Trends in Biotechnology

CellPress
REVIEWS

Special Issue: Metabolic Engineering
Review

Metabolic Engineering of *Escherichia coli* for Natural Product Biosynthesis

Dongsoo Yang,[1,4] Seon Young Park,[1,4] Yae Seul Park,[1] Hyunmin Eun,[1] and Sang Yup Lee[1,2,3,*]

Natural products are widely employed in our daily lives as food additives, pharmaceuticals, nutraceuticals, and cosmetic ingredients, among others. However, their supply has often been limited because of low-yield extraction from natural resources such as plants. To overcome this problem, metabolically engineered *Escherichia coli* has emerged as a cell factory for natural product biosynthesis because of many advantages including the availability of well-established tools and strategies for metabolic engineering and high cell density culture, in addition to its high growth rate. We review state-of-the-art metabolic engineering strategies for enhanced production of natural products in *E. coli*, together with representative examples. Future challenges and prospects of natural product biosynthesis by engineered *E. coli* are also discussed.

Highlights

E. coli has emerged as a prominent host for natural product biosynthesis.

Improved enzymes with higher activity, altered substrate specificity, and product selectivity can be obtained by structure-based or computer simulation-based protein engineering.

Balancing the expression levels of genes or pathway modules is effective in increasing the metabolic flux towards target compounds.

System-wide analysis of metabolic networks, omics analysis, adaptive laboratory evolution, and biosensor-based screening can further increase the production of target compounds.

Systems metabolic engineering allows the development of engineered *E. coli* strains that are capable of more efficiently producing diverse natural products.

E. coli as a Cell Factory for Natural Product Biosynthesis

Natural products have been widely used in food and medicine in human history. Many of these natural products have been developed as pharmaceuticals or employed as structural backbones for the development of new drugs [1], and also as food and cosmetic ingredients. Because natural products are increasingly preferred over synthetic compounds, their global demand and market size have also been increasing [2]. Other than natural products of microbial origin, most natural products are obtained by extraction from plants or animals. However, the low yields obtained by extraction from natural resources often lead to insufficient supplies and high costs [3]. Furthermore, total chemical synthesis of many of these natural products is often infeasible because of the coproduction of undesirable stereoisomers and intermediate metabolites in addition to the necessity of costly multistep reactions [4]. There has thus been much interest in producing natural products by employing metabolically engineered microorganisms [5].

Among various microbial host strains, *E. coli* has been employed as a workhorse for the production of several natural products. *E. coli*, being the best-studied organism, has several advantages as a microbial cell factory, including high growth rate, the availability of gene and genome engineering tools, established high cell density culture techniques, and various **systems metabolic engineering** (see Glossary) tools and strategies including the best-curated **genome-scale metabolic models (GEMs)** [6,7]. In addition, various high cell density culture techniques have been well established for *E. coli*, which are well discussed in a classical review paper [8]. Thus, *E. coli* has been widely employed in academia and industry as one of the most popular microbial hosts for the production of natural products [5]. However, there are many remaining hurdles, including the lack of subcellular organelles required for functional expression of eukaryotic enzymes (e.g., **cytochrome P450s**) and weak endogenous metabolic flux towards desired target natural products. In addition, possible phage infection should be avoided.

Natural products can be broadly classified into four major categories according to their structural features: terpenoids, phenylpropanoids, polyketides, and alkaloids [5]; there are also other natural products that do not fall into one of these four categories. The overall biosynthetic

[1]Metabolic and Biomolecular Engineering National Research Laboratory, Systems Metabolic Engineering and Systems Healthcare Cross Generation Collaborative Laboratory, Department of Chemical and Biomolecular Engineering (BK21 Plus Program), Institute for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea
[2]BioProcess Engineering Research Center, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea
[3]BioInformatics Research Center, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea
[4]These authors contributed equally.

*Correspondence
leesy@kaist.ac.kr (S.Y. Lee).

pathways of representative natural products belonging to the four major categories and others are shown in Figure 1 (Key Figure). In this paper, recent metabolic engineering strategies for the production of natural products by *E. coli* are reviewed together with accompanying examples. General metabolic engineering strategies for the production of different categories of natural products in *E. coli* are discussed in Box 1. Recent examples of natural products biosynthesized by metabolically engineered *E. coli* are listed in Table 1.

Enzyme Engineering Strategies to Enhance Bioconversion Efficiency

Structure-Based Enzyme Engineering

One of the most important strategies for maximizing pathway flux towards desired target natural products is enzyme engineering (Figure 2). Structure-based enzyme engineering can be employed to enhance the activity of a bottleneck enzyme and alter its substrate specificity and product selectivity. A representative example is engineering of a **promiscuous enzyme**, NphB, which can prenylate various aromatic substrates. To enhance the catalytic activity and the product selectivity of NphB, protein-docking simulation was performed to identify the key amino acids that bind to orsellinic acid in its catalytic pocket (Figure 2A) [9]. By applying the engineered NphB mutant *in vivo*, the titer of total geranylated products was increased from 80 mg/l to 300 mg/l. In another study, computational docking simulation was performed on taxadiene synthase (TXS) to generate a mutant enzyme which can convert geranylgeranyl pyrophosphate (GGPP) into taxa-4(20)-11(12)-diene, a more favorable substrate for the downstream enzyme, taxadiene oxidase [10]. Coexpression of these two enzymes could enhance taxadien-5α-ol production by 2.4-fold. *E. coli* farnesyl diphosphate synthase (IspA) is another promiscuous enzyme that can react with both dimethylallyl diphosphate (DMAPP) and geranyl diphosphate (GPP). Because IspA converts DMAPP to GPP, and also GPP to farnesyl pyrophosphate (FPP), increasing the GPP pool for enhanced monoterpene production has been challenging. Thus, IspA mutants with reduced affinity for GPP were constructed to increase the production of one of the monoterpenes, 1,8-cineole [11]. Another interesting example of enzyme engineering was performed on 3,4-dihydroxyphenylacetaldehyde synthase (DHPAAS) for the production of opioids [12], as described in Box 2. As showcased above, enzyme engineering is one of the most important strategies to increase the metabolic flux towards target natural products by improving the activities of rate-controlling enzymes or by altering their substrate specificity or product selectivity.

Random Screening of Mutant Enzymes

Although the protein structures of increasing numbers of enzymes are becoming available, there are still numerous enzymes with unknown structures. In such cases where structure-based engineering is not possible, **directed evolution** can be employed by means of random mutagenesis of target enzymes followed by high-throughput screening. A recent example is random mutagenesis and selection of 4-coumarate:CoA ligase (4CL), a key enzyme in the phenylpropanoid pathway (Figure 2B). To find an enzyme mutant with high activity through screening the random 4CL mutant library, the TtgR regulatory system responding to resveratrol was first developed as a resveratrol biosensor [13]. Enhanced production of resveratrol as well as naringenin was possible by employing a selected 4CL mutant with higher activity. In another study, directed evolution was employed on rate-controlling enzymes in the 1-deoxy-D-xylulose-5-phosphate (DXP) pathway [14]. For this purpose, lycopene was used as an indicator for colorimetric screening of enzyme variants generating higher flux towards the DXP pathway. When the selected enzyme mutants were introduced, the titer of isoprene produced by engineered *E. coli* was increased by 60%. Another example of random mutagenesis and screening of enzymes for nonnatural C_{50}-astaxanthin production in *E. coli* is given in Box 3. As described above, an enzyme mutant library can be generated for high-throughput screening of

Glossary

Adaptive laboratory evolution (ALE): a process for the rapid isolation of strains with desirable phenotypes by multiple rounds of cultivation and selection.
^{13}C-metabolic flux analysis (^{13}C-MFA): a tool for quantifying metabolic fluxes through isotopomer analysis using stable ^{13}C-labeled tracers. The propagation of ^{13}C-labeled compounds over time can be tracked using mass spectrometry to record and analyze labeling patterns in metabolic intermediates or final products.
CRISPR interference (CRISPRi): an RNA-guided target gene knockdown technology which involves the action of a catalytically dead Cas9 (dCas9) protein to inhibit target gene transcription through engineered guide RNAs.
Cytochrome P450: a ubiquitous enzyme belonging to the hemoprotein family which primarily functions as monooxygenase; requires a partner reductase for electron donation.
Directed evolution: a process for the rapid isolation of superior enzymes by multiple rounds of mutation and selection.
Feedback inhibition: a phenomenon when the activity of an enzyme is suppressed by the end product or intermediates of the pathway; an inhibitor interacts with the target enzyme by binding to an allosteric site, which is followed by conformational changes of the enzyme.
Genome-scale metabolic models (GEMs): a set of genome-wide metabolic reaction models containing gene–protein–reaction association information that describes the entire cellular metabolism of a specific organism.
Machine learning: a broad set of algorithms that are used to perform calculations and tasks based on patterns and inferences using training data.
Molecular biosensor: a genetic construct that converts the abundance of a target molecule into an observable signal in a proportional manner.
Polyketide synthase (PKS): a multidomain enzyme or a complex of enzymes that catalyze polyketide biosynthesis; these can be classified into three major types (types I, II, and III) according to their mechanism of carbon chain elongation.
Promiscuous enzyme: an enzyme which can catalyze multiple side

variant enzymes with desired traits such as higher activity, improved substrate specificity, and/or desired product selectivity. High-throughput screening of enzymes requires easily identifiable signals, and colorimetric screening or surrogate reporters yielding fluorescence have been most popularly employed.

Removing Feedback Inhibition

Several intermediate metabolites exert **feedback inhibition** on enzymes in natural product biosynthetic pathways, and this can be a severe problem for the enhanced production of natural products. This problem can be solved by mutating the key enzymes to be resistant to feedback inhibition (Figure 2C). In the biosynthetic pathway of carbapenem (a β-lactam antibiotic), either glutamate or proline can be used as an important precursor. One study employed a glutamate feeding strategy for carbapenem production. The key enzyme, glutamate 5-kinase (ProB), is responsible for the conversion of glutamate to glutamyl 5-phosphate. However, ProB is feedback-inhibited by proline, which is also an important intermediate in the carbapenem biosynthetic pathway. Therefore, to make it resistant to feedback inhibition, its proline-binding site was mutated [15], leading to a significant increase in carbapenem production in *E. coli* [16]. Similarly, feedback inhibition-resistant 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase (AroG) and anthranilate (ANT) synthase (TrpE) were developed for the enhanced production of indirubin and methyl anthranilate (MANT), respectively (Figure 2C and Box 4) [17,18]. In particular, the feedback inhibition-resistant AroG enzyme may also be widely used for the production of other natural products because several of these, including phenylpropanoids and benzylisoquinoline alkaloids (BIAs), are biosynthesized via the shikimate pathway.

Engineering Membrane-Associated Enzymes

Many enzymes, such as cytochrome P450s that are often required for the production of natural products, are membrane-associated. When producing natural products in *E. coli*, functional expression of heterologous membrane-associated enzymes is often a challenge. The activities of membrane-associated enzymes could be improved by engineering their N-terminal hydrophobic regions. For instance, a signal peptide derived from *E. coli* was fused to an N-terminal truncated β-carotene ketolase from microalgae, and this resulted in enhanced astaxanthin production (Figure 2D) [19]. Among diverse membrane-associated enzymes, plant P450s are especially troublesome because they interact with P450 reductase (CPR) on the inner membranes of the endoplasmic reticulum or mitochondria which are absent in *E. coli* [20]. Thus, appropriate engineering of the P450 N-terminal transmembrane region is important. For artemisinic acid production in *E. coli*, amorphadiene oxidase (AMO) became active by adopting an alternative P450 N-terminal peptide [21]. Yeast is more advantageous for P450 expression because of the presence of intracellular organelles, and a higher artemisinic acid titer (~25 g/l) could be achieved in yeast [22,23]. The interaction of P450 with CPR has also been considered as a way to improve the electron transfer that is required for product formation. In a study on taxadien-5α-ol production in *E. coli*, P450 was directly fused with CPR to facilitate their interaction, leading to improved conversion of taxadiene to taxadien-5α-ol [24]. Nevertheless, it was later reported that maintaining P450 expression levels above those of CPR is better than employing the P450/CPR fusion protein for the production of oxygenated taxanes in *E. coli* [25].

Chaperone Coexpression

Because the production of natural products in *E. coli* involves heterologous enzymes, functional expression of the corresponding heterologous genes is essential. However, heterologous proteins can often aggregate or become misfolded, resulting in inactive enzymes. In some, but not all, cases, correct folding of heterologous proteins can be achieved by

reactions or can utilize multiple substrates beyond its main substrate.
Substrate channeling: a strategy that spatially recruits multiple enzymes using synthetic scaffolds to efficiently deliver and convert metabolic intermediates.
Synthetic small regulatory RNA (sRNA): a *trans*-acting target-specific knockdown tool comprising a noncoding RNA (harboring a target-specific antisense sequence, a scaffold, and a terminator) and Hfq protein; it inhibits translation by binding to the translation initiation region of the target mRNA.
Systems metabolic engineering: an interdisciplinary field of study which integrates traditional metabolic engineering with systems biology, synthetic biology, and evolutionary engineering to provide a holistic approach to microbial metabolism for enhanced production of target chemicals while considering upstream to downstream bioprocesses.
Tunable intergenic regions (TIGRs): a designable intergenic region harboring two hairpin loops flanking an RNase E site; the stability of the target gene mRNA is determined by the secondary structure of the intergenic region.

CellPress
REVIEWS

Key Figure

The Overall Biosynthetic Pathways for Producing Representative Natural Products in *Escherichia coli*

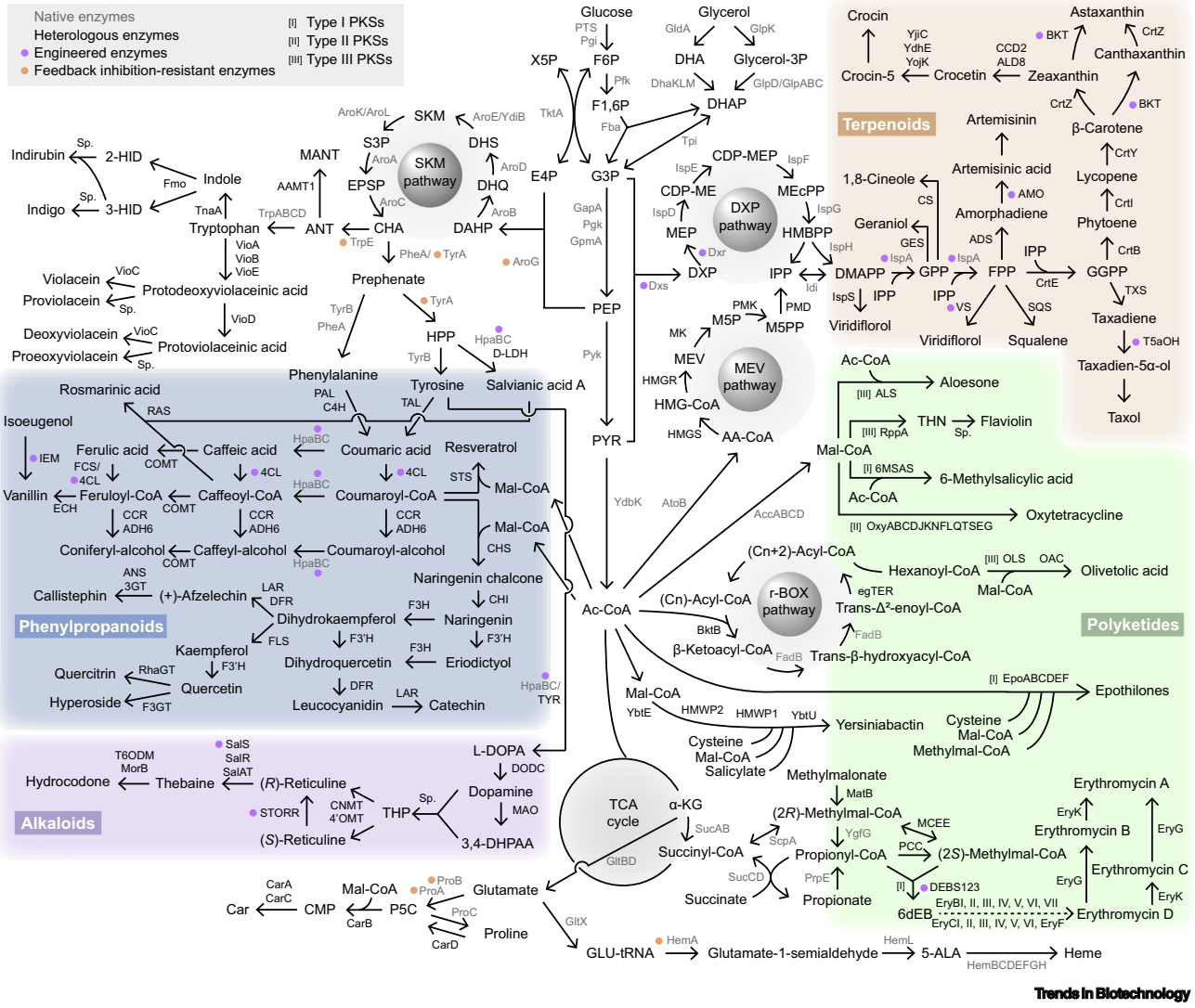

Figure 1. Metabolite abbreviations are as follows: 5-ALA, 5-aminolevulinate; AA-CoA, acetoacetyl-coenzyme A; Ac-CoA, acetyl-CoA; ANT, anthranilate; Car, (5*R*)-carbapen-2-em-3-carboxylic acid; CDP-ME, 4-(cytidine 5'-diphospho)-2-*C*-methyl-D-erythritol; CDP-MEP, 2-phospho-4-(cytidine 5'-diphospho)-2-*C*-methyl-D-erythritol; CHA, chorismate; CMP, (2*S*,5*S*)-5-carboxymethylproline; DAHP, 3-deoxy-D-arabinoheptulosonate 7-phosphate; DHA, dihydroxylacetone; DHAP, dihydroxyacetone phosphate; 6-dEB, 6-deoxyerythronolide B; 3,4-DHPAA, 3,4-dihydroxyphenylacetaldehyde; DHQ, 3-dehydroquinate; DHS, 3-dehydrohikimate; DMAPP, dimethylallyl diphosphate; DXP, 1-deoxy-D-xylulose-5-phosphate; E4P, erythrose 4-phosphate; EPSP, 5-enolpyruvyl-shikimate 3-phosphate; F1,6P, D-fructose 1,6-bisphosphate; F6P, fructose 6-phosphate; FPP, farnesyl diphosphate; G3P, glyceraldehyde 3-phosphate; GGPP, geranylgeranyl diphosphate; glycerol-3P, glycerol-3-phosphate; GPP, geranyl diphosphate; GLU-tRNA, L-glutamyl-tRNA; HMBPP, 1-hydroxy-2-methyl-2-butenyl 4-diphosphate; HMG-CoA, 3-hydroxy-3-methylglutaryl-CoA; HPP, 4-hydroxyphenylpyruvate; IPP, isopentenyl diphosphate; α-KG, α-ketoglutarate; L-DOPA, L-3,4-dihydroxyphenylalanine; Mal-CoA, malonyl-CoA; MANT, methyl anthranilate; MEcPP, 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate; MEP, 2-*C*-methyl-D-erythritol-4-phosphate; methylmal-CoA, methylmalonyl-

(Key Figure legend continued at the bottom of the next page.)

coexpression of chaperone genes (Figure 2E). For example, GroEL and GroES were used to prevent inclusion body formation of large **polyketide synthase (PKS)** proteins involved in the production of epothilone [26] and erythromycin [27]. In addition, the same chaperones allowed correct folding of three plant enzymes in the resveratrol pathway, which resulted in increased resveratrol production [28]. Other strategies such as lowering the temperature, fusion with soluble protein tags, and reducing the level of gene expression can also be employed to solubilize target proteins.

Metabolic Flux Optimization

After constructing a biosynthetic pathway for the generation of a target natural product by introducing and/or engineering key enzymes, optimization of the metabolic flux can further improve production titers. Optimization of the metabolic flux can be achieved by employing various metabolic engineering tools and strategies such as amplifying key pathway genes, downregulating competitive pathway genes, and balancing or modulating pathway gene expression levels. Representative examples are described below.

Increasing Precursor Pools

To maximize the metabolic flux towards a target product, securing sufficient precursor pools by resolving rate-controlling steps in the biosynthetic pathway is important. To this end, amplification, reduction, and optimization of metabolic fluxes are often performed (Figure 2F).

CoA; MEV, mevalonate; M5P, MEV-5-phosphate; M5PP, MEV-5-pyrophosphate; P5C, pyrroline 5-carboxylate; PEP, phosphoenolpyruvate; PYR, pyruvate; S3P, shikimate-3-phosphate; SKM, shikimate; TCA, tricarboxylic acid; THN, 1,3,6,8-tetrahydroxynaphthalene; THP, tetrahydropapaveroline; X5P, xylulose 5-phosphate. Enzyme abbreviations are as follows: AAMT1, methyltransferase1; AccABCD, acetyl-CoA carboxylase; ADH6, alcohol dehydrogenase; ADS, amorphadiene synthase; ALD8, aldehyde dehydrogenase; ALS, aloesone synthase; AMO, amorphadiene oxidase; ANS, anthocyanadin synthase; AroA, 3-phosphoshikimate-1-carboxyvinyltransferase; AroB, 3-dehydroquinate synthase; AroC, chorismate synthase; AroD, 3-dehydroquinate dehydratase; AroE, shikimate dehydrogenase; AroG, DAHP synthase; AroK, shikimate kinase 1; AroL, shikimate kinase 2; AtoB, acetyl-CoA acetyltransferase; BKT, β-carotene ketolase; BktB, thiolase; CarA, carbapenam-3-carboxylate synthase; CarB, carboxymethylproline synthase; CarC, (5*R*)-carbapenem-3-carboxylate synthase; CarD, proline dehydrogenase (putative); CCD2, carotenoid-cleaving dioxygenase; CCR, cinnamoyl-CoA reductase; C3H, 4-coumarate 3-hydroxylase; C4H, cinnamate-4-hydroxylase; CHI, chalcone isomerase; CHS, chalcone synthase; 4CL, 4-coumarate:CoA ligase; CNMT, coclaurine *N*-methyltransferase; COMT, caffeate *O*-methyltransferase; CrtB, phytoene synthase; CrtE, geranylgeranyl pyrophosphate synthetase; CrtI, phytoene dehydrogenase; CrtY, lycopene β-cyclase; CrtZ, β-carotene hydroxylase; CS, cineole synthase; DEBS, deoxyerythronolide B synthase; DFR, dihydroflavonol 4-reductase; DFR, dihydroflavonol 4-reductase; DhaKLM, dihydroxyacetone kinase; D-LDH, D-lactate dehydrogenase; DODC, dopa decarboxylase; Dxr, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; Dxs, 1-deoxy-D-xylulose-5-phosphate synthase; ECH, enoyl-CoA hydratase/aldolase; egTER, enoyl-CoA reductase; Eno, enolase; EpoABCDEF, epothilone biosynthetic enzymes; EryBI/II/III/IV/V/VI/VII,EryCI/II/III/IV/V/VI,EryF, erythromycin biosynthetic enzymes; EryG, erythromycin *O*-methyltransferase; EryK, erythromycin C-12 hydroxylase; F3′H, flavonoid 3′-hydroxylase; F3H, flavanone 3β-hydroxylase; FadB, hydroxyacyl-CoA dehydrogenase; enoyl-CoA hydratase; Fba, fructose-bishposphate aldolase; FCS, trans-feruloyl-CoA synthetase; FLS, flavonol synthase; Fmo, flavin-containing monooxygenase; GapA, glyceraldehyde-3-phosphate dehydrogenase A; GES, geraniol synthase; GldA, glycerol dehydrogenase; GlpD/GlpABC, glycerol-3-phosphate dehydrogenases; GlpK, glycerol kinase; GltBD, glutamate synthase; GltX, glutamyl-tRNA synthetase; GpmA, 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase; 3GT, 3-*O*-glycosyltransferase; HemA, glutamyl-tRNA reductase; HemBCDEFGH, heme biosynthetic enzymes; HemL, glutamate-1-semialdehyde 2,1-aminomutase; HID, hydroxyindole; HMGR, hydroxymethylglutaryl-CoA reductase; HMGS, hydroxymethylglutaryl-CoA synthase; HMWP, high-molecular-weight protein; HpaBC, 4-hydroxypropanoic acid 3-hydroxylase; Idi, isopentenyl-diphosphate Δ-isomerase; IEM, isoeugenol monooxygenase; IspA, geranyl diphosphate synthase; IspD, 2-*C*-methyl-D-erythritol 4-phosphate cytidylyltransferase; IspE, 4-(cytidine 5'-diphospho)-2-*C*-methyl-D-erythritol kinase; IspF, 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate synthase; IspG, (*E*)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase; IspH, 1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate reductase; IspS, isoprene synthase; LAR, leucoanthocyanidin reductase; MAO, monoamine oxidase; MatB, malonyl-CoA synthetase; MCEE, methylmalonyl-CoA/ethylmalonyl-CoA epimerase; MK, mevalonate kinase; MorB, morphinone reductase; 6MSAS, 6-methylsalicylic acid synthase; OAC, olivetolic acid cyclase; OLS, olivetolic acid synthase; 4'OMT, 3′-hydroxy-*N*-methylcoclaurine 4′-*O*-methyltransferase; OxyABCDJKNFLQTSEG, oxytetracycline biosynthetic enzymes; PAL, phenylalanine ammonia lyase; PCC, propionyl-CoA carboxylase; Pfk, phosphofructokinase; Pgi, G6P isomerase; Pgk, phosphoglycerate kinase; PheA/TyrA, chorismate mutase/prephenate dehydratase; PKS, polyketide synthase; PMD, phosphomevalonate decarboxylase; PMK, phosphomevalonate kinase; ProA, glutamate 5-semialdehyde dehydrogenase; ProB, glutamate 5-kinase; ProC, pyrroline-5-carboxylate reductase; PrpE, propionyl-CoA synthetase; Pyk, pyruvate kinase; RAS, rosmarinic acid synthase; RppA, 1,3,6,8-tetrahydroxynaphthalene synthase; SalAT, salutaridinol acetyltransferase; SalR, salutaridine reductase; SalS, salutaridine synthase; ScpA, methylmalonyl-CoA mutase; SQS, squalene synthase; STORR, epimerase of (*S*)- to (*R*)-reticuline; STS, stilbene synthase; SucAB, 2-oxoglutarate dehydrogenase; SucCD, succinyl-CoA synthetase; T5aOH, taxadiene 5-α hydroxylase; TAL, tyrosine ammonia-lyase; TktA, transketolase 1; TnaA, tryptophanase α subunit; T6ODM, thebaine 6-*O*-demethylase; Tpi, triose-phosphate isomerase; TrpAB, tryptophan synthase; TrpC, indole-3-glycerol phosphate synthase; TrpE/TrpD, anthranilate synthase; TYR, tyrosinase; TyrB, tyrosine aminotransferase; TXS, taxadiene synthase; YbtE, Salicyl-AMP ligase; YbtU, yersiniabactin biosynthetic protein; YdbK, pyruvate-ferredoxin/flavodoxin oxidoreductase; YdiB, shikimatede hydrogenase; YgfG, methylmalonyl-CoA decarboxylase; YjiC/YdhE/YojK, UDP-glucuronosyltransferases; VioA, flavin-dependent L-tryptophan oxidase; VioB, 2-imino-3-(indol-3-yl)propanoate dimerase; VioC, monooxygenase; VioD, protodeoxyviolaceinate monooxygenase; VioE, violacein biosynthesis enzyme; VS, viridiflorol synthase. Other abbreviations: PTS, phosphotransferase system; r-BOX, reversal of β-oxidation; Sp., spontaneous.

For the production of polyketides and phenylpropanoids, securing a sufficient malonyl-CoA pool is necessary (Figure 1). In one study, overexpression of acetyl-CoA synthase and a rate-controlling enzyme acetyl-CoA carboxylase resulted in 15-fold higher production of malonyl-CoA, which consequently led to a fourfold increase in phloroglucinol titer (Figure 2F) [29]. In another study, knocking down *pabA* (encoding *p*-aminobenzoate synthetase) using **synthetic small regulatory RNA (sRNA)** resulted in increased accumulation of malonyl-CoA, leading to enhanced production of 6-methylsalicylic acid, aloesone, resveratrol, and naringenin [30]. **CRISPR interference (CRISPRi)** was also employed to knock down fatty acid biosynthetic genes *fabB*, *fabI*, and *fabF* to increase the intracellular concentration of malonyl-CoA. This resulted in enhanced resveratrol production up to 188.1 mg/l, which is sixfold higher than that obtained with the control strain [31].

For terpenoid biosynthesis, a sufficient supply of isopentenyl diphosphate (IPP) and DMAPP is important (Figure 1). This has often been achieved by overexpression of the *idi* and *dxs* genes in the DXP pathway [32]. In addition, a balanced supply of pyruvate and glyceraldehyde 3-phosphate (G3P), the precursors of the DXP pathway, is important for high-level production of terpenoids. G3P supply is often limited in terpenoid production, and thus increasing G3P supply through overexpression of *ppsA* (encoding phosphoenolpyruvate/PEP synthase) and *pck* (encoding PEP carboxykinase), and inactivation of *pykFA* (encoding isozymes Pyk-I and -II), can lead to higher

Box 1. Production of Different Categories of Natural Products in *E. coli*

Terpenoids, which are derived from terpenes comprising isoprene units, constitute the largest class of natural products. Artemisinin and taxol (paclitaxel), which are antimalarial and anticancer agents, respectively, are well-known terpenoid drugs produced by engineered microorganisms. *E. coli* harbors the DXP pathway that synthesizes IPP and DMAPP, the building blocks of terpenoids. To further increase pathway flux towards these precursors, and to circumvent unknown regulatory mechanisms on the native DXP pathway, a heterologous MEV pathway has often been additionally introduced [88].

Phenylpropanoids are biosynthesized from either tyrosine or phenylalanine, ubiquitous aromatic amino acids that are produced through the shikimate pathway. Well-known phenylpropanoids include quercetin, kaempferol, and resveratrol. Flavonoids and stilbenoids are the two major phenylpropanoid subgroups for which common intermediates are produced by condensation of one molecule of coumaroyl-CoA with three molecules of malonyl-CoA. Because it is important to secure a sufficient malonyl-CoA pool in *E. coli*, several different biosensors have been developed to efficiently screen for malonyl-CoA overproducing strains [30,89]. In addition, mutant enzymes which are resistant to feedback inhibition by precursors (such as phenylalanine and tyrosine) can enhance phenylpropanoid production in *E. coli* [17,18].

Alkaloid is a general term that embraces nitrogen-containing natural products, of which morphine, cocaine, and vinblastine are the best known. Alkaloids are classified into multiple subgroups, including monoterpene indole alkaloids (MIAs), tropane alkaloids, and BIAs, according to their different precursors or chemical structures. Of these, only BIAs have been produced in *E. coli*, which has been found to be an excellent host organism because a high flux of tyrosine, the common BIA precursor, can be achieved [40]. Reticuline, the most important intermediate of BIAs, is synthesized from tyrosine through multiple pathways (Box 2). The BIA pathway from reticuline then splits into sanguinarine and morphine pathways. To date, only the latter has been constructed in *E. coli*. Because morphinan alkaloids can only be produced from (*R*)-reticuline, conversion of (*S*)-reticuline to (*R*)-reticuline or simultaneous production of both isomers has been reported [40].

Polyketides are formed by repetitive condensation of small carbon units (e.g., acetyl-CoA, malonyl-CoA, propionyl-CoA, methylmalonyl-CoA, and hexanoyl-CoA), catalyzed by PKS enzymes. The production of polyketides in *E. coli* has been challenging mainly because of the low availability (e.g., malonyl-CoA) or no production (e.g., methylmalonyl-CoA) of precursors without further engineering, the difficulty of functional expression of PKSs, and the need to introduce heterologous cofactors (e.g., F420) or supplementary genes for generating functional holoenzymes (e.g., *sfp* from *Bacillus subtilis* for the activation of acyl-carrier proteins). One of the most popular examples of polyketides produced in *E. coli* is erythromycin, which is synthesized by a type I modular PKS [81]. Several ingenious strategies including integration of PKS genes into the *E. coli* genome [90], chaperone overexpression [27], and increasing precursor pools [91] have been employed, in addition to typical metabolic engineering approaches, for the successful production of erythromycin in *E. coli*. However, production of aromatic polyketides by type II PKSs has been relatively unsuccessful because of the poor solubility of the core PKS enzymes [92]. It is worthwhile mentioning nonribosomal peptides (NRPs), peptide natural products that are independent of ribosomes and mRNA [93]. However, mainly because of their large BGCs and complexity, only a few NRPs have been produced in *E. coli* (e.g., echinomycin and yersiniabactin, Table 1).

lycopene production [33]. Further study demonstrated that downregulation of *gapA* (encoding G3P dehydrogenase) is more efficient than *ppsA* overexpression in increasing lycopene production [34]. Furthermore, introduction of a heterologous mevalonate (MEV) pathway into *E. coli* was shown to be effective in enhancing terpenoid production, improving the isoprene titer to 24.0 g/l, which is 20-fold higher than that of the strain without the MEV pathway [35].

For the production of opioids, a group of chemicals belonging to the alkaloids, metabolic engineering strategies to overproduce tyrosine, the common BIA precursor, is described in detail in Box 2 [36].

Increasing Cofactor Levels

In addition to increasing precursor supplies through flux manipulation and optimization, a balanced supply of cofactors required for enzyme activities is also important. For example, NADPH is an important cofactor for NADPH-dependent enzymes such as class II P450s [37]. One strategy to increase NADPH levels is to reinforce the metabolic flux of the pentose phosphate pathway, which can be done by knocking out *pgi* (encoding glucose 6-phosphate isomerase) and *ppc* (encoding PEP carboxylase) [38]. This allowed enhanced production of (+)-catechin to 754.0 mg/l, which is 943% higher than the yield (72.3 mg/l) obtained without knocking out *ppc* and *pgi* (Figure 2G). In addition, ATP, the major intracellular energy source, is required in many reactions, and thus increasing the ATP pool is also important to enhance natural product biosynthesis. Although sufficient ATP is generated by *E. coli* under aerobic culture conditions, the actual amount of ATP necessary for natural product biosynthesis can sometimes exceed this amount, even when their titers are low. Thus, replenishing ATP availability can be an effective strategy to enhance natural product biosynthesis. For example, enhanced production of pinocembrin (from 65.77 mg/l to 102.02 mg/l) has been reported by knocking down *metK* (encoding *S*-adenosyl-L-methionine synthase) and *proB* (encoding glutamate 5-kinase) using CRISPRi, which increased ATP levels in *E. coli* (Figure 2G) [39]. Although only the cases of NADPH and ATP are described above, other cofactors including FAD [19] and heme [40] may need to be considered where appropriate.

Balancing Gene Expression Levels

Increasing metabolic flux by simply overexpressing target genes often results in inefficient carbon utilization or metabolic burden [5]. Hence, balanced expression of genes is important to maximize target compound production while maintaining optimal cell growth. For example, the expression levels of the key biosynthetic genes for lycopene production (*dxs*, *idi*, and *crtE*) were fine-tuned by screening ribosome-binding site (RBS) sequences of different strengths. Employing the best RBS combinations resulted in the production of 3.52 g/l of lycopene [41].

Although fine-tuning of different components of gene expression (e.g., promoter strength, RBS strength, and 5′-untranslated region sequences) has been effective in balancing gene expression levels, it is labor-intensive and time-consuming, especially when a large number of target genes are involved. To solve this problem, a new method was proposed to balance the expression of multiple genes by high-throughput screening of different **tunable intergenic regions (TIGRs)** (Figure 2H) [42]. Balancing the expression levels of the upstream MEV pathway genes was achieved by screening TIGR libraries, and successfully increased MEV production by sevenfold. Later, the expression levels of downstream MEV pathway genes were also optimized by screening TIGR libraries [43]. Dynamically regulating the constructed MEV pathway by employing promoters that are responsive to toxic intermediates (IPP and FPP) led to enhanced zeaxanthin production by 2.1-fold. The final engineered strain produced 722.46 mg/l of zeaxanthin by fed-batch fermentation.

Table 1. Natural Products Biosynthesized by Engineered *E. coli* Strains

Product	Substrate	Concentration (mg/l)	Content (mg/gDCW[a])	Scale[b]	Refs
Terpenoids					
Lycopene	Glycerol	3520	50.6	3 l fed-batch	[41]
	Luria–Bertani (LB)	NR[c]	448	Six-well flat-bottomed microtiter plates	[106]
β-Carotene	Glycerol	3200	NR	3 l fed-batch	[98]
Zeaxanthin	Glucose	722.46	23.16	3 l fed-batch	[98]
Astaxanthin	Glycerol	432.82	7.12	1.6 l fed-batch	[19]
Amorphadiene	Glucose	30 000	NR	100 ml fed-batch	[107]
Artemisinic acid	Glycerol	105	NR	50 ml flask	[21]
Taxadiene	Glycerol	1020	NR	1 l fed-batch	[24]
Oxygenated taxanes	Glycerol	570	NR	925 ml fed-batch	[25]
Geraniol	Glucose	2000	NR	2 l fed-batch	[108]
Viridiflorol	Glucose	25 700	NR	100 ml fed-batch	[107]
Crocin-5	Glycerol	NR	NR	10 ml flask	[109]
Polyketides					
6-Methylsalicylic acid	Glycerol	440.3	NR	2 l fed-batch	[30]
Aloesone	Glucose	30.9	NR	50 ml flask	[30]
Flaviolin	Glucose	26.0	NR	50 ml flask	[30]
Epothilone	LB/propionate	<0.001	NR	NR	[26]
Erythromycin A	Glycerol/propionate	10	NR	100 ml flask	[27]
Oxytetracycline	LB	2.0	NR	25 ml flask	[110]
Olivetolic acid	Glycerol	80	NR	400 ml batch	[111]
Phenylpropanoids					
(2*S*)-Naringenin	Glucose/L-tyrosine	421.6	NR	50 ml flask	[112]
	Glycerol	103.8	NR	50 ml flask	[30]
Resveratrol	*p*-Coumaric acid	2300	NR	25 ml flask	[76]
	Glucose	304.5	NR	25 ml flask	[31]
Vanillin	Isoeugenol	~4500	NR	15 ml flask	[113]
	Glycerol	24.7	NR	50 ml flask	[114]
(+)-Afzelechin	*p*-Coumaric acid	40.7	NR	125 ml flask (coculture)	[115]
	Glucose	26.1	NR	125 ml flask (coculture)	[49]
Coniferyl alcohol	Glucose/glycerol	124.9	NR	1 l fed-batch (coculture)	[116]
Caffeyl alcohol	Glucose/glycerol	854.1	NR	1 l fed-batch (coculture)	[116]
Hyperoside (quercetin 3-*O*-galactoside)	Sucrose/quercetin	940	NR	100 ml flask	[117]
Quercitrin (quercetin 3-*O*-rhamnoside)	Sucrose/quercetin	1176	NR	100 ml flask	[117]
(+)-Catechin	Eriodictyol	910.9	NR	40 ml flask	[38]
Rosmarinic acid	Xylose/glucose	172	NR	100 ml flask (coculture)	[50]
(2*S*)-Pinocembrin	Glucose	525.8	NR	1.5 l fermentation	[118]
Salicylate 2-*O*-β-D-glucoside	Glucose	2500	NR	NR (coculture)	[51]
Alkaloids					
Thebaine	Terrific-broth	2.1	NR	100 ml flask (stepwise culture)	[40]

Table 1. (continued)

Product	Substrate	Concentration (mg/l)	Content (mg/gDCW[a])	Scale[b]	Refs
	(TB)/glycerol				
Hydrocodone	TB/glycerol	0.36	NR	100 ml flask (stepwise culture)	[40]
(*S*)-Reticuline	Turbo broth/glycerol	46.0	NR	3 l fermentation	[36]
(*R,S*)-Reticuline	TB/glycerol	16	NR	50 ml flask	[40]
(*R,S*)-THP	Glycerol	287	NR	300 ml flask (stepwise culture)	[52]
L-DOPA	Glucose	25 530	NR	3 l fed-batch	[119]
Dopamine	Glycerol	2150	NR	1 l fed-batch	[52]
Others					
Yersiniabactin	Glycerol	17.4	NR	25 ml flask	[120]
Echinomycin	Glucose	0.3	NR	1.5 l fed-batch	[121]
Methyl anthranilate	Glucose	4470	NR	1.8 l fed-batch (two-phase culture using tributyrin)	[18]
Violacein	Glucose	4070	NR	400 ml fed-batch	[69]
Deoxyviolacein	Glucose	1230	NR	400 ml fed-batch	[69]
Heme	Glucose/L-glutarate	239.2	NR	2 l fed-batch	[105]
Indigo	Glucose	640	NR	2 l fed-batch	[17]
Indirubin	Glucose	56	NR	2 l fed-batch	[17]

[a]DCW, dry cell weight.
[b]Working volumes are noted; detailed culture types are described in parentheses.
[c]NR, not reported.

Substrate Channeling

In addition to balancing the expression levels of heterologous genes, efficient conversion of substrates to target compounds can be achieved by **substrate channeling** [44], which can effectively reduce metabolic burden, toxic intermediates accumulation, and substrate diffusion (Figure 2I). For example, synthetic protein scaffolds were developed to recruit target enzymes to proximity through protein–protein interactions. By bringing the biosynthetic enzymes of the MEV pathway together using synthetic protein scaffolds (comprising SH3, PDZ, and GBD domains), the MEV titer could be increased by 77-fold [44]. By employing the same synthetic protein scaffolds, the production of (+)-catechin could be increased by 155.6% through substrate channeling (Figure 2I) [38]. In another study, synthetic DNA scaffolds were developed for recruiting enzymes fused to different zinc-finger domains, and this successfully enhanced the production of resveratrol, 1,2-propanediol, and MEV [45]. These studies suggest that substrate channeling can indeed be an effective strategy for improving the production of natural products.

Pathway Modularization

Many natural products are produced from long and intertwined biosynthetic pathways. Given the difficulty of manipulating the expression levels of several individual genes, the whole pathway can be segregated into several modules (i.e., pathway modularization) to allow more efficient pathway engineering (Figure 2J) [46,47]. For example, the DXP pathway was divided into upstream and the downstream modules, and their expression levels were balanced using different promoters. This effectively reduced the accumulation of toxic intermediates, increasing isoprene production by 4.72-fold [14]. In a study on astaxanthin production, segmentation of the heterologous MEV pathway into four modules followed by balancing their expression levels using different promoters

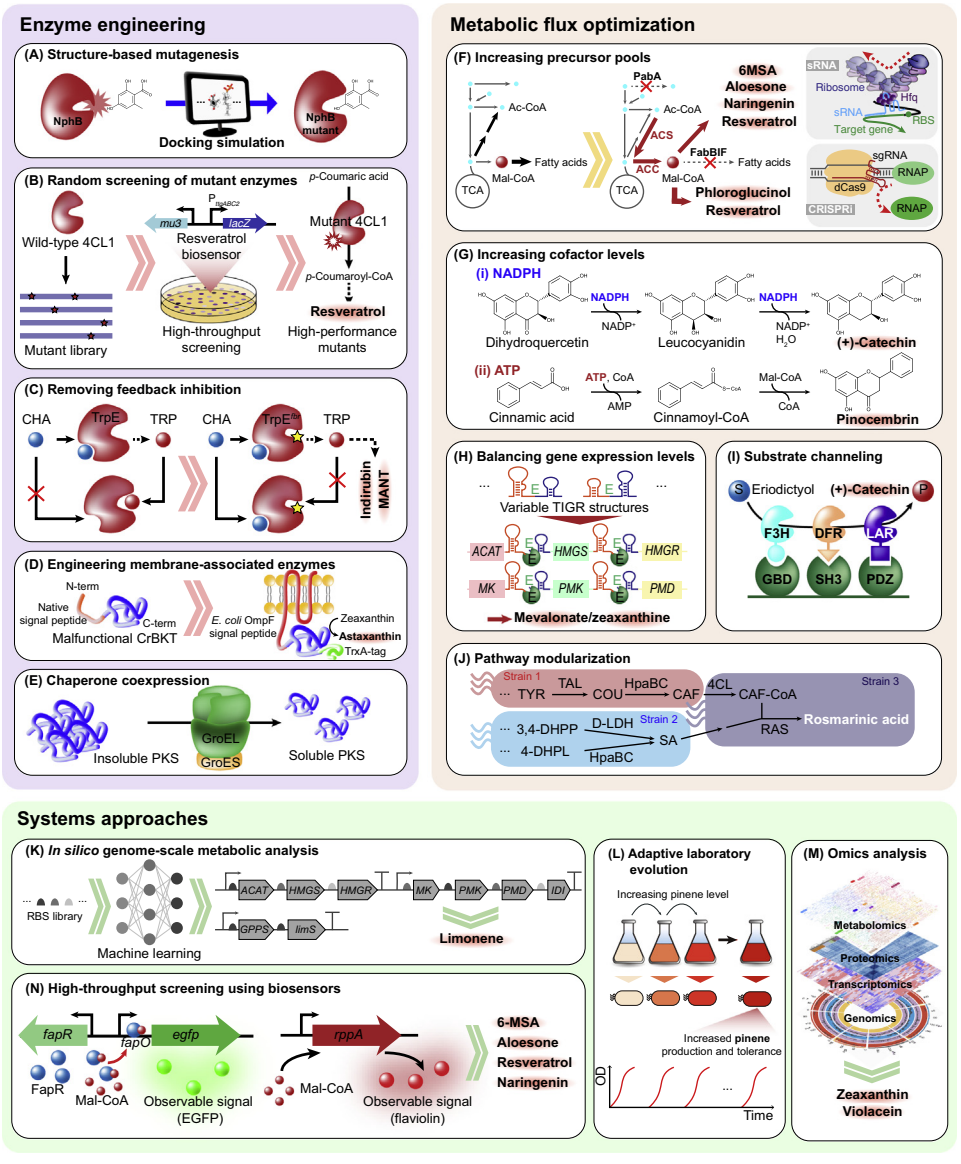

Figure 2. Tools and Strategies Employed for Heterologous Production of Natural Products in *Escherichia coli*. Enzyme engineering strategies to enhance bioconversion efficiency are described in the box at the upper left (A–E). Metabolic flux optimization strategies are described in the box at the upper right (F–J). Systems approaches are described in the box at the bottom (K–N). (A) Enzyme structure-based mutagenesis. (B) Random screening of mutant enzymes to select mutants with high performance. (C) Removing feedback inhibition. (D) Engineering membrane-associated enzymes to recover their activities. (E) Chaperone coexpression to aid in enzyme solubility. (F) Increasing precursor pools by employing tools including CRISPR interference (CRISPRi) and synthetic sRNA. (G) Increasing cofactor levels. (H) Balancing expression levels of multiple genes using tools such as TIGRs. (I) Substrate channeling to spatially recruit enzymes using synthetic scaffolds. (J) Pathway modularization and optimization. (K) *In silico* genome-scale metabolic analysis to predict gene manipulation targets for enhanced production of target compounds. (L) Adaptive laboratory evolution (ALE) to generate strains with desirable phenotypes. (M) Omics analysis to diagnose the whole cellular system from a holistic view. (N) High-throughput screening using biosensors. Enzyme abbreviations are as follows: ACS, acetyl-CoA synthetase; ACC, acetyl-CoA carboxylase; 4CL, 4-coumarate:CoA ligase; CrBKT, *Chlamydomonas reinhardtii* β-carotene ketolase; DFR, dihydroflavonol 4-reductase; FabBIF, fatty acid biosynthetic enzymes; FapR, transcriptional repressor; F3H, flavanone 3β-hydroxylase; GPPS, geranyl diphosphate synthase; HpaBC, 4-

(Figure legend continued at the bottom of the next page.)

and RBS sequences led to the production of 320 mg/l of astaxanthin [48].

Alternatively, a modularized metabolic pathway can be distributed into different strains to effectively reduce the metabolic burden. For example, the long anthocyanin biosynthetic pathway comprising 15 genes was distributed into four different *E. coli* strains, and these four strains were cocultured. This resulted in the production of 26.1 mg/l of (+)-afzelechin and 9.5 mg/l of pelargonidin-3-*O*-glucoside (callistephin) [49]. The same strategy was also employed for the production of rosmarinic acid and salicylate 2-*O*-β-D-glucoside, which resulted in higher titers than those produced by single strains (Figure 2J) [50,51]. Coculture can also overcome the inherent limitation of employing a single species, as demonstrated by the production of oxygenated taxanes by an *E. coli*–yeast consortium [47]. P450s have been known to be extremely difficult to be functionally expressed in *E. coli.* Thus, a P450 gene encoding taxadiene 5α-hydroxylase was expressed in yeast for producing oxygenated taxanes from taxadiene produced by engineered *E. coli*. In another study, the reticuline biosynthetic pathway was distributed into three different *E. coli* strains followed by their stepwise cultivation (Box 2), resulting in the production of 48 mg/l of reticuline [52]. Additional incorporation of thebaine or hydrocodone producers led to production of these opioids with titers higher than those obtained in yeast (Box 2 for further details) [40]. In general, the use of a single strain is of course best if it yields the desired performance, but, if not, one does not need to rely on single strain strategy. The above strategies can be considered as a 'divide and conquer' approach, which can be used to divide long pathways into several modules in one host strain or by employing multiple host strains capable of performing the desired reactions.

Systems Approaches

In the above, we described rational and intuitive engineering strategies to enhance the production of various natural products in *E. coli*. However, engineering *E. coli* without considering the complex metabolic and regulatory networks at the systems level can limit the production of target natural products to below the desired level. Overcoming this constraint necessitates the use of systems metabolic engineering, which integrates systems biology, synthetic biology, and evolutionary engineering with traditional metabolic engineering. Systems metabolic engineering has been shown to be an efficient strategy in developing various microbial cell factories [53,54]. Many tools and strategies are available for implementing systems metabolic engineering; several of these have been applied to natural product biosynthesis in *E. coli* and are showcased below.

In Silico Genome-Scale Metabolic Analysis

Various computational tools have been developed to predict the outcomes of a given perturbation to the microbial system (Figure 2K). The development of comprehensive GEMs for *E. coli* has allowed accurate prediction of biological functions within a cell [55]. Several algorithms and programs have also been developed to simulate GEMs with different objectives, as recently reviewed [56,57], and have been used to develop strains for natural product biosynthesis. In one

hydroxyphenylacetate 3-hydroxylase; IDI, isopentenyl-diphosphate Δ-isomerase; LAR, leucoanthocyanidin reductase; D-LDH, D-lactate dehydrogenase; NphB, prenyltransferase; PabA, *p*-aminobenzoate synthetase; PKS, polyketide synthase; RAS, rosmarinic acid synthase; TAL, tyrosine ammonia lyase; TrpE, anthranilate synthase; TrxA, thioredoxin 1. Abbreviations for metabolites are as follows: Ac-CoA, acetyl-CoA; CAF, caffeic acid; CAF-CoA, caffeoyl-CoA; CHA, chorismate; COU, *p*-coumaric acid; 4-DHPL, 4-dihydroxyphenyllactate; 3,4-DHPP, 3,4-dihydroxyphenylpyruvate; Mal-CoA, malonyl-CoA; MANT, methyl anthranilate; 6MSA, 6-methylsalicylic acid; SA, salvianic acid A; TCA, tricarboxylic acid; TRP, L-tryptophan; TYR, L-tyrosine. Gene abbreviations are as follows: *ACAT*, acetyl-CoA acetyltransferase; *HMGR*, HMG-CoA reductase; *HMGS*, hydroxymethylglutaryl-CoA (HMG-CoA) synthase; *limS*, limonene synthase; *MK*, mevalonate kinase; *PMD*, diphosphomevalonate decarboxylase; *PMK*, phosphomevalonate kinase. Other abbreviations: E, RNase E site (in green)/RNase E (in black); RBS, ribosome binding site; RNAP, RNA polymerase; TIGR, tunable intergenic region.

example, OptForce [58] was developed and employed to identify knockout and overexpression gene targets that can potentially increase the intracellular malonyl-CoA pool. A strain engineered based on *in silico* predicted targets generated a higher level of malonyl-CoA, leading to increased naringenin production. In another study, the FVSEOF (flux variability scanning based on enforced objective flux) algorithm was used to identify overexpression gene targets, and this approach was effective in increasing astaxanthin production [19].

In silico analysis can also be employed to identify enzyme candidates that are likely to be functional in *E. coli*. For instance, the MiBiG (minimum information about a biosynthetic gene cluster) repository was employed to identify type II minimal PKS pathway components that are phylogenetically close to *E. coli* [59]. Among the identified biosynthetic gene clusters (BGCs), the PKS pathway from *Photorhabdus luminiscens* TT01 enabled the production of diverse polyketides in *E. coli*.

Machine learning has recently been used to more accurately decipher genotype–phenotype relationships [60]. For example, to balance metabolic flux towards (*S*)-limonene, machine learning was used to predict optimal RBS sequences for the key biosynthetic genes [61]. RBS libraries were constructed for each gene, and the results of screening less than 3% of the libraries were used to train the algorithm. It successfully predicted a set of RBS sequences that enhanced (*S*)-limonene production from 593 mg/l to 1.15 g/l, without the need to screen the entire libraries (Figure 2K). As showcased above, *in silico* genome-scale metabolic analysis is an excellent strategy that can greatly reduce the time and effort necessary to identify effective gene manipulation targets for the construction of high-performance microbial cell factories.

Adaptive Laboratory Evolution

Because we do not fully understand the entire metabolic and regulatory mechanisms even for the best-studied organism *E. coli*, numerous traits still cannot be obtained through rational or systems-level metabolic engineering. This is why the 'random mutagenesis and

Box 2. Opioid Production in *E. coli*

Opioids are naturally derived BIAs from the opium poppy (*Papaver somniferum*) which are widely used for medical applications such as analgesics [94]. Owing to the low productivity of plant sources, both chemical synthesis and microbial production of opioids have been pursued. Although chemical synthesis of opioids is possible, their complex structures make it difficult to synthesize them in an economically viable way [95]. In this regard, genetically amenable microbial hosts such as yeast and *E. coli* have been used for opioid production. In particular, *E. coli* has been recognized as a favorable host for opioid production because of its high productivity of the common opioid precursor, tyrosine [40].

Several metabolic engineering strategies have recently been developed to minimize undesirable reactions and enzyme inhibition in a long biosynthetic pathway (Figure IA), aiming at the industrial production of opioids. In one study, *ppsA* and *tktA* (encoding transketolase) involved in the shikimate pathway were overexpressed, resulting in the production of 4.37 g/l of tyrosine from glycerol. This consequently led to the production of 46.0 mg/l of (*S*)-reticuline [36]. Monoamine oxidase (MAO), that is responsible for the conversion of dopamine into 3,4-dihydroxyphenylacetaldehyde (DHPAA), is a membrane-associated enzyme, and its expression is therefore difficult in *E. coli*. Thus, structure-based enzyme engineering of *Bombyx mori* DHPAA synthase (DHPAAS) was performed for the simultaneous production of dopamine and DHPAA, which resulted in increased (*R*,*S*)-tetrahydropapaveroline (THP) production [12]. In another study, to reduce the metabolic burden, the reticuline biosynthetic pathway was segmented and distributed into three strains (each producing dopamine, THP, and reticuline) [52]. Stepwise culture of these strains resulted in 287 mg/l of THP production. Later, the downstream pathway towards thebaine or hydrocodone was introduced into another *E. coli* strain to allow the production of opioids in four steps, yielding 2.1 mg/l of thebaine or 0.36 mg/l of hydrocodone, respectively (Figure IB) [40]. Despite higher titers than could be obtained from yeast (6.4 and 0.3 μg/l of thebaine and hydrocodone, respectively) [96], these are much lower than the titers of other natural products produced by *E. coli*, which is possibly due to weak metabolic flux towards (*R*)-reticuline [40]. In future studies, various systems metabolic engineering strategies can be applied for more efficient opioid production. In addition, researchers should be cautious in their studies because easier production of these psychotropic drugs may become a serious social threat if not properly contained.

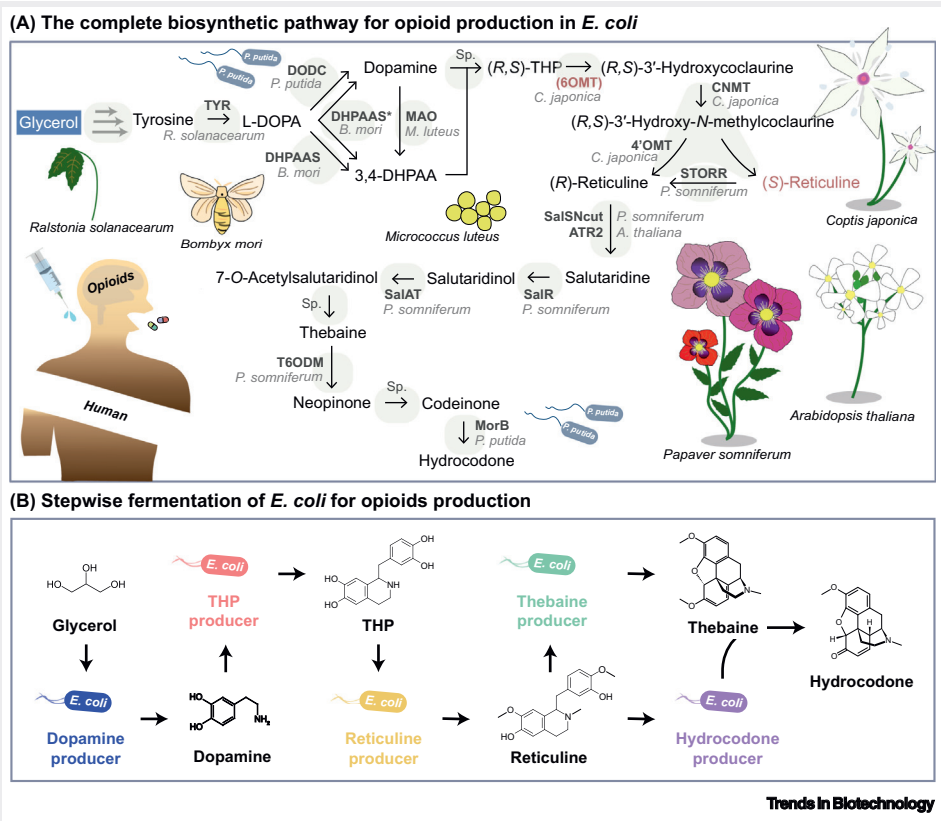

Figure I. Biosynthetic Pathways for the Production of Opioids in *Escherichia coli*. (A) The complete biosynthetic pathway for opioid production in *E. coli*. When 6OMT is not expressed, both the *S* and *R* forms of reticuline are produced, whereas only (*S*)-reticuline is produced when 6OMT is expressed. Organisms from which each enzyme originated are noted below the corresponding enzymes. (B) Stepwise fermentation of engineered *E. coli* strains for opioid production. Abbreviations: ATR2, NADPH-cytochrome P450 reductase 2; CNMT, coclaurine *N*-methyltransferase; DHPAAS, 3,4-dihydroxyphenylacetaldehyde synthase; DHPAAS*, engineered DHPAAS; 3,4-DHPAA, 3,4-dihydroxyphenylacetaldehyde; DODC, DOPA decarboxylase; MAO, monoamine oxidase; MorB, morphinone reductase; 6OMT, 6-*O*-methyltransferase; 4′ OMT, 4′-*O*-methyltransferase; *P. putida*, *Pseudomonas putida*; STORR, epimerase of (*S*)- to (*R*)-reticuline; SalSNcut, N-terminal-truncated salutaridine synthase; SalR, salutaridine reductase; SalAT, salutaridinol acetyltransferase; Sp., spontaneous; (*R,S*)-THP, (*R,S*)-tetrahydropapaveroline; T6ODM, thebaine 6-*O*-demethylase; TYR, tyrosinase.

selection' strategy has been used in industrial biotechnology sector for many years. As better equipment and robotic systems with automatic and preprogrammable capabilities have become available, **adaptive laboratory evolution (ALE)**, which mimics the process of natural selection in a laboratory, is being increasingly used for improving strain performance (Figure 2L) [62]. ALE is distinguishable from directed evolution in that it is applied to strains rather than to enzymes. ALE allows strains with improved product tolerance, growth rate, and consumption rate of nonpreferred substrates to outgrow the whole population [63]. For example, to increase pinene tolerance and production, ALE was performed in the presence of 2.0% (w/v) pinene [64]. In addition, to generate high-performance GPP synthase mutants, directed evolution was performed by error-prone PCR and DNA shuffling to further increase pinene production by 2.2-fold (Figure 2L). Unexpected mutations are often observed while performing ALE. Examining these mutations can provide additional insight into the complex metabolic and regulatory networks of microorganisms, thus providing new engineering guidelines to promote target compound production. Because DNA

sequencing costs have decreased significantly, it can be performed even for a massive number of evolved strains. Thus, reverse engineering of a production strain based on DNA sequencing of evolved strains obtained by ALE will be widely applicable to the development of high-performance strains.

Omics Analysis

Monitoring intracellular metabolic changes following the introduction of heterologous BGCs is important for the identification of beneficial gene manipulation targets that are otherwise difficult to select rationally. In this regard, interpretation of omics data can provide a holistic view on metabolism and regulation that is useful for developing other metabolic engineering strategies (Figure 2M). In a study on zeaxanthin production in *E. coli*, proteomic analysis was performed to identify differentially expressed proteins upon introduction of the MEV pathway [43]. These proteins not only included metabolic enzymes involved in precursor and cofactor availability but also those involved in nonobvious functions. For example, FtsZ, MreB, and RodZ, proteins involved in cell division and shape, were found to be downregulated in the zeaxanthin overproducer strain. Interestingly, further downregulating these proteins resulted in even higher production of zeaxanthin [43]. Instead of performing genome-scale metabolic simulation to obtain calculated metabolic flux values, actual flux values (the fluxome) can be calculated by measuring isotopomer distributions through **^{13}C-metabolic flux analysis (^{13}C-MFA)** [65,66]. ^{13}C-MFA was applied to identify gene manipulation targets to enhance the production of MEV [67] and violacein [68] in *E. coli*. With the development of other analytical technologies (e.g., RNA-seq for transcriptome analysis, mass spectrometric analysis for proteome analysis), omics analysis has become faster and easier. Such advances result in better understanding of cellular metabolism and regulation at the systems level, which in turn facilitates the development of high-performance strains.

Box 3. Production of Natural and Nonnatural Carotenoids in *E. coli*

Carotenoids, mostly tetraterpenoids, are synthesized by plants, fungi, algae, and bacteria, and are responsible for their red, orange, and yellow colors [97]. As effective antioxidants, carotenoids are utilized in diverse industries including healthcare and cosmetics. There have been numerous studies on efficiently producing carotenoids in *E. coli*. Among >600 different carotenoids that have been characterized in nature, lycopene, β-carotene, zeaxanthin, and astaxanthin are the best-studied and have been produced in *E. coli* (Figure I). To the best of our knowledge, the highest lycopene titer achieved in *E. coli* is 3.52 g/l, using the strategies of increasing intracellular NADPH and ATP supplies and modulating expression of *dxs*, *idi*, and *crt* by screening RBS libraries [41]. For the best β-carotene production strain (3.2 g/l of β-carotene production), both the DXP and the MEV pathways were employed [98]. Because the introduction of the whole MEV pathway could result in an increased metabolic burden, the expression levels of the MEV pathway genes were balanced by screening TIGR libraries for zeaxanthin production in *E. coli* [43]. Then, the genes were dynamically controlled by using an IPP/FPP-responsive promoter to prevent accumulation of these toxic intermediates. As a result, 722.46 mg/l of zeaxanthin could be produced [43]. For the production of astaxanthin, a zeaxanthin derivative, the key membrane-associated enzyme (β-carotene ketolase) derived from microalgae was engineered and *in silico*-predicted target genes were amplified. After optimization of fermentation conditions, 432.82 mg/l of astaxanthin was produced [19].

Other than these natural carotenoids, nonnatural carotenoids have also been produced in *E. coli* by enzyme engineering (Figure I). In one study, C_{50}-astaxanthin could be synthesized in *E. coli* after directed evolution of FPP synthase, carotenoid synthase, and phytoene desaturase [99]. High-throughput colorimetric screening was possible because variants producing carotenoids with different carbon backbones exhibited different colors. Later, even larger C_{60}-carotenoids were produced based on the previously developed C_{50}-carotenoid pathway [100]. In this study, heterologous genes encoding elongases and cyclases with significant activities towards C_{50}-lycopene and C_{60}-flavuxanthin, respectively, were expressed together with the C_{50}-lycopene pathway in *E. coli.* This resulted in production of the novel carotenoids C_{60}-flavuxanthin, C_{60}-sarcinaxanthin, C_{60}-sarprenoxanthin, and C_{60}-decaprenoxanthin [100].

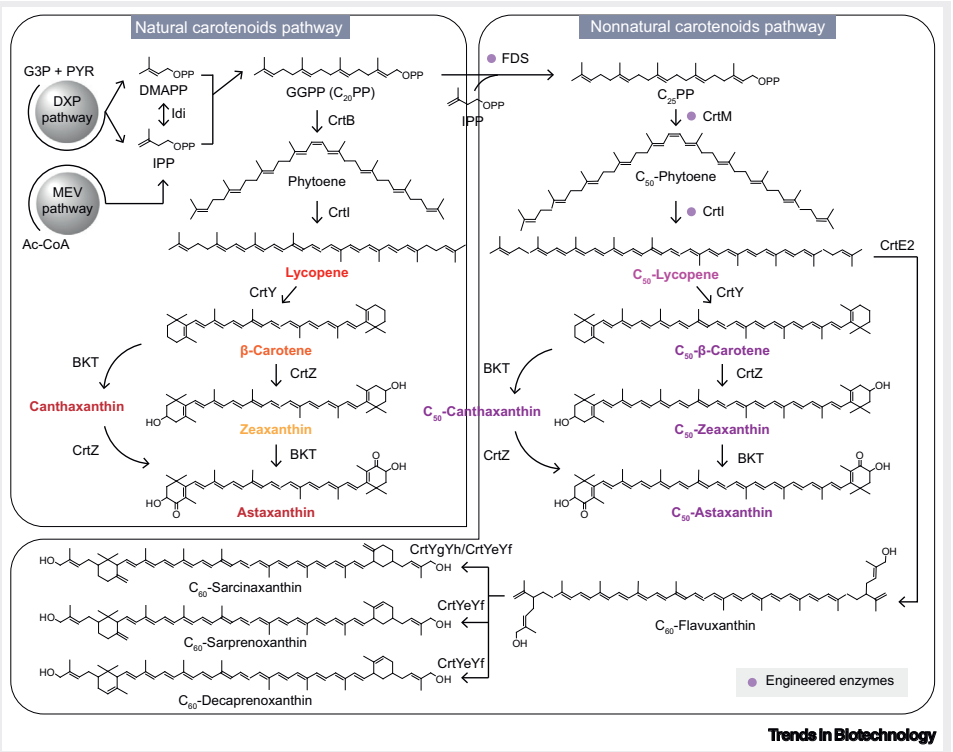

Figure I. Biosynthetic Pathways for the Production of Natural and Nonnatural Carotenoids in *Escherichia coli*. Metabolite abbreviations are as follows: Ac-CoA, acetyl coenzyme A; C_{25}PP, geranylfarnesyl diphosphate; DMAPP, dimethylallyl diphosphate; DXP, 1-deoxy-D-xylulose-5-phosphate; GGPP, geranylgeranyl diphosphate; G3P, glyceraldehyde 3-phosphate; IPP, isopentenyl diphosphate; MEV, mevalonate; OPP, diphosphate unit; PYR, pyruvate. Enzyme abbreviations are as follows: BKT, β-carotene ketolase; CrtB, phytoene synthase; CrtE2, lycopene elongase; CrtI, phytoene dehydrogenase; CrtM, C30 carotenoid backbone synthase; CrtY, lycopene cyclase; CrtYeYf, heterodimeric C50 ε/γ-cyclase; CrtYgYh, C50 γ-cyclase; CrtZ, β-carotene hydroxylase; FDS, farnesyl diphosphate synthase.

High-Throughput Screening Using Molecular Biosensors

For target compounds that display explicit colors, such as the red-colored lycopene and the purple-colored violacein, high-throughput colorimetric screening is feasible [14,69]. Otherwise, **molecular biosensors** can be applied to visualize colorless target compounds of different concentrations by converting them into observable signals (Figure 2N) [70]. Biosensors can be classified into two major categories: transcription factor-based and enzyme-based. In transcription factor-based biosensors, transcription factors bound to target compounds elicit differential expression of reporter genes [70,71]. Enzyme-based biosensors can directly convert target compounds into other surrogate metabolites that exhibit color or fluorescence [30]. Such biosensors can be combined with *in trans* target gene knockdown tools such as the synthetic sRNA technology [72,73] for system-wide screening of gene manipulation targets. For instance, a genome-scale sRNA library was employed together with a type III PKS (RppA)-based malonyl-CoA biosensor to identify gene knockdown targets that could enhance the malonyl-CoA pool [30]. Knockdown of the selected gene targets significantly augmented the production of polyketides (6-methylsalicylic acid and aloesone) and phenylpropanoids (resveratrol and naringenin). As exemplified above, high-throughput screening using biosensors has been useful in identifying target genes for manipulation. However, because only a handful of biosensors are

available, the development of new biosensors that can detect more diverse metabolites will be useful for further accelerating strain development.

Representative Examples of Natural Products Biosynthesized in *E. coli*

We have reviewed different strategies for the production of natural products in *E. coli.* In this section we highlight the production of three representative products in *E. coli* – taxol, resveratrol, and erythromycin. Important works on opioids and carotenoids are described in Boxes 2 and 3, respectively. The production of other natural products including MANT, violacein, and heme is shown in Box 4.

Taxol

Taxol is an anticancer drug isolated from the Pacific yew tree [74]. Taxadiene, an important taxol precursor, was first produced in *E. coli* with a titer of 1.3 mg/l [75]. Later, taxadiene production was increased to up to ~1 g/l by dividing the taxadiene biosynthetic pathway into two modules and balancing their expression levels [24]. As described earlier, a P450 and its partner CPR, both from *Taxus cuspidata*, were introduced for the oxygenation of taxadiene [25]. Balancing the expression ratio of the two enzymes, and N-terminal modification, resulted in 4.9-fold enhanced production of oxygenated taxanes (from 116 mg/l to 570 mg/l). Coculture was also employed using an engineered *E. coli* and a P450-expressing *Saccharomyces cerevisiae*, which resulted in 33 mg/l of oxygenated taxane production [47].

Resveratrol

Resveratrol (3,5,4'-trihydroxy-*trans*-stilbene), that is found in red wine, is a representative stilbenoid. Initial reports on resveratrol production in *E. coli* mostly relied on feeding precursors, and high-level resveratrol production (2.3 g/l) was achieved by feeding *p*-coumaric acid [76]. Later, *de novo* production of resveratrol from simple carbon sources was achieved either by integration of the resveratrol biosynthetic pathway into the chromosome (4.6 mg/l of resveratrol) [77] or by coculture of strains harboring segmented

Box 4. Production of Other Important Natural Products in *E. coli*

MANT is a grape flavoring agent which has recently been produced by engineered *E. coli* [18]. For the first step towards microbial production of MANT, ANT methyltransferase 1 (AAMT1) from *Zea mays* was introduced to convert ANT to MANT (Figure IA) [18]. To maximize the metabolic flux towards the shikimate pathway, the feedback inhibition on AroG by aromatic amino acids was removed by introducing feedback inhibition-resistant mutants generated by site-directed mutagenesis. Then, to enhance the flux towards PEP, *ppsA* was overexpressed while *pykF* and *pykA* were deleted. In addition, the supply of *S*-adenosyl-L-methionine (SAM), a cofactor for AAMT1, was enhanced by overexpression of $metA^{fbr}$, $cysE^{fbr}$, and *metK*. Because high-level accumulation of MANT was toxic to *E. coli*, a two-phase cultivation strategy was applied by addition of tributyrin (to separate MANT from the culture broth), leading to 4.47 g/l of MANT production [18].

Violacein, also derived from the shikimate pathway, is a purple compound with diverse pharmacological activities [101]. For its production in *E. coli*, *tktA* was first overexpressed to increase the metabolic flux towards erythrose 4-phosphate (E4P) [102]. To further increase the metabolic flux towards the shikimate pathway, *trpR* encoding a transcriptional repressor and *trpL* encoding a tryptophan operon leader peptide (which attenuates transcription of the tryptophan operon) were deleted, together with activation of the shikimate pathway genes, leading to 0.71 g/l of violacein production. Then, overexpression of *vioE*, encoding a bottleneck enzyme, led to production of 4.45 g/l of violacein and deoxyviolacein [103]. An *E. coli* strain more efficiently producing violacein and deoxyviolacein could be constructed without extensive metabolic engineering by employing an *E. coli* genome-scale sRNA library (Figure IB) [69]. After high-throughput colorimetric screening, a strain harboring an sRNA targeting *ytfR* (encoding sugar ABC transporter ATPase) led to production of 5.19 g/l of violacein and deoxyviolacein.

Heme is a porphyrin derivative which is a valuable pharmaceutical agent as well as an essential food supplement. Although intracellular production of heme in *E. coli* has been reported [104], its purification is a major challenge. To resolve this issue, *ccmABC* (encoding a heme exporter) was amplified to secrete heme into the extracellular medium. To produce its precursor 5-aminolevulinic acid, the C5 pathway was employed to produce heme from glucose instead of feeding with L-glutamate or succinate. Knocking out *ldhA*, *pta*, and *yfeX* (encoding a putative heme-degrading enzyme) was also performed. The final engineered strain produced 115.5 mg/l of heme from glucose, of which 73.4 mg/l of heme was secreted (Figure IC) [105].

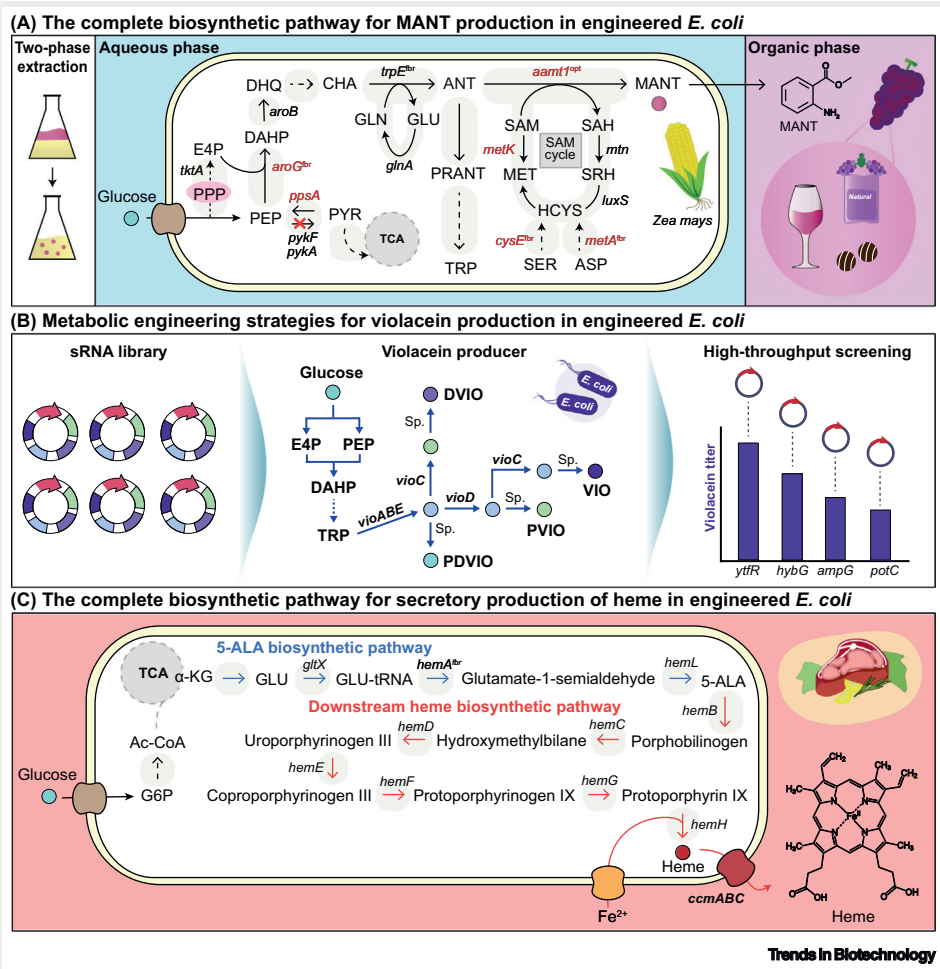

Figure I. Biosynthetic Pathways for the Production of Methyl Anthranilate (MANT), Violacein, and Heme in Engineered *Escherichia coli*. (A) The complete biosynthetic pathway for MANT production in engineered *E. coli*. (B) Metabolic engineering strategies for violacein production in engineered *E. coli*. A genome-scale synthetic small regulatory RNA (sRNA) library was employed for colorimetric screening of violacein overproducers. (C) The complete biosynthetic pathway for secretory heme production in engineered *E. coli*. The biosynthetic pathways illustrated in blue and red represent the C5 pathway for 5-ALA biosynthesis and the downstream heme biosynthetic pathway, respectively. Overexpressed genes are marked in red; red X represents gene deletion; solid arrows indicate single reactions; dashed arrows represent multiple reactions. Metabolite abbreviations are as follows: Ac-CoA, acetyl-coenzyme A; 5-ALA, 5-aminolevulinate; ANT, anthranilate; ASP, L-aspartate; CHA, chorismate; DAHP, 3-deoxy-D-arabinoheptulosonate 7-phosphate; DHQ, 3-dehydroquinate; DVIO, deoxyviolacein; GLN, L-glutamine; GLU, L-glutamate; GLU-tRNA, L-glutamyl-tRNA; G6P, glucose-6-phosphate; HCYS, L-homocysteine; α-KG, α-ketoglutarate; MET, L-methionine; PDVIO, prodeoxyviolacein; PEP, phosphoenolpyruvate; PRANT, *N*-(5-phosphoribosyl)-anthranilate; PYR, pyruvate; PVIO, proviolacein; SAH, *S*-adenosyl-L-homocysteine; SAM, *S*-adenosyl-L-methionine; SER, L-serine; SRH, *S*-ribosyl-L-homocysteine; TRP, L-tryptophan; VIO, violacein. Gene abbreviations are as follows: *aamt1*, ANT methyltransferase1; *aroB*, DHQ synthase; *cysE*^{fbr}, Ser acetyltransferase (feedback inhibition-resistant mutant); *glnA*, Gln synthetase; *gltX*, glutamyl-tRNA synthetase; *hemA*^{fbr}, feedback inhibition-resistant version of glutamyl-tRNA reductase; *hemB*, aminolevulinic acid dehydratase; *hemC*, porphobilinogen deaminase; *hemD*, uroporphyrinogen III synthase; *hemE*, uroporphyrinogen decarboxylase; *hemF*, coproporphyrinogen-III oxidase; *hemG*, protoporphyrinogen IX dehydrogenase; *hemH*, ferrochelatase; *hemL*, glutamate-1-semialdehyde 2,1-aminomutase; *luxS*, *S*-ribosylhomocysteine lyase; *metA*^{fbr}, homoserine *O*-succinyltransferase (feedback inhibition-resistant mutant); *metK*, MET adenosyltransferase; *mtn*, 5′-methylthioadenosine/*S*-adenosylhomocysteine nucleosidase; *ppsA*, PEP synthetase; *pykF*, PYR kinase I; *pykA*, PYR kinase II; *tktA*, transketolase I; *trpE*^{fbr}, ANT synthase component I (feedback inhibition-resistant mutant). Other abbreviations: PPP, pentose phosphate pathway; Sp., spontaneous reaction; TCA, tricarboxylic acid cycle.

pathways (22.6 mg/l of resveratrol) [78]. To further enhance resveratrol production, malonyl-CoA supply was increased by introducing a heterologous malonate assimilation pathway and repressing fatty acid biosynthetic genes, which resulted in a 10.2-fold higher titer (from 18.5 mg/l to 188.1 mg/l) [31]. In addition, the expression level of the rate-controlling tyrosine ammonia lyase (TAL) was enhanced by reducing the secondary structure of the 5′ region of the mRNA, and this resulted in a further increase in resveratrol titer (304.5 mg/l). In another study, resveratrol glucosides were produced to take advantage of their enhanced solubility and bioavailability. The biosynthetic pathway for resveratrol glucosides was divided into two different *E. coli* strains to alleviate the metabolic burden. Coculture of these two strains yielded 92 mg/l of resveratrol glucosides [79].

Erythromycin

Erythromycin, a complex polyketide macrolide, is natively produced from the soil bacterium *Saccharopolyspora erythraea*. This class of antibiotics is widely applied against Gram-positive bacterial pathogens, and erythromycin has several variants, of which erythromycin A is the most active compound [80]. Condensation of six (2*S*)-methylmalonyl-CoA and one propionyl-CoA molecules catalyzed by deoxyerythronolide B synthase (DEBS) generates 6-deoxyerythronolide B (6-dEB), a macrolactone precursor of erythromycin. Additional tailoring reactions followed by the addition of two deoxysugar moieties led to the complete production of erythromycin. The first successful production of 6-dEB from propionate in *E. coli* was achieved by introducing *Bacillus subtilis sfp* (encoding phosphopantetheinyl transferase), optimizing metabolic flux, and expressing large 6-dEB BGCs [81]. To produce 6-dEB directly from glucose, (2*S*)-methylmalonyl-CoA and propionyl-CoA pools were increased by knocking out *ygfH* (encoding propionyl-CoA:succinate-CoA-transferase) and activating short-chain fatty acid uptake [82]. This resulted in the production of 134.1 mg/l of 6-dEB in *E. coli.* The complete biosynthesis of erythromycin A in *E. coli* was achieved by introducing the reconstructed erythromycin BGC comprising 26 genes (>50 kb), yielding 10 mg/l of erythromycin A in flask culture [27]. Various erythromycin glycoside analogs were also produced in *E. coli* in an effort to increase antibiotic bioactivity [83].

Concluding Remarks and Future Perspectives

E. coli is an attractive microbial host for the heterologous production of natural products. Over the past two decades, various metabolic engineering tools and strategies have been developed to allow the production of diverse natural products in *E. coli*. However, one notable obstacle to producing diverse natural products is that the function of many enzymes is not known and many pathways lack enzyme annotation (see Outstanding Questions). In this regard, recent advances in machine learning are expected to aid in the generation of enzyme mutants with desired activities and characteristics [84]. Although there have been recent reports on engineering modular PKS assembly lines for the production of new compounds [85], the *de novo* creation of proteins with predictable functions remains challenging. In this regard, a breakthrough in the development of new algorithms for *de novo* creation of enzymes will greatly boost the construction of new biosynthetic reactions and pathways [86]. Metagenomes can also be excellent resources for the discovery of novel enzymes or pharmaceutically active compounds [87]. Samples taken from a wide range of environments, such as soil, river, sea, and deep sea, as well as human and animal microbiomes, can be subjected to metagenome sequencing. The development of more efficient ALE methods is also important. In particular, more explicit screening methods with increased throughput should be developed to enhance the accuracy and speed of identifying strains showing high performance and desired traits. Systems metabolic engineering will continue to play a

Outstanding Questions

How can we increase the titers of natural products to the tens of grams per liter level?

What tools and strategies do we need to develop to identify unknown reaction steps and associated enzymes ('missing enzymatic steps') in the biosynthetic pathways for natural products, to identify (including undiscovered) enzymes capable of catalyzing desired reactions, and to elucidate hypothetical enzymes with unknown functions?

How can artificial intelligence (AI) and algorithms be trained for efficient natural product biosynthesis from microbial hosts?

How can novel enzymes or natural products be efficiently mined from diverse metagenomes?

How can we deal with ethical and safety issues that might arise from the production of natural compounds that are harmful to humans, our society, and the environment?

What will be the most effective way to accelerate the ALE process? How can we minimize false-positive hits while rapidly screening for beneficial mutants?

How can we overcome unfavorable perception of the public to natural products produced by genetically engineered organisms?

central role in resolving these issues to construct high-performance *E. coli* strains and to achieve industrial-level production of natural products.

Acknowledgments

This work was supported by the Technology Development Program to Solve Climate Changes on Systems Metabolic Engineering for Biorefineries (NRF-2012M1A2A2026556 and NRF-2012M1A2A2026557) from the Ministry of Science and ICT through the National Research Foundation (NRF) of Korea. S.Y.L. and D.Y. were also supported by the Novo Nordisk Foundation (grant NNF16OC0021746).

References

1. Butler, M.S. (2004) The role of natural product chemistry in drug discovery. *J. Nat. Prod.* 67, 2141–2153
2. David, B. *et al.* (2015) The pharmaceutical industry and natural products: historical status and new trends. *Phytochem. Rev.* 14, 299–315
3. Yun, U.W. *et al.* (2012) Plant natural products: history, limitations and the potential of cambial meristematic cells. *Biotechnol. Genet. Eng. Rev.* 28, 47–59
4. Li, C.J. and Trost, B.M. (2008) Green chemistry for chemical synthesis. *Proc. Natl. Acad. Sci. U. S. A.* 105, 13197–13202
5. Park, S.Y. *et al.* (2018) Metabolic engineering of microorganisms for the production of natural compounds. *Adv. Biosyst.* 2, 1700190
6. Lee, S.Y. (1996) High cell-density culture of *Escherichia coli*. *Trends Biotechnol.* 14, 98–105
7. Kim, W.J. *et al.* (2017) Current state and applications of microbial genome-scale metabolic models. *Curr. Opin. Syst. Biol.* 2, 10–18
8. Lee, S.Y. (1996) High cell-density culture of *Escherichia coli*. *Trends Biotechnol.* 14, 98–105
9. Qian, S. *et al.* (2019) Engineering *Escherichia coli* as a platform for the *in vivo* synthesis of prenylated aromatics. *Biotechnol. Bioeng.* 116, 1116–1127
10. Edgar, S. *et al.* (2016) Engineering of taxadiene synthase for improved selectivity and yield of a key taxol biosynthetic intermediate. *ACS Synth. Biol.* 6, 201–205
11. Mendez-Perez, D. *et al.* (2017) Production of jet fuel precursor monoterpenoids from engineered *Escherichia coli*. *Biotechnol. Bioeng.* 114, 1703–1712
12. Vavricka, C.J. *et al.* (2019) Mechanism-based tuning of insect 3, 4-dihydroxyphenylacetaldehyde synthase for synthetic bioproduction of benzylisoquinoline alkaloids. *Nat. Commun.* 10, 2015
13. Xiong, D. *et al.* (2017) Improving key enzyme activity in phenylpropanoid pathway with a designed biosensor. *Metab. Eng.* 40, 115–123
14. Lv, X. *et al.* (2016) Combinatorial pathway optimization in *Escherichia coli* by directed co-evolution of rate-limiting enzymes and modular pathway engineering. *Biotechnol. Bioeng.* 113, 2661–2669
15. Pérez-Arellano, I. *et al.* (2010) Molecular mechanisms modulating glutamate kinase activity. Identification of the proline feedback inhibitor binding site. *J. Mol. Biol.* 404, 890–901
16. Shomar, H. *et al.* (2018) Metabolic engineering of a carbapenem antibiotic synthesis pathway in *Escherichia coli*. *Nat. Chem. Biol.* 14, 794–800
17. Du, J. *et al.* (2018) Metabolic engineering of *Escherichia coli* for the production of indirubin from glucose. *J. Biotechnol.* 267, 19–28
18. Luo, Z.W. *et al.* (2019) Microbial production of methyl anthranilate, a grape flavor compound. *Proc. Natl. Acad. Sci. U. S. A.* 116, 10749–10756
19. Park, S.Y. *et al.* (2018) Metabolic engineering of *Escherichia coli* for high-level astaxanthin production with high productivity. *Metab. Eng.* 49, 105–115
20. Brignac-Huber, L.M. *et al.* (2016) Cytochrome P450 organization and function are modulated by endoplasmic reticulum phospholipid heterogeneity. *Drug Metab. Disposition* 44, 1859–1866
21. Chang, M.C. *et al.* (2007) Engineering *Escherichia coli* for production of functionalized terpenoids using plant P450s. *Nat. Chem. Biol.* 3, 274–277
22. Paddon, C.J. *et al.* (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* 496, 528–532
23. Paddon, C.J. and Keasling, J.D. (2014) Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat. Rev. Microbiol.* 12, 355–367
24. Ajikumar, P.K. *et al.* (2010) Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. *Science* 330, 70–74
25. Biggs, B.W. *et al.* (2016) Overcoming heterologous protein interdependency to optimize P450-mediated taxol precursor synthesis in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 113, 3209–3214
26. Mutka, S.C. *et al.* (2006) Heterologous production of epothilone C and D in *Escherichia coli*. *Biochemistry* 45, 1321–1330
27. Zhang, H. *et al.* (2010) Complete biosynthesis of erythromycin A and designed analogs using *E. coli* as a heterologous host. *Chem. Biol.* 17, 1232–1240
28. Zhao, Y. *et al.* (2018) Combinatorial optimization of resveratrol production in engineered *E. coli*. *J. Agric. Food Chem.* 66, 13444–13453
29. Zha, W. *et al.* (2009) Improving cellular malonyl-CoA level in *Escherichia coli* via metabolic engineering. *Metab. Eng.* 11, 192–198
30. Yang, D. *et al.* (2018) Repurposing type III polyketide synthase as a malonyl-CoA biosensor for metabolic engineering in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 115, 9835–9844
31. Wu, J. *et al.* (2017) Efficient *de novo* synthesis of resveratrol by metabolically engineered *Escherichia coli*. *J. Ind. Microbiol. Biotechnol.* 44, 1083–1095
32. Choi, H.S. *et al.* (2010) *In silico* identification of gene amplification targets for improvement of lycopene production. *Appl. Environ. Microbiol.* 76, 3097–3105
33. Farmer, W.R. and Liao, J.C. (2001) Precursor balancing for metabolic engineering of lycopene production in *Escherichia coli*. *Biotechnol. Prog.* 17, 57–61
34. Jung, J. *et al.* (2016) Precise precursor rebalancing for isoprenoids production by fine control of *gapA* expression in *Escherichia coli*. *Metab. Eng.* 38, 401–408
35. Yang, C. *et al.* (2016) Synergy between methylerythritol phosphate pathway and mevalonate pathway for isoprene production in *Escherichia coli*. *Metab. Eng.* 37, 79–91
36. Nakagawa, A. *et al.* (2011) A bacterial platform for fermentative production of plant alkaloids. *Nat. Commun.* 2, 326
37. Urlacher, V.B. and Girhard, M. (2019) Cytochrome P450 monooxygenases in biotechnology and synthetic biology. *Trends Biotechnol.* 37, 882–897
38. Zhao, S. *et al.* (2015) Improvement of catechin production in *Escherichia coli* through combinatorial metabolic engineering. *Metab. Eng.* 28, 43–53
39. Tao, S. *et al.* (2018) Regulation of ATP levels in *Escherichia coli* using CRISPR interference for enhanced pinocembrin production. *Microb. Cell Fact.* 17, 147
40. Nakagawa, A. *et al.* (2016) Total biosynthesis of opiates by stepwise fermentation using engineered *Escherichia coli*. *Nat. Commun.* 7, 10390

41. Sun, T. *et al.* (2014) Production of lycopene by metabolically-engineered *Escherichia coli*. *Biotechnol. Lett.* 36, 1515–1522
42. Pfleger, B.F. *et al.* (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat. Biotechnol.* 24, 1027–1032
43. Shen, H.-J. *et al.* (2016) Dynamic control of the mevalonate pathway expression for improved zeaxanthin production in *Escherichia coli* and comparative proteome analysis. *Metab. Eng.* 38, 180–190
44. Dueber, J.E. *et al.* (2009) Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* 27, 753–759
45. Conrado, R.J. *et al.* (2011) DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. *Nucleic Acids Res.* 40, 1879–1889
46. Gao, C. *et al.* (2018) Engineering *Escherichia coli* for malate production by integrating modular pathway characterization with CRISPRi-guided multiplexed metabolic tuning. *Biotechnol. Bioeng.* 115, 661–672
47. Zhou, K. *et al.* (2015) Distributing a metabolic pathway among a microbial consortium enhances production of natural products. *Nat. Biotechnol.* 33, 377–383
48. Zhang, C. *et al.* (2018) Multidimensional heuristic process for high-yield production of astaxanthin and fragrance molecules in *Escherichia coli*. *Nat. Commun.* 9, 1858
49. Jones, J.A. *et al.* (2017) Complete biosynthesis of anthocyanins using *E. coli* polycultures. *mBio* 8 e00621-17
50. Li, Z. *et al.* (2019) Balancing the non-linear rosmarinic acid biosynthetic pathway by modular co-culture engineering. *Metab. Eng.* 54, 1–11
51. Ahmadi, M.K. *et al.* (2016) *E. coli* metabolic engineering for gram scale production of a plant-based anti-inflammatory agent. *Metab. Eng.* 38, 382–388
52. Nakagawa, A. *et al.* (2014) (*R,S*)-Tetrahydropapaveroline production by stepwise fermentation using engineered *Escherichia coli*. *Sci. Rep.* 4, 6695
53. Lee, S.Y. *et al.* (2019) A comprehensive metabolic map for production of bio-based chemicals. *Nat. Catal.* 2, 18–33
54. Choi, K.R. *et al.* (2019) Systems metabolic engineering strategies: integrating systems and synthetic biology with metabolic engineering. *Trends Biotechnol.* 37, 817–837
55. Monk, J.M. *et al.* (2017) iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.* 35, 904–908
56. Gu, C. *et al.* (2019) Current status and applications of genome-scale metabolic models. *Genome Biol.* 20, 121
57. O'Brien, E.J. *et al.* (2015) Using genome-scale models to predict biological capabilities. *Cell* 161, 971–987
58. Xu, P. *et al.* (2011) Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. *Metab. Eng.* 13, 578–587
59. Cummings, M. *et al.* (2019) Assembling a plug-and-play production line for combinatorial biosynthesis of aromatic polyketides in *Escherichia coli*. *PLoS Biol.* 17, e3000347
60. Yang, J.H. *et al.* (2019) A white-box machine learning approach for revealing antibiotic mechanisms of action. *Cell* 177, 1649–1661
61. Jervis, A.J. *et al.* (2018) Machine learning of designed translational control allows predictive pathway optimization in *Escherichia coli*. *ACS Synth. Biol.* 8, 127–136
62. Guzmán, G.I. *et al.* (2019) Enzyme promiscuity shapes adaptation to novel growth substrates. *Mol. Syst. Biol.* 15, e8462
63. Choe, D. *et al.* (2019) Adaptive laboratory evolution of a genome-reduced *Escherichia coli*. *Nat. Commun.* 10, 935
64. Niu, F.-X. *et al.* (2018) Enhancing production of pinene in *Escherichia coli* by using a combination of tolerance, evolution and modular co-culture engineering. *Front. Microbiol.* 9, 1623
65. Crown, S.B. *et al.* (2015) Integrated ^{13}C-metabolic flux analysis of 14 parallel labeling experiments in *Escherichia coli*. *Metab. Eng.* 28, 151–158
66. Chen, X. *et al.* (2011) Synergy between ^{13}C-metabolic flux analysis and flux balance analysis for understanding metabolic adaption to anaerobiosis in *E. coli*. *Metab. Eng.* 13, 38–48
67. Wada, K. *et al.* (2017) ^{13}C-metabolic flux analysis for mevalonate-producing strain of *Escherichia coli*. *J. Biosci. Bioeng.* 123, 177–182
68. Fang, M.-Y. *et al.* (2015) High crude violacein production from glucose by *Escherichia coli* engineered with interactive control of tryptophan pathway and violacein biosynthetic pathway. *Microb. Cell Fact.* 14, 8
69. Yang, D. *et al.* (2019) Expanded synthetic small regulatory RNA expression platforms for rapid and multiplex gene expression knockdown. *Metab. Eng.* 54, 180–190
70. Meyer, A.J. *et al.* (2019) *Escherichia coli* 'marionette' strains with 12 highly optimized small-molecule sensors. *Nat. Chem. Biol.* 15, 196–204
71. Kim, S.K. *et al.* (2018) A genetically encoded biosensor for monitoring isoprene production in engineered *Escherichia coli*. *ACS Synth. Biol.* 7, 2379–2390
72. Na, D. *et al.* (2013) Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. *Nat. Biotechnol.* 31, 170–174
73. Noh, M. *et al.* (2017) Gene expression knockdown by modulating synthetic small RNA expression in *Escherichia coli*. *Cell Syst.* 5, 418–426
74. Wani, M.C. *et al.* (1971) Plant antitumor agents. VI. Isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *J. Am. Chem. Soc.* 93, 2325–2327
75. Huang, Q. *et al.* (2001) Engineering *Escherichia coli* for the synthesis of taxadiene, a key intermediate in the biosynthesis of taxol. *Biorg. Med. Chem.* 9, 2237–2242
76. Lim, C.G. *et al.* (2011) High-yield resveratrol production in engineered *Escherichia coli*. *Appl. Environ. Microbiol.* 77, 3451–3460
77. Liu, X. *et al.* (2016) *De novo* biosynthesis of resveratrol by site-specific integration of heterologous genes in *Escherichia coli*. *FEMS Microbiol. Lett.* 363
78. Camacho-Zaragoza, J.M. *et al.* (2016) Engineering of a microbial coculture of *Escherichia coli* strains for the biosynthesis of resveratrol. *Microb. Cell Fact.* 15, 163
79. Thuan, N.H. *et al.* (2018) *Escherichia coli* modular coculture system for resveratrol glucosides production. *World J. Microbiol. Biotechnol.* 34, 75
80. Wu, H. *et al.* (2016) Inactivation of SACE_3446, a TetR family transcriptional regulator, stimulates erythromycin production in *Saccharopolyspora erythraea*. *Synth. Syst. Biotechnol.* 1, 39–46
81. Pfeifer, B.A. *et al.* (2001) Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* 291, 1790–1792
82. Boghigian, B.A. *et al.* (2011) Multi-factorial engineering of heterologous polyketide production in *Escherichia coli* reveals complex pathway interactions. *Biotechnol. Bioeng.* 108, 1360–1371
83. Fang, L. *et al.* (2018) Broadened glycosylation patterning of heterologously produced erythromycin. *Biotechnol. Bioeng.* 115, 2771–2777
84. Wu, Z. *et al.* (2019) Machine learning-assisted directed protein evolution with combinatorial libraries. *Proc. Natl. Acad. Sci. U. S. A.* 116, 8852–8858
85. Klaus, M. and Grininger, M. (2018) Engineering strategies for rational polyketide synthase design. *Nat. Prod. Rep.* 35, 1070–1081
86. Langan, R.A. *et al.* (2019) *De novo* design of bioactive protein switches. *Nature* 572, 205–210
87. Zipperer, A. *et al.* (2016) Human commensals producing a novel antibiotic impair pathogen colonization. *Nature* 535, 511–516
88. Ward, V.C.A. *et al.* (2018) Metabolic engineering of *Escherichia coli* for the production of isoprenoids. *FEMS Microbiol. Lett.* 365, fny079
89. Johnson, A.O. *et al.* (2017) Design and application of genetically-encoded malonyl-CoA biosensors for metabolic engineering of microbial cell factories. *Metab. Eng.* 44, 253–264
90. Wang, Y. and Pfeifer, B.A. (2008) 6-Deoxyerythronolide B production through chromosomal localization of the

deoxyerythronolide B synthase genes in *E. coli*. *Metab. Eng.* 10, 33–38
91. Zhang, H. *et al.* (2012) Improved *E. coli* erythromycin a production through the application of metabolic and bioprocess engineering. *Biotechnol. Prog.* 28, 292–296
92. Gao, X. *et al.* (2010) Engineered polyketide biosynthesis and biocatalysis in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 88, 1233–1242
93. Strieker, M. *et al.* (2010) Nonribosomal peptide synthetases: structures and dynamics. *Curr. Opin. Struct. Biol.* 20, 234–240
94. Narcross, L. *et al.* (2016) Microbial factories for the production of benzylisoquinoline alkaloids. *Trends Biotechnol.* 34, 228–241
95. Rinner, U. and Hudlicky, T. (2012) Synthesis of morphine alkaloids and derivatives. *Top. Curr. Chem.* 309, 33–66
96. Galanie, S. *et al.* (2015) Complete biosynthesis of opioids in yeast. *Science* 349, 1095–1100
97. Stahl, W. and Sies, H. (2003) Antioxidant activity of carotenoids. *Mol. Aspects Med.* 24, 345–351
98. Yang, J. and Guo, L. (2014) Biosynthesis of β-carotene in engineered *E. coli* using the MEP and MVA pathways. *Microb. Cell Fact.* 13, 160
99. Furubayashi, M. *et al.* (2015) A highly selective biosynthetic pathway to non-natural C_{50} carotenoids assembled from moderately selective enzymes. *Nat. Commun.* 6, 7534
100. Li, L. *et al.* (2019) Genetically engineered biosynthetic pathways for nonnatural C_{60} carotenoids using C_5-elongases and C_{50}-cyclases in *Escherichia coli*. *Sci. Rep.* 9, 2982
101. Duran, N. *et al.* (2007) Violacein: properties and biological activities. *Biotechnol. Appl. Biochem.* 48, 127–133
102. Rodrigues, A.L. *et al.* (2013) Systems metabolic engineering of *Escherichia coli* for production of the antitumor drugs violacein and deoxyviolacein. *Metab. Eng.* 20, 29–41
103. Zhou, Y. *et al.* (2018) Enhanced production of crude violacein from glucose in *Escherichia coli* by overexpression of rate-limiting key enzyme(s) involved in violacein biosynthesis. *Appl. Biochem. Biotechnol.* 186, 909–916
104. Kwon, S.J. *et al.* (2003) High-level production of porphyrins in metabolically engineered *Escherichia coli*: systematic extension of a pathway assembled from overexpressed genes involved in heme biosynthesis. *Appl. Environ. Microbiol.* 69, 4875–4883
105. Zhao, X.R. *et al.* (2018) Metabolic engineering of *Escherichia coli* for secretory production of free haem. *Nat. Catal.* 1, 720–728
106. Coussement, P. *et al.* (2017) Direct combinatorial pathway optimization. *ACS Synth. Biol.* 6, 224–232
107. Shukal, S. *et al.* (2019) Systematic engineering for high-yield production of viridiflorol and amorphadiene in auxotrophic *Escherichia coli*. *Metab. Eng.* 55, 170–178
108. Liu, W. *et al.* (2016) Engineering *Escherichia coli* for high-yield geraniol production with biotransformation of geranyl acetate to geraniol under fed-batch culture. *Biotechnol. Biofuels* 9, 58
109. Wang, W. *et al.* (2019) Construction of *Escherichia coli* cell factories for crocin biosynthesis. *Microb. Cell Fact.* 18, 120
110. Stevens, D.C. *et al.* (2013) Alternative sigma factor over-expression enables heterologous expression of a type II polyketide biosynthetic pathway in *Escherichia coli*. *PLoS One* 8, e64858
111. Tan, Z. *et al.* (2018) Synthetic pathway for the production of olivetolic acid in *Escherichia coli*. *ACS Synth. Biol.* 7, 1886–1896
112. Wu, J. *et al.* (2015) Enhancing flavonoid production by systematically tuning the central metabolic pathways based on a CRISPR interference system in *Escherichia coli*. *Sci. Rep.* 5, 13477
113. Zhao, L.Q. *et al.* (2018) Efficient biotransformation of isoeugenol to vanillin in recombinant strains of *Escherichia coli* by using engineered isoeugenol monooxygenase and sol-gel chitosan membrane. *Process Biochem.* 71, 76–81
114. Ni, J. *et al.* (2015) Mimicking a natural pathway for *de novo* biosynthesis: natural vanillin production from accessible carbon sources. *Sci. Rep.* 5, 13670
115. Jones, J.A. *et al.* (2016) Experimental and computational optimization of an *Escherichia coli* co-culture for the efficient production of flavonoids. *Metab. Eng.* 35, 55–63
116. Chen, Z. *et al.* (2017) Metabolic engineering of *Escherichia coli* for microbial synthesis of monolignols. *Metab. Eng.* 39, 102–109
117. De Bruyn, F. *et al.* (2015) Metabolic engineering of *Escherichia coli* into a versatile glycosylation platform: production of bio-active quercetin glycosides. *Microb. Cell Fact.* 14, 138
118. Wu, J. *et al.* (2016) Efficient biosynthesis of (2S)-pinocembrin from D-glucose by integrating engineering central metabolic pathways with a pH-shift control strategy. *Bioresour. Technol.* 218, 999–1007
119. Fordjour, E. *et al.* (2019) Metabolic engineering of *Escherichia coli* BL21 (DE3) for *de novo* production of L-DOPA from D-glucose. *Microb. Cell Fact.* 18, 74
120. Ahmadi, M.K. *et al.* (2015) Total biosynthesis and diverse applications of the nonribosomal peptide-polyketide siderophore yersiniabactin. *Appl. Environ. Microbiol.* 81, 5290–5298
121. Watanabe, K. *et al.* (2006) Total biosynthesis of antitumor nonribosomal peptides in *Escherichia coli*. *Nat. Chem. Biol.* 2, 423–428