



# Catalytic mechanism of ancestral L-lysine oxidase assigned by sequence data mining

Received for publication, April 11, 2021, and in revised form, July 26, 2021. Published, Papers in Press, August 4, 2021.  
<https://doi.org/10.1016/j.jbc.2021.101043>

Sayaka Sugiura<sup>1,‡</sup>, Shogo Nakano<sup>1,2,\*,‡</sup>, Masazumi Niwa<sup>1</sup>, Fumihito Hasebe<sup>1</sup>, Daisuke Matsui<sup>3</sup>, and Sohei Ito<sup>1</sup>

From the <sup>1</sup>Graduate Division of Nutritional and Environmental Sciences, University of Shizuoka, Suruga-ku, Shizuoka, Japan; <sup>2</sup>PREST, Japan Science and Technology Agency, Kawaguchi, Japan; <sup>3</sup>Department of Biotechnology, College of Life Sciences, Ritsumeikan University, Kusatsu, Shiga, Japan

Edited by Joseph Jez

A large number of protein sequences are registered in public databases such as PubMed. Functionally uncharacterized enzymes are included in these databases, some of which likely have potential for industrial applications. However, assignment of the enzymes remained difficult tasks for now. In this study, we assigned a total of 28 original sequences to uncharacterized enzymes in the FAD-dependent oxidase family expressed in some species of bacteria including *Chryseobacterium*, *Flavobacterium*, and *Pedobactor*. Progenitor sequence of the assigned 28 sequences was generated by ancestral sequence reconstruction, and the generated sequence exhibited L-lysine oxidase activity; thus, we named the enzyme AncLLysO. Crystal structures of ligand-free and ligand-bound forms of AncLLysO were determined, indicating that the enzyme recognizes L-Lys by hydrogen bond formation with R76 and E383. The binding of L-Lys to AncLLysO induced dynamic structural change at a plug loop formed by residues 251 to 254. Biochemical assays of AncLLysO variants revealed the functional importance of these substrate recognition residues and the plug loop. R76A and E383D variants were also observed to lose their activity, and the  $k_{cat}/K_m$  value of G251P and Y253A mutations were approximately 800- to 1800-fold lower than that of AncLLysO, despite the indirect interaction of the substrates with the mutated residues. Taken together, our data demonstrate that combinatorial approaches to sequence classification from database and ancestral sequence reconstruction may be effective not only to find new enzymes using databases of unknown sequences but also to elucidate their functions.

Amino acids are key biomolecules that regulate physiological processes. Proteins and many bioactive compounds are synthesized utilizing L-amino acids and their metabolites, whereas D-amino acids serve as precursors of peptide drugs. Many L-amino acid-metabolizing enzymes have been reported, and several of them are used to synthesize fine chemicals (1), such as amino acid racemase (2) and L-amino acid dehydrogenase (3). Among these enzymes, L-amino acid oxidase (LAAO) is expected to have many industrial and pharmaceutical applications (4, 5). LAAOs are FAD-dependent enzymes and are broadly

expressed in many species, from bacteria to mammals (6–10). LAAO catalyzes the oxidation of the main chain amino group of L-amino acids and produces imino acids; these products are released into solvents and are quickly hydrolyzed to keto acid (4). The reduced FAD is reoxidized by oxygen molecules and generates H<sub>2</sub>O<sub>2</sub> as a by-product (4). LAAO activity can be estimated by quantifying H<sub>2</sub>O<sub>2</sub> using the Trinder reaction.

LAAOs have broad substrate selectivity, which enables their varied applications. For example, the LAAOs from *Rhodococcus opacus* (8) and *Proteus myxofaciens* (11) can be utilized to deracemize racemic amino acids to D-amino acids as efficiently as L-amino acid deaminase (12–15). On the other hand, LAAOs bearing high substrate specificity could be used to quantify specific L-amino acid concentrations in various samples. These LAAOs are named based on their substrate specificity and are the main focus of this study. Examples of these LAAOs include L-aspartate oxidase (16, 17) and L-glutamate oxidase (18, 19).

The assignment of new LAAOs aids in advancing the application of these enzymes. In addition, structural and functional analyses of LAAOs would be helpful to predict how LAAOs acquired unique substrate selectivity toward different L-amino acids through the molecular evolutionary process. Currently, many research groups have succeeded in experimentally screening LAAOs from various species (4, 5, 9). Simultaneously, the recent expansion of registered protein sequence data in public databases enables us to find new LAAOs using *in silico* enzyme screening methods and ancestral sequence reconstruction (ASR). Here, ASR is a sequence-based protein redesign method that can generate ancestral proteins located at each node of the phylogenetic tree (20, 21). Ancestral proteins often have desirable properties for use in practical applications, such as high thermostability (22, 23) and broad substrate selectivity (24). Therefore, ASR is currently being adopted as a tool for protein engineering (25). For example, new LAAOs that bear broad substrate selectivity (>10 L-amino acids) could be assigned by a paralog search of L-arginine oxidase (AROD) (26–28). Artificial LAAOs can be designed by ASR utilizing six of the assigned LAAOs as a sequence library. The designed enzyme, called AncLAAO, could be produced using the *Escherichia coli* expression system with the highest yield among any previously reported LAAOs (>50 mg/l). These can then be applied to

<sup>‡</sup> These authors contributed equally to this work.

\* For correspondence: Shogo Nakano, [snakano@u-shizuoka-ken.ac.jp](mailto:snakano@u-shizuoka-ken.ac.jp).

## Structural and functional analysis of AncLLysO

deracemize dozens of racemic amino acid derivatives to their D-forms with high enantiopurity (26). On the other hand, there are thousands of protein sequences belonging to the LAAO superfamily in databases (29), and the currently available functionally characterized sequences are only the tip of the iceberg. Several sequences, such as AROD (27, 30), AncLAAO (26, 28), and L-tryptophan oxidase (VioA) (31–33), can now be functionally annotated. Thus, there is an opportunity to acquire new LAAOs from the database using *in silico* enzyme screening methods.

In this study, we attempted to assign new LAAOs from the database utilizing one of the previously designed AncLAAO sequences (28) as a template. Through combinational approaches using paralog searches, the application of previously reported original protein sequence selection methods (27, 34–36), and biochemical assays, we newly characterized LAAOs exhibiting high specificity toward L-Lys. These included L-Lysine  $\alpha$ -oxidase (LLysO) from the *Chryseobacterium*, *Flavobacterium*, and *Pedobacter* species. The assigned LLysO has low sequence identity (less than 30%) to the previously reported LLysO from *Trichoderma viridae* (TvLysOX) (37, 38) and L-amino acid oxidase/monooxygenase (L-LOX/MOG) from *Pseudomonas* sp. AIU813 (39–41); therefore, the enzymatic properties of the assigned LLysO could not be estimated only from the sequence analysis. Ancestral LLysO (AncLLysO) was designed utilizing the assigned LLysO sequences as templates. Compared with native LLysO, AncLLysO improved thermal stability and provided good quality of crystals. The combination of biochemical and structural analysis of AncLLysO and its variants revealed the substrate recognition and reaction mechanism of LLysO at a molecular level.

## Results

### Sequence classification from databases to assign a new LLysO family

In a previous study, we reported that paralog search and sequence library classification using several key residues as a motif are effective for finding new enzymes from sequence databases bearing unique properties, such as broad substrate selectivity and high thermostability (26, 35, 36). By applying combinational approaches (represented in Fig. 1), we attempted to assign new LAAOs from the databases using the following procedures. First, paralogs that bear moderate sequence identity (<30%) with AncLAAO-N5 (28) were identified by Blastp utilizing open reading frames of six *Pseudoalteromonas* genera as a library. As shown in a previous study, these genera have LAAO sequences that were utilized to design AncLAAO-N5 (26, 28). Through this analysis, a paralog sequence was assigned in *Pseudoalteromonas luteoviolacea*; the sequence identity was quite low compared with the functionally annotated LAAOs. This paralog sequence was called PIHyp (Fig. 1).

Next, we attempted to classify homologous sequences of PIHyp registered in the sequence database by applying the previously reported approach (35, 36). To achieve this, the sequences were prepared by submitting PIHyp to Blastp (42); the database and E-values were set to “nonredundant” and  $1.0 \times 10^{-6}$ , respectively. Finally, a total of 112 sequences could be

obtained by eliminating sequences that were more than 5% longer or shorter than PIHyp and that shared a greater than 90% identity with other sequences; these sequences were utilized as the library (Fig. 1). The elimination was performed using an original python script (35). We attempted to assign key residues to curate the library by following previously reported procedures (35, 36). Four residues (the 235th, 379th, 450th, and 544th residues) were assigned as key residues by the analysis. After pairwise alignment between PIHyp and each of the sequences in the library (total 1–112 sequences), the sequences were classified into two groups based on the combination of the key residues: one had Leu, Met, Glu, and Pro (LMEP, yellow bar in Fig. 1) and the other had Met, Val, Gly, and Gln (MVGQ, green bar in Fig. 1). Finally, a total of 28 LMEP and seven MVGQ sequences were acquired (Fig. 1).

The quality of the sequence libraries could be evaluated by comparing conservation energies before and after the classification (34, 36). Conservation energies at the  $i^{\text{th}}$  residues ( $E_{aa,i}$ ) and the sum of all residues ( $E_c$ ) could be calculated with the previously reported equation:

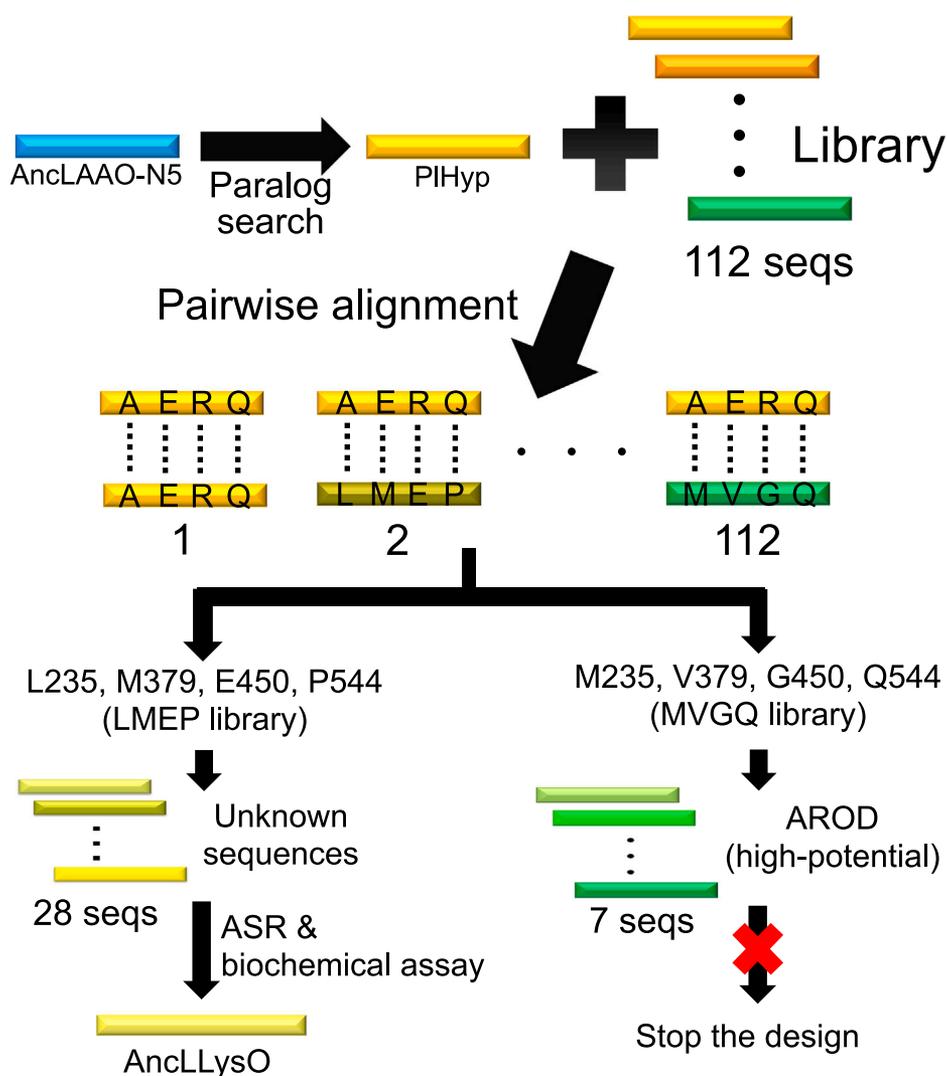
$$E_{aa,i} = -\ln f_{aa,i}, E_c = \sum_{i=1}^n E_{aa,i}$$

where  $f_{aa,i}$  is the frequency at the  $i^{\text{th}}$  positions of the target protein (PIHyp, in this case) in the alignment, and  $n$  is the total number of residues in the target protein (43). The  $E_c$  value would be close to zero if the curation is correctly performed (36). The distribution of the  $E_{aa,i}$  value for before and after the classification is represented in Fig. S1. The  $E_c$  value for the LMEP (102.6 in Fig. S1B) and MVGQ (129.5 in Fig. S1C) library was more than 3-fold lower than that of the noncurated library (397.9 in Fig. S1A), suggesting that the library could be classified to align the sequences accurately. Clearly assigned consensus residues (residues 35–44, 269–278, and 316–323) after curation also supported this point (Fig. S1).

Here, the seven MVGQ sequences would have L-arginine oxidase (AROD) activity because they shared more than 70% identity with the already reported AROD sequences, and therefore, we discontinued further design. On the other hand, the 28 LMEP sequences share less than 30% identity with the already characterized LAAOs, suggesting that the sequences had novel activities. Preliminary enzyme activity assay for one of the sequences from *Chryseobacterium angstadtii* (CaLLysO, WP\_048506397.1, Table S1) suggested that the sequences would have L-Lys oxidase (LLysO) activity. To prove this point, the ancestral protein of the 28 LMEP sequences, named AncLLysO (Table S2), was designed as a representative of the curated sequences; AncLLysO would have LLysO activity if the 28 LMEP sequences bear this activity.

### Enzyme functional analysis of CaLLysO and AncLLysO

Enzymatic properties of CaLLysO and AncLLysO were analyzed using biochemical assays and suggested that substrate specificity and pH dependency were similar to each other. Specific activity measurement of CaLLysO (red bar in Fig. 2A)



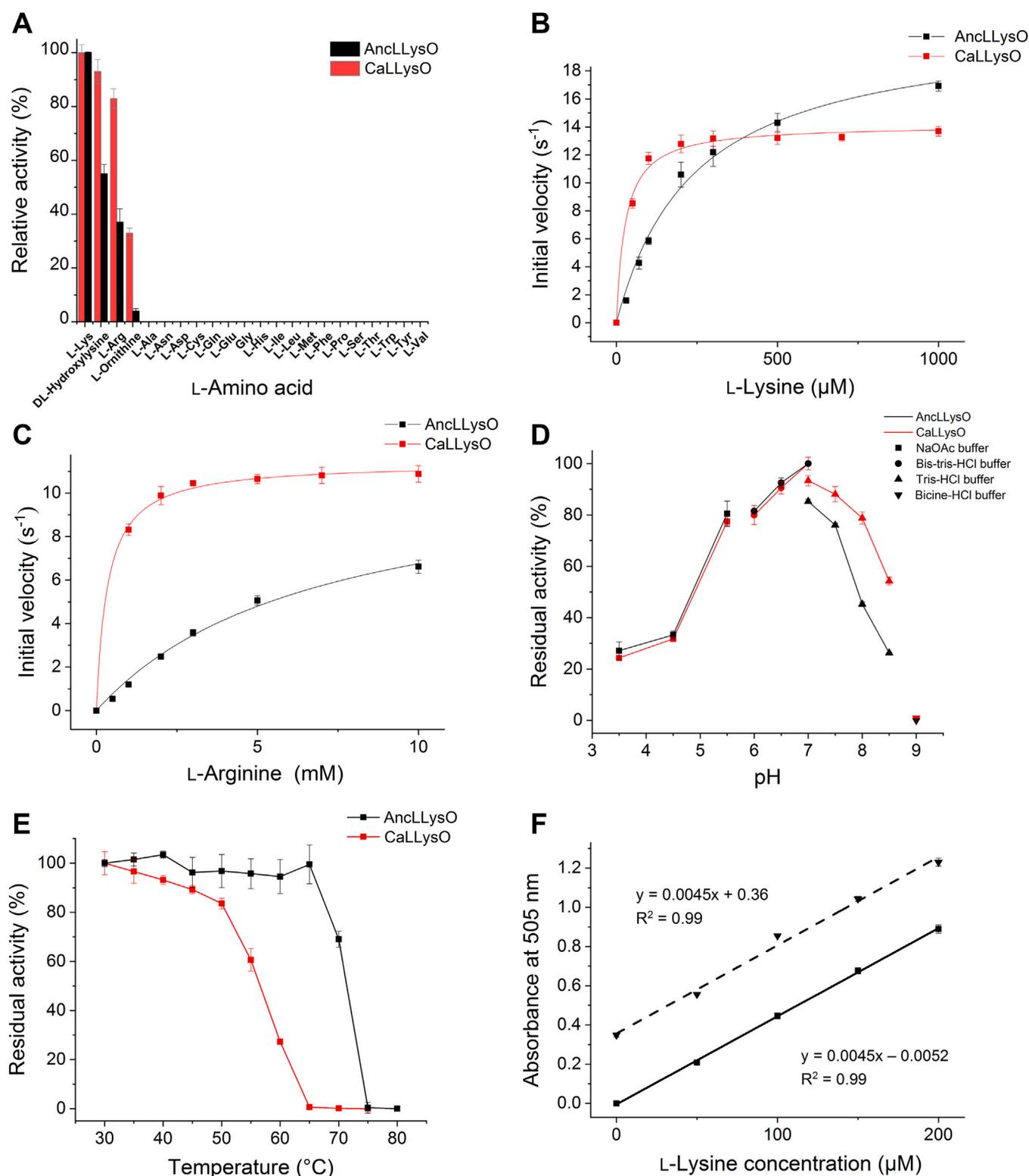
**Figure 1. Schematic view indicating how to design AncLLysO from AncLAAO-N5 sequence data.** Paralog sequence of L-amino acid oxidase (LAAO), named PIHyp (GI: WP\_063360054.1), was assigned from *Pseudoalteromonas luteoviolacea*, similar to the assignment of AncLAAO (26). A library containing a total 112 homolog sequences to PIHyp was prepared by applying Blastp analysis and the preprocessing approach. The library was classified into two curated libraries using four key residues (235, 379, 450, and 544) as sequence motifs and pairwise alignment. The key residues were assigned by applying the same procedure used in previous studies (35, 36). Sequences bearing M235, V379, G450, and Q544 (green bar, total seven sequences) exhibit L-arginine oxidase (AROD) activity because they share high sequence identity with the enzyme; thus, we halted further design. On the other hand, sequences bearing L235, M379, E450, and P544 (yellow bar, total 28 sequences) had no sequence identity to already reported enzymes. The ancestral protein of the curated sequences was generated by ancestral sequence reconstruction (ASR) the designed sequence was named AncLLysO based on the enzymatic activity.

and AncLLysO (black bar in Fig. 2A) toward 20 L-amino acids and two derivatives (L-Ornithine [L-Orn] and 5-hydroxy-DL-Lysine [5-OH-DL-Lys]) indicated that both of the LLysOs exhibited strong activity toward basic amino acids in the following order: L-Lys > 5-OH-DL-Lys > L-Arg > L-Orn (Fig. 2A). Enzyme kinetic analysis toward L-Lys (Fig. 2B) and L-Arg (Fig. 2C) indicated that both CaLLysO and AncLLysO had the highest enzyme efficiency ( $k_{cat}/K_m$ ) values toward L-Lys compared with other substrates; the values of L-Lys were >15- and 42-fold higher than those of L-Arg in CaLLysO and AncLLysO, respectively (Table 1). The pH analysis indicated that both CaLLysO and AncLLysO exhibited the highest activity at pH 7.0 (Fig. 2D). On the other hand, a functional trade-off between thermostability and  $k_{cat}/K_m$  value toward L-Lys was observed between CaLLysO and AncLLysO; the  $t_{1/2}$  value of AncLLysO was approximately 15 °C higher than that

of CaLLysO (Fig. 2E), whereas the  $k_{cat}/K_m$  value of AncLLysO was more than 7-fold lower than that of CaLLysO (Table 1). Thus, AncLLysO is more thermophilic than CaLLysO.

The highly specific LLysO can be applied to quantify the concentration of L-Lys in various samples (44). Here, AncLLysO is suitable for the application because it exhibits high thermostability and a larger relative  $k_{cat}/K_m$  value toward L-Lys in comparison with L-Arg; the  $k_{cat}/K_m$  value was 42.5 for AncLLysO and 15.5 for CaLLysO (Table 1). Utilizing AncLLysO, the L-Lys concentration was quantified for the samples containing only buffer (straight line in Fig. 2F) and plasma (dotted line in Fig. 2F), and the results were plotted in Figure 2F. The slopes of the plots were identical to each other (0.045 in Fig. 2F), indicating that AncLLysO can quantify L-Lys concentration as well as the previously reported highly specific L-amino acid oxidases.

## Structural and functional analysis of AncLLysO



**Figure 2. Biochemical analysis of CaLLysO and AncLLysO.** Relative specific activity of CaLLysO and AncLLysO toward different amino acid substrates (10 mM) (A). The parameters of CaLLysO and AncLLysO are represented as *red* and *black* in the figure. Both enzymes exhibited the highest activity toward L-Lys. Enzyme kinetic plots of CaLLysO and AncLLysO toward L-Lys (B) and L-Arg (C), respectively. Analysis of pH optima (D) and thermostability (E). Optimal pH value of CaLLysO and AncLLysO was 7.0. The  $t_{1/2}$  values of CaLLysO and AncLLysO were approximately 57 and 73 °C, respectively. Quantification of L-Lys concentration by AncLLysO (F). Samples containing human plasma (downward triangle) and only buffer (square) are shown as dashed and straight lines, respectively. The assay was performed by endpoint method. All of the experiments were performed using two biological replicates per three technical replicates.

Summarizing the results, we showed that both CaLLysO and AncLLysO have LLysO activity. In particular, AncLLysO has favorable properties for analyzing its biochemical functions,

such as high thermostability and specificity toward L-Lys. Thus, AncLLysO was adopted as a research target to reveal its enzymatic properties at a molecular level.

**Table 1**  
Enzyme kinetic parameters of AncLLysO and CaLLysO toward the substrates L-Lys, L-Arg, and L-Orn

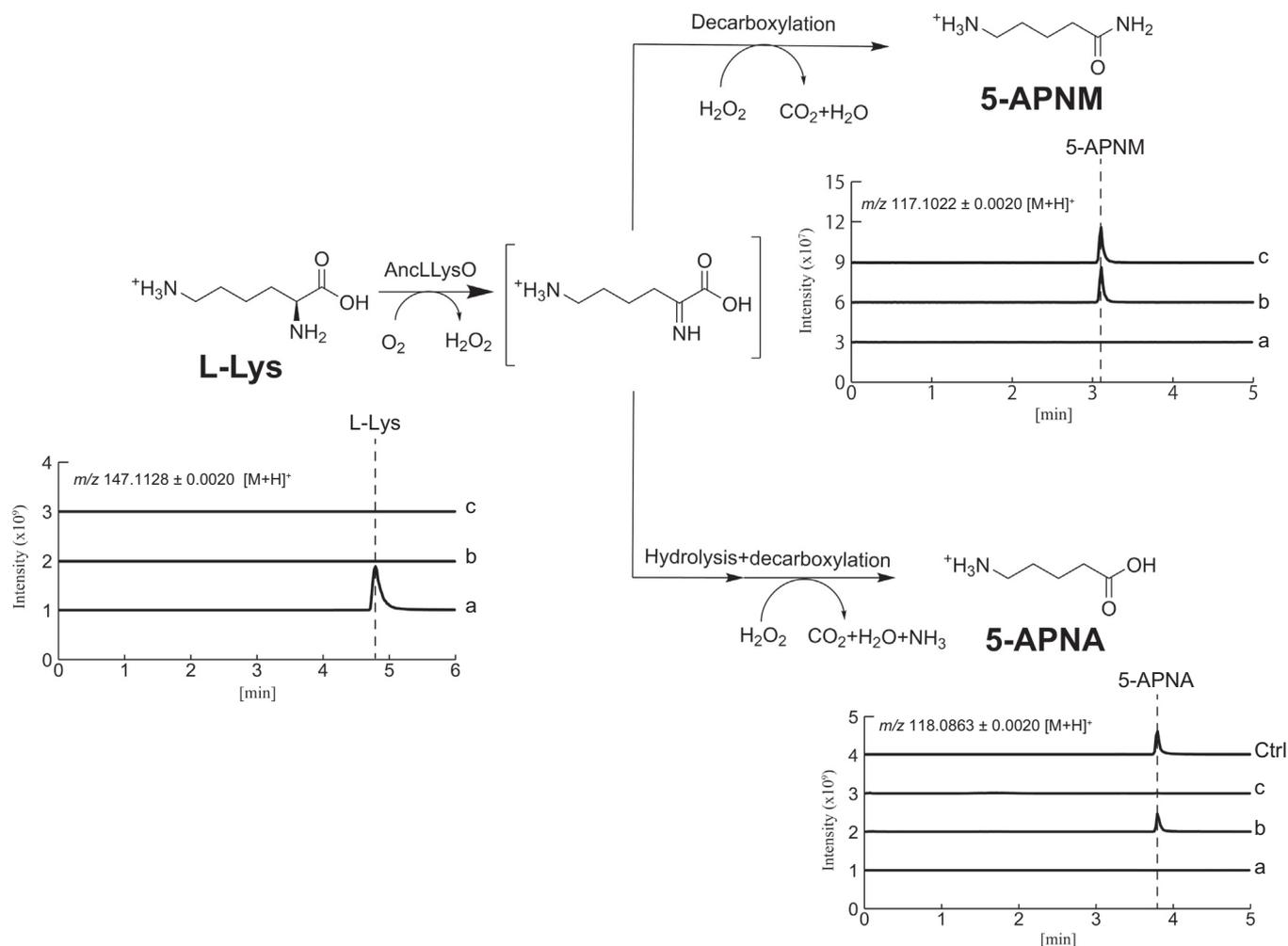
Substrate	CaLLysO			AncLLysO		
	$k_{cat}$ s <sup>-1</sup>	$K_m$ mM	$k_{cat}/K_m$ s <sup>-1</sup> mM <sup>-1</sup>	$k_{cat}$ s <sup>-1</sup>	$K_m$ mM	$k_{cat}/K_m$ s <sup>-1</sup> mM <sup>-1</sup>
L-Lys	14.2 ± 0.3	0.027 ± 0.003	516	21.7 ± 0.9	0.3 ± 0.03	72.3
L-Arg	11.4 ± 0.1	0.34 ± 0.03	33.2	11.4 ± 1.0	6.9 ± 1.1	1.7
L-Orn	3.60 ± 0.07	4.4 ± 0.4	0.81	46.4 ± 3.0	45.8 ± 5.2	1.0

The measurement of enzyme kinetic parameters was performed using two biological replicates per three technical replicates.

### LC-HRMS analysis of products generated by converting L-Lys with AncLLysO

Currently, a total of two types of LLysO have been reported: one is L-Lys  $\alpha$ -oxidase, which oxidizes the main chain amino group of L-Lys (44), and the other is L-Lys  $\epsilon$ -oxidase, which oxidizes the side chain amino group of L-Lys (45). To ensure clarity regarding the types represented by AncLLysO identified, the reaction products of AncLLysO were analyzed by LC-HRMS (Fig. 3). Here, the reactions were conducted under the following three independent conditions: reaction buffer (100 mM KPB (7.0) and 10 mM L-Lys) containing no enzyme

(a in Fig. 3), containing 0.25 mg AncLLysO (b in Fig. 3), and containing 0.25 mg AncLLysO and 1000 U catalase (c in Fig. 3). After stopping the reaction by mixing nine volumes of acetonitrile with one volume of reaction solution, the mixture was centrifuged at 20,000g at 4 °C for 10 min. The supernatant was analyzed by LC-HRMS. A peak corresponding with L-Lys completely disappeared following the addition of AncLLysO (b and c in Fig. 3), indicating that AncLLysO can utilize L-Lys as a substrate. The product peak analysis suggested that AncLLysO exhibits L-Lys  $\alpha$ -oxidase activity. In fact, a peak of 5-aminopentanoic acid (5-APNA) was detected in condition



**Figure 3. Reaction scheme of AncLLysO predicted from LC-HRMS analysis.** Extracted ion count chromatograms of  $[M + H]^+$  for each compound (L-Lys, 5-APNM, and 5-APNA). Here, each chromatogram (a, b, and c) was obtained by analyzing samples after reacting for 24 h under the following conditions: a, 10 mM L-Lys; b, 10 mM L-Lys and 0.25 mg AncLLysO; and c, 10 mM L-Lys, 0.25 mg AncLLysO, and 1000 U catalase. 5-APNA, 5-aminopentanoic acid; 5-APNM, 5-aminopentanimide.

## Structural and functional analysis of AncLLysO

“b” (Fig. 3). As shown in a previous study (41), 5-APNA was produced under the conditions containing H<sub>2</sub>O<sub>2</sub> by hydrolyzing and decarboxylating the resultant imino acids (Fig. 3). The analysis also suggested that AncLLysO generates 5-aminopentanamide (5-APNM); the peak of 5-APNM was confirmed under conditions with (b in Fig. 3) and without catalase (c in Fig. 3), as was the case with L-LOX/MOG (41).

Referring to the analysis for L-LOX/MOG (41), we estimated the relative yield of 5-APNA and 5-APNM from their peak area value. Our analysis suggested that a larger amount of 5-APNA was produced than 5-APNM. The area value for 5-APNA was more than 15-fold larger than that of 5-APNM (b in Fig. 3), suggesting that, for AncLLysO, the monooxygenase function was the promiscuous activity compared with the oxidase function. This was clearly different in the case of L-LOX/MOG, where the monooxygenase function was the main activity identified through LC-HRMS analysis (41).

### Structural analysis of AncLLysO (ligand-free) form

As shown through biochemical analysis, AncLLysO exhibited both L-Lys  $\alpha$ -oxidase and monooxygenase activity, and structural analysis is expected to reveal its reaction mechanism at a molecular level. However, this mechanism is difficult to predict because structures that share high sequence identity with AncLLysO were unavailable. In fact, the most similar

structure to AncLLysO registered in PDB was AncLAAO and these structures share only 20% identity. Construction of an accurate homology model of AncLLysO is difficult, and therefore, we were required to determine crystal structures of AncLLysO to elucidate the mechanism.

In this study, the crystal structure for the ligand-free form of AncLLysO (AncLLysO(LF)) was determined at a 2.4-Å resolution. The initial phase was determined using the iodide single anomalous dispersion (SAD) method. Crystallographic parameters are represented in Table 2. AncLLysO has a fold typical of the flavin-dependent amine oxidase (FAO) superfamily (Fig. 4A)—the Rossmann-core fold, which recognizes FAD, and a hot-dog-like fold, which is important for the recognition of substrates (29). Structural analysis of AncLLysO(LF) by DALI server (46) also supported this point as the top ten structures that have structural similarity to AncLLysO are members of the FAO superfamily (Table 3).

The active site structure of AncLLysO(LF) indicated that residues that would form interactions with the main-chain L-Lys are present including R76, G553, and W554 (Fig. 4B). These residues are highly conserved in other LAAOs (asterisk in Fig. S2) (4). On the other hand, there is no space to bind L-Lys at the site as the side chain Y254 occupies the space (Fig. 4B), inferring that dynamic conformational changes of Y254 would be induced as the reaction progressed. The crystal structure of the L-Lys-binding form of AncLLysO has to be

**Table 2**

Statistics of X-ray diffraction data collection AncLLysO for native (ligand-free) and substrate-binding forms of K387A variant (L-Lys- and L-Arg-binding form)

Parameter	Native (ligand-free)	Iodide-SAD	L-Lys binding (K387A variant)	L-Arg binding (K387A variant)
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub>	P2 <sub>1</sub>
Unit cell parameters				
a (Å)	89.4	90.3	85.8	85.9
b (Å)	80.5	80.6	80.1	80.1
c (Å)	173.8	172.8	88.0	88.0
$\alpha$ (degree)	90.0	90.0	90.0	90.0
$\beta$ (degree)	90.0	90.0	90.4	90.5
$\gamma$ (degree)	90.0	90.0	90.0	90.0
X-ray source	BL5A (PF)	BL5A (PF)	BL5A (PF)	NW12A (PF-AR)
Wavelength (Å)	1.00	1.70	1.00	1.00
Resolution (Å)	49.3–2.2 (2.31–2.20)	49.4–2.4 (2.53–2.40)	42.9–2.4 (2.53–2.40)	48.9–2.2 (2.32–2.20)
No. of reflections <sup>a</sup>	864,192	2,622,002	315,778	410,497
No. of unique reflections	63,996	50,279	46,762	60,710
Completeness (%)	98.8 (97.8)	100 (99.9)	99.8 (99.0)	99.7 (97.8)
I/ $\sigma$ (I)	18.9 (3.2)	42.3 (12.4)	16.2 (4.0)	15.2 (3.9)
CC <sub>1/2</sub> <sup>b</sup>	0.999 (0.868)	1.00 (0.990)	0.998 (0.926)	0.998 (0.899)
R <sub>merge</sub> <sup>b</sup>	0.106 (0.846)	0.092 (0.405)	0.087 (0.421)	0.084 (0.446)
B of Wilson plot (Å) <sup>2</sup>	29.0	30.1	27.6	26.8
Iodide Sites		81		
FOM before DM <sup>c</sup>		0.28		
FOM after DM <sup>c</sup>		0.71		
R <sup>d</sup>	0.216		0.185	0.192
R <sub>free</sub> <sup>e</sup>	0.255		0.231	0.231
RMSD of geometry				
Bond length (Å)	0.007		0.007	0.007
Bond angle (degree)	1.498		1.460	1.438
Geometry				
Ramachandran outlier (%)	0.3		0.3	0.1
Ramachandran favored (%)	99.7		99.7	99.9
PDB code	7EIH		7EII	7EIJ

Abbreviation: SAD, single anomalous dispersion.

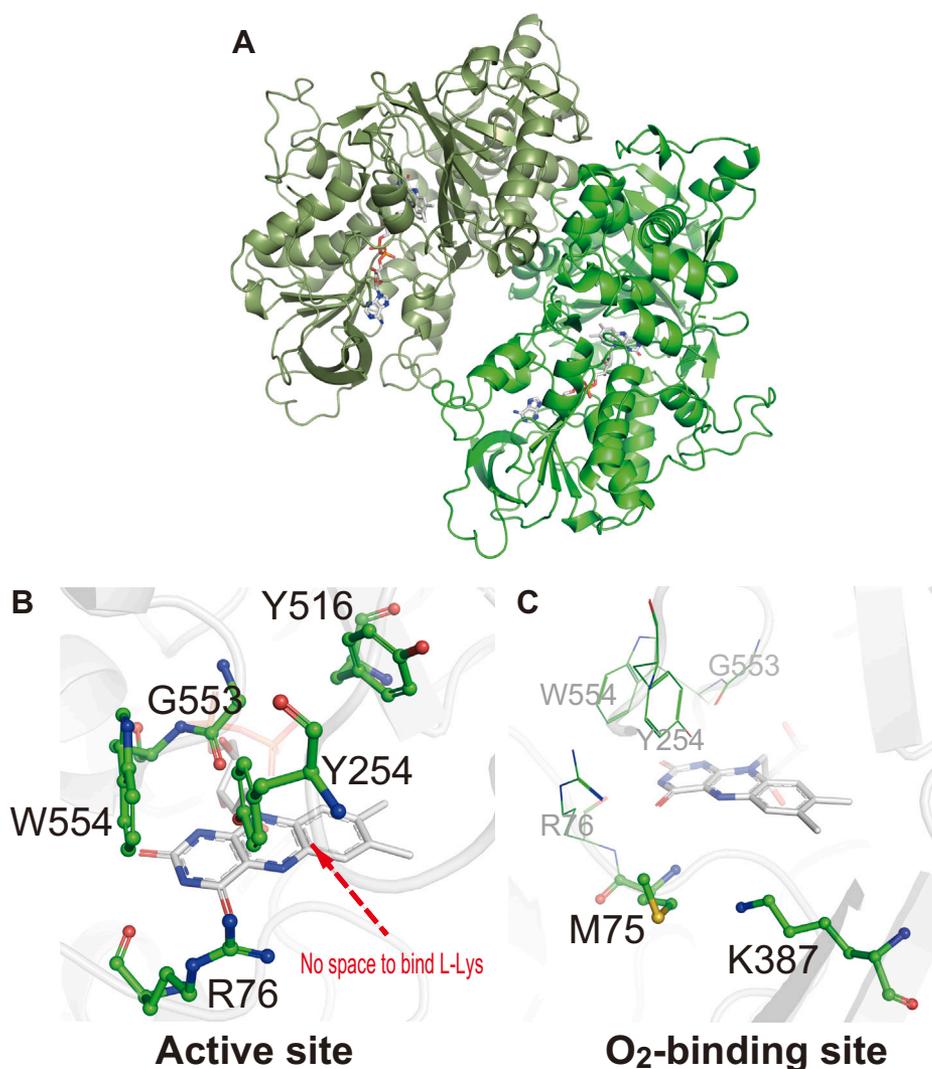
<sup>a</sup> Sigma cutoff was set to none ( $F > 0\sigma F$ ).

<sup>b</sup>  $R_{\text{merge}} = \sum_i \sum_h |I_i(h) - \langle I(h) \rangle| / \sum_h I(h)$ , where  $I_i(h)$  is the  $i^{\text{th}}$  measurement of reflection  $h$ , and  $\langle I(h) \rangle$  is the mean value of the symmetry-related reflection intensities. Values in brackets are for the shell of the highest resolution.

<sup>c</sup> FOM before/after DM means the figure of merit before/after density modification.

<sup>d</sup>  $R = \sum ||F_o| - |F_c|| / \sum |F_o|$ , where  $F_o$  and  $F_c$  are the observed and calculated structure factors used in the refinement, respectively.

<sup>e</sup>  $R_{\text{free}}$  is the  $R$ -factor calculated using 5% of the reflections chosen at random and omitted from the refinement.



**Figure 4. Structure analysis of AncLLysO ligand-free form.** Overall structure of AncLLysO ligand-free form (A). Active site structure of AncLLysO(LF) (B). Side chain of Y254 occupies active site of AncLLysO, so there is no space to bind substrate, L-Lys. O<sub>2</sub>-binding site of AncLLysO (C). As well as the previously reported LAAO family, there is a Lys residue (K387) that coordinates O<sub>2</sub> to oxidize FADH<sub>2</sub>.

determined to prove this hypothesis, and inactivated variants are required for structure determination. We attempted to design the variants by mutating the residues at the O<sub>2</sub>-binding site (Fig. 4C). Previous research indicates that, in many FAD-dependent oxidases, the Lys residue at the site would recognize the O<sub>2</sub> molecule, which oxidizes the FADH<sub>2</sub>, and their variants quietly reduce the activity (47). In AncLLysO, K387 corresponds with the residue (Fig. 4C), and the activity of the AncLLysO(K387A) variant was reduced below the detection limit of the activity assay as expected (Table 4). Utilizing AncLLysO(K387A), we attempted to determine the ligand-binding structures of AncLLysO.

#### Active site structures of the L-Lys- and L-Arg-binding forms of AncLLysO(K387A)

Crystal structures of the L-Lys- and L-Arg-binding forms of AncLLysO(K387A) were determined at 2.4- and 2.2-Å resolution, respectively, using the cocrystallization method. In

general, fitting of substrates into an electron density map would be difficult at the resolution of the AncLLysO(K387A) structure; however, the substrates appeared to be fitted into the map appropriately. In fact, the polder F<sub>o</sub>-F<sub>c</sub> omit map indicates that L-Lys (Fig. 5A) and L-Arg (Fig. 5B) coordinate at the active site using different binding modalities; a stereo view of L-Lys- and L-Arg-binding forms is shown in Fig. S3. In the L-Lys-binding form, the carboxyl and side-chain amino groups form hydrogen bonds with R76 and E383, respectively (dotted line in Fig. 5A). In the L-Arg-binding form, only the guanidino group formed interactions with E383 and Q552 (dotted line in Fig. 5B). Differences in the binding mode would affect the substrate specificity of AncLLysO. In fact, the main-chain carbon atom of L-Arg is moved to the arrowed direction approximately 1.5 Å compared with L-Lys (Fig. 5C). Because of this movement, the distance between the main-chain carbon atom of L-Arg and the N5 atom of FAD is about 1.2 Å longer than in the case of L-Lys (Fig. 5C). This causes inefficient hydride transfer of C $\alpha$ -H from L-Arg to the N5 atom of FAD

## Structural and functional analysis of AncLLysO

**Table 3**

Top ten structures that have structural similarity to AncLLysO (PDB ID: Ligand free) by Dali search against PDB25

Organism	Class	PDB code monomer	Z-score <sup>a</sup>	r.m.s.d. <sup>b</sup>	lali <sup>c</sup>	%id <sup>d</sup>
<i>Synthetic construct</i>	Ancestral L-amino acid oxidase	7C4M-A	32.1	2.5	506	25
<i>Chromobacterium violaceum</i>	Flavin-dependent L-tryptophan oxidase, VioA	6ESE-A	29.9	2.7	398	19
<i>Exiguobacterium sibiricum</i>	Protoporphyrinogen oxidase	3LOV-A	29.7	3.4	392	15
<i>Gloydus halys</i>	L-amino acid oxidase	1TDO-A	29.4	2.6	436	17
<i>Nicotiana tabacum</i>	Protoporphyrinogen oxidase	1SEZ-A	28.0	4.0	399	15
<i>Homo sapiens</i>	Renalase	3QJ4-A	27.8	2.8	312	16
<i>Pseudomonas syringae pv.tomato</i>	Amine oxidase	3KKJ-A	27.0	2.9	311	17
<i>Homo sapiens</i>	Lysine-specific histone demethylase 1	2X0L-A	26.1	2.9	399	17
<i>Homo sapiens</i>	Amine oxidase	2Z5X-A	25.9	3.2	422	14
<i>Homo sapiens</i>	Lysine-specific histone demethylase 1B	4FWE-A	25.6	3.1	399	17

<sup>a</sup> A measure of the statistical significance of the result relative to alignment of random structure.

<sup>b</sup> Root-mean-square deviation for C $\alpha$  atoms.

<sup>c</sup> Number of aligned residues.

<sup>d</sup> Sequence identity between AncLLysO and the targeted chain.

during the reaction by AncLLysO. A similar phenomenon may occur in the L-Orn-binding form of AncLLysO.

Next, we attempted to represent structural changes of AncLLysO that are induced by the binding of substrates. Structural comparison between AncLLysO(LF)- (green) and the L-Lys-binding form of AncLLysO(K387A; magenta) are shown in Figure 5D, suggesting that dynamic structural changes could be confirmed at the following residues: 251 to 254, 383, and 516 (Fig. 5D). From a structural comparison between the ligand-free form and the L-Lys- or L-Arg-binding form of AncLLysO, substrate binding appeared to induce a conformational change of a loop that is formed by residues 251 to 254 (251-GGYY-254); *cis*-*trans* transformation at G251 is remarkable in this change. The loop position is consistent with a plug loop in L-LOX/MOG, which occurs when these two structures are superimposed on each other (40). Therefore, we call the loop a “plug loop” as well. The active site residues E383 and Y516 moved to the optimal position to recognize the L-Lys associating with the conformational change of the plug loop in AncLLysO (Fig. 5D).

Structural comparison between AncLLysO and other LAAOs may be helpful to predict how AncLLysO exhibits high substrate specificity toward L-Lys. Superimposed structures of

AncLLysO(K387A), AncLAAO, and VioA are represented in Figure 5D, indicating that the size of the active site of AncLLysO is narrower than those of AncLAAO and VioA, as the site was formed mainly by bulky aromatic residues (Phe and Tyr). There are four aromatic residues in AncLLysO (magenta in Fig. 5D), whereas the number is decreased to two in AncLAAO and VioA (dark blue and yellow in Fig. 5D). The narrowing of the active site in AncLLysO made binding to amino acids other than L-Lys difficult; therefore, this could be one of the reasons that AncLLysO exhibits high specificity toward L-Lys.

### Enzyme kinetic analysis of AncLLysO variants

The structural analysis indicated that conformational rearrangement of the active site structure would be caused by the binding of substrates, triggering a dynamic conformational change of the plug loop. The next challenge is to demonstrate the functional role of the residues, which is related to the rearrangement using biochemical assays. To accomplish this, we attempted to analyze the enzyme kinetic parameters of several AncLLysO variants.

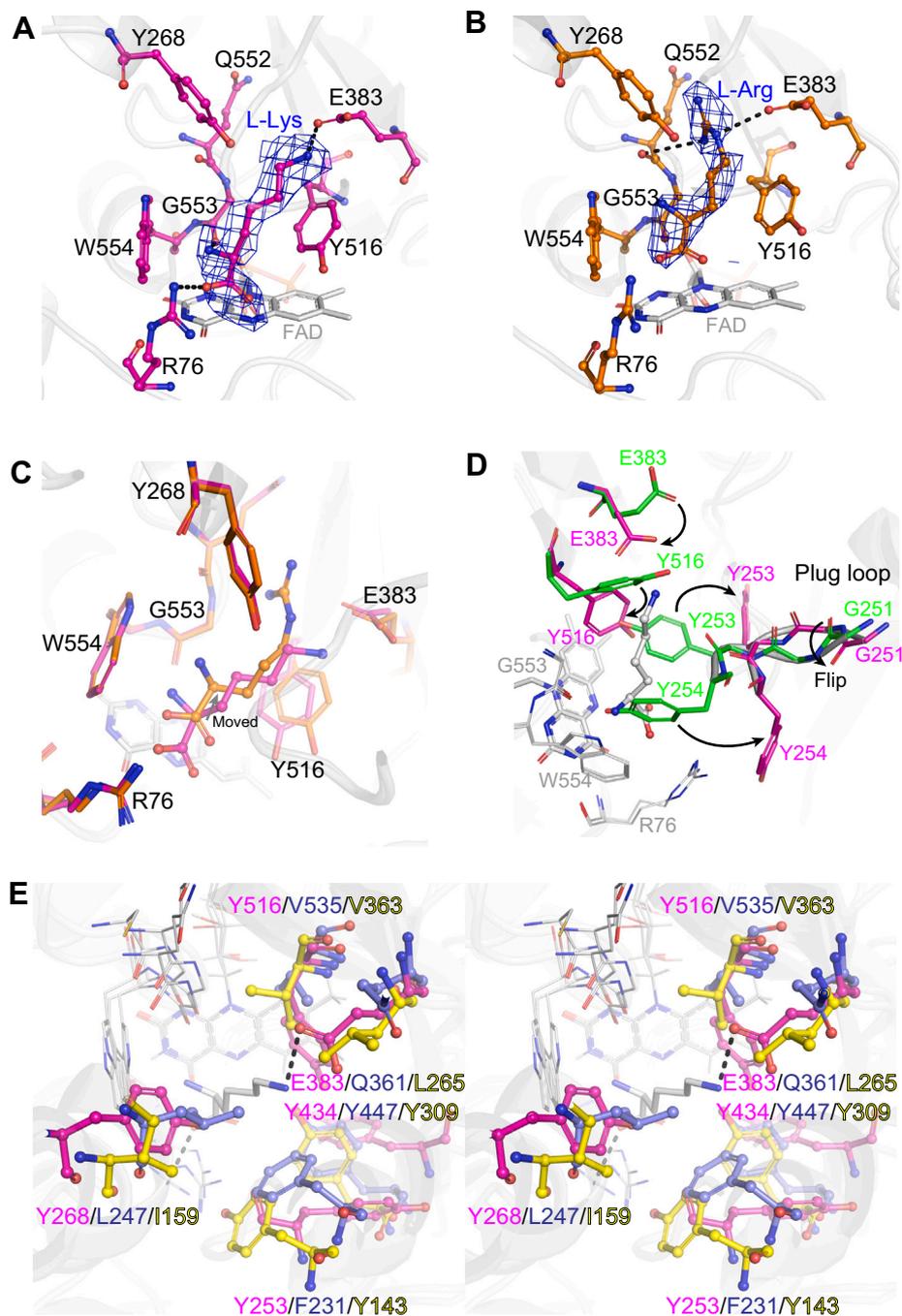
First, analysis of AncLLysO variants of the substrate recognition sites (R76A, E383D, Y268F, and Y516F) was performed,

**Table 4**

Enzyme kinetic parameters of AncLLysO variants toward L-Lys

Sample name	$k_{cat}$ s <sup>-1</sup>	$K_m$ mM	$k_{cat}/K_m$ s <sup>-1</sup> mM <sup>-1</sup>
AncLLysO	21.7 ± 0.9	0.3 ± 0.03	72.3
Residues that are located at O <sub>2</sub> -binding site			
K387A	N.D.	N.D.	N.D.
Residues that are located at substrate-recognition site			
R76A	N.D.	N.D.	N.D.
E383D	N.D.	N.D.	N.D.
Y268F	13.5 ± 1.4	1.0 ± 0.2	13.5
Y516F	9.3 ± 0.4	0.2 ± 0.03	46.5
Y516A	1.4 ± 0.05	4.1 ± 0.3	0.34
Residues that are located at plug loop			
G251A	10.4 ± 0.7	1.7 ± 0.3	6.1
G251P	1.1 ± 0.08	11.3 ± 1.3	0.09
Y253F	3.4 ± 0.2	0.15 ± 0.02	22.6
Y253A	0.21 ± 0.01	4.4 ± 0.3	0.05
Y254F	15.0 ± 1.1	0.6 ± 0.08	25
Y254A	6.0 ± 0.4	3.0 ± 0.5	2.0

The measurement of enzyme kinetic parameters was performed using two biological replicates per three technical replicates.

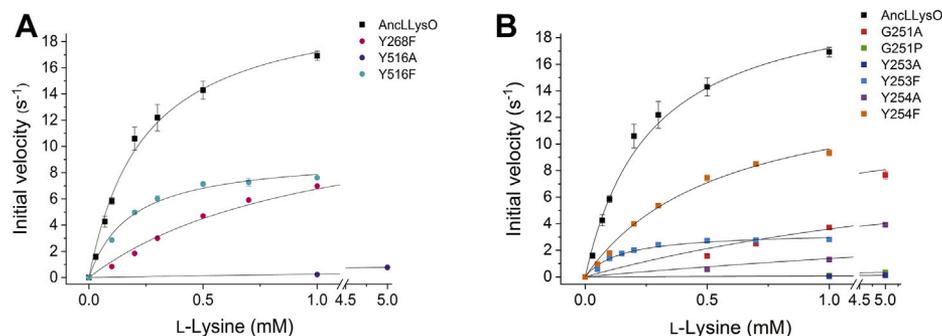


**Figure 5. Structure comparison at active site structures of ligand binding form of AncLLysO(K387A).** Active site structures for L-Lys- (A) and L-Arg- (B) binding form of AncLLysO(K387A). The L-Lys- and L-Arg-binding forms are *magenta* and *orange*, respectively. The polder Fo-Fc omit maps of L-Lys and L-Arg were contoured at 3.5 and 4.0  $\sigma$ , respectively. Structure comparison of L-Lys- and L-Arg-binding forms of AncLLysO(K387A) (C). Conformational change of plug loop and active site residues by binding of L-Lys (D). AncLLysO(LF)- and L-Lys-binding forms of AncLLysO(K387A) are *green* and *magenta*, respectively. Structural comparison at the active site among AncLLysO, AncLAAO-N5 (PDB ID: 7C4N), and VioA (PDB ID: 6FW8) (E). The structures of the L-Lys-binding forms of AncLLysO(K387A), AncLAAO-N5, and VioA are represented by *magenta*, *dark blue*, and *yellow*, respectively.

and enzyme kinetic plots and parameters are represented in Figure 6A and Table 4, respectively. This analysis indicated that mutation of residues that form hydrogen bonds with L-Lys inactivated AncLLysO. In fact, the activity of R76A and E383D variants was too low to estimate their parameters (Table 4). On the other hand, activity loss by the mutations would be attenuated if hydrophobic interactions are maintained after the mutations;  $k_{\text{cat}}/K_{\text{m}}$  values of Y268F and Y516F were approximately 20% and 55% that of AncLLysO (Table 4).

Next, an analysis of AncLLysO variants of the plug loop was performed. Kinetic plots and parameters are shown in Figure 6B and Table 4, respectively, suggesting that the side-chain phenyl groups of Y253 and Y254 are important for AncLLysO to express high activity. In fact,  $k_{\text{cat}}$  and  $k_{\text{cat}}/K_{\text{m}}$  values were largely decreased by mutating the residues to Ala compared with Phe (Table 4). In the variants, the decrease of the Y253A variant was remarkable; the  $k_{\text{cat}}$  and  $k_{\text{cat}}/K_{\text{m}}$  values were two and four orders lower than those of AncLLysO

## Structural and functional analysis of AncLLysO



**Figure 6. Enzyme kinetics plots of AncLLysO variants.** The following two types of variants were designed to estimate the enzymatic functional roles of AncLLysO; one is the residue that is located near the substrate (A) and the other is the residue located on the plug loop (B). All of the experiments were performed using two biological replicates per three technical replicates.

(Table 4). For the G251 variants, loss of activity could be confirmed as the residue was located separately from the active site. The decrease of the G251P variant was particularly apparent; the  $k_{\text{cat}}$  and  $k_{\text{cat}}/K_{\text{m}}$  values were approximately 19- and 800-fold lower than those of AncLLysO (Table 4). Here, mutation of G251 to Pro increases the energy barrier required to switch between the *cis* and *trans* forms; in fact, the energy required to induce *cis-trans* isomerization of Pro was estimated to be approximately 20 kJ/mol (48). This suggested that G251P would fix its conformation only in the *cis* or *trans* form. In this situation, the conformational change at the plug loop observed in Figure 5D was suppressed, and recognition of L-Lys and/or emission of the product may be inhibited in the G251P variant.

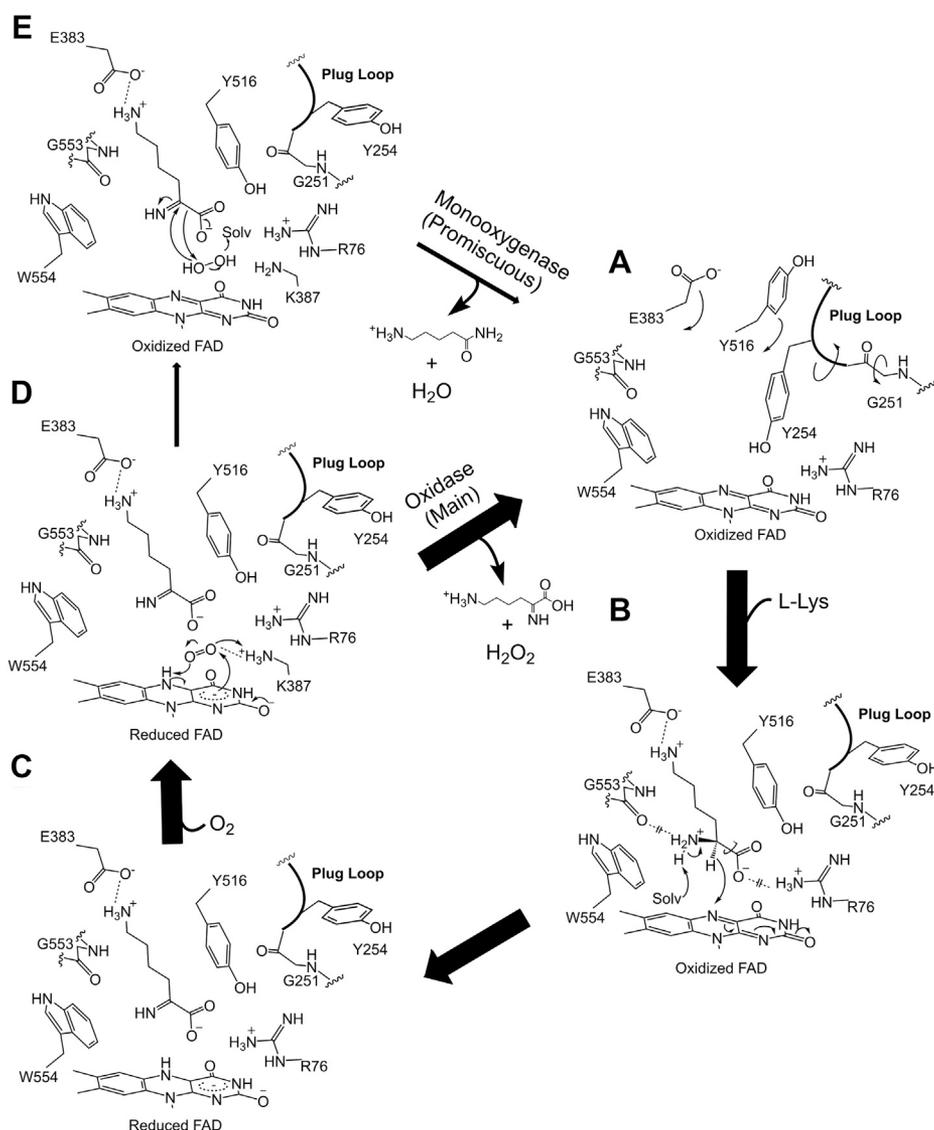
Enzyme kinetics of the monooxygenase activity of AncLLysO can be estimated by quantifying the amount of 5-APNM produced by the aminoamide oxidizing enzyme from *Aspergillus carbonarius* AIU 205 (AcAOx) (49). Referring to the previous study (40), enzyme kinetics for the monooxygenase activity of AncLLysO were assayed utilizing the purified AcAOx as the quantifying enzymes of 5-APNM. However, the concentration of 5-APNM could not be quantified by the assay because the amount of 5-APNM produced by AncLLysO was too low to detect with AcAOx. In fact, for L-LOX/MOG, of which enzyme kinetic parameters could be estimated by the assay, the main product of the reaction with L-Lys was 5-APNM (85% for 5-APNM versus 0.32% for 5-APNA), whereas the main product for AncLLysO was 5-APNA (3.2% for 5-APNM versus 66% for 5-APNA, Fig. 3).

### Discussion

Summarizing the results for biochemical and structural analysis of AncLLysO and their variants, we propose the reaction mechanism shown in Figure 7. In the ligand-free form, a cavity at the active site to recognize substrates is occupied by the side-chain Y254, which is the residue forming the plug loop (A in Fig. 7); this occupation would bring about structural changes of the substrate recognition residues, such as E383 and Y516 and made it difficult to predict the recognition mechanism solely from the structure of the AncLLysO ligand-free form. Rearrangement of the active site structure would be

induced by the binding of L-Lys. Specifically, the side-chain Y254 moved away from the active site, associating with the *cis-trans* transformation of G251, and simultaneously the side chains E383 and Y516 flipped into the active site to recognize L-Lys (A to B in Fig. 7). The  $k_{\text{cat}}/K_{\text{m}}$  value of Y516A decreased as the  $K_{\text{m}}$  value increased. This did not occur with the Y516F variant (Table 4), suggesting that the side-chain phenyl group of Y516 is important to form the substrate-binding site that recognizes L-Lys more efficiently. The functional importance of the structural change of the plug loop was shown by enzyme kinetics analysis (Table 4). All of the variants on the plug loop decreased the  $k_{\text{cat}}/K_{\text{m}}$  value compared with AncLLysO despite no residues forming interactions with L-Lys in the ligand-binding form (Table 4 and Fig. 5).

AncLLysO oxidizes L-Lys via an identical mechanism already described for LAOs (4). The main-chain amino group of L-Lys is deprotonated by an activated solvent water molecule, and the hydride on the main-chain carbon atom is transferred to FAD (B to C in Fig. 7). The oxygen molecule would bind to the active site by forming an interaction with the side-chain amino group of K387, and the oxidation of reduced FAD progressed continuously (D in Fig. 7). To predict the oxidation mechanism, we referred to the study of D-amino acid oxidase, which catalyzes oxidation of D-amino acids through a mechanism similar to AncLLysO (50). From the LC-HRMS analysis, the produced imino acid is converted to 5-APNA or 5-APNM via a different mechanism. For 5-APNA production, the imino acid and  $\text{H}_2\text{O}_2$  are released to the solvent without reacting to each other (D to A in Fig. 7). The released product is hydrolyzed and decarboxylated to 5-APNA. The mechanism generally corresponds to that of LAOs (4). On the other hand, 5-APNM is produced by the monooxygenase activity of AncLLysO via a mechanism similar to that of L-LOX/MOG (41). Utilizing  $\text{H}_2\text{O}_2$ , imino acid is decarboxylated to 5-APNM at the active site (D to E in Fig. 7). Trisrivirat *et al.*, (41) reported that a plug loop is required to exhibit the activity in L-LOX/MOG, and AncLLysO also bears this plug loop (251-GGYY-254). Although enzyme kinetic parameters for monooxygenase activity of AncLLysO and their variants could not be determined because of their promiscuous nature, the mutation at the loop may affect the monooxygenase activity. Here, AncLLysO shared low sequence



**Figure 7. Proposed reaction mechanism of AncLLysO for L-Lys  $\alpha$ -oxidase and monooxygenase activity.** In this scheme, ligand free, ligand binding, and product binding form were represented as A, B, and C, respectively. After oxidation of an FADH (D), the product imino acid and  $\text{H}_2\text{O}_2$  were released into the solvent (D to A); this is main oxidase activity. On the other hand, 5-APNM and  $\text{H}_2\text{O}$  were released into the solvent via state E; this is promiscuous monooxygenase activity confirmed in AncLLysO.

identity with L-LOX/MOG (<20%), but they exhibited similar activity. In other LAOs, the plug loop also exists (filled circle in Fig. S2), suggesting that monooxygenase activity may be confirmed in other LAOs that belong to an evolutionally distant family, similar to the case of AncLLysO and L-LOX/MOG.

Currently, four enzymes that belong to the FAO superfamily and exhibit LAO activity are reported. These include AncLAAO (26, 28), VioA (31), AROD (27, 30), and AncLLysO. These enzymes share low sequence identity to each other and to other proteins in the superfamily registered in the PDB database (<30%). Their substrate selectivity and specificity were clearly different from each other. However, their overall structures were highly conserved and the enzymes could oxidize common substrates. For example, L-Lys and L-Arg were oxidized by AncLAAO, AROD, and AncLLysO, and L-Trp was oxidized by VioA and AncLAAO. Based on

previous research of enzyme evolution (51), we predicted that the four enzymes would be derived from a common ancestor, which may exhibit broad substrate selectivity toward L-amino acids, like AncLAAO. The promiscuous enzyme functions of the current enzymes would be a functional vestige of the progenitor enzymes; therefore, it is reasonable to assume that the common ancestors of the four enzymes bear monooxygenase activity similar to AncLLysO. These enzymes could be investigated in the near future.

## Conclusion

Taken together, we assigned a novel FAD-dependent enzyme, called LLysO and bearing a main L-Lys  $\alpha$ -oxidase and promiscuous monooxygenase activity, from the database. Here, LLysO has a substrate specificity more specifically toward L-Lys than the L-LOX/MOG. The  $k_{\text{cat}}/K_m$  value of L-

## Structural and functional analysis of AncLLysO

LOX/MOG was equivalent to L-Lys and L-Orn (40), whereas, for AncLLysO and CaLLysO, the  $k_{\text{cat}}/K_m$  values toward L-Lys were greater than 15-fold higher than the values toward L-Orn (Table 1). The assignment of LLysO was achieved by the combinational usage of the sequence classification method utilizing correlatively mutated residues as the motif and ancestral sequence reconstruction. Although sequence identity among AncLLysO, L-LOX/MOG, and TvLLysO was quite low (<30%), the interaction mode is highly conserved with L-Lys. The acidic amino acids (Asp or Glu) and Arg at the active site formed a hydrogen bond with the side-chain amino group and the main-chain carboxyl group of L-Lys, respectively (Fig. S4, A–C). This suggests that, as in the case of the assignment of AncLLysO, there is a chance to find new LAAOs that have unique sequences compared with previously reported LAAOs.

AncLLysO has several characteristics that enable its enzymatic functions to be revealed by experimental approaches. These include high thermostability, resilience against mutations, and ease of crystallization, all of which allowed us to reveal the catalytic mechanism of AncLLysO at the molecular level. In the future, new enzymes could be screened from databases in a similar manner and their functions characterized by applying the combinational approach reported in this study.

### Experimental procedures

#### Reconstruction of AncLLysO

A total of 28 sequences (Table S1) that were selected by applying the procedure described in Figure 1 and one sequence bearing low identity (approximately 20%) compared with the selected sequences were aligned by MAFFT software (52). The aligned sequences were analyzed by MEGA6 (53), and phylogenetic tree data were prepared using the maximum-likelihood method. The aligned sequences and the tree data were submitted to the FastML web server (54) and JTT empirical models were adopted for the analysis. Finally, we selected a common ancestral sequence generated by the FastML as AncLLysO.

#### Overexpression and purification of AncLLysO and CaLLysO

Plasmids containing AncLLysO and CaLLysO were digested with *NcoI* and *XhoI*, and corresponding DNA sequences were subcloned into the pET28a vector, which were cut with the same two restriction enzymes. DNA encoding LLysOs was synthesized by GENEWIZ. The produced expression plasmids were transformed into the *E. coli* strain BL21(DE3). The strain was cultivated in 1 l of LB broth containing 30 µg/ml of kanamycin at 37 °C. The temperature was lowered to 18 °C when the  $A_{600}$  value reached 0.6 to 0.8, and then isopropyl- $\beta$ -D-thiogalactopyranoside was added to a final concentration of 0.5 mM. After the strains were cultivated overnight, they were collected by centrifugation. The collected cells were suspended into bufferA (20 mM Tris-HCl [pH 8.0] and 10 mM NaCl). After sonication of the cells, the supernatant was collected with centrifugation at 11,000g for 40 min. The supernatant was applied to a HisTrap-HP column (GE Healthcare), and the column was washed with 30 ml of bufferA containing 10 mM

imidazole. The samples were eluted by 15 ml of bufferA containing 30 mM imidazole. The samples were applied to a MonoQ column (GE Healthcare) that was equilibrated by bufferA and purified by linear gradient utilizing bufferA and bufferB (20 mM Tris-HCl [pH 8.0] and 500 mM NaCl) as elution buffer. The fractions exhibiting the highest  $A_{450}/A_{280}$  ratio were collected and concentrated to 500 µl. The concentrated samples were applied to a Superdex 200pg column that was equilibrated with bufferA; the purity was confirmed by SDS-PAGE. The purified samples were utilized in subsequent analysis.

UV-visible spectra of AncLLysO and the variants were measured to estimate FAD contents using the  $A_{280}/A_{450}$  ratio for each sample. The analysis indicated that, for all of the AncLLysO variants generated in this study, the contents reached greater than 80%, indicating that differences in contents had a minimal effect on the relative comparison of enzyme kinetics parameters (Table 4).

#### Site-directed mutagenesis of AncLLysO

Plasmids containing AncLLysO cloned into pET28b were utilized as a template. Site-directed mutagenesis was performed utilizing QuikChange Lightning Multi-site mutagenesis kit (Agilent Technologies). Primers utilized to design the variants are listed in Table S3. Sequence confirmation of the AncLLysO variants was performed by DNA sequencing.

#### Analysis of substrate selectivity, thermal stability, and enzyme kinetics of AncLAAO and CaLLysO

Oxidase activity of CaLLysO, AncLLysO, and their variants was measured by quantifying the concentration of  $\text{H}_2\text{O}_2$  produced by the enzymatic reaction utilizing a color-developed method. The components of the assay buffer were as follows: 10 mM amino acid, 1.5 mM 4-aminoantipyrine, 2 mM phenol, 50 U/ml horseradish peroxidase, and 100 mM buffer. The following four types of buffers and 10 mM L-Lys were utilized in the assay to estimate optimal pH value: sodium acetate (pH 3.5–4.5), Bis-Tris-HCl (pH 6.0–7.0), Tris-HCl (pH 7.0–8.5), and BICIN (pH 9.0). Substrate selectivity was estimated utilizing the following amino acid and buffer: 10 mM amino acids and 100 mM Bis-Tris-HCl (pH 7.0). Thermostability was measured utilizing the assay buffer containing 10 mM L-Lys and 100 mM Bis-Tris-HCl (pH 7.0). The initial velocity of AncLLysO and CaLLysO was calculated by monitoring the time-dependent absorption change at 505 nm, which was derived from the produced pigment bearing  $\epsilon_{505} = 12,700 \text{ M}^{-1} \text{ cm}^{-1}$  with UV-visible spectrometer (UV-2450, Shimadzu).

The kinetic parameters of CaLLysO, AncLLysO, and their variants toward L-Lys, L-Arg, and L-ornithine (L-Orn) were measured under the conditions containing the following concentrations of substrates: 0.1 to 1.0 mM L-Lys, 1.0 to 10 mM L-Arg, and 5.0 to 50 mM L-Orn, respectively. A procedure identical to the measurement of substrate selectivity was applied to determine initial velocity. The enzyme kinetic parameters were estimated by fitting the initial velocity to the

Michaelis–Menten equation with the nonlinear least-squares method by ORIGIN software; the parameters are represented in Table 1. All of the experiments were performed using two biological replicates per three technical replicates.

### Analysis of products by LC-HRMS

LC-HRMS analysis was performed using Q Exactive (Thermo Fisher Scientific), equipped with an electrospray ionization module. Here, the following columns, which are joined to the system, were utilized to detect the products by LC-HRMS: a UPLC column (XBridge BEH Amide XP column [length,  $2.1 \times 50 \text{ mm}^2$ ; inner diameter (i.d.),  $2.5 \mu\text{m}$ ; Nihon Waters K.K.]) equipped with a guard column (XBridge BEH Amide XP VanGuard cartridge [length,  $2.1 \times 5 \text{ mm}^2$ ; i.d.,  $2.5 \mu\text{m}$ ; Nihon Waters K.K.]). The column was kept at  $40^\circ\text{C}$ . The volume of injected samples was  $1 \mu\text{l}$ . The following two solutions were utilized as the mobile phase: 5 mM ammonium formate and 90% (v/v) acetonitrile (solution A) and 5 mM ammonium formate and 50% (v/v) acetonitrile (solution B). The flow rate was set to 0.4 ml/min, and the products were eluted by applying the following conditions: (1) 0% B for 1 min, (2) 0% to 100% B over 4 min, (3) 100% B for 2 min, and (4) 0% B for 5 min.

### Crystallization and X-ray data collection of AncLLysO

The purified AncLLysO was concentrated to about 15 mg/ml by centrifugation. Crystallization of AncLLysO for phase determination was performed as follows. After mixing a total of  $1.5 \mu\text{l}$  of the concentrated AncLLysO samples with  $1.0 \mu\text{l}$  of reservoir solution, composed of 12.5% (w/v) PEG3350, 0.1 M HEPES-NaOH (pH 7.5), and 0.2 M ammonium sulfate, the AncLLysO crystals appeared at  $22^\circ\text{C}$ . The crystals were soaked in a cryoprotectant reservoir quickly (25% [w/v] PEG3350, 0.1 M HEPES-NaOH [pH 7.5], 0.2 M ammonium sulfate, 20% [v/v] glycerol, and 0.2 M NaI), and the crystals were flash-cooled under a liquid nitrogen stream (100 K). X-ray diffraction data were collected using a Pilatus3 detector instrument at a BL5A beamline in the Photon Factory. Integration and scaling of the data were performed by XDS (55) and SCALA (56), respectively. The initial phase determination was achieved by the iodide single anomalous dispersion method. AutoSol, implemented in PHENIX software (57), assigned a total of 81 anomalous sites by analyzing the data. Model building was performed by AutoBuild (57) and Coot (58), and the initial structure for the ligand-free form of AncLLysO could be obtained.

Crystallization of the L-Lys- and L-Arg-binding forms of AncLLysO was performed by following procedures utilizing the AncLLysO(K387A) variant as a sample. The samples were concentrated to approximately 15 mg/ml. Samples that bind L-Lys or L-Arg to the active site were prepared by mixing a total of  $90 \mu\text{l}$  of the concentrated samples with  $10 \mu\text{l}$  of 100 mM L-Lys and 100 mM L-Arg, respectively. Crystals could be obtained by mixing a total of  $1.5 \mu\text{l}$  of the samples with  $1.0 \mu\text{l}$  of reservoir solution (25% [w/v] PEG3350 and 0.1 M Tris-HCl [pH 8.5]) and incubating at  $22^\circ\text{C}$ . The crystals were soaked

in a cryoreservoir (25% [w/v] PEG3350, 0.1 M Tris-HCl [pH 8.6], 10 mM NaCl, and 20% [v/v] ethylene glycol) containing 10 mM L-Lys or L-Arg for 60 min. The soaked crystals were flash-cooled under a cryonitrogen stream (100 K). X-ray diffraction data were collected at BL5A in the Photon Factory. Data integration and scaling were performed by XDS and SCALA, respectively, and the phase was determined by the MOLREP software (59) utilizing the structure of the ligand-free form of AncLLysO as a template. Model building and refinement were performed by COOT (58) and either REFMAC (60) or PHENIX (57), respectively. All figures were prepared by PyMOL (61). Crystallographic parameters are represented in Table 2.

### Data availability

All data are contained within this article.

**Supporting information**—This article contains [supporting information](#).

**Acknowledgments**—X-ray data were collected at the synchrotron facilities of the Photon Factory (PF) using beamline BL-5A (2018G006). The authors are grateful to the beamline staff for their assistance with the experiments performed at the Photon Factory. The authors express their appreciation to Prof. M. Yamada and Prof. K. Isobe for providing the *Aspergillus carbonarius* AIU 205 strain. This work was supported by JSPS KAKENHI Grant Numbers 16K18688 and 18K14391 and by JST, PRESTO, JSP Grant Number JPMJPR20AB.

**Author contributions**—S. N. conceptualization; S. S., S. N., and D. M. methodology; S. S. and S. N. validation; S. S. and S. N. formal analysis; S. N., M. N., D. M., S. I., and F. H. investigation; S. S., S. N., M. N., and F. H. data curation; S. N. writing – original draft; S. S. and S. N. visualization; S. N. and S. I. supervision; S. N. project administration; S. N. funding acquisition.

**Conflict of interest**—The authors declare that they have no conflicts of interest related to the contents of this article.

**Abbreviations**—The abbreviations used are: 5-APNA, 5-aminopentanoic acid; 5-APNM, 5-aminopentanamide; AncLLysO(LF), ligand-free form of AncLLysO; AncLLysO, ancestral LLysO; AROD, L-arginine oxidase; ASR, ancestral sequence reconstruction; FAO, flavin-dependent amine oxidase; L-LOX/MOG, L-amino acid oxidase/monooxygenase; LAAO, L-amino acid oxidase; LLysO, L-lysine  $\alpha$ -oxidase.

### References

- Xue, Y. P., Cao, C. H., and Zheng, Y. G. (2018) Enzymatic asymmetric synthesis of chiral amino acids. *Chem. Soc. Rev.* **47**, 1516–1561
- Baxter, S., Royer, S., Grogan, G., Brown, F., Holt-Tiffin, K. E., Taylor, I. N., Fotheringham, I. G., and Campopiano, D. J. (2012) An improved racemase/acylase biotransformation for the preparation of enantiomerically pure amino acids. *J. Am. Chem. Soc.* **134**, 19310–19313
- Paradisi, F., Collins, S., Maguire, A. R., and Engel, P. C. (2007) Phenylalanine dehydrogenase mutants: Efficient biocatalysts for synthesis of non-natural phenylalanine derivatives. *J. Biotechnol.* **128**, 408–411
- Pollegioni, L., Motta, P., and Molla, G. (2013) L-amino acid oxidase as biocatalyst: A dream too far? *Appl. Microbiol. Biotechnol.* **97**, 9323–9341

## Structural and functional analysis of AnCLLysO

- Hossain, G. S., Li, J., Shin, H. D., Du, G., Liu, L., and Chen, J. (2014) L-amino acid oxidases from microbial sources: Types, properties, functions, and applications. *Appl. Microbiol. Biotechnol.* **98**, 1507–1515
- Yang, H., Johnson, P. M., Ko, K. C., Kamio, M., Germann, M. W., Derby, C. D., and Tai, P. C. (2005) Cloning, characterization and expression of escapin, a broadly antimicrobial FAD-containing L-amino acid oxidase from ink of the sea hare *Aplysia californica*. *J. Exp. Biol.* **208**, 3609–3622
- Moustafa, I. M., Foster, S., Lyubimov, A. Y., and Vrieling, A. (2006) Crystal structure of LAAO from *Calloselasma rhodostoma* with an L-phenylalanine substrate: Insights into structure and mechanism. *J. Mol. Biol.* **364**, 991–1002
- Faust, A., Niefind, K., Hummel, W., and Schomburg, D. (2007) The structure of a bacterial L-amino acid oxidase from *Rhodococcus opacus* gives new evidence for the hydride mechanism for dehydrogenation. *J. Mol. Biol.* **367**, 234–248
- Gomez, D., Espinosa, E., Bertazzo, M., Lucas-Elio, P., Solano, F., and Sanchez-Amat, A. (2008) The macromolecule with antimicrobial activity synthesized by *Pseudoalteromonas luteoviolacea* strains is an L-amino acid oxidase. *Appl. Microbiol. Biotechnol.* **79**, 925–930
- Hahn, K., Neumeister, K., Mix, A., Kottke, T., Groger, H., and Fischer von Mollard, G. (2017) Recombinant expression and characterization of a L-amino acid oxidase from the fungus *Rhizoctonia solani*. *Appl. Microbiol. Biotechnol.* **101**, 2853–2864
- Alexandre, F.-R., Pantaleone, D. P., Taylor, P. P., Fotheringham, I. G., Ager, D. J., and Turner, N. J. (2002) Amine–boranes: Effective reducing agents for the deracemisation of DL-amino acids using L-amino acid oxidase from *Proteus myxofaciens*. *Tetrahedron Lett.* **43**, 707–710
- Molla, G., Melis, R., and Pollegioni, L. (2017) Breaking the mirror: L-amino acid deaminase, a novel stereoselective biocatalyst. *Biotechnol. Adv.* **35**, 657–668
- Rosini, E., Melis, R., Molla, G., Tessaro, D., and Pollegioni, L. (2017) Deracemization and stereoinversion of  $\alpha$ -amino acids by L-amino acid deaminase. *Adv. Synth. Catal.* **359**, 3773–3781
- Motta, P., Pollegioni, L., and Molla, G. (2019) Properties of L-amino acid deaminase: En route to optimize bioconversion reactions. *Biochimie* **158**, 199–207
- Parmeggiani, F., Rué Casamajo, A., Walton, C. J. W., Galman, J. L., Turner, N. J., and Chica, R. A. (2019) One-pot biocatalytic synthesis of substituted d-tryptophans from indoles enabled by an engineered aminotransferase. *ACS Catal.* **9**, 3482–3486
- Nasu, S., Wicks, F. D., and Gholson, R. K. (1982) L-aspartate oxidase, a newly discovered enzyme of *Escherichia coli*, is the B protein of quinoxaline synthetase. *J. Biol. Chem.* **257**, 626–632
- Mattevi, A., Tedeschi, G., Bacchella, L., Coda, A., Negri, A., and Ronchi, S. (1999) Structure of L-aspartate oxidase: Implications for the succinate dehydrogenase/fumarate reductase oxidoreductase family. *Structure* **7**, 745–756
- Arima, J., Sasaki, C., Sakaguchi, C., Mizuno, H., Tamura, T., Kashima, A., Kusakabe, H., Sugio, S., and Inagaki, K. (2009) Structural characterization of L-glutamate oxidase from *Streptomyces* sp. X-119-6. *FEBS J.* **276**, 3894–3903
- Utsumi, T., Arima, J., Sakaguchi, C., Tamura, T., Sasaki, C., Kusakabe, H., Sugio, S., and Inagaki, K. (2012) Arg305 of *Streptomyces* L-glutamate oxidase plays a crucial role for substrate recognition. *Biochem. Biophys. Res. Commun.* **417**, 951–955
- Gumulya, Y., and Gillam, E. M. (2017) Exploring the past and the future of protein evolution with ancestral sequence reconstruction: The 'retro' approach to protein engineering. *Biochem. J.* **474**, 1–19
- Merkel, R., and Sterner, R. (2016) Ancestral protein reconstruction: Techniques and applications. *Biol. Chem.* **397**, 1–21
- Akanuma, S., Nakajima, Y., Yokobori, S., Kimura, M., Nemoto, N., Mase, T., Miyazono, K., Tanokura, M., and Yamagishi, A. (2013) Experimental evidence for the thermophilicity of ancestral life. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 11067–11072
- Furukawa, R., Toma, W., Yamazaki, K., and Akanuma, S. (2020) Ancestral sequence reconstruction produces thermally stable enzymes with mesophilic enzyme-like catalytic properties. *Sci. Rep.* **10**, 15493
- Clifton, B. E., and Jackson, C. J. (2016) Ancestral protein reconstruction yields insights into adaptive evolution of binding specificity in solute-binding proteins. *Cell Chem. Biol.* **23**, 236–245
- Spence, M. A., Kaczmarek, J. A., Saunders, J. W., and Jackson, C. J. (2021) Ancestral sequence reconstruction for protein engineers. *Curr. Opin. Struct. Biol.* **69**, 131–141
- Nakano, S., Minamino, Y., Hasebe, F., and Ito, S. (2019) Deracemization and stereoinversion to aromatic d-amino acid derivatives with ancestral L-amino acid oxidase. *ACS Catal.* **9**, 10152–10158
- Nakano, S., Niwa, M., Asano, Y., and Ito, S. (2019) Following the evolutionary track of a highly specific L-arginine oxidase by reconstruction and biochemical analysis of ancestral and native enzymes. *Appl. Environ. Microbiol.* **85**, e00459-19
- Nakano, S., Kozuka, K., Minamino, Y., Karasuda, H., Hasebe, F., and Ito, S. (2020) Ancestral L-amino acid oxidases for deracemization and stereoinversion of amino acids. *Commun. Chem.* **3**, 181
- Tararina, M. A., and Allen, K. N. (2020) Bioinformatic analysis of the flavin-dependent amine oxidase superfamily: Adaptations for substrate specificity and catalytic diversity. *J. Mol. Biol.* **432**, 3269–3288
- Matsui, D., Terai, A., and Asano, Y. (2016) L-arginine oxidase from *Pseudomonas* sp. TPU 7192: Characterization, gene cloning, heterologous expression, and application to L-arginine determination. *Enzyme Microb. Technol.* **82**, 151–157
- Fuller, J. J., Ropke, R., Krausz, J., Rennhack, K. E., Daniel, N. P., Blankenfeldt, W., Schulz, S., Jahn, D., and Moser, J. (2016) Biosynthesis of violacein, structure and function of L-tryptophan oxidase VioA from *Chromobacterium violaceum*. *J. Biol. Chem.* **291**, 20068–20084
- Kameya, M., Onaka, H., and Asano, Y. (2013) Selective tryptophan determination using tryptophan oxidases involved in bis-indole antibiotic biosynthesis. *Anal. Biochem.* **438**, 124–132
- Yamaguchi, H., Tatsumi, M., Takahashi, K., Tagami, U., Sugiki, M., Kashiwagi, T., Kameya, M., Okazaki, S., Mizukoshi, T., and Asano, Y. (2018) Protein engineering for improving the thermostability of tryptophan oxidase and insights from structural analysis. *J. Biochem.* **164**, 359–367
- Nakano, S., Motoyama, T., Miyashita, Y., Ishizuka, Y., Matsuo, N., Tokiwa, H., Shinoda, S., Asano, Y., and Ito, S. (2018) Benchmark analysis of native and artificial NAD(+)-dependent enzymes generated by a sequence-based design method with or without phylogenetic data. *Biochemistry* **57**, 3722–3732
- Kido, K., Yamanaka, S., Nakano, S., Motani, K., Shinohara, S., Nozawa, A., Kosako, H., Ito, S., and Sawasaki, T. (2020) AirID, a novel proximity biotinylation enzyme, for analysis of protein-protein interactions. *Elife* **9**, e54983
- Motoyama, T., Hiramatsu, N., Asano, Y., Nakano, S., and Ito, S. (2020) Protein sequence selection method that enables full consensus design of artificial L-threonine 3-dehydrogenases with unique enzymatic properties. *Biochemistry* **59**, 3823–3833
- Amano, M., Mizuguchi, H., Sano, T., Kondo, H., Shinyashiki, K., Inagaki, J., Tamura, T., Kawaguchi, T., Kusakabe, H., Imada, K., and Inagaki, K. (2015) Recombinant expression, molecular characterization and crystal structure of antitumor enzyme, L-lysine  $\alpha$ -oxidase from *Trichoderma viride*. *J. Biochem.* **157**, 549–559
- Kondo, H., Kitagawa, M., Matsumoto, Y., Saito, M., Amano, M., Sugiyama, S., Tamura, T., Kusakabe, H., Inagaki, K., and Imada, K. (2020) Structural basis of strict substrate recognition of L-lysine  $\alpha$ -oxidase from *Trichoderma viride*. *Protein Sci.* **29**, 2213–2225
- Matsui, D., Im, D. H., Sugawara, A., Fukuta, Y., Fushinobu, S., Isobe, K., and Asano, Y. (2014) Mutational and crystallographic analysis of L-amino acid oxidase/monooxygenase from *Pseudomonas* sp. AIU 813: Interconversion between oxidase and monooxygenase activities. *FEBS Open Bio* **4**, 220–228
- Im, D., Matsui, D., Arakawa, T., Isobe, K., Asano, Y., and Fushinobu, S. (2018) Ligand complex structures of L-amino acid oxidase/monooxygenase from *Pseudomonas* sp. AIU 813 and its conformational change. *FEBS Open Bio* **8**, 314–324
- Trisrivirat, D., Lavan, N., Chenprakhon, P., Matsui, D., Asano, Y., and Chaiyen, P. (2020) Mechanistic insights into the dual activities of the

- single active site of L-lysine oxidase/monooxygenase from *Pseudomonas* sp. AIU 813. *J. Biol. Chem.* **295**, 11246–11261
42. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402
  43. Jäckel, C., Bloom, J. D., Kast, P., Arnold, F. H., and Hilvert, D. (2010) Consensus protein design without phylogenetic bias. *J. Mol. Biol.* **399**, 541–546
  44. Isobe, K., Matsui, D., and Asano, Y. (2019) Comparative review of the recent enzymatic methods used for selective assay of L-lysine. *Anal. Biochem.* **584**, 113335
  45. Okazaki, S., Nakano, S., Matsui, D., Akaji, S., Inagaki, K., and Asano, Y. (2013) X-ray crystallographic evidence for the presence of the cysteine tryptophylquinone cofactor in L-lysine epsilon-oxidase from *Marinomonas mediterranea*. *J. Biochem.* **154**, 233–236
  46. Holm, L., and Rosenstrom, P. (2010) Dali server: Conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549
  47. McDonald, C. A., Fagan, R. L., Collard, F., Monnier, V. M., and Palfey, B. A. (2011) Oxygen reactivity in flavoenzymes: Context matters. *J. Am. Chem. Soc.* **133**, 16809–16811
  48. Odefey, C., Mayr, L. M., and Schmid, F. X. (1995) Non-prolyl cis-trans peptide bond isomerization as a rate-determining step in protein unfolding and refolding. *J. Mol. Biol.* **245**, 69–78
  49. Sugawara, A., Matsui, D., Komeda, H., Asano, Y., and Isobe, K. (2014) Characterization and application of aminoamide-oxidizing enzyme from *Aspergillus carbonarius* AIU 205. *J. Biosci. Bioeng.* **117**, 263–268
  50. Saam, J., Rosini, E., Molla, G., Schulten, K., Pollegioni, L., and Ghisla, S. (2010) O<sub>2</sub> reactivity of flavoproteins: Dynamic access of dioxygen to the active site and role of a H<sup>+</sup> relay system in D-amino acid oxidase. *J. Biol. Chem.* **285**, 24439–24446
  51. Khersonsky, O., and Tawfik, D. S. (2010) Enzyme promiscuity: A mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* **79**, 471–505
  52. Yamada, K. D., Tomii, K., and Katoh, K. (2016) Application of the MAFFT sequence alignment program to large data-reexamination of the usefulness of chained guide trees. *Bioinformatics* **32**, 3246–3251
  53. Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729
  54. Ashkenazy, H., Penn, O., Doron-Faigenboim, A., Cohen, O., Cannarozzi, G., Zomer, O., and Pupko, T. (2012) FastML: A web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res.* **40**, W580–W584
  55. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132
  56. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82
  57. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221
  58. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501
  59. Vagin, A., and Teplyakov, A. (1997) MOLREP: An automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025
  60. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367
  61. Delano, W. L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA