HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2016 March 19.

Published in final edited form as:

Nature. 2015 March 19; 519(7543): 315–320. doi:10.1038/nature14227.

Visualizing Transient Watson-Crick Like Mispairs in DNA and RNA Duplexes

Isaac J. Kimsey, **Katja Petzold**, **Bharathwaj Sathyamoorthy**, **Zachary W. Stein**, and **Hashim M. Al-Hashimi***

Department of Biochemistry and Chemistry Duke University Medical Center Durham, NC 27710, USA

Abstract

Rare tautomeric and anionic nucleobases are believed to play fundamental biological roles but their prevalence and functional importance has remained elusive because they exist transiently, in low-abundance, and involve subtle movements of protons that are difficult to visualize. Using NMR relaxation dispersion, we show that wobble dG•dT and rG•rU mispairs in DNA and RNA duplexes exist in dynamic equilibrium with short-lived, low-populated Watson-Crick like mispairs that are stabilized by rare enolic or anionic bases. These mispairs can evade Watson-Crick fidelity checkpoints and form with probabilities (10^{-3}-10^{-5}) that strongly imply a universal role in replication and translation errors. Our results indicate that rare tautomeric and anionic bases are widespread in nucleic acids, expanding their structural and functional complexity beyond that attainable with canonical bases.

Nucleic acid bases exist predominantly in one neutral tautomeric form. This in turn gives rise to the strict Watson-Crick (WC) pairing rules (Fig. 1a) that govern how genetic information is replicated, transcribed, and translated. However, if bases adopt alternative energetically disfavored tautomeric or anionic forms (Fig. 1a), pairing rules can be violated and new functions can emerge. For example, although rarely observed, minor tautomeric and anionic bases can form WC-like dG•dT[1-3], dA•dC[4], and rG•rU[5,6] mispairs that are believed to contribute to spontaneous mutations[1,7-9] and translational errors[10]. Chemical modifications that stabilize or lock bases in their anionic or enol-like forms can be mutagenic[11,12] or expand the decoding capacity of tRNAs[13,14]. In addition, anionic and tautomeric forms of the bases are believed to play crucial roles in nucleic acid catalysis[15,16], RNA-ligand recognition[17,18], and in the therapeutic mechanisms of nucleic acid base analogues[19].

*Correspondence and requests for materials should be addressed to H.M.A. (hashim.al.hashimi@duke.edu). Tel: 919-684-4325, Fax: 919-684-8885.

Author Contributions I.J.K and H.M.A. conceived the project and experimental design. I.J.K prepared NMR samples as well as performed and analyzed all NMR RD experiments. I.J.K assigned resonances in all nucleic acid constructs with assistance from B.S. K.P. prepared the hp-GU-24 sample and carried out additional NMR RD experiments. I.J.K. performed all DFT calculations. Z.W.S. assisted I.J.K. with numerical Bloch-McConnell simulations. I.J.K. and H.M.A. wrote the manuscript with critical input from B.S. and K.P.

The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper.

Supplementary Information is available online at www.nature.com/nature.

Despite growing evidence that rare tautomeric and anionic bases play important roles in nucleic acids, their occurrence, stabilities, and biological significance has remained elusive. Characterizing rare tautomeric and anionic bases in polynucleotides is a longstanding problem because such energetically unfavorable species typically exist in low abundance, for short periods of time, and involve movements of protons that are difficult to visualize at the atomic level. NMR relaxation dispersion (RD) techniques[20-22] are making it possible to characterize low-populated (populations of 0.1%-10%) transient (lifetimes of micro-to-milliseconds) states of nucleic acids[23-25] that are often referred to as 'excited states' (ES). Here, we use NMR RD to characterize transient WC-like dG•dT and rG•rU mispairs in DNA and RNA that are stabilized by rare tautomeric and anionic bases and obtain evidence that they play universal roles in misincorporation during replication and translation.

Transient WC-like dG•dT tautomer mispair

dG•dT mispairs generally adopt a distinct 'wobble' (WB) geometry (Fig. 1a) since a WC geometry results in a steric clash between imino protons (Fig. 1a). However, enol tautomers of dG or dT, or their anionic form, can alleviate this steric clash, and allow formation of WC-like dG•dT mispairs (Fig. 1a). Soon after the discovery of the DNA double helix, Watson and Crick hypothesized that such WC-like mispairs could provide a basis for spontaneous mutations[7]. We used NMR rotating frame spin relaxation ($R_{1\rho}$) RD[20,26,27] to examine whether wobble dG•dT mispairs can transiently morph into such WC-like dG•dT mispairs in canonical DNA duplexes. For these studies we used a hairpin DNA duplex (hp-GT DNA) containing a site-specifically $^{13}C/^{15}N$-labeled dG•dT wobble mispair (Fig. 1b and Extended Data Fig. 1). Exchange between WB and WC dG•dT mispairs entails deprotonation of either dG-N1 or dT-N3 via tautomerization (neutral) or ionization (charged), both of which would induce large changes in N1/3 chemical shifts (CSs) and therefore give rise to significant ^{15}N RD. In contrast, because a WB-to-WC transition preserves an *anti* base and C2′-endo sugar pucker, it is expected to induce smaller changes in the sugar (dG-C1′ and dT-C1′) and base (dG-C8 and dT-C6) carbon CSs and therefore induce more limited ^{13}C RD. Indeed, we observed very significant ^{15}N RD at base imino dG-N1 and to a lesser extent at dT-N3, much less significant ^{13}C RD at base dG-C8 and dT-C6, and essentially no ^{13}C RD at sugar dG-C1′ and dT-C1′ at pH 6.9 and 25 °C (Fig. 1c and Extended Data Fig. 2-3). This unique pattern of RD is consistent with exchange directed toward a transient WC-like mispair (Fig 1a). It is inconsistent with exchange directed toward other base pair (bp) geometries such as Hoogsteen[24] or base opened states (Supplementary Discussion 1 and Extended Data Fig. 4). A second exchange process was apparent at pH 8.4 (Fig. 1d) and this will be discussed further below. Similar RD profiles were observed in a different DNA duplex (Extended Data Fig. 1-2), indicating that the observed dG•dT exchange occurs robustly in DNA duplexes.

The RD data measured at dG-N1, dT-N3, dG-C8 and dT-C6 could be globally fitted (Supplementary Table 1 and Extended Data Fig. 2-3) to a single exchange process directed toward an excited state (ES1) that has a population (p_{ES1}) of ~0.17% and a lifetime (τ_{ES1}) of ~0.38 ms (Fig. 2a). ES1 is characterized by ^{15}N CSs that are significantly downfield shifted for dG-N1 ($\Delta\omega_{N1}$ +36 ppm) and to a lesser extent dT-N3 ($\Delta\omega_{N3}$ +18 ppm)(Fig. 2b, Extended Data Fig. 4). The downfield shifted imino nitrogen CSs are unprecedented for nucleic acids,

and are directed toward the CSs of bases (dG and dT) that have been deprotonated due to ionization or modifications that lock an enol-like form ($\Delta\omega$ ~50-60 ppm)(Extended Data Fig. 5)[28-30]. On the other hand, ES1 features much smaller changes in carbon CSs (Fig. 2b), consistent with a WB-to-WC transition.

It would be highly energetically disfavored to simultaneously deprotonate dG-N1 and dT-N3 when forming ES1. Moreover, though the magnitude of the ES1 N1/N3 downfield CSs strongly suggests deprotonation, it is not as far downfield shifted as expected based on deprotonation of nucleotides in free solution (Extended Data Fig. 5). Therefore, the strongly but incompletely downfield shifted dG-N1 and dT-N3 CSs suggest that ES1 consists of at least two WC-like species in rapid exchange on the NMR timescale in which either dG-N1 (dG^{enol}•dT or dG^{-}•dT) or dT-N3 (dG•dT^{enol} or dG•dT^{-}) is deprotonated (Fig. 1a). The ES1 population and CSs are largely independent of pH within the pH range of 6.0-7.9 (Fig. 2a and Supplementary Discussion 2). This is inconsistent with exchange directed toward ionic dG^{-}•dT and dG•dT^{-} (Extended Data Fig. 5). Rather, the population of ES1 increases with temperature (Fig. 2a) as expected for tautomeric species dG^{enol}•dT and dG•dT^{enol}.

Based on the measured p_{ES1} and τ_{ES1}, the free energy difference (ΔG) between GS and ES1 is ~3.8 kcal/mol and the forward free energy barrier ($\Delta G^{\ddagger}$) is ~16.4 kcal/mol (Extended Data Fig. 6). These values are in good agreement with computationally predicted parameters (2.8-5.6 kcal/mol[31] and ~17-21 kcal/mol[31,32], respectively) for dG•dT WB-to-WC tautomer transitions.

These computational studies also predict that dG^{enol}•dT likely exists in fast exchange on the NMR timescale (free energy barrier ~5-6 kcal/mol[31]) with a minor dG•dT^{enol} (20%) species (Fig. 1a and Supplementary Discussion 3-4). Under these conditions, the measured ES1 CSs would represent a population-weighted average of the two tautomeric states (Methods). We find that all ES1 CSs (dG-N1, dT-N3, dG-C8 and dT-C6) are in quantitative agreement with values predicted by density functional theory (DFT)[33] calculations for a weighted dG^{enol}•dT(80%)⇋dG•dT^{enol}(20%) equilibrium (Fig. 2c, 3a).

Transient WC-like dG•dT^{-} anionic mispair

Interestingly, upon increasing the pH to 8.4, we observed evidence for a second excited state (ES2), which is seen as a second peak in the off-resonance RD profile of dT-N3 (Fig. 1d). Global fitting of this RD data (Extended Data Fig. 2-3) revealed two excited states (ES1 and ES2) that are most likely arranged in a linear topology (ES1⇋dG•dT⇋ES2).

Compared to ES1, ES2 ($\Delta G^{\ddagger}$ 16.4 and ΔG 4.59 kcal/mol) has a considerably lower population (p_{ES2} ~0.04%) and lifetime (τ_{ES2} ~70 μs) at pH 8.4 (Fig. 2a). The dG-N1 and dT-N3 ES2 CSs are not only ‘swapped’ relative to ES1 such that dT-N3 experiences the larger downfield shift ($\Delta\omega_{N3}$ +56 ppm) while dG-N1 experiences a smaller downfield shift ($\Delta\omega_{N1}$ +9 ppm)(Fig. 2b); they are also more asymmetric in favor of a deprotonated dT species. In addition, unlike p_{ES1}, p_{ES2} increases significantly with pH, consistent with ionization and the formation of dG•dT^{-} (Fig. 3a). The ES2 CSs are in excellent agreement with values computed using DFT assuming a WC-like dG•dT^{-} (100%) species (Fig. 3a-b). However, we cannot rule out that dG•dT^{-} is in rapid equilibrium with a WC-like dG^{-}•dT or an inverted

wobble (iWB) geometry[3,34,35] (Extended Data Fig. 4 and Supplementary Discussion 1) that falls outside detection limits (Fig. 3a).

Fingerprinting the dG•dT excited states

We adapted a mutate-and-CS fingerprint strategy[24,25] to test the proposed ES1 and ES2 (Fig. 3a). Here, chemical modifications are used to trap an ES, or induce specific perturbations to the GS⇌ES equilibrium. We trapped ES1 (dG^{enol}•dT) using the mutagenic base O^6-methyl-2′-deoxyguanosine (^{m6}dG)(Fig. 3c and Extended Data Fig. 7) which is known to adopt a distorted WC dG^{enol}•dT-like mispair[12,30]. Relative to the WB, this modification resulted in negligible changes in dG-C1′ ($\Delta\omega_{C1'}$ −0.1 ppm) and dT-C1′ ($\Delta\omega_{C1'}$ +0.7 ppm) CSs, and a small downfield shift in dG-C8 ($\Delta\omega_{C8}$ +1 ppm), consistent with the RD-derived ES1 CSs (Fig. 3b and Extended Data Fig. 7). The modification induced a small upfield shift in dT-C6 ($\Delta\omega_{C6}$ −0.5 ppm) that is inconsistent with the downfield dT-C6 CS (Fig. 3b) observed by RD. However, such a deviation is expected based on DFT calculations ($\Delta\omega_{C6}$ −2.3 ppm for dT-C6 in the ^{m6}dG•T pair)(Supplementary Discussion 5) and can be attributed to minor deviations from an ideal dG^{enol}•dT WC-like mispair geometry (Fig. 3c)[12,30]. Severe line broadening did not permit measurement of the dT-N3 CSs in these non-isotopically enriched samples.

To test the proposed dG•dT$^-$ ES2, we measured the difference in dT-N3 CS between neutral and anionic dTTP ($\Delta\omega_{N3}$ +55 ppm) and found them to be in excellent agreement with the dT-N3 CS differences measured by RD ($\Delta\omega_{N3}$ +56 ppm)(Fig. 3b and Extended Data Fig. 5). In addition, we used the mutagenic thymidine-analogue, 5-bromo-2′-deoxyuridine (^{5Br}dU) to push the equilibrium toward dG•^{5Br}dU$^-$ (Fig. 3d). This modification lowers the pK_a of ^{5Br}dU-N3 (~8.6) and favors a WC-like dG•^{5Br}dU$^-$ geometry at high pH[11]. This modification increased the population of ES2 ($\Delta G^{\ddagger}$ 15.1 and ΔG 3.37 kcal/mol) by over two orders of magnitude at the expense of ES1 ($\Delta G^{\ddagger}$ 16.0 and ΔG 4.67 kcal/mol) while minimally affecting the ES1 and ES2 CSs (Fig. 3b). The consistencies in ES2 CSs between dG•dT$^-$ and dG•^{5Br}dU$^-$ mispairs further support a dominant WC-like ES2, rather than an iWB bp, in naked DNA. The unaffected ES1 CSs indicate that ^{5Br}dU does not significantly impact the dG^{enol}•^{5Br}dU⇌dG•^{5Br}dU^{enol} equilibrium (Fig. 3d) relative to dG•dT (Fig. 3a), consistent with prior computational studies[31,36].

Transient WC-like rG•rU mispairs in RNA

If the observed ESs correspond to WC-like dG•dT mispairs, one would predict that similar ESs should arise in rG•rU mispairs in RNA where WC bps are also readily accommodated within the A-form helix. To test this hypothesis, we carried out analogous pH- and temperature-dependent RD measurements on two RNA duplexes. RD profiles measured for rG•rU in A-form RNA (Fig. 4a and Extended Data Fig. 8) are very similar to those measured for dG•dT in B-form DNA (Fig. 1c-d). Global analyses of the RD data (Supplementary Table 1) revealed an apparent 3-state exchange process at pH ≥7.9 (Extended Data Fig. 3). The RD-derived CSs (Fig. 4b and Extended Data Fig. 9), together with the pH and temperature dependence of the populations and lifetimes (Fig. 4c), are consistent with rG•rU^{enol}⇌rG^{enol}•rU as ES1 and rG•rU$^-$ as ES2 (Fig. 4d) as observed in

DNA. The rG•rU ES1 forward free energy barrier ($\Delta G^{\ddagger}$ 15.7 and ΔG 3.86 kcal/mol) is consistent with barriers measured for dG•dT ES1 (16.4 kcal/mol) and computationally predicted barriers for G•U tautomerization (17.1 kcal/mol)[32].

Computational studies[36] show that dG•dU^{enol} is stabilized relative to dG•dT^{enol}. This is predicted to tilt the rapid rG^{enol}•rU$\rightleftharpoons$rG•rU^{enol} equilibrium in favor of rG•rU^{enol} (40%) in RNA as compared to dG•dT^{enol} (20%) in DNA (Methods). We find that in RNA, the ES1 rU-N3 CSs are slightly more downfield shifted ($\Delta\omega_{N3}$ +30 ppm) than rG-N1 ($\Delta\omega_{N1}$ +26 ppm)(Fig. 4b). Reweighting the DFT-predicted CSs assuming 60:40 ratios of rG^{enol}•rU:rG•rU^{enol} gives an excellent fit to RD-derived values (Fig. 4b), and are in better agreement than 80:20 rG^{enol}•rU:rG•rU^{enol} (Extended Data Fig. 9). We note that we observe changes in the ES1 CSs at higher pH that suggest a potentially more complex exchange process (Supplementary Discussion 6). As with dG•dT^{-}, the WC-like rG•rU^{-} may exist in equilibrium with both a WC-like rG^{-}•rU and/or an iWB rG•rU^{-} (Fig. 4d).

rG•rU wobbles are widespread in RNA where they play important structural and functional roles[37]. We therefore examined whether the ESs observed here would occur in more complex RNA structural contexts. Indeed, the rU-N3 ^{15}N RD profiles measured for two wobble rG•rU mispairs in a 69-nt *Bacillus subtilis* guanine riboswitch (Extended Data Fig. 8) provide initial evidence ($\Delta\omega_{N3}$ +44-47 ppm and p_{ES} ~0.04% at pH 7.9) for WC-like rG•rU^{-} mispairs in more complex RNA structures (Extended Data Fig. 9 and Supplementary Discussion 7). Therefore, we can expect that transient WC-like rG•rU mispairs exist robustly across the RNA transcriptome.

Biological implications

Initial selection of NTPs during replication, and tRNAs during translation, strongly relies on WC stereochemical geometry as a means of discriminating against mispairs[38,39]. The low error rate (10^{-3}-10^{-6}) during initial selection accounts for most of the overall fidelity of replication (~10^{-6}-10^{-10})[40-42] and translation (~10^{-3}-10^{-5})[43-45]. By stereochemically mimicking the WC geometry, the ES WC-like dG•dT and rG•rU mispairs observed here can provide a mechanism for evading WC fidelity checks during initial substrate selection[1,8,10]. The intrinsic probabilities with which WC-like mispairs form have long been suspected as important determinants of substitution mutation probability[8,10]. By carrying out the first measurements of the intrinsic probabilities with which WC-like dG•dT and rG•rU mispairs form in native DNA/RNA systems, we are able to obtain unique insights into the mechanisms of misincorporation and the potential roles of ES1/ES2.

We find that the probabilities with which WC-like dG•dT ES1 and ES2 form in duplex DNA (10^{-3}-10^{-5}) span the dGTP•dT/dG•dTTP misincorporation and base substitution probabilities seen during replication using high-fidelity polymerases with little/no proof-reading capabilities (Fig. 5a, Extended Data Fig. 10 and Supplementary Discussion 8)[46-48]. Similarly, the WC-like rG•rU ES1 and ES2 probabilities (10^{-3}-10^{-4}) span the majority of amino acid misincorporation probabilities arising specifically due to rG•rU pairing at any codon position (10^{-3}-10^{-5})[43,44] (Fig. 5b and Supplementary Discussion 8). Note that some of the amino acid misincorporation probabilities (10^{-5}) are lower than the measured rG•rU

ES2 probability, which could be due to translational proofreading[45] and/or lower pH conditions that destabilize ES2. These results, together with previous structural studies showing that WB and WC-like mispairs can exist within polymerase[1-3] and ribosome[5,6,13] active sites, strongly suggest that energetic competition between WB and WC-like mispairs is robust and is a key determinant of misincorporation probability during replication and translation (Supplementary Discussion 9). However, additional studies are needed to determine the probabilities with which WC-like mispairs form within the unique environment of polymerases and ribosomes. A recent MD study found that rG^{enol}•rU tautomers can be stabilized in a ribosome context, but challenges their involvement in decoding errors based on predicted tRNA binding energies[49].

The pH-dependent dG•dT misincorporation probability[47,48,50] points to the involvement of an anionic species in misincorporation[1]. Our results strongly suggest that this species is most likely ES2 dG•dT^- and not the energetically disfavored dG^-•dT. We observe excellent agreement between the pK_a-predicted probability of forming ES2 dG•dT^- and pH-dependent dGTP•dT/dG•dTTP misincorporation probabilities measured for a reverse transcriptase[47] which lacks any proofreading ability (Fig. 5c and Extended Data Fig. 10). We note that the correlation is reduced at more extreme pH, near the pK_as of other ionizable groups in proteins and DNAs (Extended Data Fig. 10 and Supplementary Discussion 10). We also find that dG•$^{5Br}dU^-$ enhances the probability of forming a WC-like ES2 (Fig. 3d) and correspondingly results in an ~8-fold increase dG•^{5Br}dU misincorporation[47]. These data suggest that for this polymerase, misincorporation proceeds predominantly via a WC-like dG•dT^-. It is very likely that parameters such as polymerase types, DNA sequence, and the chemical environment can affect the relative stabilities and lifetimes of the anionic, tautomeric, and WB mispairs[3]. Therefore we can expect this to affect the flux through distinct misincorporation pathways involving different WC and WB conformations, which may help to explain the broad range of misincorporation probabilities.

Our findings suggest that unconventional enol tautomeric and anionic bases exist robustly in genomes. We anticipate these rare tautomeric and anionic bases play unique roles in DNA damage induction and repair, nucleic acid recognition, chemical modifications of nucleic acids, and catalysis. The NMR methods outlined here can immediately be applied to characterize tautomeric and anionic species, which we believe will not be restricted to dG•dT and rG•rU mispairs, but rather will be widespread across diverse nucleic acid motifs.

METHODS

Sample Preparation

NMR buffer—All duplex DNA and RNA samples were buffer exchanged using a centrifugal concentrator (EMD Millipore) into a solution containing 25 mM sodium chloride (100 mM sodium chloride for Dickerson-GT DNA), 15 mM sodium phosphate, 0.1 mM EDTA, and 10% D_2O with variable pH (6.0, 6.4, 6.8, 6.9, 7.9). pH values of 8.4 were obtained for hp-GT DNA and hp-GU-20 RNA samples by direct titration of pH 7.9 samples with filtered 0.5 M NaOH solution. Monovalent ion concentration subsequently increased by a small amount proportional to the NaOH titrated in but did not affect DNA/RNA conformation as confirmed by NMR. Natural isotopic abundance oligonucleotide sample

concentrations ranged from 2-3.5 mM. $^{13}C/^{15}N$-labeled oligonucleotide sample concentrations ranged from 0.7-3.5 mM. xpt-G riboswitch sample was diluted to ~30 μM in a solution containing saturated guanine, denatured, and annealed on ice. Sample was buffer exchanged against either potassium acetate (pH 6.7) or sodium phosphate (pH 7.9) buffer and concentrated to 0.7-1.7 mM. Mg^{2+} was titrated in until total concentration was ~5 mM.

Site-specifically $^{13}C/^{15}N$-labeled DNA samples—Selectively $^{13}C/^{15}N$-labeled DNA samples (hp-GT DNA, ^{5Br}U5-hp-GT and ^{8Br}G15-hp-GT DNA) were purchased from the Yale Keck Oligonucleotide Synthesis Facility and were synthesized using commercially available 2′-deoxyguanosine DMT-phosphoramidite (98% ^{13}C10, 98% ^{15}N5) and 2′-deoxythymidine phosphoramidite (98% ^{13}C10, 98% ^{15}N2) purchased from Cambridge Isotope Labs. hp-GT DNA was selectively $^{13}C/^{15}N$-labeled at dT5 and dG15, ^{5Br}U5-hp-GT DNA was $^{13}C/^{15}N$-labeled at dG15, and ^{8Br}G15-hp-GT was $^{13}C/^{15}N$-labeled at dT5. Samples were purified using RP-HPLC prior to buffer exchange. NMR experiments were used to confirm native folding of hp-GT, ^{5Br}U5-hp-GT, and ^{8Br}G15-hp-GT DNA constructs.

Enzymatic synthesis of $^{13}C/^{15}N$-labeled DNA samples—The Dickerson-GT DNA sample was synthesized *in vitro* using uniformly enriched $^{13}C/^{15}N$ dGTP and dTTP (Silantes GmbH) as previously described[51]. Reaction mixture was centrifuged and filtered to remove excess pyrophosphate and concentrated down to 1 mL in a 3 kDa molecular weight cut-off centrifugal concentrator (EMD Millipore). Sample was mixed with 1 mL of a formamide-based denaturing loading dye, denatured at 95 °C for 5 min, and loaded onto a 33×102 cm sequencing gel (20% polyacrylamide/8M urea) and run for 12 hours to resolve target oligonucleotide from template and other nucleic acid species. Target band was shadowed briefly using a UV hand-lamp and excised prior to gel electroelution (Whatman, GE Healthcare), followed by ethanol precipitation. Sample purity was confirmed using gel electrophoresis (20% polyacrylamide/8M urea) stained with SybrGOLD prior to buffer exchange.

Enzymatic synthesis of $^{13}C/^{15}N$-labeled RNA samples—The hp-GU-20, hp-GU-24 and xpt-G riboswitch RNA samples were prepared using *in vitro* transcription as previously described[25] using uniformly enriched $^{13}C/^{15}N$ ribonucleotide triphosphates (hp-GU-20 RNA: rGTP & rUTP only, hp-GU-24: All and xpt-G: All). Purification was carried out as described above for $^{13}C/^{15}N$-labeled DNA construct.

Unlabeled and unlabeled-modified DNA samples—hp-GT and Dickerson-GT DNA constructs at natural isotopic abundance were purchased from Integrated DNA Technologies. The O^6-Methyl-2′-deoxyguanosine mismatch constructs (^{m6}G15-hp-GT and ^{m6}G4-Dickerson-GT DNA) were purchased from the Yale Keck Oligonucleotide Synthesis Facility. hp-GT and Dickerson-GT DNA constructs were desalted prior to buffer exchange. Unlabeled-modified samples were purified using RP-HPLC prior to buffer exchange.

Isotopically enriched dNTP and rNTP samples—Uniformly $^{13}C/^{15}N$ enriched dGTP, dTTP, rGTP and rUTP samples were purchased (Silantes GmbH) and added to an NMR

buffer (25 mM sodium chloride, 15 mM sodium phosphate, 0.1 mM EDTA and 10% D_2O at pH 6.9). Sample was adjusted to pH ~12.5 directly using 5 M NaOH.

NMR experiments

Resonance assignment—The CS assignment for DNA and RNA constructs were obtained using aromatic [^{13}C, ^{1}H], aliphatic [^{13}C, ^{1}H], imino [^{15}N, ^{1}H] heteronuclear and [^{1}H, ^{1}H] NOESY homonuclear correlation experiments. The data for all DNA constructs were acquired on an 18.8T Agilent spectrometer equipped with a triple resonance HCN cryogenic probe, for the uniformly ^{13}C/^{15}N-labeled hp-GU-24 RNA construct on a 14.1T Bruker Avance spectrometer equipped with a triple-resonance HCN cryogenic probe, and for the xpt-G RNA riboswitch on a 14.1T Agilent spectrometer equipped with a Bruker HCPN cryogenic probe. All data were processed and analyzed using the software NMRpipe[52] and SPARKY (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco), respectively. Resonance assignment for exchangeable and non-exchangeable ^{1}H was performed using the 2D [^{1}H, ^{1}H] WATERGATE NOESY experiment[53] (mixing time 250 ms) as described previously[27,54], and their covalently bonded ^{13}C/^{15}N nuclei were assigned using heteronuclear single/multiple quantum coherence correlation experiments (HSQC or HMQC). For the labeled RNA constructs conventional HSQC experiments[53] were acquired for all spins, while for unlabeled DNA constructs conventional HSQC was used for the aliphatic C1′ spins and SOFAST-HMQC[55,56] were employed for the imino and aromatic spins.

^{15}N $R_{1\rho}$ relaxation dispersion—1D ^{15}N $R_{1\rho}$ RD experiments[27,57] targeting imino nitrogen resonances of interest were carried out at 14.1T (hp-GT, ^{5Br}U5-hp-GT, Dickerson-GTDNA and hp-GU-24 RNA) and 16.4T (hp-GU-20 RNA) as previously described[57]. Raw data were processed using NMRpipe[52] to generate a series of peak intensities. On- and off-resonance $R_{1\rho}$ RD profiles were recorded using spinlock powers (ω_{SL} $2\pi^{-1}$) ranging from 100-2000 Hz, with the absolute offset frequencies (Ω $2\pi^{-1}$ Hz) ranging from 0-3.5× the applied spinlock power (Supplementary Table 1). Offset frequencies greater than 3.5× the given spinlock power were not used owing to significant R_1 relaxation contributions[27]. Magnetization of the spins of interest was allowed to relax under an applied spinlock for the following durations: [0 – 120 ms] for N1/N3 of hp-GT, ^{5Br}U5-hp-GT, Dickerson-GT DNA and hp-GU-24 RNA and [0 – 100 ms] for N1/N3 of hp-GU-20 RNA, and [0 – 80 ms], [0 – 74 ms], [0 – 68 ms] for N3 of the xpt-G riboswitch.

^{13}C $R_{1\rho}$ relaxation dispersion—^{13}C $R_{1\rho}$ RD experiments targeting carbon resonances of interest were carried out at 14.1T as previously described[24,27]. On- and off-resonance $R_{1\rho}$ RD profiles were measured using spinlock powers (ω_{SL}) ranging from 150-3500 Hz, with the absolute offset frequencies (Ω $2\pi^{-1}$, Hz) ranging from 0-3.5× the applied spinlock power (Supplementary Table 1). Magnetization of the spins of interest were allowed to relax under an applied spinlock for the following durations: [0 – 60 ms] for C1′/C6/C8 of hp-GT DNA and C1′ of Dickerson-GT, [0 – 50 ms] for C6/C8 Dickerson-GT DNA.

^{13}C/^{15}N CSs of ionized dNTPs and rNTPs—Carbon and nitrogen CSs of neutral and deprotonated dNTPs and rNTPs were collected at 25 °C and pH ~6.9 and ~12.5 using a 2D

[^{15}N, ^{13}C] HMQC experiment on a 14.1T Agilent spectrometer equipped with a Bruker HNCP cryoprobe.

Analysis of $R_{1\rho}$ data

Fitting of ^{13}C and ^{15}N $R_{1\rho}$ data—$R_{1\rho}$ values were calculated by fitting the decay of peak intensity versus relaxation delay to a monoexponential[58]. Errors in $R_{1\rho}$ were estimated using spectral noise and duplicate $R_{1\rho}$ data points using a suite of Mathematica notebooks[58]. Measured on- and off-resonance $R_{1\rho}$ data were globally fit to algebraic equations describing N-site chemical exchange using a Levenberg-Marquardt method, weighted to the experimental error in the $R_{1\rho}$ data. For 2-state exchange, data was fit to the Laguerre equation (1)[59] under the valid assumption that the populations are highly asymmetric, such that $p_{GS} >> p_{ES}$ where $p_{ES} < 0.3$. For the 3-state chemical exchange model where $k_{BC} = k_{CB} = 0$, data was fit to both the 3-state Laguerre equation (2) as well as the general 3-state equation (3)[59], where $p_{GS} >> p_{ES}$ and $p_{ES} < 0.1$. Fitted parameters derived from Eqn. 2 and Eqn. 3 are in excellent agreement with one another (Supplementary Table 1). Statistical tests, F-Test and Akaike information criterion (AIC), were used to select the best-fit N-state exchange model[24] (Extended Data Fig. 3). ^{13}C and ^{15}N RD data from the dG•dT or rG•rU mispair resonances for each construct, at each temperature and pH condition were fitted globally (where k_{ex} and p_B are the shared-parameters) when possible. In the cases where ^{15}N RD data was available but ^{13}C RD data showed no chemical exchange, did not exhibit 3-state exchange, or was not collected, the ^{15}N N1/N3 RD data were globally fitted as described above. Bloch-McConnell[60] (B-M) numerical simulations were used to validate the algebraic approximations' for 2- and 3-state exchange scenarios.

(1) 2-state Laguerre equation[59,61]:

$$R_{1\rho}=R_1\,cos^2\,\theta+R_2\,sin^2\,\theta+\frac{sin^2\,\theta\,p_{GS}p_{ES1}\Delta\omega_{ES1}^2k_{ex1}}{\left\{\omega_{GS}^2\omega_{ES1}^2/\omega_{eff}^2+k_{ex1}^2-sin^2\,\theta\,p_{GS}p_{ES1}\Delta\omega_{ES1}^2\left(1+\frac{2k_{ex1}^2\left(p_{GS}\omega_{GS}^2+p_{ES1}\omega_{ES1}^2\right)}{\omega_{GS}^2\omega_{ES1}^2+\omega_{eff}^2k_{ex1}^2}\right)\right\}}$$

(2) 3-state Laguerre equation with no minor exchange[25]:

$$R_{1\rho}=R_1\,cos^2\,\theta+R_2\,sin^2\,\theta+sin^2\,\theta\left(\frac{p_{GS}p_{ES1}\Delta\omega_{ES1}^2k_{ex1}}{\left\{\omega_{GS}^2\omega_{ES1}^2/\omega_{eff}^2+k_{ex1}^2-sin^2\,\theta\,p_{GS}p_{ES1}\Delta\omega_{ES1}^2\left(1+\frac{2k_{ex1}^2\left(p_{GS}\omega_{GS}^2+p_{ES1}\omega_{ES1}^2\right)}{\omega_{GS}^2\omega_{ES1}^2+\omega_{eff}^2k_{ex1}^2}\right)\right\}}+\frac{\left\{\omega_{GS}^2\omega^2\right.\right.$$

(3) 3-state general equation with no minor exchange[59,62]:

$$R_{1\rho}=R_1\,cos^2\,\theta+R_2\,sin^2\,\theta+sin^2\,\theta\left(\frac{k_{GS\rightarrow ES1}\Delta\omega_{ES1}^2}{\Omega_{ES1}^2+\omega_{SL}^2+k_{ES1\rightarrow GS}^2}+\frac{k_{GS\rightarrow ES2}\Delta\omega_{ES2}^2}{\Omega_{ES2}^2+\omega_{SL}^2+k_{ES2\rightarrow GS}^2}\right)$$

in which R_1 and R_2 are the intrinsic longitudinal and transverse relaxation rates (s^{-1}). The exchange rates are defined as $k_{exi} = k_{GS\rightarrow ESi} + k_{ESi\rightarrow GS}$, where $k_{GS\rightarrow ESi} = p_{ESi}k_{exi}$ and $k_{ESi\rightarrow GS} = p_{GS}k_{exi}$ and where $i = 1$ or 2. The CS difference between the GS and ESs is given by $\Delta\omega_{ESi} = \Omega_{ESi} - \Omega_{GS}$, where $\Omega = \Omega_{obs} - \omega_{rf}$ defines the resonance offset from the carrier

frequency (ω_{rf}), $\Omega_{obs} = (\Omega_{GS}p_{GS} + \Omega_{ES1}p_{ES1})$ or $\Omega_{obs} = (\Omega_{GS}p_{GS} + \Omega_{ES1}p_{ES1} + \Omega_{ES2}p_{ES2})$, and where $\Omega_{obs} \approx \Omega_{GS}$ when $p_{GS} >> p_{ESi}$, as is the case in the ESs measured. The average effective spinlock field in the rotating frame is given by $\omega^2_{eff} = \Omega^2 + \omega^2_{SL}$ and $\omega^2_{GS} = (\Omega_{GS} - \omega_{rf})^2 + \omega^2_{SL}$, $\omega^2_{ESi} = (\Omega_{ESi} - \omega_{rf})^2 + \omega^2_{SL}$ and ω_{SL} is the spinlock power. The tilt angle in the rotating frame is given by $\theta = \arctan(\omega_{SL}/\Omega)$.

Analysis of the 3-state exchange model—We repeated measurements of RD profiles for hp-GT dG-N1 and dT-N3 at pH 8.4 and 25 °C on a different spectrometer and also obtained data that is consistent with 3-state exchange (Extended Data Fig. 2 and Supplementary Table 1). In addition, we collected one additional spinlock power (800 Hz) for both resonances, and find that the inclusion of this extra spinlock power has negligible effect on the fitted exchange parameters indicating that they are robustly determined by the measured data. We note that differences in the RD profiles and fitted parameters between spectrometers are largely within error, with minor differences likely arising due to small differences in temperature and/or spinlock calibrations.

The dT-N3 resonance of hp-GT DNA at 25 °C and pH 7.9 was also individually refit assuming both 2-state and 3-state exchange. The individual dT-N3 3-state fit gave very similar fitted parameters as the dT-N3 in the 3-state global fit with dG-N1. The F-test (at 0.05 significance level) favored the 3-state individual fit model over the 2-state individual fit model. AIC also favored the 3-state model, estimating the 3-state model to be 3.1×10^{15} times more likely to be the correct model than the 2-state model. The individual 2-state and 3-state fits to dG-N1 RD profiles give similar agreement (reduced χ^2 ~0.83) when the ES2 CS is fixed based on the globally fitted value; however, statistical tests slightly favor the simpler model (AIC favors 2-state by approximately two-fold).

The 3-state model is supported by statistical tests (F-test and AIC favor the 3-state over 2-state exchange model) and also by B-M simulations (data not shown) for the hp-GT dG-N1 and dT-N3 data at pH 7.9. Here, $R_{1\rho}$ values were simulated, noise corrupted, and evaluated for the ability to report on the 3-state exchange following the procedure reported in Bothe *et al.*[63].

Bloch-McConnell 3-state numerical simulations—Parameters derived from the algebraic 3-state fits, along with the ω_{rf} and ω_{SL}, were used to simulate numerical solutions to the 3-state B-M equations[60] (Extended Data Fig. 3). The B-M simulations were carried out using a procedure similar to that described recently for 2-state exchange[63]. Simulations were carried out assuming a 0.25 second relaxation delay under the applied ω_{SL}.

Thermodynamic analysis of $R_{1\rho}$ RD-derived parameters—Temperature-dependent analysis of forward and reverse exchange rates in the site-labeled (dG15•dT5 $^{13}C/^{15}N$) hp-GT DNA and G/U labeled hp-GU-20 RNA samples were carried out as previously described[24]. This analysis employed the ES populations and exchange rates obtained based on 2-state global fitting of dG15-N1 and dT5-N3 RD data measured in hp-GT DNA at 10, 20, 25 and 30 °C. For hp-GU-20 RNA, the analysis employed populations and exchange rates obtained based on 2-state and 3-state global fitting of rG16-N1 and rU5-N3 RD data measured at 10, 20, 25 and 30 °C at pH 6.9. Errors in the fitted thermodynamic parameters

are given by weighted fits of the modified van't Hoff equation[24] to the RD-derived forward and reverse exchange rates and their errors. The NMR sample temperature was calibrated using 99.8% methanol-d_4 (Cambridge Isotope Laboratories) using the equation $T = -16.7467(\Delta\delta)^2 - 52.5130(\Delta\delta) + 419.1381$, where $\Delta\delta$ is the difference in CS (ppm) between the hydroxyl and methyl proton[64].

Density functional theory geometry optimizations and CS calculations

All DFT calculations[33] were performed using Gaussian 09c (Gaussian, Inc.)[65] and carried out on the University of Michigan's Advanced Research Computing HPC cluster, Flux, as previously described[24] with minor modifications to DFT method and basis set used. Geometry optimizations were carried out using the empirical exchange-correlation functional, M06-2X[66], with the 6-31+G(d,p) basis-set. The ^{13}C and ^{15}N isotropic magnetic shielding (σ_{13C} and σ_{15N}) were calculated using the GIAO method with M06-2X/6-31+G(d,p). CSs of the nucleobases (δ_{13C} and δ_{15N}) were calculated by $\delta_{13C/15N} = \sigma_{13C/15N} - \sigma_{TMS/NH3}$, where σ_{TMS} and σ_{NH3} are the isotropic magnetic shieldings calculated for the reference compounds trimethylsilane (^{13}C) and NH_3 (^{15}N), respectively.

Input structures for DFT calculations—We generated idealized B-/A-form helices corresponding to our sequence contexts (hp-GT DNA and hp-GU-24 RNA) using make-NA (J. Stroud, *make-NA*, http://structure.usc.edu/make-na/server.html 2011). The duplexes were truncated to the trinucleotide step centered on the target mispair (GTG/CGC for hp-GT DNA and GUA/UGC for hp-GU-24 RNA). The sugar and phosphate moieties were removed and replaced with a methyl for i (dG•dT or rG•rU mispair), i+1, and i−1 base pairs to save on computational time. Although the structures lack a 2′-deoxyribose or ribose sugar moieties, they will be denoted as dG•dT or rG•rU to avoid confusion. All heavy atoms were frozen for the i+1 and i−1 base pairs while geometry optimizations were carried out for protons and heavy atoms of the central dG•dT or rG•rU base pair as well as the protons of the i+1/i−1 pairs. We performed full geometry optimizations on: dG•dT and rG•rU GS WB pairs, dG^{enol}•dT, $dG•dT^{enol}$, rG^{enol}•rU, $rG•rU^{enol}$, and dG^{-}•dT and rG^{-}•rU ES WC-like pairs. All converged to the expected WB or WC-like geometries. In the instance where the geometry optimizations of dG•dT^{-} and rG•rU^{-} starting states failed to converge to a stable WC-like dG•dT^{-}/rG•rU^{-} geometry (and instead converged to an iWB geometry, in vacuum), the WC-like $dG•dT^{enol}$/$rG•rU^{enol}$ states were converted to dG•dT^{-}/rG•rU^{-} and geometry optimizations were carried out on the protons only. CSs for each state were calculated and later used in calculating population weighted CSs assuming different populations of these mispair species. In addition, while the sugar moieties were truncated to methyl groups to save on computation time, it should be noted that previous DFT studies of nucleotides have shown that tautomerization, primarily of pyrimidines, can have an affect on the sugar conformation[67]. However, we can rule out large changes in sugar pucker arising in ES1 based on the negligibly small chemical exchange contributions to both dG-C1′ and dT-C1′ (see Extended Data Fig. 2) and only very small changes in dG-C1′ and dT-C1′ CSs upon locking the enol-like form with ^{m6}dG•dT ($|\Delta\omega_{C1'}| \leq 0.7$ ppm, see Fig. 3b and Extended Data Fig. 7). Finally, while the CSs of the anionic pairs are predicted assuming planar pair geometry, prior computational studies of G•T^{-} and G^{-}•T nucleobase pairs in isolation have shown that they can favor non-planar and non-WC geometry[35].

dG^{enol}•dT distance dependent DFT calculations—We carried out distance dependent DFT calculations[68] on a pair of WC-like dG^{enol}•dT N1/N9-methyl nucleobases in vacuum using the M06-2X method and 6-31+G(d,p) basis-set, as described above. The geometry of a dG•dT wobble pair and dG^{enol}•dT WC-like pair was optimized with no constraints prior to CS calculations. The ideal N1-N3 distance of the dG^{enol}•dT WC-like mispair was then manually varied from 2.44Å to 3.8Å in increments of 0.1Å from 2.44-3.04Å and then to 3.8Å (Extended Data Fig. 4). At every increment the proton positions alone were optimized and CSs were calculated relative to an optimized dG•dT wobble base pair.

Population-weighted average DFT-predicted CS calculations—Based on the computationally predicted energetic differences between interconverting dG^{enol}•dT and dG•dT^{enol} base mispairs in water (ΔG 0.7-0.8 kcal/mol)[31] and in a weakly polar medium (ΔG 0.99 kcal/mol)[69], we can predict that the dG•dT ES1 CSs represent a population weighted average between interconverting dG^{enol}•dT(~80%)⇋dG•dT^{enol}(~20%) states. Thus, the DFT-predicted CSs for dG^{enol}•dT and dG•dT^{enol} were summed in a population-weighted manner. It is noted that the computationally predicted energetic stabilities of the tautomeric states differ when calculated in water versus vacuum, or a weakly polar medium, with the values predicted in water giving the greatest agreement with our experimental results.

In the case of rG•rU ES1, computational studies have shown that dG•dU^{enol} is ~1 kcal/mol more stable than dG•dT^{enol} in a DNA fragment[36], suggesting that an rG^{enol}•rU⇋rG•rU^{enol} equilibrium should be titled slightly more towards rG•rU^{enol} than dG•dT^{enol} in DNA. We can qualitatively estimate the relative stability between rG^{enol}•rU and rG•rU^{enol} to be 60:40 based on a best fit to the RD-derived CSs.

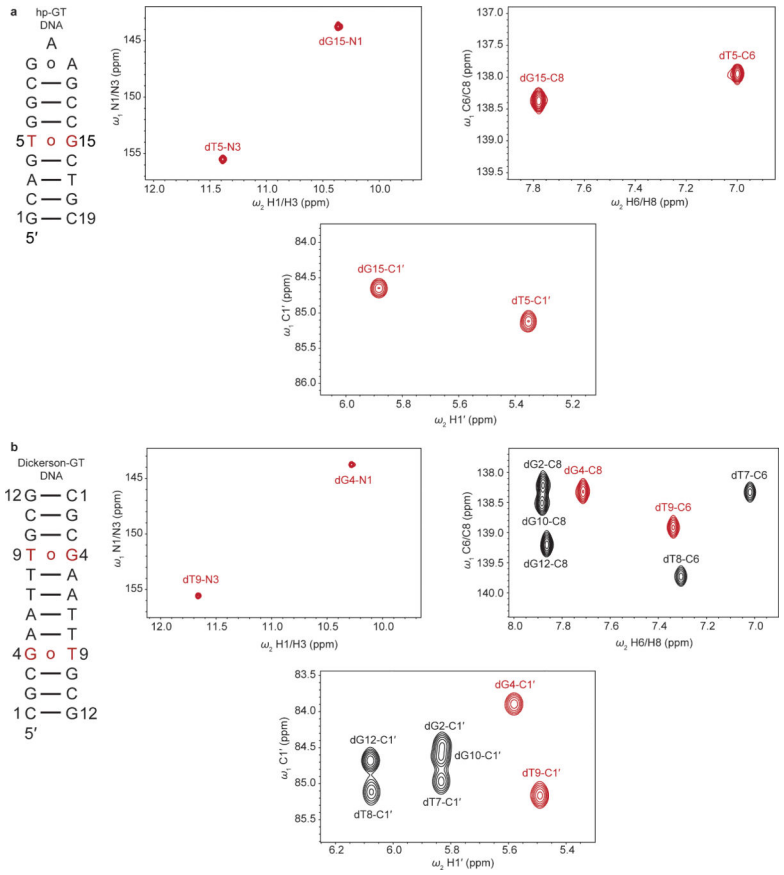
pK_a fitting and probability estimation

The apparent pK_as for hp-GT ES2, ^{5Br}U-hp-GT ES2, and dGTP•dT misincorporation (pH 6.5-8.6) were fit to the Henderson-Hasselbalch equation using a Monte-Carlo (MC) approach. Here, 10^6 p_Bs at pH 7.9 and/or 8.4 were selected from a Gaussian distribution with mean p_B value and standard deviation representing the uncertainty in p_B based on fitting of the RD data. 10^6 fits to Eqn. 4 were then carried out assuming these p_B values to generate 10^6 pK_as.

$$p_B = \left(\frac{10^{pH-pKa}}{1+10^{pH-pKa}} \right) \quad (4)$$

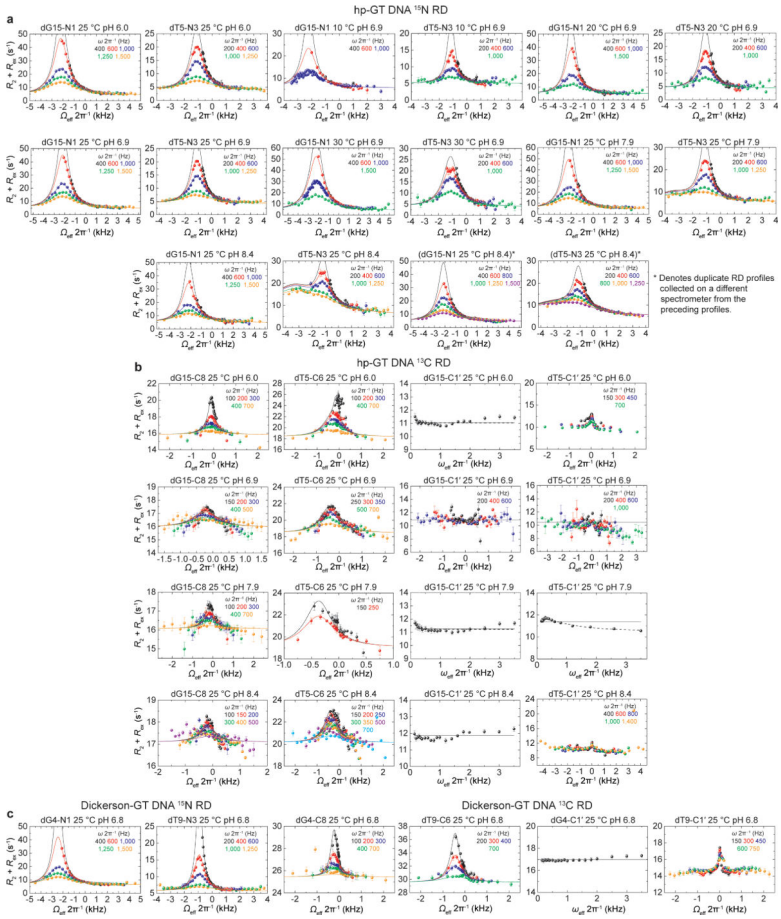
Where p_B is the probability of forming ES2 or dGTP•dT misincorporation probability at a given pH. The resulting fitted pK_a values were fitted to a Gaussian distribution. The mean value of the Gaussian distribution is the reported pK_a value and the standard deviation is assumed to be the error. An analogous approach was used to back-calculate predicted p_Bs at a given pH using the pK_a derived by the above method.

Extended Data



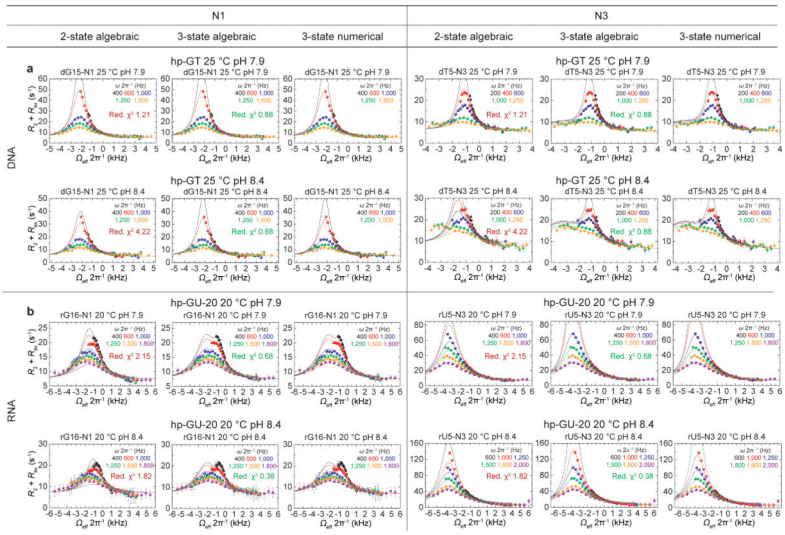
Extended Data Figure 1. NMR spectra of site- and selectively-labeled dG•dT mispair DNA constructs

a-b Shown are the **a**, hp-GT DNA and **b**, Dickerson-GT constructs with $^{13}C/^{15}N$ labeled dG•dT mispairs highlighted in red along with 2D imino [^{15}N, ^{1}H] HSQC, 2D aromatic [^{13}C, ^{1}H] HSQC and 2D C1′ [^{13}C, ^{1}H] HSQC spectra (pH 6.9, 25 °C).



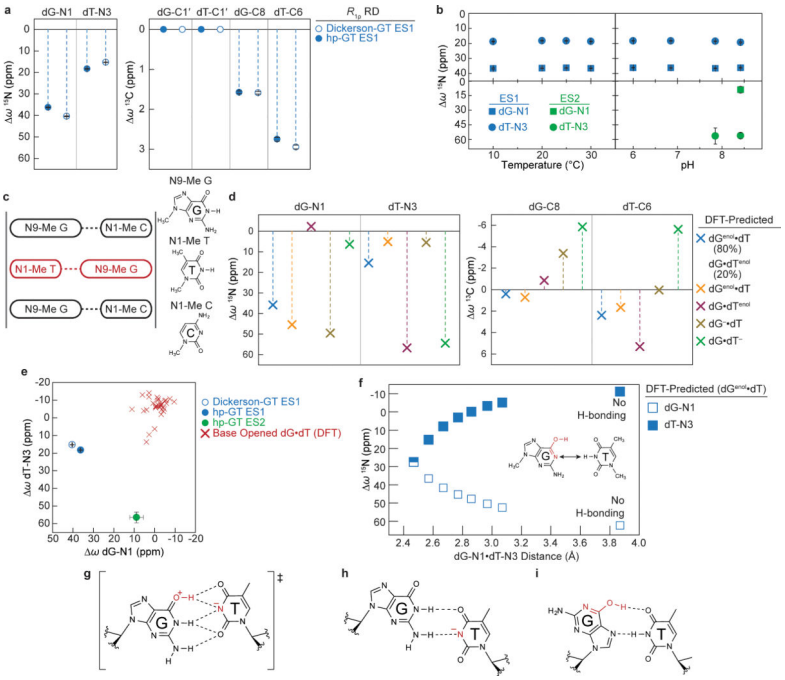
Extended Data Figure 2. Rotating frame relaxation dispersion profiles of dG•dT mispairs in hp-GT and Dickerson-GT DNA constructs

RD profiles showing chemical exchange (R_2+R_{ex}) in the dG•dT mispair as a function of the spin lock offset ($\Omega_{eff}\ 2\pi^{-1}$) and spin lock power ($\omega_{SL}\ 2\pi^{-1}$, color coded in insets). Shown are **a**, ^{15}N and **b**, ^{13}C RD profiles in hp-GT DNA in hp-GT. On-resonance profiles showing solid and dashed black lines indicate fits assuming no chemical exchange (solid) and simplified 2-state exchange process (dash). The hp-GT dG15-N1 and dT5-N3 in brackets denote duplicate profiles (with an additional 800 Hz spinlock power for each) collected at pH 8.4 and 25 °C collected on a different spectrometer from the preceding profiles. **c**, ^{15}N and ^{13}C RD profiles for Dickerson-GT. Sample conditions are indicated on each profile. Error bars represent experimental uncertainty (one s.d., see Methods).



Extended Data Figure 3. Multiple site exchange comparison and numerical solutions

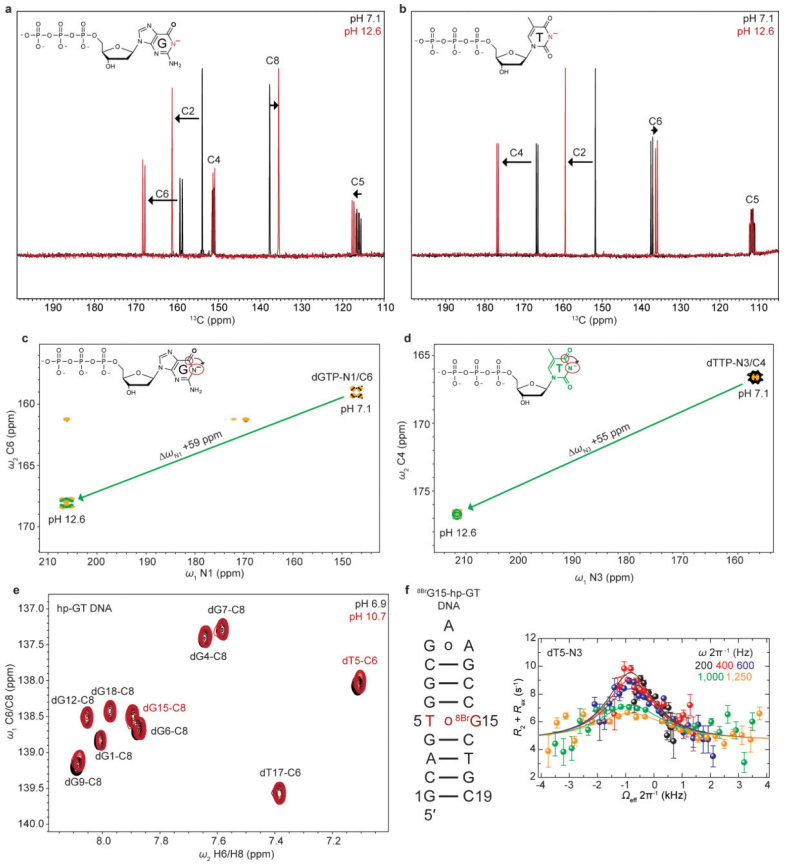
Global fitting of **a,** hp-GT DNA and **b,** hp-GU-20 RNA N1 and N3 RD profiles to 2-state algebraic equation (Eqn. 1, fit reduced χ^2 shown in inset) and 3-state algebraic equation (Eqn. 2, fit reduced χ^2 shown in inset). Numerical solutions to the Bloch-McConnell 3-state equations assuming no minor exchange and input exchange parameters obtained based on the 3-state algebraic fit are also shown to establish the validity of the 3-state expression under these exchange scenarios (Eqn. 2, see Methods). Sample conditions are indicated on each profile. Error bars represent experimental uncertainty (one s.d., see Methods).



Extended Data Figure 4. Chemical shift fingerprinting dG•dT excited states

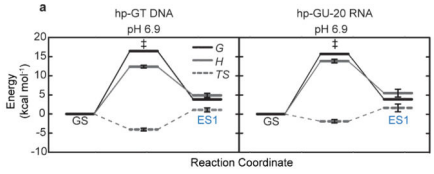
a, RD-derived dG15-N1 and dT5-N3 CSs (CSs) (referenced to GS WB) for ES1 (25 °C and pH 6.9) and ES2 (25 °C and pH 8.4) of hp-GT and ES1 of Dickerson-GT (25 °C and pH 6.9)

are shown. Errors in all RD-derived fitted parameters (eg. $\Delta\omega$) denote s.e. from the weighted global fit (see Methods). **b**, RD-derived hp-GT dG•dT ES1 (blue) and ES2 (green) ^{15}N CSs are shown as a function of temperature and pH for both dG15-N1 (square) and dT5-N3 (circle). **c**, Scheme used to calculate CSs using DFT (see Methods). Shown is a schematic representation of scenario used to for calculating CSs using DFT. Idealized B-form DNA helix is generated to give a central dG•dT mispair (red) that is flanked by canonical dG•dC pairs, analogous to the hp-GT construct. Residues are trimmed to 1-/9-methyl bases and i +1/i−1 pairs are frozen in place for subsequent geometry optimizations and NMR CS calculations. **d**, DFT-calculated CSs (referenced to an energy optimized WB geometry) are shown for various tautomeric and anionic configurations, where dG^{enol}•dT/dG•dT^{enol} represents population weighted average over dG^{enol}•dT (80%) and dG•dT^{enol} (20%). **e**, RD-derived ES1 and ES2 CSs are plotted against DFT-calculated CSs of base opened dG•dT mispairs, taken from X-ray structures and pruned to 1-/9-methyl bases. **f**, DFT-calculated CSs (referenced to an energy optimized WB geometry) are plotted as a function of dG-N1- -dT-N3 inter-atomic distance for a WC-like dG^{enol}•dT tautomeric pair. **g**, Computational studies[31,32,70] predict that the tautomeric pathway proceeds via a planar dG^{+}•dT$^{−}$ ion pair (charge delocalization is implied) that is highlighted by a network of five H-bonds. **h**, Predicted pair geometry of an anionic dG•dT$^{−}$ inverted wobble. Deprotonated dT-N3 is highlighted in red (charge delocalization is implied). **i**, Predicted pair geometry of a dG^{enol}•dT Hoogsteen mispair.



Extended Data Figure 5. Attempts to trap anionic dG

a, 1D ^{13}C spectra (without ^{13}C-^{13}C homonuclear decoupling) of the aromatic carbon region of protonated dGTP (black) and anionic dGTP (red) showing CS perturbations induced upon deprotonation of dGTP-N1. **b**, ^{13}C spectra (without ^{13}C-^{13}C homonuclear decoupling) of the aromatic carbon region of protonated dTTP (black) and anionic dTTP (red) showing CS perturbations induced upon deprotonation of dTTP-N3. **c**, 2D [^{15}N, ^{13}C] HMQC spectra of dGTP showing CS of dGTP-N1 induced upon deprotonation. The spectra is rotated by 90°, to depict ^{15}N CS along x-axis for visualization purposes. Red circles on inset structure highlight measured resonances (C6 and N1). **d**, 2D [^{15}N, ^{13}C] HMQC spectra of dTTP showing CS perturbation of dTTP-N3 induced upon deprotonation. The spectra is rotated by 90°, to depict ^{15}N CS along x-axis for visualization purposes. Red circles on inset structure highlight measured resonances (C4 and N3). **e**, hp-GT DNA spectra of the dG/dT aromatic carbons upon increase in pH from 6.9 (black) to 10.7 (red). Minor upfield CSs are observed for dT5-C6 and dG9-C8, but not dG15-C8, indicating the dT5 in the dG•dT mispair is likely undergoing deprotonation and not the paired dG15. **f**, ^{8Br}G15-hp-GT DNA construct bearing a $^{13}C/^{15}N$ site-labeled dT5 paired with a 8-bromo-2′-deoxyguanosine is shown (left) along with the ^{15}N RD profile for the paired dT5-N3. Error bars represent experimental uncertainty (one s.d., see Methods).



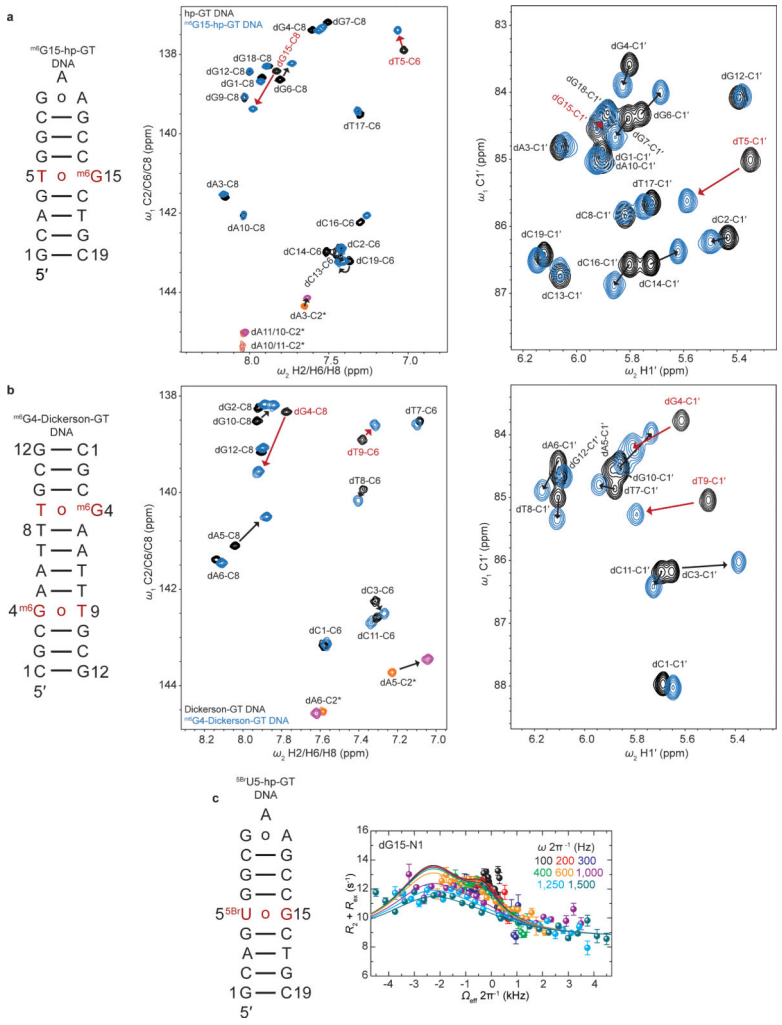
b

Construct	pH	Mispair	ES	ΔG_{AB}	ΔH_{AB}	$T\Delta S_{AB}$	$\Delta G^{\ddagger}_{A}$	$\Delta H^{\ddagger}_{A}$	$T\Delta S^{\ddagger}_{A}$	$\Delta G^{\ddagger}_{B}$	$\Delta H^{\ddagger}_{B}$	$T\Delta S^{\ddagger}_{B}$
hp-GT DNA	6.9	dG15•dT5	ES1	3.80±0.01	4.87±0.53	1.07±0.53	16.4±0.01	12.4±0.35	-4.06±0.35	12.6±0.01	7.51±0.40	-5.13±0.40
hp-GU-20 RNA	6.9	rG16•rU5	ES1	3.86±0.02	5.48±1.03	1.62±1.03	15.7±0.01	13.8±0.46	-1.87±0.46	11.8±0.01	8.36±0.92	-3.48±0.92
* ^{5Br}U-hp-GT DNA	6.9	dG15•^{5Br}dU5	ES1	4.67±0.30	--	--	16.0±0.21	--	--	11.3±0.21	--	--
* hp-GT DNA	8.4	dG15•dT5	ES2	4.59±0.24	--	--	16.4±0.17	--	--	11.8±0.17	--	--
* hp-GU-20 RNA	8.4	rG16•rU5	ES2	3.29±0.11	--	--	15.9±0.08	--	--	12.6±0.08	--	--
* ^{5Br}U-hp-GT DNA	6.9	dG15•^{5Br}dU5	ES2	3.37±0.40	--	--	15.1±0.28	--	--	11.8±0.27	--	--

(kcal mol^{-1})

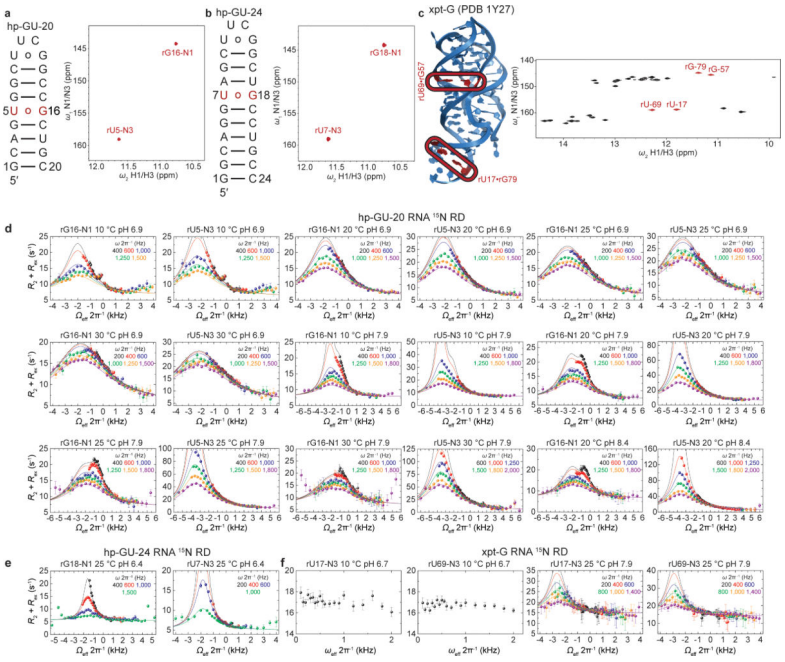
Extended Data Figure 6. Kinetic-thermodynamic plots and parameters

a, Kinetic-thermodynamic diagram for exchange between GS and ES1 via a transition state for hp-GT DNA ES1 (left) and hp-GU-20 RNA ES1 (right), showing activation ($G^{\ddagger}$) and net free energy (G), enthalpy (H), and entropy (TS) changes (referenced to 0). **b**, Kinetic-thermodynamic parameters derived from RD data. Asterisk denotes parameters calculated using only a single temperature (see Methods), wherein enthalpic and entropic parameters cannot be derived. Here, dG15•dT5 ES2 values were calculated at 25.05 °C, rG16•rU5 ES2 values were calculated at 20.05 °C, and dG15•^{5Br}dU5 ES1 and ES2 values were calculated at 10.05 °C. Error is given by the s.e. of the weighted global fits of the corresponding RD profiles. Error is propagated using the respective uncertainties in k_{ex} and p_{ES}.



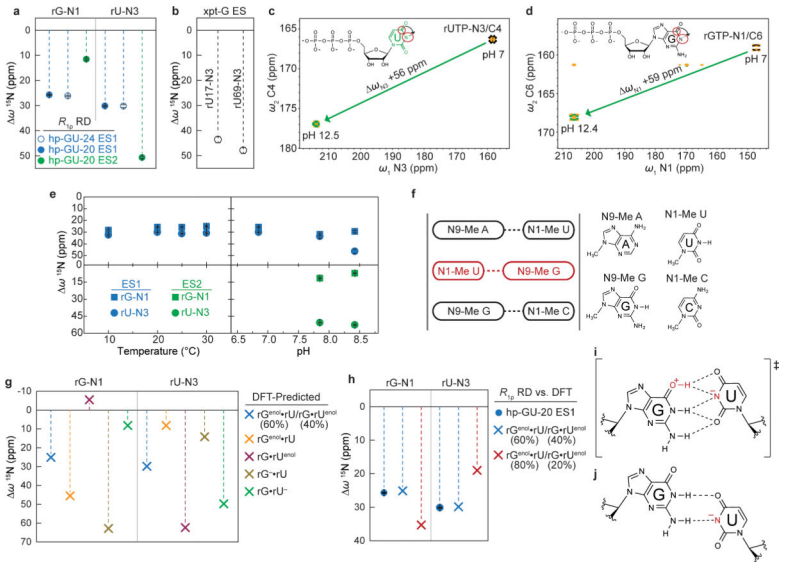
Extended Data Figure 7. Trapping or stabilizing dG•dT ES1 and ES2

a, ^{m6}G15-hp-GT DNA construct is shown (left) where dG15 is methylated at the O^6 position to trap a near-WC "dG^{enol}•dT"-like geometry (Fig. 3c). CS perturbations induced in the aromatic (center) and sugar (right) resonances upon O^6-methylation (blue) with the hp-GT DNA spectra (black) with the resonances for the dG•dT mispair from hp-GT DNA in red. ^{m6}dG•dT mispair and CSs are highlighted in red. **b**, Similarly, ^{m6}G4-Dickerson-GT DNA construct is shown (left) where dG4 is O^6-methylated to trap a WC-like state, with similar color scheme as **a**. **c**, ^{5Br}U5-hp-GT DNA construct bearing a ^{13}C/^{15}N site-labeled dG15 paired with a 5-bromo-2′-deoxythymidine is shown (left) along with the ^{15}N RD profile for the paired dG15-N1. Error bars represent experimental uncertainty (one s.d., see Methods).



Extended Data Figure 8. Rotating frame relaxation dispersion profiles for rG•rU mispairs in hp-GU-20, hp-GU-24 and xpt-G RNA constructs

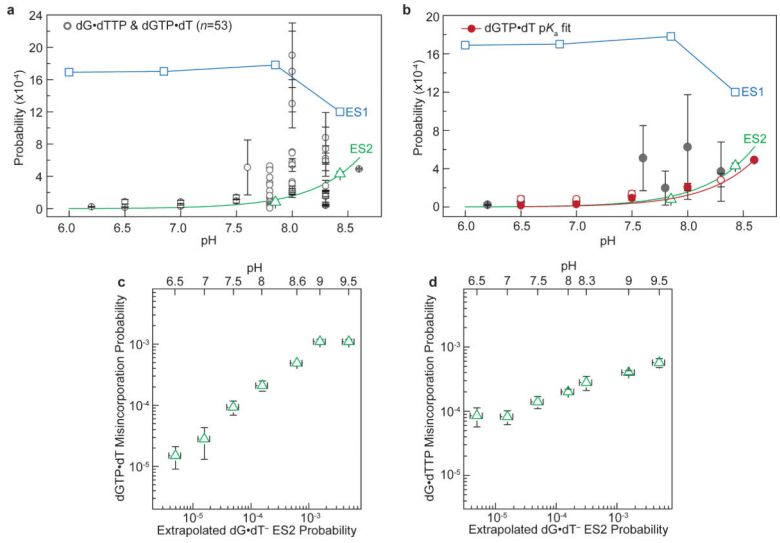
a-b RNA constructs and the imino [^{15}N, ^{1}H] HSQC zoomed into the rG•rU wobble region of the spectra for hp-GU-20 and hp-GU-24. rG•rU mispair resonances are shown in red. **c**, The *Bacillus subtilis* guanine binding riboswitch (xpt-G RNA)[71] construct and full imino [^{15}N, ^{1}H] HSQC of folded and guanine ligand-bound riboswitch. rG•rU mispair resonances are shown in red. **d-f**, ^{15}N RD profiles for **d,** hp-GU-20 **e,** hp-GU-24 and **f,** xpt-G riboswitch RNA. Error bars represent experimental uncertainty (one s.d., see Methods).



Extended Data Figure 9. CS fingerprinting rG•rU excited states

a, RD-derived rG16-N1 and rU5-N3 CSs (referenced to GS WB) are shown for ES1 (20 °C and pH 6.9) and ES2 of hp-GU-20 (20 °C and pH 7.9) and ES1 rG18-N1 and rU7-N3 CSs of

hp-GU-24 (25 °C and pH 6.9). Errors in all RD-derived fitted parameters (eg. $\Delta\omega$) denote s.e. from the weighted global fit (see Methods). **b**, RD-derived CSs (referenced to GS WB) are shown for the ES of xpt-G riboswitch (rU17-N3 and rU69-N3) at 25 °C and pH 7.9. **c**, 2D [^{15}N, ^{13}C] HMQC spectra of rUTP showing CS of rUTP-N3 induced upon deprotonation. The spectra is rotated by 90°, to depict ^{15}N CS along x-axis for visualization purposes. Red circles on inset structure highlight measured resonances (C4 and N3). **d**, 2D [^{15}N, ^{13}C] HMQC spectra of rGTP showing CS of rGTP-N1 induced upon deprotonation. The spectra is rotated by 90°, to depict ^{15}N CS along x-axis for visualization purposes. Red circles on inset structure highlight measured resonances (C6 and N1). **e**, RD-derived hp-GU-20 rG•rU ES1 (blue) and ES2 (green) CSs are shown as a function of temperature and pH for both rG16-N1 (square) and rU5-N3 (circle). **f**, Scheme used to calculate CSs using DFT. Idealized A-form RNA helix is generated to give a central rG•rU mispair (red) that is flanked by canonical rG•rC and rA•rU pairs, analogous to the hp-GU-24 construct. Residues are trimmed to 1-/9-methyl bases and i+1/i−1 pairs are frozen in place for subsequent geometry optimizations and CS calculations (see Methods). **g**, DFT-predicted CSs (referenced to an energy optimized WB geometry) are shown for various tautomeric and anionic configurations, where rG^{enol}•rU/rG•rU^{enol} represents population weighted average CSs of rG^{enol}•rU (60%) and rG•rU^{enol} (40%). **h**, ^{15}N rG-N1 and rU-N3 CS comparison between RD-derived ES1 CSs and population weighted DFT-predicted CSs (60:40 vs. 80:20). **i**, Computational studies[32] predict that the tautomeric pathway for a rG•rU pair can proceed via a planar rG^{+}•rU^{-} ion pair (charge delocalization is implied) that is highlighted by a network of five H-bonds. **j**, Pair geometry of an anionic rG•rU^{-} inverted wobble. Deprotonated rU-N3 is highlighted in red (charge delocalization is implied).



Extended Data Figure 10. dG•dT Misincorporation probabilities and correlation to WC-like excited states

a, Explicit dGTP•dT and dG•dTTP kinetic misincorporation and base substitution probabilities (n=53) and associated errors[46-48,72-77] (see Supplementary Discussion 8) are plotted against hp-GT dG•dT ES1 (blue squares) and ES2 (green triangles). The pK_a fit of ES2 probabilities to the Henderson-Hasselbalch equation (Eqn. 4, see Methods) is shown as

the green trend line. **b**, Red trend line shows the pK_a fit to dGTP•dT misincorporation probabilities[47] from pH 6.5-8.6 to the Henderson-Hasselbalch equation. The fit was weighted using reported experimental errors and gave a reduced χ^2 of 3.56. **c-d**, Extrapolated dG•dT$^-$ ES2 probability (s.e. from the weighted global fit) is plotted against dGTP•dT (left) and dG•dTTP (right) misincorporation probabilities (errors as given)[47] from pH 6.5–9.5.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank S. Horowitz, H. Zhou, J. Lee, A.M. Mustoe, and E.N. Nikolova for assistance and critical comments. We are grateful for technical support and resources from the Duke Magnetic Resonance Spectroscopy Center and University of Michigan Flux HPC Cluster. This work was supported by an NIH grant (R01GM089846) and an Agilent Thought Leader Award given to H.M.A.

REFERENCES

1. Bebenek K, Pedersen LC, Kunkel TA. Replication infidelity via a mismatch with Watson–Crick geometry. Proc. Natl. Acad. Sci. U. S. A. 2011; 108:1862–1867. [PubMed: 21233421]
2. Koag M-C, Nam K, Lee S. The spontaneous replication error and the mismatch discrimination mechanisms of human DNA polymerase β. Nucleic Acids Res. 2014; 42:11233–11245. [PubMed: 25200079]
3. Xia S, Konigsberg WH. Mispairs with Watson-Crick base-pair geometry observed in ternary complexes of an RB69 DNA polymerase variant. Protein Sci. 2014; 23:508–513. [PubMed: 24458997]
4. Wang W, Hellinga HW, Beese LS. Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. Proc. Natl. Acad. Sci. U. S. A. 2011; 108:17644–17648. [PubMed: 22006298]
5. Ogle JM, Murphy Iv FV, Tarry MJ, Ramakrishnan V. Selection of tRNA by the Ribosome Requires a Transition from an Open to a Closed Form. Cell. 2002; 111:721–732. [PubMed: 12464183]
6. Demeshkina N, Jenner L, Westhof E, Yusupov M, Yusupova G. A new understanding of the decoding principle on the ribosome. Nature. 2012; 484:256–259. [PubMed: 22437501]
7. Watson JD, Crick FHC. The Structure of DNA. Cold Spring Harbor Symp. Quant. Biol. 1953; 18:123–131. [PubMed: 13168976]
8. Topal MD, Fresco JR. Complementary base pairing and the origin of substitution mutations. Nature. 1976; 263:285–289. [PubMed: 958482]
9. Harris VH, et al. The effect of tautomeric constant on the specificity of nucleotide incorporation during DNA replication: support for the rare tautomer hypothesis of substitution mutagenesis. J. Mol. Biol. 2003; 326:1389–1401. [PubMed: 12595252]
10. Topal MD, Fresco JR. Base pairing and fidelity in codon-anticodon interaction. Nature. 1976; 263:289–293. [PubMed: 958483]
11. Sowers L, Goodman M, Eritja R, Kaplan B, Fazakerley G. Ionized and wobble base-pairing for bromouracil-guanine in equilibrium under physiological conditions: A nuclear magnetic resonance study on an oligonucleotide containing a bromouracil-guanine base-pair as a function of pH. J. Mol. Biol. 1989; 205:437–447. [PubMed: 2538629]
12. Warren JJ, Forsberg LJ, Beese LS. The structural basis for the mutagenicity of O^6-methyl-guanine lesions. Proc. Natl. Acad. Sci. U. S. A. 2006; 103:19701–19706. [PubMed: 17179038]
13. Weixlbaumer A, et al. Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. Nat. Struct. Mol. Biol. 2007; 14:498–502. [PubMed: 17496902]

14. Cantara WA, Murphy FV, Demirci H, Agris PF. Expanded use of sense codons is regulated by modified cytidines in tRNA. Proc. Natl. Acad. Sci. U. S. A. 2013; 110:10964–10969. [PubMed: 23781103]
15. Bevilacqua PC, Yajima R. Nucleobase catalysis in ribozyme mechanism. Curr. Opin. Chem. Biol. 2006; 10:455–464. [PubMed: 16935552]
16. Cochrane JC, Strobel SA. Catalytic Strategies of Self-Cleaving Ribozymes. Acc. Chem. Res. 2008; 41:1027–1035. [PubMed: 18652494]
17. Gilbert SD, Reyes FE, Edwards AL, Batey RT. Adaptive Ligand Binding by the Purine Riboswitch in the Recognition of Guanine and Adenine Analogs. Structure. 2009; 17:857–868. [PubMed: 19523903]
18. Singh V, et al. Direct Observation of Multiple Tautomers of Oxythiamine and their Recognition by the Thiamine Pyrophosphate Riboswitch. ACS Chem. Biol. 2013; 9:227–236. [PubMed: 24252063]
19. Li D, et al. Tautomerism provides a molecular explanation for the mutagenic properties of the anti-HIV nucleoside 5-aza-5,6-dihydro-2′-deoxycytidine. Proc. Natl. Acad. Sci. U. S. A. 2014; 111:E3252–E3259. [PubMed: 25071207]
20. Korzhnev DM, Orekhov VY, Kay LE. Off-Resonance $R_{1\rho}$ NMR Studies of Exchange Dynamics in Proteins with Low Spin-Lock Fields: An Application to a Fyn SH3 Domain. J. Am. Chem. Soc. 2004; 127:713–721. [PubMed: 15643897]
21. Palmer AG III. Chemical exchange in biomacromolecules: Past, present, and future. J. Magn. Reson. 2014; 241:3–17. [PubMed: 24656076]
22. Hansen AL, Kay LE. Measurement of histidine pKa values and tautomer populations in invisible protein states. Proc. Natl. Acad. Sci. U. S. A. 2014; 111:E1705–E1712. [PubMed: 24733918]
23. Hoogstraten CG, Wank JR, Pardi A. Active Site Dynamics in the Lead-Dependent Ribozyme. Biochemistry. 2000; 39:9951–9958. [PubMed: 10933815]
24. Nikolova EN, et al. Transient Hoogsteen base pairs in canonical duplex DNA. Nature. 2011; 470:498–502. [PubMed: 21270796]
25. Dethoff EA, Petzold K, Chugh J, Casiano-Negroni A, Al-Hashimi HM. Visualizing transient low-populated structures of RNA. Nature. 2012; 491:724–728. [PubMed: 23041928]
26. Massi F, Johnson E, Wang C, Rance M, Palmer AG III. NMR $R_{1\rho}$ Rotating-Frame Relaxation with Weak Radio Frequency Fields. J. Am. Chem. Soc. 2004; 126:2247–2256. [PubMed: 14971961]
27. Hansen AL, Nikolova EN, Casiano-Negroni A, Al-Hashimi HM. Extending the Range of Microsecond-to-Millisecond Chemical Exchange Detected in Labeled and Unlabeled Nucleic Acids by Selective Carbon $R_{1\rho}$ NMR Spectroscopy. J. Am. Chem. Soc. 2009; 131:3818–3819. [PubMed: 19243182]
28. Duthaler RO, Roberts JD. Effects of solvent, protonation, and N-alkylation on the nitrogen-15 chemical shifts of pyridine and related compounds. J. Am. Chem. Soc. 1978; 100:4969–4973.
29. Cho BP, Kadlubar FF, Culp SJ, Evans FE. Nitrogen-15 nuclear magnetic resonance studies on the tautomerism of 8-hydroxy-2′-deoxyguanosine, 8-hydroxyguanosine, and other C8-substituted guanine nucleosides. Chem. Res. Toxicol. 1990; 3:445–452. [PubMed: 2133096]
30. Goswami B, Gaffney BL, Jones RA. Nitrogen-15-labeled oligodeoxynucleotides. 5. Use of ^{15}N NMR to probe H-bonding in an O^6MeG•T base pair. J. Am. Chem. Soc. 1993; 115:3832–3833.
31. Nomura K, et al. DFT Calculations on the Effect of Solvation on the Tautomeric Reactions for Wobble Gua-Thy and Canonical Gua-Cyt Base-Pairs. J. Mod. Phys. 2013; 4:422–431.
32. Brovarets OO. Effect of a modification of uracil on the tautomerization barrier of the wobble Gua•^{5X}Ura base pair into the Gua*•^{5X}Ura base pair with the Watson-Crick geometry: quantum-chemical study. Rep. Natl. Acad. Sci. Ukraine. 2013; 4:154–158.
33. Xu X-P, Au-Yeung SCF. Investigation of Chemical Shift and Structure Relationships in Nucleic Acids Using NMR and Density Functional Theory Methods. J. Phys. Chem. B. 2000; 104:5641–5650.
34. Johnson SJ, Beese LS. Structures of mismatch replication errors observed in a DNA polymerase. Cell. 2004; 116:803–816. [PubMed: 15035983]
35. Brovarets OO, Zhurakivsky RO, Hovorun DM. Is there adequate ionization mechanism of the spontaneous transitions? Quantum-chemical investigation. Biopolym. Cell. 2010; 26:398–405.

36. Orozco M, Hernández B, Luque FJ. Tautomerism of 1-Methyl Derivatives of Uracil, Thymine, and 5-Bromouracil. Is Tautomerism the Basis for the Mutagenicity of 5-Bromouridine? J. Phys. Chem. B. 1998; 102:5228–5233.
37. Varani G, McClain WH. The G•U wobble base pair. EMBO Rep. 2000; 1:18–23. [PubMed: 11256617]
38. Kunkel TA, Bebenek K. DNA Replication Fidelity. Annu. Rev. Biochem. 2000; 69:497–529. [PubMed: 10966467]
39. Voorhees RM, Ramakrishnan V. Structural Basis of the Translational Elongation Cycle*. Annu. Rev. Biochem. 2013; 82:203–236. [PubMed: 23746255]
40. Kunkel TA. DNA Replication Fidelity. J. Biol. Chem. 2004; 279:16895–16898. [PubMed: 14988392]
41. Schaaper RM. Base selection, proofreading, and mismatch repair during DNA replication in Escherichia coli. J. Biol. Chem. 1993; 268:23762–23765. [PubMed: 8226906]
42. Xia S, Konigsberg WH. RB69 DNA Polymerase Structure, Kinetics, and Fidelity. Biochemistry. 2014; 53:2752–2767. [PubMed: 24720884]
43. Ogle JM, Ramakrishnan V. Structural Insights Into Translational Fidelity. Annu. Rev. Biochem. 2005; 74:129–177. [PubMed: 15952884]
44. Zhang Z, Shah B, Bondarenko PV. G/U and Certain Wobble Position Mismatches as Possible Main Causes of Amino Acid Misincorporations. Biochemistry. 2013; 52:8165–8176. [PubMed: 24128183]
45. Gromadski KB, Rodnina MV. Kinetic Determinants of High-Fidelity tRNA Discrimination on the Ribosome. Mol. Cell. 2004; 13:191–200. [PubMed: 14759365]
46. Kunkel TA, Alexander PS. The base substitution fidelity of eucaryotic DNA polymerases. Mispairing frequencies, site preferences, insertion preferences, and base substitution by dislocation. J. Biol. Chem. 1986; 261:160–166. [PubMed: 3941068]
47. Yu H, Eritja R, Bloom LB, Goodman MF. Ionization of bromouracil and fluorouracil stimulates base mispairing frequencies with guanine. J. Biol. Chem. 1993; 268:15935–15943. [PubMed: 7688001]
48. Eckert KA, Kunkel TA. Effect of reaction pH on the fidelity and processivity of exonuclease-deficient Klenow polymerase. J. Biol. Chem. 1993; 268:13462–13471. [PubMed: 8390464]
49. Satpati P, Åqvist J. Why base tautomerization does not cause errors in mRNA decoding on the ribosome. Nucleic Acids Res. 2014; 42:12876–12884. [PubMed: 25352546]
50. Driggers PH, Beattie KL. Effect of pH on the base-mispairing properties of 5-bromouracil during DNA synthesis. Biochemistry. 1988; 27:1729–1735. [PubMed: 3284589]
51. Zimmer DP, Crothers DM. NMR of Enzymatically Synthesized Uniformly $^{13}C^{15}N$-Labeled DNA Oligonucleotides. Proc. Natl. Acad. Sci. U. S. A. 1995; 92:3091–3095. [PubMed: 7724521]
52. Delaglio F, et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR. 1995; 6:277–293. [PubMed: 8520220]
53. Cavanagh, J.; Fairbrother, WJ.; Palmer, AG., III; Skelton, NJ. Protein NMR spectroscopy: principles and practice. Academic Press; 1995.
54. Wijmenga SS, van Buuren BNM. The use of NMR methods for conformational studies of nucleic acids. Prog. Nucl. Magn. Reson. Spectrosc. 1998; 32:287–387.
55. Farjon J, et al. Longitudinal-Relaxation-Enhanced NMR Experiments for the Study of Nucleic Acids in Solution. J. Am. Chem. Soc. 2009; 131:8571–8577. [PubMed: 19485365]
56. Sathyamoorthy B, Lee J, Kimsey I, Ganser L, Al-Hashimi H. Development and application of aromatic [^{13}C, ^{1}H] SOFAST-HMQC NMR experiment for nucleic acids. J. Biomol. NMR. 2014:1–7.
57. Nikolova EN, Gottardo FL, Al-Hashimi HM. Probing Transient Hoogsteen Hydrogen Bonds in Canonical Duplex DNA Using NMR Relaxation Dispersion and Single-Atom Substitution. J. Am. Chem. Soc. 2012; 134:3667–3670. [PubMed: 22309937]
58. Spyracopoulos L. A suite of Mathematica notebooks for the analysis of protein main chain ^{15}N NMR relaxation data. J. Biomol. NMR. 2006; 36:215–224. [PubMed: 17061025]

59. Palmer AG III, Massi F. Characterization of the Dynamics of Biomacromolecules Using Rotating-Frame Spin Relaxation NMR Spectroscopy. Chem. Rev. 2006; 106:1700–1719. [PubMed: 16683750]
60. McConnell HM. Reaction Rates by Nuclear Magnetic Resonance. The Journal of Chemical Physics. 1958; 28:430–431.
61. Miloushev VZ, Palmer AG III. $R_{1\rho}$ relaxation for two-site chemical exchange: General approximations and some exact solutions. J. Magn. Reson. 2005; 177:221–227. [PubMed: 16143548]
62. Trott O, Palmer AG III. Theoretical study of $R_{1\rho}$ rotating-frame and R_2 free-precession relaxation in the presence of n-site chemical exchange. J. Magn. Reson. 2004; 170:104–112. [PubMed: 15324763]
63. Bothe JR, Stein ZW, Al-Hashimi HM. Evaluating the uncertainty in exchange parameters determined from off-resonance $R_{1\rho}$ relaxation dispersion for systems in fast exchange. J. Magn. Reson. 2014; 244:18–29. [PubMed: 24819426]
64. Findeisen M, Brand T, Berger S. A ^{1}H-NMR thermometer suitable for cryoprobes. Magn. Reson. Chem. 2007; 45:175–178. [PubMed: 17154329]
65. Gaussian 09. Gaussian, Inc.; Wallingford, CT, USA: 2009.
66. Zhao Y, Truhlar D. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. Theor. Chem. Acc. 2008; 120:215–241.
67. Kochina OS, Zhurakivsky RO, Hovorun DM. Effect of the nucleotide bases tautomerization on the conformational properties of the nucleosides: quantum-mechanical investigation by the DFT method. Rep. Natl. Acad. Sci. Ukraine. 2008; 1:181–186.
68. Czernek J, Fiala R, Sklenář V. r. Hydrogen Bonding Effects on the ^{15}N and ^{1}H Shielding Tensors in Nucleic Acid Base Pairs. J. Magn. Reson. 2000; 145:142–146. [PubMed: 10873505]
69. Brovarets OO, Hovorun DM. The nature of the transition mismatches with Watson–Crick architecture: the G*•T or G•T* DNA base mispair or both? A QM/QTAIM perspective for the biological problem. J. Biomol. Struct. Dyn. 2014:1–21. [PubMed: 23252879]
70. Brovarets OO, Hovorun DM. Physicochemical mechanism of the wobble DNA base pairs Gua•Thy and Ade•Cyt transition into the mismatched base pairs Gua*•Thy and Ade•Cyt* formed by the mutagenic tautomers. Ukr. Bioorg. Acta. 2009; 8:12–18.
71. Serganov A, et al. Structural Basis for Discriminative Regulation of Gene Expression by Adenine- and Guanine-Sensing mRNAs. Chem. Biol. 2004; 11:1729–1741. [PubMed: 15610857]
72. Boosalis MS, Petruska J, Goodman MF. DNA polymerase insertion fidelity. Gel assay for site-specific kinetics. J. Biol. Chem. 1987; 262:14689–14696. [PubMed: 3667598]
73. Mendelman LV, Boosalis MS, Petruska J, Goodman MF. Nearest neighbor influences on DNA polymerase insertion fidelity. J. Biol. Chem. 1989; 264:14415–14423. [PubMed: 2474545]
74. Singer B, Chavez F, Goodman MF, Essigmann JM, Dosanjh MK. Effect of 3′ flanking neighbors on kinetics of pairing of dCTP or dTTP opposite O^6-methylguanine in a defined primed oligonucleotide when Escherichia coli DNA polymerase I is used. Proc. Natl. Acad. Sci. U. S. A. 1989; 86:8271–8274. [PubMed: 2682644]
75. Dosanjh MK, Essigmann JM, Goodman MF, Singer B. Comparative efficiency of forming ^{m4}T•G versus ^{m4}T•A base pairs at a unique site by use of Escherichia coli DNA polymerase I (Klenow fragment) and Drosophila melanogaster polymerase α-primase complex. Biochemistry. 1990; 29:4698–4703. [PubMed: 2115381]
76. Singer B, Dosanjh MK. Site-directed mutagenesis for quantitation of base-base interactions at defined sites. Mutat. Res. Fund. Mol. Mech. Mut. 1990; 233:45–51.
77. Dosanjh MK, Galeros G, Goodman MF, Singer B. Kinetics of extension of O^6-methylguanine paired with cytosine or thymine in defined oligonucleotide sequences. Biochemistry. 1991; 30:11595–11599. [PubMed: 1747377]
78. Nikolova EN, Goh GB, Brooks CL III, Al-Hashimi HM. Characterizing the Protonation State of Cytosine in Transient G•C Hoogsteen Base Pairs in Duplex DNA. J. Am. Chem. Soc. 2013; 135:6766–6769. [PubMed: 23506098]

79. Várnai P, Canalia M, Leroy J-L. Opening Mechanism of G•T/U Pairs in DNA and RNA Duplexes: A Combined Study of Imino Proton Exchange and Molecular Dynamics Simulation. J. Am. Chem. Soc. 2004; 126:14659–14667. [PubMed: 15521786]
80. Yin Y, et al. Dynamics of spontaneous flipping of a mismatched base in DNA duplex. Proc. Natl. Acad. Sci. U. S. A. 2014; 111:8043–8048. [PubMed: 24843124]
81. Giudice E, Várnai P, Lavery R. Base pair opening within B - DNA: free energy pathways for GC and AT pairs from umbrella sampling simulations. Nucleic Acids Res. 2003; 31:1434–1443. [PubMed: 12595551]
82. Padermshoke A, Katsumoto Y, Masaki R, Aida M. Thermally induced double proton transfer in GG and wobble GT base pairs: A possible origin of the mutagenic guanine. Chem. Phys. Lett. 2008; 457:232–236.
83. Koag M-C, Kou Y, Ouzon-Shubeita H, Lee S. Transition-state destabilization reveals how human DNA polymerase β proceeds across the chemically unstable lesion N7-methylguanine. Nucleic Acids Res. 2014; 42:8755–8766. [PubMed: 24966350]
84. Patel DJ, Shapiro L, Kozlowski SA, Gaffney BL, Jones RA. Structural studies of the O^6meG•T interaction in the d(C-G-T-G-A-A-T-T-C-O^6meG-C-G) duplex. Biochemistry. 1986; 25:1036–1042. [PubMed: 3964659]
85. Zaher HS, Green R. Quality control by the ribosome following peptide bond formation. Nature. 2009; 457:161–166. [PubMed: 19092806]
86. Parker J. Errors and alternatives in reading the universal genetic code. Microbiological Reviews. 1989; 53:273–298. [PubMed: 2677635]
87. Jang YH, et al. pK_a Values of Guanine in Water: Density Functional Theory Calculations Combined with Poisson-Boltzmann Continuum-Solvation Model. J. Phys. Chem. B. 2002; 107:344–357.
88. Nair DT, Johnson RE, Prakash L, Prakash S, Aggarwal AK. Human DNA polymerase ι incorporates dCTP opposite template G via a G•C^+ Hoogsteen base pair. Structure. 2005; 13:1569–1577. [PubMed: 16216587]

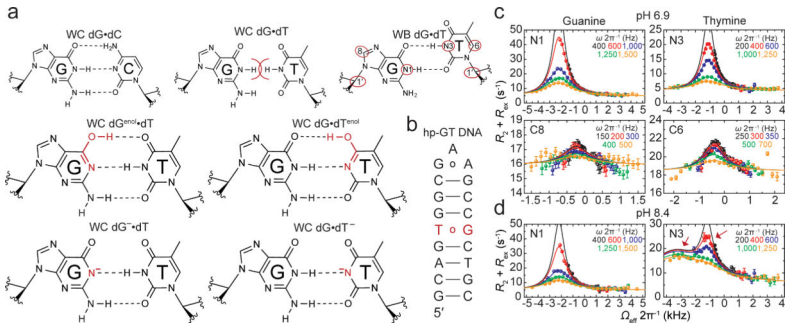


Figure 1. Chemical exchange in dG•dT mispairs

a, WC dG•dC, sterically prohibited WC dG•dT, and WB dG•dT ($R_{1\rho}$ measured nuclei highlighted in red ovals). Below are four WC-like tautomeric and anionic (implied charge delocalization) bps. **b**, DNA duplex with a $^{13}C/^{15}N$ site-labeled dG•dT mispair. **c**, RD profiles for dG•dT (25 °C and pH 6.9) showing R_2+R_{ex} as a function of the spin lock offset ($\Omega_{eff}\ 2\pi^{-1}$) and power ($\omega_{SL}\ 2\pi^{-1}$, in insets) with global fits to dG-N1, dG-C8, dT-N3, and dT-C6. Error bars represent experimental uncertainty (one s.d., see Methods). **d**, RD profiles showing 3-state exchange (25 °C and pH 8.4) and global 3-state fit to dG-N1 and dT-N3.

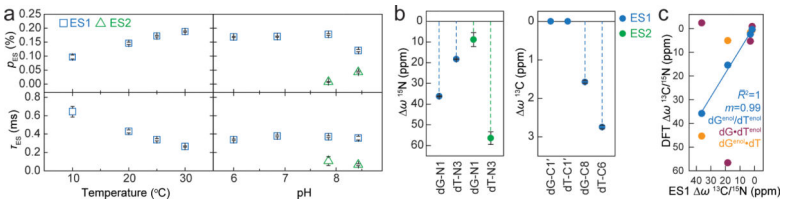


Figure 2. Characterizing WC-like transient states

a, Population and lifetime of dG•dT ES1/ES2 measured in hp-GT DNA as a function of temperature (at pH 6.9) and pH (at 25 °C). Errors in fitted parameters denote s.e. from the weighted global fit. **b**, Differences between the GS (referenced to 0 ppm) and ES CSs ($\Delta\omega=\omega_{ES}-\omega_{GS}$) for hp-GT DNA. **c**, Measured CSs for ES1 are plotted against DFT-predicted values.

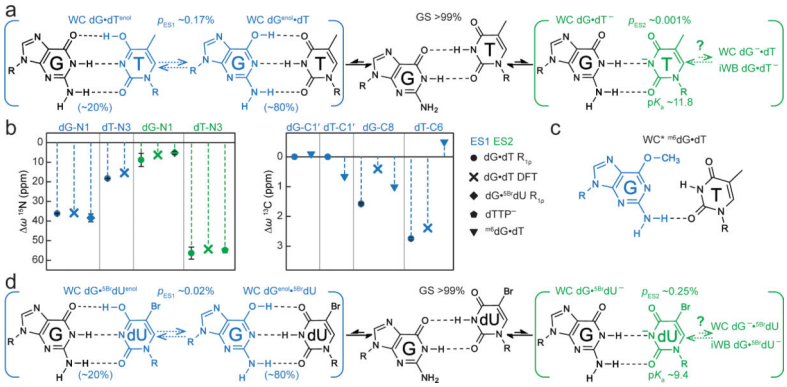


Figure 3. Mutate-and-CS fingerprinting ES1 and ES2

a, Multi-state equilibrium between WB and WC-like dG•dT mispairs. ES1/ES2 populations and weights are shown (25 °C and pH 6.9). p_{ES2} estimated based on the observed apparent pK_a. **b**, CS fingerprinting dG•dT ES1/ES2 using chemical modifications and structure-based DFT predictions of CSs. ES1 DFT CSs are given for 80:20 dG^{enol}•dT:dG•dT^{enol} weighting. **c**, ^{m6}dG•dT structure[12,30]. **d**, dG•^{5Br}dU^{-} ES2 stabilized relative to dG•dT^{-} ES2. Populations and weights (10 °C and pH 6.9) are shown.

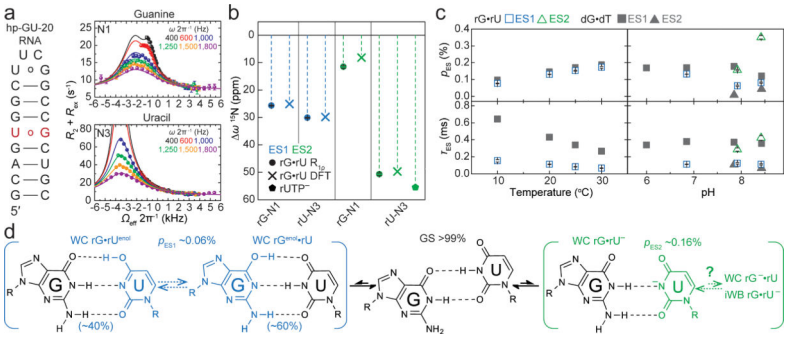


Figure 4. Transient tautomeric and anionic WC-like mispairs in A-form RNA

a, RNA duplex and $R_{1\rho}$ RD profiles with 3-state global fits to rG-N1 and rU-N3 (20 °C and pH 7.9). **b**, CSs for rG•rU ES1/ES2 compared to structure-based DFT predictions and rUTP ionization. **c**, Population and lifetime of rG•rU ES1/ES2 measured as a function of temperature (at pH 6.9) and pH (at 20 °C). dG•dT ES1 and ES2 shown (in grey) for comparison. **d**, Multi-state equilibrium between WB and WC-like rG•rU mispairs (20 °C and pH 7.9).

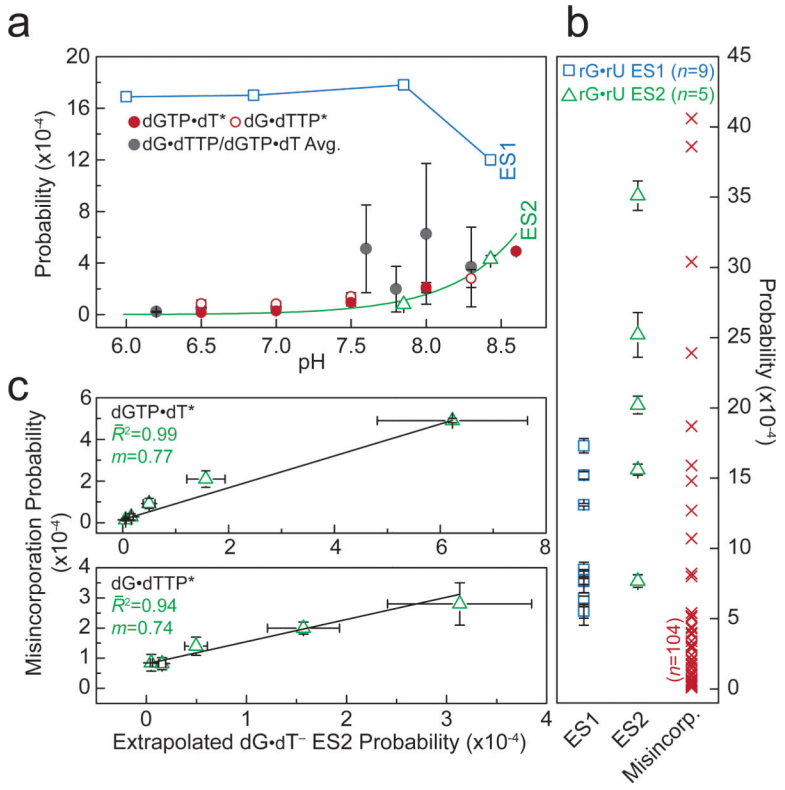


Figure 5. Correlation between WC-like bps and misincorporation probabilities

a, dG•dT ES1 (blue square) and ES2 (green triangle) probabilities. pH-dependent dGTP•dT*/dG•dTTP* misincorporation probabilities (errors as published)[47]. Averaged dGTP•dT/dG•dTTP misincorporation and base substitution probabilities; error bars reflect the s.d. (Supplementary Discussion 8). ES2 fit to Henderson-Hasselbalch equation shown as green line (see Methods). A blue line connects ES1 points for visualization purposes. **b**, rG•rU ES1 (blue square) and ES2 (green triangle) probabilities measured at varied conditions (Supplementary Table 1), and amino acid misincorporation probabilities (red 'X') due to rG•rU mispairs (Zhang *et al.*[44]). **c**, The probabilities of dGTP•dT/dG•dTTP misincorporation (error as given) by AMV RT (Yu *et al.*[47]) versus the pK_a-predicted probability (s.e.) of forming a WC-like dG•dT$^-$ mispair as a function of pH.