

Cleavage of a carbon–fluorine bond by an engineered cysteine dioxygenase

Jiasong Li¹, Wendell P. Griffith¹, Ian Davis¹, Inchul Shin¹, Jiangyun Wang², Fahui Li², Yifan Wang¹, Daniel J. Wherrett¹ ¹ and Aimin Liu¹ ^{1*}

Cysteine dioxygenase (CDO) plays an essential role in sulfur metabolism by regulating homeostatic levels of cysteine. Human CDO contains a post-translationally generated Cys93–Tyr157 cross-linked cofactor. Here, we investigated this Cys–Tyr cross-linking by incorporating unnatural tyrosines in place of Tyr157 via a genetic method. The catalytically active variants were obtained with a thioether bond between Cys93 and the halogen-substituted Tyr157, and we determined the crystal structures of both wild-type and engineered CDO variants in the purely uncross-linked form and with a mature cofactor. Along with mass spectrometry and ¹⁹F NMR, these data indicated that the enzyme could catalyze oxidative C–F or C–Cl bond cleavage, resulting in a substantial conformational change of both Cys93 and Tyr157 during cofactor assembly. These findings provide insights into the mechanism of Cys–Tyr cofactor biogenesis and may aid the development of bioinspired aromatic carbon–halogen bond activation.

Post-translational modification increases the number of possible molecular variations of proteins in living cells by several orders of magnitude and hence is known as ‘the chemistry of proteome diversifications’^{1,2}. Whereas reversible protein modifications play central roles in cellular regulation, unidirectional post-translational modifications generate novel cofactors to enhance or expand the catalytic repertoire of enzymes^{3–5}. Irreversible post-translational modifications have become a fundamental challenge for chemists in predicting the structures and functions of proteins. A protein-derived Cys–Tyr cofactor has recently been found in mammalian cysteine dioxygenase (CDO, EC 1.13.11.20)^{6,7}. Such a cofactor is only known in a few proteins^{8–12}. CDO is a nonheme iron enzyme that catalyzes the conversion of L-cysteine to cysteine sulfinic acid (CSA; Supplementary Fig. 1). The product, CSA, is ultimately catabolized to taurine and sulfate¹³. The Cys–Tyr cofactor contains a thioether (C–S) bond between the side chains of a cysteine residue (Cys93; human CDO numbering) and a tyrosine residue (Tyr157)¹⁴. The presence of such a Cys–Tyr cofactor boosts the catalytic efficiency of CDO by one order of magnitude¹⁵ owing to the concerted redox action of the metal ion and the protein-derived cofactor.

CDO occupies a central position in biological sulfur metabolism; its enzymatic activity is crucial for maintaining proper cysteine levels for protein synthesis and for initiating cysteine catabolism. It has been established that the metabolites derived from cysteine and the ratio of cysteine to sulfate and taurine exert a wide variety of physiological effects in cells, including energy balance, fat metabolism¹⁶, autoimmune and neurological conditions^{6,17}, oxidative stress¹⁸, O₂ sensing and hypoxia^{19–21}, and production of the signaling molecule H₂S²². Reduced activity of CDO results in elevated serum levels of neuroexcitatory cysteine, which has been associated with rheumatoid arthritis²³, breast cancer²⁴, and several neurological disease states, such as Alzheimer’s and Parkinson’s diseases²⁵. In mouse models, knockout of CDO results in increased activity of the alternative desulfhydration pathways, causing increased H₂S production and cytotoxicity²².

When recombinant CDO is isolated, cross-linked and uncross-linked forms are both present. Separation of the two forms can be achieved in the denatured state by SDS–PAGE; two distinct bands are observable, and the faster-migrating band is identified as the cross-linked form¹⁵. The cross-linking reaction takes place during catalysis as an autocatalytic reaction because of an uncoupled oxygen activation at the nonheme iron center of the enzyme, where O₂ activation is not linked to the oxidation of the substrate cysteine but rather to its own residues. It takes an unclarified number, likely hundreds to thousands, of turnovers to obtain a fully mature enzyme¹⁵. Depending on the substrate concentrations (L-cysteine and O₂), pH, and temperature, the time needed to reach the fully cross-linked form in solution varies. Once the Cys–Tyr cofactor is generated, the mature form of CDO performs the coupled dioxygenation reaction more efficiently with an increased capability to metabolize high levels of cysteine.

To date, understanding of the mechanism of Cys–Tyr cross-link biogenesis in CDO has stagnated because of several technical challenges in addition to the single-turnover nature discussed above. First, the uncross-linked form of active CDO, the starting material for studying cofactor biogenesis, is difficult to isolate, though it is observable in denatured gels. Thus, little spectroscopic characterization has been done, and no structural information is currently available for the uncross-linked form. Second, traditional site-directed mutagenesis approaches can support the catalytic importance of the cofactor, but mutation of Cys93 or Tyr157 leads to disruption of the formation of the Cys–Tyr cross-link and the inability of the protein variants to generate a mature enzyme active site, thus offering limited information^{15,26–28}.

In the present work, we surmounted these challenges by employing a genetic method for site-specific incorporation of unnatural amino acids²⁹ into human CDO (hCDO) at position 157 of the protein sequence. The resulting protein variants were characterized by enzymatic activity and biophysical techniques to explore the mechanism of cofactor biogenesis in CDO. An oxidative carbon–fluorine (C–F) or carbon–chlorine (C–Cl) bond cleavage was

¹Department of Chemistry, University of Texas at San Antonio, San Antonio, TX, USA. ²Institute of Biophysics, Chinese Academy of Sciences, Beijing, P.R. China. *e-mail: Feradical@utsa.edu

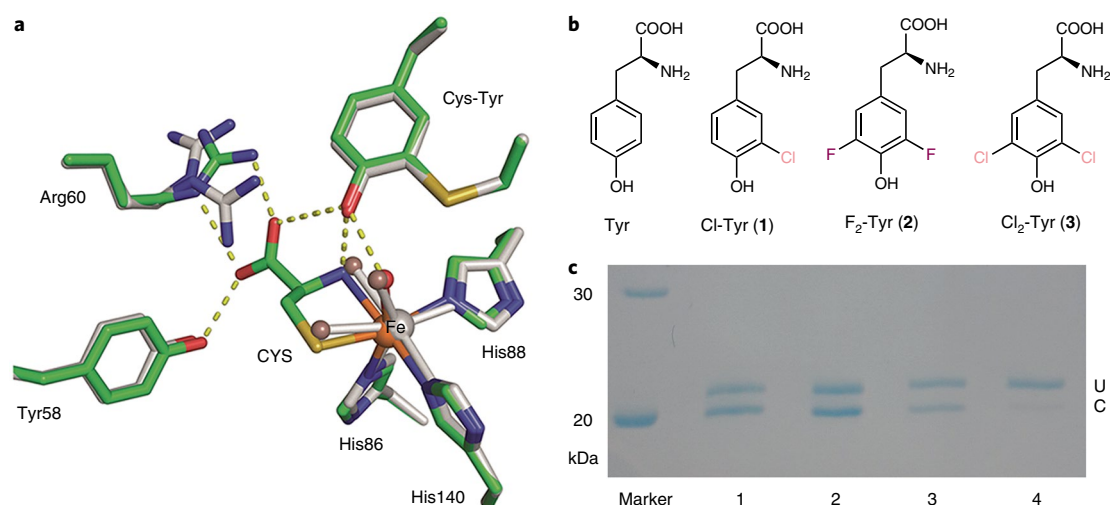


Fig. 1 | Crystal structures of human CDO and incorporation of unnatural amino acids into the catalytic active site tyrosine. **a**, Overlay of the active sites of the ligand-free (gray) and substrate (CYS)-bound complex (green) crystal structures of human CDO (PDB entries 6BGF and 6BGM). Arg60 presents two conformations in the ligand-free structure. The omit $F_o - F_c$ electron densities for the water ligands (WAT) and the L-cysteine are shown in Supplementary Fig. 2a,b. **b**, Native tyrosine and halogen-substituted unnatural tyrosine analogs used in this study; i.e., L-Tyr, 3-Cl-L-Tyr (Cl-Tyr, **1**), 3,5-F₂-L-Tyr (F₂-Tyr, **2**), and 3,5-Cl₂-L-Tyr (Cl₂-Tyr, **3**). **c**, SDS-PAGE shows two bands, with the slower moving band corresponding to the uncross-linked enzyme (U) and the faster moving band to the mature CDO with a cross-linked cofactor (C). The lanes from left to right of the molecular weight marker are as follows: (1) wild-type CDO; (2) Cl-Tyr157 CDO; (3) F₂-Tyr157 CDO; and (4) Cl₂-Tyr157 CDO. The full gel image is displayed in Supplementary Fig. 5b,c.

observed during cofactor formation in the dihalogen-substituted enzyme variants, indicating that the nonheme iron center generates an oxidant that is strong enough to oxidize the adjacent second coordination sphere residues for cross-linking whether a stronger or weaker bond is present or the chemical properties of the tyrosine are altered. When the engineered protein with a monochlorine-substituted tyrosine was exploited, the structural and MS analyses showed that steric hindrance is the next factor governing selectivity for generating a new C–S bond.

Results

Structures of ligand-free and substrate-bound human CDO. To investigate the formation and key role of the Cys93–Tyr157 cross-link, we determined the crystal structures of mature hCDO both in the ligand-free status and in complex with L-cysteine (Fig. 1a) refined to 2.20 and 1.80 Å resolution, respectively (Supplementary Table 1). Prior to this work, CDO research was primarily focused on the mouse and rat gene products (mouse and rat CDOs have identical protein sequences). Only one human CDO structure was reported at 2.7 Å resolution¹⁴, and it is ambiguous whether this is a ligand-bound structure³⁰. We determined that the ligand-free structure belongs to the $P6_5$ space group, with each asymmetric unit containing one CDO molecule. The electron density of the enzyme active site shows that the obtained structures are 100% cross-linked, with Cys93 and Tyr157 covalently connected via a thioether bond. The overall structure is analogous to that of the mouse and rat CDOs described previously⁷. Three water ligands were found to coordinate with the Fe ion in all the resting-state wild-type (WT) CDO structures (Supplementary Fig. 2) that we obtained at various resolutions, consistent with the published spectroscopic data from mouse and rat CDOs indicating that the iron ion exhibits a pseudo-octahedral coordination geometry^{7,31}. The first crystal structure, determined from Ni²⁺-reconstituted mouse CDO, shows three water ligands⁶, whereas the rat CDO structure and Mössbauer spectroscopy show a tetracoordinate iron center with only one or two water ligand(s)^{7,31,32}. The number of water ligands is relevant to the mechanistic understanding, because the first step of the catalytic cycle is substrate ligation to the iron ion. The crystal structure of

CDO in complex with L-cysteine (Fig. 1a) shows that the substrate L-cysteine displaces two water molecules, with both the amino and thiol groups coordinating to the ferrous ion, and the remaining water ligand is only weakly associated to the Fe ion at a distance of 2.6 Å and is hydrogen bonded with the hydroxyl group of the Cys–Tyr cofactor. The substrate L-cysteine forms a salt bridge with Arg60 and is hydrogen bonded to Tyr58 and the Cys–Tyr cofactor in addition to its bidentate coordination to the Fe(II) ion.

Genetic incorporation of unnatural amino acids. To investigate the biogenesis mechanism in CDO, we genetically substituted Tyr157 with three unnatural amino acids: 3-chloro-L-tyrosine (Cl-Tyr, **1**), 3,5-difluoro-L-tyrosine (F₂-Tyr, **2**), and 3,5-dichloro-L-tyrosine (Cl₂-Tyr, **3**) (Fig. 1b). The advantage of this approach is that it allows less perturbation to the enzyme active site structure than site-directed mutagenesis while specifically and chemically altering the amino acid residue of interest. The genetic incorporation of unnatural amino acids was site specific, and all six other tyrosine residues in hCDO were left unaltered. The CDO protein isolated was 100% modified at position 157 with specific tyrosine analogs. This efficient genetic incorporation of unnatural tyrosines into a specific tyrosine residue in a protein has been previously used for the study of other systems including a nonheme Fe enzyme, ribonucleotide reductase^{33–35}.

Of the unnatural tyrosine compounds available, F₂-Tyr is the most interesting for this study because fluorine is highly electronegative and small in size. Thus, we expect F₂-Tyr157 to mimic the native human enzyme with minimal structural perturbation while specifically attenuating the hydrogen bonding potential of the tyrosyl hydroxyl group because of the strong electron-withdrawing properties of fluorine. To selectively incorporate F₂-Tyr into hCDO, we used a mutant of a *Methanocaldococcus jannaschii* tyrosyl amber suppressor tRNA(*MjtRNA*_{Tyr}^{CUA})/F₂-TyrRS (Y32R, L65Y, H70G, F108N, Q109C, D158N, L162S) pair to target the TAG codon, as previously reported³⁶. We expressed the hCDO variant bearing F₂-Tyr at position 157 in *Escherichia coli* using pEVOL-F₂-TyrRS and PV16-hCDO157TAG plasmids. Likewise, we also genetically incorporated Cl-Tyr and Cl₂-Tyr into position 157 of hCDO by a similar method using pEVOL-Cl-TyrRS and pEVOL-Cl₂-TyrRS³⁷.

The ESI mass spectra of the WT, Cl-Tyr157, F₂-Tyr157, and Cl₂-Tyr157 variants under denaturing solution conditions showed an ensemble of charge states ranging from $[M+8H]^+8$ to $[M+33H]^+33$. Deconvolution of these mass spectra results in experimentally determined masses for these proteins of 22,913.20, 22,947.11, 22,949.16, and 22,982.07 Da for WT, Cl-Tyr157, F₂-Tyr157 and Cl₂-Tyr157 CDO, respectively (Supplementary Fig. 3). These measured mass values are consistent with the predicted masses of 22,913.21, 22,947.17, 22,949.19 and 22,982.13 Da, respectively, from protein sequence. These masses of Cl-Tyr157, F₂-Tyr157, and Cl₂-Tyr157 reflect a +34, +36 or +69 Da difference compared to WT CDO, as would be expected for the substitution of one or more halogen atoms for hydrogen at Tyr157.

The unnatural amino acids' function in human CDO. The fluorine- and chlorine-bearing CDO variants were catalytically active, with the as-isolated F₂-Tyr157 variant having about 10% activity and the Cl/Cl₂-Tyr157 variants having 2% activity compared to that observed for WT CDO (Supplementary Fig. 4 and Supplementary Table 2). The question motivating the next step of this study is whether or not the cross-linked Cys-Tyr cofactor could be formed in the F₂-Tyr157 and Cl₂-Tyr157 CDO variants. On the basis of bond dissociation energies, the formation of the C-S bond needed for cofactor assembly was expected to be hindered by the fluorine (but not chlorine) substitution. Because the C-F bond is often thought to be one of the strongest single bonds (second only to the C-Si bond) between formally neutral atoms³⁸, one may expect that F₂-Tyr157 substitution would greatly hinder, or prevent, the formation of the thioether bond for cofactor assembly. Oxidative formation of the cross-link with F₂-Tyr157 would require breaking a C-F bond and an S-H bond and forming a new C-S bond. Without knowing the mechanism of the cofactor formation, it was difficult to predict a priori whether or not the cross-linked Cys-Tyr cofactor would still be formed in the halogen-substituted CDO variants.

The as-isolated WT, Cl-Tyr, F₂-Tyr157, and Cl₂-Tyr157 CDO variants all exhibited two bands by SDS-PAGE, indicating that the substitutions were unable to abrogate cross-link formation (Fig. 1c). On the basis of gel analysis, we estimated that 55%, 54%, 41%, and 16% cross-links were present in these as-isolated CDO samples, respectively. Further processing of the as-isolated halogen-substituted CDO variants with L-cysteine and O₂ led to a mainly cross-linked enzyme concomitant with the weakening of the slower-migrating band (Supplementary Fig. 5). Upon closer analysis of the ESI mass spectra, we found that F₂-Tyr157 CDO in the +10 charge state and below contained an additional peak corresponding to 20 Da lower mass (Supplementary Fig. 6), which fits the expected mass of the cross-linked form of F₂-Tyr157 CDO. After cross-link formation, the protein is more resistant to acid and organic solvent denaturation, and thus is more compact, so it carries less protonic charge and appears at lower charge states.

Oxidative C-F bond cleavage during cofactor biogenesis. The gel analysis and the ESI data suggested that the Cys-Tyr cofactor was formed in the CDO variants with unnatural amino acids. Next, we considered whether they do so by breaking the C-F or C-Cl bond or by forming a differently configured Cys-Tyr, i.e., linking the thiol group to C2 or C6. To rigorously determine the nature and position of the cross-link in F₂-Tyr157 CDO, an MS/MS analysis was conducted (Fig. 2a).

After conversion to the all-cross-link form, the WT and F₂-Tyr157 CDO proteins were digested by three enzymes: trypsin, GluC, and chymotrypsin. As expected, the cross-linked peptide from WT hCDO was detected at m/z 3,226.4 (Fig. 2b). In samples of F₂-Tyr157 CDO, no peak was detected around m/z 3,262 corresponding to a cross-linked peptide containing two fluorine atoms on Tyr157. Instead, a peak at m/z 3,244.5 was found, which is

consistent with the cross-linked peptide containing only one fluorine atom (Fig. 2b).

To confirm the location of the cross-link and displacement of a fluorine atom upon cross-linking, tandem MS using LIFT technology³⁹ on a MALDI-TOF/TOF mass spectrometer was conducted. An illustration of the cross-linked peptide indicating the major fragment ions detected is depicted in Fig. 2a. Based on this schematic, any N-terminal fragment (b-ion, green colored) before His92 of peptide A (residues Gly80 through His92) and any C-terminal ion (y-ion, orange colored) of peptide B (residues Ser158 through Phe167) after Ser158 would have the same mass (or m/z value) in the cross-linked WT and F₂-Tyr157 CDO peptides. The MS data showed that this is indeed the case. Our data also showed an increase of +18 Da in F₂-Tyr157 relative to the WT peptide for any fragment containing Tyr157 (Fig. 2c). These results unambiguously confirmed the cross-link between Cys93 and Tyr157 and established that Tyr157 contains one, and only one, fluorine atom in the mature form, as the other is displaced during the formation of the cross-link. These MS analyses further suggested that the cofactor formed in F₂-Tyr157 CDO is analogous to the Cys-Tyr cofactor found in the native protein. The fluorine atom released during oxidative conversion was detected by ¹⁹F NMR (Fig. 2d). Because fluorine release should be stoichiometric with the protein, more than 10,000 scans over the course of 20 h on a 500 MHz NMR spectrometer were collected. Spiking in aqueous KF confirmed that the signal for the released fluorine is F⁻ (Fig. 2d). Such a fluoride signal is not present in the ¹⁹F NMR spectrum of F₂-Tyr157 CDO (Supplementary Fig. 7).

Achieving cofactor biogenesis in crystallo. We determined the crystal structures of F₂-Tyr157 hCDO and Cl₂-Tyr157 hCDO (Fig. 3, Supplementary Fig. 8 and Supplementary Table 1) in the absence of and in complex with L-cysteine. When F₂-Tyr157 crystals were grown anaerobically, only the uncross-linked form crystallizes, and we thus obtained the long-sought, 100% uncross-linked structure of CDO, refined to 2.40 Å resolution (Fig. 3a). Cys93 exhibits two alternative conformations in the uncross-linked CDO structure, with occupancy in a 73:27 ratio, and the sulfhydryl group in the major conformation points away from the iron ion. The catalytic Fe(II) ion is ligated by the Nε atoms of His86, His88, and His140 and three water molecules. The substrate-bound, uncross-linked crystal structure was refined to 2.10 Å resolution (Fig. 3b). A substantial rotation of F₂-Tyr157 is noted (Supplementary Fig. 9a). As shown in the wild-type structure, three water ligands are bound to the Fe ion in all ligand-free structures of the CDO variants, and only one water ligand remains associated with the iron ion in the substrate-bound structure (Supplementary Fig. 2). The B-factor of the water ligand changes from 15.5 in the ligand-free structure to 37.8 in the substrate-bound structure, indicating that the water ligand becomes flexible when L-cysteine binds to the metal ion.

We also obtained a crystal structure with both cross-linked and uncross-linked forms in the same crystal structure from the as-isolated, ligand-free F₂-Tyr157 hCDO crystalized under aerobic conditions and refined it to 1.80 Å resolution. The cross-linked and uncross-linked forms of the cofactor were present in a 53:47 ratio in this structure based on their respective occupancy (Supplementary Fig. 8a,b). The cross-linked conformation of Cys93 showed a connecting density to F₂-Tyr157, indicating that the cross-linked cofactor is present. Notably, Cys93 also exhibited two alternative conformations in this structure. In the uncross-linked conformation, the sulfhydryl group of Cys93 points away from the iron center, whereas in the cross-linked form the sulfur atom of Cys93 turns toward the iron center by nearly 3 Å. The F₂-Tyr157 ring also rotated about 46 degrees along the axis through Cβ and the OH group, with a nearly 1-Å shift toward Cys93, allowing C-S bond formation (Supplementary Fig. 9b). We also found a similar phenomenon in

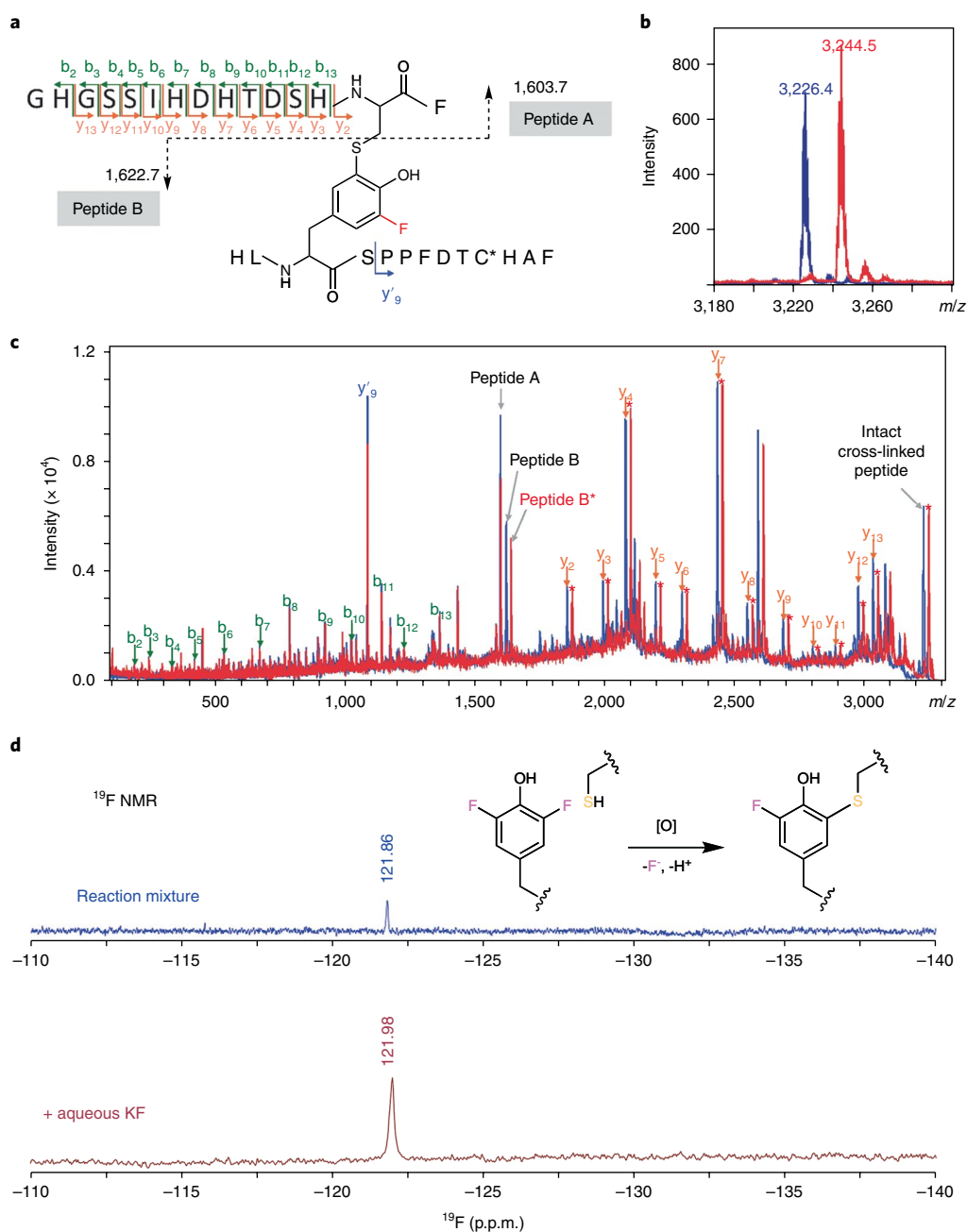


Fig. 2 | MS/MS spectra of cross-linked peptides of WT and F_2 -Tyr157 CDO proteins and ^{19}F NMR detection of leaving fluoride. **a, Proposed cross-link of peptides Gly80–Phe94 and His155–Phe167 through Cys93 and F_2 -Tyr157. **b**, MS spectrum of the cross-linked peptide Gly80–Phe94/His155–Phe167, including one carboxyamidomethyl cysteine at m/z 3,226.4 for WT and one at m/z 3,244.5 for F_2 -Tyr157 CDO. **c**, Collision-induced dissociation (CID)–MS/MS spectrum of the precursor ions at m/z 3,226.4 and m/z 3,244.5. **d**, In situ detection of fluoride during cofactor biogenesis in F_2 -Tyr157 CDO. ^{19}F NMR spectra of the fluoride ion in solution from the cross-linking reaction of F_2 -Tyr157 CDO after acid precipitation of the protein. Top, ^{19}F NMR spectrum after the conversion of F_2 -Tyr157 CDO to the all-cross-linked form after 12,480 transients. Bottom, ^{19}F NMR spectrum after spiking in aqueous KF to confirm that the signal is F^- after 128 transients. Both spectra are referenced to internal trifluoroacetic acid. MS experiments were repeated one time, and ^{19}F NMR experiments were repeated two times with similar results. The F_2 -Tyr157 CDO was also recorded once by ^{19}F NMR before the autocatalytic processing reaction (a negative control), and the resulting spectrum is shown in Supplementary Fig. 7. The asterisks in **a** and **c** indicate fragment peaks in the crosslinked peptides from F_2 -Tyr157 CDO that are +18 relative to the wild-type fragments containing the tyrosine-157 and only one fluorine.**

the crystal structure from the as-isolated Cl_2 -Tyr157 CDO crystalized under aerobic conditions (Supplementary Figs. 8c,d and 9c).

Taking advantage of the uncross-linked crystals, we subsequently investigated the possible uncoupled reaction in crystallo that leads to the oxidation of the second coordination sphere residues and the formation of the cofactor. After reacting the F_2 -Tyr157 variant crystals with L-cysteine (100 mM) under aerobic conditions

for 5 h, the single crystals of F_2 -Tyr157 hCDO showed a fully mature, cysteine-bound active site with 100% cross-link between Cys93 and Tyr157 (Fig. 3c), and the structure was refined to 1.95 Å resolution (Supplementary Table 1). An overlay of the uncross-linked F_2 -Tyr157 CDO structures with and without a substrate cysteine bound at the active site is shown in Fig. 3d, and an overlay of the cross-linked and uncross-linked F_2 -Tyr157 CDO structures

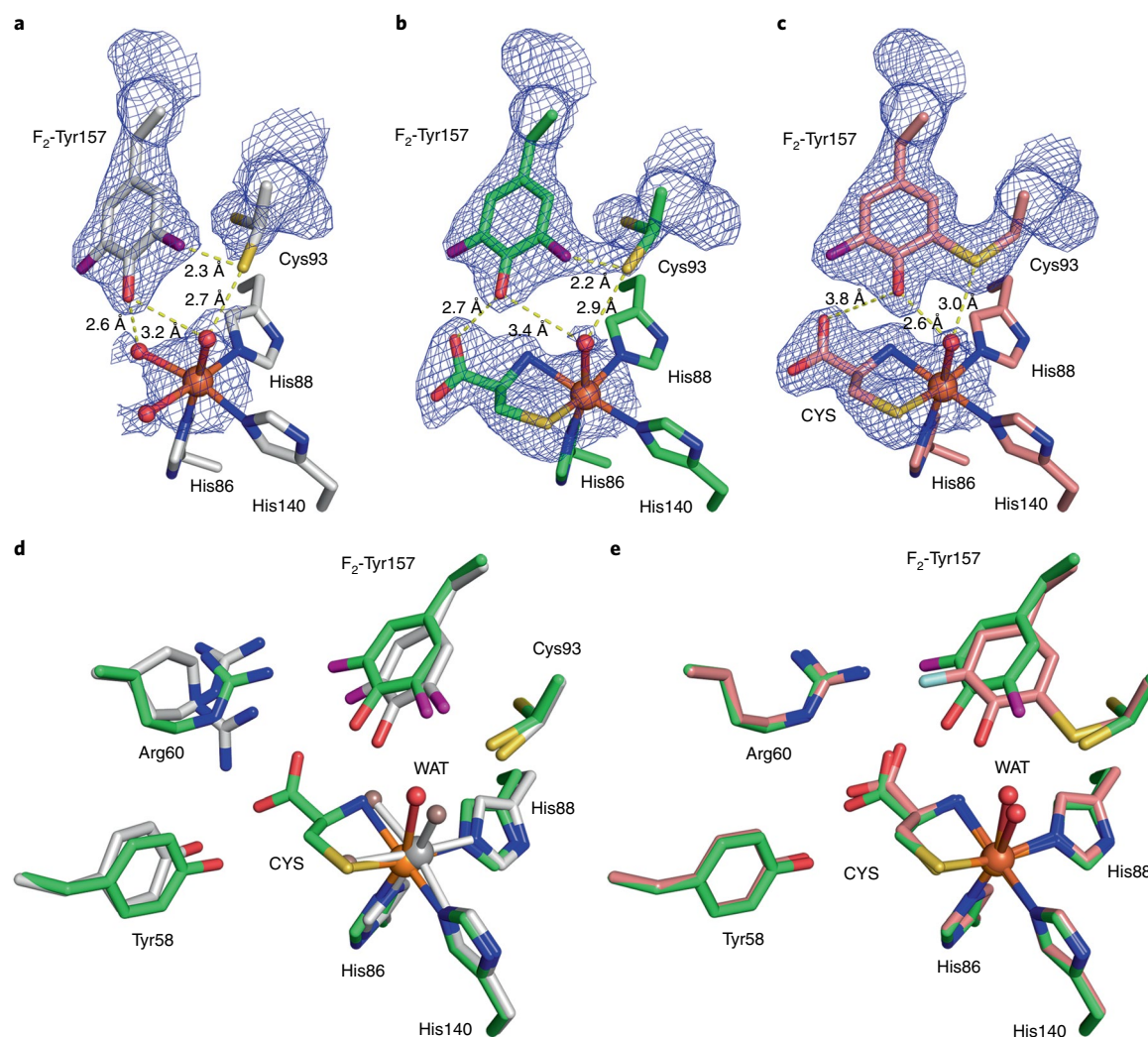


Fig. 3 | Crystal structures of F₂-Tyr157 CDO. $2F_o - F_c$ electron density was contoured at $1.2 \sigma_{\text{rms}}$ with potential hydrogen bond interactions (broken lines) shown. **a**, The active site of the 100% uncross-linked F₂-Tyr157 CDO structure in the substrate-free form. **b**, The active site of the 100% uncross-linked F₂-Tyr157 CDO structure in the substrate-bound form. **c**, The active site of the mature F₂-Tyr157 CDO structure in the substrate-bound form. **d**, Overlaid active site of the 100% uncross-linked F₂-Tyr157 CDO structure in the substrate-free (2.40 Å resolution, gray) and substrate-bound (2.10 Å resolution, green) forms. **e**, Superimposition of the substrate-bound mature F₂-Tyr CDO (1.95 Å resolution, pink) and 100% uncross-linked structures (green). The L-cysteine substrate is labeled as CYS. The omit $F_o - F_c$ electron densities for the water ligands (WAT) and the L-cysteine are shown in Supplementary Fig. 2.

in complex with L-cysteine is shown in Fig. 3e and Supplementary Fig. 9d. A substantial rotation of F₂-Tyr157 is noted between the uncross-linked and mature forms of the enzyme. Cys93 shows only one conformation in the mature form⁴⁰ but two conformations in each of the uncross-linked structures. Likewise, in crystallo cofactor maturation was also observed in Cl₂-Tyr157 CDO protein crystals.

To discern which bond would be cleaved when both C–H and C–Cl bonds are available, we analyzed the monochlorine-substituted Cl-Tyr157 variant by MS and determined its crystal structures from the purely uncross-linked and the fully cross-linked forms (Supplementary Table 1). In the as-isolated form, the uncross-linked and cross-linked portions are present in a similar occupancy. Figure 4a shows the overlay of the active site of the ligand-free Cl-Tyr157 CDO structures in the cross-linked and uncross-linked forms. The most revealing structural information for mechanistic understanding is the uncross-linked form. The crystal structure shows in the uncross-linked form that equal populations of the C–H and C–Cl bonds of the Tyr157 are positioned toward Cys93 and the iron ion, indicating that the C–H and C–Cl bonds have similar opportunity for oxidative cleavage. In the fully processed form with

a mature cofactor and bound L-cysteine substrate at the nonheme Fe center, only a single conformation of the cofactor is present with 100% occupancy of the chlorine (Fig. 4b). As corroborated by the MS analysis (Supplementary Fig. 10), the C–H bond is cleaved but the C–Cl bond is not. Therefore, we conclude that the large size of chlorine makes the C–Cl bond less accessible for oxidation.

Discussion

The success in obtaining a purely uncross-linked crystal structure of CDO is an important advancement toward unfolding the cofactor biogenesis mechanism. Previously, the uncross-linked CDO structure was acquired from the C93A mutant²⁸. The key difference is that the uncross-linked structure we report is a fully capable protein for cofactor assembly, whereas the mutant is a permanently disabled variant in terms of cofactor biogenesis. In both the substrate-free and substrate-bound C93A mutant structures, a buffer-derived chloride ion is ligated to the iron ion, but the chloride ligation is not seen in our structures. Our purely uncross-linked structure revealed important details for mechanistic understanding that were not possible to detect in the mutant structures. Cys93 shows two

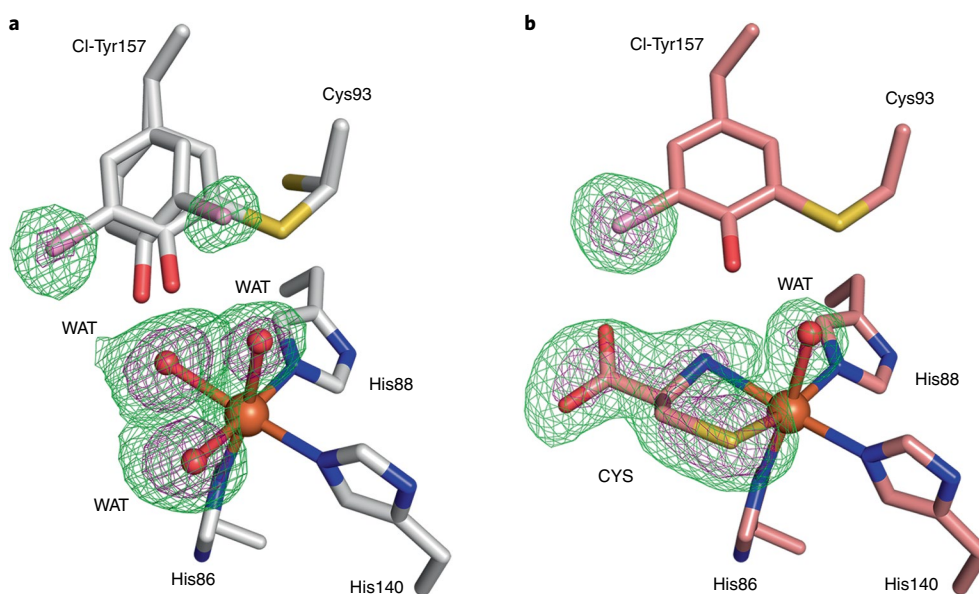


Fig. 4 | Crystal structures of Cl-Tyr157 CDO. The omit $F_o - F_c$ electron densities (green) of the WT CDO and mutants contoured at 3σ (green) and 6σ (purple), respectively. **a**, The active site of the Cl-Tyr157 CDO structure in the cross-linked and uncross-linked forms in the absence of the substrate. **b**, The active site of the substrate-bound Cl-Tyr157 CDO structure in the cross-linked form. CYS represents the L-cysteine substrate bound at the nonheme iron center.

conformations before the cross-link reaction, and the sulfur atom of Cys93 undergoes a nearly 3-Å rotation toward the iron center during autocatalytic cofactor biogenesis. The findings presented in this work may be useful in considering the mechanism of the Cys-Tyr biogenesis.

The ability of CDO to catalyze aromatic C-F bond cleavage is not a predictable result, as the energy of an aliphatic C-F bond (485 kJ mol^{-1}) is substantially higher than that of the corresponding aliphatic C-H bond (411 kJ mol^{-1})⁴¹. Because the bond energy of an aliphatic C-Cl bond (i.e., 327 kJ mol^{-1}) is lower than that of a C-H bond, the cleavage of the C-Cl bond in Cl_2 -Tyr157 CDO is no surprise. However, it is an interesting observation that the C-H bond is preferentially cleaved in mono-Cl-Tyr157 CDO even though both C-H and C-Cl are equally populated toward Cys93 and the iron center.

It should be noted that the C-F bond in this case may not be as strong as those in an aliphatic molecule. However, the aromatic C-F bond is expected to be stronger than that of the corresponding C-H bond. When a hydroxyl group is attached to the C4 ring carbon of a tyrosine, it activates the C3 and C5 positions. We propose that the C-F bond may be cleaved during a putative radical step in which Tyr157 is oxidized, either directly by the iron-bound oxidant or by a Cys93 radical (Supplementary Fig. 11). Once a tyrosyl radical is formed, it is well-known that the radical is delocalized to specific positions⁴, including the ring C3 and C5. Thus, the C-H or C-F bond activation in Tyr157 would be easier than generally thought for a C-H or C-F bond in hydrocarbons. This is especially true after a tyrosyl radical is formed during cofactor biogenesis.

The cross-link formation is apparently less efficient for the Cl_2 -Tyr157 derivative than the F_2 -Tyr157 variant and the wild-type protein. Because a C-Cl bond is less durable than a C-F bond, the less-efficient cross-linking of Cl_2 -Tyr157 CDO suggests that there may be an additional factor governing bond cleavage. We hypothesize that the size of a chlorine atom adversely affects access of the C-Cl bond. The cross-link formation in mono-substituted Cl-Tyr157 CDO is as efficient as that in the native enzyme, and in the uncross-linked structure the C-H and C-Cl

of Cl-Tyr157 are equally populated toward Cys93 and the Fe ion. However, it is the stronger C-H bond that is cleaved and the weaker C-Cl bond that is retained in the mature Cl-Tyr157 CDO variant. These results suggest that the size of chlorine makes it more difficult for Cl-Tyr157 to position closer to a putative Cys93 radical or an iron-bound oxidant relative to the fluorinated tyrosine derivative. Because of the small size of fluorine, there is no tRNA synthetase available for producing monofluorine-substituted tyrosine derivative. The heterogeneous substitution of chlorine and fluorine at the ring 3 and 5 positions of tyrosine (3-chloro-5-fluoro tyrosine) is also inaccessible by the genetic incorporation approach. Based on the current results, we envision that the C-F bond would more likely be cleaved in a hypothetical 3-chloro-5-fluoro-Tyr157 CDO variant.

The data presented in this work collectively indicate an O_2 -dependent carbon-halogen bond cleavage by a protein-bound iron center under mild, physiologically relevant conditions. Even though the C-F bond is one of the strongest covalent bonds in organic chemistry, and stronger than the corresponding C-H bonds, CDO is able to cleave a C-F bond for the purpose of cofactor biogenesis. The cleavage of an aromatic C-F bond by nitrogen-ligated nonheme iron model complexes has been described in two recent papers^{42,43}. However, to our knowledge, the observation described here for CDO is the first example of oxidative C-F bond cleavage by a protein-bound iron center. The activation and functionalization of C-F bonds is an important area of inorganic, organic, and medicinal chemistry. In biological systems, the only C-F bond cleavage previously known was the oxygen-independent reactions catalyzed by fluoroacetate dehalogenase and BzCoA reductase^{44,45}. This unexpected finding during the study of CDO cofactor biogenesis reveals new insights into the power of the nonheme iron ion in a dioxygenase and sheds light onto the interplay between the catalytic metal ion and the cofactor of CDO. It further increases our expectation of the oxidizing power of non-heme Fe enzymes. The oxidative C-F bond cleavage described in this study broadens the range of fluorine chemistry and further extends potential industrial applications of natural or engineered Fe-dependent proteins.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at <https://doi.org/10.1038/s41589-018-0085-5>.

Received: 30 November 2017; Accepted: 4 May 2018;

Published online: 25 June 2018

References

- Walsh, C. T. *Posttranslational Modification of Proteins: Expanding Nature's Inventory*. (Roberts & Company Publishers, Greenwood Village, CO, 2006).
- Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J. Jr. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem. Int. Ed. Engl.* **44**, 7342–7372 (2005).
- Klinman, J. P. & Bonnot, F. Intrigues and intricacies of the biosynthetic pathways for the enzymatic quinocofactors: PQQ, TTQ, CTQ, TPQ, and LTQ. *Chem. Rev.* **114**, 4343–4365 (2014).
- Stubbe, J. & van Der Donk, W. A. Protein radicals in enzyme catalysis. *Chem. Rev.* **98**, 705–762 (1998).
- Krebs, C., Bollinger, J. M. Jr. & Booker, S. J. Cyanobacterial alkane biosynthesis further expands the catalytic repertoire of the ferritin-like 'di-iron-carboxylate' proteins. *Curr. Opin. Chem. Biol.* **15**, 291–303 (2011).
- McCoy, J. G. et al. Structure and mechanism of mouse cysteine dioxygenase. *Proc. Natl. Acad. Sci. USA* **103**, 3084–3089 (2006).
- Simmons, C. R. et al. Crystal structure of mammalian cysteine dioxygenase. A novel mononuclear iron center for cysteine thiol oxidation. *J. Biol. Chem.* **281**, 18723–18733 (2006).
- Whittaker, J. W. Free radical catalysis by galactose oxidase. *Chem. Rev.* **103**, 2347–2363 (2003).
- Cowley, R. E. et al. Structure of the reduced copper active site in preprocessed galactose oxidase: ligand tuning for one-electron O₂ activation in cofactor biogenesis. *J. Am. Chem. Soc.* **138**, 13219–13229 (2016).
- Schnell, R., Sandalova, T., Hellman, U., Lindqvist, Y. & Schneider, G. Siroheme- and [Fe₄-S₄]-dependent NirA from *Mycobacterium tuberculosis* is a sulfite reductase with a covalent Cys-Tyr bond in the active site. *J. Biol. Chem.* **280**, 27319–27328 (2005).
- Polyakov, K. M. et al. High-resolution structural analysis of a novel octaheme cytochrome c nitrite reductase from the haloalkaliphilic bacterium *Thioalkalivibrio nitratireducens*. *J. Mol. Biol.* **389**, 846–862 (2009).
- Hromada, S. E. et al. Protein oxidation involved in Cys-Tyr post-translational modification. *J. Inorg. Biochem.* **176**, 168–174 (2017).
- Stipanuk, M. H., Ueki, I., Dominy, J. E. Jr., Simmons, C. R. & Hirschberger, L. L. Cysteine dioxygenase: a robust system for regulation of cellular cysteine levels. *Amino Acids* **37**, 55–63 (2009).
- Ye, S. et al. An insight into the mechanism of human cysteine dioxygenase. Key roles of the thioether-bonded tyrosine-cysteine cofactor. *J. Biol. Chem.* **282**, 3391–3402 (2007).
- Dominy, J. E. Jr. et al. Synthesis of amino acid cofactor in cysteine dioxygenase is regulated by substrate and represents a novel post-translational regulation of activity. *J. Biol. Chem.* **283**, 12188–12201 (2008).
- Niewiadomski, J. et al. Effects of a block in cysteine catabolism on energy balance and fat metabolism in mice. *Ann. NY Acad. Sci.* **1363**, 99–115 (2016).
- Gordon, C., Bradley, H., Waring, R. H. & Emery, P. Abnormal sulphur oxidation in systemic lupus erythematosus. *Lancet* **339**, 25–26 (1992).
- Kwon, D. Y. et al. Impaired sulfur-amino acid metabolism and oxidative stress in nonalcoholic fatty liver are alleviated by betaine supplementation in rats. *J. Nutr.* **139**, 63–68 (2009).
- Olson, K. R. et al. Thiosulfate: a readily accessible source of hydrogen sulfide in oxygen sensing. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **305**, R592–R603 (2013).
- Weits, D. A. et al. Plant cysteine oxidases control the oxygen-dependent branch of the N-end-rule pathway. *Nat. Commun.* **5**, 3425 (2014).
- White, M. D. et al. Plant cysteine oxidases are dioxygenases that directly enable arginyl transferase-catalysed arginylation of N-end rule targets. *Nat. Commun.* **8**, 14690 (2017).
- Ueki, I. et al. Knockout of the murine cysteine dioxygenase gene results in severe impairment in ability to synthesize taurine and an increased catabolism of cysteine to hydrogen sulfide. *Am. J. Physiol. Endocrinol. Metab.* **301**, E668–E684 (2011).
- Bradley, H. et al. Sulfate metabolism is abnormal in patients with rheumatoid arthritis Confirmation by in vivo biochemical findings. *J. Rheumatol.* **21**, 1192–1196 (1994).
- Jeschke, J. et al. Frequent inactivation of cysteine dioxygenase type 1 contributes to survival of breast cancer cells and resistance to anthracyclines. *Clin. Cancer Res.* **19**, 3201–3211 (2013).
- Heafield, M. T. et al. Plasma cysteine and sulphate levels in patients with motor neurone, Parkinson's and Alzheimer's disease. *Neurosci. Lett.* **110**, 216–220 (1990).
- Li, W., Blaesi, E. J., Pecore, M. D., Crowell, J. K. & Pierce, B. S. Second-sphere interactions between the C93-Y157 cross-link and the substrate-bound Fe site influence the O₂ coupling efficiency in mouse cysteine dioxygenase. *Biochemistry* **52**, 9104–9119 (2013).
- Davies, C. G., Fellner, M., Tchesnokov, E. P., Wilbanks, S. M. & Jameson, G. N. The Cys-Tyr cross-link of cysteine dioxygenase changes the optimal pH of the reaction without a structural change. *Biochemistry* **53**, 7961–7968 (2014).
- Driggers, C. M. et al. Structure-based insights into the role of the Cys-Tyr crosslink and inhibitor recognition by mammalian cysteine dioxygenase. *J. Mol. Biol.* **428**, 3999–4012 (2016).
- Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of *Escherichia coli*. *Science* **292**, 498–500 (2001).
- Driggers, C. M. et al. Cysteine dioxygenase structures from pH 4 to 9: consistent cys-persulfenate formation at intermediate pH and a Cys-bound enzyme at higher pH. *J. Mol. Biol.* **425**, 3121–3136 (2013).
- Souness, R. J. et al. Mechanistic implications of persulfenate and persulfide binding in the active site of cysteine dioxygenase. *Biochemistry* **52**, 7606–7617 (2013).
- Tchesnokov, E. P. et al. An iron-oxygen intermediate formed during the catalytic cycle of cysteine dioxygenase. *Chem. Commun. (Camb.)* **52**, 8814–8817 (2016).
- Oyala, P. H. et al. Biophysical characterization of fluorotyrosine probes site-specifically incorporated into enzymes: *E. coli* ribonucleotide reductase as an example. *J. Am. Chem. Soc.* **138**, 7951–7964 (2016).
- Ravichandran, K. R. et al. Formal reduction potentials of difluorotyrosine and trifluorotyrosine protein residues: defining the thermodynamics of multistep radical transfer. *J. Am. Chem. Soc.* **139**, 2994–3004 (2017).
- Minnihan, E. C., Young, D. D., Schultz, P. G. & Stubbe, J. Incorporation of fluorotyrosines into ribonucleotide reductase using an evolved, polyspecific aminoacyl-tRNA synthetase. *J. Am. Chem. Soc.* **133**, 15942–15945 (2011).
- Li, F. et al. A genetically encoded ¹⁹F NMR probe for tyrosine phosphorylation. *Angew. Chem. Int. Ed. Engl.* **52**, 3958–3962 (2013).
- Liu, X. et al. Significant expansion of fluorescent protein sensing ability through the genetic incorporation of superior photo-induced electron-transfer quenchers. *J. Am. Chem. Soc.* **136**, 13094–13097 (2014).
- O'Hagan, D. Understanding organofluorine chemistry. An introduction to the C-F bond. *Chem. Soc. Rev.* **37**, 308–319 (2008).
- Suckau, D. et al. A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. *Anal. Bioanal. Chem.* **376**, 952–965 (2003).
- Fellner, M., Alo, S., Tchesnokov, E. P., Wilbanks, S. M. & Jameson, G. N. Substrate and pH-dependent kinetic profile of 3-mercaptopropionate dioxygenase from *Pseudomonas aeruginosa*. *Biochemistry* **55**, 1362–1371 (2016).
- Cottrell, T. L. *The Strengths of Chemical Bonds*. 2nd edn. (Butterworths Scientific, 1958).
- Sahu, S. et al. Aromatic C-F hydroxylation by nonheme iron(IV)-oxo complexes: Structural, spectroscopic, and mechanistic investigations. *J. Am. Chem. Soc.* **138**, 12791–12802 (2016).
- Sahu, S. et al. Direct observation of a nonheme iron(IV)-oxo complex that mediates aromatic C-F hydroxylation. *J. Am. Chem. Soc.* **136**, 13542–13545 (2014).
- Chan, P. W. Y., Yakunin, A. F., Edwards, E. A. & Pai, E. F. Mapping the reaction coordinates of enzymatic defluorination. *J. Am. Chem. Soc.* **133**, 7461–7468 (2011).
- Tiedt, O. et al. ATP-dependent C-F bond cleavage allows the complete degradation of 4-fluoroaromatics without oxygen. *MBio* **7**, e00990–16 (2016).

Acknowledgements

The work is supported in part by the National Institutes of Health grants GM107529, GM108988, and MH107985, the National Science Foundation grant CHE-1623856, and the Litcher Brown Distinguished Chair Endowment fund (to A.L.). J.W. acknowledges the support of the National Science Foundation of China grants (91527302, 31370016, and U1532150). The mass spectrometry facility is sponsored by the National Institutes of Health grant G12MD007591. The MALDI-TOF and NMR spectrometers are shared instruments sponsored by the National Science Foundation under the award numbers #1126708 and 1625963, respectively. X-ray synchrotron data were collected at the beamlines of the Advanced Photon Source Section 19, Structural Biology Center user program GUP-48198, Argonne National Laboratory and at the beamline BL9-2 of the Stanford Synchrotron Radiation Lightsource (SSRL) under the user program #5B14, SLAC National Accelerator Laboratory. The beamline staff scientists are acknowledged for the assistance of the remote data collections. The Advanced Photon Source is a US Department of Energy, Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. SSRL is supported by the US Department of Energy, Office of Science, Office of Basic

Energy Sciences under Contract No. DE-AC02-76SF00515 and by the National Institutes of Health (P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of National Institutes of Health or National Science Foundation.

Author contributions

Genetic incorporation of unnatural amino acids was performed by J.L. (cloning, protein expression and purification, and enzyme assays). J.W. and E.L. provided TyrRS. W.P.G. and J.L. conducted mass spectrometry analyses. J.L. obtained all protein crystals, collected X-ray diffraction data, and interpreted and refined the structural data together with I.S. The mechanistic models were proposed and refined by J.L., I.D., Y.W., and A.L. Y.W. participated in the unnatural amino acid production and isolation by an enzymatic method. D.J.W. performed the ^{19}F NMR analysis. A.L. conceived the research and wrote

the manuscript together with J.L. All authors contributed to data analysis and to the writing and editing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41589-018-0085-5>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.L.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

General methods. Primers were synthesized by Integrated DNA Technologies. Reagents were purchased from Sigma-Aldrich, New England BioLabs (NEB), and Thermo, Inc. DNA manipulations in *Escherichia coli* were carried out according to standard procedures. Ampicillin (100 µg/mL), and chloramphenicol (30 µg/mL) were used as antibiotics for selection of recombinant strains.

Synthesis of 3,5-difluorotyrosine. To synthesize F₂-Tyr, we transformed 2,6-difluorophenol to F₂-Tyr by using *Citrobacter freundii* (ATCC8090) tyrosine phenol lyase (TPL) following established methods⁴⁶. One liter solution containing 30 mM ammonium acetate, 60 mM sodium pyruvate, 5 mM β-mercaptoethanol, 40 µM PLP, and 10 mg of TPL enzyme, pH 8.0, was stirred in the dark for 3 d at room temperature (20–22 °C). The aqueous phase was then concentrated and purified by a C-18 column from 10% to 90% methanol, 0.1% trifluoroacetic acid (v/v) in water. To test for the presence of the unnatural amino acid, a ninhydrin solution consisting of 0.2% (w/v) ninhydrin, 95% (v/v) *N*-butanol, 0.5% (v/v) acetic acid, and 4.3% (v/v) water was prepared. The enzymatic reaction mixture was first mixed with the ninhydrin solution at a 1:1 ratio and then spotted onto a TLC plate. A purple/pink spot indicates the presence of the unnatural amino acid. The F₂-Tyr prepared was used in the next experiment for genetic incorporation into hCDO.

Cloning, expression, purification and protein analysis. Expression vector pVP16-hCDO was bought from the DNASU Plasmid Repository. Expression and purification of His-MBP-tagged enzymes hCDO was performed as follows: the culture was grown at 37 °C in Luria Bertani (LB) media in a baffled flask at 200 r.p.m. with the appropriate antibiotic to an optical density of 0.8 AU at 600 nm. Isopropyl-β-thiogalactoside (IPTG) was added to a final concentration of 500 µM, and the temperature was lowered to 28 °C. After 12 h of further incubation, the cells were harvested and resuspended in lysis buffer (300 mM NaCl, 50 mM Tris-HCl, and 1 mM ferrous ammonium sulfate, pH 8.0) and then disrupted by a Microfluidizer LM20 cell disruptor; the supernatant was recovered after centrifugation (13,000g for 30 min). His-MBP-tagged protein was separated using nickel-nitrilotriacetic acid resin. After being washed with washing buffer (300 mM NaCl, 50 mM Tris-HCl, 20 mM imidazole, pH 8.0), protein was eluted with elution buffer (300 mM NaCl, 50 mM Tris-HCl, 250 mM imidazole, pH 8.0). Purified protein-containing fractions were confirmed by SDS-PAGE and then dialyzed into the storage buffer (50 mM Tris-HCl, 100 mM NaCl, 10% glycerol, pH 8.0). The dialyzed protein was concentrated and stored at –80 °C for further use. Protein concentration was determined by UV-vis absorbance at 280 nm ($\epsilon_{280\text{ nm}} = 25,440\text{ cm}^{-1}\text{M}^{-1}$). For the expression of F₂-Tyr157, Cl-Tyr157, and Cl₂-Tyr157 hCDO proteins, pEVOL-F₂-TyrRS (or pEVOL-Cl-TyrRS and pEVOL-Cl₂-TyrRS) was co-transformed with pVP16-hCDO157TAG into BL21 (DE3). A single colony was grown overnight at 37 °C in 4 mL of LB medium. The transformed cells were induced with 0.5 mM IPTG and 0.02% *L*-arabinose at OD₆₀₀ of 0.8 in the presence of 0.5 mM F₂-Tyr, Cl-Tyr157, or Cl₂-Tyr. After growing for 12 h at 30 °C, the F₂-Tyr157, Cl-Tyr157, and Cl₂-Tyr157 hCDO proteins were purified using the protocol described for WT hCDO. The MBP tag was removed by cleavage with TEV protease during the purification of the MBP-fused hCDO. The liberated native CDO and mutants were further purified by gel-filtration chromatography in 20 mM Tris-HCl and 50 mM NaCl (pH 8.0) and were ultrafiltrated to the required concentration for subsequent catalytic assays and crystallization. We used the GelAnalysis (<http://www.gelanalyzer.com>) to estimate the ratio of the cross-linked form.

Catalytic assay & maturation assay. *L*-Cysteine was dissolved in hCDO storage buffer. Freshly prepared hCDO and *L*-cysteine solutions were incubated in a 37 °C water bath for 5 min. Appropriate amounts of these two solutions were well mixed to initiate the reaction in a 37 °C water bath. The activity assay protocol for hCDO was derived from that previously described for CDO^{15,47,48}. A typical assay was conducted in a total volume of 300 µL as follows: 0.2–0.5 µM as-purified hCDO, 50 mM MES buffer, pH 6.1, 0.3 mM ammonium iron sulfate, 62.5 µM bathocuproine disulfonate and varying concentrations of *L*-cysteine (0–25 mM). The reaction was initiated by addition of cysteine with shaking to ensure proper oxygenation using a shaker (220 r.p.m.). The hCDO-catalyzed reaction was quenched by the addition of 2 µL HFBA and centrifugation for removal of precipitated proteins. All aliquots were immediately ultrafiltered through Microcon (Millipore, Bedford, MA) 10,000 Da molecular-weight cut-off ultrafiltration tubes. The flow-through was kept in an ice water bath for further high-performance liquid chromatography analysis. To detect the dioxygenase product CSA, reactions aliquots were applied to a reverse-phase column (Thermo InertSustain C-18 column, 100 × 4.6 mm, 5 µM) with a mobile phase of 20 mM sodium acetate, 0.6% methanol, 0.3% heptofluorobutyric acid, pH 2.0. Curves of initial reaction velocity versus cysteine starting concentration were fitted with the Michaelis-Menten equation in OriginPro (OriginLab, Northampton, MA) to determine k_{cat} and K_{m} values. The maturation assay was performed in 100 mM Tris-HCl (pH 8.0) buffer with 25 mM *L*-cysteine and 75 µM enzyme at 37 °C. The SDS-PAGE was replicated at least three times by the independent experiments to ensure reproducibility.

Crystallization. A 2-µL aliquot of the enzyme solution (10 mg/mL) was mixed with 2 µL of reservoir solution containing 0.1 M MES (pH 6.5), 2 M ammonium sulfate, and 2% PEG 400 and was crystallized by the hanging-drop, vapor-diffusion method at 22 °C. The crystals grew as rods in 3 d. After soaking in a cryoprotectant containing reservoir solution plus 20% glycerol for 0.5 min, the crystal was flash-frozen and stored by liquid nitrogen for data collection using synchrotron radiation. The substrate-bound structures were obtained by soaking 100 mM *L*-cysteine to the CDO crystals. The anaerobic crystallization for obtaining pure uncross-linked CDO was conducted in an O₂-free COY anaerobic chamber.

Data collection, structure determination and refinement. Crystallographic data were acquired at a temperature of 100 K at Stanford Synchrotron Radiation Lightsource beamline BL9-2 (PDB code, wavelength (Å); 6BGM, 0.97946; 6BGE, 1.0000; 6BPU, 0.97946; 6BPV, 0.97622; 6BPW, 0.97946; 6BPX, 1.0000; 6CDH, 1.0000; 6CDN, 1.0000) and at the Advanced Photon Sources (Argonne National Laboratory, Argonne, IL) beamline 19BM (PDB code, wavelength (Å); 6BPT, 0.97919; and 6BPS, 0.97919). All X-ray diffraction intensity data were integrated, scaled, and merged using HKL2000⁴⁹. Molecular replacement was performed with Phenix⁵⁰ using the crystal structure of CDO as a starting model (Protein Data Bank entry 2ICL)¹⁴. The final model was manually adjusted and refined with Coot⁵¹ and Phenix. Ramachandran statistics were analyzed using MolProbity (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>). All the phi and psi angles were located in the preferred and allowed regions without any outliers. We generated all the molecular model figures using PyMOL (W.L. deLano, The PyMOL Molecular Graphics System version 1.8.6.0. Schrödinger LLC, <http://www.pymol.org/>; 2002).

Mass spectrometry. Solutions of the intact protein were exhaustively desalted by filtration with 10 mM ammonium acetate through centrifugal filters with 10 kDa membrane cut-off. For ESI-MS analysis, samples were diluted to approximately 2 µM in a solution containing 50% methanol and 0.1% acetic acid. Mass spectra were collected on a maXis plus quadrupole-time of flight mass spectrometer equipped with an electrospray ionization source (Bruker Daltonics) and operated in the positive ionization mode. Samples were introduced via syringe pump at a constant flow rate of 3 µL/min. Source parameters are summarized as follows: capillary voltage, 3,500 V with a set end plate offset of 500 V; nebulizer gas pressure, 0.4 bar; dry gas flow rate, 4.0 L/min; source temperature, 200 °C. Mass spectra were averages of 1 min of scans collected at a rate of 1 scan per second in the range of 50 ≤ *m/z* ≤ 3,000. Compass Data Analysis software version 4.3 (Bruker Daltonics) was used to process all mass spectra.

To generate cross-linked peptide species within a mass range suitable for efficient fragmentation and large coverage of the peptide sequence, a multiple enzyme digestion approach was used as previously described by Kleffmann and co-workers⁵² with minor modification. In short, the WT and F₂-Tyr157 CDO proteins were digested in solution first with trypsin, followed by GluC protease, and then chymotrypsin. Proteomics-grade enzymes were used in 1:25 enzyme:substrate ratio and incubated at 37 °C with each enzyme overnight. For MALDI-TOF/TOF mass spectrometry, a supersaturated matrix solution was prepared by dissolving α-cyano-4-hydroxycinnamic acid (Bruker Daltonics) in a mixture of 50% acetonitrile, 0.1% TFA in water. Reconstituted digests were spotted onto a stainless-steel target using the sandwich method in which 1 µL of matrix solution was spotted and dried, followed by 1 µL of sample solution, and another 1 µL of matrix solution. All MALDI-TOF/TOF mass spectra were collected on an ultrafleXtreme MALDI-TOF/TOF mass spectrometer equipped with a smartbeam II laser (Bruker Daltonics) in the positive ionization and reflector modes. The instrument was calibrated using the peptide calibration standard mixture (from Bruker Daltonics containing angiotensin II, angiotensin I, substance P, ACTH clip 1-17, ACTH clip 18-39, and somatostatin 28); and acquisition optimized in the mass range from 500 to 4,500 *m/z*. Approximately 3,000–8,000 shots were acquired per MS and LIFT-MS/MS spectrum using 1,000 Hz acquisition speed. FlexAnalysis 3.3 software (Bruker Daltonics) was used for data processing.

Nuclear magnetic resonance spectroscopy. ¹⁹F NMR spectra were recorded on an Agilent (Billerica, MA) DD2 500 MHz spectrometer equipped with a 5-mm inverse detect probe at 300 K. Spectra were recorded in 90/10 buffer/D₂O and referenced to internal trifluoroacetic acid (–76.5 p.p.m.). One-dimension ¹⁹F spectra (s2pul) were recorded with 5 s relaxation delay, 64 k data points, and multiplied with an exponential function for a line-broadening of 5 Hz before Fourier transformation. All NMR data were processed using MestReNova NMR v11.0.3 software. The condition for conversion of F₂-Tyr157 CDO to the all cross-linked form was 50 mM pH 6.5 MES buffer, 25 mM *L*-cysteine, and 300 µM protein.

Statistics and reproducibility. All experiments, including the assays and SDS-PAGE, were obtained from replicated independent experiments to ensure reproducibility. The representative results, the numbers of independent experiments are labeled in the figure legends or associated text in Supplementary Table 2.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability. Structure factors and coordinates for the crystal structures solved in this work have been deposited in the Protein Data Bank under accession codes 6BGM, 6BGE, 6BPU, 6BPV, 6BPW, 6BPX, 6CDH, 6CDN, 6BPT, and 6BPS.

References

46. Seyedsayamdost, M. R., Yee, C. S. & Stubbe, J. Site-specific incorporation of fluorotyrosines into the R2 subunit of *E. coli* ribonucleotide reductase by expressed protein ligation. *Nat. Protoc.* **2**, 1225–1235 (2007).
47. Arjune, S., Schwarz, G. & Belaidi, A. A. Involvement of the Cys-Tyr cofactor on iron binding in the active site of human cysteine dioxygenase. *Amino Acids* **47**, 55–63 (2015).
48. Stipanuk, M. H., Dominy, J. E. Jr., Ueki, I. & Hirschberger, L. L. Measurement of cysteine dioxygenase activity and protein abundance. *Curr. Protoc. Toxicol.* **38**, 6.15.11–16.15.25 (2008).
49. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
50. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, (213–221) (2010).
51. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
52. Kleffmann, T., Jongkees, S. A. K., Fairweather, G., Wilbanks, S. M. & Jameson, G. N. L. Mass-spectrometric characterization of two posttranslational modifications of cysteine dioxygenase. *J. Biol. Inorg. Chem.* **14**, 913–921 (2009).

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study. [For final submission](#): please carefully check your responses for accuracy; you will not be able to make changes later.

► Experimental design

1. Sample size

Describe how sample size was determined.

Sample sizes were chosen based on preliminary experiments so as to provide sufficient power for statistical comparison (where appropriate). The values with standard errors were determined from three independent experiments.

2. Data exclusions

Describe any data exclusions.

In crystallographic analysis data were excluded on statistical basis according to default parameters of software employed. There is no other data excluded from the analyses.

3. Replication

Describe the measures taken to verify the reproducibility of the experimental findings.

Experimental findings could be reliably reproduced successfully.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

No randomization of samples/organisms/participants is required for the reported structural and biochemical analysis.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Blinding is not applicable, because no group allocation was carried out.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a | Confirmed

- ☐ ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- ☐ ☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ A statement indicating how many times each experiment was replicated
- ☒ ☐ The statistical test(s) used and whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- ☒ ☐ Test values indicating whether an effect is present
*Provide confidence intervals or give results of significance tests (e.g. *P* values) as exact values whenever appropriate and with effect sizes noted.*
- ☐ ☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- ☐ ☒ Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on [statistics for biologists](#) for further resources and guidance.

► Software

Policy information about [availability of computer code](#)

7. Software

Describe the software used to analyze the data in this study.

Coot v0.8.8; Phenix-dev-2926; PyMOL v1.8.6.0; CCP4 Program Suit 6.4.0; OriginPro v8.5; GelAnalyzer; MestreNova NMR 11.0.4; Bruker Compass DataAnalysis v4.3; Bruker FlexAnalysis v3.4; mMASS v5.5.0

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* [guidance for providing algorithms and software for publication](#) provides further information on this topic.

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

All materials readily available from authors, commercial sources.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used in this study.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No cell lines were used in this study.

b. Describe the method of cell line authentication used.

No cell lines were used in this study.

c. Report whether the cell lines were tested for mycoplasma contamination.

No cell lines were used in this study.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

No cell lines were used in this study.

► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.

No research animals or animal-derived materials were used in this study.

Policy information about [studies involving human research participants](#)

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

There is no experiments about human research participants.

In the format provided by the authors and unedited.

Cleavage of a carbon–fluorine bond by an engineered cysteine dioxygenase

Jiasong Li¹, Wendell P. Griffith¹, Ian Davis¹, Inchul Shin¹, Jiangyun Wang², Fahui Li², Yifan Wang¹, Daniel J. Wherritt ¹ and Aimin Liu ^{1*}

¹Department of Chemistry, University of Texas at San Antonio, San Antonio, TX, USA. ²Institute of Biophysics, Chinese Academy of Sciences, Beijing, P.R. China. *e-mail: Feradical@utsa.edu

Supplementary Information

Cofactor Biogenesis in Cysteine Dioxygenase: C-F Bond Cleavage with Unnatural Tyrosine

Jiasong Li¹, Wendell P. Griffith¹, Ian Davis¹, Inchul Shin¹, Jiangyun Wang², Fahui Li², Yifan Wang¹, Daniel J. Wherrett¹, and Aimin Liu^{1,*}

¹Department of Chemistry, University of Texas at San Antonio, San Antonio, TX 78249, USA

²Institute of Biophysics, Chinese Academy of Sciences, Beijing, 100101, P.R. China

List of Supplementary Tables

Supplementary Table 1. X-ray crystallographic data collection and refinement statistics.

Supplementary Table 2. Catalytic efficiency of wild-type human CDO and variants.

List of Supplementary Figures

- Supplementary Fig. 1.** Cysteine dioxygenase (CDO) is a nonheme iron enzyme playing a central role in cellular thiol metabolism.
- Supplementary Fig. 2.** The active site of wild-type (WT) CDO and variants.
- Supplementary Fig. 3.** Comparison of ESI-MS spectra of purified WT CDO, Cl-Tyr157 CDO, F₂-Tyr15 CDO, and Cl₂-Tyr157 CDO.
- Supplementary Fig. 4.** Catalytic activity assays of wild-type human CDO and variants.
- Supplementary Fig. 5.** SDS-PAGE of WT and CDO variants.
- Supplementary Fig. 6.** Comparison of ESI-MS spectra of purified F₂-Tyr157 CDO in two charge states.
- Supplementary Fig. 7.** ¹⁹F NMR spectra of the F₂-Tyr157 CDO.
- Supplementary Fig. 8.** Crystal structures of F₂-Tyr/Cl₂-Tyr157 CDO with both uncrosslinked and crosslinked forms.
- Supplementary Fig. 9.** Conformational changes of the active site in CDO variants.
- Supplementary Fig. 10.** Comparison of ESI-MS spectra of purified Cl-Tyr157 CDO in two charge states.
- Supplementary Fig. 11.** A working model for C-F bond cleavage during cofactor formation in F₂-Tyr157 CDO.

Supplementary Table 1. X-ray crystallographic data collection and refinement statistics

	WT CDO (6BGM) (Crosslinked)	WT CDO- CYS (6BGF) (Crosslinked)	F ₂ -Tyr CDO (6BPU) (As-purified, mixture)	F ₂ -Tyr CDO- CYS (6BPV) (Crosslinked)	Cl ₂ -Tyr CDO (6BPW) (As- purified, mixture)
Data collection					
Space group	<i>P</i> 6 ₅				
Cell dimensions					
<i>a</i> , <i>b</i> , <i>c</i> (Å)	131.8, 131.8, 34.3	131.3, 131.3, 34.2	131.4, 131.4 34.2	131.4, 131.4 34.2	131.1, 131.1 34.3
α , β , γ (°)	90°, 90°, 120°				
Resolution (Å)	50.00 - 2.20 (2.26 - 2.20)	50.00 - 2.25 (2.31 - 2.25)	50.00 - 1.80 (1.84 - 1.80)	50.00 - 1.95 (2.00 - 1.95)	50.00 - 2.43 (2.49 - 2.43)
<i>R</i> _{sym} or <i>R</i> _{merge}	18.5 (83.9)	8.2 (41.9)	20.8 (67.8)	16.3 (76.8)	21.3 (67.5)
<i>I</i> / σ <i>I</i>	11.9 (2.2)	18.9 (2.7)	12.6 (1.3)	14.6 (1.8)	11.1 (1.5)
Completeness (%)	99.9 (100.0)	98.8 (92.0)	99.7 (96.8)	99.9 (98.9)	97.7 (94.3)
Redundancy	8.1 (6.6)	5.4 (3.7)	7.3 (4.0)	9.4 (6.8)	9.7 (6.4)
Refinement					
Resolution (Å)	32.94 - 2.20	37.90 - 2.25	37.93 - 1.80	31.57 - 1.95	32.83 - 2.43
No. reflections	17462	16149	31938	25060	12767
<i>R</i> _{work} / <i>R</i> _{free}	17.9 / 21.6	16.5 / 21.9	16.8 / 18.9	16.0 / 17.9	16.0 / 20.6
No. atoms					
Protein	1540	1553	1589	1578	1538
Cys	N / A	7	N / A	7	N / A
Fe ²⁺	1	1	1	1	1
Water	72	104	156	169	79
<i>B</i> -factors					
Protein	46.2	42.2	34.0	33.1	40.1
Cys	N / A	47	N / A	38.6	N / A
Fe ²⁺	42.2	38.2	27.7	27.5	39.7
Water	51.6	46.5	45.2	43.5	44.6
R.m.s. deviations					
Bond lengths (Å)	0.007	0.007	0.007	0.007	0.008
Bond angles (°)	0.841	0.910	0.904	1.000	0.968

	Cl ₂ -Tyr CDO- CYS (6BPX) (Crosslinked)	Cl-Tyr CDO (6CDH) (Mixture)	Cl-Tyr CDO -CYS (6CDN) (Crosslinked)	F ₂ -Tyr CDO (6BPT) (Uncrosslinked)	F ₂ -Tyr CDO -CYS (6BPS) (Uncrosslinked)
Data collection					
Space group	<i>P</i> 6 ₅				
Cell dimensions					
<i>a</i> , <i>b</i> , <i>c</i> (Å)	131.8, 131.8 34.3	131.4, 131.4, 34.2	131.3, 131.3, 34.2	131.8,131.8, 34.3	131.2, 131.2, 34.1
α , β , γ (°)	90°, 90°, 120°				
Resolution (Å)	50.00 - 2.15 (2.21 - 2.15)	50.00 -1.82 (1.85 - 1.82)	50.00 - 2.05 (2.09 - 2.05)	50.00 - 2.40 (2.44 - 2.40)	50.00 - 2.10 (2.14 - 2.10)
<i>R</i> _{sym} or <i>R</i> _{merge}	14.5 (69.3)	21.7 (92.6)	43.6 (13.5)	21.7 (92.6)	16.1 (97.6)
<i>I</i> / σ <i>I</i>	13.2 (1.4)	13.5 (3.1)	15.4 (2.2)	9.2 (2.0)	12.8 (1.7)
Completeness (%)	99.8 (99.9)	99.8 (99.8)	99.2 (91.9)	98.1 (92.5)	99.8 (98.8)
Redundancy	6.0 (5.2)	13.5 (3.12)	5.7 (3.3)	10.0 (8.7)	9.4 (6.9)
Refinement					
Resolution	43.15 - 2.15	43.03 - 1.82	43.00 – 2.05	38.04 - 2.40	37.88 - 2.10
No. reflection	18829	30625	21263	13427	20110
<i>R</i> _{work} / <i>R</i> _{free} (%)	17.8 / 21.2	16.5 / 18.2	16.2 / 19.3	15.6 / 20.2	16.7 / 20.3
No. atoms					
Protein	1537	1577	1542	1541	1555
CYS	7	N / A	7	N / A	7
Fe ²⁺	1	1	1	1	1
Water	89	162	112	118	156
B-factors					
Protein	42.1	34.9	36.6	23.9	27.27
CYS	45.9	N / A	44.5	N / A	27.86
Fe ²⁺	35.8	31.9	32.3	14.8	18.99
Water	49.2	46.6	44.8	29.1	36.8
R.m.s. deviations					
Bond lengths (Å)	0.009	0.007	0.008	0.008	0.008
Bond angles (°)	1.027	0.904	0.934	0.998	1.020

Each data set was collected from a single crystal. Similar crystal structures were solved to ensure the observations are reproducible.

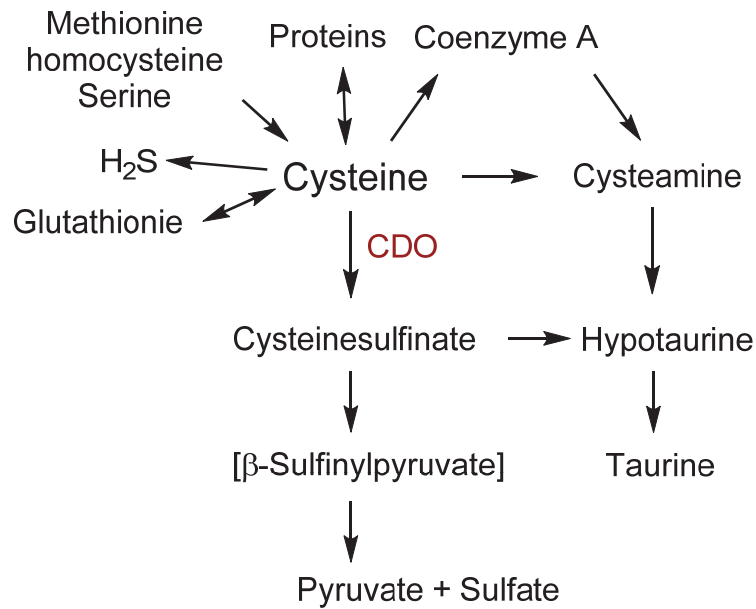
Values in parentheses are for highest-resolution shell.

Supplementary Table 2. Catalytic efficiency of wild-type CDO and variant

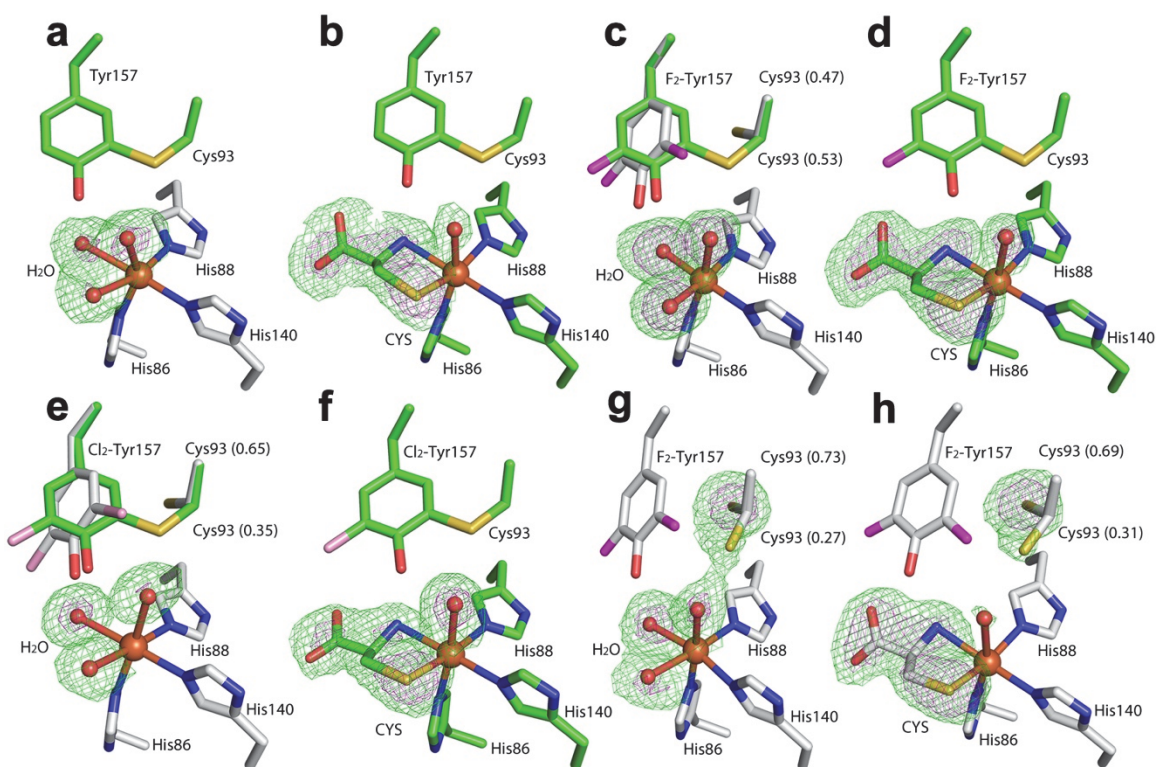
Construct	k_{cat} (s⁻¹)	K_{m} (mM)	$k_{\text{cat}}/K_{\text{m}}$ (M⁻¹s⁻¹)
WT CDO (Substrate inhibition)	9.3 ± 3.4	8.5 ± 3.6	1104
WT CDO	4.87 ± 0.86	4.05 ± 0.99	1202
Cl-Tyr157 CDO	0.19 ± 0.0094	5.59 ± 0.50	34.0
F₂-Tyr157 CDO	0.77 ± 0.025	6.09 ± 0.55	126
Cl₂-Tyr157 CDO	0.17 ± 0.038	3.60 ± 0.22	47.2

Values for $k_{\text{cat}}/K_{\text{m}}$ were calculated using activity data shown in **Supplementary Fig. 4**. The values were obtained from non-linear fitting with standard errors, and n = 3 independent experiments were repeated with similar results.

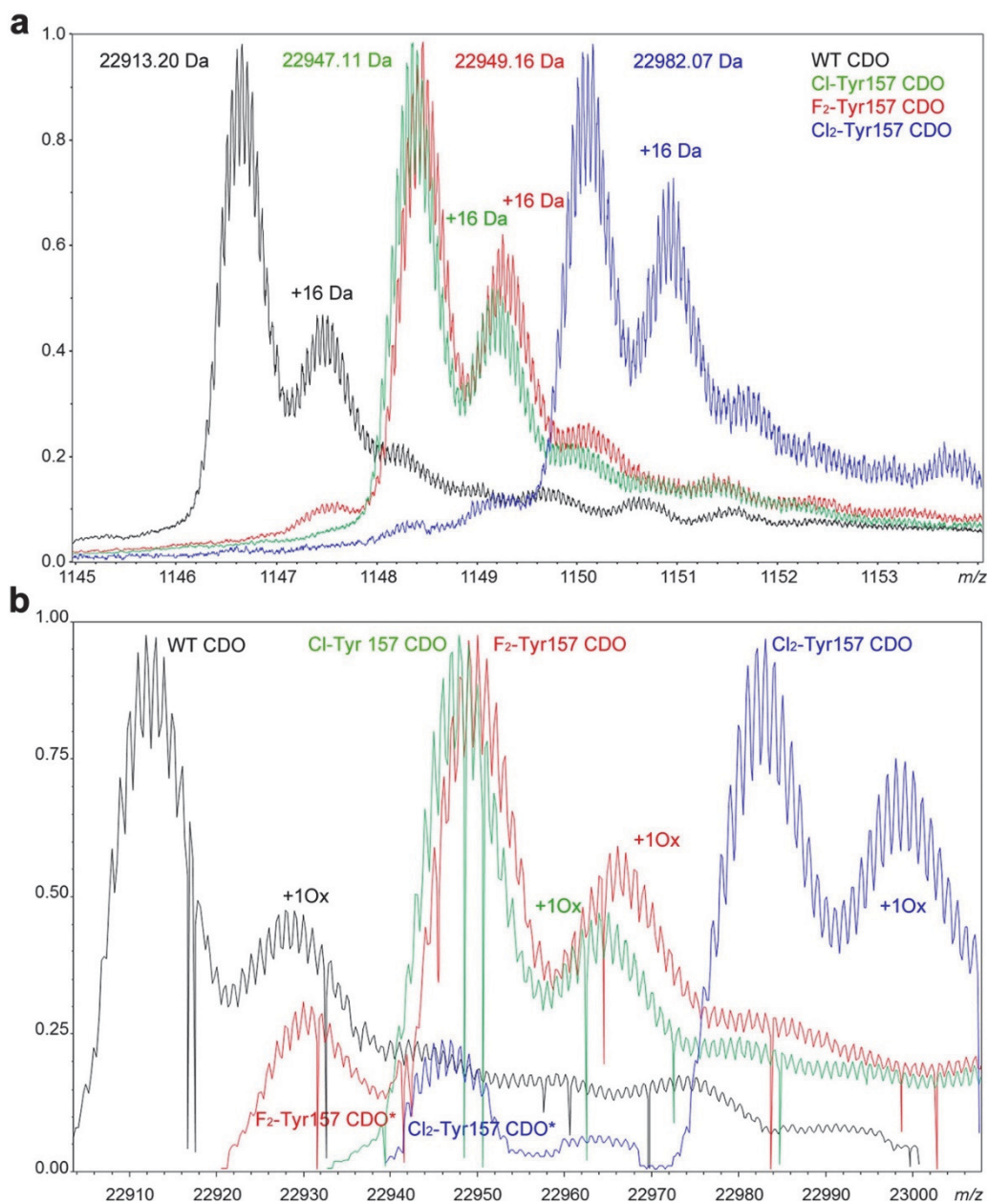
All assays were conducted in 50 mM MES buffer, pH 6.1, with 0.3 mM ferrous iron. The Activity assay for recombinant expressed human CDO was similar to the protocol used for rat CDO⁴. Because of the high amount of substrate inhibition shown by WT CDO (**Supplementary Fig. 4a**), only activity data that were obtained from low amounts of substrate (**Supplementary Fig. 4b**). WT CDO exhibited high amounts of activity within the range of concentrations (0 – 5 mM). These two mutants resulted in a reduction of CDO activity to 5~10% of that observed for WT CDO (Table 1). The decreased activity of the CDO variants could arise from a decrease in the fraction of mature enzyme or decreased pK_a of the cofactor due to halogen substitution. At supra-physiological substrate levels, however, these CDO variants demonstrated no signs of substrate inhibition (**Supplementary Fig. 4 c, d and e**). Overall, these data indicate that the modification of the phenol ring of Tyr157 does not abolish the activity of CDO. It is therefore reasonable to probe cofactor biogenesis in CDO through genetic incorporation of unnatural amino acids.



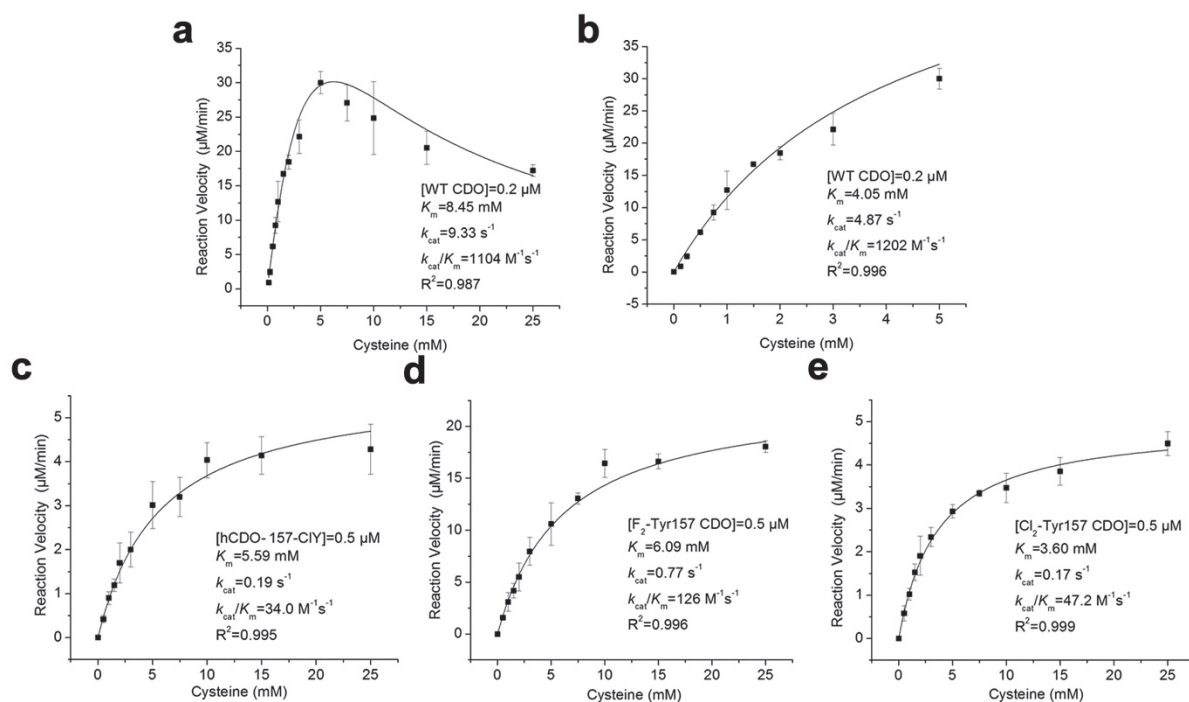
Supplementary Fig. 1. Cysteine dioxygenase (CDO) is a nonheme iron enzyme playing a central role in cellular thiol metabolism. CDO catalyzes an O₂-dependent dioxygenation converting cysteine to sulfinic acid (cysteinesulfinate), which is part of the biological sulfur cycle.



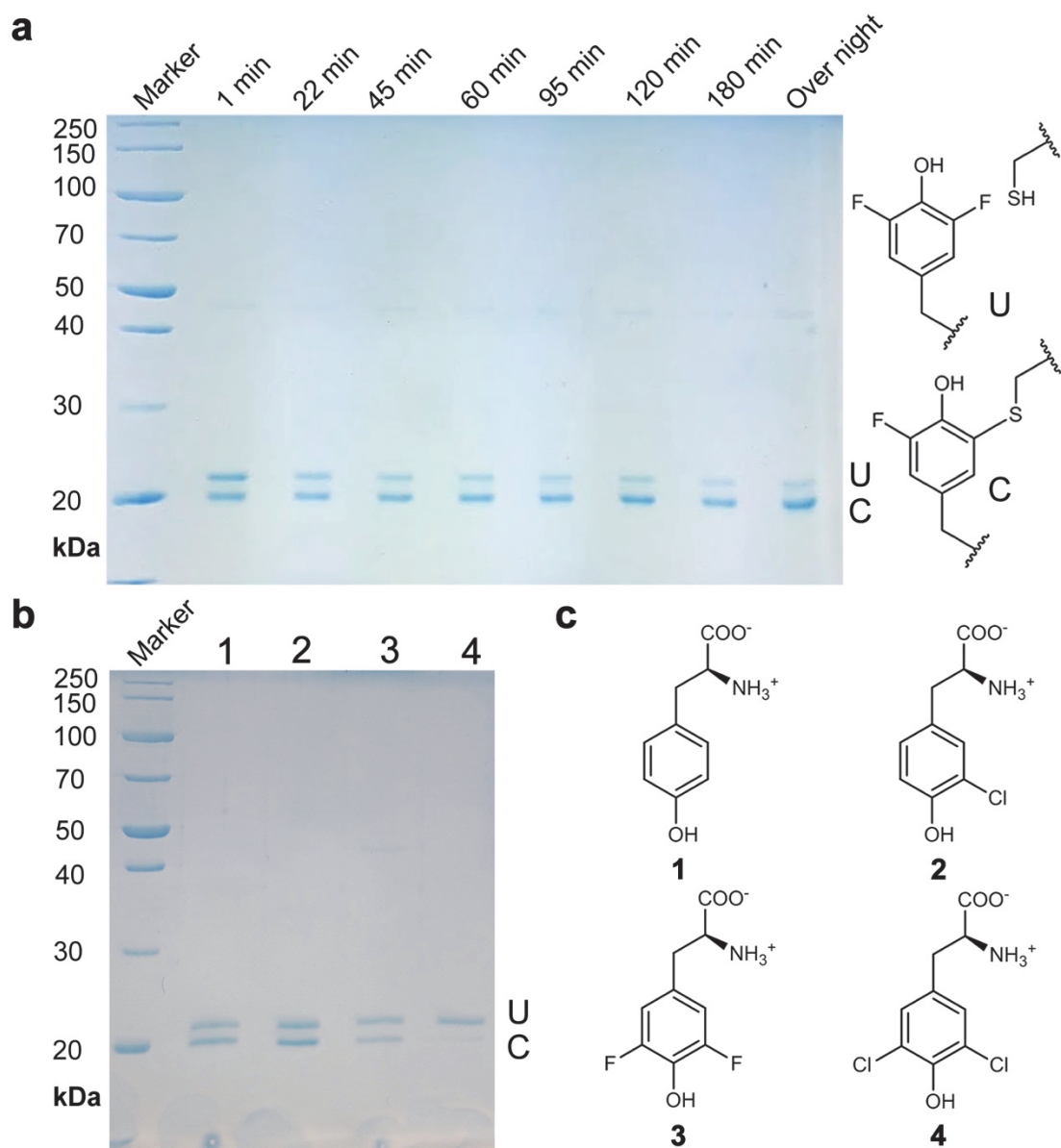
Supplementary Fig. 2. The active site of wild-type (WT) CDO and variants. The omit F_o-F_c electron densities (green) of the WT CDO & mutants contoured at 3 σ (green) and 6 σ (purple), respectively. The L-cysteine substrate is labeled as CYS. **(a)** The active site of the WT CDO structure in the substrate-free form. **(b)** The active site of the WT CDO structure in the substrate-bound form. **(c)** The active site of the F₂-Tyr157 CDO structure in the in the crosslinked & uncrosslinked forms without substrate. **(d)** The active site of the substrate bound F₂-Tyr157 CDO structure in the crosslinked form. **(e)** The active site of the Cl₂-Tyr157 CDO structure in the in the crosslinked & uncrosslinked forms without substrate. **(f)** The active site of the substrate bound Cl₂-Tyr157 CDO structure in the crosslinked form. **(g)** The active site of the 100% uncrosslinked F₂-Tyr157 CDO structure in the substrate-free form. **(h)** The active site of the 100% uncrosslinked F₂-Tyr157 CDO structure in the substrate-bound form.



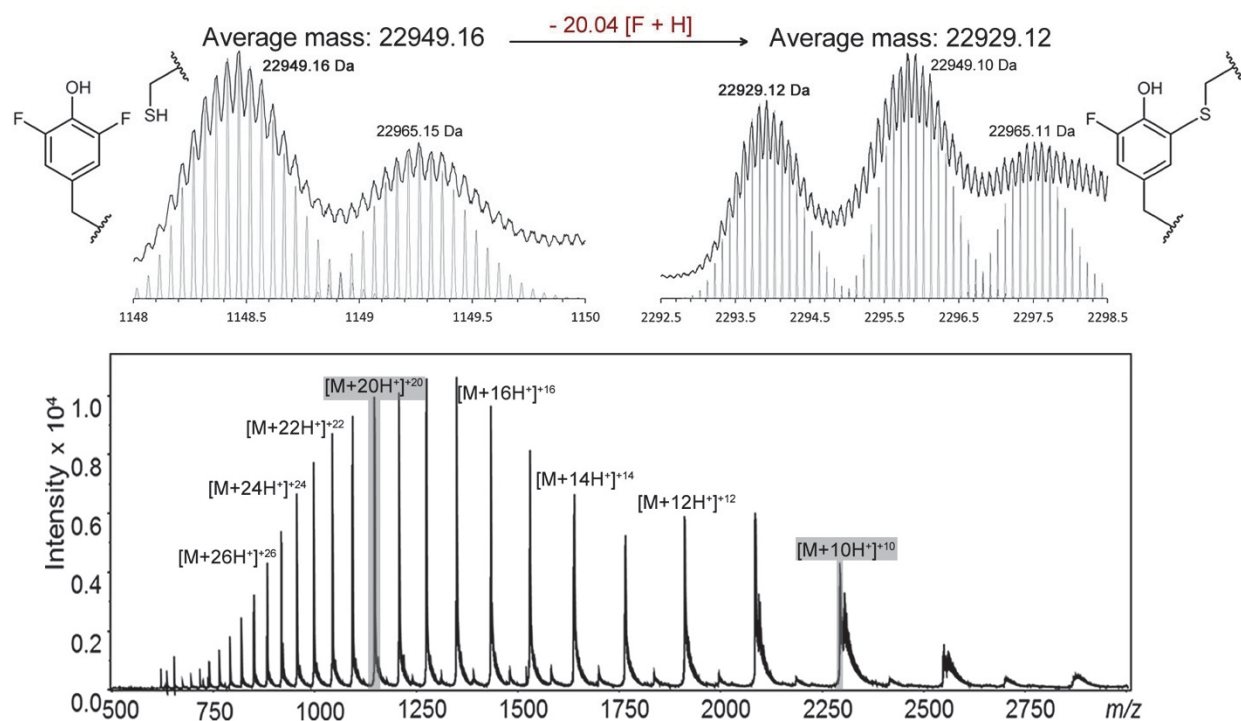
Supplementary Fig. 3. Comparison of ESI-MS spectra of purified WT CDO, Cl-Tyr157 CDO, F₂-Tyr157 CDO, and Cl₂-Tyr157 CDO. (a) Zoomed-in mass spectra of the +20 charge states with the experimentally determined masses of WT CDO, Cl-Tyr157, F₂-Tyr157 CDO and Cl₂-Tyr157 CDO. Peaks labeled with +16Da indicate the oxidized protein species most likely due to the oxidation of methionine. The Deconvolution of these mass spectra result in experimentally determined masses for these proteins of 22,929.22, 22,963.11, 22,965.15, and 22998.10 Da for the oxidized species of WT, Cl-Tyr157, F₂-Tyr157 and Cl₂-Tyr157 CDO, respectively. These measured mass values are consistent with the predicted masses of 22,929.21, 22,963.17, 22,965.19, and 22998.13 Da from protein sequence, respectively. (b) Comparison of deconvoluted mass spectra for WT CDO, Cl-Tyr157CDO, F₂-Tyr157 CDO, and Cl₂-Tyr157 CDO. Species detected are labeled. Species labeled with "*" indicate where a halogen atom was cleaved in the crosslink formation. These experiments were repeated independently two times with similar results.



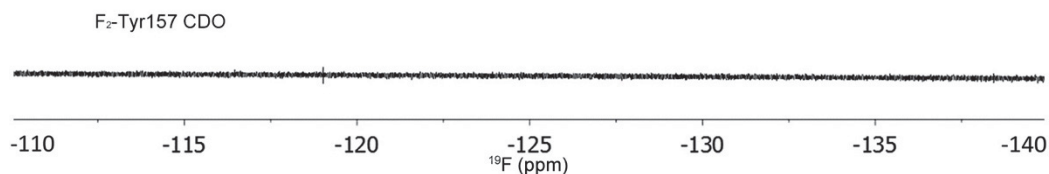
Supplementary Fig. 4. Catalytic activity assays of WT human CDO and variants. The production of cysteinesulfinic acid (CSA) was measured at various L-cysteine concentrations for the following purified recombinant proteins: WT (a) and (b), Cl-Tyr157 CDO (c), F₂-Tyr157 CDO (d), Cl₂-Tyr157 CDO (e). n = 3 independent experiments were repeated, and the data represent mean value $\pm$ s.d. Measurements were taken from distinct samples.



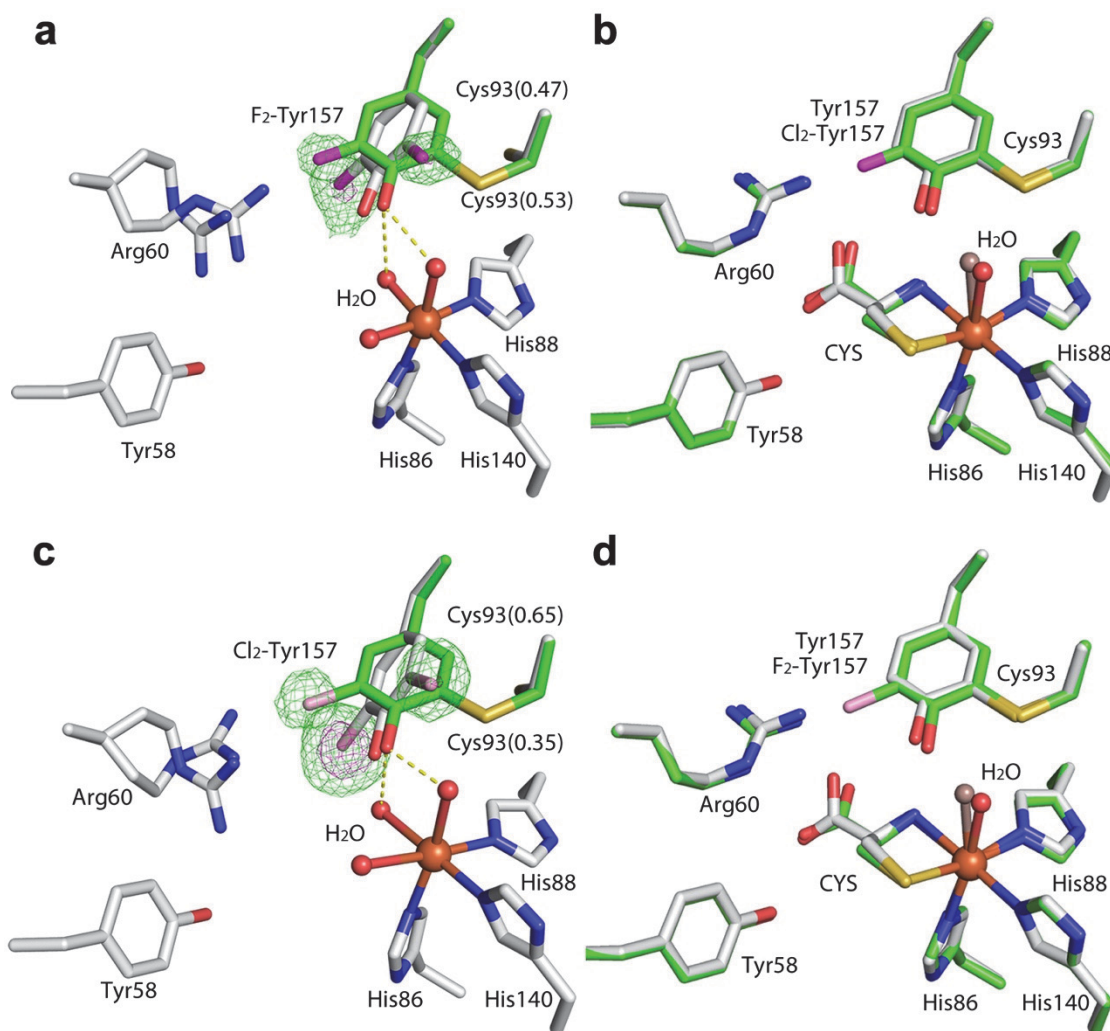
Supplementary Fig. 5. SDS-PAGE of WT and CDO variants with an unnatural tyrosine. U: Uncrosslinked form; C: Crosslinked form. **(a)** Conversion of as-isolated F₂-Tyr157 CDO to the mature form. **(b)** Full gel image for **Figure 1c** of the main text. SDS-PAGE shows two bands with the slower moving band being uncrosslinked enzyme (U) and faster moving band corresponding to the mature CDO (C). **(c)** Native tyrosine and halogen-substituted unnatural tyrosine analogous used in this study. L-Tyr (**1**), 3-Cl-L-Tyr (**2**), 3,5-F₂-L-Tyr (**3**), and 3,5-Cl₂-L-Tyr (**4**). The portion of uncrosslinked CDO decreases with time. The SDS-PAGE was replicated at least three times by the independent experiments to ensure reproducibility.



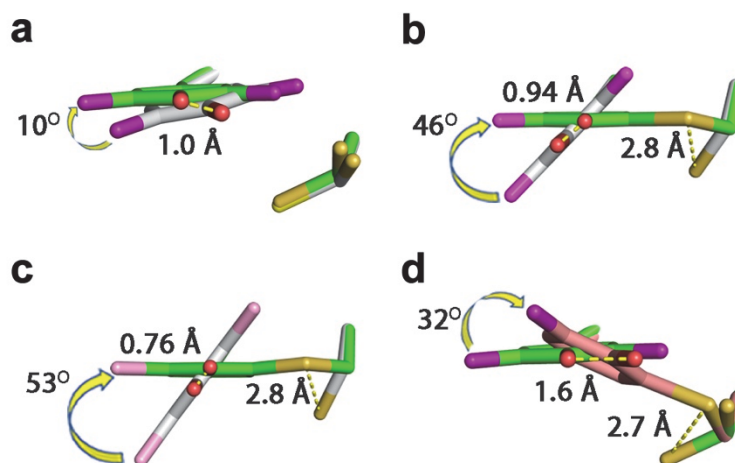
Supplementary Fig. 6. Comparison of ESI-MS spectra of purified F₂-Tyr157 CDO in two charge states. Zoomed-in mass spectra with the experimentally determined masses for the peaks in two charge states, respectively. The 20.04 Da difference of the first peak matches the mass of 1F and 1H. For the +20 peaks, the first one is for the uncrosslinked form (Measured: 22949.16 Da; Simulated: 22949.19 Da); the second one is for the oxidized and uncrosslinked form (Measured: 22965.15 Da; Simulated: 22965.19 Da). For the +10 peaks, the first one is for the crosslinked form (Measured: 22929.12 Da; Simulated: 22929.19 Da); the second one is for the uncrosslinked form (Measured: 22949.10 Da; Simulated: 22949.19 Da); the third one is for the oxidized and uncrosslinked form (Measured: 22965.11 Da; Simulated: 22965.19 Da). These experiments were repeated independently three times with similar results.



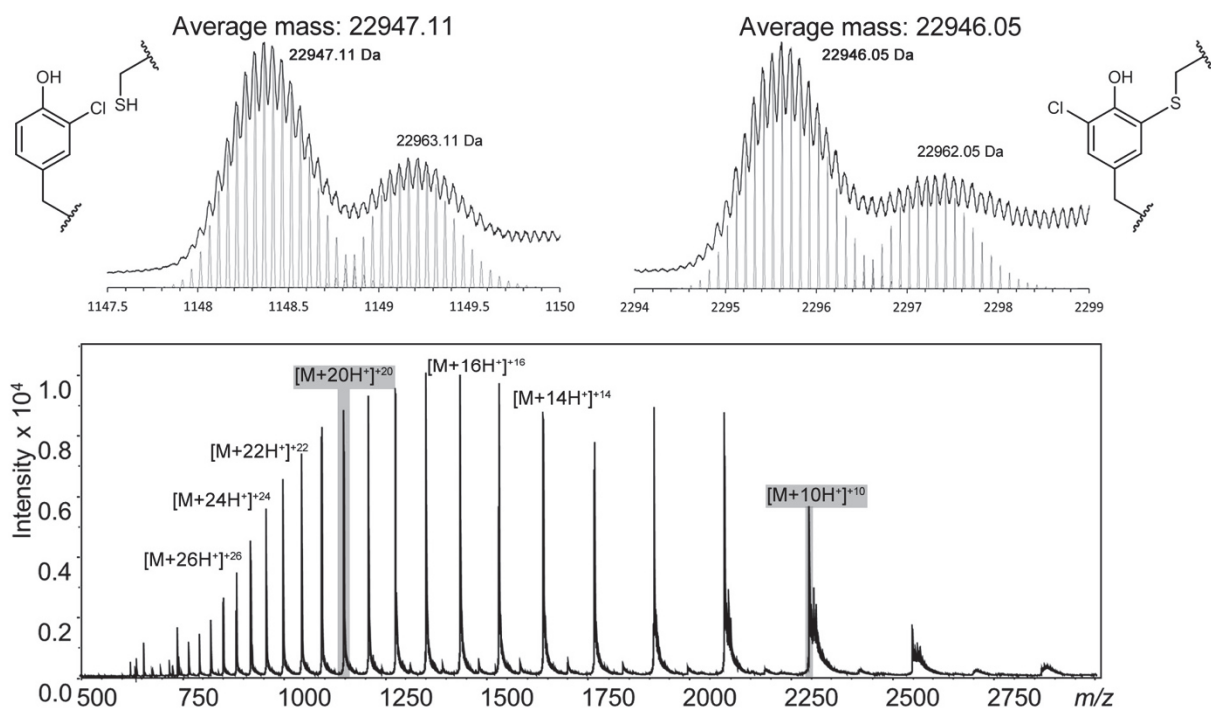
Supplementary Fig. 7. ¹⁹F NMR spectra of the F₂-Tyr157 CDO. ¹⁹F NMR spectrum of F₂-Tyr157 CDO, before reacting with L-cysteine and O₂, was collected after 12,480 transients. In this negative control, there is no signal of fluoride (as seen in the spiked NMR spectrum). These experiments were repeated independently two times with similar results.



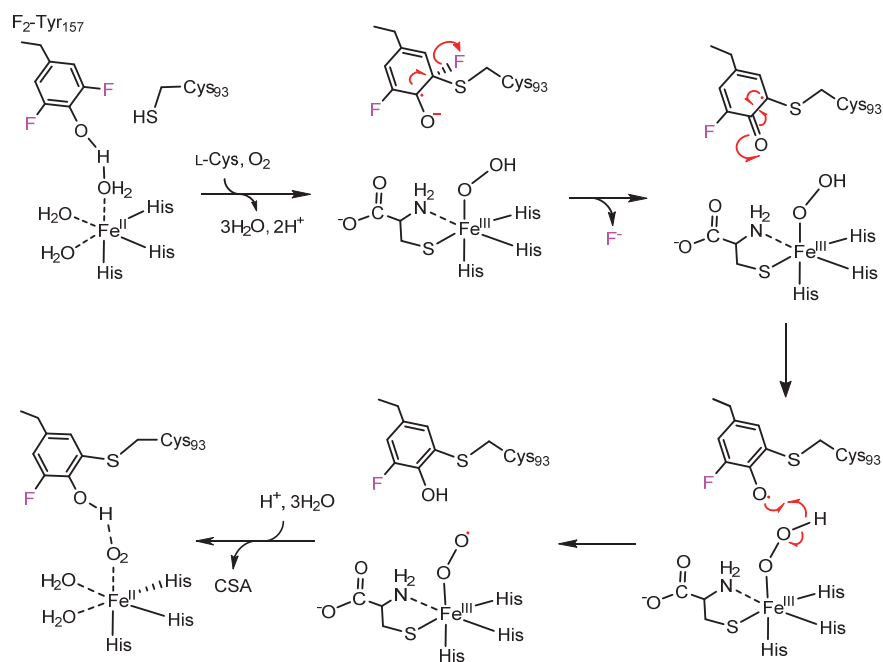
Supplementary Fig. 8. Crystal structures of F₂-Tyr/Cl₂-Tyr157 CDO with both uncrosslinked and crosslinked forms. The omit $F_o - F_c$ electron densities (green) of the fluorine or chlorine was contoured at 3 σ (green) and 6 σ (purple), respectively. The L-cysteine substrate is labeled as CYS. (a) The active site of F₂-Tyr157 CDO in the crosslinked & uncrosslinked forms. (b) Superimposition of the substrate-bound mature F₂-Tyr CDO (green) and the substrate-bound mature WT CDO (grey). (c) The active site of Cl₂-Tyr157 CDO in the crosslinked & uncrosslinked forms. (d) Superimposition of the substrate-bound mature Cl₂-Tyr (green) and the substrate-bound mature WT CDO (grey).



Supplementary Fig. 9. Conformational changes of the active site in WT CDO and variants. (a) Substrate binding caused a noticeable conformational change of both Cys93 and F₂-Tyr157. (b) The conformational difference of F₂-Tyr157 in the in F₂-Tyr157 CDO in the crosslinked & uncrosslinked F₂-Tyr157 CDO. (c) The conformational difference of Cl₂-Tyr157 in the crosslinked & uncrosslinked Cl₂-Tyr157 CDO. (d) F₂-Tyr157 rotates during cofactor formation. The fluorine atoms are shown in purple color.



Supplementary Fig. 10. Comparison of ESI-MS spectra of purified Cl-Tyr157 CDO in two charge states. Zoomed-in mass spectra with the experimentally determined masses for the peaks in two charge states, respectively. For the +20 peaks, the first one is for the uncrosslinked form (Measured: 22947.11 Da; Simulated: 22947.17 Da); the second one is for the oxidized and uncrosslinked form (Measured: 22963.11 Da; Simulated: 22963.17 Da). For the +10 peaks, the first one is most likely for the mixture of the crosslinked and uncrosslinked form (Measured: 22946.05 Da; Simulated: 22946.16 Da); the second one is for the oxidized form of the mixture (Measured: 22962.05 Da; Simulated: 22962.16 Da). This result shows that it is the C-H cleavage not the C-Cl cleavage during the cofactor biogenesis of Cl-Tyr157 CDO. These experiments were repeated independently two times with similar results.



(where CSA = cysteine sulfinic acid)

Supplementary Fig. 11. A working model for C-F bond cleavage during cofactor formation in F₂-Tyr157 CDO. In the proposed mechanism, the formation of a transient-state tyrosyl radical species in Tyr157 facilitates the C-F bond cleavage.