Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal of Mass Spectrometry 240 (2005) 261–272
International Journal of Mass Spectrometry
www.elsevier.com/locate/ijms

Insights into nucleic acid reactivity through gas-phase experimental and computational studies

Jeehiun K. Lee*

Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, 610 Taylor Road, Piscataway, NJ 08854, USA

Received 30 July 2004; accepted 27 September 2004
Available online 10 November 2004

Abstract

Accurate measurements of the acidities and basicities of nucleic bases and nucleic base derivatives is essential for understanding issues of fundamental importance in biological systems. Hydrogen bonding modulates recognition of DNA and RNA bases, and the interaction energy between two bonded complementary nucleobases is dependent on the intrinsic basicity and acidity of the acceptor and donor groups. In addition, understanding the intrinsic reactivity of nucleic bases can shed light on key biosynthetic mechanisms for which nucleobases are substrates. In this review, we highlight advances in our lab toward understanding the fundamental reactivity of DNA and RNA. In particular, we focus on our investigation of the gas phase acidities and basicities of natural and unnatural nucleobases, and the implications of our results for the mechanisms of nucleotide biosynthetic and repair enzymes.

Keywords: Nucleic acids; RNA; DNA; Enzyme nucleobase; Acidity; Proton affinity

1. Introduction

In this review, we describe ongoing research in our lab on the measurement of multiple acidities and proton affinities of DNA and RNA nucleobases (Fig. 1). Of the concepts that organic chemists use to make sense of reactivity, ideas relating to acidity and basicity have been among the most useful. The intrinsic gas-phase acidities and proton affinities of DNA and RNA nucleobases are of interest for purely chemical reasons, but are also of importance for biological reasons, since biological environs are often relatively non-polar in nature [1]. Furthermore, hydrogen bonding modulates recognition of DNA and RNA bases, and the interaction energy between two complementary nucleobases that are held together by NH–O and NH–N hydrogen bonds is dependent on the intrinsic basicity of the acceptor atoms as well as on the acidity of the NH donor groups [2,3]. Gas-phase acidities and proton affinities of the bases are largely unknown and comparison of those acidities and proton affinities to solution values will yield valuable information on intrinsic nucleobase reactivity and the role of solvent in affecting base reactivity [4–9]. In essence, gas-phase experiments can provide the link between calculations and condensed phase data.

This review is not meant to be comprehensive, but rather to highlight some advances we have made in our lab toward understanding the fundamental reactivity of DNA and RNA. The review is divided into three sections: (I) the examination of the pyrimidine base uracil and derivatives; (II) the examination of the purine base adenine and alkyl derivatives; and (III) future plans.

2. Uracil

2.1. Acidity and uracil-DNA glycosylase [10]

Uracil-DNA glycosylase (UDGase) is a genome-protecting enzyme that removes misincorporated uracil (**1**) from DNA [11,12]. Uracil can substantially disrupt specific

* Tel.: +1 732 4456562; fax: +1 732 4455312.
E-mail address: jklee@rutchem.rutgers.edu.

1387-3806/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijms.2004.09.020

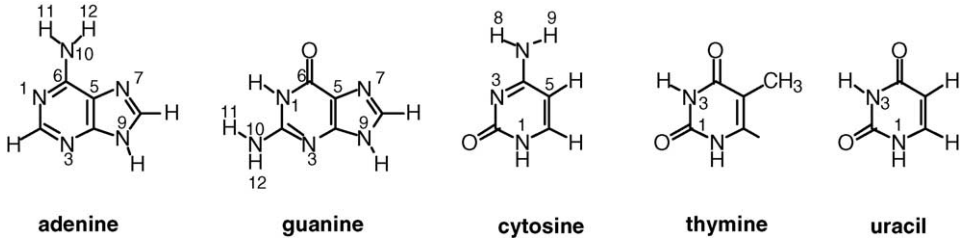
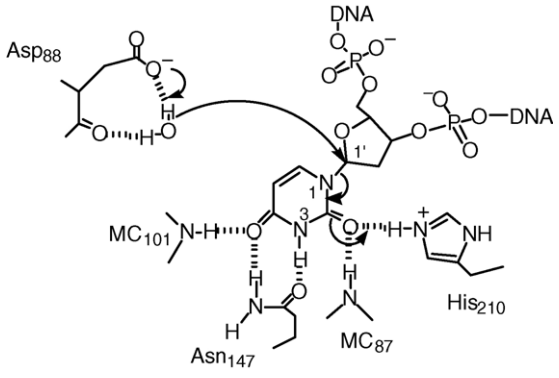
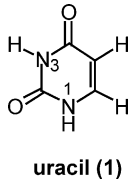


Fig. 1. Five "normal" DNA/RNA nucleobases.

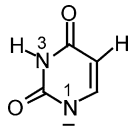


Scheme 1.

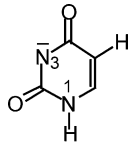
protein binding in regulatory DNA sequences, and thus poses a real threat to the genome unless it is actively removed. The proposed mechanism for uracil excision by UDGase involves nucleophilic attack by some form of activated water at C1′ (Scheme 1). This prompts an immediate question: how good a leaving group is uracil N1$^-$? More fundamentally, why does nature choose N1 for glycosylation? The condensed phase pK_a's of the N1 and N3 sites are not differentiable; uracil deprotonates with a pK_a value of 9.5 to form the N1$^-$ (**2**) species, which is in equilibrium with the N3$^-$ (**3**) species in a 1:1 ratio (throughout this review, we refer to a deprotonated or protonated ion by the site at which it has gained or lost a proton, with the resultant charge. Therefore, uracil deprotonated at N1 is the N1$^-$ ion and uracil deprotonated at N3 is the N3$^-$ ion). 3-Methyl uracil (**4**) has a higher pK_a (10.0) than 1-methyl uracil (**5**) (9.8), implying that the N3 site in uracil might be slightly more acidic. Why then is the N1 site the glycosylated position [13,14]? The N3 and N1 sites are both readily alkylated in chemical reactions, but what happens in an enzyme active site [15]? We thus became interested in pursuing the intrinsic acidity of uracil in the gas phase.



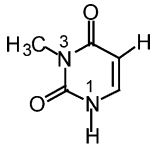
uracil (1)



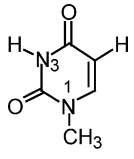
2



3



4

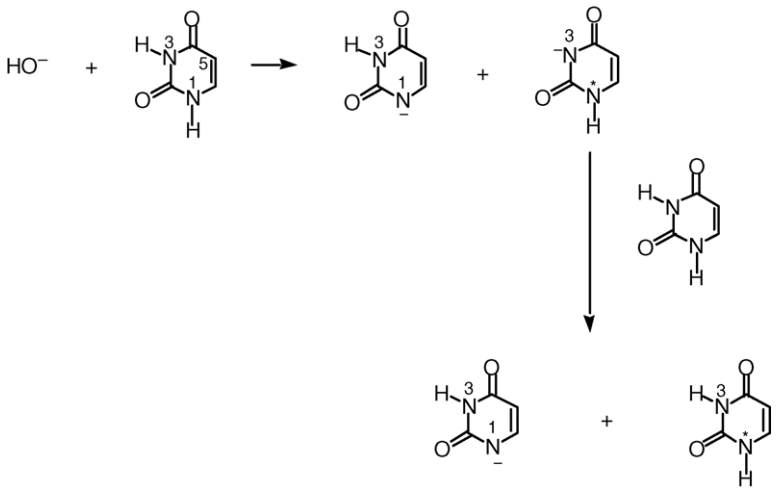


5

Table 1
Calculated ΔH_{acid} values (kcal mol^{-1}) of the N1 and N3 sites of uracil, the N3 site of 1-methyluracil and the N1 site of 3-methyluracil

Method	B3LYP/6-31+G*
N1, uracil	330.5
N3, uracil	344.1
N1, 3-methyl uracil	332.8
N3, 1-methyl uracil	345.3

Despite the proximity of pK_a values for the uracil N1 and N3 sites in solution, our and others' calculations predict that the gas phase N1 and N3 acidities should be separated by more than 10 kcal mol^{-1} (throughout this review, calculated and experimental acidities from our lab are reported as ΔH_{acid} (298 K) values) [2]. Our calculations at B3LYP/6-31+G* indicate that N1–H is very acidic, ΔH_{acid} = 330.5 kcal mol^{-1} (Table 1). The predicted acidity of N3–H, is by contrast, 344.1 kcal mol^{-1}. The tantalizing aspect of this computational result is the notion that while the solution phase N1 and N3 acidities of uracil are the same, the gas-phase acidities might actually be very different. The corollary would be

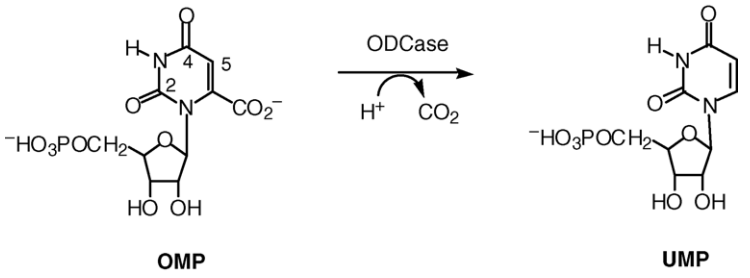


Scheme 2.

that in a non-polar environment, the N1 and N3 sites might be differentiable. Furthermore, the calculated gas-phase acidity of N1–H is predicted to be near that of HCl; the high acidity of N1–H would correlate to a stable $N1^-$ ion and indicate that in a non-polar environment such as the UDGase active site, deprotonated uracil might be a good leaving group.

We measured the most acidic site of uracil, N1–H, to be 333 ± 4 kcal mol^{-1}. This measurement was conducted using the acidity bracketing method, and is consistent with other measurements [5,8–10,16]. The novelty of our work is in measuring multiple acidic sites; however, the less acidic site of uracil is not as easily accessible as the most acidic site. Our standard experimental method involves using a strong base, HO^- ($\Delta H_{acid} = 390.7$ kcal mol^{-1}), to deprotonate both N1–H and N3–H sites of uracil. Under normal bracketing conditions, therefore, deprotonation by hydroxide produces a mixture of the $N3^-$ and the $N1^-$ ions. However, a serious problem is that the $N3^-$ ion reacts with any neutral uracil present to deprotonate N1–H and form the $N1^-$ ion; we call this uracil-catalyzed isomerization (Scheme 2) [10,17]. Therefore, to bracket the N3 site, we must perform the deprotonation of uracil under conditions that will allow the $N3^-$ to be sustained, that is, we remove the $N1^-/N3^-$ mixture from the neutral uracil environment as quickly as possible. The method that we developed takes advantage of our Finnigan dual-cell Fourier transform mass spectrometer (FTMS). First, we allow hydroxide, a strong base, to deprotonate uracil, presumably at N1 and N3, then we immediately transfer the ions to our second cell, which is free of uracil neutral. We then allowed the M-1 of uracil (*m*/*z*111) to react with reference acids. In summary, by forming a mixture of ions deprotonated at the less acidic and more acidic sites, and isolating the ions from the neutral precursor by transfer into the second cell, we avoid the neutral-catalyzed isomerization that results in loss of the ions deprotonated at the less acidic site. This new technique allowed us to measure the acidity of the N3 site of uracil to be 347 ± 4 kcal mol^{-1}.

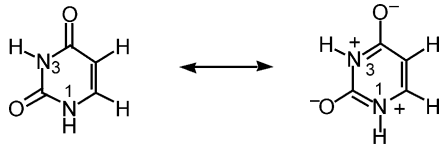
Our gas-phase theoretical and experimental studies therefore show that contrary to in solution, where the N1 and N3 protons have the same pK_a, there is an enormous difference in the N1–H and N3–H gas-phase acidities. In practical terms, N1–H of uracil in the gas phase is as acidic as HCl, while N3–H is closer to acetic acid. What this means is that in a non-polar active site, the two sites are discernible and differ in reactivity. Moreover, because N1–H is so acidic, the corresponding anion (**2**) is quite stable, and would be a good leaving group in a non-polar environment. Our results were subsequently supported by NMR studies by Stivers and co-workers [18]. They established an "unusually low pK_a" of uracil when in the active site of UDGase, and using a novel heteronuclear NMR approach, found that the bound uracil is anionic at pH 7.5. That is, the enzyme-NMR studies established the stability



Scheme 3.

of anionic uracil in the active site, contrary to the expectation based on solution phase pK_a's [1,18–22].

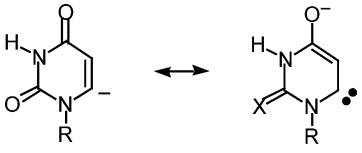
We also used theory to correlate the condensed phase to the gas phase. We conducted dielectric medium calculations on the N1 and N3 acidities to ascertain how acidities change with medium dielectric. We find that while the N1$^-$ ion is stabilized by 34.4 kcal mol^{-1} by a change in dielectric from the gas phase ($\varepsilon = 1$) to water ($\varepsilon = 78$), the N3$^-$ ion gets stabilized by a greater amount, 42.2 kcal mol^{-1}. Therefore, although the N1 and N3 sites are intrinsically quite different in acidity, the preferential solvation of the N3$^-$ site results in the two acidities coalescing in solution. If one considers the benzenoid resonance structure of uracil (**1**′), one can rationalize that N1–H, which is proximal to only one negatively charged oxygen, will be more easily removed than N3–H, which is proximal to two negatively charged oxygens. This effect would be mitigated in solution, which explains the coalescence of the N3 and N1 acidities.



uracil (1) **1'**

2.2. C5 and C6 acidities and orotidine 5′-monophosphate decarboxylase [16]

Orotidine 5′-monophosphate decarboxylase (ODCase) catalyzes the decarboxylation of orotate ribose monophosphate (OMP) to form uracil ribose monophosphate (UMP, Scheme 3) [23–25]. Decarboxylation is the last step in the de novo biosynthesis of pyrimidine nucleotides. Lack of enzyme functionality results in a genetic disease, orotic aciduria. The enzyme is an antitumor target because of its place in controlling DNA biosynthesis. Several mechanisms have been hypothesized, but this reaction remains a hotly debated mechanistic mystery [23–25]. The nature of the C6 anionic intermediate resulting from decarboxylation of orotate is the focus of ODCase mechanistic studies, how stable is it and how does the enzyme catalyze the reaction? We have established, through earlier computational studies, that the C6 anion may garner special stability because of its resonance structure **6**, a carbene-ylide [26–28].



6

In our studies of uracil, we measured the multiple acidities of a series of alkylated derivatives of uracil: 1-methyl, 3-methyl, 6-methyl, 5,6-dimethyl and 1,3-dimethyl (Fig. 2, Table 2). The intriguing aspect of our results, as relates to ODCase, is the particularly high acidity of C6–H, the C6

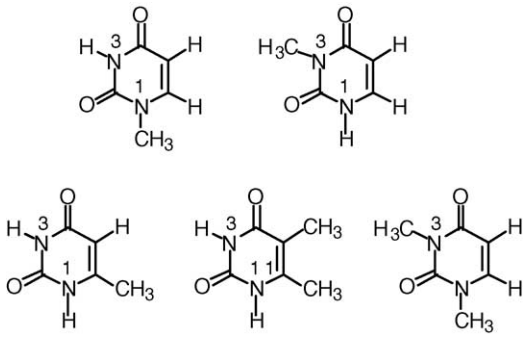


Fig. 2. Alkylated derivatives of uracil.

in 1-methyluracil and 3-methyluracil brackets to 363 and 369 kcal mol^{-1} in 1,3-dimethyluracil. Our results are also consistent with earlier work by Gronert et al., who bracketed the C6 site of 1,3-dimethyluracil (via decarboxylation of orotate) to be 369.9 ± 3.1 kcal mol^{-1}, which is very close to the value of acetone [16,29]. The high acidity of C6–H should correlate to the C6$^-$ carbene-ylide being unusually stable, as predicted by calculation. The transformation of OMP to UMP is the only known biochemical decarboxylation where the resultant anion has no pi system into which to delocalize. Our results establish that the C6 anion is not as unfavorable in a non-polar environment as one might initially think, despite its lack of pi stabilization; this C6$^-$ stability in a non-polar environment may be related to catalysis by ODCase.

2.3. O2 and O4 proton affinities and orotidine 5′-monophosphate decarboxylase [27,28,30]

While the C6$^-$ ion might be more stable than one might expect in a non-polar environment, it is still believed that catalysis by ODCase involves more than just isolating OMP in a non-polar site [23–25]. We have been at the forefront of proposals involving proton transfer, either to the 2- or the 4-oxygen, as a route to catalysis [24,26–28,30]. Our computational work supports the protonation of the 4-oxygen as the most energetically favorable route for catalysis. We sought to establish the higher proton affinity (PA) of the O4 versus the O2 of uracil as a confirmation of computational predictions.

The proton affinity (defined herein as the ΔH associated with the proton transfer) of the more basic site of uracil, the O4, was first measured in 1975 by Wilson and McCloskey to be 208.6 kcal mol^{-1} [4]. Using acidity bracketing, we measured a consistent value for the PA of O4, PA = 209 ± 3 kcal mol^{-1}.

Measuring the less basic site of a molecule is an experimental challenge. We should have been able, in principle, to apply the same method we used to measure the less acidic sites to the measurement of the PA of less basic sites. The experiment was conducted as follows: when H_3O^+ (PA (water) = 165.2 kcal mol^{-1}) is used to protonate uracil, two ions should be formed, the O2-protonated (O2H$^+$) and the O4-protonated (O4H$^+$) uracil. When the ions are allowed to stay

Table 2
Summary of experimental and (calculated) gas-phase acidities (kcal mol^{-1}) of uracil, 1-methyluracil, 3-methyluracil, 6-methyluracil, 5,6-dimethyluracil and 1,3-dimethyluracil

Structure	N1	N3	C5	C6
Uracil	333 ± 4 (330.5)	347 ± 4 (344.1)	NM (377.6)	NM (363.0)
1-Methyluracil	NA	348 ± 3 (345.3)	NM (378.8)	363 ± 3 (364.4)
3-Methyluracil	333 ± 2 (332.8)	NA	NM (379.9)	363 ± 3 (365.0)
6-Methyluracil	331 ± 3 (332.0)	352 ± 5 (345.6)	NM (379.3)	NA
5,6-Dimethyluracil	333 ± 2 (333.2)	349 ± 3 (346.1)	NA	NA
1,3-Dimethyluracil	NA	NA	384 ± 3 (380.2)	369 ± 2 (367.1)

NA: not applicable (site is alkylated), NM: not measured.

in an environment where there is a constant pressure of neutral uracil, the $O2H^+$ ion reacts with another molecule of uracil and isomerizes to $O4H^+$. In order to bracket the less basic O_2 site, we must transfer the $O2H^+/O4H^+$ mixture out of the neutral environment, into a second reaction cell, where the reference base is added at a constant pressure. Under these "less basic" conditions, the $O2H^+$ will not completely isomerize to $O4H^+$, and the $O2H^+$ can be bracketed.

We found, however, that using this method, we cannot accurately measure the PA of the less basic O2 site of uracil. We were unable to pinpoint the lower limit of the proton affinity because essentially proton transfer was always detected. We suspect that the constant proton transfer is either the result of an inability to completely eject the highly acidic H_3O^+, which when present will proton transfer to any of our reference bases, and/or insufficient cooling of the generated ions. To overcome these problems, we modified our experimental design to use a less acidic precursor ion to generate protonated uracil.

In the modified experiment, H_3O^+ (PA (water) = 165.2 kcal mol^{-1}) is used to protonate a known base, 3-pentanone (Scheme 4). We isolate the 3-pentanoneH^+ ion (PA (3-pentanone) = 200.0 kcal mol^{-1}) and use it as our precursor ion to protonate uracil. The $O2H^+/O4H^+$ mixture is then transferred to the second cell, out of the neutral uracil environment, where the reference base is cyclohexanone. The PA of cyclohexanone is 201.0 kcal mol^{-1}. If there is any $O4H^+$ present, cyclohexanone will be unable to deprotonate it, because the PA of the O4 of uracil is 209 kcal mol^{-1}. Therefore, if proton transfer occurs, it must be the reaction between cyclohexanone and the $O2H^+$ ion. We do detect proton transfer, indicating that the uracil O2 site has a basicity less than 201.0 kcal mol^{-1}. Also, since the O2 site is protonated by 3-pentanoneH^+, it must be more basic than 200.0 kcal mol^{-1}. We also performed the experiment in the opposite direction, where we used protonated cyclohexanone as the precursor ion (PA (cyclohexanone) = 201.0 kcal mol^{-1}). Protonated cyclohexanone does react with uracil to produce the $[M+H]^+$ ion. However, when the protonated uracil is allowed to react with 3-pentanone (PA (3-pentanone) = 200.0 kcal mol^{-1}) proton transfer does not occur. We, therefore, bracket the less basic site of uracil to be between 3-pentanone and cyclohexanone, at 201 ± 3 kcal mol^{-1}.

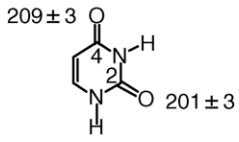
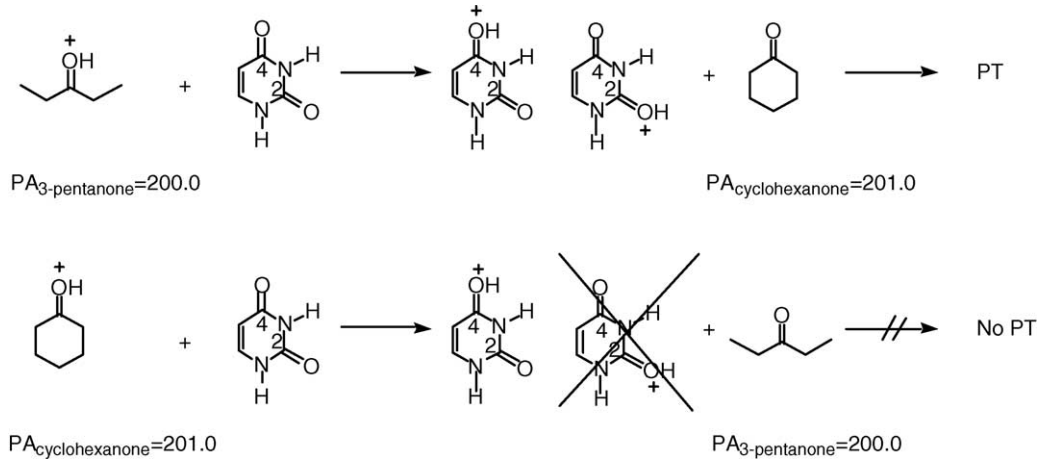


Fig. 3. Experimental proton affinities of uracil, in kcal mol^{-1}.

A summary of the gas-phase experimental proton affinity results for the O2 and O4 sites of uracil are shown in Fig. 3. We have found that the two oxygen sites of uracil are differentiable, with the O4 being more basic than the O2, by



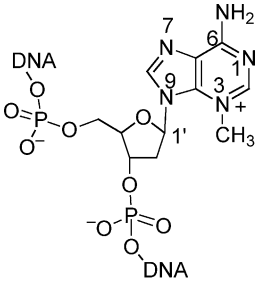
Scheme 4.

8 kcal mol^{-1}. These experimental results provide evidence for the intrinsically higher basicity of the O4 predicted by computations. This result has implications for the ODCase catalytic mechanism; since intrinsically the O4 is more basic, protonation at that site might be favored in the non-polar active site of the enzyme.

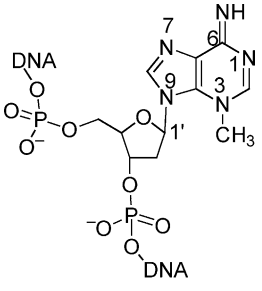
3. Adenine

3.1. 3-Methyl adenine and 3-methyl adenine DNA glycosylase [31]

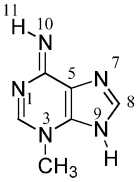
Adenine can be alkylated by cancer chemotherapeutics as well as environmental mutagens and 3-methyladenine is the most common mutation [32]. 3-Methyladenine DNA glycosylase is a non-specific enzyme that will cleave a wide range of damaged bases, including hypoxanthine, 7-methylguanine and 3-methyl- and 1, N^6-etheno-adenine from DNA [33,34]. As with uracil-DNA glycosylase, the proposed mechanism for excision of alkylated bases from DNA by human 3-methyladenine glycosylase involves nucleophilic attack at C1′ by some form of activated water (Scheme 5) [35,36]. This likewise prompts the question, how good of a leaving group is 3-methyladenine? It is not known whether the protonated substrate **7a** or the neutral substrate **7b** is the active site species [37]. The protonated substrate would undoubtedly be a good leaving group; in this study, however, we focused on the possibility that the neutral **7b** is cleaved, with 3-methyladenine N9$^-$ serving as a leaving group. A related question is, how acidic is the N9–H of 3-methyladenine **8**? The more acidic the site, then the more stable the resultant N9$^-$ ion should be, and the better a leaving group it should be.



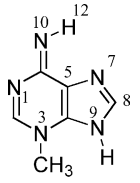
7a



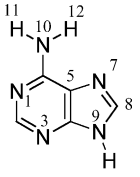
7b



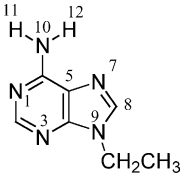
8a



8b



9



10

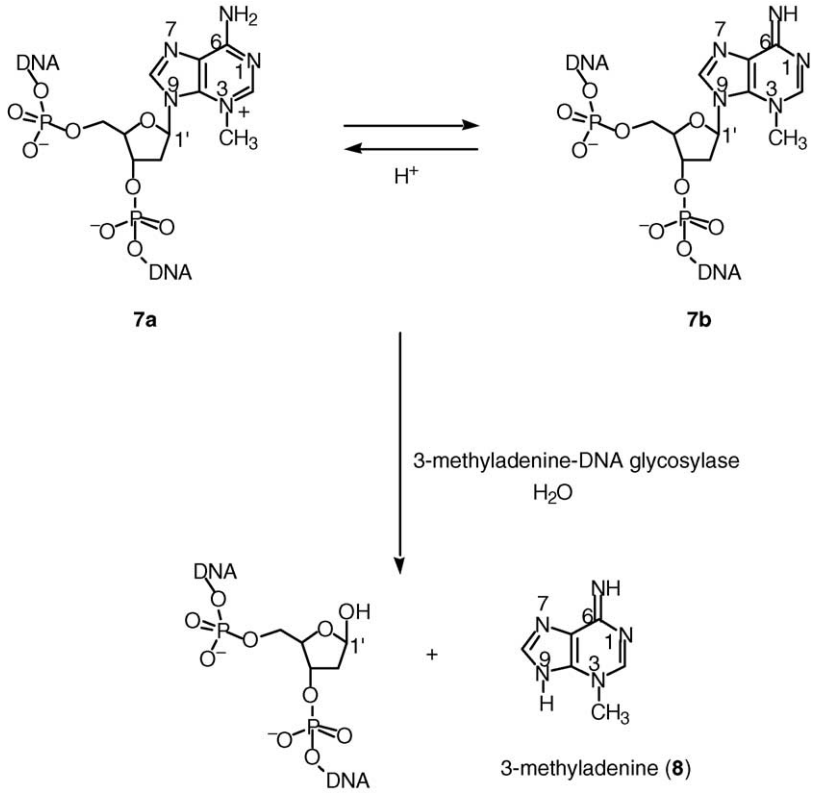
Using the same methods we developed in our uracil study, we calculated and measured the acidities of adenine (**9**), 9-ethyladenine (**10**) and 3-methyladenine (**8**). The results of our acidity calculations at B3LYP/6-31+G* for adenine, 9-ethyladenine and 3-methyladenine are shown in Table 3.

The measurement of the acidities of adenine is consistent with calculations: the N9–H is found to have an experimental ΔH_{acid} of 333 ± 2 kcal mol^{-1}. This value is in agreement with an earlier measurement of ΔG_{acid} = 325 kcal mol^{-1} [9]. The less acidic N10–H site, measured for the first time by us, is 352 ± 4 kcal mol^{-1}. Both these values are in agreement with the corresponding calculated acidity values of 334.8 and 354.2 kcal mol^{-1} (Table 3). We bracketed the acidity of 9-ethyl adenine as a means for "blocking" the N9 site and confirming our N10–H measurement of adenine. Consistent with calculations and the parent adenine results, 9-ethyl adenine has a ΔH_{acid} of 352 ± 4 kcal mol^{-1}.

3-Methyladenine is of special interest to us due to its mutagenic properties in nature. It is most often drawn, in DNA,

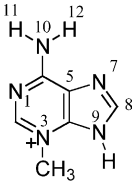
Table 3
Calculated gas-phase acidities (kcal mol^{-1}) of adenine, 9-ethyladenine and 3-methyladenine at B3LYP/6-31+G*

Structure	N9	N10H11	N10H12
Adenine, N9 tautomer (**9**)	334.8	354.2	353.5
9-Ethyladenine (**10**)	–	355.0	354.4
3-Methyladenine, imine tautomer	324.2 (**8a**) 322.1 (**8b**)	–	



Scheme 5.

as the protonated species **7a**, which can be in equilibrium with the imine form **7b** (Scheme 5). The free base analogs of **7a** and **7b** are the protonated 3-methyladenine **11** and the imine **8**. Imine **8** is predicted, computationally, to be extraordinarily acidic in the gas phase, comparable to the acidity of HBr (322–324 kcal mol^{-1}, Table 3). This highly acidic N9–H would imply that the corresponding conjugate base would be very stable, also implying that the N9$^-$ ion would serve as a good leaving group. This is of interest because if the deprotonated methylated adenine is a favorable leaving group, then it should be particularly facile for the 3-methyladenine glycosylase to excise.



11

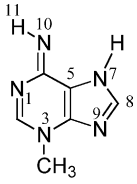
In starting the bracketing studies, we allowed the conjugate base of 3-methyladenine to react with 1,1,1-trifluoro-2,4-pentadione, an acid that we had readily available in the laboratory, and the ΔH_{acid} of which is 328.3 ± 2.9 kcal mol^{-1}. We would not expect the conjugate base of 3-methyladenine, which has a predicted acidity of 322–324 kcal mol^{-1}, to deprotonate the pentadione. To our surprise, the conjugate base of 3-methyladenine readily deprotonates the dione reference acid, while the enolate of the pentadione is unable to deprotonate 3-methyladenine! Clearly, 3-methyladenine is far less acidic than calculations predicted; we ultimately found the gas-phase acidity to be 347 ± 4 kcal mol^{-1}, bracketing between acetic acid (ΔH_{acid} (CH_3COOH) = 348.1 kcal mol^{-1}) and formic acid (ΔH_{acid} (HCOOH) = 345.3 kcal mol^{-1}).

This curious result, at odds with our calculations, led us to consider the various tautomers of 3-methyladenine. The possible tautomers are what we call the "N9 tautomer", which we initially assumed to be the most stable (**8a**, the N9H11 tautomer, and **8b**, the N9H12 tautomer), the "N7 tautomer" (**12a**, the N7H11 tautomer, and **12b**, the N7H12 tautomer) and the "N10 tautomer" (**13**). The relative gas-phase energies for each of these species are shown in Table 4. The most stable tautomer appears to be the N10 (**13**). The N9 tautomers **8a** and **8b** are about 25 kcal mol^{-1} less stable than the N10; the N7H11 tautomer **12a** is 10 kcal mol^{-1} less stable and the

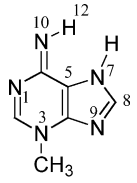
Table 4
Calculated gas-phase relative energies and acidities (kcal mol^{-1}) of the different tautomers of 3-methyladenine at B3LYP/6-31+G*

Structure	Relative E	ΔH_{acid}
N9H11 tautomer **8a**	24.5	324.2 (N9—H)
N9H12 tautomer **8b**	24.7	322.1 (N9—H)
N7H11 tautomer **12a**	10.4	338.3 (N7—H)
N7H12 tautomer **12b**	17.8	329.0 (N7—H)
N10 tautomer **13**	0	346.8 (N10—H11)
		348.7 (N10—H12)

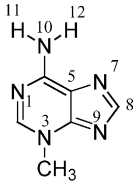
N7H12 tautomer **12b**, 18 kcal mol^{-1} less stable than the N10 tautomer **13**.



12a



12b



13

The corresponding acidities for all the tautomers are also given in Table 4. Based upon our calculated results, we would expect the N10 tautomer **13** to be the major structure in the gas phase, and indeed, the calculated acidity of that tautomer correlates with our experimental result. The calculated values are 346.8 and 348.7 kcal mol^{-1} (depending on which proton is extracted); our experimental value is 347 kcal mol^{-1}.

While this result is of fundamental interest, biologically speaking, the N10 tautomer is probably not of importance. In DNA, the adenine N9 is substituted by a ribose moiety. Therefore, post-alkylation, the N9 has no proton to lose to tautomerize to **13**. That is, the N10 tautomer of 3-methyladenine would only be accessible if there were a proton at N9; the ribose effectively blocks that site in DNA [38].

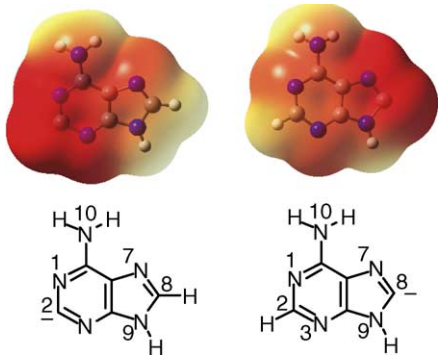


Fig. 5. Calculated electrostatic potential surfaces for adenine C2$^-$ and adenine C8$^-$. Red indicates negative potential while blue indicates positive potential.

Interestingly, if glycosylation of adenine in DNA yields the neutral 3-methyladenine tautomer **7b** (Scheme 5), the alkylated nucleobase may be a good leaving group. Our calculations show that the N9–H for **8** is highly acidic, 10 kcal mol^{-1} more so than for the parent adenine (Table 4) [33,39,40]. Herein may lie a possible reason that 3-methyladenine is particularly susceptible to excision.

3.2. C–H acidities, then ODCase again! [41]

As with uracil, we examined a series of alkylated adenine derivatives: 9-ethyladenine (**10**), 3-methyladenine (**13**), 1-methyladenine (**14**) and *N*,*N*-dimethyladenine (**15**). In our study of these species, we became intrigued with the C–H acidities. Although when thinking about hydrogen bonding, one usually focuses exclusively on X–H•••Y bonds where X and Y are heteroatoms, C–H•••Y hydrogen bonds have

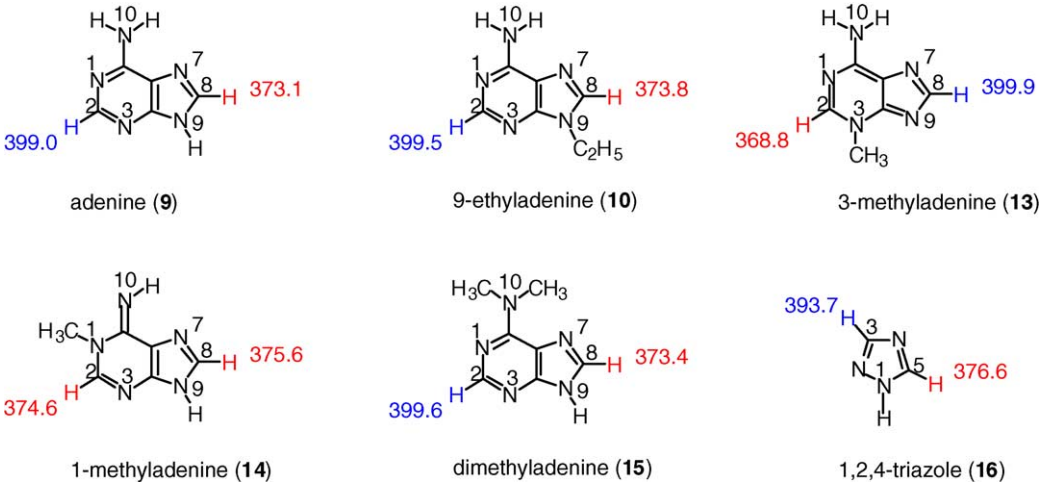
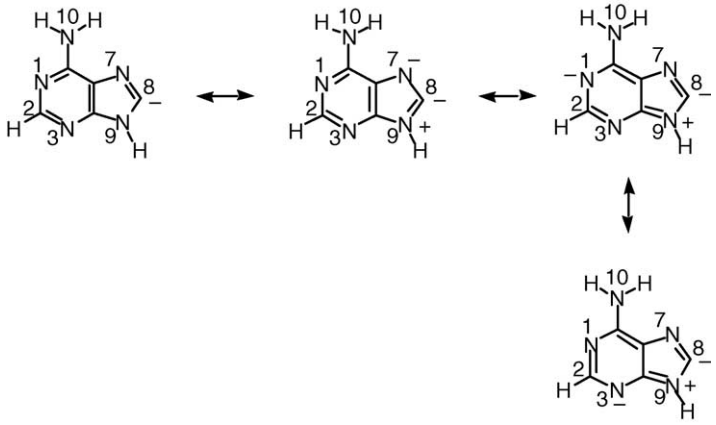
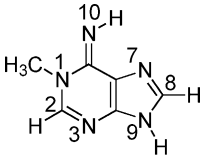


Fig. 4. Calculated C2 and C8 acidities of adenine and alkyl derivatives using B3LYP/6-31 + G*, at 298 K. Less acidic sites are indicated in blue while more acidic sites are indicated in red.

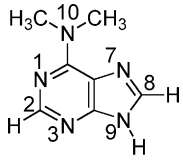


Scheme 6.

in fact found to be an important extra recognition factor in biological binding [42]. Therefore, C–H acidities are of great interest. In this section of the review, we focus on our computational results.



14

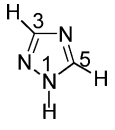


15

The acidities of the C–H protons on the C2 and C8 of each adenine derivative are of particular interest. The calculated C2–H and C8–H acidities are shown in Fig. 4. We were intrigued to find that some of these vinylic protons are quite acidic (shown in red). In adenine (**9**), the C2–H is of a typical aromatic proton acidity, 399.0 kcal mol^{-1} (the gas-phase acidity of benzene is 401.70 ± 0.50 kcal mol^{-1}) [43]. The C8–H, however, is quite acidic, 373.1 kcal mol^{-1}, a value closer to that of HF (ΔH_{acid} = 371.3 kcal mol^{-1}) or acetone (ΔH_{acid} = 369.1 kcal mol^{-1}) than of benzene [43].

In 9-ethyladenine (**10**), like adenine, the C8–H acidity is calculated to be 373.8 kcal mol^{-1} while the C2–H is much higher, at 399.5 kcal mol^{-1}. This pattern is repeated in *N*,*N*-dimethyladenine (**15**), where the C8–H is considerably more acidic than the C2–H. 3-Methyladenine (**13**), however, shows the opposite trend; the C2–H is the acidic one (368.8 kcal mol^{-1}, calculated) while the C8–H is the less acidic site, at 399.9 kcal mol^{-1}. 1-Methyladenine shows yet a different pattern, wherein both the C2–H and the C8–H are acidic (calculated values, 374.6 and 375.6 kcal mol^{-1}, respectively). These differing trends in C–H acidity among the adenine derivatives piqued our interest: why are some sites so much more acidic than others?

The pattern that emerges when one examines the five structures is that the more acidic vinylic C–H site in each molecule is always adjacent to an N–R moiety (R = H, alkyl). This appears to be true regardless of whether the proton is attached to a carbon on the five or the six membered ring of the purine. For example, in 1-methyladenine, both the C–H on the six-membered and on the five-membered rings are adjacent to N–R groups, and both are quite acidic. The values are also very consistent; all the C–H sites that are not adjacent to an N–R group essentially have a calculated acidity of 399–400 kcal mol^{-1}. The acidities of the C–H sites adjacent to an N–R group are quite consistent as well, ranging between 369 and 376 kcal mol^{-1}.

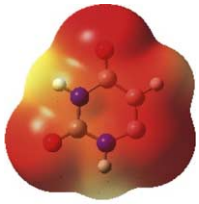


16

To test our hypothesis that vinylic C–H sites next to N–R moieties are more acidic than others, we calculated the C–H acidities on the model compound 1,2,4-triazole (**16**). Within this molecule, there are two C–H sites; by our hypothesis, the site between the two imine nitrogens (C3–H) should be less acidic than the site adjacent to the N–H (C5–H). The calculated acidity values are shown in Fig. 4 and are consistent with our hypothesis. The C3–H is less acidic than water, at 393.7 kcal mol^{-1}, while the C5–H is significantly more acidic, 376.6 kcal mol^{-1}.

Why are the C–H sites adjacent to N–R groups particularly acidic? One possibility is that for those carbanions with an adjacent N–R group, the nitrogen lone pair can delocalize into the pi system, making that N more “positive”, which could in turn electrostatically stabilize the adjacent carbanion (Scheme 6). Another possibility is that electron repulsion might have a large effect. For example, the C2 carbanion of

adenine must experience significant electrostatic repulsion from the two in-plane nitrogen lone pairs that flank it; in contrast, the C8 carbanion is adjacent to only one in-plane lone pair. To computationally explore these possibilities, we calculated the electrostatic potential for the relevant species. The electrostatic potential surfaces for adenine $C2^-$ and $C8^-$ are shown in Fig. 5. The color at each point on these surfaces reflects the interaction energy between the molecule and a positive test charge at that point. A red color indicates attractive potential while blue represents repulsive potential. These anions have an attractive potential to a positive test charge, so the overall surface is quite red. The areas of pale red indicate a less "negative" region; yellow/green indicates a more neutral or "positive" region, depending on how bluish the color is. The C2 site of the $C2^-$ ion appears to be flanked by quite a large "negative" cloud, consistent with the argument



6

Fig. 6. Calculated electrostatic potential surface for uracil $C6^-$ ion. Red indicates negative potential while blue indicates positive potential.

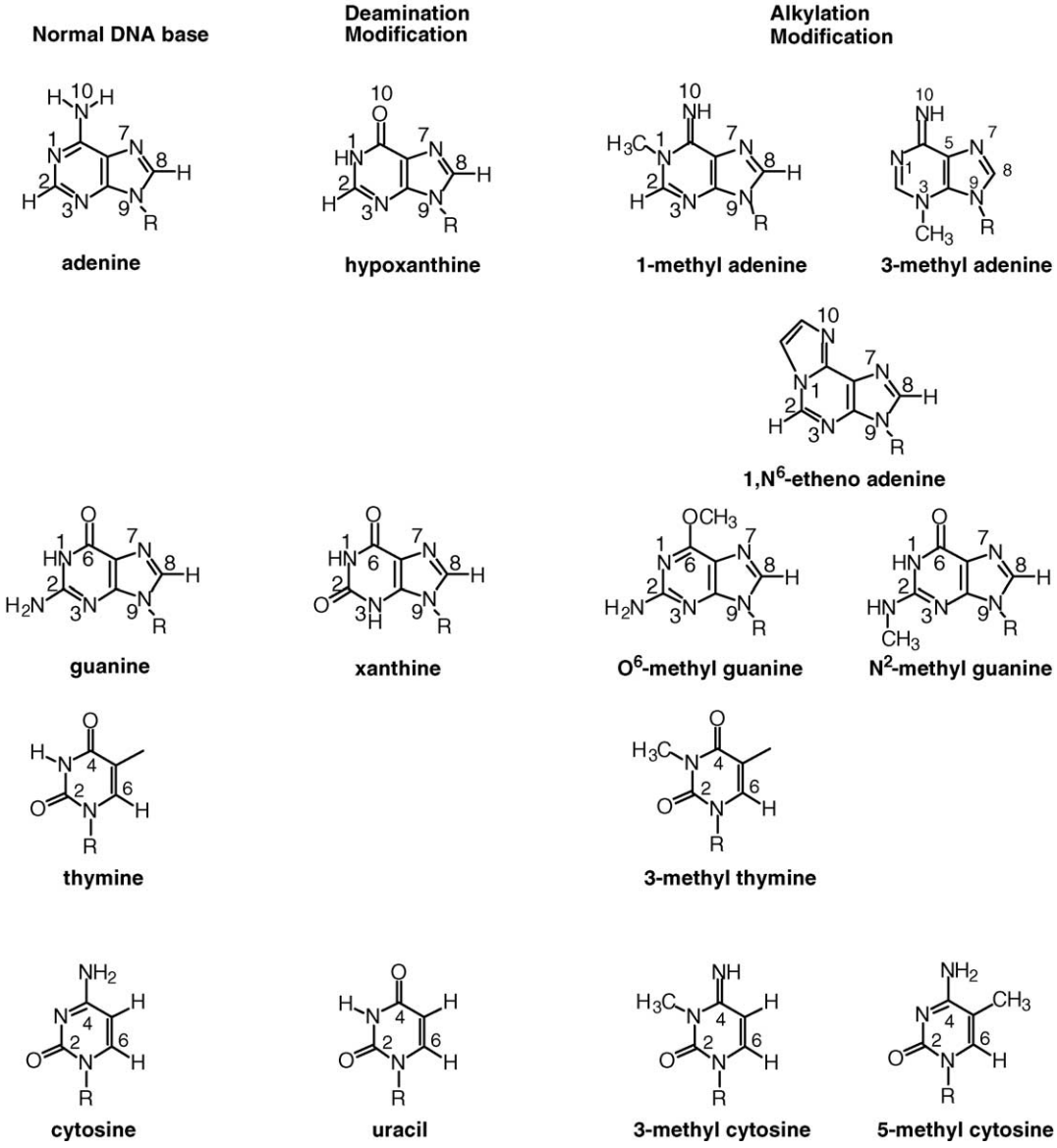


Fig. 7. Normal and damaged DNA bases.

that the two in-plane nitrogen lone pairs provide substantial electrostatic repulsion. The C8 site of the $C8^-$ ion, on the other hand, is surrounded by a very red N7 but a less red N9. Furthermore, the proton on N9 is almost green, indicating a less negative environment. The slightly positive charge on the N–H group is consistent with delocalization of the N lone pair into the pi-system. This pattern of electrostatic repulsion of carbanions flanked by two imine nitrogens, and stabilization of carbanions with an adjacent N–R group holds true for all the calculated adenine derivatives and the deprotonated triazole anions.

Study of these adenine derivatives led us back to uracil and the uracil $C6^-$ anion (vide supra, Section 2.2). This ion is also adjacent to an N–R group; could that N–R group be providing extra stabilization as we observe with the purine derivatives? Electrostatic potential calculations of the uracil $C6^-$ ion do in fact show that the C6 site appears to be stabilized by the adjacent N1–H group (Fig. 6); this would help explain why the C6–H is so unusually acidic, and has implications for the decarboxylation catalyzed by ODCase.

4. Future directions

We will continue our examination of the acidities and proton affinities of the five parent DNA and RNA nucleobases (Fig. 1). Our overall focus has now expanded to include damaged bases, particularly those modified by deamination and alkylation (Fig. 7). By measuring the multiple acidities and proton affinities of normal and damaged nucleobases, we will learn what makes damaged bases inherently different from normal bases. Are certain sites in damaged bases different in acidity and/or basicity than in their normal counterparts, making them hydrogen bond in a different way and contributing to mutagenicity? What makes some damaged bases recognizable by glycosylases? We are confident that our fundamental acidity and proton affinity studies will continue to lend insight into the overall reactivity and properties of nucleic acids and the enzymes for which they are substrates.

Acknowledgements

We thank the National Science Foundation (CHE-0092215), the Alfred P. Sloan Foundation, and the Rutgers Busch Grant Program for support, and the National Center for Supercomputing Applications, the High Performance Computing Center at the University of Kentucky and the Boston University Scientific Computing Facilities for computational resources.

References

[1] T. Simonson, C.L. Brooks III, J. Am. Chem. Soc. 118 (1996) 8452.
[2] M.T. Nguyen, A.K. Chandra, T. Zeegers-Huyskens, J. Chem. Soc., Faraday Trans. 94 (1998) 1277.
[3] A.K. Chandra, M.T. Nguyen, T. Uchimaru, T. Zeegers-Huyskens, J. Phys. Chem. A 103 (1999) 8853.
[4] M.S. Wilson, J.A. McCloskey, J. Am. Chem. Soc. 97 (1975) 3436.
[5] W. Li, I. Santos, A.G. Marshall (Eds.), 1996 ASMS Abstracts, American Society for Mass Spectrometry, Santa Fe, New Mexico, 1996.
[6] M. Meot-Ner, J. Am. Chem. Soc. 101 (1979) 2396.
[7] F. Greco, A. Liguori, G. Sindona, N. Uccella, J. Am. Chem. Soc. 112 (1990) 9092.
[8] T. Miller, S.T. Arnold, A.A. Viggiano, A.E.S. Miller, J. Phys. Chem. A 108 (2004) 3439.
[9] E.C.M. Chen, E.S. Chen, J. Phys. Chem. B 104 (2000) 7835.
[10] M.A. Kurinovich, J.K. Lee, J. Am. Chem. Soc. 122 (2000) 6258.
[11] R. Savva, K. McAuley-Hecht, T. Brown, L. Pearl, Nature 373 (1995) 487.
[12] G. Dianov, B. Sedgwick, G. Daly, M. Olsson, S. Lovett, T. Lindahl, Nucleic Acids Res. 22 (1994) 993.
[13] M.E. Jones, Ann. Rev. Biochem. 49 (1980) 253.
[14] D. Voet, J.G. Voet, Biochemistry, second ed., John Wiley & Sons, Inc., New York, 1995.
[15] E. Kimura, H. Kitamura, T. Koike, M. Shiro, J. Am. Chem. Soc. 119 (1997) 10909.
[16] M.A. Kurinovich, J.K. Lee, J. Am. Soc. Mass Spectrom. 13 (2002) 985.
[17] J.J. Grabowski, X. Cheng, J. Am. Chem. Soc. 111 (1989) 3106.
[18] A.C. Drohat, J.T. Stivers, J. Am. Chem. Soc. 122 (2000) 1840.
[19] F. Jordan, H. Li, A. Brown, Biochemistry 38 (1999) 6369.
[20] A.C. Drohat, G. Xiao, M. Tordova, J. Jagadeesh, K.W. Pankiewicz, K.A. Watanabe, G.L. Gilliland, J.T. Stivers, Biochemistry 38 (1999) 11876.
[21] A.C. Drohat, J. Jagadeesh, E. Ferguson, J.T. Stivers, Biochemistry 38 (1999) 11866.
[22] A.C. Drohat, J.T. Stivers, Biochemistry 39 (2000) 11865.
[23] B.G. Miller, R. Wolfenden, Ann. Rev. Biochem. 71 (2002) 847.
[24] K.N. Houk, J.K. Lee, D.J. Tantillo, S. Bahmanyar, B.N. Hietbrink, ChemBioChem 2 (2001) 113.
[25] J.K. Lee, D.J. Tantillo (Eds.), Orotidine Monophosphate Decarboxylase: A Mechanistic Dialogue, Vol. 238, Springer, New York, 2004.
[26] J.K. Lee, K.N. Houk, Science 276 (1997) 942.
[27] L.M. Phillips, J.K. Lee, J. Am. Chem. Soc. 123 (2001) 12067.
[28] D.A. Singleton, S.A. Merrigan, B.J. Kim, P. Beak, L.M. Phillips, J.K. Lee, J. Am. Chem. Soc. 122 (2000) 3296.
[29] S. Gronert, W.Y. Feng, F. Chew, W. Wu, Int. J. Mass Spectrom. 196 (2000) 251.
[30] M.A. Kurinovich, L.M. Phillips, S. Sharma, J.K. Lee, Chem. Comm. (2002) 2354.
[31] S. Sharma, J.K. Lee, J. Org. Chem. 67 (2002) 8360.
[32] S.A. Smith, B.P. Engelward, Nucleic Acids Res. 28 (2000) 3294.
[33] B. Singer, B. Hang, Chem. Res. Toxicol. 10 (1997) 713.
[34] B.P. Engelward, J.M. Allan, A.J. Dreslin, J.D. Kelly, M.M. Wu, B. Gold, L.D. Samson, J. Biol. Chem. 273 (1998) 5412.
[35] A.Y. Lau, O.D. Schärer, L. Samson, G.L. Verdine, T. Ellenberger, Cell 95 (1998) 249.
[36] A.Y. Lau, M.D. Wyatt, B.J. Glassner, L.D. Samson, T. Ellenberger, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 13573.
[37] In biological papers, 3-methyladenosine is most commonly shown as the protonated species **7a**. The pK_a of protonated 3-methyladenine has been measured (6.1), and the relatively low value in comparison to other purine derivatives led the authors to believe that the pK_a corresponds to protonation of the imidazole ring, implying that 13 may be the favored neutral tautomeric form of 3-methyladenine in solution. The pK_a of protonated 3-methyladenosine, where N9 is substituted with a ribose is not, to our knowledge, known;
P.D. Lawley, P. Brookes, Biochem. J. 89 (1963) 127;
N.J. Leonard, J.A. Deyrup, J. Am. Chem. Soc. 84 (1962) 2148.

[38] N.J. Leonard, J.A. Deyrup, J. Am. Chem. Soc. 84 (1961) 2148.
[39] E. Seeberg, K.G. Berdal (Eds.), Repair of Alkylation Damage to DNA, Landes Bioscience, New York, 1997.
[40] P.D. Lawley, P. Brookes, Biochem. J. 89 (1963) 127.
[41] S. Sharma, J.K. Lee, J. Org. Chem. 69 (2004) 7018.
[42] M.C. Wahl, M. Sundaralingam, Trends Biochem. Sci. 22 (1997) 97.
[43] W.G. Mallard, P.J. Linstrom, (Eds.), NIST Chemistry WebBook, NIST Standard Reference Database Number 69, March 2003, National Institute of Standards and Technology: Gaithersburg, MD 20899, 2003, http://webbook.nist.gov.