PCCP

Computational Investigations of Selected Enzymes From Two Iron and α–ketoglutarate–Dependent Families

Journal:	*Physical Chemistry Chemical Physics*
Manuscript ID	CP-PER-08-2021-003800.R1
Article Type:	Perspective
Date Submitted by the Author:	21-Sep-2021
Complete List of Authors:	Berger, Madison; University of North Texas, Chemistry Walker, Alice; Wayne State University, Chemistry Montelongo, Erik; University of North Texas, Chemistry Cisneros, Andres; University of North Texas, Chemistry

PCCP

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Computational Investigations of Selected Enzymes From Two Iron and α–ketoglutarate–Dependent Families

Madison B. Berger[a], Alice R. Walker[b], Erik Antonio Vázquez-Montelongo[a] and G. Andrés Cisneros*[a]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

DNA alkylation is used as the key epigenetic mark in eukaryotes, however, most alkylation in DNA can result in deleterious effects. Therefore, this process needs to be tightly regulated. The enzymes of the AlkB and Ten-Eleven Translocation (TET) families are members of the Fe and alpha-ketoglutarate-dependent superfamily of enzymes that are tasked with dealkylating DNA and RNA in cells. Members of these families span all species and are an integral part of transcriptional regulation. While both families catalyze oxidative dealkylation of various bases, each has specific preference for alkylated base type as well as distinct catalytic mechanisms. This perspective aims to provide an overview of computational work carried out to investigate several members of these enzyme families including AlkB, ALKB Homolog 2, ALKB Homolog 3 and Ten-Eleven Translocate 2. Insights into structural details, mutagenesis studies, reaction path analysis, electronic structure features in the active site, and substrate preferences are presented and discussed.

1 Introduction

The Fe/α–ketoglutarate (Fe/α-kg) dependent superfamily of enzymes encompasses a large number of enzyme families that catalyze a broad range of reactions.[1–5] While there are many enzymes very similar to those that make up this family in sequence and/or mechanism, no others require the substrate α-kg. These enzymes are known to act on amino acid side chains, lipids, nucleotides and a variety of small molecules.[1] The reactions carried out include hydroxylation, demethylation, epoxidation, ring fragmentation, epimerization and desaturation.[4,6,7] Several Fe/α-kg enzymes and their reaction mechanisms have been investigated by various computational approaches.[8–18] Among this superfamily, members that are involved in DNA transactions have been a particular focus.

DNA is susceptible to various types of damage and modifications that frequently arise due to endogenous and/or exogenous sources.[19] Alkylation is a common form of DNA modification and can have both negative and positive effects. Methylation of cytosine at the 5 position to yield 5-methylcytosine is the staple epigenetic marker for gene regulation[20]. Conversely, other types of alkylation can lead to instabilities and strand breakage, which has been exploited as a common form of cancer treatment .[21] Two of the enzyme families tasked with handling this type of damage or modification are the AlkB and Ten-Eleven Translocation (TET) families.[22,23] Both of these families are members of the Fe/α-kg superfamily.[1,24] These enzymes have been extensively studied using computational, experimental, and crystallographic methods because of their biochemical importance.[1] AlkB of *E.coli* is responsible for the oxidative dealkylation of nucleobases. These enzymes can act on all seven of the N-methylated Watson-Crick base pairs such as N^1- methyl adenine (1mA) and N^3-methyl cytosine (3mC).[25] Various homologues of *E. coli* AlkB exist across both prokaryotes and eukaryotes.[22]

DNA methylation is a key epigenetic modification that is conducted by the DNA methyltransferase family.[28] A balance between methylation and demethylation must be met in order to preserve the integrity of the genome. Disruptions to the demethylation of cytosine is common in many types of cancer.[29] TET enzymes catalyze the sequential oxidation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC).[30] The TET family consists of three members all located on human chromosomes: TET1, TET2 and TET3.[31] All TET enzymes are a crucial component of tumor suppressor mechanisms in cancer and it is well known that all three are easily mutated.[32] Damage to TET1 can lead to acute myeloid and lymphoid leukemias and TET2 damage can lead to several different types of myeloproliferative disorders.[33]

The reaction mechanisms for both AlkB and TET were initially inferred from TauD, another member of the Fe/α-kg dependent

[a]*Department of Chemistry, University of North Texas, Denton, Texas, 76201, USA*
[b]*Department of Chemistry, Wayne State University, Detroit, Michigan, 48202, USA*
† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/cXCP00000x/

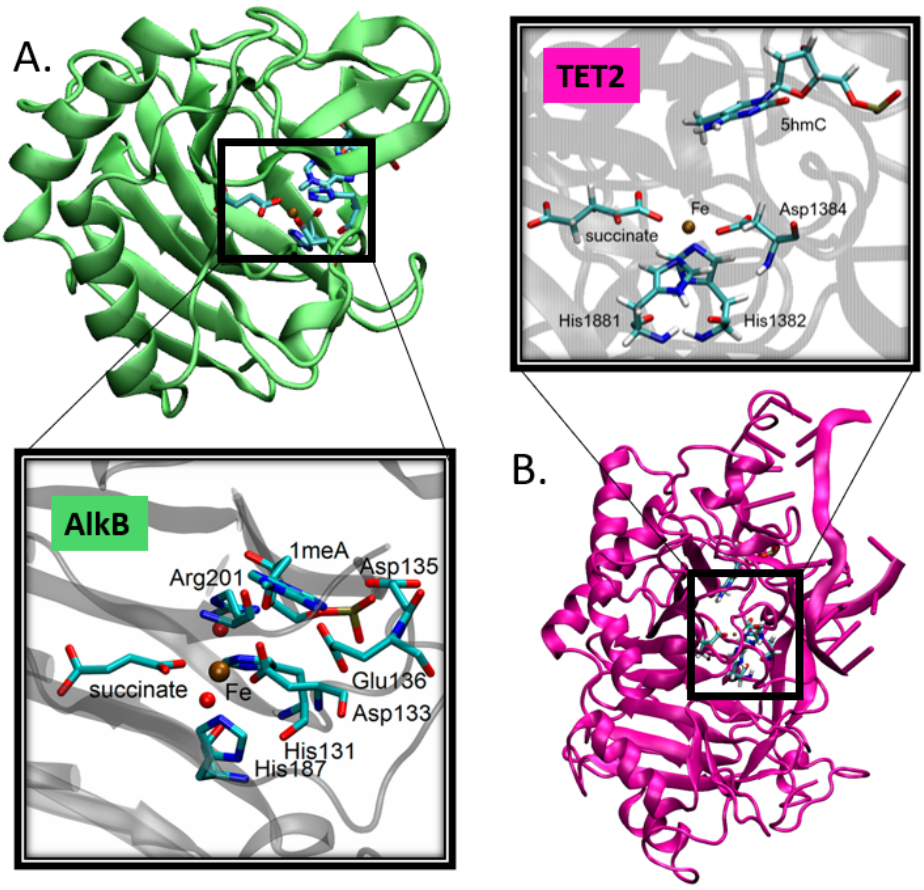


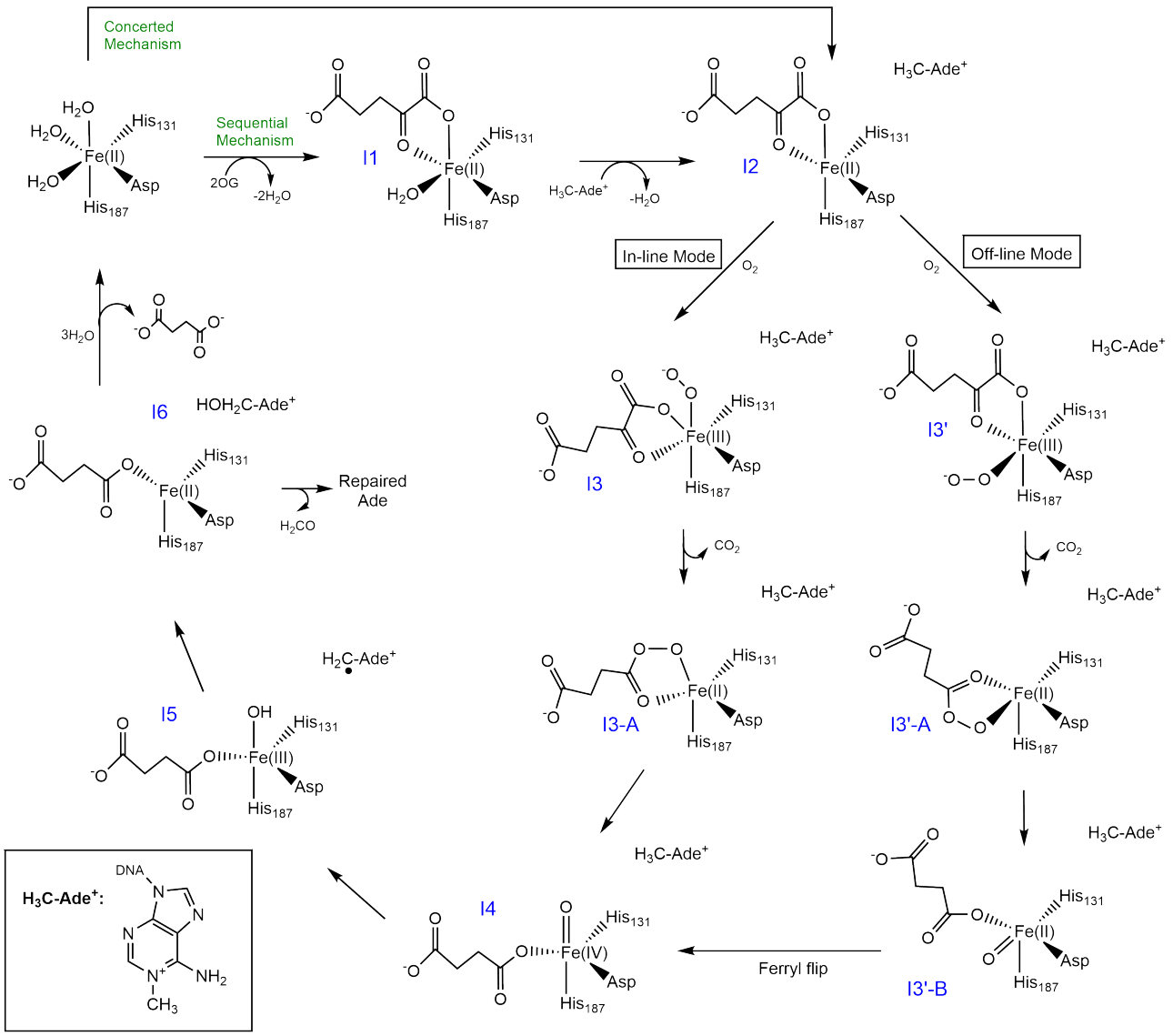
Fig. 1 A) The active site of AlkB containing succinate. Protein structure generated from pdb: 2FDG.[26] Active site image reproduced from: D. Fang, R.L. Lord, G.A. Cisneros, *J. Chem. Theory Comput.*, 2014, **10**, 5136–5148. (doi/10.1021/ct500572t) Copyright 2014 American Chemical Society. B) The active site of TET2 containing succinate. Protein structure generated from pdb: 4NM6 with 5hmC instead of 5mC.[27]

family.[1,8,34–39] There are two features that are structurally conserved in this family. A non–heme Fe, which alternates between oxidation states II, III and IV depending on the stage of the reaction. The non–heme Fe is coordinated by a triad of amino acids in the active site consisting of two histidine (His) residues and one aspartate (Asp) or glutamate (Glu) (Figure 1).[40] This feature is also referred to as the 2-His-1-carboxylate triad.[41] In addition, this triad is located inside of a double-stranded β-sheet fold, also known as a "jelly-roll" fold. The Fe maintains an octahedral geometry with three of those sites occupied by the 2-His-1-carboxylate triad and the other three by water or the required co–substrates: α–ketoglutarate and molecular oxygen.

Members from both of these enzyme families have been investigated by both classical and hybrid quantum mechanics/molecular mechanics (QM/MM) approaches to gain insights into their structure, function and catalytic mechanisms. Molecular dynamics (MD) simulations have been employed to study the structural features of AlkB and TET2, including the effects of mutations and different types of RNA/DNA.[42–47]

Quantum mechanics/molecular mechanics (QM/MM) is a useful method to study the reactivity of enzymes. Briefly, a small number of atoms in the system are treated with DFT, semi-empirical or other *ab initio* approaches and the rest of the protein is treated with (generally) a classical force field.[49] One of the many challenges of QM/MM can be separating a system into the two regions and determining which residues are important for the reaction of interest. Several methods currently suggested in the literature for building the QM region include protein sequencing and structural evolution analyses, electron localization function or charge shift analysis.[50,51] Once the regions have been determined, the boundary between the two must be specially handled, especially if there are bonds that are "cut" across the QM/MM boundary. There are a variety of methods used throughout the literature to address this issue including the pseudobond approach, link atoms and frozen localized orbitals.[49]

Another key component of QM/MM is the type of charge embedding that will be used between the QM charge density and the force field used in the MM. Three separate schemes have been designed to better represent the interactions between the two regions: mechanical, electrostatic, and polarizable embedding.[52] Mechanical embedding is the most basic method for representing interactions in that a force field is used for the QM and MM interaction. Electrostatic embedding is the most popular method and not as computationally demanding as polarizable embedding. Here, the charges from the MM are incorporated into the calculation of the QM Hamiltonian. Polarizable embedding takes elec-



Scheme 1 The proposed AlkB catalyzed dealkylation mechanism of 1mA adapted from *Waheed et al.*[48]

trostatic embedding a step further and allows for self consistent polarization between the quantum and classical subsystems.[53–55]

Once the QM/MM system has been established, it can be used to compute properties and investigate reaction pathways. There are several methods that can then be used to determine the minimum energy paths for chemical reactions, including a family of methods called chain-of-states. This family includes approaches such as the Nudged Elastic Band, replica path, and the quadratic string methods.[56–58] In these chain-of-states methods, appropriate sampling of the various states of the system must be met in order to connect the first and last state (reactant and product). QM/MM free energy calculations can also provide insight on reaction barriers, solubility, substrate binding affinity and equilibrium constants.[59,60] One approach to compute the free energy associated with the reactive process in an enzyme is called the minimum free energy path.[61,62]

In this perspective, we discuss how computational simulations have been used to elucidate the reaction mechanisms of selected AlkB and TET family enzymes. We show how combining QM/MM with other tools and techniques such as MD, non-covalent interactions, and other analyses continue to provide important insights into the reactivity and function of these important systems. The remainder of the paper is as follows: Section 2 presents the computational investigation of selected AlkB family enzymes, namely AlkB, ALKBH2 and ALKBH3. This section is separated following the different stages of the enzymatic reaction mechanism, and provides new QM/MM results on the rate limiting step catalyzed by ALKBH2 and ALKBH3. Subsequently, computational investigation of the mechanism of TET2 are presented in Section 3, again divided by the different oxidation stages catalyzed by this enzyme. Finally, concluding remarks and and a perspective are presented.

2 Mechanistic Studies of AlkB Family Enzymes

The AlkB family of enzymes have the ability to repair both monoalkyl substrates as well as certain etheno adducts of some DNA and RNA bases, with AlkB being the most versatile. The repair mechanism for both types of damage are very similar.[63–72] The reaction mechanism can be separated in two phases and is summarized in Scheme 1. The first phase involves the formation of a reactive ferryl, Fe(IV)=O, intermediate; this is then followed by the oxidation of the alkyl moiety on the substrate. This dealkylation reaction carried out by the AlkB family of enzymes has been extensively studied via MD, QM/MM and DFT methods.

Using the repair of 1mA as an example, the first step is the formation of the ferryl intermediate via binding of α-kg and O_2. This results in the formation of succinate and CO_2. From there, the oxyl atom on the Fe(IV)=O intermediate must rearrange from an axial to an equatorial position in order to align the oxygen trans to a neighboring His residue.[40,73] Subsequently, an H atom is abstracted by the oxyl from the substrate (1mA), which is also the rate limiting step of this reaction followed by an OH rebound to the radical methyl moiety. Finally, 1-hydroxymethyladenine disproportionates into formaldehyde and the dealkylated base.

2.1 Substrate Binding

Before the coordination of α-kg and O_2 to the Fe(II) atom in the active site, three coordination sites of the Fe are occupied by water (and the other three are bound to the three conserved active site residues).[2] In the first phase of the reaction, α-kg and the methylated substrate must bind in the active site. The binding of α-kg displaces two of the equatorial waters allowing for a bidentate configuration with the Fe (**I1**).

Both sequential and concerted mechanisms have been proposed for the binding of α-kg and 1mA.[74] Magnetic circular dichroism spectrosopy results show that the binding of α-kg alone does not allow for a rapid formation of Fe(IV)=O. Both substrates must be simultaneously bound in order for the dioxygen activation and substrate oxidation to continue which suggests that a concerted mechanism is preferred. Given that the concerted mechanism has been shown to be the most likely, the question can then be raised as to whether the O_2 can even enter the active site let alone bind to the Fe center without the other two substrates binding first. The authors also noted the penta-coordinate geometry around the Fe before O_2 binding. The binding of both substrates is what allows for the opening of a position on the Fe(II) that is meant for O_2. In a similar system to AlkB where pterin was used instead of α-kg, the calculated barrier for product release was 16 kcal/mol. If the pterin was not bound, this barrier increases by 5 kcal/mol, resulting in a much slower turnover rate by the enzyme.

Subsequently, O_2 must replace the third water on the Fe coordination sphere and complete the octahedral coordination to the Fe. The mechanism for the binding of the O_2 co–substrate to complete the enzyme–substrate (ES) complex was hypothesized to occur via an intra–molecular tunnel observed in the original crystal structures.[76] The proposed O_2 transport mechanism was recently investigated using polarizable and non–polarizable MD simulations paired with several other analyses (Figure 2).[75] Computational analysis suggested two possible tunnels by which the O_2 might transfer from the solvent to the active site.[75,77]

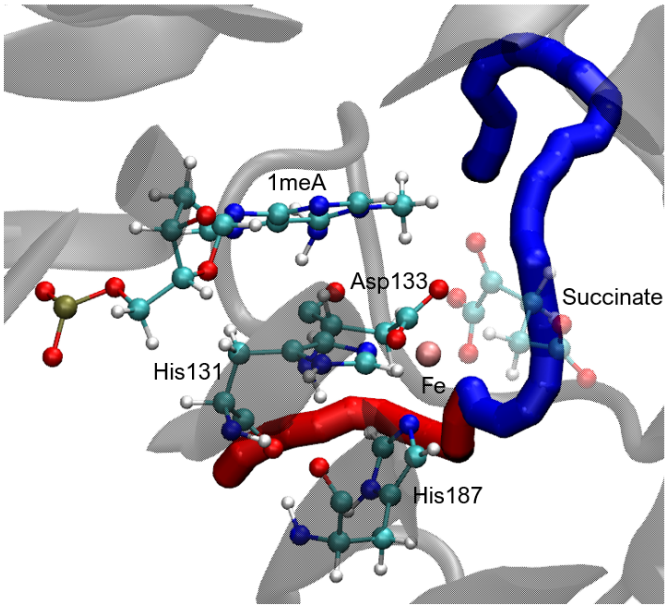


Fig. 2 The two most probable tunnels for O_2 to travel in order to reach the active site adapted from Torabifard *et al.*[75]

Based on the original tunnel observed from the crystal structure, a tyrosine at position 178 was hypothesized to act as a possible gate along the tunnel.[76] This tyrosine was computationally investigated via MD by replacing it with three different amino acids (W178Y, W178A and W178P). The computational results suggest that that W178Y and W178P behave similar to WT in O_2 occupancy, RMSF and distance correlation analyses. W178A differed the most in dynamic behavior and revealed a new path in which the O_2 molecule could be transported.

The PMF associated with the diffusion of O_2 from the surface of the protein to the active site was calculated using the ff99SB force field for both possible tunnels. The most probable tunnel (blue) shown in Figure 2 exhibited a downhill free energy barrier of 3 kcal/mol compared to 1.5 kcal/mol (red).[75] Since O_2 is neutral, but has a large polarizability, the barrier for the second tunnel was also calculated using the AMEOBA polarizable force field, resulting in a calculated free energy difference of 51.5 kcal/mol. This large discrepancy between the non–polarizable and polarizable PMFs is due exclusively to the polarization interactions. The large free energy change calculated with AMOEBA is also consistent with physiological expectations given the fact that AlkB and homologues are adaptive-response proteins localized mainly in the nuclei, where O_2 concentration needs to be tightly regulated to avoid DNA oxidative. Thus, high affinity for this substrate should be expected to enable catalytic turnover.

There are two possible orientations that have been proposed for the binding of O_2 to the Fe center, termed "in-line" and "off-line" modes (Scheme 1).[48,78] The "in-line" mode refers to the re-

arrangement of α-kg around the Fe that allows for the opening of a coordination site that the O_2 can occupy. From there, the O_2 binds trans to His187 (**I3**). Conversely, the "off-line" mode corresponds to the binding of O_2 trans to His131 (**I3'**). This path requires a ferryl flip (**I3'-B** to **I4**) in order to properly orient the oxyl moiety so that it is closer to the methylated substrated.

Several crystallographic studies have shown that the O_2 should bind in the same position that the final water occupies before leaving the active site (trans to His131).[76,79,80] An MM model was used to further analyze the binding mode of O_2.[78] Quesne *et al.* found that there was sufficient space in the binding pocket trans to His131 ("off-line" mode) to insert O_2. An additional MM model was constructed in which the O_2 was placed trans to His187 ("in-line" mode). This mode failed to optimized because the binding pocket was too dense at that position. Close contacts (<1.7Å) to the Asp133 and methyl group of the substrate were found in addition to clashes between the the O_2 and α-kg. The O_2 was found to be almost 6 Å from the methyl carbon of the substrate where the proton will be abstracted.[40] This would indicate that the proposed "in-line" mode of binding is the preferred pathway. The end point for the two tunnels proposed by Torabifard *et al.*[75] suggest preferential binding trans to His131 (**I3'**), although there appears to be sufficient space for O_2 to bind trans to His187 and thus both binding modes appear to be accessible at the end of the tunnels (Figure 2, **I3**).

2.2 Reorientation of the oxo moiety

Upon release of CO_2, the oxo moiety of Fe(IV)=O must undergo a reorientation from an axial to an equatorial position if it was added via the "in-line" mode. This will convert it from trans with respect to (w.r.t.) His131 to trans w.r.t. His187. Quesne *et al.* used QM/MM to study the reorientation of the oxo moiety in order to determine why it may be catalytically relevant for this enzyme.[78] They found that the rotation of the oxo group changes the shape of the molecular orbitals in the active site as well as the occupancies of the HOMO and LUMO orbitals.[78] In the axial position, $\sigma^*_{x^2-y^2}$ is singly occupied and $\sigma^*_{z^2}$ is virtual. The ordering of the orbitals is completely reversed when the Fe(IV)=O reorients and the oxo is in the equatorial position. Additionally, an isomerization energy difference of -6.0 kcal/mol in favor of the equatorial position was reported. It was also noted that there were very few changes in spin densities and charges of these two complexes.[78]

2.3 Hydrogen atom abstraction

Once the ferryl intermediate has been formed and the oxyl has rearranged (**I4**), the hydrogen atom transfer (HAT) step can take place from the methyl on the substrate. This is the rate limiting step for the entire reaction (**I4** to **I5**). For this step, an electron must be transferred into an orbital of the ferryl group. Two possible channels, termed σ and π have been postulated depending on the orbital that is occupied after the HAT.[81] In the σ-channel, a spin-up electron is transferred from the σ orbital of the methylated substrate to the σ^* orbital of the Fe(IV)=O. In the π-channel, a spin-down electron is transferred to the π^* orbital of the Fe(IV)=O. These two channels arise due to the possible angles that can be formed by the Fe-O-H atoms in the TS, which lead to differing orbital overlaps. The TS structure of the σ-channel has a linear geometry ($\sim$180°) while the Fe-O-H angle for the π-channel is significantly smaller ($\sim$120°).[48]

	ΔE_{reac} (kcal/mol)	$\Delta E^{\ddagger}$(kcal/mol)
AlkB[82]	-3.7	22.4 (19.8 exp.[83])
ALKBH2	-3.2	25.7 (21.3 exp.[84])
ALKBH3	-4.2	28.6

Table 1 Comparison of reaction energies and potential energy barriers for ALKBH2 and ALKBH3 compared to AlkB and to estimations from experimental kinetics data.

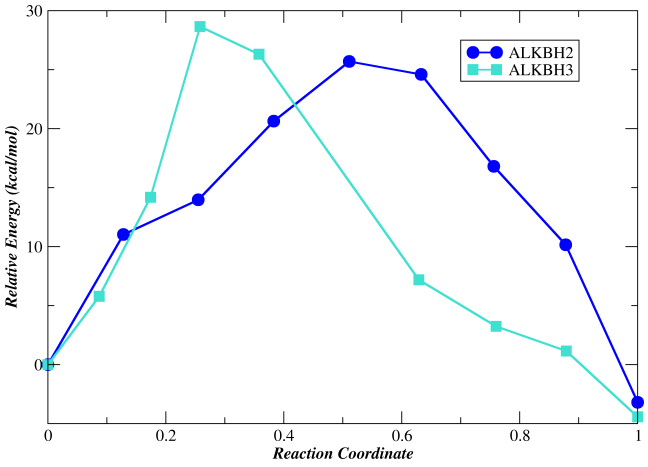


Fig. 3 ALKBH2 and ALKBH3 optimized paths connecting reactant and product structures using the quadratic string method.

Calculations for the HAT step have been performed for several spin states including triplet, quintet and septet, with the quintet state being the most stable consistent with previous work with TauD.[37,38] This intermediate can adopt two spin states: an intermediate spin state that is ferromagnetically coupled to the oxyl group (^{IS}Fe–O_F), and a high spin Fe that is antiferromagnetically coupled to the oxyl group (^{HS}Fe–O_{AF}).[82] The resulting **I4** state (Scheme 1) corresponds to the ^{IS}Fe–O_F state, with spin density populations of 3.26 au on the Fe and 0.54 au on the O. These spin densities are consistent with other reports on AlkB and other Fe=O systems (Table 2).[78,82,85,86]

Prior to the TS, an intersystem crossing (ISC) between the ^{IS}Fe–O_F and the ^{HS}Fe–O_{AF} states was observed.[82] After the minimum energy crossing point (MECP), the potential energy of the TS structure for the hydrogen abstraction step corresponds to a barrier of 22.4 kcal/mol, with a calculated Helmholtz free energy barrier of 18.9 kcal/mol.[82] Similar results for potential energy barriers were reported by Liu *et al.* (20.9 kcal/mol), Quesne *et al.* (23.4 kcal/mol), and Waheed *et al.*[40,48,78] Interestingly, several of the reported works show that the chosen level of theory has a significant impact on the calculated electronic structure description, with GGA functionals only reporting one surface (^{IS}Fe–O_F) whereas simulations that employ range–separated and dispersion corrected functionals report both spin surfaces.[40,48,78,82] These

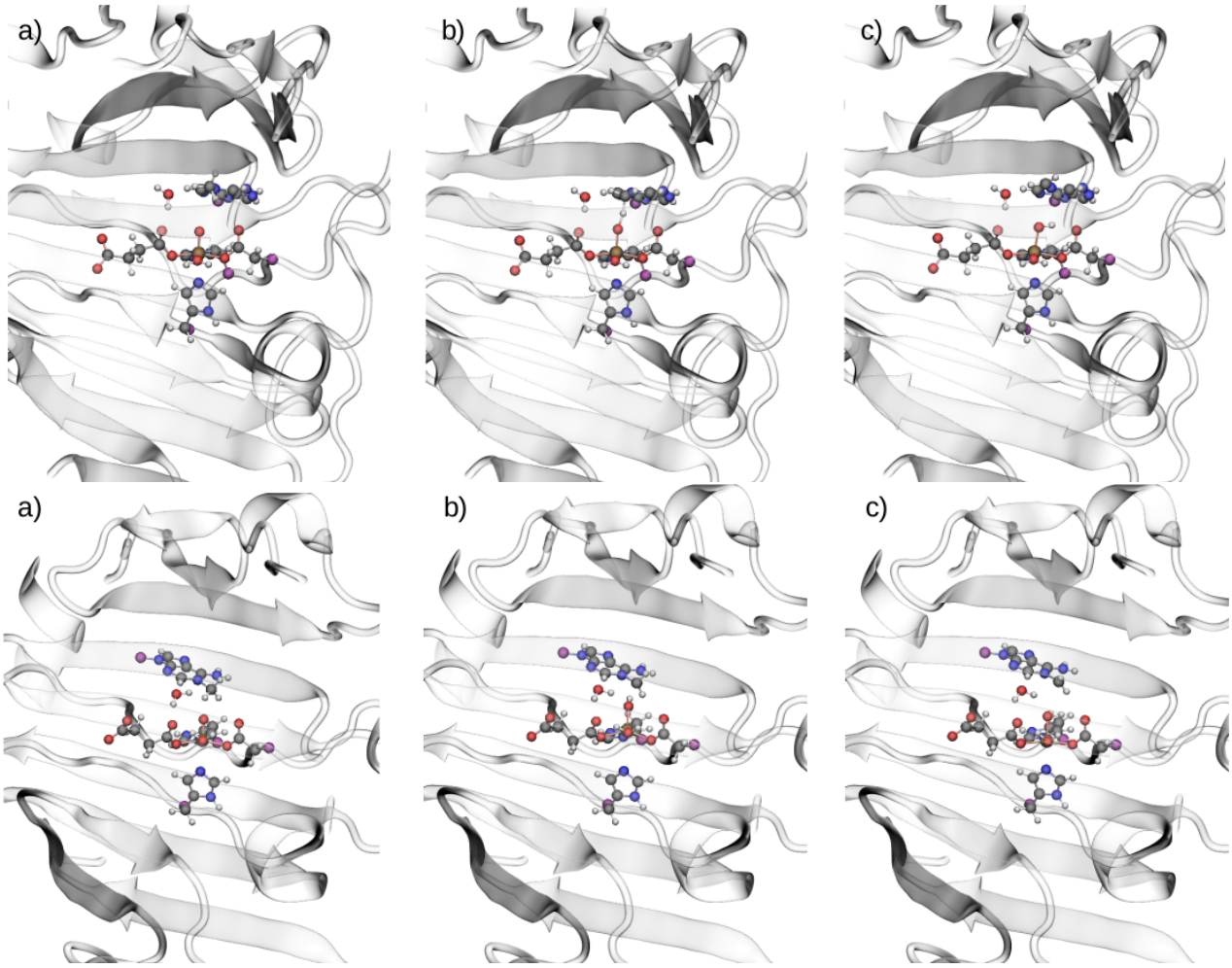


Fig. 4 Optimized reactant (a), TS (b) and product (c) structures for ALKBH2 (top) and ALKBH3 (bottom).

results are all in good agreement with the reported barrier of 19.8 kcal/mol estimated from experimental kinetic analysis.[83]

2.3.1 QM/MM Simulations of I4→I5 by ALKBH2 and ALKBH3

Humans have nine homologs of AlkB, ALKBH1-8 and FTO, that catalyze the repair of different DNA substrates. ALKBH2 and ALKBH3 can repair double-stranded DNA with a variety of lesions including 3-methylcytosine (3mC), 1-methyladenine (1mA), 3-methylthymidine (3mT), 1-methylguanine (1mG), and several etheno adducts.[22,63,67,87] ALKBH2 has a unique pincer structure to hold the double stranded DNA substrate in place, and has a very low efficiency for single stranded DNA substrates.[88] Contrary to ALKBH2, ALKBH3 has been shown experimentally to prefer ssDNA and act efficiently on several methylated RNA bases.[88,89]

We performed QM/MM simulations to investigate the reaction mechanism for the **I4** to **I5** step of the dealkylation of 1mA catalyzed by ALKBH2 and ALKBH3 following the same approach as that reported by Fang *et al.* for AlkB described in the previous subsection.[82] Briefly, crystal structures of the two enzymes with bound substrate (ALKBH2: 3BUC and ALKBH3: 2IUW) were used as starting points and examined with Molprobity followed by system preparation including solvation in TIP3P water and neutralization of the system with Na.[80,90–93] Subsequently, MD simulations were performed following the same procedure as described by Fang *et al.* using AMBER12's pmemd.cuda program and the ff99SB force field.[82,94,95] Each system was subjected to 50 ns of production in the NVT ensemble with the Berendsen thermostat and a 1 fs timestep.[96] Smooth particle mesh Ewald was used for the nonbonded interactions with an 8 Åcutoff from which selected snapshots were obtained for the subsequent QM/MM simulations (see SI for further computational methods).[97,98]

All QM/MM simulations for both systems involve QM regions containing 70 atoms including 4 pseudobond atoms.[99] The ωB97XD functional was used with a 6-31G(d,p) basis set for all non-pseudobond atoms.[82] An in-house software using TINKER for the MM part and a modified version of Gaussian16 for the QM part was used to perform additive QM/MM with electrostatic embedding for ALKBH2.[100,101] The ALKBH3 results were obtained with the same parameters and procedure using the LICHEM 1.1 software package.[102,103] Each converged structure was confirmed to have no negative frequencies except for the TS, which has one negative frequency along the reaction coordinate. The quadratic string method (QSM) was used to optimize the

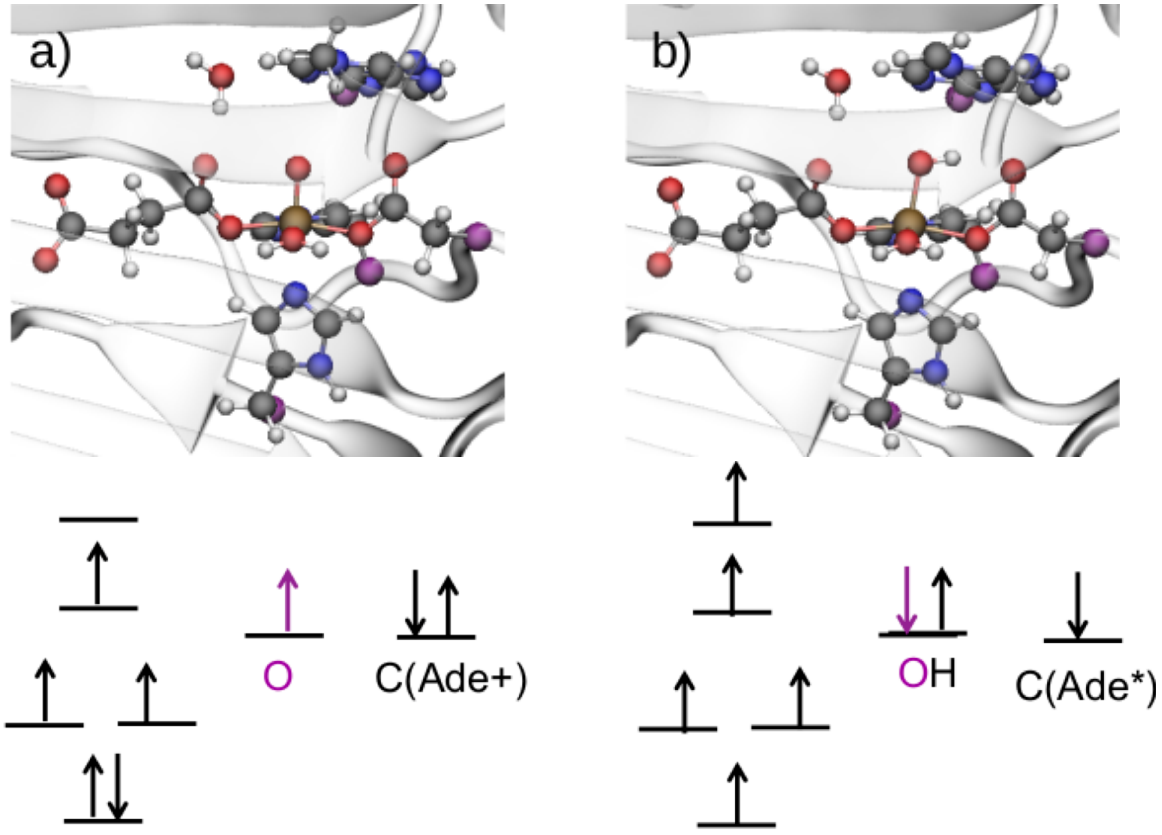


Fig. 5 Converged reactant (a) and product (b) structures with their respective spin diagrams.

Parameter	AlkB			ALKBH2			ALKBH3		
	R	TS	P	R	TS	P	R	TS	P
Fe Spin Density	3.26	4.24	4.35	3.20	4.16	4.36	3.31	4.33	4.34
O Spin Density	0.54	-0.23	0.26	0.62	-0.06	0.27	0.46	0.17	0.29
Fe-O Distance (Å)	1.61	1.77	1.84	1.60	1.74	1.84	1.59	1.86	1.83

Table 2 Comparison of electronic and structural parameters for the rate-limiting step in AlkB[82], ALKBH2 and ALKBH3.

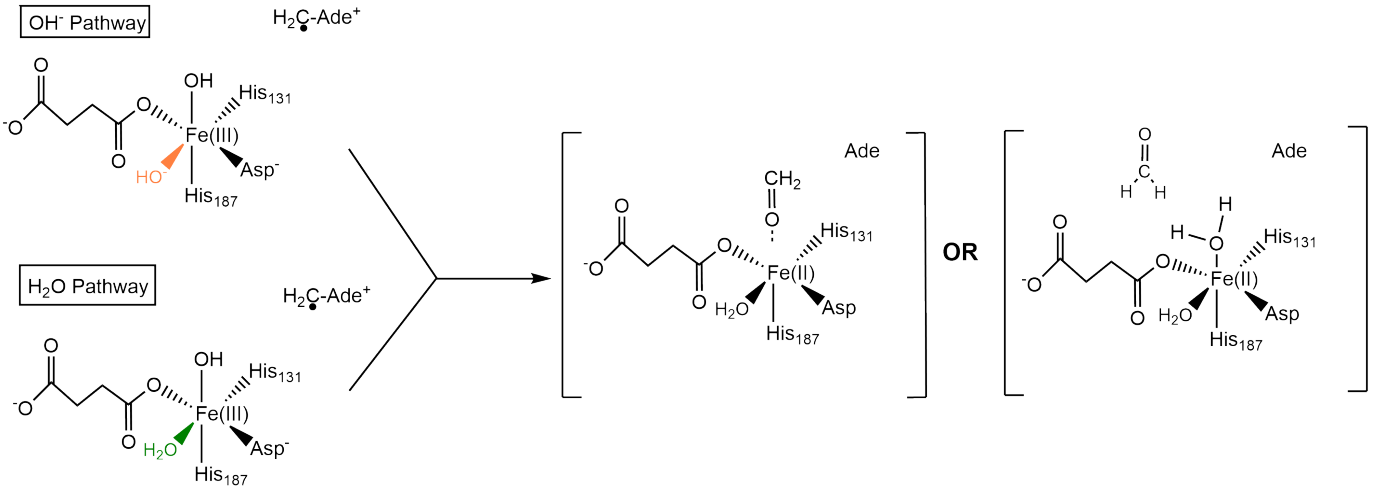
path between reactant, a guess TS and product for ALKBH3 (Figure 3).[103] The MM environment was replicated from the reactant structure for all images, and initially restrained. These restraints were gradually lowered over 11 steps.[103]

Figure 4 shows the optimized structures for the critical points (**I4**, TS and **I5**) for ALKBH2 and ALKBH3. Similar to what was reported by Fang *et al.*, the **I4** systems correspond to a ^{IS}Fe–O$_F$ state in both cases. Figure 5 presents a diagram for the spin states for the reactant (a) and product (b) on the quintet surface for ALKBH2. The minimized reactant has a calculated Mulliken spin density on the Fe of 3.20(3.31) for ALKBH2(ALKBH3), and the oxo has a spin density of 0.62(0.46) (2), as can also be seen in Table 2. Additionally, these spin densities are comparable to those obtained for an intermediate spin Fe-O inorganic complex.[104] The TS and product structures have Fe spin densities of 4.16(4.35) and 4.36(4.34) for ALKBH2(ALKBH3) and O spin densities of -0.06(0.17) and 0.27(0.29) au respectively. These values again compare quite well to previous work and literature values.[48,78,82] The observed changes in spin density (Table 2) are in agreement with the work from Fang *et. al.*,[82] suggesting the existance of an ISC between **I4** and **I5** for these systems.

For ALKBH2, the calculated ΔE barrier is 25.7 kcal/mol, which compares favorably to the experimental ΔG of $\sim$ 21.3 kcal/mol as estimated from experimental kinetics data.[84] In the case of AlkBH3, the calculated potential energy barrier is 28.6 kcal/mol. ALKBH2 has been shown to have lower processivity than AlkB for 1mA repair, and interestingly this point is captured by the calculated results (Table 1).

2.4 Hydroxyl radical rebound

The final phase of the reaction by AlkB involves the rebound of OH to the carbon radical of 1mA (**I5**). Finally, the C-N bond breaks and formaldehyde is formed (**I6**). Results from both crystal structure analysis and time-resolved Raman suggest that a zwitterion intermediate may be possible.[106,107] In addition, there is a large vacancy (shown in Figure 1A) along the plane that the succinate and Asp share. This could easily be occupied by a water molecule resulting in a H_2O pathway or a hydroxide radical if the pH is slightly basic, resulting in a OH^- pathway (Scheme 2).[105] The barrier for the abstraction of a hydrogen atom is very sim-



Scheme 2 Two possible pathways following the hydrogen atom abstraction of AlkB.[105]

ilar for both pathways, 23.2 kcal/mol (OH^- pathway) and 22.4 kcal/mol (H_2O pathway).[105] While the barriers may be similar, the respective mechanisms differ in significant ways. The rebound of the OH (**I6**) can occur via a concerted or stepwise path. The concerted path involves the coupled OH rebound and proton transfer, whereas in the stepwise mechanism the proton is first transferred followed by the oxygen. A concerted path was preferred in the OH^- path in which the OH rebounds and proton transfers simultaneously followed by the breaking of the C-N bond. The H_2O pathway favors a sequential mechanism in which the hydroxyl unbinds from the Fe and then loses its proton. The loss of the proton is paired with the C-N bond breaking. Several proton acceptors were proposed and analyzed including Asp133, Asp135, Glu136 or solvent. While the proton transfer to Glu136 had the lowest barrier, the authors could not rule out the other options. The entire study suggested that because the barrier for the C-N bond breakage was smaller in the OH^- and exterior proton acceptors are not needed, the OH^- is likely the preferred pathway.[105]

2.5 Effects of second-sphere residues and other molecules

The environment surrounding the active site has been shown to play an important role in the reactivity of many non-heme and Fe-dependent enzymes including AlkB.[108–110] As mentioned above, a second–shell water molecule that coordinates the ferryl moiety was observed to be important for the rate–limiting step in AlkB, ALKBH2 and ALKBH3. Several residues around the active site of AlkB are also known to play a crucial role in catalysis. In the case of AlkB, an energy decomposition analysis (EDA) was carried out by Fang *et al.*, which can provide insight into the Coulomb and vdW interactions calculated during MD.[82] This analysis suggested that nine residues exhibited energy differences that significantly impacted the reaction barrier: T51, R73, Y76, K127, Q132, Q155, R161, F185, and R210. Only three of the nine have been studied experimentally (Y76, T51 and R161). R161 does not appear to have a significant effect on the reaction barrier, but does aid in stabilization of 1mA. Both T51 and Y76 contribute to a decrease in the reaction rate. These two sites were subjected to mutation to an alanine residue,[111] which lead to an increase in the reaction barrier of 0.5 kcal/mol (T51) and 1.0 kcal/mol (Y76). A sequence alignment was also carried out that compared AlkB with ALKBH2 and ALKBH3, human homologues of AlkB. Two residues were structurally conserved: R210 and R161.

Waheed *et al.* used a hydrogen bond analysis to better understand the importance of the residues surrounding the active site or QM region. The TS of AlkB was found to be stabilized by a π-stacking interaction between Y76 and W69 and the substrate's cytosine ring. The Fe specifically was stabilized by a hydrogen bond interaction via R210. This same residue was proposed by Fang *et al.* to stabilize the TS.[82] In other work, it has also been suggested that R210 aids in hydrogen bonding with the oxo and succinate group.[78] This aids in the isomerization necessary for the "in-line" binding mode of O_2. Lastly, I143 and W178 aid in the stabilization of the TS in which CO_2 is being lost and the Fe reorients to an Fe(IV)=O intermediate.[48] The same site, W178, was also important for modulating the two possible tunnels for O_2 transport.[75]

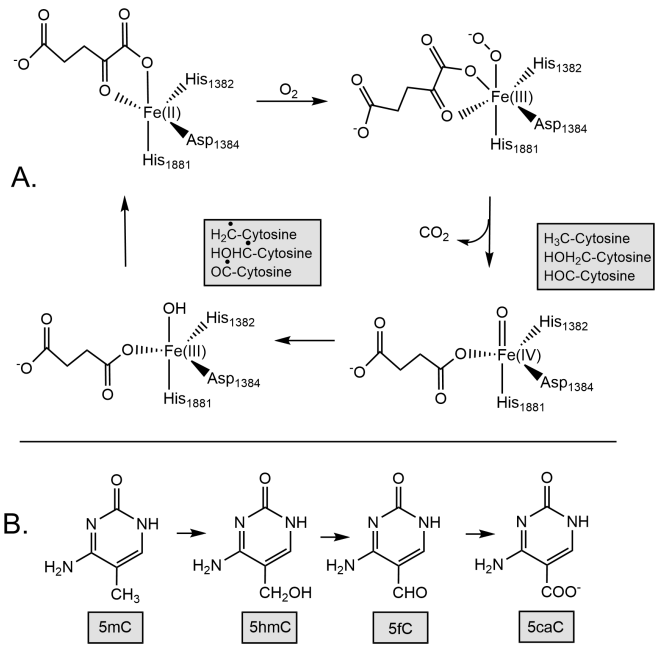
Effects of mutations far away from the active site have been shown to have a significant effect on AlkB enzymes. For example, a prostate-cancer associated mutation on ALKBH7 that results in the ALKBH7 R191Q variant was uncovered and characterized recently.[45] The mutation site at position 191 is located over 20 Å away from the active site. Interestingly, although it does not result in overall changes to the ternary structure, it affects a loop that produces minor changes that propagate to the active site. This variant was studied theoretically and experimentally and the results show that wild type ALKBH7 binds the required cofactors in the active site. However, the E191Q mutant does not bind the cofactors in the active site due to the structural changes induced by the cancer mutation.

3 Mechanistic Studies of TET2

TET2 (Figure 1B), similar to AlkB, catalyzes the demethylation of DNA, albeit with a major preference for a single substrate, 5mC, and using an iterative oxidation mechanism (Scheme 3A). Rather

than a direct demethylation of 5mC, TET2 does so indirectly in three sequential steps following single-hit kinetics: 5mC to 5hmC, 5hmC to 5fC, and 5fC to 5caC (Scheme 3B).[30,112–114] Recently, it has been shown that TET1-3 enzymes can also carry out direct demethylation with a preference for substrates that lack a 5-methyl group such as 4-N-methyl-5-methylcytosine or 4,4-N,N-dimethyl-5-methylcytosine.[115]

The indirect demethylation of each intermediate is similar to the AlkB enzyme mechanism in that these three phases are present: (1) formation of the reactive Fe(IV)=O, (2) hydrogen atom abstraction from the substrate, followed by (3) hydroxyl rebound (for 5hmC and 5caC) or second H abstraction (for 5fC) (Scheme 3A). To determine if the hydrogen atom abstraction step was also the rate-limiting step of TET2, Hu *et al.* carried out a deuterium exchange analysis.[114] Hydrogen atoms on the methylated base (5mC) were replaced by deuterium, which lead to a significant decrease in the enzymatic activity of TET2 thus confirming this is also the rate-limiting step for TET2.



Scheme 3 (A) The demethylation reaction carried out by TET2 and (B) the sequence of methylated cytosine bases resulting from one round of oxidation.

3.1 5mC to 5hmC

The first stage in the TET2 catalyzed demethylation involves the oxidation of 5mC of 5hmC. This step is similar to the AlkB catalyzed reaction, however, the main difference is that there is no disproportionation to formaldehyde and cytosine after the OH rebound step. It has been recently reported that Alkb, ALKBH2 and ALKBH3 can also carry out the iterative oxidation of 5mC, albeit with a 5–fold lower efficiency compared to the repair of 1mA.[87] Interestingly, AlkB family enzymes can also repair other neutral alkylated bases such as 3mT and 1mG.[63,64] However, similar to the reduced activity for 5mC oxidation by AlkB enzymes, the efficiency for the repair of these uncharged damaged bases is also significantly reduced compared with charged substrates repaired by AlkB family enzymes.[63,64]

Lu *et al.* performed QM/MM simulations to investigate all the oxidation steps catalyzed by TET2.[27] In the crystal structure of TET2 with 5mC used by Lu *et al.* (PDB:4NM6), it was noted that residue D1384 adopted an orientation that was different when compared with the other two crystals in which 5hmC and 5fC were bound.[27,116] In addition, no second-shell water was observed between D1384 and 5mC, which they hypothesized to be consistent with the more hydrophobic methyl of 5mC.

In the first reaction phase where the dioxygen is bound to Fe and attacks the α-kg, this reaction was found to be highly exergonic with a relatively small reaction energy barrier of 7.6 kcal/mol for 5mC. The same reorientation from axial to equatorial as seen in AlkB was also observed in TET2 for the oxidation of 5mC. Here, the dioxygen is initially trans to H1382 but reorients to trans w.r.t. H1881. This reorientation was deemed necessary because of the large electrostatic attraction between R1286 and the peroxy bridge. From there, the peroxy bridge must be broken leading to the formation of Fe(IV)=O. This phase of the reaction was found to be irreversible and quick to proceed.

Once the Fe(IV)=O has formed, the rate-limiting step then proceeds. Previous work has shown that for a quintet spin state, the σ-pathway is favored resulting in a linear attack.[117] This was also seen in the calculations by Lu *et al.* and Waheed *et al.*[27,44] The relative potential energy barrier for this step was found to be 15.5 kcal/mol for 5mC.[27] Similar results were found for this step by Waheed *et al.* (16.3 kcal/mol).[44] The final stage of the reaction for this oxidation step is the hydroxyl rebound. Lu *et al.* found a barrier of 16.9 kcal/mol, which suggests the hydroxyl rebound is slightly higher in energy than the HAT from the 5mC substrate.[27] This effect is proposed to be due to the freely rotating methyl that allows for a low hydrogen abstraction barrier that allows for a hydrogen atom to be consistently oriented toward the activated ferryl leading to an efficient abstraction.[27] Conversely, Waheed *et al.* reported a barrier of 10.1 kcal/mol for the rebound step.[44]

3.2 5hmC to 5fC

The oxidation of 5hmC to 5fC involves a hydrogen atom transfer from the hydroxyl moiety, and a proton transfer from the methylene moiety on 5hmC. Two separate groups have reported that the ferryl intermediate includes a second–shell water that hydrogen bonds to the oxygen atom on the ferryl intermediate.[27,114,118] Similar to the second–shell water for the ferryl intermediate in the AlkB family enzymes mentioned above, this water plays an important structural and electronic role.[27,50,114,118]

In the work performed by Lu *et al.*, the reaction begins with the formation of the peroxy bridge. This causes the hydroxyl group of 5hmC to be in close enough proximity to interact with R1261 and the peroxy moiety.[27] In addition, experimental work has shown that 5hmC (and 5fC) is highly likely to form intra-molecular hydrogen bonds making the reaction more difficult to proceed.[114] These additional interactions prevent the total reorientation of the peroxy bridge and ultimately affect the geometry around the Fe center.[27] In addition, the following phase of the reaction in

which the bridge is broken leads to a square pyramidal geometry around the Fe as seen in 5mC and 5fC. However, the oxo group is trans w.r.t. H1382 rather than trans w.r.t H1882 as seen in the other two systems. Due to the bond between the oxo group and the water, the reorientation is crucial for the reaction to proceed to the rate-limiting step much like AlkB. The barrier for this reorientation found by Lu *et al.* was 6.9 kcal/mol.[27] After the reorientation, a bond with the second-shell water molecule is also observed leading to an unfavorable orientation between the oxo and the hydroxyl hydrogen of 5hmC.[27] Based on this structural arrangement, the first step for the HAT/proton transfer reported by Lu *et al.* corresponds to the proton abstraction from the methylene, with a barrier of 18.7 kcal/mol, followed by the HAT from the hydroxyl with a corresponding barrier of 7.0 kcal/mol.

By contrast, Torabifard *et al.* used QM/MM calculations to investigate the rate limiting steps only (HAT and proton abstraction).[118] These simulations were based on structures obtained from previous simulations that reported the importance of an active site scaffold in TET2.[43] The optimized ferryl intermediate and the energetic features of the reaction reported by Torabifard *et al.* share several similarities with the work of Lu *et al.* including the second–shell water forming a hydrogen bond with the ferryl oxygen. However, the orientation of the 5hmC substrate in the optimized intermediate is such that the HAT from the hydroxyl moiety is found to occur first, with a barrier of 20.1 kcal/mol. This is followed by the proton transfer with a corresponding barrier of 7.3 kcal/mol.[118]. Thus, although the calculated barriers agree in magnitude, the two reported mechanisms differ with regards to the specific rate–limiting step involved in this stage. This point underscores the importance of the arrangement of all components in the active site (Scheme 3) and suggests that both mechanisms may be at play depending on the specific geometry of the system.

3.3 5fC to 5caC

Similar to the 5hmC system, the crystal structure of TET2 with 5fC contains a second shell water between D1384 and 5fC (PDB:5D9Y).[114] The barrier for the formation of the peroxy bridge was found to be 7.8 kcal/mol.[27] Similar to the 5mC system, following the formation of the peroxy bridge with Fe, the dioxygen changes from trans w.r.t. H1382 to trans w.r.t. H1881.[27] While 5fC does not require an additional reorientation as seen in the 5hmC reaction, the barrier for the hydrogen atom transfer is significantly higher than the other two steps (26.7 kcal/mol).[27] The group hypothesized that this large difference in energy may result from the restrained conformation of the formyl group. Because the σ- pathway is favored, the need for a linear attack is crucial which is difficult for the rigid formyl group of 5fC. A higher energy barrier hydroxyl rebound was also determined in this system (20.7 kcal/mol).[27]

3.4 The effects of mutations on TET2 reactivity

While it has been known that TET2 can perform the sequential oxidation of 5mC, the question was raised as to why 5fC and 5caC are formed if 5hmC is stable in cells.[43] Liu *et al.* proposed that TET2 is specifically built to provide support for these higher order oxidations and a scaffold in the active site is responsible for the sequential oxidation of these intermediates. Two key residues within this scaffold were found to be conserved among many of the TET enzymes (T1372 and Y1902) and subjected to experimental and computational mutagenesis studies.[43] When saturation mutagensis was initially performed on T1372, three distinct phenotypes were observed: 1) WT-like, 2) 5hmC dominant, and 3) catalytically dead.

Subsequent MD simulations were carried out on all four methylated intermediates for WT several T1372 TET2 variants.[43] The mutants were found to have different effects on inter-molecular interactions within the active site depending on the identity of the residue employed to replace T1371. Hydrogen bond and EDA further revealed a scaffold that consists of T1372-Y1902, which is necessary to orient the substrate in the active site for efficient catalysis. Subsequent MD simulations on Y1902 predicted that a Y1902F variant would exhibit a 5hmC-dominant phenotype and a T1372A/Y1902F variant would rescue catalytic activity. Experimental mutagenesis confirmed these predictions and provided support for the importance of this scaffold in the active site of TET2.

Furthering this work, a computational investigation into the catalytic mechanism for the oxidation of 5hmC by the T1372E TET2 variant was carried out to investigate the reason for the 5hmC-dominant phenotype for this variant.[118] It was found that the T1372E variant has a different electron configuration for the ferryl intermediate (^{HS}Fe-O_{AF}) arising from an altered orientation of the substrate compared to WT, which has a similar configuration to AlkB, ALKBH2 and ALKBH3 (^{IS}Fe-O_F) (see above). The mutation in this scaffold eliminates a crucial hydrogen bond between T1372 and Y1902 and instead, hydrogen bonds to 5hmC. This altered orientation results in a shift in the second-shell water position where the oxygen is pointed toward the oxo moiety. These changes lead to a highly stabilized product where hydrogen bond can then form after the HAT between the second-shell water and the newly formed water. By extension, the newly formed water can then form a hydrogen bond with 5fC. The orientation of 5hmC changes within the active site resulting in a almost two-fold increase in the energy barrier for the oxidation of 5hmC to 5fC by T1372E TET2. These results provide further support for the importance of the proper orientation of the substrate in the active site of TET2 facilitated by the active site scaffold.

Hu *et al.* investigated a variety of mutants and their effect on 5mC oxidation.[114] They found that the double mutant K1299E-S1303N decreases TET2 activity significantly. QM/MM calculations were carried out on this mutant by Waheed *et al.*[44] A much higher barrier (25.0 kcal/mol) was found when compared with the WT (16.3 kcal/mol) for the hydrogen atom abstraction step indicating that this double mutation decreases the rate of reaction. In addition, the K1299E-S1303N mutant proceeds via the π-pathway rather than the σ-pathway as found in the WT. Torabifard *et al.* found that K1299 has a significant impact on catalysis by EDA, and showed that K1299 aids in the stabilization of one of the transition states of the oxidation of 5hmC.[118] Additional QM/MM was done on the S1290A-Y1295A, Y1902A and

N1387A mutants and similar barriers were found.[44] In addition, Hu *et al.* studied the R1261G mutant and found that TET2 activity was again disrupted by this mutation.[114] Torabifard *et al.* found strong interactions between R1261 and α-kg indicating a large stabilization effect for the TS.[118] This same residue was found to form hydrogen bonds with the hydroxyl group of 5hmC rather than the oxo group bonding to 5hmC thus requiring a reorientation of the Fe(IV)=O moiety.[27] This required reorientation leads to an increase in the energy barrier for 5hmC.

As mentioned above, there is typically a decreasing concentration of the analogues of 5mC found within organisms and this is likely due to the conformation of the methylated substrate within the active site. Sappa *et al.* proposed that if the active site were altered, a more efficient turnover of the 5mC analogues could be found.[119] The group began by replacing a variety of hydrophobic and polar residues with alanine that were near to the active site. From there, the enzyme was exposed to 5mC and the concentrations of the intermediates (5hmC, 5fC and 5caC) were measured. Several mutants (T1372A, T1393A and T1883A) showed a high preference for 5hmC (>80%) which was speculated to occur because of the need for 5hmC to change orientation to further oxidation (5hmC-dominant phenotype). In addition, they found that the V1395A mutant synthesized 5caC as the dominant intermediate with the highest yield (94%). This mutant, when compared with WT, was also more efficient at acting upon 5mC and 5fC. Similar mutations with small amino acid side chains were implemented for residue V1395 and both V1395G and V1395S were able to successfully oxidize 5mC to 5caC. The authors believe the active site of TET2 may be crowded and smaller side chains may aid in a higher catalytic efficiency of TET2.[119]

3.5 Substrate preference

While a majority of the work done this far on the TET2 enzyme has been with respect to DNA substrates, it is worth discussing the ability of TET enzymes to repair RNA methylated bases. The various forms of methylated cytosine such as 5mC, 5hmC, 5fC and 5caC can also be found in different types of RNA (mRNA, tRNA, rRNA and ncRNA).[120–122] While concentrations may vary across different organisms, the need for repair enzymes is still crucial. Experimental and computational work done by DeNizio *et al.* looked at the preference of TET2 for ssDNA, ssRNA, dsDNA and dsRNA.[42,123] Their results indicate that while TET2 can tolerate both ssDNA and ssRNA, it prefers ssDNA which was also corroborated by Fu *et al.*.[124] Similarly, dsDNA is highly preferred to dsRNA and hybrid DNA:RNA configurations can also be tolerated if the substrate is DNA. It was speculated that this strong disfavor for dsRNA may arise from the preference of RNA for the A-form structure.

4 Summary and Perspective

Computational methods paired with experimental work have shown to be very useful in elucidating the mechanistic details of the Fe/α-kg dependent enzymes discussed here. Both the AlkB and TET families are an integral part of the DNA/RNA repair and modification enzymes found across all species and are particularly important for cancer prevention. The AlkB family still has some unanswered questions that require further work. The O_2 tunnels proposed by Torabifard *et al.* indicate that O_2 could bind in two different positions (trans w.r.t. H187 or trans w.r.t. H131) leading to different reaction paths. The binding of O_2 is a pivotal phase of the reaction that will ultimately allow for the activation of the Fe moiety.

While both enzyme families can dealkylate a variety of substrates, they each have preferences for certain bases. Mechanistically, AlkB can directly dealkylate damaged bases via one round of catalysis whereas TET2 can catalyze both direct and indirect dealkylation. The question can still be raised as to why these enzymes have specific preferences. AlkB, ALKBH2, ALKBH3 and TET2 have all been shown to demethylate 5mC and its derivatives.[87] Bian *et al.* proposed that the AlkB enzymes need to bind 5mC in the syn-conformation to allow the reaction to proceed. This binding conformation would be similar to the preferential methylated bases that AlkB enzymes commonly repair such as 3mC. Additionally, the preferred TET2 substrate is neutral, while several AlkB family preferred substrates are cationic, with possible dealkylation of neutral damaged bases. Nevertheless, the activity of AlkB family enzymes is significantly reduced for neutral alkylated bases. Additionally, the role of second–shell residues and other molecules has been shown to be important for these enzymes, similar to other inorganic systems that carry out similar C-H activation reactions[82,118,125]. These second–shell residues are also likely to play an important role in selectivity. Future work may help shed light on the drivers for the selectivity of members from both of these families.

Author Contributions

Author contributions are as follow: ARW and EVM performed data curation, formal analysis, visualization and investigation for the ALKBH2 and ALKBH3 section. MBB performed data curation, investigation, formal analysis, visualization and writing of the original draft. GAC was involved in supervision, funding securement, conceptualization, formal analysis, methodology and software. All authors contributed to reviewing and editing.

Funding

This work was funded by NIH R01GM108583 and NSF CHE-1856162. Computing time from CASCaM with partial support from NSF CHE-1531468 is gratefully acknowledged. MBB thanks MolSSI for support via a software development fellowship. E.A.V.-M. wishes to acknowledge CONACyT for funding.

Conflicts of interest

Authors declare no conflicts.

Acknowledgements

The authors thank Prof. L. Noodleman and Prof. R. Lord for insightful discussions.

Notes and references

1 R. P. Hausinger, *Crit. Rev. Biochem. Molec. Biol.*, 2004, **39**, 21–68.

2 S. Martinez and R. P. Hausinger, *J. Biol. Chem.*, 2015, **290**, 20702–20711.

3 C. Q. Herr and R. P. Hausinger, *Trends Biochem. Sci.*, 2018, **43**, 517–532.

4 M. S. Islam, T. M. Leissing, R. Chowdhury, R. J. Hopkinson and C. J. Schofield, *Annu. Rev. Biochem.*, 2018, **87**, 585–620.

5 Q. Yu, S. Liu, L. Yu, Y. Xiao, S. Zhang, X. Wang, Y. Xu, H. Yu, Y. Li, J. Yang, J. Tng, H.-C. Duan, L.-H. Wei, H. Zhang, J. Wei, Q. Tang, C. Wang, W. Zhang, Y. Wang, P. Song, Q. Lu, W. Zhang, S. Dong, B. Song, C. He and G. Jia, *Nat. Biotechnol.*, 2021.

6 S.-S. Gao, N. Naowarojna, R. Cheng, X. Liu and P. Liu, *Nat. Prod. Rep.*, 2018, **35**, 792–837.

7 *2-Oxoglutarate-Dependent Oxygenases*, ed. R. P. Hausinger and C. J. Schofield, The Royal Society of Chemistry, 2015.

8 S. Álvarez-Barcia and J. Kästner, *J. Phys. Chem. B*, 2017, **121**, 5347–5354.

9 M. Hibi and J. Ogawa, *Appl. Microbiol. Biotechnol.*, 2014, **98**, 3869–3876.

10 B. Wang, D. Usharani, C. Li and S. Shaik, *J. Am. Chem. Soc.*, 2014, **136**, 13895–13901.

11 S. Goudarzi, S. R. Iyer, J. T. Babicz Jr., J. J. Yan, G. H. J. Peters, H. E. M. Christensen, B. Hedman, K. O. Hodgson and E. I. Solomon, *P.N.A.S.*, 2020, **117**, 5152–5159.

12 C. R. Zwick III and H. Renata, *Nat. Prod. Rep.*, 2020, **37**, 1065–1079.

13 L.-F. Wu, S. Meng and G.-L. Tang, *B.B.A.*, 2016, **1864**, 453–470.

14 C. Loenarz and C. J. Schofield, *Trends Biochem. Sci.*, 2011, **36**, 7–18.

15 D. Sun, D. Gao, P. Xu, Q. Guo, Z. Zhu, X. Cheng, S. Bai, H.-M. Qin and F. Lu, *Protein. Expr. Purif.*, 2018, **149**, 1–6.

16 D. Mandal, D. Mallick and S. Shaik, *Acc. Chem. Res.*, 2018, **51**, 107–117.

17 S. Pandian, M. A. Vincent, I. H. Hillier and N. A. Burton, *Dalton Trans.*, 2009, 6201–6207.

18 T. Borowski, H. Noack, M. Radoń, K. Zych and P. E. Siegbahn, *J. Am. Chem. Soc.*, 2010, **132**, 12887–12898.

19 D. Branzei and M. Foiani, *Nat. Rev. Mol. Cell Biol.*, 2008, **9**, 297–308.

20 S. Kumar, V. Chinnusamy and T. Mohapatra, *Frontiers in Genetics*, 2018, **9**, 640.

21 F. Drabløs, E. Feyzi, P. A. Aas, C. B. Vågbø, B. Kavli, M. S. Bratlie, J. Peña-Diaz, M. Otterlei, G. Slupphaug and H. E. Krokan, *DNA Repair*, 2004, **3**, 1389–1407.

22 B. I. Fedeles, V. Singh, J. C. Delaney, D. Li and J. M. Essigmann, *J. Biol. Chem.*, 2015, **290**, 20734–20742.

23 E. M. Leddin and G. A. Cisneros, *Advances in Protein Chemistry and Structural Biology*, Elsevier, 2019.

24 G. Zheng, Y. Fu and C. He, *Chem. Rev.*, 2014, **114**, 4602–4620.

25 F. Chen, Q. Tang, K. Bian, Z. T. Humlock, X. Yang, M. Jost, C. L. Drennan, J. M. Essigmann and D. Li, *Chem. Res. Toxicol.*, 2016, **29**, 687–693.

26 B. Yu, W. C. Edstrom, J. Benach, Y. Hamuro, P. C. Weber, B. R. Gibney and J. F. Hunt, *Nature*, 2006, **439**, 879–884.

27 J. Lu, L. Hu, J. Cheng, D. Fang, C. Wang, K. Yu, H. Jiang, Q. Cui, Y. Xu and C. Luo, *Phys. Chem. Chem. Phys.*, 2016, **18**, 4728–4738.

28 F. Lyko, *Nat. Rev. Genet.*, 2018, **19**, 81–92.

29 D. Cheishvili, L. Boureau and M. Szyf, *Br. J. Pharmacol.*, 2015, **172**, 2705–2715.

30 S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He and Y. Zhang, *Science*, 2011, **333**, 1300–1303.

31 D. Li, B. Guo, H. Wu, L. Tan and Q. Lu, *Cytogenet. Genome Res.*, 2015, **146**, 171–180.

32 K. D. Rasmussen and K. Helin, *Genes Dev.*, 2016, **30**, 733–750.

33 N. J. Tsiouplis, D. W. Bailey, L. F. Chiou, F. J. Wissink and A. Tsagaratou, *Front. cell Dev. Biol.*, 2021, **8**, 1829.

34 S. Ye, J. C. Price, E. W. Barr, M. T. Green, J. M. Bollinger, Jr, C. Krebs and F. Neese, *J. Am. Chem. Soc.*, 2010, **132**, 4739–4751.

35 T. Borowski, A. Bassan and P. E. Siegbahn, *Chem. Eur. J.*, 2004, **10**, 1031–1041.

36 S. P. de Visser, *J. Am. Chem. Soc.*, 2006, **128**, 9813–9824.

37 J. M. Bollinger, J. C. Price, L. M. Hoffart, E. W. Barr and C. Krebs, *Eur. J. Inorg. Chem.*, 2005, 4245–4254.

38 J. C. Price, E. W. Barr, B. Tirupati, J. M. Bollinger and C. Krebs, *Biochem.*, 2003, **42**, 7497–7508.

39 M. Costas, M. P. Mehn, M. P. Jensen and L. Que, *Chem. Rev.*, 2004, **104**, 939–986.

40 H. Liu, J. Llano and J. W. Gauld, *J. Phys. Chem. B*, 2009, **113**, 4887–4898.

41 E. L. Hegg and L. Q. Jr., *Eur. J. Biochem.*, 1997, **250**, 625–629.

42 J. E. DeNizio, M. Y. Liu, E. M. Leddin, G. A. Cisneros and R. K. Kohli, *Biochem.*, 2019, **58**, 411–421.

43 M. Y. Liu, H. Torabifard, D. J. Crawford, J. E. DeNizio, X.-J. Cao, B. A. Garcia, G. A. Cisneros and R. M. Kohli, *Nat. Chem. Bio.*, 2017, **13**, 181–187.

44 S. O. Waheed, S. S. Chaturvedi, T. G. Karabencheva-Christova and C. Z. Christov, *ACS Catl.*, 2021, **11**, 3877–3890.

45 A. R. Walker, P. Silvestrov, T. A. Müller, R. H. Podolsky, G. Dyson, R. P. Hausinger and G. A. Cisernos, *PloS Comput. Biol.*, 2017, **13**, 1–13.

46 P. Silvestrov, T. A. Müller, K. N. Clark, R. P. Hausinger and G. A. Cisneros, *J. Mol. Graph. Model.*, 2014, **54**, 123–130.

47 S. A. Lenz, D. Li and S. D. Wetmore, *DNA Repair*, 2020, **96**, 1–10.

48 S. O. Waheed, R. Ramanan, S. S. Chaturvedi, N. Lehnert, C. J. Schofield, C. Z. Christov and T. G. Karabencheva-Christova, *ACS Cent. Sci.*, 2020, **6**, 795–814.

49 H. M. Senn and W. Thiel, *Angew. Chem. Int. Ed.*, 2009, **48**, 1198–1229.

50 M. A. Hix, E. M. Leddin and G. A. Cisneros, *J. Chem. Theory Comput.*, 2021, **17**, 4524–4537.

51 H. J. Kulik, J. Zhang, J. P. Klinman and T. J. Martínez, *J. Phys. Chem. B*, 2016, **120**, 11381–11394.

52 A. O. Dohn, *Int. J. Quantum Chem.*, 2020, **120**, 1–22.

53 M. Nottoli and F. Lipparini, *J. Chem. Phys.*, 2020, **153**, 224108.

54 J. Nochebuena, S. Naseem-Khan and G. A. Cisneros, *WIREs Comp. Mol. Sci/*, 2021, **11**, e1515.

55 D. Loco, L. Lagardère, O. Adjoua and J.-P. Piquemal, *Acc. Chem. Res.*, 2021, **54**, 2812–2822.

56 N. A. Zarkevich and D. D. Johnson, *J. Chem. Phys.*, 2015, **142**, 024106.

57 H. L. Woodcock, M. Hodošček, P. Sherwood, Y. S. Lee, H. F. S. III and B. R. Brooks, *Theor. Chem. Acc.*, 2003, **109**, 140–148.

58 S. K. Burger and W. Yang, *J. Chem. Phys.*, 2006, **124**, 054109.

59 N. Hansen and W. F. van Gunsteren, *J. Chem. Theory Comput.*, 2014, **10**, 2632–2647.

60 X. Lu, D. Fang, S. Ito, Y. Okamoto, V. Ovchinnikov and Q. Cui, *Mol. Simul.*, 2017, **42**, 1056–1078.

61 H. Hu, Z. Lu and W. Yang, *J. Chem. Theory Comput.*, 2007, **3**, 390–406.

62 H. Hu, Z. Lu, J. M. Parks, S. K. Burger and W. Yang, *J. Chem. Phys.*, 2008, **128**, 034105.

63 J. C. Delaney and J. M. Essignmann, *PNAS*, 2004, **101**, 14051–14056.

64 P. Koivisto, P. Robins, T. Lindahl and B. Sedgwick, *J. Biol. Chem.*, 2004, **279**, 40470–40474.

65 D. Li, J. C. Delaney, C. M. Page, A. S. Chen, C. Wong, C. L. Drennan and J. M. Essigmann, *J. Nucleic Acids*, 2010, 1–9.

66 D. Li, J. C. Delaney, C. M. Page, X. Yang, A. S. Chen, C. Wong, C. L. Drennan and J. M. Essigmann, *J. Am. Chem. Soc.*, 2012, **134**, 8896–8901.

67 D. Li, B. I. Fedeles, N. Shrivastav, J. C. Delaney, X. Yang, C. Wong, C. L. Drennan and J. M. Essigmann, *Chem. Res. Toxicol.*, 2013, **26**, 1182–1187.

68 J. C. Delaney, L. Smeester, C. Wong, L. E. Frick, K. Taghizadeh, J. S. Wishnok, C. L. Drennan, L. D. Samson and J. M. Essigmann, *Nat. Struct. Molec. Biol.*, 2005, **12**, 855–860.

69 A. M. Maciejewska, K. P. Ruszel, J. Nieminuszczy, J. Lewicka, B. Sokołowska, E. Grzesiuk and J. T. Kuśmierek, *Mutat. Res.*, 2010, **684**, 24–34.

70 Y. Mishina, C.-G. Yang and C. He, *J. Am. Chem. Soc.*, 2005, **127**, 14594–14595.

71 T. Duncan, S. C. Trewick, P. Koivisto, P. A. Bates, T. Lindahl and B. Sedgwick, *PNAS*, 2002, **99**, 16660–16665.

72 D. Zdżalik, A. Domańska, P. Prorok, K. Kosicki, E. van de Born, P. Falnes, C. J. Rizzo, F. P. Guengerich and B. Tudek, *DNA Repair*, 2015, **30**, 1–10.

73 C. V. Sastri, J. Lee, K. Oh, Y. J. Lee, J. Lee, T. A. Jackson, K. Ray, H. Hirao, W. Shin, J. A. Halfen, J. Kim, J. Lawrence Que, S. Shaik and W. Nam, *P.N.A.S.*, 2007, **104**, 19181–19186.

74 E. Solomon, D. DeWeese and J. Babicz Jr., *Biochem.*, 2021.

75 H. Torabifard and G. A. Cisneros, *Chem. Sci.*, 2017, **8**, 6230–6238.

76 B. Yu, W. C. Edstrom, J. Benach, Y. Hamuro, P. C. Weber, B. R. Gibney and J. F. Hunt, *Nature*, 2006, **439**, 879–884.

77 A. Pavelka, E. Sebestova, B. Kozlikova, J. Brezovsky, J. Sochor and J. Damborsky, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 2016, **13**, 505–517.

78 M. G. Quesne, R. Latifi, L. E. Gonzalez-Ovalle, D. Kumar and S. P. de Visser, *Chem. Eur. J.*, 2014, **20**, 435–446.

79 C. Yi, B. Chen, B. Qi, W. Zhang, G. Jia, L. Zhang, C. J. Li, A. R. Dinner, C.-G. Yang and C. He, *Nat. Struct. Mol. Biol*, 2012, **19**, 671–676.

80 C.-G. Yang, C. Yi, E. M. Duguid, C. T. Sullivan, X. Jian, P. A. Rice and C. He, *Nature*, 2008, **452**, 961–965.

81 C. Geng, S. Ye and F. Neese, *Ang. Chem. Intl. Ed.*, 2010, **49**, 5717–5720.

82 D. Fang, R. L. Lord and G. A. Cisneros, *J. Chem. Phys*, 2013, **117**, 6410–6420.

83 P. Koivisto, T. Duncan, T. Lindahl and B. Sedgwick, *J. Biol. Chem.*, 2003, **278**, 44348–44354.

84 J. Ringvoll, M. N. Moen, L. M. Nordstrand, L. B. Meira, B. Pang, A. Bekkelung, P. C. Dedon, S. Bjeilland, L. D. Samson, P. O. Falnes and A. Klungland, *Cancer Res*, 2008, **68**, 4142–4149.

85 G. A. Cisneros, *Interdiscip. Sci. Comput. Life Sci.*, 2010, **2**, 70–77.

86 R. Latifi, J. L. Minnick, M. G. Quesne, S. P. de Visser and L. Tahsini, *Dalton Trans.*, 2020, **49**, 4266–4276.

87 K. Bian, S. A. Lenz, Q. Tang, F. Chen, R. Qi, M. Jost, C. L. Drennan, J. M. Essigmann, S. D. Wetmore and D. Li, *Nucleic Acids Res.*, 2019, **47**, 5522–5529.

88 P. Aas, M. Otterlei, P. Falnes, C. Vågbø, F. Skorpen, M. Akbari, O. Sundhein, M. Bjørås, G. Slupphaug, E. Seeberg and H. Krokan, *Nature*, 2003, **421**, 859–863.

89 S. Dango, N. Mosammaparast, M. E. Sowa, L.-J. Xiong, F. Wu, K. Park, M. Rubin, S. Gygi, J. W. Harper and Y. Shi, *Mol. Cell.*, 2012, **44**, 373–384.

90 O. Sundheim, C. B. Vågbø, M. Bjørås, M. M. Sousa, V. Talstad, P. A. Aas, F. Brabløs, H. E. Krokan, J. A. Trainer and G. Slupphaug, *EMBO J.*, 2006, **25**, 3389–3397.

91 C. J. Williams, J. J. Headd, N. W. Moriarty, M. G. Prisant, L. L. Videau, L. N. Deis, V. Verma, D. A. Keedy, D. J. Hintze, V. B. Chen, S. Jain, S. M. Lewis, W. B. A. 3rd, J. Snoeyink, P. D. Adams, S. C. Lovell, J. S. Richardson and D. C. Richardson, *Protein Sci.*, 2018, **27**, 293–315.

92 W. Jorgensen, J. Chandrasekhar, J. Madura, R. Impey and M. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

93 C. Schafmeister, W. Ross and V. Romanovski, *LEAP*, 1995, University of California, San Francisco.

94 D. Case, T. Darden, I. T.E. Cheatham, C. Simmerling, J. Wang, R. Duke, R. Luo, R. Walker, W. Zhang, K. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. Göetz, I. Kolossvary, K. Wong, F. Paesani, J. Vanicek,

R. Wolf, J. Liu, X. Wu, S. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. Roe, D. Mathews, M. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. Kollman, *AMBER 12*, 2012, University of California, San Francisco.

95 K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror and D. E. Shaw, *PLOS One*, 2012, **7**, 1.

96 R. Berendsen, J. Postma, W. van Gunsteren, A. Di Nola and J. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

97 S. Kuwajima and A. Warshel, *J. Chem. Phys.*, 1988, **89**, 3751–3759.

98 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089.

99 J. M. Parks, H. Hu, A. J. Cohen and W. Yang, *J. Chem. Phys.*, 2008, **129**, 154106.

100 J. A. Rackers, Z. Wang, C. Lu, M. L. Laury, L. Lagardére, M. J. Schnieders, J.-P. Piquemal, P. Ren and J. W. Ponder, *J. Chem. Theory Comput.*, 2018, **14**, 5273–5289.

101 M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, G. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. Janesko, R. Gomperts, B. Mennucci, H. Hratchian, J. Ortiz, A. Izmaylov, J. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. I. nad T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. Montgomery, J. Peralta, F. Ogliaro, M. Bearpark, J. Heyd, E. Brothers, K. Kudin, V. Staroverov, T. Keith, R. Kobayashi, J. Normad, K. Raghavachari, A. Rendell, J. Burant, S. Iyengar, J. Tomasi, M. Cossi, K. Morokuma, O. Farkas, J. Foresman and D. Fox, *Gaussian16 Revision C.01*, 2016, Gaussian Inc. Wallingford, CT.

102 E. G. Kratz, A. R. Walker, L. Lagardére, F. Lipparini, J.-P. Piquemal and G. A. Cisneros, *J. Comp. Chem.*, 2016, **37**, 1019–1029.

103 H. Gökcan, E. Vazquez-Montelongo and G. Cisneros, *J. Chem. Theo. Comput.*, 2019, **15**, 3056–3065.

104 P. Verma, K. D. Vogiatzis, N. Planas, J. Borycz, D. J. Xiao, J. R. Long, L. Gagliardi and D. G. Truhlar, *J. Am. Chem. Soc.*, 2015, **137**, 5770–5781.

105 D. Fang and G. A. Cisneros, *J. Chem. Theory Comput.*, 2014, **10**, 5136–5148.

106 P. Grzyska, E. Appelman, R. Hausinger and D. Proshlyakov, *Proc. Natl. Acad. Sci. U.S.A.*, 2010, **107**, 3982–3987.

107 C. Yi, G. Jia, G. Hou, Q. Dai, Q. Zhang, G. Zheng, X. Jian, C.-G. Yang, Q. Cui and C. He, *Nature*, 2010, **468**, 330–333.

108 E. Solomon, S. Goudrazi and K. Sutherlin, *Biochem.*, 2016, **55**, 6363–6374.

109 P. Bruijnincx, G. van Koten and R. Gebbink, *Chem. Soc. Rev.*, 2008, **37**, 2716–2744.

110 A. Faponle, M. Quesne, C. Sastri, F. Banse and S. de Visser, *Chemistry*, 2015, **21**, 1221–1236.

111 P. J. Holland and T. Hollis, *PLOS ONE*, 2010, **5**,.

112 M. Tahiliana, K. Koh, Y. Shen, W. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. Iyer, D. Liu, L. Aravind and A. Rao, *Science*, 2009, **324**, 930–935.

113 Y. He, B. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. Song, K. Zhang, C. He and G. Xu, *Science*, 2011, **333**, 1303–1307.

114 L. Hu, J. Lu, J. Cheng, Q. Rao, H. Hou, Z. Lou, L. Zhang, W. Li, W. Gong, M. Liu, C. Sun, X. Yin, J. Li, X. Tan, P. Wang, Y. Wang, D. Fang, Q. Cui, P. Yang, C. He, H. Jiang, C. Luo and Y. Xu, *Nature*, 2015, **527**, 118–122.

115 U. Ghanty, T. Wang and R. M. Kohli, *Angew. Chem. Int. Ed. Engl.*, 2020, **59**, 11312–11315.

116 L. Hu, Z. Li, J. Cheng, Q. Rao, W. Gong, M. Liu, Y. G. Shi, J. Zhu, P. Wang and Y. Xu, *Cell*, 2013, **155**, 1545–1555.

117 S. Ye and F. Neese, *P.N.A.S.*, 2011, **108**, 1228–1233.

118 H. Torabifard and G. A. Cisneros, *Chem. Sci.*, 2018, **9**, 8433–8445.

119 S. Sappa, D. Dey, B. Sudhamalla and K. Islam, *J. Am. Chem. Soc.*, 2021, **143**, 11891–11896.

120 Y. Motorin, F. Lyko and M. Helm, *Nucleic Acids Res.*, 2010, **38**, 1415–1430.

121 J. Squires, H. Patel, M. Nousch, T. Sibbritt, D. Humphreys, B. Parker, C. Suter and T. Preiss, *Nucleic Acids Res.*, 2012, **40**, 5023–5033.

122 W. Huang, M. Lan, C. Qi, S. Zheng, S. Wei, B. Yuan and Y. Feng, *Chem. Sci.*, 2016, **7**, 5495–5502.

123 E. M. Leddin and G. A. Cisneros, *Advances in Protein Chemistry and Structural Biology*, Academic Press, 2019, vol. 117, pp. 91–112.

124 L. Fu, C. Guerrero, N. Zhong, N. Amato, Y. Liu, S. Liu, Q. Cai, D. Ji, S. Jin, I. Niedernhofer, G. Pfeifer, G. Xu and Y. Wang, *J. Am. Chem. Soc.*, 2014, **136**, 11582–11585.

125 J. G. Vitillo, C. C. Lu, C. J. Cramer, A. Bhan and L. Gagliardi, *ACS Catalysis*, 2021, **11**, 579–589.