ENGINEERING OF THE AROMATIC AMINO ACID BIOSYNTHETIC PATHWAYS IN ESCHERICHIA COLI FOR THE PRODUCTION OF VALUE-ADDED CHEMICALS

by

YUHENG LIN

(Under the Direction of YAJUN YAN)

ABSTRACT

Aromatic amino acids (AAAs) *L*-phenylalanine, *L*-tyrosine and *L*-tryptophan serve not only as the building blocks for protein synthesis, but also as the precursors of secondary metabolites in organisms. A number of AAA-derived metabolites have shown pharmaceutical or industrial importance. However, these molecules usually occur in nature at low abundance, which greatly limits their efficient isolation, production and broad applications. In past decades, the development of metabolic engineering and synthetic biology facilitated the construction of 'microbial chemical factories' that convert renewable carbon sources to various valuable molecules. This dissertation focuses on the engineering of the *E. coli* AAA biosynthetic pathways to achieve the heterologous production of pharmaceutically and industrially important chemicals derived from AAAs or its biosynthetic intermediates. After careful evaluation, we select several important target compounds, including caffeic acid, simple coumarins, salicylic acid, muconic acid and 5-hydroxytryptophan and explore their production in *E. coli*. For this purpose, we designed, reconstituted and optimized their respective artificial biosynthetic pathways, and also conducted the modification of the *E. coli* native shikimate pathway via combinatorial metabolic engineering approaches. This research established powerful microbial platforms for the production of AAA related value-added chemicals, exhibiting great potential for commercialized production.

INDEX WORDS: Metabolic Engineering; Aromatic Amino Acids Biosynthesis; Microbial Production; Biocatalysts; Value-added Chemicals

ENGINEERING OF THE AROMATIC AMINO ACID BIOSYNTHETIC PATHWAYS IN *ESCHERICHIA COLI* FOR THE PRODUCTION OF VALUE-ADDED CHEMICALS

by

YUHENG LIN

BE, China Agricultural University, China, 2007

MS, Graduate University Chinese Academy of Sciences, China, 2010

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2014

ENGINEERING OF THE AROMATIC AMINO ACID BIOSYNTHETIC PATHWAYS IN ESCHERICHIA COLI FOR THE PRODUCTION OF VALUE-ADDED CHEMICALS

by

YUHENG LIN

Major Professor:	Yajun Yan
Committee:	Michael W. W. Adams Jim Kastner Ramaraja Ramasamy

Electronic Version Approved:

Julie Coffield
Interim Dean of the Graduate School
The University of Georgia
December 2014

DEDICATION

I would like to dedicate this dissertation to my beloved grandfather Mr. Wenqing Lin.

谨以此博士论文献给我最亲爱的祖父林文清先生。

ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my major professor Dr. Yajun Yan, who continually and convincingly conveyed a spirit of ambition, perseverance and open-mindedness in regard to research and life. Without his guidance and persistent help this dissertation would not have been possible.

I would also like to thank my committee members, Dr. Michael W. W. Adams, Dr. Jim Kastner and Dr. Ramaraja Ramasamy, who helped me improve my research work and dissertation.

Finally, I would like to thank my family for their long-lasting and selfless support.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

L-phenylalanine, *L*-tyrosine and *L*-tryptophan are the three major aromatic amino acids (AAAs) in nature. They serve not only as the building blocks of proteins, but also as the biosynthetic precursors to secondary metabolites in animals, plants and microorganisms. Unlike animals that obtain the AAAs from foods, plants and microorganisms usually have to synthesize these amino acids by themselves via the aromatic biosynthetic pathway– a bridge between primary and secondary metabolism. So far, a number of AAA-derived secondary metabolites have been shown pharmaceutical and industrial importance. However, these compounds usually occur at low abundance in their native producers, which hampers their efficient isolation, production and broad applications. Also, the complexity in chemistry of these compounds involving chiral centers and regio-specific modifications frequently poses a great challenge to chemical synthesis approaches. Fortunately, the development of metabolic engineering and synthetic biology in combination with the accumulated knowledge in fundamental genetics, biochemistry, bioinformatics and structural biology provides new opportunities for the synthesis of natural and non-natural molecules using microbial systems. Here, we report the microbial synthesis of pharmaceutically and industrially important molecules derived from AAA pathways.

In *E. coli*, the AAA pathways share a common upper part, named shikimate pathway (Figure 1.1). The first committed step of shikimate pathway is the conversion of erythrose 4-phosphate (E4P) and phosphoenolpyruvate (PEP) into 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP). This reversible reaction is catalyzed by three DAHP synthase isoenzymes encoded by *aroF*, *aroH* and *aroG*, which are feed-back inhibited by tyrosine, tryptophan and phenylalanine, respectively. Then, DAHP is sequentially converted to 3-dehydroquinate, 3-dehydroshikimate and shikimate by the action of 3-dehydroquinate synthase, 3-dehydroquinate dehydratase and shikimate dehydrogenase, respectively. Shikimate is an

important intermediate in AAA pathways. Through phosphoralation and reation with another PEP molecule, shikimate is converted to 5-enolpyruvyl-shikimate-3-phosphate; then the latter is further converted into another important intermediate chorismate. Chorismate is the last intermediate within shikimate pathway and also serves as the branching point towards either tryptophan or phenylalanine and tyrosine biosynthesis. The first committed enzyme of each AAA branch (e.g. PheA, TyrA and TrpE) is subject to feed-back inhibition by its respective end product.

The following chapers describe the reconstitution of several artificial biosynthesis pathways in *E.coli* either by extending the AAA pathways or by redirecting the shikimate pathway from chorismate in order to achieve the production of valuable molecules (Figure 1.1). Specifically, in Chapter 2, we explored the production of caffeic acid by extending the tyrosine biosynthesis in *E.coli*. On the basis of Chapter 2, we further extended the caffeic pathway towards the production of plant-specific simple coumarins, umbelliferone and scopoletin in Chapter 3. In Chapter 6, we report the extension of tryptophan pathway. A multi-functional drug 5-hydroxytryptophan (5-HTP) was produced through the introduction of a 5-hydroxylation reaction in combination of a co-factor regeneration mechanism. In Chapter 4 and Chapter 5, we explored the redirection of shikimate pathway from chorismate, leading to the production of an anticoagulant precursor 4-hydroxycoumarin (4HC) and an important platform chemical muconic acid (MA). Both of the pathways share the upper module with salicylic acid as an intermediate. These chapters are relatively independent with each other. Each work involves the design, validation and optimization of artificial biosynthetic pathway, as well as the modification of the *E. coli* native shikimate pathway via combinatorial metabolic engineering approaches. This research established powerful microbial platforms for the production of AAA related value-added chemicals, which has laid the foundation for scale-up production.

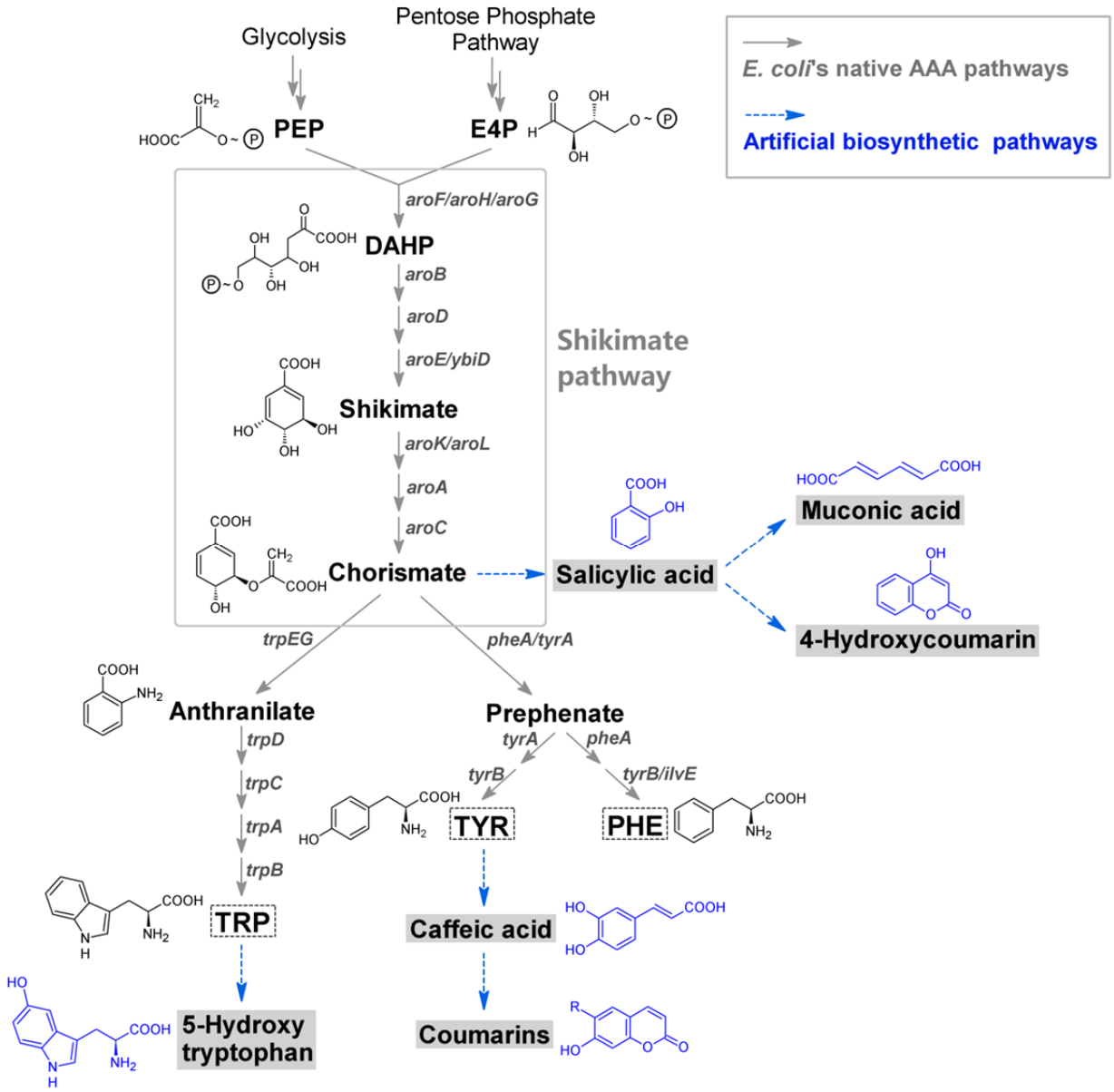


Figure 1.1 An overview of the engineering of aromatic amino acid (AAA) biosynthetic pathways in *E. coli*. Grey-colored arrows indicate native pathways. Blue-colored arrows indicate heterologous pathways.

CHAPTER 2[1]

BIOSYNTHESIS OF CAFFEIC ACIS BY METABOLICALLY ENGINEERED *ESCHERICHIA COLI*

[1] Yuheng Lin and Yajun Yan. 2012, *Microbial Cell Factories.* 11: 42
Reprinted here with permission of the publisher.

2.1 Abstract

Caffeic acid (or 3,4-dihydroxycinnamic acid) is a natural phenolic compound derived from the phenylpropanoid pathway in plants. Caffeic acid and its phenethyl ester (CAPE) have attracted increasing attention for their various pharmaceutical properties and health-promoting effects. Nowadays, large-scale production of drugs or drug precursors through microbial approaches provides a promising alternative to chemical synthesis and extraction from plant sources. In this work, we explored the production of caffeic acid using metabolically engineered *Escherichia coli.* We first identified that an *E. coli* native hydroxylase complex known as the 4-hydroxyphenylacetate 3-hydroxylase (4HPA3H) was able to convert *p*-coumaric acid to caffeic acid efficiently. This finding allowed us to obviate the use of a membrane-associated cytochrome P450 enzyme from plants, *p*-coumarate 3-hydroxylase (C3H), which is hard to be functionally expressed in prokaryotic systems. Moreover, the performances of two tyrosine ammonia lyases (TALs) from *Rhodobacter* species were compared when expressed in *E. coli*. The enzyme assay results indicated that the TAL from *R. capsulatus* (*Rc*) possesses higher activity toward both tyrosine and *L*-dopa. On this basis, we further designed a dual pathway leading from tyrosine to caffeic acid consisting of the enzymes 4HPA3H and *Rc*TAL. This heterologous pathway extended *E. coli*'s native tyrosine biosynthesis machinery and was able to produce caffeic acid (12.1 mg/L) from simple carbon sources. Further improvement in production was accomplished by boosting tyrosine biosynthesis in *E. coli*, which involved the alleviation of tyrosine-induced feedback inhibition and carbon flux redirection. Finally, the titer of caffeic acid reached 50.2 mg/L in shake flasks after 48-hour cultivation. In conclusion, we have successfully established a novel pathway and constructed an *E. coli* strain for the *de novo* production of caffeic acid. This work forms a basis for further improvement in production, and opens the possibility of microbial synthesis of more complex plant secondary metabolites derived from caffeic acid. In addition, we have determined that TAL is the rate-limiting enzyme in this pathway. Thus, exploration for more active TALs through bio-prospecting and protein engineering approaches is necessary for further improvement of caffeic acid production.

2.2 Background

Caffeic acid (3,4-dihydroxycinnamic acid) is a natural phenolic compound initially found in plants. Previous studies on its biological activities suggested that caffeic acid possesses anti-oxidant [1,2], anti-virus [3], anti-cancer [4] and anti-inflammatory properties [5]. Moreover, its derivative, caffeic acid phenethyl ester (CAPE), has drawn great attention because of its demonstrated therapeutic effects including its potential as an anti-diabetic and liver-protective agent as well as an anti-tumor drug for human breast cancer treatment [6,7].

Caffeic acid is one of the pivotal intermediates of plant phenylpropanoid pathway starting from the deamination of phenylalanine which generates cinnamic acid. Followed by a two-step sequential hydroxylation at the 4- and 3- position of the benzyl ring, cinnamic acid is converted into caffeic acid via *p*-coumaric acid [8,9]. The involved enzymes, cinnamate 4-hydroxylase (C4H) and *p*-coumarate 3-hydroxylase (C3H) are plant-specific cytochrome P450 dependent monooxygenases. Due to their instability and membrane-bound property, the purification and characterization of these enzymes are quite challenging, particularly for C3H [10]. It was also suggested that the hydroxylation at the 3-position could also occur after *p*-coumaric acid is esterified, which does not generate caffeic acid as the intermediate [8,11]. Recently, genes and enzymes involved in caffeic acid biosynthesis were also reported in the actinomycete *Saccharothrix espanaensis*. A tyrosine ammonia lyase (TAL) encoded by *sam8* and a microbial C3H encoded by *sam5* are responsible for the conversion of tyrosine to *p*-coumaric acid and then to caffeic acid, respectively [12].

Currently, caffeic acid is commonly produced by extraction from plant sources, such as coffee beans. Chemical or enzymatic hydrolysis of caffeoylquinic acid derivatives is also employed to produce caffeic acid [13,14]. Like many other secondary metabolites, caffeic acid derivatives are usually accumulated at low levels in plants and hence the isolation of these compounds is to some extent difficult and expensive. Microbial conversion provides an alternative approach to caffeic acid production. Sachan *et al.* reported the co-production of caffeic acid and *p*-hydroxybenzoic acid in *Streptomyces caeruleus* by feeding *p*-coumaric acid [15]. Over decades, advances in metabolic engineering and synthetic biology

enable the production of various plant-specific secondary metabolites in recombinant microorganisms [16-19]. Most recently, the conversion of tyrosine to caffeic acid (the titer was not reported) and ferulic acid (7.1 mg/L) in *E. coli* was achieved by the co-expression of the enzymes encoded by the *sam5* and *sam8* from *S. espanaensis* and an *O*-methyltransferase from *Arabidopsis thaliana* [20]. However, the above-mentioned studies relied on feeding the direct precursors such as tyrosine and *p*-coumaric acid, which would increase the production cost and cannot be preferred for large-scale production. Alternatively, the development of processes that can enable the biosynthesis of these high-value metabolites from simple carbon sources is much more desirable. By utilizing tyrosine-overproducing strains as hosts, the production of several natural products such as *L*-dopa, flavonoids, and benzylisoquinoline alkaloids from simple carbon sources has already been achieved [21-23].

In this study, we characterized the *E. coli* native 4-hydroxyphenlacetate 3-hydroxylase (4HPA3H) that was capable of hydroxylating *p*-coumaric acid and tyrosine in addition to its native substrate 4-hydroxyphenylacetic acid. Moreover, we found the TAL from *Rhodobacter capsulatus* was able to accept both tyrosine and *L*-dopa as substrates. Based on these findings, we further designed a novel dual pathway leading from tyrosine to caffeic acid mediated by the enzymes 4HPA3H and TAL. As shown in Figure 2.1, native tyrosine biosynthesis can be extended by the introduction of the 4HPA3H and TAL, yielding *L*-dopa and *p*-coumaric acid, respectively. Then TAL further converts *L*-dopa to caffeic acid; while 4HPA3H converts *p*-coumaric acid into caffeic acid as well. Furthermore, by grafting this dual pathway into *E. coli*, we successfully achieved *de novo* biosynthesis of caffeic acid. This work not only opens the route to the production of caffeic acid from simple carbon sources, but also paves the way to the microbial synthesis of many other phenylpropanoids derived from caffeic acid.

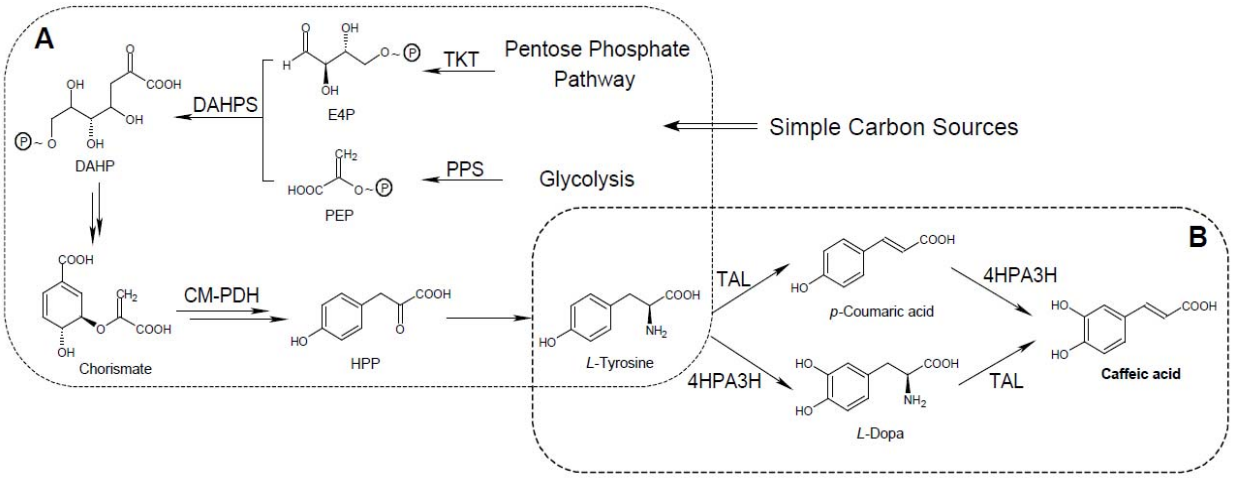


Figure 2.1 Proposed caffeic acid biosynthetic pathway. (A) Native tyrosine biosynthetic pathway in *E. coli.* (B) The artificial dual pathway mediated by 4HPA3H and TAL for caffeic acid biosynthesis from tyrosine. PPS: phosphoenolpyruvate synthase; TKT: transketolase; CM-PDH: chorismate mutase-prephenate dehydrogenase; DAHPS: 3-deoxy-D-arabino- heptulosonate-7-phosphate synthase; 4HPA3H: 4-hydroxyphenylacetate 3-hydroxylase; TAL: tyrosine ammonia lyase; E4P: *D*-erythrose-4-phosphate; PEP: phosphoenolpyruvate; HPP: 4-hydroxyphenylpyruvate.

2.3 Materials and Methods

2.3.1 Chemicals and enzymes

The following commercially available chemicals and enzymes were used in this study: *L*-dopa (ACROS Organics); tyrosine (Sigma-Aldrich), caffeic acid (TCI), *p*-coumaric acid (MP Biochemicals), IPTG (Zymo Research Co.), restriction enzymes (NEB), Hot Start KOD Plus DNA polymerase (EMD Chemicals Inc.), Rapid DNA ligase Kit (Roche). All the enzymes were used according to the instructions provided by the manufacturers.

2.3.2 Strains, plasmids, media, and growth conditions

E. coli XL1-Blue (Stratagene) was used for gene cloning and plasmid propagation. Wild type *E. coli* strain BW25113 (*E. coli* Genetic Resource Center) and its derivatives were employed for either enzyme assays or shake flask experiments. Plasmids pZE12-luc and pCS27 were used for gene over-expression in *E. coli* [24,25]. The characteristics of all the strains and plasmids used in this study are described in Table 2.1. *E. coli* cells for gene cloning, plasmid propagation, and inoculum preparation were grown in Luria-Bertani (LB) medium at 37^{o}C. The fermentation medium was modified M9 minimal salt medium containing (per liter): glycerol (10 g), glucose (2.5 g), NH_4Cl (1 g), Na_2HPO_4 (6 g), KH_2PO_4 (3 g), NaCl (0.5 g), $MgSO_4 \cdot 7H_2O$ (2 mmol), $CaCl_2 \cdot 2H_2O$ (0.1 mmol), vitamin B1 (2.0 mg), H_3BO_3 (1.25 mg), $NaMoO_4 \cdot 2H_2O$ (0.15 mg), $CoCl_2 \cdot 6H_2O$ (0.7 mg), $CuSO_4 \cdot 5H_2O$ (0.25 mg), $MnCl_2 \cdot 4H_2O$ (1.6 mg), and $ZnSO_4 \cdot 7H_2O$ (0.3 mg). For the strains carrying plasmids, 100 μg/ml of ampicillin, 50 μg/ml of kanamycin and/or 30 μg/ml of chloramphenicol were added if necessary. For all shake flask experiments, 200 μl overnight LB culture was inoculated into 10 ml fermentation medium and grown at 37^{o}C with shaking. After OD_{600} reached 0.4-0.5, IPTG was added into the cultures to a final concentration of 0.2 mM. Then the cultures were transferred to 30^{o}C in a gyratory shaker at 250 rpm. Samples were collected after 24 and 48 hours, and then analyzed by HPLC.

2.3.3 Molecular biology techniques

General molecular biology techniques and DNA manipulations were carried out according to the standard protocols [26]. Deletion of kanamycin resistant gene from *E. coli* JW1316-1 was conducted using the method described by Kirill A. Datsenko and Barry L. Wanner [27]. Host cells were transformed with the plasmids by electroporation (EPPENDORF Electroporator 2510, 1.8 kV when using 0.1 cm cuvettes).

Table 2.1 Strains and plasmids used in Chapter 2

Plasmid or Strain	Relevant characteristics	Source
Plasmids		
pZE12-luc	ColE1 ori; Amp^R; $P_L$lacO1; *luc*	[25]
pCS27	p15A ori; Kan^R; $P_L$lacO1; MCS1	[24]
pZE-*Rc*TAL	From pZE12, $P_L$lacO1; *tal(Rc)*	This study
pZE-*Rs*TAL	From pZE12, $P_L$lacO1; *tal(Rs)*	This study
pZE-*Ec*HpaBC	From pZE12, $P_L$lacO1; *hpaB(Ec)-hpaC(Ec)*	This study
pZE-TH	From pZE12, $P_L$lacO1; *tal(Rc)-hpaB(Ec)-hpaC(Ec)*	This study
pCS-TPTA	From pCS27, $P_L$lacO1; $tyrA^{fbr}$-*ppsA-tktA-aroG*^{fbr}	This study
Strains		
XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB* $lacI^qZ$ Δ*M15 Tn10* (Tet^R)]	Stratagene
BW25113	F-, Δ(*araD-araB*), ΔlacZ (::*rrnB-3*), λ-, *rph-1*, Δ(*rhaD-rhaB*), *hsdR*	Yale CGSC
JW1316-1	BW25113, Δ*tyrR*::*kan*	Yale CGSC
YL-1	BW25113, Δ*tyrR*::FRT (as JW1316-1, but kan^R gene deleted)	This study
YL-2	BW25113 harboring pZE-TH	This study
YL-3	BW25113 harboring pCS-TPTA	This study
YL-4	YL-1 harboring pCS-TPTA	This study
YL-5	BW25113 harboring pZE-TH and pCS-TPTA	This study
YL-6	YL-1 harboring pZE-TH and pCS-TPTA	This study

Table 2.2 Primers used in Chapter 2

Plasmid	Gene	Sequence (5'-3')
pZE-*Rc*TAL	*tal(Rc)*	F: gggaaaGGTACCatgctcgatgcaaccatcgg
		R: gggaaaGCATGCtcatgccgggggatcggc
pZE-*Rs*TAL	*tal(Rs)*	F: gggaaaGGTACCatgctcgccatgagccc
		R: gggaaaGCATGCtcagacgggagattgctgcaag
pZE-*Ec*HpaBC	*hpaBC(Ec)*	F: gggaaaGGTACCatgaaaccagaagatttccgcgc
		R: gggaaaGCATGCttaaatcgcagcttccatttccagc
pZE-TH	*hpaBC(Ec)*	F: gggaaaGCATGCaggagatataccatgaaaccagaagatttccgcgccag
		R: gggaaaTCTAGAttaaatcgcagcttccatttccagcatc
	tyrA/Met-53-Ile	F: gcgcgaggcatctattttggcctcgcgtcgtg
		R: cacgacgcgaggccaaaatagatgcctcgcgc
	tyrA/Ala-354-Val	F: ctggttcggcgattacgtgcagcgttttcagagtg
		R: cactctgaaaacgctgcacgtaatcgccgaaccag
	aroG/Asp-146-	F: gcaggtgagtttctcaacatgatcaccccac
	Asn	R: gtggggtgatcatgttgagaaactcacctgc
pCS-TPTA	$tyrA^{fbr}$	F: gggaaaGGTACCatggttgctgaattgaccgcattacg
		R: gggaaaGTCGACgCATATGttactggcgattgtcattcgcctgac
pCS-TPTA	*ppsA*	F: gggaaaCATATGaggagatataccatgtccaacaatggctcgtcac
		R: gggaaaGTCGACttatttcttcagttcagccaggcttaac
pCS-TPTA	*tktA*	F: gggaaaCTCGAGaggagatataccatgtcctcacgtaaagagcttgcc
		R: gggaaaGCATGCttacagcagttcttttgctttcgcaac
pCS-TPTA	$aroG^{fbr}$	F: gggaaaGCATGCaggagatataccatgaattatcagaacgacgatttacgc
		R: gggaaaAAGCTTttacccgcgacgcgcttttac

Capital letters indicate restriction sites; underlines indicate *RBS*

2.3.4 Construction of plasmids

To construct pZE-*Rc*TAL and pZE-*Rs*TAL, the genes encoding *Rc*TAL and *Rs*TAL were amplified by high-fidelity polymerase chain reaction (PCR) from the genomic DNAs of *Rhodobacter capsulatus* and *Rhodobacter sphaeroides* using the primers listed in Table 2.2 [28]. Amplified fragments and pZE12-luc were digested with *Kpn*I and *Sph*I, and then ligated with Rapid DNA ligase. To construct pZE-*Ec*HpaBC, the gene cluster *hpaBC* was amplified from *E. coli* MG1655 genome directly. The amplified *hpaBC* fragment was inserted into pZE12-luc vector between *Kpn*I and *Sph*I as well. The pZE-TH was constructed by cloning the gene cluster *hpaBC* into the pZE-*Rc*TAL using restriction enzymes *Sph*I and *Xba*I. A ribosome binding site is located upstream of each gene to facilitate protein expression. The genes *tyrA*, *aroG, ppsA,* and *tktA* were all amplified from *E. coli* MG1655 genomic DNA. Point mutations were introduced to *tyrA* (Met-53-Ile and Ala-354-Val) and *aroG* (Asp-146-Asn) by splicing and overlapping extension PCR (SOE-PCR), generating $tyrA^{fbr}$ and $aroG^{fbr}$ [29-31]. The genes $tyrA^{fbr}$ and *ppsA* were first cloned into pCS27 simultaneously via three-piece ligation using restriction enzymes *Kpn*I, *Nde*I, and *Sal*I, generating the plasmid pCS-TP. Similarly, *tktA* and $aroG^{fbr}$ were then simultaneously inserted into pCS-TP using restriction enzymes *Xho*I, *Sph*I, and *Hind*III resulting in pCS-TPTA.

2.3.5 4HPA3H *in vitro* assay

The *E. coli* strain BW25113 carrying the plasmid pZE-*Ec*HpaBC was pre-inoculated into LB liquid medium containing 100 μg/ml of ampicillin and grown at 37°C overnight with shaking at 250 rpm. In the following day, 1 ml of preinoculum was added to 50ml of fresh LB medium also containing 100 μg/ml of ampicillin. The culture was left to grow at 37°C till OD_{600} reached 0.6 and then induced with 0.5 mM IPTG. Protein expression was conducted at 30°C for another 3h. The cells were harvested and resuspended in 2 ml of buffer A (20 mM KH_2PO_4, pH=7.0), and then lysed by French Press. The soluble fraction was collected by ultra-centrifugation and used as crude enzyme extract for the enzyme assay. Total protein concentration was estimated using the BCA kit (Pierce Chemicals). The total protein concentration of the crude extract is 617.2 μg/ml. The enzyme activity was assayed according to the

protocol described by Tai. et al. with a few modifications [32]. The 1ml reaction system contained 2 mM NADH, 2 mM FAD, 2 mM substrate (tyrosine or *p*-coumaric acid) and 100 µl of crude enzyme extract in buffer A. The reaction was incubated at 30^{o}C for 1.5 min and terminated by acidification (1% HCl). The amount of products (*L*-dopa and caffeic acid, respectively) were measured and quantified by HPLC.

2.3.6 Whole-cell bioconversion by 4HPA3H

The *E. coli* strain BW25113 carrying the plasmid pZE-*Ec*HpaBC was pre-inoculated into LB liquid medium containing 100 µg/ml of ampicillin and grown at 37^{o}C overnight with shaking at 250 rpm. Then 0.1 ml of preinoculum was added to 10ml of fresh LB medium also containing 100 µg/ml of ampicillin. The culture was grown at 37^{o}C till OD_{600} reached 0.6 and then induced with 0.5 mM IPTG for 3 hours. After that, 100 uL of *p*-coumaric acid (10 g/L) was added to reach a final concentration of 100 mg/L. Samples were collected at 3 h and analyzed by HPLC

2.3.7 4HPA3H *in vivo* assay

The pre-inoculum of *E. coli* strain BW25113 carrying pZE-*Ec*HpaBC from an overnight culture was added in to 10 ml of LB medium (1:100 V/V) and grown at 37^{o}C. IPTG was added to the cultures to a final concentration of 0.5 mM until OD_{600} reached 0.6. The cultures were left at 30^{o}C for around another 3 hours with shaking for protein expression till OD_{600} reached 3.0 (approximately equivalent to 1 g/L cell). Then the cells were collected, washed, resuspended in 10 ml of NaCl (0.9%) solution. 1 mM substrate (tyrosine or *p*-coumaric acid) was added to the cell resuspensions at 30^{o}C. Samples were collected after 1 and 2 hours, and then analyzed by HPLC.

2.3.8 Enzyme assay of *Rc*TAL and *Rs*TAL

The crude enzyme extracts of *Rc*TAL and *Rs*TAL were prepared as described before [28]. But the cells were resuspended in buffer B (50 mM, Tris-HCl, pH=8.5). The 1 ml reaction system contained 2 mM substrate (tyrosine or *L*-dopa) and 100 µl crude extract in buffer B. The reaction was incubated at

30^{o}C for 1.5 min and the amount of products (*p*-coumaric acid and caffeic acid, respectively) was measured by HPLC.

2.3.9 HPLC analysis of products

Tyrosine, *L*-dopa, *p*-coumaric acid, and caffeic acid generated in enzyme assays and fermentations were quantitatively analyzed by HPLC (Dionex Ultimate 3000) with a reverse-phase ZORBAX SB-C18 column and an Ultimate 3000 Photodiode Array Detector. The compounds were separated by elution with a methanol-water gradient (water containing 0.2 % trifluoroacetic acid). The following gradient was used at a flow rate of 1 ml/min: 10 to 50 % methanol for 15min, 50 to 10 % methanol for 1 min, and 10% methanol for an additional 4min. Quantification for the four above-mentioned compounds was based on the peak areas of absorbance at 274, 280, 308 and 323 nm, respectively. The data shown in this study were taken from duplicate or triplicate independent experiments.

2.4 Results and Discussion

Plants and bacteria are very different in cell structure, physiology and genetics. One of the difficulties in reconstructing plant pathways in microbial systems is the availability of functional enzymes that are compatible with the specific microorganism. For the biosynthesis of caffeic acid in plant, two cytochrome P450-dependent monooxygenases are involved, which are C4H and C3H [10]. Due to the requirement for anchorage on endoplasmic reticulum, functional expression of these plant P450-dependent enzymes were always problematic in bacterial systems [10,33]. Fortunately, TALs identified from various sources can catalyze the direct formation of *p*-coumaric acid from tyrosine bypassing the enzymatic step catalyzed by C4H [28], and thus, the need for a C4H was not obligatory. Nevertheless, the need for C3H still remains. Although an alternative microbial C3H was identified from *S. espanaensis,* its activity seems to be low, which limits its applications [20].

2.4.1 *p*-Coumaric acid hydroxylation by 4HPA3H

One of the most challenging steps in reconstructing plant phenylpropanoid pathway in *E. coli* is the 3-hydroxylation of *p*-coumaric acid, because all C3Hs identified in plants are cytochrome P450-dependent monooxygenases and are hard to be functionally expressed in bacterial systems [10]. Therefore, the exploration of alternative enzymes compatible with *E. coli* is necessary. By examining *E. coli*'s native enzymes and pathways related to metabolism of aromatic compounds, we reasoned that the 4HPA3H complex encoded by the operon *hpaBC* involved in the 4-hydroxyphenylacetate (4-HPA) degradation may play the role of C3H [32,34]. This enzyme complex can accept a broad range of substrates and has been applied to produce *L*-dopa and hydroxytyrosol from tyrosine and 4-tyrosol, respectively [23,32,34,35]. Because *p*-coumaric acid is similar to 4-HPA, tyrosine and 4-tyrosol in molecular structure (Figure 2.2), we think that the catalytic pocket of 4HPA3H should be able to accommodate *p*-coumaric acid as well. To test this hypothesis, we first cloned *hpaBC* into a high-copy expression vector pZE12-luc. After over-expressing this enzyme complex in wild type *E. coli* BW25113, crude extract was prepared for *in vitro* enzyme assay. Our results indicate that 4HPA3H complex is capable of converting *p*-coumaric acid to caffeic acid in the presence of FAD (flavin adenine dinucleotide) and NADH (nicotinamide adenine dinucleotide hydride, Table 2.3). Its specific activity toward *p*-coumaric acid (5.37×10^{-3} U/mg protein) is much higher than its activity toward tyrosine (2.44×10^{-3} U/mg protein).

Table 2.3 *In vitro* activity of 4HPA3H complex

Enzyme	Activity toward Substrate		Ratio (A : B)
	A (tyrosine) (10^{-3} U/mg protein)	B (*p*-coumaric acid) (10^{-3} U/mg protein)	
4HPA3H	2.44 ± 0.11	5.37 ± 0.31	0.45

One U (unit) is defined as the amount (1 μmole) of product formed per minute.

Furthermore, we carried out whole-cell conversion studies which reflect the *in vivo* enzymatic activity. BW25113 harboring pZE-*Ec*HpaBC was able to completely convert 100 mg/L *p*-coumaric acid to caffeic acid within 3 hours after the induction of isopropyl *β-D*-1- thiogalactopyranoside (IPTG), indicating the *in vivo* activity toward *p*-coumaric acid is high. Meanwhile, no caffeic acid was detected in the culture of the control strain (BW25113 harboring pZE12-luc) even after 20 hours. This phenomenon suggested that although *hpaBC* exists in the genome of *E. coli*, it is not natively expressed. Thus, over-expression of *hpaBC* is necessary to obtain adequate 4HPA3H activity. The result of *in vivo* enzyme assay showed that the highest conversion rates (within the first hour) from tyrosine to *L*-dopa and from *p*-coumaric acid to caffeic acid are 112.98 and 240.80 μmol· h^{-1}·$gDCW^{-1}$, respectively (Figure 2.3). For both of the products, we did not observe obvious intracellular accumulation. Both *in vitro* and *in vivo* assay results indicate that *p*-coumaric acid is preferred by 4HPA3H. To our knowledge, this is the first report of the 4HPA3H activity toward *p*-coumaric acid.

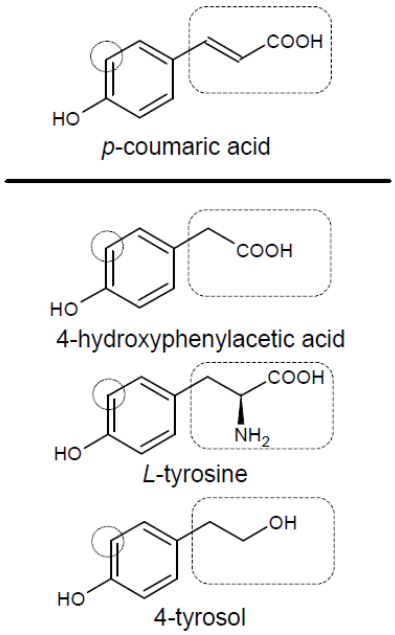


Figure 2.2 Molecular structures of *p*-coumaric acid and three known substrates of 4HPA3H. The circles indicate the hydroxylation positions. The dashed boxes indicate the difference in molecular structures.

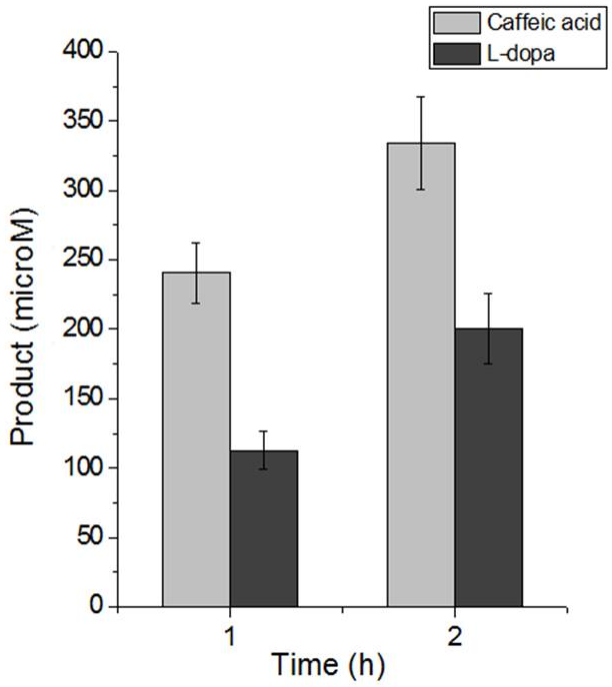


Figure 2.3 *In vivo* enzyme activity of 4HPA3H complex toward tyrosine and *p*-coumaric acid. The grey and black bars refer to the amount of caffeic acid and *L*-dopa, respectively.

2.4.2 Comparison of *Rc*TAL and *Rs*TAL

Previous studies reported that TALs from *R. capsulatus* (*Rc*) and *R. sphaeroides* (*Rs*) catalyze the deamination of tyrosine [28]. In addition, *Rs*TAL can also take *L*-dopa as a substrate [36]. But the activity of *Rc*TAL toward *L*-dopa has not been investigated. To evaluate the performance of the two TALs in *E. coli*, we performed *in vitro* enzyme assays using crude extracts. The genes encoding the two TALs were cloned and expressed in *E. coli* using the plasmids pZE-*Rc*TAL and pZE-*Rs*TAL. Interestingly, both TALs slightly prefer *L*-dopa over their native substrate tyrosine. For *Rc*TAL, the specific activities toward tyrosine and *L*-dopa were 0.93×10^{-3} and 1.54×10^{-3} U/mg protein, respectively. For *Rs*TAL, the specific activities are 0.80×10^{-3} and 1.14×10^{-3} U/mg protein, respectively. The results indicated that *Rc*TAL is

slightly more active than *Rs*TAL toward both substrates (Table 2.4). As a control, the crude extract of the wild-type *E. coli* carrying the blank vector did not exhibit any activity.

Table 2.4 Comparison of *in vitro* activity of *Rc*TAL and *Rs*TAL

Enzyme	Activity toward Substrate*		Ratio (A : C)
	A (tyrosine) (10^{-3} U/mg protein)	C (*L*-dopa) (10^{-3} U/mg protein)	
*Rc*TAL	0.93 ± 0.03	1.54 ± 0.05	0.60
*Rs*TAL	0.80 ± 0.13	1.14 ± 0.03	0.70

***** The crude extract of the wild-type *E. coli* did not show any TAL activity

2.4.3 Production of caffeic acid in *E. coli*

Based on the activities of 4HPA3H and *Rc*TAL, we proposed a novel dual pathway for caffeic acid biosynthesis from tyrosine (Figure 2.1B). Because *E. coli* natively biosynthesizes tyrosine, it is expected that the introduction of *Rc*TAL and 4HPA3H can result in the biosynthesis of caffeic acid by utilizing *E. coli*'s endogenous tyrosine (Figure 2.1). To achieve this goal, the genes encoding *Rc*TAL and 4HPA3H were amplified and consecutively cloned into a high-copy-number plasmid pZE12-luc under the control of a strong IPTG-inducible promoter $P_L$lacO1, generating the plasmid pZE-TH. A ribosome binding site (RBS) was placed upstream of each gene. Strain YL-2 was developed by introducing pZE-TH into wild type *E. coli* strain BW25113 to test this pathway. The production of caffeic acid was carried out in shake flasks using modified M9 minimal salt medium as described in "Methods and Materials". High performance liquid chromatography (HPLC) analysis of the fermentation samples showed that the retention time (10.1 min) and UV profile of the product were identical to those of the caffeic acid standard, confirming that caffeic acid was produced (Figure 2.4). The strain YL-2 was able to produce

11.1 ± 1.1 mg/L caffeic acid after 24 hours, without obvious accumulation of intermediates including tyrosine, *p*-coumaric acid and *L*-dopa. 48-hour cultivation did not lead to a great increase in caffeic acid production (12.1 ± 0.3 mg/L) (Table 2.5). However, *L*-dopa was accumulated at a concentration of 7.4 ± 0.2 mg/L.

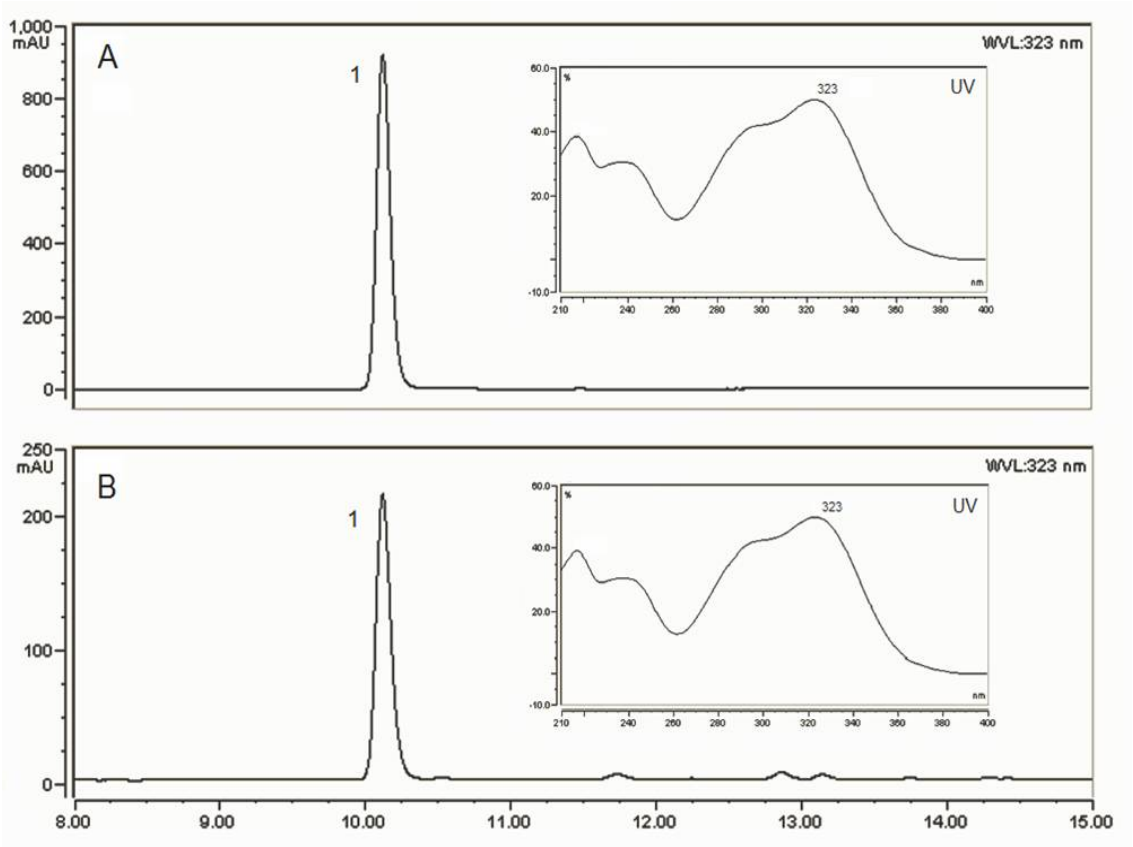


Figure 2.4 HPLC analysis of caffeic acid produced by engineered *E. coli* (A) Standard, 50 mg/L caffeic acid. (B) A sample taken from the fermentation culture of YL-2 after 24 hours. Peak 1 corresponded to caffeic acid. The retention time was 10.1 min. UV absorbance profiles are shown beside the peaks.

Table 2.5 Production of caffeic acid and tyrosine by engineered *E. coli* strains

Strain	24 hours				48 hours			
	Product (mg/L)	Intermediates (mg/L)			Product (mg/L)	Intermediates (mg/L)		
	caffeic acid	tyrosine	coumaric acid	*L*-dopa	caffeic acid	tyrosine	coumaric acid	*L*-dopa
YL-2	11.1±1.1	< 0.2	< 0.2	< 0.2	12.1±0.3	< 0.2	< 0.2	7.4±0.2
YL-2*	20.2 ±1.8	< 0.2	< 0.2	2.1±0.3	21.5 ± 4.0	< 0.2	< 0.2	14.2±1.7
YL-3	–	188.9±5.7	–	–	–	296.6±1.0	–	–
YL-4	–	218.6±8.2	–	–	–	426.7±4.9	–	–
YL-5	30.5±1.9	17.1±1.8	2.9±0.3	15.4±1.7	50.2±10.1	25.1 ± 2.5	< 0.2	75.3±13.6

–: No production

*: addition of 100 mg/L tyrosine in the cultures

2.4.4 Construction of tyrosine overproducers

The addition of 100 mg/L tyrosine into the cultures of YL-2 resulted in a two-fold increase in caffeic acid production (Table 2.5), suggesting that tyrosine is a limiting precursor for caffeic acid biosynthesis. In wild-type *E. coli*, tyrosine biosynthesis is strictly controlled by several regulatory mechanisms. Two feedback inhibition-sensitive enzymes chorismate mutase-prephenate dehydrogenase (CM-PDH, encoded by *tyrA*) and 3-deoxy-*D*-arabino-heptulosonate-7-phosphate synthase (DAHPS, encoded by *aroG*) were identified as the major regulatory components in the tyrosine pathway [37]. The feedback inhibition resistant (fbr) variants $tyrA^{fbr}$ *and* $aroG^{fbr}$ have already been developed and applied in tyrosine production [37,38]. In addition, the availability of erythrose-4-phosphate (E4P) and phosphoenolpyruvate (PEP) is extremely critical to tyrosine biosynthesis (Figure 2.1A). Over-expression of PEP synthase (PPS, encoded by *ppsA*) and transketolase (TKT, encoded by *tktA*) was able to increase

the availability of PEP and E4P, and redirect the carbon flux into the tyrosine pathway [38]. In this work, *tyrA*^{fbr}, *ppsA*, *tktA* and *aroG*^{fbr} were consecutively cloned into a medium-copy-number plasmid pCS27 under the control of $P_L$lacO1 promoter as well, generating the plasmid pCS-TPTA. By introducing pCS-TPTA into wild type *E. coli* BW25113, we obtained a recombinant strain YL-3. Compared with wild type strain which produced little tyrosine, YL-3 was able to produce 296.6 ± 1.0 mg/L tyrosine in 48 hours, which indicated that over-expression of the four enzymes was effective. Furthermore, the strain YL-1 (Δ*tyrR*) was also employed as the host to alleviate the *tyrR*-mediated regulation [23]. The introduction of pCS-TPTA into YL-1 (yielding strain YL-4) resulted in the accumulation of higher amount of tyrosine (426.7 ± 4.9 mg/L in 48h, Table 2.5). It should be noted that YL-4 exhibited only slight improvement in tyrosine production compared to YL-3 in the first 24 hours. However, its advantage was demonstrated in the following 24 hours. These results are consistent with what were reported previously [37,38].

2.4.5 Improvement of caffeic acid production with tyrosine overproducing strains

Although YL-4 was able to produce higher amount of tyrosine, this *tyrR*-deleted strain seemed to be in conflict with pZE-derived plasmids for unknown reasons and did not express the enzymes 4HPA3H and *Rc*TAL as well as expected. Only a trace amount of caffeic acid (<0.2 mg/L) and *p*-coumaric acid (<1 mg/L) but a large amount of tyrosine (>400 mg/L) were detected in the YL-6 (YL-1 harboring pZE-TH and pCS-TPTA) cultures. Thus, we employed wild type *E. coli* BW25113 as the parent strain. By transforming it with both pZE-TH and pCS-TPTA, we generated the strain YL-5. The titer of caffeic acid in the shake flask fermentation using YL-5 reached 50.2 ± 10.1 mg/L after 48h fermentation which is a 5-fold increase compared to YL-2. Moreover, we analyzed the intermediates accumulated in the culture. The presence of 25.1 ± 2.5 mg/L tyrosine indicated that tyrosine availability is no longer the limiting factor for caffeic acid production in the strain YL-5. The accumulation of a large amount of *L*-dopa (75.3 ± 13.6 mg/L) and a small amount of coumaric acid (< 0.2 mg/L) suggested that *Rc*TAL became into the rate-limiting step in this artificial pathway, especially after 24 h (Table 2.5).

2.5 Summary

We have successfully established a novel pathway and constructed an *E. coli* strain for the *de novo* production of caffeic acid via metabolic engineering approaches. We first identified that 4HPA3H can function as a C3H which exhibited decent activity toward *p*-coumaric acid and tyrosine, thus gains great potential for metabolic engineering and biocatalysis applications. In addition, we compared the TALs from *R. capsulatus* (*Rc*) and *R. sphaeroids* (*Rs*) that are able to catalyze the deamination of both tyrosine and L-dopa. RcTAL exhibited higher activities toward both substrates. Then a dual pathway leading from tyrosine to caffeic acid was proposed and introduced into *E. coli*. The artificial pathway extended the native tyrosine pathway of *E. coli* and produced 12.1 mg/L of caffeic acid from simple carbon sources. Further improvement of production was accomplished via alleviating feedback inhibition and redirecting carbon flux into tyrosine biosynthesis. Finally, the titer of caffeic acid reached 50.2 mg/L in shake flasks after 48-hour cultivation.

The established pathway obviated the use of two cytochrome P450-dependent monooxygenases (C4H and C3H) and achieved the *de novo* biosynthesis of caffeic acid, which opened the possibility of microbial synthesis of more complex plant secondary metabolites derived from caffeic acid. However, for the production system to be more economically viable, productivity has to be further improved. We have identified that *Rc*TAL is the rate-limiting enzyme in the pathway once the tyrosine availability issue was solved. To meet the process metrics and avoid the accumulation of the intermediates (tyrosine and *L*-dopa), we will explore more TALs for higher catalytic activity via bioprospecting and protein engineering approaches. In addition to the tyrosine overproducers we generated in this study, the strains employed to produce tyrosine in amino acid industry are also ready to be used as hosts for caffeic acid production. With proper process optimization, industrially relevant production should be expectable.

CHAPTER 3

COMBINATORIAL BIOSYNTHESIS OF PLANT-SPECIFIC SIMPLE COUMARINS IN *ESCHERICHIA COLI*[2]

[2] Yuheng Lin, Xinxiao Sun, Qipeng Yuan and Yajun Yan. 2013, *Metabolic Engineering*, 18: 69-77
Reprinted here with permission of the publisher

3.1 Abstract

Coumarins are plant secondary metabolites that have demonstrated a variety of important therapeutic properties, such as antibacterial, anti-inflammatory, and anti-coagulant effects, as well as anti-cancer and anti-AIDS activities. However, knowledge regarding their biosynthesis is relatively limited even for the simplest coumarin molecule, which serves as the gateway molecule to many pharmaceutically-important coumarin derivatives. Here we reported the design and validation of artificial pathways leading to the biosynthesis of plant-specific simple coumarins in bacteria. First, *Escherichia coli* strains were engineered to convert inexpensive phenylpropanoid acid precursors, 4-coumarate and ferulate to simple coumarins, umbelliferone (4.3 mg/L) and scopoletin (27.8 mg/L), respectively. Further, we assembled the complete artificial pathways in *E. coli* and achieved *de novo* biosynthesis of umbelliferone and scopoletin without addition of precursors. This study lays the foundation for microbial production of more diverse coumarin compounds.

3.2 Background

Secondary metabolites generated by plants represent a major source of natural products that play important roles in plant biochemistry and physiology and demonstrate a variety of pharmaceutical properties. They have been a reliable source for drug discovery and drug development, which is evidenced by the long history of herbal medicines and the large proportion of the drugs of plant origin in the current market [39]. However, plant-based drug discovery and drug development are still under-explored due to the issue of low availability. Isolation and purification of target molecules to establish creditable libraries of pure plant natural products for screening and trials confront the problems of low abundance and season or region-dependent sourcing. On the other hand, the complex chemistry of plant natural products frequently involving multiple chiral centers and delicate regio- and stereo-specific modifications also poses a great challenge to chemical synthesis and limits its success [39].

The convergence of metabolic engineering and fundamental plant biochemistry provides new opportunities and great potential to tackle these problems by allowing reassembly and optimization of

plant-specific biosynthetic pathways in microbial systems to achieve substantial production of desirable products from inexpensive precursors or even simple carbon sources. Recent efforts in this field have led to the biosynthesis of terpenoids, alkaloids, and phenylpropanoids in microbes such as *Escherichia coli* and *Saccharomyces cerevisiae* [39]. Although fundamental research of functional genomics and plant biochemistry will continue playing primary roles in deciphering plant natural product biosynthesis, the repertoire can also be readily expanded by combinatorial biosynthesis and retro-biosynthesis based on the massive and scattered enzyme data that has been accumulated so far [39]. In this paper, we demonstrate an effort of developing a microbial platform to achieve combinatorial biosynthesis of plant-specific simple coumarins of which the biosynthetic pathways are still incomplete [40].

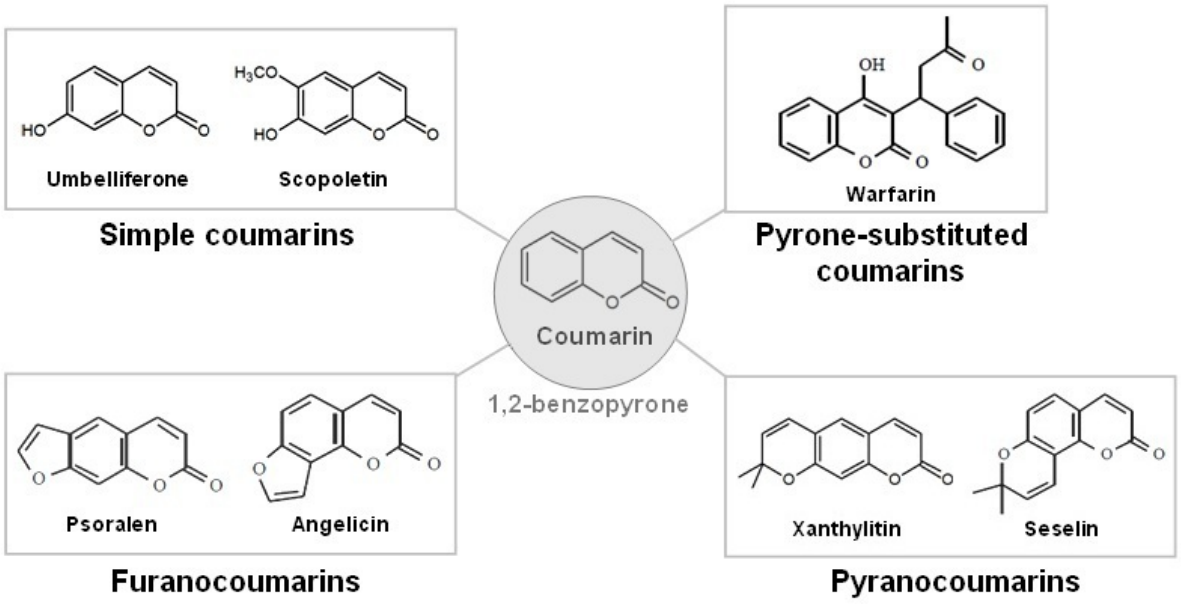


Figure 3.1 Categories of coumarin compounds.

As a major group of plant natural products, coumarins (1, 2-benzopyrones) derived from the phenylpropanoid pathway can be classified into four categories: simple coumarins, pyrone-substituted coumarins, furanocoumarins, and pyranocoumarins (Figure 3.1). Simple coumarins consist of the simplest coumarin compound and its hydroxylated, alkoxylated, alkylated, and glycosylated derivatives. Pyrone-

substituted coumarins contain substitutions and modifications on the pyrone ring. Furanocoumarins have a typical molecular structure of a furan ring attached to the coumarin nucleus. According to the position of the furan ring, furanocoumarins have both linear and angular types. Pyranocoumarins, however, contain a pyran ring and have two types similar to the furanocoumarins [41]. Coumarins and their derivatives have demonstrated a vast array of therapeutic applications including antibacterial agents, anti-inflammation, central nervous system stimulants, and anti-coagulants [39]. For instance, the synthetic derivatives of 4-hydroxycoumarin such as warfarin, phenprocoumon, and acenocoumarol are among the most widely-prescribed anticoagulant pharmaceuticals for the prevention and treatment of thrombosis and embolism. Moreover, extensive research into their pharmacological and therapeutic properties has resulted in the acknowledgment of their therapeutic roles in the treatment of cancers and AIDS [39]. For example, coumarin and umbelliferone (7-hydroxycoumarin) have been reported to have growth-inhibitory effects on human cancer cell lines including A549 (lung), ACHN (renal), H727 (lung), MCF-7 (breast) and HL-60 (leukemia) [39], and have also shown activity against malignant melanoma, metastatic renal cell carcinoma, and prostate cancer evidenced by clinical trials [39]. In addition, furanocoumarins and pyranocoumarins can be applied to anti-tumor and anti-HIV therapy [42]

Despite the pharmaceutical importance of coumarins, comparatively little information is available regarding the biosynthesis of even the simplest coumarin molecular skeleton, which serves as the gateway molecule to many other coumarin derivatives and is unquestionably a valuable starting point for further drug discovery. The general scheme of the coumarin biosynthesis was inferred several decades ago, which suggested the biosynthesis of the basic coumarin molecular skeleton starts from *L*-phenylalanine via the intermediates *trans*-cinnamate, *trans*-2-coumarate, *trans*-2-coumarate-*β*-*D*-glucoside, *cis*-2-coumarate-*β*-*D*-glucoside, and *cis*-2-coumarate (Figure 3.2) [43]. The enzymatic reaction of each step in the scheme was also confirmed by experimental evidences. The enzyme, phenylalanine ammonia-lyase (PAL) has been identified to convert *L*-phenylalanine into *trans*-cinnamate [44]. Then a membrane-bound P450 enzyme, cinnamate 2-hydroxylase (C2H) catalyzes hydroxylation at the 2-position of *trans*-cinnamate yielding *trans*-2-coumarate [45]. At the next step**,** a 2-coumarate *O*-*β*-glucosyltransferase

(2GT) transfers the glucosyl group from UDP-glucose to *trans*-2-coumarate generating *trans*-2-coumarate-*β*-*D*-glucoside [46]. The *trans*-2-coumarate-*β*-*D*-glucoside undergoes *trans/cis* isomerization yielding *cis*-2-coumarate-*β*-*D*-glucoside, which was suggested to occur photochemically. However, the existence of the corresponding light-induced isomerase is also possible [47]. A *β*-glucosidase (GBA) has been assigned to hydrolysate *cis*-2-coumarate-*β*-*D*-glucoside specifically releasing *cis*-2-coumarate and shows no activity towards the *trans*-isomer. The last step of the pathway is the spontaneous lactonization of *cis*-2-coumarate forming coumarin [48]. This pathway is responsible for coumarin biosynthesis in general, but a large challenge that has yet to be overcome is the fact that the genes encoding the enzymes C2H, 2GT, and GBA are still unidentified. Therefore, it is impossible to engineer the coumarin biosynthesis based on this incomplete pathway.

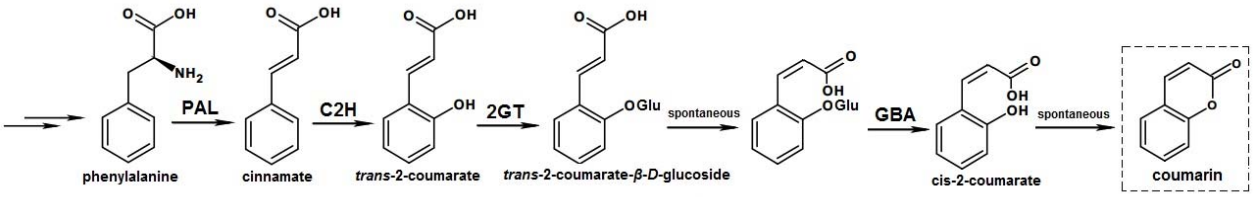


Figure 3.2 Previously proposed coumarin biosynthetic pathway in plants. PAL, phenylalanine ammonia lyase; C2H, cinnamate 2-hydroxylase; 2GT, 2-coumarate *O*-*β*-glucosyltransferase, GBA, *β*-glucosidase.

Most recently, several enzymes from *Arabidopsis thaliana*, *Ipomoea batatas* (L.) *Lam,* and *Ruta graveolens L*. were characterized, which catalyzed coumarin formation reactions. However, the enzymatic reactions do not fit into the known coumarin pathway described in Figure 3.2. In 2008, *Kai et al.* identified a Fe (II)- and 2-oxoglutarate-dependent dioxygenase (2OGD), feruloyl-CoA 6'-hydroxylase (F6'H1) from *A. thaliana* [39]. The coding sequence was functionally expressed in *E. coli*. The recombinant enzyme F6'H1 exhibited specific hydroxylase activity towards feruloyl-CoA generating 6'-hydroxyferuloyl-CoA. The presence of the CoA group with the *ortho*-hydroxyl group led to spontaneous

trans/cis isomerization and lactonization reactions, yielding a simple coumarin, scopoletin (7-hydroxy-6-methoxycoumarin) (Figure 3.3). Only trace activity of F6'H1 toward 4-coumaroyl-CoA was detected; while no activity was detected toward caffeoly-CoA, cinnamoyl-CoA, ferulate, caffeate, 4-coumarate, and cinnamate. With the coding sequence of the F6'H1 becoming available, more homologs were identified and obtained from sweet potato and rue based on bioinformatic analysis. All the homologs exhibited activity towards feruloyl-CoA; while one homolog from rue and three homologs from sweet potato (designated as C2'Hs) were found to be able to hydroxylate 4-coumaroyl-CoA efficiently, synthesizing the simple coumarin, umbelliferone (7-hydroxycoumarin) (Figure 3.3) [39]. These results indicate that these 2OGDs may either play less dominant roles in coumarin biosynthesis or only possess side-activity towards coumarin formation. However, engineering of coumarin biosynthesis can be achieved based on these reactions. Here we presented the pathway design and validation affording the bioconversion of phenylpropanoid acids into two simple coumarins, umbelliferone and scopoletin and further their *de novo* biosynthesis in *E. coli* for the first time.

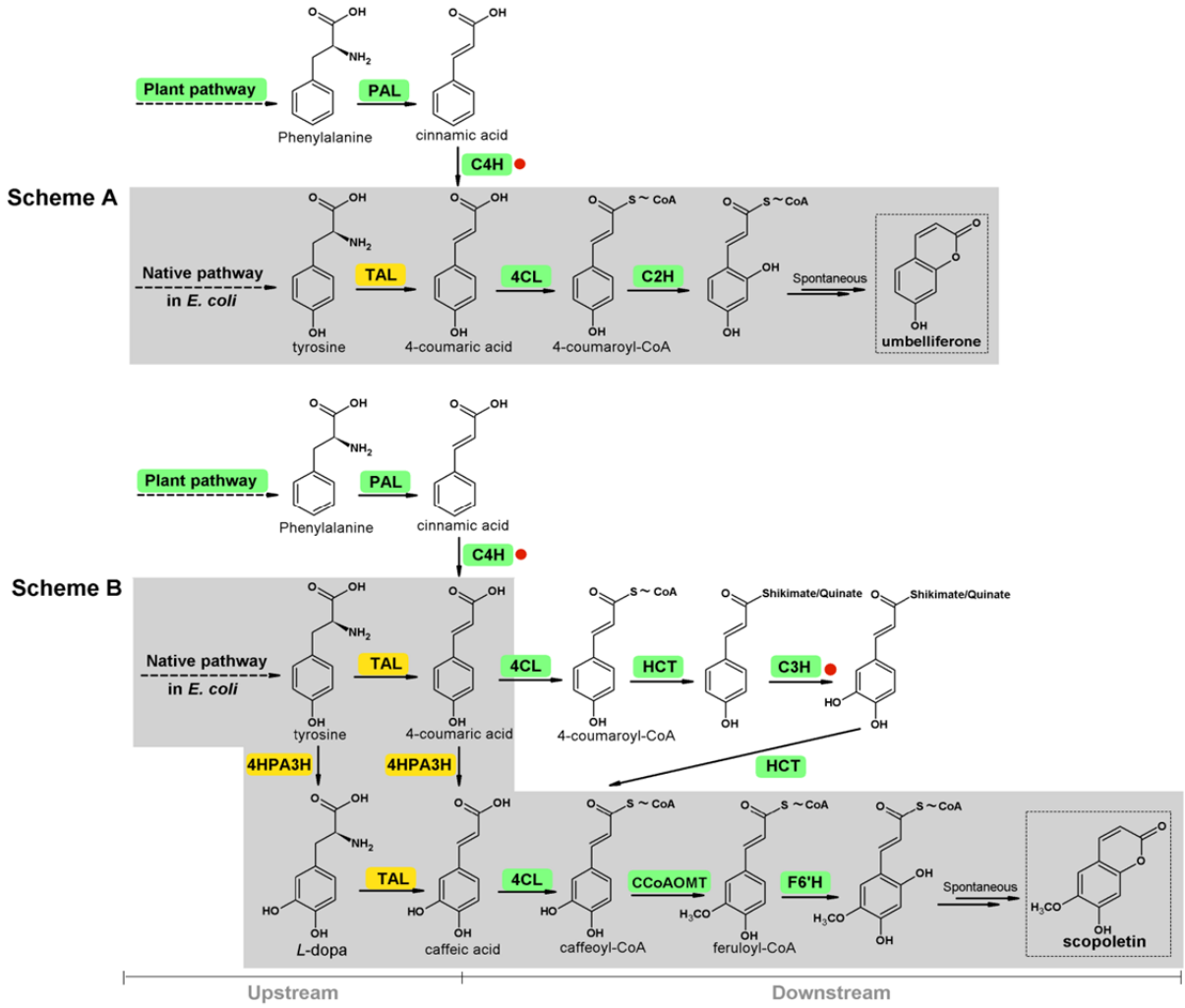


Figure 3.3 The artificial pathways and the proposed plant biosynthetic routes of umbelliferone (scheme A) and scopoletin (scheme B), based on C2'H and F6'H catalyzed reactions. The grey-colored boxes indicate the artificial biosynthetic pathways designed in this study. Enzymes in green-colored boxes are plant-originated; while the enzymes in yellow-colored boxes are originated from microorganisms including bacteria and yeast. Red-dotted enzymes indicate the cytochrome 450 dependent hydroxylases. PAL, phenylalanine ammonia lyase; TAL, tyrosine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate:CoA ligase; C2'H, coumaroyl-CoA 2'-hydroxylase; HCT, hydroxycinnamoyl-CoA: shikimate/quinate hydroxycinnamoyltransferase; 4HPA3H, 4-hydroxyphenylacetate 3-hydroxylase; C3H, coumarate 3-hydroxylase; CCoAOMT, Caffeoyl-CoA O-methyltransferase; F6'H. feruloyl-CoA 6'-hydroxylase.

3.3 Materials and Methods

3.3.1 Strains and plasmids

E. coli XL1-Blue (from Stratagene) was employed as the host for plasmid propagation and gene cloning. *E. coli* BW25113 (from *E. coli* Genetic Resource Center) was used for protein expression, enzyme assays, and shake flask experiments. The characteristics of all the strains and plasmids used in this study were detailed in Table 3.1.

Table 3.1 Strains and plasmids used in Chapter 3

Plasmid and Strain	Characteristics	Source
Plasmid		
pZE12-luc	ColE1 ori; Amp^R; $P_L$lacO1; *luc*	[49]
pCS27	p15A ori; Kan^R; $P_L$lacO1; MCS1	[49]
pZE-RgTAL	From pZE12, $P_L$lacO1; *RgTAL*	This study
pZE-C2'H-Pc4CL2	From pZE12, $P_L$lacO1; *C2'H-Pc4CL2*	This study
pZE-F6'H1-Pc4CL2	From pZE12, $P_L$lacO1; *F6'H1-Pc4CL2*	This study
pZE-Pc4CL2	From pZE12, $P_L$lacO1; *Pc4CL2*	This study
pZE-At4CL1	From pZE12, $P_L$lacO1; *At4CL1*	This study
pZE-At4CL2	From pZE12, $P_L$lacO1; *At4CL2*	This study
pZE-RgTC4	From pZE12, $P_L$lacO1; *RgTAL-C2'H-At4CL1*	This study
pZE-4FO	From pZE12, $P_L$lacO1; *At4CL2-F6'H1-CCoAOMT1*	This study
pZE-RgTH	From pZE12, PLlacO1; *RgTAL-hpaBC*	This study
pZE-RgTH-4FO	From pZE12, dual operons, both $P_L$lacO1; *RgTAL-hpaBC* and *At4CL2-F6'H1-CCoAOMT1*	This study
pCS-TPTA	From pCS27, $P_L$lacO1; $tyrA^{fbr}$-*ppsA-tktA-aroG*^{fbr}	[49]

Strain		
XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacI^{q}Z* Δ*M15 Tn10* (Tet^R)]	Stratagene
BW25113	F-, Δ(*araD-araB*), ΔlacZ (::*rrnB-3*), λ-, *rph-1*, Δ(*rhaD-rhaB*), *hsdR*	Yale CGSC

3.3.2 DNA manipulations

The plasmids pZE12-luc and pCS27 were employed for gene cloning, protein expression and pathway assembly in this work [39]. The RgTAL cDNA was synthesized with codon optimization for optimal expression in *E. coli* (Eurofins MWG Operon, AL)[21]. Pc4CL2 cDNA (GenBank Accession Number X13325) is a generous gift from Dr. Koffas group at Rensselaer Polytechnic Institute. At4CL1 (GenBank Accession Number AEE32699), At4CL2 (GenBank Accession Number AEE76480), F6'H1 (GenBank Accession Number NP187970), and CCoAOMT1 (GenBank Accession Number Q9C5D7) cDNAs were purchased from Arabidopsis Biological Resource Center (ABRC). C2'H (Ib2-1-1) (GenBank Accession Number BAL22347) cDNA was synthesized in our lab by PCR-based oligonucleotide assembly. The gene *HpaBC* was amplified from *E. coli* BW25113 genomic DNA. For the TAL enzyme assay, we created the plasmid pZE-RgTAL by subcloning the RgTAL cDNA into pZE12-luc using the restriction enzymes *Kpn* I and *Xba* I. For the 4CL enzyme assay, we subcloned Pc4CL2, At4CL1, and At4CL2 cDNAs into the plasmid pZE12-luc using *Kpn* I and *Sph* I, respectively. The plasmids pZE-F6'H1-Pc4CL2, pZE-C2'H-Pc4CL2, and pZE-4FO were constructed for step-wise validation of the pathways. More specifically, the F6'H1 and Pc4CL2 cDNAs were cloned as an operon into pZE12-luc via three-piece ligation using *Kpn* I, *Nde* I, and *Xba* I, yielding the plasmid pZE-F6'H1-Pc4CL2. The plasmid pZE-C2'H-Pc4CL2 was developed by the same cloning strategy using *Kpn* I, *Nde* I, and *Xba* I as well. To create the plasmid pZE-4FO, the At4CL2, F6'H1, and CCoAOMT1 cDNAs were simultaneously inserted into the plasmid pZE-luc as an operon using the restriction enzymes *Kpn* I, *Sph* I,

Nde I, and *Xba* I. In order to achieve *de novo* biosynthesis of the simple coumarins, the plasmids pZE-RgTH, pZE-TC4, pZE-RgTH-4FO were generated. Similarly, the plasmid pZE-RgTH carrying the RgTAL cDNA and the gene *HpaBC* as an operon was constructed by three-piece ligation using *Kpn* I, *Sph* I, and *Xba* I. The RgTAL, C2'H, and At4CL1 cDNAs were also cloned into the plasmid pZE-luc as an operon via four-piece ligation using the restriction enzymes *Kpn* I, *Sph* I, *Nde* I, and *Xba* I. The plasmid pZE-RgTH-4FO harboring two operons, RgTAL-HpaBC and At4CL2-F6'H1-CCoAOMT1 were assembled as described by Gilson [39] using the plasmids pZE-RgTH and pZE-4FO as the templates and the plasmid pZE12-luc as the backbone. The plasmid pCS-TPTA that we previously generated for caffeate biosynthesis was also used in this study [49]. PCR reactions were performed using Phusion high fidelity DNA polymerase (from New England Biolabs). The restriction enzymes and quick ligase used in this study were also purchased from New England Biolabs. The primers used in this study are listed in Table 3.2.

Table 3.2 Primers used in Chapter 3

Plasmid	**Coding Enzyme**	**Sequence (5'-3')**
pZE-RgTAL	*RgTAL*	F: gggaaaGGTACCatggcgcctcgcccgacttc
		R: gggaaaTCTAGAttatgccagcatcttcagcagaacattg
pZE-C2'H-Pc4CL2	*C2'H*	F: gggaaaGGTACCatgccgagcaccaccctgagcac
		R: gggaaaCATATGtcattcaatacgcgcaaaggcaatcg
	Pc4CL2	F: gggaaaCATATGaggagatataccatgggagactgtgtagcaccc
		R: gggaaaTCTAGAttatttgggaagatcaccggatgc
pZE-F6'H-Pc4CL2	*F6'H1*	F: gggaaaGGTACCatggctccaacactcttgacaac
		R: gggaaaCATATGtcagatcttggcgtaatcgacgg
pZE-Pc4CL2	*Pc4CL2*	F: gggaaaGGTACCatgggagactgtgtagcacccaaag
		R: gggaaaGCATGCttatttgggaagatcaccggatgc

pZE-At4CL1	*At4CL1*	F: gggaaaGGTACCatggcgccacaagaacaagcag
		R: gggaaaGCATGCtcacaatccatttgctagttttgccctc
pZE-At4CL2	*At4CL2*	F: gggaaaGGTACCatgacgacacaagatgtgatagtcaatg
		R: gggaaaGCATGCctagttcattaatccatttgctagtcttgc
pZE-RgTC4	*RgTAL*	F: gggaaaGGTACCatggcgcctcgcccgacttc
		R: gggaaaGCATGCatgccagcatcttcagcagaacattg
	C2'H	F:gggaaaGCATGCaggagatataccatgccgagcaccaccctg
		R: gggaaaCATATGtcattcaatacgcgcaaaggcaatcg
	At4CL1	F: gggaaaCATATGaggagatataccatggcgccacaagaacaagcag
		R: gggaaaTCTAGAtcacaatccatttgctagttttgccctc
pZE-RgTH-4FO	*HpaBC*	F: gggaaaGCATGCaggagatataccatgaaaccagaagatttccgcgccag
		R: gggaaaTCTAGAttaaatcgcagcttccatttccagcatc
	At4CL2	F: gggaaaGGTACCatgacgacacaagatgtgatagtcaatg
		R: gggaaaGCATGCctagttcattaatccatttgctagtcttgc
	F6'H 1	F: gggaaaGCATGCaggagatataccatggctccaacactcttgacaac
		R: gggaaaCATATGtcagatcttggcgtaatcgacgg
	CCoAOMT1	F: gggaaaCATATGaggagatataccatggctaaggatgaagccaagg
		R: gggaaaTCTAGAtcaatataacctcctgcaaatagtgatacc
DNA assembly	*RgTAL-HpaBC*	F: gcgagcggtatcagctcactcaaggcgtatcacgaggcccttc
		R: ctgtggataaccgtattaccgcctttagggcggcggatttgtcctac
	At4CL2-F6'H-	F: aaggcggtaatacggttatccacag
	CCoAOMT	R: tgagtgagctgataccgctcgc

Capital letters indicate restriction sites

3.3.3 Enzyme assays

To evaluate the activity of 4CLs, the *E. coli* strain BW25113 was transformed with the 4CL-expressing plasmids respectively. The obtained transformants were pre-inoculated in Luria-Bertani (LB) medium containing 100 μg/ml of ampicillin and grown aerobically at 37 ^{o}C overnight. In the following day, the pre-inoculums were transferred into 50 ml of fresh LB medium at a volume ratio of 1:100. The cultures were left to grow at 37 ^{o}C till OD_{600} reached 0.6-0.8 and then induced with 0.5 mM IPTG. Protein expression was conducted at 30 ^{o}C for another 3 h. The cells were harvested and concentrated 25 times in Tris-HCl buffer (500 mM, pH=7.8), and then lysed by French Press. The soluble fractions were collected by ultra-centrifugation and used as crude extracts for enzyme assays. The BCA kit (from Pierce Chemicals) was used to estimate total protein concentrations. The 4CL enzyme assays were performed according to the method reported by Ehlting et al with modifications [50]. The spectrophotometric assays to detect the formation of CoA esters were performed at 30 ^{o}C for 30 s. The 1 ml reaction system contains Tris-HCl (400 mM, pH = 7.8), $MgCl_2$ (5 mM), ATP (5 mM), CoA (0.3 mM), crude enzyme (10 μl for At4CL1 and At4CL2; 30 μl for Pc4CL2) and different amount of substrates (4-coumarate, caffeate, and ferulate). The substrate concentrations varied from 0 to 100 μM. Enzyme kinetic parameters were obtained by non-linear curve fitting with Origin 8.0 software. Each data point was the average of duplicate or triplicate experiments. The enzyme activities of TALs were measured according to the protocol as described previously [49].

3.3.4 Feeding experiments

Feeding experiments were conducted to examine the production of scopoletin and umbelliferone from phenylpropanoid acids (4-coumarate, caffeate, and ferulate). The plasmids pZE-C2'H-Pc4CL2, pZE-F6'H1-Pc4CL2, and pZE-4FO were used for this purpose. For *E. coli* BW25113 carrying pZE-C2'H-Pc4CL2 or pZE-F6'H1-Pc4CL2, method A was used; for BW25113 carrying pZE-4FO, method B was used. For both methods, the transformants were inoculated in 3 ml LB medium containing 100 μg/ml ampicillin and grown overnight at 37 ^{o}C in a shaker. Subsequently, 200 μl overnight cultures were re-

inoculated into 20 ml LB medium containing the same amount of ampicillin. The cultures were grown at 37 ^{o}C with shaking (250 rpm). The expression of the enzymes was induced by adding IPTG to a final concentration of 0.25 mM when OD_{600} reached 0.6-0.7. For method A, after 4 hours' expression at 37^{o}C, the cultures were supplemented with 1mM certain phenylpropanoid acid substrate (4-coumarate or ferulate) and 1mM $FeSO_4$. The bioconversion of the substrates to corresponding coumarins was conducted at 37 ^{o}C with shaking (250 rpm). For method B, cells were collected after 3 hours' induction at 30 ^{o}C and then resuspended in the same volume of M9Y medium, which contains 20 g/L glycerol, 2.5 g/L glucose, 5 g/L yeast extract, 1 g/L NH_4Cl, 6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.1 mM $FeSO_4$, and 1.0 mg/L vitamin B1. 1 mM caffeate and 0.5 mM IPTG were supplemented to the medium. Samples were collected every 6-8 hours to monitor substrate consumption, product formation, and cell growth.

3.3.5 *De novo* coumarin biosynthesis

Shake flask experiments were conducted to verify the *de novo* coumarin biosynthesis. The plasmids pCS-TPTA, pZE-TC4, and pZE-RgTH-4FO were involved in this studied. More specific, pZE-TC4 was introduced into *E. coli* BW25113 to create an umbelliferone producer; while pZE-RgTH-4FO was introduced into *E. coli* BW25113 to generate a scopoletin producer. To test their productivities, overnight culture of these strains were inoculated into M9Y medium containing appropriated antibiotics and cultivated at 37^{o}C with shaking, respectively. When the OD_{600} values of the cultures reached 0.6, IPTG was added to the cultures to a final concentration of 0.5mM. Then the cultures were transferred to 30 ^{o}C for coumarin biosynthesis. Samples were taken every 6-8 hours and analyzed by HPLC.

3.3.6 HPLC and mass spectrum analysis

Umbelliferone (from ACROS ORGANICS), scopoletin (from ACROS ORGANICS), 4-coumarate (from MP Biomedicals), caffeate (Tokyo Chemicals), and ferulate (from MP Biomedicals) were used as the standards. Both the standards and samples were analyzed and quantified by HPLC

(Dionex Ultimate 3000) with a reverse-phase ZORBAX SB-C18 column and an Ultimate 3000 Photodiode Array Detector (DAD). Solvent A was 20 mM ammonium acetate (pH = 5.5), and solvent B was methanol. The column temperature was set to 28 ^{o}C. The following gradient was used at a flow rate of 1 ml/min: 5 to 50 % solvent B for 15 min, 50 to 5 % solvent B for 1 min, and 5 % solvent B for an additional 4min. Quantification of 4-coumarate, caffeate, ferulate, umbelliferone, and scopoletin was based on the peak areas at absorbance of specific wavelengths (303 nm for 4-coumrate, 323 nm for caffeate, 323 nm for ferulate, 323 nm for umbelliferone, and 343 nm for scopoletin). Quantitative analysis of tyrosine and *L*-dopa was performed by an HPLC method reported by our previous study [49]. The peak of our target products were collected and used for ESI-MS analysis. Both positive and negative ion modes were used to confirm the molecular weight.

3.4 Results

3.4.1 Design of umbelliferone and scopoletin biosynthetic pathways

The goal of the pathway design is to connect the coumarin formation reactions with bacterial primary metabolism. First, we designed an umbelliferone biosynthetic pathway based on the C2'H-catalzyed reaction. In plants, 4-coumaroyl-CoA, the immediate precursor for umbelliferone is generated from *L*-phenylalanine via several steps of deamination, hydroxylation, and CoA esterification (Figure 3.3, scheme A). The intrinsical problem associated with this plant-specific pathway is the presence of a cytochrome P450-dependent monooxygenase, cinnamate 4-hydroxylase (C4H). Functional expression of plant P450 enzymes is always hard to be achieved in bacterial systems due to their requirement for anchorage on endoplasmic reticulum and redox partners. To circumvent the barrier of C4H, we replaced the enzymes PAL and C4H with the enzyme TAL, by which *L*-tyrosine, rather than *L*-phenylalanine is utilized directly as the entrance substrate. In addition to C2'H and TAL, a 4-coumarate: CoA ligase (4CL) is also required to complete the pathway by converting 4-coumarate into 4-coumaroyl-CoA. For the case of scopoletin, we designed a more complicated pathway to introduce the 6-methoxyl group, which involves another cytochrome P450-dependent enzyme, 4-coumarate 3-hydroxylase (C3H) in plants. To

obviate the need for C3H, we use the *E. coli* native 4-hydroxyphenylacetate 3-hydroxylase (4HPA3H) as a surrogate, which was identified to be able to hydroxylate 4-coumarate forming caffeate in our previous study [49]. It should be noted 4HPA3H can also hydroxylate *L*-tyrosine affording *L*-DOPA, which can be converted to caffeate by TAL as well. In addition, the enzymes 4CL and caffeoyl-CoA *O*-methyltransferase (CCoAOMT) are necessitated to complete the pathway (Figure 3.3, scheme B). Both designed pathways utilize *L*-tyrosine as the precursor, which can be generated by *E. coli* endogenously. Therefore, *de novo* biosynthesis of scopoletin and umbelliferone becomes feasible by extending the *E. coli* native aromatic amino acid biosynthesis machinery.

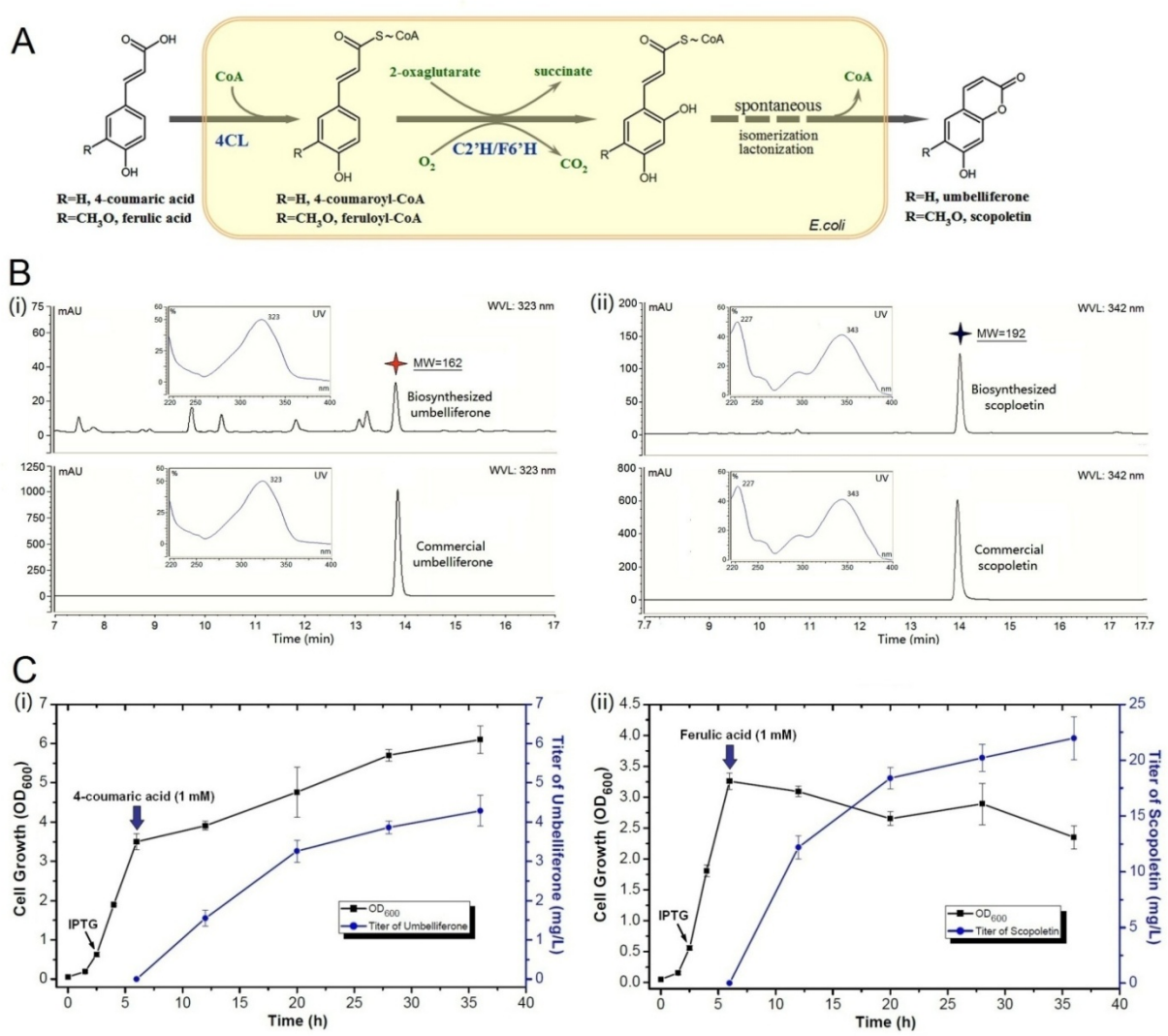


Figure 3.4 Conversion of phenylpropanoid acids to simple coumarins. (A) Schematic representation of the conversion steps from phenylpropanoid acids to simple coumarins. (B) HPLC analysis of umbelliferone (i) and Scopoletin (ii) produced by engineered *E. coli* at 12 h. The red- and black-colored asterisks indicate the biosynthesized umbelliferone and scopoletin, respectively. Meanwhile, the commercial standards (50 mg/L) of corresponding coumarin compounds were analyzed below. UV absorbance profiles are shown beside the peaks. The molecular weights (MW) of the produced compounds were measured by ESI-MS. (C) Time courses of cell growth and production of the engineered *E. coli* strains. Black- and blue-colored arrows indicate the time point of IPTG addition and phenylpropanoid acid feeding, respectively.

3.4.2 Bioconversion of phenylpropanoid acids into simple coumarins

The centerpieces of the designed pathways are the two enzymatic reactions catalyzed by C2'H and F6'H. To examine the functionality of these two reactions in *E. coli* cells, we first constructed two partial pathways leading from 4-coumarate and ferulate to umbelliferone and scopoletin, respectively, since 4-coumarate and ferulate are relatively inexpensive precursors that have been used to produce valuable phenolic compounds, such as resveratrol flavonoids and vanillin [39]. The enzyme Pc4CL2 from *Petroselinum crispum* was recruited to synthesize CoA esters for C2'H or F6'H from 4-coumarate and ferulate due to its broad substrate specificity. According to the enzyme assays, this enzyme demonstrated similar activities towards 4-coumarate (K_m = 11.8 ± 3.6 μM; Specific Activity = 0.119 ± 0.009 μmol.min^{-1}.mg^{-1}) and ferulate (K_m = 11.2 ± 1.9 μM; Specific Activity = 0.082 ± 0.005 μmol.min^{-1}.mg^{-1}) (Table 3.3). Therefore, we reason that generation of umbelliferone/scopoletin from 4-coumarate/ferulate can be achieved by coupling Pc4CL2 with C2'H/F6'H (Figure 3.4). To test this hypothesis, the cDNAs of Pc4CL2, C2'H, and F6'H1 were used to construct the plasmids pZE-C2'H-Pc4CL2 and pZE-F6'H1-Pc4CL2. The genes were consecutively cloned as operons in the plasmids regulated by a strong IPTG-inducible promoter $P_L$lacO1. A ribosomal binding site (RBS) was placed upstream of each open reading frame. The generated plasmids were introduced into *E. coli* BW25113 separately for the feeding

experiments. The samples were collected and analyzed by HPLC and ESI-MS. As shown in Figure 3.4B, HPLC analysis of the samples (taken at 12 h) showed that the retention time and UV absorbance profile of the biosynthesized products were identical to those of coumarin standards. To further confirm the identity of the produced compounds, the corresponding peaks were collected for mass spectrum analysis (Figure 3.5). Positive and negative ion modes of ESI-MS showed ion peaks at m/z 163 $(M+H)^+$ and 161 $(M-H)^-$, respectively, corresponding to the molecular weight 162 for umbelliferone (molecular formula $C_9H_6O_3$); meanwhile ion peaks at m/z 193 $(M+H)^+$ and 191 $(M-H)^-$ reflected the molecular weight of scopoletin (192, molecular formula $C_{10}H_8O_4$). Consistent with the reported in vitro study, F6'H1 also showed strict substrate activity *in vivo*. Other than ferulate, the *E.coli* strain carrying pZE-F6'H1-Pc4CL2 was not able to convert other phenylpropanoid acids including cinnamate, 4-coumarate and caffeate into the corresponding coumarins in 24 h.

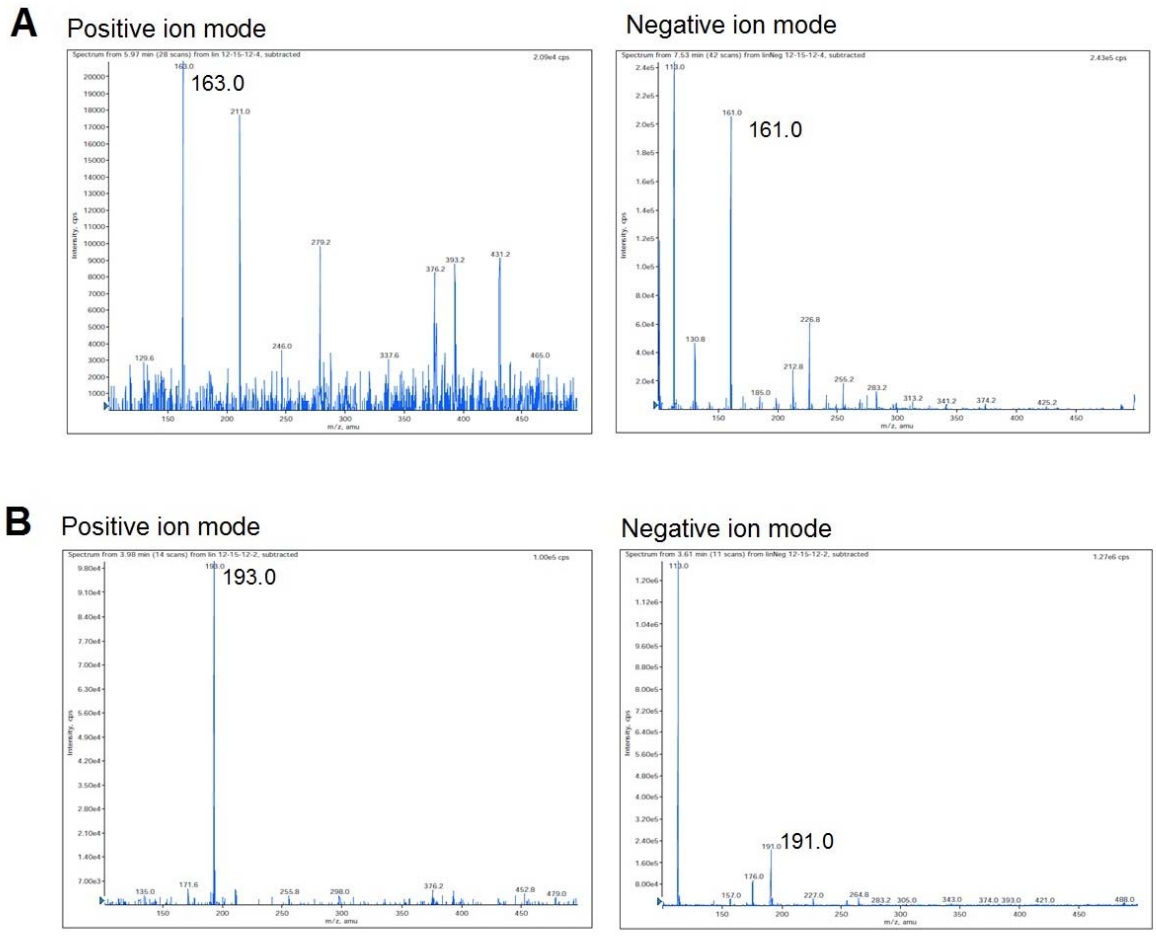


Figure 3.5 ESI-MS analysis of the biosynthesized umbelliferone (A) and scopoletin (B) using both positive and negative ion modes.

The time course bioconversion studies showed that (in Figure 3.4C) production of both umbelliferone and scopoletin occurred immediately after substrate addition, while production rates slowed down gradually as time passed for both umbelliferone and scopoletin producing strains. Difference in growth profile was also observed between these two strains. Production of umbelliferone was accompanied by slow cell growth; while cell growth stopped during the generation of scopoletin. By the end of 36 h, the titer of umbelliferone and scopoletin reached 4.29 ± 0.40 and 21.97 ± 1.93 mg/L by

consuming 17.37 ± 1.61 mg/L 4-coumarate and 55.01 ± 3.63 mg/L ferulate, respectively, reflecting molar yields of 25.0% and 40.9%. Comparatively, wild-type *E. coli* BW25113 did not consume 4-coumarate or ferulate noticeably in the 36 h. In addition, we noticed that frequent sampling from shake flasks negatively affected the product formation, which was evidenced by the scopoletin producing strain generating 27.80 ± 0.04 mg/L scopoletin within 30 hours with one-time sampling, about 39% higher than the titers (~20 mg/L at 30 h) obtained with sampling every 6-8 hours.

3.4.3 *De novo* biosynthesis of umbelliferone and scopoletin

After achieving the biosynthesis of umbelliferone and scopoletin from the corresponding phenylpropanoid acids, we proceeded with the constitution of entire umbelliferone and scopoletin pathways in *E. coli*, which allow the endogenous tyrosine to be utilized for coumarin production.

For the case of umbelliferone biosynthesis, deamination of tyrosine is the entrance step in the artificial pathway (Figure 3.3, scheme A). So a highly active TAL is desired to re-direct the carbon flux towards 4-coumarate. It has been reported that RgTAL (from *Rhodotorula glutinis*) possesses high activity towards tyrosine. So we synthesized a codon-optimized RgTAL gene for its optimal expression in *E. coli*. The enzyme assay in this study suggested that the activity of RgTAL towards tyrosine (29.39×10^{-3} μmol.min^{-1}.mg^{-1}) is more than 30-fold higher than that of RcTAL (0.93×10^{-3} μmol.min^{-1}.mg^{-1}), which was employed in caffeate production in our previous work [49]. For the CoA ligase, At4CL1 from Arabidopsis was recruited for its demonstrated high catalytic efficiency towards 4-coumarate (K_m = 28.7 ± 11.0 μM; Specific Activity = 0.693 ± 0.101 μmol.min^{-1}.mg^{-1}, Table 3.3). For the same reason, C2'H (Ib2-1-1 from Rue) was selected for its activity towards umbelliferone formation [51]. To express the synthetic pathway in *E. coli*, the genes of RgTAL, C2'H, and At4CL1 were organized as an operon and cloned into a high copy number plasmid pZE12-luc, generating an expression plasmid pZE-TC4. *E. coli* BW25113 transformed with pZE-TC4 demonstrated the ability to produce umbelliferone without addition of 4-coumarate. At the end of 48 hours, 2.43 ± 0.09 mg/L umbelliferone was generated with a non-growth-dependent pattern (Figure 2.6A). No obvious accumulation of the intermediates including tyrosine and 4-

coumarate was observed. Furthermore, we explored the possibility to improve the umbelliferone production by increasing the availability of tyrosine. The plasmid pCS-TPTA carrying the genes *ppsA* and *tktA* as well as feedback inhibition resistant variants $tyrA^{fbr}$ and $aroG^{fbr}$ has been shown to be able to improve tyrosine production in our previous work [49]. Therefore, the plasmid pCS-TPTA was co-transferred into *E. coli* BW25113 with pZE-TC4. After 48 hours' cultivation in shake flasks, the accumulation of 1.95 ± 0.26 mg/L umbelliferone was observed in the culture, slight lower than that of no tyrosine overproduction. However, we detected 170.33 ± 30.64 mg/L of tyrosine and 64.26 ± 5.21 mg/L of 4-coumarate remaining unconverted (Figure 3.7). The time course study suggested that the generation of umbelliferone is not growth-dependent in this strain as well. The cell growth reached stationary phase after 12 hours with an OD_{600} value of about 7.

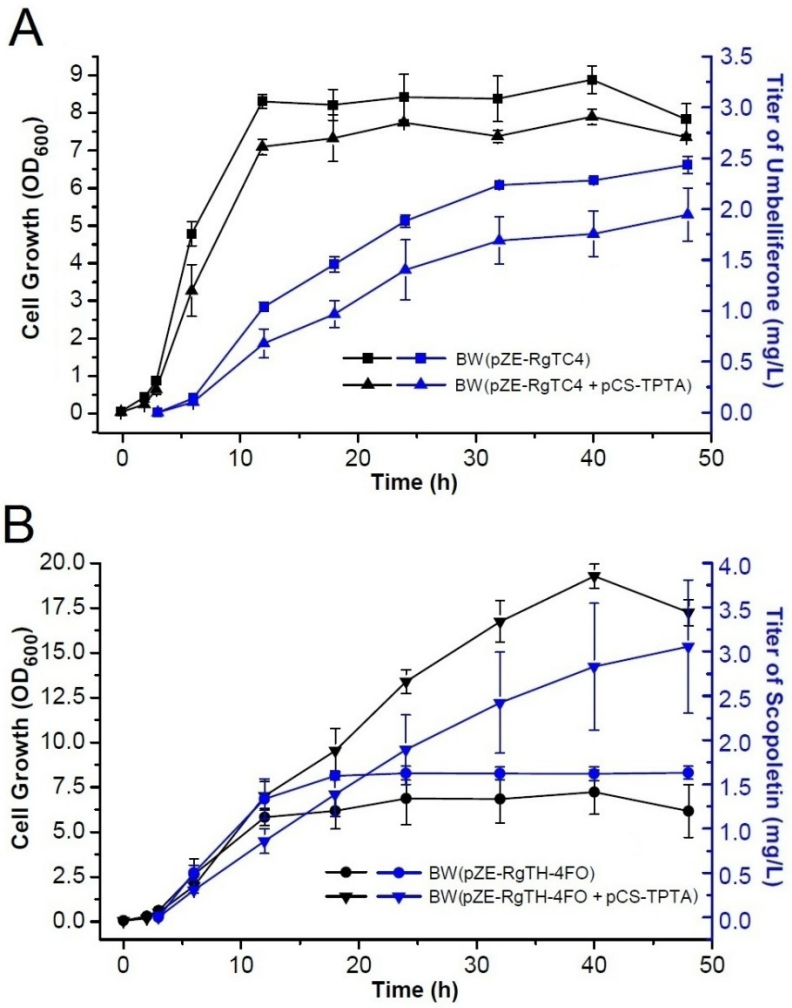


Figure 3.6 Time courses of cell growth and *de novo* biosynthesis of umbelliferone (A) and scopoletin (B). Black- and blue-colored axises correspond to cell density (OD_{600}) and titer of coumarins accumulated in the culture, respectively. BW, *E. coli* host strain BW25113. See table S1 for the information of plasmids pZE-RgTH4, pZE-RgTH-4FO and pCS-TPTA.

Table 3.3 Substrate Specificities of 4-coumarate:CoA ligases (4CLs)

Enzyme	Substrate	Km (μM)	Specific activity (SA) (μmol.min^{-1}.mg^{-1})	SA/Km (ml.min^{-1}. mg^{-1})
***Pc*4CL2**	4-Coumarate	11.8 ± 3.6	0.119 ± 0.009	10.0
	Caffeate	8.4 ± 2.5	0.092 ± 0.006	11.0
	Ferulate	11.2 ± 1.9	0.082 ± 0.005	7.3
***At*4CL1**	4-Coumarate	28.7 ± 11.0	0.693 ± 0.101	24.1
	Caffeate	15.7 ± 7.1	0.316 ± 0.043	20.1
	Ferulate	40.8 ± 4.5	0.408 ± 0.019	10.0
***At*4CL2**	4-Coumarate	45.1 ± 6.2	0.626 ± 0.025	13.9
	Caffeate	7.9 ± 1.8	0.509 ± 0.021	64.4
	Ferulate	ND	–	–

Values are given as average ± standard error of non-linear fitting. ND, not detectable.

Each number is the mean of values taken from at least two independent experiments.

Compared with umbelliferone, it is more complicated to achieve *de novo* biosynthesis of scopoletin, since additional enzymes are required to introduce a 6-methoxyl group onto the coumarin skeleton. First, the deaminated products of tyrosine, 4-coumarate, needs to be hydroxylated into caffeate, which can be performed by a promiscuous hydroxylase 4HPA3H [49]. In our previous work, we have established caffeate production in *E. coli* using RcTAL and 4HPA3H [49]. In this study, we replaced the RcTAL with the more active RgTAL to form an upstream module together with 4HPA3H (Figure 3.3, Scheme B). Additionally, three more enzymatic steps catalyzed by the enzymes 4CL, CCoAOMT, and F6'H have to be incorporated to extend the caffeate pathway to scopoletin, which are designated as the downstream module in this study (Figure 3.3, Scheme B). We then validated the downstream module by

feeding experiments. The plasmid pZE-4FO harboring the genes of At4CL2, CCoAOMT1, and F6'H1 was constructed for this purpose. The enzyme At4CL2 from Arabidopsis was selected due to its higher substrate affinity and catalytic activity towards caffeate compared with Pc4CL2 and At4CL1 (Table S3). The *E. coli* strain BW25113 expressing the downstream module was able to produce 6.61 ± 0.21 mg/L scopoletin in 60 hours when 1 mM caffeate was fed, which suggested CCoAOMT1 is functional *in vivo* and can utilize *E. coli* endogenous S-Adenosyl methionine (SAM) as the methyl donor to perform methylation. We further generated the plasmid pZE-RgTH-4FO to express both upstream and downstream modules. *E. coli* BW25113 containing the plasmid pZE-RcTH-4FO was used to perform *de novo* biosynthesis of scopoletin in shake flasks. By the end of 24 hours, 1.62 ± 0.08 mg/L scopoletin was accumulated in the culture; while no further increase was observed after 24 hours (Figure 3.6B). The intermediates including tyrosine, coumarate, caffeate, and ferulate were not detectable. We also explored the improvement of scopoletin production by increasing the availability of tyrosine. The *E. coli* strain co-transformed with pZE-RgTH-4FO and pCS-TPTA produced 3.06 ± 0.75 mg/L scopoletin after 48 hours' cultivation (Figure 3.6B). At the same time, 111.25 ± 26.31 mg/L tyrosine and 61.86 ± 17.50 mg/L *L*-dopa accumulated in the culture; no detectable accumulation was observed for 4-coumarate, caffeate and ferulate (Figure 3.7). Growth-dependent production of scopoletin was also observed for this strain. Interestingly, the introduction of pCS-TPTA resulted in better cell growth. The scopoletin biosynthesis slowed down once the cell growth entered stationary phase.

3.5 Discussion

Plants generate many pharmaceutical lead compounds, most of which are hardly achieved via chemical synthesis due to the complexity of chemistry and the existence of multiple stereo centers. On the other hand, plant extraction is also inefficient when the requirement for substantial and highly pure compounds became necessary for biomedical studies. Therefore, alternative approaches to rapidly produce these compounds are highly desired. Recent advances in metabolic engineering have allowed the microbial cells to be engineered by precise genetic manipulations for exquisite biosynthesis of target

compounds. Chemical conversion by metabolically engineered microorganisms is now an accepted reality. The synthetic capability of biological systems holds great potential for further exploitation through efforts in *de novo* design and construction of synthetic pathways via combinatorial biosynthesis or retro-biosynthesis. Such efforts require introduction of natural pathways in heterologous hosts, as well as the utilization of enzymes from various species to function in artificial pathways in specific ways.

Engineering coumarin biosynthesis in bacterial systems has not been reported before. In contrast, the production of similar plant-specific and phenylpropanoid-derived compounds such as flavonoids and resveratrol has been widely achieved in metabolically engineered microorganisms [39]. The lagging efforts in engineering coumarin biosynthesis are mainly due to the relatively limited understanding of genetic and biochemical basis of its biosynthesis. In this study, we demonstrate the design of artificial coumarin biosynthetic pathways. Further pathway validation in *E. coli* led to production of over 27 mg/L scopoletin and over 4 mg/L umbelliferone from the corresponding phenylpropanoid acids, as well as their *de novo* biosynthesis (2-3 mg/L) for the first time.

This work is based on the recent findings of two hydroxylase-catalyzed coumarin formation reactions. The further establishment of the entire artificial pathways necessitates the enzymes from both microorganisms and plants, which authorize the interaction between the pathways and the native metabolism allowing endogenous tyrosine to be used as the precursor. The productivity of the artificial pathways largely depends on the activities of the employed heterologous enzymes in *E. coli* when the precursor provided by the native metabolism is abundant. For example, in the case of umbelliferone production, the low titer of target product and accumulation of intermediates tyrosine (170.33 ± 30.64 mg/L) and 4-coumaric acid (64.26 ± 5.21 mg/L) could be explained by the relatively low activities of TAL and C2'H in *E. coli*; while in scopoletin biosynthesis, the accumulation of tyrosine (111.25 ± 26.31 mg/L) and *L*-dopa (61.86 ± 17.50 mg/L) can also reflect the low activity of TAL. In addition, the pathway operation is also driven by several cofactors which are provided by the host as well, For instance, $FADH_2$/NADPH required by 4HPA3H, CoA required by 4CL, SAM required by CCoAOMT, and 2-oxogluarate required by F6'H and C2'H. The heavy reliance of the pathways on native metabolism for

cofactor supplies could also partially explain why the titers of coumarin products are not high in the wild type *E. coli* strain. Furthermore, the low titers may also be the consequence that the pathways involve plant enzymes, which usually cannot be expressed well or have poor stability in prokaryotic hosts due to the lack of specific protein folding chaperons and post-translational modification mechanism. Besides, the collaboration among pathway enzymes and synergistic action between the pathways and the host metabolism may also be sub-optimal. The abovementioned problems can be addressed by further metabolic engineering efforts, which consist of rational and evolutionary strain improvement to improve cofactor availability, pathway enzyme optimization via bio-prospecting and protein engineering, and fine-tuning the expression level and timing of pathway enzymes.

In conclusion, this work can potentially lead to a new approach for coumarin generation through further optimization and improvement. The reconstitution of the artificial coumarin pathways can be readily utilized to facilitate understanding and characterizing the enzymes involved in the biosynthesis of more complex coumarin molecules.

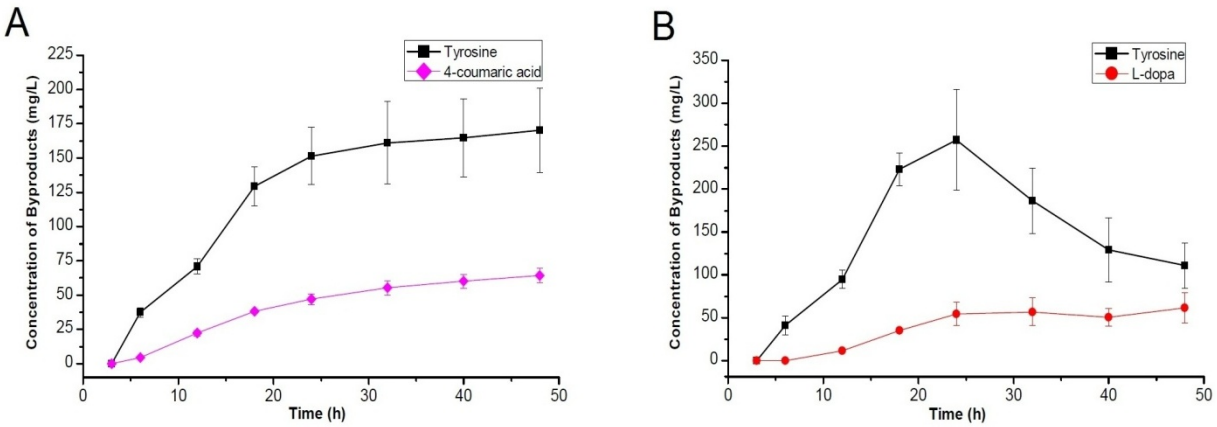


Figure 3.7 Time courses of intermediate accumulation. (A) Strain BW25113 carrying pZE-RgTC4 and pCS-TPTA for the production of umbelliferone; (B) Strain BW25113 carrying pZE-RgTH-4FO and pCS-TPTA for the production of scopoletin.

CHAPTER 4

MICROBIAL SYNTHESIS OF THE ANTICOAGULANT DRUG PRECURSOR 4-HYDROXYCOUMARIN[3]

[3] Yuheng Lin, Xiaolin Shen, Qipeng Yuan and Yajun Yan. 2013, *Nature Communications*, 4: 2603
Reprinted here with permission of the publisher

4.1 Abstract

4-Hydroxycoumarin (4HC) type anticoagulants (e.g. warfarin) are known to play a significant role in the treatment of thromboembolic diseases -a leading cause of patient morbidity and mortality worldwide. 4HC usually serves as an immediate precursor of these synthetic anticoagulants. Although 4HC was initially identified as a naturally occurring product, its biosynthesis has not been fully elucidated. Here we present the design, validation, *in vitro* diagnosis, and optimization of an artificial biosynthetic mechanism leading to the microbial biosynthesis of 4HC. Remarkably, function-based enzyme bioprospecting leads to the identification of a characteristic FabH-like quinolone synthase from *Pseudomonas aeruginosa* with high efficiency on the 4HC-forming reaction, which promotes the high-level *de novo* biosynthesis of 4HC in *Escherichia coli* (about 500 mg/L in shake flasks) and further *in situ* semi-synthesis of warfarin. This work holds scale-up potential for microbial production of 4HC and opens up the possibility of biosynthesizing diverse coumarin molecules with pharmaceutical importance.

4.2 Background

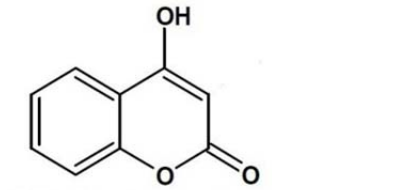
Thromboembolic diseases including venous thromboembolism (VTE) and arterial thrombosis are a leading cause of patient morbidity and mortality worldwide. Annually, VTE alone results in approximately 300,000 and 550,000 deaths in the US and Europe, respectively, and an even larger number of non-fatal events [39]. 4-Hydroxycoumarin (4HC) type oral anticoagulant drugs have been playing significant roles against thromboembolic diseases. Interestingly, anticoagulant function of 4HC derivatives was initially discovered due to its cause of a fetal animal disease manifesting as internal bleeding of the livestock fed with moldy sweet clover forage (called “sweet clover disease”). Indeed, fermentation of plant materials containing melilotoside by molds causes the formation of 4HC and its derivative dicoumarol. The latter demonstrates the blood anticoagulant property by antagonism of vitamin K and acted as a forerunner of the synthetic anticoagulants typified by warfarin [52]. Warfarin is one of the most prescribed oral anticoagulants worldwide with a $300 million global market in 2008 [53]. Besides, acenocoumarol and phenprocoumon are commonly administered in Europe [39]. These drugs

share the 4HC core structure but differ in 3-substitution on the pyrone ring (Figure 4.1), and can be chemically synthesized using 4HC as an immediate precursor [39].

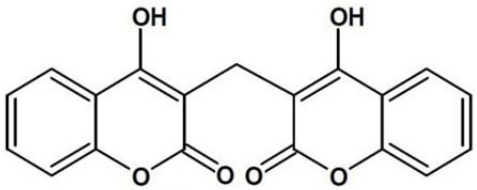
In past decades, various strategies were developed to chemically synthesize 4HC using petro-derived chemicals, such as phenol, acetosalicylate, methylsalicylate, or 2'-hydroxyacetophenone as starting materials [54]. Nevertheless, increasing concerns on petroleum depletion and environmental issues have stimulated greater efforts towards the development of biological processes utilizing renewable resources instead of petro-based chemicals. The convergence of genetics, bioinformatics, and metabolic engineering greatly promoted the engineered biosynthesis of a variety of pharmaceutically important compounds in heterologous microbial hosts, e.g. artemisinic acid [39], taxadiene [39], caffeic acid [49], benzylisoquinoline alkaloids [39], terpenoids [55], anthocyanin [39], flavonoids [21] and resveratrol [39]. All these successful cases were built on thorough understanding of the products' native biosynthetic mechanisms, especially genetic and biochemical properties of the involved enzymes. However, lack of knowledge in these aspects hindered the reconstitution of the biosynthesis of pharmaceutically important 4HC. Although it was proposed that 4HC was formed when melilotoside-containing plant materials were fermented by molds and a biosynthetic scheme was described by isotopic labeling analysis [56], involved enzymes have not been identified [57]. Several recent studies revealed that the *ortho*-hydroxylated cinnamoyl-CoA analogs can form coumarins by spontaneous *trans/cis* isomerization and lactonization [39], suggesting that the pathway might be shunted from *trans*-2-coumaroyl-CoA to generate coumarin rather than 4HC (Figure 4.2). Recently, Liu *et al* identified several biphenyl synthases (BISs) from *Sorbus aucuparia* that catalyze the formation of 3,5-dihydroxybiphenyl through decarboxylative condensation of three malonyl-CoA molecules with benzoyl-CoA. Surprisingly, when *ortho*-hydroxybenzoyl-CoA (salicoyl-CoA) was used in place of benzoyl-CoA as a substrate, only one molecule of malonyl-CoA was condensed to form 4HC, suggesting that the *ortho*-hydroxyl group facilitates the intramolecular cyclization without the condensation of another two malonyl-CoA molecules. Accordingly, a biosynthetic pathway extended from plant salicylate biosynthesis was proposed [58]. However, the same study reported that *S. aucuparia* cells cannot produce 4HC natively even with the presence of supplemented

salicylate [58], indicating the absence of a CoA ligase that can convert salicylate to salicoyl-CoA. In addition, salicylate biosynthesis in plants has not been fully elucidated [59].

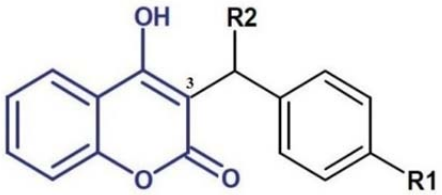
In this work, we present the design and constitution of a novel biosynthetic mechanism affording the *de novo* biosynthesis of 4HC in *E. coli*. Remarkably, a FabH-like quinolone synthase from *Pseudomonas aeruginosa* is identified by function-based bioprospecting, which eliminates the bottleneck of the biosynthetic mechanism. Preliminary optimization via metabolic engineering demonstrates its scale-up potential, leading to efficient biosynthesis of 4HC and *in situ* semi-synthesis of warfarin.



4-Hydroxycoumarin



Dicoumarol



Synthetic 4-hydroxycoumarins

	R1	R2
Warfarin	H	CH_3CHOCH_2
Phenprocoumon	H	CH_3CH_2
Acenocoumarol	NO_2	CH_3CHOCH_2

Figure 4.1 Structures of 4-hydroxycoumarin (4HC) class anticoagulants. Blue-colored moiety indicates the 4HC core structure

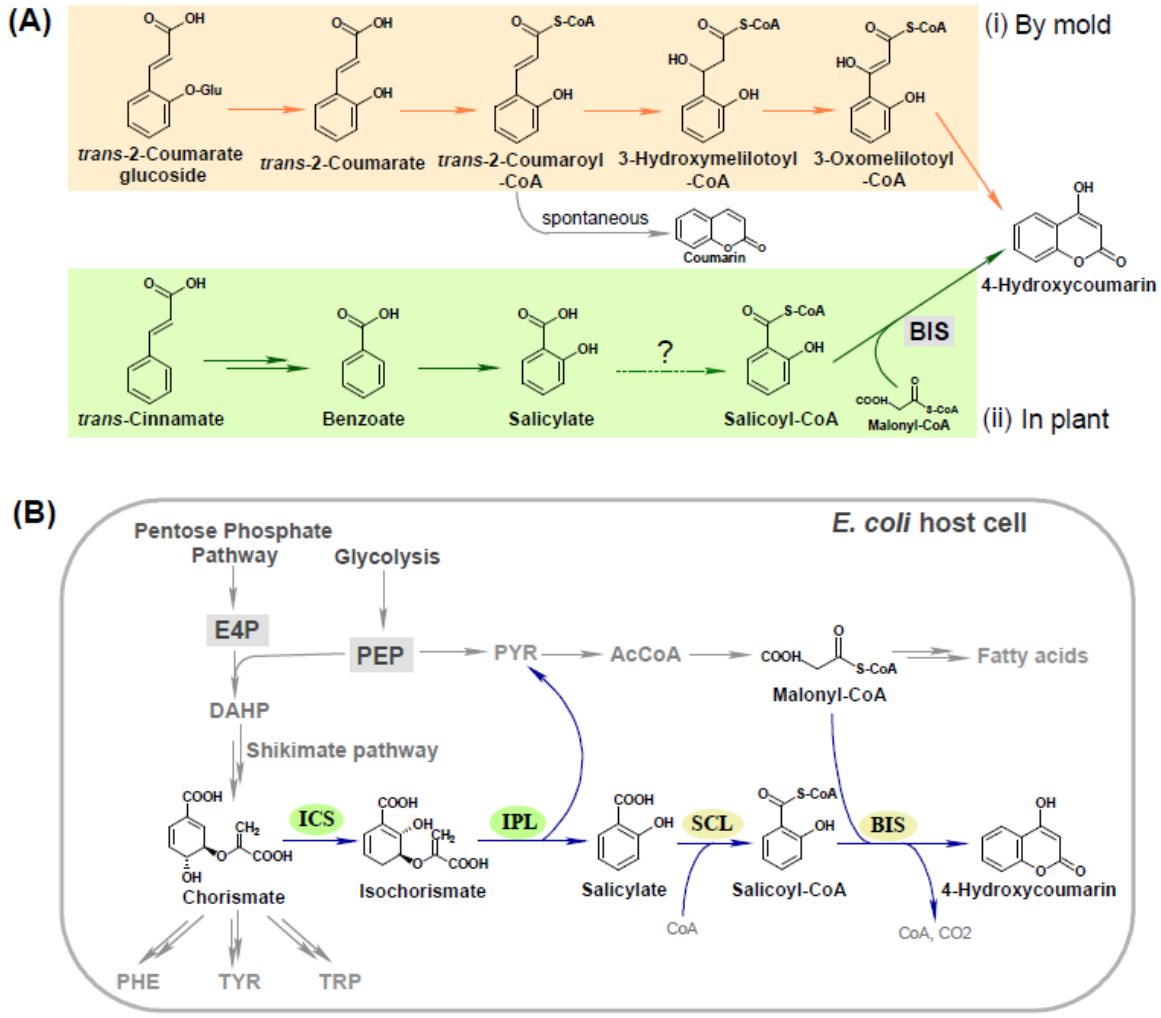


Figure 4.2 Schematic representations of natural and artificial 4HC biosynthetic pathways. (A) Previously proposed natural 4HC biosynthetic routes. Scheme i in the yellow box describes the mold-mediated 4HC biosynthesis; scheme ii in blue box represents a proposed microbe-independent pathway in plant. Question mark indicates a questionable catalytic step that was not identified. (B) The artificial 4HC biosynthetic mechanism designed in this study. Blue-colored arrows indicate non-native catalytic steps; gray-colored arrows indicate the *E. coli* endogenous metabolism. Critical enzymes ICS and IPL, and SCL and BIS are highlighted with green-and yellow-colored circles, respectively, representing upper and lower modules. E4P: *D*-erythrose-4-phosphate; PEP: phosphoenolpyruvate; PYR: pyruvate; AcCoA: acetyl-CoA

4.3 Results

4.3.1 Retro-design of 4HC biosynthesis.

As the direct precursor of natural and synthetic anticoagulants (Figure 4.1), natural 4HC biosynthesis has not been fully understood as mentioned above (Figure 4.2). However, identification of the 4HC-forming reaction catalyzed by BIS provided an opportunity to explore the combinatorial biosynthesis of 4HC. The design was firstly focused on the establishment of a reaction that can provide the substrate salicoyl-CoA for BIS. We speculated that esterification of salicylate with coenzyme A is a reaction that might be catalyzed by certain CoA transferase/ligase. By searching the enzyme database (BRENDA) and literature, we found only a few enzymes with salicylate:CoA ligase (SCL) activity, including SdgA (involved in salicylate degradation in *Streptomyces sp.* WA46), MdpB2 and SsfL1 (involved in maduropeptin and tetracycline SF2575biosynthesis in *Actinomadura madurae* ATCC39144 and *Streptomyces sp.* SF2575, respectively)[39]. Besides, some benzoate:CoA ligases also exhibited weak side activity towards salicylate [39]. To further achieve *de novo* biosynthesis of 4HC, a metabolic connection has to be established between salicylate and the host's metabolism. In nature, salicylate is produced not only by plants as a signal molecule but also by some bacteria as an intermediate in siderophore biosynthesis [39]. Compared with the intricate plant pathways, bacteria generate salicylate using more straightforward strategies. For instance, in *Pseudomonas* and *Mycobacterium* species, salicylate formation requires only two enzymes that are isochorimate synthase (ICS) and isochorismate pyruvate lyase (IPL) by shunting chorismate from shikimate pathway [39]. Taken together, a novel biosynthetic mechanism for 4HC is established by grafting the enzymatic reactions catalyzed by ICS, IPL, SCL and BIS onto the shikimate pathway (Figure 4.2B).

4.3.2 Conversion of salicylate to 4HC

Conversion of salicylate to 4HC (the lower module) by SCL and BIS is a non-natural pathway and the centerpiece of this design. The three identified BISs were reported to show different preference towards salicoyl-CoA; BIS3 was selected for pathway construction due to its higher k_{cat} value [58]. To

obtain an optimal SCL, we characterized the catalytic parameters of SdgA and MdpB2 after evaluating all the reported SCLs. The enzyme assays revealed that SdgA (K_m=4.05 μM, k_{cat} = 10.63 s^{-1}) possesses about 2-fold higher substrate affinity and 10-fold higher activity than MdpB2 (K_m= 8.53 μM, k_{cat} = 1.18 s^{-1}) (Fig. S1 and Table S1). To further test their biosynthetic potential *in vivo*, an expression vector (pZE-BIS3-SdgA) carrying the genes of BIS3 and SdgA was constructed and introduced into *E. coli*. The strain was cultured in the presence of 1 mM salicylate for 24 h. HPLC analysis showed that the strain produced 2.3 ± 0.2 mg/L of 4HC with around 3-6 mg/L salicylate consumed (Fig. S2).

4.3.3 Biosynthesis of salicylate

We borrowed the biosynthetic strategy from *Pseudomonas* involving ICS and IPL to establish the salicylate biosynthesis (the upper module) in *E. coli*. First, to screen for a potent ICS, the enzymes PchA (from *P. aeruginosa*), EntC and MenF (from *E. coli*) were over-expressed and purified for enzyme kinetic studies. The enzyme assays indicated that EntC (K_m =11.93 μM and k_{cat} = 2.12 s^{-1}) is much more active than MenF (K_m =6.75 μM and k_{cat} = 0.13 s^{-1}) and PchA (K_m=3.69 μM and k_{cat} = 0.20 s^{-1}) (Supplementary Fig. S3 and Table S1). Then the activity of IPLs from *Pseudomonas fluorescence and P. aeruginosa* were estimated by coupled enzyme assays since the substrate isochorismate is neither commercially available nor chemically stable. The results showed that the former enzyme (PfPchB, estimated turnover number = 15.8 s^{-1}) is slightly more active than the latter one (PaPchB, estimated turnover number =11.2 s^{-1}). Therefore, EntC and PfPchB were selected for the test of salicylate biosynthesis *in vivo*. We consecutively cloned the genes of EntC and PfPchB into the vector pZE12-luc as an operon, generating pZE-EP. As we expected, *E. coli* strain harboring pZE-EP obtained the capability to produce salicylate. By the end of 32h, 158.5 ± 2.5 mg/L of salicylate was accumulated in the cultures following a growth-dependent production pattern (Supplementary Fig. S4).

4.3.4 Constitution and diagnosis of the 4HC biosynthetic mechanism

With the validated upper and lower modules, further efforts were directed to the validation of the complete 4HC biosynthetic mechanism. The genes encoding EntC, PfPchB, BIS3 and SdgA were consecutively cloned into the vector pZE12-luc as an operon, generating a plasmid pZE-EPBS. However, the *E. coli* strain harboring pZE-EPBS only produced a trace amount of 4HC (< 0.2 mg/L), but accumulated a large amount of salicylate (156.2 ± 18.7 mg/L) after 48h production in shake flasks. The result suggested that the upper module performed well in the full pathway; while the bottleneck lies in the lower module.

To locate the rate-limiting step, we designed and performed an *in vitro* complementation assay in which excess amounts of purified SdgA and/or BIS3 were supplemented into the crude extract of the *E. coli* cells expressing the full pathway. As shown in Figure 4.3, without supplemented enzymes, the crude extract can only convert salicylate to 4HC at a very low rate (0.18 mg/L/h) in the presence of required cofactors; while the presence of purified SdgA and BIS3 significantly improved the rate (8.82 mg/L/h), indicating that the purified enzymes functioned well in this assay system (positive control). When purified SdgA was supplemented alone into the crude extract, the conversion rate was not obviously increased (0.24 mg/L/h). Noticeably, when purified BIS3 was added alone, the 4HC formation was recovered to a rate (7.83 mg/L/h) comparable with that of the positive control, indicating that BIS3 was a major bottleneck in the pathway. We speculated that the low *in vivo* activity of BIS3 might result from the slow kinetics, sub-optimal expression, instability, or cross-species incompatibility issues. To overcome this bottleneck, searching for a superior substitute was our first choice.

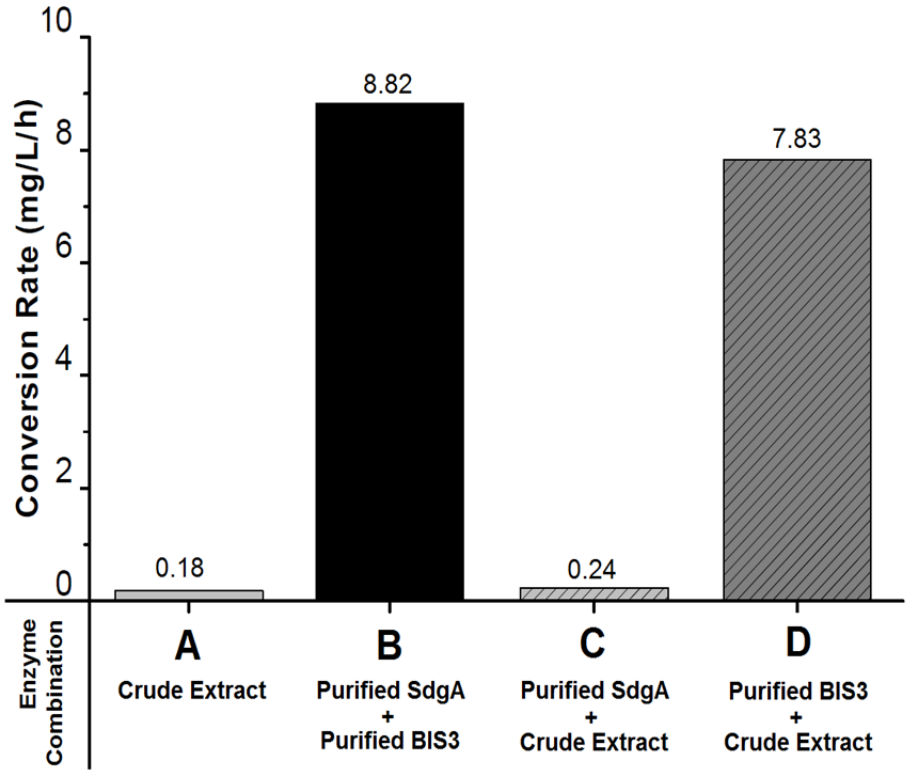


Figure 4.3 *In vitro* complementation assay for examining the rate-limiting enzyme. 4 combinations of enzymes were tested for *in vitro* 4HC formation. *Crude Extract was prepared from the lysed cells of the *E. coli* strain expressing the full pathway (*E. coli*/pZE-EPBS)

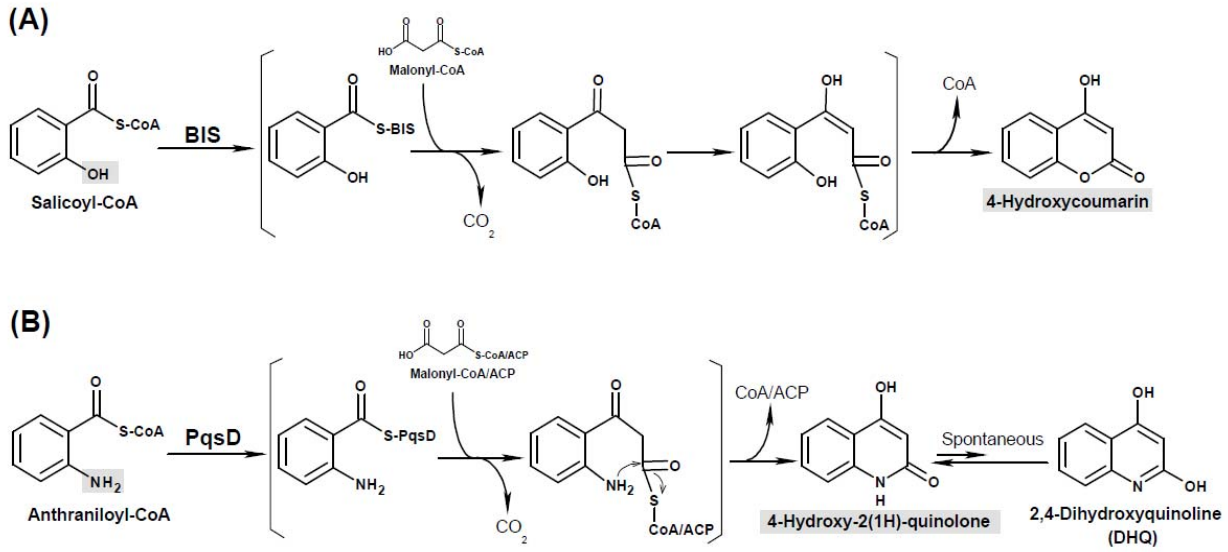


Figure 4.4 Comparison of the reaction mechanisms of biphenyl synthase (BIS) and Pseudomonas quinolone synthase (PqsD). BIS catalyzes the decarboxylative condensation of malonyl-CoA with salicoyl-CoA, after which intramolecular cyclization takes place to form 4HC. PqsD was reported to condense malonyl-CoA/malonyl-ACP and anthraniloyl-CoA. Then similar intramolecular cyclization takes place to form 4-hydroxy-2(1H)-quinolone which is spontaneously interchangeable to its tautomer 2,4-dihydroxyquinoline (DHQ). Under physiological conditions 4-hydroxy-2(1H)-quinolone is the dominant form.

4.3.5 Bioprospecting for a superior substitute to BIS

BIS is a subclass of chalcone synthase (CHS)-like type III polyketide synthases (PKS). However, no other type III PKS with sequence similarity has been identified to catalyze the 4HC-forming reaction. By structure-based examination of bacterial secondary metabolites, we identified that 4-hydroxy-2(1H)-quinolone in *P. aeruginosa* shares high structural similarity with 4HC [39]. The formation of 4-hydroxy-2(1H)-quinolone is catalyzed by a *β*-ketoacyl-ACP synthase III (FabH)-type quinolone synthase (PqsD) via decarboxylative condensation of malonyl-CoA or -ACP with anthraniloyl-CoA and spontaneous

intramolecular cyclization [39]. Despite having a tautomer 2,4-dihydroxyquinoline (DHQ), 4-hydroxy-2(1H)-quinolone is the predominant form at physiological pH [60] (Figure 4.4B). Considering the high similarity in substrate and product structures and catalytic mechanisms between the two reactions, we reasoned that PqsD may also accept salicoyl-CoA as a substrate to form 4HC, as BIS does (Figure 4.4A). To test this hypothesis, we replaced the BIS3 gene with PqsD coding sequence in the lower module, generating the plasmid pZE-PqsD-SdgA (pZE-PS). The *E. coli* strain carrying pZE-PS completely converted 2 mM of salicylate (276 mg/L) into 4HC within about 7 h with a yield of over 99 %, indicating the high activity of PqsD towards salicoyl-CoA. The produced 4HC has identical HPLC retention time and UV absorption profile with its commercial standard. Its identity was further confirmed by ESI-MS and NMR analysis (Supplementary Fig. S5-S8 and Table S2).

4.3.6 Metabolic engineering for improved 4HC biosynthesis

We first reconstituted the improved biosynthetic mechanism in *E. coli* by introducing the two modules as dual operons using the high-copy plasmid pZE-EP-PS. The *E. coli* strain carrying pZE-EP-PS (Strain A, Figure 3.5A) produced 42.3 mg/L of 4HC without addition of any intermediates. Meanwhile, a trace amount of salicylate was detected in the cultures, indicating that the lower module functioned well with this expression strategy and almost completely converted endogenous salicylate to 4HC. However, this expression strategy decreased the efficiency of the upper module. According to the stoichiometry, 42.3 mg/L of 4HC should be generated from 36.1 mg/L salicylate which is much less than the production obtained with the *E. coli* strain only expressing the upper module (pZE-EP). We speculated that the decrease in salicylate-producing capability might be attributed to the following two reasons: 1) the two adjacent operons on the same plasmid might interfere with each other; 2) the incongruous expression of the upper and lower modules may have caused metabolic imbalance. To test this hypothesis, we co-expressed the lower module operon (PqsD-SdgA) using a medium-copy plasmid (pCS-PS) together with pZE-EP in *E. coli* (Strain B, Figure 3.5B). Strain B produced 108.9 mg/L 4HC in 18 h with no measurable salicylate accumulated. Furthermore, we explored the performance of another construct pZE-

EPPS in which the genes encoding the full pathway enzymes were consecutively cloned as one operon. *E. coli* harboring pZE-EPPS (Strain C, Figure 3.5C) produced 207.7 mg/L 4HC in 24 h with 21.6 mg/L salicylate left unconverted. With this expression strategy, the production of 4HC was improved by 5 folds compared with the initial construct, suggesting that gene organization and operon configuration have great impact on the biosynthetic capability of heterologous pathways.

We further speculated that boosting the availability of chorismate and malonyl-CoA, the two major intermediates of 4HC biosynthesis, may divert more metabolic flux towards product formation. Chorismate is a critical intermediate in the shikimate pathway, of which the rate-limiting steps have been identified and the regulation mechanism has been well studied. As shown in Supplementary Fig. S9, the 3-deoxy-*D*-arabino-heptulosonate-7-phosphate synthases in *E. coli* (DAHPS) encoded by *aroG*, *aroF*, and *aroH* are feedback-inhibited by phenylalanine, tyrosine and tryptophan, respectively [61]. Moreover, the erythrose-4-phosphate (E4P) and phosphoenolpyruvate (PEP) availability limit can be alleviated by over-expressing transketolase (encoded by *tktA*) and PEP synthase (encoded by *ppsA*), respectively [39]. Besides, shikimate kinase (encoded by *aroK*/*aroL*) was proved to be another bottleneck which can be eliminated by the over-expression of *aroL* [39]. Based on this knowledge, we cloned *aroL*, *ppsA*, *tktA* and the feedback-inhibition-resistant *aroG* (*aroG*^{fbr}) into pCS27 generating a chorismate-boosting plasmid pCS-APTA. The *E. coli* strain carrying pZE-EPPS and pCS-APTA (Strain E, Figure 3.5E) produced 283.9 mg/L 4HC in 18h, a 37 % increase compared with its parent Strain C. Meanwhile, we created another construct by inserting $P_L$lacO1-APTA operon into pZE-EP generating a dual-operon plasmid pZE-EP-APTA, which was co-transferred together with pCS-PS into *E. coli* (Strain D, Figure 5 D). Remarkably, Strain D produced 483.1 mg/L 4HC in 24 h, reflecting around 4.4- and 11.4-fold increases compared with its parent (Strain B) and the initial Strain A, respectively. Meanwhile, we detected the accumulation of salicylate at the concentrations of 197.6 and 222.3 mg/L at 24 h for Strains D and E, respectively, due to the boosting of shikimate pathway.

Furthermore, we examined the impact of malonyl-CoA on the production of 4HC. Since it has been reported that over-expression of acetyl-CoA carboxylase (*accABCD*) and biotin ligase (*birA*) can

increase the availability of malonyl-CoA [39], we cloned genes *accADBC* and *birA* into pSA74 generating a malonyl-CoA-enhancing plasmid pSA-ACCB. The introduction of pSA-ACCB into Strain E led to slightly improved production of 4HC (313.4 mg/L, 11 % higher than Strain E). However, the pSA-ACCB exerted negative influence to Strain D (*E. coli*/pZE-EP-APTA/pCS-PS), manifesting a dramatic fall in the 4HC production (184.1 mg/L). The results indicated that: 1) malonyl-CoA availability might not be a dominant limiting factor in 4HC production; 2) the over-expression of *accADBC* and *birA* could improve malonyl-CoA availability but might cause metabolic burden in strain D, which offset their benefit in boosting malonyl-CoA availability.

4.3.7 Semi-synthesis of warfarin

With tentative optimization, the resulting *E. coli* strain demonstrated great scale-up potential for 4HC production. Then we explored the feasibility of *in situ* semi-synthesis of warfarin via a green chemistry approach [62]. To this end, the other precursor benzyldeneacetone and the catalyst (*S,S*)-1,2-diphenylethylenediamine were added into the supernatant of the strain D culture (containing about 500 mg/L 4HC) and incubated in a sonication bath for 3 hours. Quantitative HPLC analysis indicated that 43.7 ± 2.6 mg/L warfarin was generated in the supernatant corresponding to a molar yield of 4.6 % (Supplementary Fig. S10). The low yield might be due to the facts that: 1) the aqueous condition is not optimal for the Michael addition reaction; 2) 4HC concentration is lower than the optimal concentration for warfarin synthesis.

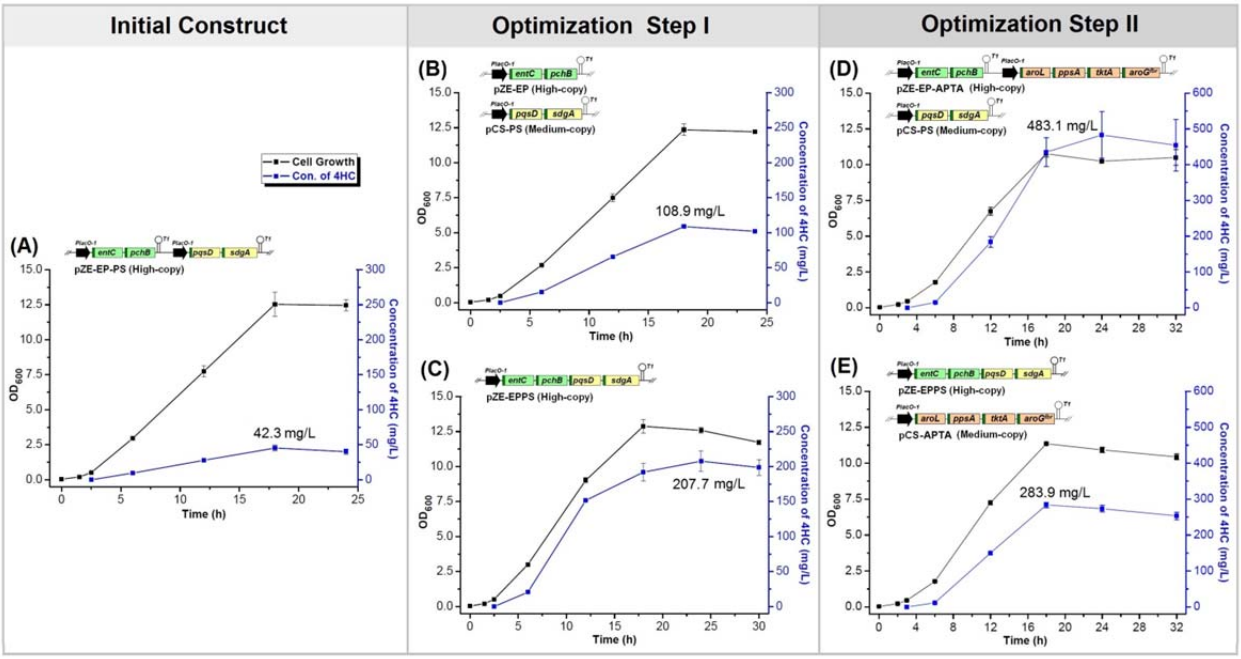


Figure 4.5 Growth and production profiles of the constructed 4HC producing *E. coli* strains. (A) Strain A (*E. coli* carrying pZE-EP-PS) expresses the upper module (EP) and lower module (PS) with 2 operons on the high-copy plasmid; (B) and (C) indicate the modular optimization by adjusting gene organization, copy number, and operon configuration. Strain B separately expresses upper module (EP) and lower module (PS) on the high-copy and the medium-copy plasmids, respectively, while Strain C expresses the full pathway within the same operon on the high-copy plasmid; (D) and (E) indicate improving precursor availability by over-expressing *aroL*, *ppsA*, *tktA* and *aroG*^{fbr} (APTA) on the high-copy and medium-copy plasmids, respectively. Characteristics of the plasmid(s) carried by *E. coli* are shown on the upper-left corner of each graph. All data are reported as mean ± s.d. from three independent experiments (n=3). Error bars are defined as s.d.

4.4 Discussion

Recent advances of metabolic engineering have allowed the microorganisms to be engineered to enable the efficient and environmental-friendly production of valuable molecules. Although the design

principles for constituting a productive pathway are explored and yet to be well-established, recruitment of catalytically superior and host-suitable enzymes should be the primary one in the principles, which is evidenced by this work and previous studies [39]. Conventionally, sequence-based bioprospecting aided by bioinformatics and computational tools is an effective approach in searching for such candidates. For instance, BLAST search using the sequence information of an enzyme with known function as a query may be employed to identify homologous enzymes from various organisms capable of catalyzing the same type of reaction but exhibiting enhanced activity and desired substrate specificity [63]. In our work, we first developed an *in vitro* complementation assay to accurately locate the rate-limiting step in the 4HC biosynthesis. To eliminate the bottleneck, we further employed a function-based bioprospecting strategy to search for a more suitable enzyme. We successfully identified 4-hydroxy-2(1H)-quinolone synthase for efficient microbial biosynthesis of 4HC totally based on the similarity in catalytic mechanism and substrate/product structures between BIS and 4-hydroxy-2(1H)-quinolone synthase. Indeed, 4-hydroxy-2(1H)-quinolone synthase shares low sequence identity with BIS (around 25 %), but exhibits functional and catalytic attributes of both CHS and FabH-like enzymes. On one hand, it can catalyze the condensation of malonyl-CoA as well as intra-molecular cyclization, which are the properties of CHS-like PKS; on the other hand, it can also condense malonyl-ACP in a manner of FabH [39]. So far, the function-based bioprospecting can only be performed manually through analyzing and comparing enzyme catalytic mechanism. However, we envision the development of computational tools that can effectively predict the catalytic substitutability of enzymes with low sequence correlation will further enhance our capability of engineering combinatorial biosynthesis.

With efficient enzymes available for each catalytic step, optimization of the pathway by adjusting expression level of individual enzymes is also critical for the pathway's overall performance. First, inharmonious expression of pathway enzymes may waste cellular resources for the formation of unnecessary RNAs, proteins or intermediates. Besides, over-expression of some exotic enzymes is stressful or toxic to host cells, which may result in growth retardation and undesired adaptive responses, hence reducing yield and productivity [64]. Methodologies have been developed to determine the ideal

expression level and have been successfully applied by fine-tuning the expression level of pathway enzymes and modules, such as the use of plasmids with different copy numbers, promoters with various transcription strength, and synthetic RBSs with different translation efficiency [39]. In our case, modular optimization by adjusting gene organization, copy number, and operon configuration also led to around 5-fold increase in the 4HC titer (Strain C *v.s.* Strain A).

In conclusion, this work achieves microbial production of the pharmaceutically important drug precursor 4HC for the first time and demonstrates great scale-up potential. The findings provide a new insight into the non-natural 4HC biosynthesis, which can serve as a starting point for expanding the molecular diversity of coumarin compounds through synthetic chemistry and biology approaches.

4.5 Materials and Methods

4.5.1 Experimental materials

E. coli strain XL1-Blue was used for plasmid propagation and gene cloning; BL21 star™ (DE3) was used for recombinant protein expression and purification; BW25113 containing F' from XL1-Blue was used as the host strain for the biosynthesis of salicylate and 4HC. The characteristics of all the strains and plasmids used in this study were described in Supplementary Table S3. Luria-Bertani (LB) medium was used to grow *E. coli* cells for plasmid construction, propagation, and inoculum preparation. The biosynthesis medium M9Y contains (per liter): glycerol (20 g), yeast extract (5 g), NH_4Cl (1 g), Na_2HPO_4 (6 g), KH_2PO_4 (3 g), NaCl (0.5 g), $MgSO_4 \cdot 7H_2O$ (2 mmol), $CaCl_2 \cdot 2H_2O$ (0.1 mmol), and vitamin B1 (1.0 mg). 100 μg/ml of ampicillin, 50 μg/ml of kanamycin, and/or 30 μg/ml of chloramphenicol were added when necessary.

4.5.2 DNA manipulation

The plasmids were generated via either regular cloning or Gibson assembly [39]. Information of the pathway enzymes used in this study is summarized in Supplementary Table S4. The primers are listed in Supplementary Table S5. The plasmid pETDUET-1 was used for the over-expression and purification of recombinant proteins with an N-terminal His tag; while pZE12-luc, pCS27 and pSA74 were compatible plasmids used for expressing multiple enzymes involved in the biosynthetic mechanism. The codon-optimized BIS3 cDNA was synthesized by Eurofins MWG Operon. The cDNA of SdgA and MdpB2 are generous gifts from Dr. Julian Davies and Dr. Ben Shen, respectively. Genes of EntC and MenF were cloned from the genomic DNA of *E. coli* MG1655. The coding sequences of PqsD, PchA, and PaPchB were cloned from the *P. aeruginosa* PAO1 genomic DNA. The PfPchB gene was cloned from the *P. fluorescence* Pf5 genomic DNA. To purify the proteins, the genes of SdgA, MdpB2, EntC, MenF, PchA, PaPchB and PfPchB were separately sub-cloned into pETDUET-1. All the genes were fused in frame with the his-tag DNA sequence using BamHI and PstI, except for PchA (using BamHI and

SalI) and MdpB2 (using BamHI and HindIII). To construct pZE-BIS3-SdgA, the genes of BIS3 and SdgA were digested with KpnI/NdeI and NdeI/XbaI, respectively, and then ligated with the KpnI/XbaI digested pZE12-luc fragment via simultaneous three-piece ligation. To construct pZE-PS, the BIS3 gene was replaced with the PqsD cDNA using the same restriction sites. For pCS-PS, the same strategy was employed but using pCS27 as the backbone and different restriction enzymesAcc65I, NdeI and BamHI. To construct pZE-EP, the genes of EntC and PfPchB were digested with KpnI/NdeI and NdeI/SphI, respectively, and then ligated with the KpnI/SphI digested pZE12-luc fragment via three-piece ligation. The plasmid pZE-EP-PS harboring two operons, $P_L$lacO1-EntC-PaPchB and $P_L$lacO1-PqsD-SdgA, was assembled as described by Gilson using the plasmids pZE-EP and pZE-PS as templates and pZE12-luc as the backbone[48]. pZE-EPBS and pZE-EPPS were generated by inserting the BIS3/PqsD and SdgA genes into pZE-EP using the restriction sites SphI, NdeI and XbaI via three-piece ligation. pCS-APTA was constructed by inserting *aroL*, *ppsA*, *tktA* and *aroG*^{fbr} through two rounds of three-piece ligation using KpnI/NdeI/SalI and XhoI/SphI/Hind III. The similar strategy was used to construct pSA-ACCB using Acc65I/PstI/SalI and SalI/EcoRI/BamHI.The plasmid pZE-EP-APTA was constructed by inserting the $P_L$lacO1-APTA operon from pCS-APTA into pZE-EP using SacI and SpeI.

4.5.3 Enzyme assays of SCLs

To evaluate the activity of SCLs (SdgA and MdpB2), the *E. coli* strain BL21 Star™ (DE3) was transformed with pET-SdgA and pET-MdpB2 separately. The obtained transformants were pre-inoculated in Luria-Bertani (LB) medium containing 100 μg/ml of ampicillin and grown aerobically at 37 °C overnight. Next day, the pre-inoculums were transferred into 50 ml of fresh LB medium at a volume ratio of 1:100. The cultures were left to grow at 37 °C till the OD_{600} values reached 0.6-0.8 and then induced with 1.0 mM IPTG. Protein expression was conducted at 30 °C for another 5 h. The cells were harvested and the proteins were purified with the His-Spin Protein Miniprep™ kit (ZYMO RESEARCH). The BCA kit (Pierce Chemicals) was used to estimate protein concentrations. The SCL enzyme assays were performed according to the method described by Ishiyama *et al.* with modifications [39]. The 1 ml

reaction system contained 785 µl of Tris-HCl (pH = 7.5, 100 mM), 5 µl of the purified enzyme (SdgA or MdpB2), 10 µl of $MgCl_2$ (0.5 M), 50 µl of ATP (100 mM), 50 µl of coenzyme A (5 mM), 100 µl of salicylate (100 µM, 200 µM, 500 µM, 1mM). The reactions lasted 0.5 min for SdgA and 2.5 min for MdpB2, respectively, and then were terminated by acidification with 20 µL of HCl (20%). The reaction rates were calculated according to the salicylate consumption at 30 °C, which was measured by HPLC.

4.5.4 Enzyme assays of ICSs

The kinetic parameters of the ICSs: EntC, MenF, and PchA were determined using coupled assays [39]. IPL from *P. aeruginosa* (PaPchB) was used to convert isochorismate to salicylate which was quantified by HPLC. The 1 ml reaction system contained 866 µl of Tris-HCl (pH=7.5, 100 mM), 20 µl of $MgCl_2$ (0.5 M), 0.1 µM of the purified ICS (EntC, MenF or PchA), 0.5 µM of PaPchB, 100 µl of chorismic acid (100 µM, 200 µM, 500 µM, 1 mM). The reactions lasted 1 min for EntC, and 5 min for MenF and PchA, respectively, and then were terminated by acidification with 20 µL of HCl (20%). The reaction rates were calculated according to the salicylate accumulation at 30 °C. The kinetic parameters were estimated by using OriginPro8 through non-linear regression of the Michaelis-Menten equation.

4.5.5 Coupled enzyme assays for IPLs

First, purified EntC enzyme was used to convert an excess amount of chorismic acid into IPL's substrate isochorismate. The 1 ml reaction system containing Tris-HCl (100 mM, pH = 7.5), $MgCl_2$ (5 mM), purified EntC (0.5 µM), Chorismic acid (100 µM) was incubated at 30 °C for 30 min. Then the purified PaPchB or PfPchB was added into the reaction system and incubated for 30 s, after which the reactions were terminated by acidification with 20 µl of HCl (20 %). The enzyme turnover numbers were estimated according to the generation of salicylate, which was measured by HPLC.

4.5.6 Feeding experiments

Feeding experiments were conducted to examine the conversion of salicylate to 4HC. The *E. coli* strain carrying pZE-BIS3-SdgA or pZE-PS was inoculated in 3 ml LB medium and grown overnight at 37 °C. Subsequently, 200 µl overnight cultures were re-inoculated into 20 ml of M9Y medium and grown at 37 °C with shaking (250 rpm). The expression of the enzymes was induced by adding IPTG to a final concentration of 0.5 mM when the OD_{600} values reached 0.6-0.7. At the same time, 1 mM of salicylate was added into the cultures and the cultures were shaken at 30 ^{o}C for several hours, which was followed by HPLC analysis.

4.5.7 *De novo* biosynthesis of salicylate and 4HC

Overnight cultures (100 µl) of salicylate or 4HC producing strains were inoculated into M9Y medium (10 ml) containing appropriate antibiotics and cultivated at 37 °C with shaking at 300 rpm. When the OD_{600} values of the cultures reached around 0.6, IPTG was added to the cultures to a final concentration of 0.5 mM. Then the cultures were transferred to 30 °C for salicylate and 4HC biosynthesis. Samples were taken every other hours and analyzed by HPLC.

4.5.8 *In vitro* complementation assay.

The *E. coli* strain carrying the plasmid pZE-EPBS was pre-inoculated into 3 ml LB liquid medium containing 100 µg/ml of ampicillin and grown at 37 ^{o}C overnight with shaking at 300 rpm. In the following day, 1 ml of the preinoculum was added to 50 ml of fresh M9Y medium. The culture was left to grow at 37 ^{o}C till the OD_{600} value reached around 0.6 and then induced with 0.5 mM IPTG. The expression of the pathway enzymes was conducted at 30 ^{o}C for another 5 h. The cells were harvested and re-suspended in 2 ml of Tris-HCl buffer (100 mM, pH = 7.5), and then lysed by French Press. The soluble fraction was collected by ultra-centrifugation and used as the crude enzyme extract for the complementation assay. The 1 ml reaction system firstly contained Tris-HCl (100 mM, pH = 7.5), $MgCl_2$ (5 mM), ATP (5 mM), coenzyme A (0.25 mM), salicylate (0.2 mM), and crude extract (50 µl)

with/without purified SdgA (20 µl). After 30 min, 100 µl of malonyl-CoA (2 mM) and 50 µl of crude extract with/without purified BIS3 (20 µl) were supplemented into the reaction system. The reactions were finally terminated in another 0.5-2 h by acidification. The reaction rates were calculated according to the generation of 4HC that was measured by HPLC. The protein concentrations of purified BIS3 and SdgA were 1200 and 395 mg/L, respectively.

4.5.9 Semi-synthesis of warfarin

The culture containing about 500 mg/L of produced 4HC was centrifuged to remove the cells. Then 1 g/L of benzyldeneacetone and 100 mg/L of catalyst (*S,S*)-1,2-diphenylethylenediamine were added into the supernatant followed by incubation in the sonication bath for 3 h. The production of warfarin was analyzed by HPLC.

4.5.10 HPLC quantitative analysis

4-Hydroxycoumarin (from ACROS ORGANICS), salicylic acid (from ACROS ORGANICS), and warfarin (from MP Biomedicals) were purchased as the standards. Both the standards and samples were quantitatively analyzed by HPLC (Dionex Ultimate 3000) with a reverse-phase ZORBAX SB-C18 column and an Ultimate 3000 Photodiode Array Detector. Solvent A was sodium acetate solution (20 mM, pH = 5.5), and solvent B was 100 % methanol. The following gradient was used for 4HC and salicylate analysis at a flow rate of 1 ml/min: 5 to 50 % solvent B for 15 min, 50 to 5 % solvent B for 1 min, and 5 % solvent B for additional 4 min, For warfarin analysis, the gradient was from 20% to 80% solvent B. Quantification was based on the peak areas referring to the commercial standards at the wavelength of 285 nm. Samples containing over 200 mg/L of products were diluted before running HPLC to maintain a linear concentration-peak area relationship.

4.5.11 ESI-MS and NMR analysis

For ESI-MS analysis, the peak corresponding to 4HC fraction was collected from HPLC, extracted with acetyl acetate, and dissolved in H_2O. ESI-MS analysis was conducted using the Perkin Elmer Sciex API I plus mass spectrometer. For NMR analysis, the biosynthesized 4HC was extracted from the culture with the same volume of acetyl acetate. Then the extract was dried by a vacuum evaporator, dissolved with DMSO, and diluted with water. Further purification was performed by collecting the 4HC fraction from HPLC. The collected fraction was extracted again with acetyl acetate, dried, and re-dissolved in DMSO. Then the purified 4HC (roughly 0.2-0.3 mg in around 50 µl DMSO) was diluted in 600 µl DMSO-*d6*. The NMR analysis was conducted using 500-MHz Varian Unity Inova with a 5 mm Broad Band Detection Probe at 25 ^{o}C. The solvent DMSO was used as the reference compound. ^{1}H, ^{13}C and gHSQC (gradient Heteronuclear Single Quantum Coherence) analysis was conducted (Supplementary Fig. S6-S8). The carbons and protons were assigned by referring to the data from Spectral Database for Organic Compounds (SDBS No.: 6281) (Supplementary Fig. S11).

4.5.12 NMR data

^{1}H NMR (500 MHz, DMSO-*d6*): δ 5.60 (s, 1H), 7.34–7.39 (m, 2H), 7.65 (t, J = 7.8 Hz, 1H), 7.82 (d, J = 7.8 Hz, 1H), 12.52 (s, br, J = 7.8 Hz, 1H); ^{13}C NMR (125 MHz, DMSO-*d6*): δ 91.44, 116.27, 116.84, 123.66, 124.39, 133.17, 153.98, 162.32, 166.11. The protons and carbons are assigned as shown in Supplementary Fig. S11 and Table S2.

4.6 Supporting Information

4.6.1 Supplementary Figures

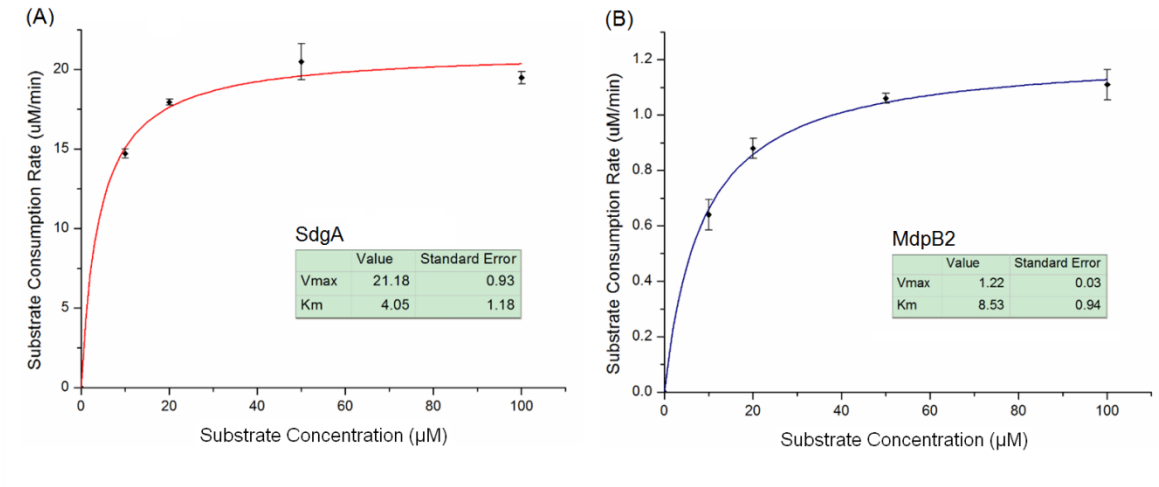


Fig. S1. Kinetic parameters of SdgA (A) and MdpB2 (B). The K_m and V_{max} values were estimated with OriginPro8 through non-linear regression of the Michaelis-Menten equation. Protein concentrations [E] of SdgA and MdpB2 in the reaction systems were 0.0332 μM and 0.0172 μM, respectively. k_{cat} values of the two enzymes were calculated according to the formula $k_{cat}= V_{max}/[E]$. All data points are reported as mean ± s.d. from two independent experiments (n=2). Error bars are defined as s.d.

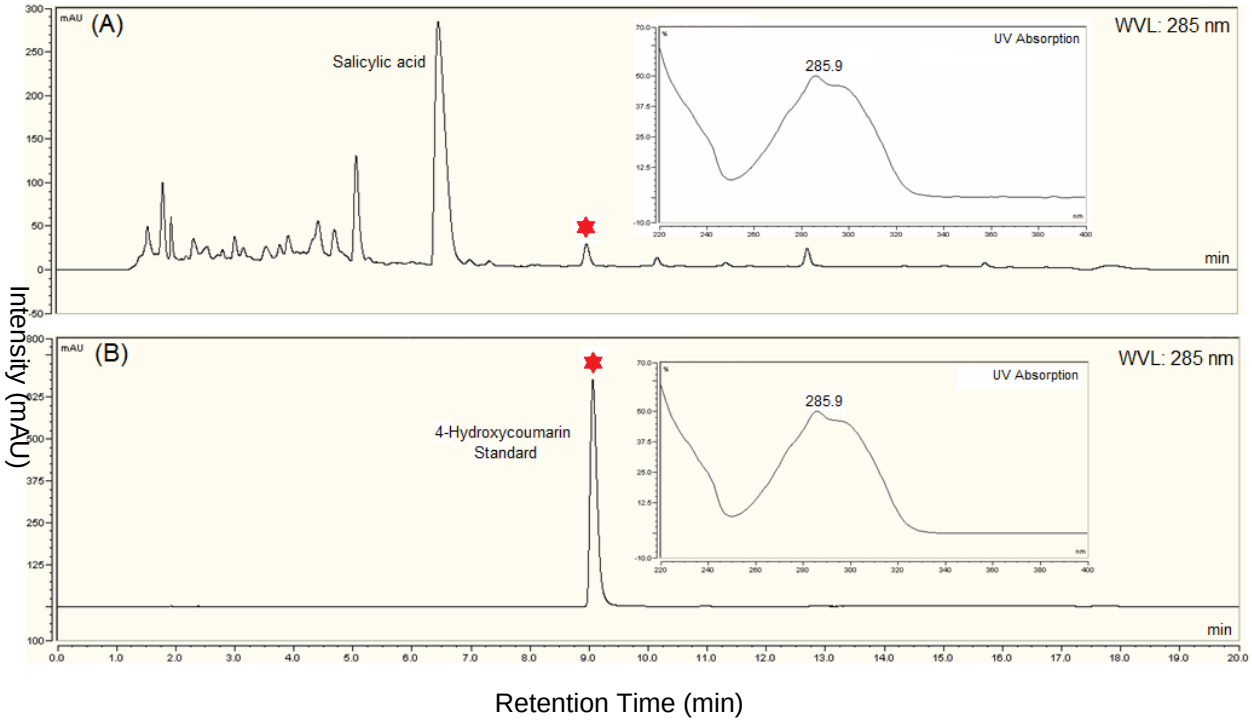


Fig.S2. HPLC analysis of 4HC produced by *E. coli* carrying pZE-BIS3-SdgA in the presence of 1 mM of salicylic acid. (A) A sample taken from the cell culture after 24 hours. (B) 50 mg/L of 4HC standard. The retention time was about 9.0 min. UV absorbance profiles are shown beside the peaks marked with red-colored asterisks.

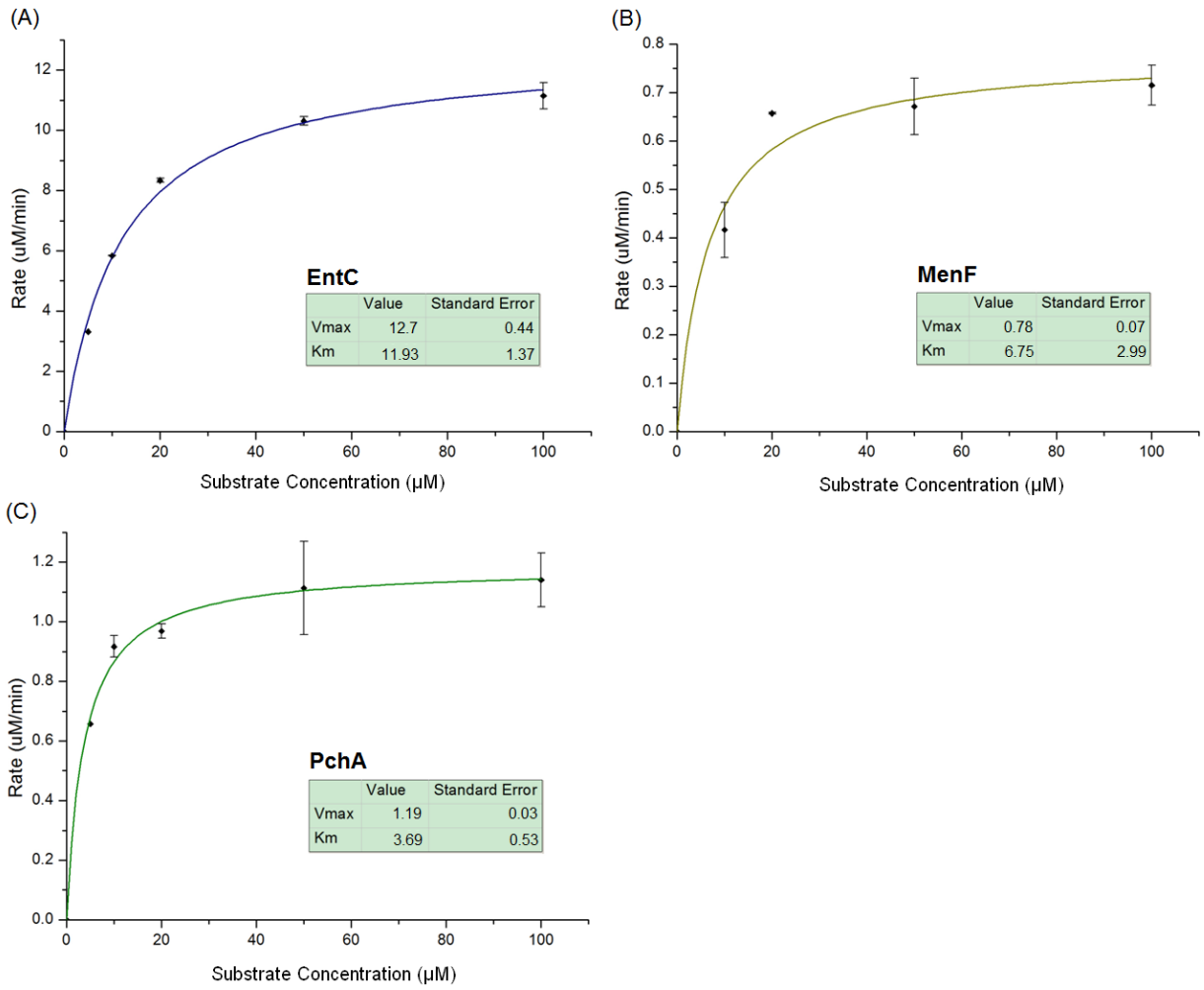


Fig. S3. Kinetic parameters of EntC (A), MenF (B) and PchA (C). The K_m and V_{max}were generated with OriginPro8 through non-linear regression of the Michaelis-Menten equation. Protein concentrations [E] of EntC, MenF and PchA in the reaction systems were all 0.1 μM, respectively. k_{cat} of the enzymes were calculated according to the formula $k_{cat}= V_{max}/[E]$. All data points are reported as mean ± s.d. from two independent experiments (n=2).Error bars are defined as s.d.

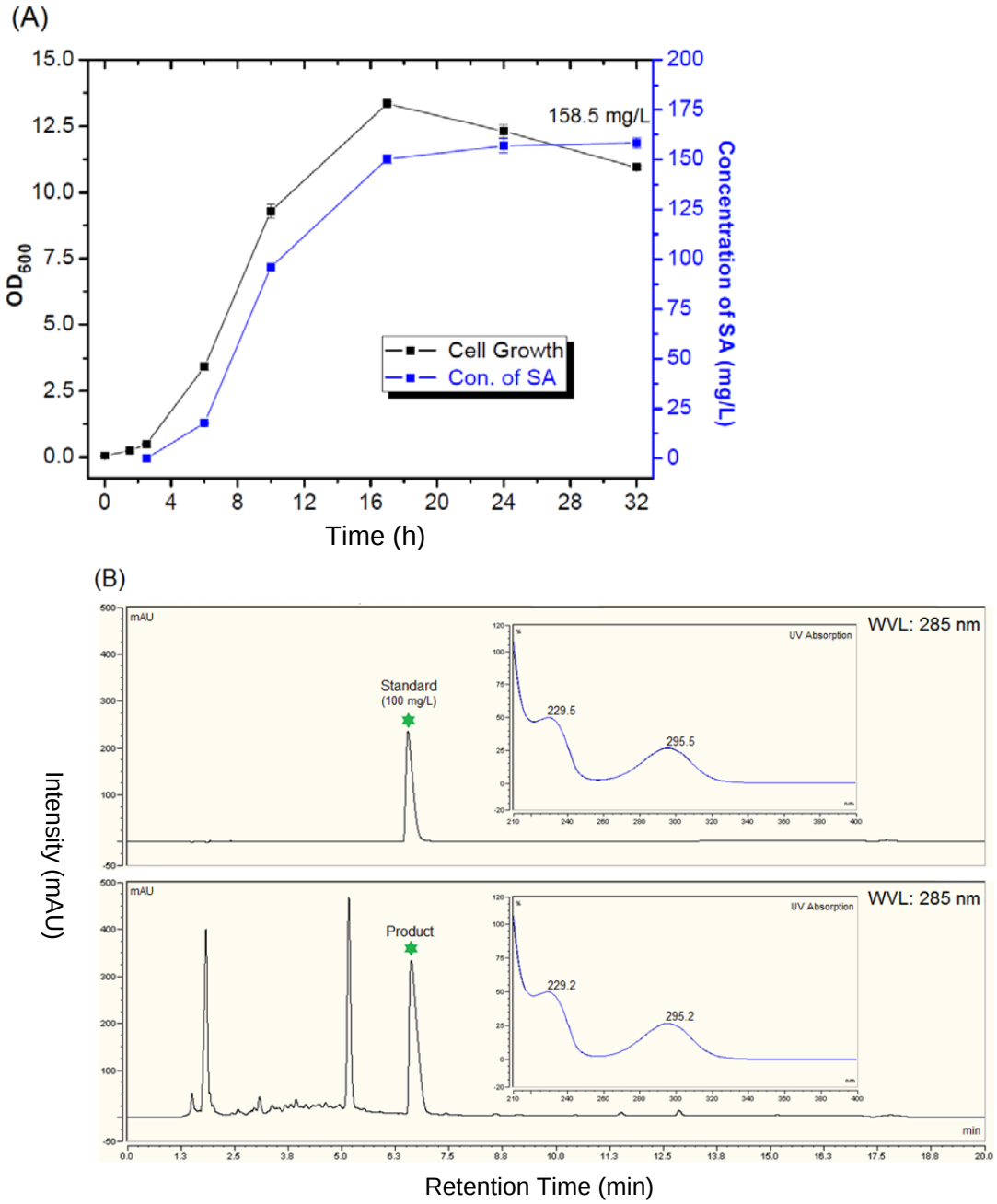


Fig.S4. Salicylate biosynthesis in *E. coli* and HPLC analysis.(A)Time courses of cell growth and salicylate biosynthesis for *E. coli* carrying pZE-EP. All data points are reported as mean ± s.d. from three independent experiments (n=3).Error bars are defined as s.d. (B) HPLC analysis of the biosynthesized salicylate. The UV absorption profiles of the standard and the biosynthesized salicylate are shown beside their peaks (indicated by green-colored asterisks).

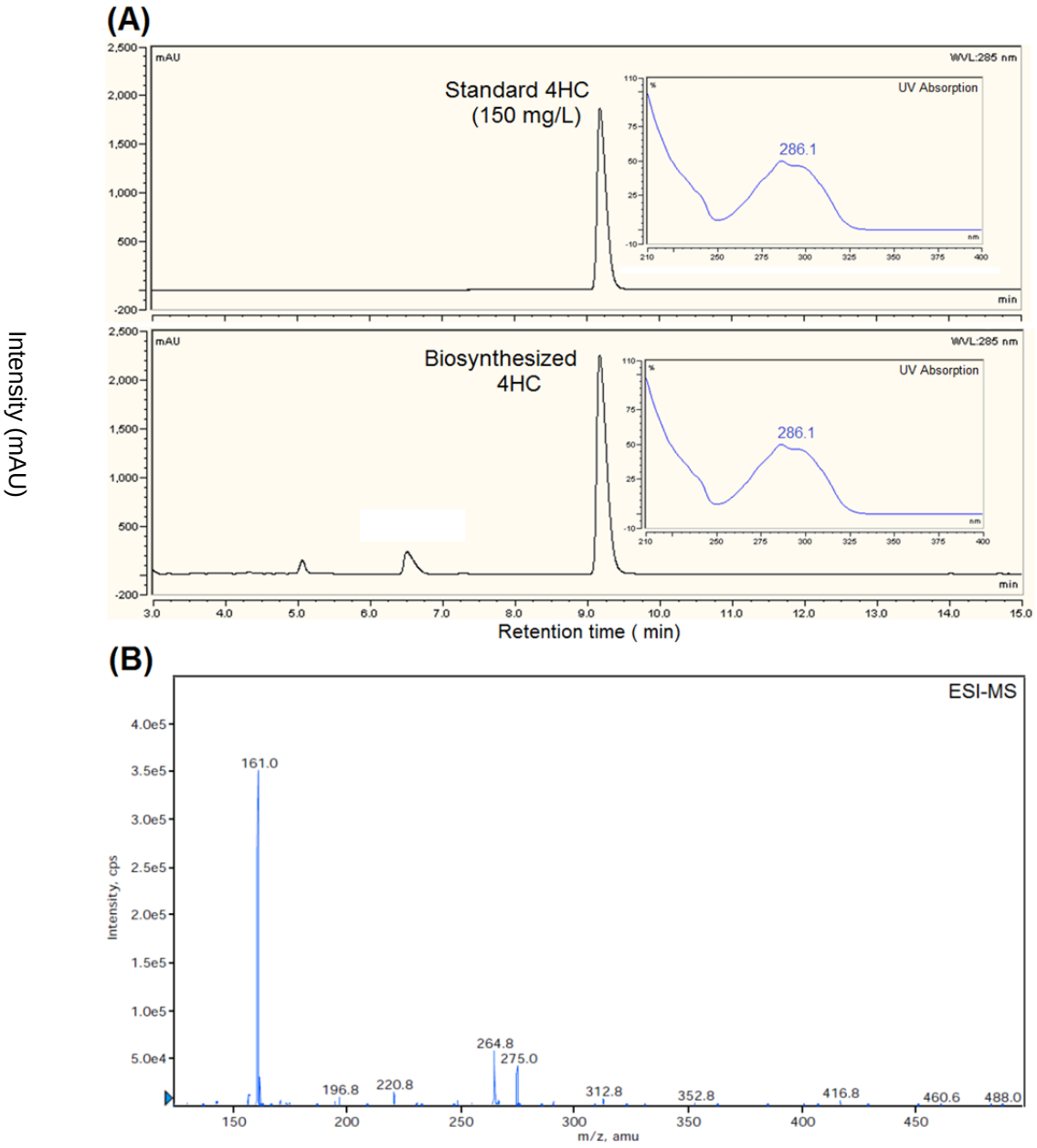


Fig.S5. HPLC and ESI-MS analysis of the biosynthesized 4HC.(A)HPLC analysis of the biosynthesized 4HC. The UV absorption profiles of the standard and the biosynthesized 4HC are shown beside their peaks. (B) ESI-MS (negative ion mode) verification of the biosynthesized 4HC collected and purified from the HPLC fraction. The peak at 161 (M-H)- correspond to the molecular weight 162 (molecular formula $C_9H_6O_3$).

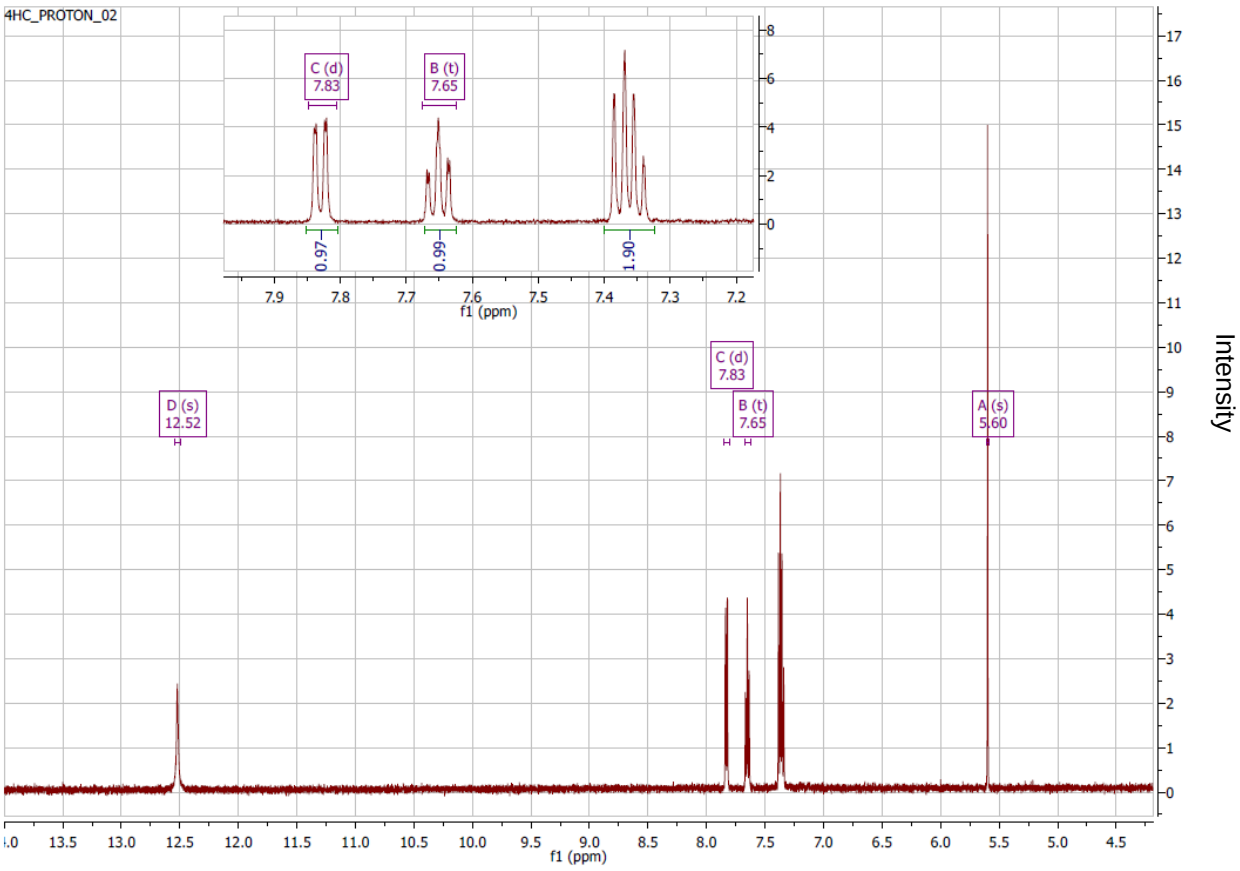


Fig. S6. ^{1}H NMR spectrum of the biosynthesized 4HC. The multiplet between 7.34 and 7.39 was determined to be the signal of two protons based on its integration value and the subsequent gHSQC spectrum.

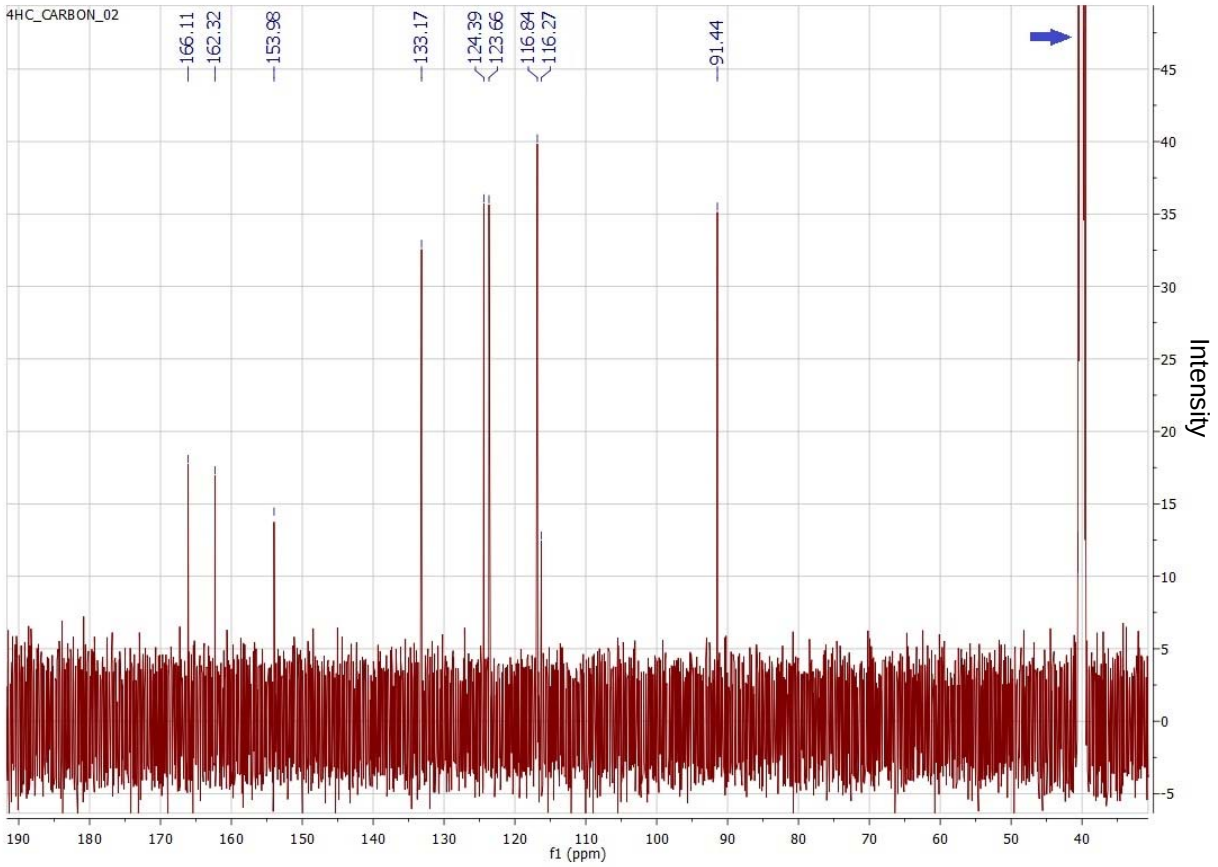


Fig. S7. ^{13}C NMR spectrum of the biosynthesized 4HC. The arrow indicates the solvent DMSO as the reference compound.

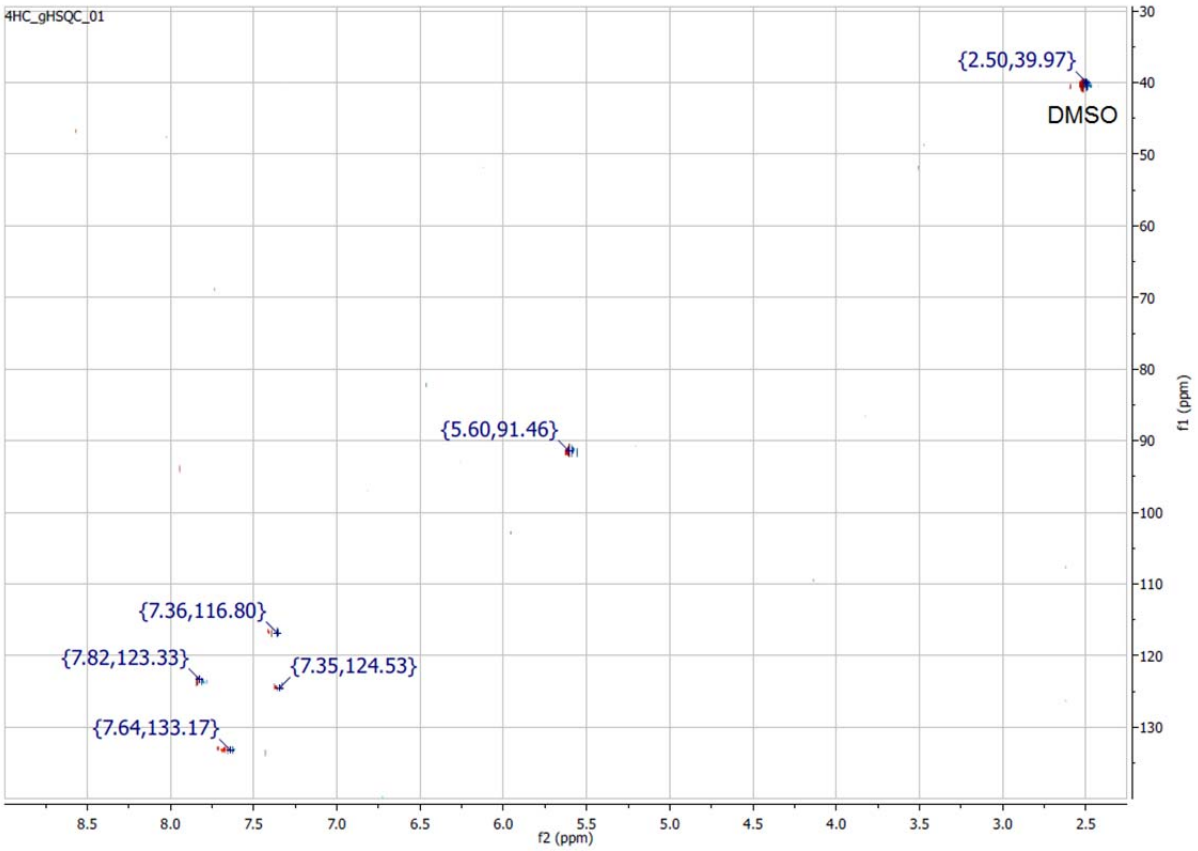


Fig. S8. gHSQC (gradient Heteronuclear Single Quantum Coherence) NMR spectrum of the biosynthesized 4HC. f1 indicates chemical shift for carbons; f2 indicates chemical shift for protons.

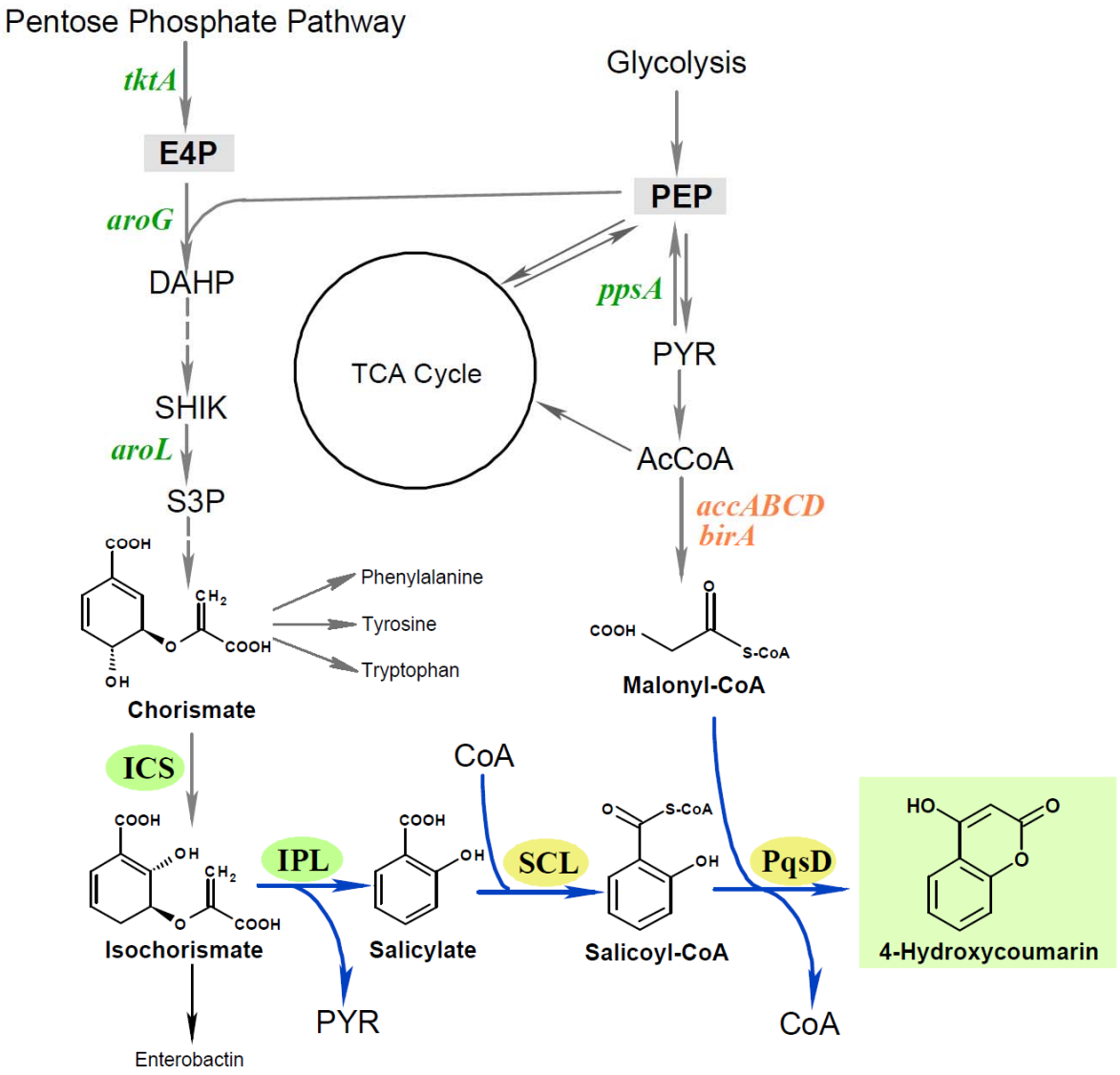


Fig. S9. Artificial 4HC biosynthetic mechanism shunted from shikimate pathway. E4P: *D*-erythrose-4-phosphate; PEP: phosphoenolpyruvate; PYR: pyruvate; AcCOA: acetyl-CoA; DAHP: 3-deoxy-*D*-arabino-heptulosonate-7-phosphate; SHIK: shikimate; S3P: shikimate-3-phosphate.

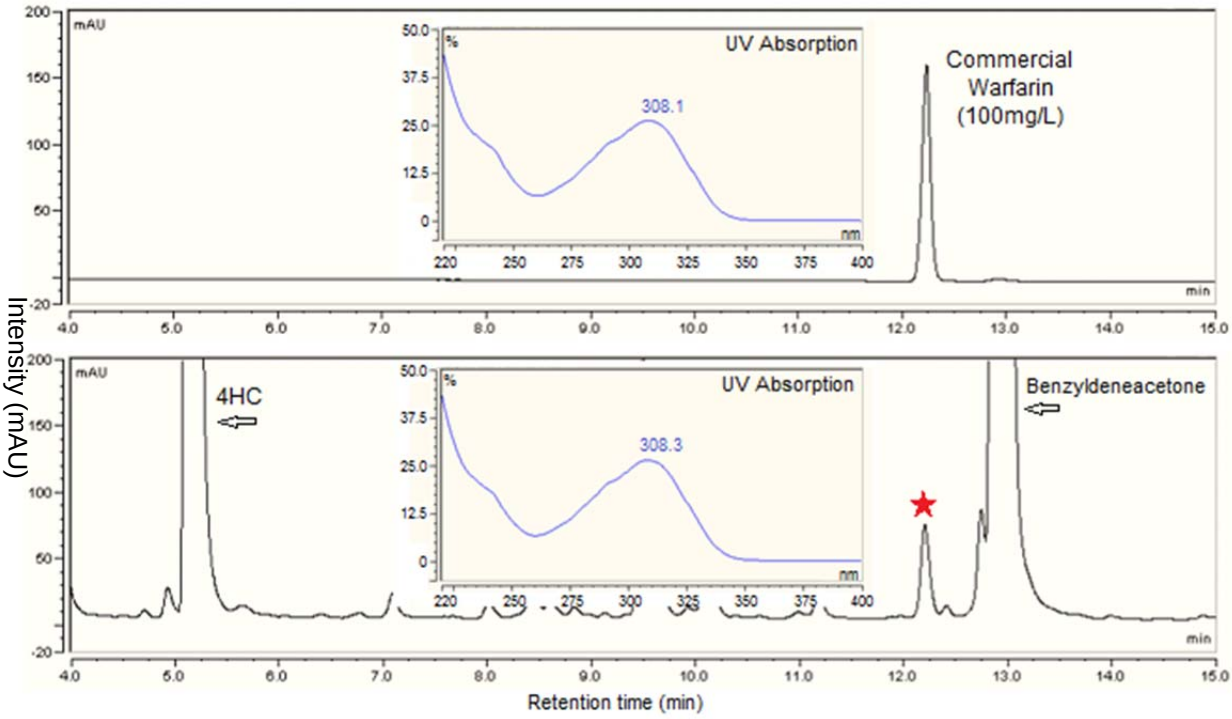


Fig. S10. HPLC chromatography of the standard and semi-synthesized warfarin. The red-colored asterisk indicates the peak of the semi-synthesized warfarin. Its retention time was about 12.2 min. UV absorption profiles are shown beside the warfarin peaks. The peaks of the precursors 4HC and benzyldeneacetone were also indicated.

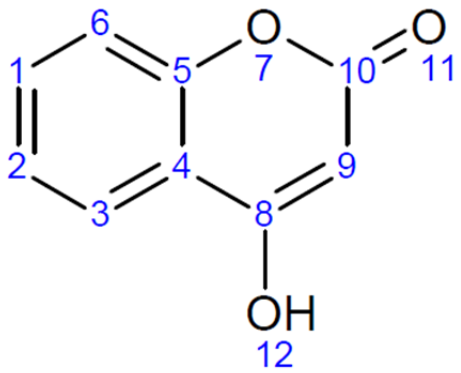


Fig. S11. Numbering of the assigned carbons and protons.

4.6.2 Supplementary Tables

Table S1. Kinetic parameters of the enzymes measured in this study

Enzyme	Kinetic Parameters*		
	k_m (μM)	k_{cat} (S^{-1})	k_{cat}/k_m (S^{-1} M^{-1})
SdgA	4.05 ± 1.18	10.63 ± 0.48	2624691
MdpB2	8.53 ± 0.94	1.18 ± 0.03	138335
EntC	11.93 ± 1.37	2.12 ± 0.07	177703
MenF	6.75 ± 2.99	0.13 ± 0.01	19259
PchA	3.69 ± 0.53	0.20 ± 0.01	54201

*All data are reported as mean ± s.d. from two independent experiments (n=2).

Table S2.NMR data of 4HC

Atom	Chemical Shift (ppm)	Splitting Pattern	J-Value (Hz)
1 C	133.17		
H	7.65	t	7.8
2 C	124.39		
H	7.35	t*	7.9
3 C	123.66		
H	7.82	d	7.8
4 C	116.27		
5 C	153.98		
6 C	116.84		
H	7.37	d*	ND
7 O			
8 C	162.32		
9 C	91.44		
H	5.60	s	
10 C	166.11		
11 O			
12 O			
H	12.52	s, br	

s: singlet; d: doublet; t: triplet; br: broad peak; ND: not determined due to the overlap.

*The multiplet between 7.34 and 7.39 was determined as the overlap of a triplet and a doublet based on the gHSQC spectrum and previously reported data.

Table S3. Strains and plasmids used in this study

Plasmid and Strain	Characteristics	Source
Plasmid		
pZE12-luc	*ColE1* ori; Amp^r; $P_L$lacO-1; *luc*	49
pCS27	*p15A* ori; Kan^r; $P_L$lacO-1; MCS*	49
pSA74	*pSC101* ori; Cm^r; $P_L$lacO-1; MCS	[65]
pETDUET-1	*pBR322* ori; Amp^r two T7 promoters; two MCS	Novagen
pZE-BIS3-SdgA	From pZE12-luc, $P_L$lacO-1; *bis3-sdgA*	This study
pZE-PS	From pZE12-luc, $P_L$lacO-1; *pqsD-sdgA*	This study
pZE-EP	From pZE12-luc, $P_L$lacO-1; *entC-pfpchB*	This study
pZE-EP-PS	From pZE12-luc, dual operons *entC-pfpchB* and *pqsD-sdgA*, both with $P_L$lacO-1	This study
pCS-PS	From pCS27, $P_L$lacO-1; *pqsD-sdgA*	This study
pZE-EPBS	From pZE12-luc, $P_L$lacO-1; *entC-pfpchB-bis3-sdgA*	This study
pZE-EPPS	From pZE12-luc, $P_L$lacO-1; *entC-pfpchB-pqsD-sdgA*	This study
pCS-APTA	From pCS27, $P_L$lacO-1; *aroL-ppsA-tktA-aroG*^{fbr}	This study
pZE-EP-APTA	From pZE12-luc, dual operons *entC-pfpchB* and *aroL-ppsA-tktA-aroG*^{fbr}, both with $P_L$lacO-1	This study
pSA-ACCB	From pSA74, $P_L$lacO-1; *accA-accD-accB-accC-birA*	This study
Strain		
*E. coli*XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacI*^{q}*Z* Δ*M15 Tn10* (Tet^R)]	Stratagene
*E. coli*BW25113	Δ(*araD-araB*), ΔlacZ (::*rrnB-3*), λ-, *rph-1*, Δ(*rhaD-rhaB*), *hsdR*	CGSC
Strain A	BW25113/F' harboring pZE-EP-PS	This study
Strain B	BW25113/F' harboring pZE-EP and pCS-PS	This study

Strain C	BW25113/F' harboring pZE-EPPS	This study
Strain D	BW25113/F' harboring pZE-EP-APTA and pCS-PS	This study
Strain E	BW25113/F' harboring pZE-EPPS and pCS-APTA	This study

*MCS: multiple cloning sites.

Table S4 Information of the pathway enzymes investigated in this work

Enzyme	**Activity**	**Information**
SdgA	Salicylate:CoA ligase	From *Streptomyces sp.* Strain WA46. Initially reported by Ishiyama et al[24]. We determined its kinetic parameters in this study for the first time.
MdpB2		From *Actinomadura madurae* ATCC39144. Initially characterized by Ling et al[25].
EntC	Isochorismate synthase	From *E. coli* MG1655. Initially characterized by Liu et al[51].
MenF		From *E. coli* MG1655. Initially characterized by Daruwala et al[52].
PchA		From *Pseudomonas aeruginosa* PAO1. Initially characterized by Serino et al[53].
PaPchB	Isochorismate pyruvate lyase	From *Pseudomonas aeruginosa* PAO1. Initially characterized by Serino et al[53].
PfPchB		From *Pseudomonas fluorescence Pf-5*. A putative IPL was not characterized before, sharing 62% identity with PaPchB.
BIS3	4HC-forming enzyme	Previously characterized as a biphenyl synthase from *Sorbus aucuparia*, also has the activity to form 4HC[22].
PqsD		From *Pseudomonas aeruginosa* PAO1. Previously identified as a quinolone synthase[31]. We identified its 4HC-forming activity *in vivo*for the first time.

Table S5. Primers used in this study

Primer	Sequence	Use
yy123(BamHI)	gggaaaggatccggatacgtcactggctgaggaagtac	To clone EntC gene
yy124(PstI)	gggaaactgcagttaatgcaatccaaaaacgttcaacatggtag	into pETDUET-1
yy125(BamHI)	gggaaaggatccgcaatcacttactacggcgctgg	To clone MenF gene
yy126(PstI)	gggaaactgcagctattccatttgtaataaagtacgcagccc	into pETDUET-1
yy127(BamHI)	gggaaaggatccgagccggctggcgcccctgagccagt	To clone PchA gene
yy128(SalI)	gggaaagtcgactcaggcgacgccgcgctgcaa	into pETDUET-1
yy131(BamHI)	gggaaaggatccgaaaactcccgaagactgcacc	To clone PaPchB
yy132(PstI)	gggaaactgcagtcatgcggcaccccgtgtct	gene into pETDUET-1
yy133(BamHI)	gggaaaggatccgctggccttcgaccccatgaatt	To clone PfPchB
yy134(PstI)	gggaaactgcagtcactcatcttgggctccttgatc	gene into pETDUET-1
yy135(BamHI)	gggaaaggatccgacgcgtgagggattcgtgccct	To clone SdgA gene
yy136(PstI)	gggaaactgcagtcacaccgcctcgacggagtct	into pETDUET-1
yy139(BamHI)	gggaaaggatccgaccagcattccgcgcatgatcc	To clone MdpB2
yy140(HindIII)	gggaaaaagctttcagcgggtcggggcggtgacgaggt	gene into pETDUET-1
yy141(BamHI)	gggaaaggatccggcccctgtggtcaagaacgagcct	To amplify BIS3
yy142(PstI)	gggaaactgcagtcagtaggtgatgaactcgctacg	gene into pETDUET-1
yy215(KpnI)	gggaaaggtaccatggcccctgtggtcaagaacg	To amplify BIS3
yy185(NdeI)	gggaaacatatgtcagtaggtgatgaactcgctacgcag	gene for pZE-BIS3-SdgA

yy186(NdeI)	gggaaacatatgaggagatataccatgacgcgtgagggattcgtgc	To amplifySdgA gene for pZE-BIS3-SdgA and pZE-PS
yy187(XbaI)	gggaaatctagatcacaccgcctcgacggagtc	
yy180(KpnI)	gggaaaggtaccatggatacgtcactggctgaggaagtac	To amplify EntC
yy181(NdeI)	gggaaacatatgttaatgcaatccaaaaacgttcaacatggtag	gene for pZE-EP
yy182(NdeI)	gggaaacatatgaggagatataccatgctggccttcgaccccatg	To amplify PfPchB
yy183(SphI)	gggaaagcatgctcactcatcttgggctccttgatccag	gene for pZE-EP
yy184(SphI)	gggaaagcatgcaggagatataccatggcccctgtggtcaagaacg	To amplify BIS3
yy185(NdeI)	gggaaacatatgtcagtaggtgatgaactcgctacgcag	gene for pZE-EPBS
yy338(KpnI)	gggaaaggtaccatgggtaatccgatcctggccg	To amplify PqsD
yy339(NdeI)	gggaaacatatgtcaacatggccggttcacctc	gene for pZE-PS
yy218	aaggcggtaatacggttatccacag	Gilson DNA assembly for constructing pZE-EP-PS
yy219	tgagtgagctgataccgctcgc	
yy220	gcgagcggtatcagctcactcaaggcgtatcacgaggccctttc	
yy221	ctgtggataaccgtattaccgcctttagggcggcggatttgtcctac	
yy186(NdeI)	gggaaacatatgaggagatataccatgacgcgtgagggattcgtgc	To amplify SdgA
yy401(BamHI)	gggaaaggatcctcacaccgcctcgacggagtc	gene for pCS-PS
yy362(SphI)	gggaaagcatgcaggagatataccatgggtaatccgatcctggccg	To amplifyPqsD
yy339(NdeI)	gggaaacatatgtcaacatggccggttcacctc	gene for pZE-EPPS
yy188(KpnI)	gggaaaggtaccatgacacaacctctttttctgatcgggc	To amplify *aroL* for
yy189(NdeI)	gggaaacatatgtcaacaattgatcgtctgtgccagggc	pCS-APTA
yy143(NdeI)	gggaaacatatgaggagatataccatgtccaacaatggctcgtcac	To amplify *ppsA* for
yy144(SalI)	gggaaagtcgacttatttcttcagttcagccaggcttaac	pCS-APTA
yy145(XhoI)	gggaaactcgagaggagatataccatgtcctcacgtaaagagcttgcc	To amplify *tktA* for
yy146(SphI)	gggaaagcatgcttacagcagttcttttgctttcgcaac	pCS-APTA

yy147(SphI)	gggaaagcatgcaggagatataccatgaattatcagaacgacgatttacgc	To amplify *aroG*^{fbr} for pCS-APTA
yy148(HindIII)	gggaaaaagcttttacccgcgacgcgcttttac	
yy303(SacI)	gggaaagagctctcttcacctcgagaattgtgagcg	To amplify APTA operon for pZE-EP-APTA
yy304(SpeI)	gggaaaactagtctactcaggagagcgttcaccg	
yy295(KpnI)	gggaaaggtaccatgagtctgaatttccttgattttgaacagc	To amplify *accA* for pSA-ACCB
yy296(PstI)	gggaaactgcagttacgcgtaaccgtagctcatcag	
yy297(PstI)	gggaaactgcagaggagatataccatgagctggattgaacgaattaaaagc	To amplify *accD* for pSA-ACCB
yy298(SalI)	gggaaagtcgactcaggcctcaggttcctgatc	
yy299(SalI)	gggaaagtcgacaggagatataccatggatattcgtaagattaaaaaactgatcgag	To amplify *accBC* for pSA-ACCB
yy300(EcoRI)	gggaaagaattcttatttttcctgaagaccgagtttttctcc	
yy301(EcoRI)	gggaaagaattcaggagatataccatgaaggataacaccgtgccac	To amplify *birA* for pSA-ACCB
yy302(BamHI)	gggaaaggatccttatttttctgcactacgcagggatatttc	

CHAPTER 5

MICROBIAL PRODUCTION OF SALICYLIC ACID AND MUCONIC ACID THROUGH METABOLIC ENGINEERING APPROACHES[4]

[4] Yuheng Lin, Xinxiao Sun, Qipeng Yuan, Yajun Yan. 2014, *Metabolic Engineering*, 23C: 62-69
Reprinted here with permission of the publisher

5.1 Abstract

Cis,cis-muconic acid (MA) and salicylic acid (SA) are naturally-occurring organic acids having great commercial value. MA is a potential platform chemical for the manufacture of several widely-used consumer plastics; while SA is mainly used for producing pharmaceuticals (for example, aspirin and lamivudine) and skincare and haircare products. At present, MA and SA are commercially produced by organic chemical synthesis using petro-derived aromatic chemicals, such as benzene, as starting materials, which is not environmentally friendly. Here, we report a novel approach for efficient microbial production of MA via extending shikimate pathway by introducing the hybrid of an SA biosynthetic pathway with its partial degradation pathway. First, we engineered a well-developed phenylalanine producing *Escherichia coli* strain into an SA overproducer by introducing isochorismate synthase and isochorismate pyruvate lyase. The engineered stain is able to produce 1.2 g/L of SA from simple carbon sources, which is the highest titer reported so far. Further, the partial SA degradation pathway involving salicylate 1-monoxygenase and catechol 1,2-dioxygenase is established to achieve the conversion of SA to MA. Finally, a *de novo* MA biosynthetic pathway is assembled by integrating the established SA biosynthesis and degradation modules. Modular optimization enables the production of 1.5 g/L MA within 48 h in shake flasks. This study not only establishes an efficient microbial platform for the production of SA and MA, but also demonstrates a generalizable pathway design strategy for the *de novo* biosynthesis of valuable degradation metabolites.

5.2 Background

Fossil fuels supply the world with not only energy but also important feedstock for chemical industry. However, the shrinking availability of fossil reserves and the deteriorating environment compel people to explore renewable alternatives for the production of fuels, chemicals, and pharmaceuticals. Fortunately, the metabolic diversity of biological systems provides us with an extremely rich chemical repertoire. In recent years, the development of metabolic engineering has enabled the establishment of microbial chemical factories by constituting heterologous or non-natural biosynthetic pathways into

genetically advantageous microbial hosts [24,49,66-72]. In this study, we report our work on extending shikimate pathway for the production of two industrially important chemicals, *cis, cis*-muconic acid (MA) and its biosynthetic precursor salicylic acid (SA) in *Escherichia coli*.

MA is a platform chemical that serves as the precursor to several bio-plastics. It can be easily converted into adipic acid by chemical hydrogenation, and the latter one is a direct building block for nylon-6,6 and polyurethane [73]. In addition, MA is a synthetic precursor to terephthalic acid, a chemical used for manufacturing polyethylene terephthalate (PET) and polyester [74]. The global production of adipic acid and terephthalic acid is 2.8 and 71 million metric tons, respectively [74]. SA is an important drug precursor mainly used for producing pharmaceuticals such as aspirin and lamivudine (an anti-HIV drug). SA esters and salts used in sunscreens and medicaments account for another large portion of SA consumption. The global market for SA products was estimated to be $292.5 million in 2012 and is expected to reach $521.2 million in 2019, growing at an annual increase of 8.6% [75]. Currently, commercial production of adipic acid, terephthalic acid, and SA predominantly relies on organic chemical synthesis using petroleum-derived chemicals such as benzene as starting materials. These chemical synthesis processes are considered nonrenewable and environmentally unfriendly. Therefore, it is of great importance to develop “green” synthetic approaches that can utilize renewable feedstock.

In fact, MA and SA are both naturally-occurring metabolites. MA is an intermediate in the microbial degradation of aromatic hydrocarbons [76]; while SA serves not only as a plant hormone [59] but also as a biosynthetic precursor of bacterial siderophore [77]. In past 20 years, many efforts have been made for the microbial production of MA. Draths and Frost reported the earliest study on the artificial biosynthesis of MA in *Escherichia coli* from renewable carbon source glucose [78]. By introducing three heterologous enzymes 3-dehydroshikimate dehydratase, protocatechuic acid decarboxylase and catechol 1,2-dioxygenase (CDO), the carbon flux was redirected from the *E. coli* native shikimate pathway to the biosynthesis of MA. Metabolically optimized strains carrying this artificial pathway were able to produce up to 2.4 g/L of MA via two-stage bioconversion in shake flasks [78] and 38.6 g/L via fed-batch fermentation [79]. Afterwards, the same pathway was reconstituted in *Saccharomyces cerevisiae* [80],

and the highest titer reported was nearly 141 mg/L [74]. Very recently, our group reported the construction of a different artificial pathway in *E. coli* by shunting tryptophan biosynthesis from anthranilate, which led to the production of 389 mg/L MA in shake flasks [73]. By contrast, much less attention has been paid on engineering microbes for the production of SA, except for our recent work, in which SA was produced as a biosynthetic precursor to 4-hydroxycoumarin. Over-expression of heterologous enzymes isochorismate synthase (ICS) and isochorismate pyruvate lyase (IPL) in a wild-type *E. coli* strain resulted in the accumulation of 158.5 mg/L of SA in the culture [81].

In this work, we first reconstitute and optimize the biosynthesis of SA in *E. coli* by engineering a well-developed phenylalanine overproducing strain, yielding 1.2 g/L of SA, which is the highest titer reported so far. Further, the partial SA degradation pathway involving salicylate 1-monoxygenase (SMO) and catechol 1,2-dioxygenase (CDO) is established to achieve the conversion of SA to MA. On these bases, a novel MA biosynthetic approach is established by introducing the hybrid of the SA biosynthetic pathway and its partial degradation pathway (Figure 5.1). Through modular optimization, the generated optimal strain produces about 1.5 g/L of MA in shake flasks.

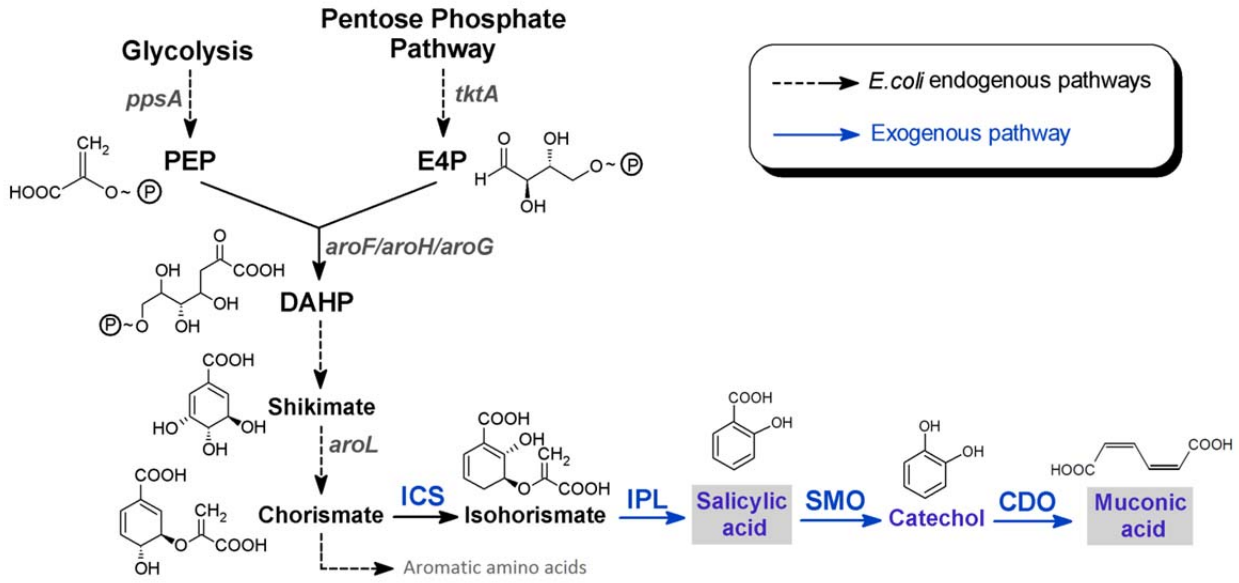


Figure 5.1 A novel artificial pathway for the biosynthesis of MA. Black-colored arrows indicate native metabolic pathways in *E. coli;* blue-colored arrows indicate the introduced artificial pathway. ICS, isochorismate synthase; IPL, isochorismate pyruvate lyase; SMO, salicylate 1-monoxygenase; CDO, catechol 1,2-dioxygenase.

5.3 Materials and Methods

5.3.1 Media, strains, and plasmids

Luria-Bertani (LB) medium was used for inoculants preparation, cell propagation, and protein expression; while modified M9 medium was used for *de novo* microbial production of SA and MA. LB medium contains 10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl. The modified M9 (M9Y) medium contains 10 g/L glycerol, 2.5 g/L glucose, 6 g/L Na_2HPO4, 0.5 g/L NaCl, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, 246.5 mg/L $MgSO_4$, 14.7 mg/L $CaCl_2$, 1 g/L yeast extract, 2 g/L MOPS, vitamin B1 (2.0 mg), H_3BO_3 (1.25 mg), $NaMoO_4{\cdot}2H_2O$ (0.15 mg), $CoCl_2{\cdot}6H_2O$ (0.7 mg), $CuSO_4{\cdot}5H_2O$ (0.25 mg), $MnCl_2{\cdot}4H_2O$ (1.6 mg), and $ZnSO_4{\cdot}7H_2O$ (0.3 mg). When needed, ampicillin, kanamycin, and chloramphenicol were added to the medium to the final concentrations of 100, 50, and 34 μg/ml,

respectively. *E. coli* XL1-Blue was used as the host strain for plasmid construction and propagation. *E. coli* BW25113 was used as a wild-type (wt) strain for *in vivo* enzyme assays and feeding experiments. *E. coli* BL21 Star (DE3) was used for protein expression and purification. *E. coli* ATCC31884 is a phenylalanine overproducing strain purchased from American Type Culture Collection (ATCC) and was used for constructing the derivative strains SXX1 and QH4 [69]. pZE12-luc, pCS27, and pSA74 are high-, medium- and low-copy plasmids, respectively, used for expressing pathway enzymes. Plasmid pETDuet-1 was employed for protein expression and purification. The details of the strains and plasmids used in this study are depicted in Table 5.1.

5.3.2 DNA manipulation

Plasmids pZE-ppbenABCD, pZE-paantABC, pZE-pfantABC were constructed in our previous study [73]. The codon-optimized gene $nahG^{opt}$ was synthesized and then subcloned into pZE12-luc vector using KpnI and XbaI, and into pETDuet-1 using NcoI and XhoI, yielding plasmids pZE-NahG and pET-NahG, respectively. The sequence of $nahG^{opt}$ is provided in the supplementary materials. The plasmids pZE-EP, pCS-APTA, pZE-EP-APTA were constructed in our previous study [81]. The expression cassette $P_L$lacO1-EP was amplified by PCR from pZE-EP and inserted into plasmids pCS27 and pSA74 between SacI and SpeI, yielding plasmids pCS-EP and pSA-EP, respectively. The gene *catA* was amplified from *Pseudomonas putida* KT2440 genomic DNA. Plasmid pZE-NC was created by subcloning $nahG^{opt}$ and *catA* into plasmid pZE12-luc using KpnI, XhoI, and XbaI. The expression cassette $P_L$lacO1-NC was amplified by PCR from pZE-NC and inserted into plasmid pCS27, pSA74, pZE-EP, and pCS-APTA between SacI and SpeI, yielding plasmids pCS-NC, pSA-NC, pZE-EP-NC (two operons) and pCS-NC-APTA (two operons).

Table 5.1 Strains and plasmids used in Chapter 5

Strain	Genotype	Source
XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F′ *proAB lacI*^{q}*ZΔM15 Tn10* (Tet^{r})]	Stratagene
BW25113	rrnBT14 ΔlacZWJ16 hsdR514 ΔaraBADAH33 ΔrhaBADLD78	CGSC
QH4	*E. coli* ATCC31884 with *pheA* and *tyrA* disrupted	[69]
SXX1	QH4 with *trpD* disrupted	This study
Plasmids	**Description**	**Reference**
pZE12-luc	$P_L$lacO1, *colE* ori, *luc*, Amp^{r}	[81]
pCS27	$P_L$lacO1, P15A ori, Kan^{r}	[81]
pSA74	$P_L$lacO1, pSC101 ori, Cm^{r}	[81]
pETDuet-1	pT7, PBR322 ori, Amp^{r}	Novagen
pZE-NahG	pZE12-luc containing *nahG*^{opt}	This study
pZE-paantABC	pZE12-luc containing *antABC* from *P. aeruginosa* PAO1	[68]
pZE-pfantABC	pZE12-luc containing *antABC* from *P. fluorescens* Migula	[68]
pZE-ppbenABCD	pZE12-luc containing *benABCD*from *P. putida* KT2440	[68]
pET-NahG	pETDuet-1 containing *nahG*^{opt}	This study
pZE-EP	pZE12-luc containing *entC* from *E. coli* and *pchB* from *P. fluorescens* Migula	[81]
pCS-EP	pCS27 containing *entC* from *E. coli* and *pchB* from *P. fluorescens* Migula	This study
pSA-EP	pSA74 containing *entC* from *E. coli* and *pchB* from *P. fluorescens* Migula	This study
pZE-NC	pZE12-luc containing *nahG*^{opt} and *catA* from *P. putida* KT2440	This study
pCS-NC	pCS27 containing *nahG*^{opt} and *catA* from *P. putida* KT2440	This study

pSA-NC	pSA74 containing *nahG*^{opt} and *catA* from *P. putida* KT2440	This study
pCS-APTA	pCS27 containing *aroL, ppsA, tktA, aroG*^{fbr} from *E. coli*	[81]
pZE-EP-APTA	pZE12-luc containing $P_L$lacO1-EP and $P_L$lacO1-APTA	[81]
pZE-EP-NC	pZE12-luc $P_L$lacO1-EP and $P_L$lacO1-NC	This study
pCS-NC-APTA	pCS27 containing $P_L$lacO1-NC and $P_L$lacO1-APTA	This study

5.3.3 Screening SMOs *in vivo*

To evaluate the activity of enzymes converting SA to catechol, *E. coli* BW25113 was transformed with the expression vectors pZE-ppbenABCD, pZE-paantABC, pZE-pfantABC and pZE-NahG, separately. The resultant transformants were inoculated in 3 ml LB liquid medium containing 100 μg/ml of ampicillin and grown aerobically at 37 ^{o}C. The overnight cultures were inoculated into 20 ml fresh LB medium and left to grow at 37 ^{o}C till OD_{600} reached 0.6 and then induced with 0.25 mM of IPTG at 37 ^{o}C for additional 3 h. Then the cells were harvested, and re-suspended in M9Y medium (OD_{600} = 2.1-2.7). SA was added into the cell suspension to a final concentration of 2 mM. The flasks were incubated with shaking at 37 ^{o}C for 10 min for the SMO encoded by *nahG*^{opt} and 1 h for ppBenABCD, paAntABC, and pfAntABC. Samples were taken by removing cell pellets and the product (catechol) concentrations were measured with HPLC. The *in vivo* activity was expressed as μM/min/OD.

5.3.4 *In vitro* SMO enzyme assay

To express and purify the enzyme, *E. coli* BL21 Star (DE3) was transformed with the expression plasmid pET-NahG. A fresh transformant was inoculated in 3 ml LB medium containing 100 μg/ml of ampicillin and grown aerobically at 37 ^{o}C. Overnight cultures were inoculated into 50 ml fresh LB medium and left to grow at 37 ^{o}C till OD_{600} reached 0.6 and then induced at 30 ^{o}C with 0.5 mM IPTG for another 3 h. Cells were then harvested and lysed by French Press. The recombinant protein with an N-terminal multi-histidine tag was purified using His-Spin protein miniprep kit (ZYMO RESEARCH). The enzyme concentration was measured using BCA kit (Pierce Chemicals). The standard enzyme assay was

performed by making an assay mixture containing 500 μM NADH, 10 μM FAD, 0.97 nM purified enzyme, and SA as the substrate. The final volume was adjusted to 1 ml with Kpi buffer (20 mM, pH = 7.0). The substrate concentrations varied from 0 to 100 μM. The reaction system was kept at 37 ^{o}C for 1 min and stopped by adding 50 μL HCl (20%). The reaction rates were calculated by measuring the formation of catechol via HPLC. The kinetic parameters were estimated with OriginPro8 through non-linear regression of the Michaelis–Menten equation.

5.3.5 Feeding experiments

Feeding experiments were conducted to examine the production of MA from SA. *E. coli* BW25113 was transformed with the plasmid pZE-NC. Single colonies were inoculated into 3 ml LB medium containing 100 μg/ml ampicillin and grown aerobically at 37^{o}C. 200 μl of overnight cultures were inoculated into 20 ml LB medium containing 100 μg/ml ampicillin. The cultures were left to grow at 37 ^{o}C till OD_{600} reached 0.6 and then induced with 0.25 mM IPTG. After 3 h induction, cells were harvested, re-suspended in 20 ml of M9Y medium containing 3 mM of SA. Then SA was continuously fed into the cultures at 3 mM/h. Samples were taken at 2 h, 5 h and 10 h and the product concentrations were analyzed by HPLC.

5.3.6 *De novo* production of SA and MA

Overnight LB cultures of the producing strains were inoculated at 3% into the M9Y medium containing appropriated antibiotics and cultivated at 37 ^{o}C with shaking at 300 rpm. IPTG was added to the cultures to a final concentration of 0.25 mM at 0 h. Samples were taken every 24 hours. OD_{600} values were measured and the concentrations of the products and intermediates were analyzed by HPLC.

5.3.7 HPLC analysis

SA (from SIGMA ALDRICH), catechol (from Alfa Aesar), MA (from ACROS ORGANICS) were used as standards. Both the standards and samples were analyzed and quantified by HPLC (Dionex

Ultimate 3000) equipped with a reverse phase ZORBAX SB-C18 column and an Ultimate 3000 Photodiode Array Detector. Solvent A was water with 0.1 % formic acid, and solvent B was methanol. The column temperature was set to 28 ^{o}C. The following gradient was used at a flow rate of 1 ml/min: 5 to 50 % solvent B for 15 min, 50 to 5 % solvent B for 1 min, and 5 % solvent B for an additional 4 min. Quantification of SA, catechol, and MA was based on the peak areas at absorbance of specific wavelengths (329 nm for SA, 274 nm for catechol, and 260 nm for MA). Glucose, glycerol, and acetate were quantified using a previously described method [72].

5.4 Results

5.4.1 A novel artificial biosynthetic pathway towards MA production

SA is a widely occurring aromatic metabolite which is produced not only by plants as a phytohormone but also by some bacteria as an intermediate in the biosynthesis of siderophore [59,77]. For its biosynthesis in bacteria, only two enzymes isochorismate synthase (ICS) and isochorismate pyruvate lyase (IPL) are required to synthesize SA from chorismate, a pivotal metabolite in shikimate pathway [77,82]. In contrast, some bacterial species such as *Pseudomonas* were reported to be capable of utilizing SA as a carbon and energy source, during which SA is degraded via catechol and MA [76,83]. In nature, however, the catabolism and anabolism of SA usually do not occur simultaneously in time and space. In this work, we reconstitute and synchronize the SA biosynthesis pathway catalyzed by ICS and IPL and its partial degradation pathway catalyzed by SMO and CDO, leading to a novel biosynthetic approach towards MA production from renewable carbon sources (Figure 5.1).

5.4.2 Transformation of a phenylalanine overproducer to an SA overproducer

Re-directing carbon flux from the native shikimate pathway towards salicylate biosynthesis is the first step towards establishing the artificial MA biosynthetic pathway. Previously, we have reported that the EntC from *E. coli* and the PchB from *P. fluorescence* are the most efficient ICS and IPL, respectively, among all the screened enzymes. When EntC and PchB were co-expressed in wild type *E. coli* host, the

resulting strain produced 158.5 mg/L of SA after 32-hour cultivation [81]. To further elevate the production of SA, we focused on engineering a well-developed phenylalanine overproducing strain *E. coli* ATCC31884, since this strain has been successfully modified to produce tyrosine and caffeic acid efficiently [69,84]. To eliminate undesired consumption of chorismate, we disrupted the competing pathway genes *pheA*, *tyrA,* and *trpD* from ATCC31884 which encode the enzymes responsible for the first committing catalytic steps in the biosynthesis of phenylalanine, tyrosine, and tryptophan, respectively, generating the strain SXX1 (Figure 5.2A). However, when EntC and PchB were over-expressed in SXX1with the high-copy plasmid (pZE-EP), the resulting strain only produced 177.95 mg/L of SA after 48 hour cultivation (Figure 5.2B), which is not significantly improved compared with the wild type host strain carrying pZE-EP. Meanwhile, we observed that the final cell density of this mutant strain was quite low (OD_{600} = 2.5), suggesting that the simultaneous disruption of *pheA*, *tyrA* and *trpD* impaired the cell growth.

Then we turned to another ATCC31884 derived strain QH4 with *pheA* and *tyrA* disrupted. QH4 carrying pZE-EP produced 778.16 mg/L SA by the end of 48 h. As we expected, the dramatic increase in SA production was accompanied by the improvement of cell growth (final OD_{600} = 4.6). Further, to test the impact of plasmid copy number on SA production, we constructed another two plasmids pCS-EP (medium copy number) and pSA-EP (low copy number). As shown in Figure 2B, QH4 carrying pCS-EP and pSA-EP produced 621.05 and 207.20 mg/L SA, respectively, indicating that the reduced copy number of plasmids used for expressing EntC and PchB resulted in lower production of SA. Therefore, high-level expression of the pathway enzymes is preferred to redirect more carbon flux into the artificial pathway. On this basis, we further boosted the availability of chorismate by eliminating the bottlenecks associated with the shikimate pathway. For this purpose, we employed a previously constructed chorismate-boosting plasmid expressing *aroL*, *ppsA*, *tktA*, and the feedback inhibition resistant mutant of *aroG* (*aroG*^{fbr}) (Figure 5.1), namely pCS-APTA (medium copy number) [81]. When pCS-APTA was co-transferred with pZE-EP into QH4, the resulting strain produced 1179.92 mg/L of SA at 48 h, a 51.6% increase in titer compared with its parent strain (QH4 containing pZE-EP). To our knowledge, this is the

highest titer for SA production via microbial production approaches. However, when the APTA expression cassette was moved into the high-copy-number plasmid pZE-EP resulting in pZE-EP-APTA, the SA production of QH4 carrying pZE-EP-APTA (813.97 mg/L) was not improved significantly (Figure 5.2B) compared with that of QH4 carrying pZE-EP alone. SA production for all the strains followed the growth-dependent pattern.

In addition, to evaluate the toxicity of SA against the host strain, a growth inhibition assay was conducted. We added SA into the cell cultures at different final concentrations ranging from 0-1000 mg/L and detected its impact on cell growth. As shown in Figure 3C, SA exhibited toxicity against QH4 cells especially when its concentration exceeded 500 mg/L. 1000 mg/L SA completely inhibited the cell proliferation. Notably, strain QH4 carrying pZE-EP and pCS-APTA produced 1179.92 mg/L SA. This result suggested that on one hand QH4 can develop some degree of tolerance towards SA when the cells were initially exposed to low concentration of the produced SA; on the other hand, this strain has probably reached its maximum SA production capacity, which was constrained by its SA tolerance.

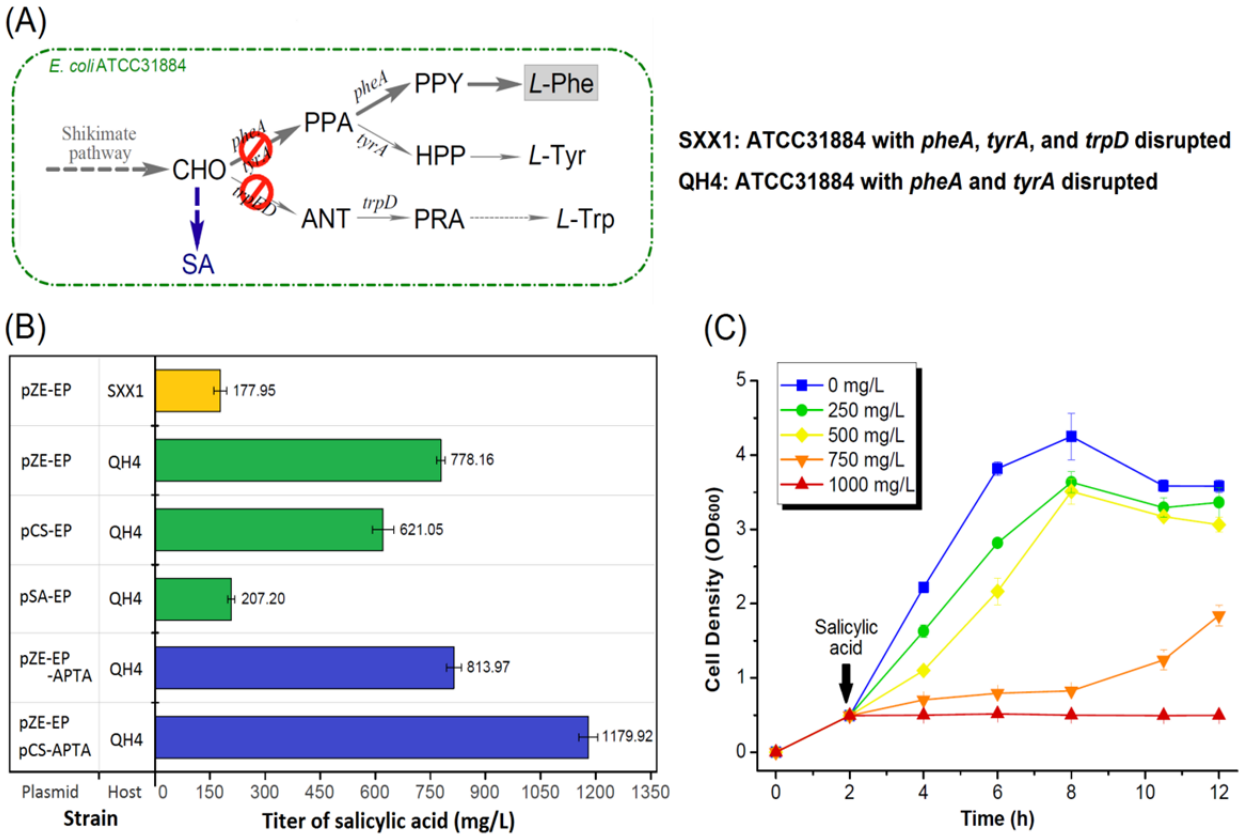


Figure 5.2 Transformation of a phenylalanine producing strain into an SA overproducer. (A) Schematic representation of the aromatic amino acid biosynthetic pathways in *E. coli*. Grey-colored arrows refer to the native carbon flow of *E. coli* strain ATCC31884, while the blue-colored arrow indicates the desired carbon flow after metabolic engineering. CHO, chorismate; PPA, prephenate; PPY, phenylpyruvate; HPP, 4-hydroxyphenylpyruvate; ANT, anthranilate; PRA, 5-phosphoribosyl-anthranilate. (B) Production of SA by metabolically engineered strains. pZE-EP, pCS-EP, and pSA-EP indicate that EntC and PchB are co-expressed using high-, medium-, and low-copy plasmids, respectively. All data points are reported as mean ± s.d. from three independent experiments. (C) Growth inhibition assay to evaluate the toxicity of SA towards QH4 strain. All data points are reported as mean ± s.d. from two independent experiments.

5.4.3 Screening for an efficient SMO

Following the reconstitution and optimization of an SA biosynthetic pathway, our subsequent effort was focused on constructing the SA degradation pathway. SMO catalyzes the first enzymatic step of SA degradation, and thus it is of great importance to have an efficient SMO available for the pathway assembly. Pseudomonas species (for example, *P. putida*) are known to demonstrate diverse metabolism, especially their capability of degrading a variety of aromatic hydrocarbon pollutants [83]. A putative SMO (encoded by *nahG*) was identified from the genome of *P. putida* DOT-T1E. For the screening purpose, we synthesized a codon-optimized gene of $nahG^{opt}$. In addition, we have previously reported two anthranilate 1,2-dioxygenases (ADOs) from *P. aeruginosa* PAO1 (encoded by *paantABC*) and *P. fluorescens Migula* (encoded by *pfantABC*) and a benzoate 1,2-dioxygenase (BDO) from *P. putida* KT2440 (encoded by *ppbenABCD*) that can catalyzed the conversion of anthranilate to catechol [73]. Given the structure similarity of the substrates anthranilate (2-aminobenzoate) and SA (2-hydroxybenzoate), these enzymes were also considered as candidates for SMO screening. To test their catalytic activities, the genes *paantABC, pfantABC, ppBenABCD,* and $nahG^{opt}$ were cloned into a high-copy-number plasmid pZE12-luc separately, resulting in the corresponding expression vectors used for *in vivo* assays. The wild type *E. coli* cells carrying these expression vectors were cultivated and then fed with 2 mM SA. Their activities were estimated by measuring the catechol formation rates. As indicated in Table 5.2, the SMO encoded by $nahG^{opt}$ showed the highest *in vivo* activity (54.78 μM/min/OD) among all the tested enzymes, while the ADOs and BDO also exhibited activity towards SA to generate catechol, although their catalytic activities were much lower.

Furthermore, to measure the kinetic parameters of the SMO encoded by $nahG^{opt}$, a multi-histidine tag was fused its N-terminus, and the recombinant protein was purified. The result of *in vitro* enzyme assay showed that the K_m and k_{cat} value of the SMO were 4.37 μM and 96.56 s^{-1}, respectively, indicating its high substrate affinity and catalytic efficiency (Figure 5.3A).

Table 5.2 *In vivo* activity of salicylate 1-monoxygenases

Genes	Source	***In vivo* activity (μM/min/OD)**
paantABC	*P. aeruginosa*	0.17 ± 0.01
pfantABC	*P. fluorescence*	0.16 ± 0.01
ppbenABCD	*P. putida*	1.89 ± 0.02
nahG^{opt}	*P. putida*	54.78 ± 1.65

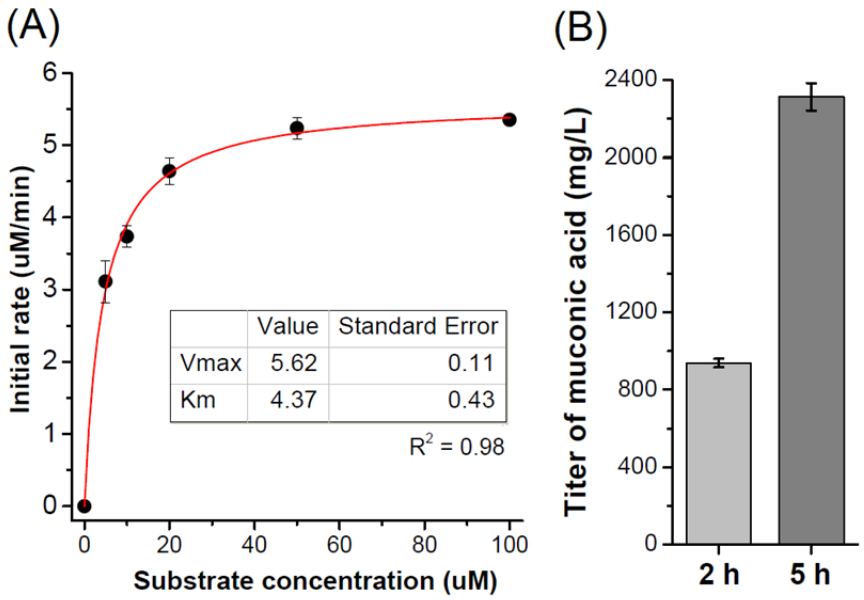


Figure 5.3 Activity of the salicylate 1-monoxygase (SMO) encoded by *nahG*^{opt}**.** (A) Kinetic parameters of the SMO. The K_m and V_{max} values were estimated with OriginPro8 through non-linear regression of the Michaelis-Menten equation. The protein concentration [E] of the SMO in the reaction systems was 0.97 nM. The k_{cat} value was calculated according to the formula $k_{cat}= V_{max}/[E]$. All data points are reported as mean ± s.d. from two independent experiments. (B) Conversion of SA to MA using the wild type *E. coli* strain carrying pZE-NC. Data points are reported as mean ± s.d. from three independent experiments. Error bars are defined as s.d.

5.4.4 Bioconversion of SA to MA

Conversion of SA to MA is a part of the SA degradation pathway, which involves the action of SMO and CDO. Given that an efficient CDO from *P. putida* KT-2440 (encoded by *catA*) has been identified previously in our lab [73], we aimed to assemble this pathway by co-expressing the SMO and the CDO in *E. coli*. For this purpose, the genes $nahG^{opt}$ and *catA* were cloned into a high-copy-number plasmid as an operon, yielding an expression vector pZE-NC. To explore its potential in the bioconversion of SA to MA, the wild type *E. coli* strain carrying pZE-NC was pre-cultivated in LB medium till the OD_{600} values reached 4 - 4.5 and then transferred to M9Y medium containing 3 mM SA. Considering the toxicity issue, the concentration of SA was maintained below 500 mg/L (3.62 mM) by continuous feeding at 3 mM/h. By the end of 2 and 5 h, the titers of MA reached 938.4 and 2313.0 mg/L, respectively (Figure 5.3B), indicating the high conversion rate from SA to MA with this plasmid construct.

5.4.5 Efficient *de novo* MA production via modular optimization

With the well-constructed SA biosynthesis and partial degradation pathways, we assembled the complete MA producing pathway by introducing the two modules simultaneously into *E. coli* using plasmid pZE-EP-NC, a high-copy expression vector with two operons. To our surprise, *E. coli* QH4 carrying pZE-EP-NC (Table 5.3, strain LS-1) only produced 5.27 mg/L MA after 48 h cultivation; while intermediates (SA and catechol) were not accumulated in the cultures. Besides, we observed that this strain underwent a longer lag phase before entering the exponential phase. We reasoned that the two modules were co-expressed in an unharmonious manner, which apparently exerted negative influence on the cell viability and productivity. To address this issue, we performed modular optimization to adjust the relative expression levels of each module. In this case, we decomposed the whole pathway into three modules (Figure 5.4A), namely EP module (expressing *entC* and *pchB* responsible for SA biosynthesis), NC module (expressing $nahG^{opt}$ and *catA* responsible for converting SA to MA) and APTA module (expressing *aroL*, *ppsA*, *tktA,* and $aroG^{fbr}$ to increase chorismate availability). To ensure the maximum

carbon flux towards the artificial pathway, the EP module was always fixed on high-copy plasmids. The expression level of NC module was optimized using high-, medium-, and low-copy plasmids. For APTA module, we attempted the high- and medium-copy expression in addition to the native expression (single copy). The plasmid combinations used for the modular optimization were listed in Table 5.3. As the result shown in Table 5.3 and Figure 5.4B, compared with APTA module, NC module had much more impact on the performance of MA production. Expression of the NC module on a low-copy plasmid resulted in dramatically improved MA production, suggesting that relatively lower expression level of the NC module was more beneficial for this artificial pathway. Furthermore, over-expression of APTA module further enhanced the titer by about 20% (from LS6 to LS7 and LS8). Placing APTA module on the high-copy plasmid resulted in slightly higher MA production (1453.64 mg/L) than that on the medium-copy plasmid (1425.71 mg/L). Through modular optimization, the MA titer was improved by 275 folds compared with that of the initial test with strain LS-1. For all the MA producing strains, no accumulation of SA and catechol was observed, indicating the high robustness of NC module. We also analyzed the by-product accumulation and carbon source consumption for the best two strains LS-7 and LS-8. By the end of 48 h, Strain LS-8 consumed all the glucose (2.5 g/L) and glycerol (10 g/L), while strain LS-7 consumed all the glucose but left around 3 g/L glycerol unconsumed. Acetate was the major by-product for both strains (1.5-2.5 g/L).

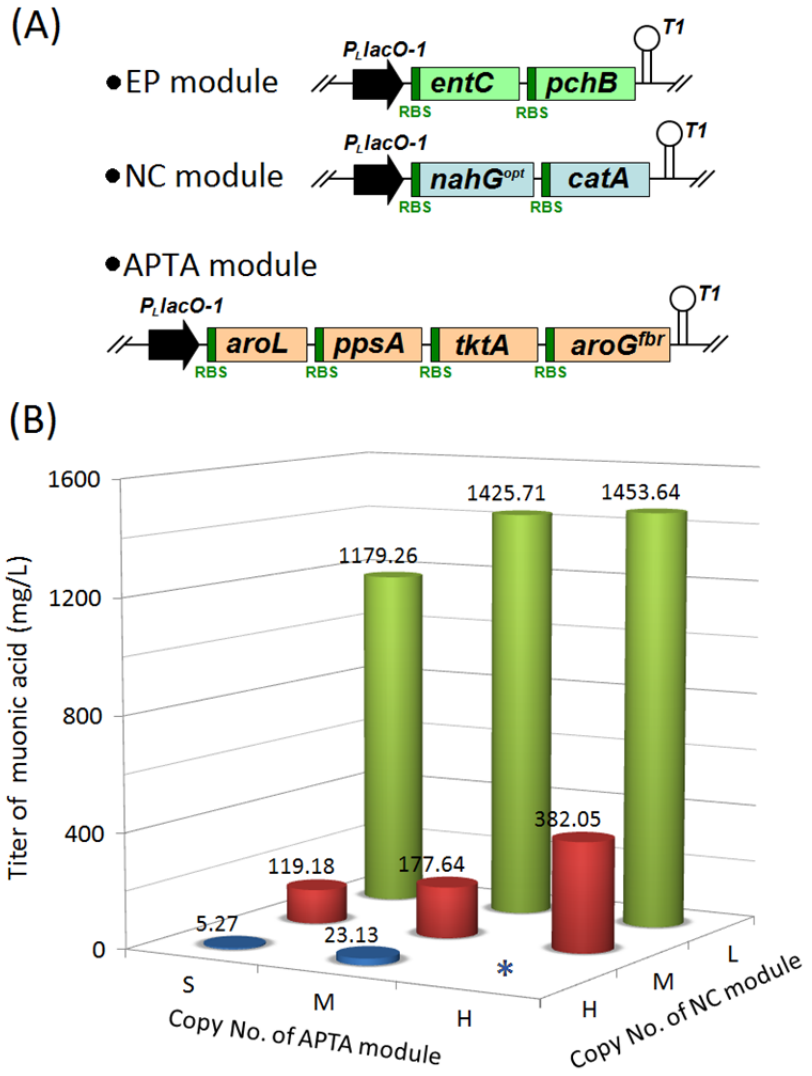


Figure 5.4 Modular optimization of the MA biosynthetic pathway. (A) Gene organization of the three modules: EP, NC, and APTA. *entC*, *pchB*, *nahG*^{opt} and *catA* encode ICS, IPL, SMO and CDO, respectively. (B) Optimization of MA production by adjusting the copy number of each module. All data points are reported as mean from three independent experiments. Asterisk mark indicates an untested construct.

Table 5.3 Combinations of plasmids for modular optimization

Strain name	Module Copy No.			Plasmid combinations	Titers (mg/L)
	EP	NC	APTA		
LS-1	H	H	S	pZE-EP-NC	5.27 ± 1.14
LS-2	H	H	M	pZE-EP-NC, pCS-APTA	23.13 ± 23.06
LS-3	H	M	S	pZE-EP, pCS-NC	119.18 ± 23.15
LS-4	H	M	M	pZE-EP, pCS-NC-APTA	177.64 ± 33.58
LS-5	H	M	H	pZE-EP-APTA, pCS-NC	382.05 ± 30.62
LS-6	H	L	S	pZE-EP, pSA-NC	1179.26 ± 61.58
LS-7	H	L	M	pZE-EP, pSA-NC, pCS-APTA	1425.71 ± 41.08
LS-8	H	L	H	pZE-EP-APTA, pSA-NC	1453.64 ± 98.88

*H, M and L indicate high-, medium-, and low-copy number plasmids;

S refers to native expression with a single copy in the *E. coli* genome

5.5 Discussion

The extreme diversity of metabolism in living organisms generates an enormous number of natural products and metabolic intermediates, a substantial portion of which are of industrial and/or pharmaceutical importance. Over the past decades, metabolic engineering promoted the microbial production of a variety of valuable molecules, including bio-fuels, bulk chemicals, and pharmaceuticals [85-88]. In this study, we devised a novel artificial pathway for the efficient production of MA by bridging the SA biosynthesis with its partial degradation pathway. In fact, biosynthesis and degradation of a specific molecule usually do not occur simultaneously in nature, since from the perspective of energy, it is not metabolically economic for organisms to survive nutrient-poor environment. But from the perspective of microbial production, it is a feasible strategy to link a degradation pathway to certain biosynthetic pathway and further to the microbial host's native metabolism, which can lead to the

expansion of native metabolism and *de novo* production of a valuable degradation intermediate from inexpensive and renewable carbon sources. In addition to MA, we think this design strategy can also be generalized to establish the artificial biosynthesis of other degradation intermediates.

Shikimate pathway is the only route in bacteria leading to the biosynthesis of aromatic compounds. Several intermediates can serve as the branch points for the biosynthesis of other aromatic compounds. Before this study, two artificial pathways derived from shikimate pathway have been reported for the microbial production of MA. Draths and Frost reported the first pathway that shunts the shikimate pathway via 3-dehydroshikimate by introducing three heterologous enzymes. To ensure high yield, the competing pathway that consumes 3-dehydroshikimate was deleted. However, the knockout of shikimate dehydrogenase (*aroE*) disrupted not only the biosynthesis of aromatic amino acids but also the formation of other important aromatic molecules, such as 4-hydroxybenzoate, the precursor of respiratory chain component, ubiquinol. Therefore, six supplements (*L*-phenylalanine, *L*-tyrosine, *L*-tryptophan, *p*-aminobenzoic acid, *p*-hydroxybenzoic acid, and 2,3-dihydroxybenzoic acid) were added in the fermentation media to maintain the cell growth and high productivity, which increased the cost of medium. Recently, our group reported another MA producing pathway that shunts the tryptophan biosynthesis from anthranilate. This pathway can keep the integrity of shikimate pathway, but involves a rate-limiting transamination step, conversion of chorismate towards anthranilate. This reaction requires the participation of glutamine, a less abundant amino acid, which limited the efficiency of the whole pathway. Comparatively, the pathway developed in this study did not disrupt the shikimate pathway either; meanwhile, all the catalytic steps in this pathway are very efficient. Although phenylalanine and tyrosine biosynthesis was deleted, cell growth can be easily restored by supplementing small amount of yeast extract (1 g/L). In addition, chorismate is an abundant precursor which led to the reported high-yield production of phenylalanine, tyrosine, etc, indicating that the carbon flux via this pathway can reach a high level if further production optimization is performed.

In general, current metabolic engineering approaches towards the reconstitution of artificial pathways in heterologous hosts involve very limited regulatory mechanism, which frequently brings

undesired effects. On one hand, unregulated expression of heterologous enzymes can cause cell toxicity and metabolic imbalance because of either the enzymes themselves or the toxic intermediates and products the enzymes generate. On the other hand, excessive expression of pathway enzymes can result in the waste of cellular resources and bring metabolic burden to the hosts, which may limit yield and productivity. Therefore, it is desirable to control the expression of pathway enzymes at a proper and balanced level. In recent years, the development of synthetic biology tools enabled the fine-tuning of protein expression level, including the adjustment of gene copy number [81], promoter strength [89], mRNA stability [90] and RBS binding efficiency[91]. In addition, dynamic regulatory circuits were also developed to control pathway enzyme expression in response to the accumulation of certain intermediates [92,93]. These strategies were proved to be very effective for those relatively simple metabolic pathways. However, as artificial biosynthetic pathways become more complex, an increasing number of enzymes are involved in the pathway assembly. It becomes much more laborious and time-consuming to exhaustively explore the optimal expression level for each of the pathway enzymes simultaneously. In this study, we employed a simplified method of modular optimization to balance pathway enzyme expression. Using this approach, a multi-step pathway can be decomposed into several modules which can be initially optimized individually. Then the relative expression level of each module can be further adjusted in the context of the whole pathway. This strategy greatly reduces the number of engineering targets and combinations and has been used to achieve the efficient production of a variety of molecules [81,94-96]. In this work, the modular optimization enabled the improvement of MA production by 275 folds, which demonstrated its generalizable potential to the optimization of other complex pathways.

In conclusion, we established a novel MA biosynthetic pathway by connecting the SA biosynthesis with its partial degradation pathway and achieved the efficient production of MA and its precursor SA. However, cellular toxicity remains to be a challenge that limits the production of these chemicals, especially for SA. To address this issue, it is necessary to seek and develop more resistant *E. coli* strains. A recent work reported the engineering of efflux pumps which successfully improved the tolerance of *E. coli* toward non-native product n-butanol [97]. Besides, some microbial species other than

E. coli exhibit high tolerance towards toxic chemicals, such as *Pseudomonas*. Given the increasing availability of genetic manipulation tools, it becomes more feasible to transfer and engineer the demonstrated pathways into these microorganisms.

CHAPTER 6

EFFICIENT PRODUCTION OF 5-HYDROXYTRYPTOPHAN THROUGH COMBINATORIAL METABOLIC AND PROTEIN ENGINEERING[5]

[5] Yuheng Lin, Xinxiao Sun, Qipeng Yuan and Yajun Yan. 2014, *ACS Synthetic Biology*, 3: 497-505
Reprinted here with permission of the publisher.

6.1 Abstract

5-Hydroxytryptophan (5-HTP) is a clinically effective drug against depression, insomnia, obesity, chronic headaches, etc. It is only commercially produced by the extraction from the seeds of *Griffonia simplicifolia* due to lack of synthetic methods. Here, we report the efficient microbial production of 5-HTP via combinatorial protein and metabolic engineering approaches. First, we reconstituted and screened prokaryotic phenylalanine 4-hydroxylase (P4H) activity in *Escherichia coli.* Then, sequence and structure-based protein engineering dramatically shifted its substrate preference, allowing for efficient conversion of tryptophan into 5-HTP. Importantly, *E. coli* endogenous tetrahydromonapterin (MH4) was able to be utilized as the coenzyme when a foreign MH4 recycling mechanismwas introduced. Whole-cell bioconversion enabled the high-level production of5-HTP (1.1-1.2 g/L) from tryptophan in shake flasks. On its basis, metabolic engineering efforts were further made to achieve the *de novo* 5-HTP biosynthesis from glucose.This work not only holds great scale-up potential but also demonstrates a strategy to expand native metabolism of microorganisms.

6.2 Background

World Health Organization (WHO) reported that depression is a common mental disorder affecting more than 350 million people globally. At the worst, it results in suicide with an estimated number of 1 million per year. Unfortunately, over 50% of sufferers over the world (over 90% in some regions) have neverreceived medical treatment [98]. Alterations inserotonin (5-hydroxytryptamine) metabolism were thought to be an important physiological factor for depression [99]. Dysfunction of the serotonergic mechanism in the central nervous system (CNS) has been implicated in the etiology of depression [100]. However, supply of serotonin via oral administration is not clinically effective against depression because it cannot pass through the brain-blood barrier. Unlike the conventional antidepressants (e.g. selective serotonin re-uptake inhibitors) acting on minimizing serotonin loss, 5-hydroxytryptophan (5-HTP) functions as the direct precursor to increase serotonin supply. Orally administered 5-HTP can easily pass through the blood-brain barrier not requiring transport molecules. Then it can be effectively

converted into serotonin in the CNS by endogenous decarboxylase [100]. Detailed clinical trials have demonstrated its efficacy in alleviating depression symptoms. Meanwhile, the therapeutic administration of 5-HTP has been shown to be effective in treating insomnia, fibromyalgia, obesity, cerebellar ataxia, and chronic headaches [101]. Importantly, relatively few adverse effects are associated with its use in treatment [99]. In most European countries, 5-HTP is a commonly prescribed drug for multiple treatment purposes; while in North America it is sold as an "over-the-counter" dietary supplement.

Due to the difficulty in region-selective hydroxylation of tryptophan via chemical approaches, the commercial production of 5-HTP only relies on the isolation from the seeds of *Griffonia simplicifolia*, a woody climbing shrub grown in West and Central Africa [100,101]. The season- and region-dependent supply of the raw materials has been largely limiting its cost-effective production and broad clinical applications. In past decades, the development of metabolic engineering and protein engineering in combination with fundamental genetics, biochemistry and bioinformatics provides new strategies to synthesize natural and non-natural molecules using microbial systems. On the basis of accumulated knowledge on natural biosynthetic mechanisms of target products, especially the genetic and biochemical information of the involved enzymes, heterologous enzymatic reactions can be reconstituted, modified and optimized in genetically superior microbial hosts to achieve efficient production of pharmaceutically important compounds that are scarce in nature [70,71,81,102-105].

5-HTP is natively produced in human and animals from *L*-tryptophan by the action of tryptophan 5-hydroxylase (T5H) and then converted to the neurotransmitter serotonin under normal physiological conditions [101]. T5Hs belong to the class of pterin-dependent aromatic amino acid hydroxylases (AAAHs) that also include two other subgroups: phenylalanine 4-hydroxylases (P4Hs) and tyrosine 3-hydroxylases (T3Hs). AAAHs were broadly identified and extensively studied in animals due to their close relationship with human diseases such as phenylkentonuria, Parkinson's disease, and neuropsychiatric disorders [106]. These enzymes consist of three domains that are the N-terminal regulatory domain, the central catalytic domain, and the C-terminal domain involved in tetramer formation, and usuallyutilize tetrahydrobiopterin (BH4) as the coenzyme (or co-substrate) [107]. Animal

T5Hs were proved to be unstable and hard to be functionally expressed in a microbial host [108,109]. A very recent patent reported the use of truncated T5H1 from *Oryctolagus cuniculus*, which produced up to 0.9 mM (equivalent to 198 mg/L) of 5-HTPfrom tryptophan in *Escherichia coli*. To supply the pterin coenzyme, the animal BH4 biosynthetic pathway coupled with a regeneration system including a total of five enzymes was required to be co-expressed in *E. coli* [110]. However, the production efficiency is still not satisfying for scale-up production.

A few AAAHs were also found in bacteria such as *Pseudomonas* and *Chromobacterium* species [111,112]. So far, all of them were identified as P4Hs with little activity for tryptophan hydroxylation; but such activity was reported to be improved *in vitro* when mutations were introduced into the P4H from *Chromobacterium violaceum* [113]. Prokaryotic P4Hs consist of only one domain corresponding to the catalytic domain of animal AAAHs [107]. Recent experimental evidence indicated that bacterial P4Hs may utilizete trahydromonapterin (MH4) instead of BH4 as the native pterin coenzyme [114], since BH4 does not naturally occur in most bacteria. Interestingly, MH4 is the major form of pterin in *E.coli,* although its function is still unknown. In this work, we report the reconstitution of bacterialP4H activity in *E. coli* through utilization and recycling of its endogenous MH4. Combined bioprospecting and protein engineering approaches enabled the development of the P4H mutants that are highly active on converting tryptophan to 5-HTP, which allowed the establishment of an efficient 5-HTP production platform via further metabolic engineering efforts. This *de novo* process does not require supplementation of expensive pterin co-factors or precursors but only utilizes renewable simple carbon sources, which holds great potential for scale-up production of 5-HTP in microorganisms.

6.3 Materials and Methods

6.3.1 Experimental materials

E. coli XL1-Blue was employed as the host strain for cloning and plasmid propagation; *E. coli* BW25113*ΔtnaA*was used as the host strain for *in vivo* enzyme assays, feeding experiments, and *de novo* production of tryptophan and 5-HTP. QH4 was previously constructed with the disruption of *pheLA* and

tyrA from a phenylalanine producer *E. coli* ATCC31884 [115]. Luria-Bertani (LB) medium containing 10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl was used for cell cultivation and enzyme expression. M9 minimal medium containing 5 g/L glycerol, 6 g/L Na_2HPO_4, 0.5 g/L NaCl, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, 246.5 mg/L $MgSO_4 \cdot 7H_2O$, 14.7 mg/L $CaCl_2$ and 27.8 mg/L $FeSO_4 \cdot 7H_2O$, was used for *in vivo* assays of P4Hs. Modified M9 (M9Y) medium was used for *de novo* production of tryptophan and 5-HTP. M9Y medium contains 10 g/L glucose, 6 g/L Na_2HPO4, 0.5 g/L NaCl, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, 246.5 mg/L $MgSO_4 \cdot 7H_2O$, 14.7 mg/L $CaCl_2 \cdot 2H_2O$, 27.8 mg/L $FeSO_4 \cdot 7H_2O$, 2 g/L yeast extract and 2 g/L sodium citrate dihydrate. When necessary, kanamycin, ampicillin and/or chloramphenicol were supplemented to the media at the final concentration of 50, 100 and 34 mgl^{-1}, respectively. pZE12-luc, pCS27, pSA74 are high-, medium- and low-copy number plasmids [116], respectively, which were used for expressing enzymes in *E. coli*. Details of the strains and plasmids used in this study are depicted in Table 6.1.

Table 6.1 Strains and plasmids used in Chapter 5

Strain	**Genotype**	**Source**
XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F′ *proAB lacI*^{q}*ZΔM15 Tn10* (Tet^r)]	Stratagene
JW3686-7	F-, *Δ(araD-araB)567*, *ΔlacZ4787*(::rrnB-3), *λ-*, *rph-1*, *ΔtnaA739::kan*, *Δ(rhaD-rhaB)568*, *hsdR514*	CGSC
QH4	*E. coli* ATCC31884 with *pheLA* and *tyrA* disrupted	[69]
BW25113*ΔtnaA*	JW3686-7 with *kan* deleted	This study
QH4	QH4 with *tnaA* deleted	This study
Plasmids	**Description**	**Reference**
pZE12-luc	$P_L$lacO1, *colE*ori, *luc*, Amp^r	[81]
pCS27	$P_L$lacO1, P15A ori, Kan^r	[81]

pSA74	$P_L$lacO1, pSC101 ori, Cm^r	[81]
pZE-PaphhA	pZE12-luc containing *phhA* from *P. aeruginosa* PAO1	This study
pZE-PaABM	pZE12-luc containing*phhA and phhB* from *P. aeruginosa* PAO1, and *folM* from *E. coli* MG1655	This study
pZE-PfABM	pZE12-luc containing*phhA* from *P. fluorescens* Migula*, phhB* from *P. aeruginosa* PAO1, and *folM* from *E. coli* MG1655	This study
pZE-PpABM	pZE12-luc containing*phhA* from *P. putida* KT2440, *phhB* from *P. aeruginosa* PAO1, and *folM* from *E. coli* MG1655	This study
pZE-ReABM	pZE12-luc containing*phhA* from *R. eutropha* H16,*phhB* from *P. aeruginosa* PAO1, and *folM* from *E. coli* MG1655	This study
pZE-XcABM	pZE12-luc containing*phhA* from *X. campetris*ATCC 33913, *phhB* from *P. aeruginosa* PAO1, and *folM* from *E. coli* MG1655	This study
pZE-XcABMW179F	pZE-XcABM with mutation W179F on the P4H	This study
pSA-XcABM2Ma	pZE-XcABM with mutations W179F and L98Y on the P4H	This study
pZE-XcABM2Mb	pZE-XcABM with mutations W179F and Y231C on the P4H	This study
pZE-XcABM3M	pZE-XcABM with mutations W179F, L98Y and Y231C on the P4H	This study
pCS-XcABMW179F	pCS27 containing*phhA* (W179F) from *X. campetris* ATCC 33913, *phhB* from *P. aeruginosa* PAO1, and *folM* from *E. coli* MG1655	This study
pSA-TrpEDCBA	pSA74 containing *trpEDCBA* with S40F onTrpE	This study

6.3.2 DNA manipulation

E. coli strain BW25113 *ΔtnaA::kan* (JW3686-7) was purchased from Coli Genetic Stock Center (GCSC). The kanamycin resistant marker was deleted according to the reported protocol [117]. Deletion of the*tnaA*gene from QH4 was performed using the reported Red disruption method [117]. Plasmid pZE-PaphhA was constructed by inserting the amplified *phhA* gene from *Pseudomonas aeruginosa*into

pZE12-luc using restriction sites Acc65I and XbaI. pZE-PaABM was constructed by inserting *phhA* and *phhB*from *P. aeruginosa* and *folM* from *E. coli* into pZE12-luc via multi-piece ligation using Acc65I/NdeI, NdeI/HindIII and HindIII/XbaI. pZE-PpABM, pZE-PfABM, pZE-ReABM and pZE-XcABM were constructed using the same approach with the respective *phhA* genes in place of the *phhA*gene from *P. aeruginosa*. pSA-trpEDCBA was constructed by inserting the DNA fragment of *trpEDCBA*from *E. coli*into pSA74 using Acc65I and BamHI. Site-directed mutagenesis was conducted by overlap PCR. Plasmids pZE-XcABMW179F, pZE-XcABM2Ma, pZE-XcABM2Mb, pZE-XcABM3M were constructed by replacing the wild type *phhA*gene from *Xanthomonas campetris* with the respective mutant genes (Table 6.1)

6.3.3 Construction of phylogenetic tree and homology modeling

The AAAH sequences were randomly picked through Genbank using "phenylalanine 4-hydroxylase", "tyrosine 3-hydroxylase" and "tryptophan 5-hydroxylase" as the search key words. The alignment of the AAAH Amino acid sequences was conducted by using ClustalX 2.1. The phylogenetic tree was constructed by Molecular Evolutionary Genetics Analysis (MEGA) version 5.02 using the neighbor-joining method [118]. Bootstrapping test was performed to evaluate the reliability (1000 replicates). All other used parameters were the default of the software. The homology model of XcP4H was built with the SWISS-MODEL online server by using the crystal structure of the P4H from *C. violaceum* (PDB code 3TK2) as a template.

6.3.4 *In vivo* assays of wild-type and mutant P4Hs

E. coli BW25113*ΔtnaA* carrying pZE-PaABM was inoculated in 50 ml of LB liquid medium containing 0.5 mM of IPTG and 100 $\mu g ml^{-1}$ of ampicillin, and grown aerobically at 37^{o}C for about 8 h till OD_{600} reached 4.5-5.5. Then the cells were harvested, and re-suspended in the M9 minimal medium (OD_{600} = 4.5-5.5). After adaption for 20 min, phenylalanine or tryptophan was added into the cell suspension to a final concentration of 500 mg/L. At the same time, 1 mM of ascorbic acid was added to

avoid the product oxidation. The flasks were incubated with shaking (300 rpm) at 37 ^{o}C for 1 h. Subsequently, samples were taken by removing cell pellets and the products (tyrosine and tryptophan) were quantitatively measured with HPLC. The same method was used to measure the *in vivo* activities of other P4Hs and XcP4H mutants. The *in vivo* activities of P4Hs were expressed as μM/min/OD_{600}.

6.3.5 Bioconversion of tryptophan to 5-HTP

E. coli strain BW25113*ΔtnaA* was transformed with plasmid pZE-XcABMW179F. Single colonies were inoculated into 50 ml LB medium containing 0. 5 mM of IPTG and grown aerobically at 37^{o}C for about 8 h till OD_{600} reached around 5.0. Then cells were harvested, re-suspended in 20 ml of M9Y medium (at OD_{600} =12-13) containing 2g/L of tryptophan and left to grow at 30 and 37^{o}C. Samples were taken at 2 h, 5 h, 10 h and 16 h. The concentrations of produced 5-HTP were analyzed by HPLC.

6.3.6 *De novo* production of tryptophan and 5-HTP

Overnight LB cultures of the producing strains were inoculated at 2% into the M9Y medium containing appropriated antibiotics and cultivated at 30 and 37^{o}C with shaking at 300 rpm. IPTG was added to the cultures to a final concentration of 0.5 mM at 0 h. Samples were taken every 12 hours. The OD_{600} values were measured and the concentrations of the products, intermediates and by-products were analyzed by HPLC.

6.3.7 HPLC analysis.

L-tyrosine (from SIGMA ALDRICH), *L*-tryptophan (from SIGMA ALDRICH) and 5-HTP (fromAcros Organics) were used as standards. Both the standards and samples were quantified by HPLC (Dionex Ultimate 3000) equipped with a reverse phase ZORBAX SB-C18 column and an Ultimate 3000 Photodiode Array Detector. A gradient elution method was used according to our previous study [119]. Quantification of tryptophan, tyrosine and 5-HTP was based on the peak areas at specific wavelength (276 nm). Glucose, acetate and pyruvate were quantified using a previously described method [72].

6.4 Results

6.4.1 Phylogenetic analysis of AAAHs

Compared with animal AAAHs that include three sub-groups, their prokaryotic counterparts were all identified or annotated as P4H only. Previous biochemical and structural studies revealed that in addition to the central catalytic domain, animal AAAHs usually consist of two additional domains that are the N-terminal regulatory domain and the C-terminal domain involved in tetramer formation; while prokaryotic AAAHs (e.g. the P4H from *C. violaceum*) are monomers with only one single domain that shares moderate sequence similarity (about 30%) with the catalytic domains of animal AAAHs [120]. To explore the evolutionary relationship among AAAHs, 25 amino acid sequences from both prokaryotes and animals were randomly selected and a phylogenetic tree was constructed using MEGA 5.02 based on the neighbor-joining method (Figure 6.1)[118]. The tree reflects a considerable evolutionary separation between prokaryotic and animal AAAHs. The three subfamilies (P4Hs, T5Hs, and T3Hs) of animal AAAHs are distinctly separated as well, among which P4Hs show closer phylogenetic relationship with T5Hs than with T3Hs. These results are consistent with a previous phylogenetic study on AAAHs [121].

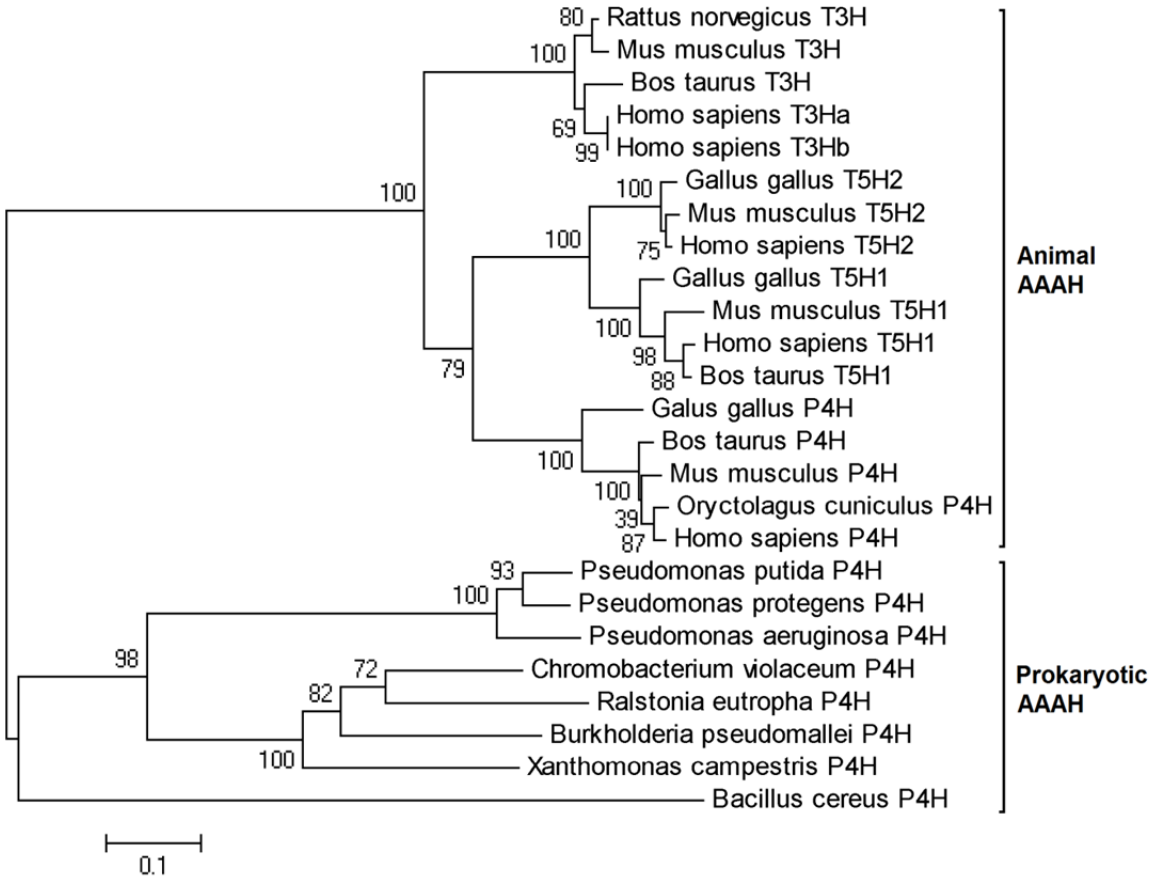


Figure 6.1 Phylogenetic relationships of prokaryotic P4Hs and animal AAAHs. The protein sequence alignment was performed using ClustalX 2.1. The phylogenetic tree was constructed with MEGA 5.02 by using the neighbor-joining method. The bootstrapping method was used for phylogeny test (1000 replications). The numbers associated with the branches refer to the bootstrap values representing the substitution frequencies per amino acid residue.

Considering the phylogenetic evidences in combination with the development of functional diversity, we inferred that the animal AAAHs were evolved from prokaryotic P4Hs through duplication and divergence. Therefore, we hypothesized that even after a long-term evolution process, prokaryotic and animal P4Hs may still share some conserved amino acid residues that determine their substrate preference towards phenylalanine. Meanwhile, animal P4Hs and T5Hs share high sequence similarity,

suggesting that the interchange of substrate preference from phenylalanine to tryptophan may only involve the substitution of a small number of residues. Based on these hypotheses, we speculated that by performing a comprehensive alignment analysis of the sequences of animal AAAHs and prokaryotic P4Hs, we may be able to identify the substrate-determining residues from the latter group and artificially evolve them into T5Hs.

6.4.2 Bio-prospection and reconstitution of prokaryotic P4Hs in *E. coli*

Before exploring the substrate-determining amino acid residues, we picked five P4Hs from different microorganisms (*P. aeruginosa*, *Pseudomonas putida*, *Pseudomonas fluorescence*, *Ralstonia eutropha* and *X. campestris*) to verify and compare their activities and substrate preferences, since most of the prokaryotic P4Hs are still putative enzymes without experimental confirmation of their function. The P4H from *P. aeruginosa* (PaP4H) was previously identified *in vitro* and its crystal structure has been resolved [122]. Some genetic and biochemical evidences suggested that PaP4H utilizes MH4 instead of BH4 as the native pterin coenzyme [114]. Thus, we first selected it as a prototype to establish its*in vivo* activity in *E. coli*, since MH4 is the major pterin produced by *E. coli* (Figure 6.2). To achieve the expression of PaP4H, its gene *phhA* was amplified from the genomic DNA of *P. aeruginosa* and cloned into a high-copy number plasmid under the control of an IPTG-inducible promoter P_L*lacO1*. The resulting expression vector pZE-PaphhA was introduced into *E. coli* strain BW25113*ΔtnaA* (abbreviated as BW*ΔtnaA*). Since tryptophanase encoded by *tnaA* was reported to catalyze the degradation of tryptophan and 5-HTP [123], the gene was knocked out from all the strains used in this study. We observed that the cell growth of BW*ΔtnaA* carrying pZE-PaphhA was significantly retarded. Its OD_{600} values only reached 0.8-1.0 after 8-hour cultivation, dramatically lower than those of the control strain (BW*ΔtnaA* carrying an empty vector) with OD_{600} values at 5.5-6.0. A similar effect was also observed in a previous study 124]. When the cells were incubated with phenylalanine (500 mg/L), almost no hydroxylated product (tyrosine) was detected. Indeed, *P. aeruginosa* possesses a pterin 4a-carbinolamine dehydratase (PCD, encoded by *phhB*) responsible for the regeneration of dihydromonapterin (MH2) which can be further reduced to

MH4. But *E. coli* does not have such a mechanism natively. To establish an artificial MH4 recycling system (Figure 6.2), *phhB* from *P. aeruginosa* and *folM* from *E.coli* (encoding dihydromonapterin reductase, DHMR) were co-expressed along with the *phhA* using the vector pZE-PaABM. Interestingly, the *E. coli* strain harboring this vector dramatically improved cell viability which was comparable with the control strain. Its OD_{600} values reached 4.5-5.5 after cultivation for 8 hours. When these cells were collected and incubated with phenylalanine, a large amount of tyrosine was produced at a rate of 83.50 $\mu M/h/OD_{600}$, as was shown in the *in vivo* assays (Table 6.2). These results indicated that introduction of the MH4 recycling system not only restored the cell growth but also enabled the *E. coli* strain to convert phenylalanine to tyrosine. Moreover, this strain was also capable of converting tryptophan into 5-HTP (Table 6.2), although the production rate (0.19 $\mu M/h/OD_{600}$) was much lower, only equivalent to 0.23% of that towards phenylalanine.

Table 6.2 *In vivo* activities of P4Hs from different microorganisms

Source of P4H	***In vivo* activity**[*] ($\mu Mh^{-1}OD_{600}^{-1}$)		**Preference** (Phe:Trp)
	Phenylalanine	**Tryptophan**	
Pseudomonas aeruginosa	83.50 ± 16.00	0.19 ± 0.02	439.5
Pseudomonas putida	76.32 ± 10.02	0.12 ± 0.03	636.0
Pseudomonas fluorescence	82.47 ± 12.05	0.20 ± 0.05	412.4
Ralstonia eutropha H16	73.33 ± 4.63	1.22 ± 0.04	60.1
Xanthomonas campestris	97.40 ± 4.42	2.91 ± 0.21	33.5

[*]All data are reported as mean ± s.d. from three independent experiments.

On this basis, another four P4Hs were also tested by replacing the PaP4H gene on pZE-PaABM with their respective genes. As shown in Table 6.2, all the identified P4Hs showed high activity and

strong substrate preference towards phenylalanine in *E. coli*. Among them, the P4H from *X. campestris* (XcP4H) exhibited the highest activity towards both phenylalanine and tryptophan. The three from *pseudomonas* species showed the most similar catalytic properties, which is consistent with their close phylogenetic relationship. Therefore, we confirmed that all the tested P4Hs can function well by utilizing the *E. coli* endogenous pterin coenzyme MH4 in the presence of a recycling system.

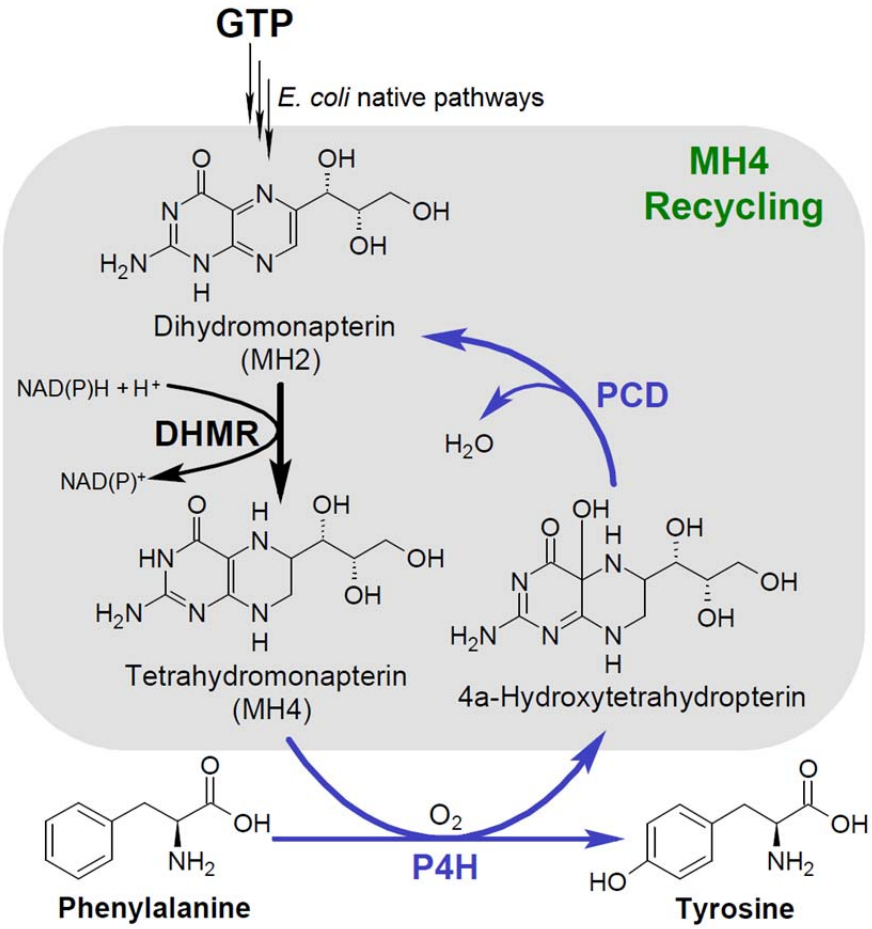


Figure 6.2 Reconstitution of prokaryotic P4H activity in *E. coli.* The black- and blue-colored arrows indicate the *E. coli* native pathways and heterologous reactions, respectively. Bold arrows refer to the over-expressed steps. The introduced MH4 recycling system is indicated by the grey-colored box. GTP, guanosine-5'-triphosphate; PCD, pterin-4α-carbinolamine dehydratase; DHMR, dihydromonapterin reductase; P4H, phenylalanine 4-hydroxylase.

6.4.3 Modification of the XcP4H substrate preference through protein engineering

XcP4H was selected for protein engineering due to its superior catalytic potential. To investigate the substrate-determining amino acid residues, its sequence was aligned with animal P4Hs and T5Hs. Comparison of the sequences of only animal P4Hs and T5Hs led to the identification of a number of residues that are conserved within each group but varied between groups. However, when these residues were aligned with the XcP4H sequence, only six of them were found to be conserved in P4Hs and probably critical to the substrate selectivity, which are Q85, L98, W179, L223, Y231, and L282 (Figure 6.3). To further investigate their locations in the enzyme structure, a homology model was built using the crystal structure of the P4H from *C. violaceum* (PDB code 3TK2) as a template. The conserved residues were well aligned with those in the crystal structures of the human P4H (PDB code 1MMK) and T5H1 (PDB code 3HF6), indicating the reliability of the model. In this structure, W179 islocated inside the catalytic pocket just at the predicted phenylalanine binding site, while L98 and Y231 are near the entrance to the pocket, which are closer to the coenzyme MH4 binding site (Figure 6.4A). However, Q85, L223 and L282 are not located near the catalytic pocket, suggesting these residues are less relevant to the enzyme's substrate selection. Therefore, we selected W179, L98 and Y231 as the targets for further mutation analysis. We hypothesized that if these residues were replaced with their respective residues in T5Hs that are F, Y and C, respectively, the mutants might exhibit stronger preference towards tryptophan. As a result, the W179F mutant of XcP4H exhibited a 17.4-fold increase in tryptophan hydroxylation activity compared with the wild-type (WT) enzyme; meanwhile its activity towards phenylalanine decreased by about 20%. The substrate preference towards phenylalanine over tryptophanwas shifted from 33.5 to 1.5 (Table 6.3). When the mutations L98Y or Y231C were combined with W179F, the substrate preference further shifted towards tryptophan, although their activities towards tryptophan were not as high as that of W179F alone. The triple mutant showed almost the same preference towards the two substrates (Figure 6.4B). As we mentioned, L98 and Y231 are closer to the MH4 binding site, suggesting that these two residues might not contribute to the aromatic amino acid substrate selection.

Table 6.3 *In vivo* activities and substrate preferences of XcP4H mutants

XcP4H mutants	Substrate				Preference
	Phenylalanine		Tryptophan		
	In vivo activity** ($\mu M\ h^{-1}OD_{600}^{-1}$)	R.A.* (%)	*In vivo* activity ($\mu M\ h^{-1}OD_{600}^{-1}$)	R.A. (%)	(Phe:Trp)
WT	97.40 ± 4.42	100	2.91 ± 0.21	100	33.5
W179F	78.05 ± 4.34	80	50.60 ± 4.72	1739	1.5
W179F/L98Y	44.49 ± 4.95	46	35.13 ± 1.67	1207	1.3
W179F/Y231C	50.92 ± 4.36	52	27.71 ± 2.99	952	1.8
W179F/L98Y/Y231C	16.56 ± 1.86	17	16.58 ± 2.59	570	1.0

*R.A., relative activity, setting the R.A. of WT XcP4H as 100%

**All data are reported as mean ± s.d. from three independent experiments.

To further explore the potential of XcP4H mutant W179F for whole-cell biocatalysis, feeding experiments were conducted by incubating pre-cultured *E. coli* cells harboring pZE-XcABMW179F (initial OD_{600} = 12-13) with 2.0 g/L of tryptophan. As shown in Figure 6.4C, the initial conversion rates were similar at 30 and 37 ^{o}C, although the cells grew slightly faster at 37^{o}C. However, the production efficiency at 30^{o}C became obviously higher after 5 hours. By the end of 16 hours, the cultures at 30 and 37^{o}C accumulated 1114.8 and 758.3 mg/Lof 5-HTP at the expense of 1503.2and 1417.1 mg/L tryptophan, respectively. Meanwhile, we observed that the color of the cultures gradually turned dark after 5 hours, especially at 37^{o}C, probably due to the oxidation of 5-HTP and tryptophan under aerobic conditions.

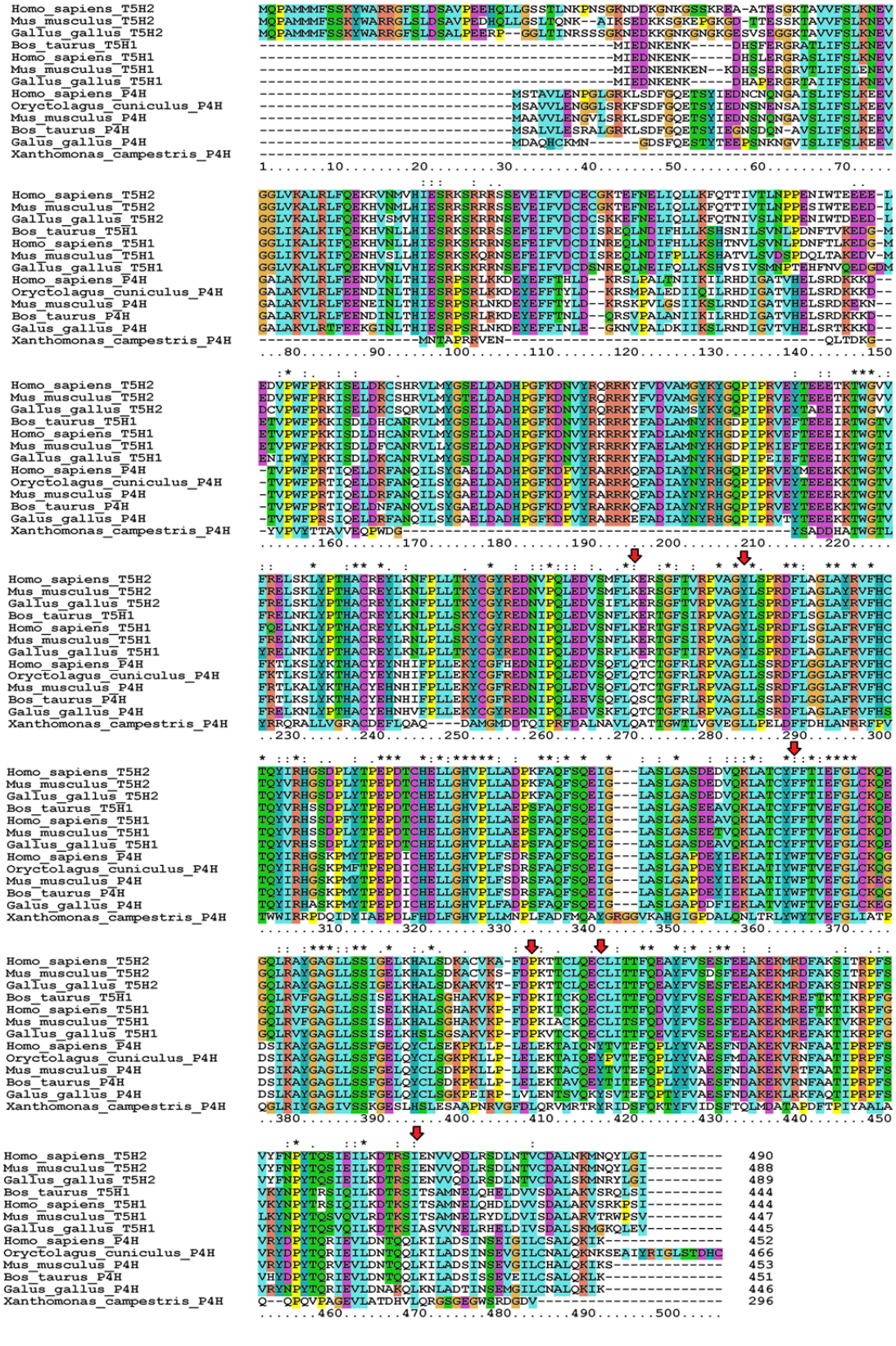


Figure 6.3 Amino acid sequence alignment of animal AAAHs and XcP4H.

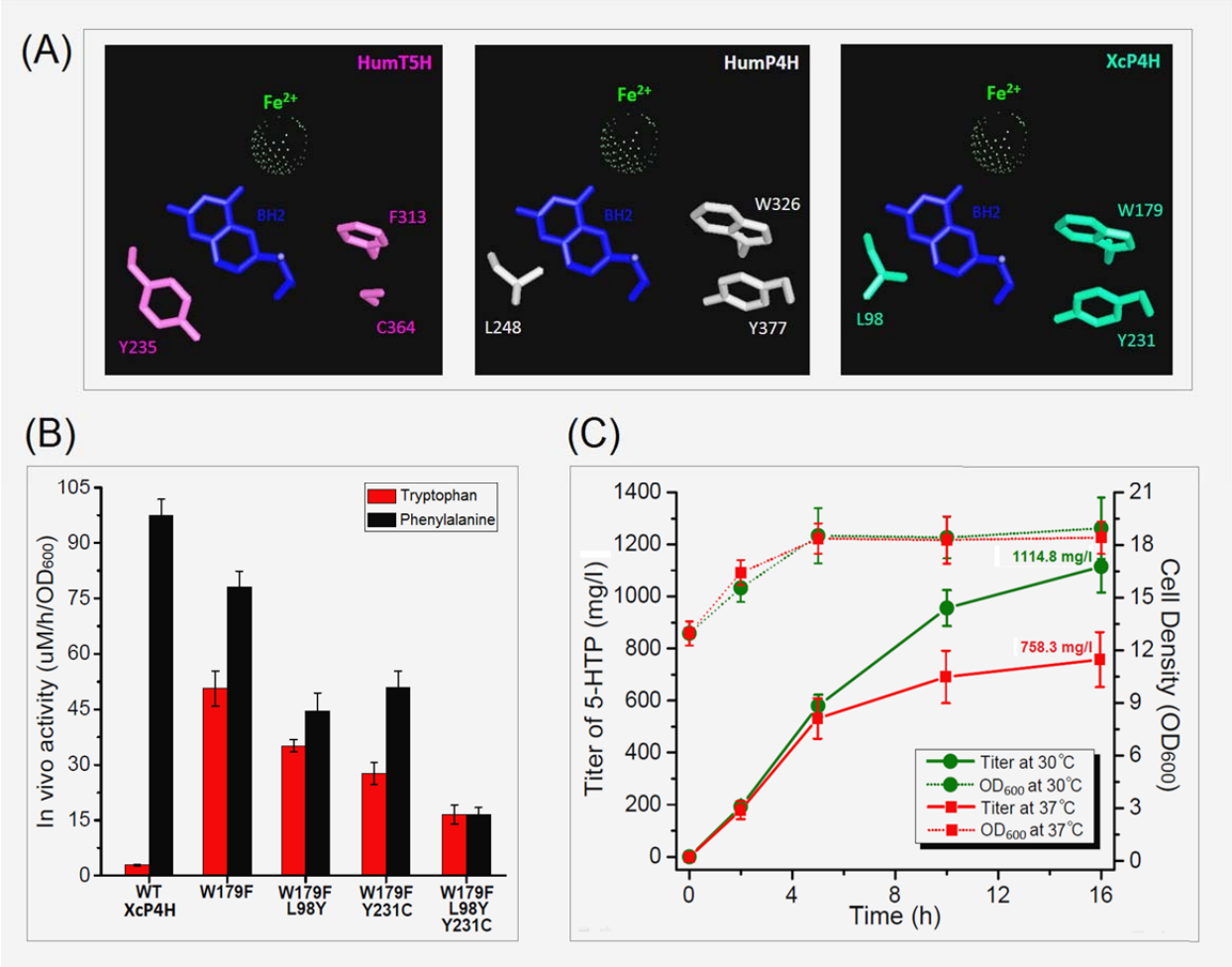


Figure 6.4 Modification of XcP4H via protein engineering. (A) Comparative illustration of the positions of the 3 critical residues (L98, W179 and Y231) in the structures of XcP4H, the P4H from human (HumP4H) and the T5H1 from human (HumT5H). (B) *In vivo* activities of wild type XcP4H and its mutants. Red- and black-colored bars indicate the activities towards tryptophan and phenylalanine, respectively. (C) Whole-cell bioconversion of tryptophan into 5-HTP using XcP4H mutant W179F. Solid and dotted lines indicate the time courses of 5-HTP production and cell density, respectively. Green- and red-colored lines indicate the profiles at 30 and 37°C, respectively. All data are reported as mean ± s.d. from three independent experiments. Error bars are defined as s.d.

6.4.4 *De novo* microbial synthesis of 5-HTP via metabolic engineering

After achieving the efficient bioconversion of tryptophan to 5-HTP, we proceeded with the construction of a 5-HTP producing strain which allows the utilization of endogenous tryptophan generated from simple carbon sources (Figure 6.5A). Our first attempt was focused on the construction of a tryptophan overproducer. In *E. coli*, tryptophan biosynthesis is branched from the shikimate pathway at chorismate by the action of the *trp* regulon (Figure 6.5A) and negatively regulated by tryptophan transcriptional repressor (TrpR) in response to intracellular tryptophan levels. To circumvent the intrinsic regulation at the transcription level, the complete *trp* operon including *trpEDCBA* was cloned into a low-copy plasmid under the control of an IPTG inducible promoter. Meanwhile, to eliminate the feed-back inhibition effect, a mutation S40F was incorporated into TrpE according to a previous study [125], resulting in plasmid pSA-TrpEDCBA. When the plasmid was introduced into *E. coli* BW*ΔtnaA,* the resulting strain produced 292.2 mg/L of tryptophan at 37 ^{o}C after 24-hour cultivation; however, the titers dramatically decreased after 48 h (74.4 mg/L) probably due to oxidative degradation [126]. This problem was solved when the growth temperature was changed to 30 ^{o}C (Figure 6.5B). In addition to BW*ΔtnaA*, we also attempted to use QH4*ΔtnaA* as the host for boosting carbon flux through the shikimate pathway, because QH4 is a derivative of the well-developed phenylalanine overproducer ATCC31884 with *pheLA* and *tyrA* disrupted and has been successfully engineered for the enhanced production of caffeic acid, salicylic acid and muconic acid in our previous studies [115,116]. However, in this study, QH4*ΔtnaA* harboring pSA-TrpEDCBA did not significantly improve the production of tryptophan but showed slightly improved titers at 30^{o}C compared with the BW*ΔtnaA* host. By the end of 48 h, up to 304.4 mg/L of tryptophan was produced (Figure 6.5B). The control strain QH4*ΔtnaA* without the over-expression of the *trp* operon did not accumulate tryptophan at either temperature.

As the tryptophan production and the bioconversion of tryptophan to 5-HTP were achieved and 30^{o}C worked better for both cases, our further efforts were directed to the establishment of *de novo* biosynthesis of 5-HTP at this temperature by integrating the two modules. When pZE-XcABMW179F was co-transferred together with pSA-TrpEDCBA into *E. coli* BW*ΔtnaA* and QH4*ΔtnaA*, the generated

strains only produced 19.9 and 11.5 mg/L of 5-HTP, respectively, without accumulating tryptophan in the cultures. Apparently, the introduction of the 5-hydroxylation reaction using a high-copy number plasmid exerted negative influence on the carbon flow through tryptophan compared with their parent strains. We speculated that the excessive expression of the XcP4H mutant with the MH4 recycling system might have resulted in metabolic imbalance and disturbed carbon flux towards tryptophan. To test this hypothesis, we cloned the coding sequences of XcP4H mutant W179F, PCD and DNMR into a medium-copy-number plasmid instead of the high-copy-number one, yielding plasmid pCS-XcABMW179F. Interestingly, we observed dramatic improvement on 5-HTP production for both BW*ΔtnaA* and QH4*ΔtnaA* harboring pCS-XcABMW179F; by the end of 48 hours, the two strains produced 128.6 and 152.9 mg/L of 5-HTP, respectively (Figure 6.5C), at the expense of 8.5 and 9.7 g/L of glucose consumption, respectively. Meanwhile, we detected the accumulation of tryptophan at the concentrations of 166.3 and 339.7 mg/L for the two strains, indicating that the carbon flux towards tryptophan was fully recovered. Other by-products were also detected as depicted in Table 6.4. The 5-HTP producing strains followed growth-dependent production patterns (Figure 6.5C).

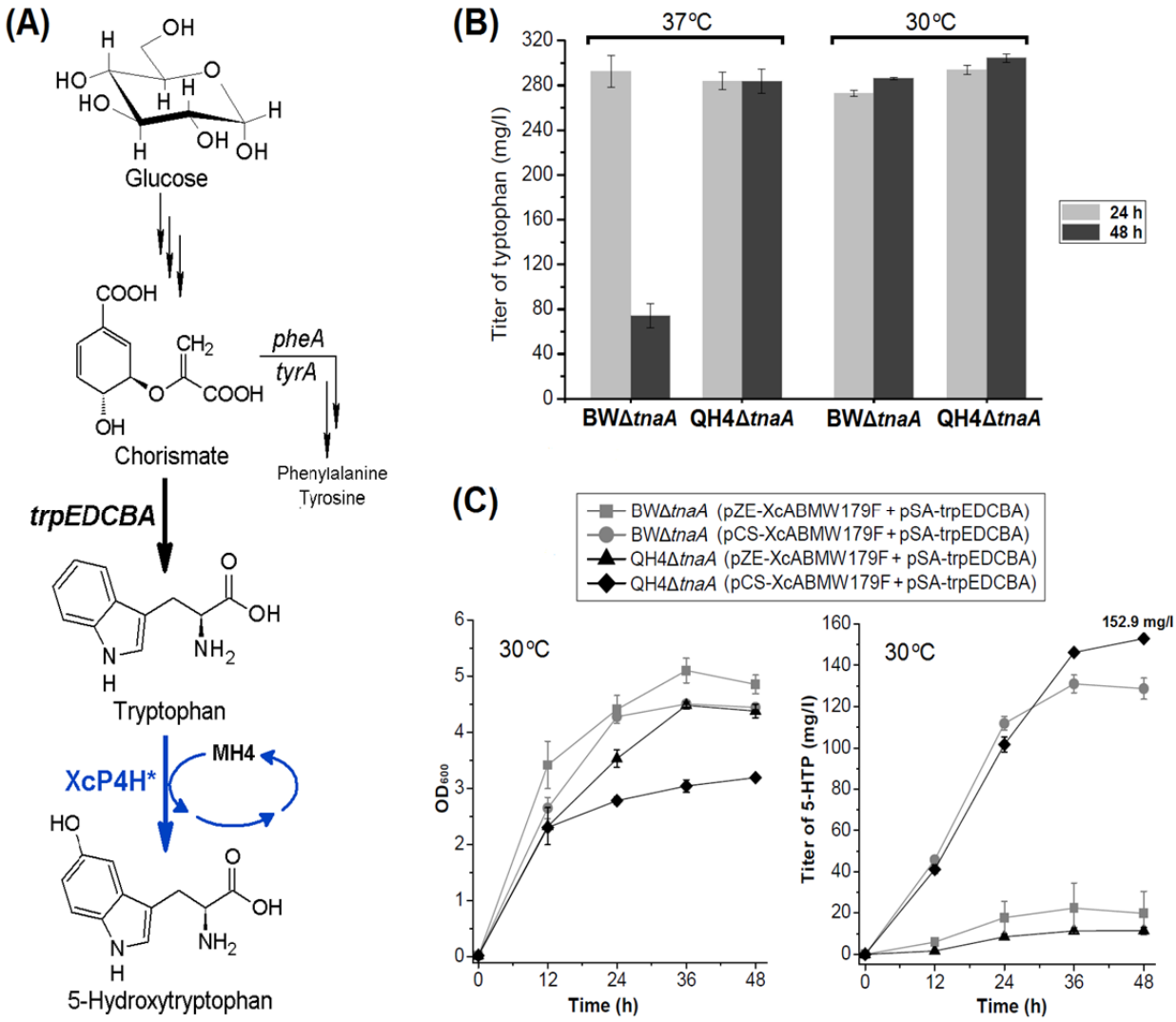


Figure 6.5 *De novo* production of 5-HTP from glucose. (A) Schematic presentation of the complete 5-HTP biosynthetic pathway. The black-and blue-colored arrows indicate the*E. coli* native pathways and heterologous reactions, respectively. (B) Production of tryptophan from glucose at 30 and 37°C. Data are reported as mean ± s.d. from two independent experiments. (C) Profiles of cell growth and 5-HTP production from glucose of two host strains BW*ΔtnaA* (grey-colored) and QH4*ΔtnaA* (black-colored).All data are reported as mean ± s.d. from three independent experiments. Error bars are defined as s.d.

Table 6.4 Intermediates and by-products produced by 5-HTP producing strains

Host strain[a]	Biomass (OD_{600})	Intermediates and by-products[b]			
		Tyrosine (mg l^{-1})	Tryptophan (mg l^{-1})	Acetate (g l^{-1})	Pyruvate (g l^{-1})
BW*ΔtnaA*	4.44 ± 0.02	52.30 ± 0.36	166.29 ± 1.52	2.72 ± 0.00	1.41 ± 0.06
QH4*ΔtnaA*	3.19 ± 0.05	34.85 ± 3.34	339.68 ± 18.15	1.55 ± 0.22	2.11 ± 0.27

[a]Both host strains containing plasmid pSA-TrpEDCBA and pCS-XcABMW179F.

[b]All data are reported as mean ± s.d. from three independent experiments. Error bars are defined as s.d.

6.5 Discussion

Expense on medication has become a major burden of families over the world. High price indeed deprives thelow-income population of access to the use of some drugs that are hard to obtain and expensive. The causes of this issue areeither the low efficiency in isolating these pharmaceuticals from natural sources orthe high costfor their chemical synthesis. Microbial biosynthesis and biocatalysis provide a facile and eco-friendly way for the production of pharmaceutically valuable compounds. The development of metabolic engineering andsynthetic biology tools enables tailored assembly of heterologous and artificial pathways in desirable host strains for the biosynthesis of target products [127].

Lack of suitable enzymes is one of the most frequently encountered problems in pathway engineering in microbes. Functional expression of eukaryotic enzymes is often problematic due to their low solubility, low stability and/or requirements for post-translational modification. For example, the tryptophan 5-hydroxylation reaction has been understood for a long time and many T5Hs has been identified and characterized from human and animals. However, animal AAAHs were hard to be expressed in *E. coli* in a soluble and stable form [108,128]. In addition, their activities are usually regulated by phosphorylation as well as their products [129]. Although the use of truncated or fusion proteins can help obtain soluble and active enzymes [128], the catalytic efficiency seems still low in the

production of 5HTP in *E. coli* using truncated animal T5Hs [110]. In recent years, protein engineering has become a potent tool for enzyme modification in order to obtain desiredcatalytic properties [130]. In our study, a highly active P4H that utilizes MH4 was successfully engineered to catalyze the tryptophan 5-hydroxylation reaction. As we were writing this manuscript, another study also reported the 5-HTP production in *E. coli* with a mutant P4H from *C. violaceum* [131]. However, its function had to completely rely on the supplementation of exogenous pterin coenzyme 6,7-dimethyl-5,6,7,8-tetrahydropterine hydrochloride [131], which is disadvantageous in terms of economic viability as we discuss below.

Self-supply or regeneration of cofactors (including co-enzymes and co-substrates) is one of the greatest advantages for whole-cell biosynthesis and biocatalysis. In *E. coli* host, many of such molecules can be natively generated along with cell growth, such as FMN/$FMNH_2$, FAD/$FADH_2$, $NAD(P)^+$/NAD(P)H, coenzymeA, acetyl-CoA, malonyl-CoA, MH4, etc. Although it is more convenient and economical to utilize these endogenous cofactors, sometimes heterologous enzymes require the cofactor(s) not produced by the host strain. To solve such a problem, one approach is to supplement exogenous cofactors into the culture medium. But it should be noted that most of the cofactors such as tetrahydropterine are so expensive that the supplementation of them is not economically viable for commercial production. Another approachis to introduce the cofactor biosynthetic and/or regeneration mechanism(s) into the host strain. As in the study using animal T5H to produce 5-HTP, a BH4 biosynthetic pathway starting from GTP was introduced into *E. coli* [110]. Meanwhile, a BH4 regeneration system was necessary to achieve continuous production. Recently, an interesting study reported that the mouse T3H can also utilize *E. coli* MH4 in the presence of a BH4 regeneration system, although the efficiency was proved to be low [132]. In this work, with minimal modifications of the host strains' metabolism, the prokaryotic P4Hs and the mutants were able to utilize and recycle *E. coli* endogenous MH4 and NAD(P)H (Figure 5. 2).

In conclusion, this work simultaneously solved two problems in the biological production of 5-HTP, which are related to enzyme compatibility and cofactor self-supply. To our knowledge, the titer of

5-HTP (1.1-1.2 g/L) generated from tryptophan in this work is significantly higher than those in previous studies, showing great scale-up potential. Moreover, this work also demonstrates the *de novo* production of 5HTP without needing to supplement precursors and coenzymes. Since the high-level production of tryptophan (up to 48.7 g/L) has already been achieved in *E. coli* [133], introduction and optimization of the 5-hydroxylation reaction into tryptophan overproducers is expected to result in efficient and low-cost production of 5-HTP.

CHAPTER 7

CONCLUSION

In conclusion, we have demonstrated the microbial production of pharmaceutically important molecules, some of which have shown great potential for commercialized production. From these studies, some general rules have been summarized for constructing an efficient artificial pathway.

First, highly efficient and compatible enzymes are the prerequisite for a productive pathway. When constructing a heterologous pathway, the non-native enzymes from other organisms are frequently incompatible with the microbial hosts. For example, a majority of CYP450 enzymes from plants and membrane proteins from animals cannot be functionally expressed in *E. coli* system. One approach to solve such problems is to seek for the isoenzymes from microorganisms, since the possibility for functional expression of a microbial enzyme is much higher than that from higher organisms. In addition to obtain related catalytic information from literature and enzyme database, BLAST search is a commonly method to prospect enzymes with the same or similar functions. Through *in vitro* or *in vivo* enzyme assays that facilitate the comparison of the kinetic parameters, that the most active candidate can be selected for use of pathway construction. However, the sequence-based approach does not work well in some cases since some enzymes from higher organisms have no counterparts in microorganism. Function-based strategy is an alternative that provides more opportunities to obtain desirable enzymes. Random mutagenesis in combination of a high-throughput selection method often led to the generation of efficient variants with preferred properties. Alternatively, the accumulated knowledge in enzyme structures allowed rational modification of catalytic pocket to endow new catalytic properties to the existing enzymes.

Second, engineering of hosts' native pathway is necessary to improve the productivity of a heterologous pathway. On one hand, some regulation mechanisms (e.g. feedback inhibition) and bottleneck steps exist in wild-type microorganisms to prevent the waste of cellular resources; on the other

hand, the native metabolic network consists of competing pathways which direct the carbon flux to undesirable metabolites. Therefore, appropriate deregulation of native metabolism and the over-expression of rate-limiting enzymes are helpful the re-direct carbon flow to the target pathway. In addition, inactivation of competing pathways by conventional gene disruption method or newly develop synthetic RNA approaches may further improve the production and decrease the generation of by-products.

Third, optimization of expression level is also critical to the performance of an artificially pathway. After long-term evolution, the native pathway enzymes are usually expressed at an optimal level to secure the metabolic balance. The introduction of heterologous pathway may frequently interfere with the host cells' native metabolism. Some negative effects can be brought in if the expression level of foreign enzymes is not well controlled. On one hand, unoptimized expression of these enzymes may cause toxicity and metabolic imbalance due to either the enzymes themselves or the toxic intermediates. On the other hand, excessive expression of enzymes can result in the waste of cellular resources and metabolic burden to host cells. Therefore, it is desirable to control the expression of pathway enzymes at a proper and balanced level. In past years, the development of synthetic biology tools allowed the fine-tuning of protein expression level through the adjustment of plasmid copy number, promoter strength, and RBS binding efficiency. Recently, artificial dynamic regulatory mechanisms were also developed to control pathway enzyme expression in response to the accumulation of certain intermediates. In addition, multivariate modular optimization is also a powerful tool to balance the metabolism. Using this approach, a multi-step pathway can be split into several modules that can be optimized individually. Then the relative expression level of each module can be further adjusted in the context of the whole pathway. This strategy greatly reduces the workload and facilitated the production improvement in many cases.

Overall, this is a proof-of concept study demonstrating the feasibility of microbial production of AAA derived molecules. It has shown great potential for scaled production. Some more efforts are necessary to make these technologies more economically viable and ready for commercialization. For example, the development of antibiotic-free and IPTG-free expression system would be helpful to lower

the production cost and potential environmental concerns. Moreover, optimization of fed-batch cultivation conditions is equally important to achieve high titer, yield and productivity.

REFERENCES

1. Mori H, Iwahashi H: Antioxidant Activity of Caffeic Acid through a Novel Mechanism under UVA Irradiation. *J Clin Biochem Nutr* 2009, 45:49-55.

2. Gulcin I: Antioxidant activity of caffeic acid (3,4-dihydroxycinnamic acid). *Toxicology* 2006, 217:213-220.

3. Ikeda K, Tsujimoto K, Uozaki M, Nishide M, Suzuki Y, Koyama AH, Yamasaki H: Inhibition of multiplication of herpes simplex virus by caffeic acid. *Int J Mol Med* 2011, 28:595-598.

4. Rajendra Prasad N, Karthikeyan A, Karthikeyan S, Reddy BV: Inhibitory effect of caffeic acid on cancer cell proliferation by oxidative mechanism in human HT-1080 fibrosarcoma cell line. *Mol Cell Biochem* 2011, 349:11-19.

5. Chao PC, Hsu CC, Yin MC: Anti-inflammatory and anti-coagulatory activities of caffeic acid and ellagic acid in cardiac tissue of diabetic mice. *Nutr Metab (Lond)* 2009, 6:33.

6. Wu J, Omene C, Karkoszka J, Bosland M, Eckard J, Klein CB, Frenkel K: Caffeic acid phenethyl ester (CAPE), derived from a honeybee product propolis, exhibits a diversity of anti-tumor effects in pre-clinical models of human breast cancer. *Cancer Lett* 2011, 308:43-53.

7. Celik S, Erdogan S, Tuzcu M: Caffeic acid phenethyl ester (CAPE) exhibits significant potential as an antidiabetic and liver-protective agent in streptozotocin-induced diabetic rats. *Pharmacol Res* 2009, 60:270-276.

8. F. Bourgaud AH, R. Larbat, S. Doerper, E. Gontier, S. Kellner and U. Matern: Biosynthesis of coumarins in plants: a major pathway still to be unravelled for cytochrome P450 enzymes. *Phytochem. Rev.* 2006, 5:293-308.

9. Kojima M, Takeuchi W: Detection and characterization of p-coumaric acid hydroxylase in mung bean, Vigna mungo, seedlings. *J Biochem* 1989, 105:265-270.

10. Kim YH, Kwon T, Yang HJ, Kim W, Youn H, Lee JY, Youn B: Gene engineering, purification, crystallization and preliminary X-ray diffraction of cytochrome P450 p-coumarate-3-hydroxylase (C3H), the Arabidopsis membrane protein. *Protein Expr Purif* 2011, 79:149-155.

11. Kneusel RE, Matern U, Nicolay K: Formation of trans-caffeoyl-CoA from trans-4-coumaroyl-CoA by Zn2+-dependent enzymes in cultured plant cells and its activation by an elicitor-induced pH shift. *Arch Biochem Biophys* 1989, 269:455-462.

12. Berner M, Krug D, Bihlmaier C, Vente A, Muller R, Bechthold A: Genes and enzymes involved in caffeic acid biosynthesis in the actinomycete Saccharothrix espanaensis. *J Bacteriol* 2006, 188:2666-2673.

13. Wang J, Lu DQ, Zhao H, Ling XQ, Jiang B, Ouyang PK: Application of response surface methodology optimization for the production of caffeic acid from tobacco waste. *African Journal of Biotechnology* 2009, 8:1416-1424.

14. Yoshimoto M, Kurata-Azuma R, Fujii M, Hou DX, Ikeda K, Yoshidome T, Osako M: Enzymatic production of caffeic acid by koji from plant resources containing caffeoylquinic acid derivatives. *Biosci Biotechnol Biochem* 2005, 69:1777-1781.

15. Sachan A, Ghosh S, Sen SK, Mitra A: Co-production of caffeic acid and p-hydroxybenzoic acid from p-coumaric acid by Streptomyces caeruleus MTCC 6638. *Appl Microbiol Biotechnol* 2006, 71:720-727.

16. Yan Y, Chemler J, Huang L, Martens S, Koffas MA: Metabolic engineering of anthocyanin biosynthesis in Escherichia coli. *Appl Environ Microbiol* 2005, 71:3617-3623.

17. Yan Y, Kohli A, Koffas MA: Biosynthesis of natural flavanones in Saccharomyces cerevisiae. *Appl Environ Microbiol* 2005, 71:5610-5613.

18. Yan Y, Huang L, Koffas MA: Biosynthesis of 5-deoxyflavanones in microorganisms. *Biotechnol J* 2007, 2:1250-1262.

19. Yan Y, Li Z, Koffas MA: High-yield anthocyanin biosynthesis in engineered Escherichia coli. *Biotechnol Bioeng* 2008, 100:126-140.

20. Choi O, Wu CZ, Kang SY, Ahn JS, Uhm TB, Hong YS: Biosynthesis of plant-specific phenylpropanoids by construction of an artificial biosynthetic pathway in Escherichia coli. *J Ind Microbiol Biotechnol* 2011.

21. Santos CN, Koffas M, Stephanopoulos G: Optimization of a heterologous pathway for the production of flavonoids from glucose. *Metab Eng* 2011, 13:392-400.

22. Nakagawa A, Minami H, Kim JS, Koyanagi T, Katayama T, Sato F, Kumagai H: A bacterial platform for fermentative production of plant alkaloids. *Nat Commun* 2011, 2:326.

23. Munoz AJ, Hernandez-Chavez G, de Anda R, Martinez A, Bolivar F, Gosset G: Metabolic engineering of Escherichia coli for improving L: -3,4-dihydroxyphenylalanine (L: -DOPA) synthesis from glucose. *J Ind Microbiol Biotechnol* 2011.

24. Shen CR, Liao JC: Metabolic engineering of Escherichia coli for 1-butanol and 1-propanol production via the keto-acid pathways. *Metab Eng* 2008, 10:312-320.

25. Lutz R, Bujard H: Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* 1997, 25:1203-1210.

26. Sambrook J FE, Maniatis T: Molecular cloning: a laboratory manual. 2nd Ed. *NY: Cold Spring Harbor Laboratory* 1989.

27. Datsenko KA, Wanner BL: One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* 2000, 97:6640-6645.

28. Xue Z, McCluskey M, Cantera K, Sariaslani FS, Huang L: Identification, characterization and functional expression of a tyrosine ammonia-lyase and its mutants from the photosynthetic bacterium Rhodobacter sphaeroides. *J Ind Microbiol Biotechnol* 2007, 34:599-604.

29. Heckman KL, Pease LR: Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat Protoc* 2007, 2:924-932.

30. Lutke-Eversloh T, Stephanopoulos G: Feedback inhibition of chorismate mutase/prephenate dehydrogenase (TyrA) of Escherichia coli: generation and characterization of tyrosine-insensitive mutants. *Appl Environ Microbiol* 2005, 71:7224-7228.

31. Zrelli H, Matsuoka M, Kitazaki S, Zarrouk M, Miyazaki H: Hydroxytyrosol reduces intracellular reactive oxygen species levels in vascular endothelial cells by upregulating catalase expression through the AMPK-FOXO3a pathway. *Eur J Pharmacol* 660:275-282.

32. Louie TM, Xie XS, Xun L: Coordinated production and utilization of FADH2 by NAD(P)H-flavin oxidoreductase and 4-hydroxyphenylacetate 3-monooxygenase. *Biochemistry* 2003, 42:7509-7517.

33. Leonard E, Runguphan W, O'Connor S, Prather KJ: Opportunities in metabolic engineering to facilitate scalable alkaloid production. *Nat Chem Biol* 2009, 5:292-300.

34. Prieto MA, Perez-Aranda A, Garcia JL: Characterization of an Escherichia coli aromatic hydroxylase with a broad substrate range. *J Bacteriol* 1993, 175:2162-2167.

35. Liebgott PP, Amouric A, Comte A, Tholozan JL, Lorquin J: Hydroxytyrosol from tyrosol using hydroxyphenylacetic acid-induced bacterial cultures and evidence of the role of 4-HPA 3-hydroxylase. *Res Microbiol* 2009, 160:757-766.

36. Noel JP, Louie GV, Bowman ME, Moore BS, Moffitt MC: Substrate switched ammonia lyases and mutases *United States Patent Application* 2009, No. 2009/0011400 A1

37. Chavez-Bejar MI, Lara AR, Lopez H, Hernandez-Chavez G, Martinez A, Ramirez OT, Bolivar F, Gosset G: Metabolic engineering of Escherichia coli for L-tyrosine production by expression of genes coding for the chorismate mutase domain of the native chorismate mutase-prephenate dehydratase and a cyclohexadienyl dehydrogenase from Zymomonas mobilis. *Appl Environ Microbiol* 2008, 74:3284-3290.

38. Lutke-Eversloh T, Stephanopoulos G: L-tyrosine production by deregulated strains of Escherichia coli. *Appl Microbiol Biotechnol* 2007, 75:103-110.

39. : !!! INVALID CITATION !!!

40. Bourgaud F, Hehn A, Larbat R, Doerper S, Gontier E, Kellner S, Matern U: Biosynthesis of coumarins in plants: a major pathway still to be unravelled for cytochrome P450 enzymes. *Phytochem Rev* 2006, 5:293-308.

41. Lacy A, O'Kennedy R: Studies on coumarins and coumarin-related compounds to determine their therapeutic role in the treatment of cancer. *Curr Pharm Des* 2004, 10:3797-3811.

42. Kostova I, Raleva S, Genova P, Argirova R: Structure-Activity Relationships of Synthetic Coumarins as HIV-1 Inhibitors. *Bioinorg Chem Appl* 2006:68274.

43. Stoker JR, Bellis DM: The biosynthesis of coumarin in Melilotus alba. *J Biol Chem* 1962, 237:2303-2305.

44. Poulton JE, McRee DE, Conn EE: Intracellular Localization of Two Enzymes Involved in Coumarin Biosynthesis in Melilotus alba. *Plant Physiol* 1980, 65:171-175.

45. Gestetner B, Conn EE: The 2-hydroxylation of trans-cinnamic acid by chloroplasts from Melilotus alba Desr. *Arch Biochem Biophys* 1974, 163:617-624.

46. Kleinhofs A, Haskins FA, Gorz HJ: Trans-o-hydroxycinnamic acid glucosylation in cell-free extracts of Melilotus alba. *Phytochemistry* 1967, 6:1313-1318.

47. Haskins FA, Williams LG, Gorz HJ: Light-Induced Trans to Cis Conversion of beta-d-Glucosyl o-Hydroxycinnamic Acid in Melilotus alba Leaves. *Plant Physiol* 1964, 39:777-781.

48. Oba K, Conn EE, Canut H, Boudet AM: Subcellular Localization of 2-(beta-d-Glucosyloxy)-Cinnamic Acids and the Related beta-glucosidase in Leaves of Melilotus alba Desr. *Plant Physiol* 1981, 68:1359-1363.

49. Lin Y, Yan Y: Biosynthesis of caffeic acid in Escherichia coli using its endogenous hydroxylase complex. *Microb Cell Fact* 2012, 11:42.

50. Ehlting J, Buttner D, Wang Q, Douglas CJ, Somssich IE, Kombrink E: Three 4-coumarate:coenzyme A ligases in Arabidopsis thaliana represent two evolutionarily divergent classes in angiosperms. *Plant J* 1999, 19:9-20.

51. Matsumoto S, Mizutani M, Sakata K, Shimizu B: Molecular cloning and functional analysis of the ortho-hydroxylases of p-coumaroyl coenzyme A/feruloyl coenzyme A involved in formation of umbelliferone and scopoletin in sweet potato, Ipomoea batatas (L.) Lam. *Phytochemistry* 2012, 74:49-57.

52. Murray RDH, Méndez J, Brown SA: The natural coumarins: occurrence, chemistry and biochemistry. *Wiley, Chichester* 1982.

53. Melnikova I: The anticoagulants market. *Nature Reviews. Drug Discovery* 2009, 8:353-354.

54. Gao WT, Hou WD, Zheng MR, Tang LJ: Clean and Convenient One-Pot Synthesis of 4-Hydroxycoumarin and 4-Hydroxy-2-Quinolinone Derivatives. *Synthetic Communications* 2010, 40:732-738.

55. Martin VJ, Pitera DJ, Withers ST, Newman JD, Keasling JD: Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nature Biotechnology* 2003, 21:796-802.

56. Lequesne PW: The Natural Coumarins - Occurrence, Chemistry and Biochemistry - Murray,Rdh, Mendez,J, Brown,Sa. *Journal of the American Chemical Society* 1983, 105:6536-6536.

57. Bye A, King HK: The biosynthesis of 4-hydroxycoumarin and dicoumarol by Aspergillus fumigatus Fresenius. *Biochemical Journal* 1970, 117:237-245.

58. Liu B, Raeth T, Beuerle T, Beerhues L: A novel 4-hydroxycoumarin biosynthetic pathway. *Plant Molecular Biology* 2010, 72:17-25.

59. Chen Z, Zheng Z, Huang J, Lai Z, Fan B: Biosynthesis of salicylic acid in plants. *Plant Signal Behav* 2009, 4:493-496.

60. Heeb S, Fletcher MP, Chhabra SR, Diggle SP, Williams P, Camara M: Quinolones: from antibiotics to autoinducers. *FEMS Microbiology Reviews* 2011, 35:247-274.

61. Kikuchi Y, Tsujimoto K, Kurahashi O: Mutational analysis of the feedback sites of phenylalanine-sensitive 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase of Escherichia coli. *Applied and Environmental Microbiology* 1997, 63:761-762.

62. Rogozinska M, Adamkiewicz A, Mlynarski J: Efficient "on water" organocatalytic protocol for the synthesis of optically pure warfarin anticoagulant. *Green Chemistry* 2011, 13:1155-1157.

63. Dhamankar H, Prather KL: Microbial chemical factories: recent advances in pathway engineering for synthesis of value added chemicals. *Current Opinion in Structural Biology* 2011, 21:488-494.

64. Zhang F, Carothers JM, Keasling JD: Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nature Biotechnology* 2012, 30:354-359.

65. Huo YX, Cho KM, Rivera JGL, Monte E, Shen CR, Yan YJ, Liao JC: Conversion of proteins into biofuels by engineering nitrogen flux. *Nature Biotechnology* 2011, 29:346-U160.

66. Atsumi S, Cann AF, Connor MR, Shen CR, Smith KM, Brynildsen MP, Chou KJ, Hanai T, Liao JC: Metabolic engineering of Escherichia coli for 1-butanol production. *Metab Eng* 2008, 10:305-311.

67. Zhang K, Sawaya MR, Eisenberg DS, Liao JC: Expanding metabolism for biosynthesis of nonnatural alcohols. *Proc Natl Acad Sci U S A* 2008, 105:20653-20658.

68. Lin Y, Sun X, Yuan Q, Yan Y: Combinatorial biosynthesis of plant-specific coumarins in bacteria. *Metab Eng* 2013, 18:69-77.

69. Huang Q, Lin Y, Yan Y: Caffeic acid production enhancement by engineering a phenylalanine over-producing Escherichia coli. *Biotechnology and Bioengineering* 2013.

70. Anthony JR, Anthony LC, Nowroozi F, Kwon G, Newman JD, Keasling JD: Optimization of the mevalonate-based isoprenoid biosynthetic pathway in Escherichia coli for production of the anti-malarial drug precursor amorpha-4,11-diene. *Metabolic Engineering* 2009, 11:13-19.

71. Ajikumar PK, Xiao WH, Tyo KEJ, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G: Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in Escherichia coli. *Science* 2010, 330:70-74.

72. Shen X, Lin Y, Jain R, Yuan Q, Yan Y: Inhibition of acetate accumulation leads to enhanced production of (R,R)-2,3-butanediol from glycerol in Escherichia coli. *J Ind Microbiol Biotechnol* 2012, 39:1725-1729.

73. Sun X, Lin Y, Huang Q, Yuan Q, Yan Y: A novel muconic acid biosynthesis approach by shunting tryptophan biosynthesis via anthranilate. *Appl Environ Microbiol* 2013, 79:4024-4030.

74. Curran KA, Leavitt JM, Karim AS, Alper HS: Metabolic engineering of muconic acid production in Saccharomyces cerevisiae. *Metab Eng* 2013, 15:55-66.

75. Research TM: Salicylic Acid Market for Pharmaceutical, Skin care, Hair care and Other Applications - Global Industry Analysis, Size, Share, Growth, Trends, and Forecast 2013-2019. Edited by; 2013.

76. Fuchs G, Boll M, Heider J: Microbial degradation of aromatic compounds - from one strategy to four. *Nat Rev Microbiol* 2011, 9:803-816.

77. Gaille C, Kast P, Haas D: Salicylate biosynthesis in Pseudomonas aeruginosa. Purification and characterization of PchB, a novel bifunctional enzyme displaying isochorismate pyruvate-lyase and chorismate mutase activities. *Journal of Biological Chemistry* 2002, 277:21768-21775.

78. Draths KM, Frost JW: Environmentally Compatible Synthesis of Adipic Acid from D-Glucose. *Journal of the American Chemical Society* 1994, 116:399-400.

79. Niu W, Draths KM, Frost JW: Benzene-free synthesis of adipic acid. *Biotechnol Prog* 2002, 18:201-211.

80. Weber C, Bruckner C, Weinreb S, Lehr C, Essl C, Boles E: Biosynthesis of cis,cis-muconic acid and its aromatic precursors, catechol and protocatechuic acid, from renewable feedstocks by Saccharomyces cerevisiae. *Appl Environ Microbiol* 2012, 78:8421-8430.

81. Lin Y, Shen X, Yuan Q, Yan Y: Microbial biosynthesis of the anticoagulant precursor 4-hydroxycoumarin. *Nat Commun* 2013, 4:2603.

82. Gaille C, Reimmann C, Haas D: Isochorismate synthase (PchA), the first and rate-limiting enzyme in salicylate biosynthesis of Pseudomonas aeruginosa. *Journal of Biological Chemistry* 2003, 278:16893-16898.

83. Seo JS, Keum YS, Li QX: Bacterial degradation of aromatic compounds. *Int J Environ Res Public Health* 2009, 6:278-309.

84. Patnaik R, Zolandz RR, Green DA, Kraynie DF: L-tyrosine production by recombinant Escherichia coli: fermentation optimization and recovery. *Biotechnol Bioeng* 2008, 99:741-752.

85. Zhang K, Woodruff AP, Xiong M, Zhou J, Dhande YK: A synthetic metabolic pathway for production of the platform chemical isobutyric acid. *ChemSusChem* 2011, 4:1068-1070.

86. Xu P, Bhan N, Koffas MAG: Engineering plant metabolism into microbes: from systems biology to synthetic biology. *Current Opinion in Biotechnology* 2013, 24:291-299.

87. Rabinovitch-Deere CA, Oliver JW, Rodriguez GM, Atsumi S: Synthetic biology and metabolic engineering approaches to produce biofuels. *Chem Rev* 2013, 113:4611-4632.

88. Oliver JW, Machado IM, Yoneda H, Atsumi S: Cyanobacterial conversion of carbon dioxide to 2,3-butanediol. *Proc Natl Acad Sci U S A* 2013, 110:1249-1254.

89. Hammer K, Mijakovic I, Jensen PR: Synthetic promoter libraries--tuning of gene expression. *Trends Biotechnol* 2006, 24:53-55.

90. Smolke CD, Martin VJ, Keasling JD: Controlling the metabolic flux through the carotenoid pathway using directed mRNA processing and stabilization. *Metab Eng* 2001, 3:313-321.

91. Salis HM, Mirsky EA, Voigt CA: Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* 2009, 27:946-950.

92. Dahl RH, Zhang F, Alonso-Gutierrez J, Baidoo E, Batth TS, Redding-Johanson AM, Petzold CJ, Mukhopadhyay A, Lee TS, Adams PD, et al.: Engineering dynamic pathway regulation using stress-response promoters. *Nat Biotechnol* 2013, 31:1039-1046.

93. Zhang F, Carothers JM, Keasling JD: Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat Biotechnol* 2012, 30:354-359.

94. Juminaga D, Baidoo EE, Redding-Johanson AM, Batth TS, Burd H, Mukhopadhyay A, Petzold CJ, Keasling JD: Modular engineering of L-tyrosine production in Escherichia coli. *Appl Environ Microbiol* 2012, 78:89-98.

95. Ajikumar PK, Xiao WH, Tyo KE, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G: Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. *Science* 2010, 330:70-74.

96. Xu P, Gu Q, Wang W, Wong L, Bower AG, Collins CH, Koffas MA: Modular optimization of multi-gene pathways for fatty acids production in E. coli. *Nat Commun* 2013, 4:1409.

97. Fisher MA, Boyarskiy S, Yamada MR, Kong N, Bauer S, Tullman-Ercek D: Enhancing Tolerance to Short-Chain Alcohols by Engineering the Escherichia coli AcrB Efflux Pump to Secrete the Non-native Substrate n-Butanol. *ACS Synth Biol* 2013.

98. WHO: World Health Organization Fact Sheet on Depression. Edited by. http://www.who.int/mediacentre/factsheets/fs369/en/index.html; 2012.

99. Byerley WF, Judd LL, Reimherr FW, Grosser BI: 5-Hydroxytryptophan: a review of its antidepressant efficacy and adverse effects. *J Clin Psychopharmacol* 1987, 7:127-137.

100. Turner EH, Loftis JM, Blackwell AD: Serotonin a la carte: supplementation with the serotonin precursor 5-hydroxytryptophan. *Pharmacol Ther* 2006, 109:325-338.

101. Birdsall TC: 5-Hydroxytryptophan: a clinically-effective serotonin precursor. *Altern Med Rev* 1998, 3:271-280.

102. Leonard E, Ajikumar PK, Thayer K, Xiao WH, Mo JD, Tidor B, Stephanopoulos G, Prather KLJ: Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107:13654-13659.

103. Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J, et al.: Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 2006, 440:940-943.

104. Zhang K, Li H, Cho KM, Liao JC: Expanding metabolism for total biosynthesis of the nonnatural amino acid L-homoalanine. *Proc Natl Acad Sci U S A* 2010, 107:6234-6239.

105. Lim CG, Fowler ZL, Hueller T, Schaffer S, Koffas MAG: High-Yield Resveratrol Production in Engineered Escherichia coli. *Applied and Environmental Microbiology* 2011, 77:3451-3460.

106. Teigen K, McKinney JA, Haavik J, Martinez A: Selectivity and affinity determinants for ligand binding to the aromatic amino acid hydroxylases. *Curr Med Chem* 2007, 14:455-467.

107. Fitzpatrick PF: Mechanism of aromatic amino acid hydroxylation. *Biochemistry* 2003, 42:14083-14091.

108. McKinney J, Knappskog PM, Pereira J, Ekern T, Toska K, Kuitert BB, Levine D, Gronenborn AM, Martinez A, Haavik J: Expression and purification of human tryptophan hydroxylase from Escherichia coli and Pichia pastoris. *Protein Expr Purif* 2004, 33:185-194.

109. Martinez A, Knappskog PM, Haavik J: A structural approach into human tryptophan hydroxylase and its implications for the regulation of serotonin biosynthesis. *Current Medicinal Chemistry* 2001, 8:1077-1091.

110. KNIGHT EM, Zhu J, FÖRSTER J, Luo H: Microorganisms for the production of 5-hydroxytryptophan. US Patent 2013.

111. Nakata H, Yamauchi T, Fujisawa H: Phenylalanine hydroxylase from Chromobacterium violaceum. Purification and characterization. *J Biol Chem* 1979, 254:1829-1833.

112. Zhao G, Xia T, Song J, Jensen RA: Pseudomonas aeruginosa possesses homologues of mammalian phenylalanine hydroxylase and 4 alpha-carbinolamine dehydratase/DCoH as part of a three-component gene cluster. *Proc Natl Acad Sci U S A* 1994, 91:1366-1370.

113. Kino K, Hara R, Nozawa A: Enhancement of L-tryptophan 5-hydroxylation activity by structure-based modification of L-phenylalanine 4-hydroxylase from Chromobacterium violaceum. *J Biosci Bioeng* 2009, 108:184-189.

114. Pribat A, Blaby IK, Lara-Nunez A, Gregory JF, 3rd, de Crecy-Lagard V, Hanson AD: FolX and FolM are essential for tetrahydromonapterin synthesis in Escherichia coli and Pseudomonas aeruginosa. *J Bacteriol* 2010, 192:475-482.

115. Huang Q, Lin Y, Yan Y: Caffeic acid production enhancement by engineering a phenylalanine over-producing Escherichia coli strain. *Biotechnol Bioeng* 2013, 110:3188-3196.

116. Lin Y, Sun X, Yuan Q, Yan Y: Extending shikimate pathway for the production of muconic acid and its precursor salicylic acid in Escherichia coli. *Metab Eng* 2014, 23C:62-69.

117. Datsenko KA, Wanner BL: One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America* 2000, 97:6640-6645.

118. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011, 28:2731-2739.

119. Lin Y, Yan Y: Biosynthesis of caffeic acid in Escherichia coli using its endogenous hydroxylase complex. *Microb Cell Fact* 2012, 11:42.

120. Erlandsen H, Kim JY, Patch MG, Han A, Volner A, Abu-Omar MM, Stevens RC: Structural comparison of bacterial and human iron-dependent phenylalanine hydroxylases: similar fold, different stability and reaction rates. *J Mol Biol* 2002, 320:645-661.

121. Cao J, Shi F, Liu X, Huang G, Zhou M: Phylogenetic analysis and evolution of aromatic amino acid hydroxylase. *FEBS Lett* 2010, 584:4775-4782.

122. Ekstrom F, Stier G, Eaton JT, Sauer UH: Crystallization and X-ray analysis of a bacterial non-haem iron-containing phenylalanine hydroxylase from the Gram-negative opportunistic pathogen Pseudomonas aeruginosa. *Acta Crystallogr D Biol Crystallogr* 2003, 59:1310-1312.

123. Gong F, Ito K, Nakamura Y, Yanofsky C: The mechanism of tryptophan induction of tryptophanase operon expression: tryptophan inhibits release factor-mediated cleavage of TnaC-peptidyl-tRNA(Pro). *Proc Natl Acad Sci U S A* 2001, 98:8997-9001.

124. Song J, Xia T, Jensen RA: PhhB, a Pseudomonas aeruginosa homolog of mammalian pterin 4a-carbinolamine dehydratase/DCoH, does not regulate expression of phenylalanine hydroxylase at the transcriptional level. *J Bacteriol* 1999, 181:2789-2796.

125. Zhao ZJ, Zou C, Zhu YX, Dai J, Chen S, Wu D, Wu J, Chen J: Development of L-tryptophan production strains by defined genetic modification in Escherichia coli. *J Ind Microbiol Biotechnol* 2011, 38:1921-1929.

126. Simat TJ, Steinhart H: Oxidation of Free Tryptophan and Tryptophan Residues in Peptides and Proteins. *J Agric Food Chem* 1998, 46:490-498.

127. Mora-Pale M, Sanchez-Rodriguez SP, Linhardt RJ, Dordick JS, Koffas MA: Biochemical strategies for enhancing the in vivo production of natural products with pharmaceutical potential. *Curr Opin Biotechnol* 2014, 25C:86-94.

128. Higgins CA, Vermeer LM, Doorn JA, Roman DL: Expression and purification of recombinant human tyrosine hydroxylase as a fusion protein in Escherichia coli. *Protein Expression and Purification* 2012, 84:219-223.

129. Daubner SC, Lauriano C, Haycock JW, Fitzpatrick PF: Site-directed mutagenesis of serine 40 of rat tyrosine hydroxylase. Effects of dopamine and cAMP-dependent phosphorylation on enzyme activity. *J Biol Chem* 1992, 267:12639-12646.

130. Marcheschi RJ, Gronenberg LS, Liao JC: Protein engineering for metabolic engineering: current and next-generation tools. *Biotechnol J* 2013, 8:545-555.

131. Hara R, Kino K: Enhanced synthesis of 5-hydroxy-l-tryptophan through tetrahydropterin regeneration. *AMB Express* 2013, 3:70.

132. Satoh Y, Tajima K, Munekata M, Keasling JD, Lee TS: Engineering of L-tyrosine oxidation in Escherichia coli and microbial production of hydroxytyrosol. *Metab Eng* 2012, 14:603-610.

133. Wang J, Cheng LK, Wang J, Liu Q, Shen T, Chen N: Genetic engineering of Escherichia coli to enhance production of L-tryptophan. *Appl Microbiol Biotechnol* 2013, 97:7587-7596.