UNIVERSITY OF LEEDS

This is a repository copy of *Computational studies of DNA base repair mechanisms by nonheme iron dioxygenases: selective epoxidation and hydroxylation pathways*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/217680/

Version: Accepted Version

Article:

Latifi, R., Minnick, J.L., Quesne, M.G. orcid.org/0000-0001-5130-1266 et al. (2 more authors) (2020) Computational studies of DNA base repair mechanisms by nonheme iron dioxygenases: selective epoxidation and hydroxylation pathways. Dalton Transactions, 49 (14). pp. 4266-4276. ISSN 1477-9226

https://doi.org/10.1039/d0dt00007h

© The Royal Society of Chemistry 2020. This is an author produced version of an article published in Dalton Transactions. Uploaded in accordance with the publisher's self-archiving policy.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

ARTICLE

Computational studies of DNA base repair mechanisms by nonheme iron dioxygenases: Selective epoxidation and hydroxylation pathways

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Reza Latifi,*[a] Jennifer L. Minnick,[a] Matthew G. Quesne,*[b,c] Sam P. de Visser*[d] and Laleh Tahsini*[a]

DNA base repair mechanisms of alkylated DNA bases is an important reaction in chemical biology and in the human body it is typically catalyzed by an α-ketoglutarate-dependent nonheme iron dioxygenase named the AlkB repair enzyme. In this work we report a detailed computational study into the structure and reactivity of AlkB repair enzymes with alkylated DNA bases. In particular, we investigate the aliphatic hydroxylation and C=C epoxidation mechanisms of alkylated DNA bases by a high-valent iron(IV)-oxo intermediate. Our computational studies use quantum mechanics/molecular mechanics methods on full enzymatic structures as well as cluster models on active site systems. The work shows that the iron(IV)-oxo species is rapidly formed after dioxygen binding to an iron(II) center and passes a bicyclic ring structure as intermediate. Subsequent, cluster models explore the mechanism of substrate hydroxylation and epoxidation of alkylated DNA bases. The work shows low energy barriers for substrate activation and consequently energetically feasible pathways are predicted. Overall, the work shows that a high-valent iron(IV)-oxo species can efficiently dealkylate alkylated DNA bases and return them into their original form.

Introduction

DNA base repair enzymes have important functions in biology to fix damaged bases and return their structure and function to their natural form.[1] DNA bases can be damaged through unnatural effects such as X-ray radiation or chemical agents. However, the body in some cases modifies DNA bases temporarily as a defence mechanism. Thus, during DNA replication, the DNA bases of the mother strand are typically methylated to protect it from damage by DNA polymerase activities. However, after the replication is complete these methyl groups need to be removed again to make the DNA functional. The enzymes involved in this process are the AlkB repair enzymes that typically oxidize the alkylated bases.[2] For instance, N^3-methyl cytosine (3-meC) is believed to be hydroxylated at the methyl group by a nonheme iron dioxygenase and released as formaldehyde. Experimental work showed these enzymes to utilize α-ketoglutarate and dioxygen on a nonheme iron centre and as such they were classified within the class of nonheme iron/α-ketoglutarate-dependent dioxygenases.[3] These nonheme iron dioxygenases are versatile enzymes in biology with functions varying from the biosynthesis of natural products, including hormones such as flavonols in plants and antibiotics, but also react through the biodegradation of toxic compounds.[4]

In general, the nonheme iron/α-ketoglutarate-dependent dioxygenases have a mononuclear iron centre that binds the protein through a typical 2-His/1-carboxylate facial ligand system whereby the carboxylate is typically the side chain of a Glu or Asp residue.[5] The active site of a representative AlkB repair enzyme is given in Fig. 1 as taken from the 3I2O protein databank (pdb) file.[6,7] The active site has a central iron(II) atom that binds the protein via the side chains of His_{131}, Asp_{133} and His_{187} and in addition binds α-ketoglutarate (αKG), the co-substrate. The position trans to His_{131} (the axial ligand) is the proposed binding site of molecular oxygen,[8] which binds as iron(III)-superoxo and reacts with αKG to form succinate and an iron(IV)-oxo species. For several nonheme iron/α-ketoglutarate-dependent dioxygenases this iron(IV)-oxo species has been characterized with UV-Vis, Raman and Mössbauer spectroscopy studies.[9]

---

[a] 107 Physical Science Building, Department of Chemistry, Oklahoma State University, Stillwater, 74078 OK, USA.
[b] Cardiff University, School of Chemistry, Main Building, Park Place, Cardiff, CF10 3AT, United Kingdom.
[c] Research Complex at Harwell, Rutherford Appleton Laboratory, Harwell Oxford, Didcot, Oxon, OX110FA, United Kingdom..
[d] Manchester Institute of Biotechnology and Department of Chemical Engineering and Analytical Science, 131 Princess Street, Manchester M1 7DB, UK.

† Footnotes relating to the title and/or authors should appear here.

Electronic Supplementary Information (ESI) available: Optimized geometries of structures described in this work as well as Tables with Energies, group spin densities and charges and a full set of Cartesian coordinates is available. See DOI: 10.1039/x0xx00000x

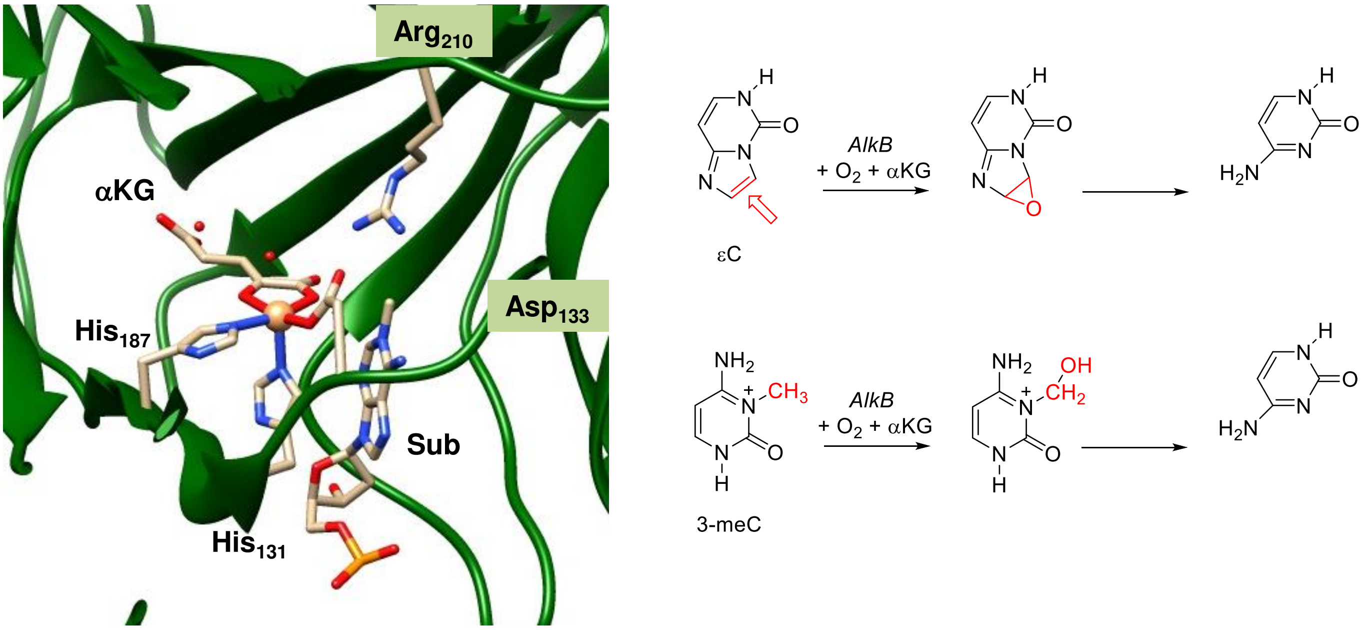


Fig. 1. Active site structure of AlkB repair enzymes as taken from the 3I2O pdb file and typical repair reactions of alkylated cytosine by AlkB repair enzymes.

Moreover, computational modelling showed this process to be stepwise with an initial electrophilic attack on the α-keto-position of αKG to form a bicyclic ring structure followed by decarboxylation to form persuccinate. In a final dioxygen bond cleavage step an iron(IV)-oxo and succinate is formed, which is the active species of the enzyme.[10] Overall, AlkB repair enzymes have been shown to activate a range of alkylated DNA bases including the exocyclic adducts 1,N^6-etheno adenine (εA), 3,N^4-etheno cytosine (εC) and 1,N^2-etheno guanine (εG). In addition, there are methylated DNA bases: N^3-methyl thymine (3-meT), N^1-methyl adenine (1-meA) and N^3-methyl cytosine (3-meC).[6] Thus, the exocyclic alkylated DNA bases are expected to be epoxidized by the active species of AlkB, while the methylated DNA bases have been suggested to be hydroxylated prior to a demethylation reaction (Fig. 1).

Previous computational studies on AlkB repair enzymes mainly focused on methylated DNA bases, such as 3-meC and 1-meA. The work established a stepwise mechanism starting with a hydrogen atom abstraction by an iron(IV)-oxo species prior to radical rebound to form alcohol.[11] In addition, detailed molecular mechanics (MM) and molecular dynamics (MD) studies established the flexibility of the protein and it was found that after iron, cofactor and substrate binding the system is relatively stable.[12] Finally, a QM/MM study on the hydroxylation of N^1-methyl adenine (1-meA) by AlkB repair enzymes proposed a mechanism where dioxygen is converted into an iron(IV)-oxo species, which undergoes an isomerization gated by Arg$_{210}$ followed by hydrogen atom abstraction from substrate.[8]

To gain further insight into the mechanistic details of alkylated DNA bases and particularly, those that contain exocyclic structures and hence react through epoxidation, we decided to do a further combined MD, QM/MM and DFT study. The work shows that a high-valent iron(IV)-oxo species will be highly efficient for converting these alkylated DNA bases back into their original forms as low reaction barriers are found for all substrates.

Methods

QM/MM calculations

Our calculations utilized a well-established set of QM/MM protocols[13] and started from the crystal structure coordinates of the substrate and αKG bound iron(II) resting state of AlkB.[14] The pdb2pqr software package was initially used to add protons to the model,[15] using PROPKA at pH 7, followed by a careful visual analysis of the protonation states of all titratable residues. The iron(II) bound resting state was manually converted into an iron(III)-superoxo species using the oxygen atom of crystal water 601 that occupies the sixth coordination site as the proximal oxygen atom of O_2. An iterative solvation procedure was performed in a sphere with radius of 35Å to give a total system of 21,611 atoms. Thereafter, the Charmm force field,[16] was adopted in a 800 ps molecular dynamics (MD) simulation, where no geometric restraints were placed on the protein components, of the system. Subsequently, snapshots along the MD trajectory were taken at 300, 400 and 500 ps: Sn_{300}, Sn_{400} and Sn_{500}.

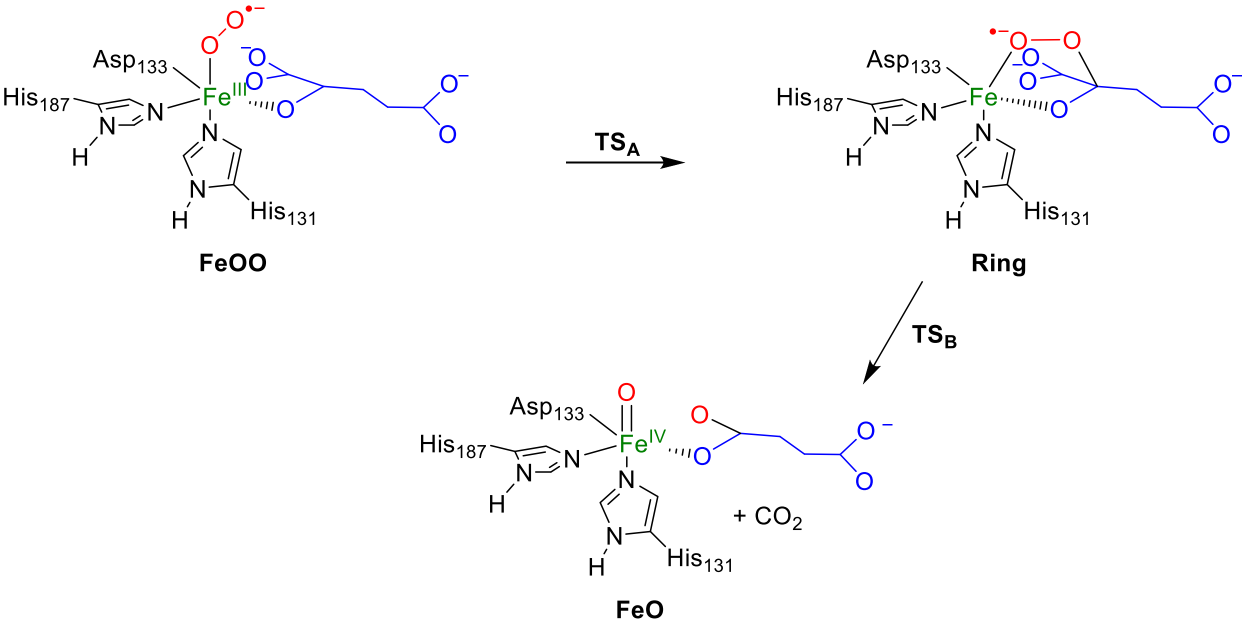
Two sets of calculations with different QM regions were performed with both including the iron(III)-superoxo group, methylimidazole groups for His_{131} and His_{187}, acetate for αKG and Asp_{133}, methylguanidinium for Arg_{210}, and the N^1-methyladenine part of the substrate. This QM region (model A) was comparable to that used in previous QM/MM work on AlkB.[8] In addition, a system with a larger QM region (model B) was explored that contained the two closest water molecules to the iron-oxygen centre in order to rule out any critical effects of the explicitly solvent molecules.

The QM region was described with the hybrid density functional method B3LYP and the full mechanism was calculated for three different snapshots Sn_{300}, Sn_{400}, Sn_{500}. Test calculation were run with the pure density functional method BP86 and showed the same trends.[17] Spin state ordering and reactivity trends were conserved across all methods and QM regions. Therefore, unless otherwise indicated, we focus here on QM region A as calculated at UB3LYP level of theory with all data being available in supporting information.[18] For all calculations the triplet-ζ basis set TZVP (basis set BS3) was added to all atoms of the QM-region, whilst the CHARMM force field[16] was applied to all groups in the MM-region via the DL-Poly software.[19]

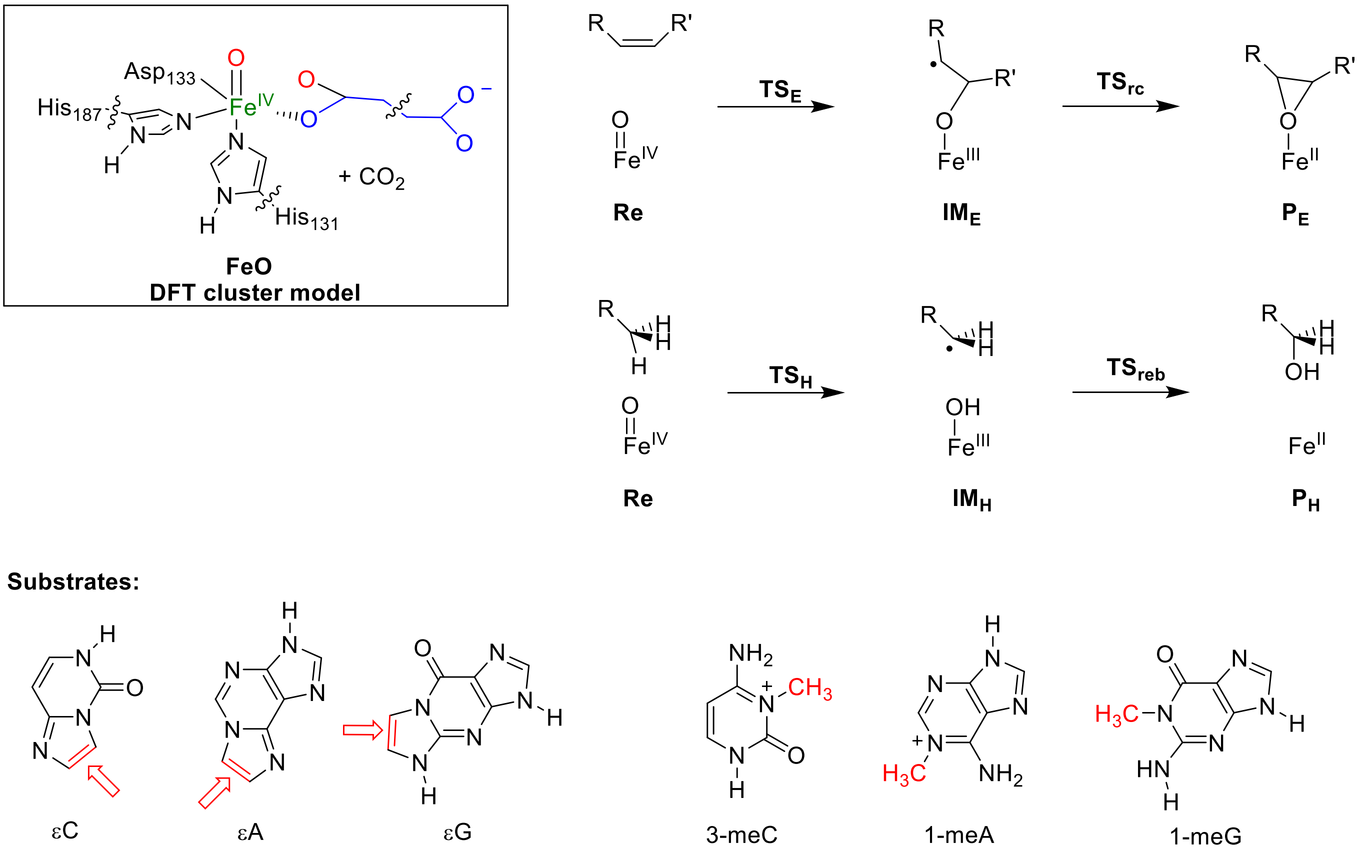
QM/MM calculations focused on the dioxygen activation by AlkB repair enzymes and studied the mechanism displayed in Scheme 1. The studies started from an iron(III)-superoxo complex (**FeOO**) and explored the attack of oxygen on αKG to form a bicyclic ring structure (**Ring**) via a transition state $\mathbf{TS_A}$. Thereafter, O-O bond cleavage leads to an iron(IV)-oxo species (**FeO**) and splits off CO_2 via transition state $\mathbf{TS_B}$.

Cluster model calculations

Alongside the QM/MM calculations also cluster models were investigated that started from an iron(IV)-oxo species and focused on the mechanisms of substrate activation of alkylated DNA bases. These contained small models of the first-coordination sphere of ligands to iron, whereby His_{131} and His_{187} were abbreviated to imidazole and Asp_{133} and succinate to acetate. The cluster model is schematically depicted in Scheme 2 alongside the mechanisms for substrate hydroxylation and epoxidation we studied. The set of alkylated DNA bases of cytosine, guanine and adenine with their nomenclature is given at the bottom of Scheme 2 and focused on εC, εA, εG, 3-meC, 1-meA and 1-meG. Substrate epoxidation was investigated as a stepwise process starting from a reactant complex (**Re**) through electrophilic attack of the oxo on one of the carbon atoms of the double bond via transition state $\mathbf{TS_E}$ to form a radical intermediate $\mathbf{IM_E}$. A subsequent ring-closure via transition state $\mathbf{TS_{rc}}$ leads to epoxide products ($\mathbf{P_E}$). For methylated DNA bases a mechanism starting with a hydrogen atom abstraction (via transition state $\mathbf{TS_H}$) leads to an iron(III)-hydroxo radical complex $\mathbf{IM_H}$ that after radical rebound (via transition state $\mathbf{TS_{reb}}$) gives alcohol product complexes ($\mathbf{P_H}$). Full details of the optimized geometries for all landscapes are given in the Electronic Supporting Information.



Scheme 1. Reaction mechanism of dioxygen activation by AlkB enzymes and the formation of a high-valent iron(IV)-oxo species.



Scheme 2. Reaction mechanisms of substrate epoxidation and hydroxylation of alkylated DNA bases by iron(IV)-oxo as calculated with DFT.

All calculations follow well established procedures in our groups and have been extensively benchmarked and validated.[20] We use the unrestricted hybrid density functional method UB3LYP,[18] for geometry optimizations, constraint geometry scans and frequency calculations in combination with a double–ζ quality LACVP basis set on iron that contains a core potential and 6–31G on all other atoms: basis set B1.[21] Full optimizations were done in *Gaussian* 0.9 without constraints and followed by an analytical frequency calculation.[22] Subsequent single point calculations with a triple–ζ quality LACV3P+ basis set on iron that contains a core potential and 6–311+G* on the rest of the atoms were done, basis set B2.

All local minima described here had real frequencies only, while the transition states were characterized by a single imaginary mode for the correct vibration.

Results and discussion

Our work on the mechanistic features of AlkB repair enzymes started with a detailed solvation and molecular dynamical equilibration setup protocol as a precursor to QM/MM calculations. Fig. 2 demonstrates that the solvation and equilibration steps were continued until almost no additional water molecules were able to migrate into the system and there was a saturation point reached with a steady equilibrium. All snapshots were taken after the RMSD had plateaued to sample a representative set of equilibrated starting structures. Whilst these different snapshots gave slightly different reaction energies they all led to the same basic reactivity trends and therefore we will be focusing mostly on the Sn_{500} snapshot here, with a detailed breakdown of all the reaction energetics available in the Electronic Supporting Information.

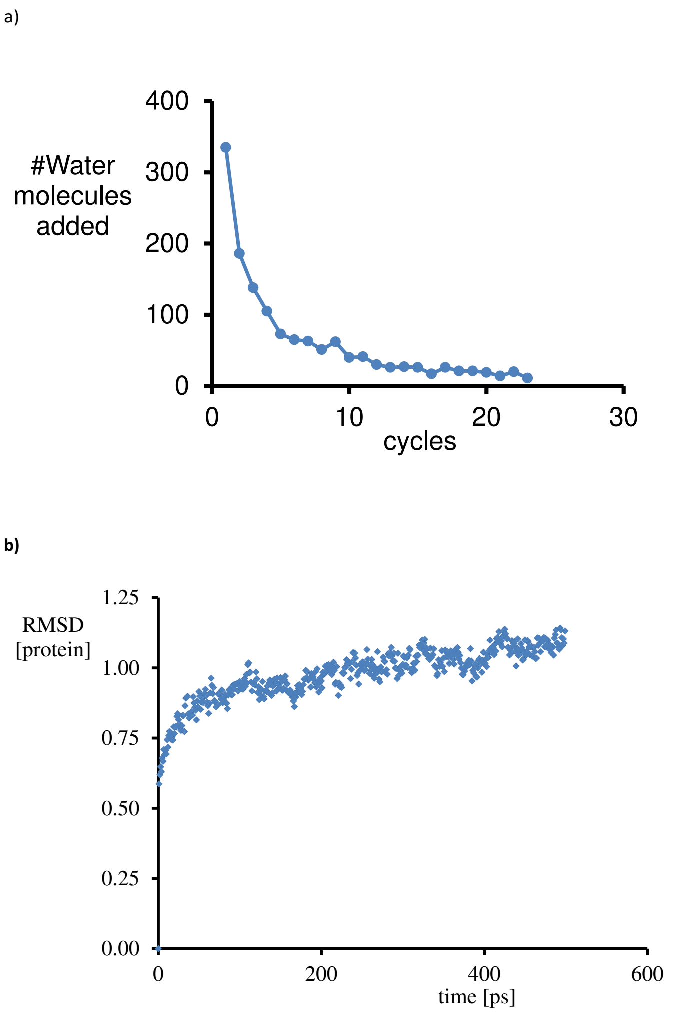
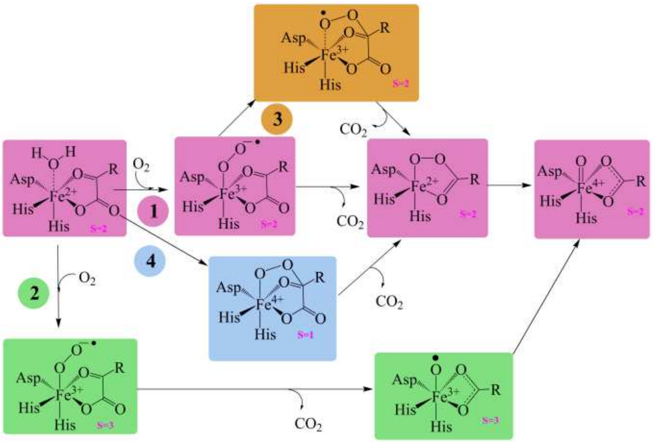


Fig. 2. a) Solvation procedure used to iteratively solvate the system Calculations performed with the CHARMM force field as implemented in the CHARMM software package b) Molecular dynamics simulation of the full model system with the RMSD value as a function of time for a period of 500 ps.

There is still a large amount of speculation concerning the critical oxygen activation step performed by this superfamily of biocatalysts, with independent groups undertaking many *in silico* studies into the formation of the active ferryl-oxo species.[23,24] However, the dominant potential energy surfaces (PES) observed, varied wildly depending on the choice of system and exact functional used, with no consensus as yet been reached. The reaction pathways of the four most prominent theories from the literature are represented in Scheme 3. The first of the four dominant theories was put forward by Borowski et al,[25] who found by using a cluster model of the active site region of clavaminic acid synthase, in conjunction with the B3LYP hybrid DFT method, that O_2 activation proceeded on a dominant quintet potential energy surface, whereby the attack of an iron(III)-superoxo radical on the β-carbon of αKG leads to a rate limiting decarboxylation of the coenzyme. The resulting Fe(II)-peracid intermediate was subsequently cleaved to form the ferryl-oxo. In the same year Borowski et al.[23a] also published a DFT/B3LYP study into O_2 activation using a model of the first coordination sphere of 4-hydroxyphenylpyruvate dioxygenase. However, whilst this new study did not exclude the possibility of mechanism 1 (Scheme 3), the accessibility of a second high-spin septet potential energy surface was also investigated. In this mechanism (mechanism 2 in Scheme 3) the now antiferromagnetic coupled Fe(III)-superoxo radical attacks the α-keto carbon leading to O–O heterolysis and decarboxylation steps in a consecutive manner, with no stable peracid intermediate predicted. Subsequent work by Borowski and co-workers has shown varying accessibility of these two pathways by different αKG dependent oxygenase models.[10e,26]

The third mechanism in Scheme 3 is the one proposed by Topol et al.,[23f] which emerged from a DFT/B3LYP study of an asparaginyl hydroxylase active site model. As with mechanism 1 a low-lying quintet potential energy surface was reported; however, this time the rate limiting step was found to involve the attack of the superoxo on the co-enzyme, which resulted in a new intermediate bicyclic species. Subsequent work by one of us added further credibility to mechanism 3.[23c] In this study a DFT/B3LYP model of the active region of TauD was used and the same rate limiting step and bicyclic superoxo intermediate was found. The final mechanism (4) was proposed in 2011 by Diebold et al.[23b] and involved modelling the active region of the enzyme with a modified BP86 density functional that included a 10% Hartree-Fock component. The authors found that the oxygen bound to both the iron and αKG forming a triplet ground state Fe(IV)-alkyl peroxo bridged intermediate. Subsequent decarboxylation of the coenzyme leads to the same quintet Fe(II)-peracid observed in both mechanisms 1 and 3.



Scheme 3. Four proposed mechanisms for O_2 activation determined by different DFT cluster models.

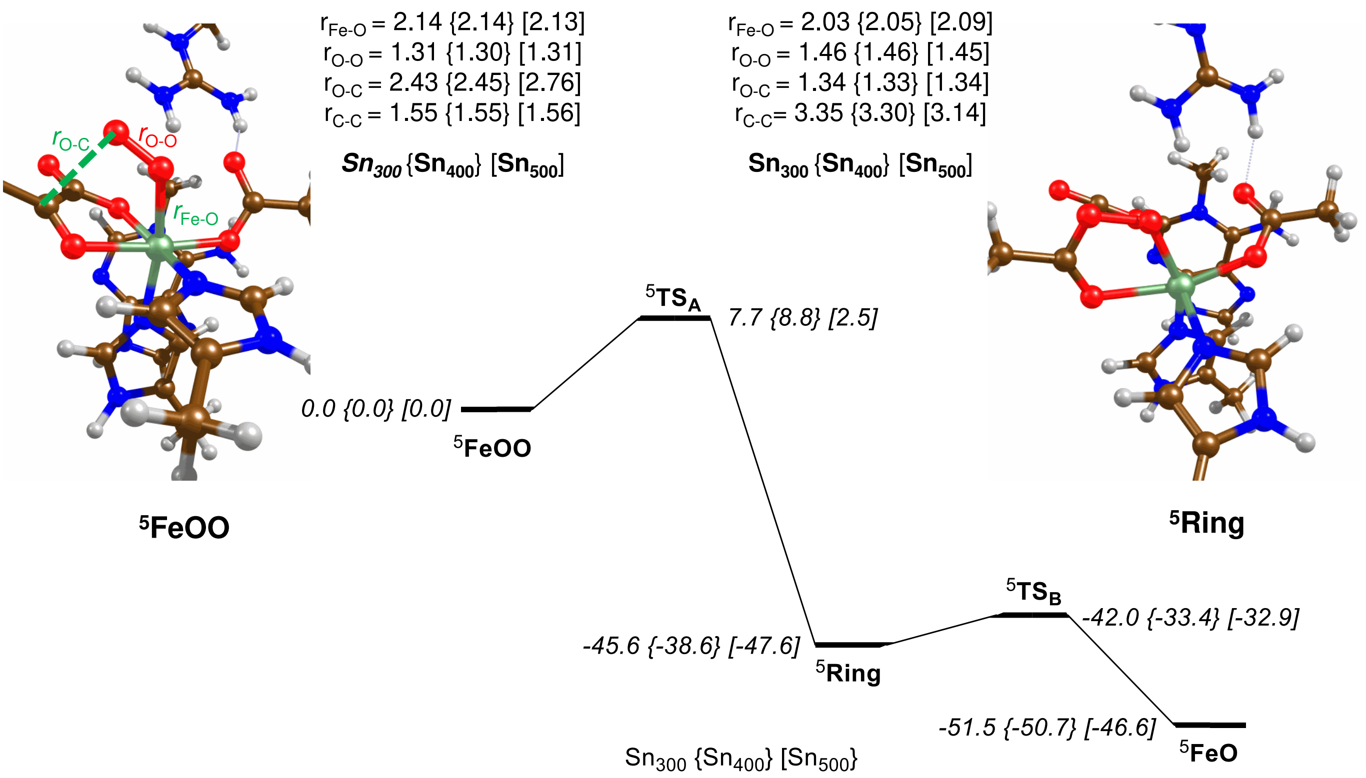


Fig. 3. QM/MM (Turbomole:Charmm) calculated potential energy landscape for the conversion of the iron(III)-superoxo into an iron(IV)-oxo species. Relative energies obtained with basis set BS3. are QM/MM energies in kcal mol^{-1}, while optimized geometries give bond lengths in angstroms. Date presented are for snapshots Sn_{300}, Sn_{400} and Sn_{500}.

In order to access which one of these hypotheses were most reasonable for AlkB, we calculated the conversion of the iron(III)-superoxo complex into an iron(IV)-oxo complex with QM/MM using snapshots Sn_{300}, Sn_{400} and Sn_{500} (see Fig. 3). In the reactant structure, we find close-lying quintet and septet spin state structures. Thus, the quintet and septet spin state have the same electronic configuration, namely $\pi^*_{xy}{}^1\, \pi^*_{xz}{}^1\, \pi^*_{yz}{}^1\, \sigma^*_{z2}{}^1\, \sigma^*_{x2-y2}{}^1\, \pi^*_{OO,xz}{}^1$. However, the $\pi^*_{OO,xz}$ is occupied with an up-spin electron in the septet spin state, while there is a down-spin electron in the quintet spin state. These close-lying spin states would imply a multistate reactivity patterns with reaction mechanisms on various spin state surfaces.[27] The triplet spin state is well higher in energy by at least 7 kcal mol^{-1} in all snapshots. In general, the reaction mechanisms on the lowest energy spin state surface is found for the quintet spin state; hence we will focus on this state in Fig. 3. The septet spin barriers are well higher in energy due to the lack of available 3d orbitals for the transfer of one electron. The spin-state ordering matches previous DFT calculations using small model complexes well.[10a,10b]

The geometry of the ^{5}**FeOO** complex as optimized in snapshots Sn_{300}, Sn_{400} and Sn_{500} and the results are given in Fig 3. As can be seen the structures are very similar with Fe–O distances of 2.14, 2.14 and 2.13Å in Sn_{300}, Sn_{400} and Sn_{500}, respectively. Similar small deviations are seen for the O–O and C–C distances, while the intermolecular distance O–C shows a bit more fluctuation as would be expected from different protein structures. Nevertheless, the different snapshots give similar structures and electronic configurations.

Subsequently, we ran constraint geometry scans for the attack of the terminal oxygen atom of the iron(III)-superoxo on the α-keto position of αKG and located transition states for all three snapshots (^{5}**TS$_A$**). Barriers of 7.7, 8.8 and 2.5 kcal mol^{-1} are obtained in snapshots Sn_{300}, Sn_{400} and Sn_{500}, respectively.

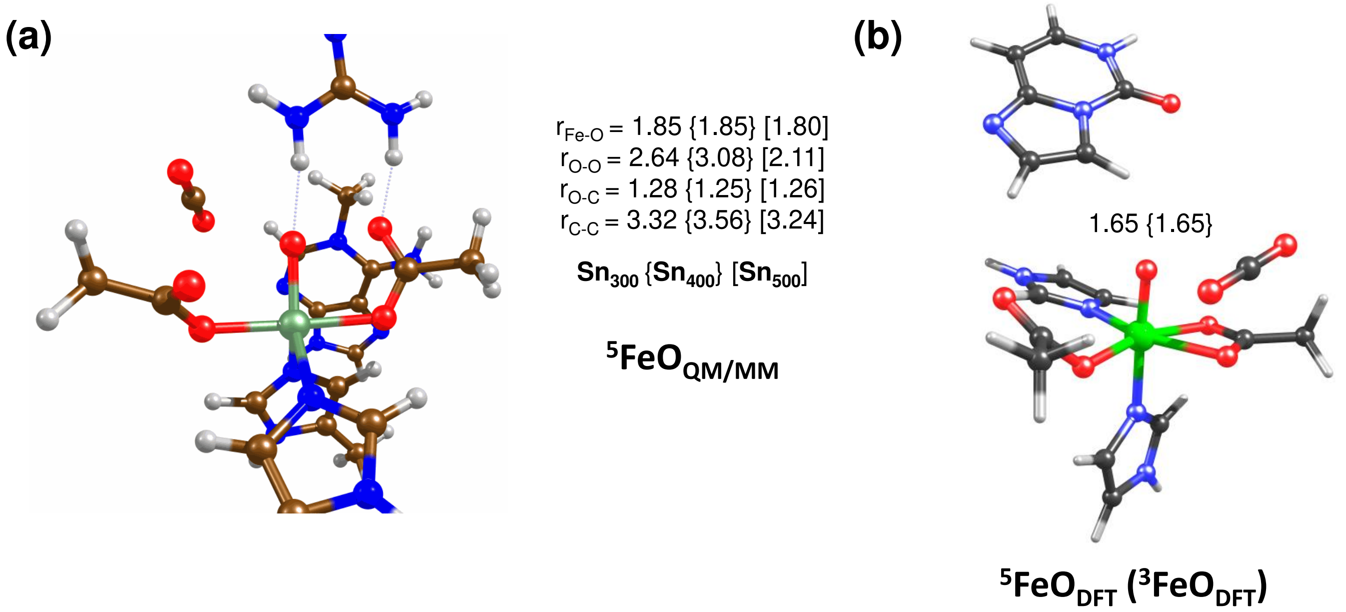


Fig. 4. QM/MM and DFT optimized geometries of the iron(IV)-oxo species (^{5}**FeO**). Bond lengths are in angstroms.

Crucially, this reaction step was inaccessible with the septet species because the transfer of a β-electron from the $\pi^*_{OO,xz}{}^1$ to the singly occupied $\pi^*_{xy}{}^1$ orbital is spin-forbidden. Therefore, whilst our calculations point to an equilibrium between S = 2 and S = 3 states, only the quintet superoxo could be considered as active for αKG decarboxylation. Previously, using gas-phase DFT cluster models a barrier of 10.4 kcal mol^{-1} was reported.[10a,10b,28] As such, the QM/MM approach gives only minor differences in the oxygen activation barriers and consequently the second-coordination sphere effect and the long-range interactions of the protein contribute little to the energies of this reaction step.

After ^{5}**TS$_A$** the structure relaxes to persuccinate structure through release of CO$_2$; hence no stable bicyclic ring structure was observed in the QM/MM calculations. This structure is considerably lower in energy than the iron(III)-superoxo reactant and hence, passing **TS$_A$** will be an irreversible process. The optimized geometries of the ring-structures are given in Fig. 3 and show elongation of the O–O bond from about 1.31Å in the ^{5}**FeOO** structure to 1.46Å. As such the peroxo bond changes from formally a double bond to a single bond. Interestingly, formation of the ring-structure strengthens the Fe-O bond and it is shortened by about 0.1Å. The QM/MM optimized geometries match those previously reported with DFT cluster methods well.[10,23–26]

The persuccinate structure is followed by a small barrier (^{5}**TS$_B$**) of 3.6 and 5.2 kcal mol^{-1} for snapshots Sn_{300} and Sn_{400}, while for Sn_{500} a bit larger barrier of 14.7 kcal mol^{-1} is found. Nevertheless, the pathways all lead to an iron(IV)-oxo species with high exothermicity. Indeed, experimental studies characterized the iron(IV)-oxo species using resonance Raman, Mössbauer and UV-Vis spectroscopy methods.[8] Therefore, in agreement with the energy landscape, the iron(IV)-oxo species is the most stable structure.

Optimized geometries of the **FeO** complexes as obtained with QM/MM and DFT cluster models are shown in Fig. 4. The structures match previously calculated iron(IV)-oxo species calculated with QM/MM or DFT well and give the same electronic configuration.

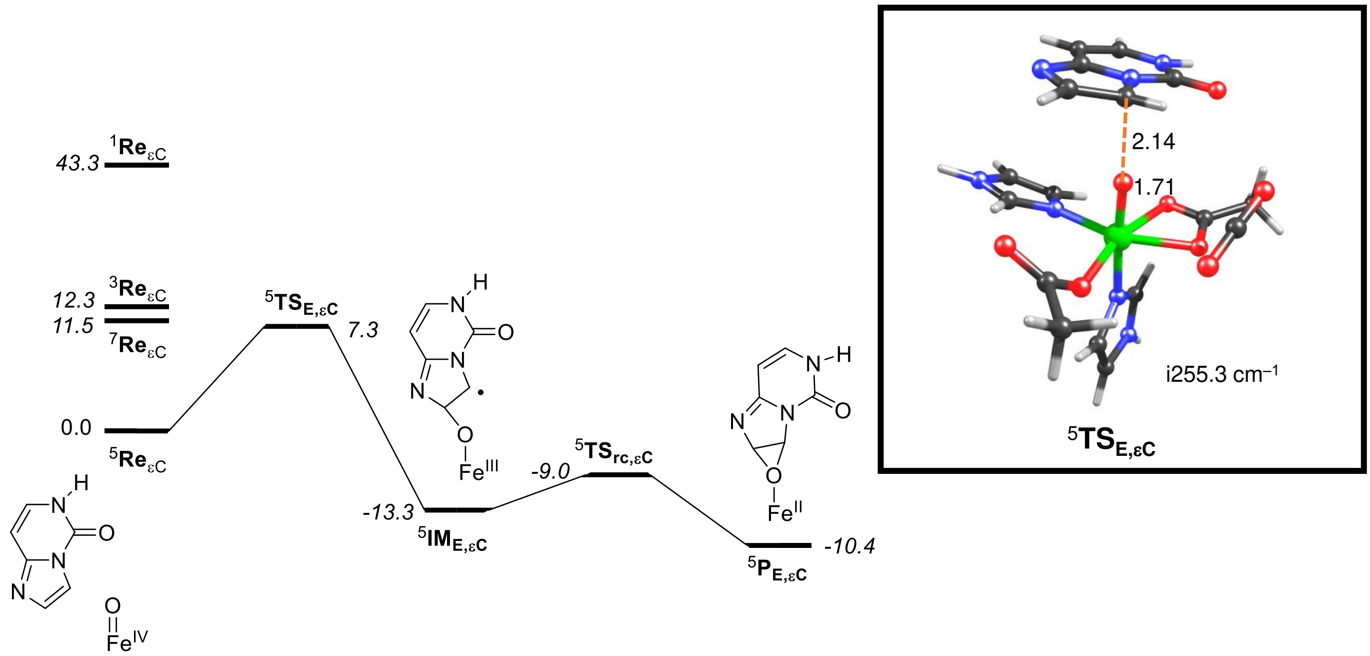


Fig. 5. Potential energy profile for εC activation by an iron(IV)-oxo model complex of AlkB. Free energies contain, zero-point, thermal, solvent and entropic corrections (at 298 K) and are given in kcal mol^{-1}. The optimized geometry of $^{5}\mathbf{TS}_{\mathbf{E,\varepsilon C}}$ gives bond lengths in angstroms and the imaginary frequency in cm^{-1}.

Subsequently, we decided to calculate dealkylation of alkylated DNA bases using small gas-phase DFT cluster models. In particular, we calculated substrate activation and particularly dealkylation by an iron(IV)-oxo model of AlkB. Fig 5 displays the potential energy landscape of oxygen atom transfer to the double bond of εC by an iron(IV)-oxo complex. The reactant complex is in a quintet spin ground state with the triplet, septet and singlet spin states higher lying by ΔG = 12.3, 11.5 and 43.3 kcal mol^{-1}, respectively. The ordering of the spin states of the iron(IV)-oxo species matches experimental EPR and Mössbauer measurements of analogous nonheme iron dioxygenases that characterized these species as quintet spin ground states.[9,29] Computational studies on analogous nonheme iron enzymes and biomimetic model complexes also found a quintet spin state ground state for pentacoordinated iron(IV)-oxo complexes.[23,27,30]

The quintet pathway has the lowest barriers and we find a free energy of activation of oxygen attack on one of the carbon atoms of εC of 7.3 kcal mol^{-1} to obtain a radical intermediate. The radical intermediate is more stable than the reactant complex by ΔG = 13.3 kcal mol^{-1} and through a small ring-closure barrier leads to epoxide products. The obtained stepwise mechanism is similar to that seen before by biomimetic iron(IV)-oxo complexes and enzymatic reaction mechanisms such as by cytochrome P450 Compound I.[31,32]

The optimized geometry of $^{5}\mathbf{TS}_{\mathbf{E,\varepsilon C}}$ is given in Fig 5 and is characterized by a long C–O distance of 2.14Å and a relatively short Fe–O distance of 1.71Å. As such the structure is reactant-like or "early" on the potential energy surface. The transition state has an imaginary frequency of i255.3 cm^{-1}, which is typical for substrate epoxidation barriers by nonheme iron(IV)-oxo oxidants.

Thereafter, we calculated the epoxidation reaction of εA and εG by the iron(IV)-oxo model of AlkB and the same reaction mechanism as that obtained by εC was found. The rate-determining transition states for activation of εA and εG are depicted in Fig. 6, namely $^{5}\mathbf{TS}_{\mathbf{E,\varepsilon A}}$ and $^{5}\mathbf{TS}_{\mathbf{E,\varepsilon G}}$. The free energies of activation of εA and εG are 8.3 and 9.5 kcal mol^{-1}, respectively. The barriers for activation of εA, εG and εC; therefore, range from 7.3 to 9.5 kcal mol^{-1}, which are small barriers in compared to typical olefin epoxidation by iron(IV)-oxo species.[32,33] Consequently, the epoxidation of exocyclic DNA bases by AlkB repair enzymes is expected to proceed fast and will lead to epoxide products rapidly.

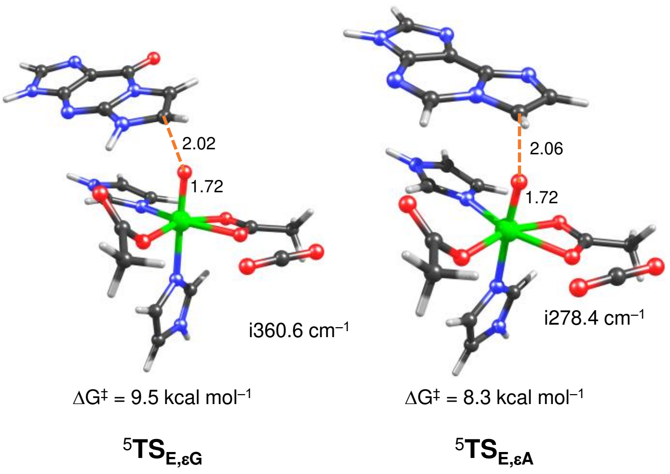


Fig. 6. Optimized geometries of epoxidation transition states for εG and εA substrates. Free energies of activation (in kcal mol^{-1}) include zero-point, thermal (298K), solvent and entropic corrections. Bond lengths are in angstroms and the imaginary frequency in cm^{-1}.

The structures are very similar to the $^{5}\mathbf{TS}_{\mathbf{E,\varepsilon C}}$ shown in Fig. 5, with a long C–O distance of 2.02 and 2.06Å for $^{5}\mathbf{TS}_{\mathbf{E,\varepsilon G}}$ and $^{5}\mathbf{TS}_{\mathbf{E,\varepsilon A}}$, respectively. Both structures have a short Fe–O distance of 1.72Å and, hence, are both “early” on the potential energy landscape.

Next, we investigated the hydroxylation of methylated DNA bases by AlkB and particularly focused on the activation of 3-meC, 1-meA and 5-meG bases by an iron(IV)-oxo model complex of AlkB. All reactions are stepwise with an initial hydrogen atom abstraction to form an iron(III)-hydroxo intermediate and are completed with a radical rebound step to form alcohol products. In all cases the rebound barriers are small and the initial hydrogen atom abstraction step is rate-determining. This matches previous studies on heme and nonheme iron(IV)-oxo reactivities.[34]

Optimized geometries of the hydrogen atom abstraction transition states of 3-meC, 1-meA and 5-meG activation by the iron(IV)-oxo species of AlkB are shown in Fig. 7. Energetically, the methylated cytosine group incurs a small hydrogen atom abstraction barrier of 6.5 kcal mol^{-1}, while for methylated adenine and guanine barriers of 19.3 and 10.6 kcal mol^{-1} are obtained, respectively. As such these processes will happen with dramatically different rate constants; although the size and shape of the protein pocket that binds substrate will play a role too.

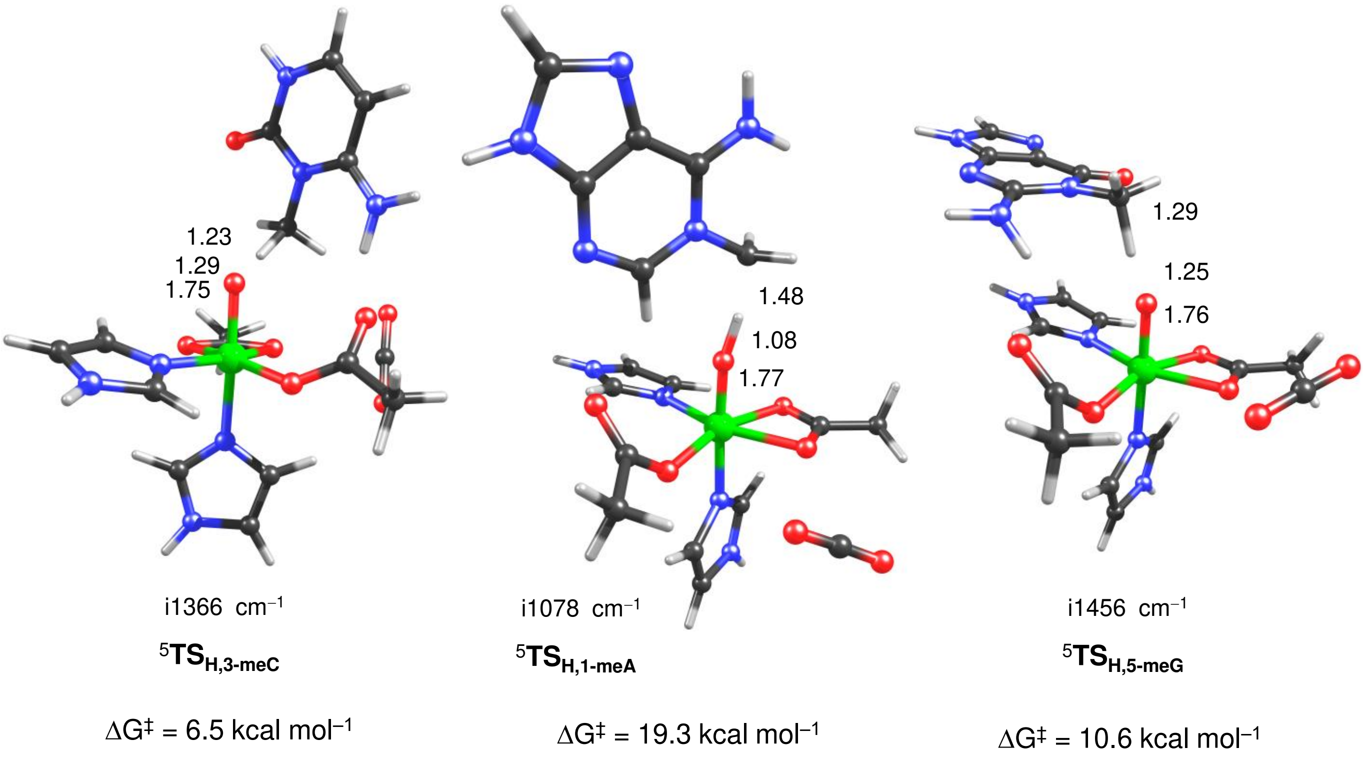


Fig. 7. Hydrogen atom abstraction barriers from methylated DNA bases 3-meC, 1-meA and 5-meG as calculated in Gaussian-09. Transition states have bond lengths in angstroms and the imaginary frequency in cm^{-1}. Free energies of activation (in kcal mol^{-1}) include zero-point, thermal (298K), solvent and entropic corrections.

Geometrically, in all transition states have similar Fe–O distances as the epoxidation barriers described in Figs. 5 and 6. The C-H and O-H distances of the transferring hydrogen atom; however, shows large fluctuation between the various structures. In $^{5}\mathbf{TS}_{\mathbf{H,3\text{-}meC}}$ the transferring hydrogen atom is almost midway in between the donor and acceptor groups and C–H and O–H distances of 1.23 and 1.29Å, respectively are obtained. In $^{5}\mathbf{TS}_{\mathbf{H,5\text{-}meG}}$, by contrast, the transferring hydrogen is also almost midway in between donor and acceptor group but slightly closer to the oxo: C–H distance of 1.29Å and O–H distance of 1.25Å are found. Interestingly, the $^{5}\mathbf{TS}_{\mathbf{H,1\text{-}meA}}$ structure is product like with a very short O–H distance of 1.08Å and a relatively long C–H distance of 1.48Å.

All three hydrogen atom abstraction transition states are characterized with a large imaginary frequency for the C–H–O stretch vibration with values well over i1000 cm^{-1}. This implies that the reaction will have a large kinetic isotope effect for the replacement of hydrogen by deuterium atom.

Conclusions

We present a systematic QM/MM and cluster model DFT study on the mechanistic features of AlkB repair enzyme. Our results show that the stepwise conversion of iron(iii)–superoxo to iron(iv)–oxo initiates with the rate determining attack of the terminal oxygen atom of ^{5}**Fe(III)OO** on the α-keto position of αKG. The reaction proceeds through relatively small barrier on the quintet spin state surface, $^{5}\mathbf{TS}_{\mathbf{A}}$. Subsequent relaxation of the transition state geometry yields ^{5}**Ring**, persuccinate structure with an elongated O–O bond to 1.46 Å, that is followed by releases of a CO_2 molecule through a very small barrier generating the iron(iv)–oxo species. We have also investigated the dealkylation of DNA bases through both epoxidation and hydroxylation reactions by small DFT cluster models. Electrophilic addition of oxygen atom of Fe=O group to the double bond in epoxidation and hydrogen abstraction by Fe= in hydroxylation are the initial and rate determining steps that generates radical intermediates. Our calculations show that the iron(iv)–oxo model of AlkB can dealkylate and repair DNA bases more efficiently than biomimetic models.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We gratefully acknowledge Oklahoma State University start-up funding (LT). The computing for this project was performed at the OSU High Performance Computing Center at Oklahoma State University supported in part through the National Science Foundation grant OAC-1126330.

The authors would like to thank Dr Kakali Sen and Prof. Walter Thiel from Max-Planck für Kohlenforschung, Mülheim for the support during the setting up of the QM/MM models. The UK Catalysis Hub is thanked for resources and support provided via membership of the UK Catalysis Hub Consortium and funded by EPSRC (Grants EP/R026815/1, EP/K014854/1, and EP/M013219/1).

Notes and references

1 a) A. E. Pegg, *Chem. Res. Toxicol*., 2011, **24**, 618; b) C. Escudé and J.-S. Sun. *Top. Curr. Chem*., 2005, **253**, 109.

2 a) P. J. O'Brien, *Chem. Rev*., 2006, **106**, 720; b) C. Yi, C. G. Yang and C. He, *Acc. Chem. Res*., 2009, **42**, 519.

3 a) M. A Kurowski, A. S. Bhagwat, G. Papaj and J. M. Bujnicki, *BMC Genomics*, 2003, **4**, 48; b) Y. Mishina, L. X. Chen and C. He, *J. Am. Chem. Soc*., 2004, **126**, 16930; c) C. Yi, G. Jia, G. Hou, Q. Dai, W. Zhang, G. Zheng, X. Jian, C.-G. Yang, Q. Cui and C. He, *Nature*, 2010, **468**, 330; d) R. W. D. Welford, I. Schlemminger, L. A. McNeill, K. S. Hewitson and C. J. Schofield, *J. Biol. Chem*., 2003, **278**, 10157; e) P. Ø. Falnes, R. F. Johansen and E. Seeberg, *Nature*, 2002, **419**, 178; f) S. C. Trewick, T. F. Henshaw, R. P. Hausinger, T. Lindahl and B. Sedgwick, *Nature*, 2002, **419**, 174; g) N. C. Giri, H. Sun, H. Chen, M. Costa and M. J. Maroney, *Biochemistry*, 2011, **50**, 5067.

4 a) E. I. Solomon, T. C. Brunold, M. I. Davis, J. N. Kemsley, S. K. Lee, N. Lehnert, F. Neese, A. J. Skulan, Y. S. Yang and J. Zhou, *Chem. Rev*., 2000, **100**, 235; b) M. Costas, M. P. Mehn, M. P. Jensen and L. Que Jr, *Chem. Rev*., 2004, **104**, 939; c) D. Buongiorno and G. D. Straganz, *Coord. Chem. Rev*., 2013, **257**, 541; d) S. P. de Visser, *Chem. Record*, 2018, **18**, 1501; e) K. Ray, F. F. Pfaff, B. Wang and W. Nam, *J. Am. Chem. Soc*., 2014, **136**, 13942.

5 a) L. Que Jr, *Nat. Struct. Biol*., 2000, **7**, 182; b) E. G. Kovaleva and J. D. Lipscomb, *Nat. Chem. Biol*., 2008, **4**, 186; c) P. C. A. Bruijnincx, G. van Koten and R. J. M. Klein Gebbink, *Chem. Soc. Rev*., 2008, **37**, 2716.

6 B. Yu and J. F Hunt, *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 14315.

7 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucl. Acids Res*., 2000, **28**, 235.

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

9 a) D. A. Proshlyakov, T. F. Henshaw, G. R. Monterosso, M. J. Ryle and R. P. Hausinger, *J. Am. Chem. Soc*., 2004, **126**, 1022; b) J. M. Bollinger Jr, J. C. Price, L. M. Hoffart, E. W. Barr and C. Krebs, *Eur. J. Inorg. Chem*., 2005, 4245; c) P. J. Riggs-Gelasco, J. C. Price, R. B. Guyer, J. H. Brehm, E. W. Barr, J. M. Bollinger Jr and C. Krebs, *J. Am. Chem. Soc*., 2004, **126**, 8108.

10 a) T. Borowski, A. Bassan and P. E. M. Siegbahn, *Chem. Eur. J*., 2004, **10**, 1031; b) S. P. de Visser, *Chem. Commun*., 2007, 171; c) A. V. Nemukhin, I. A. Topol, R. E. Cachau and S. K. Burt, *Theor. Chem. Acc*., 2006, **115**, 348; d) S. Sinnecker, N. Svensen, E. W. Barr, S. Ye, J. M. Bollinger Jr, F. Neese and C. Krebs, *J. Am. Chem. Soc*., 2007, **129**, 6168; e) A. Wójcik, M. Radoń and T. Borowski, *J. Phys. Chem. A*, 2016, **120**, 1261.

11 a) H. Liu, J. Llano and J. W. Gauld, *J. Phys. Chem. B*, 2009, **113**, 4887; b) D. Fang, R. L. Lord and G. A. Cisneros, *J. Phys. Chem. B*, 2013, **117**, 6410.

12 a) A. M. Maciejewska, J. Poznański, Z. Kaczmarska, B. Krowisz, J. Nieminuszczy, A. Polkowska-Nowakowska, E. Grzesiuk and J. T. Kuśmierek, *J. Biol. Chem*., 2013, **288**, 432; b) B. Bleijlevens, T. Shivarattan, K. S. van den Boom, A. de Haan, G. van der Zwan, P. J. Simpson and S. J. Matthews, *Biochemistry*, 2012, **51**, 3331; c) X. Pang, K. Han and Q. Cui, *J. Comp. Chem*., 2013, **34**, 1620; d) H. Torabifard and G. A. Cisneros, *Chem. Sci*., 2017, **8**, 6230.

13 a) L. Ji, A. S. Faponle, M. G. Quesne, M. A. Sainna, J. Zhang, A. Franke, D. Kumar, R. van Eldik, W. Liu and S. P. de Visser,

Chem. Eur. J., 2015, **21**, 9083; b) A. S. Faponle, M. G. Quesne and S. P. de Visser, *Chem. Eur. J*., 2016, **22**, 5478.

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

15 T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe and N. A. Baker, N*ucleic Acids Res*., 2007, **35**, 522.

16 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, S. Swaminathan and M. Karplus, *J. Comput. Chem*., 1983, **4**, 187.

17 a) A. D. Becke, *Phys. Rev. A: At., Mol., Opt. Phys*., 1988, **38**, 3098; b) J. P. Perdew, *Phys. Rev. B: Condens. Matter Mater. Phys*., 1986, **33**, 8822.

18 a) A. D. Becke, *J. Chem. Phys*., 1993, **98**, 5648; b) C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785.

19 W. Smith and T. R. Forester, *J. Mol. Graphics*, 1996, **14**, 136.

20 a) M. G. Quesne, T. Borowski and S. P. de Visser, *Chem. Eur. J*., 2016, **22**, 2562; b) D. Kumar, W. Thiel and S. P. de Visser, J. Am. Chem. Soc. 2011, 133, 3869; c) A. S. Faponle, F. P. Seebeck and S. P. de Visser, *J. Am. Chem. Soc*., 2017, **139**, 9259.

21 a) P. J. Hay and W. R. Wadt, *J. Chem. Phys*., 1985, **82**, 270; b) R. Ditchfield, W. J. Hehre and J. A. Pople, *J. Chem. Phys*., 1971, **54**, 724.

22 *Gaussian* 09, Revision D.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, Gaussian, Inc., C. T. Wallingford, 2009.

23 a) T. Borowski, A. Bassan and P. E. M. Siegbahn, *Biochemistry*, 2004, **43**, 12331; b) A. R. Diebold, C. D. Brown-Marshall, M. L. Neidig, J. M. Brownlee, G. R. Moran and E. I. Solomon, *J. Am. Chem. Soc*., 2011, **133**, 18148; c) S. P. de Visser, *Chem. Commun*., 2007, **2007**, 171; d) A. Wójcik, M. Radoń and T. Borowski, *J. Phys. Chem. A*, 2016, **120**, 1261; e) T. Borowski, A. Bassan and P. E. M. Siegbahn, *Chem. Eur. J*., 2004, **10**, 1031; f) I. A. Topol, A. V. Nemukhin, K. Salnikow, R. E. Cachau, Y. G. Abashkin, K. S. Kasprzak and S. K. Burt, *J. Phys. Chem. A*, 2006, **110**, 4223; g) S. Ye, C. Riplinger, A. Hansen, C. Krebs, J. M. Bollinger and F. Neese, *Chem. Eur. J*., 2012, **18**, 6555; h) M. R. A. Blomberg, T. Borowski, F. Himo, R.-Z. Liao and P. E. M. Siegbahn, *Chem. Rev*., 2014, **114**, 3601; i) S. Ghafoor, A. Mansha and S. P. de Visser, *J. Am. Chem. Soc*., 2019, **141**, 20278.

24 a) E. Godfrey, C. S. Porro and S. P. de Visser, *J. Phys. Chem. A*, 2008, **112**, 2464; b) H. Chen, W. Lai, J. Yao and S. Shaik, *J. Chem. Theory Comput*., 2011, **7**, 3049; c) N. Zeb, M. H. Rashid, M. Q. E. Mubarak, S. Ghafoor and S. P. de Visser, *J. Inorg. Biochem*., 2019, **198**, 110728; d) A. Timmins, M. Saint-André and S. P. de Visser, *J. Am. Chem. Soc*., 2017, **139**, 9855.

25 T. Borowski, A. Bassan and P. E. M. Siegbahn, *Chem. Eur. J*., 2004, **10**, 1031.

26 T. Borowski, A. Bassan and P. E. M. Siegbahn, *Inorg. Chem*., 2004, **43**, 3277.

27 a) H. Hirao, D. Kumar, L. Que Jr and S. Shaik, *J. Am. Chem. Soc.*, 2006, **128**, 8590; b) S. P. de Visser, *J. Am. Chem. Soc*., 2006, **128**, 15809.

28 S. P. de Visser, *The Chem. Record*, 2018, **18**, 1501.

29 J. C. Price, E. W. Barr, B. Tirupati, J. M. Bollinger Jr and C. Krebs, *Biochemistry*, 2003, **42**, 7497.

30 a) S. Álvarez-Barcia and J. Kästner, *J. Phys. Chem. B*, 2017, **121**, 5347; b) R. N. Manna, T. Malakar, B. Jana and A. Paul, *ACS Catal*., 2018, **8**, 10043; c) S. R. Iyer, V. D. Chaplin, M. J. Knapp and E. I. Solomon, *J. Am. Chem. Soc*., 2018, **140**, 11777; d) J. Xue, J. Lu and W. Lai, *Phys. Chem. Chem. Phys*., 2019, **21**, 9957.

31 a) D. Kumar, B. Karamzadeh, G. N. Sastry and S. P. de Visser, *J. Am. Chem. Soc*., 2010, **132**, 7656; b) D. Kumar, R. Latifi, S. Kumar, E. V. Rybak-Akimova, M. A. Sainna and S. P. de Visser, *Inorg. Chem*., 2013, **52**, 7968.

32 a) S. P. de Visser, F. Ogliaro, N. Harris and S. Shaik, *J. Am. Chem. Soc*., 2001, **123**, 3037; b) D. Kumar, B. Karamzadeh, G. N. Sastry and S. P. de Visser, *J. Am. Chem. Soc*., 2010, **132**, 7656; c) X.-X. Li, V. Postils, W. Sun, A. S. Faponle, M. Solà, Y. Wang, W. Nam and S. P. de Visser, *Chem. Eur. J*., 2017, **23**, 6406.

33 M. A. Sainna, S. Kumar, D. Kumar, S. Fornarini, M. E. Crestoni and S. P. de Visser, *Chem. Sci*., 2015, **6**, 1516.

34 a) R. Latifi, M. Bagherzadeh and S. P. de Visser, *Chem. Eur. J*., 2009, **15**, 6651; b) S. Kumar, A. S. Faponle, P. Barman, A. K. Vardhaman, C. V. Sastri, D. Kumar and S. P. de Visser, *J. Am. Chem. Soc*., 2014, **136**, 17102.