



**HAL**  
open science

# Structural Studies based on two Lysine Dioxygenases with Distinct Regioselectivity Brings Insights Into Enzyme Specificity within the Clavamate Synthase-Like Family

Karine Bastard, Tatiana Isabet, Enrico Stura, Pierre Legrand, Anne Zaparucha

## ► To cite this version:

Karine Bastard, Tatiana Isabet, Enrico Stura, Pierre Legrand, Anne Zaparucha. Structural Studies based on two Lysine Dioxygenases with Distinct Regioselectivity Brings Insights Into Enzyme Specificity within the Clavamate Synthase-Like Family. *Scientific Reports*, 2018, 8 (1), 10.1038/s41598-018-34795-9. cea-02099557

**HAL Id: cea-02099557**

**<https://cea.hal.science/cea-02099557>**

Submitted on 15 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# SCIENTIFIC REPORTS



OPEN

## Structural Studies based on two Lysine Dioxygenases with Distinct Regioselectivity Brings Insights Into Enzyme Specificity within the Clavamate Synthase-Like Family

Karine Bastard<sup>1</sup>, Tatiana Isabet<sup>2</sup>, Enrico A. Stura<sup>3</sup>, Pierre Legrand<sup>2</sup> & Anne Zaparucha<sup>1</sup>

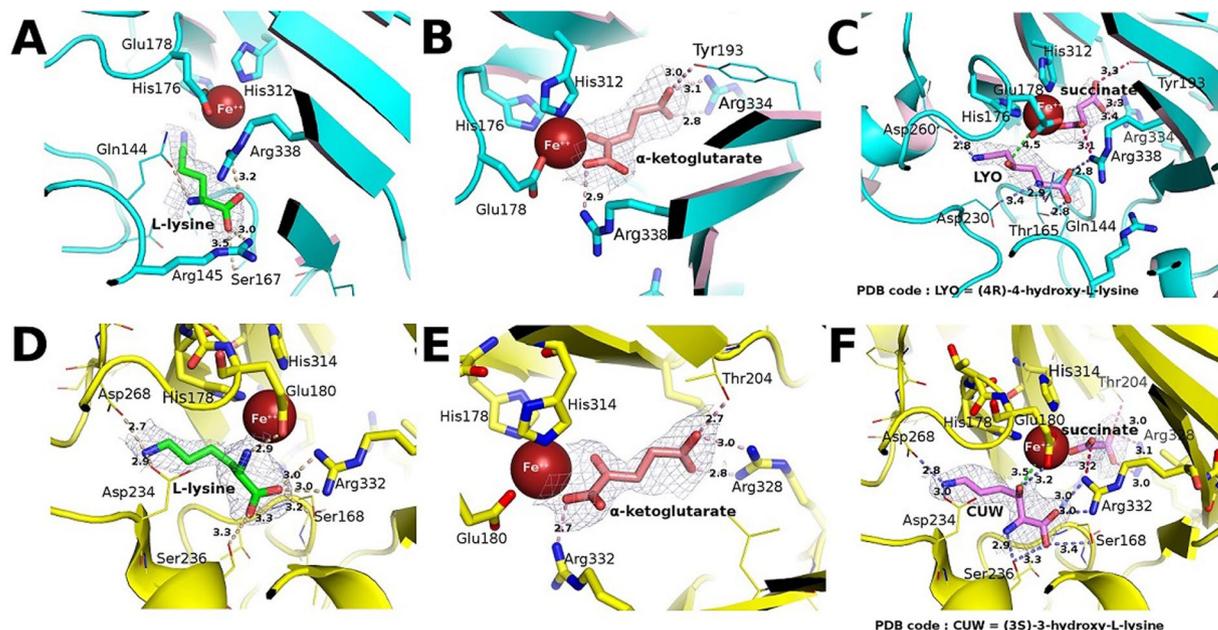
Iron(II)/ $\alpha$ -ketoacid-dependent oxygenases ( $\alpha$ KAOs) are enzymes that catalyze the oxidation of unactivated C-H bonds, mainly through hydroxylation. Among these, those that are active towards amino-acids and their derivatives are grouped in the Clavamate Synthase Like (CSL) family. CSL enzymes exhibit high regio- and stereoselectivities with strict substrate specificity. This study reports the structural elucidation of two new regiodivergent members, KDO1 and KDO5, active towards lysine, and the structural and computational analysis of the whole family through modelling and classification of active sites. The structures of KDO1 and KDO5 in complex with their ligands show that one exact position in the active site controls the regioselectivity of the reaction. Our results suggest that the substrate specificity and high stereoselectivity typical of this family is linked to a lid that closes up in order to form a sub-pocket around the side chain of the substrate. This dynamic lid is found throughout the family with varying sequence and length and is associated with a conserved stable dimeric interface. Results from this study could be a starting-point for exploring the functional diversity of the CSL family and direct *in vitro* screening in the search for new enzymatic activities.

Enzymes derived from microorganisms are particularly appealing for the design of chemical processes that are environmentally friendly and sustainable<sup>1</sup>. Enzymes catalyze interconversions of many functional groups with well-defined chemo-, regio- and stereoselectivities, resulting in high atom economy and reduced waste during synthesis. Over the past 25 years, the tremendous progress in molecular biology, bioinformatics and process optimization have lowered the biotechnological barriers that were limiting the industrial use of enzymes<sup>2-4</sup>. Much research has looked for novel enzymes able to catalyze challenging reactions such as the functionalization of non-activated (i.e. not alpha to C-C or C-heteroatom double bonds) C-H bonds. Such reactions still resist most efforts made by organic chemists. However, enzymes that can catalyze the formation of various C-heteroatom bonds have been found<sup>5,6</sup>. We have focused our work on the iron(II)/ $\alpha$ -KetoAcid-dependent Oxygenases ( $\alpha$ KAOs), a mechanistically diverse superfamily of enzymes that use dioxygen as the oxidant to catalyze several reactions via C-H bond activation, mostly hydroxylation<sup>7-9</sup>. These enzymes operate with three substrates, dioxygen,  $\alpha$ -ketoglutarate ( $\alpha$ -KG) and a primary substrate. One oxygen atom is transferred to a primary substrate to give an oxidized product, and the second oxygen is given to the  $\alpha$ -KG, which upon oxidation, forms succinate and CO<sub>2</sub><sup>10,11,12</sup>.

Many  $\alpha$ KAOs are tailoring enzymes involved in secondary metabolite biosynthesis. Such enzymes, often found in bacteria, have been reported to hydroxylate the side chains of free amino acids and derivatives, or tether peptides in non-ribosomal peptide biosynthesis<sup>13</sup>. Since hydroxylated amino acids contain several stereogenic centers, they are valuable chiral building blocks for fine chemical synthesis, which justifies the intensive effort

<sup>1</sup>Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057, Evry, France. <sup>2</sup>Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin, BP 48, 91192, Gif-sur-Yvette, France. <sup>3</sup>CEA, Institut des Sciences du Vivant Frédéric Joliot, Service d'Ingénierie Moléculaire des Protéines (SIMOPRO), Université Paris-Saclay, Gif-sur-Yvette, 91190, France. Karine Bastard and Tatiana Isabet contributed equally. Correspondence and requests for materials should be addressed to A.Z. (email: [anne.zaparucha@genoscope.cns.fr](mailto:anne.zaparucha@genoscope.cns.fr))





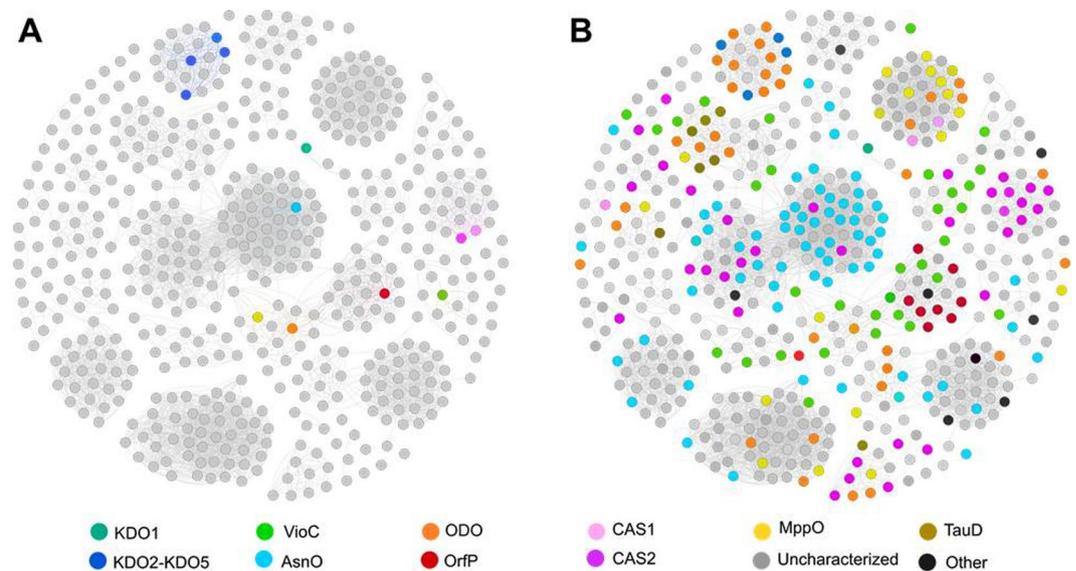
**Figure 2.** Comparison of the binding sites of KDO5 and KDO1, with the  $2F_o - F_c$  electron density (contoured at  $0.5 \sigma$ ) of the bound substrates and products. (A) KDO5 in complex with L-lysine (monomer C). (B) KDO5 in complex with  $\alpha$ -KG (monomer A). (C) KDO5 in complex with (4R)-4-hydroxy-L-lysine and succinate (monomer D). (D) KDO1 in complex with L-lysine (monomer C). (E) KDO1 in complex with  $\alpha$ -KG (monomer B). (F) KDO1 in complex with (3S)-3-hydroxy-L-lysine and succinate (monomer B). Among the four molecules in the asymmetric unit, only the binding site with the best positioning in the electron density for the ligands has been illustrated. Identifying codes assigned for the small ligand compounds are indicated on Panels B and E.

iron ion is present in the active site chelated by two His (176 and 312) and one Glu (178). In KDO5-lysine crystal structure within the binding site of monomer C (Fig. 2A), the carboxylate of L-lysine substrate is stabilized by salt bridges with the Arg338 (3.2 Å) and 145 (3.0 Å) in the active site, and positioned by one H-bond with Ser167 (3.5 Å). In  $\alpha$ -KG bound structure,  $\alpha$ -KG is best defined for monomer A (Fig. 2B). The carboxylate moieties of the co-substrate form salt bridges with Arg338 and 334 (2.9 Å; 2.8 Å and 3.1 Å); the  $\alpha$ -ketoacid moiety binds the iron as bidentate; the terminal carboxylate interacts by H-bond with Tyr193 (3.0 Å). In KDO5, the ligands in the product complex are less well defined, reflecting their ability to escape from the active site at the completion of the catalytic event. The better defined products are located within monomer D. The (4R)-4-hydroxy-L-lysine is stabilized by salt bridges with Arg338 (2.8 Å), Asp230 (3.4 Å) and Asp260 (2.8 Å), and makes H-bonds with Gln144 (2.8 Å), Thr165 (2.8 Å). The C4-OH group is close to the iron (4.5 Å) (Fig. 2C). The succinate resultant from the oxidative decarboxylation of the  $\alpha$ -KG, makes H-bond Tyr193 (3.3 Å), and establishes salt bridges with Arg338 (3.1 Å), Arg334 (3.3 and 3.4 Å) and with the iron (Fig. 2C).

The binding sites are different between KDO5 and KDO1. In KDO5 the pocket is characterized by a neck defined by four residues, namely, Trp112, Arg145, Asn232 and Asp230. Of these, Arg145 is instrumental in the recognition of L-lysine, as it is able to attract the substrate towards the end of the pocket where it can be maintained during the reaction (Fig. 2A).

While a ferrous ion is present in the KDO5 apo structure, it is not the case in KDO1 apo as the metal is lost during the purification from the KDO1 preparation used to grow the crystals. After crystal soaking with  $\text{FeSO}_4$  the two His residues (178 and 314) and Glu180 in the KDO1 active site recapture an iron. The entrance of the catalytic pocket of KDO1 is bifurcated, and the main opening does not match the one of KDO5. In KDO1, the L-lysine complex shows a well ordered lysine in monomer C (Fig. 2D). The interacting residues are Glu180, Arg332, Asp 234, Asp 268 (salt bridges 2.9 Å; 3.0 Å and 3.0 Å; 2.9 Å; 2.7 Å), Ser168 and Ser236 (H-bonds 3.2 Å and 3.3 Å; 3.3 Å). The  $\text{N}_\epsilon$  is restrained at the bottom of the pocket by two salt bridges with the  $\text{O}\delta_2$  of Asp268 and the  $\text{O}\delta_1$  of Asp234, while in KDO5 the  $\text{N}_\epsilon$  is stabilized *via* a water molecule in interaction with Asp260. The size of the KDO1 pocket is smaller than that of KDO5 due to Ser168 (instead of Gly). In the  $\alpha$ -KG crystals, the structure shows a well ordered  $\alpha$ -KG in monomer B. The  $\alpha$ -KG makes salt bridges with Arg332 (2.7 Å), Arg328 (2.8 Å and 3.0 Å) and H-bond with Thr204 (2.7 Å), while its  $\alpha$ -ketoacid moiety binds the iron as bidentate (Fig. 2E). For KDO1, the best product is located within the active site of monomer B (Fig. 2F). The (3S)-3-hydroxy-L-lysine makes salt bridges with Glu180 (3.2 Å), Arg332 (3.0 Å and 3.0 Å), Asp234 (3.0 Å) and Asp268 (2.8 Å), and H-bonds with Ser168 (3.4 Å) and Ser236 (2.9 Å and 3.3 Å). The C3-OH is located close to the iron (3.5 Å). The succinate makes H-bond with Thr204 (3.0 Å), and salt bridges with Arg332 (3.2 Å), Arg328 (3.0 and 3.1 Å) and the iron (Fig. 2F).

In comparing the product complexes of KDO1 and KDO5, we found that the electron densities of the ligands are generally better for KDO1, except for monomer D of KDO5. The hydroxyllysines overlap is slightly off by 3 Å, as would be expected given that different positions are hydroxylated (Fig. S4). In KDO1, the  $\text{N}_\epsilon$  of hydroxyllysine



**Figure 3.** Sequence Similarity Network (SSN) of the Clavaminate synthase-like family (IPR014503 InterPro family) from the  $\alpha$ KAOs superfamily. Nodes represent proteins and edges are the links between those proteins that have at least 50% sequence identity. **(A)** Experimental data retrieved from Swissprot mapped on the SSN. KDO5 and other KDO2-4, characterized previously<sup>16</sup> are colored in dark blue. **(B)** Annotations from Uniprot mapped on the SSN. The network is visualized with Gephi (Bastian, “Gephi: An Open Source Software for Exploring and Manipulating Networks”, AAAI Publications, Third International AAAI Conference on Weblogs and Social Media, 2009).

is better stabilized than in KDO5, because the two H-bonds (Asp234 and Asp268) (Fig. 2E) stabilize the substrate after interacting with a water molecule (H<sub>2</sub>O-562: Table S5). In the case of KDO5, the N $\alpha$  of the product does not maintain the interaction with the equivalent water molecule as the peptide amide becomes positioned where the water molecule was located, deeper in the pocket. The position of the active site residues in both KDO1 and KDO5 remains unchanged between the substrate and product complexes.

**Crystallographic considerations.** Although the resolution of the KDO5 and KDO1 structures in complex with the lysine substrate are comparatively high, the electron density for the substrates is not well defined. For both KDO1 and KDO5, in the presence of  $\alpha$ -KG and lysine, the reaction takes place and it was therefore impossible to obtain crystals with both substrates simultaneously. However, we could obtain crystals in the presence of the two products resulting from the reaction. In the absence of the  $\alpha$ -KG, unlike what is observed for the lysine-KDO1 complexes, the lysine is positioned differently in the four molecules of lysine-KDO5, suggesting that this substrate is not properly stabilized. The flexibility of the N-terminus of KDO5 might have a role in the access to the active site. The first residue that can be positioned in any KDO5 structure is Ala13 in monomer C of the apo form; for KDO1 the first ordered residue is Glu30 in monomers A and D of the apo form.

**Relationship between members of the Clavaminate synthase-like family.** According to our knowledge,  $\alpha$ KAOs encompass 11 phylogenetically unrelated enzyme families according to the InterPro database<sup>20</sup> (Table S6). Enzymes that transform free L-amino acids, such as KDO1 and KDO5, are found in four different families represented by enzymes called VioC, MpDO, SadA and GriE (Table S6)<sup>17,21–24</sup>. Because they possess different folds (Table S6), alignment of their whole structures is not possible. However, a structural alignment of the 2-His-1-carboxylate facial triad of their structures is possible and can help to define a spatial arrangement of functional groups (Fig. S5).

KDO1 and KDO5 belong to the Clavaminate synthase-like family (CSL), which is classified in the IPR014503 InterPro family. In order to analyze the sequence diversity and the relationships between proteins of this family, we performed a sequence similarity network (SSN) using the EFI-EST<sup>25</sup>. In SSNs, the most related proteins are grouped together in clusters. The network reveals five highly populated sub-families for which there was no reported experimental data (Fig. 3A). Mapping the annotation retrieved from Uniprot shows mis-annotation for the family as close homologs were assigned different functions or remote homologs are assigned identical functions (Fig. 3B). For instance, the sub-family highly connected to the sub-family containing AsnO, brings together CAS annotated proteins. KDO5’s subfamily is fully disconnected from other CSL dioxygenases, showing an early evolution away from them. However, sequences in the KDO5 sub-family share on average 70% identity and are found in two different taxonomic classes (data not shown), indicating that, although these sequences have evolved, the proteins maintain the same enzymatic activity<sup>16,26</sup>.

**Active site profiles within CSL family.** Enzymes catalyzing the same chemical transformation while sharing low sequence identity, present identical residues in their active site as evolutionary pressure tends to conserve the residues playing a critical role in the catalysis<sup>27</sup>. Because loop or subdomain insertions can disrupt sequence

alignment, these conserved residues are better identified in protein structures. Previously, it has been shown that, in general, substrate specificity is linked to the presence of a few distinctive amino acids in the active site<sup>28</sup>.

To understand what makes CSL dioxygenases specific to a range of substrates, we carried out an analysis of the active sites of all its members. Homology modeling and subsequent classification of active sites was performed using ASMC (Active Site Modeling and Clustering) software<sup>29</sup>. The resulted “Active site tree” was divided in 12 groups (referred to as G1 to G12), each containing between 15 and 91 sequences. The three-dimensional superposition of modeled and crystallographic-resolved active sites within each group was projected linearly to form conservation patterns called logos (Figs 4A and S6). Thus, 3D-motifs found in the active sites can be easily read on the logos. All logos have the conserved triade His, Glu/Asp, His (respectively in position P8, P9 and P17 of the logo) for iron binding. Arg in P18 is highly conserved and forms a salt bridge with the carboxylate group of  $\alpha$ -KG/succinate. Two other residues are very conserved: Leu in P7 and Arg in P19, which delimit two distinct pockets that host  $\alpha$ -KG/succinate and the substrate/product (Fig. S7). Arg in P19, conserved in all ASMC groups, interacts with the carboxylate group of the substrates. A large majority of the groups have a glutamine or a glutamate in P2. These residues (Gln in KDO5, VioC, OrfP, ODO, or Glu for AsnO, and KDO1) interact with the  $\alpha$ -NH<sub>2</sub> of the amino acid substrates. Only a few exceptions are noted in the family, like CAS 1 from G7 (Leu in P2), where proclavaminate substrate is found deeply in the pocket, or proteins from G10.

**C4 regio-selectivity in KDO5 and other CSL dioxygenases.** The logo of KDO5's group (*i.e.* G12) is highly conserved and shows an arginine in position P3 which corresponds to Arg145 (Fig. 4A). As observed in crystallographic snapshots, it interacts through a salt bridge with the carboxylate group of the L-lysine substrate and might be responsible for the regioselectivity of the reaction (Fig. 4B). Indeed, this residue, absent in KDO1 (C3-regioselective) (Fig. 4B), seems to be a key residue for positioning the substrate deeply in the pocket and bringing the C4 close to the iron, resulting in a C4-regioselectivity. Force constants have been calculated on crystallographic structures of different members of the CSL family (VioC, AsnO, CAS, OrfP, KDO1 and KDO5)<sup>17,19,30–32</sup>. Our analysis revealed that Arg145, even when not situated on a rigid secondary element, is part of this very rigid core, underlying its high importance for KDO5 activity (Fig. S8).

Arg in P3 of the logo is also found in CAS1 (G7) crystallographic structure<sup>18</sup>. The side chain of this residue (Arg115) points toward the carboxylate of the substrate, and might be involved in binding with the substrate *via* a water molecule<sup>31</sup>. As G12 group, CSL dioxygenases from G4 also present an Arg in P3 of the logo (Fig. 4B), suggesting that these enzymes might also catalyze the C4-hydroxylation of an amino acid. Arg in P3 of G12 and G4 is part of an electropositive patch that includes the Lys in P11 (G12) and His in P11 (G4) (Fig. 4B).

Interestingly, groups from Clade I (*i.e.* G1, G2, and G3) show a Lys in P3. This Lys does not influence the electrostatic properties of the pocket region that hosts the carboxylate group of the substrate, because this residue is buried by Asn in P6, a residue conserved in this clade (Fig. 4B). Consequently, the type of residue in P3 for proteins of Clade I should have no direct influence on the regioselectivity of the reaction.

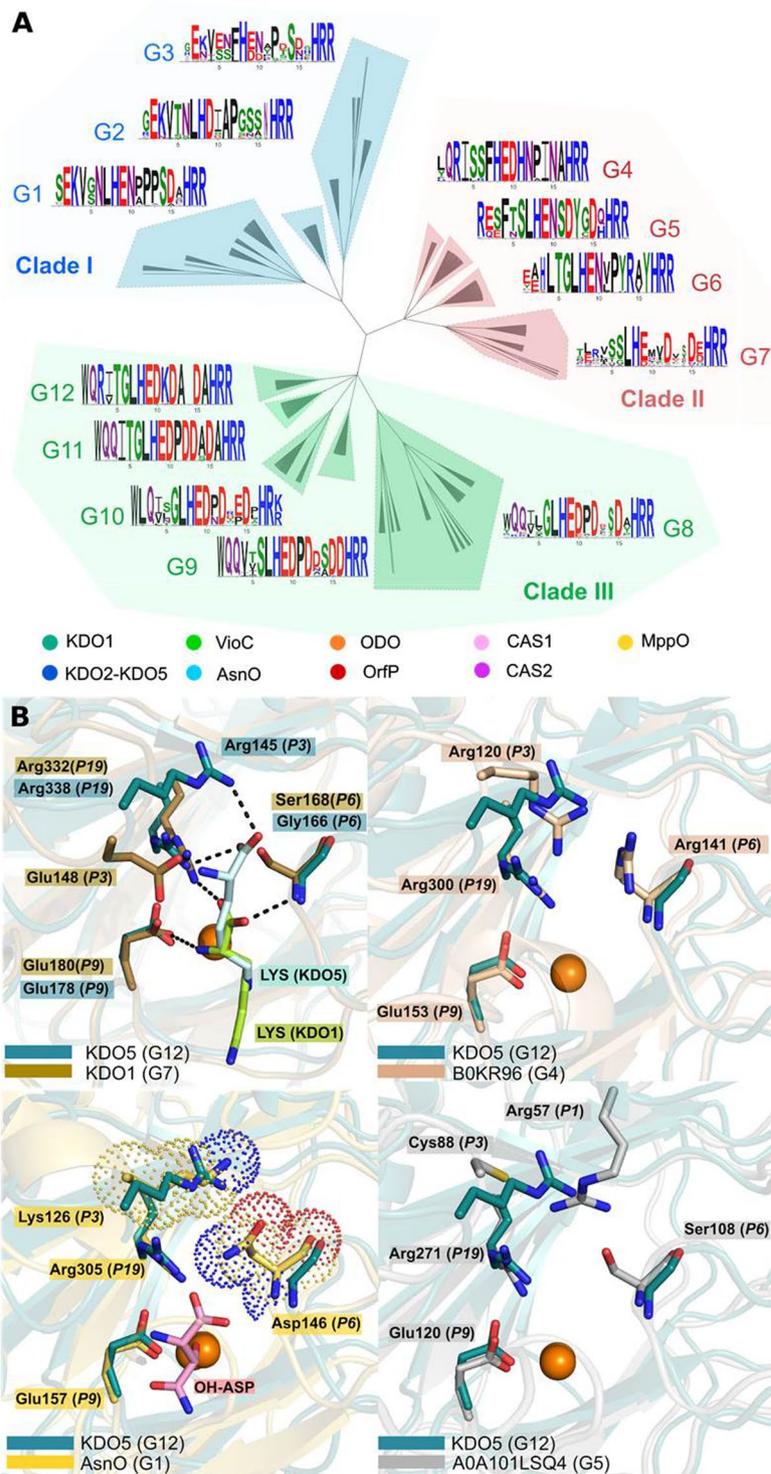
For G5 proteins, the three-dimensional motif formed by Arg in P1 and the catalytic residues (His in P8, His in P8, Glu/Asp in P17) is the same as the one found in the proteins of G4 and G12, where an Arg is found in P3. Even if these two Arg (in P1 of G5 and P3 of G4 and G12) are not found at the same position in sequence, their side chains point in the same direction in 3D (Fig. 4B). Thus, because enzymes from G4 and G12 groups share this common 3D-motif with G5, it is expected that they will have the same regioselectivity towards the substrate.

Taxonomic analysis of ASMC groups shows that enzymes predicted to hydroxylate in C4 (*i.e.* G4, G5 and G12 groups), are found in restricted taxonomic branches (Fig. S9) and might have appeared relatively recently (*i.e.* Proteobacteria). By contrast, the oldest taxonomic levels, such as Actinobacteria phylum, present a great diversity of CSL dioxygenases.

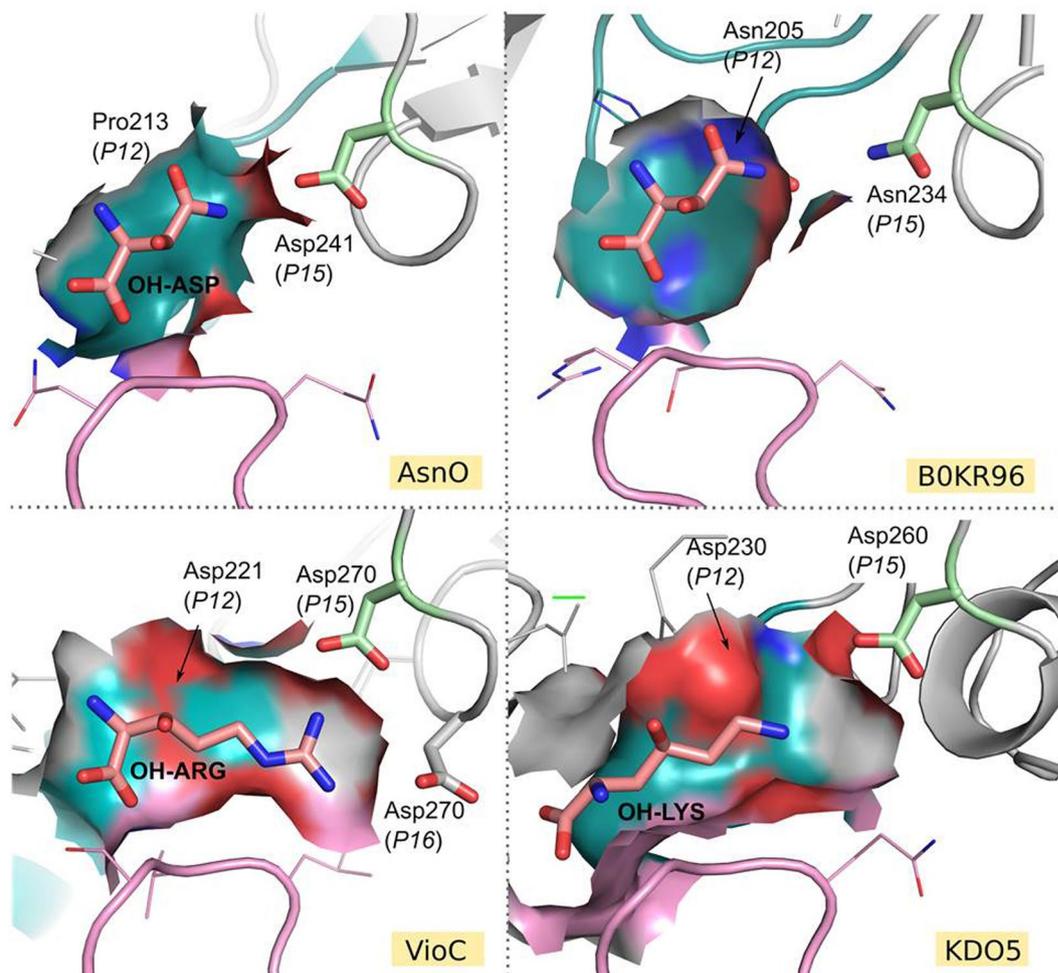
**A sub-pocket for selecting the type of substrate.** A conserved 3D-motif DXDX in P12 to P15 of the logo is found in all groups of Clade III and in two groups of Clade II (G5 and G7). Asp in P12 and Asp in P15, conserved respectively in 55% and 76% in CSL proteins are spatially close, the latter being also very conserved in these same ASMC groups (and in G1). These two Asp form a negatively charged sub-pocket into which the side-chain of the substrate fits, as observed in the crystallographic structures of KDO1, KDO5, VioC, OrfP and CAS (Fig. 5). DXDX is also observed in models of ODO and MppO, both active towards substrates with electropositive side chains. More broadly, it is expected that this motif/sub-pocket is present in CSL dioxygenases active towards substrates with positively charged side chains. In Clade III, a highly conserved Asp in P10 takes part of the electronegative potential of the sub-pocket and interacts through electrostatic interactions with the residues in P3 (*i.e.* Arg in G12 or Gln/Glu in the other groups of Clade III).

Interestingly, G4 and G6 groups do not display this negatively charged sub-pocket, but a hydrophobic (motif NPIN from P12 to P15) and positively charged (motif PYRA/V) sub-pocket respectively (Fig. 5). The consequent assumption is that enzymes from G4 might catalyze the hydroxylation of hydrophobic substrates, while those from G6 might transform amino acids with a negatively charged side chain, such as glutamate or aspartate. The sub-pocket we describe here cannot be found in the crystallographic structures of other  $\alpha$ KAO known to transform hydrophobic substrates or aspartate, such as P3H2, P4H, AspH, all classified in another InterPro family (IPR007803) (Table S6). Indeed, the substrate/product binding pocket of these enzymes are not structured the same way as in CSL enzymes (Fig. S10).

In summary, the sub-pocket disclosed in our models, formed by the residue P12 of the flexible loop and P15, is critical for the selection of the type of substrate. In VioC structure, this sub-pocket, negatively charged, is strengthened by the presence of an additional Asp in position P16, very conserved in the G10 group, interacting with the bidentate guanidinium of the arginine substrate side chain (Fig. 5). In G6, a very conserved Tyr is found in P16, suggesting that this residue might play an important role, like Asp in P16 of VioC, in interacting with a large substrate (sterically equivalent to an arginine). In groups of Clade I, DXDX motif from P12 to



**Figure 4.** Dividing the CSL family into groups with similar active sites. **(A)** ASMC hierarchical tree containing 523 proteins, grouped in 3 clades (I, II, III). For each ASMC group (from G1 to G12), a logo sequence is depicted and represents the conservation of residues in the active site pocket. Proteins with known activities are indicated by a circle in the tree. **(B)** Regioselectivity among the CSL proteins. Superimposition of crystallographic structure of KDO5, chain C, in complex with lysine, with crystallographic structure of KDO1, chain C, in complex with lysine (top left), with a representative of group G4 (model with Uniprot ID B0KR96) (top right), with crystal structure of AsnO in complex with hydroxy-aspartate (PDB code 2OG7) (down left) and with a representative of G5 (model with Uniprot ID A0A101LSQ4) (down right). Residues interacting with the carboxylate and amino group of the substrates are indicated in stick, along with their position in the logo. Black dashed lines indicate interactions between substrates and proteins.



**Figure 5.** Depiction of the sub-pockets in the CSL family. For Clade I, represented by AsnO, the sub-pocket is formed by 3D motif PPSD in *P12* to *P15* of the active site logo. For Clade II, a representative of the G4 group, a model of protein with Uniprot ID B0KR96, was chosen. The sub-pocket is composed of residues NPIN. To get an idea of the size of substrate that can host the protein, Asn from AsnO was positioned at the center of the binding pocket. For Clade III, represented by VioC and KDO5, the sub-pocket is formed by 3D motifs DSDD and DA-D respectively. Residue in *P15* of the logo is represented in green. The flexible lid is colored deep blue teal with atoms colored by elements. The flexible lid occupies a large part of the binding pocket, which is represented in surface. The adjacent loop is colored in pink.

*P15* is absent. However, an Asp is found in *P15* of G1, and partially in G3, which might play a role in substrate positioning, as exemplified by the short distance (3.6 Å) observed in AsnO structure in complex with (2S,3S) hydroxy-asparagine.

**Induced lid motion upon substrate binding.** Previous studies have highlighted that the flexible lid can change its conformation upon substrate binding<sup>19</sup>, be partly disordered when complexed with ligands<sup>18</sup>, or completely disordered in the apo-form, but become ordered when the substrate is bound<sup>17</sup>. It was hypothesized that the lid shields the substrate from bulk solvent access to enable the reaction and is involved in the substrate selection<sup>17–19</sup>. The diversity in terms of sequences, shape and length of this flexible lid among the family could be another element to explain the high substrate specificity of CSL dioxygenases. Models belonging to the G1 group show the same lid length as AsnO, whereas models from the G2 group have a loop without a well-defined structure which is the same length as the KDO5 lid. Thus, despite a large variability in terms of length and sequence (Table S7), a conserved region is involved in the sub-pocket formation after substrate binding.

Unfortunately, the lack of resolution for the lid region of the KDO5 crystallographic structures makes it difficult to interpret the exact roles of residues other than *P12*. In VioC, a Ser (Ser224, *P14*) establishes a H-bond with a Gln (Gln137, *P2*) which, in turn, positions the latter in the active site by interacting with the N $\alpha$  of L-arginine substrate. This induces the stereoselectivity of the reaction. In the monomer C bound to lysine (Fig. S11), the flexible loop is better defined until residue Tyr233 (not in logo). On it, Asn232 (a residue which should appear in *P14* of the logo, but is not shown on it as it could not be superimposed correctly with the reference pocket) (Fig. S11) is orientated in a certain way, so that its interaction with the N $\alpha$  group of L-lysine substrate positions the latter to

perform the observed stereoselectivity. Thus, lid closure could be directed by a H-bond between *P14* and the N $\alpha$  of the amino acid substrate, either directly, as exemplified by KDO5, or indirectly, as observed for VioC.

**Influence of the adjacent loop.** Neighboring loops to the flexible lid might also be involved in substrate selection. For KDO1, the lid is well defined both in the apo and complexed forms, but the adjacent loop, from Ala160 to Ser171 (not in the logo except for *P5* and *P6*, respectively residues 167 and 168) undergoes large conformational changes. Among all CSL dioxygenases whose crystal structures have been reported so far, this is the first time such a loop motion is observed. Its motion seems to be induced by  $\alpha$ -KG binding as this loop is closed in apo and complexed forms with products and lysine, but opened when  $\alpha$ -KG is bound (Fig. S12). In the open form, lysine or  $\alpha$ -KG can easily access the active site and the iron. Induced fit motion between the adjacent loop and the flexible lid seems to be correlated with polar interactions between His164 (not in logo) and Glu235 (*P13*), both of which are distant from 4.1 Å (Fig. S12A). In the closed form, the lysine or hydroxy-lysine/succinate are covered by the adjacent loop and the active site is locked by interactions between Thr167 (*P5*) and Ser236 (*P14*) (Fig. S12B). In KDO5's monomer C in complex with lysine, the active site is locked by interaction between Thr165 (*P5*) and Asn232 (supposed to be on *P14*, see above paragraph). *P5* is a conserved position in the CSL family where Thr, Ser or Leu are frequently found (34%, 25%, 13% respectively). On the contrary, *P13* and *P14* are not conserved in the logo probably because of the large variety of length, shape and sequence compositions of the lid. For KDO1, the adjacent loop has a direct impact on positioning the substrate in the active site as Ser168 (*P6* in the logo) makes a H-bond with the carboxylate group of the lysine. In KDO5, Gly166 (*P6*) and Ser167 (*P7*) also stabilize the carboxylate group of the substrate, like Asn146 (*P6*) in AsnO<sup>19</sup>.

**Conservation in the dimeric interface.** Dimers have been described previously for other  $\alpha$ KAOs<sup>33</sup>. The similarity of the dimeric interface between KDO1 and KDO5 (Fig. 1) suggests that for the proper functioning of these enzymes, an interface may be needed. Such an assembly is also observable for the crystallographic structures of VioC, OrfP and AsnO, in which the same dimeric arrangement is mediated by two helices. A strong dimeric interface may provide a stabilizing force to allow flexible elements, such as the loops surrounding the active site, to adapt to structural requirements for a fast and stereoselective catalysis<sup>34</sup>.

Throughout the family, the residues that are part of the interface are found to be very diverse (Fig. S13A). However, an Asp, positioned at the middle of the interface for VioC, AsnO, KDO1, KDO5 and OrfP crystal structures, is conserved in 70% of the proteins (Fig. S13B). In KDO5, this Asp (Asp187) is involved in a H-bond network connecting residues from one chain with two residues from the other chain (Fig. S13C). This network creates a small cavity in which a hydrophobic residue (Ile in KDO5) from the other chain is hosted (Fig. S13D).

## Conclusion

The CSL family brings together Fe(II)/ $\alpha$ KAOs, some of which are known to catalyze the hydroxylation of amino acids and derivatives with a high specificity. We have determined the structures of two regio-divergent lysine  $\alpha$ KAOs in various states. These results allowed us to enrich the structural and computational analysis in the whole family to try to decipher the key determinants of selectivity. Our analysis strongly suggests that the substrate specificity together with the high stereoselectivity observed in this family are related to two structural features: a flexible loop (lid) and a sub-pocket. Our work pinpointed a particular residue that correlates with the regioselectivity. Although the sequence and length of the lid vary greatly within the family, our crystallographic structures confirm its dynamic nature and suggest a common behavior of closing upon substrate binding to form a sub-pocket surrounding the substrate, and re-opening to release the products. Our results also highlight the flexibility of an adjacent loop, enclosing the active site. A constant feature observed in this family is a conserved dimeric interface, with the flexible lid positioned at its opposite end. We therefore suggest that this stable interface is a key feature necessary to allow the lid flexibility and therefore a fast, regio- and stereoselective reaction. In summary, our analysis provides new insights into the specificity of the CSL enzymes and can guide further functional exploration for the discovery of  $\alpha$ KAOs with new activities.

## Methods

**Protein preparation.** The cloning, enzyme production and purification have been described elsewhere<sup>16</sup>. Oligonucleotides were purchased from Sigma-Genosys (St. Quentin Fallavier, France) and specific extensions added for cloning into pET-22b(+) (Novagen) including a hexa-histidine tail. The plasmid was introduced into *E. coli* BL21(DE3) for over-expression. The KDOs were purified on a Ni-NTA affinity column as reported previously<sup>35</sup>. Protein concentration was estimated using the Bradford method, and the purity was evaluated by SDS-PAGE electrophoresis. Aliquots of the purified protein were stored at  $-80^{\circ}\text{C}$ .

**Protein oligomeric state determination.** Analytic gel filtrations were performed on a Superdex 200 Increase 10/300 GL column by injecting 100  $\mu\text{L}$  of the purified protein with the following mobile phase: 50 mM Tris-HCl pH 8.0, 50 mM NaCl, 10% glycerol at a flow rate of 0.3 ml/min. The oligomerization state of the targeted protein was determined according to its elution volume compared to a calibration curve obtained with proteins of known global molecular weight.

**Crystallization, data collection and structure determination of KDO1 and KDO5.** The initial screening was carried out on a mosquito crystal liquid handler (TTP Labtech, Cambridge, UK) using sitting drops of 150  $\mu\text{L}$  of protein. For KDO1, larger crystals were grown using equal volume drops of KDO1 at 9.3 mg/ml in 0.050 M Tris-HCl, 0.2 M NaCl, 0.001 M DTT, pH 8.0 and of reservoir solution, by hanging drop vapor diffusion. For KDO5, crystallization experiments were carried out using equal volume drops of KDO5 at 5 mg/ml in buffer, 0.050 M Tris-HCl, 0.2 M NaCl, 0.001 M DTT, pH 8.0 and of reservoir solution, by sitting drop vapor diffusion. Crystals were improved by the clever technique of being transferred to a cryosolution (Tables S1 and S2),

and then flash cooled in liquid nitrogen. To obtain the KDO1 complexes, the crystals were soaked in either a solution of FeSO<sub>4</sub>/dithionite and L-lysine or FeSO<sub>4</sub>/dithionite and α-ketoglutarate or FeSO<sub>4</sub>/dithionite, Na succinate and (3S)-3-hydroxy-L-lysine. The ligand soaking was carried out directly in the cryosolution. To obtain the KDO5 complexes, the crystals were soaked in either a solution of FeSO<sub>4</sub>/dithionite and L-lysine, or FeSO<sub>4</sub>/dithionite and α-ketoglutarate, or both simultaneously. The concentration of the ligands and the soak time were optimized to obtain a good occupancy of the ligands with minimal damage to the crystals (Tables S1 and S2).

X-ray diffraction data sets were collected at the SOLEIL Synchrotron (Saint Aubin, France), on beamline Proxima 1 of a Dectris Pilatus 6M detector, or on beamline Proxima 2 A of a Dectris EigerX 9M or ADSC Q315r detector (Tables S1 and S2).

Data processing was carried out at the synchrotron facility using XDS with the xdsme script<sup>36</sup>. To solve the KDO1 structure, molecular replacement was carried out using PDB entry 2WBO<sup>17,37</sup> as the model. The electron density maps were viewed and fitted in COOT<sup>17,37</sup> and fitted with the sequence corresponding to UNIPROT C7QJ42. To solve the KDO5 structure, since molecular replacement with various models failed, a heavy-atom soaking approach was followed. The crystals used to solve the structure were obtained by being soaked for 20 mins. in 26% polyethylene glycol 3350, 0.2 M Imidazole malate, 0.15 M Li<sub>2</sub>SO<sub>4</sub> at pH 7.0 with 0.2 mM K<sub>2</sub>ReCl<sub>6</sub> in a cryosolution composed of 30% CM5. A data set spanning 360° of data was recorded at the rhenium L3 absorption edge. Four rhenium sites were located using SHELXD<sup>38</sup>. Initial Re-SAD phases were estimated and refined using PHASER<sup>39,40</sup> and then improved and extended by density modification, including NCS averaging, using the PARROT program<sup>39,40</sup>. A first model could be built automatically by BUCCANEER<sup>37,41</sup> and the rebuilding completed using COOT<sup>17,37</sup> with the sequence corresponding to UNIPROT J3BZS6.

The structures were subjected to various cycles of rebuilding COOT and refinement with BUSTER<sup>42</sup>. To take into account the anisotropic fall-off in resolution for KDO5-Lys and KDO5-α-KG datasets, STARANISO<sup>43</sup> was used. These data were used in all subsequent model building and refinement cycles.

The structures have been analyzed with PISA and PLIP<sup>44</sup> servers to calculate the surface buried in the interfaces and the interactions made by the ligands<sup>45</sup>. Of the four monomers in the asymmetric unit, the monomer with the lowest Local Ligand Density Fit (LLDF) score and the strongest electron density for the ligand is considered to be the most representative<sup>46</sup>.

**Sequence similarity network.** The initial set of proteins, containing all members of the Clavaminate synthase-like InterPro domain (IPR014503), was collected in May 2017 (587 sequences). Homologs to KDO1 and KDO5 retrieved from Uniprot using a Z-score threshold value of 150 were added to the set. The list of proteins (UniProt IDs) was provided to SSN-EFI (<http://efi.igb.illinois.edu/efi-est/>)<sup>25</sup>. Sequences and annotation information from the UniProt database were retrieved and mapped on the SSN.

**Active site classification.** To the set of CSL family, we added 153 homologs to KDO1 and KDO5 retrieved from Uniprot using a threshold Z-score value of 150. The entire set was clustered based on 100% sequence identity to obtain a non-redundant set of 608 proteins. Only proteins sharing at least 22% of sequence identity with at least one αKAO for which structures are available were kept (see following paragraph). In total 523 proteins were used for the active site classification, using ASMC software with three main steps<sup>29</sup>. Firstly, homology models for our set were built using structures templates, the crystal structures of CAS1 in complex with α-KG and proclavaminate acid (PDB 1DRT, chain A), VioC with (2S, 3S)-hydroxy-arginine (2WBQ, A), AsnO with succinate and (2S, 3S)-hydroxy-asparagine (2OG7, A), OrfP with succinate and (3R, 4R)-dihydroxy-L-arginine (4M2G, A), KDO1 and KDO5<sup>17,19,31,32</sup>. Structures of VioC with L-arginine and α-KG (6ALM) and NapI (6DAW) were not yet released in the PDB<sup>9,11</sup> at the time of our homology modeling work. Secondly, detection of pockets was performed on KDO5. Above the 12 predicted pockets, we selected the pocket that contained the catalytic (His, His, Glu/Asp) and the residues binding L-lysine. This pocket contains 28 residues and corresponds to the pocket surrounding the substrate. Among them, we kept residues whose Cα atom is within a distance of 12.5 Å from the catalytic glutamate. In total, the reference pocket is composed of 19 residues. The 19 equivalent residues were selected on VioC crystal structure to form the active site reference. VioC was chosen as the reference structure as it is average sequence according to the SSN. Finally, the 523 sequences were structurally aligned on VioC structure. Residues aligned with the 19 residues the active site reference were extracted to form a multiple sequence alignment of 523 sequences which were subsequently classified.

**Comparing lids and dimer interfaces between members of the family.** Homology models built were structurally aligned on KDO5 structure. We examined the residues that are structurally aligned with the extremities of the lid loop highlighted in this work. The lid corresponds to residues from 230 to 242 in KDO5.

Dimer interfaces formed by αKAOs monomers for each homology model were examined using an alternative version of the ASMC option (unpublished). In this version, the residues in contact between chain A and B of the KDO5 crystal structure are provided to ASMC to form logo sequences. Interface residues are retrieved using the BioPython/GetInterfaces.py script (provided by [www.blopg.com](http://www.blopg.com)). Residues are considered in contact if any of their atoms are less than 4.5 Å from any atoms of residues of the other chain. The interface neighbor cutoff was defined as 10 Å.

## References

- Sheldon, R. A. & Woodley, J. M. Role of Biocatalysis in Sustainable Chemistry. *Chem. Rev.* **118**, 801–838 (2018).
- Bornscheuer, U. T. *et al.* Engineering the third wave of biocatalysis. *Nature* **485**, 185–194 (2012).
- Wells, A. S., Finch, G. L., Michels, P. C. & Wong, J. W. Use of Enzymes in the Manufacture of Active Pharmaceutical Ingredients. A Science and Safety-Based Approach To Ensure Patient Safety and Drug Quality. *Org. Process Res. Dev.* **16**, 1986–1993 (2012).
- Sun, H., Zhang, H., Ang, E. L. & Zhao, H. Biocatalysis for the synthesis of pharmaceuticals and pharmaceutical intermediates. *Bioorg. Med. Chem.* **16**, 1275–1284 (2018).

5. Wan, J.-P., Gan, L. & Liu, Y. Transition metal-catalyzed C-H bond functionalization in multicomponent reactions: a tool toward molecular diversity. *Org. Biomol. Chem.* **15**, 9031–9043 (2017).
6. Lewis, J. C., Coelho, P. S. & Arnold, F. H. Enzymatic functionalization of carbon-hydrogen bonds. *Chem. Soc. Rev.* **40**, 2003–2021 (2011).
7. Gerlt, J. A. & Babbitt, P. C. Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu. Rev. Biochem.* **70**, 209–246 (2001).
8. Hausinger, R. P. FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Crit. Rev. Biochem. Mol. Biol.* **39**, 21–68 (2004).
9. Dunham, N. P. *et al.* Two Distinct Mechanisms for C-C Desaturation by Iron(II)- and 2-(Oxo)glutarate-Dependent Oxygenases: Importance of  $\alpha$ -Heteroatom Assistance. *J. Am. Chem. Soc.* **140**, 7116–7126 (2018).
10. Mukherjee, A. *et al.* Oxygen activation at mononuclear nonheme iron centers: a superoxo perspective. *Inorg. Chem.* **49**, 3618–3628 (2010).
11. Mitchell, A. J. *et al.* Visualizing the Reaction Cycle in an Iron(II)- and 2-(Oxo)-glutarate-Dependent Hydroxylase. *J. Am. Chem. Soc.* **139**, 13830–13836 (2017).
12. Martinez, S. & Hausinger, R. P. Catalytic Mechanisms of Fe(II)- and 2-Oxoglutarate-dependent Oxygenases. *J. Biol. Chem.* **290**, 20702–20711 (2015).
13. Wu, L.-F., Meng, S. & Tang, G.-L. Ferrous iron and  $\alpha$ -ketoglutarate-dependent dioxygenases in the biosynthesis of microbial natural products. *Biochim. Biophys. Acta* **1864**, 453–470 (2016).
14. Roper, L. & Grogan, G. Chapter 8 - Biocatalysis for Organic Chemists: Hydroxylations. in *Organic Synthesis Using Biocatalysis* (eds Goswami, A. & Stewart, J. D.) 213–241 (Academic Press, 2016).
15. Hibi, M. & Ogawa, J. Characteristics and biotechnology applications of aliphatic amino acid hydroxylases belonging to the Fe(II)/ $\alpha$ -ketoglutarate-dependent dioxygenase superfamily. *Appl. Microbiol. Biotechnol.* **98**, 3869–3876 (2014).
16. Baud, D. *et al.* Synthesis of Mono- and Dihydroxylated Amino Acids with New  $\alpha$ -Ketoglutarate-Dependent Dioxygenases: Biocatalytic Oxidation of C-H Bonds. *ChemCatChem* **6**, 3012–3017 (2014).
17. Helmetag, V., Samel, S. A., Thomas, M. G., Marahiel, M. A. & Essen, L.-O. Structural basis for the erythro-stereospecificity of the L-arginine oxygenase VioC in viomycin biosynthesis. *FEBS J.* **276**, 3669–3682 (2009).
18. Zhang, Z. *et al.* Crystal structure of a clavamate synthase-Fe(II)-2-oxoglutarate-substrate-NO complex: evidence for metal centered rearrangements. *FEBS Lett.* **517**, 7–12 (2002).
19. Strieker, M., Kopp, F., Mahler, C., Essen, L.-O. & Marahiel, M. A. Mechanistic and structural basis of stereospecific C $\beta$ -hydroxylation in calcium-dependent antibiotic, a daptomycin-type lipopeptide. *ACS Chem. Biol.* **2**, 187–196 (2007).
20. Finn, R. D. *et al.* InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* **45**, D190–D199 (2017).
21. Joint Center For. Crystal structure of a bsmA homolog (Mpe\_A2762) from *Methylobium petroleophilum* PM1 at 1.91 Å resolution, <https://doi.org/10.2210/pdb3pl0/pdb> (2010).
22. Xu, Q. *et al.* Crystal structure of a member of a novel family of dioxygenases (PF10014) reveals a conserved cupin fold and active site. *Proteins* **82**, 164–170 (2014).
23. Qin, H. M. *et al.* Crystal Structure of a Novel N-Substituted L-Amino Acid Dioxygenase in complex with alpha-KG from *Burkholderia ambifaria* AMMD, <https://doi.org/10.2210/pdb3w21/pdb> (2013).
24. Lukat, P., Blankenfeldt, W. & Mueller, R. GriE in complex with cobalt, alpha-ketoglutarate and l-leucine, <https://doi.org/10.2210/pdb5nci/pdb> (2017).
25. Gerlt, J. A. *et al.* Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta* **1854**, 1019–1037 (2015).
26. Baud, D. *et al.* Biocatalytic Approaches towards the Synthesis of Chiral Amino Alcohols from Lysine: Cascade Reactions Combining alpha-Keto Acid Oxygenase Hydroxylation with Pyridoxal Phosphate-Dependent Decarboxylation. *Adv. Synth. Catal.* **359**, 1563–1569 (2017).
27. Furnham, N., Dawson, N. L., Rahman, S. A., Thornton, J. M. & Orengo, C. A. Large-Scale Analysis Exploring Evolution of Catalytic Machineries and Mechanisms in Enzyme Superfamilies. *J. Mol. Biol.* **428**, 253–267 (2016).
28. Bastard, K. *et al.* Revealing the hidden functional diversity of an enzyme family. *Nat. Chem. Biol.* **10**, 42–49 (2014).
29. de Melo-Minardi, R. C., Bastard, K. & Artiguenave, F. Identification of subfamily-specific sites based on active sites modeling and clustering. *Bioinformatics* **26**, 3075–3082, <http://www.genoscope.fr/ASMC/> (2010).
30. Sacquin-Mora, S. Bridging Enzymatic Structure Function via Mechanics: A Coarse-Grain Approach. *Methods Enzymol.* **578**, 227–248 (2016).
31. Zhang, Z. *et al.* Structural origins of the selectivity of the trifunctional oxygenase clavaminic acid synthase. *Nat. Struct. Biol.* **7**, 127–133 (2000).
32. Chang, C.-Y. *et al.* Biosynthesis of Streptolidine Involved Two Unexpected Intermediates Produced by a Dihydroxylase and a Cyclase through Unusual Mechanisms. *Angew. Chem. Int. Ed Engl.* **126**, 1974–1979 (2014).
33. Clifton, I. J. *et al.* Structural studies on 2-oxoglutarate oxygenases and related double-stranded  $\beta$ -helix fold proteins. *J. Inorg. Biochem.* **100**, 644–669 (2006).
34. Haliloglu, T. & Bahar, I. Adaptability of protein structures to enable functional interactions and evolutionary implications. *Curr. Opin. Struct. Biol.* **35**, 17–23 (2015).
35. Kreimeyer, A. *et al.* Identification of the last unknown genes in the fermentation pathway of lysine. *J. Biol. Chem.* **282**, 7191–7197 (2007).
36. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
37. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
38. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
39. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
40. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 470–478 (2010).
41. Cowtan, K. TheBuccaneersoftware for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011 (2006).
42. Bricogne, G., Blanc, E., Brandl, M., Flensburg, C. & Keller, P. BUSTER version 2.11. 2. Cambridge, United Kingdom (2011).
43. Tickle, I. J. *et al.* Staraniso. Cambridge, United Kingdom: Global Phasing Ltd (2017).
44. Salentin, S., Schreiber, S., Joachim Haupt, V., Adasme, M. F. & Schroeder, M. PLIP: fully automated protein-ligand interaction profiler. *Nucleic Acids Res.* **43**, W443–W447 (2015).
45. Bleuler, S., Laumanns, M., Thiele, L. & Zitzler, E. PISA — A Platform and Programming Language Independent Interface for Search Algorithms. in *Lecture Notes in Computer Science*, 494–508 (2003).
46. Read, R. J. *et al.* A new generation of crystallographic validation tools for the protein data bank. *Structure* **19**, 1395–1412 (2011).

## Acknowledgements

The authors would like to thank V. Pellouin, C. Pellé, P. Sirvain and A. Fossey-Jouenne for excellent technical assistance. This work was supported by Commissariat à l'énergie atomique et aux énergies alternatives (CEA), the CNRS and the University of Evry Val d'Essonne. The authors are grateful to Dr. Alain Perret, Dr. Carine Vergne-Vaxelaire and Dr. Bret Church for helpful discussion and to The Faculty of Medicine and Health, University of Sydney, for hosting KB as a visitor scientist. We acknowledge synchrotron SOLEIL (Saint-Aubin, France) for provision of radiation facilities on the PROXIMA-1 and PROXIMA-2A beamlines.

## Author Contributions

A.Z. conceived the project. T.I. and E.A.S. achieved the crystallization and heavy atom derivatization. T.I., E.A.S. and P.L. have performed the crystallographic structure determination, refinement and analysis. K.B. conducted the structural bioinformatics and taxonomic analysis with input of A.Z. A.Z., K.B., T.I., E.A.S. wrote the manuscript, which was critically revised by all authors. All authors have read and approved the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-34795-9>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018