Sara Selma García¹,²,*, Nikolaos Ntelkis¹,²,*, Trang Hieu Nguyen¹,², Alain Goossens¹,²,⁵

¹ Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium.
² VIB Center for Plant Systems Biology, Ghent, Belgium.
* These authors contributed equally to this work
⁵ For correspondence (e-mail alain.goossens@psb.vib-ugent.be).

SUMMARY

Plants are the most sophisticated biofactories and sources of food and biofuels present in nature. Engineering the plant metabolism can increase the production of desired compounds and improve the nutritional or commercial value of the plant species. However, this can be challenging because of the complexity of the regulation of multiple genes and the involvement of different protein interactions. To improve metabolic engineering (ME) capabilities, different tools and strategies for rerouting the metabolic pathways have been developed, including genome editing and transcriptional regulation approaches. In addition, cutting-edge technologies have provided new methods for understanding uncharacterized biosynthetic pathways, protein degradation mechanisms, protein-protein interactions, or allosteric feedback, enabling the design of novel ME approaches.

Keywords: CRISPR/Cas, subcellular organelles, posttranslational modification, plant production platforms, interactomics, transcription factors

INTRODUCTION

Plants are champions in producing small natural products, also known as metabolites. Primary metabolites, such as amino acids, nucleotides, vitamins, lipids, and carbohydrates, are essential for the survival of the plant and also serve as the building blocks for an enormous variety of chemical structures, the so-called secondary or specialized metabolites, such as alkaloids, phenolics, terpenes, etc. (Desmet et al., 2021). The latter metabolites typically have more specific roles in plant development, adaptation to different environmental conditions,
and defense against biotic attackers (Marone et al., 2022). In comparison to the group of primary metabolites, which are generally conserved across the plant kingdom, specialized metabolites are much more diverse and different between taxa and species, and even between organs and cell types within a species. A tentative estimate suggests that there are around 200,000 to 1,000,000 specialized metabolites produced within the plant kingdom, from which only a few thousands are known, and from which only a small fraction has been characterized for their physiological functions and/or bioactivity (Saito and Matsuda, 2010; Afendi et al., 2012).

Humans rely on plant-based resources for food, energy, fibers, drugs, cosmetics, pesticides, etc. As the human population is increasing, the demand for plant-based products is also rising, exceeding the natural capacity of plants, even despite current agricultural practices. This is particularly true for the many bioactive specialized metabolites that are typically produced in small amounts by (medicinal) plants (D’Amelia et al., 2021). This supply–demand gap became a driving force to increase the production of high-value metabolites, or, conversely, to reduce that of unwanted compounds by engineering the plant metabolic system, a strategy that is booming thanks to the recent advances in genetic modification and genome editing technologies, and the elucidation of biosynthetic pathways and regulatory factors. Plant metabolic engineering (PME) is defined as the practice of optimizing genetic and regulatory processes to modify the metabolism of the host plant, and consequently trigger a change in its metabolic profile (Farré et al., 2014). Generally, four main goals can be defined: (1) enhancing the content of desired/beneficial metabolites, (2) reducing/depleting the presence of unwanted/toxic compounds, (3) producing natural metabolites that are new to the host plant, and (4) producing new-to-nature metabolites (Arendt et al., 2016; Barone et al., 2020; Zhu et al., 2021).

For PME to be successful, it is often imperative to acquire a profound fundamental knowledge of both the metabolic network, i.e. the individual enzymes involved in a biosynthetic pathway, and the elaborate regulatory networks that have been installed during evolution to steer metabolic networks. Within this special issue, several aspects of metabolic regulation, including gene expression, protein degradation, protein–protein interaction, redox and allosteric feedback, as well as metabolic responses to specific nutrients, cofactors, and developmental and environmental conditions are discussed. In our review, we exemplify how to apply and translate this knowledge to engineer the plant metabolic system. We provide an update of the state-of-the-art in PME, with a special focus on cutting-edge technologies. Finally, we also point to some challenges and considerations when engineering the plant metabolic system, and suggest possible solutions.

**STRATEGIES FOR REWRITING METABOLIC PATHWAYS**

In brief, PME is generally oriented towards the optimization of phytochemical production. To achieve this, ME strategies can for instance involve the transcriptional activation or repression of genes encoding enzymes involved in the pathway or the introduction of genes that encode
a more active version of an endogenous enzyme or that add additional steps of the pathway to obtain a novel metabolite (Figure 1A). The classical strategy to regulate gene expression consists of the overexpression of transcription factors (TFs) (Aharoni and Galili, 2011). TFs usually comprise a DNA-binding domain and an effector domain that recruits the transcriptional machinery or other effectors for the control of gene expression (Hong, 2016). One of the first findings that highlighted the role of TFs as regulators of metabolic pathways was the discovery of the TFs that induce the production of anthocyanins in Arabidopsis and Nicotiana tabacum (Lloyd et al., 1992). Since then, many studies have unraveled the role of different TFs in plant-specialized metabolic pathways, such as the phenylpropanoids (Anwar et al., 2021), the flavonoids (Hassani et al., 2020), the carotenoids (Stanley and Yuan, 2019) or the alkaloids (Yamada and Sato, 2021). One remarkable example is the employment of orthologous TFs that control the different branches of the flavonoid pathway, to generate a collection of transgenic tomato plants that accumulate different levels of anthocyanins, flavonols and flavanones (Butelli et al., 2008; Zhang et al., 2015). This impressive work of ME provides tomato plants with increased nutraceutical values because anthocyanins have been associated with protection against a broad range of human diseases (Ullah et al., 2020). The binding specificity/sensitivity or transcriptional regulation capacities of TFs can also be improved. For example, an engineered CrMYC2 is a de-repressed version of the MYC2a TF that increases the production of monoterpene indole alkaloids (MIAs) in Catharanthus roseus (Schweizer et al., 2018). MIAs are a wide range of compounds with pharmaceutical activities, including antineoplastic compounds, whose production has already been optimized employing heterologous production systems (Liu et al., 2021; Dudley et al., 2022; Huang and Huang, 2022).

As a further optimization, the expression of the target TFs can also be induced with a specific compound, enabling an optimized control of specialized metabolism (De Geyter et al., 2012). This situation occurs naturally in the plant in response to biotic or abiotic stresses, inducing the expression of specific TFs and signaling cascades to regulate specialized metabolism (Zhou and Memelink, 2016). The elucidation of these inducible regulatory sequences coupled with the expression of specific TFs has provided interesting approaches to increase the production of plant metabolites (Memelink et al., 2001; Naoumkina et al., 2008; Colinas and Goossens, 2018). For instance, many jasmonate-inducible TFs, such as MYC2, have been identified to regulate the production of interesting pharmacological compounds that include glucosinolates (Schweizer et al., 2013), terpenes (Ribeiro et al., 2022), or alkaloids (Sui et al., 2018). Remarkably, the production of artemisinin, a sesquiterpene used for the treatment of malaria, can be stimulated in the glandular trichomes of Artemisia annua by jasmonate (Maes et al., 2011), through a signaling cascade that involves many different TFs (Hassani et al., 2020).

Another popular approach for PME is RNA interference (RNAi), which acts post-transcriptionally by silencing or downregulating undesirable genes of the target or competing pathways (Figure 1A) (Pandita, 2022). Some examples include the increase of cottonseed oil by downregulating phosphoenolpyruvate carboxylase 1 (GhPEPC1) in Gossypium hirsutum (Xu et al., 2016), or the adaptation of this crop as a safe source of proteins by decreasing its content of gossypol, a toxic polyphenolic bi-sesquiterpene present in the cotton plant, by
targeting the δ-cadinene synthase-encoding gene (GhCAD1) by RNAi (Rathore et al., 2012). Similarly, the amount of artemisinin can be increased by RNAi-mediated silencing of the genes involved in competing pathways, such as β-caryophyllene synthase (AaCPS), β-farnesene synthase (AaBFS) squalene synthase (AaSQS), and germacrene A synthase (AaGAS) (Lv et al., 2016).

**PROGRAMMABLE TOOLS FOR GENOME EDITING AND TRANSCRIPTIONAL REGULATION**

To maximize the accumulation of interesting compounds, more sophisticated tools need to be applied in PME approaches to redirect the metabolic flux by targeting specific enzymatic steps and avoiding potentially unwanted secondary activities. The emergence of genome editing tools enables the specific manipulation of the genome inducing targeted mutations, but also they can be employed for the efficient control of the transcriptional activity of the target (Jaiswal et al., 2021).

**Zinc-finger (ZFs) and transcription activator-like effectors (TALEs)**

ZFs and TALEs have been engineered as efficient tools for targeted genome editing (Figure 1B). They contain a specific DNA-binding domain fused to the FokI nuclease that generates a DNA double-strand break (DSB), exploiting the error-prone non-homologous end joining (NHEJ) DNA repair pathway to generate, in general, a loss-of-function mutation (Zheng et al., 2020).

The first reported use of Zinc-finger nuclease (ZFNs) in Arabidopsis for genome editing (Lloyd et al., 2005) suggested that this approach is suitable for performing heritable NHEJ-based gene disruptions. ZFNs also could be applied in other plant species, both for basic science and for improving agricultural traits (Novak, 2019). For example, the editing of the maize ZmIPK1 gene (Shukla et al., 2009), which encodes inositol-1,3,4,5,6-pentakisphosphate 2-kinase that catalyzes the final step in phytate biosynthesis, results in a lower accumulation of phytate in the seeds, which is a non-nutritional and environmentally harmful metabolite.

Similarly, TALENs (transcription activator-like effector nucleases) have been employed in plants for ME purposes, showing an improvement in specificity and versatility compared with the ZFNs (Gupta and Musunuru, 2014). In rice, a fragrant variety was generated from a non-fragrant one by altering its volatile pattern via a targeted knock-out of betaine aldehyde dehydrogenase 2 (OsBADH2), a gene involved in 2-acetyl-1-pyrroline (2AP) emissions in rice (Shan et al., 2015). The nutritional characteristics of soybean was improved by increasing the amount of oleic acid and decreasing that of linolenic acid via targeting of the gene GmFAD2 encoding fatty acid desaturase 2 (Haun et al., 2014). In sugarcane (Saccharum spp. hybrids), targeted mutations of the gene encoding caffeic acid O-methyltransferase (SoCOMT) tweak the lignin pathway in this crop, resulting in altered lignocellulosic biomass allowing more efficient bioethanol production (Jung and Altpeter, 2016). In Nicotiana benthamiana, the model crop for molecular farming, a knock-out line was generated for α(1,3)-fucosyltransferase (NbFucT) and β(1,2)-xylosyltransferase (NbXylT), genes that remove the plant-specific N-glycans, thus providing an improved platform to produce biopharmaceuticals in plants (Li et al., 2016).
In addition to inducing loss-of-function mutations, this technology is also suitable to obtain controlled gene insertions, as such avoiding potentially detrimental effects of random insertions. The engineered nuclease generating DSBs is coupled to a DNA donor carrying the sequence of homology and the cassette to be inserted by homologous recombination. Although this approach is less efficient in plants than in other organisms (Schuermann et al., 2005), the challenging homologous recombination can be mediated by TALENs as reported in tomato (Čermák et al., 2015). This work combines the TALENs with a geminivirus replicon carrying the DNA donor that contains *anthocyanin 1* (*SlANT1*), encoding a TF that controls antioxidative anthocyanin production, resulting in tomatoes with high levels of anthocyanins. This approach, although less efficient, allows the controlled insertion of genes into the plant genome, avoiding the potentially detrimental effects of random insertions.

Beyond the employment of ZFNs and TALENs for genome editing, these tools have also been proven suitable as programmable TFs to control target gene expression, thereby avoiding the limitations of the classical TF-based strategies. To enable sequence-specific control of the transcriptional activity, it was necessary to change the coupled Fok-1 nuclease by translational fusions of transcriptional regulation domains (Gaj et al., 2013). Several nice examples employing the ZFs and TALEs in plants for programmed transcriptional regulation have been published (Stege et al., 2002; Lindhout et al., 2006; Gao et al., 2014; Liu et al., 2014). Synthetic ZFs enabled reporter expression in tobacco and maize protoplasts (Stege et al., 2002). Furthermore, synthetic TALEs were used to increase anthocyanin accumulation through the overexpression of *PRODUCTION OF ANTHOCYANIN PIGMENT 1* (*AtPAP1*) in transgenic tobacco plants (Liu et al., 2014). In a similar manner, designer TALEs (dTALEs) were successfully tested as transcriptional activators of endogenous genes in Arabidopsis genes (Morbitzer et al., 2010). Yet, ZFNs and TALENs have not been extensively used as tools in PME due to their low versatility and the need to design a new protein with each target.

**CRISPR/Cas**

To overcome the low versatility of the ZFNs and TALENs, the CRISPR (clustered regularly interspaced short palindromic repeats)/Cas (CRISPR-associated) system emerged as a specific and versatile alternative to perform targeted genome editing and transcriptional regulation, opening the door to novel and sophisticated approaches in PME. The CRISPR/Cas system is derived from ancient immunity systems adopted by some prokaryotic organisms and comprises a nuclease (Cas9) and a guide RNA (gRNA). The gRNA drives the Cas9 to the target site, generating a specific binding and cleavage that results in a DSB. This system is widely used in different species to efficiently mutate the target genes by exploiting, like the previous editing tools, the error-prone NHEJ DNA repair pathway (Jiang and Doudna, 2017).

Currently, many examples of CRISPR-mediated genome editing exist in plants (Razzaq et al., 2019; Wada et al., 2022), proving its efficiency and ease of use. Specific CRISPR knock-outs have been generated to reroute metabolic pathways and obtain biofortified crops or optimized production platforms for added-value compounds. For instance, engineering the phytosterol biosynthesis pathway provides biofortified tomatoes with a high accumulation of provitamin D3 (Li et al., 2022). In this work, the CRISPRend tool was used to target 7-DR2 encoding an enzyme involved in the conversion of provitamin D3 to cholesterol. Consequently,
a higher accumulation of the precursor of vitamin D in fruits was achieved with a minimal impact on other important branches of the pathway, such as the phytosterol and brassinosteroid biosynthesis (Li et al., 2022). Other examples of ME are the efficient manipulation of the γ-aminobutyric acid (GABA) content in tomato fruits employing two different approaches. Nonaka et al. (2017) performed a targeted mutagenesis of the glutamate decarboxylase genes (SIGAD2 and SIGAD3) by CRISPR, deleting the autoinhibitory domain contained in these enzymes and thus increasing their activity. Li et al. (2018) carried out a multiplexed approach to reroute the GABA shunt by targeting simultaneously four key genes, SIGABA-TP1, SIGABA-SITP3, SICAT9 and SISSADH. In both studies, the content of GABA was efficiently increased in the fruits, reaching a 15-fold and 19-fold of accumulation respectively, compared with the control. Another interesting pathway in tomato fruit that was manipulated by CRISPR was that of lycopene, promoting its synthesis through targeting the SlSGR1 gene for loss of function, while inhibiting its conversion into β- and α-carotene by targeting the LCY genes. Similarly was the accumulation of β-carotene in banana fruits increased by CRISPR-based editing of the GN-LCYε gene, simultaneously decreasing the α-carotene and lutein contents (Kaur et al., 2020). Also in other crops like the soya bean, the simultaneous CRISPR-based editing of GmF3H1, GmF3H2 and GmFNSII-1 was carried out to increase its isoflavone content and thus improve its resistance to the soya bean mosaic virus (Zhang et al., 2020). In Medicago truncatula, the rerouting of the saponin pathway was achieved by knocking out the MtCYP93E2 and MtCYP72A61 genes, thus accumulating hemolytic sapogenins and avoiding the accumulation of sapogenols (Confalonieri et al., 2021).

Additionally, the CRISPR tool can also be employed for the regulation of target gene expression. In order to avoid the endonuclease activity of the Cas protein and maintain its binding capacity, a mutation of specific amino acids in the RuvC1 and HNH nuclease domains was carried out (Qi et al., 2013). The resulting protein, named “dead Cas” or dCas can be directed to the promoter of target genes by the gRNA and generate a transcriptional response, either activation or repression, depending on the regulatory domain to which it is attached (Larson et al., 2013; Maeder et al., 2013). Since the emergence of the initial CRISPR-based TFs, several optimizations of the effector domain have been reported to maximize the transcriptional response (Karlson et al., 2021). Currently, there are some studies that have developed CRISPR-based programable TFs robust enough to carry out a simultaneous activation of several genes involved in a metabolic pathway. For example, the CRISPR-Act3.0 strategy was reported as a strong and efficient tool for maximizing transcriptional regulation in plants, activating the multiple enzymatic steps of proanthocyanin and β-carotene biosynthesis (Pan et al., 2021). Also, the dCasEV2.1 tool, also an engineered version of the dCas9 protein to perform targeted transcriptional activation, was efficiently reported for ME approaches, targeting multiple genes involved in the flavonoid pathway that results in a precise accumulation of the targeted flavonol or flavanone in the leaves (Selma et al., 2022). These tools present a great potential both for activation or repression approaches and thus rewriting the endogenous pathways in plants. The transcriptional activation or repression can even be combined using different Cas proteins (Tang et al., 2017) to accumulate the target metabolite, while avoiding unwanted branches of the pathways.
SUBCELLULAR COMPARTMENTALIZATION FOR PME

The biosynthetic pathway leading to the production of a specific specialized metabolite in plants often comprises several enzymatic steps that can take place in different subcellular localizations such as the cytosol, chloroplast, vacuole, nucleus, peroxisome, or endoplasmic reticulum (ER) (Shih and Morgan, 2020). One well-characterized example of multiple-organelle compartmentalization of a metabolic pathway is the MIA biosynthesis in *C. roseus* (Yamamoto et al., 2016). This characteristic of metabolic networks adds complexity to the regulation of the target pathway but also opens the door for improving the metabolic output.

The optimization of subcellular compartmentalization for PME approaches can include the re-localization of the enzyme production in the organelles, the alteration of the subcellular precursor availability, or the regulation of the transporters associated with biosynthesis (Heinig et al., 2013; Jaramillo-Madrid et al., 2022). For instance, retargeting the expression of the enzymes to the compartment with the highest precursor pool can be used to increase the targeted metabolite accumulation. Although there are not so many examples of this approach, it was reported that the overexpression of patchoulool synthase along with farnesyl diphosphate (FPP) synthase in tobacco plastids results in a 100-fold higher accumulation of patchoulool compared with its cytosolic production (Wu et al., 2006). Also, the overexpression of *FaNES1*, encoding the strawberry linalool/nerolidol synthase, in different subcellular localizations can produce an altered pattern of metabolite accumulation. This enzyme requires FPP to produce nerolidol and geranyl diphosphate to produce linalool. The overexpression of *FaNES1* in the mitochondria enables nerolidol to accumulate up to 30-fold higher compared with plastid overexpression, in which the accumulation of FPP is less abundant. In contrast, the plants that overexpress *FaNES1* in the plasmids, in which the pool of geranyl diphosphate is more abundant, show a higher accumulation of linalool (Aharoni et al., 2003; Kappers et al., 2005). Another recent example exemplifies the reconstruction of the momilactone biosynthetic pathway in *N. benthamiana*, involving rerouting of diterpene biosynthesis from the chloroplast to the cytosol, thereby improving heterologous momilactone production by more than 10-fold (De La Peña and Sattely, 2021).

Another approach consists of the engineering of the transporters present in the organelles (Jain and Zoncu, 2022). Because a metabolic pathway can be partitioned between different subcellular compartments, a flux of metabolic intermediates can be generated between the organelles. The activity of the transporters can determine the bottleneck when the metabolic flux through a pathway is increased by a PME approach. For example, the overexpression in the plastid of the glucose-6-phosphate/phosphate translocator and adenylate translocator, the transporters that enable starch formation in the cell, allows an increased carbon skeleton and energy supply into the plastids and thus augments the pool of starch precursors (Zhang et al., 2008). Following the same approach, the overexpression of the transporters *AaLTP3* and *AaPDR2* in *N. benthamiana* plants that express the artemisinin biosynthetic pathway increases artemisinin and arteannuin B production (Wang et al., 2016). This finding expands the strategies to improve terpene production in other heterologous systems.

In a further step, it is possible to engineer the organelles to, for example, increase the foreign protein expression, avoiding gene silencing and thus improving metabolic production (Jensen
and Scharff, 2019). In a work reported by Lu et al. (2013), the total tocopherol and tocotrienol accumulation in tobacco were increased by adding a three-gene biosynthetic pathway in chloroplasts, showing that it is possible to manipulate the genome of plastids to maximize the target metabolic outcome. Another approach of organelle engineering consists of the generation of a macro-chloroplast by the overexpression of the \textit{AtFtsZ1} gene, which is involved in protoplast division (Stokes et al., 2000; Occhialini et al., 2020). The increased chloroplast size can offer more suitable structures for maximizing metabolic production, such as terpene production (Nagegowda and Gupta, 2020). Also the ER is an interesting target to be engineered, since it is the center of the synthesis and transport of a wide diversity of metabolites (Jaramillo-Madrid et al., 2022). Although currently there are not many studies to have reported the modification of the ER to optimize metabolic production in plants, the studies carried out in yeast showed that ER expansion, through the knock-out of the \textit{phosphatidate phosphatase 1 (ScPAH1)} gene, increases its metabolic capacity (Arendt et al., 2017; Zhao et al., 2021). This achievement inspires PME to develop a similar strategy to increase ER proliferation. However, the \textit{PAH} mutation in plants generates a negative effect on plant growth and viability (Eastmond et al., 2010), requiring other strategies to engineer the ER for PME purposes.

**POSTTRANSLATIONAL MODIFICATIONS (PTMs)**

An equally important step in metabolic flux regulation takes place at the protein level, as proteins act as enzymes or (ultimately) regulators of enzyme activities and levels. Pathway-related proteins, such as biosynthetic enzymes, regulators or TFs, are constantly used, renewed and replaced. Fine-tuning of this process, which ultimately affects metabolite production, happens with PTMs, which are chemical modifications of specific amino acids typically in accessible sequences at the protein surface. These modifications can be classified as enzymatic or non-enzymatic, with the former requiring enzymes to perform a modification at a specific residue, while the latter occur spontaneously, because of the environmental conditions. Examples of enzymatic PTMs include phosphorylation, acetylation, ubiquitination, sulfation and N-glycosylation, with for each a specific enzyme that is responsible for either the addition (e.g. kinase) or the removal (e.g. phosphatase) of the chemical group.

Well-studied examples of PTMs regulating plant metabolism include the glycolysis pathway (O’Leary et al., 2011), the Calvin-Benson cycle (Michelet et al., 2013) and photorespiration (Bartsch et al., 2010; Hodges et al., 2013) for primary metabolism, as well as the isoprenoid pathway (Banerjee and Sharkey, 2014) for specialized metabolism. In most examples, a combination of several enzymatic and non-enzymatic PTMs of different enzymes are required for the proper functioning of the pathway. However, these examples represent exhaustively studied metabolic pathways that have been under the spotlight for decades, which is generally not the case for novel biosynthetic pathways of specialized metabolism. To design optimized PME programs for increased metabolite yield, it is important to identify, predict and characterize PTMs occurring in a metabolic pathway of interest. This requires a combination of -omic datasets, modeling and screening technologies to identify not only which PTMs exist,
but also which PTMs are relevant to a specific pathway. Such an example can be found for *Escherichia coli* metabolism, where whole-genome metabolic modeling, transcriptomics, high-throughput screening and molecular dynamics were used to identify relevant PTMs for specific enzymes within pathways of interest (Brunk *et al*., 2018). While relevant tools, such as CRISPR pooled screens (Gaillochet *et al*., 2021) and whole-genome metabolic models (Töpfer and Niokoloski, 2013), exist in plants, PTM engineering is currently focused mainly on N-glycosylation, to create human-like glycan structures (Strasser *et al*., 2014; Montero-Morales and Steinkellner, 2018), with the aim to “humanize” plant PTMs instead of engineering them toward increased metabolic production.

**TECHNOLOGIES FOR PME**

**Production platforms**

Choosing the production platform in a PME project is crucial to achieve a high yield of the desired metabolite (Figure 2A). Unlike microbial platforms, where *E. coli* and *Saccharomyces cerevisiae* are typically the production platforms for a range of metabolites, plant diversity enables the use of different hosts for different target metabolic pathways, either *in vivo* or *in vitro*. For example, several high-value biopharmaceuticals (e.g. paclitaxel) are produced *in vitro* from plant cell cultures of naturally overproducing plants (Wilson and Roberts, 2012) even at commercial scale. Natural harvest of whole plants (*in vivo*) for metabolite extraction is also performed, although with lower yields and with high environmental impact (Wu *et al*., 2021). Plant cell cultures enable the usage of bioreactors and thus can be compared with microbe-based industrial biotechnology processes, which allow a precisely monitored and contained production under good manufacturing practices, which are highly desirable parameters in an industrial-scale setting.

Among *in vitro* production platforms, there are several options that lead to metabolite overproduction. First, the plant cell culture can stem from the naturally overproducing plant (homologous overproduction) or not (heterologous overproduction). Engineering a naturally overproducing cell line requires less effort, because the pathway in question already exists in the genome and is expressed. Such approaches have been utilized and reached commercial application, for example with paclitaxel plant cell cultures (Malik *et al*., 2011). Many general classes of specialized metabolites that are used as dyes, from anthocyanins to betacyanins, or as drug-like molecules, such as scopolamines, are represented in this category of natural product commercialization via plant cell cultures and we direct the reader to a thorough review on the matter (Ochoa-Villarreal *et al*., 2016). Heterologous production occurs in cell cultures of other model plants such as *A. thaliana* and *N. tabacum*, where tools for genetic engineering are widely available (Wu *et al*., 2021).

Another consideration is whether to create stable or transient lines. Transient transformation usually requires a plant organ, such as the leaf, to perform the transformation. A popular and straightforward transient transformation technique is leaf infiltration of *N. benthamiana*.
leaves (Chincinska, 2021), which yields rapid results and has the potential for scale-up. Such a high-throughput synthetic biology platform was set-up to yield gram-scale quantities of economically important specialized metabolites, such as beta-amyrin, a precursor to oxygenated triterpenes (Reed et al., 2017). Hairy root transformation is a stable transformation method that requires Agrobacterium rhizogenes to infect the roots (Hu and Du, 2006). Proper infection increases both lateral root formation and biomass, eventually increasing the level of metabolites produced (Chandra and Chandra, 2011). Hairy root cultures can then be maintained in the dark by supplying the required nutrient sources, in flasks of batch-fed bioreactors. The versatility of A. rhizogenes strains to infect a range of different plant species is outstanding and can be used both on model and non-model organisms. For example, overexpression of AtMYB12, a master transcriptional regulator of flavonoid biosynthesis in Arabidopsis, in licorice (Glycyrrhiza inflata) hairy roots leads to a 2- to 5-fold accumulation of bioactive flavonoids licochalcone A and echinatin, which entail anti-inflammatory, antimicrobial and antioxidant properties (Wu et al., 2022). Similarly, overexpression of the leaf color (LC) TF that regulates anthocyanin production in maize and of AtPAP1, a member of the MYB TF family in Arabidopsis, in hairy roots of the Asian medicinal plant Scutellaria baicalensis, results in an enhanced total flavone content, and especially of baicalin, baicalein, and wogonin, all comprising health-promoting traits (Park et al., 2021).

Advances in PME have also been observed in legumes. Fine-tuning of the mevalonate pathway was achieved in Medicago truncatula, a model legume (Mertens et al., 2016). Overexpression of two native triterpene saponin master regulators (MtTSAR1 and 2) in M. truncatula hairy roots results in the upregulation of several enzymes in the saponin pathway and eventually leads to an increased accumulation of those specialized metabolites (Mertens et al., 2016). Overexpression of the same regulators in hairy roots of the medicinal legume Trigonella faenum graecum similarly altered the expression levels of the saponin biosynthetic genes and increased saponin content (Garagounis et al., 2020). In the same plant, overexpression of Δ24-reductase, the native rate-limiting enzyme of diosgenin biosynthesis, steers the metabolic flow toward cholesterol production, resulting in an increased diosgenin content, an economically important metabolite for progesterone and cortisol biosynthesis (Nasiri et al., 2022).

**Enzyme engineering**

An essential component of a PME project is the set of enzymes that will eventually produce the metabolite of interest from simple precursor molecules. Since enzymes have likely evolved to perform a specific biological function and not to catalyze chemical reactions at maximum efficiency, enzyme kinetics often limit the production of certain metabolites. Enzyme engineering allows the optimization of key or rate-limiting enzymes to increase metabolic flux through the pathway and eventually increase metabolite yields (Figure 2B). Optimization could be done by deleting negative feedback regulation or by increasing catalytic activity. For example, allosteric feedback inhibition by the end product naringenin was removed from the phenylpropanoid pathway to increase its metabolic flux in tomato (Alberstein et al., 2012).
The authors targeted 4-coumarate:CoA ligase (SI4CL), an upstream enzyme of the phenylpropanoid pathway that is negatively regulated by naringenin (Alberstein et al., 2012). Directed evolution by PCR mutagenesis and screening for naringenin chalcone, which produces a yellow color in the visible spectrum and allows for easy colorimetric measurement of metabolic accumulation, enabled the identification of superior 4-coumarate:CoA ligase variants that can be used for the increased production of phenylpropanoids.

Also more sophisticated approaches, such as continuous directed evolution (Molina et al., 2022), have been employed for plant enzymes, where the enzyme is optimized ex planta in compatible systems. The main difference between continuous and classical directed evolution is that continuous directed evolution is performed continuously in vivo, by exploiting the cellular machinery of microorganisms like E. coli or yeast, whereas classical directed evolution is performed in batches or in vitro, which is labor-intensive and limiting in terms of throughput. Using this technology, arogenate dehydratase (AtADT2) from A. thaliana was optimized in E. coli by creating feedback-resistant variants that can be used for plant breeding (Leong and Hanson, 2023). Work from the same group resulted in the optimization of the thiamine thiazole synthesis enzyme THI4, which is an energetically expensive enzyme, as it is destroyed after one catalytic cycle. The authors explored the possibility of optimizing a prokaryotic homolog of THI4 (TaTHI4) for increased catalytic activity in plants, while also creating a high-throughput cloning platform for other target genes (García-García et al., 2020; García-García et al., 2022). Another example of rational engineering is the tomato GABA shunt engineering work, where feedback-resistant variants of SlGAD2 and SlGAD3 were created using CRISPR/Cas9, which resulted in a 7- to 15-fold increase in GABA accumulation in the fruit (Nonaka et al., 2017). These works highlight the importance of enzyme engineering and the adaptation of cutting-edge tools for PME.

Besides semi-rational or evolutionary based wet-lab approaches to optimize enzymes, in silico tools have emerged to aid in this quest and possibly reduce the labor and resources required. Molecular docking is a proven tool that can predict possible interactions between enzymes and their substrate or an allosteric inhibitor, and is based on thermodynamic principles to arrive at an optimal prediction (Fan et al., 2019). However, docking requires a protein structure to perform docking simulations, which until recently was the limiting factor, because many plant specialized metabolism enzyme structures are not characterized. Alphafold, a protein structure prediction tool may provide a solution for this (Jumper et al., 2021). Its performant ability to predict protein structure could be used to predict the structure of novel enzymes in a pathway of interest. Coupled with molecular docking, multi-omics data and whole-genome metabolic models, enzyme engineering could become more streamlined and bioinformatically aided.

**Data mining**

Another fundamental tool for the contemporary metabolic engineer is data mining, as there is abundant information on biosynthetic genes/clusters and genes in general from metagenomics or plant genome databases (Figure 2C). By having access to a huge amount of
genomic data that can be mined for novel enzymes and regulatory elements, as well as mining tools like antiSMASH (Blin et al., 2019), it is possible to find new pathway components that perform better and use those as is or as starting material for optimization and future use in plant metabolic pathways. For example, transcriptomic mining was recently used to elucidate the metabolic pathway of the FDA-approved plant alkaloid colchicine from naturally overproducing Colchium autumnale and Gloriosa superba (Nett et al., 2020). The authors utilized the fact that biosynthetic genes usually are co-expressed to identify and verify candidate genes by reconstructing a near-complete pathway of 16 enzymatic steps in N. benthamiana, paving the way for ME of this alkaloid class (Nett et al., 2020). In a similar example from the same group, metabolomic and transcriptomic data were used to gain insight into the biosynthesis and metabolic regulation of huperzine A, an acetylcholinesterase inhibitor with clinical interest as a treatment for neurological disease (Nett et al., 2021).

When it comes to enzyme engineering, we do not always need to reinvent the wheel and engineer enzymes or pathways to perform in a more efficient manner – nature has maybe already done it. A stellar example is Rubisco. Rubisco performs the carboxylation step in the carbon fixation process, which is also rate-limiting in most modern agricultural conditions. However, optimizing Rubisco by rational engineering has proved inefficient. Data mining of metagenomic databases allowed the discovery of highly active Rubisco enzymes, that outperformed plant Rubiscos in vitro (Davidi et al., 2020). In a different example, the evolution of metabolic pathways and genomic/metabolomic data from different tomato varieties was exploited to increase the biosynthesis of acylsugar, a specialized defense compound (Ning et al., 2015). This was achieved by identifying an evolved feedback-resistant variant of ISOPROPYLMALATE SYNTHASE (SlIPMS3), which acts in Leu biosynthesis. This variant had a C-terminal deletion, which abolished regulation by Leu, without affecting catalytic activity. By creating transgenic lines with the evolved IPMS variant, increased acylsugar production was observed, as well as a change in the structural diversity of the acylsugars produced (Ning et al., 2015).

**Proximity labeling and metabolite–protein interactions**

In order to elucidate the regulation of a metabolic pathway, it may be crucial to identify the protein–protein interactions (PPIs) or metabolite–protein interactions (MPIs) that enzymes or regulators have with other proteins or metabolites within the producing cell (Figure 2D). Identification of the interactions is the first step toward engineering the pathway to increase the flux e.g. by removing allosteric inhibition of an enzyme or deleting domains responsible for PPIs that lead to protein degradation or inactivation. Tools like yeast two-hybrid and immunoprecipitation provide standard in vitro tools for the identification of PPIs, while proximity-dependent biotin labeling (PBL) is a new tool that allows the high-throughput identification of proximal interactions of a bait protein. PBL systems such as TurboID have been also applied and optimized for use in plant systems, for the identification of interactors of a nucleotide-binding leucine-rich repeat immune receptor (Zhang et al., 2019) and of the GSK3 signaling network in A. thaliana (Kim et al., 2019). Although there are still no examples of PBL in PME projects, the technology is definitely capable of assisting in the optimization of
plant metabolic pathways, as it has been tested for use in several major plant models (Arora et al., 2020).

In addition to PPIs, MPIs also play an essential role in metabolic regulation, with the most basic example being allosteric inhibition by downstream products of a metabolic pathway. Engineering enzymes and regulators to be feedback-resistant can be tricky, because the exact allosteric site and amino acid residues responsible for an interaction are difficult to pinpoint. Technologies such as limited proteolysis coupled with mass spectrometry (MS) (Piazza et al., 2018) or cellular thermal shift assay coupled with MS (Dziekan et al., 2019) enable the identification of such interactions with high level of detail (Venegas-Molina et al., 2021), and have been applied in plants to identify novel and existing interactions (Piazza et al., 2020). The advantage of such a tool is that it is metabolite-centered, which provides insights into and information about the metabolite of interest directly. For example, the intermediate of final metabolites of a pathway of interest could be screened for proteome-wide interactions, to identify not only closely related negative metabolic regulation but also interactions that are related to regulation upstream of the pathway or at the signaling level.

These technologies allow a metabolite-centric view of metabolism and metabolic regulation, but what if one wishes to characterize a novel pathway where intermediate or final metabolites are unknown? This requires global or interactome-wide approaches, such as PROMIS (Veyel et al., 2018), which combines co-fractionation and size exclusion chromatography with simultaneous protein–metabolite extraction and liquid chromatography-MS to identify all PMIs in a tissue/cell lysate. Although there are no direct applications to PME yet, the wealth of information one can acquire with such a dataset is immense and has direct applications for the characterization of existing and novel pathways. As an example, PROMIS was used to create a multi-omics time series of protein and metabolite accumulation in Arabidopsis leaves, which can be mined for novel metabolic regulators (Omidbakhshfard et al., 2021).

OUTLOOK

With the increasing demand for plant natural products, our better understanding of the plant's metabolic systems thanks to recent advances in omics approaches, and cutting-edge genome engineering technologies, PME may become a hot area in the green industry in the twenty-first century. As highlighted in this review, the past two decades, many successful stories have been reported on the improvement of metabolism in crops and medicinal plants (Zhu et al., 2020; Mora-Vásquez et al., 2022). However, this is not always straightforward, because there are still large gaps in our knowledge of plant metabolic networks. The complexity of the metabolic networks, being interconnected and tightly regulated at multiple levels, which is particularly true for the central metabolism, impedes accurate predictions of engineering outcomes or limits the successful production of desired compounds owing to growth penalties and/or vulnerability to stresses. Also, we have only scratched the surface of the potentially hundreds of thousands of metabolites that exist in the plant kingdom, many of
which may have yet unanticipated roles as plant growth regulators or signaling molecules, next to possibly being bioactive compounds of human interest. So, understanding the plant metabolism as a whole is the key for a successful intervention strategy, which can be referred to as knowledge-driven approaches for PME (Farré et al., 2015). In addition to the identification of enzymes and regulators of the pathway, mapping the interaction between metabolites and their target proteins and characterizing metabolite interacting domains is needed not only to help resolving the complexity of plant metabolism, but also to identify novel targets for engineering. Although this is still a challenging task, we have highlighted several techniques that are currently being developed and optimized for implementation in plants (Venegas-Molina et al., 2021).

Another hardly explored frontier in PME is that of metabolite biosensors. Traditionally, in a PME project, compounds of interest are analyzed using chromatographic separation, using an inert gas or liquid as a mobile phase (gas/liquid chromatography, GC/LC), coupled to various types of MS. These approaches are typically laborious and time consuming, and usually require biochemical and analytical chemistry expertise. Genetically encoded biosensors offer a nondestructive alternative to the metabolomics analysis techniques, that would be more amenable for design–build–test–learn (DBTL) cycles, an increasingly adopted framework that represents a more systematic and efficient approach in ME and synthetic biology programs in microbial chassis. The development of biosensors requires advanced knowledge of metabolite–target protein interactions, i.e. the metabolite-interacting protein domains need to be identified and fused to a reporter domain, such as fluorescence proteins. This principle has been applied to develop several plant hormone biosensors (Brunoud et al., 2012; Larrieu et al., 2015). However, this research area is still relatively unexplored for most of plant specialized metabolites. Recently, Beltrán et al. (2022) adopted a system based on the abscisic acid-sensing system with the PYR1 receptor, which functions through chemically induced dimerization. Using computationally designed libraries of PYR1, the authors created high-affinity receptors for 21 structurally diverse ligands, including natural and synthetic cannabinoids and several organophosphates, setting the stage for large-scale user-interested metabolite biosensor development.

Last but not least, artificial intelligence is changing many aspects of life, and has the potential to function as the brain in future ME projects. Machine learning can use the exponentially growing amounts of omics data to choose the best compound to produce, suggest possible pathways or eventually design new enzymes and pathways to produce it, and interpret the resulting experimental data to troubleshoot the ME effort (Lawson et al., 2021; Jang et al., 2022). A pioneering application of machine learning in improving butanol production in a cell-free system (Karim et al., 2020) pointed to the great potential of this approach to design and improve metabolic pathways (for a review, see Lawson et al., 2021). Although there is no example of artificial intelligence application in PME, we envisage that it will become a tool used on a regular basis in future PME projects.
ACKNOWLEDGEMENTS

We thank Annick Bleys for the critical reading and help in preparing the manuscript. This work has been supported by the European Community’s H2020 Program under grant agreement 760331-Newcotiana to A.G. and the Research Foundation Flanders with research project grant G0E1122N to A.G.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest associated with this work.

REFERENCES


Leong, B.J. and Hanson, A.D. (2023) Continuous directed evolution of a feedback-resistant Arabidopsis arogenate dehydratase in plantized Escherichia coli. ACS Synthetic Biology 12 (1), 43-50. https://doi.org/10.1021/acssynbio.2c00511

Li, J., Scarano, A., Gonzalez, N.M., D’Orso, F., Yue, Y., Nemeth, K. et al. (2022) Biofortified tomatoes provide a new route to vitamin D sufficiency. Nature Plants 8 (6), 611-616. https://doi.org/10.1038/s41477-022-01154-6


Wada, N., Osakabe, K. and Osakabe, Y. (2022) Expanding the plant genome editing toolbox with recently developed CRISPR–Cas systems. Plant Physiology 188 (4), 1825-1837. https://doi.org/10.1093/plphys/kiac027


Figure 1. Tools and strategies for rewriting metabolic pathways. A) Classical approaches employed for PME that include the addition of heterologous steps of the pathway, the transcriptional gene regulation mediated by TFs or the gene silencing/knock-down induced by RNAi. B) Efficient tools for genome editing and transcriptional regulation employed for rewriting the metabolic pathways in plants. The genome editing tools can be applied for both generating knock-outs or more active versions of the edited enzymes (E.2.2). The programable TF can perform either a targeted transcriptional activation or repression to modulate the metabolic flux. Green, gray, and red squares represent the production of a metabolite, repression of a metabolic branch, and the absence of the metabolite in the pathway, respectively. Faint colors in the squares reflect decreased metabolite levels. The enzymes are represented by the label "E". The green/orange/red labeling in the enzymes annotations is employed for representing the presence/ decrease/absence of the enzyme, respectively. The arrows indicate an increase or decrease of the enzyme in the pathway. Figure created with BioRender.com.
Figure 2. Technologies for PME. A) Plant transformation tools, like Agrobacterium transformation of whole plants or specific organs (leaves or roots) allow the engineering of metabolic regulation depending on the end goal. B) Enzyme engineering utilizes dry and wet lab approaches to create superior enzymes, with increased catalytic activity, lower turnover rate or feedback resistance that alter metabolic regulation of specific metabolic pathways. C) Bioinformatics and especially data mining can aid in engineering metabolic regulation by providing alternative pathway enzymes or regulators with a desired function. This technology should be combined with enzyme engineering to adapt mined genes to a specific organism, or to produce similar changes in native genes based on data mining information. D) Characterizing metabolite–protein interactions is crucial to understand the complex networks of protein and ligand binding in a target pathway. This technology indirectly helps engineering metabolic regulation by providing a wealth of information that can be utilized in other technologies to directly affect metabolism. Created with BioRender.com