

# Supplementary Information

## Structural and Stereoelectronic Insights into Oxygenase Catalyzed Formation of Ethylene from 2-Oxoglutarate

Zhihong Zhang<sup>‡</sup>, Tristan J. Smart<sup>‡</sup>, Hwanho Choi<sup>‡</sup>, Florence Hardy<sup>‡</sup>, Christopher T. Lohans<sup>‡</sup>, Martine I. Abboud<sup>‡</sup>, Melodie S. W. Richardson<sup>‡</sup>, Robert S. Paton<sup>‡</sup>, Michael A. McDonough<sup>‡\*</sup>, Christopher J. Schofield<sup>‡\*</sup>

<sup>‡</sup>Department of Chemistry, University of Oxford, 12 Mansfield Road, Oxford, OX1 3TA, United Kingdom.

\***Corresponding Authors:** Prof. Christopher J. Schofield, christopher.schofield@chem.ox.ac.uk and Dr. Michael A. McDonough, michael.mcdonough@chem.ox.ac.uk.

## Table of Contents

### Supplementary Methods

1. *Protein production and purification*
2. *PsEFE variants*
3. *Crystallography*
  - a. *PsEFE co-crystallization with BTP or 2OG (I222 crystal form)*
  - b. *PsEFE co-crystallization with NOG and L-Arg (P1 crystal form)*
  - c. *SAD structure solution, model building and refinement*
4. *Enzyme activity measurements*
  - a. *GC/MS assays for ethylene*
  - b. *Steady state kinetics*
  - c. *NMR Spectroscopy*
5. *Computational Chemistry*
  - a. *MD simulation methods*
  - b. *Quantum mechanical calculations*

## Supplementary Tables

1. *Data collection and refinement statistics.*
2. *RMSD of protein chains*
3. *Ethylene production by PsEFE variants.*
4. *Differences in the transition energies for Grob fragmentations.*

## Supplementary Figures

1. *GC-MS analysis of PsEFE catalysis.*
2. *LC-MS detection of pyrroline-5-carboxylate (P5C).*
3. *NMR analysis of PsEFE-catalyzed formation of pyrroline-5-carboxylic acid (P5C).*
4. *Arginine analogues tested for PsEFE substrate activity.*
5. *Secondary structure based sequence alignment of PsEFE.*
6. *Structural comparison of PsEFE with ACCO and ANS.*
7. *Structural changes observed upon ligand binding to PsEFE.*
8. *Views of the  $2mF_o - DF_c$  electron density maps from the PsEFE.Fe.NOG.L-Arg complex structure for each of the four protein molecules in the asymmetric unit.*
9. *Comparison of the substrate binding mode of PsEFE with those of other 2OG oxygenases.*
10. *The hydrophobic binding pocket for 2OG in PsEFE.*
11. *Views of 2-oxoglutarate (2OG) binding modes in selected 2OG oxygenase structures.*
12. *Time-course evolutions of the torsion angle between the O3-C5-O4 plane and the C3-C4-C5 plane of 2OG.*
13. *Analysis of Grob type fragmentation*
14. *Sequence alignment of PsEFE and related enzymes*
15. *Sequence phylogeny of PsEFE and related enzymes*

## Supplementary Methods

### 1. Protein production and purification

A synthetic *PsEFE* DNA sequence optimized for expression in *E. coli* based on the EFE gene sequence from *Pseudomonas syringae* pv. phaseolicola (UNIPROT ID P32021) was obtained (GeneArt ThermoFisher). The synthetic *PsEFE* DNA was PCR amplified and subcloned into the pCold-1 vector (Clontech), encoding for *N*-terminally His-tagged *PsEFE*, and the pETite vector (Lucigen), encoding for *N*-terminally His-SUMO-tagged *PsEFE*. The sequences of the plasmids were verified by DNA sequencing. The plasmids were transformed into *E. coli* BL21(DE3). After optimizing for *PsEFE* production at different temperatures on a small scale (100 mL), large scale cultures were carried out for production of protein for crystallography and biochemical characterization. Large scale growths were carried out in 2xTY medium supplemented with the appropriate antibiotic concentration using a shaker incubating at (37°C, 200 rpm). The temperature was lowered to 18°C when the OD<sub>600</sub> of the cells reached 0.8. After incubation at 18°C for 40 min, isopropyl-β-*D*-thiogalactopyranoside (IPTG) was added to a final concentration of 0.5 mM to induce protein production. The induced culture was incubated overnight at 18°C, then harvested by centrifugation. The cell paste was stored at -80°C prior to protein purification. Seleno-L-methionine (SeMet) substituted *PsEFE* was produced using LeMaster medium containing 50 μg racemic seleno- methionine (1). The growth conditions for SeMet *PsEFE* were the same as those for growing wild type *PsEFE*.

The frozen cell paste was thawed and re-suspended into binding buffer (20 mM Tris pH7.8, 20 mM imidazole, 500 mM NaCl, 4°C) with a 1 g cell mass : 4 mL buffer ratio. Cells were lysed by sonication (5x10 s intervals with 20 s pause between intervals) while being cooled using a dry-ice-ethanol bath using a Sonics Vibra-Cell sonicator equipped with a micro tip. Unbroken cells and cell debris were removed by centrifugation (23,000 rpm, Beckman-Coulter JA25.50 rotor, 4°C, 20 min). The supernatant was loaded onto a His-Trap (GE Healthcare Life Sciences) column pre-equilibrated with binding buffer. Contaminating proteins were eluted by increasing the concentration of imidazole to 30 mM. *PsEFE* was eluted with Eluting Buffer (20 mM Tris-HCl pH 7.8, 500 mM imidazole, 500

mM NaCl, 22°C). Fractions containing *PsEFE* were pooled and concentrated using an Amicon spin concentrator, then buffer exchanged using a PD-10 desalting column (GE Healthcare Life Sciences) into protease cleavage buffer (20 mM Tris-HCl, 150 mM NaCl, 10% glycerol, pH 8.0). The *N*-terminal His-SUMO tag was cleaved by adding human SenP2 protease to the purified *PsEFE* protein at a 1:1500 ratio and incubated at 4°C for 5 h on a wheel mixer (Stuart Rotator, Keison Products, UK). The *N*-terminal His-SUMO-tag was removed using His-Trap (GE Healthcare Life Sciences) chromatography. The unbound-eluted *PsEFE* was concentrated and loaded onto a Superdex S75 (GE Healthcare Life Sciences) gel filtration column that had been pre-equilibrated with 25 mM Tris-HCl pH 7.8. *PsEFE* was eluted using the same buffer. Fractions containing *PsEFE* were pooled, concentrated and stored at -80°C.

## **2. *PsEFE* variants**

*PsEFE* variants were made by site-directed mutagenesis using the polymerase chain reaction (PCR). The synthetic *PsEFE* gene (GeneArt, ThermoFisher) in a pCold-1 vector was used as the template for producing variants. Protein production and purification of variants was performed as for wildtype *PsEFE*.

## **3. Crystallography**

### **a. *PsEFE* crystallization (I222 crystal form)**

For crystallization in the *PsEFE* I222 crystal form, sitting drop vapour diffusion high-throughput crystallization trials were set up at 20°C using a Phoenix/RE automated liquid dispenser in 96-well/3-subwell low profile Intelliplates (Art Robbins). Optimization of a hit condition [PACT Premier G10 (Molecular Dimensions)] resulted in diffraction quality crystals that were obtained by hanging drop vapour diffusion in 24-well Linbro plates using plastic coverslips (3 µL drop/0.5 mL well) (3 µL drop; 2 µL protein:1 µL well solution) for full-length *PsEFE* native or the selenomethionine derivative (protein conc. 12-20 mg·mL<sup>-1</sup>) produced using 6X-His-tagged *PsEFE*, using the following crystallization conditions: 17.5-25% (w/v) polyethylene glycol (PEG) 3350 and 6000, 0.1 M bis-tris-propane (BTP) pH 7.0, 0.04 M sodium/potassium phosphate, 3-5 mM manganese (II) chloride.

Mn(II) was used as an Fe(II) surrogate. *PsEFE:Mn:2OG* crystals were grown with the addition of 10mM 2OG to the above condition.

#### **b. *PsEFE* co-crystallization with NOG and L-Arg (*P1* crystal form)**

For co-crystallization of *PsEFE* with *N*-oxalylglycine (NOG) and L-Arg, optimization of a hit (PACT Premier D7) from the sitting drop high-throughput crystallization trials at 20°C by addition of 10 mM NOG and 20 mM L-Arg resulted in well diffracting crystals for full-length *PsEFE* (protein conc. 15-25 mg·mL<sup>-1</sup>) produced using the *N*-terminally His-SUMO tagged construct; crystallization conditions: 0.2 M sodium chloride, 0.1 M Tris-HCl pH 8.0 and 20% (w/v) PEG 6000. This condition was optimized by microseeding using the hanging drop method and 24-well Linbro plates 3 µL drop (2 parts protein : 1 part well solution) and 0.5 mL well volumes. Crystals were cryo-protected with glycerol (20-25% (v/v) glycerol added to well solution), then harvested using nylon loops mounted on SPINE pins (Hampton Research) and plunged into, then stored in liquid nitrogen until data collection. Crystals were mounted in a cryo-stream 100K and data were collected at the Diamond Synchrotron Light Source (Didcot UK) on beamlines I04-1 and I02.

#### **c. SAD structure solution, model building and refinement**

High-resolution data (1.67 Å) were collected at Diamond beamline I04-1 in fine slice mode (0.2° oscillation, 0.044 s exposure, beamsizes 60 x 50 µm, 44% transmission, wavelength 0.9174 Å, 1800 images for a 360° total sweep) using a selenomethione *PsEFE* derivative single crystal (75 x 30 µm). The fixed wavelength at beamline I04-1 was sufficiently close to the selenium K edge (0.9795 Å) to measure single wavelength anomalous difference (SAD) data (measured  $\Delta F/F$  0.107) and to identify a substructure as indicated by the automated beamline data processing software (FastDP and FastEP) while at the beamline to 2.0 Å resolution. The following day, analysis of the more computationally extensive xia2 auto-processed data in iSpyB (2) enabled higher quality data to 1.67 Å to be extracted (Table S1)(3-5). These data were then input into AUTOSOL (6) in PHENIX (7) which resulted in the identification of 11 of 9 possible selenium sites in the substructure with a FOM of 0.33 (the two additional sites were later found to originate from the bound active site Mn and an alternative conformation for a selenomethionine side chain). The initial phases were then improved

by density modification in RESOLVE (8) (FOM of 0.74) and AUTOBUILD (9). Refinement resulted in a preliminary model containing 328 out of 350 residues with  $R_{\text{work}}$  0.2064 and  $R_{\text{free}}$  0.2348.

High-resolution data (1.55 Å) were collected at Diamond beamline I04-1 in fine slice mode (0.2° oscillation, 0.2 s exposure, beamsize 60x50 μm, 100% transmission, wavelength 0.9174 Å, 1800 images for a 360° total sweep) using a wildtype *PsEFE* crystal (200 x 75 μm). The model derived from the selenomethionine structure solution was then refined against the isomorphous higher resolution native dataset and manual cycles of model fitting in COOT (10) and refinement in PHENIX (7) continued until decreasing  $R_{\text{work}}$  and  $R_{\text{free}}$  no longer converged (Table 1).

Ultra-high-resolution data (1.08 Å) were also collected on Diamond beamline I02 in fine slice mode (0.2° oscillation, 0.05 s exposure, beamsize 82x28 μm, 40% transmission, wavelength 0.9795 Å, 1800 images for a 360° total sweep) using a *PsEFE* crystal (250 x 150 μm) grown in the presence of L-Arg substrate and NOG. Data were auto-processed using XIA2 (DIALS run) (4). Molecular replacement using PHASER (11) using the *I*222 crystal form of *PsEFE* as a search model located the four molecules present in the asymmetric unit. The four *PsEFE* chains were refined independently by manual cycles of model fitting in COOT (10) and refinement in PHENIX including anisotropic displacement parameters until  $R_{\text{work}}$  and  $R_{\text{free}}$  no longer converged (Table 1). Coordinates and structure factors are deposited in the Protein Data Bank under accession codes 5LSQ (*PsEFE*:Mn:BTP), and 5LUN (*PsEFE*:Fe:NOG:L-Arg).

## 4. Enzyme activity measurements

### a. GC/MS assays for ethylene

*PsEFE* activity was initially assayed using an optimized modified version of the reported GC method (12). Purified *PsEFE* (between 10 to 15  $\mu\text{M}$ ) was added to the 150  $\mu\text{L}$  of aqueous solution containing 50 mM HEPES buffer, pH 7.2, containing 5 mM ascorbate or DTT, 5 mM 2OG, 2 mM L-Arg, and 200  $\mu\text{M}$  of  $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$ . Reactions were carried out in 2.0 mL mass spectrometry vials with gas-tight rubber stoppers and incubated at 30°C while shaking (50 rpm; 30 min). Reactions were quenched by rapid freezing on dry ice, and then allowed to reach room temperature; 200  $\mu\text{L}$  of head space gas was then removed with a gas tight syringe for ethylene and carbon dioxide measurements. Ethylene was assayed for using a gas chromatograph (GC) machine (Model 7820A, Agilent Technologies) equipped with a flame ionization detector (FID) and Rt-Q-BOND column (0.32 mm x 30 mm, RESTEK). The amount of ethylene produced was calculated by comparing the integrated peak area of the enzyme reaction to that of standard 1000 ppm ethylene. The masses of ethylene and carbon dioxide were identified using either a Shimadzu GC-MS (GC-2010 plus) or an Agilent GC (GC 7820A) machine.

### b. Steady State Kinetics

Steady state kinetic assays were performed using the GC method as above with the following conditions; 125 nM *PsEFE*, 40 mM HEPES pH 7.5, 0.2 mM  $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$ , 0.4 mM sodium L-ascorbate. L-Arg concentrations varied from 0-640  $\mu\text{M}$  while keeping 2OG at 0.5 mM. 2OG concentrations varied from 0-640  $\mu\text{M}$  while keeping L-Arg at 0.5 mM. Reactions were quenched at specified time points and 100  $\mu\text{L}$  of head space gas was measured using the Agilent GC.

### c. NMR Spectroscopy

$^1\text{H}$ -NMR assays were performed using a Bruker Avance III 700 MHz spectrometer equipped with a TCI inverse cryoprobe. Some  $^1\text{H}$  spectra were acquired at 298 K using a Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence; water suppression was accomplished *via* presaturation. Experimental parameters: 16 scans, 40 ms total echo time, and 2.0 s relaxation delay. Spectra were processed using TopSpin 3.1 with a line broadening of 0.5 Hz. Reaction mixtures consisted of 1  $\mu\text{M}$  *PsEFE*,

400  $\mu\text{M}$  2OG, 500  $\mu\text{M}$  ascorbate, 500  $\mu\text{M}$  L-Arg and 20  $\mu\text{M}$   $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$  buffered with 50 mM Tris- $\text{D}_{11}$ , pH 7.5, in 10 %  $\text{D}_2\text{O}$ . Time courses consisted of 15 spectra acquired consecutively, with the first acquisition beginning 210 s after initiating reaction. Bruker Match<sup>TM</sup> tubes (3.0 mm x 100 mm; CortecNet) were used.

## 5. Computational Chemistry

### a. MD simulations

MD simulations of *PsEFe* in complex with 2OG used the AMBER program (version 12) and the force field parameters of Cornell et al. (13). To obtain potential parameters for 2OG, which are not available in the force field database, the procedure of Fox and Kollman was used, to be consistent with the standard AMBER force field (14). The *PsEFe:Fe:NOG:L-Arg* crystal structure was used as starting coordinates. To obtain the all-atom model including hydrogens for *PsEFe*, the protonation states of ionizable residues were assigned by inspecting their hydrogen bonding patterns; e.g. Asp and Glu side chains were assigned as neutral if either of their carboxylate oxygens was directed within 3.5 Å of hydrogen-bond accepting group (15). Similarly, Lys side chains were assumed as being protonated unless the  $\text{N}^\epsilon$  atom is in proximity to a hydrogen-bond donating group. The same procedure was applied for determining the protonation states of His residues.

After the addition of 3 chloride ions to neutralize the total charge, the all-atom model for the *PsEFe*-succinate complex was placed in a cube with edge length 70.9 Å using TIP3P water model (16). After 2000 minimization cycles to remove poor steric contacts, the system was equilibrated beginning with 20 ps equilibration dynamics of solvent molecules at 300 K. The next step involved equilibration of the solutes with a fixed configuration of the solvent molecules consecutively at 10, 50, 100, 150, 200, 250, and 300 K for 10 ps at each temperature. Equilibration dynamics of the entire system were then performed at 300 K for 100 ps. Following equilibration, 10.8 ns production dynamics simulations were carried out with periodic boundary conditions in the NPT ensemble. The temperature and pressure were kept at 300 K and 1 atm using Berendsen temperature coupling and isotropic molecule-based scaling, respectively (17).



## **b. Quantum mechanical calculations**

For quantum mechanical studies, structures corresponding to energy minima and transition states for the reactions of 3-chloropropanoate were optimized at the B3LYP/6-31+G\* level using Gaussian 09(18). The electronic energies computed in this way were used to calculate the relative free energies ( $\Delta G$ ) given by:  $\Delta G = \Delta E_{\text{elec}} + \Delta H' - T\Delta S$ , where  $\Delta H'$  denotes the enthalpy change due to thermal motions of the nuclei including the zero-point vibrational energies, and  $\Delta S$  is the entropy change. Electronic energies ( $E_{\text{elec}}$ ) were evaluated and vibrational frequencies were used to calculate  $\Delta H'$  and  $\Delta S$  at the B3LYP/6-31+G\* level of theory. The effects of solvation were described with an implicit description of acetic acid using the CPCM treatment (19); the United Atom Topological Model (UAHF) was used to define the solute cavity. All optimized species were verified as either minima or transition structures by the presence of zero or a single imaginary vibrational frequency. Gibbs free energies were evaluated at the 298 K using vibrational frequencies.

## Supplementary Table S1: Data collection and refinement statistics.

	<i>PsEFE</i> (SeMet)	<i>PsEFE:Mn:BTP</i>	<i>PsEFE:Mn:2OG</i>	<i>PsEFE:Fe:NOG:L-Arg</i>
PDB ID	†	5LSQ	5MOF	5LUN
<b>Data collection</b>				
Beamline	Diamond I04-1	Diamond I04-1	Diamond I03	Diamond I02
Space group	<i>I</i> 222	<i>I</i> 222	<i>I</i> 222	<i>P</i> 1
Wavelength (Å)	0.91741	0.91741	0.97960	0.97949
Unit Cell Dimensions				
a,b,c (Å)	79.710, 97.230, 98.310	79.590, 97.750, 98.550	79.483, 97.850, 98.094	49.839, 79.056, 97.859
$\alpha, \beta, \gamma$ (°)	90, 90, 90	90, 90, 90	90, 90, 90	91.563, 93.416, 100.796
Resolution (Å)	49.15 - 1.67 (1.71 - 1.67)	39.79 - 1.55 (1.59 - 1.55)	69.28 - 1.45 (1.54 - 1.45)	77.60 - 1.08 (1.12 - 1.08)
No. of molecules/ASU	1	1	1	4
No. of unique reflections	44579 (3246)	55915 (4090)	67866 (2697)	584165 (55024)
Completeness (%)	99.9 (100)	99.9 (99.9)	100 (100)	93.0 (89.3)
Multiplicity	13.4 (13.5)	13.2 (13.8)	6.3 (6.1)	3.5 (3.5)
$R_{\text{merge}}$	0.185 (2.252)	0.088 (0.769)	0.103 (1.438)	0.053 (0.822)
Mean $I/\sigma(I)$	11.5 (1.2)	19.2 (3.9)	10.8 (2.0)	7.7 (1.4)
CC half	0.998 (0.479)	ND	0.998 (0.592)	0.994 (0.701)
Wilson B value (Å <sup>2</sup> )	14.5	13.2	8.9	11.6
<b>Phasing</b>				
Figure of merit				
Centric	0.181			
Acentric	0.346			
Density Modification	0.74			
<b>Refinement</b>				
$R_{\text{work}}$		0.1346	0.1726	0.1620
$R_{\text{free}}$		0.1566	0.1894	0.1848
R.m.s.d.				
Bond length (Å)		0.013	0.004	0.011
Bond angle (°)		1.433	0.727	1.394
No. of atoms				
Protein		2730	2847	10974
Ligand/ion		33	31	204
Water		401	475	2128
$\langle B_{\text{factor}} \rangle$ (Å <sup>2</sup> ) all atoms		20.0	18.0	21.0

Numbers in parentheses refer to the highest resolution shell.

$R_{\text{merge}}$  is the unweighted R-value on  $I$  between symmetry mates.

$R_{\text{cryst}} = \sum hkl | |F_{\text{obs}}(hkl)| - k |F_{\text{calc}}(hkl)| | / \sum hkl |F_{\text{obs}}(hkl)|$  for the working set of reflections;  $R_{\text{free}}$  is the R-value for ~5% of the reflections excluded from refinement.

† Coordinates were not deposited for the selenomethionine data; however, the structure factors for the selenomethionine data are included in the 5LSQ deposit.

## Supplementary Table S2: RMSD (Å) matrix based on C $\alpha$ atoms for *PsEFE* chains.

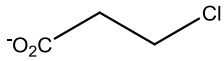
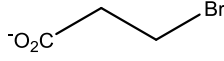
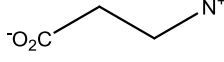
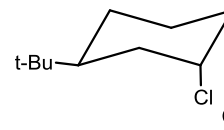
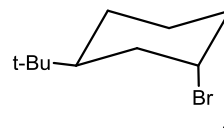
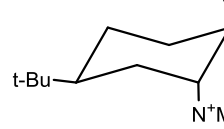
	<i>PsEFE:Mn:BTP</i>	<i>PsEFE:Mn:2OG</i>	<i>PsEFE:Fe:NOG:L-Arg chain A</i>	<i>PsEFE:Fe:NOG:L-Arg chain B</i>	<i>PsEFE:Fe:NOG:L-Arg chain C</i>	<i>PsEFE:Fe:NOG:L-Arg chain D</i>
<i>PsEFE:Mn:BTP</i>	0	0.127	0.633	0.471	0.578	0.534
<i>PsEFE:Mn:2OG</i>		0	0.612	0.459	0.594	0.516
<i>PsEFE:Fe:NOG:L-Arg chain A</i>			0	0.220	0.158	0.146
<i>PsEFE:Fe:NOG:L-Arg chain B</i>				0	0.177	0.185
<i>PsEFE:Fe:NOG:L-Arg chain C</i>					0	0.170
<i>PsEFE:Fe:NOG:L-Arg chain D</i>						0

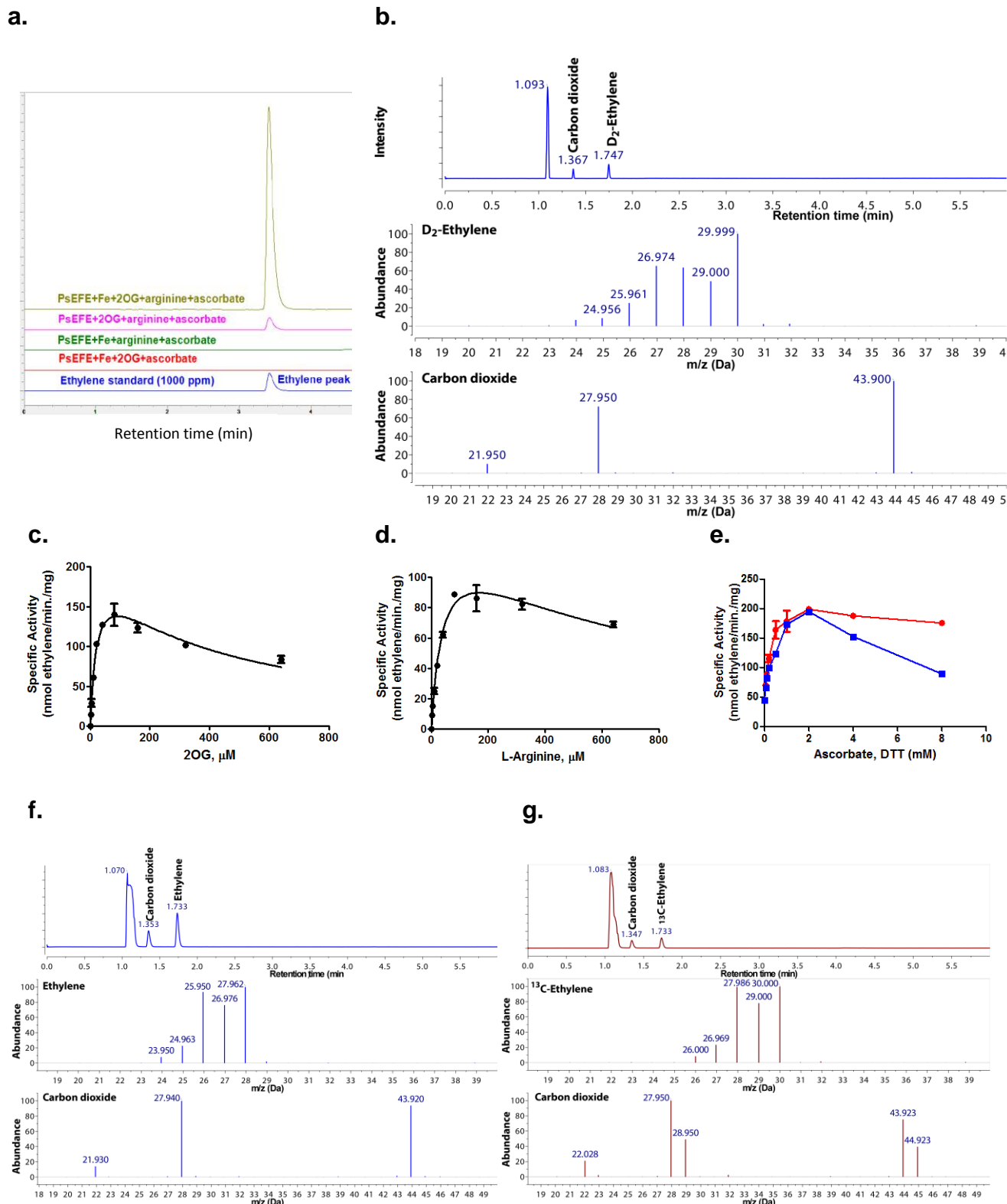
**Supplementary Table S3: Ethylene production by *PsEFE* variants.**

<i>PsEFE</i> variants	Relative ethylene forming activity (%)	Proposed role
Wild type	100±3.6	
E84Q	Not detected	L-Arg binding
E84D	Not detected	L-Arg binding
T86S	31.3 ±0.7	L-Arg binding
T86V	Not detected	L-Arg binding
R171K	Not detected	L-Arg binding
R171A	Not detected	L-Arg binding
Y192F	5.6 ±0.2	L-Arg binding
R316K	13.1 ±0.3	L-Arg binding
R316A	3.7 ±0.1	L-Arg binding
C317S	21.4 ±0.5	L-Arg binding
C317A	34.0 ±0.7	L-Arg binding
Y318F	65.6 ±1.2	L-Arg binding
F175Y	18.6 ±0.5	2OG binding
V270T	4.3 ±0.2	2OG binding

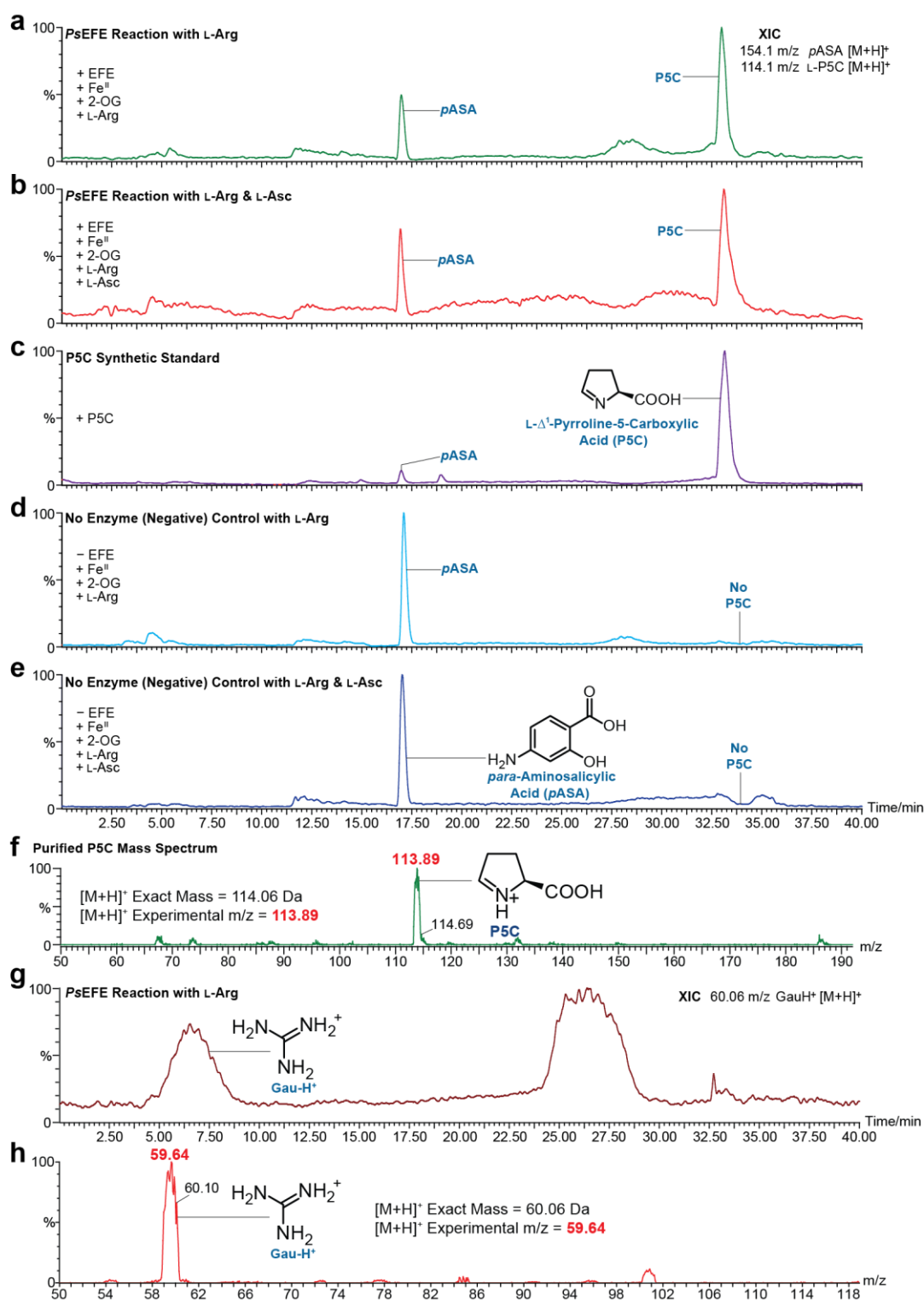
The production of ethylene by wild type *PsEFE* is normalized to 100%. Standard errors of the mean are calculated from three replicates.

**Supplementary Table S4:** Calculated Gibbs free energy for Grob fragmentation to give ethylene with torsion angles of 0°, 45°, and 90° (units kcal·mol<sup>-1</sup>). The torsion angle ( $\phi$ ) is defined as that between the plane of the carboxylate and the C-1-2-3 plane. The *tert*-butyl group in 4,5,6 'locks' the other substituents diaxial. Note that reaction is preferred for  $\phi = 90^\circ > 45^\circ > 0^\circ$ .

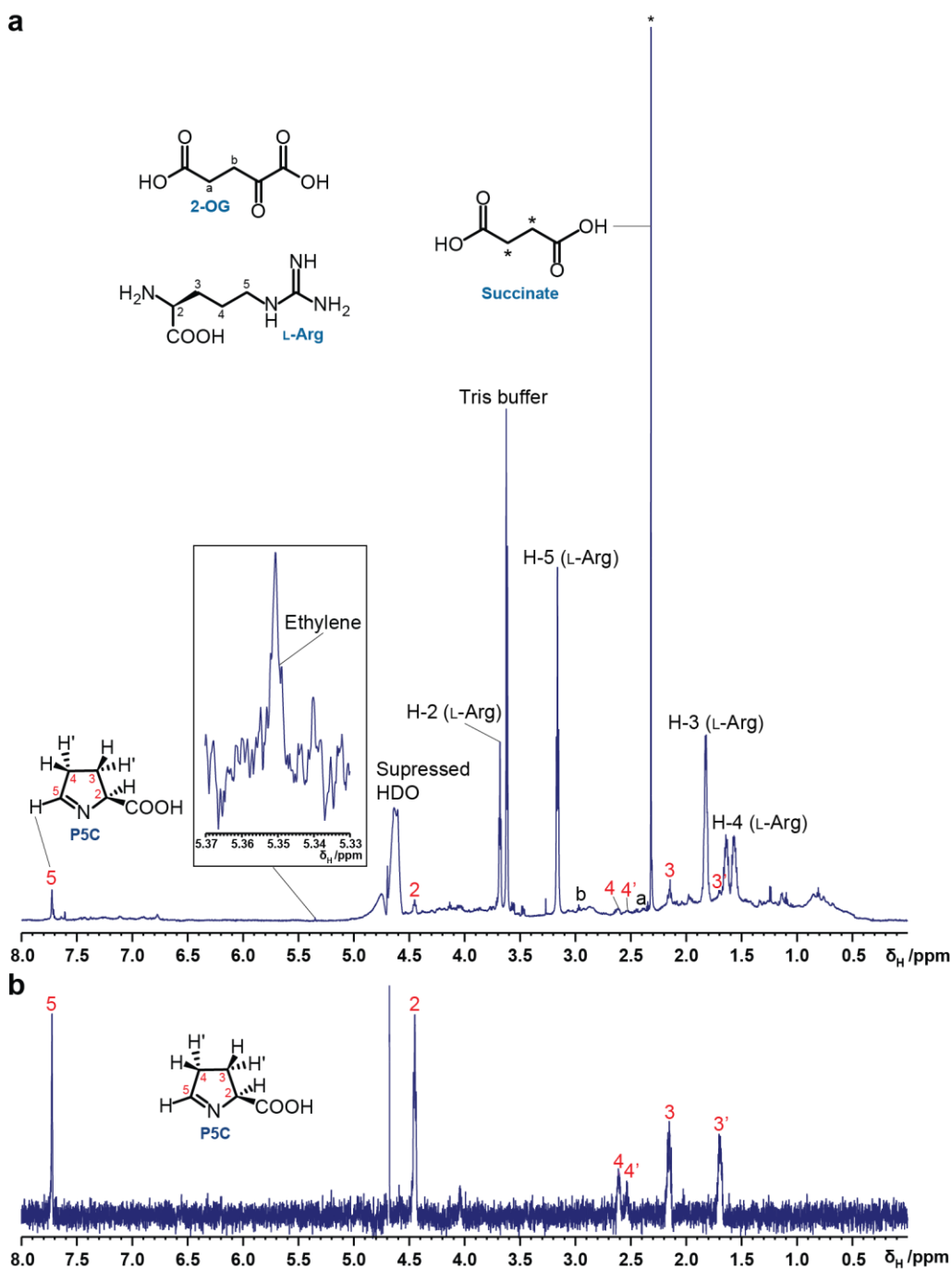
Reactant	$\phi_{\text{coco}}=0^\circ$	$\phi_{\text{coco}}=90^\circ$	$\phi_{\text{coco}}=45^\circ$
1 	15.82	12.27	14.05
2 	14.10	10.61	12.24
3 	8.01	5.34	6.87
4 	12.21	8.12	10.44
5 	11.19	7.23	9.48
6 	5.35	1.41	3.87



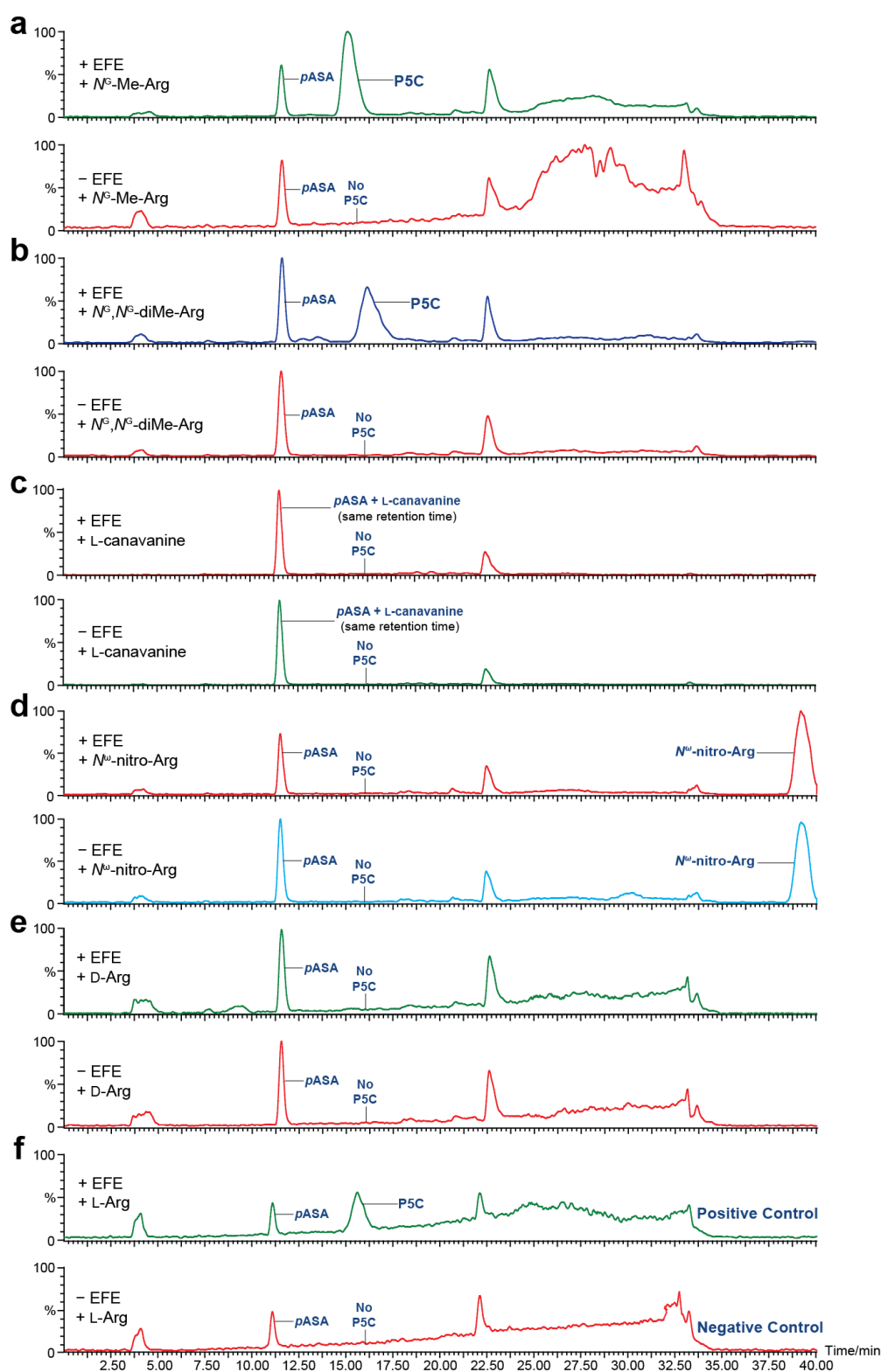
**Supplementary Figure S1: GC-MS analysis of *PsEFE* catalysis.** (a) GC chromatograms of *PsEFE* activity under the indicated conditions. (b) GC-MS analysis of ethylene production from [3,3- $^2\text{H}$ ]-labelled 2OG; the ethylene peak is at 1.75 min (top). MS spectra of  $^2\text{H}_2$ -ethylene (partially  $^2\text{H}$ -labelled) (middle) and  $\text{CO}_2$  peak is at 43.9 (bottom). The peak at 27.95 is  $\text{CO}$  likely deriving from  $\text{CO}_2$  fragmentation. (c) Dependence of *PsEFE* activity on variation in 2OG (d) and L-Arg concentrations with curve fitting using GraphPad Prism 7.0 for Michaelis-Menten kinetics with substrate inhibition.  $K_m$  values for 2OG and L-Arg:  $22.22 \pm 3.5 \mu\text{M}$  and  $41.31 \pm 5.4 \mu\text{M}$ , respectively;  $V_{\text{max}}$  values for 2OG and L-Arg:  $206.2 \pm 14.93 \mu\text{M}\cdot\text{min}^{-1}$  and  $134.0 \pm 8.61 \mu\text{M}\cdot\text{min}^{-1}$ , respectively. (e) Ascorbate (red) or DTT (blue) enhance *PsEFE* ethylene forming activity. (f) and (g) incubation with unlabelled 2OG or 1,2,3,4- $^{13}\text{C}$ -labelled 2OG, respectively. Note, the production of labelled ethylene and both labelled and unlabelled  $\text{CO}_2$  in (g), though some of the unlabelled  $\text{CO}_2$  is likely atmosphere derived.



**Supplementary Figure 2: LC/MS detection of pyrroline-5-carboxylate (P5C).** PsEFE reactions (100  $\mu$ L) were at 25  $^{\circ}$ C for 14 h in 50 mM Tris-HCl, pH 7.5. Reactions were quenched with 100  $\mu$ L methanol (1:1) containing *para*-aminosalicylic acid (pASA) standard, then microcentrifuged. Analysis employed a Waters Quattro Micro ESI-MS system (positive ion mode) and HPLC using a SiELC Primesep 100 HPLC column (100  $\text{\AA}$  pore size, 5  $\mu$ m particle size; 4.6 mm  $\times$  250 mm). Extracted-ion chromatograms (XICs) (analysing for presence of P5C, pASA [standard], or guanidinium ions) are shown for samples containing: **(a)** 15  $\mu$ M PsEFE, 100  $\mu$ M  $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$ , 1 mM 2OG, 1 mM L-Arg; **(b)** 15  $\mu$ M PsEFE, 100  $\mu$ M  $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$ , 1 mM 2OG, 1 mM L-Arg, 1 mM sodium L-ascorbate (L-Asc); **(c)** Synthetic P5C standard; **(d)** 100  $\mu$ M  $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$ , 1 mM 2OG, 1 mM L-Arg; **(e)** 100  $\mu$ M  $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$ , 1 mM 2OG, 1 mM L-Arg, 1 mM sodium L-ascorbate. **(f)** The PsEFE reaction (with L-Arg and L-Asc) was scaled up (2 $\times$ 1 mL) and P5C was purified (retention time of  $\sim$ 12 min) by an LC/MS method to give (f). **(g)** Formation of protonated guanidinium ions (Gau-H<sup>+</sup>) was observed (XIC: 60.06 m/z) in cases where P5C was produced; conditions were as in sample (a). **(h)** Mass spectrum for the guanidinium ions (Gau-H<sup>+</sup>) produced in sample (a) [shown eluting from the HPLC column after  $\sim$ 7 min in (g)].

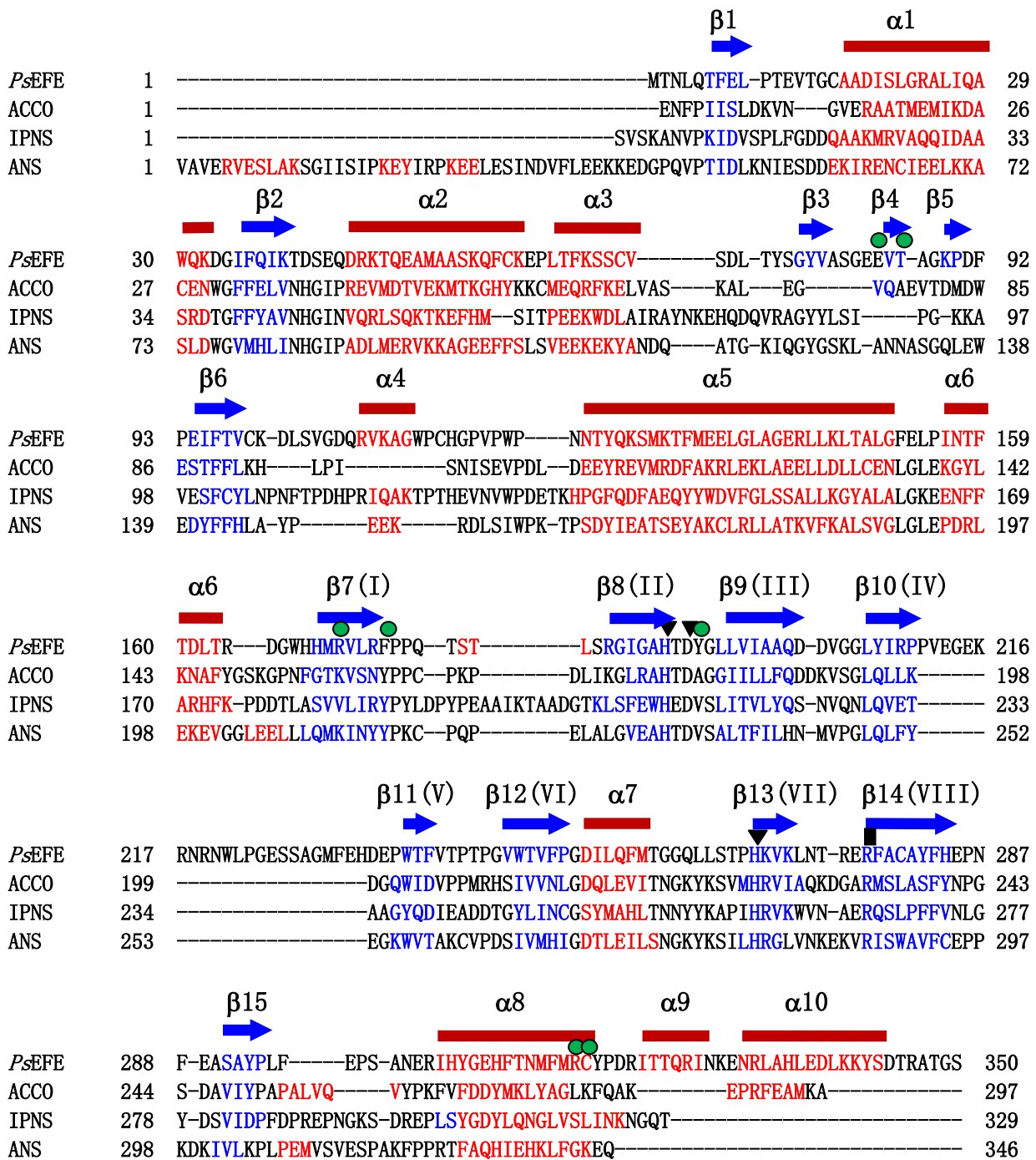


**Supplementary Figure S3: NMR detection of *PsEFE*-catalyzed formation of pyrroline-5-carboxylic acid (P5C).** Reactions (165  $\mu$ L) were carried out in Eppendorf tubes, and then transferred to 3 mm NMR tubes. Reactions were incubated overnight at room temperature in 50 mM Tris- $d_4$ -DCl, pH 7.5 [conditions: 15  $\mu$ M *PsEFE*, 100  $\mu$ M  $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$ , 1 mM 2OG, 1 mM L-Arg]. Spectra were measured at 298 K using a 700 MHz ( $^1\text{H}$ ) Bruker AVIII 700 NMR spectrometer [equipped with a  $^1\text{H}/^{13}\text{C}/^{15}\text{N}$  TCI  $\text{N}_2$  CryoProbe];  $^1\text{H}$ -NMR spectra employed a pulse sequence with presaturation (to suppress [HDO] solvent peaks) and baseline optimization ('zgpr' pulse sequence – with O1P = 4.701 ppm). **(a)**  $^1\text{H}$ -Spectrum of the crude *PsEFE* reaction mixture. **(b)** The 1D TOCSY spectrum of the crude *PsEFE* reaction mixture (with irradiation at 4.45 ppm) is consistent with the reported assignment for P5C. Note that the ethylene peak is small relative to P5C in this assay due to overnight incubation in a gas permeable Eppendorf tube, leading to ethylene loss.

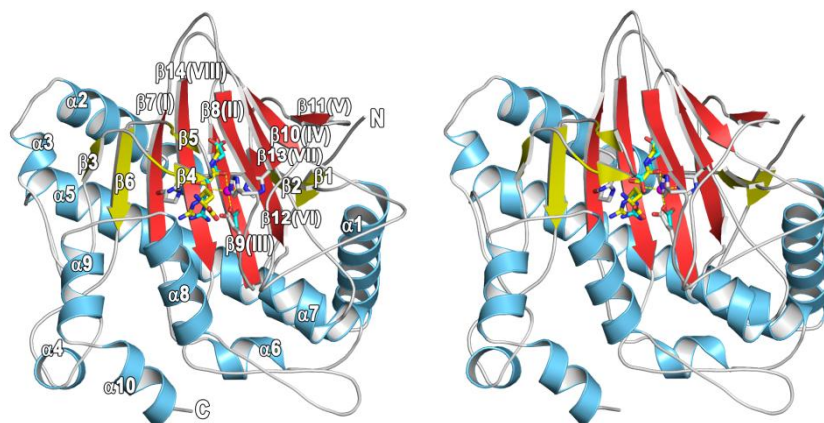
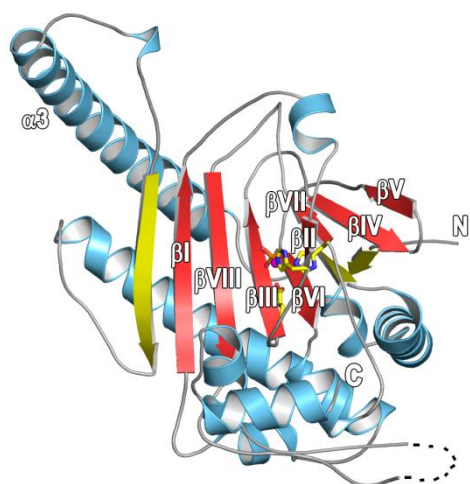
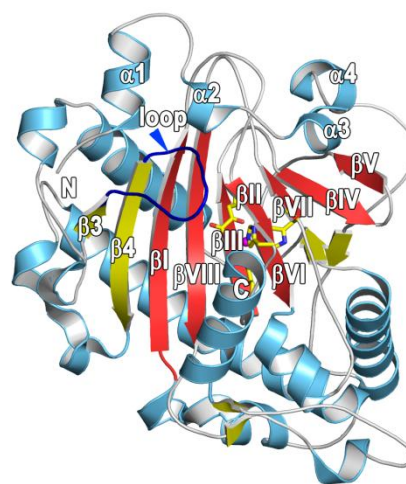


**Supplementary Figure S4: Arginine analogues tested with PsEFE.** Reactions (100  $\mu$ L) were performed at room temperature overnight in 10 mM HEPES-NaOH, pH 7.5; conditions were 15  $\mu$ M PsEFE, 100  $\mu$ M  $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$ , 1 mM 2OG, 1 mM substrate. Samples were analysed by LC/MS by a similar method to that of Supplementary Figure 3. Extracted-ion chromatograms (XICs) (showing 114.1  $m/z$  [P5C] and 154.1  $m/z$  [pASA] ions) are shown for reactions and 'no enzyme' controls. Substrate analogues tested: **(a)**  $N^6$ -methyl-L-arginine ( $N^6$ -Me-Arg); **(b)** (asymmetric)  $N^6,N^6$ -dimethyl-L-arginine ( $N^6,N^6$ -diMe-Arg); **(c)** L-canavanine ( $\text{SO}_4$ ); **(d)**  $N^6$ -nitro-L-arginine ( $N^6$ -nitro-Arg); **(e)** D-arginine (D-Arg); and **(f)** L-Arg. Preliminary assays for ethylene production similarly indicated high selectivity for L-Arg; no ethylene formation was observed with D-Arg,  $N^6$ -nitro-L-arginine, 2-keto-4-methyl-thiobutyric acid, guanidine with L-alanine, L-( $\alpha$ -amino)adipic acid, diaminopimelic acid,  $N^6$ -acetyl-L-arginine, L-histidine, L-citrulline, or L-lysine. When compared with L-Arg (100%  $\pm$  13% activity), reduced levels of activity were observed with  $N^6$ -methyl-L-arginine (52%  $\pm$  7%), (asymmetric)  $N^6,N^6$ -dimethyl-L-arginine (42%  $\pm$  10%),  $N^6$ -hydroxy-L-arginine (17%  $\pm$  9%), L-canavanine (5-10%), and L-homoarginine (3.3%  $\pm$  0.7%). All experiments were carried out in triplicate, with the same results.



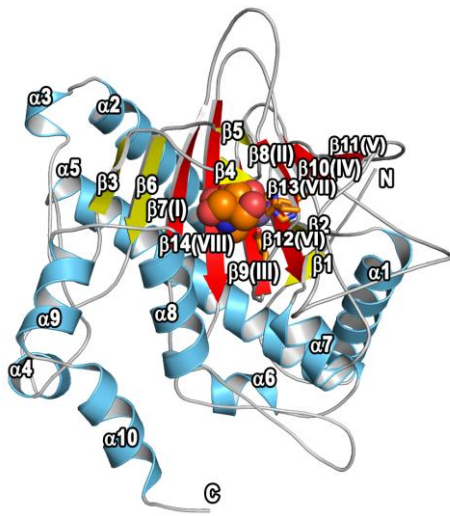


Supplementary Figure S5: Secondary structure based sequence alignment of *PsEFE* with ACC oxidase (ACCO) from *Petunia hybrida* (PDB 1WA6), isopenicillin *N* synthase (IPNS) from *Aspergillus nidulans* (PDB 1IPS), and anthocyanidin synthase (ANS) from *Arabidopsis thaliana* (PDB 1GP5). Residues in blue correspond to  $\beta$ -strands and in red to  $\alpha$ -helices. *PsEFE* secondary structure elements (blue arrows for  $\beta$ -strands and red rectangles for  $\alpha$ -helices) above the sequences are based on the *PsEFE* structure. Residues ligating to Fe are indicated by black triangles; green circles show residues involved in hydrogen bond or electrostatic interactions with L-Arg substrate. Arg277, which interacts with the NOG or 2OG C-5 carboxylate, is indicated with a black rectangle. The major DSBH strands are labelled with roman numerals (I-VIII);  $\alpha$ -helices and  $\beta$ -strands are sequentially numbered.

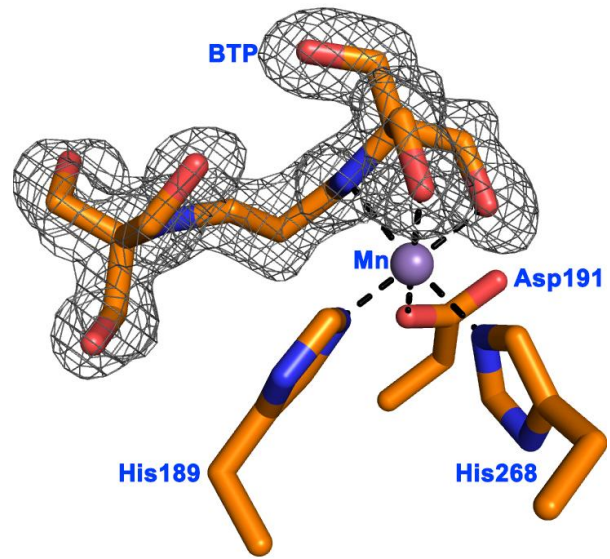
**a****b****c**

**Supplementary Figure S6: The overall fold of (a) PsEFE (stereoview) compared to (b) 1-aminocyclopropane-1-carboxylate oxidase (ACCO) and (c) anthocyanidin synthase (ANS). (a)** Stereoview of the PsEFE fold. The eight  $\beta$ -strands of the double stranded  $\alpha$ -helix (DSBH) core are in red and indicated by numerals (I to VIII). Non-DSBH  $\beta$ -strands are yellow, helices marine blue and loops gray. The active site iron ligating residues and substrates are shown as sticks. **(b)** View of the structure of ACCO from *Petunia hybrida* (PDB 1WA6)(20). Note the ACCO structure has an unusually long  $\alpha$ -3 helix and an apparently open active site. **(c)** View from a crystal structure of ANS from *Arabidopsis thaliana* (PDB 1GP5)(21). The four *N*-terminal  $\alpha$ -helices ( $\alpha$ 1-4) and the loops between them wrap around the DSBH core and help to enclose the active site. The loop ('loop', dark blue) located between  $\beta$ 3 and  $\beta$ 4 also folds over the active site, but is not directly involved in substrate binding.

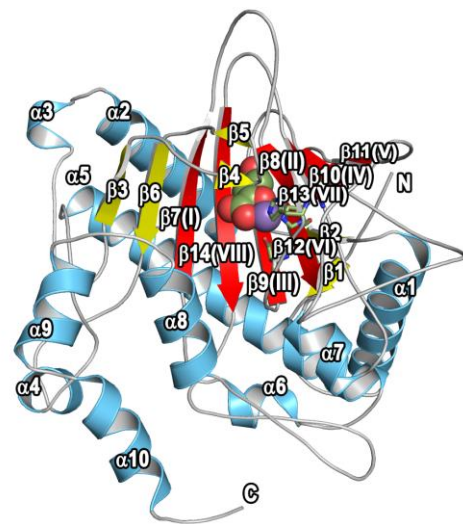
a



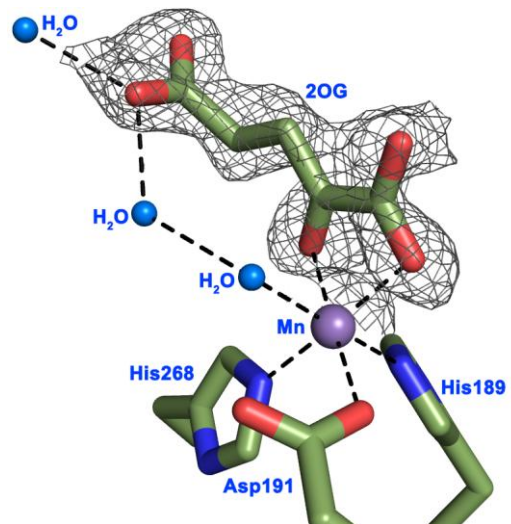
b



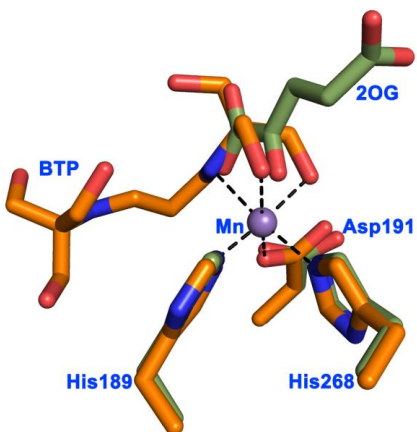
c



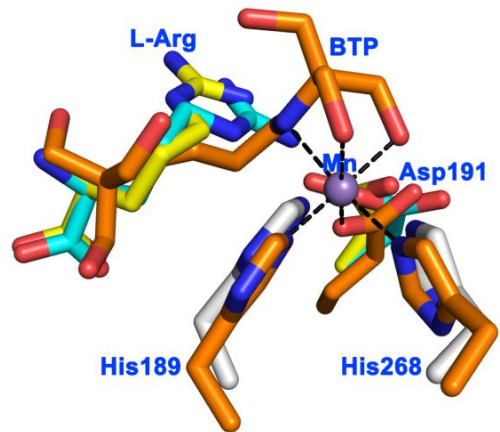
d



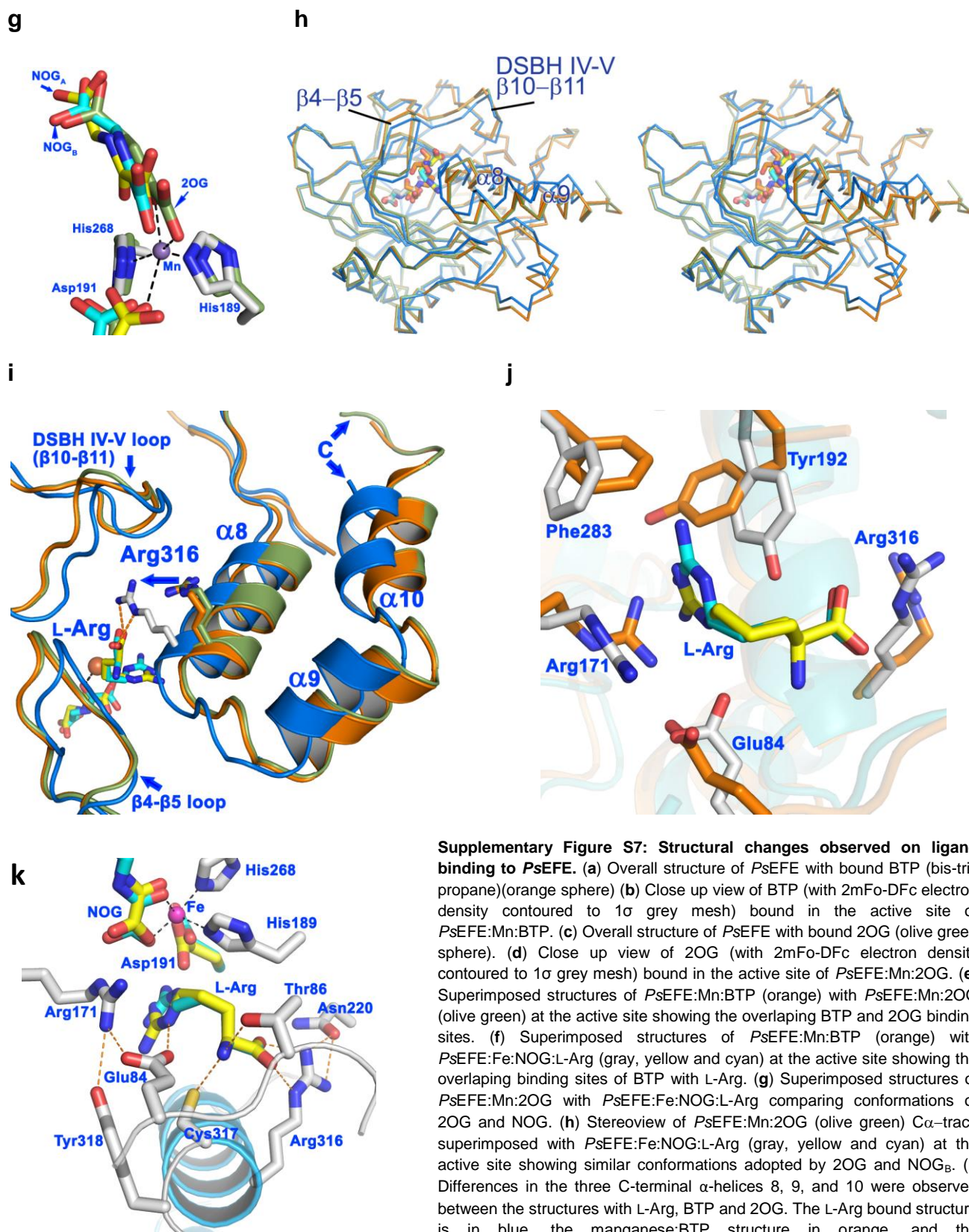
e



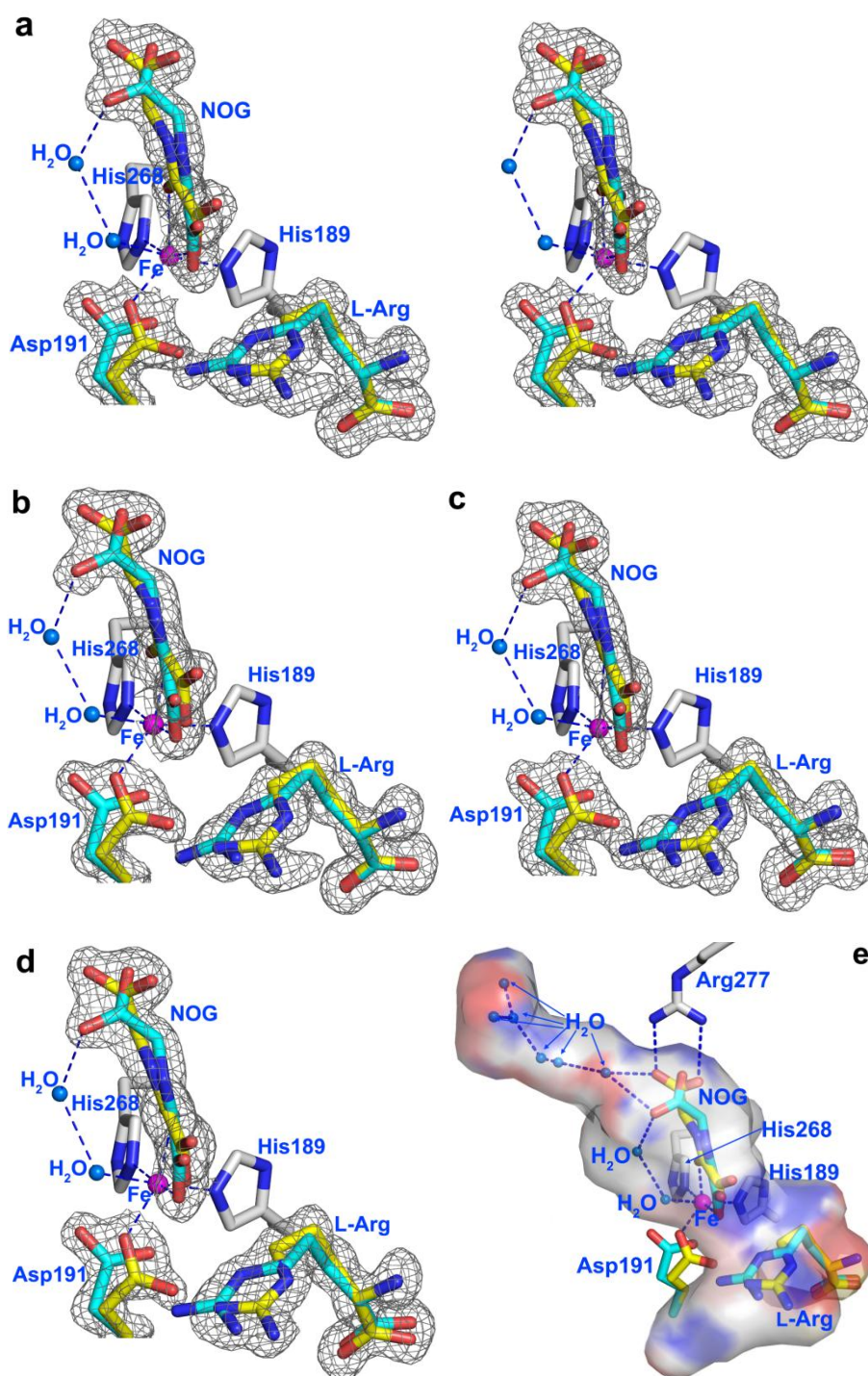
f



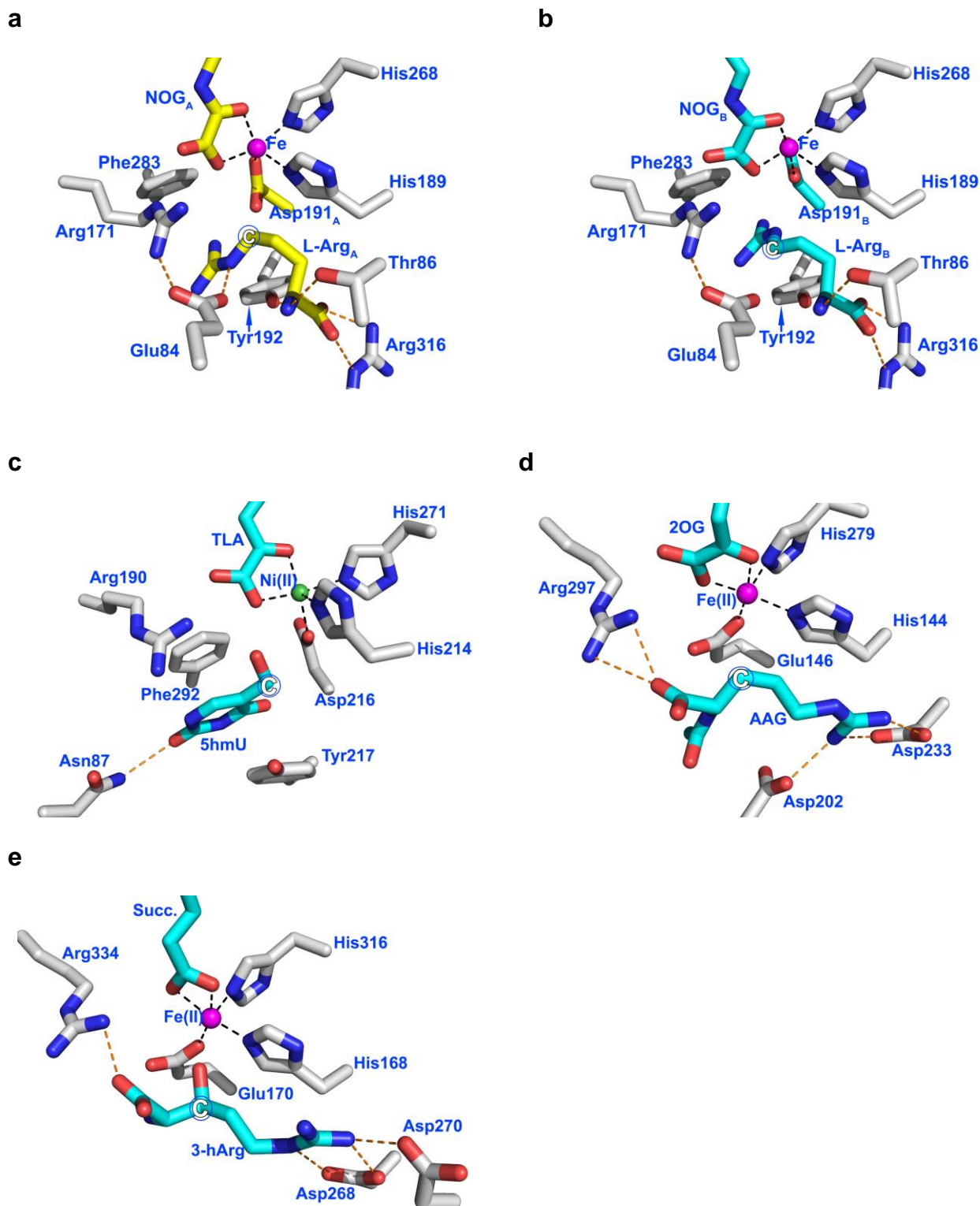




**Supplementary Figure S7: Structural changes observed on ligand binding to PsEFE.** (a) Overall structure of PsEFE with bound BTP (bis-tris propane)(orange sphere) (b) Close up view of BTP (with 2mFo-DFc electron density contoured to 1 $\sigma$  grey mesh) bound in the active site of PsEFE:Mn:BTP. (c) Overall structure of PsEFE with bound 2OG (olive green sphere). (d) Close up view of 2OG (with 2mFo-DFc electron density contoured to 1 $\sigma$  grey mesh) bound in the active site of PsEFE:Mn:2OG. (e) Superimposed structures of PsEFE:Mn:BTP (orange) with PsEFE:Mn:2OG (olive green) at the active site showing the overlapping BTP and 2OG binding sites. (f) Superimposed structures of PsEFE:Mn:BTP (orange) with PsEFE:Fe:NOG:L-Arg (gray, yellow and cyan) at the active site showing the overlapping binding sites of BTP with L-Arg. (g) Superimposed structures of PsEFE:Mn:2OG with PsEFE:Fe:NOG:L-Arg comparing conformations of 2OG and NOG. (h) Stereoview of PsEFE:Mn:2OG (olive green) C $\alpha$ -trace superimposed with PsEFE:Fe:NOG:L-Arg (gray, yellow and cyan) at the active site showing similar conformations adopted by 2OG and NOG<sub>B</sub>. (i) Differences in the three C-terminal  $\alpha$ -helices 8, 9, and 10 were observed between the structures with L-Arg, BTP and 2OG. The L-Arg bound structure is in blue, the manganese:BTP structure in orange, and the manganese:2OG structure in olive green. (j) Conformational changes upon L-Arg binding. Gray sticks are with L-Arg bound and orange sticks with BTP bound. (k) Residues involved in the binding of L-Arg.

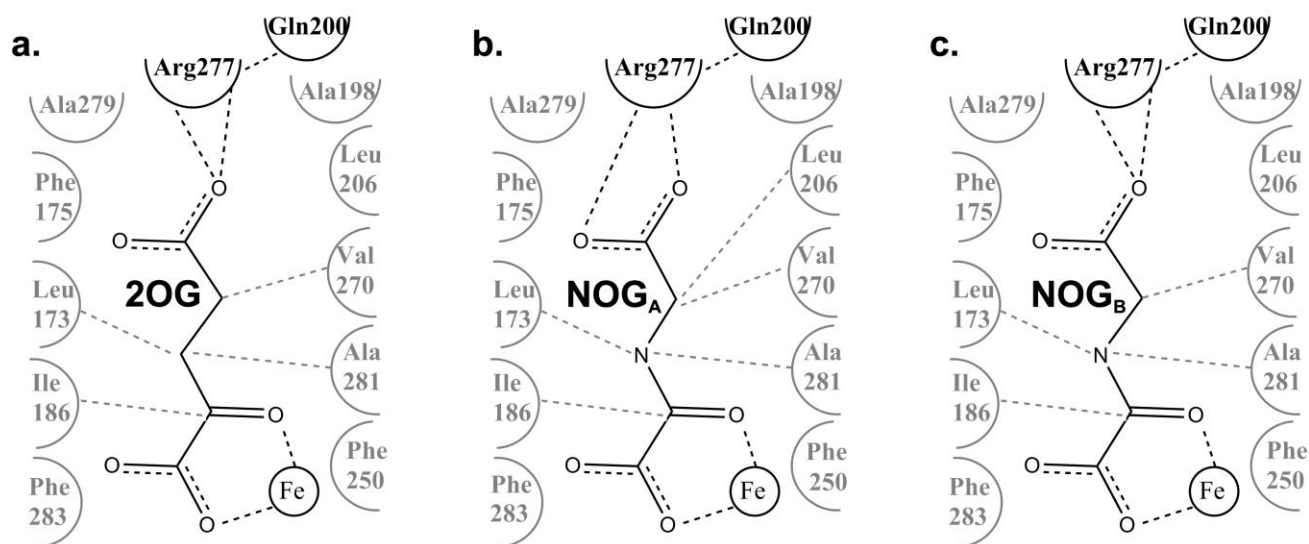


**Supplementary Figure S8: Views of the  $2mF_o-DF_c$  of PsEFE.Fe.NOG.L-Arg complex structure electron density maps for NOG and L-arginine of the four protein chains in the asymmetric unit.** The views reveal conservation of the two observed binding modes (Conformations A (yellow) and B (cyan) for chains A-D, see main text and Fig. 2b-e) for L-Arg and NOG in each of the four PsEFE chains in the asymmetric unit. (a) Stereoview for chain A; (b) (c) and (d) views for chains B, C and D (with  $2mF_o-DF_c$  electron density contoured to  $1\sigma$  grey mesh). (e) PsEFE surface representation of the active site cavity leading to an unprecedented water filled channel traversing the DSBH (red represents oxygen, blue nitrogen and white hydrophobic regions of the protein) (see main text).

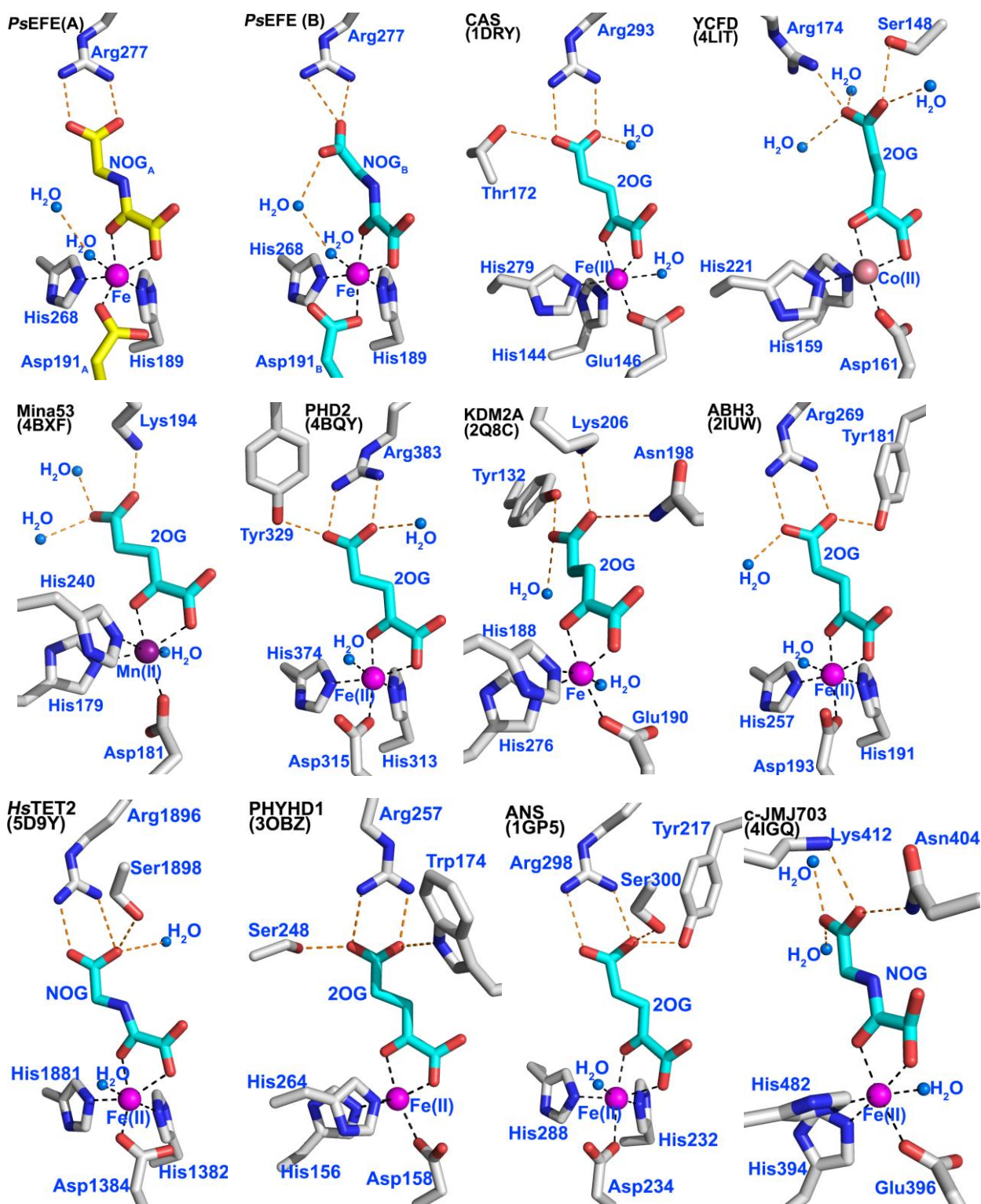


**Supplementary Figure S9: Comparison of the substrate binding mode of *PsEFE* with those of other 2OG oxygenases. (a, b)** Conformations A and B observed in the *PsEFE*.Fe.NOG.L-Arg complex. **(c)** T7H (human TET homologue) from *Neurospora crassa* in complex with Ni(II) and 5-hydroxymethyl uracil (5hmU) (PDB ID 5C3R); **(d)** CAS from *Streptomyces clavuligerus* in complex with Fe(II), 2OG and *N*-acetylarginine (AAG) (PDB ID 1DRY); **(e)** L-Arginine oxygenase VioC from *Streptomyces vinaceus* in complex with (2*S*,3*S*)-3-hydroxyarginine (3-hArg)(PDB ID 2WBO):The hydroxylated carbon is labelled with a circled white "C".



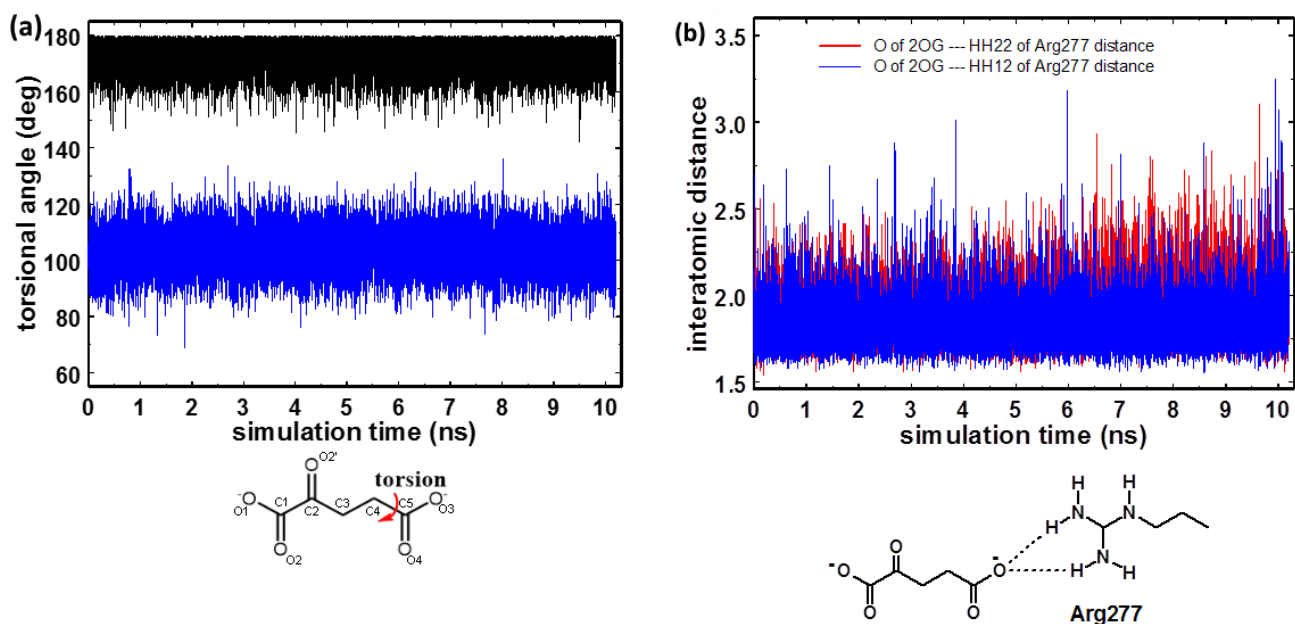


**Supplementary Figure S10: Interactions in the hydrophobic binding pocket for 2OG in *PsEFE*.** Residues within 6 Å of 2OG/NOG are depicted as semi-circles. (a) 2OG, (b) NOG<sub>A</sub> and (c) NOG<sub>B</sub>. Predicted hydrogen bonds and metal chelation are depicted in bold dashed lines and van der Waals contacts ( $\leq 4.0$  Å) are shown as gray dashed lines. Figure was made using ChemDraw Professional 15.1.

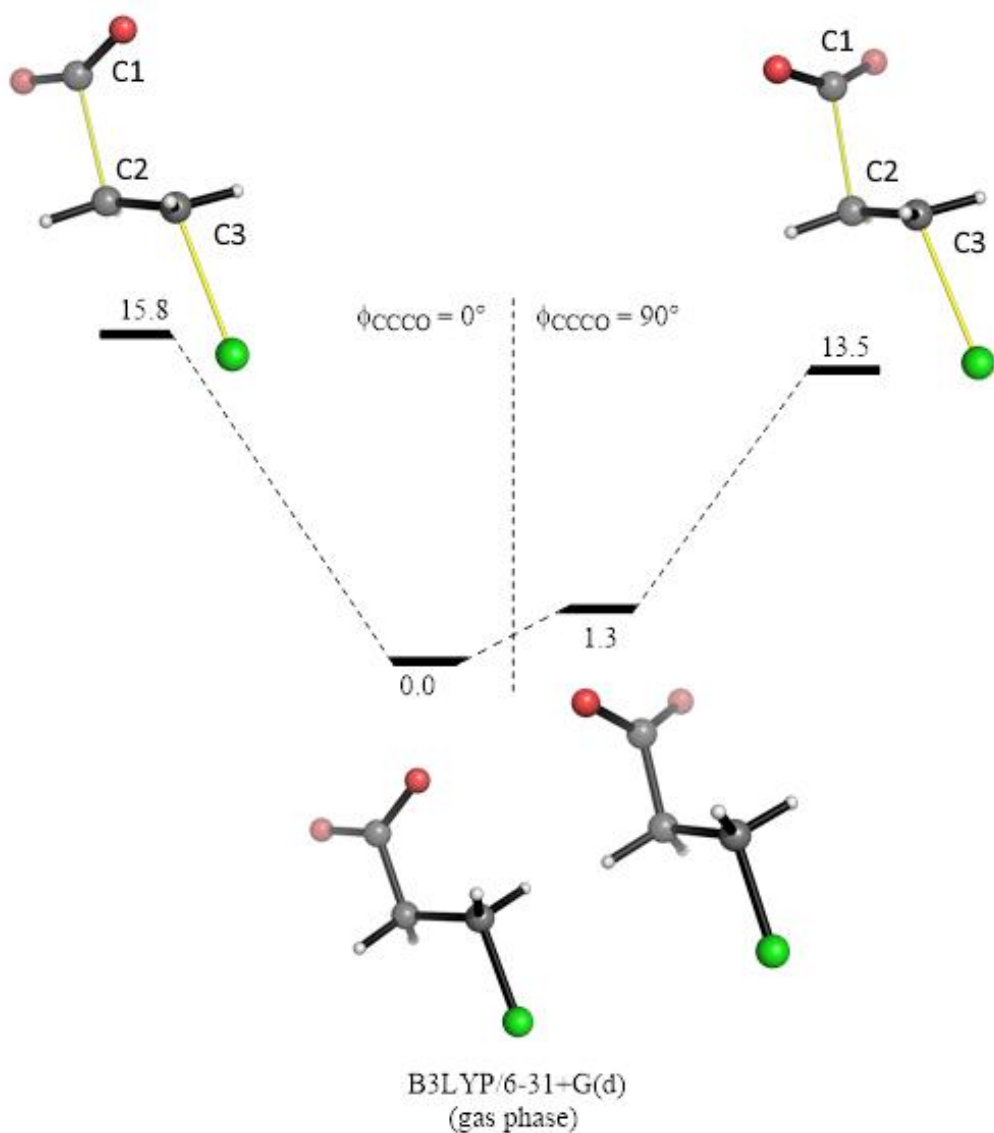


**Supplementary Figure S11: Views of 2-oxoglutarate (2OG) binding modes from selected 2OG oxygenase structures.** *PsEFE*(A) and *PsEFE*(B): Conformations A and B of *N*-oxalylglycine (NOG), L-Arg and Asp191; CAS: clavaminic acid synthase from *Streptomyces clavuligerus* (PDB ID 1DRY); YCFD: 50S ribosomal protein L16 arginine hydroxylase from *E. coli* (PDB ID 4LIT); Mina53: bifunctional lysine-specific demethylase and histidyl-hydroxylase Mina53 from *Homo sapiens* (PDB ID 4BXF); PHD2: HIF prolyl hydroxylase domain 2 from *Homo sapiens* (PDB ID 4BQY); KDM2A(FBXL11): JmjC domain-containing histone demethylase from *Homo sapiens* (PDB ID 2Q8C); ABH3: alkylated repair protein alkB homolog 3 from *Homo sapiens* (PDB ID 2IUW); TET2: methylcytosine dioxygenase TET2 (Ten-eleven translocation) from *Homo sapiens* (PDB ID 5D9Y); PHYHD1: a phytanol-CoA hydroxylase related dioxygenase from *Homo sapiens* (PDB ID 3OBZ); ANS: anthocyanidin synthase from *Arabidopsis thaliana* (PDB ID 1GP5); c-JMJ703: H3K4 demethylase from *Oryza sativa subsp. Japonica* (PDB ID 4IGQ). Note, that in all of the structures, except for *PsEFE*, the torsion angle formed between the plane of the C-5 carboxylate and the plane formed by C-3/C-4/C-5 of 2OG/NOG is similarly low.

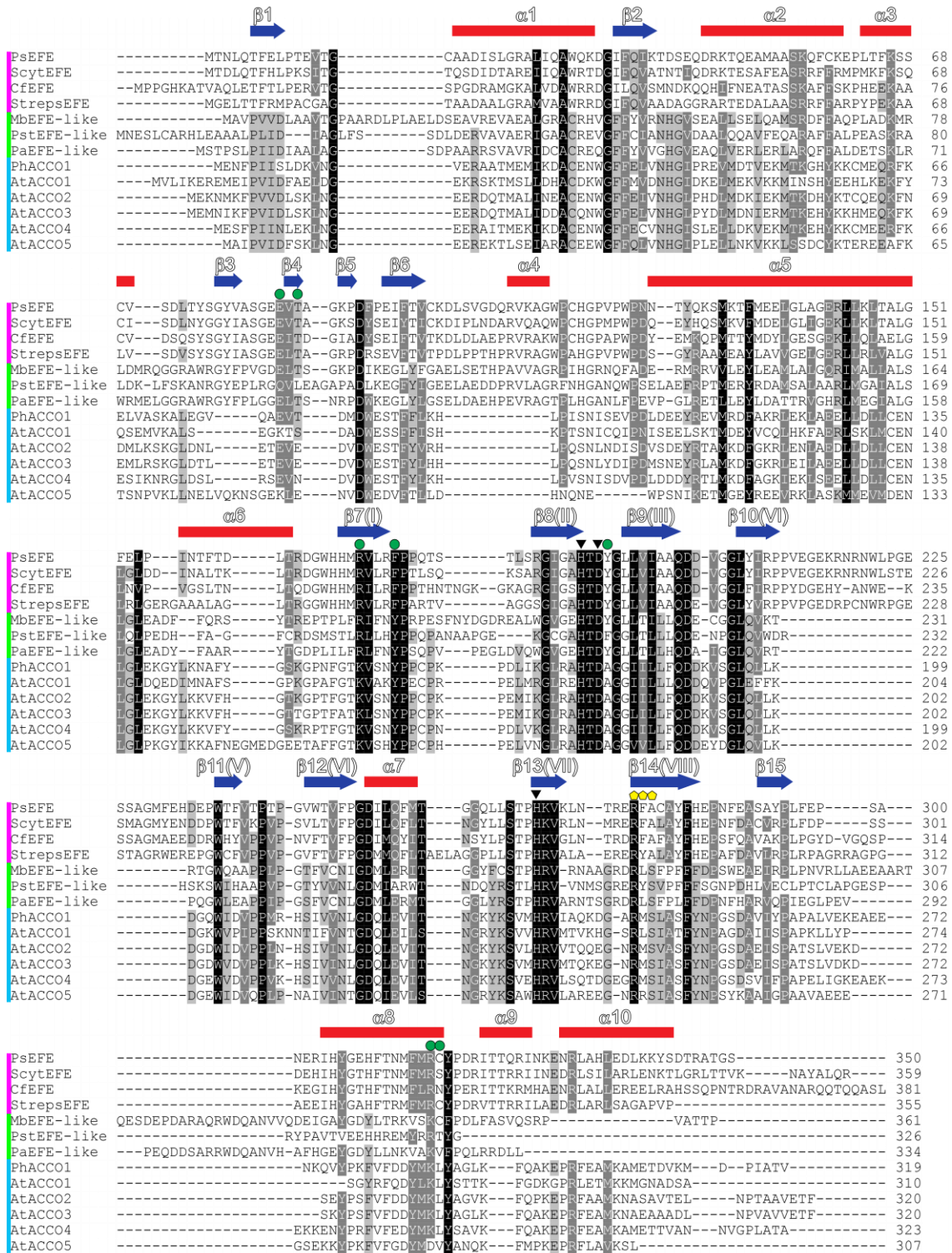




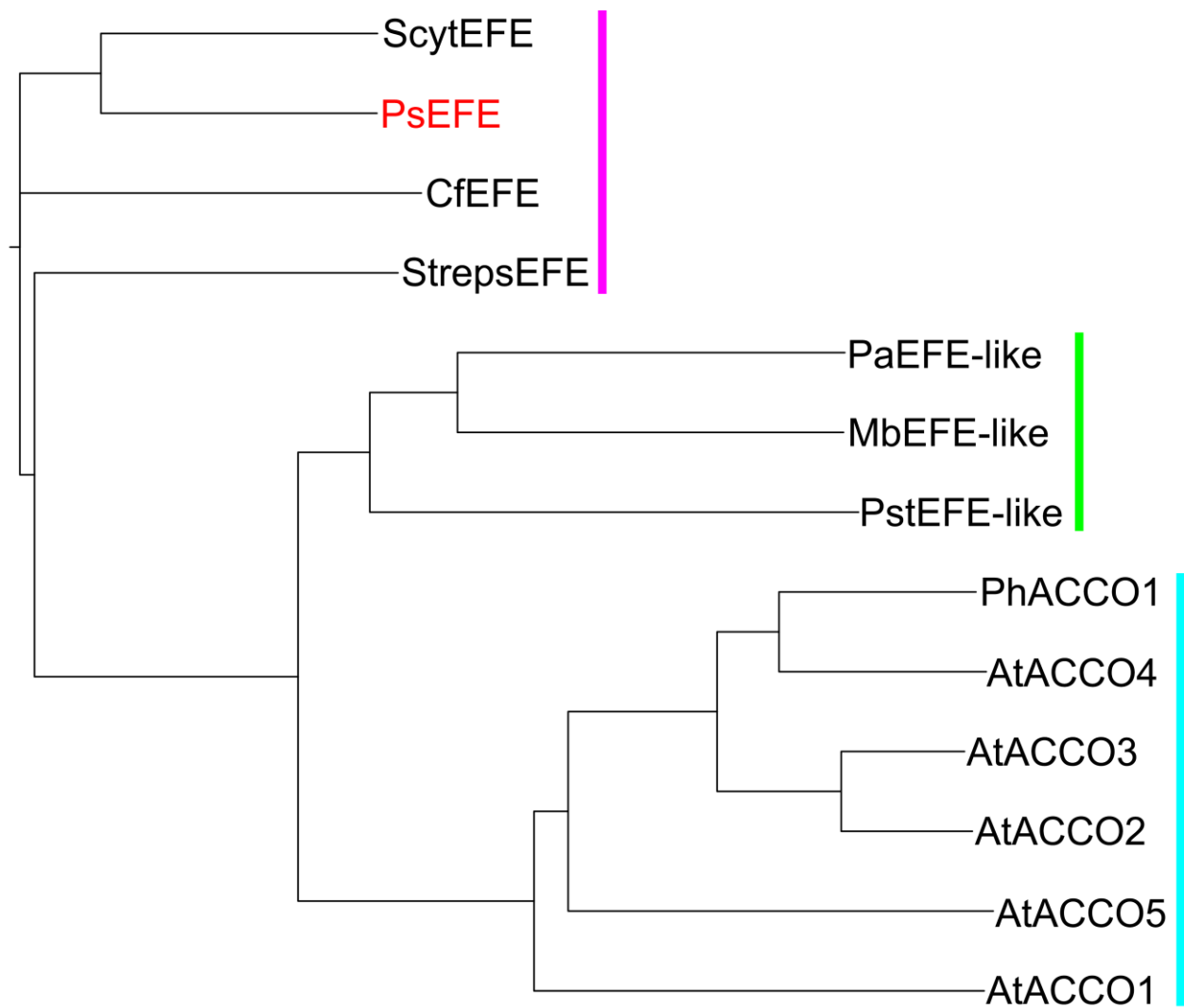
**Supplementary Figure S12: MD simulation studies on the conformation of 2OG in complex with PsEFE.** A PsEFE:Fe:2OG:L-Arg model was calculated based on the high resolution PsEFE:Fe:NOG:L-Arg structure (a) Time-course evolutions of the torsion angle between the O3-C5-O4 plane and the C3-C4-C5 plane of 2OG in water (black) and when complexed with PsEFE (blue); (b) the two hydrogen bonds between 2OG and Arg277 of PsEFE (see Supplementary Fig. S11) were maintained for 94% of the 10 ns simulation time.



**Supplementary Figure S13: Analysis of the preferred torsion angle for Grob type fragmentation to give ethylene, CO<sub>2</sub>, and a chloride ion from 3-chloropropanoate.** The figure shows the free energy profile of the intrinsic reaction coordinates for concerted fragmentation of 3-chloropropanoate to give ethylene. The torsion angle ( $\phi$ ) is defined as that between the plane of the carboxylate and the C-2 C-3 bond. Calculations were carried out for  $\phi_{\text{CCCO}}=0^\circ$ ,  $\phi_{\text{CCCO}}=45^\circ$ , and  $\phi_{\text{CCCO}}=90^\circ$  (Supplementary Table 3). The results imply  $\phi_{\text{CCCO}}=90^\circ$  is preferred for fragmentation to ethylene.



**Supplementary Figure S14: Clustal Omega sequence alignment of assigned and predicted ethylene-forming enzymes.** Sequences sharing 65-50% Seq ID to PsEFE highlighted to left in pink and 22-20% in green. Secondary structure elements for PsEFE shown above, helices red rectangles, strands blue arrows. L-Arg interacting residues indicated by green circle above, iron binding residues with black triangle, and "RXS" motif with yellow pentagons. ACCOs from *Petunia hybrida* and *Arabidopsis thaliana* highlighted to left in light blue. Identity to PsEFE in [square brackets]. PsEFE [100%], *Pseudomonas syringae* pv. phaseolicola (UniProt Q549K5); ScytEFE [65%], *Scytonema* sp. HK-05 (Refseq WP\_073633445.1); CfEFE [50%], *Colletotrichum fioriniae* PJ7 (Refseq XP\_007602399.1); StrepsEFE [54%], *Streptomyces* sp. NRRL S-87 (Refseq WP\_030204587.1); MbEFE-like [22%], *Monosiga brevicollis* MX1 (Refseq XP\_001748043.1); PstEFE-like [20%], *Pseudomonas stutzeri* (Refseq WP\_053527348.1); PaEFE-like [21%], *Pseudomonas aeruginosa* (Refseq WP\_034076857.1); PhACCO1 [17%], *Petunia hybrida* (UniProt Q08506); AtACCO1-5 [16%, 17%, 17%, 16%, 16%], *Arabidopsis thaliana* (UniProt Q9ZUN4, Q41931, Q06378, Q06588, Q0WPW4).



**Supplementary Figure S15: Clustal Omega sequence phylogeny of assigned and predicted ethylene-forming enzymes.** PsEFE, *Pseudomonas syringae* pv. phaseolicola (UniProt Q549K5); ScytEFE, *Scytonema* sp. HK-05 (Refseq WP\_073633445.1); CfEFE, *Colletotrichum fioriniae* PJ7 (Refseq XP\_007602399.1); StrepsEFE, *Streptomyces* sp. NRRL S-87 (Refseq WP\_030204587.1); MbEFE-like, *Monosiga brevicollis* MX1 (Refseq XP\_001748043.1); PstEFE-like, *Pseudomonas stutzeri* (Refseq WP\_053527348.1); PaEFE-like, *Pseudomonas aeruginosa* (Refseq WP\_034076857.1); PhACCO1, *Petunia hybrida* (UniProt Q08506); AtACCO1-5, *Arabidopsis thaliana* (UniProt Q9ZUN4, Q41931, O65378, Q06588, Q0WPW4). Biochemical ethylene forming activity has been shown for PsEFE (this work) and for various ACCOs.

## Supplementary References

1. LeMaster DM & Richards FM (1985) H-1-N-15 Heteronuclear NMR-studies of *Escherichia coli* thioredoxin in samples isotopically labeled by residue type. *Biochemistry* 24(25):7263-7268.
2. Delageniere S, *et al.* (2011) ISpyB: An information management system for synchrotron macromolecular crystallography. *Bioinformatics* 27(22):3186-3192.
3. Bailey S (1994) The CCP4 suite - programs for protein crystallography. *Acta Crystallogr D* 50:760-763.
4. Winter G (2010) Xia2: An expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* 43:186-190.
5. Zhang Z, Sauter NK, van den Bedem H, Snell G, & Deacon AM (2006) Automated diffraction image analysis and spot searching for high-throughput crystal screening. *J Appl Crystallogr* 39:112-119.
6. Terwilliger TC, *et al.* (2009) Decision-making in structure solution using bayesian estimates of map quality: The Phenix AUTOSOL wizard. *Acta Crystallogr D* 65:582-601.
7. Adams PD, *et al.* (2011) The Phenix software for automated determination of macromolecular structures. *Methods* 55(1):94-106.
8. Terwilliger TC (2000) Maximum-likelihood density modification. *Acta Crystallogr D* 56(Pt 8):965-972.
9. Terwilliger TC, *et al.* (2008) Iterative model building, structure refinement and density modification with the Phenix AUTOBUILD wizard. *Acta Crystallogr D* 64(Pt 1):61-69.
10. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of COOT. *Acta Crystallogr D* 66(Pt 4):486-501.
11. McCoy AJ, Storoni LC, & Read RJ (2004) Simple algorithm for a maximum-likelihood sad function. *Acta Crystallogr D* 60(Pt 7):1220-1228.
12. Nagahama K, *et al.* (1991) L-Arginine is essential for the formation *in-vitro* of ethylene by an extract of *Pseudomonas syringae*. *J Gen Microbiol* 137(7):1641-1646.
13. Case DA, *et al.* (2005) The amber biomolecular simulation programs. *J Comput Chem* 26(16):1668-1688.
14. Fox T & Kollman PA (1998) Application of the RESP methodology in the parametrization of organic solvents. *J Phys Chem B* 102(41):8070-8079.
15. Jeffrey GA (1997) *An introduction to hydrogen bonding* (Oxford University Press, New York ; Oxford).
16. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, & Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79(2):926-935.
17. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, & Haak JR (1984) Molecular-dynamics with coupling to an external bath. *J Chem Phys* 81(8):3684-3690.
18. Frisch MJ, *et al.* (2009) Gaussian 09 (Gaussian, Inc., Wallingford, CT, USA).
19. Cossi M, Rega N, Scalmani G, & Barone V (2003) Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J Comput Chem* 24(6):669-681.
20. Zhang ZH, Ren JS, Clifton IJ, & Schofield CJ (2004) Crystal structure and mechanistic implications of 1-aminocyclopropane-1-carboxylic acid oxidase - the ethylene-forming enzyme. *Chem Biol* 11(10):1383-1394.
21. Wilmouth RC, *et al.* (2002) Structure and mechanism of anthocyanidin synthase from *Arabidopsis thaliana*. *Structure* 10(1):93-103.