Cytochrome P450s (P450s) are the most versatile catalysts and are widely used by plants to synthesize a vast array of structurally diverse specialized metabolites that not only play essential ecological roles but also constitute a valuable resource for the development of new drugs. To accelerate the metabolic engineering of these high-value metabolites, it is imperative to identify and characterize pathway P450s, and to further improve their activities through protein engineering. In this review, we focus on P450 engineering and summarize the major strategies for enhancing the stability and activity of P450s and successful cases of P450 engineering. Studies in which the functions of P450s were altered to create de novo metabolic pathways or novel compounds are discussed as well. We also overview emerging tools, specifically DNA synthesis, machine learning, and de novo protein design, as well as the evolutionary patterns of P450s unveiled from a massive number of DNA sequences that could be integrated to accelerate the engineering of these enzymes. These approaches would greatly aid in the exploitation of plant-specialized metabolites or derivatives for various uses including medical applications.

Key words: plant cytochrome P450, plant natural product, protein engineering, amino acid co-evolution, protein design

Shang Y. and Huang S. (2020). Engineering Plant Cytochrome P450s for Enhanced Synthesis of Natural Products: Past Achievements and Future Perspectives. Plant Comm. 1, 100012.

INTRODUCTION

In adaptation to constantly changing environments, plants synthesize over 200,000 structurally and functionally diverse metabolites, such as alkaloids, terpenoids, and phenols. These specialized compounds serve not only as a rich reservoir of chemicals with essential ecological functions, but also as a valuable source of new pharmaceuticals, cosmetics, flavors, as well as pesticides (Wink, 2010; Kroymann, 2011). To exploit the potentials of plant-derived metabolites, it is imperative to understand how these bioactive compounds are synthesized and how their diverse structures and activities are formed in plants.

Generally, plants employ signature enzymes (e.g., cyclases) to form the basic carbon backbones, followed by tailoring enzymes (e.g., P450s, acyltransferases, and glycosyltransferases) that add specific functional groups to form various metabolites. Among these biocatalysts, the P450 superfamily is the most versatile, participating in irreversible and often rate-limiting reactions, including regio- and stereoselective hydroxylation, as well as complex reactions, including epoxidation, sulfoxidation, aryl–aryl coupling, nitration, and oxidative and reductive dehalogenation (Guengerich and Munro, 2013). Because of their pivotal roles in the biosynthesis of plant-specialized products, P450s are of great interest and have been selected as essential components for the metabolic engineering of plants (Urlacher and Girhard, 2019). As several reviews have addressed the progress made on genome mining and functional characterization of plant P450s, this review focuses on the topic of P450 engineering and its potential application in the metabolic engineering of plant-derived natural products.
The importance of modulating P450s in plant metabolite engineering

Enzymes with high catalytic efficiency and selectivity are crucial to the reconstruction of metabolic pathways in plants or heterologous hosts. Unfortunately, studies have shown that plant-specialized metabolism is inherently slow when compared with central metabolism (Bar-Even et al., 2011). There has been speculation that this is due to central metabolism being more conserved across phylogenetic groups, and thus the related enzymes tend to have been better optimized during evolution. In contrast, secondary metabolism aids in the adaptation to changing environments, which requires the associated enzymes to be more flexible regarding substrate recognition and product spectrum at the cost of lower catalytic rates (Nam et al., 2012). This is the case for plant P450s, which usually exhibit low catalytic efficiencies ($k_{cat} < 5 \text{ s}^{-1}$) and are well known for their substrate promiscuity (Bar-Even and Tawfik, 2013; Bernhardt and Uricher, 2014). In addition to low rates and poor specificities, plant P450s have two other characteristics that hinder their widespread use in engineering applications. First of all, these enzymes tend to be incompatible with non-native hosts; statistical data suggest that ~40% of plant P450s are poorly expressed in yeast (Renault et al., 2014). Secondly, the electron flux from NAD(P)H, the ultimate electron source in cells, to P450s is usually mediated by an additional reduct partner, such as cytochrome P450 reductase (CPR). This introduces several more optimization requirements, as the inefficient electron transfer resulting from either the lack of sufficient NAD(P)H levels or poor interactions between P450s and CPRs would reduce the enzymatic activity (Lundemo and Woodley, 2015).

An emerging area of P450 engineering is the creation of abiotic reactions for the de novo synthesis of chemical products and their derivatives. In plant genomes, approximately 1% of the coding genes are P450s (Nelson and Werck-Reichhart, 2011). Although recent advances have been made in genome mining using multiomics data to accelerate the identification of enzymatic pathways, this process is still tedious and challenging. For example, the biosynthetic pathway of taxol, an anti-cancer agent originally isolated from the bark of pacific yew trees, has not been fully elucidated. The missing steps from taxa-4(5),11(12)-diene-10β-ol to 10-deacetyl baccatin III are thought to be catalyzed by P450s (Kusari et al., 2014). Based on the fact that intermediates produced in the taxol pathway should be structurally similar, this information could be utilized to search for the missing P450s from genomic data of yew trees using the method described by Zhao et al. (2013). In addition, it is also possible to repurpose two known P450s in the taxol pathway, T5αH and T10βH, that have substrates structurally similar to those of the missing P450s (Schoendorf et al., 2001; Jennewein et al., 2004), thereby completing the taxol biosynthesis pathway (Figure 1). Apart from pathway gap-filling, plant P450s could be functionally modified to add functional groups to high-value metabolites to mitigate their unwanted side effects (e.g., cytotoxicity). For example, gossypol, a sesquiterpene aldehyde that accumulates in cotton seeds, has been demonstrated to exhibit anti-tumor activity against several cancer cells (Neergheen et al., 2010), but there is also evidence pointing to its anti-nutritional properties (Sunilkumar et al., 2006). Fortunately, several related intermediates with potentially similar bioactivities that are less toxic have been identified recently (Tian et al., 2018, 2019). In these studies, it was shown that cotton uses CYP71BE79, a highly active enzyme, to decorate one of the intermediates (8-hydroxy-7-keto-6-cadinene) and prevent its phytotoxicity during gossypol synthesis (Tian et al., 2018, 2019).

Interestingly, while genes involved in secondary metabolic pathways tend to be clustered in the plant genome (Field et al., 2011; Chae et al., 2014), the gossypol pathway genes are dispersed, and the three P450s identified belong to different families, suggesting that phylogenetically distant P450s can also act on structurally similar substrates (Tian et al., 2018).
ENGINEERING OF P450 PROTEINS FOR METABOLIC ENHANCEMENT

Rational design and directed evolution are the strategies commonly used in protein engineering (Chen and Zeng, 2016). Rational design uses physics-based energy functions calculated from models or crystal structures of proteins to guide modifications that alter the substrate specificity, structural stability, or kinetic properties of enzymes. As conformational dynamics of a protein provides a basis for it to acquire new functions (Tokuriki and Tawfik, 2009), qualitatively accurate methods for computing the protein energy landscape are required for efficient rational design (Ranganathan, 2018). In addition, as some protein families, such as P450s, are notoriously difficult to crystallize, accurate protein modeling methods are also important. Directed evolution is a nature-inspired protein engineering approach originally developed in the early 1990s (Chen and Arnold, 1993). The basic logic of this approach is to generate sequence diversity, then iteratively select variants with desired properties. As nature has spent millions of years optimizing protein functions in different species, successful application of directed evolution largely relies on the ability to generate diverse protein sequences and select proper variants using high-throughput systems (Jung et al., 2011; Bassalo et al., 2016). A number of new enzymes, especially P450s, with novel functions have been created using this approach (Arnold, 2015).

The Enzyme-Host Compatibility Improvement

In terms of engineering plant P450s, one major challenge lies in improving enzyme-host compatibility, as P450s involved in the biosynthesis of plant-specialized products are membrane-bound and different from the soluble cytoplasmic P450s in prokaryotes (Munro et al., 2007). To this end, the residues within the N-terminal transmembrane anchor region of eukaryotic P450s could be deleted or modified to improve protein solubility without affecting activity. The first crystal structure of a mammalian P450 (CYP2C5) was obtained by applying this method (Williams et al., 2000). Moreover, plant P450s involved in secondary metabolism are usually poorly expressed in yeast or bacteria. Thus, the promoter, codon usage, as well as copy number, also need to be optimized to increase the yield of desired compounds (Figure 2A). For example, an induced promoter and a codon optimized gene were used in a study to successfully achieve higher P450 expression, which resulted in a 35% increase in the production of perillyl alcohol in yeast (Brown et al., 2015). One alternative to microbial platforms, is using tobacco as an efficient plant-based host to optimize the expression of plant P450s (Reed et al., 2017). Finally, substrate specificity is another factor that affects the metabolic flux toward target compounds. Although the broad
Plant Communications

shade specificity of P450s increases its chance to evolve new functions, this substrate promiscuity could impose a kinetic barrier that blocks the synthesis of desired products, thus requiring substrate recognition site engineering. For instance, a 5.6-fold higher activity was achieved for Coleus forskohlii CYP76AH15 by tuning its substrate preference (Forman et al., 2018).

The Electron Flux Optimization during Oxidation Reactions

Most eukaryotic P450s are membrane-bound proteins mainly expressed in the endoplasmic reticulum (ER) (Werck-Reichhart and Feyereisen, 2000). The lipid composition and amount of ER in the cell determine the localization as well as expression of ER-resident P450 enzymes, which in turn affects their behaviors (Strobel et al., 1970; Negretti et al., 2014; Park et al., 2014). For instance, oleaginous yeasts that are rich in membrane compartments provide an ideal environment for introducing plant-derived P450s involved in the biosynthesis of flavonoids (Lv et al., 2019). Another way to improve the activity of P450s is to enhance the physical proximity between a P450 enzyme and its redox partner on the ER membrane, leading to better electron flux during oxidation reactions (Figure 2B). This was demonstrated previously when the P450 isoflavone synthase 1 from Glycine max was fused to the bacterial P450 reductase domain from Rhodococcus sp. to form the first active P450 fusion enzyme (Sabbadin et al., 2010). Optimization of taxol precursor production by engineering the N-terminal sequence of T5-H and linking this enzyme with a CPR from Taxus represents another elegant example (Ajikumar et al., 2010). Although both the activity and solubility of plant P450s could be increased through protein fusion, introduction of these artificial enzymes into host cells may lead to uncoupled electron transfer or excessive production of reactive oxygen species, both of which are toxic to cells (Renault et al., 2014). Therefore, balancing the expression ratio between P450 and CPR is another way to efficiently optimize the electron transfer process while mitigating the aforementioned issues. To demonstrate the effectiveness of this idea, a cytochrome b₅ was introduced into yeast with low CPR and normal CYP71AV1 expression levels. The results showed that both yeast viability and the yield of artemisinic acid, the precursor to artemisinin, were improved (Paddon et al., 2013). The importance of regulating the electron transfer process by cytochrome b₅ can also be seen in the engineering of flower color. For example, overexpression of a cytochrome b₅ and CYP75A is required for the accumulation of blue pigment in carnations (Tanaka and Bruglieri, 2013), whereas a pink color is achieved by the downregulation of CYP75B in cyclamen flowers (Boase et al., 2010).

Creation of the Abiotic Reactions for P450s

At present, engineering P450s to catalyze abiotic reactions is a field of primary interest. The catalytic cycle and the reactive intermediates produced during the process are essential to the activity of a P450 enzyme (Krest et al., 2013). For example, the iron-oxo species, also known as compound I, tends to abstract hydrogens from substrates, which favor proton-coupled oxidation (Green, 2009). The iron-peroxo, or hydroperoxide, works as an oxidant in epoxidation, sulfoxidation, and C–C cleavage reactions (Jin et al., 2003; Cryle and De Voss, 2006). The ferric-superoxo species can mediate nitrination reactions (Barry et al., 2012). Residue mutations that affect either the formation or stabilization of these reactive intermediates would result in novel P450 reactions (Figure 2C). As an example, a cysteine universally conserved among P450s is involved in ligating the heme iron from the axial side. Mutating this important residue to serine abolishes the native monooxygenase activity of CYP102A1 (P450 BM3) and enables the variant to have new catalytic abilities, such as the ability to catalyze C–H amination and cyclopropanation reactions (Coelho et al., 2013a; Wang et al., 2014). By introducing this cysteine mutation as well as other residue changes, researchers have enabled P450 BM3 to perform additional reactions, including sulfimidation to generate S–N bonds (Farwell et al., 2014), aziridination of aryl olefins (Farwell et al., 2015), and carbene transfer to alkynes (Chen et al., 2018). Another conserved residue is a threonine residue that is located within the I helix on the distal side of the heme and usually acts as a proton shuttle for O–O bond scission to produce compound I (Krest et al., 2013). Olefin cyclopropanation activity was introduced to P450 BM3 by mutating residues including this threonine residue (Coelho et al., 2013b). Currently, efforts to engineer P450s to create de novo reactions primarily focus on enzymes originating from microbes and mammals with very few successful examples related to plant enzymes. Due to the limited understanding of P450s (e.g., enzyme dynamics), the targeted engineering of the regio- and stereoelectivity of plant P450s still remains a major challenge.

TOOLBOXES AND PERSPECTIVES FOR ENGINEERING PLANT P450S

Protein engineering approaches, especially those developed based on directed evolution, require the capacity to generate a large number of sequences, and this is usually achieved by performing ep-PCR (Cadwell and Joyce, 1992), site-directed mutagenesis (Wang and Malcolm, 1999), multiplex automated genome engineering (Wang et al., 2009), or genome shuffling (Zhang et al., 2002). With the advances in DNA synthesis technology and decreases in associated costs, it has become much easier to perform direct synthesis of DNA sequences to introduce sequence mutations, optimize codon usage for specific hosts, and piece together gene elements as a starting point for protein engineering. However, the increased ability to create diverse sequences also requires corresponding high-throughput screening methods for the selection process. Allosteric transcription factor or RNA-based biosensors using either fluorescent proteins or antibiotic resistance genes as reporters have recently been developed and applied to accelerate protein engineering (Bassalo et al., 2016; Wan et al., 2019). In addition, as cell-dependent systems have several innate drawbacks, cell-free screening methods, such as droplet- and microchamber-based platforms have been developed as efficient alternatives (Longwell et al., 2017). Overall, applications of these ultra-high-throughput technologies could accelerate protein engineering and transform it into a less painstaking process.

As the cost of next-generation DNA sequencing technology has dramatically decreased, a tremendous amount of sequence data have been generated by big genome sequencing projects,
such as the 10KP (10 000 plants) Genome Sequencing Project (Cheng et al., 2018) and metagenome sequence projects (Rigden and Fernández, 2017). This has provided unprecedented opportunities to study both co-evolution of protein families and co-varying residues within proteins through the combined application of different statistical analyses (De Juan et al., 2013), giving insights into the strategies used by nature in creation of the functional domains as well as residues important for substrate binding, catalysis, and allosteric regulation. With the constantly growing pool of genomic data (>350 000 P450 sequences available in databases) (Nelson, 2018) and the development of new statistical analysis tools, such as machine learning for data mining (Yang et al., 2019), the predicted functional domains or co-varying residues revealed from these analyses are becoming more and more accurate, thus narrowing the choice of beneficial mutations to a much smaller pool, thereby accelerating protein engineering.

De novo computational protein design is an approach used to introduce new structures or functions to enzymes. The software Rosetta, developed in David Baker’s lab, is one of the most popular computational tools for protein design. This method integrates protein structure, folding, binding, and assembly information with powerful computing of all-atom energy functions to create novel proteins that have extremely high stability (Kellogg et al., 2011). As suggested by the de novo protein design, a protein tends to fold to its lowest free energy state (Baker, 2019). This core principle could be used to create stable enzyme variants that serve as a starting point to incorporate a greater number of simultaneous mutations during engineering. In addition, Rosetta structure calculation combined with co-varying residue–residue contacts could be applied to accurately predict protein structure (Ovchinnikov et al., 2017), which is extremely useful for rational redesign of plant P450s that are hard to crystallize or lack homologs with known structures for protein modeling. As a proof of concept, Rosetta and information about co-evolving amino acids were utilized to engineer a multifunctional P450 (CYP87D20) into a monooxygenase (Li et al., 2019), with the aim of creating a de novo pathway to direct metabolic flux from cucurbitacin C (a bitter triterpenoid from cucumber) toward mogrol (the precursor to the natural sweetener mogroside in Siraitia grosvenorii) (Figure 2D).

Although progress has been made in the identification and engineering of plant P450s to accelerate research related to plant natural products, there is still a large gap between academic research and industrial applications. Nature has created a diverse and
Plant Communications

fascinating world of enzymes that remains largely unexplored, and is waiting to be better understood, uncovered, and utilized. Emerging new tools, especially DNA synthesis, machine learning, and de novo computational protein design, combined with the massive amount of DNA sequencing data for P450s, could be integrated to accelerate the engineering of these enzymes (Figure 3). This will eventually lead to better usage of plant-specialized metabolites in various applications.

FUNDING
This work was funded by the National Key Research and Development Program of China (2018YFA0901800) and the National Natural Science Foundation of China (31972433). This work was also supported from Yunnan Provincial and Shenzhen Municipal Governments.

ACKNOWLEDGMENTS
We thank Prof. Xiaoyan Chen (Shanghai Institute of Plant Physiology and Ecology, Chinese Academy of Sciences), Dr. Per Greisen (enEvol), and Dr. Nian Liu (Department of Chemical Engineering, Massachusetts Institute of Technology) for their critical editing of the manuscript. The authors declare no conflict of interest.

REFERENCES
Ajkumar, P.K., Xiao, W.-H., Tyo, K.E., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T.H., Pfeifer, B., and Stephanopoulos, G. (2010). Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. Science 330:70–74.
Alonso-Gutierrez, J., Chan, R., Batth, T.S., Adams, P.D., Keasling, J.D., Petzold, C.J., and Lee, T.S. (2013). Metabolic engineering of Escherichia coli for limonene and perillyl alcohol production. Metab. Eng. 19:33–41.
Arnold, F.H. (2015). The nature of chemical innovation: new enzymes by evolution. Q. Rev. Biophys. 48:404–410.
Baker, D. (2019). What has de novo protein design taught us about protein folding and biophysics? Protein Sci. 28:678–683.
Bar-Even, A., and Tawfik, D.S. (2013). Engineering specialized metabolic pathways—is there a room for enzyme improvements? Curr. Opin. Biotechnol. 24:310–319.
Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D.S., and Milo, R. (2011). The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. Biochem. 50:4402–4410.
Barry, S.M., Kers, J.A., Johnson, E.G., Song, L., Aston, P.R., Patel, B., Krasnoff, S.B., Crane, B.R., Gibson, D.M., and Loria, R. (2012). Cytochrome P450-catalyzed L-tryptophan nitration in thaxtomin within dynamic chromosomal regions. Proc. Nat. Acad. Sci. U S A 109:16116–16121.
Bassolo, M.C., Liu, R., and Gill, R.T. (2016). Directed evolution and synthetic biology applications to microbial systems. Curr. Opin. Biotechnol. 39:126–133.
Bernhardt, R., and Urlacher, V.B. (2014). Cytochromes P450 as promising catalysts for biotechnological application: chances and limitations. Appl. Microbiol. Biotechnol. 98:6185–6203.
Boase, M.R., Lewis, D.H., Davies, K.M., Marshall, G.B., Patel, D., Schwinnow, K.E., and Deroles, S.C. (2010). Isolation and antisense suppression of flavonoid 3′, 5′-hydroxylase modifies flower pigments and colour in cyclamen. BMC Plant Biol. 10:170.
Brown, S., Clastre, M., Courdavault, V., and O’Connor, S.E. (2015). De novo production of the plant-derived alkaloid strictosidine in yeast. Proc. Nat. Acad. Sci. U S A 112:3205–3210.
Cadwell, R.C., and Joyce, G.F. (1992). Randomization of genes by PCR mutagenesis. Genome Res. 2:28–33.
Chae, L., Kim, T., Nilo-Poyanco, R., and Rhee, S.Y. (2014). Genomic signatures of specialized metabolism in plants. Science 344:510–513.
Chen, K., and Arnold, F.H. (1993). Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. Proc. Nat. Acad. Sci. U S A 90:5618–5622.
Chen, Z., and Zeng, A.-P. (2016). Protein engineering approaches to chemical biotechnology. Curr. Opin. Biotechnol. 42:198–205.
Chen, K., Huang, X., Kan, S.J., Zhang, R.K., and Arnold, F.H. (2018). Enzymatic construction of highly strained carbohydrates. Science 360:71–75.
Cheng, S., Melkonian, M., Smith, S.A., Brockington, S., Archibald, J.M., Delaux, P.-M., Li, F.-W., Melkonian, B., Mavrodiev, E.V., Sun, W., et al. (2018). 10KP: a phylodiversity genome sequencing plan. Gigascience 7:1–9.
Coelho, P.S., Wang, Z.J., Ener, M.E., Baril, S.A., Kannan, A., Arnold, F.H., and Brustad, E.M. (2013a). A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins in vivo. Nat. Chem. Biol. 9:485.
Coelho, P.S., Brustad, E.M., Kannan, A., and Arnold, F.H. (2013b). Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. Science 339:307–310.
Cryle, M.J., and De Voss, J.J. (2006). Is the ferric hydroxyperoxide species responsible for sulfur oxidation in cytochrome P450s? Angew. Chem. Int. Ed. 45:8221–8223.
De Juan, D., Pazos, F., and Valencia, A. (2013). Emerging methods in protein co-evolution. Nat. Rev. Genet. 14:249.
Farwell, C.C., McIntosh, J.A., Hyster, T.K., Wang, Z.J., and Arnold, F.H. (2014). Enantoiospecific imidation of sulfoxides via enzyme-catalyzed intermolecular nitrogen-atom transfer. J. Am. Chem. Soc. 136:8766–8771.
Farwell, C.C., Zhang, R.K., McIntosh, J.A., Hyster, T.K., and Arnold, F.H. (2015). Enantoiospecific enzyme-catalyzed aziridination enabled by active-site evolution of a cytochrome P450. ACS Cent. Sci. 1:89–93.
Field, B., Fiston-Lavrner, A.-S., Kemen, A., Geisler, K., Quesneville, H., and Osbourn, A.E. (2011). Formation of plant metabolic gene clusters within dynamic chromosomal regions. Proc. Nat. Acad. Sci. U S A 108:16116–16121.
Forman, V., Bjerg-Jensen, N., Dyekjar, J.D., Moller, B.L., and Pateraki, I. (2016). Engineering of CYP76A1H5 can improve activity and specificity towards forskolin biosynthesis in yeast. Microb. Cell Fact. 17:181.
Green, M.T. (2009). OH bond activation in heme proteins: the role of thiolate ligation in cytochrome P450. Curr. Opin. Chem. Biol. 13:84–88.
Guengerich, F.P., and Munro, A.W. (2013). Unusual cytochrome P450 enzymes and reactions. J. Biol. Chem. 288:17065–17073.
Jennewein, S., Long, R.M., Williams, R.M., and Croteau, R. (2004). Cytochrome P450 taxadiene 5α-hydroxylase, a mechanistically unusual monooxygenase catalyzing the first oxygenation step of taxol biosynthesis. Chem. Biol. 11:379–387.
Jin, S., Makris, T.M., Bryson, T.A., Sligar, S.G., and Dawson, J.H. (2003). Epoxidation of olefins by hydroperoxy-ferric cytochrome P450. J. Am. Chem. Soc. 125:3406–3407.
Jung, S.T., Lauchli, R., and Arnold, F.H. (2011). Cytochrome P450: taming a wild type enzyme. Curr. Opin. Biotechnol. 22:809–817.
Kellogg, E.H., Leaver-Fay, A., and Baker, D. (2011). Role of conformational sampling in computing mutation-induced changes in protein structure and stability. Proteins 79:830–838.
Krest, C.M., Onderko, E.L., Yosca, T.H., Calixto, J.C., Karp, R.F., Livada, J., Rittle, J., and Green, M.T. (2013). Reactive intermediates in cytochrome P450 catalysis. J. Biol. Chem. 288:17074–17081.
