HHS Public Access

Author manuscript

Trends Biotechnol. Author manuscript; available in PMC 2021 July 01.

Published in final edited form as:

Trends Biotechnol. 2020 July ; 38(7): 729–744. doi:10.1016/j.tibtech.2019.12.008.

Protein engineering for improving and diversifying natural products biosynthesis

Chenyi Li, **Ruihua Zhang**, **Jian Wang**, **Lauren Marie Wilson**, **Yajun Yan***

School of Chemical, Materials and Biomedical Engineering, College of Engineering, The University of Georgia, Athens, GA 30602, USA

Abstract

Proteins found in nature have traditionally been the most frequently used biocatalysts to produce numerous natural products ranging from commodity chemicals to pharmaceuticals. Protein engineering has emerged as a powerful biotechnological toolbox in the development of metabolic engineering, particularly for the biosynthesis of natural products. Recently, protein engineering has become a favored method to improve enzymatic activities, strengthen enzyme stabilities, and expand product spectra in natural products biosynthesis. This review summarizes recent advances and typical strategies in protein engineering, while highlighting the paramount role of protein engineering in improving and diversifying biosynthesis of natural products. Future prospects and research directions are also discussed.

Keywords

protein engineering; natural products; engineered biosynthesis

Challenges in engineering natural products biosynthesis

Due to the high value of natural products (see Glossary) in pharmaceutical, cosmetic and food industries, engineering their biosynthesis has attracted extensive attention [1–3]. The past decades have witnessed a rapid development of natural products biosynthesis in both microbial and cell-free systems [4–6]. By utilizing natural enzymes, with the assistance of strain development, modularized gene expression, dynamic regulations and other metabolic engineering strategies, it is possible to synthesize various classes of value-added natural products and their derivatives, such as fatty acids, isoprenoids, alkaloids, and flavonoids [7–11]. These advances have promoted further research on the biosynthesis of natural products.

However, there are some limitations hindering the development of natural products biosynthesis. When integrating natural enzymes in biosynthetic pathways, their applications are sometimes deterred by limited enzymatic activities, narrow substrate ranges, poor

*Corresponding author: yajunyan@uga.edu (Y. Yan).

stabilities and even loss of function in heterologous hosts. These drawbacks constrain the ability to build cell factories using natural enzymes and make it challenging to further improve productivity or expand product spectra. There have been some efforts at mining novel enzymes from natural resources to solve these problems [12–14], but this approach is still time-consuming and labor-intensive to conduct. Therefore, improving existing enzymes through protein engineering has become attractive and effective to develop biosynthetic approaches to produce natural products. This review summarizes the recent achievements in engineering natural products biosynthesis through protein engineering (Table 1). This paper also highlights the paramount role of protein engineering in improving and diversifying biosynthesis of natural products.

Improving Biosynthesis of Natural Products Through Protein Engineering

Improving catalytic activity of rate-limiting enzymes

One of the main bottlenecks in the biosynthesis of natural products is limited enzyme activity. Integrating heterologous enzymes into microbial chassis may decrease catalytic activities or even result in loss of function. Enhancing enzyme activity to accelerate production processes is a major goal in this area. Improving catalytic activities of enzymes towards specific substrates through random mutagenesis is a typical strategy used in protein engineering (Figure 1). A yeast-active tyrosine hydroxylase was identified and randomly mutated by error-prone PCR to improve its catalytic activity by 4.3-fold. Further coexpression of L-3,4-dihydroxyphenylalanine (L-DOPA) decarboxylase and the engineered tyrosine hydroxylase enabled the demonstration of *de novo* dopamine production in yeast [15]. Similarly, the catalytic activity of isopentenyl diphosphate isomerase (IDI) from *Saccharomyces cerevisiae* (*S.c*) was enhanced by 2.53-fold through PCR-based random mutagenesis. The IDI with enhanced activity resulted in a 2.1-fold increase in lycopene titer (1.24 g/L) compared with the wild-type [16].

Although desired variants have been obtained in several cases through random mutagenesis, most variants generated from this strategy have decreased-to-no activities and the enzyme libraries generated by random mutagenesis are usually enormous. These processes normally require high-throughput screening, otherwise they can be time-consuming and inefficient. Thus, a preferred method is to analyze the structure of target enzymes and design site-directed mutations near the binding sites or active pockets to improve catalytic activities (Figure 1). Recently, protein engineering has benefitted significantly from the rapid development of computer-aided molecular simulation (see Box 1). By resolving crystal structure of target proteins or simulating possible structures based on their homologs, appropriate substitutions can be found in wild-type enzymes [17, 18]. Modification of glycosyltransferase UGT51 in yeast through crystal structure-based rational design enabled an 1800-fold enhanced catalytic efficiency for producing ginsenoside Rh2, a potential anticancer drug isolated from plant ginseng. By further blocking the degradation pathway of Rh2 and increasing the availability of the precursor UDP-glucose, the titer of Rh2 reached 300 mg/L with an over 900-fold increase compared with the starting strain [19]. Site-directed mutagenesis guided by computational simulation was also applied to taxadiene synthase to enable a 2.4-fold improvement in yield for the taxadien-5α-ol, an important

taxol precursor [20]. Similarly, Valliere and colleagues designed several NphB candidates assisted by computational modeling. Two promising candidates M23 and M31 were successfully identified and both increased the k_{cat} value of NphB towards cannabigerolic acid (CBGA) by 1000-fold. M23 was applied in a cell-free system, combined with an optimized flow control strategy, for synthesizing CBGA from olivetol acid (OA) and glucose. The final titer of CBGA reached 1.25 g/L. This study further explored the production of cannabidivarinic acid (CBGVA), a precursor for many rare cannabinoids, by replacing OA with divarinic acid (DA). Because M31 showed 15-fold higher catalytic efficiency than M23, M31 was applied in the final system and resulted in a titer of 1.74 g/L for CBGVA [21]. While the titers are promising, the difficulties of scaling-up cell-free systems hinder further application in industry.

As these examples demonstrate, implementing engineered proteins with enhanced activities in biosynthetic pathways resulted in new breakthroughs that could have substantial impacts on natural products biosynthesis and contribute to innovative biomanufacturing and biopharmaceutical applications. Although improving enzymatic activity through protein engineering is a well-developed area, it still needs support from innovative technologies, such as high-throughput screening techniques and computational biology, and new knowledge from structural biology and biochemistry. For example, it is still difficult to obtain crystal structures for complex proteins and membrane proteins, but revealing such information may help discover new reaction mechanisms and provide valuable insights for future protein engineering. Crystal structures and computer-guided molecular simulation have become increasingly necessary in precise sequence alterations when modulating the properties of enzymes. The revelation of more protein structures and continuing optimization of simulation functions would benefit future research for accurate prediction in protein engineering.

Colocalization to form enzyme complexes

Assembling metabolic pathways with multiple enzymes often results in flux imbalance, where the overall conversion efficiency is limited by specific pathway enzymes. The bottleneck steps result in accumulated intermediates, not only lowering overall conversion efficiency, but also causing potential toxic effects to host cells. In addition to studying the internal properties of enzymes to seek better variants, designing spatial arrangements of artificial enzyme complexes has proven effective for improving overall pathway efficiency. In some cases, enzymes needed to be spatially separated. To optimize morphine biosynthetic route in *S. cerevisiae*, the codeinone reductase was attached to a yeast organelle. The resulting arrangement separated it from the previous enzymatic catalysis and facilitated the spontaneous reaction of neopinone to codeinone [22]. While more often, the researchers have been focused on the assembling of enzyme complexes. With enzyme assembly approaches, fusion enzymes can reduce substrate diffusion, minimize intermediate toxicity and increase carbon flux through target pathways without deleting competing pathways [23–25]. One typical approach to improve diterpene production is to create a fusion enzyme of the yeast farnesyl diphosphate synthase (Erg20p) and geranylgeranyl diphosphate (GGPP) synthase (Bts1p). With Erg20p-Bts1p fusion enzyme, the substrates dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP) were significantly channeled to

diterpene products, including geranylgeraniol, sclareol, cis-abienol and abietadiene [26, 27]. More interestingly, expression of two fusion enzymes, Erg20p-Bts1p and SmCPS-SmKSL (alabdadienyl/copalyl diphosphate synthase fused with akaurene synthase-like), channeled GGPP into the diterpenoid miltiradiene with titer reaching 365 mg/L in a 15-L bioreactor [28]. For downstream products, fusion of the 3, 3′ β-carotene hydroxylase (CrtZ) and 4, 4′ β-carotene oxygenase (CrtW) improved astaxanthin production by 1.4-fold in *E. coli*, reaching 610.4 μg/gDCW (dry cell weight) [29]. Very recently, a tridomain enzyme (CrtB, CrtI, CrtY) harboring the full β-carotene producing pathway decreased precursor accumulation and doubled β-carotene production (2.5 mg/ gDCW) in *S. cerevisiae* than the strain carrying the natural configuration [30]. Engineering enzyme fusion could also improve the performance of the first enzyme with the fused partner enzyme. This especially stands out in cytochrome P450-redox systems, where most P450 enzymes need redox partner systems for their functions [31–33] (see Box 2). These results further suggested the superiority and potential of linking multiple enzymes in improving the performance of natural biosynthetic machinery.

Not limited to linear arrangements of fusion enzymes, assembling synthetic enzyme complexes can benefit more from restrained intermediate diffusion and partial isolation from cellular environment, attenuating toxic effects. Another approach is to bring enzymes that catalyze consecutive reactions physically closer to each other. Naturally occurring protein binding domains have inspired synthetic “scaffold” structures for enzyme engineering (Figure 2). For instance, pathway genes were fused to cellulose binding domain (CBD) with leucine zipper peptides pairs in *E. coli* [34]. A similar self-assembly organization was designed based on decameric methanol dehydrogenase (Mdh) for methanol conversion into fructose-6-phosphate (F6P), where the Mdh enzyme serves as the scaffold [35]. Sequential enzymes 3-hexulose-6-phosphate synthase (Hps) and 6-phospho-3-hexuloseisomerase (Phi) were fused to Mdh via Src homology 3 domains (SH3) linkages and achieved a significant increase in methanol consumption and F6P production. Multiple binding domains were also used simultaneously to study complexing effects. In addition to SH3, PSD95/DlgA/Zo-1 (PDZ) and GTPase binding domain (GBD) were also commonly used [36–38]. This approach was first demonstrated in the mevalonate (MVA) pathway, which includes high-expression acetoacetyl-CoA thiolase (AtoB) and hydroxy-methylglutaryl-CoA synthase (HMGS), and a low-expression hydroxymethylglutaryl-CoA reductase (HMGR) [36]. By attaching varied numbers of HMGR to HMGS via SH3, MVA production improved up to 10-fold, while further increased linkages (more than 6) lowers titer. By attaching AtoB, HMGS and HMGR to GBD, SH3 and PDZ respectively, an enzyme ratio of 1:2:2 led to a dramatic 77-fold increase in titer [36]. The same design was applied in glucaric acid biosynthesis to achieve a 5-fold titer increase with the scaffold unit ratio of 1:4:2 [37], and also in *Saccharomyces cerevisiae* to improve resveratrol production up to 5-fold with a ratio of 1:2:4 [38].

Synthetic nucleic acid scaffolds have also been demonstrated in natural products bioconversion (Figure 2c). By fusing resveratrol biosynthetic enzymes to Zif268 and PBSII zinc-finger (ZF) domains, a 5-fold increase in titer of resveratrol was achieved when expressing 4 repeats of the DNA scaffolds. This design was further studied by a the 3-enzyme pathway for MVA biosynthesis, where 16 repeats of scaffolds containing 1:4:2 ZF

domains led to a 3-fold titer enhancement [39]. Similarly, the binding of transcription activator like effector (TALE) to DNA was exploited for enzyme colocalization. Complexing of tryptophan-2-mono-oxygenase and indole-3-acetimide hydrolase via TALE-based scaffolds resulted in a 9.6-fold increase in indole 3-acetic acid (IAA) production [40]. Other biomolecular scaffolds have been also explored but with limited implementation in natural product biosynthesis, such as aptamer-based RNA scaffolds and viral structural protein-based lipid scaffolds [41, 42]. Although DNA scaffolds have proven effective in enhancing biosynthesis of natural products, more knowledge is needed for efficiently identifying the optimized configurations. Meanwhile, RNA- and lipid-based scaffolds are interesting alternatives that await research in biosynthetic settings.

In addition to enzyme complexes built around synthetic scaffolds, another approach to recruiting enzymes inside of a subcellular space has been studied by mimicking bacterial microcompartments (BMCs) (Figure 2d). BMCs are large protein complexes that assemble with shell subunits and features a major class of protein shells [43]. Carboxysomes are typical BMCs and recruit metabolic pathway enzymes for enhanced conversion rates from hard-to-retain substrates and toxic intermediates [44]. Some well-characterized examples include CO_2 fixation and 1,2-propanediol degradation [45]. Heterologous expression of carboxysome with enzymes has been achieved in *E. coli*, where pyruvate decarboxylase and alcohol dehydrogenase were fused with carboxysome-targeting sequence [46]. Other types of BMCs features smaller polyhedral protein complexes and vaults [43]. Application of BMCs in biosynthesis of natural products is currently hindered by limited knowledge of assembly and functionalization mechanisms. Additional challenges may include burden from structural protein overexpression, physical occupation of BMCs inside the cells, selective permeability of intermediates and the interactions between foreign enzymes and shell proteins. However, enzyme encapsulation still holds great potential as it provides a large and relatively rigid subcellular space allowing intense engineering for high local concentration of compounds while minimizing disruption to the entire host cells. So far, various approaches have been developed for enzyme colocalization, including enzyme fusion, scaffold-based complexing and encapsulation. With further understanding of those nature-inspired schemes, we expect more intriguing engineering that contributes to natural product biosynthesis.

Improving protein stability

Although enzymes have been engineered to improve catalytic activities, enzyme engineering remains challenging as enzymes have distinct structures and functions that are not often easy to predict and modify for desired properties. Another aspect of protein engineering is to improve stability that improves the lifespan of each protein molecule and therefore the overall turnover rate. More stable enzymes are also needed in industrial settings, where enzymes adapted to environments of high temperature and extreme pH are sometimes preferred.

The stability of proteins is multifaceted, including thermal stability, pH tolerance, solubility and tolerance to salts and organic solvents. Increasing thermostability will allow certain enzymes to function in high temperature, as extreme thermophiles have emerged for

biosynthetic applications [47]. Variants of a tryptophan 7-chlorination halogenase RebH were subject to three rounds of mutagenesis-based screening. The best variant carried 8 mutation positions and had a higher melting temperature than the wild type by 18°C. Molecular analysis suggested that thermostability benefited from more surface charge and lower structural flexibility [48]. Guided by a structure-guided consensus concept, thermostable variants of glucose 1-dehydrogenase (GDH) were generated and gained greatly improved half-life to 65°C for 3.5 days and were later found to tolerate high concentrations of salts and organic solvents as well. The mutations were designed by investigating the identity percentage of consensus amino acid sequences and combining of the resulting effective single substitutions [49, 50]. Meanwhile, some natively thermally stable enzymes can inspire reverse engineering for those with similar structure or function. For example, two thermostable sesquiterpene synthases remaining active at up to 78°C have been characterized. Aligning of both enzymes to a τ-murol synthase suggested removal of C-terminal residues of the latter and resulted in moderately increased thermostability, 44.9 to 45.8°C [51].

Enzymes have also been engineered to adapt to desired environmental pH. For instance, laccases, the major players in lignin degradation, prefer higher temperature and pH conditions [52]. Laccases are crucial as they initiate microbial consumption of resistant lignin substrates and allow conversion of aromatic-rich molecules into valuable products. Their catalytic properties therefore need be tuned to work with other enzymes for lignin valorization. A fungal laccase from *Botrytis aclada* was subject to directed evolution for higher optimal pH and gained up to 5-fold activity at neutral pH [53]. To alter pH, multiple mutations are often needed, and the copper sites were important in this example for retaining catalytic activity and redox potentials [53]. The emerging efforts of altering protein stability pioneered the engineering of difficult enzymes for natural product biosynthesis by enabling enzymes to function in harsh conditions, as well as increasing activities in many cases.

Engineering transcriptional regulators-based biosensor

In addition to directly engineering biosynthetic enzymes, engineering accessory proteins, such as genetically encoded biosensors, is also beneficial for production of natural products. Most biosensors are allosteric transcription factors (aTFs) with specific inducers and can recognize certain promoter sequences to repress or activate the downstream expression [54]. Besides real-time monitoring of intermediates or products, aTFs are now enabling synthetic biology applications in dynamic pathway regulating to enhance titer of natural products production and biosensor-based high-throughput screening of efficient enzymes or high-performance strains [55–61]. However, the number of natural aTFs are limited, and they are sometimes not optimal for dynamic controls or biosensor-based screening. Although identifying and characterizing novel sensors has been fruitful [62–66], mining natural aTFs is difficult to catch up with the discovery of new compounds and meet complex regulation or screening requirements. Thus, it is important to engineer existing sensor-regulators with enhanced dynamic range or expanded scope of responsive ligands to improve the efficiency of biosensor-based screening and engineering.

Engineering existing aTFs to broaden the ligand scope of biosensors to sense new chemicals is of considerable interest in the field of regulator engineering (Figure 3a) [67]. For example, the traditional L-arabinose biosensor AraC has been modified to sense a series of ligands, including D-arabinose [68], MVA [69], triacetic acid lactone [70], and ectoine [71]. Site-saturated mutagenesis at certain amino acid positions enabled these AraC variants to respond to ectoines and the best ectoine dose responsive variants were selected. Through biosensor-based high-throughput screening, the specific activity of rate-limiting enzyme L-2,4-diaminobutyric acid (DABA) aminotransferase (EctB) was enhanced by up to 4.1-fold [71]. With a similar approach, the well-known *lac* repressor LacI was engineered to respond to four new inducers: fucose (Q291T), gentiobiose (H70D, H74S), lactitol (I79T, R101H) and sucralose (D149G, I160V, H163Y, S193E), through computational-guided protein design and site-saturated mutagenesis (a typical variant for each inducer is in the corresponding parentheses) [72]. Besides these well-characterized aTFs, many unconventional biosensors have also been engineered to sense natural products and thus can be used to regulate and improve their biosynthesis. TtgR from *Pseudomonas putida* (*P.p*), which is normally responsible for resistance to multiple antibiotics and plant secondary metabolites, was engineered to specifically respond towards resveratrol. A variant with only a single amino acid substitution A38T enabled TtgR to respond to resveratrol and was applied to screen for *p*-coumarate:CoA ligase with improved activity [73]. One caveat is that crosstalk effects might exist when modifying the inducer scopes of biosensors. If the mechanism for ligand binding of an engineered biosensor is not fully understood, the regulator could possibly be induced by effectors with similar structures as the ligand. The crosstalk effect might hinder the application of the modified biosensor. Careful examination of crosstalk effects should be conducted after ligand scope engineering.

Expanding the dynamic range of biosensors could enhance efficiency in biosensor-aided screening or dynamic regulation (Figure 3b). BenM, a prokaryotic transcription factor for *cis, cis*-muconic acid (CCM), was engineered to function in *Saccharomyces cerevisiae* through PCR-based random mutagenesis on the previously identified effector binding domains. The substitutions of H110R, F211V, and Y286N together showed doubled GFP output with the CCM induction. This variant was further applied in screening the best enzyme combinations for the rate-limiting protocatechuic acid decarboxylase in synthesizing CCM [74]. However, the dynamic range of a biosensor is typically affected by both the DNA binding affinity and the activity of the corresponding promoter. A repressor can bind to the specific promoter to inhibit the transcription unless alleviated by its inducer(s). When the binding between the repressor and the promoter is completely eliminated, the promoter will reach its maximum activity, which is independent of the regulator. Dynamic range expansion should cooperate with promoter engineering or RBS engineering as well. Combined with promoter engineering and directed evolution of a vanillate sensor VanR from *Caulobacter crescentus*, the dynamic range of the regulator was improved 14-fold. The evolved biosensor was applied in screening natural catechol *O*-methyltransferases, and three active O-methyltransferases were successfully identified [75]. Engineering biosensors rapidly expanded the number of detectable compounds and optimized the responsive signals of these regulators. Biosynthesis of natural products often encounters complex conditions and thus the applicable biosensors need to meet some strict criteria, such as little or no crosstalk

effect, sensitivity to inducers (some natural products only can be produced to mg/L concentrations) and an obvious on-off switch (appropriate dynamic range). Future research for engineering sensor-regulators in synthesizing natural products should prioritize these directions.

Diversifying biosynthesis of natural products through protein engineering

Besides improving enzyme activity for boosting production yields, the range of natural products can also be diversified using enzyme promiscuity or novel enzymes [76–79]. However, some desired reactions have yet been found to be enzymatically catalyzed [80]. Expanding the substrate scope of natural enzymes through protein engineering enables creating noncanonical biochemistry, and therefore generating non-natural pathways for natural products and their high-value analogs in microbial systems (Figure 4).

New natural products can be obtained from engineered enzymes with broad substrate specificity. A typical example is the engineering of an endogenous thioesterase TesA in *E. coli*. A ‘leaderless’ mutant (with the truncated N-terminal periplasmic signal peptide, which is also referred as leader peptide) of TesA, designated as ‘TesA, showed expanded chain length preference from long-chain (C14-C18) acyl-ACP to medium-chain (C8-C12) acyl-ACP [81, 82]. Through further structure-guided mutagenesis, ‘TesA could be engineered with enhanced activities and substrate specificities towards C12 or C8 substrates [83]. The donor selectivity and prenylation specificity of an aromatic prenyltransferase AtaPT from *Aspergillus terreus* (*A.t*) were modified through site-directed mutagenesis at the substrate binding pocket to enable the utilization of a series of substrates, such as genistein, sophoricoside, and 7-hydroxycoumarin (7-HC), to yield novel prenylated products [84]. The reaction mechanism of the Fe^{II}/α-ketoglutarate-dependent dioxygenase AsqJ from *Aspergillus nidulans* was discovered by implementing computational-guided molecular simulation. AqsJ natively catalyzes the desaturation and epoxidation reaction of 4’-methoxycyclopeptin to 4’-methoxyviridicatin, a quinolone alkaloid with significant biomedical application. The engineered AsqJ V72I mutant can use a non-methylated surrogate, which also has great potential in the biopharmaceutical industry [85].

One of the challenges for modifying enzymes to create noncanonical biochemistry is that there is limited information available to guide the engineering. Generally, some insights can be obtained from available enzymes with similar substrates/products or reaction mechanisms, but sometimes it is still hard to find a related enzymatic reaction or similar substrates/products, let alone deciding whether the engineering based on these pieces of information will yield desired variants. Nevertheless, protein engineering has still enabled enormous catalytic improvement to yield novel catalytic powers. These activities have not only broadened the scope of biosynthesized natural compounds and their analogs, but also can be applied to diversify the biosynthesis pathways through engineering for existing natural products. For example, hydroxytyrosol is a valuable natural phenolic compound with potent antioxidant activity and various pharmaceutical properties that has been previously engineered in *E. coli* [86]. Recently, a flavin-dependent monooxygenase HpaBC was engineered using directed divergent evolution strategy. A variant with three consecutive substitutions (S210T, A211M, Q212G) exhibited dual tyrosol and tyramine hydroxylase

functionality, resulting in a hydroxytyrosol titer of 1.89 g/L and a yield of 82%. Further molecular analysis suggested that the introduction of M211 and G212 possibly increased the flexibility of the loop, allowing the entrance of either tyrosol or tyramine [87]. More interestingly, changing the microenvironment near active sites or binding pockets, instead of altering the key functional residues directly, could also possibly help to expand substrate specificity. Site-directed mutations were carried out on S247 of cytochrome oxidase CYP199A2 based on its crystal structure. A variant (S247D) was identified to have novel substrate specificity towards *p*-methylbenzylalcohol and phenol. This mutant can regioselectively oxidized *p*-methylbenzylalcohol, *p*-cresol and phenol to 1,4-benzenedimethanol, *p*-hydroxybenzylalcohol and hydroquinone, respectively, while the wild-type enzyme exhibited no activity for these compounds. Further engineering on site 97 yielded S97D, which can also catalyze *p*-cresol to *p*-hydroxybenzylalcohol [88]. Through protein engineering to expand enzymatic substrate spectra, distinct enzymes that only yields specific products are equipped with new catalytic capacities and could serve to generate novel natural compounds as well as metabolic pathways.

Concluding remarks and future perspectives

The continuous and rapid development of biocatalysis and biotransformation of natural products significantly depends on mining, characterizing and engineering new enzymes. Protein engineering focuses on the essential features of these biocatalysts. It can permanently alter the catalytic and physical characteristics of enzymes and potentially drastically boost the potential of biomanufacturing, paving the way for developing industrially feasible systems for biosynthesis of natural products. As highlighted in this review, these powerful strategies provide versatile toolkits and could serve as guidelines for engineering critical enzymes for natural products biosynthesis.

However, many challenges of engineering enzymes or regulators still remain to be addressed (see Outstanding Questions). Random mutagenesis is a traditional method in protein engineering, but this strategy typically generates huge libraries of variants and often requires highly efficient screening methods. A refined method is directed evolution. Using several iterative rounds of mutagenesis with specific selective pressures or standards, directed evolution can form relatively small libraries and is usually more efficient than random mutagenesis. Besides, biosensor-based selection can easily identify the protein variants with desired properties and thus enable high-throughput screening to further increase the process efficiency. Nonetheless, the analysis or screening methods used in each study are often highly specific to a particular characteristic and therefore are not always transferable to other cases. Therefore, rational protein engineering is an important future trend in protein engineering. It relies on the in-depth knowledge of protein structures or computational simulations, and their catalytic mechanisms [89, 90]. A series of computational tools were developed to rationally engineer proteins, such as homology modeling (for determining potential protein structure) [91], molecular docking (for examine interaction between proteins and substrates) [92], metadynamics (for estimation of free energy and other state functions of a system) [93], and Markov state models (for understanding the conformational dynamics of proteins and for quantitatively analyzing molecular dynamic simulations) [94]. However, these simulation processes are strongly influenced by the input parameters. For

example, the output of homology modeling and calculation results of metadynamics are highly dependent on the selected homologous protein or the input values, respectively. Therefore, computational modeling, even though effective in many situations, cannot always provide precise guidance for enzyme engineering [18]. Even when the protein structure and its reaction mechanism are fully understood, the effect of each amino acid substitution is still challenging to calculate and predict.

The further development of protein engineering still needs assistance from structural biology, biochemistry and computational biology. The crystal structures for complex proteins (e.g. enzyme moieties) or integral membrane proteins are still difficult to obtain, but elucidating these pieces of information may be helpful for discovery of new reaction mechanisms and further understanding of protein machinery. Structural information and detailed reaction mechanisms are required for more precise predictions and rational design in protein engineering. The algorithms for modeling the interaction of ligands and proteins need to be streamlined and optimized to better simulate larger and more sophisticated proteins. Biosensors are now expanded to detect not only specific compounds but also a variety of environmental signals such as pH, temperature and light [57]. Since biosensors are capable of converting intracellular characteristics to easily detectable outputs, they will become increasingly important in effective high-throughput screening. Besides engineering existing proteins, *de novo* protein design is becoming increasingly viable to tackle the challenges in natural products biosynthesis because the structural space explored by natural proteins is limited [95, 96]. Depending on the physical and chemical principles of protein folding and guided by computational molecular simulation, a wide range of structures, such as hydrogen bonding networks and helical protein filaments, can now be designed from scratch with atomic-level accuracy [96–100]. Despite the above promising progress, it is still hard to completely design a novel protein with sophisticated functions, especially with current limited accuracy of simulation function. Nonetheless, computationally predicting the *de novo* protein structures or functions is easier than analyzing the changes in protein structure accompanying sequence changes [101]. *De novo* protein design often has to deal with large scale of structure alterations and this can cause substantial changes in system state functions, while few amino acids substitution may only cause small fluctuations, which is harder to detect and predict [101]. *De novo* protein design provides a new research direction for future protein engineering and potentially could contribute to further improvement and diversification of natural product biosynthesis.

Acknowledgements

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R35GM128620. We also acknowledge the support from the College of Engineering, The University of Georgia, Athens.

Glossary

Allosteric transcription factors (aTFs)
proteins that can regulate the transcription process. They can bind to specific DNA sequences and corresponding ligands to activate or repress the downstream expression. There is usually conformation changes of aTFs in the regulating process

Directed evolution
a method often used in protein engineering that mimics the process of natural selection to drive genes, RNAs or proteins to a pre-defined direction

Dynamic range
The induction fold—the ratio of the maximum singal to the minimum—of a biosensor. It should be distinguished with operational range, which is often referred to the concentration range of ligands in which the sensor shows an induction change

Enzyme chimeras
also called fusion proteins. They are proteins constructed through the combining of two or more separate proteins, resulting in a single or multiple polypeptide with functional properties derived from each of the original proteins

Enzyme promiscuity
the ability of an enzyme to catalyze a side reaction in addition to its main function. Even though enzymes are remarkably specific catalysts, they can often perform side reactions in addition to their main, native catalytic activity

Molecular simulation
a computational method usually used for studying the physical interactions of atoms and molecules. In protein engineering, it is usually referred to simulating the interaction of proteins and ligands

Natural products
a series of compounds that can be found in nature, especially those secondary metabolites usually produced by plants, animals or fungi

Site-saturated mutagenesis
substitution a key residue through mutated DNA coding sequences with 20 natural amino acids

References

1. Koehn FE and Carter GT (2005) The evolving role of natural products in drug discovery. Nature reviews Drug discovery 4 (3), 206. [PubMed: 15729362]
2. Kole PL et al. (2005) Cosmetic potential of herbal extracts.
3. Del Nobile MA et al. (2012) Food applications of natural antimicrobial compounds. Frontiers in microbiology 3, 287. [PubMed: 23060862]
4. Rodrigues T et al. (2016) Counting on natural products for drug design. Nat Chem 8 (6), 531–41. [PubMed: 27219696]
5. O'Connor SE (2015) Engineering of Secondary Metabolism. Annu Rev Genet 49, 71–94. [PubMed: 26393965]
6. Krivoruchko A and Nielsen J (2015) Production of natural products through metabolic engineering of Saccharomyces cerevisiae. Curr Opin Biotechnol 35, 7–15. [PubMed: 25544013]
7. Wang J et al. (2018) Recent advances in microbial production of aromatic natural products and their derivatives. Appl Microbiol Biotechnol 102 (1), 47–61. [PubMed: 29127467]
8. Ferreira R et al. (2018) Redirection of lipid flux toward phospholipids in yeast increases fatty acid turnover and secretion. Proc Natl Acad Sci U S A 115 (6), 1262–1267. [PubMed: 29358378]

9. Luo X et al. (2019) Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. Nature 567 (7746), 123–126. [PubMed: 30814733]
10. Minami H et al. (2008) Microbial production of plant benzylisoquinoline alkaloids. Proceedings of the National Academy of Sciences 105 (21), 7393–7398.
11. Liu X et al. (2018) Engineering yeast for the production of breviscapine by genomic analysis and synthetic biology approaches. Nat Commun 9 (1), 448. [PubMed: 29386648]
12. Mak WS et al. (2015) Integrative genomic mining for enzyme function to enable engineering of a non-natural biosynthetic pathway. Nat Commun 6, 10005. [PubMed: 26598135]
13. Colin PY et al. (2015) Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. Nat Commun 6, 10008. [PubMed: 26639611]
14. Sévin DC et al. (2017) Nontargeted in vitro metabolomics for high-throughput identification of novel enzymes in Escherichia coli. Nature Methods 14 (2), 187–194. [PubMed: 27941785]
15. DeLoache WC et al. (2015) An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. Nat Chem Biol 11 (7), 465–71. [PubMed: 25984720]
16. Chen H et al. (2018) Enhancement of the catalytic activity of Isopentenyl diphosphate isomerase (IDI) from Saccharomyces cerevisiae through random and site-directed mutagenesis. Microb Cell Fact 17 (1), 65. [PubMed: 29712558]
17. Verma R et al. (2012) Computer-Aided Protein Directed Evolution: a Review of Web Servers, Databases and other Computational Tools for Protein Engineering. Computational and structural biotechnology journal 2, e201209008–e201209008. [PubMed: 24688649]
18. Kuhlman B and Bradley P (2019) Advances in protein structure prediction and design. Nat Rev Mol Cell Biol.
19. Zhuang Y et al. (2017) Biosynthesis of plant-derived ginsenoside Rh2 in yeast via repurposing a key promiscuous microbial enzyme. Metab Eng 42, 25–32. [PubMed: 28479190]
20. Edgar S et al. (2017) Engineering of Taxadiene Synthase for Improved Selectivity and Yield of a Key Taxol Biosynthetic Intermediate. ACS Synth Biol 6 (2), 201–205. [PubMed: 27794603]
21. Valliere MA et al. (2019) A cell-free platform for the prenylation of natural products and application to cannabinoid production. Nat Commun 10 (1), 565. [PubMed: 30718485]
22. Thodey K et al. (2014) A microbial biomanufacturing platform for natural and semisynthetic opioids. Nat Chem Biol 10 (10), 837–44. [PubMed: 25151135]
23. Chen R et al. (2014) Biomolecular scaffolds for enhanced signaling and catalytic efficiency. Current opinion in biotechnology 28, 59–68. [PubMed: 24832076]
24. Li M et al. (2019) Recent advances of metabolic engineering strategies in natural isoprenoid production using cell factories. Natural product reports.
25. Zerbe P and Bohlmann J (2015) Plant diterpene synthases: exploring modularity and metabolic diversity for bioengineering. Trends in biotechnology 33 (7), 419–428. [PubMed: 26003209]
26. Ignea C et al. (2015) Efficient diterpene production in yeast by engineering Erg20p into a geranylgeranyl diphosphate synthase. Metabolic engineering 27, 65–75. [PubMed: 25446975]
27. Ohto C et al. (2010) Production of geranylgeraniol on overexpression of a prenyl diphosphate synthase fusion gene in Saccharomyces cerevisiae. Applied microbiology and biotechnology 87 (4), 1327–1334. [PubMed: 20393702]
28. Zhou YJ et al. (2012) Modular pathway engineering of diterpenoid synthases and the mevalonic acid pathway for miltiradiene production. Journal of the American Chemical Society 134 (6), 3234–3241. [PubMed: 22280121]
29. Nogueira M et al. (2019) Construction of a fusion enzyme for astaxanthin formation and its characterisation in microbial and plant hosts: A new tool for engineering ketocarotenoids. Metabolic engineering 52, 243–252. [PubMed: 30578862]
30. Rabeharindranto H et al. (2019) Enzyme-fusion strategies for redirecting and improving carotenoid synthesis in S. cerevisiae. Metabolic engineering communications 8, e00086. [PubMed: 30723675]
31. Girvan HM and Munro AW (2016) Applications of microbial cytochrome P450 enzymes in biotechnology and synthetic biology. Current opinion in chemical biology 31, 136–145. [PubMed: 27015292]

32. McLean KJ et al. (2015) Microbial cytochromes P450 In Cytochrome P450, pp. 261–407, Springer.
33. Ciaramella A et al. (2017) Catalytically self-sufficient cytochromes P450 for green production of fine chemicals. Rendiconti Lincei 28 (1), 169–181.
34. Zhang W et al. (2017) Engineering the ribosomal DNA in a megabase synthetic chromosome. Science 355 (6329).
35. Price JV et al. (2016) Scaffoldless engineered enzyme assembly for enhanced methanol utilization. Proceedings of the National Academy of Sciences 113 (45), 12691–12696.
36. Dueber JE et al. (2009) Synthetic protein scaffolds provide modular control over metabolic flux. Nature biotechnology 27 (8), 753.
37. Moon TS et al. (2010) Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered E. coli. Metabolic engineering 12 (3), 298–305. [PubMed: 20117231]
38. Wang Y and Yu O (2012) Synthetic scaffolds increased resveratrol biosynthesis in engineered yeast cells. Journal of biotechnology 157 (1), 258–260. [PubMed: 22100267]
39. Conrado RJ et al. (2011) DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. Nucleic acids research 40 (4), 1879–1889. [PubMed: 22021385]
40. Zhu L. y. et al. (2016) Spatial organization of heterologous metabolic system in vivo based on TALE. Scientific reports 6, 26065. [PubMed: 27184291]
41. Delebecque CJ et al. (2011) Organization of intracellular reactions with rationally designed RNA assemblies. Science 333 (6041), 470–474. [PubMed: 21700839]
42. Myhrvold C et al. (2016) Synthetic lipid-containing scaffolds enhance production by colocalizing enzymes. ACS synthetic biology 5 (12), 1396–1403. [PubMed: 27487319]
43. Lee H et al. (2012) Spatial organization of enzymes for metabolic engineering. Metabolic engineering 14 (3), 242–251. [PubMed: 21946160]
44. Yeates TO et al. (2010) Bacterial microcompartment organelles: protein shell structure and evolution. Annual review of biophysics 39, 185–205.
45. Bobik TA et al. (2015) Bacterial microcompartments: widespread prokaryotic organelles for isolation and optimization of metabolic pathways. Molecular microbiology 98 (2), 193–207. [PubMed: 26148529]
46. Lawrence AD et al. (2014) Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. ACS synthetic biology 3 (7), 454–465. [PubMed: 24933391]
47. Zeldes BM et al. (2015) Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals. Frontiers in microbiology 6, 1209. [PubMed: 26594201]
48. Poor CB et al. (2014) Improving the stability and catalyst lifetime of the halogenase RebH by directed evolution. ChemBioChem 15 (9), 1286–1289. [PubMed: 24849696]
49. Vázquez-Figueroa E et al. (2007) Development of a thermostable glucose dehydrogenase by a structure-guided consensus concept. ChemBioChem 8 (18), 2295–2301. [PubMed: 17990264]
50. Vazquez-Figueroa E et al. (2008) Thermostable variants constructed via the structure-guided consensus method also show increased stability in salts solutions and homogeneous aqueous-organic media. Protein engineering, design & selection 21 (11), 673–680.
51. Styles MQ et al. (2017) Characterization of the first naturally thermostable terpene synthases and development of strategies to improve thermostability in this family of enzymes. The FEBS journal 284 (11), 1700–1711. [PubMed: 28371548]
52. Christopher LP et al. (2014) Lignin biodegradation with laccase-mediator systems. Frontiers in Energy Research 2, 12.
53. Scheiblbrandner S et al. (2017) Evolving stability and pH-dependent activity of the high redox potential Botrytis aclada laccase for enzymatic fuel cells. Scientific reports 7 (1), 13688. [PubMed: 29057958]
54. Mahr R and Frunzke J (2016) Transcription factor-based biosensors in biotechnology: current state and future prospects. Appl Microbiol Biotechnol 100 (1), 79–90. [PubMed: 26521244]

55. Rogers JK et al. (2016) Biosensor-based engineering of biosynthetic pathways. Curr Opin Biotechnol 42, 84–91. [PubMed: 26998575]
56. Zhang F and Keasling J (2011) Biosensors and their applications in microbial metabolic engineering. Trends Microbiol 19 (7), 323–9. [PubMed: 21664818]
57. Shen X et al. (2019) Dynamic gene expression engineering as a tool in pathway engineering. Curr Opin Biotechnol 59, 122–129. [PubMed: 31063878]
58. Liu D et al. (2015) Negative feedback regulation of fatty acid production based on a malonyl-CoA sensor-actuator. ACS Synth Biol 4 (2), 132–40. [PubMed: 24377365]
59. Eggeling L et al. (2015) Novel screening methods--biosensors. Curr Opin Biotechnol 35, 30–6. [PubMed: 25578902]
60. Schallmey M et al. (2014) Looking for the pick of the bunch: high-throughput screening of producing microorganisms with biosensors. Curr Opin Biotechnol 26, 148–54. [PubMed: 24480185]
61. Dahl RH et al. (2013) Engineering dynamic pathway regulation using stress-response promoters. Nat Biotechnol 31 (11), 1039–46. [PubMed: 24142050]
62. Ganesh I et al. (2015) Construction of malate-sensing Escherichia coli by introduction of a novel chimeric two-component system. Bioprocess Biosyst Eng 38 (4), 797–804. [PubMed: 25376367]
63. Ganesh I et al. (2013) Engineered fumarate sensing Escherichia coli based on novel chimeric two-component system. J Biotechnol 168 (4), 560–6. [PubMed: 24056083]
64. Liu C et al. (2018) New Intracellular Shikimic Acid Biosensor for Monitoring Shikimate Synthesis in Corynebacterium glutamicum. ACS Synth Biol 7 (2), 591–601. [PubMed: 29087704]
65. Siedler S et al. (2014) Novel biosensors based on flavonoid-responsive transcriptional regulators introduced into Escherichia coli. Metab Eng 21, 2–8. [PubMed: 24188962]
66. Liu CL et al. (2018) A novel DMAPP-responding genetic circuit sensor for high-throughput screening and evolving isoprene synthase. Appl Microbiol Biotechnol 102 (3), 1381–1391. [PubMed: 29243081]
67. Libis V et al. (2016) Sensing new chemicals with bacterial transcription factors. Curr Opin Microbiol 33, 105–112. [PubMed: 27472026]
68. Tang S-Y et al. (2008) AraC regulatory protein mutants with altered effector specificity. Journal of the American Chemical Society 130 (15), 5267–5271. [PubMed: 18355019]
69. Tang SY and Cirino PC (2011) Design and application of a mevalonate-responsive regulatory protein. Angew Chem Int Ed Engl 50 (5), 1084–6. [PubMed: 21268200]
70. Tang SY et al. (2013) Screening for enhanced triacetic acid lactone production by recombinant Escherichia coli expressing a designed triacetic acid lactone reporter. J Am Chem Soc 135 (27), 10099–103. [PubMed: 23786422]
71. Chen W et al. (2015) Design of an ectoine-responsive AraC mutant and its application in metabolic engineering of ectoine biosynthesis. Metab Eng 30, 149–155. [PubMed: 26051748]
72. Taylor ND et al. (2016) Engineering an allosteric transcription factor to respond to new ligands. Nat Methods 13 (2), 177–83. [PubMed: 26689263]
73. Xiong D et al. (2017) Improving key enzyme activity in phenylpropanoid pathway with a designed biosensor. Metab Eng 40, 115–123. [PubMed: 28111248]
74. Skjoedt ML et al. (2016) Engineering prokaryotic transcriptional activators as metabolite biosensors in yeast. Nature chemical biology 12 (11), 951. [PubMed: 27642864]
75. Kunjapur AM and Prather KLJ (2019) Development of a Vanillate Biosensor for the Vanillin Biosynthesis Pathway in E. coli. ACS Synth Biol.
76. Drummond L et al. (2019) Expanding the Isoprenoid Building Block Repertoire with an IPP Methyltransferase from Streptomyces monomycini. ACS Synth Biol 8 (6), 1303–1313. [PubMed: 31059642]
77. Ignea C et al. (2018) Synthesis of 11-carbon terpenoids in yeast using protein and metabolic engineering. Nat Chem Biol 14 (12), 1090–1098. [PubMed: 30429605]
78. Eiben CB et al. (2019) Mevalonate Pathway Promiscuity Enables Noncanonical Terpene Production. ACS Synth Biol.

79. Chen Z et al. (2017) Metabolic engineering of Escherichia coli for microbial synthesis of monolignols. Metab Eng 39, 102–109. [PubMed: 27816771]
80. Renata H et al. (2015) Expanding the Enzyme Universe: Accessing Non-Natural Reactions by Mechanism-Guided Directed Evolution. Angewandte Chemie International Edition 54 (11), 3351–3367. [PubMed: 25649694]
81. Steen EJ et al. (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature 463 (7280), 559. [PubMed: 20111002]
82. Choi YJ and Lee SY (2013) Microbial production of short-chain alkanes. Nature 502 (7472), 571. [PubMed: 24077097]
83. Grisewood MJ et al. (2017) Computational Redesign of Acyl-ACP Thioesterase with Improved Selectivity toward Medium-Chain-Length Fatty Acids. ACS Catal 7 (6), 3837–3849. [PubMed: 29375928]
84. Chen R et al. (2017) Molecular insights into the enzyme promiscuity of an aromatic prenyltransferase. Nat Chem Biol 13 (2), 226–234. [PubMed: 27992881]
85. Mader SL et al. (2018) Catalytic mechanism and molecular engineering of quinolone biosynthesis in dioxygenase AsqJ. Nat Commun 9 (1), 1168. [PubMed: 29563492]
86. Li X et al. (2018) Establishing an Artificial Pathway for Efficient Biosynthesis of Hydroxytyrosol. ACS Synth Biol 7 (2), 647–654. [PubMed: 29281883]
87. Chen W et al. (2019) Promiscuous enzymatic activity-aided multiple-pathway network design for metabolic flux rearrangement in hydroxytyrosol biosynthesis. Nat Commun 10 (1), 960. [PubMed: 30814511]
88. Furuya T et al. (2015) Alteration of the substrate specificity of cytochrome P450 CYP199A2 by site-directed mutagenesis. J Biosci Bioeng 119 (1), 47–51. [PubMed: 24982017]
89. Bornscheuer UT and Pohl M (2001) Improved biocatalysts by directed evolution and rational protein design. Current opinion in chemical biology 5 (2), 137–143. [PubMed: 11282339]
90. Lutz S (2010) Beyond directed evolution—semi-rational protein engineering and design. Current opinion in biotechnology 21 (6), 734–743. [PubMed: 20869867]
91. Schwede T et al. (2003) SWISS-MODEL: an automated protein homology-modeling server. Nucleic acids research 31 (13), 3381–3385. [PubMed: 12824332]
92. Morris GM and Lim-Wilby M (2008) Molecular docking In Molecular modeling of proteins, pp. 365–382, Springer.
93. Gervasio FL et al. (2005) Flexible docking in solution using metadynamics. Journal of the American Chemical Society 127 (8), 2600–2607. [PubMed: 15725015]
94. Bowman GR et al. (2013) An introduction to Markov state models and their application to long timescale molecular simulation, Springer Science & Business Media.
95. Dahiyat BI and Mayo SL (1997) De novo protein design: fully automated sequence selection. Science 278 (5335), 82–87. [PubMed: 9311930]
96. Huang P-S et al. (2016) The coming of age of de novo protein design. Nature 537 (7620), 320. [PubMed: 27629638]
97. Woolfson DN et al. (2015) De novo protein design: how do we expand into the universe of possible protein structures? Current Opinion in Structural Biology 33, 16–26. [PubMed: 26093060]
98. Shen H et al. (2018) De novo design of self-assembling helical protein filaments. Science 362 (6415), 705. [PubMed: 30409885]
99. Park K et al. (2015) Control of repeat-protein curvature by computational protein design. Nature structural & molecular biology 22 (2), 167.
100. Brunette T et al. (2015) Exploring the repeat protein universe through computational protein design. Nature 528 (7583), 580. [PubMed: 26675729]
101. Baker D (2019) What has de novo protein design taught us about protein folding and biophysics? Protein Sci 28 (4), 678–683. [PubMed: 30746840]
102. Wu Z et al. (2019) Machine learning-assisted directed protein evolution with combinatorial libraries. Proceedings of the National Academy of Sciences 116 (18), 8852.
103. St-Jacques AD et al. (2019) Computational Design of Multisubstrate Enzyme Specificity. ACS Catalysis 9 (6), 5480–5485.

104. Bonet J et al. (2019) rstoolbox - a Python library for large-scale analysis of computational protein design data and structural bioinformatics. BMC Bioinformatics 20 (1), 240. [PubMed: 31092198]
105. Qi R et al. (2019) Computational and Experimental Studies of Inhibitor Design for Aldolase A. The Journal of Physical Chemistry B 123 (28), 6034–6041. [PubMed: 31268712]
106. Osuna S et al. (2015) Molecular dynamics explorations of active site structure in designed and evolved enzymes. Accounts of chemical research 48 (4), 1080–1089. [PubMed: 25738880]
107. Ugalde SO et al. (2018) Engineering a self-sufficient Mycobacterium tuberculosis CYP130 by gene fusion with the reductase-domain of CYP102A1 from Bacillus megaterium. Journal of inorganic biochemistry 180, 47–53. [PubMed: 29232638]
108. Zuo R et al. (2016) An artificial self-sufficient cytochrome P450 directly nitrates fluorinated tryptophan analogs with a different regio-selectivity. Biotechnology journal 11 (5), 624–632. [PubMed: 26743860]
109. Zuo R et al. (2017) Engineered P450 biocatalysts show improved activity and regio-promiscuity in aromatic nitration. Scientific reports 7 (1), 842. [PubMed: 28405004]
110. Mellor SB et al. (2016) Fusion of ferredoxin and cytochrome P450 enables direct light-driven biosynthesis. ACS chemical biology 11 (7), 1862–1869. [PubMed: 27119279]
111. Winzer T et al. (2015) Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein. Science 349 (6245), 309–312. [PubMed: 26113639]
112. Farrow SC et al. (2015) Stereochemical inversion of (S)-reticuline by a cytochrome P450 fusion in opium poppy. Nature chemical biology 11 (9), 728. [PubMed: 26147354]
113. Chaves JE et al. (2017) Engineering isoprene synthase expression and activity in cyanobacteria. ACS synthetic biology 6 (12), 2281–2292. [PubMed: 28858481]
114. Betterle N and Melis A (2018) Heterologous leader sequences in fusion constructs enhance expression of geranyl diphosphate synthase and yield of β-phellandrene production in cyanobacteria (Synechocystis). ACS synthetic biology 7 (3), 912–921. [PubMed: 29397685]

Box 1.

Computational molecular simulation-aided enzyme engineering

In terms of future approaches to protein engineering, computational methods are very attractive as they save resources such as time and money [102]. One challenge in designing enzymes is ensuring that they have the specificity for the desired substrate or substrates. Multichemical state analysis (MCSA) was applied to aid the redesign of the target multisubstrate enzyme [103]. Ultimately this methodology could be used to create biocatalysts with broader specificity. A further challenge with computational approaches to protein engineering is that there needs to be a comprehensive large-scale library for protein structures and sequences. Bonet and colleagues present rstoolbox, which is ideal for not only software developers but also users. It allows users to compare various enzyme design protocols [104]. Additionally, molecular dynamics simulation is a valuable tool when evaluating both structural and energy-based parameters for protein-inhibitor/enzyme-substrate systems. This provides insight into mechanistic features of how the designed inhibitor interacts with its environment [105, 106]. Finally, one of the most recent approaches used to design enzymes is machine learning. As presented by Wu et al., machine learning can be used to more efficiently directly mutate the protein of interest to ultimately create and screen a library of variants. This library of mutants is then tested experimentally based on their fitness scores [102].

Box 2.

Other functions of fusion enzymes

Natural or synthetic fusion P450-reductase enzymes have been demonstrated with higher catalytic performance than individual enzymes. The P450 BM3 (P450 BM3; CYP102A1) from *Bacillus megaterium* is a natural fusion enzyme of a heme and reductase-domain and catalyzes the oxidation of arachidonic acid with a high turnover number, mainly attributed to the self-sufficient electron transfer from the reductase-domain and the covalently linked heme-domain [31]. Mimicking the natural design, a catalytically active fusion protein of CYP130 from *Mycobacterium tuberculosis* and the NADPH reductase domain of P450 BM3 from *B. megaterium* was created with improved dextromethorphan N-demethylation activity [107]. Similarly, fusion of cytochrome P450 TxtE with the reductase domains of P450 BM3 created active **enzyme chimeras** with improved nitration activity towards tryptophan analogs [108, 109]. Interestingly, the fusion of ferredoxin with the P450 CYP79A1 from *Sorghum bicolor* rendered CYP79A1 to obtain electrons for catalysis by interacting directly with the photosystem I, enabling direct light-driven biosynthesis of the *p*-hydroxyphenylacetaldoxime [110]. However, not all oxidoreductase module of P450 fusion enzymes function as redox partners. The natural CYP82Y2-oxidoreductase fusion from opium poppy sequentially catalyzes the two-step epimerization of (S)- to (R)-reticuline via 1,2-dehydroreticuline [111, 112]. To improve enzyme performance, a minor and underestimated approach is to create fusion enzymes that increase the expression of rate-limiting enzymes. This approach employs a highly expressed proteins or protein fragments to augment enzyme expression, which has been especially widely exploited in cyanobacteria. Fusion of CpcB, the highly expressed β-subunit of phycocyanin in cyanobacteria, with isoprene synthase (IspS) improved IspS expression by 254-fold and Isoprene yield by 27-fold [113]. Fusion of the highly expressed kanamycin resistant protein NptI to a heterologous geranyl-diphosphate synthase (GPPS) not only promoted GPPS expression in cyanobacteria, but also conferred a selectable marker for the screening of transformants [114].

Outstanding Questions box

How to improve the accuracy when predicting the effect of a site-directed mutation on enzyme characteristics such as catalytic activity, substrate binding affinity, and stability, if the crystal structure of the target protein is not available?

Engineering enzyme chimeras will cause rigid protein formations, drastically decreasing the flexibility and potential for further engineering. How can we increase the flexibility of enzyme chimeras?

How can we effectively involve unnatural amino acids in protein engineering? How will unnatural amino acids affect the protein engineering output?

Amino acid substitution sometimes can lead to global effects on protein characteristics. For example, by changing the key residues, the efficiency of the enzyme can be improved, but meanwhile its specificity or stability could also be influenced. How to engineer target proteins with minimum or no unfavorable effects on other properties of proteins?

Engineered CRISPR-Cas9 with improved cutting efficiency, lowered off-target effects, and different PAM-recognition sequences has reached single base pair precision. How could optimized CRISPR-Cas9 be implemented in future engineering of biosynthesis of natural products?

How to minimize or eliminate undesirable crosstalk effects when expanding substrate specificity?

Transcription regulators usually contains two kinds of binding domains: one for DNA binding and one for ligand binding. Changing either one of the binding domains may affect both the sensitivity and dynamic range. Moreover, engineering the ligand binding domain could potentially change the ligand preference of the biosensor. How can we orthogonally engineer the sensitivity and dynamic range?

Highlights

In engineering biosynthesis of natural products, protein engineering is paramount for modifying characteristics of enzymes or genetically encoded biosensors.

Protein engineering has improved the biosynthesis of natural products through enhancement of enzymatic activities, colocalization of enzyme complexes, improvement of protein stabilities, and engineering of sensor-regulators for better screening or dynamic regulations.

Engineering existing proteins will yield variants with novel catalytic functions. These advances can expand the spectra of products and thus diversify the biosynthesis of natural products.

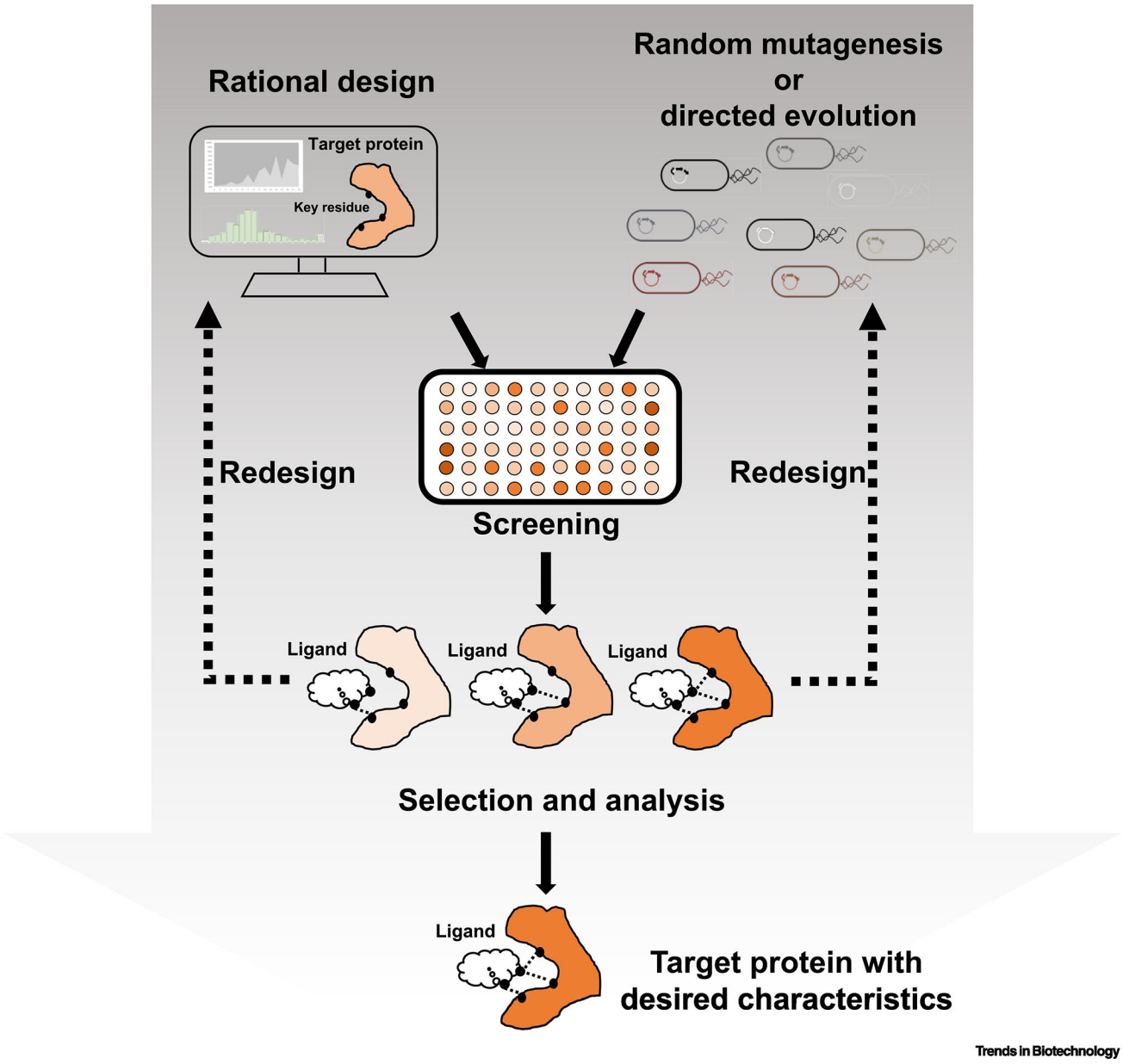

Figure 1. General protein engineering processes.

The library can be established through either rational design aided by computational simulation and/or crystal structures, or random mutagenesis or directed evolution. With the assist of efficient screening methods, variants with different characteristics, such as enhanced catalytic activity and increased stability towards heat, acid or alkaline, can be selected and analyzed. If no candidate shows the desired feature, it may be necessary to regenerate the library and repeat the process until the desired variant is obtained.

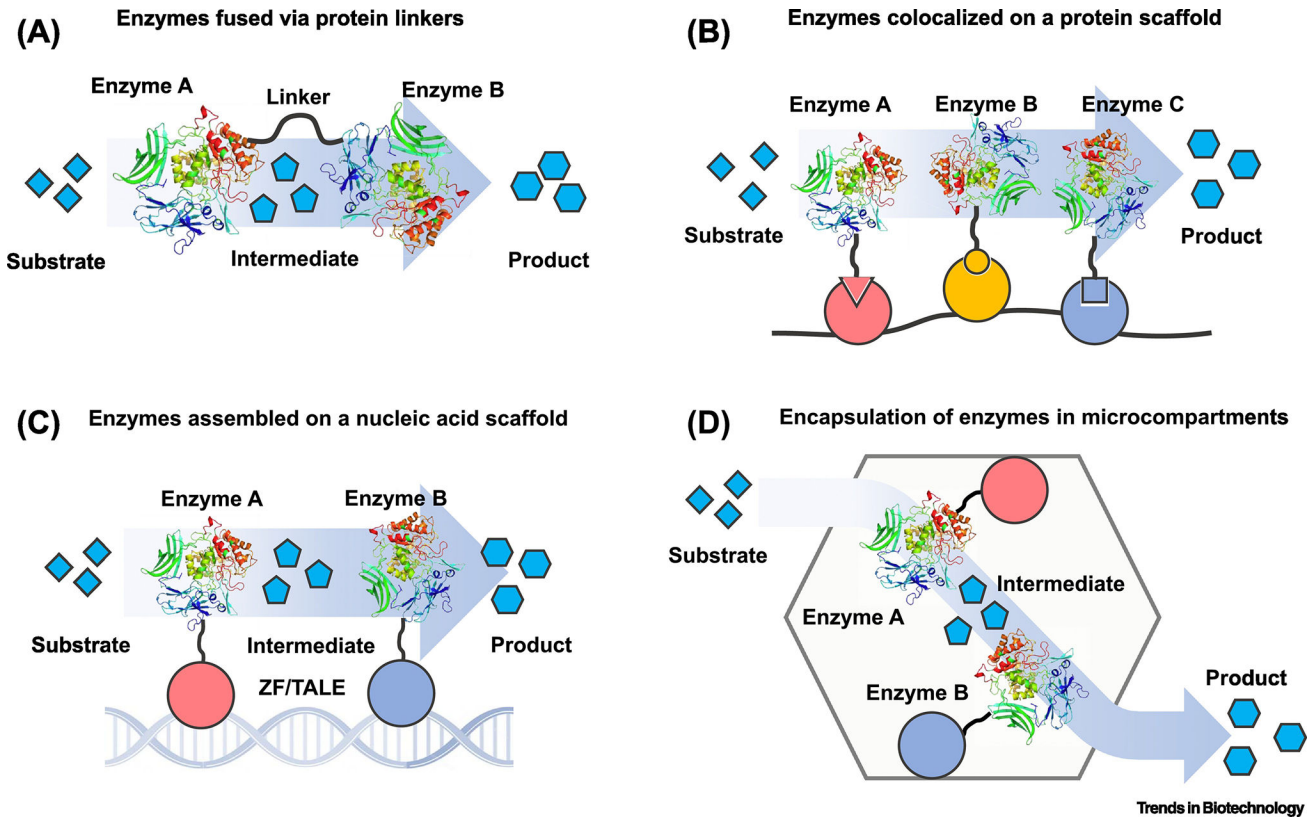

Figure 2. Typical strategies for protein colocalization.

a) Enzymes fused via protein linkers, generating a “substrate channel”. **b)** Enzymes colocalized on a protein scaffold. **c)** Enzymes assembled on a nucleic acid scaffold. For DNA scaffolds, enzymes were fused to zinc finger (ZF) or TALE domains that bind DNA; for RNA scaffold, enzymes were fused to protein adaptor that binds RNA aptamer region on the scaffold. **d)** Encapsulation of enzymes in microcompartments, for instance, carboxysomes.

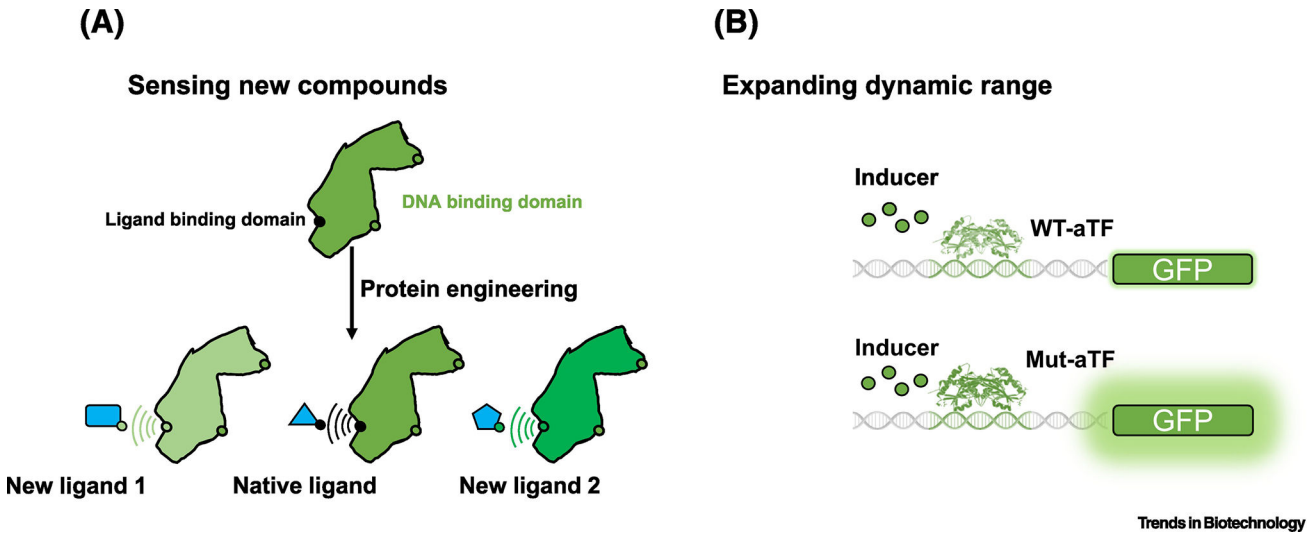

Figure 3. Protein engineering aids the optimization of genetically encoded biosensors.

a) Engineering biosensors for responding to novel inducers. Through protein engineering, variant regulators with novel binding functions can be generated and applied in metabolic engineering of natural products biosynthesis. **b)** Engineering biosensors for expanded dynamic range. The enhanced regulators usually have stronger binding affinities and thus can result in more obvious output.

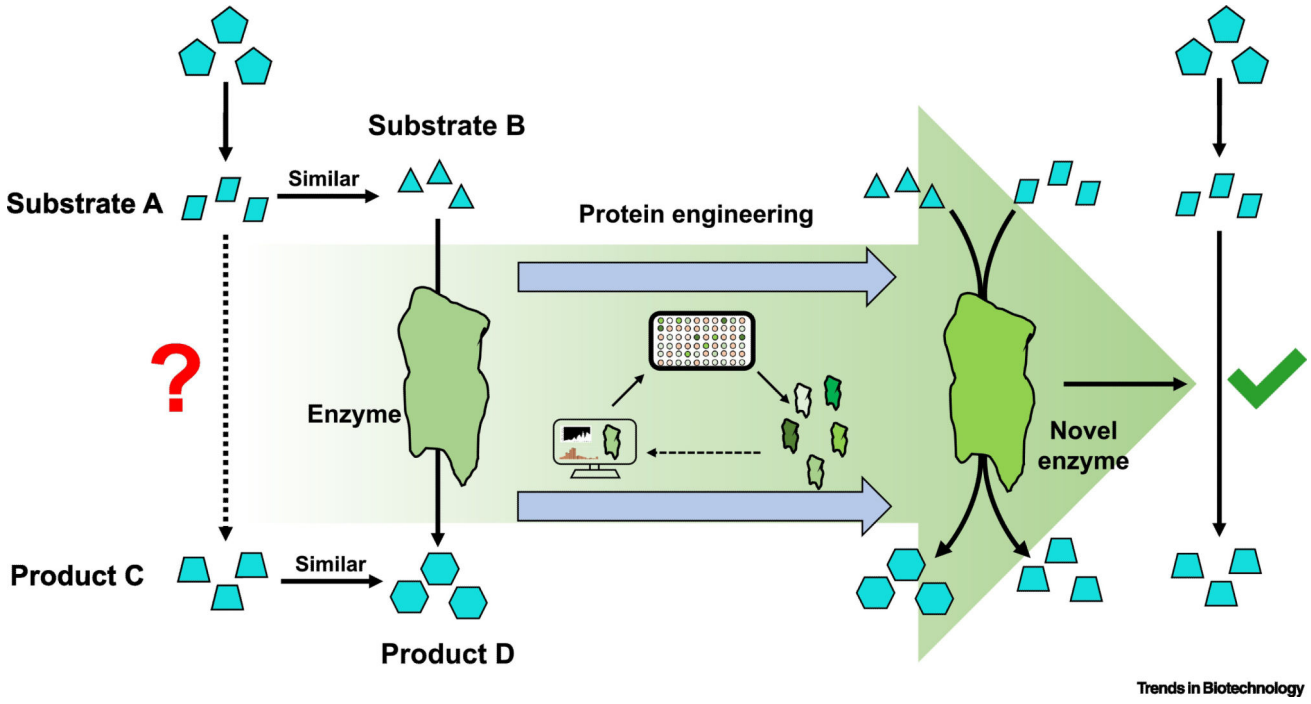

Figure 4. Protein engineering helps to diversify biosynthesis of natural products.

Through engineering existing enzymes, variants with novel catalytic function or expanded substrate spectrum can be obtained. These new enzymes can then be applied in constructing novel pathways for synthesizing natural products or producing analogs of natural products.

Table 1

Examples of protein engineering to improve and diversify natural products biosynthesis

Enzyme	Chassis	Substrate (Ligand)	Product	Mutation(s)*	Biological effects (Results)	Reference
Improving catalytic activity of rate-limiting enzymes						
Tyrosine hydroxylase	*S.c*	Tyrosine	L-DOPA	W13L/ F309L	4.3-fold improvement in catalytic activity	[15]
IDI	*S.c*	IPP	DMAPP	L141H/ Y195F/ W256C	2.53-fold improvement in specific activity	[16]
glycosyltransferase UGT51	*S.c*	Protopanaxadiol	ginsenoside Rh2	S81A/ L82A/ V84A/ E96K/ K92A/ S129A/ N172D	1800-fold increase in catalytic activity	[19]
taxadiene synthase	*E.c*	GGPP	taxadiene	Y688L	2.4-fold improvement in taxadien-5α-ol yield	[20]
Prenyltransferase NphB	N/A	a: OA b: DA	a: CBGA b: CBGVA	M23: G286S/ Y288A; M31: A232S/ Y288V;	Over 1000-fold increase of the k_{cat} value towards CBGA and CBGVA	[21]
Colocalization to form enzyme complexes						
Erg20p: Bts1p & SmCPS: SmKSL	S.c	IPP, DMAPP, GGPP	diterpenes	N/A	365 mg/L miltiradiene	[28]
CrtZ: CrtW	*E.c*	β-carotene	astaxanthin	N/A	1.4-fold improvement in astaxanthin production	[29]
CrtB: CrtI: CrtY	*S.c*	phytoene and ζ-carotene	β-carotene	N/A	Doubled β-carotene production	[30]
Mdh: Hps: Phi	*E.c*	Methanol	F6P	N/A	More than 27-fold enhancement in F6P production	[35]
a: HMGR: HMGS b: AtoB: HMGR: HMGS (linked via CBD)	*E.c*	a: acetoacetyl-CoA b: acetyl-CoA	MVA	N/A	a: 10-fold increase in MVA production b: 77-fold increase in MVA production	[36]
myo-inositol-1-phosphate synthase: myo-inositol oxygenase: urinate dehydrogenase	*E.c*	D-glucose	D-glucaric acid	N/A	5-fold titer increase in glucaric acid production	[37]
p-coumarate-CoA ligase: stilbenesynthase	*S.c*	*p*-coumarate	resveratrol	N/A	5-fold titer increase in resveratrol production	[38]
AtoB: HMGR: HMGS (linked via DNA scaffold)	*E.c*	acetyl-CoA	MVA	N/A	3-fold enhancement in titer	[39]
tryptophan-2-mono-oxygenase: indole-3-acetimide hydrolase	*E.c*	L-tryptophan	IAA	N/A	9.6-fold increase in titer of IAA	[40]
carbonic anhydrase: ribulose-1,5-bisphosphate carboxylase	N/A	CO_2	3-phospho-glyceric acid	N/A	Maintain high local concentration of CO_2 and benefit cell growth	[45]
pyruvate decarboxylase: alcohol dehydrogenase	*E.c*	pyruvate	ethanol	N/A	20-fold increase in specific activity	[46]

Enzyme	Chassis	Substrate (Ligand)	Product	Mutation(s)*	Biological effects (Results)	Reference
Improving protein stability						
tryptophan 7-chlorination halogenase (RebH)	/	tryptophan	7-chloro-L-tryptophan	S2P/ M71V/ K145M/ E423D/ E461G/ S130L/ N166S/ Q494R	18°C higher melting temperature than the wild type	[48]
GDH	*E.c*	glucose	glucono-1,5-lactone	P45A/ F155A/ V227A	improved half-life to 65°C for 3.5 days;	[49, 50]
τ-muurolol synthase	*E.c*	farnesyl diphosphate	τ-muurolol	/	remaining active at up to 78°C	[51]
Laccase	*Botrytis aclada*	Phenols	N/A	T383I	2.6-fold increased half-life at pH 4.0	[53]
Engineering transcriptional regulators-based biosensor						
AraC	*E.c*	L-arabinose	N/A	P8R/T24D/ H80L/ Y82Q; P8G/ F15W/ T24P/ H80A	Respond to D-arabinose	[68]
				P8P/ T24L/ H80L/ Y82L/ H93R	Respond to mevalonate	[69]
				P8V/ T24I/ H80G/ Y82L/ H93R	Respond to triacetic acid lactone	[70]
				P8Y/ T24M/ H80T/ Y82V/ H93V	Respond to ectoine	[71]
LacI	*E.c*	Lactose/Isopropyl β-d-1-thiogalactopyranoside (IPTG)	N/A	Q291T	Respond to fucose	[72]
				H70D/ H74S	Respond to gentiobiose	
				I79T/ R101H	Respond to lactitol	
				D149G/ I160V/ H163Y/ S193E	Respond to sucralose	
TtgR	*P.p*	antibiotics	N/A	A38T	Respond to resveratrol	[73]
BenM	*S.c*	*cis, cis*-muconic acid	N/A	H110R/ F211V/ Y286N	Doubled dynamic range	[74]
VanR	*E.c*	vanillate	N/A	T49I/ A111V/ P179S	14-fold increase in dynamic range	[75]
Diversifying biosynthesis of natural products through protein engineering						
Thioesterase TesA	*E.c*	2,3,4-saturated fatty acyl CoA	2,3,4-saturated fatty acid	N-terminal truncated	Expanded chain length preference from long-chain (C_{14}-C_{18}) acyl-ACP	[81, 82]
				C_{12}: S122K/ Y145K/ L146K C8: M141L/ Y145K/ L146K	Altered chain length preference from long-chain (C_{12} and C_8) acyl-ACP	[83]
aromatic prenyltransferase AtaPT	*A.t*	Aromatic compounds/ isoprenoids	prenylated products	E91A; E91Q	Capable of using genistein, sophoricoside, and 7-HC as substrate	[84]
FeII/α-ketoglutarate-dependent dioxygenase AsqJ	*A.n*	4'-methoxy-cyclopeptin	4'-methoxy-viridicatin	V72I	Capable of using non-methylated surrogate as substrate	[85]
monooxygenase HpaBC	*E.c*	tyrosol	Hydroxy-tyrosol	S210T/ A211M/ Q212G	Exhibiting novel dual tyrosol and tyramine hydroxylase	[87]

Enzyme	Chassis	Substrate (Ligand)	Product	Mutation(s)*	Biological effects (Results)	Reference
cytochrome oxidase CYP199A2	*E.c*	4-methoxy-benzoate	4-hydroxy-benzoate	S97D; S247D	S97D: catalyze p-cresol to p-hydroxy-benzylalcohol; S247D: regioselectively oxidized p-methyl-benzylalcohol to 1,4-benzene-dimethanol	[88]

*The symbol "/" in this column means these mutations are present at the same time. Symbol ";" was used to separate multiple functional variants in a study. "N/A" means no mutation was made in the experiment or the mutation was not specified in the reference.