSSION

To determine how the UB is situated in duplex DNA, solution NMR and MD calculations were used to probe the structure and dynamics of the UB in an oligonucleotide. The standard restrained MD NMR procedure resulted in a single high-resolution solution structure that maintains a remarkably unaltered B-type helical geometry with the UB base well accommodated, intrahelical, involved in base stacking, yet exhibiting an unusual base pair conformation. Further probing with supercooled aqueous NMR provided evidence of an alternative UB5:T14 base pair conformation correlating to the previously predicted and energetically favorable configuration. Additionally, several diminished 2D NMR crosspeak intensities provide evidence that the UB duplex exhibits unusual dynamics on the NMR time-scale (Figure 4B, D, Supplementary Figures S5, S6 and S7). Notwithstanding these indications of dynamics, all observed 1D resonances that could be identified exhibit sharp peaks (the only exceptions being the missing UB H2 and T14 imino proton resonances), all NOESY and TOCSY crosspeaks were assigned, and no exchange peaks were observed (Supplementary Figure S1). Although quite low CORMA Rx values were obtained for both structures, the higher AMBER distance restraint penalties and deviations for the UB structure compare to the control indicate an increase in dynamics (Table 2).

Of the weak 2D crosspeaks intensities, the low intensity of UB5 H3'–G6 P HETCOR crosspeak is insightful. A reduced H3'–P crosspeak intensity can be the result of small $^3J_{H3'-P}$ values, yet for the UB sequence, the $^3J_{H3'-P}$ coupling values were measured and are in the anticipated values for B-type DNA (UB5 H3'–G6 P $^3J_{H3'-P}$ = 3.8 Hz). Thus, the weak crosspeak is not a result of a small $^3J_{H3'-P}$ coupling, but due to backbone dynamics (Supplementary Table S5).

The rMD NMR solved structure conforms to the NMR data and does not exhibit altered backbone torsion angles for the UB or any of the surrounding residues despite the fact that the backbone torsion angles for the UB (aside from the experimentally determined ϵ and sugar puckering) were unconstrained during structure calculations (Supplementary Figure S3). MDtar calculations were utilized to probe for conformational changes that may be contained in the NMR data but not readily apparent. The calculations reveal concerted α/γ and B_I/B_{II} sampling in the backbone torsion angles around G4 and UB5 that did not manifest themselves in the conventional structure calculation protocol. This sampling was also observed in the long term unconstrained MD simulations. In contrast, the control sequence does not show any significant deviations in the backbone configuration in either simulation indicating the effect originated from the UB.

These predicted motions are supportive of the diminished UB5 COSY sugar proton and UB5 H3'–P HETCOR crosspeaks. Additionally, the B_{II} torsion angle populations for UB5 and T14 predicted by the long-term molecular dynamics simulations are consistent with the ^{31}P chemical shift resonances. ^{31}P resonances are partially dependent on B_I/B_{II} populations with downfield shifts generally attributed to a higher B_{II} population (53–56). In the molecular dynamics simulation, UB5-P-G6 exhibits 0% B_{II} population whereas the control sequence's A5-P-G6 sampled the B_{II} conformational space 4% of the trajectory. This trend is consistent with the upfield shifted ^{31}P resonance of −1.04 ppm (Δppm = −0.52 ppm as compared to the control sequence). Similarly, the T14-P-C15 of the UB sequence samples the B_{II} population 8.6% compared to 0.7% in the control sequence; this again is consistent with the downfield shifted signal (Δppm = +0.56 ppm as compared to the control. Although the 31P resonance for G4-P-UB5 does not deviate much from the control value despite the enhanced G4 BII sampling found in the molecular dynamics simulations, the complex motions exhibited in the UB5 α, β and γ torsion angles may offset the impact on the 31P data) (Figure 2B, Supplementary Figure S13).

Although most of the imino proton resonances were overlapped and prevented a direct comparison of residues, the base pair opening analysis does suggest a heightened base opening rate on the 5' side of the UB (G4) compared to the 3' side: ∼3 ms for the UB sequence G4:C16 compared to 40–50 ms for the control sequence G6:C13. For the control sequence, the G4 imino proton signal recovers similarly to G6 (Supplementary Figure S20) suggesting similar base pair lifetimes, however, this was not observed for the UB sequence where G4 appears to recover more quickly than the G6/G12 peak (Supplementary Figure S19) suggesting that G4 has a shorter base pair lifetime. This trend is also in agreement with the ^{13}C T_1 analysis revealing faster dynamics on the 5' side. Backbone flexibility on the 5' side of the UB may in part be lending to the accommodation of UB insertion into the duplex environment.

The most glaring anomaly exhibited in the NMR data, however, is the missing UB H2 resonance, which defied detection in all NMR spectra under annealed conditions, yet could be readily observed under denaturing conditions and for the single strand (Figure 4C). The UB5 H2 proton is the

only signal that exhibits this behavior. Because only a single base proton is affected, this behavior cannot be explained solely by the heightened backbone torsion angle sampling.

The supercooled aqueous NMR data provide evidence of a weak base pair with corresponding higher opening rates. The 'missing' UB H2 proton resonance can then be rationalized in context of the surrounding structure and chemical exchange. The 'anti' conformation of the UB and the sequence context of G4-UB5-G6 places the H2 proton in the major groove flanked by two carbonyl groups. Carbonyl shielding effects can cause quite dramatic chemical shift changes of protons [57].

With this, the data support at least 2 different UB5:T14 base pair configurations (Figures 3A, C, 4B and Supplementary Figure S14B). These expose the UB5 H2 proton to different environments from the adjacent carbonyls' shielding cones resulting in chemical exchange for the UB5 H2 proton (Supplementary Figure S14A). The appearance of resonances in chemical exchange is dependent on the difference in chemical shifts and the time-scale of the motion. The fact that the H2 peak disappears from NMR spectra can provide limits on the time frame of the motion.

The broadest a NMR resonance can be due to chemical exchange is at the coalescence point, which is determined by the chemical shifts difference of the proton in the two environments:

$$k_c = (\pi \; \Delta v) / 2^{1/2} \qquad (1)$$

where $\Delta\nu$ is the difference in chemical shift (Hz) and k_c is the exchange rate (s^{-1}) at the coalescence point. The difference in chemical shifts must be relatively large as smaller differences would have resulted in an observable broad peak. Because there are no observable NMR resonances under annealed conditions for the UB H2 proton and our NMR and MD calculations indicate two possible orientations, a simple two state model (Equation 1) was used to estimate the timescale of the coalescence point. A symmetrical exchange was assumed since the absence of any signals did not allow us to estimate populations (Supplementary Figures S15 and S16).

Chemical shift differences were calculated using the NMR solved structure and the unconstrained MD model, which was supported by imino proton data collected under supercooled aqueous conditions. In addition, a third theoretical conformation where UB5 was flipped out from the flanking G4 and G6 bases was also used representing an extreme structural limit [57,58].

The difference in calculated UB H2 chemical shifts from the two NMR and MD base conformations is estimated to be 450 Hz resulting in a lifetime ($1/k_c$) of ∼1 ms. Considering the completely flipped out model the difference in chemical shifts is ∼700 Hz that reduces the lifetime to ∼0.6 ms. Compared to the experimentally determined A5:T14 base pair lifetime of 3 to 5 ms, these theoretical values are not unreasonable given the inability to determine the T2 and T11 base opening rates and the ability to slow dynamics of the UB5–T14 base pair under supercooled aqueous conditions [48–50,59]. Further support for the predicted timescale comes from the reduced HSQC intensities of UB5, G6 and T14 (Supplementary Figures S6 and S7). Comparison of HSQC peak intensities have been found to

give qualitative information of local dynamics on the µs to ms time frame and has been used to characterize broad protein motions and protein–protein interactions [60,61]. Additionally, spectra obtained under different NMR field strengths (400, 500 and 600 MHz) did reveal any signals for the UB H2; this was in line with our simulated spectra as well as our estimates above (Supplementary Figures S15 and S16).

The UB's impact on the dynamics and geometry of the DNA duplex are somewhat similar to that of DNA mismatches. As with many mismatches and other lesions, the DNA B-type conformation is robust and accommodates the UB:dT inclusion with little impact on the global geometry showing some similarity to mismatches with an observed increased base opening. The presence of a mismatch in a DNA duplex enhances the dynamics of the mismatch and surrounding residues similar to our findings. However, unlike many mismatches, there is not an observed transferred distortion of helical parameters to other residues outside of the UB5:T14 base pair; the UB disruption to the conformation of the duplex is confined to that base pair with flanking residues quite comparable to the control structure (Figure 3B) [48,62].

Additionally, the UB presents a similar reduction in thermostability as that of the 3-nitropyrrol and 5-nitrolindole universal bases. Both the UB and the nitropyrrol and nitrolindole gain thermal stability through conformational sampling yet achieve this in different ways. The nitropyrrol and nitrolindole bases are not involved in hydrogen bonding with their base pairs, but stabilize via a dynamic intercalation with the base pair causing an increase in rise between both the 3' and 5' flanking residues [16]. In contrast, the UB presents a less disruptive effect on local base conformation, yet exhibits a heightened breathing rate and dynamic sampling in the phosphodiester backbone.

Taken together, the UB:dT base pair is accommodated within the duplex via entropic freedom in the backbone and base pairing dynamics, yet remarkably does not overly impact the rest of the DNA duplex or the overall B-type conformation about the UB. With the insight of the UB:dT duplex, the ΔH and ΔS data implies a similar stabilization scheme for the UB opposite the dG and dA. Intriguingly, the UB opposite dC is more enthalpically favored suggesting a more rigid structure, and should be characterized in future studies. Moreover, based on our solved structure, the UB's unique N8 connection and 'anti' orientation presents a quite different landscape of major and minor groove binding determinants with respects to canonical G-C and A-T base pairs. This unique presentation may be an attribute that could be exploited with custom groove binders. Our data show that the UB remains a viable universal base candidate.

ACCESSION CODES

Coordinates and NMR data have been deposited in the PDB and BMRB under accession codes 2n5o and 2n5p for the UB and control structures, respectively (BRMB codes: 25723 and 25724 for the UB and control, respectively).

REFERENCES

1. Nakano,S., Fujii,M. and Sugimoto,N. (2011) Use of nucleic Acid analogs for the study of nucleic Acid interactions. *J. Nucleic Acids*, **2011**, 967098.
2. Hirao,I., Kimoto,M. and Yamashige,R. (2012) Natural versus artificial creation of base pairs in DNA: Origin of nucleobases from the perspectives of unnatural base pair studies. *Acc. Chem. Res.*, **45**, 2055–2065.
3. Kimoto,M., Kawai,R., Mitsui,T., Yokoyama,S. and Hirao,I. (2008) Efficient PCR amplification by an unnatural base pair system. *Nucleic Acids Symp. Ser. (Oxf)*, 469–470.
4. Chiaramonte,M., Moore,C.L., Kincaid,K. and Kuchta,R.D. (2003) Facile polymerization of dNTPs bearing unnatural base analogues by DNA polymerase α and Klenow fragment (DNA polymerase I). *Biochemistry*, **42**, 10472–10481.
5. Loakes,D., Van Aerschotl,A., Brown,D.M. and Hill,F. (1996) Enzymatic recognition of acyclic universal base analogues in oligonucleotides. *Nucleosides Nucleotides Nucleic Acids*, **15**, 1891–1904.
6. Too,K. and Loakes,D. (2008) Universal base analogues and their applications to biotechnology. In: Herdewijn,P (ed). *Modified Nucleosides and Medicine*. Wiley-VCH Verlag GmbH Co. KGaA, Weinheim, pp. 277–303.
7. Smith,C.L., Simmonds,A.C., Felix,I.R., Hamilton,A.L., Kumar,S., Nampalli,S., Loakes,D., Hill,F. and Brown,D.M. (1998) DNA polymerase incorporation of universal base triphosphates†. *Nucleosides Nucleotides*, **17**, 541–554.
8. Sengupta,S., Onodera,K., Lai,A. and Melcher,U. (2003) Molecular detection and identification of influenza viruses by oligonucleotide microarray hybridization. *J. Clin. Microbiol.*, **41**, 4542–4550.
9. Guschin,D.Y., Mobarry,B.K., Proudnikov,D., Stahl,D.A., Rittmann,B.E. and Mirzabekov,A.D. (1997) Oligonucleotide microchips as genosensors for determinative and environmental studies in microbiology. *Appl. Environ. Microbiol.*, **63**, 2397–2402.
10. Ryabinin,V.A., Kostina,E.V., Maksakova,G.A., Neverov,A.A., Chumakov,K.M. and Sinyakov,A.N. (2011) Universal oligonucleotide microarray for sub-typing of Influenza A virus. *PLoS One*, **6**, e17529.
11. Pinheiro,A.V., Han,D., Shih,W.M. and Yan,H. (2011) Challenges and opportunities for structural DNA nanotechnology. *Nat. Nanotechnol.*, **6**, 763–772.
12. Loakes,D. (2001) Survey and summary: The applications of universal DNA base analogues. *Nucleic Acids Res.*, **29**, 2437–2447.
13. Liang,F., Liu,Y.-Z. and Zhang,P. (2013) Universal base analogues and their applications in DNA sequencing technology. *RSC Adv.*, **3**, 14910–14928.
14. Loakes,D., Hill,F., Brown,D.M. and Salisbury,S.A. (1997) Stability and structure of DNA oligonucleotides containing non-specific base analogues. *J. Mol. Biol.*, **270**, 426–435.
15. Martin,F.H., Castro,M.M., Aboul-ela,F. and Tinoco,I. Jr (1985) Base pairing involving deoxyinosine: Implications for probe design. *Nucleic Acids Res.*, **13**, 8927–8938.
16. Gallego,J. and Loakes,D. (2007) Solution structure and dynamics of DNA duplexes containing the universal base analogues 5-nitroindole and 5-nitroindole 3-carboxamide. *Nucleic Acids Res.*, **35**, 2904–2912.
17. He,J. and Seela,F. (2003) Oligonucleotides incorporating 8-aza-7-deazapurines: Synthesis and base pairing of nucleosides with nitrogen-8 as a glycosylation position. *Org. Biomol. Chem.*, **1**, 1873–1883.
18. Seela,F. and Debelak,H. (2000) The N(8)-(2′-deoxyribofuranoside) of 8-aza-7-deazaadenine: a universal nucleoside forming specific hydrogen bonds with the four canonical DNA constituents. *Nucleic Acids Res.*, **28**, 3224–3232.
19. Seela,F. and Kaiser,K. (1988) 8-Aza-7-deazaadenine N8- and N8-(β-D-2′-Deoxyribofuranosides): Building blocks for automated DNA synthesis and properties of oligodeoxyribonucleotides. *Helv. Chim. Acta*, **71**, 1813–1823.
20. Seela,F., Jawalekar,A.M., Sun,L. and Leonard,P. (2005) Oligonucleotides containing pyrazolo[3,4-d]pyrimidines: 8-Aza-7-deazaadenines with bulky substituents in the 2- or 7-position. *Nucleosides Nucleotides Nucleic Acids*, **24**, 1485–1505.
21. Seela,F., Zulauf,M. and Debelak,H. (2000) Base-Pairing Properties of 8-aza-7-deazaadenine Linked via the 8-Position to the DNA Backbone. *Helv. Chim. Acta*, **83**, 1437–1453.
22. Zhang,X., Budow,S., Leonard,P., Eickmeier,H. and Seela,F. (2006) N8-(2′-O-methylribofuranosyl)-8-aza-7-deazaadenine monohydrate. *Acta Crystallogr. C*, **62**, o79–o81.
23. Aramini,J.M. and Germann,M. (1999) Solution structure of a DNA⊙ RNA hybrid containing an α-anomeric thymidine and polarity reversals: d (ATGG-3′-3′-αT-5′-5′-GCTC)⊙ r (gagcaccau). *Biochemistry*, **38**, 15448–15458.
24. Plateau,P. and Gueron,M. (1982) Exchangeable proton NMR without base-line distortion, using new strong-pulse sequences. *J. Am. Chem. Soc.*, **104**, 7310–7311.
25. Schroeder,K.T., Skalicky,J.J. and Greenbaum,N.L. (2005) NMR spectroscopy of RNA duplexes containing pseudouridine in supercooled water. *RNA*, **11**, 1012–1016.
26. Szyperski,T. and Mills,J.L. (2011) NMR-based structural biology of proteins in supercooled water. *J. Struct. Funct. Genomics*, **12**, 1–7.
27. Aramini,J.M., Cleaver,S., Pon,R., Cunningham,R. and Germann,M. (2004) Solution structure of a DNA duplex containing an [alpha]-anomeric adenosine: Insights into substrate recognition by endonuclease IV. *J. Mol. Biol.*, **338**, 77–91.
28. Johnson,C.N., Spring,A.M., Sergueev,D., Shaw,B.R. and Germann,M.W. (2011) Structural basis of the RNase H1 activity on stereo regular borano phosphonate DNA/RNA hybrids. *Biochemistry*, **50**, 3903–3912.
29. Johnson,C.N., Spring,A.M., Desai,S., Cunningham,R.P. and Germann,M.W. (2012) DNA sequence context conceals alpha-anomeric lesions. *J. Mol. Biol.*, **416**, 425–437.
30. Wu,Z., Tjandra,N. and Bax,A. (2001) Measurement of 1H3′-31P dipolar couplings in a DNA oligonucleotide by constant-time NOESY difference spectroscopy. *J. Biomol. NMR*, **19**, 367–370.
31. Zweckstetter,M. and Bax,A. (2001) Characterization of molecular alignment in aqueous suspensions of Pf1 bacteriophage. *J. Biomol. NMR*, **20**, 365–377.
32. Case,D.A., Cheatham,T.E. 3rd, Darden,T., Gohlke,H., Luo,R., Merz,K.M. Jr, Onufriev,A., Simmerling,C., Wang,B. and Woods,R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.
33. Pearlman,D.A., Case,D.A., Caldwell,J.W., Ross,W.S., Cheatham,T.E. III, DeBolt,S., Ferguson,G., Seibel,G. and Kollman,P. (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp. Phys. Commun.*, **91**, 1–41.
34. Borgias,B.A. and James,T.L. (1988) COMATOSE, a method for constrained refinement of macromolecular structure based on two-dimensional nuclear Overhauser effect spectra. *J. Magn. Reson.*, **79**, 493–512.
35. Borgias,B.A. and James,T.L. (1990) MARDIGRAS-A procedure for matrix analysis of relaxation for discerning geometry of an aqueous structure. *J. Magn. Reson*, **87**, 475–487.

36. Keepers,J.W. and James,T.L. (1984) A theoretical study of distance determinations from NMR. Two-dimensional nuclear overhauser effect spectra. *J. Magn. Reson.*, **57**, 404–426.

37. Thomas,P.D., Basus,V.J. and James,T.L. (1991) Protein solution structure determination using distances from two-dimensional nuclear Overhauser effect experiments: effect of approximations on the accuracy of derived structures. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 1237–1241.

38. James,T.L. (1991) Relaxation matrix analysis of two-dimensional nuclear Overhauser effect spectra. *Curr. Opin. Struct. Biol.*, **1**, 1042–1053.

39. Liu,H., Spielmann,H.P., Ulyanov,N.B., Wemmer,D.E. and James,T.L. (1995) Interproton distance bounds from 2D NOE intensities: Effect of experimental noise and peak integration errors. *J. Biomol. NMR*, **6**, 390–402.

40. Cheatham,T.E. 3rd, Cieplak,P. and Kollman,P.A. (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.*, **16**, 845–862.

41. Lavery,R., Moakher,M., Maddocks,J.H., Petkeviciute,D. and Zakrzewska,K. (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.*, **37**, 5917–5929.

42. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.

43. Couch,G.S., Hendrix,D.K. and Ferrin,T.E. (2006) Nucleic acid visualization with UCSF Chimera. *Nucleic Acids Res.*, **34**, e29.

44. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.

45. Aramini,J.M., Mujeeb,A., Ulyanov,N.B. and Germann,M.W. (2000) Conformational dynamics in mixed alpha/beta-oligonucleotides containing polarity reversals: a molecular dynamics study using time-averaged restraints. *J. Biomol. NMR*, **18**, 287–302.

46. Ulyanov,N.B., Schmitz,U., Kumar,A. and James,T.L. (1995) Probability assessment of conformational ensembles: sugar repuckering in a DNA duplex in solution. *Biophys J.*, **68**, 13–24.

47. Zgarbová,M., Otyepka,M., Šponer,J., Lankaš,F. and Jurečka,P. (2014) Base pair fraying in molecular dynamics simulations of DNA and RNA. *J. Chem. Theory Comput.*, **10**, 3177–3189.

48. Mazurek,A., Johnson,C., Germann,M. and Fishel,R. (2009) Sequence context effect for hMSH2-hMSH6 mismatch-dependent activation. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 4177–4182.

49. Gueron,M. and Leroy,J.L. (1995) Studies of base pair kinetics by NMR measurement of proton exchange. *Methods Enzymol.*, **261**, 383–413.

50. Moe,J.G. and Russu,I.M. (1992) Kinetics and energetics of base-pair opening in 5′-d(CGCGAATTCGCG)-3′ and a substituted dodecamer containing G.T mismatches. *Biochemistry*, **31**, 8421–8428.

51. Pearlman,D.A. and Kollman,P.A. (1991) Are time-averaged restraints necessary for nuclear magnetic resonance refinement? A model study for DNA. *J. Mol. Biol.*, **220**, 457–479.

52. Torda,A.E., Scheek,R.M. and van Gunsteren,W.F. (1990) Time-averaged nuclear Overhauser effect distance restraints applied to tendamistat. *J. Mol. Biol.*, **214**, 223–235.

53. Gorenstein,D.G. (1992) 31P NMR of DNA. *Methods Enzymol.*, **211**, 254–286.

54. Gorenstein,D.G., Lai,K. and Shah,D.O. (1984) 31P and two-dimensional 31P/1H correlated NMR spectra of Duplex d(Ap[17O]Gp[18O]Cp[16O]T) and assignment of 31P signals in d(ApGpCpT)2-actinomycin D complex. *Biochemistry*, **23**, 6717–6723.

55. Precechtelova,J., Novak,P., Munzarova,M.L., Kaupp,M. and Sklenar,V. (2010) Phosphorus chemical shifts in a nucleic acid backbone from combined molecular dynamics and density functional calculations. *J. Am. Chem. Soc.*, **132**, 17139–17148.

56. Heddi,B., Foloppe,N., Bouchemal,N., Hantz,E. and Hartmann,B. (2006) Quantification of DNA BI/BII backbone states in solution. Implications for DNA overall structure and recognition. *J. Am. Chem. Soc.*, **128**, 9170–9177.

57. Martin,N.H., Allen,N.W. 3rd, Brown,J.D., Kmiec,D.M. Jr and Vo,L. (2003) An NMR shielding model for protons above the plane of a carbonyl group. *J. Mol. Graph. Model.*, **22**, 127–131.

58. Martin,N.H., Loveless,D.M., Main,K.L. and Pyles,A.K. (2006) Computation of through-space NMR shielding effects by functional groups common to peptides. *J. Mol. Graph. Model.*, **25**, 1–9.

59. Nonin,S., Leroy,J.L. and Gueron,M. (1995) Terminal base pairs of oligodeoxynucleotides: imino proton exchange and fraying. *Biochemistry*, **34**, 10652–10659.

60. Granovsky,A.E., Clark,M.C., McElheny,D., Heil,G., Hong,J., Liu,X., Kim,Y., Joachimiak,G., Joachimiak,A., Koide,S. *et al.* (2009) Raf kinase inhibitory protein function is regulated via a flexible pocket and novel phosphorylation-dependent mechanism. *Mol. Cell. Biol.*, **29**, 1306–1320.

61. Smrcka,A.V., Kichik,N., Tarrago,T., Burroughs,M., Park,M.S., Itoga,N.K., Stern,H.A., Willardson,B.M. and Giralt,E. (2010) NMR analysis of G-protein betagamma subunit complexes reveals a dynamic G(alpha)-Gbetagamma subunit interface and multiple protein recognition modes. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 639–644.

62. Rossetti,G., Dans,P.D., Gomez-Pinto,I., Ivani,I., Gonzalez,C. and Orozco,M. (2015) The structural impact of DNA mismatches. *Nucleic Acids Res.*, **43**, 4309–4321.