

# Divergent Mechanisms for Enzymatic Excision of 5-Formylcytosine and 5-Carboxylcytosine from DNA

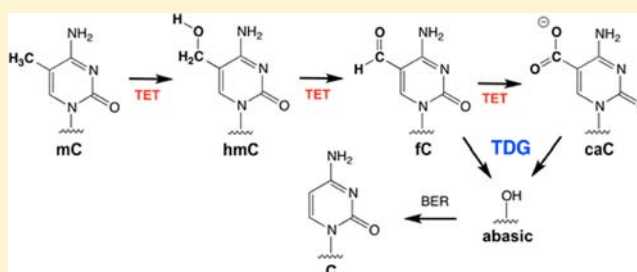
Atanu Maiti,<sup>†</sup> Anna Zhachkina Michelson,<sup>‡</sup> Cherece J. Armwood,<sup>†</sup> Jeehiun K. Lee,<sup>‡,\*</sup> and Alexander C. Drohat<sup>†,\*</sup>

<sup>†</sup>Department of Biochemistry and Molecular Biology, University of Maryland School of Medicine, Baltimore, Maryland 21201, United States

<sup>‡</sup>Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, New Brunswick, New Jersey 08854, United States

## S Supporting Information

**ABSTRACT:** 5-Methylcytosine (mC) is an epigenetic mark that impacts transcription, development, and genome stability, and aberrant DNA methylation contributes to aging and cancer. Active DNA demethylation involves stepwise oxidation of mC to 5-hydroxymethylcytosine, 5-formylcytosine (fC), and potentially 5-carboxylcytosine (caC), excision of fC or caC by thymine DNA glycosylase (TDG), and restoration of cytosine via follow-on base excision repair. Here, we investigate the mechanism for TDG excision of fC and caC. We find that 5-carboxyl-2'-deoxycytidine ionizes with  $pK_a$  values of 4.28 (N3) and 2.45 (carboxyl), confirming that caC exists as a monoanion at physiological pH. Calculations do not support the proposal that G-fC and G-caC base pairs adopt a wobble structure that is recognized by TDG. Previous studies show that *N*-glycosidic bond hydrolysis follows a stepwise ( $S_N1$ ) mechanism, and that TDG activity increases with pyrimidine N1 acidity, that is, leaving group quality of the target base. Calculations here show that fC and the neutral tautomers of caC are acidic relative to other TDG substrates, but the caC monoanion exhibits poor acidity and likely resists TDG excision. While fC activity is independent of pH, caC excision is acid-catalyzed, and the pH profile indicates that caC ionizes in the enzyme–substrate complex with an apparent  $pK_a$  of 5.8, likely at N3. Mutational analysis reveals that Asn191 is essential for excision of caC but dispensable for fC activity, indicating that N191 may stabilize N3-protonated forms of caC to facilitate acid catalysis and suggesting that N191A-TDG could potentially be useful for studying DNA demethylation in cells.



## INTRODUCTION

The conversion of cytosine (C) to 5-methylcytosine (mC) constitutes the major type of DNA methylation in vertebrates, and this epigenetic signal has a profound influence on biological processes including transcription, development, and genome stability.<sup>1</sup> Moreover, aberrant DNA methylation is implicated in aging and in human diseases including cancer.<sup>2</sup> The DNA methyltransferases that catalyze DNA methylation are well-characterized, but the enzymes responsible for actively reversing the methylation mark (i.e., converting mC back to C) had remained elusive. However, recent biochemical and biological studies have established a pathway for active DNA demethylation involving oxidation of mC and processing of mC derivatives via thymine DNA glycosylase (TDG) and base excision repair (BER), as illustrated in Figure 1.

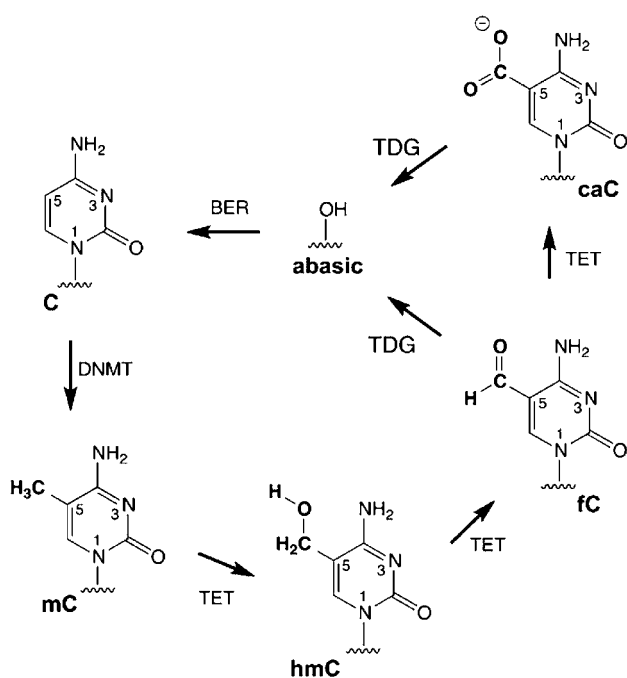
It was recently discovered that the TET (ten-eleven translocation) family of dioxygenases can oxidize mC to 5-hydroxymethylcytosine (hmC),<sup>3,4</sup> a base that was previously detected in mammalian DNA.<sup>5</sup> Subsequent studies revealed that TET enzymes can further oxidize hmC to 5-formylcytosine (fC) and fC to 5-carboxylcytosine (caC) in a stepwise manner.<sup>6–8</sup> We showed that TDG can rapidly excise fC from

DNA *in vitro*,<sup>9</sup> and this activity was subsequently found in mammalian cells.<sup>10–13</sup> TDG also excises caC *in vitro*,<sup>6</sup> albeit substantially slower than fC,<sup>9</sup> and this activity is found in mammalian cells.<sup>6,11,13</sup> The abasic nucleotide generated by TDG excision of fC or caC is replaced by cytosine via downstream BER, completing the demethylation process.

No mammalian glycosylase can excise hmC, and TDG is the only one that can remove fC or caC.<sup>1,9,14–16</sup> TDG, and its glycosylase activity in particular, is essential for embryonic development, indicating a critical role for TDG-mediated DNA demethylation in regulating developmental genes.<sup>17,18</sup> TDG also functions to maintain genome integrity; it was discovered as an enzyme that selectively excises T from G·T mismatches, an activity needed to protect against mutations caused by mC deamination.<sup>19</sup> Here, we investigate the mechanism by which TDG excises fC and caC from DNA and the chemical properties of these bases that dictate the catalytic requirements for their excision.

Received: June 25, 2013

Published: September 24, 2013



**Figure 1.** Pathway for active DNA demethylation involving TET enzymes and TDG-initiated BER. Details and abbreviations are provided in the main text.

Previous kinetic isotope effect (KIE) studies show that 2'-deoxynucleotide hydrolysis reactions, including non-enzymatic and those catalyzed by UNG, MutY, and ricin, follow a stepwise ( $S_N1$ ) mechanism that involves rupture of the *N*-glycosidic bond to yield a short-lived oxocarbenium ion intermediate and subsequent addition of the nucleophile.<sup>20–24</sup> As such, the rate depends on the stability (leaving group ability) of the departing nucleobase, and catalysis can be achieved by activation (protonation) or electrostatic stabilization of the leaving group.<sup>22,23,25,26</sup> This chemical precedent indicates that the TDG reaction likely follows a stepwise mechanism<sup>23</sup> rather than a concerted ( $S_N2$ ) mechanism as recently suggested.<sup>27</sup> Consistent with this conclusion, we previously showed that TDG activity ( $k_{max}$ ) depends on N1 acidity of the target pyrimidine; that is, activity increases with leaving group ability of the excised base.<sup>16</sup> For example,  $k_{max}$  is much greater for 5-chlorouracil ( $pK_a^{N1} = 8.1$ ) relative to thymine ( $pK_a^{N1} = 10.2$ ), even though these bases have very similar steric and electrostatic properties.<sup>16</sup> This work also revealed that TDG can excise cytosine analogues harboring a C5 substituent (F, Br, OH) that enhances N1 acidity relative to cytosine, which is not excised.<sup>16</sup>

These previous findings had suggested that TDG could excise fC, given the expected enhancement in N1 acidity afforded by the electronic effect of a formyl group ( $\sigma_m = 0.35$  for CHO),<sup>28</sup> and this prediction was confirmed.<sup>9</sup> We evaluate the chemical properties of fC that are potentially relevant to its enzymatic excision here, including N1 acidity and resonance stabilization of the N1-deprotonated fC anion, and we examine the proposal that fC forms an imino tautomer and a wobble structure with guanine that is recognized by TDG.<sup>29,30</sup>

The chemical properties of caC that dictate the catalytic requirements for its excision are more complex, due to a number of factors. First, the effect of the carboxyl substituent on N1 acidity depends on its ionization state; acidity is expected to be enhanced by a protonated carboxyl group ( $\sigma_m =$

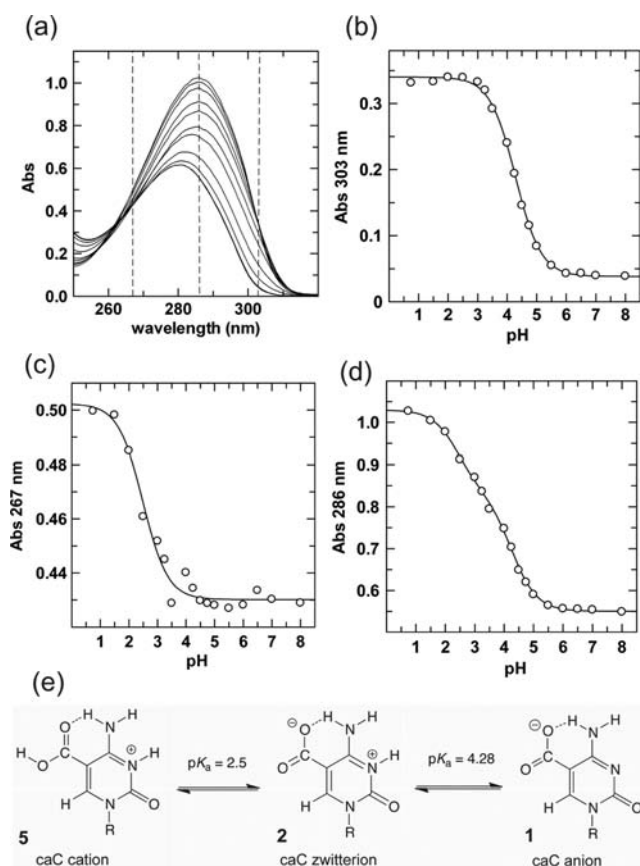
0.37)<sup>28</sup> and reduced by a deprotonated carboxyl ( $\sigma_m = -0.10$ ). A previous study reported the calculated N1 acidity for caC with COOH,<sup>31</sup> but the carboxyl is expected to be deprotonated (COO<sup>-</sup>) at physiological pH,<sup>32</sup> and N1 acidity for the caC monoanion is unknown. In addition, N3 ionization is expected to dramatically impact N1 acidity, but this has not been previously examined. Here, we report calculated acidities for several different ionization and tautomeric states of caC and for fC and other relevant pyrimidines. Because *N*-glycosidic bond hydrolysis can be acid-catalyzed, the ionization sites and  $pK_a$  values are relevant to enzymatic catalysis. A previous study on ionization of the 5-carboxyl-2'-deoxycytidine (5-ca-dC) nucleoside reported a  $pK_a$  for N3 but not for the carboxyl group (proposed to be <1).<sup>32</sup> We re-examined 5-ca-dC ionization here and provide  $pK_a$  values for both N3 and the carboxyl group.

We also determined the pH dependence of catalysis for fC and caC, and the results inform the mechanism by which TDG excises these bases. We also determined the role of conserved active-site residues in TDG excision of fC and caC, using site-directed mutagenesis and enzyme kinetics and equilibrium binding experiments. Together, the computational and biochemical results reveal key differences in the chemical properties of fC and caC that are relevant to enzymatic excision and uncover differences in the mechanism used by TDG to excise these bases, which differ by only a single atom. We also reveal a TDG variant that exhibits normal fC activity but no detectable caC activity, which provides a novel tool to investigate the mechanism of DNA demethylation in cells.

## RESULTS AND DISCUSSION

**Ionization of 5-Carboxyl-2'-deoxycytidine.** Given that *N*-glycosidic bond hydrolysis can be acid-catalyzed, the  $pK_a$  values for ionization at N3 and the carboxyl of caC are important for understanding enzymatic excision of this base from DNA. We collected UV absorbance spectra for 5-ca-dC under conditions of pH 0.5 to 8 (Figure 2a), and the results indicate two ionization events. Observation of isosbestic points for scans at pH  $\leq 2.5$  (257 nm, 303 nm) and pH  $> 4.5$  (267 nm) provide wavelengths that can be used to unambiguously determine each  $pK_a$ . The pH dependence of  $A_{303}$  reflects a single ionization, yielding a  $pK_a = 4.28 \pm 0.02$  (Figure 2b), and the same result is obtained for  $A_{257}$  ( $pK_a = 4.30 \pm 0.06$ , not shown). The pH dependence of  $A_{267}$  gives  $pK_a = 2.49 \pm 0.09$  (Figure 2c). Other wavelengths reflect both ionizations. Fitting the pH dependence of  $A_{286}$  gives  $pK_a^1 = 4.25 \pm 0.04$  and  $pK_a^2 = 2.41 \pm 0.09$  (Figure 2d), in excellent agreement with the values determined at the isosbestic points (Figure 2b,c). For comparison, we find that 2'-deoxycytidine (dC) ionizes at N3 with a  $pK_a = 4.31 \pm 0.01$  (Supporting Information Figure S1), which is identical to a previous finding.<sup>33</sup>

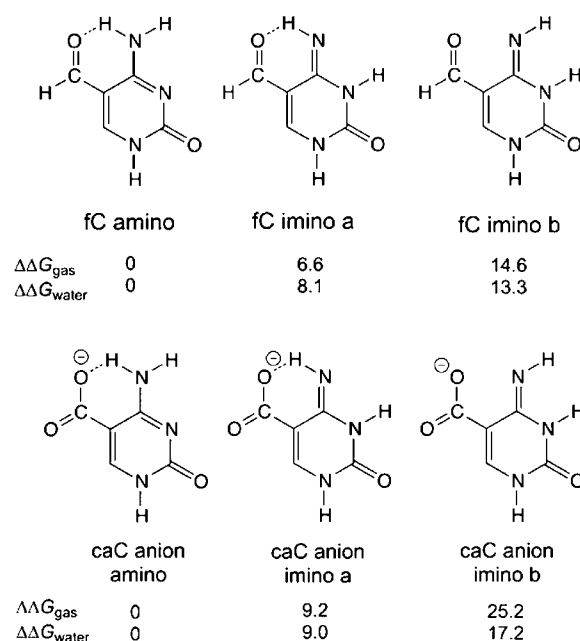
Regarding 5-ca-dC ionization, we assign the  $pK_a$  of 4.28 to ionization at N3 (Figure 2e). Given the small electronic effect for a carboxylate substituent ( $\sigma_m = -0.10$  for COO<sup>-</sup>),<sup>28</sup> which is probably offset by hydrogen bonding to the vicinal NH2 (Figure 2e), it is reasonable to find similar N3  $pK_a$  values for 5-ca-dC and dC. Assignment of the  $pK_a = 4.25$  to N3 is also supported by observation that this ionization exhibits a similar isosbestic point (267 nm) as that observed for N3 of dC (265 nm, Supporting Information Figure S1). Thus, we assign the  $pK_a = 2.45$  to the carboxyl of caC, which is consistent with previously determined  $pK_a$  values for a carboxyl with a vicinal NH2 in aromatic systems.<sup>32</sup> We consider the  $pK_a$  values



**Figure 2.** Ionization of N3 and the carboxyl of 5-ca-dC. (a) UV absorbance spectra for 5-ca-dC at pH 0.75–6.0; the absorbance is essentially unchanged for pH 6.0–8.0. Isosbestic points are observed for a subset of the scans at 257, 267, and 303 nm; absorbance at these wavelengths depends on only one of the two ionizations. (b) The pH dependence of  $A_{303}$  was fitted to eq 1, giving  $pK_a = 4.28 \pm 0.02$ ; (c) pH dependence of  $A_{267}$  was fitted to eq 1, giving  $pK_a = 2.49 \pm 0.09$ ; (d) pH dependence of  $A_{286}$  was fitted to eq 2, giving  $pK_a^1 = 4.25 \pm 0.04$  and  $pK_a^2 = 2.41 \pm 0.09$ . (e) Ionization of 5-ca-dC as indicated by our findings.

reported here to be more accurate than previously reported values of  $pK_a = 4$  for N3 and  $pK_a < 1$  for the carboxyl of 5-ca-dC, which were obtained from UV absorbance monitored at a single wavelength ( $A_{300}$ ).<sup>32</sup> Our findings confirm that caC exists as a monoanion at physiological pH and suggest that acid catalysis of 5-ca-dC hydrolysis is more likely to involve protonation at N3 than at the carboxylate group. Additional studies are needed to determine the  $pK_a$  values for N3 and the carboxyl of caC in duplex DNA, but they are not expected to be dramatically perturbed from the values reported here.

**Amino Tautomers of fC and caC Likely Predominate in DNA.** We next consider the tautomeric state of fC and caC that is likely to predominate in DNA under physiological conditions. It was proposed that fC and caC favor an imino tautomeric state (Figure 3) and thereby adopt a wobble structure when paired with guanine in DNA, similar to the structure of G·T and G·U mismatches, and that the wobble structure is a unifying feature of substrate recognition by TDG.<sup>29,30</sup> To examine this idea, we calculated the relative stability of the amino and imino tautomers for fC and caC (monoanion). As shown in Figure 3, the amino tautomers of fC and caC are much more stable than their imino counterparts in the gas phase and in water. Notably, NMR studies find that the



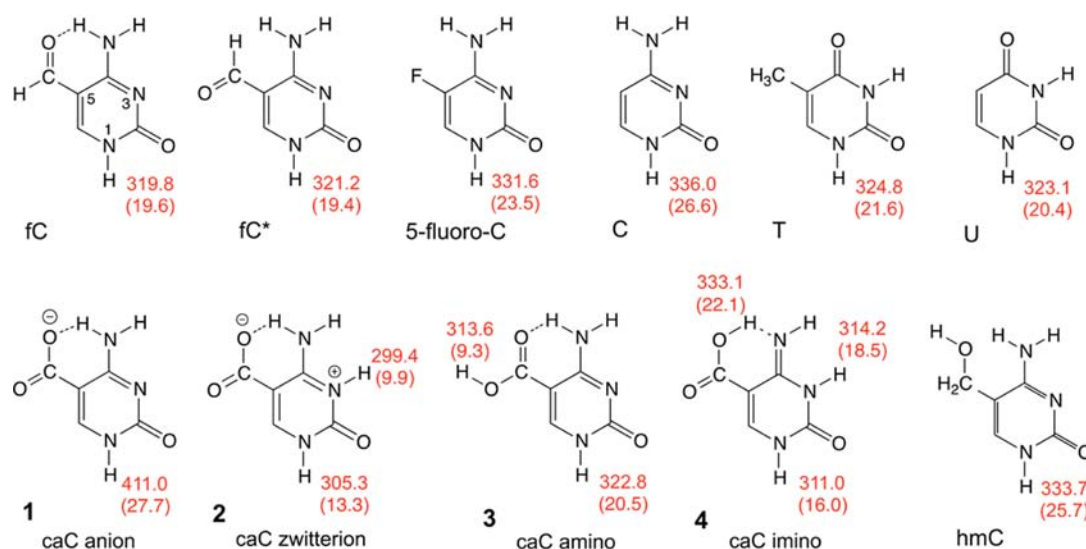
**Figure 3.** Calculated relative stabilities of the amino and imino tautomers of fC and the caC anion in the gas phase and water. The values are reported as the difference in free energy ( $\Delta\Delta G$ , kcal mol<sup>-1</sup>) with respect to the most stable species ( $\Delta\Delta G = 0$ ).

amino tautomer is the predominant form of 5-formyl-2'-deoxycytidine (in DMSO).<sup>34</sup> Moreover, DNA melting studies show that the stability of G·fC and G·caC base pairs is equal to or greater than that of G·C pairs.<sup>32,35</sup> These findings indicate that G·fC and G·caC pairs adopt the canonical Watson–Crick structure rather than the destabilized wobble structure that would be favored by imino tautomers of fC and caC. Together, these observations do not support the proposal that amino–imino tautomerization of fC and caC can explain how TDG selectively recognizes G·fC and G·caC pairs in DNA.<sup>29,30</sup>

#### Calculated Acidities of fC, caC, and Other Pyrimidines.

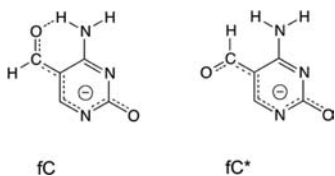
As discussed above, we previously showed that TDG activity depends on pyrimidine N1 acidity; that is, it increases with leaving group quality of the excised base.<sup>9,16</sup> Although active-site interactions can play a significant role, pyrimidine N1 acidity is an important factor in TDG activity. To understand the catalytic requirements for excision of fC and caC, we calculated the acidities for these bases and other relevant pyrimidines (Figure 4). The acidities are reported as the free energy required for deprotonation ( $\Delta G$ , kcal mol<sup>-1</sup>), where a lower value indicates greater acidity. Given the biochemical and structural evidence that the TDG active site is relatively nonpolar,<sup>16,36–38</sup> we calculated the acidities in the gas phase as well as in water. Notably, the acidity trends are the same for both media, though the differences are more pronounced in the gas phase. Our results are consistent with previous calculations for limited subsets of these pyrimidines.<sup>16,31,39,40</sup> However, N1 acidities have not previously been reported for the caC anion (1) and two of the neutral caC tautomers (2 and 4). As shown below, this new information is important for understanding the catalytic requirements for excision of caC from DNA.

The calculations indicate that fC is remarkably acidic (N1) compared to other pyrimidines that are excised by TDG, including U, T, and 5-fluorocytosine (5FC). Notably, an intramolecular hydrogen bond involving the formyl oxygen and the vicinal NH<sub>2</sub> of fC<sup>34,35</sup> stabilizes one rotamer over the other



**Figure 4.** Acidity for N1 and other sites for pyrimidines in the gas phase and bulk solution. The calculated acidities are reported as the free energy ( $\Delta G$ , kcal mol<sup>-1</sup>) required for deprotonation in the gas phase and in water (parentetical values).

by about 5 kcal/mol (Supporting Information Figure S2), but it does not substantially alter N1 acidity (compare fC and fC\* in Figure 4). The robust acidity of fC is attributable to the electronic effect of the formyl substituent ( $\sigma_m = 0.35$  for CHO)<sup>28</sup> and resonance stabilization of the fC anion, via charge delocalization to the formyl oxygen and O2 (Figure 5) ( $\sigma_p =$



**Figure 5.** Resonance stabilization of the fC anion.

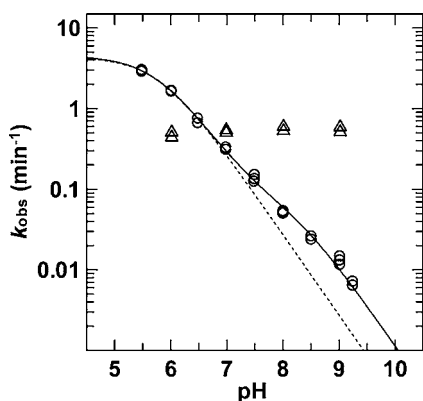
0.42,  $\sigma_p^- = 1.03$ ; para values given because  $\sigma_m^-$  is not available, to our knowledge).<sup>28</sup> The resonance effect likely accounts for the greater acidity of fC relative to 5FC, given the equivalent electronic effects for fluoro and CHO ( $\sigma_m$  of 0.34 and 0.35, respectively).<sup>28</sup> Previous findings that TDG excises fC some 17-fold faster than 5FC<sup>9,16</sup> can likely be explained by the much greater N1 acidity of fC relative to 5FC and perhaps by electrostatic catalysis of fC excision via stabilization of negative charge resonating to the formyl oxygen of the departing fC anion.<sup>9,16</sup>

The poor N1 acidity of hmC (Figure 4) likely accounts in large part for its resistance to excision by TDG, given previous findings that TDG rapidly excises 5-hydroxymethyluracil (hmU) and can therefore accommodate a hydroxymethyl group at the C5 position of pyrimidines.<sup>16</sup>

The predominant form of caC expected under physiological conditions, the monoanion (Figure 2), is likely to resist enzymatic excision, given that the dianion is expected to be unstable and thus a poor leaving group. Consistent with this notion, the caC anion **1** is much less acidic than pyrimidines that are excised by TDG, including fC, U, T, and 5FC (Figure 4). By contrast, the neutral forms of caC exhibit remarkable acidity, including the zwitterion **2** and the uncharged amino **3** and imino **4** tautomers. Indeed, **2** and **4** are substantially more acidic than U, T, and fC, and **3** exhibits similar acidity to these

bases. These relative acidities have implications for catalysis of caC excision, as discussed below. Our findings confirm that ionization of the carboxyl group has a major impact on N1 acidity (Figure 4, compare **1** and **3**), consistent with a large difference in the substituent electronic effects ( $\sigma_m$  is  $-0.10$  for COO<sup>-</sup> and 0.37 for COOH).<sup>28</sup> The calculations also show that protonation at N3 greatly increases N1 acidity of caC, even if the carboxyl remains deprotonated (compare **1** and **2**). As shown above, this N3 ionization occurs with  $pK_a = 4.3$  for 5-ca-dC (Figure 2). Resonance effects likely contribute to acidity of the caC tautomers, as noted above for fC (Figure 5). Thus, our calculations indicate that protonation of the caC monoanion to give one of the neutral caC tautomers (**2**, **3**, or **4**), that is, acid catalysis, could be an effective strategy for enzymatic excision of caC from DNA.

**pH Dependence of Catalysis for Excision of fC and caC.** To examine the potential role for acid catalysis of caC excision, we determined the pH dependence of TDG activity for caC and fC. The kinetics experiments were performed under saturating enzyme conditions such that the rate constants are not influenced by enzyme–substrate association or events after the chemical step (product release or inhibition).<sup>16</sup> As shown in Figure 6, the fC excision activity of TDG is essentially unchanged for pH 6.0–9.0. This is consistent with expectations; while N3 protonation would likely enhance fC excision, it is unlikely to occur given a  $pK_a$  value of 2.6 for N3 of fC.<sup>33</sup> Notably, we previously shown that TDG activity for G·U and G·T substrates is also nearly constant for pH 6–9.<sup>41</sup> These findings indicate no evidence of acid or base catalysis for G·fC, G·U, or G·T substrates. In stark contrast, caC activity increases with decreasing pH;  $k_{obs}$  is 420-fold higher at pH 5.5 than at pH 9.25 (Figure 6). A previous study found that caC activity is 10-fold higher at pH 5.5 versus pH 8.0 for the *catalytic domain* of TDG (residues 111–308).<sup>29</sup> However, we find that full-length TDG (410 residues) exhibits a much higher 57-fold difference in caC activity for the same pH values (Figure 6). This discrepancy could indicate that a full-length enzyme should be used for examining the pH dependence of catalysis. Previous studies show that the N-terminal region of TDG contributes to binding and base excision, particularly for G·T substrates,<sup>42–44</sup> and it could potentially contribute to caC activity.

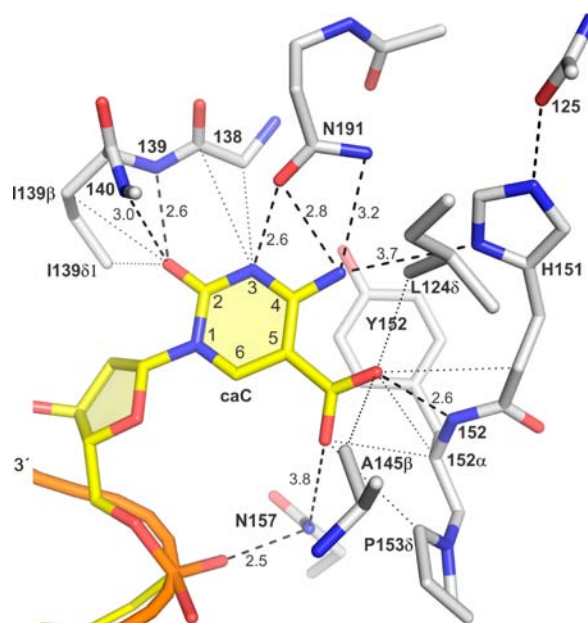


**Figure 6.** pH dependence of TDG activity for G-fC ( $\Delta$ ) and G-caC (O) substrates. Fitting the G-caC data to a model for ionization of an essential protonated group (eq 4; dotted line) gives an apparent  $pK_a = 5.80 \pm 0.03$  and a limiting  $k_{obs}$  of  $4.4 \pm 0.1 \text{ min}^{-1}$ , but fitting is poor for  $\text{pH} > 7$ . Fitting to a model with an essential protonated group and a second, non-essential group (eq 5, solid line) gives apparent  $pK_a$  values of  $pK_a^1 = 5.75 \pm 0.03$  and  $pK_a^2 = 8.2 \pm 0.7$ , a limiting  $k_{obs}$  of  $4.5 \pm 0.1 \text{ min}^{-1}$ , and a rate enhancement factor ( $1 + \alpha$ ) of 4.5 (i.e., the fold increase in  $k_{obs}$  resulting from deprotonation of the second group).

Because the kinetics experiments were conducted under saturating enzyme conditions, the pH profile for caC excision reflects the ionization of one or more groups in the enzyme–substrate (ES) complex. We fitted the caC data to a standard equation for ionization of a single essential protonated group (Figure 6, dotted line); this model assumes that no activity remains when the ionizing group is deprotonated (i.e., log linear, slope =  $-1$ ). While this yields an apparent  $pK_a$  of  $5.80 \pm 0.03$ , the  $k_{obs}$  values are displaced increasingly above the fitted curve for  $\text{pH} > 7$  ( $k_{obs}$  is about 1.7-, 2.8-, and 4.3-fold greater than predicted for  $\text{pH}$  8.0, 8.5, and 9.0, respectively). This deviation could reflect ionization of a second, non-essential group that leads to modestly ( $\sim 4.3$ -fold) higher activity upon deprotonation. The fitting is improved for a double-ionization model (Figure 6, solid line), giving apparent  $pK_a$  values of  $pK_a^1 = 5.75 \pm 0.03$  for the essential protonated group and  $pK_a^2 = 8.2 \pm 0.7$  for the non-essential deprotonated group.

We conclude that the essential protonated group ( $pK_a^1 = 5.75$ ) is the caC base, rather than a TDG side chain serving as a general acid (i.e., to protonate caC). Crystal structures reveal only two residues that could potentially serve as an essential general acid, H151 and Y152 (Figure 7). However, the H151A and Y152F variants exhibit pH profiles for caC excision that are very similar to that of native TDG (Supporting Information Figure S3). Thus, H151 and Y152 do not serve as an essential general acid nor can they be assigned as the second (non-essential) ionizing group. Together, the data support a model whereby TDG excises a neutral form of the caC base but not the caC anion, and deprotonation of a second group, likely from the enzyme, enhances this activity. Such a model is consistent with the very poor calculated N1 acidity of the caC anion, which suggests that it is unlikely to be excised by TDG.

It seems reasonable that the proton needed for ionization of caC in the ES complex is derived from solvent, given that caC excision is relatively slow (minute time scale), and that the TDG active site is relatively permissive and protected from solvent by two loops that move with nucleotide flipping and by residues in the disordered N-terminal region (L124, Figure 7). Observation that the free 5-ca-dC nucleoside ionizes with  $pK_a$



**Figure 7.** Previously reported structure of TDG (catalytic domain) with a 5-carboxyl-dC analogue (noncleavable) flipped into the active site (PDBID: 3UOB).<sup>38</sup> Hydrogen bonds (dashed lines) and van der Waals contacts (dotted lines,  $d \leq 3.7 \text{ \AA}$ ) are shown. Similar interactions are observed for a structure of the N140A-TDG variant (catalytic domain) bound to DNA containing an A-caC mismatch (PDBID: 3UO7).<sup>38</sup>

values of 4.28 for N3 and 2.45 for the carboxyl (Figure 2) suggests that protonation of caC in the ES complex ( $pK_a^1 = 5.75$ ) occurs at N3 rather than the carboxyl, to give either the zwitterion 2 or the uncharged imino tautomer 4 (Figure 4). This is supported by findings that activity is much higher for caC versus fC at low pH (Figure 6) and that N3-protonated forms of caC (2 and 4) are more acidic than fC while the carboxyl (COOH) form 3 is less acidic (Figure 4). Additional evidence for protonation of caC at N3 rather than the carboxyl is provided by the mutational studies below. Our finding that caC excision is acid-catalyzed accounts for the initially puzzling observation that caC is excised  $\sim 5$ -fold slower than fC at  $\text{pH}$  7.5,<sup>9</sup> even though the N3-protonated forms of caC (2 and 4) are more acidic than fC (Figure 4); the apparent  $pK_a^1 = 5.75$  indicates a low population of N3-protonated caC and a high population of the excision-resistant anion 1 in the ES complex at  $\text{pH}$  7.5.

#### Mutational Analysis of fC and caC Excision by TDG.

We previously characterized the role of conserved side chains in TDG excision of uracil or thymine (G·U or G·T mispairs),<sup>36</sup> but their contribution to fC and caC excision was unknown. Crystal structures of DNA-bound TDG, with either caC (Figure 7)<sup>38</sup> or uracil<sup>36</sup> flipped into its active site, reveal potential catalytic interactions between the flipped base and enzyme backbone and side chain groups. A structure is not available for the TDG-fC complex, but existing structures suggest interactions that could facilitate fC excision. We used mutagenesis and kinetics experiments to examine the role of four conserved side chains in the excision of fC and caC, and the results are given in Table 1. Because the experiments were performed under saturating enzyme conditions, the rate constants ( $k_{max}$ ) reflect the maximal rate of product formation without influence from enzyme–substrate association or product release or product inhibition.<sup>16,45</sup>

Table 1. Glycosylase Activity for TDG and Variants<sup>a</sup>

| substrate | enzyme | $k_{\max}$ (min <sup>-1</sup> ) | relative to wild-type TDG |
|-----------|--------|---------------------------------|---------------------------|
| G·fC      | TDG    | 0.61 ± 0.04                     |                           |
|           | A145G  | 0.22 ± 0.02                     | 0.36                      |
|           | H151A  | 0.35 ± 0.06                     | 0.57                      |
|           | Y152F  | 0.91 ± 0.09                     | 1.5                       |
|           | N157A  | 0.14 ± 0.03                     | 0.23                      |
|           | N191A  | 1.00 ± 0.03                     | 1.6                       |
| G·caC     | TDG    | 0.14 ± 0.01                     |                           |
|           | A145G  | 0.60 ± 0.01                     | 4.3                       |
|           | H151A  | 0.072 ± 0.002                   | 0.51                      |
|           | Y152F  | 0.29 ± 0.04                     | 2.1                       |
|           | N157A  | 0.13 ± 0.001                    | 0.89                      |
|           | N191A  | ND                              |                           |

<sup>a</sup>The  $k_{\text{obs}}$  values are the mean and standard deviation for three independent experiments collected at 22 °C under saturating enzyme conditions and analyzed by HPLC. ND, activity not detected. Data for wild-type TDG was previously reported.<sup>9</sup>

Regarding fC excision, we find small (<2-fold) effects on  $k_{\text{obs}}$  for the H151A, Y152F, and N191A mutations (Table 1), indicating that the H151 and N191 side chains and the hydroxyl of Y152 contribute minimally to fC excision. The A145G mutation causes a 2.7-fold decrease in  $k_{\text{obs}}$ . Thus, fC is the only substrate identified to date for which A145 facilitates catalysis; A145 curtails activity for G·T<sup>36</sup> and G·caC substrates (Table 1). The crystal structure with flipped caC raises the possibility that the A145 methyl contacts the formyl group of fC (Figure 7), which could potentially help position the flipped base to optimize catalytic interactions with its formyl or O2 oxygen. The N157A mutation causes a 4-fold loss in  $k_{\text{obs}}$  for G·fC activity (Table 1). Previous studies found that the same mutation diminishes activity for G·U, G·hmU, and G·T substrates.<sup>27,30</sup> This damaging effect for numerous substrates can be reasonably explained by disruption of the contact between the N157 side chain NH2 and the 5'-phosphate of the flipped nucleotide (Figure 7). Consistent with this idea, the N157D mutation causes a greater loss in G·U and G·hmU activity than the N157A mutation,<sup>42,46</sup> as expected for electrostatic repulsion between negatively charged Asp side chain and the 5'-phosphate. Notably, this Asn is conserved in prokaryotic MUG enzymes that are homologous to TDG and in the more distally related UNG enzymes. Moreover, the Asn-phosphate contact is seen in all DNA-bound structures of TDG, MUG, and UNG.<sup>29,36–38,46,47</sup>

Our finding that A145, H151, Y152, N157, and N191 contribute minimally or not at all to excision of fC is consistent with its robust N1 acidity (Figure 4), which renders it inherently amenable to enzymatic excision. Nevertheless, crystal structures suggest that three backbone amide groups of TDG could contact the O2 and formyl oxygens of fC (Figure 7). These contacts could help retain the flipped fC in the active site and promote C–N bond cleavage by stabilizing the departing fC anion. While these interactions could be sufficient to catalyze fC excision, we cannot rule out a potential role for side chains other than the five examined here.

We next consider the role of the same five active-site residues in caC excision by TDG. We find a negligible role for H151 in caC excision (Table 1), which is consistent with findings that the imidazole is distal from the caC base (Figure 7). Our finding does not support the proposal that the imidazole (protonated) of H151 contributes to recognition of the caC

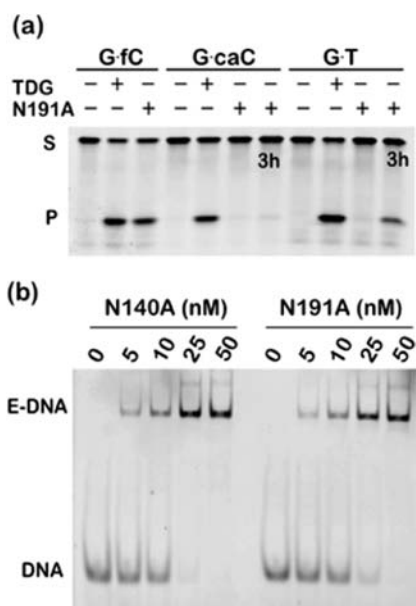
anion.<sup>38</sup> The Y152F mutation gives a 2.1-fold increase in activity, demonstrating that the hydroxyl of Y152 is dispensable and suggesting that caC excision is modestly enhanced by the more hydrophobic environment (benzyl versus phenol). Remarkably, the A145G mutation gives a 4-fold increase in caC activity (Table 1). Crystal structures indicate that the A145 methyl contacts the caC carboxyl (Figure 7).<sup>38</sup> The nonpolar methyl could potentially disfavor flipping of the caC anion or impede the formation of optimal catalytic interactions with other active-site groups. This finding is reminiscent of the 13-fold increase in G·T activity caused by the A145G mutation,<sup>36</sup> raising the question of why a residue that curtails activity for two biological substrates is strictly conserved in TDG enzymes (vertebrates). Our previous studies indicate that A145 counters aberrant excision of T from A·T pairs,<sup>36</sup> and we find here that it contributes to fC excision.

The N157A mutation has no significant effect on caC excision (Table 1), even though it does adversely impact activity for G·fC (Table 1), G·U, G·hmU, and G·T substrates.<sup>27,30</sup> These findings do not support the proposal that the N157A mutation has no effect on substrate specificity.<sup>30</sup> As noted above, N157 contacts the 5'-phosphate of the flipped nucleotide (Figure 7) and is highly conserved in TDG, MUG, and UNG enzymes. The absence of a damaging effect for N157A on caC activity suggests that the expected loss could be offset by an effect for Ala at position 157 that favors excision of caC but not the other bases (fC, U, hmU, T). One possibility is that the neutral carboxyl (COOH) of caC is stabilized to a greater extent by Ala versus Asn; that is, Ala could favor neutral forms of caC (3 and 4) that are more acidic than the monoanion 1.

Strikingly, the N191A variant has no detectable caC excision activity (Table 1 and Figure 8a), even for extended reaction times (up to 3 h) and enzyme concentrations that greatly exceed the saturating level for nonspecific DNA (5–10 μM ≫  $K_d^{\text{NS}} = 0.2 \mu\text{M}$ ).<sup>48</sup> This is not due to a mutational effect on protein stability because N191A-TDG retains full fC activity (Table 1 and Figure 8a) and substantial activity for G·U mispairs.<sup>36</sup> Moreover, an electrophoretic mobility shift assay (EMSA) shows that the N191A mutation does not substantially weaken the binding of TDG to DNA containing a G·caC pair (Figure 8b). These observations indicate that N191 plays a key role in the chemical step of the reaction for excision of caC, a role that is not needed for fC excision.

**Implications for Catalysis.** Together, the calculated acidities, mutational studies, and previous crystal structures suggest that N191 facilitates acid catalysis of caC excision, a function that is not required for fC excision (Figure 9). Crystal structures indicate that the N191 side chain oxygen ( $O_{\delta 1}$ ) contacts N3 of caC (Figure 7). This interaction would favor the N3-protonated forms of caC (2 and 4) but not the anion or the neutral amino tautomer 3 and could account for findings that the N3  $pK_a$  of caC is increased by 1.5 units in the enzyme–substrate complex relative to the 5-ca-dC nucleoside.

While the zwitterion 2 is acidic and should be highly amenable to TDG excision, a nonpolar active site could favor the facile conversion of 2 to the imino tautomer 4 (Figure 9 and Supporting Information Figure S4). Calculations show that 2 is more stable than 4 in water ( $\Delta\Delta G_{\text{water}} = 5 \text{ kcal mol}^{-1}$ ), while 4 is much more stable in the gas phase ( $\Delta\Delta G_{\text{gas}} = 9 \text{ kcal mol}^{-1}$ ; Supporting Information Figure S4). As such, 4 might be favored in the TDG active site, which is relatively nonpolar.<sup>16,36–38,62</sup> Indeed, TDG forms many nonpolar contacts,



**Figure 8.** Effect of the N191A mutation on substrate binding and base excision. (a) Electrophoretic glycosylase assay shows that N191A-TDG has normal G-fC activity, markedly reduced G-T activity, and no significant G-caC activity, even for a reaction time of 3 h. Reactions were performed with 0.20  $\mu\text{M}$  enzyme, 0.10  $\mu\text{M}$  substrate, and were quenched after 5 min or 3 h (as indicated). (b) Electrophoretic mobility shift assay (EMSA) shows that N191A-TDG binds G-caC DNA with high affinity, similar to that observed for N140A-TDG. Previous studies show that N140A-TDG binds G-caC and other substrates with essentially the same affinity and/or enzyme–substrate contacts as native TDG,<sup>38,45</sup> but it does not excise caC (under the EMSA conditions here). DNA (10 nM) was incubated with enzyme (5–50 nM) at 22 °C for 30 min prior to running the EMSA.

one hydrogen bond, and no ionic interactions with the carboxyl of caC (Figure 7). Notably, 4 could also be derived directly from the caC anion via coupled abstraction of an amino ( $\text{NH}_2$ ) proton by  $\text{COO}^-$  and protonation at N3 (Figure 9). This would likely be facilitated by N191 and perhaps by the nonpolar contacts with the carboxyl (Figure 7), which could favor the  $\text{COOH}$  of 4 over the  $\text{COO}^-$  of 1.

**Concluding Remarks.** The studies reported here advance our understanding of the chemical properties of fC and caC that dictate the catalytic requirements for their excision from DNA and the mechanism by which TDG excises these oxidized forms of mC. We show that the 5-ca-dC nucleoside ionizes with  $\text{pK}_a$  values of 4.28 (N3) and 2.45 (carboxyl), confirming that caC exists as a monoanion at physiological pH. The calculated stabilities of the amino and imino tautomers for fC and caC (anion) do not support the proposal that G-fC and G-caC pairs adopt a wobble structure that is recognized by TDG.<sup>29,30</sup> The calculated N1 acidities for fC and the neutral forms of caC are comparable to other TDG substrates, but the caC anion exhibits poor acidity, indicating resistance to enzymatic excision. We find that TDG excision of fC is pH-independent, while excision of the caC anion is acid-catalyzed (Figure 9). The pH profile indicates that caC ionizes in the enzyme–substrate complex with an apparent  $\text{pK}_a$  of 5.75, likely at N3. Mutational analysis of five conserved active-site side chains reveals that none are critical for excision of fC, in keeping with its robust N1 acidity. Nevertheless, fC excision likely requires electrostatic catalysis, and crystal structures suggest this could involve backbone amide groups contacting the formyl and O2

oxygens of fC (Figure 9). In contrast to the results for fC, N191 is essential for excision of caC. The results suggest N191 facilitates acid catalysis by stabilizing an N3-protonated form of caC (2 and 4).

Our finding that N191A-TDG possesses normal fC activity but no detectable caC activity suggests that this variant could be useful for examining the role of TDG in active DNA demethylation in mammalian cells. In particular, N191A-TDG could potentially be used to examine the possibility that excision of fC, rather than caC, constitutes the major role of TDG in a TET-initiated pathway for active DNA demethylation. Observation that the N191A variant also has diminished G-T activity<sup>36</sup> should not preclude its potential utility in this regard, because G-T mispairs do not arise in the TET-initiated demethylation pathway. A recent study found that N157D-TDG exhibits selective activity for caC over fC at pH 6, but at neutral pH, this variant has greatly reduced activity for both caC and fC substrates compared to native TDG.<sup>30</sup> Thus, the utility of N157D-TDG for cellular studies is unclear. Nevertheless, a TDG variant that exhibits selective activity for caC over fC under physiological conditions could be useful for studying the function of TDG in active DNA demethylation; our findings could inform the design of such a variant.

## EXPERIMENTAL SECTION

**Materials.** The duplex DNA substrates consisted of 5'-GGGA-GAAGAGGAGGAAXGAAGAGAGCTC, where x = fC, caC, or T, and a complementary strand that places G opposite the target base (x). This 28 bp DNA construct also places the target base (x) in a CpG context, consistent with the specificity of TDG. For reactions analyzed by denaturing PAGE, the target strand included a 3'-fluorescein-dT. The oligodeoxynucleotides were synthesized, purified, and quantified as described.<sup>9</sup> Full-length human TDG was expressed in *Escherichia coli* and purified as described.<sup>49</sup> The TDG variants A145G, H151A, and N191A were prepared as described.<sup>36</sup> Expression plasmids for Y152F-TDG and N157A-TDG were generated using site-directed mutagenesis<sup>45</sup> (primers provided in Supporting Information Table S1), and the mutation was confirmed by DNA sequencing. The Y152F-TDG and N157A-TDG variants were expressed and purified using the protocol for native TDG.<sup>36,49</sup> 5-Carboxyl-2'-deoxycytidine was from Berry & Associates.

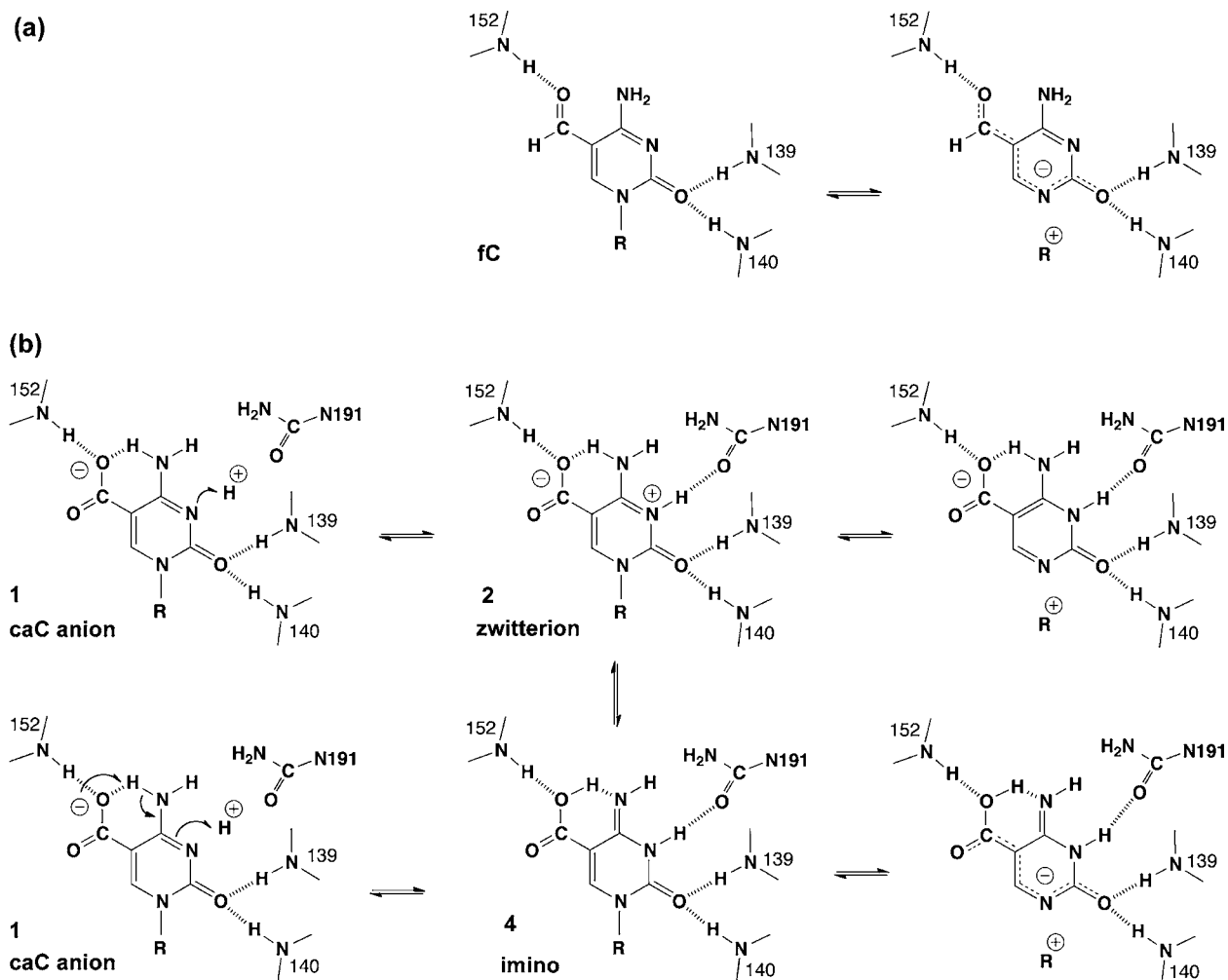
**$\text{pK}_a$  Determinations.** Ionization of 5-ca-dC (100  $\mu\text{M}$ ) was monitored by UV absorbance in buffers of varying pH (0.5–8.0) consisting of 25 mM sodium phosphate, 25 mM sodium acetate, and 50 mM NaCl. The pH dependence of absorbance at a given wavelength was fitted by nonlinear regression using equations for a single ionization (eq 1) or double ionization (eq 2) with Graft 6.0.<sup>50</sup>

$$A = A_1 + \frac{(A_2 - A_1)10^{(\text{pH}-\text{pK}_a)}}{10^{(\text{pH}-\text{pK}_a)} + 1} \quad (1)$$

$$A = A_1 + \frac{(A_2 - A_1)10^{(\text{pH}-\text{pK}_{a1})}}{10^{(\text{pH}-\text{pK}_{a1})} + 1} + \frac{(A_3 - A_2)10^{(\text{pH}-\text{pK}_{a2})}}{10^{(\text{pH}-\text{pK}_{a2})} + 1} \quad (2)$$

**Glycosylase Activity.** The kinetics experiments used to determine the glycosylase activity of TDG and the TDG variants were performed essentially as described,<sup>36</sup> using a saturating concentration of enzyme (5 or 10  $\mu\text{M}$ ) and DNA substrate concentrations of 0.5 or 1.0  $\mu\text{M}$ , in HEMN.1 buffer (0.02 M HEPES pH 7.5, 0.1 M NaCl, 0.2 mM EDTA, 2.5 mM  $\text{MgCl}_2$ ) at 22 °C. The reactions were monitored by electrophoresis<sup>9</sup> or anion-exchange HPLC under denaturing conditions.<sup>16</sup> The HPLC data were fitted by nonlinear regression to a single exponential equation (eq 3).

$$\text{fraction product} = A(1 - e^{-k_{\text{obs}}t}) \quad (3)$$



**Figure 9.** Potential mechanisms for TDG excision of fC and caC as suggested by the results here and previous structural studies.<sup>38</sup> (a) Excision of fC does not require a large role for the side chain of any active-site residue examined here, consistent with the robust N1 acidity of fC (Figure 4). Nevertheless, it seems likely that fC excision involves electrostatic catalysis, where backbone amide groups (139, 140, 152; Figure 7) stabilize the departing fC anion. (b) In contrast to the results for fC, our findings indicate that excision of the caC anion, the predominant species at neutral pH, requires acid catalysis. Findings that N191 is essential for caC excision but dispensable for fC activity suggest that N191 is needed to stabilize an N3-protonated form of caC, which could be the zwitterion 2 or the neutral imino tautomer 4. It seems likely that excision of caC also involves electrostatic catalysis, via the same three backbone amide groups.

where  $A$  is the amplitude,  $k_{\text{obs}}$  is the observed rate constant, and  $t$  is reaction time. Because the experiments were performed under saturating enzyme conditions ( $[E] > [S] \gg K_d$ ), the rate constants ( $k_{\text{obs}}$ ) reflect the maximal rate of product formation (i.e.,  $k_{\text{obs}} \approx k_{\text{max}}$ ) and are not influenced by product release or product inhibition. The attainment of saturating enzyme conditions was confirmed by observation that using a 2-fold higher enzyme concentration yielded the same  $k_{\text{obs}}$  value (within error). Notably, previous findings from kinetics and equilibrium binding experiments show that TDG binds very tightly to G·T, G·fC, and G·caC substrates and tightly to nonspecific DNA ( $K_d = 0.1\text{--}0.3 \mu\text{M}$ ).<sup>16,36,45,48,49,51</sup>

**pH Dependence of Activity.** The dependence of TDG activity ( $k_{\text{obs}}$ ) on pH was monitored for G·fC and G·caC substrates at 22 °C using a buffer consisting of 0.01 M NaMES, 0.01 M NaHEPES, 0.01 M Tris, 0.01 M NaCHES, 0.1 M NaCl, 2.5 mM MgCl<sub>2</sub>, and 0.2 mM EDTA. The caC activity of TDG could not be determined for pH < 5.5 or pH > 9.25, due likely to loss of TDG structural integrity, consistent with our previous observations for pH dependence of activity for G·U and G·T substrates and pH effects on enzyme stability.<sup>41</sup> As noted above, kinetics experiments were performed under saturating enzyme conditions, as confirmed by observation of equivalent  $k_{\text{obs}}$  values for multiple TDG concentrations (ranging from 2.5 to 15  $\mu\text{M}$ ). The pH profile for caC activity was fitted by nonlinear regression to equations

involving ionization of a single group (eq 4) or two groups (eq 5) in the enzyme–substrate complex:

$$k_{\text{obs}} = \frac{k_1}{10^{(\text{pH}-\text{pKa})} + 1} \quad (4)$$

$$k_{\text{obs}} = \frac{k_1 + \alpha k_1 / (10^{(\text{pKa}2-\text{pH})} + 1)}{10^{(\text{pH}-\text{pKa}1)} + 1} \quad (5)$$

For eqs 4 and 5,  $k_{\text{obs}}$  is the observed rate constant and  $k_1$  is the limiting rate constant.

**Equilibrium Binding Experiments.** The equilibrium binding of enzyme to DNA was monitored using an electrophoretic mobility shift assay (EMSA), essentially as previously described.<sup>48</sup> The binding reactions (30  $\mu\text{L}$ ) were performed by incubating G·caC DNA (10 nM) with varying concentrations of enzyme (5–50 nM) at room temperature for 30 min in binding buffer (25 mM HEPES pH 7.5, 0.1 M NaCl, 1 mM EDTA, 1 mM DTT, 0.1 mg/mL BSA, 5% glycerol). Samples were loaded to a precast 6% native polyacrylamide gel (Invitrogen), and electrophoresis was performed for 60 min at 100 V, 5 °C. The gels were analyzed using a Typhoon 9400 imager (GE Healthcare) in the fluorescence mode to detect the 3'-fluorescein-labeled DNA.



**Computational Studies.** The gas-phase calculations were conducted at the B3LYP/6-31+G(d) level of theory using Gaussian 03 and Gaussian 09.<sup>52–56</sup> Structures were fully optimized in the gas phase and frequencies calculated (no imaginary frequencies were found). Gas-phase acidity and relative stability values are reported as  $\Delta G$  in kcal mol<sup>-1</sup>. The only exception is the caC zwitterion **2** (Figure 4), which is not a stable minimum in the gas phase and therefore is partially optimized. For the caC zwitterion **2**,  $\Delta G$  values are estimated from  $\Delta E$  values. Dielectric medium calculations were done using the conductor-like polarizable continuum solvent model (CPCM, fully optimized at B3LYP/6-31+G(d) with UFF cavity) as implemented in Gaussian 03.<sup>57–59</sup> The one exception is the N3–H acidity of **4** (Figure 4); the deprotonated structure is not stable in solution, so the gas-phase optimized structure was used. The “total free energy in solution” ( $\Delta G$ ) values are reported, and these account for the free energy of solvation of a proton (–265.9 kcal mol<sup>-1</sup>).<sup>60,61</sup>

## ■ ASSOCIATED CONTENT

### ● Supporting Information

Supplementary Figures S1–S4, Supplementary Table S1, and the Cartesian coordinates and energies (in Hartree) for the calculated species. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*Address correspondence to [adrohat@som.umaryland.edu](mailto:adrohat@som.umaryland.edu), [jeehiun@rci.rutgers.edu](mailto:jeehiun@rci.rutgers.edu).

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by the U.S. National Institutes of Health (GM72711 to A.C.D.) and the U.S. National Science Foundation (to A.Z.M. and J.K.L.).

## ■ REFERENCES

- (1) Nabel, C. S.; Manning, S. A.; Kohli, R. M. *ACS Chem. Biol.* **2012**, *7*, 20.
- (2) Jones, P. A.; Baylin, S. B. *Nat. Rev. Genet.* **2002**, *3*, 415.
- (3) Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. *Science* **2009**, *324*, 930.
- (4) Ito, S.; D'Alessio, A. C.; Taranova, O. V.; Hong, K.; Sowers, L. C.; Zhang, Y. *Nature* **2010**, *466*, 1129.
- (5) Penn, N. W.; Suwalski, R.; O'Riley, C.; Bojanowski, K.; Yura, R. *Biochem. J.* **1972**, *126*, 781.
- (6) He, Y. F.; Li, B. Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; Sun, Y.; Li, X.; Dai, Q.; Song, C. X.; Zhang, K.; He, C.; Xu, G. L. *Science* **2011**, *333*, 1303.
- (7) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. *Science* **2011**, *333*, 1300.
- (8) Pfaffeneder, T.; Hackner, B.; Truss, M.; Munzel, M.; Muller, M.; Deiml, C. A.; Hagemeyer, C.; Carell, T. *Angew. Chem., Int. Ed.* **2011**, *50*, 7008.
- (9) Maiti, A.; Drohat, A. C. *J. Biol. Chem.* **2011**, *286*, 35334.
- (10) Raiber, E. A.; Beraldi, D.; Ficiz, G.; Burgess, H.; Branco, M. R.; Murat, P.; Oxley, D.; Booth, M. J.; Reik, W.; Balasubramanian, S. *Genome Biol.* **2012**, *13*, R69.
- (11) Nabel, C. S.; Jia, H.; Ye, Y.; Shen, L.; Goldschmidt, H. L.; Stivers, J. T.; Zhang, Y.; Kohli, R. M. *Nat. Chem. Biol.* **2012**, *8*, 751.
- (12) Song, C. X.; Szulwach, K. E.; Dai, Q.; Fu, Y.; Mao, S. Q.; Lin, L.; Street, C.; Li, Y.; Poidevin, M.; Wu, H.; Gao, J.; Liu, P.; Li, L.; Xu, G. L.; Jin, P.; He, C. *Cell* **2013**, *153*, 678.
- (13) Shen, L.; Wu, H.; Diep, D.; Yamaguchi, S.; D'Alessio, A. C.; Fung, H. L.; Zhang, K.; Zhang, Y. *Cell* **2013**, *153*, 678. Shen, L.; Wu,

H.; Diep, D.; Yamaguchi, S.; D'Alessio, A. C.; Fung, H. L.; Zhang, K.; Zhang, Y. *Cell* **2013**, *153*, 692.

(14) Masaoka, A.; Matsubara, M.; Hasegawa, R.; Tanaka, T.; Kurisu, S.; Terato, H.; Ohyama, Y.; Karino, N.; Matsuda, A.; Ide, H. L. *Biochemistry* **2003**, *42*, 5003.

(15) Morera, S.; Grin, L.; Vigouroux, A.; Couve, S.; Henriot, V.; Saparbaev, M.; Ishchenko, A. A. *Nucleic Acids Res.* **2012**, *40*, 9917.

(16) Bennett, M. T.; Rodgers, M. T.; Hebert, A. S.; Ruslander, L. E.; Eisele, L.; Drohat, A. C. *J. Am. Chem. Soc.* **2006**, *128*, 12510.

(17) Cortazar, D.; Kunz, C.; Selfridge, J.; Lettieri, T.; Saito, Y.; Macdougall, E.; Wirz, A.; Schuermann, D.; Jacobs, A. L.; Siegrist, F.; Steinacher, R.; Jiricny, J.; Bird, A.; Schar, P. *Nature* **2011**, *470*, 419.

(18) Cortellino, S.; Xu, J.; Sannai, M.; Moore, R.; Caretti, E.; Cigliano, A.; Le Coz, M.; Devarajan, K.; Wessels, A.; Soprano, D.; Abramowitz, L. K.; Bartolomei, M. S.; Rambow, F.; Bassi, M. R.; Bruno, T.; Fanciulli, M.; Renner, C.; Klein-Szanto, A. J.; Matsumoto, Y.; Kobi, D.; Davidson, I.; Alberti, C.; Larue, L.; Bellacosa, A.; et al. *Cell* **2011**, *146*, 67.

(19) Wiebauer, K.; Jiricny, J. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 5842.

(20) Werner, R. M.; Stivers, J. T. *Biochemistry* **2000**, *39*, 14054.

(21) Dinner, A. R.; Blackburn, G. M.; Karplus, M. *Nature* **2001**, *413*, 752.

(22) McCann, J. A. B.; Berti, P. J. *J. Am. Chem. Soc.* **2008**, *130*, 5789.

(23) Berti, P. J.; McCann, J. A. *Chem. Rev.* **2006**, *106*, 506.

(24) Chen, X. Y.; Berti, P. J.; Schramm, V. L. *J. Am. Chem. Soc.* **2000**, *122*, 1609.

(25) Drohat, A. C.; Stivers, J. T. *J. Am. Chem. Soc.* **2000**, *122*, 1840.

(26) Drohat, A. C.; Jagadeesh, J.; Ferguson, E.; Stivers, J. T. *Biochemistry* **1999**, *38*, 11866.

(27) Hashimoto, H.; Zhang, X.; Cheng, X. *DNA Repair* **2013**, *12*, 535.

(28) Hansch, C.; Leo, A.; Taft, R. W. *Chem. Rev.* **1991**, *91*, 165.

(29) Hashimoto, H.; Hong, S.; Bhagwat, A. S.; Zhang, X.; Cheng, X. *Nucleic Acids Res.* **2012**, *40*, 10203.

(30) Hashimoto, H.; Zhang, X.; Cheng, X. *J. Mol. Biol.* **2013**, *425*, 971.

(31) Williams, R. T.; Wang, Y. *Biochemistry* **2012**, *51*, 6458.

(32) Sumino, M.; Ohkubo, A.; Taguchi, H.; Seio, K.; Sekine, M. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 274.

(33) La Francois, C. J.; Jang, Y. H.; Cagin, T.; Goddard, W. A.; Sowers, L. C. *Chem. Res. Toxicol.* **2000**, *13*, 462.

(34) LaFrancois, C. J.; Fujimoto, J.; Sowers, L. C. *Chem. Res. Toxicol.* **1998**, *11*, 75.

(35) Munzel, M.; Lischke, U.; Stathis, D.; Pfaffeneder, T.; Gnerlich, F. A.; Deiml, C. A.; Koch, S. C.; Karaghiosoff, K.; Carell, T. *Chemistry* **2011**, *17*, 13782.

(36) Maiti, A.; Noon, M. S.; Mackerell, A. D., Jr.; Pozharski, E.; Drohat, A. C. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 8091.

(37) Maiti, A.; Morgan, M. T.; Pozharski, E.; Drohat, A. C. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 8890.

(38) Zhang, L.; Lu, X.; Lu, J.; Liang, H.; Dai, Q.; Xu, G. L.; Luo, C.; Jiang, H.; He, C. *Nat. Chem. Biol.* **2012**, *8*, 328.

(39) Zhachkina, A.; Lee, J. K. *J. Am. Chem. Soc.* **2009**, *131*, 18376.

(40) Liu, M.; Li, T. T.; Amegayibor, F. S.; Cardoso, D. S.; Fu, Y. L.; Lee, J. K. *J. Org. Chem.* **2008**, *73*, 9283.

(41) Maiti, A.; Drohat, A. C. *DNA Repair* **2011**, *10*, 545.

(42) Gallinari, P.; Jiricny, J. *Nature* **1996**, *383*, 735.

(43) Steinacher, R.; Schar, P. *Curr. Biol.* **2005**, *15*, 616.

(44) Guan, X.; Madabushi, A.; Chang, D. Y.; Fitzgerald, M.; Shi, G.; Drohat, A. C.; Lu, A. L. *Nucleic Acids Res.* **2007**, *35*, 6207.

(45) Maiti, A.; Morgan, M. T.; Drohat, A. C. *J. Biol. Chem.* **2009**, *284*, 36680.

(46) Barrett, T. E.; Scharer, O. D.; Savva, R.; Brown, T.; Jiricny, J.; Verdine, G. L.; Pearl, L. H. *EMBO J.* **1999**, *18*, 6599.

(47) Parikh, S. S.; Walcher, G.; Jones, G. D.; Slupphaug, G.; Krokan, H. E.; Blackburn, G. M.; Tainer, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5083.

(48) Morgan, M. T.; Maiti, A.; Fitzgerald, M. E.; Drohat, A. C. *Nucleic Acids Res.* **2011**, *39*, 2319.

(49) Morgan, M. T.; Bennett, M. T.; Drohat, A. C. *J. Biol. Chem.* **2007**, *282*, 27578.

(50) Leatherbarrow, R. J. *GraFit 5.0*; Erithacus Software Ltd.; Staines, U.K., 1998.

(51) Fitzgerald, M. E.; Drohat, A. C. *J. Biol. Chem.* **2008**, *283*, 32680.

(52) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(53) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(54) Kohn, W.; Becke, A. D.; Parr, R. G. *J. Phys. Chem.* **1996**, *100*, 12974.

(55) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, K. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, 2004.

(56) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, K. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 09*; Gaussian, Inc.; Wallingford, CT, 2009.

(57) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995.

(58) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669.

(59) Takano, Y.; Houk, K. N. *J. Chem. Theory Comput.* **2005**, *1*, 70.

(60) Camaioni, D. M.; Schwerdtfeger, C. A. *J. Phys. Chem. A* **2005**, *109*, 10795.

(61) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2007**, *111*, 408.

(62) Kurinovich, M. A.; Lee, J. K. *J. Am. Chem. Soc.* **2000**, *122*, 6258.