THE JOURNAL OF PHYSICAL CHEMISTRY A
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCA

Article

Calculated pK_a Values for a Series of Aza- and Deaza-Modified Nucleobases

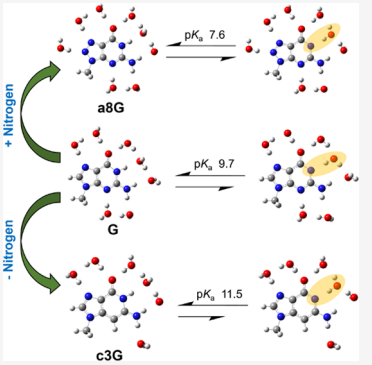
Alan J. Mlotkowski, H. Bernhard Schlegel, and Christine S. Chow*

Cite This: *J. Phys. Chem. A* 2023, 127, 3526−3534

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: A variety of synthetic modified nucleobases have been used to investigate the structure and function of RNA and DNA or act as enzyme inhibitors. A set of these modifications involves the addition or removal of a nitrogen atom in the ring. These aza and deaza modifications have garnered interest as useful biochemical tools, but information on some of their physical characteristics is lacking. In this study, the B3LYP density functional with the 6-31+G(d,p) basis set and an implicit−explicit solvent model was used to perform *ab initio* quantum mechanical studies to estimate pK_a values of aza- and deaza-modified nucleobases. A comparison between theoretical and known experimental pK_a values was carried out, and adjustment factors were applied to 57 pK_a values in the purine and pyrimidine data sets.



INTRODUCTION

An important characteristic of a nucleic acid is its acid−base properties, which can dictate both intra- and intermolecular interactions.[1−3] Each nucleobase within RNA or DNA has unique hydrogen-bond donor and acceptor capabilities due to its chemical composition and influences of the neighboring nucleotides.[4−6] The pK_a values of the acceptor and donor pairs influence the hydrogen-bonding strength,[7] which could impact the stability of nucleotide pairs within RNA or DNA structures.[2,3,8] Naturally occurring modifications have been reported to shift the pK_a values of nucleobases by as much as 3 pK_a units.[9] The use of RNA modifications as probes for higher-order structures could therefore have unintended effects on global or local RNA folding if the nucleobase pK_a values are altered.

To develop a greater understanding of RNA structure and function relationships and the impact of RNA modifications, a variety of techniques have been employed. Modified nucleosides have been used to probe RNA interactions and the roles of specific atoms within the nucleobase. Growing subsets of modified nucleobases that are useful in the nucleic acid field are aza- and deaza-modified bases (Figure 1). Deaza-modified nucleosides have been found in nature as either hypermodification scaffolds or biosynthetic precursors;[10−12] however, to the best of our knowledge, the deaza modifications have not been found in naturally occurring RNAs. Deazapurines and deazapyrimidines lack nitrogen atoms at specific positions of the nucleobase. As such, the physical properties of the nucleobase can be altered at the single-atom level. Single-atom mutagenesis has been carried out with many nucleic acids to gain a better understanding of their structure−

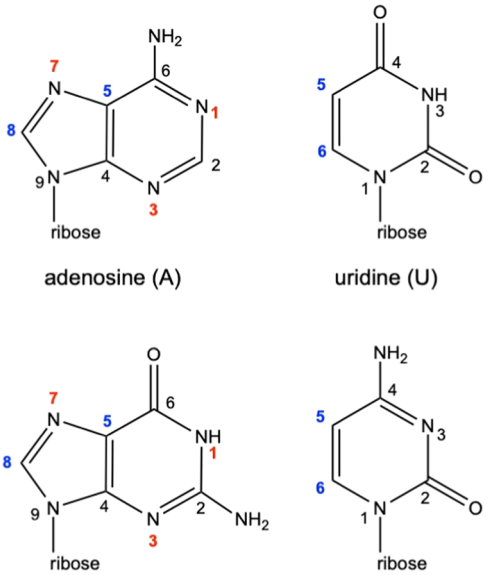
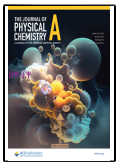


Figure 1. Structures of the four canonical nucleosides/nucleobases. The structures and numbering schemes of pyrimidines (uridine and cytidine) and purines (adenosine and guanosine) are shown. Positions with reported aza modifications are numbered in blue, while deaza-modified positions are numbered in red.

Received: February 27, 2023
Revised: March 26, 2023
Published: April 10, 2023



function relationships.[13−18] In contrast, the addition of a nitrogen atom to a purine nucleobase gives the azapurine class of modification. The addition of a nitrogen atom has led to unique pH-dependent fluorescent properties of the nucleobase while maintaining Watson−Crick (WC) base-pairing capabilities.[19,20] Azapyrimidines (occasionally referred to as triazine nucleosides) have demonstrated utility in probing immunosuppression and purine metabolism pathways.[21,22] Aza nucleosides also have reported anticancer,[23,24] antiviral,[25,26] and antibacterial[27] capabilities.

The aza and deaza sets of modifications are of high interest to the research community because of their ability to probe RNA or DNA interactions and mimic canonical nucleotides. Therefore, determining their physical and chemical characteristics and understanding their specific roles in these processes is important. One quantitative measure that is the focus of this study is the pK_a value. A few experiments have been conducted on aza and deaza nucleobases to determine their pK_a values,[16,17] but many of the synthetically available analogues are lacking experimental pK_a values, in particular, for the sites on the non-WC faces.

Computational chemistry has been shown to give good estimates of pK_a values for alcohols, phenols, and nucleobases.[9,28−30] In this current study, the B3LYP density functional with the 6-31+G(d,p) basis set was used in conjunction with a hybrid implicit−explicit solvation methodology to determine the pK_a values of 20 aza- and deaza-modified nucleobases. Every hydrogen-bonding position on the nucleobase was solvated with an explicit water molecule, and the entire structure was then surrounded by implicit continuum solvation. A two-parameter linear correction factor was generated by comparing our calculated theoretical pK_a values to experimentally determined ones and used to obtain adjusted pK_a values for this set of modified nucleobases.

COMPUTATIONAL METHODS

The Gaussian suite of programs[31] was utilized in the present study to calculate the free energies of aza- and deaza-modified nucleobases. The B3LYP density functional with the 6-31+G(d,p) basis set was utilized for the series of calculations along with the implicit−explicit solvation system previously developed.[28,29,32−34] For the aza- and deaza-modified nucleobases, a water molecule was placed at every position that could hydrogen bond, as done in previous studies with the natural modifications.[9] The bulk water in the environment was modeled by the SMD implicit solvation method of Marenich, Cramer, and Truhlar.[35] The specific locations and orientations of the water molecules surrounding each aza and deaza nucleobase structure are given in the Supporting Information. As with earlier reports, the sugar moiety was substituted with a methyl group to reduce the computational cost.[9,28]

Computational methods to determine pK_a values have been widely used,[36−38] and similar approaches were used here with the aza and deaza nucleobases. The theoretical pK_a values were determined by first finding the lowest energy structure of the neutral form of each modified nucleobase. Many different combinations of water molecule placements along with any varying structural conformations were examined, and the lowest energy structure was selected. This structure was then protonated or deprotonated, depending on the position of interest while maintaining the same solvation orientation and structural conformation as the neutral species. The calculated difference in free energy of the structures along with the experimental estimate of the free energy of the proton in aqueous solution (−270.297 kcal/mol) was used to determine the free energy for acid dissociation, $\Delta G_{\mathrm{dissoc(aq)}}$.[39,40] This value was then used to determine the pK_a values, as shown in eq 1.

$$\mathrm{p}K_a = \frac{\Delta G_{\mathrm{dissoc\ (aq)}}}{2.303RT} \tag{1}$$

A linear correlation factor was used to adjust the theoretical pK_a values. The adjustment is based on the line of best fit between the calculated pK_a values and nucleobases with experimentally determined pK_a values. Equations 2 and 3 were used for the pK_a value adjustment for purines and pyrimidines, respectively. The plots of these linear regressions include values for the unmodified nucleobases along with naturally occurring modifications reported previously.[9]

$$\text{adjusted p}K_a = \frac{(\text{calculated p}K_a + 0.06)}{1.16} \tag{2}$$

$$\text{adjusted p}K_a = \frac{(\text{calculated p}K_a + 0.54)}{1.19} \tag{3}$$

RESULTS AND DISCUSSION

Determination of Nucleobase pK_a Values. Applying the B3LYP/6-31+G(d,p) level of theory with a combined implicit−explicit solvation method has been shown previously to give accurate pK_a values of methyl-substituted nucleobases.[28] This method was also used to determine the pK_a values of a series of naturally occurring modified nucleobases with good agreement between computed and experimental values.[9] Consistent methodologies were used for each aza or deaza nucleobase, with substitution of a methyl group for the ribose or deoxyribose sugar and varying the number of associating water molecules with each nucleobase. Each aza or deaza nucleobase had a unique number of potential hydrogen-bonding interactions because of the atom substitutions. Therefore, when each position of the nucleobase was explicitly solvated, the numbers of water molecules varied. A few of the modified nucleobases could not accommodate explicit water molecules at all possible hydrogen-bonding positions within the structure. During geometry optimization, some of the water molecules would occasionally drift away from the expected positions. Such events occurred more often with structures that had water molecules "crowding" each other. To overcome this challenge, problematic water molecules were removed, and the structures were reoptimized. The optimized structures with the explicit water molecule orientations and locations are provided in the Supporting Information.

As with a previous study,[9] the known experimental and theoretical pK_a values were compared to obtain a linear adjustment factor for the calculated results. This adjustment was done to reduce the systematic errors that arise from using a combination of theoretical methods. The applied adjustment factors are different for the purine and pyrimidine data sets, which accounts for the larger variations in their chemical and electronic structures. The adjustment employed values from unnatural modified nucleobases in the present work and natural modifications in prior work.[9] The previous work had 12 and 14 experimental data points for purines and pyrimidines, respectively, that were used to obtain the adjustment parameters. The present study includes 12 and 3 additional experimental values for purines and pyrimidines, respectively,

https://doi.org/10.1021/acs.jpca.3c01358
J. Phys. Chem. A 2023, 127, 3526−3534

for a total of 24 and 17 data points used for the adjustments. As observed previously, the adjusted theoretical data show a similar trend to that of the experimental pK_a values. The mean absolute error (MAE) obtained for both the purine and pyrimidine data sets is 0.6. From 24 different unmodified and modified methyl-substituted nucleobases, 57 pK_a values are reported, of which only 20 had previously known experimental values.

For the linear adjustments, the deaza and aza pK_a values were added to previously reported linear adjustments to strengthen the correlations. The additional values in the plots lead to slight changes in the slopes and y-intercepts. As such, the pK_a values for the unmodified nucleobases differ from the values reported previously.[9] The adenosine N1 pK_a value is now closer and the guanosine value is further from the experiment values (both by 0.3 units). The uridine and cytidine pK_a values are comparable to the previous report.[9] Although these differences in the values are small, they do indicate limitations of the applied method. Despite these limitations, our reported pK_a values fall within the MAEs, and the changes in the pK_a values resulting from modifications can be still elucidated with confidence.

Purines. A total of 13 different methyl-substituted purine deaza/aza modifications were examined. Many previous pK_a studies on nucleobases focused on the WC face (N1 position), but tertiary structures of RNA include other interactions (e.g., N3 and N7 positions). Therefore, we studied all three positions of interest in these modified purines (N1, N3, and N7). Additionally, the pK_a values of position 8 were determined for six of the purines that had a nitrogen substitution (i.e., N8) and can therefore be protonated at this site. By plotting the experimental pK_a data against the calculated values (Figure 2), a small systematic 0.06 pK_a unit

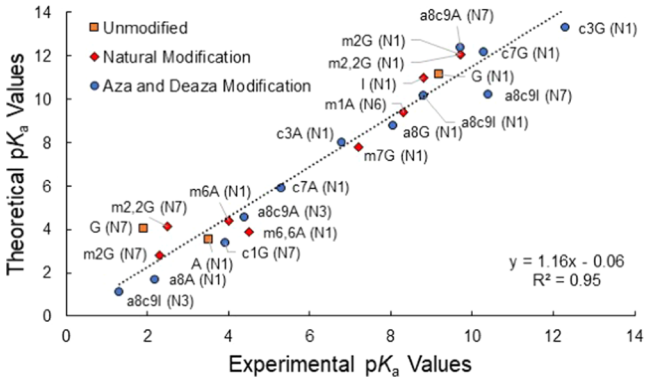


Figure 2. Plot and linear regression fit for the theoretical and experimental pK_a values of the purine aza/deaza modifications. The symbols indicate unmodified nucleobases (orange squares), natural modifications (red diamonds),[9] and aza/deaza modifications (blue circles). The MAE for this set of data is 0.6.

shift for the theoretical data is observed. A two-parameter adjustment for the purine modifications was used, and a linear fit of $1.16x - 0.06$ with an R^2 value of 0.95 was obtained.

Applying the adjustment to the theoretical values, a MAE of 0.6 pK_a units was obtained for the data shown in Figure 2, indicating good agreement between the experimental and calculated pK_a values (Tables 1 and 2) for the modified purine nucleobases (structures are shown in the Supporting Information). The pK_a values that were calculated to be less than −1.0 are labeled as "unfavorable" because protonation at those positions would not be energetically favorable. The trend in the adjusted pK_a values for the N1 position of adenosine modifications is as follows: a8A ≪ formycin A < A < a8c7A ≪ c7A ≪ c3A (in the range of 1.5−7.0) (Table 1). The corresponding pK_a trend for the N1 position of guanosine modifications is as follows: a5c7G ≪ a8G ≪ formycin B ≈ a8c7G < G < c7G < c3G (in the range of 2.5−11.5) (Table 2).

Adenosine Modifications. The deazaadenosine modifications examined in this study have higher pK_a values than the parent adenosine (Table 1). The biggest difference in the pK_a value is observed for the N1 position of c3A (ΔpK_a +3.9; ΔpK_a = N1 pK_a of c3A−N1 pK_a of A), which will exist in the protonated state at neutral pH. The c7A modification has a smaller effect at N1 (ΔpK_a +2.1). RNA melting studies showed that both c3A and c7A have destabilizing effects on base pairing with U, but with c7A being less destabilizing than c3A,[16] mirroring the trend in the ΔpK_a values. For the N3 position, the largest difference in the pK_a value is observed for the c1A modification (ΔpK_a +4.1), with a smaller difference for c7A (ΔpK_a +2.4). The pK_a values for the N7 position of c3A and c1A are also higher compared to A (ΔpK_a +2.0 and +2.6, respectively). With its higher N7 pK_a value along with a lack of hydrogen-bonding capabilities at position 1, changes in RNA stability and/or structure could result from A to c1A substitutions. For example, preferences could shift from WC to Hoogsteen pairing, along with impacts on Hoogsteen pairing strength.

The 8-azaadenosine (a8A) nucleobase (Figure 3) has a nitrogen atom at position 8. This modification has been used as a probe to study the catalytic roles of A residues in ribozymes, with focus on general acid/base mechanisms.[45] The computed pK_a values for the N1 and N3 positions are 1.5 and −0.5, respectively. Compared to A, the N1 pK_a value is lower by 1.6 units. The pK_a values for N7 and N8 are less than −1.0; therefore, protonation of these positions is unfavorable. The pK_a values for the a8A modification are useful to know because this nucleobase has been reported to exhibit moderate fluorescence when in the protonated state and comparatively weaker fluorescence in the neutral state.[46] Also of note are thermal melting studies that revealed that a8A destabilizes WC base pairs within the context of a DNA duplex.[47] This result with a8A supports the "ΔpK_a rule" in biological systems, in which a larger difference in the pK_a values of a donor and acceptor correlates with weaker hydrogen bonding.[7] In the case of a8A, the decrease in the N1 pK_a value compared to A would be expected to destabilize the duplex. However, factors other than pK_a (e.g., changes in hydration) also need to be considered when examining impacts of base modification on DNA or RNA stabilities.[16]

For the combined modification a8c7A with both aza and deaza substitutions (Figure 3), the N1 and N3 pK_a values are close to those of A (ΔpK_a +0.3 and 0, respectively). Protonation of the N8 position of a8c7A is possible but only under highly acidic conditions (pK_a value of −0.5). Because the pK_a values of A and a8c7A are found to be similar, these results suggest that a8c7A could serve as a good substitute for A in nucleic acids. Indeed, thermal melting studies with DNAs containing either a8c7A or A at the same locations revealed similar stabilities.[48]

Formycin A, a purine analogue with antibacterial properties, also contains aza (C8) and deaza (N9) substitutions (i.e., a8c9A) (Figure 3). Despite having the aza and deaza substitutions, the N1 pK_a value of formycin A is close to

https://doi.org/10.1021/acs.jpca.3c01358
J. Phys. Chem. A 2023, 127, 3526−3534

Table 1. Experimental and Theoretical pK_a Values for Modified Deaza/Azaadenosines

modification	abbreviation	position	experimental pK_a value[a]	adjusted theoretical pK_a value[b]
8-azaadenosine	a8A	N1	2.2[19]	1.5
		N3		−0.5
		N7		unfavorable
		N8		unfavorable
formycin A	a8c9A	N1		2.9
		N3	4.4[41]	4.0
		N7	9.7[41]	10.7
		N8		0.4
adenosine	A	N1	3.5[42]	3.1
		N3		0.8
		N7		0.9
8-aza-7-deazaadenosine	a8c7A	N1		3.4
		N3		0.8
		N8		−0.5
7-deazaadenosine	c7A	N1	5.3[16]	5.2
		N3		3.2
3-deazaadenosine	c3A	N1	6.8[16]	7.0
		N7		2.9
1-deazaadenosine	c1A	N3		4.9
		N7		3.5

[a]The pK_a values were determined by UV titration[19,42] and NMR spectroscopy[16,41] with typical errors of 0.1−0.2 pK_a units. [b]The pK_a values less than −1.0 are denoted as unfavorable.

Table 2. Experimental and Theoretical pK_a Values for Modified Deaza/Azaguanosines

modification	abbreviation	position	experimental pK_a value[a]	adjusted theoretical pK_a value[b]
5-aza-7-deazaguanosine	a5c7G	N1		2.5
		N3		unfavorable
8-azaguanosine	a8G	N1	8[19]	7.6
		N3		unfavorable
		N7		unfavorable
		N8		unfavorable
formycin B	a8c9I	N1	8.8[43]	8.8
		N3	1.3[43]	1.0
		N7	10.4[43]	8.9
		N8		−0.5
8-aza-7-deazaguanosine	a8c7G	N1		8.9
		N3		0.7
		N8		−0.2
guanosine	G	N1	9.2[44]	9.7
		N3		−0.6
		N7	1.9[44]	3.5
7-deazaguanosine	c7G	N1	10.3[16]	10.6
		N3		0.7
3-deazaguanosine	c3G	N1	12.3[16]	11.5
		N7		3.9
1-deazaguanosine	c1G	N3		0.5
		N7	3.93[17]	3.0

[a]The pK_a values were determined by UV titration[17,19] and NMR spectroscopy[16,43,44] with typical errors of 0.1−0.2 pK_a units. [b]The pK_a values less than −1.0 are denoted as unfavorable.

that of A (ΔpK_a −0.2). In contrast, the N7 position of formycin A contains a unique imino proton with a pK_a value of 10.7. This feature may be important for formycin A activity, as it has been reported that tautomerization resulting in protonation of the N8 position rather than N7 is favored when the nucleoside is bound in the *Escherichia coli* purine nucleoside phosphorylase binding pocket.[49] While the modest shift in the N1 pK_a value would have minimal impact on WC interactions within nucleic acids, the difference at N7 could alter interactions on the Hoogsteen edge. As such, the incorporation of a8c9A into RNA structures might be useful for probing noncanonical interactions. Impacts on RNA tertiary interactions would need to be considered when using a8c9A as a tool for structure analysis.

Guanosine Modifications. The calculated pK_a shifts for the deazaguanosine modifications are variable, displaying both higher and lower values compared to the unmodified guanosine (Table 2). Each pK_a value of the modified

https://doi.org/10.1021/acs.jpca.3c01358
J. Phys. Chem. A 2023, 127, 3526−3534

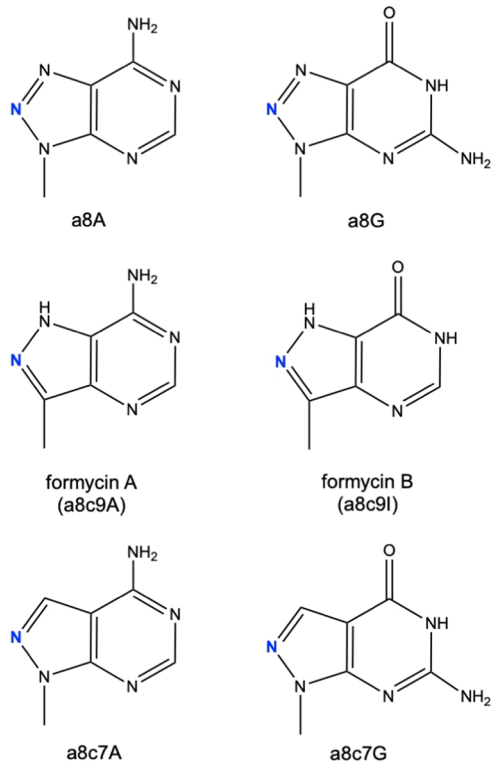


Figure 3. Structures of aza-modified nucleobases. The modified positions are highlighted in blue. The deaza/aza nomenclature is given for the purine analogues formycin A and B (a8c9A and a8c9I, respectively).

nucleobases was compared to that of the parent G. Both c3G and c7G have larger pK_a values at the N1 position (ΔpK_a +1.8 and +0.9, respectively). The N3 position pK_a value is also greater for c1G and c7G (ΔpK_a +1.1 and +1.3, respectively). The N7 pK_a value for c3G is slightly higher (ΔpK_a +0.4), whereas the value for c1G is slightly lower (ΔpK_a −0.5). Changes in pK_a values can influence base-pairing preferences, resulting in downstream structural changes of nucleic acids. With its modest increase in the pK_a value at the N1 position compared to G, c7G is a good mimic for WC interactions. This modification slightly destabilizes canonical base pairing,[16] thus providing another example that is consistent with the ΔpK_a rule.[7] The c7G nucleobase may also be a useful mimic for non-WC pairing such as sugar-edge interactions at the N3 position, although the N3 pK_a value varies by 1.3 units. In contrast, c3G would be expected to have similar behavior as G at the N7 position, whereas the altered pK_a value at the N1 position is predicted to impact canonical base pairing. Indeed, melting studies showed that c3G is more destabilizing than c7G in RNA duplexes.[16,50]

As with the aza modification a8A, incorporation of a nitrogen atom at the C8 position in guanosine (a8G) (Figure 3) leads to different pH-dependent fluorescence properties. In this case, the a8G nucleobase exhibits higher fluorescence in the deprotonated state compared to the neutral state.[46] The pK_a value for the N1 position of a8G is lower by 2.1 units compared to G. Since the N1 pK_a value for a8G is close to physiological pH, this modified nucleobase is a useful probe to determine the roles of protonated states of guanosine residues in enzyme and ribozyme catalysis.[51] As with a8A, protonation of the N3, N7, and N8 positions of a8G is unfavorable. Previous studies showed that oligonucleotides containing a8G form more stable base pairs with C compared to the parent G, while G quartet formation with a8G is disfavored.[52] The observed lower N1 pK_a value of a8G compared to G is consistent with the more stable pairing with C, whereas the altered N7 pK_a value would disfavor tertiary structures such as G quartets.

The pK_a values of mixed aza and deazaguanosine modifications display the biggest variability. The N1 of a5c7G has a ΔpK_a of −7.2 when compared to G. The N1 position of a5c7G is deprotonated at physiological pH, a major difference from the unmodified nucleobase (Figure 4). In

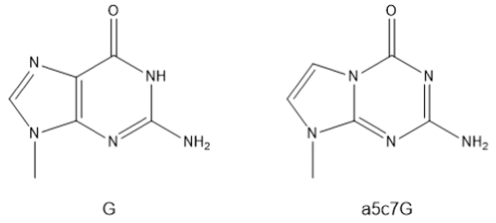


Figure 4. Structures of G and a5c7G with different protonation status at N1.

contrast, the N1 pK_a value of a8c7G (Figure 3) is closer to that of G (ΔpK_a −0.8). The N3 pK_a value of a8c7G is higher compared to that of G (ΔpK_a +1.3). The modest decrease of the N1 pK_a value is consistent with the small increase in duplex DNA base-pair stability that was reported.[48] Formycin B (a8c9I; I is inosine) is structurally similar to a8G (Figure 3), but this nucleobase lacks the amine at the N2 position and contains a carbon at position 9 rather than nitrogen. The pK_a changes for a8c9I mirror those observed with formycin A; the N1 pK_a is lower (ΔpK_a −0.9), and the N7 value is higher (ΔpK_a +5.4). As with formycin A, the N7 position of formycin B follows a similar trend with the pK_a value being significantly more basic than that of G.

Pyrimidines. A total of seven different methyl-substituted pyrimidine aza modifications were examined. Of the 13 positions studied, only three had previously reported pK_a values. The primary site of interest in the pyrimidine series is the N3 position, although certain modifications result in additional N-containing sites such as N5 and N6. The experimental pK_a data were plotted against the calculated values (Figure 5). A two-parameter adjustment for the pyrimidine modifications was used, and a linear fit of $1.19x - 0.54$ with an R^2 value of 0.97 was obtained.

The adjusted theoretical pK_a values resulted in a MAE of 0.6 for the data shown in Figure 5. The individual errors are larger at the extreme ends of the pK_a range, possibly due to the lack of experimental pK_a values less than 4 and greater than 10. Positions that highly disfavor protonation were determined to have negative adjusted theoretical pK_a values (pK_a values less than −1.0 are listed as unfavorable in Table 3). The trend in the pK_a values for the N3 position for uridine modification is as follows: a6U < a6T < a5U ≪ U ≪ DHaU ≈ DHaT (in the range of 6.6−11.9) and the trend for cytidine is as follows: a5C ≪ C ≪ DHaC (in the range of 1.4−6.4) (Table 3).

Both azauridine modifications (Figure 6) have lower pK_a values for the N3 position compared to U (ΔpK_as −1.8 for a5U and −3.1 for a6U). Protonation of the N5 and N6 positions in a5U and a6U, respectively, is unfavorable. The a6T modification also has a lower pK_a value at the N3 position compared to U (ΔpK_a −2.1) and unfavorable protonation at

https://doi.org/10.1021/acs.jpca.3c01358
J. Phys. Chem. A 2023, 127, 3526−3534

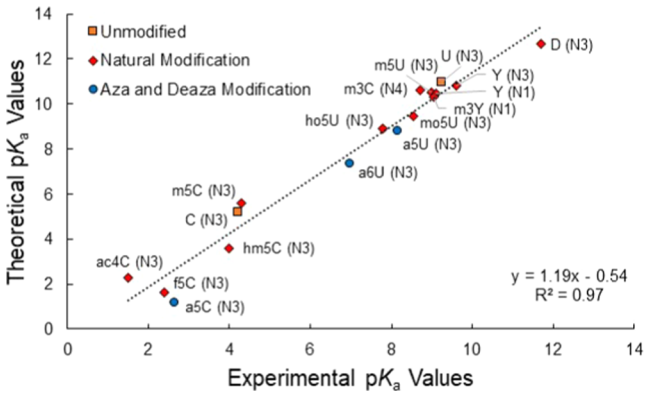


Figure 5. Plot and linear regression fit for the theoretical and experimental pK_a values of the aza modifications of the pyrimidine data set. The symbols indicate unmodified nucleobases (orange squares), natural modifications (red diamonds),[9] and aza modifications (blue circles). The MAE for this set of data is 0.6.

the N6 position. In this case, oligonucleotides containing a6T have decreased stability.[57]

Saturation of a5U at the N5−C6 position leads to generation of a second imino group (DHaU, Figure 6). As such, the N5 of a5U is acidic, whereas the corresponding N5 of DHaU is basic with a concomitant higher pK_a value for the N3 imino group compared to that of U (ΔpK_a +2.1). Saturation of the analogous a6T leads to DHaT. The N3 position of DHaT is also observed to be more basic with a pK_a value of 11.9.

The DHaU nucleobase can participate in nonstandard interactions. Structurally related to pseudouridine, DHaU has an imino group on the Hoogsteen face (N5) that can participate in unique interactions such as water-mediated hydrogen bonding observed with the N1 in pseudouridine-containing RNAs.[58−60] Such interactions are predicted to stabilize RNA duplexes. The N5 position of DHaU has a significantly higher pK_a value (12.6) than the corresponding N1 position of pseudouridine (9.2), which could further stabilize interactions with water molecules.

As with the aza modifications on U, the 5-aza modification of C and the corresponding saturated nucleobase have shifted pK_a values. The pK_a value of the N3 position of a5C is lower

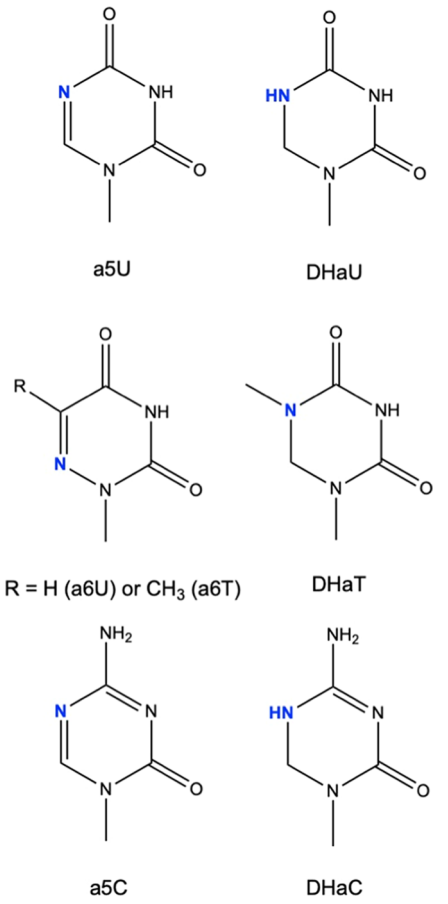


Figure 6. Structures of aza-modified pyrimidines. Saturation of the N5−C6 double bond of a5U and a5C results in a change in N5 protonation.

by 3.4 units compared to that of C. The added nitrogen at position 5 is more acidic than the N3. This acidic pK_a is reminiscent of the shifted N3 pK_a values of several purines mentioned earlier.

Saturation of the N5−C6 double bond of a5C (DHaC, Figure 6) leads to a higher N3 pK_a value compared to that of C

Table 3. Experimental and Theoretical pK_a Values for Modified Deaza/Azapyrimidines

modification	abbreviation	position	experimental pK_a value[a]	adjusted theoretical pK_a value[b]
6-azauridine	a6U	N3	6.99[53]	6.6
		N6		unfavorable
6-azathymidine	a6T	N3		7.6
		N6		unfavorable
5-azauridine	a5U	N3	8.15[53]	7.9
		N5		unfavorable
uridine	U	N3	9.25[54]	9.7
5,6-dihydroazauridine	DHaU	N3		11.8
		N5		12.6
5,6-dihydro-5-azathymidine	DHaT	N3		11.9
5-azacytidine	a5C	N3	2.64[55]	1.4
		N5		−0.7
cytidine	C	N3	4.2[56]	4.8
5,6-dihydroazacytidine	DHaC	N3		6.4
		N5		11.8

[a]The pK_a values were determined by UV titration[53−56] with typical errors of 0.1−0.2 pK_a units. [b]The pK_a values less than −1.0 are denoted as unfavorable.

https://doi.org/10.1021/acs.jpca.3c01358
J. Phys. Chem. A 2023, 127, 3526−3534

(ΔpK_a +1.6). The N5 imino is basic with a pK_a value of 11.8. In contrast to the N3 pK_a value of unsaturated a5C, the DHaC N3 position is partially protonated at physiological pH. The aza-C series of modifications has been used to probe various enzymatic activities and cellular processes. For example, DNA methylase activity is lower in cells treated with a5C, whereas hypomethylation in cells treated with DHaC is less apparent.[61] Additionally, the N5 of DHaC was observed to participate in a water-mediated hydrogen bond, which could play a role in its inhibition of the DNA (cytosine C5)-methyltransferase enzyme.[62] In contrast, the protonated N3 of DHaC was reported to have a weaker interaction with a glutamic acid chain[62] compared to the unmodified counterpart. This observation supports the ΔpK_a rule,[7] in which the greater pK_a value of N3 of DHaC compared to that of C would correspond to a weaker interaction with the amino acid side chain.

CONCLUSIONS

In this study, a computational method was employed to provide pK_a values for a series of 20 nucleobases with aza and deaza modifications. The list of computationally determined pK_a values for modified nucleobases was expanded using an implicit−explicit hybrid solvation system with the B3LYP density functional and the 6-31+G(d,p) basis set. A comparison between theoretical and known experimental pK_a values was carried out, and adjustment factors for purine and pyrimidine data sets were derived. The comparison yielded R^2 values of 0.95 (MAE of 0.6) and 0.97 (MAE of 0.6) for purines and pyrimidines, respectively. A total of 57 adjusted theoretical pK_a values were obtained. This method provides reliable pK_a values, although we observe larger discrepancies between the experimental and theoretical values at lower (pK_a < 4) and higher (pK_a > 10) ranges. The larger availability of experimental pK_a values in the 4−10 range correlates with higher accuracy with the adjusted theoretical pK_a values. More experimentally determined values at lower and higher ranges would strengthen the calculations and reduce discrepancies between the two methods.

Aza and deaza modifications are highly useful for probing interactions and functions of specific atoms in DNA and RNA systems.[16−18] Not surprisingly, the impact of the modification on the nucleobase pK_a value is dictated by its location. Furthermore, when specific aza/deaza substitutions are made, the pK_a values of multiple functional groups within the nucleobase can be impacted. In this study, the observed trend is that deaza modifications increase nucleobase pK_a values compared to those of the unmodified counterparts, whereas aza modifications decrease the pK_a values. In contrast, although mixed aza and deaza modifications deviate structurally from their unmodified counterparts, the pK_a values of these modified nucleobases tend to be closer to those of the parent nucleosides. This feature makes the mixed modifications ideal mimics for canonical nucleosides.

Application of the "ΔpK_a rule" in the aza/deaza series of modifications was often consistent with base-pair stabilities in studies with duplex DNAs or RNAs. However, these observations do not take into account other factors such as base stacking, solvation effects, or interactions with metal ions that may impact base-pair stability.[63] Water-mediated backbone interactions may also impact stability as observed with pseudouridylated RNAs.[64,65] Similar interactions are likely to occur with DHaC and DHaU, thus impacting stabilities of nucleic acids containing these modifications. While qualitative correlations between base-pair stabilities and the differences of the pK_a values for aza and deaza nucleobases have been observed, more in depth studies are needed to understand the physical and chemical contributions that impact the nucleic acid structure.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.3c01358.

Optimized structures and Cartesian coordinates of methyl-substituted nucleobases (PDF)

Details on calculations of pK_a values (XLSX)

AUTHOR INFORMATION

Corresponding Author

Christine S. Chow − *Department of Chemistry, Wayne State University, Detroit, Michigan 48202, United States;* orcid.org/0000-0002-5538-4553; Email: cchow@wayne.edu

Authors

Alan J. Mlotkowski − *Department of Chemistry, Wayne State University, Detroit, Michigan 48202, United States;* orcid.org/0000-0002-7384-6432

H. Bernhard Schlegel − *Department of Chemistry, Wayne State University, Detroit, Michigan 48202, United States;* orcid.org/0000-0001-7114-2821

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.3c01358

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Evan Jones for helpful discussions. This material is based upon work supported by (while C.S.C. was serving at) the U.S. National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF. This work was supported by grants from the National Science Foundation (CHE1856437 to H.B.S.; CHE1904754 supported A.J.M.). The Wayne State University High Performance Computing Services provided computational support.

REFERENCES

(1) Kapinos, L. E.; Operschall, B. P.; Larsen, E.; Sigel, H. Understanding the Acid−Base Properties of Adenosine: the Intrinsic Basicities of N1, N3 and N7. *Chem. − Eur. J.* **2011**, *17*, 8156−8164.

(2) Acharya, P.; Cheruku, P.; Chatterjee, S.; Acharya, S.; Chattopadhyaya, J. Measurement of Nucleobase pK_a Values in Model Mononucleotides Shows RNA-RNA Duplexes to be More Stable than DNA-DNA Duplexes. *J. Am. Chem. Soc.* **2004**, *126*, 2862−2869.

(3) Tang, C. L.; Alexov, E.; Pyle, A. M.; Honig, B. Calculation of pK_as in RNA: on the Structural Origins and Functional Roles of Protonated Nucleotides. *J. Mol. Biol.* **2007**, *366*, 1475−1496.

(4) Xia, T.; SantaLucia, J., Jr.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X.; Cox, C.; Turner, D. H. Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation

of RNA Duplexes with Watson−Crick Base Pairs. *Biochemistry* **1998**, *37*, 14719−14735.
(5) Richardson, K. E.; Znosko, B. M. Nearest-Neighbor Parameters for 7-Deaza-Adenosine·Uridine Base Pairs in RNA Duplexes. *RNA* **2016**, *22*, 934−942.
(6) Serra, M. J.; Lyttle, M. H.; Axenson, T. J.; Schadt, C. A.; Turner, D. J. RNA Hairpin Loop Stability Depends on Closing Base Pair. *Nucleic Acids Res.* **1993**, *21*, 3845−3849.
(7) Shan, S.-o.; Loh, S.; Herschlag, D. The Energetics of Hydrogen Bonds in Model Systems: Implications for Enzymatic Catalysis. *Science* **1996**, *272*, 97−101.
(8) Šponer, J.; Leszczynski, J.; Hobza, P. Electronic Properties, Hydrogen Bonding, Stacking, and Cation Binding of DNA and RNA Bases. *Biopolymers* **2001**, *61*, 3−31.
(9) Jones, E. L.; Mlotkowski, A. J.; Hebert, S. P.; Schlegel, H. B.; Chow, C. S. Calculations of pK_a Values for a Series of Naturally Occurring Modified Nucleobases. *J. Phys. Chem. A* **2022**, *126*, 1518−1529.
(10) Morris, R. C.; Elliot, M. S. Queuosine Modification of tRNA: A Case for Convergent Evolution. *Mol. Genet. Metab.* **2001**, *74*, 147−159.
(11) McCarty, R. M.; Somogyi, Á.; Lin, G.; Jacobsen, N. E.; Bandarian, V. The Deazapurine Biosynthetic Pathway Revealed: In Vitro Enzymatic Synthesis of PreQ$_0$ from Guanosine 5′-Triphosphate in Four Steps. *Biochemistry* **2009**, *48*, 3847−3852.
(12) McCarty, R. M.; Bandarian, V. Biosynthesis of Pyrrolopyrimidines. *Bioorg. Chem.* **2012**, *43*, 15−25.
(13) Polacek, N. Atomic Mutagenesis of the Ribosome: Towards a Molecular Understanding of Translation. *Chimia* **2022**, *67*, 322−326.
(14) Lang, K.; Erlacher, M.; Wilson, D. N.; Micura, R.; Polacek, N. The Role of 23S Ribosomal RNA Residue A2451 in Peptide Bond Synthesis Revealed by Atomic Mutagenesis. *Chem. Biol.* **2008**, *15*, 485−492.
(15) Hoernes, T. P.; Clementi, N.; Juen, M. A.; Erlacher, M. D.; et al. Atomic Mutagenesis of Stop Codon Nucleotides Reveals the Chemical Prerequisites for Release Factor-Mediated Peptide Release. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E382−E389.
(16) Bereiter, R.; Himmelstoß, M.; Renard, E.; Mairhofer, E.; Egger, M.; Breuker, K.; Kreutz, C.; Ennifar, E.; Micura, R. Impact of 3-Deazapurine Nucleobases on RNA Properties. *Nucleic Acids Res.* **2021**, *49*, 4281−4293.
(17) Bereiter, R.; Renard, E.; Breuker, K.; Kreutz, C.; Ennifar, E.; Micura, R. 1-Deazaguanosine-Modified RNA: The Missing Piece for Functional RNA Atomic Mutagenesis. *J. Am. Chem. Soc.* **2022**, *144*, 10344−10352.
(18) Doherty, E. E.; Karki, A.; Wilcox, X. E.; Mendoza, H. G.; Manjunath, A.; Matos, V. J.; Fisher, A. J.; Beal, P. A. ADAR Activation by Inducing a *Syn* Conformation at Guanosine Adjacent to an Editing Site. *Nucleic Acids Res.* **2022**, *50*, 10857−10868.
(19) Wierzchowski, J.; Wielgus-Kutrowska, B.; Shugar, D. Fluorescence Emission Properties of 8-Azapurines and their Nucleosides, and Application to the Kinetics of the Reverse Synthetic Reaction of Purine Nucleoside Phosphorylase. *Biochim. Biophys. Acta* **1996**, *1290*, 9−17.
(20) Da Costa, C. P.; Fedor, M. J.; Scott, L. G. 8-Azaguanine Reporter of Purine Ionization States in Structured RNAs. *J. Am. Chem. Soc.* **2007**, *129*, 3426−3432.
(21) Čihák, A.; Veselý, J.; Škoda, J. Azapyrimidine Nucleosides: Metabolism and Inhibitory Mechanisms. *Adv. Enzyme Regul.* **1985**, *24*, 335−354.
(22) Škoda, J. Mechanism of Action and Application of Azapyrimidines. *Prog. Nucleic Acid Res. Mol. Biol.* **1963**, *2*, 197−219.
(23) Christman, J. K. 5-Azacytidine and 5-Aza-2′-Deoxycytidine as Inhibitors of DNA Methylation: Mechanistic Studies and their Implications for Cancer Therapy. *Oncogene* **2002**, *21*, 5483−5495.
(24) Izbicka, E.; MacDonald, J. R.; Davidson, K.; Lawrence, R. A.; Gomez, L.; Von Hoff, D. D. 5,6 Dihydro-5′-Azacytidine (DHAC) Restores Androgen Responsiveness in Androgen-Insensitive Prostate Cancer Cells. *Anticancer Res.* **1999**, *19*, 1285−1291.
(25) Crance, J. M.; Scaramozzino, N.; Jouan, A.; Garin, D. Interferon, Ribavirin, 6-Azauridine and Glycyrrhizin: Antiviral Compounds Active Against Pathogenic Flaviviruses. *Antiviral Res.* **2003**, *58*, 73−79.
(26) Rada, B.; Dragún, M. Antiviral Action and Selectivity of 6-Azauridine. *Ann. N. Y. Acad. Sci.* **1977**, *284*, 410−417.
(27) Negrya, S. D.; Efremenkova, O. V.; Solyev, P. N.; Chekhov, V. O.; Ivanov, M. A.; Sumarukova, I. G.; Karpenko, I. L.; Kochetkov, S. N.; Alexandrova, L. A. Novel 5-Substituted Derivatives of 2′-Deoxy-6-Azauridine with Antibacterial Activity. *J. Antibiot.* **2019**, *72*, 535−544.
(28) Thapa, B.; Schlegel, H. B. Calculations of pK_a's and Redox Potentials of Nucleobases with Explicit Waters and Polarizable Continuum Solvation. *J. Phys. Chem. A* **2015**, *119*, 5134−5514.
(29) Thapa, B.; Schlegel, H. B. Improved pK_a Prediction of Substituted Alcohols, Phenols, and Hydroperoxides in Aqueous Medium Using Density Functional Theory and a Cluster-Continuum Solvation Model. *J. Phys. Chem. A* **2017**, *121*, 4698−4706.
(30) Thapa, B.; Raghavachari, K. Accurate pK_a Evaluations for Complex Bio-Organic Molecules in Aqueous Media. *J. Chem. Theory Comput.* **2019**, *15*, 6025−6035.
(31) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.et al. *Gaussian 09*, revision D.01; Gaussian Inc.: Wallingford, CT, 2014.
(32) Bryantsev, V. S.; Diallo, M. S.; Goddard, W. A. Calculation of Solvation Free Energies of Charged Solutes Using Mixed Cluster/Continuum Models. *J. Phys. Chem. B* **2008**, *112*, 9709−9719.
(33) Zhang, S. A Reliable and Efficient First Principles-Based Method for Predicting pK_a Values. III. Adding Explicit Water Molecules: Can the Theoretical Slope Be Reproduced and pK_a Values Predicted More Accurately? *J. Comput. Chem.* **2012**, *33*, 517−526.
(34) da Silva, E. F.; Svendsen, H. F.; Merz, K. M. Explicitly Representing the Solvation Shell in Continuum Solvent Calculations. *J. Phys. Chem. A* **2009**, *113*, 6404−6409.
(35) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378−6396.
(36) Shields, G. C.; Seybold, P. G. *Computational Approaches for the Prediction of pK_a Values*; CRC Press, 2013.
(37) Ho, J.; Coote, M. L. A Universal Approach for Continuum Solvent pK_a Calculations: Are We There Yet? *Theor. Chem. Acc.* **2010**, *125*, 3−21.
(38) Ho, J. Predicting pK_a in Implicit Solvents: Current Status and Future Directions. *Aust. J. Chem.* **2014**, *67*, 1441−1460.
(39) Camaioni, D. M.; Schwerdtfeger, C. A. Comment on "Accurate Experimental Values for the Free Energies of Hydration of H^+, OH^-, and H_3O^+". *J. Phys. Chem. A* **2005**, *109*, 10795−10797.
(40) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Aqueous Solvation Free Energies of Ions and Ion-Water Clusters Based on an Accurate Value for the Absolute Aqueous Solvation Free Energy of the Proton. *J. Phys. Chem. B* **2006**, *110*, 16066−16081.
(41) Cho, B. P.; McGregor, M. A. Tautomerism, Protonation, and Ionization of Formycin in Aqueous Solution by the pH Dependence of ^{13}C Chemical Shifts and ^{13}C-^{1}H Coupling Constants. *Nucleosides Nucleotides* **1994**, *13*, 481−490.
(42) Harkins, T. R.; Freiser, H. Adenine-Metal Complexes. *J. Am. Chem. Soc.* **1958**, *80*, 1132−1135.
(43) Luyten, I.; Thibaudeau, C.; Chattopadhyaya, J. The Determination of the Ionization Constants of C-Nucleosides. *Tetrahedron* **1997**, *53*, 6903−6906.
(44) Levene, P. A.; Bass, L. W. *Nucleic Acids*; Chemical Catalog Company: New York, 1931.
(45) Cottrell, J. W.; Scott, L. G.; Fedor, M. J. The pH Dependence of Hairpin Ribozyme Catalysis Reflects Ionization of an Active Site Adenine. *J. Biol. Chem.* **2011**, *286*, 17658−17664.

https://doi.org/10.1021/acs.jpca.3c01358
J. Phys. Chem. A 2023, 127, 3526−3534

(46) Wierzchowski, J.; Antosiewicz, J. M.; Shugar, D. 8-Azapurines as Isosteric Purine Fluorescent Probes for Nucleic Acid and Enzymatic Research. *Mol. BioSyst.* **2014**, *10*, 2756−2774.
(47) Seela, F.; Münster, I.; Lüchner, U.; Rosemeyer, H. 8-Azaadenosine and Its 2′-Deoxyribonucleoside: Synthesis and Oligonucleotide Base-Pair Stability. *Helv. Chim. Acta* **1998**, *81*, 1139−1155.
(48) He, J.; Seela, F. 8-Aza-7-Deazapurine−Pyrimidine Base Pairs: The Contribution of 2- and 7-Substituents to the Stability of Duplex DNA. *Tetrahedron* **2002**, *58*, 4535−4542.
(49) Kierdaszuk, B.; Modrak-Wójcik, A.; Wierzchowski, J.; Shugar, D. Formycin A and its N-Methyl Analogues, Specific Inhibitors of *E. coli* Purine Nucleoside Phosphorylase (PNP): Induced Tautomeric Shifts on Binding to Enzyme, and Enzyme→Ligand Fluorescence Resonance Energy Transfer. *Biochim. Biophys. Acta* **2000**, *1476*, 109−128.
(50) Seio, K.; Sasami, T.; Tawarada, R.; Sekine, M. Synthesis of 2′-O-Methyl-RNAs Incorporating a 3-Deazaguanine, and UV Melting and Computational Studies on its Hybridization Properties. *Nucleic Acids Res.* **2006**, *34*, 4324−4334.
(51) Liu, L.; Cottrell, J. W.; Scott, L. G.; Fedor, M. J. Direct Measurement of the Ionization State of an Essential Guanine in the Hairpin Ribozyme. *Nat. Chem. Biol.* **2009**, *5*, 351−357.
(52) Seela, F.; Lampe, S. Synthesis, Base Pairing, and Structural Transitions of Oligodeoxyribonucleotides Containing 8-Aza-2′-Deoxyguanosine. *Helv. Chim. Acta* **1994**, *77*, 1003−1017.
(53) Gut, J. Aza Analogs of Pyrimidine and Purine Bases of Nucleic Acids. *Adv. Heterocycl. Chem.* **1963**, *1*, 189−251.
(54) Hall, R. *The Modified Nucleosides in Nucleic Acids*; Columbia University Press: New York, 1971.
(55) Notari, R. E.; Deyoung, J. L. Kinetics and Mechanisms of Degradation of the Antileukemic Agent 5-Azacytidine in Aqueous Solutions. *J. Pharm. Sci.* **1975**, *64*, 1148−1157.
(56) Shugar, D.; Fox, J. J. Spectrophotometric Studies of Nucleic Acid Derivatives and Related Compounds as a Function of pH: I. Pyrimidines. *Biochim. Biophys. Acta* **1952**, *9*, 199−218.
(57) Sanghvi, Y. S.; Hoke, G. D.; Freier, S. M.; Zounes, M. C.; Gonzalez, C.; Cummins, L.; Sasmor, H.; Cook, P. D. Antisense Oligodeoxynucleotides: Synthesis, Biophysical and Biological Evaluation of Oligodeoxynucleotides Containing Modified Pyrimidines. *Nucleic Acids Res.* **1993**, *21*, 3197−3203.
(58) Davis, D. R. Stabilization of RNA Stacking by Pseudouridine. *Nucleic Acids Res.* **1995**, *23*, 5020−5026.
(59) Auffinger, P.; Westhof, E. RNA Hydration: Three Nanoseconds of Multiple Molecular Dynamics Simulations of the Solvated $tRNA^{Asp}$ Anticodon Hairpin. *J. Mol. Biol.* **1997**, *269*, 326−341.
(60) Newby, M. I.; Greenbaum, N. L. Investigation of Overhauser Effects Between Pseudouridine and Water Protons in RNA Helices. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12697−12702.
(61) Jones, P. A.; Taylor, S. M. Cellular Differentiation, Cytidine Analogs and DNA Methylation. *Cell* **1980**, *20*, 85−93.
(62) Sheikhnejad, G.; Brank, A.; Christman, J. K.; Goddard, A.; Alvarez, E.; Ford, H., Jr.; Marquez, V. E.; Marasco, C. J.; Sufrin, J. R.; O'Gara, M.; Cheng, X. Mechanism of Inhibition of DNA (Cytosine C5)-Methyltransferases by Oligodeoxyribonucleotides Containing 5,6-Dihydro-5-Azacytosine. *J. Mol. Biol.* **1999**, *285*, 2021−2034.
(63) Šponer, J. E.; Vázquez-Mayagoitia, Á.; Sumpter, B. G.; Leszczynski, J.; Šponer, J.; Otyepka, M.; Banáš, P.; Fuentes-Cabrera, M. Theoretical Studies on the Intermolecular Interactions of Potentiallly Primordial Base-Pair Analogues. *Chem. − Eur. J.* **2010**, *16*, 3057−3065.
(64) Newby, M. I.; Greenbaum, N. L. A Conserved Pseudouridine Modification in Eukaryotic U2 snRNA Induces a Change in Branch-Site Architecture. *RNA* **2001**, *7*, 833−845.
(65) Sumita, M.; Desaulniers, J.-P.; Chang, Y.-C.; Chui, H. M.-P.; Clos, L.; Chow, C. S. Effects of Nucleotide Substitution and Modification on the Stability and Structure of Helix 69 From 28S rRNA. *RNA* **2005**, *11*, 1420−1429.

https://doi.org/10.1021/acs.jpca.3c01358
J. Phys. Chem. A 2023, 127, 3526−3534