




# Characterization and protein engineering of glycosyltransferases for the biosynthesis of diverse hepatoprotective cycloartane-type saponins in *Astragalus membranaceus*

Kuan Chen<sup>1,2,†</sup> , Meng Zhang<sup>1,†</sup>, Baihan Gao<sup>1,†</sup>, Aobulikasimu Hasan<sup>1</sup>, Junhao Li<sup>3</sup>, Yang'oujie Bao<sup>1</sup>, Jingjing Fan<sup>1</sup>, Rong Yu<sup>1</sup>, Yang Yi<sup>1</sup>, Hans Ågren<sup>3</sup>, Zilong Wang<sup>1</sup>, Haiyang Liu<sup>4</sup>, Min Ye<sup>1,\*</sup>  and Xue Qiao<sup>1,\*</sup> 

<sup>1</sup>State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, Beijing, China

<sup>2</sup>Beijing Institute of Clinical Pharmacy, Beijing Friendship Hospital, Capital Medical University, Beijing, China

<sup>3</sup>Department of Physics and Astronomy, Uppsala University, Uppsala, Sweden

<sup>4</sup>State Key Laboratory of Phytochemistry and Plant Resources in West China and Yunnan Key Laboratory of Natural Medicinal Chemistry, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China

Received 9 July 2022;

revised 15 November 2022;

accepted 4 December 2022.

\*Correspondence (Tel +86 10 82801516;

fax +86 10 82802024; email

qiaoxue@bjmu.edu.cn (X.Q.) and Tel +86 10

82801516; fax +86 10 82802024; email

yemin@bjmu.edu.cn (M.Y.)

<sup>†</sup>These authors contributed equally to this work.

## Summary

Although plant secondary metabolites are important source of new drugs, obtaining these compounds is challenging due to their high structural diversity and low abundance. The roots of *Astragalus membranaceus* are a popular herbal medicine worldwide. It contains a series of cycloartane-type saponins (astragalosides) as hepatoprotective and antiviral components. However, astragalosides exhibit complex sugar substitution patterns which hindered their purification and bioactivity investigation. In this work, glycosyltransferases (GT) from *A. membranaceus* were studied to synthesize structurally diverse astragalosides. Three new GTs, AmGT1/5 and AmGT9, were characterized as 3-O-glycosyltransferase and 25-O-glycosyltransferase of cycloastragenol respectively. AmGT1<sub>G146V/I</sub> variants were obtained as specific 3-O-xylosyltransferases by sequence alignment, molecular modelling and site-directed mutagenesis. A combinatorial synthesis system was established using AmGT1/5/9, AmGT1<sub>G146V/S</sub> and the reported AmGT8 and AmGT8<sub>A394F</sub>. The system allowed the synthesis of 13 astragalosides in *Astragalus* root with conversion rates from 22.6% to 98.7%, covering most of the sugar-substitution patterns for astragalosides. In addition, AmGT1 exhibited remarkable sugar donor promiscuity to use 10 different donors, and was used to synthesize three novel astragalosides and ginsenosides. Glycosylation remarkably improved the hepatoprotective and SARS-CoV-2 inhibition activities for triterpenoids. This is one of the first attempts to produce a series of herbal constituents *via* combinatorial synthesis. The results provided new biocatalytic tools for saponin biosynthesis.

**Keywords:** astragalosides, biosynthesis, glycosyltransferases, protein engineering, *Astragalus membranaceus*.

## Introduction

Plant-derived natural products are important source for therapeutic agents (Najmi *et al.*, 2022; Patridge *et al.*, 2016). As secondary metabolites, compounds in one plant usually exhibit high structural diversity with similar backbones and various substitution groups. However, obtaining these homologues are usually challenging due to their similar physicochemical properties and relatively low abundance. Usually, kilograms of plant materials are used to afford milligrams of compounds (Corsello and Garg, 2015; Guerra-Bubb *et al.*, 2012), which was laborious and time-consuming. As an alternative method, biosynthesis is green and efficient to produce structurally diverse natural products (Corsello and Garg, 2015; Guerra-Bubb *et al.*, 2012).

Astragalosides are known as effective compounds of the herbal medicine *Astragalus* root (roots of *Astragalus membranaceus*; Ionkova *et al.*, 2014; Su *et al.*, 2021). They exhibit remarkable bioactivities in hepatoprotection, anti-virus, cardiovascular protection and immunomodulatory (Chu *et al.*, 2010; Li *et al.*, 2018;

Liang *et al.*, 2020). Astragalosides possess a cycloartane-type triterpenoid skeleton named cycloastragenol (CA, **1**), which was modified by glucosylation at C3/C6/C25-OH and/or xylosylation at C3-OH (Su *et al.*, 2021). At least 22 astragalosides have been reported from *Astragalus* root, containing 12 different glycosylation patterns (Table S1). They suffer from low abundance and poor UV (ultraviolet) absorption, which challenged their purification, and hindered further study on their biological activities. Therefore, it is necessary to employ biosynthetic approaches to prepare astragalosides.

Though we have identified the scaffold-forming enzyme for astragalosides (Chen *et al.*, 2022), the downstream steps remain unclear. Glycosylation, usually catalysed by uridine diphosphate (UDP)-sugar-dependent glycosyltransferases (GT), plays an important role to synthesize astragalosides (Kurze *et al.*, 2021; Rahimi *et al.*, 2019). Up to now, at least 67 plant triterpenoid glycosyltransferases have been reported, most of which catalyse the glycosylation of oleanane-type and dammarane-type triterpenoids (Table S2). For cycloartane-type triterpenoids, only

AmGT8 and its mutants reported by our group could catalyse their glycosylation (Zhang *et al.*, 2022). Other GTs modifying the cycloartane backbones remain unknown. Meanwhile, a number of astragalosides contain xylosyl groups, while specific *O*-xylosyltransferase has rarely been reported, except for Pn3-32-i5 from *Panax notoginseng* (Wang *et al.*, 2020), GmUGT73F4 from *Glycine max* (Sayama *et al.*, 2012) and PgUGT94Q13 from *Panax ginseng* (Li *et al.*, 2021). The latter two enzymes could still use UDP-glucose (UDP-Glc) with a lower activity. Due to the large size of the substrate binding pocket, it is difficult to modify the function of triterpenoid GTs. Usually the desired function was obtained by screening thousands of mutants (Li *et al.*, 2020; Wei *et al.*, 2015; Zhang *et al.*, 2020a, 2022). Thus, the discovery and engineering of biosynthetic catalysts for astragalosides remain challenging.

In this work, we used transcriptome mining and protein engineering to obtain a series of new GTs modifying cycloartane-type triterpenoids. AmGT5, AmGT1<sub>G146V/I</sub> and AmGT9 were characterized as 3-*O*-glucosyltransferase, 3-*O*-xylosyltransferases and 25-*O*-glucosyltransferase of cycloastragenol respectively. These enzymes were used in combination with the 2'-*O*-glucosyltransferase AmGT8 and 6-*O*-glucosyltransferase AmGT8<sub>A394F</sub> to produce 13 astragalosides in *Astragalus root* with the conversion rates ranged from 22.6% to 98.7%. They covered most of the sugar-substitution patterns for astragalosides. This is one of the first attempts to produce a whole series of herbal constituents *via* combinatorial synthesis. The products were tested for inhibition activities against SARS-CoV-2 proteins and hepatoprotective activities against acetaminophen-induced cell death, where several bioactive astragalosides were discovered.

## Results and discussion

### Chemical analysis of astragalosides in *A. membranaceus*

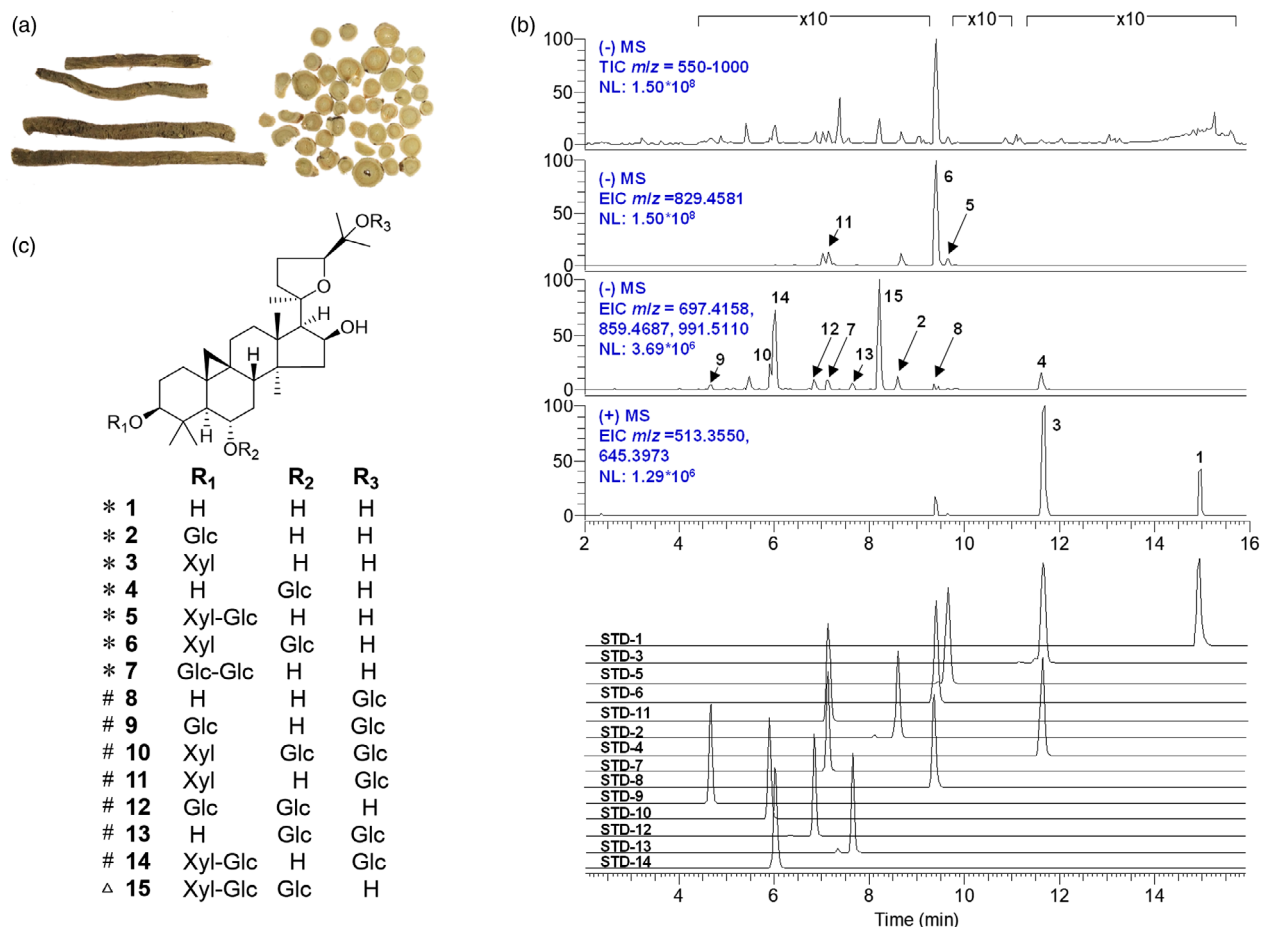
The roots of *Astragalus membranaceus* (*Astragalus root*) contain a series of cycloartane-type saponins. Among them, 22 out of 48 possess a cycloastragenol skeleton and are named as astragalosides (Su *et al.*, 2021). Astragalosides are the major bioactive components of *Astragalus root* (lonkova *et al.*, 2014). According to the substitution of sugar moieties, they could be divided into 12 groups (Table S1), with glucosylation at C3/C6/C25-OH and/or xylosylation at C3-OH. For example, astragaloside IV is one of the most abundant saponins in *Astragalus root* which contain a 3-*O*-xyloside and a 6-*O*-glucoside (3X,6G). In order to elucidate the structures of various astragalosides, *Astragalus roots* were analysed using ultra-high performance liquid chromatography coupled with mass spectrometry (UHPLC/MS). The samples were treated with ammonia hydrolysis (13% NH<sub>3</sub>-H<sub>2</sub>O, 65 °C, 1 h) to degrade the downstream acylation products before analysis. In total, 14 astragalosides (2–15) were unambiguously characterized by comparing the retention time and tandem mass spectrometry (MS/MS) spectra with reference standards except for 15 (Figure 1, Figure S1–S6). They covered all 12 substitution patterns reported in *Astragalus root*. They also contained two extra patterns, 3GG (7) and 25G (8), which have been reported from other *Astragalus* plants (Barbic *et al.*, 2010; Gulcernal *et al.*, 2012). These results indicated that *Astragalus root* contains at least 14 astragalosides with different glycosylation patterns, which covered almost all natural cycloastragenol-derived saponins identified so far. These astragalosides should be synthesized by different glucosyltransferases in *A. membranaceus*.

### Candidate gene screening and functional characterization

To explore the GTs synthesizing astragalosides, candidate genes were identified from *A. membranaceus* transcriptome by BLASTn (basic local alignment search tool for nucleic acid,  $e < 10^{-50}$  and ORF length > 1200 bp), using plant triterpenoid GTs (Table S2) as probes. Three GT genes were found and named as AmGT1, AmGT5 and AmGT9 (Genbank Accession Nos. ON075039, ON075040 and ON075041), together with the reported AmGT8 (Zhang *et al.*, 2022). Phylogenetic analysis (Figure 2a) showed that AmGT1 and AmGT5 grouped together with 3-*O*-glucosyltransferases from *Glycyrrhiza uralensis* (Chen *et al.*, 2019; Zhang *et al.*, 2020a), indicating their similar function. AmGT9 clustered with 3-*O*-glucosyltransferases from *Barbarea vulgaris* (Erthmann *et al.*, 2018; Rahimi *et al.*, 2019) and 28-*O*-glucosyltransferases from *Centella asiatica* (Costa *et al.*, 2017; Kim *et al.*, 2017), which was away from legume triterpenoid GTs, suggesting a distinct function. In addition, AmGT8, reported as a 3/6/2'-*O*-glucosyltransferase from *A. membranaceus*, was grouped with 2'-*O*-glucosyltransferases from *G. uralensis* (Nomura *et al.*, 2019) and *G. max* (Takagi *et al.*, 2018) as expected. These results suggested the diversity of GTs in *A. membranaceus* which might contribute to the diversity of astragalosides. According to the results of real-time quantitative polymerase chain reaction (qRT-PCR) analysis, AmGT8 showed high expression mainly in roots, where astragalosides 2–15 were detected, indicating its potential role in the biosynthesis of these compounds. AmGT1 and AmGT9 expressed in roots, stems and leaves, while the relative expression level of AmGT5 was low (Figure S7). AmGT1/5/9 were cloned from *A. membranaceus* and expressed in *Escherichia coli* using pET28a(+) vectors. The proteins were then purified and used for functional characterization *in vitro* (Figure S8).

The biochemical characteristics of AmGT1/5/9 proteins (5 μg) were studied using CA (0.1 mM) as the substrate, and UDP-Glc (0.5 mM) as the sugar donor. AmGT1/5/9 showed their maximum activity at pH 8.0 (50 mM Na<sub>2</sub>HPO<sub>4</sub>-NaH<sub>2</sub>PO<sub>4</sub>) and 37 °C, and were independent of divalent cations (Figure S9). AmGT1 could catalyse CA to produce compounds 2 (conversion rate 69.2%) and 9 (30.8%) (Figure 2c). According to MS/MS analysis, 2 and 9 were identified as mono-*O*-glucoside ([M-H + HCOOH]<sup>-</sup> *m/z* 697.4125, [M + Na]<sup>+</sup> *m/z* 675.4107) and di-*O*-glucoside ([M-H + HCOOH]<sup>-</sup> *m/z* 859.4646, [M-H-Glc]<sup>-</sup> *m/z* 651.4080, [M-H-2Glc]<sup>-</sup> *m/z* 489.3548) of CA respectively (Figures S2, S4). Compound 2 was characterized as CA-3-*O*-glucoside by comparing with a reference standard (Figure S1). Compound 9 was prepared from a scaled-up reaction and characterized by nuclear magnetic resonance (NMR) spectroscopy. The glycosylation sites were characterized at C-3 and C-25 according to the heteronuclear multiple bond correlation (HMBC) of C-3/H-1' ( $\delta_C$  89.4/ $\delta_H$  5.02) and C-25/H-1'' ( $\delta_C$  79.0/ $\delta_H$  5.08) (Appendix S1). Similarly, the products of AmGT5 and AmGT9 were CA-3-*O*-glucoside (2, 65.6%) and CA-25-*O*-glucoside (8, 84.1%) respectively (Figure 2c). These results suggested that AmGT1/5 and AmGT9 are 3-*O*-glucosyltransferase and 25-*O*-glucosyltransferase of cycloastragenol respectively. They are the only GTs modifying cycloartane-type triterpenoids except for AmGT8, and AmGT9 is the first GT modifying the side chain of cycloartane-type triterpenoids.

Since several astragalosides contain *O*-xylosyl groups at C-3, we then studied the function of AmGTs using alternative sugar



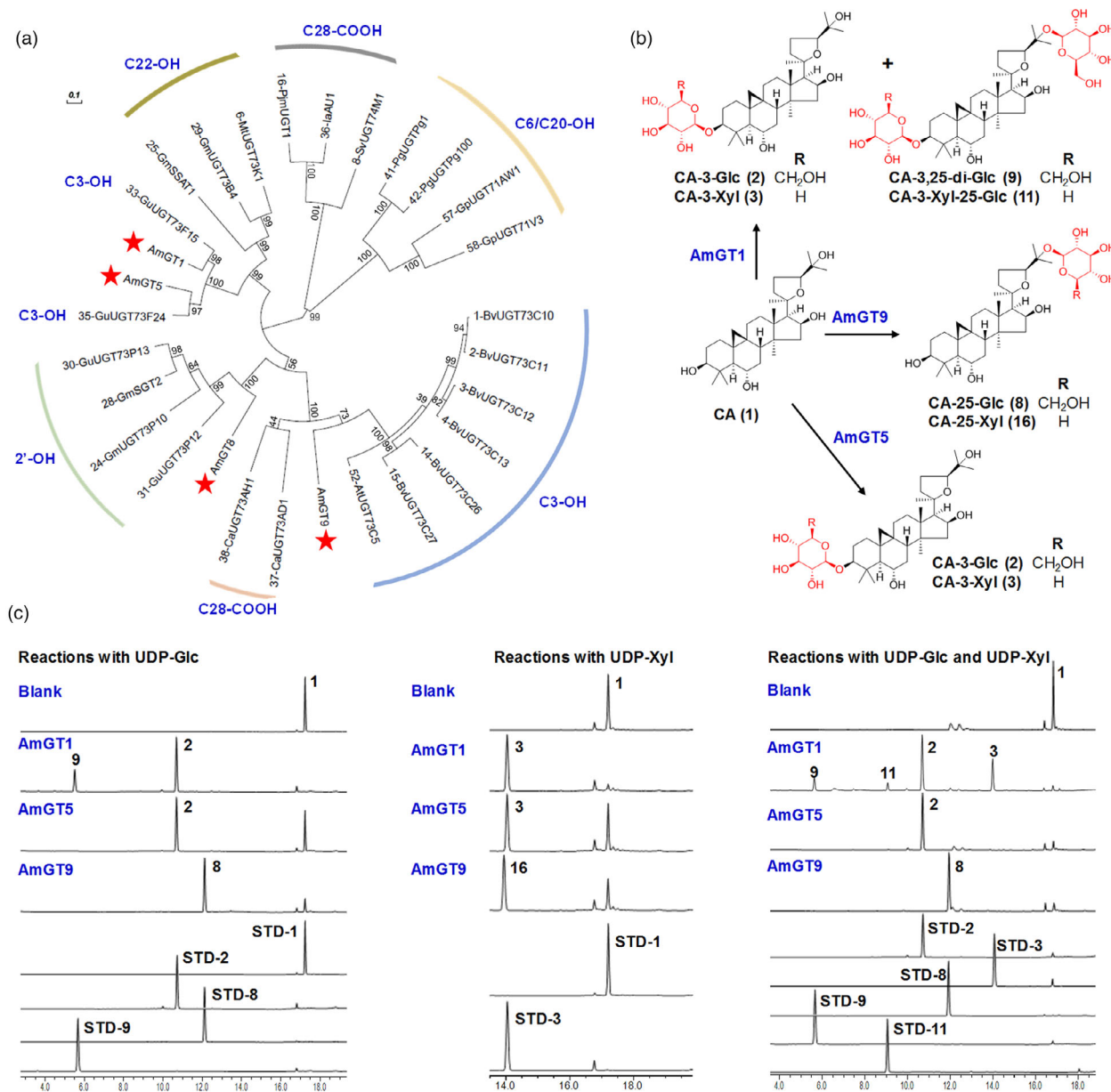
**Figure 1** Chemical analysis of astragalosides in *Astragalus* root. (a) Raw materials and the decoction pieces of *Astragalus* root. (b) LC/MS analysis of *Astragalus* root extracts after ammonia hydrolysis. TIC, total ion chromatogram; EIC, extracted ion chromatogram. In EIC,  $m/z$  513.3550 ( $[(CA + Na)^+]$ ),  $m/z$  645.3973 ( $[(CA + Xyl + Na)^+]$ ),  $m/z$  697.4158 ( $[(CA + Glc + HCOO)^-]$ ),  $m/z$  829.4581 ( $[(CA + Glc + Xyl + HCOO)^-]$ ),  $m/z$  859.4687 ( $[(CA + 2Glc + HCOO)^-]$ ) and  $m/z$  991.5110 ( $[(CA + 2Glc + Xyl + HCOO)^-]$ ) were extracted. (c) Structures of compounds **1–15**. \* (**1–7**) were identified by comparing with reference standards (Figure S1–S6); # (**8–14**) were purified and identified by NMR (Appendix S1); Δ (**15**) was identified by HR-MS/MS (Figure S6). Glc, glucosyl group. Xyl, xylosyl group.

donors. When UDP-xylose (UDP-Xyl) was used, the product of AmGT1/5 was CA-3-*O*-xyloside (**3**, 93.3% for AmGT1 and 63.4% for AmGT5), while the product of AmGT9 was tentatively identified as CA-25-*O*-xyloside (**16**, 71.1%) (Figure S6). When a mixed sugar donor (UDP-Glc:UDP-Xyl:CA = 5:5:1) was used, AmGT1 could catalyse the formation of mono-*O*-glucoside **2**, mono-*O*-xyloside **3**, di-*O*-glycosides **9** (CA-3-*O*-glucoside-25-*O*-glucoside) and **11** (CA-3-*O*-xyloside-25-*O*-glucoside) at the same time, while AmGT5/9 could not produce any xyloside (Figure 2c). Compounds **2** and **3** were identified by comparison with reference standards (Figure S2), while **8/9/11** were identified by NMR spectroscopic analysis (Appendix S1). These results suggested that AmGT5 is a specific 3-*O*-glucosyltransferase, while AmGT1 could use both UDP-Xyl and UDP-Glc as the sugar donor with a similar selectivity, which is rare for triterpenoid GTs (Table S2).

### Protein engineering of AmGT1 using site-directed mutagenesis

Efficient production of 3-*O*-xylosyl astragalosides *in vitro* requires highly specific 3-*O*-xylosyltransferase. AmGT1 catalysed CA to

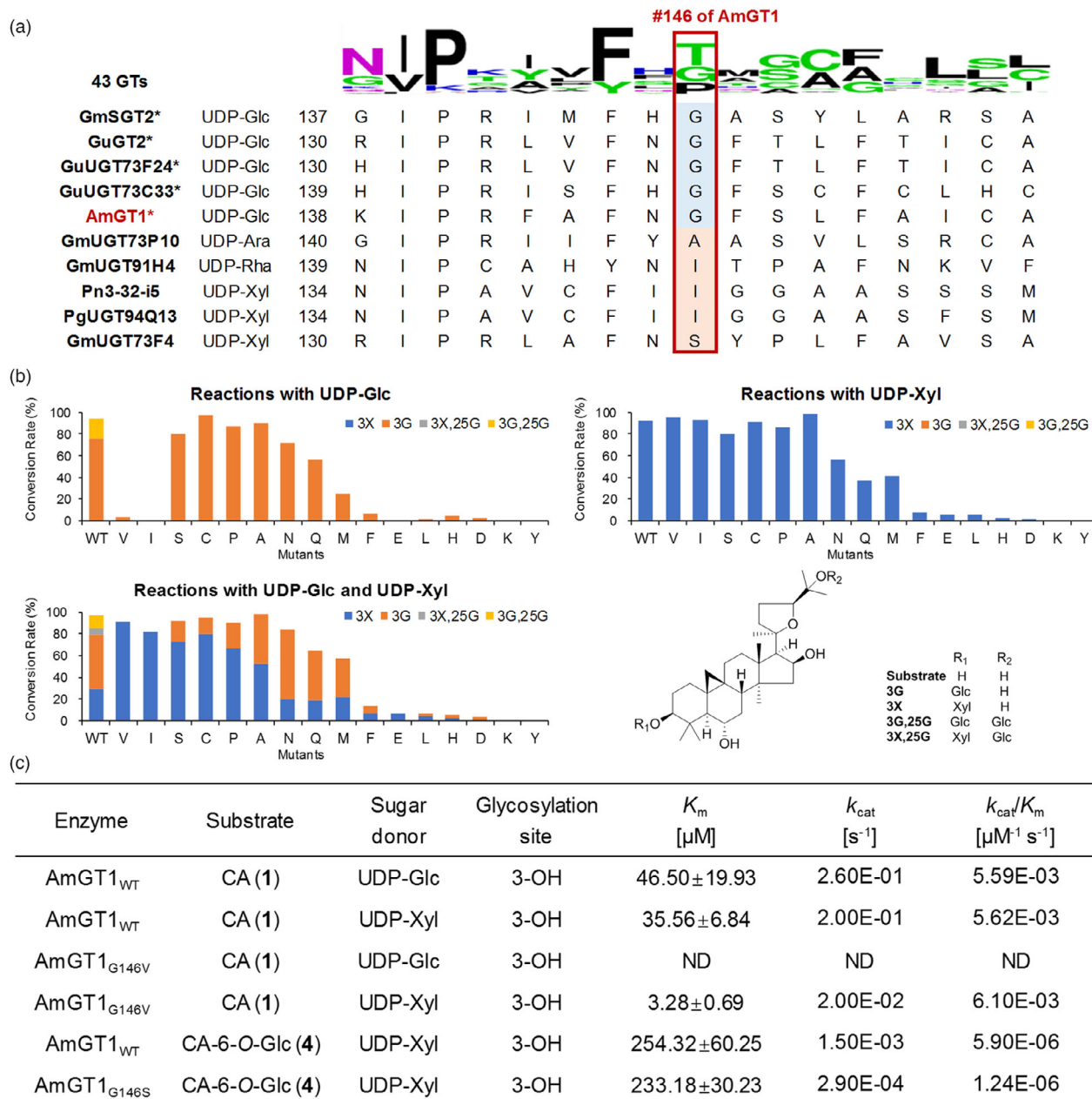
produce 3-*O*-xyloside (conversion rate ~30%) and 3-*O*-glucoside (~50%) using a mixed sugar donor (Figure 2c), which is not satisfying for combinatorial biosynthesis. Thus, site-directed mutagenesis was carried out to improve the sugar donor mutagenesis was carried out to improve the sugar donor specificity for AmGT1. First, we aligned the protein sequence of 53 triterpene GTs to discover the key residues. Among them, 43 GTs are specific to UDP-Glc, 5 GTs could tolerate other sugar donors besides UDP-Glc, and 5 GTs are specific to donors other than UDP-Glc (Figure S10). We found one residue, #146, which is conserved in GTs using UDP-Glc as T/P/G, and distinct in GTs utilizing other donors (A/I/S) (Figure 3a). To predict the function of #146, molecular docking of AmGT1 with UDP-Glc was conducted, using an AmGT1 structure generated by Alphafold2 (Tunyasyunakool *et al.*, 2021). The result showed that G146 was located close to the C6-OH of UDP-Glc (Figure 4a). Interestingly, the residue corresponding to G146 in AmGT1 is T145 in GgCGT, which has been reported to form a hydrogen bond with the C6-OH of the glucosyl group (Zhang *et al.*, 2020b). Since the difference between UDP-Xyl and UDP-Glc lies in C6-OH, it was speculated that #146 may affect the selectivity of UDP-Xyl/UDP-Glc for AmGT1.



**Figure 2** Phylogenetic analysis and functional characterization of AmGT1, AmGT5 and AmGT9. (a) Phylogenetic analysis of four AmGTs with 25 previously characterized plant triterpenoid GTs (Table S2). The tree was established by MEGA6 using the Maximum Likelihood method (bootstrap = 1000) with default parameters. (b) Functions of AmGTs with cycloastragenol (CA, **1**) as the substrate and UDP-Glc and/or UDP-Xyl as the sugar donor. (c) UHPLC/CAD chromatograms of the reaction mixtures. Products were characterized by comparing with reference standards or NMR analysis except for **16**.

To explore its function, G146 was mutated into 16 other residues with different properties in AmGT1. Mutants G146V and G146I lost their activity to UDP-Glc, but retained high conversion rate using UDP-Xyl (Figure 3b, Figure S11). These variants became specific xylosyltransferases which are rare in nature (Table S2). In kinetic analysis, AmGT1<sub>G146V</sub> showed higher affinity ( $K_m = 3.28 \mu\text{M}$ ) with UDP-Xyl than wild-type AmGT1 ( $35.56 \mu\text{M}$ ) when CA was used as the substrate (Figure 3c, Figure S12). More importantly, AmGT1<sub>WT</sub> used UDP-Xyl and UDP-Glc with a similar selectivity ( $k_{\text{cat}}/K_m = 5.62\text{E-}03$  vs.  $5.59\text{E-}03$ ), while AmGT1<sub>G146V</sub> could only use UDP-Xyl ( $k_{\text{cat}}/K_m = 6.10\text{E-}03$  vs. undetectable for UDP-Glc).

AmGT1<sub>G146S/C/P/A</sub> showed high conversion rates using both UDP-Glc and UDP-Xyl, but UDP-Xyl was preferred when a mixture (UDP-Glc:UDP-Xyl = 1:1) was used (Figure 3b). For example, G146S preferentially utilized UDP-Xyl (conversion rate 72.2%) and retained weak UDP-Glc activity (19.8%). The ratio of xylosides for G146C/P/A was 79.9%/66.3%/52.4%, respectively, which was higher than the wild-type (35.3%). The activity of G146N/Q/M decreased, while the selectivity to sugar donors was unchanged. To sum up, residues with shorter side chain (V/I/S/C/P/A) enhanced the specificity to UDP-Xyl and reduced the activity to UDP-Glc, while the longer side chain (F/E/L/H/D/K/Y) reduced the activity to both donors. In addition, it is unexpected that

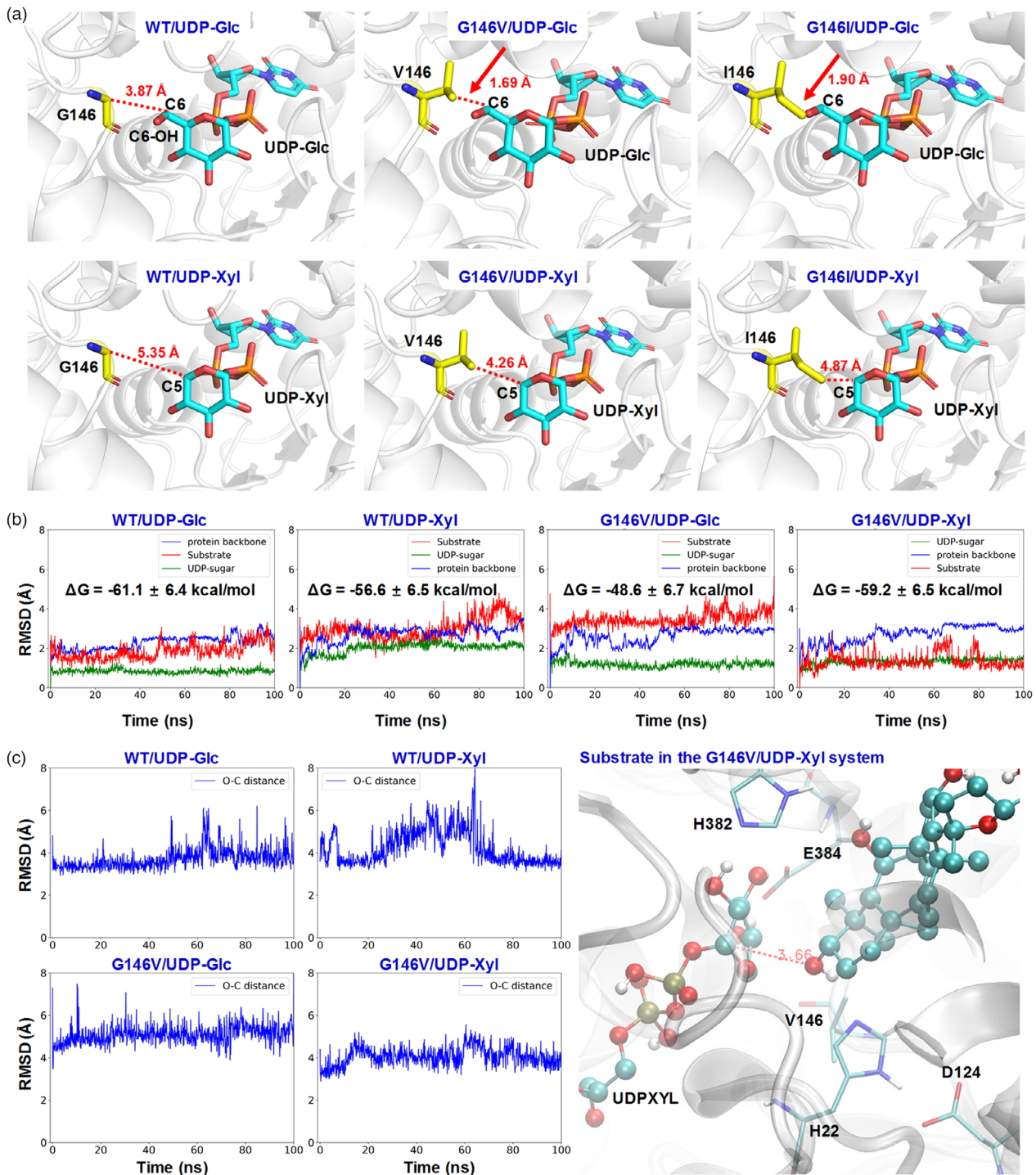


**Figure 3** Protein engineering of AmGT1 using site-directed mutagenesis. (a) Protein sequence alignment of 53 triterpene GTs. GTs without \* were specific to UDP-Glc; GTs with \* could tolerate other sugar donors besides UDP-Glc (Figure S10). (b) Functional analysis of AmGT1 variants with CA as the substrate, while UDP-Glc and/or UDP-Xyl as the sugar donor. WT, wild type. 3G, CA-3-O-glucoside. 3X, CA-3-O-xyloside. 3X,25G, CA-3-O-xyloside-25-O-glucoside. 3G,25G, CA-3,25-di-O-glucoside. (c) Kinetic parameters of AmGT1 and its mutants. ND, product under the limit of quantitation.

mutation of G146 significantly eliminate its glycosylation activity towards C25-OH (Figure S11). These results suggested that engineering of G146 could improve both UDP-Xyl selectivity and regio-selectivity of AmGT1.

To further explain the mechanism of function for residue #146, we simulated the interaction between AmGT1 and UDP-sugar donors. According to molecular docking, #146 is close to the C6 of UDP-Glc and C5 of UDP-Xyl (Figure 4a). AmGT1 and its variants G146V/I could all use UDP-Xyl since the reaction pocket could well accommodate the sugar donor, with a distance of 5.35 Å, 4.26 Å and 4.87 Å between C5 and G/V/I respectively.

For UDP-Glc, AmGT1<sub>WT</sub> could accommodate its C6-OH due to the smaller size of Gly, with a distance of 3.87 Å. When Gly was mutated to Val or Ile, steric bulk from long side chains (distance 1.69 Å/1.90 Å between C6 and V/I) hindered the appropriate binding of UDP-Glc. In addition, G146K/Y completely lost their activities towards UDP-Xyl due to their even longer side chains. The 100 ns molecular dynamics simulations showed that UDP-Glc is more stable than UDP-Xyl in the wild-type AmGT1, while the substrate CA is more stable in AmGT1<sub>G146V</sub> with UDP-Xyl than that with UDP-Glc (Figure 4b). We also calculated the binding-free energies of CA with the enzymes using the molecular



**Figure 4** Molecular docking and molecular dynamics to explain the mechanism of function for #146. (a) Molecular docking of AmGT1, AmGT1<sub>G146V</sub> and AmGT1<sub>G146I</sub> with UDP-Glc or UDP-Xyl. (b) Time evolution of root mean square deviation (RMSD) of the protein backbone, UDP-sugar donor and substrate in molecular dynamics simulations, coloured in blue, green and red lines respectively. The MM/GBSA binding free energies of substrate (CA) with AmGT1/AmGT1<sub>G146V</sub> and UDP-Glc/UDP-Xyl were shown in each subgraph. (c) Time evolution of the distance between C3-OH of CA and C1 of UDP-Xyl/Glc in AmGT1 and AmGT1<sub>G146V</sub>. The distance was illustrated in the right panel.

mechanics generalized Born surface area (MM/GBSA) method (Kollman *et al.*, 2000). The binding free energies of CA with AmGT1<sub>WT</sub> were  $-61.1 \pm 6.4$  and  $-56.6 \pm 6.5$  kcal/mol, in the presence of UDP-Glc or UDP-Xyl respectively. For AmGT1<sub>G146V</sub>, the values were  $-48.6 \pm 6.7$  or  $-59.2 \pm 6.5$  kcal/mol when

UDP-Glc or UDP-Xyl was used (given in Figure 4b). These results also supported that AmGT1<sub>WT</sub>/AmGT1<sub>G146V</sub> favoured for UDP-Glc/UDP-Xyl. In addition, the C3-OH is closer to UDP-Xyl ( $\sim 4$  Å) than UDP-Glc ( $\sim 5$  Å) in AmGT1<sub>G146V</sub>, leading to a larger probability for the xylosylation reaction (Figure 4c). These results

indicated that steric bulk and sugar acceptor/donor binding state both contribute to the altered sugar donor specificity for #146 variants.

### Combinatorial biosynthesis of astragalosides using AmGTs

Several AmGTs and their mutants could catalyse the glycosylation of CA. For combinatorial synthesis, it is necessary to elucidate the order of glycosylation at different sites. Therefore, the function of AmGTs was investigated using various CA glycosides as substrates, including **2–6**, **8** and **13** (Figure 5a).

For compounds **2** (CA-3-O-glucoside) and **3** (CA-3-O-xyloside) with a C3-OH substitution, AmGT8 and 9 could catalyse the glucosylation at C2'-OH and C25-OH respectively. However, the C6-OH glucosylation activity of AmGT8<sub>A394F</sub> decreased significantly. For compound **4** (CA-6-O-glucoside), AmGT1/5 and AmGT9 could catalyse the glucosylation at C3-OH and C25-OH respectively. These results indicated that C6-O-glucosylation should occur before C3/C25-O-glycosylation. For compounds **5** (CA-3-O-xyloside-2'-O-glucoside) and **6** (CA-3-O-xyloside-6-O-glucoside), only AmGT9 showed high glycosylation activity at C25-OH. When compound **8** (CA-25-O-glucoside) was used, the C3-OH glucosylation activity retained for AmGT1/5, while the activities of AmGT8 and AmGT8<sub>A394F</sub> were reduced (Figure 5b). These results indicated that C2'/C6-O-glycosylation should occur before C25-O-glycosylation, while the order of C3/C25-O-glycosylation was interchangeable (Appendix S1).

Since that specific 3-O-xylosylation is important for astragalosides, AmGT1 and its variants were also tested using UDP-Xyl as the sugar donor (Figure 5c). For wild-type AmGT1, the substitution of C6-OH (**4** and **13**) affected its catalytic activity, while the substitution of C25-OH (**8**) did not. For AmGT1 variants, only G146S/P/A showed considerable xylosylation activity (69.3%–76.1%) to CA-6-O-glucoside (**4**), though the  $K_m$  value of AmGT<sub>G146S</sub> was only slightly lower than the wild-type (Figure 3c, Figure S12). These results suggested that different AmGT1 variants should be used for the xylosylation of different substrates: AmGT1<sub>WT/G146V/I</sub> for CA, and AmGT1<sub>G146S/P/A</sub> for CA-6-O-glucoside.

Based on the above results, we constructed the combinatorial synthesis network of astragalosides **2–14** by AmGTs (Figure 5d). Xylosylation of C3-OH could be catalysed by AmGT1<sub>G146V/I</sub> for compound **1**, AmGT1<sub>G146S/P/A</sub> for compound **4** and AmGT1 for compound **8**. Glucosylation of C3-OH could be catalysed by AmGT1/5, while C25-OH by AmGT9. Meanwhile, glucosylation at C2'-OH was achieved by AmGT8, while at C6-OH by AmGT8<sub>A394F</sub>. The order of glycosylation for most astragalosides should be C6-OH first, then C3-OH, C2'-OH and C25-OH. For astragalosides without C6/C2'-OH, glycosylation order at C3-OH and C25-OH could be switched.

Astragalus root contains at least 14 astragalosides (**2–15**) with different glycosylation patterns. Among them, **2**, **3**, **4** and **8** are

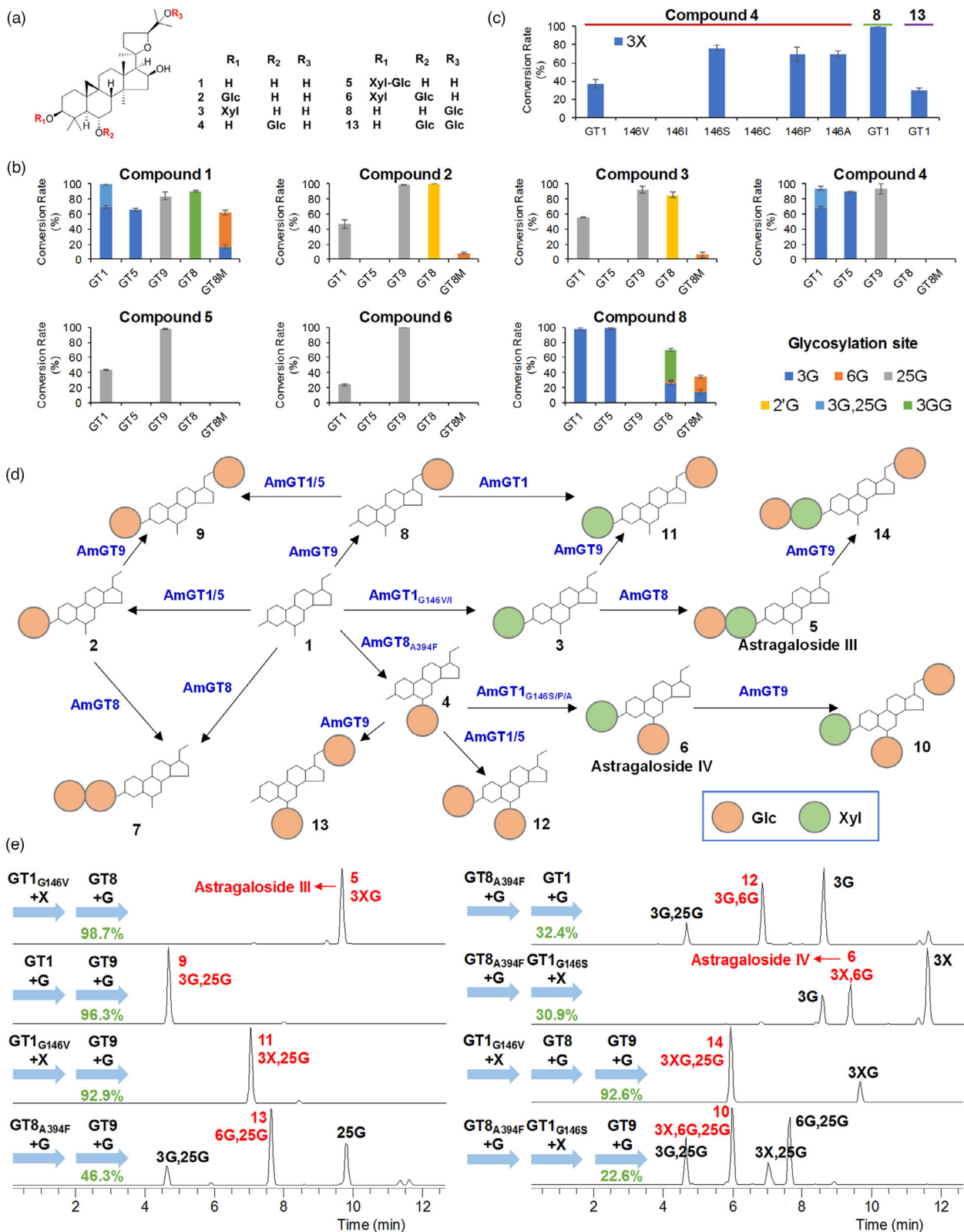
mono-glycosides which could be synthesized by AmGT1<sub>G146S</sub>, AmGT1<sub>G146V</sub>, AmGT8<sub>A394F</sub> and AmGT9 with a conversion rate up to 97.2%, 95.5%, 45% and 84.1% respectively. The di-glycoside **7** could be produced using AmGT8 with a conversion rate of >90% (Zhang et al., 2022). For other di-glycosides (**5**, **6**, **9**, **11–13**) and tri-glycosides (**10**, **14**), they could be synthesized using combinatorial biosynthesis (Figure 5e). According to the biochemical characteristics, AmGT1/5/9 favoured for Na<sub>2</sub>HPO<sub>4</sub>-NaH<sub>2</sub>PO<sub>4</sub> buffer at pH = 8, while AmGT1<sub>G146V/S</sub> favoured for Tris-HCl buffer at pH = 7/8 (Figure S9). Since that AmGT5 suffered from very poor activity in Tris-HCl, we used Na<sub>2</sub>HPO<sub>4</sub>-NaH<sub>2</sub>PO<sub>4</sub> as a compatible buffer. One-pot, step-by-step reactions were established as shown in Figure 5e. The reaction mixture was analysed by UHPLC/MS to confirm the product and by ultra-high-performance liquid chromatography coupled with charged aerosol detection (UHPLC/CAD) to calculate the conversion rate (Figure S13). As a result, conversion rates for astragalosides **5**, **6** and **9–14** were from 22.6% to 98.7%. For the reactions with conversion rates <50%, by-products (3G from AmGT8<sub>A394F</sub>, 25G from AmGT9) were detected which warrants further optimization. This is the first report to produce the whole series of natural products in an herb using combinatorial biosynthesis.

### Structural modification of triterpenoids by AmGTs

Besides astragalosides, AmGTs could also be used to expand the structural diversity of saponins by attaching various glycosyl groups. When CA (**1**) or CA-25-O-Glc (**8**) was used as the substrate, AmGT1 could utilize 10 sugar donors, including UDP-Glc, UDP-Xyl, UDP-glucosamine (UDP-GlcN), UDP-N-acetylglucosamine (UDP-GlcNAc), UDP-glucuronic acid (UDP-GlcA), UDP-galactose (UDP-Gal), UDP-galactosamine (UDP-GalN), UDP-arabinose (UDP-Ara), UDP-rhamnose (UDP-Rha) and thymidine diphosphate glucose (TDP-Glc) (Figure 6a). The conversion rates were above 85% except for UDP-GlcA and UDP-GalN. However, when CA-6-O-Glc (**4**) and CA-6,25-di-O-Glc (**13**) were used, the sugar donor promiscuity of AmGT1 decreased remarkably. CA-3-O-galactoside (**17**) and CA-3-O-acetylglucosamine (**18**) were purified from scaled-up reactions using CA and AmGT1, indicating the glycosylation occurred at C3-OH. Their structures were characterized by NMR as new compounds (Appendix S1). To our best knowledge, AmGT1 was the triterpene GT with highest sugar donor promiscuity (Table S2). In contrast, AmGT5 and AmGT9 could mainly utilize UDP-Glc and UDP-Xyl.

AmGTs could also be employed to synthesize ginsenosides, which are tetracyclic triterpenoid saponins similar to astragalosides, and are bioactive components from Ginseng roots (Zhao et al., 2020). Most of the ginsenosides contain two or more sugar moieties, while ginsenosides with only one sugar residue are rare. In this study, AmGTs were used to catalyse the glycosylation of protopanaxadiols (PPD, **19–20**) and protopanaxatriols (PPT, **21–22**) to obtain ginsenosides Rh1 and Rh2 (**23–26**, Figure S14).

**Figure 5** Substrate promiscuity of AmGTs and the combinatorial synthesis of astragalosides. (a) Structures of substrates. (b) Conversion rates for different substrates catalysed by AmGTs. AmGT was abbreviated as GT, and AmGT8<sub>A394F</sub> was abbreviated as GT8M. 3G/6G/25G/2'G means C3/C6/C25/C2'-O-glucoside. (c) Functional characterization of AmGT1 (GT1) and its variants with compounds **4**, **8** and **13** as substrates and UDP-Xyl as the sugar donor. 3X, C3-O-xyloside. (d) Combinatorial synthesis network of astragalosides **2–14** using AmGTs. The tetracyclic skeleton represents cycloastragenol (CA, **1**). Glc, glucosyl group. Xyl, xylosyl group. (e) UHPLC/MS chromatograms for combinatorial biosynthetic reaction samples. Extracted ion chromatograms was shown with extraction of  $m/z$  667.4052 ([CA + Xyl + HCOO]<sup>-</sup>),  $m/z$  697.4158 ([CA + Glc + HCOO]<sup>-</sup>),  $m/z$  829.4581 ([CA + Glc + Xyl + HCOO]<sup>-</sup>),  $m/z$  859.4687 ([CA + 2Glc + HCOO]<sup>-</sup>) and  $m/z$  991.5110 ([CA + 2Glc + Xyl + HCOO]<sup>-</sup>). Conversion rates noted were calculated using UHPLC/CAD (Figure S13). AmGT was abbreviated as GT. Peaks with red labels indicated target products in each reaction.



Meanwhile, both AmGT1 and AmGT8 showed good sugar donor promiscuity to use more than four different donors (Figure 6c). In total, nine catalytic products are rare ginsenosides, and 21 monoglycosides are unreported structures. A new compound, 3-O-

xyloside of 20R-PPD (**27**) was purified from the scaled-up reaction using AmGT1, and was characterized by NMR (Appendix S1).

AmGT1 could also catalyse the glycosylation of a variety of triterpenoids or steroids (Figure S15). Interestingly, it showed high



conversion rate to compounds with 3- $\beta$ -OH (**28–31**) and poor activity to 3- $\alpha$ -OH and phenolic 3-OH (**32** and **33**). The product of oleanolic acid (**30**) was prepared and characterized as oleanolic acid-3-O-glucoside (**34**) by NMR (Appendix **S1**).

### Biological activities of glycosylated products

As the effective components of *Astragalus* root, the saponin extract exhibited hepatoprotective and anti-virus activities (Li *et al.*, 2018; Liang *et al.*, 2020; Sun *et al.*, 2007; Tang *et al.*, 2010). However, the bioactivity of various astragalosides has rarely been reported except for a few major ones such as astragaloside IV. Thus, we investigated the bioactivities of the 13 astragalosides obtained through combinatorial synthesis. Compounds **6**, **9**, **10**, **11** and **14** showed protective effects against acetaminophen (APAP)-induced HepG2 cell death. The cell viabilities were from 67.6% to 76.0%, which was significantly higher than their aglycone (**1**, 62.3%) and the model group (45.7%) (Figure **7a**). Meanwhile, the inhibitory activities against key proteins of severe acute respiratory syndrome coronavirus (SARS-CoV-2) were evaluated *in vitro* (Figure **7b–e**, Figure **S16**). The inhibition of receptor-binding domain of the spike protein (S-RBD) significantly increased for the tri-glycoside **14** (IC<sub>50</sub> 15.17  $\mu$ M), compared to the aglycone (**1**, IC<sub>50</sub> > 40  $\mu$ M). In addition, the glycosylated product **34** also exhibited increased inhibitory effect against 3C-like protease (3CL<sup>pro</sup>, IC<sub>50</sub> 5.95  $\mu$ M) and papain-like protease (PL<sup>pro</sup>, IC<sub>50</sub> 7.57  $\mu$ M), compared to its aglycone **30**. These results indicated that glycosylation could improve the pharmacological activities of triterpenoids.

## Methods

### Plant materials and chemicals

Two-week-old *A. membranaceus* seedlings were obtained as previously reported (Chen *et al.*, 2022) and used for gene cloning (Figure **S17a**). A 3-year-old *A. membranaceus* plant was collected from Daxing'anling (Heilongjiang, China) and used for chemical analysis (Figure **S17b**). Reference standards **1**, **5**, **6**, **19–26** and

**28–33** were purchased from Chengdu Desite Biotechnology (China), Shanghai TargetMol Biotechnology (China), Solarbio Technology (Beijing, China) or Sigma-Aldrich (China). Reference standards **2**, **3**, **4** and **7** were previously purified and identified by NMR in our group (Zhang *et al.*, 2022). Other 12 glycosylated products (**8–14**, **17**, **18**, **27** and **34**) were purified and identified by NMR in this study (Appendix **S1**). UDP-sugar donors were purchased from Aifly Biotechnology (China). LC/MS grade methanol and acetonitrile were purchased from Fisher Scientific (USA).

### Chemical analysis of *A. membranaceus*

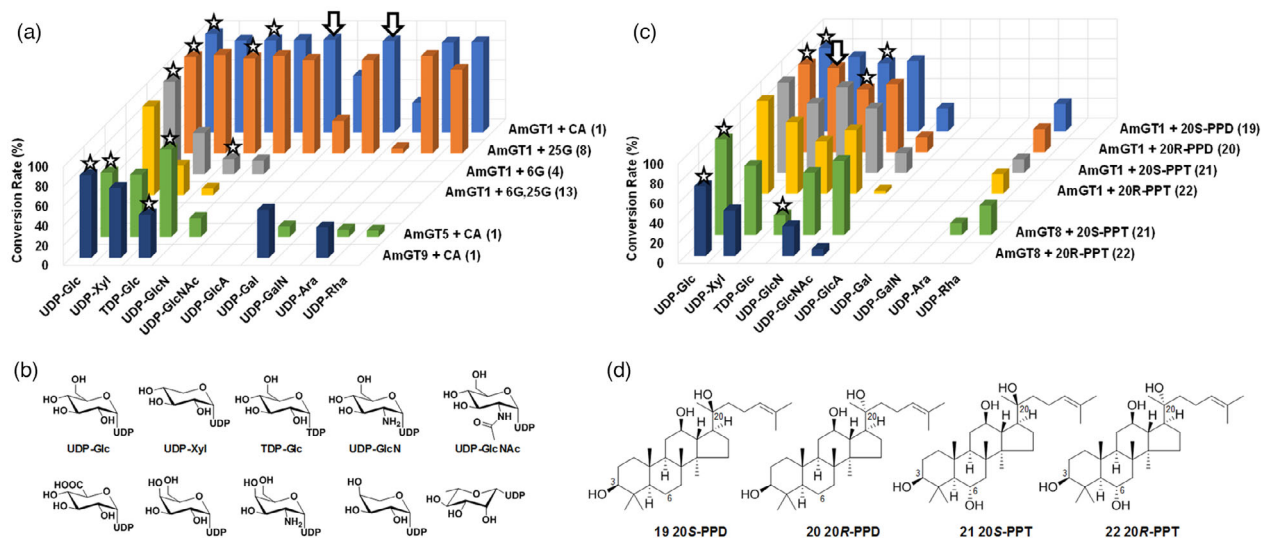
For chemical analysis, 4 mg dried powder of *A. membranaceus* roots was extracted by ultrasonication for 30 min in 1 mL of 75% methanol. After centrifugation, 200  $\mu$ L supernatant was mixed with 200  $\mu$ L of 13% NH<sub>3</sub>-H<sub>2</sub>O and heated at 65  $^{\circ}$ C for 1 h. After dryness, the mixture was redissolved with 1 mL of 75% methanol and centrifuged at 21 130 g for 30 min. The supernatant was analysed by LC/MS as described in 'UHPLC/CAD and UHPLC/MS analysis' of the Experiment.

### Transcriptome mining, phylogenetic analysis and transcript level analysis

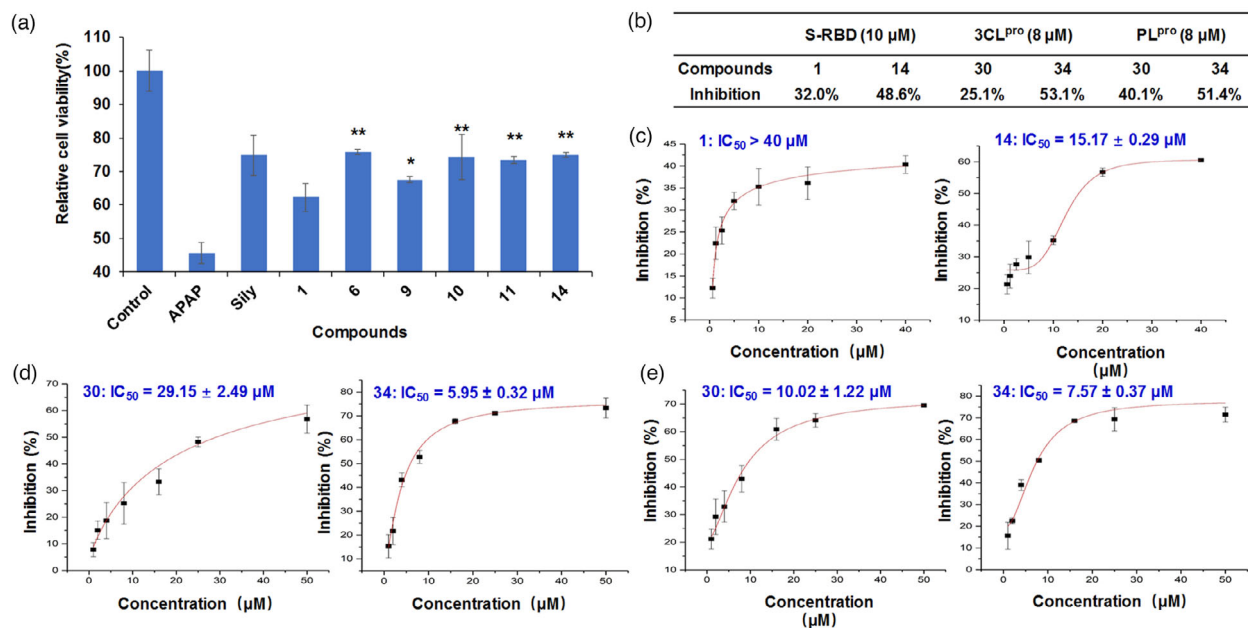
Two transcriptome datasets (NCBI Accession No. ERR706814 and SRR5343992) were used for gene screening. Plant triterpenoid GTs in Table **S2** were used as templates. For phylogenetic analysis, 25 selected plant triterpenoid GTs were analysed together with AmGT1/5/9/8. The phylogenetic tree was constructed using MEGA6 Software with the Maximum Likelihood method based on ClustalW multiple alignments (Jones *et al.*, 1992; Tamura *et al.*, 2013). The bootstrap consensus tree inferred from 1000 replicates was used.

### Molecular cloning, mutagenesis and expression of AmGTs

Total RNA was extracted from roots of 2-week-old *A. membranaceus* seedlings using the TranZol kit (Transgen Biotech,



**Figure 6** The sugar donor promiscuity of AmGT1/5/9. (a) Conversion rates for cycloastragenol (**1**), CA-6-O-Glc (**4**), CA-25-O-Glc (**8**) and CA-6,25-di-O-Glc (**13**) using different sugar donors. (b) Structures of UDP-sugar donors. (c) Conversion rates for 20 *S/R*-PPD (**19/20**) and 20 *S/R*-PPT (**21/22**) using different sugar donors. (d) Structures of compounds **19–22**. Asterisks in (a) and (c) indicated structures confirmed by comparing with reference standards, while arrows indicated compounds purified and characterized by NMR.



**Figure 7** Biological activities of the glycosylated products. (a) Hepatoprotective activities of catalytic products against APAP-induced injury of HepG2 cells. Silymarin (Sily) was used as the positive drug. For all groups,  $n = 3$ . \* $P < 0.05$ , \*\* $P < 0.01$  compared with compound 1 (Student's  $t$ -test). (b) Inhibitory activities of selected compounds against S-RBD, 3CL<sup>pro</sup> and PL<sup>pro</sup> of SARS-CoV-2. For all groups,  $n = 3$ . (c–e) Dose-dependent inhibition against S-RBD (c), 3CL<sup>pro</sup> (d) and PL<sup>pro</sup> (e) of selected catalytic products. For all groups,  $n = 3$ .

China) according to the manufacturer's instructions. The cDNA was synthesized with TransScript One-step gDNA Removal and cDNA Synthesis SuperMix (Transgen Biotech, China). The coding regions of *AmGT1*, *AmGT5* and *AmGT9* were amplified using primers in Table S4 with TransStart FastPfu DNA Polymerase (Transgen Biotech, China) and then cloned into pET28a(+) vectors using Quick-Change method (Bok and Keller, 2012). The mutants of *AmGTs* were constructed by a Fast MultiSite Mutagenesis System (Transgen Biotech, China) using the primers in Table S5.

After verification of the sequences, the recombinant plasmids were introduced into *E. coli* BL21(DE3) for heterologous expression. The procedures for protein expression and purification were same as reported (Chen *et al.*, 2019). Briefly, single colonies were grown in 500 mL LB medium containing kanamycin (50  $\mu$ g/mL) at 37 °C. After the OD<sub>600</sub> reached 0.4–0.6, the cells were induced with 0.1 mM isopropyl  $\beta$ -D-thiogalactoside at 18 °C for 18 h. The cells were then harvested by centrifugation. The recombinant proteins were purified using a Ni-NTA column and stored in storage buffer (20 mM Tris, 500 mM NaCl, 20% glycerol, pH = 7.5).

### Functional characterization, kinetic studies and combinatorial biosynthesis

The analytical reactions were carried out in a 100  $\mu$ L system, containing 50 mM Na<sub>2</sub>HPO<sub>4</sub>-NaH<sub>2</sub>PO<sub>4</sub> buffer (pH 8.0), 0.1 mM substrate, 0.5 mM UDP-sugar donor and 20  $\mu$ g of purified enzyme. After incubation at 37 °C for 2 h, the reactions were terminated with 100  $\mu$ L methanol. The mixtures were centrifuged at 21 130 g for 20 min and then analysed by UHPLC/CAD or UHPLC/MS. The conversion rates (%) were calculated by CAD peak area ( $A_{\text{product}}/A_{\text{substrate+product}} \times 100\%$ ). Blank samples were prepared from reactions without sugar donors.

Kinetic studies were performed in a 100  $\mu$ L system, containing 50 mM Na<sub>2</sub>HPO<sub>4</sub>-NaH<sub>2</sub>PO<sub>4</sub> buffer (pH 8.0), 0.33  $\mu$ g of purified enzyme, 2.5 mM of the substrate and 1–200  $\mu$ M of UDP-sugar.

The reactions were incubated at 37 °C for 10 min, and then terminated and analysed as described in the above paragraph. Products were quantified by LC/MS using a calibration curve. Kinetic parameters of *AmGTs* were calculated by fitting in the Michaelis–Menten plot.

For the combinatorial biosynthesis of di-/tri-glycosides, one-pot, step-by-step reactions were carried out in 100  $\mu$ L systems, containing 50 mM Na<sub>2</sub>HPO<sub>4</sub>-NaH<sub>2</sub>PO<sub>4</sub> buffer (pH 8.0) and 0.1 mM CA. For each step, 20  $\mu$ g of purified enzyme and 0.5 mM UDP-sugar donor were added. The mixture was incubated at 37 °C for 2 h before the enzyme and sugar donor were added for the next step. Finally, reactions were terminated and analysed as described in the first paragraph of this section. The conversion rates (%) were calculated by CAD peak area.

### Scaled-up reactions to prepare catalytic products

The scaled-up reactions were performed at 37 °C overnight using a total volume of 100 mL phosphate buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub>-NaH<sub>2</sub>PO<sub>4</sub>, pH 8.0). The reaction contained 10 mg of substrate ( $n$  mol), ~20 mg sugar donor (equivalent to  $2n$  mol) and approximate 30 mg purified *AmGT*. The reactions were terminated by adding 200 mL methanol, and centrifuged at 21 130 g for 20 min. The supernatants were concentrated and then redissolved in 2 mL of 50% methanol. The solution was subjected to an Agilent Zorbax SB-C18 column (9.4 mm  $\times$  250 mm, 5  $\mu$ m) and separated using a semi-preparative HPLC/ELSD (ELSD 6000, Alltech, USA), with acetonitrile and water containing 0.03% trifluoroacetic acid ( $v/v$ ) as the mobile phase. NMR spectra of the purified products were obtained on AVANCE III-400 or AVANCE III-600 instrument (Bruker, USA) (Appendix S1).

### Homology modelling, molecular docking and binding free energy calculations

The structure models of *AmGT1* and its mutants were generated by Alphafold2 (Tunyasuvunakool *et al.*, 2021). Molecular docking

of protein with ligands was carried out by Autodock (Trott and Olson, 2010). Models were visualized by AutoDocktools v1.5.6 and Pymol.

The protonation states of ionizable residues were determined using PROPKA3 (implemented in Schrödinger Suite, version 2021-3) prior to the MD simulations (Olsson *et al.*, 2011). According to the predicted pKa values, all ASP and GLU residues were deprotonated, while LYS and ARG were all protonated. For HIS residues, HIS 22, 38, 221 and 282, and 382 were  $\delta$ -protonated, HIS 228, 232, 300, 351, 360 and 463 were  $\epsilon$ -protonated, and all the other HIS residues were protonated at both  $\delta$ - and  $\epsilon$ -nitrogen sites. All the MD simulations were carried out using the Desmond package (Desmond Molecular Dynamics System, D. E. Shaw Research, New York) on the NVIDIA Tesla T4 equipped GPU nodes at the supercomputer centre of Linköping University. The OPLS4 force field was used for both protein and ligand atoms with the default atomic charge scheme (Lu *et al.*, 2021). In a MD simulation, the solvent molecules were added to a 12-Å buffering region surrounding the protein, which generated an orthorhombic box containing ~17 300 TIP3P water molecules with certain amount of counter ions Na<sup>+</sup> and Cl<sup>-</sup> to neutralize the system as well as to mimic the physical salt concentration of 0.15 M (Jorgensen *et al.*, 1983). To maintain the temperature and pressure of simulations at 300.0 K and 1.0 atm, the Nose-Hoover chain thermostat (Martyna and Klein, 1992) and Martyna-Bobias-Klein barostat (Martyna, 1994) were used respectively. The default procedure for minimization and equilibration implemented in Desmond was used before the 100-ns production simulation for each protein-ligand system. The trajectory was saved every 50 ps. Analysis of the trajectory files was completed using the Visual Molecular Dynamics (VMD) package, version 1.9.3 (Humphrey *et al.*, 1996).

The MM/GBSA (molecular mechanics, the generalized Born model and solvent accessibility) binding free energy was calculated using the Prime module of Schrödinger Suite (version 2021-3). A total of 400 snapshots evenly extracted from 100-ns trajectory were used for MM/GBSA calculations. We used the OPLS4 force field to refine the residues within 8 Å of the ligand in the continuum solvation VSGB (variable dielectric surface generalized Born) (Li *et al.*, 2011) before MM/GBSA calculations. The mean and deviation values calculated from the extracted snapshots were reported in this study.

### UHPLC/CAD and UHPLC/MS analysis

UHPLC/CAD analysis was performed on a Vanquish Flex UHPLC Dual System (ThermoFisher Scientific, USA) coupled with a Charged Aerosol Detector (CAD). Samples (5  $\mu$ L) were separated on an Acquity UHPLC HSS T3 column (100 mm  $\times$  2.1 mm, 1.8  $\mu$ m, Waters, USA). The mobile phase consisted of water containing 0.1% formic acid (*v/v*, A) and acetonitrile (B). A gradient elution programme was used: 0–1 min, 30% B; 1–12 min, 30%–50% B; 12–15 min, 50%–100% B; 15–18 min, 100% B. The flow rate was 0.3 mL/min, and the column temperature was 55 °C. An automatic inverse gradient was used to eliminate the impact of changes in solvent composition during gradient elution.

UHPLC/MS analysis was performed on a Vanquish UHPLC system coupled with a Q-Exactive quadrupole-Orbitrap mass spectrometer equipped with a heated electrospray ionization source (ThermoFisher Scientific, USA). Samples (2  $\mu$ L) were separated on an Acquity UHPLC HSS T3 column (150 mm  $\times$  2.1 mm, 1.8  $\mu$ m, Waters, USA) at 50 °C. The mobile

phase and gradient programme were the same as UHPLC/CAD method. The mass spectrometer was operated in the negative and positive ion modes. The MS parameters were as follows: sheath gas pressure, 45 arb; auxiliary gas pressure, 10 arb; discharge voltage, 4.5 kV; capillary temperature, 350 °C.

### Hepatoprotective activities against APAP-induced cell death

To evaluate the hepatoprotective activities against APAP-induced cell death, HepG2 cells were treated with 20  $\mu$ M compounds and 14 mM APAP for 24 h, and the cell viabilities were measured using CCK-8 method (Kuang *et al.*, 2017). Silymarin (20  $\mu$ M) was used as the positive control.

### Inhibitory activity assay against S-RBD, 3CL<sup>pro</sup> and PL<sup>pro</sup>

The inhibition activity against S-RBD was measured by ELISA as described in our previous study (Yi *et al.*, 2022a,b). Briefly, the spike protein DRA49 (0.3  $\mu$ g/mL) was incubated in the coating solution (Na<sub>2</sub>CO<sub>3</sub> 1.59 mg/mL, NaHCO<sub>3</sub> 2.93 mg/mL, pH 9.6) on 96-well microplates at 4 °C overnight. The plates were washed for three times, incubated at 37 °C for 40 min using wash buffer containing 2% bovine serum albumin, and washed again for three times. Candidate compounds (10  $\mu$ M) or positive drug (S-hlgG1, 5  $\mu$ g/mL) were added and incubated for 1 h. ACE2 (angiotensin-converting enzyme 2, 0.3  $\mu$ g/mL) was then added and incubated for another 30 min. At last, the enzyme-linked antibody SA-HRP (1 : 10 000) was added. Tetramethyl benzidine was used as the chromogenic agent. Finally, the reaction was stopped by 1 M HCl and the enzyme activity was determined at OD<sub>450nm</sub> on LabServ K3 Microplate Reader (Thermo Fisher Scientific, USA; Yi *et al.*, 2022a,b).

The inhibition activity against 3CL<sup>pro</sup> and PL<sup>pro</sup> was measured by fluorescence resonance energy transfer method. For 3CL<sup>pro</sup>, the reaction mixture contained 12.5  $\mu$ g/mL protein, 3 mM substrate, 8  $\mu$ M candidate compound or positive drug (GC376), in 20 mM Tris-HCl (pH 7.0) as the reaction buffer. After incubated at 25 °C for 10 min, the fluorescence was detected (emission at 490 nm and excitation at 340 nm) on a FlexStation 3 Multi-Mode Microplate Reader (Molecular Devices, China; Dai *et al.*, 2020; Yi *et al.*, 2022a,b). For PL<sup>pro</sup>, the reaction mixture contained 12.5  $\mu$ g/mL protein, 5 mM substrate, 8  $\mu$ M candidate compound or positive drug (GRL0617), in 50 mM HEPES (pH 7.5) as the reaction buffer. After incubated at 25 °C for 10 min, the fluorescence was detected (emission at 535 nm and excitation at 340 nm) on a FlexStation 3 Multi-Mode Microplate Reader (Yi *et al.*, 2022a,b).

### Statistical analysis

For substrate promiscuity analysis, qRT-PCR analysis, hepatoprotective activities and S-RBD/3CL<sup>pro</sup>/PL<sup>pro</sup> inhibition assays, all measurements were in triplicate and the data were expressed as mean  $\pm$  SD. The significance of difference between groups was determined by two-tailed, paired Student's *t*-tests for bioactivity assays.

### Acknowledgements

The authors thank Yu Cui and Ruiping Chai from Thermo Fisher Scientific for providing technical support during data collection. The computations were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at the National Supercomputer Center at Linköping University (Sweden).

This work was supported by the National Natural Science Foundation of China (82122073, 81973448, 81725023), Beijing Natural Science Foundation (JQ18027), Swedish Research Council through the Grant (2021-3-22) and State Key Laboratory of Phytochemistry and Plant Resources in West China (P2020-KF07).

## Conflict of interest statement

The authors declare no conflict of interests.

## Authors' contributions

XQ, MY, KC and MZ planned and designed the research. KC, MZ, BG, AH, YB, JF, RY, YY, ZW and HL performed experiments. JL, HÅ and MZ contributed to the theoretical calculation. KC, MZ, XQ and MY analysed the data. XQ and KC wrote the manuscript.

## References

- Barbic, M., Macabeo, A.P.G., Kreft, S. and Heilmann, J. (2010) Cycloastragenol glycosides from *Astragalus illyricus*. *Biochem. Syst. Ecol.* **28**, 460–462.
- Bok, J.W. and Keller, N.P. (2012) Fast and easy method for construction of plasmid vectors using modified quick-change mutagenesis. *Methods Mol. Biol.* **944**, 163–174.
- Chen, K., Hu, Z.M., Song, W., Wang, Z.L., He, J.B., Shi, X.M., Cui, Q.H. et al. (2019) Diversity of O-glycosyltransferases contributes to the biosynthesis of flavonoid and triterpenoid glycosides in *Glycyrrhiza uralensis*. *ACS Synth. Biol.* **8**, 1858–1866.
- Chen, K., Zhang, M., Xu, L.L., Yi, Y., Wang, L.L., Wang, H.T., Wang, Z.L. et al. (2022) Identification of oxidosqualene cyclases associated with saponin biosynthesis from *Astragalus membranaceus* reveals a conserved motif important for catalytic function. *J. Adv. Res.* <https://doi.org/10.1016/j.jare.2022.03.014>
- Chu, C., Qi, L.W., Liu, E.H., Li, B., Gao, W. and Li, P. (2010) Radix Astragalii (*Astragalus*): latest advancements and trends in chemistry, analysis, pharmacology and pharmacokinetics. *Curr. Org. Chem.* **14**, 1792–1807.
- Corseello, M.A. and Garg, N.K. (2015) Synthetic chemistry fuels interdisciplinary approaches to the production of artemisinin. *Nat. Prod. Rep.* **32**, 359–366.
- Costa, F., Barber, C.J.S., Kim, Y.B., Reed, D.W., Zhang, H.X., Fett-Neto, A.G. and Covello, P.S. (2017) Molecular cloning of an ester-forming triterpenoid: UDP-glucose 28-O-glucosyltransferase involved in saponin biosynthesis from the medicinal plant *Centella asiatica*. *Plant Sci.* **262**, 9–17.
- Dai, E.H., Zhang, B., Jiang, X.M., Su, H.X., Li, J. and Zhao, Y. (2020) Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science*, **368**, 1331–1335.
- Erthmann, P.O., Agerbirk, N. and Bak, S. (2018) A tandem array of UDP-glycosyltransferases from the UGT73C subfamily glycosylate saponins, forming a spectrum of mono- and bisdesmosidic saponins. *Plant Mol. Biol.* **97**, 37–55.
- Guerra-Bubb, J., Croteau, R. and Williams, R.M. (2012) The early stages of taxol biosynthesis: an interim report on the synthesis and identification of early pathway metabolites. *Nat. Prod. Rep.* **29**, 683–696.
- Gulcernal, D., Masullo, M., Bedir, E., Festa, M., Karayildirim, T., Alankus-Caliskan, O. and Piacente, S. (2012) Triterpene glycosides from *Astragalus angustifolius*. *Planta Med.* **78**, 720–729.
- Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 3–38.
- Ionkova, I., Shkondrov, A., Krasteva, I. and Ionkov, T. (2014) Recent progress in phytochemistry, pharmacology and biotechnology of *Astragalus* saponins. *Phytochem. Rev.* **13**, 343–374.
- Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992) The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282.
- Jorgensen, W.L., Chandrasekhar, J. and Madura, J.D. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.
- Kim, O.T., Jin, M.L., Lee, D.Y. and Jetter, R. (2017) Characterization of the asiatic acid glucosyltransferase, UGT73AH1, involved in asiaticoside biosynthesis in *Centella asiatica* (L.) Urban. *Int. J. Mol. Sci.* **18**, 11.
- Kollman, P.A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M. et al. (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33**, 889–897.
- Kuang, Y., Lin, Y., Li, K., Song, W., Ji, S., Qiao, X., Zhang, Q.Y. et al. (2017) Screening of hepatoprotective compounds from licorice against carbon tetrachloride and acetaminophen induced HepG2 cells injury. *Phytomedicine*, **34**, 59–66.
- Kurze, E., Wust, M., Liao, J.R., McGraphery, K., Hoffmann, T., Song, C.K. and Schwab, W. (2021) Structure-function relationship of terpenoid glycosyltransferases from plants. *Nat. Prod. Rep.* **39**, 389–409.
- Li, J.N., Abel, R., Zhu, K., Cao, Y.X., Zhao, S.W. and Friesner, R.A. (2011) The VSG 2.0 model: a next generation energy model for high resolution protein structure modeling. *Proteins*, **79**, 2794–2812.
- Li, L., Huang, W.X., Wang, S.K., Sun, K.C., Zhang, W.X., Ding, Y.M., Zhang, L. et al. (2018) Astragaloside IV attenuates acetaminophen-induced liver injuries in mice by activating the Nrf2 signaling pathway. *Molecules*, **23**, 2032.
- Li, J., Yang, J.G., Mu, S.C., Shang, N., Liu, C., Zhu, Y.M., Cai, Y. et al. (2020) Efficient O-glycosylation of triterpenes enabled by protein engineering of plant glycosyltransferase UGT74AC1. *ACS Catal.* **10**, 3629–3639.
- Li, X.D., Wang, Y.M., Fan, Z.J., Wang, Y., Wang, P.P., Yan, X. and Zhou, Z.H. (2021) High-level sustainable production of the characteristic protopanaxatriol-type saponins from *Panax* species in engineered *Saccharomyces cerevisiae*. *Metab. Eng.* **66**, 87–97.
- Liang, Y.X., Zhang, Q.Y., Zhang, L.J., Wang, R.F., Xu, X.Y. and Hu, X.H. (2020) *Astragalus membranaceus* treatment protects raw264.7 cells from influenza virus by regulating G1 phase and the TLR3-mediated signaling pathway. *Evid-Based Compl. Alt.* **2019**, 2971604.
- Lu, C., Wu, C.J., Ghoreishi, D., Chen, W., Wang, L.L., Damm, W., Ross, G.A. et al. (2021) OPLS4: improving force field accuracy on challenging regimes of chemical space. *J. Chem. Theory Comput.* **17**, 4291–4300.
- Martyna, G.J. (1994) Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177–4189.
- Martyna, G.J. and Klein, M.L. (1992) Nosé–Hoover chains: the canonical ensemble via continuous dynamics. *J. Chem. Phys.* **97**, 2635–2643.
- Najmi, A., Javed, S.A., Al Bratty, M. and Alhazmi, H.A. (2022) Modern approaches in the discovery and development of plant-based natural products and their analogues as potential therapeutic agents. *Molecules*, **27**, 349.
- Nomura, Y., Seki, H., Suzuki, T., Ohyama, K., Mizutani, M., Kaku, T., Tamura, K. et al. (2019) Functional specialization of UDP-glycosyltransferase 73P12 in licorice to produce a sweet triterpenoid saponin, glycyrrhizin. *Plant J.* **99**, 1127–1143.
- Olsson, M.H.M., Sondergaard, C.R., Rostkowski, M. and Jensen, J.H. (2011) PROPKA3: consistent treatment of internal and surface residues in empirical pK(a) predictions. *J. Chem. Theory Comput.* **7**, 525–537.
- Patridge, E., Gareiss, P., Kinch, M.S. and Hoyer, D. (2016) An analysis of FDA-approved drugs: natural products and their derivatives. *Drug Discov. Today*, **21**, 204–207.
- Rahimi, S., Kim, J., Mijakovic, I., Jung, K.H., Choi, G., Kim, S.C. and Kim, Y.J. (2019) Triterpenoid-biosynthetic UDP-glycosyltransferases from plants. *Biotechnol. Adv.* **37**, 18.
- Sayama, T., Ono, E., Takagi, K., Takada, Y., Horikawa, M., Nakamoto, Y., Hirose, A. et al. (2012) The Sg-1 glycosyltransferase locus regulates structural diversity of triterpenoid saponins of soybean. *Plant Cell*, **24**, 2123–2138.
- Su, H.F., Shaker, S., Kuang, Y., Zhang, M., Ye, M. and Qiao, X. (2021) Phytochemistry and cardiovascular protective effects of Huang-Qi (*Astragalus Radix*). *Med. Res. Rev.* **41**, 1999–2038.
- Sun, W.Y., Wei, W., Wu, L., Gui, S.Y. and Wang, H. (2007) Effects and mechanisms of extract from *Paeonia lactiflora* and *Astragalus membranaceus* on liver fibrosis induced by carbon tetrachloride in rats. *J. Ethnopharmacol.* **112**, 514–523.
- Takagi, K., Yano, R., Tochigi, S., Fujisawa, Y., Tsuchinaga, H., Takahashi, Y., Takada, Y. et al. (2018) Genetic and functional characterization of Sg-4

- glycosyltransferase involved in the formation of sugar chain structure at the C-3 position of soybean saponins. *Phytochemistry*, **156**, 96–105.
- Tamura, K., Stecher, G., Peterson, D., Filipi, A. and Kumar, S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729.
- Tang, L., Liu, Y., Wang, Y.L. and Long, C.L. (2010) Phytochemical analysis of an antiviral fraction of *Radix Astragalii* using HPLC-DAD-ESI-MS/MS. *J. Nat. Med.* **64**, 182–186.
- Trott, O. and Olson, A.J. (2010) Software news and update AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461.
- Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Zidek, A., Bridgland, A. et al. (2021) Highly accurate protein structure prediction for the human proteome. *Nature*, **596**, 590–596.
- Wang, D., Wang, J.H., Shi, Y.S., Li, R.S., Fan, F.Y., Huang, Y., Li, W.X. et al. (2020) Elucidation of the complete biosynthetic pathway of the main triterpene glycosylation products of *Panax notoginseng* using a synthetic biology platform. *Metab. Eng.* **61**, 131–140.
- Wei, W., Wang, P.P., Wei, Y.J., Liu, Q.F., Yang, C.S., Zhao, G.P., Yue, J.M. et al. (2015) Characterization of *Panax ginseng* UDP-glycosyltransferases catalyzing protopanaxatriol and biosyntheses of bioactive ginsenosides F1 and Rh1 in metabolically engineered yeasts. *Mol. Plant*, **8**, 1412–1424.
- Yi, Y., Li, J.H., Lai, X.Y., Zhang, M., Kuang, Y., Bao, Y.O.J., Yu, R. et al. (2022a) Natural triterpenoids from licorice potently inhibit SARS-CoV-2 infection. *J. Adv. Res.* **36**, 201–210.
- Yi, Y., Zhang, M., Xue, H., Yu, R., Bao, Y.O.J., Kuang, Y., Chai, Y. et al. (2022b) Schaftoside inhibits 3CL<sup>pro</sup> and PL<sup>pro</sup> of SARS-CoV-2 virus and regulates immune response and inflammation of host cells for the treatment of COVID-19. *Acta. Pharm. Sin. B.* **12**, 4154–4164.
- Zhang, L., Ren, S.C., Liu, X.F., Liu, X.C., Guo, F., Sun, W.T., Feng, X.D. et al. (2020a) Mining of UDP-glycosyltransferases in licorice for controllable glycosylation of pentacyclic triterpenoids. *Biotechnol. Bioeng.* **117**, 3651–3663.
- Zhang, M., Li, F.D., Li, K., Wang, Z.L., Wang, Y.X., He, J.B., Su, H.F. et al. (2020b) Functional characterization and structural basis of an efficient di-C-glycosyltransferase from *Glycyrrhiza glabra*. *J. Am. Chem. Soc.* **142**, 3506–3512.
- Zhang, M., Yi, Y., Gao, B.H., Su, H.F., Bao, Y.O.J., Shi, X.M., Wang, H.D. et al. (2022) Functional characterization and protein engineering of a triterpene 3-*O*-glycosyltransferase reveal a conserved residue critical for the regioselectivity. *Angew. Chem. Int. Ed.* **61**, e202113587.
- Zhao, J.N., Wang, R.F., Zhao, S.J. and Wang, Z.T. (2020) Advance in glycosyltransferases, the important bioparts for production of diversified ginsenosides. *Chin. J. Nat. Med.* **18**, 643–658.
- mongholicus with cycloastragenol as the skeleton. Acylation was not included.
- Table S2** Plant triterpenoid glycosyltransferases reported.
- Table S3** Primers for qRT-PCR experiment.
- Table S4** Primers of AmGTs for recombinant plasmid construction with pET28a(+).
- Table S5** Primers used in site-directed mutagenesis experiments.
- Figure S1** LC/MS analysis of *Astragalus* roots after ammonia hydrolysis.
- Figure S2–S6** HR-MS and HR-MS/MS data of compounds 1–16.
- Figure S7** Relative transcript level of AmGTs in different tissues through qRT-PCR analysis.
- Figure S8** SDS-PAGE analysis of AmGTs characterized in this study.
- Figure S9** Effects of various conditions on the activity of AmGTs.
- Figure S10** Protein sequence alignment for 53 triterpene GTs.
- Figure S11** The functional characterization of mutants of AmGT1 with CA as the substrate, UDP-Glc or UDP-Xyl as the sugar donor.
- Figure S12** The kinetic parameters of AmGT1 and its mutants.
- Figure S13** Combinatorial biosynthesis of di-/tri- glycosides analyzed by UHPLC/CAD.
- Figure S14** Production of rare ginsenosides catalyzed by AmGTs.
- Figure S15** Glucosylation of other triterpenoids or steroids catalyzed by AmGT1.
- Figure S16** Inhibitory activities of selected compounds against S-RBD, 3CL<sup>pro</sup>, and PL<sup>pro</sup> of SARS-CoV-2.
- Figure S17** Plant materials used in this study.
- Figure S18–S24** Glycosylation of compounds 1–6 and 8 catalyzed by AmGTs with UDP-Glc as the sugar donor.
- Figure S25** Glycosylation of compounds 4/8/13 catalyzed by AmGT1 and its mutants with UDP-Xyl as the sugar donor.
- Figure S26–S33** Glycosylation of compound 1 catalyzed by AmGTs with different sugar donors.
- Figure S34–S36** Glycosylation of compounds 8, 4, and 13 catalyzed by AmGT1 with different sugar donors.
- Figure S37–S40** Glycosylation of compounds 19–22 catalyzed by AmGTs with UDPGlc as the sugar donor.
- Figure S41–S44** Glycosylation of compounds 19–22 catalyzed by AmGTs with UDPXyl as the sugar donor.
- Figure S45–S48** Glycosylation of compounds 19–22 catalyzed by AmGTs with TDPGlc as the sugar donor.
- Figure S49–S52** Glycosylation of compounds 19–22 catalyzed by AmGTs with UDPGlcN as the sugar donor.
- Figure S53–S56** Glycosylation of compounds 19–22 catalyzed by AmGTs with UDPGlcNac as the sugar donor.
- Figure S57–S60** Glycosylation of compounds 19–22 catalyzed by AmGTs with UDPAra as the sugar.

## Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Table S1** Glycosylation patterns of astragalosides reported in *Astragalus membranaceus* and *Astragalus membranaceus* var.