Supplemental information

In vivo assembly in tobacco cells to elucidate and engineer the biosynthesis of 4-hydroxydihydrocinnamaldehyde from *Gloriosa superba*

Zhiqiang Xiong[1, 2, 3†], Zhuoying Ding[1, 2, 3 †], Jingyi Sun[1, 2, 3], Xuefei Jiang[3], Hanqing Cong[1, 2], Huapeng Sun[1, 2*], Fei Qiao[1, 2*]

[1] National Key Laboratory for Tropical Crops Breeding, Sanya 572024, China

[2] Key Laboratory of Crop Gene Resources and Germplasm Enhancement in Southern China, Ministry of Agriculture, Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China

[3] Hainan Key Laboratory of Sustainable Utilization of Tropical Bioresources, Key Laboratory for Quality Regulation of Tropical Horticultural Plants of Hainan Province, Sanya Nanfan Research Institute, College of Horticulture, Hainan University, Haikou 570228, China

* Correspondence: huapeng_sun@catas.cn; fei.qiao@foxmail.com

† These authors contributed equally to this article.

Experimental Section

Bioinformatics analysis

The amino acid sequences of candidate genes were predicted by NCBI local CDD (Conserved Domain Database) (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). The physicochemical properties of candidate genes were analyzed by the PROTPARAM program (https://web.expasy.org/protparam/). DNAMAN 2.0 software (lynnon biosoft, USA) was used to analyze sequencing results. The genetic relationship between candidate genes and other species was analyzed by MEGA 7.0 based on a neighbour-joining (NJ) method with 1000 replications of bootstrapping. The protein secondary structure of candidate genes was predicted using the SOPMA online website (https://npsa-prabi.ibcp.fr/). The three-dimensional protein structure of candidate genes was established by SWISS-MODEL online program (http://swissmodel.expasy.org/). Then, the data were processed and analysed through CLUSTAL (https://www.genome.jp/tools-bin/clustalw) and ESPript 3.0 (https://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi) online software.

The online database of Chemspider (http://www.chemspider.com/) was used for searching structure of 4-coumaraldehyde (ID: 556592). The three-dimensional structural of *Gs*DBR1 was homology predicted as template structure *At*DBR (PDB, 2j3k). The Molecular Operating Environment (MOE) 2009 software was used for molecular docking. After all the docking work was completed, the conformation of the lowest binding energy was selected, and the Python Molecule (PyMoL) software was used to analyze the interaction of *Gs*DBR1 protein structure and ligand (Xiong et al., 2022).

Assembly of tandem gene clusters

In Module I and Module II tandem gene clusters, each gene carried independent promoter and terminator sequences. The 35S promoter and NOS terminator sequences were amplified from pK7FWH2, then assembled with candidate genes into pUC19 between *EcoR I* and *Hind III* sites by using NEBuilder HiFi DNA Assembly Mix (New England Biolabs) as the manufacturer's instructions. *E. coli* DH5α competent cells were used for plasmid amplification and isolation in the gene cluster assembled process.

Identification of transgenic positive cell lines

For the identification of positive strains in transgenic strains, details of the establishment and maintenance of BY2 cell lines are described in our previous publication (Xiong et al., 2022). We used ZEISS Apotome microscope (Zeiss, Germany) to observe the results of subcellular localization of overexpressed cell lines in stably transformed suspension tobacco BY2 cells. Meanwhile, we also performed full-length identification of target genes by obtaining cDNAs of transgenic strains and wild-type tobacco BY2 cell lines as templates using the Primer STAR Max DNA Polymerase (TAKARA Biomedical, Japan) kit, according to the manufacturer's instructions. The wild-type tobacco BY2 cell strain as negative control was observed in the same condition. Packed cell volume (PCV) was used as an indicator of cell growth (Liu et al., 2013). For *in vivo* experiments, all flasks with suspended culture cells were placed on an oscillator (25 mm) at 150 rpm and 25 °C for 7-12 days in the dark.

Quantitative real-time PCR analysis

Primers at the 3' and 5' ends were designed according to Primer 5.0 software for key genes of the 4-HDCA pathway (Table S1), and the Quantitative real-time PCR (qRT-PCR) reactions were performed using the PikoReal System (Thermo Fisher, USA). Total RNA (2 μg) was obtained from different *G. superba* tissues (root, stem, leave, flower), and the first strand of cDNA was synthesized using PrimeScriptTM RT Master Mix (Takara, Japan). qRT-PCR reactions were performed using a total system of 10 μL of reaction mixture consisting of 1 μL cDNA template, 0.3 μL forward primer (10 μM), 0.3 μL reverse primer (10 μM), 5 μL 2X SYBR qPCR Master Mix (VAZYME, China) and 3.4 μL ddH2O. ddH_2O was used as the template for all negative controls. The results of the experiments were processed using the $2^{-\Delta\Delta CT}$ method. *GsActin7* (GenBank accession no. OM816676) was used as a reference gene, and the expression level of leave was set to '1'. Three technical replicates and three biological replicates were used for each sample treatment.

References

Xiong, Z.Q., Wang, L., Sun, J.Y., Jiang, X.F., Cong, H.Q., Sun, H.P. and Qiao, F. Functional characterization of a *Colchicum autumnale* L. double-bond reductase (CaDBR1) in colchicine biosynthesis. *Planta*. 256: 95 (2022).

Liu, Q., Qiao, F., Ismail, A., Chang, X. and Nick, P. The plant cytoskeleton controls regulatory volume increase. *BBA-Biomembranes*. 1828 (9) (2013).

Supplemental Table S1. The primers were used in this study.

Primer name	Sequence (5′–3′)
GsPAL1-F	ATGGAGTCTGTGTATGAGAACGGC
GsPAL1-R	GCATATGGGGATGGGAGCCCC
GsPAL2-F	ATGGAGTTTGTCCACCAA
GsPAL2-R	ACAGATAGGCAGAGGAGCACC
Gs4CL1-F	ATGGGCTCCACACCGCCG
Gs4CL1-R	GTTATGACCATTTGGGATTCC
Gs4CL2-F	ATGGACCAAACCCTAGCA
Gs4CL2-R	CATCCTCGAAGACCTGTT
Gs4CL3-F	ATGGCCTCATCTCGCTCCGGC
Gs4CL3-R	CAACTTTGATCTCACTTTCGC
GsCCR1-F	ATGCCTGCCATCACCACC
GsCCR1-R	AGGCATGACGGGGAGGAC
GsCCR2-F	ATGCCCGCCGACGCCCCC
GsCCR2-R	TGCAACAGTTGTCTCTTC
GsDBR1-F	ATGGCGGAGGTGAGGGCGAA
GsDBR1-R	CTCATGGGCTAATACGACCAA
GsDBR2-F	ATGGCAGGAGAGAAGGTGAGC
GsDBR2-R	TTCGTGCGCAACTGTTACAAC
GsC4H1-F	ATGGATCTCTTACTCCTCGAG
GsC4H1-R	GAAAACCCTAGGCTTTGCAAC
pET28a-GsPAL1-F	AAGAAGGAGATATACATGGAGTCTGTGTATGAGAACGGC
pET28a-GsPAL1-R	GACGGAGCTCGAATTGCATATGGGGATGGGAGCCCC
pET28a-GsPAL2-F	AAGAAGGAGATATACATGGAGTTTGTCCACCAA
pET28a-GsPAL2-R	GACGGAGCTCGAATTACAGATAGGCAGAGGAGCACC
pET28a-Gs4CL1-F	AAGAAGGAGATATACATGGGCTCCACACCGCCG
pET28a-Gs4CL1-R	GACGGAGCTCGAATTGTTATGACCATTTGGGATTCC
pET28a-Gs4CL2-F	AAGAAGGAGATATACATGGACCAAACCCTAGCA
pET28a-Gs4CL2-R	GACGGAGCTCGAATTCATCCTCGAAGACCTGTT
pET28a-Gs4CL3-F	AAGAAGGAGATATACATGGCCTCATCTCGCTCCGGC
pET28a-Gs4CL3-R	GACGGAGCTCGAATTCAACTTTGATCTCACTTTCGC
pET28a-GsCCR1-F	AAGAAGGAGATATACATGCCTGCCATCACCACC
pET28a-GsCCR1-R	GACGGAGCTCGAATTAGGCATGACGGGGAGGAC
pET28a-GsCCR2-F	AAGAAGGAGATATACATGCCCGCCGACGCCCCC
pET28a-GsCCR2-R	GACGGAGCTCGAATTTGCAACAGTTGTCTCTTC
pET28a-GsDBR1-F	AAGAAGGAGATATACATGGCGGAGGTGAGGGCGAA
pET28a-GsDBR1-R	GACGGAGCTCGAATTCTCATGGGCTAATACGACCAA
pET28a-GsDBR2-F	AAGAAGGAGATATACATGGCAGGAGAGAAGGTGAGC
pET28a-GsDBR2-R	GACGGAGCTCGAATTTTCGTGCGCAACTGTTACAAC
pYeDP60-GsC4H1-F	CTAAATTACCGGAATGGATCTCTTACTCCTCGAG
pYeDP60-GsC4H1-R	GATCCCCCGCGAATTCGAAAACCCTAGGCTTTGCAAC

GsPAL1-qRT-F	ACAACCAGGACGTGAACT
GsPAL1-qRT-R	CGACATCAGCTTCAGGAT
GsPAL2-qRT-F	TGTGATGGGAAGGTGATT
GsPAL2-qRT-R	TCAACAGATAGGCAGAGG
Gs4CL1-qRT-F	GGCTACAGCGAACACAAT
Gs4CL1-qRT-R	ATCTCATCATCATCATCCACAT
Gs4CL2-qRT-F	TACATACTGGACCGCTTGA
Gs4CL2-qRT-R	GACAGCAACATCAATAATCTCAG
Gs4CL3-qRT-F	AGCCTCTTCCACCTAATC
Gs4CL3-qRT-R	TTGTTGCCTGTGTATTGTT
GsCCR1-qRT-F	CTGTAAGACTGTTTGTGT
GsCCR1-qRT-R	CGATGTATCCCTTCTCTA
GsCCR2-qRT-F	GGAAGCAGCCATACAAAT
GsCCR2-qRT-R	TCATAGAGGCACTGTGTTA
GsDBR1-qRT-F	GAAGGTGAATCTATTGAAG
GsDBR1-qRT-R	TTAAGAAGAACAGCATCT
GsDBR2-qRT-F	CCATAACTTGTTCTGTCT
GsDBR2-qRT-R	TTCTACGCTATCTTCAAC
GsC4H1-qRT-F	CGAGTATAACTATGGCGATT
GsC4H1-qRT-R	CCTCTCCTTAACCTCCTTA
GsActin7-qRT-F	ACTGATGCTCTGATGAAG
GsActin7-qRT-R	TCTCCTTGATGTCCCTTA
pDNOR-Module I-F	GGGGACAAGTTTGTACAAAAAAGCAG GCTTCATGGAGTCTGTGTATGAGAA
pDNOR-Module I-R	GGGGACCACTTTGTACAAGAAAGCTGG GTCGAAAACCCTAGGCTTTGC
pDNOR-Module II-F	GGGGACAAGTTTGTACAAAAAAGCAG GCTTCATGGGCTCCACACCGCCGGA
pDNOR-Module II-R	GGGGACCACTTTGTACAAGAAAGCTG GGTCCTCATGGGCTAATACGACCA
pDNOR-Module III-F	GGGGACAAGTTTGTACAAAAAAGCAG GCTTCATGGAGTCTGTGTATGAGAA
pDNOR-Module III-R	GGGGACCACTTTGTACAAGAAAGCTG GGTCCTCATGGGCTAATACGACCA

Supplemental Table S2. Analysis of various physicochemical properties of 4-HDCA biosynthesis-related genes.

GenBank ID	Genes Name	Annotation ["species"]	ORF (bp)	Amino Acids	Molecular Weight	Theoretical isoelectric point	Instability index (II)	Aliphatic index	Grand average of hydropathicity
ON324201	*GsPAL1*	phenylalanine ammonia-lyase	2166	721	78.06	5.93	34.10	90.82	-0.137
ON324202	*GsPAL2*	phenylalanine ammonia-lyase	2163	720	77.47	6.21	30.30	94.51	-0.054
ON324196	*GsC4H1*	cinnamate 4-hydroxylase	1518	505	57.96	9.23	45.62	99.41	-0.219
ON324193	*Gs4CL1*	4-coumarate: coenzyme A	1647	548	59.36	5.59	33.68	98.52	0.099
ON324194	*Gs4CL2*	4-coumarate: coenzyme A	1568	555	61.02	7.67	39.61	88.90	-0.064
ON324195	*Gs4CL3*	4-coumarate: coenzyme A	1629	542	59.42	7.69	46.79	101.09	0.070
ON324197	*GsCCR1*	cinnamoyl-CoA reductase	987	328	35.83	8.24	32.18	87.92	-0.141
ON324198	*GsCCR2*	cinnamoyl-CoA reductase	1005	334	36.31	5.44	27.97	89.88	-0.081
ON324199	*GsDBR1*	2-alkenal reductase (NADP (+)-dependent)	1041	346	38.18	5.68	29.16	85.30	-0.062
ON324200	*GsDBR2*	2-alkenal reductase (NADP (+)-dependent)	1041	346	38.07	6.61	31.42	90.58	-0.011

Supplemental Table S3. *In vitro* activity validation protocols for *Gs*PALs, *Gs*4CLs, *Gs*CCRs, *Gs*C4H, *Gs*DBRs.

Genes	Substrate	Mobile phase	Chemicals	Determine wavelength	Assay condition	Elution program
GsPALs	L-Phenylalanine (500 μM)	A: 0.1 % HCOOH aqueous B: methanol Flow velocity: 0.5 mL/min	pH 7.5 Tris-Hcl (50 mM)	254 nm 280 nm	Incubate at 30 ℃ for 30 min	0 min 5 % B 5 min 55 % B 14 min 90 % B 18 min 90 % B 19 min 5 % B 21 min 5 % B
Gs4CLs	4-Coumaric acid (500 μM) or Cinnamic acid (500 μM)	A: 1 % H_2PO_4 aqueous B: acetonitrile Flow velocity: 1 mL/min	pH 7.5 Tris-Hcl (200 mM), ATP (5 mM), $Mgcl_2$ (5 mM), CoA (2.5 mM)	300 nm 333 nm	Incubate at 25 ℃ for 20 min	0 min 5 % B 5 min 5 % B 35 min 30 % B 45min 100% B 47 min 5 % B 50 min 5 % B
GsCCRs	4-coumaroyl CoA (500 μM) or Cinnamoyl CoA (500 μM)	A: 1 % H_2PO_4 aqueous B: acetonitrile Flow velocity: 1 mL/min	pH 6 Tris-Hcl (200 mM), NADPH (100 μM)	280 nm 300 nm 333 nm	Incubate at 30 ℃ for 25 min	0 min 5 % B 5 min 5 % B 35 min 30 % B 45 min 100 % B 47 min 5 % B 50 min 5 % B
GsC4H	Cinnamic aciad (500 μM) or Dihydrocinnamaldehyde (500 μM)	A: 0.1 % HCOOH aqueous B: methanol Flow velocity: 0.5 mL/min	None	280 nm 310 nm	None	0 min 5 % B 5 min 55 % B 14 min 90 % B 18 min 90 % B 19 min 5 % B 25 min 5 % B
GsDBRs	4-coumaraldehyde (500 μM) or Cinnamaldehyde (500 μM)	A: 0.1 % HCOOH aqueous B: methanol Flow velocity: 0.5 mL/min	pH 7.5 Tris-Hcl (50 mM), NADPH (100 μM)	280 nm	Incubate at 30 ℃ for 30 min	0 min 5 % B 5 min 55 % B 14 min 90 % B 18 min 90 % B 19 min 5 % B 25 min 5 % B

Supplemental Table S4. Detection of Module I/II/III stable transgenic cells

Transformation cell strain	Inducing substance	Mobile phase	Determine wavelength	Extract method	Elution program
Module I	L-Phenylalanine (500 μM) or MeJA (100 μM)	A: 0.1 % HCOOH aqueous B: methanol Flow velocity: 1 mL/min	310 nm	Ultrasound with methanol at 30 °C	0 min 20 % B 20 min 50 % B 25 min 90 % B 30 min 90 % B 31 min 20 % B 35 min 20 % B
Module II/III	L-Phenylalanine (500 μM), MeJA (100 μM) or 4-Coumaric acid (500 μM)	A: 0.1 % HCOOH aqueous B: methanol Flow velocity: 0.8 mL/min	280 nm	Ultrasound with methanol at 30 °C	0 min 10 % B 10 min 60 % B 14 min 100 % B 18 min 100 % B 19 min 10 % B 23 min 10 % B

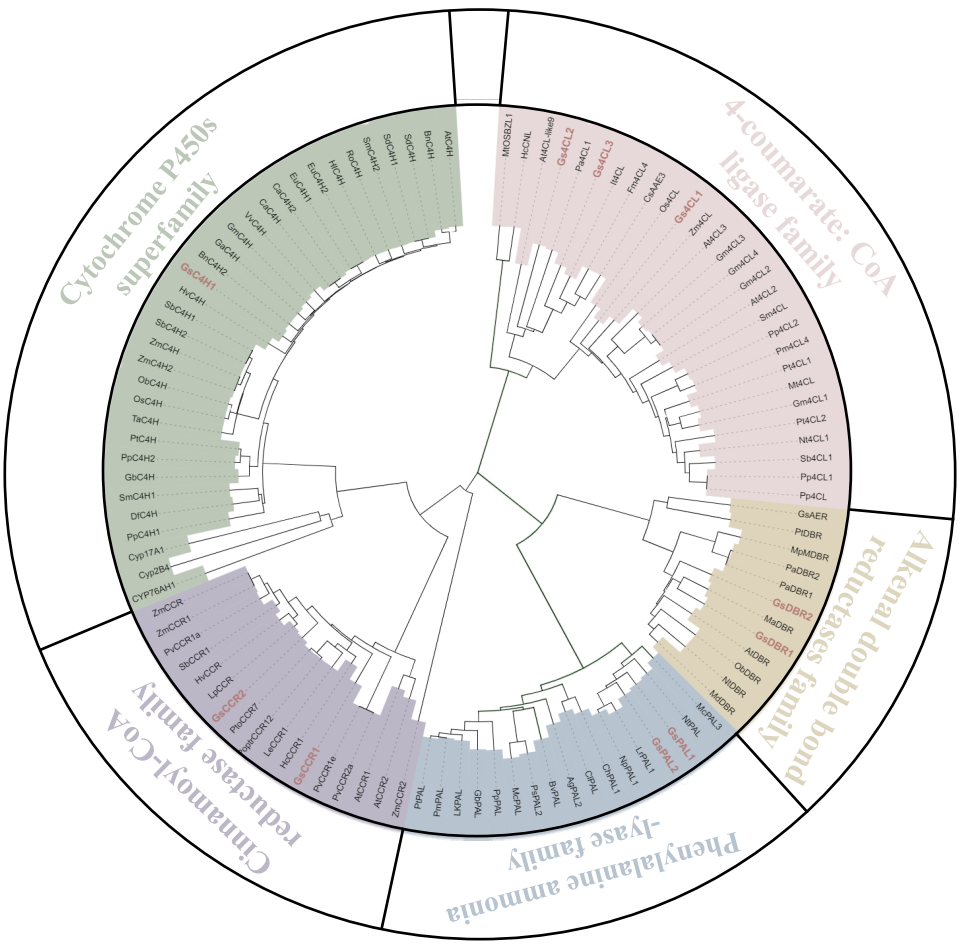

Figure S1. Phylogenetic analysis of genes involved in 4-HDCA biosynthetic pathway. Sequences from this study are indicated with red lines. The following sequences were analyzed in this study:

The following sequences were analyzed in this study: *At*DBR (*Arabidopsis thaliana*; AAY25470.1), *Nt*DBR (*Nicotiana tabacum*; NP_001313179.1), *Md*DBR (*Malus domestica*; 6YTZ_A), *Pt*DBR (*Pinus taeda*; ABG91753.1), *Ob*DBR (*Ocimum basilicum*; AY879288), *Ma*DBR (*Musaacuminata*; XM_009402863.1), *Mp*MDBR (*Marchantia paleacea*; MH427075.1), *Pa*DBR1 (*Plagiochasma appendiculatum*; KF051271), *Pa*DBR2 (*Plagiochasma appendiculatum*; KF051272) and *Gs*AER (*Gloriosa superba*; MT512051). *Mc*PAL3(*Morinda citrifolia*; MN833288.1), *Nt*PAL(*Nicotiana tabacum*, XP_016480992.1), *Lr*PAL1 (*Lycoris radiata*, AWW24969.1), *Np*PAL1 (*Narcissus papyraceus*, AXU39892.1), *Ch*PAL1 (*Cephalotaxus hainanensis*, MN398159.1), *Cl*PAL (*Cunninghamia lanceolate*, AFX98071.1), *Ag*PAL (*Anthoceros agrestis*, SPO49995.1), *Bv*PAL (*Botrypus virginianus*, AAW80641.1), *Ps*PAL2 (*Papaver somniferum*, XP_026431031.1), *Mc*PAL (*Macleaya cordata*, OVA08384.1), *Pp*PAL (*Pinus pinaster*, AAP85251.1), *Gb*PAL (*Ginkgo biloba*, ABU49842.1), *Lk*PAL (*Larix kaempferi*, AHA44840.1), *Pt*PAL (*Pinus taeda*, AHX74218.2), *Pp*PAL (*Pinus massoniana*, ACS28225.2), *Zm*CCR (*Zea mays*, CAA74071), *At*CCR1 (*Arabidopsis thaliana*, AAG46037.1), *At*CCR2 (*Arabidopsis thaliana*, AAG53687.1), *Pto*CCR7 (*Populus tomentosa*, KF145198.1), *Hv*CCR (*Hordeum vulgare*, AAN71760.1), *Lp*CCR (*Lolium perenne*, AAG09817.1), *Poptr*CCR12 (*Populus trichocarpa*, CAA12276.1), *Pv*CCR1e (*Panicum virgatum*, GQ450301.1), *Pv*CCR1a (*Panicum virgatum*, GQ450297.1), *Pv*CCR2a (*Panicum virgatum*, GQ450302.1), *Sb*CCR1 (*Sorghum bicolor*, XP_002445566.1), *Zm*CCR1 (*Zea mays*, NP_001105488), *Zm*CCR1 (*Zea mays*, NP_001105715) , *Hc*CCR1 (*Hibiscus cannabinus*, HM151381.1), *Le*CCR1 (*Lycopersicon esculentum*, AAY41879), *At*C4H (*Arabidopsis thaliana*, NP_180607.1), *Bc*C4H (*Brassica napus*, XP_013748807.1), *Sd*C4H1 (*Scoparia dulcis*, AGX85600.1), *Sm*C4H2 (*Salvia miltiorrhiza*,ABC75596.1), RoC4H(*Rubus occidentalis*, ACM17896.1), *Eu*C4H1 (*Eucalyptus urophylla*, AGJ71350.1), *Hl*C4H(*Humulus lupulus*, ACM69364.1), *Eu*C4H2(*Eucalyptus urophylla*, AGJ71350.1), *Ca*C4H2 (Canarium album,ACR10242.1), *Ca*C4H, (Canarium album, ACR10242.1), *Vv*C4H (*Vitis vinifera*, XP_002266238.1), *Gm*C4H (*Glycine max*, NP_001237317.1), *Ga*C4H (*Gossypium arboretum*, AAG10197.1), *Bn*C4H2 (*Boehmeria nivea*, AJT59581.1), *Hv*C4H (*Hordeum vulgare*, AHJ61295.1), *Sb*C4H1 (*Sorghum bicolor*, 002G126600), *Sd*C4H2 (*Scoparia dulcis*, AGX85600.1), *Zm*C4H (*Zea mays*, PWZ11997.1), *Zm*C4H2 (*Zea mays*, NP_001149158.1), *Ob*C4H (*Oryza brachyantha*, XP_006654260.1), *Os*C4H (*Oryza sativa Japonica Group*, BAF45113.1), *Ta*C4H (*Triticum aestivum*, AAG17469.1), *Pt*C4H (*Pinus taeda*, AAD23378.1), *Pp*C4H2 (*Physcomitrella patens*, ADF28535), *GbC4H* (*Ginkgo biloba*, AAW70021.1), *Sm*C4H1 (*Selaginella moellendorffii*, XP_002977018.1), *Df*C4H (*Dryopteris fragrans*, KF830705.2), *Pp*C4H1 (*Pinus pinaster*, AFL65041.1), Cyp17A1 (*Danio rerio*, 4R1Z), Cyp2B4 (*Oryctolagus cuniculus*, 1PO5), CYP76AH1 (*Salvia miltiorrhiza*, 5YLW), *Pp*4CL1(*Prunus persica*, XP007215519), *Pp*4CL (*Physcomitrella patens*, ABY21312), *Pp*4CL2 (*Prunus persica*, XP_007222242), *Sb*4CL1(*Scutellaria baicalensis*, BAD90936.1), *Nt*4CL1 (*Nicotiana tabacum*, NP_001312667.1), *Pt*4CL2 (*Populus tomentosa*, AAL02144.1), *Gm*4CL1 (*Glycine max*, NP_001236418.1), *Mt*4CL (Medicago truncatula, XP_003637266.1), *Pt*4CL1 (*Pinus taeda*, AAB42383.1), *Pm*4CL4 (*Physcomitrium magdalenae*, ABY21315.1), *Sm*4CL (*Selaginella moellendorffii*, EFJ29005), *At*4CL2 (*Arabidopsis thaliana*, AAD47192.1), *Gm*4CL2 (*Glycine max*, AAC97389.1), *Gm*4CL3 (*Glycine max*, NP_001236236.2), *Gm*4CL3 (*Glycine max*, X69955), *Zm*4CL (*Zea mays,* AAS67644.1), *Os*4CL (*Oryza sativa Japonica Group*, CAA36850.1), *Cs*AAE3 (*Cannabis sativa*, AFD33347.1), *Fm*4CL4 (*Fraxinus mandshurica*, AHL44983.1), *It*4CL (*Isatis tinctoria*, ADG46006.1), *Pa*4CL1(*Plagiochasma appendiculatum*, KJ944317), *At*4CL-like9 (*Arabidopsis thaliana*, NP_201143.1), *Hc*CNL (*Hypericum calycinum*, AFS60176.1), *Mt*OSBZL1 (*Medicago truncatula*, XP_003600627.1), *At*4CL3 (*Arabidopsis thaliana*, NP_849844.1)

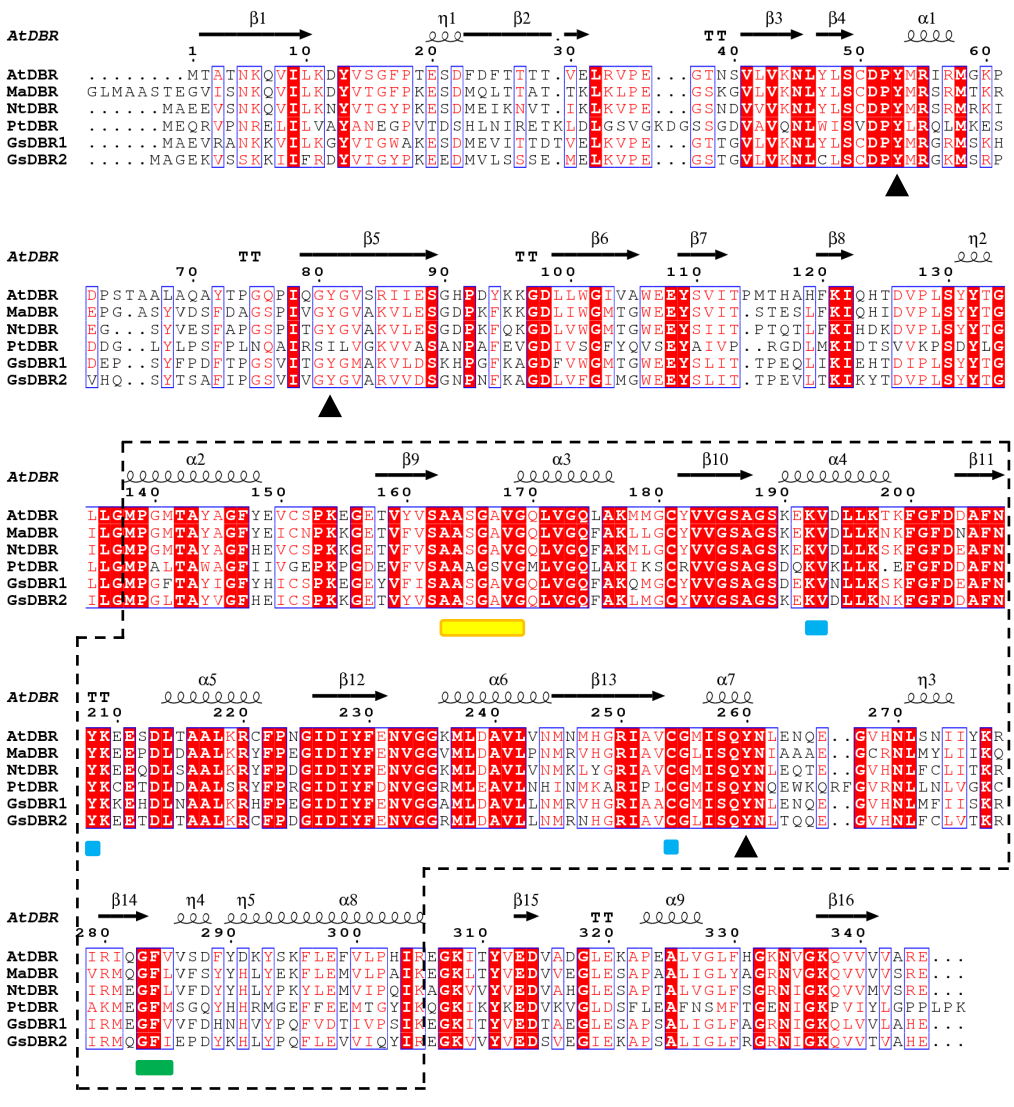

Figure S2. Multiple sequence alignment analysis of *Gs*DBRs and other DBRs. The dotted-line region indicates the conserved domain feature of the MDR family; the black triangle and blue square indicate crucial amino acids; the yellow rectangle marks a conserved domain of $NADP^+$- or NADP(H) binding sites; and the green rectangle marks a domain of substrate-binding sites. The following sequences were analyzed in this study: *At*DBR (*Arabidopsis thaliana*; AAY25470.1), *Nt*DBR (*Nicotiana tabacum*; NP_001313179.1), *Md*DBR (*Malus domestica*; 6YTZ_A), *Pt*DBR (*Pinus taeda*; ABG91753.1).

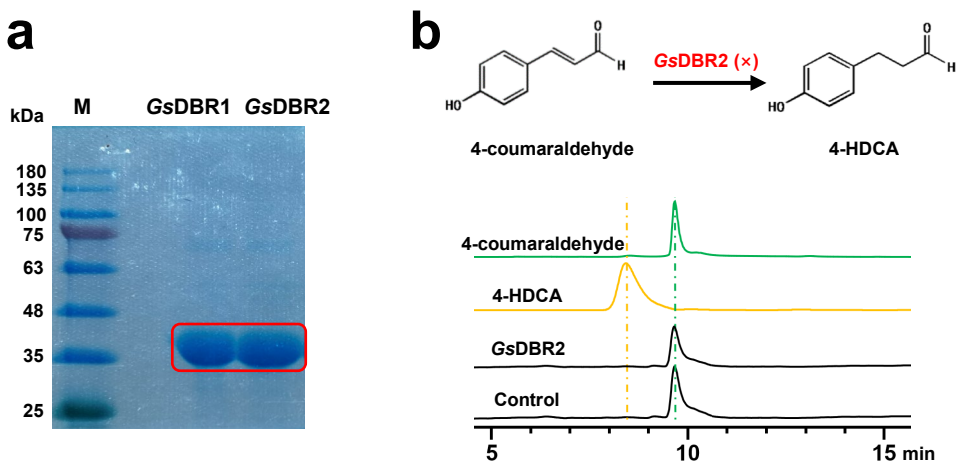

Figure S3. Heterologous expression and enzymatic assays result of *Gs*DBR2. **(a)** Purification of *Gs*DBR2 protein. SDS–PAGE with a 10 % separation gel showed the molecular weight, and lane M was the Color Mixed Protein Marker 180 (10-180 kDa). **(b)** *in vitro* enzymatic assays result of *Gs*DBR2 with feeding 4-coumaraldehyde as substrate.

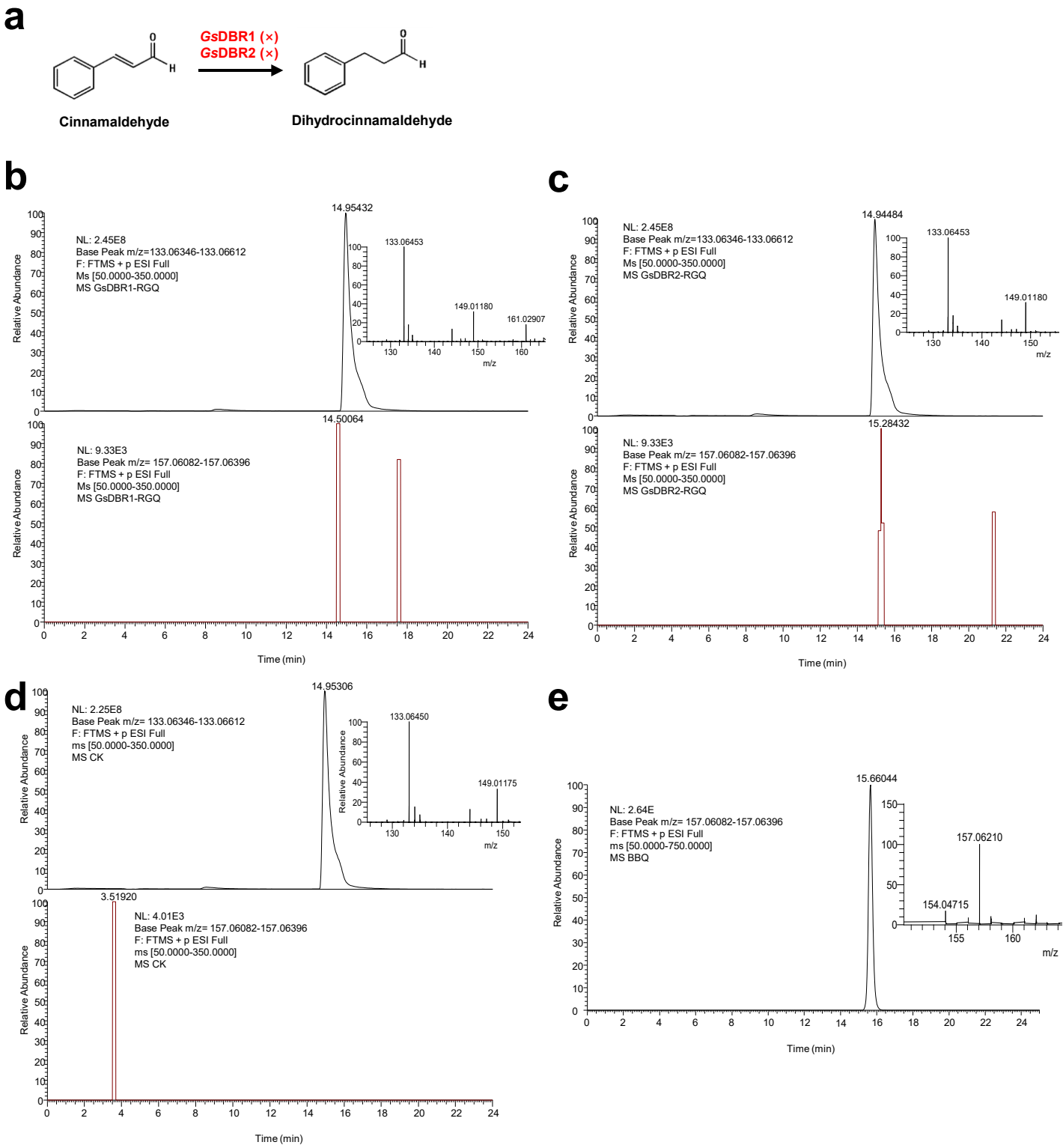

Figure S4. The UPLC-MS determination results of *Gs*DBRs enzymatic assays with cinnamaldehyde. **(a)** Chemical structure of the *in vitro* enzymatic reaction of *Gs*DBRs proteins. **(b)** The catalytic results of *Gs*DBR1. **(c)** The catalytic results of *Gs*DBR2. **(d)** EIC and MS fragments for cinnamaldehyde m/z value (133.06450 $[M+H]^+$) and dihydrocinnamaldehyde m/z value (157.06210 $[M+Na]^+$) at the indicated retention time in the CK enzyme assay sample. **(e)** EIC and MS fragments for the dihydrocinnamaldehyde m/z value (157.06210 $[M+Na]^+$) at the indicated retention time.

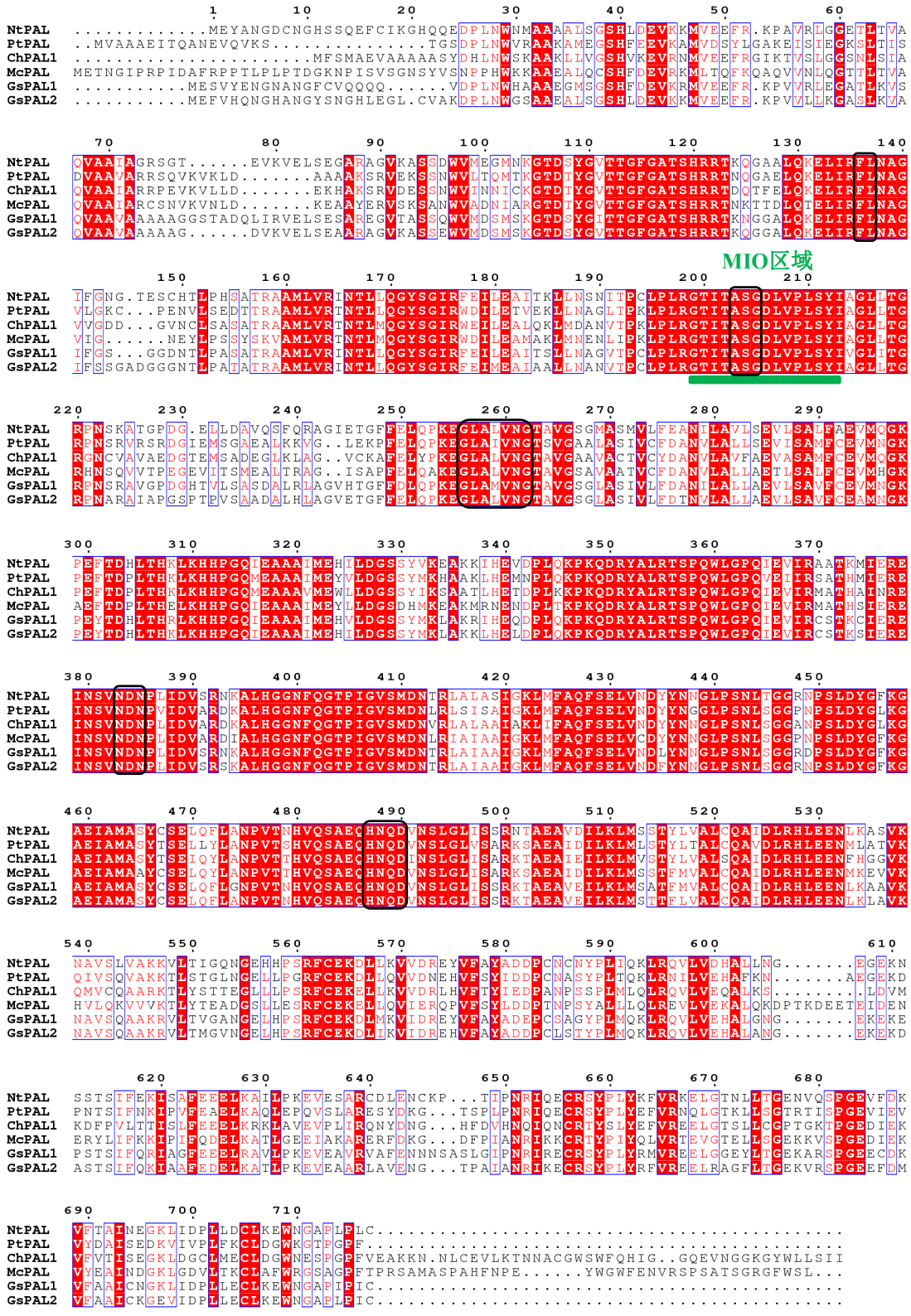

Figure S5. Multiple sequence alignment analysis of *GsPALs* and other PALs. The black rectangle indicates the four conserved units of PAL, namely the central fragment (catalytic center), the N-terminal fragment, the C-terminal fragment and the C-terminal extension containing the additional insertion; The 3,5-dihydro-5-methylidene-4H-imidazol-4-one (MIO) region represents a repair motif consisting of a conserved amino acid sequence (Ala-Ser-Gly). The following sequences were analyzed in this study: *Nt*PAL(*Nicotiana tabacum*, XP_016480992.1), *Pt*PAL (*Pinus taeda*, AHX74218.2), *Ch*PAL1 (*Cephalotaxus hainanensis*, MN398159.1), *Mc*PAL (*Macleaya cordata*, OVA08384.1).

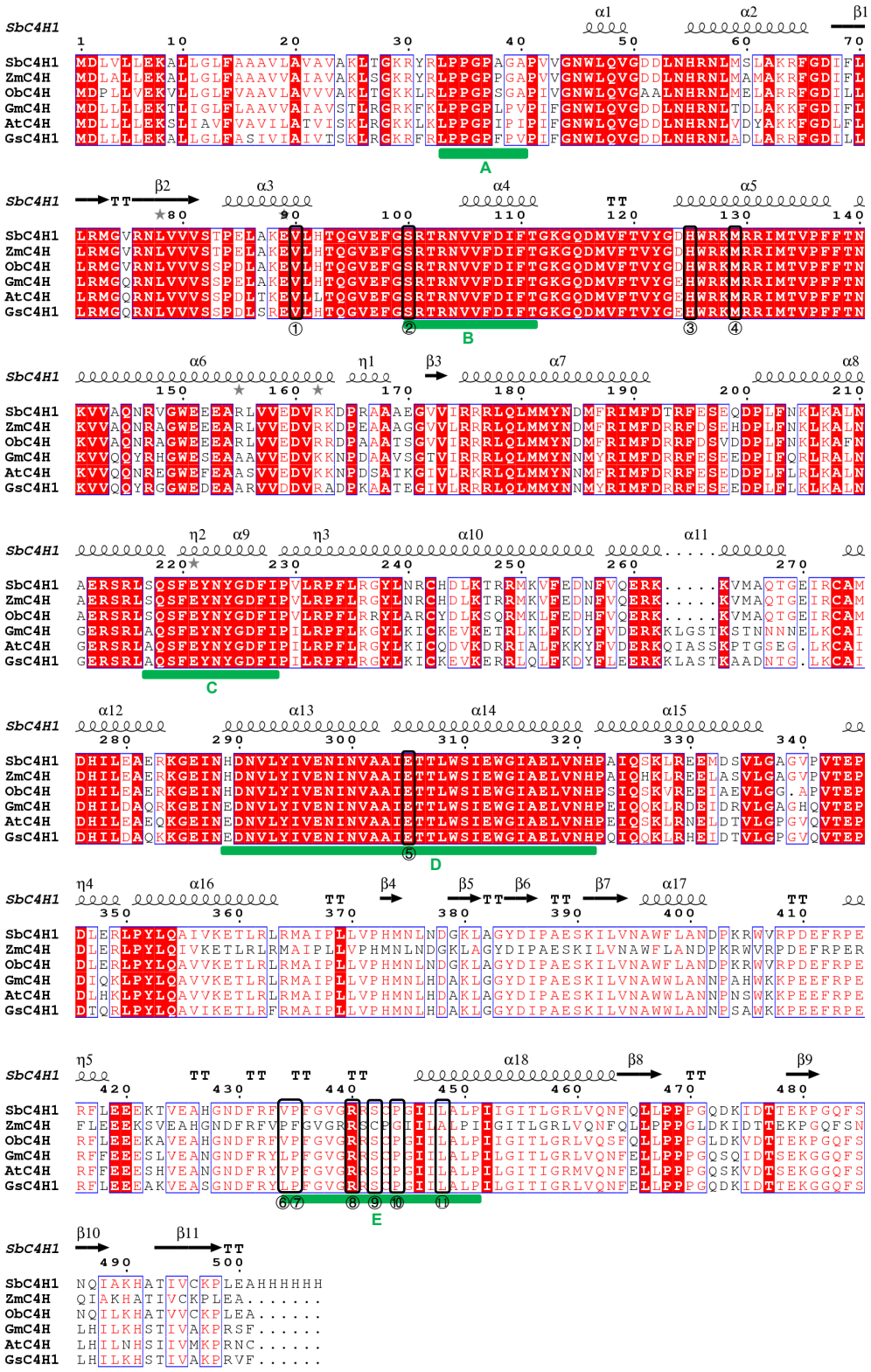

Figure S6. Multiple sequence alignment analysis of *GsC4H* and other C4Hs. The green shapes of A-E represent the typical six substrate recognition site regions of cytochrome P450 enzymes that are predicted as a signal anchor sequence, and were used to identify P450 enzymes in different subcellular localizations and different physiological functions; 1-11 indicates that cytochrome P450 enzymes are located at 11 conserved amino acid residues around heme. The following sequences were analyzed in this study: *Sb*C4H1 (*Sorghum bicolor*, 002G126600), *Zm*C4H (*Zea mays*, PWZ11997.1), *Gm*C4H (*Glycine max*, NP_001237317.1), *Ob*C4H (*Oryza brachyantha*, XP_006654260.1), *At*C4H (*Arabidopsis thaliana*, NP_180607.1).

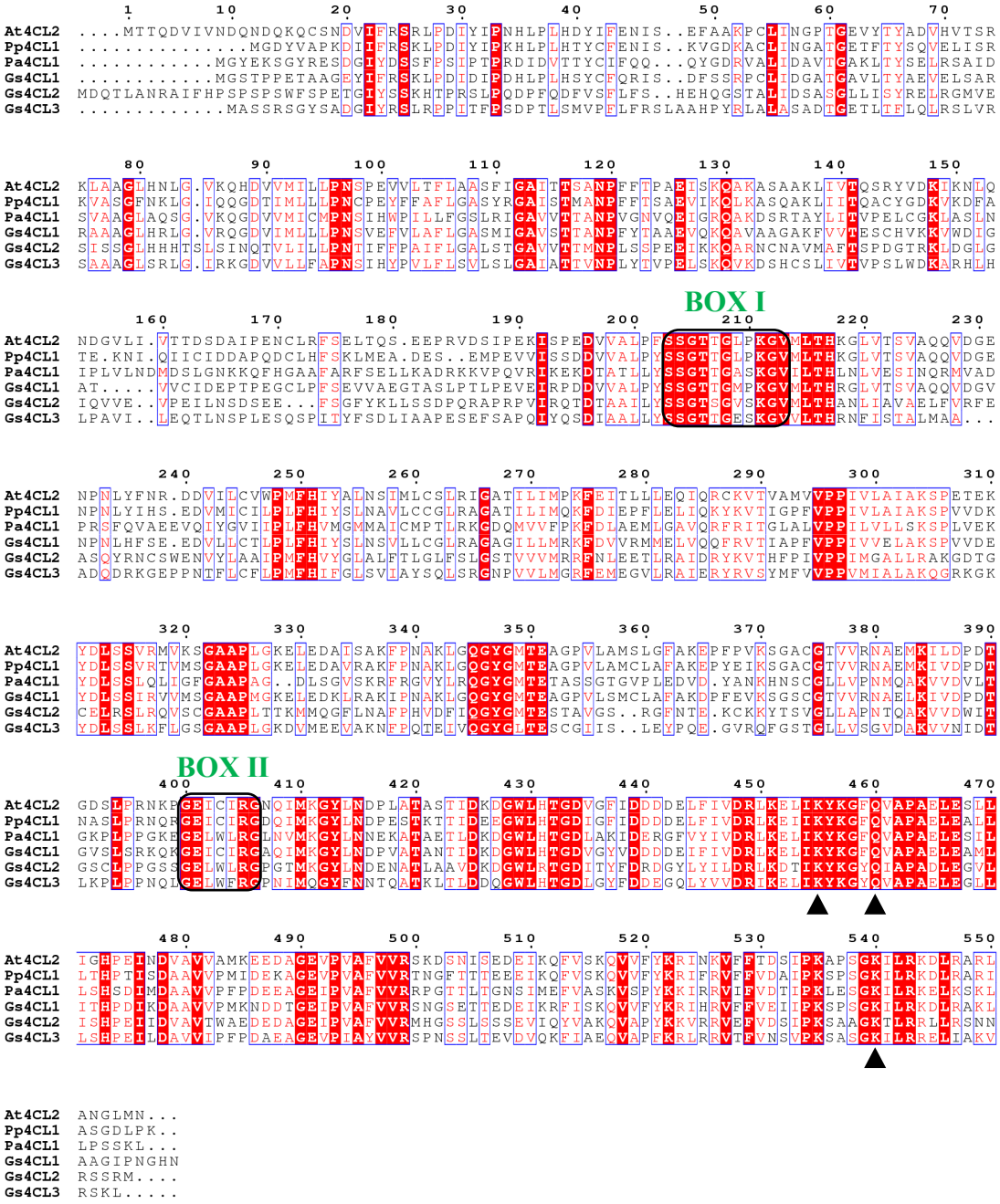

Figure S7. Multiple sequence alignment analysis of *Gs4CLs* and other 4CLs. The conserved peptide motifs BOX I and BOX II structural domains represent the conserved catalytic residues of the 4CL enzyme bound to the substrate. The black triangles indicate conserved amino acid residues, which are identified as catalytic centers for the thioester forming part of the adenylation reaction. The following sequences were analyzed in this study: *At*4CL2 (*Arabidopsis thaliana*, AAD47192.1), *Pp*4CL1(*Prunus persica*, XP007215519), *Pa*4CL1(*Plagiochasma appendiculatum*, KJ944317).

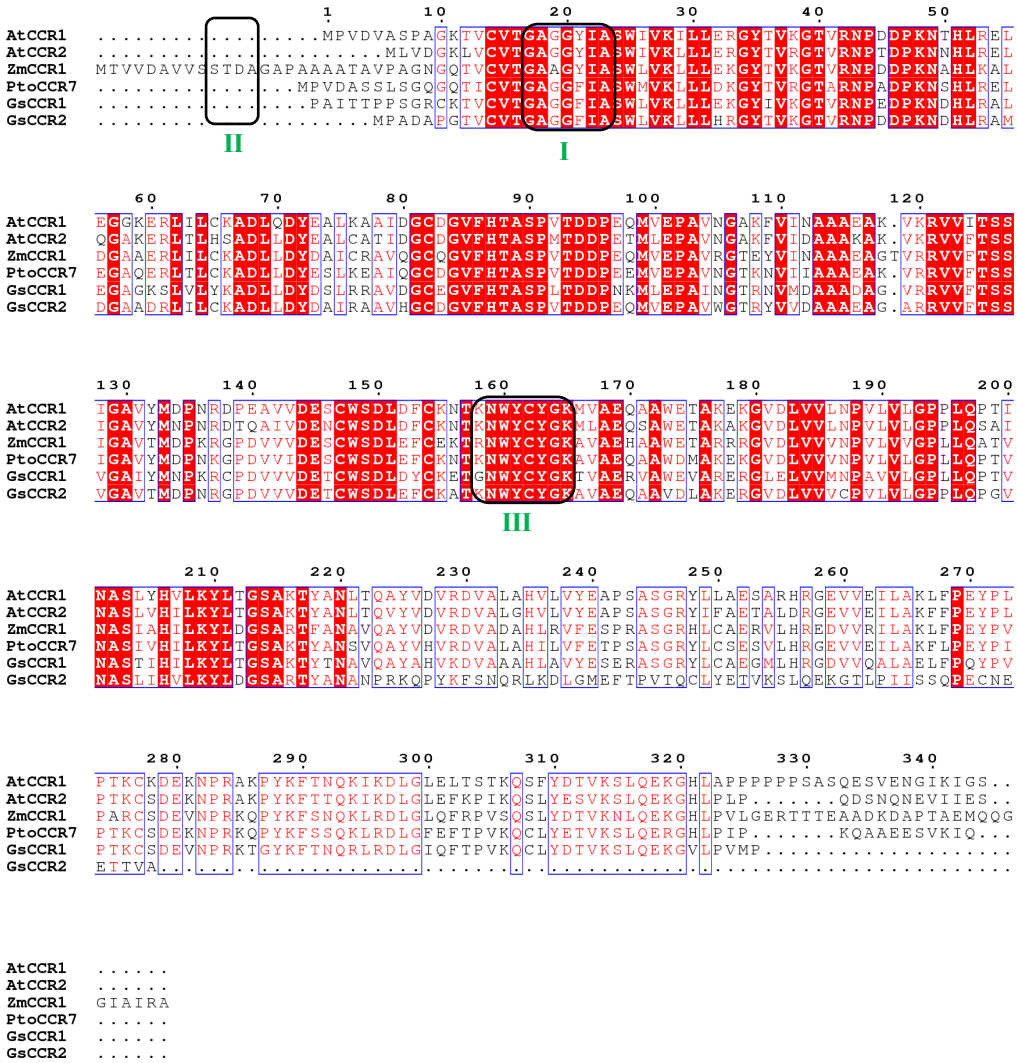

Figure S8. Multiple sequence alignment analysis of *GsCCRs* and other CCRs. I and II denote conserved regions involved in the N-terminal portion of the NADP(H) cofactor binding site (GXXGXXA, DXXD); III indicates a conserved structural domain (KNWYCYGK) distinguished from the NAD (H)-dependent short-chain dehydrogenase superfamily (SDR) gene. The following sequences were analyzed in this study: *At*CCR1 (*Arabidopsis thaliana*, AAG46037.1), *At*CCR2 (*Arabidopsis thaliana*, AAG53687.1), *Zm*CCR1 (*Zea mays*, NP_001105488), *Pto*CCR7 (*Populus tomentosa*, KF145198.1).

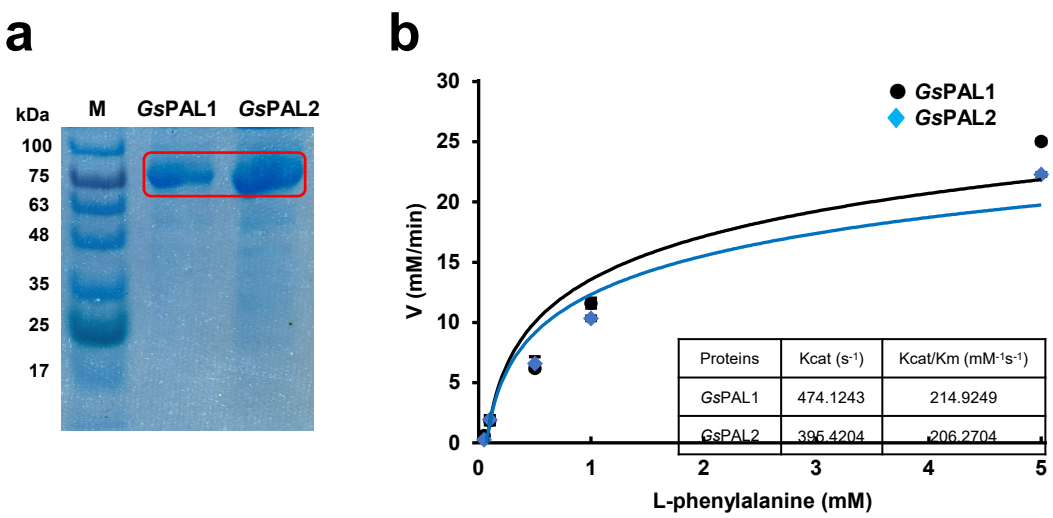

Figure S9. Heterologous expression and enzymatic assays result of *Gs*PALs. **(a)** Purification of *Gs*PALs proteins. SDS–PAGE with a 10 % separation gel showed the molecular weight, and lane M was the Color Mixed Protein Marker 180 (10-180 kDa). **(b)** The enzymatic kinetic curve of *Gs*PALs with feeding L-phenylalanine as substrate. ●/◆; n= 3±SD, V indicates the reaction rate of the enzyme at a certain amount.

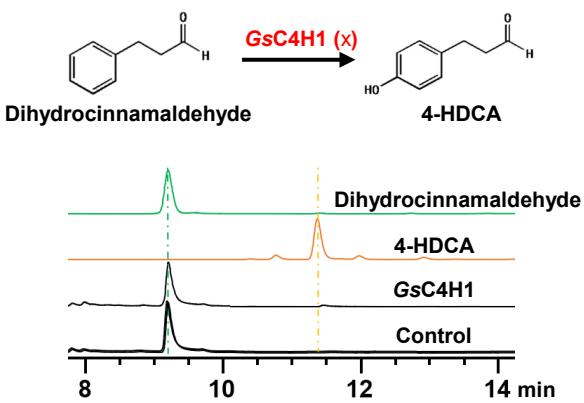

Figure S10. UPLC analysis results of *Gs*C4H1 enzymatic assays with feeding dihydrocinnamaldehyde as substrate.

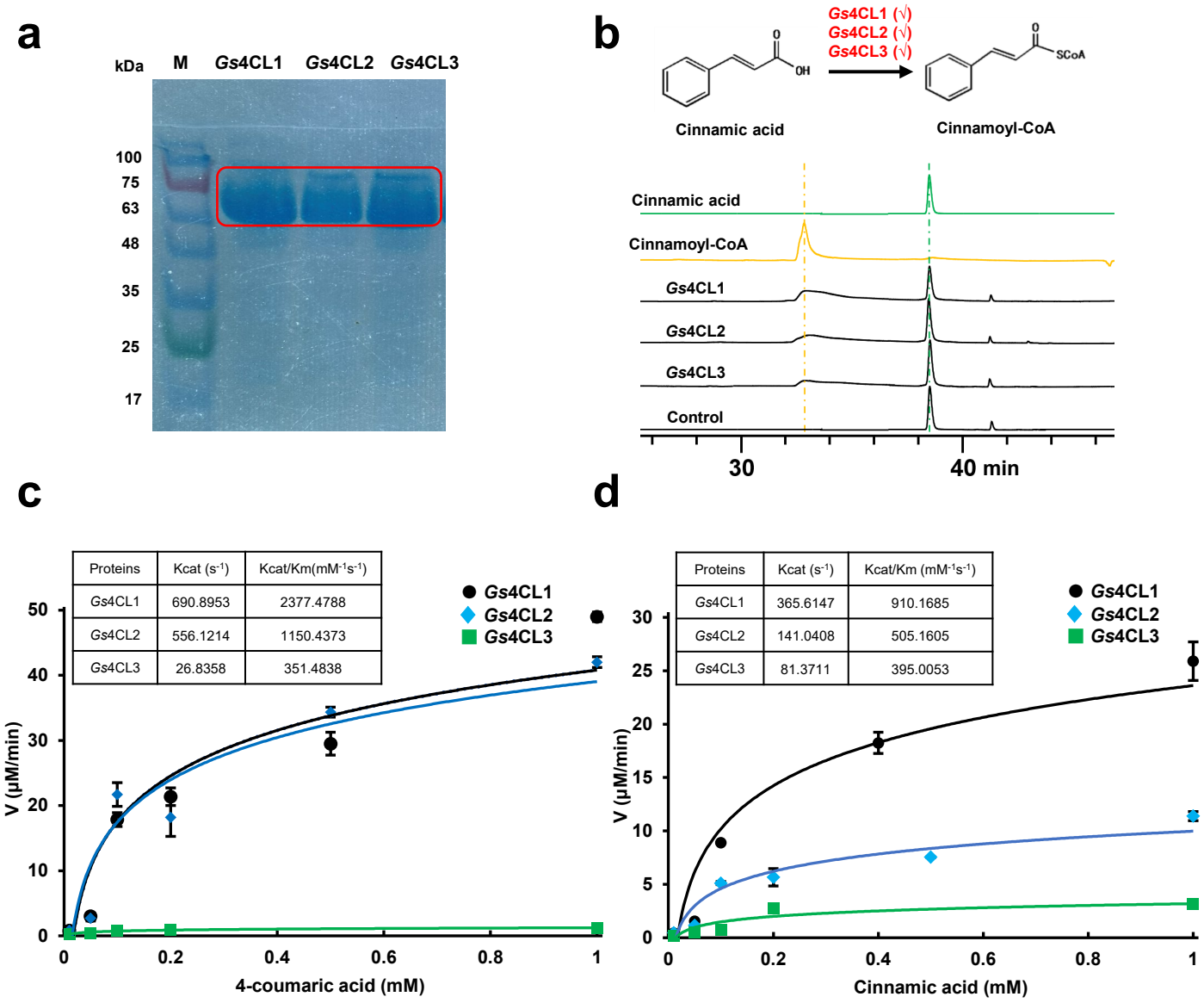

Figure S11. Heterologous expression and enzymatic assays result of *Gs*4CLs. **(a)** Purification of *Gs4C*Ls protein. SDS–PAGE with a 10 % separation gel showed the molecular weight, and lane M was the Color Mixed Protein Marker 180 (10-180 kDa). **(b)** *in vitro* enzymatic assays result of *Gs*4CLs with feeding cinnamic acid as substrate. **(c)** The enzymatic kinetic curve of *Gs*4CLs with feeding 4-coumaric acid as substrate. **(d)** The enzymatic kinetic curve of *Gs*4CLs with feeding cinnamic acid substrate. ●/◆/■; n= 3±SD, V indicates the reaction rate of the enzyme at a certain amount.

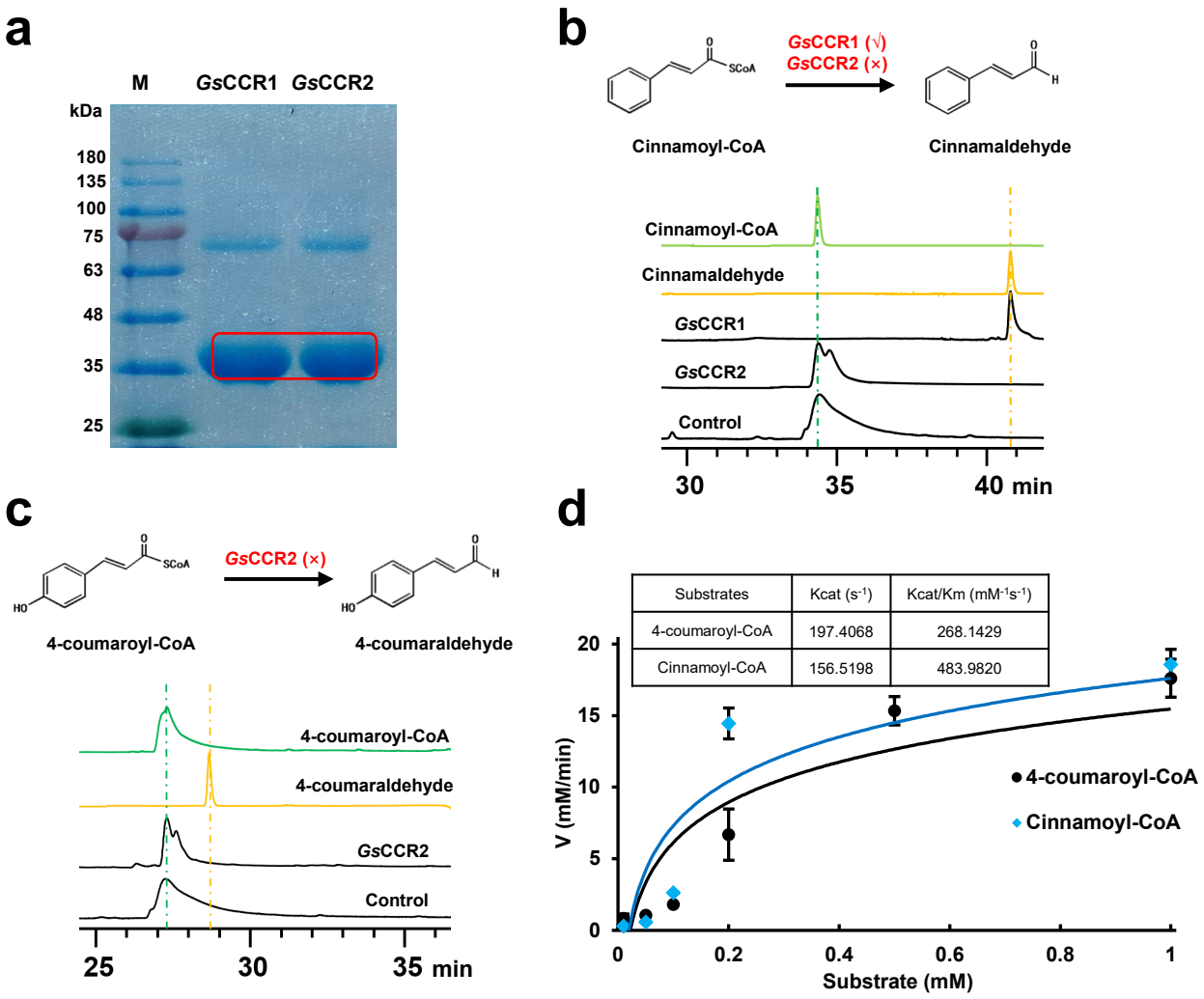

Figure S12. Heterologous expression and enzymatic assays result of *Gs*CCRs. **(a)** Purification of *Gs*CCRs protein. SDS–PAGE with a 10 % separation gel showed the molecular weight, and lane M was the Color Mixed Protein Marker 180 (10-180 kDa). **(b)** *in vitro* enzymatic assays result of *Gs*CCRs with feeding cinnamoyl-CoA as substrate. **(c)** *in vitro* enzymatic assays result of *Gs*CCR2 with feeding 4-coumaroyl-CoA as substrate. **(d)** The enzymatic kinetic curve of *Gs*CCR1s with feeding cinnamoyl-CoA and 4-coumaroyl-CoA as substrate. ●/◆; n= 3±SD, V indicates the reaction rate of the enzyme at a certain amount.

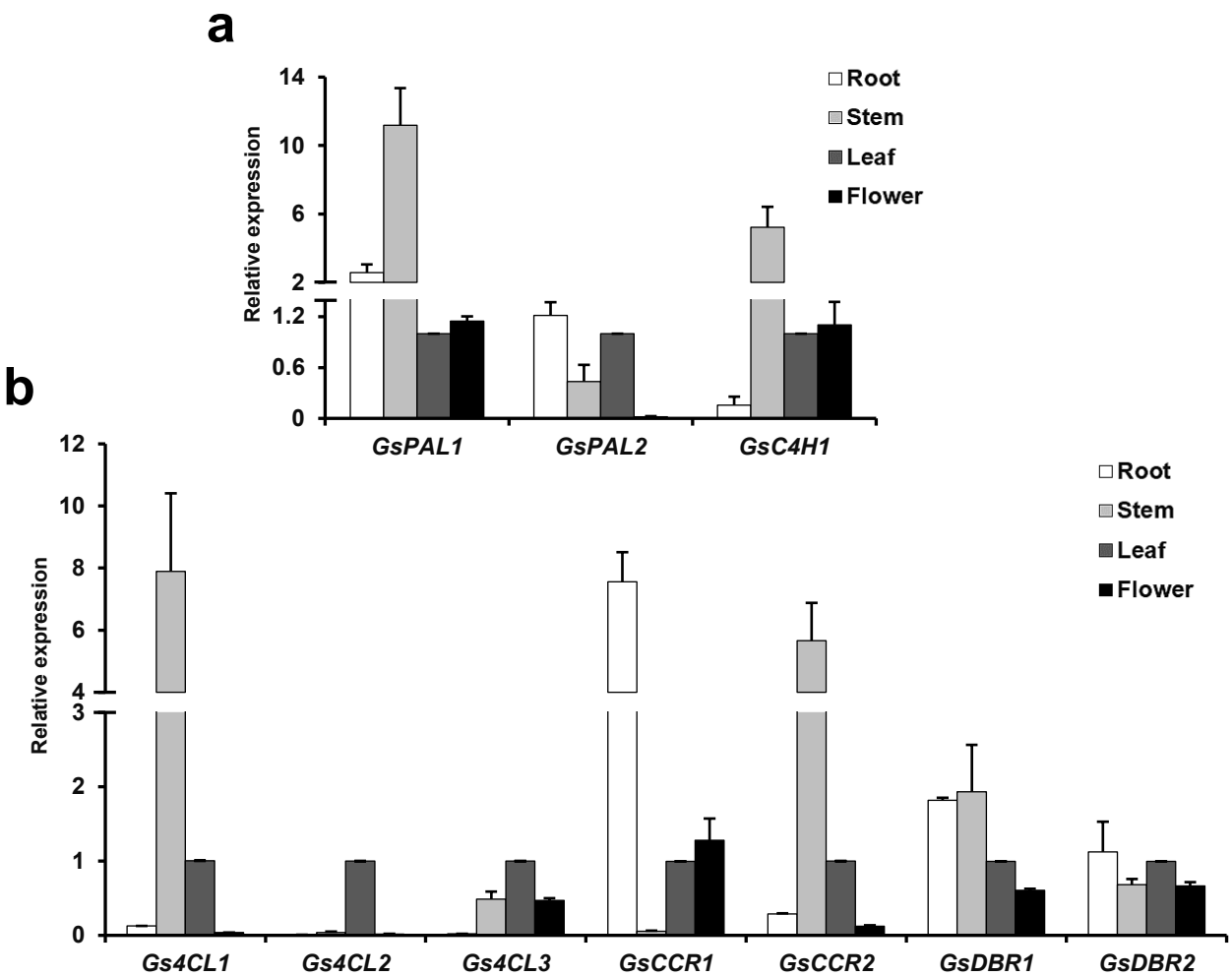

Figure S13. Expression analysis of candidate genes related to 4-HDCA biosynthesis in *G. superba*. The error line is expressed as the standard deviation obtained from at least three replicates. *Gs*, *Gloriosa superba* L.; *PAL*, phenylalanine ammonia-lyase; *C4H*, cinnamate 4-hydroxylase; *4CL*, 4-coumarate CoA ligase; *CCR*, cinnamoyl CoA reductase; *DBR*, NADPH-dependent double bond reductases.

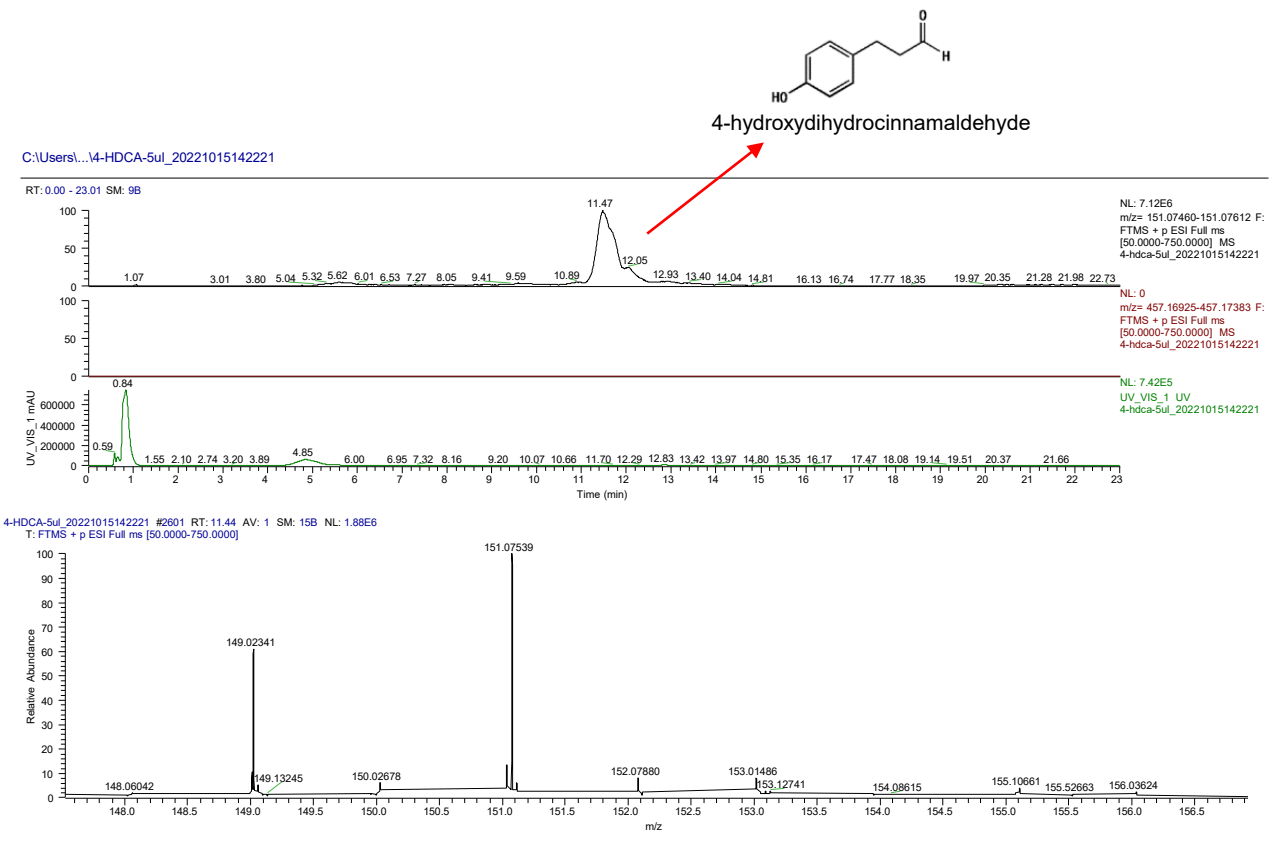

Figure S14. The catalytic product of 4-HDCA was further confirmed by UPLC-MS in $[M+H]^+$ mode (m/z 151.07539).

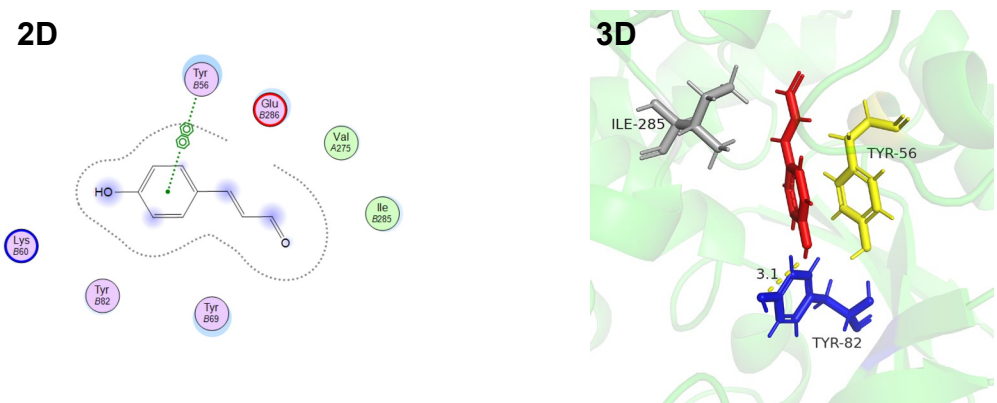

Figure S15. Molecular docking and modeling results of *Gs*DBR2 protein.
Note: 2D and 3D maps of *Ca*DBR2 protein docking with 4-coumaraldehyde. Small molecules of 4-coumaraldehyde are modified to red; the receptor skeleton is displayed in 80 % transparent green cartoon form; other colors are molecular docking amino acid residues.