Following the Evolutionary Track of a Highly Specific L-Arginine Oxidase by Reconstruction and Biochemical Analysis of Ancestral and Native Enzymes

Shogo Nakano,^{a,c} Masazumi Niwa,^a Yasuhisa Asano,^{b,c} Sohei Ito^a

^aGraduate Division of Nutritional and Environmental Sciences, University of Shizuoka, Shizuoka, Japan

^bBiotechnology Research Center and Department of Biotechnology, Toyama Prefectural University, Imizu, Toyama, Japan

^cAsano Active Enzyme Molecule Project, Exploratory Research for Advanced Technology, Japan Science and Technology Agency, Imizu, Toyama, Japan

ABSTRACT Following the evolutionary track of enzymes can help elucidate how enzymes attain their characteristic functions, such as thermal adaptation and substrate selectivity, during the evolutionary process. Ancestral sequence reconstruction (ASR) is effective for following evolutionary processes if sufficient sequence data are available. Selecting sequences from the data to generate a curated sequence library is necessary for the successful design of artificial proteins by ASR. In this study, we tried to follow the evolutionary track of L-arginine oxidase (AROD), a flavin adenine dinucleotide (FAD)-dependent amino acid oxidase (LAAO) that exhibits high specificity for L-arginine. The library was generated by selecting sequences in which the 15th, 50th, 332nd, and 580th residues are Gly, Ser, Trp, and Thr, respectively. We excluded sequences that are either extremely short or long and those with a low degree of sequence identity. Three ancestral ARODs (AncARODn0, AncARODn1, and AncARODn2) were designed using the library. Subsequently, we expressed the ancestral ARODs as well as native *Oceanobacter kriegii* AROD (OkAROD) in bacteria. AncARODn0 is phylogenetically most remote from OkAROD, whereas AncARODn2 is most similar to OkAROD. Thermal stability was gradually increased by extending AROD sequences back to the progenitor, while the temperature at which the residual activity is half of the maximum measured activity ($T_{1/2}$) of AncARODn0 was >20°C higher than that of OkAROD. Remarkably, only AncARODn0 exhibited broad substrate selectivity similar to that of conventional promiscuous LAAO. Taken together, our findings led us to infer that AROD may have evolved from a highly thermostable and promiscuous LAAO.

IMPORTANCE In this study, we attempted to infer the molecular evolution of a recently isolated FAD-dependent L-arginine oxidase (AROD) that oxidizes L-arginine to 2-ketoarginine. Utilizing 10 candidate AROD sequences, we obtained a total of three ancestral ARODs. In addition, one native AROD was obtained by cloning one of the candidate ARODs. The candidate sequences were selected utilizing a curation method defined in this study. All the ARODs were successfully expressed in *Escherichia coli* for analysis of their biochemical functions. The catalytic activity of our bacterially expressed ancestral ARODs suggests that our ASR was successful. The ancestral AROD that is phylogenetically most remote from a native AROD has the highest thermal stability and substrate promiscuity. Our findings led us to infer that AROD evolved from a highly thermostable and promiscuous LAAO. As an application, we can design artificial ARODs with improved functions compared with those of native ones.

KEYWORDS L-arginine oxidase, ancestral sequence reconstruction

Citation Nakano S, Niwa M, Asano Y, 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. <https://doi.org/10.1128/AEM.00459-19>.

Editor Haruyuki Atomi, Kyoto University

Copyright © 2019 American Society for Microbiology. All Rights Reserved.

Address correspondence to Shogo Nakano, snakano@u-shizuoka-ken.ac.jp.

S.N. and M.N. contributed equally to this work.

Received 24 February 2019

Accepted 27 March 2019

Accepted manuscript posted online 12 April 2019

Published 30 May 2019

LAAO (L-amino acid oxidase) is an amino acid-metabolizing enzyme that catalyzes the oxidation of the main-chain amino group of L-amino acids, thereby generating α -keto acid and ammonia (1). Two types of LAAOs have been reported: one is flavin adenine dinucleotide (FAD) dependent (2–4) while the other is not FAD dependent (5–8). The former type of LAAO is the main target of this study. Therefore, we refer to the FAD-dependent type as LAAO. LAAOs with broad substrate selectivity, or promiscuous LAAOs, have been discovered in a diverse range of species, including bacteria (9), fungi (10), and vipers (11).

LAAOs are biotechnologically important enzymes. In fact, they could be used in bioconversion of pure L-amino acids to D-amino acids, which are precursors for the synthesis of pharmaceutical compounds (1). On the other hand, LAAOs bearing high substrate specificity, also referred to as specific LAAOs, have been reported. They are named depending on the difference in specificity, such as L-phenyl alanine oxidase (2), L-arginine oxidase (AROD) (12), and L-glutamate oxidase (3, 4). These enzymes are applied to the quantification of L-amino acid concentrations in various samples.

Specific and promiscuous LAAOs oxidize their substrates using a similar mechanism (11, 13, 14), suggesting that there is a common evolutionary relationship between them. However, the evolutionary relationship between the two types of LAAOs is unclear because they share a low degree of sequence identity. We may obtain clues to find new LAAOs with novel specificities by determining when specific and promiscuous LAAOs are generated in the process of evolution (15, 16).

One question follows: how can we infer the molecular evolution of LAAOs? Ancestral sequence reconstruction (ASR) is one approach for answering that question. Evolutionary and phylogenetic contexts in enzymes can be analyzed by expression and biochemical analysis of ancestral sequences, which are generated at each node of the phylogenetic tree (17, 18). Several groups succeeded in reconstructing ancestral enzymes/proteins (19–26). The results indicated that ancestral enzymes have high thermal stability (20, 23) and substrate promiscuity (26). On the other hand, there have been several instances of ancestral sequences resulting in insoluble proteins when expressed, which made ASR difficult to perform. The insoluble expression of ancestral proteins could be caused by incorrect design when performing ASR (18). We believe that curation of candidate sequences based on certain criteria may help avoid this problem.

Development of a new curation method would help broaden the application of ASR to following molecular evolution, especially for proteins whose functions have not been well elucidated because of a lack of structural and mutational analysis data. In a previous study, we reported that the method of selecting sequences which have “key residues” from the noncurated library (27) is effective for designing a full consensus protein, which is an artificial protein designed by mutating all residues of the target protein to consensus residues. The “key residues” can be assigned utilizing only primary sequence data and can reduce the sum of the conservation energy (E_c), which was defined by Jäckel et al. (28). A full consensus protein generated by utilizing a curated library exhibited high solubility, thermal stability, and reactivity (27). The merit of curation is that sequences bearing low identity compared with the target protein can be omitted by using key residues (27). Excluding these sequences can improve the accuracy of multiple-sequence alignment (MSA), which is crucial to the successful design of full-length consensus proteins (29). Here, full consensus design and ASR are similar approaches (17) that both depend on MSA results. Therefore, we consider the curation method effective for performing ASR.

Based on our results, we postulate that the molecular evolution of various proteins can be inferred by ASR based on a sequence library generated by the curation method. In this study, we inferred the evolutionary track of specific LAAOs using AROD as an example. AROD, an orphan oxidase which was recently assigned by Matsui et al., can oxidize only L-Arg and L-Lys out of 20 L-amino acids (12). In addition, there are no kinetic data about mutants and crystal structures of AROD; instead, only sequence data are available. Through expression and biochemical analysis of three ancestral ARODs (AncARODn0, AncARODn1, and AncARODn2) and one native AROD (that from *Oceano-*

FIG 1 Schematic model illustrating the curation of the sequence library and the design of the ancestral ARODs AncARODn0, AncARODn1, and AncARODn2. First, sequences homologous to that of PtAROD were obtained through blastp analysis (procedure 1). After eliminating sequences that were either too long or too short, a sequence pair between PtAROD and one of the sequences in the library was prepared, and sequence alignment was performed (procedure 2). If the sequence had “key residues,” the sequence was selected (procedure 3), and this process was applied to all the sequences in the library. ASR was performed with FastML utilizing the curated library (procedure 4). Consequently, three ancestral ARODs (red circles) and one native AROD (OkAROD) (black circle) were generated.

bacter kriegii [OkAROD]), we tried to hypothesize about the evolutionary track of specific LAAOs such as AROD.

RESULTS

Ancestral AROD design. The ancestral sequence of ARODs must be reconstructed in order to perform molecular evolutionary analysis. Curation of AROD sequences has to be performed based only on sequence data. To tackle this difficult task, we performed the curation as described below.

A sequence library was prepared by submitting the AROD protein sequence from *Pseudomonas* sp. strain TPU 7192 (PtAROD) to the blastp web server (30) (Fig. 1, procedure 1). A total of 751 homologous sequences could be obtained with the E value set to 0.001. The GenBank accession numbers and taxa of the sequences obtained are listed in Table S1 in the supplemental material. Among these are more than 80 eukaryotic and 15 viral sequences which bear less than 35% sequence identity to PtAROD. Furthermore, there are extremely short (171 amino acids [aa] for GI no. 908496689) or long (1,154 aa for GI no. 551606006) sequences. These sequences made it difficult to obtain plausible MSA results. MSA results are crucial to the design of ancestral proteins, and therefore, it is necessary to exclude these sequences during curation. Thus, to perform curation, we applied the following approach. First, identical or irregular sequences were omitted using our original Python script that adopts the algorithm described in Fig. S1 in the supplemental material. The script calculates the sequence identity for all sequence pairs in the 751 homologous sequences, and sequences bearing >90% identity were deleted (Fig. S1). Finally, we were left with 306 sequences that we used as the library (Fig. 1). Next, library curation was performed in the following manner: (i) preparation of all pairs (a total of 306 pairs) formed by a PtAROD and one sequence in the library, (ii) sequence alignment of all the pairs, and (iii) selection of sequences bearing “key residues” (Fig. 1, procedures 2 and 3). Here, the key residues must be able to minimize the sum of conservation energy (E_c), defined by Jäcker et al. (28), after the curation. The E_c value is the sum of conservation energies for amino acid residues at the i th position in the target protein ($E_{aa,i}$):

$$E_{aa,i} = -\ln f_{aa,i}, \quad E_c = \sum_1^i E_{aa,i}$$

where $f_{aa,i}$ is the frequency of an amino acid appearing as a consensus residue at the i th position in PtAROD obtained by multiple-sequence alignment (MSA). In this study, four residues (the 15th, 50th, 332nd, and 580th residues, which were Gly, Ser, Trp, and Thr, respectively) were assigned as key residues, thereby allowing us to generate a curated library containing 10 sequences (see Fig. S2A [red bars] in the supplemental material). There were more than 700 sequences bearing low identity to PtAROD prior to curation (Fig. S2A, black bars), which suggests that our curation approaches can exclude these sequences, as we expected. Furthermore, the E_c value could be decreased by curation: the value was changed from 839.3 to 241.7 (Fig. S2B). The four residues were randomly selected to reduce the E_c value by selection, so the functional importance of these residues remains unknown.

Next, we attempted to acquire AROD homologous sequences by hidden Markov model (HMM) profiling with the HMMER suite (31). The *hmmsearch* implemented in HMMER was performed by submitting aligned sequences which were selected from 10 of the sequences most homologous to PtAROD and aligned by MAFFT. The E value was set to 0.001, and finally, we could obtain a total of 1,657 sequences. After application of the procedure indicated in Fig. S1 to the obtained sequences, a total of 42 homologous sequences could be obtained. Here, there are more than 30 sequences bearing low sequence identity (<35%) to PtAROD among the obtained sequences (Fig. S2C). The E_c value was 683.5 (Fig. S2D), and the value was >2-fold higher than that calculated utilizing the curated library based on four residues. Taking the results together, we consider that the curated approach suggested in this study can omit sequences bearing low sequence identity, which are hard to exclude by the HMMER approach.

In the curated library, we expressed a native AROD candidate from *Oceanobacter kriegii* (OkAROD). In addition, three ancestral ARODs (AncARODn0, AncARODn1, and AncARODn2 [red circles in Fig. 1]) were designed by utilizing FastML software (32); MSA was performed by MAFFT (33). The phylogenetic tree was generated with MEGA6 (34) using the maximum-likelihood (ML) method. Sequence information for the designed ARODs and OkAROD is listed in Table S2 in the supplemental material. The MSA for 10 homologous sequences (red in Table S1) and the three ancestral ARODs are shown in Fig. S3 in the supplemental material. The sequence identities of AncARODn0, AncARODn1, and AncARODn2 to OkAROD were 67%, 73%, and 78%, respectively. Phylogenetic analysis was performed on 139 sequences (see Fig. S4 in the supplemental material); the method for the selection of sequences and the generation of the tree is described in Materials and Methods. The phylogenetic analysis indicated that all of the AncARODs were positioned on the ancestral node (red in Fig. S4). Among the designed ARODs, AncARODn0 is phylogenetically most remote from a native AROD, OkAROD, whereas AncARODn2 is most phylogenetically similar to OkAROD (Fig. S4). Posterior probabilities (PPs) appear to be a criterion for whether reconstructed ancestral sequences provide a plausible estimation of the phenotypes of ancestral proteins (22, 23). The average PPs for AncARODn0, AncARODn1, and AncARODn2 were 0.88, 0.90, and 0.89, respectively (Fig. S4). These values were not inferior to those for other ancestral proteins (22, 23), suggesting that a plausible model to infer molecular evolution of ARODs can be designed.

Estimation of enzymatic properties of ancestral and native ARODs. The enzymatic activity of ancestral and native ARODs was analyzed. Measurement of substrate specificity toward the 20 L-amino acids indicated that all of the ARODs exhibit the highest activity toward L-Arg (Table 1), suggesting that the enzymes have AROD activity, as expected. The ARODs exhibit activity toward L-Lys, as does PtAROD. Remarkably, only AncARODn0 can oxidize other L-amino acids, i.e., L-His, L-Phe, L-Leu, L-Met, and L-Tyr (Table 1); we could not detect activity toward these L-amino acids for AncARODn1, AncARODn2, OkAROD, or PtAROD (12).

Next, biochemical functional analysis of ancestral and native ARODs was performed with other methods. The analysis of pH dependency indicated that the optimal pH

TABLE 1 Specific activity of ancestral ARODs (AncARODn0, AncARODn1, and AncARODn2) and a native AROD (OkAROD) toward the 20 L-amino acids

Substrate	Relative activity (%; mean $\pm$ SD)			
	AncARODn0	AncARODn1	AncARODn2	OkAROD
L-Arginine	100 $\pm$ 3.7	100 $\pm$ 1.7	100 $\pm$ 4.2	100 $\pm$ 5.1
L-Lysine	72.5 $\pm$ 2.4	7.5 $\pm$ 0.4	16.4 $\pm$ 0.2	15 $\pm$ 1.0
L-Histidine	4.3 $\pm$ 0.4	<0.0	<0.0	<0.0
L-Phenylalanine	3.6 $\pm$ 0.1	<0.0	<0.0	<0.0
L-Leucine	1.6 $\pm$ 0.2	<0.0	<0.0	<0.0
L-Methionine	0.8 $\pm$ 0.2	<0.0	<0.0	<0.0
L-Tryptophan	<0.0	<0.0	<0.0	<0.0
L-Isoleucine	<0.0	<0.0	<0.0	<0.0
L-Valine	<0.0	<0.0	<0.0	<0.0
Glycine	<0.0	<0.0	<0.0	<0.0
L-Proline	<0.0	<0.0	<0.0	<0.0
L-Alanine	<0.0	<0.0	<0.0	<0.0
L-Tyrosine	0.6 $\pm$ 0.0	<0.0	<0.0	<0.0
L-Cysteine	<0.0	<0.0	<0.0	<0.0
L-Glutamine	<0.0	<0.0	<0.0	<0.0
L-Asparagine	<0.0	<0.0	<0.0	<0.0
L-Serine	<0.0	<0.0	<0.0	<0.0
L-Threonine	<0.0	<0.0	<0.0	<0.0
L-Aspartic acid	<0.0	<0.0	<0.0	<0.0
L-Glutamic acid	<0.0	<0.0	<0.0	<0.0

value would be 7.0 in OkAROD and AncARODn2 and 7.5 in AncARODn0 and AncARODn1 (Fig. 2A). Several ancestral ARODs are resistant to basic pH; under basic conditions (pH > 10.0), the residual activity of AncARODn0 and AncARODn2 was more than 20% of the maximum activity (Fig. 2A).

Thermostability was estimated by analyzing the residual activity utilizing heat-treated samples (Fig. 2B). The estimated thermostability of the enzymes is in the following order: OkAROD ($T_{1/2}$ = 65°C) < AncARODn2 ($T_{1/2}$ = 72°C) < AncARODn1 ($T_{1/2}$ = 78°C) < AncARODn0 ($T_{1/2}$ = 88°C) (Fig. 2B), where $T_{1/2}$ was defined as the temperature at which the residual activity of the AROD is half of the maximum measured activity. The gradual improvement in thermostability may indicate the successful design of AncARODs with ASR; it is also possible that the improvement may

FIG 2 (A) pH dependency of ancestral (black) and native (red) ARODs. The activities of AncARODn0, AncARODn1, AncARODn2, and OkAROD are represented as filled squares, circles, upward-facing triangles, and downward-facing triangles, respectively. The relative activity was calculated by normalizing the highest activity as 100%. (B) Temperature-dependent activity changes for ancestral and native ARODs. The representation is identical to that in panel A. To estimate $T_{1/2}$, the dotted line was drawn at the point where the relative activity would be 50%.

FIG 3 (A to D) Temperature-dependent enzyme kinetic plots of ancestral ARODs (AncARODn0 [A], AncARODn1 [B], and AncARODn2 [C]) and native AROD (OkAROD) (D). The initial velocity at 10, 20, 30, and 40°C is represented as filled squares, circles, upward-facing triangles, and downward-facing triangles, respectively. (E) Arrhenius plots for k_{cat} values of ancestral and native ARODs. Activation energies were calculated from the plots; the parameters are given in Table 2.

have been achieved by introducing a consensus mutation that stabilizes the designed protein (35).

Enzyme kinetic analysis for ancestral and native ARODs. Enzyme kinetic analysis of ancestral (AncARODn0 [Fig. 3A], AncARODn1 [Fig. 3B], and AncARODn2 [Fig. 3C]) and native (OkAROD [Fig. 3D]) ARODs was performed utilizing L-Arg as a substrate at different temperatures (ranging from 10 to 40°C). The kinetic parameters (Table 2) indicate that the k_{cat} values toward L-Arg ($k_{cat,L-Arg}$) were similar to each other, whereas the K_m value toward L-Arg ($K_{m,L-Arg}$) differed between the ARODs. In fact, the $k_{cat,L-Arg}$

TABLE 2 Enzymatic properties of ancestral and native ARODs using L-Arg as a substrate at various temperatures^a

Temp (°C)	AncARODn0 (activation energy, 41.8 kJ/mol)			AncARODn1 (activation energy, 36.2 kJ/mol)			AncARODn2 (activation energy, 24.8 kJ/mol)			OkAROD (native) (activation energy, 34.3 kJ/mol)		
	k_{cat} (s ⁻¹)	K_m (μM)	k_{cat}/K_m (s ⁻¹ μM ⁻¹)	k_{cat} (s ⁻¹)	K_m (μM)	k_{cat}/K_m (s ⁻¹ μM ⁻¹)	k_{cat} (s ⁻¹)	K_m (μM)	k_{cat}/K_m (s ⁻¹ μM ⁻¹)	k_{cat} (s ⁻¹)	K_m (μM)	k_{cat}/K_m (s ⁻¹ μM ⁻¹)
10	2.1 ± 0.1	22.0 ± 2.9	0.09	2.87 ± 0.1	573 ± 69	0.005	3.4 ± 0.0	135 ± 2.1	0.03	3.2 ± 0.1	172 ± 7.0	0.02
20	4.3 ± 0.1	16.4 ± 1.6	0.3	5.7 ± 0.1	1,008 ± 76	0.005	4.7 ± 0.2	119 ± 12	0.04	7.2 ± 0.1	196 ± 6.2	0.04
30	6.9 ± 0.1	11.3 ± 0.8	0.6	9.1 ± 0.1	913 ± 51	0.01	7.5 ± 0.1	58.1 ± 2.5	0.13	8.6 ± 0.2	164 ± 8.4	0.05
40	9.0 ± 0.2	12.3 ± 1.2	0.7	12.6 ± 0.7	1548 ± 270	0.008	11.1 ± 0.2	113 ± 5.3	0.09	19.0 ± 0.8	715 ± 58	0.03

^aThe measurement of enzyme kinetics parameters was performed more than three times. Values are means ± standard deviations.

value ranged from 2.1 to 19.0 s⁻¹ at 10 to 40°C; however, the K_m ranged from 11.3 to 1,548 μM in the ARODs (Table 2). These differences would affect enzymatic efficiency, measured by k_{cat}/K_m . The values measured were in the following order at every temperature: AncARODn0 > AncARODn2 > OkAROD > AncARODn1 (Table 2). Activation energies were calculated from Arrhenius plots (Fig. 3E) as being in the order AncARODn0 > AncARODn1 > OkAROD > AncARODn2 (Table 2), indicating that, in ancestral ARODs, activation energies gradually increased as they became more similar to the ancestral sequence.

Next, kinetic analysis of AROD activity toward different L-amino acids (L-Lys, L-His, L-Phe, and L-Leu) was performed (Table 3). The K_m values toward L-Lys ($K_{m,Lys}$) were 1 to 2 orders of magnitude higher than the value of $K_{m,L-Arg}$ for all of the ARODs. Furthermore, the K_m values toward L-His, L-Phe, and L-Leu were more than 40-fold higher than the $K_{m,Lys}$ value for AncARODn0 (Table 3). On the other hand, the k_{cat} toward these L-amino acids was >0.7 s⁻¹ (Table 3), and several of them were not inferior to $k_{cat,L-Arg}$ compared with the difference in K_m .

Taken together, these results indicate that only AncARODn0, which is phylogenetically most remote from a native AROD, OkAROD, exhibits substrate promiscuity. Several mutations affect the substrate specificity of AncARODs; however, assigning the mutations from among hundreds of mutated residues is difficult. In fact, functionally important residues, such as active-site residues, cannot be assigned because neither crystal structures nor homology models are available for ARODs. Among the kinetic parameters measured for all the ARODs, K_m was clearly changed, whereas the change in k_{cat} was small. The activation energy for AncARODn0 was the highest among the ARODs.

DISCUSSION

In this study, we obtained three ancestral ARODs and one native AROD; the phylogenetic relationship of the ARODs is represented in Fig. 4 as stars with reference to the phylogenetic tree described by the ML method (see Fig. S4 in the supplemental material). The change in thermal stability ($T_{1/2}$) is represented as a thick line with a color gradient ranging from green (low temperature) to red (high temperature), indicating that stability was gradually improved by extending ARODs remote from the native AROD sequence (OkAROD) of the tree (Fig. 4 and S4). Gradual improvement in stability without a decay in enzyme activity has also been reported by another group (23). The results, especially for the expression of catalytically active AncARODs, indicate that ASR

TABLE 3 Comparison of enzyme kinetic parameters of ancestral and native ARODs toward L-Lys, L-His, L-Phe, and L-Leu^a

Substrate	AncARODn0			AncARODn1			AncARODn2			OkAROD (native)		
	k_{cat} (s ⁻¹)	K_m (mM)	k_{cat}/K_m (s ⁻¹ μM ⁻¹)	k_{cat} (s ⁻¹)	K_m (mM)	k_{cat}/K_m (s ⁻¹ μM ⁻¹)	k_{cat} (s ⁻¹)	K_m (mM)	k_{cat}/K_m (s ⁻¹ μM ⁻¹)	k_{cat} (s ⁻¹)	K_m (mM)	k_{cat}/K_m (s ⁻¹ μM ⁻¹)
L-Lys	5.1 ± 0.1	0.4 ± 0.0	0.012	16.7 ± 1.4	24.6 ± 3.7	7.0E-4	2.6 ± 0.1	11.5 ± 0.8	2.0E-4	3.5 ± 0.2	8.6 ± 0.1	4.0E-4
L-His	0.7 ± 0.0	16.5 ± 1.4	4.0E-5	ND ^b	ND	ND	ND	ND	ND	ND	ND	ND
L-Phe	1.2 ± 0.1	48.8 ± 3.7	2.0E-5	ND	ND	ND	ND	ND	ND	ND	ND	ND
L-Leu	ND	ND	9.0E-6	ND	ND	ND	ND	ND	ND	ND	ND	ND

^aMeasurement of enzyme kinetic parameters was performed more than three times at 30°C. Values are means ± standard deviations.

^bND, not determined.

FIG 4 Schematic model representing the biochemical parameters of ancestral and native ARODs on a phylogenetic tree. Thermal stability is colored gradually based on the difference in $T_{1/2}$; the value is indicated in the star. Relative activities toward seven of the 20 L-amino acids are represented as bar graphs.

of ARODs was successfully performed. The success or failure of ASR is strongly dependent on preparation of the sequence library (18). Moreover, our results suggest that library curation based on lowering the E_c is effective for ASR. This concept can broaden application of ASR from proteins for which structural and mutational data are available to less-studied ones, such as AROD, for which only sequence information is available (12). At the same time, there are concerns that the curation method excludes overlapped sequences judged by the script shown in Fig. S1 in the supplemental material. In fact, three sequences in the phylogenetic tree (GI no. 800980956, 923029671, and 640706981 in Fig. S4) were omitted from the design of ancestral ARODs. In a recent study, this concern, expressed as subsampling of taxa, did not greatly affect the design of ancestral sequence (36, 37). Several of the incorrectly inferred ancestral residues appeared not to affect ancestral phenotypes (36), and therefore, we believe that the phenotypes of AncARODs are encoded correctly even if the three sequences are omitted.

Bar graphs showing substrate specificity toward L-amino acids are presented in Fig. 4, indicating that AncARODn0 is the most remote from a native AROD, OkAROD, that exhibits substrate promiscuity (red bar in Fig. 4). Here, the reactivity of LAAO from *Rhizoctonia solani* (RsLAAO) (38) is similar to that of AncARODn0. The LAAOs can oxidize

seven of the L-amino acids shown in Fig. 4. In RsLAAO, k_{cat} is comparable in value for all L-amino acids, whereas k_{cat}/K_m was clearly changed, as in the case of AncARODn0. To represent this point, we calculated the relative enzyme efficiency for L-Arg and L-Lys, $E_{\text{LArg/LLys}}$, using the following equation:

$$E_{\text{LArg/LLys}} = \frac{(k_{\text{cat,LArg}}/K_{m,\text{LArg}})}{(k_{\text{cat,LLys}}/K_{m,\text{LLys}})}$$

Among the designed ARODs, AncARODn2 has the highest $E_{\text{LArg/LLys}}$: the value was >5-, 12-, and 40-fold higher than those of OkAROD, AncARODn0, and AncARODn1, respectively (Fig. 4). Here, comparing the k_{cat}/K_m values obtained with different substrates can be helpful, since the resurrected proteins may be either promiscuous or specific enzymes. Risso et al. reported that resurrected β -lactamase exhibits high k_{cat}/K_m values toward different substrates compared with modern β -lactamase (39). AncARODn0 has the highest k_{cat}/K_m values toward L-Arg, L-Lys, L-His, L-Phe, and L-Leu among the ARODs (Tables 2 and 3), suggesting that AncARODn0 is a promiscuous enzyme, like the resurrected β -lactamase, and, moreover, that ARODs may have evolved from promiscuous LAAOs. Large-scale comprehensive analysis of sequence and functional relationships revealed that promiscuous activities are broadly distributed among extant enzymes, and the functions of these enzymes appeared to connect to each other via enzyme promiscuity even if they shared low sequence identity (40). Thus, other highly specific LAAOs may be generated through a process identical to that for ARODs. Simultaneously, we inferred the existence of new LAAOs that bear substrate specificity toward other L-amino acids. We are now performing database searches to find these LAAOs.

From the viewpoint of applicability, we expected our designed ARODs to be useful for quantifying L-Arg concentrations in various samples, which requires high specificity toward L-Arg (12). To confirm this point, we quantified L-Arg concentrations in samples both with and without human plasma utilizing the four ARODs (AncARODn0, AncARODn1, AncARODn2, and OkAROD) (see Fig. S5A to D in the supplemental material). The results indicated that only AncARODn0 is not suitable to quantify the L-Arg concentration; in fact, the slope of a line with plasma (Fig. S5A, dashed line) was >1.6-fold larger than the slope of a line without plasma (Fig. S5A, straight line), whereas the slopes were almost identical to each other in other ARODs (Fig. S5B to D). Taking the findings together, we predicted that three ARODs (AncARODn1, AncARODn2, and OkAROD) have the potential to quantify L-Arg concentrations in various samples.

During the process of changing enzyme functions, the original function is often sacrificed to acquire new enzymatic functions, such as thermostability, solubility, and structural stability. However, Khersonsky and Tawfik suggested that inherently evolvable enzymes (in other words, those with mutational “robustness”) achieve changes with a minimal negative functional trade-off (41). The high thermostability of AncARODs may partly be due to the mutational robustness of AncARODs. At the same time, we agree that one key part of its evolvability is lacking because there is no evidence thus far that AncARODs have the ability to vary amino acid residues near their active site. We intend to determine through structural and mutational analysis whether AncARODs possess mutational robustness and are therefore good targets for designing enzymes for diagnostic applications.

ARODs are difficult targets with which to perform ASR because there is little information that is helpful to the design. Despite the difficulty, we can infer the evolutionary track of ARODs from a native AROD, OkAROD, and evaluate potential applications, not limited to generating thermostable enzymes but also generating highly specific enzymes, through the design and biochemical analysis of ancestral and native ARODs. We believe that our report can contribute to widening the application of ASR to other proteins in the future.

MATERIALS AND METHODS

Phylogenetic analysis and reconstruction of ancestral ARODs. The 10 selected sequences (red in Table S1 in the supplemental material) were aligned using MAFFT software (33). The aligned sequences were analyzed with MEGA6 (34), and a phylogenetic tree was generated by the maximum-likelihood method. Aligned sequences and tree data were submitted to FastML (32); JTT empirical models were adopted for the analysis. Finally, we generated three ancestral ARODs: AncARODn0, AncARODn1, and AncARODn2.

Next, we tried to represent the phylogenetic location of 10 sequences among the homologous sequences. The homologous sequences were processed by the script shown in Fig. S1 in the supplemental material with the cutoff value set to 90%. Sequences containing fewer than 540 aa or greater than 648 aa were excluded. Finally, we obtained a total of 139 sequences. The sequences that we obtained were aligned using MAFFT (33), and the results were analyzed with MEGA6 (34). A phylogenetic tree was generated by the ML method with the bootstrap value set to 500. The tree is shown in Fig. S4 in the supplemental material.

Overexpression and purification of one native and three ancestral AROD candidates. Genes encoding a native AROD candidate from *Oceanobacter kriegii* (OkAROD) (GI no. 654842541) and three ancestral ARODs (AncARODn0, AncARODn1, and AncARODn2) (see Table S2 in the supplemental material) were synthesized and cloned into the pET15b vector via the NdeI and BamHI sites by using GeneScript. The plasmids were introduced into the strain BL21(DE3), which was grown at 37°C in 1 liter of LB broth for 4 h. The temperature was decreased to 23°C, IPTG (isopropyl- β -D-thiogalactopyranoside) was added to a final concentration of 0.5 mM, and the cells were grown overnight. After harvesting, the cells were resuspended in buffer A and sonicated. The supernatant was collected by centrifugation at $11,000 \times g$ for 30 min and applied to an Ni²⁺-Sepharose 6 Fast Flow column (GE Healthcare, Uppsala, Sweden). After washing the column with buffer A containing 70 mM imidazole, the samples were eluted with buffer A containing 300 mM imidazole. The eluate containing AROD was concentrated and applied to a Superdex 200-pg increase column (GE Healthcare, Uppsala, Sweden) equilibrated with buffer A. AROD purity was determined by SDS-PAGE. The purified protein was then utilized in biochemical assays.

Determination of substrate specificity and kinetic parameters for two AROD candidates. AROD activity was measured by quantification of H₂O₂ produced by an enzymatic reaction using a color development method. The assay reagent was composed of 1.5 mM aminoantipyrine, 2 mM phenol, 50 U ml⁻¹ horseradish peroxidase, and 100 mM bis-Tris HCl, pH 7.5. To measure the specific activity of ARODs, the reaction was started by adding 10 mM substrate (20 L-amino acids) and purified AROD sample to the assay reagent. The initial velocity of each AROD was calculated by measuring the time-dependent absorbance change at 505 nm in the concentration of *N*-ethyl-*N*-(2-hydroxy-3-sulfopropyl)aniline (extinction coefficient at 505 nm, 12,700 M⁻¹ cm⁻¹ [42]) with a UV-visible spectrometer (DU800; Beckman). The relative activity of ARODs toward the 20 L-amino acids was calculated by setting the activity toward L-Arg to 100% (Table 1).

The kinetic parameters for L-Arg, L-Lys, L-His, L-Phe, and L-Leu were measured by using the following substrate concentrations: 0.01 to 5 mM L-Arg, 0.3 to 10 mM L-Lys, 1.0 to 50 mM L-His, 1.0 to 30 mM L-Phe, and 1.0 to 30 mM L-Leu. For L-Arg, the enzyme activity was measured by changing the temperature (from 10 to 40°C) to determine activation, whereas the temperature was set to 30°C to measure the kinetic parameters for the other substrates. The procedure to determine initial velocity was identical to that for the measurement of specific activity. The parameters were calculated with the Michaelis-Menten equation and by applying the nonlinear least-squares method.

Estimation of thermal stability and pH dependency of AROD candidates. The thermal stability of ARODs was determined as follows. The enzyme solution was incubated at 30 to 80°C for 10 min without substrate. The heat-treated samples were subsequently transferred to an ice bath. The remaining activity of ARODs was measured by applying a procedure identical to that for the measurement of specific activity. The pH dependency of ARODs was also determined under the same conditions, except that sodium acetate buffer (pH 3.5 to 5.5), bis-Tris-HCl (pH 6.0 to 7.0), Tris-HCl (pH 7.0 to 8.5), Bicine (pH 9.0 to 9.5), and glycine-KOH (pH 10.0 to 11.5) were used as buffers.

Quantification of L-Arg concentration utilizing ARODs. A total of 180 μ l of substrate solution containing 0 to 200 μ M L-Arg dissolved in 100 mM bis-Tris-HCl (7.0), 2 mM phenol, and 1.5 mM 4-aminoantipyrine with 50 U peroxidase was incubated for 3 min at 37°C. To prepare a solution containing human plasma, we added a total of 20 μ l human plasma to the solution. The reaction was started by adding 20 μ l of ARODs (20 mU of enzyme) to the solution, followed by incubation for 40 min at 37°C. After incubation, the absorbance at 505 nm was measured using a UV-visible spectrometer (DU800; Beckman). Changes in absorbance depending on L-Arg concentration were plotted with ORIGIN. The assay was performed in triplicate.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/AEM.00459-19>.

SUPPLEMENTAL FILE 1, PDF file, 1.9 MB.

ACKNOWLEDGMENTS

We are deeply appreciative of Daisuke Matsui's suggestion concerning the measurement of AROD enzymatic activity.

This work was supported by JSPS KAKENHI grant numbers 16K18688, 17K06931, 17H06169, and 18K14391 and JST/ERATO grant number JPMJER1102.

S.N. developed the method to curate the sequence library, designed artificial ARODs, designed the research study, and wrote the manuscript. M.N. performed the experiments. S.N., M.N., Y.A., and S.I. performed the data analysis.

REFERENCES

- Pollegioni L, Motta P, Molla G. 2013. L-Amino acid oxidase as biocatalyst: a dream too far? *Appl Microbiol Biotechnol* 97:9323–9341. <https://doi.org/10.1007/s00253-013-5230-1>.
- Ida K, Kurabayashi M, Suguro M, Hiruma Y, Hikima T, Yamamoto M, Suzuki H. 2008. Structural basis of proteolytic activation of L-phenylalanine oxidase from *Pseudomonas* sp. P-501. *J Biol Chem* 283:16584–16590. <https://doi.org/10.1074/jbc.M800366200>.
- Arima J, Sasaki C, Sakaguchi C, Mizuno H, Tamura T, Kashima A, Kusakabe H, Sugio S, Inagaki K. 2009. Structural characterization of L-glutamate oxidase from *Streptomyces* sp. X-119-6. *FEBS J* 276:3894–3903. <https://doi.org/10.1111/j.1742-4658.2009.07103.x>.
- Utsumi T, Arima J, Sakaguchi C, Tamura T, Sasaki C, Kusakabe H, Sugio S, Inagaki K. 2012. Arg305 of *Streptomyces* L-glutamate oxidase plays a crucial role for substrate recognition. *Biochem Biophys Res Commun* 417:951–955. <https://doi.org/10.1016/j.bbrc.2011.12.033>.
- Okazaki S, Nakano S, Matsui D, Akaji S, Inagaki K, 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. <https://doi.org/10.1093/jb/mvt070>.
- Matsui D, Asano Y. 2015. Heterologous production of L-lysine epsilon-oxidase by directed evolution using a fusion reporter method. *Biosci Biotechnol Biochem* 79:1473–1480. <https://doi.org/10.1080/09168451.2015.1034654>.
- Sehanobish E, Williamson HR, Davidson VL. 2016. Roles of conserved residues of the glycine oxidase GoxA in controlling activity, cooperativity, subunit composition, and cysteine tryptophylquinone biosynthesis. *J Biol Chem* 291:23199–23207. <https://doi.org/10.1074/jbc.M116.741835>.
- Andreo-Vidal A, Mamounis KJ, Sehanobish E, Avalos D, Campillo-Brocal JC, Sanchez-Amat A, Yukl ET, Davidson VL. 2018. Structure and enzymatic properties of an unusual cysteine tryptophylquinone-dependent glycine oxidase from *Pseudoalteromonas luteoviolacea*. *Biochemistry* 57:1155–1165. <https://doi.org/10.1021/acs.biochem.8b00009>.
- Gomez D, Espinosa E, Bertazzo M, Lucas-Elio P, Solano F, 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. <https://doi.org/10.1007/s00253-008-1499-x>.
- Nuutinen JT, Marttinen E, Soliymani R, Hilden K, Timonen S. 2012. L-Amino acid oxidase of the fungus *Hebeloma cylindrosporum* displays substrate preference towards glutamate. *Microbiology* 158:272–283. <https://doi.org/10.1099/mic.0.054486-0>.
- Moustafa IM, Foster S, Lyubimov AY, 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. <https://doi.org/10.1016/j.jmb.2006.09.032>.
- Matsui D, Terai A, 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. <https://doi.org/10.1016/j.enzmictec.2015.10.002>.
- Faust A, Niefind K, Hummel W, 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. <https://doi.org/10.1016/j.jmb.2006.11.071>.
- Saam J, Rosini E, Molla G, Schulten K, Pollegioni L, Ghisla S. 2010. O₂ reactivity of flavoproteins: dynamic access of dioxygen to the active site and role of a H⁺ relay system in D-amino acid oxidase. *J Biol Chem* 285:24439–24446. <https://doi.org/10.1074/jbc.M110.131193>.
- Khersonsky O, Roodveldt C, Tawfik DS. 2006. Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr Opin Chem Biol* 10:498–508. <https://doi.org/10.1016/j.cbpa.2006.08.011>.
- Jensen RA. 1976. Enzyme recruitment in evolution of new function. *Annu Rev Microbiol* 30:409–425. <https://doi.org/10.1146/annurev.mi.30.100176.002205>.
- Gumulya Y, Gillam EM. 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. <https://doi.org/10.1042/BCJ20160507>.
- Merkel R, Sterner R. 2016. Ancestral protein reconstruction: techniques and applications. *Biol Chem* 397:1–21. <https://doi.org/10.1515/hsz-2015-0158>.
- Gaucher EA, Govindarajan S, Ganesh OK. 2008. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* 451:704–707. <https://doi.org/10.1038/nature06510>.
- Akanuma S, Nakajima Y, Yokobori S, Kimura M, Nemoto N, Mase T, Miyazono K, Tanokura M, Yamagishi A. 2013. Experimental evidence for the thermophilicity of ancestral life. *Proc Natl Acad Sci U S A* 110:11067–11072. <https://doi.org/10.1073/pnas.1308215110>.
- Busch F, Rajendran C, Heyn K, Schlee S, Merkel R, Sterner R. 2016. Ancestral tryptophan synthase reveals functional sophistication of primordial enzyme complexes. *Cell Chem Biol* 23:709–715. <https://doi.org/10.1016/j.chembiol.2016.05.009>.
- Clifton BE, Jackson CJ. 2016. Ancestral protein reconstruction yields insights into adaptive evolution of binding specificity in solute-binding proteins. *Cell Chem Biol* 23:236–245. <https://doi.org/10.1016/j.chembiol.2015.12.010>.
- Nguyen V, Wilson C, Hoemberger M, Stiller JB, Agafonov RV, Kutter S, English J, Theobald DL, Kern D. 2017. Evolutionary drivers of thermoadaptation in enzyme catalysis. *Science* 355:289–294. <https://doi.org/10.1126/science.aah3717>.
- Clifton BE, Kaczmarek JA, Carr PD, Gerth ML, Tokuriki N, Jackson CJ. 2018. Evolution of cyclohexadienyl dehydratase from an ancestral solute-binding protein. *Nat Chem Biol* 14:542–547. <https://doi.org/10.1038/s41589-018-0043-2>.
- Kaltenbach M, Burke JR, Dindo M, Pabis A, Munsberg FS, Rabin A, Kamerlin SCL, Noel JP, Tawfik DS. 2018. Evolution of chalcone isomerase from a noncatalytic ancestor. *Nat Chem Biol* 14:548–555. <https://doi.org/10.1038/s41589-018-0042-3>.
- Devamani T, Rauwerdink AM, Lunzer M, Jones BJ, Mooney JL, Tan MA, Zhang ZJ, Xu JH, Dean AM, Kazlauskas RJ. 2016. Catalytic promiscuity of ancestral esterases and hydroxynitrile lyases. *J Am Chem Soc* 138:1046–1056. <https://doi.org/10.1021/jacs.5b12209>.
- Nakano S, Motoyama T, Miyashita Y, Ishizuka Y, Matsuo N, Tokiwa H, Shinoda S, Asano Y, 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. <https://doi.org/10.1021/acs.biochem.8b00339>.
- Jäckel C, Bloom JD, Kast P, Arnold FH, Hilvert D. 2010. Consensus protein design without phylogenetic bias. *J Mol Biol* 399:541–546. <https://doi.org/10.1016/j.jmb.2010.04.039>.
- Porebski BT, Buckle AM. 2016. Consensus protein design. *Protein Eng Des Sel* <https://doi.org/10.1093/protein/gzw015>.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. <https://doi.org/10.1093/nar/25.17.3389>.
- Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39:W29–W37. <https://doi.org/10.1093/nar/gkr367>.
- Pupko T, Pe'er I, Shamir R, Graur D. 2000. A fast algorithm for joint reconstruction of ancestral amino acid sequences. *Mol Biol Evol* 17:890–896. <https://doi.org/10.1093/oxfordjournals.molbev.a026369>.
- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059–3066. <https://doi.org/10.1093/nar/gkf436>.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6:

- Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30:2725–2729. <https://doi.org/10.1093/molbev/mst197>.
35. Trudeau DL, Kaltenbach M, Tawfik DS. 2016. On the potential origins of the high stability of reconstructed ancestral proteins. *Mol Biol Evol* 33:2633–2641. <https://doi.org/10.1093/molbev/msw138>.
36. Randall RN, Radford CE, Roof KA, Natarajan DK, Gaucher EA. 2016. An experimental phylogeny to benchmark ancestral sequence reconstruction. *Nat Commun* 7:12847. <https://doi.org/10.1038/ncomms12847>.
37. Rosenberg MS, Kumar S. 2001. Incomplete taxon sampling is not a problem for phylogenetic inference. *Proc Natl Acad Sci U S A* 98:10751–10756. <https://doi.org/10.1073/pnas.191248498>.
38. Hahn K, Neumeister K, Mix A, Kottke T, Groger H, 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. <https://doi.org/10.1007/s00253-016-8054-y>.
39. Risso VA, Gavira JA, Mejia-Carmona DF, Gaucher EA, Sanchez-Ruiz JM. 2013. Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian beta-lactamases. *J Am Chem Soc* 135:2899–2902. <https://doi.org/10.1021/ja311630a>.
40. Baier F, Copp JN, Tokuriki N. 2016. Evolution of enzyme superfamilies: comprehensive exploration of sequence-function relationships. *Biochemistry* 55:6375–6388. <https://doi.org/10.1021/acs.biochem.6b00723>.
41. Khersonsky O, Tawfik DS. 2010. Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu Rev Biochem* 79:471–505. <https://doi.org/10.1146/annurev-biochem-030409-143718>.
42. Tamaoku K, Ueno K, Akiura K, Ohkura Y. 1982. New water-soluble hydrogen donors for the enzymatic photometric determination of hydrogen peroxide. II. N-ethyl-N-(2-hydroxy-3-sulfopropyl)aniline derivatives. *Chem Pharm Bull* 30:2492–2497. <https://doi.org/10.1248/cpb.30.2492>.