Elucidation and *de novo* Reconstitution of Glyceollin Biosynthesis

Yunlong Sun[1,#], Cong Chen[2,#], Chao Lin[1], Hao Zhang[1], Jiazhang Lian[2,3,]* Benke Hong[1,]*

[1]Zhejiang Key Laboratory of Precise Synthesis of Functional Molecules, Department of Chemistry, School of Science and Research Center for Industries of the Future, Westlake University, Hangzhou, China.

[2]Key Laboratory of Biomass Chemical Engineering of Ministry of Education, College of Chemical and Biological Engineering, Zhejiang University, Hangzhou, China.

[3]ZJU-Hangzhou Global Scientific and Technological Innovation Center, Zhejiang University, Hangzhou, China

[#]These authors contributed equally

*Corresponding authors

Abstract

Glyceollins are phytoalexins produced by soybeans in response to stressors such as pathogen invasion, injury, and environmental challenges. In addition to their antibacterial and antifungal activities, these compounds have attracted significant attention for their potential anticancer, anti-inflammatory, and antioxidant properties. However, their limited accessibility-due to the challenges and high costs associated with organic synthesis or purification from treated soybean seedlings-has hindered further physiological and biochemical studies. Moreover, the incomplete understanding of glyceollin biosynthesis, particularly the final cyclization steps, remains a major barrier to elucidating their physiological functions and achieving sustainable production through synthetic biology. In this study, we identified previously uncharacterized genes encoding two reductases for 7,2',4'-trihydroxy-isoflavanol (THI) synthesis and five P450 enzymes responsible for the final oxidative cyclization in glyceollins I, II, and III biosynthesis, thereby completing the entire biosynthetic pathway. By reconstructing the pathway *de novo* through synthetic biology, we achieved the successful production of glyceollins from simple carbon sources in Baker's yeast. This work advances understanding of glyceollin biosynthesis, facilitates sustainable production in microbial hosts and offers new opportunities for their applications in agriculture and biology.

Introduction

Soybean (*Glycine max*), one of the most important crop plants in the world for dietary protein and oil production, often suffers yield losses due to pathogen attacks such as *Phytophthora sojae*.[1-3] In response to these threats, soybeans leverage their natural resistance mechanisms to produce isoflavonoid phytoalexins like glyceollins.[4-5] In addition to their antibacterial and antifungal activities,[6-7] glyceollins have garnered significant attention due to their anticancer, anti-inflammatory, and antioxidative properties,[8-10] with particular potential for targeting breast cancer through the estrogen receptor signaling pathway.[11-14] Unlike ubiquitous isoflavonoids daidzein and genistein, which are constitutively present in soybeans,[15] glyceollins are synthesized only under specific environmental triggers, such as pathogen infection or exposure to other elicitors. Soybean predominantly produces three glyceollin isomers: glyceollin I, glyceollin II and glyceollin III. Pure forms of glyceollins can be purified from soybeans, but the methods for large-scale isolation are time-consuming and unsuitable for mass production, which has driven interest in alternative production methods.[16-17] Although chemical synthesis of glyceollins I and II has been reported,[17-21] the complex synthetic routes and high manufacturing costs have limited their practical applications for biological use.

Since the discovery of glyceollin I in soybean in 1972,[22] the biosynthetic mechanism of glyceollins has been extensively investigated (**Figure 1**).[10] They are synthesized from phenylalanine (**1**), which is converted into the isoflavone daidzein *via* eight known genes. Daidzein (**2**) is transformed into 2'-hydroxydaidzein (**3**) by CYP81E,[23] followed by reduction to 2'-OH dihydrodaidzein (**4**) by isoflavone reductase (IFR).[24,25] A subsequent reduction, mediated by an uncharacterized reductase, produces 7,2',4'-trihydroxyisoflavanol (THI, **5**), which is converted to 3,9-OH pterocarpan (**6**) by pterocarpan synthase (PTS) [25] and further modified to glycinol (**7**) by CYP93A1.[26] Glycinol (**7**) undergoes dimethyl allylation at C2 or C4 position, catalyzed by prenyltransferase G2DT[27] or G4DT[28,29], resulting in the formation of 4-glyceollidin (**8**) or glyceocarpin (**9**), respectively. However, the final cyclization steps remain unresolved, as the genes encoding the responsible enzymes have not been identified. It is unclear whether these steps are catalyzed by a single gene or multiple paralogous genes. Studies with soybean cell culture microsomes suggested the involvement of cytochrome P450 (CYP) enzyme(s) capable of converting 4-glyceollidin (**8**) and glyceocarpin (**9**) into glyceollins I (**10**), II (**11**), and III (**12**).[30]

In this study, we elucidated the complete biosynthetic pathway for gyceollins I, II, and III by identifying two reductases responsible for synthesizing THI and five CYP enzymes catalyzing the final

steps of glyceollin biosynthesis. We confirmed these findings through both heterologous expression in *N. benthamiana* and *in vitro* enzyme assays. Leveraging these insights, we successfully reconstituted the entire glyceollin biosynthetic pathways in *Saccharomyces cerevisiae* (Baker's yeast), enabling *de novo* production of glyceollins from simple carbon sources. This platform provides a foundation for optimizing glyceollin production at an industrial scale, paving the way for future agricultural and biological investigations.

Results

Identification and characterization of THI synthase (THIS)

The enzyme responsible for converting 2'-OH dihydrodaidzein (**4**) to THI (**5**) in soybeans remains unidentified. Similar reductive reactions have been reported with vestitone reductase (*Ms*VR) from *Medicago sativa* and sophorol reductase (*Ps*SR) from *Pisum sativum* (**Supplementary Fig. 1**).[31,32] To identify candidate genes encoding this reductase, we performed a BLAST search against the *Glycine max* genome using the protein sequences of *Ms*VR as a query. The search revealed seven candidate genes with sequence identities to *Ms*VR ranging from 63.6% to 79.2% **(Supplementary Table 1)**. Among these, two genes, reductase1 and reductase2, which shared 97.6% identity, were cloned from Triton X-100 treated soybean roots [33] into a binary vector for plant expression. To characterize their function, the genes were transiently expressed in *N. benthamiana* leaves, followed by infiltration of synthetic substrate **4**. Analysis of crude leaf extracts using quadrupole-time-of-flight (QTOF)-HRMS revealed the production of THI (**5**) and 3,9-OH pterocarpan (**6**), based on retention time and mass fragmentation patterns (**Figure 2a, 2b, Supplementary Figure 2**). Notably, it has been reported that the microsomal supernatant from elicitor-induced soybean cell cultures could catalyze the direct conversion of **4** to **6**.[34] To determine whether reductase1 and reductase2 might be bifunctional enzymes capable of catalyzing both steps (**4** → **5** → **6**), synthetic intermediate **5** was infiltrated into the *N. benthamiana* leaves that expressed an empty vector. Successful detection of 3, 9-OH pterocarpan (**6**) suggests that **5** could be spontaneously converted to **6** within the cellular environment or through the action of an endogenous *N. benthamiana* enzyme (**Figure 2a, 2b**). To further investigate, we expressed and purified N-terminal His6-tagged reductase1 and reductase2 proteins for *in vitro* enzyme assays. These assays confirmed the formation of **5**, with no detectable formation of **6**, suggesting that reductase1 and reductase2 are specifically responsible for the reduction of **4** to **5** (**Figure 2a, 2c**). These enzymes were therefore named THI synthase 1 (*Gm*THIS1) and THI synthase 2 (*Gm*THIS2). Additionally, both *Ms*VR and *Ps*SR were shown to catalyze the conversion of **4** to **5** in *N. benthamiana*

transient expression system and *in vitro* assays using purified enzymes (**Supplementary Figure 3**). However, chromatograms revealed undesired and uncharacterized byproducts, indicating possible nonspecific activities.

Identification and characterization of glyceollin I synthases (GISs).

The final steps in the biosynthesis of glyceollins I, II, and III are proposed to involve oxidation of the prenyl unit followed by cyclization catalyzed by cytochrome P450 enzymes, to form 5- and 6-membered rings.[30] As most biosynthetic genes are up-regulated upon infection by *Phytophthora sojae*, we used publicly available transcriptomic data from soybean hairy roots treated with a wall glucan elicitor from *P. sojae*[35,36] to identify candidate CYP genes. From this dataset, 28 upregulated CYP genes with expression levels at least two-fold higher than those of the control were selected for further characterizations **(Supplementary Table 2)**. We cloned these genes from Triton X-100-induced soybean cDNA and transiently expressed them in *N. benthamiana* leaves. When two of these candidates, CYP1 and CYP2 were expressed in the presence of synthetic 4-glyceollidin (**8**), a new peak with an *m/z* $(M+H)^+$ = 339.12 and retention time (Rt) = 6.57 was observed. This peak, absent in the empty vector control, along with a fragment ion at *m/z* $(M+H-H_2O)^+$ = 321.11, matched the synthetic standard of glyceollin I (**10**) (**Figure 3a, Supplementary Figure 4**). Therefore, we named these two CYPs as glyceollin I synthase 1 and 2 (*Gm*GIS1 and *Gm*GIS2). To confirm their functions, we conducted feeding experiments in *S. cerevisiae* WAT11 strain using substrate **8**. While *Gm*GIS1 and *Gm*GIS2 share 87.3% protein sequence identity and exhibited comparable activity in *N. benthamiana*, glyceollin I (**10**) was undetectable when *Gm*GIS1 was expressed in *S. cerevisiae* (**Figure 3a**).

We identified a homolog of *Gm*GIS1/2 in the soybean genome (**Supplementary Figure 5**), CYP3, which is downregulated upon WGE treatment (**Supplementary Table 2**) and could not be cloned from cDNA. To investigate its potential function, we synthesized the gene and transiently expressed it in *N. benthamiana*. Surprisingly, CYP3 (designated as *Gm*GIS3) exhibited significantly lower activity compared with *Gm*GIS1/2 in this system (**Figure 3a**). Protein sequences comparison of *Gm*GIS1, *Gm*GIS2, and *Gm*GIS3 revealed amino acids variations at 15 positions, with *Gm*GIS3 differing from the other two (**Supplementary Figure 5**). Homology models revealed that *Gm*GIS3 possess three distinct residues within their active sites (V120A, S307A, P312S) compared with *Gm*GIS1 and *Gm*GIS2 (**Figure 3b**). These differences may contribute to the decreased activity observed in *Gm*GIS3. We hypothesized these natural mutations to have altered the volume of the active site pocket, affecting substrate accessibility and the reaction process. To test this hypothesis, we performed site-directed

mutagenesis to replace V120, S307, and P312 in *Gm*GIS3 with residues found in *Gm*GIS1/2 (A120, A307, and S312) either individually or in different combinations (**Figure 3b**). Transient expression of the mutants in tobacco has shown that the single-site mutant *Gm*GIS3_P312S exhibits a two-fold increase in product formation compared to the wild-type *Gm*GIS3. In contrast, the mutant *Gm*GIS3_V120A demonstrates a six-fold increase, while the mutant GmGIS3_S307A displayed a five-fold enhancement. Notably, the dual-site mutant GmGIS3_V120A-S307A exhibits an approximately eight-fold increase in product formation compared to the wild-type *Gm*GIS3. (**Figure 3c**). Finally, the triple-site mutation (*Gm*GIS3_V120A-S307A-P312S) also enhanced product formation, although to a lesser extent than the *Gm*GIS3_V120A-S307A variant., Thses results highlighting the crucial role of these residues in optimizing enzymatic activity.

Identification and characterization of glyceollin II (GIIS) and III (GIIIS) synthases.

Glyceollins II (**11**) and III (**12**) are isomeric pterocarpinoid phytoalexins that share a characteristic 6α-hydroxy-pterocarpan ring system. The distinction between these two compounds lies in the arrangement of the C2 prenyl group in the biosynthetic precursor glyceocarpin (**9**), which leads to the formation of distinct heterocyclic structures: a dihydropyran ring in glyceollin II (**11**) and a hydrofuran ring in glyceollin III (**12**). Transient expression of CYP4 with synthetic glyceocarpin (**9**) as a substrate revealed a peak at Rt = 6.48 min that co-eluted with the synthetic standard of glyceollin II (**11**) (**Figure 3d, Supplementary Figure 6**). Therefore, CYP4 was named as glyceollin II synthase (*Gm*GIIS). Transient expression of CYP5 in *N. benthamiana*, combined with infiltration of glyceocarpin (**9**), resulted in the emergence of a novel peak at Rt = 6.25 min, distinct from those of glyceollins I and II. Mass spectral analysis of this peak revealed a fragmentation pattern consistent with glyceollin I or II, showing a parent ion at *m/z* $(M+H)^+$ = 339.12 and a fragment ion at *m/z* $(M-H_2O+H)^+$ = 321.11 (**Figure 3d**, **Supplementary Figure 7b**). We proposed this compound as glyceollin III, as its retention time corresponds to the peak detected in LC-MS data from elicitor-induced soybean roots (**Supplementary Figure 7, 8**) and aligned with previously reported data in the literature.[35,37] Accordingly, we designated CYP5 as glyceollin III synthase (*Gm*GIIIS). To further confirm their functions, we conducted feeding experiments in *S. cerevisiae*, which demonstrated their respective roles in the biosynthesis of glyceollins II and III (**Figure 3d**).

The soybean genome contains a homolog of *Gm*GIIIS, CYP6, which is upregulated in response to *P. sojae* infection (**Supplementary Table 2**). However, transient expression of CYP6 in *N. benthamiana* leaves revealed no detectable activity, indicating its function as a pseudo-glyceollin III synthase and

was designated *Gm*GIIISp. To identify candidate amino acid residues critical for the inactivity of *Gm*GIIISp, we generated a structural model of the enzyme. We found four amino acids in the binding pocket to differ between *Gm*GIIISp and *Gm*GIIIS (**Figure 3e**). These residues from *Gm*GIIIS were introduced into *Gm*GIIISp in attempts to recover its activity. We performed assays by transient expression of the resulting mutants in *N. benthamiana* leaves with the co-infiltration of **9**. Single mutation, L122V or G382A, failed to restore activity. However, mutagenesis of residues 496-497 (GV to RA) restored approximately 10% of *Gm*GIIIS activity, indicating the importance of these two amino acids for enzymatic function (**Figure 3f**). Indeed, reverse mutations at positions 497-498 (RA497-498GV) in *Gm*GIIIS resulted in a loss of function (**Figure 3f**). Further investigation involved additional mutations to the *Gm*GIIISp-GV496-497RA mutant. Introducing the L122V mutation abolished the residual activity, whereas substituting G382A instead of L122V slightly improved the activity compared to *Gm*GIIISp-GV496-497RA. Finally, combing all four mutations (G496R, V497A, L122V and G382A) into *Gm*GIIISp, resulted in the quadruple mutant achieving more than 40% of the activity of *Gm*GIIIS (**Figure 3f**). These results demonstrate that the naturally inactive *Gm*GIIISp can regain substantial enzymatic activity through targeted mutations at the predicted critical residues, providing insights into the structural basis of enzyme functionality. While this manuscript was in preparation, an independent study reported the enzymatic activities of *Gm*GIS1/2 and *Gm*GIIIS through *in vitro* assays and supported by *in planta* gene silencing experiments.[37]

***De novo* biosynthesis of daidzein in yeast.**

Following the complete elucidation of the biosynthetic pathway of glyceollins, we aimed to achieve *de novo* biosynthesis of glyceollins in *S. cerevisiae*. Considering the lengthy pathway (16 genes) and the presence of a rate-limiting prenylation step,[38-40] we segmented the whole pathway into three modules: the daidzein module (from *L*-tyrosine to daidzein), the glycinol module (from daidzein to glycinol), and the glyceollin module (from glycinol to glyceollins). Unlike the biosynthetic pathway in soybean, which starts from *L*-phenylalanine, we reconstituted the pathway in yeast starting from L-tyrosine, as it already contains a 4-OH group.

For the daidzein module (**Figure 4a**), we first redirected the metabolic flux toward *L*-tyrosine by integrating copies of $ARO4^{K229L}$ (the feedback inhibition-insensitive 3-deoxy-D-arabino-heptulosonic acid 7-phosphate synthase mutant) and $ARO7^{G141S}$ (the feedback inhibition-insensitive mutant of chorismate mutase) while knocking out a competing enzyme-encoding gene *ARO10* (phenylpyruvate decarboxylase). We subsequently inserted expression cassettes containing *FjTAL* (tyrosine ammonia

lyase from *Flavobacterium johnsoniae*), *Pc4CL2* (4-coumaroyl-CoA synthase 2 from *Petroselinum crispum*), *PhCHS~GmCHR5* (*Petunia hybrida* chalcone synthase fused with *Glycine max* chalcone reductase CHR5), *GmCHR5* (an extra copy of chalcone reductase CHR5 from *G. max*), and *PhCHI* (chalcone isomerase from *Petunia hybrida*). The resultant strain GYN1 produced 17.12 mg/L isoliquiritigenin, 0.64 mg/L liquiritigenin, and 5.62 mg/L naringenin, respectively (**Figure 4b**). The inefficiency of isoliquiritigenin isomerization prompted us to explore other CHI orthologs with higher activity. Liu et al. previously achieved 77.7 mg/L liquiritigenin production using *GmCHIB2* (chalcone isomerase from *G. max*).[41] Consequently, we integrated *GmCHIB2* into GYN1, resulting in the construction of strain GYN2 with a notable increase in liquiritigenin titer (24.57 mg/L, **Figure 4b**). To construct a daidzein-producing strain, we further introduced two downstream genes, *2-HIS* (2-hydroxyisoflavanone synthase) and *HID* (2-hydroxyisoflavanone dehydratase), into GYN2 to generate strains GYN3 and GYN4. The former included an optimized combination of *Ge*2-*HIS* (*Glycyrrhiza echinata*) and *GmHID* (*G. max*) as previously reported,[41] while the latter incorporated *AmIFS* (isoflavone synthase from *Astragalus membranaceus*) and *GmHID*. After one day of induction with galactose, both strains exhibited similar daidzein (**3**) titers of approximately 24 mg/L, reflecting comparable catalytic efficiency between *Ge*2-HIS and *Am*IFS (**Figure 4b**). Therefore, we chose GYN3 as our platform strain for subsequent studies.

***De novo* biosynthesis of glycinol in yeast**

In the glycinol module (**Figure 5**, **Supplementary Figure 9**), we sequentially integrated *GmCYP81E*, *GmIFR*, *GmTHIS1*, and *GmPTS* expression cassettes into GYN3, resulting in the construction of strains GYN5, GYN6, GYN7, and GYN8 **(Supplementary Table 3)**. As shown in **Figure 5**, most of the daidzein (**2**) in GYN5 was converted to 2'-OH daidzein (**3**), which was subsequently transformed into 2'-OH dihydrodaidzein (**4**) in GYN6. However, the conversion ratio of 2'-OH daidzein (**3**) remained below 50%. Notably, in GYN7, we detected both 7,2',4'-trihydroxyisoflavanol (**5**) and 3,9-OH pterocarpan (**6**), consistent with the results observed in *N. benthamiana* (**Figure 2**). With the introduction of *Gm*PTS, we observed a higher abundance of 3,9-OH pterocarpan (**6**) in GYN8. Ultimately, we integrated *GmCYP93A* expression cassette into GYN8 to construct strain GYN9, which converted more than 92% of 3,9-OH pterocarpan (**6**) into glycinol (**7**). Given the substantial abundance of glycinol (**7**) in yeast, we proceeded to reconstitute the glyceollin module.

***De novo* biosynthesis of glyceollins in yeast.**

The glyceollin module encompasses two steps: prenylation and cyclization (**Figure 6a**). Akashi et al. and Yoneyama et al. demonstrated that *Gm*G4DT and *Gm*G2DT catalyzed the prenylation of glycinol.[27,28] Unfortunately, when we introduced both prenyltransferases (PTs) in their full lengths **(Supplementary Table 3)** to construct strains GYN11 and GYN10, we failed to detect the corresponding prenylated products, 4-glyceollidin (**8**) and glyceocarpin (**9**), in both strains (**Supplementary Figure 10**). Previous studies suggested that the signal peptide of prenyltransferases might hinder their function in yeast.[38,40,42] Therefore, we truncated 44 amino acids from the *N*-terminus of *Gm*G2DT and *Gm*G4DT, leading to the construction of strain GYN12 with truncated *GmG4DT* (*tGmG4DT*) and GYN13 with truncated *GmG2DT* (*tGmG2DT*).[27,28] However, prenylated products remained undetectable in these strains (**Supplementary Figure 10**), indicating that prenylation is a major rate-limiting step in our engineered yeast strains.

To discern whether instability of prenylated products in yeast contributes to their undetectability, we introduced the last genes, *GmGIS1*, *GmGIS2*, *GmGIIS*, and *GmGIIIS*, resulting the construction of strains GYN14 to GYN17 (Supplementary Table 3), to synthesize the end products (glyceollins I/II/III). However, we could not extract the ion pairs corresponding to any of the end products in these four strains (**Supplementary Figure 11**). When we fed 4-glyceollidin (**8**) or glyceocarpin (**9**) at a final concentration of 1 mg/L, we could detect all three types of glyceollins by LC-MS, except for the strain containing *Gm*GIS1, despite the titer of glyceollin I catalyzed by *Gm*GIS2 remaining low (**Supplementary Figure 12**). These results suggested that *Gm*GIS2, *Gm*GIIS, and *Gm*GIIIS were functional in yeast and the ultra-low titer of the prenylated products is the primary issue.

In the mevalonate pathway, ERG20 is a crucial node preferentially catalyzing dimethylallyl diphosphate (DMAPP) and isopentenyl pyrophosphate (IPP) to form farnesyl pyrophosphate (FPP) in *S. cerevisiae* (**Figure 6a**). It has been reported that decreasing ERG20 expression could enhance the biosynthesis of prenylated products.[38-40,43-44] To evaluate whether the native ERG20 in yeast impaired the prenylation of glycinol (**7**), we transformed both PTs and their truncated versions into strain AM3-1,[45] where the native promoter of *ERG20* (*ERG20p*) was replaced with a dynamically regulated promoter *HXT1p* **(Supplementary Table 3)**. When we fed 0.5 mg/L glycinol (**7**) into AM3-1 expressing PTs, we were able to detect 4-glyceollidin (**8**) and glyceocarpin (**9**), with the truncated PTs demonstrating superior efficiency (**Supplementary Figure 13**). Thus, we replaced the native *ERG20p* in GYN12 and GYN13 with *HXT1p*, generating yeast strains GYN18 and GYN19, respectively (**Supplementary Table 3**). Consequently, we observed a notable chromatographic peak, with the

retention time matching that of the authentic standards for 4-glyceollidin (**8**) and glyceocarpin (**9**), although the ion abundance remained relatively low (**Figure 6b**).

Regarding cyclization, when we integrated the final genes into GYN18 and GYN19 to construct strains GYN20 with *GmGIS2*, GYN21 with *GmGIIS*, and GYN22 with *GmGIIIS* (**Supplementary Table 3**), we again failed to detect all three glyceollins (**Supplementary Figure 14**). In a previous study, Yang et al. demonstrated that fusing IDI1 with PT could significantly enhance the titers of prenylated products.[40] Accordingly, we constructed strains GYN23, GYN24, and GYN25 by increasing a copy of *t*HMG1 and fusing IDI1 with PTs in the relevant strains (Supplementary Table 3). As a result, the titers of 4-glyceollidin (**8**) and glyceocarpin (**9**) were markedly increased by approximately 22- and 9-fold, respectively (**Figure 6b, c**).

Finally, we successfully detected glyceollin III with an estimated titer of 0.06-0.17 mg/L (based on synthetic standards of glyceollins I and II) in strain GYN25 (**Figure 6d, Supplementary Figure 15**). In strain GYN24, we could also detect glyceollin II, although its titer (~1.1 μg/L) was much lower than that of glyceollin III (**Figure 6d, Supplementary Figure 16**). In strain GYN23, the titer of glyceollin I was significantly lower (~0.04 μg/L) than that of glyceollin II and glyceollin III (**Figure 6d, Supplementary Figure 17**), which was consistent with the feeding experiments (**Supplementary Figure 12**). These results indicated that the three P450s were also rate-limiting steps for *de novo* biosynthesis of glyceollins, especially for *Gm*GIS2 and *Gm*GIIS.

Discussion

This study successfully elucidates the complete biosynthetic pathway for glyceollins I, II, and III, resolving long-standing gaps in the understanding of these critical soybean phytoalexins. By identifying and characterizing THI synthase and multiple cytochrome P450 enzymes, we have elucidated the precise molecular basis of glyceollin biosynthesis. Importantly, this work establishes a *de novo* production platform in *S. cerevisiae*, marking a significant milestone toward the sustainable and scalable production of these biologically active compounds.

Our findings reveal that *Gm*THIS1 and *Gm*THIS2 catalyze the conversion of 2'-hydroxy-dihydrodaidzein (**4**) to THI (**5**), a reaction previously hypothesized but not experimentally demonstrated. These enzymes exhibit high specificity toward 2'-OH dihydrodaidzein (**4**), distinguishing them from related vestitone and sophorol reductases, which displayed non-specific activity and byproduct formation. The functional characterization of cytochrome P450 enzymes,

including *Gm*GIS1, *Gm*GIS2, and their paralog *Gm*GIS3, reveals their pivotal roles in the final oxidative cyclization leading to glyceollin I (**10**). Notably, the decreased activity of *Gm*GIS3 compared to *Gm*GIS1 and *Gm*GIS2 highlights functional divergence among these paralogs. Restoration of *Gm*GIS3 activity through site-directed mutagenesis underscores the importance of active site architecture in determining enzyme functionality and efficiency. Furthermore, the identification of glyceollins II and III synthases (*Gm*GIIS and *Gm*GIIIS) demonstrates the diversification of cytochrome P450 enzymes in soybean. The structural basis for their regioselective cyclization of glyceocarpin to distinct glyceollin isomers reflects the intricate evolution of enzyme-substrate specificity. Interestingly, *Gm*GIIISp, a homolog of *Gm*GIIIS, functions as a pseudo-enzyme due to critical amino acid substitutions in its active site, providing new insights into the evolutionary trajectory of glyceollin biosynthetic genes. Collectively, these findings offer a molecular basis for further enzyme engineering to enhance glyceollin production.

Through the reconstitution and optimization of the glyceollin biosynthetic pathway in a modular manner, we achieved *de novo* biosynthesis of glyceollins I, II, and III from simple carbon sources for the first time. Nevertheless, the titers of glyceollins are still far from satisfaction, particularly for glyceollin I and glyceollin II, indicating the necessity for further pathway optimization and metabolic engineering. Considering the accumulation of prenylated products in our engineered yeast strain and variation in the titer of glyceollins, the last step enzymes (*Gm*GIS2/*Gm*GIIS/*Gm*GIIIS) become the major bottleneck for efficient biosynthesis. Therefore, protein engineering and chassis engineering should be performed to enhance the expression, folding, and/or activity of P450 enzymes.[46] For example, *de novo* protein design and amino acid co-evolutionary information could be combined for molecular engineering of P450 enzymes in a structural and data-driven approach.[47] In addition, the microenvironment for P450 activities could be systematically optimized, including overexpressing the redox partner genes, enhancing NADPH supply, expanding the endoplasmic reticulum, as well as regulating iron and heme homeostasis.[47,48] The successful reconstitution and potentially further optimization of the glyceollins pathway in yeast demonstrates the feasibility of bulk production of these compounds for functional assays and biotechnological applications.

In conclusion, our study provides a comprehensive framework for understanding glyceollin biosynthesis. By bridging gaps in pathway elucidation and translating these insights into synthetic biology, we have unlocked new opportunities for the sustainable production and utilization of these valuable natural products in agriculture, medicine, and biotechnology.

References

(1) Prayogo, Y.; Bayu, M.; Susanto, G. W. A.; Indiati, S. W. Impact of Biopesticide Inundation on the Diversity of Soybean Pests and Diseases. *IOP Conf. Ser.: Earth Environ. Sci.* **2021**, *743*, 012003.

(2) Rao, W.; Wan, L.; Wang, E. Plant Immunity in Soybean: Progress, Strategies, and Perspectives. *Mol. Breed.* **2023**, *43*, 1-12.

(3) Whitham, S. A.; Qi, M.; Innes, R. W.; Ma, W.; Lopes-Caitar, V.; Hewezi, T. Molecular Soybean-Pathogen Interactions. *Annu. Rev. Phytopathol.* **2016**, *54*, 443-468.

(4) Burden, R. S.; Bailey, J. A. Structure of the Phytoalexin from Soybean. *Phytochemistry* **1975**, *14*, 1389-1390.

(5) Lyne, R. L.; Mulheirn, L. J.; Leworthy, D. P. New Pterocarpinoid Phytoalexins of Soybean; *J. Chem. Soc., Chem. Commun.*, **1976**, 497

(6) Kim, H. J.; Suh, H. J.; Lee, C. H.; Kim, J. H.; Kang, S. C.; Park, S.; Kim, J. S. Antifungal Activity of Glyceollins Isolated from Soybean Elicited with Aspergillus Sojae. *J. Agric. Food Chem.* **2010**, *58*, 9483-9487.

(7) Araya-Cloutier, C.; Vincken, J. P.; Van De Schans, M. G. M.; Hageman, J.; Schaftenaar, G.; Den Besten, H. M. W.; Grup-pen, H. QSAR-Based Molecular Signatures of Prenylated (Iso)Flavonoids Underlying Antimicrobial Potency against and Membrane-Disruption in Gram Positive and Gram Negative Bacteria. *Sci. Rep.* **2018**, *8*, 9267.

(8) Bamji, S. F.; Corbitt, C. Glyceollins: Soybean Phytoalexins That Exhibit a Wide Range of Health-Promoting Effects. Journal of Functional Foods. Elsevier Ltd July 1, **2017**, pp 98-105.

(9) Pham, T. H.; Lecomte, S.; Efstathiou, T.; Ferriere, F.; Pakdel, F. An Update on the Effects of Glyceollins on Human Health: Possible Anticancer Effects and Underlying Mechanisms. *Nutrients* **2019**, *11*, 79.

(10) Yue, Z.; He, S.; Wang, J.; Jiang, Q.; Wang, H.; Wu, J.; Li, C.; Wang, Z.; He, X.; Jia, N. Glyceollins from Soybean: Their Pharma-cological Effects and Biosynthetic Pathways. *Heliyon* **2023**, *9*, e21874.

(11) Lecomte, S.; Chalmel, F.; Ferriere, F.; Percevault, F.; Plu, N.; Saligaut, C.; Surel, C.; Lelong, M.; Efstathiou, T.; Pakdel, F. Glyceollins Trigger Anti-Proliferative Effects through Estradiol-Dependent and Independent Pathways in Breast Cancer Cells. *Cell Commun. Signal.* **2017**, *15*, 26.

(12) Salvo, V. A.; Boué, S. M.; Fonseca, J. P.; Elliott, S.; Corbitt, C.; Collins-Burow, B. M.; Curiel, T. J.; Srivastav, S. K.; Shih, B. Y.; Carter-Wientjes, C.; Wood, C. E.; Erhardt, P. W.; Beckman, B. S.;

McLachlan, J. A.; Cleveland, T. E.; Burow, M. E. Antiestrogenic Glyceollins Suppress Human Breast and Ovarian Carcinoma Tumorigenesis. *Clin. Cancer Res.* **2006**, *12*, 7159-7164.

(13) Zimmermann, M. C.; Tilghman, S. L.; Boué, S. M.; Salvo, V. A.; Elliott, S.; Williams, K. Y.; Skripnikova, E. V.; Ashe, H.; Payton-Stewart, F.; VanhoyRhodes, L.; Fonseca, J. P.; Corbitt, C.; Collins-Burow, B. M.; Howell, M. H.; Lacey, M.; Shih, B. Y.; Carter-Wientjes, C.; Cleveland, T. E.; McLachlan, J. A.; Wiese, T. E.; Beckman, B. S.; Burow, M. E. Glyceollin I, a Novel Antiestrogenic Phytoalexin Isolated from Activated Soy. *J. Pharmacol. Exp. Ther.* **2010**, *332*, 35-45.

(14) Payton-Stewart, F.; Khupse, R. S.; Boué, S. M.; Elliott, S.; Zimmermann, M. C.; Skripnikova, E. V.; Ashe, H.; Tilghman, S. L.; Beckman, B. S.; Cleveland, T. E.; McLachlan, J. A.; Bhatnagar, D.; Wiese, T. E.; Erhardt, P.; Burow, M. E. Glyceollin I Enantiomers Distinctly Regulate ER-Mediated Gene Expression. *Steroids* **2010**, *7*, 870-878.

(15) Medic, J.; Atkinson, C.; Hurburgh, C. R. Current Knowledge in Soybean Composition. *J. Am. Oil Chem. Soc.* **2014**, *91*, 363-384.

(16) Luniwal, A.; Khupse, R.; Reese, M.; Liu, J.; ElDakdouki, M.; Malik, N.; Fang, L.; Erhardt, P. Multigram Synthesis of Glyceollin I. *Org. Process. Res. Dev.* **2011**, *15*, 1149-1162.

(17) Ciesielski, P.; Metz, P. Asymmetric One-Pot Transformation of Isoflavones to Pterocarpans and Its Application in Phytoalexin Synthesis. *Nat. Commun*. **2020** *11*. 3091

(18) Kohno, Y.; Koso, M.; Kuse, M.; Takikawa, H. Formal Synthesis of Soybean Phytoalexin Glyceollin I. *Tetrahedron Lett.* **2014**, *55*, 1826-1828.

(19) Khupse, R. S.; Erhardt, P. W. Total Syntheses of Racemic, Natural and Unnatural (+) Glyceollin I. *Org. Lett.* **2008**, *10*, 5007-5010.

(20) Malik, N.; Zhang, Z.; Erhardt, P. Total Synthesis of (±)-Glyceollin II and a Dihydro Derivative. *J. Nat. Prod*. **2015**, *78*, 2940-2947.

(21) Khupse, R. S.; Sarver, J. G.; Trendel, J. A.; Bearss, N. R.; Reese, M. D.; Wiese, T. E.; Boue, S. M.; Burow, M. E.; Cleveland, T. E.; Bhatnagar, D.; Erhardt, P. W. Biomimetic Syntheses and Antiproliferative Activities of Racemic, Natural (-), and Unnnatural (+)-Glyceollin I. *J. Med. Chem.* **2011**, *54*, 3506-3523.

(22) Sims, J. J.; Keen, N. T.; Honwad, V. K. Hydroxyphaseollin, an Induced Antifungal Compound from Soybeans. *Phytochemistry* **1972**, *11*, 827-828.

(23) Uchida, K.; Akashi, T.; Aoki, T. Functional Expression of Cytochrome P450 in Escherichia Coli: An Approach to Functional Analysis of Uncharacterized Enzymes for Flavonoid Biosynthesis. *Plant Biotechnol.* **2015**, *32*, 205-213.

(24) Fischer, D.; Ebenau-Jehle, C.; Grisebach, H. Phytoalexin Synthesis in Soybean: Purification and Characterization of NADPH:2′-Hydroxydaidzein Oxidoreductase from Elicitor-Challenged Soybean Cell Cultures. *Arch. Biochem. Biophys.* **1990**, *276*, 390-395.

(25) Uchida, K.; Akashi, T.; Aoki, T. The Missing Link in Leguminous Pterocarpan Biosynthesis Is a Dirigent Domain-Containing Protein with Isoflavanol Dehydratase Activity. *Plant Cell Physiol* **2017**, *58*, 398-408.

(26) Schopfer, C. R.; Kochs, G.; Lottspeich, F.; Ebel, J. Molecular Characterization and Functional Expression of Dihydroxyptero-carpan 6a-Hydroxylase, an Enzyme Specific for Pterocarpanoid Phytoalexin Biosyn-thesis in Soybean (*Glycine Max L.*). *FEBS Lett.* **1998**, *432*, 182-186.

(27) Yoneyama, K.; Akashi, T.; and Aoki, T. Molecular Characterization of Soybean Pterocarpan 2-Dimethylallyltransferase in Glyceollin Biosynthesis: Local Gene and Whole-Genome Duplications of Prenyltransferase Genes Led to the Structural Diversity of Soybean Prenylated Isoflavonoids. *Plant Cell Physiol.* **2016**, *57*, 2497-2509.

(28) Akashi, T.; Sasaki, K.; Aoki, T.; Ayabe, S.; and Yazaki, K. Molecular Cloning and Characterization of a cDNA for Pterocarpan 4-Dimethylallyltransferase Catalyzing the Key Prenylation Step in the Biosynthesis of Glyceollin, a Soybean Phytoalexin. *Plant Physiol*. **2009**, *149*, 683-693.

(29) Sukumaran, A.; McDowell, T.; Chen, L.; Renaud, J.; Dhaubhadel, S. Isoflavonoid-Specific Prenyltransferase Gene Family in Soybean: GmPT01, a Pterocarpan 2-Dimethylallyltransferase Involved in Glyceollin Biosynthesis. *Plant J.***2018**, *96*, 966-981.

(30) Welle, R.; Grisebach, H. Induction of Phytoalexin Synthesis in Soybean: Enzymatic Cyclization of Prenylated Pterocarpans to Glyceollin Isomers. *Arch. Biochem. Biophys.* **1988**, *263*, 191-198.

(31) Guo, L.; Dixon, R. A.; Paiva, N. L. Con-version of Vestitone to Medicarpin in Alfalfa (Medicago Sativa L.) Is Catalyzed by Two Independent Enzymes: Identification, Purification, and Characterization of Vestitone Reductase and 7,2′-Dihydroxy-4′-Methoxyisoflavanol Dehydratase. *J. Bio. Chem.* **1994**, *269*, 22372-22378.

(32) DiCenzo, G. L.; VanEtten, H. D. Studies on the Late Steps of (+)-Pisatin Biosynthesis: Evidence for (-) Enantiomeric Intermediates. *Phytochemistry* **2006**, *67*, 675-683.

(33) Yoshikawa, M. Diverse modes of action of biotic and abiotic phytoalexin elicitors. *Nature* **1978**, **275**, 546-547.

(34) Fischer, D.; Ebenau-Jehle, C.; Grisebach, H. Purification and Characterization of Pterocarpan Synthase from Elicitor-Challenged Soybean Cell Cultures. *Phytochemistry* **1990**, *29*, 2879–2882.

(35) Jahan, M. A.; Harris, B.; Lowery, M.; Infante, A. M.; Percifield, R. J.; Kovinich, N. Glyceollin Transcription Factor GmMYB29A2 Regulates Soybean Resistance to *Phytophthora Sojae*. *Plant Physiol.* **2020**, *183*, 530-546.

(36) Khatri, P.; Wally, O.; Rajcan, I.; Dhaubhadel, S. Comprehensive Analysis of Cytochrome P450 Monooxygenases Reveals Insight Into Their Role in Partial Resistance Against *Phytophthora sojae* in Soybean. *Front. Plant Sci.* **2022**, *13*, 862314.

(37) Khatri, P., Kuflu, K., McDowell, T., Lin, J., Kovinich, N., and Dhaubhadel, S. Discovery of the cytochrome P450 monooxygenase prenyl cyclases that catalyze the final step of glyceollin biosynthesis in soybean *Mol. Plant*, **2024**, doi: https://doi.org/10.1016/j.molp.2025.01.022.

(38) Li, H.; Ban, Z.; Qin, H.; Ma, L.; King, A. J.; Wang, G. A Heteromeric Membrane-Bound Prenyltransferase Complex from Hop Catalyzes Three Sequential Aromatic Prenylations in the Bitter Acid Pathway. *Plant Physiol.* **2015**, *167*, 650-659.

(39) Luo, X.; Reiter, M. A.; d'Espaux, L.; Wong, J.; Denby, C. M.; Lechner, A.; Zhang, Y.; Grzybowski, A. T.; Harth, S.; Lin, W.; Lee, H.; Yu, C.; Shin, J.; Deng, K.; Benites, V. T.; Wang, G.; Baidoo, E. E. K.; Chen, Y.; Dev, I.; Petzold, C. J.; Keasling, J. D. Complete Biosynthesis of Cannabinoids and Their Unnatural Analogues in Yeast. *Nature* **2019**, *567*, 123-126.

(40) Yang, S.; Chen, R.; Cao, X.; Wang, G.; Zhou, Y. J. *De Novo* Biosynthesis of the Hops Bioactive Flavonoid Xanthohumol in Yeast. *Nat. Commun.* **2024**, *15*, 253

(41) Liu, Q., Liu, Y., Li, G., Savolainen, O., Chen, Y., and Nielsen, J. *De novo* biosynthesis of bioactive isoflavonoids by engineered yeast cell factories. *Nat. Commun.* **2021**, *12*, 6085

(42) Sayed, H. M. B.; Nassar, S.; Kaufholdt, D.; Beerhues, L.; Liu, B.; El-Awaad, I. Biosynthesis of Polyprenylated Xanthones in Hypericum Perforatum Roots Involves 4-Prenyltransferase. *Plant Physiol.* **2023**, *192*, 2971-2988.

(43) Fischer, M. J. C.; Meyer, S.; Claudel, P.; Bergdoll, M.; Karst, F. Metabolic Engineering of Monoterpene Synthesis in Yeast. *Biotechnol. Bioeng.* **2011**, *108*, 1883-1892.

(44) Zhao, J.; Li, C.; Zhang, Y.; Shen, Y.; Hou, J.; Bao, X. Dynamic Control of ERG20 Expression Combined with Minimized Endogenous Downstream Metabolism Contributes to the Improvement of Geraniol Production in Saccharomyces Cerevisiae. *Microb. Cell Fact.* **2017**, *16*, 17.

(45) Liu, T.; Huang, Y.; Jiang, L.; Dong, C.; Gou, Y.; Lian. J. Efficient production of vindoline from tabersonine by metabolically engineered *Saccharomyces cerevisiae*. *Commun. Biology* **2021**, *4*, 1089.

(46) Jiang, L.; Huang, L.; Cai, J.; Xu, Z.; Lian, J. Functional Expression of Eukaryotic Cytochrome P450s in Yeast. *Biotechnol. Bioeng*. **2020**, *118*, 1050-1065.

(47) Li, D.; Ma, Y.; Zhou, Y.; Gou, J.; Zhong, Y.; Zhao, L.; Han, L.; Ovchinnikov, S.; Ma, L.; Huang, S.; Greisen, P.; Shang, Y. A Structural and Data-Driven Approach to Engineering a Plant Cytochrome P450 Enzyme. *Sci. China Life Sci.* **2019**, *62*, 873-882.

(48) Cheng, Y.; Luo, L.; Tang, H.; Wang, J.; Ren, L.; Cui, G.; Zhao, Y.; Tang, J.; Su, P.; Wang, Y.; Hu, Y.; Ma, Y.; Guo, J.; Huang, L. Engineering the Microenvironment of P450s to Enhance the Production of Diterpenoids in Saccharomyces Cerevisiae. *Acta. Pharm. Sin. B* **2024**, *14*, 4608-4618.

Data availability

Data supporting the findings of this work are available within the paper and its Supplemental Information files. The sequence of genes characterized in this article will be deposited in the NCBI database prior to publication. Source data are provided with this paper. Correspondence and requests for materials should be addressed to B.H. or J.L.

Acknowledgements

This research was supported by Zhejiang Provincial Natural Science Foundation of China under Grant No. XHD24B0701 (to B. H.) and National Natural Science Foundation of China under Grant No. 22278361 (to J.Z.). We are grateful for financial support from Westlake University startup, Research Center for Industries of the Future (RCIF) at Westlake University, and Zhejiang Key Laboratory Construction Project (to B. H.). We thank Yuechen Bai (Fudan University) for providing soybean seeds. We thank Westlake University Instrumentation and Service Center for Molecular Sciences (ISCMS) for the facility support and technical assistance. We thank Dr. Yinjuan Chen for assistance with mass spectrometry, Dr. Xingyu Lu and Dr. Xiaohuo Shi for assistance with NMR analysis.

Author Contributions

Y.S. performed the compounds synthesis, cloning, transient expression, protein-expression and protein-assay experiments; C.C. carried out the metabolic engineering in yeast, C.L. performed the

NMR experiments and structural assignments. H.Z. constructed the CYP mutants. J.L. and B.H. designed and directed the work. Y.S., C.C., J.L. and B.H. wrote the manuscript with contributions from all authors. All authors discussed the results and commented on the paper.

Competing interests

The authors declare no competing interests.

Figure Legends

Figure 1: Proposed biosynthetic pathway of glyceollins I, II, and III. Dash arrows represent uncharacterized step, black arrows represent characterized steps.

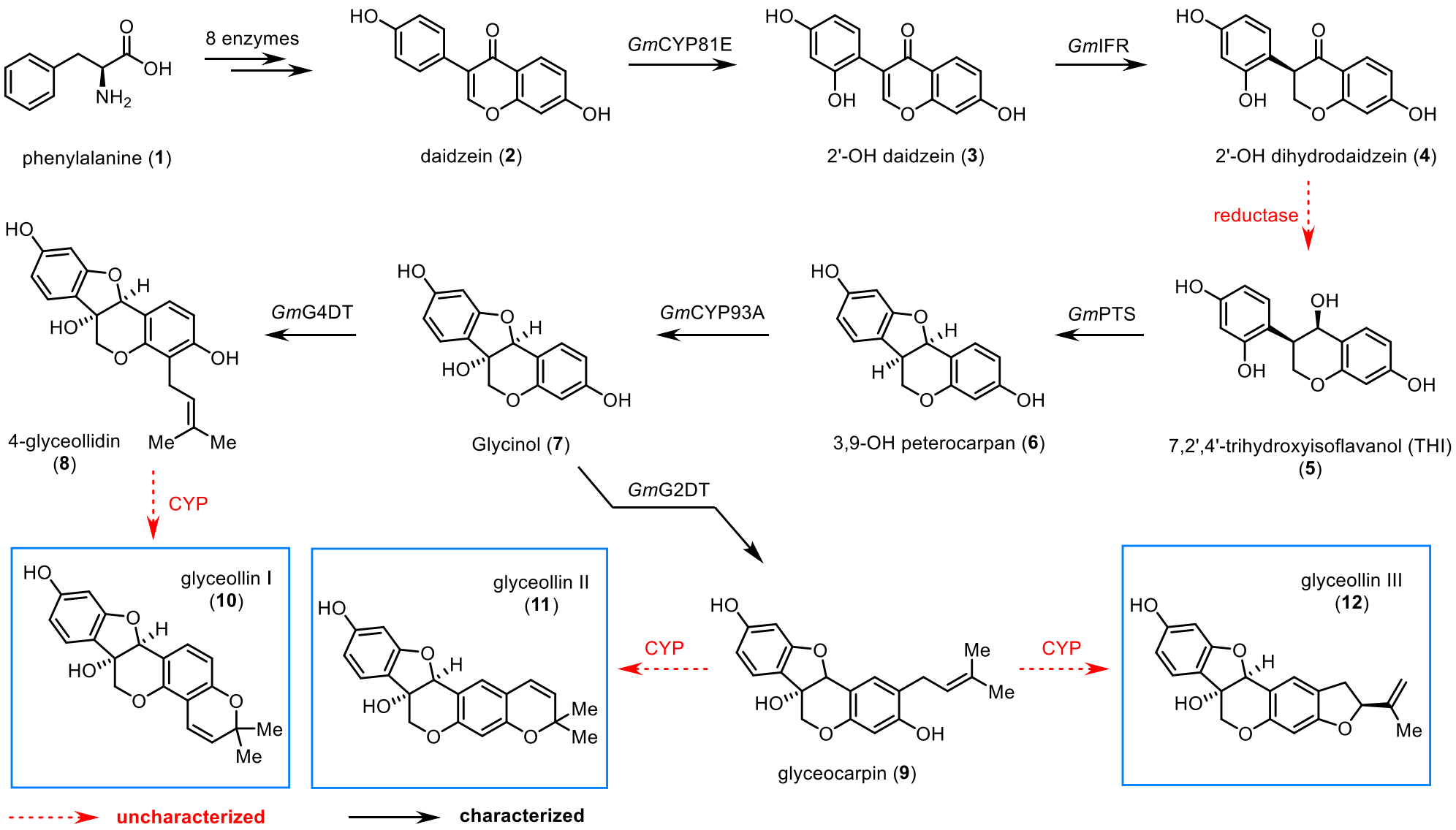
Figure 2: Functional characterization of *Gm*THIS1/2. (a) Scheme of the conversion from **4** to **5**, where the step from **5** to **6** represents a shunt in *N. benthamiana*. (b) Transient expression of *Gm*THIS1/2 with the substrate **4**. Extracted ion chromatograms (EICs) for **4** ($[M-H_2O+H]^+$ = *m/z* 257.0814) and **5** ($[M+H]^+$ = *m/z* 257.0814) are displayed. (c) *In vitro* experiments of *Gm*THIS1/2 using purified proteins. EICs for **4** and **5** are displayed.

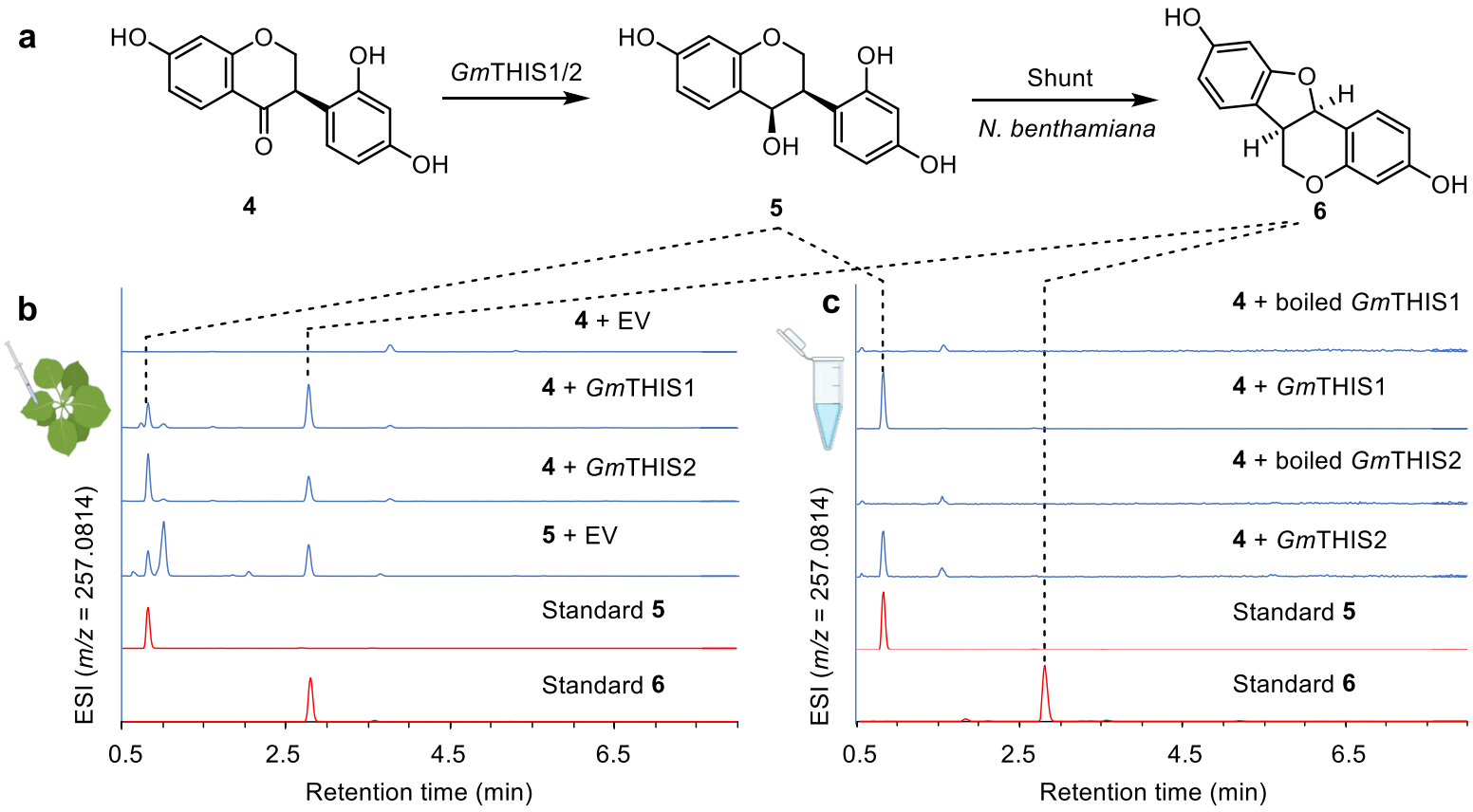
Figure 3: Functional characterization of *Gm*GIS1/2/3 and *Gm*GIIS, *Gm*GIIIS, *Gm*GIIISp. (a) Transient expression in *N. benthamiana* (blue) and feeding experiments in *S. cerevisiae* (black) showing the conversion of **8** to **10**.EIC for **10** ($[M-H_2O+H]^+$ = *m/z* 321.1127) is displayed. (b) A predicted model of *Gm*GIS2 and *Gm*GIS3, highlighting the interacting substrates and cofactors. (c) Comparison of activity for *Gm*GIS2, *Gm*GIS3, and their mutants. Average LC-MS peak area of product **10** are shown (d) Transient expression in *N. benthamiana* (blue) and feeding experiments in *S. cerevisiae* (black), showing *Gm*GIIS and *Gm*GIIIS converting **9** into **11** and **12**, respectively. EICs for **11** ($[M-H_2O+H]^+$ = *m/z* 321.1127) and **12** ($[M-H_2O+H]^+$ = *m/z* 321.1127) are presented (e) A predicted model of *Gm*GIIIS and *Gm*GIIISp, focusing on the substrates and cofactors. (f) Comparison of activity for *Gm*GIIIS, *Gm*GIIISp, and their mutants. Average LC-MS peak area of product **12** are shown. For bar graphs, plotted values represent the mean value ± standard deviations (n = 3 biologically independent samples).

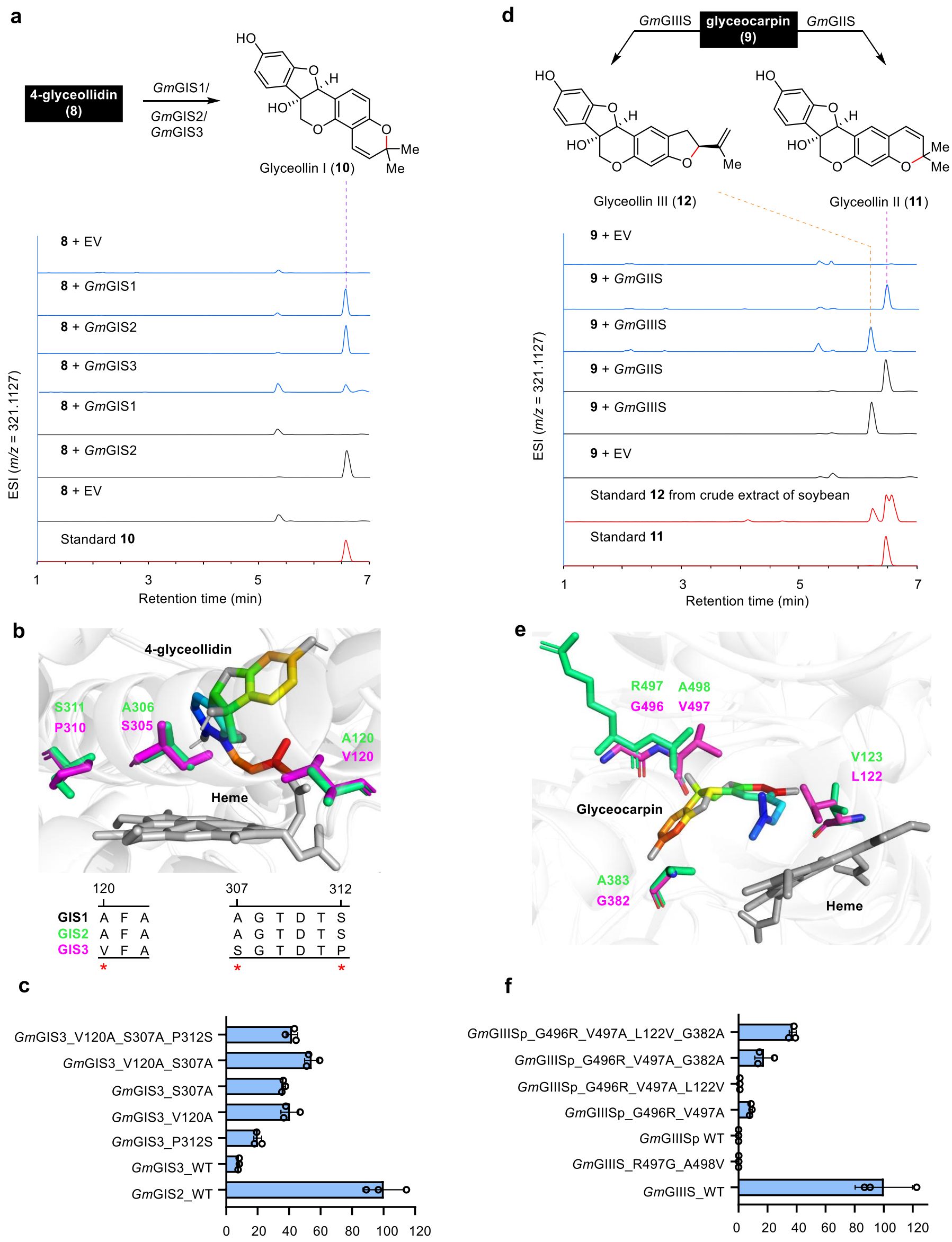
Figure 4: Reconstitution of the daidzein module for *de novo* biosynthesis of daidzein. (a) Schematic diagram illustrating the engineered metabolic pathway for the construction of daidzein-producing strains. (b) Metabolic modification of the daidzein module improved the titers of isoliquiritigenin, liquiritigenin, naringenin, and daidzein. DAHP: 3-deoxy-D-arabino-heptulosonic acid 7-phosphate, CHA: chorismic acid, PPA: prephenate, PPY: phenylpyruvate, HPP: para-hydroxy-phenylpyruvate, *p*-PAC: para-phenylacetaldehyde, L-TYR: L-tyrosine, *p*-HCA: *p*-coumaric acid, *p*-HCA-CoA: p-coumaroyl-CoA, ARO4*: DAHP synthase mutant (K229L), ARO7*: chorismate mutase mutant (G141S), ARO10: phenylpyruvate decarboxylase, *Fj*TAL: *F. johnsoniae* tyrosine ammonia lyase, *Pc*4CL2: *P. crispum* 4-coumaroyl-CoA synthase 2, *Ph*CHS-GmCHR5: *P. hybrida* chalcone synthase fused with *G. max* chalcone reductase, *Gm*CHR5: *G. max* chalcone reductase, *Gm*CHIB2: *G. max* chalcone isomerase, *Ge*2-HIS: *G. echinate* 2-hydroxyisoflavanone synthase, *Am*IFS: *A. membranaceus* isoflavone synthase, *Gm*HID: *G. max* 2-hydroxyisoflavanone dehydratase. Data are shown as mean value ± standard deviations (n = 3 biologically independent samples).

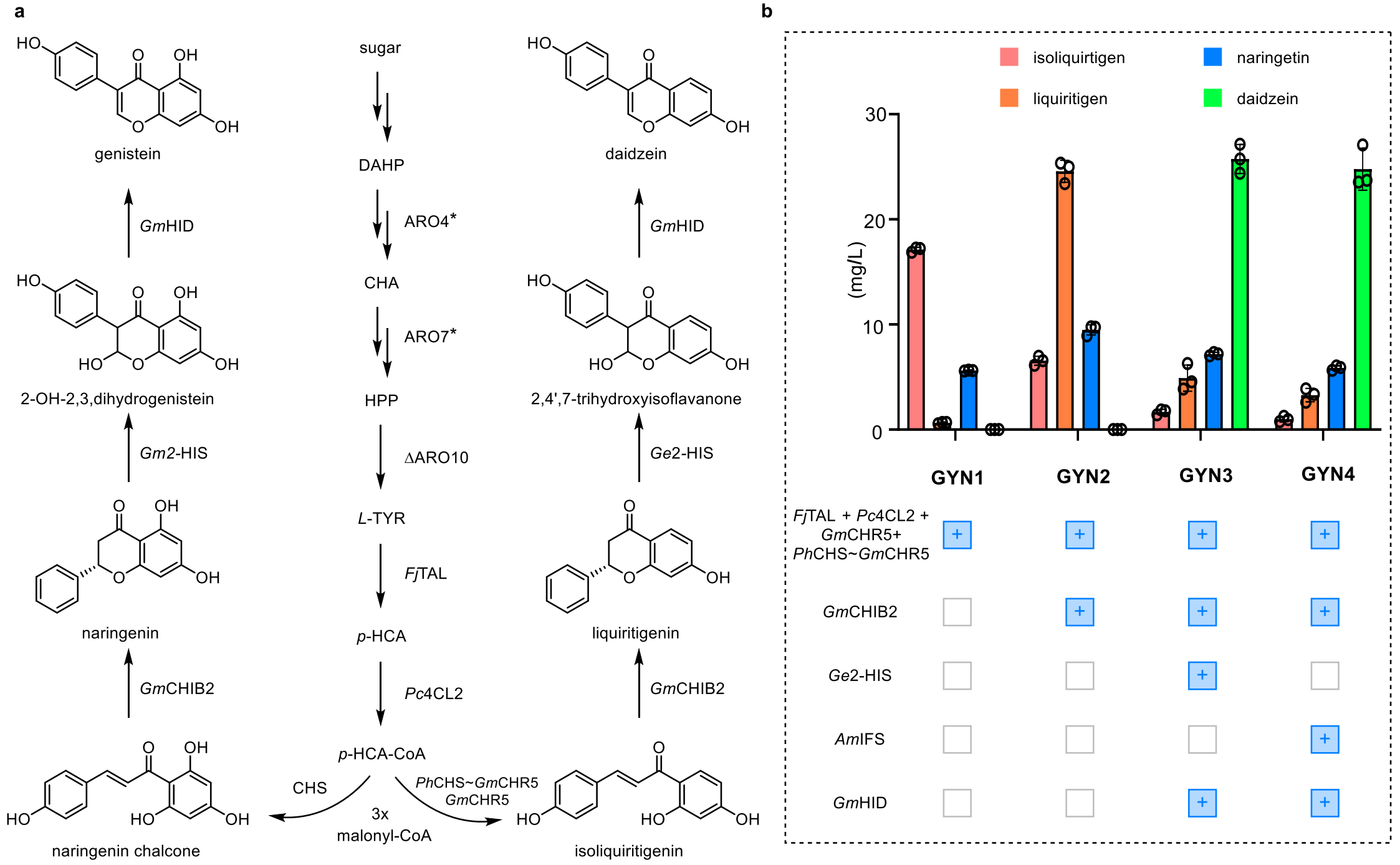
Figure 5: Reconstitution of the glycinol module for *de novo* biosynthesis of glycinol. Scheme of the glycinol module engineering and average LC–MS peak area of relative products produced in yeast expressing the indicated enzymes. Data are shown as mean value ± standard deviations (n = 3 biologically independent samples)**.**

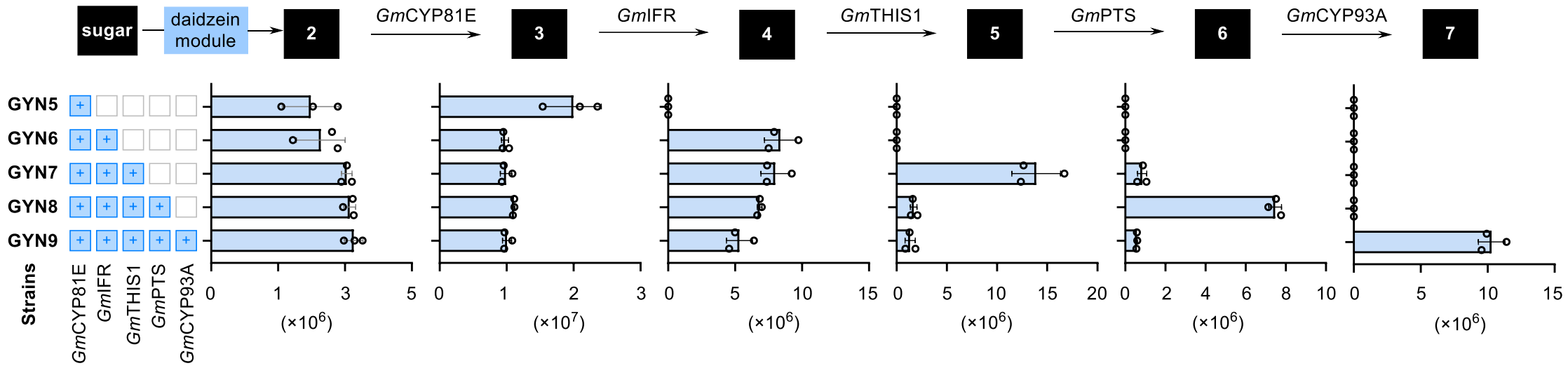
Figure 6: Reconstitution of the glyceollin module for *de novo* biosynthesis of glyceollins. (a) Scheme of the glyceollin module engineering. (b) Fusion expression of IDI1 with *tGm*G2DT/*tGm*G4DT, remarkably enhancing the titer of **8** and **9**. EICs for **8** and **9** ($[M-H_2O+H]^+$ = *m/z* 323.1283) are presented (**c**) Changes in the titer of **8** and **9** in strains GYN18, GYN23, GYN19, and GYN25. (d) LC-MS profiles of glyceollins I, II, and III-producing strains. EICs for **10**, **11**, and **12** ($[M-H_2O+H]^+$ = *m/z* 321.1127) are presented. Data are shown as mean value ± standard deviations (n = 3 biologically independent samples).

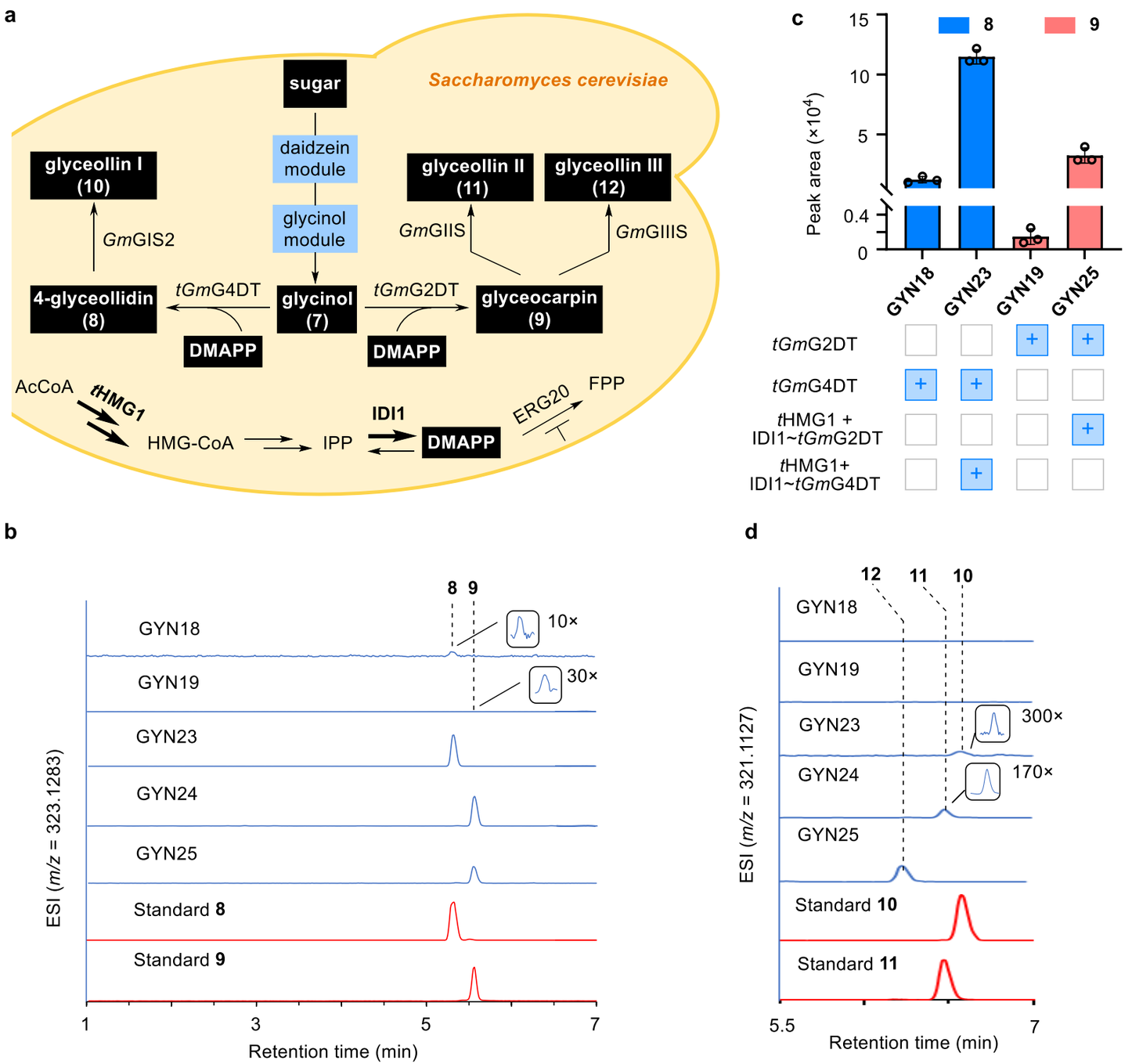

8 enzymes
phenylalanine (1)
daidzein (2)
GmCYP81E
2'-OH daidzein (3)
GmIFR
2'-OH dihydrodaidzein (4)
reductase
7,2',4'-trihydroxyisoflavanol (THI) (5)
GmPTS
3,9-OH peterocarpan (6)
GmCYP93A
Glycinol (7)
GmG4DT
4-glyceollidin (8)
CYP
glyceollin I (10)
glyceollin II (11)
GmG2DT
glyceocarpin (9)
CYP
CYP
glyceollin III (12)
uncharacterized
characterized


a
HO
O
OH
O
OH
4
GmTHIS1/2
HO
O
OH
OH
OH
5
Shunt
N. benthamiana
HO
O
H
H
O
OH
6
b
4 + EV
4 + GmTHIS1
4 + GmTHIS2
5 + EV
Standard 5
Standard 6
ESI (m/z = 257.0814)
0.5
2.5
4.5
6.5
Retention time (min)
c
4 + boiled GmTHIS1
4 + GmTHIS1
4 + boiled GmTHIS2
4 + GmTHIS2
Standard 5
Standard 6
ESI (m/z = 257.0814)
0.5
2.5
4.5
6.5
Retention time (min)


a
4-glyceollidin (8)
GmGIS1/
GmGIS2/
GmGIS3
Glyceollin I (10)
8 + EV
8 + GmGIS1
8 + GmGIS2
8 + GmGIS3
8 + GmGIS1
8 + GmGIS2
8 + EV
Standard 10
ESI (m/z = 321.1127)
Retention time (min)
d
GmGIIIS
glyceocarpin (9)
GmGIIS
Glyceollin III (12)
Glyceollin II (11)
9 + EV
9 + GmGIIS
9 + GmGIIIS
9 + GmGIIS
9 + GmGIIIS
9 + EV
Standard 12 from crude extract of soybean
Standard 11
ESI (m/z = 321.1127)
Retention time (min)
b
4-glyceollidin
S311
P310
A306
S305
A120
V120
Heme
120
307
312
GIS1 A F A A G T D T S
GIS2 A F A A G T D T S
GIS3 V F A S G T D T P
e
R497 A498
G496 V497
V123
L122
Glyceocarpin
A383
G382
Heme
c
GmGIS3_V120A_S307A_P312S
GmGIS3_V120A_S307A
GmGIS3_S307A
GmGIS3_V120A
GmGIS3_P312S
GmGIS3_WT
GmGIS2_WT
f
GmGIIISp_G496R_V497A_L122V_G382A
GmGIIISp_G496R_V497A_G382A
GmGIIISp_G496R_V497A_L122V
GmGIIISp_G496R_V497A
GmGIIISp WT
GmGIIIS_R497G_A498V
GmGIIIS_WT


a
sugar
DAHP
ARO4*
CHA
ARO7*
HPP
ΔARO10
L-TYR
FjTAL
p-HCA
Pc4CL2
p-HCA-CoA
CHS
3x
malonyl-CoA
PhCHS~GmCHR5
GmCHR5
genistein
GmHID
2-OH-2,3,dihydrogenistein
Gm2-HIS
naringenin
GmCHIB2
naringenin chalcone
daidzein
GmHID
2,4',7-trihydroxyisoflavanone
Ge2-HIS
liquiritigenin
GmCHIB2
isoliquiritigenin
b
isoliquirtigen
naringetin
liquiritigen
daidzein
(mg/L)
30
20
10
0
GYN1
GYN2
GYN3
GYN4
FjTAL + Pc4CL2 + GmCHR5+ PhCHS~GmCHR5
GmCHIB2
Ge2-HIS
AmIFS
GmHID


sugar
daidzein module
2
GmCYP81E
3
GmIFR
4
GmTHIS1
5
GmPTS
6
GmCYP93A
7
GYN5
GYN6
GYN7
GYN8
GYN9
Strains
GmCYP81E
GmIFR
GmTHIS1
GmPTS
GmCYP93A
(×10^6)
(×10^7)
(×10^6)
(×10^6)
(×10^6)
(×10^6)


a
Saccharomyces cerevisiae
sugar
daidzein module
glycinol module
glycinol (7)
glyceollin I (10)
GmGIS2
4-glyceollidin (8)
tGmG4DT
DMAPP
tGmG2DT
DMAPP
glyceocarpin (9)
GmGIIS
GmGIIIS
glyceollin II (11)
glyceollin III (12)
AcCoA
tHMG1
HMG-CoA
IPP
IDI1
DMAPP
ERG20
FPP
b
GYN18
GYN19
GYN23
GYN24
GYN25
Standard 8
Standard 9
8 9
10×
30×
ESI (m/z = 323.1283)
Retention time (min)
c
Peak area (×10^4)
8
9
GYN18
GYN23
GYN19
GYN25
tGmG2DT
tGmG4DT
tHMG1 + IDI1~tGmG2DT
tHMG1+ IDI1~tGmG4DT
d
12 11 10
GYN18
GYN19
GYN23
GYN24
GYN25
Standard 10
Standard 11
300×
170×
ESI (m/z = 321.1127)
Retention time (min)