Discovery and Engineering of Cytochrome P450s for Terpenoid Biosynthesis

Han Xiao,1,4,* Yue Zhang,2,3,4 and Meng Wang2,3,*

Terpenoids represent 60% of known natural products, including many drugs and drug candidates, and their biosynthesis is attracting great interest. However, the unknown cytochrome P450s (CYPs) in terpenoid biosynthetic pathways make the heterologous production of related terpenoids impossible, while the slow kinetics of some known CYPs greatly limit the efficiency of terpenoid biosynthesis. Thus, there is a compelling need to discover and engineer CYPs for terpenoid biosynthesis to fully realize their great potential for industrial application. This review article summarizes the current state of CYP discovery and engineering in terpenoid biosynthesis, focusing on recent synthetic biology approaches toward prototyping CYPs in heterologous hosts. We also propose several strategies for further accelerating CYP discovery and engineering.

Terpenoids and Their Biosynthesis
Terpenoids represent the most diverse and widespread family of natural products. They not only play vital roles in complex biochemical processes such as stress response and cytophylaxis [1] but also have various bioactivities and extensive applications in the food, cosmetics, and pharmaceutical industries [2,3]. Some terpenoids in particular may function as drugs to treat various human diseases. For example, the sesquiterpenoid artemisinin is used as an antimalarial drug [4], and the diterpenoid Taxol was developed to be an important anticancer chemotherapy drug [5].

Terpenoid biosynthesis can be divided into four stages: building block formation, condensation, core structure assembly, and post-modification (Figure 1). All terpenoids are derived from the universal C5 building blocks isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). In the first stage, IPP and DMAPP are synthesized via the mevalonate (MVA) pathway or a mevalonate-independent pathway (see Figure 1 for details) [6,7]. In the second stage, condensation of one molecule of DMAPP with one, two, or three molecules of IPP leads to geranyl pyrophosphate (GPP), farnesyl pyrophosphate (FPP), or geranylgeranyl pyrophosphate (GGPP), the direct precursor of monoterpenoids, sesquiterpenoids, or diterpenoids, respectively. Alternatively, condensation of two molecules of FPP generates long-chain C30 triterpenoids. In the third stage, these direct precursors are further assembled into the core structures of different terpenoids via terpene synthases (see Glossary) [8]. Finally, terpenoid core structures undergo a series of post-modifications in which cytochrome P450s (CYPs) are usually heavily involved, yielding final terpenoid products with wide chemical diversity. Until now, CYPs as a superfamily of heme monooxygenases have been found to modify more than 97% of all terpenoids, leading to structural diversity and corresponding bioactivity [9]. These CYPs mainly carry out selective oxidations, providing chemical functionality for subsequent acetylation, esterification, alkylation, and other reactions. In addition, a growing number of CYPs have been reported to exhibit promiscuity in converting different substrates [10] or yielding different products [11].

Highlights
CYP discovery and engineering are crucial for terpenoid biosynthesis, which has recently attracted great academic and industrial interest.

Major obstacles to traditional CYP discovery methods are the inaccessibility of genetic tools in native hosts and the low activity of CYPs.

Synthetic biology approaches based on multi-omics studies, heterologous CYP expression, and metabolic engineering of chassis cells have revolutionized CYP discovery platforms and significantly accelerated their discovery rate.

Harnessing the great potential of genome engineering and robotic automation will further increase the likelihood of discovering or engineering CYP with desired activity as well as enhance the production of terpenoids with commercial value.

1State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic & Developmental Sciences, and Laboratory of Molecular Biochemical Engineering, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 800 Dong-chuan Road, Shanghai, 200240, China
2Key Laboratory of Systems Microbial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China
3Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China
4Co-first author with equal contribution

*Correspondence: smallihan@sjtu.edu.cn (H. Xiao) and wangmeng@tib.cas.cn (M. Wang).
Bottlenecks for Efficient Biosynthesis of Terpenoids: Cryptic CYPs and Low Activity of CYPs

Understanding the post-modification steps in terpenoid biosynthesis is of particular importance for their efficient production, because such selective reactions performed by CYPs are extremely hard to achieve via conventional chemical routes. In contrast to the ample research on the first three stages of terpenoid biosynthesis [12,13], the detailed elucidation of post-modification steps of terpenoid biosynthesis is insufficient. To date, biosynthetic pathways of many bioactive terpenoids, such as paclitaxel and ganoderic acids, remain elusive due to the vast proportion of biosynthetic steps that may involve uncharacterized CYPs [14,15]. In addition, except for Bacillus megaterium-derived CYP102A1 (P450BM3) and Pseudomonas putida-derived CYP101A1 (P450cam), most characterized CYPs have low $k_{cat}$ values in the range of ~1–300 min$^{-1}$, which are too slow to have practical applications in industry [16]. Furthermore, most CYPs require an additional protein partner, cytochrome P450 reductase (CPR), to obtain electrons from NAD(P)H for oxygen activation. The insufficient electron transfer from CPR to CYP, as usually observed in many CYP-catalyzed processes, may lead to diminished catalytic rates [17]. Moreover, CYP catalytic activity can be greatly affected by the uncoupling between product formation and NAD(P)H consumption. The subsequent leakage of NAD(P)H may result in production of reactive oxygen species that, in turn, can destabilize CYPs (Box 1) [16].

Cryptic CYPs and the low activity of CYPs can significantly hinder the efficient biosynthesis of terpenoids. In the following sections, we summarize strategies for CYP discovery and engineering for terpenoid biosynthesis. First, traditional CYP discovery approaches, including gene deletion, gene silencing, and in vitro enzymatic assay, as well as their drawbacks, are introduced. While traditional approaches are still commonly used and sometimes critical in determining CYP functions in native hosts, their limitations make them insufficient to explore CYPs from genetically intractable species. With guidance from multi-omics information, prototyping CYP candidates in an engineered genetically tractable host using synthetic biology approaches has provided a clear systemic workflow to address the key issues in CYP discovery. Next, we review CYP engineering studies related to terpenoid biosynthesis. Finally, we propose perspectives on how to further accelerate CYP discovery and engineering.

Traditional Approaches for CYP Discovery

Traditional approaches for CYP discovery include gene deletion, gene silencing, and in vitro enzymatic assays (Box 2; Table 1; Figure 2). Successful examples are commonly found in microorganisms with well-understood genetic backgrounds and easy access to genetic toolkits.

Gene Deletion

Gene deletion is a method commonly used to deduce CYP functions in native hosts with well-established genetic manipulation tools, such as Aspergillus and Penicillium. For example, fumagillin is a meroterpenoid that is synthesized via a polyketide-terpenoid hybrid pathway (fma biosynthetic gene cluster) from Aspergillus fumigatus. Deletion of the CYP gene Afi510 in the fma cluster abolished fumagillin production but increased the accumulation of β-trans-b ergamotene, indicating that Afi510 is a β-trans-b ergamotene oxidase in fumagillin biosynthesis [18]. In another study, deletion of two CYP genes, vrtE and vrtK, was performed to investigate their functions in the meroterpenoid viridicatumtoxin biosynthesis in Penicillium aethiopicum. Both vrtE and vrtK knockout mutants were not able to produce viridicatumtoxin but exhibited accumulation of the new intermediates naphthaceninedione and...
Figure 1. Terpenoid Biosynthesis Pathways. Terpenoid biosynthesis consists of four stages: building block formation, condensation, core structure assembly, and functionalization.

(Figure legend continued on the bottom of the next page.)
Box 1. Uncoupling in CYPs and Possible Solutions
CYP reactions take place at the heme domain, where substrate binds and electron transfer occurs. CPR-mediated insufficient electron transfer from NADPH to the heme domain, also known as an uncoupling event, leads to byproduct formation without substrate oxidation. The subsequent leakage of reducing equivalents generates reactive oxygen species that can induce cellular stress and result in decreased cell growth, protein expression, and terpenoid production. One way to solve the problem is to create self-sufficient CYPs by fusing CYPs with their partner CPRs. Another possible strategy is to colocalize CYPs and CPRs in the same subcellular compartment to maximize their electron transfer efficiency.

Box 2. CYP Discovery Approaches
CYP discovery is a prerequisite to elucidate terpenoid biosynthetic mechanisms and to efficiently synthesize terpenoids. Although many approaches for CYP discovery have been adopted in the past, synthetic biology approaches circumvent some limitations of traditional methods, leading to a higher success rate of CYP discovery.
1. Traditional methods for CYP discovery involve gene deletion and gene silencing in native hosts. After this process, metabolites from both mutant and wild type are carefully compared, and detailed functions of target CYPs are deduced accordingly.
2. A few CYPs can be overexpressed and purified while their substrates are easily obtained, and then in vitro enzymatic assay is a good way to confirm their activities.
3. Recently developed synthetic biology approaches, including multi-level omics analysis, heterologous CYP expression in genetically tractable hosts, and metabolic engineering of chassis cells, greatly accelerate CYP discovery. Multi-level omics analysis narrows down the CYP candidate pool for a particular pathway. Heterologous CYP expression provides key CYP expression platforms and access to some previously unobtainable substrates. Advances in metabolic engineering of model microorganisms, including E. coli and S. cerevisiae, establish platforms for the discovery and characterization of CYPs and by extension enable the de novo biosynthesis of interesting terpenoids.
4. Rapid development in the field of genome editing provides ample tools for further engineering of heterologous hosts that will significantly increase substrate or precursor supply, leading to a higher success rate for CYP discovery and high yield of target terpenoid for potential industrial applications.
5. CYP engineering through directed evolution or rational design is an alternative way to obtain CYPs with desired properties for a particular terpenoid biosynthetic pathway. It also expands the substrate scope of known CYPs and widens their applications in various situations.

Despite the indispensable role of gene deletion in elucidation of CYP function, it is impossible to perform gene deletion in many terpenoid producers with immature genetic manipulation strategies. Even for some genetically tractable species (such as plants), their large unknown genomes and multi-copy alleles make it very difficult to perform proper gene deletion. However, development of gene editing tools, such as the emerging CRISPR-Cas technology [20], particularly the CRISPR-Cas ribonuclease protein editing technology [21], may provide a novel way to address this problem.

previridicatumtoxin, respectively. Accordingly, VrtE and VrtK were speculated to be responsible for C5-hydroxylation and spirobicyclic ring formation, respectively [19].
Gene Silencing

To circumvent the difficulty in gene deletion, gene silencing is an alternative technique to identify the roles of CYP candidates in plants. This method was used to downregulate the expression of CYPs in Catharanthus roseus that were potentially involved in the biosynthesis of secoliganin, a precursor for the assembly of multiple monoterpenoid indole alkaloids. CrDL7H was characterized as a 7-deoxyloganic acid 7-hydroxylase involved in secoliganin biosynthesis [22]. In another study, silencing cyp76ah1 in Salvia miltiorrhiza significantly increased the accumulation of mitradiene and decreased the amount of tanshinones compared to the wild type, demonstrating the pivotal role of CYP76AH1 in the biosynthesis of tanshinones.

Table 1. Typical Examples for CYP Discovery

<table>
<thead>
<tr>
<th>Technique</th>
<th>Organism(s)</th>
<th>Enzyme(s)</th>
<th>Produced terpenoid(s)</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional approach</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene deletion</td>
<td>Aspergillus fumigatus</td>
<td>Af510</td>
<td>Fumagillin</td>
<td>[18]</td>
</tr>
<tr>
<td></td>
<td>Penicillium aethiopicum</td>
<td>VntE, VntK</td>
<td>Viridicatumtoxin</td>
<td>[19]</td>
</tr>
<tr>
<td>Gene silencing</td>
<td>Catharanthus roseus</td>
<td>CrDL7H</td>
<td>Secologanin</td>
<td>[22]</td>
</tr>
<tr>
<td></td>
<td>Salvia miltiorrhiza</td>
<td>CYP76AH1</td>
<td>Tanshinones</td>
<td>[23]</td>
</tr>
<tr>
<td>In vitro enzymatic assay</td>
<td>Western redcedar</td>
<td>CYP750B1</td>
<td>Sabinol</td>
<td>[24]</td>
</tr>
<tr>
<td></td>
<td>Taxus</td>
<td>CYP725A4</td>
<td>Taxadiene-5α-ol</td>
<td>[25]</td>
</tr>
<tr>
<td>Synthetic biology approach</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multi-level omics-guided CYP discovery</td>
<td>Salvia pomifera Heterologous host: Saccharomyces cerevisiae</td>
<td>Fifteen CYP candidates</td>
<td>Carnosic acid</td>
<td>[30]</td>
</tr>
<tr>
<td></td>
<td>Maesa lanceolata Heterologous host: S. cerevisiae</td>
<td>CYP716A75, CYP87D16</td>
<td>C-28 and C-16α oxidized β-amyrin</td>
<td>[31]</td>
</tr>
<tr>
<td></td>
<td>Sorangium celulosum So ce56 Heterologous host: Escherichia coli</td>
<td>CYP260A1, CYP264B1</td>
<td>Eremophilane-, humulane-, and cedrane-type structures; allylic alcohols</td>
<td>[33]</td>
</tr>
<tr>
<td></td>
<td>Mycobacterium HKN-1500 Heterologous host: E. coli</td>
<td>ahpGHI</td>
<td>(--)α-Perillylalcohol</td>
<td>[34]</td>
</tr>
<tr>
<td></td>
<td>Ganoderma lucidum Heterologous host: S. cerevisiae</td>
<td>CYP5150L8</td>
<td>Ganoderic acid</td>
<td>[15]</td>
</tr>
<tr>
<td></td>
<td>Medicago truncatula Heterologous host: S. cerevisiae</td>
<td>CYP93E2, CYP716A12</td>
<td>4-epi-Hederagenin</td>
<td>[38]</td>
</tr>
<tr>
<td>Heterologous expression in genetically tractable hosts</td>
<td>Lotus japonicas Heterologous host: N. benthamiana</td>
<td>LJCPY71D353</td>
<td>20-Hydroxylupeol</td>
<td>[41]</td>
</tr>
<tr>
<td></td>
<td>oats Heterologous host: N. benthamiana</td>
<td>AsCPY51H10</td>
<td>12,13β-Epoxy-16β-hydroxy-β-amyrin</td>
<td>[42]</td>
</tr>
<tr>
<td></td>
<td>Emericella variecolor NBRC 32302 Heterologous host: Aspergillus oryzae NSAR1</td>
<td>Gnn-P450</td>
<td>Quiannulatic acid</td>
<td>[43]</td>
</tr>
<tr>
<td>Metabolic engineering of chassis cells</td>
<td>S. cerevisiae</td>
<td>–</td>
<td>Amorpha-4,11-diene</td>
<td>[48]</td>
</tr>
<tr>
<td></td>
<td>S. cerevisiae and Pichia pastoris</td>
<td>–</td>
<td>trans-Nootkatol</td>
<td>[51]</td>
</tr>
<tr>
<td></td>
<td>S. cerevisiae</td>
<td>–</td>
<td>β-Amyrin and its derivatives</td>
<td>[54]</td>
</tr>
<tr>
<td></td>
<td>E. coli</td>
<td>–</td>
<td>β-Carotene</td>
<td>[55]</td>
</tr>
</tbody>
</table>
Traditional approaches

- **Gene deletion**
  - CYP gene
  - ![Schematic diagram of gene deletion](image)

- **Gene silencing**
  - iRNA
  - Target mRNA
  - mRNA degradation
  - ![Schematic diagram of gene silencing](image)

- **In vitro enzymatic assay**
  - CYP+substrate
  - ![Schematic diagram of in vitro enzymatic assay](image)

Synthetic biology approaches

- **Multi-level omics-guided CYP discovery**
  - Genome/transcriptome data
  - Computational analysis
  - Promising CYP candidates
  - ![Schematic diagram of multi-level omics-guided CYP discovery](image)

- **Heterologous expression in genetically tractable hosts**
  - CYP candidates of various sources
  - Expression in heterologous cell factories
  - Fermentation and product analysis
  - ![Schematic diagram of heterologous expression](image)

- **Metabolic engineering of chassis cells**
  - Pathway engineering for sufficient substrates
  - Optimization of cofactor and electron transfer
  - Membrane engineering
  - ![Schematic diagram of metabolic engineering](image)

**Figure 2. Approaches for CYP Discovery.** Traditional approaches consist of gene deletion, gene silencing, and in vitro enzymatic assays, while synthetic biology approaches include multi-level omics-guided CYP discovery, heterologous expression in genetically tractable hosts, and metabolic engineering of chassis cells. Each approach is represented by a schematic diagram. CoA, coenzyme A; CYP, cytochrome P450; FDP, farnesyl diphosphate; FPP, farnesyl pyrophosphate.
of tanshinones [23]. However, similar to gene deletion, gene silencing is often infeasible for genetically intractable hosts.

**In Vitro Enzymatic Assays**

*In vitro* enzymatic assays can characterize the function of CYPs with relatively good solubility and stability. Because of the absence of genetic regulation or transport barriers, the cell-free environment facilitates substrate access and product detection. The catalytic role of *Western redcedar* CYP750B1 was confirmed by an *in vitro* assay using cell-free tissue extracts in the presence of NADPH and both (+)-trans-sabinol and (+)-cis-sabinol. Only the conversion of (+)-trans-sabinol to sabinone was detected, demonstrating the stereo-selectivity of CYP750B1 [24]. Recently, the *in vitro* lipid nanodisc assay has been successfully used to characterize the function of CYP725A4 from *Taxus*. The purified C-5 hydroxylase CYP725A4 was encapsulated together with CPR, NADPH, and taxadiene in a lipid nanodisc for *in vitro* bioconversion, and taxadiene-5α-ol was identified as the primary product [25]. Such a one-pot *in vitro* assay using purified CYP together with its partner CPR, cofactor, and substrate could avoid CYP deactivation and the possible interference from intracellular substrate during microsomal preparation.

Despite the successful examples mentioned above, difficulties still persist in *in vitro* characterization of CYPs. For instance, during purification, membrane-bound CYPs usually exhibit poor solubility or aggregation, which may occlude the active sites of enzymes and block their functions.

**Synthetic Biology Approaches for CYP Discovery**

Given that the traditional methods mentioned above have various shortcomings, recent advances in synthetic biology have created alternative methods for CYP discovery. Synthetic biology is distinguished by its potential for rationally designing novel biological systems with standardized components, aiming to construct a formulated objective or an efficient cell factory that can tackle challenges. Multi-level omics analysis reveals the most promising candidates and promotes successful and efficient CYP discovery. Benefitting from the efficient genetic manipulations of heterologous hosts, *heterologous expression* of CYPs for subsequent characterization is standardized and convenient. With a modularized and fine-tuned supply of precursors, cofactors, and energy, metabolic engineering of *chassis cells* maximizes the likelihood of product detection. Taken together, these synthetic biology approaches illustrate a clear systematic workflow that can accelerate CYP discovery (Box 2; Table 1; Figure 2).

**Multi-level Omics-Guided CYP Discovery**

The first aspect of CYP discovery using a synthetic biology approach is the identification of candidate CYPs. Rapid development in genome and transcriptome sequencing technologies makes it much easier to obtain a large CYP candidate pool. However, it is very difficult to pinpoint the exact CYPs involved in a target pathway. Detailed transcriptome analyses of samples from different tissues, times, and mutants, combined with metabolomics studies targeting specific terpenoids, have been used to narrow down the CYP candidate pool [24,26–29]. For example, to elucidate the biosynthesis of carnosic acid and related diterpenoids in *Salvia pomifera*, 81 contigs similar to CYP genes were generated via genome sequencing. Based on the elevated transcription in the trichome of *S. pomifera* and sequence similarity with other plant-derived CYP71 and CYP76 family members, 15 CYP candidates were short-listed for further analysis [30]. In addition to looking for clues in samples with internal high transcript abundance, external stimuli can also induce terpenoid production to reveal cryptic key CYPs. For example, transcription profiling on methyl jasmonate (MeJA)-treated *Maesa lanceolata* shoot cultures was performed to identify maesasaponin biosynthesis genes.
After MeJA treatment for 24 h, seven transcriptional upregulated CYPs were retained for functional analyses. Finally, CYP716A75 and CYP87D16 were identified to be responsible for catalyzing the C-28 and C-16α oxidations of β-amyrin, respectively [31].

Heterologous Expression in Genetically Tractable Hosts
Because they have known genome sequences, ample genetic tools, and diverse precursors for terpenoid biosynthesis, *Escherichia coli* and *Saccharomyces cerevisiae* have been extensively developed for decades as important hosts to biosynthesize terpenoids [32]. Sesquiterpene hydroxylases CYP260A1 and CYP264B1 from *Sorangium cellulosum* So ce56 were expressed in *E. coli* to compare their substrate specificities and regio-selectivities. As a result, CYP260A1 was identified to catalyze sesquiterpene conversion for eremophilane-, humulane-, and cedrane-type structures, and CYP264B1 was characterized as a highly selective sesquiterpene hydroxylase that catalyzed zerumbone, α-humulane, β-caryophyllene, and eremophilane-type substrates to produce allylic alcohols [33]. Most native CYPs involved in limonene biotransformation have a substrate preference for (+)-d-limonene. To use (−)-l-limonene, ahpGHI from *Mycobacterium* HXN-1500 was expressed in *E. coli* to convert (−)-l-limonene to (−)-l-perillylalcohol [34]. In addition, since the endoplasmic reticulum and post-translational modification system of yeast are suitable for expressing membrane-bound eukaryotic CYPs, functional expression of eukaryotic CYP candidates was often conducted in yeast [15,26, 35–37]. For example, to identify a CYP for converting lanosterol to triterpenoid ganoderic acid, a systematic screening of CYP candidates from *Ganoderma lucidum* was performed in *S. cerevisiae*. The CYP515OL8-overexpressed strain generated 3-hydroxy-lanosta-8, 24-dien-26 oic acid (ganoderic acid HLDOA), and the final titer in 120-h fermentation reached 14.5 mg/l [15]. In addition to biosynthesis of the known terpenoids, rare triterpenoids, which are absent in the native host, could also be produced in engineered yeasts. Expression of bAS, CPR, CYP93E2, and CYP716A12 enabled production of 4-epi-hederagenin in the recombinant yeast, which is a triterpenoid not reported in *Medicago truncatula* [38].

The transient expression system of *Agrobacterium tumefaciens*-infected *Nicotiana benthamiana* is an alternative platform to study plant-derived CYPs [26,39,40]. Using this platform, candidate CYPs from *Lotus japonicus* were heterologously expressed in *N. benthamiana* leaves infiltrated with *A. tumefaciens*, and gas chromatography (GC)-mass spectrometry analysis of extracted metabolites indicated that LjCYP71D353 catalyzed the formation of 20-hydroxyypeol from dihydroolpeol [41]. Similarly, heterologous coexpression with β-amyrin synthase in *N. benthamiana* leaves helped identifying an oat-derived CYP AsCYP51H10 [42].

*Aspergillus* is another potential host for CYP expression. The Qnn-P450 of *Emericella variecolor* NBRC 32302 was heterologously coexpressed with sesterterpene synthase EvQS in *Aspergillus oryzae* NSAR1 and able to successively oxidize C-19 of quianulatene to quiannulatic acid [43]. Another promising microbial host for terpenoid biosynthesis is *Bacillus subtilis*. This microbe has an inherent methylefrtythiol phosphate (MEP) pathway and a wide substrate spectrum available for various terpenoid biosyntheses [44]. Furthermore, there are a few successful examples of heterologous biosynthesis of terpenoids in the moss *Physcomitrella patens*, due to its natural tolerance toward terpenoids, distinguished endogenous terpenoid producing profile that facilitates the characterization of exogenous products, and accessibility to genome editing tools [45].
### Metabolic Engineering of Chassis Cells

CYP discovery is highly dependent on how well a host is engineered. First, the host must provide sufficient rare and sometimes unstable substrates to confirm CYP activity, which otherwise has to rely on grueling purification from native hosts. Second, host metabolic engineering provides efficient cofactor regeneration, electron transfer, and CYP expression to maximize CYP activity. Last, membrane engineering facilitates the efficient storage of hydrophobic terpene products so that product titers can be significantly improved.

Host engineering for precursor supply is a generally applicable approach for efficient terpenoid biosynthesis [46]. Acetyl-coenzyme A (CoA) is an important intermediate in the MVA pathway for terpenoid biosynthesis, and its concentration in yeast mitochondria is estimated to be 20–30-fold higher than that in the cytosol [47]. The eight-gene FPP biosynthetic pathway was expressed in yeast mitochondria to produce amorpha-4,11-diene; this strategy overcame the low acetyl-CoA concentration and bypassed competition in the cytosol [48]. In another example, an engineered yeast platform incorporated different modules of prenyl diphosphate substrate synthesis, terpene skeleton synthesis, and terpene skeleton decoration, and heterologous CPR was used to identify the catalytic roles of CYP76AH24, CYP71BE52, CYP76AK6, and CYP76AK8 in camosic acid biosynthesis [30].

The oxidation of CYP requires a complicated source of electrons from NAD(P)H via redox partners; thus, these redox partners are often heterologously expressed to support the activities of CYPs [49,50]. To ensure sufficient suitable CPRs for supporting the catalytic capacity of CYPs, a novel positive effector ICE2 (the type III membrane protein) was identified with the ability to stabilize CPRs in both S. cerevisiae and Pichia pastoris. Overexpression of ice2 improved the conversion of (+)-valencene to trans-nootkatol by 40–50%, indicating the potential of ICE2 as a general tool for improving the performance of recombinant CYPs in yeasts [51].

Plant-derived CYPs and CPRs are usually insoluble and localized at the endoplasmic reticulum (ER) membrane by a hydrophobic N-terminal domain, hindering their reconstitution and heterologous expression in microbial hosts. Hence, engineering the ER or cell membrane is a promising strategy for improving the expression of eukaryotic CYPs and enhancing the storage of hydrophobic terpenoid products [52,53]. Disrupting the phosphatidic acid phosphatase gene pan1 led to the expansion of the ER and enhanced the accumulation of β-amyrin and its derivatives [54]. The membrane of E. coli was engineered by a combination of over-expressing membrane-bending proteins and enhancing the membrane synthesis pathway, leading to a 2.9-fold increase of β-carotene production [55].

### Engineering CYPs in Terpenoid Biosynthesis

To obtain a CYP with improved or novel functions, several strategies can be generally adopted for CYP engineering, including directed evolution [56] and rational design [57]. In addition, when eukaryotic CYPs are expressed in heterologous hosts, the membrane-anchored domains are usually truncated to promote the catalytic performance of CYPs (Figure 3).

#### Directed Evolution

Directed evolution is conducted to obtain CYP variants with desired characteristics, including enhanced enzyme activity, stability, or altered substrate specificity. For example, to avoid using the expensive cofactor NADPH in cell-free, CYP-catalyzed reactions, the fusion protein P450cin-ADD-CinC was engineered to use zinc/cobalt(III) sepulchrate as electron delivery system for an increased hydroxylation activity toward 1,8-cineole. After two rounds of
CYP engineering

Figure 3. CYP Engineering Methods. CYP engineering methods include directed evolution, rational design, and terminal modification, each of which are represented by schematic diagrams. Enzyme* represents engineered enzyme by rational design.

sequence saturation mutagenesis, with each followed by one round of multiple-site saturation mutagenesis, the variant KB8 exhibited a 3.8-fold increase in catalytic efficiency (from 7 to 28 μM⁻¹ min⁻¹) and a 1.5- and 4-fold increase in product yield and product formation rate, respectively, compared to the control enzyme P450cin-ADD-CinC with platinum/cobalt (III) sepulchrate as the electron delivery system [58]. In another study, using iterative saturation mutagenesis, the CYP BM3 (F87A) mutants were able to convert testosterone to either 2β- or 15β-alcohols in a highly selective manner [59]. The most outstanding variant, KSA-14, exhibited a 152-fold increase in product formation rate and significant improvement in the $k_{cat}$ value (from 0.31 to 6.64 s⁻¹) [59]. In the engineering of P450BM3, efficient directed evolution was conducted on the basis of the mutability landscapes and molecular dynamics simulations, generating small but high-quality libraries. The resultant P450BM3 mutants can hydroxylate different steroids specifically at the C16-position in both regio- and diastereoselective manners [60].

Rational Design

Rational protein design is usually used to improve CYP performance with a known crystal structure or an existing homology model. Based on docking analysis of a homology model with progesterone, saturation mutagenesis followed by site-directed mutagenesis was performed to significantly change the regio- and stereoselectivities of CYP106A2 from the 15β-position to the 11α-position in progesterone hydroxylation [61]. In another work, docking studies of CYP106A2 indicated that A243S mutation in the binding pocket could enable a new hydrogen bond to the carbonyl oxygen at position 21 and stabilize a corresponding conformation to facilitate 6β hydroxylation, which was further confirmed by experimental results [62]. In the engineered P450BM3, a phenylalanine residue in the substrate channel to the heme iron was first replaced by smaller hydrophobic amino acids, and then the first-shell amino acids with
maximum possibility to directly interact with the substrate were subjected to site mutagenesis to alter the regio- and stereoselectivity of this enzyme. These efforts enabled P450\textsubscript{BM3} variants to hydroxylate \(\beta\)-cembrenediol at C9-and C10-positions with 100 and 97\% regioselectivity, respectively [63]. Moreover, the potential of unnatural amino acid substitution was investigated to modulate the activity and selectivity of CYP. Based on the crystal structure of a promiscuous CYP102A1 variant, 11 sites were selected for substitution with four unnatural amino acids, which represent diversified aromatic side-chain groups. The resultant variants were observed with large shifts in regioselectivity in the oxidation of small-molecule drug (S)-ibuprofen methyl ester and the bulkier natural product (+)-nootkatone [64]. In addition, P450 fingerprinting is a useful tool for CYP engineering. Using semisynthetic chromogenic probes, the active site configuration of engineered P450 variants was mapped to generate reliable prediction of activity–substrate structure relationship, which may provide convenient and effective guidelines for CYP engineering [65].

**Terminal Modification**

To avoid insolubility, instability, or total loss of activity in heterologous CYP expression, modification of the membrane anchor region is commonly required. The N-terminal membrane anchor region of *Rosmarinus officinalis* CYP76A1-H4 was replaced with the leader peptide MAKKTSSKGK to guarantee effective function in *E. coli*, and this enzyme was characterized as a key enzyme in the production of the phenolic diterpenoid ferruginol [66]. To achieve heterologous biosynthesis of opiates in yeast, the key CYP enzyme SalSyn was engineered to avoid the misprocessing and to increase the corresponding activity. The N-terminal \(\alpha\) helices of SalSyn were replaced with those of cheilanthifoline synthase, and the chimeric protein was yeast codon-optimized, which successfully rescued the N-linked glycosylation caused by the misprocessing of the nascent protein [67].

The electron transfer uncoupling event is often an obstacle for good performance of CYP in heterologous microorganism hosts (Box 1). To produce protopanaxadiol in *S. cerevisiae*, the protopanaxadiol synthase (PPDS) from *Panax ginseng* was fused with the CPR (ATR1) from *Arabidopsis thaliana*. Compared with PPDS and ATR1 coexpression, the fusion enzymes achieved approximately 4.5-fold increase in catalytic activity and 71.1\% improvement in protopanaxadiol production [68].

**Concluding Remarks and Future Perspectives**

Discovering CYPs that are predicted to exist, but have not yet been characterized, is crucial but challenging for efficient terpenoid biosynthesis. Early efforts in CYP discovery focused on gene deletion, gene silencing, and *in vitro* enzymatic assays, which heavily rely on genetic manipulations of the native host, high CYP enzyme activity, and the availability of substrates or precursors. These limitations made the discovery of key CYPs difficult, unpredictable, and sporadic. Recently, with the aid of multi-omics analysis, heterologous CYP expression, and metabolic engineering, the synthetic biology approach has allowed rapid prototyping of CYP candidates in a genetically tractable host with a much higher success rate. Under this paradigm, well-characterized elements (e.g., promoter, ribosome binding site) combined with codon optimization have been adopted for the heterologous expression of CYPs.

As a significant host in synthetic biology, *S. cerevisiae* offers distinct advantages for CYP discovery and terpenoid biosynthesis. First, it naturally generates multiple terpenoid precursors via the MVA pathway, which may save extensive effort in metabolically engineering a host strain. Second, its post-translational modification system and endoplasmic reticulum may support the expression of membrane-bound eukaryotic CYPs. Third, its native CYPs are less...
likely to interfere with the discovery of exogenous CYPs, because there are only a few native CYPs in *S. cerevisiae* and their functions have been well characterized. In addition to *S. cerevisiae*, expression of plant-derived CYPs in heterologous hosts with higher evolutionary affinity, such as in *Aspergillus*, *A. thaliana*, and tobacco, is a promising alternative. These hosts may have more mature regulation systems and native compartments to support the activity of the plant-derived CYPs, but interference from their native CYPs could be a problem. To biologically synthesize complicated terpenoids, simultaneous overexpression of multiple CYPs is usually required but challenging in a heterologous host, where those exogenous CYPs may have severe competition for substrate, CPR, and cofactor. In the future, we believe that microenvironment engineering approaches [69,70], such as improved electron transfer coupling, subcellular compartment targeting and engineering, and multi-enzyme complex engineering, may be required to reconstitute and improve the activity of CYP in a heterologous host, thereby accelerating CYP discovery and improving terpenoid biosynthetic efficiency.

Making the potential products of CYP-catalyzed reactions detectable in heterologous hosts is critical to discover CYPs and subsequently engineer them via synthetic biology approaches. Beyond the aforementioned strategies, fine-tuning host metabolism to supply enough substrates, cofactors, and partners for CYPs and overcoming potential CYP instability and product toxicity problems are highly desirable but cumbersome. Recently emerging genome engineering technologies enable multiplex genome editing with high efficiency and accuracy [71–73]. These primary studies demonstrated the great potential to engineer heterologous hosts both in extent and depth, which will accelerate CYP discovery and engineering in the near future.

Another important direction to further improve CYP discovery and engineering is to develop high-throughput screening methods to confirm their activity. Current color- or fluorescence-based assays are cost effective but difficult to generalize [57], while the most frequently used HPLC or GC analyses are labor intensive [74,75] (see Outstanding Questions). This problem can be partially solved by harnessing the great power of robotic automation [76,77] or high-throughput CYP candidate cloning, expression, and activity screening. A product-based biosensor would provide much higher throughput in CYP prototyping and engineering, and it would be an excellent tool in metabolic engineering for improving target terpenoid production.

**Combinatorial biosynthesis** of terpenoids is another interesting direction in exploring applications of CYPs beyond their native activities. By harnessing the promiscuity of CYPs, combined with terpenoid cyclases and empowered by high precursor supply in heterologous hosts, combinatorial biosynthesis may rapidly generate more terpenoids than nature offers [78–80].

**Acknowledgments**

The authors gratefully acknowledge the National Natural Science Foundation of China (31600071), the Shanghai Municipal Natural Science Foundation (17ZR1448900, 18ZR1420300), the Key Project of Chinese Academy of Sciences (QYZDB-SSW-SMC012), the 100 Talents Program of Chinese Academy of Sciences, the Bioresource service network project of the Chinese Academy of Sciences (ZYSYS-018), the Key Research Program of the Chinese Academy of Sciences (KFZD-SW-219), the International cooperation project of the Chinese Academy of Sciences (153D31KYSB20170121), and the Key Projects in the Tianjin Science & Technology Pillar Program(15PTCYSY00020) for financial support. We thank Prof. Jian-Jiang Zhong (Shanghai Jiao Tong University) for valuable suggestions on improving the language quality of the manuscript.
References

22. Salim, V. et al. (2013) Virus-induced gene silencing identifies Catharanthus roseus 7-deoxoyuridic acid-7-hydroxylase, a step in iridoid and monoterpene indole alkaloid biosynthesis. Plant J. 76, 754–765
37. Forman, V. et al. (2017) Production of putative diterpene carbocyclic acid intermediates of triptolide in yeast. Molecules 22, 981