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

pubs.acs.org/JPCB

Article

Oxidation Enhances Binding of Extrahelical 5-Methyl-Cytosines by Thymine DNA Glycosylase

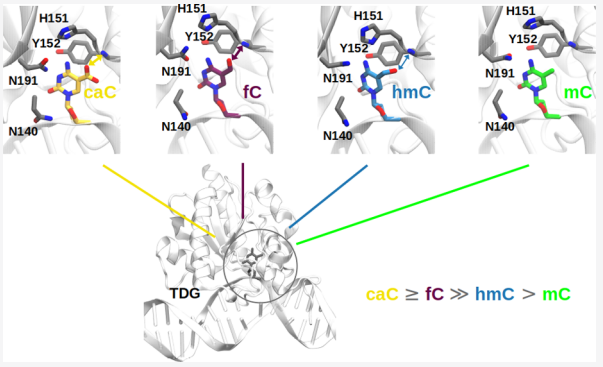
Frank Beierlein, Senta Volkenandt, and Petra Imhof*

Cite This: J. Phys. Chem. B 2022, 126, 1188−1201

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The DNA repair protein thymine DNA glycosylase (TDG) removes mispaired or damaged bases, such as oxidized methyl-cytosine, from DNA by cleavage of the glycosidic bond between the sugar and the target base flipped into the enzyme's active site. The enzyme is active against formyl-cytosine and carboxyl-cytosine, whereas the lower oxidized hydroxymethyl-cytosine and methyl-cytosine itself are not processed by the enzyme. Molecular dynamics simulations with thermodynamic integration of TDG complexed to DNA carrying one of four different (oxidized) methyl-cytosine bases in extrahelcial conformation, methyl-cytosine (mC), hydroxymethyl-cytosine (hmC), formyl-cytosine (fC), or carboxyl-cytosine (caC), show a more favorable binding affinity of the higher oxidized forms, fC and caC, than the nonsubstrate bases hmC and mC. Despite rather comparable, reaction-competent conformations of the flipped bases in the active site of the enzyme, more and stronger interactions with active site residues account for the preferred binding of the higher oxidized bases. Binding of the negatively charged caC and the neutral fC are strengthened by interactions with positively charged His151. Our calculated proton affinities find this protonation state of His151 the preferred one in the presence of caC and conceivable in the presence of fC as well as increasing the binding affinity toward the two bases. Discrimination of the substrate bases is further achieved by the backbone of Tyr152 that forms a strong hydrogen bond to the carboxyl and formyl oxygen atoms of caC and fC, respectively, a contact that is completely lacking in mC and much weaker in hmC. Overall, our computational results indicate that the enzyme discriminates the different oxidation forms of methyl-cytosine already at the formation of the extrahelical complexes.



INTRODUCTION

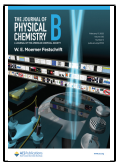
The base excision repair (BER) system is a machinery of enzymes recognizing, removing, and correcting mispairs and damage in the DNA. In the first step of the base excision repair system, glycosylases such as human thymine DNA glycosylase recognize a damaged or mispaired base and remove it via glycosidic C1′-N1 bond hydrolysis. The resulting abasic site is further processed by apurinic/apyrimidinic endonuclease that cleaves the DNA backbone to allow subsequent insertion of a new, correct nucleotide by polymerase β and ultimately sealing of the backbone by a ligase enzyme.

One possible DNA damage is the deamination of cytosine or 5-methyl-cytosine (mC) that results in a mismatch in the DNA with G:U or G:T mispairs instead of the Watson−Crick pair (m)C:G and finally in mutations in the encoded proteins. Such mispairs are recognized and removed by uracil DNA glycosylase (UDG) or human thymine DNA glycosylase (TDG), respectively, in the first step of base excision. Maintaining CpG sites (regions in the DNA where a cytosine is next to a guanine), in both, their nucleotide composition and their methylation status, is of particular importance since methylation of cytosine at CpG sites is a means of epigenetic regulation and alterations in DNA methylation are now known to be related to genetic diseases and cancer.[1,2]

In this context, thymine DNA glycosylase has been reported to play a role in a demethylation pathway, resulting in the removal of mC: Enzymes such as the ten-eleven translocation (TET) methyl-cytosine dioxygenases transform mC by stepwise oxidation into hydroxymethyl-cytosine (hmC), formyl-cytosine (fC), and carboxyl-cytosine (caC).[3−5] Whereas hmC and mC are not processed by the TDG enzyme, the higher-oxidized forms, fC and caC, are recognized and expelled by TDG and, following the base excision repair pathway, ultimately replaced by unmethylated cytosine (C)[3,6,7] (Figure 1).

Biochemical binding data show DNA with substrate bases to be bound more strongly to TDG than C, mC, and hmC,[6,8]

Received: November 18, 2021
Revised: January 15, 2022
Published: February 3, 2022



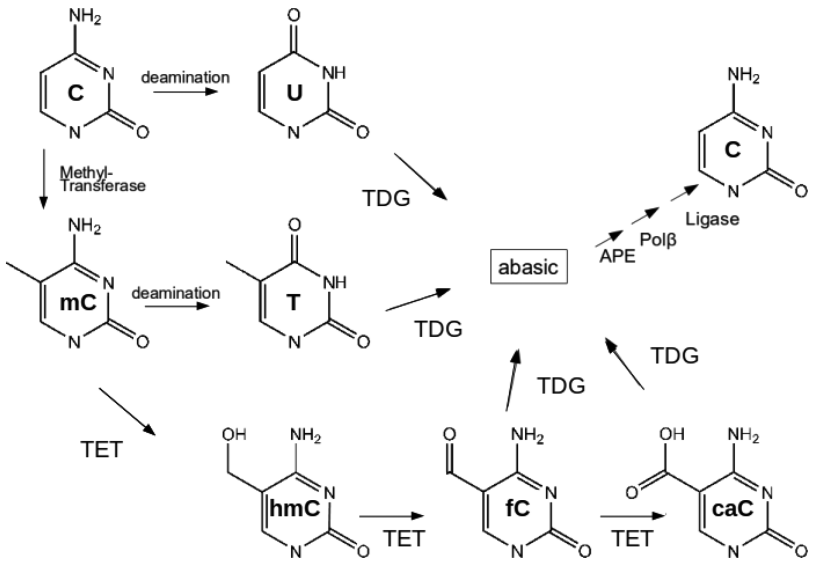


Figure 1. Cytosine can be methylated by methyl-transferase enzymes. Deamination of cytosine (C) or methyl-cytosine (mC) leads to uracil (U) or thymine (T), respectively, which are recognized and removed by thymine DNA glycosylase (TDG). Methylcytosine dioxygenases (TET) transform methyl-cytosine into its oxidized forms hydroxymethyl-cytosine (hmC), formyl-cytosine (fC), and carboxyl-cytosine (caC). TDG recognizes and excises fC and caC, but not mC or hmC. The abasic product of TDG is further processed by other enzymes of the base excision repair pathway.

which could be interpreted as recognition already taking place at the stage of initial complex formation, as suggested for discrimination of mispaired thymine against paired thymine or methyl-cytosine.[9,10] Our previous simulations of uncomplexed DNA and intrahelical TDG-DNA complexes with different XC bases (mC, hmC, fC, or caC) in their amino forms paired with guanine do not show significant differences in DNA conformation or interaction with the TDG protein.[11,12] In their imino forms, conformational differences exist between cognate and noncognate XC forms. However, computed relative binding affinities render intrahelical TDG-DNA complexes with imino forms of the XC unlikely.[12] Thus, discrimination of the XC bases likely takes place at a later stage.

According to the crystal structures of TDG glycosylases complexed to lesioned DNA,[7,13−16] the target base is in an extrahelical conformation, flipped into the enzyme's active site (see Figure 2). The base flip as a possible means of discrimination between substrate and nontarget bases has been studied by molecular dynamics simulations for mispaired vs paired thymine and for fC and caC.[17] In these simulations, the conformational transition of Arg275 that intercalates into the DNA at the void left by the flipped out base has been found to be an important part of the process.[17−19] For T:G mispairs, their local deformation, which is already present in free DNA and stabilized in the complex with the TDG enzyme,[10,18,20] facilitates and hence accelerates the transition to the fully extrahelical state. Extrusion of fC and caC, though, has been found to follow different dynamics than the base flip of thymine.[17,18] In particular, for flipping caC out of the DNA helix and into the active site, interactions of the carboxyl group with a guiding protein residue, Lys201, are important.[18]

Crystal structures show that also the detailed conformations of the substrate bases embedded in the active site of TDG differ between mispaired bases (uracil or thymine analogues) and damaged bases, e.g., oxidized methyl-cytosine bases XC (fC and caC).[7,13−16]

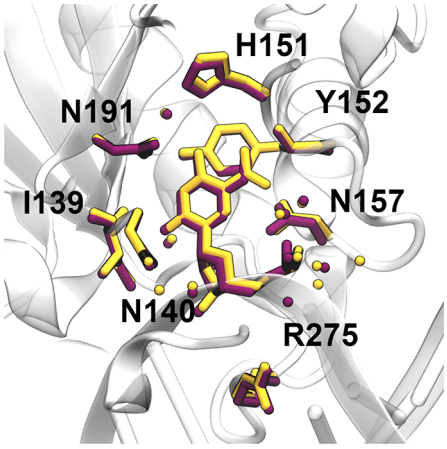


Figure 2. Superimposition of the crystal structures of TDG bound to DNA with $5caC^F$ (6U17,[16] orange) and $5fC^F$ (5T2W,[21] purple) with the extrahelical base flipped into the active site of TDG. Note that the structure of $5caC^F$ has been solved using the one of $5fC^F$ as a template.[16] The base flipped into the active site and active site residues Ile139, Asn140, Asn157, His151, Tyr152, Asn191, and Arg275 are shown in licorice.

A common feature of glycosidic bond cleavage reactions, though, is their stepwise character:[22−27] first, as the rate-determining step, the C1′-N1 bond dissociates, then a nucleophilic water molecule attacks the sugar C1′ atom, accompanied or followed by proton transfer to the departed base.

With the leaving group departure being the rate-determing step, the propensity of the bases to leave as anions, as quantified by their pK_a^{N1} (the microscopic ionization constant for the N1 site), may be decisive for the bond hydrolysis reaction.[22,28−30] In particular, the electron-withdrawing C5-substituents in fC and caC have been attributed to the increase of the leaving group ability of these bases and to account for TDG's activity against only these higher oxidized forms and not hmC or mC. For caC, only the form with a protonated

https://doi.org/10.1021/acs.jpcb.1c09896
J. Phys. Chem. B 2022, 126, 1188−1201

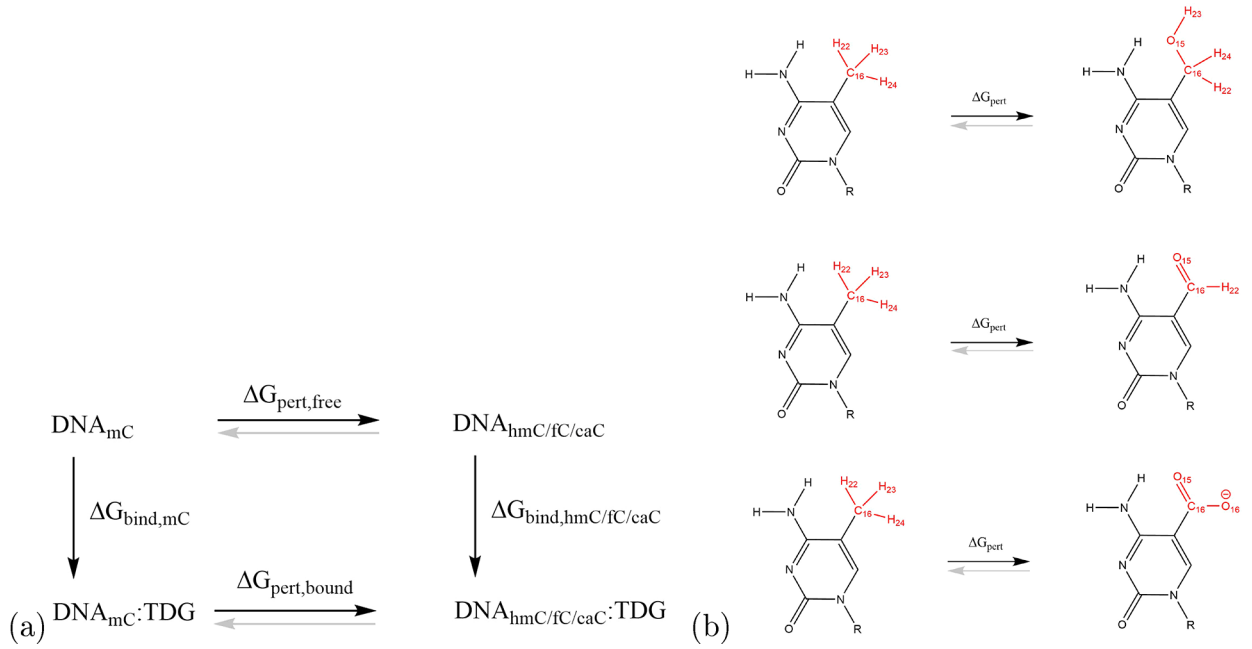


Figure 3. (a) Thermodynamic cycle used to determine relative binding free energies of TDG-DNA complexes with different forms of oxidized methyl-cytosine flipped out into the active site of TDG by perturbations from mC to hmC, fC, and caC (black arrows) and vice versa (gray arrows). (b) Definition of soft-core region: Atoms treated by a soft-core potential in the perturbations are shown in red.

carboxyl group has an increased leaving group ability,[6] rendering leaving group protonation catalytically important. Indeed, the cleavage rate for caC is enhanced at low pH[31] and the mechanism for caC excision has therefore been classified as acid-catalyzed. Protonation prior to leaving group departure has also been suggested for cleavage of mispaired thymine[23] but does not play a role in fC excision.[24]

Regarding nucleophilic attack, there is no obvious general base in the TDG enzyme that could activate a water molecule. The aspartate residue that is discussed to play such a role in uracil DNA glycosylase (though this role has also been attributed to a histidine residue[32]) is replaced by an asparagine, Asn140, in TDG. While Asn140 cannot act as a general base, there is a consensus that this residue plays a role in at least positioning the nucleophile, and computations[24] suggest strong hydrogen bonds of the nucleophilic water molecule to this residue to help its activation.

We use molecular dynamics simulations to obtain a microscopic view of the interactions between the XC bases and the active site of the TDG protein and to explain the measured differences in binding data. Whether and how the detailed binding mode of the extrahelical base in the active site of TDG contributes to recognition and thus discrimination of the substrate bases caC and fC over the nonsubstrate bases hmC and mC, in the prereactive complex, is the subject of the present paper.

METHODS

Model Setup. We modeled thymine DNA glycosylase (TDG) based on PDB[33] structure 6U17,[16] which is a 1:1 complex of the protein and DNA, with 5-caC^F flipped out of the DNA strand into the active site of the protein. For our simulations, we used the protein complexed to a shortened DNA fragment (sequence 5′-GCTCTGTAC**G**TGAGC-GATG-3′/3′-CGAGACATG[**XC**]ACTCGCTAC-5′), where [**XC**] marks the (oxidized) methyl cytosine. Additionally, we constructed B-DNA with the same sequence using NAB from the AmberTools suite[34] for simulations of uncomplexed DNA with intrahelical XC in solution (termed "free DNA" in the following). These simulations were also used as a reference in our previous paper.[12] In the protein structure, hydroxycysteine 276 (CSO 276) was replaced by cysteine. Protonation states and flips of histidine residues were checked by Whatif[35] and by visual inspection; histidine residues for which the protonation state deviates from the Amber default (HIE) were named accordingly. In particular, histidine residues 150 and 179 were treated as HID, i.e., with the proton located at the Nδ atom, and for His151 we performed simulations with this residue treated as both HIE151 (proton at the Nε atom) and as HIP151 (protons at the Nδ and Nε atoms). His151 with a proton at the Nδ atom only was not considered since the crystal structures of TDG bound to DNA with extrahelical 5-caC^F[16] or 5-fC[21] show a short distance between the Nε atom of His151 and the backbone O atom of Pro125 that strongly suggests a hydrogen bond between the two residues with the hydrogen atom located at the Nε atom of His151. According to our previous work on TDG complexed to DNA carrying a T:G mispair, this short distance is not maintained with Nδ-protonated His151.[23]

Methyl-cytosine (mC) and its oxidized derivatives 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC), and 5-carboxymethylcytosine (caC) were parametrized following a protocol previously established.[12,36,37] For the modified bases (and the C1′/H1′ atoms; the charges of all other nucleotide atoms were kept identical to their parmbsc1[38] values), we used RESP[39,40] charges based on quantum chemical calculations (HF/6-31G(d)//B3LYP/6-31G(d);[41−45] optimizations were performed in polarizable continuum model (PCM) water)[46,47] with Gaussian 16[48] in agreement with the Amber force fields.[40,49] Missing force field parameters of the bases were amended using values from Gaff (version 1.81)[50,51] or parmbsc1/ff14SB.[38,52] The DNA part of the system was

https://doi.org/10.1021/acs.jpcb.1c09896
J. Phys. Chem. B 2022, 126, 1188−1201

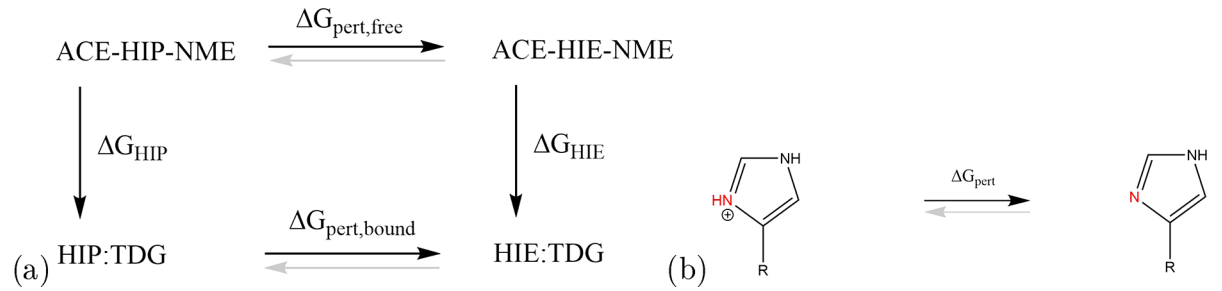


Figure 4. (a) Thermodynamic cycle used to compute protonation affinities of His151 in TDG with different forms of oxidized 5-methyl-cytosine flipped out into the active site of TDG (mC, hmC, fC, and caC) by perturbations from protonated to unprotonated His151. Simulations were performed both in solution (above) and in the protein−DNA complex (HIP → HIE, black arrow) and vice versa (HIE → HIP, gray arrow). (b) Definition of soft-core region: Atoms treated by a soft-core potential in the perturbations are shown in red.

described by the parmbsc1[38] force field and the protein by ff14SB.[52]

Prior to solvation, the systems were initially geometry optimized using implicit solvent (GB/SA) for 100 steepest descent and 400 conjugate gradient minimization steps. Then, the systems were solvated with TIP3P[53] water, using truncated octahedral box geometries exceeding the solutes by 15 Å in either direction, resulting in approximate box dimensions of 99 Å (protein−DNA complex) and 92 Å (free DNA). Then sodium counterions[54] were added to neutralize the system, and additional NaCl was added until a sodium concentration of 150 mM was obtained. Periodic boundary conditions were used throughout and the distance cutoff for all nonbonding interactions was set to 10 Å. Long-range electrostatics were described by the particle-mesh Ewald method.[55,56] For van der Waals interactions beyond those included in the direct sum, a continuum model correction for energy and pressure was used, as implemented in Amber.

Molecular Dynamics Simulations. After initial geometry optimization (first 5000 steps with restraints (50 kcal·mol^{-1} Å^{-2}) on the DNA and protein, then 5000 optimization steps without restraints, switch from steepest descent to conjugate gradients after 500 steps in either case), the solvated systems were heated to 298 K during a 500 ps simulation with weak restraints (10 kcal·mol^{-1} Å^{-2}) on the protein and the DNA in the NVT ensemble. After that, 600 ns unrestrained NPT Langevin dynamics were performed for each system at 298 K and 1 bar (weak pressure coupling, isotropic position scaling, pressure relaxation time 2 ps, collision frequency of 2 ps^{-1}) of which the first 100 ns were used for NPT equilibration of the system. SHAKE constraints were applied to bonds involving hydrogen atoms.[57] Watson−Crick distance restraints were imposed on the DNA termini (20 kcal · mol^{-1} Å^{-2}, allowing ±0.1 Å movement from the equilibrium bond distance) to prevent fraying of the DNA termini.[58]

Three independent simulation runs were performed for each system. Simulation frames were saved every 100 ps. All simulations were performed with Amber 18[34] and Amber 20[59] using pmemd.cuda on GPUs.

Thermodynamic Integration Simulations. Relative binding free energies of the complexes of the DNA oligonucleotides containing the four (oxidized) variants of methyl-cytosine with TDG were obtained using the thermodynamic cycle depicted in Figure 3a. The perturbations were performed with Amber 20[59] pmemd.cuda following a dual-topology thermodynamic integration (TI) approach.[60−65] Methyl-cytosine (mC) was perturbed into its oxidized forms caC, fC, and hmC, respectively, using a lambda coordinate of 21 windows (0.00, 0.05, ..., 0.95, 1.00), in both the bound state (DNA complexed with TDG) and the free state (DNA with intrahelical XC solvated in water, 150 mM NaCl, as above).[54] Perturbations were performed in one step, using a van-der-Waals (vdW) and electrostatic soft-core potential with Amber 20 default soft-core parameters[60] and the soft-core regions are indicated in Figure 3b. For comparison, we also used a 3-step perturbation protocol in which the vanishing atoms are first discharged, then the actual perturbation takes place using a vdW soft-core, and finally the appearing atoms are charged. As the results obtained with this alternative protocol were practically identical to the ones obtained with the one-step protocol, we decided to proceed with the one-step protocol.

The perturbation free energies $\Delta G_{\text{pert,bound}}$ and $\Delta G_{\text{pert,free}}$ (scheme in Figure 3a) were obtained from the free energy gradients by trapezoidal numerical integration. Starting structures for the TI simulations were taken from the unbiased MD simulations of caC, fC, hmC, and mC after 10.5 ns equilibration for both protonation states of His151, HIE and HIP, respectively.

Again, Watson−Crick (NMR) distance restraints on DNA termini (see above) were employed to prevent fraying of the DNA termini. After initial geometry optimization (5000 steps), each lambda window was heated to 298 K during 200 ps NVT simulation with weak Cartesian restraints (5 kcal·mol^{-1}·Å^{-2}) on non-hydrogen DNA/protein atoms, followed by 200 ps NPT equilibration without restraints. An integration time step of 1 fs was used, with SHAKE[57] constraints on all bonds involving hydrogen except the perturbed residues (in addition to SHAKE being removed between bonds containing one common and one unique atom). A Monte Carlo barostat was used for pressure (1 bar) control. Energy output was written every ps, while geometries were written every 10 ps. All other simulation parameters were chosen as suggested in the Amber TI tutorial.[63,64] Each lambda window was simulated for 30.4 ns, of which the last 20 ns were used for integration and for each perturbation three individual runs were performed. To check for possible hysteresis effects, we also performed each perturbation in the opposite direction, that is starting from an equilibrated structure corresponding to the system at the $\lambda = 1$ window of the perturbation (indicated by gray arrows in Figure 3).

In order to investigate different protonation states of His151 (HIP or HIE) when bound to TDG with different forms of oxidized methyl-cytosine (mC, hmC, fC and caC) flipped out into the active site of TDG, we employed the thermodynamic cycle shown in Figure 4[66,67] where HIP was perturbed to HIE, both in solution (ACE- and NME-protected) and in the

https://doi.org/10.1021/acs.jpcb.1c09896
J. Phys. Chem. B 2022, 126, 1188−1201

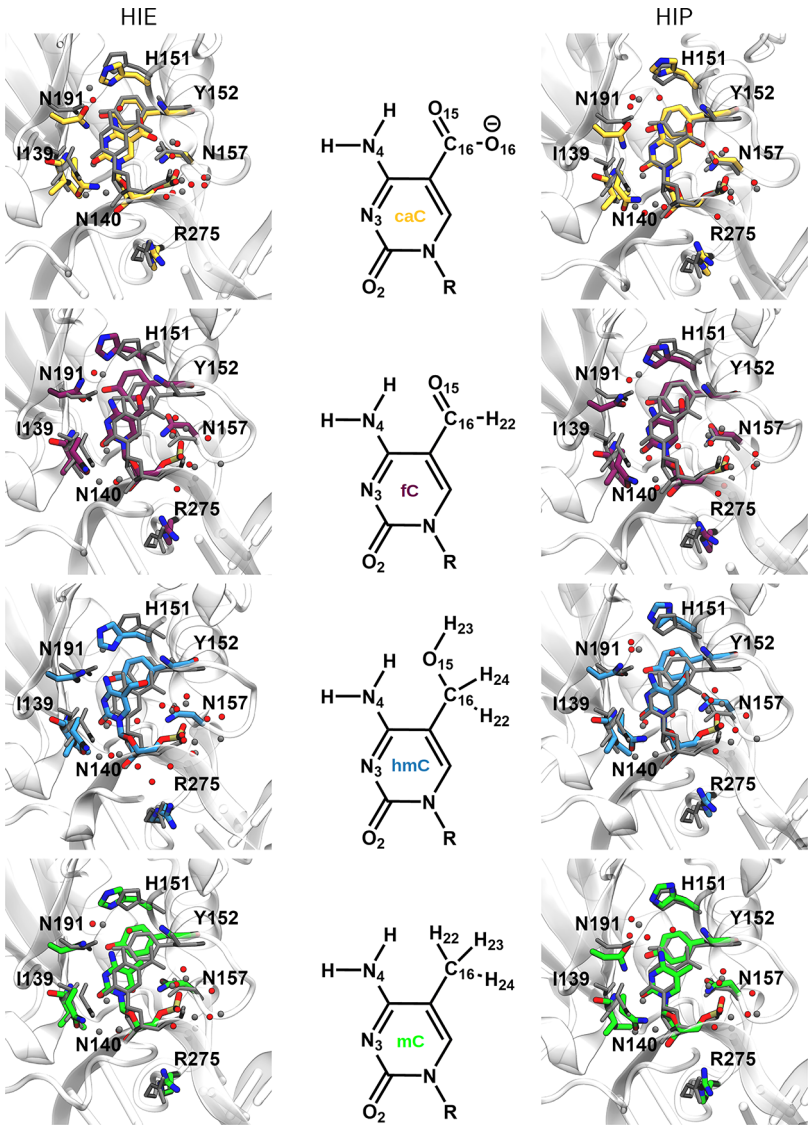


Figure 5. Active site of the TDG protein with extrahelical XC base bound shown as superimposition of median structures of the simulations of caC (orange), fC (purple), hmC (blue), and mC (green) with neutral His151 (HIE, left) or protonated His151 (HIP, right) with the crystal structure of 5caC^{F} (6U17,[16] gray). Residues Ile139, Asn140, Asn157, His151, Tyr152, Asn191, and Arg275 are shown in stick representation. Hydrogen atoms are omitted for clarity. Only water molecules within 4 Å of the XC nucleotides are shown.

protein−DNA complexes. Again, we performed the perturbations from HIP to HIE and also from HIE to HIP, to check for possible hysteresis. Thermodynamic integration simulation details were as described above, the choice of the soft-core region is depicted in Figure 4. The reported double free energy differences, $\Delta\Delta G$, describe the difference of the deprotonation free energy of HIP in the protein case and in solution

$$\Delta\Delta G(\mathrm{HIP} \rightarrow \mathrm{HIE}) = \Delta G_{\mathrm{HIE}} - \Delta G_{\mathrm{HIP}}$$
$$= \Delta G_{\mathrm{pert,bound}} - \Delta G_{\mathrm{pert,free}}$$

For the opposite direction of the perturbation, the double free energy difference is

$$\Delta\Delta G(\mathrm{HIE} \rightarrow \mathrm{HIP}) = \Delta G_{\mathrm{HIP}} - \Delta G_{\mathrm{HIE}}$$

Analysis. Only the last 500 ns was used for analysis. Cpptraj[68] from the AmberTools suite, VMD 1.9.3,[69] and Curves+/Canal[70,71] were used for the analyses. A flip angle, describing how much the XC base is rotated out of the DNA helix, is defined as the pseudo-dihedral formed by the XC base, the sugar of the XC nucleotide, the sugar of the next nucleotide downstream, and the next base and its complementary base, a definition we have used previously.[9,10] A hydrogen bond was defined based on geometric criteria, i.e., a donor−acceptor distance not larger than 3.2 Å and a donor-hydrogen-acceptor angle deviating from linearity by not more than 42°. Residue interaction energies were calculated using the LIE command of cpptraj[68] with default settings.

Further data analysis was performed by custom-made Jupyter Notebooks using NumPy[72] with Python.[73] All error estimates are the standard deviation from the mean of the three individual runs per simulation setup. Median structures for visualization were determined from the simulation trajectories by custom-made scripts and MDtraj.[74] Matplotlib[75] was used for plotting and molecular figures were prepared with VMD 1.9.3.[69]

https://doi.org/10.1021/acs.jpcb.1c09896
J. Phys. Chem. B 2022, 126, 1188−1201

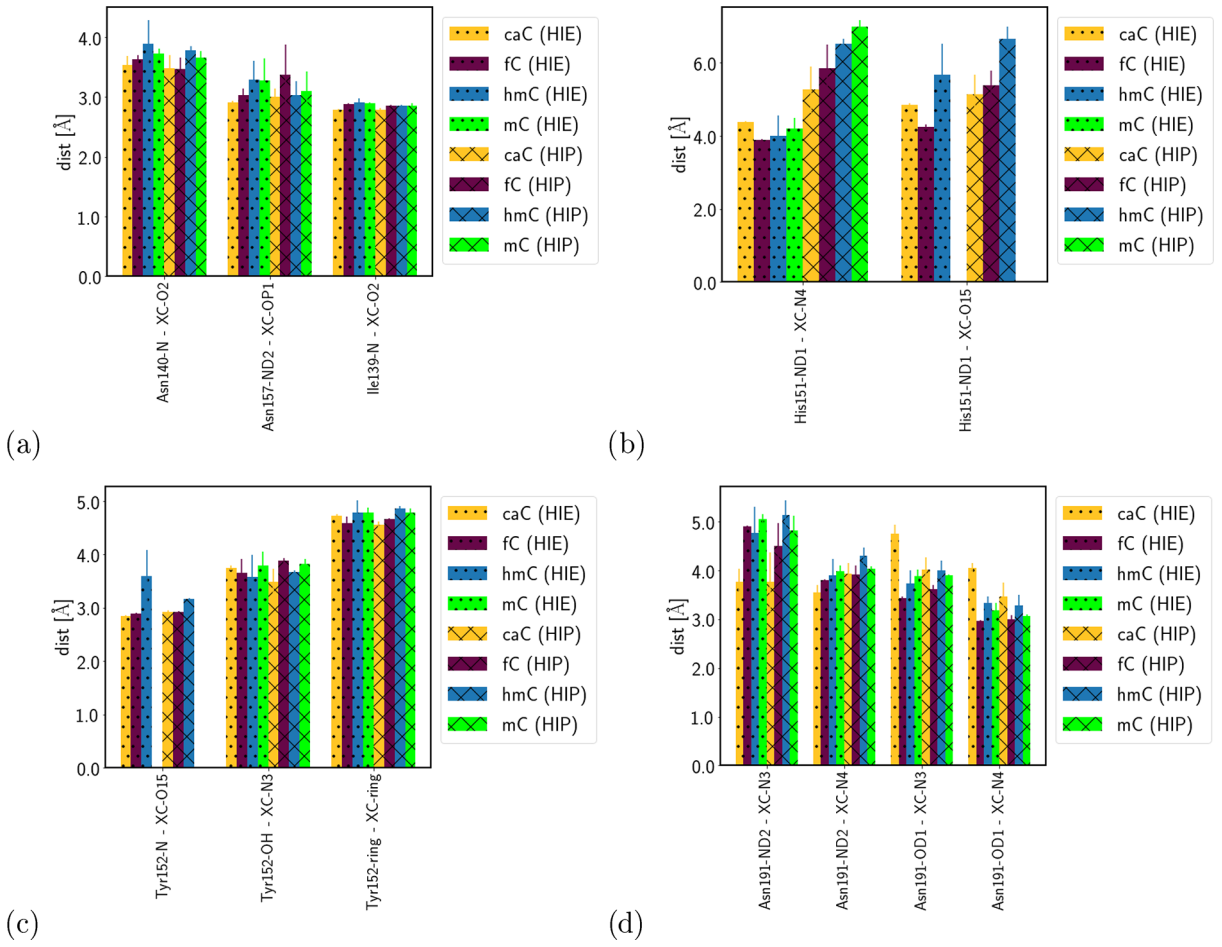


Figure 6. Average distances between the XC base and active site residues: (a) Asn140, Asn157, Ile139, (b) His151, (c) Tyr152, and (d) Asn191. For the distance distributions see Figures S4–S8.

RESULTS

TDG-DNA Complex Conformations. All TDG-DNA complexes, caC, fC, hmC, and mC, exhibit an extrahelical conformation of the XC base that is flipped into the active site of TDG (see Figure 5).

Accordingly, the base flip angles and the opening angle show values that are characteristic for an extrahelical conformation and are similar in all models (see Figure S1). Moreover, none of the other DNA base pair or helical parameters indicates significant differences in the DNA conformation between the different models (see Figure S2).

The XC base is "on top" of residue Tyr152 and is surrounded by residues Ile139, Asn140, Asn157, and Asn191, as is the case also in the crystal structures of TDG bound to caC and to fC (cf. Figure 5). Arg275 is surrounded by the phosphate groups of the neighboring nucleotides, thus filling the void left by the flipped-out base, as also observed in the crystal structures (cf. Figure 5). Some of the individual distances between the XC bases and active site residues, however, vary between the different models (see Figure 6). Whereas distances to Ile139, Asn140, and Asn157 are rather similar (Figure 6a), significant differences can be observed for the distance between His151 or Tyr152 and the oxygen atom of the different oxidized methyl groups, O15 (see Figure 6b and c), and in the models with protonated His151 also in the distances of this residue to the amino group of the XC bases (see Figure 6b). Asn191 is closer to the N3 atom of the base with its ND2 atom in the caC models, whereas the OD1 atom is somewhat farther from the XC base in these models, indicating a preference for a switched conformation of Asn191 when caC is bound to the active site of TDG.

TDG-DNA Interactions. *Binding Free Energies and Proton Affinities.* With neutral His151 (HIE), TDG binds most preferably to DNA carrying a fC lesion whereas the binding free energies of hmC and mC are comparable (see Table 1). Binding of DNA with a caC lesion, on the other hand, is much less favorable than binding an intact mC when His151 is neutral. The difference between the binding free energy differences calculated from perturbing mC to caC (HIE) and vice versa, as well as from perturbations between mC and hmC (HIP) are somewhat larger than the error of

Table 1. Relative Binding Free Energies (in kcal/mol) of the DNA with Oxidized Methyl-Cytosine (hmC, fC, or caC, Respectively) to TDG Bound to TDG with Neutral His151 (HIE) or Protonated His151 (HIP)

	HIE	HIP
mC → caC	6.35 ± 0.62	−6.05 ± 0.80
caC → mC	−2.40 ± 0.41	7.52 ± 0.39
mC → fC	−2.92 ± 0.09	−6.11 ± 0.05
fC → mC	3.70 ± 0.23	6.43 ± 0.46
mC → hmC	−0.24 ± 0.26	−0.98 ± 0.36
hmC → mC	0.23 ± 0.10	3.62 ± 0.18

https://doi.org/10.1021/acs.jpcb.1c09896
J. Phys. Chem. B 2022, 126, 1188−1201

Table 2. Protonation Affinities (in kcal/mol) as Calculated from Relative Free Energies (see Methods) of His151 in TDG-DNA Complexes with Different Forms of Oxidized Methyl-Cytosine[a]

	caC	fC	hmC	mC
HIP → HIE	6.98 ± 0.73	3.13 ± 0.52	3.21 ± 1.19	1.91 ± 0.78
HIE → HIP	−7.50 ± 1.07	1.37 ± 0.31	−4.75 ± 1.00	3.26 ± 0.84

[a]HIE represents neutral and HIP protonated His151, respectively.

Table 3. Electrostatic Interaction Energies (in kcal/mol) of TDG and the XC Base (for Arg275 with the Backbone of XC and the Backbone of Its Neighboring Residues) in Complexes with Different Forms of Oxidized Methyl-Cytosine (for van der Waals Interaction energies, see Table S1)[a]

HIE	caC	fC	hmC	mC
Arg275	−139.86 ± 4.92	−140.18 ± 4.21	−135.22 ± 7.38	−139.15 ± 0.62
AS	−45.66 ± 0.97	−18.41 ± 1.58	−17.91 ± 5.55	−13.19 ± 3.14
Ile139	−15.14 ± 0.14	−9.88 ± 0.17	−10.19 ± 0.59	−11.12 ± 0.14
Asn140	4.50 ± 1.66	5.52 ± 0.23	6.93 ± 0.35	6.02 ± 1.96
His151	−6.21 ± 0.16	−3.96 ± 0.07	−4.67 ± 2.47	−2.65 ± 0.55
Tyr152	−18.23 ± 0.63	−7.57 ± 2.41	−8.22 ± 6.36	−3.35 ± 1.98
Asn191	−10.58 ± 1.86	−2.52 ± 0.62	−1.76 ± 4.64	−2.08 ± 1.18
HIP	caC	fC	hmC	mC
Arg275	−141.87 ± 6.47	−132.99 ± 10.57	−140.80 ± 5.49	−138.81 ± 2.62
AS	−75.15 ± 4.54	−24.65 ± 3.24	−17.30 ± 2.09	−10.74 ± 2.76
Ile139	−14.52 ± 0.36	−9.87 ± 0.21	−10.69 ± 0.15	−11.15 ± 0.18
Asn140	7.73 ± 2.30	5.50 ± 0.75	5.16 ± 1.38	5.95 ± 0.45
His151	−35.53 ± 3.30	−7.72 ± 0.16	−2.47 ± 0.20	3.31 ± 0.27
Tyr152	−24.26 ± 3.63	−5.85 ± 0.87	−6.38 ± 0.57	−3.12 ± 0.46
Asn191	−8.58 ± 5.09	−6.70 ± 2.95	−2.92 ± 2.06	−5.74 ± 2.78

[a]HIE and HIP refer to TDG with neutral and protonated His151, respectively. AS means active site residues Ile139, Asn140, His151, Tyr152, Asn191 together.

these TI simulations. This can be explained by different interactions with water molecules or Tyr152, depending on the starting point of the perturbation simulations (see the Supporting Information).

Looking also at the protonation affinities, in the complex with caC, neutral His151 is computed to be so highly unfavorable (Table 2), compared to protonated His151, that neutral His151 can be ruled out as very unlikely when caC is bound.

As for the protonation state of His151 in the other models, the calculated protonation affinities (Table 2) are not informative for models fC and mC, due to the large hysteresis: Performing alchemical perturbation calculations from protonated to unprotonated His151 or vice versa, both result in unfavorable energies. This may be indicative of the XC base being differently accommodated in the active site in the respective starting conformation of the simulations, that is with protonated or unprotonated His151, respectively, and the limitations of the perturbation simulations do not allow a full relaxation to the optimal conformation in the other protonation state. Comparison of the starting and end points of the perturbation simulations ($\lambda = 0$ and $\lambda = 1$) indeed reveal differences in structure and some interactions of the fC and mC base, respectively, with Tyr152 and water molecules, depending on the starting point of the perturbation simulations (see Supporting Information). Such a behavior can furthermore be interpreted as both protonation states of His151, HIP and HIE, being conceivable in the fC and in the mC model, depending on the conformation.

When His151 is protonated (HIP), the calculated relative binding affinities (see Table 1) suggest a strong preference for binding of caC or fC over binding of mC, whereas the relative binding affinity of hmC does not suggest DNA with this base to be favored over mC. His151 is not only clearly protonated in the complex with caC bound, but this is also the preferred form when hmC is bound to TDG (see Table 2).

Comparing the interaction energies of the active site with the XC nucleotide, these are most favorable with caC. This can be due to the different number of atoms contributing to the interactions and also due to the different (negative vs neutral) charge of the caC base. For the caC system, the interactions are significantly weaker when complexed to TDG with neutral His151 (∼−46 kcal/mol) than with protonated His151 (∼−75 kcal/mol, see Table 3). The interaction energy with His151 itself (∼−6 vs ∼−36 kcal/mol for neutral and protonated His151, respectively, see Table 3), corresponds to this difference. The attractive interactions between the two oppositely charged residues, protonated His151 and the negatively charged caC base, are likely the major contribution to the more favorable binding affinity of caC over the non-cognate forms, hmC and mC, the latter of which shows an even repulsive interaction with protonated His151 (i.e. positive interaction energy, see Table 3). It is interesting to note that the interaction energy of the intercalating residue Arg275 with the backbone of the XC nucleotide and the backbone of its neighboring residues is comparable in all models and in both protonation states of His151, HIE and HIP. Also the combined interactions of the active site residues with fC suggest a preference over the hmC and mC models when His151 is protonated, whereas this is not the case with neutral His151, further indicating that this protonation state cannot be ruled out for fC, though being of a lesser importance than in the case of caC.

https://doi.org/10.1021/acs.jpcb.1c09896
J. Phys. Chem. B 2022, 126, 1188−1201

Table 4. Probabilities for Hydrogen Bonds between TDG and the XC Nucleotide[a]

	HIE			
acceptor - donor	caC	fC	hmC	mC
XC:O15 - TYR152:N	0.97 ± 0.00	0.98 ± 0.01	0.49 ± 0.39	
XC:O2 - ILE139:N	0.84 ± 0.13	0.92 ± 0.04	0.80 ± 0.16	0.93 ± 0.02
XC:OP1 - ASN157:ND2	0.94 ± 0.03	0.87 ± 0.05	0.56 ± 0.47	0.71 ± 0.23
ASN191:OD1 - XC:N4		0.66 ± 0.05		0.53 ± 0.12
XC:N3 - ASN191:ND2				
	HIP			
acceptor - donor	caC	fC	hmC	mC
XC:O15 - TYR152:N	0.91 ± 0.10	0.95 ± 0.03	0.49 ± 0.11	
XC:O2 - ILE139:N	0.74 ± 0.23	0.78 ± 0.20	0.91 ± 0.08	0.85 ± 0.10
XC:OP1 - ASN157:ND2	0.57 ± 0.50	0.72 ± 0.25	0.62 ± 0.44	0.78 ± 0.27
ASN191:OD1 - XC:N4	0.31 ± 0.26	0.74 ± 0.13	0.31 ± 0.21	0.40 ± 0.27
XC:N3 - ASN191:ND2	0.51 ± 0.36	0.21 ± 0.23		0.20 ± 0.21

[a]Only those hydrogen bonds that have a probability of at least 0.5 in at least one of the models are listed. HIE and HIP refer to TDG with neutral and protonated His151, respectively. For atom labels of the XC bases, see Figures 3 and 5.

Table 5. Average Number of Hydrogen Bonds between the XC Nucleotide and Water (W)[a]

	HIE			
acceptor - donor	caC	fC	hmC	mC
XC:O15 - W	0.58 ± 0.20			
XC:O16 - W	1.81 ± 0.10			
XC:O2 - W	0.61 ± 0.53		0.84 ± 0.12	
XC:OP1 - W	1.72 ± 0.23	1.98 ± 0.17	1.53 ± 0.41	1.93 ± 0.12
XC:OP2 - W	2.55 ± 0.17	2.29 ± 0.06	2.35 ± 0.08	2.37 ± 0.11
W - XC:O15			0.88 ± 0.07	
	HIP			
acceptor - donor	caC	fC	hmC	mC
XC:O15 - W	0.54 ± 0.19			
XC:O16 - W	1.91 ± 0.07			
XC:O2 - W	0.66 ± 0.40	0.81 ± 0.11		0.92 ± 0.02
XC:OP1 - W	1.52 ± 0.18	1.88 ± 0.04	2.00 ± 0.36	1.86 ± 0.23
XC:OP2 - W	2.49 ± 0.12	2.41 ± 0.08	2.36 ± 0.15	2.51 ± 0.04
W - XC:O15			0.89 ± 0.04	

[a]Only those hydrogen bonds that have a probability of at least 0.5 in at least one of the models are listed. HIE and HIP refer to TDG with neutral and protonated His151, respectively. For atom labels of the XC bases see Figures 3 and 5.

Another large difference in interaction energies between caC and the other models can be observed for the electrostatic interactions with the active site residue Tyr152. caC interacts most favorably with this residue and this interaction is, again, stronger in the complex with protonated His151 whereas the interaction between XC base and Tyr152 is not affected by the protonation state of His151 in the other models (see below for a detailed analysis of the hydrogen-bonded interactions). The other two oxidized forms, fC and hmC, exhibit interaction energies with Tyr152 that are similar, but weaker than in the caC model, and the interaction of mC with Tyr152 is even weaker.

Hydrogen-Bonded Interactions. There are several residues that form direct hydrogen bonds with the XC nucleotide (see Table 4). Ile139 has a high probability for a hydrogen bond with the O2 atom of the XC base in all models. And also hydrogen bonds between Asn157 and the phosphate backbone of XC are observed, though with large errors in some models. Average distances between the XC base and either of the two residues, though, are, within error, comparable in all models (see Figure 6a).

Large differences can be observed for the hydrogen-bonded interactions between Asn191 and the XC base. In the fC models, with protonated and unprotonated His151, and in the mC model, with neutral His151, the OD1 atom of Asn191 accepts a hydrogen bond from the N4 atom of the XC base with a higher probability than that in the other models. Asn191 as a hydrogen bond donor to the N3 atoms of the bases, in contrast, is observed with significant probability only in the caC model with protonated His151 (see Table 4). Correspondingly, the caC model exhibits a shorter average distance from Asn191 to the N3 atom of the base than in the other models (see Figure 6d).

A hydrogen-bond interaction with the oxygen atom of the oxidized methyl group (O15) and the backbone NH atom of Tyr152 is present almost throughout the simulations of the caC and fC models, but it has only about half this probability in the hmC model which can be explained by the hydroxyl group acting partially as the hydrogen bond acceptor, with Tyr152 or water (see Table 5), and partially as the hydrogen bond donor, to water (Table 5). The average distance between Tyr152 and the XC base is also significantly shorter in the caC and fC models than in the hmC model (see Figure 6c). This

https://doi.org/10.1021/acs.jpcb.1c09896
J. Phys. Chem. B 2022, 126, 1188−1201

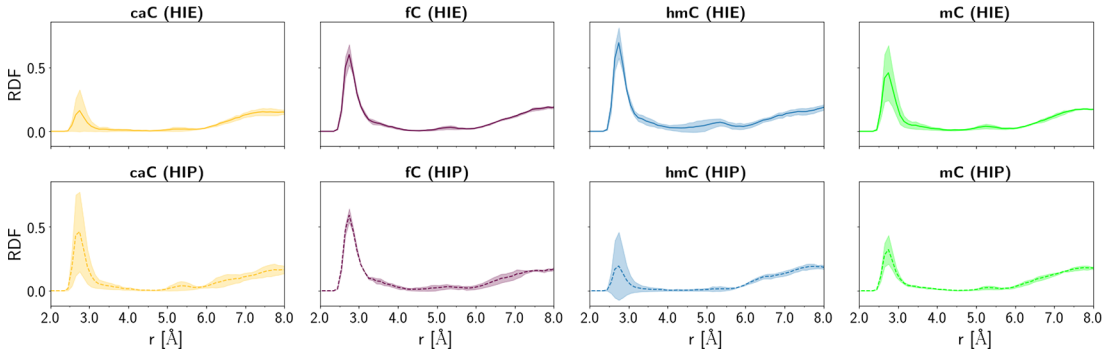


Figure 7. Radial distribution functions (RDF) of oxygen atoms of water molecules around the OD1 atom of Asn140. HIE and HIP refer to TDG with neutral and protonated His151, respectively.

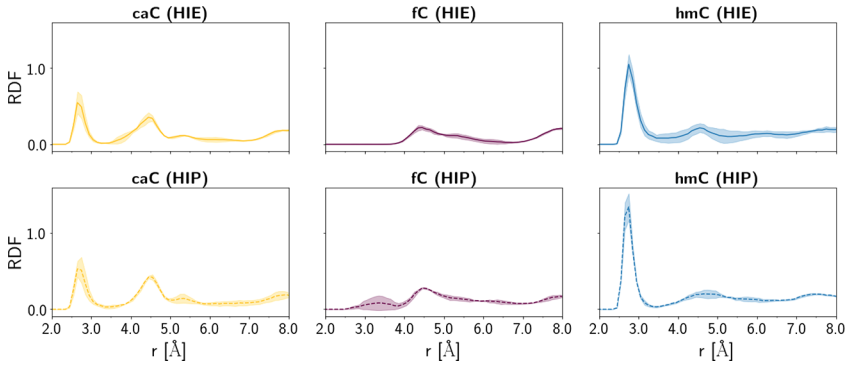


Figure 8. Radial distribution functions (RDF) of oxygen atoms of water molecules around the O15-atom of the oxidized methyl group of the XC base. HIE and HIP refer to TDG with neutral and protonated His151, respectively.

hydrogen bond between the oxidized forms, caC, fC, and hmC and Tyr152 likely accounts for their more favorable interaction energies with this residue, compared to mC which lacks oxygen atom O15 and can therefore not form such a hydrogen bond.

It is interesting to note that the average distances between His151 and the O15 atom of the XC base follow this trend (see Figure 6b), although no direct hydrogen bond is observed between His151 and the XC base. A similar trend is observed for the distance between His151 and the N4 atom of the base. These distances are about the same in the models with unprotonated His151 and shorter than in the protonated case, indicating some repulsion between the His151-ND-H and the amino group of the XC base.

Hydrogen bonds between the XC nucleotide and water are, in all models, predominantly formed with the phosphate backbone, resulting in at least two hydrogen bonds between water molecules, one to each of the nonester oxygen atoms (OP1 and OP2), respectively (see Table 5). Almost two hydrogen bonds donated by water molecules are accepted by the O16 atom of the carboxyl group in the caC models. The other carboxyl oxygen atom of caC, O15, exhibits a much lower probability for hydrogen bonds with water molecules, as can be explained by this atom already being engaged in a hydrogen bond to Tyr152. For the hmC model, for which the O15 atom shows a lower probability for hydrogen bonds with Tyr152, hydrogen bonds to water can be observed with O15 as acceptor. In the fC models, in contrast, the O15 atom forms hydrogen bonds exclusively with Tyr152 and not with water.

There is also some probability for a hydrogen bond between water and the O2 atom of the XC base, but no clear trend with the oxidation levels of the models can be observed. This is also reflected in the rather large errors of the radial distribution functions of water around the O2 atom (see Figure S9), suggesting no stable water position at this site which is partially precluded by the hydrogen bond between Ile139 and the O2 atom (see Table 4).

The hydrogen-bonded interactions between the intercalating residue Arg275 and the nucleotides A28 and G30, i.e. the neighbors of the lesion (XC29), are comparable in all models and agree well with the similar distances between Arg 275 and these residues (see Figure S6) and the similar interaction energies calculated for this residue with the DNA (see Table 3). These interactions indicate a stable conformation with the extrahelical base that is further confirmed by the similar narrow distribution of the opening and flip angles in all models (see Figure S1).

Hydrogen-bonded interactions between Lys232 and residues T31 and A32 further downstream, which have been discussed earlier[10] as an important TDG−-DNA interaction that may contribute to the complex binding energy, are also present in all models with high probability (see Figure S3). The only remarkable differences in hydrogen-bonded interactions are those formed between TDG and the DNA strand complementary to that carrying the XC nucleotide are by Cys233 with T31 and between Gln278 and residues G30, next to XC, or T31. Differences in the protein interactions with the XC-containing strand can only be observed for DNA residues that are far away from the lesion site.

Water in the Active Site. *Putative Nucleophilic Water Molecules.* Asn140 has been discussed to be responsible for placing a water molecule close to the XC base that can act as a nucleophile.[13,14,23,24,30] The radial distribution functions around the OD1 atom of Asn140 (Figure 7) indeed show peaks at typical hydrogen-bond distances (~2.8 Å), albeit with

https://doi.org/10.1021/acs.jpcb.1c09896
J. Phys. Chem. B 2022, 126, 1188−1201

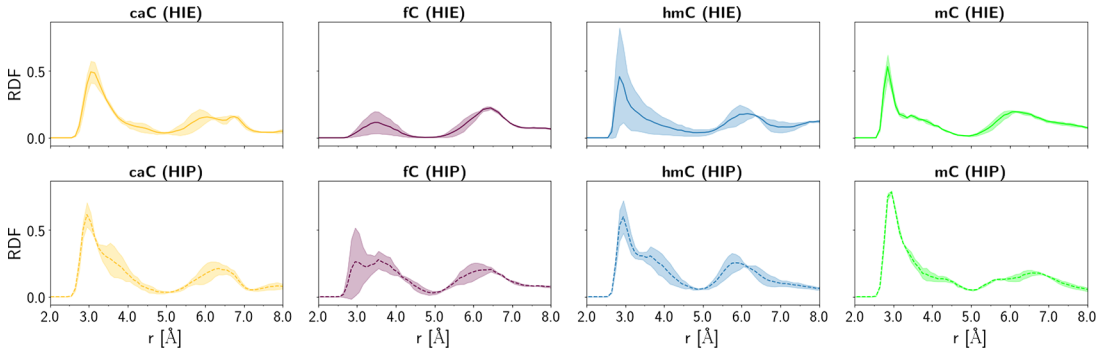


Figure 9. Radial distribution functions (RDF) of oxygen atoms of water molecules around the N4-atom of the XC base. HIE and HIP refer to TDG with neutral and protonated His151, respectively.

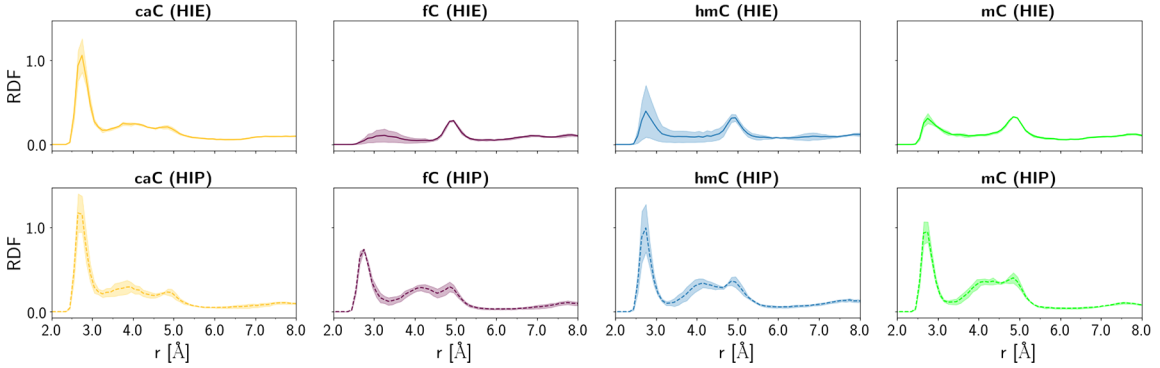


Figure 10. Radial distribution functions (RDF) of oxygen atoms of water molecules around the OD1 atom of Asn191. HIE and HIP refer to TDG with neutral and protonated His151, respectively.

varying intensity. Both caC models and the hmC model with protonated His151 exhibit large errors for first hydration peaks around the OD1 atom of Asn140, indicating a less ordered position of putative nucleophilic water molecules. The fC model, and the hmC model with neutral His151 exhibit the sharpest distribution with least errors.

The radial distribution functions of water around the O2 atom of the XC bases all show a peak at hydrogen bonding distance to this atom, indicating that the presence of another water molecule on the side of the XC sugar opposite to Asn140 is probable (see Figure S9).

Other Water Molecules. An analysis of the radial distribution functions of water molecules around the oxygen atom of the oxidized methyl group, O15, shows that only caC and hmC exhibit peaks below 3 Å (see Figure 8). Their intensities agree well with hydrogen bonds of atom O15 with water molecules that are only observed for these models (cf. Table 5).

The radial distribution functions of water molecules around the nitrogen atom N4 show a peak at about a hydrogen bonded distance (~3 Å) in all models, albeit with varying intensity and errors (see Figure 9). There is, however, no significant probability for a hydrogen bond between the amino group of the base and water molecules. Instead, the N4 atom of the mC and in particular fC bases forms hydrogen bonds with Asn191 (cf. Table 4).

Still, water molecules are also observed around Asn191 (Figure 10). Water around Asn191 is clearly within the hydrogen-bonding distance with OD1 in both caC models, i.e., with neutral and protonated His151, and in the other models with protonated His151, corresponding to the observed high probabilities for hydrogen bonds with water (see Table S2).

DISCUSSION

Our simulation data show that all XC bases are well embedded in the active site of TDG and remain in an extrahelical conformation throughout the simulations. The active site conformations, also those of the noncognate complexes, do not differ substantially from the crystal structures of TDG bound to fC or caC. According to our calculated relative binding affinities, DNA with a caC lesion is most preferably bound, but also fC is favored over the noncognate forms, hmC and mC. This finding is in agreement with the experimentally observed weaker binding of mC and hmC by TDG compared to binding of DNA with cognate bases.[6,8] The preference of fC over hmC and mC is much more pronounced for TDG with protonated His151 than with neutral His151. Although the calculated protonation free energies for these residues do not allow a conclusion about the preferred protonation state of His151 in a TDG-DNA complex with fC, the more favorable relative binding affinity suggests that His151 protonation might be relevant. Neither pK_a-dependent kinetic experiments[6] nor computations of the mechanism of hydrolytic fC cleavage[24] find His151 protonation important. His151 protonation is thus not strictly necessary for processing fC, but when His151 is protonated it contributes to preferred binding of fC over hmC or mC.

In TDG complexed to caC, His151 is clearly protonated, as can be seen from the computed proton affinities and the interaction energies of His151 with caC. This is in agreement with the observation of an acid catalyzed enzyme activity.[8,15]

https://doi.org/10.1021/acs.jpcb.1c09896
J. Phys. Chem. B 2022, 126, 1188−1201

Besides the interactions of His151 with the XC base that are stronger for the higher oxidized forms, caC and fC, further discrimination between cognate and noncognate forms of methyl-cytosines is achieved via interactions with Tyr152. The crystal structure of TDG complexed to DNA with fC[F21] and with caC[F16] also shows a short distance between active site residue Tyr152 and the fC base, suggesting a stable hydrogen bond between these residues. The hydrogen bond between NH and O15 discriminates oxidized from unoxidized methyl-cytosine due to the latter lacking the oxygen atom. hmC, on the other hand, contains such an oxygen atom. It is, however, engaged in hydrogen bonds with water as the donor, and therefore exhibits weaker and less frequent hydrogen bonds with Tyr152. The other oxidized XC bases, caC and fC, could in principle accept hydrogen bonds from water molecules with their O15 atom, but this is observed with moderate probability only for the charged caC which forms another strong hydrogen bond to water via its other oxygen atom, O16.

fC, in contrast, does not form such hydrogen bonds, and, indeed, according to the radial distribution functions, there is little probability for water molecules to be close enough for such bonds. Significant hydrogen bonded interactions of fC that may contribute to its discrimination are observed between Asn191 and the amino group of the fC base. This interaction is more probable for fC than for any of the other XC bases, including caC. At least in models with protonated His151, this interaction, together with that to His151 itself, contributes most to the more favorable interaction energy of the active site with fC than in the systems with the noncognate XC bases. Mutation experiments, however, showed that mutation of His151 or Asn191 to alanine has little or no effect on the enzymatic activity,[15] indicating that these residues do not contribute to the enzymatic reaction with fC. For caC, though, mutation of Asn191 leads to an inactive enzyme.[15] There are no significant hydrogen bond interactions between Asn191 and the caC base in our simulations, suggesting that this residue plays an important role in the chemical step rather than in substrate binding.

Besides the more favorable interactions, and hence preferred binding of the caC and fC bases, also the reaction competence of the complexes with the XC base flipped into TDG's active site contributes to the overall discrimination. That is, a base that is strongly bound, but does not undergo hydrolytic cleavage of the glycosidic bond, e.g., because it is not well placed for catalysis, may even act as an inhibitor for the enzyme. The requirements are a positioning of the nucleophile, likely by Asn140, and possibly also its activation. And, at least in the case of caC, leaving group activation through protonation, e.g., via His151, the presence of a positive charge, or strong hydrogen bonds to surrounding active site residues (including water molecules) can help leaving group departure.

The distributions of water molecules around the target for nucleophilic attack, the C1′ atom, and around Asn140 show a sufficiently high probability for a water molecule at reaction competent distances, placed by Asn140, for all models, albeit with different errors. A similar observation can be made for water molecules close to O2, that is on the other side of the sugar, and not held by Asn140. The presence of a water molecule here, though perhaps not acting as the nucleophile, might help the dissociation of the XC base by formation of hydrogen bonds or even proton transfer, upon or after departure of the base. Yet, such a scenario would be possible in all XC models simulated in this work and can therefore not be important for XC base discrimination.

As for the leaving group activation, calculations of the fC cleavage mechanism[24] show that the fC base can depart without further activation by a proton. caC cleavage, on the other hand, has been shown to be most likely acid catalyzed, or at least is faster at lower pH.[15] This is in agreement with our calculated proton affinities that predict a charged His151 when caC is bound to TDG. Moreover, the probability distributions of water around the carboxyl group and the hydrogen bond probabilities of this group with water render a proton transfer from His151 to caC via one or more water molecules possible. In fact, such a protonation step, either to neutralize the carboxyl group, or to the N3 atom, forming a zwitter ion, has been discussed as preceding the base dissociation.[16] Protonation prior to C1′−N1 bond scission would enable the caC base to depart as an anion and not a dianion, rendering it a likely first step in the chemical mechanism of caC cleavage.

Calculations of the glycosidic bond cleavage of a mispaired thymine base[23] have shown that also in the enzymatic environment departure of a neutral base is strongly favored over departure of an anionic base. fC seems to manage without protonation, i.e., can depart as an anion. hmC, that has been calculated to have a lower N1 acidity, and hence lower leaving group ability,[8,15,30] could benefit from such a proton transfer to either the hydroxyl group or the N3 atom. According to our simulations, both possibilities are rather unlikely since there is no hydrogen bond (with significant probability) observed between the N3 atom and water and the indeed probable hydrogen bonds between the hydroxyl group and water involve the hmC base as the donor. Though in TDG complexes with hmC protonated His151 is preferred over neutral His151, and thus a proton donor would be available, hmC seems to be placed not well enough for such a proton transfer to occur. The hmC base would then have to dissociate as anion and this is less likely than for the higher oxidized fC and neutral caC. mC could, in principle, accept a proton at its N3 atom but there is no indication of hydrogen bonds to this atom, either, and the leaving group abilities of methyl-cytosine rule out its departure as an anion.

Taken together, a large part of the discrimination between the different XC bases may still be necessary in the chemical step, exploiting different leaving group abilities and/or affinities for accepting a proton (or a strong hydrogen bond) prior to departure by the XC bases. The active site conformation in all the models studied in this work indicate nucleophile activation to follow similar mechanisms, though of course nothing can be concluded about the active site conformation after base dissociation, based on the present simulations of models with intact C1′−N1 bond. Even if the chemical step is the last resort for TDG to prevent aberrant cleavage of noncognate XC bases, the more favorable binding affinities of the cognate systems, fC and caC, render formation of the extrahelical TDG-DNA complex and, with that, binding of the XC base in the active site of TDG, an important means of substrate discrimination.

CONCLUSION

According to our simulation data, discrimination of cognate and noncognate methyl-cytosine bases is achieved to a large extent upon binding to TDG in extrahelical conformation. Whereas the positioning of the XC base in the active site is comparable in all XC models, the interaction strength with active site residues, via hydrogen bonds or undirected

https://doi.org/10.1021/acs.jpcb.1c09896
J. Phys. Chem. B 2022, 126, 1188−1201

electrostatic interactions, accounts for large parts of the differences in relative binding affinities. In particular, hydrogen-bonded interaction of the backbone of Tyr152 with the oxygen atom of the oxidized forms caC and fC is more favorable than in hmC and thus favors binding of the higher oxidized forms. Interaction with His151 is especially important for the negatively charged caC, in line with the high protonation affinity of His151 in that model.

Assuming that glycosidic bond cleavage essentially follows the same mechanism in all models, that is, C1′−N1 bond dissociation taking place before the attack of a nucleophilic water molecule, our simulations render all complexes equally reaction competent. They all exhibit the presence of putative nucleophilic water molecules and no obvious route for leaving group activation through, e.g., protonation. The exception is caC for which proton acceptance and thus neutralization is likely a prerequisite for glycosidic bond scission. Further discrimination in the chemical reaction of the (then) neutral XC bases, may be due to differences in the detailed mechanism or made possible by different leaving group abilities, intrinsic to the different XC bases. It is the same intrinsic differences in the chemistry of the XC bases that enable different interactions with the TDG enzyme, despite (seemingly) very similar accommodation of the XC bases in the active site, and that favor binding of the higher oxidized substrate bases over the noncognate XC bases.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.1c09896.

Conformational parameters of the DNA, table with van der Waals contributions to interaction energies, distance distributions between the XC base and protein residues, hydrogen bond probabilities between TDG and the DNA other than the XC bases, hydrogen bond probabilities of TDG residues and water, radial distribution functions of water molecules around the XC bases, analysis of interaction energies and hydrogen bond probabilities from thermodynamic integration perturbation simulations, and complete citation of reference (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Petra Imhof − *Department for Chemistry and Pharmacy Computer Chemistry Centre, Friedrich-Alexander University (FAU) Erlangen Nürnberg, 91052 Erlangen, Germany; Department of Physics, Freie Universität Berlin, 14195 Berlin, Germany;* orcid.org/0000-0001-9820-4578; Email: petra.imhof@fau.de

Authors

Frank Beierlein − *Department for Chemistry and Pharmacy Computer Chemistry Centre, Friedrich-Alexander University (FAU) Erlangen Nürnberg, 91052 Erlangen, Germany; Erlangen National High Performance Computing Center (NHR@FAU), Friedrich-Alexander University (FAU) Erlangen Nürnberg, 91058 Erlangen, Germany*

Senta Volkenandt − *Department for Chemistry and Pharmacy Computer Chemistry Centre, Friedrich-Alexander University (FAU) Erlangen Nürnberg, 91052 Erlangen, Germany; Department of Physics, Freie Universität Berlin, 14195 Berlin, Germany*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.1c09896

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research has been funded by the Deutsche Forschungsgemeinschaft through Grant IM141/1-2. The authors gratefully acknowledge the compute resources and support provided by the Erlangen Regional Computing Center (RRZE). The authors gratefully acknowledge the compute resources and support provided by NHR@FAU. F.B. thanks E. Beierlein for constant support. The authors thank Vincent Stegmaier and Anselm Horn for helpful discussions.

■ REFERENCES

(1) Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Rev. Gen.* **2012**, *13*, 484−492.

(2) Baylin, S.; Jones, P. A decade of exploring the cancer epigenome − biological and translational implications. *Nature Rev. Cancer* **2011**, *11*, 726−734.

(3) He, Y.; et al. TET-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **2011**, *333*, 1303−1307.

(4) Kohli, R. M.; Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **2013**, *502*, 472−479.

(5) Ito, S.; Shen, L.; Dai, Q.; Wu, S.; Collins, L.; Swenberg, J.; He, C.; Zhang, Y. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **2011**, *333*, 1300−1303.

(6) Maiti, A.; Drohat, A. C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine potential implications for active demethylatio of CpG sites. *J. Biol. Chem.* **2011**, *286*, 35334−35338.

(7) Hashimoto, J. H.; Hong, S.; Bhagwat, A. S.; Zhang, X.; Cheng, X. Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. *Nucl. Acid. Res.* **2012**, *40*, 10203−10214.

(8) Maiti, A.; Drohat, A. Dependance of substrate binding and catalysis on pH, ionic strength, and temperature for thymine DNA glycosylase: Insights into recognition and provessing ofG:T mispairs. *DNA repair* **2011**, *10*, 545−553.

(9) Imhof, P.; Zahran, M. The effect of a G:T mispair on the dynamics of DNA. *PLoS One* **2013**, *8*, e53305.

(10) Kanaan, N.; Imhof, P. Interactions of the DNA repair enzyme human Thymine DNA Glycosylase with cognate and non-cognate DNA. *Biochemistry* **2018**, *57*, 5654−5665.

(11) Bagherpoor Helabad, M.; Kanaan, N.; Imhof, P. Base-flip in DNA studied by molecular dynamics simulations of differently oxidised forms of methyl-cytosine. *Int. J. Mol. Sci.* **2014**, *15*, 11799−11816.

(12) Volkenandt, S.; Beierlein, F.; Imhof, P. Interaction of thymine DNA glycosylase with oxidised 5-methyl-cytosines in their amino- and imino-Forms. *Molecules* **2021**, *26*, 5728.

(13) Maiti, A.; Morgan, M.; Pozharski, E.; Drohat, A. Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 8890−8895.

(14) Hashimoto, H.; Zhang, X.; Cheng, X. Activity and crystal structure of human thymine DNA glycosylase mutant N140A with 5-carboxylcytosine DNA at low pH. *DNA repair* **2013**, *12*, 535−540.

(15) Maiti, A.; Michelson, A. Z.; Armwood, C. J.; Lee, J. K.; Drohat, A. C. Divergent mechanisms for enzymatic excision of 5-

formylcytosine and 5-carboxylcytosine from DNA. *J. Am. Chem. Soc.* **2013**, *135*, 15813−15822.
(16) Pidugu, L. S.; Dai, Q.; Malik, S. S.; Pozharski, E.; Drohat, A. C. Excision of 5-carboxylcytosine by Thymine DNA glycosylase. *J. Am. Chem. Soc.* **2019**, *141*, 18851−18861.
(17) Da, L.-T.; Yu, J. Base-flipping dynamics from an intrahelical to an extrahelical state exerted by thymine DNA glycosylase during DNA repair process. *Nucl. Acid. Res.* **2018**, *46*, 5410−5425.
(18) Dodd, T.; Yan, C.; Kossmann, B. R.; Martin, K.; Ivanov, I. Uncovering universal rules governing the selectivity of the archetypal DNA glycosylase TDG. *Proc. Nat. Am. Soc.* **2018**, *115*, 5974−5979.
(19) Tian, J.; Wang, L.; Da, L.-T. Atomic resolution of short-range sliding dynamics of thymine DNA glycosylase along DNA minor-groove for lesion recognition. *Nucl. Acid. Res.* **2021**, *49*, 1278−1293.
(20) Li, S.; Da, L.-T. Key structural motifs in thymine DNA glycosylase responsible for recognizing certain DNA bent conformation revealed by atomic simulations. *Biochem. Biophys. Res. Commun.* **2020**, *526*, 953−959.
(21) Pidugu, L. S.; Flowers, J. W.; Coey, C. T.; Pozharski, E.; Greenberg, M. M.; Drohat, A. C. Structural basis for excision of 5-formylcytosine by thymine DNA glycosylase. *Biochemistry* **2016**, *55*, 6205−6208.
(22) Bennett, M.; Rodgers, M.; Hebert, A.; Ruslander, L.; Eisele, L.; Drohat, A. Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability. *J. Am. Chem. Soc.* **2006**, *128*, 12510−12519.
(23) Kanaan, N.; Crehuet, R.; Imhof, P. Mechanism of the glycosidic bond cleavage of mismatched thymine in human thymine DNA glycosylase revealed by quantum mechanical/molecular mechanical calculations. *J. Phys. Chem. B* **2015**, *119*, 12365−12380.
(24) Naydenova, E.; Dietschreit, J. C. B.; Ochsenfeld, C. Reaction mechanism for the N-glycosidic bond cleavage of 5-formylcytosine by thymine DNA glycosylase. *J. Phys. Chem. B* **2019**, *123*, 4173−4179.
(25) Kaur, R.; Nikkel, D. J.; Wetmore, S. D. Computational studies of DNA repair: Insights into the function of monofunctional DNA glycosylases in the base excision repair pathway. *WIREs Comput. Mol. Sci.* **2020**, *10*, e1471.
(26) Jeong, Y. E. R.; Lenz, S. A. P.; Wetmore, S. D. DFT Study on the deglycosylation of methylated, oxidized, and canonical pyrimidine nucleosides in water: Implications for epigenetic regulation and DNA repair. *J. Phys. Chem. B* **2020**, *124*, 2392−2400.
(27) Williams, R. T.; Wang, Y. A density functional theory study on the kinetics and thermodynamics of N-glycosidic bond cleavage in 5-substituted 2′-deoxycytidines. *Biochemistry* **2012**, *51*, 6458−6462.
(28) Shapiro, R.; Kang, S. Uncatalyzed hydrolysis of deoxyuridine, thymidine and 5-bromodeoxyuridine. *Biochemistry* **1969**, *8*, 1806−1810.
(29) Shapiro, R.; Danzig, M. Acidic hydrolysis of deoxycytidine and deoxyuridine derivatives. General mechanism of deoxyribonucleoside hydrolysis. *Biochemistry* **1972**, *11*, 23−29.
(30) Drohat, A. C.; Maiti, A. Mechanisms for enzymatic cleavage of the N-glycosidic bond in DNA. *Org. Biomol. Chem.* **2014**, *12*, 8367−8378.
(31) Hashimoto, H.; Hong, S.; Bhagwat, A.; Zhang, X.; Cheng, X. Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: Its structural basis and implications for active DNA demethylation. *Nucl. Acid. Res.* **2012**, *40*, 10203−10214.
(32) Naydenova, E.; Roßbach, S.; Ochsenfeld, C. QM/MM Study of the uracil DNA glycosylase reaction mechanism: A competition between Asp145 and His148. *J. Chem. Theory Comput.* **2019**, *15*, 4344−4350.
(33) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucl. Acid. Res.* **2000**, *28*, 235−242.
(34) Case, D. A. et al. *AMBER 2018*; University of California: San Francisco, 2018.
(35) Vriend, G. WHAT IF: A molecular modeling and drug design program. *J. Mol. Graph.* **1990**, *8*, 52−56.
(36) Hardwick, J. S.; Haugland, M. M.; El-Sagheer, A. H.; Ptchelkine, D.; Beierlein, F. R.; Lane, A. N.; Brown, T.; Lovett, J. E.; Anderson, E. A. 2′-Alkynyl spin-labelling is a minimally perturbing tool for DNA structural analysis. *Nucl. Acid. Res.* **2020**, *48*, 2830−2840.
(37) Beierlein, F. R.; Paradas Palomo, M.; Sharapa, D. I.; Zozulia, O.; Mokhir, A.; Clark, T. DNA-dye-conjugates: conformations and spectra of fluorescence probes. *PLoS One* **2016**, *11*, e0160229.
(38) Ivani, I.; et al. Parmbsc1: a refined force field for DNA simulations. *Nat. Methods* **2016**, *13*, 55−58.
(39) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(40) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* **1995**, *16*, 1357−1377.
(41) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(42) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-consistent molecular-orbital methods. IX. An extended Gaussian-type basis for molecular-orbital studies of organic molecules. *J. Chem. Phys.* **1971**, *54*, 724−728.
(43) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; DeFrees, D. J.; Pople, J. A.; Gordon, M. S. Self-consistent molecular orbital methods. 23. A polarization-type basis set for 2nd-row elements. *J. Chem. Phys.* **1982**, *77*, 3654−3665.
(44) Hariharan, P. C.; Pople, J. A. Influence of polarization functions on molecular-orbital hydrogenation energies. *Theor. Chem. Acc.* **1973**, *28*, 213−222.
(45) Hariharan, P. C.; Pople, J. A. Accuracy of AH equilibrium geometries by single determinant molecular-orbital theory. *Mol. Phys.* **1974**, *27*, 209−214.
(46) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999−3094.
(47) Mennucci, B. Polarizable continuum model. *WIREs: Comput. Mol. Sci.* **2012**, *2*, 386−404.
(48) Frisch, M. J. et al. *Gaussian 16*, rev. B.01: Gaussian Inc.: Wallingford, CT, 2016.
(49) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(50) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(51) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247−260.
(52) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.
(53) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(54) Joung, I. S.; Cheatham, T. E., III Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **2008**, *112*, 9020−9041.
(55) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(56) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(57) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.

https://doi.org/10.1021/acs.jpcb.1c09896
J. Phys. Chem. B 2022, 126, 1188−1201

(58) Tutorial 9.1 NMR refinement of DNA and RNA duplexes. *The Amber Project*. http://ambermd.org/tutorials/advanced/tutorial4/index.php (accessed July 30, 2021).
(59) Case, D. A. et al. AMBER 2020. University of California, San Francisco, 2020.
(60) Steinbrecher, T.; Mobley, D. L.; Case, D. A. Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. *J. Chem. Phys.* **2007**, *127*, 214108.
(61) Beierlein, F. R.; Kneale, G. G.; Clark, T. Predicting the effects of basepair mutations in DNA-protein complexes by thermodynamic integration. *Biophys. J.* **2011**, *101*, 1130−1138.
(62) Lee, T.-S.; et al. Alchemical binding free energy calculations in AMBER20: Advances and best practices for drug discovery. *J. Chem. Inf. Model.* **2020**, *60*, 5595−5623.
(63) Tutorial 7.3 Thermodynamic Integration using soft core potentials/Side-chain mini tutorial. The Amber Project. http://ambermd.org/tutorials/advanced/tutorial9/index.html#sidechain_mini (accessed July 30, 2021).
(64) Tutorial 7.3 Thermodynamic Integration using soft core potentials/Small molecule binding to T4-lysozyme L99A. The Amber Project. http://ambermd.org/tutorials/advanced/tutorial9/index.html. (accessed July 30, 2021).
(65) Eberlein, L.; Beierlein, F. R.; van Eikema Hommes, N. J. R.; Radadiya, A.; Heil, J.; Benner, S. A.; Clark, T.; Kast, S. M.; Richards, N. G. J. Tautomeric equilibria of nucleobases in the Hachimoji expanded genetic alphabet. *J. Chem. Theory Comput.* **2020**, *16*, 2766−2777.
(66) Tutorial 7.2 pKa Calculations using Thermodynamic Integration. *The Amber Project*. http://ambermd.org/tutorials/advanced/tutorial6/index.php (accessed July 30, 2021).
(67) Simonson, T.; Carlsson, J.; Case, D. A. Proton binding to proteins: pKa calculations with explicit and implicit solvent models. *J. Am. Chem. Soc.* **2004**, *126*, 4167−4180.
(68) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **2013**, *9*, 3084−3095.
(69) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(70) Blanchet, C.; Pasi, M.; Zakrzewska, K.; Lavery, R. CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucl. Acid. Res.* **2011**, *39*, W68−W73.
(71) Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucl. Acid. Res.* **2009**, *37*, 5917−5929.
(72) Oliphant, T. *NumPy: A guide to NumPy*; Trelgol Publishing, 2006. http://www.numpy.org/.
(73) Van Rossum, G.; Drake, F. L. *Python 3 Reference Manual*; CreateSpace: Scotts Valley, CA, 2009.
(74) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* **2015**, *109*, 1528−1532.
(75) Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90−95.

https://doi.org/10.1021/acs.jpcb.1c09896
J. Phys. Chem. B 2022, 126, 1188−1201