Supporting Information

Formation of Monofluorinated Radical Cofactor in Galactose Oxidase through Copper-Mediated C−F Bond Scission

Jiasong Li, Ian Davis, Wendell P. Griffith, and Aimin Liu*

Department of Chemistry, The University of Texas at San Antonio, San Antonio, TX 78249

List of Supplementary Tables

List of Supplementary Figures

Experimental Procedure

Materials

3,5-Diflurotyrosine (F$_2$-Tyr) was synthesized using an enzymatic method and isolated as described previously.[1] 3,5-Dichlorotyrosine (Cl$_2$-Tyr) (98% purity) was purchased from Ark Pharm, Inc and used for cell culture without further purification. All primers were synthesized at the Integrated DNA Technologies. 2,2'-Azino-*bis*(3-ethylbenzthiazoline-6-sulfonic acid) (ABTS), D-galactose, and horseradish peroxidase (HRP) were from Sigma-Aldrich (St Louis, MO). Other reagents were purchased from Sigma-Aldrich, New England Biolabs, and Thermo Fisher Scientific with the reagent grade or better and used as received. 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. The pEVOL-F$_2$/Cl$_2$-TyrRS and pET22b-GAO^{V}272TAG plasmids, a mutant of *Methanocaldococcus jannaschii* tyrosyl amber suppressor tRNA(MjtRNA^{Tyr}$_{CUA}$)/F$_2$-TyrRS (Tyr32Arg, Leu65Tyr, His70Gly, Phe108Asn, Gln109Cys, Asp158Asn, Leu162Ser) pair[2] or Cl$_2$-TyrRS (Tyr32Arg, Leu65Tyr, His70Gly, Phe108Asn, Gln109Cys, Asp158Asn, Leu162Ser) pair[3] were used to recognize the TAG codon in the 272 position of GAO^V. F$_2$-Tyr or Cl$_2$-Tyr was selectively incorporated into GAO^V at position 272 as expressed in *Escherichia coli* cells. Cu(II)-reconstituted GAO^V was achieved by adding CuSO$_4$ during the purification using Ni-NTA agarose beads.

Bacterial strain, genetic manipulation, and plasmids

Compared to the wild-type enzyme from *Fusarium austroamericanum*, GAO^V contains six mutations, S10P, M70V, G195E, V494A, N535D, and Pro136 as a silent mutation at DNA level.

The pET22b-GAO^V Tyr272TAG plasmid was generated by following the primers:

> GAO^V Tyr272TAG forward: (GGCTCGTGGGTAGCAGTCATCAGCTACCATG)
>
> GAO^V Tyr272TAG reverse: (CTGATGACTGCTACCCACGAGCCACTTGCATG)

The pEVOL-F$_2$-/Cl$_2$-TyrRS and pET22b-GAO^{V}272TAG plasmids, a mutant of *Methanocaldococcus jannaschii* tyrosyl amber suppressor tRNA(MjtRNA^{Tyr}$_{CUA}$)/F$_2$-TyrRS (Tyr32Arg, Leu65Tyr, His70Gly, Phe108Asn, Gln109Cys, Asp158Asn, Leu162Ser) pair[2] or Cl$_2$-TyrRS (Tyr32Arg, Leu65Tyr, His70Gly, Phe108Asn, Gln109Cys, Asp158Asn, Leu162Ser) pair[3] were used to recognize the TAG stop codon in the 272 position of GAO^V. F$_2$-Tyr or Cl$_2$-Tyr was selectively incorporated into GAO^V at the position 272 as expressed in *Escherichia coli* cells. Cu(II)-reconstituted GAO^V was achieved by adding CuSO$_4$ during the purification using Ni-NTA agarose beads.

Protein purification and characterization

The expression and purification of GAO^V was described previously.[4] The cell culture was prepared at 37 °C in Luria Bertani (LB) media within a baffled flask at 200 rpm with 100 mg ampicillin per liter. The cells were induced with 0.5 % lactose at 25 °C when the optical density reached 0.6 AU at 600 nm. After overnight culture, the cells were harvested and resuspended in the lysis buffer, i.e., 100 mM NaPi at pH 7.0, and then disrupted by a Microfluidizer LM20 cell disruptor. The supernatant was recovered after centrifugation (13,000×g for 30 min) at 4 °C. The *C*-terminal His-tagged protein was purified using Ni-NTA agarose beads. After buffer

exchanging with five CVs of washing buffer (100 mM NaPi, 20/50 mM imidazole, pH 7.0), the isolated protein was eluted with elution buffer (100 mM NaPi, 300 mM imidazole, pH 7.0). The GAO proteins were further purified by Superdex 200 gel-filtration column in pH 7.0 100 mM NaPi buffer with 5% glycerol and stored at -80 °C for the future use. The protein concentration was determined based on the extinction coefficient of $\varepsilon_{280\ nm}$ = 124,135 $cm^{-1}M^{-1}$. For the expression of F_2/Cl_2-Tyr157 variants, pEVOL-F_2/Cl_2-TyrRS was co-transformed with pET22b-GAO^V272TAG into BL21(DE3). The transformed cells were induced with 0.5% lactose and 0.02% L-arabinose at OD (600 nm) 0.6 in the presence of 0.5 mM F_2-Tyr or Cl_2-Tyr. GelAnalysis (http://www.gelanalyzer.com) was used to estimate the ratio of the cross-linked mature form. Cu(II)-reconstituted GAO^V was achieved by adding $CuSO_4$ during the purification using Ni-NTA agarose beads. The conversion of as-isolated F_2-Tyr272 GAO^V to the mature form was derived from that previously described for GAO in 50 mM piperazine-*N,N-bis*(2-ethanesulfonic acid) (PIPES), pH 6.5.[5] To convert as-isolated F_2-Tyr272 GAO^V to the mature form Tetrakis (acetonitrile) Cu(I), hexafluorophosphate [Cu(I)-$(CH_3CN)_4·PF_6$] was dissolved in anaerobic acetonitrile immediately before addition of 1.5 equivalents to degassed 10 μM F_2-Tyr272 GAO^V in the glovebox. Then the oxygen-saturated buffer was added rapidly. The samples were taken at different time to run the SDS-PAGE.

Mass Spectrometry
Solutions of the intact protein were exhaustively desalted by filtration with 10 mM ammonium bicarbonate through centrifugal filters with 10 kDa membrane cut-off. For ESI-MS analyses, samples were diluted to approximately 10 μM in a solution containing 50% methanol and 3% 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 a 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 endplate 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 one minute of scans collected at a rate of 1 scan per second in the range 50 ≤ m/z ≤ 3000. Compass Data Analysis software version 4.3 (Bruker Daltonics) was used to process all mass spectra.

Electron paramagnetic resonance (EPR) spectroscopy
The samples were transferred to quartz EPR tubes and slowly frozen in liquid nitrogen. EPR spectra were recorded on a Bruker E560 X-band spectrometer equipped with a cryogen-free 4 K temperature system with an SHQE high-Q resonator as described previously[6-7] at 100 kHz modulation frequency, 0.05 mW microwave power, 0.6 mT modulation amplitude at 30 K, and were averaged over four scans for each spectrum. Oxidation of GAO was performed in 100 mM NaPi, pH 7.0, by incubation with 100 mM $K_3[Fe(CN)_6]$ for 10 min followed by removal of the oxidant on a desalting column. The samples were transferred to quartz EPR tubes and slowly frozen in liquid nitrogen. After the measurement, the oxidized samples were treated with 20 mM hydroxyurea and remeasured. A sample of Cu(II)-EDTA (1 mM) was used as the standard sample to quantitate the EPR-active Cu(II) species in galactose oxidase and the genetically modified variants.

Activity assay

The coupled assay system measuring hydrogen peroxide production was used, and the steady-state kinetic analysis was derived from that previously described for GAO by Baron *et al.*[8] All assays were conducted at 25 °C in a total volume of 500 µL as follows: 0.1–5 nM reconstituted GAO, 5 U HRP, 1 mM ABTS in 50 mM NaPi buffer, pH 7.0, and varying concentrations of galactose (0.25–480 mM). The oxidation of 2 µM of ABTS per min equals the consumption of 1 µM of galactose and O_2 per min, under the conditions described above.

Crystallization, data collection, model building, and refinement

Crystals of the GAO[V] and unnatural amino acids variants were grown at 22 °C by using the sitting-drop vapor-diffusion technique against a mother liquor composed of 0.05 M ammonium acetate (pH 6.0), 0.1 M calcium chloride, 10% PEG 6000, 5% glycerol, 10 mM *N*-acetyl-D-glucosamine. A 2-µL aliquot of the enzyme solution (10 mg/ml) was mixed with a 2-µL of reservoir solution and crystallized by the sitting drop, vapor-diffusion method at 22°C. The crystals grew in 3 days. After soaking in a cryoprotectant containing reservoir solution plus 20% glycerol for a few seconds, the crystals were flash-cooled and stored in liquid nitrogen for data collection using synchrotron radiation.

Data collection, structure determination, and refinement.

Crystallographic data were acquired at 100 K temperature at Stanford Synchrotron Radiation Lightsource (SSRL) beamline BL9-2 and the Advanced Photon Sources (Argonne National Laboratory, Argonne, IL) beamline 19BM. All X-ray diffraction intensity data were integrated, scaled, and merged using HKL2000.[9] Molecular replacement was performed with Phenix[10] using the crystal structure of GAO from *Fusarium austroamericanum* as a starting model (Protein Data Bank entry 1GOG).[11] The final model was manually adjusted and refined with Coot[12] and Phenix. Ramachandran statistics were analyzed using MolProbity.[13] 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).

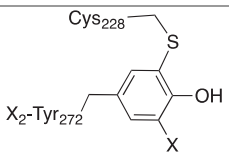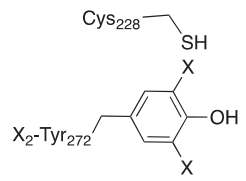**Table S1. X-ray crystallographic data collection and refinement statistics**

Description	GAO[V]	F$_2$-Tyr272 GAO[V]	Cl$_2$-Tyr272 GAO[V]
PDB ID	**6XLT**	**6XLS**	**6XLR**
Data collection			
Beamline	SBC-19-BM (0.97919 Å)	SSRL-BL9-2 (0.97946 Å)	SBC-19-BM (0.97919 Å)
Space group	*C*2		
Wavelength (Å)	50.00 – 1.48 (1.51 – 1.48)	50.00 – 1.80 (1.84 – 1.80)	50.00 – 1.23 (1.26 – 1.23)
a, *b*, *c* (Å)	97.3, 89.1, 86.2	97.2, 89.2, 86.2	97.2, 89.0, 86.0
α, β, γ (°)	90, 118, 90		
Completeness (%)	99.9 (99.0)	98.4 (97.7)	99.7 (99.9)
No. of total reflections	535369	227576	904404
No. of unique reflections	108128	59316	186045
I / σ	25.7 (1.1)	12.4 (1.3)	20.8 (1.73)
$CC_{1/2}$ last shell	0.646	0.555	0.940
Redundancy	5.0 (4.2)	3.8 (3.5)	4.9 (4.2)
R_{merge}	6.0 (95.9)	12.5 (83.0)	8.1 (48.9)
Refinement			
Resolution (Å)	28.47 – 1.47	41.63 – 1.80	32.84 – 1.23
No. of reflections	108106	59302	186037
R_{work} / R_{free}	14.8 / 16.9	14.8 / 18.4	15.5 / 17.2
RMSD for bond lengths (Å)	0.006	0.007	0.007
RMSD for bond angles (°)	1.1	1.1	1.2
Ramachandran statistics[2]			
Preferred (%)	96.7	97.0	97.0
Allowed (%)	3.3	3.0	3.0
Outliers (%)	0	0	0
No. of atoms			
Protein	4947	4850	4917
Cu[II] (Occupancy)[3]	1 (0.54)	1 (0.50)	1 (0.51)
Ca[II] (Occupancy)	1 (0.93)	1 (1.00)	1 (1.00)
Acetate	8	16	16
Glycerol	6	6	72
Water	755	626	806
Average *B*-factors (Å^2)			
Protein	31.2	27.3	17.8
Cu[II]	20.0	31.7	14.2
Ca[II]	31.6	33.4	25.6
Acetate	42.5	43.0	41.2
Glycerol	59.1	43.4	41.0
Water	38.6	39.4	33.7

[1] Values in parentheses are for the highest-resolution shell.

[2] Ramachandran statistics were analyzed using MolProbity.[14]

[3] The incomplete copper occupancy in the three structures may be due to the loss of copper during purification and crystallization.

Table S2. The mass difference between GAO^V and unnatural tyrosine variants at different charge states

Molecular weight difference (Da)	Uncrosslinked F$_2$-Tyr272 GAO^V minus crosslinked GAO^V (Crosslinked F$_2$-Tyr272 GAO^V minus crosslinked GAO^V)	Uncrosslinked Cl$_2$-Tyr272 GAO^V minus crosslinked GAO^V (Crosslinked Cl$_2$-Tyr272 GAO^V minus crosslinked GAO^V)
Theoretical difference	38 (18)	71 (34.5)
Experimental difference (z = +45)	35	55
Experimental difference (z = +47)	33	67
Experimental difference (z = +53)	37	72
Chemical structures (X = H/F/Cl)	Crosslinked X$_2$-Tyr272 GAO^V	Uncrosslinked X$_2$-Tyr272 GAO^V

The mass difference shows that both F$_2$-Tyr272 GAO^V and Cl$_2$-Tyr272 GAO^V are mixtures of crosslinked and uncrosslinked isoforms. The entries in these rows reflect the experimental data from experiments results shown in Figure S2. The chemical structures at the bottom are included to illustrate the mass difference. The expected difference between crosslinked F$_2$-Tyr272 GAO^V and crosslinked GAO^V is 18 Da, and the expected difference between uncrosslinked F$_2$-Tyr272 GAO^V and crosslinked GAO^V is 38 Da. The experimental difference at all three charge states is between 18 and 38 Da. Regarding the Cl$_2$-Tyr272 GAO^V, the expected difference between crosslinked Cl$_2$-Tyr272 GAO^V and crosslinked GAO^V is 34.5, and the expected difference between uncrosslinked F$_2$-Tyr272 GAO^V and crosslinked GAO^V is 71 Da. The experimental mass difference decreases as the charge state decreases. After crosslink formation, the protein is more resistant to acid and organic solvent denaturation, and thus is more compact, so it carries a less protonic charge and appears at lower charge states.

Table S3. Relative EPR-active Cu(II) signal in galactose oxidase variants

Enzyme	Cu(II)-reconstituted sample	$K_3Fe(CN)_6$-treated sample	Hydroxyurea-treated oxidized sample
GAO^V	60.9%	40.0%	59.9%
F_2-Tyr272 GAO^V	79.1%	65.8%	82.9%

The EPR signal intensities were obtained by double integration after baseline correction. The quantitation was carried out by comparing to a Cu(II)-EDTA standard and presented as a percentage of the corresponding protein concentration of the samples. The Cu(II) EPR signal represents the copper center without a spin coupled ligand radical. Thus, the increase of the copper signal intensity indicates the hidden portion of the Cu(II)-radical became EPR active, i.e., Cu(II) center without a spin-coupled ligand radical.
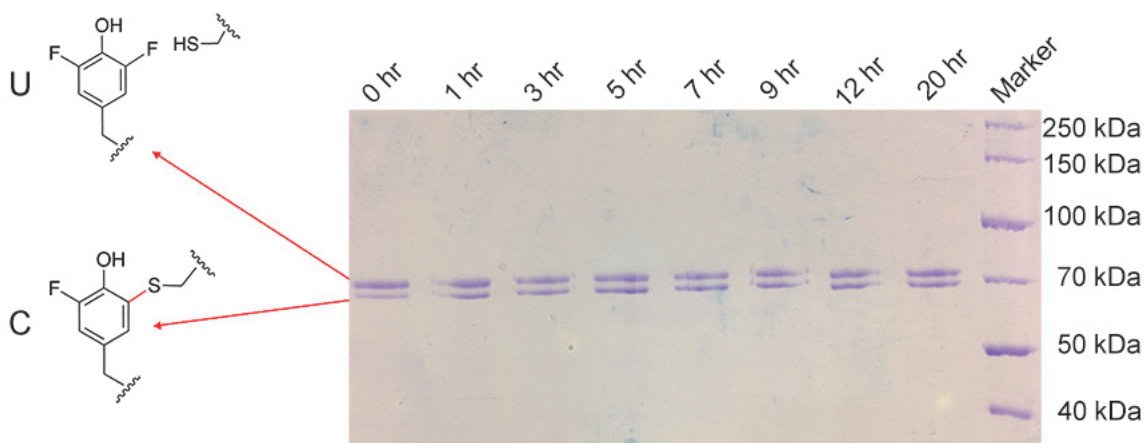
Figure S1. Conversion of as-isolated F$_2$-Tyr272 GAO^V upon incubation with Cu(I) and O$_2$. After purification and degassing, 1.5 eq of Cu(I) was added to GAO^V (10 μM) followed by addition of an equal volume of oxygen saturated buffer. Aliquots were taken at various times, and the reaction was quenched by dilution into SDS-containing buffer. Analysis by SDS-PAGE was replicated at least three times in independent experiments to ensure reproducibility. Time 0 corresponds to a sample quenched immediately (<10 s) after addition of oxygen-saturated buffer.
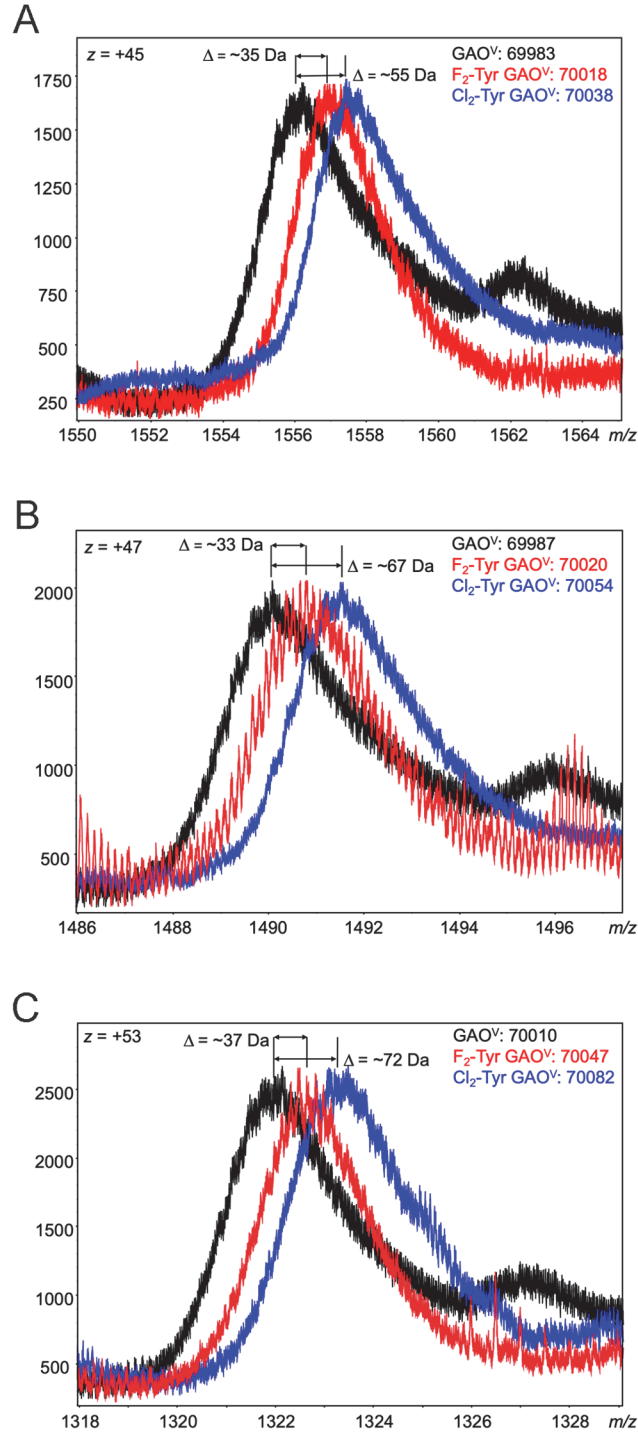
Figure S2. **ESI-MS analysis of GAO^V and unnatural tyrosine variants at different charge states**. Black, GAO^V; red, F$_2$-Tyr272 GAO^V; blue, Cl$_2$-Tyr272 GAO^V. The expected mass of GAO^V is 69,739 Da. The experimentally determined mass is larger due to nonspecific adducts and modifications such as oxidation of residues like methionine. Relative differences between the protein variants show that both F$_2$-Tyr272 GAO^V and Cl$_2$-Tyr272 GAO^V are mixtures of the crosslinked and uncrosslinked forms.
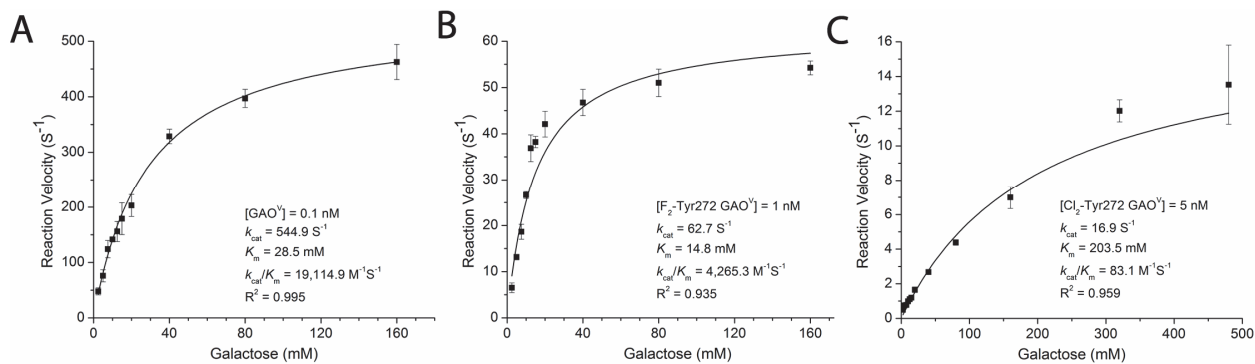
Figure S3. Catalytic activity assays of GAO^V and unnatural amino acid variants. The production of hydrogen peroxide (H$_2$O$_2$) was measured at various galactose concentrations for the following purified recombinant proteins: GAO^V, F$_2$-Tyr272 GAO^V (**B**), Cl$_2$-Tyr272 GAO^V (**C**). n = 3 independent experiments were repeated, and the data represent the mean value ± s.d.
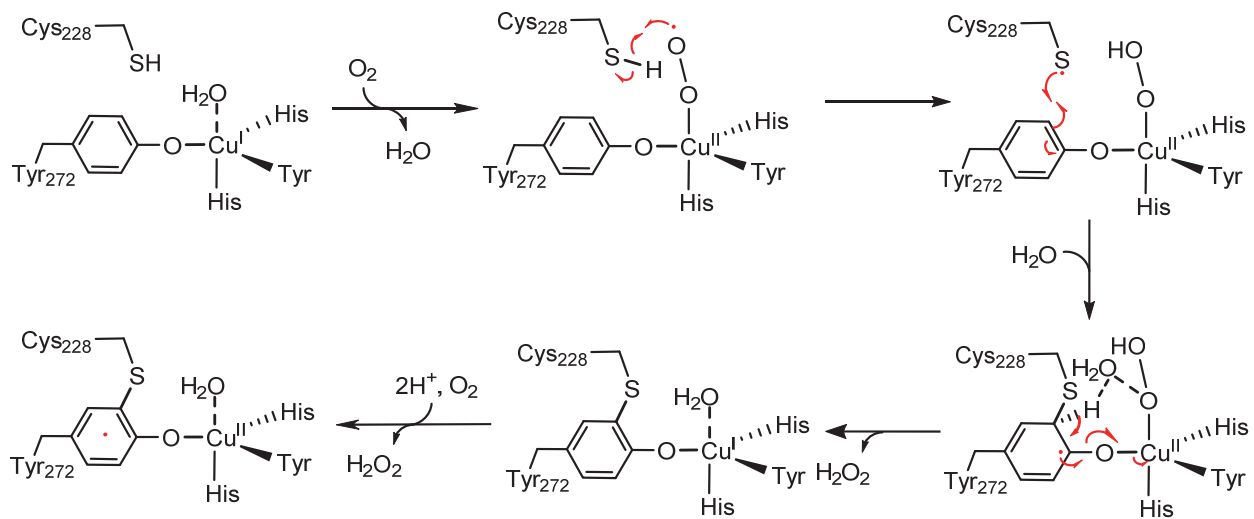
Figure S4. A proposed mechanistic model for cofactor biogenesis in GAO

References

(1) Li, J.; Griffith, W. P.; Davis, I.; Shin, I.; Wang, J.; Li, F.; Wang, Y.; Wherritt, D. J.; Liu, A., Cleavage of a carbon-fluorine bond by an engineered cysteine dioxygenase. *Nat. Chem. Biol.* **2018**, *14*, 853-860.

(2) Li, F.; Shi, P.; Li, J.; Yang, F.; Wang, T.; Zhang, W.; Gao, F.; Ding, W.; Li, D.; Li, J.; Xiong, Y.; Sun, J.; Gong, W.; Tian, C.; Wang, J., A genetically encoded ^{19}F NMR probe for tyrosine phosphorylation. *Angew. Chem. Int. Ed.* **2013**, *52*, 3958-62.

(3) Liu, X. H.; Jiang, L.; Li, J. S.; Wang, L.; Yu, Y.; Zhou, Q.; Lv, X. X.; Gong, W. M.; Lu, Y.; Wang, J. Y., Significant expansion of fluorescent protein sensing ability through the genetic incorporation of superior photo-induced electron-transfer quenchers. *J. Am. Chem. Soc.* **2014**, *136*, 13094-13097.

(4) Choosri, W.; Paukner, R.; Wuhrer, P.; Haltrich, D.; Leitner, C., Enhanced production of recombinant galactose oxidase from *Fusarium graminearum* in E. coli. *World J. Microbiol. Biotechnol.* **2011**, *27*, 1349-1353.

(5) Whittaker, M. M.; Whittaker, J. W., Cu(I)-dependent biogenesis of the galactose oxidase redox cofactor. *J. Biol. Chem.* **2003**, *278*, 22090-22101.

(6) Dornevil, K.; Davis, I.; Fielding, A. J.; Terrell, J. R.; Liu, A., Crosslinking of dicyclotyrosine by the cytochrome P450 enzyme CYP121 from *Mycobacterium tuberculosis* proceeds through a catalytic shunt pathway. *J. Biol. Chem.* **2017**, *292*, 13645-13657.

(7) Fielding, A. J.; Dornevil, K.; Ma, L.; Davis, I.; Liu, A., Probing ligand exchange in the P450 enzyme CYP121 from *Mycobacterium tuberculosis*: Dynamic equilibrium of the distal heme ligand as a function of pH and temperature. *J. Am. Chem. Soc.* **2017**, *139*, 17484-17499.

(8) Baron, A. J.; Stevens, C.; Wilmot, C.; Seneviratne, K. D.; Blakeley, V.; Dooley, D. M.; Phillips, S. E.; Knowles, P. F.; McPherson, M. J., Structure and mechanism of galactose oxidase. The free radical site. *J. Biol. Chem.* **1994**, *269*, 25095-25105.

(9) Otwinowski, Z.; Minor, W., Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1997**, *276*, 307-326.

(10) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallog. Sec. D-Biol. Crystallog.* **2010**, *66* (Pt 2), 213-221.

(11) Ito, N.; Phillips, S. E. V.; Stevens, C.; Ogel, Z. B.; Mcpherson, M. J.; Keen, J. N.; Yadav, K. D. S.; Knowles, P. F., Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase. *Nature* **1991**, *350*, 87-90.

(12) Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* **2004**, *60*, 2126-2132.

(13) Williams, C. J.; Headd, J. J.; Moriarty, N. W.; Prisant, M. G.; Videau, L. L.; Deis, L. N.; Verma, V.; Keedy, D. A.; Hintze, B. J.; Chen, V. B.; Jain, S.; Lewis, S. M.; Arendall, W. B., 3rd; Snoeyink, J.; Adams, P. D.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C., MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **2018**, *27*, 293-315.

(14) Lovell, S. C.; Davis, I. W.; Arendall, W. B., 3rd; de Bakker, P. I.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C., Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* **2003**, *50*, 437-450.