Decoding Regioselective Reaction Mechanism of the Gentisic Acid Catalyzed by Gentisate 1,2-Dioxygenase Enzyme

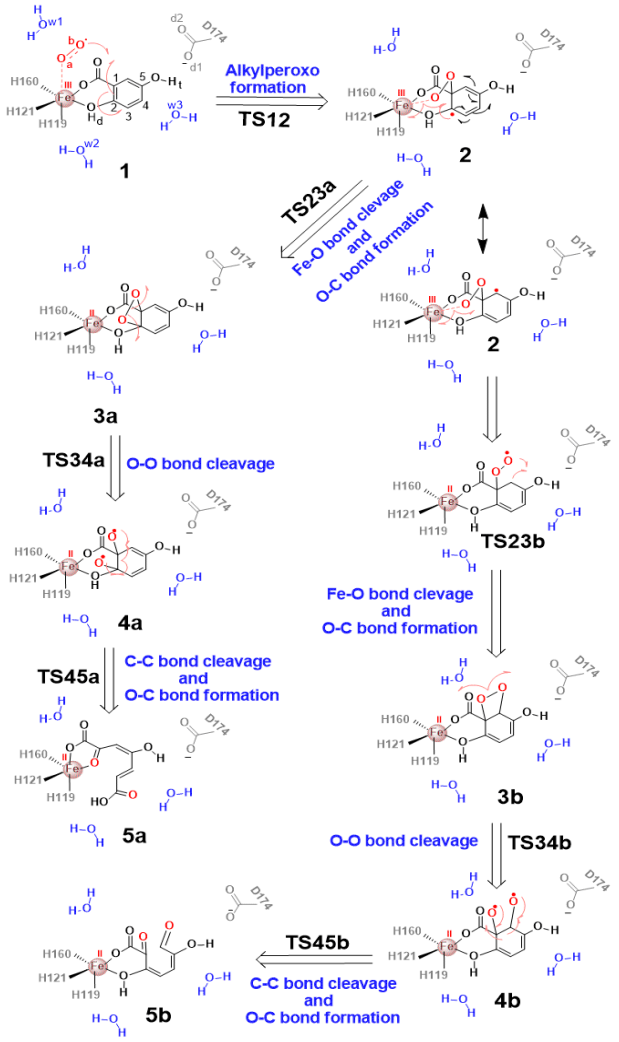
Rounak Nath,[#] Rabindra Nath Manna,[#] Ankan Paul*

School of Chemical Sciences, Indian Association for the Cultivation of Sciences, Kolkata, 700032, India.

KEYWORDS: Non-Heme Iron Dioxygenase, Metalloenzyme, GDO, QM/MM Calculation, Enzyme Catalysis Reactions.

ABSTRACT: Gentisate 1,2-dioxygenase (GDO), a ring-fission non-heme dioxygenase enzyme, displays a unique regioselective reaction of gentisic acid (GTQ) in the presence of molecular oxygen. GTQ is an important intermediate in the aerobic biodegradation pathways of recalcitrant polyaromatic hydrocarbons (PAHs) pollutants. Classical molecular dynamics simulations of wild-type GDO and its mutated variants (Asp174Glu and Asp174Ala) explored the presence of three active water molecules at the active site which plays pivotal roles in facilitating the oxidative cleavage of an aromatic C-C bond of the GTQ substrate. Three distinct reaction mechanisms using the QM/MM calculations decoded for the regioselective reaction of the GTQ catalyzed by GDO enzyme. The formation of the main product as a maleylpyruvate along with pathway A, which is the most favourable one. The first step for the conversion to an alkyl peroxo intermediate is a rate-determining step with an associated barrier of 21.4 kcal/mol at the uB3LYP-D3/def2-TZVP/OPLS level of theory on the quintet spin surface. Our study illustrates the crucial role of active water molecules in the stabilization of the O_2 molecule, the O-O, and C-C bond cleavage steps and additionally uncovered the important anchor role of the Asp174 in the enzymatic cycle. Essentially, our findings paved a new route in the mechanism of degradation processes of PAHs pollutants by dioxygenase enzymes, and provide molecular insights to design iron-containing biomimetic catalysts.

Non-heme iron dioxygenase proteins constitute an important class of O_2 activating oxidative enzymes with diverse applications spanning biosynthesis, pharmaceuticals, and pollutant remediation.[1-4] Gentisate 1,2-dioxygenase (GDO), a non-heme ring-fission dioxygenase enzyme has emerged as a wide-ranging effective catalyst for both intra and extradiol cleavage for the aerobic catabolism of the toxic and xenobiotic compounds.[5-6] GDO enzyme-containing micro-organisms utilize gentisic acid (GTQ), a vital intermediate for the biodegradation of polyaromatic hydrocarbons (PAHs) pollutants, as the sole source of carbon, metabolic energy, and convert them into less toxic compounds. Thus, biodegradation techniques may offer an excellent economical avenue for the removal of PAHs from the environment. Additionally, GTQ is used as an antioxidant excipient in the formation of pharmaceutical compounds.[7] The GDO enzyme is an unique enzyme that regioselectively cleaves an aromatic C-C bond in the GTQ substrate bearing the -OH and -COOH groups to yield 2-oxo-4-hydroxyhepta-3,5-dienedioic acid (maleylpyruvic acid) as a major product in the presence of molecular oxygen through an intradiol cleavage.[6] This type of cleavage process is mostly observed for the intra-diol dioxygenases.[8] Furthermore, 2,7-dioxo-3,6-dihydroxyhepta-3,5-dienoic acid is also formed as a minor product from the GTQ via extradiol cleavage which is commonly observed for the extradiol dioxygenases.[6,8] In contrast to other substrates (such as catechol, salicylic acid) of the dioxygenases experimental studies showed that the substrate for the GDO enzyme carries a hydroxyl group at the 5-position of the aromatic ring which happens to be a crucial requirement for facilitating aromatic cleavage.[9-11] Several GDO biomimetic model catalysts [5,12] had been prepared and characterized to provide crucial mechanistic insights of GDO catalyzed enzymatic reactions. The underlying cause dictating the regioselectively oxidative ring C-C bond cleavage of the GTQ substrate by the GDO enzyme is still not known. The role of binding site residues in the oxidative C-C ring cleavage reaction is yet to be understood. The knowledge of mechanistic intricacies at the atomic-level is a necessary requirement to develop efficient biomimetic catalysts. Recent theoretical studies of the different enzymatic reactions conjectured that active water molecule plays a pivotal role in enzymatic catalysis.[13-15] It is also interesting to probe, if there is any role of water molecules in effectuating the oxidative catalytic process. In this present venture, wild-type GDO along with two mutated variants (for instance, Asp174Ala, Asp174Glu) have been considered to elucidate the catalyzing role of 174th residue at the active site of the GDO enzyme, whose active role has been suggested in previous experimental investigations.[6] The present study unveils the mechanistic paradigms at play in this crucial enzymatic cycle and reveals the mechanistic origin of regioselective cleavage of the aromatic C-C bonds of the GTQ substrate.



Scheme 1. Proposed mechanisms for the oxidative aromatic C-C bond cleavage reaction of the GTQ substrate in the active site of GDO enzyme. All numberings of the displayed atoms and nomenclatures of the stationary states and transition states used throughout this study.

Computational Details

(a) Classical Molecular Dynamics Simulations

We have taken the initial geometry of the GTQ docked GDO enzyme from the protein data bank (PDB ID: 3NL1).[16] Eppinger et al. resolved the structure of GDO enzyme from the analysis of mutants of the salicylate 1,2-dioxygenase from *Pseudaminobacter salicylatoxidans*.[6] The crystal structure of GDO does not contain a dioxygen molecule. To prepare the enzyme model system, a dioxygen is manually added to the iron center. Since, the crystallographic water molecules are not around catalytic ferrous ion, hence, they removed. Moreover, His119, His121, His160 residues are assigned as singly protonated on their N_{D1} atoms using Gromacs 5.1 package.[17] Atomic charges of the catalytic Fe^{2+} ion, dioxygen, and GTQ are calculated using Gaussian 09 program[18] at the B3LYP/6-31++G(d, p) (C,H,N,O), LANL2DZ(Fe) level of theory.[19-22] The classical molecular dynamics (MD) simulations were performed to equilibrated the GTQ bound enzyme complex. The OPLS-AA force field [23] was used for the protein for all classical MD simulations as well as hybrid quantum mechanics/molecular mechanics (QM/MM) calculations. The three model systems were prepared in this present investigation, such as, wild type GDO–GTQ substrate, two mutated variants (Asp174Ala–GTQ and Asp174Glu-GTQ). In the case of a mutated system, the aspartic acid (D) residue in the 174th position of the GDO is mutated with either alanine (A) or glutamic acid (E). Then, the protein-ligand complex is solvated in a cubic box using TIP3P water [24] molecules with a 10 Å radius buffer zone of water molecules around the enzyme complex in each direction. After solvation, the whole protein-ligand complex is neutralized by adding ten sodium ions. The long-range cut-off for nonbonded interactions is taken as 10 Å for all the above minimization steps, and also for all subsequent MD simulations. Then all the systems are minimized with 50000 steps in steepest descent method. Then, each system is simulated at 300 K for 1 ns using the NVT ensemble followed by another 1 ns simulation using NPT ensemble. The temperature is fixed at 300 K with the V-rescale algorithm[25] and Parrinello−Rahman barostat [26] is used to maintain the pressure. The particle mesh Ewald summation method [27] is used to measure electrostatic interactions. The integration step for all MD simulations is 1 fs. Subsequently, 50 ns of production run simulations are performed for all the three aforementioned systems using NPT ensemble. During the energy minimization and classical MD simulations, the coordinates of iron ligating ligands were kept frozen. Monitoring of root means square deviation (RMSD), radial distribution function, the radius of gyration, and trajectory analysis were performed using GROMACS 5.1 toolkit.[17] Visual Molecular Dynamics (VMD) 1.9.2 was used in order to analyze the trajectories and explore the atomic-level insights.[28]

(b) QM/MM calculation

Several configurations of the GTQ substrate bound wild-type GDO enzyme complex, obtained from our classical MD simulations are used as initial structures for the hybrid QM/MM optimization calculations. We have prepared three reaction mechanisms. The details of the mechanism will be discussed in the results and discussion section. In all the three mechanisms, the number of QM atoms are the same (70 atoms). The QM region consists of Fe^{2+} ion, dioxygen, GTQ, three water molecules and side chains of His119, His121, His160, and Asp174. Four link atoms are placed between the C_γ and C_β atoms of the His119, His121, His160, and Asp174 residues to saturate the valence of the QM subsystem. The rest of the protein, water molecules, and ten Na^+ ions are part of the MM region (where, OPLS-AA and TIP3P potential were used for enzyme and water, respectively). Based on the reported studies on non heme iron dioxygenases [13, 29-31] the QM region is described by the B3LYP functional along with Pople's 6-31+G(d) basis set [32] for atoms like C, N, O, H and for iron the Los Alamos pseudopotential (LANL2DZ) basis set [33-34] consisting of an effective core potential (ECP) was used. The QM/MM optimization computations were performed using the micro-macro iteration scheme as implemented in the f-DYNAMO library[35] and its interface with Gaussian 09 which had been previously used to investigate several complex enzymatic reactions.[36-38] The positions of all atoms beyond 20 Å from the substrate were fixed and a cut off radii switch between 14.5 Å and 16

Å was used for all non-bonded interactions. The transition state (TS) structures were verified by characterizing whether it has a unique imaginary frequency representing the desired reaction coordinate (Table S2). Finally, the energies of all the optimized structures are corrected by single-point calculations using dispersion included uB3LYP-D3/def2-TVZP/OPLS and uB3LYP-D3/6-311+G(d) (C, H, O, N), LANL2TZ(Fe)/OPLS level of theories.[39-40]

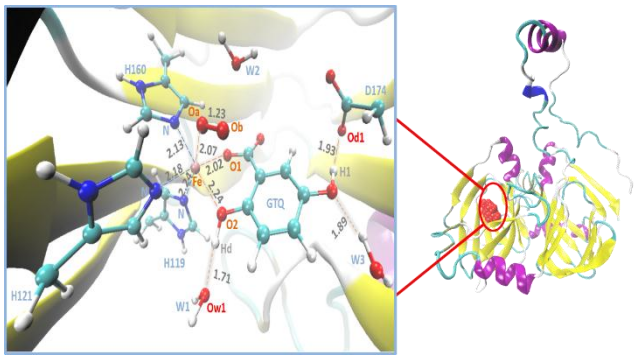


Figure 1. The structure of the GDO-GTQ enzyme complex is shown. In the active site of optimized structure 1, the catalytic Fe^{2+} ion is surrounded by His119, His121, His160, GTQ, molecular dioxygen and three water molecules (for instance, W1, W2 and W3) The Asp174 interacts with the H_1 atom of the OH motif of the GTQ substrate. Herein, "H" and "D" are denote as the single letter code of the histidine and aspartic amino acid.

Results and Discussion

We initiated our investigations by performing classical MD simulations of wild type GDO and two mutated variants, such as, Asp174Ala and Asp174Glu model systems. To check whether the simulations lead to a stable enzyme-substrate complex, we have monitored the root-mean-square-deviations (RMSD) of the protein backbone atoms relative to a starting crystal structure (Figure S1). Careful inspection from RMSF plots (Figure S2) of the systems revealed that residues which range from number 50 to 350 are less fluctuated for all the three model systems, whereas, terminal residues were found to be more fluctuated for the Asp174Glu mutated system. The radius of gyration (Rg) of protein was steady throughout 50 ns of simulations for all the aforementioned systems (Figure S3). This implies that the protein complex is stable, equilibrated, and stayed in natively compact form throughout the simulations. The MD simulations have revealed that catalytic ferrous (Fe^{2+}) ion has formed an octahedral coordination sphere in the binding site of the GDO. More precisely, the catalytic Fe^{2+} ion coordinates with three nitrogen atoms (N_{E2}) of the histidine residues (for instance, His119, His121 and His160) along with two coordination with oxygen atoms of carboxylate and hydroxyl group (O_1 and O_2 respectively) of the GTQ substrate and also forms one non-covalent bond with the O_a oxygen atom of molecular oxygen (O_2), respectively (Figure 1). We have considered the binding of the GTQ substrate in the neutral form, because, we did not find any potential catalytic residues to abstract the H_d atom of the GTQ. Molecular dynamics simulations revealed the presence of strong hydrogen bond interaction between the H_d atom of the GTQ substrate and the oxygen atom (O_{w1}) of the water molecule, W1 (Figure S4). A model was prepared to check any role of the W1 water molecule for the de-protonation process of the GTQ substrate. However, we did not obtain any optimized stationary and transition state structures for the deprotonation process of the H_d atom of the GTQ substrate by the O_{w1} atom of the W1 water. Hence, our calculations suggested that the W1 water does not act as a base rather it provides stabilization to the GTQ substrate through hydrogen bond which helps the substrate to remain in an optimum position to facilitate the reaction. Additionally, the coordination number plot (Figure S5) displays the presence of two water molecules in the range of 3.5 to 4.3 Å of the catalytic Fe^{2+} ion. Moreover, there is one water molecule (W3) around para hydroxyl (OH) group of the GTQ within ~3.0 Å as detected (Figure S6). Recent theoretical studies of the different enzymatic reactions conjectured that water molecule plays a pivotal role in enzymatic catalysis.[13-15] In this present venture, it will be noteworthy to investigate whether water molecules in the vicinity of the catalytic sites play any vital role in the oxidative C-C bond cleavage reaction of the GTQ. Additionally, the substrate is unique in the sense that it contains an electron-donating group, such as OH moiety in the 5th position. To investigate further, we have analyzed the inter-atomic distances between the H_1 atom of the para OH moiety of the GTQ and either the O_{d1} atom of the Asp174 of the wild type GDO enzyme or O_{e1} atom of Glu174 residue of mutated Asp174Glu system (Figure 2). In the wild type GDO system, the inter-atomic distances of the H_1(GTQ)-O_{d1}(Asp174) decrease to ~2.0 Å with the progression of the MD simulations which implies that a strong hydrogen bond is formed between H_1(GTQ) and the O_{d1} atom of the Asp174. The distance of H_1(GTQ)-O_{e1}(Glu174) in the mutated Asp174Glu system remains around ~5.0 Å which is a much greater distance to form an effective hydrogen bond. In the Asp174Ala system, where instead of the aspartate in the 174th residue, there is an alanine residue, which is unable to form a hydrogen bond with the substrate. Moreover, the hydrogen bonding between the para positioned OH of the GTQ and O_{d1} atom of Asp174 occurs in the wild-type system that helps the substrate to remain in optimal orientation for facilitating the oxidative cleavage reaction of the GTQ substrate. Thus, the findings from our MD simulations display excellent agreement with the experimental observations, where Asp174 was shown to be essential for the enzymatic activity of GDO based on mutagenesis studies.[6]

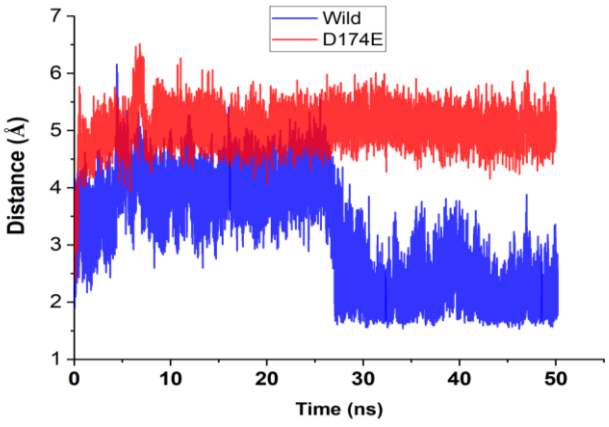
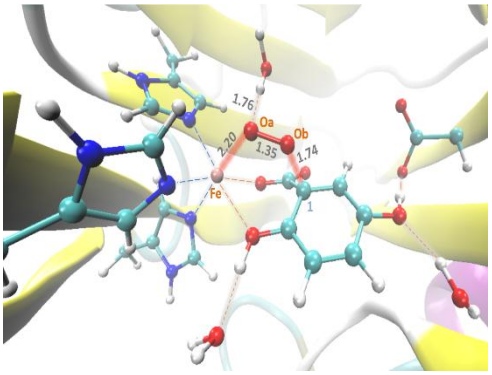
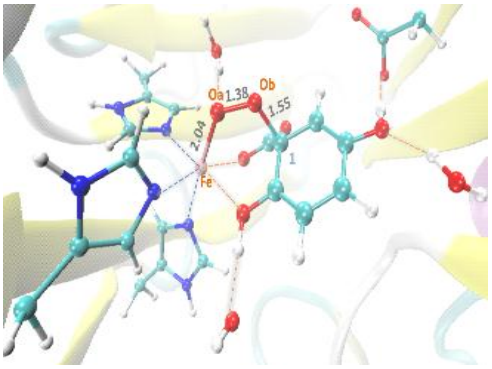


Figure 2. Time evolution of inter-atomic bond distances of H_1(GTQ)-O_{d1}(Asp174) and H_1(GTQ)-O_{e1}(Glu174) are represented as blue and red lines respectively during the classical MD simulations.

Our MD simulations have evinced the presence of a water molecule (W2) near the dioxygen molecule in the active site of the GDO, which plays an important role in the O_2 activation process during the reaction (Figure S7). Similar mode of activation can be found in other mononuclear non-heme iron dioxygenases.[13-15] On the contrary, Roy et al. did not found any water molecule for O_2 activation in their salicylate dioxygenase system.[8,11] Instead of a water molecule, they had found that the Arg127 residue played an important role in the process of activation of the O_2 molecule. However, our MD simulations revealed that the Arg127 residue is far away from the O_2 molecule in the GDO active site (Figure S8). Several reaction pathways were considered (Schemes 1 and 2) to probe the oxidative C-C aromatic bond cleavage reaction of the GTQ. The molecular di-oxygen binds to the sixth coordination site of catalytic ferrous ion (Fe^{2+}) in an end-on fashion to form an enzyme-substrate-O_2 complex (1). The coupling between a quintet Fe^{2+} ion and a stable triplet O_2 gives rise to three possible spin states, namely triplet, quintet, and septet, respectively. The triplet spin state of the complex (1) is 16.1 kcal/mol higher in energy than the quintet state of (1) at the uB3LYP-D3/def2-TZVP/OPLS level of theory (Figure S13). On the other hand, the septet spin state of (1) is 7.3 kcal/mol lower than the quintet spin state of complex (1). It is noteworthy to mention that the overestimation of the stability of septet spin complexes of Fe^{2+} compared to the lower spin species are often observed at the B3LYP functional due to the presence of high percentage of exact exchange.[41] We also reckoned that the quintet spin state complex becomes stable than the high spin septet state at the BP86/def2-TZVP/OPLS level (Figure S26), as has been found earlier in other allied systems.[42-43] The barrier for TS12 on the quintet spin surface (21.4 kcal/mol) is lower than the septet spin surface (26.4 kcal/mol). Thus, subsequent steps for this reaction proceed on the quintet spin surface. Similar trends pertaining to spin state ordering were also predicted by Roy et al. in their system.[8] The interatomic distances of Fe-O_1 and Fe-O_2 in the optimized structure (1) on the quintet spin state are 2.02 and 2.24 Å, respectively which imply that the substrate is strongly bound with the catalytic iron (Figure 1). The interatomic distance between the catalytic iron and nitrogen atoms of three histidine residues (for instance, His119, His121, and His160) are in the range of 2.12-2.26 Å which are similar to the corresponding values provided by the Eppinger et al. for the SDO, HPCD enzyme.[6] The optimized structure of the complex (1) on the quintet spin surface displayed that the Fe-O_a bond distance is 2.07 Å which is longer than the Fe-O(O_2) bond length (1.98 Å) in that of the SDO enzyme.[11] Although, the obtained Fe-O(O_2) bond distance in the GDO system follows a similar trend with other non-heme iron dioxygenases, for instance, HPCD,[29] CDO,[44] and DAD[13]. The O_a-O_b bond length of the complex 1 (1.23 Å) is slightly elongated compared to that in free molecular oxygen (O_2) (1.20 Å) depicting super-oxo character of the bound oxygen.



TS12



2

Figure 3. Snapshot pictures of the optimized structures of the TS12 and 2 on the quintet spin state with the key inter-atomic distances (Å) are presented.

We have considered the quintet spin surface to compute the further steps of the oxidative cleavage of the aromatic C-C bond of the GTQ substrate. As stated earlier that the Asp174 residue displayed a strong hydrogen bond interaction with the para-hydroxyl group of the GTQ substrate. Hence, we attempted several times to model the hydrogen (H_1) abstraction process by the Asp174 residue. However, we did not get any transition state or stationary state for the abstraction of proton from the para-hydroxyl group of the GTQ substrate to the Asp174 residue. Our findings suggested that the Asp174 residue assists in binding GTQ substrate, rather than participating in the deprotonation process of the para-hydroxyl group of the substrate. The formation of Fe(III) alkylperoxo intermediate (2) is an endothermic process with the associated activation energy barrier of 21.4 kcal/mol at the uB3LYP-D3/def2-TZVP/OPLS level of theory which happens to be the rate-determining step for both the pathway A and B (Figure 4). We have also conducted our calculation at the uB3LYP-D3/6-311+G(d) (C, H, O, N), LANL2TZ(Fe)/OPLS level of theory (Figure S11 and S12) and the obtained results follow similar trends as the reported one. Herein, the obtained barrier for the first step from our calculation is similar in the range with the barrier found by Roy et al.[8,11] for the SDO enzyme catalysed reaction. In our optimized TS12 structure, the bond distances of O_a-O_b and O_b-C_1 are 1.35 and 1.74 Å respectively (Figure 3), which indicate that O_b-C_1 bond begins to

form and O_a-O_b bond of the O_2 molecule is elongated. Interestingly, we have noticed that the O_a atom of O_2 in TS12 exhibits a strong hydrogen bond (1.76 Å) with one of the active water molecules (W2). In our study, this particular water molecule plays pivotal role in stabilizing molecular oxygen during the reaction.[13-15] In the optimized alkylperoxo intermediate (2), the key interatomic distance of Fe-O_a decreases to 2.04 Å and the O_a-O_b bond elongates to 1.38 Å (Figure 3). Moreover, a new O_b-C_1 bond (1.55 Å) in the intermediate (2) is formed. The formation of alkylperoxo intermediate was also proposed by Paine and his co-workers and suggested by Chen et al. for the GDO from S. pomeroyi organism.[3] The alkylperoxo intermediate (2) comprises a radical which can either form 3a intermediate through the transition state TS23a along the path A to the main product (maleylpyruvate) or 3b intermediate via transition state TS23b along the path B leading to the side product.

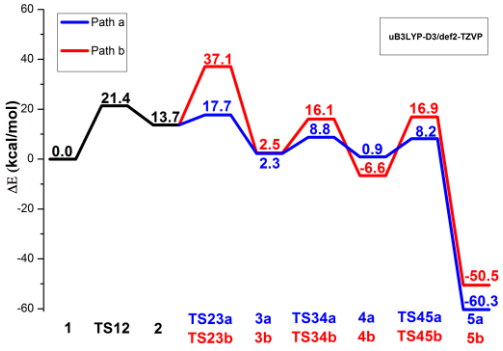
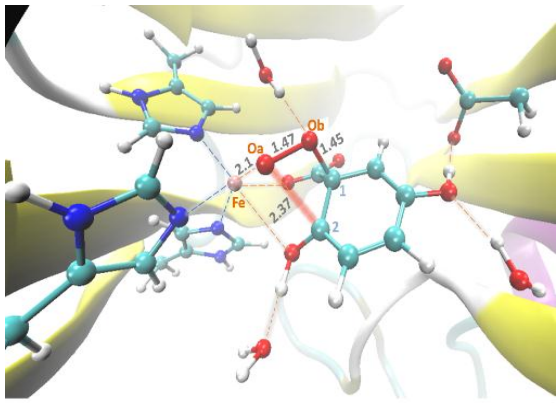
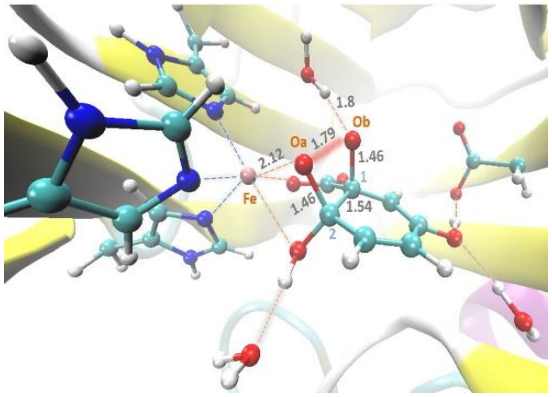


Figure 4. The potential energy profile for the oxidative cleavage of the C-C aromatic bond of the GTQ substrate in the active site of the GDO enzyme following pathway A and B at the uB3LYP-D3/def2-TZVP/OPLS level of theory.

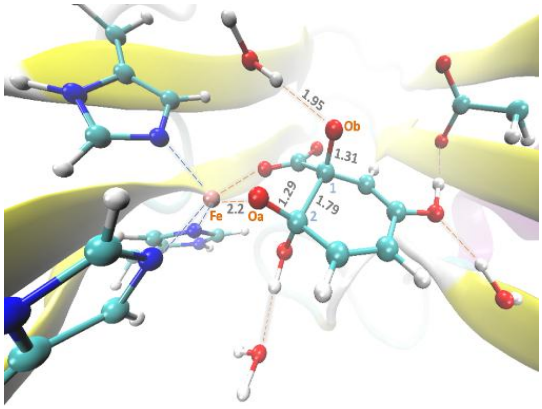
Both 3a and 3b intermediates are dioxetane like compounds with different orientations of the four-membered ring. Eppinger et al. mentioned that the formation of the side products was easily explained from the dioxetane intermediate.[6] However, no detailed mechanistic picture was furnished. Intriguingly, their observation prompted us to further probe the fate of the dioxetane intermediate. The activation barrier for the formation of 3a intermediate is 17.7 kcal/mol through the transition state TS23a. While the barrier is much higher (37.1 kcal/mol) for the formation of 3b intermediate compared to the 3a intermediate (Figure 4). In the TS23a structure, the bond distances of O_a-C_2 and O_a-O_b are 2.37 and 1.47 Å, respectively (Figure 5). The O_a-C_2 interatomic distance shortens indicating the formation of a partial bond. The Fe-O_a bond distance extends to 2.10 Å. Additionally, the W2 water makes a strong hydrogen bond with the O_b atom to stabilize the TS23a. The formation of 3a intermediate is endothermic in nature with the reaction energy of 2.3 kcal/mol with respect tothe reactants (Figure 4). The O_b-C_6 and O_a-O_b bond distances are 2.36 and 1.40 Å respectively at the TS23b structure (Figure 6). The main difference between the TS23a and TS23b is the inter-atomic interaction between the Fe and the O_a atoms is absent in the TS23b. The O_a-O_b bond distance in the 3a and 3b is very similar to each other (1.49 Å). The bond distances of C_1-C_2 (1.55 Å), O_a-C_2 (1.47 Å) and O_b-C_1 (1.50 Å) at the intermediate 3a (Figure S16) clearly showed that the formation of a four-membered ring and the same feature is observed for the intermediate 3b, where we get the indication of formation of a four-membered ring by analyzing the bond distances of C_1-C_6 (1.54 Å), O_a-C_6 (1.48 Å) and O_b-C_1 (1.48 Å), respectively (Figure S17).



TS23a



TS34a



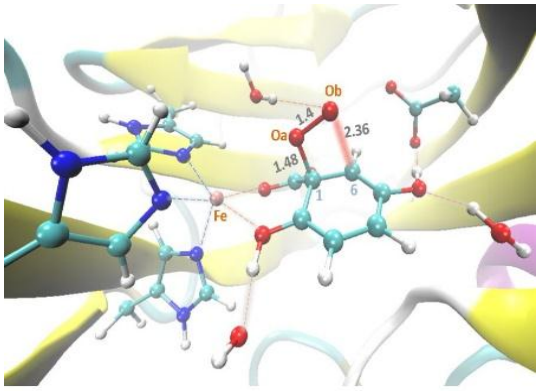
TS45a

Figure 5. Optimized structures of TS23a, TS34a, and TS45a on the quintet spin state along Path A with the key inter-atomic distances (Å) are presented.

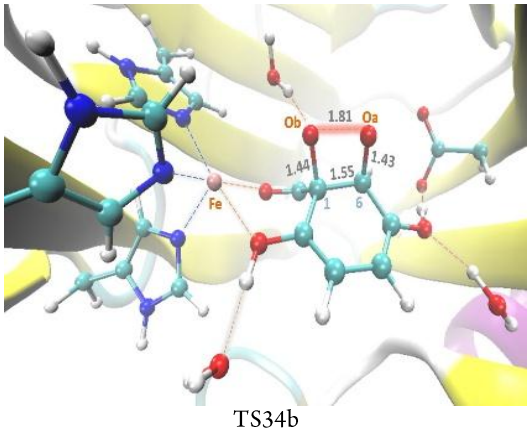
The formation of dioxetane intermediate (3b) is 2.5 kcal/mol uphill on the potential energy landscape (Figure 4) of the quintet surface and involves a new bond formation between the radical (C_6) atom of the GTQ and oxygen (O_a) atom. In the next

step, the O-O bond in the dioxetane (such as 3a and 3b) intermediates homolytically cleaved through their corresponding TS34a and TS34b transition states to form their corresponding intermediates 4a and 4b with the associated energy barriers of 8.8 and 16.1 kcal/mol, respectively (Figure 4). The barrier for the homolytic O_a-O_b bond cleavage event along the path B is 7.3 kcal/mol higher compared to that of path A. Elongated O_a-O_b bonds (TS34a: 1.79 Å and TS34b: 1.81 Å) are indicative of the desired bond scission happening in both of the transition state structures (Figure 5 and 6). For both TS34a and TS34b, the W2 water forms strong hydrogen bonds with the O_a atom, which provide further stabilization for those transition states. Moreover, the O_a-C_2 and O_b-C_1 bonds (for instance, 1.46 Å and 1.46 Å) in TS34a and O_b-C_1 and O_a-C_6 bonds (1.44 Å and 1.43 Å) in TS34b (Figure 5 and 6) are decreased leading to our next intermediates 4a along path A and 4b along Path B with the reaction energies of 0.9 and -6.6 kcal/mol respectively (Figure 4). The formation of 4a intermediate is more favourable than the formation of 4b intermediate in terms of corresponding TS barriers. Both the intermediates 4a and 4b are di-radical in nature and the major bond lengths of the O_a-C_6, O_b-C_1, and C_1-C_6 are 1.41, 1.40, 1.54 Å respectively in the 4b structure (Figure S20). Interestingly, iron in the 4a intermediate structure becomes five coordinated with the exclusion of O_a atom from its coordination site. The bond distance of O_a-C_2, O_b-C_1 and C_1-C_2 are 1.40, 1.40, 1.54 Å respectively in the 4a structure (Figure S19). Intra-molecular hydrogen bonding takes place between the W2 water and the O_a radical in both 4a and 4b intermediates that suggest the importance of the water molecule in the QM region. The intermediates 4a and 4b are followed by the homolytic cleavage of C-C bonds and O-C double-bond formations leading to the 5a and 5b, final products of pathway A and B through the TS45a and TS45b respectively. Indeed our findings depict that the main product 5a (6-carboxy-4-hydroxy-2-oxohexa-3,5-dienoate) is energetically more favourable compared to the side product 5b (3,6-dihydroxy-2,7-dioxoheptа-3,5-dienoate) (see Figure 4). In the TS45a, the C_1-C_2 bond elongated to 1.79 Å (Figure 5) that describes the homolytic bond breaking. Simultaneously, the O-C bond shrinks to 1.29 Å leading to the formation of the main product (5a). In the optimized geometry of 5a, O_b-C_1 and O_a-C_2 bond distances are 1.24 and 1.22 Å suggesting the formation of O-C double bond with desired complete cleavage of C_1-C_2 bond. The C_5-C_6 and C_3-C_4 bond lengths are 1.36 Å and 1.34 Å that indicate the formation of C-C double bonds in alternate positions (Figure S22). The transition state TS45b is 16.9 kcal/mol uphill in the potential energy landscape (Figure 4) and the larger barrier height of TS45b compared to that of TS45a (8.2 kcal/mol) suggests the pathway A is favourable over the pathway B. The elongated C_1-C_6 bond (1.86 Å) and reduced bond lengths of O_a-C_6 and O_b-C_1 (1.31 and 1.3 Å respectively) refer homolytic breaking of corresponding C-C bond along with the commencement of O-C double bond formations (Figure 6). In the side product, the bond lengths of O_a-C_6, O_b-C_1, C_2-C_3 and C_4-C_5 are 1.22, 1.23, 1.36, and 1.34 Å respectively (Figure S23) clearly indicating the formation of double bonds between the respective atoms with the desired cleavage of C_1-C_6 bond. The active site residues His119, His120, His 161 and Asp174 are highly conserved among the several bacterial species as revealed from our MSA analysis suggesting their crucial roles in such catalysis (Figure S29).

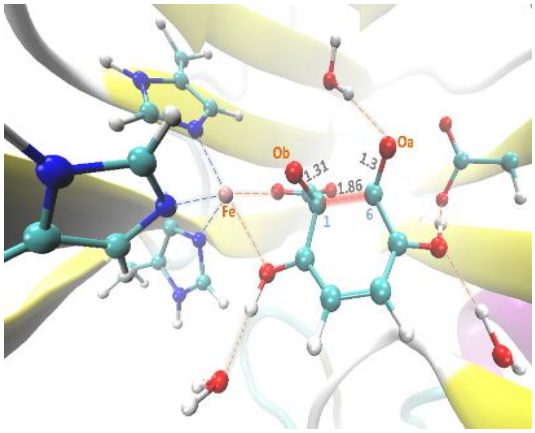
Since SDO enzyme is homologous to GDO,[6] we have also followed the footsteps of Roy et al. for the catalytic mechanism of the SDO enzyme to design our 3rd pathway (such as pathway C in scheme 2) for the GDO enzyme leading towards the main product only. Here, we have performed a comparative study to check the feasibility of our proposed pathway with the SDO enzyme by Roy et al.[11] although both substrates are different.



TS23b

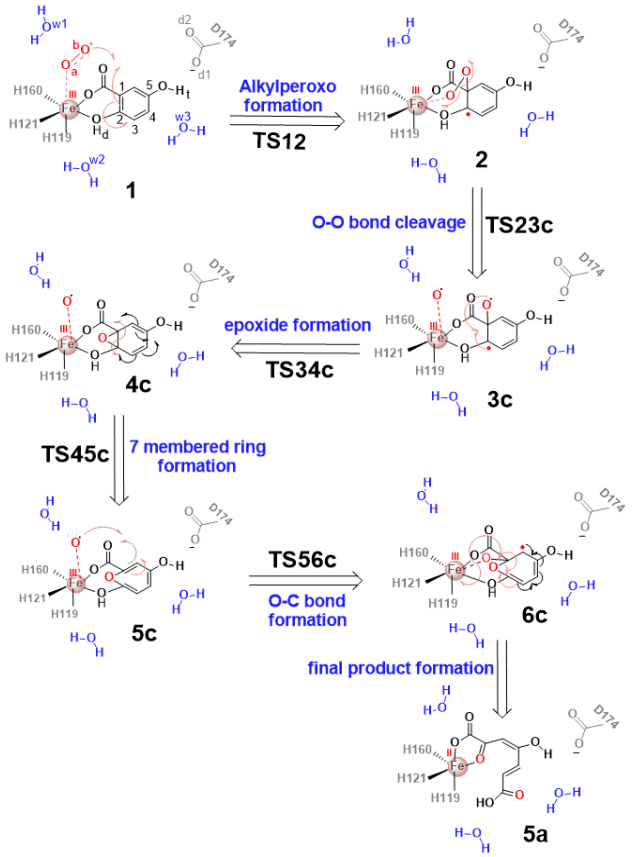


TS34b



TS45b

Figure 6. Optimized structures of the TS23b, TS34b and TS45b on the quintet spin state along Path B with the key inter-atomic distances (Å) are displayed.



Scheme 2. Proposed mechanisms for the oxidative aromatic C-C bond cleavage reaction of the GTQ substrate in the active site of the GDO enzyme via seven membered ring intermediate.

The formation of the alkylperoxo intermediate (2) along the pathway C is identical to earlier aforementioned pathways as shown in scheme 1. The distinction first appears when the dioxygen (O_a-O_b) bond is cleaved homolytically to form two oxygen radicals attached with Fe and C_1 centers (scheme 2). Our calculation has identified that in the transition state TS23c the key interatomic distances of Fe-O_a decreases to 1.71 Å and the O_a-O_b bond elongates to 1.89 Å along with the shortening of the O_b-C_1 bond to 1.42 Å (Figure 7) and these distances are very similar to the corresponding values provided by Roy et al. for the SDO enzyme.[11] The energy of TS23c is 24.1 kcal/mol higher (Figure 9) compared to the starting substrate (1). This suggests that our proposed pathway A is favourable where the barrier height of transition state after alkylperoxo intermediate (2) via TS23a is 17.7 kcal/mol (Figure 4) over pathway C. The unprotonated geminal dihydroxy intermediate (3c) is stabilized by strong intermolecular hydrogen bonding between the W2 water and the O_b atom of dioxygen moiety. The key bond distances of Fe-O_a (1.61 Å), O_b-C_1 (1.39 Å) (Figure S18) again closely resembles the corresponding values provided by Roy et al. for SDO enzyme.[11] Furthermore, the formation of 3c intermediate is endothermic like 3a and 3b. However, the extent of endothermicity is much higher in case of 3c (13 kcal/mol) compared to 3a and 3b (2.3 and 2.5 kcal/mol respectively). These results have demonstrated that our proposed pathway is more favorable over the pathway C. Intermediate 4c is an epoxide like structure formed through TS34c which is 19.8 kcal/mol uphill in the potential energy profile. The bond distances of O_b-C_1 and C_1-C_2 in TS34c (1.40 Å and 1.51 Å respectively) (Figure 7) are similar to the corresponding values provided by Roy et al. [11] for SDO enzyme although the O_2-C_2 bond length for our system elongates to 2.21 Å.

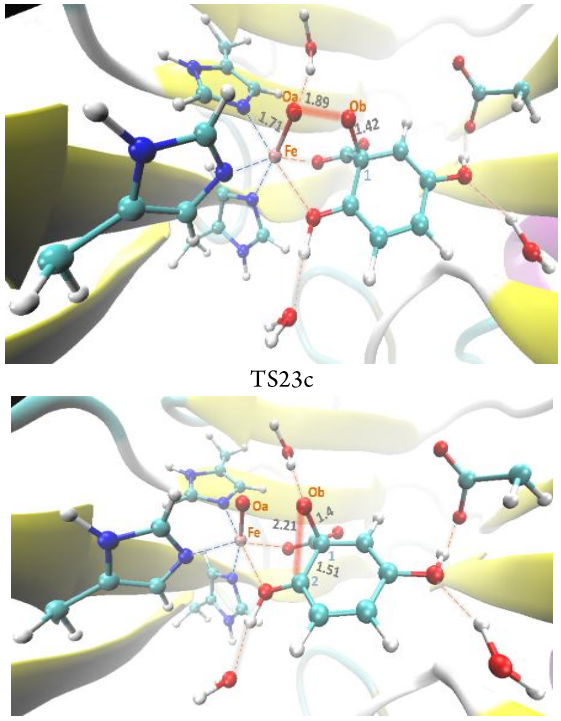
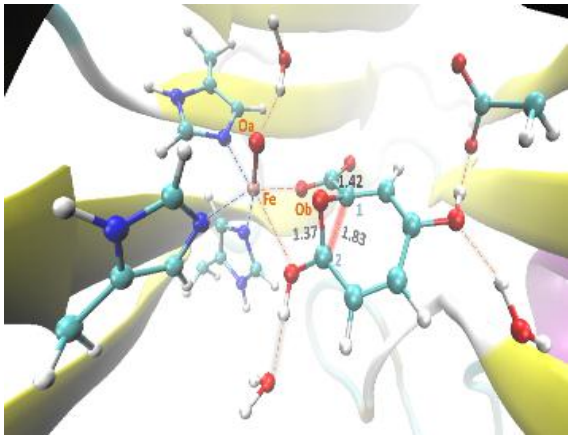


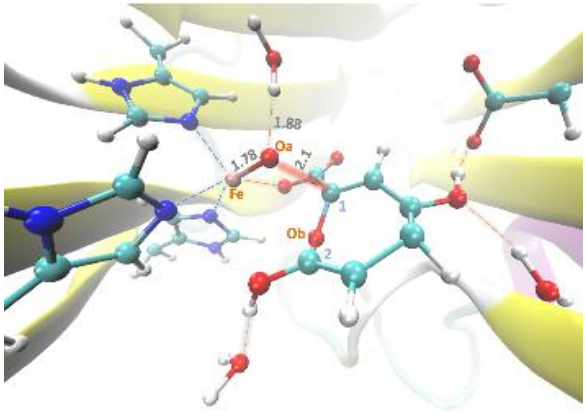
Figure 7. Optimized structures of the TS23c and TS34c on the quintet spin state along the Path C with the key inter-atomic distances (Å) are displayed.

The epoxide intermediate (4c) is stabilized by -11.9 kcal/mol (Figure 9) mainly due to the fulfilment of valency of the O_b atom with a new O_b-C_2 bond (1.41 Å). The intermolecular hydrogen-bonding with the W2 water has shifted to an iron, which coordinates with the O_a atom to attain the required valency in intermediate 4c. The other bond lengths of O_b-C_1 (1.46 Å) and C_1-C_2 (1.51 Å) along with O_b-C_2 indicate the formation of a three-membered epoxide ring and elongation of Fe-O_2 bond to 2.57 Å has happened (Figure S21). Subsequently, a C-C homolytic cleavage via TS45c occurs resulting a highly stabilized (-18.5 kcal/mol) seven-membered ring (5c) (Figure 9). TS45c is -8.8 kcal/mol stabilized compared to the starting substrate (1) and with the elongation of the C_1-C_2 bond (1.83 Å) both O_b-C_1 (1.42 Å) and O_b-C_2 (1.37 Å) (Figure 8) bonds shorten. The Fe-O_2 bond has further elongated to 2.83 Å to form penta-coordination sphere of the catalytic iron centre. As the O_2 centre for our system is not deprotonated, a neutral seven-membered lactone intermediate (5c) is formed unlike the mono radical intermediate depicted by Roy et al.[11] The unique alternate bond lengths of C_2-C_3, C_4-C_5, and C_6-C_1 (1.37, 1.37 and 1.36 Å respectively) (Figure S24) depict the formation of conjugated double bonds in 5c and elongated Fe-O_2 bond length (3.17 Å) indicates the iron centre to be penta coordinated. In the seven membered ring (5c), the O_a radical is attached to the iron atom (Fe-O_a ~1.62 Å). In the next step, the O_a radical forms a new bond with

the C_1 atom of GTQ through TS56c and the bond length of O_a-C_1 (2.1 Å) along with a stretched Fe-O_a bond (1.78 Å) in TS56c (Figure 8) infer the initiation of formation of the new O_a-C_1 bond. The barrier of this step is moderately high (19.6 kcal/mol) compared to the previous intermediate 5c (Figure 9) due to the rapture of the C-C double bond and subsequent conversion to a radical species from a stable neutral seven membered ring. TS56c eventually converts to the next intermediate 6c towards the final major product (5a). This radical intermediate 6c is stabilized by -27.8 kcal/mol (Figure 9) and it can subsequently rearrange to deliver the final product (5a). This rearrangement is essentially a barrierless process. In the intermediate 6c, the bond lengths of Fe-O_a and O_a-C_1 are 2.05 Å and 1.40 Å, respectively (Figure S25). Like the other intermediates and transition states in TS56c and 6c, the W2 water formed inter-molecular hydrogen bonding with the O_a atom to provide stabilization to the TS56c and the radical intermediate (6c).



TS45c



TS56c

Figure 8. Optimized structures of the TS45c and TS56c on the quintet spin state along Path C with the key inter-atomic distances (Å) are displayed.

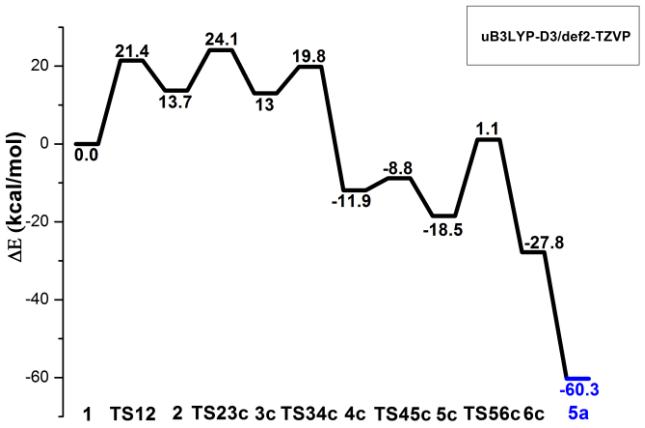


Figure 9. The potential energy profile for the oxidative cleavage of C-C aromatic bond of the GTQ substrate in the active site of the GDO enzyme following pathway C at the uB3LYP-D3/def2-TZVP/OPLS level of theory.

Conclusion

In this present venture, classical MD simulations along with the hybrid QM/MM calculations have been carried out to probe the oxidative cleavage of the aromatic C-C bond of the gentisic acid (GTQ) catalyzed by GDO, a non-heme dioxygenase enzyme. The equilibrium MD simulation ascertained that there are three essential water molecules present in the active site of GDO. Those water molecules stabilize the O_2 molecule and the GTQ substrate during the degradation reaction. Asp174 residue activates the GTQ substrate with strong hydrogen bond interaction, which is situated between the para-hydroxy group of the GTQ and carboxylate group of the Asp174. Moreover, this hydrogen bonding interaction occurs predominantly in the only wild-type GDO enzyme, which plays an important role to stabilize the substrate during the reaction. Herein, three types of degradation mechanisms (such as Pathway A, B, and C, respectively) of the GTQ substrate have been thoroughly discussed. Our calculation evinced that the formation of alkylperoxo intermediate (2) is the rate-determining step (21.4 kcal/mol) for the oxidative cleavage of the C-C aromatic bond of the GTQ catalyzed by GDO enzyme for the pathway A and B. Both pathway A and B head towards the main and side products respectively through dioxetane like intermediates. While, the Path C leads to the main product through a seven-membered intermediate. Interestingly, we have found that the activation energy barrier for the TS23a along Path A is significantly lower compared to the barrier for the TS23b along Path B and TS23c along Path C. Our QM/MM study indeed displayd that the GDO enzyme regio-selectively cleaves the aromatic C-C bond of GTQ forming two products. Path A is more favourable pathway for the formation of the main product due to the lower activation energy barrier at the 2nd step. Essentially, our findings evinced an alternative distinct mechanistic perspective for the regioselective cleavage of the aromatic C-C bond of the GTQ catalyzed by GDO enzyme incorporating two molecular oxygen atoms.

ASSOCIATED CONTENT

Supporting Information is available free of charge on the ACS Publications Web site. The computational details, discussion of QM/MM models, plots, structures, and Cartesian coordinates are included in the supporting information

AUTHOR INFORMATION

Corresponding Author

* A. Paul Email: rcap@iacs.res.in

Author Contributions

‡ R. N and R. N. M contributed equally, A. P. designed research. R. N and R. N. M performed research. R. N. M, R. N and A. P analyzed, discussed, wrote the manuscript.

Funding Sources

R.N is thankful to CSIR, India for providing fellowship (Award No. 09/080(1149)/2020-EMR-I). R. N. M is thankful to Indian Association for the Cultivation of Science for providing fellowship. We also thank the SERB-DST, India (Grant No. CRG/2019/006685) for financial support.

ACKNOWLEDGMENT

The authors gratefully acknowledge the HPC and CRAY cluster of Indian Association for the Cultivation of Science, Kolkata for providing computational resources. We also thank Dr. Boyli Ghosh for helpful discussions.

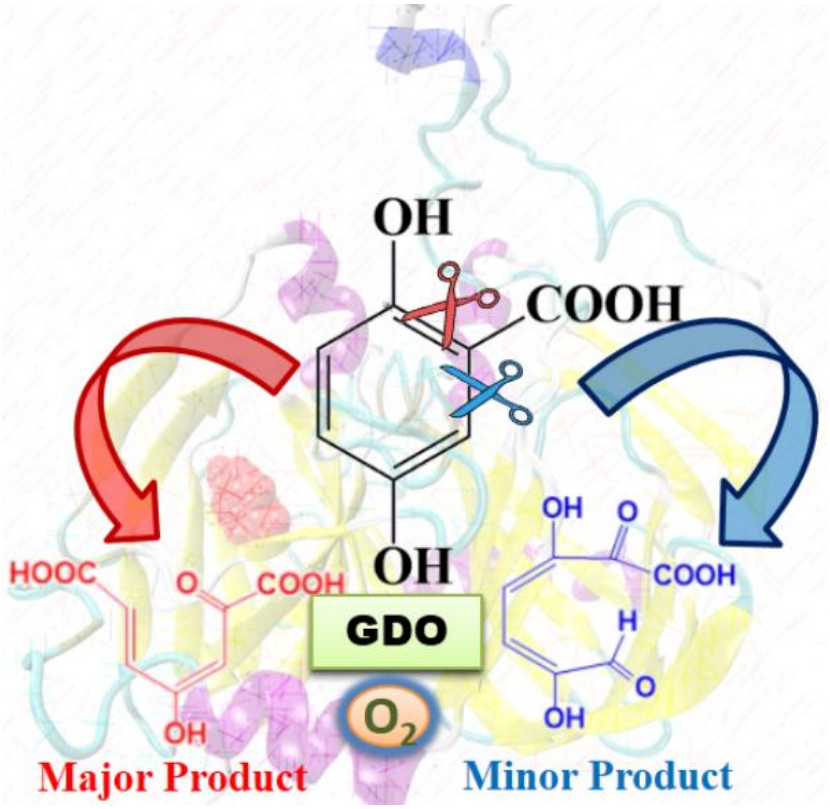
ABBREVIATIONS

QM/MM, Quantum Mechanics/Molecular Mechanics; MD, molecular dynamics; GTQ, Gentisic acid; GDO, Gentisate 1,2-dioxygenase; SDO, Salicylate 1,2-dioxygenase.

REFERENCES

1. Costas, M.; Mehn, M. P.; Jensen, M. P.; Que, L. Dioxygen Activation at Mononuclear Nonheme Iron Active Sites: Enzymes, Models, and Intermediates. *Chem. Rev.* **2004**, *104*, 939-986.
2. Ray, K.; Pfaff, F. F.; Wang, B.; Nam, W. Status of Reactive Non-Heme Metal–Oxygen Intermediates in Chemical and Enzymatic Reactions. *J. Am. Chem. Soc*. **2014**, *134*, 13942–13958.
3. Chen, H.; Cho, K. B.; Lai, W.; Nam, W.; Shaik. S. Dioxygen Activation by a Non-Heme Iron(II) Complex: Theoretical Study toward Understanding Ferric–Superoxo Complexes. *J. Chem. Theory Comput*. **2012**, *8*, 915–926.
4. Lin, Y. T.; Stańczak, A.; Manchev, Y.; Straganz, G. D.; Visser, S. P. Can a Mononuclear Iron(III)- Superoxo Active Site Catalyze the Decarboxylation of Dodecanoic Acid in UndA to Produce Biofuels. *Chem. Eur. J.* **2020,** *26*, 2233-2242.
5. Rahaman, R.; Chakraborty, B.; Paine, T. K. Mimicking the Aromatic Ring Cleavage Activity of Gentisate-1,2-Dioxygenase by a Nonheme Iron Complex. *Angew. Chem. Int. Ed.* **2016**, *55*, 13838 –13842
6. Eppinger, E.; Ferraroni, M.; Bürger, S.; Steimer, L.; Peng, G.; Briganti, F.; Stolz, A. Function of different amino acid residues in the reaction mechanism of gentisate 1,2-dioxygenases deduced from the analysis of mutants of the salicylate 1,2-dioxygenase from Pseudaminobacter salicylatoxidans. *Biochim. Biophys. Acta.* **2015**, *1854*,1425–1437.
7. Sokala, A.; Pindelska, E.; Szeleszczuk, L.; Kolodziejski, W. Pharmaceutical properties of two ethenzamide-gentisic acid cocrystal polymorphs: Drug release profiles, spectroscopic studies and theoretical calculations. *Int. J. Pharm***. 2017**, *522*, 80-89.
8. Roy, S.; Kastner, J. Synergistic Substrate and Oxygen Activation in Salicylate Dioxygenase Revealed by QM/MM Simulations. *Angew. Chem. Int. Ed.* **2016**, *55*, 1168 –1172.
9. Borowski, T.; Siegbahn, P. E. M. Mechanism for Catechol Ring Cleavage by Non Heme Iron Intradiol Dioxygenases: A Hybrid DFT Study. *J. Am. Chem. Soc*. **2006**, *128*, 12941-12953.
10. Harpel, M. R.; Lipscomb, J. D. Gentisate 1,2-Dioxygenase from Pseudomonas. *J. Bio. chem.* **1990**, *265*, 22187-22196.
11. Roy, S.; Kastner, J. Catalytic Mechanism of Salicylate Dioxygenase: QM/MM Simulations Reveal the Origin of Unexpected Regioselectivity of the Ring Cleavage. *Chem. Eur. J.* **2017**, *23*, 8949 – 8962.
12. Chakraborty, B.; Ghosh, I.; Jana, R. D.; Paine, T. K. Oxidative C–N bond cleavage of (2-pyridylmethyl) amine-based tetradentate supporting ligands in ternary cobalt(II)–carboxylate complexes. *Dalton Trans*., **2020**, *49*, 3463–3472.
13. Manna, R. N.; Malakar, T.; Jana, B.; Paul, A. Unraveling the Crucial Role of Single Active Water Molecule in the Oxidative Cleavage of Aliphatic C–C Bond of 2,4′-Dihydroxyacetophenone Catalyzed by 2,4′-Dihydroxyacetophenone Dioxygenase Enzyme: A Quantum Mechanics/Molecular Mechanics Investigation. *ACS Catal.* **2018,** *8* , 10043-10050.
14. Borišek, J.; Pintar, S.; Ogrizek, M.; Turk, D.; Perdih, A.; Novič, M. A Water-Assisted Catalytic Mechanism in Glycoside Hydrolases Demonstrated on the Staphylococcus aureus Autolysin E. *ACS Catal*. **2018**, *8*, 4334–4345.
15. Wang, Y.; Li, J.; Liu, A. Oxygen activation by mononuclear nonheme iron dioxygenases involved in the degradation of aromatics. *J. Biol. Inorg. Chem.* **2017**, *22*, 395–405.
16. Ferraroni, M.; Matera, I.; Steimer, L.; Bürger, S.; Scozzafava, A.; Stolz, A.; Briganti, F. Crystal structures of salicylate 1,2-dioxygenase-substrates adducts: A step towards the comprehension of the structural basis for substrate selection in class III ring cleaving dioxygenases. *J. Struct. Biol*. **2012**, *177*, 431-438.
17. Abraham, M. J.; Murtola, T.; Schulz, R.; Pall, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *Software X*. **2015**, *1–2*, 19– 25.
18. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg,J.J.; Dapprich,S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, Revision A.02; Gaussian, Inc., Wallingford, CT, 2009.
19. Lee, C.; Yang, W.; R. G. Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785-789.
20. Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self – consistent Molecular Orbital Methods.xx.A Basis Set for Correlated Wave Functions. *J. Chem. Phys.* **1980**, *72*, 650–654.
21. Pritchard, B. P.; Altarawy, D.; Didier, B.; Gibsom, T. D.; Windus, T. L. A New Basis Set Exchange: An Open, Up-to-date Resource for the Molecular Sciences Community. *J. Chem. Inf. Model*., **2019**, *59*, 4814-4820.
22. Hay, P. J.; Wadt, W. R. Ab initio effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals. *J. Chem. Phys.*, **1985**, *82*, 299-310.
23. Jorgensen, W.L; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] Potential Functions for Proteins, Energy

Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc*. **1988**, *110*, 1657-1666.
24. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *Journal of Physical Chemistry*, **1987**, *91*, 6269-6271.
25. Berendsen, H. C. J.; Postma, J. P. M.; Gunsteren, W. F. V.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys*. **1984**, *81*, 3684–3690.
26. Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
27. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *J. Chem. Phys*., **1993**, *98*, 10089-10092.
28. Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph*., **1996**, *14*, 33-38.
29. Dong, G.; Lai, W. Reaction Mechanism of Homoprotocatechuate 2,3-Dioxygenase with 4-Nitrocatechol: Implications for the Role of Substrate. *J. Phys. Chem. B* **2014**, *118*, 1791−1798.
30. Wang, B.; Cao, Z.; Rovira, C.; Song, J.; Shaik, S. Fenton-Derived OH Radicals Enable the MPnS Enzyme to Convert 2-Hydroxyethylphosphonate to Methylphosphonate: Insights from Ab Initio QM/MM MD Simulations. *J. Am. Chem. Soc.* **2019**, *141*, 9284−9291.
31. Dong, G.; Shaik, S.; Lai, W. Oxygen Activation by Homoprotocatechuate 2,3-Dioxygenase:a QM/MM Study Reveals the Key Intermediates in the Activation Cycle. *Chem. Sci*., **2013**, *4*, 3624–3635.
32. Marenich, A. V.; Cramer C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B*, **2009,** *113*, 6378-6396.
33. Wadt, W. R.; Hay, P. J. *Ab initio* Effective Core Potentials for Molecular Calculations. Potentials for Main Group Elements Na to Bi. *J. Chem. Phys*. **1985**, *82*, 284-298**.**
34. Hay P. J.; Wadt, W. R. *Ab initio* effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals. *J. Chem. Phys*. **1985**, *82*, 299-310.
35. Field, M. J.; Albe, M.; Bret, C.; Proust-De Martin, F.; Thomas, A. The dynamo library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. *J. Comput. Chem.* **2000**, *21*, 1088– 1100.
36. Swiderek, K.; Tunon, I.; Williams, I. H.; Moliner, V. Insights on the Origin of Catalysis on Glycine N-Methyltransferase from Computational Modeling. *J. Am. Chem. Soc.* **2018**, *140*, 4327– 4334.
37. García-Meseguer, R.; Martí, S.; Ruiz-Pernía, J. J.; Moliner, V.; Tuñón, I. Studying the role of protein dynamics in an S_N2 enzyme reaction using free-energy surfaces and solvent coordinates. *Nat. Chem.* **2013**, *5*, 566– 571.
38. Manna, R. N.; Zinovjev, K.; Tunon, I.; Dybala-Defratyka, A. Dehydrochlorination of Hexachlorocyclohexanes Catalyzed by the LinA Dehydrohalogenase. A QM/MM Study. *J. Phys. Chem. B* **2015**, *119*, 15100– 15109.
39. Roy, L. E.; Hay, P. J.; Martin, R. L. Revised Basis Sets for the LANL Effective Core Potentials. *J. Chem. Theory Comput.* **2008**, *4*, 1029-1031.
40. Florian, W.; Reinhart, A. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys*. **2005**, *7*, 3297-3305.
41. Reiher, M. Theoretical Study of the Fe(phen)2(NCS)2 Spin-Crossover Complex with Reparametrized Density Functionals. *Inorg. Chem.* **2002**, *41*, 6928−6935.
42. Pau, M. Y. M.; Davis, M. I.; Orville, A. M.; Lipscomb, J. D. Solomon, E. I.; Spectroscopic and Electronic Structure Study of the Enzyme-Substrate Complex of Intradiol Dioxygenases: Substrate Activation by a High Spin Ferric Non-Heme Iron Site. *J Am Chem Soc*. **2007**, *129*, 1944–1958.
43. Dawson, W. K.; Jono, R.; Terada, T.; Shimizu, K. Electron Transport in a Dioxygenase-Ferredoxin Complex: Long Range Charge Coupling between the Rieske and Non-Heme Iron Center. *PLOS ONE* **2016**, *11*, 1-41.
44. Christian, G. J.; Ye, a. S.; Neese, F. Oxygen activation in extradiol catecholate dioxygenases – a density functional study. *Chem. Sci*. **2012**, *3*, 1600–1611.


OH
COOH
OH
HOOC
O
COOH
OH
GDO
O_2
OH
O
COOH
H
O
OH
Major Product
Minor Product