npg

LETTER TO THE EDITOR

Production of bioactive ginsenoside compound K in metabolically engineered yeast

Dear Editor,

Ginseng has been traditionally used as herbal medicine in Asia for thousands of years to enhance physical performance and to increase resistance to stress and aging, and has been developed into various kinds of dietary supplement with increasing market demand [1]. Its active constituents are ginsenosides, a group of triterpene saponins (*ca.* 2% in *Panax ginseng* dried roots) [1]. However, the main functional component detected in mammalian blood or organs after oral administration of ginseng or ginsenosides is compound K (CK) [2], which presents bioactivities of anti-inflammation, hepatoprotection, anti-diabetes and anti-cancer in experiments of *in vitro* cell biology and/or *in vivo* animal models (Supplementary information, Table S1). Recently, CK was approved by the China Food and Drug Administration to commence clinical trials (CDEL20130379) for arthritis prevention and treatment.

CK, which has never been identified in *Panax* plants [2], is currently manufactured by bio-deglycosylation of major protopanaxadiol (PPD)-type ginsenosides (mainly Rb1, Rb2, Rd and Rc, which normally bear sugar chains with two or more glucose/arabinose moieties at the C-3 and/or C-20 of PPD) (Supplementary information, Figure S1A) [3]. However, due to the long cultivation periods required for the growth of qualified roots (5-7 yrs) and long rotational tillage for replantation of *Panax* plants (≥ 5 yrs) [4], the availability of ginsenosides for CK manufacturing is limited. On the other hand, large scale production of CK by chemical total synthesis is not practical particularly due to the difficulty for selective glycosylation of the spatially hindered C-20*S*-OH [5]. In this study, we explored a synthetic biology strategy to construct chassis cells expressing heterologous genes encoding the whole pathway(s) for CK synthesis so that large quantities of CK may be produced from cheap monosaccharide via microbial fermentation, as in the case of artemisinic acid [6].

As briefly illustrated in Supplementary information, Figure S1B, the biosynthesis of PPD (1a) from dam-marenediol II (DM) (1b) was achieved recently via co-expressing cDNAs encoding a cytochrome P450 (CY-P716A47) identified from *P. ginseng* and an NADPH-cytochrome P450 reductase ATR2-1 from *Arabidopsis thaliana* [7]. Catalyzed by a dammarenediol synthase isolated from *P. ginseng* (PgDDS), DM (1b) could be synthesized from 2,3-(*S*)-oxidosqualene (1c) [8], which, together with squalene (1d) and other intermediates, could be traced back to the isoprenoid precursors produced by the well studied mevalonate (MEV) pathway in yeast. Therefore, it is reasonable to hypothesize that CK may be produced in a PPD-producing chassis if a novel UDP-glycosyltransferase (UGT) enabling the selective glycosylation of C-20*S*-OH of PPD is expressed in the cell.

A proprietary cDNA database including 479 689 assembled cDNA contigs was established based on 9 *Panax* EST datasets available from the NCBI GenBank (Supplementary information, Table S2). Among them, we identified 512 contigs potentially encoding plant UGTs (with PSPG-box, the consensus sequence of Plant Secondary Product Glycosyltransferase); and the open reading frames (ORFs) predicted for 158 UGT contigs > 1 320 bp in length were clustered into 42 operational taxonomic units (OTUs) based on protein sequences at 95% similarity cutoff. Among these OTUs, 16 ORFs of the *Panax* UGTs (UGTPgs) were amplified from mRNAs prepared from a *P. ginseng* callus by RT-PCR and assigned to several plant UGT clusters with functions in biosynthesis of triterpene, flavonol, anthocyanin, sterol and other natural products as shown in the databases of CAZy and/or UniProtKB/Swiss-Prot (Supplementary information, Table S3 and Figure S2).

Chassis vectors expressing ORFs of the 16 UGTPgs (pHCD-UGTPg1 to pHCD-UGTPg16), were transformed into a chassis yeast, *Saccharomyces cerevisiae* strain BA21 individually [7] (Supplementary information, Figure S3) and yielded 16 recombinant yeast strains, designated AK1 to AK16 correspondingly. The individual *n*-butanol extract of the supernatants of the disrupted yeast cells cultured in galactose media was analyzed by HPLC
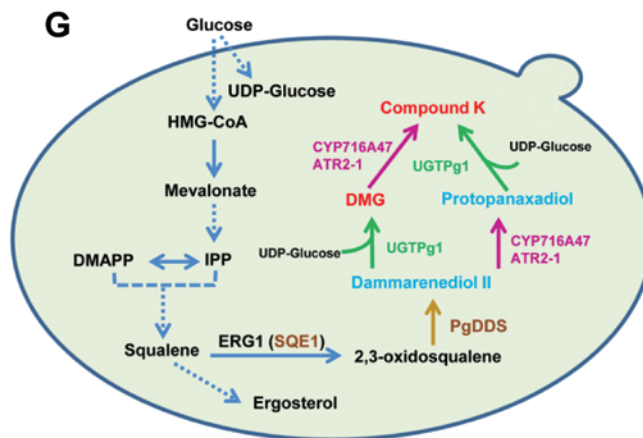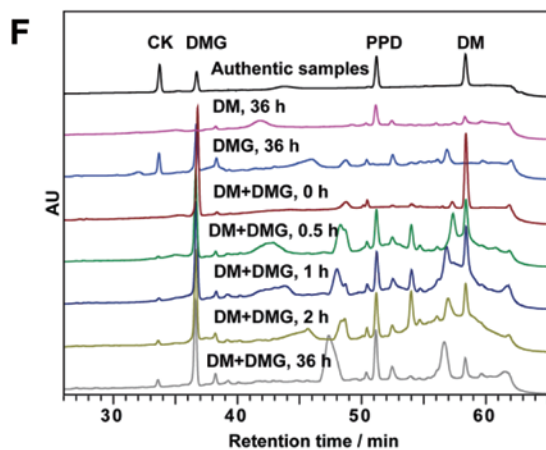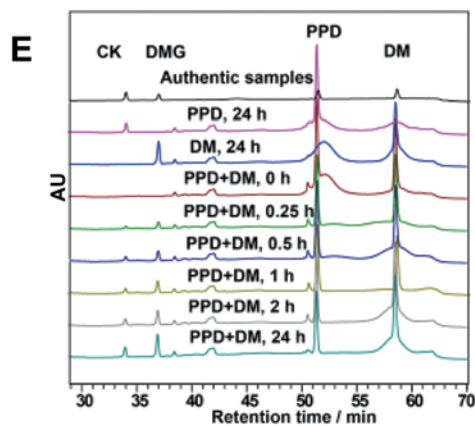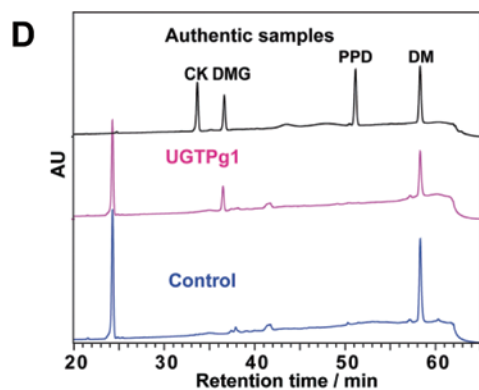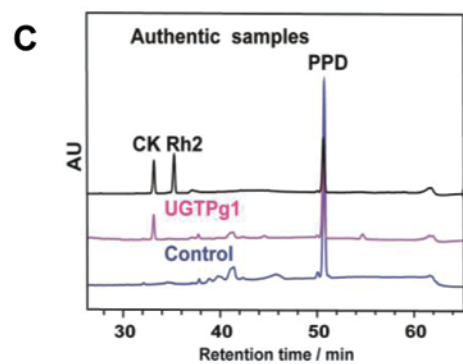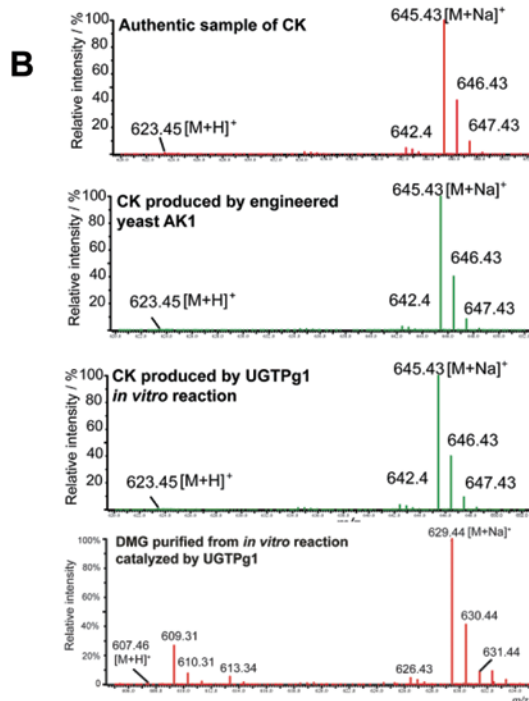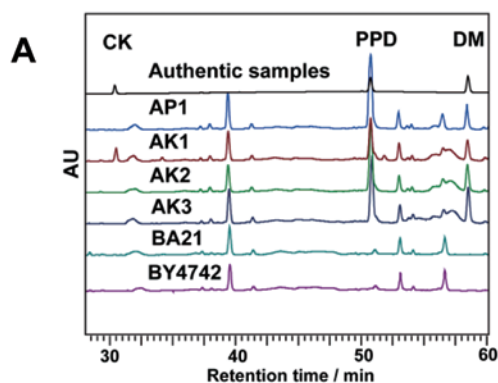
(Figure 1A) and a perspective CK compound (*ca.* 240 μg/L) was detected in AK1 harboring the UGTPg1 gene, but not in any of the other recombinant strains (Figure 1A). After semi-preparative HPLC purification, the structure of the perspective CK compound was verified by HPLC/ESIMS (Figure 1B), ^{1}H and ^{13}C NMR analyses (Supplementary information, Figure S4A and S4B), to be chemically identical to the authentic CK.

By replacing the galactose inducible promoters in AK1 to the constitutive promoters, *TEF1p* and *GPM1p*, CK was produced from glucose with a yield of *ca.*150 μg/L in this new chassis strain, BK1 (Supplementary information, Table S4). We further improved the productivity of isoprenoids and/or terpenoids in the chassis yeasts by overexpression of tHMGR (truncated HMGR) and UPC2.1(a semidominant mutant allele of UPC2 which is a global transcription factor for sterols biosynthesis) via transforming the pLLeu-tHMGR-UPC2.1 plasmid (Supplementary information, Data S1) into AK1 and BK1. As expected, this resulted in about 5-fold increase of the yield of CK, reaching up to 1.4 and 0.8 mg/L in the new chassis strains AKE and BKE, respectively (Supplementary information, Table S4). It is expected that the CK yield could be increased by further pathway engineering and optimization of fermentation conditions, as in the case of artemisinic acid [6] and PPD [9].

We successfully demonstrated that the heterologously expressed UGTPg1 (Supplementary information, Figure S5) could readily transfer a glucosyl moiety to the C-20*S*-OH of PPD *in vitro* and thus convert PPD into CK as monitored by TLC (Supplementary information, Figure S6A), HPLC (Figure 1C) and HPLC/ESIMS (Figure 1B) and ^{1}H NMR determination further conformed the

CK structure (Supplementary information, Figure S4C). We also found that incubating UGTPg1 and UDP-glucose with DM yielded another product (Supplementary information, Figure S6B), which was identified to be a new compound, namely 20*S*-*O*-β-(D-glucosyl)-dammarenediol II (DMG), by HPLC (Figure 1D) and HPLC/ESIMS (Figure 1B) as well as 1D and 2D NMR confirmation (Supplementary information, Figure S7). In particular, its glucosyl moiety was assigned to the C-20 by the key correlations from H-1′ and H-21 to C-20 in the 2D NMR spectrum (Supplementary information, Figure S7C). DMG was also detected in the *n*-butanol extract of the four engineered yeast strains together with squalene, DM, PPD and CK (Supplementary information, Figure S8 and Table S4). These results indicate that UGTPg1 also catalyzes the glucosylation of C-20*S*-OH of DM, both *in vitro* and *in vivo*. Similar incubations with a C-20*S* PPD-type ginsenoside, Rh2, yielded ginsenoside F2 (Supplementary information, Figure S6A). UGTPg1 also converts another PPD-type ginsenoside, Rg3, into ginsenoside Rd (Supplementary information, Figure S6A). When PPD, Rh2 and Rg3 were mixed with UGTPg1, CK, F2 and Rd were all detected after 15 mins incubation. However, UGTPg1 prefers Rg3 and Rh2 over PPD as its substrate (Supplementary information, Figure S9).

Although UGTPg1 readily transfers a glucosyl moiety to the free C-20*S*-OH of dammarane derivative substrates, it neither catalyzes the glucosylation of C-3-OH of CK, nor transfers a glucosyl moiety to further extend the sugar chain at C-20*S* (Supplementary information, Figure S6A). UGTPg1 does not glucosylate C-20*R* PPD, the epimer of C-20*S* PPD (Supplementary information, Figure S6C) or other triterpenoids with free hydroxyl

Figure 1 Functional characterization of UGTPg1 and its application in the biosynthesis of compound K from monosaccharide. **(A)** HPLC analysis of CK produced from the genetically engineered yeast AK1. Also analyzed are AK2, AK3 and the chassis BY4742, BA21 and AP1. **(B)** the MS spectrum of authentic CK sample, CK produced from the genetically engineered yeast AK1 and from the *in vitro* reaction of PPD catalyzed by UGTPg1, and DMG produced from the *in vitro* reaction of DM catalyzed by UGTPg1, respectively. **(C)** HPLC analysis of *in vitro* products from the incubation of UGTPg1 with PPD and UDP-glucose. Red, the crude UGTPg1 from cell extract of recombinant *E. coli* harboring pET28a-UGTPg1; blue, negative control, cell extract of recombinant *E. coli* harboring pET28a; black, authentic chemical samples of CK, Rh2 and PPD. **(D)** HPLC analysis showing the *in vitro* production of DMG by incubating UGTPg1 with DM. **(E)** HPLC analysis of products collected in time course of the *in vitro* reactions of PPD and DM catalyzed by UGTPg1. Black, the authentic samples; pink and blue, incubating UGTPg1 with PPD (0.5 mM) and DM (0.5 mM) for 24 h, respectively; other colors, incubating UGTPg1 with the mixed PPD (0.5 mM) and DM (0.5 mM) for 0 h, 0.25 h, 0.5 h, 1 h, 2 h and 24 h. **(F)** HPLC analysis of products collected in time course of the *in vitro* reactions of DM and DMG catalyzed by CYP716A47. Black, the authentic samples; purple and blue, incubating the microsomes from the yeast strain AC1 with DM (0.5 mM) and DMG (0.5 mM) for 36 h, respectively; other colors, incubating the microsomes with the mixed DM (0.5 mM) and DMG (0.5 mM) for 0 h, 0.5 h, 1 h, 2 h and 36 h. **(G)** a schematic presentation of *in vivo* CK production via two potential pathways. Blue arrows represent the biochemical pathway that already existed in *S. cerevisiae*; brown, green and purple arrows represent the heterologous pathways from plants, mainly *P. ginseng*, incorporated into *S. cerevisiae*. Solid line, one-step reaction; dashed line, more than one-step reactions. CYP716A47: a *P. ginseng* cytochrome P450; ATR2-1: an NADPH-cytochrome P450 reductase from *Arabidopsis thaliana*; UGTPg1: a novel *P. ginseng* UDP-glycosyltransferase identified in this study.

groups at C-3 and/or C-28, such as oleanane (pentacyclic triterpene) (Supplementary information, Figure S6D). These findings indicate that UGTPg1 specifically glucosylates the C-20S-OH of dammarane-type triterpenoids, and is a regioselective and stereospecific glycosyltransferase. The V_{max}, K_m, k_{cat} and k_{cat}/K_m for UGTPg1 to catalyze the glucosylation of PPD and DM (Supplementary information, Table S5) indicate that UGTPg1 prefers DM over PPD as its substrate.

The UGTPg1-catalyzed *in vitro* conversion of PPD to CK (Figure 1B and 1C) confirms the proposed pathway for CK production (Supplementary information, Figure S1B). However, the unexpected detection of DMG, in all four CK-producing strains (Supplementary information, Figure S8 and Table S4) and the *in vitro* catalytic activity of UGTPg1 to selectively transfer a glucosyl moiety to the C-20S-OH of DM with higher affinity and turn-over number compared with those of PPD (Figure 1E and Supplementary information, Table S5), suggest a possible alternative pathway for CK biosynthesis. Because it was confirmed that the extracted microsomes from a yeast strain harboring CYP716A47 oxidized DMG to produce CK *in vitro* (Figure 1F), CK could be produced from the glucosylation of DM, followed by the oxidation of DMG by CYP716A47. Therefore two parallel biosynthetic pathways for CK production may co-exist in our chassis yeasts (Figure 1G).

The above findings are intriguing because the transcripts encoding CYP716A47 and UGTPg1 are present in *P. ginseng* tissues (Supplementary information, Figure S10), so are those encoding all the enzymes of the MEV-driven terpenoid biosynthesis pathway [10]. Besides, two newly cloned NADPH-cytochrome P450 reductases from *P. ginseng* did contribute to PPD production in recombinant yeasts in place of *ATR2-1* (Supplementary information, Figure S11). It is thus possible that the two synthetic pathways for CK production demonstrated in this study may actually occur in *Panax* plants despite the fact that CK has never been detected in *Panax* plants. A potential scenario might be that UGTPg1 prefers Rg3 and Rh2 over PPD as its substrate (Supplementary information, Figure S9) and/or that CK may be transformed into other ginsenosides via extensive glycosylations and thus little CK is produced and/or accumulated in *P. ginseng*.

In summary, the identification of UGTPg1, the first characterized UGT for glucosylation of tetracyclic triterpenoid substrates from plants, was the key to the success in biosynthesis of CK in a one-pot reaction from simple sugars. Our study provides not only a cheap CK-manufacturing method for its potential clinical applications but also adds to our understanding of the biosynthetic pathways of ginsenosides within the *Panax* plants.

Acknowledgements

Xing Yan[1, *], Yun Fan[1, 3, *], Wei Wei[1, *], Pingping Wang[1, *], Qunfang Liu[2], Yongjun Wei[1], Lei Zhang[1], Guoping Zhao[1, 3, 4, 5], Jianmin Yue[2], Zhihua Zhou[1]

[1]*CAS-Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China;* [2]*State Key Laboratory of Drug Research, Institute of Materia Medica, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, 201203, China;* [3]*State Key Laboratory of Genetic Engineering, Department of Microbiology, School of Life Sciences and Institute of Biomedical Sciences, Fudan University, Shanghai 200032, China;* [4]*Shanghai-MOST Key Laboratory for Health and Disease Genomics, Chinese National Human Genome Center, Shanghai 201203, China;* [5]*Department of Microbiology and Li KaShing Institute of Health Sciences, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong SAR, China*
*These four authors contributed equally to this work.
Correspondence: Zhihua Zhou[a], Jianmin Yue[b], Guoping Zhao[c]
[a]Tel: 86-21-54924050, Fax: 86-21-54924049
E-mail: zhouzhihua@sippe.ac.cn
[b]Tel/Fax: 86-21-50806718
E-mail: jmyue@mail.shcnc.ac.cn
[c]Tel/Fax: 86-21-54924002
E-mail: gpzhao@sibs.ac.cn

References

1   Yun T-K. *Lancet Oncol* 2001; **2**:49-55.
2   Hasegawa H. *J Pharmacol Sci* 2004; **95**:153-157.
3   Zhou W, Yan Q, Li JY, *et al. J Appl Microbiol* 2008; **104**:699-706.
4   Sivakumar G, Yu K, Paek K. *Eng Life Sci* 2005; **5**:527-533.
5   Dürr C, Hoffmeister D, Wohlert SE, *et al. Angew Chem Int Ed Engl* 2004; **43**:2962-2965.
6   Paddon C, Westfall P, Pitera D, *et al. Nature* 2013; **496**:528-532.
7   Han JY, Kim HJ, Kwon YS, *et al. Plant Cell Physiol* 2011; **52**:2062-2073.
8   Tansakul P, Shibuya M, Kushiro T, *et al. FEBS Lett* 2006; **580**:5143-5149.
9   Dai Z, Liu Y, Zhang X, *et al. Metabolic Eng* 2013; **20**:146-156
10 Li C, Zhu Y, Guo X, *et al. BMC Genomics* 2013; **14**:245.

(**Supplementary information** is linked to the online version of the paper on the *Cell Research* website.)