Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

J Am Chem Soc. Author manuscript; available in PMC 2021 June 24.

Published in final edited form as:

J Am Chem Soc. 2020 June 24; 142(25): 11183–11191. doi:10.1021/jacs.0c03774.

Environmental Effects on Guanine-Thymine Mispair Tautomerization Explored with Quantum Mechanical/Molecular Mechanical Free Energy Simulations

Pengfei Li[1], **Atul Rangadurai**[2], **Hashim M. Al-Hashimi**[2], **Sharon Hammes-Schiffer**[1,*]

[1]Department of Chemistry, Yale University, 225 Prospect Street, New Haven, CT 06520

[2]Department of Biochemistry, Duke University, Durham, NC, 27710

Abstract

DNA bases can adopt energetically unfavorable tautomeric forms that enable the formation of Watson-Crick-like (WC-like) mispairs, which have been proposed to give rise to spontaneous mutations in DNA and misincorporation errors in DNA replication and translation. Previous NMR and computational studies have indicated that the population of WC-like guanine-thymine (G-T) mispairs depends on the environment, such as the local nucleic acid sequence and solvation. To investigate these environmental effects, herein G-T mispair tautomerization processes are studied computationally in aqueous solution, in A-form and B-form DNA duplexes, and within the active site of a DNA polymerase λ variant. The wobble G-T (wG-T), WC-like G-T*, and WC-like G*-T forms are considered, where * indicates the enol tautomer of the base. The minimum free energy paths for the tautomerization from the wG-T to the WC-like G-T* and from the WC-like G-T* to the WC-like G*-T are computed with mixed quantum mechanical/molecular mechanical (QM/MM) free energy simulations. The reaction free energies and free energy barriers are found to be significantly influenced by the environment. The wG-T→G-T* tautomerization is predicted to be endoergic in aqueous solution and the DNA duplexes but slightly exoergic in the polymerase, with Arg517 and Asn513 providing electrostatic stabilization of G-T*. The G-T*→G*-T tautomerization is also predicted to be slightly more thermodynamically favorable in the polymerase relative to a DNA duplex. These simulations are consistent with an experimentally driven kinetic misincorporation model suggesting that G-T mispair tautomerization occurs in the ajar polymerase conformation or concertedly with the transition from the ajar to the closed polymerase conformation. Furthermore, the order of the associated two proton transfer reactions is predicted to be different in the polymerase than in aqueous solution and the duplexes. These studies highlight the impact of the environment on the thermodynamics, kinetics, and fundamental mechanisms of G-T mispair tautomerization, which plays a role in a wide range of biochemically important processes.

*Corresponding author sharon.hammes-schiffer@yale.edu.

The authors declare no competing financial interest.

Supporting Information

Computational details and supplementary figures and tables (PDF)

Equilibrated aqueous system with wG-T mispair (PDB)

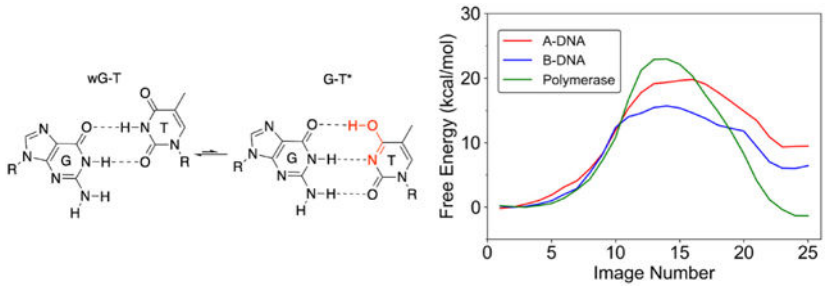
Equilibrated A-DNA duplex system with wG-T mispair (PDB)

Equilibrated B-DNA duplex system with wG-T mispair (PDB)

Equilibrated variant of polymerase λ with wG-T mispair (PDB)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Graphical Abstract



Introduction

In DNA, typically the bases G, C, A, and T, which correspond to guanine, cytosine, adenine, and thymine, respectively, form the canonical Watson-Crick (WC) base pairs G-C and A-T. In some cases, however, the bases can adopt energetically less favorable tautomeric forms, enabling the formation of WC-like mispairs, such as G*-T, G-T*, A*-C, and A-C*, where * denotes the energetically unfavorable enolic (for G and T) or imino (for A and C) tautomeric forms of the bases. By mimicking the canonical Watson-Crick geometry,[1] these mispairs could potentially evade fidelity checkpoints and lead to spontaneous mutations and misincorporation errors in DNA replication and translation.[2-5] Misincorporation can also arise through WC-like anionic G-T^{-} mispairs that form through ionization of T,[6] although this pathway is thought to play a minor role at physiological pH.[7] Indeed, WC-like mispairs have been observed in the active sites of DNA polymerases[8-10] and ribosomes[11-13] in catalytically competent conformations. Mutations from replicative errors are proposed to account for a large fraction of mutations that contribute to various forms of cancer.[14] Moreover, the tautomeric forms of bases are proposed to play important roles in DNA damage and repair,[15-16] as well as nucleic acid recognition,[17-18] chemical modifications,[19-22] and catalysis.[23] Thus, understanding the mechanism and thermodynamics of tautomerization is important for a wide range of biomedical applications.

Characterizing the tautomeric forms of the bases experimentally is challenging because of the low abundance of these species.[24-25] Recently, Al-Hashimi and co-workers studied the shortlived, low abundance WC-like G-T and G-U mispairs in B-DNA and A-RNA duplexes using NMR relaxation dispersion.[7, 15, 26] Their results indicated that at near neutral pH, wobble G-T mispairs (wG-T) are in equilibrium with a low population (~10^{-3}) of tautomeric WC-like G-T mispairs composed of the G*-T and G-T* forms in rapid exchange with each other (Figure 1).[7] They also found evidence suggesting that the frequencies of replication and translational errors are influenced by the free energy associated with forming the tautomeric WC-like mispairs.[7]

The tautomerization process of G-T mispairs has been shown to depend significantly on the environment. NMR studies indicated that the prevalence of the wobble and WC-like structures for G-T mispairs in B-DNA duplexes is sequence dependent, which could potentially translate into certain sequences being more prone to replicative errors than others.[7] In addition, no evidence for exchange to a WC-like mispair was found in G-U

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

mispairs when placed next to flexible non-duplex like environments such as bulges and apical loops.[7] X-ray crystallographic studies illustrated that the G-T mispair adopts a wobble structure in the context of A-DNA[27] and B-DNA.[28] Furthermore, crystal structures show that the G-T mispair can adopt different configurations depending on polymerase conformation (Table S2).[9-10, 29-31] For example, the nascent G-T mispair adopts the WC-like structure in a human DNA polymerase λ variant[9] and in DNA polymerase β[10] when it is in the closed state. In contrast, the nascent G-T mispair adopts a wobble structure in a DNA polymerase BF1 variant when it is in the ajar state,[29] in variants of the RB69 polymerase, [32-33] and in Y-family DNA polymerases η[30] and DPO4.[34] In addition, a systematic crystallographic study of a DNA polymerase RB69 variant illustrated that the local microenvironment plays a significant role in modulating the equilibrium between wobble and WC-like G-T mispairs following incorporation into the DNA.[31] Computational studies have also indicated that the kinetics and thermodynamics of tautomerization of isolated G-T mispairs can be influenced by different dielectric continuum solvent environments.[35-36] In addition, recently the G-T mispair tautomerization has been studied in a DNA polymerase λ variant using the free energy perturbation (FEP) approach, suggesting that the G*-T form is the most prevalent form of the mispair in this polymerase.[37]

Despite these earlier studies, a complete understanding of the environmental effects on G-T mispair tautomerization has still not been established. Calculations of the tautomerization free energies have not yet been performed in the context of DNA/RNA duplexes, for which experimental data are available, to serve as a benchmark for the calculations. Furthermore, although it has been proposed that polymerases can greatly alter tautomerization thermodynamics relative to DNA,[8] quantitative estimates of this effect remain elusive. To further elucidate these environmental effects, we examined the tautomerization process of G-T mispairs in aqueous solution, an A-DNA duplex, a B-DNA duplex, and a DNA polymerase λ variant using mixed quantum mechanical/molecular mechanical (QM/MM) free energy simulations based on the finite-temperature string method with umbrella sampling.[38-39] The free energy profiles corresponding to the G-T tautomerization process were generated for each of these systems, and the impact of electrostatic interactions, duplex geometry, and polymerase environment was analyzed. This work highlights the significance of the environmental effects on the G-T mispair tautomerization process and provides fundamental insights into this critical biochemical reaction.

Methods

We performed QM/MM free energy simulations of G-T mispairs in aqueous solution, A-form and B-form DNA duplexes, and a DNA polymerase λ variant. The crystal structure associated with PDB entry 1D92[27] corresponds to an A-DNA duplex composed of eight base pairs with two G-T mispairs occurring between residues 3 and 14 and between residues 6 and 11. After mutating residue 6 from T to C to retain only a single G-T mispair, this structure served as the initial structure for the G-T mispair in an A-DNA duplex. Moreover, residues 3 and 14 of this structure were used to generate an initial structure for the G-T mispair in aqueous solution. The initial structure of the G-T mispair in a B-DNA duplex, which was chosen to have the same sequence as the A-DNA duplex, was generated using the nucleic acid builder (NAB) program in AmberTools[40] because a crystal structure was not

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

available. The initial structure for the G-T mispair in the DNA polymerase λ variant was obtained from PDB entry 3PML.[9] Each system was solvated in a cuboid water box with ~0.15 M NaCl after neutralizing with Mg^{2+} ions for the DNA duplexes and Na^{+} for the polymerase. PDB files of the equilibrated systems are provided in the Supporting Information (SI).

A QM/MM approach with QM regions illustrated in Figure 2 and Figure S1 was employed to model all of these systems. The hydrogen link atom approach[41] was used to treat the QM/MM interface. The QM region, which contained 50–61 atoms, was described by the ωB97X-D3 functional[42] and the 6-311G** basis set.[43-44] Benchmarking of different levels of theory is provided in Table S3. For the MM region, the nucleosides and nucleotides were represented by the OL15 force field,[45] and the TIP3P water model[46] was utilized to model the solvent. The van der Waals parameters of the Na^{+} and Cl^{-} ions were obtained from the hydration free energy (HFE) parameter set for the TIP3P water model.[47] For the polymerase λ variant system, the protein was described by the AMBER ff14SB force field,[48] and additional polyphosphate parameters[49] were used to model the bound GTP ligand. Following a full equilibration procedure, the protein residues, water molecules, and ions with no atoms within 18 Å of the N1 atom of the guanine base in the G-T mispair were frozen during the QM/MM free energy simulations. The Q-Chem program[50] was used to perform the QM calculations. The QM/MM minimizations were performed using the AMBER software package[40] with the Q-Chem/AMBER interface,[51] while the QM/MM molecular dynamics (MD) simulations were performed using the CHARMM program[52] with the Q-Chem/CHARMM interface.[53]

The finite temperature string method with umbrella sampling[38-39] was utilized to conduct the QM/MM free energy simulations. In this approach, the string is defined in terms of reaction coordinates that are typically inter-atomic distances relevant to the reaction studied. The string is composed of a series of images connecting the reactant and product, where each image is associated with specified values of the reaction coordinates. For each iteration, an umbrella sampling simulation is performed for each image with harmonic restraints on the reaction coordinates. The average values of the reaction coordinates for each image are used to determine the restraints for the next iteration. This process is continued until the string is determined to be converged, as described in more detail below.

For each system, the string corresponding to the tautomerization from the wobble G-T (wG-T) structure to the WC-like G-T* structure, denoted the wG-T→G-T* string, and the string corresponding to the tautomerization from the WC-like G-T* structure to the WC-like G*-T structure, denoted the G-T*→G*-T string, were calculated (Figure S2). The 13 reaction coordinates used to describe the tautomerization process include all interatomic distances at the interface, as depicted in Figure 3. In each case, the initial string was generated by quadratic interpolation of structures generated for the reactant state, an approximate transition state, and the product state using QM/MM energy minimization in which only the QM region was allowed to move in a fixed MM environment.

The wG-T→G-T* and G-T*→G*-T strings were represented by 25 and 15 images, respectively, and each image was defined in terms of the values of the 13 reaction

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

coordinates used for the umbrella sampling restraints. For a given iteration, a 100 fs MD trajectory was propagated for each image while imposing the associated harmonic restraints on the reaction coordinates. Prior to the subsequent iteration, the string images were redistributed evenly along the updated string that was fitted to the average values of the reaction coordinates sampled in the latest iteration. This cycle continued until the string was converged. Each string was computed with 24 or 25 iterations, producing a total of 397.5 ps of QM/MM MD sampling (Table S4). Convergence of the computed strings is illustrated in Figures S3 and S4. For each string, the combined data for all iterations were analyzed using the binless weighted histogram analysis method (WHAM) approach.[54] The final string represents the minimum free energy path (MFEP), and the free energy along the MFEP was generated for each string using the combined data from all iterations. Further details about the simulations are provided in the Supporting Information (SI).

These simulations did not consider the direct conversion from wG-T→G*-T but rather assumed a stepwise process involving the G-T* intermediate, as supported by prior calculations.[35-36] Moreover, these simulations do not address the possibility of ionized G-T mispairs because NMR experiments have suggested that under physiological conditions, misincorporation proceeds primarily via the tautomeric and not the anionic species.[7] The anionic mispairs were also found to be thermodynamically unfavorable in a theoretical study of a DNA polymerase λ variant.[37] The main goal of the present work is to compute the relative free energies of the wG-T, G-T*, and G*T tautomers, which do not depend on the reaction paths. The free energy barriers should be interpreted within the limitations of the mechanistic assumptions.

The statistical uncertainties associated with the free energy calculations were analyzed using a bootstrapping error analysis (see Figure S5). However, this analysis does not account for the errors arising from the level of theory used to generate the potential energy surface and the limitations in the conformational sampling. In particular, the functional and basis set used in the DFT calculations are associated with errors in the relative free energies of at least 1 kcal/mol, and the molecular mechanical force field describing the environment increases this error. The errors introduced by the limited conformational sampling are more difficult to estimate because they depend on the accuracy of the starting structure and the conformational flexibility of the system. Altogether, the uncertainties in the relative free energies are estimated to be at least 3 kcal/mol. As discussed below, the consistency of our results with previous calculations and with experimental data provide validation for the qualitative conclusions emerging from this study.

Results and Discussion

Tautomerization from the wG-T state to the G-T* state

The free energy profiles along the MFEP for the wG-T→G-T* tautomerization in the four different environments studied are depicted in Figure 4. The free energies of tautomerization, ΔG(wG-T→G-T*), and the free energy barriers, $\Delta G^{\ddagger}$(wG-T→G-T*), computed in aqueous solution, the A-DNA duplex, the B-DNA duplex, and the DNA polymerase λ variant in the closed conformation are given in Table 1. The uncertainty analysis of these free energy profiles is given in Figure S5. Our computed ΔG(wG-T→G-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

T*) of 6.0 kcal/mol for the B-DNA duplex is qualitatively consistent with previous NMR studies on B-DNA duplexes indicating that the ΔG(wG-T→G-T*) is 4.3–4.7 kcal/mol at 298 K.[7, 15] In addition, our computed free energy barrier of 15.7 kcal/mol for the B-DNA duplex is consistent with the barrier of 16.0 – 17.0 kcal/mol at room temperature estimated from NMR measurements for the wG-T→G-T* transition in B-DNA duplexes.[7] The slightly negative free energy change of −1.3 kcal/mol associated with the wG-T→G-T* tautomerization in the DNA polymerase λ variant is also consistent with the observation of the WC-like G-T configuration in the X-ray structure of this system in the same closed conformation[9], the observation of WC-like mispairs in structures of other DNA polymerases in the closed conformation[8-10] and the tendency of the polymerase active site to be tuned to the WC shape.[1] Furthermore, the "transition state" (wG-T↔G-T*) structures in Figure 5 are consistent with previous computational studies on isolated base pairs indicating that the transition state structure is the zwitterion (G^+-T^-) formed after proton transfer from T to G.[35, 55]

To identify the factors determining the different free energy profiles, we examined the structures associated with the reactant state (wG-T), the "transition state" (wG-T↔G-T*) and the product state (G-T*) for the wG-T→G-T* tautomerization in the four systems (Figure 5 and Figure S6). Note that these structures are not stationary points on the free energy surfaces and therefore should be viewed as representative configurations rather than true minima or transition states. Although the wG-T structures are similar for the four different systems, the G-T* structures exhibit more significant differences. In particular, the N1-H1-N3, N2-H21-O2, and O6-H3-O4 hydrogen bonds are more linear in the polymerase variant than in aqueous solution (Table S5). Similarly, the hydrogen bond donor-acceptor distances associated with these two hydrogen bonds are significantly shorter in the polymerase variant than in aqueous solution (Table S5). These trends are consistent with the trends in ΔG(wG-T→G-T*), which indicate that the G-T* tautomer is the most stable in the polymerase variant and the least stable in aqueous solution compared to its wG-T analog. Note that these trends are followed for some but not all of these hydrogen-bonding parameters for the A-DNA and B-DNA systems, which exhibit reaction free energies in between those of the polymerase and the aqueous systems. Thus, these reaction free energy differences are explained in part by differences in hydrogen-bonding interactions between the base pairs.

We also analyzed the steric and electrostatic impact of the A-DNA, B-DNA, and polymerase environments on the G-T mispair tautomerization process. This analysis entailed the calculation of the average energies of the G-T mispair, denoted $E_{\text{G-T}}$, and the average electrostatic interaction energy between the G-T mispair and the environment, denoted E_{elec}, during the tautomerization. Specifically, these energies were obtained by averaging the energy of the QM region and the QM/MM electrostatic interaction energy over the last iteration of the string calculation for specific images associated with the wG-T, wG-T↔G-T*, and G-T* states. All of the configurations were weighted to account for the umbrella sampling restraints, and the average energies of the wG-T↔G-T* and G-T* states were computed relative to the average energy of the wG-T state for each system to enable comparisons across systems of different sizes (Table S6). The qualitative trends were reproducible when the same analysis was performed for the second-to-last iteration of the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

string calculations (Table S6). Differences in $E_{G\text{-}T}$ primarily reflect the impact of environmental restraints on hydrogen-bonding and other internal interactions within the G-T mispair (Figure 5A and Table S5), while differences in E_{elec} account for the distinct electrostatic environments. Interestingly, the ordering of $E_{G\text{-}T} + E_{elec}$ for wG-T↔G-T* and G-T* among the DNA duplexes and the polymerase are consistent with the ordering of $\Delta G^{\ddagger}$ (wG-T→-G-T*) and ΔG(wG-T→G-T*), respectively (Table 1). Nevertheless, given the sampling limitations, this analysis should be viewed as providing only qualitative rather than quantitative information.

Although the sequences are identical, the A-DNA and B-DNA duplex structures are different, leading to different interactions of the G-T mispair with the neighboring nucleotides and thus differences in the reaction free energies and free energy barriers associated with tautomerization (Figure 4 and Figure S7). These differences are quantified by analysis of the energies $E_{G\text{-}T}$ and E_{elec} defined above (Table S6). $E_{G\text{-}T}$ for G-T* is positive for A-DNA and negative for B-DNA, suggesting significant differences in the hydrogen-bonding and other internal interactions within the G-T mispair for these two environments. The positive value of E_{elec} for the wG-T↔G-T* and G-T* states for both the A-DNA and B-DNA duplexes indicates that the electrostatic environment stabilizes the wG-T state more than the wG-T↔G-T* and G-T* states for both duplexes, although their magnitudes are different. As mentioned above, $E_{G\text{-}T} + E_{elec}$ is consistent with the higher free energy barrier and more endoergic tautomerization for A-DNA compared to B-DNA. This analysis illustrates that a combination of environmental constraints and electrostatic interactions with the duplex environments, which are structurally distinct, is most likely responsible for the differences in the free energy profiles for tautomerization.

The DNA polymerase λ variant has the largest $\Delta G^{\ddagger}$(wG-T→G-T*) and has the only negative ΔG(wG-T→G-T*) among the four systems. The P–P distances in the polymerase variant are ~2 Å smaller than those in the DNA duplexes, indicating a more compact structure in the polymerase variant (Table S7). More important are the interactions of the G-T mispair with the protein environment. Analysis of $E_{G\text{-}T}$ and E_{elec} indicates that a combination of environmental constraints and electrostatic effects results in a more energetically unfavorable wG-T↔G-T* state for the DNA polymerase λ variant compared to the DNA duplexes (Table S6). This analysis is consistent with the observation that the free energy barrier for wG-T→G-T* tautomerization is higher in the DNA polymerase λ variant than in the DNA duplexes (Figure 4). Moreover, analysis of E_{elec} illustrates that the electrostatic environment of the polymerase stabilizes the G-T* state relative to the wG-T state, in contrast to the DNA duplex environments, which destabilize the G-T* state relative to the wG-T state. This analysis is consistent with the observation that ΔG(wG-T→G-T*) is negative for the DNA polymerase λ variant and positive for the DNA duplexes. This analysis illustrates that the polymerase environment plays an important role in determining the thermodynamics and kinetics for the wG-T→G-T* tautomerization. The computed free energy barrier of ~23 kcal/mol (Table 1) in this closed conformation of the polymerase is much higher than expected based on the rate of misincorporation. This finding is consistent with our experimental studies[7] implying that tautomerization most likely occurs in an ajar conformation of the polymerase or in a manner that may be coupled to closing of the polymerase. This issue will be discussed further below.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

We further examined the polymerase environment to identify the key residues that impact the reaction free energy and free energy barrier. Asn513 and Arg517 exhibit hydrogen-bonding interactions with the guanine base (Figure 6). To elucidate the electrostatic effects, we estimated the electrostatic contributions from these two residues and the two Mg^{2+} ions coordinated to the GTP ligand (Table S8). These calculations indicate that Arg517 significantly stabilizes the G-T* state relative to the wG-T state by ~10 kcal/mol and that Asn513 and the two Mg^{2+} ions also contribute to this stabilization by a smaller amount. This observation is consistent with a previous study in which mutation of Asn279 of DNA polymerase β, which is analogous to Asn513 in the DNA polymerase λ variant, was found to destabilize the WC-like structure of the G-T mispair, converting it to a wobble mispair instead.[56] In addition, these observations are also consistent with kinetic studies showing that mutation of Arg283 in DNA polymerase β, which is analogous to Arg517 in the DNA polymerase λ variant, reduces the rate of G-T misincorporation by ~10 fold,[57] presumably by destabilizing the WC-like G-T mispair. This analysis also indicates that the two Mg^{2+} ions destablize the wG-T↔G-T* state, while Arg517 stabilizes the wG-T↔G-T* state (Table S8), leading to an overall destabilization of the wG-T↔G-T* state relative to the wG-T state. These electrostatic effects of the Mg^{2+} ions are consistent with the proximity of these positively charged ions to the slightly positive charge on the G in the zwitterionic G^{+}-T^{-} "transition state" (Figure 6). Overall, the electrostatic interactions of Arg517, Asn513 and the two Mg^{2+} ions significantly impact the thermodynamics and kinetics of the wG-T→G-T* tautomerization in the DNA polymerase variant.

Tautomerization from the G-T* mispair to the G*-T mispair

The G-T*→G*-T tautomerization involves two proton transfer (PT) reactions (Figure 5B): the PT reaction between two oxygen atoms, denoted as PT-O, and the PT reaction between two nitrogen atoms, denoted as PT-N. These two reactions were found to be concerted in that a single barrier connects the reactant and product without a stable intermediate (Figure 7A), consistent with previous studies of the double proton transfer in A-T and G-C tautomerization[58-60] and prior studies on G-T mispairs.[35-36] Moreover, our results also indicate that the two PT reactions occurring in the G-T*→G*-T tautomerization process are asynchronous, consistent with a previous computational study.[36] However, further analysis indicates that the environment can alter the order of the two proton transfer reactions. Specifically, PT-O occurs first in the aqueous and DNA duplex environments, whereas PT-N occurs first in the DNA polymerase λ variant (Figure 7B). Thus, the polymerase environment alters the fundamental mechanism of this tautomerization process. Moreover, we examined the structures associated with the reactant state (G-T*), "transition state" (G-T*↔G*-T), and product state (G*-T) along the G-T*→G*-T tautomerization process in the four systems (Figure S9) and performed a hydrogen-bonding analysis (Table S5). Our analysis indicates that the G-T*→G*-T tautomerization exhibits smaller differences among the four systems compared to the differences observed for the wG-T→G-T* tautomerization.

The computed free energy profiles for the G-T*→G*-T conversion indicate that the environment has a relatively modest influence on ΔG(G-T*→G*-T) and the associated free energy barrier $\Delta G^{\ddagger}$(G-T*→G*-T) (Table 1 and Figure 7A). In particular, the ΔG(G-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

T*→G*-T) values in aqueous solution and both DNA duplexes are small, suggesting that this tautomerization is nearly isoergic in these environments, consistent with previous QM calculations using an implicit solvent model[35, 55] and NMR measurements on G-T mispairs.[7] In contrast, this tautomerization is slightly exoergic for the DNA polymerase λ variant (Figure 7A and Table 1). The free energy barriers in all four environments are 5 – 7 kcal/mol.

Our results are qualitatively consistent with a previous FEP study[37] of tautomerization in the same polymerase variant. In this previous study, the free energy change for the wG-T→G-T* tautomerization was computed to be 1.8 kcal/mol, and the free energy change for the wG-T→G*-T tautomerization was computed to be −0.2 kcal/mol, in contrast to the values of −1.3 and −3.7 kcal/mol, respectively, obtained in the present work. In both studies, the G*-T form was found to be the most stable form in this polymerase variant and was more stable than the G-T* form by ~2 kcal/mol. The quantitative differences for the free energies of the enol tautomers relative to the wG-T form are most likely due to limitations in conformational sampling and differences in the potential energy surfaces. Specifically, the previous FEP study[37] used a reparameterized force field to describe the mispairs, whereas the present work uses DFT within a QM/MM framework. Despite these differences in methodology, however, the qualitative results are consistent.

Thermodynamics of conformational transitions in polymerase

Based on kinetic modeling, the equilibrium between wobble G-T mispairs in an ajar polymerase conformation and WC-like G-T mispairs in a closed polymerase conformation has been proposed to be disfavored by ~0.7 kcal/mol (Figure 8A).[7] The results obtained in this study indicate that the closed polymerase favors the WC-like G-T mispair by ~1–3 kcal/mol relative to the wobble G-T mispair (Figures 4 and 7A). Since the overall reaction between ajar wobble and closed WC-like mispair is disfavored by ~0.7 kcal/mol, the transition between ajar and closed polymerase with a wobble G-T mispair is deduced to be disfavored by ~1.7–3.7 kcal/mol (Figure 8B), consistent with the crystal structure of DNA polymerase BF1 variant in the ajar state with a wobble G-T mispair.[29] Compared to a preformed DNA duplex, which favors the Watson-Crick geometry, the replicating polymerase expends energy (~2–4 kcal/mol) to create an environment that favors the WC-like shape by closing on the wobble. This expense of energy is compensated by more strongly favoring the WC-like tautomer relative to the wobble in the closed state as compared to the duplex, leading to a net +0.7 kcal/mol destabilization in both cases.

Conclusions

We investigated the environmental effects on G-T mispair tautomerization by studying this process in aqueous solution, an A-DNA duplex, a B-DNA duplex, and a DNA polymerase λ variant using QM/MM free energy simulations. We found that the environment can significantly influence the thermodynamics, kinetics, and reaction mechanisms. The wG-T→G-T* tautomerization is predicted to be endoergic for the aqueous and duplex systems and to be slightly exoergic in the DNA polymerase λ variant. The associated free energy barrier is relatively high for the tautomerization in the closed conformation of the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

polymerase, supporting prior kinetic models[7] implying that tautomerization occurs in the ajar conformation or in a concerted manner coupled to the transition from the ajar to the closed conformation. The free energy profiles for the G-T*→G*-T tautomerization are not as sensitive to the environment, although again the tautomerization is most thermodynamically favorable in the DNA polymerase λ variant. Furthermore, we found that the order of the two proton transfer reactions associated with the G-T*→G*-T tautomerization is different in the polymerase environment than in the aqueous and duplex environments, illustrating that the environment can change the fundamental mechanism as well as the thermodynamics and kinetics of tautomerization.

The differences among the reaction free energies and free energy barriers for the G-T tautomerization processes can be explained in terms of hydrogen-bonding interactions between the G-T mispairs and electrostatic interactions between the G-T mispairs and the environment. Analysis of the process in the DNA polymerase λ variant identified specific residues and metal ions that influence the free energy profiles. For the wG-T→G-T* tautomerization, electrostatic stabilization of G-T* relative to wG-T by Arg517 provides an explanation for the thermodynamic favorability of this reaction in the polymerase environment. Arg517 also stabilizes the “transition state” relative to wG-T, but the more dominant electrostatic destabilization by the two Mg^{2+} ions coordinated to the GTP ligand leads to a higher free energy barrier in the polymerase environment. These analyses suggest that mutating Arg517 could impact the probability of G-T mispair tautomerization in the polymerase variant. Moreover, our results also suggest that Asn513 can slightly stablize the G-T* state relative to the wG-T state, consistent with a previous study regarding the analogous Asn279 in DNA polymerase β.[56] Moreover, the free energy barrier may be lower for the wG-T→G-T* tautomerization when a TTP fragment rather than a GTP fragment is inserted during the polymerization process because of favorable electrostatic interactions of T^- in the zwitterionic G^+-T^- transition state with the two Mg^{2+} ions.

Overall, this computational study highlights the significant effects of the environment on G-T mispair tautomerization. The methods outlined in this study can also be applied to the study of G-U mispair tautomerization in the context of A-RNA, for which Watson-Crick like mispairs have also been observed experimentally.[7] In addition, they can be used to explore other proposed WC-like mispairs[3, 61] that have proven elusive to study experimentally, to understand the mechanism associated with chemical modifications that are proposed to enhance miscoding via altered tautomerization propensities,[19-22, 62] and to examine the role of mutagenic metals such as Mn^{2+} in enhancing tautomerization.[63] The insights provided by this work could enhance fundamental understanding of biologically important processes such as spontaneous mutations and therapeutic design of nucleic acid base analogues.[64] This enhanced understanding could also have broader implications for the design of more effective nucleic acid catalysts and nucleic acid base analogues, as well as strategies to prevent spontaneous mutations that may cause certain types of cancer.[14] Similar computational approaches could be used to study other types of DNA mispairs in a variety of environments. Furthermore, these methods could be used to study RNA-ligand recognition[17-18] and catalysis,[23] where the involvement of tautomers has been well-characterized experimentally.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

We acknowledge helpful discussions with Alexander Soudackov and Clorice Reinhardt. We are grateful for financial support from the National Institutes of Health Grants GM056207 (S.H.-S.) and R01GM089846 (H.M.A.). We acknowledge computational support from the Extreme Science and Engineering Discovery Environment (XSEDE) platform[65] that is supported by the National Science Foundation (NSF). Specifically, this study used resources on Comet at the San Diego Supercomputer Center (SDSC) with an allocation number of TG-MCB120097.

Funding Sources

National Institutes of Health Grant GM056207 to Sharon Hammes-Schiffer and US National Institutes of Health Grant R01GM089846 to Hashim M. Al-Hashimi.

References

(1). Kool ET, Active site tightness and substrate fit in DNA replication. Annu. Rev. Biochem 2002, 71, 191–219. [PubMed: 12045095]

(2). Watson JD; Crick FHC, Genetical Implications of the Structure of Deoxyribonucleic Acid. Nature 1953, 171, 964–967. [PubMed: 13063483]

(3). Topal MD; Fresco JR, Complementary base pairing and the origin of substitution mutations. Nature 1976, 263, 285–289. [PubMed: 958482]

(4). Rozov A; Demeshkina N; Westhof E; Yusupov M; Yusupova G, New Structural Insights into Translational Miscoding. Trends Biochem. Sci 2016, 41, 798–814. [PubMed: 27372401]

(5). Raper AT; Reed AJ; Suo Z, Kinetic Mechanism of DNA Polymerases: Contributions of Conformational Dynamics and a Third Divalent Metal Ion. Chem. Rev 2018, 118, 6000–6025. [PubMed: 29863852]

(6). Yu H; Eritja R; Bloom LB; Goodman MF, Ionization of bromouracil and fluorouracil stimulates base mispairing frequencies with guanine. J. Biol. Chem 1993, 268, 15935–15943. [PubMed: 7688001]

(7). Kimsey IJ; Szymanski ES; Zahurancik WJ; Shakya A; Xue Y; Chu C-C; Sathyamoorthy B; Suo Z; Al-Hashimi HM, Dynamic basis for dG• dT misincorporation via tautomerization and ionization. Nature 2018, 554, 195–201. [PubMed: 29420478]

(8). Wang W; Hellinga HW; Beese LS, Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. Proc. Natl. Acad. Sci. U.S.A 2011, 108, 17644–17648. [PubMed: 22006298]

(9). Bebenek K; Pedersen LC; Kunkel TA, Replication infidelity via a mismatch with Watson–Crick geometry. Proc. Natl. Acad. Sci. U.S.A 2011, 108, 1862–1867. [PubMed: 21233421]

(10). Koag M-C; Nam K; Lee S, The spontaneous replication error and the mismatch discrimination mechanisms of human DNA polymerase β. Nucleic Acids Res. 2014, 42, 11233–11245. [PubMed: 25200079]

(11). Demeshkina N; Jenner L; Westhof E; Yusupov M; Yusupova G, A new understanding of the decoding principle on the ribosome. Nature 2012, 484, 256–259. [PubMed: 22437501]

(12). Rozov A; Demeshkina N; Westhof E; Yusupov M; Yusupova G, Structural insights into the translational infidelity mechanism. Nat. Commun 2015, 6, 7251. [PubMed: 26037619]

(13). Rozov A; Wolff P; Grosjean H; Yusupov M; Yusupova G; Westhof E, Tautomeric G• U pairs within the molecular ribosomal grip and fidelity of decoding in bacteria. Nucleic Acids Res. 2018, 46, 7425–7435. [PubMed: 29931292]

(14). Tomasetti C; Li L; Vogelstein B, Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. Science 2017, 355, 1330–1334. [PubMed: 28336671]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(15). Kimsey IJ; Petzold K; Sathyamoorthy B; Stein ZW; Al-Hashimi HM, Visualizing transient Watson–Crick-like mispairs in DNA and RNA duplexes. Nature 2015, 519, 315–320. [PubMed: 25762137]

(16). Freudenthal BD; Beard WA; Cuneo MJ; Dyrkheeva NS; Wilson SH, Capturing snapshots of APE1 processing DNA damage. Nat. Struct. Mol. Biol 2015, 22, 924–931. [PubMed: 26458045]

(17). Gilbert SD; Reyes FE; Edwards AL; Batey RT, Adaptive Ligand Binding by the Purine Riboswitch in the Recognition of Guanine and Adenine Analogs. Structure 2009, 17, 857–868. [PubMed: 19523903]

(18). Singh V; Peng CS; Li D; Mitra K; Silvestre KJ; Tokmakoff A; Essigmann JM, Direct Observation of Multiple Tautomers of Oxythiamine and their Recognition by the Thiamine Pyrophosphate Riboswitch. ACS Chem. Biol 2014, 9, 227–236. [PubMed: 24252063]

(19). Suen W; Spiro TG; Sowers LC; Fresco JR, Identification by UV resonance Raman spectroscopy of an imino tautomer of 5-hydroxy-2′-deoxycytidine, a powerful base analog transition mutagen with a much higher unfavored tautomer frequency than that of the natural residue 2′-deoxycytidine. Proc. Natl. Acad. Sci. U.S.A 1999, 96, 4500–4505. [PubMed: 10200291]

(20). Harris VH; Smith CL; Jonathan Cummins W; Hamilton AL; Adams H; Dickman M; Hornby DP; Williams DM, The Effect of Tautomeric Constant on the Specificity of Nucleotide Incorporation during DNA Replication: Support for the Rare Tautomer Hypothesis of Substitution Mutagenesis. J. Mol. Biol 2003, 326, 1389–1401. [PubMed: 12595252]

(21). Cantara WA; Murphy FV; Demirci H; Agris PF, Expanded use of sense codons is regulated by modified cytidines in tRNA. Proc. Natl. Acad. Sci. U.S.A 2013, 110, 10964–10969. [PubMed: 23781103]

(22). Vendeix FA; Murphy IV FV; Cantara WA; Leszczyńska G; Gustilo EM; Sproat B; Malkiewicz A; Agris PF, Human tRNALys3UUU Is pre-structured by natural modifications for cognate and wobble codon binding through Keto–Enol tautomerism. J. Mol. Biol 2012, 416, 467–485. [PubMed: 22227389]

(23). Singh V; Fedeles BI; Essigmann JM, Role of tautomerism in RNA biochemistry. RNA 2015, 21, 1–13. [PubMed: 25516996]

(24). Katritzky A; Waring A, 299. Tautomeric azines. Part I. The tautomerism of 1-methyluracil and 5-bromo-1-methyluracil. J. Chem. Soc 1962, 1540–1544.

(25). Wolfenden RV, Tautomeric equilibria in inosine and adenosine. J. Mol. Biol 1969, 40, 307–310. [PubMed: 5365014]

(26). Szymanski ES; Kimsey IJ; Al-Hashimi HM, Direct NMR evidence that transient tautomeric and anionic states in dG· dT form Watson–Crick-like base pairs. J. Am. Chem. Soc 2017, 139, 4326–4329. [PubMed: 28290687]

(27). Hunter WN; Kneale G; Brown T; Rabinovich D; Kennard O, Refined crystal structure of an octanucleotide duplex with G· T mismatched base-pairs. J. Mol. Biol 1986, 190, 605–618. [PubMed: 3783714]

(28). Hunter WN; Brown T; Kneale G; Anand NN; Rabinovich D; Kennard O, The structure of guanosine-thymidine mismatches in B-DNA at 2.5-A resolution. J. Biol. Chem 1987, 262, 9962–9970. [PubMed: 3611072]

(29). Wu EY; Beese LS, The structure of a high fidelity DNA polymerase bound to a mismatched nucleotide reveals an “ajar” intermediate conformation in the nucleotide selection mechanism. J. Biol. Chem 2011, 286, 19758–19767. [PubMed: 21454515]

(30). Zhao Y; Gregory MT; Biertümpfel C; Hua Y-J; Hanaoka F; Yang W, Mechanism of somatic hypermutation at the WA motif by human DNA polymerase η. Proc. Natl. Acad. Sci. U.S.A 2013, 110, 8146–8151. [PubMed: 23630267]

(31). Xia S; Konigsberg WH, Mispairs with Watson-Crick base-pair geometry observed in ternary complexes of an RB69 DNA polymerase variant. Protein Sci. 2014, 23, 508–513. [PubMed: 24458997]

(32). Xia S; Wang M; Lee HR; Sinha A; Blaha G; Christian T; Wang J; Konigsberg W, Variation in mutation rates caused by RB69pol fidelity mutants can be rationalized on the basis of their kinetic behavior and crystal structures. J. Mol. Biol 2011, 406, 558–570. [PubMed: 21216248]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(33). Xia S; Wang J; Konigsberg WH, DNA mismatch synthesis complexes provide insights into base selectivity of a B family DNA polymerase. J. Am. Chem. Soc 2013, 135, 193–202. [PubMed: 23214497]

(34). Vaisman A; Ling H; Woodgate R; Yang W, Fidelity of Dpo4: effect of metal ions, nucleotide selection and pyrophosphorolysis. EMBO J. 2005, 24, 2957–2967. [PubMed: 16107880]

(35). Nomura K; Hoshino R; Shimizu E; Hoshiba Y; Danilov VI; Kurita N, DFT calculations on the effect of solvation on the tautomeric reactions for wobble Gua-Thy and canonical Gua-Cyt base-pairs. J. Mod. Phys 2013, 4, 422–431.

(36). Brovarets' O. h. O.; Hovorun DM, The nature of the transition mismatches with Watson–Crick architecture: the G*· T or G· T* DNA base mispair or both? A QM/QTAIM perspective for the biological problem. J. Biomol. Struct. Dyn 2015, 33, 925–945. [PubMed: 24842163]

(37). Maximoff SN; Kamerlin SCL; Florián J, 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 2017, 121, 7813–7822. [PubMed: 28732447]

(38). Rosta E; Nowotny M; Yang W; Hummer G, Catalytic mechanism of RNA backbone cleavage by ribonuclease H from quantum mechanics/molecular mechanics simulations. J. Am. Chem. Soc 2011, 133, 8934–8941. [PubMed: 21539371]

(39). Ganguly A; Thaplyal P; Rosta E; Bevilacqua PC; Hammes-Schiffer S, Quantum mechanical/molecular mechanical free energy simulations of the self-cleavage reaction in the hepatitis delta virus ribozyme. J. Am. Chem. Soc 2014, 136, 1483–1496. [PubMed: 24383543]

(40). Case DA; Cheatham TE; Darden T; Gohlke H; Luo R; Merz KM Jr; Onufriev A; Simmerling C; Wang B; Woods RJ, The Amber biomolecular simulation programs. J. Comput. Chem 2005, 26, 1668–1688. [PubMed: 16200636]

(41). Groenhof G, Introduction to QM/MM simulations In Biomolecular Simulations, Springer: 2013; pp 43–66.

(42). Lin Y-S; Li G-D; Mao S-P; Chai J-D, Long-range corrected hybrid density functionals with improved dispersion corrections. J. Chem. Theory Comput 2012, 9, 263–272. [PubMed: 26589028]

(43). Krishnan R; Binkley JS; Seeger R; Pople JA, Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. J. Chem. Phys 1980, 72, 650–654.

(44). McLean A; Chandler G, Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z= 11–18. J. Chem. Phys 1980, 72, 5639–5648.

(45). Zgarbová M; Sponer J; Otyepka M; Cheatham III TE; Galindo-Murillo R; Jurecka P, Refinement of the sugar–phosphate backbone torsion beta for AMBER force fields improves the description of Z-and B-DNA. J. Chem. Theory Comput 2015, 11, 5723–5736. [PubMed: 26588601]

(46). Jorgensen WL; Chandrasekhar J; Madura JD; Impey RW; Klein ML, Comparison of simple potential functions for simulating liquid water. J. Chem. Phys 1983, 79, 926–935.

(47). Li P; Song LF; Merz KM Jr, Systematic Parameterization of Monovalent Ions Employing the Nonbonded Model. J. Chem. Theory Comput 2015, 11, 1645–1657. [PubMed: 26574374]

(48). Maier JA; Martinez C; Kasavajhala K; Wickstrom L; Hauser KE; Simmerling C, ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J. Chem. Theory Comput 2015, 11, 3696–3713. [PubMed: 26574453]

(49). Meagher KL; Redman LT; Carlson HA, Development of polyphosphate parameters for use with the AMBER force field. J. Comput. Chem 2003, 24, 1016–1025. [PubMed: 12759902]

(50). Shao Y; Gan Z; Epifanovsky E; Gilbert AT; Wormit M; Kussmann J; Lange AW; Behn A; Deng J; Feng X, Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. Mol. Phys 2015, 113, 184–215.

(51). Götz AW; Clark MA; Walker RC, An extensible interface for QM/MM molecular dynamics simulations with AMBER. J. Comput. Chem 2014, 35, 95–108. [PubMed: 24122798]

(52). Brooks BR; Brooks CL; MacKerell AD; Nilsson L; Petrella RJ; Roux B; Won Y; Archontis G; Bartels C; Boresch S, CHARMM: the biomolecular simulation program. J. Comput. Chem 2009, 30, 1545–1614. [PubMed: 19444816]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(53). Woodcock HL; HodoščDek M; Gilbert AT; Gill PM; Schaefer HF; Brooks BR, Interfacing Q-Chem and CHARMM to perform QM/MM reaction path calculations. J. Comput. Chem 2007, 28, 1485–1502. [PubMed: 17334987]

(54). Souaille M; Roux B, Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. Comput. Phys. Commun 2001, 135, 40–57.

(55). Brovarets' O. h. O.; Hovorun DM, How many tautomerization pathways connect Watson–Crick-like G*· T DNA base mispair and wobble mismatches? J. Biomol. Struct. Dyn 2015, 33, 2297–2315. [PubMed: 25932960]

(56). Koag M-C; Lee S, Insights into the effect of minor groove interactions and metal cofactors on mutagenic replication by human DNA polymerase β. Biochem. J 2018, 475, 571–585. [PubMed: 29301983]

(57). Ahn J; Werneburg BG; Tsai M-D, DNA polymerase β: structure– fidelity relationship from pre-steady-state kinetic analyses of all possible correct and incorrect base pairs for wild type and R283A mutant. Biochemistry 1997, 36, 1100–1107. [PubMed: 9033400]

(58). Florian J; Hrouda V; Hobza P, Proton Transfer in the Adenine-Thymine Base Pair. J. Am. Chem. Soc 1994, 116, 1457–1460.

(59). Florian J; Leszczyński J, Spontaneous DNA Mutations Induced by Proton Transfer in the Guanine·Cytosine Base Pairs: An Energetic Perspective. J. Am. Chem. Soc 1996, 118, 3010–3017.

(60). Gorb L; Podolyan Y; Dziekonski P; Sokalski WA; Leszczynski J, Double-proton transfer in adenine– thymine and guanine– cytosine base pairs. A post-hartree– fock ab initio study. J. Am. Chem. Soc 2004, 126, 10119–10129. [PubMed: 15303888]

(61). Topal MD; Fresco JR, Base pairing and fidelity in codon–anticodon interaction. Nature 1976, 263, 289–293. [PubMed: 958483]

(62). Ikeuchi Y; Kimura S; Numata T; Nakamura D; Yokogawa T; Ogata T; Wada T; Suzuki T; Suzuki T, Agmatine-conjugated cytidine in a tRNA anticodon is essential for AUA decoding in archaea. Nat. Chem. Biol 2010, 6, 277–282. [PubMed: 20139989]

(63). Vashishtha AK; Wang J; Konigsberg WH, Different divalent cations alter the kinetics and fidelity of DNA polymerases. J. Biol. Chem 2016, 291, 20869–20875. [PubMed: 27462081]

(64). Li D; Fedeles BI; Singh V; Peng CS; Silvestre KJ; Simi AK; Simpson JH; Tokmakoff A; Essigmann JM, Tautomerism provides a molecular explanation for the mutagenic properties of the anti-HIV nucleoside 5-aza-5,6-dihydro-2′-deoxycytidine. Proc. Natl. Acad. Sci. U.S.A 2014, 111, E3252–E3259. [PubMed: 25071207]

(65). Towns J; Cockerill T; Dahan M; Foster I; Gaither K; Grimshaw A; Hazlewood V; Lathrop S; Lifka D; Peterson GD, XSEDE: accelerating scientific discovery. Comput. Sci. Eng 2014, 16, 62–74.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

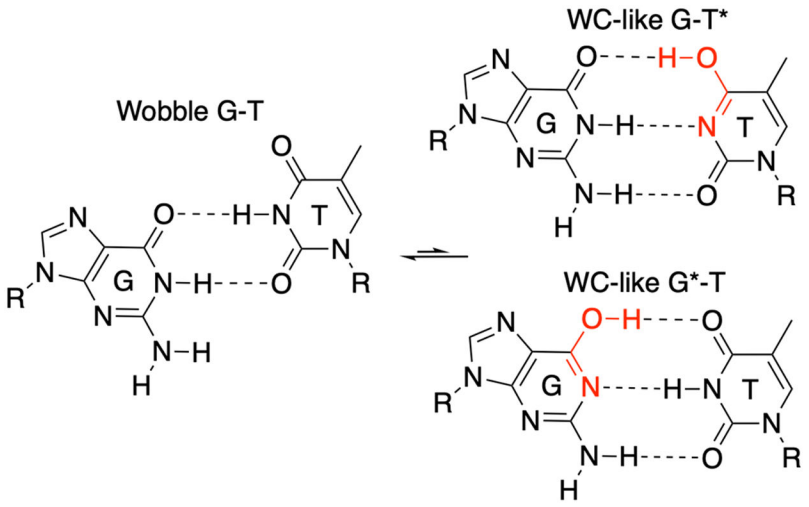


Figure 1.

Tautomerization of the wobble G-T mispair into WC-like structures. The asterisk denotes the enol tautomeric form of G or T.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

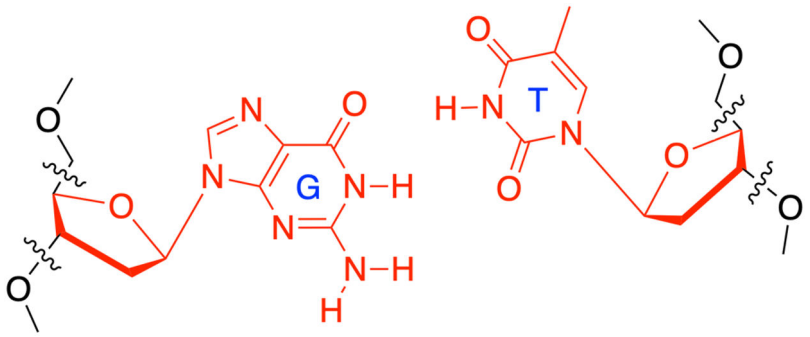


Figure 2.

Illustration of the QM region for the QM/MM simulations of the G-T mispair in the DNA duplexes. The atoms depicted in red are in the QM region, and the cut bonds are indicated by wavy curves. Illustration of the QM regions for the G-T mispair in aqueous solution and the DNA polymerase λ variant are shown in Figure S1.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

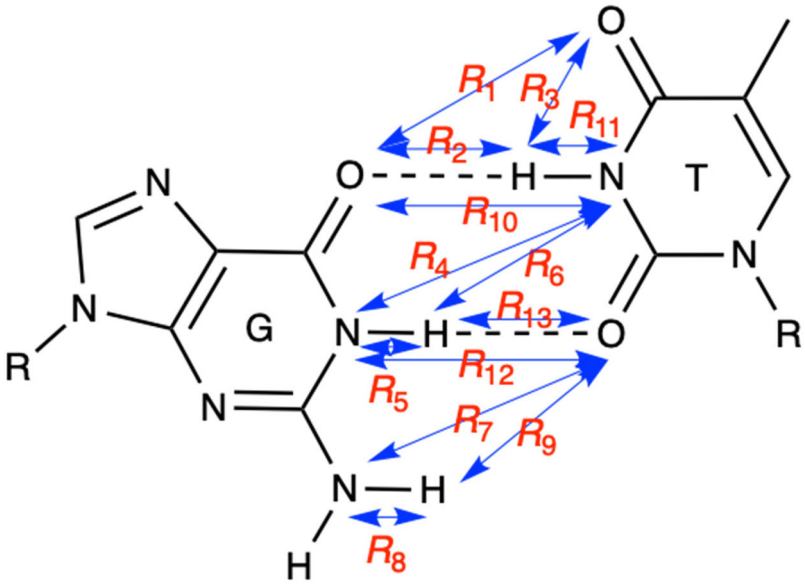


Figure 3.

Depiction of the 13 reaction coordinates used to describe the tautomerization process in the QM/MM finite temperature string simulations with umbrella sampling.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

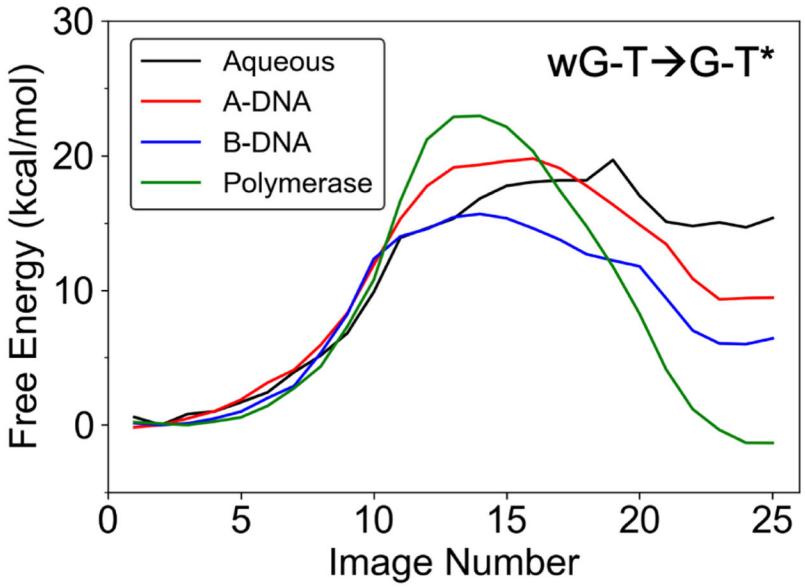


Figure 4.

Free energy profiles along the MFEPs computed from the strings corresponding to the wG-T→G-T* tautomerization for the four systems investigated. The images correspond to equally spaced points along the MFEP. The aqueous system is less well-defined because of the flexibility of the DNA base pair.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

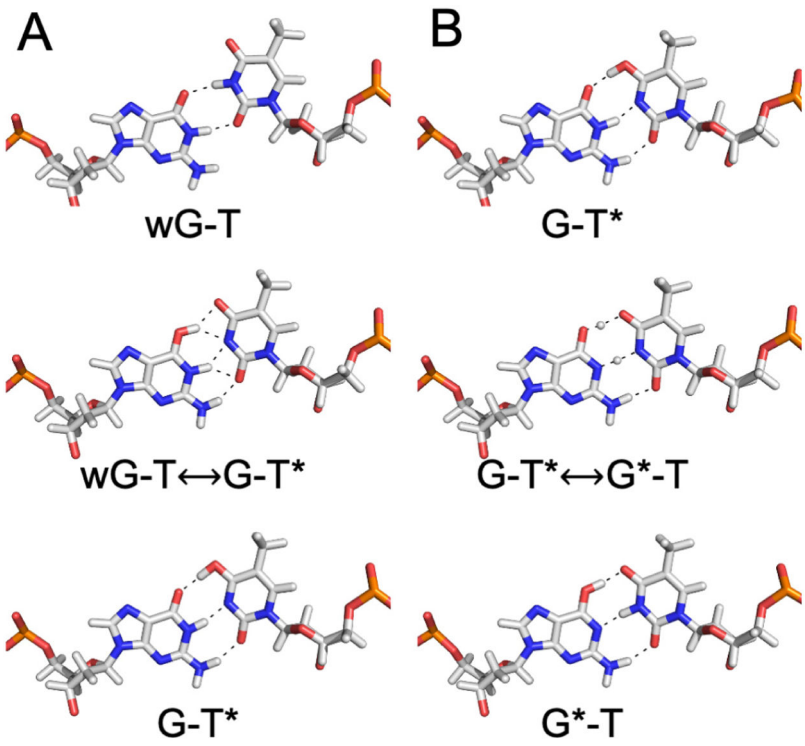


Figure 5.

(A) Representative configurations along the wG-T→G-T* tautomerization process in the A-DNA duplex. These structures are the last configurations of the last iteration of each string simulation for the images corresponding to the reactant state, the “transition state” (i.e., top of the free energy barrier), and the product state in Figure 4. Specifically, these configurations were obtained from images 1, 16, and 23 of the wG-T→G-T* string for the A-DNA duplex. (B) Analogous representative configurations along the G-T*→G*-T tautomerization process in the A-DNA duplex. These configurations were obtained from images 2, 9, and 14 of the G-T*→G*-T string for the A-DNA duplex. Analogous configurations along the wG-T→G-T* tautomerization process for the other systems are given in Figure S6.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

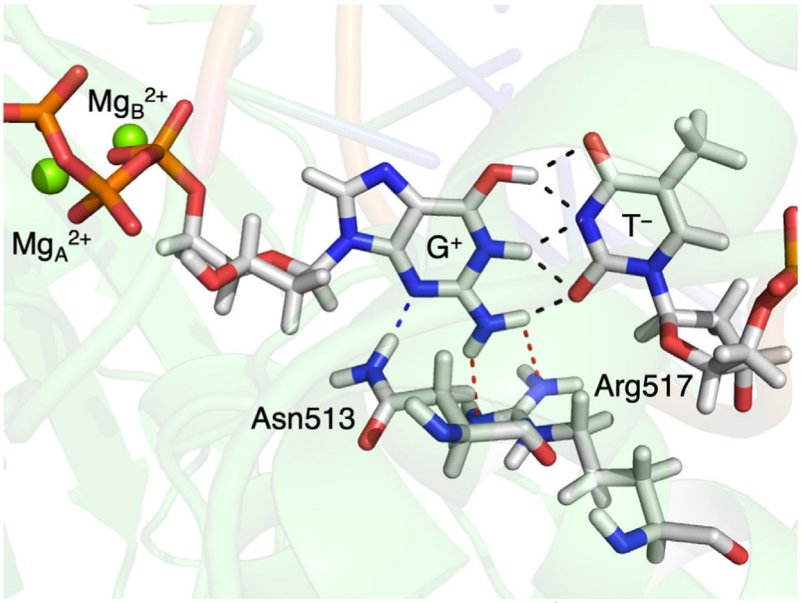


Figure 6.

Interactions of Asn513, Arg517, and the two Mg^{2+} ions with the G-T mispair in the DNA polymerase λ variant. This structure is the last configuration of the last iteration for image 14 of the DNA polymerase variant, corresponding to the wG-T↔G-T* state in Figure 4. The hydrogen bonds within the G-T mispair are indicated with black dashed lines, the hydrogen bond between Asn513 and G is indicated with a blue dashed line, and the hydrogen bonds between Arg517 and G are indicated with red dashed lines. The analogous configurations corresponding to the wG-T and G-T* states are shown in Figure S8.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

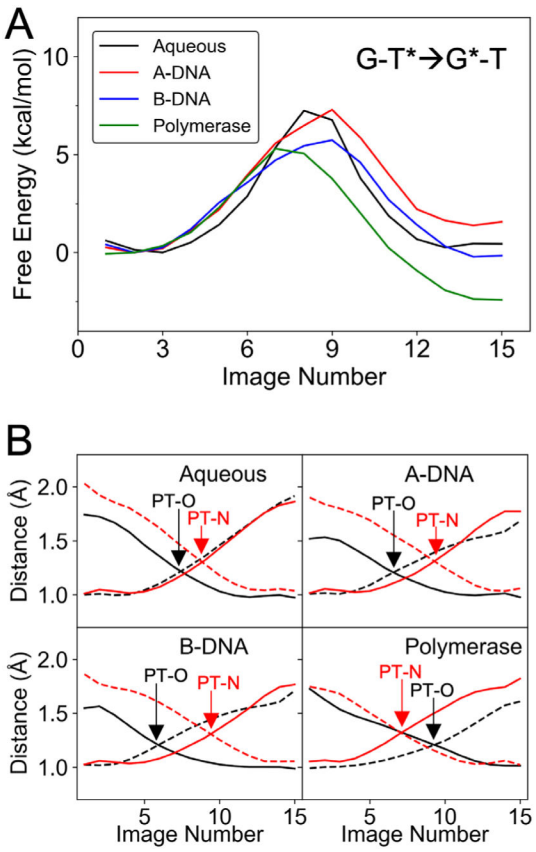


Figure 7.

(A) Free energy profiles along the MFEPs computed from the strings corresponding to the G-T*⟶G*-T tautomerization for the four systems investigated. The images correspond to equally spaced points along the MFEP. (B) The key reaction coordinates associated with the G-T*⟶G*-T tautomerization along the MFEP, where R2, R3, R5, and R6 are represented by black solid, black dashed, red solid, and red dashed curves, respectively. Note that R2 and R3 are associated with PT-O (i.e., proton transfer between the two oxygen atoms), and R5 and R6 are associated with PT-N (i.e., proton transfer between the two nitrogen atoms).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

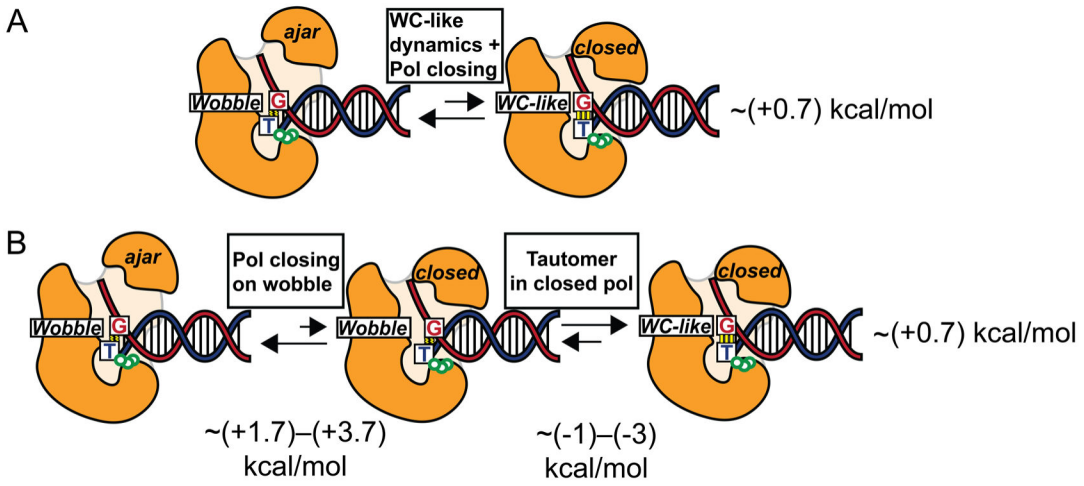


Figure 8.

(A) Equilibrium between DNA polymerase in an ajar conformation with a wobble G-T mispair and DNA polymerase in a closed conformation with a WC-like G-T mispair. A value of ~(+0.7) kcal/mol is deduced based on kinetic studies of misincorporation of DNA polyemrase β at 37 °C.[7] (B) Decomposition of the equilibrium process in A into two steps involving closing of the polymerase on a wobble G-T mispair and subsequent conversion of the wobble mispair into a WC-like conformation within the closed polymerase, as computed in this work. The overall change in free energy is independent of path and therefore is the same for the paths depicted in parts A and B, allowing the deduction of the free energy associated with the first step in part B.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Calculated Free Energy Barriers and Reaction Free Energies for G-T Mispair Tautomerization in Aqueous Solution, the A-DNA Duplex, the B-DNA Duplex, and the DNA Polymerase λ Variant.[a]

	$\Delta G^{\ddagger}$ (wG-T→G-T*)	ΔG (wG-T→G-T*)	$\Delta G^{\ddagger}$ (G-T*→G*-T)	ΔG (G-T*→G*-T)
Aqueous	19.7	14.7	7.2	0.3
A-DNA	20.0	9.5	7.3	1.4
B-DNA	15.7	6.0	5.7	−0.2
DNA polymerase λ variant	23.0	−1.3	5.4	−2.4

[a]Energies given in kcal/mol, and the temperature is 300 K.