How Human TET2 Enzyme Catalyzes the Oxidation of Unnatural Cytosine Modifications in Double-Stranded DNA

Sodiq O. Waheed,[&] Ann Varghese,[&] Shobhit S. Chaturvedi,[&] Tatyana G. Karabencheva-Christova,[&#] and Christo Z. Christov[&#]

[&]Department of Chemistry, Michigan Technological University, Houghton, Michigan 49931, United States.

[#]**Corresponding co-authors:** christov@mtu.edu, tatyanak@mtu.edu

Abstract

Methylation of cytosine bases is strongly linked to gene expression, imprinting, aging, and carcinogenesis. The Ten-eleven translocation (TET) family of enzymes, which are Fe(II)/2-oxoglutarate (2OG)-dependent enzymes, employ Fe(IV)=O species to dealkylate the lesioned bases to an unmodified cytosine. Recently, it has been shown that the TET2 enzyme can catalyze promiscuously DNA substrates containing unnatural alkylated cytosine. Such unnatural substrates of TET can be used as direct probes for measuring the TET activity or capturing TET from cellular samples. Herein, we studied the catalytic mechanisms during the oxidation of the unnatural C5-position modifications (5-ethylcytosine (5eC), 5-vinylcytosine (5vC) and 5-ethynylcytosine (5eyC)) and the demethylation of N4-methylated lesions (4-methylcytosine (4mC) and 4,4-dimethylcytosine(4dmC)) of the cytosine base by the TET2 enzyme using molecular dynamics (MD) and combined quantum mechanics and molecular mechanics (QM/MM) computational approaches. The results reveal that the chemical nature of the alkylation of the double-stranded (ds) DNA substrates induces distinct changes in the interactions in the binding site, the second coordination sphere, and long-range correlated motions of the ES complexes. The rate-determining hydrogen atom transfer (HAT) is faster in N4-methyl substituent substrates than in the C5-alkylations. Importantly, the calculations show the preference of hydroxylation over desaturation in both 5eC and 5vC substrates. The studies elucidate the post-hydroxylation rearrangements of the hydroxylated intermediates of 5eyC and 5vC to ketene and 5-formylmethylcytosine (5fmC), respectively, and hydrolysis of hemiaminal intermediate of 4mC to formaldehyde and unmodified cytosine proceed exclusively in aqueous solution outside of the enzyme environment. Overall, the studies show that the chemical nature of the unnatural

alkylated cytosine substrates exercises distinct effects on the binding interactions, reaction mechanism, and dynamics of TET2.

Keywords: DNA repair, demethylation, TET2 enzyme, non-heme enzymes, QM/MM calculations, reaction mechanism, molecular dynamics

1.0 Introduction

Ten-eleven translocation (TET) enzymes are important epigenetic modulators with important roles in various physiological and pathological processes, especially in embryogenesis, cell differentiation, carcinogenesis, and neural development.[1–6] Three mammalian family members of TET proteins have been identified, namely, TET1, TET2, and TET3.[2,6,7] All three TET proteins contain a similar catalytic C-terminal domain which harbors the conserved double-stranded β-helix (DSBH) fold and a cysteine-rich (Cys) region.[1,3,4] TET enzymes belong to non-heme Fe(II) and 2-oxoglutarate (2OG) dependent oxygenases that convert 5-methyl-cytosine (5mC) to 5-hydroxymethylcytosine (5hmC) and subsequently convert 5hmC into 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC).[3,5,7,8] These oxidized 5mC bases act as intermediates in active DNA demethylation pathways, involving their conversion to unmodified cytosine via either active excision or passive replication-dependent processes.[7,9,10] In addition, the different oxidized intermediates exercise their own epigenetic rules.[3,6,8] Studies have revealed that the catalytic domains in TET enzyme preferentially bind to cytosines on CpG islands without interacting with adjacent bases.[1,7] TET loss-of-function is firmly associated with cancer; TET2 loss-of-function mutations are frequently observed in myelodysplastic syndromes, myeloid malignancies, and peripheral T-cell lymphomas (PTCLs), which are a heterogeneous and poorly understood group of aggressive non-Hodgkin lymphomas that are resistant to conventional therapies.[1,6]

The crystal structure of the human TET2-double-stranded (ds) DNA complex [Figure 1] reported by Hu et al. revealed that the structure contains loop 1, loop 2, GS linker, and three Zn finger regions of which two of the Zn(II) regions (Zn2 and Zn3 finger regions) bring the DSBH and the Cys-rich domains together to facilitate a compact catalytic domain formation,[7] unlike other Fe(II)/2OG DNA modifying enzymes, AlkB and its human homologues (AlkBH1-8, AlkBH9/FTO), which lack Zn(II) finger regions.[8,11a,b] Recent computational studies on TET2 and its various

mutants have shown the importance of these structural elements' correlated motions on the interaction of DNA with the TET2 enzyme and the overall structural formation.[12]

The demethylation process in 2OG-dependent oxygenases involves the activation of dioxygen followed by substrate oxidation.[8,11a,13–15a] The dioxygen activation mechanism, which passes through the Fe(II)peroxysuccinate intermediate[16–18] starts with the binding of molecular oxygen to the Fe(II) center, leading to the formation of highly reactive Fe(IV)=O species.[17,19–21] Various experimental and computational studies have addressed the mechanisms of dioxygen activation by non-heme Fe(II) enzymes.[18,20,21] After the formation of the Fe(IV)=O intermediate, it abstracts a hydrogen atom from the substrate to generate an Fe(III)-hydroxyl intermediate which undergoes a hydroxyl rebound reaction with the substrate radical carbon to complete the hydroxylation.[17,19-21]

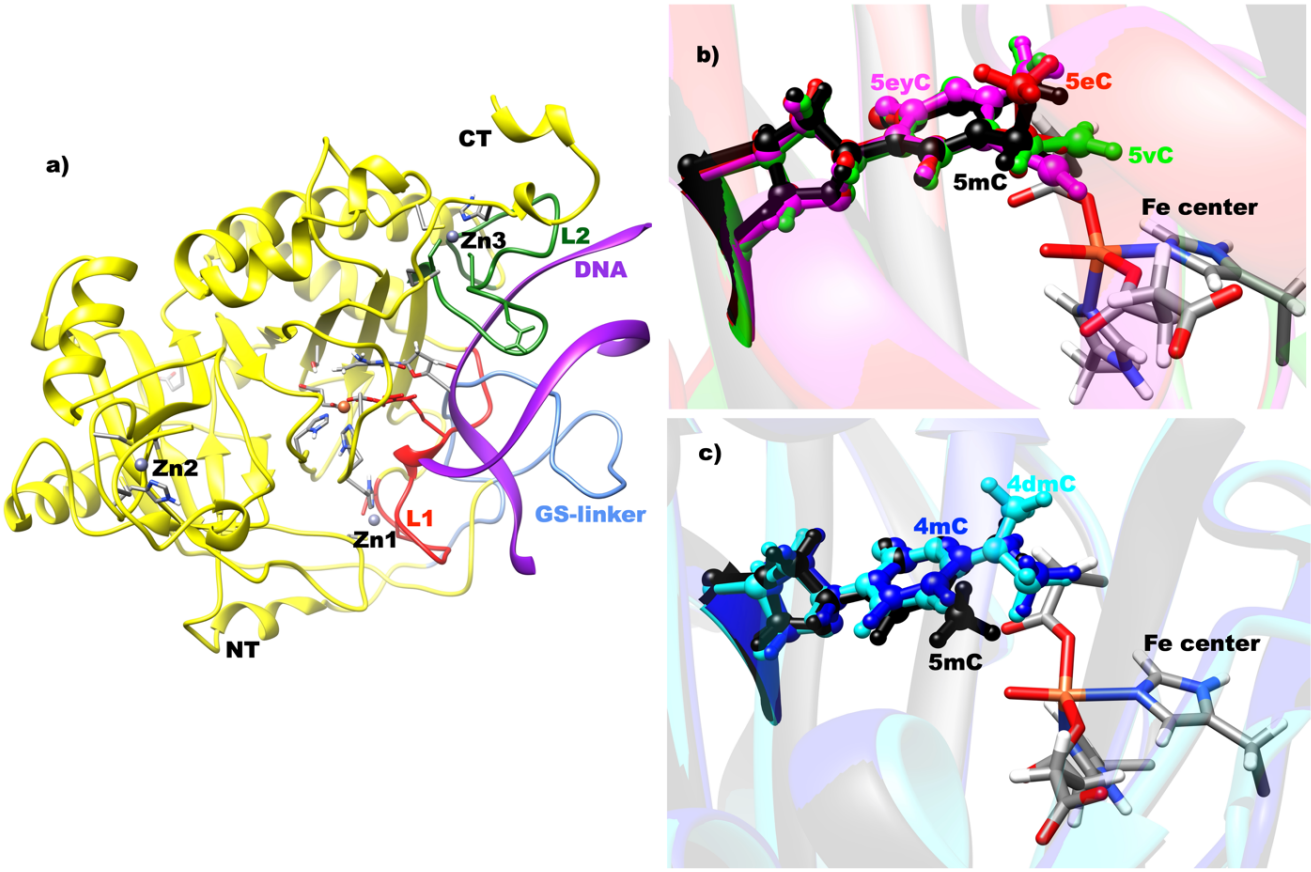


Figure 1: (a) Protein structure for the human TET2-dsDNA complex derived from the average molecular dynamics (MD) structure of ferryl intermediate simulations with the 5eC substrate. (b) Overlaid structure of the C5-alkylated 5eC, 5vC, and 5eyC unnatural substrates with the natural

5mC substrate. (c) Overlaid structure of the N4-methylated 4mC, 4dmC substrates with the natural C5-methylated 5mC substrate. In both (b) and (c), the Fe center depicts the one of 5mC dsDNA-bound TET2.

A recent work by Islam and co-workers has reported that TET2 can oxidize 5-ethylcytosine (5eC) to 5-hydroxyethylcytosine (5heC) in a site- and degree-specific manner to offer the mono-hydroxylated product via benzylic C—H activation.[22] The study revealed that C—H hydroxylation in both the natural 5mC substrate and the unnatural 5eC showed comparable activity, and further oxidation of 5heC was nearly undetectable, implying that 5heC cannot undergo successive oxidations as in 5hmC generated after the first oxidation of the 5mC natural substrate of TET2.[22] The hydroxylated product of 5eC acts as a docking site for the protein implicated in transcription and has the potential to control chromatin-dependent processes.[22]

Another study by Kohli and co-workers has demonstrated that TET enzymes can also hydroxylate unnatural nucleobases at the 5-position on the cytosine base, 5-vinylcytosine (5vC), and 5-ethynylcytosine (5eyC) aside from 5mC in DNA in a similar manner as another Fe(II) and 2OG dependent dioxygenase,[9] thymine hydroxylase, which acts on the free thymine nucleobase.[23,24] TET catalyzes the oxidation of 5vC to produce 5-formylmethylcytosine (5fmC) as the major product of the reaction. In contrast to 5vC, the 5eyC DNA substrate serves as an alternative probe, functioning as an activity-based cross-linker of TET enzymes.[9,25] Oxidation of the 5eyC substrate by TET produces a high-energy ketene intermediate, which reacts and forms a covalent cross-link with the TET enzyme. These unnatural substrates have been reported to be used to detect and quantify the TET activity either in vitro or in cell extracts.[9,25]

TET enzymes are different from other DNA repair enzymes because the alkyl group to be oxidized is not linked to a heteroatom,[7,8,10] but the C5-position carbon of the cytosine base. However, the latest studies by Kohli and co-workers show that TET enzymes are also proficient as direct N-demethylases of the cytosine bases in a similar manner as the AlkB-type of other DNA demethylases. They reported that N-demethylase activity can be observed on substrates that lack a 5-methyl group, and as such, TET enzymes can be similarly proficient in either oxidation of

5mC or demethylation of 4-methylcytosine (4mC) and 4,4-dimethylcytosine (4dmC) substrates, which have potential biological and mechanistic implications.[10]

Although the experimental studies discovered the catalytic promiscuity of TET2 with unnatural substrates, knowledge about the respective catalytic mechanisms and the potential differences from the natural 5mC substrates is still missing. To complete the missing information, we performed molecular dynamics (MD) simulations and combined quantum mechanics and molecular mechanics (QM/MM) studies on the catalytic mechanisms of TET2 with these new substrates. Also, the accommodation of these substrates with unnatural alkylations in the active site of TET2 bound to dsDNA would involve conformational changes; it is important to explore how these changes influence the active site and second sphere residues interactions, long-range correlated motions in the TET2 enzyme. Hence, this current study aims at completing this missing knowledge in comparison to our previously reported study on the 5mC natural substrate of the TET2 enzyme.[12] Furthermore, the study aims to explore the mechanisms of the post-hydroxylation reactions for 5vC and 5eyC that are currently missing as well if these reactions take place within the enzyme environment or in the water solvent. In addition, this work aims to reveal that the role of the second coordination sphere (SCS) and long-range correlated motions in the catalysis of TET2 with unnatural substrates might be important; however, it is completely unexplored. The study also delineates the role of the SCS residues in stabilizing the TSs with respect to the RCs in the catalyzed reactions.

2.0 Computational Methods

2.1 System Setup:

The initial structure was obtained from the X-ray crystal structure of human TET2-DNA (PDB code, 4NM6, in complex with the 5-methylcytosine (5mC) substrate)[7] and the missing residues were added using Modeller.[26] The N-oxalylglycine (NOG) used for crystallization was converted to 2OG, and the methyl group of the 5mC substrate was modified to ethyl, vinyl, and ethynyl groups to give the unnatural 5eC, 5vC, and 5eyC substrates, respectively, using GaussView 6.0. However, in 4mC and 4dmC substrates, the methyl group at position 5 of the cytosine in 5mC was removed, and the hydrogen atoms at the exocyclic N4 position were replaced with one and two methyl

groups to give the 4mC and 4dmC substrates, respectively. The protonation state of the ionizable side chain was estimated using Propka,[27] while the Fe-ligating histidine residues were assigned the protonation state based on visual inspection of their local environment. The Fe(IV) ion is coordinated by a monodentate succinate, two histidine residues (His1382 and His1881), and one aspartate residue (Asp1384). The missing hydrogen atoms in the crystal structure were added to the protein followed by neutralization with Na+ counterions using the Leap module in Amber18.[28] The parameters for the non-standard residues/ligands, succinate, 5eC, 5vC, 5eyC, 4mC, and 4dmC substrates were generated using Antechamber[29] as implemented in Amber package. To generate the Fe(IV)=O intermediate metal center parameter, we used the Metal Center Parameter Builder (MCPB.py)[30] module in Amber as used in various studies.[11b,15b,31,32] The high spin quintet state of Fe was assigned to the ground state as various studies have shown the preference of non-heme Fe(II) and 2OG oxygenases for this spin state.[11b,33-35] The 4-coordinated zinc metal centers and the coordinating residues in the Zn finger regions were described using Amber's Zinc Amber force field (ZAFF) method.[36] The systems were surrounded by a rectangular box solvated with TIP3P water molecules[37] within 10 Å from the protein's surface.

2.2 MD Simulations

After the initial setup, the systems were minimized in two steps to remove the bad contacts. In the first step, only the solvent molecules and Na+ counterions were minimized with the solute molecules restrained with harmonic potential of 500 kcal/mol/Å^2 while in the second stage of minimization, all of the systems were geometrically minimized without any constraints. The minimizations were performed using 5000 steps of steepest descent followed by 5000 steps of conjugate gradient minimizer. The CPU version of the Amber18 code (SANDER) was used for the energy minimization. The systems were gently annealed by gradually increasing the temperature from 0 to 300 K in an NVT ensemble for 50 ps using a Langevin thermostat.[38] Subsequently, density equilibration was done at a constant temperature of 300 K for 1 ns in an NPT ensemble. The solute molecules were restrained with a weak restraint of 10 kcal/mol/ Å^2. Thereafter, all of the restraints were removed, and the systems were further equilibrated for 3 ns in an NPT ensemble at a fixed temperature and pressure of 300 K and 1 bar, respectively. Production MD runs were performed for 1000 ns in an NPT ensemble with the target pressure set at 1 bar and a

constant pressure coupling of 2 ps. The pressure was held constant using the Berendsen barostat,[39] and the covalent bonds that contain hydrogens were constrained using the SHAKE algorithm.[40] Long-range electrostatic interactions were treated using the Particle Mesh Ewald (PME) method[41] with a vdW cutoff of 10 Å. The GPU version of the PMEMD engine integrated with Amber18[42] was used for the productive MD simulations. The FF14SB[43] force field was used in all of the simulations, and the periodic boundary conditions were also used in all of the simulations. The obtained trajectories were analyzed using the cpptraj module[44] in Ambertools utilities, and the dynamic cross-correlation analysis was done with Bio3D.[45]

2.3 QM/MM Calculations

The reaction mechanism was performed using combined QM/MM methods. The QM region included the Fe, oxygen atom of the ferryl, the side chain of Asp1384, the imidazole groups of the Fe-ligating His1382 and His1881 residues, succinate, and the alkylated bases of the substrates (5-ethylcytosine, 5-vinylcytosine, 5-ethynylcytosine, 4-methylcytosine, and 4,4-dimethylcytosine parts of 5eC, 5vC, 5eyC, 4mC, and 4dmC substrates, respectively). The ChemShell suite[46] of programs was used for the QM/MM calculations. The QM calculations were performed with Turbomole[47] and the MM calculations with DL_POLY[48] software. The MM region was described using the Amber ff14SB force field,[43] and the electronic embedding scheme[49] was used to describe the interaction between the QM and MM regions. The QM/MM boundary was treated using hydrogen link atoms. The B3LYP functional with def2-SVP basis set was used for all geometry optimizations as in previous studies.[14,17] Afterward, a relaxed potential energy scan was performed along the reaction coordinate with 0.1 A increment to locate the transition states. The DL-find optimizer[50] implemented in ChemShell was used for reactant optimization and scanning while the transition states were reoptimized using the partitioned rational function optimization (P-RFO) algorithm in the HDLC code.[51] The local minimum and the first-order saddle points were verified with frequency calculations. The energies of optimized stationary points were recalculated via single point calculation using def2-TZVP basis set (labeled BS) for all of the atoms. Grimme's D3 dispersion correction[52] was applied in single-point calculations. Energy decomposition analysis (EDA) calculations were then performed on the optimized stationary

point geometries using the method developed by Cisneros and co-workers[53-55] to determine the energetic contributions of the individual residues.

3.0 Results and Discussion

3.1 Structure and Dynamics of the ES Complex for Substrate Oxidation (Fe(IV)=O of Human TET2 with dsDNA Containing Different Alkylated Substituted Bases: 5eC,5vC, and 5eyC)

The Fe(IV)=O intermediate is the reactive oxidizing complex that performs the oxidation of alkylated substrates in 2OG oxygenases; [8,11a,15a,b,56] hence we performed MD simulation of 1 μs on the ferryl complexes of TET2 bound to 5eC, 5vC and 5eyC alkylated dsDNA substrates, enabling us to gain insight into the conformational dynamics of the complexes. Figure S1 shows the root mean square deviations (RMSDs) of the protein-DNA complexes of the systems. The RMSD profiles show the stability of the complexes during simulations. Furthermore, the O-C distance and Fe-O-C angle, which are crucial for efficient substrate oxidation[57a,58] show average values between 3.43 to 4.67 Å and 115.89 to 140.48°, respectively [Figures S2 and S3]. Analyses of the primary coordination sphere show that the Fe(IV) center is stabilized by several interactions. For example, the backbones of the Fe-coordinating histidines (His1382 and His1881) are stabilized via hydrogen bonding interactions with each other [Figure S4] similarly as TET2 bound to the natural 5mC substrate. In 5mC substrate complex, we observed that the non-coordinating oxygen of the aspartate ligand is stabilized by a solvent-mediated hydrogen bonding interaction with Asn1387 (in 74.7% of the trajectories).[12] However, in 5eC substrate complex, the interaction weakens and reduces to 10.2%, likely due to steric effects arising from the variations in the size of the alkyl substituents. The stabilization of the non-coordinating oxygen of Fe-coordinating Asp or Glu residues is a common feature in 2OG enzymes.[8,57b] For example, in AlkB and AlkBH2, an Arg residue forms direct hydrogen bonding interactions with the non-coordinating oxygen of the Fe-coordinating carboxylate residue while in histone JmjC demethylases, KDM4A/E and KDM7B, Asn is employed.[15b,17,18] The networks of hydrogen bonding interactions [Figure S4] of iron coordinating His1382, Asp1384, and His1881 residues with the backbones of His1380, Ala1876, and Glu1879, respectively, in 5eC substrate are preserved as in the natural substrate, enhancing the stability of the Fe center (Table S1). The non-coordinating carboxylate part of the succinate is also stabilized via hydrogen bonding interactions with Arg1896 (70.8%) and Ser1898 (84.4%)

like in 5mC substrate.[12] The heteroaromatic cytosine ring of the substrates are stabilized by π-stacking interactions with Tyr1902, which is conserved during the simulations in both 5mC and 5eC substrates [Figure S5] and thereby aid the orientation of the substrate in the active site. The substrates' exocyclic amine (N4) in both complexes are stabilized by hydrogen bonding interactions with Asn1387. These interactions are very stable in both complexes, especially during the last 500 ns of the simulation [Figure S6]. The stable hydrophobic interactions [Figure S7] between the methylene group of the ethyl substituent in the 5eC substrate and methylene C2 of the Fe-coordinated succinate and Val1900 likely help in the proper orientation of the reactant complex.

The principal component analysis (PCA) [Figure 2a], which shows the flexible parts of the protein, and the direction of motions[59,60a] reveal major motions in the GS-linker and loop 2 (L2) regions in TET2-dsDNA-5eC, unlike in 5mC[12] where in addition to those two motions, the cysteine-rich N-terminal (Cys-N) region and the DNA also experienced some motions. The GS-linker and L2 in TET2-dsDNA-5eC complex move towards the solvents; in contrast to TET2-dsDNA-5mC complex, the same regions move towards the binding surface to aid better binding of the protein to DNA. Also, the observed motions in the DNA and Cys-N regions, which move toward the protein, help in the compaction of the protein-DNA complex in TET2-dsDNA with 5mC dsDNA substrate. These observed differences might be related to the pressure on the enzyme exercised by the unnatural 5eC substrate.

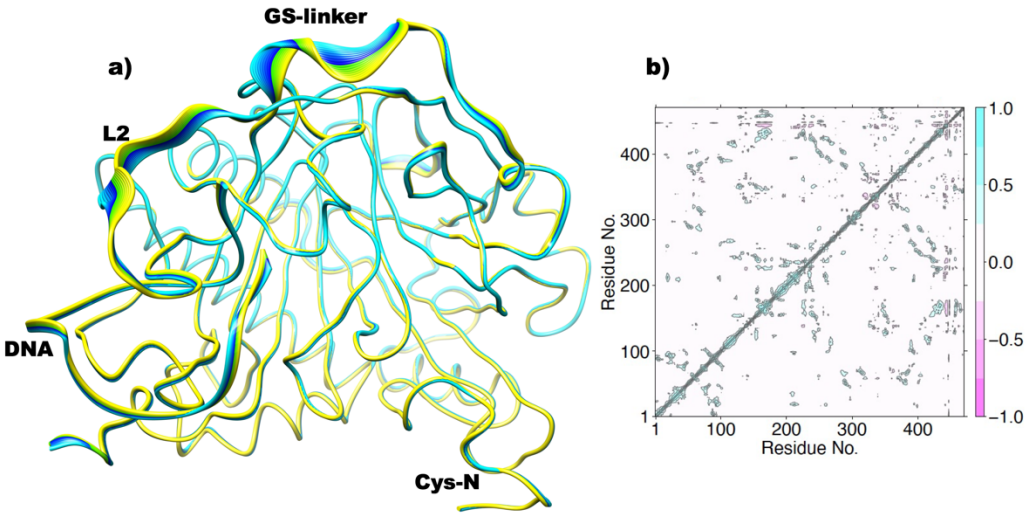


Figure 2: Principal component analysis (a) and dynamic cross correlation (b) of the ferryl complex of TET2 bound to 5eC dsDNA substrate. Residues numbers are as follows: 1-445 (TET protein), 446-448 (Zn), 449 (Fe), 450 (O), 451 (succinate), and 452-475 (DNA). In part (b), residues numbers range 6-17, 157-181, 332-363, and 452-472 on both axes denote the Cys-N, L2, GS-linker, and DNA, respectively. Yellow to blue represents the direction of motion of residues in part (a).

In addition, the distance between the center of mass of the protein and that of the DNA [Figure S8] reveals average values of 22.3 and 27.1 Å for TET2-dsDNA-5mC and TET2-dsDNA-5eC complexes, respectively, indicating that the former complex is more structurally compact than the latter. The observed motions in TET2-dsDNA-5eC complex likely explain the observed expansion of the protein-DNA complex.

The dynamic cross-correlation analysis (DCCA) [Figure 2b] demonstrates that several correlated motions are similar to the TET2 complexed with the natural 5mC substrate. For example, the positively correlated motion of GS linker and L2 with L1, DNA, DSBH core antiparallel β7 and β17, and Zn3 site are also preserved in TET2 bound to 5eC substrate. However, the characteristic correlated motions between L2 and the Fe-coordinating HxD motif loop observed in TET2 with the natural substrate disappear.[12] The long-range correlated motions observed in TET2-dsDNA-5eC complex might assist in the interaction and binding of DNA to the TET2 protein, thereby enhancing the overall structural formation and stability via long-range interactions.

3.2 Dynamics of Fe(IV)=O species of TET2 bound to N4 methylated 4mC and 4dmC dsDNA substrates

The dynamics of the TET2 bound to 4mC and 4dmC substrates reveal similar interactions in the active center as well as in the second sphere region with the details presented in the SI (section 1). In TET2-dsDNA-4dmC complex, the PCA [Figure 3b] show motions in the GS-linker, cys-rich N-terminal (Cys-N), and GS-linker supporting β12 and β13, while in TET2-dsDNA-4mC [Figure 3a], the cys-rich N-terminal (Cys-N), β12 and β13 regions become rigid with new motions in the DNA and loop 2 (L2) regions. These observed motions differ from TET2 bound to natural 5mC substrate, which reveals motion in three different regions (i.e., GS-linker, DNA, and Cys-N). In TET2-dsDNA-4dmC, the Cys-N moves towards the loop containing the binding residues from the

Zn1 finger region. β12 and β13 sheets, which support the GS-linker, move towards Zn2 finger binding residues. These motions aid in stabilizing the first two Zn finger regions via long-range interactions, which in turn assist in the overall structural stability of the protein-DNA complex. However, in TET2-dsDNA-4mC, the GS-linker region moves toward β12 to enhance its stability while L2 and DNA move in parallel and potentially weaken the interactions between the L2 and the DNA. The DCCA of TET2-dsDNA-4dmC [Figure S9] confirms that the Cys-N region has a positive correlation with the Zn1 finger region and its coordinating residues while the GS-linker has a positive correlation with L1, L2, DNA, β6 and β7 elements of DSBH core, supporting the experimental data on the importance of the linker in DNA binding. The antiparallel β12 and β13 regions that support the GS-linker positively correlate with each other, thus increasing the linker stability.

However, in TET2-dsDNA-4mC complex [Figure S10], the L2 positively correlates with the Zn3 finger region and its coordinating residues, DNA, and the iron-coordinating HxD loop, while the GS-linker shows a positive correlation with DNA, L1 and β12. DNA shows a positive correlation with L1, L2, GS linker, β16 and β17 of DSBH core, Fe-coordinating His1382, Asp1384, and loop containing the HxD residues. Overall, the long-range correlated motions in both complexes might aid in DNA recognition and binding to TET2 and specifically in the overall structural stability of the ES complex.

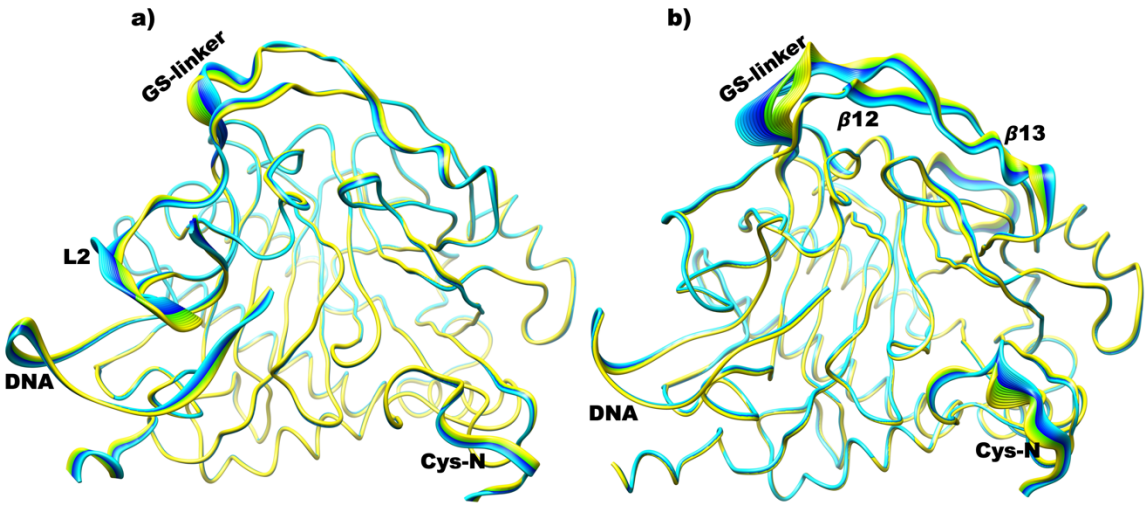
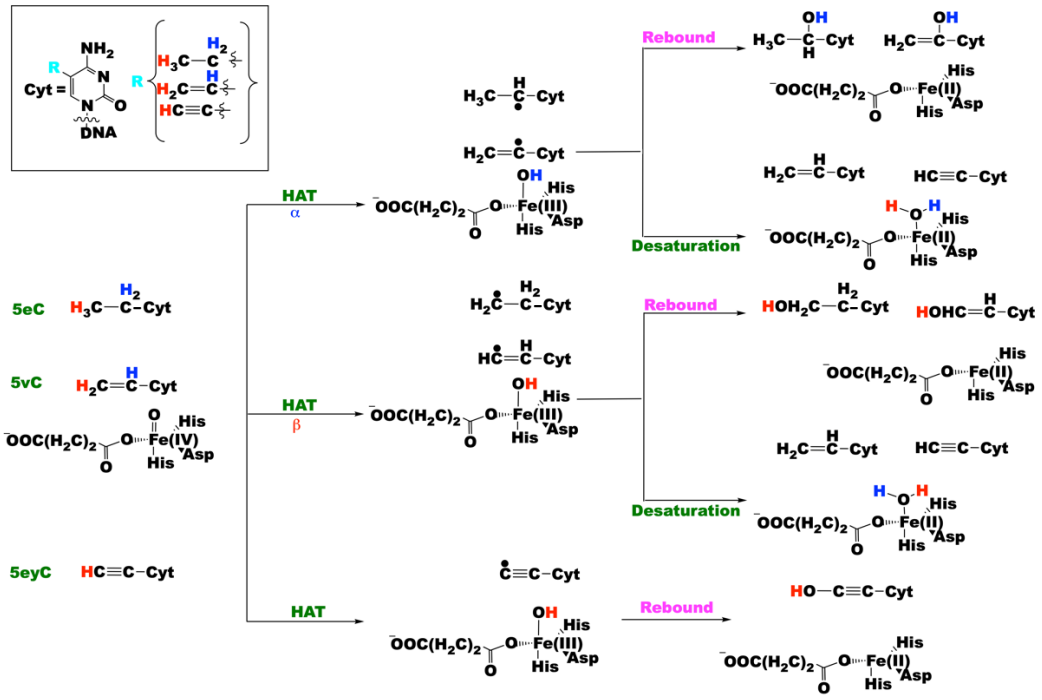


Figure 3: Principal component analysis of the ferryl complex of TET2 bound to 4mC dsDNA (a) and 4dmC (b) substrates. Residues numbers are as follows: 1-445 (TET protein), 446-448 (Zn), 449 (Fe), 450 (O), 451 (succinate), and 452-475 (DNA). Yellow to blue represents the direction of motion of residues.

3.3 Reactions Mechanisms of TET2 with dsDNA containing 5eC, 5vC, and 5eyC

To understand the mechanisms of 5eC, 5vC, and 5eyC dsDNA substrates oxidation by the highly electrophilic Fe(IV)=O intermediate of human TET2, we performed computational studies using combined QM/MM methods as depicted in scheme 1. The substrates oxidation has been proposed to occur via two stages:[8,11a,19,35,61] i) hydrogen atom abstraction (HAT) and ii) rebound hydroxylation or desaturation. Due to the chemical nature of the alkyl substituents in 5eC and 5vC substrates, different alternative pathways for HAT can occur to generate the radical substrates intermediates, which either undergo rebound hydroxylation or desaturation via the second HAT to give the alcohol or desaturated products, respectively (Scheme 1). Previous experimental and computational studies have reported a high spin quintet state as the ground state for the ferryl species,[17,18,33,34,62] hence, this study was conducted at this spin state.



Scheme 1: Possible reaction mechanisms pathways explored for the oxidation of the unnatural 5eC, 5vC, and 5eyC dsDNA substrates by TET2.

3.3.1 Mechanism of 5eC Substrate Oxidation

3.3.1.1 Hydrogen Atom Abstraction

To explore the oxidation of 5eC, we first performed the HAT from the alkyl group of the lesioned DNA substrate by the highly reactive Fe(IV)=O species, leading to the Fe(III)—OH intermediate and a radical substrate. Various studies have reported HAT to be the rate-determining step in 2OG oxygenases.[12,17,19,15b,35,61,62] The HAT was carried out on both the CH_2 (C_α) and CH_3 (C_β) carbons of the lesioned substrate to determine the regioselectivity of oxidation on the ethyl group of the 5eC substrate, resulting in the formation of the Fe(III)—OH center, and a secondary radical (R—$^\bullet$CH—CH_3) (**IM1$_{e\alpha}$**) and a primary radical substrate (R—CH_2—$^\bullet CH_2$) (**IM1$_{e\beta}$**), respectively. To evaluate the effect of conformational flexibility on the rate-determining HAT step, we used five (5) well-equilibrated structures from the last 500ns production MD trajectories by taking the average O—C distance and Fe—O—C angle into account, as these have been shown to determine the substrate oxidation efficiency.[35,60b] The calculated energy barriers at the B3LYP/def2-TZVP (BS) + ZPE level for the five (5) snapshots vary between 15.8 and 23.3 kcal/mol, and 17.4 and 25.7 kcal/mol for HAT from CH_2 (C_α) and CH_3 (C_β) carbons, respectively. The Boltzmann weighted average was found to be 16.7 and 18.4 kcal/mol for HAT from CH_2 (C_α) and CH_3 (C_β), respectively, which is in agreement with the previously reported HAT barriers for Fe(II)/2OG dependent enzymes.[12,17,19,37,61] The slightly higher barrier obtained in the latter likely reflects the preference for HAT from the methylene (CH_2) carbon of the ethyl substituent of 5eC substrate, which is also partly confirmed from the O—C_α and O—C_β distances plots from the productive simulation [Figure S11]. Subsequent calculations were done using the structure that gives the lowest energy barrier for the HAT. Figures 4, 5, and S12 summarize the calculated energy profile and the stationary point geometries for the hydroxylation and desaturation reactions of 5eC dsDNA to give the alcohol and desaturated products, respectively.

Both transition states, **TS$_{H1e\alpha}$** and **TS$_{H1e\beta}$** for HAT from C_α and C_β, respectively, have the Fe—O bond elongated from 1.62 Å to a nearly similar distance of 1.76 and 1.78 Å, respectively [Figure 5]. The Fe—O bond polarizes at the TS to form an efficient electron acceptor, ferric-oxyl species.[34]

In addition, the C—H and O—H distances are 1.24 Å and 1.44 Å, respectively, in **$TS_{H1e\alpha}$** but are 1.24 and 1.34 Å, respectively, in **$TS_{H1e\beta}$**. The obtained electron spin density of 4.00 and 4.05 for the Fe at **$TS_{H1e\alpha}$** and **$TS_{H1e\beta}$**, respectively, confirm the ferric nature of the Fe center in both pathways at the transition states. The two transition states show a difference in the orientation of the substrate ethyl group with respect to the iron-oxo species. The Fe—O—H_α angle in **$TS_{H1e\alpha}$** is 152.56° while the corresponding Fe-O-H_β angle is 143.06° in **$TS_{H1e\beta}$**. The variation in the angles is likely due to the type of carbon where the hydrogen is to be abstracted from (i.e., CH_2 (C_α) vs CH_3 (C_β)), as the substrate ethyl group has to undergo changes to favor the respective HAT.

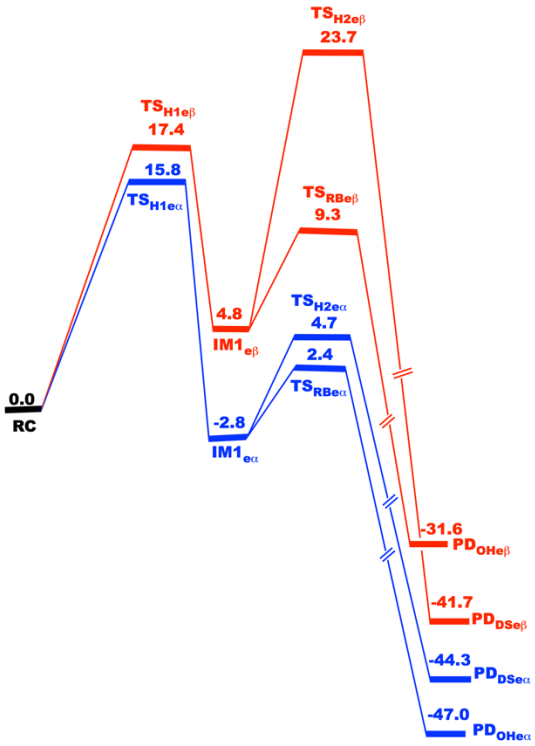


Figure 4: QM/MM potential energy profile for the hydroxylation and desaturation reactions of 5eC dsDNA substrate by TET2, calculated using UB3LYP/def2-TZVP with ZPE. The relative energies are in kcal/mol.

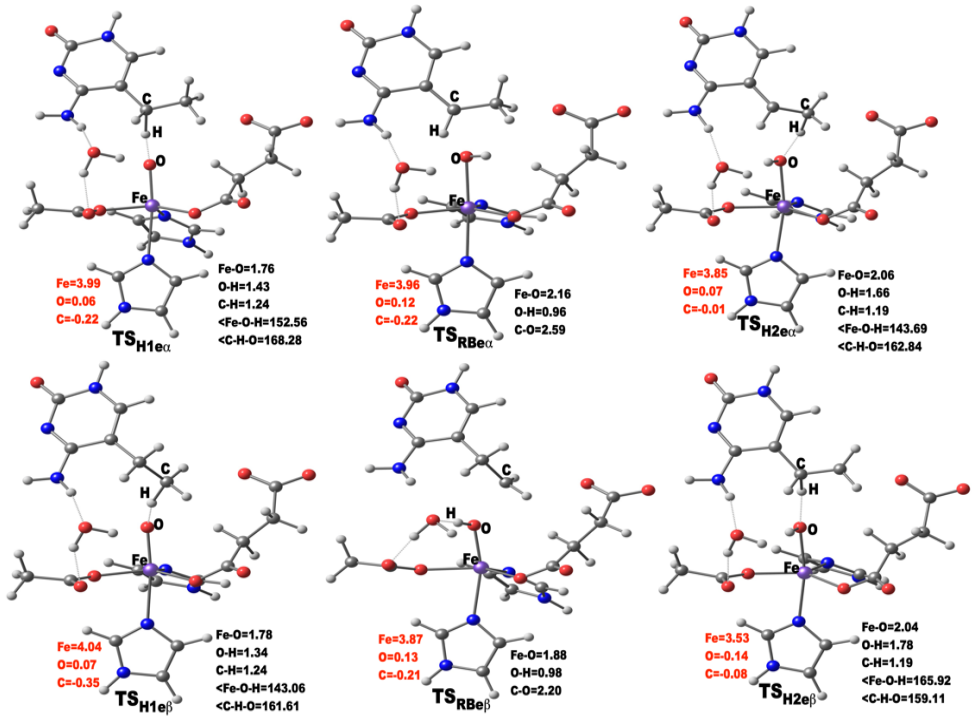


Figure 5: The transition states structures obtained during hydroxylation and desaturation processes of 5eC dsDNA by TET2. Distances (Å), the spin densities are in black and red, respectively, while the Fe—O—H and C—H—O angles are in degrees.

During the hydrogen atom abstraction, one electron from the substrate is transferred to the Fe(IV)=O species to give the Fe(III)—OH intermediate. Two possible pathways have been proposed in the literature for the electron transfer[17,63,64] [Figure 6]: i) σ-pathway, where an α-spin electron is being transferred from the substrate to the unoccupied antibonding $\sigma^*_{z^2}$ orbital of the Fe, resulting in a fully exchange-enhanced Fe 3d orbitals with five unpaired electrons and leaves a β-electron on the substrate. ii) π-pathway, where a β-spin electron is transferred into one of the singly occupied antibonding π* orbitals of the Fe(IV)=O, resulting in an α-electron remaining on the substrate.

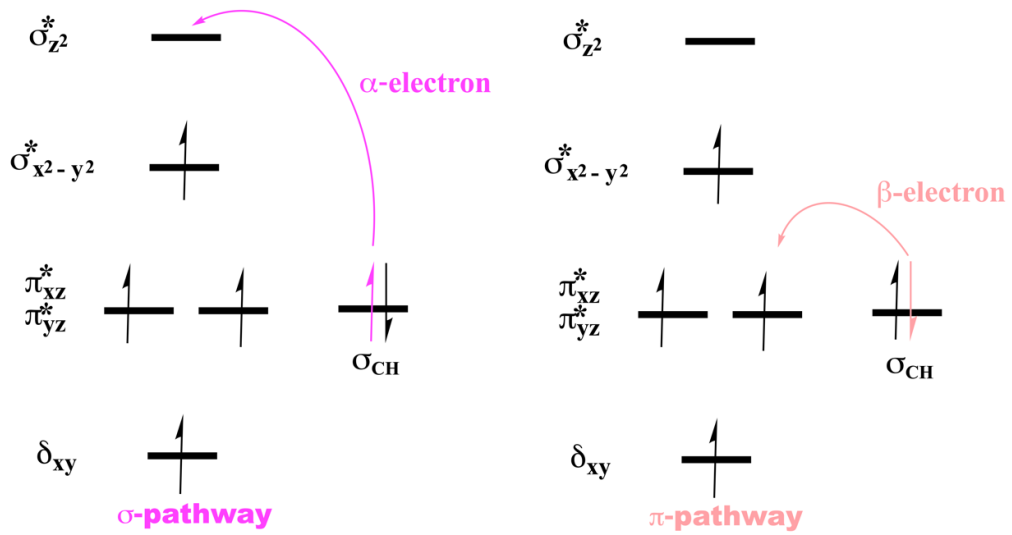


Figure 6: Possible reaction pathways for the HAT by the ferryl complex.

During the HAT via both C_α and C_β, the electron transfer proceeds via σ-pathway as an α-electron is transferred from the substrate to the $\sigma^*_{z^2}$ orbital Fe(IV)=O, leaving a β-electron on the radical substrate, thereby strengthening stabilizing exchange interactions with other unpaired electrons of the Fe 3d-orbitals. The calculated spin densities for the C_α and C_β at the transition states are -0.224 and -0.354, respectively, supporting the transfer of an α-electron to the Fe(IV)=O $\sigma^*_{z^2}$ orbital. The spin natural orbital (SNO) of the two HAT transition states presented in Figures S13 and S14 also supports the electron transfer mechanism. Also, the values of Fe—O—H_α and Fe—O—H_β angles - 152.56° and 143.06°, respectively, additionally support the σ-transfer mechanism. Even though in an ideal case, a Fe—O—H angle of 180° is preferred for an effective σ-transfer[63,64] but numerous cases in the literature are reported for σ-transfer at angles quite lower than the ideal 180°.[17,19,35] Geometric constraints (described in the MD analysis) facilitate the stability of the 5eC and its proper orientation in the protein environment. Studies on other 2OG oxygenase enzymes have also reported similar Fe—O—H angle values for HATs that proceed via σ-pathway.[17,19,35] The σ-pathway is conserved in all the snapshots [Table S2] for the HAT via both the C_α and C_β, implying that the change in conformational flexibility does not alter the MO mechanism of the HAT. Thus the σ-transfer HAT is conserved across 5eC and 5mC substrates.

Similar to the natural 5mC substrate, $\mathbf{TS_{H1e\alpha}}$ and $\mathbf{TS_{H1e\beta}}$ are stabilized by second coordination sphere residues. For example, the cytosine base of the substrate is stabilized via stacking interactions with Tyr1902 and His1904 while the Asn1387 sidechain and NH group of His1904

form hydrogen bonding interactions with exocyclic N4 amine and N3 of the substrate, respectively, aiding in the proper orientation of the substrate in the active site. A series of hydrophobic interactions of residues Thr1393, Val1395, and Val1900 in the vicinity of the Fe center and the substrate also enhance the stability of the two transition states [Figures S15 and S16].

Importantly, in the 5eC HAT transition states, the non-coordinating oxygen of the Fe-ligating Asp1384 is stabilized by hydrogen bonding interaction with the hydroxyl group of Thr1393, which was not observed in 5mC.[12] Furthermore, the stabilization of the methylene (CH_2) and methyl (CH_3) groups of the ethyl substituent of 5eC substrate by the isopropyl sidechain of Val1900 varies in both $\mathbf{TS_{H1e\alpha}}$ and $\mathbf{TS_{H1e\beta}}$. In the $\mathbf{TS_{H1e\alpha}}$, the methyl part of the ethyl group is better stabilized by hydrophobic interaction with the Val1900. In contrast, in $\mathbf{TS_{H1e\beta}}$, the methylene moiety of the ethyl group is more favorably stabilized one via the same interaction.

The dynamic correlation analysis of these TSs stabilizing residues using the dynamic correlation plot from the simulation performed for the ferryl species of the enzyme complex shows that residues Thr1393, Val1395, and Val1900, which are involved in hydrophobic interactions, have a positive correlation with anti-parallel DSBH core β7 and β17, the loop containing Fe-chelating HxD motif, and Cys-rich β6 while Asn1387 shows a positive correlation with DNA interacting loop (L2) residues, Fe center, and the ferryl oxo group, suggesting their importance in catalysis and DNA binding. Tyr1902 and His1904 that stabilize the cytosine ring of the substrate have a positive correlation with Zn3 site, GS-linker, the loop containing HxD Fe-chelating residues, and DNA. These correlated motions are similar to what was observed in the HAT transition state for TET2 bound to the natural 5mC dsDNA substrate, implying that TET2 uses a universal network of correlated motions of catalytically important interactions for both the 5mC and the unnatural 5eC substrate.

After the formation of the transition states, the reaction leads to Fe(III)—OH and radical substrate intermediates, $\mathbf{IM1_{e\alpha}}$ and $\mathbf{IM1_{e\beta}}$ for respective HAT from C_α and C_β. In both intermediates, the O—H distance reduces to 0.97 and 0.98 Å in both complexes, respectively, and the Fe—O bond elongates to 1.88 and 1.87 Å, respectively, when compared with the starting distance of 1.62 Å in the **RC**. These values signify the formation of the ferric-hydroxo intermediate with spin

densities of 4.24 for Fe, pointing to the intermediates' Fe(III) character. The **IM1$_{e\alpha}$** is 2.8 kcal/mol lower in energy than the **RC** while **IM1$_{e\beta}$** is 4.8 kcal/mol above the **RC**, implying that the formed radical intermediate is exothermic and better stabilized via HAT from C_α than from C_β. The difference in the stabilities of the intermediates is likely due to the formation of a benzylic radical substrate intermediate in **IM1$_{e\alpha}$** in comparison to the formation of primary radical substrate intermediate in **IM1$_{e\beta}$**. The benzylic radical is more stable than the primary radical due to the resonance stabilization effect of the benzylic radical. Overall, the HAT from C_α is both kinetically and thermodynamically favorable when compared to C_β as the former occurs via a lower barrier than the latter with a more stable relative energy of the formed radical substrate intermediate.

The QM/MM calculations of the HAT step were also performed in solution, without the protein environment. The rate-determining barrier (18.8 kcal/mol) is consistent with those reported within the protein environment (15.8 kcal/mol). The Fe(III)—OH - radical substrate intermediate, however, is endothermic, unlike in the protein, likely illustrating the importance of the protein environment in the stabilization of the Fe(III)—OH – substrate radical intermediate [Table S3]. We performed additional QM/MM calculations with two different QM region sizes of atoms 56 and 83 in the QM region (for more details, see Table S3 in the SI). The QM/MM calculations for 5eC substrate show slight changes in the HAT energy barrier (values of 14.3 and 16.7 kcal/mol in models with 56 and 83 QM region atoms, respectively) and the stability of Fe(III)—OH intermediate [Table S3]. The results were also compared with the one of QM-only model where the rate-determining HAT barrier is higher by ~8kcal/mol and the Fe(III)-OH radical intermediate is destabilized by ~9kcal/mol [Table S3] when compared to the QM/MM model, highlighting the vital importance of the MM environment in the catalysis.

3.3.1.2 Hydroxylation and Desaturation of 5eC Radical Substrate Intermediates

After the formation of the Fe(III)—OH species and radical substrate intermediates, the hydroxyl group from the ferric-hydroxo species can either undergo rebound to the radical to produce alcohol products or abstract second hydrogen atom from the adjacent C—H bond to form the desaturated products with Fe(II)—OH_2 center. From the **IM1$_{e\alpha}$**, the process of OH rebound passes through **TS$_{RBe\alpha}$** with a barrier of 5.2 kcal/mol from the intermediate. However, the abstraction of the second hydrogen atom that initiates the competitive desaturation reaction passes through

$TS_{H2e\alpha}$ that is 2.3 kcal/mol higher than the rebound transition state **$TS_{RBe\alpha}$**, suggesting that the rebounding of OH to the radical substrate might be more preferable than the abstraction of the second hydrogen from the C—H bond of the substrate radical to form the desaturated product. Due to the small difference in the activation barriers, however, both types of reaction could still occur with close probability. Analysis of the **$TS_{RBe\alpha}$** and **$TS_{H2e\alpha}$** geometries reveal that in **$TS_{RBe\alpha}$**, the Fe—O bond elongates to 2.16 Å while the C_α—O bond shorten to 2.59 Å when compared to 3.51 Å in **$IM1_{e\alpha}$**, pointing to the readiness to form the hydroxylated product. In the desaturation pathway involving the second HAT from C_β, in the **$TS_{H2e\alpha}$**, the Fe—O bond lengthens to 2.06 Å with O—H and C_β —H distance of 1.66 and 1.19 Å, respectively, depicting a more reactant-type transition state structure. In the TS for the second HAT, Fe—O—H angle is 143.69, which deviates from the first HAT from C_α (152.56°), implying that the process of abstracting second hydrogen to form Fe(II)—OH_2 involves a slight change in the orientation of the substrate to favor the desaturation.

Furthermore, we explored the possibilities of both rebound and desaturation from **$IM1_{e\beta}$**. The desaturation barrier, **$TS_{H2e\beta}$**, via second HAT, is 14.4 kcal/mol higher than the rebound one, **$TS_{RBe\beta}$**, indicating that the rebound occurs through a lower energy path than the desaturation. The rebound and desaturation reactions from **$IM1_{e\beta}$** being less favored than **$IM1_{e\alpha}$** could be due to the structural differences in the radical intermediate formed after HAT and, in particular, the orientation of the alkyl substituent in the substrate. In addition, long-range interactions could contribute towards the preference of **$IM1_{e\alpha}$** vs. **$IM1_{e\beta}$**. Analyses of the key distances for both transition states are similar to the observed results from **$IM1_{e\alpha}$** (Figure 5). The hydroxylation and the desaturation from **$IM1_{e\alpha}$** result in highly exothermic **$PD_{OHe\alpha}$** and **$PD_{DSe\alpha}$** products, with the energy of -47.0 and -44.3 kcal/mol, respectively, indicating that the hydroxylated substrate is both kinetically and thermodynamically preferred. The distance of 1.42 Å for C_α—O confirms the formation of hydroxylated product (**$PD_{OHe\alpha}$**), while in the case of desaturation (**$PD_{DSe\alpha}$**), C_α—C_β and Fe—O distance of 1.34 and 2.13 Å, depict the formation of C_α=C_β unsaturated and Fe(II)—OH_2 bonds, respectively.

Similarly, the formed **$PD_{OHe\beta}$** and **$PD_{DSe\beta}$** from **$IM1_{e\beta}$** are also highly exothermic by -31.6 and -41.7 kcal/mol, respectively, but the desaturated product is ~10 kcal/mol stable than the alcohol

product even though the alcohol product occurs via a lower rebound barrier than the one for desaturation [Figure 4]. The result indicates that in the C_β-centered reaction, rebound hydroxylation is kinetically preferred while desaturation is thermodynamically preferred. This observed reactivity pattern from **IM1$_{e\beta}$** implies that the rebound is favored over the desaturation, but the relative energies of the corresponding product give a more stable **PD$_{DSe\beta}$** than **PD$_{OHe\beta}$**.

Overall, the substrate oxidation either via CH_2 (C_α) or CH_3 (C_β) results in more stable products than the **RC.** The C—H activation from C_α gives both stable **PD$_{OHe\alpha}$** and **PD$_{DSe\alpha}$** products than the ones from C_β. The C_α hydroxylation product is more stable than the desaturation product, while in C_β, the reverse is the case. In both cases, the rebound hydroxylation proceeds with lower activation energy than the desaturation reaction. The stability of products formed via CH_2 (C_α) and the lower barriers involved when compared to CH_3 (C_β) likely show the preference for substrate oxidation at the CH_2 position over the terminal CH_3 position of the ethyl group.

The DFT-D3 dispersion corrected energies at BS + ZPE level show the same trends as the calculations without D3 correction with slight differences in both reaction barriers and reaction energies (for more details, see Tables S4-S7 in the SI).

3.3.1.3 Energetic Effects of Individual Residues on the Stabilization of the Transition States and Products

To obtain further insight into the energetic contributions towards the stabilization of the TSs of HAT, rebound and desaturation reactions, we performed energy decomposition analysis (EDA)[54,55] on the QM/MM optimized **RC**, **TS$_{H1e\alpha}$**, **TS$_{H1e\beta}$**, **TS$_{RBe\alpha}$**,**TS$_{H2e\alpha}$**, **TS$_{H2e\beta}$**, **TS$_{RBe\beta}$**, **PD$_{OHe\alpha}$**, **PD$_{DSe\alpha}$**, **PD$_{OHe\beta}$**, and **PD$_{DSe\beta}$**. We calculated the differences in the non-bonded intermolecular interaction energies (i.e., the Coulomb and van der Waals) between the individual residues of the MM environment and the QM region when the system goes from the reactants to the transition states and the products.[54,55] A negative contribution refers to a stabilizing contribution to the transition states and products, while the opposite applies to a positive one.[54,55]

The EDA [Figure 7 and Table S8] based on the optimized structure of HAT transition states, **TS$_{H1e\alpha}$** and **TS$_{H1e\beta}$**, and the **RC** reveal that C1263, R1269, H1380, and E1874 contribute sensitively to the stabilization of **TS$_{H1e\alpha}$**. In contrast, the transition state **TS$_{H1e\beta}$** is stabilized by T1372 and T1393. The variations in the magnitude of energetic contributions and the residues likely account for the

observed lower barrier for **$TS_{H1e\alpha}$** than **$TS_{H1e\beta}$**. For the rebound reaction, the **$TS_{RBe\alpha}$** is stabilized by H1380, C1263, and charged residues K1299, E1874, and E1879 while in **$TS_{RBe\beta}$**, charged residues R1261 and E1874 destabilize the rebound transition state. The desaturation **$TS_{H2e\alpha}$** and **$TS_{H2e\beta}$** are stabilized by similar residues, C1263, K1299, H1380, T1393, E1874, and E1879 but with different magnitude of energetic contributions, and this likely contribute to the lower barrier observed for **$TS_{H2e\alpha}$** when compared to **$TS_{H2e\beta}$**.

The alcohol product of the rebound to C_β, **$PD_{OHe\beta}$**, is stabilized by both positively and negatively charged residues, R1261, R1269, K1299, D1376, K1826, E1874, and E1879 while in **$PD_{OHe\alpha}$**, K1310 and K1905 further stabilized the product in addition to the one observed for **$PD_{OHe\beta}$**. The energetic contributions of these stabilizing residues are significantly greater in **$PD_{OHe\alpha}$**, than **$PD_{OHe\beta}$**, and this might likely be one of the reasons for a more thermodynamically stable **$PD_{OHe\alpha}$** over **$PD_{OHe\beta}$** product. Notably, the desaturated products, **$PD_{DSe\alpha}$** and **$PD_{DSe\beta}$**, are stabilized by different residues, R1359, V1371, and T1393, while R1261 significantly destabilizes the two products.

Experimental mutagenesis studies have shown that mutation of some of these transition states/products stabilizing residues either results in reduction/loss of enzymatic activity and might lead to refractory anemia and chronic myelomonocytic leukemia.[7,65,66] For example, mutant K1299E/S1303N has been reported to significantly reduce TET2 activity and also result in refractory anemia.[7] T1372E mutant has been demonstrated both experimentally and computationally to halt the successive oxidation of 5hmC to 5fC/5caC [54,67] while variant T1393A reduces the oxidation of 5mC to 5hmC by 75% in TET2.[66] The identified residues are part of the network of electrostatic and Van der Waals interactions that most strongly contribute towards the energetic stabilization of the TS and PD in respect to RC. These residues may be important as potential targets for mutagenesis studies.

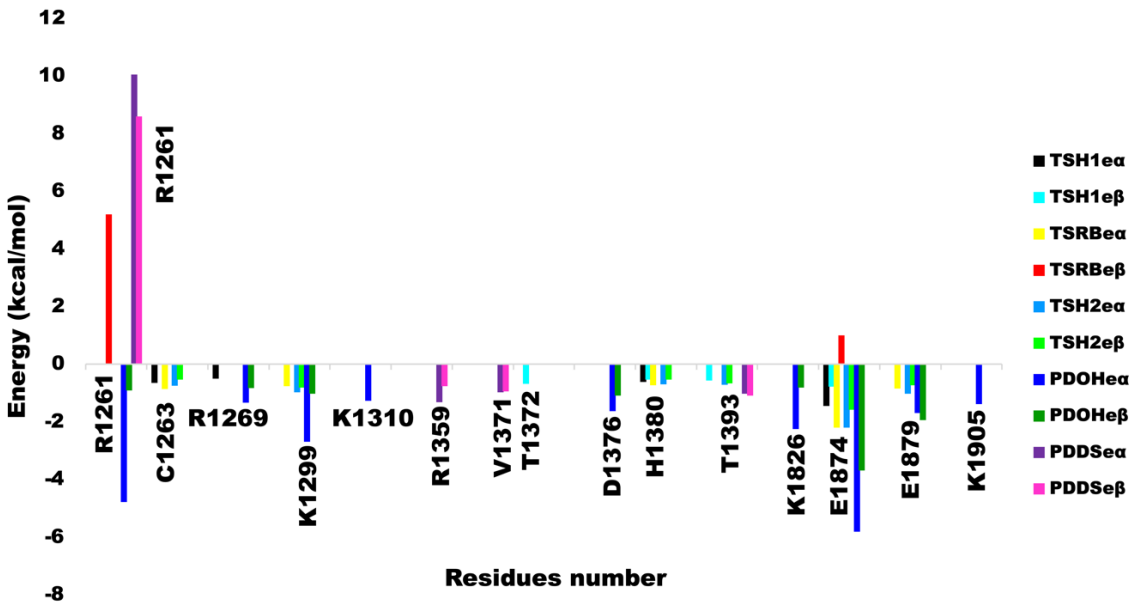


Figure 7: Energy decomposition analysis (EDA) of the residues stabilizing the transition states and the products.

3.3.2 Oxidation of Unsaturated Substrates by TET2

In a recent study, TET2 enzyme has been reported to be catalytically active towards unsaturated 5-vinylcytosine (5vC) and 5-ethynylcytosine (5eyC) substrate similar to saturated 5-ethylcytosine (5eC) and 5-methylcytosine (5mC) substrates.[9] We explored the oxidation of the 5vC and 5eyC substrates. The Fe(IV)=O intermediate can abstract a hydrogen atom from either carbon (.i.e., C_α or C_β) of the 5vC vinyl group in a similar manner as the ethyl group in 5eC while in 5eyC substrate, the intermediate can only abstract the terminal hydrogen atom [Scheme 1]. After the HAT, the radical substrate intermediate undergoes rebound to give the hydroxylated product in 5eyC while in 5vC, both alcohol and desaturated products are possible.

3.3.2.1 Hydrogen Atom Abstraction in 5vC and 5eyC Substrates

The HAT step was studied using 5 well-equilibrated snapshots from the MD to understand the effect of conformational dynamics on this rate-determining step. The Boltzmann weighted average barrier of 22.7 kcal/mol is found for 5eyC while 19.4 kcal/mol and 23.3 kcal/mol are obtained for HAT from C_α and C_β, respectively, for the 5vC. These barriers are slightly higher than the reported HAT barriers for 5mC and 5eC substrate, likely due to the unsaturated nature of the

vinyl and ethynyl groups of 5vC and 5eyC, respectively, making the C—H bonds to be slightly stronger in both than in 5mC and 5eC substrates. Despite the variations in the barriers, they are all in good agreement with the previously reported values for the HAT for non-heme Fe(II) and 2OG dependent oxygenases.[12,18,54,61] Subsequent calculations were then performed using RCs that showed the lowest barriers.

In the 5vC substrate, HAT from C_β is 3.6 kcal/mol higher than the one from C_α, implying that the rate-determining HAT from C_α is faster than C_β, and thus likely shows the preference of HAT from C_α over C_β. This can be due to the combined effects of the closer proximity of the Cα-H bond to the Fe(IV)=O species than Cβ-H and the restriction in the conformational freedom of the vinyl group inside the TET2 pocket, which may be partly due to the rigid C=C bond of the vinyl substituent.

Figure S17 shows the transition state geometries, $\mathbf{TS_{H1v\alpha}}$ and $\mathbf{TS_{H1v\beta}}$, for HAT from C_α and C_β, respectively, for 5vC substrate. $\mathbf{TS_{H1v\alpha}}$ has a symmetric HAT transition state with nearly equal O—H and C—H distances of 1.28 and 1.27 Å, respectively, whereas in $\mathbf{TS_{H1v\beta}}$, the transition state is product-like, i.e., late TS with the transferring hydrogen atom closer to the acceptor atom than the donor atom. The values obtained for O—H and C—H distances are 1.14 and 1.39 Å, respectively. In both TSs, the Fe—O bond elongated to similar distances of 1.79 Å for $\mathbf{TS_{H1v\alpha}}$ and 1.75 Å for $\mathbf{TS_{H1v\beta}}$. The Fe—O—H angle varies from 135.22 to 150.88° in the $\mathbf{TS_{H1v\alpha}}$ from the five scans, whereas in $\mathbf{TS_{H1v\beta}}$, the values between 114.52 and 120.65° are obtained, indicating that an α-electron might be shifted from the substrate to the Fe(IV)=O orbital in $\mathbf{TS_{H1v\alpha}}$, while a β-electron might be transferred in the case of $\mathbf{TS_{H1v\beta}}$. These observations are further confirmed by calculating the spin densities values of both C_α and C_β of the vinyl group at the respective transition state. In $\mathbf{TS_{H1v\alpha}}$, the C_α spin density varies between -0.373 and -0.247, while values from 0.557 to 0.646 are obtained for C_β in $\mathbf{TS_{H1v\beta}}$ for all the snapshots. These results reveal that a β-electron is left on the vinyl group of the 5vC substrate in $\mathbf{TS_{H1v\alpha}}$ while in $\mathbf{TS_{H1v\beta}}$, an α-electron remains in the substrates, suggesting that HAT from C_α and C_β proceed via σ- and π-pathway, respectively [Table S9]. By modifying the structural determinants for the σ- and π-pathways, one could switch between the two mechanisms, which is an example of how the protein environment

might influence the orbital mechanism of HAT. The analysis of the Spin Natural Orbitals[63] [Figures S18 and S19] confirms the spin density results.

Similarly, in 5eyC, the abstraction of the only terminal hydrogen by the active Fe(IV)=O complex passes through **$TS_{H1ey\beta}$**, which possess C—H and O—H distances of 1.44 and 1.09 Å, respectively, in a similar manner as HAT from the terminal C_β in 5vC substrate. The results indicate that the HAT from the terminal carbon (C_β) in both unsaturated 5vC and 5eyC dsDNA substrates possess late transition states with product-like properties. The observed Fe—O—H angle in **$TS_{H1ey\beta}$** varies between 127.16 to 138.79° and the calculated spin densities of the terminal carbon at the transition state range between -0.214 to -0.117, pointing to a σ-electron transfer mechanism for the HAT in 5eyC substrate [Table S10].

The HAT results in the reduction of Fe(IV)=O species to Fe(III)—OH complex with the generation of radical substrates intermediates, **$IM1_{ey\beta}$** for 5eyC, **$IM1_{v\alpha}$** and **$IM1_{v\beta}$**, for HAT from C_α and C_β of the vinyl group of 5vC, respectively. The formed **$IM1_{ey\beta}$** is endergonic with an energy of 7.9 kcal/mol at the BS + ZPE level of theory, while in **$IM1_{v\alpha}$** and **$IM1_{v\beta}$** of 5vC substrate, the intermediates are also endergonic with the energy of 3.4 and 9.3 kcal/mol, respectively. In all the intermediates, the elongation of the Fe—O bond by ~0.25 Å when compared to **RC** and the O—H distances of ~0.98 Å confirms the formation of the ferric-hydroxo intermediate. The calculated spin densities between 4.23 and 4.24 in all the intermediates support the +3-oxidation state assigned for Fe at the reaction state.

3.3.2.2 Rebound Hydroxylation and Desaturation Reactions in 5vC and 5eyC Radical substrates Intermediates

Similarly to 5eC, the formed of the Fe(III)—OH complex and the substrates radical intermediates, **$IM1_{v\alpha}$** and **$IM1_{v\beta}$** of 5vC substrate can undergo radical rebound processes via **$TS_{RBv\alpha}$** and **$TS_{RBv\beta}$**, respectively, in which the hydroxyl (OH) group is transferred from the ferric-hydroxo complex to the radical carbons, resulting in the formation of hydroxylated intermediates, **$IM2_{OHv\alpha}$** and **$IM2_{OHv\beta}$**, and reduction of Fe(III) to Fe(II). Alternatively, **$IM1_{v\alpha}$** and **$IM1_{v\beta}$** can undergo desaturation by abstracting another hydrogen atom from the adjacent carbon via **$TS_{H2v\alpha}$** and **$TS_{H2v\beta}$** to form the desaturated products, **$IM2_{DSv\alpha}$** and **$IM2_{DSv\beta}$** and Fe(II)—OH_2 center, similarly to

5eC. The imaginary frequency values for all the obtained transition states are presented in Table S11.

The rebound reaction occurs more rapidly in **$TS_{RBv\alpha}$** with a barrier of 11.8 kcal/mol and the overall reaction is highly exothermic with energy of -35.5 kcal/mol, implying that the formed **$IM2_{OHv\alpha}$** is very stable. However, the rebound barrier leading to the formation of **$IM2_{OHv\beta}$** is 17.7 kcal/mol, and thus slower than the one of **$IM2_{OHv\alpha}$** with overall reaction energy of -29.2 kcal/mol. The calculated desaturation pathways where a second hydrogen atom is abstracted from the adjacent carbon pass through **$TS_{H2v\alpha}$** and **$TS_{H2v\beta}$** and result in higher barriers of 20.9 and 25.1 kcal/mol, respectively. The resultant desaturated products, **$IM2_{DSv\alpha}$** and **$IM2_{DSv\beta}$** are ~22.0 kcal/mol more stable than the **RC** but less stable than the hydroxylated intermediates. Overall, the results imply that the rebound hydroxylation of 5vC is both kinetically and thermodynamically preferred over the competitive desaturated reaction, possibly due to the change in the orientation of C_β methine radical group in the RC for the rebound reaction. In addition, the study reveals that the rebound hydroxylation on C_α-based radical is preferred (again both kinetically and thermodynamically) over a rebound hydroxylation of C_β-based radical.

The rebound hydroxylation in 5eyC passes through **$TS_{RBey\beta}$** possesses a barrier of 17.4 kcal/mol and the formed terminal hydroxylated intermediate, **$IM2_{OHey\beta}$** is 32.9 kcal/mol below the **RC**, suggesting the formation of a stable product. The optimized geometries of the stationary points and QM/MM energy profile are presented in Figures S17, S20, S21, and 8, respectively. In summary, the rebound hydroxylation proceeds with lowest activation barrier (5.2 kcal/mol) in C_α atom of 5eC followed by 5mC (10.1 kcal/mol),[12] followed by C_α atom of 5vC (11.8 kcal/mol) and 17.4 kcal/mol in 5eyC.

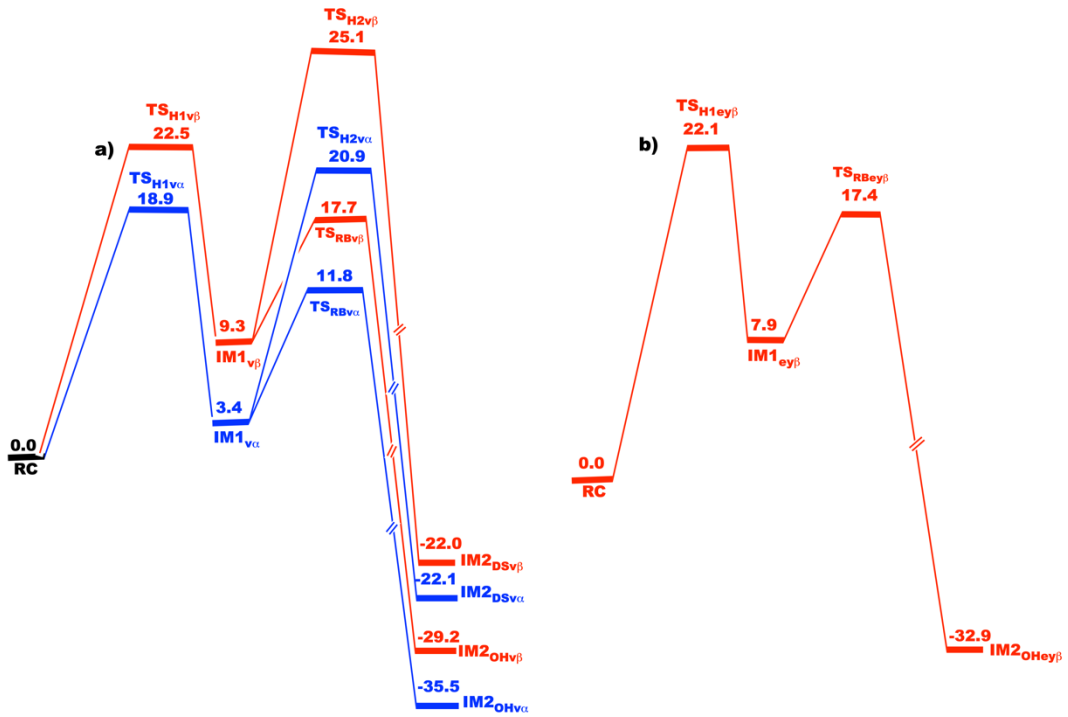


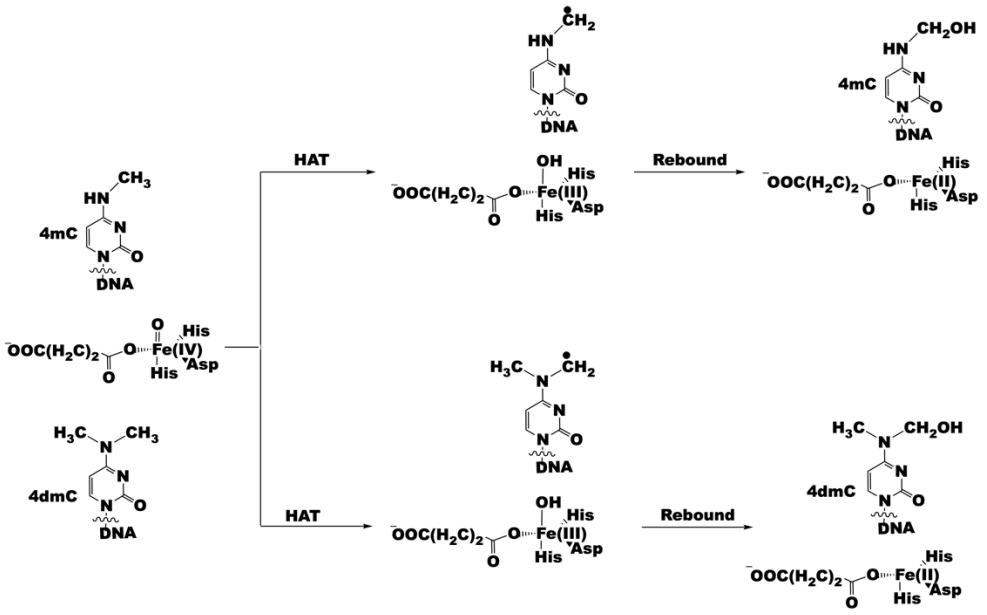
Figure 8: QM/MM potential energy profile for the hydroxylation and desaturation reactions of 5vC (a) and for the hydroxylation of 5eyC (b) dsDNA substrates by TET2, calculated using UB3LYP/def2-TZVP with ZPE. The relative energies are in kcal/mol.

Even though various studies have shown that non-heme Fe(II) and 2OG dependent enzymes prefer high spin quintet spin state as the ground state,[17,18,33,34] we also performed calculations at septet higher spin state for 5eC, 5vC, 5eyC, and 4mC substrates, and the result indeed revealed that the quintet Fe(IV)=O species is more potent oxidants for C—H activation and thus the oxidation of the substrate is faster at quintet than septet spin state in all cases [Tables S4-S7]. Comparison of the Fe(IV)=O complex energies at both spin states revealed that the quintet ground state is more stable than the septet by 16.4, 16.9, 14.3, and 15.2 kcal/mol in 5eC, 5vC, 5eyC, and 4mC substrates, respectively, thus supporting the previous experimental and computational studies on the favorable ground state for non-heme 2OG enzymes.[20,21,34]

3.4 Mechanism of 4mC and 4dmC dsDNA substrates Oxidation by TET2

The discovery that TET2 oxidizes 4mC and 4dmC (lacking C5 alkylations) demonstrate the potential of TET2 to act as indirect (on C5 alkylations) and direct demethylase (on 4mC and 4dmC). Aiming to provide insight into how TET2's catalytic mechanism and interactions of direct

demethylation might differ from indirect demethylation, we performed QM/MM studies of TET2 with dsDNA containing 4mC and 4dmC (Scheme 2).



Scheme 2: Reaction mechanisms for the hydroxylation of the unnatural 4mC and 4dmC dsDNA substrates by TET2.

The oxidation of the substrates using 5 well-equilibrated snapshots gave Boltzmann weighted average of 10.8 and 14.0 kcal/mol for the rate-determining HAT step of 4mC and 4dmC substrates, respectively. During the HAT, the electron transfer occurred via σ-channel, where an α-electron was transferred into the unoccupied antibonding orbital of the Fe. The HAT results in exothermic Fe(III)—OH intermediates with energies of -5.7 and -7.7 kcal/mol for 4mC and 4dmC, respectively, at BS + ZPE level of theory. The rebound hydroxylation of the radical intermediates led to the formation of highly exothermic hydroxylated intermediates, **IM2$_{OH4m}$** and **IM2$_{OH4dm}$** for 4mC and 4dmC with overall energies of -37.6 and -42.8 kcal/mol, respectively. **IM2$_{OH4dm}$** is more stable than **IM2$_{OH4m}$** by 5.2 kcal/mol as the OH group of the **IM2$_{OH4dm}$** is locked in a strong hydrogen bonding interaction with the negatively charged oxygen atom of the C4 carboxylate of the succinate, unlike in **IM2$_{OH4m}$** where the OH group forms weaker hydrogen bonding interaction with the non-coordinating carbonyl oxygen of the Fe-ligating Asp1382. Detailed analyses are provided in the SI (section 2).

3.5 Post Hydroxylation Reactions for hydroxylated intermediates of $IM2_{OHey\beta}$, $IM2_{OHv\beta}$, and $IM2_{OH4m}$ of 5eyC, 5vC, and 4mC Substrates

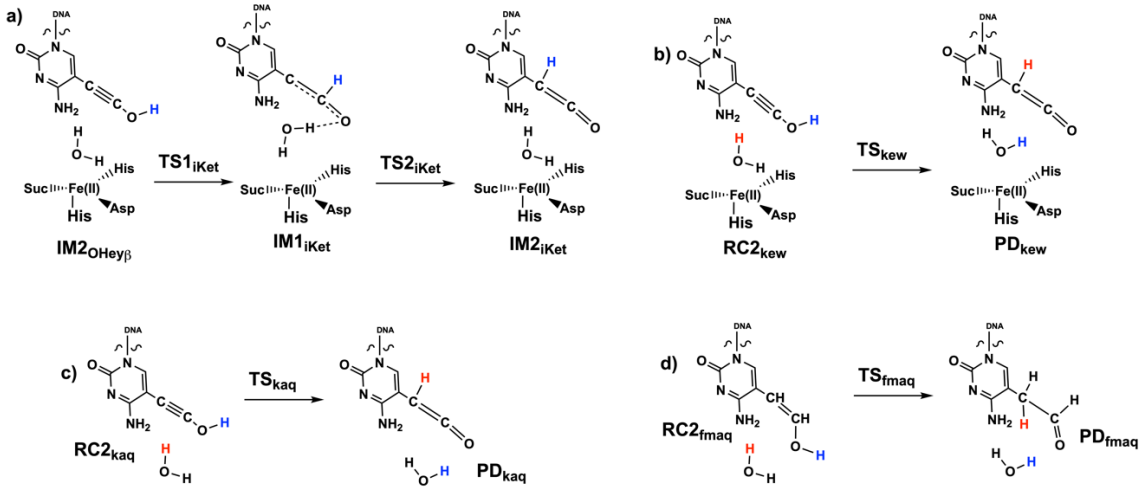
The decomposition of the hydroxylated hemiaminal intermediate of the DNA base has been a subject of discussion about whether the hydrolysis to the unmodified base would proceed in the enzyme or an aqueous solution. A recent study on FTO has reported that the decomposition of the hemiaminal intermediate to formaldehyde and unmodified DNA base proceed faster in an aqueous solution.[61] Another study demonstrated that the stabilities and reactivities of the hemiaminal intermediate vary depending on the position of the N-methyl group on the substrate. The fragmentation of the hemiaminal intermediate where the hydroxymethyl is linked to the endocyclic nitrogen was found to be faster than the exocyclic counterparts, which are very slow due to their stability.[68]

3.5.1 Rearrangement Reactions of $IM2_{OHey\beta}$ and $IM2_{OHv\beta}$ in Enzyme and Water

Studies on the ability of TET enzymes to act on unnatural modifications of cytosine base at C5 position by Kohli and co-workers[9] have suggested that the **$IM2_{OHey\beta}$**, formed during the oxidation of 5eyC can undergo rearrangement to give a ketene intermediate. To achieve this, we explored using QM/MM methods two possible pathways: i) rearrangement in the enzyme where the proton from the hydroxyl (OH) group is transferred to the C_α of the substituent with or without mediation via water molecule; ii) rearrangement in aqueous solution outside of the enzyme.

The rearrangement in (i) involving the transfer of the OH proton to the C_α generated an intermediate, **$IM1_{iKet}$** where the proton is first temporarily transferred to the C_β [Scheme 3a]. This passes through **$TS1_{iKet}$** with a 53.4 kcal/mol barrier, which is too high to favor such a pathway. After the formation of **$IM1_{iKet}$**, the proton is transferred to the C_α to generate the ketene product. This transfer proceeds via **$TS2_{iKet}$** with a barrier of 24.7 kcal/mol, resulting in a highly exothermic ketene product, **$IM2_{iKet}$**. This calculated pathway has a remarkably high energy barrier, with the **$TS1_{iKet}$** being the rate-determining step, even though the formed product is very stable. Next, we explored the possibility of mediating the rearrangement via water molecules in the enzyme [Scheme 3b]. The assistance with water resulted in a barrier of 28.4 kcal/mol, which is 25.0 kcal/mol lower than the rate-determining step barrier observed for the rearrangement without water molecules but still too high to be feasible.[61] Various studies have reported the role of water

molecules in the reduction of activation barriers.[61,69,70] For example, a study by Shaik and co-workers on FTO reported a 25.6 kcal/mol barrier for the water-assisted decomposition of hydroxylated intermediate of 6-methyladenine substrate in contrast to 59.0 kcal/mol obtained without the assistance of water molecules.[61] Also, a study on the catalytic mechanism of HIV-1 protease revealed that the rate of degradation of the substrate was lowered by almost 10 kcal/mol in the water-assisted pathway.[69] Our study, in agreement with previous works, demonstrates the role of water in enhancing the reaction rate.



Scheme 3: Reaction scheme for a) formation of ketene inside the enzyme without the assistance of water molecule; b) formation of ketene with the assistance of water inside the enzyme; c) formation of ketene with the assistance of water outside the enzyme in the water solvent; d) formation of 5-formylmethylcytosine (5fmC) outside the enzyme in the water solvent.

The rearrangement results in a highly stable product with energy of -26.5 kcal/mol. We, therefore, performed a QM/MM study of the rearrangement in an aqueous solution outside of the enzyme [Scheme 3c], where the hydroxylated 5eyC substrate was embedded in a water box containing 896 water molecules. The hydroxylated cytosine part was included in the QM part with one water molecule, and the remaining water molecules remained in the MM. We are conscious that QM/MM calculations using a single structure might not provide the most accurate value of the activation and reaction energies in highly disordered media and protein environment.[12,17,71,72] The formation of the ketene proceeded faster with a barrier of 10.4 kcal/mol, suggesting that the rearrangement of **IM2$_{OHeyβ}$** to ketene takes place in aqueous

solution outside of the enzyme. Figures 9 and S22 present the potential energy profile for the rearrangement of **IM2$_{OHey\beta}$** as well as the optimized stationary point geometries involved in these reactions.

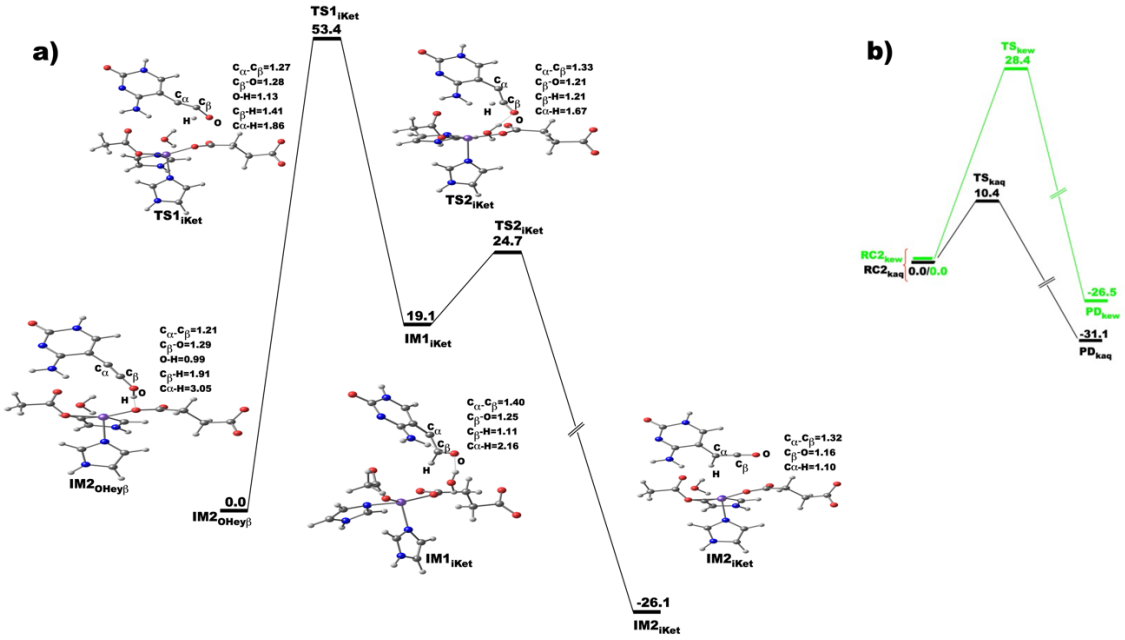


Figure 9: QM/MM potential energy profiles for the formation of ketene inside the enzyme without the assistance of water molecule (a) and the ketene formation with the assistance of water (b) inside (green) and outside (black) of the enzyme. The relative energies and the distances are in kcal/mol and Å, respectively.

Similarly, we carried out the tautomerization of **IM2$_{OHv\beta}$** to generate 5-formylmethylcytosine (5fmC) product which has been reported by Ghanty et al., to be the predominant product of the oxidation of 5vC by TET2.[9] We performed the calculations via direct transfer of the OH proton to the C_α of the substituent with and without the assistance of water molecule in a similar manner as above ketene intermediate formation. The QM/MM potential energy scan shows an increase in energy without leading to the formation of the desired product. Furthermore, we explored the possibility of mediating the proton transfer via the noncoordinating oxygen of the iron coordinating succinate (Suc) as the hydroxyl group is locked in hydrogen bonding interaction with it. The proton transfer to the succinate resulted in protonated succinate (SucH) product, and it passes through a transition state with an energy barrier of 1.62 kcal/mol, indicating a very rapid process [Figure S23]. On an attempt to transfer the proton from the SucH to the C_α of the

substrate substituent, it falls back to the hydroxylated **IM2$_{OH\nu\beta}$** intermediate. We then explored the tautomerization mechanism in an aqueous solution outside the enzyme [Scheme 3d]. The calculations showed the formation of a very stable and exothermic 5-formylmethylcytosine (5fmC), with a reaction barrier and overall reaction energy of 17.8 and -12.1 kcal/mol, respectively [Figure S23], implying that as the ketene formation reaction, the formation of 5-formylmethylcytosine proceeds in water and not in the enzyme.

3.5.2 Decomposition of hemiaminal intermediate of 4mC

The hemiaminal intermediate **RC3** formed after the oxidation of 4mC undergoes decomposition to form the demethylated cytosine base and formaldehyde.[10] The demethylation is first studied in the enzyme with the aid of a water molecule, as presented in Figures 10 and S24. The decomposition starts from **RC3**, where the proton from the hydroxyl group is first transferred to the N3 of the base. This transfer is mediated via a water molecule with an energy barrier of 17.6 kcal/mol, resulting in the generation of an endothermic **IM1** with an energy of 6.0 kcal/mol. Thereafter, the C4—N4 bond cleaves to give **IM2** intermediate, which passes through **TS2** of energy barrier 14.6 kcal/mol. This bond cleavage leads to the elimination of the formaldehyde from the hemiaminal intermediate. Subsequently, a water-assisted proton shift from N3 to N4 via **TS3** led to the demethylated cytosine product (**PD$_{dm}$**). The calculated overall barrier of 50.9 kcal/mol for the rate-determining step is too high to be feasible even though the formed product is stable by the energy of -3.3 kcal/mol. Hence, we explored the possibility of decomposing the hemiaminal intermediate in an aqueous solution outside of the enzyme environment using QM/MM.

As presented in Figure S25, the results reveal that the demethylation in aqueous solution proceeds with a barrier of 20.4 kcal/mol. This result is in good agreement with the previously reported experimental[56] and computational studies by Shaik and co-workers on the decomposition of N^6-hydroxymethyladenosine (hm^6A) by other human N-alkylated nuclei acid demethylase (FTO), where the hydrolysis of the hm^6A proceeded in aqueous solution outside of the enzyme after the in-enzyme hydroxylation of the methyl group of the adenine base.[61] The product formed here is endothermic with an energy of 4.1 kcal/mol in contrast to -3.3 kcal/mol

obtained inside the enzyme, indicating the role of the protein environment in the stabilization of the product in the latter.

The studies demonstrate the post-hydroxylation steps, e.g., ketene formation after the 5eyC hydroxylation, the formation of 5-formylmethylcytosine after hydroxylation of 5vC, and the formaldehyde separation from hemiaminal of 4mC proceeds in the water solvent and not in the enzyme environment, nevertheless of the diversity of the chemical transformations of the respective hydroxylated intermediates. The finding implies that TET2 as a Fe(II)/2OG enzyme is specialized in providing the suitable catalytic environment only for the reactions of the dioxygen activation and the substrate oxidation (HAT and rebound hydroxylation) of several different alkylated forms of cytosine in dsDNA, while the post-hydroxylation reactions proceed non-enzymatically in the water solvent.

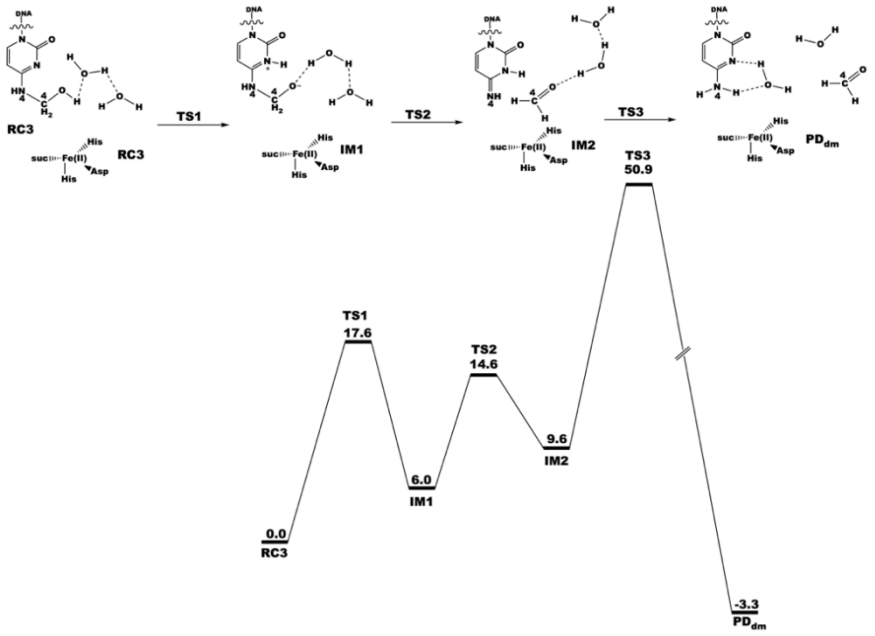


Figure 10: Reaction scheme and the QM/MM potential energy profile for the decomposition of hemiaminal intermediate of 4mC hydroxylation inside the enzyme, calculated using UB3LYP/def2-TZVP with ZPE. The relative energies are in kcal/mol.

4.0 Conclusions

We report the MD and QM/MM studies on the dynamics and the reaction mechanisms involved in the oxidation of unnatural C5-position modifications of cytosine (5eC, 5vC, and 5eyC) as well as demethylation of exocyclic N4 mono- and di-methylated (4mC and 4dmC) lesions of cytosine

by TET2 enzyme. We studied the possible hydroxylation and desaturation pathways in the oxidation of 5eC and 5vC dsDNA substrates, while in the case of 5eyC, 4mC, and 4dmC, we explored the hydroxylation pathway. Due to the nature of the ethyl and vinyl substituents of 5eC and 5vC, respectively, we explored oxidation at both Cα and Cβ carbons of the two substituents. We then carried out QM/MM modeling of the rearrangement reactions of the hydroxylated intermediates of 5eyC and 5vC substrates to produce ketene intermediate and 5-formylmethyl-cytosine (5fmC), respectively. The hydroxylated intermediate of 4mC (hemiaminal) was demethylated to unmodified cytosine and formaldehyde. These post-hydroxylated reactions were studied both inside and outside of the enzyme with the participation of water molecules, showing that in all cases, the post-hydroxylation reactions proceed in the water solvent outside the enzyme environment.

The dynamics studies reveal that the variations in the alkylation status on the cytosine base influence the overall collective motions of the key structural motifs of the enzyme and the second coordination sphere interactions. The QM/MM calculations reveal that HAT from the N-methylated substrates is faster than the counterparts from C5-position modifications. HAT from Cα has lower barriers than Cβ in 5eC and 5vC, resulting in more stable Fe(III)—OH intermediates than the one from Cβ. Furthermore, we explored the effects of conformational flexibilities on the HAT by using multiple snapshots. The results show variations in the HAT reaction barriers, implying that the conformational changes affect the HAT rate. The calculations reveal that σ-pathway is used for the HAT electron transfer mechanism except for HAT from Cβ in 5vC substrate, where a π-pathway is favored. The electron transfer pathways are preserved in all the snapshots for all the systems, suggesting that the change in conformations does not affect the electron transfer pathways used during the HAT, nevertheless, the conformational changes can influence the energetic changes along the reaction path. The calculations also identified the key residues crucial for the HAT and their contributions to catalysis, DNA interaction, and structural stability via long-range interactions.

Furthermore, the hydroxylation pathway in 5eC and 5vC is energetically favored than the desaturation one, showing the preference for the former pathway. The rearrangement of the hydroxylated intermediates of 5eyC to ketene intermediate proceeds with a barrier of 28.4 and

10.4 kcal/mol with the assistance of water molecules in and outside of the enzyme, respectively, indicating that the rearrangement is faster outside of the enzyme as in 5vC. Our calculated barrier for the decomposition of the hemiaminal intermediate of 4mC in an aqueous solution outside of the enzyme is 20.4 kcal/mol, which is lower than the calculated one inside the enzyme, but the demethylated product is more stable inside the enzyme, possibly due to the extra stabilization from the protein environment.

The results delineate the fine and delicate changes in interactions in the active site and beyond that are involved in the TET2-catalyzed oxidation of unnatural cytosine alkylations in dsDNA. The study provides an insight into the nature of the delicate but important differences between the key catalytic interactions involved in the catalysis of the unnatural substrates in comparison to the natural substrates. Overall, the study explains the atomistic and electronic structural mechanism of the substrate promiscuity of TET2 and its catalytic strategy as an indirect and direct demethylase.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Analysis of molecular dynamics and QM/MM optimized QM geometries with spin densities and Cartesian coordinates of the QM region in QM/MM calculations are provided.

Acknowledgments

C.Z.C. acknowledges NSF grant 1904215 (testing the methodology) and NIH grant 1R15GM139118-01A1.

References

(1) Tsiouplis, N. J.; Bailey, D. W.; Chiou, L. F.; Wissink, F. J.; Tsagaratou, A. TET-Mediated Epigenetic Regulation in Immune Cell Development and Disease. *Front. Cell Dev. Biol.* **2021**, *8*, 623948.

(2) Mahfoudhi, E.; Talhaoui, I.; Cabagnols, X.; Della Valle, V.; Secardin, L.; Rameau, P.; Bernard, O. A.; Ishchenko, A. A.; Abbes, S.; Vainchenker, W.; Saparbaev, M.; Plo, I. TET2-Mediated 5-Hydroxymethylcytosine Induces Genetic Instability and Mutagenesis. *DNA Repair* **2016**, *43*, 78–88.

(3) Bird, A. DNA Methylation Patterns and Epigenetic Memory. *Genes Dev.* **2002**, *16* (1), 6–21.

(4) Cimmino, L.; Abdel-Wahab, O.; Levine, R. L.; Aifantis, I. TET Family Proteins and Their Role in Stem Cell Differentiation and Transformation. *Cell Stem Cell* **2011**, *9* (3), 193–204.
(5) Branco, M. R.; Ficz, G.; Reik, W. Uncovering the Role of 5-Hydroxymethylcytosine in the Epigenome. *Nat. Rev. Genet.* **2012**, *13* (1), 7–13.
(6) Pastor, W. A.; Aravind, L.; Rao, A. TETonic Shift: Biological Roles of TET Proteins in DNA Demethylation and Transcription. *Nat. Rev. Mol. Cell Biol.* **2013**, *14* (6), 341–356.
(7) Hu, L.; Li, Z.; Cheng, J.; Rao, Q.; Gong, W.; Liu, M.; Shi, Y. G.; Zhu, J.; Wang, P.; Xu, Y. Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation. *Cell* **2013**, *155* (7), 1545–1555.
(8) Zheng, G.; Fu, Y.; He, C. Nucleic Acid Oxidation in DNA Damage Repair and Epigenetics. *Chem. Rev.* **2014**, *114* (8), 4602–4620.
(9) Ghanty, U.; DeNizio, J. E.; Liu, M. Y.; Kohli, R. M. Exploiting Substrate Promiscuity To Develop Activity-Based Probes for Ten-Eleven Translocation Family Enzymes. *J. Am. Chem. Soc.* **2018**, *140* (50), 17329–17332.
(10) Ghanty, U.; Wang, T.; Kohli, R. M. Nucleobase Modifiers Identify TET Enzymes as Bifunctional DNA Dioxygenases Capable of Direct N-Demethylation. *Angew. Chem. Int. Ed.* **2020**, *59* (28), 11312–11315.
(11) (a) Fedeles, B. I.; Singh, V.; Delaney, J. C.; Li, D.; Essigmann, J. M. The AlkB Family of Fe(II)/α-Ketoglutarate-Dependent Dioxygenases: Repairing Nucleic Acid Alkylation Damage and Beyond. *J. Biol. Chem.* **2015**, *290* (34), 20734–20742. (b) Waheed, S. O.; Ramanan, R.; Chaturvedi, S. S.; Ainsley, J.; Evison, M.; Ames, J. M.; Schofield, C. J.; Christov, C. Z.; Karabencheva-Christova, T. G. Conformational Flexibility Influences Structure–Function Relationships in Nucleic Acid *N* -Methyl Demethylases. *Org. Biomol. Chem.* **2019**, *17* (8), 2223–2231.
(12) Waheed, S. O.; Chaturvedi, S. S.; Karabencheva-Christova, T. G.; Christov, C. Z. Catalytic Mechanism of Human Ten-Eleven Translocation-2 (TET2) Enzyme: Effects of Conformational Changes, Electric Field, and Mutations. *ACS Catal.* **2021**, *11* (7), 3877–3890.
(13) Hausinger, R. P. Fe(II)/α-Ketoglutarate-Dependent Hydroxylases and Related Enzymes. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39* (1), 21–68.
(14) Schofield, C. J.; Zhang, Z. Structural and Mechanistic Studies on 2-Oxoglutarate-Dependent Oxygenases and Related Enzymes. *Curr. Opin. Struct. Biol.* **1999**, *9* (6), 722–731.
(15) (a) Chaturvedi, S. S.; Ramanan, R.; Waheed, S. O.; Karabencheva-Christova, T. G.; Christov, C. Z. Structure-Function Relationships in KDM7 Histone Demethylases. In *Advances in Protein Chemistry and Structural Biology*; Elsevier, 2019; Vol. 117, pp 113–125. (b) Ramanan, R.; Waheed, S. O.; Schofield, C. J.; Christov, C. Z. What Is the Catalytic Mechanism of Enzymatic Histone N-Methyl Arginine Demethylation and Can It Be Influenced by an External Electric Field? *Chem. – Eur. J.* **2021**, *27* (46), 11827–11836.
(16) Mitchell, A. J.; Dunham, N. P.; Martinie, R. J.; Bergman, J. A.; Pollock, C. J.; Hu, K.; Allen, B. D.; Chang, W.; Silakov, A.; Bollinger, J. M.; Krebs, C.; Boal, A. K. Visualizing the Reaction Cycle in an Iron(II)- and 2-(Oxo)-Glutarate-Dependent Hydroxylase. *J. Am. Chem. Soc.* **2017**, *139* (39), 13830–13836.
(17) Waheed, S. O.; Ramanan, R.; Chaturvedi, S. S.; Lehnert, N.; Schofield, C. J.; Christov, C. Z.; Karabencheva-Christova, T. G. Role of Structural Dynamics in Selectivity and Mechanism of

Non-Heme Fe(II) and 2-Oxoglutarate-Dependent Oxygenases Involved in DNA Repair. *ACS Cent. Sci.* **2020**, *6* (5), 795–814.

(18) Chaturvedi, S. S.; Ramanan, R.; Lehnert, N.; Schofield, C. J.; Karabencheva-Christova, T. G.; Christov, C. Z. Catalysis by the Non-Heme Iron(II) Histone Demethylase PHF8 Involves Iron Center Rearrangement and Conformational Modulation of Substrate Orientation. *ACS Catal.* **2020**, *10* (2), 1195–1209.

(19) Quesne, M. G.; Latifi, R.; Gonzalez-Ovalle, L. E.; Kumar, D.; de Visser, S. P. Quantum Mechanics/Molecular Mechanics Study on the Oxygen Binding and Substrate Hydroxylation Step in AlkB Repair Enzymes. *Chem. – Eur. J.* **2014**, *20* (2), 435–446.

(20) Wójcik, A.; Radoń, M.; Borowski, T. Mechanism of O_2 Activation by α-Ketoglutarate Dependent Oxygenases Revisited. A Quantum Chemical Study. *J. Phys. Chem. A* **2016**, *120* (8), 1261–1274.

(21) Song, X.; Lu, J.; Lai, W. Mechanistic Insights into Dioxygen Activation, Oxygen Atom Exchange and Substrate Epoxidation by AsqJ Dioxygenase from Quantum Mechanical/Molecular Mechanical Calculations. *Phys. Chem. Chem. Phys.* **2017**, *19* (30), 20188–20197.

(22) Kavoosi, S.; Sudhamalla, B.; Dey, D.; Shriver, K.; Arora, S.; Sappa, S.; Islam, K. Site- and Degree-Specific C–H Oxidation on 5-Methylcytosine Homologues for Probing Active DNA Demethylation. *Chem. Sci.* **2019**, *10* (45), 10550–10555.

(23) Thornburg, L. D.; Stubbe, J. Mechanism-Based Inactivation of Thymine Hydroxylase, an .Alpha.-Ketoglutarate-Dependent Dioxygenase, by 5-Ethynyluracil. *Biochemistry* **1993**, *32* (50), 14034–14042.

(24) Thornburg, L. D.; Lai, M. T.; Wishnok, J. S.; Stubbe, J. A Non-Heme Iron Protein with Heme Tendencies: An Investigation of the Substrate Specificity of Thymine Hydroxylase. *Biochemistry* **1993**, *32* (50), 14023–14033.

(25) Ghanty, U.; Serrano, J. C.; Kohli, R. M. Harnessing Alternative Substrates to Probe TET Family Enzymes. In *TET Proteins and DNA Demethylation*; Bogdanovic, O., Vermeulen, M., Eds.; Methods in Molecular Biology; Springer US: New York, NY, 2021; Vol. 2272, pp 265–280.

(26) Fiser, A.; Šali, A. Modeller: Generation and Refinement of Homology-Based Protein Structure Models. In *Methods in Enzymology*; Elsevier, 2003; Vol. 374, pp 461–491.

(27) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K_a Predictions. *J. Chem. Theory Comput.* **2011**, *7* (2), 525–537.

(28) Case, D.A.; Betz, R.M.; Curetti, D.S.; Cheatham, T.E.; Daeden, T.A.; Duke, R.E.; Giese, T.J.; Gohlke, H.; Goetz, A.W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T.S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B. P. A. AMBER 2018. University of California: San Francisco; **2018**.

(29) 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* (2), 247–260.

(30) Li, P.; Merz, K. M. MCPB.Py: A Python Based Metal Center Parameter Builder. *J. Chem. Inf. Model.* **2016**, *56* (4), 599–604.

(31) Chaturvedi, S. S.; Ramanan, R.; Waheed, S. O.; Ainsley, J.; Evison, M.; Ames, J. M.; Schofield, C. J.; Karabencheva-Christova, T. G.; Christov, C. Z. Conformational Dynamics Underlies Different Functions of Human KDM7 Histone Demethylases. *Chem. – Eur. J.* **2019**, *25* (21), 5422–5426.
(32) Bian, K.; Lenz, S. A. P.; Tang, Q.; Chen, F.; Qi, R.; Jost, M.; Drennan, C. L.; Essigmann, J. M.; Wetmore, S. D.; Li, D. DNA Repair Enzymes ALKBH2, ALKBH3, and AlkB Oxidize 5-Methylcytosine to 5-Hydroxymethylcytosine, 5-Formylcytosine and 5-Carboxylcytosine in Vitro. *Nucleic Acids Res.* **2019**, *47* (11), 5522–5529.
(33) Solomon, E. I. Geometric and Electronic Structure Contributions to Function in Bioinorganic Chemistry: Active Sites in Non-Heme Iron Enzymes. *Inorg. Chem.* **2001**, *40* (15), 3656–3669.
(34) Solomon, E. I.; Light, K. M.; Liu, L. V.; Srnec, M.; Wong, S. D. Geometric and Electronic Structure Contributions to Function in Non-Heme Iron Enzymes. *Acc. Chem. Res.* **2013**, *46* (11), 2725–2739.
(35) Fang, D.; Lord, R. L.; Cisneros, G. A. *Ab Initio* QM/MM Calculations Show an Intersystem Crossing in the Hydrogen Abstraction Step in Dealkylation Catalyzed by AlkB. *J. Phys. Chem. B* **2013**, *117* (21), 6410–6420.
(36) Peters, M. B.; Yang, Y.; Wang, B.; Füsti-Molnár, L.; Weaver, M. N.; Merz, K. M. Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF). *J. Chem. Theory Comput.* **2010**, *6* (9), 2935–2947.
(37) 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.
(38) Davidchack, R. L.; Ouldridge, T. E.; Tretyakov, M. V. New Langevin and Gradient Thermostats for Rigid Body Dynamics. *J. Chem. Phys.* **2015**, *142* (14), 144114.
(39) Bresme, F. Equilibrium and Nonequilibrium Molecular-Dynamics Simulations of the Central Force Model of Water. *J. Chem. Phys.* **2001**, *115* (16), 7564–7574.
(40) 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* (3), 327–341.
(41) Deserno, M.; Holm, C. How to Mesh up Ewald Sums. I. A Theoretical and Numerical Comparison of Various Particle Mesh Routines. *J. Chem. Phys.* **1998**, *109* (18), 7678–7693.
(42) Götz, 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.
(43) 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.
(44) 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* (7), 3084–3095.
(45) Grant, B. J.; Rodrigues, A. P. C.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. D. Bio3d: An R Package for the Comparative Analysis of Protein Structures. *Bioinformatics* **2006**, *22* (21), 2695–2696.
(46) Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.;

Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjøvoll, M.; Fahmi, A.; Schäfer, A.; Lennartz, C. QUASI: A General Purpose Implementation of the QM/MM Approach and Its Application to Problems in Catalysis. *J. Mol. Struct. THEOCHEM* **2003**, *632* (1–3), 1–28.

(47) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic Structure Calculations on Workstation Computers: The Program System Turbomole. *Chem. Phys. Lett.* **1989**, *162* (3), 165–169.

(48) Smith, W.; Forester, T. R. DL_POLY_2.0: A General-Purpose Parallel Molecular Dynamics Simulation Package. *J. Mol. Graph.* **1996**, *14* (3), 136–141.

(49) Sousa, S. F.; Ribeiro, A. J. M.; Neves, R. P. P.; Brás, N. F.; Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Ramos, M. J. Application of Quantum Mechanics/Molecular Mechanics Methods in the Study of Enzymatic Reaction Mechanisms. *WIREs Comput. Mol. Sci.* **2017**, *7*, e1281.

(50) Kästner, J.; Carr, J. M.; Keal, T. W.; Thiel, W.; Wander, A.; Sherwood, P. DL-FIND: An Open-Source Geometry Optimizer for Atomistic Simulations. *J. Phys. Chem. A* **2009**, *113* (43), 11856–11865.

(51) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A. A consistent and accurate ab initio parametrization of density functional dispersion correlation (DFT-D) for the 94 elements H-Pu. *Phys. Chem. Chem. Phys.* **2000**, *2* (10), 2177–2186.

(52) Billeter, S. R.; Turner, A. J.; Thiel, W. Linear Scaling Geometry Optimisation and Transition State Search in Hybrid Delocalised Internal Coordinates. *J. Chem. Phys.* **2010**, *1*32, 154104.

(53) Cisneros, G. A.; Perera, L.; Schaaper, R. M.; Pedersen, L. C.; London, R. E.; Pedersen, L. G.; Darden, T. A. Reaction Mechanism of the ε Subunit of *E. Coli* DNA Polymerase III: Insights into Active Site Metal Coordination and Catalytically Significant Residues. *J. Am. Chem. Soc.* **2009**, *131* (4), 1550–1556.

(54) Torabifard, H.; Cisneros, G. A. Insight into Wild-Type and T1372E TET2-Mediated 5hmC Oxidation Using *Ab Initio* QM/MM Calculations. *Chem. Sci.* **2018**, *9* (44), 8433–8445.

(55) Leddin, E. M.; Cisneros, G. A. Comparison of DNA and RNA Substrate Effects on TET2 Structure. In *Advances in Protein Chemistry and Structural Biology*; Elsevier, 2019; Vol. 117, pp 91–112.

(56) Fu, Y.; Jia, G.; Pang, X.; Wang, R. N.; Wang, X.; Li, C. J.; Smemo, S.; Dai, Q.; Bailey, K. A.; Nobrega, M. A.; Han, K.-L.; Cui, Q.; He, C. FTO-Mediated Formation of N6-Hydroxymethyladenosine and N6-Formyladenosine in Mammalian RNA. *Nat. Commun.* **2013**, *4*, 1798.

(57) (a) Ma, G.; Zhu, W.; Su, H.; Cheng, N.; Liu, Y. Uncoupled Epimerization and Desaturation by Carbapenem Synthase: Mechanistic Insights from QM/MM Studies. *ACS Catal.* **2015**, *5* (9), 5556–5566. (b) Light, K. M.; Hangasky, J. A.; Knapp, M. J.; Solomon, E. I. Spectroscopic Studies of the Mononuclear Non-heme Fe^{II} Enzyme FIH: Second-Sphere Contributions to Reactivity. *J. Am. Chem. Soc.* **2013**, 135, 9665-9674.

(58) Su, H.; Sheng, X.; Zhu, W.; Ma, G.; Liu, Y. Mechanistic Insights into the Decoupled Desaturation and Epoxidation Catalyzed by Dioxygenase AsqJ Involved in the Biosynthesis of Quinolone Alkaloids. *ACS Catal.* **2017**, *7* (8), 5534–5543.

(59) Balsera, M. A.; Wriggers, W.; Oono, Y.; Schulten, K. Principal Component Analysis and Long Time Protein Dynamics. *J. Phys. Chem.* **1996**, *100* (7), 2567–2572.

(60) (a) Hünenberger, P. H.; Mark, A. E.; van Gunsteren, W. F. Fluctuation and Cross-Correlation Analysis of Protein Motions Observed in Nanosecond Molecular Dynamics Simulations. *J. Mol. Biol.* **1995**, *252* (4), 492–503. (b) Su, H.; Sheng, X.; Zhu, W.; Ma, G.; Liu, Y. Mechanistic Insights into the Decoupled Desaturation and Epoxidation Catalyzed by Dioxygenase AsqJ Involved in the Biosynthesis of Quinolone Alkaloids. *ACS Catal.* **2017**, *7* (8), 5534–5543.

(61) Wang, B.; Cao, Z.; Sharon, D. A.; Shaik, S. Computations Reveal a Rich Mechanistic Variation of Demethylation of *N* -Methylated DNA/RNA Nucleotides by FTO. *ACS Catal.* **2015**, *5* (12), 7077–7090.

(62) Wang, B.; Usharani, D.; Li, C.; Shaik, S. Theory Uncovers an Unusual Mechanism of DNA Repair of a Lesioned Adenine by AlkB Enzymes. *J. Am. Chem. Soc.* **2014**, *136* (39), 13895–13901.

(63) Shaik, S.; Chen, H.; Janardanan, D. Exchange-Enhanced Reactivity in Bond Activation by Metal–Oxo Enzymes and Synthetic Reagents. *Nat. Chem.* **2011**, *3* (1), 19–27.

(64) Neidig, M. L.; Decker, A.; Choroba, O. W.; Huang, F.; Kavana, M.; Moran, G. R.; Spencer, J. B.; Solomon, E. I. Spectroscopic and Electronic Structure Studies of Aromatic Electrophilic Attack and Hydrogen-Atom Abstraction by Non-Heme Iron Enzymes. *Proc. Natl. Acad. Sci.* **2006**, *103* (35), 12966–12973.

(65) Sappa, S.; Dey, D.; Sudhamalla, B.; Islam, K. Catalytic Space Engineering as a Strategy to Activate C–H Oxidation on 5-Methylcytosine in Mammalian Genome. *J. Am. Chem. Soc.* **2021**, *143* (31), 11891–11896.

(66) Sudhamalla, B.; Wang, S.; Snyder, V.; Kavoosi, S.; Arora, S.; Islam, K. Complementary Steric Engineering at the Protein–Ligand Interface for Analogue-Sensitive TET Oxygenases. *J. Am. Chem. Soc.* **2018**, *140* (32), 10263–10269.

(67) Liu, M. Y.; Torabifard, H.; Crawford, D. J.; DeNizio, J. E.; Cao, X.-J.; Garcia, B. A.; Cisneros, G. A.; Kohli, R. M. Mutations along a TET2 Active Site Scaffold Stall Oxidation at 5-Hydroxymethylcytosine. *Nat. Chem. Biol.* **2017**, *13* (2), 181–187.

(68) Shishodia, S.; Zhang, D.; El-Sagheer, A. H.; Brown, T.; Claridge, T. D. W.; Schofield, C. J.; Hopkinson, R. J. NMR analyses on N-hydroxymethylated nucleobases - implications for formaldehyde toxicity and nuclei acid demethylases. *Org. Biomol. Chem.* **2018**, 16(21), 4021-4032.

(69) Lawal, M. M.; Sanusi, Z. K.; Govender, T.; Tolufashe, G. F.; Maguire, G. E. M.; Honarparvar, B.; Kruger, H. G. Unraveling the concerted mechanism of the human immunodeficiency virus type 1 (HIV-1) protease: a hybrid QM/MM study. *Struct. Chem.* **2019**, 30, 409-417.

(70) Wallin, G.; Aqvist, J. The transition state for peptide bond formation reveals the ribosome as a water trap. *Proc. Natl. Acad. Sci USA* **2010**, 107(5), 1888-1893.

(71) Christov, C. Z.; Lodola, A.; Karabencheva-Christova, T. G.; Wan, S.; Coveney, P. V.; Mulholland, A. J. Conformational effects on the pro-S hydrogen abstraction reaction in cyclooxgenase-1: an integrated QM/MM and MD study. *Biophys. J.* **2013**, 104(5), L5-7.

(72) Claeyssens, F.; Ranaghan, K. E.; Manby, F. R.; Harvey, J. N.; Mulholland, A. J. Multiple high-level QM/MM reaction paths demonstrate transition-state stabilization in chorismate mutase: correlation of barrier height with transition-state stabilization. *Chem. Commun*. **2005**, 5068-5070.

Table of Contents Only