**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

*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

**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 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 $<\Delta\omega_{G\text{-}N1}>$ and $<\Delta\omega_{T/U\text{-}N3}>$ 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 d$G^{enol}$•dT in DNA ($K_t$ = 2.1-4.6) whereas the populations of r$G^{enol}$•rU and rG•r$U^{enol}$ 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 $G{\bullet}T^{enol}/U^{enol}\rightarrow G^{enol}{\bullet}T/U$ 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 $G{\bullet}T^{enol}/U^{enol}$ and $G^{enol}{\bullet}T/U$ 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 $G^{enol}{\bullet}T/U$ 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 $G^{enol}{\bullet}T/U$ population compared to $G{\bullet}T^{enol}/U^{enol}$.

**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 $F_{pol}= (k_{pol}/K_d)_{incorrect}/(k_{pol}/K_d)_{correct} \approx 10^{-4}\text{-}10^{-5}$ 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 $\approx 10^{-1}\text{-}10^{-2}$ 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 $F_{pol}^{dG\bullet 5BrdUTP}$ measured for AMV RT when increasing the pH from 7.0 to 8.4 (Fig. 5c; left). This can be attributed to the lower p$K_a$ of dG•$^{5Br}$dU$^-$ (p$K_a$ ≈9)[14] relative to dG•dT$^-$ (p$K_a$ ≈11.8)[8]. We further verified the robustness of these predictions by measuring $k_{obs}^{dG\bullet 5BrdUTP}$ and $k_{obs}^{dTTP}$ for human DNA polymerase β at high pH (8.4). The model accurately predicts the ≈4-fold enhancement in $k_{obs}^{dG\bullet 5BrdUTP}$ 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 $k_{obs}^{dG\bullet dCTP}$ varied weakly (<1.2-fold) with sequence, $k_{obs}^{dG\bullet dTTP}$ 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 $k_{obs}^{dG\bullet dTTP}$, this is not too surprising considering that other

microscopic steps could also vary with sequence. The predictions do recapitulate the lower $k_{obs}^{dG\bullet dTTP}$ 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 $k_{ex} = k_{GS\rightarrow ES1} + k_{ES1\rightarrow GS}$, 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].

## REFERENCES


1. Watson, J. D. & Crick, F. H. The structure of DNA. *Cold Spring Harb. Symp. Quant. Biol.* **18,** 123–131 (1953).
2. Oertell, K. *et al.* Kinetic selection vs. free energy of DNA base pairing in control of polymerase fidelity. *Proc. Natl Acad. Sci. USA* **113**, E2277–E2285 (2016).
3. Voorhees, R. M. & Ramakrishnan, V. Structural basis of the translational elongation cycle. *Annu. Rev. Biochem.* **82,** 203–236 (2013).
4. Kool, E. T. Active site tightness and substrate fit in DNA replication. *Annu. Rev. Biochem.* **71**, 191-219 (2002).
5. Demeshkina, N., Jenner, L., Westhof, E., Yusupov, M. & Yusupova, G. A new understanding of the decoding principle on the ribosome. *Nature* **484**, 256-259 (2012).
6. Wang, W., Hellinga, H. W. & Beese, L. S. Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. *Proc. Natl Acad. Sci. USA* **108,** 17644–17648 (2011).
7. Bebenek, K., Pedersen, L. C. & Kunkel, T. A. Replication infidelity via a mismatch with Watson-Crick geometry. *Proc. Natl Acad. Sci. USA* **108,** 1862–1867 (2011).
8. Kimsey, I. J., Petzold, K., Sathyamoorthy, B., Stein, Z. W. & Al-Hashimi, H. M. Visualizing transient Watson-Crick-like mispairs in DNA and RNA duplexes. *Nature* **519,** 315–320 (2015).
9. Rozov, A., Demeshkina, N., Westhof, E., Yusupov, M. & Yusupova, G. Structural insights into the translational infidelity mechanism. *Nat. Commun.* **6,** 7521 (2015).
10. Topal, M. D. & Fresco, J. R. Base pairing and fidelity in codon-anticodon interaction. *Nature* **263**, 289-293 (1976).
11. Topal, M. D. & Fresco, J. R. Complementary base pairing and the origin of substitution mutations. *Nature* **263**, 285-289 (1976).
12. Loveland, A. B., Demo, G., Grigorieff, N. & Korostelev, A. A. Ensemble cryo-EM elucidates the mechanism of translation fidelity. *Nature* **546**, 113–117 (2017).
13. Singh, V., Fedeles, B. I. & Essigmann, J. M. Role of tautomerism in RNA biochemistry. *RNA* **21**, 1-13 (2015).
14. Yu, H., Eritja, R., Bloom, L. B. & Goodman, M. F. Ionization of bromouracil and fluorouracil stimulates base mispairing frequencies with guanine. *J. Biol. Chem.* **268,** 15935–15943 (1993).
15. Koag, M. C., Nam, K. & Lee, S. The spontaneous replication error and the mismatch discrimination mechanisms of human DNA polymerase β. *Nucleic Acids Res.* **42,** 11233–11245 (2014).
16. Rozov, A. Demeshkina, N., Westhof, E. Yusupov, M. & Yusupova, G. New Structural Insights into Translational Miscoding. *Trends Biochem. Sci.* **41**, 798-814 (2016).
17. Tomasetti, C., Li. L., & Vogelstein, B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science* **355**, 1330-1334 (2017).
18. Eckert, K. A. & Kunkel, T. A. Effect of reaction pH on the fidelity and processivity of exonuclease-deficient Klenow polymerase. *J. Biol. Chem.* **268,** 13462–13471 (1993).
19. Warren, J. J., Forsberg, L. J. & Beese, L. S. The structural basis for the mutagenicity

of $O^6$-methyl-guanine lesions. *Proc. Natl Acad. Sci. USA* **103,** 19701–19706 (2006).
20. Dosanjh, M. K., Galeros, G., Goodman, M. F. & Singer, B. Kinetics of extension of $O^6$-methylguanine paired with cytosine or thymine in defined oligonucleotide sequences. *Biochemistry* **30,** 11595–11599 (1991).
21. Mendelman, L. V., Boosalis, M. S., Petruska, J. & Goodman, M. F. Nearest neighbor influences on DNA polymerase insertion fidelity. *J. Biol. Chem.* **264,** 14415–14423 (1989).
22. Lee, H. R. & Johnson, K. A. Fidelity of the human mitochondrial DNA polymerase. *J. Biol. Chem.* **281,** 36236–36240 (2006).
23. Xia, S. & Konigsberg, W. H. Mispairs with Watson-Crick base-pair geometry observed in ternary complexes of an RB69 DNA polymerase variant. *Protein Sci.* **23,** 508–513 (2014).
24. Ogle, J. M., Murphy, F. V., Tarry, M. J. & Ramakrishnan, V. Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell* **111,** 721–732 (2002).
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.* **4,** 422–431 (2013).
26. Brovarets', O. O. & Hovorun, D. M. 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.* **33,** 925–945 (2015).
27. Trott, O., Abergel, D. & Palmer, A. G. III. An average-magnetization analysis of $R_{1\rho}$ relaxation outside of the fast exchange limit. *Mol. Phys.* **101,** 753–763 (2003).
28. Korzhnev, D. M., Orekhov, V. Y. & Kay, L. E. 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.* **127,** 713–721 (2005).
29. Hansen, A. L., Nikolova, E. N., Casiano-Negroni, A. & Al-Hashimi, H. M. 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.* **131,** 3818–3819 (2009).
30. Mulder, F. A. A., Mittermaier, A., Hon, B., Dahlquist, F. W. & Kay, L. E. Studying excited states of proteins by NMR spectroscopy. *Nat. Struct. Biol.* **8,** 932–935 (2001).
31. Szymanski, E. S., Kimsey, I. J. & Al-Hashimi, H. M. Direct NMR Evidence that Transient Tautomeric and Anionic States in dG•dT Form Watson-Crick-like Base Pairs. *J. Am. Chem. Soc.* **139**, 4326–4329 (2017).
32. Orozco, M., Hernández, B. & Luque, F. J. 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* **102,** 5228–5233 (1998).
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.* **221,** 1–4 (2012).
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.* **7**, 3062–3613 (2016).

35. Delaney, J. C & Essigmann, J. M. Effects of Sequence Context of $O^6$-methylguanine Repair and Replication in Vivo. *Biochemistry* **40**, 14968–14975 (2001).
36. Trott, O., Palmer, A. G. 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.* **170**, 104–112 (2004).
37. Sekhar, A. *et al.* Thermal fluctuations of immature SOD1 lead to separate folding and misfolding pathways. *Elife* **4,** e07296 (2015).
38. Kunkel, T. A. & Alexander, P. S. The base substitution fidelity of eukaryotic DNA polymerases. *J. Biol. Chem.* **261,** 160–166 (1986).
39. Zahurancik, W. J., Klein, S. J. & Suo, Z. Significant contribution of the 3′→5′ exonuclease activity to the high fidelity of nucleotide incorporation catalyzed by human DNA polymerase ε. *Nucleic Acids Res.* **42,** 13853–13860 (2014).
40. Loeb, L. & Kunkel, T. A. Fidelity of DNA synthesis. *Annu. Rev. Biochem.* **51**, 429–457 (1982).
41. Zhang, J., Ieong, K.-W., Johansson, M. & Ehrenberg, M. Accuracy of initial codon selection by aminoacyl-tRNAs on the mRNA-programmed bacterial ribosome. *Proc. Natl Acad. Sci. USA* **112,** 9602–9607 (2015).
42. Rothwell, P. J. & Waksman, G. Structure and mechanism of DNA polymerases. *Adv. Protein Chem.* **71**, 401–440 (2005).
43. Tsai, Y. C. & Johnson, K. A. A new paradigm for DNA polymerase specificity. *Biochemistry* **45**, 9675–9687 (2006).
44. Patel, S. S., Wong, I. & Johnson, K. A. Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. *Biochemistry* **30**, 511–525 (1991).
45. Zahurancik, W. J., Klein, S. J. & Suo, Z. Kinetic mechanism of DNA polymerization catalyzed by human DNA polymerase ε. *Biochemistry* **52**, 7041–7049 (2013).
46. Balbo, P. B., Wang, E. C. & Tsai, M. D. Kinetic mechanism of active site assembly and chemical catalysis of DNA polymerase β. *Biochemistry* **50**, 9865–9875 (2011).
47. Maximoff, S. N., Kamerlin, S. C. L. & 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* **121**, 7813–7822 (2017).
48. Freudenthal, B. D., Beard, W. A., Shock, D. D. & Wilson, S. H. Observing a DNA polymerase choose right from wrong. *Cell* **154**, 157–168 (2013).
49. Freudenthal, B. D., Beard, W. A., Cuneo, M. J., Dyrkheeva, N. S. & Wilson, S. H. Capturing snapshots of APE1 processing DNA damage. *Nat. Struct. Mol. Biol.* **22**, 924–931 (2015).

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

**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.

**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.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Z.S. (suo.3@osu.edu) or H.M.A. (hashim.al.hashimi@duke.edu).

**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.

**Extended Data legends**

**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).

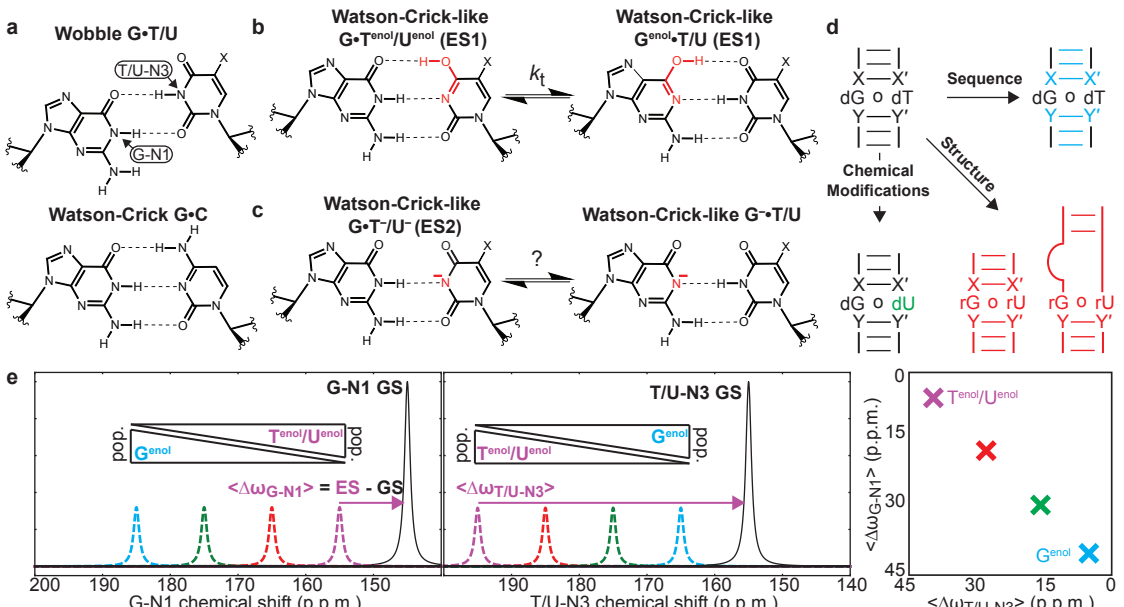
**Extended Data Figure 7 | Kinetic mechanisms used to model misincorporation**. Rate constants for each step are listed in Supplementary Table 7.

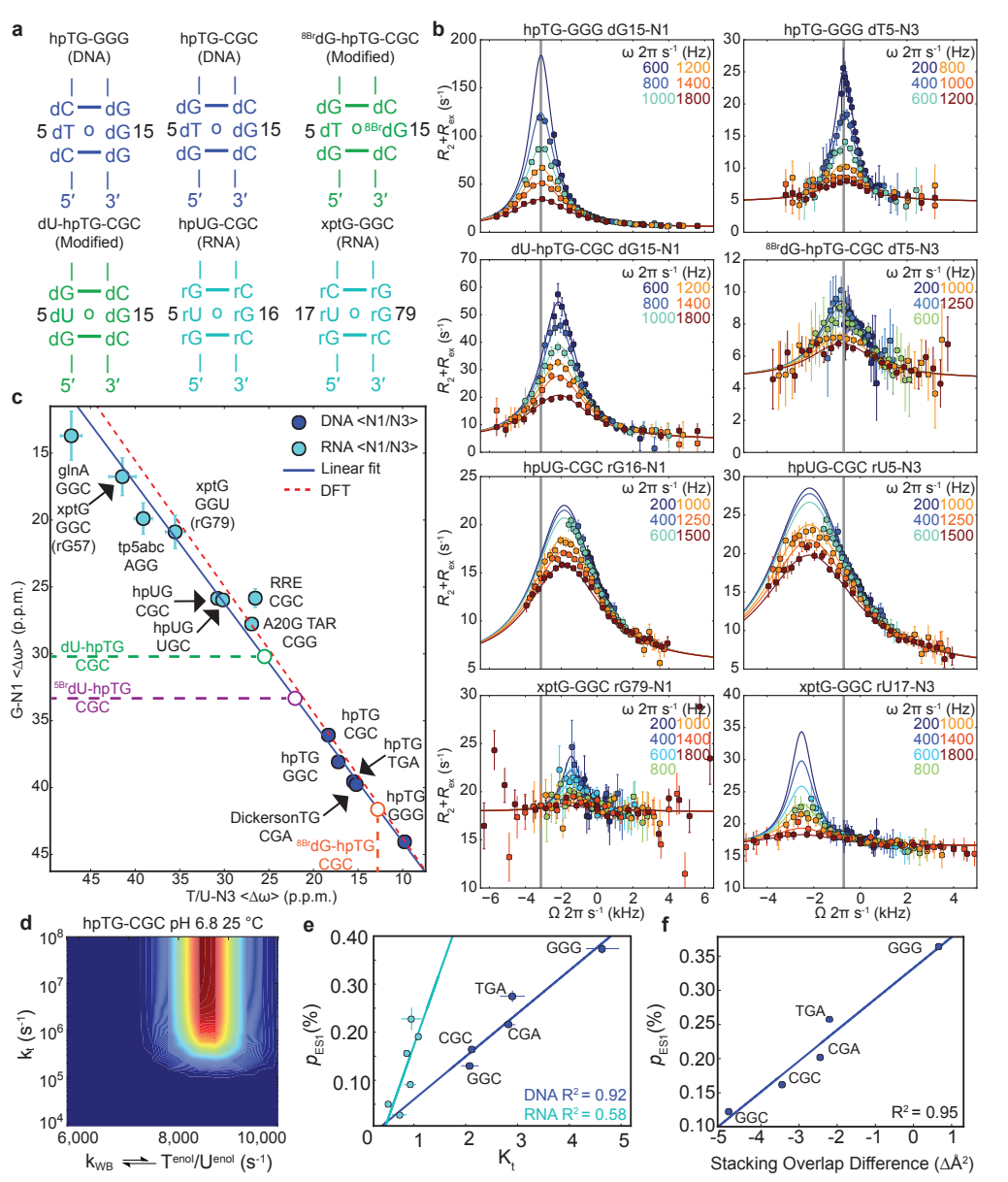
**Extended Data Figure 8 | Benchmarking kinetic simulations of misincorporation a,** Comparison of $k_{pol}$ and $K_d$ values for correct incorporation measured experimentally for human DNA polymerase ε with values computed based on pre-steady state simulations using the microscopic rate constants provided in ref. 39. Error bars reflect fitting uncertainty as previously published[39]. **b,** Robustness of calculated $k_{pol}$ values for human

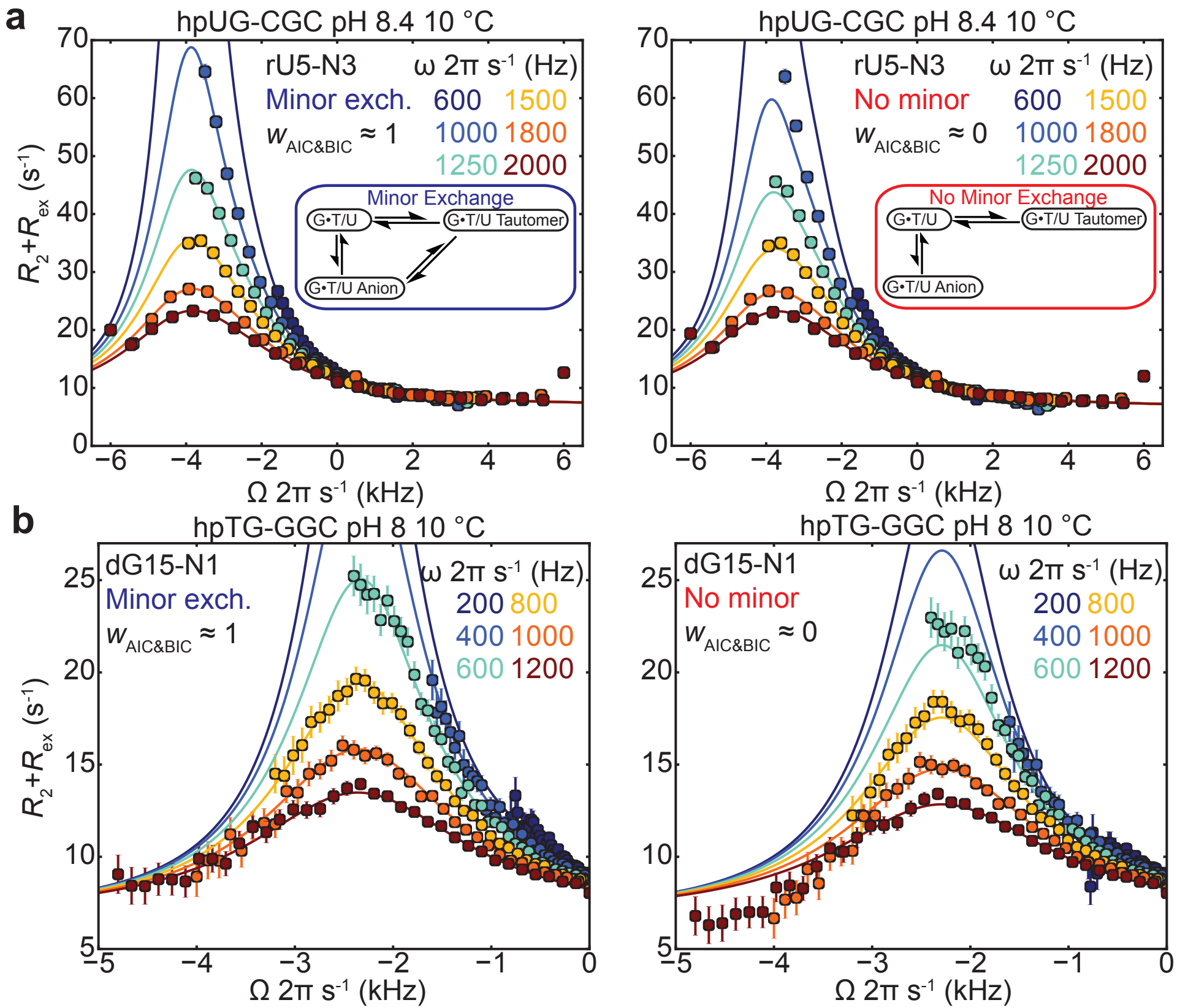
DNA polymerase ε when varying rate constants (forward rate constant: blue; and reverse rate constant: orange) for steps other than tautomerization/ionization by 2-fold ($n$ = 200 independent simulations in which rate constants were randomly varied within 2-fold). As expected, the only rate constant with a substantial effect on the reported $k_{pol}$ values was the rate limiting $k_2$ conformational change step (middle).

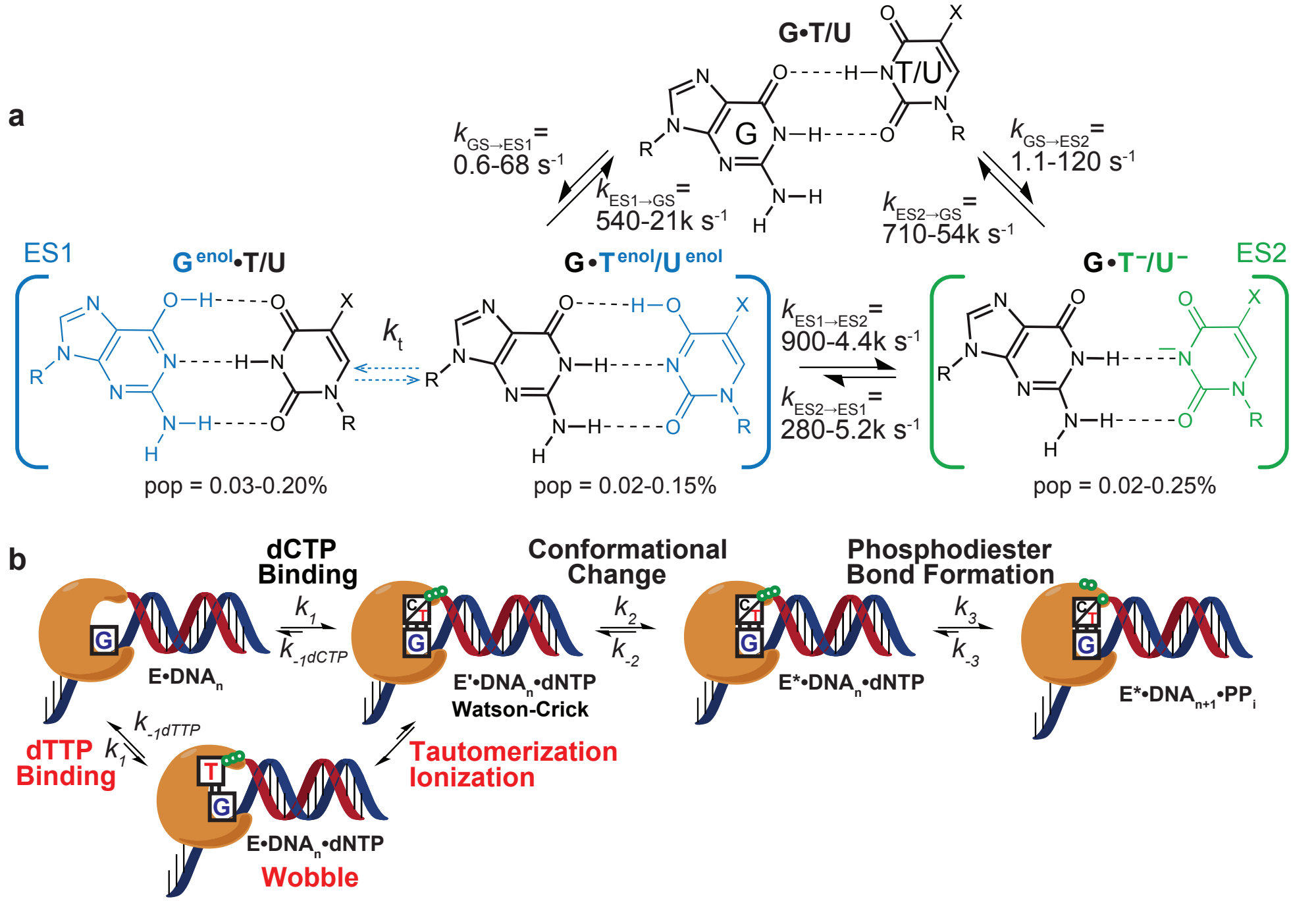
**Extended Data Figure 9 | $k_{obs}$ values measured for human DNA polymerase β dCTP•dG (left) or dTTP•dG (right) insertion at pH 8.4, 25 °C, and 100 μM dNTP.** DNA template sequence (5′ to 3′) is read from bottom (n+1 position) to top (n-1 position). Individual replicates ($n$ = 3 independent experiment) 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 $F_{pol}$.


a
Wobble G•T/U
T/U-N3
G-N1
Watson-Crick G•C
b
Watson-Crick-like
G•Tenol/Uenol (ES1)
kt
Watson-Crick-like
Genol•T/U (ES1)
c
Watson-Crick-like
G•T−/U− (ES2)
?
Watson-Crick-like G−•T/U
d
Sequence
Structure
Chemical
Modifications
dG o dT
dG o dU
rG o rU
e
G-N1 GS
T/U-N3 GS
pop.
dod
Genol
Tenol/Uenol
<ΔωG-N1> = ES - GS
<ΔωT/U-N3>
G-N1 chemical shift (p.p.m.)
T/U-N3 chemical shift (p.p.m.)
<ΔωG-N1> (p.p.m.)
<ΔωT/U-N3> (p.p.m.)


a
hpTG-GGG (DNA)
hpTG-CGC (DNA)
8Brdg-hpTG-CGC (Modified)
dU-hpTG-CGC (Modified)
hpUG-CGC (RNA)
xptG-GGC (RNA)
b
hpTG-GGG dG15-N1
hpTG-GGG dT5-N3
dU-hpTG-CGC dG15-N1
8BrdG-hpTG-CGC dT5-N3
hpUG-CGC rG16-N1
hpUG-CGC rU5-N3
xptG-GGC rG79-N1
xptG-GGC rU17-N3
ω 2π s-1 (Hz)
R2+Rex (s-1)
Ω 2π s-1 (kHz)
c
DNA <N1/N3>
RNA <N1/N3>
Linear fit
DFT
glnA GGC
xptG GGC (rG57)
tp5abc AGG
xptG GGU (rG79)
hpUG CGC
hpUG UGC
RRE CGC
A20G TAR CGG
dU-hpTG CGC
5Brdu-hpTG CGC
hpTG CGC
hpTG GGC
hpTG TGA
hpTG GGG
DickersonTG CGA
8BrdG-hpTG CGC
G-N1 <Δω> (p.p.m.)
T/U-N3 <Δω> (p.p.m.)
d
hpTG-CGC pH 6.8 25 °C
kt (s-1)
kWB ⇌ Tenol/Uenol (s-1)
e
pES1(%)
Kt
DNA R2 = 0.92
RNA R2 = 0.58
f
pES1(%)
Stacking Overlap Difference (ΔÅ2)
R2 = 0.95


a
hpUG-CGC pH 8.4 10 °C
rU5-N3
Minor exch.
$w_{AIC\&BIC} \approx 1$
ω 2π s-1 (Hz)
600 1500
1000 1800
1250 2000
Minor Exchange
G•T/U
G•T/U Tautomer
G•T/U Anion
$R_2+R_{ex}$ (s-1)
Ω 2π s-1 (kHz)
hpUG-CGC pH 8.4 10 °C
rU5-N3
No minor
$w_{AIC\&BIC} \approx 0$
ω 2π s-1 (Hz)
600 1500
1000 1800
1250 2000
No Minor Exchange
G•T/U
G•T/U Tautomer
G•T/U Anion
Ω 2π s-1 (kHz)
b
hpTG-GGC pH 8 10 °C
dG15-N1
Minor exch.
$w_{AIC\&BIC} \approx 1$
ω 2π s-1 (Hz)
200 800
400 1000
600 1200
$R_2+R_{ex}$ (s-1)
Ω 2π s-1 (kHz)
hpTG-GGC pH 8 10 °C
dG15-N1
No minor
$w_{AIC\&BIC} \approx 0$
ω 2π s-1 (Hz)
200 800
400 1000
600 1200
Ω 2π s-1 (kHz)


a
G•T/U
$k_{GS\rightarrow ES1}$ = 0.6-68 s$^{-1}$
$k_{ES1\rightarrow GS}$ = 540-21k s$^{-1}$
$k_{GS\rightarrow ES2}$ = 1.1-120 s$^{-1}$
$k_{ES2\rightarrow GS}$ = 710-54k s$^{-1}$
ES1
G$^{enol}$•T/U
$k_t$
G•T$^{enol}$/U$^{enol}$
$k_{ES1\rightarrow ES2}$ = 900-4.4k s$^{-1}$
$k_{ES2\rightarrow ES1}$ = 280-5.2k s$^{-1}$
G•T$^-$/U$^-$
ES2
pop = 0.03-0.20%
pop = 0.02-0.15%
pop = 0.02-0.25%
b
dCTP Binding
Conformational Change
Phosphodiester Bond Formation
$k_1$
$k_{-1dCTP}$
$k_2$
$k_{-2}$
$k_3$
$k_{-3}$
E•DNA$_n$
E'•DNA$_n$•dNTP
Watson-Crick
E*•DNA$_n$•dNTP
E*•DNA$_{n+1}$•PP$_i$
$k_{-1dTTP}$
$k_1$
dTTP Binding
Tautomerization
Ionization
E•DNA$_n$•dNTP
Wobble

**Supplementary Information for**

# 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]**

## TABLE OF CONTENTS

1) Supplementary Discussion
   - Tilting tautomeric equilibrium using chemical modifications
2) Methods
   - Enzymatic synthesis of uniformly $^{13}C/^{15}N$-labeled RNA samples
   - Enzymatic synthesis of selectively $^{13}C/^{15}N$-labeled DNA samples
   - Choice of DNA duplexes
   - NMR sample conditions
   - NMR resonance assignments.
   - $R_{1\rho}$ measurements.
   - Fitting RD profiles using the B-M equations.
   - Model comparison and model selection.
   - Expressions for chemical shifts and tautomeric equilibria used to interpret equilibrium tilting experiments.
   - Resolving the individual tautomeric by constrained liner 3-state B-M fits.
   - Calculating $K_t$.
   - Defining the lower limits of rapid tautomeric exchange.
   - Calculating stacking overlap.
   - Kinetic modeling.
   - Kinetic Assays
   - Code Availability
   - Data Availability
3) Additional References
4) Supplemental Table Legends
   - Supplemental Table Legends 1-9

## SUPPLEMENTARY DISCUSSION

### Tilting tautomeric equilibrium using chemical modifications.

We exploited bias for either enolic T/U or enolic G and introduced chemical modifications that are designed to specifically tilt the tautomeric equilibrium in DNA (Fig. 1b). We omitted the thymidine methyl group, an electron donating moiety, and indeed observed that dG•dU tilts the equilibrium in favor of dG•dT$^{enol}$ (Fig. 2c and Supplementary Table 1). In contrast, replacement of the methyl group with a moderately electron-withdrawing bromine tilted the equilibrium to a lesser degree in favor of dG•$^{5Br}$dU$^{enol}$ (Fig. 2c and Supplementary Table 1). These results are in general agreement with computational studies showing a preference for forming the T$^{enol}$/U$^{enol}$ tautomer in 1-methyl nucleobases follows the general order: U$^{enol}$ > $^{5Br}$U$^{enol}$ > T$^{enol}$[32]. Conversely, replacing the 2′-deoxyguanosine H8 with bromine tilted the equilibrium more in favor of dG$^{enol}$•dT (Fig. 2c and Supplementary Table 1). While the RNA data are less biased towards rG$^{enol}$•rU, it also features slightly larger variations in the $<\Delta\omega_{ES1}>$ values; likely due to the much more varied structural contexts (Fig. 2c). Taken together, these results establish the existence of two statistically distinct[50] G$^{enol}$•T/U and G•T$^{enol}$/U$^{enol}$ species (Supplementary Table 2) in ultra-fast chemical exchange within the nucleic acid helical environment.

## METHODS

**Enzymatic synthesis of uniformly $^{13}C/^{15}N$-labeled DNA and RNA samples.** The uniformly $^{13}C/^{15}N$-labeled "DickersonTG-CGA" DNA sample was synthesized using the Crothers-Zimmerman primer extension procedure as described previously[8,51] using uniformly $^{13}C/^{15}N$-labeled deoxynucleotide triphosphates (Cambridge Isotope Labs). The sample was purified using polyacrylamide gel electrophoresis (PAGE) (20% polyacrylamide in 1X Tris/Borate/EDTA buffer) and gel electroelution (Whatman, GE Healthcare). RNA samples were prepared using *in vitro* transcription using uniformly $^{13}C/^{15}N$-labeled ribonucleotide triphosphates (Silantes GmbH) as described previously[52,53]. Samples were purified using PAGE (10-20% polyacrylamide in 1X Tris/Borate/EDTA buffer) and gel electroelution (Whatman, GE Healthcare).

**Chemical synthesis of selectively $^{13}C/^{15}N$-labeled DNA samples.** DNA samples containing residue-specific $^{13}C/^{15}N$-labeled nucleotides at the dG•dT mismatch positions (hpTG-CGC, hpTG-GGC, hpTG-GGG, hpTG-TGA, $^{5Br}$dU5-hpTG-CGC, dU5-hpTG-CGC, and $^{8Br}$dG15-hpTG-CGC) were purchased from the Yale Keck Oligonucleotide Synthesis Facility. The $^{13}C/^{15}N$-selectively labeled phosphoramidites were purchased from Cambridge Isotope Labs (98% $^{13}C/^{15}N$ 2′-deoxyguanosine DMT-phosphoramidites and 98% $^{13}C/^{15}N$ 2′-deoxythymidine phosphoramidites). All synthesized oligos were purified using RP-HPLC or GlenPak cartridges (Glen Research). The purity of each sample was assessed and confirmed by HPLC, mass spectrometry, and NMR. NMR experiments were also used to confirm proper duplex formation and folding.

**Choice of DNA duplexes**. The DNA duplexes used in this study (Extended Data Fig. 2a) were selected based on the following criteria: hpTG-CGC was the focus of our previous study[8] and was the subject of NMR studies in unmodified and modified form[54]. hpTG-GGC and hpTG-GGG were selected to study the impact of swapping one or both neighboring pyrimidines into purines. Based on melting data performed on a number of duplexes containing dG•dT wobbles (data not shown), hpTG-TGA was selected because it was the least stable. In contrast hpTG-CGC was the second most stable sequence context. DickersonTG-CGA is the Dickerson dodecamer sequence context, and it was selected because it has been studied previously by X-ray crystallography[55] with dG•dT mismatches. It also allows us to assess the impact of swapping 5′ G > A on the central mismatch.

**NMR sample conditions.** All DNA and RNA samples were exchanged into their respective buffers (listed in Supplementary Table 1) using a centrifugal concentrator (EMD Millipore). The exchange was repeated to ensure >99.5% exchange into the target buffer. When the desired pH was >8.0, the sample pH was directly adjusted using stock solutions of 50 mM NaOH and/or 50 mM HCl. All buffers contained 0.1 mM EDTA except the glnA and xptG riboswitch samples, which contained 5 and 10 mM magnesium, respectively, and which was directly added to the samples to the desired concentration. $D_2O$ (≈5-10%) was added to all NMR samples. The DNA and RNA concentration in the NMR samples ranged between ≈0.5 and 3.5 mM.

**NMR resonance assignments.** The H1/H3 and N1/N3 resonance assignments for all DNA and RNA constructs were obtained using imino [$^{15}$N, $^{1}$H] heteronuclear and [$^{1}$H, $^{1}$H]

NOESY homonuclear correlation experiments as previously described[8]. Assignment experiments were collected on 600 and 700 MHz Bruker Avance spectrometers equipped with HCNP and HCN cryogenic probes, respectively, and an 800 MHz Agilent spectrometer equipped with a HCN cryogenic probe. Data was processed using NMRpipe software package[56] and analyzed using SPARKY (T.D. Goddard and D.G. Kneller, SPARKY 3, University of California, San Francisco).

***$R_{1\rho}$* measurements.** All RD data were measured on 600 and 700 MHz Bruker Avance spectrometers equipped with HCNP and HCN cryogenic probes, respectively, using the 1D $^{15}$N $R_{1\rho}$ experiment[29] as described previously[8,53]. Raw data were processed as a pseudo-2D assuming a Lorentzian peak-shape using NMRpipe[56] to generate a series of time-dependent peak intensities. Data was collected at varying spinlock powers ($\omega$ $2\pi^{-1}$ Hz) and offset frequencies ($\Omega$ $2\pi^{-1}$ Hz, where $\Omega = \Omega_{obs} - \omega_{rf}$ and $\omega_{rf}$ is the carrier frequency) given in the respective RD profiles (Fig. 2b and 3a-c, Extended Data Fig. 3 and 5). Representative spin lock power and offsets are provided in Supplementary Table 9. The spins of interest were allowed to relax under an applied spinlock for the varying durations, ranging from 0 – 120 ms. $R_{1\rho}$ values for a given spinlock power and offset combination were calculated by fitting the decaying peak intensities to a monoexponential decay[8,53]. The $R_{1\rho}$ uncertainty was calculated using a Monte-Carlo approach (500 iterations) using the root-mean-square spectral noise to estimate uncertainty in intensity values as described previously[57].

**Fitting RD profiles using the B-M equations.** The chemical exchange parameters of interest were obtained by fitting the experimentally measured $^{15}$N $R_{1\rho}$ values to numerical

solutions of the Bloch–McConnell (B-M) equations[28,36] describing *n*-site chemical exchange. Briefly, for a given set of chemical exchange parameters and combinations of $\Omega$ and $\omega_{SL}$, $R_{1\rho}$ can be simulated by solving the BM equations[28,57]. Exchange parameters were obtained by finding values that best reproduce the experimentally measured $R_{1\rho}$ and their associated uncertainties. The residual sum of squares of the experimental and simulated $R_{1\rho}$ data were minimized using a bounded[58] least-squares algorithm to give the exchange parameters that best-fit the experimental data. The uncertainty in the fitted chemical exchange parameters were computed based on the standard error of the fit, or by a Monte-Carlo approach[57] in the case of the rapid tautomer B-M fits shown in Supplementary Table 2. For all RD data reported, fits were carried out assuming equivalent intrinsic longitudinal ($R_1$) and transverse ($R_2$) relaxation rates for the GS and ES. Here, $R_{1(GS)}=R_{1(ES1)}=R_{1(ES2)}$ and $R_{2(GS)}=R_{2(ES1)}=R_{2(ES2)}$[27,36,59], which is a good approximation for low-populated ES ($<\approx 0.5\%$).

When available, both the RD data measured in G-N1 and T/U-N3 of the same mismatch under identical conditions were globally fitted, sharing the ES populations and exchange rates ($k_{GS\rightarrow ES1}$, $k_{GS\rightarrow ES2}$, $k_{ES1\rightarrow ES2}$, $k_{ES1\rightarrow GS}$, $k_{ES2\rightarrow GS}$, $k_{ES2\rightarrow ES1}$, $k_{T^{enol}\rightarrow G^{enol}}$, and/or $k_{G^{enol}\rightarrow T^{enol}}$) as described previously[8]. Very similar chemical exchange parameters were generally obtained from individual versus global fits. We also note that the linear correlation shown in Fig. 2c is also observed when individually fitting the G-N1 and T/U-N3 RD data, giving identical results within error (Supplementary Table 3). Individual fits were reported in cases where only one base in the G•T/U mismatch was $^{15}$N-labeled (dG in $^{5Br}$dU5-hpTG-CGC and dU5-hpTG-CGC; dT in $^{8Br}$dG15-hpTG-CGC; Supplementary Table 1).

**Model comparison and model selection.** Model selection (2-state or 3-state with four distinct topologies, Extended Data Fig. 6) in fitting the $R_{1\rho}$ data was carried out by calculating the Akaike's ($w_{AIC}$) and Bayesian information criterion ($w_{BIC}$) weights[50,60] for each model and selecting the model with the highest relative probability. AIC and BIC weights shown in Fig. 3a,b and Extended Data Fig. 5 reflect comparison between starlike and triangular topologies. The DNA and RNA mismatches that did not call for a triangular topology (Supplementary Table 5) were best fit with the 3-state starlike topology (Extended Data Figs. 5 and 6d, and Supplementary Tables 5, 6).

**Expressions for chemical shifts and tautomeric equilibria used to interpret equilibrium tilting experiments.** The $<\Delta\omega_{G\text{-}N1}>$ and $<\Delta\omega_{T/U\text{-}N3}>$ values measured for ES1 at near neutral pH (Supplementary Table 1) from two state fitting of the RD data are given by,

$$<\Delta\omega_{G\text{-}N1}> = <\omega_{G\text{-}N1(ES1)}> - \omega_{G\text{-}N1(GS)}$$

$$<\Delta\omega_{T/U\text{-}N3}> = <\omega_{T/U\text{-}N3(ES1)}> - \omega_{T/U\text{-}N3(GS)}$$

Where $\omega_{(ES1)}$ and $\omega_{(GS)}$ represent the $^{15}N$ chemical shifts of the ES1 and GS, respectively. The angular bracket denotes a population ($p$) weighted average over the $G^{enol}$•T/U and G•$T^{enol}$/$U^{enol}$ species. Below we use $f_{Genol}$ and $f_{Tenol/Uenol}$ to denote the relative populations of $G^{enol}$ and $T^{enol}$/$U^{enol}$, whose sum is equal to 1.

$$<\Delta\omega_{G\text{-}N1}> = f_{Genol} \times \Delta\omega_{Genol\text{-}N1} + f_{Tenol/Uenol} \times \Delta\omega_{Tenol/Uenol\text{-}N1}$$

$$= f_{Genol} \times \Delta\omega_{Genol\text{-}N1} + (1\text{-}f_{Genol}) \times \Delta\omega_{Tenol/Uenol\text{-}N1}$$

$$<\Delta\omega_{T/U\text{-}N3}> = f_{Genol} \times \Delta\omega_{Genol\text{-}N3} + f_{Tenol/Uenol} \times \Delta\omega_{Tenol/Uenol\text{-}N3}$$

$$= f_{Genol} \times \Delta\omega_{Genol\text{-}N3} + (1\text{-}f_{Genol}) \times \Delta\omega_{Tenol/Uenol\text{-}N3}$$

in which:

$$\Delta\omega_{Genol\text{-}N1} = \omega_{Genol\text{-}N1(ES1)} - \omega_{G\text{-}N1(GS)}$$

$$\Delta\omega_{Tenol/Uenol\text{-}N1} = \omega_{Tenol/Uenol\text{-}N1(ES1)} - \omega_{G\text{-}N1(GS)}$$

$$\Delta\omega_{Genol\text{-}N3} = \omega_{Genol\text{-}N3(ES1)} - \omega_{G\text{-}N3(GS)}$$

$$\Delta\omega_{Tenol/Uenol\text{-}N3} = \omega_{Tenol/Uenol\text{-}N3(ES1)} - \omega_{T/U\text{-}N3(GS)}$$

$\omega_{Genol\text{-}N1(ES1)}$ and $\omega_{Tenol/Uenol\text{-}N1(ES1)}$ are the G-N1 chemical shifts in the $G^{enol}$•T/U and G•$T^{enol}$/$U^{enol}$ tautomeric species, respectively. $\omega_{Genol\text{-}N3(ES1)}$ and $\omega_{Tenol/Uenol\text{-}N3(ES1)}$ are the T/U-N3 chemical shifts in the $G^{enol}$•T/U and G•$T^{enol}$/$U^{enol}$ tautomeric species, respectively. For simplicity we drop the "ES1" and "GS" arguments below.

Substituting $f_{Genol} = (<\Delta\omega_{T/U\text{-}N3}>-\Delta\omega_{Tenol/Uenol\text{-}N3})\ (\Delta\omega_{Genol\text{-}N3} - \Delta\omega_{Tenol/Uenol\text{-}N3})^{-1}$ into the expression for $<\Delta\omega_{G\text{-}N1}>$ yields the following relationship between $<\Delta\omega_{G\text{-}N1}>$ and $<\Delta\omega_{T/U\text{-}N3}>$,

$$\langle\Delta\omega_{G\text{-}N1}\rangle = \langle\Delta\omega_{U\text{-}N3}\rangle \frac{\Delta\omega_{Genol\text{-}N1} - \Delta\omega_{Uenol\text{-}N1}}{\Delta\omega_{Genol\text{-}N3} - \Delta\omega_{Uenol\text{-}N3}} + \frac{\Delta\omega_{Genol\text{-}N3} \times \Delta\omega_{Uenol\text{-}N1} - \Delta\omega_{Genol\text{-}N1} \times \Delta\omega_{Uenol\text{-}N3}}{\Delta\omega_{Genol\text{-}N3} - \Delta\omega_{Uenol\text{-}N3}}$$

Eq. 1

Equivalent expressions for DNA are obtained by replacement of "U" with "T". The above equation shows that a plot of $<\Delta\omega_{G\text{-}N1}>$ versus $<\Delta\omega_{T/U\text{-}N3}>$ is linear with slope and intercept determined by the four fundamental chemical shifts of the tautomeric species: $\Delta\omega_{Genol\text{-}N1}$, $\Delta\omega_{Genol\text{-}N3}$, $\Delta\omega_{Tenol/Uenol\text{-}N3}$, and $\Delta\omega_{Tenol/Uenol\text{-}N1}$.

The measured $<\Delta\omega_{G\text{-}N1}>$ and $<\Delta\omega_{T/U\text{-}N3}>$ values were fitted to the above expression under the assumption that the chemical shifts for the enolic bases ($\Delta\omega_{Genol\text{-}N1}$ and $\Delta\omega_{Tenol/Uenol\text{-}N3}$) range between -10 to +70 ppm while the corresponding chemical shift for paired bases ($\Delta\omega_{Genol\text{-}N3}$ and $\Delta\omega_{T/Uenol\text{-}N1}$) range between -10 to +10 ppm. These ranges are plausible given that deprotonation of G-N1 and T/U-N3 is known to give rise to large downfield chemical shifts[8], whereas small changes in H-bonding (as seen for $\Delta\omega_{Genol\text{-}N3}$ and $\Delta\omega_{Tenol/Uenol\text{-}N1}$) and stacking are not expected to give rise to deviations beyond that observed in the *BioMagResBank*[61]. This analysis also assumes that the underlying basis tautomer chemical shifts ($\Delta\omega_{Genol\text{-}N1}$; $\Delta\omega_{Genol\text{-}N3}$; $\Delta\omega_{Tenol/Uenol\text{-}N3}$; and $\Delta\omega_{Tenol/Uenol\text{-}N1}$; Supplementary Table 3) are, within error, the same for various mismatches in DNA and RNA and that sequence-specific variations are negligible because they are likely comparable to those sequence specific variations seen in the GS (≈2.5/4 p.p.m. for G-N1 and T/U-N3, respectively). We carried out a grid search of over all possible combinations of $\Delta\omega_{Genol\text{-}N1}$, $\Delta\omega_{Genol\text{-}N3}$, $\Delta\omega_{Tenol/Uenol\text{-}N3}$, and $\Delta\omega_{Tenol/Uenol\text{-}N1}$ in 1 p.p.m. increments over their respective ranges (2,893,401 possible combinations). For each combination we used orthogonal distance regression[62] to calculate the residual variance between the experimental $<\Delta\omega_{G\text{-}N1}>$ and $<\Delta\omega_{T/U\text{-}N3}>$ chemical shifts, and their respective uncertainties (Supplementary Table 1), and those calculated using equation 1. The top 1% of these

combinations with the lowest residual variance were selected. The values of $\Delta\omega_{\text{Genol-N1}}$, $\Delta\omega_{\text{Genol-N3}}$, $\Delta\omega_{\text{Tenol/Uenol-N3}}$, and $\Delta\omega_{\text{Tenol/Uenol-N1}}$ were given by the mean of their respective distributions. The errors in each case were given by the standard deviation of the distributions. The deduced tautomeric $\Delta\omega_{\text{Genol-N1}}$, $\Delta\omega_{\text{Genol-N3}}$, $\Delta\omega_{\text{Tenol/Uenol-N3}}$, and $\Delta\omega_{\text{Tenol/Uenol-N1}}$ chemical shifts and the measured values of $<\Delta\omega_{\text{G-N1}}>$ and $<\Delta\omega_{\text{T/U-N3}}>$ were then used to solve for $p_{\text{Genol}}$ for each mismatch, from which an initial estimate of $K_{\text{taut}} = p_{\text{Genol}} / (1-p_{\text{Genol}})$ was calculated and used in B-M fitting simulations described below.

**Resolving the individual tautomers by constrained linear 3-state B-M fits.** We also carried out a 3-state fit assuming a linear wobble⇌G•T$^{\text{enol}}$/U$^{\text{enol}}$⇌G$^{\text{enol}}$•T/U pathway[25] to the RD data measured at low pH for constructs listed in Supplementary Table 1 using the B-M equations. The purpose of this analysis was to (i) verify that the tautomeric chemical shifts ($\Delta\omega_{\text{Genol-N1}}$; $\Delta\omega_{\text{Genol-N3}}$; $\Delta\omega_{\text{Tenol/Uenol-N3}}$; $\Delta\omega_{\text{Tenol/Uenol-N1}}$) and $K_{\text{taut}}$ values obtained from linear analysis of the $<\Delta\omega_{\text{G-N1}}>$ and $<\Delta\omega_{\text{T/U-N3}}>$ values are consistent with the measured RD data, (ii) allow for further optimization of the mismatch specific $K_{\text{taut}}$ values, (iii) to relax the assumption of a single global set of tautomeric chemical shifts and allow for variations across different mismatches, and (iv) determine rate constants for a linear 3-state model supported by computational methods[8,25] in which the wobble initially transitions into G•T$^{\text{enol}}$/U$^{\text{enol}}$ ($k_{\text{Tenol/Uenol}}$, described by rate constants $k_{GS\rightarrow ES1*}$ and $k_{ES1\rightarrow GS*}$) followed by rapid tautomeric exchange (described by rate constants $k_{\text{t}}$ and $k_{\text{-t}}$), i.e. wobble⇌G•T$^{\text{enol}}$/U$^{\text{enol}}$⇌ G$^{\text{enol}}$•T/U. In this analysis, initial values for the tautomeric chemical shifts were equal to those deduced from the linear plot $<\Delta\omega_{\text{G-N1}}>$ versus $<\Delta\omega_{\text{T/U-N3}}>$ (Fig. 2c) and were allowed to float by an amount determined by the uncertainty in the

fitted values (Supplementary Table 3). The initial values for $G^{enol}$•T/U and G•$T^{enol}$/$U^{enol}$ were obtained using the $K_{taut}$ value obtained in the linear fit (see section above) and the ES1 population ($p_{ES1}$) deduced from a 2-state fit to the RD data (Supplementary Table 1). The initial exchange rates were those given in Supplementary Table 1. The initial value for $k_{taut}$ was set to 1,000,000 $s^{-1}$ but was allowed to vary from 1-15,000,000 $s^{-1}$. All parameters, except $\Delta\omega_{Genol\text{-}N1}$, $\Delta\omega_{Genol\text{-}N3}$, $\Delta\omega_{Tenol/Uenol\text{-}N3}$, and $\Delta\omega_{Tenol/Uenol\text{-}N1}$, were allowed to float within a large range of parameter space ($p_{ES1}$ and $p_{ES2}$ from 0-0.5, $k_{GS\rightleftharpoons T^{enol}/U^{enol}}$ from 1-1,000,000 $s^{-1}$, $R_1$ from $10^{-6}$-20 $s^{-1}$ and $R_2$ from $10^{-6}$-200 $s^{-1}$. While the fits converged to a singular $k_{taut}$ value, we can only assign a lower bound to the rapid tautomeric exchange (see below) owing to experimental limitations that result in equally good fits for rates of $k_{taut}$ >≈500,000-1,000,000 $s^{-1}$ (Extended Data Fig. 4).

**Calculating $K_t$.** In cases where we were unable to fit for the individual tautomeric probabilities owing to additional anionic exchange, $K_t$ was calculated by directly solving for $p_{Genol}$ and $p_{Tenol/Uenol}$ using the back-calculated fundamental tautomeric chemical shifts (Supplementary Table 3) and the fitted $\langle\Delta\omega_{G\text{-}N1}\rangle$ and $\langle\Delta\omega_{T/U\text{-}N3}\rangle$ chemical shifts.

$$p_{Genol} = p_{ES1}\,[(\langle\Delta\omega_{G\text{-}N1}\rangle - \Delta\omega_{Tenol/Uenol\text{-}N1})\,(\Delta\omega_{Genol\text{-}N1} - \Delta\omega_{Tenol/Uenol\text{-}N1})^{-1}]$$

$$p_{Tenol/Uenol} = p_{ES1}\,[(\langle\Delta\omega_{T/U\text{-}N3}\rangle - \Delta\omega_{Genol\text{-}N3})\,(\Delta\omega_{Tenol/Uenol\text{-}N3} - \Delta\omega_{Genol\text{-}N3})^{-1}]$$

To further verify the consistency of these values with the RD data, and also refine the mismatch specific exchange parameters, we used these parameters and their uncertainties as initial guesses to re-fit the RD profiles to a 3-state model with a specific linear topology (wobble⇌G•$T^{enol}$/$U^{enol}$⇌$G^{enol}$•T/U) that is supported by computational modeling. A

satisfactory fit was obtained in all cases (Supplementary Table 4) yielding $p_{\mathrm{Genol}}$, $p_{\mathrm{Tenol/Uenol}}$, and $K_t$ values similar to the initial estimates.

**Defining the lower limits of rapid tautomeric exchange.** Owing to experimental limitations, it was not possible to uniquely determine $k_t$ for each fitted resonance (Supplementary Table 2). Instead, we were able to determine the lower bounds for $k_t$ by performing a grid search[57] over $k_{\mathrm{Tenol/Uenol}}$ versus $k_t$ parameter space while all other parameters were held constant to those values reported in Supplementary Table 2. $k_{\mathrm{Tenol/Uenol}}$ was iterated near the B-M fitted value reported (Supplementary Table 3) in 31 logarithmically spaced increments. Simultaneously, $k_t$ was iterated from $10^3$ to $10^8$ $s^{-1}$ in 31 logarithmically spaced increments. For each of the 961 combinations of $k_{\mathrm{Tenol/Uenol}}$ and $k_t$ we calculated $\overline{\chi}^2$ for the simulated versus known $R_{1\rho}$ values. The combinations of $k_{\mathrm{Tenol/Uenol}}$ versus $k_t$ were each given a weight, calculated as:

$$Z_i(\overline{\chi}^2) = \frac{e^{-0.5\cdot\Delta\overline{\chi}_i^2}}{\sum_{k=1}^{K} e^{-0.5\cdot\Delta\overline{\chi}_k^2}} \qquad \text{Eq. 2}$$

Where $\Delta\overline{\chi}_i^2 = \overline{\chi}_i^2 - min\,\overline{\chi}^2$. Here, the largest Z value reflects the most probable fit compared to all other possible combinations. Values are then scaled from 0 to 1, with 1 (red) the most and 0 (blue) the least probable solution (Fig. 2d and Extended Data Fig. 4).

**Calculating stacking overlap.** Stacking overlap (measured as the $Å^2$ geometric overlap) between dG•dT in wobble or Watson-Crick-like conformation with 5′ and 3′ neighboring Watson-Crick base pairs was computed for the five DNA duplexes analyzed in this study (Extended Data Fig. 2a) using 3DNA[63]. We report (Fig. 2e) the gain or loss in stacking

overlap in transitioning from a wobble dG•dT mismatch to a WC-like dG•dT ($\Delta Å^2 = Å^{2(WC)} - Å^{2(WB)}$). To calculate this, we first obtain DNA duplexes containing wobble dG•dT mismatches from the Protein Databank (PDB)[64]. An in-house program was used to identify high-resolution X-ray structures with ≤3Å resolution for DNA duplexes containing wobble dG•dT mismatches in different sequence contexts. The $Å^2$ geometric overlap was calculated using 3DNA and averaged across duplicate sequences. The average for each wobble sequence context was then saved. The corresponding WC-like dG•dT mismatch duplexes were built using make-NA (J. Stroud, make-NA, http://structure.usc.edu/make-na/server.html; 2011). The data shown in Fig. 2f was calculated when excluding exocyclic groups from the geometric overlap calculation; however similar results were obtained when including the exocyclic groups.

**Kinetic modeling.**

*Implementation.*

Simulations were performed to model kinetic mechanisms (Fig. 4b and Extended Data Fig. 7) using the Python programming language to solve for the network of differential equations describing the time- and [dNTP]-dependent species populations for correct and incorrect incorporation events. Input for the simulations included the microscopic rate constants (Supplementary Table 7) for nucleotide binding and dissociation ($k_1$ and $k_{-1}$); the forward and reverse conformational changes ($k_2$ and $k_{-2}$); phosphodiester bond formation and breakage ($k_3$ and $k_{-3}$); and NMR-derived rate constants describing the triangular tautomer/anion network ($k_{GS\rightarrow ES1}$, $k_{GS\rightarrow ES2}$, $k_{ES1\rightarrow GS}$, $k_{ES2\rightarrow GS}$, $k_{ES1\rightarrow ES2}$, $k_{ES2\rightarrow ES1}$). Rate constants for conformational change ($k_2$ and $k_{-2}$) were assumed to be the same for

incorporation of ES1 and ES2. Rate constants for phosphodiester bond formation for dG•dT were assumed to be identical to that of dG•dC. The microscopic rate constants for correct dG•dC incorporation were obtained from prior studies[39,43,46] for a variety of polymerases and pH/temperature conditions and are listed in Supplementary Table 7.

The kinetic simulations were performed in a manner analogous to a benchtop pre-steady-state kinetic experiment, where the kinetic parameters for correct and incorrect incorporation are determined by two separate assays rather than by direct competition. The incorporation time was varied at multiple dNTP concentrations in order to derive $k_{obs}$, $k_{pol}$, and $K_d$ values for a given incorporation scheme. Time points for correct incorporation were: 0.001, 0.005, 0.01, 0.05, 0.1, 0.1, 0.2, 0.3, 0.5, and 1 second. Time points for incorrect incorporation were: 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 10.0, 15.0, 30.0, and 60.0 seconds. Time points were 'collected' by recording the product species population at each time point. dNTP-concentrations were varied by multiplying the [dNTP] value by the estimated dNTP association rate constant ($k_1$ = 100 $s^{-1}$ $\mu M^{-1}$), which is assumed to be limited by diffusion. Free dNTP concentration was assumed to be invariant over the course of the simulation. dNTP concentrations used for correct incorporation were: 0.625, 1.25, 2.5, 5, 10, 20, 40, 80, 200 μM; and for incorrect were: 50, 100 200, 300, 500, 750, 1000, and 1500 μM. It is assumed that the [E•DNA] binary complex is pre-formed for both correct and incorrect incorporation models. The time dependence of product ($DNA_{n+1}$) concentration was recorded at multiple time points for each dNTP concentration and fit to,

$$[product, t] = A(1 - e^{-k_{obs}t})$$

where [product, t] is the concentration of the product ($DNA_{+1}$) species at time = $t$, and $k_{obs}$ is the dNTP-concentration dependent rate of polymerization and A is a pre-exponential factor. The dependence of $k_{obs}$ on [dNTP] was fit to,

$$k_{obs} = \frac{k_{pol} \times [dNTP]}{K_d + [dNTP]}$$

where $k_{pol}$ is the maximum nucleotide incorporation rate constant. $K_d$ is the apparent equilibrium dissociation constant for the dNTP, which does not reflect the true nucleotide equilibrium dissociation constant and should be regarded as a $K_m$ value[65]. Calculations were done for both correct and incorrect incorporation pathways and used to compute the misincorporation frequency ($F_{pol}$):

$$F_{pol} = \frac{\left(\frac{k_{pol}}{K_d}\right)_{incorrect}}{\left(\frac{k_{pol}}{K_d}\right)_{correct}}$$

For simplicity, all kinetic simulations treat the very rapid equilibrium tautomeric $dG{\bullet}dT^{enol} \rightleftharpoons dG^{enol}{\bullet}dT$ equilibrium as a single aggregate ES1 species ($dG^{enol}{\bullet}dT + dG{\bullet}dT^{enol}$). However, similar results were obtained when explicitly treating the two tautomeric species using the NMR-derived $K_t$ and $k_t$ assuming the computationally supported linear topology (wobble$\rightleftharpoons dG{\bullet}dT^{enol} \rightleftharpoons dG^{enol}{\bullet}dT$; data not shown).

*Microscopic rate constants used in kinetic simulations.*

Kinetic simulations reported in Fig. 5a utilize the microscopic rate constants reported for correct incorporation of a dG•dCTP base pair by human DNA polymerase ε (left), rat DNA polymerase β (middle), and T7 DNA polymerase (right)[39,43-46]. $F_{pol}$ values ($M_{ES1+ES2}$, $M_{ES1}$, $M_{ES2}$) in Fig. 5c (left) are calculated with the Pol ε mechanisms using GGC sequence data, and similar trends are observed with other models and sequences. $F_{pol}$ models (M) were compared with pH-titration data for AMV RT[14], however lack of a comprehensive description of AMV RT precluded direct simulation of an AMV RT mechanism. Therefore, $k_{obs}$ values in Fig. 5c,d were calculated using the Polymerase β mechanism, to directly compare with experimental results; when necessary ES2 populations and kinetics were extrapolated to match experimental conditions (See Methods). Finally, the model was independently verified using KinTek Explorer (Kintek)[66].

*Exchange parameters for kinetic simulations.*

Unless stated otherwise, exchange parameters for the CGC sequence deduced by NMR RD were used in the kinetic simulations given the availability of robust RD data at four different pH conditions. Similar results were obtained when using other sequence contexts (data not shown). In some cases, the populations and rate constants were interpolated to match the polymerase experimental conditions based on the linear dependence of ln($k$) (where $k$ is a rate constant) versus temperature or pH. This was done for the data shown in Fig. 5a. In general, similar results (within <2-fold) were obtained when using exchange parameters measured under conditions that most resemble the polymerase experimental conditions without interpolation. In addition, interpolation was used to determine rate

constants for negligible population of ES2 at low pH <8.0 in Fig 5a. The ES2 populations at a desired pH were computed based on p$K_a$ values obtained from fitting the measured ES2 populations versus pH while forward rate constants were determined from a linear fit to ln($k_{GS\rightarrow ES2}$) with pH while ($k_{ES2\rightarrow GS}$) is pH-independent. The population of ES1 and GS⇌ES1 rates of exchange were assumed to be pH-independent as supported by the RD NMR data (Supplemental Table 5). Interpolation was also used to determine rate constants and populations for ES2 at high temperatures/pH for CGA, TGA, GGG (indicated with *) to allow comparison to experimental data in Fig. 5d.

*Parameters and system of differential equations used in kinetic simulations*

Below is a description of all the parameters and system of differential equations used in the simulations.

<u>Rate constants as listed in Extended Data Fig. 7.</u>

$k_1$ = dNTP-on rate constant ($s^{-1}$ $uM^{-1}$)

$k_{-1}$ = dNTP-off rate constant. ($s^{-1}$)

$k_2$ = forward conformational change rate constant. Subscript *t* or *a* indicate transition from either tautomeric ES1 or anionic ES2.

$k_{-2}$ = reverse conformational change rate constant. Subscript *t* or *a* indicate transition from either tautomeric ES1 or anionic ES2.

$k_3$ = forward phosphodiester bond formation rate constant.

$k_{-3}$ = reverse phosphodiester bond formation rate constant.

$k_{GS\rightarrow ES1}$= forward rate constant of tautomeric ES1 formation.

$k_{ES1 \rightarrow GS}$ = reverse rate constant of tautomeric ES1 formation.

$k_{GS \rightarrow ES2}$ = forward rate constant of anionic ES2 formation.

$k_{ES2 \rightarrow GS}$ = reverse rate constant of anionic ES2 formation.

$k_{ES1 \rightarrow ES2}$ = rate constant for minor exchange from tautomeric ES1 to anionic ES2.

$k_{ES2 \rightarrow ES1}$ =rate constant for minor exchange from anionic ES2 to tautomeric ES1.

<u>Species</u>

E•DNA = Binary complex of a polymerase and primer-template DNA complex.

E•DNA•*dNTP* = Initial ternary complex following dNTP binding of a correct dNTP.

E•DNA•*dNTP*(WB) = Initial ternary complex following dNTP binding of an incorrect dNTP, assumed to adopt a wobble configuration.

E•DNA•*dNTP*(ES1) = Ternary complex with a Watson-Crick-like incorrect base pair adopting a tautomeric ES1 conformation.

E•DNA•*dNTP*(ES2) = Ternary complex with a Watson-Crick-like incorrect base pair adopting an anionic ES2 conformation.

E′•DNA•*dNTP* = Ternary DNA, dNTP, polymerase complex that has undergone conformational change and is poised for catalysis with either a Watson-Crick or Watson-Crick-like base pair.

E′•$DNA_{n+1}$•$PP_i$ = Product $DNA_{(i+1)}$ species following phosphodiester bond formation.

<u>Differential Equations for Correct Incorporation</u>

$$\frac{d[\text{E•DNA}]}{dt} = k_{-1}[\text{E•DNA•}dNTP] - k_1[\text{E•DNA}]$$

$$\frac{d[\text{E•DNA•}dNTP]}{dt} = k_1\,[\text{E•DNA}][\text{d}NTP] - k_{-1}[\text{E•DNA•}dNTP] + k_{-2}[\text{E}'\text{•DNA•}dNTP] - k_2[\text{E•DNA•}dNTP]$$

$$\frac{d[\mathrm{E'\bullet DNA\bullet}dNTP]}{dt} = k_2[\mathrm{E\bullet DNA\bullet}dNTP] - k_{-2}[\mathrm{E'\bullet DNA\bullet}dNTP] + k_{-3}[\mathrm{E'\bullet DNA_{n+1}\bullet}PP_i] - k_3[\mathrm{E'\bullet DNA\bullet}dNTP]$$

$$\frac{d[E'\bullet DNA_{n+1}\bullet PP_i]}{dt} = k_3[\mathrm{E'\bullet DNA\bullet}dNTP] - k_{-3}[\mathrm{E'\bullet DNA_{n+1}\bullet}PP_i]$$

<u>Differential Equations for Incorrect Incorporation</u>

$$\frac{d[\mathrm{E\bullet DNA}]}{dt} = k_{-1}[\mathrm{E\bullet DNA\bullet}dNTP(\mathrm{WB})] - k_1[\mathrm{E\bullet DNA}]$$

$$\frac{d[E\bullet DNA\bullet dNTP(WB)]}{dt} = k_1[\mathrm{E\bullet DNA}][dNTP] - k_{-1}[\mathrm{E\bullet DNA\bullet}dNTP(\mathrm{WB})] - k_{\mathrm{GS\rightarrow ES1}}[\mathrm{E\bullet DNA\bullet}dNTP(\mathrm{WB})] - k_{GS\rightarrow \mathrm{ES2}}[\mathrm{E\bullet DNA\bullet}dNTP(\mathrm{WB})] + k_{\mathrm{ES1\rightarrow GS}}[\mathrm{E\bullet DNA\bullet}dNTP(\mathrm{ES1})] + k_{\mathrm{ES2\rightarrow GS}}[\mathrm{E\bullet DNA\bullet}dNTP(\mathrm{ES2})]$$

$$\frac{d[E\bullet DNA\bullet dNTP(ES1)]}{dt} = k_{\mathrm{GS\rightarrow ES1}}[\mathrm{E\bullet DNA\bullet}dNTP(\mathrm{WB})] + k_{-2\mathrm{t}}[\mathrm{E'\bullet DNA\bullet}dNTP] + k_{\mathrm{ES2\rightarrow ES1}}[\mathrm{E\bullet DNA\bullet}dNTP(\mathrm{ES2})] - k_{\mathrm{ES1\rightarrow GS}}[\mathrm{E\bullet DNA\bullet}dNTP(\mathrm{ES1})] - k_{2\mathrm{t}}[\mathrm{E\bullet DNA\bullet}dNTP(\mathrm{ES1})] - k_{\mathrm{ES1\rightarrow ES2}}[\mathrm{E\bullet DNA\bullet}dNTP(\mathrm{ES1})]$$

$$\frac{d[E\bullet DNA\bullet dNTP(ES2)]}{dt} = k_{\mathrm{GS\rightarrow ES2}}[\mathrm{E\bullet DNA\bullet}dNTP(\mathrm{WB})] + k_{-2\mathrm{a}}[\mathrm{E'\bullet DNA\bullet}dNTP] + k_{\mathrm{ES1\rightarrow ES2}}[\mathrm{E\bullet DNA\bullet}dNTP(\mathrm{ES1})] - k_{\mathrm{ES2\rightarrow ES1}}[\mathrm{E\bullet DNA\bullet}dNTP(\mathrm{ES2})] - k_{2\mathrm{a}}[\mathrm{E\bullet DNA\bullet}dNTP(\mathrm{ES2})] - k_{\mathrm{ES2\rightarrow GS}}[\mathrm{E\bullet DNA\bullet}dNTP(\mathrm{ES2})]$$

$$\frac{d[E'\bullet DNA\bullet dNTP]}{dt} = k_{2\mathrm{t}}[\mathrm{E\bullet DNA\bullet}dNTP(\mathrm{ES1})] + k_{2\mathrm{a}}[\mathrm{E\bullet DNA\bullet}dNTP(\mathrm{ES2})] + k_{-3}[\mathrm{E'\bullet DNA_{n+1}\bullet}PP_i] - k_{-2\mathrm{t}}[\mathrm{E'\bullet DNA\bullet}dNTP] - k_{-2\mathrm{a}}[\mathrm{E'\bullet DNA\bullet}dNTP] - k_3[\mathrm{E'\bullet DNA\bullet}dNTP]$$

$$\frac{d[E'\bullet DNA_{n+1}\bullet PP_i]}{dt} = k_3[\mathrm{E'\bullet DNA\bullet}dNTP] - k_{-3}[\mathrm{E'\bullet DNA_{n+1}\bullet}PP_i]$$

**Kinetic Assays.** All DNA substrates (Supplementary Table 8) used in pre-steady-state kinetic assays were purchased from Integrated DNA Technologies, INC. (Coralville, IA, USA), purified by denaturing PAGE, and recovered using the crush and soak method. Purified DNA oligonucleotide primers were 5′-radiolabeled by incubating with Optikinase and [$\gamma$-$^{32}$P]ATP for 3 hr at 37 °C and then purifying from free [$\gamma$-$^{32}$P]ATP by passing through a Bio-Spin 6 column (Bio-Rad). The 5′-radiolabeled primers were annealed to DNA oligonucleotide templates by mixing the primer with a 1.15-fold molar excess of a complementary template and incubating the mixture at 95 °C for 5 min before cooling slowly to room temperature over several hours. For pre-steady-state kinetic assays, a pre-incubated solution of 300 nM human DNA polymerase β and 30 nM 5′-radiolabeled DNA substrate in reaction buffer (50 mM Tris-Cl, pH 8.4 at 25 °C; 50 mM NaCl; 0.1 mM EDTA; 5 mM $MgCl_2$; 5 mM DTT; 0.1 mg/mL BSA; and 10% glycerol) was rapidly mixed with reaction buffer containing either 100 μM dCTP or 100 μM dTTP. Reactions were quenched after varying incubation times with the addition of 0.37 M EDTA. All reported concentrations in the reaction mixtures are final. Fast reactions were performed using a rapid chemical quench-flow apparatus (KinTek). Reaction products were separated by denaturing polyacrylamide gel electrophoresis (17% acrylamide, 8 M urea) and quantitated using a Typhoon Trio or RGB (GE Healthcare) and ImageQuant (Molecular Dynamics). All kinetic data were fit by nonlinear regression using KaleidaGraph (Synergy Software). Product concentration was plotted against reaction time and a single-exponential equation, $[\text{product}] = A[1 - \exp(-k_{obs}t)]$, was fit to the data to yield the observed nucleotide incorporation rate constant $k_{obs}$. All reactions were carried out in triplicate and the reported

errors reflect the standard deviation from the average value determined from three measurements of $k_{obs.}$

**Code Availability.** Custom Python 2.7 programs written in-house were used to simulate and fit both NMR and kinetic data. Code for NMR fitting, simulation of relaxation dispersion data, and kinetics simulations are available on GitHub (https://github.com/alhashimilab).

**Data Availability.** The data that support the findings of this study are available upon reasonable request from the corresponding authors.

**Additional References**


50. Wagenmakers, E.-J. & Farrell, S. AIC model selection using Akaike weights. *Psychon. Bull. Rev.* **11,** 192–196 (2004).

51. Smith, D. E., Su, J. Y. & Jucker, F. M. Efficient enzymatic synthesis of $^{13}C$,$^{15}N$-labeled DNA for NMR studies. *J. Biomol. NMR* **10,** 245–253 (1997).

52. Dethoff, E. A., Petzold, K., Chugh, J., Casiano-Negroni, A. & Al-Hashimi, H. M. Visualizing transient low-populated structures of RNA. *Nature* **491,** 724–728 (2012).

53. Xue, Y. *et al.* Characterizing RNA excited states using NMR relaxation dispersion. *Meth. Enzymol.* **558,** 39–73 (2015).

54. Sowers, L. C., Goodman, M. F., Eritja, R., Kaplan, B. & Fazakerley, G. V. Ionized

and wobble base-pairing for bromouracil-guanine in equilibrium under physiological conditions. *J. Mol. Biol.* **205**, 437–447 (1989).

55. Hunter, W. N., Brown, T., Kneale, G., Anand, N. N., Rabinovich, D., Kennard, O. The structure of guanosine-thymidine mismatch in B-DNA at 2.5-A resolution. *J. Biol. Chem.* **262,** 9962–9970 (1987).

56. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–293 (1995).

57. Bothe, J. R., Stein, Z. W. & Al-Hashimi, H. M. Evaluating the uncertainty in exchange parameters determined from off-resonance $R_{1\rho}$ relaxation dispersion for systems in fast exchange. *J. Magn. Reson.* **244,** 18–29 (2014).

58. Branch, M. A., Coleman, T. F. & Li, Y. A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems. *SIAM J. Sci. Comput.* **21,** 1–23 (1999).

59. Baldwin, A. J. & Kay, L. E. An $R_{1\rho}$ expression for a spin in chemical exchange between two sites with unequal transverse relaxation rates. *J. Biomol. NMR* **55,** 211–218 (2013).

60. Burnham, K. P. & Anderson, D. R. Multimodel inference. Understanding AIC and BIC in model selection. *Socio. Meth. Res.* **33,** 261–304 (2004).

61. Markley, J. L. *et al.* BioMagResBank (BMRB) as a partner in the Worldwide Protein Data Bank (wwPDB): new policies affecting biomolecular NMR depositions. *J. Biomol. NMR* **40,** 153–155 (2008).

62. Boggs, P. T., Byrd, R. H. & Schnabel, R. B. A stable and efficient algorithm for nonlinear orthogonal distance regression. *SIAM J. Sci. Stat. Comput.* **8,** 1052–1078

(1987).

63. Lu, X. J & Olson, W. K. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc*. **3**, 1213-1227 (2008).

64. Berman, H. M. The Protein Data Bank. *Nucleic Acids Res*. **28**, 235–242 (2000).

65. Johnson, K. A. The kinetic and chemical mechanism of high-fidelity DNA polymerases. *Biochim. Biophys. Acta.* **1804**, 1041-1048 (2010).

66. Johnson, K. A. Fitting Enzyme Kinetic Data with KinTek Global Kinetic Explorer. *Methods Enzymol*. **467**, 601–626 (2009).

## SUPPLEMENTARY TABLE LEGENDS

**Supplementary Table 1 | Exchange parameters and statistics for fitting tautomeric ES1 RD profiles in DNA and RNA at near-neutral pH in unmodified and modified constructs.**

B-M 2-state fits of RD profiles measured in unmodified and modified DNA and RNA at near-neutral pH for data shown in Fig. 2. Listed are the B-M fitted chemical exchange parameters, measurement conditions, and buffer composition. tp5abc-AGG undergoes slow exchange with a second independent ES characterized elsewhere[53].

**Supplementary Table 2 | Exchange parameters and statistics for explicit rapid-tautomer 3-state fit in DNA and RNA at neutral pH.**

B-M 3-state fits of RD profiles in Supplementary Table 1 assuming the linear wobble⇌G•$T^{enol}$/$U^{enol}$⇌$G^{enol}$•T/U topology[25]. Listed are the B-M fitted chemical exchange parameters, measurement conditions, and buffer composition. The mismatch-specific fundamental tautomeric chemical shifts obtained from this analysis vary by ≈9/4 p.p.m. for $\Delta\omega_{Genol\text{-}N1}/\Delta\omega_{Genol\text{-}N3}$ and ≈10/9 p.p.m. for $\Delta\omega_{Tenol/Uenol\text{-}N3}/\Delta\omega_{Tenol/Uenol\text{-}N1}$. These variations are ≈2-4-fold greater than those observed for the corresponding GS chemicals shift (Extended Data Fig. 3c). Data for the glnA riboswitch is included for completeness but left out of analysis due to a second confounding slow exchange process, which introduces large errors in the fitting.

**Supplementary Table 3 | Fundamental tautomer chemical shifts.**

A table of the fundamental tautomer chemical shifts derived from experimental measurements and from density functional theory predictions[8].

**Supplementary Table 4 | Statistical comparisons of exchange parameters obtained from B-M fitting RD profiles assuming starlike, linear, and triangular topologies.**

Shown are the fitted exchange parameters and statistical weights ($w_{AIC}$ and $w_{BIC}$) for DNA and RNA RD constructs given in Supplementary Table 5, assuming either 3-state starlike, linear, or triangular topologies.

**Supplementary Table 5 | Comprehensive list of 3-state fits of RD profiles measured in DNA and RNA assuming starlike and triangular topologies at various temperatures and pH conditions.**

All measured B-M 3-state fits (with and without minor exchange) for data at various temperatures and pH (≥7.8) conditions. Listed are the B-M fitted chemical exchange parameters, measurement conditions, buffer composition, flux calculations for excited state formation either directly or via minor exchange.

**Supplementary Table 6 | Comparison of ES1 $<\Delta\omega_{G\text{-}N1}>$ and $<\Delta\omega_{T/U\text{-}N3}>$ chemical shifts in starlike versus triangular topologies.**

Reference ES1 $<\Delta\omega_{G\text{-}N1}>$ and $<\Delta\omega_{T/U\text{-}N3}>$ chemical shifts derived from low pH (<7.8) RD data in the absence of the influence of ES2 are shown. ES1 $<\Delta\omega_{G\text{-}N1}>$ and $<\Delta\omega_{T/U\text{-}N3}>$ chemical shifts for each corresponding construct at high pH (≥7.8) fit to starlike and triangular topologies are shown for comparison. Calculated residuals between starlike and

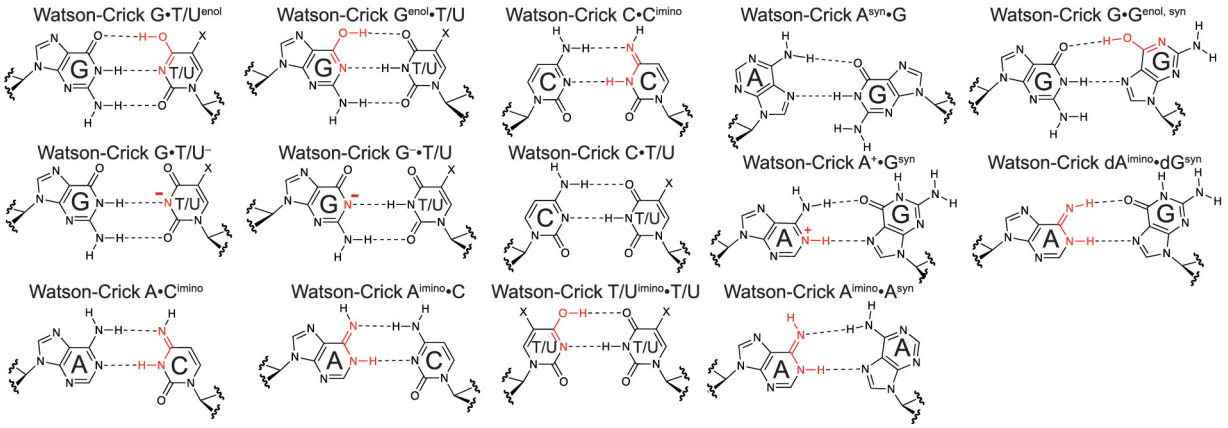
reference; linear and reference chemical shifts reflect similarity between starlike/triangular and reference chemical shifts. The lower residuals indicate that inclusion of minor exchange lowers the ES1 chemical shift peculiarities and brings them closer to the reference ES1 chemical shifts.

**Supplementary Table 7 | Rate constants and results from kinetic modeling.**

**Supplementary Table 8 | $k_{obs}$ values and constructs from *in vitro* assays with human DNA polymerase β.**

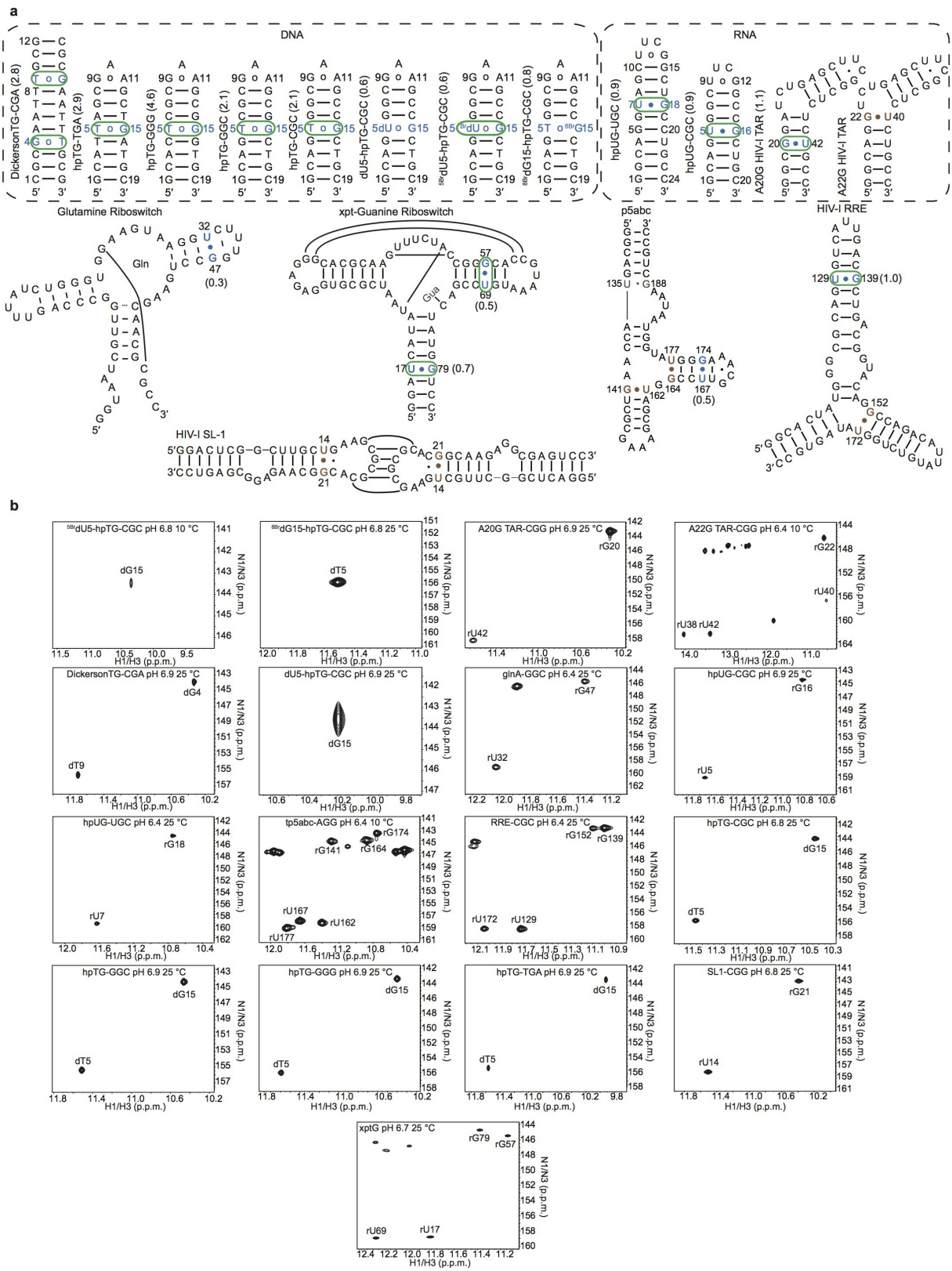
Values reported are the average ± one s.d. from $n = 3$ biologically independent experiments.

**Supplementary Table 9 | Spinlock powers and offsets for hpTG-GGC and hpUG-CGC pH 6.9 / 8.4 at 10 °C.**



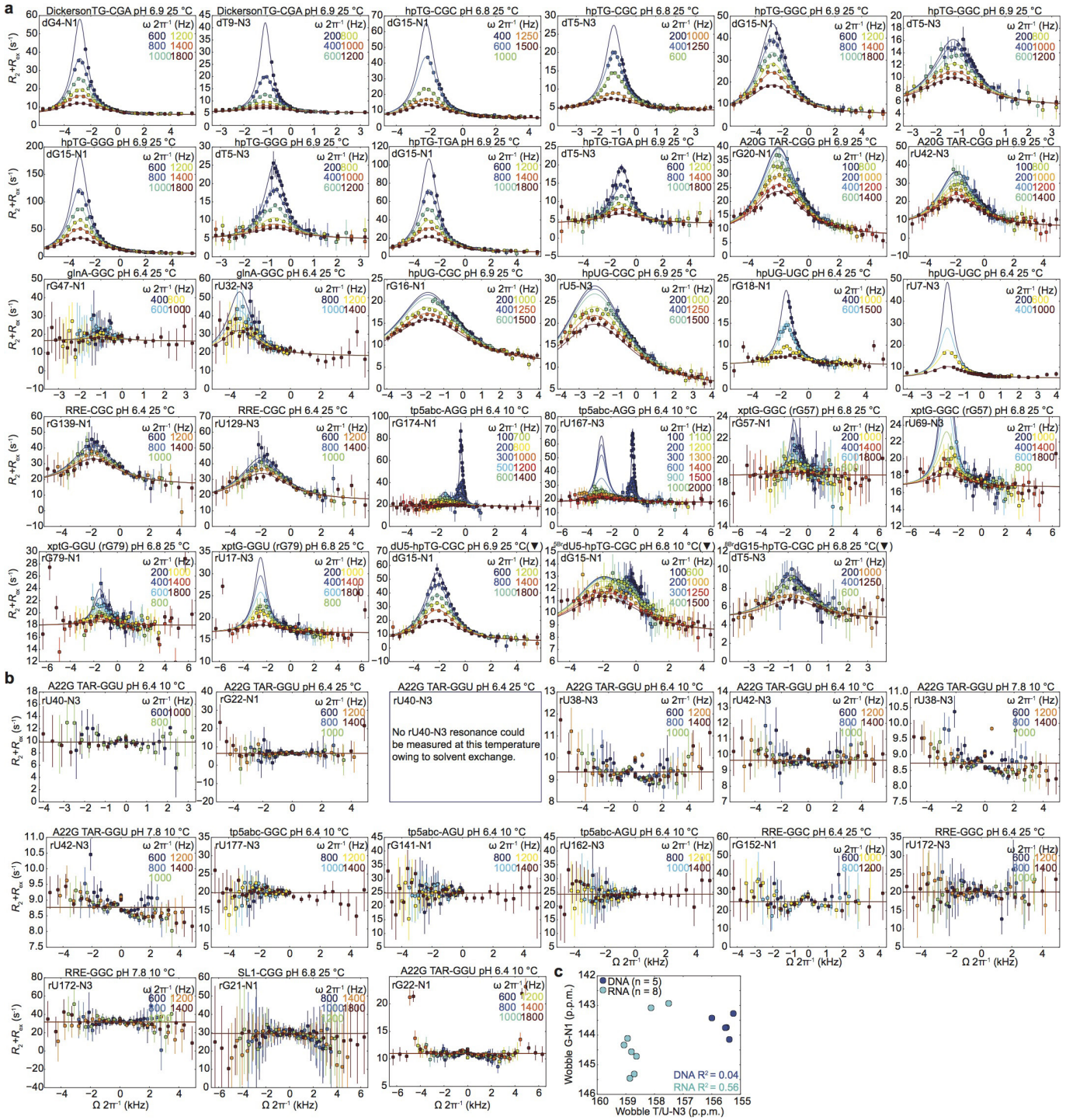

**Extended Data Figure 1 | Watson–Crick-like mismatches. a**, Watson–Crick-like mismatches stabilized by tautomeric and ionic base forms.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.




**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 towards 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 relaxation dispersion 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 previous 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 relaxation dispersion measurements. The spectrum shown for xptG was collected at pH 6.7 and 25 °C in potassium acetate buffer as described previously[8].

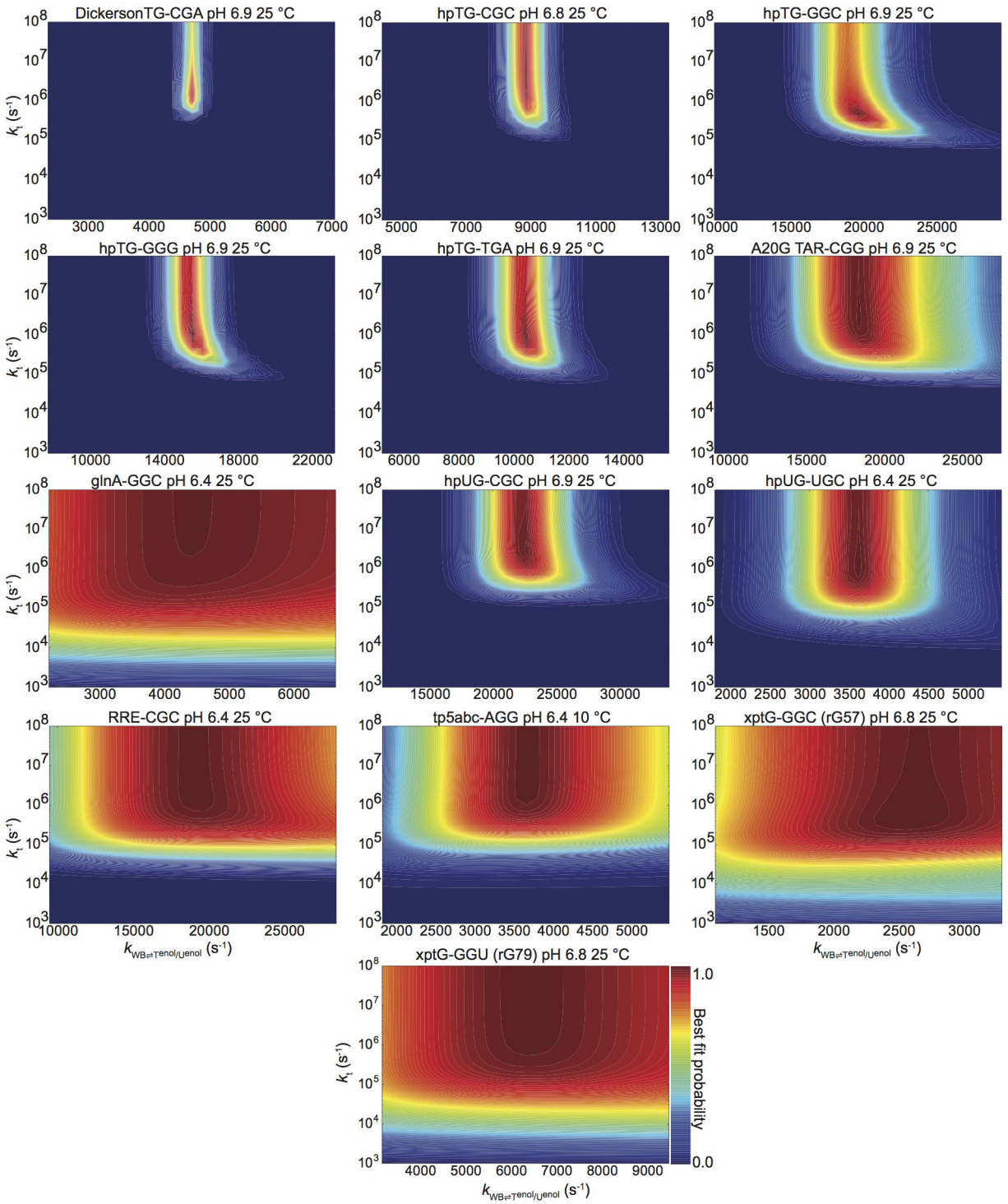
© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.




**Extended Data Figure 3 | Relaxation dispersion profiles measured in DNA and RNA at near-neutral pH. a**, $^{15}$N G-N1 and T/U-N3 relaxation dispersion 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[50]. The trend lines represent B–M two- or three-state fits. Constructs containing a chemically modified base are indicated by ▼. **b**, The absence of $^{15}$N relaxation dispersion for rG•rU mismatches near bulges, apical loops or three-way junctions. Error bars reflect experimental uncertainty (one s.d., see Supplementary 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** and **b** reflect experimental uncertainty (one s.d., see Supplementary Methods).

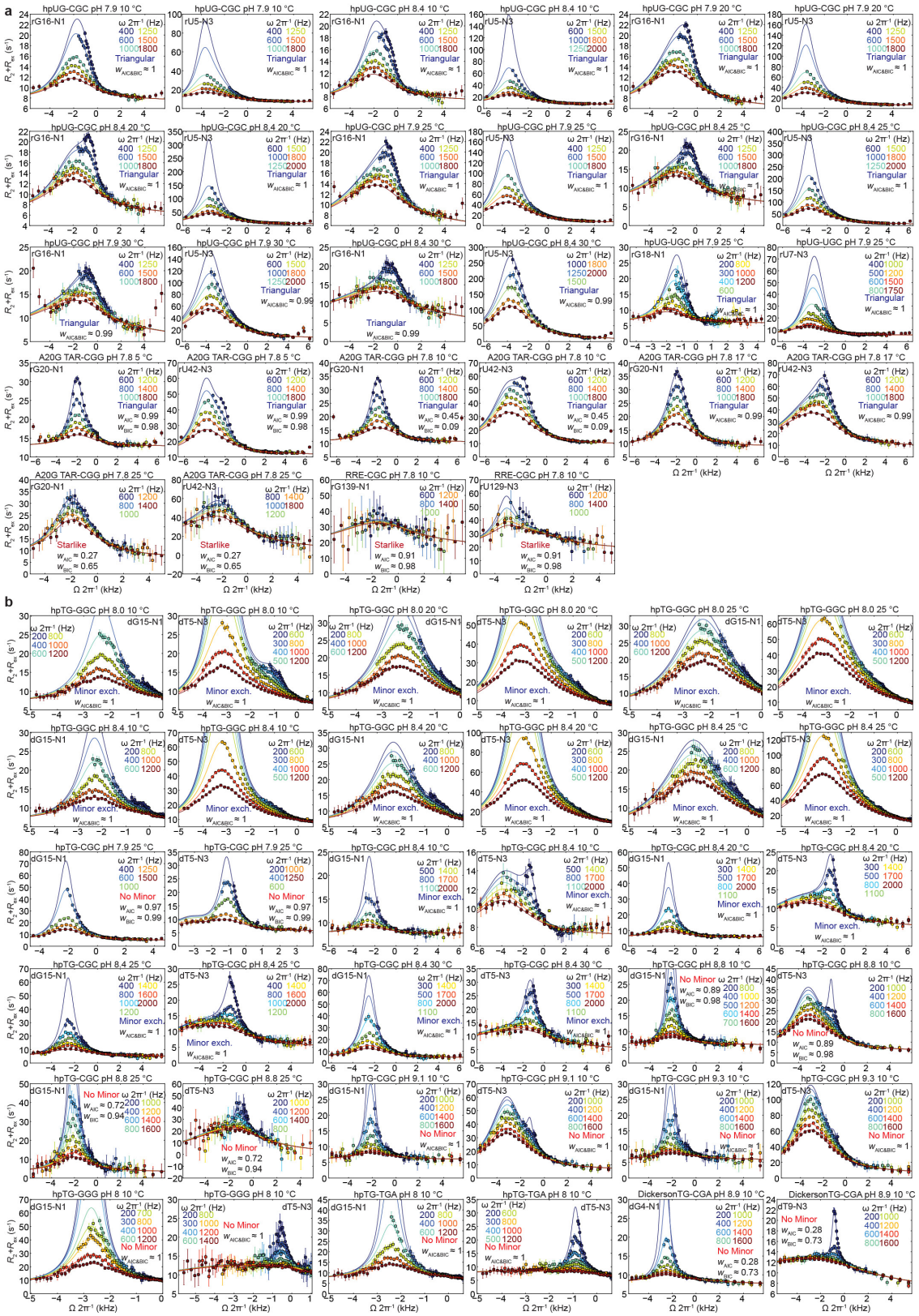
50. Xue, Y. *et al.* Characterizing RNA excited states using NMR relaxation dispersion. *Methods Enzymol.* **558,** 39–73 (2015).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.




**Extended Data Figure 4 | Establishing lower limits for the rates of base pair tautomeric exchange.** Agreement between measured and predicted $R_{1\rho}$ values (scaled $\overline{\chi}^2$ weight, equation (2) in Supplementary Methods) when varying the exchanges rates of wobble $\rightleftharpoons$ G•T$^{enol}$/U$^{enol}$ ($k_{WB\rightleftharpoons T^{enol}/U^{enol}}$) and G•T$^{enol}$/U$^{enol}$ $\rightleftharpoons$ G$^{enol}$•T/U ($k_t$). See Supplementary Methods for additional details.

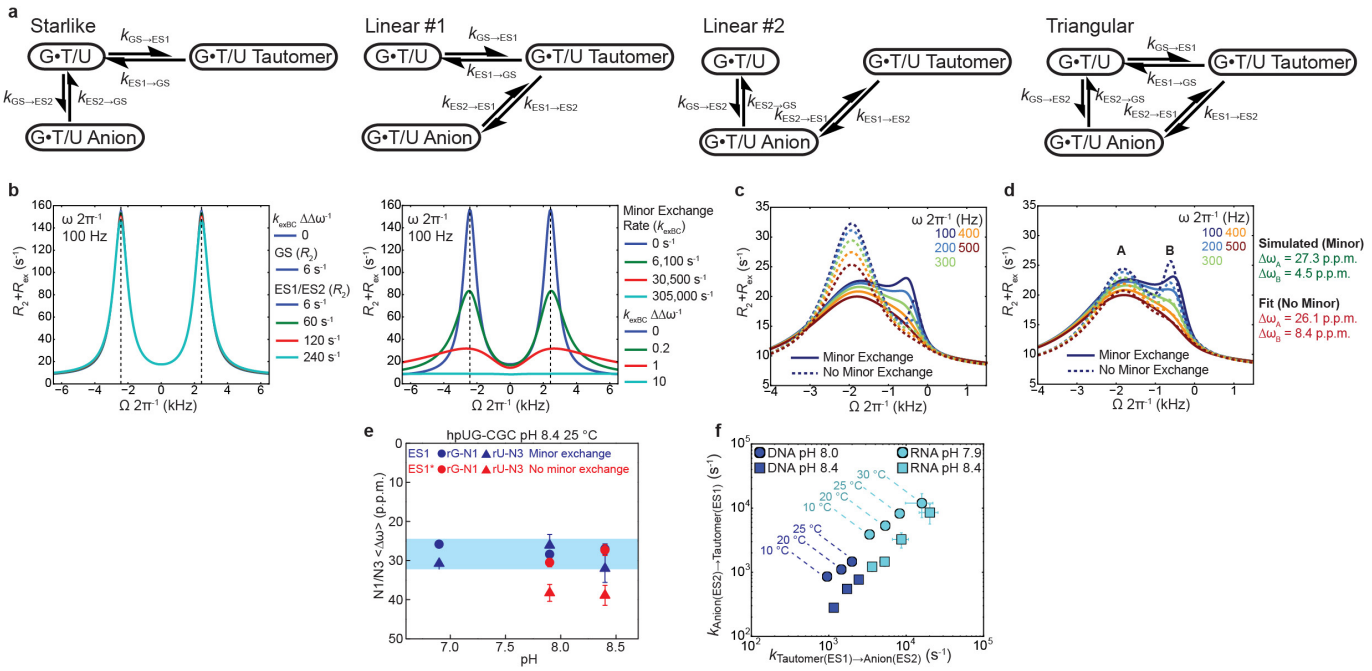
© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



**Extended Data Figure 5 | Relaxation dispersion profiles measured in DNA and RNA at high pH. a**, **b**, B–M three-state fits of RNA (**a**) and DNA (**b**) $^{15}$N relaxation dispersion data for starlike and triangular topologies (as indicated within the plots). The relative statistical weights $w_{AIC}$ and $w_{BIC}$ (refs 51, 52) 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 Supplementary Methods).

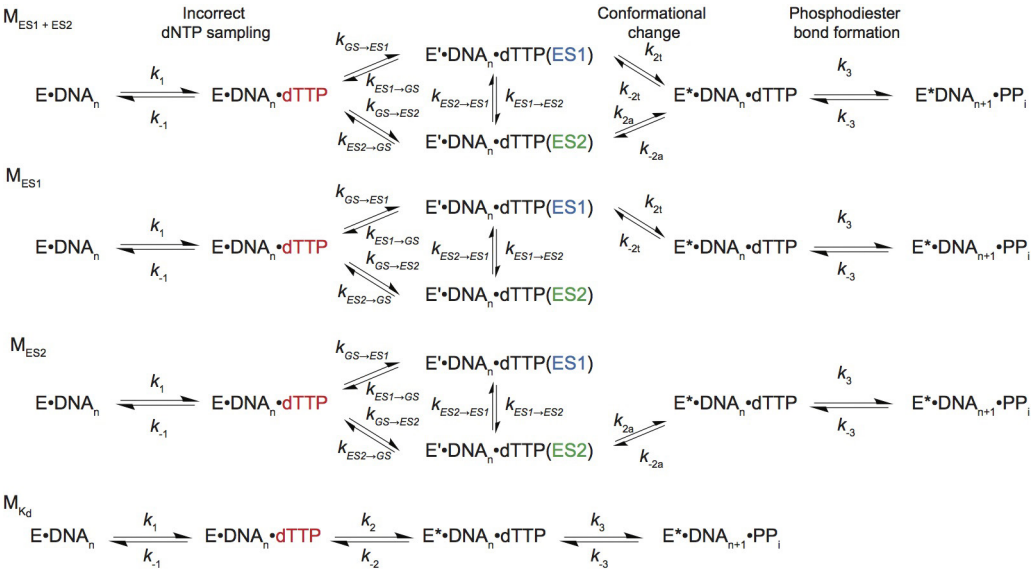

51. Wagenmakers, E.-J. & Farrell, S. AIC model selection using Akaike weights. *Psychon. Bull. Rev.* **11,** 192–196 (2004).
52. Burnham, K. P. & Anderson, D. R. Multimodel inference. Understanding AIC and BIC in model selection. *Sociol. Methods Res.* **33,** 261–304 (2004).


© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



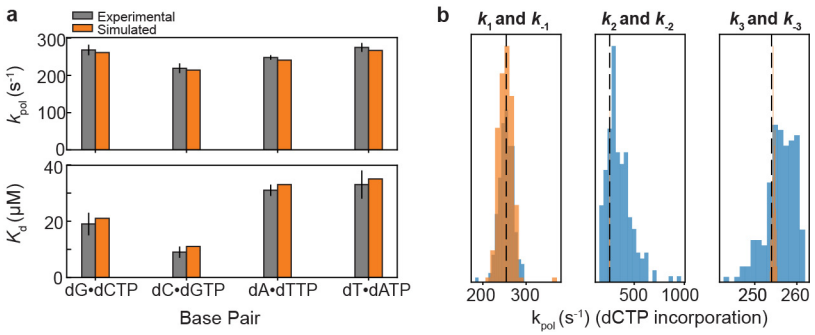

**Extended Data Figure 6 | Discerning minor exchange between WC-like tautomeric and anionic G•T/U mismatches a**, Topologies used to model chemical exchange. Individual rate constants are shown for each process of the different topologies. **b**, Left, B–M simulations showing that when $R_{2(GS)} \neq R_{2(ES1)}$ 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 relaxation dispersion 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 ($k_{ES1 \rightarrow ES2} = 0$ and $k_{ES2 \rightarrow ES1} = 0$) are also shown (dashed lines). **d**, Dashed lines denote three-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- or underestimation of the true (green) versus fitted (red) ES chemical shifts when fitting relaxation dispersion profiles that have triangular topology with minor exchange to a starlike model that has no minor exchange. **e**, The ES1 $\langle \Delta\omega_{rG\text{-}N1} \rangle$ and $\langle \Delta\omega_{rU\text{-}N3} \rangle$ values as a function of pH derived from the three-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** and **f** reflect experimental uncertainty (one s.d., Supplementary Methods).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



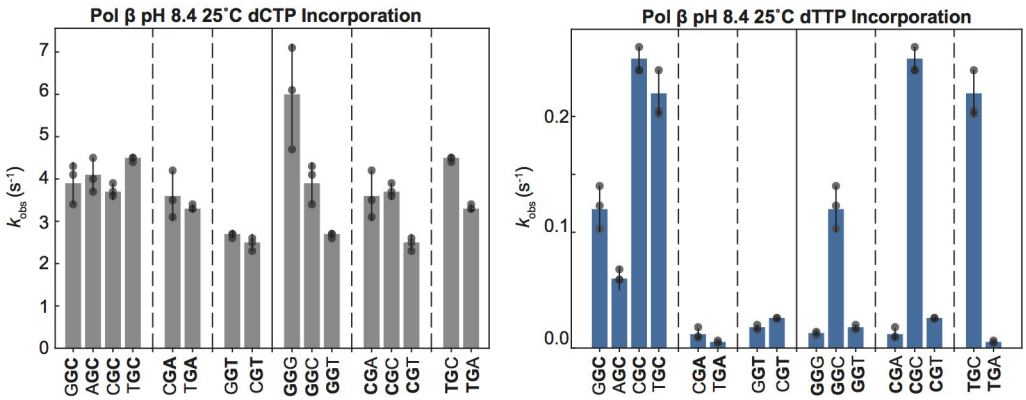

**Extended Data Figure 7 | Kinetic mechanisms used to model misincorporation.** Rate constants for each step are listed in Supplementary Table 7.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



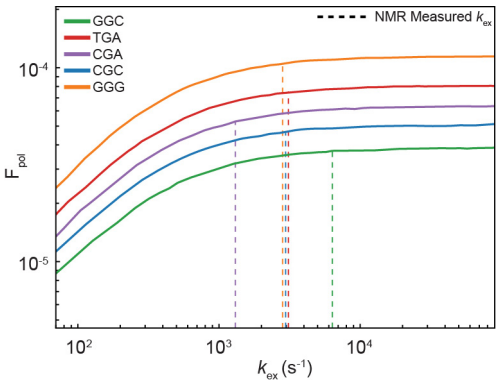

**Extended Data Figure 8 | Benchmarking kinetic simulations of misincorporation. a**, Comparison of $k_{pol}$ and $K_d$ values for correct incorporation measured experimentally for human DNA polymerase ε with values computed on the basis of pre-steady state simulations using the microscopic rate constants 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 upon varying the rate constants (forward, blue; reverse, orange) for steps other than tautomerization or ionization by twofold ($n = 200$ independent simulations in which rate constants were varied randomly by up to twofold). As expected, the only rate constant with a substantial effect on the reported $k_{pol}$ values was the rate-limiting conformational change step $k_2$ (middle).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.




**Extended Data Figure 9 | $k_{obs}$ values measured for human DNA polymerase β insertion.** Incorporation of dCTP (dCTP•dG) is shown on the left, that of dTTP (dTTP•dG) on the right. pH 8.4, 25 °C, 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 experiments) are indicated by grey circles. Bar height reflects average of replicates, and error bars reflect one s.d.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.




**Extended Data Figure 10 | $F_{pol}$ is primarily governed by ES1 populations.** Simulated $F_{pol}$ values as a function of scaling up or scaling down of 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 experimentally in this study (green dotted line) minimally affects $F_{pol}$; decreasing the $k_{ex}$ within the range measured experimentally in this study (purple dotted line) also affects the value of $F_{pol}$ only minimally. Much larger decreases in $k_{ex}$ are required to significantly reduce the value of $F_{pol}$.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.