Directed evolution of key enzymes in the synthesis of 4-hydroxymandelate pathway

Qianwen Jin, Xuanye Li, Yanjun Ji, Ruilin Zhang, Yuting Gao, Haiming Yuan, Dengke Fan, Zichang Lu, Siyu Li, Hui Wang, Mengjin Zhang, Ge Song, Jiaxing Liu, Pengchao Wang, Guangchao Sui

Abstract:

In this study, we employed metabolic engineering and directed evolution strategies to investigate and address two key challenges, including insufficient precursor supply and limited activity of crucial enzymes, in the biosynthesis of 4-hydroxymandelate (HMA), a high-value aromatic compound. Through gene overexpression and multi-level gene interference using the CRISPRi approach, we successfully engineered an *E. coli* chassis strain capable of stably producing the essential substrate 4-hydroxyphenylpyruvate (HPP) and the titer reached 0.91 g/L. In a high-throughput screening process using a biosensor based on allosteric transcription factors, we eliminated the reliance on expensive equipment. The activity of the key enzyme and the viability of the strain were coupled through directed evolution. We eventually obtained a mutant clone, $HmaS^{V152G}$, with a 125% improvement in the catalytic rate. In its fermentation, we achieved the high production of 3.63 g/L HMA in 24 h with a single supply of 20 g/L glucose.

Keywords:

Directed evolution, 4-hydroxymandelate, metabolic engineering, CRISPRi

1. Introduction

Four-hydroxymandelate (HMA) is an aromatic molecule that has a high application value in various fields, such as medicine, cosmetics, food and chemical industry[1]. In the pharmaceutical field, HMA serves as a precursor for the synthesis of various drugs, including the antihypertensive drug atenolol, and a variety of medications for cancers, allergies and diabetes. In the cosmetics industry, HMA is also frequently added to beauty products due to its relatively high antioxidant capacity compared to α-tocopherol. In the food and flavor synthesis field, ethyl vanillin, a derivative of HMA, has great prospects due to its sweet chocolate aroma[2].

The existing methods to obtain HMA include natural extraction, chemical synthesis and biological synthesis. Generally, major chemical methods for extraction involve the condensation reaction of phenol and glyoxylic acid, as well as electrochemical reactions of phenol and trichloroacetaldehyde. Clearly, these methods give rise to serious environmental pollution and the discharged wastewater can corrode various equipment. In actual production, it is also necessary to consider the issues regarding high cost of raw materials and low purity of products. In the biocatalytic synthesis of HMA, the key step is the conversion of 4-hydroxyphenylpyruvate (HPP) to HMA catalyzed by hydroxymandelate synthase (HmaS). Nevertheless, the activity of HmaS is a limiting factor exhibiting a conversion rate of only 19.4%. This bottleneck in the crucial conversion step limits the production of HMA and has become the focus of our research.

In this study, we planned to increase HMA yield by two methods: constructing an upstream chassis strain with a high-yield

substrate and boosting the key enzyme activity downstream. Directed evolution has been perceived as an efficient renovation method to improve the activity of downstream key enzymes. For example, for the directed evolution of the PccB protein in the biosynthesis pathway of succinic acid, the growth of the host bacteria and the enzyme activity were coupled together[3]. Therefore, only when the enzyme activity was significantly improved, and sufficient acetyl-CoA was provided as a substrate, the bacteria could survive in the minimal culture medium.

Through optimizing systematic metabolic engineering process, unsatisfactory bioproduction efficiency can be significantly enhanced. For example, researchers could achieve L-leucine yield of 63.29 g/L, the highest yield ever reported, by releasing feedback inhibition of key enzymes in the L-leucine biosynthesis pathway, enriching essential precursors including pyruvate and acetyl-CoA, adjusting the redox flux, and enhancing L-phenylalanine efflux[4].

The substrate HPP of the key reaction in the 4-hydroxymandelate synthesis pathway is also a high-value aromatic compound. Thus, we decided to engineer an *Escherichia coli* (*E. coli*) strain with high HPP levels in complete biosynthesis. In previous studies, a *de novo* pathway for the HMA synthesis has been established in *Saccharomyces cerevisiae* (*S. cerevisiae*)[2]. However, the research of developing *E. coli* as the host strain remains as a gap. Compared to *S. cerevisiae* and other bacterial microorganisms, *E. coli* has several advantages as a production host, including its fast growth rate, clear genetic background and ease of genetic manipulation. Therefore, it is highly demanded to engineer this chassis.

2. Results

2.1 Construction and Test of a Dual-Module Screening Platform

Two key aspects of directed evolution include the construction of a diverse mutant library and developing an appropriate high-throughput screening method. While methods for building mutant libraries are well-established, selecting an effective screening method can expedite the analysis of experimental results. Flow cytometry-based droplet sorting technology allows rapid screening of dominant clones, but its application is still limited by high investment of the equipment, particularly in newly established laboratories. Therefore, we planned to develop a low-cost and user-friendly high-throughput screening biosensor based on allosteric transcription factors[5].

We constructed the gY9s-HmaS-Bio177 plasmid containing two essential gene modules, the catalytic module and the sensing module. In the catalytic module, the *hmaS* gene is regulated by a constitutive promoter. The sensing module contains two important reporter genes: the red fluorescent protein *mCherry* gene and the cytosine deaminase *codA* gene that converts 5-fluorocytosine (5-FC) to 5-fluorouracil (5-FU), which kills bacteria. The catalytic activity is coupled with the strength of the reporter gene signal. HPP levels decrease in response to high HmaS enzymatic activity, resulting in weak reporter signal, and the strain shows a light color and can survive in the medium containing 5-FC (Fig. 1a). Meanwhile, we demonstrated that the biosensor based on an allosteric transcription factor specifically responded to the substrate HPP (Supplementary Fig. 1). To test its functionality, we used bacteria harboring the Bio177 plasmid to simulate the scenario of low activity, and bacteria containing the gY9s-HmaS-Bio177 plasmid for high enzyme activity (Fig. 1b). Both strains were spread on selection plates containing 0.01, 0.05 and 0.1 mg/L of 5-FC to simulate the screening process under stress conditions (Supplementary Fig. 2). We found that, in medium with 0.01 mg/L 5-FC, all 10 randomly selected clones were verified as positive by colony PCR and contained the

hmaS gene, yielding a 100% selection rate to distinguish desirable *hmaS*-containing clones from others (Fig. 1b). Under condition of 0.05 mg/L 5-FC, the growth of the wild-type HmaS strain was inhibited, resulting in a reduced colony number on the plate. Ultimately, the WT strain could not grow under 0.1 mg/L 5-FC condition. Overall, our results not only validated the feasibility of the dual-module system for the screening, but also primarily determined 5-FC concentrations used to distinguish mutant and WT clones.

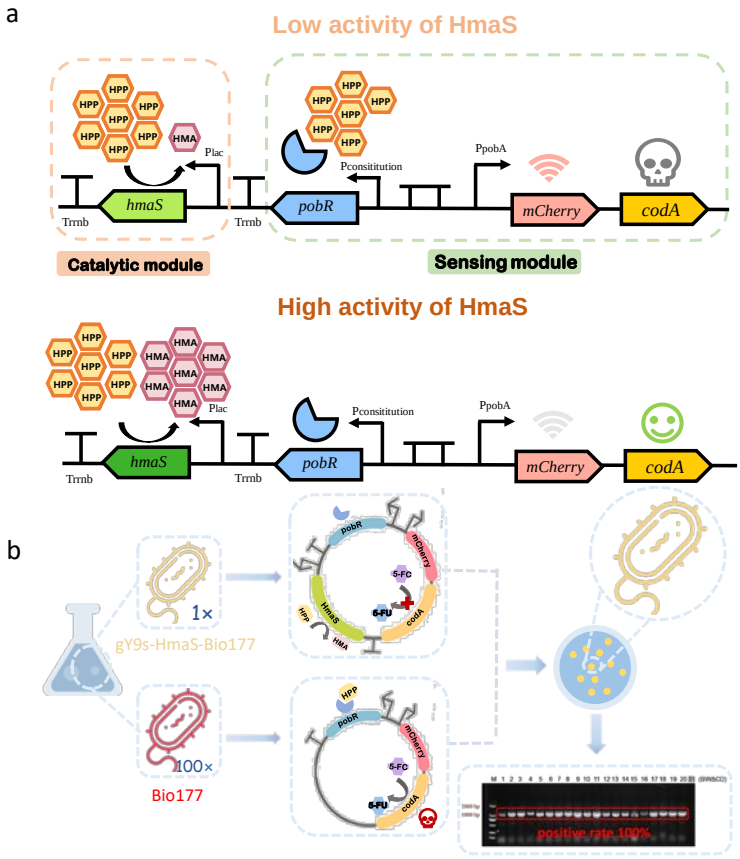

Figure. 1 The strategy of the dual-module screening platform and the principle of the screening. **a.** Schematic diagrams of the constructed plasmids to screen for clones with high HMA production. pobR: the gene encoding the allosteric transcription factor PobR protein that specifically responds to 4-hydroxyphenylpyruvate (HPP). p_{pobA}: a promoter that specifically binds the PobR complex to control downstream reporter gene expression. The activity of the HmaS enzyme is essential to the remaining levels of HPP in the medium, subsequently determining the strength of the reporter signal, as well as the color and viability of the bacteria. **b.** Schematic diagram to screen for clones expressing high levels of HMA. Bio177 and gY9s-HmaS-Bio177 were mixed at a 100:1 ratio, and the bacteria were cultivated on the screening plates. The colonies on the plates were evaluated by colony PCR.

2.2 Construction and Screening of the Random Mutant Library

Based on error-prone PCR, we used a low-fidelity DNA polymerase and adjusted Mn^{2+} and Mg^{2+} concentrations and ratio to manipulate the mutation rate. We optimized the Golden Gate assembly procedure to increase the probability of inserting mutated fragments into the vector. Eventually, we constructed a mutant library containing 1×10^5 clones (Supplementary Fig. 3) and the rate of positive clones was 100% (Supplementary Fig. 4). We determined the final concentrations of 0.02 mM Mn^{2+} and 0.5 mM Mg^{2+} for mutagenesis and achieved a mutation rate of 2 mutations/kb, producing a high variety of mutations.

Next, we conducted the screening in liquid media and refined 5-FC concentration gradient to 0.1, 0.2 and 0.3 mg/L for the selection. After inducing protein expression, we transferred the bacterial cultures into the selection media and cultivated the bacteria at 30°C until the bacteria reached the exponential phase (OD_{600} = 0.6). We extracted whole-cell catalysate and analyzed them using HPLC. Based on colony fluorescence and HMA production, we obtained a clone with significantly increased enzyme activity under the conditions of 0.6 g/L HPP and 0.3 mg/L 5-FC. This condition was subsequently used in the following screening process (Fig. 2a).

We found that HMA production of mutant clones could increase up to 125% compared to that of WT (Fig. 2b). However, a significant ratio of false positive clones incapable to produce any HMA were identified. After DNA sequencing analysis, we discovered that the catalytic module and/or recognition module was lost in these bacterial strains. To reduce this phenomenon, we changed the timing for colony isolation to OD_{600} = 0.2 of bacterial density.

Thus, in the next round of screening, we tested individual colonies at OD_{600} = 0.2. Under this condition, 60% of the strains showed increased production compared to the WT, the false positive ratio was decreased by 84.6%, and the screening positivity rate was 75%.

We continued multiple rounds of screening under the optimized condition and selected the individual clone of gY9s-$HmaS^{V152G}$-Bio177 with the highest enzyme activity. We sequenced the mutant clone and discovered that the codon GTG for Val152 of the *hmaS* gene was replaced by Gly, forming the $HmaS^{V152G}$ mutant.

2.3 Modeling and Molecular Docking

We conducted homology modeling and molecular docking to analyze whether the selected mutant $HmaS^{V152G}$ altered the local structure of the HmaS protein and changed its binding affinity and reaction activity with HPP molecule. We employed the Alphafold2[6-7] for homology modeling of both the $HmaS^{WT}$ and mutant $HmaS^{V152G}$ (Fig. 2c). Next, we used the Autodock Vina to dock the HPP molecule to HmaS proteins. The docking energy of HPP with WT and V152G mutant was determined as -3.8 and -4.7 kcal/mol, respectively, indicating that HPP is likely better oriented towards the HmaS mutant. Following docking analysis, we evaluated the ligand-protein interactions using Pymol and Ligplus. After HPP docking with $HmaS^{V152G}$, HPP forms hydrogen bonds with D154, G199 and A204, and also engages in hydrophobic interactions with L153, G203, T206 and R318 (Fig. 2d).

In the surface model, the binding pocket of the V152G mutant seems to better accommodate the HPP molecule than WT, which possibly allows hydrophobic forces to better secure the molecule. Moreover, an additional hydrogen bond is formed between HPP and V152G versus HPP-WT interaction. In summary, the V152G mutation appears to confer HmaS a better ability to bind HPP, thereby enhancing its catalytic activity in converting HPP into HMA. This finding provides valuable insights into how mutations can impact the structure and function of enzymes, advancing the understanding of enzymatic catalysis.

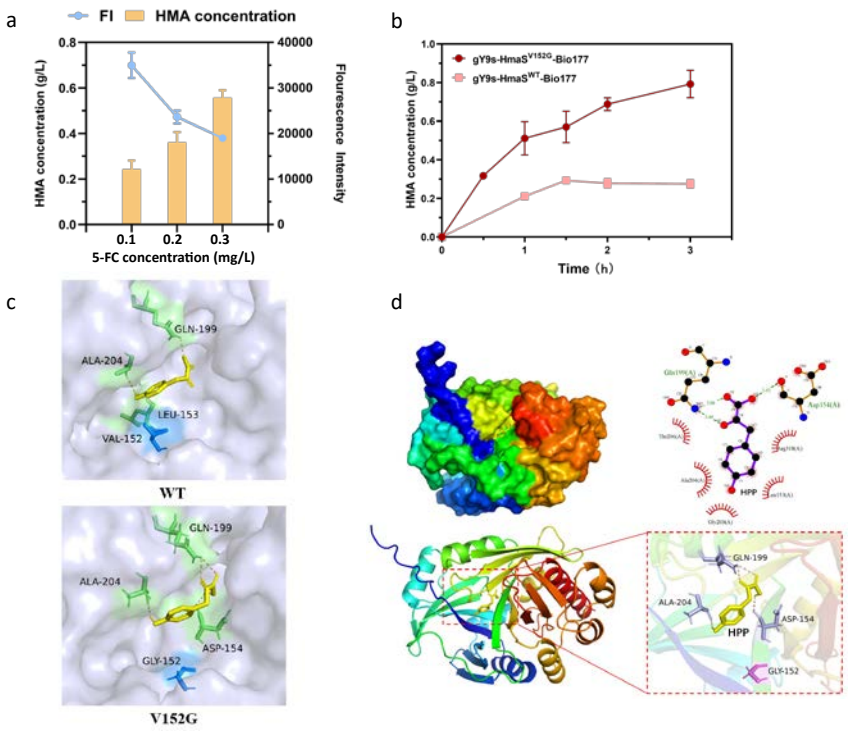

Figure 2. Selection and analysis of the *hmaS* mutant library. **a.** Graph of the correlation between 5-FC concentration in the medium and HMA production. With increased 5-FC concentrations, the selection pressure accelerated, which enhanced the performance of viable individuals, improved HMA production, and reduced fluorescence strength. The orange bars represent HMA levels and blue line depicts fluorescence intensity. **b.** Graph of HMA production by the bacteria harboring $HmaS^{WT}$ and $HmaS^{V152G}$ clones. With 9 mM HPP as the substrate for whole-cell catalysis, HMA accumulation by the gY9s-$HmaS^{V152G}$-Bio177 and gY9s-$HmaS^{WT}$-Bio177 strains was monitored for 3 h. **c.** The surface models for the spatial structures of the $HmaS^{WT}$ and $HmaS^{V152G}$ mutants in the binding domain of HmaS. The residue 152 is colored in dark blue. The V152G mutation alters the positions of adjacent residues and subsequently increases the wideness of the pocket to easily accommodate HPP molecules. **d.** Simulated structures of $HmaS^{V152G}$-HPP binding analyzed by Pymol and Ligplus. Compared to $HmaS^{WT}$, $HmaS^{V152G}$ uses G152, L154 and E199 to form additional hydrogen bonds with a HPP molecule.

2.4 Engineering an *E. coli* chassis with High-HPP Production

In this study, we aimed to utilize *E. coli* endogenous glycolytic pathway, pentose phosphate pathway and shikimate pathway, which can achieve *de novo* HPP synthesis with glucose as a substrate. However, under natural conditions, the accumulation of aromatic compounds in *E. coli* is limited, which prevents us from using HPP as a catalytic substrate[8]. Therefore, we designed two strategies: overexpressing and silencing key genes in biosynthesis pathways, to improve the capability of bacteria for HPP synthesis.

2.4.1 Overexpression of Key Genes to Enhance the Precursor Supply

Phosphoenolpyruvate (PEP) and erythrose-4-phosphate (E4P) are two important precursor molecules for the synthesis of aromatic compounds. However, the fact of limited flux inhibits the synthesis of downstream products. To address it, we constructed the pSB1c-$aroG^{fbr}$-tktA-ppsA (pATP) plasmid, in which the expression of three key genes was controlled by an arabinose-inducible promoter. Overexpression of the phosphoenol-pyruvate synthase (*ppsA*) gene reduced PEP flowing towards pyruvate, leading to PEP accumulation[9]. To enhance E4P supply, we overexpressed the transketolase (*tktA*) gene to promote fructose-6-phosphate consumption[10]. With overexpressed *$aroG^{fbr}$*[11] gene, both crucial precursor molecules showed flux into the shikimate pathway (Fig. 3d). We introduced the pATP plasmid into the expression strain BWΔCD

and conducted shake-flask fermentation followed by HPLC analysis. Through the modifications of genetic overexpression, we obtained a host strain that could produce up to 0.72 g/L HPP.

2.4.2 Building Multi-level Inhibition Platform Based on CRISPRi to Reduce Competitive Pathway

CRISPRi-mediated gene suppression may result in cellular adaptability defects, especially when targeted genes are essential for cell growth[12-13]. In order to achieve multi-level inhibition of competitive genes for HPP production, we employed the CRISPRi technique based on mismatch sgRNAs[14]. To alter the binding efficiency of the dCas9 protein to the target genes and generate various levels of inhibition, we introduced mutations in the 7-8 bp of sgRNA spacers to create mismatched sgRNAs, which totally encompassed 16 possible variants (Fig. 3a). Through assessing the impact of different sgRNA variants on HPP production, we successfully selected the optimal inhibition strength for the competitive genes, leading to enhanced HPP production and minimized growth stress caused by the gene knockout.

Initially, we constructed a plasmid containing both the dCas9 coding sequence and sgRNA, induced by IPTG. Mutations in sgRNA spacers were introduced using reverse primers. To test the feasibility of this method, we co-transformed *E. coli* BW25113 with R6K-dCas9-eGFP plasmid containing the sgRNA targeting the eGFP coding sequence (+40bp) and pYB1a-eGFP plasmid. Meanwhile, characterization tests were performed (Supplementary Fig. 6). We tested the effectiveness of sgRNAs with different mismatches in mediating eGFP suppression. DNA sequencing results from 24 randomly selected colonies revealed 13 mutation types, accounting for 81.25%, with eGFP repression levels from 6.2% to 80.3% (Fig. 3b, Supplementary Fig.5). This pre-study proved that the CRISPRi method using double mismatched sgRNAs could achieve multi-level inhibition in a relatively straightforward and controllable manner.

To increase the accumulation of the precursor HPP in the bacteria and promote *de novo* synthesis of HMA, we selected 4 relevant genes *pykF*, *tyrB*, *tyrR* and *pheA*[15-16] as targets of inhibition. These four genes are involved in the pathways leading to the synthesis of alanine, tyrosine and phenylalanine[17]. We designed corresponding mismatch sgRNA libraries for these four genes, and added BsaI restriction sites at both ends of all sgRNAs. We mixed four mismatch sgRNA libraries and a carrier plasmid containing dCas9 and other necessary parts, and used the Golden Gate method to randomly assemble the four mismatch sgRNAs into one plasmid, leading to the construction of a multiple sgRNA library for polygenic inhibition (Fig. 3c). We individually co-transformed the plasmids of these R6K-dCas9-sgRNA libraries with the pSB1c-aroG^{fbr}-tktA-pheA plasmid into *E. coli* BWΔCD, induced them in the ZYM5052 medium, and transferred them into M9 medium containing 20 g/L glucose for fermentation. After 24 h of cultivation, we performed HPLC analysis to test their HPP production. We observed that the inhibition of these four genes to different degrees could increase the strain's ability to accumulate HPP. Especially, when *tyrR* was silenced, a significant increase of HPP was achieved with an improvement of 113.6% (0.704 g/L) compared to the strains only expressing the enhanced upstream pathways. Ninety percent of the mismatched sgRNA variants exhibited to different degrees of improved HPP production. Among them, we selected a non-mutant *tyrR*i strain for flask fermentation, and the HPP titer reached 0.91 g/L. Noteworthily, in these multi-sgRNA mutant library, we did not find any clone with a superior HPP increase, possibly due to the limitations of library capacity and/or screening efficiency.

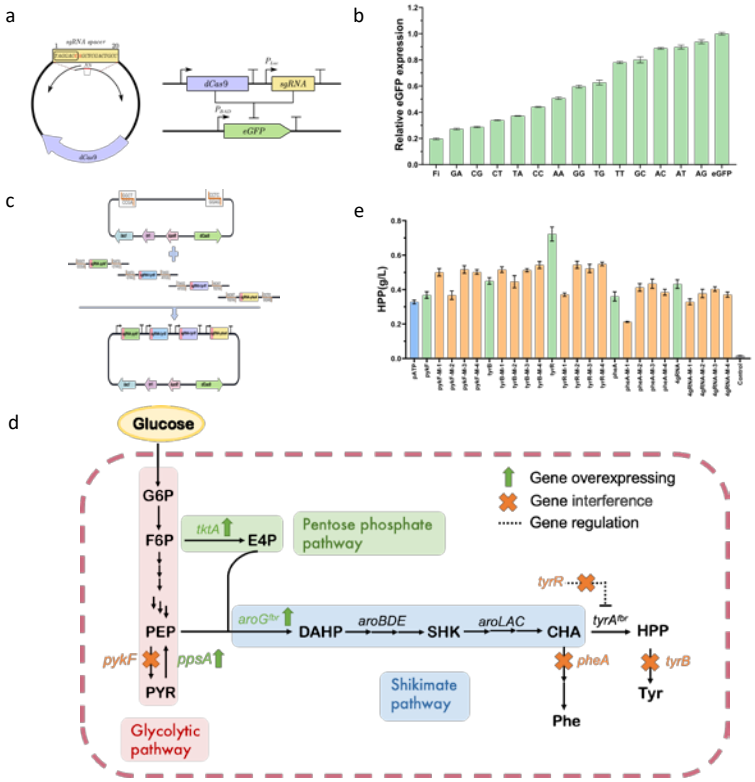

Figure. 3 The construction of sgRNA library and its uses in improving HPP production. **a.** Schematic diagram of generating sgRNAs for Multi-level Inhibition based on CRISPRi and mismatch sgRNAs. By co-expressing sgRNAs and the dCas9 protein in the same vector, point mutations at the 7-8bp spacer of sgRNA seed region were introduced, which enabled interference of target gene expression. **b.** Bar graph of the Inhibitory efficiency by the random double-mismatched eGFP-sgRNAs. **c.** Construction of the plasmids containing sgRNAs targeting four different genes using the Golden Gate method. The generated plasmid allows random combinations of these sgRNAs with differential inhibitory activities to create a mutant library. **d.** The biosynthetic pathway of HPP using glucose as a substrate in *E. coli*. Green arrows represent overexpressed genes, red crosses indicate inhibited genes, and dashed lines signify feedback inhibition. G6P: Glucose 6-phosphate, F6P: Fructose 6-phosphate, PEP: phosphoenolpyruvate, E4P: D-erythrose 4-phosphate, PYR: pyruvate, DAHP: 3-deoxy-D-arabino-heptulosonate-7-phosphate, SHK: shikimate, CHA: chorismite, HPP: 4-hydroxyphenylpyruvate, Phe: Phenylalanine, Tyr: tyrosine, *tktA*: transketolase I, *pykF*: pyruvate kinase, *ppsA*: phosphoenolpyruvate synthase, *tyrA*: chorismate mutase/prephenate dehydrogenase, *tyrB*: aromatic-amino-acid transaminase, *tyrR*: a transcriptional regulator, *pheA*: chorismate mutase/prephenate dehydratase, *aroG*: 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase, *aroBDE*: *aroB*: 3-dehydroquinate synthase, *aroD*: 3-dehydroquinate dehydratase, *aroE*: dehydroshikimate reductase, *aroLAC*: *aroL*: shikimate kinase, *aroA*: EPSP synthase, *aroC*: chorismate synthase. **e.** Bar graph of HPP production with different genes inhibited by CRISPRi. The blue column represents the strain transformed by pATP, the green column is for the strain transformed by pATP and R6K-dCas9-sgRNA, and the orange columns indicate the strain transformed by pATP and R6K-dCas9-sgRNA mutants. The control group was the BWΔCD strain carrying no plasmid. Fi means fully interference, transformed into the strain that containing the non-mutated R6K-dCas9-eGFP plasmid, which had CA in the 7-8bp spacer of the sgRNA.

2.5 Evaluation of Metabolic Engineering and Directed Evolution for HMA Production Enhancement

The optimized clones (gY9s-HmaS^{V152G}-Bio177) obtained from the screening were transformed into the bacterial host (pATP+*tyrR*i) that was created by metabolic engineering modifications. After cultivation and induced protein expression, the bacteria were transferred to a fermentation medium using glucose as the substrate and fermented in a shake flask at 30°C for 8 h. Simultaneously, we established control groups in the Table 1. By comparing HMA1, HMA3 and HMA5 clones, we found that gene overexpression or silencing alone had limited effects on overall fermentation promotion. However, the combination of these two strategies could lead to a remarkable increase in the fermentation yield. When two approaches are combined, it allowed the endogenous HPP synthesis pathway to achieve balanced, stable and

sustained expression. Importantly, when comparing the HMA1 and HMA2, HMA3 and HMA4, HMA5 and HMA6 pairs, we found that the essential limiting factor was the enzymatic activity of hydroxymandelate synthase, which significantly restricted HMA production. Fermentation was carried out for a long time using the HMA5 and HMA6 clones (Fig. 4a). Aliquots were collected every two hours, and product detection was performed using HPLC analysis to assess the performance of these two modification methods. Compared with $HmaS^{WT}$, the $HmaS^{V152G}$ obtained through the screening, showed a 4.73-fold increase in HMA production and produced 3.63 g/L HMA after 24 h of fermentation (Fig. 4b). This also confirmed our initial hypothesis that the limited factor are insufficient precursor supply and low enzyme activity of crucial enzymes. Importantly, we identified the critical targets to improve the HMA production. The extremely high yield in HMA6 demonstrated that our proper combination of upstream chassis modification to increase precursor supply and downstream evolution of key enzymes to improve conversion rates is very innovative and effective.

Table 1. List of plasmids used to generate different strains in each group for fermentation.

Groups	Constructed plasmids
HMA1	gY9s-$HmaS^{WT}$-Bio177+ pATP
HMA2	gY9s-$HmaS^{V152G}$-Bio177+ pATP
HMA3	gY9s-$HmaS^{WT}$-Bio177+ *tyrR*i
HMA4	gY9s-$HmaS^{V152G}$-Bio177+ *tyrR*i
HMA5	gY9s-$HmaS^{WT}$-Bio177+ pATP+*tyrR*i
HMA6	gY9s-$HmaS^{V152G}$-Bio177+ pATP+*tyrR*i

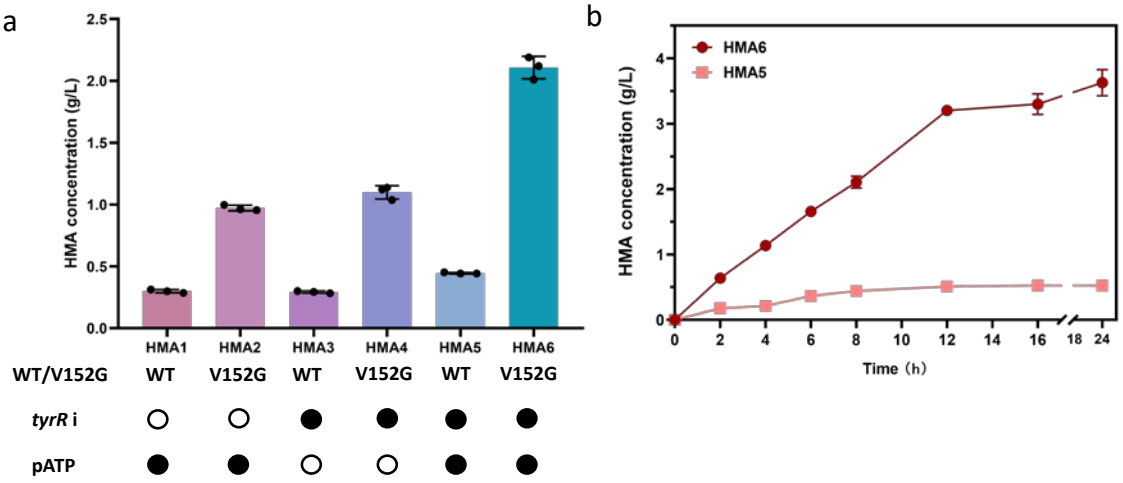

Figure. 4 Fermentation of HMA producing strains and evaluation of metabolic engineering and directed evolution. **a.** Bar graph to compare relative HMA levels produced by different strains. Solid black dots indicate the presence of the plasmids, hollow black dots indicate their absence. **b.** Comparison of HMA production by the HMA5 and HMA6 strains. The bacteria were fermented in shaking flasks for 24 h with a single supply of 20 g/L glucose as the substrate, and HMA levels monitored at indicated time points.

3. Discussion

In this project, we applied metabolic engineering to obtain a high-HPP expression chassis strain (pATP+ *tyrR*i) that increased the conversion efficiency of hydroxymandelate synthase to 125% (gYB2s-$HmaS^{V152G}$-Bio177). We fermented the bacteria in a shake flask using glucose as the substrate. Successfully, we achieved significantly improved field and the highly efficient HMA biosynthesis: producing 3.63 g/L HMA after 24 h of fermentation. The analysis of this significant improvement revealed the following effective approaches: overexpressing the genes involved in the

synthesis of key precursors to adjust the metabolic flux in *E. coli*, and silencing the *tyrR* gene that is a critical feedback inhibition factor. These led to a greater flow of glucose into the mandelate pathway for aromatic compound synthesis. Furthermore, through directed evolution, we obtained mutant clones ($HmaS^{V152G}$) with enhanced enzymatic activity. Changes in key amino acid residues significantly reduced the energy required for hydroxymandelate synthase to bind HPP, which enlarged the binding pocket allowing easy substrate entry to the catalytic center. The significant increase in enzymatic activity enabled continuous and efficient HPP-to-HMA conversion, overcoming the intrinsic balance constraints of *E. coli* for HPP production and reducing the pressure on glucose flow towards HPP. Additionally, since the hydroxymandelate synthase was ectopically expressed, its expression is barely restricted in the host bacteria.

We also conducted a thorough analysis of various phenomena during the experimental process: 1. When constructing a random mutant library, we employed a discontinuous random mutagenesis method that allowed us to conveniently control the mutation rates. To address the limitation of this method in generating mutations continuously, we constructed the eMuta T7 plasmid, which can express both T7 RNA polymerase and cytidine deaminase[18-19]. We added dual-T7 promoters to the target gene. The high specificity of T7 RNA polymerase and T7 promoters enabled the specific mutations. This continuous mutation tool has been successfully applied to low-copy plasmids previously, and we anticipate its application to high-copy target plasmids by adjusting protein expression levels. 2. Screening in the logarithmic growth phase of the strains caused a higher false-positive rate. Based on the sequencing results, we speculated that 5-FU interfered with bacterial growth when OD_{600} = 0.6, causing severe growth stress in *E. coli*, which could subsequently resort to gene loss as a survival strategy. When we screened the clones at an earlier time point, *E. coli* experienced relatively low stress, which did not allow them to adjust most of its genes, leading to a significant reduction in the false-positive rate of the screening. In summary, we obtained continuous and balanced HPP production chassis strains through two metabolic engineering approaches: gene overexpression and gene interference, which ensured an adequate supply of precursors. Moreover, in the critical downstream conversion process, we used a simple high-throughput screening technique based on biosensors to select clones with significantly improved enzymatic activity from a massive mutant library. This unprecedented high yield demonstrated that through proper experimental design and timely adjustments of experimental procedures, we can reduce the reliance on high investment of laboratory equipment, improve the efficiency of manual screening, and achieve our goal to obtain high-yield strains.

4.Methods

4.1 Chemicals, strains and materials

The strains and plasmids used in this study were listed in Supplementary Table 1. The primers used in this study were synthesized by Sangon Biotech (Shanghai, China), and listed in Supplementary Table 2. *E. coli* DH5α was used as the host strain to construct all recombinant plasmids, and *E. coli* BW25113 was used for HPP biosynthesis and protein expression. All kits used in this study were from Omega Bio-tek (Guangzhou, China). P-hydroxyphenylpyruvate was purchased from Aladdin, and 5-fluorocytosine was from Macklin. Other chemicals were purchased from OXOID (Beijing, China) or Sigma-Aldrich (Shanghai, China). High-fidelity enzyme premix (YEASEN, Shanghai, China) was used in PCR amplification. Restriction enzymes were purchased from New England Biolabs (Beijing, China). Luria-Bertani (LB) broth (10 g/L tryptone, 5 g/L yeast extract, 10 g/L

NaCl) was used for bacteria construction. Screening was performed using the media supplemented with 50 × 5052 (25 g/L Glycerol, 2.5 g/L Glucose), 5 × M9, 1 M $MgSO_4$ and 1M $CaCl_2$. Protein expression was performed using the ZY medium, 50 × 5052, 50 × M, 1000 × trace element, 1 M $MgSO_4$, 0.2% arabinose and 1mM IPTG. Chloramphenicol (30 mg/L), Streptomycin (50 mg/L), Kanamycin (50 mg/L) and Ampicillin (100 mg/L) were used where appropriate.

In this study, bacterial growth and fluorescence values were detected using a Multimode Plate Reader (TECAN SPARK, Shanghai, China). The concentration of the generated 4-hydroxymandelic acid (HMA) was analyzed using an HPLC system (Agilent Technologies 1100, Shanghai, China) equipped with a C18 reverse-phase column (Welchrom® series, Shanghai, China).

4.2 Construction of the *hmaS* mutant library

The mutant library was constructed by introducing random mutations to the *hmaS* gene using error-prone PCR and adjusted the concentration of Mn^{2+} (0.5 mM/μL) and Mg^{2+} (0.02 mM/μL) in the reaction system to obtain desirable mutation rates. Next, the hmaS mutant fragment was connected with Bio177 fragment by the Golden Gate assembly method, using T4 DNA ligase and BsaI restriction enzyme. To optimize the procedure of ligation, we continuously tested T4 DNA ligase and restriction endonucleases operating temperatures to alter their functions. The recombinant plasmids were transformed into *E. coli* DH5a and cultured on plates containing streptomycin for 12 h.

4.3 Screening of mutant library

The colonies on the plate were rinsed by liquid LB and then cultured in 100 mL LB containing streptomycin in a shaker for 12 h to extract gy9s-HmaS-Bio177 random mutant library plasmids. The bacteria harboring the gy9s-HmaS-Bio177 random mutant library plasmids were transferred into the expression strain BWΔCD and cultured at 37°C for 12 h. The bacteria were then cultured in configured ZYM5052 medium to induce protein expression at 30℃ for 20 h. The concentrations of 0.3 mg/L 5-FC and 0.6 g/L HPP were found to be optimal. The induced strain was transferred to the screening medium to examine bacterial growth. When the strain grew to logarithmic phase (OD_{600} = 0.2 or 0.6), the fluorescence intensity was measured. A single colony was isolated from the bacteria and subjected for 12 h of extended culture in a shaker at 37 ℃. Bacterial fluid of 6 OD was harvested for whole-cell catalysis. The product accumulation was determined using HPLC analysis.

4.4 Construction of mutant library using the CRISPRi platform

To create a CRISPRi platform based on dCas9 and mismatched sgRNA, pSC101-dCas9-eGFP was modified based on the Cas9 protein expression vector through mutating Cas9 to make dCas9, and SgRNA targeted the coding sequence (+40 bp) of eGFP was amplified by PCR, and the PCR product was inserted into pSC101-dCas9-eGFP, then we replaced the ori of this plasmid to obtain R6K-dCas9-eGFP as the base plasmid for subsequent characterization. We replaced sgRNA spacers to individually target *pykF*, *tyrB*, *tyrR* and *pheA* in R6K-dCas9-eGFP, and thus generated R6K-dCas-pykF, R6K-dCas-tyrB, R6K-dCas-tyrR, R6K-dCas-pheA. These four sgRNA fragments were linked together by the Golden Gate method to generate the R6K-dCas9-pykF-tyrB-tyrR-pheA plasmid. Using degenerate primers in the amplifications of these sgRNAs, we obtained their mutated fragments, which were subsequently subcloned into vectors to construct sgRNA-mutated plasmids leading to the construction of the mutant library.

4.5 Characterization of CRISPRi platform

To assess the capacity of the CRISPRi system in repressing eGFP gene expression, the R6K-dCas9-eGFP plasmid was co-transformed with the reporter plasmid PyB1a-eGFP into *E. coli* BW25113, which were cultured in LB medium containing 100 mg/L Amp, 50 mg/L Kana at 37 °C and 200 rpm for 12 h. Subsequently, 50 μL of the bacteria was inoculated into 5 mL ZYM5052 medium containing 0.2 g/L Ara, 1 mM IPTG, 100 mg/L Amp, 50 mg/L Kana at 30℃, and cultured at 200 rpm for 24 h to induce expression. Cell density (OD_{600}) and green fluorescence intensity (with 430 nm as the excitation wavelength and 510 nm as the emission wavelength) were measured using a 96-well plate reader.

To test the effect of CRISPRi-mediated gene inhibition on HPP production, pATP and R6K-dCas9-sgRNA were co-transformed into *E. coli* BWΔCD. Colonies of the generated bacteria harboring these two plasmids were grown in 5 mL LB medium containing Chl and Kana for 12 h. Next, 5 μL the bacterial culture was transferred into 5 ml ZYM5052 medium containing 0.2 g/L Ara, 1 mM IPTG, 30 mg/L Chl, 50 mg/L Kana, and cultured at 30℃ until OD_{600} reached 5. The bacteria were collected by centrifugation at 4,000 rpm for 10 min, and resuspended in 10 mL M9 medium containing 20 g/L glucose, 0.2 g/L Ara, 1 mM IPTG, 30 mg/L Chl and 50 mg/L Kana. After further cultivation at 30℃ for 24 h, HPP levels were detected by HPLC.

4.6 High performance liquid chromatography (HPLC) detection

The concentration of HMA was measured using a HPLC system. The mobile phases consisted of 10% methanol, 0.1% formic acid, 90% water and 20 mM PBS buffer, with a flow rate of 0.7 mL/min. The injection volume was set to 10 μL, column temperature at 30°C, and detection wavelength as 210 nm. All HPLC analyses were quantified using an HMA standard curve with an R^2 coefficient over 0.99.

4.7 Sample Preparation

The optical density (OD) value of the induced products was determined using a spectrophotometer.

In the assay of whole-cell catalysis, bacteria were cultured in induction medium for 20 h and the harvested by centrifugation at 4,000 rpm for 10 min. The collected bacteria were washed with PBS buffer (pH 7.0) and suspended in 200 μL of a reaction mixture containing 9 mM HPP and 50 mM HEPES. The final catalysis system should contain bacteria solution with an OD_{600} of 6.

For fermentation, bacteria were induced for 20 h and collected by centrifugation at 4,000 rpm for 10 min, followed by suspension in 10 mL of a reaction mixture, and further cultivation in a shaker at 30°C and 200 rpm. Next, 10 mL fermentation system consisted of 2 mL of 5 × M9, 20 μL of 1 M $MgSO_4$, 1 μL of 1 M $CaCl_2$, 500 μL of 20% glucose, 10 μL antibiotics and double distilled water.

Ultimately, the products of whole-cell catalysis or fermentation were centrifuged at 10,000 g for 10 min. A 100 μL of the supernatant was diluted by 10 times with water and filtered through a 0.22 μm filter.

4.8 eMuta T7-mediated sequential directed evolution of proteins

eMuta T7 and target plasmids were transformed into *E. coli* BW25113 and cultured at 37°C for 12 h. Single colonies on the plates were picked for expansion culture. Next, 1% of the bacterial fluid was inoculated into ZYM5052 medium (ZY medium, 50 × 5052, 50 × M, 1000 × trace elements, 1 M $MgSO_4$, 0.2% arabinose, 1 mM IPTG, 30 mg/L Chloramphenicol, and 50 mg/L Streptomycin) to induce protein expression. Fresh nutrients and antibiotics were added to the culture medium every 24 h, and the bacteria in cycle #6, #9, #12 and #21

(every 4 h constitutes one cycle of induction) were selected to isolate single colonies on plates containing 30 mg/L Chloramphenicol and 50 mg/L Streptomycin. These isolated clones were randomly selected and cultured in liquid LB. Plasmids were extracted and retransformed into *E. coli* DH5α, which were randomly selected for DNA sequencing.

4.9 Modeling and docking

Pymol was employed to construct the HmaS mutant model using the PobR wild-type which was simulated by Alphafold2 as a template. The structure files of the ligands HPP and HMA were obtained from the organic small molecule database Pubchem. (https://pubchem.ncbi.nlm.nih.gov/).
Autodock Vina is used to simulate molecular docking. Autogrid search space coordinates were set as center_x = 0, center_y = 2.454 and center_z = 0. Dimensions of the search space were set as size_x = 20, size_y = 20 and size_z = 20, and exhaustiveness was set at 10. The 10 conformational conditions in a score based on the lowest binding energy were listed as the docking results. The interaction between a small molecule and predicted protein receptors was examined by Ligplus to evaluate the accuracy of the docking. The three-dimensional schematics of the protein and its ligand were portrayed using PyMol Version 2.2.0.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgments

This research was supported by funding from Northeast Forestry University. We sincerely thank the strong support from the College of Life Sciences and the Aulin College, Northeast Forestry University. We are grateful for the experimental platform provided by the university, the guidance of our teachers, and the assistance of our mentors.

References

[1] F.-F. Li, Y. Zhao, B.-Z. Li, J.-J. Qiao, G.-R. Zhao, Engineering Escherichia coli for production of 4-hydroxymandelic acid using glucose-xylose mixture. Microbial Cell Factories 15, 90 (2016). DOI: https://doi.org/10.1186/s12934-016-0489-4

[2] M. Reifenrath, E. Boles, Engineering of hydroxymandelate synthases and the aromatic amino acid pathway enables de novo biosynthesis of mandelic and 4-hydroxymandelic acid with Saccharomyces cerevisiae. Metabolic Engineering 45, 246-254 (2018). DOI: https://doi.org/10.1016/j.ymben.2018.01.001

[3] X. Liu et al., Characterization and directed evolution of propionyl-CoA carboxylase and its application in succinate biosynthetic pathway with two CO2 fixation reactions. Metabolic Engineering 62, 42-50 (2020). DOI: https://doi.org/10.1016/j.ymben.2020.08.012

[4] X. Ding et al., High-level and-yield production of L-leucine in engineered Escherichia coli by multistep metabolic engineering. Metabolic Engineering 78, 128-136 (2023). DOI: https://doi.org/10.1016/j.ymben.2023.06.003

[5] Y. Liang et al., Directed evolution of the PobR allosteric transcription factor to generate a biosensor for 4-hydroxymandelic acid. World Journal of Microbiology & Biotechnology 38, 104 (2022). DOI: https://doi.org/10.1007/s11274-022-03286-5

[6] M. Varadi et al., AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Research 50, D439-D444 (2022). DOI: https://doi.org/10.1093/nar/gkab1061

[7] J. Jumper et al., Highly accurate protein structure prediction with AlphaFold. Nature 596, 583-589 (2021). DOI: https://doi.org/10.1038/s41586-021-03819-2

[8] F. Wu, P. Cao, G. Song, W. Chen, Q. Wang, Expanding the repertoire of aromatic chemicals by microbial production. Journal of Chemical Technology and Biotechnology 93, 2804-2816 (2018). DOI: https://doi.org/10.1002/jctb.5690

[9] S.-S. Choi et al., Cell Factory Design and Culture Process Optimization for Dehydroshikimate Biosynthesis in Escherichia coli. Frontiers in Bioengineering and

Biotechnology 7, 241 (2019). DOI: https://doi.org/10.3389/fbioe.2019.00241

[10] K. Li, M. R. Mikola, K. M. Draths, R. M. Worden, J. W. Frost, Fed-batch fermentor synthesis of 3-dehydroshikimic acid using recombinant Escherichia coli. Biotechnology and Bioengineering 64, 61-73 (1999). DOI: https://doi.org/10.1002/(SICI)1097-0290(19990705)64:1<61::AID-BIT7>3.0.CO;2-G

[11] J. Bongaerts, M. Kramer, U. Muller, L. Raeven, M. Wubbolts, Metabolic engineering for microbial production of aromatic amino acids and derived compounds. Metabolic Engineering 3, 289-300 (2001). DOI: https://doi.org/10.1006/mben.2001.0196

[12] L. Cui et al., A CRISPRi screen in E. coli reveals sequence-specific toxicity of dCas9. Nature Communications 9, 1912 (2018). DOI: https://doi.org/10.1038/s41467-018-04209-5

[13] V. de Bakker, X. Liu, A. M. Bravo, J.-W. Veening, CRISPRi-seq for genome-wide fitness quantification in bacteria. Nature Protocols 17, 252-281 (2022). DOI: https://doi.org/10.1038/s41596-021-00639-6

[14] J. Wang, C. Li, T. Jiang, Y. Yan, Biosensor-assisted titratable CRISPRi high-throughput (BATCH) screening for over-production phenotypes. Metabolic Engineering 75, 58-67 (2023). DOI: https://doi.org/10.1016/j.ymben.2022.11.004

[15] J. Pittard, H. Camakaris, J. Yang, The TyrR regulon. Molecular Microbiology 55, 16-26 (2005). DOI: https://doi.org/10.1111/j.1365-2958.2004.04385.x

[16] A. Rodriguez et al., Constitutive expression of selected genes from the pentose phosphate and aromatic pathways increases the shikimic acid yield in high-glucose batch cultures of an Escherichia coli strain lacking PTS and pykF. Microbial Cell Factories 12, 86 (2013). DOI: https://doi.org/10.1186/1475-2859-12-86

[17] M. Kanehisa, M. Furumichi, Y. Sato, M. Kawashima, M. Ishiguro-Watanabe, KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Research 51, D587-D592 (2023). DOI: https://doi.org/10.1093/nar/gkac963

[18] H. Park, Gene-specific mutagenesis enables rapid continuous evolution of enzymes in vivo. Nucleic Acids Research 49, e32 (2021). DOI: https://doi.org/10.1093/nar/gkac1117

[19] N. M. Gaudelli et al., Programmable base editing of A.T to G.C in genomic DNA without DNA cleavage. Nature 551, 464-471 (2017). DOI: https://doi.org/10.1038/nature24644