Biochemical and Crystallographic Investigations into Isonitrile Formation by a Non-Heme Iron-Dependent Oxidase/Decarboxylase

Rohan Jonnalagadda[†§], Antonio Del Rio Flores[†‡], Wenlong Cai[‡], Rimsha Mehmood[&,#], Maanasa Narayanamoorthy[♦], Chaoxiang Ren[‡], Jan Paulo T. Zaragoza[♦^], Heather J. Kulik[#], Wenjun Zhang[‡u*], and Catherine L. Drennan[§,&,+*]

[§]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 United States

[+&]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 United States

[#]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 United States

[+]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 United States

[‡]Department of Chemical and Biomolecular Engineering, University of California Berkeley, Berkeley, California 94720 United States

[♦]Department of Chemistry, University of California Berkeley, Berkeley, California 94720 United States

[^]California Institute for Quantitative Biosciences, Berkeley, California 94720 United States

[u]Chan Zuckerberg Biohub, San Francisco, California, 94158, United States

[†]Co-first authors

*For correspondence: cdrennan@mit.edu (CLD); wjzhang@berkeley.edu (WZ)

Abstract

The isonitrile moiety is found in marine sponges and some microbes, where it plays a role in processes such as virulence and metal acquisition. Until recently only one route was known for isonitrile biosynthesis, a condensation reaction that brings together a nitrogen atom of L-Trp/L-Tyr with a carbon atom from ribulose-5-phosphate. With the discovery of ScoE, a mononuclear Fe(II) α-ketoglutarate dependent dioxygenase from *Streptomyces coeruleorubidus*, a second route was identified. ScoE forms isonitrile from a glycine adduct, with both the nitrogen and carbon atoms coming from the same glycyl moiety. This reaction is part of the nonribosomal biosynthetic pathway of isonitrile lipopeptides. Here, we present structural, biochemical and computational investigations of the mechanism of isonitrile formation by ScoE, an unprecedented reaction in the mononuclear Fe(II) α-ketoglutarate dependent dioxygenase superfamily. The full stoichiometry of this enzymatic reaction was measured, showing that for every glycyl substrate consumed, 1 molecule of isonitrile product is generated, and 2 molecules each of α-ketoglutarate and molecular oxygen are consumed. We also present multiple high-resolution (1.45-1.96 Å resolution) crystal structures of Fe(II)-bound ScoE, both with and without the glycyl-adduct containing substrate, (*R*)-3-((carboxylmethyl)amino)butanoic acid (CABA), and both with and without co-substrate α-ketoglutarate. We further present a structure of ScoE bound to Fe(IV)=O mimic oxovanadium in the presence of substrate CABA to 2.10 Å resolution. We find α-ketoglutarate can bind in an unexpected conformation and potentially low affinity state that impinges on the substrate binding site. Comparison to previously published crystal structure of ScoE indicates that a substantial conformational change may be necessary to form a higher affinity α-ketoglutarate binding site. Together, these structures with accompanying data from site-directed mutagenesis and computation, provide insight into the mechanism of isonitrile formation and how the structure of ScoE has been adapted to perform this unusual chemical reaction.

Introduction

The isonitrile, an electron-rich functional group, is a hallmark of a variety of natural products such as xanthocillin and rhabduscin, and plays a role in diverse processes such as metal acquisition and virulence (**Fig. S1**)[1]. The biosynthesis of the isonitrile group was thought to be restricted to the isonitrile synthase (IsnA) enzyme family. These enzymes catalyze the formation of isonitriles using the nitrogen atom of an α-amino group of L-Trp/L-Tyr and a carbon atom from ribulose-5-phosphate (**Fig. S1**)[2]. Recent studies of a conserved gene cluster in some Actinobacteria species, including *M. tuberculosis*, revealed a novel biosynthetic pathway for isonitrile synthesis as part of the formation of nonribosomally synthesized isonitrile lipopeptide (INLP). Furthermore, a study of the gene cluster from *Streptomyces coeruleorubidus* showed that isonitrile biosynthesis is catalyzed by the enzyme ScoE[3]. Notably, unlike the IsnA enzymes, ScoE forms an isonitrile from a single substrate: a β-glycine adduct attached to a short fatty acyl chain, as confirmed by isotope labeling[3]. ScoE therefore represents a novel enzymatic mechanism of isonitrile biosynthesis[3].

Bioinformatic analysis of the INLP-producing gene clusters suggested that the *scoE* gene encodes a non-heme iron(II) α-ketoglutarate (α-KG) dependent dioxygenase (Fe(II)/αKG-dioxygenase)[3]. This prediction was confirmed when it was shown that both iron(II) and α-KG were required by ScoE *in vitro* for isonitrile formation on an untethered substrate: (R)-3-carboxyaminobutanoic acid (CABA) to form an untethered product: (R)-3-isocyano butanoic acid (INBA) (**Fig. 1**)[4]. This finding firmly places ScoE in the Fe(II)/αKG-dioxygenase superfamily, a family of enzymes that uses a mononuclear Fe(II) cofactor, α-KG, and molecular oxygen to catalyze a diverse set of reactions, such as hydroxylation, halogenation, carbon-carbon bond desaturation, and ring contraction[5]. Notably, isonitrile formation is a new activity in this enzyme family.

Independent of the particular reaction, all Fe(II)/αKG-dioxygenases are thought to share a common mechanism of activation (**Fig. S2**)[5]. This activation chemistry occurs at the mononuclear Fe(II) cofactor, which is coordinated by a conserved 2 histidine-1 carboxylate facial triad motif (2His-1Asp for ScoE) that acts as three ligands to the metal. The co-substrate α-KG binds the metallocofactor in a bidentate fashion, leaving

one open coordination site, usually found in close proximity to the substrate binding site[6]. This final coordination site is the binding site of molecular oxygen, at which point chemistry begins. Specifically, molecular oxygen is cleaved, with one oxygen atom inserting into α-KG to form succinate and carbon dioxide, and the other atom remaining bound to the metallocofactor to form a highly reactive Fe(IV)-oxo species[7-9] (**Fig. S2**). This reaction step is thought to occur primarily in the presence of substrate, but it has been observed in the absence of substrate as well, generally at a lower rate (uncoupled reaction)[10]. The Fe(IV)-oxo species performs a hydrogen atom abstraction at the substrate to form a Fe(III)-OH species and a substrate radical (**Fig. S2**)[11]. The resolution of this substrate radical differs between enzymes of this family. The best understood reaction of this enzyme family is hydroxylation, in which the Fe(III)-OH bond is homolytically cleaved such that the hydroxyl group "rebounds" to the substrate radical to yield a hydroxylated product and restore the Fe(II) cofactor (**Fig. S3**)[12]. The hydroxylase TauD is the most studied Fe(II)/αKG-dioxygenase, and is considered to be the archetypal enzyme of the family[5].

Inspection of the reaction catalyzed by ScoE (**Fig. 1**) has led to the proposal that like TauD, ScoE catalyzes hydroxylation reactions as part of the four-electron oxidation of CABA (**Fig. 2**)[4, 13]. The site(s) of the hydroxylation(s) are unknown, but the alpha-carbon of the glycyl moiety has been proposed to be both singly and doubly hydroxylated, and the nitrogen of the glycyl moiety has also been proposed to be a site of hydroxylation (**Fig. 2**). Chang and colleagues proposed a C5-hydroxylated CABA compound as an intermediate in the reaction from ScoE's homologue, SfaA, from *Streptomyces thioluteus* based on a ^{13}C-NMR spectrum obtained by using a 5-^{13}C-CABA substrate through C-H coupling and decoupling experiments[13]. However, this putative intermediate was not chemically synthesized and tested *in vitro* to determine whether it can be directly converted to INBA by ScoE. Regardless of the site(s) of hydroxylation, these mechanistic proposals would require 2 equivalents of α-KG and 2 equivalents of molecular oxygen. Chang and colleagues found that 4.5 equivalents of succinate were produced for every equivalent of isonitrile product generated by ScoE and suggested that this unexpectedly high succinate to isonitrile ratio could be due to

uncoupling of α-KG cleavage from INBA production[13]. Here we investigate the stoichiometry of this reaction and confirm that one CABA is consumed and one INBA is generated per two equivalents each of α-KG and molecular oxygen. Two equivalents of succinate are generated.

Structures of ScoE have also been determined. The first crystal structure of ScoE revealed the archetypal fold, showing that it is consistent with this family of enzymes, containing a double stranded β-helix fold (also known as the jelly roll fold)[4]. The structure also showed the expected 2-His-1-carboxylate motif for Fe(II) binding[4]. However, this structure lacked physiological ligands and the physiological metal, with Zn(II), choline and acetate observed instead of Fe(II), CABA and α-KG, respectively[4]. Another study reported a crystal structure of ScoE bound to Fe(II), CABA, and α-KG analog tartrate bound at 2.18 Å resolution[13]. Here we present a higher resolution structure of ScoE with Fe(II) and CABA bound at 1.7 Å resolution and evaluate through mutagenesis its binding interactions with the protein. We go on to solve crystal structures of ScoE with α-KG bound in an unexpected conformation that impinges on the CABA binding site. We use structural comparisons between the ScoE structures to identify a conformational change that repositions the residues Arg157 and His299 by approximately 10 Å in the active site, creating what appears to be an inducible binding site for α-KG that does not impinge on the CABA binding site. Together with a crystal structure of ScoE bound with oxovanadium, a stable mimic of the ferryl-oxo intermediate, we consider how these structural, biochemical, and computational data inform our understanding of this unprecedented reaction for this enzyme superfamily.

Results

The complete stoichiometry of the ScoE reaction was determined

Metal-free ScoE of *S. coeruleorubidus* was purified heterologously from *E. coli*, with the Fe(II) cofactor reconstituted *in vitro,* and used to determine reaction stoichiometry. Product INBA formation was indirectly detected and quantified using 3,6-Di-2-pyridyl-1,2,4,5-tetrazine (Py-tetrazine), which reacts with INBA to generate 3,5-

di(pyridine-2-yl)-1H-pyrazol-4-amine (Py-aminopyrazole)[14]. No free INBA was detected after the tetrazine click reaction (data not shown), indicating complete conversion of INBA to Py-aminopyrazole. Oxygen consumption was quantified using an Oxygraph Hansatech probe at room temperature (**Fig. S4**). Enzymatic reactions were incubated until the oxygen signal plateaued, indicating the completion of the reaction. Part of the assay mixture was subsequently quenched with cold MeOH and the other part was quenched with cold MeOH supplemented with Py-tetrazine to obtain the enzymatic stoichiometry using liquid chromatography-high resolution mass spectrometry (LC-HRMS). Standards containing varying concentrations of α-KG, succinate, CABA, and Py-aminopyrazole, dissolved in 50 mM HEPES pH 8.0, were prepared to quantify their relative concentrations in the enzymatic assays (**Fig. S5**).

The results from these stoichiometry quantification experiments show the expected 1:1:1 correlation between α-KG consumption, oxygen consumption and succinate formation (**Table 1**). CABA consumption and INBA production are also 1:1 as expected. The stoichiometry between α-KG consumed/succinate produced and CABA consumed/INBA produced is 2:1 (summarized in **Table 1**). A previous study reported a 4.5:1 stoichiometry between succinate and INBA, which as the authors' suggested was artifactually high in succinate, most likely due to unproductive cleavage of α-KG[13]. However, in the absence of other measurements of products formed and reactants consumed, the degree of α-KG reaction uncoupling was unknown, and it was impossible to determine whether 4.5:1 stoichiometry was actually 2:1 or 1:1. In this study, the complete quantification of reactants and products involved in the ScoE enzymatic reaction has been determined, showing that two equivalents of α-KG are necessary for INBA production, indicating that two α-KG-dependent half-reactions must be involved in isonitrile biosynthesis.

CO_2 is a reaction product of ScoE reaction with CABA

The carbon and two oxygen atoms of the carboxylate group of CABA are lost in the formation of INBA (**Fig. 1**), but no experimental evidence has been previously reported to establish whether this group is lost as a result of decarboxylation,

decarbonylation or deformylation[4, 13]. To address this question, [6-^{13}C]-CABA was chemically synthesized and used as a substrate for an *in vitro* ScoE assay, where the presence of [^{13}C]-CO_2 or [^{13}C]-CO was monitored using gas chromatography-mass spectrometry (GC-MS). The release of [^{13}C]-CO_2 as a product of the ScoE catalyzed INBA formation was confirmed (**Fig. S6**). In contrast, [^{13}C]-CO was not detected in the GC-MS experiments nor was CO detected spectroscopically using reduced myoglobin, which exhibits a Soret shift when CO binds to the myoglobin heme (**Fig. S7**)[15]. To test possible deformylation of CABA, two independent analytical assays were performed and both experiments yielded a negative result, failing to detect formate in our ScoE reaction. Specifically, a Sigma-Aldrich Formate Assay kit failed to detect formate; and no formate-adduct was detected through LC-HRMS following a click reaction with 2-nitrophenylhydrazine (**Fig. S8**)[15].

H_2O_2 is not produced by ScoE, indicating that water is the likely product of O_2 reduction

We additionally investigated whether O_2 is reduced to hydrogen peroxide (H_2O_2) as a result of the ScoE catalyzed reaction. Two independent analytical assays were used to test for the presence of H_2O_2. First, a commercial Amplex Red H_2O_2 Detection kit was used. Second, production of H_2O_2 by ScoE was probed through a coupled assay with the enzyme catalase. Catalase converts H_2O_2 to O_2, the production of which can be monitored using an Oxygraph Hansatech according to a previously described method[16]. Both sensitive analytical methods yielded a negative result for H_2O_2, indicating that oxygen is likely reduced to water.

CABA binds to ScoE in a similar orientation to Taurine in TauD

To probe the structural basis of isonitrile formation by an Fe(II)/αKG-dioxygenase, a structure of ScoE from *S. coeruleorubidus* was solved to 1.70 Å resolution with CABA bound (**Table S1**). In the ScoE active site, clear electron density was observed for a bound CABA (**Fig. 3A**) in the same location that was proposed to be the CABA binding site from the choline-bound ScoE structure[4] and shown to be the

CABA binding site in the recent ScoE structure[13]. As expected[4, 13], Fe(II) is bound at the conserved 2-His-1-carboxylate facial triad motif. An acetate molecule from the crystallization buffer occupies the putative α-KG binding site. The CABA binding site in ScoE is reminiscent of the taurine binding site in the archetypal α-KG dependent dioxygenase TauD from *E. coli* (**Fig. 3B**)[17]. In both active sites, the respective substrate is bound axial to a histidine residue in the conserved facial triad. A conserved arginine (Arg310 in ScoE; Arg270 in TauD) coordinates the carboxylate moiety of CABA in ScoE and the sulfonate moiety of taurine in TauD. CABA is further locked in the active site by an interaction between the other substrate carboxylate moiety and Lys193.

A conformational change in Lysine 193 occurs upon CABA binding

A comparison of the choline-bound and CABA-bound ScoE structures reveals a conformational change of Lys193 that positions the primary amine of the side chain to interact with CABA (**Fig. 4**). In the presence of non-substrate choline, the side chain of Lys193 is flipped away from the ScoE active site. As Lys193 is at the protein surface, the "flipped out" conformation of this side chain exposes the CABA binding site to solvent. When CABA binds, Lys193 flips in, burying the untethered CABA[4] in the active site (**Fig. 4B**).

An ordered water network extends from the CABA binding site to the protein surface

The observed orientation of CABA in the ScoE active site appears to be maintained by electrostatic interactions with Lys193 and Arg310 on either end of the CABA substrate and by a hydrogen bond between Tyr96 and the CABA secondary amine in the middle of the CABA molecule (**Fig. 5**). To test the importance of this interaction, we prepared a ScoE variant in which this tyrosine residue was substituted with phenylalanine (ScoE-Y96F), and detected no product formation from this variant in an *in vitro* assay (**Fig. 5B**). This protein:CABA interaction helps organize a larger hydrogen bonding network that extends to the protein surface. Tyr96, Tyr97, Tyr101, and the backbone of Lys193 form hydrogen bonds to stabilize three ordered water

molecules within the protein (**Fig. 5A**). Furthermore, the backbone amine of Lys193 makes a hydrogen bond with the backbone carbonyl of Arg195, aiding in positioning this side chain for hydrogen bonding to the side chains of Asp198 and Glu209. Both of these side chains also make hydrogen bonds with Arg201 at the surface of the protein (**Fig. 5A**).

We examined the importance of this hydrogen bonding network by altering residues in this network (ScoE-Y101F and ScoE-R195Q) (**Fig. 5A**). Product formation was not detected from *in vitro* ScoE activity assays with either variant (**Fig. 5B**). Product formation was however detected *in vitro* with another variant, ScoE-Y97F, but the yield of product appears to be lower than that of wild-type ScoE, suggesting that the hydrogen bonding network is compromised but still intact in this variant. Strikingly, the observed loss of activity in both ScoE-Y101F and ScoE-R195Q variants despite the preservation of the hydrogen bond at Tyr96 suggests that the hydrogen bonding network plays a catalytic role in isonitrile formation assuming that the structural integrity of the protein is maintained.

Without a second Arg in the active site, α-KG binding appears to be weak

Following characterization of our structure of ScoE bound to CABA, we sought to structurally characterize the α-KG bound state. We obtained crystals by anaerobic co-crystallization of ScoE in the presence of α-KG and Fe(II), and the structure was solved to 1.85 Å resolution (**Table S1**). Weak electron density for a bidentate-bound molecule was observed at the metal center that was modeled as a single molecule of α-KG (**Fig. 3C**).

The molecule coordinates the metal center in a bidentate fashion, with the 1-carboxylate and the α-keto moieties each acting as ligands to the mononuclear Fe(II). The α-KG molecule is also partially disordered, and electron density is not observed for C4 and C5 of this molecule, consistent with a lack of interactions being made between α-KG and protein side chains. However, modeling of α-KG was aided by the observation of a hydrogen bond between the 1-carboxylate moiety of α-KG and the side chain of Thr127. This binding mode results in C4 and C5 of α-KG intruding upon the

CABA binding site. Furthermore, a comparison to other α-KG dependent dioxygenase active sites, such as that of TauD, reveals that ScoE lacks a conserved arginine (Arg266 in TauD) that is responsible for interacting with the 5-carboxylate moiety of α-KG (**Fig. 3**). Due to the unexpected nature of this α-KG binding mode, we have termed this configuration of the co-substrate an off-site configuration.

CABA binding does not appear to stabilize the α-KG binding site

After observing α-KG bound in an off-site configuration, we hypothesized that CABA binding may promote a conformational change to form a more stable α-KG binding site. To test this idea, we solved a crystal structure of ScoE in the presence of both α-KG and CABA to 1.45 Å resolution (**Table S1)**. The observed electron density in the active site was again consistent with α-KG bound in an off-site configuration (**Fig. 3D**). Furthermore, weak electron density is also observed for CABA in the previously described CABA binding site. Critically, we did not observe conformational changes in active site side chains that would indicate a new α-KG binding site (**Fig. 3D**). The observation of the respective binding modes of CABA and α-KG was unexpected due to the substantial steric clashes expected from bringing the two molecules into such close proximity. In order to prevent these substantial clashes while refining the structure, the CABA bound state and the off-site α-KG state were modeled as alternate conformations of each other, suggesting that the two states are mutually exclusive: CABA cannot bind to ScoE when α-KG is bound in this configuration and vice versa. Although typically in this enzyme superfamily substrate and α-KG both bind before the reaction of α-KG with O_2 begins[5], there is precedent for α-KG binding and the generation of an offline Fe(IV)=O prior to substrate binding[18, 19]. Here, substrate binding is thought to cause the off-line Fe(IV)=O to reposition to an on-line configuration that is appropriate for hydrogen atom abstraction[20]. In this case, the enzyme could use the same binding for substrate and α-KG. Thus, we investigated whether α-KG can react with O_2 in the absence of CABA and if the ScoE active site can accommodate an off-line Fe(IV)=O intermediate.

α-KG can react with O_2 in the absence of CABA

We tested whether CABA binding is required for the consumption of oxygen and α-KG by measuring oxygen consumption in the presence and absence of CABA. As a control, we measured the consumption of oxygen in the absence of α-KG, and under this condition, no oxygen consumption was observed (**Fig. S9**). Following this control experiment, we observed clear consumption of oxygen by ScoE in the presence of α-KG and in the absence of CABA, with a possible plateau in the oxygen consumption after approximately 300 seconds. Addition of CABA results in further oxygen consumption, and the rate of this oxygen consumption appears to be greater when CABA is present (**Fig. S9**). However, we did not measure the rate of this reaction in this work.

A ScoE structure with a ferryl-oxo mimic demonstrates a potential off-line oxo species

With the knowledge that α-KG can react with O_2 in the absence of CABA, we investigated if the ScoE active site can accommodate an off-line Fe(IV)=O intermediate. Direct structural characterization of a protein-bound Fe(IV)=O species is precluded by the high reactivity of this intermediate. However, vanadium(IV) in the presence of oxygen has been demonstrated to be a suitable and stable mimic of the Fe(IV)=O intermediate for use in X-ray crystallography[21, 22]. Crystals of ScoE with oxovanadium were obtained by aerobic crystallization of ScoE in the presence of vanadium(IV) sulfate and CABA, and the structure was solved to 2.1 Å resolution (**Table S1**). CABA binds as observed previously. Oxovanadium appears bound in the active site, coordinated by the conserved 2-His-1-carboxylate facial triad (**Fig. 3E**). Strikingly, the oxygen atom of the mimic is not oriented toward CABA, and is instead observed oriented away from the CABA binding site, in what would be an off-line configuration (**Fig. 3E**). This orientation is stabilized by a hydrogen bond made between the oxovanadium species and Arg310. Critically, the observation of this off-line configuration is consistent with the observation of an off-site bound α-KG molecule, suggesting that prior to CABA binding an offline Fe(IV)=O species may form as a result of the reaction of molecular oxygen and α-KG.

Once CABA or a CABA intermediate is bound, however, another binding site for α-KG would appear to be required.

A ScoE structure with tartrate bound reveals a new potential α-KG binding site

A vital insight into the binding of α-KG was gained from the recent reporting of another crystal structure of ScoE in the presence of CABA and tartrate, solved to 2.18 Å resolution by Chang and colleagues[13]. As mentioned above, the CABA binding site is identical to the site we have observed in the 1.7 Å resolution structure we report here. As a result of the precipitant solution used, a molecule of tartrate was observed bound to Fe(II), and unlike what we report above for α-KG, this tartrate molecule does not intrude upon the CABA binding site. Tartate binding in this ScoE structure is stabilized by a striking conformational change that repositions His299 and Arg157 such that their side chains are stacked with each other and are pointing toward the carboxylate moiety of tartrate (**Fig. 3F**). Modeling α-KG in place of tartrate demonstrates that the binding site defined by Arg157/His299 is large enough to both accommodate α-KG and preserve the interaction between this molecule and the side chain (**Fig. S10**). Additionally, this putative binding site for α-KG in ScoE is reminiscent of the α-KG binding site in TauD (**Fig. 3B,F, S10**).

In all crystal structures lacking tartrate both reported here and previously[4, 13], Arg157 and His299 are both flipped away from the ScoE active site and do not interact with one and other (**Fig. 3F**). These outward facing positions are observed even when CABA is present, indicating that this conformational change is not induced by CABA binding. The change itself is non-trivial with the repositioning of His299 and Arg157 side chains being a consequence of a dramatic alteration in the hydrogen bonding patterns between two strands in which side chain (Asn158, Arg298) to backbone hydrogen bonds are replaced by backbone to backbone hydrogen bonds (**Fig. 6A,B**). With His299 and Arg157 pointed out, there is an access channel to the Fe(II), and with His299 and Arg157 pointing in, tartrate is sequestered in the active site (**Fig. 6C,D**). We suspect that in solution, α-KG binding induces the conformational change that brings His299 and Arg157 into the active site, but that lattice contacts in our crystals between His299 and a

neighboring molecule (**Fig. S11**) block the ability of α-KG to induce this conformational state, hindering our ability to visualize an Arg157/His299: α-KG interaction. We also suspect that the outward facing conformations of these residues is physiological relevant, since there must be a route for α-KG to enter the active site and for succinate and CO_2 to leave.

If we assume that Arg157 and His299 are the anchors for α-KG binding, then the question arises if both α-KG molecules that are required for INBA generation will interact with Arg157/His299. It is possible that prior to CABA binding, α-KG could bind to ScoE in the observed off-site binding mode and react to form a protected, off-line Fe(IV)=O intermediate. In this case, CABA binding would trigger a rearrangement of this intermediate and the first hydroxylation would occur (**Fig. S12**). The presence of a bound CABA-hydroxylated intermediate would preclude off-site α-KG binding due to the extensive steric clashes that this binding would entail, necessitating the movement of His299 and Arg157 into the active site to form a new α-KG binding site, allowing for the second α-KG-dependent reaction.

Computational analyses suggest that movement of His299 and Arg157 are energetically feasible

To test the hypothesis that Arg157 and His299 could dynamically reposition and form interactions with the active site, we carried out classical molecular dynamics (MD). All structures were modeled to contain an off-line Fe(IV)=O intermediate and bound succinate, with His299 and Arg157 modeled in both inward positions from the structure of Chang and colleagues[13], only His299 or both in outward positions for the structure from the present work, as well as around nine favorable rotameric states supported by the local protein environment (Fig. S13 and Table S2). In total, over 3 μs of MD from these simulations suggest high flexibility of His299, which readily interconverts between the outward and inward orientations in almost all trajectories (Fig. S15). While reorientation of Arg157 is less frequently observed in simulations, movement from the outward to the inward orientation in conjunction with His299 motion to an inward state does occur on the ns timescale (Fig. S16). Once in the

inward orientations, both His299 and Arg157 are sufficiently proximal to form hydrogen bonds with the carboxylate tail active site substrates (Fig. S14 and Fig. S15).

Substitution of residues Arg157 and His299 in the putative α-KG binding site leads to loss of enzyme activity

To probe the potential α-KG binding site described above, we generated three ScoE variants: His299Gln, Arg157Gln, and Arg157Glu and investigated the ability of these enzyme variants to produce both INBA (**Fig. 6E**) and succinate (**Fig. 6F**) when reacted with α-KG and CABA. We found that His299Gln completely eliminated production of INBA and succinate in these assays and Arg157 variants were less active compared to the wild type counterpart (**Fig. 6E,F**). These results suggest that His299 and Arg157 play a role in catalysis with His299 being the more important of the two residues. Given the positioning of these residues far from the CABA binding, even when these residues are swung-in (**Fig. 3F**), we attribute the loss in activity of our ScoE variants to one or more of the reactions with α-KG. As discussed above, the first half-reaction with α-KG could occur using the off-site binding mode that exists prior to CABA binding (**Fig. S12B**). Thus, we next considered whether Arg157/His299 are required for the ability of ScoE to react with α-KG in the absence of CABA (**Fig. 6G**). Mimicking the results with CABA, we found that without CABA, His299Gln ScoE was completely unable to produce succinate from α-KG and the Arg157 variants had decreased ability to produce succinate (**Fig. 6G**). These data suggest that there is only one binding site for α-KG and that this site is not the off-site binding mode observed when lattice contacts prevent His299/Arg157 movement, but rather, the binding site for α-KG is inducible and requires His299.

Discussion

The discovery that isonitrile biosynthesis can occur by desaturation of a carbon-nitrogen bond[3] prompted immediate speculation as to the mechanism of the enzyme

responsible. Bioinformatics, structural and biochemical characterizations firmly placed one of the responsible enzymes, ScoE, into the large and diverse Fe(II)/α-KG dioxygenase superfamily[3, 4]. Best known for catalyzing hydroxylation reactions, enzymes in this superfamily had not previously been associated with isonitrile formation, raising the question as to how ScoE is able to perform this unprecedented chemistry.

A key piece of missing data when we started this study was the stoichiometry of the ScoE reaction. In particular, we wanted to know the number of α-KG molecules that were required for isonitrile formation. These data were particularly important as we tried to understand our structural data, which showed electron density for α-KG overlapping with the CABA binding site (established both by our work and that of others)[4, 13]. We reasoned that α-KG could bind to this overlapping site in the absence of CABA, but if the reaction required two α-KG molecules, then there must be another α-KG binding site that was eluding detection. The alternative that a newly formed CABA intermediate must depart ScoE and later rebind to allow for a second α-KG reaction seemed unlikely. The experiments described here have now firmly established a stoichiometry of 2 α-KG molecules per INBA formed, leading us to believe that there must be an additional α-KG binding site on ScoE.

Inspection of our ScoE structures revealed that the arginine typically associated with α-KG binding was missing in ScoE. However, in a recent ScoE structure with tartrate[13], an arginine was found in the active site. This arginine (Arg157), along with His299, moved over 10 Å from a position on the surface of the protein into the active site. Molecular dynamics (MD) trajectories initialized from either crystal structure pose supported the expectation that interconversion between these two poses could occur at physiological temperatures. Conversion from the outward to inward poses of both His299 and Arg157 were directly observed on modest timescales (< 250 ns). Here we show that substitution of these residues abolished or decreased both the cleavage of α-KG into succinate and INBA production. It is surprising that His299, and not Arg157, is the more important of the two residues, given that an Arg residue is typically associated with α-KG binding and that Arg157 in the "in" position occupies the equivalent site as the α-KG-binding arginine in TauD (**Fig. 3B,F**). Thus, it appears that ScoE differs from

canonical α-KG -Fe(II) enzymes in that the residue(s) involved in α-KG cleavage are mobile and are not exclusively arginine, but ScoE shares similarities too. The discovery of this 'inducible' α-KG binding site means that ScoE is similar to other family members[5] in that it appears to have an α-KG binding site that does not overlap with the substrate binding site, allowing ScoE to form a ternary complex (α-KG-CABA-ScoE). Formation of a ternary complex is a common feature in this superfamily[5] that limits the generation of a highly reactive Fe(IV)=O species before substrate is available for reaction. Enzymes that do not form ternary complexes rely on the formation of an off-line Fe(IV)=O species that is protected in the absence of substrate to prevent unwanted chemistry. Our structure with oxovanadium bound to ScoE indicates that the ScoE active site is capable of accommodating an off-line Fe(IV)=O species. However, we favor the mechanistic option that has the greater precedence in this superfamily, in which ScoE forms a ternary complex with both α-KG and CABA bound at the same time using the 'inducible' site for α-KG binding (**Fig. 7**).

As far as we know, no other Fe(II)/α-KG dioxygenase has an 'inducible' α-KG binding site with Arg157/His299 swinging 10 Å to create a binding site. The reason for this difference could lie in the nature of the two co-substrates: CABA and α-KG. These molecules are very similar to each other and an enzyme that binds both of them must have evolved so that CABA binds in its binding site and not where α-KG should bind and vice versa. Our structural data of ScoE in the presence of both α-KG and CABA, in which neither co-substrate is bound with high occupancy, may reflect on the design challenge represented by using such similar co-substrates. Therefore, we speculate that to prevent CABA from binding between the Fe(II) and Arg157/His299, these residues are not permanently placed in the active site.

Another motivation for our structural studies was to determine which atoms on CABA might be accessible to the Fe(IV)=O cofactor for hydroxylation, given that the CABA modification site(s) have not been firmly established. The two atoms of CABA that are closest to the Fe ion are N and C5 (**Fig 3A**). From a structural prospective exclusively, C5 of CABA appears the most likely site of a first hydroxylation, which favors the mechanistic schemes shown in **Fig. 2i** and **2ii**. However, the scheme shown

in **Fig. 2iii**, cannot be ruled out as a small arrangement of CABA would make the N accessible. Additionally, this work has revealed that the residue that contacts N of CABA, Tyr96, is catalytically essential.

Although there is still much to learn about ScoE, these experiments have started to unravel ScoE's enigmatic reaction mechanism. Regardless of what future experiments show, it is not possible to imagine results that would allow one to draw a reaction mechanism in which each step has precedence in this enzyme class. Additionally, regardless of mechanism, this reaction is nothing like the previously known biological method for isonitrile formation that involves a remarkable condensation reaction. Although click chemistry that utilizes isonitriles might seem a bit mundane due to its ubiquitous use in biotechnology, there appears to be nothing about the biological formation of isonitriles that is mundane. The chemistry involved in producing isonitriles in biology is nothing short of fascinating.

Materials and Methods

Construction of plasmids for protein expression in E. coli Individual genes were PCR amplified from genomic DNA and cloned into either pET24B, pETCDFDuet-1, or pET30 by means of restriction enzyme digestion (ThermoFisher) and ligation with Quick T4 DNA Ligase (New England Biolabs). The primers used in this study are reported in Table M1. Plasmids were extracted using a Zyppy Miniprep Kit (Zymo Research) and confirmed by DNA sequencing at the UC Berkeley Sequencing Facility.

Point mutations of ScoE were generated by PCR using Phusion HighFidelity DNA Polymerase (Thermo Scientific) and the primers listed in table M1. pET30- ScoE was used as a template for all mutants with the exception that R157Q was used as a template to generate R157E. Reactions were conducted according to the manufacturer's protocol with supplied buffer using DMSO and each pair of primers. The PCR program began at 98 °C for 30 s, followed by 30 cycles of 72 °C for 10 s, 72 °C for 5 min, and final extension at 72 °C for 5 min. The template DNA was digested with 10 units of DpnI (Thermo Scientific) for 1 h at 37 °C, and the remaining PCR product was transformed into *E. coli* XL-1 Blue competent cells by heat-shock. The introduction of point mutation was confirmed with DNA sequencing to yield pET30-ScoE-R157E, H299Q, Y101F, Y96F, Y97F, and R195Q.

Table M1: List of primers used.	
Primer	**Sequence (5' -> 3')**
pET30 scoE-F	AAAGGATCCGATGCAGATCGACGAACAGCC
pET30 scoE-R	TATAAGCTTTCATGCCGCCTGGATCCCGT
pET30 R157E-F	AGGTAATTCCGGAGAAGAACGAGGGCACGTACTTCATCGACA
pET30 R157E-R	TGTCGATGAAGTACGTGCCCTCGTTCTTCTCCGGAATTACCT
pET30 R157Q-F	AGGTAATTCCGGAGAAGAACCAGGGCACGTACTTCATCGACA
pET30 R157Q-R	TGTCGATGAAGTACGTGCCCTGGTTCTTCTCCGGAATTACCT
pET30 H299Q-F	TCCACCGCGCCCGGCAGACGACCACACCCGAGCC
pET30 H299Q-R	GGCTCGGGTGTGGTCGTCTGCCGGGCGCGGTGGA
pET30 Y101-F	CCTACTACGAGCCGATGTTCCAGCACCCGGAGGTCA
pET30 Y101-R	TGACCTCCGGGTGCTGGAACATCGGCTCGTAGTAGG
pET30 Y96F-F	GGCCGCCCGGAGGCCTTCTACGAGCCGATGTACCAGCA
pET30 Y96F-R	TGCTGGTACATCGGCTCGTAGAAGGCCTCCGGGCGGCC
pET30 Y97F-F	GGCCGCCCGGAGGCCTACTTCGAGCCGATGTACCAGCA
pET30 Y97-R	TGCTGGTACATCGGCTCGAAGTAGGCCTCCGGGCGGCC
pET30 R195Q-F	GCAAGTACTTCAAGATCCAGCCCCACGATGTCTACCG
pET30 R195Q-R	CGGTAGACATCGTGGGGCTGGATCTTGAAGTACTTGC
pETCDFDuet-1 scoA-F	AAAGAATTCGATGTCACCGCATGACGACGC
pETCDFDuet-1 scoA-R	TATAAGCTTCTACTTGGCGGGCATTGCCG
pET24B scoB-F	AAACATAATGCCTGCTCCCCTCACGCT
pET24B scoB-R	AAACTCGAGTCATGCGGTGACATGGCCCG
pETCDFDuet-1 scoC-F	AAAGGATCCGATGGACCGGCTCCACCACCC
pETCDFDuet-1 scoC-R	AAACTGCAGTCAGTTGACCTTGCGTGCGG

Overexpression and purification of ScoE for crystallization

The expression and purification for all proteins used in this study followed the same general procedure for His6-tag purification as detailed here. BL21 cells were grown at 37°C in 1L of LB in a shake flask supplemented with 50 μg/mL of kanamycin to an OD_{600} of 0.5 at 250 rpm. The shake flask was then placed over ice for 10 mins and induced with 120 µM of isopropyl-β-D-thiogalactopyranoside (IPTG). The cells were then incubated for 16 hours at 16°C and 200 rpm to undergo protein expression.

Subsequently, the cells were harvested by centrifugation (6,371 x g, 15 min, 4°C), and the supernatant was removed. The cell pellet was resuspended in 30 mL of lysis buffer (25 mM HEPES pH 8, 500 mM NaCl, 5 mM imidazole) and cells were lysed by sonication on ice. Cellular debris was removed by centrifugation (27,216 x g, 1 hour, 4°C) and the supernatant was filtered with a 0.45 µm filter before batch binding. Ni-NTA resin (Qiagen) was added to the filtrate at 1.5 mL/L of cell culture, and the samples were allowed to nutate for 1 hour at 4°C. The protein-resin mixture was loaded onto a gravity flow column. The flow through was discarded and the column was then washed with approximately 25 mL of wash buffer (25 mM HEPES pH 8, 100 mM NaCl, 20 mM imidazole) and tagged protein was eluted in approximately 20 mL of elution buffer (25 mM HEPES pH 8, 100 mM NaCl, 250 mM imidazole). The whole process was monitored using a Bradford assay. Purified proteins were concentrated using Amicon Ultra spin filters to yield 3 mL of protein. ScoE was dialyzed at 4°C using a 10 kDa Slide-A-Lyzer Dialysis Cassettes in 1 L of dialysis buffer (25 mM HEPES pH 8, 100 mM NaCl, 1 mM EDTA). The dialysis buffer was changed twice over 9 hours and dialyzed overnight. After dialysis, ScoE was desalted to remove EDTA using a GE PD-10 column and eluted into a buffer containing 25 mM HEPES pH 8 and 100 mM NaCl. ScoE was concentrated using a 10 kDa Amicon Ultra spin filter until the protein concentration reached 40 mg/mL and 10% v/v glycerol was added. The proteins were flash frozen in liquid nitrogen and stored at -80°C.

Overexpression and purification of proteins for in vitro assays

The expression and purification for all proteins used in this study followed the same general procedure for His6-tag purification as detailed here. BL21 cells (BAP1 for ScoB) were grown at 37°C in 1L of LB in a shake flask supplemented with 50 μg/mL of kanamycin to an OD_{600} of 0.5 at 250 rpm. The shake flask was then placed over ice for 10 mins and induced with 120 µM of isopropyl-β-D-thiogalactopyranoside (IPTG). The cells were then incubated for 16 hours at 16°C and 200 rpm to undergo protein expression. Subsequently, the cells were harvested by centrifugation (6,000 x g, 15 min, 4°C), and the supernatant was removed. The cell pellet was resuspended in 30 mL of

lysis buffer (25 mM HEPES pH 8, 500 mM NaCl, 5 mM imidazole) and cells were lysed by sonication on ice. Cellular debris was removed by centrifugation (15,000 x g, 30 min, 4°C) and the supernatant was filtered with a 0.45 µm filter before batch binding. Ni-NTA resin (Qiagen) was added to the filtrate at 1.5 mL/L of cell culture, and sample were allowed to nutate for 1 hour at 4°C. The protein-resin mixture was loaded onto a gravity flow column. The flow through was discarded and the column was then washed with approximately 25 mL of wash buffer (25 mM HEPES pH 8, 100 mM NaCl, 20 mM imidazole) and tagged protein was eluted in approximately 20 mL of elution buffer (25 mM HEPES pH 8, 100 mM NaCl, 250 mM imidazole). The whole process was monitored using a Bradford assay. Purified proteins were concentrated using Amicon Ultra spin filters to yield 3 mL of protein. ScoE was dialyzed at 4°C using a 10 kDa Slide-A-Lyzer Dialysis Cassettes in 1 L of dialysis buffer (25 mM HEPES pH 8, 100 mM NaCl, 10 mM EDTA). The dialysis buffer was changed twice over 9 hours, where the final buffer exchange contained 1 mM EDTA. ScoE was dialyzed overnight with this final buffer. After dialysis, ScoE was concentrated using a 10 kDa Amicon Ultra spin filter and 10% v/v glycerol was added. The proteins were flash frozen in liquid nitrogen and stored at -80°C. For ScoA, B, and C, these proteins were concentrated in appropriate Amicon Ultra spin filters and exchanged in exchange buffer (25 mM HEPES pH 8, 100 mM NaCl). After three rounds of exchange, the purified proteins were flash frozen and stored as mentioned previously. Presence and purity of enzymes was assessed using SDS-PAGE (**Figure M1**) and their concentration was determined using a NanoDrop UV-Vis spectrophotometer (ThermoFisher). The approximate protein yields were 10.4 mg/L for ScoA (158 kDa), 3.4 mg/L for ScoB (12 kDa), 17.2 mg/L for ScoC (57 kDa), 5.2 mg/L for ScoE (37 kDa), 4.5 mg/L for Y96F ScoE (37 kDa), 4.7 mg/L for Y97F ScoE (37 kDa), 5.0 mg/L for R195Q ScoE (37 kDa), 4.8 mg/L for Y101F ScoE (37 kDa), 4.6 mg/L for H299Q ScoE (37kDa), 5.5 mg/L for R157Q ScoE (37 kDa), and 5.4 mg/L for R157E ScoE (37 kDa).

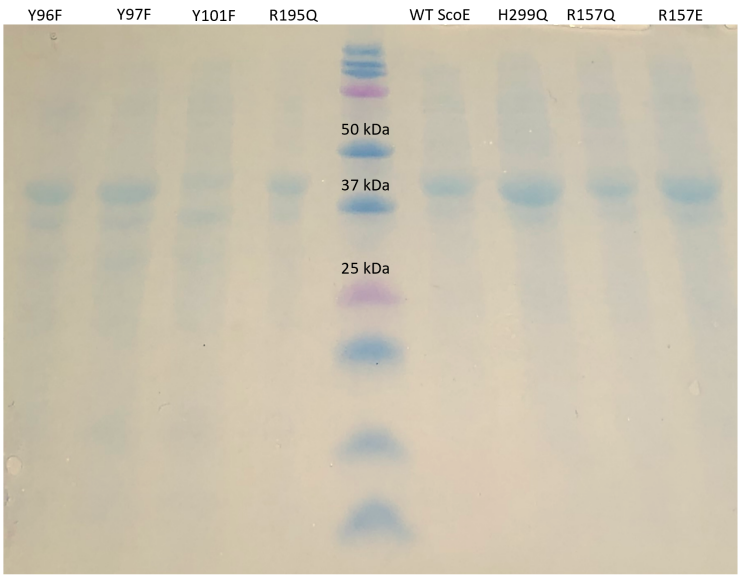

Figure M1: SDS-PAGE analysis of recombinant *E. coli* proteins purified and utilized in this study. ScoE and single amino acid variants used for *in vitro* biochemical assays. Any kD Mini-PROTEAN TGX gels (precast, Biorad) were used for analysis of single amino acid variants of ScoE in this study. All recombinant proteins contain an N-hexahistidine tag and were largely soluble in *E. coli*.

ScoE reconstitution to yield holo-enzyme

Immediately before each assay, a Bio-Rad Bio-Gel P-6 gel column was equilibrated in exchange buffer containing 25 mM HEPES pH 8 and 100 mM NaCl. 50-90 μL of ScoE was desalted to remove EDTA following the manufacturer's protocol. ScoE was immediately added to the biochemical assay of interest and supplemented with ammonium iron(II) sulfate hexahydrate matching 90% of the protein concentration.

Synthesis of isotope labeled CABA

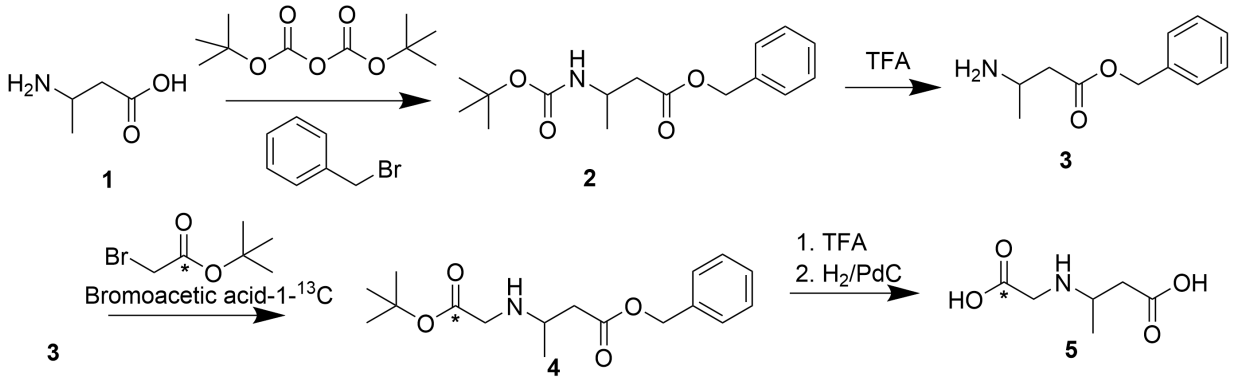

Figure M2: Synthesis of isotope labeled CABA. Compound **3** was mixed with labeled bromoacetic acid to give **5**.

In general, the synthesis of CABA was followed as previously reported[4]. Bromoacetic acid-2-^{13}C was used for CABA-6-^{13}C (**5**) synthesis (**Figure M2**). The product was a colorless oily liquid and was confirmed by HRMS and ^{1}H NMR[4]. The product was a colorless oily liquid. Calculated for $C_6H_{12}NO_4$ (6-^{13}C) $[M+H]^+$ 163.0794, found: 163.0794 (0 ppm error). ^{1}H NMR (900 MHz, DMSO-*d6*) δ 3.77 (s, 2H), 3.51 (dd, J = 13.2, 6.6 Hz, 1H), 2.81 (dd, J = 16.6, 4.4 Hz, 1H), 2.48 (s, 1H), 1.23 (d, J = 6.6 Hz, 3H); ^{13}C NMR (226 MHz, DMSO-*d6*) δ 171.70, 168.51, 50.16, 49.94, 45.23, 16.30.

O_2 Consumption from ScoE in vitro reaction

To investigate O_2 consumption from the ScoE *in vitro* reaction, a 0.5 mL single-turnover assay containing 50 mM HEPES pH 8, 500 µM CABA, 250 *µ*M α-KG, 100 µM Apo-ScoE, and 90 µM $(NH_4)_2Fe(SO_4)_2$ was conducted in a sealed reaction chamber with an integrated oxygen electrode unit (Oxygraph Plus System, Hansatech Instruments, UK) . The instrument was calibrated with 0.5 mL of 50 mM HEPES pH 8.0 at room temperature and a mixing rate of 50 RPM. The assay components were added to the reaction chamber except for the iron. The oxygen signal was allowed to equilibrate before addition of 90 µM $(NH_4)_2Fe(SO_4)_2$ with an airtight needle to initiate the reaction. Oxygen consumption was collected as a function of time, and the reaction was allowed to incubate until the profile reached a steady state. The difference between the

starting and final oxygen concentrations was used to estimate O_2 consumption. Two 100 µL aliquots of the reaction mixture were quenched with 200 µL of cold methanol and 200 µL of 667 µM 3,6-di(pyridine-2-yl)-1,2,4,5-tetrazine dissolved in cold methanol, respectively for coupled stoichiometry quantification experiments.

Detection and quantification of isonitrile butanoic acid (INBA) using tetrazine click reaction

A ScoE *in vitro* assay containing the same reaction components as mentioned previously was quenched with 200 µL of 667 µM 3,6-di(pyridine-2-yl)-1,2,4,5-tetrazine dissolved in cold methanol, gently mixed, and incubated for 1 hour at room temperature. The reaction mixture was vortexed briefly and centrifuged to remove aggregated protein. 100 µL standards containing 50 mM HEPES pH 8, 500 μM CABA, 250 μM α-KG, and varying concentrations of 3,5-di(pyridine-2-yl)-1H-pyrazol-4amine (Py-aminopyrazole) were quenched using the same quench method to develop a standard curve using LC-HRMS. LC-HRMS analysis was performed on an Agilent Technologies 6545 Accurate-Mass Q-TOF LC-MS instrument and an Eclipse Plus C18 column (100 x 4.6 mm). Chromatography was performed using a linear gradient of 10-50% acetonitrile (vol/vol) with 0.1% formic acid over 12 minutes in water with 0.1% formic acid (vol/vol) at a flow rate of 0.5 mL/min. A Py-aminopyrazole standard was synthesized and characterized by NMR as previously described[14] and was observed by LC-HRMS as $[M+H]^+$=238.1095 (calculated $[M+H]^+$: 238.1087, 3.4 ppm error). A standard curve was used to quantify the production of INBA.

Quantification of succinate production, CABA consumption, and α-KG consumption

A ScoE *in vitro* assay containing the same reaction components as mentioned previously was quenched with cold methanol, gently mixed, vortexed, and centrifuged to remove the aggregated protein. 100 µL standards containing 50 mM HEPES pH 8 and varying concentrations of CABA, α-KG, and succinate were quenched with 200 µL of cold methanol to develop a standard curve using LC-HRMS. LC-HRMS analysis was

performed on an Agilent Technologies 6545 Accurate-Mass Q-TOF LC-MS instrument and an Eclipse Plus C18 column (100 x 4.6 mm). Chromatography was performed using a linear gradient of 10-50% acetonitrile (vol/vol) with 0.1% formic acid over 12 minutes in water with 0.1% formic acid (vol/vol) at a flow rate of 0.5 mL/min. Extracted ion chromatograms (EICs) for CABA, α-KG, and succinate where generated using the following *m/z*: $[M+H]^+$=162.0761 for CABA, $[M-H]^-$=145.0142 for α-KG, and $[M-H]^-$=117.0193 for succinate. The following m/z values were observed: $[M+H]^+$=162.0763 for CABA (1.2 ppm error), $[M-H]^-$=145.0142 for α-KG (0 ppm error), and $[M-H]^-$=117.0193 for succinate (0 ppm error). Standard curves for CABA, α-KG, and succinate were utilized to quantify their respective consumption and production

H_2O_2 Detection Methods

To detect H_2O_2, the 0.5 mL ScoE enzymatic assay mentioned earlier was conducted with the same oxygen electrode unit. After the concentration of oxygen reached equilibrium, 120 U/mL of catalase was added to the reaction chamber. No appreciable increase of oxygen was observed when compared to adding solely water or buffer, indicating that no H_2O_2 was present in the reaction mixture. In addition, H_2O_2 was again undetected using the Amplex Red Hydrogen Peroxide/Peroxidase Assay Kit by following the manufacturer's instructions.

Formate Detection Methods

A 200 µL in vitro assay containing 50 mM HEPES pH 8, 500 µM CABA, 250 µM α-KG, 100 μM Apo-ScoE, and 90 μM $(NH_4)_2Fe(SO_4)_2$ was performed and incubated at room temperature for 10 minutes. This assay was adapted from previous work[2]. The reaction was quenched with 400 µL of cold MeOH and 100 µL of the supernatant was reacted with 10 µL of 290 mM 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and 10 µL of 120 mM 2-nitrophenylhydrazine (dissolved in 250 mM HCl). The reaction was incubated at 60 ˚C for 15 minutes and centrifuged. The supernatant was analyzed with LC-HRMS analysis on an Agilent Technologies 6545 Accurate-Mass Q-TOF LC-MS instrument and an Eclipse Plus C18 column (100 x 4.6 mm). Chromatography was

performed using a linear gradient of 10-50% acetonitrile (vol/vol) with 0.1% formic acid over 12 minutes in water with 0.1% formic acid (vol/vol) at a flow rate of 0.5 mL/min. A standard curve of all the previously mentioned components (except enzyme) was constructed using different sodium formate concentrations. Extracted ion chromatograms were constructed with $[M+H]^+$= 182.0561 (observed: $[M+H]^+$= 182.0558, 1.6 ppm error). Additionally, a Sigma-Aldrich Formate Assay Kit was utilized according to the manufacturer's protocol, resulting in no detectable amounts of formate.

CO Detection Method

A 200 µL in vitro assay containing 50 mM HEPES pH 8, 500 µM CABA, 250 µM α-KG, 100 μM Apo-ScoE, and 90 μM $(NH_4)_2Fe(SO_4)_2$ was performed and incubated at room temperature for 10 minutes in a sealed septum. This assay was adapted from previous work[2]. A solution of myoglobin and sodium dithionite was added to give 10 μM and 20 mM final concentrations, respectively, and the reaction mixture was allowed to incubate for 10 minutes. The septum was immediately removed, and the absorption spectrum was recorded right after (150 µL) using a flat bottom 96 well plate and Biotek plate reader. The Soret band remained at 434 nm for both the reaction-Fe control, and -ScoE control, supporting the absence of CO.

GC-MS Experiments

For the detection of CO_2 and CO, a 4.5 mL of reaction mixture contained 50 mM HEPES pH 8, 90 µM $(NH_4)_2Fe(SO_4)_2$, 2 mM (8 or unlabeled), 1 mM α-KG (1,2,3,4- ^{13}C labeled or unlabeled), and 100 µM ScoE. The reaction was performed in a 10-ml sealed headspace vial (Agilent) and initiated by adding the iron. 1 mL of the headspace gas was acquired using a gastight syringe (Hamilton) and injected into Agilent 5977A GCMS system equipped with a HP-5ms column. Injector temperature was set at 120 °C, and helium was used as the carrier gas at a flow-rate of 3 mL/min. The temperature gradient was as follows: 40-100 °C over 5 minutes. The mass spectrometer was operated in electron ionization mode with automatically tuned parameters, and the acquired mass range was m/z = 15-100. The CO_2 and CO signals were confirmed using authentic

standards, and the production of $^{13}CO_2$ and ^{13}CO was determined by extracting m/z = 45 and 29, respectively.

Crystallization

Crystals, which were used in the soaking experiments described below as well as to obtain the structure of Fe(II)-ScoE with an off-site α-KG, were grown by the sitting drop vapor diffusion technique at 24°C in an anaerobic environment (95% Ar, 5% H_2; Coy Laboratory Products, Inc.), following optimization of the previously published precipitant solution[4]. A 1 μL aliquot of protein solution (8 mg/mL metal-chelated ScoE, 50 mM HEPES pH 8, 100 mM NaCl, and 1 mM α-KG) was added to 2 μL of precipitant solution (205 mM sodium acetate, 100 mM Tris pH 8.5, 24% (w/v) PEG 4,000) supplemented with 250 μM Fe(II)Cl_2 in a 24-well sitting drop tray. Crystals formed within 1 week and reached a maximum size of approximately 50 μm after approximately 2 weeks. Paraffin oil was used for cryoprotection during crystal harvesting. Crystals were manually looped and streaked through paraffin oil then flash-frozen in liquid N_2.

To obtain the structure of (*R*)-3-((carboxymethyl)amino)butanoic acid (CABA)-Fe(II)-ScoE, crystals that were grown as described above were transferred into 2 μL of a soaking solution containing 205 mM sodium acetate, 100 mM Tris pH 8.5, 24% (w/v) PEG 4,000, 1 mM CABA, and 250 μM Fe(II)Cl_2 and lacking α-KG at 24°C in an anaerobic environment (95% Ar, 5% H_2; Coy Laboratory Products, Inc.).

To obtain a structure of CABA-Fe(II)-ScoE with α-KG, crystals that were grown as described above were transferred into 2 μL of soaking solution containing 205 mM sodium acetate, 100 mM Tris pH 8.5, 24% (w/v) PEG 4,000, 1 mM CABA, 1 mM α-KG, and 250 μM Fe(II)Cl_2 [room temperature] in an anaerobic environment. Crystals were soaked for 24 hours prior to cryo-protection by manual looping and streaking through paraffin oil followed by flash-freezing in liquid N_2. This soaking procedure lead to the structure of CABA-Fe(II)-ScoE with off-site α-KG.

To obtain the structure of CABA-oxovanadium-ScoE, ScoE was co-crystallized with CABA and oxovanadium using the sitting drop vapor diffusion technique at 24°C in aerobic conditions. A 1 μL aliquot of protein solution (8 mg/mL metal-chelated ScoE, 50

mM HEPES pH 8, 100 mM NaCl, 1 mM succinate, 1 mM CABA, and 1 mM vanadium(IV) sulfate was added to 2 μL of precipitant solution (205 mM sodium acetate, 100 mM Tris pH 8.5, 24% (w/v) PEG 4,000). Crystals formed within 1 week and reached a maximum size after approximately 2 weeks. Paraffin oil was used as a cryoprotectant for crystal harvesting. Crystals were manually looped and streaked through paraffin oil then flash-frozen in liquid N_2.

Data collection and processing

Data for structures of CABA-Fe(II)-ScoE; and CABA-Fe(II)-ScoE with off-site α-KG; Fe(II)-ScoE with off-site α-KG were collected at the Advanced Photon Source (Argonne, Illinois, USA) on beam line 24ID-C using a Pilatus 6M pixel array detector at a temperature of 100 K at a wavelength of 0.9791 Å in a single 180° wedge, and were processed in were indexed, integrated, and scaled in XDS[23]. Data for the structure of CABA-oxovanadium-ScoE were collected at the Advanced Photon Source (Argonne, Illinois, USA) on beam line 24ID-E using an Eiger 16M pixel array detector at a temperature of 100 K at a wavelength of 0.9791 Å in a single 360° wedge, and indexed, integrated, and scaled in XDS[23]. Resolution cutoffs were determined based on a combination of R_{merge}, $CC_{1/2}$, I/σ, and data completeness in the highest resolution bin.

Structure Determination and Refinement

All ScoE structures were solved by molecular replacement using Phaser[24] to the full extent of resolution using the previously published 1.8 Å resolution structure of ScoE (PDB 6DCH)[4] with ligands and metals removed. Molecular replacement resulted in a solution with one ScoE protomer within the asymmetric unit for all data sets. Test sets for R_{free} calculations were chosen to be comprised of the same reflections as were used for refinement of the previously published crystal structure of ScoE for all data sets. Initial rounds of refinement included simulated annealing to reduce model bias from molecular replacement. Iterative rounds of model building were performed in Coot[25], followed by positional refinement and individual *B*-factor refinement performed in Phenix[26]. Mononuclear Fe(II) and oxovanadium were modeled in early stages of

refinement based on local environment, coordination geometry and positive $F_O - F_C$ difference density. Restraint files for CABA and α-KG were generated using the Grade web-server[27]. Positive $F_O - F_C$ difference density was used to assist modeling of CABA, α-KG, chloride, and oxovanadium. Density for α-KG was not of good quality in any structure. When density was present that could not be explained by presence of water molecules or other compounds in the crystallization buffer, α-KG was modeled and refined. In no case did the refinement indicate full occupancy for α-KG, as strong negative $F_O - F_C$ difference density would appear when α-KG was refined at full occupancy. Thus α-KG was refined at partial occupancy in one conformation with extra positive $F_O - F_C$ difference density modeled as water molecules at partial occupancy, although this positive difference density might also indicate alternative α-KG positions. Regardless, we believe that lattice contacts prevent α-KG from assuming its relevant position in our structures and that each α-KG are bound in an off-site position (see main text), which would explain the poor density. The final occupancies of these molecules were calculated by fixing the B-factors to those of atoms in the surrounding environment followed by refinement of partial occupancies. Remaining water molecules were added manually during advanced stages of refinement. Final models for all structures were confirmed with simulated annealing composite omit maps generated in Phenix. All structures contain residues 28-323 (of 326 residues). All software used for refinement was compiled by SBGrid. Final refinement statistics can be found in Table 2. All figures were made using the PyMOL molecular graphics system version 2.0.7 (Schrodinger, LLC).

Computational Analysis

Protein structure and preparation. The present work's crystal structure of ScoE with Fe, α–KG and CABA was modified to model the Fe(IV)-oxo intermediate with succinate bound using Avogadro[28] and PyMOL[29] (structure labeled CS1). This structure represents the outwards orientation of Arg157 and His299, in which these residues are pointing away from succinate's tail. The charge state of ScoE was assigned using the H++ webserver[30-32] assuming a pH of 7.0 with all other defaults applied. After manual

charge assignment of residues adjacent to cofactors/substrates, ScoE has a net charge of -15. Two further structures (CS2 and CS3) were prepared by aligning CS1 to another crystal structure of ScoE (PDB: 6L6X[33]) in PyMOL, the latter representing Arg157 and His299 in inward state, pointing towards the succinate tail. For CS2, the coordinates of Arg157 from the crystal structure (PDB: 6L6X) were aligned to CS1. For CS3, the coordinates of both Arg157 and His299 from the latter structure (PDB: 6L6X) were copied onto CS1. Consequently, CS1 represents both Arg157 and His299 in outward state, CS2 represents Arg157 in inward and His299 in outward state, and CS3 represents both Arg157 and His299 in inward state (**Fig. S13**).

To obtain further different configurations of Arg157, PyMOL's mutagenesis tool was employed to generate different rotamers of Arg157 residue, starting with CS1. PyMOL found 22 backbone dependent and 81 backbone independent rotamers for Arg157 (and none for His299). Based on visual inspection and closest distance between the tail of succinate and sidechain of Arg157, three backbone dependent (BD6, BD20, BD21) and six backbone independent (ID16, ID21, ID29, ID38, ID39 and ID52) rotamers were selected for collecting dynamics (**Table S2**).

For non-standard residues (succinate and CABA), we use the generalized AMBER force field (GAFF)[34] with restrained electrostatic potential (RESP)[35] charges obtained at the Hartree-Fock/6-31G*[36] level using Gaussian/16[37]. AMBER's Python utility for the Metal Center Parameter Builder (MCPB.py)[38] was used to obtain force field parameters to describe the metal active site. The QM geometry optimization, force constant calculation and RESP charge calculation needed for MCPB.py were performed using Gaussian/16[37] with functional UB3LYP[39-41] and basis set LANL2DZ effective core potential[42] on Fe and 6-31G*[36] for the remaining atoms.

The protein structures were solvated in a periodic rectangular prism box with at least a 10 Å buffer of TIP3P[43] water and neutralized with 15 Na^+ counterions for a total simulation of 42,923 atoms (4,741 protein/substrate atoms).

MM Equilibration and Dynamics. The structures were equilibrated with MM MD using the GPU-accelerated PMEMD code in AMBER/18[44]. Equilibration steps were: i) restrained (1000 steps) and unrestrained (2000 steps) minimizations, ii) 10-ps NVT

heating to 300 K with a Langevin thermostat with collision frequency of 5.0 ps^{-1} and a random seed, and iii) 1-ns NpT equilibration using the Berendsen barostat with a pressure relaxation time of 2 ps. Production dynamics were collected for at least 250 ns for each structure (**Table S2**). The SHAKE algorithm[45] was applied with a 2-fs timestep for all MD, and the particle mesh Ewald method was used for long range electrostatics with a 10-Å electrostatic cutoff.

Analysis of MD Trajectories. Snapshots spaced 1 ns apart were analyzed for all configurations using the cpptraj utility in AMBER/18[44]. For each trajectory, four distances were computed between the two oxygen atoms of succinate's COO^- tail, and the two nitrogen atoms on side chains of Arg157 (the terminal NH_2 groups) and His299 (the nitrogen atoms on sidechain ring). The minimum of the former four distances was labeled d_{R157} and the minimum of the latter four distances was labeled d_{H299} (**Fig. S14 and Fig. S15**).

Acknowledgements
We thank David Born and Nick Harris for initial guidance with this project. This research was supported by National Institutes of Health Grants R35 GM126982 (CLD) and NIH Pre-Doctoral Training Grant T32GM007287 (RJ). C.L.D. is a Howard Hughes Medical Institute Investigator. This research was also financially supported by grants to W.Z. from the NIH (R01GM136758) and the Chan Zuckerberg Biohub Investigator Program. This research was also financially supported by grants to HJK from the NSF (CBET-1704266) and a Burroughs Wellcome Fund Career Award at the Scientific Interface. ADRF is funded by the UC Berkeley Chancellor's Fellowship. We would like to thank Tzu-Yang Huang and Bryan McCloskey at UC Berkeley for their assistance with initial GC-MS experiments and consultation in designing experiments. We want to express our gratitude to Judith Klinman for helpful discussions concerning non-heme iron(II) oxidases and usage of the Oxygraph Hansatech probe. We would like to also thank Dr. Miao Zhang from the UC Berkeley Catalysis Center for training with the GC-MS apparatus and providing assistance with data interpretation. The GC-MS work was made possible by the Catalysis Facility of Lawrence Berkeley National Laboratory, supported by the Director, Office of Science, of the U.S. Department of Energy (contract no. DE-AC02-05CH11231. This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P30 GM124165). The Pilatus 6M detector on 24-ID-C beam line is funded by a NIH-ORIP HEI grant (S10 RR029205). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-

References

1. Clarke-Pearson, M.F. and S.F. Brady, *Paerucumarin, a new metabolite produced by the pvc gene cluster from Pseudomonas aeruginosa.* J Bacteriol, 2008. **190**(20): p. 6927-30.
2. Brady, S.F. and J. Clardy, *Cloning and heterologous expression of isocyanide biosynthetic genes from environmental DNA.* Angew Chem Int Ed Engl, 2005. **44**(43): p. 7063-5.
3. Harris, N.C., M. Sato, N.A. Herman, F. Twigg, W. Cai, J. Liu, X. Zhu, J. Downey, R. Khalaf, J. Martin, H. Koshino, and W. Zhang, *Biosynthesis of isonitrile lipopeptides by conserved nonribosomal peptide synthetase gene clusters in Actinobacteria.* Proceedings of the National Academy of Sciences, 2017. **114**(27): p. 7025-7030.
4. Harris, N.C., D.A. Born, W. Cai, Y. Huang, J. Martin, R. Khalaf, C.L. Drennan, and W. Zhang, *Isonitrile Formation by a Non-Heme Iron(II)-Dependent Oxidase/Decarboxylase.* Angew Chem Int Ed Engl, 2018. **57**(31): p. 9707-9710.
5. Martinez, S. and R.P. Hausinger, *Catalytic Mechanisms of Fe(II)- and 2-Oxoglutarate-dependent Oxygenases.* J Biol Chem, 2015. **290**(34): p. 20702-11.
6. Hausinger, R.P., *FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes.* Crit Rev Biochem Mol Biol, 2004. **39**(1): p. 21-68.
7. Hanauske-Abel, H.M. and V. Gunzler, *A stereochemical concept for the catalytic mechanism of prolylhydroxylase: applicability to classification and design of inhibitors.* J Theor Biol, 1982. **94**(2): p. 421-55.
8. Muthukumaran, R.B., P.K. Grzyska, R.P. Hausinger, and J. McCracken, *Probing the iron-substrate orientation for taurine/alpha-ketoglutarate dioxygenase using deuterium electron spin echo envelope modulation spectroscopy.* Biochemistry, 2007. **46**(20): p. 5951-9.
9. Price, J.C., E.W. Barr, B. Tirupati, J.M. Bollinger, Jr., and C. Krebs, *The first direct characterization of a high-valent iron intermediate in the reaction of an alpha-ketoglutarate-dependent dioxygenase: a high-spin FeIV complex in taurine/alpha-ketoglutarate dioxygenase (TauD) from Escherichia coli.* Biochemistry, 2003. **42**(24): p. 7497-508.
10. TUDERMAN, L., R. MYLLYLÄ, and K.I. KIVIRIKKO, *Mechanism of the Prolyl Hydroxylase Reaction.* European Journal of Biochemistry, 1977. **80**(2): p. 341-348.
11. Price, J.C., E.W. Barr, T.E. Glass, C. Krebs, and J.M. Bollinger, Jr., *Evidence for hydrogen abstraction from C1 of taurine by the high-spin Fe(IV) intermediate detected during oxygen activation by taurine:alpha-ketoglutarate dioxygenase (TauD).* J Am Chem Soc, 2003. **125**(43): p. 13008-9.
12. Bollinger Jr., J.M., J.C. Price, L.M. Hoffart, E.W. Barr, and C. Krebs, *Mechanism of Taurine: α-Ketoglutarate Dioxygenase (TauD) from Escherichia coli.* European Journal of Inorganic Chemistry, 2005. **2005**(21): p. 4245-4254.
13. Chen, T.Y., J. Chen, Y. Tang, J. Zhou, Y. Guo, and W.C. Chang, *Pathway from N-alkylglycine to Alkylisonitrile Catalyzed by Iron(II) and 2-Oxoglutarate-Dependent Oxygenases.* Angew Chem Int Ed Engl, 2020.

14. Huang, Y.-B., W. Cai, A. Del Rio Flores, F.F. Twigg, and W. Zhang, *Facile Discovery and Quantification of Isonitrile Natural Products via Tetrazine-Based Click Reactions.* Analytical Chemistry, 2020. **92**(1): p. 599-602.
15. Warui, D.M., N. Li, H. Norgaard, C. Krebs, J.M. Bollinger, Jr., and S.J. Booker, *Detection of formate, rather than carbon monoxide, as the stoichiometric coproduct in conversion of fatty aldehydes to alkanes by a cyanobacterial aldehyde decarbonylase.* J Am Chem Soc, 2011. **133**(10): p. 3316-9.
16. Rui, Z., X. Li, X. Zhu, J. Liu, B. Domigan, I. Barr, J.H. Cate, and W. Zhang, *Microbial biosynthesis of medium-chain 1-alkenes by a nonheme iron oxidase.* Proc Natl Acad Sci U S A, 2014. **111**(51): p. 18237-42.
17. O'Brien, J.R., D.J. Schuller, V.S. Yang, B.D. Dillard, and W.N. Lanzilotta, *Substrate-induced conformational changes in Escherichia coli taurine/alpha-ketoglutarate dioxygenase and insight into the oligomeric structure.* Biochemistry, 2003. **42**(19): p. 5547-54.
18. Trewick, S.C., T.F. Henshaw, R.P. Hausinger, T. Lindahl, and B. Sedgwick, *Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage.* Nature, 2002. **419**(6903): p. 174-178.
19. Myllylä, R., K. Majamaa, V. Günzler, H.M. Hanauske-Abel, and K.I. Kivirikko, *Ascorbate is consumed stoichiometrically in the uncoupled reactions catalyzed by prolyl 4-hydroxylase and lysyl hydroxylase.* Journal of Biological Chemistry, 1984. **259**(9): p. 5403-5.
20. Mitchell, A.J., Q. Zhu, A.O. Maggiolo, N.R. Ananth, M.L. Hillwig, X. Liu, and A.K. Boal, *Structural basis for halogenation by iron- and 2-oxo-glutarate-dependent enzyme WelO5.* Nat Chem Biol, 2016. **12**(8): p. 636-40.
21. Martinie, R.J., C.J. Pollock, M.L. Matthews, J.M. Bollinger, Jr., C. Krebs, and A. Silakov, *Vanadyl as a Stable Structural Mimic of Reactive Ferryl Intermediates in Mononuclear Nonheme-Iron Enzymes.* Inorg Chem, 2017. **56**(21): p. 13382-13389.
22. Davis, K.M., M. Altmyer, R.J. Martinie, I. Schaperdoth, C. Krebs, J.M. Bollinger, Jr., and A.K. Boal, *Structure of a Ferryl Mimic in the Archetypal Iron(II)- and 2-(Oxo)-glutarate-Dependent Dioxygenase, TauD.* Biochemistry, 2019. **58**(41): p. 4218-4223.
23. Kabsch, W., *XDS.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.
24. McCoy, A.J., R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, and R.J. Read, *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
25. Emsley, P., B. Lohkamp, W.G. Scott, and K. Cowtan, *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
26. Adams, P.D., P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, and P.H. Zwart, *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
27. Smart, O.S.W., T.O., *No Title.* Global Phasing Ltd. , 2011.
28. Hanwell, M.D., D.E. Curtis, D.C. Lonie, T. Vandermeersch, E. Zurek, and G.R. Hutchison, *Avogadro: an advanced semantic chemical editor, visualization, and analysis platform.* Journal of cheminformatics, 2012. **4**(1): p. 17.
29. Schrodinger, L.L.C., *The PyMOL Molecular Graphics System, Version 1.7.4.3*. 2010.

30. Anandakrishnan, R., B. Aguilar, and A.V. Onufriev, *H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations.* Nucleic Acids Research, 2012. **40**(W1): p. W537-W541.
31. Gordon, J.C., J.B. Myers, T. Folta, V. Shoja, L.S. Heath, and A. Onufriev, *H++: a server for estimating pKas and adding missing hydrogens to macromolecules.* Nucleic Acids Research, 2005. **33**(suppl 2): p. W368-W371.
32. Myers, J., G. Grothaus, S. Narayanan, and A. Onufriev, *A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules.* Proteins: Structure, Function, and Bioinformatics, 2006. **63**(4): p. 928-938.
33. Chen, T.Y., J. Chen, Y. Tang, J. Zhou, Y. Guo, and W.c. Chang, *Pathway from N‐Alkylglycine to Alkylisonitrile Catalyzed by Iron (II) and 2‐Oxoglutarate‐Dependent Oxygenases.* Angewandte Chemie, 2020. **132**(19): p. 7437-7441.
34. Wang, J., R.M. Wolf, J.W. Caldwell, P.A. Kollman, and D.A. Case, *Development and testing of a general amber force field.* Journal of Computational Chemistry, 2004. **25**(9): p. 1157-1174.
35. Bayly, C.I., P. Cieplak, W. Cornell, and P.A. Kollman, *A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model.* The Journal of Physical Chemistry, 1993. **97**(40): p. 10269-10280.
36. Harihara, P.C. and J.A. Pople, *Influence of Polarization Functions on Molecular-Orbital Hydrogenation Energies.* Theoretica Chimica Acta, 1973. **28**(3): p. 213-222.
37. Frisch, M.J., G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A.V. Marenich, J. Bloino, B.G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, J.V. Ortiz, A.F. Izmaylov, J.L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V.G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M.J. Bearpark, J.J. Heyd, E.N. Brothers, K.N. Kudin, V.N. Staroverov, T.A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A.P. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, J.M. Millam, M. Klene, C. Adamo, R. Cammi, J.W. Ochterski, R.L. Martin, K. Morokuma, O. Farkas, J.B. Foresman, and D.J. Fox, *Gaussian 16 Rev. C.01*. 2016: Wallingford, CT.
38. Li, P. and K.M. Merz Jr, *MCPB. py: A python based metal center parameter builder*. 2016, ACS Publications.
39. Stephens, P.J., F.J. Devlin, C.F. Chabalowski, and M.J. Frisch, *Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields.* The Journal of Physical Chemistry, 1994. **98**(45): p. 11623-11627.
40. Lee, C., W. Yang, and R.G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density.* Physical Review B, 1988. **37**: p. 785--789.
41. Becke, A.D., *Density-functional thermochemistry. III. The role of exact exchange.* Journal of Chemical Physics, 1993. **98**(7): p. 5648-5652.
42. Hay, P.J. and W.R. Wadt, *Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg.* The Journal of Chemical Physics, 1985. **82**(1): p. 270-283.

43. Jorgensen, W.L., J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, *Comparison of simple potential functions for simulating liquid water.* The Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
44. Case, D., I. Ben-Shalom, S. Brozell, D. Cerutti, T. Cheatham III, V. Cruzeiro, T. Darden, R. Duke, D. Ghoreishi, and M. Gilson, *AMBER 2018; 2018.* University of California, San Francisco.
45. Ryckaert, J.-P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.* Journal of Computational Physics, 1977. **23**(3): p. 327-341.