

REVIEW ARTICLE

<https://doi.org/10.1038/s41467-019-09848-w>

OPEN

Synthetic biology strategies for microbial biosynthesis of plant natural products

Aaron Cravens^{1,3}, James Payne^{1,3} & Christina D. Smolke^{1,2}

Metabolic engineers endeavor to create a bio-based manufacturing industry using microbes to produce fuels, chemicals, and medicines. Plant natural products (PNPs) are historically challenging to produce and are ubiquitous in medicines, flavors, and fragrances. Engineering PNP pathways into new hosts requires finding or modifying a suitable host to accommodate the pathway, planning and implementing a biosynthetic route to the compound, and discovering or engineering enzymes for missing steps. In this review, we describe recent developments in metabolic engineering at the level of host, pathway, and enzyme, and discuss how the field is approaching ever more complex biosynthetic opportunities.

The field of metabolic engineering endeavors to create a green manufacturing industry based on bioproduction of commodity chemicals in cell factories. Plant natural products (PNPs) are especially important targets because of their utility as flavors, fragrances, and medicines, but can be challenging to synthesize due to stereochemical complexity. PNPs are produced via specialized plant metabolism involving numerous enzymes from diverse classes that enzymatically transform central metabolites into secondary metabolite compounds such as the analgesic morphine and antimalarial artemisinin. Many PNPs are obtained from processed plant biomass, requiring substantial land, water, and time investment, and which introduces insecurity in supply chains due to variability in crop yields resulting from pests or extreme weather. Furthermore, intermediates in a PNP biosynthetic pathway are often unavailable from the host plant, thus complicating efforts to produce novel derivatives of the PNP of interest. Microbial production of PNPs can overcome these challenges by enabling (1) on-demand production capabilities associated with microbial cells, (2) scalable and controlled production in fermentation facilities, and (3) the capacity to produce PNPs and PNP intermediates at higher purity or yield than those provided by the native plant host. In addition, microbial production of PNPs can serve as a discovery platform to synthesize novel derivatives of PNPs and gain insight into enzymes involved in plant secondary metabolism.

Metabolic engineering to produce a particular PNP relies on iterative engineering cycles of design, build, and test referred to as the DBT cycle¹. At the level of the host, DBT includes selection and engineering of the host to overproduce PNP-precursor metabolites in sufficient quantity; at the pathway level, a biosynthetic route to produce the PNP is determined and candidate enzymes are tested or discovered; and at the enzyme level, protein engineering may be warranted to improve function or produce derivative compounds (Fig. 1a).

Advances in synthetic biology and enabling technologies like DNA synthesis², sequencing³, and analytical techniques⁴ have accelerated the DBT cycles for metabolic and protein engineering to the point where both can be deployed to engineer the biosynthesis of a particular

¹Department of Bioengineering, Stanford University, 443 Via Ortega, MC 4245, Stanford, CA 94305, USA. ²Chan Zuckerberg Biohub, 499 Illinois St, San Francisco, CA 94158, USA. ³These authors contributed equally: Aaron Cravens, James Payne. Correspondence and requests for materials should be addressed to C.D.S. (email: csmolke@stanford.edu)

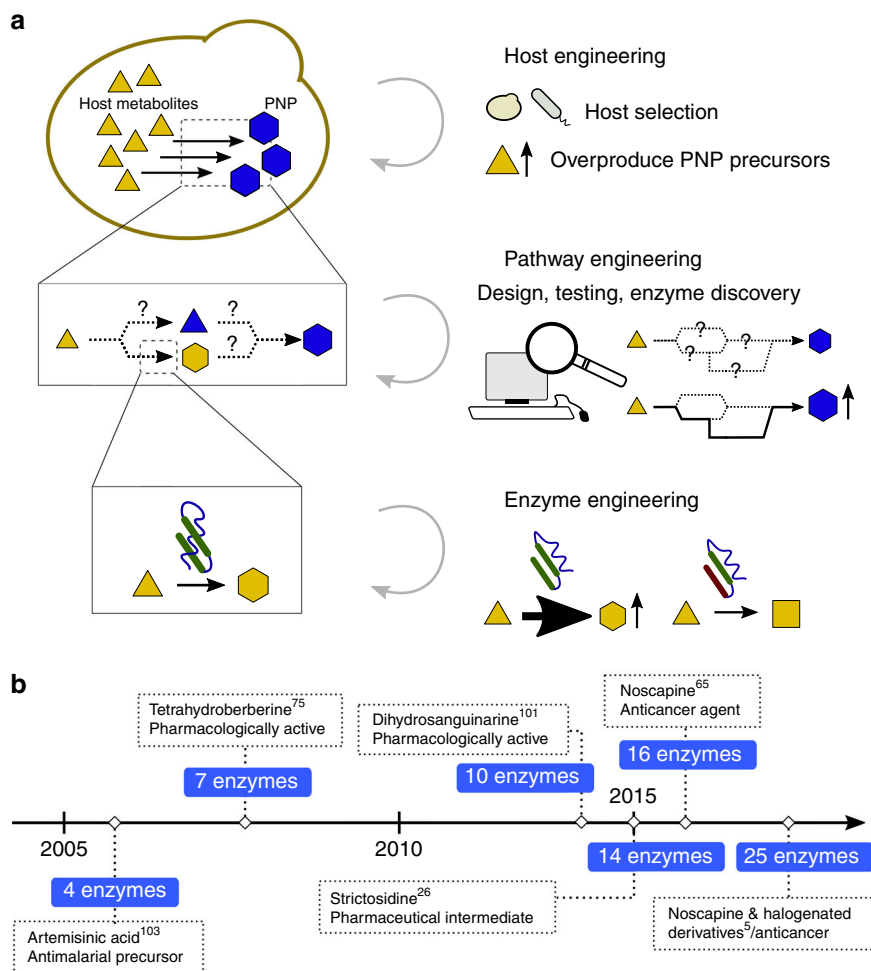


Fig. 1 Metabolic engineering at multiple levels has enabled engineering of increasingly complex heterologous PNP pathways. **a** Heterologous production of a PNP in a microbial host can involve engineering at three different scales: host, pathway, and enzyme. **b** Timeline of examples of notable yeast-produced PNPs, showing increasing pathway length accomplished over time. Labels show product name or compound utility

molecule. Indeed, the complexity of PNP pathways being discovered and engineered has steadily increased over the past 20 years (Fig. 1b), highlighted by a recent example of a 25-enzyme pathway for the anticancer compound noscapine reported in 2018⁵. Once a strain has been developed that produces a small amount of the desired product, strategies to engineer that strain for industrial-scale titers of the product can then be employed; these strategies have been reviewed elsewhere⁶, and the focus of this review will be on initial engineering strategies producing at least detectable concentrations of the desired PNP and/or novel PNP derivative which verify pathway viability. Heterologous PNP biosynthesis and application of DBT require judicious selection and engineering of the production host, the biosynthetic pathway, and the individual enzymes composing the pathway. In this review we discuss recent examples and technologies that enable engineering of hosts, pathways, and enzymes to make PNPs and novel PNP derivatives and the technological advances on the horizon that are expected to further accelerate this field. In the coming years, we expect researchers to increasingly employ metabolic and protein engineering to solve a range of ever more complex biosynthetic challenges.

Identifying and engineering a suitable host organism

A PNP may be selected as a metabolic engineering target for a variety of reasons including medicinal utility, industrial application, or scientific interest. For a given PNP, the first step towards

heterologous production is selection of an appropriate host species in which to engineer the pathway. Within a species, use of previously developed strains that overproduce necessary metabolites can greatly accelerate progress. And lastly, within a given strain, preliminary engineering of the host prior to incorporation of heterologous enzymes can facilitate implementation of the non-native pathway in a new context.

Selecting the host species for a heterologous plant pathway.

When selecting a host species for a heterologous pathway, properties such as ease of cloning, ease of culturing, and suitability of the host for the new enzymes and compounds are considered. Organisms with a long history of use in research, and particularly in metabolic engineering, often have well developed techniques for cloning, culturing, and industrial scale-up that make them attractive choices.

A first choice of host for production of PNPs may be plant cells, where plant specific subcellular compartments and protein processing are conserved, a topic recently reviewed elsewhere⁷. Indeed, model plants such as *Nicotiana benthamiana* are useful for transient expression of plant pathway enzymes during preliminary testing and discovery, as enzyme function, necessary cofactors, and substrate pools are likely to be maintained in planta^{8,9}. However genetic manipulation of plants, even well-established model plants, remains unwieldy and slow compared to microorganisms and thus a microbial host is often preferable.

Microorganisms such as *Escherichia coli* and *Saccharomyces cerevisiae* have a wealth of well-established tools available for genetic manipulation, are easily cultured, have a range of developed platform strains available (see following section, “Selecting a host strain that overproduces PNP precursors”), and are amenable to production scale-up. Other microorganisms are often employed to purposes which evolution has made them especially well-suited: *Streptomyces* is often used for the production of antibiotics originally derived from *Streptomyces* species¹⁰; *Corynebacterium glutamicum* is widely used for the high-titer production of amino acids¹¹; *Yarrowia lipolytica* is frequently employed when using lipids as a substrate¹². Yet for most applications involving the production of PNPs in a microbial host, *E. coli* or *S. cerevisiae* is employed.

Thus, the most immediate question for a metabolic engineer seeking to produce a compound in a heterologous host is often whether to use *E. coli* or *S. cerevisiae*. Distinct advantages of *S. cerevisiae* are its ease of genomic integration, owing to a high rate of homology directed recombination, and that as a eukaryote yeast contains many organelles found in plants. Some enzymes from PNP biosynthesis pathways, such as cytochrome P450s, are transmembrane proteins and require the presence of an appropriate membrane, such as the endoplasmic reticulum (ER), for proper anchoring and folding. This potential roadblock was demonstrated during the Semi-synthetic Artemisinin Project, a landmark achievement in metabolic engineering in which *S. cerevisiae* was engineered to produce high titers of artemisinic acid, a precursor to the important antimalarial artemisinin. In this project, both *E. coli* and *S. cerevisiae* were considered as potential hosts, and while impressive titers of the intermediate amorphadiene (25 g/L) were achieved in *E. coli*¹³, the subsequent step in the pathway is carried out by P450_{AMO}, a plant cytochrome P450. High activity of this enzyme could not be attained in *E. coli* necessitating a switch to production in *S. cerevisiae*¹⁴. While strategies exist to modify transmembrane proteins for function in the cytosol¹⁵, using *S. cerevisiae* as a host for pathways containing transmembrane proteins avoids the added labor necessary to modify those enzymes. Furthermore, *S. cerevisiae* contains cellular microcompartments (e.g., mitochondria and peroxisomes) that can be used to mimic the subcellular localization employed in PNP biosynthesis in plants¹⁶. Conversely, *E. coli* has a doubling time that is 3–4 times shorter than *S. cerevisiae*, is well suited to very high expression of enzymes, and has a different profile of native metabolites available compared to *S. cerevisiae*. For example, the presence of a native pathway for certain isoprenoid compounds was used to engineer *E. coli* strains with 2,400-fold higher production of taxadiene (a precursor to the PNP drug taxol) compared to strains of *S. cerevisiae*¹⁷ engineered for taxadiene production.

One additional avenue when choosing a host organism for PNP biosynthesis is to utilize multiple organisms in a co-culture with components of a metabolic pathway split between distinct organisms of the same or different species^{18–21}. Merits of this approach include reducing burden on the host from the heterologous pathway, the ability to utilize the species most suited to expression of specific enzymes in the pathway, and modularity associated with being able to mix pathways by growing distinct strains together. In one example, benzylisoquinoline alkaloids (BIAs) were synthesized in an *E. coli* and *S. cerevisiae* co-culture system²⁰. *E. coli* were engineered for biosynthesis of the branchpoint intermediate (S)-reticuline, and *S. cerevisiae* strains were engineered to express membrane-bound P450 enzymes that derivatized (S)-reticuline to other PNPs. In another example, high titers of an anthocyanin PNP were achieved by splitting the metabolic burden of the pathway across four *E. coli* strains which were co-cultured¹⁹. Limitations of co-

cultures are pathway specific and include inefficiencies in the transport and/or diffusion of intermediate metabolites between cells in the co-culture and the need to balance growth of multiple hosts in as single culture, which may differ in optimal growth conditions and rates.

Selecting a host strain that overproduces PNP precursors.

Following selection of a host species, engineering the host to increase titers of native metabolites that are biosynthetic precursors to the product of interest can greatly facilitate downstream production of PNP molecules. The core metabolic networks of model organisms are well-characterized and can be used to guide overexpression and knockout modifications for overproduction of central metabolite precursors and to address common challenges (e.g., feedback inhibition or other metabolic regulation). One of the advantages of biosynthesis over chemical synthesis is how readily biosynthetic strains are distributed; once a strain has been engineered to produce a compound, researchers looking to expand on that work in the future need not repeat tedious syntheses of starting material.

Strains of *E. coli* and *S. cerevisiae* that overproduce alkaloids, fatty acids, terpenes, and other valuable compound classes have been engineered (Table 1). Platform strains that overproduce central metabolites or a heterologous secondary metabolite can both be useful: central metabolites, such as geranyl pyrophosphate or amino acids, provide a starting point for the production of potentially thousands of diverse PNP compounds, while secondary metabolites can provide an easy starting point from which to engineer biosynthesis of a specific PNP product. For example, platform strains that produce the key branch point alkaloid (S)-reticuline^{22,23} have enabled microbial biosynthesis of a wide range of BIAs produced by *Papaver somniferum* (opium poppy), including morphine²⁴ and noscapine²⁵. Likewise, stricoidine producing strains²⁶ provide a key branch point metabolite for the biosynthesis of monoindole alkaloids (MIAs), which include vincristine, ibogaine, yohimbine, and thousands of others.

A platform strain can be useful not only because it produces a valuable starting material, but also because the means of production of said starting material are particularly inexpensive, sustainable, or offer easy handling for the researcher or industrial producer. This is demonstrated by the engineering of an efficient simultaneous saccharification and co-fermentation (SSCF) strain for bioethanol production in *E. coli* which utilizes lignocellulosic biomass²⁷, an inexpensive waste product from agriculture and forestry, in place of expensive refined sugars. In another example, an enzyme was designed that allows for assimilation of formate into central metabolism²⁸, potentially allowing the biosynthesis of medicines and commodity chemicals from formate, which is expected to be abundantly available from electrochemical reduction of CO₂. Lastly, researchers generated a strain of *E. coli* that can produce its own biomass from CO₂ via photosynthesis²⁹. Although there is significant interest in utilizing natural photosynthesizing microorganisms (e.g., cyanobacteria) for the production of PNPs³⁰, it could be advantageous to engineer the ability to fix CO₂ in well studied, genetically tractable industrial microorganisms such as *E. coli* and *S. cerevisiae*. While the aforementioned strains have not yet been used for the production of PNPs, the other PNP-producing strains discussed throughout this review could potentially be integrated into these platforms to produce complex PNPs de novo from agricultural waste, formate produced with renewable energy³¹, or directly from atmospheric CO₂. This principle has been demonstrated through the engineering of *E. coli* to utilize the one-carbon feedstock methanol and ultimately convert it into the flavanoid

Table 1 Common platform strains^a

Compound	Titer	Production host	Utility	Reference
Geranylgeranyl pyrophosphate (GGPP)	> 1 g/L	<i>E. coli</i>	Intermediate for diterpene and diterpenoid biosynthesis and precursor to carotenoids, gibberellins, tocopherols	17
Acetyl-CoA & Farnesene	130 g/L	<i>S. cerevisiae</i>	Precursor metabolites to many PNPs including artemisinin acid and anti-malarial artemisinin	33
L-ornithine	> 1 g/L	<i>S. cerevisiae</i>	Precursor metabolite for tropane alkaloids	105
Aromatic amino acids	10 – 100 g/L	<i>E. coli</i>	Precursor to amino acid derived PNPs including alkaloids like reticuline and flavonoids like naringenin	104
Taxadiene	1 g/L	<i>E. coli</i>	Precursor to widely used anticancer drug Taxol (Paclitaxel)	17
Strictosidine	0.5 mg/L	<i>S. cerevisiae</i>	Monoterpene indole alkaloid PNP branchpoint	26
Reticuline	160 mg/L	<i>E. coli</i>	Alkaloid PNP branchpoint compound	102
Naringenin	110 mg/L	<i>E. coli</i>	Flavonoid PNP branchpoint	67
Amorphaniene	> 40g/L	<i>S. cerevisiae</i>	Precursor to anti-malarial artemisinin	103

^aExamples of engineered strains producing different compounds or compound classes that can be used as platform strains for the production of diverse downstream compounds. Yellow, core metabolite platform; blue, secondary metabolite platform

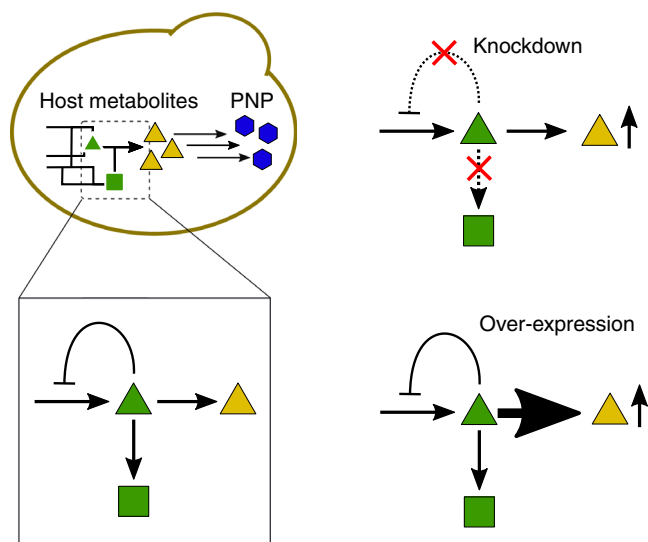


Fig. 2 Common host engineering strategies to increase titer of a PNP precursor compound. Yellow triangles, core metabolite platform; blue hexagons, secondary metabolite platform

naringenin³². Such strategies could support more sustainable bioprocesses for producing increasingly diverse products, including PNPs, at industrial scale.

Engineering host metabolism to facilitate PNP biosynthesis.

After selection of a host or existing platform strain, the supply of biosynthetic precursors may be enhanced by modifications to the host, such as gene deletions, swapping of endogenous enzymes with more active homologues, or overexpression of endogenous metabolic genes (Fig. 2). A recent *tour de force*³³ combined all of these techniques to reprogram yeast central metabolism to overproduce acetyl-CoA for isoprenoid and fatty acid biosynthesis - molecules which are a starting point for many PNPs such as the antimalarial artemisinin. A model of the yeast reaction stoichiometries for acetyl-CoA, redox cofactors, and sugar was used to determine a more favorable reaction stoichiometry, which was

defined as having a reduced ATP requirement, reduced loss of carbon to side reactions, and improved pathway redox balance. The optimal acetyl-CoA stoichiometry was implemented by augmenting acetyl-CoA biosynthesis with expression of four enzymes involved in acetyl-CoA biosynthesis in other organisms, allowing the yeast to produce 25% more of the isoprenoid farnesene with an equal supply of sugar while requiring less oxygen, an important consideration for oxygen-constrained industrial fermentation environments.

Optimization for tyrosine and *p*-coumaric acid overproduction, from which many PNPs including some alkaloids, polyphenols, and flavonoids are derived, has also been pursued in the context of *E. coli* and *S. cerevisiae*. For example, researchers engineered yeast producing 1.9 g/L of *p*-coumaric acid through a combination of six genetic modifications to yeast native metabolism. These included engineering feedback-resistant enzymes, over-expressing enzymes at bottlenecks, and removing competing side pathways³⁴.

Deletion of competing or undesired side pathways in the host is a common strategy to increase precursor titers (Fig. 2). In work on the *de novo* production of strictosidine, a plant-derived alkaloid, researchers monitored biosynthetic intermediates in their engineered pathway to identify competing side pathway²⁶. Finding that geraniol, an intermediate in strictosidine biosynthesis, was metabolized by the yeast through esterification, deletions were made to ATF1 and OYE2 which reduced undesired host interactions and resulted in a 6-fold increase in strictosidine production.

Finally, evolution has emerged as a powerful approach for host optimization, although it has not yet been directly applied to PNP biosynthesis. In the aforementioned work on altering yeast metabolism from alcoholic fermentation to lipogenesis, researchers also employed laboratory evolution methods to improve lipogenic growth on glucose³⁵. Deletion of pyruvate decarboxylase genes (PDC1, 5, and 6) involved in alcoholic fermentation resulted in strains unable to grow on glucose as a carbon source. Adaptive laboratory evolution was applied to evolve FFA producing strains lacking ethanol fermentation for growth on glucose by gradually shifting the carbon source from ethanol to glucose over 200 generations. New methods like SCRaMble have enabled inducible control of host genetic variation³⁶. SCRaMble

utilizes a synthetic yeast chromosome V with recombination sites introduced in all non-essential genes such that when recombination is induced these genes are shuffled within chromosome V. SCRaMble was applied in *S. cerevisiae* and shown to improve host strain background for improved production of violacein, penicillin, and utilization of xylose as a carbon source.

Strategies for planning and engineering a metabolic pathway

Following selection of a suitable host, a route to the desired PNP can be planned and implemented. A candidate pathway is first outlined through selection of stepwise chemical intermediates leading from host metabolism to the target compound, followed by selection of enzymes to carry out each specified reaction. For certain PNPs, detailed knowledge of the native biosynthetic pathway is available and can be used to outline all intermediates and enzymes in a pathway, facilitating pathway engineering into a heterologous host. However, such detailed knowledge can require years or even decades of dedicated research in planta and is frequently unavailable or incomplete. In such cases, candidate pathway design, enzyme selection, and pathway testing all offer distinct challenges which are discussed in the following sections.

Computational tools for global pathway design. Literature on a given PNP biosynthetic route can be instrumental to outlining a pathway, although even for well-studied PNPs there are often gaps in our knowledge. One way to overcome the restriction of needing plant biochemical data for each enzymatic step is to use an approach agnostic to the natural product in question. When a reaction path to a chemical entity is unknown, retrosynthetic analysis can be used such that the target molecule is transformed into simpler precursor structures without making assumptions about starting material availability. Resulting precursors are in turn transformed into simpler structures until available starting constituents are reached. By breaking a target molecule into potential precursors, it is then possible to select enzymes which interconvert in the other direction.

Retrosynthetic pathway design deconstructs a PNP one step at a time and utilizes reaction/enzyme pairs from databases such as MetaCyc³⁷ to propose biosynthetic routes to the target. Of ten available retrosynthesis-based pathway design tools³⁸, only RetroPath³⁹ has been experimentally tested. RetroPath takes starting compounds, a target, and reaction rules to generate potential pathways and was experimentally validated on the design of biosynthetic routes to pinocembrin⁴⁰, a flavonoid four enzymatic steps from *E. coli* central metabolism. RetroPath narrowed down a list of nine million in silico pathways to twelve top-ranked candidates, with one providing 24 mg/L pinocembrin after construction and optimization in *E. coli*. Notably, RetroPath and similar tools such as BNICE.ch²⁴ only consider the type of reaction occurring when considering enzymatic matches. If the substrate of the desired reaction is very different from that of the known reaction to which it is being compared, ranking the results by some measure of substrate similarity, such as Tanimoto distance, might be advantageous⁴¹.

Retrosynthesis can also be performed manually, without the aid of automated tools. To characterize the rapidity of heterologous biosynthesis for the production of valuable compounds, a group of researchers recently performed a pressure test to produce 10 molecules of interest in 90 days⁴². The 10 molecules were a mix of PNPs (carvone and vincristine) and non-PNPs; the fungal metabolite epicolactone provides an example of a retrosynthetic approach that could be applied to PNPs to identify potential pathways. The genomic sequence of the native producer of epicolactone was unavailable and so the researchers based their enzymatic retrosynthesis on a previously developed eight-step

non-enzymatic chemical synthesis⁴³. Enzyme classes were assigned to each reaction manually, guided by literature and pathway databases. Multiple enzymes were identified for each of the eight steps based on reaction type, and to narrow down the candidates, enzyme hits were limited to tropolone-like biosynthetic gene clusters identified from the biosynthetic gene cluster databases MIBiG⁴⁴ and antiSMASH⁴⁵. However, no pathways were experimentally validated within the 90 day time frame.

Nature has developed a limited set of biosynthetic tools; a retrosynthetic scheme might envision chemical transformations which no known enzyme class is able to carry out, and even enzymes which are known to perform the desired chemistry may only do so on a very different substrate. Enzyme evolution to alter substrate scope is still a time-consuming endeavor, and designing enzymes capable of entirely new chemistries has had very limited demonstration thus far⁴⁶. A key question left unanswered for retrosynthetic methods is what strategies can be used for the design of long pathways when some or many steps are non-functional during in vivo testing. This is especially important for the long pathways common in plant secondary metabolism. If automated retrosynthesis tools are to gain more use for PNP biosynthesis, it would be of benefit if not only hypothetical pathways are generated, but also modules containing several enzymatic steps for orthogonal testing, as discussed in the following section “Strategies for the construction of candidate pathways”.

Computational approaches for enzyme candidate discovery.

If the retrosynthetic approach fails to identify functional enzymes for steps in a pathway, enzyme discovery is essential. High-throughput sequencing has enabled efforts to comprehensively profile the genomes and transcriptomes of plant species with important medicinal, industrial, or scientific applications, and these data have fed computational approaches to enzyme discovery, such as plantiSMASH⁴⁷ and the 1KP Project⁴⁸, which leverage genomic information to prioritize biosynthetic gene clusters and enzymes for pathway discovery. Comparison of omics data between species which either produce or lack specific compounds can also help elucidate which enzymes are important for their biosyntheses (Fig. 3).

Demonstrating this approach, two enzymes involved in the biosynthesis of breviscapine flavonoids were discovered entirely via transcriptomic and genomic analysis and subsequently incorporated into an engineered biosynthetic pathway in *S. cerevisiae*⁴⁹. Prior to this work, the enzymes for two steps thought to be catalyzed by a UDP-glycosyltransferase and a P450 were unknown. The researchers identified 83 putative UDP-glycosyltransferases (UDPGTs) from the *Erigeron breviscapus* genome and divided them into 15 gene families. Previous work allowed the researchers to narrow the list from 83 to one likely candidate in the UGT88 family⁵⁰. The function of this lone candidate was validated in vitro and then introduced into a yeast strain producing apigenin, the substrate of UDPGT, resulting in an engineered strain that produced apigenin-7-O-glucuronide. The P450 enzyme was discovered by narrowing down 312 putative P450s in the *E. breviscapus* genome to a list of 134 candidates by comparison with P450s from non-breviscapine producing plant species. Of 134 candidates, 36 were selected and screened for activity using the aforementioned strain producing apigenin-7-O-glucuronide. One P450, CYP706X, resulted in a new peak by HPLC matching the expected product scutellarin, and was subsequently verified via mass spectrometry. This work highlights the ability to identify one enzyme, incorporate it into an engineered strain, and use that new strain to discover enzymes that perform a subsequent reaction, thus leveraging

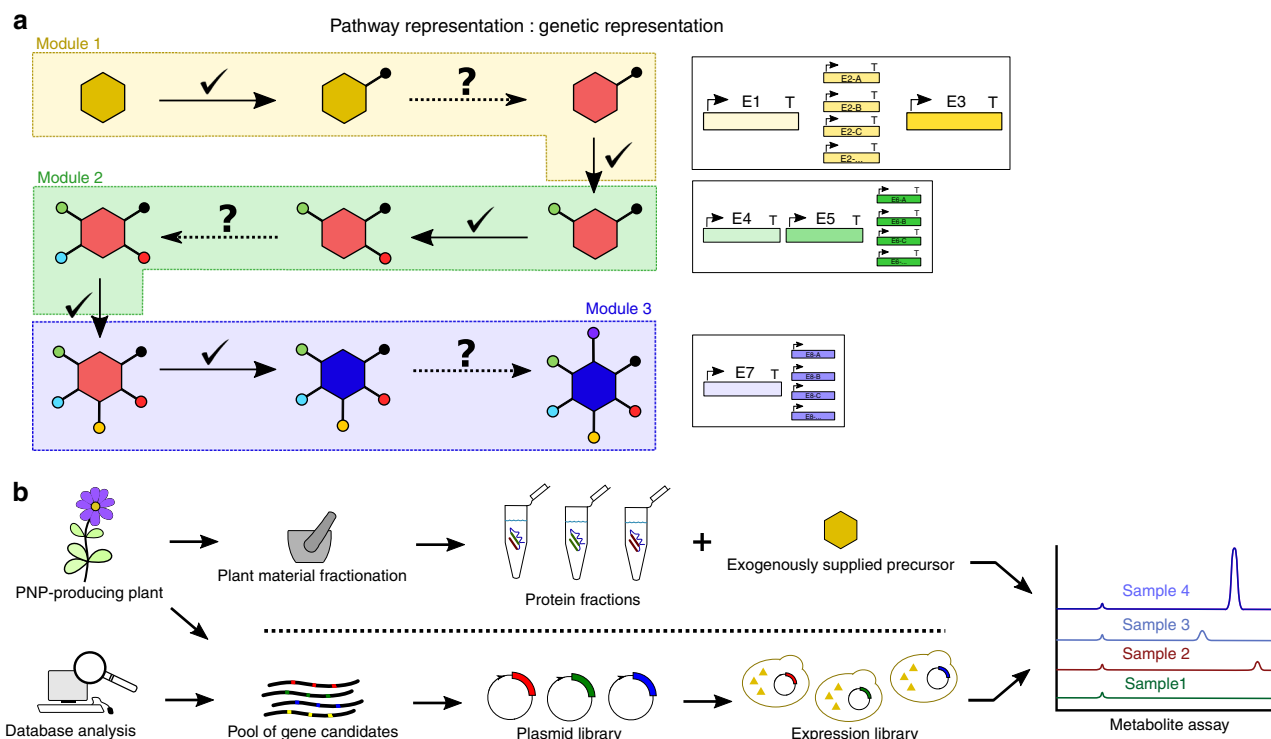


Fig. 3 Pathway engineering can be broken down into enzyme module and discovery components. **a** Example of a modular approach to construction of an eight-enzyme pathway; show at the level of pathway and corresponding genetic modules for testing. Note that each module contains 2–3 enzymes, only one of which is unknown (represented by dotted arrow with question mark), thus distributing critical unknown steps amongst the modules to be tested independently. **b** Strategies for identification of an unknown enzyme in a pathway. Starting either with the native PNP-producing plant or a database of genes containing known or putative enzyme activity, a pool of potential genes can be identified and then cloned into a library of plasmids, which are then used to generate a library of host cells expressing the gene candidates. Gene candidates can then be screened for production of the metabolite of interest. Conversely, in the absence of guiding genetic information, the plant material can be fractionated and separated as best as possible into its constituent proteins, to which the precursor for the desired enzymatic transformation is added, and then screened for production of the metabolite of interest. The latter strategy requires additional steps to determine the sequence of the gene encoding the active protein

intermediate strains developed over the course of a project. As DNA synthesis costs continue to drop, one can envision simply synthesizing and testing an entire panel of candidate genes without needing to computationally prioritize the list beforehand.

A similar strategy was employed to discover a key enzyme required for the biosynthesis of morphine in *Papaver somniferum*. During morphinan alkaloid biosynthesis, (*S*)-reticuline is converted to (*R*)-reticuline by an epimerase. In 2015, three different groups reported the discovery of a two-component epimerase, consisting of a reductase and oxidase, using distinct discovery strategies^{51–53}. The computational and synthetic biology driven approach taken by one team of researchers⁵³ relied on the IKP Project⁴⁸ and PhytoMetaSyn²⁵ databases to search for enzymes similar to a codeinone reductase that had been previously identified. Without accessing plant material, epimerase candidates were synthesized and expressed in an engineered yeast strain producing (*S*)-reticuline, thus affording conversion to (*R*)-reticuline and ultimately enabling de novo biosynthesis of opioids in yeast.

Experimental approaches for enzyme candidate discovery.

Computational approaches to enzyme discovery based on enzyme class, expression, or phylogenetic comparison require a putative enzyme class assignment and/or detailed genomic or transcriptomic data. In the absence of this information, or when the exact nature of the reaction(s) being carried out is unclear, enzyme discovery can be performed with native plant material.

This approach is especially important when a reaction involves unique metabolites and/or a catalytic mechanism that is not well represented in enzyme databases.

One powerful approach to discovery is to isolate an unknown enzyme from native plant material via protein fractionation and functional assay^{54–56}. The active protein fraction can then be identified using protein-mass spectrometry (protein-MS) followed by transcriptomic or genomic mapping (Fig. 3b). Researchers used this approach to discover an enzyme responsible for the ultimate step in the biosynthesis of thebaine, an opiate alkaloid which is converted to the medicinal opiates codeine and morphine in *P. somniferum*⁵⁵. The conversion of (*7S*)-salutaridinol-7-O-acetate (7SOA) to thebaine can occur spontaneously⁵⁷, but the potential role of an unidentified enzyme in *P. somniferum* had been hypothesized⁵⁸. When latex extract from opium poppy was added to 7SOA an increase in thebaine was measured, indicative of enzymatic activity (thebaine synthase, THS). Because this enzyme catalyzed a reaction previously unknown to biocatalysis, a transcriptomic search based on homologous enzymes was not possible. To isolate the enzyme, protein chromatography was used to enrich fractions with THS activity. Six major proteins were present in the THS active fractions as revealed by protein-MS and comparison with predicted translation products of opium poppy. Each candidate gene was expressed in *E. coli* and tested in vitro, but only one (Bet v1-1) displayed THS activity. Subsequently, the THS variant was expressed in an engineered yeast strain, demonstrating an improved biosynthetic route from fed norlaudanosoline to thebaine.

Another team of researchers used a similar method to discover a UDP-glucose:indoxyl glucosyltransferase (UGIG) for *E. coli* based production of indican⁵⁶, a water soluble indigo precursor with potential application for production of sustainable indigo dye. Purification of UGIG from leaves of *Persicaria tinctoria*, one of the highest yielding plants for indican, led to identification of a UGIG gene. The purified UGIG was analyzed via protein-MS and fragments were matched to transcriptome-predicted sequences. *E. coli* was chosen as a production host based on a prior platform for production of the precursor indoxyl, and UGIG expression in indoxyl producing *E. coli* led to accumulation of indican, validating the role of the discovered enzyme.

Putative enzyme function can also be confirmed using virus-induced gene silencing (VIGS) in planta⁵⁴. This approach was used to identify two unknown enzymes in a seven-step pathway from the MIA tabersonine to the anticancer drug precursor vindoline in *Catharanthus roseus*⁵⁹. Researchers used tissue specific qPCR to determine that only the two terminal steps of the vindoline pathway occur outside of the leaf epidermis. To discover earlier genes in the pathway responsible for a net hydration of the substrate, candidate genes suspected to possess hydratase activity and preferentially expressed in leaf epidermis were queried and two candidate genes were selected. VIGS was used to validate the function of a new oxidase (T3O) and reductase (T3R) enzyme in planta and recombinant enzyme assay showed that product formation was only possible via the coupled action of T3O and T3R. The researchers then used the discovered enzymes to complete a seven-gene pathway in yeast producing vindoline from tabersonine, further validating the functions of T3O and T3R and providing a platform for microbial vindoline production. VIGS has also been used to characterize enzyme activities involved in the biosynthesis of noscapine from *P. somniferum*⁶⁰ and the etoposide aglycone from *Podophyllum hexandrum*^{8,60} among many other examples.

Strategies for the construction of candidate pathways. Techniques generally used to construct pathways in heterologous hosts are discussed extensively elsewhere⁶¹; these include discussions of methods for rapid multi-gene integration⁶², gene editing methodologies⁶³, and techniques for combinatorial enzyme expression⁶⁴. However, strategies specifically for the organization and testing of long metabolic pathways (defined here as >5 heterologous genes) have not been clearly defined. Pathway planning and enzyme identification, as described in the preceding sections, are useful for selecting enzyme candidates, but transitioning from an outlined pathway to a functional biosynthetic route expressed in a heterologous host is non-trivial⁴². Challenges include the proper expression of candidate enzymes, which may not be functional when expressed in a heterologous host, and the assembly and validation of multi-enzyme pathways when chemical intermediates are not commercially available.

One approach involves breaking a pathway down into biosynthetic modules, where each module's set of enzymes can be tested and optimized independently in a heterologous host and only combined once validated. Each module ideally begins and concludes with substrates that are commercially available, and steps requiring enzyme discovery are isolated into individual modules, such that a single module is not contingent upon multiple unknown steps (Fig. 3a).

In one example of the utility of clearly defined modules, researchers engineered yeast strains for (S)-reticuline production through the use of four genetic modules containing 17 biosynthetic enzymes⁵³. The modules focused on overproduction of pathway precursors, cofactor recycling enzymes, production of the intermediate (S)-norcoclaurine from native metabolism, and

conversion of (S)-norcoclaurine to (S)-reticuline. This genetic design allowed for independent analysis of each module's role in the pathway and any limitations. In the same work, a fifth module for thebaine biosynthesis was later designed, possessing additional enzymes that were discovered and engineered independently from the first four modules by feeding (S)-reticuline. Ultimately, module five was incorporated into the (S)-reticuline producing strain for fully de novo thebaine biosynthesis. A similar modular strategy was employed for the development of other long biosynthetic pathways in heterologous hosts, including for the production of the alkaloids noscapine⁶⁵, sanguinarine^{66,67}, strictosidine²⁶, and breviscapine flavonoids⁴⁹.

Alongside the pathway specific approach described above, a new frontier for pathway construction is the use of highly automated foundries⁶⁸—collections of wet-lab robotics and software designed to standardize the synthesis, assembly, and testing of DNA parts in microbes. It is unclear if the enzyme discovery components required for some long biosynthetic pathways can be automated, given that discovery methods are frequently tailored to an individual pathway. To date, existing foundry-based approaches have only afforded short pathways (<5 enzymes) or pathways which were already validated^{42,69}.

Enzyme engineering to enable enzyme function in new contexts. When introducing enzymes into a heterologous host, an enzyme may function suboptimally or not at all for reasons that include the new host context (improper folding, post-translational modifications, mislocalization, missing cofactors) or the new chemical context (suboptimal pH, non-natural substrate present, product feedback inhibition). Sub-optimal function of a heterologous enzyme may result in bottlenecks of carbon flux from central metabolism into the PNP pathway. Many of these modes of failure can be alleviated through enzyme engineering.

In plants, localization can cluster pathway enzymes, separate reaction intermediates in the pathway, and provide specific pH or substrate conditions. Localization can be a powerful tool when expressing plant enzymes in microbial contexts for the same reasons. In engineering a heterologous yeast strain for the production of morphine and its semi-synthetic derivatives, researchers observed substantial accumulation of the undesired side product neomorphine⁷⁰. Neomorphine accumulation results from activity of codeinone reductase (COR) on the direct product of T6ODM, neopinone, prior to a spontaneous double bond shift. ER localization tags were fused to COR, thereby sequestering it in the ER and allowing more time for cytosolic neopinone to spontaneously rearrange to codeinone before interacting with COR. The localization strategy ultimately increased morphine titer by sevenfold while decreasing production of the undesired intermediate neomorphine by fourfold. Engineering spatial organization of enzymes can also be accomplished through the use of synthetic scaffolds, which can increase product titers through enzyme clustering or increased local substrate concentration. In one example, researchers constructed a three enzyme pathway in *E. coli* from acetyl-CoA to the intermediate mevalonate, a precursor to the important PNP artemisinin⁷¹. Mevalonate titers were improved by 77-fold to 5 mM by using SHL/SH3 association domains to cluster the three enzymes in the pathway. RNA-based scaffolds have also been applied to cluster enzymes on RNA-scaffolds using RNA-binding domains fused to enzymes⁷².

Another common problem faced when expressing plant enzymes in a new context is host misprocessing of post-translational modifications or signal peptides. In the course of engineering the biosynthesis of opioids in yeast, researchers

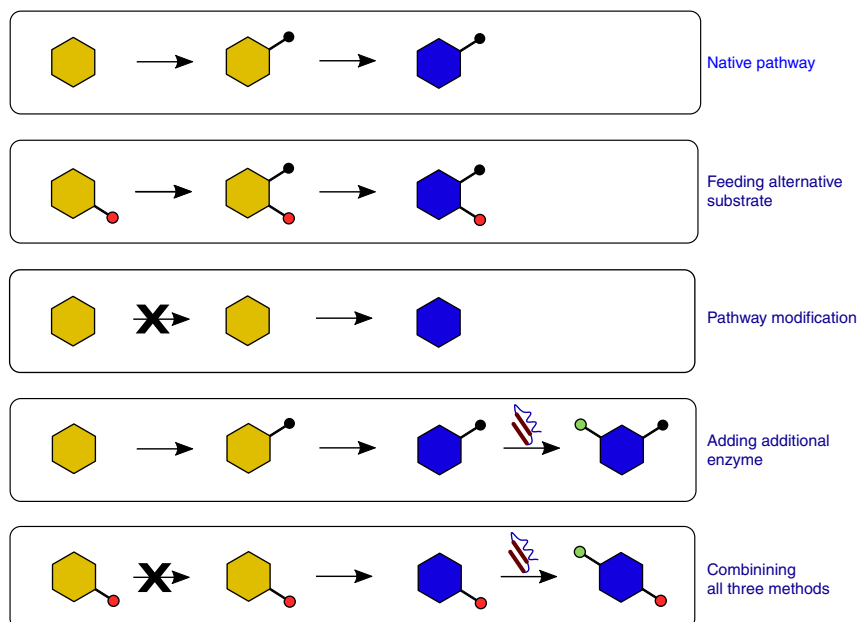


Fig. 4 Potential means of producing novel metabolites once a heterologous pathway to a natural PNP is produced in a microbial host. An example heterologous PNP biosynthetic pathway (top row) can be leveraged to produce novel products in a number of ways: (1) feeding unnatural substrates which are then functionalized by downstream enzymes (second row), (2) removing pathway enzymes to exclude biosynthetic steps (third row), or (3) adding additional enzymes to functionalize intermediates or the final product (fourth row). These three strategies can then be combined in order to afford a wealth of additional products (fifth row)

encountered low activity in the enzyme salutaridine synthetase (SalSyn)⁵³. Western blotting indicated yeast-expressed SalSyn was present as three distinct molecular weights resulting from improper N-linked-glycosylation, indicative of improper localization to the lumen of the ER instead of the ER outer membrane. Protein engineering corrected the improper N-terminal sorting of SalSyn, allowing it to localize to the ER outer membrane and preventing N-linked glycosylation. The engineered enzyme improved conversion of (*R*)-reticuline to salutaridine by sixfold. In another example, researchers engineered brewer's yeast for biosynthesis of aromatic monoterpene molecules (linalool and geraniol) native to the hop plant and important components to the flavor of beer⁷³. In plants, monoterpene biosynthesis occurs in chloroplasts and plant monoterpene synthases typically contain N-terminal plastid targeting sequences (PTSs) of 20–80 amino acids which are cleaved to yield mature protein. In the absence of PTS cleavage, enzyme function is decreased. The researchers tested truncated linalool synthases using bioinformatic and structural information to predict the PTS sites for removal. In one instance, truncation of the PTS motif resulted in a 15-fold improvement in linalool titers. In the same work, additional enzyme engineering was carried out on HMG-CoA reductase, a key rate-limiting step in the pathway for monoterpene biosynthesis. HMG-CoA reductase is controlled by an allosteric domain which responds to product accumulation by inhibiting enzyme function. The researchers truncated the yeast HMGR protein removing an inhibitory domain, thereby increasing flux towards end products.

In addition to the rational modifications discussed, additional enzyme engineering may be required for increasing yield or enabling an enzyme to act on a non-native substrate. The latter might be encountered for enzymes candidates selected through RetroPath or similar computational approaches. Techniques for engineering enzymes with higher activity or for promiscuity toward non-native substrates is a subject covered in the next section and detailed in other recent reviews of enzyme engineering⁷⁴.

Leveraging engineered strains to make novel PNP derivatives

The previous sections have discussed engineering microorganisms to make a desired PNP. Once established in a genetically tractable microbial host, heterologous biosynthetic pathways are an invaluable resource for the synthesis of new-to-nature molecules. While PNPs often possess useful biological activities and accordingly are frequently employed directly as drugs, a higher percentage of drugs are derivatives of PNPs⁷⁵. Derivatization can enhance biological properties of a compound, but PNPs are often synthetically complex, precluding practical syntheses of novel derivatives. With an established heterologous biosynthetic pathway, one can readily replace, add, or remove enzymes or feed in alternative starting materials to synthesize functionalized derivatives. The following sections discuss strategies by which an established PNP-producing engineered microorganism can be leveraged to produce new-to-nature molecules. Each strategy is potentially complementary and in theory can afford a wealth of novel chemical entities from a single starting heterologous pathway (Fig. 4).

Novel PNP derivatives via unnatural substrate feeding.

In addition to the ease of genetic modification of industrial microorganisms, including *E. coli* and *S. cerevisiae*, liquid cultures are easily fed exogenous substrates for incorporation into engineered biosynthetic routes. By extension, derivatives of pathway intermediates can be fed to access derivatives of downstream products. For example, feeding the unnatural intermediate norlaudanosoline to yeast and expressing three methyltransferases resulted in the native intermediate (*S*)-reticuline, demonstrating the flexibility of some enzymes to accept derivatives of their native substrates⁷⁶.

However, substrates which differ more radically from the native substrate are less likely to be accepted at high efficiency, though enzymes differ greatly in their promiscuities. Promiscuity can be assayed *in vitro*, as was done for the BIA biosynthetic enzyme norcoclaurine synthase⁷⁷ and the enzymes for (*S*)-

reticuline epimerization to (*R*)-reticuline⁵¹, or in vivo, as was done for the MIA biosynthetic enzyme strictosidine synthase⁷⁸. In the former cases, the enzyme in question was purified and reacted with derivatives of the native substrate in vitro, thus elucidating which substrates are likely to be accepted in vivo. A cell-free system was developed to assay the ability of prenyltransferases to produce PNPs and novel derivatives and was demonstrated with cannabinoids from *Cannabis sativa*; by feeding in divarinic acid in place of olivetolic acid, cannabinoids that are typically minor products in planta were produced in high titers and prenyltransferase mutants were quickly assayed for substrate selectivity⁷⁹. Promiscuity is also readily probed in vivo directly in the pathway context. Novel isoflavonoids have been produced through feeding of flavanones to engineered yeast⁸⁰, while novel flavonoids and stilbenes were similarly generated from carboxylic acids fed to engineered *E. coli*⁷⁵. In their work on the de novo biosynthesis of noscapine, researchers showed that several halogenated derivatives of the early intermediate tyrosine were accepted by seven downstream enzymes in the pathway, affording halogenated derivatives of pathway intermediates up to (*S*)-reticuline⁵. The substitution of a hydrogen atom for a halogen is relatively sterically conservative, but is a ubiquitous modification in medicinal chemistry, with nearly a quarter of all pharmaceuticals containing at least one halogen⁸¹. However, the titer of the halogenated reticuline derivatives was either too low for halogenated derivatives of further downstream intermediates to be observed, or the subsequent enzyme, berberine bridge enzyme, possesses too narrow a substrate scope. When derivatization occurs at the terminus of a pathway, enzyme promiscuity may not be required. For example, researchers produced novel betalain pigments in yeast by feeding diverse amine scaffolds⁸². A yeast strain was engineered for betalamic acid production, which then spontaneously condensed with the fed primary and secondary amines resulting in new-to-nature pigments.

Novel PNP derivatives via combinatorial biosynthesis. Once a heterologous biosynthetic pathway is established, the enzymes in that pathway and/or analogues of those enzymes can be employed in different combinations to afford distinct products. In this way, no unnatural substrates or novel enzymes need be introduced; all of the necessary tools to make new products are present from the initial engineering effort. For example, in biosynthesis of noscapine from canadine in *S. cerevisiae*, expressing one cytochrome P450, CYP82Y1, in the absence of the preceding enzyme, an N-methyltransferase, afforded 1-hydroxycanadine in place of the usual product, 1-hydroxy-N-methylcanadine⁶⁵. Similarly, swapping CYP82Y1 with CYP82X2, which is downstream in the native biosynthetic pathway, resulted in the production of N-methylphiocarpine, an isomer of the native product, 1-hydroxy-N-methylcanadine. Neither of these two products had previously been identified in the native plant host, *P. somniferum*⁶⁵. The combinatorial space around terpene biosynthetic pathways has been similarly probed using transient expression in *Nicotiana benthamiana* to generate novel sesquiterpenoids derived from the parthenolide biosynthetic pathway from feverfew (*Tanacetum parthenium*)⁸³. This space can be further expanded through the introduction of analogues of the native pathway enzymes. Researchers reconstituted the rebeccamycin biosynthetic pathway from the soil bacterium *Lechevalieria aerocolonigenes* in *Streptomyces albus*, which natively contains RebH, a tryptophan 7-halogenase⁸⁴. By exchanging RebH with pyrH and thal, a tryptophan 5-halogenase and a tryptophan 6-halogenase, respectively, and expressing the other pathway genes in different combinations, a total of 32 different compounds were produced. A similar

strategy could be applied to PNPs using *S. cerevisiae* or *E. coli* as a host to rapidly explore structure space around the native PNP.

Novel PNP derivatives via novel enzyme incorporation. The chemical space accessed by heterologous pathways rebuilt from plants can be further expanded by addition of new enzymes to the pathway. In this way, a natural product can be directly transformed via halogenation, hydroxylation, methylation, prenylation, or any other chemistry available via an enzyme that will accept the natural product as a substrate. Since natural products are often large and complex, an enzyme which natively performs the desired transformation at high efficiency may not be available. However, protein engineering can be used to expand the substrate scope of existing enzymes to accept larger substrates, as has been demonstrated for cytochromes P450⁸⁵, halogenases⁸⁶, and aminotransferases⁸⁷, to name only a few classes. Through similar efforts, type III polyketide synthases have been engineered to produce novel and often larger products by accepting novel substrates and performing additional chain elongation steps prior to cyclization^{88–91}. Given the large effort traditionally required for protein engineering and metabolic engineering, though, relatively few examples in which the two techniques are successfully combined have been reported. In light of advances that have accelerated both and furnished a wealth of engineered enzymes and pathways with which to work, we can expect to see the two utilized in concert increasingly in the future. For example, a recent perspective discussed valuable drugs that could be produced by leveraging opioid biosynthetic *S. cerevisiae* strains with engineered enzymes and subsequent semi-synthesis⁹²; these include the pharmaceuticals cisatracurium, levorphanol, and butorphanol.

To date, metabolic pathways have been more frequently modified with natural enzymes. For example, in work on the combinatorial biosynthesis of the bacterial metabolite rebeccamycin in *E. coli*⁴², researchers identified 21 additional genes known to modify the bisindole core and predicted that combinatorial expression of those enzymes could access to a total of 540 bisindole derivatives, 98% of which have not yet been reported in PubChem. Early concrete demonstrations of addition of derivatizing enzymes to metabolic pathways mostly utilize pathways in their native host, rather than in heterologous hosts. For example, researchers added a halogenase to the marine bacterium *Streptomyces coeruleorubidus*, which natively produces pacidamycin, in order to produce the new-to-nature derivative chloropacidamycin⁹³. The researchers used the introduced chlorine as a synthetic handle for cross-coupling reactions to make a range of novel products. While the two preceding examples are of bacterial natural products, the strategies could readily be applied to PNP pathways expressed in microbial hosts. In a notable achievement, researchers reported the integration of two halogenases into the medicinal plant *Catharanthus roseus* and observed that chlorinated catharanthine alkaloids were produced⁹⁴. Both of the preceding examples rely on the halogenation of tryptophan, an early intermediate in the pathway; in *C. roseus*, 7-chlorotryptophan accumulation was observed in the plant and suspected to adversely affect the growth rate. To alleviate this, researchers engineered RebH to act on tryptamine, the immediate downstream metabolite in the biosynthetic pathway, rather than on tryptophan⁹⁵; integration of this engineered RebH variant showed no accumulation of 7-chlorotryptophan. More recently, researchers engineered *E. coli* to produce resveratrol, a stilbenoid produced by several plants, and then added the halogenase Rdc2 to produce 2-chlororesveratrol⁹⁶. This work utilized a heterologous pathway in an industrial microorganism with an additional enzyme

introduced to produce a new-to-nature natural product derivative, albeit with a relatively simple chemical structure. Given the dramatically more complex PNPs that have been biosynthesized heterologously in recent years, future demonstrations of non-native enzyme incorporation will furnish increasingly synthetically complex novel products.

Conclusion and future directions

The past 20 years of PNP metabolic engineering have seen increasingly sophisticated pathway engineering, with engineered pathways composed of two to seven enzymes in the early 2000s progressing to pathways containing 20 or more enzymes at present. The more enzymatic steps in a heterologous pathway, the more formidable the challenge for construction of the pathway, discovery of the requisite enzymatic components, and overcoming interdependencies introduced between the many enzymatic steps. Taken at face value, this should mean that long, complex PNP biosynthetic pathways take significantly longer to engineer than simple pathways. However, recent years have seen the rapid implementation of long pathways as a result of advances in DNA synthesis, sequencing, and genome engineering. These enabling technologies, along with the emergence of PNP platform strains, have allowed the discovery and engineering of increasingly long PNP pathways.

Following these trends, the field of metabolic engineering leverages sequencing and synthesis to more rapidly discover pathways and enzymes and engineer those pathways into metabolic hosts. Decreasing costs of high-throughput sequencing continue to allow comprehensive profiling of plant genomes and transcriptomes, providing plentiful putative enzyme targets that can be mined via comparison with existing databases of enzymes of known function. Inexpensive DNA synthesis enables wholesale synthesis of dozens of predicted enzymes for any given step in a pathway. This approach has the advantage that hundreds of hypothesized enzymes can be tested for a given step as opposed to testing individual enzymatic hypotheses in planta. Platform strains producing important metabolites for a given specialized PNP pathway will be used to screen single enzymes and combinations of predicted enzymes to reconstitute partial metabolic pathways. Importantly, once a platform strain for a given intermediate is made, the discovery and assembly of downstream pathways are greatly facilitated.

Throughout this review, we have highlighted instances in which protein engineering was employed to solve challenges (e.g., low activity, product inhibition, and poor functioning in new host conditions) encountered during the reconstruction of metabolic pathways in heterologous hosts. As the design-build-test cycles associated with both metabolic engineering and protein engineering have been greatly accelerated in recent years, we expect to see protein engineering employed more frequently during the construction of heterologous metabolic pathways. Although these cycles have accelerated, engineering enzymes for altered small-molecule production still relies on screening, typically via LC or GC, rather than selection, resulting in a disproportionate amount of project time required for sample analysis, even for limited library sizes. One exciting means to engineer selections for small molecules is through the use of genetically-encoded biosensors linking concentration of a compound of interest to an output such as fluorescence-protein expression or cell fitness. Protein and RNA biosensors can be engineered to recognize a range of small molecules and control genetic output accordingly^{97–99}, and early examples have demonstrated their application in enzyme evolution^{23,100,101}. As methods for developing biosensors improve we anticipate that they will be increasingly employed in enzyme engineering to explore wider ranges of sequence space

more rapidly than traditional screening methods allow. Furthermore, we expect protein engineering to be used to solve a wider range of problems in the future—not only to adapt existing enzymes to their new conditions, but also to develop completely novel activities which fill in gaps in biosynthetic pathways or expand pathways in new directions for the production of novel PNP derivatives. This will allow existing and future heterologous biosynthetic pathways to be leveraged for the production of innumerable valuable and novel chemical entities. As metabolic engineering accelerates towards increasingly complex and tailored pathways in the coming years, we expect protein engineering to become an increasingly dominant force for the production of known and novel molecules on both the research and industrial scales.

Received: 19 January 2019 Accepted: 4 April 2019

Published online: 13 May 2019

References

- Nielsen, J. & Keasling, J. D. Engineering cellular metabolism. *Cell* **164**, 1185–1197 (2016).
- Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* **11**, 499 (2014).
- Heather, J. M. & Chain, B. The sequence of sequencers: the history of sequencing DNA. *Genomics* **107**, 1–8 (2016).
- Majors, R. E. Historical developments in HPLC and UHPLC column technology: the past 25 years. *LCGC North Am.* **33**, 818–840 (2015).
- Li, Y. et al. Complete biosynthesis of noscapine and halogenated alkaloids in yeast. *Proc. Natl Acad. Sci. USA* **115**, E3922–E3931 (2018). **This paper is the longest known example of a plant biosynthetic pathway reconstructed in a heterologous host, as well as an example of using PNP-producing platforms for producing unnatural PNPs.**
- Jones, J. A. & Koffas, M. A. G. Optimizing metabolic pathways for the improved production of natural products. *Methods Enzym.* **575**, 179–193 (2016).
- Espinosa-Leal, C. A., Puente-Garza, C. A. & García-Lara, S. In vitro plant tissue culture: means for production of biological active compounds. *Planta* **248**, 1–18 (2018).
- Lau, W. & Sattely, E. S. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science* **349**, 1224–1228 (2015).
- Jeon, J.-E. et al. A pathogen-responsive gene cluster for the production of highly modified fatty acids in tomato. Preprint at <https://doi.org/10.1101/408518> (2018).
- Demain, A. L. Pharmaceutically active secondary metabolites of microorganisms. *Appl. Microbiol. Biotechnol.* **52**, 455–463 (1999).
- Wendisch, V. F., Jorge, J. M. P., Pérez-García, F. & Sgobba, E. Updates on industrial production of amino acids using *Corynebacterium glutamicum*. *World J. Microbiol. Biotechnol.* **32**, 105 (2016).
- Wolf, K. *Nonconventional Yeasts in Biotechnology*. (Springer, Berlin, 1996).
- Tsuruta, H. et al. High-level production of amorpha-4,11-diene, a precursor of the antimalarial agent artemisinin, in *Escherichia coli*. *PLoS ONE* **4**, e4489 (2009).
- Paddon, C. J. & Keasling, J. D. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat. Rev. Microbiol.* **12**, 355–367 (2014).
- Mizrachi, D. et al. A water-soluble DsbB variant that catalyzes disulfide-bond formation *in vivo*. *Nat. Chem. Biol.* **13**, 1022–1028 (2017).
- Hammer, S. K. & Avalos, J. L. Harnessing yeast organelles for metabolic engineering. *Nat. Chem. Biol.* **13**, 823–832 (2017).
- Ajikumar, P. K. et al. Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. *Science* **330**, 70–74 (2010).
- Fang, Z., Jones, J. A., Zhou, J. & Koffas, M. A. G. Engineering *Escherichia coli* co-cultures for production of curcuminoids from glucose. *Biotech. J.* **13**, 1700576 (2018).
- Jones, J. A. et al. Complete biosynthesis of anthocyanins using *E. coli* polycultures. *mBio* **8**, e00621-17 (2017).
- Minami, H. et al. Microbial production of plant benzylisoquinoline alkaloids. *Proc. Natl Acad. Sci. USA* **105**, 7393–7398 (2008).
- Camacho-Zaragoza, J. M. et al. Engineering of a microbial coculture of *Escherichia coli* strains for the biosynthesis of resveratrol. *Microb. Cell Fact.* **15**, 163 (2016).

22. Trenchard, I. J., Siddiqui, M. S., Thodey, K. & Smolke, C. D. De novo production of the key branch point benzyloquinoline alkaloid reticuline in yeast. *Metab. Eng.* **31**, 74–83 (2015).
23. DeLoache, W. C. et al. An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. *Nat. Chem. Biol.* **11**, 465–471 (2015). **This paper is an example of applying biosensor-based screening methods to engineering the production of the key BIA branchpoint alkaloid reticuline.**
24. Hadadi, N., Hafner, J., Shajkofci, A., Zisaki, A. & Hatzimanikatis, V. ATLAS of biochemistry: A repository of all possible biochemical reactions for synthetic biology and metabolic engineering studies. *ACS Synth. Biol.* **5**, 1155–1166 (2016).
25. Xiao, M. et al. Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *J. Biotechnol.* **166**, 122–134 (2013).
26. Brown, S., Clastre, M., Courdavault, V. & O'Connor, S. E. De novo production of the plant-derived alkaloid strictosidine in yeast. *Proc. Natl Acad. Sci. USA* **112**, 3205–3210 (2015).
27. Tai, Y.-S. et al. Engineering nonphosphorylative metabolism to generate lignocellulose-derived products. *Nat. Chem. Biol.* **12**, 247–253 (2016).
28. Siegel, J. B. et al. Computational protein design enables a novel one-carbon assimilation pathway. *Proc. Natl Acad. Sci. USA* **112**, 3704–3709 (2015).
29. Antonovsky, N. et al. Sugar synthesis from CO₂ in *Escherichia coli*. *Cell* **166**, 115–125 (2016).
30. Xue, Y. & He, Q. Cyanobacteria as cell factories to produce plant secondary metabolites. *Front. Bioeng. Biotechnol.* **3**, 57 (2015).
31. Yishai, O., Lindner, S. N., de la Cruz, J. G., Tenenboim, H. & Bar-Even, A. The formate bio-economy. *Curr. Opin. Chem. Biol.* **35**, 1–9 (2016).
32. Whitaker, W. B. et al. Engineering the biological conversion of methanol to specialty chemicals in *Escherichia coli*. *Metab. Eng.* **39**, 49–59 (2017).
33. Meadows, A. L. et al. Rewriting yeast central carbon metabolism for industrial isoprenoid production. *Nature* **537**, 694–697 (2016).
34. Rodriguez, A., Kildegaard, K. R., Li, M., Borodina, I. & Nielsen, J. Establishment of a yeast platform strain for production of *p*-coumaric acid through metabolic engineering of aromatic amino acid biosynthesis. *Metab. Eng.* **31**, 181–188 (2015).
35. Yu, T. et al. Reprogramming yeast metabolism from alcoholic fermentation to lipogenesis. *Cell* **174**, 1549–1558.e14 (2018).
36. Blount, B. A. et al. Rapid host strain improvement by in vivo rearrangement of a synthetic yeast chromosome. *Nat. Commun.* **9**, 1932 (2018). **This paper is the first known example of engineering host metabolism through use of inducible chromosome recombination synthetic biology tools.**
37. Caspi, R. et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* **42**, D459–D471 (2014).
38. Wang, L., Dash, S., Ng, C. Y. & Maranas, C. D. A review of computational tools for design and reconstruction of metabolic pathways. *Synth. Syst. Biotechnol.* **2**, 243–252 (2017).
39. Delépine, B., Duigou, T., Carbonell, P. & Faulon, J.-L. RetroPath2.0: a retrosynthesis workflow for metabolic engineers. *Metab. Eng.* **45**, 158–170 (2018).
40. Fehér, T. et al. Validation of RetroPath, a computer-aided design tool for metabolic pathway engineering. *Biotechnol. J.* **9**, 1446–1457 (2014).
41. Hadadi, N. & Hatzimanikatis, V. Design of computational retrosynthesis tools for the design of de novo synthetic pathways. *Curr. Opin. Chem. Biol.* **28**, 99–104 (2015).
42. Casini, A. et al. A pressure test to make 10 molecules in 90 days: external evaluation of methods to engineer biology. *J. Am. Chem. Soc.* **140**, 4302–4316 (2018).
43. Ellerbrock, P., Armanino, N., Ilg, M. K., Webster, R. & Trauner, D. An eight-step synthesis of epilactone reveals its biosynthetic origin. *Nat. Chem.* **7**, 879–882 (2015).
44. Medema, M. H. et al. Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.* **11**, 625–631 (2015).
45. Blin, K. et al. antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **45**, W36–W41 (2017).
46. Röthlisberger, D. et al. Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).
47. Kautsar, S. A., Suarez Duran, H. G., Blin, K., Osbourn, A. & Medema, M. H. plantSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Res.* **45**, W55–W63 (2017).
48. Matasci, N. et al. Data access for the 1,000 Plants (1KP) project. *Gigascience* **3**, 17 (2014).
49. Liu, X. et al. Engineering yeast for the production of breviscapine by genomic analysis and synthetic biology approaches. *Nat. Commun.* **9**, 448 (2018). **This paper is an example of the de novo biosynthesis of medicinal alkaloids and demonstrates the application of PNP platform strains for enzyme discovery.**
50. Nagashima, S., Hirotani, M. & Yoshikawa, T. Purification and characterization of UDP-glucuronate: baicalein 7-O-glucuronosyltransferase from *Scutellaria baicalensis* Georgi. cell suspension cultures. *Phytochemistry* **53**, 533–538 (2000).
51. Farrow, S. C., Hagel, J. M., Beaudoin, G. A. W., Burns, D. C. & Facchini, P. J. Stereochemical inversion of (S)-reticuline by a cytochrome P450 fusion in opium poppy. *Nat. Chem. Biol.* **11**, 728–732 (2015).
52. Winzer, T. et al. Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein. *Science* **349**, 309–312 (2015).
53. Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M. & Smolke, C. D. Complete biosynthesis of opioids in yeast. *Science* **349**, 1095–1100 (2015). **This paper is the first known example of the complete biosynthesis of opioids in yeast and demonstrates the application of PNP platform strains for enzyme discovery.**
54. Caputi, L. et al. Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle. *Science* **360**, 1235–1239 (2018).
55. Chen, X. et al. A pathogenesis-related 10 protein catalyzes the final step in thebaine biosynthesis. *Nat. Chem. Biol.* **14**, 738–743 (2018).
56. Hsu, T. M. et al. Employing a biochemical protecting group for a sustainable indigo dyeing strategy. *Nat. Chem. Biol.* **14**, 256–261 (2018).
57. Lenz, R. & Zenk, M. H. Acetyl coenzyme A: salutaridinol-7-O-acetyltransferase from papaver somniferum plant cell cultures. The enzyme catalyzing the formation of thebaine in morphine biosynthesis. *J. Biol. Chem.* **270**, 31091–31096 (1995).
58. Barton, D. H. R., Bhakuni, D. S., James, R. & Kirby, G. W. Phenol oxidation and biosynthesis. Part XII. Stereochemical studies related to the biosynthesis of the morphine alkaloids. *J. Chem. Soc. C: Organic* **0**, 128–132 (1967).
59. Qu, Y. et al. Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proc. Natl Acad. Sci. USA* **112**, 6224–6229 (2015).
60. Winzer, T. et al. A Papaver somniferum 10-gene cluster for synthesis of the anticancer alkaloid noscapine. *Science* **336**, 1704–1708 (2012).
61. Luo, Y., Enghiad, B. & Zhao, H. New tools for reconstruction and heterologous expression of natural product biosynthetic gene clusters. *Nat. Prod. Rep.* **33**, 174–182 (2016).
62. Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A Highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth. Biol.* **4**, 975–986 (2015).
63. Ryan, O. W., Poddar, S. & Cate, J. H. D. CRISPR–Cas9 genome engineering in *Saccharomyces cerevisiae* cells. *Cold Spring Harb. Protoc.* **2016**, <https://doi.org/10.1101/pdb.prot086827> (2016).
64. Jeschek, M., Gerngross, D. & Panke, S. Rationally reduced libraries for combinatorial pathway optimization minimizing experimental effort. *Nat. Commun.* **7**, 11163 (2016).
65. Li, Y. & Smolke, C. D. Engineering biosynthesis of the anticancer alkaloid noscapine in yeast. *Nat. Commun.* **7**, 12137 (2016).
66. Trenchard, I. J. & Smolke, C. D. Engineering strategies for the fermentative production of plant alkaloids in yeast. *Metab. Eng.* **30**, 96–104 (2015).
67. Fossati, E. et al. Reconstitution of a 10-gene pathway for synthesis of the plant alkaloid dihydrosanguinarine in *Saccharomyces cerevisiae*. *Nat. Commun.* **5**, 3283 (2014).
68. Chao, R., Mishra, S., Si, T. & Zhao, H. Engineering biological systems using automated biofoundries. *Metab. Eng.* **42**, 98–108 (2017).
69. Carbonell, P. et al. An automated design-build-test-learn pipeline for enhanced microbial production of fine chemicals. *Commun. Biol.* **1**, 66 (2018).
70. Thodey, K., Galanie, S. & Smolke, C. D. A microbial biomanufacturing platform for natural and semisynthetic opioids. *Nat. Chem. Biol.* **10**, 837–844 (2014).
71. Dueber, J. E. et al. Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* **27**, 753–759 (2009).
72. Sachdeva, G., Garg, A., Godding, D., Way, J. C. & Silver, P. A. In vivo colocalization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner. *Nucleic Acids Res.* **42**, 9493–9503 (2014).
73. Denby, C. M. et al. Industrial brewing yeast engineered for the production of primary flavor determinants in hopped beer. *Nat. Commun.* **9**, 965 (2018).
74. Zeymer, C. & Hilvert, D. Directed evolution of protein catalysts. *Annu. Rev. Biochem.* **87**, 131–157 (2018).
75. Katsuyama, Y., Funai, N., Miyahisa, I. & Horinouchi, S. Synthesis of unnatural flavonoids and stilbenes by exploiting the plant biosynthetic pathway in *Escherichia coli*. *Chem. Biol.* **14**, 613–621 (2007).
76. Hawkins, K. M. & Smolke, C. D. Production of benzyloquinoline alkaloids in *Saccharomyces cerevisiae*. *Nat. Chem. Biol.* **4**, 564–573 (2008).
77. Ruff, B. M., Bräse, S. & O'Connor, S. E. Biocatalytic production of tetrahydroisoquinolines. *Tetrahedron Lett.* **53**, 1071–1074 (2012).
78. McCoy, E. & O'Connor, S. E. Directed biosynthesis of alkaloid analogs in the medicinal plant *Catharanthus roseus*. *J. Am. Chem. Soc.* **128**, 14276–14277 (2006).

79. Valliere, M. A. et al. A cell-free platform for the prenylation of natural products and application to cannabinoid production. *Nat. Commun.* **10**, 565 (2019).
80. Chemler, J. A., Lim, C. G., Daiss, J. L. & Koffas, M. A. G. A versatile microbial system for biosynthesis of novel polyphenols with altered estrogen receptor binding activity. *Chem. Biol.* **17**, 392–401 (2010).
81. Herrera-Rodriguez, L. N., Khan, F., Robins, K. T. & Meyer, H.-P. Perspectives on biotechnological halogenation Part I: Halogenated products and enzymatic halogenation. *Chem. Today* **29**, 31–33 (2011).
82. Grewal, P. S., Modavi, C., Russ, Z. N., Harris, N. C. & Dueber, J. E. Bioproduction of a betalain color palette in *Saccharomyces cerevisiae*. *Metab. Eng.* **45**, 180–188 (2018).
83. Kashkooli, A. B., van der Krol, A., Rabe, P., Dickschat, J. S. & Bouwmeester, H. Substrate promiscuity of enzymes from the sesquiterpene biosynthetic pathways from *Artemisia annua* and *Tanacetum parthenium* allows for novel combinatorial sesquiterpene production. *Metab. Eng.* **54**, 12–23 (2019).
84. Sánchez, C. et al. The biosynthetic gene cluster for the antitumor rebeccamyacin: characterization and generation of indolocarbazole derivatives. *Chem. Biol.* **9**, 519–531 (2002).
85. Fasan, R., Chen, M. M., Crook, N. C. & Arnold, F. H. Engineered alkane-hydroxylating cytochrome P450BM3 exhibiting natively catalytic properties. *Angew. Chem. Int. Ed.* **46**, 8414–8418 (2007).
86. Payne, J. T., Poor, C. B. & Lewis, J. C. Directed evolution of RebH for site-selective halogenation of large biologically active molecules. *Angew. Chem. Int. Ed. Engl.* **54**, 4226–4230 (2015).
87. Savile, C. K. et al. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **329**, 305–309 (2010).
88. Morita, H. et al. Synthesis of unnatural alkaloid scaffolds by exploiting plant polyketide synthase. *Proc. Natl Acad. Sci. USA* **108**, 13504–13509 (2011).
89. Wanibuchi, K., Morita, H., Noguchi, H. & Abe, I. Enzymatic formation of an aromatic dodecaketide by engineered plant polyketide synthase. *Bioorg. Med. Chem. Lett.* **21**, 2083–2086 (2011).
90. Bhan, N., Cress, B. F., Linhardt, R. J. & Koffas, M. Expanding the chemical space of polyketides through structure-guided mutagenesis of *Vitis vinifera* stilbene synthase. *Biochimie* **115**, 136–143 (2015).
91. Bhan, N. et al. Enzymatic formation of a resorcylic acid by creating a structure-guided single-point mutation in stilbene synthase. *Protein Sci.* **24**, 167–173 (2015).
92. Ehrenworth, A. M. & Peralta-Yahya, P. Accelerating the semisynthesis of alkaloid-based drugs through metabolic engineering. *Nat. Chem. Biol.* **13**, 249–258 (2017).
93. Deb Roy, A., Grünschow, S., Cairns, N. & Goss, R. J. M. Gene expression enabling synthetic diversification of natural products: chemogenetic generation of pacidamycin analogs. *J. Am. Chem. Soc.* **132**, 12243–12245 (2010).
94. Runguphan, W., Qu, X. & O'Connor, S. E. Integrating carbon–halogen bond formation into medicinal plant metabolism. *Nature* **468**, 461–464 (2010). **This paper is an example of unnatural PNP production via novel enzyme incorporation into the native plant producer, demonstrating that chlorinated precursor metabolites can transit through a biosynthetic pathway to the terminal products.**
95. Glenn, W. S., Nims, E. & O'Connor, S. E. Reengineering a tryptophan halogenase to preferentially chlorinate a direct alkaloid precursor. *J. Am. Chem. Soc.* **133**, 19346–19349 (2011).
96. Wang, S. et al. Metabolic engineering of *Escherichia coli* for the biosynthesis of various phenylpropanoid derivatives. *Metab. Eng.* **29**, 153–159 (2015).
97. Townshend, B., Kennedy, A. B., Xiang, J. S. & Smolke, C. D. High-throughput cellular RNA device engineering. *Nat. Methods* **12**, 989–994 (2015).
98. Feng, J. et al. A general strategy to construct small molecule biosensors in eukaryotes. *Elife* **4**, e10606 (2015).
99. Abatamarco, J. et al. RNA-aptamers-in-droplets (RAPID) high-throughput screening for secretory phenotypes. *Nat. Commun.* **8**, 332 (2017).
100. Michener, J. K. & Smolke, C. D. High-throughput enzyme evolution in *Saccharomyces cerevisiae* using a synthetic RNA switch. *Metab. Eng.* **14**, 306–316 (2012).
101. Raman, S., Rogers, J. K., Taylor, N. D. & Church, G. M. Evolution-guided optimization of biosynthetic pathways. *Proc. Natl Acad. Sci. USA* **111**, 17803–17808 (2014).
102. Matsumura, E. et al. Microbial production of novel sulphated alkaloids for drug discovery. *Sci. Rep.* **8**, 7980 (2018).
103. Ro, D.-K. et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943 (2006).
104. Rodriguez, A. et al. Engineering *Escherichia coli* to overproduce aromatic amino acids and derived compounds. *Microb. Cell Fact.* **13**, 126 (2014).
105. Qin, J. et al. Modular pathway rewiring of *Saccharomyces cerevisiae* enables high-narlevel production of L-ornithine. *Nat. Commun.* **6**, 8224 (2015).

Acknowledgements

We thank Benjamin Kotopka for valuable feedback in the preparation of the manuscript. This work was supported by the National Institutes of Health (grant to C.D.S., AT007886, fellowship to J.T.P., F32 AT009509–03), Novartis Institutes for Biomedical Research (grant to C.D.S., IC2013–1373), and Agilent Foundation (fellowship to A.C.).

Author contributions

A.C. provided draft text for the section “Strategies for planning and engineering a metabolic pathway”. J.P. provided draft text for the sections “Identifying and engineering a suitable host organism” and “Leveraging engineered strains to make novel PNP derivatives”. A.C., J.P., and C.D.S. wrote the other sections, and revised and finalized the manuscript text for publication.


Additional information

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Journal Peer Review Information: *Nature Communications* thanks Mattheos Koffas, Jens Nielsen, and other anonymous reviewer(s) for their contribution to the peer review of this work.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019