Catalytic Mechanism of Human Ten-Eleven Translocation -2 (TET2) Enzyme: Effects of Conformational Changes, Electric Field and Mutations

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

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

#Corresponding author: christov@mtu.edu

Abstract

Ten-eleven translocation (TET) family of enzymes are non-heme Fe(II)- and 2-oxoglutarate (2OG)-dependent oxygenases that perform oxidation of the methyl group of the 5-methylcytosine (5mC) on DNA. TET enzymes play a crucial role in epigenetic modifications and have been linked to malignant transformation and various forms of cancer such as prostate, lung, and breast cancer. In this study, Molecular Dynamics (MD) and Combined Quantum Mechanics/Molecular Mechanics (QM/MM) approaches were used to explore the catalytic mechanism, conformational dynamics, and the effects of mutations during the first oxidation from 5mC substrate to 5hmC by human TET2 enzyme. The studies reveal that a correlated motion between the main structural elements in TET2 -the glycine-serine (GS) linker and the Cys-rich N-terminal (Cys-N) subdomain plays a key role in the orientation of the DNA substrate in the wild-type (WT) TET2. This correlated motion is affected in the mutant forms of TET2. The conformational changes in the WT TET2 influence the rate of the hydrogen atom abstraction (HAT) step; however, its mechanism via σ-channel remains unchanged. The results enabled us to identify key residues that are crucial for the HAT and to delineate their crucial energy contributions and long-range correlated interactions. Notably, several remote mutations, far away from the TET2 enzymes' active site, unexpectedly exercise a substantial effect on the HAT step by (i) increasing the required activation barrier and (ii) switching the electron transfer mechanism from σ- to π-channel. Remarkably, the mutations alter the internal electric fields along the Fe=O bond that in synergy with changes in the geometric factors (e.g., the hydrogen abstraction distance and the angle) influence the reactivity of the TET2 mutant forms. The kinetic isotope effect (KIE) calculations indicate weak tunneling contributions in the WT, with variations in the mutant forms. The double mutant form K1299E-S1303N which has clinical implications in patients with refractory anemia exercises a substantial effect on the activation barrier, electric field, and the KIE. This study offers a novel insight into molecular biophysics and pathology of the human TET2 enzyme and asserts the vital effects of the protein residues in the second sphere and beyond on the catalytic process.

Keywords: TET2 enzyme, 2-oxoglutarate-dependent enzyme, non-heme iron enzymes, QM/MM, Molecular Dynamics, DNA methylation

1.0 Introduction

Methylation at cytosine DNA bases in the genome is strongly linked to various normal and pathological processes in the cell, including the regulation of gene expression, genomic imprinting, embryonic development, active DNA demethylation, oncogenesis, and other epigenetic inheritance processes.[1–7] The methylation of DNA is catalyzed by DNA methyltransferases (DNMTs), acting on cytosine and adenine bases, respectively.[3,5] TET enzymes oxidize C-methylation of cytosine bases[8,9a] and recent study show that the enzymes are also proficient as direct N-demethylases of cytosine bases at the exocyclic amine N4 position[9b] in a similar manner as other non-heme Fe(II)- and 2OG - dependent enzymes such as bacterial AlkBs and human AlkBs (AlkBH1-AlkBH8 and FTO).[8,9a] The TET enzymes possess enzymatic activity toward 5-methylcytosine (5mC), and they include TET1-3 and the AID/APOBEC subfamily of enzymes.[10a,11,12] TET proteins contain a conserved double stranded beta helix (DSBH) domain, a cysteine rich domain, and binding sites for Fe(II) and 2OG co-substrate that together form the core catalytic region in the C-terminus.[10a,13,14] [Figure 1] The overall structure of TET2 [Figure 1] is stabilized by three zinc cations (zinc fingers) coordinated by residues from both the cysteine rich and the DSBH domains, thus bringing flexible regions from the two domains together to facilitate the stability of the overall structure.[13] TET enzymes belong to the Fe(II)- and 2OG - dependent oxygenase family of enzymes, and catalyze the successive oxidation of 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and finally 5-carboxylcytosine (5caC).[10a,b,13,15] TET2 has been referred to as a *de novo* 5mC dioxygenase whose the iterative oxidation of the 5mC to the various oxidized forms is not controlled by the opposite strand cytosine-guanine dinucleotides (CpG).[10b] Crystallographic studies revealed that substrate binding by TET enzymes involves flipping of the methylated base from the DNA strand into the iron center for demethylation,[9a,13] similarly to the nucleic acids N-methyl demethylases (AlkB and its human homologs).[8,9a,16,17]

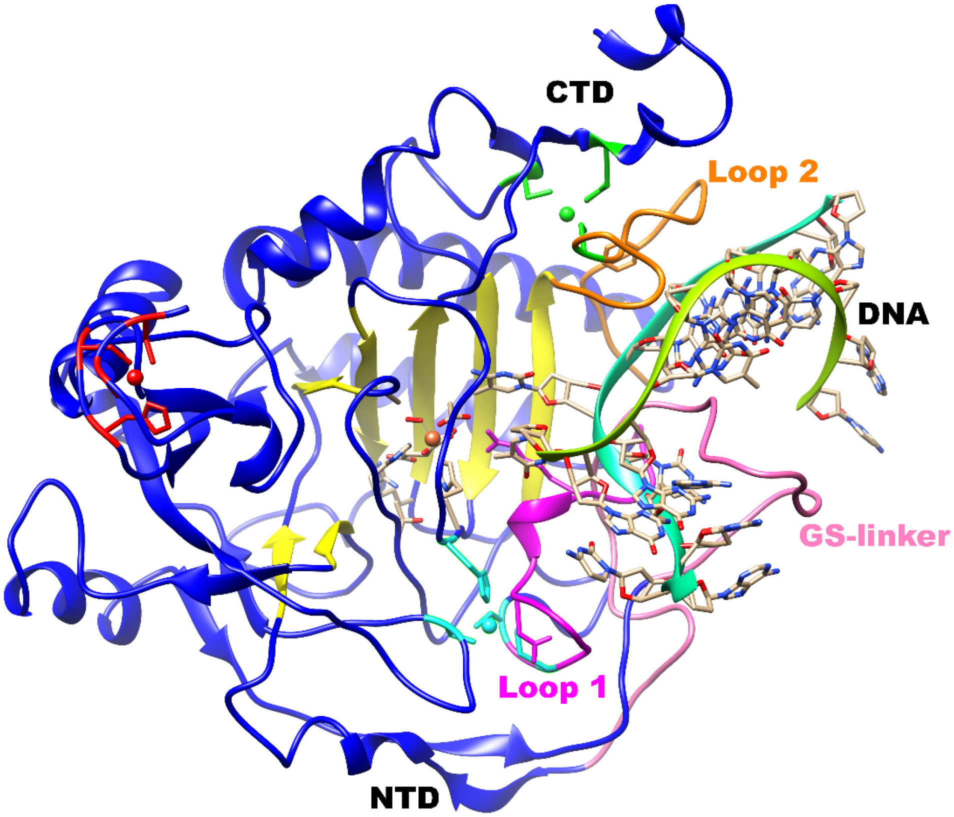
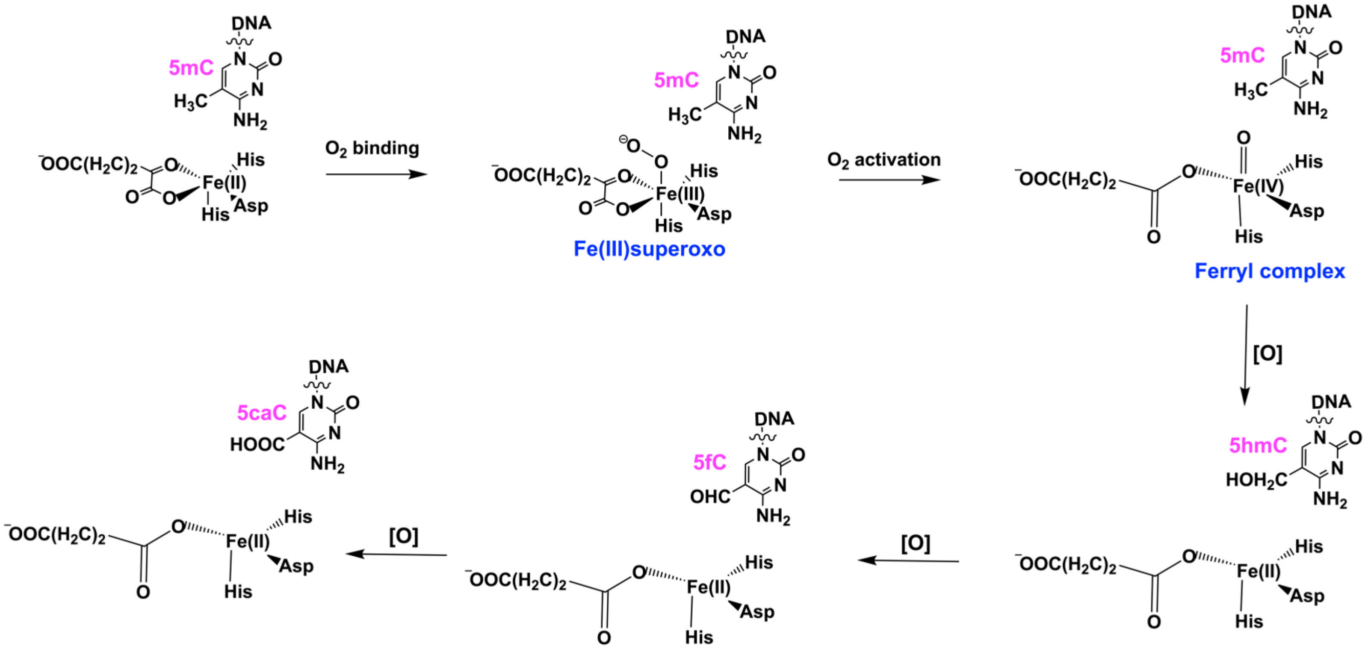


Figure 1. Average MD structure of Human TET2 bound to dsDNA derived from the Fe(III)-O_2 intermediate simulations. Coloring: Double Stranded Beta Helix (DSBH) fold (yellow), and the three zinc finger regions (Zn1 (cyan), Zn2 (red) and Zn3 (green))

The catalysis commences with dioxygen (O_2) binding to the five-coordinate Fe (II) center, and formation of the Fe(III)-superoxo complex. The first step of the reaction mechanism is a dioxygen activation and involves the formation of an active Fe(IV)-oxo (ferryl) intermediate coupled with decarboxylation of 2OG to succinate and CO_2. The formed ferryl intermediate performs hydrogen atom abstraction (HAT) from the methyl group at the C5 position of 5mC, followed by rebound hydroxylation, producing 5hmC [Scheme 1].

Oxidation of 5mC to 5hmC is the first reaction step catalyzed by TET enzymes. Some studies have shown that the oxidation reaction rate is significantly reduced for the 5hmC as it is stable in mammalian genome DNA.[15,18,19] In particular, studies reported that the hydrogen atom abstraction (HAT) process is more efficient in 5mC than 5hmC and 5fC.[15,18,19]



Scheme 1. The overall proposed reaction mechanism for successive oxidation of 5mC to 5hmC, 5fC and 5caC by TET enzymes

The catalytic mechanism of TET enzymes and nucleic acids N-methyl demethylases (NADMs) has been extensively studied both experimentally and computationally which provided an understanding of the three-dimensional structures, substrate binding, and their kinetics.[7,12,13,15,17,18,20-28] However, the effects of the conformational dynamics on the catalytic process in the flexible TET2-dsDNA complex are still poorly understood. It is particularly important to reveal how long-range correlated motions in the TET2 enzyme influence its catalysis. The current study aims to provide this missing knowledge by exploring the mechanism of the first oxidation step of 5mC dsDNA substrate to 5hmC intermediate.

Several experimental studies have shown that the TET2-dsDNA enzyme is frequently mutated in myeloid malignancies, such as myelodysplastic syndrome (MDS), chronic myelomonocytic leukemia (CMML), acute myeloid leukemia (AML), and myeloproliferative neoplasms (MPN).[2,11,29,30] Most of the leukemia-associated mutations inhibit or abolish the TET2 catalytic activity.[31] The mutations may alter active site interactions, the catalytic core domain and/or influence long-range interactions, thus affecting the TET2 activity.[11,13,32] Such clinically important mutations include the single substitution W1291A and the double substitution K1299E-S1303N [Figure 2A], causing myelodysplastic syndrome and refractory anemia, respectively.[13,32] Also, several experimental mutations performed in vitro such as Y1902A, N1387A, H1904R, S1290A-Y1295A, and M1293A-Y1294A [Figure 2A] demonstrated to influence substrate binding, catalysis, and enzyme turnover.[13] Recent studies elegantly demonstrated the power of the computational and experimental methods to provide a unique insight into the effect of a single mutation of TET2 (T1372E) on the oxidation of 5hmC to 5fC/caC.[21a,33] The QM/MM calculations by Torabifard *et al* on the effect of T1372E mutation in halting the iterative oxidation of 5hmC by TET2 revealed the orientation of the water molecule in the active site is essential for iterative oxidation in the wild-type (WT). The electron localization function (ELF) and non-covalent interaction (NCI) revealed that the process involves a hydrogen atom transfer from the hydroxyl moiety, followed by a proton transfer with the participation of the water molecule. The study found that unfavorable orientation of the substrate in T1372E mutant due to the formation of a hydrogen bond between the substrate and the mutant site which is absent in the WT resulted in the much larger energy barrier observed in the mutant.[21a] The calculations corroborate the experimental study by Liu *et al* that first reported T1372E as a “5hmC-stalling” mutant.[33] Nevertheless, the effects of the above clinical and experimental mutants on the TET2-dsDNA structure, and reaction mechanism are unknown, which is an important barrier to understanding their role and molecular pathology. Therefore, it is important to complete the knowledge gap about how mutations linked to diseases influence the long-range interactions in TET2 and thus in turn influence its reaction mechanism during the first oxidation of 5mC to 5hmC.

The present study *tests the hypothesis* that the conformational changes in the wild type (WT) TET2 influence the mechanism by affecting the active site interactions, the local electric field, and long-range correlation motions in TET2-dsDNA complex. Furthermore, we hypothesize that mutations linked to diseases influence the long-range interactions observed in WT TET2 and affect the reaction mechanism by altering key structural determinants and the internal electric field.

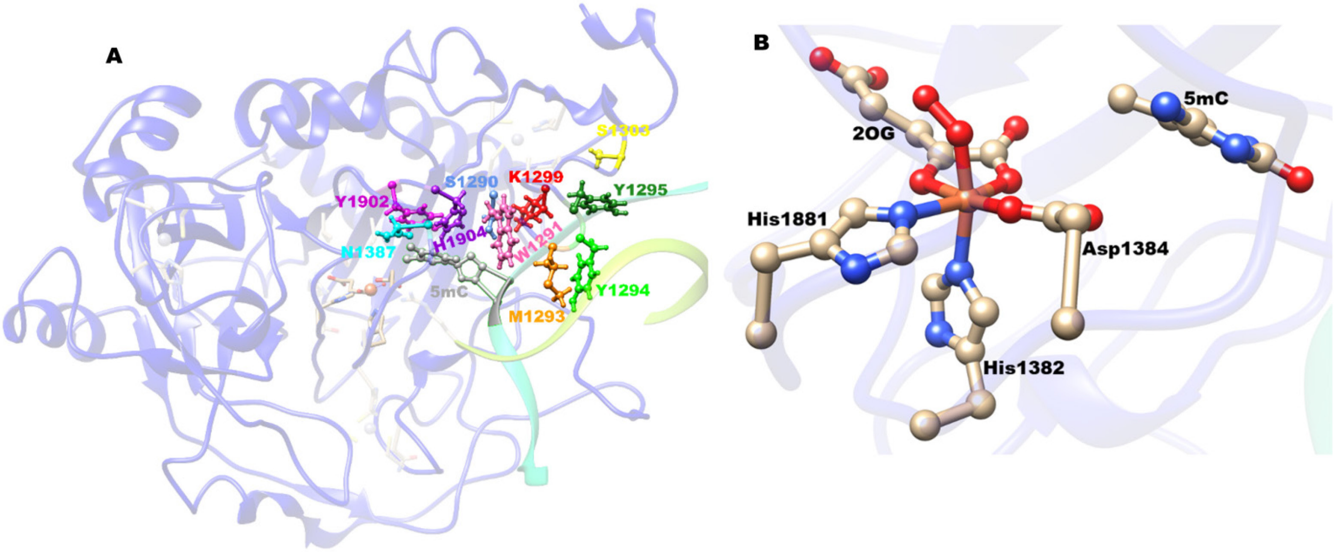


Figure 2. The structure of TET2 showing the locations of the individual substituted residues (A) and QM region used for the QM/MM calculations (B).

2.0 Computational Methods

2.1 System Preparation

X-ray crystal structure of human TET2-dsDNA (PDB code, 4NM6[13]) was used as the initial structure, and the missing residues were added using Modeller.[34] The N-oxalylglycine (NOG) used for crystallization was converted to 2OG by replacing the NH in NOG with methylene (CH_2) group using GaussView 6.0. The protonation states of the ionizable side chains were evaluated with the Propka software,[35] and the histidine residues that are coordinated to the Fe(II) center were assigned protonation states based on visual inspection of their local environments.

The active site Fe(II) is octahedrally coordinated to the co-substrate 2OG in a bidentate manner, to two histidines (His1382 and His1881), an aspartate (Asp1384) residue, and a water molecule. To generate parameters for the Fe(III)-superoxo complex, the water molecule was substituted with O_2, bound in an end on manner. The active site parameters were prepared using the Metal Center Parameter Builder (MCPB.py),[36] in Amber 18. Bond and angle force constants were derived using the Seminario method;[37] point charge parameters for the electrostatic potential were obtained using the ChgModB method. The zinc ion and its coordinating residues in the Zn finger regions were described using the Zinc Amber force field (ZAFF) method.[38] The parameters for 2OG and the substrate (5mC) were generated using the Antechamber module of Amber 18.[39] The Leap module in Amber was used to add hydrogen atoms to the protein systems and then neutralized using Na^+ counter ions. The systems were surrounded by a rectangular box solvated with TIP3P water molecules[40] within a distance of at least 10 Å from the protein's surface. Previous molecular dynamics simulations using these parameter generation procedures have successfully reproduced the geometry of non-heme iron systems.[41–43] Similar procedures were used to prepare the parameters for the ferryl complex, where the 2OG co-substrate was substituted with succinate, which was modeled with monodentate binding of the coordinating carboxylate. The parameters for all the mutants were developed by making use of the wild-type parameter after manually substituting the various mutants' residues.

2.2 Molecular Dynamic Simulations

A two-stage minimization of the geometries using MM was first performed before the MD simulations. Only water molecules and Na^+ were minimized in the first stage of the minimization, while the solute molecules were restrained with 500 $\text{kcal.mol}^{-1}\text{Å}^{-2}$ harmonic potential and all atoms were optimized without restraints during the final minimization stage. The systems were subjected to 5000 steps of steepest descent, followed by 5000 steps of conjugate gradient energy minimization.

The minimized systems were first heated for 50 ps by linearly increasing the temperature from 0 to 300 K in a canonical ensemble (NVT) using Langevin thermostat.[44] The heated systems were further subjected to constant temperature heating (at 300 K) for 1 ns in an NPT ensemble. The solute molecules were restrained with the harmonic potential of 10 kcal mol^{-1} $Å^2$ during the heating processes. After that, the systems were equilibrated within an NPT ensemble at a fixed temperature and pressure of 300 K and 1 bar, respectively, for 3 ns without any restraints on solute molecules. The MD productive runs were performed for 1 μs in an NPT ensemble with a target pressure set at 1 bar and constant pressure coupling of 2 ps. The pressure was held constant using Berendsen barostat,[45] and the SHAKE algorithm[46] was used to constrain the bond lengths of those bonds involving hydrogen atoms. The simulations were performed using the GPU version[47] of the PMEMD engine integrated with Amber18.[48] The FF14SB[49] force field was used in all the simulations, and periodic boundary conditions were employed in all simulations. Long-range electrostatic interactions were calculated using the Particle Mesh Ewald (PME) method[50] with a direct space and vdW cut-off of 10 Å.

The hydrogen bonding analysis was done using CPPTRAJ,[51] and the dynamic cross correlation analysis was done with Bio3D.[52] The binding free energy values were calculated using molecular mechanics/ generalized Born surface area (MM/GBSA)[53] method as implemented in Amber18.

2.3 QM/MM Calculations

QM/MM calculations were performed using the ChemShell package,[54] that combines Turbomole[55] (for QM region) and DL_POLY[56] (for MM region). The electronic embedding method was used to describe the interaction between the QM and MM regions. Hydrogen linked atoms were used to complete valences of bonds spanning between the two regions. The MM region was described with the Amber force field, and the QM part was represented with the unrestricted B3LYP (UB3LYP) functional. The non-heme iron center, the primary coordinating sphere residues, and the substrate-bound in the active site were included in the QM region [Figure 2B]. The QM/MM geometry optimizations were performed with the def2-SVP basis set (labeled as B1) for all the atoms. After that, linear transit scans along the reaction coordinate were performed with 0.1 Å increments to locate the transition states using DL-find optimizer.[57] Transition states were reoptimized using the partitioned rational function optimization (P-RFO) algorithm implemented in the HDLC code.[58] The fully optimized geometries of the minima and the transition states were characterized via frequency calculations. The energies of the optimized stationary points were recalculated via single-point energy calculation using a larger basis set, def2-TZVP (labeled as B2) for all the atoms. Electric field calculations were done using TITAN code[59] as in other studies.[60,61] Energy decomposition analysis (EDA)[21a,b] calculations were then carried out on the optimized reactants, and transition states geometries to determine the non-bonded interactions of all the residues.

The primary kinetic isotope effects (KIEs) were computed and compared with the experiment. The tunneling correction was considered during the calculations using the zero-curvature tunneling (ZCT) method as implemented in ChemShell. Hessian calculations were carried out in DL-FIND using ChemShell. The rate constant and KIEs were calculated on both deuterated and non-deuterated systems using transition state theory (TST).

3.0 Results and Discussion

3.1 Conformational Dynamics Facilitates the Formation of a Reactive Complex between TET2 Fe(IV)=O and the 5mC (dsDNA) substrate (TET2·Fe(IV)=O/dsDNA)

The active oxidizing species necessary for the oxidation of the methylated substrates in non-heme Fe(II) and 2OG-dependent enzymes is the ferryl intermediate. We performed MD simulations to explore the role of the structural dynamics of this reactive complex. TET2-Fe(IV)=O forms a stable complex with dsDNA (an average RMSD of 2.57 Å) [Figure 3A]. The plots of the time-dependent fluctuations of the distance between the oxygen (O) from the oxo group of Fe(IV)=O and the methyl carbon (C) of the substrate [Figure 3B], and the Fe-O-C angle [Figure 3C], which determine the efficiency of the substrate oxidation showed average values of 3.66 Å and 143.32°, respectively. The proper orientation of the substrates' 5-

methylcytosine ring with respect to the iron-oxygen bond of Fe(IV)=O is stabilized by a stacking interaction with the phenyl ring of Tyr1902. The dsDNA substrate in the TET2·Fe(IV)=O/dsDNA complex exhibits similar shape and orientation as in the crystal structure [Figure S3] supported by a hydrogen bonding network with Ser1828 (72.4%), His1904 (60.2%), and Arg1262 (64.6%). The substrate and the Fe(IV) center coordinated ligands are stabilized by a series of second sphere residues via hydrogen bonding interactions. For example, the non-coordinating (C4) succinate carboxylate group is stabilized by hydrogen bonding interactions with Arg1896 (89.6%), and Ser1898 (54.9%). In contrast, in the human AlkBH2-dsDNA (a human non-heme Fe(II)- and 2OG-dependent enzyme that demethylates N-methylated bases in dsDNA), the C4 succinate carboxylate is stabilized by hydrogen bonds with Arg248 (84%) and Tyr161 (84%).[22] The lower stability of the hydrogen bond with Ser1898 in TET2 compared to the one with Tyr161 in AlkBH2 might lead to more flexible succinate binding in TET2-dsDNA when compared to AlkBH2-dsDNA. The backbones of the Fe-coordinating histidines (His1382 and His1881) are stabilized additionally via hydrogen bonding interactions (87.8%) with each other. In contrast to AlkBH2-dsDNA, in TET2-dsDNA the non-coordinating oxygen of the aspartate ligand (Asp1384) is stabilized by Asn1387 -23% of the trajectory and by a stable solvent-mediated hydrogen bonding interaction (74.7%).

Comparing the MD simulations and the crystal structures of TET2-dsDNA with AlkBH2-dsDNA reveals key differences in the DNA interactions and binding in both enzymes[13,17,22] (The detailed comparison is presented in the SI).

The principal component analysis (PCA) [Figure 4], which shows the main direction of motions in a protein[62,63a] reveals that the cysteine-rich N-terminal (Cys-N) region, the GS linker region, loop 2 (L2), and the DNA are flexible regions in TET2·Fe(IV)=O/dsDNA. The Cys-N region moves toward the DSBH core for more significant interaction and better stability. The L2 and GS linker residues move towards DNA and the L2 forms scissors around DNA. These motions collectively enhance the binding of the DNA to the protein. The dynamic cross-correlation provides insights into the collective, correlated motions of the different domains of the proteins. Long-range correlated motions have been implicated in substrate binding, allosteric regulation, product release and protein folding.[63b,c] The DCCA map illustrates the covariance matrix of α carbon atoms and ranges from +1 to -1, denoting correlated and anti-correlated motions, respectively. The dynamic cross-correlation analysis (DCCA) [Figure 5], of the flexible regions observed in the PCA reveals that the residues that contribute to the motion in the Cys-N region have a positive correlation with the Zn2 finger and its four coordinating residues (Cys1193, Cys1271, Cys1273, and His1380), and two β- sheets from the DSBH core (β8 and β17), which are directly opposite to the Fe center and its first coordinating sphere residues.

Furthermore, the residues from the L2 and GS linker regions of the protein, which participate in the essential motion with the DNA substrate show a positive correlation with each other, the Zn3 finger, and its coordinating residues as well as DNA. The GS linker also shows a positive correlation with the loop 1 (L1) region, and the iron coordinating HxD loop. Even though the function of the GS linker remains elusive, its correlated motions with L1, L2, DNA, Zn3 finger, and the iron-coordinating HxD loop suggests about its importance in DNA binding. Overall, the long-range correlated motions observed in the WT ferryl complex might aid catalysis as well as DNA interactions and binding, resulting in stabilizing the overall structure of the WT TET2-dsDNA complex. In comparison, the studies on the AlkBH2-dsDNA complex show main direction of motions at the hydrophobic β-hairpin region which moves toward the Fe center, resulting in a complex that likely favor catalysis.[22]

3.2 Single/Double Mutations in the Second Sphere and Beyond Influence the Substrate Binding in TET2·Fe(IV)=O/dsDNA Complex

TET2 mutations influence substrate binding, catalysis, and DNA interactions,[13] and can also result in various diseases such as myelodysplastic syndrome and refractory anemia.[13,32] The TET2 mutations subject of this study have either been performed experimentally in vitro or clinically related to diseases. They include second sphere substitutions (W1291A, N1387A, Y1902A, and H1904R), substitutions in remote to the active site regions (M1293A-Y1294A, and K1299E-S1303N), or a combination of both (S1290A-Y1295A) [Figure 2A]. Mutant forms W1291A and K1299E-S1303N are mutations clinically linked to myelodysplastic syndrome and refractory anemia, respectively,[13,32] while the rest of the studied mutant forms are experimentally-performed in vitro mutations found to influence the substrate binding and the TET2 enzymatic activity.[13] The distance between the oxo group (O) of the Fe(IV)=O intermediate and the alpha carbon (CA) of the respective substituted residue varies between 9.2 and 25.9 Å in the mutant forms of TET2· Fe(IV)=O/dsDNA (Table S1). To understand how these mutant forms affect the binding of dsDNA substrate to TET2, we performed a series of 1 μs MD simulations of all the seven single and double mutants: H1904R, K1299E-S1303N, M1293A-Y1294A, N1387A, S1290A-Y1295A, W1291A, and Y1902A. The mutation sites are located close to the 5mC dsDNA substrate binding region and loop2 (L2), as depicted in Figure 2A. The simulations show stable systems with average RMSDs [Figure 3A] between 2.61 and 3.35 Å (Table S2) compared to 2.57 Å in the WT TET2, suggesting some structural perturbations upon the mutation with the most profound effect in M1293A-Y1294A and W1291A mutant forms.

3.2.1. Effects of the Mutations on the Structure of the Substrate Binding Site and the Iron Center

The average O - C distances [Figure 3B] varies between 3.78 and 11.27 Å (Table S2) while the average Fe-O-C angles [Figure 3C] ranges between 111.56 and 140.06° (Table S2) in comparison with 3.66 Å and 143.32°, respectively, for the WT TET2·Fe(IV)=O/dsDNA. The changes in the key HAT geometric parameters (distance and angle) in the mutant forms can lead to changes in the rate and mechanism of the HAT reaction. Furthermore, in all mutant forms, the substrate participates in a smaller number of hydrogen bonds when compared to the WT, suggesting that the residue substitutions might influence its stability. For example, while in the WT TET2·Fe(IV)=O/dsDNA complex, the substrate is stabilized by interactions with Ser1828 (72.4%), His1902 (60.2%), and Arg1262 (64.6%), in K1299E-S1303N, N1387A, and S1290A-Y1295A mutants, the substrate is supported by hydrogen bonds with Ser1828 (25.1%, 43.4%, and 32.6%, respectively). In contrast in S1290A-Y1295A and Y1902A mutants, the hydrogen bonds between the substrate and Arg1262 are found to be 21.6% and 38.4%, respectively. The N1387A mutant form shows a hydrogen bonding interaction between the substrate and His1904 in 42.4% of the MD trajectories. Interestingly, the substitute Arg in the H1904R mutant stabilized the substrate (89.9%) more than His1904 in the WT, however other hydrogen bonding interactions observed in the WT are lost in this mutant. All these results indicate weaker stabilization of the substrate in the mutant forms when compared to the WT. The mutations also influence the overall stability of the iron center where the non-coordinating (C4) succinate carboxylate group is stabilized by a hydrogen bond with Arg1896 with a presence between 15.3% and 42.3% (Table S2) and Ser1898 (between 11.8% and 35.8%) (Table S2) in comparison to WT values of 89.6% and 54.9%, respectively. Also, the hydrogen bonding interaction between the two coordinating histidines (His1382 and His1881) backbones varies between 14.3% and 61.6% in comparison to 87.8% in the WT (Table S2). The above interactions are weaker when compared to the observed in the WT. This indicates that in both mutant forms the substrates are less stabilized in the active sites.

3.2.2. Effects of Mutations on the Substrate Binding Affinity and Long-Range Correlated Motions

The effects of the mutations on the stability of the TET2·Fe(IV)=O/dsDNA complex were further confirmed by calculating the binding free energies between the substrate and the protein using MM/GBSA.[53] The binding free energies of the mutant forms suggest a weaker substrate binding (see the SI) and vary between -108.43 kcal/mol and -63.98 kcal/mol (Table S2) in reference to -122.09 kcal/mol in the WT. The most profound effect is observed in the mutant form W1291A. Even though there are variations in the stability of the active site of the protein-DNA complexes, in the mutant forms and the WT, the radius of gyration (ROG) [Figure 3D] shows a similar volume of TET2 enzyme for all WT and mutation forms. This indicates that the mutations only influence the delicate interactions in the tertiary structure of TET2 and dsDNA, however, does not lead to a more dramatic disruption of the overall structure.

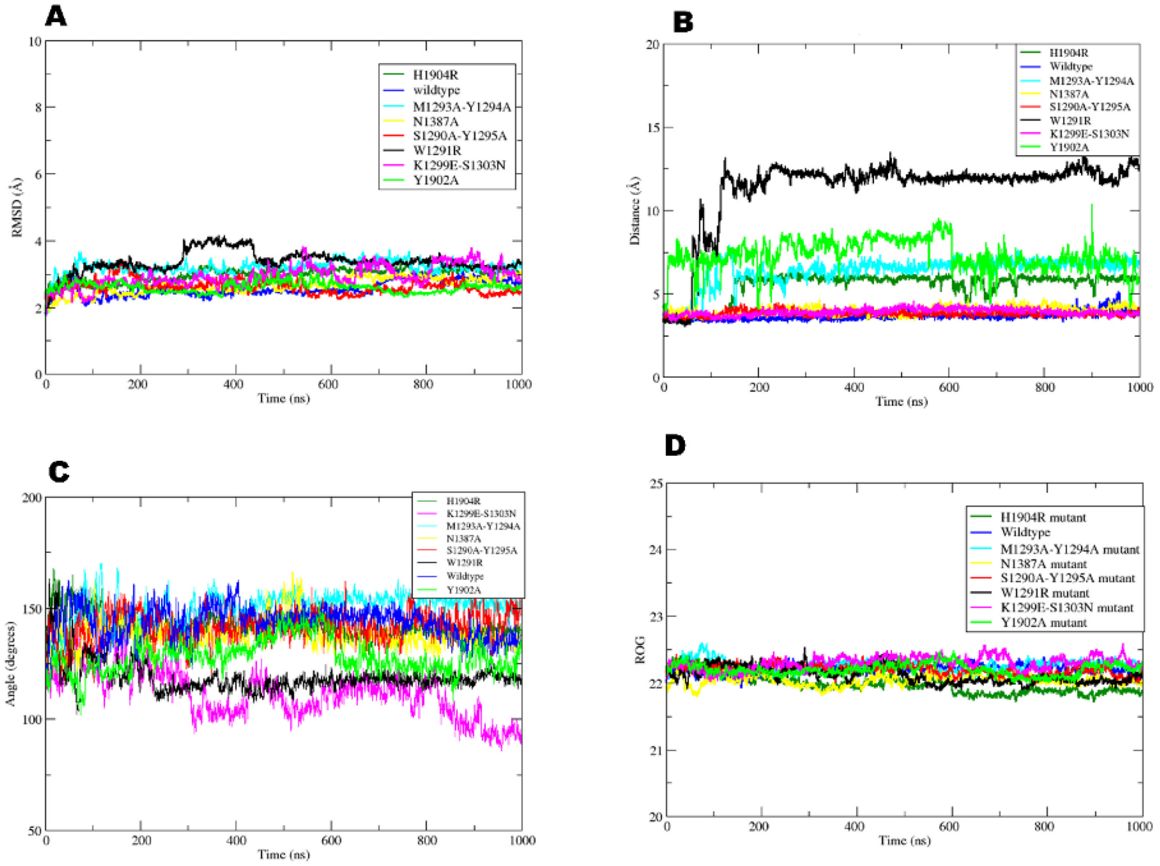


Figure 3. Plots of RMSD (A), distance between oxo group (O) of the ferryl and the substrate methyl carbon (B), angle between Fe, O and the substrate methyl carbon (C), and the radius of gyration (D) of both the WT TET2·Fe(IV)=O/dsDNA and the mutants.

The PCA of the double mutant forms [Figure 4] K1299E-S1303N, M1293A-Y1294A, and S1290A-Y1295A show a motion in the GS linker region, which moves it *away* from the DNA in K1299E-S1303N and S1290A-Y1295A forms, thus weakening the DNA-linker interactions. Interestingly, in the M1293A-Y1294A mutant form, the GS linker moves *towards* DNA as in the WT complex, thus enhancing the interactions of DNA with the protein. DCCA [Figure S7-S9] of the GS linker in these mutant forms reveals a positive correlation with the Cys-N region in M1293A-Y1294A mutant, while in K1299E-S1303N and S1290A-Y1295A mutants, the GS linker has a positive correlation with β12 in the GS linker vicinity. The

Cys-N region harbors the Zn1 finger region, which stabilizes the overall protein structure while the β12 only provides support for the GS linker region. Thus, the long-range correlated motion observed in the M1293A-Y1294A mutant form helps for the stabilization of this complex more than the K1299E-S1303N and S1290A-Y1295A. Besides, the PCA shows that the Cys-N region moves toward the protein (in M1293A-Y1294A) and away from it (in S1290A-Y1295A) supporting the DCCA results.

In the single mutant forms [Figure 4, S10-S13], Y1902A, and H1904R, the GS linker becomes more rigid. In contrast, in N1387A and W1291A, the linker remains flexible with the direction of motion towards the DNA and positively correlated with the DNA interacting loop 1 (L1). These correlated motions might enhance DNA stabilization. Also, in N1387A and Y1902A mutant forms, there is reduced flexibility around the Cys-N region, and this region becomes completely rigid in W1291A. However, in H1904R, the Cys-N region moves away from the DSBH, resulting in weakening the interactions within the DSBH region. The substitutions N1387A and H1904R show additional motions at β13 and α6, respectively, which all have a positive correlation with β12 that supports the GS linker.

Overall, DCCA and PCA analyses of the WT TET2·Fe(IV)=O/dsDNA and the mutant forms reveal that the correlated motions with the participation of the GS linker and Cys-N regions contribute less to the DNA binding to the enzyme in the mutant forms than in the WT.

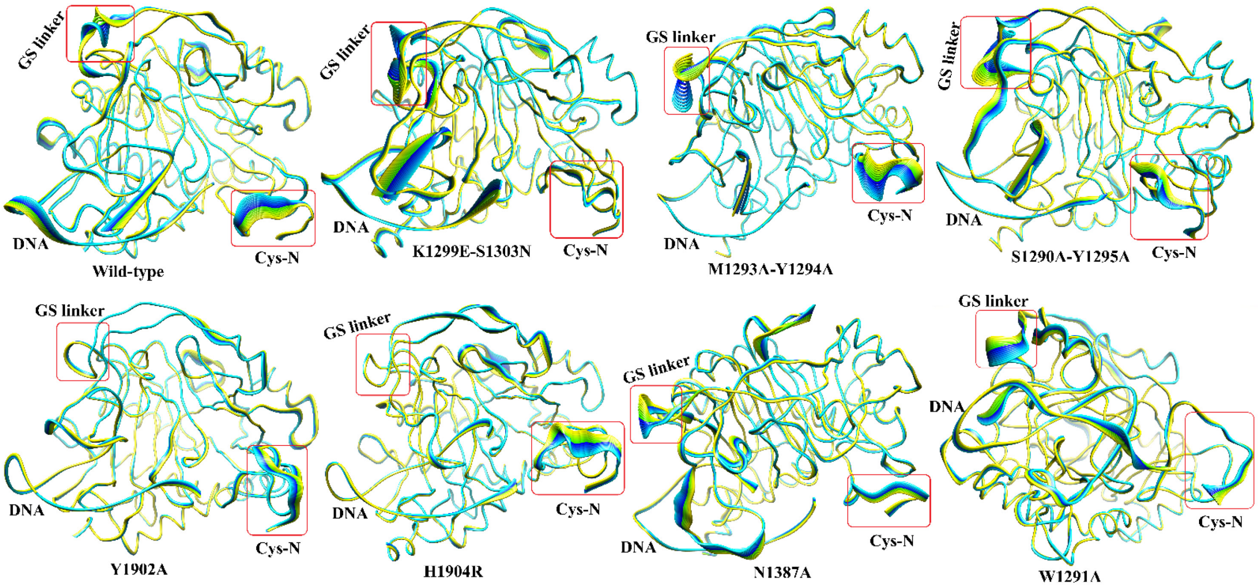


Figure 4. Principal component analysis for the WT TET2·Fe(IV)=O/dsDNA and its various mutants. Residues numbers are as follows: 1-445 (protein), 446-448 (Zn), 449 (Fe), 450 (O), 451 (succinate), 452-475 (DNA). Yellow to blue represents the direction of motion of residues.

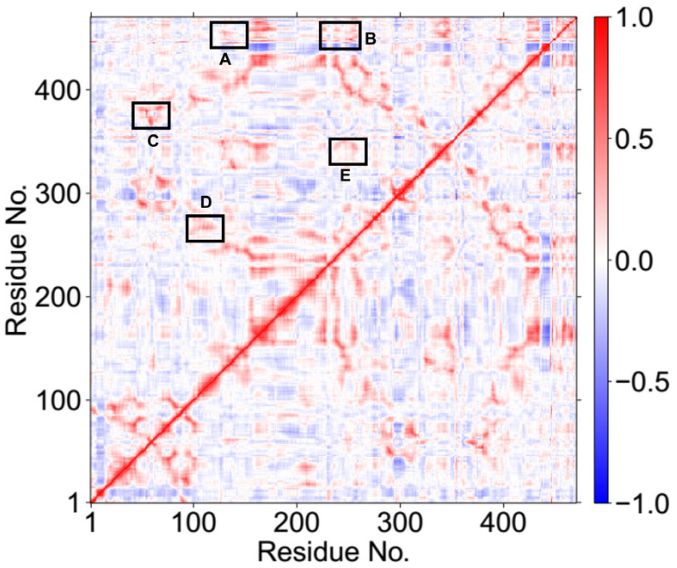


Figure 5. Dynamics cross correlation for the WT TET2·Fe(IV)=O/dsDNA MD simulations. Residue numbers 1-445 (protein), 446-448 (Zn), 449 (Fe), 450 (O), 451 (succinate), 452-475 (DNA). Boxes in the figure denote: (A) L2 residues show correlation with the DNA. (B) GS linker residues have

correlation with the DNA. (C) Correlation of Zn3 coordinating site with L1. (D) L1 residues have correlation with GS linker residues. (E) Iron coordinating HxD loop residues show correlation with GS linker residues.

3.3 Mechanism of the Hydrogen Atom Abstraction (HAT) in the WT TET2-dsDNA

HAT involves the abstraction of a hydrogen atom from the methyl group of the substrate by the electrophilic Fe(IV)=O intermediate, resulting in Fe(III)-OH center and a reactive radical formation at the methylene carbon of the substrate.[8,9,64] This step has been reported as the rate-limiting one during the substrate oxidation by Fe(II)/2OG dependent enzymes.[21a-27,65,66a] Studies have shown that the relative structure of ferryl to the substrates determine the reaction and the electron transfer mechanism.[66b-e] We used five snapshots from MD to explore this mechanistic step. The calculations were done at the quintet state of the Fe(IV)=O intermediate as it was previously demonstrated that non-heme Fe (II)- and 2OG-dependent enzymes favor this spin state.[27,67–70] The results revealed that the reaction barrier varies from 16.3 to 19.1 kcal/mol [Table S3] at the B2+ZPE level of theory, with Boltzmann weighted averaged[22,71] value of 17.1 kcal/mol, which is within the range of previously reported barriers for HAT for 2OG dependent oxygenases.[24-28,65,72] The TS_{HAT} Fe-O-H angle varies from 149.4 to 160.1°, suggesting that in all reaction paths, an α-electron is transferred from the substrate to the vacant 3d-orbital of the Fe. This is further confirmed by the TS_{HAT} spin density values of the methyl carbon of the substrate, which vary between -0.322 and -0.253, showing that a β-electron is left on the substrate and the reaction proceeds via *σ-pathway* [Figure 6]. Recent computational studies on other DNA repair enzymes, AlkB and AlkBH2, bound to duplex DNA, have shown that the HAT proceeds via the same pathway.[22] The TS_{HAT} geometry is stabilized by several key interactions and in particular Arg1261 stabilizes the non-coordinating oxygen of the succinate. A network of hydrophobic interactions including residues Val1900, Val1395, Leu1872, and Thr1393 surrounding the Fe center and the methyl group of the substrate enhance the stabilization of the TS_{HAT}. Besides, the cytosine ring of the substrate is stabilized by stacking interaction with Tyr1902. This substrate stabilization and proper orientation are further enhanced by stacking interaction of the imidazole group of His1904 and the phenyl ring of Tyr1902 and a hydrogen bonding interaction between N3 of the substrate and NH group of His1904. Further, the exocyclic nitrogen (N4) of the substrate is stabilized via a hydrogen bonding interaction with Asn1387, enhancing its stability in the WT.

HAT results in the formation of the Fe(III)-OH complex, and methylene substrate radical with an O-H distance of 0.97 Å, and the Fe-O is elongated from 1.62 to 1.85 Å. This elongation indicates the formation of the ferric-hydroxo complex and the spin density of 4.23 indicates about +3 oxidation of the Fe at this reaction state. The formed intermediate is almost thermoneutral with energy of -0.4 kcal/mol at the B2+ZPE level of theory, unlike in AlkBH2-dsDNA, where the formed Fe(III)-OH intermediate is slightly endothermic with energy of 6.5 kcal/mol at the B2+ZPE level of theory.[22] This implies that the ferric-hydroxo intermediate is more energetically favorable and stable in TET2 than in AlkBH2.

3.3.1 Energy Decomposition Analysis and Long-Range Correlated Motions of the HAT-Transition State-Stabilizing Residues

To gain insights into the role of the individual residues in the reaction steps during catalysis, we performed energy decomposition analysis (EDA) as developed by Cisneros[21a,b] on the QM/MM optimized RC and TS_{HAT} geometries. This was achieved by calculating the differences in the nonbonded interaction energies (i.e., Coulomb and van der Waals interaction energies) between the individual residues of MM environment and the QM for the RC and the TS structures.[21a-d] This analysis is only obtained in an approximate manner and thus provides a qualitative assessment of individual residues in the reaction steps.[21c,d] A negative energy difference between the TS and the RC structures refers to a stabilizing contribution to the TS and vice versa. EDA has been employed in various QM/MM and MD simulations to study numerous protein systems.[21a-f] The results as presented in Table 1 show that residues Thr1393, Val1395, Leu1872, and Arg1261, which take part in the stabilization of the TS_{HAT} when compared to the RC in the WT TET2·Fe(IV)=O/dsDNA, contribute with energies of -1.36, -0.82, -0.60, and -4.46 kcal/mol, respectively.

The DCCA of the residues that stabilize the TS_{HAT} shows that residues that are involved in hydrophobic interactions have a positive correlation with the DNA-interacting loops residues (L1), the succinate-stabilizing residues, the Fe center, and the loop (iron-chelating motif 1) from the DSBH core [Figure 4]. Also, Arg1261 that stabilizes the succinate non-coordinating oxygen has a positive correlation with both the DNA-interacting loops (L1 and L2), a loop that harbors the coordinating His1881 and GS linker residues. Asn1387 involved in stabilizing the exocyclic nitrogen of the substrate has a positive correlation with the Fe center and its oxo group, implying that this residue might be important in catalysis. Tyr1902 also has a positive correlation with GS linker residues, the loop containing iron HxD motif, the third zinc finger (Zn3), and its coordinating residues (Cys1289, Cys1298, Cys1358, and His1912). The DNA-interacting loops are crucial for DNA binding and the Zn finger is important for the overall structural stabilization.[13] These results show that complex correlation motions between the residues that stabilize the TS_{HAT} with loop 1, loop 2, Zn3 finger, and GS linker regions, ensure the overall stabilization of the transition state and the metal center. In comparison, in AlkBH2, the HAT transition state is stabilized by residues Thr252, Arg110, Tyr122, Arg254, and Glu175, which all have a correlation with the iron center and its coordinating residues and with the substrate recognition lid, suggesting that modification of the second sphere residues in the substrate binding lid could influence HAT in AlkBH2[22] but not in the TET2.

3.3.2 Influence of Electric Field and Kinetic Isotope Effects on the HAT

Recent studies have shown that the internal electric fields along the reaction coordinate in small molecules and enzymes can be an important factor that defines the reaction mechanism and specificity.[59-61,73] To evaluate how the conformational changes involved in HAT can influence the electric field, we performed calculations of the local electric field along the direction of Fe=O bond in the ferryl complex reactant geometries using TITAN.[59] The results, as shown in Table S4, reveal that in all the five snapshots used for the WT calculations, the value of the electric field intensity exhibits slight variations as a function of the conformational changes in the RCs. Importantly in RCs with similar HAT distance and angle, it is the change in the electric field that contributes to the differences in the activation energies of the reaction paths.

For example, snapshots 1 and 4 with almost the same distance between the oxo group oxygen (O) and the hydrogen atom (H) show an electric field intensity of -0.0479 au for snapshot 4, which is -0.0034 au lower than in snapshot 1 with the lowest barrier. The higher barrier observed in snapshot 4, although its favorable electric field value could be due to the constraint in Fe-O-H angle (127.8°) in snapshot 4 compared to the corresponding value of 149.7° in snapshot 1. The calculated electric field values of the transition states (Table S4) follow the similar trend as the various reactant complexes but becomes less negative at this reaction state in all cases.

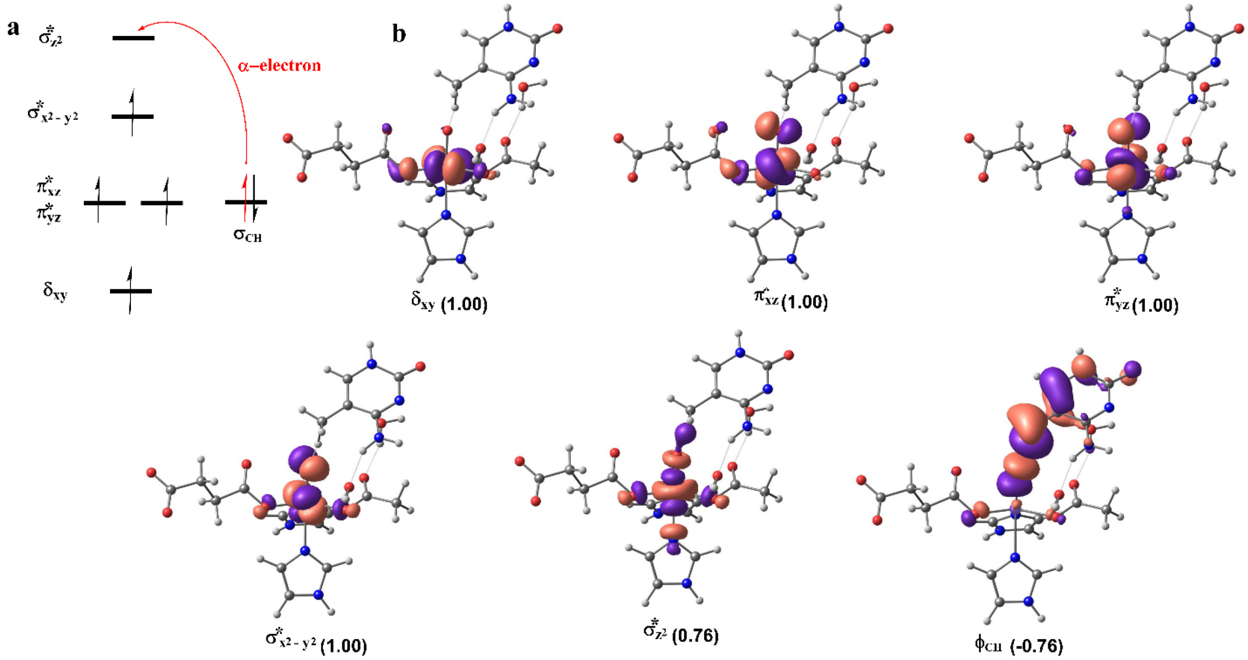


Figure 6. Electron shift diagram (a) and the spin natural orbitals (SNOs) (b) with their respective populations (in parentheses) during HAT in the WT TET2·Fe(IV)=O/dsDNA, calculated from the transition state structure.

The hydrogen atom abstraction barriers in enzymes can exhibit a kinetic isotope effect (KIE) when the transferred hydrogen atom is substituted with a heavier isotope active site dynamics.[72] To estimate the KIE contributions and to evaluate the tunneling effect on the system, we performed primary KIE calculations by replacing the transferred hydrogen atom with deuterium as tunneling is very sensitive to the mass of the tunneling particle. The calculated KIE value of 8.55 at 303 K for the WT is in very good agreement with the experimental value[74] of 9 for KIE in TET2 and also with other observed experimental KIE for Fe/2OG enzyme substrates[15]. The calculated KIE indicates a low tunneling contribution in the HAT reaction in the WT TET2.

3.4 Mechanism of the Rebound Hydroxylation in the WT TET2-dsDNA

After the formation of Fe(III)-OH intermediate and substrate radical, the intermediate undergoes a rebound reaction, where the hydroxyl group (OH) of the ferric-hydroxo intermediate is transferred to the methylene radical, leading to the formation of the hydroxylated product. The process passes through a TS_{Reb} with a barrier of 10.1 kcal/mol, including ZPE correction. The rebound TS barrier in TET2 is 2.5 kcal/mol lower than the barrier obtained for another human DNA repair enzyme, AlkBH2,[22] indicating that the process of rebounding of the hydroxyl group from the Fe(III)-OH to the substrate radical is faster in TET2 than in AlkBH2. The observed lower TS barrier in TET2 could be due to a better orientation of the Fe-OH to methylene carbon radical (O_p-C distance) compared to AlkBH2. The rebound process occurs faster than the HAT, confirming the rate-limiting nature of the HAT in substrate oxidation. The rebound hydroxylation results in a highly stable product (PD) with energy of -24.9 kcal/mol. The optimized geometries of the stationary points and the potential energy profile are presented in Figures 7 and 8, respectively.

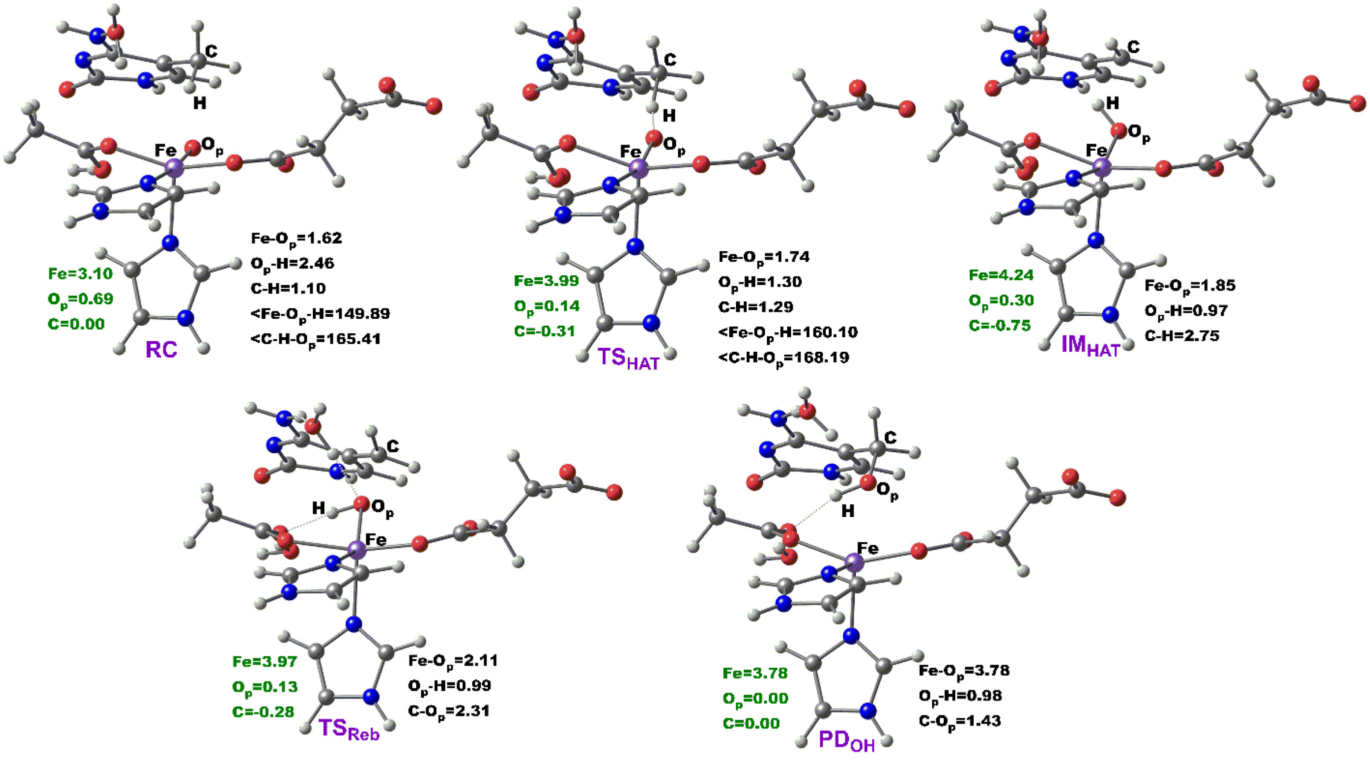


Figure 7. Stationary points geometries along the HAT and Rebound hydroxylation steps in the WT TET2·Fe(IV)=O/dsDNA. Distances (Å) and spin densities are in black and green, respectively.

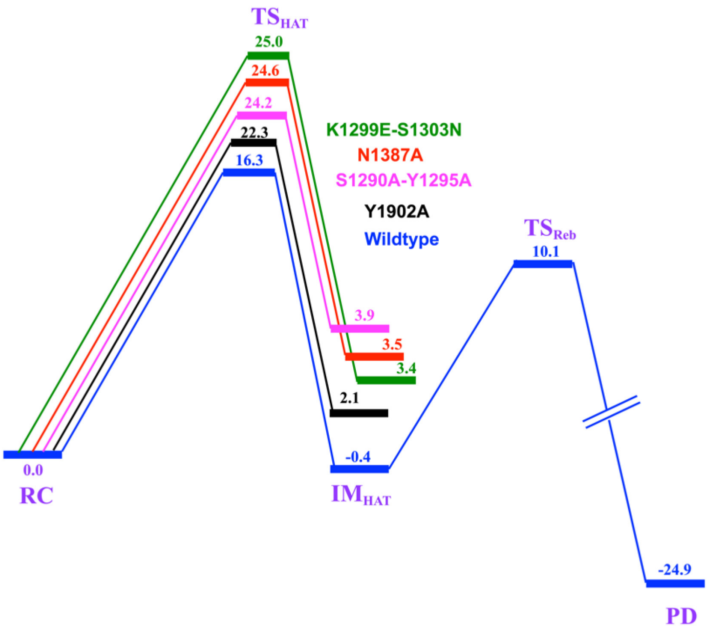


Figure 8. The energetic reaction profile for the hydroxylation of 5-methylcytosine substrate by TET2 as calculated at UB3LYP/def2-TZVP+ZPE level. The relative energies are in kcal/mol.

3.5 Effect of Mutations on Hydrogen Atom Abstraction Reaction of TET2·Fe(IV)=O/dsDNA Complex

To quantify the effect of the mutations on the rate-limiting HAT step of the reaction mechanism of the oxidation of 5mC to 5hmC by human TET2 enzyme, we performed QM/MM calculations on K1299E-S1303N, S1290A-Y1295A, Y1902A, and N1387A. In these mutant forms, the distance between the oxo-group of Fe(IV)=O intermediate and the substrate's methyl carbon is favorable for substrate oxidation.[75] The calculations revealed barriers of 25.0 kcal/mol, 24.2 kcal/mol, 22.3 kcal/mol, and 24.6 kcal/mol at the B2+ZPE level of theory for K1299E-S1303N, S1290A-Y1295A, Y1902A, N1387A mutants, respectively, in comparison to 16.3 kcal/mol obtained for the WT. The activation barriers for the mutant forms are higher than the one obtained for the WT, implying that the mutations decrease the rate of the rate-limiting HAT step in comparison to the WT complex. In the optimized TS_{HAT} structures, the Fe-O-H angles are 113.7, 145.3, 149.5, and 146.1° for K1299E-S1303N, S1290A-Y1295A, Y1902A, N1387A, respectively, in comparison to 160.10° for the WT. Lower Fe-O-H angle value in the K1299E-S1303N suggests for HAT proceeding via a π-pathway, where a β-electron is transferred into the 3d-orbital of Fe, leaving an α-electron on the substrate (Figure 9). In the remaining three mutants, HAT proceeds via a σ-pathway with an α-electron transfer to the 3d-orbital of Fe, similar to the WT, where the Fe-O-H angle and spin density at TS_{HAT} are 160.10° and -0.312, respectively. These observations are validated by the methyl carbon spin density values of 0.365, -0.260, -0.294, and -0.291 for K1299E-S1303N, S1290A-Y1295A, Y1902A, N1387A, respectively. The results demonstrate that residue substitution has a strong influence on the HAT transfer mechanism, switching the nature of the overlapping MOs.

The HAT leads to the formation of the Fe(III)-OH complex intermediate with reaction energies of 3.4, 3.9, 2.1, and 3.5 kcal/mol at B2+ZPE level of theory for K1299E-S1303N, S1290A-Y1295A, Y1902A, N1387A mutants, respectively, in respect to -0.4 kcal/mol obtained for the WT complex. This clearly shows that the formation of the ferric-hydroxo intermediate is slightly endothermic in the mutants and the formed products are less stable than in the WT. Figure 8 and (Figures 10 and S14) show the potential energy profile and the optimized geometries of the stationary points, respectively.

3.5.1 Stabilization of the HAT Transition State in the Mutant Forms: Local Interactions, Energy Decomposition and Long-Range Correlated Motions

In all the mutant forms, the TS_{HAT} is stabilized by hydrophobic interactions between Val1900, Val1395, and Thr1393 around the Fe center and the substrate's methyl group unlike in the WT where Leu1872 is inclusive. In K1299E-S1303N and N1387A, the stacking interactions between the substrate's cytosine ring and Tyr1902 are preserved as in the WT complex. However, the extra stacking interactions between His1904 and Tyr1902, which enhance the substrate stabilization are lost in these mutant forms, and the His1904 imidazole group is re-oriented to form hydrogen bonding interaction with the O2 of the substrate in the N1387A mutant. In the Y1902A mutant form, the substrate is very unstable due to the substitution Tyr1902Ala, which causes the loss of extra stacking interaction between the Tyr and the imidazole group of His1904 and thus affect the orientation of the substrate (Figure S15). The substituted Tyr1902 residue helps in the stabilization of the substrate's cytosine ring in the WT complex, and its substitution to Ala weakens the substrate stability, with Ala forming stable hydrophobic interaction with a nearby Thr1393 residue.

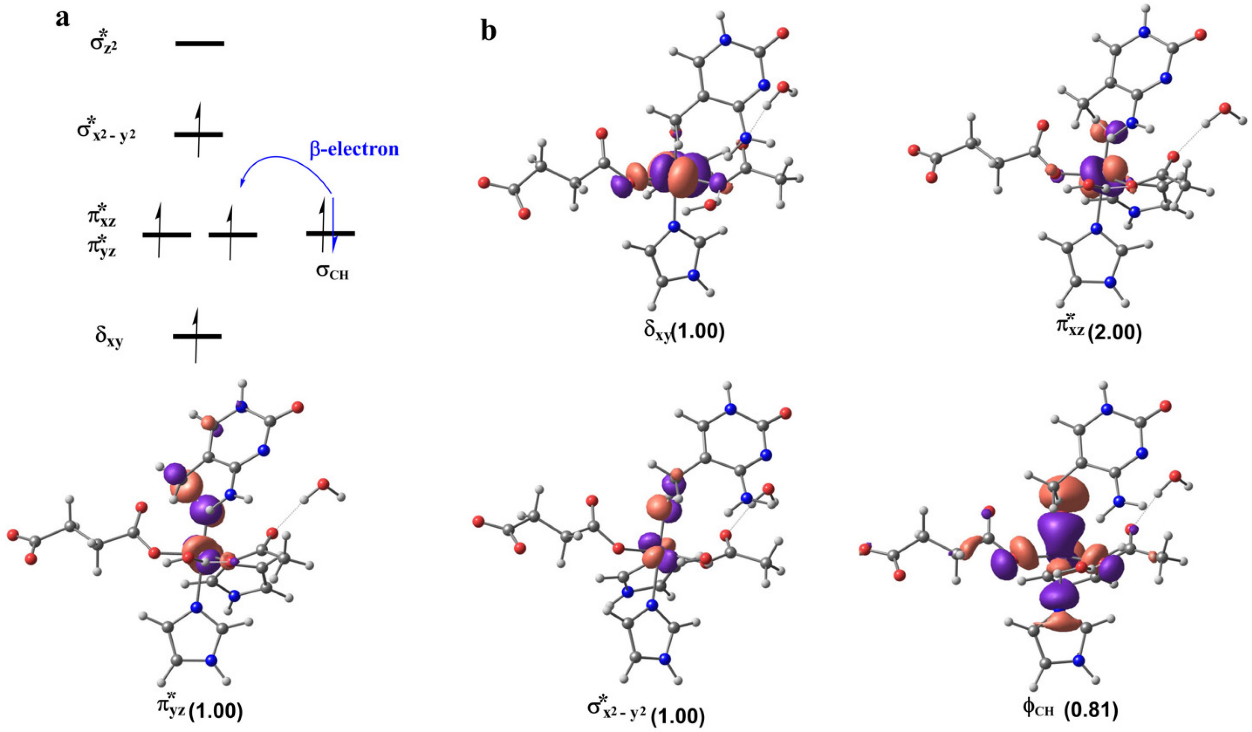


Figure 9. Electron shift diagram (a) and the spin natural orbitals (SNOs) (b) with their respective populations (in parentheses) during HAT in the K1299E-S1303N TET2·Fe(IV)=O/dsDNA mutant, calculated from the transition state structure.

The residues involved in hydrophobic interactions in all transition states of mutant forms have positive correlated motions with iron, the loop containing iron ligating HxD motif, and the residues that stabilize the succinate. The result implies that the residues enhance the stabilization of the metal center mainly via long-range interactions in contrast to the WT where they also involve a correlated motion with the DNA substrate. This difference contributes to the weaker binding of the DNA in all the mutants which further explains the observed differences in the free energies of binding. Tyr1902, a residue involved in the cytosine ring stabilization in K1299E-S1303N and N1387A mutant forms shows correlated motion with the DNA-interacting loop2, the Zn3 finger region (in N1387A), and the iron-ligand HxD motif loop (in K1299E-S1303N). This correlated motion involving Tyr1902 probably enhances the interaction of DNA to the protein in N1387A.

EDA of the mutant forms shows that residues Arg1261, His1904, and Val1900 contribute with higher energies than in the WT (Table 1), which leads to the increased reaction barrier observed in the mutants. For example, in K1299E-S1303N, residues Thr1393, Val1900, Tyr1902, Asn1387, His1904, and Arg1261 have energies 0.65, 1.11, 0.73, 0.50, 1.11, and 2.67 kcal/mol, respectively. Also, in N1387A and Y1902, the substitutions led to destabilization of the systems by 1.32 and 1.85 kcal/mol, respectively. This contributes to the higher barriers observed in the mutants with respect to the WT (where N1387 and Y1902 contribute energies -0.05 and 0.04 kcal/mol, respectively). A change in sign of the individual residues interaction energy (from positive to negative) from the reactant complex to the transition state results in the lowering of the reaction barrier and vice versa.[21a] Our results indicate that residues involved in transition state stabilization, contribute with more energy to the stabilization of the TS_{HAT} in the WT than in the mutants (Table 1).

3.5.2 How the Mutations Alter the HAT Electric Field and KIE?

The mutation of residues in the second sphere with respect to the active site and beyond influenced the electric field along the reaction coordinate (Table S5), with similar trend observed at both the reactants and transition states complexes. The differences in electric field likely result in a higher activation energy barrier in K1299E-S1303N and Y1902A mutant forms than WT snapshots with similar O-H distances and Fe-O-H angles. The complete analysis is presented in the SI. There are only small differences in the KIE values obtained for the mutant forms when compared to the WT, indicating that KIE would not reflect sensitively the effect of the various substitutions in the system at 303 K. The temperature dependence plot of KIE, however, show that the double mutant forms are more strongly dependent on temperature than the single ones (Figure S16).

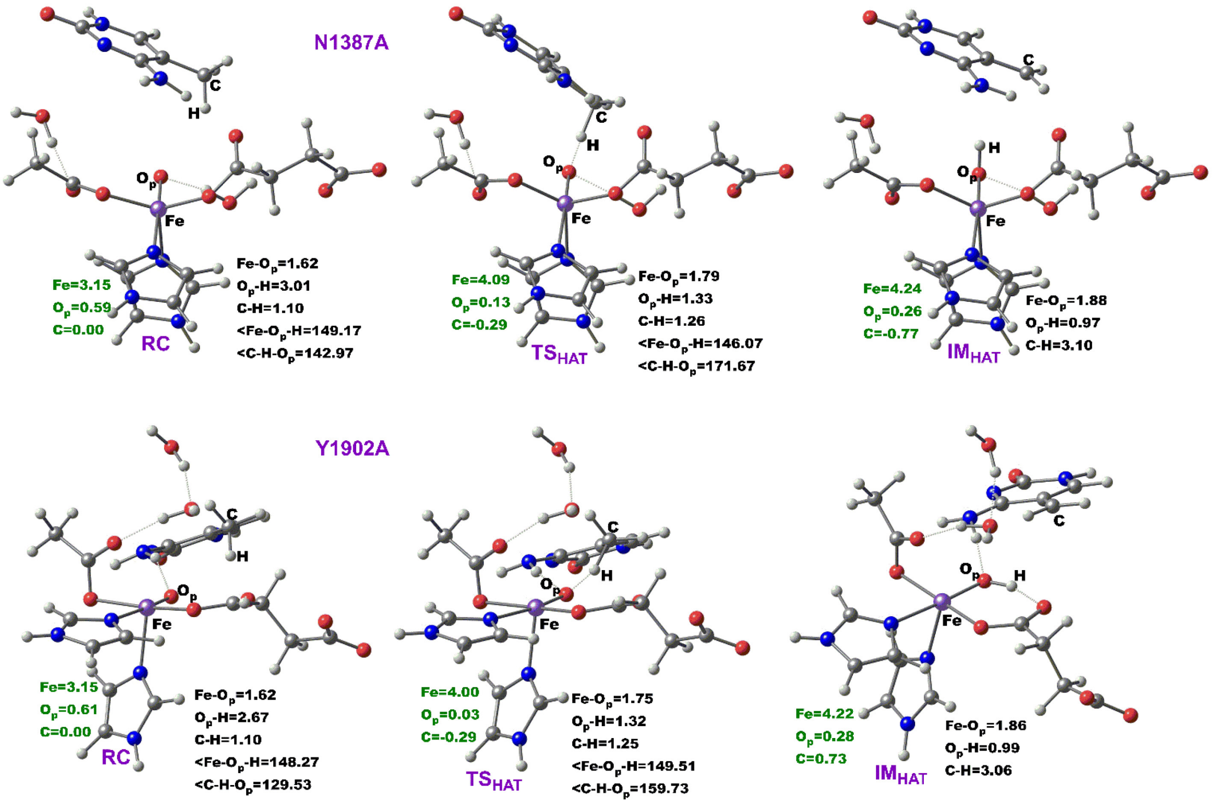


Figure 10. Stationary points geometries along the HAT step in the N1387A and Y1902A mutants. Distances (Å) and spin densities are in black and green, respectively.

Table 1. Energy Decomposition Analysis (EDA) (in kcal/mol) of the residues that stabilize the TS_{HAT} in the wild-type and the mutants

Residues	WT	N1387A	Y1902A	K1299E-S1303N	S1290A-Y1295A
Thr1393	-1.36	0.52	0.08	0.65	-0.14
Val1395	-0.82	-0.28	-0.01	-0.02	0.06
Val1900	0.08	0.51	0.27	1.11	0.69
Leu1872	-0.60	0.05	0.02	0.01	-0.02
Arg1261	-4.46	3.37	-1.63	2.67	2.53
Tyr1902	0.04	0.81	1.85	0.73	0.68
His1904	0.01	0.41	0.47	1.11	0.90
Asn1387	-0.05	1.32	-0.01	0.50	0.29

4.0 Conclusions

In this study, we have investigated how conformational dynamics of the WT TET2·Fe(IV)=O/dsDNA complex and single and double mutations (generated in vitro and with clinical implications in human patients) influence key structural determinants and the catalytic mechanism of the oxidation of 5mC dsDNA substrate to 5hmC using MD and hybrid QM/MM computational approaches.

The studies reveal that the substitutions affect the substrate's binding and the DNA interactions to the TET2 enzyme, thereby affecting the overall structural stability of the protein-DNA complexes. We identified hydrogen bonding interactions that are not present in the binding sites of the substrate and the succinate binding in the WT. The GS linker and the Cys-rich N-terminal (Cys-N) subdomain correlated motions contribute stronger to the DNA interactions to TET2 in the WT than in the mutant forms, thus enhancing the overall structural formation in the WT. The calculated energy barriers show that the mutations decrease the rate of the 5mC oxidation to 5hmC in comparison to the WT. The calculated rate-determining HAT barrier for the WT is in good agreement with other reported Fe(II) and 2OG dependent enzymes. Remarkedly, the results enabled us to identify key residues that are crucial for the HAT during the oxidation and to delineate their energy contributions and crucial long-range correlated interactions. The calculations reveal that the K1299E-S1303N variant uses π-pathway for the HAT. In contrast, the WT and other mutants use σ-pathway for the HAT, implying that residue substitutions can influence the molecular orbital interactions. The electric field calculations reveal the importance of conformational changes and mutations to alter the electric field along the direction of Fe=O bond, thus contributing to efficient catalysis. For example, the K1299E-S1303N and Y1902A mutants with similar O-H distances and Fe-O-H angles as WT show a higher HAT barrier, as the intensity of the local electric field along the direction of the Fe=O bond are reduced in mutant forms. The synergistic effects between the substrate orientation, internal electric field and long-range interactions contribute to the overall stability of the transition states in comparison to the reactant complexes. Kinetic isotope effect (KIE) calculations for the rate-determining step are in good agreement with the energy barrier, suggesting low tunneling contributions. The KIEs for the single mutant forms are weakly dependent on temperature, while in the double mutant forms, the KIEs strongly depend on temperature, especially at lower temperatures. Our calculations provide a molecular basis to understand the molecular pathology involved in the human TET2 enzyme and emphasize the crucial effects of protein sites in the second sphere and beyond on the catalytic events.

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

This research was supported by Michigan Technological University Graduate Teaching Assistantship to S.O.W. and start-up grants to C.Z.C., and T.G.K. C. C.Z.C. also acknowledges NSF grant 1904215 (testing the methodology).

Conflicts of interest

The authors declare no competing financial interest.

References

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

(2) 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, 93–204.

(3) Branco, M. R.; Ficz, G.; Reik, W. Uncovering the Role of 5-Hydroxymethylcytosine in the Epigenome. *Nat. Rev. Genet.* **2012**, *13* (1), 7–13.

(4) Williams, K.; Christensen, J.; Helin, K. DNA Methylation: TET Proteins-Guardians of CpG Islands?

EMBO Reports. **2012,** 13, 28–35.

(5) Pastor, W. A.; Aravind, L.; Rao, A. TETonic Shift: Biological Roles of TET Proteins in DNA Demethylation and Transcription. *Nature Reviews Molecular Cell Biology*. **2013,** 14, 341–356.

(6) Wu, H.; Zhang, Y. Mechanisms and Functions of Tet Proteinmediated 5-Methylcytosine Oxidation. *Genes and Development*. **2011,** 25, 2436–2452.

(7) Tan, L.; Shi, Y. G. Tet Family Proteins and 5-Hydroxymethylcytosine in Development and Disease. *Development.* **2012**, *139* (11), 1895–1902.

(8) 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.

(9) (a) Zheng, G.; Fu, Y.; He, C. Nucleic Acid Oxidation in DNA Damage Repair and Epigenetics. *Chem. Rev.* **2014**, *114* (8), 4602–4620. (b) Ghanty, U.; Wang, T.; Kohli, R. M. Nucleobase Modifiers Identity TET enzymes as Biofunctional DNA Dioxygenases Capable of Direct N-Demethylation. *Angew. Chem. Int. Ed.* **2020,** 59, 11312-11315.

(10) (a) Rasmussen, K. D.; Helin, K. Role of TET Enzymes in DNA Methylation, Development, and Cancer. *Genes and Development*. **2016,** 30. 733-750. (b) Crawford, D. J.; Liu, M. Y.; Nabel, C. S.; Cao, X.; Garcia, B. A.; Kohli, R. M. TET2 Catalyzes Stepwise 5-methylcytosine Oxidation by an Iterative and *de novo* Mechanism. *J. Am. Chem. Soc.* **2016,** 138, 730-733.

(11) An, J.; Rao, A.; Ko, M. TET Family Dioxygenases and DNA Demethylation in Stem Cells and Cancers. *Experimental and Molecular Medicine*. **2017**, 49, 1-12.

(12) 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.

(13) 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.

(14) Xue, J. H.; Chen, G. D.; Hao, F.; Chen, H.; Fang, Z.; Chen, F. F.; Pang, B.; Yang, Q. L.; Wei, X.; Fan, Q. Q.; Xin, C.; Zhao, J.; Deng, X.; Wang, B. A.; Zhang, X. J.; Chu, Y.; Tang, H.; Yin, H.; Ma, W.; Chen, L.; Ding, J.; Weinhold, E.; Kohli, R. M.; Liu, W.; Zhu, Z. J.; Huang, K.; Tang, H.; Xu, G.

L. A Vitamin-C-Derived DNA Modification Catalysed by an Algal TET Homologue. *Nature* **2019**, *569* (7757), 581–585.

(15) Hu, L.; Lu, J.; Cheng, J.; Rao, Q.; Li, Z.; Hou, H.; Lou, Z.; Zhang, L.; Li, W.; Gong, W.; Liu, M.; Sun, C.; Yin, X.; Li, J.; Tan, X.; Wang, P.; Wang, Y.; Fang, D.; Cui, Q.; Yang, P.; He, C.; Jiang, H.; Luo, C.; Xu, Y. Structural Insight into Substrate Preference for TET-Mediated Oxidation. *Nature.* **2015**, *527* (7576), 118–122.

(16) Lee, D. H.; Jin, S. G.; Cai, S.; Chen, Y.; Pfeifer, G. P.; O'Connor, T. R. Repair of Methylation Damage in DNA and RNA by Mammalian AlkB Homologues. *J. Biol. Chem.* **2005**, *280* (47), 39448–39459.

(17) Yi, C.; Chen, B.; Qi, B.; Zhang, W.; Jia, G.; Zhang, L.; Li, C. J.; Dinner, A. R.; Yang, C. G.; He, C. Duplex interrogation by a direct DNA repair protein in search of base damage. *Nat. Struct. Mol. Biol.* **2012**, *19* (7), 671–676.

(18) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science.* **2011**, *333* (6047), 1300–1303.

(19) Bachman, M.; Uribe-Lewis, S.; Yang, X.; Williams, M.; Murrell, A.; Balasubramanian, S. 5-Hydroxymethylcytosine Is a Predominantly Stable DNA Modification. *Nat. Chem.* **2014**, *6* (12), 1049–1055.

(20) 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.

(21) (a) 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. (b) Leddin, E. M.; Cisneros, G. A. Comparison of DNA and RNA substrate on TET2 structure. *Adv. Protein Chem. Strut. Biol*. **2019**, 117, 91-112. (c) Liu, H.; Zhang, Y.; Yang, W. How is the Active Site of Enolase Organized To Catalyze Two Different Reaction Steps?. *J. Am. Chem. Soc.* **2000,** 122(28), 6560-6570. (d) 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. (e) Cui, Q.; Karplus, M. Catalysis and Specificity in Enzymes: A Study of Triosephosphate Isomerase and Comparison with Methyl Glyoxal Synthase. *Adv. Protein Chem.,* **2003,** 66, 315-372. (f) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán. J. Preorganization and reorganization as related factors in enzyme catalysis: the chorismite mutase case. *Chem. -Eur. J.,* **2003,** 9, 984-991.

(22) 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.

(23) 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.

(24) Liu, H.; Llano, J.; Gauld, J. W. A DFT Study of Nucleobase Dealkylation by the DNA Repair Enzyme AlkB. *J. Phys. Chem. B.* **2009**, *113* (14), 4887–4898.

(25) 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. - A Eur. J.* **2014**, *20* (2), 435–446.

(26) 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.

(27) 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.

(28) Lu, J.; Hu, L.; Cheng, J.; Fang, D.; Wang, C.; Yu, K.; Jiang, H.; Cui, Q.; Xu, Y.; Luo, C. A Computational Investigation on the Substrate Preference of Ten-Eleven-Translocation 2 (TET2). *Phys. Chem. Chem. Phys.* **2016**, *18* (6), 4728–4738.

(29) Ko, M.; An, J.; Pastor, W. A.; Koralov, S. B.; Rajewsky, K.; Rao, A. TET Proteins and 5-Methylcytosine Oxidation in Hematological Cancers. *Immunol. Rev.* **2015**, *263* (1), 6–21.

(30) Huang, Y.; Rao, A. Connections between TET Proteins and Aberrant DNA Modification in Cancer. *Trends Genet.* **2014**, *30* (10), 464–474.

(31) Ko, M.; Huang, Y.; Jankowska, A. M.; Pape, U. J.; Tahiliani, M.; Bandukwala, H. S.; An, J.; Lamperti, E. D.; Koh, K. P.; Ganetzky, R.; Liu, X. S.; Aravind, L.; Agarwal, S.; MacIejewski, J. P.; Rao, A. Impaired Hydroxylation of 5-Methylcytosine in Myeloid Cancers with Mutant TET2. *Nature* **2010**, *468* (7325), 839–843.

(32) Langemeijer, S. M. C.; Kuiper, R. P.; Berends, M.; Knops, R.; Aslanyan, M. G.; Massop, M.; Stevens-Linders, E.; Van Hoogen, P.; Van Kessel, A. G.; Raymakers, R. A. P.; Kamping, E. J.; Verhoef, G. E.; Verburgh, E.; Hagemeijer, A.; Vandenberghe, P.; De Witte, T.; Van Der Reijden, B. A.; Jansen, J. H. Acquired Mutations in TET2 Are Common in Myelodysplastic Syndromes. *Nat. Genet.* **2009**, *41* (7), 838–842.

(33) 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.

(34) Fiser, A.; Šali, A. Modeller: Generation and Refinement of Homology-Based Protein Structure Models. *Methods Enzymol.* **2003**, *374*, 461–491.

(35) Olsson, M. H.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory. Comput.* **2011,** 7(2), 525-537.

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

(37) Seminario, J. M. Calculation of Intramolecular Force Fields from Second-Derivative Tensors. *Int. J. Quantum Chem.* **1996,** 60(7), 1271-1277.

(38) 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.

(39) 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.

(40) 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.

(41) 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. - A Eur. J. 2019, 25 (21), 5422–5426.

(42) 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 Research 2019, 47(11), 5522-5529.

(43) Pabis, A.; Geronimo, I.; York, D. M.; Paneth, P. Molecular Dynamics Simulation of Nitrobenzene Dioxygenase Using AMBER Force Field. J. Chem. Theory Comput. 2014, 10 (6), 2246–2254.

(44) 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.

(45) Bresme, F. Equilibrium and Nonequilibrium Molecular-Dynamics Simulations of the Central Force Model of Water. *J. Chem. Phys.* **2001**, *115* (16), 7564–7574.

(46) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. 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.

(47) 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.

(48) 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**.

(49) 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.

(50) Deserno, M.; Holm, C. How to Mesh up Ewald Sums. I. a Theoretical and Numerical Comparison of Various Particle Mesh Routines. *Journal of Chemical Physics*. **1998,** 109(18), 7678–7693.

(51) 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.

(52) 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.

(53) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.Py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8* (9), 3314–3321.

(54) 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*, 1–28.

(55) 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.

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

(57) 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.

(58) Billeter, S. R.; Turner, A. J.; Thiel, W. Linear Scaling Geometry Optimisation and Transition State Search in Hybrid Delocalised Internal Coordinates. *Phys. Chem. Chem. Phys.* **2000**, *2* (10), 2177–2186.

(59) Stuyver, T.; Huang, J.; Mallick, D.; Danovich, D.; Shaik, S. TITAN: A Code for Modeling and Generating Electric Fields-Features and Applications to Enzymatics Reactivity. *J. Comput. Chem.* **2020**, 41, 74-82.

(60) Laconsay, C. J.; Tsui, K. Y.; Tantillo, D. J. Tipping the balance: theoretical interrogation of divergent extended heterolytic environment. *Chem. Sci.* **2020,** 11, 2231-2242.

(61) Dubey, K. D.; Stuyver, T.; Kalita, S.; Shaik, S. Solvent Organization and Rate Regulation of a Menshutkin Reaction by Oriented External Electric Fields are Revealed by Combined MD and QM/MM Calculations. *J. Am. Chem. Soc.* **2020,** 142(22), 9955-9965.

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

(63) (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, 492-503. (b) Arnold, A. E.; Ornstein, R. L. Molecular Dynamics Study of Time-Correlated Protein Domain Motions and Molecular Flexibility: Cytochrome P450BM-3. *Biophys J.* **1997,** 73, 1147-1159. (c) Tang, Q.; Kaneko, K. Long-range correlation in protein dynamics: Confirmation by structural data and normal mode analysis. *PLoS Comput. Biol.* **2020,** 16(2), e1007670.

(64) Chaturvedi, S. S.; Ramanan, R.; Waheed, S. O.; Karabencheva-Christova, T. G.; Christov, C. Z. Structure-function relationships in KDM7 histone demethylases. *Adv. Protein Chem. Strut. Biol.* **2019**, 117, 113-125.

(65) 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.

(66) (a) Ghafoor, S.; Mansha, A.; de Visser, S. P. Selective Hydrogen Atom Abstraction from Dihydroflavonol by a Non-heme Iron Center is the Key Step in the Enzymatic Flavonol Synthesis and Avoids Byproducts. *J. Am. Chem. Soc.* **2019,** 141, 20278-20292. (b) Yan, L.; Liu, Y. Insights into the Mechanism and Enantioselectivity in the Biosynthesis of Ergot Alkaloid Cycloclavine Catalyzed by Aj_EasH from *Aspergillus japonicus*. *Inorg. Chem.* **2019,** 58, 13771-13781. (c) Bai, J.; Yan, L.; Liu, Y. Catalytic mechanism of the PrhA (V150L/A232S) double mutant involved in the fungal meroterpenoid biosynthetic pathway: a QM/MM Study. *Phys. Chem. Chem. Phys.* **2019,** 21, 25658-25668. (d) Su, H.; Sheng, X.; Zhu, W.; Ma, G.; Liu, Y. Mechanistic Insights into the Decoupling Desaturation and Epoxidation Catalyzed by Dioxygenase AsqJ involved in the Biosynthesis of Quinolone Alkaloids. *ACS Catal.* **2017,** 7, 5534-5543. (e) Ma, G.; Zhu, W.; Su, H.; Cheng, N.; Liu, Y. Uncoupling Epimerization and Desaturation by Carbapenem Synthase: Mechanistic Insights from QM/MM Studies. *ACS Catal.* **2015,** 5, 5556-5566.

(67) 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.

(68) Krebs, C.; Fujimori, D. G.; Walsh, C. T.; Bollinger, J. M. Non-Heme Fe(IV)-Oxo Intermediates. *Acc. Chem. Res*. **2007,** 40, 484–492.

(69) 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.

(70) 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.

(71) Logunov, I.; Schulten, K. Quantum chemistry: Molecular dynamics study of the dark-adaptation process in bacteriorhodopsin. *J. Am. Chem. Soc*. **1996**, 118(40), 9727-9735.

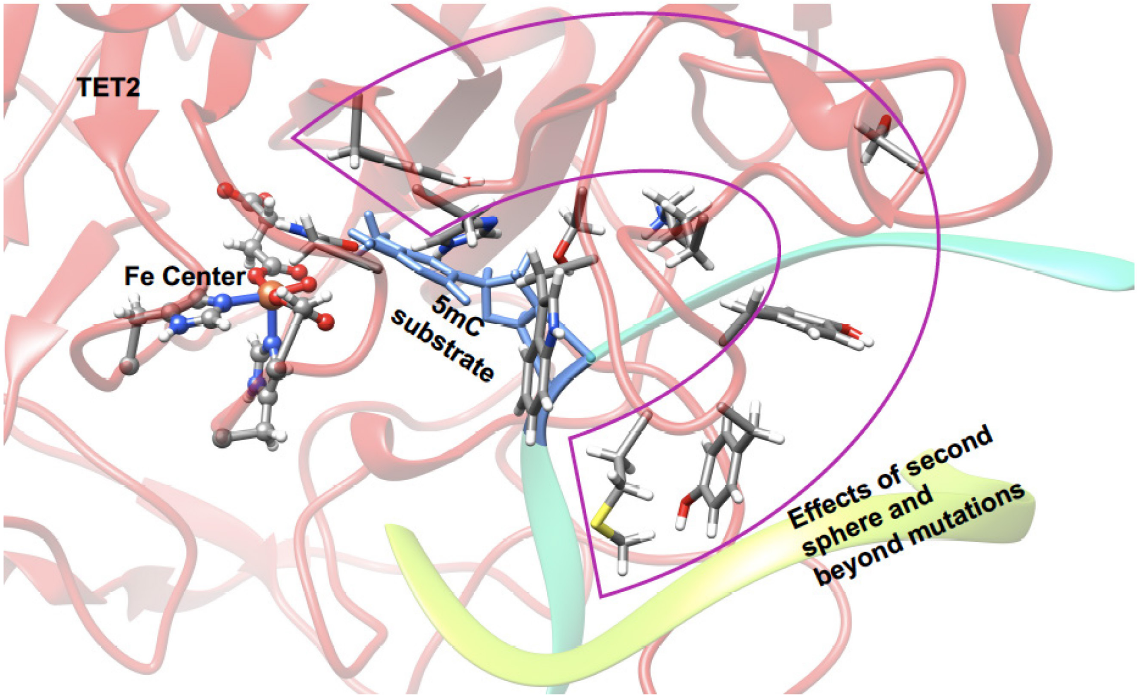
(72) Álvarez-Barcia, S.; Kästner, J. Atom Tunneling in the Hydroxylation Process of Taurine/α-Ketoglutarate Dioxygenase Identified by Quantum Mechanics/Molecular Mechanics Simulations. *J. Phys. Chem. B* **2017**, *121* (21), 5347–5354.

(73) Meir, R.; Chen, H.; Lai, W.; Shaik, S. Oriented Electric Fields Accelerate Diels-Alder Reactions and Control the endo/exo Selectivity. *ChemPhysChem,* **2010**, 11, 301-310.

(74) Jonasson, N. S. W.; Daumann, L. J. 5-Methylcytosine Is Oxidized to the Natural Metabolites of TET Enzymes by a Biomimetic Iron(IV)-Oxo Complex. *Chem. - A Eur. J.* **2019**, *25* (52), 12091–12097.

(75) Koski, M. K.; Hieta, R.; Bollner, C.; Kivirikko, K. I.; Myllyharju, J.; Wierenga, R. K. The Active Site of an Algal Prolyl 4-Hydroxylase Has a Large Structural Plasticity. *J. Biol. Chem.* **2007,** 282, 37112-37123.

TOC only


TET2
Fe Center
5mC substrate
Effects of second sphere and beyond mutations