*Published online 20 October 2021*

*Nucleic Acids Research, 2021, Vol. 49, No. 20* ***11823–11833***
*https://doi.org/10.1093/nar/gkab947*

# Constraints on error rate revealed by computational study of G•U tautomerization in translation

**Andriy Kazantsev*** **and Zoya Ignatova***

Institute of Biochemistry and Molecular Biology, Department of Chemistry, University of Hamburg, Martin-Luther-King-Platz 6, 20146 Hamburg, Germany

Received April 23, 2021; Revised September 30, 2021; Editorial Decision September 30, 2021; Accepted October 01, 2021

## ABSTRACT

**In translation, G•U mismatch in codon-anticodon decoding is an error hotspot likely due to transition of G•U from wobble (wb) to Watson-Crick (WC) geometry, which is governed by keto/enol tautomerization (wb-WC reaction). Yet, effects of the ribosome on the wb-WC reaction and its implications for decoding mechanism remain unclear. Employing quantum-mechanical/molecular-mechanical umbrella sampling simulations using models of the ribosomal decoding site (A site) we determined that the wb-WC reaction is endoergic in the open, but weakly exoergic in the closed A-site state. We extended the classical 'induced-fit' model of initial selection by incorporating wb-WC reaction parameters in open and closed states. For predicted parameters, the non-equilibrium exoergic wb-WC reaction is kinetically limited by the decoding rates. The model explains early observations of the WC geometry of G•U from equilibrium structural studies and reveals discrimination capacity for the working ribosome operating at non-equilibrium conditions. The equilibration of the exoergic wb-WC reaction counteracts the equilibration of the open-closed transition of the A site, constraining the decoding accuracy and potentially explaining the persistence of the G•U as an error hotspot. Our results unify structural and mechanistic views of codon-anticodon decoding and generalize the 'induced-fit' model for flexible substrates.**

## INTRODUCTION

Accurate expression of genetic information is essential for cell fitness (1,2), although base pair recognition errors during DNA replication and mRNA translation inevitably occur. In translation, the G•U mismatch in the middle position of the codon-anticodon pair has a fairly large error rate of $\sim 10^{-4}$–$10^{-3}$ (3–8). G•U(T) base pair formed with the canonical keto tautomers of the nucleobases is unable to adopt the Watson-Crick (WC) geometry and in aqueous environment predominantly exists in the wobble (wb) geometry (9). This structural difference between G•U(T) wb and the canonical base pairs is believed to be the basis of their selective recognition by DNA polymerases and the ribosome (10,11). The tautomeric hypothesis by Watson and Crick (12) suggests that a source of base pair recognition errors can be attributed to keto/enol tautomerism of nucleobases, which allows G*•U(T) or G•U*(T*) (where asterisk denotes the enol tautomer) to adopt the WC geometry.

Using computational chemistry methods, Brovarets' and Hovorun predicted a tautomerization reaction connecting wb and WC geometries in G•U(T) base pair (13,14) (wb-WC reaction, Figure 1 A), which was experimentally confirmed by NMR in DNA and RNA duplexes in water solution (9). Recently, the incorporation of tautomeric populations in DNA duplexes in water solution into a kinetic model of DNA replication reached a good agreement between predicted and experimentally determined misincorporation rates (15). It might indicate a negligible role of the active site environment on the properties of the wb-WC reaction. However, both experimental (16) and computational (17,18) studies demonstrate stabilization of the WC geometry of G•T in closed state of the active site in several DNA polymerases, implying a strong effect of the active site environment on the wb-WC reaction and thus, a more complicated mechanism of base pair recognition. The WC geometry of G•U at the first two codon-anticodon positions was also observed in closed state of the ribosomal decoding site (A site) in both X-ray (19–22) and in cryo-EM structures (23). Previous computational studies have concluded that at the first two codon positions the WC geometry of G*•U/ G•U* is either not stabilized in closed decoding site (24,25), or, if stabilized, the complex is less favorable compared to the complex with cognate tRNA (24). These studies were performed with classical force field approximation, and reached opposing conclusions to the structural observations at equilibrium conditions (19). Yet, the effect of the

*To whom correspondence should be addressed. Tel: +49 40 42838 2332; Email: zoya.ignatova@uni-hamburg.de
Correspondence may also be addressed to Andriy Kazantsev. Email: kazantsev.and@gmail.com
Present address: Andriy Kazantsev, Institute of Biological Information Processing IBI-1, Forschungzentrum Jülich, 52425 Jülich, Germany.

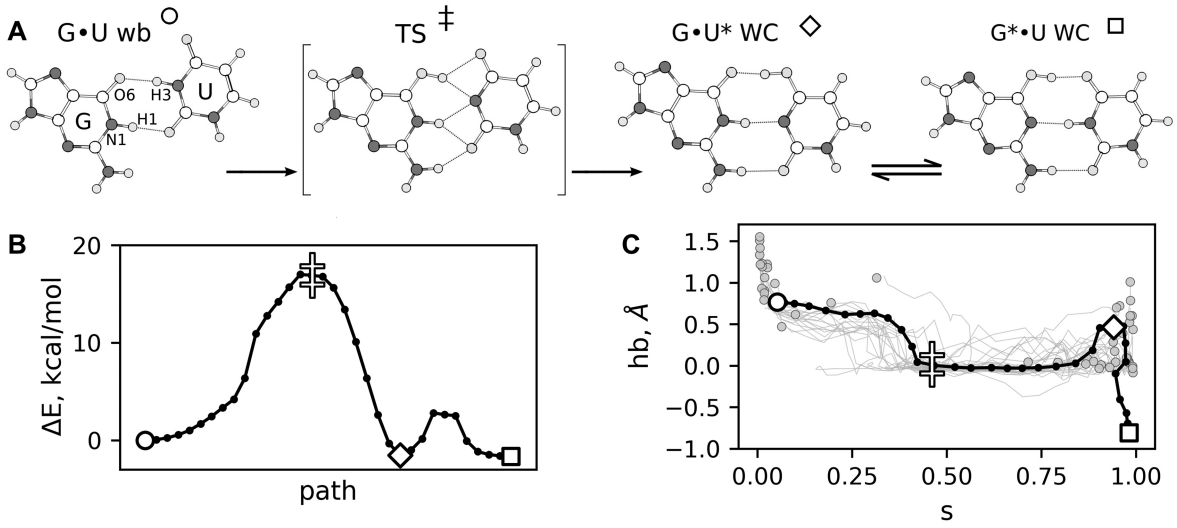


**Figure 1.** The wb-WC reaction. (**A**) Local minima and the transition state (TS) along the wb-WC reaction. Asterisks denote enol tautomers. Each state is designated by a symbol according to the states in panels (B) and (C). The wb-WC reaction consists of the slow (activation free energy $\Delta G_{\ddagger} = 15 - 20$ kcal/mol) tautomerization reaction from G•U wb to G•U* WC, followed by fast ($\Delta G_{\ddagger} \approx 5$ kcal/mol) double proton transfer reaction between G•U* and G*•U. (**B**) Reference potential energy profile of the wb-WC reaction calculated with nudge elastic band method (NEB) in gas phase. (**C**) Reference wb-WC reaction path from NEB, projected on the collective variables used in umbrella sampling calculations *s* (path collective variable) and *hb* (proton transfer variable). Gray lines and gray circles denote trajectories from the committor analysis, initiated from the TS.

ribosomal decoding site environment on the wb-WC reaction remains unclear.

The structural observations (19–23) challenge the classical 'induced fit' model of substrate recognition in replication and translation, which assumes rigid substrates and fixed base pair geometries during the recognition. This model also predicts a seemingly suboptimal codon–anticodon decoding. The open-closed conformational transition of the ribosomal decoding site, the molecular mechanism of the 'induced fit' model (26), provides the highest decoding accuracy in the equilibrium limit, but is shifted far from equilibrium due to its high forward rates, resulting in a great loss of accuracy (27–29). While trade-offs between rate and accuracy can explain evolutionary optimization of some translation rates (29), seemingly suboptimal open-closed transition rates cannot be easily justified this way, because this decoding step is not rate-limiting in translation elongation *in vitro* (30). These problems indicate that the induced fit model fails to explain the mechanism of codon-anticodon decoding.

Hovorun *et al.* have suggested a non-equilibrium exoergic wb-WC reaction in the closed state of a DNA polymerase active site, kinetically limited by the lifetime of the recognition state (14). Such mechanism, applied to codon-anticodon decoding, may solve the problem with the induced fit model, but requires formalization and detailed analysis. Overall, improving our understanding of the errors induced by G•U mismatch in codon-anticodon decoding would require accurate calculations of the effects of ribosomal decoding site on the wb-WC reaction energetics and a revised kinetic model of decoding to incorporate such effects.

We used quantum-mechanical/molecular-mechanical (QM/MM) umbrella sampling (US) method to calculate the potential of mean force (PMF) of the wb-WC reaction in the ribosomal A-site models. Focusing only on free energy of the WC geometry formation ($\Delta G_{\text{wc}}$), we benchmarked this approach on simpler systems with available $\Delta G_{\text{wc}}$ data from experiments or high levels of QM theory and obtained satisfactory agreement with the data. Our results reveal positive $\Delta G_{\text{wc}}$ (endoergic wb-WC reaction) in open A site, and negative $\Delta G_{\text{wc}}$ (exoergic wb-WC reaction) in the closed A site. The dielectric constant ε of the environment decreased in the closed A site, which might affect the wb-WC reaction. To study the implications of the exoergic wb-WC reaction in closed A site, we incorporated the wb-WC reaction into the kinetic model of initial selection in translation and derived the analytical solutions for our model. Our results are consistent with previously reported WC geometry of G•U in the structural studies at equilibrium conditions (19–23) and provide means for G•U discrimination by the working ribosome at non-equilibrium conditions. Furthermore, it can potentially explain the seemingly suboptimal decoding by revealing constraints on the decoding accuracy.

## MATERIALS AND METHODS

### QM calculations

We performed quantum chemistry (QM) calculations to obtain the reference dependence of the wb-WC equilibrium on the dielectric constant of a solvent, and to select the optimal QM level of theory for our QM/MM calculations. Using the geometries of the wb-WC reaction minima and transition state (TS) optimized at density functional theory (DFT) levels ωB97X-D3/def2-TZVP, B3LYP-D3BJ/def2-TZVP and BLYP-D3BJ/def2-SVP in gas phase and implicit water model, we calculated single-point energies with multiple DFT and semiempirical (SE) methods. The methods were assessed based on the energies ($\Delta E_{\text{wc}}$,

Supplementary Table S1) or enthalpies ($\Delta H_{wc}$, Supplementary Table S2) of the WC geometry formation ($\Delta E_{wc}$ = $\Delta E$(G*•U WC) – $\Delta E$(G•U wb)). PM3 method was rejected due to the spurious pathway of the wb-WC reaction it predicted (Supplementary Figure S2). Our benchmark calculations showed PM7 (31) as an optimal QM method for our QM/MM approach, as it was able to accurately reproduce $\Delta H_{wc}$ calculated with DFT (Supplementary Table S2), while allowing vast sampling due to its low computational cost. The activation enthalpies obtained with PM7 were less accurate (Supplementary Table S2), which precluded the use of activation free energies derived in our QM/MM simulations. Orca 4.2.1 (32) was used for all DFT and *ab initio* QM calculations, while MOPAC-2016 (James J. P. Stewart, Stewart Computational Chemistry, USA, http://openmopac.net) was used for all SE calculations.

### System setup and MD simulations

We studied the environmental effects on the wb-WC reaction in the following systems: DNA heptamer in water solution and a single G•T base pair in benzene (benchmark systems); ribosomal A-site models: closed, abasic and open; human DNA polymerase β (pol-β) and DNA polymerase from T7 virus (T7-pol).

DNA heptamer model was prepared from the solution NMR structure of DNA dodecamer (PDB ID: 1BJD), centered on one of the G•T wobble base pairs in the dodecamer. Benzene model was prepared by solvating a single G•T base pair from the DNA heptamer system in a benzene box.

pol-β model was prepared from the crystal structure of pol-β with a G•C base pair in the closed active site (PDB ID: 4KLF) (33), in which we mutated dCTP to dTTP. T7-pol was prepared in a similar way using the crystal structure with PDB ID 1T7P (34) as the initial coordinates.

To prepare the A-site models, we used the crystal structure of *T. thermophilus* 70S ribosome bound to $tRNA^{Thr}$ in the A site (PDB ID: 6GSK) with U•G base pair in the second codon (AUC)–anticodon (GGU) position (22). All residues within 35 Å radius of the second codon nucleotide were selected for the model (Figure 2 A). The outer shell of the selected sphere was restrained in all simulations. In the 'closed' (native) A-site model, the rRNA residues A1492, A1493 and G530 were in *out* conformation, surrounding the codon-anticodon helix (Figure 2B). In the 'abasic' model, the nucleobases of these residues were deleted. In the 'open' model, used only in QM/MM US simulations (see below), we applied harmonic biases from (25) to restrain A1492 and A1493 to the *in* position. All $Mg^{2+}$ ions from the ribosome crystal structure were deleted in the models.

Unless specified otherwise, all models were solvated in TIP3P (35) water box with 0.15 M of NaCl. We employed a standard equilibration protocol (see see Supplementary Methods), comprising steepest-descent minimization, solvent equilibration in NPT ensemble at standard conditions and gradual heating of the solutes to 298 K. Production MD simulations were performed in NVT ensemble using Langevin thermostat with 2 fs integration step. CHARMM36 force field (36–38) in NAMD 2.12 (39) was used for all classical MD simulations.

### Dielectric constant calculations

To analyze the dielectric properties around the base pairs in the simulated models, we applied Kirkwood-Fröhlich formula (KFF), which relates the dielectric constant (ε) to dipole moment fluctuations in a selected volume (40). Using the classical MD trajectories, we calculated ε of all nucleic acid, protein and water residues in spheres of radii ranging from 5 to 12 Å centered on the base pair of interest. Since ε calculated in a TIP3P water box using this approach depended on the probe volume radius until its convergence at ∼20 Å (Supplementary Figure S3), the approach only allows a qualitative comparison between the studied systems.

### QM/MM US simulations

We applied QM/MM umbrella sampling simulations (US) to calculate the free energy profiles (potential of mean force, PMF) of the wb-WC reaction in the models of ribosomal A site and active site of DNA polymerases, as well as benchmark systems. All QM/MM calculations were performed on PM7/CHARMM36 level. The QM region comprised only nucleobases in a G•U(T) base pair of interest, with the QM-MM boundary placed at the glycosidic bonds. To select collective variables (CV) for the US calculations, we first optimized the wb-WC reaction path in G•U in gas phase using the nudged elastic band (NEB) method at B3LYP-D3BJ/def2-TZVP level (Figure 1B, C). The coordinate frames of this path became the reference for path collective variables (41) (pathCV), which were used to calculate the position on the path ($s$) of the wb-WC reaction in the US simulations. $s$ variable described the geometry change from wobble to WC, while a distance difference $hb = d$(O6–H3) – $d$(N1–H1) was used to describe proton transfers during the reaction. Selected CVs clearly distinguished all three local minima of the wb-WC reaction (Figure 1 C). Committor analysis was performed to verify the TS, revealing approximately equal partition between G•U wb and G•U* WC (Figure 1C). Using $s$ and $hb$ as CVs, and adding a boundary potential to prevent large deviations from path $z$, we collected 100–400 ps of QM/MM US trajectories for each of 48 US windows per each system (∼120 ns in total) and calculated PMF using weighted histogram analysis method (WHAM). We focused only on the relative energies and positions of local minima in the wb-WC reaction in our US simulations

### Kinetic modeling

Decoding rate constants for a codon-anticodon combination with a G•U mismatch in the first position were obtained earlier (42), but only in the low-fidelity buffer conditions, in which the accuracy of the initial selection was extremely low (1.1), contrasting the *in vivo* measurements (3–7). Therefore, for our kinetic modeling we selected decoding rate constants measured at high-fidelity conditions, which result in more realistic accuracy (43,44). In the absence of high-fidelity measurements corresponding to a G•U mismatch, we used the set of rate constants corresponding to the A•C mismatch in the first codon-anticodon position (43,44) (Supplementary Table S3). Rate constants $k_2$, $q_3^c$ and

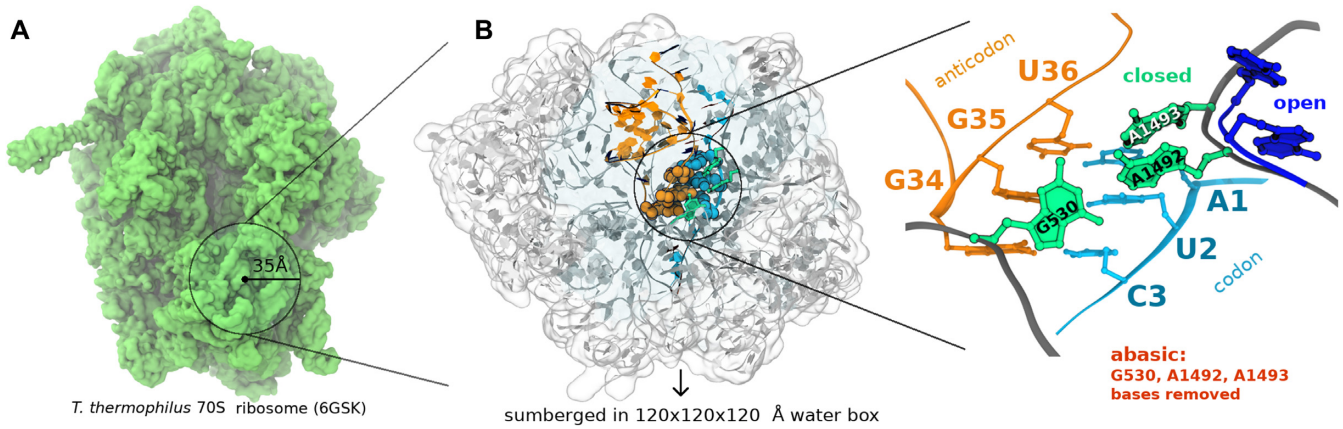


**Figure 2.** Preparation of the A-site models. (**A**) Visualization of the A-site model (right) and its position and size in the 70S ribosome (left). (**B**) Close-up view on the codon-anticodon helix in the decoding center. The 'closed' and 'open' A-site models differ by the conformations of the rRNA residues A1492 and A1493 (green and blue in the closed and open states, respectively). In the abasic A-site model nucleobases of A1492, A1493 and G530 were deleted, leaving only the sugar-phosphate backbone.

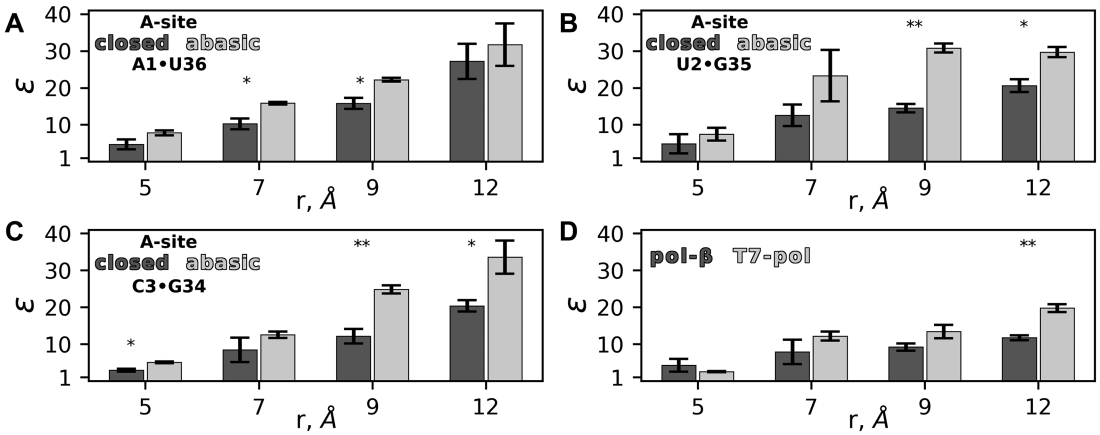


**Figure 3.** Dielectric constant ε of the base pair surroundings. (**A**–**C**) ε difference between 'closed' and 'abasic' A-site models around each codon-anticodon base pair (A1–U36, U2–G35 and C3-G34) measured in spheres of radius r. (**D**) ε difference between pol-β and T7-pol around the G•T base pair in the active site. Data are means ± SD ($n \geq 2$ replica per model). *$P < 0.05$; **$P < 0.01$ (Student's $t$-test)

$q_3^{nc}$ were assigned arbitrary. Numerical solutions of the kinetic model were obtained by numerical integration of ordinary differential equations (ODE) in Python. We describe the derivations of the analytical solutions in Appendix.

Further details on the methods are provided in the see Supplementary Methods.

## RESULTS

### Dielectric constant of base pair surroundings

Previously, it was hypothesized that closing of the ribosomal A site, via conformational change of the rRNA residues A1492, A1493 and G530, desolvates the codon-anticodon helix, increasing base-pairing selectivity by energy penalty from the lost H-bonds with water in mismatches (26,45). We hypothesized that desolvation may also bring an opposite contribution to the accuracy of G•U mismatch recognition. Our benchmark calculations demonstrate that the wb-WC reaction is exoergic in gas phase and very nonpolar implicit solvents (Supplementary Figure S1), corroborating previous computational studies (13,14,17,46). We reasoned that a decreased polarity of the closed ribosomal A site may be responsible for the apparent stabilization of the WC geometry of G•U observed earlier in structural studies (19–23).

We applied Kirkwood-Fröhlich formula (KFF) to qualitatively compare dielectric constant ε of the decoding site environment between the closed and abasic models. In the latter model, nucleobases of residues A1492, A1493 and G530 were deleted (Figure 2B). For all codon-anticodon positions, we observed a consistent decrease of ε in the closed model (Figure 3). However, only at some distance cutoffs and positions the difference was statistically significant, which can be explained by the limited length and replica numbers of the MD trajectories (Supplementary Figure S4). From our KFF calculations we conclude that the A site closing decreased ε around all three codon-anticodon positions.

To compare the effects of the ribosomal decoding site environment with the active site environment of DNA polymerases, we applied the same approach to low fidelity human DNA polymerase-β (pol-β) and and high-fidelity

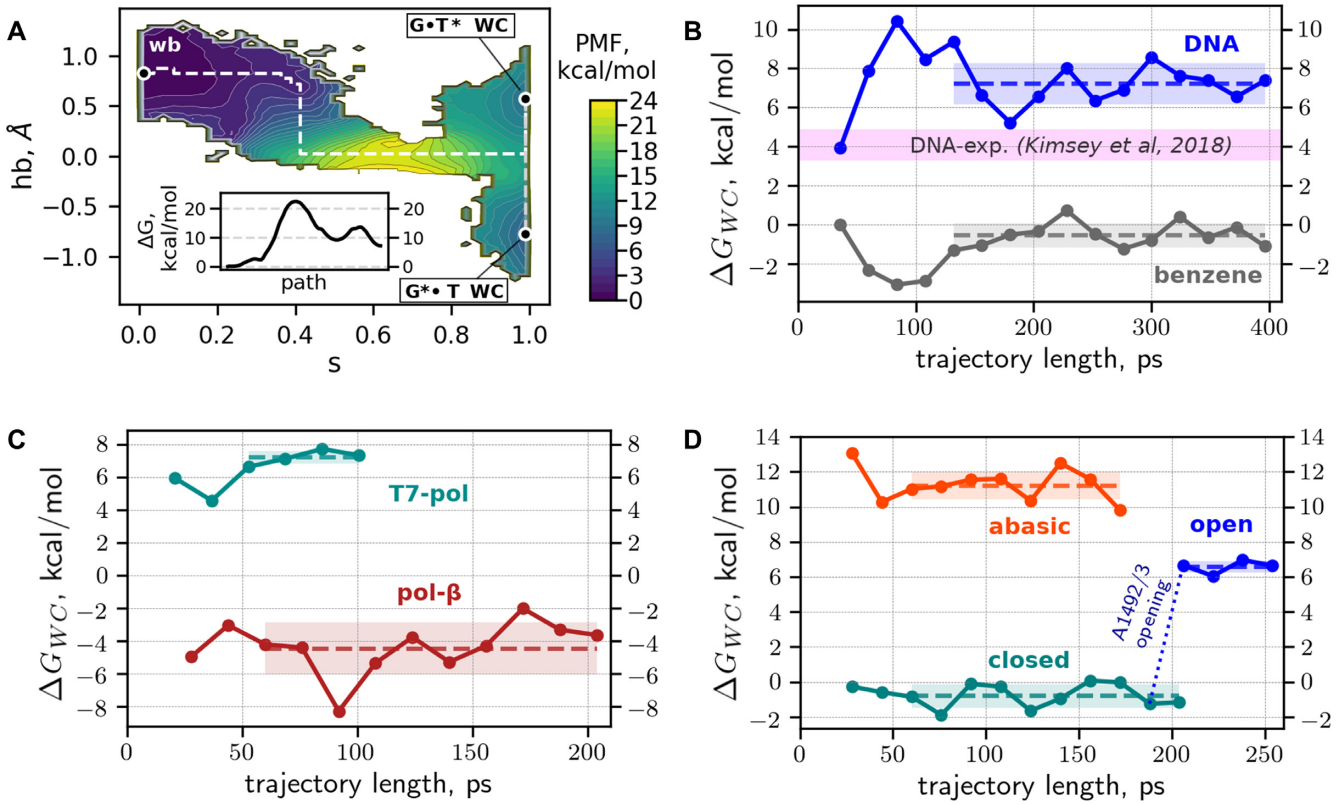


**Figure 4.** PMF from US calculations and its convergence. (**A**) US-derived PMF of the wb-WC reaction in the DNA heptamer system. Black circles denote the minima corresponding to wb and WC structures. White dashed line denotes the minimal free energy path of the wb-WC reaction, from G•T wb to G*•T WC. The inset shows the free energy profile along the minimal free energy path. PMFs from the other six systems are shown on the Supplementary Figure S5. (B–D) Convergence of the $\Delta G_{wc}$ in benchmark systems (**B**), DNA polymerases (**C**) and A-site models with G•U in the middle codon position (**D**). For each $\Delta G_{wc}$ trajectory, the shaded area and the dashed line of the corresponding color denote the standard deviation and mean of the set of batches selected for convergence evaluation, respectively. Magenta-shaded area in (B) denotes the range of $\Delta G_{wc}$ in DNA duplexes in solution measured by NMR (15).

DNA polymerase from T7 virus (T7-pol). The active site environment in pol-β was generally less polar than that of T7-pol, albeit the difference was statistically significant only at 12 Å cutoff (Figure 3). In sum, our KFF calculations revealed decreased polarity in pol-β compared to T7-pol and in the closed compared to the abasic A site of the ribosome.

### Free energy change of the wb-WC reaction estimated from QM/MM US calculations

To directly probe the effects of the studied environments on the wb-WC reaction energetics, we calculated the potential of mean force (PMF) of the wb-WC reaction using QM/MM umbrella sampling (US) calculations.

*Benchmark systems.* To verify our approach, we first applied it to benchmark systems: G•T base pair in the DNA heptamer duplex in water and a single G•T base pair in benzene. In all systems, the only quantitative property we derived from the PMFs was the total free energy change of the wb-WC reaction $\Delta G_{wc} = \Delta G$(G*•U WC) – $\Delta G$(G•U wb). $\Delta G_{wc}$ was used to evaluate the convergence of the US simulations by calculating PMF separately for consecutive batches of ∼20 ps. $\Delta G_{wc}$ trajectories of the benchmark systems and the converged values of $\Delta G_{wc}$ are shown in Figure 4B and Table 1, respectively. After initial fluctuations from 4 to 10 kcal/mol, $\Delta G_{wc}$ of the DNA model converged to 7.2 ± 1.1 kcal/mol. Our calculations overestimated the experimental $\Delta G_{wc}$ range of 3.3-4.9 kcal/mol (15), but were consistent with the recent computational result of ∼6 kcal/mol derived from (ωB97X-D3/6-311G**)/AMBER calculations (17). To assess the ability of our QM/MM setup to reproduce $\Delta G_{wc}$ dependence on ε observed in implicit solvent models, we applied it to a single G•T base pair in explicit benzene environment (ε = 2.3). $\Delta G_{wc}$ in benzene converged to –0.5 ± 0.6 kcal/mol, which is quantitatively consistent with the QM calculations in implicit solvent model of similar ε (Supplementary Figure S1). In both benchmark systems, the positions of the minima on the PMF were not qualitatively altered compared to the reference reaction path (Supplementary Figure S5, Figure 1C). From our benchmark calculations, we conclude that our setup is able to faithfully estimate the environmental effects on the wb-WC reaction, reaching qualitative agreement with experiments and quantitative agreement with higher levels of QM theory.

*DNA polymerases.* A recent QM/MM study has addressed the effects of the DNA polymerase λ environment on the wb-WC reaction (17). Here, we considered two other

**Table 1.** $\Delta G_{wc}$ values from the US simulations

| System | $\Delta G_{wc}$ |
|---|---|
| DNA | 7.2 ± 1.1 |
| Benzene | − 0.5 ± 0.6 |
| pol-β | − 4.4 ± 1.6 |
| T7-pol | 7.2 ± 0.4 |
| A site closed | − 0.7 ± 0.6 |
| A site abasic | 11.2 ± 0.9 |
| A site open | 6.6 ± 0.3 |

Mean $\Delta G_{wc}$ ± standard deviation from the regions selected for convergence estimation on Figure 4B–D, kcal/mol.

DNA polymerases: pol-β and T7-pol. In both pol-β and T7-pol, the wb-WC reaction was simulated in G•T base pair in closed state of a polymerase active site. $\Delta G_{wc}$ in pol-β converged to –4.4 ± 1.6 kcal/mol, indicating that the wb-WC reaction is largely exoergic (Figure 4C, Table 1), corroborating previous structural studies showing WC-like G•T base pair in the closed active site of pol-β (16). $\Delta G_{wc}$ in T7-pol converged to 7.2 ± 0.4 kcal/mol, quantitatively resembling $\Delta G_{wc}$ value in the DNA duplex (Figure 4B, C, Table 1). Comparing pol-β and T7-pol PMFs revealed a striking shift in the wb minimum position towards the higher value ($s \approx 0.2$) of the path variable in pol-β (Supplementary Figure S5). We attribute this shift to steric effects in pol-β; steric effects constrain a base pair geometry in the closed active site.

*Ribosomal A site.* We started the QM/MM US simulations of the A site effects on the wb-WC reaction in the second codon-anticodon U•G base pair in closed and abasic A site. $\Delta G_{wc}$ in the closed A site converged to –0.7 ± 0.6 kcal/mol, while in the abasic A site it converged to 11.2 ± 0.9 kcal/mol (Figure 4D, Table 1). However, the abasic model is only a rough approximation of the open state of the A site. To accurately estimate the effects of the rRNA nucleotides A1492 and A1493, we created the 'open' A-site model. In this model, harmonic restraints used in previous studies (25) were applied to change A1492 and A1493 conformations from *out* to *in* (Figure 2B). G530 remained in the closed-like conformation in the open model. These restraints were applied to the coordinates from the end part of the closed state US trajectories and maintained for ∼70 ps in each US window. $\Delta G_{wc}$ in the open A site converged to 6.6 ± 0.3 kcal/mol (Figure 4D, Table 1). While in the abasic model the position of the wb minimum was not qualitatively altered compared to the reference reaction path ($s = 0$), in the closed and open models it shifted to higher values of the path variable ($s \approx 0.1$), suggesting a slight destabilizing effect on the wobble geometry, likely exerted by G530 (Supplementary Figure S5). This effect was not enough to shift the wb-WC equilibrium towards WC in the open model, indicating a critical role of A1492 and A1493 residues.

In sum, using QM/MM US calculations, we demonstrate that the wb-WC reaction in G•U at the middle codon-anticodon position in the closed A site is weakly exoergic, while it was endoergic in the open and abasic models, corroborating structural studies (19–23). Decreased ε in the closed state of the decoding site can be one of the contributions to this effect. Shifts in the wb minimum position on the PMF in the closed and open models might indicate steric constraints on the base pair geometry. Similar effect was observed in pol-β active site, in which the wb-WC reaction was highly exoergic, in contrast to T7-pol, where the reaction equilibrium was not altered compared to the DNA duplex in water. However, further studies are needed to delineate the environmental contributions to the properties of the wb-WC reaction.

### The wb-WC reaction in the kinetic model of decoding in translation

The exoergic wb-WC reaction in closed A site requires reconsideration of the classical model of decoding in translation, as the previously suggested 'energy expenditure' of tautomerization (19–22) cannot be the basis of selectivity. Assuming geometric selection as the mechanism of base pair recognition (10,11,47), the error rate of G•U mismatches at the second codon position would be close to 1 with the exoergic wb-WC reaction in equilibrium settings. Instead, the error rate is around $10^{-3}$ (3–7). To reconcile the exoergic wb-WC reaction with the model of decoding, in closed A site the wb-WC reaction should be out of equilibrium, which requires its explicit introduction into the kinetic model of codon-anticodon decoding

Starting from the classical kinetic model of initial selection (Supplementary Figure S6), explained in details in (29), we introduced the wb-WC reaction by separating each near-cognate (nc) state into wb and WC states, leaving the cognate (c) branch unchanged (Figure 5A). Rate constants between nc-wb and nc-WC states corresponded to the rate constants of the wb-WC reaction in a given environment of the A site. To simplify the model for the sake of deriving analytical solutions, we assumed an instant wb-shifted wb-WC equilibrium in states *C*2 (initial binding; thus, the 'wb-WC equilibrium' here corresponds to the keto/enol equilibrium of G in aqueous environment) and *C*3 (codon recognition in the open A site). This approximation is well justified if, as suggested by our US calculations, the open state of the decoding site does not shift the equilibrium towards WC, and thus, the equilibration time is negligible. In this approximation, the wb-WC equilibrium in *C*2 does not affect the model and needed only for the model completeness and clear interpretation of the states. The wb-WC reaction in *C*4 (codon recognition in the closed A site) was modeled explicitly via its forward and reverse rate constants $k_f$, $k_r$. We used decoding rate constants measured previously at 20°C and high-fidelity buffer conditions (43,44). The nc decoding rate constants in this set corresponded not to a G•U mismatch, but to the A•C mismatch in the first position. This and other approximations preclude quantitative predictions of the error rate from the model. We analyzed effects of the wb-WC reaction on decoding that are not restricted to any particular combination of the decoding rate constants.

Inspired by the previously applied numerical kinetic model of selection in replication (15), we made the following assumptions: (i) nc-WC states are characterised with the same decoding rate constants as the cognate states, since the WC geometries are assumed to be indistinguishable for the decoding site; (ii) GTPase activation rate constant ($k_4$) in $C4^{wb}_{nc}$ state is set to zero, assuming that the major contri-

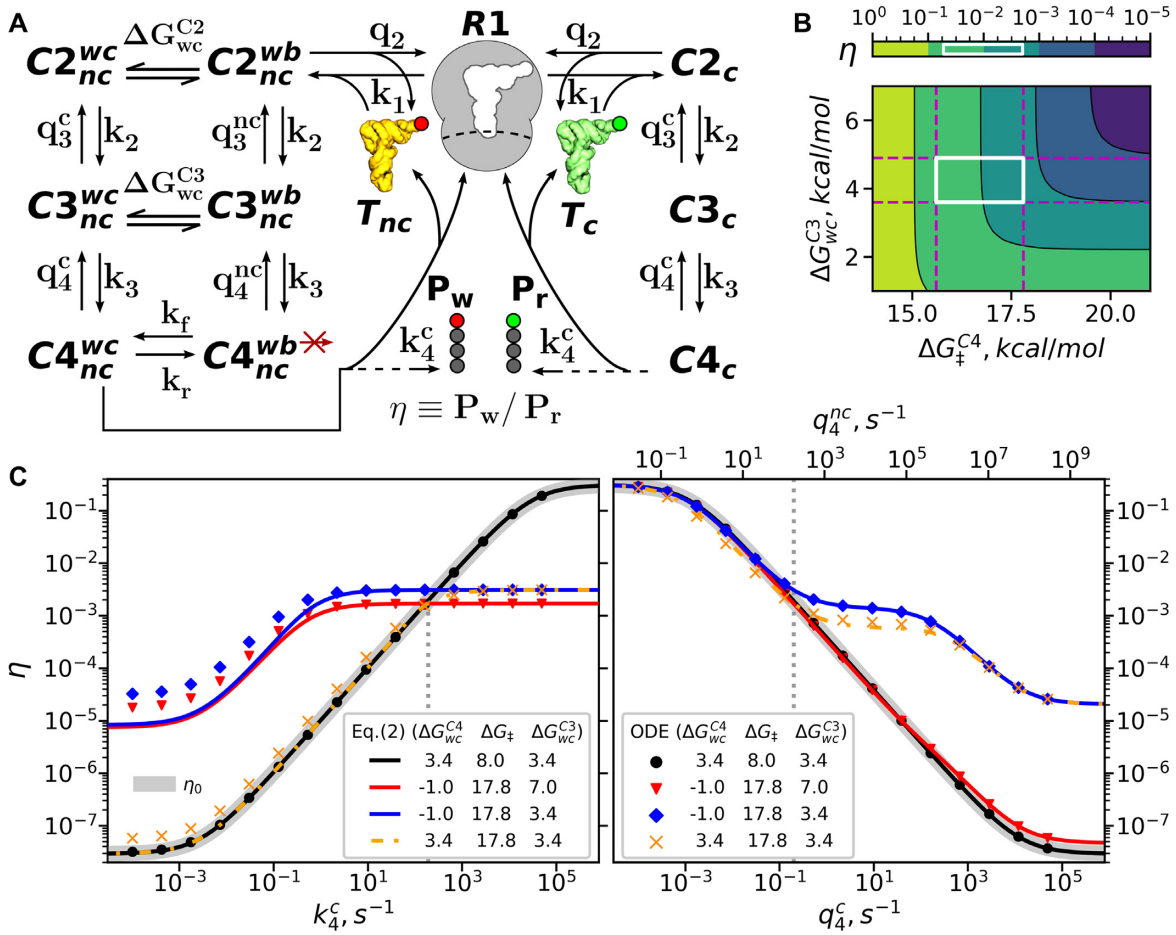


**Figure 5.** The wb-WC reaction in the kinetic model of decoding in translation. (**A**) Scheme of the kinetic model of the initial selection in decoding with incorporated wb-WC reaction (see Supplementary Figure S6 for detailed explanation of the decoding states and transitions). (**B**) $\eta$ as a function of $\Delta G_{\ddagger}$ and $\Delta G^{C3}_{wc}$, calculated using Equation (2) for $\Delta G^{C4}_{wc} = -1$ kcal/mol. The highlighted region corresponds to the experimental range of $\Delta G_{\ddagger}$ and $\Delta G_{wc}$ measured in RNA duplexes (15). (**C**) Error rate as a function of $k^c_4$ (left) and $q^{c/nc}_4$ (right), calculated by Eq. (A.1) for the classical system ($\eta_0$), and, for the specified wb-WC parameters, from numerical simulations (ODE) and by Equation (2). $\eta(q^{c/nc}_4)$ and $\eta_0(q^{c/nc}_4)$ were obtained by varying $q^c_4$ and $q^{nc}_4$ simultaneously at constant $q^c_4/q^{nc}_4$ ratio. $\eta_0(k^c_4)$ was also obtained at constant $k^c_4/k^{nc}_4$ ratio. $\Delta G$ units in the plot legend are kcal/mol. The gray dotted vertical lines denote experimental values of $k^c_4$ and $q^{c/nc}_4$. The deviation of predicted $\eta$ (Equation 2) from $\eta$(ODE) for low values of $k^c_4$ is due to increased $C3$ state population when the wb-WC reaction approaches equilibrium.

bution to this reaction rate comes from the nc-WC state, thus the contribution from nc-wb can be ignored; (iii) escape rate constants of the nc-wb states are taken from the classical nc states ($q^{nc}_i$), assuming that the cumulative escape rate is dominated by the nc-wb states, thus the nc→nc-wb rescaling can be neglected. Given assumptions (i) and (ii), the apparent near-cognate $k_4$:

$$k^{nc}_4 = P_{wc} k^c_4 \tag{1}$$

where $P_{wc}$ is the population of the $C4^{wc}_{nc}$ state—a function of the equilibrium WC populations in $C3$ and $C4$ states, and constraints imposed by the decoding rates on the reaction kinetics in $C4$. Using the equation for product concentration in a first-order reversible reaction, we derived (see Appendix) the equation for the error rate $\eta$ induced by the wb-WC reaction:

$$\eta = \frac{[C4_{nc}]}{[C4_c]} \left( P^{eq}_{WC} + (P^0_{WC} - P^{eq}_{WC}) \exp\left( -\frac{k_f + k_r}{k^c_4 + q^c_4} \right) \right) \tag{2}$$

where square brackets denote steady-state concentrations of the $C4_i$ states, $P^{eq}_{WC}$ is the equilibrium WC population in $C4$ for a given ($k_f$, $k_r$), and $P^0_{WC}$ is the equilibrium WC population in $C3$, kinetically partitioned into $C4$:

$$P^0_{WC} = P^{C3}_{WC} \frac{q^{nc}_4}{k^c_4 + q^c_4} \tag{3}$$

For convenience, the variables $P^{eq}_{WC}$, $P^{C3}_{WC}$ and $k_f$ are expressed below in terms of free energy differences $\Delta G^{C4}_{wc}$, $\Delta G^{C3}_{wc}$ and $\Delta G_{\ddagger}$ (activation free energy), respectively, using Eq. (A.5) and Eq. (A.6) at standard conditions.

First, we analyzed how $\Delta G^{C3}_{wc}$ and $\Delta G_{\ddagger}$ of the exoergic ($\Delta G^{C4}_{wc} = -1$ kcal/mol) wb-WC reaction affect $\eta$ predicted

from Equation (2) (Figure 5 B). For the range of $\Delta G_{wc}$ and $\Delta G_{\ddagger}$ in RNA duplexes in solution reported in (15), η overlaps with the *in vitro* error range of G•U mismatches (∼$10^{-3}$) (3–7). However, this result clearly cannot be interpreted as a validation of Equation (2). To test Equation (2), the decoding rate constants for G•U mismatches should be obtained, and $\Delta G_{\ddagger}$ in the decoding site accurately estimated.

Next, we addressed the dependence of η on the decoding rate constants $k_4^c$ (GTPase activation) and $q_4^{c/nc}$ (cognate and near-cognate escape rate constants of the *C*4 state). These rate constants determine the deviation of the open↔closed transition in decoding from equilibrium conditions, thereby affecting the deviation of the actual accuracy from the intrinsic accuracy (28,30). Equation (2) suggests two possible kinetic regimes of the wb-WC reaction for the case of positive $\Delta G_{wc}^{C3}$. In the 'fast' regime ($(k_f + k_r) >> (k_4^c + q_4^c)$), the kinetics of the wb-WC reaction and $P_{WC}^{C3}$ are irrelevant, and $P_{wc} = P_{WC}^{eq}$. For $P_{WC}^{eq} = k_4^{nc}/k_4^c$ (corresponds to $\Delta G_{wc}^{C4} \approx 3.4$ kcal/mol), the fast regime is equivalent to the classical error rate $\eta_0$ (Eq. (A.1)) in both $k_4^c$ and $q_4^{c/nc}$ dependencies, which is confirmed by numerical calculations (Figure 5 C).

A more intriguing and relevant, given the predicted wb-WC parameters, is the 'slow' regime, where the wb-WC equilibrium in *C*4 is shifted to the WC geometry ($k_f > k_r$), but the kinetics of the reaction is limited by the decoding rates. To visualize the 'slow' regime, we chose wb-WC parameters ($\Delta G_{wc}^{C4} = -1$ kcal/mol, $\Delta G_{\ddagger} = 17.8$ kcal/mol), which approximately correspond to the predicted $\Delta G_{wc}^{C4}$ (Table 1) and experimentally observed $\Delta G_{\ddagger}$ in RNA duplexes (15). For $k_f < k_4^c$, Equation (2) predicts a virtually flat $\eta(k_4^c)$ curve, as confirmed by numerical solutions (Figure 5 C). $\Delta G_{wc}^{C3}$ affects the offset of the curves (Figure 5 C). The reason for the flat $\eta(k_4^c)$ curve in the 'slow' regime is derived in Appendix and visualized in Supplementary Figure S7. The linear approximations of $P_{wc}$ and $\frac{[C4_{nc}]}{[C4_c]}$ inversely depend on $k_4^c$, which cancels in η. Equilibration of the exoergic wb-WC reaction in *C*4 with unfavorable contribution to the accuracy counteracts the equilibration of the open-closed transition with favorable contribution, thereby constraining the decoding accuracy.

In $\eta(q_4^{c/nc})$, the 'slow' kinetic regime with a very low $P_{WC}^{C3}$ ($\Delta G_{wc}^{C3} = 7.0$ kcal/mol) closely follows $\eta_0$. However, the same kinetic regime with $\Delta G_{wc}^{C3} = 3.4$ kcal/mol results in an almost flat curve for $q_4^{nc} > k_4^c > q_4^c$ (Figure 5C). This effect is explained by the kinetic partitioning term in $P_{WC}^{C3}\frac{q_4^{nc}}{k_4^c+q_4^c}$, which grows proportionally with $q_4^{nc}$ and cancels out the classical equilibration process. Lastly, we compared this kinetic regime comprising exoergic wb-WC reaction to the regime with the same $\Delta G_{\ddagger}$ and $\Delta G_{wc}^{C3}$, but with $\Delta G_{wc}^{C4} = 3.4$ kcal/mol. In the latter regime, $\eta(q_4^{c/nc})$ is similar to the former, as it is constrained by $P_{WC}^0$ (Figure 5C). However, $\eta(k_4^c)$ of the last regime is only constrained at $k_4^c \gtrsim \frac{k_f}{P_{WC}^0}$; at lower $k_4^c$ the reaction is at equilibrium, and η follows the classical curve (Figure 5C).

In sum, Equation (2), confirmed by numerical calculations, predicts that a high $\Delta G_{\ddagger}$ of the weakly exoergic wb-WC reaction in *C*4 causes a flat $\eta(k_4^c)$ curve for $k_f < k_4^c$, while high $P_{WC}^{C3}$ causes an almost flat $\eta(q_4^{c/nc})$ for $q_4^{nc} > k_4^c > q_4^c$. Two potentially independent parameters of the wb-WC reaction, in the vicinity of their experimentally obtained and/or predicted values, constrain the decoding accuracy of the G•U mismatch recognition in translation for a biologically-relevant range of the decoding rate constants. It is also worth noting that our model does not contradict the experiments with varying $Mg^{2+}$ concentrations. In their experiments, Zhang et al (48) reported that $Mg^{2+}$ concentration affects $q_2$, resulting in linear rate-accuracy trade-offs. For all considered wb-WC parameters, our model also predicts linear rate-accuracy trade-offs by varying $q_2$ (Supplementary Figure S8).

## DISCUSSION

Structural studies revealed stabilization of the WC geometry of the G•U base pairs in closed ribosomal A site (19–23), but this stabilization has been lacking a physichochemical explanation, and moreover, its implications for the decoding mechanism remained unexplored. The structural studies propose the term 'energy expenditure' of tautomerization to explain the discrimination capacity of the ribosome (19). However, this model does not explain how the WC geometry of G•U is observed at equilibrium conditions. These observations (19,20) were opposed by a previous computational study addressing G•U* WC in the first codon position using classical force fields approximation, which concluded that it is not stabilized in the closed A site (25). Satpati and Åqvist conclude the same for the first codon position, but suggest that G*•U WC is stabilized in the second codon position by 4 kcal/mol compared to G•U wb (24). The authors argue that such stabilization would not lead to decoding errors due to largely unfavorable interaction energy with the closed decoding site compared to the cognate complex (24). This model opposes the role of geometric selection in translation (10,11,47). Moreover, enol tautomers of G and U in the study of Satpati and Åqvist (24) were not parameterized, making an interpretation of their corresponding free energy calculation results not possible.

Here, we applied the QM/MM US approach, in which the wb-WC reaction was simulated explicitly on semiempirical level of QM theory PM7. We demonstrated the validity of this approach by performing benchmark calculations, which reached a qualitative agreement with experiments and a quantitative agreement with high levels of QM theory. This renders our approach of calculating $\Delta G_{wc}$ more accurate compared to previously published study (25). The latter greatly overestimated keto/enol tautomerization energy of uracil in water compared to the experimental data and high levels of QM theory (38 kcal/mol versus 10 kcal/mol, respectively) (25). We applied the QM/MM US approach to study the effects of the closed active site environment of two DNA polymerases, pol-β and T7-pol, on the wb-WC reaction in G•T. The wb-WC reaction in pol-β is exoergic, corroborating a previous structural study (16). The wb-WC reaction equilibrium in T7-pol is not affected compared to the DNA duplex in water. In line with our results, the WC geometry of G•T in T7-pol is not observed in any structural studies. Of note, this cannot be used as validation.

Finally, our QM/MM US simulations show that the wb-WC reaction at the middle codon position in the open A site is endoergic, but closure of the rRNA residues A1492 and A1493 shifts the equilibrium towards the WC geometry. This effect may, at least in part, be explained by the decreased polarity of the closed decoding site.

The model of the ribosomal A site used in our simulations is a simplification of the real conditions that may affect the wb-WC reaction properties during decoding. The model includes the following major simplifications: (i) only a part of the ribosome is considered, with an outer shell that was restrained during all simulations; (ii) $Mg^{2+}$ ions are absent; (iii) the 'open' model differed from the 'closed' one only by conformations of A1492 and A1493 residues. However, distant structural elements and their change during the 30S subunit closure are unlikely to dramatically affect the wb-WC reaction. Also, $Mg^{2+}$ ions have not been reported to directly interact with the second codon-anticodon position. The role of G530, which is unaffected in the 'open' model, is not fully explored in our study. It's 'latching' is suggested to be critical upon 30S closure (23), but represents a more challenging modeling task than applied here to A1492 and A1493, thus, requires future in-depth studies. It also remains to be investigated of how the wb-WC reaction is affected at the first and third codon-anticodon positions. Overall, our simplified model of the A site captures major effects of the 'open/closed' conformational changes on the wb-WC reaction at the middle codon position.

The wb-WC reaction switches its equilibrium upon conformational changes of the decoding site. It violates the rigid substrate approximation of the classical induced-fit model of codon–anticodon decoding, calling for a new model. Inspired by the suggestion of the out-of-equilibrium exoergic wb-WC reaction during DNA replication, expressed by Hovorun and coworkers (14), we modeled such a scenario in the codon-anticodon decoding. We introduced the wb-WC reaction into the classical model of initial selection in translation as a 'flexible substrate' and derived analytical solutions for the decoding accuracy governed by the wb-WC reaction. The analytical solutions fit the numerical calculations and explain the error rate of G•U mismatch decoding in terms of the wb-WC reaction parameters in 'open' and 'closed' states of the A site. For $\Delta G_{\ddagger}$ around the experimental range (∼18 kcal/mol), the exoergic wb-WC reaction is kinetically limited by the decoding rates, as predicted for the case of DNA polymerases (14). In contrast to two models (19,24), our model enables reconciling structural studies with the mechanism of codon-anticodon decoding. The model is consistent with the WC geometry of G•U reported in structural studies at equilibrium conditions ($P_{wc} \to 1$ as $k_4^c \to 0$) (19–23), and it explains the discriminating capacity of the working ribosome at non-equilibrium conditions ($k_4^c >> 0\,s^{-1}$). Thus, in our model the mechanism of decoding is no longer at odds with the geometric selection model (10,11,47)

A number of approximations used in our model restrict its application to only qualitative interpretation. Quantitative predictions would require a set of decoding rate constants for a G•U mismatch and the accurate $\Delta G_{wc}^{C3}$ and $\Delta G_{\ddagger}$, which are currently not available. Although the assumptions (i) and (ii), used to build the kinetic model, were already successfully employed in the study that indirectly confirmed the tautomeric hypothesis in DNA replication (15), validity of these assumptions in codon-anticodon decoding has yet to be proven. Moreover, an accurate error rate prediction would require expanding the model to include possible proofreading steps; in our study it was restricted to the initial selection only.

Theoretical studies based on the rigid substrate approximation have concluded that decoding in translation has been evolutionary optimized for higher rate at the expense of accuracy (28,30,49). Since the rate constants affecting the 'open-closed' transition (e.g. $k_4^c$) are not rate-limiting in translation elongation (30), the advantage of the high $k_4^c$ value (∼200 $s^{-1}$ *in vitro* (43)) remained unclear. Our model suggests an alternative explanation to the apparently suboptimal decoding. It reveals possible constraints on the codon-anticodon decoding accuracy imposed by the slow exoergic wb-WC tautomerization reaction, whereby the equilibration of the reaction counteracts the equilibration of the 'open-closed' transition. We emphasize that these predicted constraints are not restricted to specific combinations of rate constants, and arise from introducing the wb-WC reaction explicitly into the model. These constraints suggest that the ribosome is incapable of substantially reducing the error rate of G•U mismatch decoding by equilibrating the 'open-closed' transition, potentially explaining the evolutionary persistence of this dominant error hotspot.

We suggest that the model derived here for codon-anticodon decoding in translation can also be applied for base pair recognition in DNA replication (see Supplementary Discussion). Our model generalizes the classical 'induced-fit' model, as the former reduces to the latter in the equilibrium limit in the 'flexible substrate' (illustrated on Supplementary Figure S9). Thus, we believe that our theoretical and computational results are a step towards a more complete model of substrate recognition.

## DATA AVAILABILITY

Optimized base pair geometries, coordinate frames of the NEB trajectory, initial system topologies and coordinates in the QM/MM US simulations, as well an example setup of the QM/MM US simulations are available on Figshare: https://doi.org/10.6084/m9.figshare.c.5178074.v1.

## CODE AVAILABILITY

Python and Tcl scripts for the umbrella sampling setup, dielectric permittivity calculations, kinetic modeling and visualizations are available on GitHub: https://github.com/andkaz/wbwc_paper.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

This paper is dedicated to the memory of Prof. Dmytro Hovorun, who profoundly advanced the field of nucleobase

tautomerism and together with his colleagues co-discovered the wb-WC reaction. As the head of the Department of molecular and quantum biophysics in Kyiv (Ukraine), Prof. Hovorun inspired many theoretical biophysicists, including A.K. We also thank the Norddeutscher Verbund für Hoch- und Höchstleistungsrechnen (HLRN) [hhc00026] and the Hummel cluster at the University of Hamburg.

## FUNDING

Deutscher Akademischer Austauschdienst (DAAD) [to A.K.]; EU Horizon 2020 program (Marie Skłodowska-Curie grant) [764591 to Z.I.]; Funding for open access charge: EU Horizon 2020 program (Marie Skłodowska-Curie grant) [764591].
*Conflict of interest statement.* None declared.

## REFERENCES

1. Santos,M., Pereira,P.M., Varanda,A.S., Carvalho,J., Azevedo,M., Mateus,D.D., Mendes,N., Oliveira,P., Trindade,F., Pinto,M.T. *et al.* (2018) Codon misreading tRNAs promote tumor growth in mice. *RNA Biol.*, **15**, 773–786.
2. Ke,Z., Mallik,P., Johnson,A.B., Luna,F., Nevo,E., Zhang,Z.D., Gladyshev,V.N., Seluanov,A. and Gorbunova,V. (2017) Translation fidelity coevolves with longevity. *Aging Cell*, **16**, 988–993.
3. Pernod,K., Schaeffer,L., Chicher,J., Hok,E., Rick,C., Geslain,R., Eriani,G., Westhof,E., Ryckelynck,M. and Martin,F. (2020) The nature of the purine at position 34 in tRNAs of 4-codon boxes is correlated with nucleotides at positions 32 and 38 to maintain decoding fidelity. *Nucleic Acids Res.*, **48**, 6170–6183.
4. Garofalo,R., Wohlgemuth,I., Pearson,M., Lenz,C., Urlaub,H. and Rodnina,M.V. (2019) Broad range of missense error frequencies in cellular proteins. *Nucleic Acids Res.*, **47**, 2932–2945.
5. Mordret,E., Dahan,O., Asraf,O., Rak,R., Yehonadav,A., Barnabas,G.D., Cox,J., Geiger,T., Lindner,A.B. and Pilpel,Y. (2019) Systematic detection of amino acid substitutions in proteomes reveals mechanistic basis of ribosome errors and selection for translation fidelity. *Mol. Cell*, **75**, 427–441.
6. Manickam,N., Nag,N., Abbasi,A., Patel,K. and Farabaugh,P.J. (2014) Studies of translational misreading in vivo show that the ribosome very efficiently discriminates against most potential errors. *RNA*, **20**, 9–15.
7. Zhang,Z., Shah,B. and Bondarenko,P.V. (2013) G/U and certain wobble position mismatches as possible main causes of amino acid misincorporations. *Biochemistry*, **52**, 8165–8176.
8. Joshi,K., Cao,L. and Farabaugh,P.J. (2019) The problem of genetic code misreading during protein synthesis. *Yeast*, **36**, 35–42.
9. Kimsey,I.J., Petzold,K., Sathyamoorthy,B., Stein,Z.W. and Al-Hashimi,H.M. (2015) Visualizing transient Watson-Crick-like mispairs in DNA and RNA duplexes. *Nature*, **519**, 315–320.
10. Khade,P.K., Shi,X. and Joseph,S. (2013) Steric complementarity in the decoding center is important for tRNA selection by the ribosome. *J. Mol. Biol.*, **425**, 3778–3789.
11. Westhof,E., Yusupov,M. and Yusupova,G. (2014) Recognition of Watson-Crick base pairs: constraints and limits due to geometric selection and tautomerism. *F1000Prime Reports*, **6**, 19.
12. Watson,J.D. and Crick,F. H.C. (1953) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.
13. Brovarets,O.O. and Hovorun,D.M. (2009) Physicochemical mechanism of the wobble DNA base pairs Gu·Thy and Ade·Cyt transition into the mismatched base pairs Gua*·Thy and Ade·Cyt* formed by the mutagenic tautomers. *Ukrain. Bioorg. Acta 2*, **7**, 12–18.
14. Brovarets,O.O. and Hovorun,D.M. (2015) How many tautomerization pathways connect Watson-Crick-like G*·T DNA base mispair and wobble mismatches? *J. Biomol. Struct. Dyn.*, **33**, 2297–2315.
15. Kimsey,I.J., Szymanski,E.S., Zahurancik,W.J., Shakya,A., Xue,Y., Chu,C.C., Sathyamoorthy,B., Suo,Z. and Al-Hashimi,H.M. (2018) Dynamic basis for dG· dT misincorporation via tautomerization and ionization. *Nature*, **554**, 195–201.
16. Koag,M.C., Nam,K. and Lee,S. (2014) The spontaneous replication error and the mismatch discrimination mechanisms of human DNA polymerase β. *Nucleic Acids Res.*, **42**, 11233–11245.
17. Li,P., Rangadurai,A., Al-Hashimi,H.M. and Hammes-Schiffer,S. (2020) Environmental effects on guanine-thymine mispair tautomerization explored with quantum mechanical/molecular mechanical free energy simulations. *J. Am. Chem. Soc.*, **142**, 11183–11191.
18. Maximoff,S.N., Kamerlin,S. C.L. and Florian,J. (2017) DNA polymerase λ 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.
19. Demeshkina,N., Jenner,L., Westhof,E., Yusupov,M. and Yusupova,G. (2012) A new understanding of the decoding principle on the ribosome. *Nature*, **484**, 256–259.
20. Demeshkina,N., Jenner,L., Westhof,E., Yusupov,M. and Yusupova,G. (2013) New structural insights into the decoding mechanism: translation infidelity via a G.U pair with Watson-Crick geometry. *FEBS Lett.*, **587**, 1848–1857.
21. Rozov,A., Demeshkina,N., Westhof,E., Yusupov,M. and Yusupova,G. (2015) Structural insights into the translational infidelity mechanism. *Nat Commun.*, **6**, 7251.
22. Rozov,A., Wolff,P., Grosjean,H., Yusupov,M., Yusupova,G. and Westhof,E. (2018) Tautomeric G•U pairs within the molecular ribosomal grip and fidelity of decoding in bacteria. *Nucleic Acids Res.*, **46**, 7425–7435.
23. Loveland,A.B., Demo,G., Grigorieff,N. and Korostelev,A.A. (2017) Ensemble cryo-EM elucidates the mechanism of translation fidelity. *Nature*, **546**, 113–117.
24. Satpati,P. and Aqvist,J. (2014) Why base tautomerization does not cause errors in mRNA decoding on the ribosome. *Nucleic Acids Res*, **42**, 12876–12884.
25. Zeng,X., Chugh,J., Casiano-Negroni,A., Al-Hashimi,H.M. and Brooks Iii,C.L. (2014) Flipping of the ribosomal A-site adenines provides a basis for tRNA selection. *J. Mol. Biol.*, **426**, 3201–3213.
26. Ogle,J.M., Murphy IV,F.V., Tarry,M.J. and Ramakrishnan,V. (2002) Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell*, **111**, 721–732.
27. Johansson,M., Lovmar,M. and Ehrenberg,M. (2008) Rate and accuracy of bacterial protein synthesis revisited. *Curr. Opin. Microbiol.*, **11**, 141–147.
28. Savir,Y. and Tlusty,T. (2013) The ribosome as an optimal decoder: a lesson in molecular recognition. *Cell*, **153**, 471–479.
29. Pavlov,M.Y. and Ehrenberg,M. (2018) Substrate-induced formation of ribosomal decoding center for accurate and rapid genetic code translation. *Ann. Rev. Biophys.*, **47**, 525–548.
30. Wohlgemuth,I., Pohl,C., Mittelstaet,J., Konevega,A.L. and Rodnina,M.V. (2011) Evolutionary optimization of speed and accuracy of decoding on the ribosome. *Philos. Trans. Roy. Soc. B*, **366**, 2979–2986.
31. Stewart,J.J. (2013) Optimization of parameters for semiempirical methods. VI. More modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.*, **19**, 1–32.
32. Neese,F. (2018) Software update: the ORCA program system, version 4.0. *WIRES Comput. Mol. Sci.*, **8**, e1327.
33. Freudenthal,B.D., Beard,W.A., Shock,D.D. and Wilson,S.H. (2013) Observing a DNA polymerase choose right from wrong. *Cell*, **154**, 157.
34. Doublié,S., Tabor,S., Long,A.M., Richardson,C.C. and Ellenberger,T. (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature*, **391**, 251–258.
35. Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
36. Huang,J. and Mackerell,A.D. (2013) CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.*, **34**, 2135–2145.
37. Best,R.B., Zhu,X., Shim,J., Lopes,P.E.M., Mittal,J., Feig,M. and MacKerell,A.D. (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ1 and χ2 dihedral angles. *J. Chem. Theory. Comput.*, **8**, 3257–3273.

38. Denning,E.J., Priyakumar,U.D., Nilsson,L. and Mackerell,A.D. (2011) Impact of 2′-hydroxyl sampling on the conformational properties of RNA: update of the CHARMM all-atom additive force field for RNA. *J. Comput. Chem.*, **32**, 1929–1943.
39. Phillips,J.C., Hardy,D.J., Maia,J.D., Stone,J.E., Ribeiro,J.V., Bernardi,R.C., Buch,R., Fiorin,G., Hénin,J., Jiang,W. *et al.* (2020) Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.*, **153**, 044130.
40. Neumann,M. (1984) The dielectric constant of water. Computer simulations with the MCY potential. *J. Chem. Phys.*, **82**, 5663–5672.
41. Branduardi,D., Gervasio,F.L. and Parrinello,M. (2007) From A to B in free energy space. *J. Chem. Phys.*, **126**, 054103.
42. Pape,T., Wintermeyer,W. and Rodnina,M.V. (2000) Conformational switch in the decoding region of 16S rRNA during aminoacyl-tRNA selection on the ribosome. *Nat. Struct. Biol.*, **7**, 104–107.
43. Rudorf,S., Thommen,M., Rodnina,M.V. and Lipowsky,R. (2014) Deducing the kinetics of protein synthesis in vivo from the transition rates measured in vitro. *PLoS Comput. Biol.*, **10**, e1003909.
44. Gromadski,K.B., Daviter,T. and Rodnina,M.V. (2006) A uniform response to mismatches in codon-anticodon complexes ensures ribosomal fidelity. *Mol. Cell*, **21**, 369–377.
45. Satpati,P., Sund,J. and Åqvist,J. (2014) Structure-based energetics of mRNA decoding on the ribosome. *Biochemistry*, **53**, 1714–1722.
46. Nomura,K., Hoshino,R., Shimizu,E., Hoshiba,Y., Danilov,V.I. and Kurita,N. (2013) 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.
47. Schrode,P., Huter,P., Clementi,N. and Erlacher,M. (2016) Atomic mutagenesis at the ribosomal decoding site. *RNA Biol.*, **14**, 104–112.
48. Zhang,J., Pavlov,M.Y. and Ehrenberg,M. (2018) Accuracy of genetic code translation and its orthogonal corruption by aminoglycosides and $Mg^{2+}$ ions. *Nucleic Acids Res.*, **46**, 1362–1374.
49. Mallory,J.D., Kolomeisky,A.B. and Igoshin,O.A. (2019) Trade-offs between error, speed, noise, and energy dissipation in biological processes with proofreading. *J. Phys. Chem. B*, **123**, 4718–4725.