SUPPORTING INFORMATION

Structural Basis for the Role of Tyrosine 257 of Homoprotocatechuate 2,3-Dioxygenase in Substrate and Oxygen Activation

Elena G. Kovaleva[*,†] and John D. Lipscomb*[*,§]

[†]Institute of Molecular and Cellular Biology, University of Leeds, Leeds, LS2 9JT, UK

[§]Department of Biochemistry, Molecular Biology and Biophysics and Center for Metals in Biocatalysis, University of Minnesota, Minneapolis, Minnesota 55455, United States

Ligand Refinement and Molecular Modeling. Ligand refinement was initiated in the final stages of protein structure refinement by examining ligand-omit difference maps and using catecholic substrates with standard geometry as the initial models. Several factors were evaluated such as a) resolution and quality of the X-ray electron density data (e.g. weighting factor used in the refinement), and b) ligand occupancy in each subunit, including total and relative to the components typically present in the resting state. As illustrated in Figure S2, iron coordination sites for solvent (in the resting state) and aromatic substrate (in complex) differ slightly, and chloride anion is present in the anion-binding pocket in the resting state structures. Based on these markers, total observed ligand omit difference density in each active site was examined to determine the appropriate ligand refinement strategy. If total occupancy of aromatic substrate was estimated to be less than 50%, only components of the resting state were refined in the subunit. When higher occupancy of aromatic substrate (~ 60%) was estimated from the observed electron density in the subunit, refinement of a mixture was initiated (60% 4NC, 40% solvent/Cl^-), such as was the case for FeHPCD-4NC complex in subunit D (Figure S1B). However, in this case, refinement of strained conformation of 4NC was not warranted and standard geometry restraints were used in the refinement to avoid conformational bias due to significant presence of non-4NC components. When aromatic substrate is present in 70-100% occupancy, positions of substrate coordination sites to iron are well-defined as evidenced by the observed electron density (e.g. Figure 4, 6 and 7A), consistent with presence of a single major species in the active site. In these cases, following the initial refinement using standard ligand

geometry and evaluation of ligand omit electron density maps, refinement of strained conformations of HPCA and 4NC was initiated by gradually relaxing geometric restraints.

Model bias in refinement of side-on bound dioxygen in the Y257F-4NC-O_2 complex structure was evaluated using a single solvent at full occupancy refined in subunits C and D. This refinement protocol resulted in excess of positive $F_{obs}-F_{calc}$ electron density at a site (consistent with refined position of dioxygen in the final model), and temperature factors for refined solvent significantly lower than the average values. The $F_{obs}-F_{calc}$ difference map calculated when a single solvent at 100% occupancy is present in the O_2-binding site, instead of a 50:50 mixture of O_2 and solvent, is illustrated in Figure S5A,B.

The distribution, occupancy and conformation of the ligands reported here are representative of several (3-5) additional independently refined structural datasets obtained from different crystal complexes prepared using the same *in crystallo* reaction procedures as described in the Experimental Procedures.

Structural Comparisons. Additional structural comparisons are shown in Figures S2-4 and 6. Active site comparison of the resting state structures and complexes with substrates (HPCA and 4NC) illustrate negligible effects that binding of substrates have on structural integrity of the residues in the active site and in the substrate binding cleft for both FeHPCD and Y257F enzymes (Figure S2). Similarly, comparison of the active site structures of FeHPCD and Y257F in complex with either optimal substrate HPCA (Figure S3A) or alternative substrate 4NC (Figure S3B) shows that substitution of Tyr257 for Phe has no effect on positions of either the residues in the active site or on the binding mode of iron-bound substrate. As depicted in Figure S4, HPCA and 4NC coordinate in the active sites of FeHPCD and Y257F enzymes in similar orientation, including resultant positions of the aromatic ring despite differences in steric and electrostatic properties of aromatic ring substituents in the 4 position. Comparison of the refined conformations of 4NC in the active sites of FeHPCD and Y257F before and after exposure to O_2 is shown in Figure S6. These illustrate that in the case of the Y257F enzyme, the conformation of 4NC remains unchanged (Figure S6A), whereas significant localized distortion of the substrate ring (at C2) is observed for FeHPCD-4NC complex in the presence of O_2, consistent with formation of the FeHPCD-[4NC-semiquinone]-$O_2^{\bullet-}$ intermediate (Figure S6B).

Table S1. X-ray data collection and refinement statistics for anaerobic complexes of FeHPCD with HPCA and 4NC [a].

Dataset (PDB Code)	FeHPCD-HPCA (PDB 4GHG)	FeHPCD-4NC (PDB 4GHH)
Wavelength	0.9801 Å	0.9801 Å
Synchrotron (beamline)	Synchrotron Soleil (Proxima I)	Synchrotron Soleil (Proxima I)
Spacegroup	$P2_12_12$	$P2_12_12$
Cell dimensions (Å)	110.5, 150.8, 96.3	110.2, 150.5, 96.2
Cell angles (deg)	90, 90, 90	90, 90, 90
Resolution range[a] (Å)	48.1 – 1.50 (1.58)	48.1 – 1.55 (1.63)
Reflections (observed/unique)	1132656/252890	1360219/231042
R_{rim} [a, b] (%)	5.9 (74.0)	8.2 (87.3)
R_{pim} [a, c] (%)	3.0 (38.4)	3.7 (41.3)
Mean $\langle I\rangle/\sigma\langle I\rangle$ [a]	14.5 (2.3)	12.1 (2.0)
Completeness (%) [a]	98.9 (98.7)	99.9 (99.8)
R, R_{free}, test (%) [d]	12.7, 16.1, 5.0	12.5, 16.4, 5.0
RMSD [e] bond length (Å)	0.009	0.012
RMSD [e] angles (deg)	1.373	1.477
ESU [f] (Å)	0.038	0.044
Ramachandran Plot		
Allowed regions (%)	99.7	99.7
Additional regions (%)	0.3	0.3

[a] Values for the highest resolution shell are given in parentheses. [b] Redundancy-independent merging R factor: $R_{r.i.m} = \sum_{hkl}[N/(N-1)]^{1/2}\sum_i|I_i(hkl)-\langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the mean value of $I(hkl)$ [S1]. [c] Precision-indicating merging R factor: $R_{p.i.m} = \sum_{hkl}[1/(N-1)]^{1/2} \sum_i|I_i(hkl)-\langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the mean value of $I(hkl)$ [S1]. [d] $R = (\sum|F_{obs}-kF_{calc}|)/\sum|F_{obs}|$, where k is a scale factor. The R_{free} value was calculated with the indicated percentage of reflections not used in the refinement. [e] Root-mean-square deviation (RMSD) from ideal geometry in the final models. [f] Estimated overall coordinate error (ESU) based on maximum likelihood.

Table S2. X-ray data collection and refinement statistics for Y257F in free form, anaerobic complexes with HPCA and 4-NC, and aerobic complex with 4NC [a].

	Free form	anaerobic	anaerobic	reacted with O_2
Dataset	Y257F	Y257F-HPCA	Y257F-4NC	Y257F-4NC
(PDB Code)	(PDB 4GHC)	(PDB 4GHD)	(PDB 4GHE)	(PDB 4GHF)
Wavelength	0.9790 Å	0.9801 Å	0.9801 Å	0.9999 Å
Synchrotron (beamline)	APS (19-BM)	Synchrotron Soleil (Proxima I)	Synchrotron Soleil (Proxima I)	SLS (X10SA)
Spacegroup	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$
Cell dimensions (Å)	110.4, 151.7, 96.3	110.2, 150.7, 96.1	110.5, 150.1, 96.2	110.5, 150.3, 96.1
Cell angles (deg)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution range[a] (Å)	22.5 – 1.55 (1.61)	48.1 – 1.85 (1.95)	42.3 – 1.60 (1.69)	89.0 – 1.67 (1.76)
Reflections (observed/unique)	1389275/233050	792903/134568	686954/207338	874975/184533
R_{rim} [a, b] (%)	8.5 (56.1)[g]	10.0 (95.8)	6.5 (65.7)	6.8 (85.1)
R_{pim} [a, c] (%)		4.5 (42.4)	4.1 (40.9)	3.5 (42.7)
Mean $\langle I\rangle/\sigma\langle I\rangle$ [a]	17.8 (1.8)	13.4 (1.9)	12.8 (2.1)	13.5 (2.0)
Completeness (%) [a]	99.5 (98.0)	98.6 (97.3)	98.8 (99.1)	99.6 (99.5)
R, R_{free}, test (%) [d]	11.7, 14.8, 5.0	15.2, 18.9, 5.0	15.5, 18.6, 5.0	15.1, 18.1, 5.0
RMSD [e] bond length (Å)	0.010	0.013	0.012	0.0146
RMSD [e] angles (deg)	1.415	1.541	1.520	1.616
ESU [f] (Å)	0.036	0.080	0.058	0.060
Ramachandran Plot				
Allowed regions (%)	99.7	99.7	99.7	99.7
Additional regions (%)	0.3	0.3	0.3	0.3

[a] Values for the highest resolution shell are given in parentheses. [b] Redundancy-independent merging R factor: $R_{r.i.m} = \sum_{hkl}[N/(N-1)]^{1/2}\sum_i|I_i(hkl)-\langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the mean value of $I(hkl)$ [S1]. [c] Precision-indicating merging R factor: $R_{p.i.m} = \sum_{hkl}[1/(N-1)]^{1/2} \sum_i|I_i(hkl)-\langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the mean value of $I(hkl)$ [S1]. [d] $R = (\sum||F_{obs}|-k|F_{calc}||)/\sum|F_{obs}|$, where k is a scale factor. The R_{free} value was calculated with the indicated percentage of reflections not used in the refinement. [e] Root-mean-square deviation (RMSD) from ideal geometry in the final models. [f] Estimated overall coordinate error (ESU) based on maximum likelihood. [g] $R_{merge} = \sum|I_i - \langle I\rangle|/\sum|I_i|$, where I_i is the integrated intensity of a given reflection, and $\langle I\rangle$ is the mean value for that reflection [S2].

Table S3. Comparison of selected distances for FeHPCD and Y257F enzymes in the resting state and in complex with substrates.

	Resting state		Enzyme-HPCA		Enzyme-4NC		Enzyme-4NC+O_2	
	FeHPCD[a]	Y257F[b]	FeHPCD[c]	Y257F[d]	FeHPCD[e]	Y257F[f]	FeHPCD[g]	Y257F[h]
Fe-H155 (Å)	2.24 ± 0.07	2.19 ± 0.02	2.18	2.22	2.19 ± 0.00	2.24	2.23	2.21 ± 0.04
Fe-H214 (Å)	2.23 ± 0.04	2.17 ± 0.01	2.16	2.26	2.16 ± 0.02	2.18	2.29	2.16 ± 0.03
Fe-E267 (Å)	2.04 ± 0.02	2.05 ± 0.02	2.01	2.10	2.04 ± 0.01	2.01	2.09	2.04 ± 0.03
Fe-Wat1[i] (Å)	2.23 ± 0.06	2.20 ± 0.03						
Fe-Wat2[i] (Å)	2.19 ± 0.05	2.27 ± 0.01						
Fe-Wat3[i] (Å)	2.14 ± 0.08	2.13 ± 0.03	2.43	2.36	2.39 ± 0.06	2.37		(2.28 ± 0.04)[j]
Fe-$O_{C1\text{-}LIGAND}$ (Å)			2.19	2.22	2.13 ± 0.01	2.15	2.16	2.15 ± 0.02
Fe-$O_{C2\text{-}LIGAND}$ (Å)			2.09	2.05	2.17 ± 0.01	2.10	2.24	2.06 ± 0.06
$C1_{SUB}$-$O_{C1\text{-}LIGAND}$ (Å)			1.38	1.38	1.37 ± 0.00	1.37	1.27	1.36 ± 0.01
$C2_{SUB}$-$O_{C2\text{-}LIGAND}$ (Å)			1.36	1.34	1.34 ± 0.00	1.34	1.41	1.33 ± 0.01
$O_{C1\text{-}SUB}$-$N(\varepsilon2)_{H200}$ (Å)			3.05	3.16	2.98 ± 0.07	3.13	3.21	3.08 ± 0.09
$O_{C2\text{-}SUB}$-O_{Y257} (Å)			2.57		2.54 ± 0.08		2.51	
$C2_{SUB}$-O_{Y257} (Å)			2.95		2.94 ± 0.00		3.24	
Fe-$O1_{OXY}$ (Å)							2.53	2.8 ± 0.2
Fe-$O2_{OXY}$ (Å)							2.43	2.4 ± 0.1
$O1_{OXY}$-$O2_{OXY}$ (Å)							1.30	1.20[j]

[a] average values and standard deviations for distances calculated for 4 monomeric subunits of FeHPCD (PDB 3OJT and 2IG9). [b] average values and standard deviations for distances calculated for 4 monomeric subunits of Y257F variant (PDB 4GHC). [c] values for distances for FeHPCD-HPCA complex (PDB 4GHG, subunit C). [d] values for distances for Y257F-HPCA complex (PDB 4GHD, subunit C). [e] average values and standard deviations for distances calculated for FeHPCD-4NC complex (PDB 4GHH, subunits C and D). [f] values for distances for Y257F-4NC complex (PDB 4GHE, subunit C). [g] values for distances for FeHPCD-[4NC-semiquinone]- $O_2^{\bullet-}$ complex (PDB 2IGA, subunit C). [h] average values and standard deviations for distances calculated Y257F-4NC-O_2 complex (PDB 4GHF, subunits C and D). [i] Wat1, Wat2 and Wat3 designate solvent-derived ligands coordinated to the iron in the resting state of the enzyme in the positions trans to the residues H214, H155 and E267, respectively. In the high-resolution structures of all 2,3-HPCD enzymes, slightly asymmetric rather than spherical electron density is apparent at Wat2 coordination sites. Since no small ligand accounting for the observed density is present in purification or crystallization solutions, a single solvent was modeled at these sites. [j] A 50:50 mixture of dioxygen and solvent refined in subunits C and D.

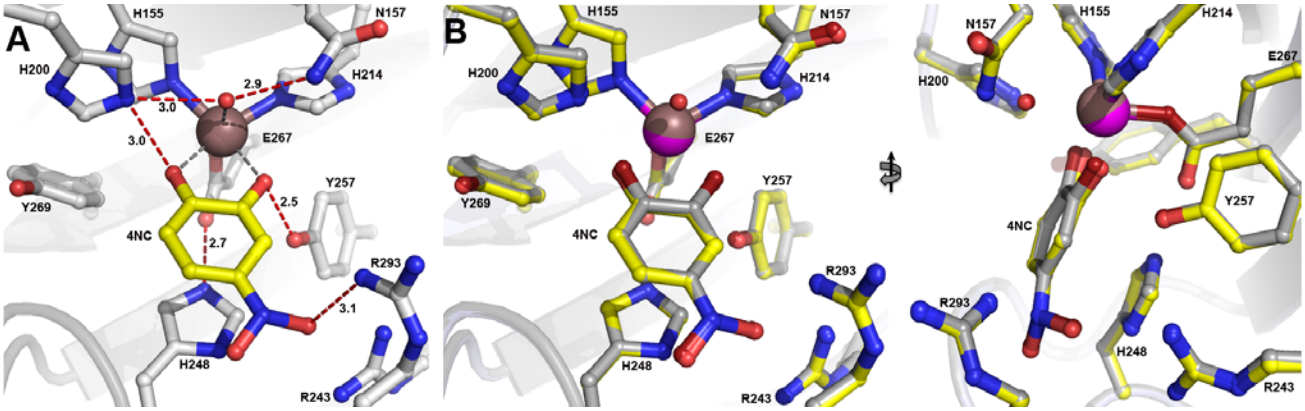

Figure S1. Active site environment of FeHPCD in complex with the alternative substrate 4NC (PDB 4GHH). (A) Hydrogen bonding interactions between bound 4NC and residues in the active site of subunit C (same model as used in Figure 6B in the main text). Atom color code: gray, carbon (enzyme); yellow, carbon (4NC); blue, nitrogen; red, oxygen; bronze, iron. Red dashed lines show hydrogen bonds (Å). Gray dashed lines indicate bonds or potential bonds to iron (Å). Additional distances are given in Table S3. The cartoon depicts the secondary structure elements. (B) Front and side views for comparison between active sites environments for FeHPCD-4NC complex in subunits C and D. The 4NC occupancy in subunit D was estimated to be only 60%, and refinement of strained conformation was not attempted to avoid bias due to solvent. Atom color code: dark gray, carbon (subunit C); gray, carbon (subunit D); blue, nitrogen; dark red, oxygen (subunit C); red, oxygen (subunit D); bronze, iron (subunit D); purple, iron (subunit C).

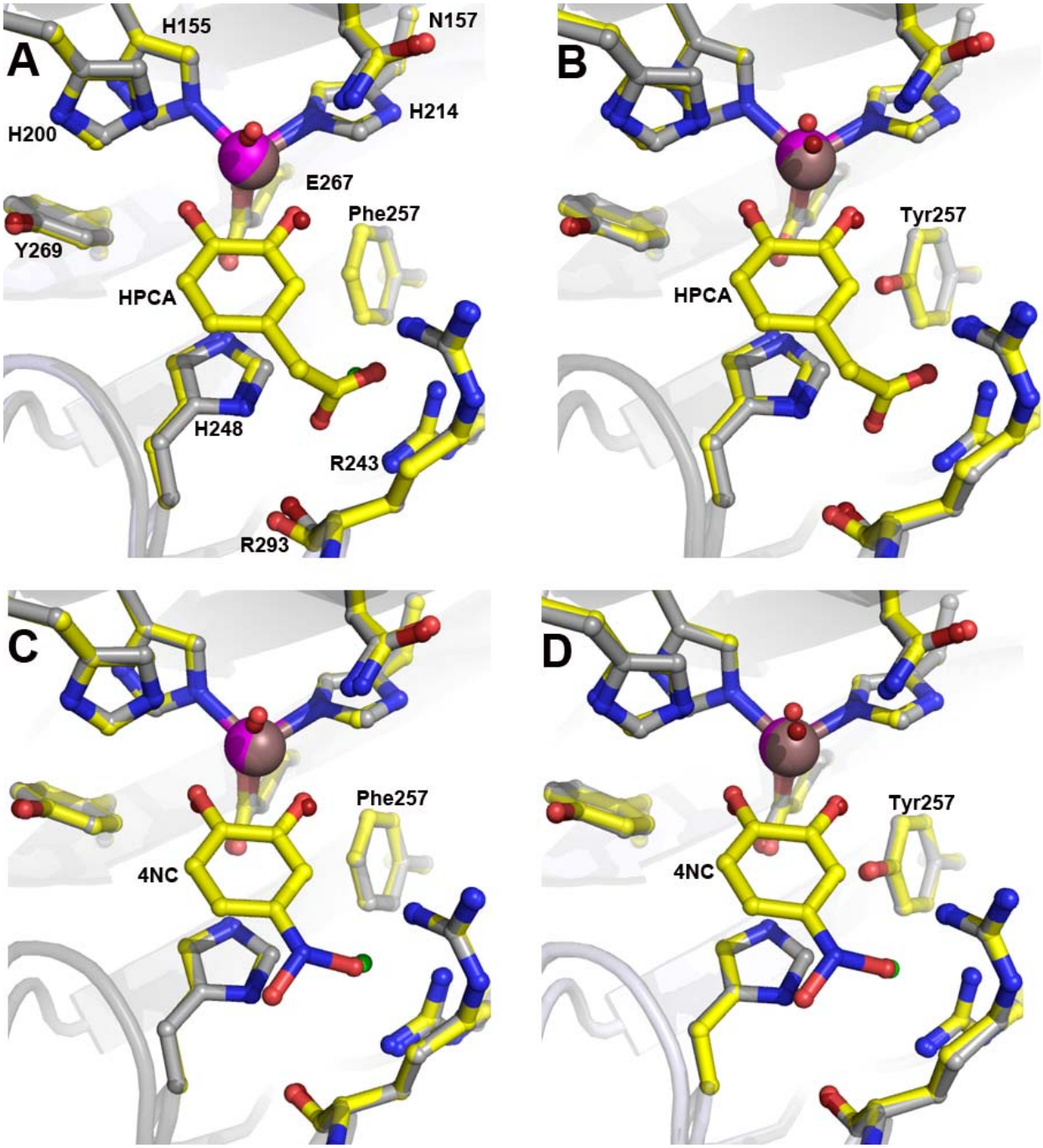

Figure S2. Active site comparison of FeHPCD and Y257F in the resting state and in complex with aromatic substrates (subunits C shown). (A) Y257F variant (PDB 4GHC) and Y257F-HPCA complex (PDB 4GHD). (B) FeHPCD (PDB 3OJT), FeHPCD-HPCA complex (PDB 4GHG). (C) Y257F variant (PDB 4GHC), Y257F-4NC complex (PDB 4GHE). (D) FeHPCD (PDB 3OJT), FeHPCD-4NC complex (PDB 4GHH). Atom color code: gray, carbon (resting state); yellow, carbon (complex); blue, nitrogen; dark red, oxygen (resting state); red, oxygen (complex); bronze, iron (resting state); purple, iron (complex); green, chlorine (resting state).

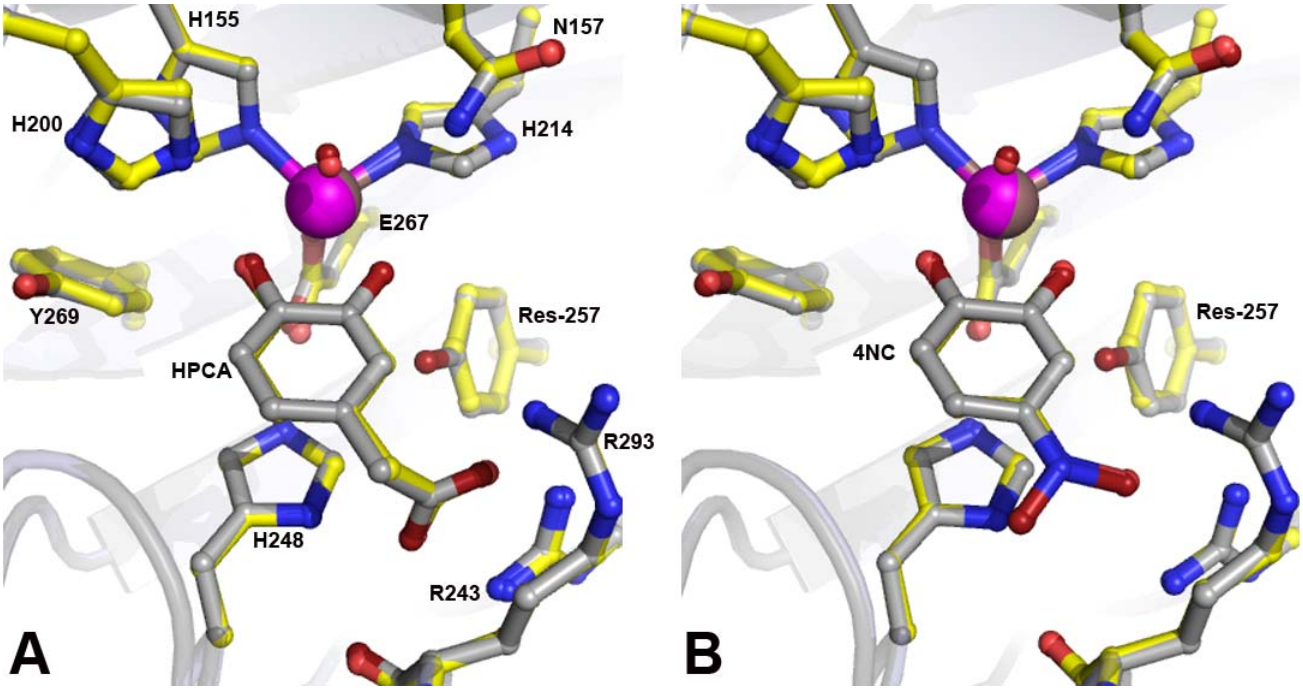

Figure S3. Additional views for comparison of the active site environments in FeHPCD and Y257F, in complex with the same aromatic substrates. The overlays show that substitution of Tyr257 for Phe has no effect on the conformation of residues in the substrate-binding cleft or on binding mode of each substrate. (A) Comparison of the Y257F-HPCA (PDB 4GHD) and FeHPCD-HPCA (PDB 4GHG) complex structures, shown in rotated view from that in Figure 3B of main text. (B) Comparison of the Y257F-4NC (PDB 4GHE) and FeHPCD-4NC (PDB 4GHH) complex structures, shown in rotated view from that in Figure 5B of main text. Atom color code: gray, carbon (FeHPCD); yellow, carbon (Y257F); blue, nitrogen; dark red, oxygen (FeHPCD); red, oxygen (Y257F); bronze, iron (FeHPCD); purple, iron (Y257F).

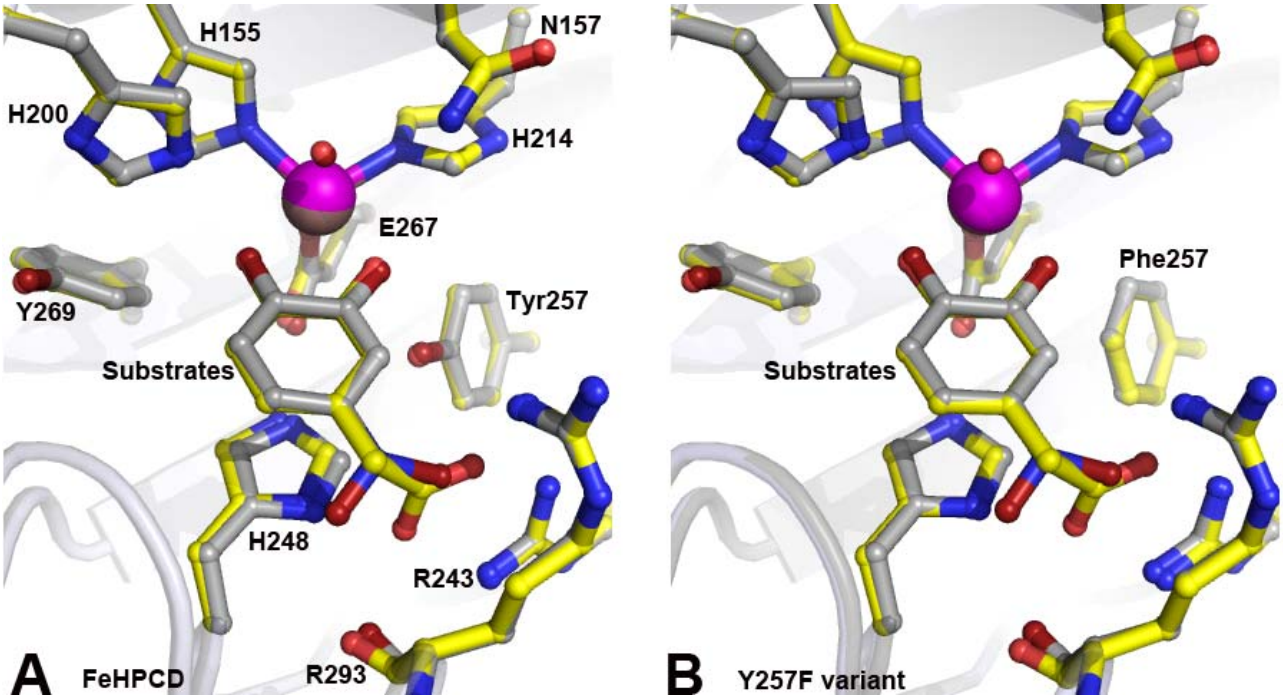

Figure S4. Comparison of active site environment of FeHPCD (A) and Y257F variant (B) in complex with HPCA and 4NC substrates. These show that binding of the optimal and alternative substrates does not elicit significant changes in the conformation of residues in the substrate-binding cleft or metal coordination of the catecholic substrates. A small shift in position of H248 and carbonyl moiety of R293 occurs, which is associated with optimal positioning of the longer substituent of HPCA. Atom color code: gray, carbon (complex with 4NC); yellow, carbon (complex with HPCA); dark red, oxygen (complex with 4NC); red, oxygen (complex with HPCA); blue, nitrogen; bronze, iron (complex with 4NC); purple, iron (complex with HPCA).

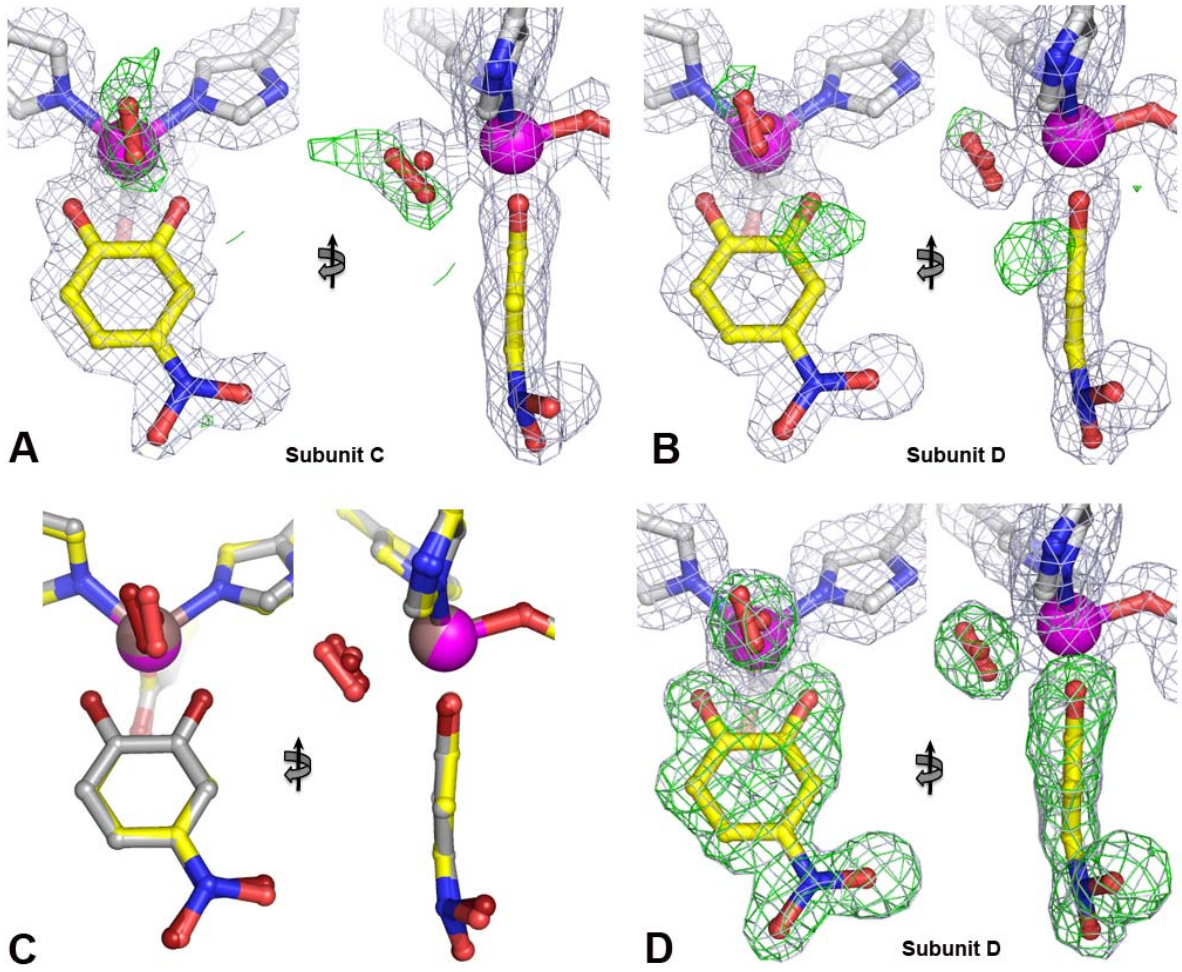

Figure S5. Models and X-ray electron density maps for Y257F-4NC-O_2 complex structures in subunits C and D (PDB 4GHF). Y257F-4NC-O_2 complex observed in subunits C (panel A) and D (panel B), with a 50:50 mixture of dioxygen and solvent refined in the O_2-binding site. The blue $2F_{obs}-F_{calc}$ electron density map is contoured at 1 σ. The green $F_{obs}-F_{calc}$ difference map was calculated when single solvent at 100% occupancy is present in the O_2-binding site of the model and it is contoured at 3.5 σ. Atom color code: gray, carbon (enzyme); yellow, carbon (4NC); blue, nitrogen; red, oxygen; purple, iron. (C) Comparison of Y257F-4NC-O_2 structures observed in subunits C and D. Atom color code: gray, carbon (subunit D); yellow, carbon (subunit C); blue, nitrogen; dark red, oxygen (subunit D); red, oxygen (subunit C); bronze, iron (subunit D); purple, iron (subunit C). (D) Front and side views for complex observed in subunit D (same panel as Figure 7A for subunit C in the main text). The blue $2F_{obs}-F_{calc}$ electron density map is contoured at 1 σ. The green ligand-omit $F_{obs}-F_{calc}$ difference map was calculated in the absence of all substrates in the final model and it is contoured at 6.5 σ. Atom color code: same as for panels A and B.

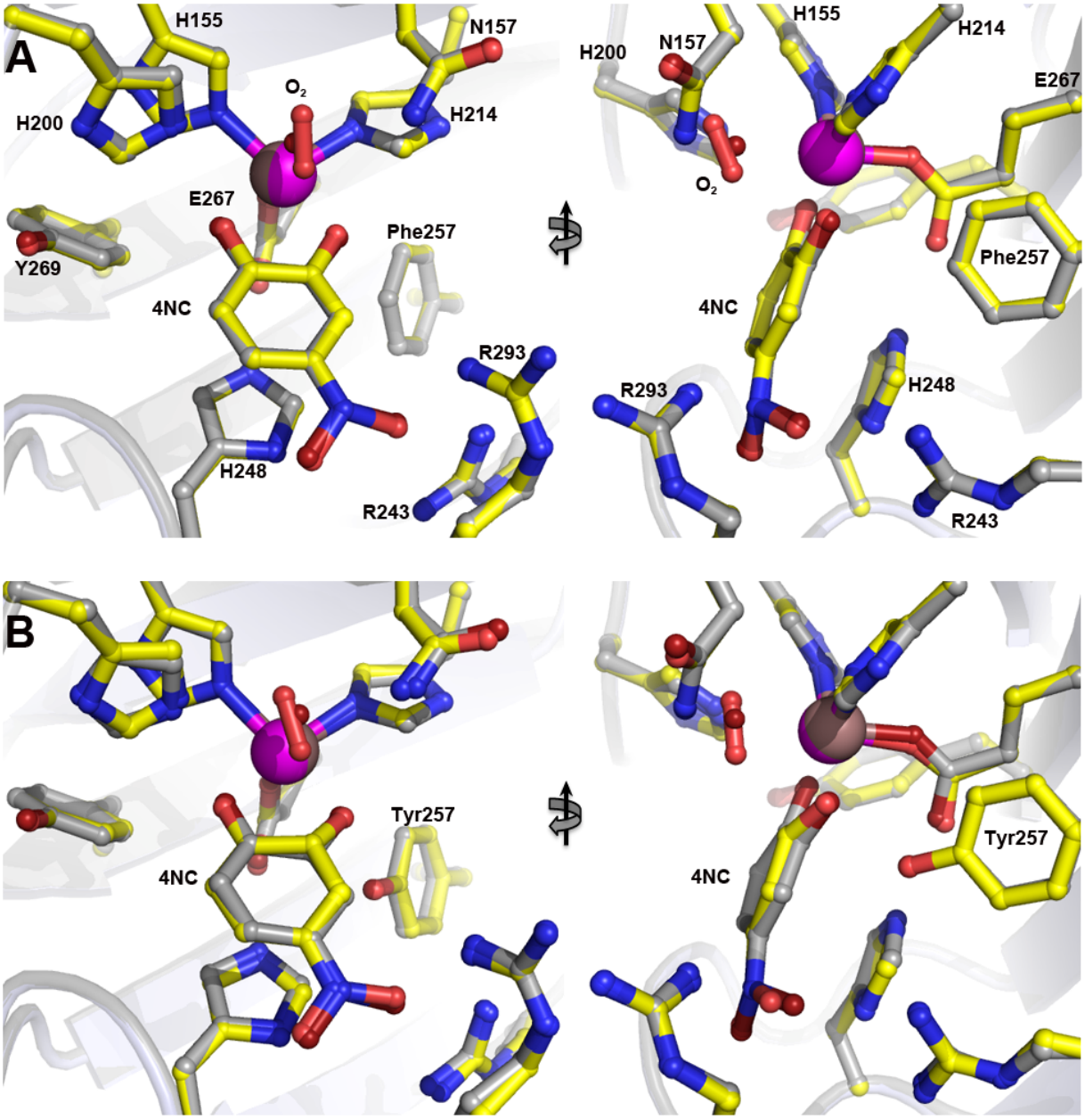

Figure S6. Comparison of 4NC conformations in the active sites of FeHPCD and Y257F enzymes before and after exposure to O_2 *in crystallo* (subunits C shown). (A) Structure overlay for Y257F-4NC complex (PDB 4GHE) and Y257F-4NC-oxy complex (PDB 4GHF). (B) Structure overlay for FeHPCD-4NC complex (PDB 4GHH) and FeHPCD-[4NC-semiquinone]-$O_2^{\bullet -}$ complex (PDB 2IGA). Atom color code: gray, carbon (enzyme-4NC complex); yellow, carbon (enzyme-4NC-oxy complex); dark red, oxygen (enzyme-4NC complex); red, oxygen (enzyme-4NC-oxy complex); blue, nitrogen; bronze, iron (enzyme-4NC complex); purple, iron (enzyme-4NC-oxy complex).

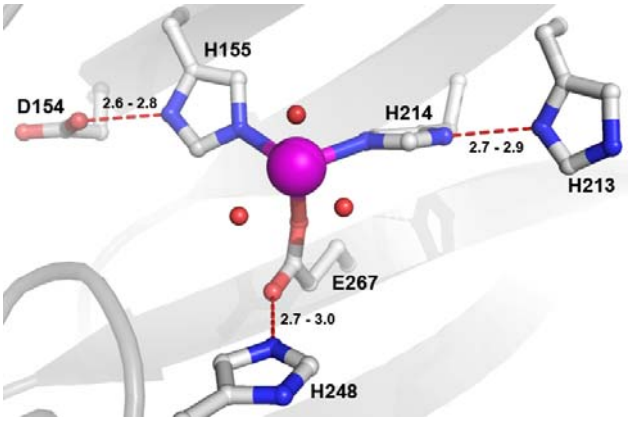

Figure S7. Representative structure of the Fe^{II}-2-His-1-carboxylate triad in the active site of FeHPCD showing second sphere residues involved in stabilizing hydrogen bonding interactions with the first coordination sphere ligands. Atom color code: gray, carbon; blue, nitrogen; red, oxygen; purple, iron. Red dashed lines show hydrogen bonds (Å), indicating the range of observed distances within the 4 subunits of both FeHPCD (PDB 3OJT) and Y257F (PDB 4GHC) enzymes in the resting state.

REFERENCES

(S1) Weiss, M. S. (2001) Global indicators of X-ray data quality. *J. Appl. Cryst. 34*, 130-135.

(S2) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A 276*, 307-326.