THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Cite This: J. Phys. Chem. B 2018, 122, 1152−1160

pubs.acs.org/JPCB

Computational Study of Uracil Tautomeric Forms in the Ribosome: The Case of Uracil and 5-Oxyacetic Acid Uracil in the First Anticodon Position of tRNA

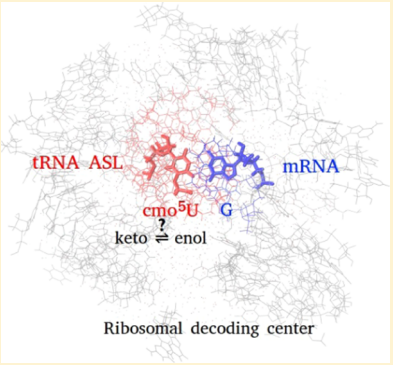
Yossa Dwi Hartono,[†,‡] Mika Ito,[†] Alessandra Villa,[†] and Lennart Nilsson*[,†]

[†]Department of Biosciences and Nutrition, Karolinska Institutet, SE-141 83 Huddinge, Sweden

[‡]Division of Structural Biology and Biochemistry, School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, 637551 Singapore

Supporting Information

ABSTRACT: Tautomerism is important in many biomolecular interactions, not least in RNA biology. Crystallographic studies show the possible presence of minor tautomer forms of transfer-RNA (tRNA) anticodon bases in the ribosome. The hydrogen positions are not resolved in the X-ray studies, and we have used ab initio calculations and molecular dynamics simulations to understand if and how the minor enol form of uracil (U), or the modified uracil 5-oxyacetic acid (cmo^{5}U), can be accommodated in the tRNA−messenger-RNA interactions in the ribosome decoding center. Ab initio calculations on isolated bases show that the modification affects the keto−enol equilibrium of the uracil base only slightly; the keto form is dominant (>99.99%) in both U and cmo^{5}U. Other factors such as interactions with the surrounding nucleotides or ions would be required to shift the equilibrium toward the enol tautomer. Classical molecular simulations show a better agreement with the X-ray structures for the enol form, but free energy calculations indicate that the most stable form is the keto. In the ribosome, the enol tautomers of U and cmo^{5}U pair with a guanine forming two hydrogen bonds, which do not involve the enol group. The oxyacetic acid modification has a minor effect on the keto−enol equilibrium.



INTRODUCTION

During protein synthesis on the ribosome, the transfer-RNA (tRNA) is selected to match the messenger-RNA (mRNA) in the ribosomal A-site.[1−3] The ribosomal scaffold ensures the fidelity of translation by rejecting the unfavorable non-complementary tRNA−mRNA interactions. This selection is tighter for the first and second codon positions, whereas some mismatches are allowed for the third position.[4,5] One example is the guanine−uracil (G−U) mismatch, which typically forms the wobble configuration with two hydrogen bonds, instead of three in a complementary Watson−Crick G−C match.

RNA bases can exist in diverse tautomeric forms. Uracil has a carboxyl group, which may exchange between an enol and keto form, guanine and cytosine have both carboxyl and amino groups, which can exchange between enol−keto or amino−imino forms. Most of the time, the keto and amino forms (major tautomer) are predominant under physiological conditions, and the imino and enol forms are very rare (minor tautomer). Nevertheless, tautomerization of RNA bases is known to influence RNA biochemistry (see ref 6 for a recent review). On one hand, the minor tautomers may be advantageous in the biochemical activities of RNA enzymes and aptamers; on the other hand, they may be implicated as the source of error during transcription and translation. In the context of interactions in ribosomal translation and in other relevant interactions requiring base complementarity, the minor tautomers can be detrimental to fidelity because the tautomers would have different sets of hydrogen bond donor and acceptors, causing errors in these processes.

Investigating base tautomerization at physiological condition is experimentally challenging due to extremely low abundance of the minor tautomers, fast exchange rate, and high chemical and structural similarity. Base tautomerization in RNA duplexes has been recently studied using NMR relaxation dispersion,[7,8] a highly specialized experimental technique, where the authors were able to measure quantitatively the populations of the rare tautomeric forms. There are also examples of approaches that combine experiments and quantum mechanical (QM) calculation.[9−11] Singh et al.,[9] for instance, used two-dimensional IR, variable temperature Fourier transform infrared and NMR spectroscopies, binding isotope effects, and computational methods to identify and quantify tautomers of a ligand bound to an RNA aptamer. There are also examples of purely computational approaches,[12−16] which typically utilize ab initio and/or free energy calculations. Maximoff et al.[12] noted a G−T mismatch with the Watson−Crick geometry in DNA polymer-

Received: November 3, 2017
Revised: December 17, 2017
Published: December 20, 2017

DOI: 10.1021/acs.jpcb.7b10878
J. Phys. Chem. B 2018, 122, 1152−1160

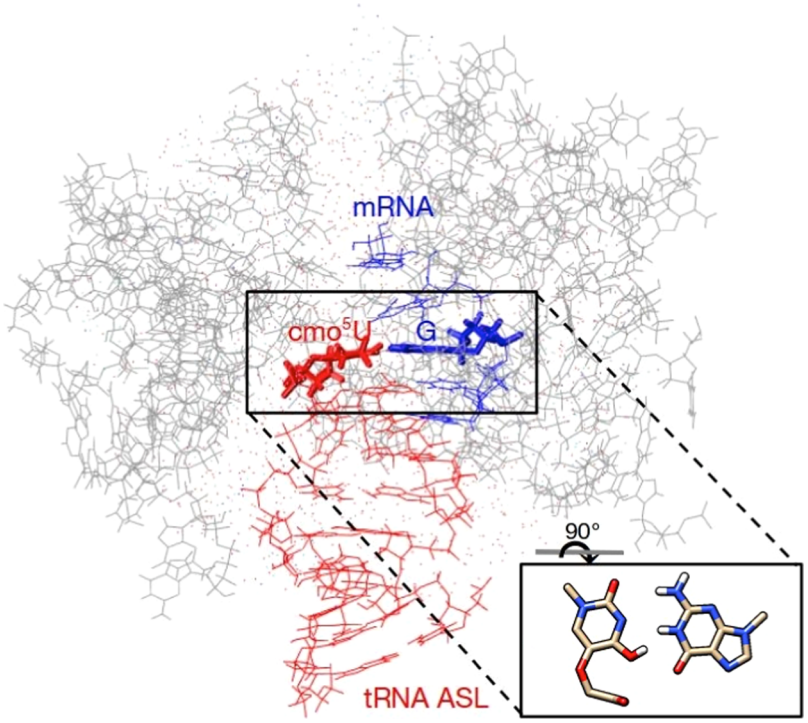


Figure 1. Representation of the subset of the X-ray structure used for MD simulation, a 25 Å radius sphere centered at cmo^{5}U−G. Shown are mRNA (blue), tRNA anticodon stem loop (ASL) fragment (red), various A-site ribosomal components (gray), water, and ions (dots). Inset shows the cmo^{5}U−G base pair configuration of the X-ray crystal structure.

ase λ and through molecular dynamics (MD) simulations, free energy perturbation, and ab initio calculations concluded that the enol G accounts for this observation. Satpati and Åqvist[13] examined G−U mismatches in the first and second codon positions in the ribosomal A-site and through the same set of computational tools also concluded that enol G is the most likely explanation.

In crystallographic studies of ribosome complexes,[17−20] minor enol tautomers have been put forward as possible explanations for the observed Watson−Crick geometries for U−G base pairs. For example, U−G base pairs with a Watson−Crick geometry were observed in the first and second codon positions in the ribosome A- and P-sites[17,18] for unmodified uridines, but there are also examples in the third/wobble codon position, where U is modified and is on the tRNA anticodon.[19,20] The 5-oxyacetic acid (cmo^5) modification of uridine in the first anticodon position (U34) of a tRNA^{Val}, accepts all four bases in the third codon position.[21] Weixlbaumer et al.[19] resolved the X-ray crystal structures of tRNA^{Val} in the ribosomal A-site with the four possible pairings and observed the expected geometries with A/U/C, but the Watson−Crick, instead of wobble, geometry with G (Figure 1). The authors proposed the enol form of cmo^{5}U as the explanation for this observation with three main arguments: (1) *More favorable cmo^{5}U−G interaction*. In the Watson−Crick geometry, there would be three hydrogen bonds, instead of two in the wobble geometry, with cmo^{5}U in the keto form. (2) *The modification*. Hillen et al.[22] observed that the crystal structure of 5-methoxyuridine showed an increase in the C4−O4 bond length and a decrease in the N3−C4 bond length, indicating that the group attached to C5 affects the keto−enol equilibrium of uridine. Analogously, the similar 5-oxyacetic acid modification is expected to affect the keto−enol equilibrium of uridine. (3) *The ribosomal environment*. Dąbkowska et al.[14] showed that interactions with a glycine zwitterion decreased the energy difference between the keto and enol tautomers. The ribosomal decoding center environment, particularly base pairing to G, might also stabilize the enol tautomer and lower the energy difference.

However, because the hydrogen positions are not resolved in the structures, and though the electron density seems to rule out the wobble geometry, it is not entirely conclusive at 3.1 Å resolution. In this study, we examined the hypothesis that the enol form of cmo^{5}U is consistent with the observed Watson−Crick geometry, taking into account the above arguments. First, we investigated the stability of the keto and enol forms of U and cmo^{5}U for the isolated base and base paired with a guanine in aqueous environments using ab initio calculations to find out whether the modification stabilizes the enol form. To account for the ribosomal environment surrounding the base pair, we sampled the conformations of keto and enol forms of U and cmo^{5}U in the ribosomal decoding center using classical molecular dynamics, and we calculated the relative free energy of tautomerization between modified and unmodified U using a molecular mechanics (MM) description with a QM/MM energy correction. Finally, because the molecular dynamics simulation explores conformations not observed in the crystal structure, particularly the base-pair geometry and the orientation of the modification, we proposed alternative conformations for cmo^{5}U.

METHODS

Ab Initio Calculations. Initial coordinates of the bases (U and cmo^{5}U) and the base pairs (U−G and cmo^{5}U−G) were taken from the X-ray crystal structure (protein data bank (PDB) ID: 2UU9).[19] Hydrogen atoms were added with the CHARMM program[23] using the CHARMM36 force field for nucleic acids[24−26] and the force field created for the modified

DOI: 10.1021/acs.jpcb.7b10878
J. Phys. Chem. B 2018, 122, 1152−1160

uridine.[27] Sugars were substituted by hydrogen atoms. Atom numbering for U and cmo^{5}U is shown in Figure S1.

Geometry optimizations were carried out first using the Hartree−Fock method,[28] followed by geometry optimizations using the second-order Møller−Plesset (MP2) perturbation method[29] with the 3-21+G** basis set. Solvent effects were included in the geometry optimization using the polarizable continuum model (PCM)[30,31] with a dielectric constant of 78.4. For the base pairs, the Watson−Crick geometries, in which two bases interact with their Watson−Crick edges, were targeted for geometry optimizations. For the enol form, although there are two directions of the H atom of the hydroxy group, only the conformation with the H atom directed to the Watson−Crick edge was used in this study because the H atom is considered to be involved in the interaction between the modified uridine and guanosine in the previous experimental study.[19] Energies of the structures were calculated at the MP2/6-31++G**//MP2/3-21+G** level. Solvent effects were included using the polarizable continuum model (PCM)[30,31] with a dielectric constant = 78.4. QM calculations were performed with the Gaussian 03 program.[32]

Molecular Dynamics Simulations. Molecular dynamics simulations were performed for the modified and unmodified uracil base alone and base paired to guanine in solution and in the ribosome complex using the CHARMM (version 41a2) program[23] and the CHARMM36 force field for nucleic acids,[24−26] ions,[33,34] and modified nucleotides.[27]

The ribosome complex was simulated as a spherical system with radius 25 Å, centered on tRNA-U34-O4, using the X-ray structure (PDB ID: 2UU9)[19] as the starting structure and with a spherical boundary potential[35] to prevent ions and water from leaving the sphere (Figure 1). Atom deletion outside the sphere is done by residue to prevent residue fragmentation. Atoms that ended up outside the sphere were restrained with a force constant of 2 kcal/mol Å^2 throughout all of the simulations and minimizations. The systems were solvated using TIP3P water molecules.[36] Ions were added to neutralize the system. For more details on the setup of the starting structure, see ref 37.

For simulations of the uracil base and of the U−G base pair (with and without modifications), the solutes were solvated using TIP3P water molecules[36] in cubic 25 × 25 × 25 Å^3 and orthogonal 40 × 20 × 20 Å^3 boxes, respectively. Ions were added to neutralize the system, and additional sodium and chloride ions were included to reach an ionic concentration of approximately 0.1 M. Simulations were performed using periodic boundary conditions. To keep the base pair planar, harmonic restraints were applied on base atoms (also on O7, C8, C9 for cmo^{5}U) in the z-direction with a force constant of 50 kcal/mol Å^2, such that the base pair can only move freely in the x−y plane.

The structures were minimized with steepest descent and adopted-basis Newton−Raphson methods with position restraints on the heavy atoms with a force constant of 85 kcal/mol Å^2. A nonbonded list cutoff of 16 Å was used with the electrostatics force switch and van der Waals switch functions between 10 and 12 Å. Simulations were performed in the NVT ensemble with Langevin dynamics with the collision frequency of 10 ps^{-1}. In all of the simulations, the SHAKE algorithm was used to constrain the bonds involving hydrogens.[38] A lookup table was used for nonbonded interactions.[39] The leapfrog integrator was used with an integration time-step of 2 fs. The length of simulations is 10 ns for ribosome.

Potential of Mean Force (PMF). The potential of mean force was calculated along the dihedral $d3$(N3−C4−O4−H4) of the enol form of uracil using umbrella sampling (see Figure S1 for atom numbering). For the umbrella sampling, a harmonic potential was applied to the dihedral $d3$(N3−C4−O4−H4) in windows separated by 5° in the range of 0−360°. Initial structures were generated with 0.6 ps simulation for each window with the force constant of 1000 kcal/mol rad^2. The final structure in each window was used as the starting structure in the next. The production run for each window is 500 ps, with the harmonic force constant of 500 kcal/mol rad^2 at 300 K. The potential of mean force (PMF) was constructed from the dihedrals in all of the windows from the last 250 ps using the weighted histogram analysis method.[40,41]

A harmonic restraint with the force constant of 5 kcal/mol Å^2 was applied to the N1−N3 distance to keep the base pair together during the sampling. For cmo^{5}U, additional harmonic restraints with the force constant of 5 kcal/mol Å^2 were applied to dihedrals C6−C5−O7−C8 and C5−O7−C8−C9 to keep the modification moiety in the position observed in the crystal structure.

Free Energy Calculations. The free energy of tautomerization between the two tautomers, $\Delta G_{\text{keto}\rightarrow\text{enol}}$, was calculated with the Bennet acceptance ratio (BAR) method.[42] The simulations were done at both states, and the potential energies of coordinates for each of the two trajectories were calculated for both states. The BAR equation is

$$\Delta A(0 \rightarrow 1) = k_B T \ln\left(\frac{\langle f(U_0 - U_1 + C)\rangle_1}{\langle f(U_1 - U_0 - C)\rangle_0}\right) + C \tag{1}$$

where k_B is the Boltzmann constant, T is the temperature, U_i is the potential energy of the coordinates for state i, $\langle\ \rangle_i$ denotes an average over state i, $f(x)$ is the Fermi function $f(x) = \frac{1}{1+\exp\left(\frac{x}{k_B T}\right)}$, and C is the free energy, which is to be solved for in a self-consistent manner.

Because the electronic configuration of the two forms may be differently influenced by the ribosomal environment, we also calculated the QM/MM potential energy using the same MM trajectory. The QM/MM energy is then used to reweight the MM potential. This calculation, QM-non-Boltzmann BAR (QM-NBB), has been shown to improve the free energy calculations.[43] The equation introduces reweighting terms to the BAR equation (eq 1), such that

$$\Delta A(0 \rightarrow 1) = k_B T \ln\left(\frac{\left\langle f(U_0 - U_1 + C)\exp\left(\frac{V_1^b}{k_B T}\right)\right\rangle_{1,b}}{\left\langle f(U_1 - U_0 - C)\exp\left(\frac{V_0^b}{k_B T}\right)\right\rangle_{0,b}} \frac{\left\langle \exp\left(\frac{V_0^b}{k_B T}\right)\right\rangle_{0,b}}{\left\langle \exp\left(\frac{V_1^b}{k_B T}\right)\right\rangle_{1,b}}\right) + C \tag{2}$$

where the biasing potential V^b is the difference between the MM and QM/MM energies

$$V^b = U^{MM} - U^{QM/MM} \tag{3}$$

DOI: 10.1021/acs.jpcb.7b10878
J. Phys. Chem. B 2018, 122, 1152−1160

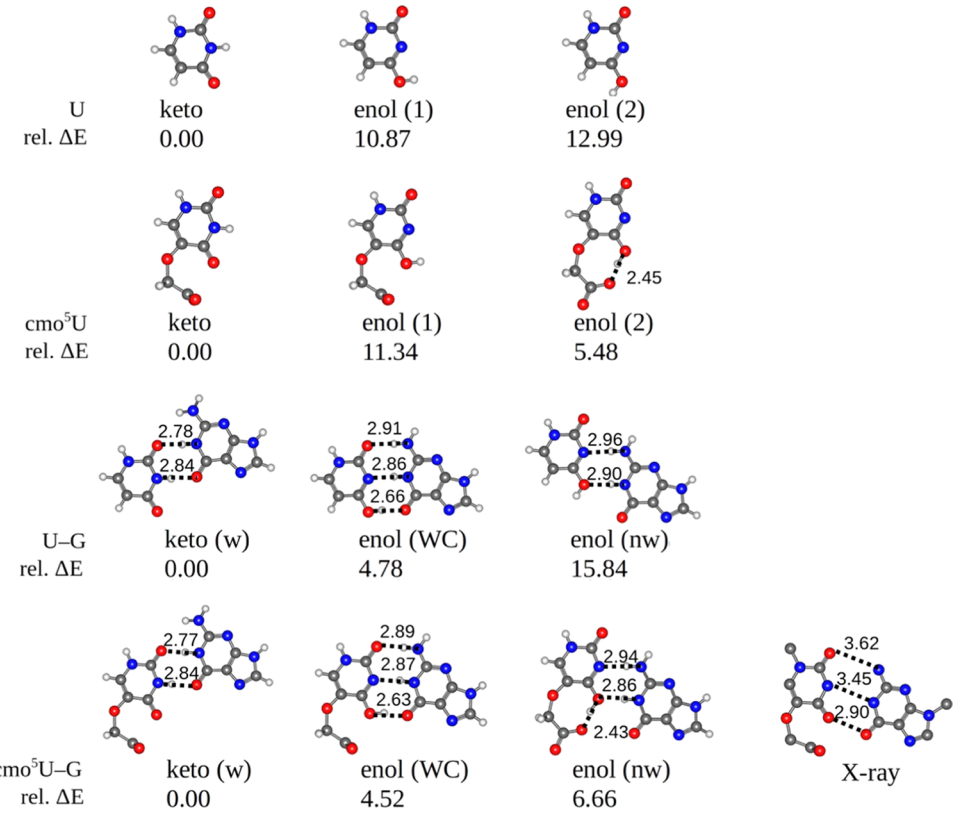


Figure 2. Ab initio optimized structures of the keto and enol forms of U and cmo^{5}U. For the enol forms, two possible orientations for the hydroxyl H are considered. Δ*E* is the difference in energy of the optimized structure relative to that of the keto tautomer, in kcal/mol. Donor−acceptor lengths of hydrogen bonds are shown in angstrom. w corresponds to wobble, WC to Watson−Crick, and nw to "new wobble" base pair geometry. The base pair geometry in the X-ray crystal structure (PDB ID: 2UU9)[19] is also shown for comparison.

QM/MM reweighting can be done just on the end states (indirect) or all of the states (direct). If alchemical intermediate states are to be reweighted, the QM/MM energy is obtained by linear scaling

$$U_\lambda^{\mathrm{QM/MM}} = (1-\lambda)U_0^{\mathrm{QM/MM}} + \lambda U_1^{\mathrm{QM/MM}} \quad (4)$$

For the free energy calculations, 10 ns MD simulations for each λ window were done with U and modified U nucleosides in aqueous and ribosome phases using the same setup as for the classical MD simulations. Thirteen λ points were used for aqueous phases (λ = 0.00, 0.10, 0.20, 0.25, 0.30, 0.40, 0.50, 0.60, 0.70, 0.75, 0.80, 0.90, and 1.00) and eleven λ points were used for ribosome phase (λ = 0.00, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, and 1.00).

QM/MM Calculations. QM/MM single-point calculations were performed at the B3LYP/6-31G* level with the Q-Chem/CHARMM interface[44,45] on the snapshots taken every 100 ps from the last 9 ns of each λ window. The nucleobase U−cmo^{5}U was designated as the QM region with a linking hydrogen atom between the nucleobase and the ribose C1′. We checked that this level of theory and basis set (B3LYP/6-31G*) gave a good accuracy compared to the single-point ab initio calculations at MP2/6-31++G by calculating the QM relative energy of one system used in ab initio calculations and noted that the value of relative keto−enol energy only differs by 0.1 kcal/mol.

RESULTS AND DISCUSSION

Keto−Enol Equilibrium in Aqueous Environment. We perform quantum mechanical calculations for the keto and enol forms of uracil (U) and 5-oxyacetic uracil (cmo^{5}U) alone and when they are base paired with guanine (G). Uracil and guanine are used as simple models of uridine and guanosine, respectively, and the solvent effect is included using the polarizable continuum model (PCM). The major tautomeric form of uridine is the diketo form, and two enol tautomers are possible. We focus on the form involving O4 because it is the form proposed by the X-ray study[19] and previous theoretical studies on C5 substituents of uridine show that this enol form is more stable than that involving O2.[15]

In the crystal structure,[19] the 5-oxyacetic acid modification adopts a particular conformation, with the four atoms C5, O7, C8, and C9 co-planar with the pyrimidine ring, regardless of which base (uridine, cytidine, adenosine, or guanosine) is paired to the modified uridine. The dihedral angles *d*1(C6−C5−O7−C8) and *d*2(C5−O7−C8−C9) (Figure S2) of the modified uridine bound to guanosine in the X-ray crystal structure are 178.39 and 0.80°, respectively.

To examine whether the modified uracil (cmo^{5}U) is stable with the modification in this conformation, energies of the two forms of the modified uracil, cmo^{5}U$_{keto}$ and cmo^{5}U$_{enol}$, are calculated with the dihedral angle *d*1 fixed at 180°, whereas the dihedral angle *d*2 is held fixed at 30° intervals from −180 to +180° (Figure S2). cmo^{5}U$_{keto}$ is more stable than cmo^{5}U$_{enol}$ at all of the values of *d*2. It is also clear that *d*2 = 0° is the least-stable conformation for both cmo^{5}U$_{keto}$ and cmo^{5}U$_{enol}$. In the X-ray crystal structure, this unstable conformation of the modification may be stabilized by hydrogen bonds between the modified uridine and the neighboring nucleotides between O5 of the modified uridine and O2′ of a neighboring uridine, and

DOI: 10.1021/acs.jpcb.7b10878
J. Phys. Chem. B 2018, 122, 1152−1160

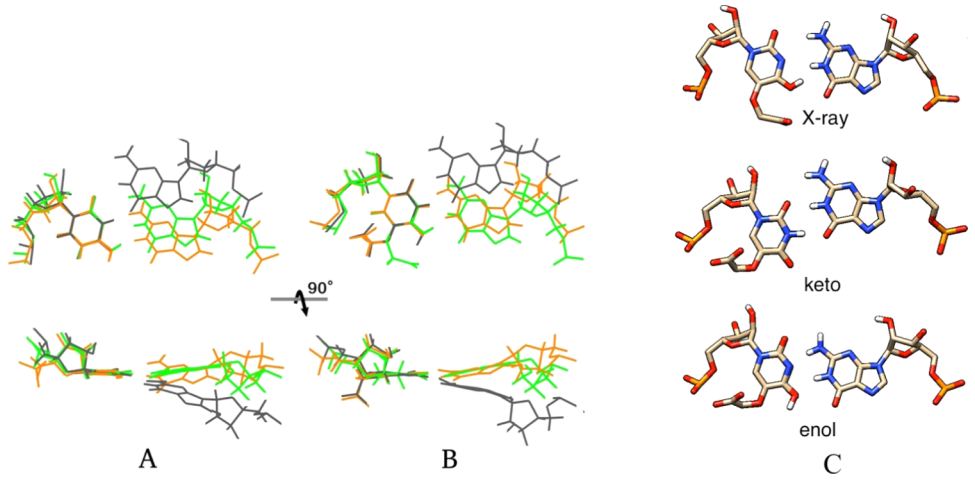


Figure 3. Representative snapshots at 5 ns from simulations of the uridine−guanine base pair in the ribosome (green: X-ray, gray: keto, orange: enol). Representation for (A) U−G and (B) cmo^{5}U−G obtained by overlapping the uracil ring (C) cmo^{5}U−G structures using CPK coloring.

between O8 of the modified uridine and N6 of a neighboring adenosine.

In our calculations, the dihedral angles *d*1 and *d*2 are fixed at 180 and 0°, respectively, to take into consideration the effects of the modification in this particular conformation. The results for the isolated base show that the keto form is always the most favorable form and the keto−enol equilibrium is only slightly affected by the modifications. Having the hydroxyl group pointing toward the modification stabilizes the enol structure by 5−6 kcal/mol compared to other enol forms (Figure 2).

In the geometry optimization, the base pairs with the keto tautomer of U shifted from the starting Watson−Crick geometry into a classic wobble geometry—a clear indication that the Watson−Crick geometry is not optimal for this tautomer. The energy calculations for the optimized geometries show this quantitatively (Figure 2). In the base pair between the unmodified U and G, two intermolecular hydrogen bonds are formed between U_{keto} and G and three intermolecular hydrogen bonds are formed between U_{enol} and G. Despite these, the U_{keto}−G pair is about 4 kcal/mol more stable than the U_{enol}−G pair (Figure 2). The energy difference between U_{keto}−G and U_{enol}−G is smaller than the energy difference between U_{keto} and U_{enol}, indicating that the keto−enol equilibrium is affected by the additional hydrogen bond in the base pair with G.

Base pairing of the modified U with G is similar to that of the unmodified U. The cmo$^5U_{keto}$−G pair is about 5 kcal/mol more stable than the cmo$^5U_{enol}$−G pair, which at room temperature corresponds to an abundance of about 5×10^{-4} for the cmo$^5U_{enol}$−G form. There is no major effect of the modification on the keto−enol stability.

Our results for the keto−enol equilibrium of U and cmo^{5}U using the PCM solvent model are comparable to the previous gas-phase results.[14] We found that U_{keto} is 10.87 kcal/mol (ΔE) more stable than U_{enol} in water at the MP2/6-31++G** level; and cmo$^5U_{keto}$ is 11.34 kcal/mol (ΔE) more stable than cmo$^5U_{enol}$ in water. Dąbkowska et al.[14] showed that U_{keto} is 12 kcal/mol (ΔE) more stable than U_{enol} in the gas phase at the B3LYP/6-31++G** level. Additionally, both studies showed that, even though the energy difference between U_{keto} and U_{enol} is decreased by interactions with a neighboring molecule, U_{keto} is still more stable than U_{enol}. A more recent study[15] also found that U_{keto} is more stable than U_{enol} by 10.8 kcal/mol in the water solution (at B3LYP-GD3/6-311++G(3df,2p)//B3LYP/6-31+G(d) level). The U−G interaction is stabilized by around 2 kcal/mol for the keto form of unmodified U and by 3 kcal/mol for U with an analogous modification in the 5 position (5-methoxy U).

Keto−Enol Equilibrium in the Ribosomal Decoding Center. To take into account the environment in the ribosomal decoding center, we run MD simulations with a ribosomal region centered at the A-site first anticodon position. The keto form of cmo^{5}U was recently parameterized in the CHARMM force field as part of a set of parameters for naturally occurring modified nucleotides.[37] The parameters for the enol form of cmo^{5}U and U are also included in this set precisely due to the speculation of the enol form from the structural observation.

The keto forms of U and cmo^{5}U converge to the expected wobble orientation (Figure 3A−C). However, for the enol forms, we encounter difficulties in accommodating a U−G Watson−Crick geometry with three hydrogen bonds, as proposed by the crystallographic study. When adding the hydroxyl hydrogen atom H4 to O4 to build the enol cmo^{5}U, we notice that the positions of cmo^{5}U O4 and guanine O6 in the experimental structure is such that the uracil enol group (O4−H4) and the hydrogen bond acceptor G O6 are not collinear (angle(O4−H4···O6) = 108.2°), which is not an optimal hydrogen bond geometry. Additionally, there might also be steric repulsion between guanine H1 and uracil H4.

We attempt various strategies to stabilize the Watson−Crick geometry. We try distance restraints based on the Watson−Crick hydrogen bond distances, but the geometry does not hold once the restraints are released. We attempt tuning the C−O−H angle parameter such that O4−H4···O6 are closer to collinearity. We also attempt these strategies in a reduced system containing just the tRNA anticodon stem loop (ASL) and mRNA. However, none of these strategies preserve the canonical Watson−Crick geometry with three hydrogen bonds. One persistent observation is that the hydroxyl H4 prefers to be oriented away from the Watson−Crick edge (dihedral *d*3(N3−C4−O4−H4) = 180°). We check the force field parameter for this dihedral (Figure S3), and 180° is not the global energy minimum. We thus investigate further to determine what causes the stabilization of this hydroxyl orientation.

DOI: 10.1021/acs.jpcb.7b10878
J. Phys. Chem. B 2018, 122, 1152−1160

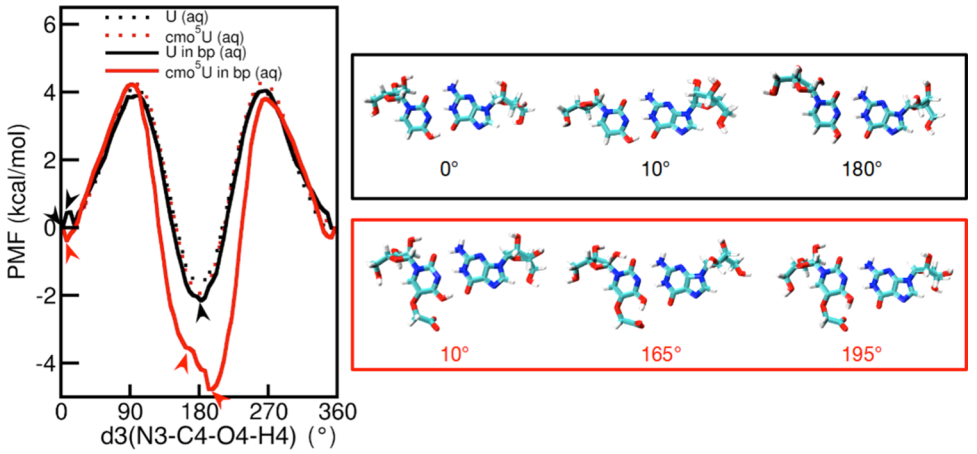


Figure 4. Potential of mean force along the dihedral N3−C4−O4−H4 of the uracil enol form in aqueous environment (U_{enol}: black and cmo^5U_{enol}: red). The dotted line is for the uracil base; the solid line is for the uracil base paired with G. Restraints were applied to maintain base pairing with G and the orientation of the modification moiety. Structural snapshots at indicated dihedral angles are shown to the right for unmodified U_{enol} and cmo^5U_{enol} in the upper and lower panels, respectively.

We perform umbrella sampling along *d*3 of U−cmo^5U for the isolated base as well as base paired to guanine in water. The gas-phase PMF is consistent with the profile of the model system 5-methoxy uracil used for the parameterization as expected (data not shown). The PMF profile in an aqueous solution shows that the global minimum is at 180° (Figure 4), consistent with what is observed in the MD simulation. Hydrogen bond analysis showed that at *d*3 = 180°, there is a high probability of a water-bridged hydrogen bond between N3 and O4 (Figure S4). When U is base paired with G, the global minimum is even lower in energy. Notably, the local minimum previously at 0° shifted slightly such that the minimum is at a configuration where the H4 atom is not co-planar with the rest of the base, indicating a steric constraint of the canonical Watson−Crick geometry.

From MD simulations, we observe that when the enol forms of U and cmo^5U are base paired to G in the ribosome, they do not stay in a typical Watson−Crick geometry with three hydrogen bonds (Figure 3). This geometry is halfway between the Watson−Crick and what Sochacka et al. call a new wobble orientation,[15] in which G shifts toward the major groove side instead of toward the minor groove side as in a typical wobble orientation (Figure 3). The simulated configurations can be accommodated in the experimentally calculated electron density (Figure S5) except for the position of the mutation (as previously discussed), indicating that a new wobble geometry for U−G base pair agrees with the crystallographic data.

To verify if the new wobble base pair geometry makes the enol base pair energetically more favorable in the solution than the keto, the new wobble orientations of U_{enol}−G and cmo^5U_{enol}−G, taken from the MD simulations, are geometry-optimized at the ab initio level (MP2/6-31++G**//MP2/3-21+G**). In the ab initio calculations, the conformation of the modification moiety of cmo^5U_{enol} is taken from the X-ray structure. In both U and cmo^5U, the enol form is still unfavorable compared to their respective keto forms (Figure 2). In addition, the enol form in the new wobble geometry is even less energetically favorable than in the Watson−Crick geometry because the latter has one more hydrogen bond. For U, the energy difference is 11.06 kcal/mol, whereas for cmo^5U, the intramolecular hydrogen bond reduces this energy gap to 2.14 kcal/mol. In the optimized geometry, the carboxyl group of the oxyacetic acid is rotated to around 90° with respect to the X-ray structure and forms an intramolecular hydrogen bond with the enol group.

To evaluate the effect of the ribosome on the relative stability of enol and keto forms, we evaluate the relative energy $\Delta E_{keto\rightarrow enol}$ at the QM/MM level in the decoding center and the relative keto−enol free energy between water and ribosomal environment. We geometry-optimize at the QM/MM level all the MD snapshots both for enol and keto forms accounting for the decoding center. The so-calculated average relative energies yield values in the same order of magnitude as values obtained at the MP2/6-31++G**//MP2/3-21+G** level for an isolated base pair in water solution. For U in the decoding center, this is 18.2 kcal/mol, whereas for cmo^5U, this is 5.7 kcal/mol (estimated error of ±10 kcal/mol). Free energy calculations are performed using BAR and QM-NBB direct and indirect approaches. The energy distribution overlap between the windows is always higher than 50% based on MM using BAR, and when the MM free energies are reweighted by the QM/MM contribution, the overlap drastically decreases to values lower than 13% for QM-NBB indirect (only for two reweighted windows) and becomes 0 for QM-NBB direct for all of the windows. The results show that the contribution of the ribosomal environment is around 1−3 kcal/mol in favor of enol in respect to water for unmodified U, whereas the $\Delta G_{keto\rightarrow enol}$ is almost the same for aqueous and ribosomal environments for cmo^5U (Table 1). The QM/MM reweighting of free energy does not affect the result in a relevant way. In general, the results show that the ribosome environment shifts the keto−

Table 1. Relative Free Energies of U and cmo^5U

	$\Delta\Delta G_{keto\rightarrow enol}$ [a]		
base	BAR	QM-NBB direct	QM-NBB indirect
U	−1.9 ± 0.5	−1.0 ± 1.3	−3.1 ± 0.6
cmo^5U	−0.7 ± 0.5	0.0 ± 1.2	0.2 ± 0.3

[a] $\Delta\Delta G_{keto\rightarrow enol} = \Delta G_{keto\rightarrow enol}$ (ribosome) − $\Delta G_{keto\rightarrow enol}$ (aq). All values are in kcal/mol. Error is estimated by calculating standard error of the mean of BAR/QM-NBB energies on three equal parts of divided trajectories.

DOI: 10.1021/acs.jpcb.7b10878
J. Phys. Chem. B 2018, 122, 1152−1160

enol equilibrium in favor of the enol form of U, but this is not enough to compensate the keto−enol electronic difference (around 5 kcal/mol) calculation for U in an aqueous solution. For cmo^{5}U, the ribosomal environment has no effect on the keto−enol equilibrium, keeping the keto form more favorable.

Alternative Hypotheses. MD simulations show that the enol hydroxyl H4 does not participate in hydrogen bonding, but is oriented away from the Watson−Crick edge. This is observed both in U and cmo^{5}U. The unmodified U does not have a modification moiety that might possibly interact with H4, whereas the cmo^{5}U modification moiety potentially can form an intramolecular hydrogen bond with H4, but this is not observed in the MD simulation. Instead, the oxyacetic acid group prefers to point away from the Watson−Crick edge toward the solvent environment. The observation that the modification does not maintain the orientation it has in the X-ray structure is also mentioned in a previous simulation study.[37] The simulations show that the hydrogen bond between U33 O2′ and the modification O5, as hypothesized in the X-ray study, is not enough to maintain the orientation. We note here that we assume that the modified acetic acid group is deprotonated, as is expected at neutral pH, but the local ribosomal environment might cause a pK_a shift, which could cause the acetic group to become protonated. In the future, it will be interesting to address this pK_a shift by using simulation techniques,[46] such as constant pH molecular dynamics,[47] that enable the direct coupling between tautomerization processes and conformational dynamics.

We also consider that the experimental orientation might be stabilized by interactions with the entities not visible at this resolution, such as ions or other ligands. For instance, a recent QM study showed how Mg^{2+} coordination affects the base/base pair charge distribution and geometries.[48] As a proof of concept that an ion could influence the conformation of the modification, we simulate the base pair cmo^{5}U−G with an Mg^{2+} ion close to the carboxyl oxygen and the guanine O6. After 5 ns, the modification moiety keeps a conformation close to what is observed in the X-ray structure (Figure S6).

We also note that the pK_a of uridine at the N3 protonation site is 9.3,[49] and a pK_a of 8.54 is observed for N3 of 5-methoxy U;[15] thus, the deprotonated U−cmo^{5}U at N3 may also be a plausible alternative consistent with the X-ray electron density. We attempt to construct a N3-deprotonated U−cmo^{5}U by redistributing the atomic partial charges without the usual validation steps of force field parameterization for a preliminary inquiry. However, we observe that N3-deprotonated U−cmo^{5}U does not form the Watson−Crick geometry, either, but new wobble geometry. Both the keto and the deprotonated forms deviate more from the crystallographic structure than the enol form (Figure S7).

CONCLUSIONS

We examined the hypothesis that the enol form of cmo^{5}U is consistent with the observed base pair geometry in the crystal structure. First, we addressed whether the modification lowers the keto−enol energy difference. Our ab initio calculations showed that in aqueous environment, the modification moiety of cmo^{5}U does not affect the keto−enol energy difference compared to that of U. The keto−enol energy difference is indeed lowered if an intramolecular hydrogen bond is formed, but in this case, the three hydrogen bonds in the Watson−Crick geometry cannot be maintained.

To take into account the ribosomal environment, we sampled the base pair in the ribosomal decoding environment, as well as performed free energy calculations. Molecular dynamics simulation showed that the enol form cannot maintain the Watson−Crick geometry. Instead, it formed a new wobble configuration with only two hydrogen bonds: the enol hydroxyl is not participating in the base pair interactions as would have been expected from the analogous C−G type Watson−Crick base pairing. Free energy calculation showed that the ribosomal environment does not lower the keto−enol energy difference in the presence of the modification, even when we used molecular mechanics description corrected with QM/MM calculations. For unmodified U, the keto−enol equilibrium is shifted toward the enol form of 2−3 kcal/mol in the ribosomal compared to the aqueous environment.

We conclude that the enol form of cmo^{5}U in the anticodon can be accommodated in the decoding center by base pairing with two hydrogen bonds with G at the third codon position. The observed base pair geometry is consistent with the crystallographic data, but the modification moiety prefers to be rotated by 180° with respect to the position in crystal structure. But, our results show that the keto form is thermodynamically more stable in the ribosomal environment. We have to note that we looked only at the possible tautomeric forms of cmo^{5}U, but we cannot exclude that the G at the codon position is in the enol form instead. We also assumed that the modification is deprotonated (conjugate base form), but it could be in its acid form.

Finally, we offered several hypotheses for future directions of inquiry such as a deprotonated form of cmo^{5}U and the presence of chemical entities, like Mg^{2+}, not resolved in the structure that may explain the position of the modification moiety.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.7b10878.

Atom numbering and dihedral definitions for U and cmo^{5}U, dihedral energy scans, water-bridge probabilities, and structural snapshots from the simulations (Figures S1−S7) (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: Lennart.Nilsson@ki.se. Tel: +46-8-52481099.

ORCID

Yossa Dwi Hartono: 0000-0002-7829-9247
Alessandra Villa: 0000-0002-9573-0326
Lennart Nilsson: 0000-0002-5067-6397

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by Nanyang Technological University Research Scholarship (Y.D.H.), and the Swedish Research Council (2015-04992).

DOI: 10.1021/acs.jpcb.7b10878
J. Phys. Chem. B 2018, 122, 1152−1160

REFERENCES

(1) Wohlgemuth, I.; Pohl, C.; Rodnina, M. V. Optimization of speed and accuracy of decoding in translation. *EMBO J.* **2010**, *29*, 3701−3709.

(2) Kramer, E. B.; Farabaugh, P. J. The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. *RNA* **2007**, *13*, 87−96.

(3) Johansson, M.; Zhang, J.; Ehrenberg, M. Genetic code translation displays a linear trade-off between efficiency and accuracy of tRNA selection. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 131−136.

(4) Ogle, J. M.; Brodersen, D. E.; Clemons, W. M., Jr.; Tarry, M. J.; Carter, A. P.; Ramakrishnan, V. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **2001**, *292*, 897−902.

(5) Ogle, J. M.; Murphy, F. V.; Tarry, M. J.; Ramakrishnan, V. Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell* **2002**, *111*, 721−732.

(6) Singh, V.; Fedeles, B. I.; Essigmann, J. M. Role of tautomerism in RNA biochemistry. *RNA* **2015**, *21*, 1−13.

(7) Kimsey, I. J.; Petzold, K.; Sathyamoorthy, B.; Stein, Z. W.; Al-Hashimi, H. M. Visualizing transient Watson−Crick-like mispairs in DNA and RNA duplexes. *Nature* **2015**, *519*, 315−320.

(8) Szymanski, E. S.; Kimsey, I. J.; Al-Hashimi, H. M. Direct NMR evidence that transient tautomeric and anionic states in dG· dT form Watson−Crick-like base pairs. *J. Am. Chem. Soc.* **2017**, *139*, 4326−4329.

(9) Singh, V.; Peng, C. S.; Li, D.; Mitra, K.; Silvestre, K. J.; Tokmakoff, A.; Essigmann, J. M. Direct observation of multiple tautomers of oxythiamine and their recognition by the thiamine pyrophosphate riboswitch. *ACS Chem. Biol.* **2014**, *9*, 227−236.

(10) Berdakin, M.; Féraud, G.; Dedonder-Lardeux, C.; Jouvet, C.; Pino, G. A. Excited states of protonated DNA/RNA bases. *Phys. Chem. Chem. Phys.* **2014**, *16*, 10643−10650.

(11) Rypniewski, W.; Banaszak, K.; Kuliński, T.; Kiliszek, A. Watson−Crick-like pairs in CCUG repeats: evidence for tautomeric shifts or protonation. *RNA* **2016**, *22*, 22−31.

(12) Maximoff, S. N.; Kamerlin, S. C. L.; 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.

(13) Satpati, P.; Åqvist, J. Why base tautomerization does not cause errors in mRNA decoding on the ribosome. *Nucleic Acids Res.* **2014**, *42*, 12876−12884.

(14) Dąbkowska, I.; Gutowski, M.; Rak, J. Interaction with glycine increases stability of a mutagenic tautomer of uracil. A density functional theory study. *J. Am. Chem. Soc.* **2005**, *127*, 2238−2248.

(15) Sochacka, E.; Lodyga-Chruscinska, E.; Pawlak, J.; Cypryk, M.; Bartos, P.; Ebenryter-Olbinska, K.; Leszczynska, G.; Nawrot, B. C5-substituents of uridines and 2-thiouridines present at the wobble position of tRNA determine the formation of their keto-enol or zwitterionic forms − a factor important for accuracy of reading of guanosine at the 3′-end of the mRNA codons. *Nucleic Acids Res.* **2017**, *45*, 4825−4836.

(16) Šponer, J. E.; Špačková, N.; Kulhánek, P.; Leszczynski, J.; Šponer, J. Non-Watson−Crick base pairing in RNA. Quantum chemical analysis of the cis Watson−Crick/sugar edge base pair family. *J. Phys. Chem. A* **2005**, *109*, 2292−2301.

(17) 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.

(18) Rozov, A.; Demeshkina, N.; Westhof, E.; Yusupov, M.; Yusupova, G. Structural insights into the translational infidelity mechanism. *Nat. Commun.* **2015**, *6*, No. 7251.

(19) Weixlbaumer, A.; Murphy, F. V.; Dziergowska, A.; Malkiewicz, A.; Vendeix, F. A. P.; Agris, P. F.; Ramakrishnan, V. Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. *Nat. Struct. Mol. Biol.* **2007**, *14*, 498−502.

(20) Kurata, S.; Weixlbaumer, A.; Ohtsuki, T.; Shimazaki, T.; Wada, T.; Kirino, Y.; Takai, K.; Watanabe, K.; Ramakrishnan, V.; Suzuki, T. Modified uridines with C5-methylene substituents at the first position of the tRNA anticodon stabilize U· G wobble pairing during decoding. *J. Biol. Chem.* **2008**, *283*, 18801−18811.

(21) Mitra, S. K.; Lustig, F.; Akesson, B.; Lagerkvist, U. Codon−anticodon recognition in the valine codon family. *J. Biol. Chem.* **1977**, *252*, 471−478.

(22) Hillen, W.; Lindner, E. E.; Joerg, H.; Gassen, H. G.; Vorbrüggen, H. 5-Methoxyuridine: the influence of 5-substituents on the keto-enol tautomerism of the 4-carbonyl group. *J. Carbohydr., Nucleosides, Nucleotides* **1978**, *5*, 23−32.

(23) Brooks, B. R.; Brooks, C. L.; MacKerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.

(24) MacKerell, A. D., Jr.; Banavali, N. K. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.* **2000**, *21*, 105−120.

(25) Foloppe, N.; MacKerell, A. D., Jr. All-Atom Empirical Force Field for Nucleic Acids: I . Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. *J. Comput. Chem.* **2000**, *21*, 86−104.

(26) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; MacKerell, A. D., Jr. 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.* **2011**, *32*, 1929−1943.

(27) Xu, Y.; Vanommeslaeghe, K.; Aleksandrov, A.; MacKerell, A. D., Jr.; Nilsson, L. Additive CHARMM force field for naturally occurring modified ribonucleotides. *J. Comput. Chem.* **2016**, *37*, 896−912.

(28) Roothaan, C. C. J. New developments in molecular orbital theory. *Rev. Mod. Phys.* **1951**, *23*, 69.

(29) Møller, C.; Plesset, M. S. Note on an approximation treatment for many-electron systems. *Phys. Rev.* **1934**, *46*, 618−622.

(30) Miertuš, S.; Tomasi, J. Approximate evaluations of the electrostatic free energy and internal energy changes in solution processes. *Chem. Phys.* **1982**, *65*, 239−245.

(31) Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilization of AB initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* **1981**, *55*, 117−129.

(32) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; et al. *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, 2004.

(33) Beglov, D.; Roux, B. Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J. Chem. Phys.* **1994**, *100*, 9050−9063.

(34) Allnér, O.; Nilsson, L.; Villa, A. Magnesium ion−water coordination and exchange in biomolecular simulations. *J. Chem. Theory Comput.* **2012**, *8*, 1493−1502.

(35) Brooks, C. L., III; Karplus, M. Deformable stochastic boundaries in molecular dynamics. *J. Chem. Phys.* **1983**, *79*, 6312.

(36) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(37) Allnér, O.; Nilsson, L. Nucleotide modifications and tRNA anticodon−mRNA codon interactions on the ribosome. *RNA* **2011**, *17*, 2177−2188.

(38) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.

(39) Nilsson, L. Efficient table lookup without inverse square roots for calculation of pair wise atomic interactions in classical simulations. *J. Comput. Chem.* **2009**, *30*, 1490−1498.

(40) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

DOI: 10.1021/acs.jpcb.7b10878
J. Phys. Chem. B 2018, 122, 1152−1160

(41) Boczko, E. M.; Brooks, C. L., III Constant-temperature free energy surfaces for physical and chemical processes. *J. Phys. Chem.* **1993**, *97*, 4509−4513.
(42) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245−268.
(43) König, G.; Hudson, P. S.; Boresch, S.; Woodcock, H. L. Multiscale free energy simulations: An efficient method for connecting classical MD simulations to QM or QM/MM free energies using non-Boltzmann Bennett reweighting schemes. *J. Chem. Theory Comput.* **2014**, *10*, 1406−1419.
(44) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; et al. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2015**, *113*, 184−215.
(45) Woodcock, H. L., III; Hodošček, M.; Gilbert, A. T.; Gill, P. M.; Schaefer, H. F.; Brooks, B. R. Interfacing Q-Chem and CHARMM to perform QM/MM reaction path calculations. *J. Comput. Chem.* **2007**, *28*, 1485−1502.
(46) Wallace, J. A.; Shen, J. K. Predicting pKa values with continuous constant pH molecular dynamics. *Methods Enzymol.* **2009**, *466*, 455−475.
(47) Goh, G. B.; Knight, J. L.; Brooks, C. L. Constant pH molecular dynamics simulations of nucleic acids in explicit solvent. *J. Chem. Theory Comput.* **2012**, *8*, 36−46.
(48) Halder, A.; Roy, R.; Bhattacharyya, D.; Mitra, A. How does Mg^{2+} modulate the RNA folding mechanism: A case study of the G:C W:W trans basepair. *Biophys. J.* **2017**, *113*, 277−289.
(49) Egli, M.; Saenger, W. *Principles of Nucleic Acid Structure*; Springer Science & Business Media, 2013.

DOI: 10.1021/acs.jpcb.7b10878
J. Phys. Chem. B 2018, 122, 1152−1160