Supporting Information

Assessing the contribution of rare DNA states to cancer mutational signatures using sequence-specific conformational fingerprinting

Or Szekely[1]*, Yeongjoon Lee[2], Atul K. Rangadurai[3], Serafima Guseva[2], Joshua Cooksey[2], Edgar M. Faison[1], Nikita Zalenski[4], Qi Zhang[1], Zucai Suo[4], and Hashim M. Al-Hashimi[2]*

[1]*Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA*

[2]*Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA*

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

[4]*Department of Biomedical Sciences, Florida State University College of Medicine, Tallahassee, FL 32306, USA*

**To whom correspondence should be addressed:*

Email: ha2639@cumc.columbia.edu

Email: or_szekely@med.unc.edu

Keywords Nucleic acid dynamics; cancer mutational signatures; deprotonation; wobble mismatches; chemical exchange; ^{19}F NMR; pK_a.

Supporting Discussion

Supporting Discussion 1: Analysis of sequence-dependent chemical shifts

H1 and H3 in G•T and G•^{5F}dU: In the G•T and G•^{5F}dU wobble pairs, the H3 imino proton of T/^{5F}dU experiences an upfield shift when positioned beneath a 5' purine neighbor, due to ring current induced shielding that reduces the local magnetic field shifting the resonance upfield (Fig. S3A, top). Conversely, when the 5' neighbor is a pyrimidine, H3 residues near the edge of the aromatic ring, leading to a downfield shift as ring currents increase the local magnetic field shifting the resonance downfield (Fig. S3A, top). The guanine H1 imino proton exhibits the opposite pattern: it shifts downfield when the 5' neighbor of T/^{5F}dU is a purine because it resides near the edge of the partner pyrimidine and upfield when the 5' neighbor of T/^{5F}dU is a pyrimidine since it resides beneath its 5' purine neighbor. For both T/^{5F}dU(H3) and G(H1) protons, the 3' neighbor has little effect on the chemical shift, as each proton always lies at the periphery of a purine base (Fig. S3A, bottom).

F5 in *G•^{5F}dU:* Similar to ^{5F}dU(H3), the F5 chemical shift in the G•^{5F}dU wobble primarily depends on the identity of the 5' neighbor (Fig. S4A) but exhibits a weaker and opposite sequence dependence, shifting downfield when the 5' neighbor is a purine (p = 0.02). This reversed trend arises from distinct ring currents at the F5 position (Fig. S4A): whereas H3 resides beneath adjacent bases and experiences shielding, the F5 lies at their periphery, where stronger purine ring currents induce deshielding and consequently larger downfield shifts. A weaker dependence is observed with the 3' neighbor, where F5 shifts slightly downfield in the presence of purines (p = 0.05), likely reflecting greater spatial separation between the 3' base and the F5 atom.

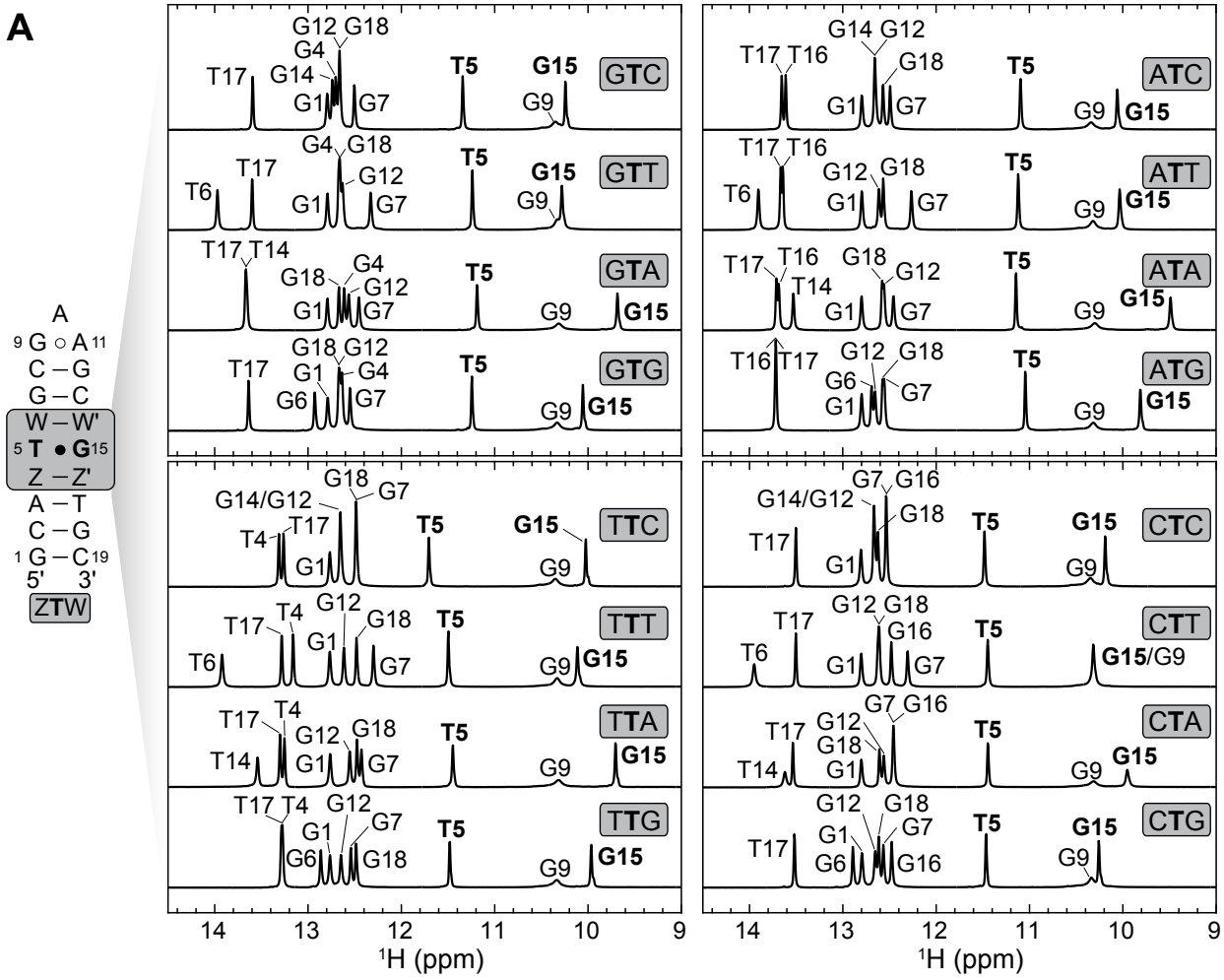
F5 in anionic *G•^{5F}dU^{-}:* The F5 chemical shift in anionic G•^{5F}dU^{-} strongly depends on the 5' neighbor of ^{5F}dU (Fig. S4B), shifting upfield when the 5' neighbor is a purine (p = 0.001). This

trend cannot readily be explained by ring currents in a B-form helix (Fig. S4B) and likely reflects additional contributions from deprotonation and possibly conformational changes in neighboring base pairs (bps). In an ideal Watson-Crick geometry, F5 lies near the periphery of the 5' neighboring base irrespective of its identity; the larger purine base would normally cause greater ring current deshielding producing a downfield shift. The observed upfield shift therefore suggests a structural or electrostatic perturbation unique to the anionic state. Because the anionic WC-like G•^{5F}dU^{-} carries a negative charge on the ^{5F}dU(N3), it might also modify stacking interactions with the 5' purine causing it to shift towards the fluorine atom, altering the ring currents experienced by the ^{5F}dU(F5) and leading to the observed upfield chemical shift (see Fig. S4C).

Supporting Discussion 2: Fingerprinting cancer mutational signatures using sequence-specific G•T^{-}

When G serves as the template during replication i.e. G(template)•dTTP, it is flanked by 5' and 3' neighboring bases, whereas incoming dTTP only experiences its 5' neighbor. The position normally occupied by the 3' neighbor is instead occupied by a polymerase side chain typically a tyrosine or phenylalanine residue that stacks against the nascent bp. Conversely, when T is the template (T(template)•dGTP), the templating T experiences both its 5' and 3' neighbors, whereas the incoming dGTP only experiences its 5' neighbor. Thus, for signatures dominated by nucleotide misincorporation (e.g. SBS14), we would expect the G•T^{-} fingerprint to manifest as sequence-specific misincorporation of T(template)•dGTP, giving rise to T>C substitutions, noting that T>C will also have contributions from A(template)•dCTP misincorporation which cannot currently be accounted for in this work. The sequence-specificity of dG(template)•dTTP is harder to predict without fingerprints measured within the polymerase context. Nevertheless, a sequence dependence similar to the G•T^{-} fingerprint could still emerge due to extension past the lesion, when the 3' neighbor comes into play, or through proofreading and repair processes[1-3].

Supporting Figures



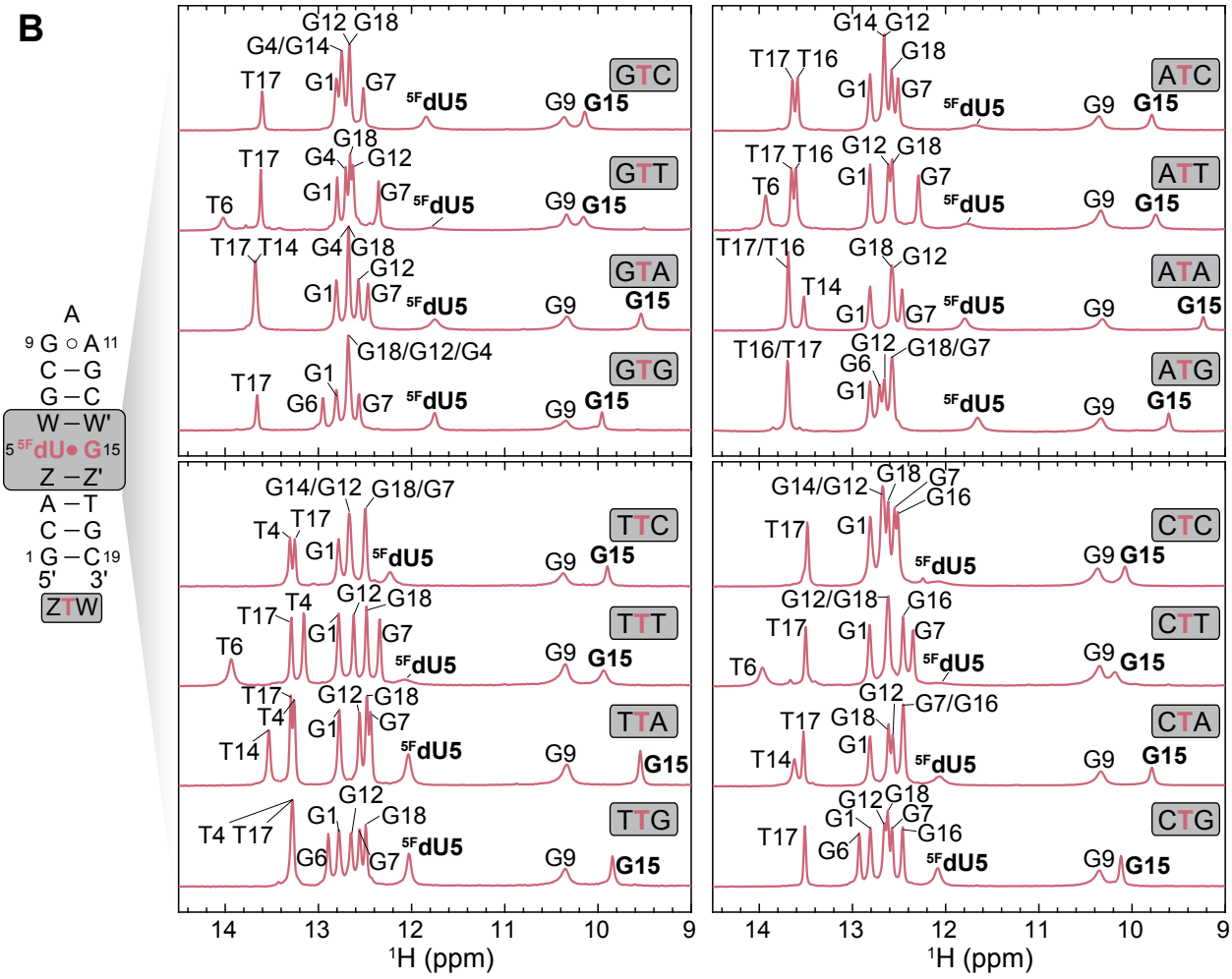
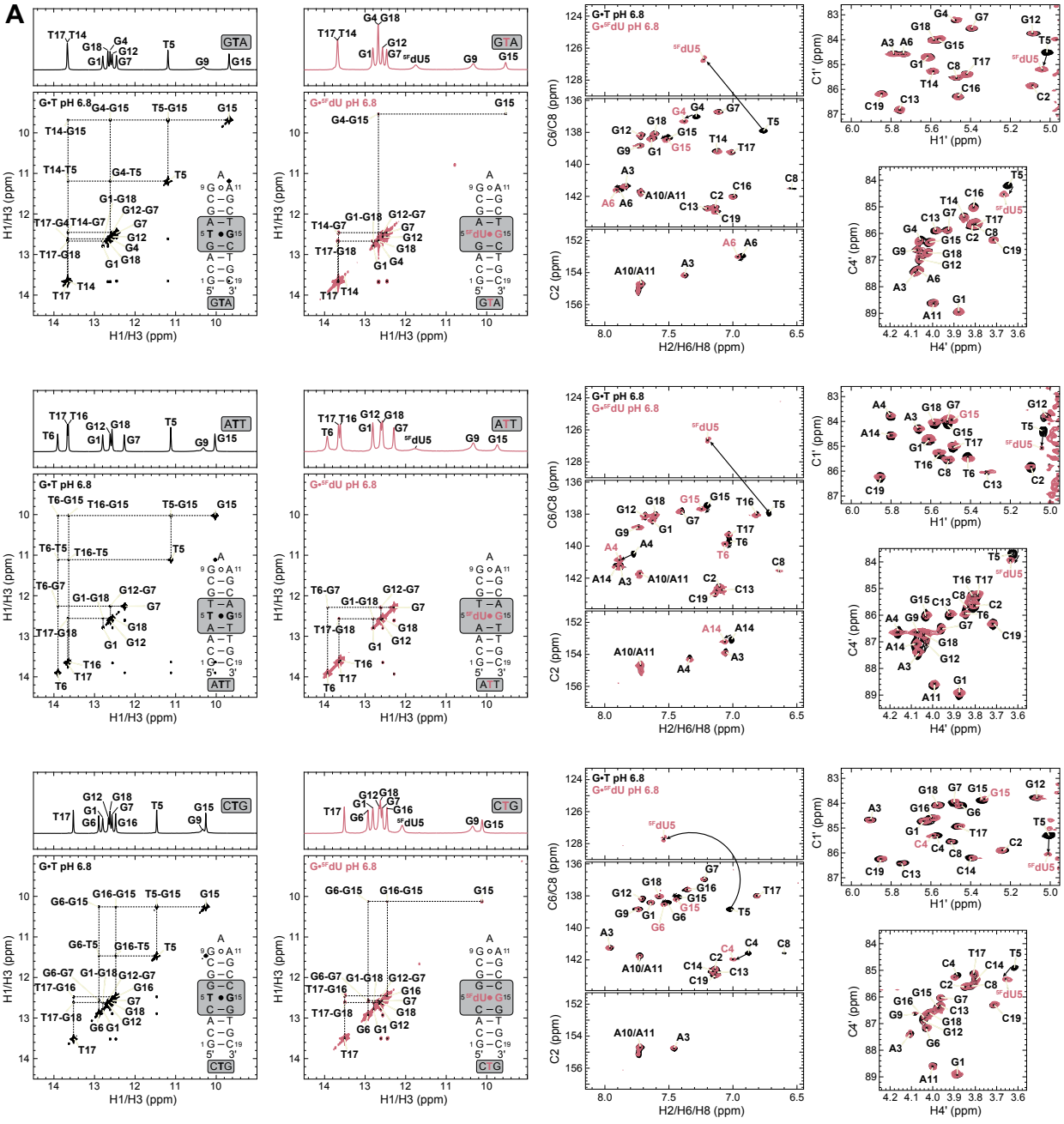
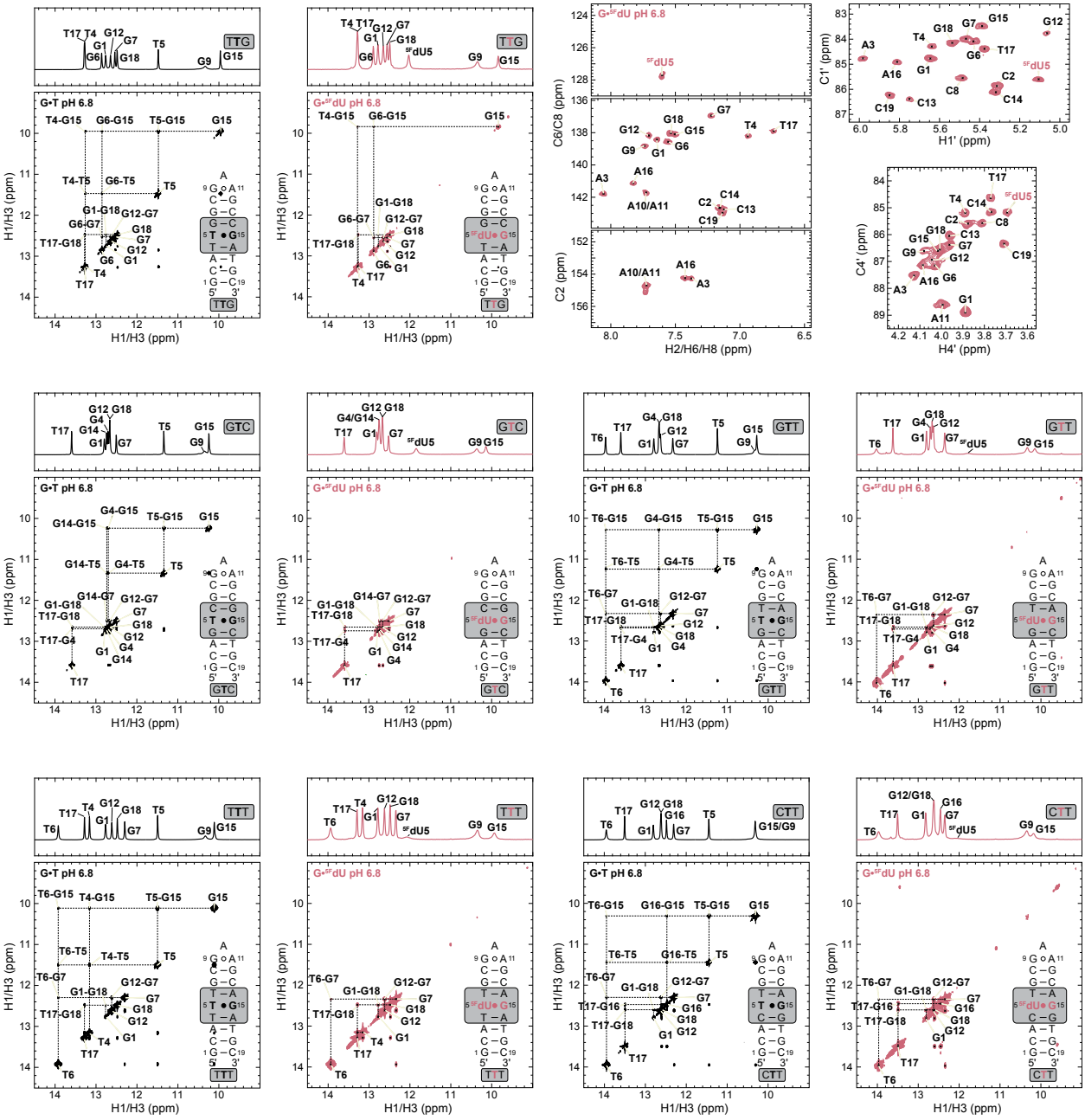
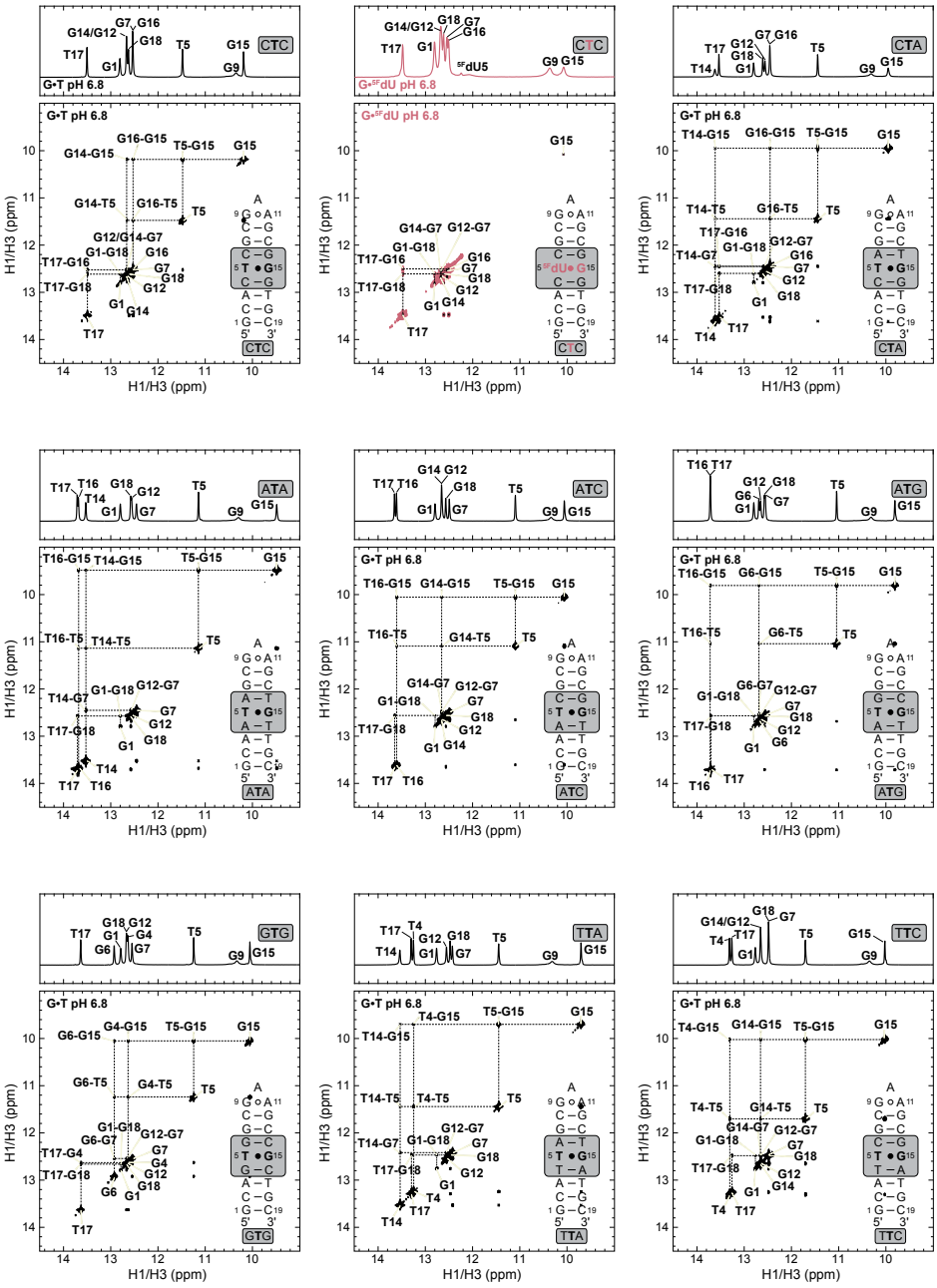


Figure S1. Sequence dependence of imino ^{1}H chemical shifts in G•T and G•^{5F}dU wobble pairs. 1D ^{1}H spectra showing sequence-dependent changes in the imino proton chemical shifts for (A) unmodified and (B) 5F modified duplexes. Shown on the left are the secondary structures of the 16 DNA hairpin constructs, with varying trinucleotide sequence contexts. Z-Z' and W-W' represent the four Watson-Crick (A-T, T-A, G-C and C-G) neighbors.


A
GTA
G•T pH 6.8
G•5FdU pH 6.8
H1/H3 (ppm)
H2/H6/H8 (ppm)
C6/C8 (ppm)
C2 (ppm)
H1' (ppm)
C1' (ppm)
H4' (ppm)
C4' (ppm)
ATT
CTG


TTG
G•T pH 6.8
G•5FdU pH 6.8
H1/H3 (ppm)
H2/H6/H8 (ppm)
H1' (ppm)
H4' (ppm)
C6/C8 (ppm)
C2 (ppm)
C1' (ppm)
C4' (ppm)
GTC
GTT
TTT
CTT
CTT


CTC
G•T pH 6.8
G•5FdU pH 6.8
CTA
ATA
ATC
ATG
GTG
TTA
TTC
H1/H3 (ppm)
H1/H3 (ppm)

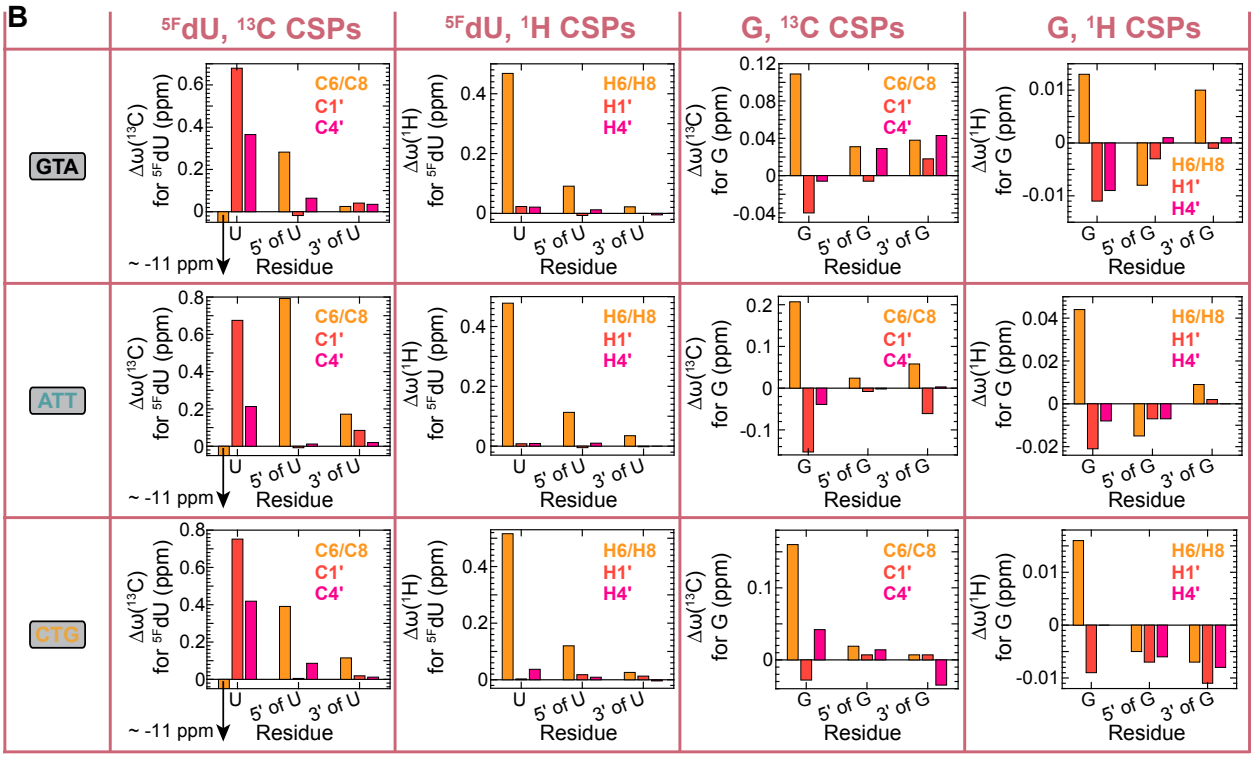


Figure S2. NMR spectra and resonance assignments of DNA duplexes and additional CSPs. (A) 1D ^{1}H imino, 2D ^{1}H-^{1}H NOESY, and 2D ^{1}H-^{13}C HSQC spectra used in assigning the resonances of the DNA hairpins examined in this study. 2D ^{1}H-^{1}H NOESY spectra were measured for all 16 unmodified (G•T, black) and 9 modified (G•^{5F}dU, pink) hairpin constructs. 2D ^{1}H-^{13}C HSQC spectra are shown for three unmodified (GTA, ATT and CTG) and four modified (GTA, ATT, CTG and TTG) sequences. Black arrows indicate chemical-shift changes arising from the 5F substitution. All spectra were collected at 600 MHz, 25 mM NaCl, pH 6.8 and T = 1 ºC. (B) Chemical shift perturbations (CSPs) between 5F-modified and unmodified sequences measured for sugar and base carbons and protons of the central G•T/^{5F}dU mismatch and its immediate 5' and 3' neighbors in three sequence contexts (GTA, ATT and CTG).

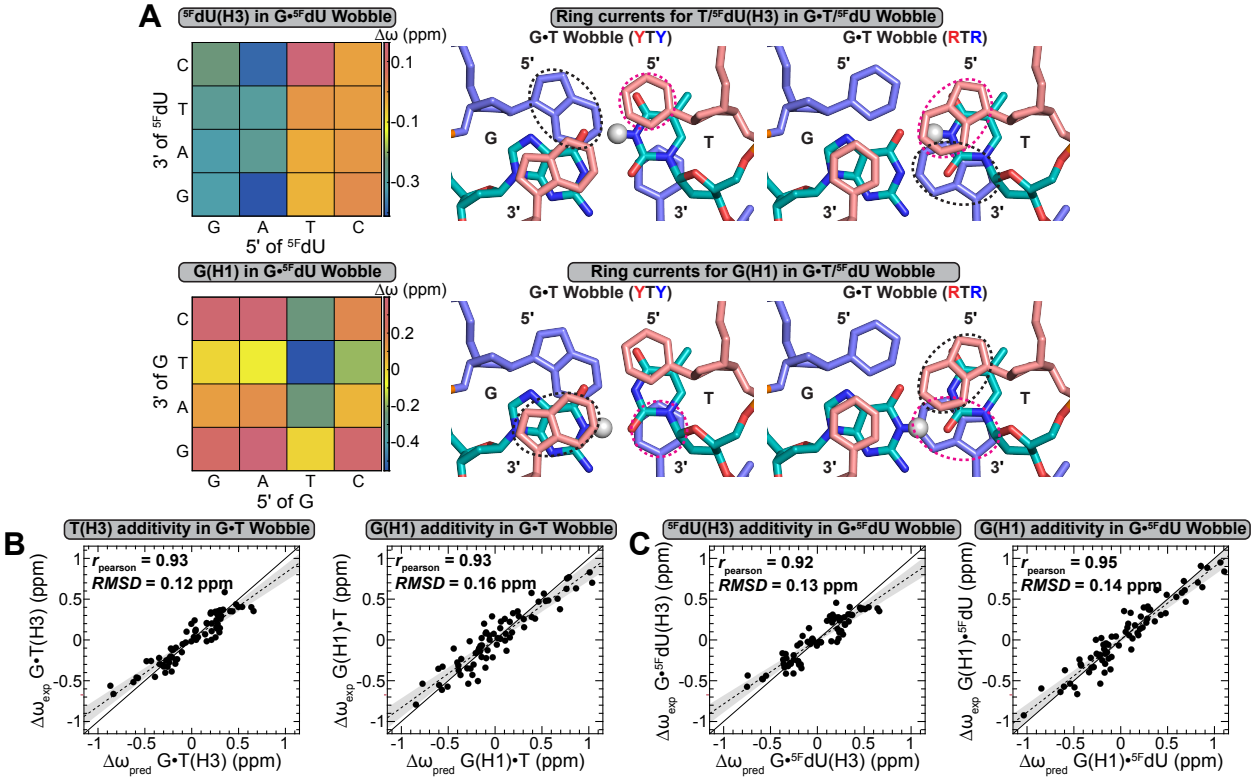


Figure S3. Analysis of imino proton chemical shifts in the G•T/^{5F}dU wobble ground state. (A) Heatmaps showing sequence-specific ^{5F}dU(H3) (left, top) and G(H1) (left, bottom) imino ^{1}H chemical shifts ($\Delta\omega = \omega(\text{seq}) - \omega(\text{ref})$; reference = CTA sequence context) in the G•^{5F}dU mismatch. $\Delta\omega$ values are color-coded. Division into two groups by the 5' neighbor is statistically significant based on Wilcoxon rank-sum test (p = 10^{-4} for ^{5F}dU(H3), p = 0.05 for G(H1). Also shown are structural models for a G•T/^{5F}dU wobble base pair in different trinucleotide sequence contexts. The central base pair (bp) is colored in turquoise. The bps on the 5' and 3' of T/^{5F}dU are colored pink and blue, respectively. The neighbors of T/^{5F}dU are denoted Y = pyrimidine and R = purine. Imino protons are shown in grey spheres. Dashed ellipses highlight nearby nucleobases that exert ring-current effects on the imino proton chemical shifts (pink, bases responsible for the 5′-neighbor dependence; black, purine bases contributing to the insensitivity of imino proton chemical shifts to the 3′-neighbor; see Supporting Discussion). (B) Testing additivity of imino chemical shifts for T(H3) (left) and G(H1) (right) protons in the G•T wobble ground state, by comparing the CSP ($\Delta\omega$)

arising from changing both Watson-Crick neighbors with the sum of contributions from changing the individual 3' and 5' neighbors. The black line indicates the y=x line with slope one. Shown are fit to a linear model (black, dashed) with the region encompassing the 95% confidence intervals for slope and y-intercept shaded in grey, as well as the root-mean-square deviation (RMSD), and the Pearson correlation coefficient ($r_{pearson}$). (C) Testing chemical shift additivity for ^{5F}dU(H3) (left) and G(H1) (right) in the G•^{5F}dU wobble ground state by comparing the CSP ($\Delta\omega$) arising from changing both Watson-Crick neighbors with the sum of contributions from changing the individual 3' and 5' neighbors. The black line indicates the y=x line with slope one. Shown are fit to a linear model (black, dashed) with the region encompassing the 95% confidence intervals for slope and y-intercept shaded in grey, as well as the root-mean-square deviation (RMSD), and the Pearson correlation coefficient ($r_{pearson}$).

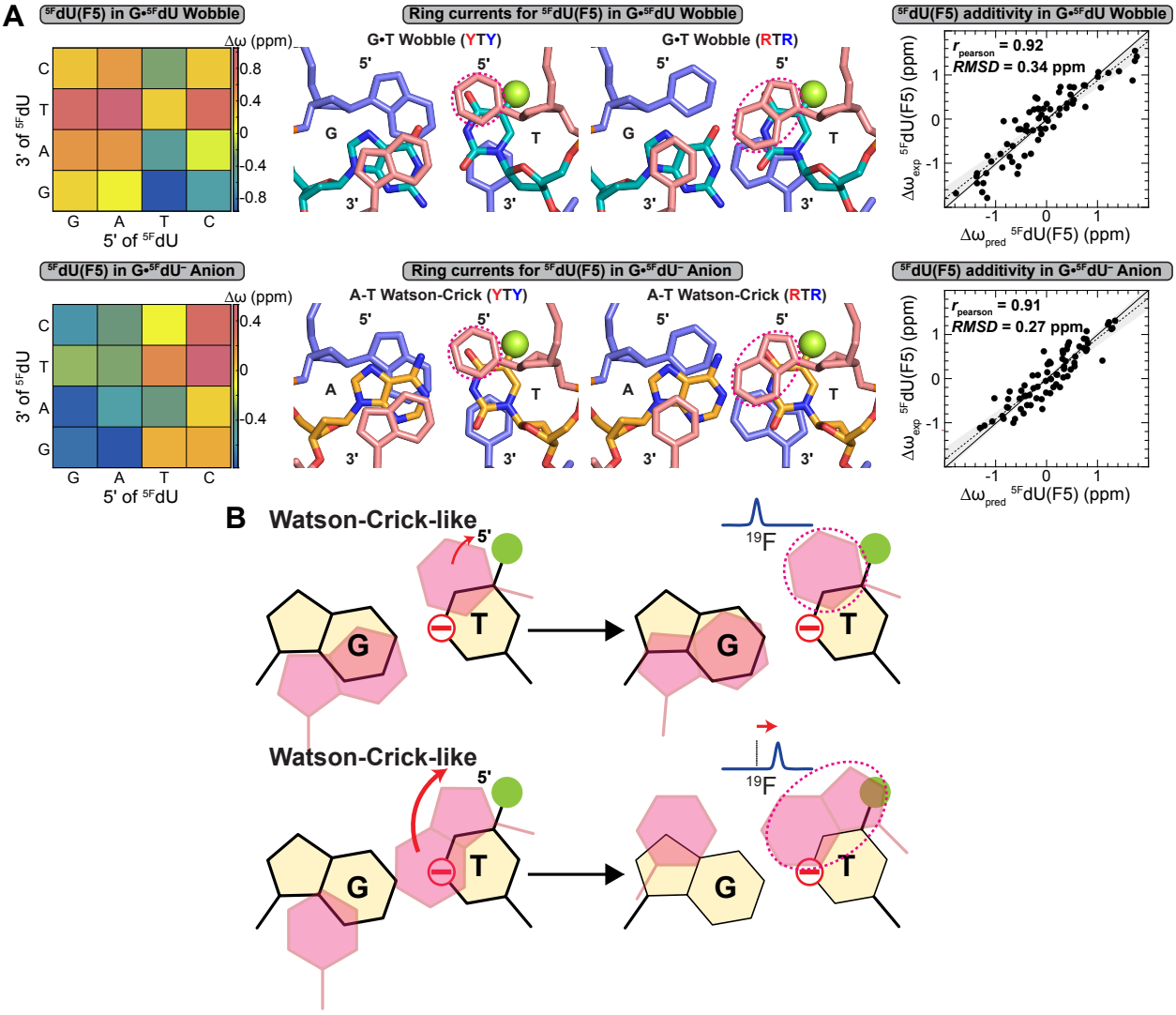
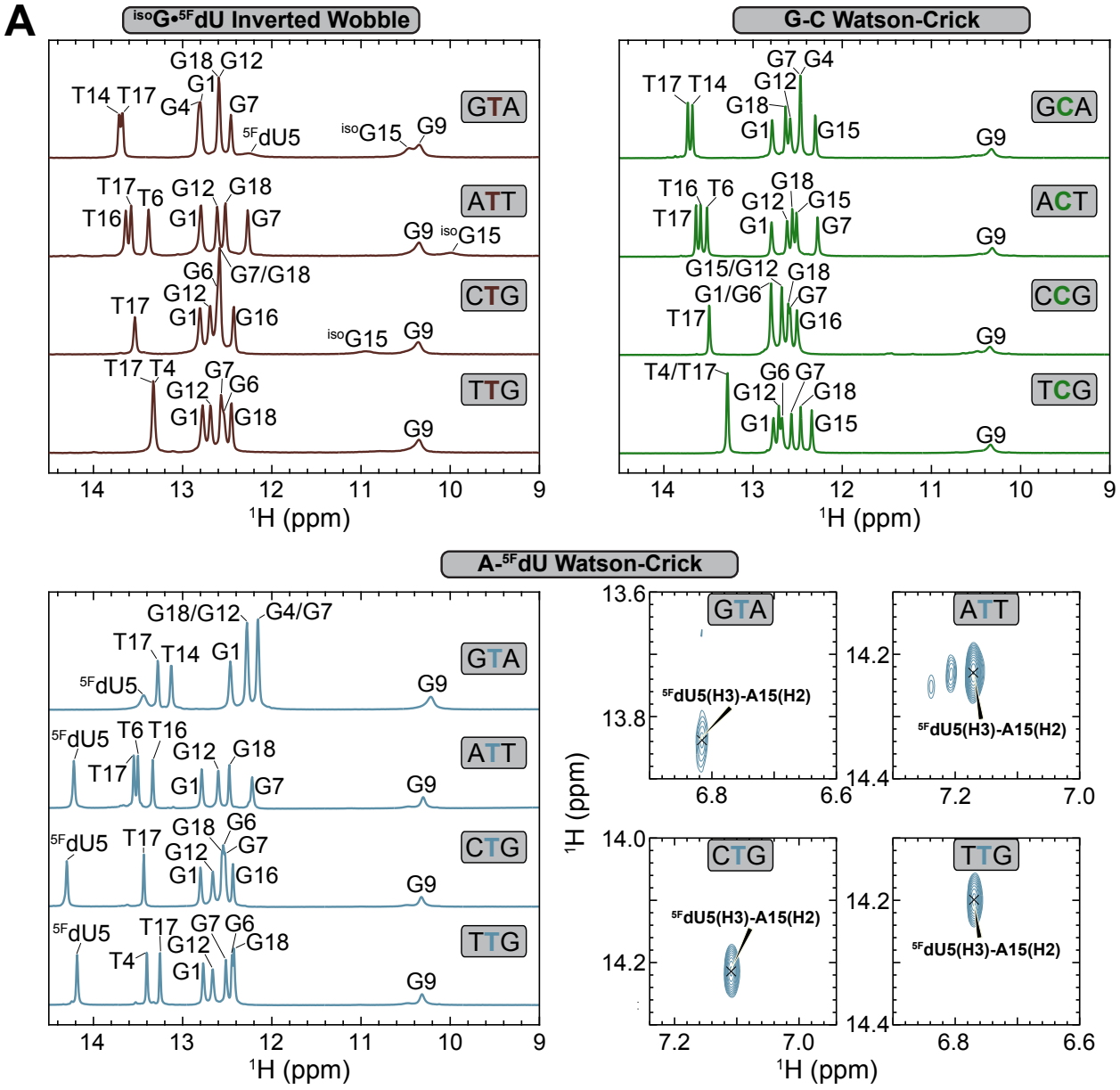
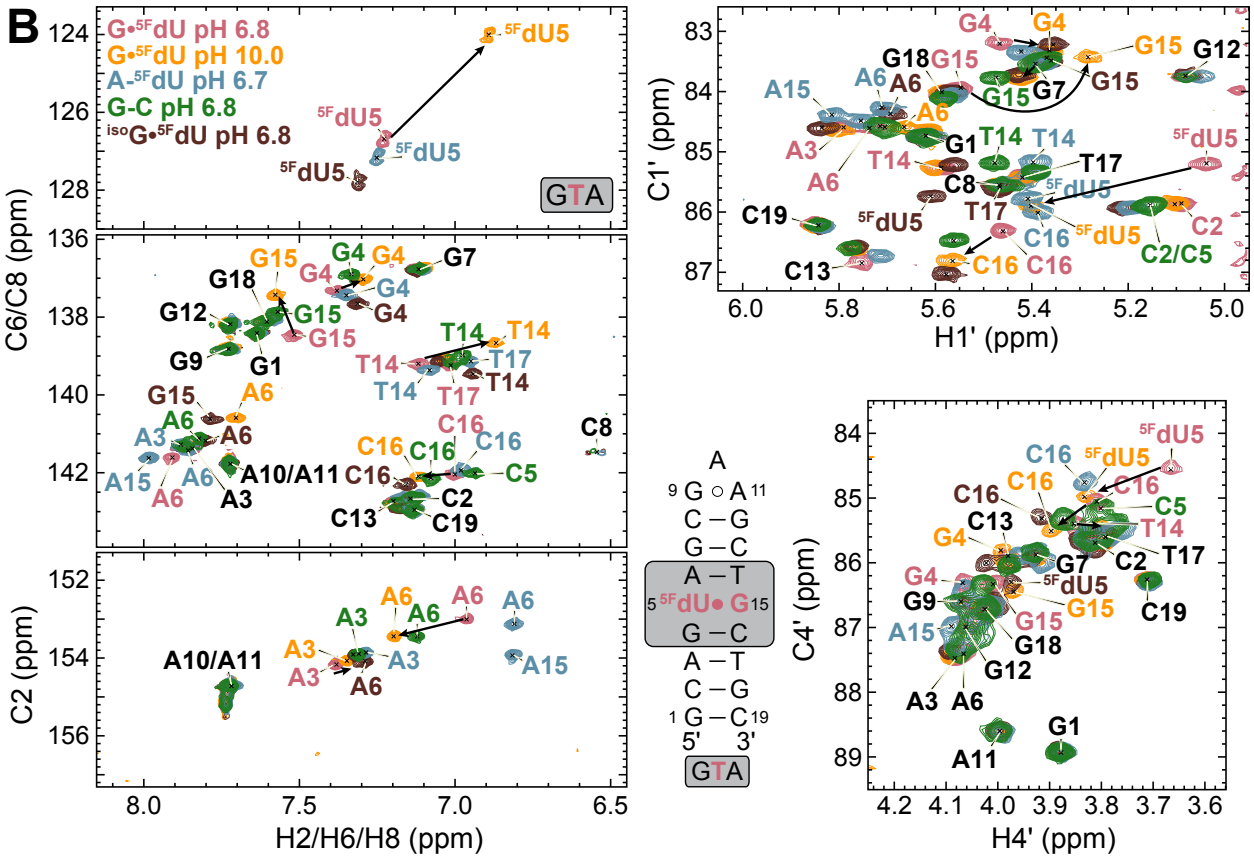
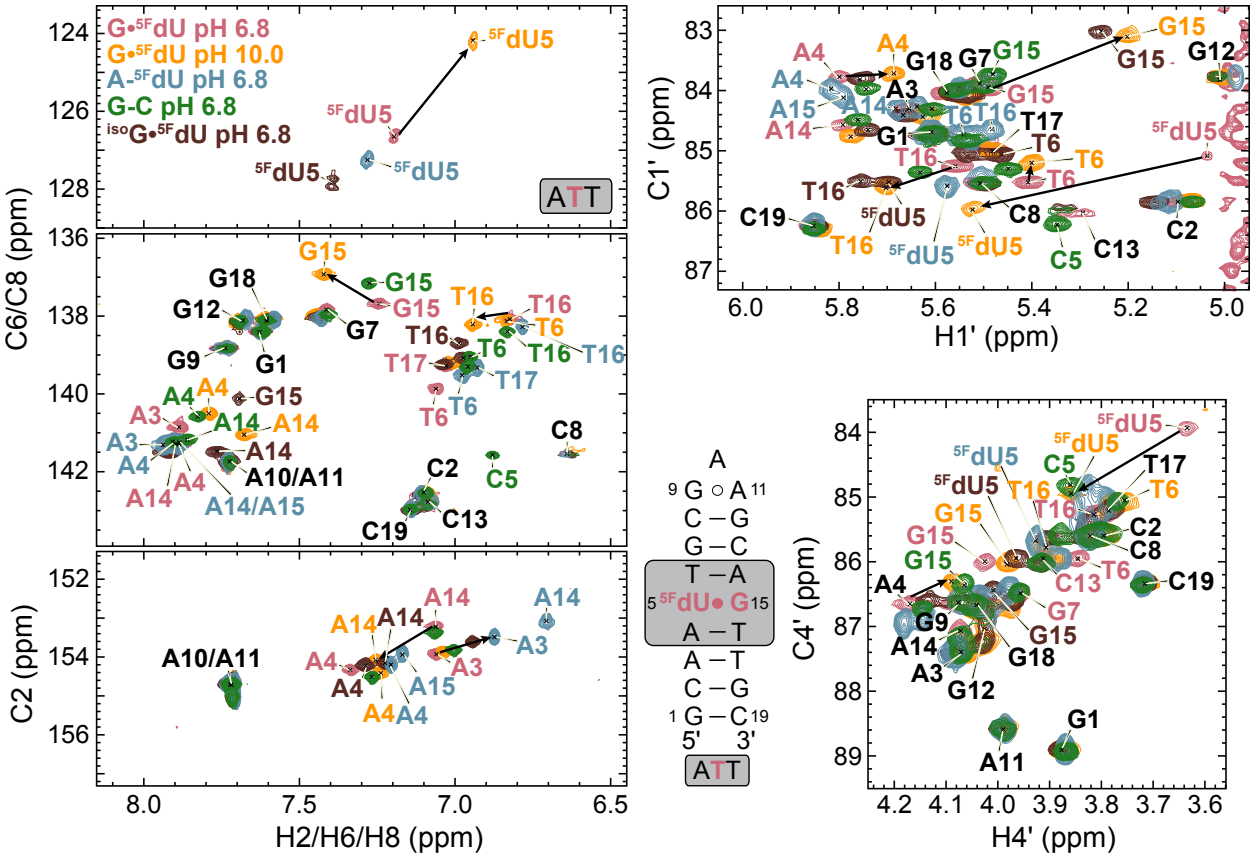


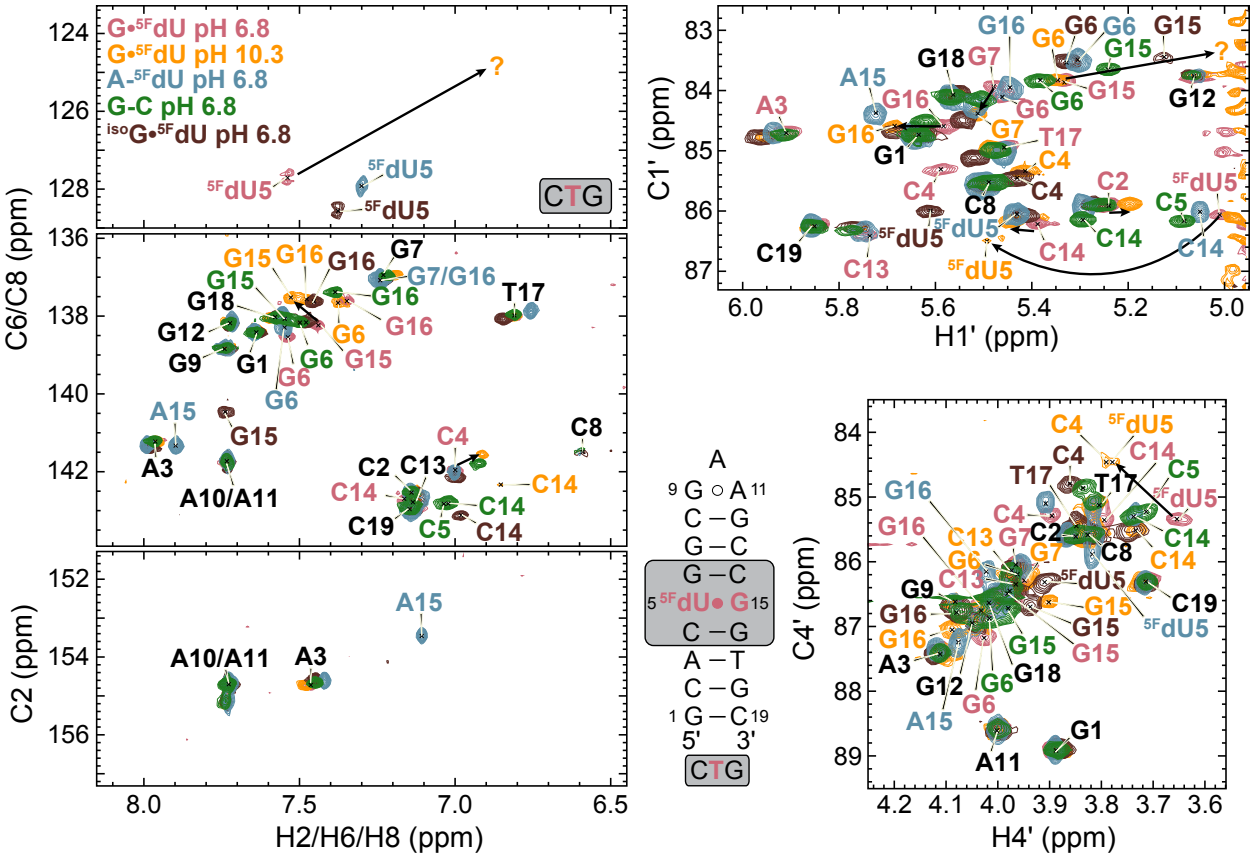
Figure S4. Analysis of 5F chemical shifts in the G•^{5F}dU wobble and anionic WC-like G•^{5F}dU^{-}. (A) (left) Heatmaps showing sequence-specific ^{5F}dU(F5) ^{19}F chemical shifts ($\Delta\omega = \omega(seq) - \omega(ref)$; reference = CTA sequence context) for the G•^{5F}dU wobble and anionic WC-like G•^{5F}dU^{-}. $\Delta\omega$ values are color-coded. In both cases, division into two groups by the 5' neighbor is statistically significant based on Wilcoxon rank-sum test (p = 0.02 and p = 0.001, respectively). Also shown are idealized structural models for the (top) G•^{5F}dU wobble (top) and WC-like G•^{5F}dU^{-} (bottom) using A-^{5F}dU as mimic in different trinucleotide sequence contexts. The central base pair (bp) is colored in turquoise (top) or orange (bottom). The bps on the 5' and 3' of ^{5F}dU are colored pink and blue, respectively. The neighbors of ^{5F}dU are denoted Y = pyrimidine and R = purine. The

fluorine atom is shown in green spheres. Dashed ellipses highlight key nucleobases that exert ring-current effects, leading to the dependence of the 5FdU (F5) chemical shift on the 5′ neighbor (see Supporting Discussion). Also shown are tests of chemical shift additivity for ^{5F}dU(F5) in (top) G•^{5F}dU (top) and (bottom) WC-like G•^{5F}dU^{-}. Additivity was assessed by comparing the CSP ($\Delta\omega$) arising from changing both Watson-Crick neighbors with the sum of contributions from changing the individual 3' and 5' neighbors. The black line indicates the y=x line with slope one. Shown are fit to a linear model (black, dashed) with the region encompassing the 95% confidence intervals for slope and y-intercept shaded in grey, as well as the root-mean-square deviation (RMSD), and the Pearson correlation coefficient ($r_{pearson}$). (C) Proposed displacement of 5' neighbors to adjust stacking interactions with the anionic thymine. Unfavorable interactions with anionic base cause a larger displacement of the purine 5' neighbor relative to the pyrimidine 5' neighbor. This displacement results in ring currents (pink dashed ellipses) that can explain the observed ^{5F}dU(F5) shifts in the anionic WC-like G•^{5F}dU^{-}.


A
isoG•5FdU Inverted Wobble
G-C Watson-Crick
A-5FdU Watson-Crick
GTA
ATT
CTG
TTG
GCA
ACT
CCG
TCG
1H (ppm)
5FdU5(H3)-A15(H2)


B
G•5FdU pH 6.8
G•5FdU pH 10.0
A-5FdU pH 6.7
G-C pH 6.8
isoG•5FdU pH 6.8
GTA
C6/C8 (ppm)
C2 (ppm)
H2/H6/H8 (ppm)
C1' (ppm)
H1' (ppm)
C4' (ppm)
H4' (ppm)
A
9 G ○ A 11
C – G
G – C
A – T
5 5FdU● G 15
G – C
A – T
C – G
1 G – C 19
5' 3'
GTA


G•5FdU pH 6.8
G•5FdU pH 10.0
A-5FdU pH 6.8
G-C pH 6.8
isoG•5FdU pH 6.8
ATT
C6/C8 (ppm)
C2 (ppm)
H2/H6/H8 (ppm)
C1' (ppm)
H1' (ppm)
C4' (ppm)
H4' (ppm)
A
9 G ○ A 11
C – G
G – C
T – A
5 5FdU● G 15
A – T
A – T
C – G
1 G – C 19
5' 3'
ATT


G•5FdU pH 6.8
G•5FdU pH 10.3
A-5FdU pH 6.8
G-C pH 6.8
isoG•5FdU pH 6.8
CTG
C6/C8 (ppm)
C2 (ppm)
H2/H6/H8 (ppm)
C1' (ppm)
H1' (ppm)
C4' (ppm)
H4' (ppm)
A
9 G ○ A 11
C – G
G – C
G – C
5 5FdU• G 15
C – G
A – T
C – G
1 G – C 19
5' 3'
CTG

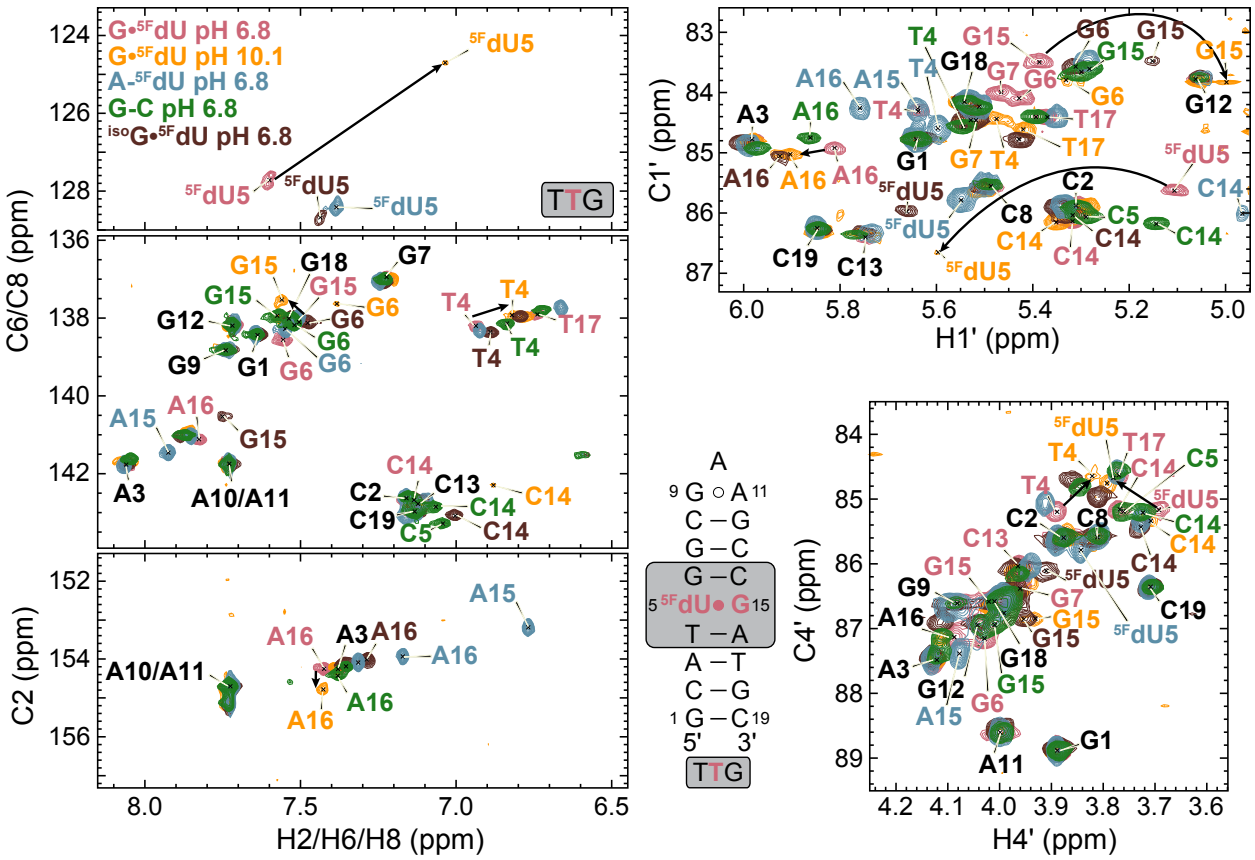
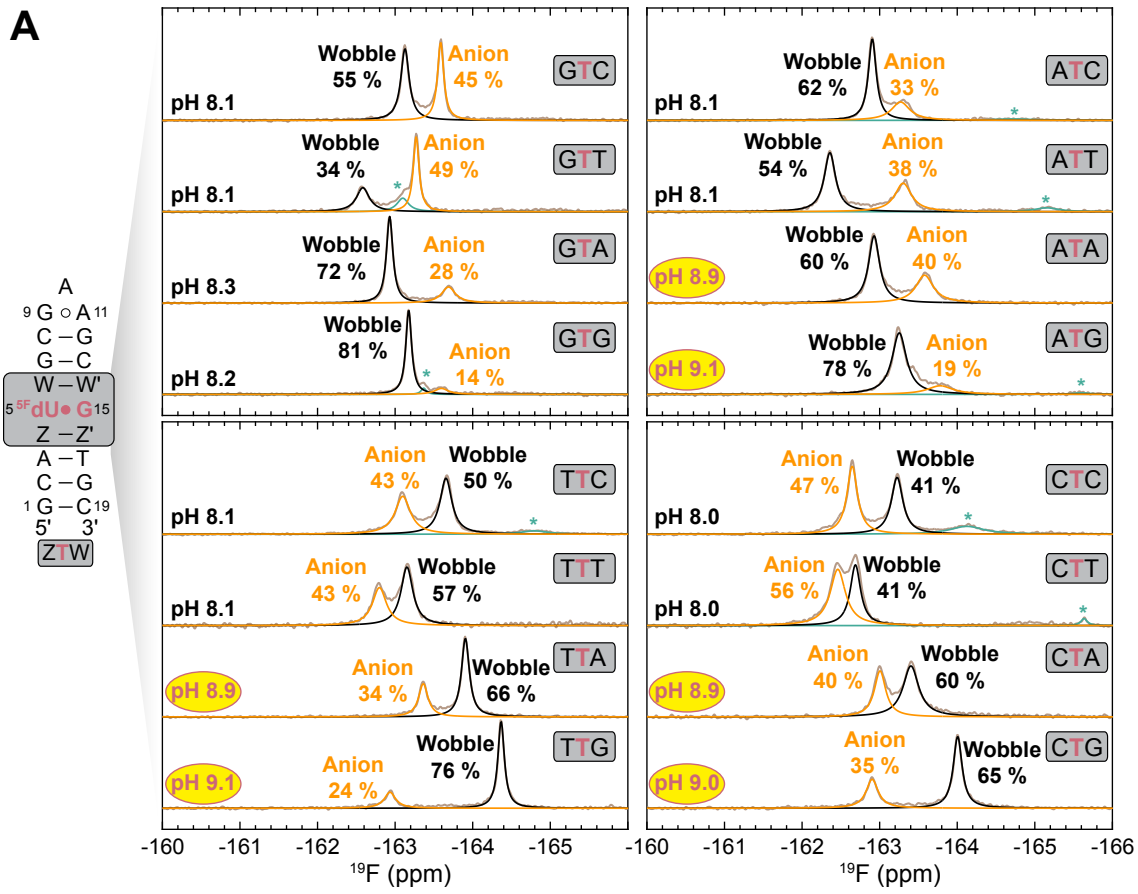
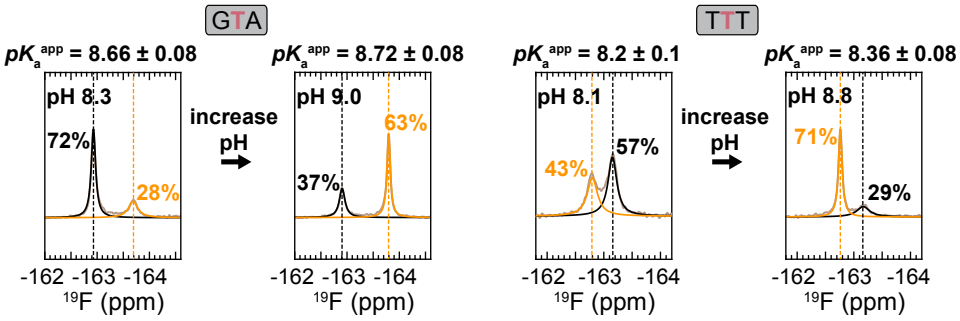


Figure S5. Chemical shift fingerprinting approach to rule out the anionic inverted wobble conformation. ^{iso}G•^{5F}dU (^{iso}G = isoguanosine) was used as a mimic of the inverted wobble (top left) and G-C (top right) / A-^{5F}dU (bottom) as a mimic of the Watson-Crick base pair. (A) 1D ^{1}H spectra measured at pH = 6.8 and T = 1 °C showing the imino resonances for conformational mimics in four representative sequence contexts GTA, ATT, CTG and TTG. For the A-^{5F}dU Watson-Crick mimic, the ^{5F}dU(H3)···A(H2) NOE cross peak is also shown (bottom right), indicating that the A-^{5F}dU base pair forms a Watson-Crick base pair. (B) 2D ^{1}H-^{13}C HSQC spectra showing the aromatic (C2H2//C6H6/C8H8) and sugar (C1'H1' and C4'H4'C4') regions for the conformational mimics; the G•^{5F}dU wobble ground state (pH 6.8); and the anionic G•^{5F}dU^{-} (pH ≥10.0). Black arrows indicate chemical-shift changes with increasing pH. All spectra were measured at 600 MHz or 700 MHz, and T = 1 °C.


A
Wobble 55 %
Anion 45 %
GTC
pH 8.1
Wobble 34 %
Anion 49 %
GTT
pH 8.1
Wobble 72 %
Anion 28 %
GTA
pH 8.3
Wobble 81 %
Anion 14 %
GTG
pH 8.2
Wobble 62 %
Anion 33 %
ATC
pH 8.1
Wobble 54 %
Anion 38 %
ATT
pH 8.1
Wobble 60 %
Anion 40 %
ATA
pH 8.9
Wobble 78 %
Anion 19 %
ATG
pH 9.1
Anion 43 %
Wobble 50 %
TTC
pH 8.1
Anion 43 %
Wobble 57 %
TTT
pH 8.1
Anion 34 %
Wobble 66 %
TTA
pH 8.9
Anion 24 %
Wobble 76 %
TTG
pH 9.1
Anion 47 %
Wobble 41 %
CTC
pH 8.0
Anion 56 %
Wobble 41 %
CTT
pH 8.0
Anion 40 %
Wobble 60 %
CTA
pH 8.9
Anion 35 %
Wobble 65 %
CTG
pH 9.0
A
9 G ○ A 11
C – G
G – C
W – W'
5 5F dU● G 15
Z – Z'
A – T
C – G
1 G – C 19
5' 3'
ZTW
-160 -161 -162 -163 -164 -165
19F (ppm)
-160 -161 -162 -163 -164 -165 -166
19F (ppm)


B
GTA
pKa app = 8.66 ± 0.08
pH 8.3
72%
28%
increase pH
pKa app = 8.72 ± 0.08
pH 9.0
63%
37%
-162 -163 -164
19F (ppm)
TTT
pKa app = 8.2 ± 0.1
pH 8.1
43%
57%
increase pH
pKa app = 8.36 ± 0.08
pH 8.8
71%
29%
-162 -163 -164
19F (ppm)

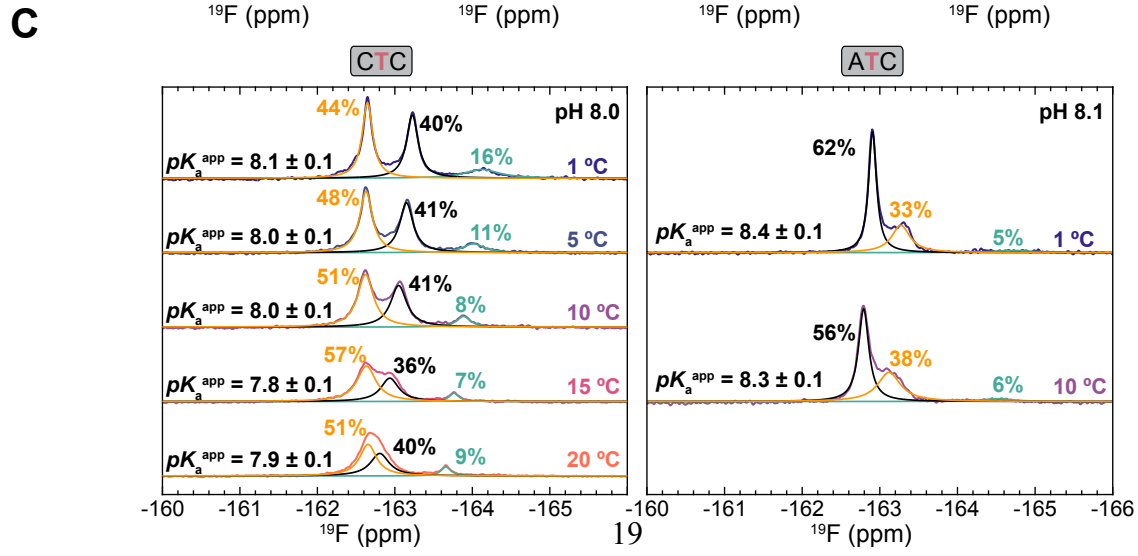

C
CTC
pH 8.0
44%
40%
16%
pKa app = 8.1 ± 0.1
1 °C
48%
41%
11%
pKa app = 8.0 ± 0.1
5 °C
51%
41%
8%
pKa app = 8.0 ± 0.1
10 °C
57%
36%
7%
pKa app = 7.8 ± 0.1
15 °C
51%
40%
9%
pKa app = 7.9 ± 0.1
20 °C
ATC
pH 8.1
62%
33%
5%
pKa app = 8.4 ± 0.1
1 °C
56%
38%
6%
pKa app = 8.3 ± 0.1
10 °C
-160 -161 -162 -163 -164 -165
19F (ppm)
-160 -161 -162 -163 -164 -165 -166
19F (ppm)

Figure S6. Detection of anionic WC-like G•^{5F}dU^{-} using pH-dependent ^{19}F NMR. (A) 1D ^{19}F NMR spectra for all 16 DNA hairpins containing a central G•^{5F}dU (in pink) in sixteen different trinucleotide sequence contexts (in grey). Z-Z' and W-W' represent the four Watson-Crick (A-T, T-A, G-C and C-G) neighbors. Shown are the spectra (light brown lines) overlaid with fits to two (black, orange) or three (black, orange, teal) Lorentzian lines (see Methods). The relative populations of the wobble and anionic G•^{5F}dU^{-} conformations are indicated. (B) 1D ^{19}F NMR spectra measured for two representative sequence contexts (GTA and TTT) at two pH conditions. Peaks were fit to Lorentzian line shapes (see Methods) to determine the relative populations of the wobble and anionic G•^{5F}dU^{-} conformations. The pK_a^{app} values are consistent across the two pH conditions. (C) 1D ^{19}F NMR spectra for two representative sequence contexts (CTC and ATC) measured at varying temperatures. Comparable pK_a^{app} values were obtained between 1°C and 20°C.

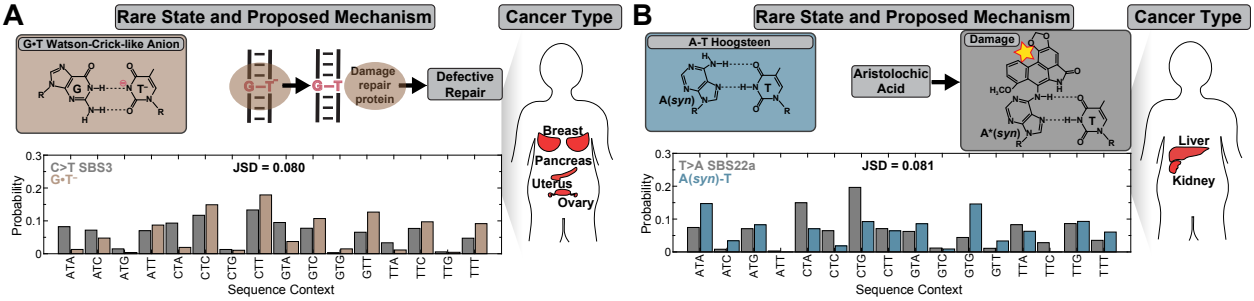


Figure S7. Conformational fingerprinting cancer mutational signatures. Shown are the cancer mutational signatures from the Catalogue of Somatic Mutations in Cancer (COSMIC)[4] database, along with their histogram correlation with the conformational fingerprints of rare mutagenic states. Also shown are the proposed mechanisms for mutations arising from these rare mutagenic states based on the aetiologies of the mutational signatures, as well as the cancer types linked to these signatures. The similarities with conformational propensities were calculated using Jensen-Shannon divergence (JSD). (A) C>T substitution from SBS3 (grey) correlates with WC-like anionic G•T^{-} propensity (brown). (B) T>A substitution from SBS22a (grey) correlates with A(*syn*)-T Hoogsteen propensity (blue).

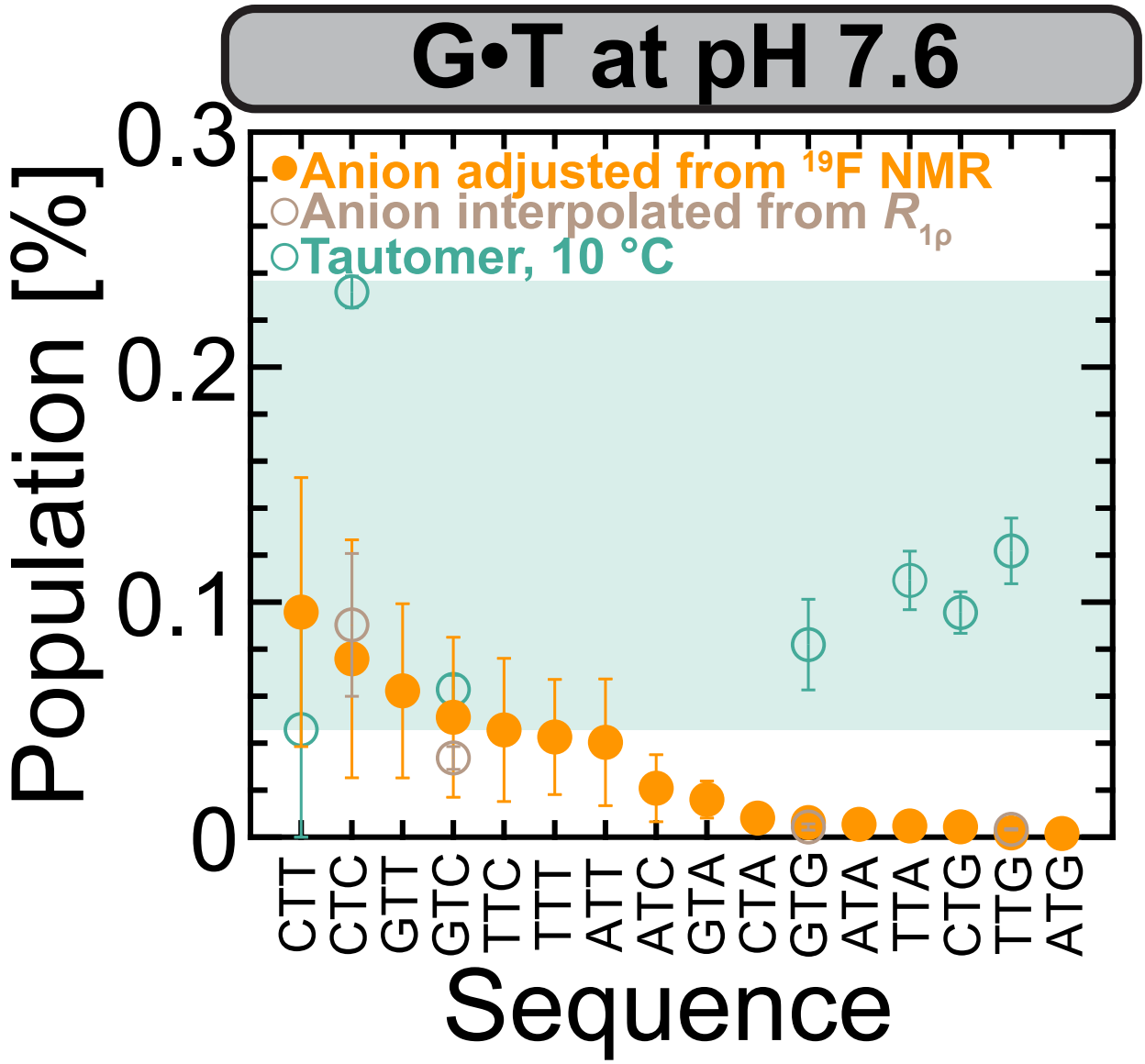


Figure S8. Comparing the relative populations of anionic and tautomeric WC-like G•T conformations. Comparing the populations of WC-like anionic G•T^{-} state (orange filled circles) obtained using ^{19}F NMR with WC-like anionic G•T^{-} (light brown open circles) previously measured by $R_{1\rho}$ at high pH and 10 °C[5] and WC-like tautomeric states G^{enol}•T and G•T^{enol} (teal open circles) previously measured by $R_{1\rho}$ at pH 6.9 and 10 °C[5] as a function of trinucleotide sequence context. The WC-like anionic G•T^{-} population was adjusted according to the energetic offset of c = $\Delta G°_{conf}$ (anion, F5) – $\Delta G°_{conf}$ (anion) = -3.4 kcal/mol and interpolated to pH 7.6. The available sequence

dependent tautomeric state population spans the teal rectangle. In certain sequence contexts such as CTT, CTC, GTT, GTC, and TTC and at pH 7.6, G•T^- achieves populations comparable to or even exceeding the populations of the tautomeric states.

Table S1. Oligonucleotides sequences used in this study. The central base pair in each construct is in bold.

Central base pair	Trinucleotide Sequence Context	Experiments	Sequence (5' – 3')
G•T	GTC	^{1}H NMR	GCAG**T**CGCGAAGCG**G**CTGC
G•T	GTT	^{1}H NMR	GCAG**T**TGCGAAGCA**G**CTGC
G•T	GTA	^{1}H NMR	GCAG**T**AGCGAAGCT**G**CTGC
G•T	GTG	^{1}H NMR	GCAG**T**GGCGAAGCC**G**CTGC
G•T	ATC	^{1}H NMR	GCAA**T**CGCGAAGCG**G**TTGC
G•T	ATT	^{1}H NMR	GCAA**T**TGCGAAGCA**G**TTGC
G•T	ATA	^{1}H NMR	GCAA**T**AGCGAAGCT**G**TTGC
G•T	ATG	^{1}H NMR	GCAA**T**GGCGAAGCC**G**TTGC
G•T	TTC	^{1}H NMR	GCAT**T**CGCGAAGCG**G**ATGC
G•T	TTT	^{1}H NMR	GCAT**T**TGCGAAGCA**G**ATGC
G•T	TTA	^{1}H NMR	GCAT**T**AGCGAAGCT**G**ATGC
G•T	TTG	^{1}H NMR	GCAT**T**GGCGAAGCC**G**ATGC
G•T	CTC	^{1}H NMR	GCAC**T**CGCGAAGCG**G**GTGC
G•T	CTT	^{1}H NMR	GCAC**T**TGCGAAGCA**G**GTGC
G•T	CTA	^{1}H NMR	GCAC**T**AGCGAAGCT**G**GTGC
G•T	CTG	^{1}H NMR	GCAC**T**GGCGAAGCC**G**GTGC
G•^{5F}dU	GTC	^{19}F NMR, ^{1}H NMR	GCAG**^{5F}dU**CGCGAAGCG**G**CTGC
G•^{5F}dU	GTT	^{19}F NMR, ^{1}H NMR	GCAG**^{5F}dU**TGCGAAGCA**G**CTGC
G•^{5F}dU	GTA	^{19}F NMR, ^{1}H NMR	GCAG**^{5F}dU**AGCGAAGCT**G**CTGC
G•^{5F}dU	GTG	^{19}F NMR, ^{1}H NMR	GCAG**^{5F}dU**GGCGAAGCC**G**CTGC
G•^{5F}dU	ATC	^{19}F NMR, ^{1}H NMR	GCAA**^{5F}dU**CGCGAAGCG**G**TTGC
G•^{5F}dU	ATT	^{19}F NMR, ^{1}H NMR	GCAA**^{5F}dU**TGCGAAGCA**G**TTGC
G•^{5F}dU	ATA	^{19}F NMR, ^{1}H NMR	GCAA**^{5F}dU**AGCGAAGCT**G**TTGC
G•^{5F}dU	ATG	^{19}F NMR, ^{1}H NMR	GCAA**^{5F}dU**GGCGAAGCC**G**TTGC
G•^{5F}dU	TTC	^{19}F NMR, ^{1}H NMR	GCAT**^{5F}dU**CGCGAAGCG**G**ATGC

G•^{5F}dU	TTT	^{19}F NMR, ^{1}H NMR	GCAT**^{5F}dU**TGCGAAGCA**G**ATGC
G•^{5F}dU	TTA	^{19}F NMR, ^{1}H NMR	GCAT**^{5F}dU**AGCGAAGCT**G**ATGC
G•^{5F}dU	TTG	^{19}F NMR, ^{1}H NMR	GCAT**^{5F}dU**GGCGAAGCC**G**ATGC
G•^{5F}dU	CTC	^{19}F NMR, ^{1}H NMR	GCAC**^{5F}dU**CGCGAAGCG**G**GTGC
G•^{5F}dU	CTT	^{19}F NMR, ^{1}H NMR	GCAC**^{5F}dU**TGCGAAGCA**G**GTGC
G•^{5F}dU	CTA	^{19}F NMR, ^{1}H NMR	GCAC**^{5F}dU**AGCGAAGCT**G**GTGC
G•^{5F}dU	CTG	^{19}F NMR, ^{1}H NMR	GCAC**^{5F}dU**GGCGAAGCC**G**GTGC
G-C	GCA	^{1}H NMR	GCAG**C**AGCGAAGCT**G**CTGC
G-C	ACT	^{1}H NMR	GCAA**C**TGCGAAGCA**G**TTGC
G-C	CCG	^{1}H NMR	GCAC**C**GGCGAAGCC**G**GTGC
G-C	TCG	^{1}H NMR	GCAT**C**GGCGAAGCC**G**ATGC
A-^{5F}dU	GTA	^{1}H NMR	GCAG**^{5F}dU**AGCGAAGCT**A**CTGC
A-^{5F}dU	ATT	^{1}H NMR	GCAA**^{5F}dU**TGCGAAGCA**A**TTGC
A-^{5F}dU	CTG	^{1}H NMR	GCAC**^{5F}dU**GGCGAAGCC**A**GTGC
A-^{5F}dU	TTG	^{1}H NMR	GCAT**^{5F}dU**GGCGAAGCC**A**ATGC
^{iso}G•^{5F}dU	GTA	^{1}H NMR	GCAG**^{5F}dU**AGCGAAGCT**^{iso}G**CTGC
^{iso}G•^{5F}dU	ATT	^{1}H NMR	GCAA**^{5F}dU**TGCGAAGCA**^{iso}G**TTGC
^{iso}G•^{5F}dU	CTG	^{1}H NMR	GCAC**^{5F}dU**GGCGAAGCC**^{iso}G**GTGC
^{iso}G•^{5F}dU	TTG	^{1}H NMR	GCAT**^{5F}dU**GGCGAAGCC**^{iso}G**ATGC
G•C/T	GTC strand 1	UV melting	CAGCAG**(C/T)**CGCGC
G•C/T	GTC strand 2	UV melting	GCGCG**G**CTGCTG
G•C/T	GTT strand 1	UV melting	CAGCAG**(C/T)**TGCGC
G•C/T	GTT strand 2	UV melting	GCGCA**G**CTGCTG
G•C/T	GTA strand 1	UV melting	CAGCAG**(C/T)**AGCGC
G•C/T	GTA strand 2	UV melting	GCGCT**G**CTGCTG
G•C/T	GTG strand 1	UV melting	CAGCAG**(C/T)**GGCGC
G•C/T	GTG strand 2	UV melting	GCGCC**G**CTGCTG
G•C/T	ATC strand 1	UV melting	CAGCAA**(C/T)**CGCGC
G•C/T	ATC strand 2	UV melting	GCGCG**G**TTGCTG
G•C/T	ATT strand 1	UV melting	CAGCAA**(C/T)**TGCGC
G•C/T	ATT strand 2	UV melting	GCGCA**G**TTGCTG
G•C/T	ATA strand 1	UV melting	CAGCAA**(C/T)**AGCGC
G•C/T	ATA strand 2	UV melting	GCGCT**G**TTGCTG
G•C/T	ATG strand 1	UV melting	CAGCAA**(C/T)**GGCGC
G•C/T	ATG strand 2	UV melting	GCGCC**G**TTGCTG

G•C/T	TTC strand 1	UV melting	CAGCAT**(C/T)**CGCGC
G•C/T	TTC strand 2	UV melting	GCGCG**G**ATGCTG
G•C/T	TTT strand 1	UV melting	CAGCAT**(C/T)**TGCGC
G•C/T	TTT strand 2	UV melting	GCGCA**G**ATGCTG
G•C/T	TTA strand 1	UV melting	CAGCAT**(C/T)**AGCGC
G•C/T	TTA strand 2	UV melting	GCGCT**G**ATGCTG
G•C/T	TTG strand 1	UV melting	CAGCAT**(C/T)**GGCGC
G•C/T	TTG strand 2	UV melting	GCGCC**G**ATGCTG
G•C/T	CTC strand 1	UV melting	CAGCAC**(C/T)**CGCGC
G•C/T	CTC strand 2	UV melting	GCGCG**G**GTGCTG
G•C/T	CTT strand 1	UV melting	CAGCAC**(C/T)**TGCGC
G•C/T	CTT strand 2	UV melting	GCGCA**G**GTGCTG
G•C/T	CTA strand 1	UV melting	CAGCAC**(C/T)**AGCGC
G•C/T	CTA strand 2	UV melting	GCGCT**G**GTGCTG
G•C/T	CTG strand 1	UV melting	CAGCAC**(C/T)**GGCGC
G•C/T	CTG strand 2	UV melting	GCGCC**G**GTGCTG

Table S2. List of spin-lock powers ($\omega_1/2\pi$ in Hz) and offsets ($\Omega_{eff}/2\pi$ in Hz) used in the off-resonance ^{15}N $R_{1\rho}$ experiment for GTA at pH 8.5 T= 1 °C.

Nucleus	$\omega_1/2\pi$ (Hz), [$\Omega_{eff}/2\pi$ (Hz)]
G(N1)	500, [-1749.0, -1590.0, -1431.0, -1272.0, -1113.0, -954.0, -795.0, -636.0, -477.0, -318.0, -159.0, -10.0, 10.0, 159.0, 318.0, 477.0, 636.0, 795.0, 954.0, 1113.0, 1272.0, 1431.0, 1590.0, 1749.0] 800, [-2805.0, -2550.0, -2295.0, -2040.0, -1785.0, -1530.0, -1275.0, -1020.0, -765.0, -510.0, -255.0, -10.0, 10.0, 255.0, 510.0, 765.0, 1020.0, 1275.0, 1530.0, 1785.0, 2040.0, 2295.0, 2550.0, 2805.0] 1200, [-4202.0, -3820.0, -3438.0, -3056.0, -2674.0, -2292.0, -1910.0, -1528.0, -1146.0, -764.0, -382.0, -10.0, 10.0, 382.0, 764.0, 1146.0, 1528.0, 1910.0, 2292.0, 2674.0, 3056.0, 3438.0, 3820.0, 4202.0] 1500, [-5247.0, -4770.0, -4293.0, -3816.0, -3339.0, -2862.0, -2385.0, -1908.0, -1431.0, -954.0, -477.0, -10.0, 10.0, 477.0, 954.0, 1431.0, 1908.0, 2385.0, 2862.0, 3339.0, 3816.0, 4293.0, 4770.0, 5247.0] 2000, [-6996.0, -6360.0, -5724.0, -5088.0, -4452.0, -3816.0, -3180.0, -2544.0, -1908.0, -1272.0, -636.0, -10.0, 10.0, 636.0, 1272.0, 1908.0, 2544.0, 3180.0, 3816.0, 4452.0, 5088.0, 5724.0, 6360.0, 6996.0]
T(N3)	500, [-2001.0, -1914.0, -1827.0, -1740.0, -1653.0, -1566.0, -1479.0, -1392.0, -1305.0, -1218.0, -1131.0, -1044.0, -957.0, -870.0, -783.0, -696.0, -609.0, -522.0, -435.0, -348.0, -261.0, -174.0, -87.0, -10.0, 10.0, 87.0, 174.0, 261.0, 348.0, 435.0, 522.0, 609.0, 696.0, 783.0, 870.0, 957.0, 1044.0, 1131.0, 1218.0, 1305.0, 1392.0, 1479.0, 1566.0, 1653.0, 1740.0, 1827.0, 1914.0, 2001.0] 800, [-3197.0, -3058.0, -2919.0, -2780.0, -2641.0, -2502.0, -2363.0, -2224.0, -2085.0, -1946.0, -1807.0, -1668.0, -1529.0, -1390.0, -1251.0, -1112.0, -973.0, -834.0, -695.0,

	-556.0, -417.0, -278.0, -139.0, -10.0, 10.0, 139.0, 278.0, 417.0, 556.0, 695.0, 834.0, 973.0, 1112.0, 1251.0, 1390.0, 1529.0, 1668.0, 1807.0, 1946.0, 2085.0, 2224.0, 2363.0, 2502.0, 2641.0, 2780.0, 2919.0, 3058.0, 3197.0] 1200, [-4807.0, -4598.0, -4389.0, -4180.0, -3971.0, -3762.0, -3553.0, -3344.0, -3135.0, -2926.0, -2717.0, -2508.0, -2299.0, -2090.0, -1881.0, -1672.0, -1463.0, -1254.0, -1045.0, -836.0, -627.0, -418.0, -209.0, -10.0, 10.0, 209.0, 418.0, 627.0, 836.0, 1045.0, 1254.0, 1463.0, 1672.0, 1881.0, 2090.0, 2299.0, 2508.0, 2717.0, 2926.0, 3135.0, 3344.0, 3553.0, 3762.0, 3971.0, 4180.0, 4389.0, 4598.0, 4807.0] 1500, [-6003.0, -5742.0, -5481.0, -5220.0, -4959.0, -4698.0, -4437.0, -4176.0, -3915.0, -3654.0, -3393.0, -3132.0, -2871.0, -2610.0, -2349.0, -2088.0, -1827.0, -1566.0, -1305.0, -1044.0, -783.0, -522.0, -261.0, -10.0, 10.0, 261.0, 522.0, 783.0, 1044.0, 1305.0, 1566.0, 1827.0, 2088.0, 2349.0, 2610.0, 2871.0, 3132.0, 3393.0, 3654.0, 3915.0, 4176.0, 4437.0, 4698.0, 4959.0, 5220.0, 5481.0, 5742.0, 6003.0] 2000, [-8004.0, -7656.0, -7308.0, -6960.0, -6612.0, -6264.0, -5916.0, -5568.0, -5220.0, -4872.0, -4524.0, -4176.0, -3828.0, -3480.0, -3132.0, -2784.0, -2436.0, -2088.0, -1740.0, -1392.0, -1044.0, -696.0, -348.0, -10.0, 10.0, 348.0, 696.0, 1044.0, 1392.0, 1740.0, 2088.0, 2436.0, 2784.0, 3132.0, 3480.0, 3828.0, 4176.0, 4524.0, 4872.0, 5220.0, 5568.0, 5916.0, 6264.0, 6612.0, 6960.0, 7308.0, 7656.0, 8004.0] 2500, [-10005.0, -9570.0, -9135.0, -8700.0, -8265.0, -7830.0, -7395.0, -6960.0, -6525.0, -6090.0, -5655.0, -5220.0, -4785.0, -4350.0, -3915.0, -3480.0, -3045.0, -2610.0, -2175.0, -1740.0, -1305.0, -870.0, -435.0, -10.0, 10.0, 435.0, 870.0, 1305.0, 1740.0, 2175.0, 2610.0, 3045.0, 3480.0, 3915.0, 4350.0, 4785.0, 5220.0, 5655.0, 6090.0, 6525.0, 6960.0, 7395.0, 7830.0, 8265.0, 8700.0, 9135.0, 9570.0, 10005.0]

Table S3. Exchange parameters obtained from fitting ^{15}N $R_{1\rho}$ data in GTA at pH 8.5, 4 °C. GS corresponds to the G•T wobble ground state, ES1 corresponds to the tautomeric WC-like state G^{enol}•T and G•T^{enol} in rapid equilibrium, and ES2 corresponds to anionic WC-like G•T^{-}. Red χ^2 is the reduced χ^2 obtained from fitting the $R_{1\rho}$ data.

	Parameter	GTA pH 8.5	
		G(N1)	T(N3)
3-state Shared Fitting	p_{ES1} (%)	0.08 ± 0.01	
	p_{ES2} (%)	0.06 ± 0.01	
	$k_{ex,GS:ES1}$ (s^{-1})	1400 ± 500	
	$k_{ex,GS:ES2}$ (s^{-1})	41000 ± 4000	
	$k_{ex,ES2:ES2}$ (s^{-1})	2000 ± 1000	
	$\Delta\omega_{ES1}$ (ppm)	35.8 ± 0.7	18.7 ± 0.7
	$\Delta\omega_{ES2}$ (ppm)	2 ± 20	66 ± 3
	R_1 (s^{-1})	1.81 ± 0.02	1.90 ± 0.02
	R_2 (s^{-1})	7.8 ± 0.2	6.3 ± 0.8
	Reduced χ^2	0.99	

Table S4. Single-base substitution (SBS) cancer mutational signatures from the Catalogue of Somatic Mutations in Cancer (COSMIC) database[4] showing strong similarities to the anionic Watson-Crick-like anionic G•T$^-$ probability distribution obtained from ^{19}F NMR measurements. Similarity assessed using the Jensen-Shannon Divergence (JSD ≤0.090, *p*-value <0.05, false discovery rate (FDR) <5%).

Transition mutation	Mutational Signature	Aetiology	JSD	*p*-value	FDR (%)
C>T	SBS23	Unknown	0.051	0.00006	0.5
C>T	SBS11	Exposure to alkylating agents, temozolomide treatment	0.072	0.0006	2.5
C>T	SBS3	Defective homologous recombination-based DNA damage repair	0.079	0.001	3.2
C>T	SBS40c	Unknown	0.089	0.002	4.9
T>A	SBS14	Concurrent polymerase epsilon mutation and defective DNA mismatch repair	0.065	0.0003	2.5

Table S5. Single-base substitution (SBS) cancer mutational signatures from the Catalogue of Somatic Mutations in Cancer (COSMIC) database[4] showing strong similarities to the A(syn)-T Hoogsteen probability distribution obtained from NMR measurements[6]. Similarity assessed using the Jensen-Shannon Divergence (JSD ≤0.090, *p*-value <0.05, false discovery rate (FDR) <5%).

Transition mutation	Mutational Signature	Aetiology	JSD	*p*-value	FDR (%)
T>A	SBS92	Tobacco smoking	0.025	0.000006	0.05
T>A	SBS42	Exposure to haloalkanes	0.032	0.00004	0.2
T>A	SBS4	Tobacco smoking	0.062	0.002	2.6
T>A	SBS19	Unknown	0.069	0.003	2.6
T>A	SBS25	Chemotherapy treatment	0.071	0.004	2.6
T>A	SBS29	Tobacco chewing	0.075	0.005	3.1
T>A	SBS22a	Aristolochic acid exposure	0.081	0.008	3.9
T>C	SBS89	Unknown	0.028	0.00002	0.1
T>C	SBS26	Defective DNA mismatch repair	0.030	0.00003	0.1
T>C	SBS10c	Defective PolD1 proofreading	0.039	0.0001	0.4
T>C	SBS5	Unknown (clock-like signature)	0.042	0.0002	0.4
T>C	SBS6	Defective DNA mismatch repair	0.044	0.0003	0.5
T>C	SBS44	Defective DNA mismatch repair	0.049	0.0005	0.6
T>C	SBS98	Unknown	0.049	0.0005	0.6
T>C	SBS25	Chemotherapy treatment	0.050	0.0006	0.6
T>C	SBS40b	Unknown	0.051	0.0006	0.6
T>C	SBS42	Exposure to haloalkanes	0.057	0.001	0.9
T>C	SBS36	Defective base excision repair	0.057	0.001	0.9
T>C	SBS4	Tobacco smoking	0.061	0.002	1.1
T>C	SBS40a	Unknown	0.065	0.003	1.3
T>C	SBS92	Tobacco smoking	0.076	0.006	2.4
T>C	SBS14	Concurrent polymerase epsilon mutation and defective DNA mismatch repair	0.081	0.008	2.9
T>C	SBS19	Unknown	0.086	0.01	3.8
T>G	SBS92	Tobacco smoking	0.042	0.0002	1.4
T>G	SBS94	Unknown	0.050	0.0005	1.4
T>G	SBS97	Unknown	0.051	0.0007	1.4
T>G	SBS36	Defective base excision repair	0.055	0.001	1.7
T>G	SBS3	Defective homologous recombination-based DNA damage repair	0.058	0.001	1.7
T>G	SBS4	Tobacco smoking	0.066	0.003	2.6

T>G	SBS96	Unknown	0.075	0.006	3.8
T>G	SBS24	Exposure to aflatoxin	0.080	0.008	4.5
T>G	SBS22b	Aristolochic acid exposure	0.082	0.009	4.5

Table S6. Experimental single-base substitution (SBS) mutational signatures from the Catalogue of Somatic Mutations in Cancer (COSMIC) database[4] for exposure of human tissue to benzo[a]pyrene (BaP) or aristolochic acid (AA-I and AA-II), showing strong similarities to the A(*syn*)-T Hoogsteen probability distribution obtained from NMR measurements[6]. Similarity assessed for T>A, T>C, and T>G substitutions using the Jensen-Shannon Divergence (JSD ≤0.090, *p*-value <0.05, false discovery rate (FDR) <5%). Greyed out rows indicate failure to pass one of the thresholds.

Experimental Signature	Transition mutation	JSD	*p*-value	FDR (%)	Experimental Study
BaP	T>A	0.046	0.0003	1.6	Mingard et al.[7]
	T>C	0.075	0.006	2.5	
	T>G	0.051	0.0006	2.2	
BaP	T>A	0.073	0.005	3.4	Kucab et al.[8]
	T>C	0.064	0.002	1.3	
	T>G				
BaP	T>A	0.081	0.009	4.2	Zhivagui et al.[9]
	T>C	0.091	0.02	4.7	
	T>G	0.075	0.006	5.1	
AA-I	T>A	0.065	0.003	3.0	Lu et al.[10]
	T>C				
	T>G				
AA-I	T>A	0.073	0.005	3.5	Kucab et al.[8]
	T>C	0.090	0.01	4.6	
	T>G	0.067	0.003	4.2	
AA-II	T>A	0.057	0.001	3.0	Kucab et al.[8]
	T>C	0.066	0.003	1.3	
	T>G				

References

1. Haracska, L., et al., Role of human DNA polymerase κ as an extender in translesion synthesis. *PNAS* **2002,** *99* (25), 16000-16005.
2. Choi, J.-Y., et al., Translesion Synthesis across O^6-Alkylguanine DNA Adducts by Recombinant Human DNA Polymerases. *J Biol Chem* **2006,** *281* (50), 38244-38256.
3. Choi, J.-Y.; Guengerich, F. P., Kinetic Analysis of Translesion Synthesis Opposite Bulky N^2- and O^6-Alkylguanine DNA Adducts by Human DNA Polymerase REV1. *J Biol Chem* **2008,** *283* (35), 23645-23655.
4. Tate, J. G., et al., COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **2018,** *47* (D1), D941-D947.
5. Kimsey, I. J., et al., Dynamic basis for dG·dT misincorporation via tautomerization and ionization. *Nature* **2018,** *554* (7691), 195-201.
6. Manghrani, A., et al., Quantitative and Systematic NMR Measurements of Sequence-Dependent A–T Hoogsteen Dynamics in the DNA Double Helix. *Biochemistry* **2025,** *64* (5), 1042-1054.
7. Mingard, C., et al., Dissection of Cancer Mutational Signatures with Individual Components of Cigarette Smoking. *Chemical Research in Toxicology* **2023,** *36* (4), 714-723.
8. Kucab, J. E., et al., A Compendium of Mutational Signatures of Environmental Agents. *Cell* **2019,** *177* (4), 821-836.e16.
9. Zhivagui, M., et al., Experimental and pan-cancer genome analyses reveal widespread contribution of acrylamide exposure to carcinogenesis in humans. *Genome Research* **2019,** *29* (4), 521-531.
10. Lu, H., et al., Aristolochic acid mutational signature defines the low-risk subtype in upper tract urothelial carcinoma. *Theranostics* **2020,** *10* (10), 4323-4333.