Preorganized Internal Electric Field Promotes a Double-displacement Mechanism for the Adenine Excision Reaction by Adenine DNA Glycosylase

Wenwen Diao[1,2,3], James D. Farrell[3,4,5], Binju Wang[6], Fangfu Ye[2,4,7]*, and Zhanfeng Wang[1]*

[1] Dr. W. Diao, and Dr. Z. Wang, Center for Advanced Materials Research, Beijing Normal University, Zhuhai 519087, China

Email: zfwang@bnu.edu.cn

[2] Dr. W. Diao, and Prof. F. Ye, Wenzhou Institute, University of Chinese Academy of Sciences, Wenzhou, Zhejiang 325000, China

[3] Dr. W. Diao, Dr. J. D. Farrell, and Prof. F. Ye, School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

[4] Dr. J. D. Farrell, and Prof. F. Ye, Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China

E-mail: fye@iphy.ac.cn

[5] Dr. J. D. Farrell, Songshan Lake Materials Laboratory, Dongguan, Guangdong 523808, China

[6] Prof. B. Wang, State Key Laboratory of Physical Chemistry of Solid Surfaces and Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China

[7] Prof. F. Ye, Oujiang Laboratory (Zhejiang Lab for Regenerative Medicine, Vision and Brain Health), Wenzhou, Zhejiang 325000, China

KEYWORDS: Adenine DNA glycosylase • Base excision repair • Catalytic mechanism • Internal electric field • QM/MM calculations• MD simulations

ABSTRACT: Adenine DNA glycosylase (MutY) is a monofunctional glycosylase, removing adenines (A) misinserted opposite 8-oxo-7,8-dihydroguanine (OG), a common product of oxidative damage to DNA. Through multiscale calculations, we decipher a detailed adenine excision mechanism of MutY that is consistent with all available experimental data, involving an initial protonation step and two nucleophilic displacement steps. During the first displacement step, N-glycosidic bond cleavage is accompanied by the attack of residue Asp144 at the anomeric carbon (C1′), forming a covalent glycosyl-enzyme intermediate to stabilize the fleeting oxocarbenium ion. After departure of the excised base, water nucleophiles can be recruited to displace Asp144, completing the catalytic cycle with retention of stereochemistry at the C1′ position. Unsurprisingly, in the absence of the protein environment, the first displacement reaction, where Asp144 acts as the nucleophile, is highly exothermic with a negative barrier, yet the second, where an un-activated water molecule acts as the nucleophile, is prohibitive both kinetically and thermodynamically. Intriguingly, we find that the enzyme modulates these two reactions by coupling them together through an internal electric field at its active site, which reduces the barrier of the difficult one at the expense of raising that of the easy one, thereby allowing both reactions to occur. These findings not only increase our understanding of the strategies used by DNA glycosylases to repair DNA lesions, but also have important implications for how internal/external electric field can be applied to modulate multi-step reactions.

Introduction

Genomic DNA is prone to oxidation by reactive oxygen species (ROS) derived from normal metabolic functions such as mitochondrial respiration or from exposure to environmental pollutants and chemicals.[1] The most common form of ROS-induced DNA-base lesion is 8-oxo-7,8-dihydroguanine (OG) due to the low redox potential of guanine (G).[2] The OG lesion is especially problematic due to its ability to mimic thymine in its *syn* conformation during DNA replication, leading to a high rate of G:C to T:A transversion mutations. Adenine DNA glycosylase (MutY in bacteria, MUTYH in humans) acts as a last defense,

removing A from the OG:A mismatch, thus initiating a repair pathway that ultimately restores the original G:C base pair (refer to Figure 1 for the OG repair pathway).[3-8] The pivotal role of adenine DNA glycosylase was recognized upon the discovery of a direct association between defective MUTYH and colorectal cancer,[9] a syndrome now referred to as MUTYH-associated polyposis (MAP).[10-12]

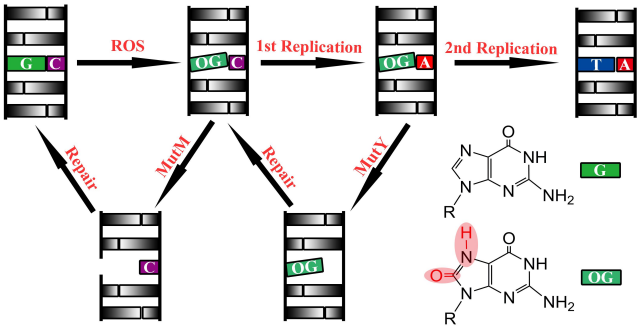


Figure 1. The OG repair pathway illustrating the role of MutY in preventing DNA mutations.

The importance of MUTYH in MAP and the promise of targeting DNA repair enzymes as therapeutic strategies have attracted tremendous efforts to understand the chemistry of MutY.[11-15] Insight into the adenine excision mechanism by MutY has been obtained through structural,[16-20] kinetic isotope effect (KIE),[21] mutational,[19, 22-23] and computational studies.[24-28] Initially, on the basis of the crystal structure of a lesion recognition complex (LRC) of MutY (PDB ID: 1RRQ, Figure S1)[16] and KIE-based transition-state (TS) analysis,[21] it was proposed that the reaction is initiated through N7 protonation of dA by protonated Glu43 via a water bridge, followed by rapid and reversible formation of an oxocarbenium ion intermediate with heterolytic N-glycosidic bond cleavage, which becomes irreversible after water attack (**Scheme 1a**). Residue Asp144 stabilizes the oxocarbenium ion, while residue Glu43 activates the water nucleophile, forming an abasic site and regenerating neutral Glu43.

However, Lee and Verdine argued that the captured LRC structure is not catalytically competent as there is no direct contact between the substrate adenine and active site residues.[17] They proposed, based on the crystal structure of a fluorinated LRC (FLRC, PDB ID: 3G0Q, Figure S2), a different water-attack mechanism, in which Glu43 directly protonates adenine before the cleavage of the glycosidic bond, and Asp144, rather than Glu43, accepts a proton to activate the water nucleophile (**Scheme 1b**).[17] Although the essential roles of Glu43 and Asp144 in this process were later demonstrated by mutagenic studies,[22] it is still hard to distinguish which mechanism is correct. Through detailed MD simulations, Wetmore and coworkers demonstrated that, of LRC and FLRC, only FLRC is catalytically competent.[27] Moreover, the mechanism shown in **Scheme 1b** is supported by their QM/MM calculations using FLRC as the starting structure.[27]

In both **Schemes 1a** and **1b**, the water nucleophile attacks from the 3′ face, leading to an inversion of the stereochemistry at C1′. However, MutY-catalyzed methanolysis studies by Woods et al. implied retention of the stereochemistry at C1′ upon adenine removal.[18] On the basis of a new crystal structure of a transition state analog complex (TSAC, PDB ID: 6U7T, Figure S3),[18, 29] in which a potential water nucleophile is found on the 3′ face rather than the 5′ face, David and coworkers proposed yet another mechanism for MutY that involves two nucleophilic displacement steps (**Scheme 1c**).[18] After protonation at N7, the N-glycosyl bond breaks via an SN1 mechanism, yielding an oxocarbenium ion intermediate. The oxocarbenium ion is then attacked by Asp144 to form a covalent glycosyl-enzyme intermediate, which survives long enough to allow the departure of the excised adenine base and the arrival of a water nucleophile. Finally, the newly-recruited water nucleophile, activated by deprotonated Glu43, can attack C1′ from the 5′ face of the sugar ring to displace Asp144, forming the abasic site with retention of the stereochemistry at C1′. Recently, enzyme-bound product structures have been successfully captured through crystallization in N146S mutants by David and coworkers,[30] which feature β-anomer abasic sites and thus support the mechanism shown in **Scheme 1c**. However, the key covalent glycosyl-enzyme intermediate has not been captured in crystal structures.

Therefore, the molecular details of MutY-catalyzed adenine removal are still incomplete, and it is interesting to know why it does not follow the canonical pathway shown in **Scheme 1b** if the double-displacement reaction mechanism shown in **Scheme 1c** is correct. Key questions include: (1) can a covalent bond form between Asp144 and the anomeric carbon C1′? (2) why not water nucleophile attack from the 3′ side? (3) which residue is the proton acceptor to activate the water nucleophile, Asp144 or deprotonated Glu43? (4) what is the role of Tyr126? (5) what is the role of the protein environment? Additionally, given the critical role of the internal electric field (IEF) in the typical DNA glycosylase, uracil-DNA glycosylase (UDG),[31] we wonder whether there is a nonnegligible IEF at the active site of MutY and what its function is.

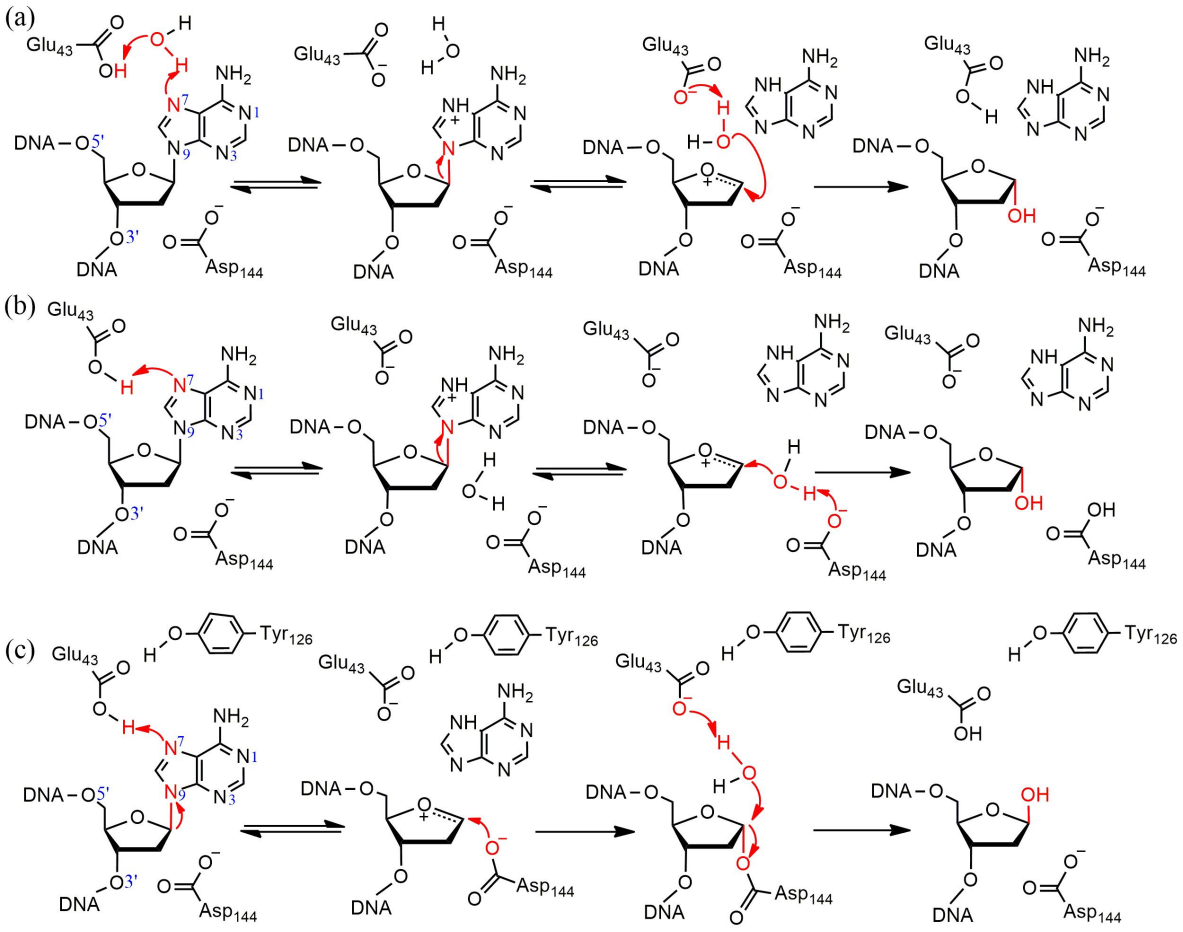
In the present study, we combine MD simulations, QM calculations, and QM/MM calculations to explore the possibility of different base excision pathways of MutY. A mechanism involving two nucleophilic displacement steps is revealed, which provides strong support for the double-displacement mechanism proposed by David and coworkers.[18] On the basis of our identified reaction pathway, roles of key residues near the active site and the protein environment are further discussed. Intriguingly, we find that an IEF in MutY, pointing toward the 5′ direction that is opposite to the one in UDG, plays crucial roles in the modulation of involved chemical reactions.

Computational Methods

System Setup. Our computational models were constructed on the basis of the fluorinated lesion recognition complex (FLRC) crystal structure (PDB ID:

3G0Q).[17] Missing residues (residue ID: 230-233 and 288-291) were added through SWISS-MODEL (available at: swissmodel.expasy.org). The first model (system I) constructed directly from the crystal structure where the adenine has not been excised was used to investigate the initial protonation step as well as the first displacement reaction, and the generally-accepted water-attack pathway in DNA glycosylases. After the excision of the adenine base, two new models were constructed on the basis of the QM/MM-optimized structure of the excised species (Int2 in Figure 2) to investigate the departure process of the excised adenine base (system II) and the second displacement reaction (system III), respectively. The excised adenine base is retained in system II, while it is removed in system III to make room for water molecules. The $[4Fe\text{-}4S]^{2+}$ cluster was parameterized using the "MCPB.py" tool.[32-33] The general AMBER force field (GAFF)[34] was used to generate the parameters for the OG, the excised adenine base, and the sugar ring, with partial atomic charges obtained using the RESP method[35] at the HF/6-31G(d) level of theory. The parmchk2 utility from AmberTools was used to generate the missing parameters. The Amber ff14SB force field[36] was employed for the normal protein residues and nucleotides. The protonation states of titratable residues (histidine, glutamate, and aspartate) were assigned on the basis of the pKa values from the PROPKA software[37] in combination with careful visual inspection of local hydrogen-bond networks. All histidine residues (residue ID: 96, 305, 309, 323, and 348) were protonated at the ε position (HIE). With the exception of residue 43, all glutamate and aspartate were deprotonated. Glutamate 43 was protonated in system I, and deprotonated in systems II and III. The resulting systems were then solvated in a rectangular box of TIP3P[38] waters extending up to a minimum cutoff of 12 Å from the protein surface. Finally, sodium ions (Na+) were added to make our systems electrically neutral.

Scheme 1. Proposed catalytic mechanisms of MutY. (a) A water-attack mechanism based on the lesion recognition complex (LRC) structure, in which Glu43 acts as the general acid and base. (b) A water-attack mechanism based on the fluorinated LRC (FLRC) structure, in which Glu43 acts as the general acid yet Asp144 acts as the general base. (c) A double-displacement mechanism based on the transition state analog complex (TSAC) structure.



Molecular Dynamics Simulation. All systems were energy minimized using a combination of the steepest descent and conjugate-gradient methods. Subsequently, the system was gently heated from 0 to 300 K in the canonical ensemble for 300 ps, using a weak restraint of 25 kcal·mol^{-1}·Å^{-2} on the non-solvent atoms. A 1 ns of density equilibration followed, employing an isothermal-isobaric Langevin thermostat[39] and a Berendsen barostat[40] at 300 K and 1.0 atm with a collision frequency of 2 ps^{-1} and a pressure relaxation time of 1 ps to achieve a uniform density. For system I, a 40ns restraint-MD simulation with both protein and DNA atoms restrained with a weak potential of 5 kcal·mol^{-1}·Å^{-2} was performed to relax the solvent and counterions (Na^+). For system II, 5 independent 100 ns classical MD simulations were performed to investigate the stability of the excised adenine base. For system III, one 100 ns classical MD simulation was performed to relax water molecules around the remaining sugar ring. In all MD simulations, an integration step of 2 fs was used, and covalent bonds containing hydrogen were constrained by the SHAKE algorithm.[41] Nonbonded interactions were treated with the Particle Mesh Ewald method[42] with a 10 Å cutoff. All MD simulations were performed with the GPU version of the Amber 18 package.[43-45]

Umbrella Sampling Simulation. Due to the high stability of the excised adenine base during classical MD simulations, umbrella sampling[46] combined with restrained MD simulations was used to explore its departure process. The center of mass (COM) distance between 4 nitrogen atoms of the adenine base (N1, N3, N7 and N9, refer to **Scheme 1** for the labels of these four nitrogen atoms) and the anomeric carbon C1′ was chosen as the collective variable (CV). The starting structure was taken from an MD-equilibrated snapshot of system II, in which the CV is around 5.0 Å. The selected CV was varied from 5.0 to 20.0 Å, with an initial interval of 0.3 Å for two adjacent windows, giving 51 windows. To smooth the final potential of mean force (PMF) profile, 36 more windows were added, giving 87 windows in total. For each window, a 5 ns MD simulation with a biasing harmonic potential (with a force constant of 50 kcal·mol^{-1}·Å^{-2}) was carried out to relax the system. Thereafter, a 15 ns production MD simulation with the same biasing harmonic potential was performed for each window to collect data needed for the PMF analysis. In the production simulations, a dump frequency of 5 ps^{-1} of the value of CV was used. The PMF was constructed with the weighted histogram analysis method (WHAM),[47-48] using the CV data from the production simulations. Good convergence has been achieved since the final PMF has nearly no change when it is recalculated with the first 80% of the data, indicating that the production simulation time is long enough to ensure sufficient overlap between two adjacent windows and further elongation is not needed.[49-50]

Quantum Mechanical/Molecular Mechanical (QM/MM) Calculation. The last snapshot of the restraint simulation of system I was used as the input of QM/MM calculations to investigate the first displacement reaction and the water-attack pathway, and a representative snapshot extracted from the MD simulation trajectory of system III was used as the input structure to investigate the second displacement reaction. All QM/MM calculations were performed with ChemShell software,[51] combining Turbomole[52] for the QM region and DL_POLY[53] for the MM region. For the QM/MM calculation of system I, the QM region consists of the substrate dA, the phosphodiester group of its 3′ side (labeled as PO_4^-), Trp30, Arg31, Glu43, Tyr126, Asp144, Asn146, Glu188, and four vicinal water molecules (131 atoms in total, refer to Figure 4a). For the QM/MM calculation of system III, the QM region includes the sugar ring of the excised dA, the phosphodiester group of its 3′ side (labeled as PO_4^-), Glu43, Tyr126, Asn146, the covalently bonded residue Asp144, and three important water molecules (71 atoms in total, refer to Figure 4c). The electronic embedding scheme[54] was used to account for the polarizing effect of the enzyme environment on the QM region. Hydrogen link atoms with the charge-shift model were applied to treat the QM/MM boundary. The hybrid B3-LYP density functional[55-57] with two levels of theory was applied to the QM region, while the MM region was treated with the Amber parameters. B3LYP was chosen due to its good performance and wide application in the studies of DNA glycosylases.[58-64] Transition states were located with relaxed potential energy surface scans using a small basis set of 6-31+G(d) (labeled B1). All reactant complex and intermediate species were fully optimized at B3LYP/B1 level. The electronic energies of all stationary and transition state species were further corrected with a larger basis set of 6-311+G(d,p) (labeled B2). Besides B3LYP, several other functionals, BP86,[65-66] M06L,[67] and M06-2X[68] were also tested in single-point calculations. Dispersion corrections computed with Grimme's D3 method[69-71] were included in the QM region of all QM/MM calculations. The electronic energies at the B3LYP-D3/B2 level were treated as the estimate of Gibbs free energies due to the observation that "the differences between QM/MM electronic energy and free energy profiles in enzymatic reactions tend to be rather small as long as local chemical events are investigated".[72] Indeed, the entropic contribution to the rate-limiting C-N bond cleavage step in DNA glycosylases has been found to be small in both experiments[73] and computations.[31, 74-76] Besides, through metadynamics simulations,[77-78] we confirmed that the obtained energy barrier in above way is a quite good estimation to the free energy barrier.

QM calculations. The effect of key residues near the active site on the catalytic mechanism was investigated with QM calculations. The geometries of key species were taken from our QM/MM-obtained structures. Moieties directly involved in a reaction were included in the QM model to obtain its intrinsic energetic character in the absence of the influence of surrounding residues. For the initial protonation reaction of dA by protonated Glu43 (RC → Int1 in Figure 2), the QM model (denoted as QM_{S1}, Figure S4) includes only the side chain of Glu43, as well as the base and sugar ring of the dA substrate. For the first displacement reaction where Asp144 acts as a nucleophile (Int1 → Int2 in Figure 2), the QM model (denoted as QM_{S2},

Figure S5) consists of protonated adenine base and sugar ring of the substrate dA, as well as the side chain of Asp144. For the second displacement reaction (Int2′ → Int3 in Figure 3), the QM model (labeled QM_{S3}, Figure S6) includes the remaining sugar ring of the excised dA, the side chain of covalently bonded Asp144, and the water nucleophile. The effect of a particular residue/group can be estimated by the energy profile difference before and after its addition into the QM model. For example, the effect of residue Tyr126 on the reaction barrier of the initial protonation reaction can be estimated with the change of the energy difference between RC and TS1 species before and after the addition of Tyr126, $[(E_{TS1}\text{-}E_{RC})_{After} - (E_{TS1}\text{-}E_{RC})_{Before}]$. Owing to the large number of residues in the protein, only the key vicinal residues that were included in the QM region of QM/MM calculations, were investigated in this way. The collective effect of residues excluded from the QM region was estimated by taking the difference between the results of the QM/MM calculations and the QM calculations with the bare QM region.

Groups/residues can be added individually or cumulatively, which may, in principle, lead to different results. In the current case, the results obtained with these two different methods are generally similar with the exception of Tyr126, due to its strong hydrogen bonding interaction with Glu43. In addition to the explicit addition of residue atoms, the effect of a particular residue/group can be also examined in terms of the background charges of its atoms, which can be viewed as an estimate of the purely electrostatic effect of the residue/group.

The effect of oriented external electric fields (OEEFs) was studied with the "Field = *M* ± *N*" keyword in Gaussian 16, which defines the axis of the OEEF, its direction along that axis (*M*) and its magnitude (*N* a.u., 1 a.u. = 5140 MV/cm). It is worth noting that the positive direction of the electric field in Gaussian 16 is defined from the negative to the positive charge, which is opposite to the conventional definition in physics. To be consistent with the direction of the calculated internal electric field, the direction of the external electric field has been inverted in related figures. The effects of OEEFs were calculated using single-point calculations at the B3LYP-D3/B2 level. For the first displacement reaction, the QM_{S2} model (shown in Figure S5) was used, while the QM_{S3} model (shown in Figure S6) was used for the second displacement reaction.

All QM calculations were performed with the Gaussian 16 (G16) software.[79]

Internal electric field (IEF) estimation. Calculating the IEF in the active site of an enzyme is not an easy task to due to its anisotropy.[31] The problem can be simplified by computing the IEF at a representative point and assuming that the electric field is isotropic. In our case, we are interested in the effect of the IEF on the two displacement reactions. Therefore, the representative point was chosen to be the anomeric carbon (C1′), which is the center atom of the two displacement reactions involved. The electric field was calculated with Coulomb's law, treating the atoms as point charges with positions taken from QM/MM-optimized structures. To investigate the effect of IEF on the first and the second displacement reactions, the QM/MM-optimized structures of Int1 and Int2′ species were used, respectively. Protein point charges were taken from the Amber ff14SB forcefield,[36] and water point charges from the TIP3P model.[38] To avoid a double-counting error, the atomic charges of the groups/residues directly involved in the reaction (QM region) were set to zero during the calculations of the IEF, the effect of which has been inherently included in the QM region. We note that precise calculation of the effect of IEF needs more accurate methods for the calculation of both the bond dipole moment[80] and the electric field.[81]

QM/MM MD and metadynamics simulations. One representative snapshot extracted from the classical MD trajectory of system I was used for the QM/MM MD simulations to compare with the results of static QM/MM calculations. All QM/MM MD simulations were performed with the CP2K package (2022.2 version),[82] which combines the QM program QUICKSTEP[83] and the MM driver FIST. A real space multigrid technique is used to compute the electrostatic coupling between the QM and MM regions.[84-85] The QM region was treated at the DFT(B3LYP-D3) level, employing a dual basis set of Gaussian and plane-waves (GPW) formalism,[83] whereas the MM region was modelled at the classical level using the same force-field as in the classical MD simulations. The QM region was chosen as the same in the static QM/MM calculations (refer to Figure 4a). The wave function was expanded in a Gaussian double-ζ valence polarized (DZVP) basis set,[86] while an auxiliary plane-wave basis set with a cutoff of 360 Ry was used to converge the electron density, in conjunction with Geodecker-Teter-Hutter (GTH) pseudopotentials[87-88] for treating the core electrons. To accelerate the calculation of the Hartree-Fock exchange within B3LYP, the auxiliary density matrix method (ADMM) was used.[89] All QM/MM MD simulations were performed in the NVT ensemble using an integration time step of 0.5 fs. The systems were firstly equilibrated without any constraint for 2.0 ps. Then, the well-tempered metadynamics method was used to explore the N-glycosidic bond cleavage process.[77-78] The collective variable (CV) were selected as the distance between C1′ and N9 atoms. The width of the Gaussian-shaped potential hills was set as 0.2 Å. The Gaussian height was set to 0.6 kcal/mol, while the time deposition interval between two consecutive Gaussians was set to 10 fs.

Results and Discussion

1. Identified reaction pathway. 1.1. Initial protonation of the adenine base. It is generally accepted that the protonation of dA facilitates its release, and that the catalytic reaction starts with protonation of dA by a protonated residue Glu43.[16-18, 21, 24-28] Besides, it has been shown both experimentally[17] and theoretically[27] that a direct hydrogen bond forms between the N7 atom of dA and the side chain of Glu43 in the catalytically competent complex (FLRC). To investigate the process of this vital proton transfer (PT) reaction, neutral adenine base and protonated Glu43 (state before the PT) were assigned in the reactant complex (refer to RC species in Figure 2). As

shown in Figure 2b, the PT process (RC → Int1 via TS1) is quite facile with a rather small barrier of 1.1 kcal/mol (refer to Figure S7 for the QM/MM-obtained TS1 structure) and a reaction energy of -0.3 kcal/mol. The PT process is reversible as evidenced by the small barrier as well as the relatively long N7-H bond distance (1.08 Å) in the Int1 species (Figure 2c).

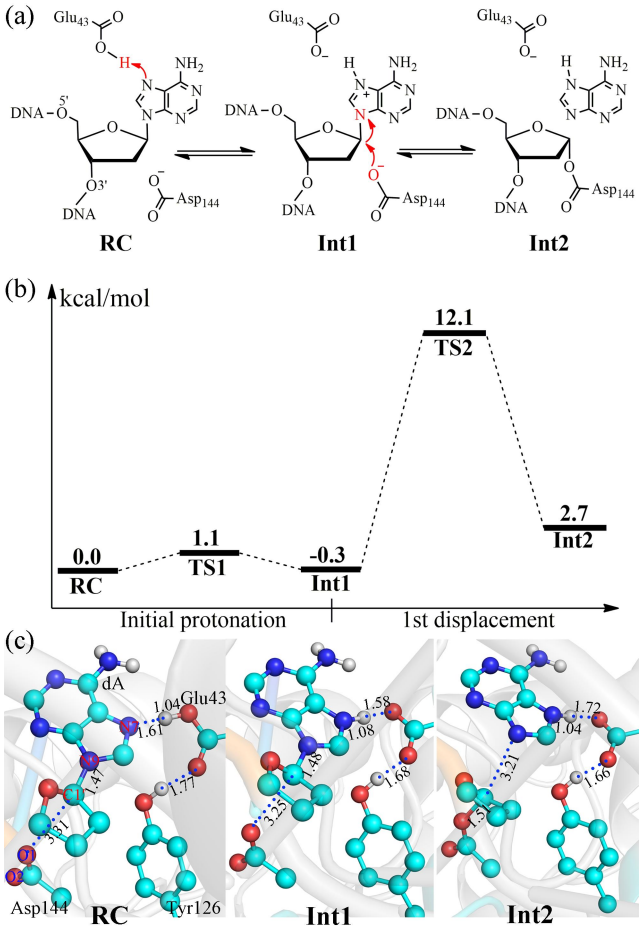


Figure 2. The initial protonation step and the first displacement step. (a) Schematic illustration of these two steps. (b) QM(B3LYP-D3/B2)/MM-calculated energy profile (in kcal/mol). (c) QM(B3LYP-D3/B1)/MM-optimized geometries of key intermediates. Distances are given in Å.

1.2. First displacement. After protonation, heterolytic N-glycosidic bond breakage occurs, affording a cationic oxocarbenium ion and a neutral leaving nucleobase. The bond breakage barrier is calculated to be 12.4 kcal/mol relative to Int1 species (TS2 in Figure 2b; refer to Figure S7 for its structure), significantly lower than that (> 20 kcal/mol) estimated by Wetmore and coworkers for the water-attack pathway shown in **Scheme 1b**.[27] At the local minimum (Int2), which is 2.7 kcal/mol higher in energy than the RC species, a covalent bond forms between the anomeric carbon (C1′) and the O1 atom of Asp144. The C1′-O1 distance shortens from 3.31 Å to 1.51 Å, while the C1′-N9 distance elongates from 1.47 Å to 3.21 Å (refer to Figure 2c). It is worth noting that this reaction mainly involves the movement of the sugar ring towards the carboxyl group of Asp144 (refer to Figure S8); the adenine base is not significantly displaced due to the intricate hydrogen-bond network it forms with surrounding residues (refer to Figure S9). Clearly, *the negatively charged residue Asp144 has a more direct role that has not been seen in other DNA glycosylase enzymes*: it acts as a nucleophile, attacking the oxocarbenium ion, and neutralizes the positive charge developed during the glycosidic bond cleavage process. The role of Asp144 revealed here is consistent with the experimental observation that MutY retains activity as long as the pH maintains the anionic charge of this position when it is replaced with glutamate and cysteine.[22]

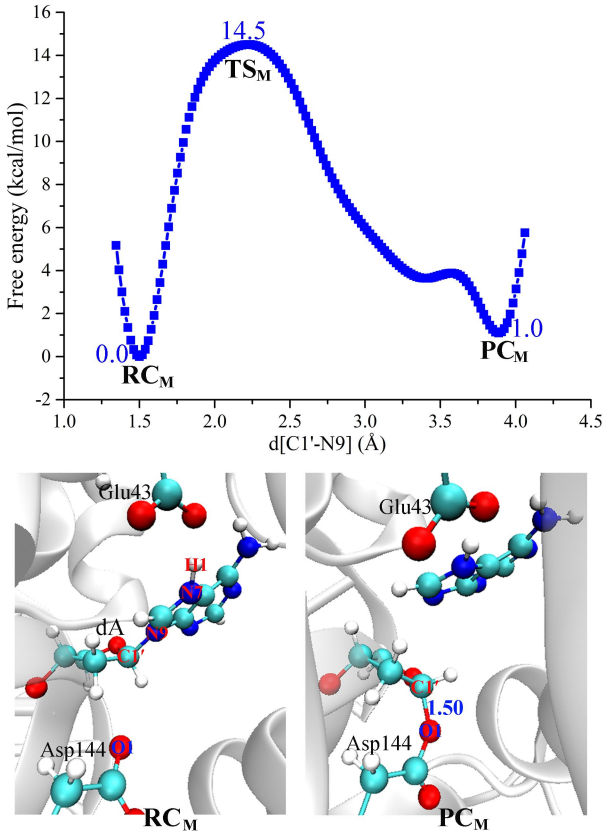


Figure 3. Free energy profile for the N-glycosidic bond breakage, obtained with QM/MM-based metadynamics simulations, shown along with the geometries of the RC_M and PC_M species. Notably, after the cleavage of the glycosidic bond, a covalent bond formed between the O1 atom of residue Asp144 and the anomeric carbon C1′.

The formation of the covalent glycosyl-enzyme intermediate is further confirmed by QM/MM MD simulations. Before the cleavage of the glycosylic bond, the proton of the residue Glu43 fluctuates between the N7 atom of the adenine base during the equilibration process (Figure S10). After the cleavage of the glycosidic bond, the proton has transferred to the N7 position and a covalent bond forms between O1 atom of residue Asp144 and the

anomeric carbon C1′ (refer to Figure S11). The free energy barrier of the whole process is estimated to be 14.5 kcal/mol via well-tempered metadynamics simulations (Figure 3),[77-78] which is close to the one shown in Figure 2b. It is noteworthy that the collective variable used in the well-tempered metadynamics simulation only involves the distance between C1′ and N9 atoms. Therefore, the bond formation between O1 atom of Asp144 and C1′ in the product complex (PC_M in Figure 3) supplies a strong support for the formation of a covalent glycosyl-enzyme intermediate.

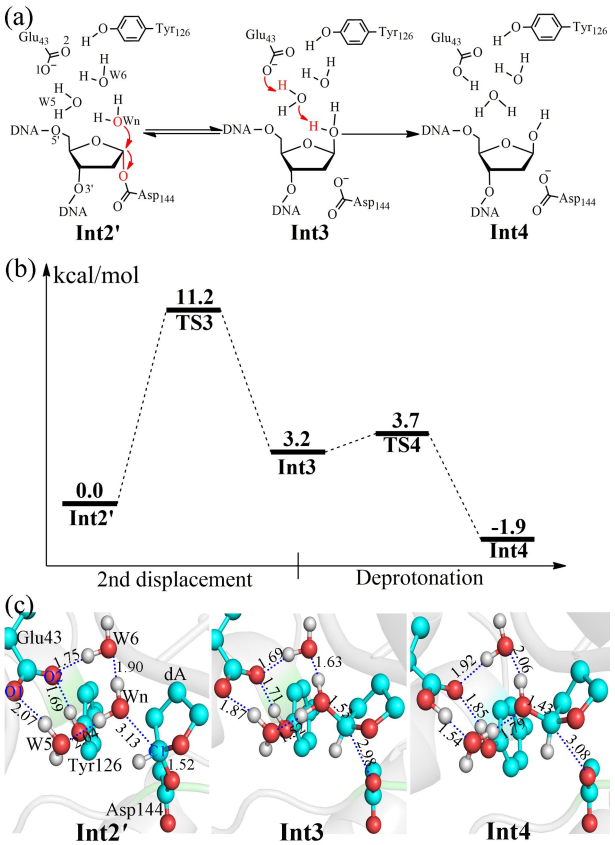


Figure 4. The second displacement step. (a) Schematic of the second displacement step, wherein two proton transfer pathways are involved. (b) QM(B3LYP-D3/B2)/MM-calculated energy profile (in kcal/mol). (c) QM(B3LYP-D3/B2)/MM-optimized geometries of key intermediates. Distances are given in Å.

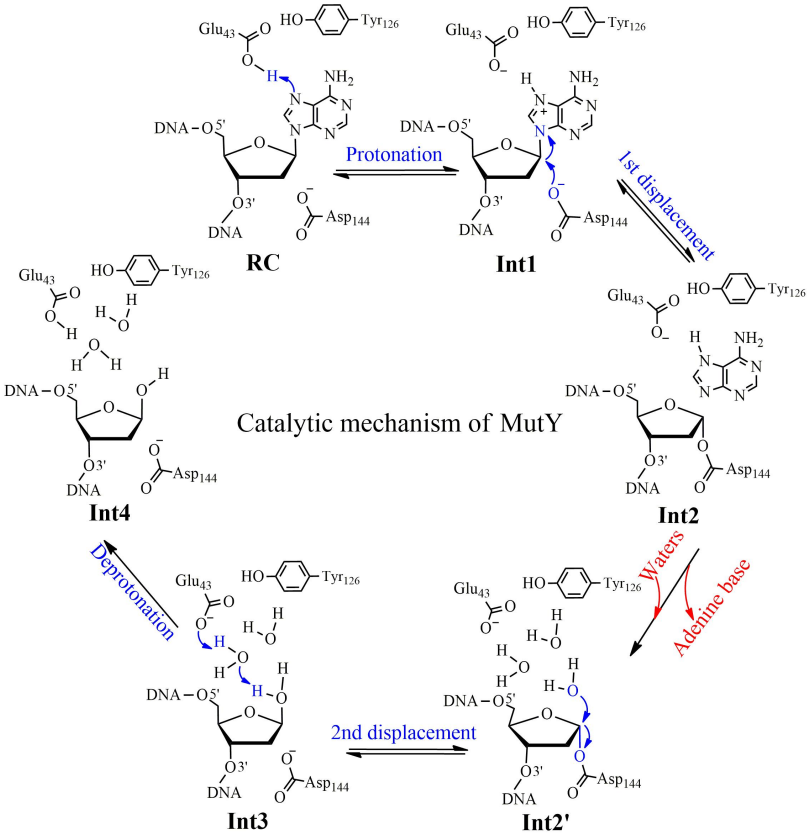
1.3. Adenine base departure. To facilitate the dissociation of the enzyme from the double-stranded DNA, the covalent bond formed between the enzyme and the sugar ring needs to be broken and a second displacement reaction in which a water molecule acts as the nucleophile to displace Asp144 is expected. Since Asp144 is bonded to the 3′ side of the sugar ring, the water nucleophile can only attack from the 5′ side and the deprotonated residue Glu43 becomes the only potential proton acceptor to activate the water nucleophile. However, the cleaved adenine base forms a stable H-bond with Glu43 (refer to Int2 species in Figure 2, as well as Figure S12), which not only prevents approach of water molecules but also impedes the ability of Glu43 to be a proton acceptor. Therefore, departure of the adenine base is expected to occur before the second displacement. However, the excised adenine base was found to be rather stable in 5 independent 100-ns classical MD simulations, the distance between C1 and N9 atoms maintaining a value of ca. 3.6 Å (Figure S13), due to the H-bonds formed with vicinal residues Arg31, Glu43, and Glu188 (Figure S9). Through umbrella sampling simulations, the departure barrier of the nucleobase is estimated to be ca. 10.0 kcal/mol (Figure S14), which is comparable to the barrier of the first displacement step (12.4 kcal/mol, shown in Figure 2). Notably, no excised adenine base has been observed in any MutY crystal structure. Therefore, we propose that the departure process is free-energy decreasing and entropy-driven.

1.4. Second displacement. After departure of the excised adenine base, several water molecules were recruited to the 5′ face of the sugar ring in a 100-ns classical MD simulation. As shown in the Int2′ species of Figure 4c, a water molecule (Wn), residing ca. 3.1 Å above the anomeric carbon in a representative snapshot of the MD simulation trajectory, is appropriately located for nucleophilic attack. Intriguingly, we find that the formation of the C-O bond precedes significantly the removal of the proton from the nucleophilic water in MutY, which is in contrast with the case of UDG where they occur almost simultaneously.[31, 64, 90-91] As shown in Figure 4, the attack sub-step (Int2′ → Int3) has a moderate barrier of 11.2 kcal/mol (TS3 in Figure 4b; refer to Figure S15 for the QM/MM-obtained TS3 structure). Without activation the Int3 intermediate lies 3.2 kcal/mol higher than the Int2′ species. Notably, the newly-formed C-O bond distance of 1.53 Å is slightly longer than its typical value of ca. 1.45 Å. In common with the first displacement step, this process mainly involves the movement of the sugar ring toward the nucleophile (Figure S16). Subsequently, activation via proton transfer to the O1 atom of the deprotonated residue Glu43 reduces the energy of the complex and completes the whole catalytic cycle. Our calculations show that the deprotonation process is quite facile with a small barrier of only 0.5 kcal/mol relative to the Int3 species (TS4 in Figure 4b; refer to Figure S15 for its QM/MM-obtained structure). Notably, upon deprotonation of the added water molecule, the C-O bond distance decreases from 1.53 Å to 1.43 Å (refer to Int3 and Int4 species in Figure 3c), the bond becoming stronger. Besides, it is noteworthy that the unactivated water molecule serving as a nucleophile is consistent with the experimental fact that methanol can be used as a nucleophile to detect the stereochemistry of the anomeric carbon.[18]

1.5. Summary of the identified pathway. Overall, the MutY-catalyzed adenine excision reaction is shown to be triggered by the initial protonation of the adenine leaving group and followed by two displacement steps (summarized in **Scheme 2**). During the first displacement step, negatively charged residue Asp144 acts as a nucleophile to attack and then stabilize the oxocarbenium

cation ion. The formation of a covalent glycosyl-enzyme intermediate has been validated by both QM/MM calculations and QM/MM MD simulations. After departure of the adenine base, water molecules can be recruited to attack the anomeric carbon from the 5′ side of the sugar ring to displace Asp144, completing the whole catalytic cycle. The neutral Glu43 has been regenerated by proton transfer from the water nucleophile. Notably, the highest barrier (12.4 kcal/mol, Int1 → TS2) of the whole catalytic process is lower than the estimated barrier (ca. 19 kcal/mol) from experiments.[92-94] Although we find that the barrier is underestimated by B3LYP functional[55-57] compared with M06-2X (Table S1 and Table S2),[68] we cannot rule out the possibility that either the recognition of the OG:A mispair or the alignment of the substrate dA in the active site is the true rate-limiting step, which is more consistent with the experimental result that rapid alignment with the catalytic residues correlates with fast cleavage of the substrate analogue.[95]

Scheme 2. Proposed reaction mechanism for Mut-Y catalysed glycosyl bond cleavage in this study.



2. Necessity of the initial protonation step. Despite the consensus that protonation of N7 position of the adenine occurs prior to its excision,[16-18, 21, 24-28] there is still no direct evidence to show that this pre-protonation step is necessary. We demonstrate this by a QM/MM calculation where the protonated Glu43 group is removed from the QM region during the first displacement process (refer to Figure S17 for the QM region used). In such a case, the H-bond between Glu43 and adenine is preserved, yet no proton transfer occurs. It is found that without protonation the barrier of the first displacement step increases from 12.4 kcal/mol to 22.2 kcal/mol (Figure S18). In addition, the reaction becomes highly endothermic with a reaction energy of 17.7 kcal/mol (Figure S18), in sharp contrast with the case of the same reaction when adenine is pre-protonated.

3. Possibility of the water-attack pathway. The role of Asp144 as a nucleophile to attack the anomeric carbon of the oxocarbenium cation makes MutY unique among DNA glycosylases. Therefore, it is necessary to compare with the canonical pathway, in which an activated water molecule serves as the nucleophile. To investigate the possibility of a competing water-attack pathway, a water molecule below the sugar ring (W1 in Figure 5a) is chosen as the nucleophile. Before it can attack, a strong H-bond formed between Asp144 and Asn146 must be broken to allow further approach of the water nucleophile toward the anomeric carbon, which needs to surmount an energy penalty of 7.0 kcal/mol (refer to Figure S19). Subsequently, N-glycosidic bond cleavage concomitant with the nucleophilic attack of the water molecule occurs with a barrier of 25.5 kcal/mol and a reaction energy of -2.2 kcal/mol relative to the Int1 species, affording the final products, an abasic site with inverted stereochemistry, and a neutral adenine base (refer to Figure S19). The calculated barrier is close to the result (23.5 kcal/mol) of Wetmore and coworkers,[27] yet significantly higher than the highest barrier (12.4 kcal/mol) identified for the double-

displacement mechanism. Moreover, the inverted stereochemistry of the abasic site is inconsistent with the experimental results of Woods et al.[18, 30] Therefore, the canonical water-attack pathway is supported neither theoretically nor experimentally.

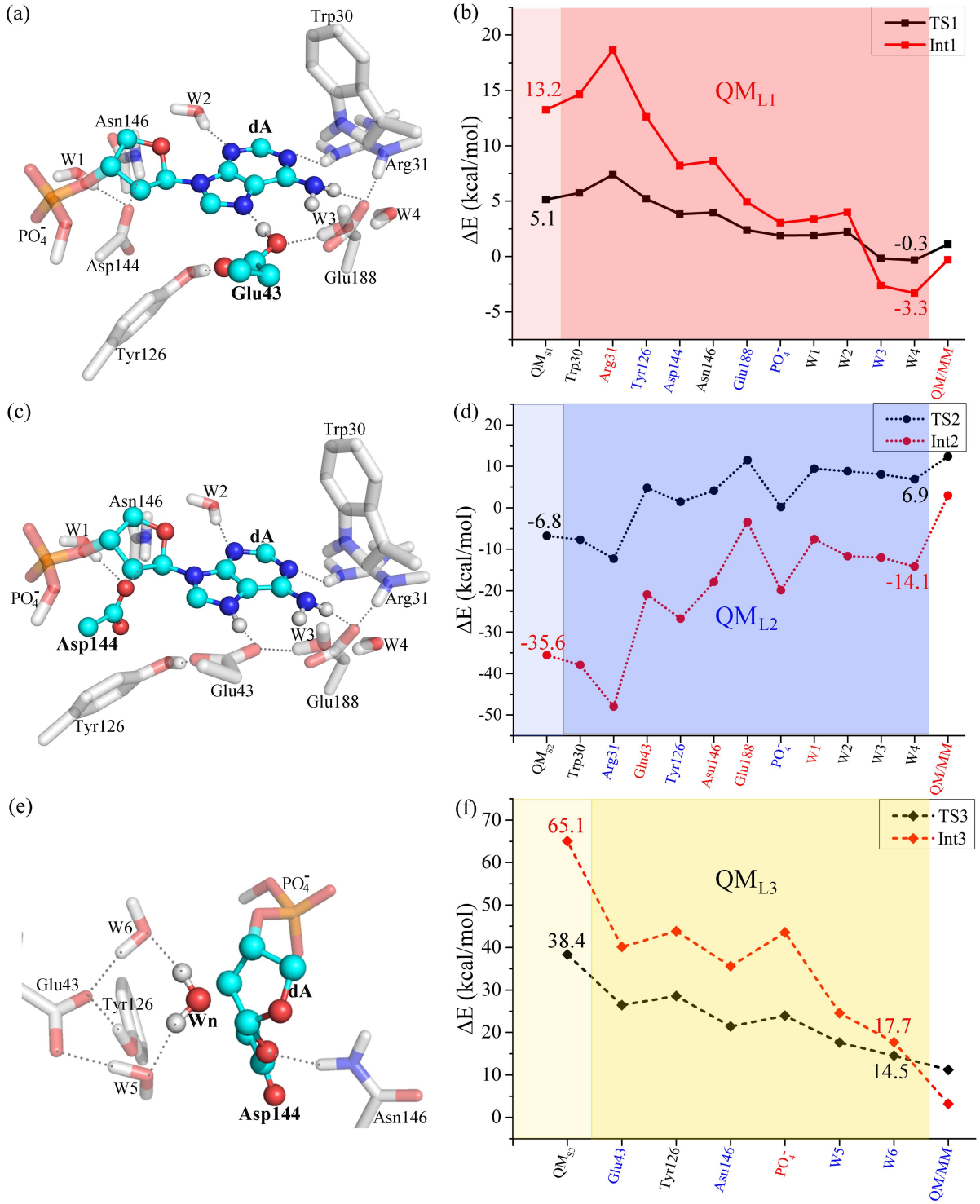


Figure 5. Roles of key residues near the active site and the protein environment. Illustrations of residues/groups included in the corresponding QM_S and QM_L regions are shown in (a), (b) and (c) for the initial PT, the first displacement, and the second displacement reaction, respectively. Residues included in QM_S region are shown in solid balls and sticks; residues included in QM_L region are shown in transparent sticks. Notably, the region of (QM_S + QM_L) is exactly the QM region of QM/MM calculations for the corresponding reaction. Nonpolar hydrogen atoms are not shown for clarity. Variation of the reaction barrier and reaction energy when residues included in the QM_L region are added into corresponding QM_S cumulatively (from left to right) for the initial PT, the

first displacement, and the second displacement reaction, are shown in (b), (d), and (f), respectively. Residues/groups with visible catalytic effect are labeled in blue, while these with apparent anticatalytic effect are labeled in red.

4. Roles of key residues near the active site and the protein environment. Having identified a putative reaction pathway, we turn to the roles of key residues near the active site as well as the protein environment. Residues directly involved in each step were grouped into small QM models, denoted $QM_{S1,}$ QM_{S2}, and QM_{S3} for the initial PT, first displacement and second displacement steps, respectively. Calculations with these small QM models can obtain the intrinsic energetic character of the reactions without the influence of surrounding residues. The residues/groups excluded from the small QM model yet included in the QM region of the corresponding QM/MM calculations (denoted as $QM_{L1,}$ QM_{L2}, and QM_{L3} for the PT, first displacement and second displacement steps, respectively) were considered as key residues near the active site and their effects were investigated in detail. Note that (QM_S + QM_L) is the same as the QM region of the corresponding QM/MM calculations. Calculation details can be found in the computational methods section of the supporting information.

For the initial protonation reaction, the small QM model (QM_{S1}) includes only the base and sugar ring of the dA substrate and residue Glu43 (refer to Figure 5a and Figure S4). Using the QM_{S1} model, the PT process is found to be intrinsically unfavorable with a reaction barrier of 5.1 kcal/mol and a reaction energy of 13.2 kcal/mol (refer to Figure 5b). QM scanning calculations confirm the electronic energy of the PT process to be an increasing function of the reaction coordinate (refer to Figure S20). When the distance between the N7 atom of the adenine base and the proton decreases to 1.06 Å, the energy of the system has increased by 11.2 kcal/mol. This agreement indicates that the result obtained by simple single-point calculations with the small QM model is a good approximation to the corresponding nonenzymatic reaction, which is at least qualitatively correct if the reaction route is not altered. Notably, the reaction barrier decreases to -0.3 kcal/mol when all residues in QM_{L1} are included, namely Trp30, Arg31, Tyr126, Asp144, Asn146, Glu188, the 3′ side phosphate group, and four water molecules (refer to Figure 5a). Key residues driving this process include Tyr126, Asp144, Glu188, the 3′ phosphate group, and a water molecule W3 that forms an H-bond with Glu43 (labeled in blue in Figure 5b). On the other hand, the effect of residue Arg31 is found to be anticatalytic (labeled in red in Figure 4b). The effects of these residues/groups have been confirmed by the individual results (refer to Figure S21), wherein they were added to the QM_{S1} individually. It was suggested that Tyr126 facilitates the PT process through stabilization of the charge generated on Glu43,[27] which should also apply to the case of W3. Herein, the catalytic role of Tyr126 is further demonstrated through two additional scannning calculations (refer to Figure S20). When phenylalanine (Phe) was added, the scanning energy profile was virtually unchanged, yet the energies along the reaction coordinate were significantly decreased upon addition of tyrosine (Tyr). Intriguingly, we find that the effects of these residues/groups can be attributed almost entirely to the electrostatic effect of their atomic charges (refer to Figure S21). When the effect of residues outside of the QM_{L1} region is included the reaction occurs with a small barrier of 1.1 kcal/mol.

Unsurprisingly, the negatively charged residue Asp144 is found to be a perfect nucleophile to the oxocarbenium. Using the QM_{S2} model that includes only Asp144 and the sugar ring and base of the dA substrate (refer to Figure 5c and Figure S5), the reaction barrier and the reaction energy of the first displacement reaction where Asp144 acts as a nucleophile to attack the anomeric carbon of the oxocarbenium ion are calculated to be -6.8 kcal/mol and -35.6 kcal/mol, respectively (refer to Figure 5d). Key residues in QM_{L2} region that further promotes the first displacement reaction include Arg31, Tyr126, and the phosphate group (labeled in blue in Figure 5d). However, the effects of Glu43, Asn146, Glu188, and W1 water molecule are found to be anticatalytic (labeled in red in Figure 5d). Similarly, these effects can largely be attributed to their atomic charge effects except for the residue Tyr126, the effect of which comes from its interaction with Glu43 as its effect becomes negligible when it is added individually (refer to Figure S22). The collective effect of residues included in QM_{L2} not only raises the reaction barrier to a positive value of 6.9 kcal/mol, but also increases the reaction energy to -14.1 kcal/mol (refer to the energetic values of W4 in Figure 5d). Residues outside of QM_{L2} region serve to further increase the barrier and reaction energy to 12.4 kcal/mol and 3.0 kcal/mol, respectively.

Through the calculations of the QM_{S3} model that includes only Asp144, the sugar ring of dA, and the water nucleophile (refer to Figure 5e and Figure S6), the second displacement reaction is found to be intrinsically unfeasible both kinetically and thermodynamically (Figure 5f). Key vicinal residues facilitating this reaction in the QM_{L3} region include Glu43, Asn146, as well as the two water molecules W5 and W6 (labeled in blue in Figure 5f). Conversely, the phosphate group is found to hinder the reaction (labeled in red in Figure 5f). The roles of these residues/groups are confirmed by the individual results (shown in Figure S23). As before, their effects can be mostly attributed to their atomic charges (refer to Figure S23). The negatively charged residue Glu43 not only serves as a general base for accepting protons, but also facilitates the movement of the sugar ring toward the water nucleophile through long-range attractive electrostatic interactions. On the contrary, the 3′ phosphate group (PO_4^-) located on the other side of the sugar ring impedes its movement and thus hinders the second displacement reaction. Residue Asn146 stabilizes both the TS3 and Int3 species through an H-bond with Asp144, which becomes stronger when Asp144 dissociates from the ribose ring

(Figure S24). Intriguingly, the two water molecules (W5 and W6 shown in Figure 5e) not only form proton transfer paths but also greatly facilitate the second displacement reaction, which can be attributed to the dipole-dipole interaction between these two organized water molecules and groups included in the QM_{S3} model. Finally, in contrast to the cases of the initial protonation reaction and the first displacement reaction, residues outside of the QM_{L3} region are found to further facilitate the second displacement reaction.

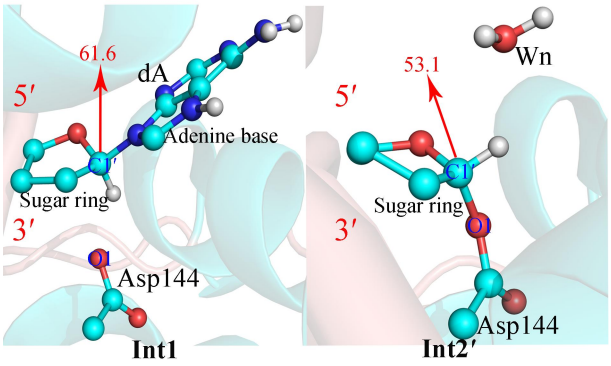


Figure 6. The internal electric field (IEF) at the active site of MutY. The IEFs were calculated at the C1′ position using the QM/MM-optimized structures of Int1 and Int2′ species. The magnitude of the IEF projected along the O1 → C1′ direction was calculated to be 61.6 MV/cm and 53.1 MV/cm for Int1 and Int2′, respectively.

Notably, the first displacement reaction of Asp144 as a nucleophile is inherently much more favorable than the second displacement reaction of an unactivated water molecule as a nucleophile (refer to the energetic values of QM_{S2} and QM_{S3} in Figure 5d and Figure 5f, respectively). However, within the protein environment, the barrier of the second displacement reaction becomes comparable to that of the first. Therefore, the effect of protein environment is different for these two displacement reactions, which facilitates the second yet hinders the first. An intriguing question is how the enzyme achieves this function. In view of the crucial role of the IEF found previously in our previous study of UDG,[31] we calculated the IEF at the position of the anomeric carbon in MutY. Interestingly, the IEF was found to point to the 5′ side in MutY (refer to Figure 6), which is opposite to the direction of the IEF found in UDG.[31] According to Coulomb's law, such an IEF will facilitate the movement of the oxocarbenium cation toward the 5′ direction while hindering its inverse movement toward the 3′ direction. Recall that the main structural change in the first displacement reaction is the movement of the oxocarbenium ion to the negatively charged residue Asp144 (3′ direction), while the second one mainly involves its movement to the water nucleophile Wn (5′ direction). As such, the IEF promotes the second displacement reaction while hindering the first one, which is further confirmed by oriented external electric field calculations (refer to Figure S25 and Figure S26). Additionally, the presence of such an IEF also provides a rational for the necessity of the initial protonation of the adenine base. Without protonation the leaving base is negatively charged, and so any motion in the 5′ direction during the first displacement reaction would be impeded by the IEF. Overall, the IEF acts as an important modulator in MutY, impeding the easy step (the first displacement reaction) while promoting the difficult step (the second displacement reaction), coupling them together so that both reactions can occur. The role of the IEF revealed in MutY may have general implications on how internal/external electric fields can be used to control chemical reactions.[96-110]

Conclusion

In summary, our detailed computational study confirms the formation of a covalent glycosyl-enzyme intermediate during the base excision process of MutY, and thus provides strong support for the double-displacement reaction mechanism. Roles of key residues near the active site and the protein environment have been discussed in detail. Unsuprisingly, aspartic acid is found to be a quite good nucleophile for oxocarbenium ion, yet a water molecule is not, which is usually activated by a proton acceptor in other DNA glycosylases, such as UDG. Intriguingly, we find that a strong IEF, pointing toward 5′ direction at the active site of the enzyme, enables the attack of an unactivated water nucleophile from the 5′ side to complete the otherwise unfeasible, second displacement reaction, at the expense of elevating the reaction energy of the first one. Accordingly, we propose that the base excision mechanism catalyzed by a strong preorganized IEF also applies to the case of thymine DNA glycosylase (TDG), in which there is no nearby proton accepter to activate water molecules. Additionally, the relatively low barriers of the steps involved in the excision process indicates that either the recognition of OG:A mispair or the alignment of the substrate in the active site of the enzyme is more likely the rate-limiting step of the overall catalytic cycle of MutY. These findings increase our understanding of the strategies DNA glycosylases use to repair lesions and how electric fields can be used to modulate chemical reactions.

ASSOCIATED CONTENT

Supporting Information. This material is available free of charge via the Internet at http://pubs.acs.org.

Figures S1-S26, Table S1-S2, and the cartesian coordinates of the QM region for all computed species (PDF)

AUTHOR INFORMATION

Corresponding Author

*Zhanfeng Wang, zfwang@bnu.edu.cn
*Fangfu Ye, fye@iphy.ac.cn

ORCID

Zhanfeng Wang: orcid.org/0000-0001-6722-2298
Fangfu Ye: orcid.org/0000-0003-1279-105X
James D. Farrell: orcid.org/0000-0002-1893-1773

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (No. 22203007 to Z.W., No. 12090054 to F.Y.), National Key Research and Development Program of China (No. 2020YFA0908200 to F.Y.), the Strategic Priority Research Program of Chinese Academy of Sciences (No. XDB33030300 to F.Y.), the Start-up Founding from Beijing Normal University (No. 310432104 to Z.W.). Dr. Z. Wang thanks supports from the Interdisciplinary Intelligence Super Computer Center of Beijing Normal University at Zhuhai.

ABBREVIATIONS

ROS, reactive oxygen species; OG, 8-oxo-7,8-dihydroguanine; MAP, MUTYH-associated polyposis; MD, molecular dynamics; QM/MM, quantum mechanical/molecular mechanical; KIE, kinetic isotope effect; LRC, lesion recognition complex; RC, reactant complex; TS, transition state; IC, intermediate complex; TSAC, transition state analog complex; IEF, internal electric field; PDB, protein data bank; UDG, uracil-DNA glycosylase; FLRC, fluorinated lesion recognition complex; TSAC, transition state analog complex; GAFF, general AMBER force field; RESP, restrained electrostatic potential; COM, center of mass; CV, collective variable; PMF, potential of mean force; WHAM, weighted histogram analysis method; G16, Gaussian 16; IEF, internal electric field; PT, proton transfer; TDG, thymine DNA glycosylase.

REFERENCES

1. Hoeijmakers, J. H. J., Genome maintenance mechanisms for preventing cancer. *Nature* **2001,** *411* (6835), 366-374.
2. Steenken, S.; Jovanovic, S. V., How easily oxidizable is DNA? One-electron reduction potentials of adenosine and guanosine radicals in aqueous solution. *J Am Chem Soc* **1997,** *119* (3), 617-618.
3. Au, K. G.; Cabrera, M.; Miller, J. H.; Modrich, P., Escherichia-Coli Muty Gene-Product Is Required for Specific a-G-]C.G Mismatch Correction. *P Natl Acad Sci USA* **1988,** *85* (23), 9163-9166.
4. Nghiem, Y.; Cabrera, M.; Cupples, C. G.; Miller, J. H., The Muty Gene - a Mutator Locus in Escherichia-Coli That Generates G.C-]T.A Transversions. *P Natl Acad Sci USA* **1988,** *85* (8), 2709-2713.
5. Michaels, M. L.; Pham, L.; Nghiem, Y.; Cruz, C.; Miller, J. H., Muty, an Adenine Glycosylase Active on G-a Mispairs, Has Homology to Endonuclease-Iii. *Nucleic Acids Res* **1990,** *18* (13), 3841-3845.
6. Michaels, M. L.; Cruz, C.; Grollman, A. P.; Miller, J. H., Evidence That Muty and Mutm Combine to Prevent Mutations by an Oxidatively Damaged Form of Guanine in DNA. *P Natl Acad Sci USA* **1992,** *89* (15), 7022-7025.
7. Michaels, M. L.; Miller, J. H., The Go System Protects Organisms from the Mutagenic Effect of the Spontaneous Lesion 8-Hydroxyguanine (7,8-Dihydro-8-Oxoguanine). *J Bacteriol* **1992,** *174* (20), 6321-6325.
8. Michaels, M. L.; Tchou, J.; Grollman, A. P.; Miller, J. H., A Repair System for 8-Oxo-7,8-Dihydrodeoxyguanine. *Biochemistry-Us* **1992,** *31* (45), 10964-10968.
9. Al-Tassan, N.; Chmiel, N. H.; Maynard, J.; Fleming, N.; Livingston, A. L.; Williams, G. T.; Hodges, A. K.; Davies, D. R.; David, S. S.; Sampson, J. R.; Cheadle, J. R., Inherited variants of MYH associated with somatic G : C -> T : A mutations in colorectal tumors. *Nat Genet* **2002,** *30* (2), 227-232.
10. David, S. S.; O'Shea, V. L.; Kundu, S., Base-excision repair of oxidative DNA damage. *Nature* **2007,** *447* (7147), 941-950.
11. Wallace, S. S.; Murphy, D. L.; Sweasy, J. B., Base excision repair and cancer. *Cancer Lett* **2012,** *327* (1-2), 73-89.
12. Mazzei, F.; Viel, A.; Bignami, M., Role of MUTYH in human cancer. *Mutat Res-Fund Mol M* **2013,** *743*, 33-43.
13. David, S. S.; Wiliams, S. D., Chemistry of glycosylases and endonucleases involved in base-excision repair. *Chem Rev* **1998,** *98* (3), 1221-1261.
14. Chmiel, N. H.; Livingston, A. L.; David, S. S., Insight into the functional consequences of inherited variants of the hMYH adenine glycosylase associated with colorectal cancer: Complementation assays with hMYH variants, and pre-steady-state kinetics of the corresponding mutated E-coli enzymes. *J Mol Biol* **2003,** *327* (2), 431-443.
15. Raetz, A. G.; David, S. S., When you're strange: Unusual features of the MUTYH glycosylase and implications in cancer. *DNA Repair* **2019,** *80*, 16-25.
16. Fromme, J. C.; Banerjee, A.; Huang, S. J.; Verdine, G. L., Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. *Nature* **2004,** *427* (6975), 652-656.
17. Lee, S.; Verdine, G. L., Atomic substitution reveals the structural basis for substrate adenine recognition and removal by adenine DNA glycosylase. *P Natl Acad Sci USA* **2009,** *106* (44), 18497-18502.
18. Woods, R. D.; O'Shea, V. L.; Chu, A.; Cao, S.; Richards, J. L.; Horvath, M. P.; David, S. S., Structure and stereochemistry of the base excision repair glycosylase MutY reveal a mechanism similar to retaining glycosidases. *Nucleic Acids Res* **2016,** *44* (2), 801-810.
19. Guan, Y.; Manuel, R. C.; Arvai, A. S.; Parikh, S. S.; Mol, C. D.; Miller, J. H.; Lloyd, S.; Tainer, J. A., MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily. *Nat Struct Biol* **1998,** *5* (12), 1058-1064.
20. Manuel, R. C.; Hitomi, K.; Arvai, A. S.; House, P. G.; Kurtz, A. J.; Dodson, M. L.; McCullough, A. K.; Tainer, J. A.; Lloyd, R. S., Reaction intermediates in the catalytic mechanism of Escherichia coli MutY DNA glycosylase. *J Biol Chem* **2004,** *279* (45), 46930-46939.
21. McCann, J. A. B.; Berti, P. J., Transition-state analysis of the DNA repair enzyme MutY. *J Am Chem Soc* **2008,** *130* (17), 5789-5797.
22. Brinkmeyer, M. K.; Pope, M. A.; David, S. S., Catalytic Contributions of Key Residues in the Adenine Glycosylase MutY Revealed by pH-dependent Kinetics and Cellular Repair Assays. *Chem Biol* **2012,** *19* (2), 276-286.
23. Wright, P. M.; Yu, J. A.; Cillo, J.; Lu, A. L., The active site of the Escherichia coli MutY DNA adenine glycosylase. *J Biol Chem* **1999,** *274* (41), 29011-29018.
24. 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* (5).

25. Tiwari, S.; Agnihotri, N.; Mishra, P. C., Quantum Theoretical Study of Cleavage of the Glycosidic Bond of 2 '-Deoxyadenosine: Base Excision-Repair Mechanism of DNA by MutY. *J Phys Chem B* **2011,** *115* (12), 3200-3207.
26. Brunk, E.; Arey, J. S.; Rothlisberger, U., Role of Environment for Catalysis of the DNA Repair Enzyme MutY. *J Am Chem Soc* **2012,** *134* (20), 8608-8616.
27. Kellie, J. L.; Wilson, K. A.; Wetmore, S. D., Standard Role for a Conserved Aspartate or More Direct Involvement in Deglycosylation? An ONIOM and MD Investigation of Adenine-DNA Glycosylase. *Biochemistry-Us* **2013,** *52* (48), 8753-8765.
28. Michelson, A. Z.; Rozenberg, A.; Tian, Y.; Sun, X. J.; Davis, J.; Francis, A. W.; O'Shea, V. L.; Halasyam, M.; Manlove, A. H.; David, S. S.; Lee, J. K., Gas-Phase Studies of Substrates for the DNA Mismatch Repair Enzyme MutY. *J Am Chem Soc* **2012,** *134* (48), 19839-19850.
29. Russelburg, L. P.; Murray, V. L. O.; Demir, M.; Knutsen, K. R.; Sehgal, S. L.; Cao, S.; David, S. S.; Horvat, M. P., Structural Basis for Finding OG Lesions and Avoiding Undamaged G by the DNA Glycosylase MutY. *Acs Chem Biol* **2020,** *15* (1), 93-102.
30. Demir, M.; Russelburg, L. P.; Lin, W. J.; Trasvina-Arenas, C. H.; Huang, B. L.; Yuen, P. K.; Horvath, M. P.; David, S. S., Structural snapshots of base excision by the cancer-associated variant MutY N146S reveal a retaining mechanism. *Nucleic Acids Res* **2023,** *51* (3), 1034-1049.
31. Diao, W. W.; Yan, S. H.; Farrell, J. D.; Wang, B. J.; Ye, F. F.; Wang, Z. F., Preorganized Internal Electric Field Powers Catalysis in the Active Site of Uracil-DNA Glycosylase. *Acs Catal* **2022,** *12* (20), 12488-12499.
32. Li, P. F.; Merz, K. M., MCPB.py: A Python Based Metal Center Parameter Builder. *J Chem Inf Model* **2016,** *56* (4), 599-604.
33. Li, P. F.; Merz, K. M., Metal Ion Modeling Using Classical Mechanics. *Chem Rev* **2017,** *117* (3), 1564-1686.
34. Wang, J. M.; 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* (9), 1157-1174.
35. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A., A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *J Phys Chem-Us* **1993,** *97* (40), 10269-10280.
36. 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* (8), 3696-3713.
37. Sondergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H., Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK(a) Values. *J Chem Theory Comput* **2011,** *7* (7), 2284-2295.
38. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* **1983,** *79* (2), 926-935.
39. Izaguirre, J. A.; Catarello, D. P.; Wozniak, J. M.; Skeel, R. D., Langevin stabilization of molecular dynamics. *J Chem Phys* **2001,** *114* (5), 2090-2098.
40. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R., Molecular-Dynamics with Coupling to an External Bath. *J Chem Phys* **1984,** *81* (8), 3684-3690.
41. Krautler, V.; Van Gunsteren, W. F.; Hunenberger, P. H., A fast SHAKE: Algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J Comput Chem* **2001,** *22* (5), 501-508.
42. 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* (12), 10089-10092.
43. Gotz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J Chem Theory Comput* **2012,** *8* (5), 1542-1555.
44. Salomon-Ferrer, R.; Gotz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J Chem Theory Comput* **2013,** *9* (9), 3878-3888.
45. Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Huang, Y.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *AMBER*, 2018; University of California: San Francisco: 2018.
46. Torrie, G. M.; Valleau, J. P., Nonphysical Sampling Distributions in Mente Carlo Free-energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977,** *23*, 187-199.
47. Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M., The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. *J Comput Chem* **1992,** *13* (8), 1011-1021.
48. Souaille, M.; Roux, B., Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput Phys Commun* **2001,** *135* (1), 40-57.
49. Wang, Z. F.; Shaik, S.; Wang, B. J., Conformational Motion of Ferredoxin Enables Efficient Electron Transfer to Heme in the Full-Length P450(TT). *J Am Chem Soc* **2021,** *143* (2), 1005-1016.
50. Wang, Z. F.; Liu, Y. R.; Wang, P. Y.; Xie, P., An intermediate state of T7 RNA polymerase provides another pathway of nucleotide selection. *Chinese Phys B* **2017,** *26* (10).
51. Metz, S.; Kastner, J.; Sokol, A. A.; Keal, T. W.; Sherwood, P., ChemShell-a modular software package for QM/MM simulations. *Wires Comput Mol Sci* **2014,** *4* (2), 101-110.
52. Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C., Electronic-Structure Calculations on Workstation Computers - the Program System Turbomole. *Chem Phys Lett* **1989,** *162* (3), 165-169.
53. Smith, W., DL_POLY-applications to molecular simulation II. *Mol Simulat* **2006,** *32* (12-13), 933-933.
54. Bakowies, D.; Thiel, W., Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J Phys Chem-Us* **1996,** *100* (25), 10580-10594.
55. Lee, C. T.; Yang, W. T.; Parr, R. G., Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Phys Rev B* **1988,** *37* (2), 785-789.
56. Becke, A. D., Density-Functional Thermochemistry .2. The Effect of the Perdew-Wang Generalized-Gradient Correlation Correction. *J Chem Phys* **1992,** *97* (12), 9173-9177.
57. Becke, A. D., Density-Functional Thermochemistry .3. The Role of Exact Exchange. *J Chem Phys* **1993,** *98* (7), 5648-5652.
58. Di Laudo, M.; Whittleton, S. R.; Wetmore, S. D., Effects of hydrogen bonding on the acidity of uracil. *J Phys Chem A* **2003,** *107* (48), 10406-10413.

59. Whittleton, S. R.; Hunter, K. C.; Wetmore, S. D., Effects of hydrogen bonding on the acidity of uracil derivatives. *J Phys Chem A* **2004,** *108* (38), 7709-7718.
60. Millen, A. L.; Archibald, L. A. B.; Hunter, K. C.; Wetmore, S. D., A kinetic and thermodynamic study of the glycosidic bond cleavage in deoxyuridine. *J Phys Chem B* **2007,** *111* (14), 3800-3812.
61. Lenz, S. A. P.; Kellie, J. L.; Wetmore, S. D., Glycosidic Bond Cleavage in Deoxynucleotides: Effects of Solvent and the DNA Phosphate Backbone in the Computational Model. *J Phys Chem B* **2012,** *116* (49), 14275-14284.
62. Lenz, S. A. P.; Kellie, J. L.; Wetmore, S. D., Glycosidic Bond Cleavage in DNA Nucleosides: Effect of Nucleobase Damage and Activation on the Mechanism and Barrier. *J Phys Chem B* **2015,** *119* (51), 15601-15612.
63. Lenz, S. A. P.; Kohout, J. D.; Wetmore, S. D., Hydrolytic Glycosidic Bond Cleavage in RNA Nucleosides: Effects of the 2 '-Hydroxy Group and Acid-Base Catalysis. *J Phys Chem B* **2016,** *120* (50), 12795-12806.
64. 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* (8), 4344-4350.
65. Perdew, J. P., Density-Functional Approximation for the Correlation-Energy of the Inhomogeneous Electron-Gas. *Phys Rev B* **1986,** *33* (12), 8822-8824.
66. Becke, A. D., Density-Functional Exchange-Energy Approximation with Correct Asymptotic-Behavior. *Phys Rev A* **1988,** *38* (6), 3098-3100.
67. Zhao, Y.; Truhlar, D. G., A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J Chem Phys* **2006,** *125* (19), 194101.
68. Zhao, Y.; Truhlar, D. G., The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor Chem Acc* **2008,** *120* (1-3), 215-241.
69. Grimme, S., Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J Comput Chem* **2006,** *27* (15), 1787-1799.
70. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J Chem Phys* **2010,** *132* (15), 154104.
71. Grimme, S.; Ehrlich, S.; Goerigk, L., Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J Comput Chem* **2011,** *32* (7), 1456-1465.
72. Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W., P450 Enzymes: Their Structure, Reactivity, and Selectivity-Modeled by QM/MM Calculations. *Chem Rev* **2010,** *110* (2), 949-1017.
73. Kuznetsov, N. A.; Kuznetsova, A. A.; Vorobjev, Y. N.; Krasnoperov, L. N.; Fedorova, O. S., Thermodynamics of the DNA Damage Repair Steps of Human 8-Oxoguanine DNA Glycosylase. *Plos One* **2014,** *9* (6), e98495.
74. Chen, Z. Q.; Zhang, C. H.; Xue, Y., Theoretical Studies on the Thermodynamics and Kinetics of the N-Glycosidic Bond Cleavage in Deoxythymidine Glycol. *J Phys Chem B* **2009,** *113* (30), 10409-10420.
75. Lenz, S. A. P.; Wetmore, S. D., QM/MM Study of the Reaction Catalyzed by Alkyladenine DNA Glycosylase: Examination of the Substrate Specificity of a DNA Repair Enzyme. *J Phys Chem B* **2017,** *121* (49), 11096-11108.
76. Sebera, J.; Hattori, Y.; Sato, D.; Reha, D.; Nencka, R.; Kohno, T.; Kojima, C.; Tanaka, Y.; Sychrovsky, V., The mechanism of the glycosylase reaction with hOGG1 base-excision repair enzyme: concerted effect of Lys249 and Asp268 during excision of 8-oxoguanine. *Nucleic Acids Res* **2017,** *45* (9), 5231-5242.
77. Laio, A.; Parrinello, M., Escaping free-energy minima. *P Natl Acad Sci USA* **2002,** *99* (20), 12562-12566.
78. Barducci, A.; Bussi, G.; Parrinello, M., Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys Rev Lett* **2008,** *100* (2).
79. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J., Gaussian 16. *Gaussian 16, Revision B.01*, Wallingford CT, 2016.
80. Chen, K. X.; Li, W. L.; Head-Gordon, T., Linear Combination of Atomic Dipoles to Calculate the Bond and Molecular Dipole Moments of Molecules and Molecular Liquids. *J Phys Chem Lett* **2021,** *12* (51), 12360-12369.
81. Wang, L.; Fried, S. D.; Markland, T. E., Proton Network Flexibility Enables Robustness and Large Electric Fields in the Ketosteroid Isomerase Active Site. *J Phys Chem B* **2017,** *121* (42), 9807-9815.
82. *CP2K version 2022.2, CP2K is freely available from https://www.cp2k.org/.*
83. VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J., QUICKSTEP: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput Phys Commun* **2005,** *167* (2), 103-128.
84. Laio, A.; VandeVondele, J.; Rothlisberger, U., A Hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations. *J Chem Phys* **2002,** *116* (16), 6941-6947.
85. Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M., An efficient real space multigrid OM/MM electrostatic coupling. *J Chem Theory Comput* **2005,** *1* (6), 1176-1184.
86. VandeVondele, J.; Hutter, J., Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J Chem Phys* **2007,** *127* (11).
87. Goedecker, S.; Teter, M.; Hutter, J., Separable dual-space Gaussian pseudopotentials. *Phys Rev B* **1996,** *54* (3), 1703-1710.
88. Hartwigsen, C.; Goedecker, S.; Hutter, J., Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys Rev B* **1998,** *58* (7), 3641-3662.
89. Guidon, M.; Hutter, J.; VandeVondele, J., Auxiliary Density Matrix Methods for Hartree-Fock Exchange Calculations. *J Chem Theory Comput* **2010,** *6* (8), 2348-2364.
90. Dinner, A. R.; Blackburn, G. M.; Karplus, M., Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature* **2001,** *413* (6857), 752-755.

91. Przybylski, J. L.; Wetmore, S. D., A QM/QM Investigation of the hUNG2 Reaction Surface: The Untold Tale of a Catalytic Residue. *Biochemistry-Us* **2011,** *50* (19), 4218-4227.
92. Francis, A. W.; Helquist, S. A.; Kool, E. T.; David, S. S., Probing the requirements for recognition and catalysis in fpg and MutY with nonpolar adenine isosteres. *J Am Chem Soc* **2003,** *125* (52), 16235-16242.
93. Chepanoske, C. L.; Lukianova, O. A.; Lombard, M.; Golinelli-Cohen, M. P.; David, S. S., A residue in MutY important for catalysis identified by photocross-linking and mass spectrometry. *Biochemistry-Us* **2004,** *43* (3), 651-662.
94. Schroeder, G. K.; Wolfenden, R., Rates of spontaneous disintegration of DNA and the rate enhancements produced by DNA glycosylases and deaminases. *Biochemistry-Us* **2007,** *46* (47), 13638-13647.
95. Majumdar, C.; McKibbin, P. L.; Krajewski, A. E.; Manlove, A. H.; Lee, J. K.; David, S. S., Unique Hydrogen Bonding of Adenine with the Oxidatively Damaged Base 8-Oxoguanine Enables Specific Recognition and Repair by DNA Glycosylase MutY. *J Am Chem Soc* **2020,** *142* (48), 20340-20350.
96. Warshel, A., Energetics of Enzyme Catalysis. *P Natl Acad Sci USA* **1978,** *75* (11), 5250-5254.
97. Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H. B.; Olsson, M. H. M., Electrostatic basis for enzyme catalysis. *Chem Rev* **2006,** *106* (8), 3210-3235.
98. Shaik, S.; Mandal, D.; Ramanan, R., Oriented electric fields as future smart reagents in chemistry. *Nat Chem* **2016,** *8* (12), 1091-1098.
99. Fried, S. D.; Boxer, S. G., Electric Fields and Enzyme Catalysis. *Annu Rev Biochem* **2017,** *86*, 387-415.
100. Ciampi, S.; Darwish, N.; Aitken, H. M.; Diez-Perez, I.; Coote, M. L., Harnessing electrostatic catalysis in single molecule, electrochemical and chemical systems: a rapidly growing experimental tool box. *Chem Soc Rev* **2018,** *47* (14), 5146-5164.
101. Shaik, S.; Ramanan, R.; Danovich, D.; Mandal, D., Structure and reactivity/selectivity control by oriented-external electric fields. *Chem Soc Rev* **2018,** *47* (14), 5125-5145.
102. Wang, Z. F.; Danovich, D.; Ramanan, R.; Shaik, S., Oriented-External Electric Fields Create Absolute Enantioselectivity in Diels-Alder Reactions: Importance of the Molecular Dipole Moment. *J Am Chem Soc* **2018,** *140* (41), 13350-13359.
103. Welborn, V. V.; Pestana, L. R.; Head-Gordon, T., Computational optimization of electric fields for better catalysis design. *Nat Catal* **2018,** *1* (9), 649-655.
104. Stuyver, T.; Danovich, D.; Joy, J.; Shaik, S., External electric field effects on chemical structure and reactivity. *Wires Comput Mol Sci* **2020,** *10* (2).
105. Stuyver, T.; Ramanan, R.; Mallick, D.; Shaik, S., Oriented (Local) Electric Fields Drive the Millionfold Enhancement of the H-Abstraction Catalysis Observed for Synthetic Metalloenzyme Analogues. *Angew Chem Int Edit* **2020,** *59* (20), 7915-7920.
106. Leonard, N. G.; Dhaoui, R.; Chantarojsiri, T.; Yang, J. Y., Electric Fields in Catalysis: From Enzymes to Molecular Catalysts. *Acs Catal* **2021,** *11* (17), 10923-10932.
107. Cassone, G.; Sponer, J.; Saija, F., Ab Initio Molecular Dynamics Studies of the Electric-Field-Induced Catalytic Effects on Liquids. *Top Catal* **2022,** *65* (1-4), 40-58.
108. de Visser, S. P.; Mukherjee, G.; Ali, H. S.; Sastri, C. V., Local Charge Distributions, Electric Dipole Moments, and Local Electric Fields Influence Reactivity Patterns and Guide Regioselectivities in alpha-Ketoglutarate-Dependent Non-heme Iron Dioxygenases. *Accounts Chem Res* **2022,** *55* (1), 65-74.
109. Hennefarth, M. R.; Alexandrova, A. N., Advances in optimizing enzyme electrostatic preorganization. *Curr Opin Struc Biol* **2022,** *72*, 1-8.
110. Wang, Z. F.; Wang, B. J.; Diao, W. W.; Wu, P.; Li, J. F.; Fu, Y. Z.; Guo, Z. Y.; Cao, Z. X.; Shaik, S., How the Conformational Movement of the Substrate Drives the Regioselective C-N Bond Formation in P450 TleB: Insights from Molecular Dynamics Simulations and Quantum Mechanical/ Molecular Mechanical Calculations. *J Am Chem Soc* **2023,** *145* (13), 7252-7267.

Table of Contents