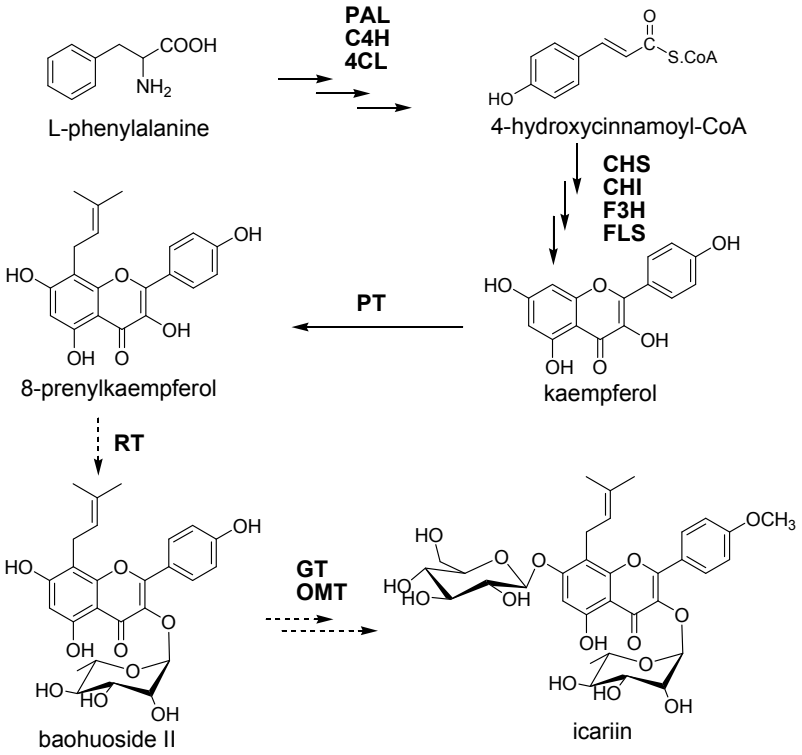
Supporting Information for

A regiospecific rhamnosyltransferase from *Epimedium pseudowushanense* catalyzes the 3-O-rhamnosylation of prenylflavonols

Keping Feng,[a] Ridao Chen,[a] Kebo Xie,[a] Dawei Chen,[a] Baolin Guo,[b] Xiao Liu,[c] Jimei Liu,[a] Min Zhang,[a] and Jungui Dai [a]

[a] State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Institute of Materia Medica, Peking Union Medical College & Chinese Academy of Medical Sciences, 1 Xian Nong Tan Street, Beijing 100050, China.
E-mail: jgdai@imm.ac.cn
[b] Institute of Medicinal Plant Development, Peking Union Medical College & Chinese Academy of Medical Sciences, 151 Malianwa North Road, Beijing 100193, China.
[c] Modern Research Center for Traditional Chinese Medicine, Beijing University of Chinese Medicine, 11 North Third Ring Road, Beijing 100029, China.

Table of Contents



Scheme. S1 Proposed biosynthetic pathway of icariin in *Epimedium*. PAL, phenylalanine ammonia lyase; C4H, cinnamic acid 4-hydroxylase; 4CL, 4-coumaric acid ligase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3β-hydroxylase; FLS, flavonol synthase; PT, prenyltransferase; RT, rhamnosyltransferase; GT, glucosyltransferase; OMT, *O*-methyltransferase.

Table S1. The HPLC method used in analysis of enzymatic reactions *in vitro* in this study

Time (min)	Solvent A (%) 0.1% formic acid	Solvent B (%) CH_3OH	flow rate (mL min^{-1})
0.00	90.0	10.0	1.0
10.00	50.0	50.0	1.0
30.00	0.0	100.0	1.0
45.00	0.0	100.0	1.0

Table S2. The HPLC method used in analysis of whole cell bioconversion in this study

Time (min)	Solvent A(%) 0.1% formic acid	Solvent B(%) CH_3OH	flow rate (mL min^{-1})
0.00	50.0	50.0	1.0
20.00	0.0	100.0	1.0

Table S3. Details of the plasmids and strains used in this study

plasmids or *E. coli* strain	relavent properties
	Plasmids
pET-28a	pBR322 ori, Kan^r
pCDFDuet	CloDE13 ori, Str^r
pE-PF3RT	pET-28a + *EpPF3RT*
pE-RhS	pET-28a + *EpRhS*
pC-PF3RT-RhS	pCDFDuet + *EpPF3RT* + *EpRhS*
	Primers
pE-PF3RT forward	CAAATGGGTCGCGGATCCATGAGTTGTATTTCAGTC
pE- PF3RT reverse	GTGGTGGTGGTGCTCGAGTCAAACACCTTTAGCAAGTTT
pC- PF3RT forward	GAATTCATGAGTTGTATTTCAGTC
pC- PF3RT reverse	GTCGACTCAAACACCTTTAGCAAGTT
pE-RhS forward	CAAATGGGTCGCGGATCCATGGCAACTTATACGCCAAAAAACA TC
pE- RhS reverse	GTGGTGGTGGTGCTCGAGTTAGATAGTTGTTTTCTTGTTGGGTTC
pC- RhS forward	GAAGGAGATATACATATGGCAACTTATACGCCAAAAAAC
pC- RhS reverse	TTCTTTACCAGACTCGAGTTAGATAGTTGTTTTCTTG
	Strains
T-pE	Transetta (DE3) harboring empty pET-28a
T-RT	Transetta (DE3) harboring pE- PF3RT
T-RhS	Transetta (DE3) harboring pE-RhS
B-pC	BL21(DE3) harboring empty pCDFDuet
B-PKR	BL21(DE3) harboring pC- PF3RT -RhS

Table S4. ^{1}H and ^{13}C NMR data of **1a** and **1**.

H	δ_H		C	δ_c	
	1a	**1**		**1a**	**1**
H-6	6.30 (1H, s)	6.25 (1H, s)	C-2	156.8	146.22
H-3',5'	6.90 (2H, d, *J* = 8.8 Hz)	7.14 (2H, d, *J* = 8.0 Hz)	C-3	134.4	136.13
H-2', 6'	7.74 (2H, d, *J* = 8.8 Hz)	8.16 (2H, d, *J* = 8.0 Hz)	C-4	177.9	176.46
H-11	3.48 (2H, m)	3.45 (2H, m)	C-5	161.6	161.80
H-12	5.14 (1H, t, *J* = 7.2 Hz)	5.20 (1H, t)	C-6	98.3	99.31
H-14	1.61 (3H, s)	1.63 (3H, s)	C-7	161.3	164.25
H-15	1.67 (3H, s)	1.73 (3H, s)	C-8	105.9	108.93
H-3',5'	6.90 (2H, d, *J* = 8.8 Hz)	7.14 (2H, d, *J* = 8.0 Hz)	C-9	153.8	157.72
H-1"	5.26 (1H, d, *J* = 1.6, Hz)		C-10	104.2	105.11
protons in rhamnose	3.10–3.98		C-1'	122.4	123.44
			C-2', 6'	130.4	128.91
			C-3', 5'	114.1	116.46
			C-4'	158.8	160.58
			C-11	21.2	21.71
			C-12	122.3	121.57
			C-13	131.0	131.95
			C-14	25.4	25.49
			C-15	17.8	19.18
			C-1"	101.8	

	C-2"	70.1
	C-3"	70.6
	C-4"	70.3
	C-5"	71.1
	C-6"	17.5

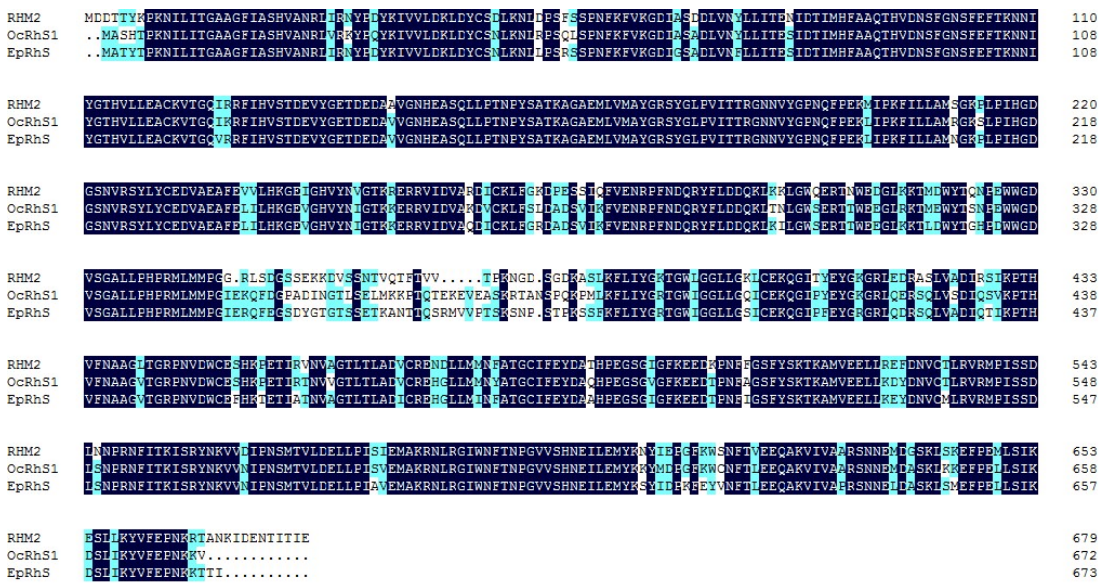


Fig. S1. Multiple alignment of the amino acid sequences of EpRhS, OcRhS (GenBank accession number ANK57460, from *Ornithogalum longebracteatum*) and RHM2 (GenBank accession number NM_104228, from *Arabidopsis. thaliana*).

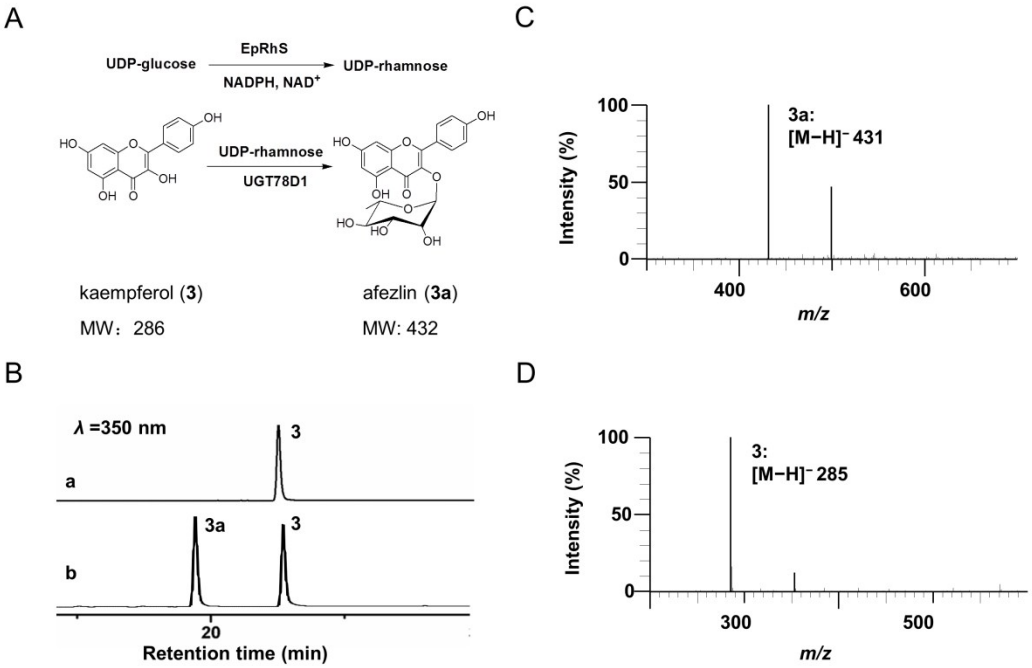


Fig. S2. Functional characterization of recombinant EpRhS combined with AtUGT78D1. A) Enzymatic reaction catalysed by EpRhS and AtUGT78D1; B) HPLC/ESI-MS analysis of the enzymatic reaction of AtUGT78D1 combined with kaempferol (**3**). a) Standard of **3**; b) Recombinant AtUGT78D1 with EpRhS reaction mixture; C), D) MS spectrum of **3** and **3a** at negative mode.

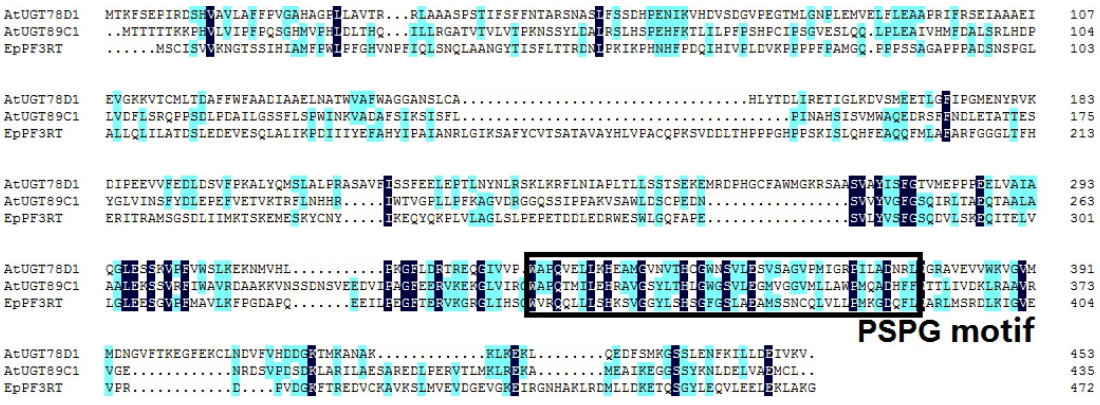


Fig. S3 Multiple alignment of the amino acid sequences of EpPF3RT, AtUGT78D1 (GenBank accession number NM_102790, from *Arabidopsis thaliana*) and AtUGT89C1 (GenBank accession number XP_002892315, from *A. thaliana*). The black box indicates the conserved region of plant secondary product glycosyltransferases (PSPG motif)

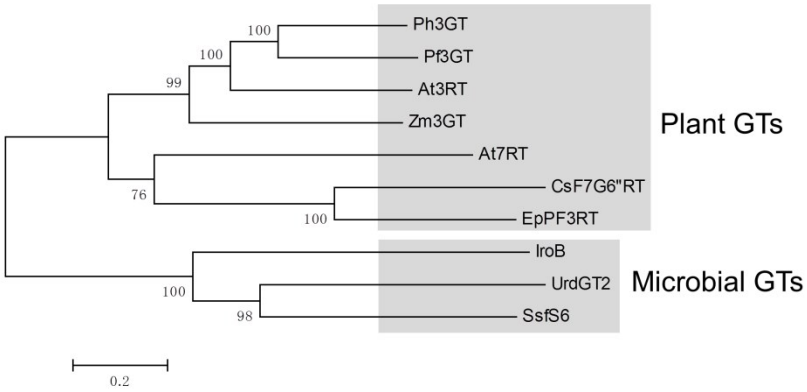


Fig. S4. The phylogenetic relationships between EpPF3RT and other plant GTs and microbial GTs. The results were calculated using the MEGA version 5 software. Method: Neighbor-joining, Bootstrap: 1000. The branch lengths represent relative genetic distances. The protein sequences and corresponding accession numbers that were used for this comparison are as follows: Ph3GT (GenBank accession number AB027454, from *Petunia x hybrida*); Pf3GT (GenBank accession number AB002818, from *Perilla frutescens*); At3RT (GenBank accession number NM_102790, from *Arabidopsis thaliana*); Zm3GT (GenBank accession number X13501, from *Zea mays*); CsF7G6"Rt (GenBank accession number NP_001275829, from *Citrus sinensis*); At7RT (GenBank accession number XP_002892315, from *A. thaliana*);. UrdGT2 (GenBank accession number AAF00209, from *Streptomycesı fradiae*); IroB (GenBank accession number CAE55724, from *Escherichia coli*); SsfS6 (GenBank accession number ADE34512, from *Streptomyces sp*);

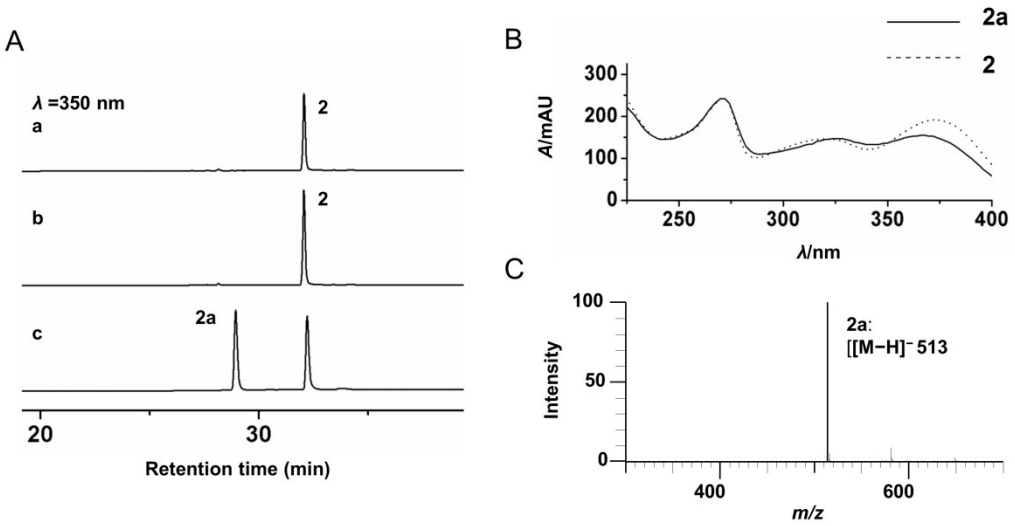


Fig. S5. HPLC-DAD/ESIMS analysis of EpPF3RT enzyme products using **2** as an aglycon acceptor. A) HPLC/ESI-MS analysis of the enzymatic reaction with anhydroicaritin (**2**). a) Standard of **2**; b) Control group; c) Recombinant EpPF3RT; B) UV spectra of **2** and **2a**. C) MS spectrum of **2a** at negative mode.

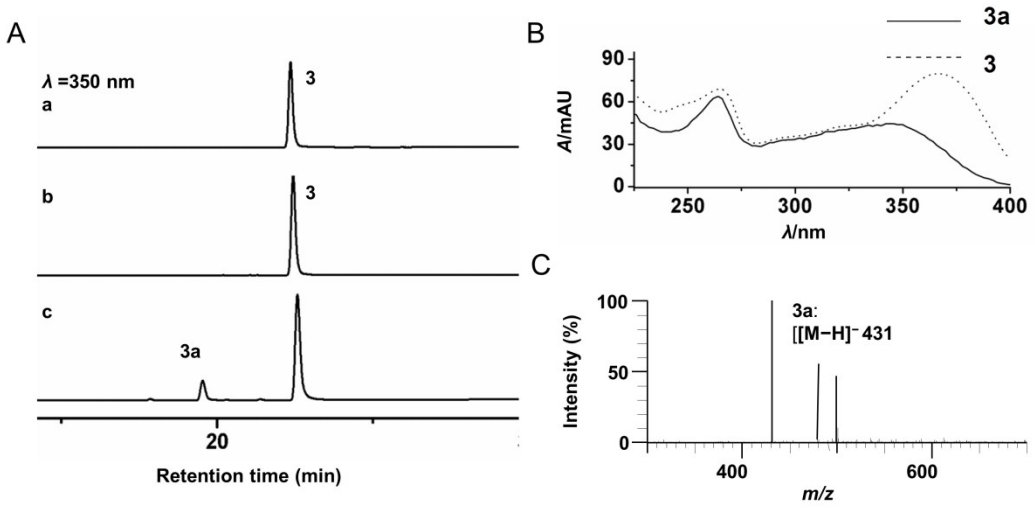


Fig. S6. HPLC-DAD/ESIMS analysis of EpPF3RT enzyme products using **3** as an aglycon acceptor. A) HPLC/ESI-MS analysis of the enzymatic reaction with kaempferol (**3**). a) Standard of **3**; b) Control group; c) Recombinant EpPF3RT; B) UV spectra of **3** and **3a**. C) MS spectrum of **3a** at negative mode.

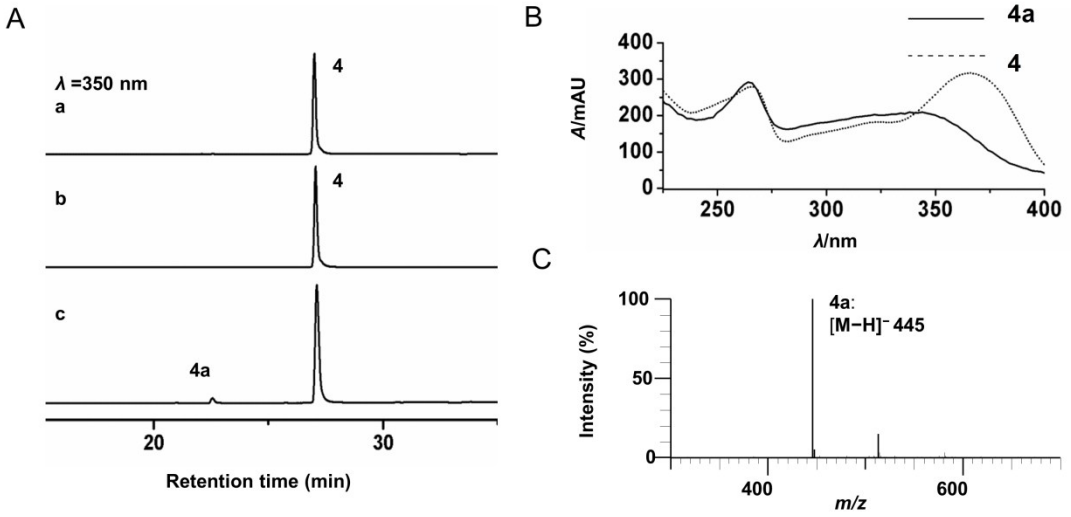


Fig. S7. HPLC-DAD/ESIMS analysis of EpPF3RT enzyme products using **4** as an aglycon acceptor. A) HPLC/ESI-MS analysis of the enzymatic reaction with kaempferide (**4**). a) Standard of **4**; b) Control group; c) Recombinant EpPF3RT; B) UV spectra of **4** and **4a**. C) MS spectrum of **4a** at negative mode.

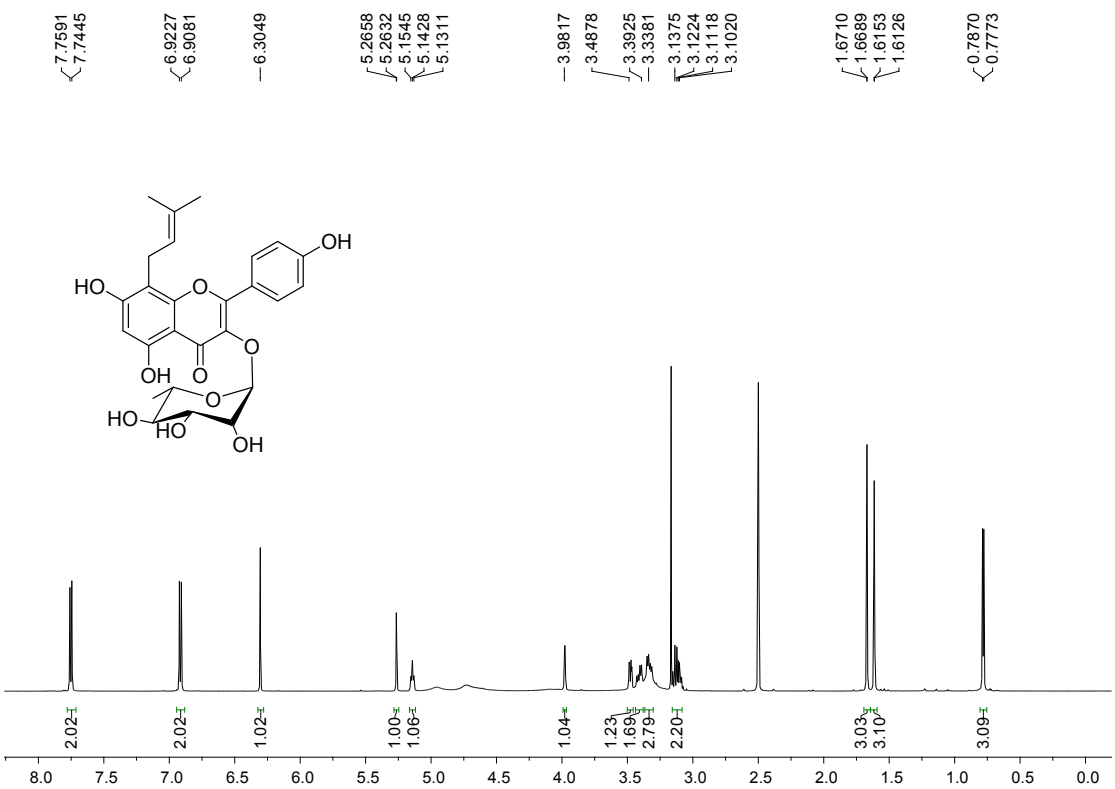


Fig. S8. ^{1}H NMR spectrum of **1a** (DMSO-d_6, 600 MHz).

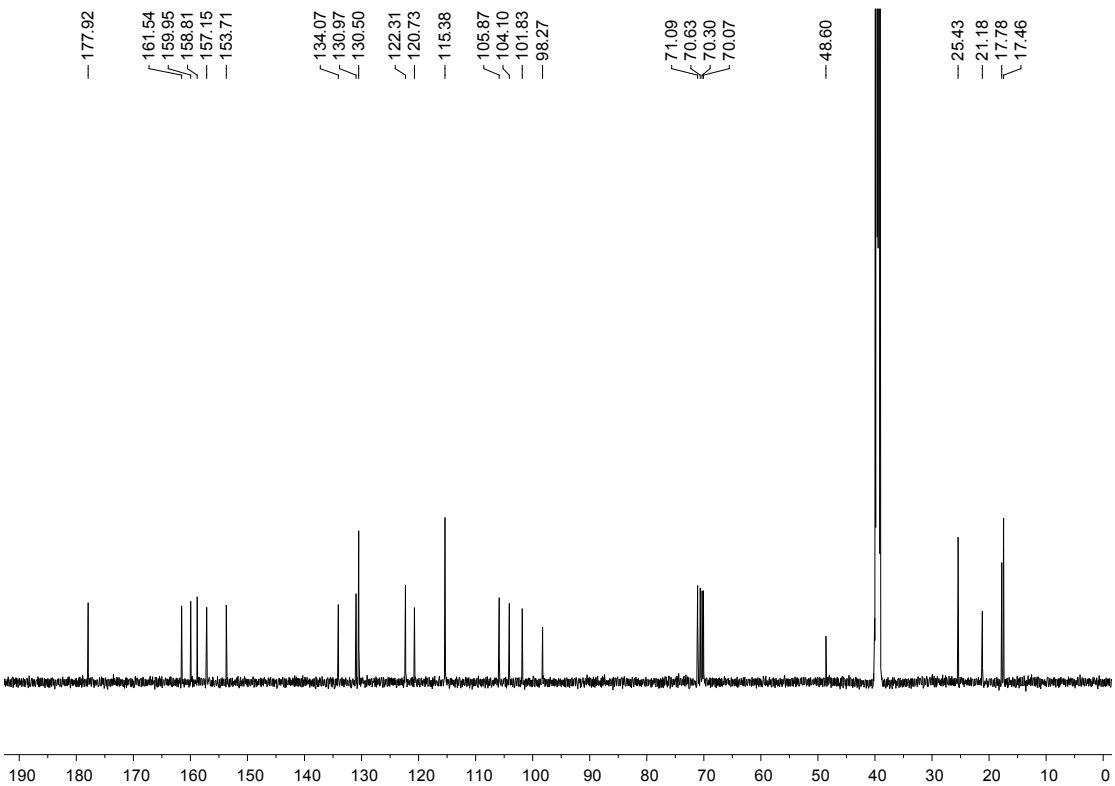


Fig. S9. ^{13}C NMR spectrum of **1a** (DMSO-d_6, 125 MHz).

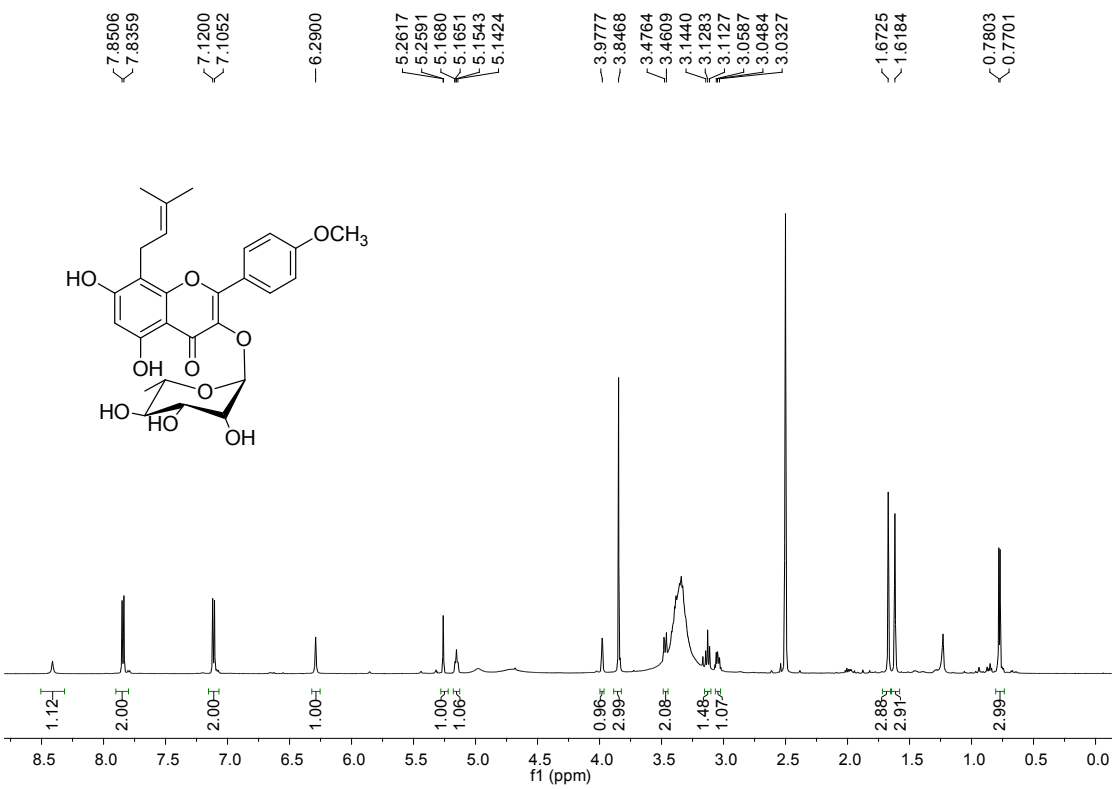


Fig. S10. ^{1}H NMR spectrum of **2a** (DMSO-d_6, 600 MHz).

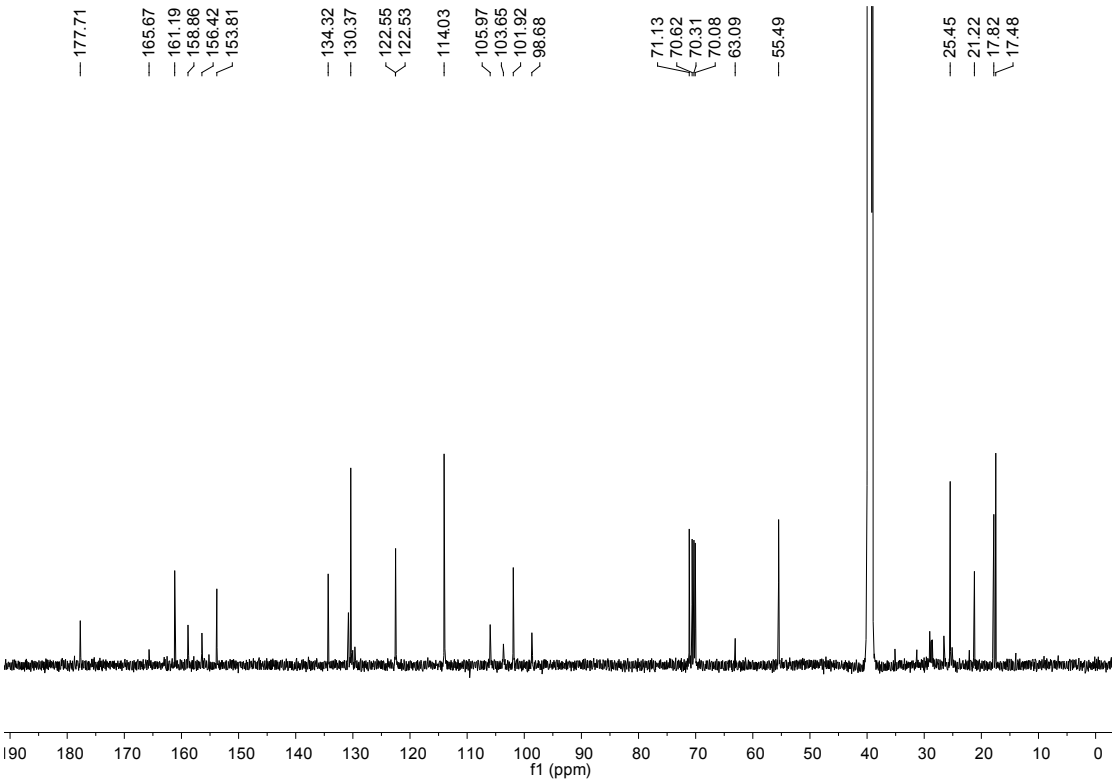


Fig. S11. ^{13}C NMR spectrum of **2a** (DMSO-d_6, 125 MHz).

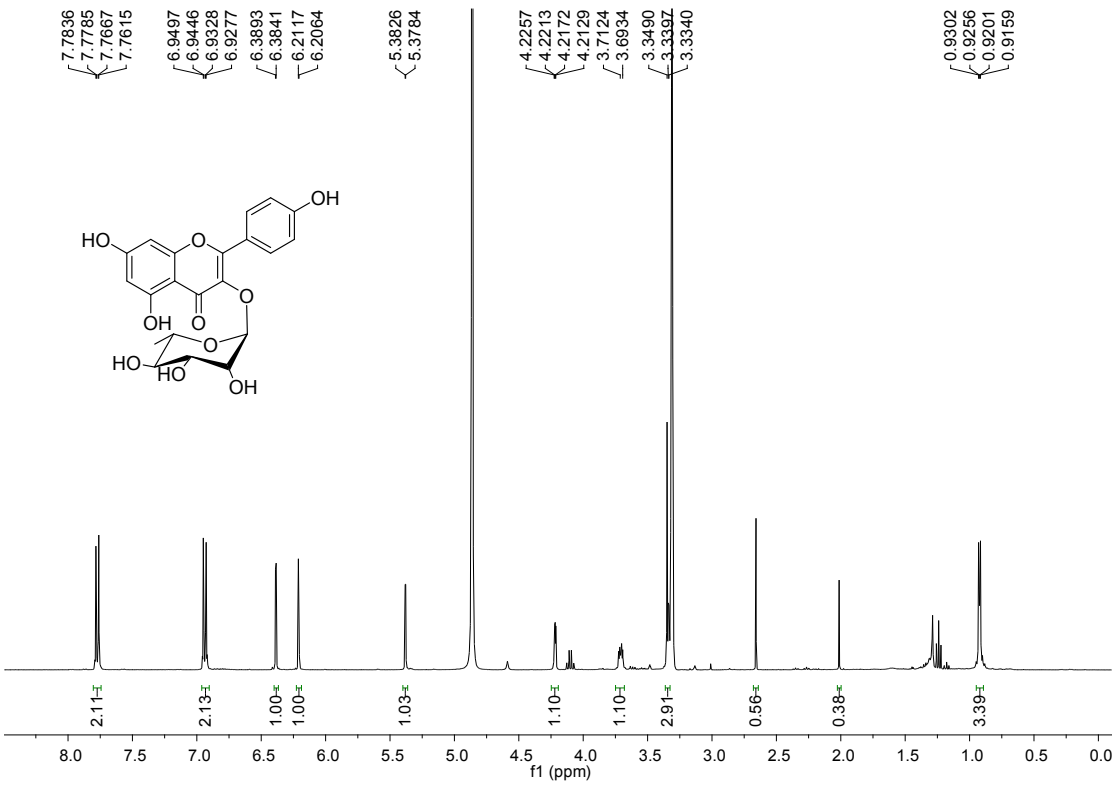


Fig. S12. ^{1}H NMR spectrum of **3a** (Methanol-d_4, 400 MHz).

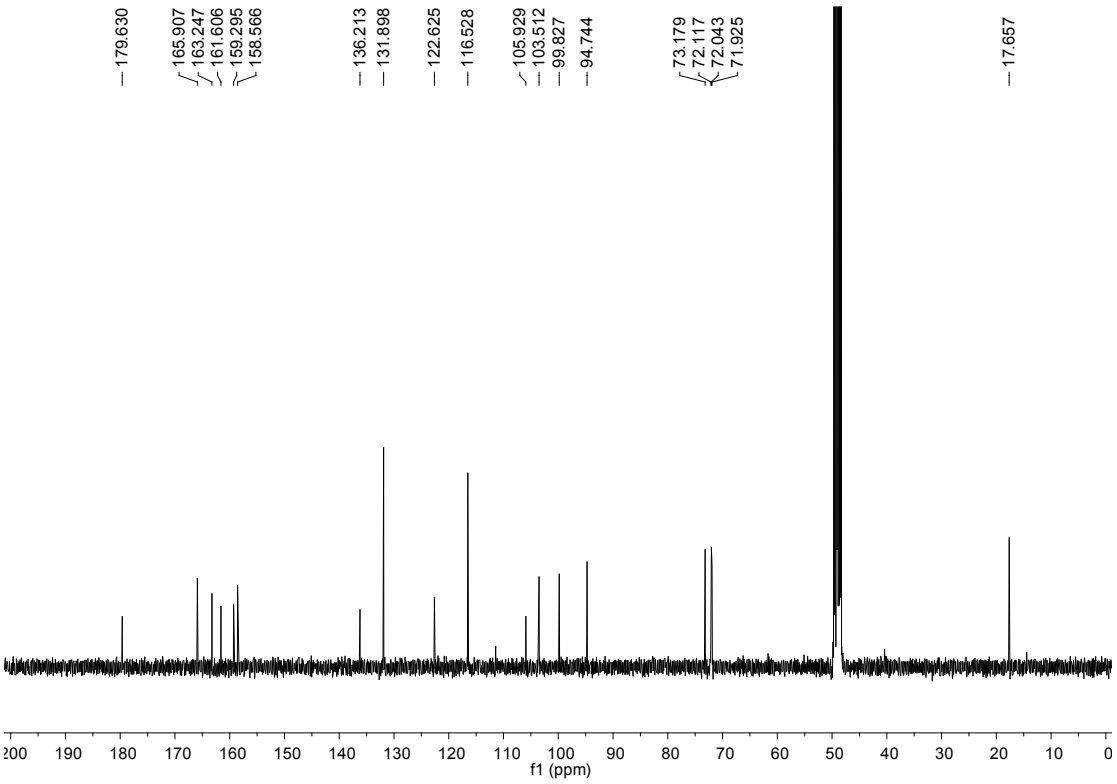


Fig. S13. ^{13}C NMR spectrum of **3a** (Methanol-d_4, 100 MHz).

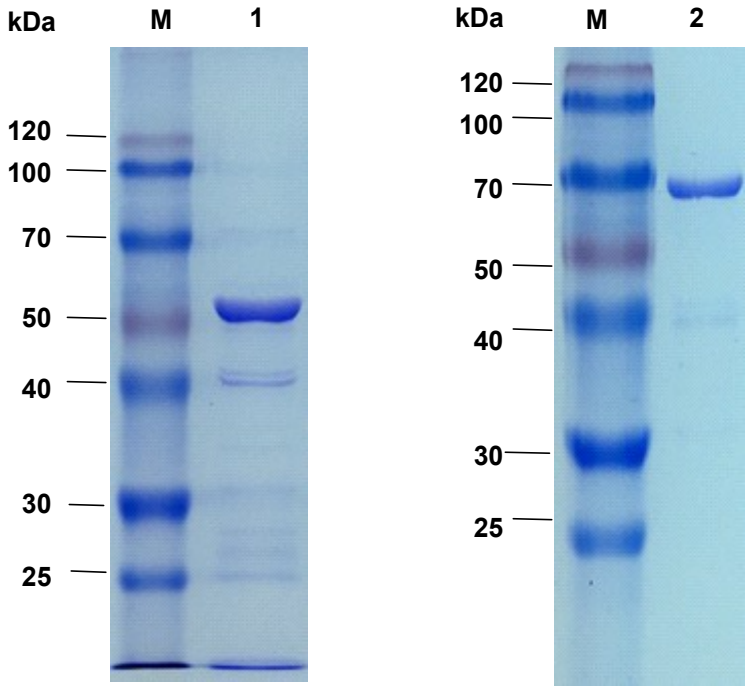


Fig. S14. SDS-PAGE of recombinant His_6-EpPF3RT and His_6-EpRhS purified by affinity chromatography. Lane M: Protein Marker; Lane 1: His-tagged EpPF3RT (predicted M.W., 52.1 kDa) purified on Ni Sepharose. Lane 2: His-tagged EpRhS (predicted M.W., 75.6 kDa) purified on Ni Sepharose.

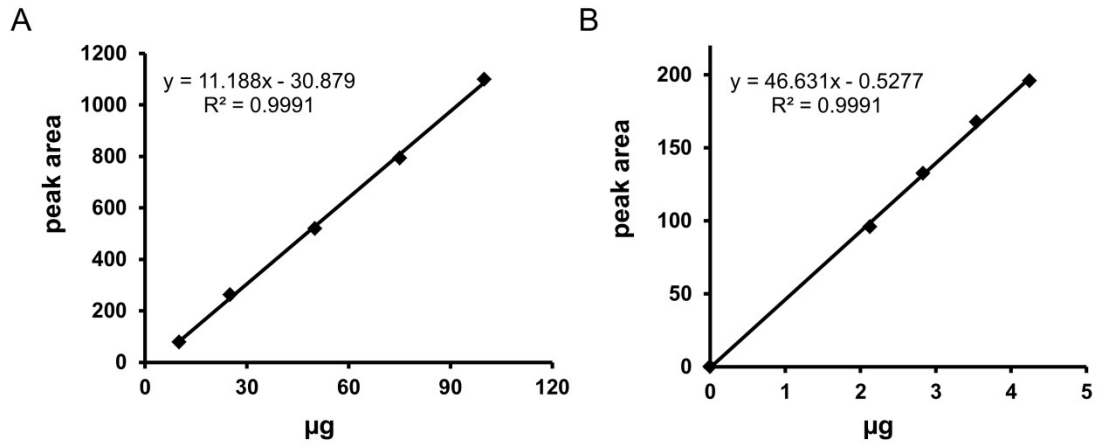


Fig. S15. The linear regression models and the regression equations of the external standard methods established for the quantitative analysis of baohuoside II (A) and 8-prenylkaempferol (**1**) (B).