

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

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

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

**\*Corresponding Authors:**

Prof. Christopher J. Schofield, phone: +44(0)1865275625, christopher.schofield@chem.ox.ac.uk

Dr. Michael A. McDonough, phone: +44(0)1865275629, michael.mcdonough@chem.ox.ac.uk

**Keywords:** ethylene-forming enzyme • 2-oxoglutarate dependent oxygenases • hydroxylase • plant development and signalling • biocatalysis

**Abstract:** Ethylene is important in industry and biological signaling. In plants, ethylene is produced by oxidation of 1-aminocyclopropane-1-carboxylic acid, as catalyzed by 1-aminocyclopropane-1-carboxylic acid oxidase. Bacteria catalyze ethylene production, but *via* the four electron oxidation of 2-oxoglutarate to give ethylene in an arginine dependent reaction. Crystallographic and biochemical studies on the *Pseudomonas syringae* ethylene-forming enzyme reveal a branched mechanism. In one branch, an apparently typical 2-oxoglutarate oxygenase reaction to give succinate, carbon dioxide and, sometimes, pyrroline-5-carboxylate occurs. Alternatively, Grob-type oxidative fragmentation of a 2-oxoglutarate derived intermediate occurs to give ethylene and carbon dioxide. Crystallographic and quantum chemical studies reveal that fragmentation to give ethylene is promoted by binding of L-arginine in a non-oxidized conformation and of 2-oxoglutarate in an unprecedented high-energy conformation that favors ethylene rather than succinate formation.

**Significance:** The plant-signaling molecule ethylene is biosynthesized from 1-aminocyclopropane-1-carboxylic acid, as catalyzed by ACC oxidase, which is homologous to the 2-oxoglutarate (2OG) oxygenases, but which does not employ a 2OG cosubstrate. Bacteria produce ethylene in a highly unusual reaction but which involves oxidative 2OG fragmentation. Biophysical studies on a *Pseudomonas* ethylene-forming enzyme (EFE) reveal how structural and stereoelectronic factors enable it to bias reaction away from normal 2OG oxygenase catalysis involving two-electron substrate oxidation concomitant with succinate formation, to arginine-dependent four-electron oxidation of 2OG to give ethylene. The results imply that negative catalysis with respect to ethylene formation, has operated during evolution of 2OG oxygenases and will be useful in protein engineering aimed at optimising ethylene production.

/body

Ethylene is of industrial importance and is a vital signalling molecule in plants, where it has roles in germination, senescence, and stress responses (1). Commercial manipulation of the natural ethylene response is agriculturally important in controlling fruit ripening (2). In higher plants, ethylene is produced from methionine, *via* oxidation of 1-aminocyclopropane-1-carboxylic acid (ACC) in an unusual reaction catalyzed by the Fe(II) dependent ACC oxidase (ACCO)(3, 4), which is part of the 2-oxoglutarate (2OG)-dependent oxygenase superfamily, though it does not use a 2OG cosubstrate (Fig. 1a)(5-7). Ethylene is also produced in some microorganisms by oxidation of 2-oxo-4-methylthiobutyric acid in a reaction not directly enzyme catalyzed (8, 9).

In work aimed at producing industrial ethylene by biocatalysis, *Pseudomonas* strains, including plant pathogens, were shown to produce large amounts of ethylene (10-14). Bacteria engineered to produce ethylene using the *Pseudomonas syringae* pv. phaseolicola ethylene-forming enzyme (PsEFE) have been developed to ripen fruit as an alternative to use of synthetic ethylene (15, 16). Ethylene-forming enzymes are being explored for biocatalysis in cyanobacteria (17-19). PsEFE catalyzed ethylene production is 2OG dependent, and stimulated by the addition of L-arginine (L-Arg), which is also converted by PsEFE into pyrroline-5-carboxylate (P5C)(Fig. 1b)(13, 20). In contrast to the consensus 2OG oxygenase mechanism, which involves sequential binding of 2OG, substrate, then oxygen, an unprecedented 'dual circuit' mechanism is proposed for PsEFE (13).

We describe biochemical, structural, and modelling studies supporting a branched mechanistic pathway for PsEFE that can lead either to ethylene *via* oxidative fragmentation of 2OG, or to succinate *via* a more typical 2OG oxygenase reaction, which sometimes results in P5C formation (Fig. 1b). The unusual Grob-type fragmentation of 2OG to give ethylene is reliant on L-Arg binding in a conformation unsuited for oxidation and of the 2OG C-5 carboxylate binding in an unprecedented high-energy conformation that promotes ethylene formation. The results will enable protein engineering efforts to optimise biocatalytic ethylene production by

optimising the branch of the bifurcating mechanism for ethylene and have implications for the evolution of ethylene-forming enzymes.

## Results

**Biochemical Studies.** A recombinant form of the ethylene-forming enzyme from *Pseudomonas syringae* pv. phaseolicola (PsEFE) was produced in *Escherichia coli* and purified to near homogeneity. Using NMR and GC/MS based assays we found that, as reported (14, 21), PsEFE-catalyzed ethylene production is 2OG dependent, stimulated by addition of Fe(II) [some Fe(II) likely co-purifies with PsEFE], and is increased by addition of ascorbate or dithiothreitol (Figs. 1c-d, [SI Appendix, Fig. S1a-e](#)). Use of  $^{13}\text{C}$  and  $^2\text{H}$  labelled 2OG demonstrates ethylene is derived from 2OG and that  $\text{CO}_2$  is coproduced with ethylene ([SI Appendix, Fig. S1f-g](#)). Using a synthetic standard, we validated PsEFE catalyzed P5C production from L-Arg ([SI Appendix, Fig. S2a-f](#)); However, by NMR we only observed relatively low levels of P5C formation compared to succinate and ethylene (Fig. 3); guanidine formation was observed in the presence of L-Arg ([SI Appendix, Fig. S2g-f](#))(22). Consistent with the literature (21), ethylene and succinate production depends on L-Arg, with no activity being observed with D-Arg ([SI Appendix, Fig. S4e-f](#)). Of the other amino acids tested for ability to promote ethylene formation, only close L-Arg derivatives, including homoarginine and *N*-methylated/hydroxylated guanidino group derivatives, manifested substantial ethylene production ([SI Appendix, S4a-d](#)). In all cases, lower activities were observed compared to L-Arg. Kinetic analyses imply apparent substrate inhibition for L-Arg and 2OG, with  $K_m$  values of  $41.31 \pm 5.39$  and  $22.22 \pm 3.45$   $\mu\text{M}$ , respectively ([SI Appendix, Fig. S1c-d](#)). Similar conclusions have recently been made by Martinez and Hausinger in work on the PsEFE from *Pseudomonas syringae* pv. phaseolicola PK2 (21). These results imply a novel mechanism for ethylene production in which binding of arginine plays a key role. To investigate this, we performed structural studies on PsEFE.

**Crystallographic Analyses of PsEFE.** We obtained three *PsEFE* crystal structures: in complex with manganese and 2OG (to 1.45 Å resolution, *PsEFE:Mn:2OG*); in complex with manganese and the buffer bis-tris-propane (BTP) (to 1.55 Å resolution, *PsEFE:Mn:BTP*), and in complex with iron, L-Arg and the unreactive 2OG analogue, *N*-oxalylglycine (NOG) (to 1.08 Å resolution, *PsEFE:Fe:NOG:L-Arg*) (*SI Appendix, Table S1*). A structure was solved and a preliminary model of the *PsEFE:Mn:BTP* structure was built using single wavelength anomalous diffraction (SAD) data obtained from selenomethionine derivatized *PsEFE:Mn:BTP* crystals. The preliminary model was used to obtain initial phases for isomorphous higher resolution native data for *PsEFE:Mn:BTP* crystals and the model was fully refined against these data. This structure was used in molecular replacement to determine the *PsEFE:Fe:NOG:L-Arg* structure. The *PsEFE:Mn:BTP* and *PsEFE:Mn:2OG* crystals have the same *I*222 space group, with one molecule in the asymmetric unit; eight residues at the termini were not modelled due to disorder (a.a. 1-2 and 345-350). The *PsEFE:Fe:NOG:L-Arg* crystals have a different morphology and crystal form (space group *P*1) with four molecules in the asymmetric unit; disordered regions not modelled (chain A, a.a. 1, 298, and 341-350; chain B, 299 and 341-350; chain C, 1-2 and 343-350; and chain D, 342-350). There are small differences in the overall folds between the two crystal forms, with C $\alpha$  RMSDs ranging from 0.13 to 0.63 Å (*SI Appendix, Table S2*). *PsEFE* is predominantly a monomer in solution as shown by gel filtration.

**Overall Structure.** The overall *PsEFE* fold comprises 10  $\alpha$ -helices and 14  $\beta$ -strands, of which 8  $\beta$ -strands (I-VIII) form the ‘major’ and ‘minor’  $\beta$ -sheets of the conserved distorted double stranded beta helix (DSBH) - the 2OG oxygenase characteristic fold (Fig. 2a, *SI Appendix, Figs. S5 and S6*) (23-25).  $\beta$ -Strands  $\beta$ 1 and  $\beta$ 2 at the *N*-terminus extend the major  $\beta$ -sheet at the end of the DSBH away from the active site;  $\beta$ -strands  $\beta$ 3 and  $\beta$ 6, extend the other end of the major  $\beta$ -sheet close to the active site.  $\alpha$ -Helices  $\alpha$ 2 and  $\alpha$ 5 bind across the surface of the major  $\beta$ -sheet and likely stabilise it. A loop region (residues 80-93), located between  $\beta$ 3 and

$\beta$ 6, which harbours  $\beta$ 4 and  $\beta$ 5, acts as a lid partially covering the active site, and provides residues that bind the L-Arg cofactor/substrate. Three  $\alpha$ -helices ( $\alpha$ 8,  $\alpha$ 9,  $\alpha$ 10) at the C-terminus also contribute to the active site.

The C-terminal  $\alpha$ -helices  $\alpha$ 8,  $\alpha$ 9,  $\alpha$ 10, the  $\beta$ 4- $\beta$ 5 loop, and the ( $\beta$ 10- $\beta$ 11) DSBH IV-V 'insert' loop directly interact with L-Arg (e.g. Arg316 in  $\alpha$ 8 with the L-Arg carboxylate) (*SI Appendix, Fig S7i-k*). The role of these regions in *PsEFE* substrate binding is notable, because substrate binding by similar structural elements is 2OG oxygenase subfamily characteristic (24) (26). The involvement of all of these substrate recognition elements is very rare in other 2OG oxygenases. Thus, *PsEFE* should be regarded as a hybrid of subgroups I and II in terms of its classification (25).

**Active Site Geometry.** The overall metal-binding mode of *PsEFE* is relatively typical for 2OG oxygenases (Fig. 2), with the metal coordinated by His189 (C-terminus of DSBH II), Asp191 (loop linking DSBH II and III), and His268 (*N*-terminus of DSBH VII), consistent with mutagenesis studies (27). However, in the high resolution *PsEFE*:Fe:NOG:L-Arg structure, Asp191 adopts two conformations (see below). In the *PsEFE*:Mn:2OG and *PsEFE*:Fe:NOG:L-Arg structures, the metal is coordinated by 2OG/NOG in a bidentate manner; a single 2OG/NOG C-1 carboxylate oxygen ligates *trans* to His268 and the 2OG/NOG C-2 carbonyl ligates *trans* to Asp191 (*SI Appendix, Fig S8*). The implied typical mode of Fe-chelation by 2OG is supported by spectroscopic studies (21). A water occupies the site *trans* to His189, likely where dioxygen binds. This arrangement suggests rearrangement to position the Fe(IV)=O intermediate, likely responsible for L-Arg oxidation (giving P5C), proximal to C-5 of the L-Arg; such rearrangement may be unnecessary for ethylene formation (see below)(28-30). In the *PsEFE*:Mn:BTP structure, two BTP hydroxyls and its amino group form tridentate metal coordination; the two hydroxyl groups ligate *trans* to His268 and Asp191, and the BTP amino group *trans* to His189.

**Evidence for Induced Fit.** Despite the low overall RMSD for C $\alpha$  atoms, comparison of structures with and without L-Arg reveals striking local differences in the backbone conformations (*SI Appendix, Fig. S7h-i*), with regions surrounding the active site appearing to move to clamp L-Arg. In the *PsEFE:Fe:NOG:L-Arg* structure, the sidechain of Arg171 (located on DSBH I) is positioned to ‘ $\pi$ - $\pi$  stack’ with the L-Arg guanidine and form hydrogen bond and electrostatic interactions with the NOG C-1 carboxylate and Glu84 sidechain (an arginine at a similar position is involved in substrate/cofactor binding in some other 2OG oxygenases)(Fig. 2, *SI Appendix, Fig. S9c-e*)(25). Glu84 (on the  $\beta$ 4- $\beta$ 5 loop) rotates  $\sim 90^\circ$  about its C $\alpha$ -C $\beta$  bond (relative to the structure without L-Arg) into the active site to interact with the Arg171 guanidino group and both  $\alpha$ -amino and guanidino groups of Conformation A (see below) of L-Arg (*SI Appendix, Fig. S7k*). Consistent with the proposed catalytic importance of these residues, the R171K, R171A, E84Q and E84D variants do not produce ethylene (*SI Appendix, Table S3*). NMR assays indicate neither succinate nor P5C are produced by the R171K and E84Q variants. Compared to its position in the absence of L-Arg, the sidechain of Tyr192 (on the DSBH II-III loop immediately after the HXD motif) rotates  $\sim 90^\circ$  about its C $\alpha$ -C $\beta$  bond into the active site to form a hydrogen bonding interaction with the L-Arg carboxylate and offset  $\pi$ - $\pi$  stack with the L-Arg guanidino group (*SI Appendix, Fig. S7j*); consistent with a role for Tyr192 in catalysis, the Y192F variant manifests substantially reduced ethylene production ( $\sim 5\%$  of wildtype) (*SI Appendix, Table S3*).

The  $\alpha$ 8,  $\alpha$ 9, and  $\alpha$ 10 C-terminal helical region shows the largest local movement towards the active site (up to 3.0 Å), as observed in all molecules in the asymmetric unit (*SI Appendix, Fig. S7h-i*). The observed movements involve Arg316 (on  $\alpha$ 8), the sidechain of which rotates  $\sim 180^\circ$  about its C $\alpha$ -C $\beta$  bond to form a salt bridge with the L-Arg carboxylate and hydrogen bond to the sidechain of Asn220 located on the DSBH IV-V loop (*SI Appendix, Fig. S7k*). The R316K and R316A variants are less active ( $\sim 13$  and 4% of wildtype ethylene-forming activity) (*SI Appendix, Table S3*). The sidechain of Phe283 (on DSBH VIII) rotates  $\sim 17^\circ$  about its C $\alpha$ -C $\beta$

bond and  $\sim 45^\circ$  about its C $\beta$ -C $\gamma$  bond to form part of the binding pocket for the L-Arg guanidino group (*SI Appendix, Fig. S7j*). Collectively, these results imply substantial induced fit upon L-Arg binding and, since L-Arg is required for ethylene production, suggest that evolution of loop residues was important in the development of the ethylene-forming activity.

**Alternative Substrate Conformations.** In all four molecules in the asymmetric unit of the high resolution *PsEFe:Fe:NOG:L-Arg* structure, L-Arg and NOG are observed in two conformations (L-Arg<sub>A</sub> and L-Arg<sub>B</sub>, NOG<sub>A</sub> and NOG<sub>B</sub>), which correlate with two sidechain conformations of the metal-ligating Asp191 (each conformation was assigned an occupancy of 50%)(Fig. 2b-e and *SI Appendix, Fig. S8*). Analysis of the structures reveals L-Arg<sub>A</sub>/NOG<sub>A</sub>/Asp191<sub>A</sub> and L-Arg<sub>B</sub>/NOG<sub>B</sub>/Asp191<sub>B</sub> correlate since the alternative combinations are not possible due to steric clashes. In both cases, monodentate metal coordination by Asp191 is observed and the shift in its carboxylate oxygen determines its interaction with the L-Arg guanidino group. Metal coordination of Asp191 in the two structures without L-Arg bound is most like conformation B in the L-Arg bound structure.

Both the  $\alpha$ -amino acid and guanidino groups of L-Arg appear precisely bound in L-Arg<sub>A</sub> and L-Arg<sub>B</sub>. Key residues involved in L-Arg binding include Tyr192 and Arg316 (to its carboxylate), and Cys317 and Thr86 (to its  $\alpha$ -amino group). Glu84 (L-Arg<sub>A</sub> only), Tyr192, Tyr318, Arg171, and Asp191 are important for interactions with the L-Arg guanidine (*SI Appendix, Fig. S7k*). Substitutions of these residues manifest substantially decreased ethylene production activity, except for the semi-conserved Y318F variant ( $\sim 65\%$  of wildtype) (*SI Appendix, Table S2*).

The positions of the L-Arg  $\alpha$ -amino acid group (up to the C-3/C-4 methylenes) are very similar in L-Arg<sub>A</sub> and L-Arg<sub>B</sub>, but the positions of their C-5 methylene and guanidino groups clearly differ (Fig. 2b-e). In Conformation B, the C-5 methylene of L-Arg<sub>B</sub> is directed away from the metal, *i.e.* it is not positioned for hydroxylation (Fig. 2e and *SI Appendix, Figs. S8a-d and S9b*). Thus, L-Arg<sub>B</sub> likely correlates with ethylene formation (Fig. 3). In contrast, L-Arg<sub>A</sub> projects its C-5 hydrogens towards the metal (Fe to L-Arg C5 distance 4.5 Å) (Fig. 2d *SI Appendix, Fig.*

[S9a](#)). L-Arg<sub>A</sub> is thus predicted to be the conformation yielding C-5 hydroxylation of L-Arg leading to P5C (Fig 3).

The distance and geometry between the L-Arg<sub>A</sub> and metal (4.3 - 4.5 Å) are similar to that in substrate structures of other 2OG oxygenases catalysing hydroxylation, including L-Arg (derivative) hydroxylases VioC [PDB 2WBO (31)] and clavaminic acid synthase, (CAS)[PDB 1DRY (28)] ([SI Appendix, Fig. S9d-e](#)). However, VioC and CAS both catalyze hydroxylation at C-3 and their substrates are bound flipped ~180° relative to the *PsEFE* binding mode. Most residues involved in L-Arg binding in *PsEFE* are not conserved in VioC/CAS except for the Arg171 equivalents, which bind the α-amino acid group as opposed to the guanidine group in *PsEFE*.

*PsEFE* catalyzed C-5 L-Arg hydroxylation gives an unstable C-5 hydroxylated product, equivalent to the hemiaminal intermediate in histone demethylase catalysis (32). This *PsEFE* intermediate, as yet not observed by NMR, can fragment to give guanidine and glutamate semi-aldehyde, which is in equilibrium with 5-hydroxyproline and P5C with the latter being preferred at neutral pH (33). Fragmentation of the nascent C-5 hydroxylated product could occur in solution or at the active site; in the latter regard, the apparently constrained bent conformation adopted by the arginine sidechain (in both L-Arg<sub>A</sub> and L-Arg<sub>B</sub>) (Fig. 2b) is notable, suggesting aldehyde formation and cyclisation at the active site (22, 33).

**2-Oxoglutarate Binding.** There are striking differences between the *PsEFE* 2OG pocket and those of other 2OG oxygenases ([SI Appendix, Fig S11](#)) (25). The *PsEFE* 2OG binding pocket is highly and unusually hydrophobic, being lined by the sidechains of Leu173, Phe175, Ile186, Ala199, Leu206, Val270, Ala279, Ala281 and Phe283 ([SI Appendix, Fig. S10](#)). A single, polar residue, Arg277, is positioned at the base of the pocket, the guanidino group which interacts with the 2OG/NOG C-5 carboxylate. For NOG<sub>A</sub> this interaction is bidentate, while for 2OG and NOG<sub>B</sub> it is monodentate. Other 2OG oxygenases have a similarly positioned Arg or Lys but have additional polar residues (Tyr/Ser/Thr) involved in 2OG binding, *via* hydrogen bonds, as

exemplified in the 2OG C-5 carboxylate binding “RXS motif” in the clavaminic acid synthase (CAS) subfamily (*SI Appendix, Fig. S11*)(23).

There are several waters in the 2OG pocket (Fig. 2b-e and *SI Appendix Fig. S11*). In the NOG structure, a water hydrogen bonds with the NOG C-5 carboxylate; In the *PsEFE:Mn:2OG* structure, two waters hydrogen bond with the C-5 carboxylate (*SI Appendix Fig. S7d*). A water filled channel leads from the exterior to the 2OG C-5 carboxylate, possibly reflecting an oxygen entry/CO<sub>2</sub> exit route (*SI Appendix Fig. S8e*). Another water in the NOG/2OG pocket bridges between the metal ligated water and the C-5 carboxylate in 2OG/NOG<sub>A</sub> (2.8 Å) (Fig. 2b-d); the combined binding of the C-5 bound water and that ligating the metal may in part reflect a dioxygen binding pocket. During 2OG oxygenase catalysis, after ligating to Fe(II), the dioxygen reacts with the 2OG ketone to give a cyclic peroxide, which then collapses to give the well characterized Fe(IV)=O intermediate effecting substrate hydroxylation (Fig. 3). With *PsEFE*, the results imply that the Fe(IV)=O intermediate (or possibly, but less likely, a cyclic peroxide) undergoes Grob-type fragmentation to give ethylene and three CO<sub>2</sub> molecules (Fig. 3). The question arises as to why ethylene formation occurs in *PsEFE* and not other 2OG oxygenases, where it has not been reported.

With *PsEFE* the methylenes of 2OG and NOG adopt an extended conformation as observed in other 2OG oxygenases (*SI Appendix, Fig. S11*)(25). Strikingly, however, by contrast to other 2OG oxygenase structures, in which the C-5 carboxylate plane (C-5/O-3/O-4) of 2OG/NOG is coplanar with the C-5/C-4/C-3 methylene carbon plane (*SI Appendix, Fig. S11*), with both 2OG and NOG, the C-5 carboxylate is not coplanar with the 2OG/NOG C-5/C-4/C-3 methylenes. In the refined conformations of NOG, in the four chains (A-D) in the asymmetric unit of *PsEFE:Fe:NOG:L-Arg* structure, the torsion angles between the carboxylate O/C-5 and C-4/C-3 bonds are consistently close to 45°; in chain A: (NOG<sub>A</sub>: 39.7°, 38.5°; NOG<sub>B</sub>: 45.3°, 43.7°); chain B: (NOG<sub>A</sub>: 40.8°, 41.2°; NOG<sub>B</sub>: 46.0°, 44.5°), chain C (NOG<sub>A</sub>: 44.7°, 43.8°; NOG<sub>B</sub>: 46.0°,

45.0°), chain D (NOG<sub>A</sub>: 47.6°, 46.8°; NOG<sub>B</sub>: 40.5°, 42.0°). In the *PsEFE*:Mn:2OG complex, the corresponding torsion angles for 2OG are 71.0°.

**MD Simulations and QM.** We investigated the stability of the unusual ‘strained’ 2OG conformation observed in the *PsEFE* structures by molecular dynamics (MD) simulations using a *PsEFE*:Fe:2OG:L-Arg model based on the high-resolution *PsEFE*:Fe:NOG:L-Arg structure (*SI Appendix*, Fig. S12). 2OG maintained the strained conformation and its interactions with Arg277 for 94% of the simulation time (assuming an O-H hydrogen bond distance limit of 2.2 Å) (*SI Appendix*, Table S4 and Fig. S12), suggesting the crystallographically observed conformation is catalytically relevant.

Surprisingly, the conformational preference for Grob-type fragmentation (34), involving entropically favoured decarboxylation to give alkenes, has not been investigated in detail, even for small molecules. To compare the reactivity of the coplanar and strained conformations, we carried out quantum calculations of propionate derivatives with chloro, bromo, and trimethylamino bearing groups with appropriate torsion angles 0°, 45°, and 90° (*SI Appendix*, Table S4 and Fig. S13). We also studied analogous reactions with (*tert*-butyl)-cyclohexanecarboxylate derivatives with the carboxylate and leaving group locked in a *trans* diaxial arrangement. The differences in the  $\Delta G^\ddagger$  (transition state free energy difference) values for fragmentation were consistently in favour of a torsion angle of 90°, with differences in  $\Delta G^\ddagger$  between the two conformations ranging from -2.7 to -4 kcal·mol<sup>-1</sup> (mean -3.6) (*SI Appendix*, Table S4). A similar, though smaller, trend was observed for the 45° torsion angle (*SI Appendix*, Table S4). These results support the potential for fragmentation of the crystallographically implied strained 2OG cosubstrate conformation to give ethylene.

**Discussion.** The overall fold of *PsEFE* is relatively typical of 2OG oxygenases (Fig. 2a); However, it contains highly distinctive features, including elements from group I and II 2OG oxygenases (25), which reveal *PsEFE* to have a hybrid fold (*SI Appendix*, Fig S6). Notably, the conformation of *PsEFE* and the inclusion of a  $\beta$ (IV)- $\beta$ (V) insert loop involved in substrate

recognition is clearly distinct from that of ACCO, the only other enzyme known to directly catalyze ethylene production (35). Interestingly, sequence alignments of predicted EFEs (65-50% sequence identity [ID] to *PsEFE*), a related EFE-like subfamily (22-20% ID) and ACCOs (16-17% ID) show conservation in some of the key L-Arg interacting motifs, and differences in the “RXS” motif suggesting they may not share a *PsEFE* 2OG fragmenting mechanism, but do share a common ancestor (*SI Appendix, Figs. S14-S15*). The difference in the overall folds of *PsEFE* and ACCO (as well as more typical 2OG oxygenases) is mirrored in specific differences at their active sites, which reflect the different reactions they catalyze. The active site differences leading to oxidative fragmentation of 2OG to give ethylene rather than the normal succinate formation, relate to the enzyme-bound conformation of 2OG and subsequently derived intermediates. The overall shape of the *PsEFE* active site is similar to that of normal 2OG oxygenases; indeed *PsEFE* catalyzes a normal reaction, i.e. L-Arg hydroxylation. Thus, it appears likely that most 2OG oxygenases have actively evolved *not* to catalyze fragmentation of 2OG to give ethylene. The evolution of such ‘negative catalysis’ (36) may have occurred since succinate is a useful byproduct, which can be recycled (e.g. by the TCA cycle), whereas CO<sub>2</sub> and gaseous ethylene are not primary metabolites (as far as is known). In support of this, recent work on cancer metabolism suggests that, at least in animals, 2OG oxygenase catalysis may be metabolically linked to the TCA cycle (37).

A standout feature at the *PsEFE* active site concerns its 2OG/NOG-binding mode (Fig. 2b-e and *SI Appendix, Figs. S7, S10, S11*). 2OG/NOG binds in an unusual highly hydrophobic pocket and adopts a strained conformation that promotes Grob-type fragmentation of the iron-ligated succinate at the ferryl intermediate stage in the reaction (Fig. 3) to give two molecules of CO<sub>2</sub> and ethylene. Ethylene formation is promoted by the presence of L-Arg, which is not oxidized during ethylene formation. Thus, in the ethylene-forming pathway, 2OG undergoes a four-electron oxidation. This differs from the ACCO reaction, where ACC undergoes a two-electron oxidation, with reaction cycles for ACCO being proposed to be completed by

ascorbate oxidation (38). Four-electron substrate oxidations are rare in 2OG oxygenase superfamily catalysis, but occurs in isopenicillin N synthase (IPNS) catalyzed oxidation of a dipeptide to isopenicillin-N; ironically, the IPNS structure is more closely related to ACCO than *PsEFE* (39).

Quantum calculations reveal that the crystallographically implied conformation of the 2OG derived intermediate in *PsEFE* is crucial in enabling fragmentation to give ethylene, rather than substrate hydroxylation and succinate, as in normal 2OG oxygenase catalysis ([SI Appendix, Fig. S12](#)). There may also be other features at the iron centre that are important in promoting ethylene formation, as evidenced by the different ligation predicted for the iron-ligating Asp191 in the ethylene-forming pathway (Figs. 2b-e).

In addition to ethylene production, *PsEFE* also catalyzes a typical 2OG oxygenase type hydroxylation, albeit one leading to an interesting cyclization / fragmentation to give P5C and guanidine (Fig. 3). However, as shown by NMR analyses (Fig. 1d), the extent of P5C formation is lower than that of ethylene and succinate in our current assays. Thus, P5C formation appears substantially uncoupled from succinate formation.

Most, crystallographic analyses of 2OG oxygenases imply that the 2OG C-5 carboxylate is near coplanar with the C-5/C-4/C-3 plane, *i.e.* there is a near 0° syn-torsion angle ([SI Appendix, Fig. S11](#)). In *PsEFE*, the corresponding torsion angle has a mean of 44° in the NOG structure and is 71° in the 2OG structure, which although is not the predicted optimum of 90° for ethylene formation, will promote fragmentation to ethylene, (the torsion angle during solution catalysis by *PsEFE* is unknown and may be closer to 90°). The observation of a torsion angle less than the optimal of 90° for ethylene production is consistent with the observation of succinate formation and some hydroxylation to give P5C (Fig. 1). The balance between fragmentation to give either ethylene or succinate formation, thus appears poised.

The structural insights into *PsEFE* will help in optimising biocatalytic ethylene production by minimizing succinate formation to promote ethylene formation and by developing *PsEFE*

variants not requiring arginine for ethylene-forming activity. Such work is of interest both with respect to industrial biocatalytic ethylene production (17) and to produce biological ethylene for agricultural applications (15).

*PsEFE* is proposed to produce ethylene to regulate plant physiology (e.g. ripening) beneficial to the bacteria (16). Some 2OG oxygenases have metabolic (e.g. in fatty acid metabolism) or sensing (e.g. in hypoxia sensing) roles in eukaryotes (40, 41). *PsEFE* has considerable potential for a sensing role, since in addition to employing 2OG and oxygen and producing ethylene, succinate, and CO<sub>2</sub>, it requires a redox sensitive metal cofactor and converts arginine to P5C, a proline/glutamate/glutamine precursor.

The results raise the possibility that other 2OG oxygenases make ethylene, perhaps in addition to catalysing more typical reactions. It is possible that the unusual conformations of 2OG observed with *PsEFE* are diagnostic of ethylene formation. Interestingly, the *PsEFE* active site architecture has similarities with those of the ten-eleven translocation (TET) 2OG oxygenases (for *PsEFE* and the TET from *Neurospora crassa* C $\alpha$  RMSD = 1.2 Å over 185 residues) (42). Arg171 of *PsEFE* is conserved in the two enzymes; in the *NcTET* structure, the analogous Arg190 interacts with its nucleosome substrate and with the 2OG C-1 carboxylate (*SI Appendix, Fig. S9c*). Substrate binding by *NcTET* involves Tyr217/Phe292, in a similar manner to which Tyr192/Phe283 of *PsEFE* interact with L-Arg (*SI Appendix, Fig. S9c*). However, the RXS motif which binds the 2OG C-5 carboxylate in many 2OG oxygenases, including the TETs, is not present in *PsEFE*, where the serine is replaced by an alanine (*SI Appendix, Fig. S5 and S14*). Thus, although it is unlikely that TETs catalyze ethylene production, other 2OG oxygenases may do so. The possibility of additional ethylene-forming enzymes is of interest in plant biology / agriculture, and it cannot be ruled out that some animals make ethylene. An early report describes ethylene formation in animal tissues, though the mechanism has yet to be validated (microbial/non-enzymatic processes are possible) (43).

We are currently exploring the possibility of other 2OG oxygenase related ethylene-forming enzymes.

## Experimental Section

A synthetic PsEFE DNA sequence optimized for *E. coli* expression based on the gene sequence from *Pseudomonas syringae* pv. phaseolicola (UNIPROT ID P32021) (GeneArt ThermoFisher) was subcloned into pCold-1 (Clontech) or pETite (Lucigen) vectors. Variants were made by mutagenesis using the polymerase chain reaction. Vectors were transformed into *E. coli* strain BL21(DE3) for protein production. Proteins were purified by standard nickel affinity and gel filtration chromatography. Crystals were grown by the sitting drop vapour diffusion method (200-300 nL drops) and data collected at the Diamond Light Source (Didcot UK). The initial PsEFE:Mn:BTP structure was determined by SAD and refined in Phenix. Initially, PsEFE activity was assayed using an optimized version of the reported GC method (20). Purified PsEFE (between 10 to 15  $\mu$ M) was added to 150  $\mu$ L of 50 mM HEPES buffer, pH 7.2, containing 5 mM ascorbate or DTT, 5 mM 2-oxoglutarate, 2 mM L-Arg and 200  $\mu$ M of  $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$ .  $^1\text{H}$ -NMR assays (at 298 K) were performed using a Bruker Avance III 700 MHz machine with a TCI inverse cryoprobe; water suppression employed presaturation. MD simulations used the AMBER program (version 12) and force field parameters of Cornell *et al.* (44). The PsEFE:Fe:NOG:L-Arg crystal structure was used for QM studies; all structures corresponding to energy minima and transition states for 3-chloropropanoate fragmentation were optimized at the B3LYP/6-31+G\* level using Gaussian 09 (45). For full experimental details see [Supplementary Information](#).

**Data deposition:** Atomic coordinates and structure factors are deposited in the Protein Data Bank under accession codes 5LSQ (PsEFE:Mn:BTP), 5MOF (PsEFE:Mn:2OG) and 5LUN (PsEFE:Fe:NOG:L-Arg).

**Author contributions:** C.J.S. conceived the research. Z.Z. and T.J.S. cloned and produced the enzymes used in biochemical studies. Z.Z., M.S.W.R., F.H. carried out GC/MS. T.J.S. carried out LC/MS. Z.Z. performed crystallization. M.A.M. collected X-ray data and determined the crystal structures. T.J.S., C.T.L. and M.I.A. designed and performed the NMR experiments. H.C. and R.S.P.

performed MDS and QM analysis. Z.Z., M.A.M. and C.J.S. wrote the manuscript. All authors discussed the results and contributed to writing the final manuscript.

## Acknowledgements

We thank the BBSRC NIBB Metals in Biology POC award BB/L013711/1 and the Wellcome Trust for funding, and the staff at the Diamond Light Source for access and technical assistance.

## Figure Legends

**Figure 1.** Ferrous iron-dependent oxygenase catalyzed ethylene formation. **(a)** 1-Aminocyclopropane-1-carboxylate oxidase (ACCO), which catalyzes ethylene formation in plants and fungi, does not use 2OG, but belongs to the 2OG oxygenase superfamily. ACCO catalysis is stimulated by CO<sub>2</sub>, which is a coproduct; ascorbate is a proposed cosubstrate. **(b)** *PsEFE* catalyzes ethylene production in a manner stimulated by L-Arg; pyrroline-5-carboxylate (P5C) is a coproduct. **(c)** <sup>1</sup>H NMR showing the dependence of *PsEFE* on L-Arg for ethylene and succinate formation from 2OG (30 min). **(d)** Partial <sup>1</sup>H NMR spectra (700 MHz) showing time-dependent formation of P5C, ethylene, and succinate by *PsEFE*. Note the relatively low P5C levels compared to succinate/ethylene.

**Figure 2.** The overall fold of *PsEFE* and the two conformations observed for L-Arg and NOG. **(a)** The eight  $\beta$ -strands of the core double-stranded  $\beta$ -helix (DSBH) are in red and labelled I - VIII. Non-DSBH  $\beta$ -strands are yellow, helices in marine blue, and loops gray. The iron-ligating residues and substrates are sticks. **(b)**  $2mF_o - DF_c$  electron density map contoured to  $1\sigma$  showing the two refined conformations of L-Arg and NOG. **(c)** Residues ligating to the iron (magenta) and involved in substrate binding. Two waters are present in the 2OG binding pocket with one coordinating to the metal. Conformations A (yellow) and B (cyan) are shown. **(d and e)** Conformations A and B of NOG and L-Arg. Conformation A, in which an L-Arg C-5 methylene hydrogen projects toward the metal enables hydroxylation (leading to pyrroline-5-carboxylate); conformation B leads to ethylene. Note that Asp191 conformations A and B also interact differently with the metal.

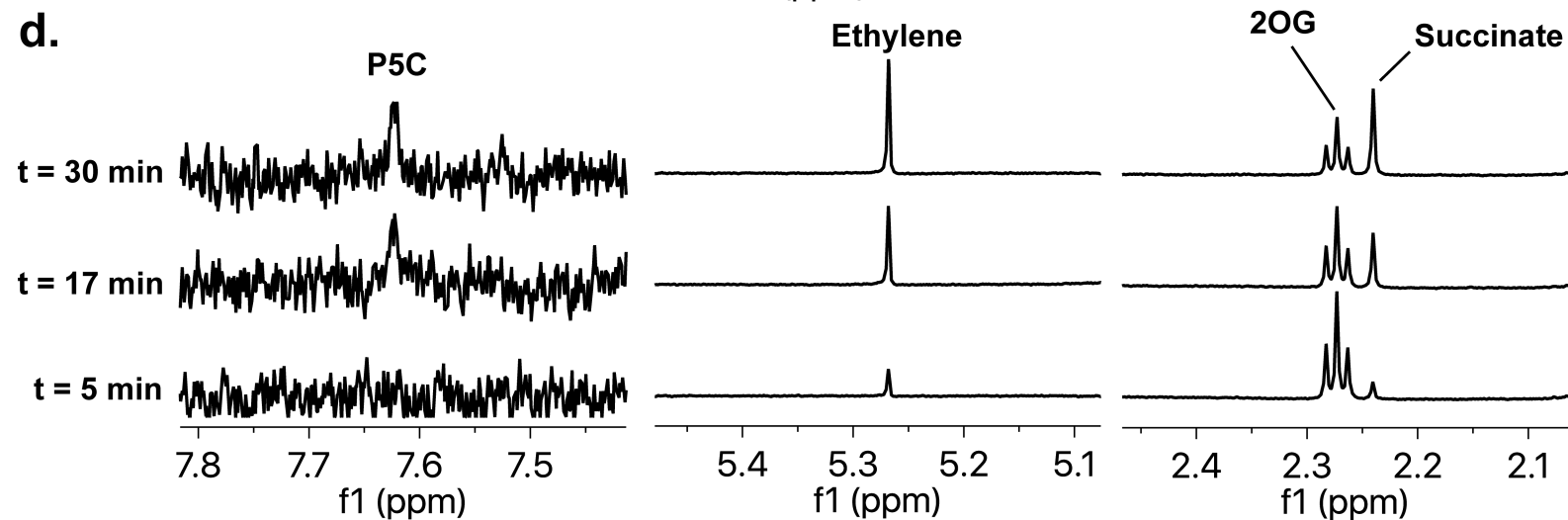
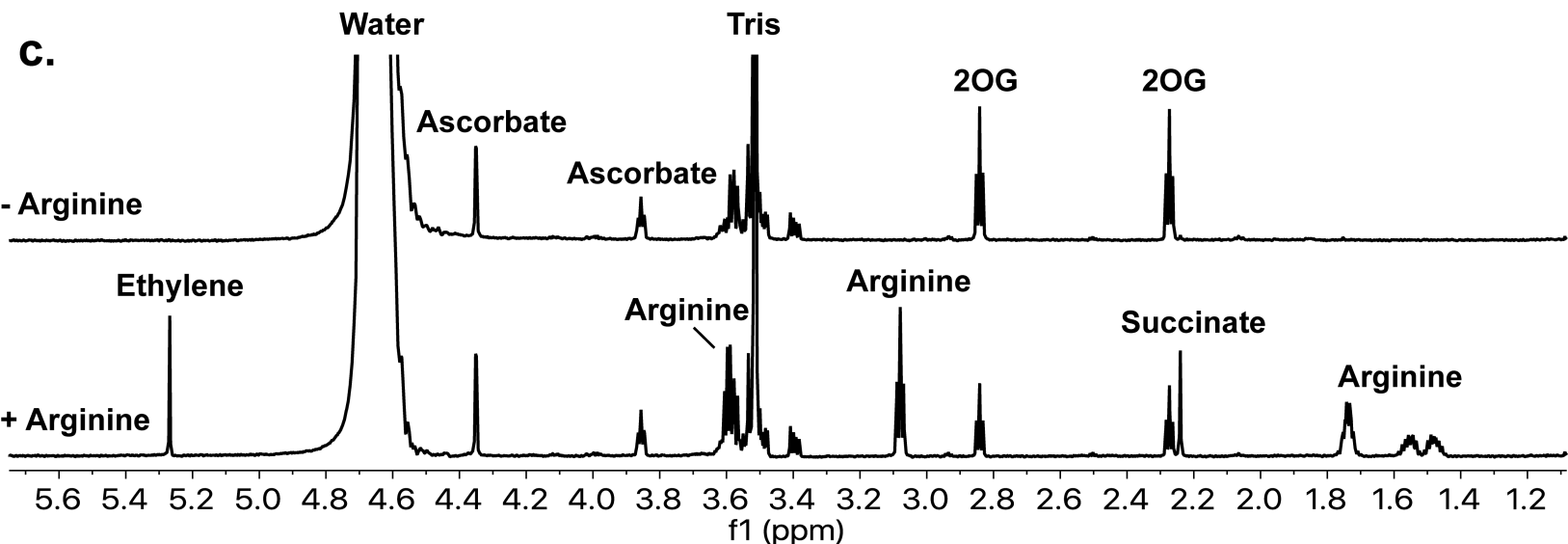
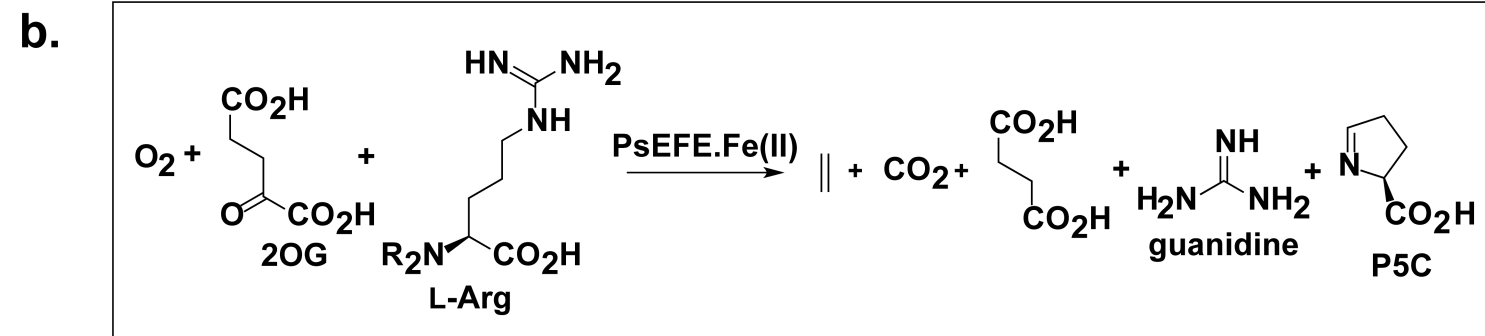
**Figure 3. Proposed outline mechanism.** Ethylene and succinate formation occur in the presence of L-Arg (Fig. 1). Variations are possible, e.g., fragmentation of the cyclic peroxide intermediate **II** to give ethylene (21), but are less likely, given the prevalence of intermediates of type **III** in 2OG oxygenase catalysis. NOG (and by implication 2OG) and L-Arg were observed in two conformations correlating with different Asp191 binding modes (Fig. 2). Formation of the bidentate 2OG intermediate complex **I** is

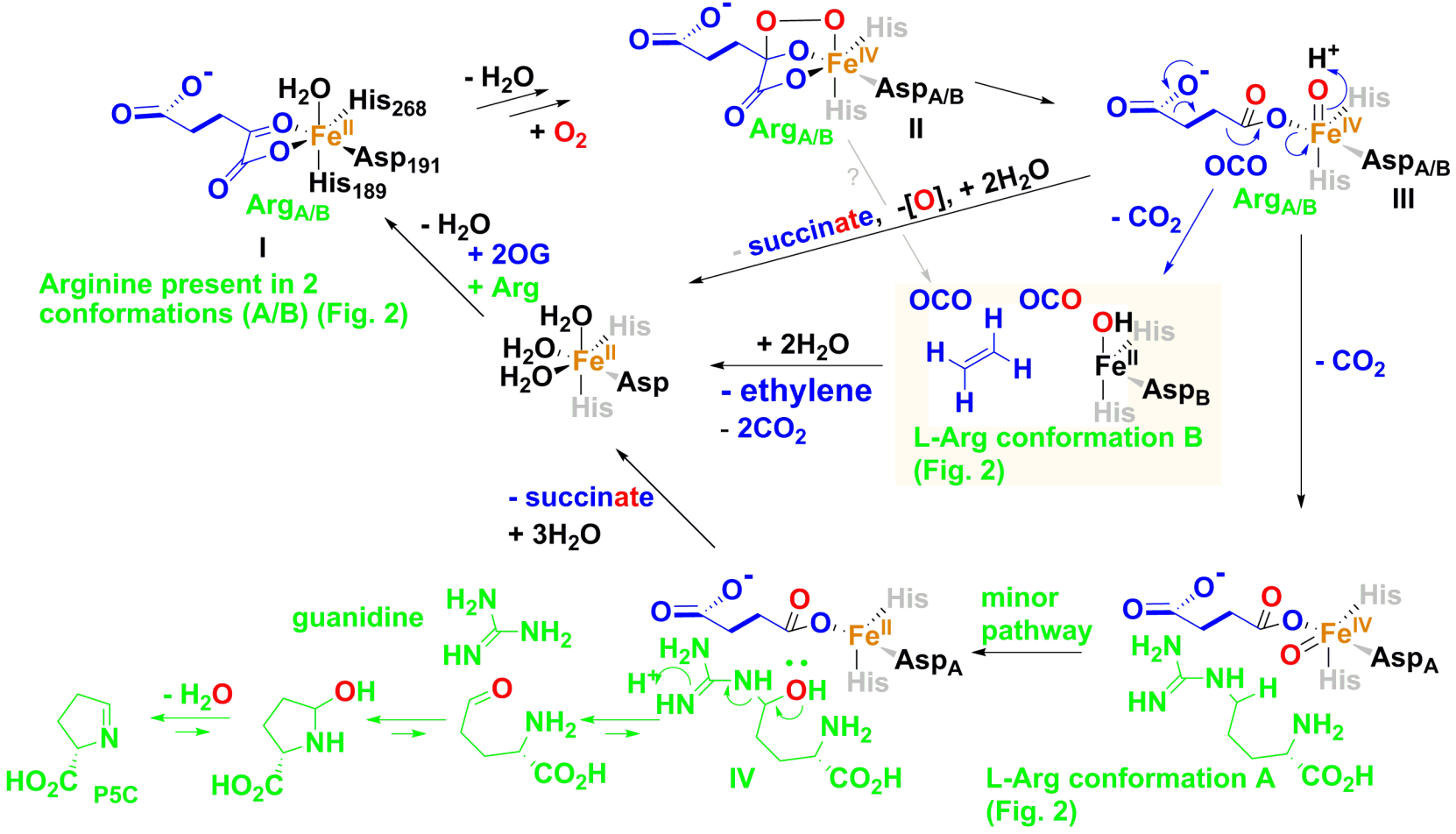
supported by spectroscopic work (21). Turnover of 2OG to succinate also occurs (Fig. 1); succinate can be released from **III**, possibly with subsequent formation of H<sub>2</sub>O<sub>2</sub>. Formation of pyrroline-5-carboxylate (P5C) occurs at a lower level than ethylene/succinate formation, *via* reaction of an Fe(IV)=O species (**III**), with L-Arg C-5. The hydroxylated product (**IV**) fragments and condenses to form P5C (21).

## References

1. Abeles FB, Morgan PW, & Saltveit ME (1992) *Ethylene in plant biology* (Academic Press, San Diego; London) 2nd Ed.
2. Bleecker AB & Kende H (2000) Ethylene: A gaseous signal molecule in plants. *Annu Rev Cell Dev Bi* 16:1-18.
3. Yang SF & Hoffman NE (1984) Ethylene biosynthesis and its regulation in higher-plants. *Annu Rev Plant Phys* 35:155-189.
4. Kende H (1993) Ethylene biosynthesis. *Annu Rev Plant Phys* 44:283-307.
5. McRae DG, Coker JA, Legge RL, & Thompson JE (1983) Bicarbonate CO<sub>2</sub>-facilitated conversion of 1-amino-cyclopropane-1-carboxylic acid to ethylene in model systems and intact tissues. *Plant Physiol* 73(3):784-790.
6. Dong JG, Fernández-Maculeit JC, & Yang SF (1992) Purification and characterization of 1-aminocyclopropane-1-carboxylate oxidase from apple fruit. *Proc Natl Acad Sci USA* 89(20):9789-9793.
7. 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.
8. Primrose SB (1977) Evaluation of role of methional, 2-keto-4-methylthiobutyric acid and peroxidase in ethylene formation by *Escherichia coli*. *J Gen Microbiol* 98(2):519-528.
9. Ogawa T, Takahashi M, Fujii T, Tazaki M, & Fukuda H (1990) The role of NADH-Fe(III)-EDTA oxidoreductase in ethylene formation from 2-keto-4-methylthiobutyrate. *J Ferment Bioeng* 69(5):287-291.
10. Goto M & Hyodo H (1987) Ethylene production by cell-free-extract of the Kudzu strain of *Pseudomonas-syringae* pv. Phaseolicola. *Plant Cell Physiol* 28(3):405-414.
11. Goto M, Ishida Y, Takikawa Y, & Hyodo H (1985) Ethylene production by the Kudzu strains of *Pseudomonas syringae* pv. Phaseolicola causing halo blight in *Pueraria-lobata* (Willd) Ohwi. *Plant Cell Physiol* 26(1):141-150.
12. Fukuda H, Fujii T, & Ogawa T (1986) Preparation of a cell-free ethylene-forming system from *Penicillium digitatum*. *Agr Biol Chem Tokyo* 50(4):977-981.
13. Fukuda H, *et al.* (1992) 2 reactions are simultaneously catalyzed by a single enzyme - the arginine-dependent simultaneous formation of 2 products, ethylene and succinate, from 2-oxoglutarate by an enzyme from *Pseudomonas syringae*. *Biochem and Biophys Res Commun* 188(2):483-489.
14. Nagahama K, *et al.* (1991) Purification and properties of an ethylene-forming enzyme from *Pseudomonas syringae* pv Phaseolicola-PK2. *J Gen Microbiol* 137(10):2281-2286.
15. Digiacomo F, *et al.* (2014) Ethylene-producing bacteria that ripen fruit. *Acs Synth Biol* 3(12):935-938.
16. Weingart H, Ullrich H, Geider K, & Völksch B (2001) The role of ethylene production in virulence of *Pseudomonas syringae* pvs. Glycinea and Phaseolicola. *Phytopathology* 91(5):511-518.
17. Eckert C, *et al.* (2014) Ethylene-forming enzyme and bioethylene production. *Biotechnol Biofuels* 7(1).
18. Xiong W, *et al.* (2015) The plasticity of cyanobacterial metabolism supports direct CO<sub>2</sub> conversion to ethylene. *Nat Plants* 1(5053).
19. Ungerer J, *et al.* (2012) Sustained photosynthetic conversion of CO<sub>2</sub> to ethylene in recombinant cyanobacterium *Synechocystis* 6803. *Energ Environ Sci* 5(10):8998-9006.
20. 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.
21. Martinez S & Hausinger RP (2016) Biochemical and spectroscopic characterization of the non-heme Fe(II) and 2-oxoglutarate dependent ethylene-forming enzyme from *Pseudomonas syringae* pv. Phaseolicola PK2. *Biochemistry* 55(43):5989-5999.
22. Batchelar ET, *et al.* (2008) Thioester hydrolysis and C-C bond formation by carboxymethylproline synthase from the crotonase superfamily. *Angew Chem Int Ed* 47(48):9322-9325.
23. Clifton IJ, *et al.* (2006) Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J Inorg Biochem* 100(4):644-669.

24. Aik W-S, McDonough MA, Thalhammer A, Chowdhury R, & Schofield CJ (2012) Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Curr Opin Struc Biol* 22(6):691-700.
25. Aik W-S, *et al.* (2015) Introduction to structural studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *2-oxoglutarate-dependent oxygenases*, RSC Metallobiology, eds Hausinger RP & Schofield CJ (The Royal Society of Chemistry, London), 1 Ed, pp 59-94.
26. McDonough MA, *et al.* (2006) Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2). *Proc Natl Acad Sci USA* 103(26):9814-9819.
27. Que L (2000) One motif - many different reactions. *Nat Struct Biol* 7(3):182-184.
28. Zhang ZH, *et al.* (2002) Crystal structure of a clavamate synthase-Fe(II)-2-oxoglutarate-substrate-NO complex: Evidence for metal centred rearrangements. *Febs Lett* 517(1-3):7-12.
29. Bollinger JMJ, Chang W-C, Matthews RJ, Boal AK, & Krebs C (2015) Mechanisms of 2-oxoglutarate-dependent oxygenases: The hydroxylation paradigm and beyond. *2-oxoglutarate-dependent oxygenases*, RSC Metallobiology, eds Hausinger RP & Schofield CJ (The Royal Society of Chemistry, London), 1 Ed, pp 95-122.
30. Hausinger RP (2004) Fe(II)/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Crit Rev Biochem Mol* 39(1):21-68.
31. Helmetag V, Samel SA, Thomas MG, Marahiel MA, & Essen LO (2009) Structural basis for the erythro-stereospecificity of the L-arginine oxygenase VioC in viomycin biosynthesis. *Febs J* 276(13):3669-3682.
32. Hopkinson RJ, Hamed RB, Rose NR, Claridge TD, & Schofield CJ (2010) Monitoring the activity of 2-oxoglutarate dependent histone demethylases by NMR spectroscopy: Direct observation of formaldehyde. *Chembiochem* 11(4):506-510.
33. Sleeman MC & Schofield CJ (2004) Carboxymethylproline synthase (CarB), an unusual carbon-carbon bond-forming enzyme of the crotonase superfamily involved in carbapenem biosynthesis. *J Biol Chem* 279(8):6730-6736.
34. Grob CA & Baumann W (1955) Die 1,4-eliminierung unter fragmentierung. *Helv Chim Acta* 38(3):594-610.
35. Simaan AJ & Réglér M (2015) 1-aminocyclopropane-1-carboxylic acid oxidase. *2-oxoglutarate-dependent oxygenases*, RSC Metallobiology, eds Hausinger RP & Schofield CJ (Royal Society of Chemistry, London), 1st Ed, pp 425-437.
36. Retey J (1990) Enzymatic-reaction selectivity by negative catalysis or how do enzymes deal with highly reactive intermediates. *Angew Chem Int Ed* 29(4):355-361.
37. Losman JA & Kaelin WG, Jr. (2013) What a difference a hydroxyl makes: Mutant IDH, (R)-2-hydroxyglutarate, and cancer. *Genes Dev* 27(8):836-852.
38. Rocklin AM, Kato K, Liu HW, Que L, & Lipscomb JD (2004) Mechanistic studies of 1-aminocyclopropane-1-carboxylic acid oxidase: Single turnover reaction. *J Biol Inorg Chem* 9(2):171-182.
39. Rutledge P (2015) Isopenicillin N synthase. *2-oxoglutarate-dependent oxygenases*, RSC Metallobiology, eds Hausinger RP & Schofield CJ (Royal Society of Chemistry, London), 1 Ed, pp 414-424.
40. Wanders RJA, Ferdinandusse S, Ebberink MS, & Waterham HR (2015) Phytanoyl-CoA hydroxylase: A 2-oxoglutarate-dependent dioxygenase crucial for fatty acid alpha-oxidation in humans. *2-oxoglutarate-dependent oxygenases*, RSC Metallobiology, eds Hausinger RP & Schofield CJ (Royal Society of Chemistry, London), 1 Ed Vol 2015-January, pp 338-349.
41. Schofield CJ & Ratcliffe PJ (2004) Oxygen sensing by HIF hydroxylases. *Nat Rev Mol Cell Bio* 5(5):343-354.
42. Li WJ, Zhang TL, & Ding JP (2015) Molecular basis for the substrate specificity and catalytic mechanism of thymine-7-hydroxylase in fungi. *Nucleic Acids Res* 43(20):10026-10038.
43. Chandra GR & Spencer M (1963) Ethylene production by subcellular particles from rat liver, rat intestinal mucosa and *Penicillium digitatum*. *Nature* 197(4865):366-367.
44. Case DA, *et al.* (2005) The AMBER biomolecular simulation programs. *J Comput Chem* 26(16):1668-1688.
45. Frisch MJ, *et al.* (2009) Gaussian 09 (Gaussian, Inc., Wallingford, CT, USA).





# 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. *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 : 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, beamsize 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  $\mu$ M 2OG, 500  $\mu$ M ascorbate, 500  $\mu$ M L-Arg and 20  $\mu$ 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> : <i>(SeMet)</i>	<i>PsEFE</i> : <i>Mn</i> : <i>BTP</i>	<i>PsEFE</i> : <i>Mn</i> : <i>2OG</i>	<i>PsEFE</i> : <i>Fe</i> : <i>NOG</i> : <i>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 <sub>merge</sub>	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 <sub>work</sub>		0.1346	0.1726	0.1620
R <sub>free</sub>		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
<B <sub>factor</sub> > (Å <sup>2</sup> ) all atoms		20.0	18.0	21.0

Numbers in parentheses refer to the highest resolution shell.

R<sub>merge</sub> is the unweighted R-value on I between symmetry mates.

R<sub>cryst</sub> =  $\sum hkl | |F_{obs}(hkl)| - k |F_{calc}(hkl)| | / \sum hkl |F_{obs}(hkl)|$  for the working set of reflections; R<sub>free</sub> 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.**

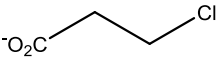
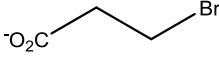
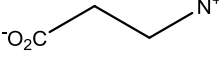
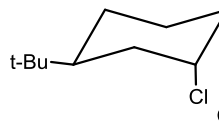
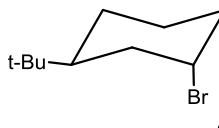
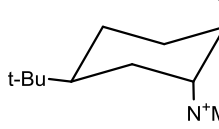
	<i>PsEFE</i> : <i>Mn</i> : <i>BTP</i>	<i>PsEFE</i> : <i>Mn</i> : <i>2OG</i>	<i>PsEFE</i> : <i>Fe</i> : <i>NOG</i> : L-Arg chain <b>A</b>	<i>PsEFE</i> : <i>Fe</i> : <i>NOG</i> : L-Arg chain <b>B</b>	<i>PsEFE</i> : <i>Fe</i> : <i>NOG</i> : L-Arg chain <b>C</b>	<i>PsEFE</i> : <i>Fe</i> : <i>NOG</i> : L-Arg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

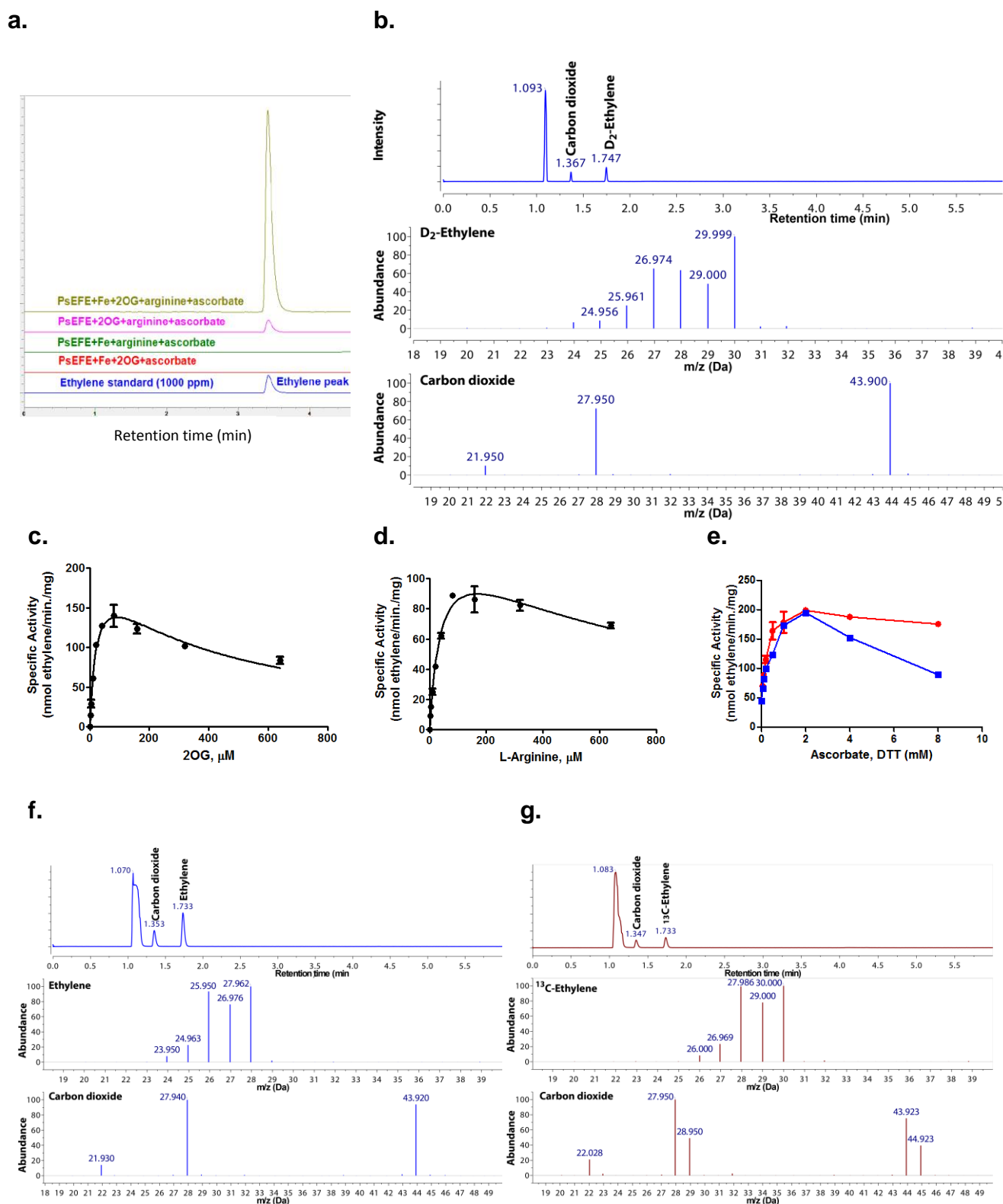
**Supplementary Table S3: Ethylene production by *PsEFE* variants.**

<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

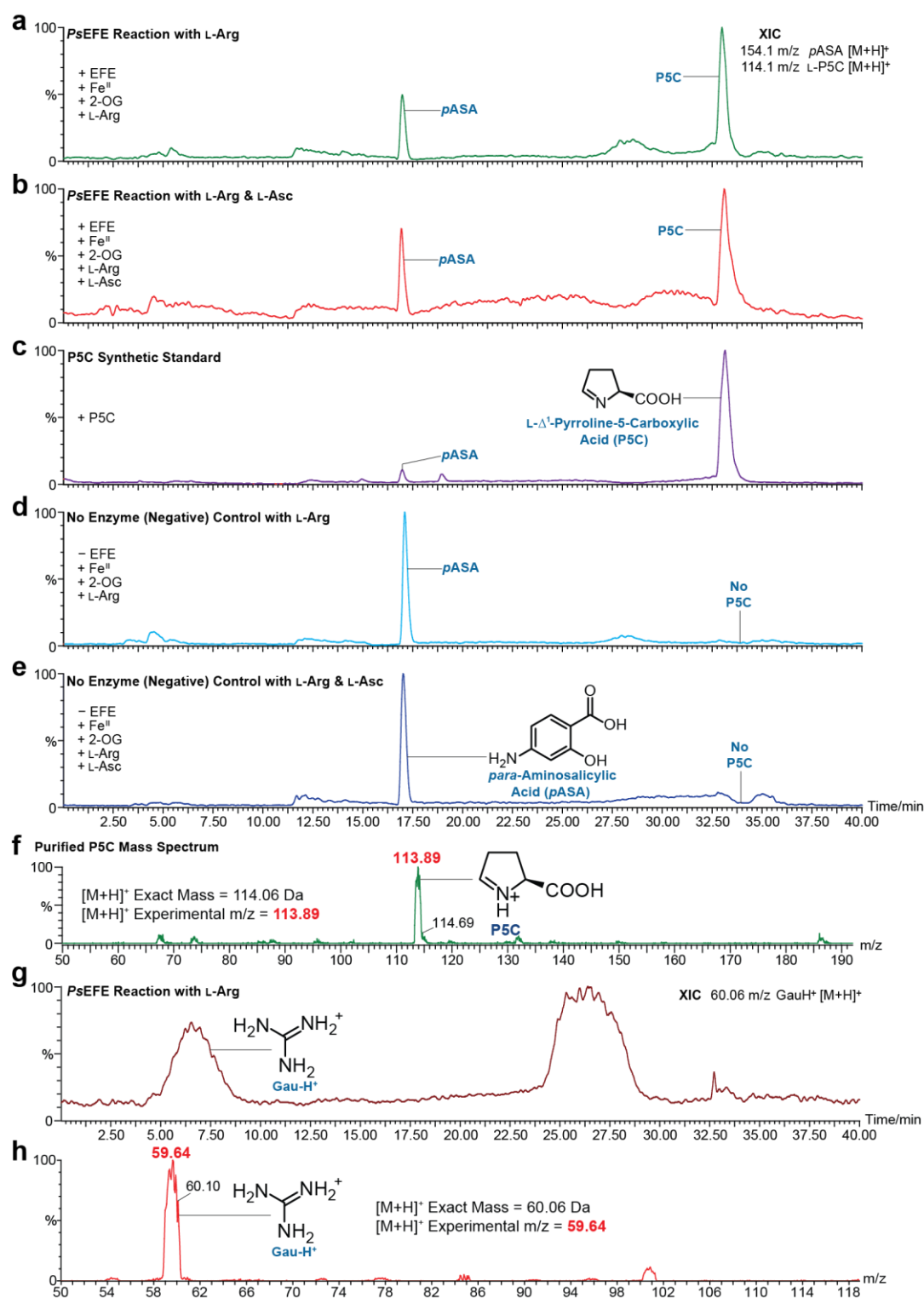
The production of ethylene by wild type *PsEFE* is normalized to 100%. Standard errors of the mean are calculated from three replicates.

**Supplementary Table S4:** Calculated Gibbs free energy in 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  $\phi = 90^\circ > 45^\circ > 0^\circ$ .

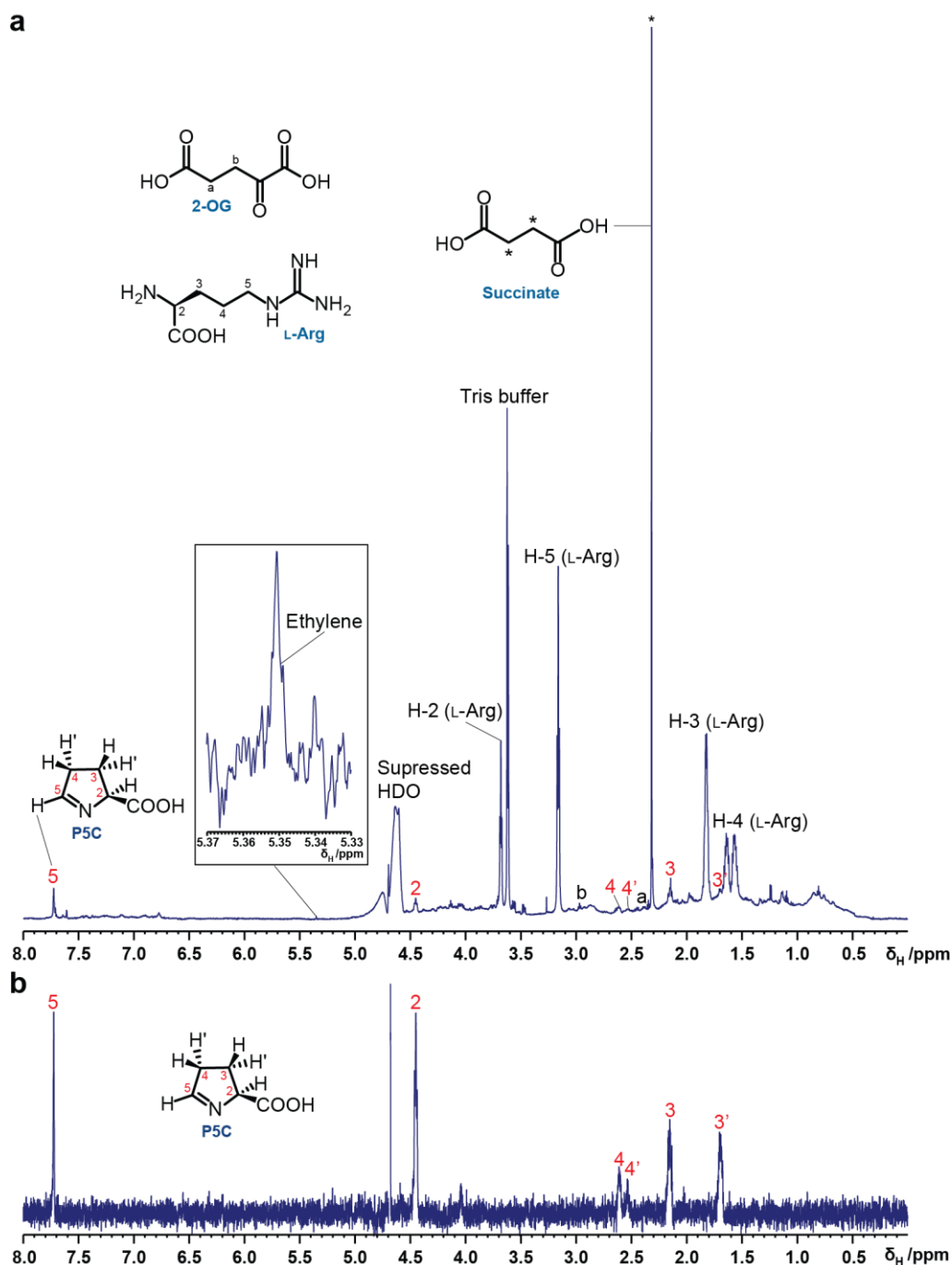
	Reactant	$\phi_{\text{ccco}}=0^\circ$	$\phi_{\text{ccco}}=90^\circ$	$\phi_{\text{ccco}}=45^\circ$
1		15.82	12.27	14.05
2		14.10	10.61	12.24
3		8.01	5.34	6.87
4		12.21	8.12	10.44
5		11.19	7.23	9.48
6		5.35	1.41	3.87



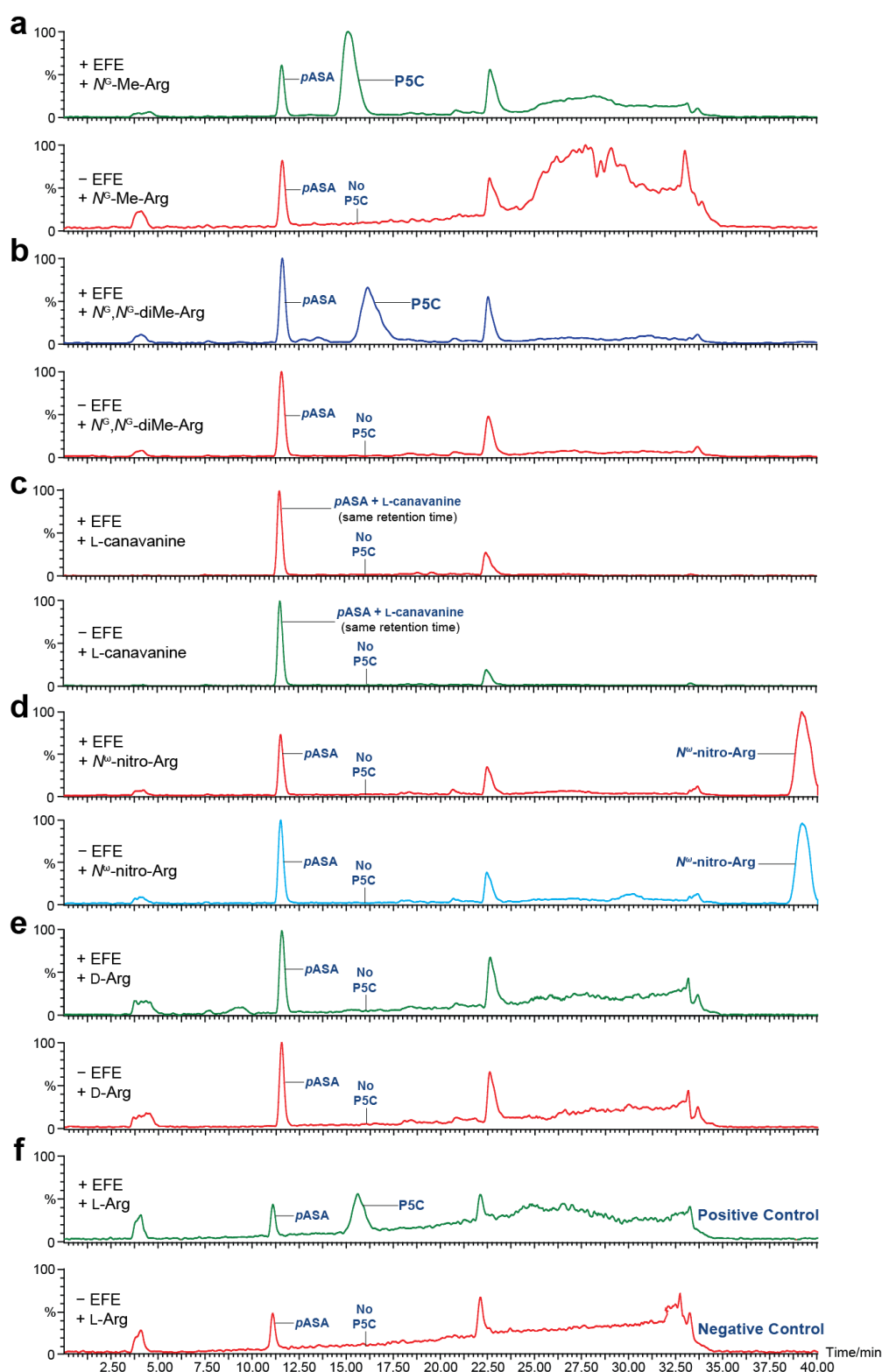
**Supplementary Figure S1: GC-MS analysis of *PsEFE* catalysis.** (a) GC chromatograms of *PsEFE* activity under the indicated conditions. (b) GC-MS analysis of ethylene production from [3,3- $^2\text{H}$ ]-labelled 2OG; the ethylene peak is at 1.75 min (top). MS spectra of  $^2\text{H}_2$ -ethylene (partially  $^2\text{H}$ -labelled) (middle) and  $\text{CO}_2$  peak is at 43.9 (bottom). The peak at 27.95 is CO 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;  $V_{\max}$  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 or DTT 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.



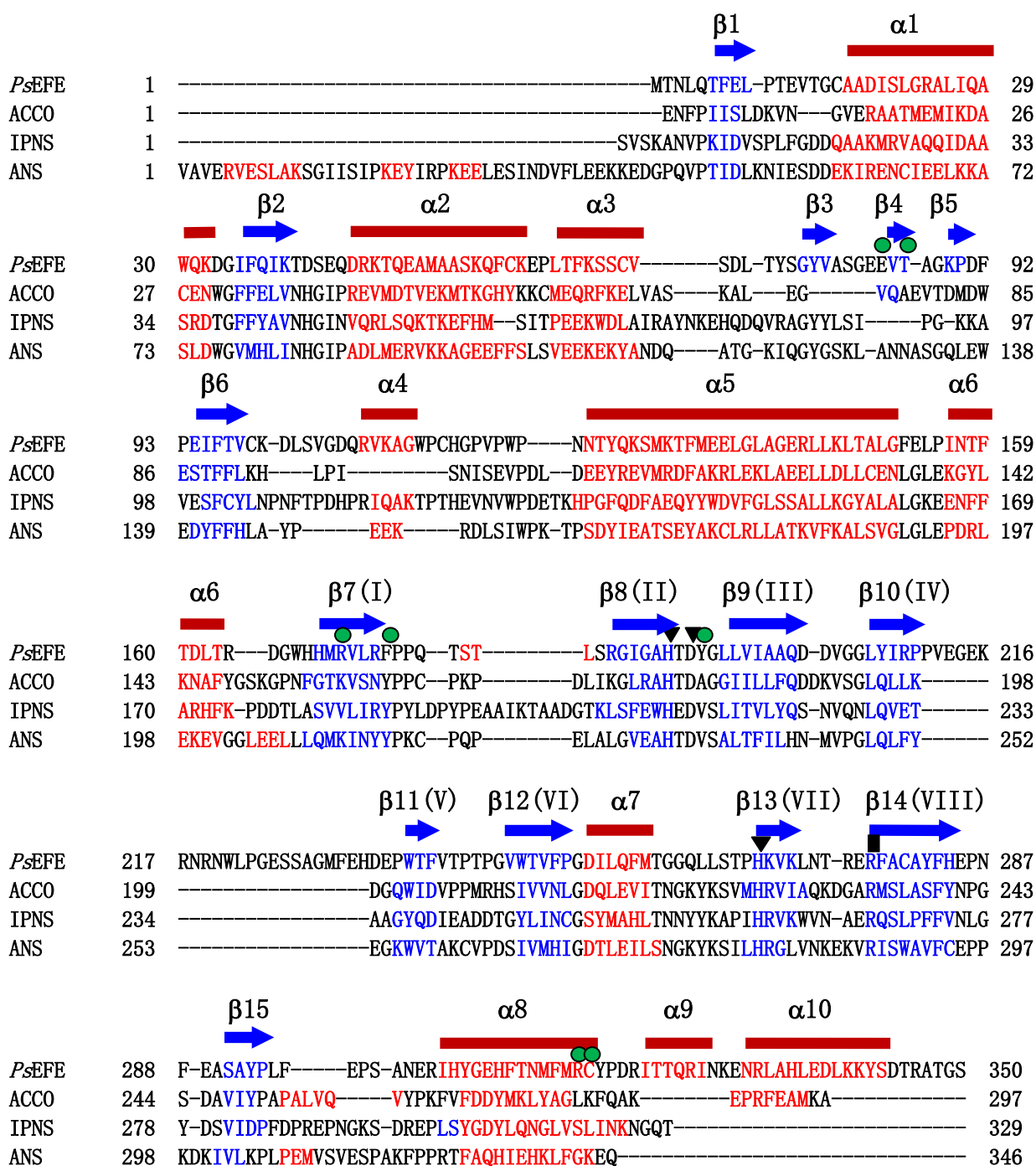
**Supplementary Figure 2: LC/MS detection of pyrroline-5-carboxylate (P5C).** PsEFE reactions (100  $\mu$ L) were at 25  $^{\circ}$ C for 14 h in 50 mM Tris-HCl, pH 7.5. Reactions were quenched with 100  $\mu$ L methanol (1:1) containing *para*-aminosalicylic acid (pASA) standard, then microcentrifuged. 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.

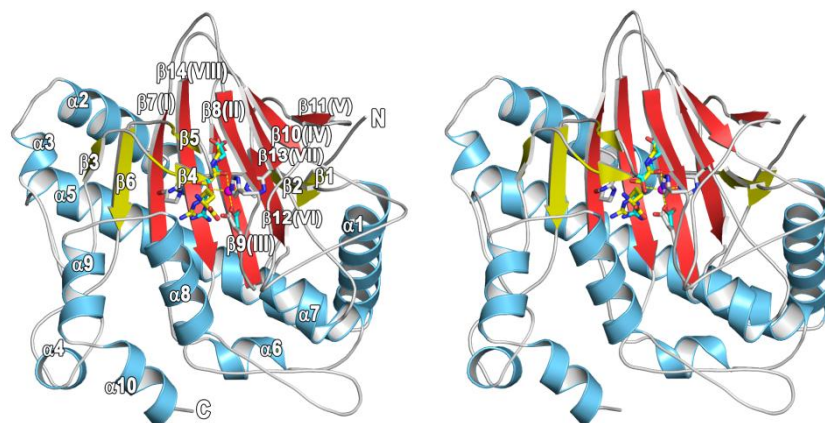


**Supplementary Figure S4: Arginine analogues tested with PsEFE.** Reactions (100  $\mu$ L) were performed at room temperature overnight in 10 mM HEPES-NaOH, pH 7.5; conditions were 15  $\mu$ M PsEFE, 100  $\mu$ M  $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$ , 1 mM 2OG, 1 mM substrate. Samples were analysed 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)**  $N^3$ -methyl-L-arginine ( $N^3$ -Me-Arg); **(b)** (asymmetric)  $N^3,N^6$ -dimethyl-L-arginine ( $N^3,N^6$ -diMe-Arg); **(c)** L-canavanine ( $\text{SO}_4$ ); **(d)**  $N^{\omega}$ -nitro-L-arginine ( $N^{\omega}$ -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,  $N^{\omega}$ -nitro-L-arginine, 2-keto-4-methyl-thiobutyric acid, guanidine with L-alanine, L-( $\alpha$ -amino)adipic acid, diaminopimelic acid,  $N^{\omega}$ -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  $N^3$ -methyl-L-arginine (52%  $\pm$  7%), (asymmetric)  $N^3,N^6$ -dimethyl-L-arginine (42%  $\pm$  10%),  $N^{\omega}$ -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.

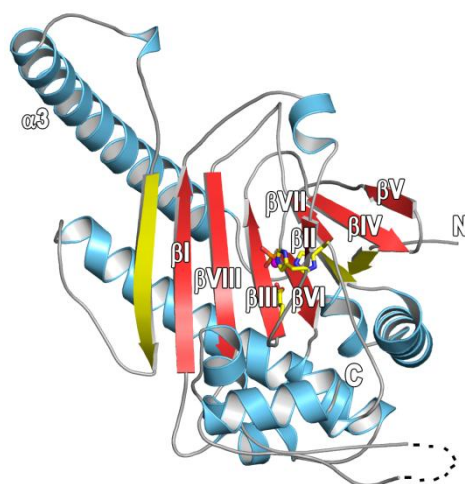


**Supplementary Figure S5: Secondary structure based sequence alignment of PsEFE with ACC oxidase (ACCO) from *Petunia hybrida* (PDB 1WA6), isopenicillin N synthase (IPNS) from *Aspergillus nidulans* (PDB 1IPS), and anthocyanidin synthase (ANS) from *Arabidopsis thaliana* (PDB 1GP5).** Residues in blue 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.

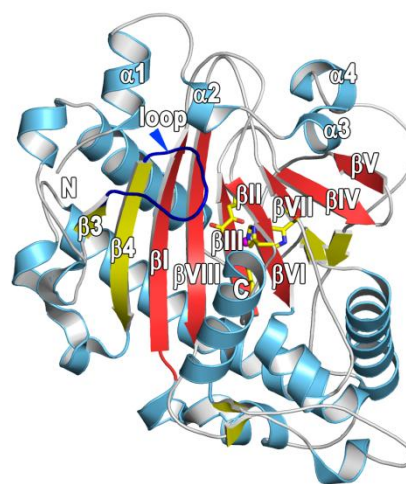
**a**



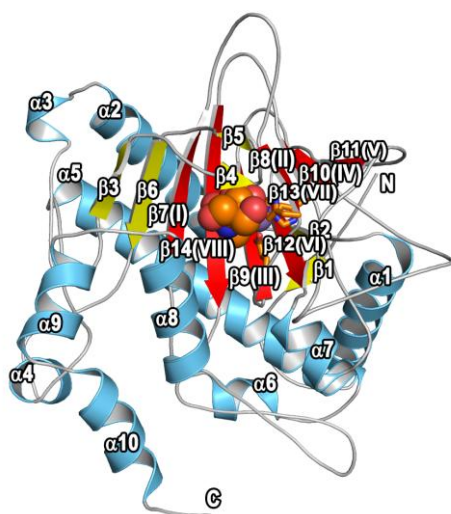
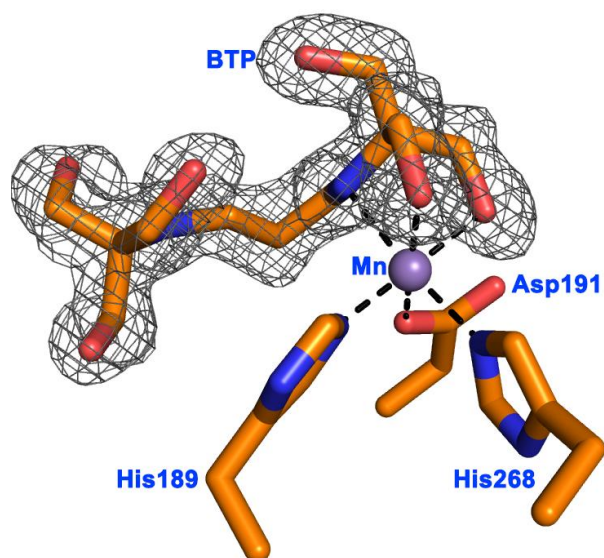
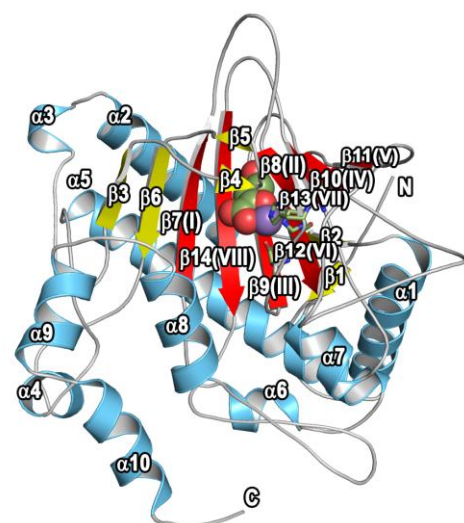
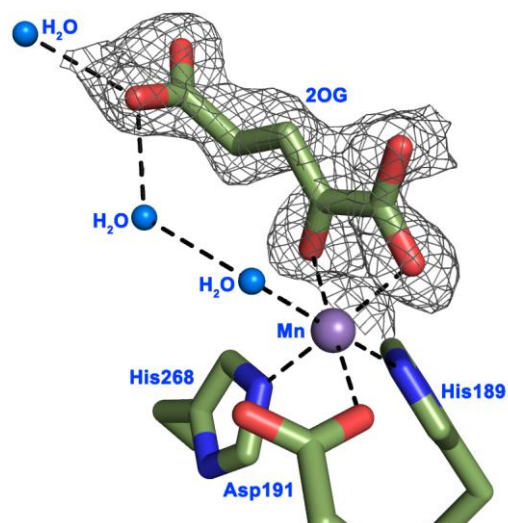
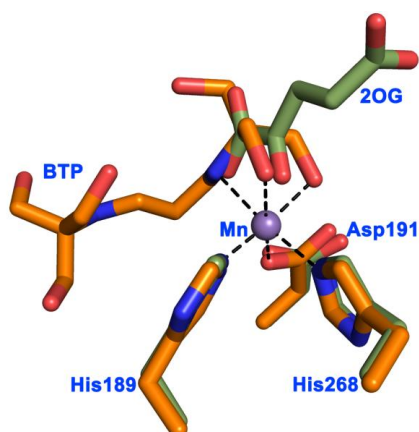
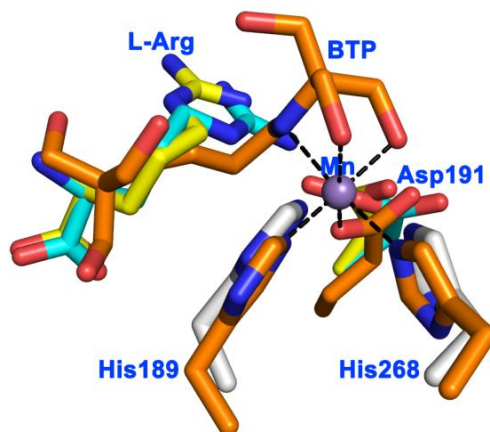
**b**

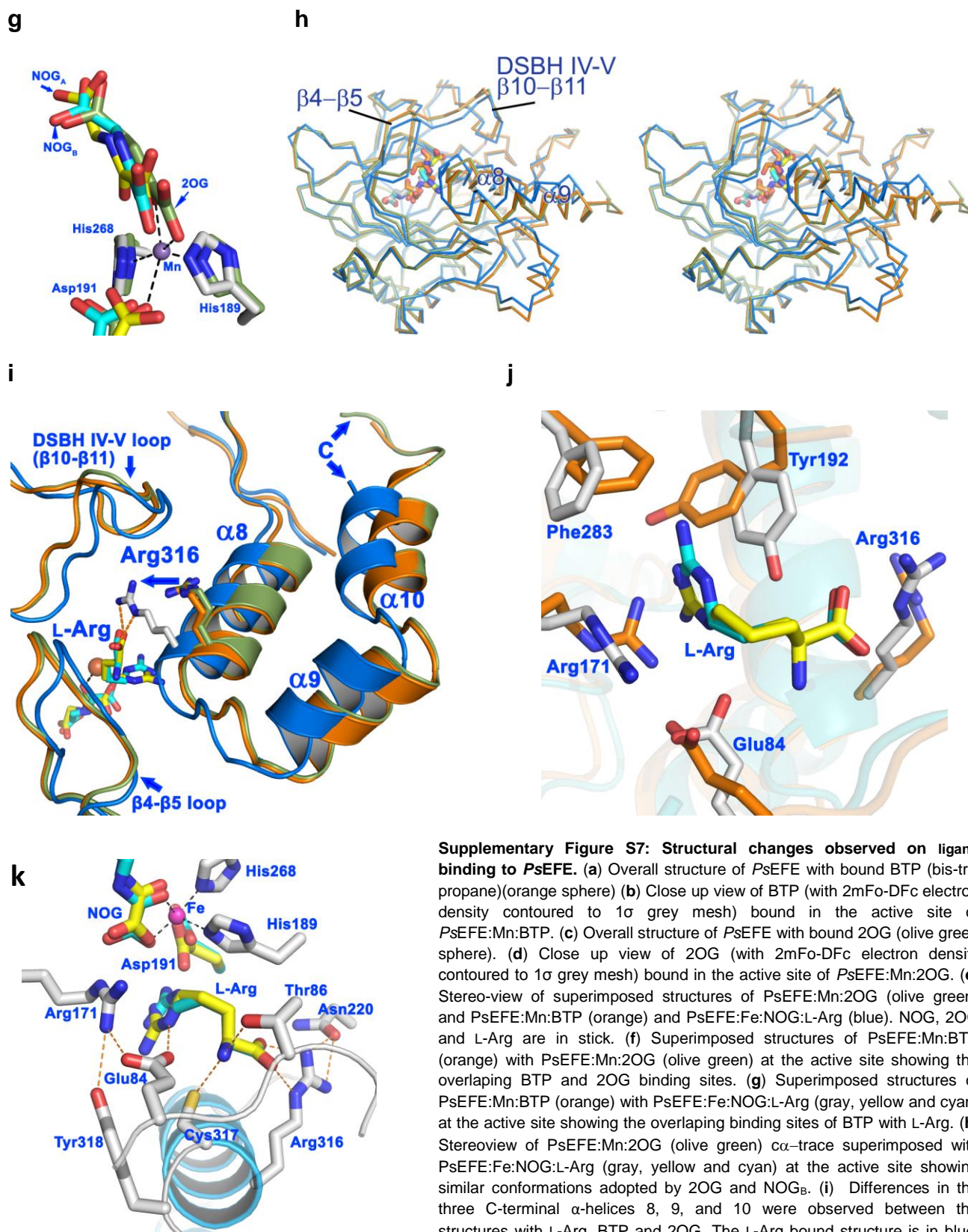


**c**

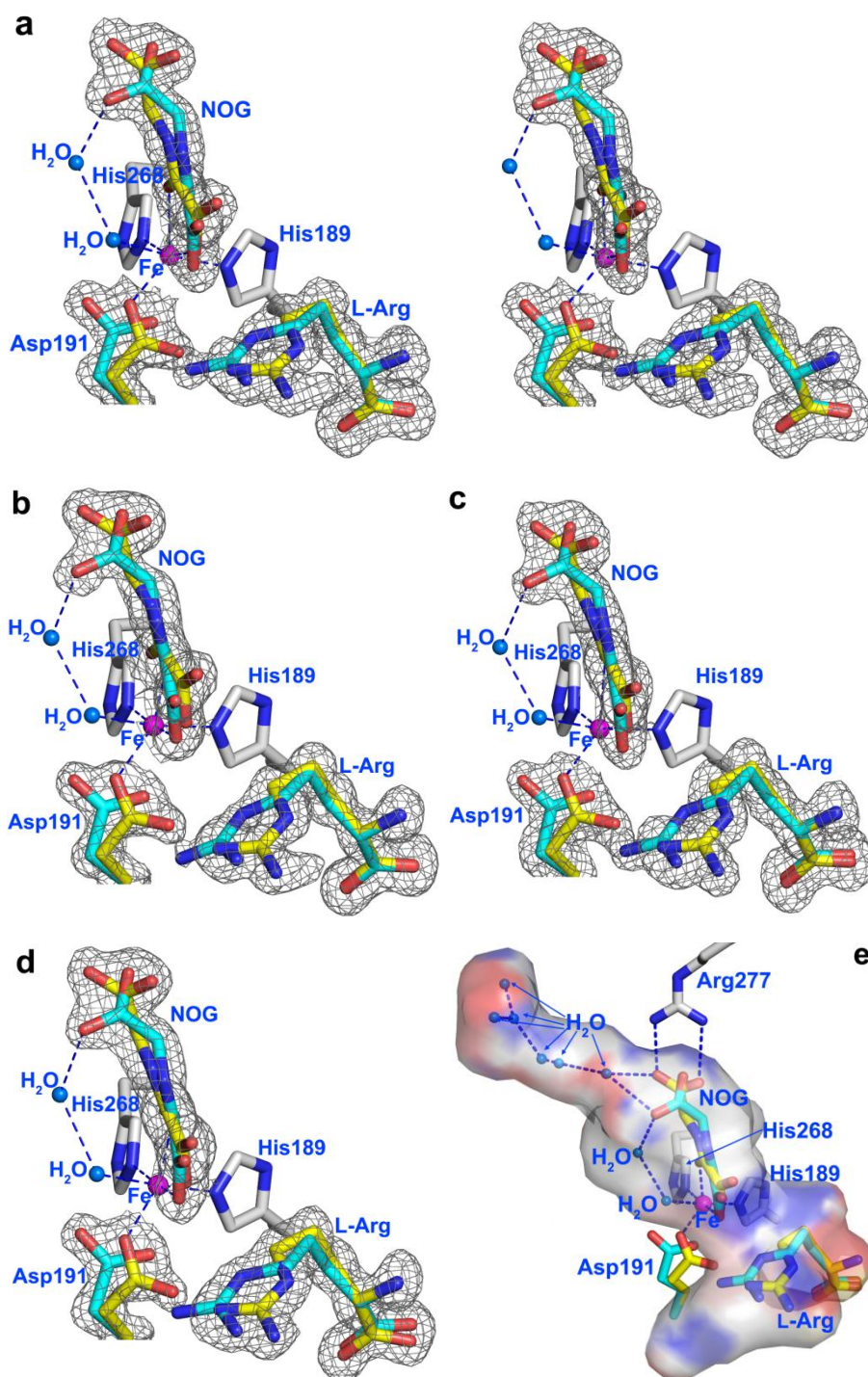


**Supplementary Figure S6: The overall fold of (a) PsEFE (stereoview) compared to (b) 1-aminocyclopropane-1-carboxylate oxidase (ACCO) and (c) anthocyanidin synthase (ANS). (a) Stereoview of the PsEFE fold. The eight  $\beta$ -strands of the double stranded  $\alpha$ -helix (DSBH) core are in red and 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 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.**

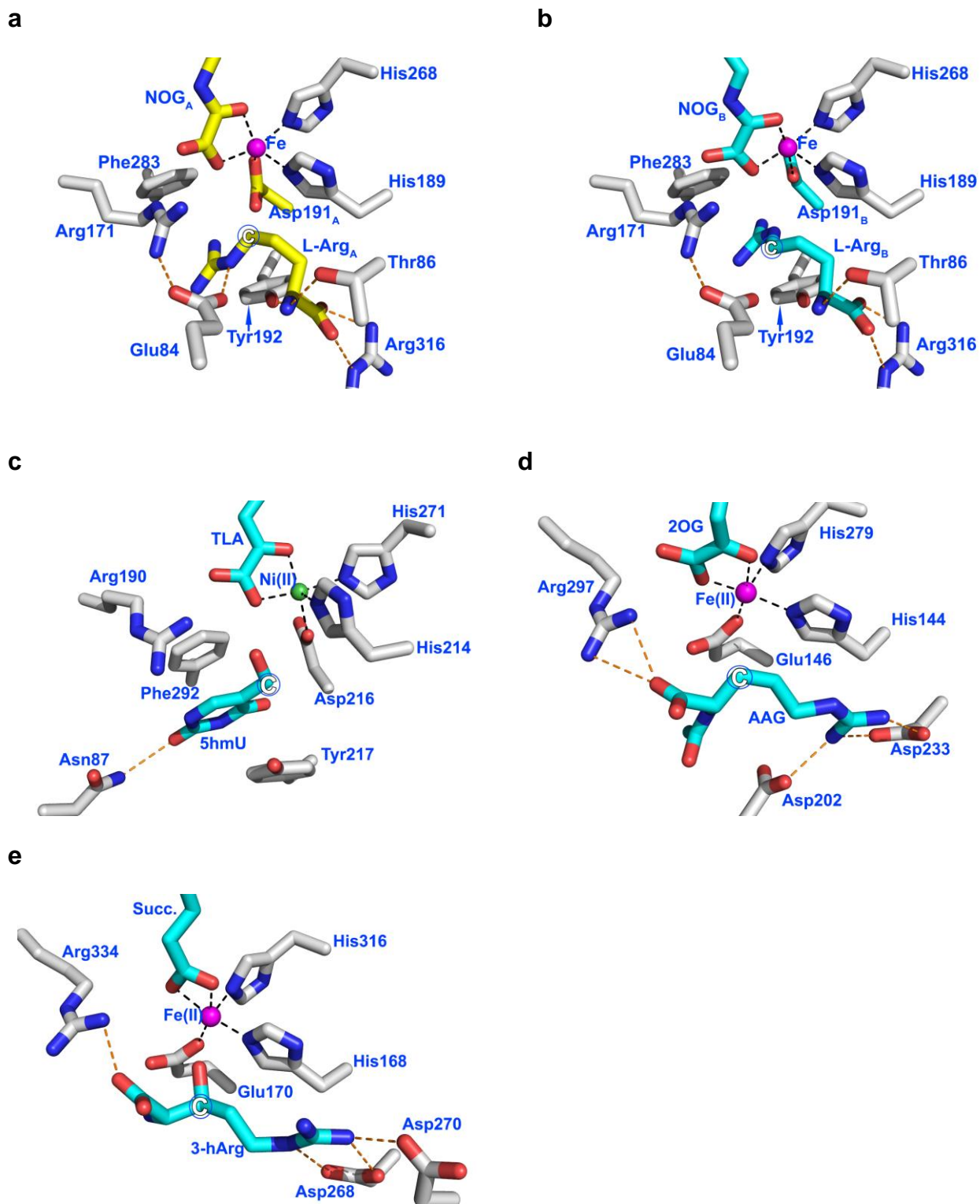
**a****b****c****d****e****f**



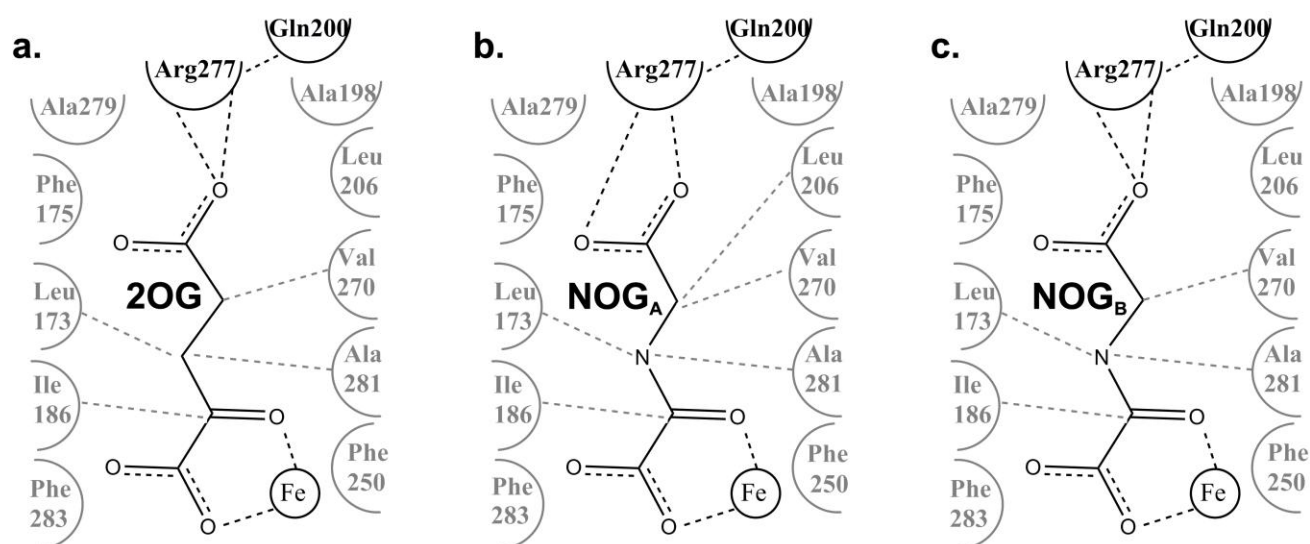
**Supplementary Figure S7: Structural changes observed on ligand binding to PsEFE.** (a) Overall structure of PsEFE with bound BTP (bis-tris propane)(orange sphere) (b) Close up view of BTP (with 2mFo-DFc electron density contoured to 1 $\sigma$  grey mesh) bound in the active site of PsEFE:Mn:BTP. (c) 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) Stereo-view of superimposed structures of PsEFE:Mn:2OG (olive green) and PsEFE:Mn:BTP (orange) and PsEFE:Fe:NOG:L-Arg (blue). NOG, 2OG and L-Arg are in stick. (f) 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. (g) 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. (h) Stereoview of PsEFE:Mn:2OG (olive green)  $\alpha$ -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.



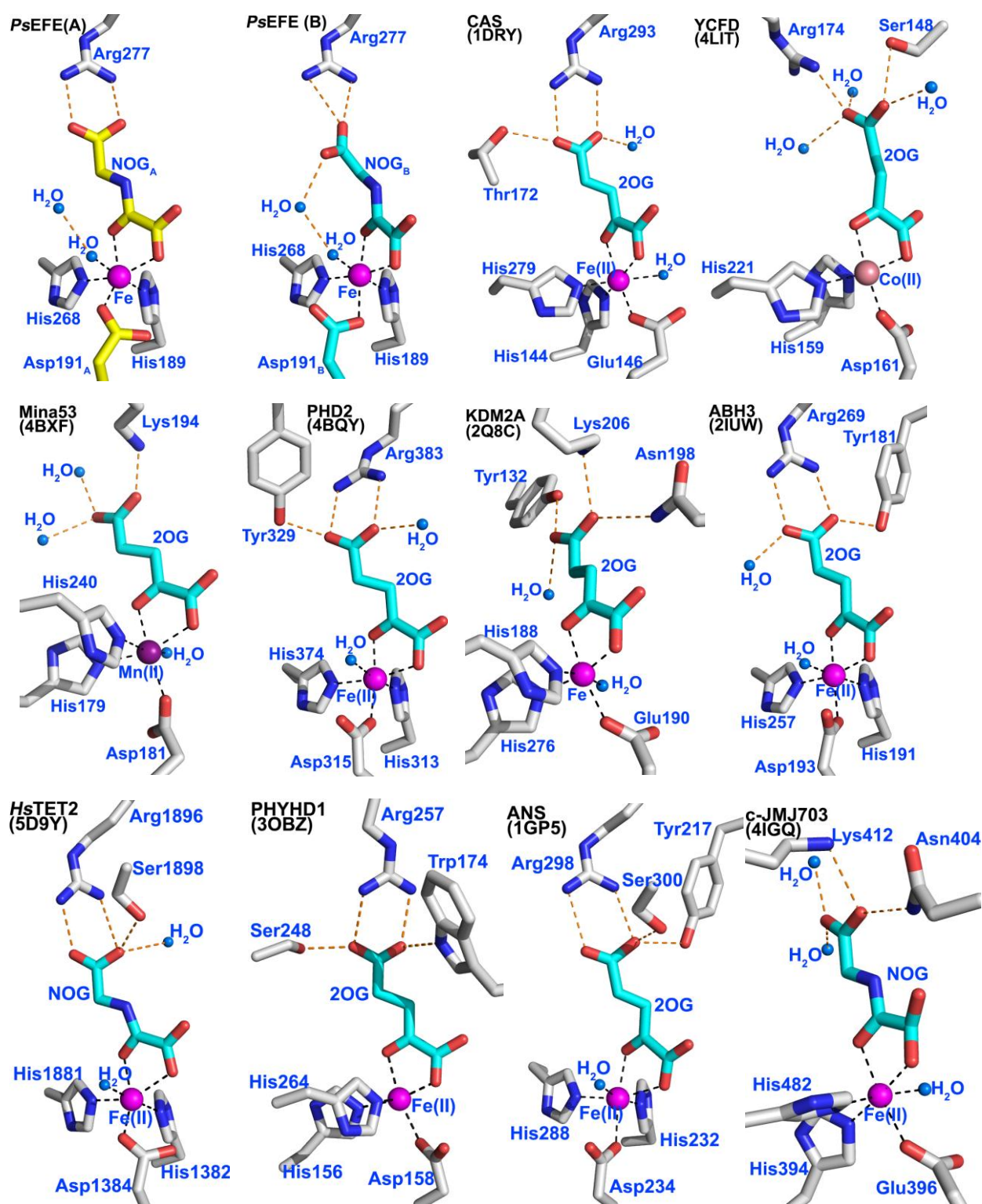
**Supplementary Figure S8: Views of the  $2mF_o-DF_c$  of *PsEFE.Fe.NOG.L-Arg* complex structure electron density maps for NOG and L-arginine of the four protein chains in the asymmetric unit.** The views reveal conservation of the two observed binding modes (Conformations A (yellow) and B (cyan) for 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).



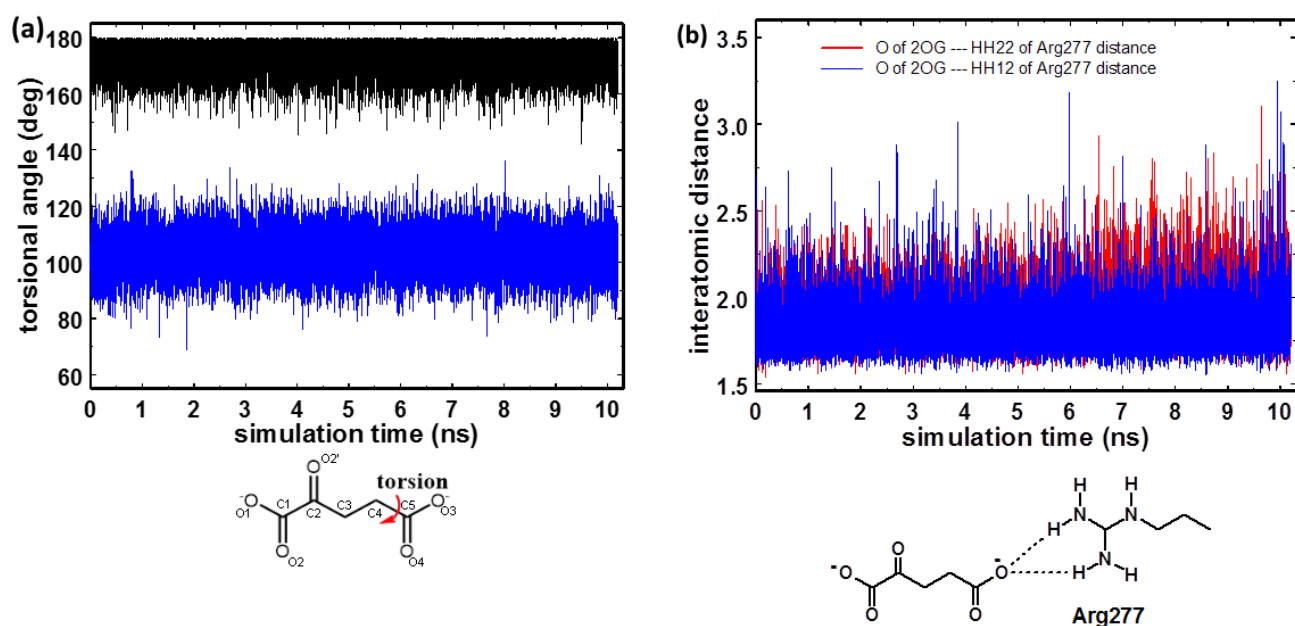
**Supplementary Figure S9: Comparison of the substrate binding mode of PsEFE with those of other 2OG oxygenases.** (a, b) Conformations A and B observed in the PsEFE.Fe.NOGL-Arg complex. (c) T7H (human TET homologue) from *Neurospora crassa* in complex with Ni(II) and 5-hydroxymethyl uracil (5hmU) (PDB 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 (2S,3S)-3-hydroxyarginine (3-hArg)(PDB ID 2WBO):The hydroxylated carbon is labelled with a circled white "C".



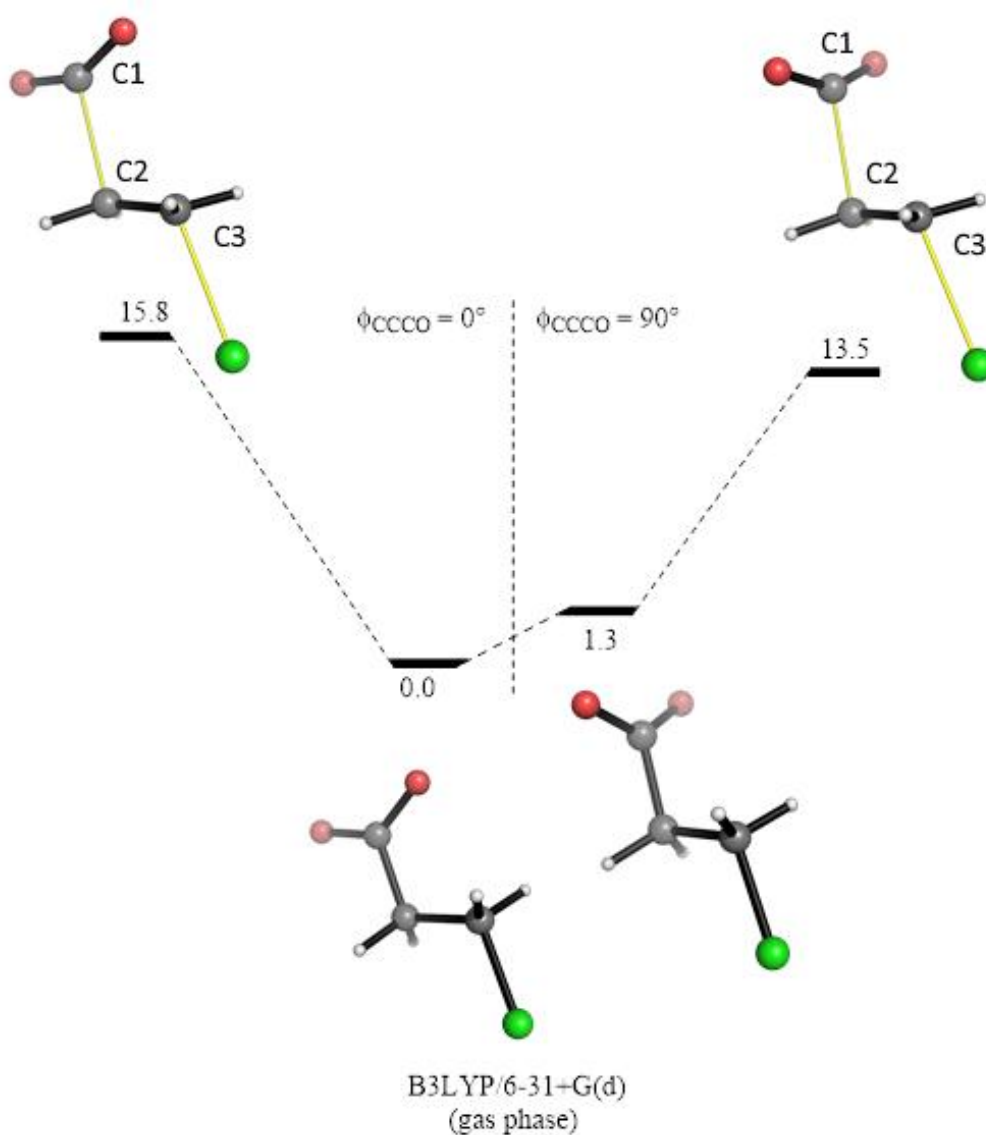
**Supplementary Figure S10: Interactions in the hydrophobic binding pocket for 2OG in *PsEFE*.** Residues within 6 Å of 2OG/NOG are depicted as semi-circles. **(a)** 2OG, **(b)** NOG<sub>A</sub> and **(c)** NOG<sub>B</sub>. Predicted hydrogen bonds and metal chelation are depicted in bold dashed lines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



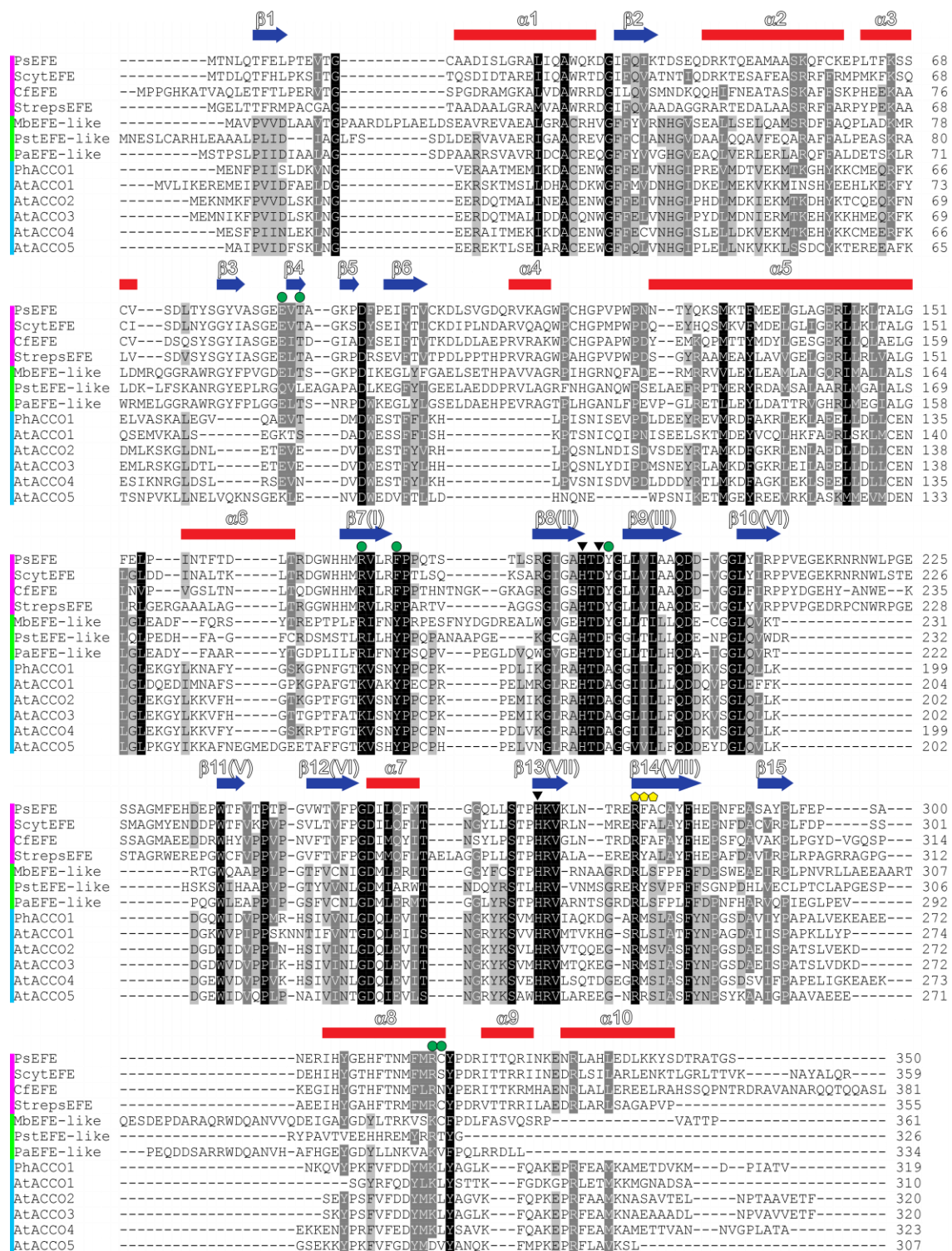
**Supplementary Figure S11: Views of 2-oxoglutarate (2OG) binding modes from selected 2OG oxygenase structures.** PsEFE(A) and PsEFE(B): Conformations A and B of *N*-oxalylglycine (NOG), L-Arg and Asp191; CAS: clavaminic synthase from *Streptomyces clavuligerus* (PDB ID 1DRY); YCFD: 50S ribosomal protein L16 arginine hydroxylase from *E. coli* (PDB ID 4LIT); Mina53: bifunctional lysine-specific demethylase and histidyl-hydroxylase Mina53 from *Homo sapiens* (PDB ID 4BFX); 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 plane of the C-5 carboxylate and the plane formed by C-3/C-4/C-5 of 2OG/NOG is similarly low.

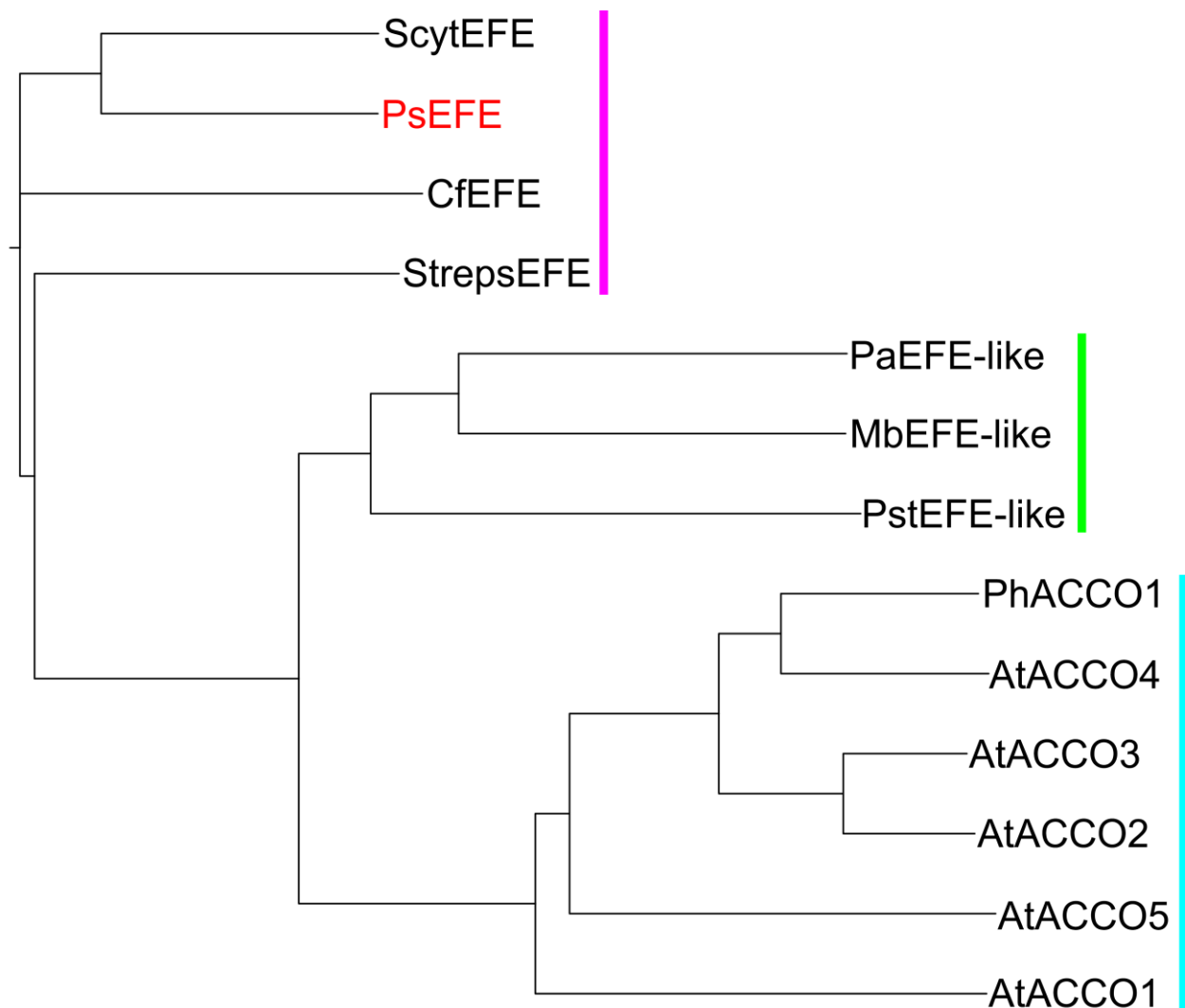


**Supplementary Figure S12: MD simulation studies on the conformation of 2OG in complex with PsEFE.** A PsEFE:Fe:2OG:L-Arg model was calculated based on the high resolution PsEFE:Fe:NOG:L-Arg structure (a) Time-course evolutions of the torsion angle between the O3-C5-O4 plane and the C3-C4-C5 plane of 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 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 fiorinae* 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, Q0WFW4). 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.