HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2018 July 31.

Published in final edited form as:

Nature. 2018 February 08; 554(7691): 195–201. doi:10.1038/nature25487.

Dynamic Basis for dG•dT misincorporation via tautomerization and ionization

Isaac J. Kimsey[1,*,†], **Eric S. Szymanski**[1,*], **Walter J. Zahurancik**[2,3], **Anisha Shakya**[4,@], **Yi Xue**[1,+], **Chia-Chieh Chu**[1], **Bharathwaj Sathyamoorthy**[1,#], **Zucai Suo**[2,3], and **Hashim M. Al-Hashimi**[1,5]

[1]Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710, USA

[2]Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH 43210, USA

[3]The Ohio State Biochemistry Program, The Ohio State University, Columbus, OH 43210, USA

[4]Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA

[5]Department of Chemistry, Duke University, Durham, North Carolina 27710, USA

Abstract

Tautomeric and anionic Watson-Crick-like mismatches play important roles in replication and translation errors through mechanisms that are not fully understood. Using NMR relaxation dispersion, we resolved a sequence-dependent kinetic network connecting G•T/U wobbles with three distinct Watson-Crick mismatches consisting of two rapidly exchanging tautomeric species (G^{enol}•T/U⇌G•T^{enol}/U^{enol}; population <0.4%) and one anionic species (G•T^{-}/U^{-}; population ≈0.001% at neutral pH). Inserting the sequence-dependent tautomerization/ionization step into a minimal kinetic mechanism for correct incorporation during replication following initial nucleotide binding leads to accurate predictions of dG•dT misincorporation probability across different polymerases, pH conditions, and for a chemically modified nucleotide, and provides mechanisms for sequence-dependent misincorporation. Our results indicate that the energetic

Correspondence and requests for materials should be addressed to Z.S. (suo.3@osu.edu) or H.M.A. (hashim.al.hashimi@duke.edu).
*These authors contributed equally to this work.
†Present Address: Nymirum, Durham, North Carolina 27713, USA
@Present Address: Institute of Basic Science, Center for Soft and Living Matter, Ulsan, South Korea
+Present Address: School of Life Sciences, Tsinghua University, Beijing, China
#Present Address: Department of Chemistry, Indian Institute of Science Education and Research Bhopal, Bhopal 462 066, India

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions I.J.K and E.S.S. contributed equally to this work. I.J.K., E.S.S., and H.M.A. conceived the research and experimental design and performed data interpretation. I.J.K, E.S.S., W.J.Z., Z.S., and H.M.A. wrote the manuscript. I.J.K. synthesized all DNA constructs as well as the RNA hpUG-CGC, hpUG-CGU, and xptG riboswitch constructs and performed all NMR RD data collection and analyses. E.S.S. constructed and tested the models of misincorporation. W.J.Z. performed all kinetic experiments. A.S. synthesized A20G and A22G HIV-I TAR constructs Y.X. synthesized and assigned p5abc and glnA riboswitch constructs. C-C.C. synthesized and assigned the HIV-I RRE construct. B.S. synthesized and assigned the HIV-I SL1 dimer complex.

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

penalty for tautomerization/ionization accounts for ≈10^{-2}–10^{-3}–fold discrimination against misincorporation, which proceeds primarily via tautomeric dG^{enol}•dT and dG•dT^{enol} with contributions from anionic dG•dT^{-} dominating at pH ≥8.4 or for some mutagenic nucleotides.

In their paper describing the structure of the DNA double helix[1], Watson and Crick proposed that if nucleotide bases adopted their energetically unfavorable tautomeric forms, mismatches (Fig. 1a) could pair up in a Watson-Crick-like (WC-like) geometry (Fig. 1b) and potentially give rise to spontaneous mutations. Decades later, it is well established that the replicative and translational machineries form a tight grip around the WC geometry to discriminate against mismatches[2–5]. There is also evidence that both tautomeric[6–13] (Fig. 1b) and anionic[7–9,14,15] (Fig. 1c) WC-like mismatches can evade such fidelity checkpoints and give rise to replication[6,7] and translation errors[16]. Despite their centrality to the fidelity of information transfer in the central dogma of molecular biology, and growing evidence showing the involvement of spontaneous mutations in generating cancer causing mutations[17], the very existence of these species and their contributions to replication and translation errors remain to be definitively established.

Tautomeric and anionic mismatches come in a variety of chemical forms (Extended Data Fig. 1). For example, WC-like G•T/U mismatches can form when either the guanine (G^{enol}•T/U and G^{-}•T/U) or thymidine/uridine (G•T^{enol}/U^{enol} and G•T^{-}/U^{-}) base assumes a rare enolic (Fig. 1b) or anionic (Fig. 1c) form. While it remains unclear which WC-like mismatch contributes to replication and translation errors, factors (e.g. changes in pH[7,8,14,18] and chemical modifications[19]) that stabilize different forms have been shown to increase misincorporation probabilities[14,20]. Misincorporation probabilities can also vary significantly with sequence context through mechanisms that remain poorly understood[21,22]. Resolving these different WC-like mismatches and their chemical dynamics is key for elucidating their potential roles in replication, transcription, and translation errors. However, this presents a formidable challenge to current biophysical methods because these mismatches differ by the placement of a single proton and π-bond (Fig. 1b,c and Extended Data Fig. 1). Protons are generally invisible to X-ray crystallography and cryo-EM[12], and consequently it has not been possible to unambiguously resolve the identity of WC-like mismatches captured within active sites of polymerases[6,7,15,23] and the ribosome decoding site[9,24]. Moreover, WC-like mismatches are predicted to exist in rapid tautomeric (G^{enol}•T/U⇌G•T^{enol}/U^{enol})[25,26] equilibria (Fig. 1b,c and Extended Data Fig. 1) making them exceptionally difficult to capture experimentally.

Techniques based on NMR relaxation dispersion (RD)[27–29] make it possible to characterize low-abundance short-lived conformational states, or ‘excited states’ (ESs), in biomolecules[30]. Using NMR RD, we recently provided evidence that wobble G•T/U mismatches exist in dynamic equilibrium with tautomeric (ES1) and anionic (ES2) WC-like mismatches within DNA and RNA duplexes[8,31]. The guanine N1 (G-N1) and thymidine/uridine N3 (T/U-N3) chemical shifts measured for tautomeric ES1 were consistent with G^{enol}•T/U, but were partially skewed toward G•T^{enol}/U^{enol}. This was interpreted as evidence for a rapid (on the chemical shift timescale) equilibrium between a major G^{enol}•T/U and minor G•T^{enol}/U^{enol} species[8]. The anionic ES2 was only detectable at high pH (≥7.8) and

was heavily skewed in favor of G•T^{-}/U^{-} with no evidence for G^{-}•T/U. The roles of these various WC-like mismatches in replication and translation errors remains unknown. Here, by combining NMR RD and measurements of misincorporation rates, we resolved a kinetic network connecting two distinct tautomeric and one anionic WC-like mismatches, and established their relative contributions to dG•dTTP misincorporation.

Tilting the tautomeric equilibrium

If ES1 does indeed represent two tautomeric species in rapid equilibrium (Fig. 1b), it should be feasible to tilt the equilibrium ($K_t = p_{Genol}/p_{Tenol/Uenol}$) by changing the local sequence or structural context around the mismatch, or by using base modifications (Fig. 1d). This in turn should lead to very specific changes in the ES1 G-N1 and T/U-N3 chemical shifts, which are population weighted averages over the two species (Fig. 1e; left). Tilting the equilibrium in favor of G^{enol}•T/U should induce a downfield shift in the ES1 G-N1, because it increases the population of deprotonated G^{enol}, and an upfield shift in ES1 T/U-N3, because it decreases the population of deprotonated T^{enol}/U^{enol}, and *vice versa* (Fig. 1e; left). $\Delta\omega_{T/U\text{-}N3}$ versus $\Delta\omega_{G\text{-}N1}$ is predicted to be linear (Fig. 1e; right) with negative slope and intercept determined by the fundamental chemical shifts of the tautomeric species (Equation 1).

We measured ^{15}N RD for five dG•dT mismatches within distinct sequence contexts and for thirteen rG•rU mismatches in nine structurally unique non-coding RNAs (Fig. 2a and Extended Data Fig. 2a). Experiments were carried out at near-neutral pH (6.4–6.9) so as to lower the anionic ES2 below detection[8] (Extended Data Fig. 2b). The RD experiments measure spin-relaxation rates in the rotating frame ($R_{1\rho}$) during a relaxation period in which a radiofrequency field is applied with variable offset ($\Omega\ 2\pi^{-1}$, in Hz) and power ($\omega\ 2\pi^{-1}$, in Hz) to suppress the chemical exchange contribution (R_{ex}) to the transverse spin relaxation rate (R_2) arising due to chemical exchange between the energetically more stable ground state (GS) and ES[27,28].

We observed G-N1 and T/U-N3 RD consistent with WC-like ES1 exchange for all five dG•dT and eight rG•rU mismatches located within helical environments (Fig. 2b and Extended Data Fig. 3a), thus establishing their robust occurrence in DNA and RNA. No RD was observed (Extended Data Fig. 3b) for rG•rU mismatches adjacent to apical loops, three-way junctions, or bulges (Extended Data Fig. 2a). This could be due to the lower abundance of WC-like mismatches when outside the grip of the helical environment, though we cannot rule out that the exchange is orders of magnitude faster and beyond detection.

As predicted based on variably tilting the G^{enol}•T/U⇌G•T^{enol}/U^{enol} equilibrium (Fig. 1e; right), the fitted $\langle\Delta\omega_{G\text{-}N1}\rangle$ and $\langle\Delta\omega_{T/U\text{-}N3}\rangle$ values obtained from two-state analysis (GS⇌ES1) of the RD profiles (Fig. 2b; Extended Data Figure 3a; and Supplementary Table 1) fell along a line with negative slope (Fig. 2c). As a negative control, no correlation was observed between the corresponding GS G-N1 and T/U-N3 chemical shifts (Extended Data Fig. 3c). We confirmed these linear trends using chemical modifications that tilt the tautomeric equilibrium toward enolic dT (dG•dU and dG•^{5Br}dU) or enolic dG (^{8Br}dG•dT) (Fig. 2c; Extended Data Fig. 3a; and Supplementary Discussion 1).

Sequence-dependent G^{enol}•T/U⇌G•T^{enol}/U^{enol}

A linear fit to the $<\Delta\omega_{T/U\text{-}N3}>$ versus $<\Delta\omega_{G\text{-}N1}>$ values, assuming physically reasonable ranges, yielded fundamental chemical shifts for the tautomeric species that are in excellent agreement with values predicted by DFT calculations (Fig. 2c and Supplementary Tables 2,3)[8]. The tautomeric equilibria (Supplementary Table 2) obtained from this analysis and from re-fitting the RD data using a 3-state model with linear topology[25] (wobble⇌G•T^{enol}/U^{enol}⇌G^{enol}•T/U) are slightly tilted in favor of dG^{enol}•dT in DNA (K_t = 2.1-4.6) whereas the populations of rG^{enol}•rU and rG•rU^{enol} are more comparable in RNA (K_t = 0.5-1.1). These differences may be attributed to the electron-donating methyl group in dT which destabilizes dT^{enol} relative to rU^{enol}[32]. The RD data also allowed us to estimate a lower bound for the fast tautomeric exchange rate $k_t = k_{G^{enol} \rightarrow T^{enol}} + k_{T^{enol} \rightarrow G^{enol}}$ >≈500,000-1,000,000 s^{-1} (Fig. 2d and Extended Data Fig. 4)[33] and a G•T^{enol}/U^{enol}⟶G^{enol}•T/U transition state barrier <9-10 kcal mol^{-1} (pre-exponential factor = $k_B T h^{-1}$[34] and $\kappa = 1$) that is in good agreement with values (≈11.5 kcal mol^{-1}) reported using computational methods[25]. These results establish the existence of G•T^{enol}/U^{enol} and G^{enol}•T/U in an ultra-fast equilibrium, each of which can potentially contribute to replication and translation errors.

Interestingly, the exchange parameters vary significantly with sequence context (Supplementary Table 1). The ES1 population ($p_{ES1} = p_{Genol} + p_{Tenol/Uenol}$) varies 3-fold in DNA and 8-fold in RNA while the forward ($k_{GS \rightarrow ES1}$) and reverse ($k_{ES1 \rightarrow GS}$) rate constants vary by 4- and 5-fold, respectively, for DNA, and by 38- and 6-fold, respectively, for RNA (Supplementary Table 1). A linear correlation is observed between p_{ES1} and K_t (Fig. 2e and Supplementary Table 2), indicating that the G^{enol}•T/U population dominates these variations with sequence and structural context. In DNA, these variations can potentially be explained by sequence-specific changes in stacking with the immediate neighbors that accompany the transition from the wobble to Watson-Crick geometry (Fig. 2f). For example, GGG, has the highest p_{ES1} and is predicted to gain stacking overlap whereas CGG has the lowest p_{ES1} and is predicted to lose stacking overlap. Similar sequence-dependent effects have been reported in lesion repair by methyltransferases[35]. Interestingly, dG dominates the changes in stacking, potentially explaining the stronger sequence dependence of the G^{enol}•T/U population compared to G•T^{enol}/U^{enol}.

Sequence-dependent anion equilibria

Next, we examined whether anionic WC-like G•T^-/U^- (Fig. 1c) also form robustly in DNA and RNA and whether anionic G^-•T/U remains undetectable under these different environments. We measured RD at high pH (≥7.8) for G•T/U mismatches in a subset of our RNA (Fig. 3a and Extended Data Fig. 5a) and DNA constructs (Fig. 3b and Extended Data Fig. 5b). In all cases, we observed RD consistent with wobble⇌anion exchange, thus establishing the robustness of this process across different sequence contexts (Fig. 3a,b and Extended Data Fig. 5). 3-state fitting of the RD data assuming a starlike topology (Extended Data Fig. 6a) yielded large $\Delta\omega_{T/U\text{-}N3(ES2)}$ ≈55 p.p.m. and much smaller $\Delta\omega_{G\text{-}N1(ES2)}$ ≈5 p.p.m., consistent with a dominant G•T^-/U^- species and with no evidence for G^-•T/U. Again, we observe strong sequence-specific variations in ES2 population (p_{ES2}) and the

$k_{GS\rightarrow ES2}$ and $k_{ES2\rightarrow GS}$ rates robustly across different temperatures and pH conditions (Supplementary Tables 4,5).

In a previous study[8], the emergence of anionic ES2 at high pH was accompanied by unexpected changes in the ES1 tautomeric chemical shifts. Similar deviations are observed here for both RNA and DNA (Supplementary Table 4). We postulated that 'minor' exchange[36] between ES1 and ES2 could 'mix' their chemical shifts and give rise to such deviations (Extended Data Fig. 6a,b and Supplementary Table 4). Indeed, all five RD profiles with peculiar ES1 chemical shifts showed a statistically significant improvement when fitting data to a 3-state model with minor exchange in a triangular rather than starlike topology (Fig. 3a,b; Extended Data Fig. 6c,d; and Supplementary Tables 4–6). The resulting ES1 ^{15}N rG-N1 and rU-N3 chemical shifts vary less significantly with pH (Extended Data Fig. 6e and Supplementary Table 4) and the rate constants ($k_{ES1\rightarrow ES2}$ and $k_{ES2\rightarrow ES1}$) exhibit the expected temperature dependence (Extended Data Fig. 6f), neither of which is expected if the data were being spuriously over fitted. Additionally, limited or poor quality RD data can make it difficult to resolve different topologies[37] (Supplementary Table 6).

Tautomerization/ionization during misincorporation

dG•dT misincorporation is the most frequent base substitution error committed by high fidelity DNA polymerases with misincorporation frequency $F_{pol}= (k_{pol}/K_d)_{incorrect}/(k_{pol}/K_d)_{correct} \approx 10^{-4}–10^{-5}$ for most studied polymerases[38,39], i.e. an error is committed with frequency of 1 in $10^4–10^5$ nucleotide incorporations, in which k_{pol} is the maximum rate of nucleotide incorporation and K_d is the apparent nucleotide equilibrium dissociation constant. Differences in nucleotide binding affinities ($K_d^{-1(incorrect)}/K_d^{-1(correct)}$) only account for a factor of $\approx 10^{-1}–10^{-2}$ in discrimination[40], whereas differences in the polymerization rates ($k_{pol}^{incorrect}/k_{pol}^{correct}$) account for $\approx 10^{-3}$.

The mechanisms that lower the values of $k_{pol}^{incorrect}$ relative to $k_{pol}^{correct}$ remain poorly understood. Many decades ago, Topal and Fresco postulated that the frequency of tautomerization may be an important determinant of misincorporation probability[11]. Interestingly, the population of the tautomeric species $\approx 10^{-3}$ is comparable to the $k_{pol}^{incorrect}/k_{pol}^{correct}$ values. In addition, the rate at which the wobble dG•dT forms either the WC-like tautomeric ($k_{GS\rightarrow ES1}$ = 0.3-10 s^{-1}; Supplementary Tables 1,5) or anionic ($k_{GS\rightarrow ES2}$ = 1.1-124 s^{-1}; Supplementary Tables 1,5) mismatches (Fig. 4a) is comparable to $k_{pol}^{incorrect}$ (0.16-1.16 s^{-1}) measured for incorrect dG•dTTP or dGTP•dT misincorporation[22,39] whereas it is up to ≈1000-fold slower than $k_{pol}^{correct}$ (25-275 s^{-1}) measured for correct dG•dCTP or dGTP•dC[22,39]. If formation of WC-like dG•dT mismatches (Fig. 4a) is required for misincorporation following initial dNTP binding in a wobble conformation, it could provide a mechanism for lowering $k_{pol}^{incorrect}$ relative to $k_{pol}^{correct}$. Indeed, prior studies have shown that DNA polymerases cannot undergo the necessary conformational changes needed for catalysis when dG•dT is in a wobble conformation[7] and all available structures of catalytically active polymerases with bound mismatches within the active site feature WC-like dG•dT or dA•dC geometries[6,7]. Similarly, WC-like rG•rU mismatches have been shown to form in the first and second codon positions

of catalytically active ribosomes[9], in which wobbles are typically rejected[5], potentially helping to explain translational error hotspots[41].

To examine this possibility, we built a kinetic model for dG•dTTP misincorporation by inserting a tautomerization/ionization step (Fig. 4a) following initial nucleotide binding in a wobble conformation and prior to the pre-chemistry conformational change in the existing minimal kinetic model for correct incorporation[42] (Fig. 4b). All other steps, including the pre-chemistry conformational change and phosphodiester bond formation, are assumed to have identical kinetic parameters as measured for correct nucleotide incorporation[39,43–46] (Supplemental Table 7). The model assumes that misincorporation directly from the wobble conformation is negligible and that the tautomerization and ionization rates measured in duplex DNA by NMR approximate the rates in the polymerase active site. We tested models (Extended Data Fig. 7) in which either the tautomeric (M_{ES1}), anionic (M_{ES2}), or both ($M_{ES1+ES2}$) species can be misincorporated as well as models that excluded the triangular network all together (M_{Kd}).

Strikingly, the most general $M_{ES1+ES2}$ model robustly predicts the measured F_{pol} values for three polymerases (T7, polymerase ε, and polymerase β) that have varying rate limiting steps and microscopic rate constants (Fig. 5a; Extended Data Fig. 8; and Supplementary Table 7). Similar results are obtained with M_{ES1} under these neutral conditions in which the ES2 population is insignificant ($<10^{-6}$ at pH 6.9) (Fig. 5a and Extended Data Fig. 8). In contrast, M_{ES2} consistently underestimates F_{pol} by one to two orders of magnitude whereas M_{Kd} overestimates F_{pol} by one to two orders of magnitude (Fig. 5a)[47]. Variants of the M_{Kd} model in which only pre-formed tautomeric dNTP with populations of 10^{-4}–10^{-5} bind in a productive WC-like geometry overestimates k_{pol} and K_d by several orders of magnitude (data not shown). These data indicate that formation of tautomeric WC-like dG^{enol}•dT and dG•dT^{enol} with population of ≈0.1% can account for the ≈10^2–10^3–fold lower value of $k_{pol}^{incorrect}$ relative to $k_{pol}^{correct}$ and that at neutral pH, >99% of misincorporation proceeds via the tautomeric species, which form predominantly via direct exchange from the wobble (Fig. 5b).

Impact of pH, modifications, and sequence

We also examined whether the $M_{ES1+ES2}$ model can reproduce the dependence of misincorporation probability on pH, base modification, and sequence. $M_{ES1+ES2}$ accurately predicts the ≈3-fold increase in misincorporation probability observed with increasing pH (Fig. 5c; left). This can be attributed to an increase in the population of dG•dT$^-$, which accounts for >70% of the net misincorporation at pH 8.4 (Fig. 5b,c). In contrast, M_{ES1} fails to predict this increase in misincorporation probability (Fig. 5c, left; M_{Kd} not shown due to absence of pH-dependent K_d values). Under high pH, the tautomeric and anionic species have comparable populations, and there is significant flux (>20%) toward both tautomeric and anionic species through the indirect minor exchange pathway (Fig. 5b). In this manner, the tautomeric and anionic contributions to misincorporation are coupled.

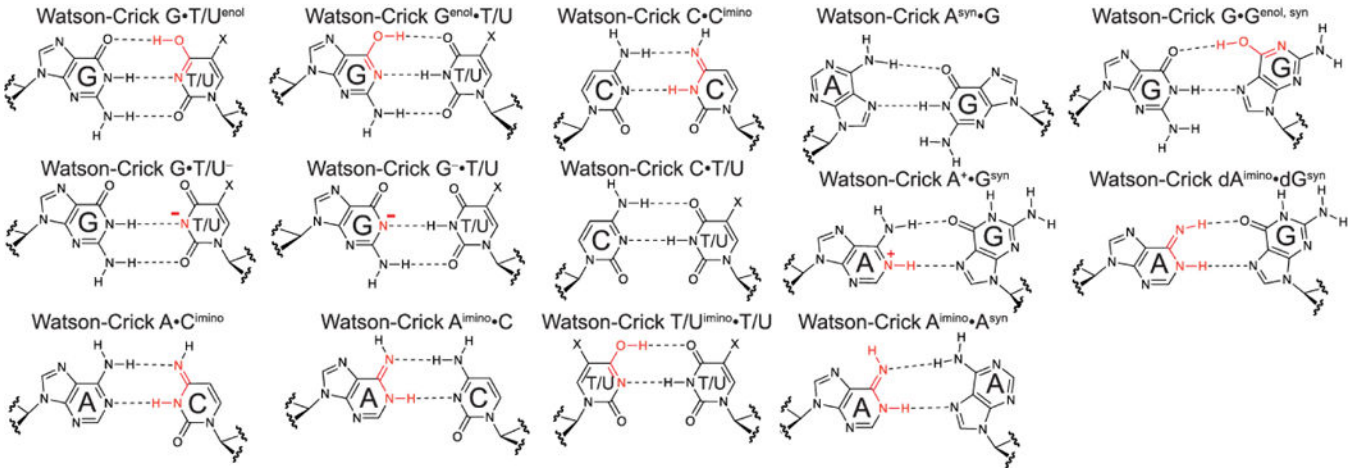
$M_{ES1+ES2}$ and NMR RD measurements also accurately predict F_{pol} and k_{obs} for 5-bromo-2′-deoxyuridine triphosphate (^{5Br}dUTP) (Fig. 5c). This includes a steeper ≈6-fold increase in

$F_{pol}^{dG•5BrdUTP}$ measured for AMV RT when increasing the pH from 7.0 to 8.4 (Fig. 5c; left). This can be attributed to the lower pK_a of dG•^{5Br}dU$^-$ (pK_a ≈9)[14] relative to dG•dT$^-$ (pK_a ≈11.8)[8]. We further verified the robustness of these predictions by measuring $k_{obs}^{dG•5BrdUTP}$ and k_{obs}^{dTTP} for human DNA polymerase β at high pH (8.4). The model accurately predicts the ≈4-fold enhancement in $k_{obs}^{dG•5BrdUTP}$ relative to k_{obs}^{dTTP} (Fig. 5c; right). Again, M_{ES1} fails to predict these variations (Fig. 5c). Indeed, at both neutral and high pH, ^{5Br}dUTP is predicted to be predominantly misincorporated via the more populated dG•dT$^-$ (Fig. 5b,c). These data indicate that misincorporation due to dG•dT$^-$ can dominate at pH≥8.4 or for chemically modified nucleotides at neutral pH.

Importantly, due to the sequence-dependence of tautomerization and ionization, $M_{ES1+ES2}$ also predicts ≈8-fold sequence-specific variations in F_{pol} at pH 8.4 (Fig. 5d). Comparable (5-fold) sequence-specific variations have been reported previously[21]. We tested these predictions using human DNA polymerase β at pH 8.4 for nine different sequence contexts (Supplementary Table 8). While $k_{obs}^{dG•dCTP}$ varied weakly (<1.2-fold) with sequence, $k_{obs}^{dG•dTTP}$ varied ≈45-fold (Extended Data Fig. 9), with larger changes observed when varying the base pair at the n-1 position, which stacks with dG•dTTP in the polymerase active site (Fig. 5d). While the $M_{ES1+ES2}$ predictions slightly underestimate the sequence-specific variations in $k_{obs}^{dG•dTTP}$, this is not too surprising considering that other microscopic steps could also vary with sequence. The predictions do recapitulate the lower $k_{obs}^{dG•dTTP}$ for CGA and comparable values for GGC and CGC (Fig. 5d). Interestingly, the two major outliers (TGA and GGG) arise primarily due to a large ES2 population. It is likely that the polymerase environment, including absence of base pairs at the n+1 position (Fig. 5d), can influence the sequence-specific dependence of tautomerization/ionization and consequently misincorporation.

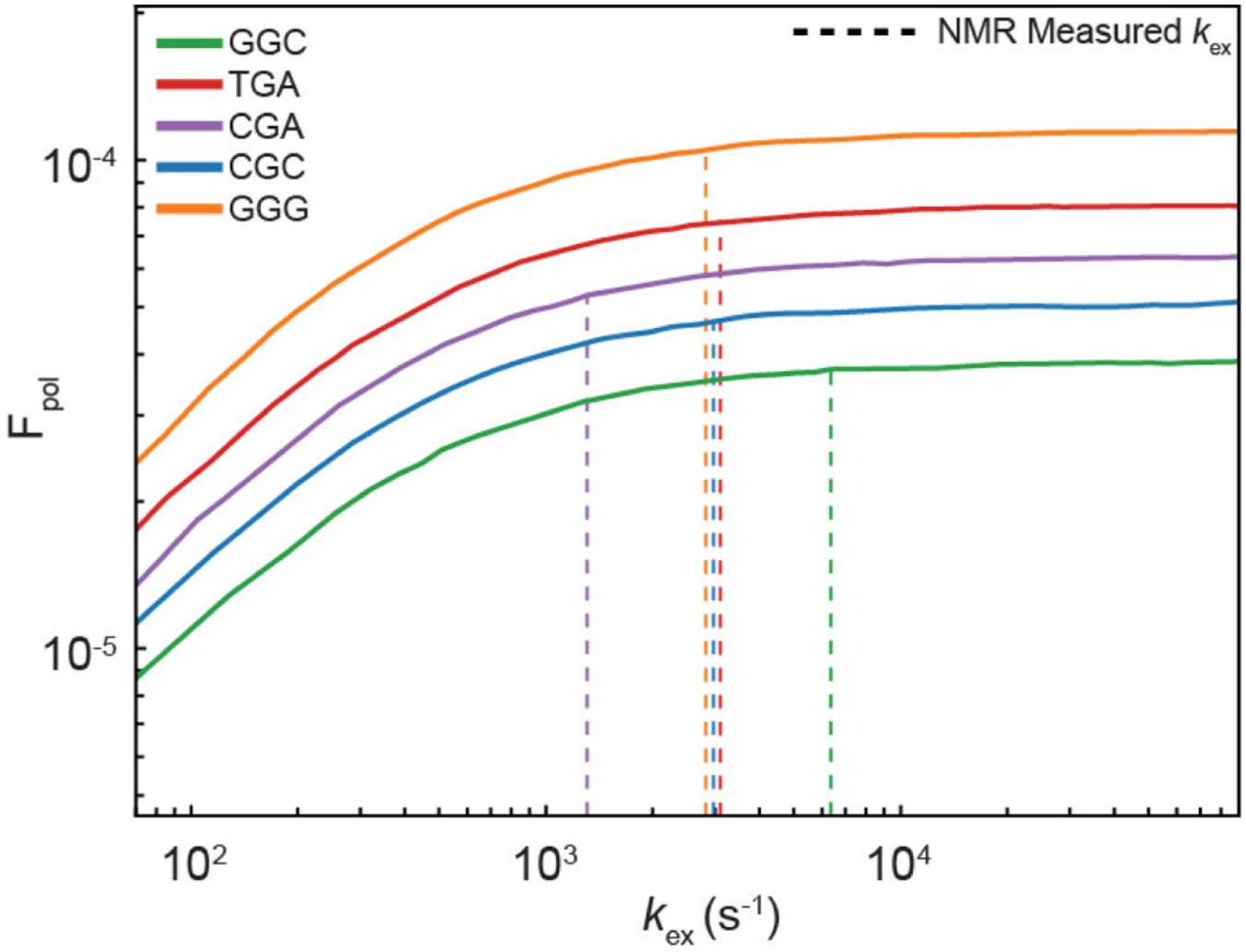
Our data indicate that formation of WC-like anionic and tautomeric mismatches help determine the frequency of dG•dT misincorporation and its dependence on pH, chemical modifications, and possibly sequence. Our analysis indicates that F_{pol} is determined primarily by the ES1 population and that significant reductions in $k_{ex} = k_{GS→ES1} + k_{ES1→GS}$, outside the range detected here, would be required to significantly reduce F_{pol} (Extended Data Fig. 10). While it is likely that differences in the polymerase active site environment will tune tautomerization/ionization dynamics, the robustness of the predictions across different polymerases, pH conditions, and modified nucleotides suggests that it will not cause significant perturbations relative to the broad kinetic range examined here. Indeed, very small differences in tautomerization/ionization dynamics are observed for DNA and RNA, which have different helical structures and stabilities. It is possible that tautomerization/ionization is dominated by the energetics of hydrogen-bonding and proton transfer and that the natural grip for the WC geometry in the double helix is similar to that achieved by the polymerase in the context of an isolated dNTP paired to the template. Other mechanisms may be applicable for purine-purine mismatches where alterations in the active site have been proposed rather than adoption of a WC-like base pair[43,48]. The approach presented here can be applied to examine the roles of other tautomeric and anionic mismatches in replication, transcription, translation, and mismatch repair[49].

Extended Data



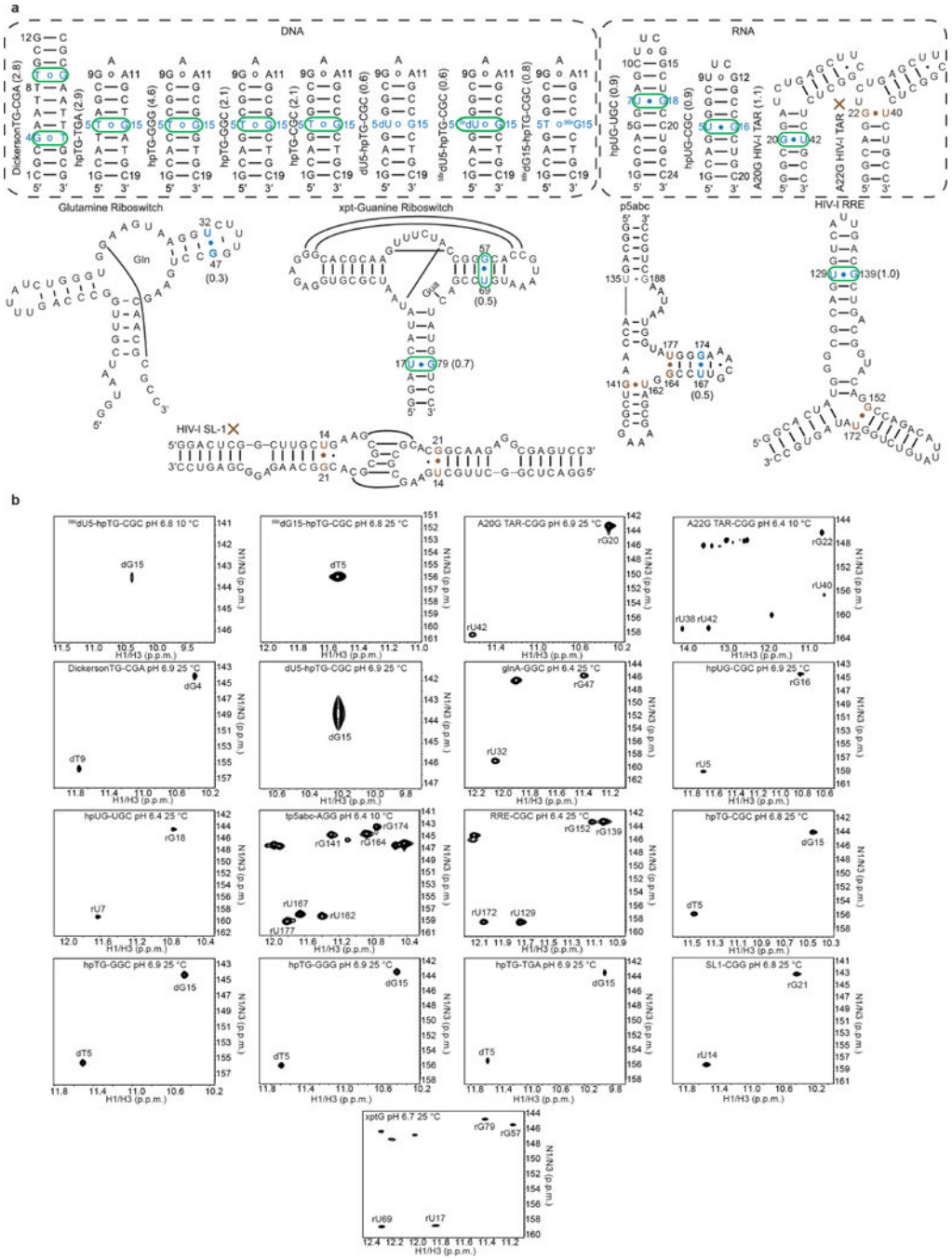
Extended Data Figure 1. Watson-Crick-like mismatches

a, Watson-Crick-like mismatches stabilized by tautomeric and ionic base forms. Tautomeric purine•pyrimindine and purine•purine mismatches were first proposed by Topal and Fresco[11]. For G•T/U mismatches, X = H or CH_3 for uridine and thymidine, respectively.



Extended Data Figure 2. DNA and RNA constructs used in this study

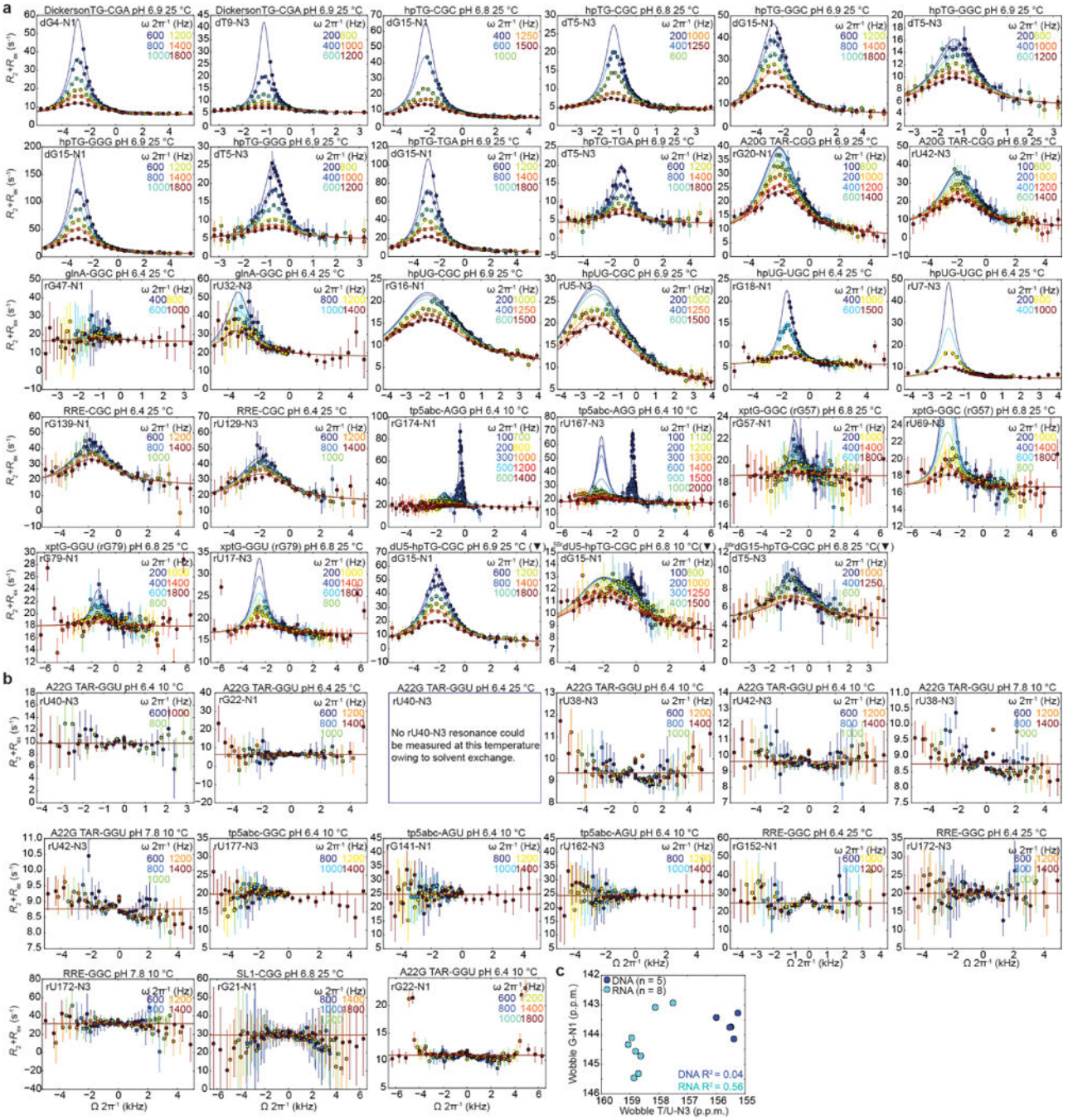
a, Secondary structures of the various DNA and RNA constructs used in this study. G•T/U mismatches that show signs of chemical exchange directed toward tautomeric and/or anionic WC-like mismatches are highlighted in blue and green, respectively. G•T/U mismatches that show no evidence for WC-like RD are highlighted in brown. The value of K_t measured at near-neutral pH is shown next to each mismatch. The DickersonTG-CGA, hpTG-CGC and hpUG-CGC sequences contexts were studied in a prior publication[8]. **b,** 2D [^{15}N, ^{1}H] HSQC spectra of DNA and RNA constructs used in this study showing the imino resonances of G-N1/H1 and T/U-N3/H3 targeted for RD measurements. Spectra shown for xptG was collected at pH 6.7 and 25 °C in potassium acetate buffer described elsewhere[8].



Extended Data Figure 3. RD profiles measured in DNA and RNA at near-neutral pH

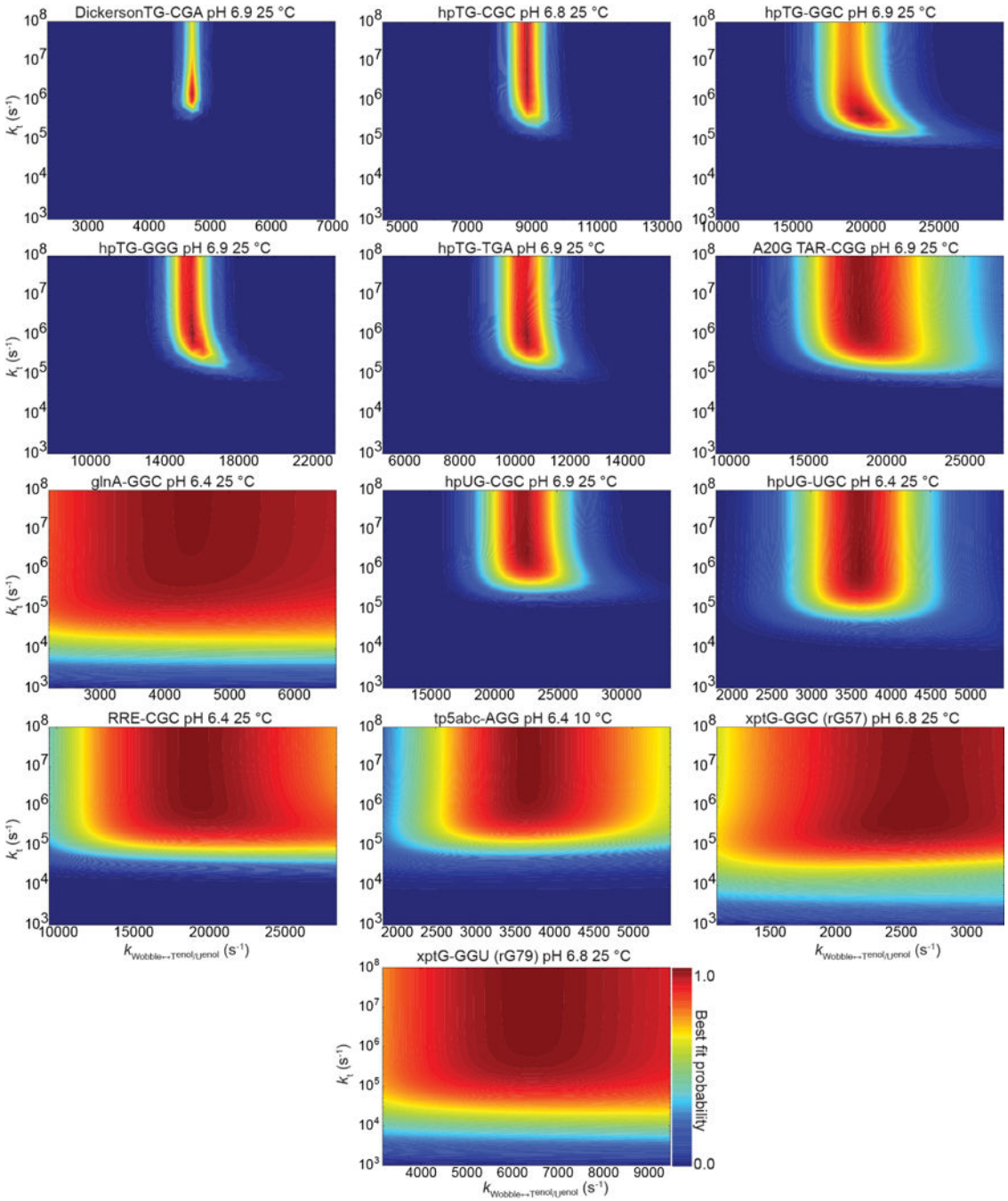
a, ^{15}N G-N1 and T/U-N3 RD measured for G•T/U mismatches at pH 6.4-6.9 and 10-25 °C showing wobble⇌tautomer exchange. Note that in addition to wobble⇌tautomer exchange, tp5abc undergoes an independent slower exchange process involving a change in secondary structure that is described in detail elsewhere[53]. The trend lines represent B-M 2-/3-state fits. Constructs containing a chemically modified base are indicated with (▼).**b**, Absence of ^{15}N RD for rG•rU mismatches near bulges, apical loops, or three-way junctions. Error bars reflect experimental uncertainty (one s.d., see Methods). **c**, No correlation is observed

between ground state (GS) wobble G-N1 and T/U-N3 chemical shifts for DNA (n = 5) or RNA (n = 8). Error bars in **a**, **b** reflect experimental uncertainty (one s.d., see Methods).



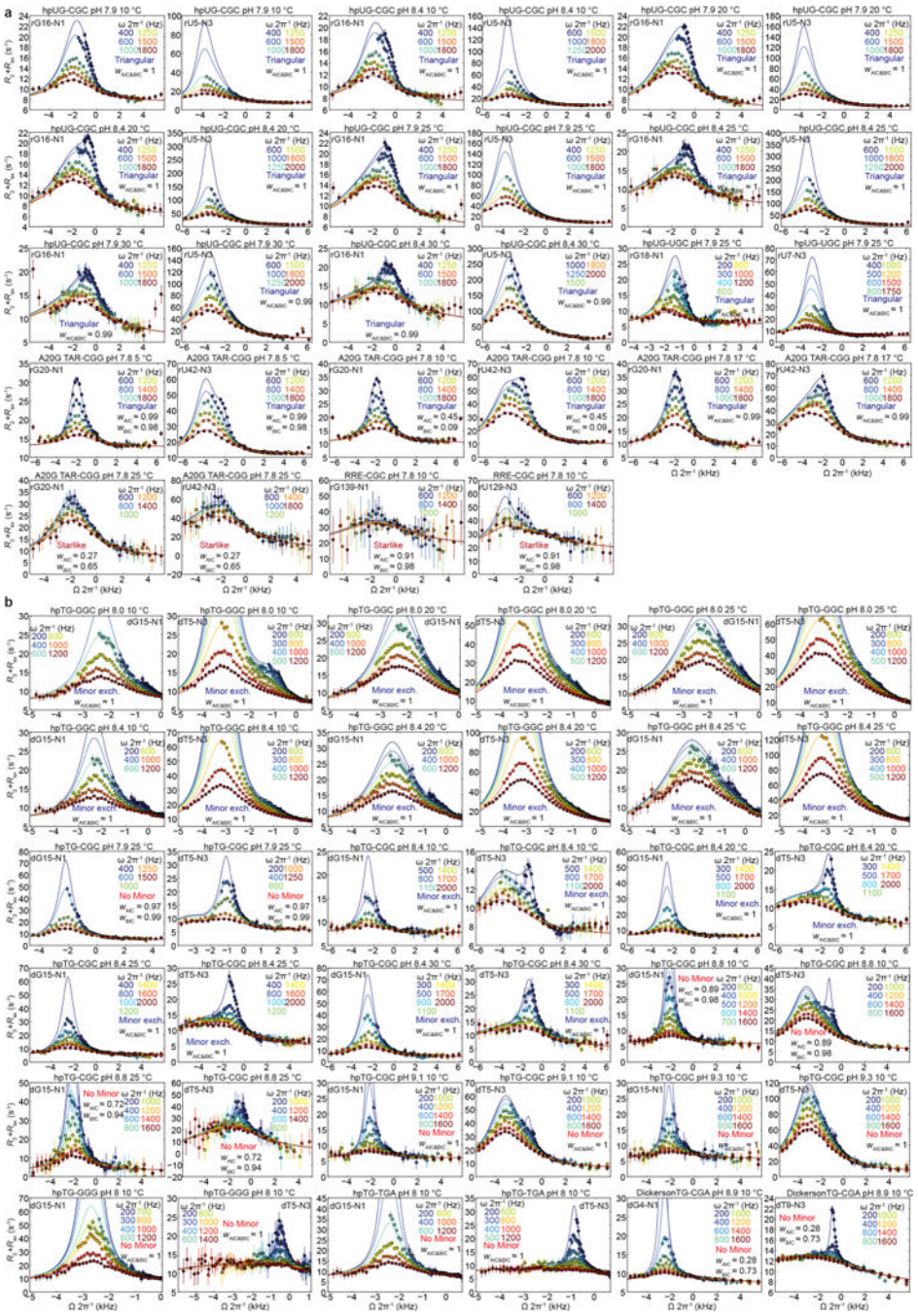
Extended Data Figure 4. Establishing lower limits for rates of base pair tautomeric exchange

Agreement between measured and predicted $R_{1\rho}$ values (scaled $\overline{\chi}^2$ weight, Eq. 2) when varying the wobble⇌G•T^{enol}/U^{enol} ($k_{GS \rightleftharpoons T^{enol}/U^{enol}}$) and G•T^{enol}/U^{enol} ⇌G^{enol}•T/U (k_t) rate of exchange. See Methods for additional details.



Extended Data Figure 5. RD profiles measured in DNA and RNA at high pH

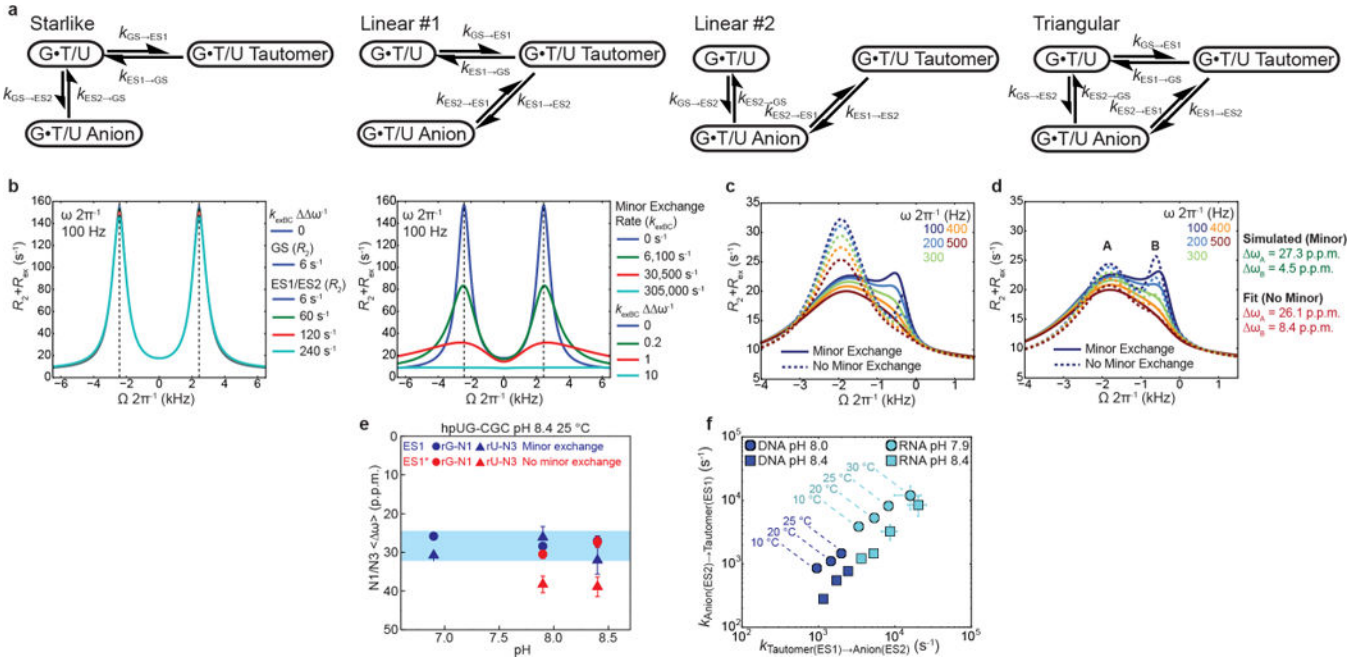
B-M 3-state fits of **a**, RNA and **b**, DNA ^{15}N RD data for starlike and triangular topologies (indicated within the plots). The relative statistical Akaike's Information Criterion (w_{AIC}) and Bayesian Information Criterion (w_{BIC}) weights[50,60] for each fit were used to select the model (representative starlike versus triangular, comparisons with linear models shown in Supplementary Table 6). Error bars reflect experimental uncertainty (one s.d., see Methods).



Extended Data Figure 6. Discerning minor exchange between WC-like tautomeric and anionic G•T/U mismatches

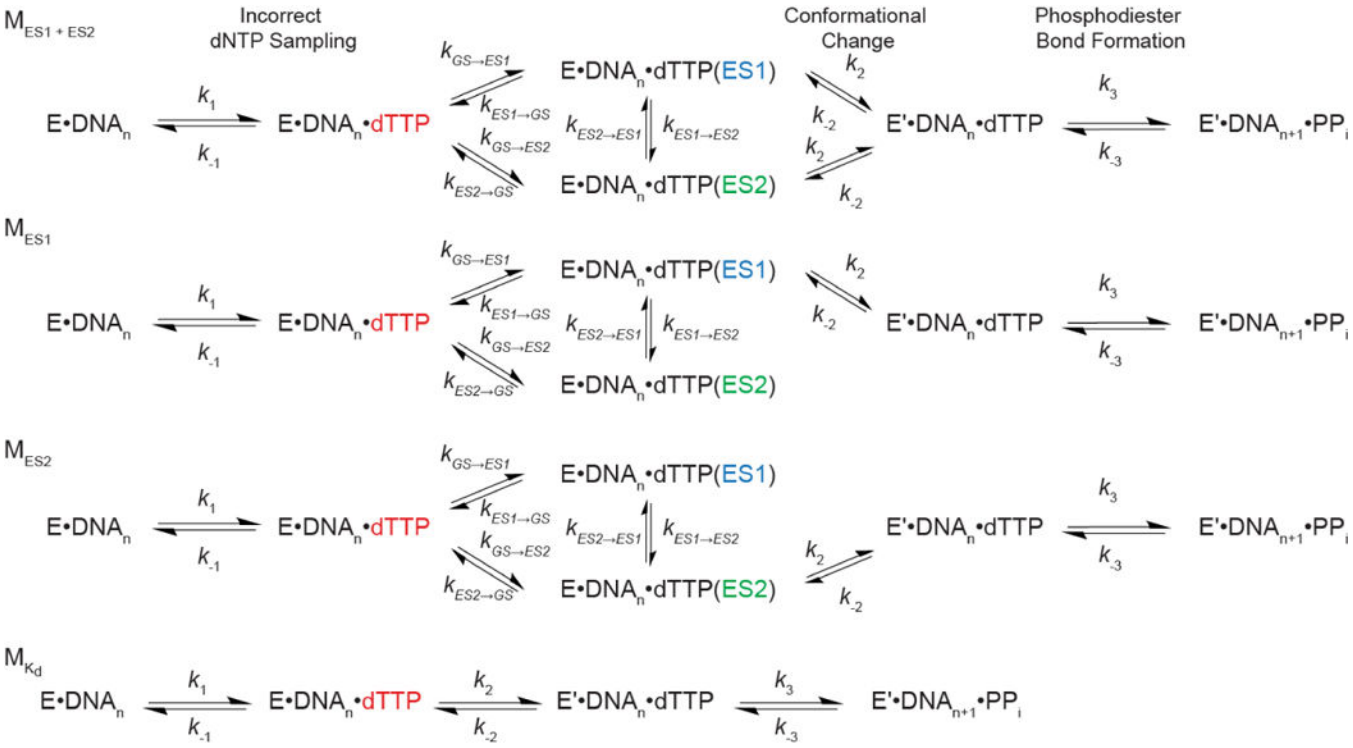
a, Topologies used to model chemical exchange. Individual rate constants are shown for each leg of the different topologies. **b**, Left: B-M simulations showing that when $R_{2(GS)} \neq R_{2(ES1)}$ and $R_{2(GS)} \neq R_{2(ES2)}$ no apparent peak asymmetry is observed. Right: B-M simulations showing that minor exchange between two ESs in a triangular topology induces asymmetry in the RD profiles and opposite changes in the apparent chemical shift for the two ESs. **c**, B-M simulations (solid lines) showing the fitted exchange parameters for hpUG-

CGC at pH 8.4 and 10 °C (Supplementary Table 5) when including minor exchange in a triangular topology. For comparison, simulations using the same parameters without minor exchange (k_3 = 0 and k_{-3} = 0) are also shown (dashed lines). **d**, Dashed lines denote 3-state B-M best fit to starlike topology ($k_{ES1\rightarrow ES2}$ = 0 and $k_{ES2\rightarrow ES1}$ = 0) to data simulated with triangular topology with minor exchange (solid lines). Shown to the right is the over/under estimation of the true (green) versus fitted (red) ES chemical shifts when fitting RD profiles with triangular topology with minor exchange to a starlike model that has no minor exchange. **e**, The ES1 $<\Delta\omega_{rG\text{-}N1}>$ and $<\Delta\omega_{rU\text{-}N3}>$ values as a function of pH derived from the 3-state B-M fit with triangular and starlike topology. **f**, Forward ($k_{ES1\rightarrow ES2}$) and reverse ($k_{ES2\rightarrow ES1}$) minor exchange rate constants for hpTG-GGC (pH 8 and 8.4) and hpUG-CGC (pH 7.9 and 8.4) as a function of temperature. Error bars in **e**, **f** reflect experimental uncertainty (one s.d., Methods).



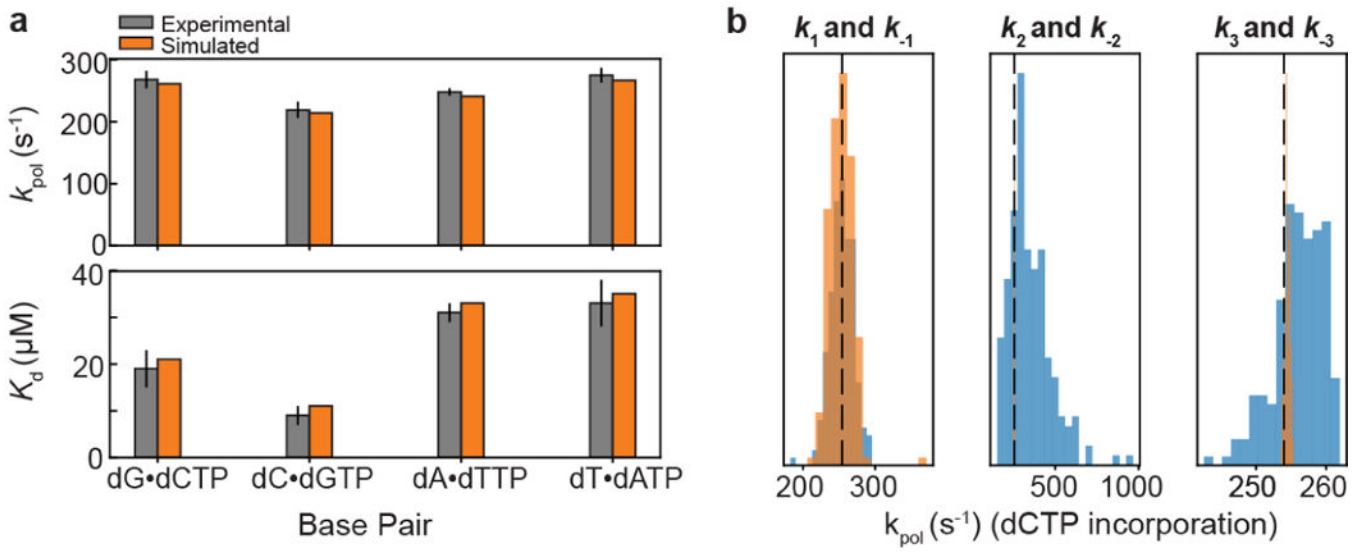
Extended Data Figure 7. Kinetic mechanisms used to model misincorporation

Rate constants for each step are listed in Supplementary Table 7.



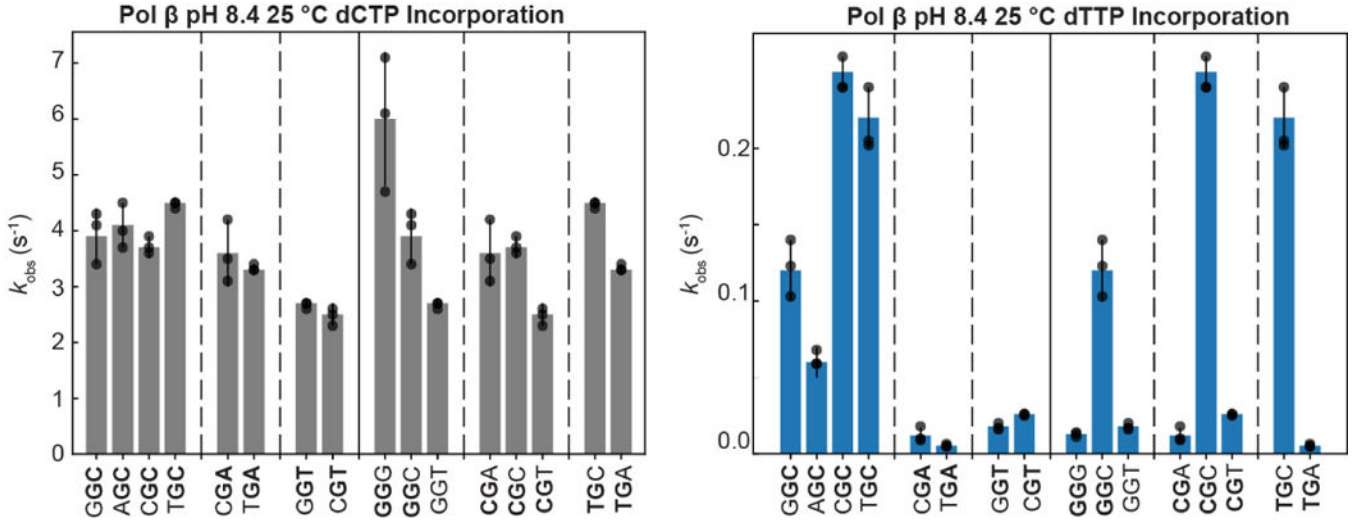
Extended Data Figure 8. Benchmarking kinetic simulations of misincorporation

a, Comparison of k_{pol} and K_d values for correct incorporation measured experimentally for human DNA polymerase ε with values computed based on pre-steady state simulations using the microscopic rate constants provided in ref. 39. Error bars reflect fitting uncertainty as previously published[39]. **b**, Robustness of calculated k_{pol} values for human DNA polymerase ε when varying rate constants (forward rate constant: blue; and reverse rate constant: orange) for steps other than tautomerization/ionization by 2-fold (n = 200 independent simulations in which rate constants were randomly varied within 2-fold). As expected, the only rate constant with a substantial effect on the reported k_{pol} values was the rate limiting k_2 conformational change step (middle).



Extended Data Figure 9. k_{obs} values measured for human DNA polymerase β dCTP•dG (left) or dTTP•dG (right) insertion at pH 8.4, 25 °C, and 100 μM dNTP

DNA template sequence (5′ to 3′) is read from bottom (n+1 position) to top (n-1 position). Individual replicates (n = 3 independent experiment) are indicated by grey circles. Bar height reflect average of replicates and error bars reflect one s.d.



Extended Data Figure 10. F_{pol} is primarily governed by ES1 populations

Simulated F_{pol} values as a function of scaling up or down the kinetic exchange rate for ES1 formation (k_{ex} = $k_{GS\rightarrow ES1}$ + $k_{ES1\rightarrow GS}$) without altering the ES1 population. Increasing k_{ex} beyond values measured in this study experimentally (green dotted line) minimally affects F_{pol}. Decreasing the rate of exchange within the range measured experimentally in this study (purple dotted line) also minimally affects the value of F_{pol}. Much larger decreases in k_{ex} are required to significantly reduce the value of F_{pol}.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank members of the Al-Hashimi laboratory and Prof. T. Oas (Duke University) for stimulating discussions and critical input. We acknowledge technical support and resources from the Duke Magnetic Resonance Spectroscopy Center and the Duke Shared Cluster Resource. This work was supported by grants from the National Institutes of Health (NIH R01GM089846, P01GM0066275, and P50GM103297) and an Agilent Thought Leader Award given to H.M.A., and a grant from the National Science Foundation (MCB-1716168) to Z.S. W.J.Z. was supported by a Pelotonia Graduate Fellowship from The Ohio State University Comprehensive Cancer Center. The authors declare no conflict of interest.

References

1. Watson JD, Crick FH. The structure of DNA. Cold Spring Harb Symp Quant Biol. 1953; 18:123–131. [PubMed: 13168976]
2. Oertell K, et al. Kinetic selection vs. free energy of DNA base pairing in control of polymerase fidelity. Proc Natl Acad Sci USA. 2016; 113:E2277–E2285. [PubMed: 27044101]
3. Voorhees RM, Ramakrishnan V. Structural basis of the translational elongation cycle. Annu Rev Biochem. 2013; 82:203–236. [PubMed: 23746255]
4. Kool ET. Active site tightness and substrate fit in DNA replication. Annu Rev Biochem. 2002; 71:191–219. [PubMed: 12045095]
5. 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]
6. Wang W, Hellinga HW, Beese LS. Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. Proc Natl Acad Sci USA. 2011; 108:17644–17648. [PubMed: 22006298]
7. Bebenek K, Pedersen LC, Kunkel TA. Replication infidelity via a mismatch with Watson-Crick geometry. Proc Natl Acad Sci USA. 2011; 108:1862–1867. [PubMed: 21233421]
8. Kimsey IJ, Petzold K, Sathyamoorthy B, Stein ZW, Al-Hashimi HM. Visualizing transient Watson-Crick-like mispairs in DNA and RNA duplexes. Nature. 2015; 519:315–320. [PubMed: 25762137]
9. Rozov A, Demeshkina N, Westhof E, Yusupov M, Yusupova G. Structural insights into the translational infidelity mechanism. Nat Commun. 2015; 6:7521. [PubMed: 26138272]
10. Topal MD, Fresco JR. Base pairing and fidelity in codon-anticodon interaction. Nature. 1976; 263:289–293. [PubMed: 958483]
11. Topal MD, Fresco JR. Complementary base pairing and the origin of substitution mutations. Nature. 1976; 263:285–289. [PubMed: 958482]
12. Loveland AB, Demo G, Grigorieff N, Korostelev AA. Ensemble cryo-EM elucidates the mechanism of translation fidelity. Nature. 2017; 546:113–117. [PubMed: 28538735]
13. Singh V, Fedeles BI, Essigmann JM. Role of tautomerism in RNA biochemistry. RNA. 2015; 21:1–13. [PubMed: 25516996]
14. 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]
15. Koag MC, 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]
16. Rozov A, Demeshkina N, Westhof E, Yusupov M, Yusupova G. New Structural Insights into Translational Miscoding. Trends Biochem Sci. 2016; 41:798–814. [PubMed: 27372401]
17. Tomasetti C, Li L, Vogelstein B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. Science. 2017; 355:1330–1334. [PubMed: 28336671]
18. 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]

19. Warren JJ, Forsberg LJ, Beese LS. The structural basis for the mutagenicity of O^6-methyl-guanine lesions. Proc Natl Acad Sci USA. 2006; 103:19701–19706. [PubMed: 17179038]
20. 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]
21. 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]
22. Lee HR, Johnson KA. Fidelity of the human mitochondrial DNA polymerase. J Biol Chem. 2006; 281:36236–36240. [PubMed: 17005554]
23. 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]
24. Ogle JM, Murphy 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]
25. 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.
26. 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. 2015; 33:925–945. [PubMed: 24842163]
27. Trott O, Abergel D, Palmer AG III. An average-magnetization analysis of $R_{1\rho}$ relaxation outside of the fast exchange limit. Mol Phys. 2003; 101:753–763.
28. 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. 2005; 127:713–721. [PubMed: 15643897]
29. 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]
30. Mulder FAA, Mittermaier A, Hon B, Dahlquist FW, Kay LE. Studying excited states of proteins by NMR spectroscopy. Nat Struct Biol. 2001; 8:932–935. [PubMed: 11685237]
31. Szymanski ES, Kimsey IJ, Al-Hashimi HM. Direct NMR Evidence that Transient Tautomeric and Anionic States in dG•dT Form Watson-Crick-like Base Pairs. J Am Chem Soc. 2017; 139:4326–4329. [PubMed: 28290687]
32. 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.
33. Ban D, et al. Exceeding the limit of dynamics studies on biomolecules using high spin-lock field strengths with a cryogenically cooled probehead. J Magn Reson. 2012; 221:1–4. [PubMed: 22743535]
34. Vallurupalli P, Chakrabarti N, Pomes R, Kay L. Atomistic picture of conformational exchange in a T4 lysozyme cavity mutant: an experiment-guided molecular dynamics study. Chem Sci. 2016; 7:3062–3613.
35. Delaney JC, Essigmann JM. Effects of Sequence Context of O^6-methylguanine Repair and Replication in Vivo. Biochemistry. 2001; 40:14968–14975. [PubMed: 11732917]
36. 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]
37. Sekhar A, et al. Thermal fluctuations of immature SOD1 lead to separate folding and misfolding pathways. Elife. 2015; 4:e07296. [PubMed: 26099300]
38. Kunkel TA, Alexander PS. The base substitution fidelity of eukaryotic DNA polymerases. J Biol Chem. 1986; 261:160–166. [PubMed: 3941068]
39. Zahurancik WJ, Klein SJ, Suo Z. Significant contribution of the $3'\rightarrow5'$ exonuclease activity to the high fidelity of nucleotide incorporation catalyzed by human DNA polymerase ε. Nucleic Acids Res. 2014; 42:13853–13860. [PubMed: 25414327]

40. Loeb L, Kunkel TA. Fidelity of DNA synthesis. Annu Rev Biochem. 1982; 51:429–457. [PubMed: 6214209]
41. Zhang J, Ieong KW, Johansson M, Ehrenberg M. Accuracy of initial codon selection by aminoacyl-tRNAs on the mRNA-programmed bacterial ribosome. Proc Natl Acad Sci USA. 2015; 112:9602–9607. [PubMed: 26195797]
42. Rothwell PJ, Waksman G. Structure and mechanism of DNA polymerases. Adv Protein Chem. 2005; 71:401–440. [PubMed: 16230118]
43. Tsai YC, Johnson KA. A new paradigm for DNA polymerase specificity. Biochemistry. 2006; 45:9675–9687. [PubMed: 16893169]
44. Patel SS, Wong I, Johnson KA. Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. Biochemistry. 1991; 30:511–525. [PubMed: 1846298]
45. Zahurancik WJ, Klein SJ, Suo Z. Kinetic mechanism of DNA polymerization catalyzed by human DNA polymerase ε. Biochemistry. 2013; 52:7041–7049. [PubMed: 24020356]
46. Balbo PB, Wang EC, Tsai MD. Kinetic mechanism of active site assembly and chemical catalysis of DNA polymerase β. Biochemistry. 2011; 50:9865–9875. [PubMed: 22010960]
47. Maximoff SN, Kamerlin SCL, Florian J. DNA Polymerase lambda Active Site Favors a Mutagenic Mispair between the Enol Form of Deoxyguanosine Triphosphate Substrate and the Keto Form of Thymidine Template: A Free Energy Perturbation Study. J Phys Chem B. 2017; 121:7813–7822. [PubMed: 28732447]
48. Freudenthal BD, Beard WA, Shock DD, Wilson SH. Observing a DNA polymerase choose right from wrong. Cell. 2013; 154:157–168. [PubMed: 23827680]
49. Freudenthal BD, Beard WA, Cuneo MJ, Dyrkheeva NS, Wilson SH. Capturing snapshots of APE1 processing DNA damage. Nat Struct Mol Biol. 2015; 22:924–931. [PubMed: 26458045]

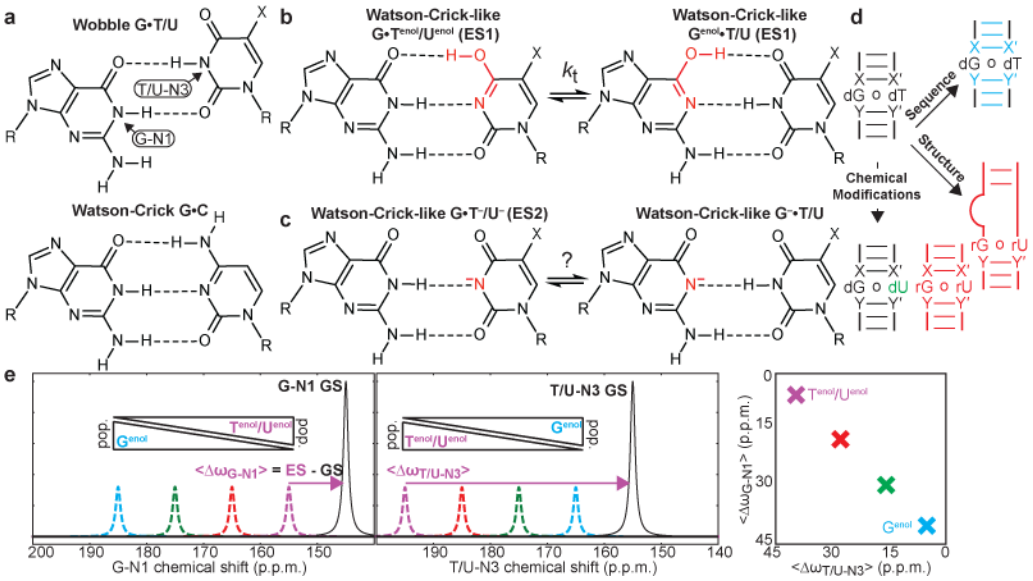


Figure 1. Tilting rapid tautomeric equilibria in excited state WC-like mismatches

Chemical structures of ground state wobble (**a**) and excited state tautomeric (**b**) and anionic (**c**) WC-like G•T/U (X = H or CH_3 for uridine and thymidine, respectively) mismatches. **d**, Tilting rapid tautomeric equilibrium using sequence (cyan), structure (red), and chemical modifications (green). X-X′ and Y-Y′ denote WC base pairs adjacent to the G•T/U mismatches. **e**, Perturbations that differentially tilt the tautomeric equilibrium are expected to give rise to anti-correlated linear changes in the $<\Delta\omega_{G\text{-}N1}>$ and $<\Delta\omega_{T/U\text{-}N3}>$ values (color-coded).

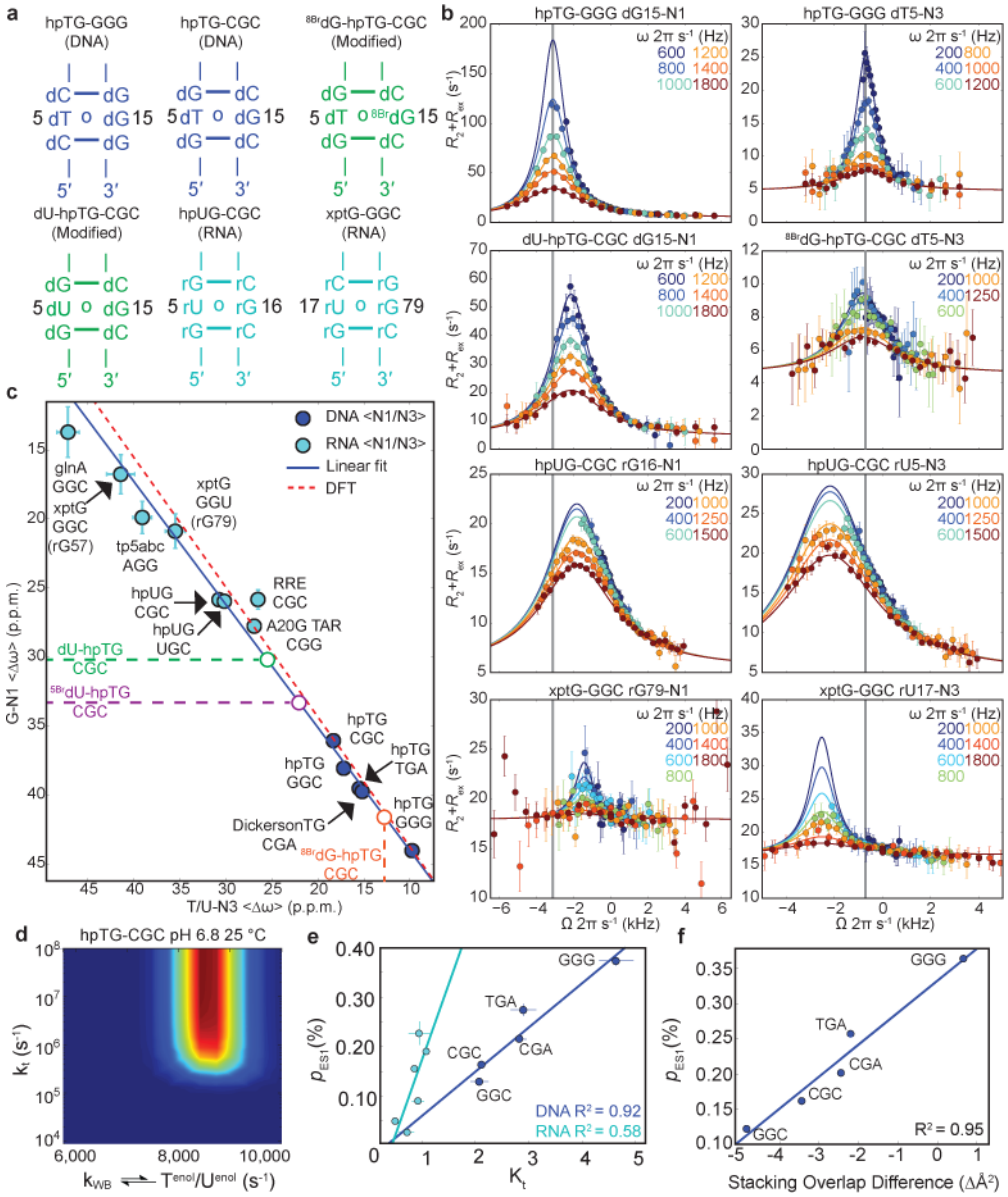


Figure 2. Resolving rapidly interconverting tautomers

a, Representative hairpin (“hp”) DNA, RNA, and chemically modified constructs. Name denotes mismatch and sequence context (5′-3′). **b**, Representative G-N1 and T/U-N3 RD profiles (pH 6.8-6.9 and 25 °C). Best fits to the B-M equations are shown. **c**, $<\Delta\omega_{T/U\text{-}N3}>$ versus $<\Delta\omega_{G\text{-}N1}>$ ES1 chemical shift differences measured by NMR RD. Blue line indicates fit to the Δω’s using the fundamental tautomer chemical shifts as variables (Methods). Red line indicates DFT-predictions. **d**, Lower bounds for the rate of tautomeric G^{enol}•T/U⇌G•T^{enol}/U^{enol} exchange. Contour plots showing scaled $\overline{\chi}^2$ weights for combinations of $k_{Tenol/Uenol}$ versus k_t; red indicates better fit. **e**, A plot of K_t versus tautomeric ES1 population for DNA (blue, n=5) and RNA (cyan, n=6) constructs determined at pH 6.9 and 25 °C. **f**, Change in stacking overlap ($\Delta Å^2 = Å^{2(WC)} - Å^{2(WB)}$) between wobble and Watson-Crick-like mismatches versus ES1 population (pH 6.9 and 25 °C) for five DNA sequence

contexts (Methods). Error bars in **b**, **c**, **e** reflect experimental uncertainty (one s.d., Methods).

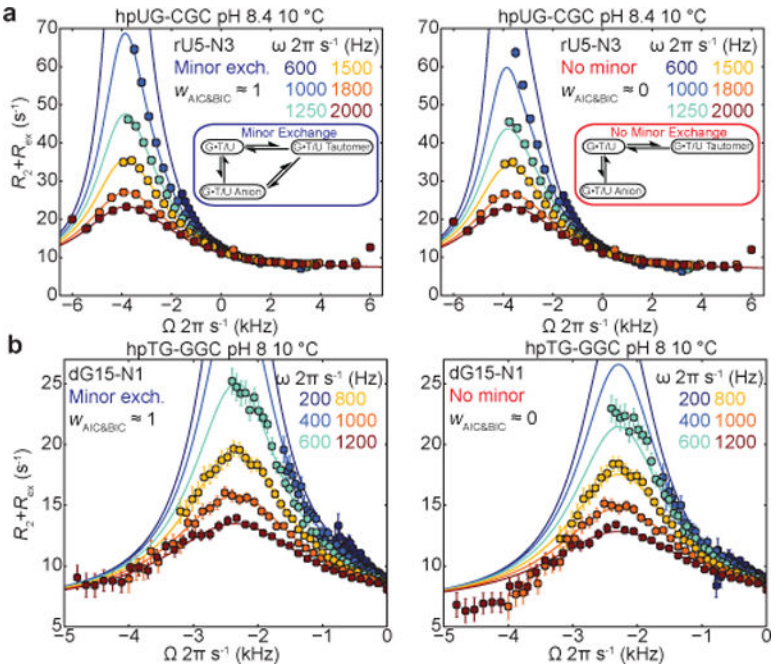


Figure 3. 3-State exchange with triangular topology and minor exchange between tautomeric and anionic WC-like excited states

Comparison of 3-state B-M fit with triangular (left) and starlike (right) topology to the RD profiles measured in (**a**) hpUG-CGC RNA and (**b**) hpTG-GGC DNA. Statistical AIC (w_{AIC}) and BIC (w_{BIC}) weights comparing starlike and triangular topologies are shown. Error bars reflect experimental uncertainty (one s.d., Methods).

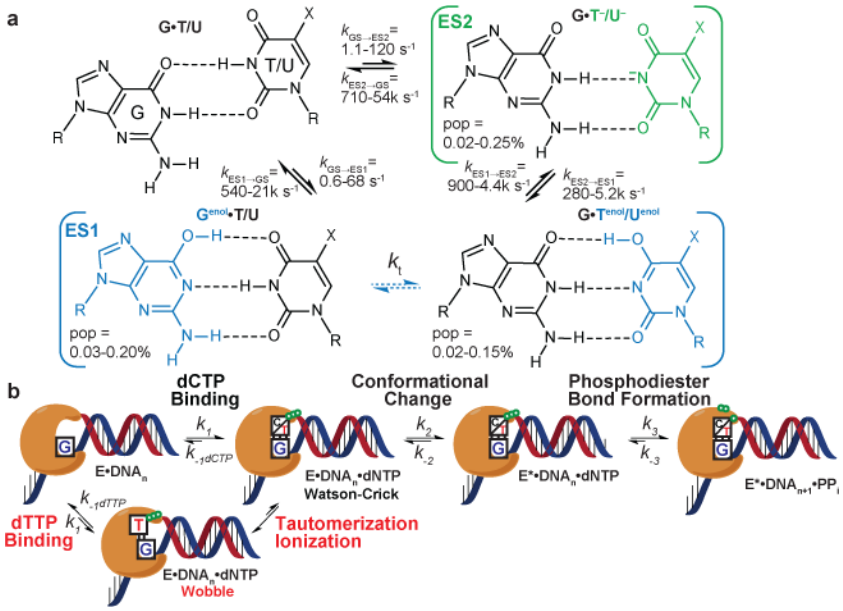


Figure 4. Kinetic mechanism for dG•dT misincorporation

a, Triangular exchange between wobble G•T/U mispair, rapidly interconverting WC-like tautomers (G^{enol}•T/U⇌G•T^{enol}/U^{enol}), and anionic WC-like G•T^{-}/U^{-}. Exchange between anionic G•T^{-}/U^{-} and a low-abundance short-lived anionic G^{-}•T/U or other non-WC species that fall outside RD detection cannot be ruled out. WC-like G•T/U populations and ranges at pH 6.4-8.9 and 10-25 °C (Supplementary Tables 1, 5). **b**, Minimal kinetic mechanism for polymerization[42]. Incorporation of an incorrect dNTP includes an additional tautomerization/ionization step allowing for the formation of a Watson-Crick-like dG•dT mismatch. Discriminatory steps are in red.

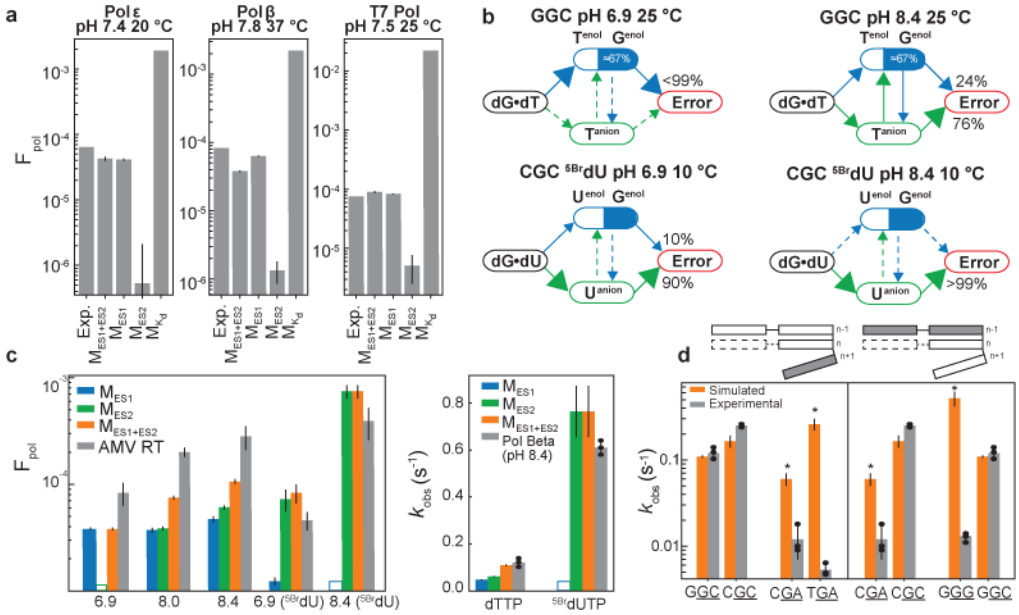


Figure 5. Measured versus predicted misincorporation probability and rates

a, F_{pol} measured experimentally for dTTP•dG misincorporation for human DNA polymerase ε, rat DNA polymerase β, and T7 DNA polymerase with values simulated using M_{ES1}, M_{ES2}, $M_{ES1+ES2}$ and M_{Kd} (error bars represent one s.d., Methods). **b**, Flux pathways for dT•dG(GGC) and ^{5Br}dU•dG(CGC). **c**, Left. Measured and simulated F_{pol} values for dTTP/^{5Br}dUTP in AMV RT[14]. Right, measured and simulated k_{obs} values for dTTP/^{5Br}dUTP misincorporation for human DNA polymerase β. Error is s.d. of n=3 biological replicates for kinetic assays, or previously published error for AMV RT[14]. Error for kinetic simulations is described in Methods. **d**, Measured and simulated k_{obs} for dTTP misincorporation for human DNA polymerase β in different sequence contexts. (*) indicates that ES2 exchange rates were extrapolated (Methods). Error is s.d. of n=3 biological replicates for kinetic assays. Error for kinetic simulations is described in Methods.