Functional expression of a plant hydroxynitrile lyase in *Escherichia coli* by directed evolution: creation and characterization of highly in vivo soluble mutants

Yasuhisa Asano¹, Mohammad Dadashipour, Mizue Yamazaki, Nobutaka Doi and Hidenobu Komeda

Biotechnology Research Center and Department of Biotechnology, Toyama Prefectural University, 5180 Kurokawa, Imizu, Toyama 939-0398, Japan

¹To whom correspondence should be addressed. E-mail: asano@pu-toyama.ac.jp

Received November 16, 2010; revised May 19, 2011; accepted May 27, 2011

Edited by Liz Meiering

Low protein solubility of recombinantly expressed proteins in *Escherichia coli* is a major factor hindering their application and analysis. We generated highly in vivo soluble mutants of a hydroxynitrile lyase in *E.coli* using protein engineering. Structure-guided saturation mutagenesis caused high solubility of single Lys–Pro mutations at positions 176, 199 and 224 of this low soluble wild-type enzyme. The triple Lys–Pro mutant generated at these surface conserved residues showed up to 8-fold increase in specific activity in the cell-free extract. Random mutagenesis also created a mutant of His103Met with 18.5-fold increase. The main expression form was reversed from insoluble to the soluble fraction following both types of above-mentioned mutations in *E.coli* at 37°C. The findings challenge the rationale of producing recombinant proteins in this host at 37°C. Formerly wild type low soluble protein was then present as soluble protein by these mutations, which also elevated the total soluble protein fraction in *E.coli*. Saturation mutagenesis of His103 provided other highly soluble mutants with hydrophobic substitutions. These mutations caused only minor secondary structural changes as determined by circular dichroism and Fourier-transform infrared spectroscopy and affected catalytic efficiency slightly for the purified mutants (0.82–1.6-fold for benzaldehyde and 0.9−1.9-fold for mandelonitrile). The stability of the mutants was differed from that of the wild type at high temperatures and at pH >8. Exchanging the buried basic-polar residue His103 with hydrophobic amino acids is in line with the overall structure of the enzyme, i.e. having hydrophilic residues in solvent-exposed areas and hydrophobic residues in the core.

**Keywords:** directed evolution/functional expression/hydroxynitrile lyase/mutagenesis/protein engineering/protein solubility

**Introduction**

Numerous advantages of *Escherichia coli* render it the first choice for recombinant gene expression, but there can be the problem of low or poor protein solubility (Makrides, 1996). It seems that the evolutionary-imposed sequence and conformation of a protein correspond to its optimum function in an organism. Therefore, in order to produce functional proteins in *E.coli*, approaches should be found to tackle this obstacle.

Although various selection methods such as the use of different fusion proteins or folding assistants including molecular chaperones exist, our focus in this research is on the application of protein engineering for enhancing in vivo protein solubility of recombinantly expressed genes in *E. coli*. Directed evolution has increasingly being used in academic and industrial laboratories for improving protein stability, altering substrate specificity and stereoselectivity as well as designing new activities, i.e. enzyme promiscuity (May et al., 2000; Kazlauskas and Bornscheuer, 2009; Turner, 2009). A combination of directed evolution and high-throughput screening of mutant libraries enables selection of re-designed molecules (Yuan et al., 2005). In one of the first directed evolution reports, Arnold and Moore (1996) enhanced the total activity of an esterase 50−60-fold, where the selected purified mutants were also studied for their kinetic parameters. Production of the food additive 5'-IMP (5'-inosine mono phosphate) by a mutant nucleoside pyrophosphate phosphotransferase was one of the first successful uses of directed evolution in industrial processes (Mihara et al., 2000; Ishikawa et al., 2002). Mutant horseradish peroxidase was expressed in *E.coli* at very low levels; therefore, it needed to be expressed in *Saccharomyces cerevisiae* to achieve a considerable increase in total enzymatic activity (Lin et al., 1999; Morawski et al., 2000). Other examples of the successful use of directed evolution in the functional expression of a fungal laccase and pig liver esterase were also limited to *S. cerevisiae* and *Pichia pastoris* (Bulter et al., 2003; Musidlovska-Persson and Bornscheuer, 2003). The physicochemical and kinetic properties of mutants were studied only for the above-mentioned fungal laccase (Bulter et al., 2003). Several research groups have tried to enhance the in vitro solubility of pharmaceutical proteins such as human versions of growth hormone, granulocyte colony-stimulating factor, insulin and various monoclonal antibodies (Trevino et al., 2005) and only limited cases have dealt with in vivo protein solubility using *E.coli*.

5'-hydroxynitrile lyases (HNLs) catalyze the reversible synthesis and degradation of cyanohydrins (Fig. 1), which are important chiral intermediates for various pharmaceuticals,
Materials and methods

Materials, enzymes, chemicals and bacterial strains

Potassium cyanide was purchased from Wako Pure Chemical Industries, Ltd (Osaka, Japan). Oligonucleotides were synthesized by Hokkaido System Science Co., Ltd (Tsukuba, Japan). Takara Ex Taq DNA polymerase was obtained from Takara Shuzo (Shiga, Japan) and the restriction endonucleases from Takara Shuzo, Toyobo (Osaka, Japan) and New England Biolabs (Beverly, MA, USA). Shrimp alkaline phosphatase and Pwo DNA polymerase were purchased from Roche Diagnostics GmbH (Penzberg, Germany) and ligation Kit, ver. 2 was from Takara Shuzo. All other chemicals were purchased from commercial sources and used without further purification. Escherichia coli JM109 and BL21 (DE3) were used as host cells for the expression plasmids, derived from pUC19 (Takara Shuzo) and pET28b (Novagen, Madison, WI, USA) as cloning vectors, respectively.

Synthesis of the gene for hydroxynitrile lyase from M. esculenta (cassava) by assembly PCR, cloning and expression in E. coliJM109

The cDNA sequence of the hnl gene from M. esculenta (GenBank: Z29091) was used to construct the full-length gene with oligonucleotides listed in Supplementary Table SI, under conditions and programs described in the Supplementary materials, Sections 1–3. Twenty primers (F01–F10 and R01–R10) were designed to overlap by 20 bases. The hnl gene with a plant codon was ligated into the PsiI and BamHI sites of a pUC19 to construct the expression plasmid pUMESD and then used to transform E. coli JM109. For the synthesis of an E. coli codon-optimized hnl gene, we used 30 primers (SY01–SY30, Supplementary Table SI) and cloned the gene in the SphiI and BamHI sites of pUC19. A Shine-dalgarno sequence was added to both constructs, as mentioned in Section 2 of Supplementary materials (Kane, 1995; Siemmer et al., 1995; Baedeker and Schulz, 1999).

Cloning of the latter gene in pUC19 resulted in the plasmid pUMESDsy (sy refers to a synthetic or E. coli codon-optimized gene). Pre-cultures grown at 37°C were used to inoculate Luria–Bertani (LB) broth supplemented with ampicillin (50 μg/ml) and incubated for 4–5 h at 37°C, then induced by 0.1 mM isopropyl β-D-thiogalactoside (IPTG) and incubated overnight. After disruption the cells with a Yasui Kikai Multi-Beads Shocker MB 601U (S) (Osaka, Japan), cell-free extracts were used to measure activity, protein concentration and the protein expression profile by 12–15% concentration and the protein expression profile by 12–15% agarose gel electrophoresis. During the screening of the library of mutants, we also used an enzymatic-freeze–thaw shock for the preparation of cell-free extracts, which is presented in details in the Supplementary materials Section 6.

Mutagenesis of MeHNL and its mutants and high-throughput screening of the resulting library for highly active and soluble mutant enzymes

Briefly, for the mutagenesis of conserved surface lysine residues (Fig. 2A and B), a polymerase chain reaction (PCR) was carried out with the hnl gene as a template using the primers listed in Supplementary Table SII. The NNS codons were used to carry out saturation mutagenesis of these six lysine residues with a QuikChange® site-directed mutagenesis kit (Stratagene, La Jolla, CA, USA). The PCR product was digested by DpnI and used to transform E. coli JM109. For high-throughput screening, the colonies were cultivated in 96-deep well plates (ABgene, Epsom, UK) containing 160 μl of LB broth supplemented with ampicillin and 0.1 mM IPTG at 37°C for 12 h on a Bioshaker MBR-024 (Taitec, Saitama, Japan) with agitation at 900 rpm. After master plates were provided, the cells were harvested by centrifugation (2000 × g for 10 min), lysed with a lysozyme solution (37°C, 1 h) and by three cycles of freezing–thawing, and suspended in a hypotonic solution (10 mM potassium phosphate buffer and 5 mM MgCl2). A cell-free extract was prepared by centrifugation of the suspension. These methods were explained in more details in Section 6 of Supplementary materials. The resulting library was screened for highly active mutants (see activity measurements). The single Lys–Pro mutations at positions 176, 199 and 224 were combined in both plant and E. coli codon-optimized genes in the plasmids pUMESD and pUMESDsy, respectively. Details are described in Section 4 of Supplementary materials.

Random mutagenesis was carried out on the hnl gene by error prone PCR (Supplementary materials, Section 5) and the resulting library of mutant genes was screened as described for Lys–Pro mutations. Saturation mutagenesis of MeHNL-His103 was done with the QuikChange® site-directed mutagenesis kit (Stratagene). The mutagenic primers containing NNS at the center were used and all substitutions were determined by sequencing. Details are given in Section 6 of Supplementary materials.

Activity measurement

Activity was measured using two methods. First, during the screening step, activity was measured by recording the kinetic slope at 280 nm photometrically in 96-well UV-Star
Microplates (Greiner Bio-One, Frickenhausen, Germany) with a Tecan GENios Microplate reader (Osaka, Japan). The data were analyzed using the software LS-PLATE Manager2001 (Wako, Osaka, Japan) in a cyanohydrin cleavage reaction containing 10 mM mandelonitrile as substrate, 100 mM citrate buffer pH 5.5 and the enzyme sample. The amount of benzaldehyde produced was monitored for 10 min, then the data plotted and compared with the control, i.e. wild-type MeHNL expressed in E.coli JM109. An extinction coefficient of \( e_{280} = 1.4 \text{ mM}^{-1} \text{ cm}^{-1} \) was used for measurement of the benzaldehyde.

Second, after selection of the highly in vivo soluble mutants, activity was determined by a high-performance liquid chromatography (HPLC)-based method in a cyanohydrin synthesis reaction, by measuring the optically active mandelonitrile synthesized from benzaldehyde and cyanide (Asano et al., 2005). The solvent system used was modified slightly (\( n \)-hexane: isopropanol = 85:15 v/v). The retention times of benzaldehyde, \((R)\)- and \((S)\)-mandelonitrile were \( \sim 5, 10 \) and 12 min, respectively. One unit of activity was defined as the amount of enzyme that produced 1 mmol of optically active mandelonitrile from benzaldehyde per min under standard assay conditions.

SDS-PAGE and protein concentrations

SDS-PAGE slab gels (12–15% acrylamide) were used to monitor the expression profile of the proteins in soluble and insoluble fractions as described by Laemmli (1970). Coomassie Brilliant Blue G-250 was utilized for staining the gels. Protein concentrations were determined using a Bio-Rad protein assay Kit (Hercules, CA, USA) with bovine serum albumin (Wako, Osaka, Japan) as a standard and measuring the absorbance at 595 nm (Bradford, 1976).

Whole-cell reaction and activity measurements for the His103-saturated mutants in cell-free extracts

Pre-cultures of each E.coli transformant containing one His103 mutant of MeHNL with both plant and synthetic (E.coli codon-optimized) genes were cultivated in 3 ml of LB broth supplemented with ampicillin, incubated at 37° C for 12 h, and then induced by 0.1 mM IPTG and incubated at 37° C for a further 12 h. The cells were harvested by centrifugation and re-suspended in 20 mM potassium phosphate buffer, pH 7.0, and the suspension was used for a whole-cell reaction at 25° C for 5 min. Then, 100 µl of the reaction mixture was extracted in 0.9 ml of organic solvent as mentioned in the activity measurements section.

Subcloning of MeHNL and the mutants into pET28b, expression in E.coli BL21 (DE3) and purification of the enzymes

Genes encoding the highly soluble single mutant MeHNL-Lys176Pro, triple mutant MeHNL-Lys176Pro, Lys199Pro, Lys224Pro, single mutants MeHNL-His103Met and His103Leu, and inactive mutant His103Tyr were amplified, digested with XhoI and Ndel, cloned into pET28b (Novagen, Madison, WI, USA), expressed in E. coli BL21...
De(3), and induced by IPTG. The wild type and mutant enzymes were purified from the cells of E. coli BL21 (DE3) transformants with a fast-flow nickel-sepharose column, according to the manufacturer’s instructions (GE Healthcare, Uppsala, Sweden) following disruption of the cells by sonication using an Insonator 201M (Kubota, Tokyo, Japan). Thrombin protease (GE Healthcare) was used to remove the fused part derived from the pET28b vector.

Temperature and pH stability of MeHNL and its highly active and inactive mutants
Stability was determined following 60 min of incubation of the enzymes at a temperature of 20–80°C (pH 6.0). The same amounts of the enzymes were also incubated for 60 min in various pH values (pH 3–8 in citrate-phosphate and 8–11 in glycine-NaOH buffers at 20°C), then the mixtures were used for activity measurements by the HPLC method. All experiments were done in triplicate and average results are shown, only otherwise stated.

Kinetic experiments on MeHNL and its mutants in the cyanohydrin synthesis and decompositions reactions
Various concentrations of mandelonitrile (0.5–15 mM) and benzaldehyde (1–150 mM) were used to compare the kinetic parameters of MeHNL and its mutants at room temperature (22°C).

Circular Dichroism, fluorescence, and FT-IR spectra of the wild type and mutant enzymes
The secondary structure and conformational changes of the highly active mutants His103Leu, His103Met, triple Lys–Pro, and inactive mutant His103Tyr were compared with that of the wild type by Circular Dichroism (CD), Fourier-transform infrared (FT-IR), and fluorescence spectrometry. A concentration of 0.4 mg/ml of each protein in the presence of 20 mM potassium phosphate buffer, pH7.0 was used for the analysis in a final volume of 500 μl. CD spectra were recorded at 37 and 20°C with a 2-mm path length rectangular quartz cuvette at Far-UV (260–200 nm) in a denaturant-free solution (0.5 M Guanidium HCl) using a Jasco 720 CD spectropolarimeter (Tokyo, Japan). The CD ellipticity was calculated according to the equation \[ \theta = (\text{CD signal} \times 100)/C \times l \times n \] (Maki et al., 1999), where \( \theta \) is mean residue ellipticity in deg cm²/mmol, CD signal is raw output in millidegrees, C is the protein concentration in mm, n is the number of amino acid residues in the protein, and l is the cell path length in cm (Myers et al., 1997; Kelly et al., 2005).

Each enzyme was monitored in denaturant-free solution for conformational changes by recording fluorescence spectra (\( \lambda_{ex} = 280 \text{ nm} \) and \( \lambda_{em} = 300–400 \text{ nm} \)) with a Shimadzu RF-5300PC (Kyoto, Japan). The graphs were normalized by the KaleidaGraph (Synergy Software, Reading, PA, USA). The Amide-I region (1580–1720/cm) of FT-IR spectra of the purified enzymes were recorded using a PerkinElmer IR Spectrum 100 (Waltham, MA, USA).

Results
The creation of highly soluble mutants of an industrially important enzyme using protein engineering methods and their characterization will be described hereafter. These mutations include both the surface Lys–Pro and substitutions on the buried His103 residues of HNL from M. esculenta.

Creation of the highly in vivo soluble mutants of recombinant MeHNL in E. coli
Primary structure of MeHNL contains 20 lysine residues, out of which 15 are conserved in that of Hevea brasiliensis, HbHNL, probably indicating a structural or functional role for these residues. There are also 10 conserved Lys residues between the S-MeHNL and the newly discovered S-HNL from Baliospermum montanum (Dashadipour et al., 2011). Resistance of these enzymes to aldehydes as substrates would be an important factor in biocatalysis. There is also a relationship between decreasing the number of solvent-exposed (surface) lysine residues and increasing the enzymatic resistance to aldehydes in previous studies (Habeb and Hiramoto, 1968). Based on this, we employed the crystal structure of the S-MeHNL (1DWP) to identify the conserved surface lysines that are located on the loops ending in the β-sheets because there is a possibility of increasing the temperature stability by increasing of β-sheet content of the enzyme (Yan et al., 2003). Meanwhile we excluded the lysine residues located on the α-helices or β-sheets of the enzyme. RasMol was used to observe the residues located on the structure of the enzyme. Then, these five (i) conserved, (ii) solvent exposed and (iii) close to β-sheet lysines were subjected to site-saturation mutagenesis to alter them in order to monitor their effects on protein property and structure. The Lys237 was also chosen which was close to the active site residue His236. In another strategy, a random mutagenesis was resulted to a dramatic enhancement in protein solubility in E.coli, which has been described properly below.

Creation of surface Lys–Pro mutants
Fig. 2A shows the alignment of these three molecularly cloned S-HNLs accompanying the catalytic triad and Lys residues in the primary structure. The conserved surface lysine residues selected for site-saturation mutagenesis are depicted in Fig. 2B (molecule A of the homodimer MeHNL, PDB 1DWP). These structure-guided site-saturation mutagenesis reactions resulted in highly in vivo soluble and active mutants of recombinant MeHNL heterologously expressed in E.coli, instead of aldehyde resistance or temperature stability. A high-throughput screening was used to search the resulting libraries of site-saturation mutagenesis of these six lysine residues. Random clones were picked and sequenced to ascertain the diversity of the library and the mutants were tested primarily for their activity compared with the wild-type MeHNL. The most active mutants were all found to have Lys–Pro mutations at positions 176, 199 and 224. To reconfirm the high activity observed in the screening step, the selected clones containing the Lys176Pro, Lys199Pro, and Lys224Pro mutations were assayed with the more sensitive method of HNL activity determination, i.e. by an HPLC. Subsequent evaluation of protein expression profiles of the wild type and these Lys–Pro mutants revealed a change in the solubility of the enzyme after these mutations. Then the double and triple mutants of Lys–Pro were generated at positions 176, 199 and 224. Fig. 3A shows the soluble fractions of the single, double and triple Lys–Pro mutants at above-mentioned positions. The triple mutant MeHNL_Lys176Pro, Lys199Pro and
The same Lys–Pro mutations were also generated in the pUMESD encoding the MeHNL from the natural plant codon and resulted in a 8.2-fold increase in the specific activity for the triple mutant Lys176Pro, Lys199Pro and Lys224Pro in the cell-free extract (Table I). Although the specific activity levels and protein expression profiles (in SDS-PAGE) were similar for each of the E. coli codon-optimized and plant genes encoding MeHNL and its mutants, the mutations of the plant hnl genes showed higher improvement. The minor differences possibly resulted from variations in the gene and mRNA structures.

**Creation of mutants of buried residue His103**

Given the encouraging results on the in vivo solubility obtained with the Lys-Pro mutations, the hnl gene encoding MeHNL was randomly mutated by error-prone PCR and ~5000 clones were tested for ‘very high enzymatic activity’ in the cell-free extract using high-throughput screening. Ultimately, a ‘highly active mutant’ was identified as MeHNL-His103Leu, although there were some other more active mutants than the wild-type enzyme, which had been ignored. Details of the screening and assay were presented in the Materials and methods section. According to the bands on SDS-PAGE, soluble and insoluble fractions were estimated to account for ~75 and 25% of the His103Leu mutant, and <5 and >95% of the wild type, respectively (Supplementary Figure S1).

The contribution of all 20 t-amino acids to enzymatic activity and in vivo solubility was evaluated based on the results of site-saturation mutagenesis of buried residue His103 in both plant and E. coli codons. Following a whole-cell activity measurement (Supplementary Figure S2), activity levels of the MeHNL-His103 substitutes were compared and the results were shown in Table I. The greatest improvement in specific activity (>18-fold increase) was obtained for the mutant His103Met: increased production of the enzyme resulted in higher specific activity in the cell-free extract in comparison with the wild-type enzyme. Besides, there were high levels of specific activity for Ile→Leu>Val>Cys (12 to 17.6-fold increase), and for Gln>Ser>Thr>Ala (2.5–5.5-fold) in the cell-free extract. These results were reconfirmed by SDS-PAGE of the corresponding soluble fractions of the mutants expressed in E. coli JM109, as illustrated in Fig. 3C.

For the E.coli codon-optimized gene encoding the mutant MeHNL-His103, there was a maximum 11.4-fold improvement of specific activity in the cell-free extract for the His103Leu substitution. There were other highly active mutants including those with substitutions of His103 for other aliphatic hydrophobic residues such as Met, Ile and Val and Cys (8.1–10.4-fold), as well. There were also high levels of specific activity increase for Gln, Thr and Ser mutants of the E. coli codon-optimized gene (4, 3.4 and 2.9-fold, respectively). Table I shows a detailed comparison between the wild type and >50 mutants generated in this study. It contains the data of Lys–Pro substitutions at positions 176, 199 and 224 and also those of the saturated mutants of His103 residue both for the plant and the E. coli optimized codons. Furthermore, the Lys→Pro and His103Leu mutants were combined, but with no further improvements in activity (data not shown).

**Fig. 3.** Expression profiles of MeHNL following the mutations. (A) SDS-PAGE of soluble fractions of the single and multiple Lys–Pro mutations of MeHNL (in 10 μg of protein for each well). M: Bio-Rad low-molecular-weight protein marker; single mutants 1: Lys176Pro; 2: Lys199Pro; 3: Lys224Pro; double mutants 4: Lys176Pro, Lys224Pro; 5: Lys199Pro, Lys224Pro; triple mutant 6: Lys176Pro, Lys199Pro, Lys224Pro; 7: MeHNL (wild type, E.coli codon-optimized gene); 8: E.coli JM109/pUC19; the arrow shows location of the MeHNL. (B) Soluble and insoluble fractions of expression of MeHNL, His103Leu and triple Lys-Pro at 37 (1–6) and 30°C (7–12), respectively. Marker was shown in the middle. (C) Soluble fractions of the MeHNL-His103 saturated mutants (plant gene) appeared in the SDS-PAGE. (See also Table I). Highly in vivo soluble mutants were shown especially for Val, Leu, Ile, Met, Cys and in a lower level for Gln, Ser, Thr and Ala.
obtained from the purified enzymes, beside Table I compares this mutant. All the data presented in this paper were this finding. Therefore, it was possible to purify a form of the specific activity of almost 50 proteins with the wild type.

Hydroxynitrile lyase mutants

<table>
<thead>
<tr>
<th>Plant codon</th>
<th>E. coli optimized codon</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total activity (U/mg)</td>
</tr>
<tr>
<td>MeHNL-His103</td>
<td>6.5 1.54 4.22 2.48</td>
</tr>
<tr>
<td>Ala</td>
<td>0.33 1.63 0.12 0.12</td>
</tr>
<tr>
<td>Arg</td>
<td>0.8 1.72 0.27 0.27</td>
</tr>
<tr>
<td>Asn</td>
<td>0.33 1.68 0.12 0.12</td>
</tr>
<tr>
<td>Asp</td>
<td>36 1.76 20.5 12</td>
</tr>
<tr>
<td>Cys</td>
<td>0.54 1.64 0.33 0.19</td>
</tr>
<tr>
<td>Gly</td>
<td>15 1.64 9.15 5.38</td>
</tr>
<tr>
<td>Glu</td>
<td>0.74 1.50 0.5 0.29</td>
</tr>
<tr>
<td>His (wt)</td>
<td>2.55 1.55 1.70 1</td>
</tr>
<tr>
<td>Ile</td>
<td>46.5 1.55 30 17.6</td>
</tr>
<tr>
<td>Leu</td>
<td>56.9 1.90 29.9 17.6</td>
</tr>
<tr>
<td>Lys</td>
<td>0.41 1.62 0.25 0.15</td>
</tr>
<tr>
<td>Met</td>
<td>58.8 1.87 31.4 18.5</td>
</tr>
<tr>
<td>Phe</td>
<td>1.09 1.63 0.67 0.39</td>
</tr>
<tr>
<td>Pro</td>
<td>0.12 1.60 0.08 0.05</td>
</tr>
<tr>
<td>Ser</td>
<td>10.5 1.65 6.36 3.74</td>
</tr>
<tr>
<td>Thr</td>
<td>7.95 1.56 5.1 3</td>
</tr>
<tr>
<td>Trp</td>
<td>0.9 1.48 0.6 0.35</td>
</tr>
<tr>
<td>Tyr</td>
<td>0.05 1.0 0.03 0.02</td>
</tr>
<tr>
<td>Val</td>
<td>45 1.71 26.3 15.5</td>
</tr>
<tr>
<td>MeHNL-Lys–Pro 176</td>
<td>6.68 1.78 3.75 2.2</td>
</tr>
<tr>
<td>199</td>
<td>5.3 1.78 2.9 1.75</td>
</tr>
<tr>
<td>224</td>
<td>8.3 1.76 4.71 2.77</td>
</tr>
<tr>
<td>176, 199</td>
<td>14.1 1.93 7.3 4.29</td>
</tr>
<tr>
<td>176, 224</td>
<td>17.6 1.81 9.72 5.72</td>
</tr>
<tr>
<td>199, 224</td>
<td>13.5 1.59 8.5 5</td>
</tr>
<tr>
<td>176, 199, 224</td>
<td>24.3 1.73 14 8.24</td>
</tr>
</tbody>
</table>

aAll assays have been monitored using a HPLC at 254 nm (synthesis of mandelonitrile from benzaldehyde as substrate).
bU/mL of cell-free extract prepared from a 3-ml LB culture in triplicate; wt, wild type; ND, not determined.
cSpecific activity (U/mg) of these selected mutant enzymes were properly compared in their purified form (plant codon) with that of the wild type in the Table II.

Another interesting result of the site-saturation mutagenesis of the His103 residue was the generation of very weakly active or inactive mutants altered at residue 103, e.g. with Tyr, Arg, Phe, etc. (Supplementary Figure S2 illustrates the whole-cell reactions of the plant and E. coli codon of MeHNL-His103). Further to characterize the increase in solubility, the soluble single mutant Lys176Pro, triple mutant Lys176Pro, Lys199Pro, Lys224Pro and buried His103Leu, His103Met, and inactive His103Tyr mutants were purified to homogeneity as well as the wild-type enzyme (data not shown). MeHNL-His103Tyr is an inactive mutant (Table I) and despite it showed no activity in E. coli JM109, weak activity was detected when the mutant was subcloned in pET28b vector and expressed in E. coli BL21 (DE3). The presence of strong promoter (T7) and also a short fusion sequence containing the His6-tag might be responsible for this finding. Therefore, it was possible to purify a form of this mutant. All the data presented in this paper were obtained from the purified enzymes, beside Table I compares the specific activity of almost 50 proteins with the wild type in the cell-free extract of E. coli to show robustness of the mutations.

Protein concentrations of the total cell-free extract, soluble and insoluble fractions were also measured in the mutants as well as the wild types with both the plant and E. coli optimized codons (Supplementary Table III shows the ratio of soluble and insoluble proteins to the total protein for the selected highly soluble His103 mutants as well as the wild type). Both 96-well transparent plate and regular (one by one using spectrometer at 595 nm) methods were used to measure the protein concentrations. The soluble fractions were used to calculate the specific activity of these enzymes, as presented in Table I. To have a reliable parameter, we calculated the ratios of soluble/total protein and insoluble/total proteins for each mutant. Although we have observed not so much clear differences for some mutants in comparison with the wild type, we could detect differences among the wild type and highly soluble mutants (Supplementary Table III).

Characterization of the mutants

After creation of these highly in vivo soluble mutants of the recombinantly expressed MeHNL in E. coli, purified forms of the proteins were used to characterize the enzymes in terms of temperature and pH stability, biochemical kinetics, CD, FT-IR and fluorescence spectra to monitor the probable structural changes. The characterized enzymes include single mutants with His103Leu and His103Met and triple Lys–Pro (at 176, 199 and 224) and single inactive mutant with His103Tyr as well as the wild-type MeHNL.

Temperature and pH stability profiles

The temperature stability of the wild-type MeHNL was compared with those of the four selected mutants. There were clear differences among these enzymes with wild type proving most stable. Highly soluble mutants His103Leu and His103Met exhibited similar temperature profiles to and were more stable
Protein solubility enhancement in E. coli by protein engineering

Table II. Comparison of kinetic parameters and specific activity of MeHNL and its mutants in the purified form*

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Parameter</th>
<th>MeHNL (wt, low in vivo soluble)</th>
<th>MeHNL-Lys176Pro (soluble)</th>
<th>MeHNL-Lys-Pro (highly soluble)</th>
<th>MeHNL-His103Leu (highly soluble)</th>
<th>MeHNL-His103Met (highly soluble)</th>
<th>MeHNL-His103Tyr (inactive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benzaldehyde</td>
<td>$V_{\text{max}}$ ($\mu$mol min$^{-1}$ mg$^{-1}$)</td>
<td>187</td>
<td>161</td>
<td>181</td>
<td>175</td>
<td>186</td>
<td>91.8</td>
</tr>
<tr>
<td></td>
<td>$K_{\text{m}}$ (mM)</td>
<td>23</td>
<td>21.6</td>
<td>27</td>
<td>13.4</td>
<td>27.9</td>
<td>12.3</td>
</tr>
<tr>
<td></td>
<td>$k_{\text{cat}}$ (s$^{-1}$)</td>
<td>96.6</td>
<td>83.2</td>
<td>93.5</td>
<td>90.4</td>
<td>96.1</td>
<td>47.4</td>
</tr>
<tr>
<td></td>
<td>$k_{\text{cat}}/K_{\text{m}}$ (mM$^{-1}$ s$^{-1}$)</td>
<td>4.2</td>
<td>3.85</td>
<td>3.46</td>
<td>6.75</td>
<td>3.44</td>
<td>3.85</td>
</tr>
<tr>
<td></td>
<td>Specific activity (U mg$^{-1}$) (mM substrate)</td>
<td>145 (75)</td>
<td>122 (50)</td>
<td>137 (50)</td>
<td>132 (45)</td>
<td>126 (50)</td>
<td>68.6 (30)</td>
</tr>
<tr>
<td>Mandelonitrile</td>
<td>$V_{\text{max}}$ ($\mu$mol min$^{-1}$ mg$^{-1}$)</td>
<td>66.8</td>
<td>80.9</td>
<td>76.8</td>
<td>41.6</td>
<td>46.2</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td>$K_{\text{m}}$ (mM)</td>
<td>5.17</td>
<td>6.98</td>
<td>5.54</td>
<td>2.55</td>
<td>1.86</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td>$k_{\text{cat}}$ (s$^{-1}$)</td>
<td>34.5</td>
<td>41.8</td>
<td>39.7</td>
<td>21.5</td>
<td>23.9</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td>$k_{\text{cat}}/K_{\text{m}}$ (mM$^{-1}$ s$^{-1}$)</td>
<td>6.67</td>
<td>6</td>
<td>7.2</td>
<td>8.43</td>
<td>12.8</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td>Specific activity (U mg$^{-1}$) (mM substrate)</td>
<td>43 (8)</td>
<td>46 (9)</td>
<td>55 (9)</td>
<td>31.6 (8)</td>
<td>43.8 (9)</td>
<td>ND</td>
</tr>
</tbody>
</table>

*A Hanes-Woolf plot was used for calculation of the kinetic parameters for the wild type and mutant MeHNLs in cyanohydrin synthesis and decomposition reactions. Substrate saturation curve was prepared by Michaelis-Menten curve for each purified enzyme.

The triple mutant MeHNL-Lys176Pro, Lys199Pro, Lys224Pro (surface substitutions).

Benzaldehyde as substrate in a cyanohydrin synthesis reaction, monitored by HPLC at 254 nm.

$k_{\text{cat}}/K_{\text{m}}$ ratio of the mutant to that of the wild type.

Mandelonitrile as substrate in a cyanohydrin cleavage reaction, monitored by spectrophotometer at 280 nm. ND, not determined.

![Fig. 4](https://academic.oup.com/peds/article-abstract/24/8/607/1508245/613)

than the triple mutant Lys176Pro, Lys199Pro and Lys224Pro (Supplementary Figure S3). Inactive mutant MeHNL-His103Tyr was the least stable of the proteins tested with standard deviations of 15–25% (compared with <5% for all other enzymes studied). There was no significant difference among the proteins in their stability at pH values 3–8, although beyond pH 8 the difference emerged among the wild type and mutants. Inactive mutant MeHNL-His103Tyr was the least stable enzyme at various pH values (data not shown).

**Kinetic experiments**

Table II demonstrates the results of kinetic studies on the enzymes with benzaldehyde and mandelonitrile as substrates of cyanohydrin synthesis and cleavage reactions, respectively. $V_{\text{max}}$ was the same as the wild type for the highly soluble mutants His103Leu, His103Met and triple Lys–Pro at positions 176, 199 and 224 while it was slightly diminished for highly soluble single mutant Lys176Pro and to almost half for the inactive mutant His103Tyr using benzaldehyde. Consequently, $k_{\text{cat}}$ showed the same pattern for the enzymes. Among the studied mutants, His103Leu and His103Tyr had smaller $K_{\text{m}}$ values for benzaldehyde (almost half) than did the wild type. When mandelonitrile was used as a substrate, the $V_{\text{max}}$ increased ~1.2-fold for the triple Lys–Pro mutant and decreased for His103Leu and His103Leu mutants. Finally, $k_{\text{cat}}/K_{\text{m}}$ (ratio of catalytic constant to the Michaelis constant) was enhanced for His103Met, His103Leu and triple Lys–Pro compared with the wild-type enzyme (1.1–1.9-fold), while it was decreased slightly for the single Lys167Pro mutant with mandelonitrile as substrate. Ratio of $k_{\text{cat}}/K_{\text{m}}$ of mutants to those of wild type clearly elevated for the His103Leu with both substrates.
CD, FT-IR, and Fluorescence spectra
The wild type and mutant enzymes showed slight differences in secondary structure in a denaturant-free solution (0 M Guanidium HCl) as illustrated in the Far-UV CD spectra (Fig. 4A). The α-helix content was monitored at 222 nm as described in the Materials and methods section. The highly soluble mutants with buried substitutions including His103Met and His103Leu showed spectra similar to the wild-type enzyme, although there were slight differences among them. The inactive mutant His103Tyr showed greater difference compared with the wild type. The highest change in mean residue ellipticity was observed for the highly soluble triple Lys176Pro, Lys199Pro Lys224Pro mutant, although this decrease in α-helix content did not apparently result in an increase in β-sheet (Supplementary Figure S4 shows FT-IR spectra of the enzymes). Denaturant-free CD spectra of these proteins at 37°C exhibited clearer differences among the proteins. The highly soluble mutants His103Leu and His103Met exhibited greater difference at 20°C (Supplementary Figure S5).

Fluorescence spectra of these proteins are depicted in Fig. 4B. The mutants His103Met, His103Tyr, His103Leu underwent a red-shift (331, 332 and 333 nm, respectively) compared with the wild type and the triple mutant Lys176Pro, Lys199Pro and Lys224Pro (both 329 nm). His103 mutants and multiple Lys–Pro also exhibited higher fluorescence intensity. The amide-I region of FT-IR spectra of the mutants was very similar to that of the wild-type enzyme, indicating no significant changes in β-sheet content after the mutations. All of the examined enzymes exhibited a peak at 1634/cm (Supplementary Figure S4).

Discussion
Directed evolution and rational design are two main strategies of protein engineering. Directed evolution is a powerful tool for achieving desirable properties of proteins and especially enzymes, although there are limited numbers of reports on the functional expression of proteins using this technique and mainly in host cells other than E. coli (Butler et al., 2003; Musidlovska-Persson and Bornscheuer, 2003) without detailed enzymatic characterizations.

Our first attempts failed to generate soluble MeHNL using techniques such as DNA shuffling and co-expression with molecular chaperones (data not shown). Then, we employed structure-guided mutagenesis to create mutants with increased resistance to the aldehyde and probably temperature, but these experiments interestingly resulted in highly active and soluble mutants. In another strategy, directed evolution was used to create highly active and in vivo soluble mutants of the enzyme in E. coli by screening in relatively small library of mutants. The enzyme in its wild type form was weakly expressed in E. coli, but its level of activity and in vivo solubility improved following the mutations. Indeed, the protein expression profile was reversed, with the main form of MeHNL present in the soluble fraction as a result of the mutations Lys176Pro, Lys199Pro and Lys224Pro alone and especially in combination (Fig. 3B, lanes 5–6), which reflected the production of the enzyme.

According to the general rule for the formation of a Schiff-base between lysine residues and aldehydes, we speculated that decreasing the number of surface lysine residues may enhance the resistance to aldehydes (Habeeb and Hiramoto, 1968; Matsumura et al., 1999). The 15 conserved lysine residues (Fig. 2A) were primarily considered as candidates for mutation since thought to be important to the structure-function of the enzyme. Then, five of these conserved solvent-exposed lysine residues, which were close to β-sheets, were subjected to site-saturation mutagenesis and then highly aldehyde- and temperature-resistant mutants were screened in the resulting library (data not shown). Although the aldehyde resistance or temperature stability did not increase following a decrease in the number of lysine residues, we found transformants with substitutions that changed a low functionally expressed enzyme into a highly active soluble form. These Lys residues play a crucial role in the in vivo solubility of this heterologously expressed eukaryotic enzyme in the prokaryotic host E. coli (Lys176, Lys199 and Lys224). A positively charged hydrophilic lysine residue with one of the highest negative hydrophobicity values (−4.6) was substituted with a small imino acid proline with low hydrophobicity (+0.3). These changes at the surface of the enzyme appeared to contribute extensively to the in vivo solubility of this recombinant protein in E. coli. Although it is thought that X to Pro substitutions stabilize proteins through destabilizing effects on unfolded states (Pro residue usually breaks or kinks α-helix), more studies are necessary to clarify the mechanism of solubility enhancement for these Lys–Pro mutants.

From the other side, random mutagenesis resulted in the His103 mutants, which was a residue buried close to the active site (Fig. 2C). His103 is the last residue of a β-sheet and located far from the substrate entrance cavity; therefore, it is thought that replacing it with Met, Val, Leu, Ile and Cys has no effect on the entrance. Substitution of this deeply buried His with Leu, Met, Val, Ile, Cys, Gln and Ser, which are expected to be more favorable for non-solvent-exposed parts of the enzyme, resulted in high activity and in vivo solubility in E. coli. The specific activity levels of the purified mutant enzymes were not high enough to justify the high activity exhibited in the cell-free extract. Increased catalytic efficiency ($k_{cat}/K_m$) was observed for His103Leu compared with the wild-type enzyme, although a slight decrease was observed for other highly soluble mutants for the substrate benzaldehyde. Three mutants had higher ratio of $k_{cat}/K_m$ including triple Lys–Pro at positions 176, 199, 224, His103Leu and His103Met when mandelonitrile was used (Table II). The improvement in catalytic efficiency alone cannot explain the increased activity and solubility of the mutant enzymes, although it was not improved for all highly in vivo soluble mutants. Briefly, we conclude that the enhanced activity and in vivo solubility are the results of mutations which led to less inclusion bodies being formed, i.e. increased solubility with larger numbers of active MeHNL molecules produced, where the newly synthesized enzymes overcome intrinsic and extrinsic factors preventing correct folding. This higher activity results in the higher specific activity in the cell-free extract. It is quite interesting that this phenomenon occurs at high local protein concentrations due to overexpression of the recombinant gene in the cytoplasm of E. coli at 37°C, but the expression level was not very different between the wild type and mutant MeHNL at 30°C (Fig. 3B). There is merit in the functional expression
of these genes at 37°C, as most proteins reported so far have been expressed in *E. coli* at lower temperatures e.g. 15–30°C.

The mutant enzymes His103Met, His103Leu and His103Tyr of *MeHNHL* exhibited increased florescence accompanying a red-shift (Fig. 4B), indicating changes in the environment around His103 as a result of these substitutions. The reason would likely be intra-molecular re-arrangements of the enzyme as a result of the mutations and the red-shift could be attributed to the subsequent minor 3D conformational changes.

All highly active mutants of *MeHNHL* including those with His103Leu, Ile, Val, Cys and Met have aliphatic hydrophobic residues substituted for the deeply buried and hydrophilic-basic residue His. They all have the highest hydrophacity in the hydropathy index table (Kyte and Doolittle, 1982). It is important in a protein that the hydrophobic amino acids occur within the structure, whereas the hydrophilic residues are more commonly found toward the surface. Therefore, the substitution of 1–4 carbon linear hydrophobic residues with the basic hydrophilic His103 in a non-solvent exposed area of the enzyme is consistent with the overall structure. We observed very low levels of activity and solubility for the His103Tyr, His103Trp, His103Phe and His103Pro mutants, which contain one or two ring structures and generate a rigid conformation following these mutations.

In a comparable study with the current research, Jenkins and colleagues improved the solubility of the catalytic domain (50–212) of a human immunodeficiency virus type 1 integrase [HIV-1 IN-(50–212)] through the systematic replacement of hydrophobic residues with Ala or Lys within the core domain. Phe185Lys was found to cause a remarkable increase, while all other 28 substitutions decreased the solubility of the enzyme (Jenkins et al., 1995). Wang et al. (2008) reported a single-site Cys57Met substitution for a human lectin Galectin-2 (hGal-2) expressed in *E. coli* (DE3). In comparison to the current study, the wild-type protein was more soluble and the mutant Cys57Met improved the solubility less than our results for the Lys→Pro (surface) and His103 (buried) substitutions.

A 2.5-fold increase was also achieved in the yield of a 169-amino acid human apolipoprotein D expressed in the periplasmic space of *E. coli*. Systematic mutagenesis directed at exposed hydrophobic side chains of this unglycosylated recombinant protein resulted in the triple mutant Trp99His, Ile118Ser and Leu129Ser (Nasreen et al., 2006). The mutant exhibited no significant changes in CD spectra compared with the wild type. Only a slight *in vivo* improvement was reported for a reverse transcriptase from leukemia virus following N-terminal truncation and a Leu435Lys mutation in the wild type and mutants, and detailed studies of the crystal structure will also provide valuable information. Investigations of these issues are presently underway in this laboratory.

### Supplementary data

Supplementary data are available at PEDS online.

### Acknowledgments

We thank Professor Teruhiko Beppu (Nihon University) for his valuable discussion on the propensities of the mutants. The authors appreciate Professor Yasuhiro Isogai, Toyama Prefectural University, for useful discussions on CD spectra. We appreciate Dr Praveen Kaul for critical reading of the manuscript and useful discussions. The authors also acknowledge Dr Shigeki Kawabata for helping in performance and plotting the fluorescence spectra.

### Funding

This work was supported in part by a Grant-in-Aid for scientific research B from Japan Society for the Promotion of Science to Y.A. (20380053). The financial support to M.D. by Ministry of Education, Culture, Sports, Science and Technology of Japan is deeply appreciated. This work also was supported in part by a grant from the Institute for Fermentation, Osaka, Japan.

### References


