

# Rational Design of the Soluble Variant of L-Pipecolic Acid Hydroxylase using the $\alpha$ -Helix Rule and the Hydropathy Contradiction Rule

Suguru Shinoda, Aoi Itakura, Haruka Sasano, Ryoma Miyake, Hiroshi Kawabata, and Yasuhisa Asano\*

Cite This: *ACS Omega* 2022, 7, 29508–29516

Read Online

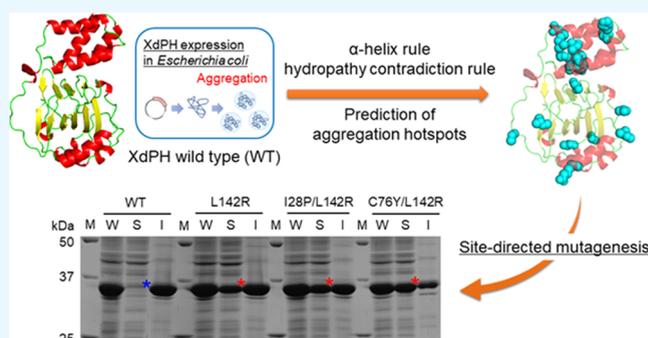
ACCESS |

Metrics &amp; More

Article Recommendations

Supporting Information

**ABSTRACT:** The production of recombinant proteins in *Escherichia coli* is an important application of biotechnology. 2-Oxoglutarate-dependent L-pipecolic acid hydroxylase derived from *Xenorhabdus doucetiae* (XdPH) is an excellent biocatalyst that catalyzes the hydroxylation of L-pipecolic acid to produce *cis*-5-hydroxy-L-pipecolic acid. However, the enzyme tends to form aggregates in the *E. coli* expression system. Our group established two rules, namely, the “ $\alpha$ -helix rule” and the “hydropathy contradiction rule,” to select residues to be altered for improving the heterologous recombinant production of proteins, by analyzing their primary structure. We rationally designed XdPH variants that are expressed in highly soluble and active forms in the *E. coli* expression system using these hotspot prediction methods, and the L142R variant showed a remarkably high soluble expression level compared to the wild-type XdPH. Further mutations were introduced into the L142R gene by site-directed mutagenesis. Moreover, the I28P/L142R and C76Y/L142R double variants displayed improved soluble expression levels compared to the single variants. These variants were also more thermostable than the wild-type XdPH. To analyze the effect of the alteration on one of the hotspots, L142 was replaced with various hydrophilic and positively charged residues. The remarkable increase in soluble protein expression caused by the alterations suggests that the decrease in the hydrophobicity of the protein surface and the enhancement of the interaction between nearby residues are important factors determining the solubility of the protein. Overall, this study demonstrated the effectiveness of our protocol in identifying aggregation hotspots for recombinant protein production and in basic biochemical research.



## INTRODUCTION

Heterologous expression systems using *Escherichia coli* are widely used in basic and applied research in biochemistry. *E. coli* is advantageous as it is inexpensive, fast, and gives high yields of the recombinant protein.<sup>1</sup> However, many limitations, such as the formation of aggregates called “inclusion bodies” caused by incorrect protein folding, are also associated with their use.<sup>2</sup> Several strategies have been investigated to overcome these limitations. One of the most widely used methods is the optimization of the culture conditions of the transformant by lowering the cultivation temperature and changing the medium composition.<sup>1,3,4</sup> Other strategies include codon optimization, coexpression with chaperones, and the use of promoters with different strengths to control the rate of protein synthesis.<sup>5,6</sup> However, these strategies are time-consuming and do not always yield positive results.

We established two rules, namely, the  $\alpha$ -helix rule and the hydropathy contradiction rule to identify residues (called aggregation hotspots) to improve solubility of the recombinant proteins by analyzing their primary structure. Replacing the amino acid residues at the hotspot with appropriate residues

via directed evolution leads to more efficient protein folding, resulting in higher expression levels of the genes.<sup>7</sup> To date, we have been successful in improving the production levels of the soluble forms of various proteins in *E. coli* by alteration of their residues using the  $\alpha$ -helix and hydropathy contradiction rules.

In this study, we targeted 2-oxoglutarate-dependent L-pipecolic acid hydroxylase (EC1.14.11.4) derived from *Xenorhabdus doucetiae* (XdPH) for use as a biocatalyst for the hydroxylation of L-pipecolic acid. Hydroxypipecolic acids (HyPips) are naturally occurring six-membered heterocyclic hydroxy amino acids that are components of some peptide antibiotics, terpenoids, and alkaloids. HyPips have been used as chiral building blocks in the synthesis of pharmaceuticals. For example, *cis*-4-hydroxy-L-pipecolic acid is a component of

Received: July 6, 2022

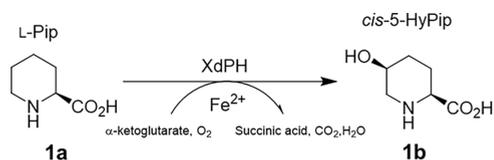
Accepted: July 28, 2022

Published: August 11, 2022



palinavir,<sup>8</sup> an HIV protease inhibitor, and *cis*-5-hydroxy-*L*-pipecolic acid (*cis*-5-HyPip) is a precursor for the synthesis of the  $\beta$ -lactamase inhibitor, MK-7655.<sup>9</sup> In previous studies, proline hydroxylases belonging to the Fe(II)/2-oxoglutarate-dependent dioxygenase superfamily were shown to catalyze the hydroxylation of *L*-pipecolic acid (*L*-Pip).<sup>10</sup> *L*-Proline *trans*-4-hydroxylase of the *Dactylosporangium* sp. strain RH1 converts *L*-Pip into *trans*-5-HyPip,<sup>11</sup> while *L*-proline *cis*-3-hydroxylase of the *Streptomyces* sp. strain TH1 converts *L*-Pip into *cis*-3-HyPip.<sup>12</sup> XdPH has been reported to be useful for catalyzing the production of *cis*-5-HyPip from *L*-Pip (Scheme 1).<sup>13</sup>

### Scheme 1. Enzymatic Synthesis of *cis*-5-HyPip by XdPH

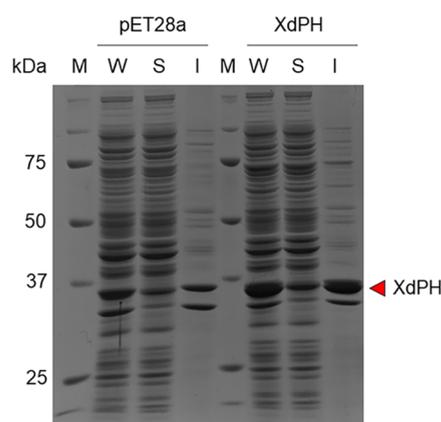


Quantum mechanical/molecular mechanical (QM/MM) studies on the catalytic mechanism of Fe(II)/2-oxoglutarate-dependent enzymes have been reported.<sup>14,15</sup> Various enzymes have been produced at high levels for industrial applications, using recombinant gene technology with *E. coli* as a host.<sup>16,17</sup> However, the insolubility of the enzymes in *E. coli* has been a limiting factor in the functional analysis, protein structure analysis, and the XdPH-catalyzed synthesis of *cis*-5-HyPip for industrial applications.

In this study, we introduced rational alterations to XdPH using hotspot prediction methods developed in our laboratory, to improve the soluble expression of the enzyme in *E. coli*. The soluble expression of XdPH was achieved by introducing single or double mutations. These variants showed not only improved soluble expression but also higher thermostability. Biochemical analysis indicated that the L142R variant was highly soluble in the *E. coli* expression system compared to the wild-type protein. Saturation mutagenesis and bioinformatic analysis of the L142R variant suggested that the decrease in hydrophobicity of the protein surface and the enhancement of the interaction with nearby residues are important factors for the soluble expression of XdPH. This study demonstrated that the two aforementioned rules can be used to improve soluble expression of recombinant proteins in the field of biochemistry.

## RESULTS AND DISCUSSION

**Heterologous Expression and 2-Oxoglutarate-Dependent *L*-Pipecolic Acid Hydroxylase Activity of Wild-Type XdPH.** In this study, the pET-28a(+) vector and *E. coli* BL21 (DE3) were used as expression systems for the *XdPH* gene, producing an *N*-terminal hexahistidine tag-fused recombinant XdPH. After gene expression using an auto-induction medium, the expression of the proteins induced by wild-type XdPH was confirmed using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and the enzyme activity assay described in the Experimental Section. The hydroxylation of *L*-Pip by the crude extract of wild-type XdPH was assayed by determining the conversion of *L*-Pip to *cis*-5-HyPip. The product was labeled with *N* $\alpha$ -(5-fluoro-2,4-dinitrophenyl)-*L*-leucinamide (*L*-FDLA) and analyzed by ultra-performance liquid chromatography (UPLC) (Figure S2). SDS-PAGE revealed the presence of recombinant XdPH in the insoluble fraction (Figure 1). The activity of XdPH was



**Figure 1.** SDS-PAGE analysis of the heterologous expression of XdPH using *E. coli* as a host. Each fraction was prepared as a transformant after induction. The method for preparing each fraction is described in Figure S1. W, whole-cell lysate; S, soluble fraction; I, insoluble fraction. pET28a: pET-28a(+) was used as the vector in the *E. coli* expression system; XdPH: *N*-terminus hexahistidine-tagged XdPH; red triangle: molecular mass of *N*-terminus hexahistidine-tagged XdPH.

determined by UPLC analysis, and 1.4 mU/mg soluble protein was obtained (Figure S3). Based on these results, recombinant XdPH was considered to form inclusion bodies after the heterologous expression of the gene.

**Prediction of the Aggregation Hotspots in XdPH by the  $\alpha$ -Helix rule and the Hydropathy Contradiction Rule.** Aggregation hotspots were predicted by the  $\alpha$ -helix rule and the hydropathy contradiction rule based on the amino acid sequence of XdPH. Hotspot residues were identified by the hydropathy contradiction rule based on the consensus design method, which improves protein function by replacing certain residues of the target protein with residues that are highly conserved in proteins of the same family. In this study, highly conserved residues were selected based on appearance rates, which were calculated for each residue through multiple sequence alignment with homologous proteins in a database. The HiSol scores were calculated for each residue of the target protein. The score was negative if the residue of the target protein was hydrophilic and the residue of the consensus protein was hydrophobic, and the score was positive if it was the other way around. Thus, the score identified residues with contradictory hydrophobicity by comparing them with the consensus amino acid residues. Using this method, appropriate amino acid residues can be altered for increasing the soluble expression of the recombinant protein. The hydropathy contradiction rule predicts aggregation hotspots according to the following three criteria:

- (1) High absolute value of HiSol score.
- (2) Highly conserved residues different from those of the target protein.
- (3) Alteration of the hydropathy index from a negative value to a positive value, or vice versa.

The appearance rate and HiSol score of each amino acid residue were calculated using the INTMSAlign\_HiSol program, based on the amino acid sequences of XdPH and 112 other proteins obtained by a BLAST sequence similarity search (Supporting Data S1). However, hotspot residues can be identified by the  $\alpha$ -helix rule by converting a hydrophobic amino acid present in the hydrophilic region of the  $\alpha$ -helix into

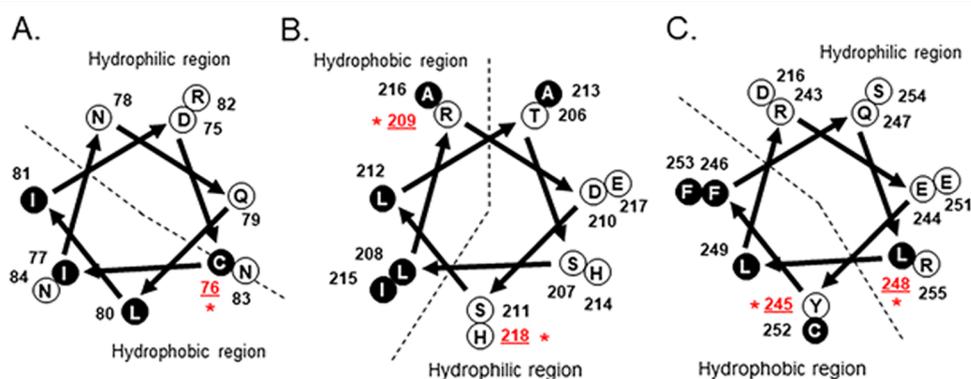
the protein into a hydrophilic amino acid, or by converting a hydrophilic amino acid existing in the hydrophobic region into a hydrophobic amino acid (Figure S4). Based on the aggregation hotspot prediction, 19 candidate residues were identified among a total of 294 residues (Table 1). Out of

**Table 1. Predicted Aggregation Hotspots, HiSol Scores, and Conserved Residues Corresponding to Residues in XdPH<sup>a</sup>**

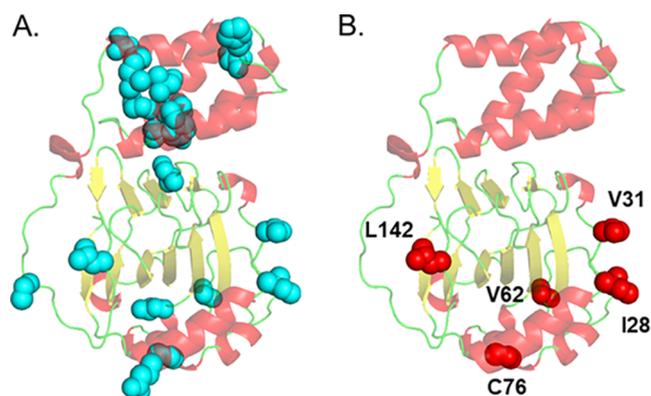
target residue	HiSol score	conserved residue	appearance rate (%)	position	prediction method
Glu15	-1.103	Ala	9 → 18.4	helix	HiSol
Ile28	2.042	Pro	1.4 → 13.6	coil	HiSol
Val31	2.035	Glu	3.6 → 34.5	coil	HiSol
Val62	2.24	Asp	0 → 29.1	coil	HiSol
Cys76	1.06	Tyr	0 → 54.8	helix	HiSol + $\alpha$
Leu89	2.021	Arg	5.5 → 25.6	helix	HiSol
Ile108	2.487	Arg	0.2 → 64.2	coil	HiSol
Leu142	2.228	Arg	1.2 → 53.1	coil	HiSol
Cys189	1.132	Asp	0 → 11.5	coil	HiSol
Leu208	1.63	Glu	12.7 → 26.6	helix	HiSol + $\alpha$
Arg209	-1.355	Leu	24.8 → 18.4	helix	HiSol + $\alpha$
His218	-1.784	Val	0.5 → 24	helix	HiSol
Arg243	-1.205	Gly	3.8 → 21.2	helix	HiSol
Glu244	-1.077	Ala	14.8 → 19.5	helix	HiSol
Tyr245	-1.181	Val	0 → 26.1	helix	HiSol + $\alpha$
Phe246	1.051	Tyr	18.9 → 49.9	helix	HiSol
Leu248	1.244	Trp	18.8 → 45.9	helix	HiSol + $\alpha$
Arg262	-2.271	Val	3.6 → 28.7	helix	HiSol
Ile271	1.682	Tyr	0 → 37.5	helix	HiSol

<sup>a</sup> $\alpha$ :  $\alpha$ -helix rule; HiSol: hydropathy contradiction rule.

these, 14 candidate residues were identified by the hydropathy contradiction rule and five were identified by both the  $\alpha$ -helix rule and the hydropathy contradiction rule (Figure 2). Twelve candidate residues on the  $\alpha$ -helix and six candidate residues on the coil structures were displayed in the homology model of XdPH. The candidate residues were located on the protein surface (Figure 3). On the basis of these results, we prepared 19 variants to improve the solubility of XdPH for heterologous expression.



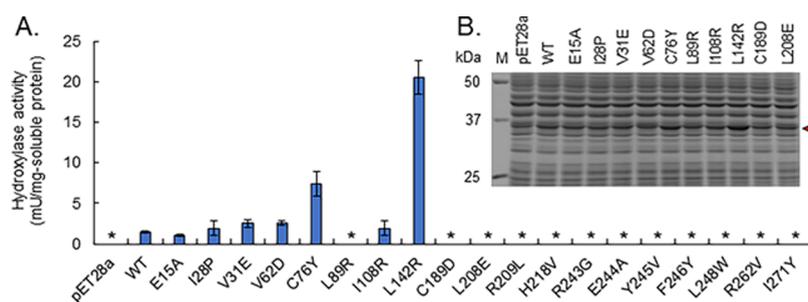
**Figure 2.** Helical wheel depiction of the  $\alpha$ -helix regions in XdPH that contributed to the improvement in soluble expression. Helical wheels are depicted for the following  $\alpha$ -helix regions of XdPH: residues 75–84 (DCINQLIRNN) (A), residues 206–218 (TSLRDSLAIHAEH) (B), and residues 243–255 (REYFQLLDECFSR) (C). Hydrophobic and hydrophilic residues are shown in black and white letters, respectively. The numbers represent the amino acid residues. The alterations in the residues are indicated as asterisks (\*), whose polar properties are opposite to those of the hydrophobic or hydrophilic region. The hydrophobic residues, C76 (A) and L248 (C), in XdPH were located in the hydrophilic regions of the  $\alpha$ -helix. However, the hydrophilic residues, R209 (B) and Y245 (C), were located in the hydrophobic regions of the same secondary structure. Residues to be modified are presented according to the hydropathy contradiction rule.



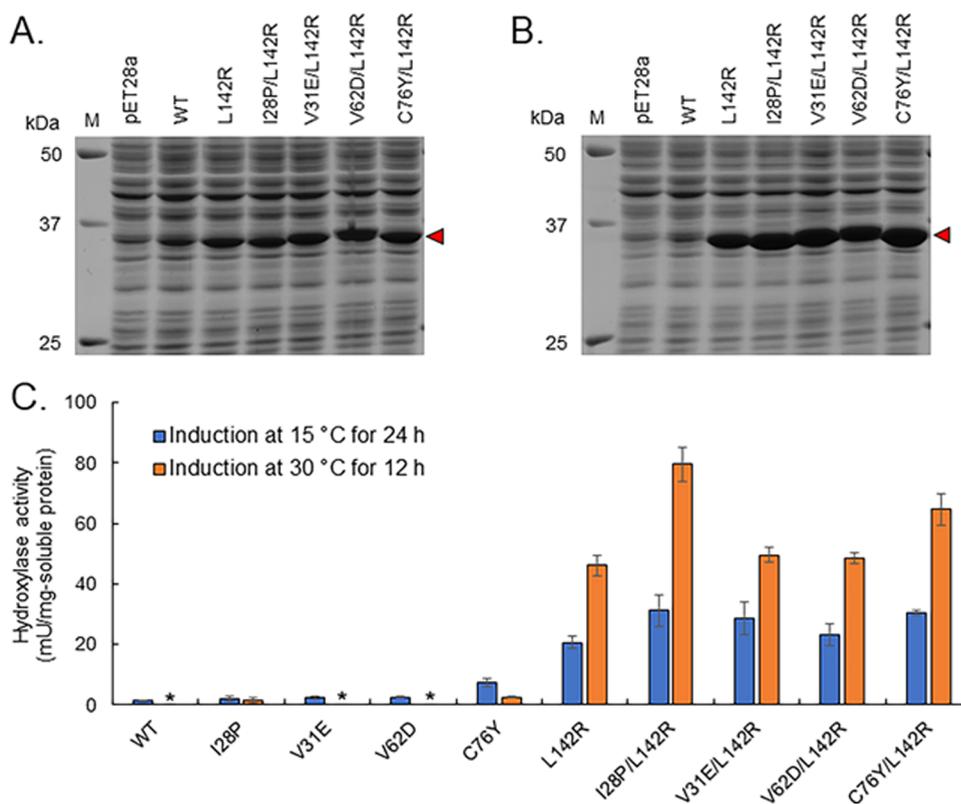
**Figure 3.** Location of the aggregation hotspots in the homology model of XdPH. Predicted hotspots in the homology models of XdPH are indicated as cyan spheres, which are located on the surface of the protein (A). I28, V31, V62, C76, and L142 are indicated as red spheres (B).

**First Screening of the Soluble Variants of XdPH with Single Amino Acid Alterations.** XdPH variants were generated using site-directed mutagenesis. The genes encoding wild-type XdPH and its variants were expressed in *E. coli* BL21 (DE3). Enzyme activity in the soluble fraction of the proteins was subsequently measured to compare the soluble expression of wild-type XdPH with that of its variants (Figure 4A). A crude enzyme extract was used in the activity assay. The soluble fractions of the wild-type, V31E, V62D, C76Y, and L142R variants showed activities of 1.4, 2.53, 2.51, 7.36, and 20.5 mU/mg soluble protein, respectively. Increased enzyme activity was observed in V31E, V62D, C76Y, and L142R variants. In particular, the activity of the L142R variant was more than 14.6-fold higher per unit of soluble protein relative to that of the wild-type XdPH. The soluble fractions of the proteins were analyzed by SDS-PAGE to compare the solubility of the wild-type XdPH with that of its variants (Figure 4B). Higher levels of C76Y and L142R variants were observed compared to those of the wild-type XdPH (Figure 4B).

**Second Screening of the Soluble XdPH Variants with Double Alterations in Amino Acids.** To further improve



**Figure 4.** Screening of the soluble XdPH variants with hydroxylase activity assay and SDS-PAGE analysis. The hydroxylase activity assay using *L*-Pip as a substrate (A). One unit (U) of hydroxylase activity is defined as the amount of enzyme required to convert 1  $\mu$ mol of *L*-Pip to *cis*-5-HyPip per minute. An asterisk (\*) indicates that the activity is less than 0.5 mU/mg soluble protein. SDS-PAGE analysis of the soluble fraction of the XdPH variants expressed using the pET expression system at 15 °C for 24 h (B). Red triangle: molecular mass of the target protein. This experiment was independently performed three times ( $n = 3$ ) with respect to the expression of each gene and the measurement of the enzyme activity.

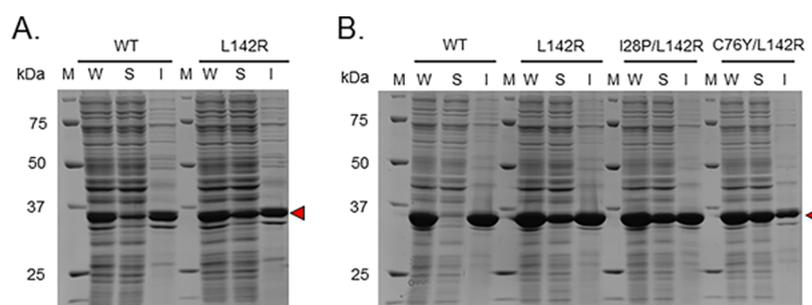


**Figure 5.** Screening of the XdPH double variants using SDS-PAGE analysis and hydroxylase activity assay. SDS-PAGE of the soluble fraction of the wild-type, L142R variant, and double variants expressed using the pET expression system at 15 °C for 24 h (A) and 30 °C for 12 h (B). Red triangle: molecular mass of the target protein. The hydroxylase activity assay performed using *L*-Pip as the substrate (C). One unit (U) of hydroxylase activity is defined as the amount of enzyme required to convert 1  $\mu$ mol of *L*-Pip to *cis*-5-HyPip per minute. Orange bar: expression at 15 °C for 24 h. Blue bar: expression at 30 °C for 12 h. Asterisk (\*) indicates enzyme activity <0.5 mU/mg soluble protein. This experiment was independently performed three times ( $n = 3$ ) from the expression of each gene to the measurement of enzyme activity.

the soluble expression of the protein, four double variants (I28P/L142R, V31E/L142R, V62D/L142R, and C76Y/L142R), each containing a unique second alteration in addition to L142R, were constructed using site-directed mutagenesis. The gene encoding the variants was expressed by induction at 15 °C for 24 h. Amounts of the proteins in the soluble fraction were analyzed by SDS-PAGE, and enzyme activities were measured to compare the solubilities of the L142R variant and variants containing the additional alterations at different sites. Enzyme activity was measured in the crude enzyme extracts. SDS-PAGE analysis showed that the soluble expression levels

of the L142R variant and double variants at 15 °C were identical (Figure 5A). Variants I28P/L142R, V31E/L142R, V62D/L142R, and C76Y/L142R had activities of 30.1, 28.7, 23.3, and 30.3 mU/mg soluble protein, respectively (Figure 5C).

Because the double variants showed improved soluble expression, they were further studied at an elevated temperature of 30 °C for 12 h. SDS-PAGE analysis revealed that the soluble expression of the L142R and double variants increased at 30 °C compared with that at 15 °C (Figure 5A,B). Variants I28P, C76Y, L142R, I28P/L142R, V31E/L142R, V62D/



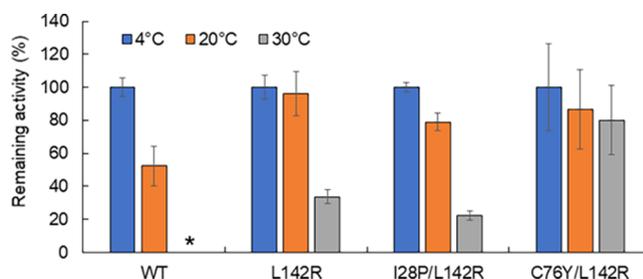
**Figure 6.** Solubility of the wild-type XdPH and its soluble variants expressed at 15 °C for 24 h (A) and 30 °C for 12 h (B). The whole-cell lysate (W), soluble fraction (S), and insoluble fraction (I) were analyzed by SDS-PAGE. The method for preparing each fraction is described in the Supporting Information. Red triangle: molecular mass of the target protein.

L142R, and C76Y/L142R had activities of 1.39, 2.45, 46.2, 79.6, 49.6, 48.5, and 64.8 mU/mg soluble protein, respectively (Figure 5C). The enzyme activity of the L142R variant was 2-fold higher at 30 °C than that at 15 °C. Furthermore, the enzymatic activities of the I28P and C76Y variants were detected at 30 °C. The I28P/L142R and C76Y/L142R double variants showed a 1.7-fold and 1.4-fold increase in activity, respectively, compared to the L142R variant. These results indicate a significant improvement in the thermostability of the L142R and double variants, as the enzyme activity of the wild-type XdPH was not detected at 30 °C. After induction at 30 °C for 12 h, the whole-cell lysate, soluble fraction, and insoluble fraction were analyzed by SDS-PAGE to compare the solubilities of the wild-type XdPH, L142R, I28P/L142R, and C76Y/L142R variants (Figure 6B). The results indicated that more soluble variants could be obtained using any combination of more than one mutation. These results suggest that the soluble expression level was improved by the cumulative effects of the alterations, in addition to the effect of L142R alteration alone.

**Temperature–Stability Relationship of XdPH and Its Soluble Variants.** As the temperature–stability relationship of the L142R variant and double variants was markedly improved, that of the wild-type and the double variants with improved soluble expression was investigated. Enzyme activity was measured under the same conditions as those of the standard enzyme assay after incubation of the enzyme at 4, 20, and 30 °C for 60 min. The enzyme activity of the wild-type XdPH was between 4 °C and 20 °C; however, a significant decrease in activity was observed at temperatures above 30 °C (Figure 7). L142R and the double variants showed activity at temperatures between 4 and 30 °C. When the enzyme activity at 4 °C was set to 100%, the relative activity of the C76Y/L142R variant at 30 °C was 80%, whereas that of the other variants was 20%. Hence, the C76Y/L142R variant exhibited improved stability at high temperatures. These results indicate that the C76Y alteration may improve thermostability through the substitution of free cysteine on the protein surface.<sup>18</sup>

#### Saturation Mutagenesis and Bioinformatic Analysis of the Positive Variants of the Residue at Position 142.

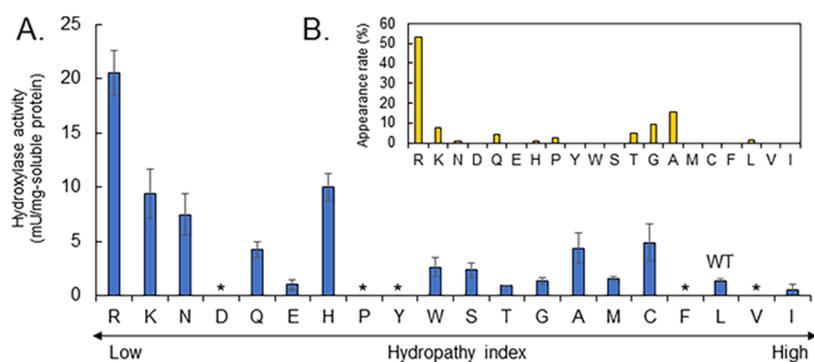
To investigate the effect of alterations of residues on soluble expression, residue L142 of XdPH was replaced by 19 different residues using site-directed mutagenesis. All genes encoding the variants were expressed in *E. coli* BL21(DE3). The enzymatic activity of the soluble fractions was measured. Nine variants, namely, L142R, L142K, L142N, L142Q, L142H, L142W, L142S, L142A, and L142C, were expressed in a soluble form, whereas the other variants did not show any



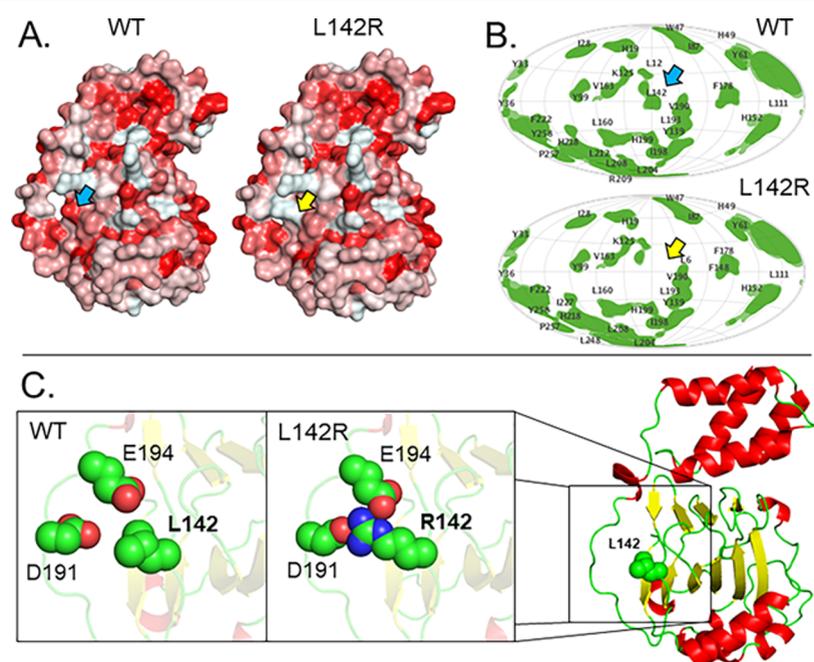
**Figure 7.** Temperature stabilities of the wild-type XdPH and its soluble variants. The enzyme activities were measured with crude enzyme extracts. The whole-cell lysate was desalted using a PD SpinTrap G-25 (Cytiva, MA) and equilibrated with 100 mM MES buffer (pH 6.5). The desalted enzyme solutions (2 mg/mL) of the wild-type, L142R, I28P/L142R, V31E/L142R, and C76Y/L142R variants were incubated for 1 h at 4, 20, and 30 °C in 100 mM MES (pH 6.5). Thereafter, the enzyme activity was measured under the same conditions as those in the enzyme assay method. The variants were further diluted 10-fold to equalize their enzyme activities with that of wild-type XdPH. One unit of hydroxylase activity is defined as the amount of enzyme required to convert 1  $\mu$ mol of L-Pip to *cis*-5-HyPip per minute. The relative activity of the enzyme at 4 °C was set to 100%. The wild-type XdPH (WT), L142R, I28P/L142R, and C76Y/L142R variants had activities of 1.34, 2.2, 3.7, and 2.88 mU/mL reaction mixture, respectively. The asterisk (\*) indicates enzyme activity <0.5 mU/mL reaction mixture. The measurement of enzyme activity was independently performed three times ( $n = 3$ ).

increase in solubility compared to wild-type XdPH (Figure 8A). In particular, substitution with negatively charged amino acids decreases enzyme activity. The amino acid sequence of XdPH was compared with those of proteins with high sequence homology (Figure S5). The protein with the highest similarity had 38% identity. Multiple sequence alignment revealed that the most frequently appearing amino acid residues corresponding to L142 were lysine, glutamic acid, glutamine, and arginine, all of which are hydrophilic amino acids with low hydropathy indices. Using the INTMSAlign program, the most frequently appearing amino acid residue corresponding to L142 in the database was found to be arginine (Figure 8B). These results suggest that hydrophobicity and charge of the amino acid residues replacing L142 are important factors for the soluble expression of XdPH.

The crystal structure of L-proline *cis*-4-hydroxylase from the *Mesorhizobium japonicum* strain LMG 29417 (PDB: 4P7W)<sup>19</sup> was employed to model the wild-type XdPH and the L142R variants, whose soluble expression had improved. The sequence identity between XdPH and L-proline *cis*-4-hydroxylase from the *M. japonicum* strain LMG 29417 is 30.3%. The



**Figure 8.** Enzyme activities displayed by the variants generated by saturation mutagenesis of L142 (A) and calculation of the appearance rate of L142 using the INTMSAlign program (B). The hydroxylase activity was assayed using L-Pip as a substrate. The calculation was performed using the INTMSAlign program according to a previous study.<sup>26</sup> The variants are arranged in an ascending order based on the hydrophobicity index from the left.<sup>27</sup> An asterisk (\*) indicates enzyme activity <0.5 mU/mg soluble protein. This experiment was independently performed three times ( $n = 3$ ) from the expression of each gene to the measurement of enzyme activity.



**Figure 9.** Three-dimensional model of XdPH based on its similarity with L-proline *cis*-4-hydroxylase of the *M. japonicum* strain LMG 29417. Comparison of the hydrophobicity on the molecular surfaces of wild-type XdPH (WT) and the L142R variant (L142R) using PyMOL (A). Hydrophilic regions are colored white, while hydrophobic regions are colored red. Analysis of the hydrophobic patches of WT XdPH and the L142R variant using the MOE program (B). Hydrophobic patches are colored green. Blue and yellow arrows indicate the positions of L142 and R142, respectively. Interaction of L142 (in WT XdPH) and R142 (in the L142R variant) with D191 and E194 (C). Carbon, nitrogen, and oxygen atoms are shown as green, blue, and red spheres, respectively.

L142R variant model was constructed by replacing L142 using the SWISS-MODEL.

The major factors causing inclusion body formation in *E. coli* are the protein's amino acid composition, its hydrophobicity, and overall net charge.<sup>20–23</sup> In this study, we demonstrated that the decrease in hydrophobicity through replacement of hydrophobic L142 with hydrophilic residues improves its soluble expression. Hydrophobicity analysis of the surface and hydrophobic patch of XdPH by homology modeling indicated a decrease in the surface hydrophobicity of the L142R variant (Figure 9A,B). Replacement with a hydrophilic residue may have reduced hydrophobicity of the hydrophobic site, resulting in improved solubility.<sup>24,25</sup> These results suggest that the hydrophilicity of amino acid residues on the surface of the protein is an important factor determining the soluble

expression of XdPH. When L142 was replaced with arginine, R142 was predicted to interact with residues D191 and E194 (Figure 9C). The bonds between R142 and the surrounding residues were analyzed using Ring 2.0 and Cytoscape. R142 showed increased hydrogen bonding with E194 (Figure S6). These results suggest that stabilization of the structure by strengthening the interaction between L142 and its surrounding residues is also an important factor for the soluble expression of XdPH.

## CONCLUSIONS

Recombinant gene technology using *E. coli* as a host has been successfully used for basic research and large-scale industrial production of proteins. However, many recombinant proteins often fail to fold correctly, resulting in aggregates known as

inclusion bodies. We established two rules: the  $\alpha$ -helix rule and the hydrophathy contradiction rule. By analyzing the target proteins to identify aggregation hotspots in the primary structures, rationally selected mutations can be introduced to improve the solubility of recombinant proteins in the *E. coli* expression system.<sup>7</sup> In this study, we introduced rationally selected alterations into XdPH using hotspot prediction methods to improve its soluble expression. To increase the soluble expression level of XdPH, aggregation hotspots identified using the  $\alpha$ -helix and hydrophathy contradiction rules were subjected to site-directed mutagenesis. For rational mutagenesis, 19 aggregation hotspots were predicted using the  $\alpha$ -helix and hydrophathy contradiction rules. Site-directed mutagenesis of the aggregation hotspots on XdPH was used to produce the V31E, V62D, C76Y, and L142R variants, which showed considerable improvements in the solubility of the enzyme in the *E. coli* expression system compared with the wild-type XdPH. The enzyme activity per unit of soluble protein displayed by the L142R variant was 14.6-fold higher than that of the wild-type XdPH. Furthermore, double mutations were introduced into the gene encoding the enzyme by site-directed mutagenesis, and the resulting double variants showed significantly improved soluble expression levels compared to the single variants. This indicates that the synergistic effects of these alterations increased the soluble expression of XdPH. The temperature–stability relationship between the wild-type XdPH and soluble XdPH variants was investigated. The XdPH variants with improved soluble expression retained their enzyme activity, whereas the wild-type showed no activity at 30 °C. Among the soluble XdPH variants, C76Y/L142R exhibited improved thermostability.

To investigate the effect of these alterations on soluble expression, L142 in XdPH was altered by site-directed mutagenesis. Replacement of L142 with hydrophilic residues improved its soluble expression; however, replacement of L142 with negatively charged residues resulted in a decrease in soluble protein expression. Homology modeling analysis also suggested that the hydrophilicity of the residue at position 142 might determine the solubility of XdPH when expressed in *E. coli*. Moreover, when L142 was replaced with arginine, homology modeling indicated that R142 interacted with the residues D191 and E194. The bonds between the residues surrounding L142 were examined, and the L142R variant was found to possess more hydrogen bonds than the wild-type XdPH. These results suggest that the stabilization of the protein structure by interaction between the residue at position 142 and its surrounding residues is an important factor determining the soluble expression of XdPH.

In this study, we demonstrated the effectiveness of the  $\alpha$ -helix rule and the hydrophathy contradiction rule in the rational design of soluble biocatalysts. Solubility was found to be an important factor determining the efficiency of biocatalytic production, from basic research to industrial applications.

## ■ EXPERIMENTAL SECTION

**Chemicals and Bacterial Strains.** *cis*-5-HyPip, L-Pip, and L-FDLA were “special grade” and purchased from Tokyo Chemical Industry Co., Ltd. (Tokyo, Japan). All other chemicals were also “special grade” and purchased from Kanto Chemical Co., Inc. (Tokyo, Japan) or Nacalai Tesque Co., Inc. (Kyoto, Japan), unless otherwise stated. *E. coli* DH5 $\alpha$  and BL21(DE3) were purchased from Nippon Gene Co., Ltd.

(Tokyo, Japan). The pET-28a(+) vector was purchased from Novagen (Darmstadt, Germany).

**Prediction of the Aggregation Hotspots in XdPH by the  $\alpha$ -Helix Rule and the Hydrophathy Contradiction Rule.** Based on the amino acid sequences of XdPH (accession no. CDG16639), aggregation hotspots were predicted according to the  $\alpha$ -helix rule and the hydrophathy contradiction rule. These calculations were performed using the INTMSAlign\_HiSol program according to a previous study.<sup>7</sup> The library file for the analysis using the INTMSAlign\_HiSol was created by collecting amino acid sequences similar to that of XdPH using a BLAST database search. Prediction of the secondary structure of XdPH by homology modeling was performed using the SWISS-MODEL.<sup>28</sup>

**Construction of the Expression Plasmid and Site-Directed Mutagenesis.** The gene encoding XdPH was codon-optimized, synthesized, and cloned into pJexpress411 by a service provider (ATUM, DNA2.0, CA). Based on the sequences of the XdPH genes, a primer pair was designed to construct the expression plasmid (Table S1). The XdPH gene was amplified by PCR, and a plasmid containing the synthesized gene was used as the template (Supporting Method S1). The DNA fragments were purified and ligated into a pET-28a(+) vector pretreated with *Nde*I and *Xho*I to obtain pET28-NHXdPH. In this study, the gene was expressed with an *N-terminal* hexahistidine tag fused to the recombinant protein. The insertion of the gene was confirmed by DNA sequencing using an ABI PRISM 310 genetic analyzer.

Site-directed mutagenesis of the XdPH gene was performed using the megaprimer PCR method.<sup>29</sup> The primers used for site-directed mutagenesis were designed based on the sequence of the XdPH gene (Table S1). The site-directed mutagenesis methods are described in the Supporting Method S2.

**Gene Expression and Analysis of Soluble Proteins.** *E. coli* BL21(DE3) cells were transformed with the constructed plasmid to express the enzyme. The experimental procedure for gene expression and soluble protein analysis is described in Figure S1. A glycerol stock for inoculation was prepared using a Luria-Bertani (LB) medium containing 50  $\mu$ g/mL kanamycin. Glycerol stocks (30  $\mu$ L) were inoculated into 3 mL of LB-Autoinduction medium (Overnight Express Auto-induction System1, Merck) containing 50  $\mu$ g/mL kanamycin and cultivated at 30 °C with shaking at 250 rpm. The bacteria were grown to an OD600 of ~0.6 to 0.8. Protein expression was induced at 15 °C for 24 h or at 30 °C for 12 h, with shaking at 250 rpm. After induction, 2 mL of the cultures was centrifuged at 10 000g for 5 min at 4 °C. The cell pellets were resuspended in 500  $\mu$ L of 100 mM 2-morpholinoethanesulfonic acid (MES) buffer (pH 6.5). Whole-cell lysates were prepared by disrupting the cell suspension via ultrasonication for 10 min on ice using a Bioruptor UCD-250 (TOSHO DENKI, Japan). Soluble and insoluble fractions were prepared from the whole-cell lysate by centrifugation at 20 000g for 20 min at 4 °C. The supernatant was collected as the soluble fraction, and the precipitate was resuspended in the same volume of MES buffer as the supernatant (insoluble fraction). Each fraction (2.5  $\mu$ L) was analyzed by SDS-PAGE using a 10% (w/v) polyacrylamide gel.

**Measurement of Enzymatic Activity.** The hydroxylation of L-Pip by the crude enzyme extract in the soluble fraction was assayed by determining the conversion of L-Pip to *cis*-5-HyPip using UPLC (ACQUITY UPLC, Waters, MA) and comparing the results to those of authentic standards. To determine

whether the enzyme recognized L-Pip as a substrate, the hydroxylation was conducted as follows: a 50  $\mu$ L reaction mixture containing 150 mM MES buffer (pH 6.5), 20 mM L-Pip, 40 mM  $\alpha$ -ketoglutarate, 1 mM L-ascorbate, 0.5 mM FeSO<sub>4</sub>, and 1 mg/mL of each crude enzyme extract was incubated at 20 °C for 10 min. After incubation, the amount of *cis*-5-HyPip in the reaction mixture was analyzed by UPLC after derivatization with L-FDLA.<sup>30</sup> The analytical methods are described in the [Supporting Method S3](#). The concentration of *cis*-5-HyPip was measured, and a peak was observed at 8.57 min ([Figure S2](#)). The amount of enzyme required to produce 1  $\mu$ mol of *cis*-5-HyPip per minute, using L-Pip as a substrate, was set to 1 U. Protein concentrations were estimated by the Bradford method using a Quick Start Bradford protein assay kit (Bio-Rad, CA) and bovine serum albumin standard.

**Structural Homology Modeling and Bioinformatic Analysis.** Structural models of the wild-type XdPH and the L142R variants of XdPH were constructed based on the crystal structure of L-proline *cis*-4-hydroxylase from the *M. japonicum* strain LMG 29417, obtained from the Protein Data Bank (PDB accession code 4P7W). Amino acid replacement was performed using the SWISS-MODEL. The PyMOL program was used to display the protein structure,<sup>31</sup> and the Molecular Operating Environment program (MOE version 2019, Montreal, Canada) was used for the hydrophobic patch analysis. The hydrogen bonds between adjacent ligands were analyzed to reveal the residue interaction network using Ring 2.0 and Cytoscape.<sup>32,33</sup> Amino acid sequence-based alignment of XdPH and other hydroxylases was performed and illustrated using GENETYX ver.12 (GENETYX, Tokyo, Japan).

## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.2c04247>.

Construction of expression plasmid (Supporting Method S1); outline of the megaprimer PCR method (Supporting Method S2); product analysis (Supporting Method S3); experimental procedure of the protein solubility test in the *E. coli* expression system (Figure S1); UPLC chromatogram of L-FDLA-derivatized *cis*-5-HyPip (Figure S2); UPLC analysis of enzyme reaction products by the XdPH wild-type (WT) and L142R variant (Figure S3); helical wheel depictions for  $\alpha$ -helix regions and mutations based on the  $\alpha$ -helix rule (Figure S4); comparison of the amino acid sequence of XdPH and other hydroxylases (Figure S5); network analyses and the comparison of XdPH wild type and L142R with Ring 2.0 and Cytoscape (Figure S6); list of primers (Table S1); and the amino acid sequences for the analysis using the INTMSAlign\_HiSol program (Supporting Data S1) (PDF)

### Accession Codes

XdPH: CDG16639, L-proline *cis*-4-hydroxylase from the *Mesorhizobium japonicum* strain LMG 29417: Q989T9 and 4P7W.

## ■ AUTHOR INFORMATION

### Corresponding Author

Yasuhisa Asano – Biotechnology Research Center and Department of Biotechnology, Toyama Prefectural University,

Imizu-shi, Toyama 939-0398, Japan; [orcid.org/0000-0003-3645-3952](https://orcid.org/0000-0003-3645-3952); Email: [asano@pu-toyama.ac.jp](mailto:asano@pu-toyama.ac.jp)

### Authors

Suguru Shinoda – Biotechnology Research Center and Department of Biotechnology, Toyama Prefectural University, Imizu-shi, Toyama 939-0398, Japan

Aoi Itakura – Biotechnology Research Center and Department of Biotechnology, Toyama Prefectural University, Imizu-shi, Toyama 939-0398, Japan

Haruka Sasano – Science & Innovation Center, Mitsubishi Chemical Corporation, Yokohama-shi, Kanagawa 227-8502, Japan

Ryoma Miyake – Science & Innovation Center, Mitsubishi Chemical Corporation, Yokohama-shi, Kanagawa 227-8502, Japan; [orcid.org/0000-0002-8094-1614](https://orcid.org/0000-0002-8094-1614)

Hiroshi Kawabata – Science & Innovation Center, Mitsubishi Chemical Corporation, Yokohama-shi, Kanagawa 227-8502, Japan; API Corporation, Tokyo 101-0047, Japan; Present Address: Innovation & Entrepreneurship Division, Mitsubishi Chemical Corporation, 1-1, Marunouchi 1-chome, Chiyoda-ku, Tokyo 100-8251, Japan

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsomega.2c04247>

### Author Contributions

S.S., R.M., and Y.A. designed the study. S.S., A.I., H.S., and R.M. performed the experiments. S.S., A.I., and H.K. performed data analysis. S.S., A.I., and Y.A. wrote the manuscript.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This research was supported by a Grant-in-Aid for Scientific Research (S) and (A) from the Japan Society for the Promotion of Science (Grant Nos. 17H06169 and 22H00361, respectively) awarded to Y.A.

## ■ ABBREVIATIONS

XdPH, 2-oxoglutarate-dependent L-pipecolic acid hydroxylase from *Xenorhabdus doucetiae*; *cis*-5-HyPip, *cis*-5-hydroxy-L-pipecolic acid; L-Pip, L-pipecolic acid; L-FDLA, *N*-(5-fluoro-2,4-dinitrophenyl)-L-leucinamide

## ■ REFERENCES

- (1) Sørensen, H. P.; Mortensen, K. K. Soluble Expression of Recombinant Proteins in the Cytoplasm of *Escherichia coli*. *Microb. Cell Fact.* **2005**, *4*, 1.
- (2) Ventura, S.; Villaverde, A. Protein Quality in Bacterial Inclusion Bodies. *Trends Biotechnol.* **2006**, *24*, 179–185.
- (3) Sørensen, H. P.; Mortensen, K. K. Advanced Genetic Strategies for Recombinant Protein Expression in *Escherichia coli*. *J. Biotechnol.* **2005**, *115*, 113–128.
- (4) Vincentelli, R.; Bignon, C.; Gruez, A.; Canaan, S.; Sulzenbacher, G.; Tegoni, M.; Campanacci, V.; Cambillau, C. Medium-Scale Structural Genomics: Strategies for Protein Expression and Crystallization. *Acc. Chem. Res.* **2003**, *36*, 165–172.
- (5) Rosano, G. L.; Ceccarelli, E. A Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* **2014**, *5*, 172.
- (6) Tokuriki, N.; Tawfik, D. S. Chaperonin Overexpression Promotes Genetic Variation and Enzyme Evolution. *Nature* **2009**, *459*, 668–673.

- (7) Matsui, D.; Nakano, S.; Dadashipour, M.; Asano, Y. Rational Identification of Aggregation Hotspots Based on Secondary Structure and Amino Acid Hydrophobicity. *Sci. Rep.* **2017**, *7*, No. 9558.
- (8) Lamarre, D.; Croteau, G.; Wardrop, E.; Bourgon, L.; Thibeault, D.; Clouette, C.; Vaillancourt, M.; Cohen, E.; Pargellis, C.; Yoakim, C.; Anderson, P. C. Antiviral Properties of Palinavir, a Potent Inhibitor of the Human Immunodeficiency Virus Type 1 Protease. *Antimicrob. Agents Chemother.* **1997**, *41*, 965–971.
- (9) Miller, S. P.; Zhong, Y. L.; Liu, Z.; Simeone, M.; Yasuda, N.; Limanto, J.; Chen, Z.; Lynch, J.; Capodanno, V. Practical and Cost-Effective Manufacturing Route for the Synthesis of a  $\beta$ -Lactamase Inhibitor. *Org. Lett.* **2014**, *16*, 174–177.
- (10) Shibasaki, T.; Sakurai, W.; Hasegawa, A.; Uosaki, Y.; Mori, H.; Yoshida, M.; Yoshida, M.; Ozaki, A. Substrate Selectivities of Proline Hydroxylases. *Tetrahedron Lett.* **1999**, *40*, 5227–5230.
- (11) Shibasaki, T.; Mori, H.; Chiba, S.; Ozaki, A. Microbial Proline 4-Hydroxylase Screening and Gene Cloning. *Appl. Environ. Microbiol.* **1999**, *65*, 4028–4031.
- (12) Mori, H.; Shibasaki, T.; Yano, K.; Ozaki, A. Purification and Cloning of a Proline 3-Hydroxylase, a Novel Enzyme Which Hydroxylates Free L-Proline to *cis*-3-Hydroxy-L-Proline. *J. Bacteriol.* **1997**, *179*, 5677–5683.
- (13) Ryoma, M.; Yasumasa, D. Method for Manufacturing *cis*-5-Hydroxy-L-Pipecolic Acid. WIPO Patent WO2016/076159 A1, 2016.
- (14) Chaturvedi, S. S.; Rajeev, R.; Jian, H.; Robert, P. H.; Christo, Z. C. Atomic and Electronic Structure Determinants Distinguish between Ethylene Formation and L-Arginine Hydroxylation Reaction Mechanisms in the Ethylene-Forming Enzyme. *ACS Catal.* **2021**, *11*, 1578–1592.
- (15) Wojdyla, Z.; Borowski, T. Properties of the Reactants and Their Interactions within and with the Enzyme Binding Cavity Determine Reaction Selectivities. The Case of Fe(II)/2-Oxoglutarate Dependent Enzymes. *Chemistry* **2022**, *28*, No. e202104106.
- (16) Kleman, G. L.; Strohl, W. R. Developments in High Cell Density and High Productivity Microbial Fermentation. *Curr. Opin. Biotechnol.* **1994**, *5*, 180–186.
- (17) Ventura, S. Sequence Determinants of Protein Aggregation: Tools to Increase Protein Solubility. *Microb. Cell Fact.* **2005**, *4*, 11.
- (18) Nakaniwa, T.; Fukada, H.; Inoue, T.; Gouda, M.; Nakai, R.; Kirii, Y.; Adachi, M.; Tamada, T.; Segawa, S.; Kuroki, R.; Tada, T.; Kinoshita, T. Seven Cysteine-Deficient Mutants Depict the Interplay between Thermal and Chemical Stabilities of Individual Cysteine Residues in Mitogen-Activated Protein Kinase c-Jun N-Terminal Kinase 1. *Biochemistry* **2012**, *51*, 8410–8421.
- (19) Koketsu, K.; Shomura, Y.; Moriwaki, K.; Hayashi, M.; Mitsuhashi, S.; Hara, R.; Kino, K.; Higuchi, Y. Refined Regio- and Stereoselective Hydroxylation of L-Pipecolic Acid by Protein Engineering of L-Proline *cis*-4-Hydroxylase Based on the X-ray Crystal Structure. *ACS Synth. Biol.* **2015**, *4*, 383–392.
- (20) Yang, J. K.; Park, M. S.; Waldo, G. S.; Suh, S. W. Directed Evolution Approach to a Structural Genomics Project: Rv2002 from *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 455–460.
- (21) Wigley, W. C.; Stidham, R. D.; Smith, N. M.; Hunt, J. F.; Thomas, P. J. Protein Solubility and Folding Monitored In Vivo by Structural Complementation of a Genetic Marker Protein. *Nat. Biotechnol.* **2001**, *19*, 131–136.
- (22) Pédelacq, J.-D.; Piltch, E.; Liong, E. C.; Berendzen, J.; Kim, C. Y.; Rho, B. S.; Park, M. S.; Terwilliger, T. C.; Waldo, G. S. Engineering Soluble Proteins for Structural Genomics. *Nat. Biotechnol.* **2002**, *20*, 927–932.
- (23) Esteban, O.; Zhao, H. Directed Evolution of Soluble Single-Chain Human Class II MHC Molecules. *J. Mol. Biol.* **2004**, *340*, 81–95.
- (24) Mosavi, L. K.; Peng, Z. Y. Structure-Based Substitutions for Increased Solubility of a Designed Protein. *Protein Eng. Des. Sel.* **2003**, *16*, 739–745.
- (25) Wurth, C.; Guimard, N. K.; Hecht, M. H. Mutations that Reduce Aggregation of the Alzheimer's A $\beta$ 42 Peptide: an Unbiased Search for the Sequence Determinants of A $\beta$  Amyloidogenesis. *J. Mol. Biol.* **2002**, *319*, 1279–1290.
- (26) Nakano, S.; Asano, Y. Protein Evolution Analysis of S-Hydroxynitrile Lyase by Complete Sequence Design Utilizing the INTMSAlign Software. *Sci. Rep.* **2015**, *5*, No. 8193.
- (27) Kyte, J.; Doolittle, R. F. A Simple Method for Displaying the Hydrophobic Character of a Protein. *J. Mol. Biol.* **1982**, *157*, 105–132.
- (28) Biasini, M.; Bienert, S.; Waterhouse, A.; Arnold, K.; Studer, G.; Schmidt, T.; Kiefer, F.; Gallo, C. T.; Bertoni, M.; Bordoli, L.; Schwede, T. SWISS-MODEL: Modelling Protein Tertiary and Quaternary Structure using Evolutionary Information. *Nucleic Acids Res.* **2014**, *42*, W252–258.
- (29) Ke, S. H.; Madison, E. L. Rapid and Efficient Site-Directed Mutagenesis by Single-tube 'Megaprimer' PCR Method. *Nucleic Acids Res.* **1997**, *25*, 3371–3372.
- (30) Fujii, K.; Ikai, Y.; Oka, H.; Suzuki, M.; Harada, K. A Nonempirical Method Using LC/MS for Determination of the Absolute Configuration of Constituent Amino Acids in a Peptide: Combination of Marfey's Method with Mass Spectrometry and Its Practical Application. *Anal. Chem.* **1997**, *69*, 5146–5151.
- (31) Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 2.3.4, 2019.
- (32) Piovesan, D.; Minervini, G.; Tosatto, S. C. The RING 2.0 Web Server for High Quality Residue Interaction Networks. *Nucleic Acids Res.* **2016**, *44*, W367–W374.
- (33) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504.