Identification of Homophenylalanine Biosynthetic Genes from Cyanobacterium *Nostoc punctiforme* PCC73102 and Its Application to Microbial Production by *Escherichia coli*

Kento Koketsu#, Satoshi Mitsuhashi, Kazuhiko Tabata

Bioprocess Development Center, Kyowa Hakko Bio Co., Ltd., Tsukuba, Ibaraki, Japan

#Corresponding author: kento.koketsu@kyowa-kirin.co.jp
Abstract

L-Homophenylalanine (L-Hph) is a useful chiral building block for synthesis of several drugs, including angiotensin-converting enzyme inhibitors and the novel proteasome inhibitor carfilzomib. While the chemo-enzymatic route of synthesis is fully developed, we investigated microbial production of L-Hph to explore the possibility of a more efficient and sustainable approach to L-Hph production. We hypothesized that L-Hph is synthesized from L-Phe via a mechanism homologous to 3-methyl-2-oxobutanoic acid conversion to 4-methyl-2-oxopentanoic acid during leucine biosynthesis. Based on bioinformatic analysis, we found three putative homophenylalanine biosynthesis genes, \textit{hphA} (Npun_F2464), \textit{hphB} (Npun_F2457), and \textit{hphCD} (Npun_F2458), in the cyanobacterium \textit{Nostoc punctiforme} PCC73102, located around the gene cluster responsible for anabaenopeptin biosynthesis. We constructed \textit{Escherichia coli} strains harboring \textit{hphABCD}-expressing plasmids and achieved the fermentative production of L-Hph from L-Phe. To our knowledge, this is a first identification of the genes responsible for the homophenylalanine synthesis in any organisms. Furthermore, to improve the low conversion efficiency of the initial strain, we optimized the expression of \textit{hphA}, \textit{hphB}, and \textit{hphCD}, which increased the yield to ~630 mg/L. The L-Hph biosynthesis and L-Leu biosynthesis genes from \textit{E. coli} were also compared. This analysis revealed that HphB has comparatively relaxed substrate specificity and can perform the function of LeuB, but HphA and HphCD show a tight substrate specificity and cannot
complement the LeuA and LeuC/LeuD functions, and vice versa. Finally, the range of substrate tolerance of the L-Hph-producing strain was examined, which showed m-Fluorophenylalanine, o-Fluorophenylalanine, and L-Tyrosine were accepted as substrates, and that the corresponding homoamino acids were generated.

Introduction

L-Homophenylalanine (L-Hph) is a nonproteinogenic amino acid, and contains an integrated one-carbon-extended side chain compared to that of L-Phe. L-Hph is a useful chiral building block for synthesis of several pharmaceutical drugs, such as angiotensin-converting enzyme (ACE) inhibitors, and the novel proteasome inhibitor carfilzomib, which is used for treating multiple myeloma (1, 2). The majority of ACE inhibitors, including enalapril, lisinopril, and benazepril, contain a common L-Hph core in their structure. Furthermore, carfilzomib, which was recently approved by the United States Food and Drug Administration, is a synthetic tetrapeptide epoxyketone that contains an L-Hph residue. Because of the importance of L-Hph as a pharmaceutical ingredient, several processes for industrial scale production of L-Hph have been developed (3). These include transamination from 2-oxo-phenylbutanoate using transaminase in a stereo-selective manner, and optical resolution of the hydantoin derivative by hydantoinase, followed by hydrolysis of the carbamoyl moiety, catalyzed by carbamoylase (4, 5, 6). Although chemo-enzymatic methods seem to be quite
efficient processes to supply L-Hph for pharmaceutical drug synthesis, production cost depends on material derived from fossil fuel. To explore more efficient and sustainable production of L-Hph, we investigated microbial production of L-Hph from L-Phe and simple carbon and nitrogen sources such as glucose and ammonia.

In nature, some secondary metabolites contain aromatic homoamino acids, for example benzyl glucosinolate from mustard oil (7) and non-ribosomally synthesized cyclic peptides from cyanobacteria (8). Biosynthetic studies on benzyl glucosinolate using radiolabeled precursors indicated that Hph exists as an intermediate during benzyl glucosinolate biosynthesis, although the genes responsible for glucosinolate synthesis have not been identified (9). Recently, the gene cluster for a protease inhibitor depsipeptide, anabaenopeptin, containing L-Hph residues, produced by the cyanobacterium *Anabaena sp.* strain 90, was discovered (10). However, previous studies have focused on these pathways in respect to non-ribosomal peptide synthetases (11), and have no mention of the homoamino acid biosynthetic genes.

In this study, we focused on the analogy of the biosynthetic pathway between L-Hph and L-Leu in regard to the side-chain carbon elongation mechanism. Based on this speculation, we searched the homologous genes of the leucine biosynthesis in the anabaenopeptin gene cluster of the cyanobacterium *Nostoc punctiforme* PCC73102 and found three putative genes responsible for L-Hph biosynthesis. We constructed *Escherichia coli* strains expressing these
genes and successfully demonstrated the microbial production of L-Hph. Next, we performed complementary comparisons between L-Hph biosynthesis genes and the L-Leu biosynthetic genes, LeuA, LeuB, and LeuC/D. Additionally, we determined substrate specificity, which reveals that the L-Hph-producing strain can also produce fluorinated Hph analogs and homotyrosine (Hty). Finally, we suggest that microbial production of L-Hph using genetically engineered *Escherichia coli* is a viable alternative process to existing chemo-enzymatic syntheses.

 Materials and Methods

**Chemicals.** L-Homophenylalanine was purchased from Tokyo Chemical Industry (Tokyo, Japan). Phenylalanine, tyrosine, and other amino acid derivatives were obtained from Wako Pure Chemical Industries (Osaka, Japan).

**Cloning of genes and construction of L-Hph-producing strains.** The L-Hph biosynthesis genes *hphA* (*Npun_F2464*), *hphB* (*Npun_F2457*), and *hphCD* (*Npun_F2458*) were amplified from the genome of *Nostoc punctiforme* PCC73102 using PrimeSTAR MAX polymerase (Takara Bio, Otsu, Japan) according to the manufacturer’s instructions, with the oligonucleotide primer pairs given in Table 1. The amplified *hphCD* fragment was digested with *Eco*RI and *Kpn*I and cloned into pTrc99a (Invitrogen, Carlsbad, CA, USA), which had been digested with the corresponding restriction enzymes, and subsequently dephosphorylated.
with shrimp alkaline phosphatase (Takara Bio), to obtain pTrc99a-Ptrc-hphCD named as pHPH01. The PCR fragment containing hphA was digested with KpnI and BamHI and introduced into pHPH01 digested with KpnI and BamHI, to construct pTrc99a-Ptrc-hphCD-hphA (pPH02). The PCR fragment containing hphB was digested with BamHI and PstI and cloned into pSTV29 (Takara Bio) digested with BamHI and PstI, to provide pSTV29-Plac-hphB (pPH03). E. coli strain W3110 (12) was transformed with pHPH02 and pPH03 to generate the L-Hph-producing strain system 1. To improve Hph-productivity, other production systems were also constructed using the oligonucleotide primers as listed in Table 1, according to a procedure similar to that described above (Table 2).

**Production of L-Hph.** Single colonies from L-Hph-producing strains were inoculated into 5 mL Luria-Bertani (LB) medium containing 100 mg/L ampicillin and 25 mg/L chloramphenicol and incubated at 30°C overnight. An inoculum of the resulting overnight culture (100 μL) was transferred into production medium (5 mL) composed of 20 g/L glucose, 2 g/L MgSO₄, 16 g/L KH₂PO₄, 14 g/L K₂HPO₄, 2 g/L NH₄SO₄, 1 g/L citric acid, 5 g/L casamino acids (Difco, Franklin Lakes, NJ, USA), 50 mg/L FeSO₄.7H₂O, 10 mg/L thiamine-HCl, and 10 mg/L MnSO₄.5H₂O (pH 7.2). L-Phe was added to the medium at a final concentration of 1 g/L (6.1 mM). Following 6 hours of cultivation at 30°C, cultures were induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and further cultivated at
30°C for 30 or 48 hours. L-Hph was produced from glucose and ammonia by direct fermentation, using the L-Hph-producing strain with L-Phe-production plasmid, *Ptrp-phe*<sup>Δ</sup>*-aroG(P150L)-pMW219* (named as pMWF01), the genes of which are amplified from the genome of *E. coli* W3110 using the oligonucleotide primers listed in Table 1. Culture conditions were the same as described above, without the addition of L-Phe (13, 14). For Hph analog production, Phe analogs were added to the culture and cultivated as above.

**Product analysis.** Culture supernatants were diluted with water and analyzed by high-performance liquid chromatography (HPLC) on a LaChrom Elite HPLC system (Hitachi High-Technologies, Tokyo, Japan) equipped with an Inertsil ODS-3 column (GL sciences, 4.6 × 250 mm, 5 µm), at a flow rate of 1.0 mL/min at 50°C, with spectrophotometric detection of substrates and products at 220 nm. Solvent gradients were as follows: L-Hph and fluorinated derivatives: 0–15 min, 25–35% MeOH/0.1% TFA; 15–16 min, 35–90% MeOH/0.1% TFA; 16–18 min, 90% MeOH/0.1% TFA; 18–19 min 90%–25% MeOH/0.1% TFA in H<sub>2</sub>O/0.1% TFA; homotyrosine: 0–15 min, 5–15% MeOH/0.1% TFA; 15–16 min, 15–90% MeOH/0.1% TFA; 16–18 min, 90% MeOH/0.1% TFA; 18–19 min 90%–5% MeOH/0.1% TFA in H<sub>2</sub>O/0.1% TFA. Mass analyses were carried out using a 3200 QTRAP LC/MS/MS system (AB Sciex, Massachusetts, U. S.).

**Complementation experiment between hph<sub>ABCD</sub> and leu<sub>ABCD</sub>.** Complementation of *leuABCD* function by *hphABCD* was examined by the following procedure. Strains BW25113
ΔleuA, BW25113 ΔleuB, BW25113 ΔleuC, and BW25113 ΔleuD were obtained from the KEIO library (15). pTrc99a plasmids harboring each hph gene were introduced into the corresponding leu gene-deletion mutant as ΔleuA-hphA, ΔleuB-hphB, ΔleuC-hphCD, and ΔleuD-hphCD (Table 3A). The resulting strains were cultured in M9 medium (16) at 30°C for 24 hours and growth was checked by OD₆₆₀ measurement using a UV-1800 spectrophotometer (Shimadzu, Kyoto, Japan). Complementation of hphABCD function by leuABCD from E. coli W3110 was examined as follows. Several plasmid systems that contained each hph gene exchanged with the corresponding leu gene were constructed to examine L-Hph production (Table 3B). Production and analysis were performed as described above.

Results

Discovery of L-Hph biosynthesis genes in Nostoc punctiforme PCC73102. We hypothesized that if aromatic homoamino acids are synthesized via a one-carbon extension mechanism, similar to L-leucine biosynthesis from 3-methyl-2-oxobutanoic acid to 4-methyl-2-oxopentanoic acid (17), the homologous genes encoding isopropylmalate synthase (LeuA), 3-isopropylmalate dehydrogenase (LeuB), and 3-isopropylmalate isomerase (LeuC, LeuD) might exist within or around the anabaenopeptin biosynthetic gene cluster of the cyanobacterium Anabaena sp. strain 90 (10). But, we could not find the expected genes
because the sequence information of the gene cluster was thought to be not completed at this time. Instead, we searched the anabaenopeptin NZ 857 (nostamide) biosynthesis gene cluster from *N. punctiforme* PCC73102, the genome of which has been completely determined (GenBank accession number AAAY02000000) (18). Subsequently, we found three putative homophenylalanine biosynthetic genes, Npun_F2464 (gene for benzylmalate synthase), Npun_F2457 (gene for 3-benzylmalate dehydrogenase), and Npun_F2458 (gene for 3-benzylmalate isomerase), named *hphA*, *hphB*, and *hphCD*, respectively, located around the anabaenopeptin gene cluster (Fig. 1A). Previously, Npun_F2464 was annotated as *aptE*, which is located downstream of a gene encoding non-ribosomal peptide synthetases. Npun_F2457 and Npun_F2458 are adjacent to the cluster, and annotated as encoding 3-isopropylmalate dehydrogenase and 3-isopropylmalate isomerase, respectively. Although *hphB* and *hphCD* could not be found in the anabaenopeptin gene cluster from *Anabaena sp.* strain 90 at this time, we used the BLAST server (19) to confirm the presence of the homologous genes in the anabaenopeptin gene cluster of *Nodularia spumigena* CCY9414, and in the genomes of other cyanobacterial species *Synechococcus* sp. PCC 7335 and *Oscillatoria* sp. PCC 6506, from which secondary metabolite-containing homoamino acid residues have not been isolated (Table 4A-C). We compared the amino acid sequences of HphA, HphB, and HphCD from *N. punctiforme* to those of LeuA, LeuB, and LeuC/D from *E. coli*, respectively, using ClustalW (20). Alignment of HphA and LeuA indicated that HphA
does not have a regulatory domain, which is located in the C-terminal region of bacterial
LeuA, and contains a leucine binding site (21). HphB shows 42% sequence identity with E.
coli LeuB. HphCD appears to be a fused structure composed of a large and a small subunit,
whereas bacterial isopropylmalate isomerase is a heterodimer composed of a large subunit,
LeuC, and a small subunit, LeuD (22). Based on these analyses, we hypothesized the L-Hph
biosynthetic pathway from L-Phe as described in Figure 1B.

To demonstrate that these gene products synthesize L-Hph, we designed a system to
produce L-Hph in a genetically engineered E. coli strain by expressing hphA, hphB, and
hphCD. First, two plasmids, pHPH02 and pHPH03, were constructed and introduced into E.
coli W3110. The resulting strain was cultivated in production medium containing 1 g/L L-Phe
at 30°C. After 48 hours of incubation, the culture supernatant was analyzed by HPLC. As
shown in Fig. 2, only the culture supernatant obtained from the strain containing all hph genes
exhibited a peak that coincides with the peak of the L-Hph standard. Furthermore, LC/MS/MS
analysis indicated that the molecular mass of this peak ([M+H]^+ = 180.2) is identical to the
theoretical mass of L-Hph ([M+H]^+ = 180.1). The 14 unit mass increase compared to the
molecular weight of L-Phe corresponds to the mass of a methylene carbon. Additionally,
fragment ions at 134.2 and 117.2 clearly showed the presence of carboxyl and amino moieties,
respectively. Thus, we demonstrated that HphA, HphB, and HphCD are responsible for the
biosynthesis of L-Hph in N. punctiforme PCC73102. It is also noteworthy that unknown peak
was detected in the culture broth, and also generated in the culture broth of the strain harboring the *hphA* gene only. LCMS/MS analysis of the peak clearly showed that this unknown compound is 2-benzylmalic acid (2-BMA), theoretical mass; ([M-H]$^-$ = 223.1), observed mass; ([M-H]$^-$ = 223.1), which is an intermediate produced by HphA (Fig. 1B).

**Optimization of gene expression to increase L-Hph productivity.** Although we identified the L-Hph biosynthesis genes, the efficiency of L-Hph production was relatively low (~154 mg/L at 48 hours cultivation) at 1 g/L L-Phe and 30°C, which we considered insufficient for industrial scale production of L-Hph. It has been known that the optimal balance of the gene expression in the host cell resulting in the increase of the product yield, and several methods have been developed to fix the metabolic flow (23). Thus, we attempted the simple method to optimize expression of *hphA*, *hphB*, and *hphCD* to increase the yield of L-Hph. And, five new two-plasmid expression systems, with different arrays of genes, were constructed (Table 1, 2) and their L-Hph production was examined. As shown in Fig. 3, optimization dramatically increased L-Hph yields, especially of Systems 5 and 6, which had yields of 630 mg/L and 560 mg/L, respectively, at 48 hours cultivation. We also used these systems to investigate the influence of temperature on L-Hph production. Fermentation was carried out at 25, 30, and 35°C, and HPLC analyses showed yields of 900 mg/L, 260 mg/L, and 0 mg/L, respectively, with 1 g/L L-Phe. At 35°C, L-Hph production was completely abolished, while 2-benzylmalic acid (2-BMA), which is a product of the HphA reaction, accumulated in the culture broth.
This may indicate that either HphB or HphCD, or both, is thermosensitive.

Next, we examined production of L-Hph from glucose and ammonia, without the addition of L-Phe, by combining the HphABCD-expressing plasmids pH11 and pH12, with a L-Phe-producing strain as a host. This L-Phe-producing strain was constructed by introducing pMWF01, which contains the genes encoding feedback resistant mutants of PheA and AroG, into *E. coli* W3110 (13, 14). The resulting strain produced a similar level of L-Hph (396 mg/L at 30 hours) from glucose and ammonia to the L-Phe-feeding culture (395 mg/L at 30 hours).

**Complementation experiment between hphABCD and leuABCD.** Based on their respective reaction schemes, the catalytic properties of HphA, HphB, and HphCD are expected to be similar to LeuA, LeuB, and LeuC/LeuD respectively. To evaluate if these genes can compensate for each other’s function, we generated several strains with the corresponding *hph* and *leu* genes exchanged, as listed in Table 3A and 3B. First, we examined if *leu* gene-deficient mutants supplemented with the corresponding *hph*-gene could grow in M9 medium (Fig. 4A). The *leuB*-deletion mutant complemented with *hphB* grew as efficiently as the wild-type strain, whereas the Δ*leuA-hphA*, Δ*leuC-hphCD*, and Δ*leuD-hphCD* strains did not grow. Second, we determined whether L-Hph-producing cells, in which one of the *hph* genes was complemented with the corresponding *leu*-gene, could produce L-Hph (Fig. 4B). The strain in which *hphB* was exchanged with *leuB* produced almost identical amounts of
L-Hph as the non-exchanged strain. However, the other strains did not produce a detectable amount of L-Hph. These results indicate that HphB has relatively relaxed substrate specificity and can perform the functions of LeuB, whereas HphA and HphCD have tight substrate specificity and cannot complement LeuA and LeuC/LeuD function, and vice versa.

**Substrate specificity of the L-Hph-producing system.** Finally, we explored the range of substrate specificity for homoamino acid synthesis by engineered *E. coli* with *hph* genes. L-Phe analogs, including L-Tyr, DL-\textit{m}-Fluoro-Phe, DL-\textit{o}-Fluoro-Phe, β-Methyl-Phe, L-Phenylglycine, L-p-Iodo-Phe, L-p-Nitro-Phe, L-Trp, were added to the culture medium of the L-Hph-producing strain (Fig. 5). Following cultivation for 48 hours, the culture supernatants were analyzed with HPLC and LC-MS/MS. DL-\textit{m}-Fluoro-Phe, DL-\textit{o}-Fluoro-Phe, and L-Tyr were successfully converted into their corresponding homoamino acid, whereas β-Methyl-Phe, L-Phenylglycine, L-p-Iodo-Phe, L-p-Nitro-Phe and L-Trp were not. The identities of the homoamino acids of DL-\textit{m}-Fluoro-Phe (theoretical [M+H]+ = 198.1, experimental [M+H]+ = 198.1), DL-\textit{o}-Fluoro-Phe (theoretical [M+H]+ = 198.1, experimental [M+H]+ = 198.1), and L-Tyr (theoretical [M+H]+ = 196.1, experimental [M+H]+ = 196.1) were verified by MS analysis. The 14 mass unit increase of the homoamino acids compared to their amino acid substrates clearly indicated extension of the amino acid side-chain by a methylene carbon. In the case of β-Methyl-Phe, L-Phenylglycine, L-p-Iodo-Phe, L-p-Nitro-Phe and L-Trp, the corresponding homoamino acids could not generated and it is
supposed that HphA, which is first enzyme for the Hph biosynthesis, could not accept these substrates. This is supported by the fact that both substrate consumption and HphA-product generation could not be observed, although we did not determine whether the host amino acid aminotransferases (24, 25) could convert these amino acids to their corresponding keto acids.

Discussion

In this study, we identified L-Hph biosynthesis genes *hphA*, *hphB*, and *hphCD* in *N. punctiforme* PCC73102, and demonstrated microbial production of L-Hph by genetically engineered *E. coli* strains employing *hphA*, *hphB*, and *hphCD* expression (Fig.1, Fig.2). Genome projects for several cyanobacterial species are ongoing, and among the draft genome sequences, the homologous genes for *hphA*, *hphB*, and *hphCD* can be found in the genomes of *Nodularia spumigena* CCY9414 (GenBank accession number: AAVW00000000), *Synechococcus* sp. PCC 7335 (GenBank accession number: ABRV00000000), and *Oscillatoria* sp. PCC 6506 (GenBank accession number: NZ_CACA00000000) (26) (Table 4A-C). *N. spumigena* CCY9414 produces the cyclic peptides nodulapeptin B and C, which contain both Hph and Hty residues in their structures (10). While *Oscillatoria* sp. PCC 6506 and *Synechococcus* sp. PCC 7335 are not known as cyclic peptide-producers, we propose that these organisms possess a capacity for producing non-ribosomally synthesized peptides involving aromatic homoamino acid residues, which must be supplied by *hphABCD*-like
genes. Interestingly, the cyclic peptides nostamide, from *Nostoc punctiforme* PCC73102, and nodulapeptins and the linear nonribosomal peptide spumidins (27), from *Nodularia spumigena* CCY9414, contain both Hph and Hty residues despite the fact that these organisms possess only one set of *hphABCD* genes. Our finding that Hph-producing strains could also generate Hty from Tyr clearly indicates that the cyanobacterial aromatic homoamino acid biosynthesis enzymes have dual substrate specificity, accepting both phenylpyruvic acid and 4-hydroxyphenyl pyruvic acid, and supply L-Hph and L-Hty as building blocks for non-ribosomal peptide synthesis. Apart from homoamino acid biosynthesis in cyanobacteria, there is some evidence that other plant and fungal organisms can also produce aromatic homoamino acids. Previous studies of benzyl glucosinolate biosynthesis using feeding experiments clearly indicated the presence of genes for Hph production in some specific plants, although the genes have not been isolated. Most recently, Cacho and co-workers identified the gene cluster for synthesis of Hty, which is a precursor of antifungal cyclic peptide echinochandins from fungus *Emericella rugulosa* NRRL 11440 (28). The Hty gene cluster is composed of isopropyl malate synthase (*htyA*), transaminase (*htyB*), isopropyl malate dehydrogenase (*htyC*), and aconitase (*htyD*), along with two oxygenases (*htyF, htyE*) for further modification of Hty, although the latter two are not involved in the heterologous production of Hty. Considering these bioinformatic analyses and the existence of various secondary metabolites containing aromatic homoamino acids, the gene sets responsible for
aromatic homoamino acid biosynthesis appear to be distributed across a wide variety of organisms, such as fungi and plant, in addition to cyanobacteria. Interestingly, cyclic peptides containing dihomotyrosine, named largamides, have been isolated from the cyanobacterium *Oscillatoria* sp. (29). Although genes involved in the synthesis of dihomotyrosine have not been identified, it is possible that an *hph*-like gene is involved, which should catalyze two rounds of methylene carbon extension in a similar manner to dihomomethionine synthesis catalyzed by *leu*-like genes from the plant *Arabidopsis thaliana* (30, 31). Discovery of largamide-producing enzymes and comparison of the crystal structures of enzymes for homoamino acid and dihomoamino acid synthesis will provide a good opportunity to study control of side chain-elongation on homoamino acid, and allow rational protein engineering to generate novel enzymes that synthesize unnatural homoamino acids, like the method described previously (32).

The remarkable accumulation of 2-BMA during fermentation of the L-Hph-producing strain implied that the HphA-catalyzed reaction is not inhibited by L-Hph, which is consistent with the absence of a C-terminal regulatory domain in HphA. Although it is possible that HphB or HphCD plays a key role in regulating this pathway, the high level of L-Hph production may indicate that this pathway does not contain critical regulatory machinery. The relatively high sequence identity of HphB and *E. coli* LeuB (42%) is consistent with the finding that HphB can function in place of LeuB, and vice versa (Figs. 4).
HphCD is a fused structure composed of a large and a small subunit, as opposed to the bacterial 3-isopropylmalate isomerase, which is a heterodimer of a large subunit, LeuC, and a small subunit, LeuD. We could not determine whether the fusion of the two subunits has any particular role in HphCD. Like LeuC/D, HphCD requires an iron-sulfur cluster in its catalytic center for activity (33). We initially suspected that the existing iron-sulfur machinery of *E. coli*, such as the Icu or Suf pathways (34), could not deliver the iron-sulfur cluster to the cyanobacterial enzyme because of the structural differences between the HphCD and LeuC/LeuD complexes. However, L-Hph production by the engineered *E. coli* clearly indicates that the *E. coli* iron-sulfur machinery is sufficient to provide the iron-sulfur cluster for apo-HphCD, although we were not able to estimate the efficiency.

To improve the low yield of L-Hph produced by the initial strain, we constructed five strains harboring the plasmid set with altered gene arrangements. This successfully elevated the yield from 154 mg/L to 630 mg/L (Fig. 3). In a recent report on taxol precursor overproduction in *E. coli*, optimization of the isoprenoid pathway was critical for both efficient production of the terpenoid, and for decreasing formation of the byproduct, indole (35). Like the optimization of the taxol pathway, the optimization of the Hph pathway might lead to increase the L-Hph production. SDS-PAGE analysis of enzyme expression indicated that lower expression of HphCD may improve yields, because higher expression of HphCD reduced growth (data not shown). We also carried out direct fermentative production
of L-Hph from glucose and ammonia by introducing the plasmid (pMWF01) harboring feedback resistant mutants of *pheA* and *aroG*, which are key enzymes for L-Phe biosynthesis. The constructed strain provided almost the same amount of L-Hph as fermentative production with feeding of 1 g/L L-Phe. Thus, this technique may represent an alternative method for direct fermentative production of L-Hph.

Additionally, we investigated the substrate specificity of this system and revealed that fluorinated Phe analogs and L-Tyr were also accepted as substrates, whereas analogs substituted with larger functional groups in the aromatic ring were not (Fig. 5). D-amino acids also could not be converted into homoamino acids, as half of the amount of substrate remained when racemic amino acids were used as a substrate. However, it is possible that the D-form could be utilized with the combination of the specific D-amino acid-oxidase or -racemase (36), leading to the efficient generation of various L-homoamino acid analogs from racemic materials.

In conclusion, we identified the L-Hph biosynthesis genes in *N. punctiforme* PCC73102 and applied them for microbial production of L-Hph in *E. coli*. To our knowledge, this is a first identification of the genes responsible for the homophenylalanine synthesis in any organisms. We also demonstrated that several aromatic homoamino acid analogs, including L-homotyrosine, could be produced using this system, even though the range of substrate specificity is moderately limited. We propose that various fine chemical materials,
such as keto acids, hydroxy acids, and alcohols, derived from aromatic homoamino acids can be produced by combining the L-Hph-producing strain with various enzymes that have been developed for industrial production of useful chiral building blocks (37). Additionally, detailed structure-based comparison between enzymes for producing L-Hph and other homoamino acids will allow us to generate novel homoamino acid producing systems through protein engineering in the future.

Acknowledgments

We thank Dr. Satoshi Koizumi for revision of the manuscript and useful advice, and Nobuo Yokota and Kazuki Ishikawa for LC-MS analysis of homoamino acids. We also thank laboratory members for their advice.
REFERENCES


25. Powell JT, Morrison JF. 1978. The purification and properties of the aspartate aminotransferase and aromatic-amino-acid aminotransferase from \( Escherichia \) \( coli \). Eur.


**TABLE 1** Oligonucleotides used to construct plasmids for expression of hph genes and pheA, aroG genes

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>Forward Sequence</th>
<th>Reverse Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH01: pTrc99a-Ptrc-hphCD</td>
<td>hphCD_Fw_EcoRI 5’-ATAGAATTCAGAGAAATAACGGTTTTG-3’</td>
<td>hphCD_Rv_KpnI 5’-AGAGGATCCATGAAATTATACGCTAATTTTTC-3’</td>
</tr>
<tr>
<td>pH02: pTrc99a-Ptrc-hphCD-hphA (derived from pH01)</td>
<td>hphA_Fw_KpnI 5’-AGAGGATCCATGAAATTATACGCTAATTTTTC-3’</td>
<td>hphA_Rv_BamHI 5’-AGAGGATCCATGAAATTATACGCTAATTTTTC-3’</td>
</tr>
<tr>
<td>pH03: pSTV29-Plac-hphB</td>
<td>hphB_Fw_BamHI 5’-AGAGGATCCATGAAATTATACGCTAATTTTTC-3’</td>
<td>hphB_Rv_PstI 5’-AGAGGATCCATGAAATTATACGCTAATTTTTC-3’</td>
</tr>
<tr>
<td>pH04: pSTV29-Ptrc-hphB (Ptrc-hphB fragment was amplified from pH08)</td>
<td>Ptrc_Fw_SacI 5’-ACAGAGCTCAGAATTACGCTAATTTTTC-3’</td>
<td>hphB_Rv_PstI 5’-AGAGGATCCATGAAATTATACGCTAATTTTTC-3’</td>
</tr>
<tr>
<td>pH05: pTrc99a-Ptrc-hphA</td>
<td>hpha_A_Fw_KpnI 5’-ACAGAGCTCAGAATTACGCTAATTTTTC-3’</td>
<td>hphA_Rv_BamHI 5’-ACAGAGCTCAGAATTACGCTAATTTTTC-3’</td>
</tr>
<tr>
<td>pH06: pTrc99a-Ptrc-hphA-hphB (derived from pH05)</td>
<td>hphB_Fw_NcoI 5’-ACAGAGCTCAGAATTACGCTAATTTTTC-3’</td>
<td>hphB_Rv_EcoRI 5’-ACAGAGCTCAGAATTACGCTAATTTTTC-3’</td>
</tr>
<tr>
<td>pH07: pSTV29-Ptrc-hphCD (Ptrc-hphCD fragment was amplified from pH01)</td>
<td>Ptrc_Fw_SacI 5’-ACAGAGCTCAGAATTACGCTAATTTTTC-3’</td>
<td>hphCD_Rv_KpnI 5’-ACAGAGCTCAGAATTACGCTAATTTTTC-3’</td>
</tr>
<tr>
<td>pH08: pTrc99a-Ptrc-hphB</td>
<td>hphB_Fw_NcoI 5’-ACAGAGCTCAGAATTACGCTAATTTTTC-3’</td>
<td>hphB_Rv_EcoRI 5’-ACAGAGCTCAGAATTACGCTAATTTTTC-3’</td>
</tr>
<tr>
<td>pH09: pTrc99a-Ptrc-hphB-hphA (derived from pH08)</td>
<td>hpha_A_Fw_KpnI 5’-ACAGAGCTCAGAATTACGCTAATTTTTC-3’</td>
<td>hpha_Rv_BamHI 5’-ACAGAGCTCAGAATTACGCTAATTTTTC-3’</td>
</tr>
<tr>
<td>pH10: pTrc99a-Ptrc-hphA-hphCD (derived from pH05)</td>
<td>hphCD_Fw_BamHI 5’-ACAGAGCTCAGAATTACGCTAATTTTTC-3’</td>
<td>hphCD_Rv_HindIII 5’-ACAGAGCTCAGAATTACGCTAATTTTTC-3’</td>
</tr>
</tbody>
</table>
pHPH11; pTrc99a-Ptrc-hphB-hphCD (derived from pHPH08)

hphCD_Fw_EcoRI  5’-ATA GAATTCAGTCACGAAAATAACGTTTTG-3’

hphCD_Rv_KpnI  5’-AGAGGTACCTTTATTAATCAAAGCGATCGCTAATTTTTC-3’

pHPH12; pSTV29-Ptrc-hphA (Ptrc-hphA fragment was amplified from pHPH05)

Ptrc_Fw_SacI  5’-ACAAGCTCGCATAATTCGTGCTGACTGAAATACGTTTTG-3’

hphA_Rv_BamHI  5’-AGAGGTACCGATGAAATTATTTAAACACCCCATTTTTCC-3’

pMWF01; pMW219-Ptrp-pheAfbr-aroG (Ptrp-pheAfbr-aroG(P150L))

pTrp_Fw_EcoRI  5’-ATCGGAATTCTCAAGGTATATCATCAGCTTTT-3’

pTrp_Rv_BamHI  5’-ATCGGAATTCTCAAGGTATATCATCAGCTTTT-3’

pheA_Fw_BamHI  5’-ATCGGAATTCTCAAGGTATATCATCAGCTTTT-3’

pheA330_Rv_SalI  5’-ATCGGAATTCTCAAGGTATATCATCAGCTTTT-3’

aroG_Fw_SalI  5’-ATCGGAATTCTCAAGGTATATCATCAGCTTTT-3’

aroG_Rv_SphI  5’-ATCGGAATTCTCAAGGTATATCATCAGCTTTT-3’

aroG_P150L_1  5’-CTTGGCGATGCACTCAGCTTTT-3’

aroG_P150L_2  5’-CTTGGCGATGCACTCAGCTTTT-3’

*pItalics indicate the sites digested by restriction enzymes. Lower cases indicate the mutational site.

**TABLE 2** L-Hph-producing systems

<table>
<thead>
<tr>
<th>System</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>E. coli</em> W3110/pHPH02, pHPH03</td>
</tr>
<tr>
<td>2</td>
<td><em>E. coli</em> W3110/pHPH02, pHPH04</td>
</tr>
<tr>
<td>3</td>
<td><em>E. coli</em> W3110/pHPH06, pHPH07</td>
</tr>
<tr>
<td>4</td>
<td><em>E. coli</em> W3110/pHPH09, pHPH07</td>
</tr>
<tr>
<td>5</td>
<td><em>E. coli</em> W3110/pHPH10, pHPH04</td>
</tr>
<tr>
<td>6</td>
<td><em>E. coli</em> W3110/pHPH11, pHPH12</td>
</tr>
</tbody>
</table>

**TABLE 3 (A)** Strains constructed to evaluate cell growth by exchanging *leu* genes with corresponding *hph* genes (B) Strains constructed to examine L-Hph production by removing *hph* or replacing it with a corresponding *leu* gene

<table>
<thead>
<tr>
<th>Experiment A</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>wild</td>
<td><em>E. coli</em> BW25113</td>
<td></td>
</tr>
<tr>
<td>ΔleuA</td>
<td><em>E. coli</em> BW25113 <em>leuA::Km</em></td>
<td></td>
</tr>
<tr>
<td>ΔleuA - hphA</td>
<td><em>E. coli</em> BW25113 <em>leuA::Km/hph</em></td>
<td></td>
</tr>
</tbody>
</table>

28
ΔleuB  E. coli BW25113 leuB::Km<sup>+</sup>
ΔleuB - hphB  E. coli BW25113 leuB::Km<sup>+</sup>pHPH08
ΔleuC  E. coli BW25113 leuC::Km<sup>+</sup>
ΔleuC - hphCD  E. coli BW25113 leuC::Km<sup>+</sup>pHPH01
ΔleuD  E. coli BW25113 leuD::Km<sup>+</sup>
ΔleuD - hphCD  E. coli BW25113 leuD::Km<sup>+</sup>pHPH01

Experiment B

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>hphA</td>
<td>E. coli W3110/pHPH05</td>
</tr>
<tr>
<td>hphD-hphB</td>
<td>E. coli W3110/pHPH11</td>
</tr>
<tr>
<td>hphA-hphCD</td>
<td>E. coli W3110/pHPH10</td>
</tr>
<tr>
<td>hphA-hphB</td>
<td>E. coli W3110/pHPH06</td>
</tr>
<tr>
<td>leuA-hphCD-hphB</td>
<td>E. coli W3110/pHPH11, pSTV29-Ptrc-leuA</td>
</tr>
<tr>
<td>hphA-hphCD-leuB</td>
<td>E. coli W3110/pHPH10, pSTV29-Ptrc-leuB</td>
</tr>
<tr>
<td>hphA-leuC/D-hphB</td>
<td>E. coli W3110/pHPH06, pSTV29-Ptrc-leuC/D</td>
</tr>
<tr>
<td>hphA-hphCD-hphB</td>
<td>E. coli W3110/pHPH10, pHPH04</td>
</tr>
</tbody>
</table>

### TABLE 4 (A) HphA and its homologues found in various cyanobacteria, and *E. coli* LeuA.

### (B) HphB and its homologues found in various cyanobacteria, and *E. coli* LeuB. (C) HphCD and its homologues found in various cyanobacteria, and *E. coli* LeuC/D.

#### (A) HphA and its homologues

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Organisms</th>
<th>Length (aa)</th>
<th>Homology, Similarity</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>HphA</td>
<td><em>Nostoc punctiforme</em> PCC73102</td>
<td>392</td>
<td>100%</td>
<td>YP_001865968</td>
</tr>
<tr>
<td>putative HphA</td>
<td><em>Anabaena</em> sp. 90</td>
<td>392</td>
<td>82%, 92%</td>
<td>ACZ55947</td>
</tr>
<tr>
<td>putative HphA</td>
<td><em>Nodularia spumigena</em> CCY9414</td>
<td>392</td>
<td>80%, 90%</td>
<td>ZP_01629954</td>
</tr>
<tr>
<td>putative HphA</td>
<td><em>Oscillatoria</em> sp. PCC 6506</td>
<td>394</td>
<td>60%, 79%</td>
<td>ZP_07113410</td>
</tr>
<tr>
<td>putative HphA</td>
<td><em>Synechococcus</em> sp. PCC 7335</td>
<td>411</td>
<td>56%, 74%</td>
<td>ZP_05037378</td>
</tr>
<tr>
<td>LeuA</td>
<td><em>Escherichia coli</em> W3110</td>
<td>523</td>
<td>26%, 44%</td>
<td>YP_488380</td>
</tr>
</tbody>
</table>

#### (B) HphB and its homologues

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Organisms</th>
<th>Length (aa)</th>
<th>Homology, Similarity</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>HphB</td>
<td><em>Nostoc punctiforme</em> PCC73102</td>
<td>363</td>
<td>100%</td>
<td>YP_001865961</td>
</tr>
<tr>
<td>putative HphB</td>
<td><em>Nodularia spumigena</em> CCY9414</td>
<td>422</td>
<td>79%, 90%</td>
<td>ZP_01629946</td>
</tr>
<tr>
<td>putative HphB</td>
<td><em>Oscillatoria</em> sp. PCC 6506</td>
<td>352</td>
<td>70%, 88%</td>
<td>ZP_07113411</td>
</tr>
<tr>
<td>protein</td>
<td>organism</td>
<td>accession</td>
<td>identity</td>
<td>similarity</td>
</tr>
<tr>
<td>-------------</td>
<td>---------------------------------</td>
<td>-----------</td>
<td>----------</td>
<td>------------</td>
</tr>
<tr>
<td>putative HphB</td>
<td><em>Synechococcus</em> sp. PCC 7335</td>
<td>349</td>
<td>64%, 76%</td>
<td>ZP_05038014</td>
</tr>
<tr>
<td>LeuB</td>
<td><em>Escherichia coli</em> W3110</td>
<td>363</td>
<td>42%, 56%</td>
<td>YP_488379</td>
</tr>
</tbody>
</table>

(C) HphCD and its homologues

<table>
<thead>
<tr>
<th>protein</th>
<th>organism</th>
<th>accession</th>
<th>identity</th>
<th>accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>HphCD</td>
<td><em>Nostoc punctiforme</em> PCC73102</td>
<td>569</td>
<td>100%</td>
<td>YP_001865962</td>
</tr>
<tr>
<td>putative HphCD</td>
<td><em>Nodularia spumigena</em> CCY9414</td>
<td>574</td>
<td>82%, 91%</td>
<td>ZP_01629947</td>
</tr>
<tr>
<td>putative HphCD</td>
<td><em>Oscillatoria</em> sp. PCC 6506</td>
<td>581</td>
<td>70%, 83%</td>
<td>ZP_07113412</td>
</tr>
<tr>
<td>putative HphCD</td>
<td><em>Synechococcus</em> sp. PCC 7335</td>
<td>567</td>
<td>69%, 83%</td>
<td>ZP_05037298</td>
</tr>
<tr>
<td>LeuC</td>
<td><em>Escherichia coli</em> W3110</td>
<td>466</td>
<td>33%, 48%</td>
<td>YP_488378</td>
</tr>
<tr>
<td>LeuD</td>
<td><em>Escherichia coli</em> W3110</td>
<td>201</td>
<td>59%, 68%</td>
<td>YP_488377</td>
</tr>
</tbody>
</table>
FIG 1 Location of L-Hph biosynthesis genes in the anabaenopeptin synthetase cluster of *Nostoc punctiforme* PCC73102 (A), and the proposed L-Hph biosynthetic pathway mediated by HphA, HphCD, HphB, and aromatic aminotransferase (ATase) in a similar route to Leucine biosynthesis (B). The Hph Biosynthetic pathway is proposed as follows; L-Phe is converted into phenylpyruvic acid via a transamination reaction mediated by the aminotransferase, such as tyrosine aminotransferase (TyrB). Phenylpyruvic acid is condensed with an acetyl-CoA catalyzed by HphA, and the resulting thioester is spontaneously hydrolyzed, leading to 2-BMA. 2-BMA is converted into 3-BMA via isomerization of hydroxyl group mediated by HphCD. Hydroxyl moiety of 3-BMA is oxidized by HphB and followed by spontaneous decarboxylation, providing 2-OPB. Finally, 2-OPB is converted into L-Hph via a transamination reaction mediated by the aminotransferase, such as TyrB.

2-BMA; 2-Benzylmalic acid, 3-BMA; 3-Benzylmalic acid, 2-OPB; 2-Oxo-phenylbutanoic acid.

FIG 2 HPLC chromatogram of supernatants obtained from fermentative broths of *E. coli* W3110 expressing the corresponding Hph proteins. Lane 1; HphCD expression only, lane 2; HphA and HphCD expression, lane 3; HphA, HphCD and HphB expression, lane 4; L-Hph
standard (1 g/L). L-Hph and L-Phe elute at 8.1 and 5.2 min, respectively. The peak at 16.1 min is 2-BMA.

FIG 3 Improvement of L-Hph production by *E. coli* W3110 strains harboring the various plasmids listed in Table 2. The error bars represent s.d. of the mean of three independent experiments.

FIG 4 Complementation experiments with Hph and Leu biosynthesis genes. Growth (OD at 660 nm) of *leu* gene-deficient mutants and mutants complemented with the corresponding *hph* gene, as listed in Table 3, in M9 medium. The error bars represent s.d. of the mean of three independent experiments (A). L-Hph production by strains containing *hph* and *leu* gene combinations as listed in Table 4. The yields of L-Hph are calculated as an average of two independent experiments, the values of which are within 10% errors (B).

FIG 5 Classification of substrates used to examine the substrate specificity of the L-Hph-producing system. DL-*m*-Fluoro-Phe, DL-*o*-Fluoro-Phe and L-Tyr are converted into corresponding homoamino acids. β-Methyl-Phe, L-Phenylglycine, L-*p*-Iodo-Phe, L-*p*-Nitro-Phe and L-Trp could not produce the homoamino acids or the biosynthetic intermediates.
Acceptable substrates

- L-Phe
- L-Tyr
- p-F-Phe
- o-F-Phe

Generated homoamino acids

- L-Hph
- L-Hty
- L-p-F-Hph
- L-o-F-Hph

Non-acceptable substrates

- β-MePhe
- L-PhenylGlycine
- p-I-Phe
- p-Nitro-Phe
- L-Trp