**Title:** Structure and functional differences of cysteine and 3-mercaptopropionate dioxygenases. A computational study.

**Authors:** Chieh-Chih Yeh; Christos Pierides; Guy Jameson; Samuel de Visser, PhD

This is the author manuscript accepted for publication. It has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record.

**To be cited as:** 10.1002/chem.202101878

**Link to VoR:** https://doi.org/10.1002/chem.202101878
Structure and functional differences of cysteine and 3-mercaptopropionate dioxygenases. A computational study.

C.-C. George Yeh,[a,b] Christos Pierides,[a,b] Guy N. L. Jameson*[c] and Sam P. de Visser*[a,b]

Abstract: Thiol dioxygenases are important enzymes for human health involved in the detoxification and catabolism of toxic thiols containing natural products such as cysteine, which has relevance to the development of Alzheimer’s and Parkinson’s diseases in the brain. Recent crystal structure coordinates of cysteine and 3-mercaptopropionate dioxygenase (CDO and MDO) showed major differences in the second-coordination spheres of the two enzymes. To understand the difference in activity between these two analogous enzymes, we created large active site cluster models. We show that CDO and MDO have different iron(III)-superoxo bound structures due to differences in ligand coordination. Furthermore, our studies show that the differences in the second-coordination sphere and particularly the position of a positively charged Arg residue results in changes of substrate positioning, mobility and enzymatic turnover. Furthermore, the substrate scope of MDO is explored with cysteinate and 2-mercaptosuccinic acid and their reactivity is predicted.

Introduction

Mononuclear nonheme iron enzymes play pivotal roles to human health and take part in various biosynthesis and biodegradation pathways in the body.[1] For instance, the DNA repair enzymes take alkylated DNA bases and use one molecule of dioxygen on a nonheme iron center perform an N-dealkylation reaction to return the bases to their original structures.[2] A relevant reaction related to aging in humans/mammals is the collagen biosynthesis that requires the amino acid R-4-hydroxyproline, which is synthesized in the body by the nonheme iron dioxygenase prolyl-4-hydroxylase through the activation of a proline residue stereoselectively via a negative catalysis process that avoids more stable reaction products.[3] In plants the nonheme iron dioxygenases take part in various steps along the flavonol biosynthesis pathways, which are important signaling molecules related to flowering and fruit growth,[4] while in bacteria the synthesis of many natural products such as antibiotics involve nonheme iron dioxygenases. Although the majority of reactions catalyzed by mononuclear nonheme iron dioxygenases cover substrate hydroxylation, actually this class of enzymes is very versatile and many alternative reaction pathways have been characterized including substrate epoxidation, ring-closure processes and desaturation.[5]

In addition to those pathways there are also sulfur-activating nonheme iron dioxygenases that take part in the biodegradation and catabolism of thiolate compounds across kingdoms. Those enzymes include cysteine dioxygenase (CDO) in humans and 3-mercaptopyruvate dioxygenase (MDO) in bacteria. Scheme 1. These enzymes take one molecule of O2 and react with their substrate to form cysteine sulfenic acid (CysO2) and 3-sulfinopropionic acid (MPO2), respectively. Thus, CDO is a nonheme iron dioxygenase that initiates the biodegradation of toxic cysteine as under elevated concentrations it can trigger neurological problems such as Alzheimer’s and Parkinson’s diseases.[6] Therefore, understanding the catalytic mechanism of thiol dioxygenases is important for human health. On the other hand, recent studies on plant CDOs showed the cysteine sulfenic acid products are linked to responses to hypoxia and maybe related to oxygen sensing mechanisms.[7] Many details on the exact pathways of cysteine degradation in the body are still missing and controversial, and several possible mechanisms have been suggested. One major difference between the thiol dioxygenases and nonheme iron hydroxylases relates to the first-coordination sphere ligands of the metal.
Most nonheme iron dioxygenases have a typical facial 2-His/1-carboxylate ligand architecture around the iron atom, where it is bound to the side chains of two His residues and a carboxylate group from either a Glu or Asp residue of the protein.\[9\] Interestingly, CDO has the carboxylate ligand missing and the iron is bound through the interactions of three neutral histidine side chains only (3-His ligand feature).\[9\] It has been proposed that this 3-His motif gives thiol dioxygenases its catalytic efficiency as it binds the cysteinate ligand stronger and also affects the O–O cleavage step in the catalytic cycle.\[9\]

Experimental studies on CDO using \[16O\] in the reaction confirmed both oxygen atoms of CysO₂ to originate from dioxygen.\[10\] The CDO resting state and substrate-bound forms have been trapped and characterized with spectroscopic techniques, including UV-Vis absorption, resonance Raman and electron paramagnetic resonance (EPR) studies as well as X-ray crystallography.\[11\]

Studies of NO binding to CDO established a mechanism proceeding through an iron(III)-superoxo intermediate.\[12\] Furthermore, using UV-Vis and EPR/Mössbauer techniques a short-lived dioxygen-bound intermediate of the catalytic cycle of CDO was characterized, which presumably is the iron(III)-superoxo species.\[13\]

Similar computational studies with either density functional methods on cluster models or quantum mechanics/molecular mechanics on the full enzymatic structure indeed confirmed a stepwise mechanism of sequential oxygen atom transfer reactions by an iron(II)-superoxo and iron(IV)-oxo species to form cysteine sulfenic acid products.\[14\]

A substrate-bound crystal structure of CDO was resolved\[15\] and the active site is highlighted in Figure 1a. It has a central iron atom linked to the protein through a facial 3-His coordination through His₈₆, His₈₈ and His₁₄₈. Substrate cysteinate binds as a bidentate ligand through the amine and thiolate groups to the metal. The substrate is tightly bound in the substrate binding pocket through a salt-bridge with Arg₆₀ and interactions with His₁₅₅ and Tyr₁₅₇.

An interesting feature of the CDO binding pocket is the covalent linkage between the Cys₉₃ and Tyr₁₅₇ groups at the ortho-position of the phenol, which has been proposed to be linked with the tight substrate binding pocket and the positioning of substrate.\[16\]

MDO similarly to CDO has a nonheme iron active center with three coordinating His residues, Figure 1b. Several biochemical studies investigated the function and properties of MDO and established its substrate selectivity, pH range and characterized the iron-bound structure spectroscopically.\[17\] The active site structure of MDO as taken from the 4TLF pdb file\[18\] is shown on the right-hand-side of Figure 1. Similarly to CDO, the metal in MDO is held in a facial 3-His orientation through interactions with His₈₉, His₉₁, and His₁₄₂. Nevertheless, there are distinct differences in the substrate binding pocket residues between CDO and MDO. In particular, the position of Arg₁₆₈ in CDO has been replaced by a Gin residue in MDO (Gln₂₀), see Figure 1. As such, the MDOs were initially labeled as a “Gln”-type CDO, but they are now considered as a separate class of dioxygenases.\[19\] There is; however, an Arg residue in position 168 in MDO (not seen in CDO), and it has been suggested that the carboxylate group of the substrate, i.e. cysteinate (Cys) in CDO and 3-mercaptopropionate (MP) in MDO binds an active site Arg residue through a salt bridge, i.e. Arg₁₆₈ in CDO and Arg₁₆₈ in MDO.\[20\]

Based on these differences in position of the Arg residues it is clear substrate is bound differently in the two enzymes, which could give functional and mechanistic differences. Figure 1b displays the proposed substrate binding in MDO as docked into the substrate-free MDO structure.\[20\] In addition to the difference in position of an active site Arg residue, the Tyr-Cys linkage is missing in MDO although a Tyr is positioned in the same position (Tyr₁₉₉). Clearly, these differences in active site residues must have an influence on substrate binding and positioning in CDO versus MDO and maybe even influence the catalytic mechanism.

---

**Figure 1.** Active site structures of CDO and MDO as taken from the 2IC1 and 4TLF pdb files. Substrate MP (in pink) was docked into the 4TLF file.
To gain insight into the substrate dioxygenation mechanism of MDO and how it compares to CDO, we created a large active site model and did a density functional theory (DFT) study alongside a large active site CDO cluster model. The work shows that the protein environment directs and guides the reaction mechanism by positioning substrate and oxidant in the active site. Furthermore, our studies presented here show that the carboxylate ligand of substrate in MDO is bound to the iron and donates electron density to the metal and thereby affects the thermodynamics and kinetics of the reaction and the electronic configuration of the various intermediates.

Results and Discussion

We created large active site cluster models of CDO and MDO based on the crystal structure coordinates deposited in the 2IC1 and 4TLF pdb files.\cite{15,20} These cluster models were created using previously described methods and take key hydrogen bonding and electrostatic interactions of the co-factor and substrate binding pocket into consideration.\cite{21} Scheme 2 shows the 308 atom CDO and 324 atom MDO models used in this work. We took the iron atom with its direct ligands bound to short peptide chains that surround the metal and substrate, which will give the model its constraints seen in the enzyme. In particular, a long peptide chain in CDO from Ser94 to Leu95 links the metal to two of its His ligands (His_{86} and His_{88}) and also connects it to the Tyr-Cys active site cross-linked group via Cys_{93}. In addition, the CDO model (A) includes residues that determine the positioning of the substrate and the shape and size of the substrate-binding pocket and the hydrogen-bonding interactions positioning the substrate and oxidant. These residues consist of the Arg_{89} side chain that forms a salt bridge with the carboxylate group of Cys, but also the hydrogen bonding residues Tyr_{157}, Ser_{94}, Tyr_{157} and His_{95}. The iron(II) atom was replaced by iron(III)-superoxo to create the starting structure of the oxygen-activation mechanism, which gives model A an overall charge of zero.

Model B is the MDO cluster model, which has similar features as the CDO model. It also includes a peptide chain from His_{84} to Ile_{97}, incorporating two of the iron ligands, namely His_{84} and His_{91}. As can be seen from Scheme 2, the Cys_{93} residue of CDO is missing in MDO and in its position is a Gly residue (Gly_{93}) located. Nevertheless, we kept the chain as it surrounds the substrate and active site and therefore determines the substrate positioning. The model is completed by the charged residue Arg_{168} and some hydrogen bonding residues in the active pocket such as Tyr_{160}, His_{157} and Tyr_{155}. We inserted the substrate in the position as recommended by Jameson et al.\cite{22} and replaced iron(II) by iron(III)-superoxo to complete a charge-neutral model. As our cluster models are close to spherical shape and contain all hydrogen bonding interactions between first and second-coordination sphere, we ran the calculations without geometric constraints.

We started with a series of calculations on the dioxygenation mechanism of CDO, whereby we compare the results of the large DFT cluster model of 308 atoms with the results of the QM/MM calculations from the literature.\cite{14b,4b,14a} The latter used a QM region of 81 atoms only and hence, our cluster model described here contains considerably more atoms calculated with a QM method and therefore will describe interactions between the first and second-coordination sphere better. A geometry optimization of the iron(III)-superoxo complexes of CDO in various spin states was done and the results are given in Figure 2. In agreement with previous DFT and QM/MM studies on CDO,\cite{14b} we find an open-shell singlet ground state (1\text{Re}_s). The quintet and triplet spin states are higher in energy by \Delta E +ZPE = 4.6 and 6.5 kcal mol^{-1}, respectively. As such, there are close-lying spin state surfaces and a multistate reactivity pattern can be expected with reactions and transition states on each of the individual spin state surfaces.\cite{15} The relative energies match the results obtained with QM/MM excellently and reproduce an open-shell singlet ground state. The average singlet-triplet gap with QM/MM averaged over seven snapshots was 4.7 kcal mol^{-1}, while the singlet-quintet gap for the same series came out at 14.2 kcal mol^{-1}.\cite{14b}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Scheme2.png}
\caption{Cluster models of CDO and MDO studied in this work. Residues were cut after the peptide bond and capped with a hydrogen atom.}
\end{figure}
An overlay of the quintet spin reactant complex with the crystal structure coordinates (Figure 2), shows that even without geometric constraints on the model structure, our model matches the original crystal structure coordinates almost perfectly. In particular, most peptide backbone groups are in the same position and the iron has the same coordination environment. This probably is due to the fact that various β-sheets and α-helix components around the active site have been included into the model. As such, the model appears to be an excellent representation of the actual system and no dramatic changes have occurred from the crystal structure orientation. Moreover, the overlay between the optimized geometry and the original crystal structure coordinates shows that the CDO active site is highly rigid and groups are positioned tightly through hydrogen bonding interactions.

Structurally, all reactant structures are in the end-on superoxo form with a short O–O distance of 1.29 Å for \( \text{Re}_A \), 1.31 Å for \( \text{Re}_{\alpha} \) and 1.30 Å for \( \text{Re}_A \). The Fe–O distance is short in the singlet spin state (1.76 Å), while it is much longer in the quintet spin state (2.24 Å). Interestingly, the metal–ligand distances in the equatorial plane are virtually the same in the singlet and quintet spin states. The equatorial Fe–amide distances are short in all spin states (2.04 – 2.06 Å), while the axial Fe–N distances range from 2.088 Å in \( \text{Re}_A \) to 2.363 Å in \( \text{Re}_A \). These differences in Fe–O and Fe–N bond lengths probably originate from the molecular orbital occupation in the singlet and quintet spin states. Thus, the singlet spin has electronic configuration \( d_{yz}^2 d_{x^2}^1 d_{y^2}^1 \) \( \sigma_{O=O} \) \( \pi_{O=O} \) \( \sigma_{O=O}^* \), while in the quintet spin state it is \( d_{yz}^2 d_{x^2}^1 d_{y^2}^1 \) \( \pi_{O=O} \) \( \sigma_{O=O} \) \( \sigma_{O=O}^* \). These orbitals represent the interactions of the metal 3d orbital with its ligands, where we take the z-axis along the O–Fe–N(His) bond.

In the singlet spin state the five electrons on the metal have the \( d_{yz} \) orbital singly occupied and antiferromagnetically coupled to the electron in the antibonding O–O interaction parallel to the xy-plane, i.e. the \( \pi_{O=O} \) orbital. In the quintet spin state all electrons are ferromagnetically coupled. We located also an alternative quintet spin state structure (\( \text{Re}_{\alpha} \)), which has the \( \sigma_{xy} \) orbital singly occupied instead of the \( \sigma_{yz} \) orbital. This state is 6.9 kcal mol\(^{-1}\) higher in energy than \( \text{Re}_A \) and was not considered further.

Next, we investigated the full reaction mechanism of dioxygenation of cysteine by our CDO model A following the mechanism shown in Scheme S3 (Supporting Information) and calculated before using alternative cluster models and QM/MM.[14] This mechanism starts from an iron(III)-superoxo complex (\( \text{Re}_A \)) where the terminal oxygen atom attacks the substrate sulfur atom via transition state \( \text{TS}_1 \) to form a ring-structure (Fe–O–O–S ring) \( \text{IM}_1 \). Next the dioxygen bond breaks via transition state \( \text{TS}_2 \) to form an iron(IV)-oxo species and cysteine sulfoxide (\( \text{IM}_2 \)). This step is followed by a second oxygen atom transfer via transition state \( \text{TS}_3 \) to form cysteine sulfenic acid bound products (\( \text{Pr}_A \)).

We tested the mechanism on the low lying singlet, triplet and quintet spin states from the reactant complexes shown in Figure 3. For MDO a similar mechanism was considered but for MDO the structures have the label B as subscript.

The full potential energy landscape for dioxygenation of CDO by cluster model A is shown in Figure 3. As can be seen from Figure 3 the dioxygenation of Cys starts from the singlet spin iron(III)-superoxo with S–O bond formation, whereby a spin-state-crossing to the quintet spin transition state takes place. Thus, the relative energy with zero point energy (ZPE) included for \( \text{TS}_1 \) is above the reactants structure by \( \Delta E + ZPE = 9.4 \) kcal mol\(^{-1}\).
Figure 3. Potential energy landscape of cysteine dioxygenation by the CDO model A as obtained at UB3LYP in Gaussian. Out of parenthesis refer to $\Delta E+$ZPE values, while relative free energies are in parenthesis. Data obtained at UB3LYP/BS2//B3LYP/BS1 level of theory with values in kcal mol$^{-1}$.

This value is close in energy to the small model complex reported previously, where a barrier of 10.0 kcal mol$^{-1}$ was reported, while with QM/MM a value of 14.4 kcal mol$^{-1}$ was obtained.$^{[14]}$ Consequently, the enlarged model has little effect on the energetics of the dioxygen activation step and is a good model for the CDO reaction mechanism.

The singlet spin transition state for S–O bond formation ($^1$TS1$_A$) was also located but found to be 21.3 kcal mol$^{-1}$ higher in energy than $^1$Re$_A$. Therefore, the singlet spin iron(III)-superoxo will be a sluggish oxidant and cannot react with sulfides and a spin state crossing to the more reactive quintet spin state will be needed. As a matter of fact the full singlet spin surface is well above the quintet spin state beyond the reactant conformation and after the first transition state another high energy transition state needs to be crossed (via $^5$TS2$_A$) to form cysteine sulfoxide intermediates with a barrier of $\Delta E+$ZPE = 21.5 kcal mol$^{-1}$. Consequently, the singlet spin surface will play little role in the substrate dioxygenation reaction in CDO enzymes.

The quintet spin barrier for S–O bond formation is the lowest in energy and leads to the ring-closure complex with Fe–O–O–S orientation $^5$IM$_1A$ with a small exothermicity of $\Delta E+$ZPE = -3.6 kcal mol$^{-1}$ with respect to $^1$Re$_A$. The structure will have a short lifetime as the O–O cleavage barrier to form cysteine sulfinate and an iron(IV)-oxo species via $^5$TS2$_A$ is only 3.0 kcal mol$^{-1}$ above $^5$IM$_1A$. Similarly to previous DFT and QM/MM studies on CDO enzymes, the iron(IV)-oxo species has close lying triplet and quintet spin states (within 2 kcal mol$^{-1}$) with the triplet spin slightly lower. This result matches studies on hexacoordinated biomimetic iron(IV)-oxo species that usually give a triplet spin ground state.$^{[23]}$ By contrast, alternative pentacoordinated iron(IV)-oxo complexes such as those in nonheme iron enzymes with 2-His/1-Asp ligand features have a quintet spin ground state.$^{[24,25]}$ Nevertheless, the triplet spin iron(IV)-oxo species has the lowest oxygen atom transfer barrier to form the cysteine sulfinic acid product complexes. However, the product complexes are the most stable in the quintet spin state. Depending on the lifetime of the iron(IV)-oxo intermediate and the possibility of spin state change from quintet to triplet, the second oxygen atom transfer will either proceed via $^5$TS3$_A$ or $^3$TS3$_A$ to form cysteine sulfinic acid products.

The triplet spin mechanism was also calculated and at the reactants stage it is only 4.6 kcal mol$^{-1}$ above the singlet spin state with the same orbital occupation. Similarly to the singlet spin state, the S–O bond formation transition state is high in energy and found to be well above 23 kcal mol$^{-1}$. Nevertheless, the triplet spin state surface becomes the lowest energy at the iron(IV)-oxo-sulfoxide intermediate ($^3$IM$_2A$), which is slightly lower in energy than the analogous quintet spin state. However, the second oxygen atom transfer on the triplet surface is well higher to that of the quintet spin state and will lead to a quintet spin product complex.

The optimized geometries of the three quintet spin transition states are shown in Figure 4, where the two oxygen atoms in FeO$_2$O$_2$ are labelled as distal and proximal.
The S–O bond formation transition state (5TS1A) has an imaginary frequency of i358 cm⁻¹ for the S–O stretch vibration. Indeed, the Fe–O and O–O distances have changed from 2.237 and 1.298 Å in 5ReB, to 2.107 and 1.324 Å in 5TS1A, respectively. Moreover, the S–O distance is shortened from 3.623 to 2.163 Å for the same structures. The optimized geometries match those of the small model complex reported in Ref [14a] well, see Figure 4 where the comparative data are given. The only difference appears to be for the S–O distance that is well longer with the smaller cluster. However, a comparison of the large DFT structure optimized geometries with those obtained with QM/MM shows that the Fe–O, O–O and O–S distances are all within 0.1 Å. Consequently, our DFT cluster model gives almost identical S distances are all within 0.1 Å.

After the S–O bond formation transition state the system relaxes to a ring-structure 5IM1A, which is slightly lower in energy than reactants (ΔE/ZPE = −3.6 kcal mol⁻¹). A small O–O cleavage barrier 5TS2A of only ΔE/ZPE = 2.0 kcal mol⁻¹ above 5IM1A converts the system to an iron(V)-oxo cysteine sulf oxide complex 5IM2A that is well lower in energy than reactants by −15.7 kcal mol⁻¹. The 5TS2A structure has a well separated dioxygen interaction of 1.885 Å, while the S–O and Fe–O distances have shortened to 1.589 and 1.762 Å. These distances are in good agreement with those obtained previously, as highlighted in Figure 4. The imaginary frequency of i443 cm⁻¹ in 5TS2A corresponds to the O–O stretch vibration. The 5TS3A structure is also shown in Figure 4 and has the S–O distance for the bond that is being formed at 2.385 Å, which is at a slightly longer distance than in 5TS1A. The imaginary frequency in 5TS2A is i302 cm⁻¹ for the S–O bond stretch vibration. Overall, the structures and potential energy profile calculated with the large model of CDO match previous work on smaller model complexes and QM/MM excellently. [14a]

Next, we investigated the reaction mechanism of 3-mercapto propionate activation by the MDO model and started with analyzing the optimized geometries of the reactant complexes (see Figure 5). An overlay of the optimized geometry of 5ReB with the crystal structure coordinates is shown in Figure S15 (Supporting Information). The backbone atoms of the protein are in virtually the same position in the model as in the crystal and therefore the calculations predict the correct fold and secondary structure of the active site. Similarly to the CDO structures, also for MDO the open-shell singlet spin state is the ground state with electronic configuration dₓ²dᵧ²d₂z²σₒₒ⁺1σₒₒ⁻1. Its molecular orbitals are schematically depicted in Figure 6. They look alike to those of CDO and analogous nonheme iron(III)-super oxide complexes in enzymes and model systems. [14b,26] The electronic configuration of the end-on iron(III)-superoxide complexes are determined by the metal 3d orbitals and their interactions with first-coordination sphere ligands and the antibonding interactions in the superoxide group. Lowest in energy are the dₓ and dᵧ orbitals that are virtually nonbonding. The d₂z orbital; however, interacts with the σ-orbital on the super oxide group perpendicular to the xy-plane as 2.107 Å. Higher in energy are the antibonding interactions of the iron 3dₓ and 3dᵧ orbitals with ligands along the z-axis (σₒₒ⁺1) and in the xy-plane (σₒₒ⁻1). Finally, there is a π-orbital along the superoxide interaction that is parallel to the xy-plane: πₒₒ⁻1. The triplet spin state with the same configuration as 5ReB is higher lying by 6.7 kcal mol⁻¹, which is almost identical to what is obtained for CDO. By contrast, the quintet spin state with electronic configuration dₓ²dᵧ²d₂z²σₒₒ⁺1σₒₒ⁻1σₒₒ⁻1; is only 0.8 kcal mol⁻¹ higher in energy than 5ReB and as such is significantly stabilized with respect to the CDO model. We swapped molecular orbitals and also converged the wavefunction to a quintet spin configuration dₓ²dᵧ²d₂z²σₒₒ⁺1σₒₒ⁻1σₒₒ⁻1; however, we found it 15.5 kcal mol⁻¹ (ΔE at BS2 level of theory) in energy hence did not consider this state further.

An overlay of the two quintet spin reactant complexes is shown on the right-hand-side of Figure 5. Thus, the CDO model binds a substrate (Cys) with a carboxylate group that binds to the conserved Arg90 and Tyr158 residues and hence is tightly locked in position.
Figure 5. Optimized reactant geometries of the iron(III)-superoxo complex of MDO for the large cluster model B as calculated at UB3LYP/BS1 in Gaussian-09. Energies are in kcal mol$^{-1}$ and optimized geometries give bond lengths in angstroms and angles in degrees. The right-hand side shows an overlay of the $^{1}$Re$_{A}$ and $^{1}$Re$_{B}$.

$\Delta E$+ZPE = 0.0 (6.7) [0.8]

$^{1}$Re$_{B}$ ($^{3}$Re$_{B}$) [5Re$_{B}$]

Figure 6. Relevant molecular orbitals and orbital occupations of $^{1}$Re$_{B}$ and $^{3}$Re$_{B}$.
Furthermore, CDO has the Tyr-Cys linked group in the active site, which is missing in MDO as well. Finally, CDO was characterized with a cis-peptide bond through residues Ser169-Pro159.20 The corresponding loop in MDO; however, has Gly and Ala in those positions and hence does not show a kink in the protein. As a result of these second coordination sphere effects the metal with its first-coordination sphere ligands have been affected as well. In particular, the Fe–S distance is 2.270 Å in $^{5}\text{ReO}_5$ while it has elongated to 2.631 Å in $^{5}\text{ReO}_5$. In addition, the proximal oxygen atom of the superoxo group in $^{5}\text{ReO}_5$ is held in position by a hydrogen bond from the phenol group of Tyr159, which is missing in CDO. Instead, the Tyr residue in that position in CDO (Tyr157) is part of the Tyr-Cys linkage and points further away and forms a hydrogen bond to the carboxylate group of substrate Cys instead.

It may be that the Tyr-Cys linkage in CDO functions to hold the Tyr away from the superoxo group and to position the substrate carboxylate group in the active site. As a result of the hydrogen bonding interaction of superoxo with Tyr159 in $^{5}\text{ReO}_5$, the Fe–O bond length in the singlet spin state is 1.90 Å, while it was 1.76 Å in $^{5}\text{ReO}_5$. The change in Fe–O bond length is much more pronounced in the quintet spin reactants, where it drops slightly from 2.24 Å in $^{5}\text{ReO}_5$ to 2.22 Å in $^{5}\text{ReO}_5$. On the other hand, the equatorial Fe–S bond length is considerably weakened from 2.27 Å to 2.63 Å for the same structures. Clearly, the local environment and particularly, the hydrogen bonding interactions, of, e.g., Tyr159 affect first-coordination sphere interactions between metal and ligands. This will further affect spectroscopic parameters such as IR, resonance Raman as well as Mössbauer parameters of the various complexes. Moreover, work of Jameson et al170 on the G95S variant of MDO showed complete lack of reactivity probably due to repositioning the Tyr159 group in the active site.

To investigate the substrate-binding position and orientation of 3-mercaptopyrrole in the pocket of MDO, we considered several alternative binding poses. In particular, we geometry optimized a monodentate ligated 3-mercaptopyrrole-bound reactant complex through the sulfur atom to iron only. The structure optimized to a local minimum with a protein fold close to that of $^{5}\text{ReO}_5$ (see Supporting Information, Figure S16). Consequently, a monodentate-binding orientation will fit the substrate binding pocket of MDO; however, it is 15 kcal mol$^{-1}$ higher in energy than the bidentate-bound reactant complex. This is most likely the result from the lack of the salt bridge between the carboxylate group of substrate with the Arg168 side chain in the monodentate orientation. The carboxylate group of 3-mercaptopyrrole in the monodentate orientation; however, interacts with the phenol group of Tyr150 and protons from the imidazole group of a His ligand. Nevertheless, as it is well higher in energy than the bidentate-ligated substrate-bound structure, it was not considered further.

Subsequently, the dioxygen activation process of 3-mercaptopyrrole by MDO model complex B was studied on the singlet, triplet and quintet spin state surfaces and the energy landscape is shown in Figure 7. Similarly to the CDO mechanism described above in Figure 3, also for MDO a spin state crossing from singlet to quintet takes place during the S–O bond formation en route to TS1. The quintet spin structures for MDO; however, are the lowest in energy along the full landscape apart from the iron(III)-superoxo structure where the open-shell singlet spin state is slightly lower in energy. Consequently, the reaction will take place on a dominant quintet spin state surface. The first barrier is small on the quintet spin state and in a large exothermic step leads to the ring-structure intermediate $^{6}\text{M1}_2$ with $\Delta E+\text{ZPE} = -20.3$ kcal mol$^{-1}$. This structure is followed by a transition state (TS2$^{B}$) of 5.7 kcal mol$^{-1}$ relative to $^{6}\text{M1}_2$ and again a large exothermic step to form the iron(V)-oxo sulfenate intermediate $^{6}\text{M2}_2$. Interestingly, the second oxygen atom transfer reaction in MDO is the rate-determining step with a magnitude of 15.8 kcal mol$^{-1}$ above $^{6}\text{M2}_2$. As such, the CDO and MDO dioxygenation mechanism in both cases starts from an open-shell singlet iron(III)-superoxo complex that reacts through spin crossover from singlet to quintet during the S–O bond formation reaction step. Furthermore, in both mechanisms the subsequent steps leading to dioxygenated product is on a dominant quintet spin state surface although the iron(IV)-oxo structure of CDO also has a close-lying triplet conformation.

We also tested the full mechanism of dioxygen activation by MDO on the singlet and triplet spin state surfaces. The triplet spin starts at 6.7 kcal mol$^{-1}$ above reactants and undergoes three small barriers to form MPO$_2$ products. On the singlet spin state surface, by contrast, a significant S–O barrier is encountered first (of 14.8 kcal mol$^{-1}$), while the next barrier is only 9.4 kcal mol$^{-1}$ and the final barrier only 3.3 kcal mol$^{-1}$. Therefore, on the singlet spin state, the rate-determining step is via $^{5}\text{TS}_1$, although this barrier is significantly lower in energy than the one found for CDO. As such, CDO will need a spin crossover for the first S–O bond formation step to the quintet spin state surface, while the other spin states will play little role of importance. On the other hand, for MDO all spin states can lead to products with barriers well below 15 kcal mol$^{-1}$ maximum. Therefore, on each spin state surface the reaction can take place, although the quintet spin is the lowest in energy.

Optimized geometries of the quintet spin transition states for the MDO pathway are shown in Figure 8. The first transition state is low in energy but characterized with a frequency that gives a mode for the S–O stretch vibration with magnitude 1213 cm$^{-1}$. Similarly, $^{5}\text{TS}_1$, shown above in Figure 4, $^{5}\text{TS}_2$ has a short O–O distance of 1.313 Å and long Fe–O interaction of 2.172 Å. The MDO structure $^{5}\text{TS}_2$; however, has a S–O distance of 2.414 Å, while it was only 2.163 Å in $^{5}\text{TS}_1$. Most probably, the hydrogen bonding interaction of Tyr159 is responsible for the S–O elongation in $^{5}\text{TS}_2$, but also appears to give a very low reaction barrier on the quintet spin state surface.

Despite the fact that the S–O bond formation appears to be easier in MDO than in CDO that is not the case for the rest of the mechanism and the O–O cleavage barrier is raised in energy with respect to CDO as well as the second oxygen atom transfer barrier. Geometrically, $^{5}\text{TS}_2$ has a shortened Fe–O bond of 1.838 Å, which is not surprising as an iron(IV)-oxo intermediate is formed. This distance is somewhat longer as the one found in the CDO model. The O–O stretch vibration has an imaginary frequency of 1524 cm$^{-1}$ and takes place at a distance of 1.841 Å. These values are of the same order of magnitude as those found for the CDO model. Therefore, the only major difference of $^{5}\text{TS}_2$, and $^{5}\text{TS}_2$ appears to be the Tyr159 hydrogen bond that may slow the O–O cleavage step down.

Interestingly, in MDO the rate-determining step in the mechanism appears to be the second oxygen atom transfer from the iron(IV)-oxo species. This is mostly because of the high stability of the iron(IV)-oxo complex of MDO. Thus, $^{5}\text{M2}_2$ is more stable than $^{5}\text{ReO}_5$ by $\Delta E+\text{ZPE} = 42.1$ kcal mol$^{-1}$, while the same step in CDO only releases 17.5 kcal mol$^{-1}$. 

8
As MDO binds a substrate with a carboxylate ligand to iron this finding implies that its corresponding iron(IV)-oxo species has the metal bound through three neutral N-based ligands and two anionic ligands of the substrate. By contrast, in CDO the iron(IV)-oxo species binds a single anionic ligand (the sulfur of Cys), while all other ligands are neutral N-based groups. This additional anionic ligand in MDO leads to a considerably more stable iron(IV)-oxo species than the one in CDO and shows that nonheme iron enzymes need a 2-His/1-carboxylate ligand to efficiently generate an iron(IV)-oxo species. Furthermore, it explains why most nonheme iron dioxygenases have a 2-His/1-Asp or 2-His/1-Glu ligand conformation as it leads to an iron(IV)-oxo species from an iron(III)-peroxo more efficiently. The final transition state in the MDO mechanism (5^TS3_A) is shown on the right-hand side of Figure 8 and is characterized with an imaginary of i466 cm\(^{-1}\) for the second oxygen atom transfer to sulfur. The structure shows similarities to the CDO model transition state for the second oxygen atom transfer.

Overall the reaction mechanisms for substrate activation by CDO and MDO show a similar reaction profile with some distinct differences. As discussed above both CDO and MDO form an open-shell singlet spin iron(III)-peroxo structure with two unpaired electrons. However, they have distinctly different close-lying quintet spin states with either \(d_{\sigma}\) or \(d_{\sigma}^{-1} d_{\pi}^{-1} \sigma_{\pi x y z}^{-1} \sigma_{\pi z}^{-1} \pi_{\pi \sigma}^{-1} (5^ReA)\) or \(d_{\pi}^{-1} d_{\pi}^{-1} d_{\pi}^{-1} \sigma_{\pi x y z}^{-1} \sigma_{\pi z}^{-1} \pi_{\pi \sigma}^{-1} (5^ReB)\) electronic configuration. These differences will affect the substrate activation mechanism. Indeed, MDO has a small spin-state crossing to the quintet spin state and will quickly converge to an iron(IV)-oxo species and sulfoxide, while the barrier for CDO is considerably larger for that step. The electronic differences between the oxygen-bound CDO and MDO complexes also affect the thermodynamics to form the iron(IV)-oxo species and its stability, whereby the MDO variant is considerably lower in energy.

To understand the differences between the 5^IM2_A and 5^IM2_B structures we give the optimized geometries side-by-side in Figure 9. The overlay of 5^IM2_A and 5^IM2_B shows that the protein environment is in approximately the same positions in both structures and the main coordination environment does not seem to have changed dramatically. In CDO the cysteine sulfoxide is held by the same hydrogen bonding interactions as the cysteine reactant in ReA, namely a salt bridge of Arg60 with its carboxylate ligand. In addition, there are also hydrogen bonds to the salt bridge from Tyr58 and Tyr157. As highlighted by the structural drawings at the bottom of Figure 9, in MDO the substrate carboxylate points in the different orientation than in CDO and as a consequence the active site Arg is located at a different position, as shown by the relative orientation of Arg60 and Arg168.

---

**Figure 7.** Potential energy landscape of 3-mercaptopropionate dioxygenation by the MDO model B as obtained at UB3LYP in Gaussian. Out of parenthesis refer to \(\Delta E + ZPE\) values, while relative free energies are in parenthesis. Data obtained at UB3LYP/BS2//UB3LYP/BS1 with values in kcal mol\(^{-1}\).
Figure 8. Optimized geometries of UB3LYP/BS1 calculated transition states along the mechanism of dioxygenation of 3-mercaptopropionate by MDO. Bond lengths are in angstroms and the imaginary frequency is in cm$^{-1}$.

Figure 9. Optimized geometries of $^5$IM2$_A$ and $^5$IM2$_B$ with bond lengths in Å. The right-hand-side gives an overlay of the two structures.

In all MDO structures the substrate carboxylate group forms a hydrogen bonding interaction with this Arg, although only with one of its CO groups. The second CO group of the carboxylate hydrogen bonds with Tyr$_{60}$ instead. The differences in position of the substrate to Arg salt bridge has a dramatic effect on the substrate binding and particularly the Fe–S bond length. Thus, in CDO the salt bridge is almost parallel to the Fe–S bond and a dihedral $\text{N}_\text{Arg60}-\text{O}-\text{S}-\text{Fe}$ of $-159^\circ$ is found in $^5$IM2$_A$, while in MDO the dihedral is $41^\circ$. In a parallel configuration, the salt bridge will donate electron density into the Fe–S bond and strengthen the interaction. Indeed, a short Fe–S bond of 2.275 (2.270) Å is seen in $^1$Re$_A$ ($^1$Re$_A$). By contrast, in MDO the salt bridge is in a different plane more perpendicular to the Fe–S bond and pushes the sulfur away from iron and a much longer Fe–S distance is found: 2.631 Å in $^5$Re$_B$. Therefore, the cysteinate substrate in CDO is stronger bound to the metal center than the 3-mercaptopropionate in MDO.
Upon sulfoxidation, the Fe–S distance is weakened and elongated to 2.588 Å in \(^5\)IM2, while it is 2.948 Å in \(^5\)IM2. As such, the sulfoxide is much weaker bound in MDO than CDO and may be released prior to the final oxygen atom transfer to form MPO\(_2\) products. It would be interesting to see if sulfoxide products can be detected in MDO enzymes.

The overall reaction mechanism established for CDO and MDO with their natural substrate is shown in Figure 10. Both enzymes proceed through double oxygen atom transfer in the same mechanism, whereby dioxygen binds as an end-on iron(III)-superoxo form with an open-shell singlet spin ground state with two unpaired electrons in \(d_{xz}, d_{yz}\) and \(\pi_{||}\). During the first oxygen attack on the sulfur of substrate a spin-state crossing to the quintet spin state takes place to form a bicyclic structure (IM1) with high-spin conformation: \(\pi_{xy}^{1}\ \pi_{xz}^{1}\ \pi_{yz}^{1}\ \sigma_{z1}^{2}\ \sigma_{x2-y2}^{1}\). Heterolytic cleavage of the dioxygen bond gives an iron(IV)-oxo species and sulfoxide products. A final oxygen atom transfer gives cysteine sulfenic acid products bound to an iron(II) center with configuration \(\pi_{xy}^{1}\ \pi_{xz}^{1}\ \pi_{yz}^{1}\ \sigma_{z1}^{2}\ \sigma_{x2-y2}^{1}\). The overall reaction is much more exothermic for MDO than for CDO and has the singlet and quintet spin states close in energy in the reactant configuration. As a consequence, we expect the reaction to be faster for MDO than for CDO. However, the substrate appears to be much weaker bound due to a carboxylate group equatorial to the sulfur.

To understand the reaction exothermicity of CDO and MDO better, we decided to calculate substrate and product binding energies. Thus, we took the optimized geometries of \(^4\)Re, \(^5\)Re, \(^6\)Pr, and \(^6\)Pr and used the BS2 energies. Subsequently, we removed the substrate (either Cys or MP) from the model and did a UB3LYP/BS2 single point calculation on the isolated substrate and protein with substrate removed. The difference in energies was then considered as the substrate binding energy (BDE\(_{\text{sub}}\)) for Cys in CDO and MP in MDO. Thereafter, the same was done for the product complexes to obtain the product binding energy (BDE\(_{\text{prod}}\)) for CysO\(_2\) in CDO and MPO\(_2\) in MDO. The obtained substrate and product binding energies are shown in Figure 11.

**Figure 10.** Electronic configurations of the structures along the mechanisms of substrate dioxygenation in CDO and MDO.

**Figure 11.** Substrate and product binding energies in the CDO and MDO models A and B. Values in kcal mol\(^{-1}\) calculated at UB3LYP/BS2.
As can be seen the product binding energies are very close for the CDO and MDO models at 150 and 149 kcal mol$^{-1}$, respectively. However, the substrate in MDO is much weaker bound than in CDO: BDE$^{\text{Sub,CDO}}$ = 143 kcal mol$^{-1}$ and BDE$^{\text{Sub,MDO}}$ = 128 kcal mol$^{-1}$. Therefore, the larger reaction exothermicity in MDO as compared to CDO is the result of weaker binding of the substrate in MDO that leads to the larger driving force to form the dioxygenated products.

Finally, we tested the substrate-bound scope of MDO by manually replacing 3-mercaptopropionate in $^{20}$Re$_{\text{B}}$ by either cysteinate ($^{20}$Re$_{\text{B,Cys}}$) or 2-mercaptosuccinic acid ($^{20}$Re$_{\text{B,MSA}}$). The optimized geometries of these two structures are shown in Figure 12 and compared to $^{20}$Re$_{\text{C}}$. Both structures converge to a local minimum with a protein fold close to what is seen for the wildtype system with 3-mercaptopropionate bound. As such, both cysteinate and 2-mercaptosuccinic acid fit the substrate-binding pocket and can bind to the metal center like 3-mercaptopropionate. Cysteinate is found to bind MDO as a bidentate ligand through the amine and sulfur groups, while the dangling carboxylate interacts with side chains Tyr$_{159}$, His$_{157}$ and Arg$_{168}$ through a tight hydrogen bonding network. A similar situation is seen for the R-2-mercaptosuccinic acid bound MDO structure, where one carboxylate binds iron and Arg$_{168}$ simultaneously, while the other carboxylate interacts with the superoxo group and His$_{157}$. Indeed an overlay of $^{20}$Re$_{\text{B}}$ and $^{20}$Re$_{\text{B,Cys}}$ (left of center in Figure 12) and an overlay of $^{20}$Re$_{\text{B}}$ and $^{20}$Re$_{\text{B,MSA}}$ (right of center in Figure 12) show only minor differences of the protein-atoms and orientations between each pair of structures. In all three structures the substrate binds in a bidentate fashion with an iron-sulfur distance of 2.562–2.631 Å, while the Fe–O$_2$ distance is 2.102 – 2.215 Å. However, there are some differences in the second coordination sphere of the three substrate-bound complexes that will affect their reactivities. As mentioned above the hydrogen bond from Tyr$_{159}$ to the distal oxygen atom of the iron(III)-superoxo group points it toward sulfur in $^{20}$Re$_{\text{B}}$ with a dihedral angle S–Fe–O–O of 16°. The MDO model with cysteinate bound has a similar small dihedral angle S–Fe–O–O of 6°; however, the carboxylate group of the substrate in the R-2-mercaptosuccinic acid bound MDO model pushes the superoxo group away leading to a dihedral angle of 86°. As a matter of fact the carboxylate group of R-2-mercaptosuccinic acid-bound MDO will shield the sulfur atom from attack from the iron(III)-superoxo group and hamper the sulfoxidation reaction. Indeed, no reactivity of MDO with 2-mercaptosuccinic acid was observed. [50] By contrast, the accessibility of the sulfur is not influenced in the cysteinate-bound MDO model and full dioxygenation is expected from this substrate-bound orientation. Experimental work on MDO with cysteinate obtained reasonable enzymatic turnover of the substrate. [29]

**Conclusion**

We present a computational study on the oxygen activation mechanisms of the thiol dioxygenases CDO and MDO using large active site cluster models of well over 300 atoms. The work shows that both enzymes should efficiently convert their natural substrates to dioxygenated products and for both enzymes the second oxygen atom transfer has the highest barriers with magnitudes of 14.2 and 15.8 kcal mol$^{-1}$. Both enzymes have an open-shell singlet spin iron(III)-superoxo reactant with substrate bound as a bidentate ligand in the equatorial plane, although in MDO the quintet spin state is within 1 kcal mol$^{-1}$. However, MDO binds substrate through two anionic bonds of the substrate carboxylate and thiolate groups, while CDO binds substrate to iron with one anionic group of thiolate and the neutral amide group. This gives the iron slightly different properties and changes the
spin-state orderings as well as the stability of the substrate bound complex. In particular, the second negatively charged ligand in the equatorial plane will lower the energy of the \( \alpha^*_{\text{Cu}} \) orbital and stabilize the quintet spin state dramatically and keeps the mechanism on this spin state surface. By contrast, in CDO with its 3-His ligand system there are close-lying singlet, triplet and quintet spin state surfaces along the mechanism and the reaction will be influenced by the equilibration between these spin states and the easy of spin state change. The work also identifies a strong hydrogen bonding interaction of a Tyr residue towards the superoxo group in MDO that is missing in CDO. It is proposed that this hydrogen bond directs the superoxo to the thiolate group for selective oxygen atom transfer. Our studies identify important interactions that position the substrate and oxidant in the catalytic cycles of CDO and MDO. Removal of these residues is expected to affect the selectivity and enzymatic turnover dramatically.

**Experimental Section**

**Model set-up.** We use cluster models of CDO and MDO structures that are based on the first and second coordinate spheres of these enzymes and take all local electrostatic interactions into consideration. Those types of models have been used extensively and the set-up and accuracy of these models have been reviewed previously.\(^{[21]}\) Our CDO model (model A, Scheme 2) is an enlarged cluster model with respect to previous DFT studies on the catalytic cycle of CDO\(^{[14]}\) and has a total of 308 atoms. We created the model based on the 2IC1 pdb file,\(^{[15]}\) using previously described procedures.\(^{[27]}\) The pdb is a substrate bound CDO monomer. We added hydrogen atoms under pH 7 conditions in Chimera,\(^{[28]}\) which resulted in all Arg and Lys residues being protonated and all Glu and Asp residues deprotonated. Furthermore, all histidine residues were singly protonated. We replaced the iron(II) center manually to iron(III)-superoxide. Our CDO model contained several protein chains that surround the cysteinate substrate binding pocket, namely the chains Tyr-Thr-Arg, Ser-His-His-Asp-His-Asp-Thr-Asn-His-Ser-Asn-Arg-Leu, His-Arg-Val-Leu and His-Arg-Val-Leu-Tyr. All side-chains that point away from the substrate binding pocket were replaced by Gly, i.e. the residues Thr92, Ile89, Asp90, Asn91, His92, Phe93, Arg94 and Leu95 while peptide termini were capped at the methyl group. In addition, we included the side-chains of residues Glu104 and Phe111 as butanoic acid and ethylbenzene, respectively. This model is charge neutral and was calculated in the quintet, triplet and singlet spin state.

The MDO model is a large active site model based on Chain C of the 4TLF pdb file,\(^{[16]}\) which is a tetramer. We manually inserted substrate 3-MPA in the binding position as described by Jameson et al using the Cartesian coordinates from their Supporting Information.\(^{[20]}\) We added hydrogen atoms under pH 7 conditions in Chimera,\(^{[28]}\) which resulted in all Arg and Lys residues being protonated and all Glu and Asp residues deprotonated. Furthermore, all histidine residues were singly protonated. We replaced the iron(II) center manually to iron(III)-superoxide, which gave us a model of a total of 324 atoms (Model B, Scheme 2). Thus, we selected several protein chains that surround the cysteamine substrate binding pocket, namely the chains Tyr-Gln-Gln-Glu, His-Arg-Asp-Asp-Arg-Glu-Val-Tyr-Val-Leu-Val-Leu, Asp-Val-Leu-Val-Glu-Val-Val-Leu-Tyr and Phe111-Asp113. All side-chains that point away from the substrate binding pocket were replaced by Gly, i.e. the residues Glu114, Asp115, Arg116, Val117, Tyr118 and Arg119 as aspartate and n-butylguanidine, respectively. This model is charge neutral and was calculated in the quintet, triplet and singlet spin state.

**Procedures.** Density Functional Theory (DFT) calculations were performed using the Gaussian-09 software package,\(^{[28]}\) and utilized the unrestricted B3LYP hybrid density functional method.\(^{[29]}\) Geometry optimizations, frequencies, intrinsic reaction coordinate scans and geometry scans were performed with an LANL2DZ basis set on iron with core potential and 6-31G* on the rest of the atoms: basis set BS1.\(^{[31]}\) No constraints were placed on any of the atoms and a comparison of the optimized geometries with the crystal structure coordinates showed little changes. To correct the energies single point calculations on the optimized geometries were done with the LACVP* basis set on iron with core potential and 6-311+G* on the rest of the atoms: basis set BS2. Solvent was included with the continuum polarized conductor model (CPCM) with a dielectric constant mimicking chlorobenzene.\(^{[32]}\)

To test the reproducibility of the calculations, we ran a series of test calculations using dispersion corrected B3LYP,\(^{[30,33]}\) but very little differences in optimized geometries and potential energy landscapes were seen.

**Acknowledgements**

The Computational Shared Facilities at the University of Manchester is acknowledged for providing CPU time.

**Keywords:** Nonheme iron enzymes • enzyme catalysis • inorganic reaction mechanisms • sulfonation • density functional theory

**References**


Insert text for Table of Contents here: Although cysteine and 3-mercaptopropionate dioxygenase bind and react similar substrates, there are key differences in the first- and second-coordination sphere that affect their catalytic reaction mechanism and reactivity as highlighted by large density functional theory studies on enzyme clusters.

Institute and/or researcher Twitter usernames: The University of Manchester (@OfficialUoM); Department of Chemical Engineering and Analytical Science (@UoMSciEng); Manchester Institute of Biotechnology (@UoMMIB)
Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Yeh, C-CG; Pierides, C; Jameson, GNL; de Visser, SP

Title:
Structure and Functional Differences of Cysteine and 3-Mercaptopropionate Dioxygenases: A Computational Study

Date:
2021-08-24

Citation:

Persistent Link:
http://hdl.handle.net/11343/298874