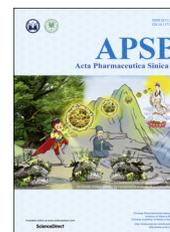




Chinese Pharmaceutical Association  
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

[www.elsevier.com/locate/apsb](http://www.elsevier.com/locate/apsb)  
[www.sciencedirect.com](http://www.sciencedirect.com)



ORIGINAL ARTICLE

# The enzymatic biosynthesis of acylated steroidal glycosides and their cytotoxic activity



Ming Liu, Jian-Qiang Kong\*

State Key Laboratory of Bioactive Substance and Function of Natural Medicines & Ministry of Health Key Laboratory of Biosynthesis of Natural Products, Institute of Materia Medica, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100050, China

Received 28 February 2018; received in revised form 9 April 2018; accepted 16 April 2018

## KEY WORDS

Steroidal glycosyltransferase;  
Steroidal glycoside acyltransferase;  
*Ornithogalum saundersiae*;  
Acylated steroidal glycosides;  
LacA

**Abstract** Herein we describe the discovery and functional characterization of a steroidal glycosyltransferase (SGT) from *Ornithogalum saundersiae* and a steroidal glycoside acyltransferase (SGA) from *Escherichia coli* and their application in the biosynthesis of acylated steroidal glycosides (ASGs). Initially, an SGT gene, designated as OsSGT1, was isolated from *O. saundersiae*. OsSGT1-containing cell free extract was then used as the biocatalyst to react with 49 structurally diverse drug-like compounds. The recombinant OsSGT1 was shown to be active against both 3 $\beta$ - and 17 $\beta$ -hydroxyl steroids. Unexpectedly, in an effort to identify OsSGT1, we found the bacteria *lacA* gene in *lac* operon actually encoded an SGA, specifically catalyzing the acetylations of sugar moieties of steroid 17 $\beta$ -glucosides. Finally, a novel enzymatic two-step synthesis of two ASGs, acetylated testosterone-17-*O*- $\beta$ -glucosides (AT-17 $\beta$ -Gs) and acetylated estradiol-17-*O*- $\beta$ -glucosides (AE-17 $\beta$ -Gs), from the abundantly available free steroids using OsSGT1 and EcSGA1 as the biocatalysts was developed. The two-step process is characterized by EcSGA1-catalyzed regioselective acylations of all hydroxyl groups on the sugar unit of unprotected steroidal glycosides (SGs) in the late stage, thereby significantly streamlining the synthetic route towards ASGs and thus forming four monoacylates. The improved cytotoxic activities of 3'-acetylated testosterone-17-*O*- $\beta$ -glucoside towards seven human tumor cell lines were thus observable.

**Abbreviations:** 6'-AE-17 $\beta$ -G, 6'-acetylated estradiol 17-*O*- $\beta$ -glucoside; 6'-AT-17 $\beta$ -G, 6'-acetylated testosterone 17-*O*- $\beta$ -glucoside; AE-17 $\beta$ -G, acetylated estradiol-17-*O*- $\beta$ -glucoside; ASGs, acylated steroidal glycosides; AT-17 $\beta$ -G, acetylated testosterone-17-*O*- $\beta$ -glucoside; E-17 $\beta$ -G, estradiol-17-*O*- $\beta$ -glucoside; EcSGA1, *E. coli* steroidal glucoside acetyltransferase; HPLC–SPE–NMR, high-performance liquid chromatography–solid phase extraction–NMR spectroscopy; IPTG, isopropyl- $\beta$ -D-thiogalactoside; ORF, open reading frame; PSBD, putative steroid-binding domain; PSPG, plant secondary product glycosyltransferase box; RIN, RNA integrity number; RP-HPLC, reversed phase high-performance liquid chromatography; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; SGs, steroidal glycosides; SGAs, steroidal glycoside acyltransferases; SGEs, steroidal glycoside esters; SGTs, steroidal glycosyltransferases; T-17 $\beta$ -G, testosterone-17-*O*- $\beta$ -glucoside; UDP-Glc, UDP-D-glucose; UDP-Gal, UDP-D-galactose; UDP-GlcA, UDP-D-glucuronic acid; UDP-GlcNAc, UDP-N-acetylglucosamine; UDP-Xyl, UDP-D-xylose; UDP-Ara, UDP-L-arabinose; UDP-GalA, UDP-D-Galacturonic acid; UTR, untranslated region

\*Corresponding author.

E-mail address: [jianqiangk@imm.ac.cn](mailto:jianqiangk@imm.ac.cn) (Jian-Qiang).

Peer review under responsibility of Institute of Materia Medica, Chinese Academy of Medical Sciences and Chinese Pharmaceutical Association.

<https://doi.org/10.1016/j.apsb.2018.04.006>

2211-3835 © 2018 Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Steroidal glycosides (SGs) are characterized by a steroidal skeleton glycosidically linked to sugar moieties, which can be further acylated with aliphatic and aromatic acids thus forming complex acylated steroidal glycosides (ASGs)<sup>1</sup>. The resulting steroidal glycoside esters (SGEs) exhibit a wide variety of biological activities, like cholesterol-lowering effect<sup>2</sup>, anti-diabetic properties<sup>3</sup>, anti-complementary activity<sup>4</sup>, immunoregulatory functions<sup>5,6</sup> and anti-cancer actions<sup>7-9</sup>, which made ASGs promising compounds with pharmaceutical potential. Numerous methods, including direct extraction<sup>9</sup>, chemical synthesis<sup>10,11</sup> and biosynthesis<sup>12</sup>, have been developed to synthesize these acylated steroidal glycosides. Direct extraction from varied organisms is one of the main methods to obtain ASGs<sup>7-9</sup>. However, the content of ASGs was usually low in natural sources<sup>7-9</sup>. Moreover, the extraction routes were highly time-consuming and required laborious purification procedures<sup>7-9,13</sup>, resulting in poor yields and/or low purity of the final products. The production of ASGs was also achieved by chemical synthesis previously<sup>10-14</sup>. However, these efforts often encounter a fundamental challenge, namely, regioselective acylation of single hydroxyl group of unprotected SGs in the late stage of the chemical synthesis of ASGs. SGs generally possess multiple hydroxyl groups with similar reactivity. Regioselective acylation of a particular one of multiple hydroxyl groups generally requires multi-step protection/deprotection procedures, which makes the synthetic pathway of these SGEs costly, wasteful, long and time-consuming, and results in low yield in the end.

The biosynthesis of ASGs from free steroids based on enzymatic catalysis was deemed to reduce the number of protection/deprotection steps due to the high selectivity of enzymes. Theoretically, the biosynthesis of ASGs includes two steps. In the first reaction, the sugar moiety from nucleotide-activated glycosyl donors was attached to steroids at different positions, most commonly at the C-3 hydroxyl group (OH), under the action of nucleotide dependent SGTs<sup>15,16</sup>. The glycosylation of a hydroxyl group at the C-3 position of steroids was well characterized and a few of steroidal 3 $\beta$ -glucosyltransferases were isolated from diverse species<sup>17,18</sup>. However, the reports of SGTs specific for positions other than C-3 of steroids are limited.

The sugar moieties of the resultant SGs can further be acylated by SGAs to form ASGs in the next step<sup>1,16</sup>. Compared to SGTs, surprisingly little is known about SGAs. Up to date, no SGA genes has yet been cloned, which in turn limit the enzyme-mediated biosynthesis of ASGs. Hence, the successful gene isolation and functional characterization of SGAs have become a prominent challenge for bioproduction of ASGs. Herein, the functional characterizations of a plant-derived SGT with activity against both 3 $\beta$ - and 17 $\beta$ -hydroxyl steroids, and a bacterial SGA, as well as their application in the biosynthesis of ASGs are reported. Initially, a steroidal glycosyltransferase OsSGT1 was isolated from medicinal plant *O. saundersiae* and showed activities for 3 $\beta$ - and 17 $\beta$ -hydroxyl steroids. Unexpectedly, in an effort to identify the function of OsSGT1, we characterized LacA protein (designated

as EsSGA1) from *E. coli* as a SGA, catalyzing testosterone-17-*O*- $\beta$ -glucoside (T-17 $\beta$ -G) and estradiol-17-*O*- $\beta$ -glucoside (E-17 $\beta$ -G) to form corresponding acylates. Further, under the synergistic actions of OsSGT1 and EsSGA1, the biosynthetic preparation of two acylated steroidal glycosides (ASGs), namely acetylated T-17 $\beta$ -Gs (AT-17 $\beta$ -Gs) and E-17 $\beta$ -Gs (AE-17 $\beta$ -Gs), was first achieved, thereby yielding four monoacetylated steroidal glycosides, namely 2'-*O*, 3'-*O*, 4'-*O* and 6'-*O*-acylates (Scheme 1). The cytotoxic activities of these monoacylates were evaluated against seven human tumor cell lines (HCT116, Bel7402, MGC803, Capan2, NCI-H460, NCI-H1650 and A549) and 3'-acetylated testosterone-17-*O*- $\beta$ -glucoside was observed to display improved cytotoxic activity against these seven cell lines (Scheme 1).

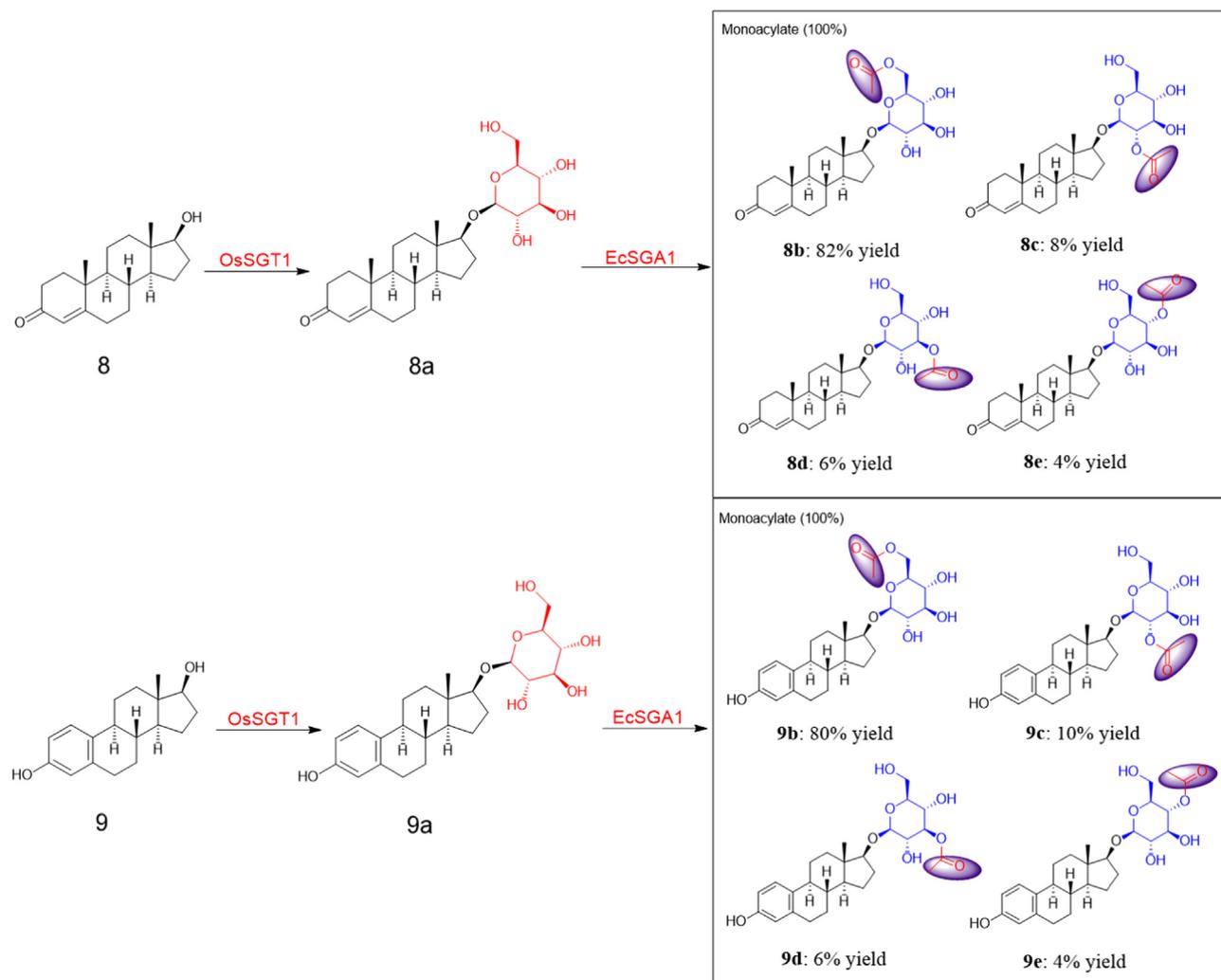
## 2. Results

### 2.1. Functional annotation and retrieval of unigenes encoding SGT

The species *O. saundersiae* is a monocotyledonous plant rich in steroidal glycosides, suggesting that it may contain SGTs responsible for the glycosylation of steroidal aglycons<sup>7-9</sup>. *O. saundersiae* is thus selected as the candidate plant for SGTs isolation. The transcriptome of *O. saundersiae* was thus sequenced with the aim of isolating genes encoding SGTs. A total of 92,995,146 raw reads were generated after the transcriptome sequencing of *O. saundersiae*. After removal of dirty reads with adapters, unknown or low quality bases, a total of 82,518,740 clean reads were retained. These clean reads were combined by assembling soft trinity to form longer unigenes. Finally, an RNA-seq database containing 107,084 unigenes with mean length of 766 bp was obtained. Next, these unigenes were aligned to publicly available protein databases for functional annotations, retrieving unigenes displaying the highest sequence similarity with SGTs. Unigene 32070 with 2182 bp in length was thus retrieved from the unigene database for its high similarity with SGTs (Supplementary Information Fig. S1). Moreover, ORF Finder result showed that this unigene contained a complete open reading frame (ORF) of 1773 bp, starting at nucleotide 99 with an ATG start codon and ending at position 1871 with a TGA stop codon. The unigene contained 98 bp of 5'-UTR (untranslated region) and 311 bp of 3'-UTR. Therefore, unigene 32070 was selected for further investigation.

### 2.2. Sequence identification of cDNA encoding steroidal glycosyltransferase

To verify the identity of unigene 32070, a nested PCR assay was therefore carried out to amplify the cDNA corresponding to the ORF of unigene 32070 using gene-specific primers (Supplementary Information Table S1). An expected band with approximately 1.7 kb was obtained, as observed in agarose gel electrophoresis (Supplementary Information Fig. S2A). The amplicon was then inserted into pEASY<sup>TM</sup>-Blunt plasmid



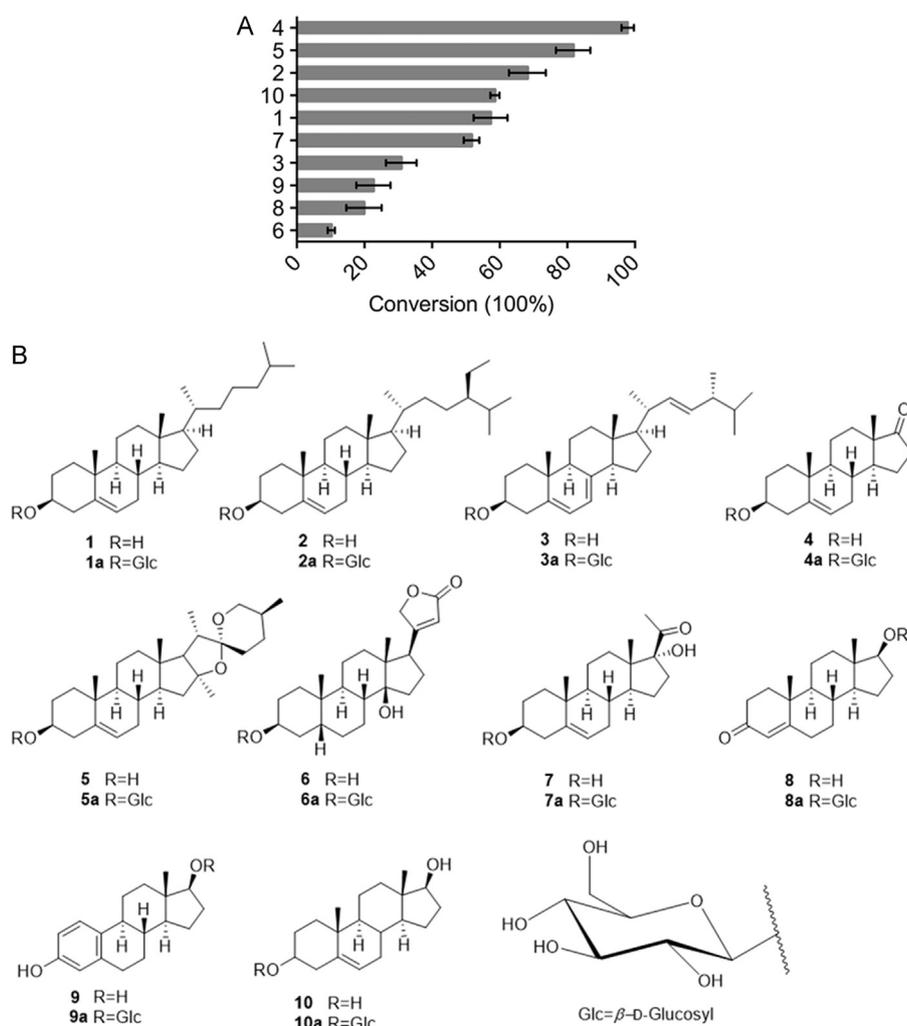
**Scheme 1** An enzymatic two-step synthesis of AT-17 $\beta$ -G (**8b–8e**) and AE-17 $\beta$ -G (**9b–9e**) from the free steroids testosterone (**8**) and estradiol (**9**). Firstly, two SGs, T-17 $\beta$ -G (**8a**) and E-17 $\beta$ -G (**9a**), were prepared from their corresponding steroidal substrates testosterone (**8**) and estradiol (**9**) in the presence of a steroidal glycosyltransferase OsSGT1 from *O. saundersiae*. The resulting T-17 $\beta$ -G (**8a**) was further regioselectively acetylated under the action of an acyltransferase EcSGA1 from *E. coli*, thereby yielding four monoacylated steroidal glycosides (**8b–8e**) with the yield ratio of 82:8:6:4. Likewise, E-17 $\beta$ -G (**9a**) was acetylated by EcSGA1 to form monoacylated products **9b–9e** in a ratio of 80:10:6:4.

(Supplementary Information Table S2) to form a recombinant vector for sequencing. Results indicated that the amplified product was 100% identity with that of unigene 32070, confirming unigene 32070 was a *bona fide* gene in *O. saundersiae* genome. The 1773-bp ORF encoded a polypeptide of 590 amino acids (aa) with a predicted molecular mass of 64.61 kDa and *pI* of 6.16. Blast analysis of the deduced protein revealed its predominant homology with sterol 3 $\beta$ -glucosyltransferase from *Elaeis guineensis* (XP\_010933168.1, 85%), *Musa acuminata* subsp. *Malaccensis* (XP\_009397543.1, 83%) and *Anthurium amnicola* (JAT62313.1, 81%). The cDNA was therefore designated as OsSGT1 and submitted to GenBank library with an accession number of MF688776.

The sequence analyses of OsSGT1 were first assessed with the aim to direct its expression and functional verification. No putative trans-membrane domain was observed in OsSGT1 based on the prediction results by TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>), suggesting OsSGT1 is a cytoplasmic SGT and may be expressed heterologously in *E. coli* in a soluble form.

Multiple alignment of OsSGT1 and other plant SGTs indicated that the middle and C-terminal parts of these SGTs were more conservative than the N-terminal region (Supplementary Information Fig. S1), consistent with previous notion<sup>19</sup>. Moreover, two conservative motifs, namely a putative steroid-binding domain (PSBD) and a plant secondary product glycosyltransferase box (PSPG), were observed in OsSGT1 (Supplementary Information Fig. S1). The region named PSBD located in the middle part of OsSGT1 and was thought to be involved in the binding of steroidal substrates<sup>19</sup>. PSPG box is about 40 aa in length and close to the carboxy-terminus. This box is a characteristic “signature sequence” of UDP glycosyltransferase and deduced to be responsible for the binding of the UDP moiety of the nucleotide sugar<sup>20</sup>. The presence of PSBD and PSPG boxes suggests that OsSGT1 may be involved in secondary metabolism, catalyzing the transfer of UDP-sugars to steroidal substrates thereby forming steroidal glycosides.

The phylogenetic tree based on deduced amino acid sequences of OsSGT1 and other SGT1 was generated by MEGA 7.0. As



**Figure 1** The sugar acceptor promiscuity of OsSGT1 with 10 reactive steroidal substrates. (A) The bars show the maximum percentage conversion of each substrate under identical reaction conditions. These steroidal substrates are listed in descending order of percent conversion. (B) The detailed structures of 10 reactive steroidal substrates. The numbers from 1 to 10 represented cholesterol (1),  $\beta$ -sitosterol (2), ergosterol (3), dehydroepiandrosterone (4), diosgenin (5), digitoxigenin (6), 17 $\alpha$ -hydroxypregnenolone (7), testosterone (8), estradiol (9) and androstenediol (10).

shown in [Supplementary Information Fig. S3](#), all selected SGTs were clustered into four clades, Mon, Di, Ba and Fun clades. The four clades included SGTs from monocots, dicots, bacteria and fungi, respectively. OsSGT1 belonged to Mon clade, suggesting that OsSGT1 was most similar to SGTs from monocots.

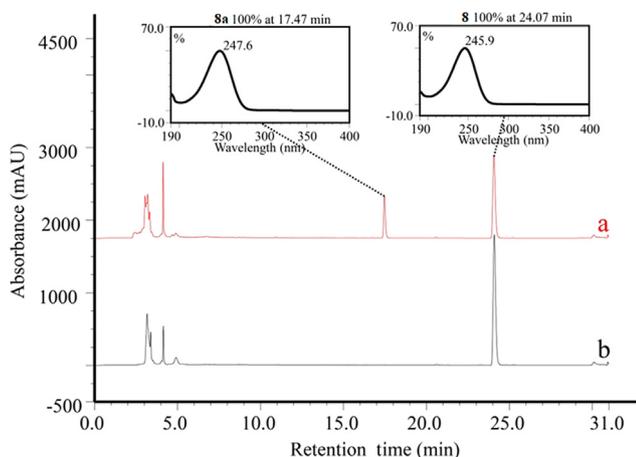
### 2.3. Prokaryotic expression of OsSGT1 in *E. coli*

OsSGT1 was then inserted into pET-28a(+) to yield a recombinant pET28a-OsSGT1 ([Supplementary Information Table S2](#)), which was transformed into *Transsetta*(DE3) (TransGen, Beijing, China) for heterologous expression. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) indicated that most of the expressed OsSGT1 protein was present in the form of insoluble inclusion body, which was regarded to be devoid of bioactivity. It was well known that chaperone proteins were able to assist protein folding and thus increase production of active protein<sup>21</sup>. Therefore, a chaperone plasmid pGro7 (Takara Biotechnology Co., Ltd., Dalian, China) was applied to be co-expressed with pET28aOsSGT1 in BL21(DE3) (TransGen, Beijing, China), facilitating the soluble expression of OsSGT1. As shown in

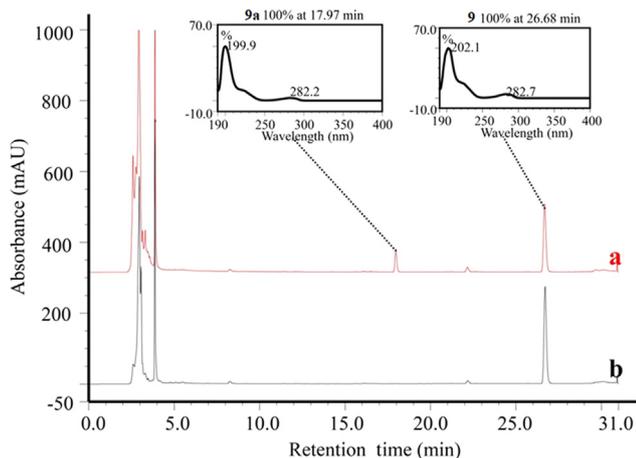
[Supplementary Information Table S2](#), the plasmid pGro7 contains two genes encoding chaperone proteins GroES and GroEL. Under the synergistic action of chaperones GroES and GroEL, an intense band with an apparent molecular mass of 64 kDa was present in the crude extract of BL21(DE3)[pET28aOsSGT1+pGro7], but not in the crude proteins of the control strain BL21(DE3)[pET-28a(+)+pGro7] ([Supplementary Information Fig. S2B](#)). The immunoblot analysis with an anti-polyhistidine tag antibody showed a bound band, but the control extract did not cross-react with the antibody ([Supplementary Information Fig. S2C](#)). These data collectively indicated that OsSGT1 was successfully expressed in *E. coli* in a soluble form ([Supplementary Information Figs. S2B and 2C](#)) in accord with the predicted result of soluble expression of OsSGT1.

### 2.4. Functional characterization of OsSGT1

To identify the activity of OsSGT1, the OsSGT1-containing crude protein was used as the biocatalyst for glycosylation reactions. Each member of the acceptor library (1–49, [Fig. 1B](#) and [Supplementary Information Fig. S4](#)) was first assessed as sugar



**Figure 2** HPLC–DAD analysis of OsSGT1-catalyzed testosterone (**8**) glycosylation. HPLC chromatogram of reaction product of testosterone (**8**) incubated with OsSGT1 protein (a) or without OsSGT1 (b). UV spectra of **8** and enzymatic product **8a** are shown in upper panels. The HPLC conditions are available in [Supplementary Information Table S3](#).



**Figure 3** HPLC–DAD analysis of OsSGT1-catalyzed estradiol (**9**) glycosylation. HPLC chromatogram of reaction product of estradiol (**9**) incubated with OsSGT1 protein (a) or without OsSGT1 (b). UV spectra of **9** and enzymatic product **9a** are shown in upper panels. The HPLC conditions are available in [Supplementary Information Table S3](#).

acceptor for OsSGT1-mediated glycosylation with UDP-Glc as the donor. The glycosyl-transferring reactions were monitored by HPLC–UV/MS analysis ([Supplementary Information Table S3](#)). Of the 49 substrates, only 10 steroids were observed to be glycosylated by OsSGT1, forming corresponding monoglycosides ([Figs. 1–3](#) and [Supplementary Information Figs. S5–14](#)). The ten steroids included seven  $3\beta$ -hydroxysteroids (**1–7**, [Supplementary Information Figs. S5–11](#)), two  $17\beta$ -hydroxysteroids (**8–9**, [Figs. 2](#) and [3](#)) and one  $3\beta,17\beta$ -dihydroxysteroid (**10**, [Supplementary Information Fig. S14](#)). The reaction activities of OsSGT1 towards the ten substrates indicated that OsSGT1 was an SGT showing activities towards both  $3\beta$ - and  $17\beta$ -hydroxysteroids, consistent with the predicted result by bioinformatics analyses ([Supplementary Information Figs. S1](#) and [S3](#)). In fact, the reports of SGTs with glycosylation activity against steroids at positions other than C-3 were limited and only three SGTs from yeast were

**Table 1**  $^1\text{H}$  and  $^{13}\text{C}$  NMR data for **8a** (600 MHz for  $^1\text{H}$  NMR and 150 MHz for  $^{13}\text{C}$  NMR, methanol- $d_4$ ,  $\delta$  in ppm).

Position	$\delta_{\text{C}}$	$\delta_{\text{H}}$
1	36.7	1.73–1.61 (5H, m, H-16b, 15a, 11a, 1, overlap)
2	34.7	2.50 (1H, m, H-2a), 2.39–2.29 (2H, m, H-2b, 6b, overlap)
3	202.4	–
4	124.1	5.73 (1H, s, H-4)
5	175.3	–
6	33.9	2.39–2.29 (2H, m, H-2b, 6b, overlap)
7	32.8	1.91 (1H, dt, $J = 9.4, 7.4\text{Hz}$ , H-7a), 1.08–0.99 (3H, m, H-14, 9, 7b, overlap)
8	36.8	2.11–2.06 (2H, m, H-8, 16a, overlap)
9	55.4	1.08–0.99 (3H, m, H-14, 9, 7b, overlap)
10	40.0	–
11	21.8	1.73–1.61 (5H, m, H-16b, 15a, 11a, 1, overlap), 1.53 (1H, dd, $J = 13.2, 4.1\text{Hz}$ , H-11b)
12	38.5	2.04 (1H, d, $J = 3.2\text{Hz}$ , H = 12a), 1.28–1.23 (4H, m, H-19, 12b)
13	44.2	–
14	51.7	1.08–0.99 (3H, m, H-14, 9, 7b, overlap)
15	24.2	1.73–1.61 (5H, m, H-16b, 15a, 11a, 1, overlap), 1.36–1.33 (1H, m, H-15b)
16	29.8	2.11–2.06 (2H, m, H-8, 16a, overlap), 1.73–1.61 (5H, m, H-16b, 15a, 11a, 1, overlap)
17	89.6	3.78 (1H, $J = 8.6\text{Hz}$ , H-17)
18	12.0	0.93 (3H, m, H-18)
19	17.7	1.28–1.23 (4H, m, H-19, 12b)
1'	104.7	3.35 (1H, d, $J = 7.8\text{Hz}$ , H-1')
2'	75.4	3.68 (1H, m, H-2')
3'	77.9	3.18 (1H, dd, $J = 9.0, 7.9\text{Hz}$ , H-3')
4'	71.7	3.30 (1H, t, $J = 9.1\text{Hz}$ , H-4')
5'	78.2	3.36 (1H, dd, $J = 9.0, 5.5\text{Hz}$ , H-5')
6'	62.8	4.35 (1H, d, $J = 7.8\text{Hz}$ , H-6'a), 3.86 (1H, dt, $J = 8.2, 4.1\text{Hz}$ , H = 6'b)

–: not applicable.

verified to exhibit selectivity towards both  $3\beta$ - and  $17\beta$ -hydroxysteroids<sup>17</sup>. OsSGT1 was therefore viewed as the first plant SGT with selectivity towards both  $3\beta$ - and  $17\beta$ -hydroxysteroids ([Figs. 2](#) and [3](#) and [Supplementary Information Figs. S5–14](#)). However, the glycosylation activity of OsSGT1 towards  $17\beta$ -hydroxyl group would be lost if additional hydroxyl group at  $2\beta$ - ( $2\beta$ -OH-testosterone, **13**),  $15\beta$ - ( $15\beta$ -OH-testosterone, **14**),  $16\beta$ - ( $16\beta$ -OH-testosterone, **15**), or  $16\alpha$ -position ( $16\alpha$ -OH-testosterone, **16**), even a methyl group at C17-position (methyltestosterone and its derivatives, **17–20**) was attached to testosterone (**8**), generating not any glycosylated products. Moreover, OsSGT1 has no activity towards other compounds, including steroids without  $3\beta$ - and  $17\beta$ -hydroxyl groups (**11–12** and **21–24**), flavonoids (**25–31**), alkaloids (**32–38**), triterpenoids (**39–42**), phenolic acids (**43–47**) and coumarins (**48–49**) as shown in [Supplementary Information Fig. S4](#).

Among 10 reactive  $3\beta$ - and  $17\beta$ -hydroxysteroids, dehydroepiandrosterone (**4**) had a maximum conversion approaching 100%, followed by diosgenin (**5**) with 82% conversion and the other compounds having conversion below 80% ([Fig. 1A](#)).

To produce sufficient glycosylated products for structural characterization, scale-up of OsSGT1-mediated reactions to preparative scale (6 mL) was conducted. The resultant glycosides

were prepared by HPLC and subjected to NMR analysis for structural elucidation. To determine the glycosylation sites of **1a–6a**, **8a** and **9a**, the  $^{13}\text{C}$  NMR analyses of the corresponding aglycons **1–6**, **8** and **9** were also performed (Supplementary Information Figs. S15–22 and Table S4). The  $^{13}\text{C}$  NMR glycosylation shifts ( $\Delta\delta$ ,  $\delta_{\text{glucoside}} - \delta_{\text{aglycon}}$ ) of these glycosides were thus examined to ascertain the glycosylation position (Table 1, and Supplementary Information Tables S4–11). The steroidal glycosides were observed to have significant glycosylation shift  $\Delta\delta$ s for C-3 (glycosides **1–6**) or C-17 position (compounds **8–9**), showing their 3- or 17-glycosides. For **7a** and **10a**, the location of glucose group was determined to be at C-3 based on their HMBC correlations between H-1' and C-3 (Supplementary Information Figs. S23 and 24). The  $\beta$ -anomeric configuration of the D-glucose unit in these ten glucosides (**1a–10a**) was determined from the large anomeric proton-coupling constants of H-1' ( $J = 7.8$  Hz) (Table 1 and Supplementary Information Tables S5–13). The structures of these glucosides were thus assigned to  $3\beta$ -glucosides (**1a–7a** and **10a**) or  $17\beta$ -glucosides (**8a** and **9a**) of steroids based on  $^1\text{H}$  NMR (**1a–10a**) and  $^{13}\text{C}$  NMR (**1a–10a**) signals, HSQC (**1a–2a**, **4a–7a**, **9a–10a**), HMBC (**7a–8a**, **10a**) and DEPT (**7a**, **10a**) spectra (Table 1, Supplementary Information Figs. S23–45 and Tables S5–13). These data collectively showed that OsSGT1 was an inverting-type glycosyltransferase.

In the preparation of T-17 $\beta$ -G (**8a**), when the concentrated reaction mixture was separated by reversed phase high-performance liquid chromatography (RP-HPLC), we accidentally discovered that in addition to the major peak representing T-17 $\beta$ -G (**8a**), a minor peak (**8b**) was also present in the HPLC profile (Supplementary Information Fig. S46A). The minor product with a  $t_{\text{R}}$  of 20.53 min was also subjected to LC-MS analysis. Surprisingly, the  $[\text{M} + \text{H}]^+$  value of the minor product was assigned to 493.3, 42 more than that of monoglycosylated testosterone (Supplementary Information Fig. S46B). This finding hints that the minor product may be an acetylated testosterone glucoside. To characterize the exact structure of **8b**, the minor product was prepared in bulk for NMR experiment (Supplementary Information Figs. S47–51). Details of  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra were tabulated in Table 2. The minor product was thus identified as 6'-acetylated testosterone 17-O- $\beta$ -glucoside (6'-AT-17 $\beta$ -G, **8b**).

To test if the acetylated product **8b** was from glucoside **8a**, the purified glucoside **8a** was used as the substrate to incubate with crude extract of *E. coli* expressing pET-28a(+) or pET28OsSGT1, respectively. In both conditions, we observed the presence of **8b** (Supplementary Information Fig. S52). On the contrary, no acetylated product **8b** was detected in the *E. coli* lysate without the addition of substrate **8a** (Supplementary Information Fig. S52). We therefore inferred that testosterone (**8**) was first glycosylated at the 17 $\beta$ -hydroxyl group by OsSGT1 to form T-17 $\beta$ -G (**8a**), which was then selectively acetylated at C-6' of sugar moiety to yield the 6'-AT-17 $\beta$ -G (**8b**) by a soluble bacterial acetyltransferase (Supplementary Information Fig. S52).

Likewise, two metabolites, E-17 $\beta$ -G (**9a**) and 6'-acetylated estradiol 17-O- $\beta$ -glucoside (6'-AE-17 $\beta$ -G, **9b**), were detected in the concentrated OsSGT1-catalyzed reaction mixture of estradiol (**9**) as shown in Supplementary Information Figs. S53–55, Tables S11 and S14. These data collectively revealed that *E. coli* cell contained at least one SGAs specific for the acetylation of steroidal 17 $\beta$ -glycosides.

Moreover, the sugar donor promiscuity of OsSGT1-catalyzed glycosylation reactions was also investigated.  $\beta$ -Sitosterol (**2**) and testosterone (**8**) were chosen as the sugar acceptors to react with

**Table 2**  $^1\text{H}$  and  $^{13}\text{C}$  NMR data for **8b** (600 MHz for  $^1\text{H}$  NMR and 150 MHz for  $^{13}\text{C}$  NMR, methanol- $d_4$ ,  $\delta$  in ppm).

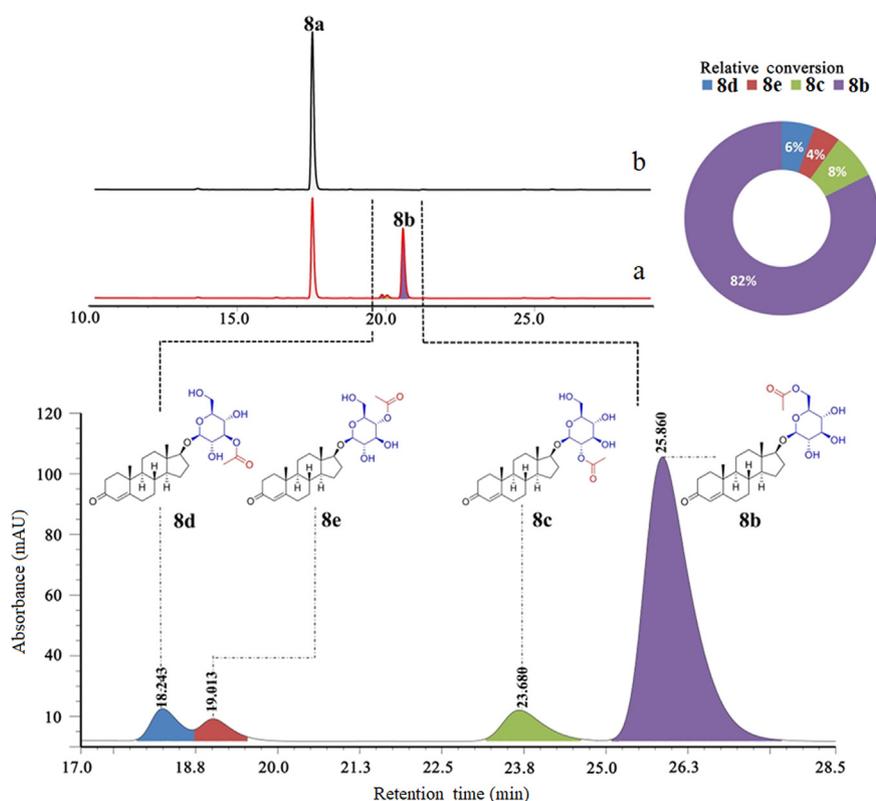
Position	$\delta_{\text{C}}$	$\delta_{\text{H}}$
1	36.7	1.66–1.60 (5H, m, H-16b, 15a, 11a, 1 overlap)
2	34.7	2.49–2.46 (2H, m, H-2a, 6a overlap) 2.30–2.27 (2H, m, H-2b, 6b overlap)
3	202.4	–
4	124.1	5.71 (1H, s, H-4)
5	175.2	–
6	33.9	2.49–2.46 (2H, m, H-2a, 6a overlap) 2.30–2.27 (2H, m, H-2b, 6b overlap)
7	32.8	1.89 (1H, m, H-7a) 1.04–0.97 (3H, m, H-14, 9, 7b, overlap)
8	36.8	2.08 (1H, m, H-8)
9	55.5	1.04–0.97 (3H, m, H-14, 9, 7b, overlap)
10	40	–
11	21.7	1.66–1.60 (5H, m, H-16b, 15a, 11a, 1 overlap) 1.50 (1H, m, H-11b)
12	38.4	2.03–2.01 (H-16a, 12a overlap)
13	44.1	–
14	51.6	1.04–0.97 (3H, m, H-14, 9, 7b, overlap)
15	24.3	1.66–1.60 (5H, m, H-16b, 15a, 11a, 1 overlap) 1.39 (1H, s, H-15b)
16	30	2.03–2.01 (H-16a, 12a overlap) 1.66–1.60 (5H, m, H-16b, 15a, 11a, 1 overlap)
17	90.3	3.65 (1H, d, $J = 8.1$ Hz, H-17)
18	12	0.89 (3H, s, H-18)
19	17.7	1.24–1.17 (4H, s, H-19, 12b overlap)
1'	104.9	4.31 (1H, d, $J = 7.7$ Hz, H-1')
2'	75.1	3.41 (1H, m, H-2')
3'	78	3.16 (1H, m, H-3')
4'	71.7	3.27 (1H, m, H-4')
5'	75.3	3.33 (1H, m, H-5')
6'	64.8	4.36 (1H, dd, $J = 11.8, 2.1$ Hz, H-6'a) 4.20 (1H, dd, $J = 11.8, 6.1$ Hz, H-6'b)
1''	172.7	–
2''	20.8	2.06 (2H, s, H-2'')

–: not applicable.

varied sugar donors listed in the Supplementary Experimental Section, respectively. Results demonstrated that both  $\beta$ -sitosterol (**2**) and testosterone (**8**) had no reactive activity towards other UDP-activated nucleotides except UDP-Glc under the action of OsSGT1, indicating OsSGT1 was specific for UDP-Glc.

## 2.5. Sequence isolation of genes encoding bacterial SGAs

To characterize the genes encoding SGAs, the first task was to analyze the genome sequence of BL21(DE3), which was public in NCBI database (Accession No. CP001509.3). This bacterial strain contains at least 35 putative acetyltransferase genes, in which *lacA*, *maa* and *wecH*, were predicted to encode O-acetyltransferase (Supplementary Information Table S15). As shown in Supplementary Information Fig. S56, further sequence analyses revealed WecH protein was a membrane-bound protein with a total of 10 membrane-spanning helices, inconsistent with the above results, in which the candidate acetyltransferase was determined to be a soluble protein in bacterial lysate. LacA and Maa proteins were predicted to have no transmembrane helices, suggesting their soluble form in bacterial. Thus, the remaining two genes, *lacA* and *maa*, were further investigated. First, the entire ORFs of the two



**Figure 4** HPLC profile of EcSGA1-catalyzed acetylation of **8a**. Upper panel, HPLC chromatogram of reaction product of **8a** incubated with purified EcSGA1 (a) or without EcSGA1 (b) using method E (Supplementary Information Table S3). Lower panel, HPLC profile of acetylated products of **8a** separated by chromatographic method I (Supplementary Information Table S3).

genes were isolated from bacterial genome using gene-specific primers (Supplementary Information Fig. S57A and Table S1). The ORFs of *lacA* and *maa* genes were 612 and 552 bp encoding polypeptides of 203 and 183 aa, respectively. The predicted molecular weights of the two proteins were 22.77 and 20.1 kDa. The two genes were then inserted into pET-28a(+) to generate two recombinant vectors, which were introduced into BL21(DE3) for heterologous expression. After isopropyl- $\beta$ -D-thiogalactoside (IPTG) induction, the accumulation of approximately 23 or 20 kDa was observed in the lysate of bacterial strain harboring pET28a-*lacA* or pET28a-*maa* (Supplementary Information Fig. S57B). Moreover, the presence of bacterially-expressed His-LacA or His-Maa fusion protein in the bacterial lysate was verified by Western-blot with anti-His antibody (Supplementary Information Fig. S57C). The expressed LacA or Maa protein was then purified to near homogeneity by affinity chromatography (Supplementary Information Fig. S57B).

## 2.6. Enzymatic activity characterization of EcSGA1

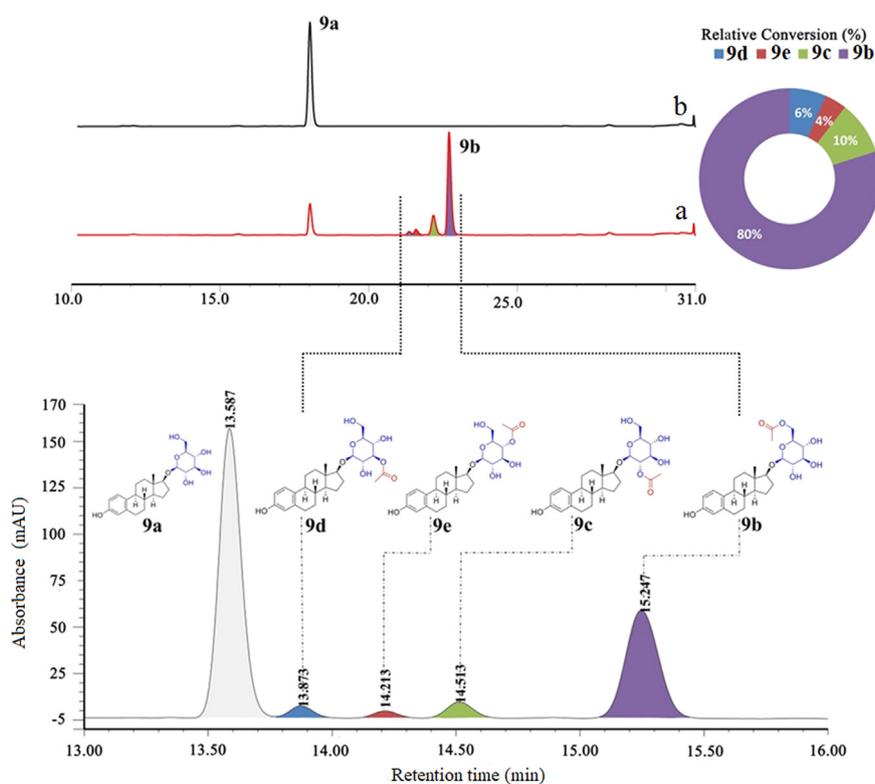
The purified His-LacA or His-Maa fusion protein was used as the biocatalyst to react with T-17 $\beta$ -G (**8a**) and acetyl-CoA. The reactions were monitored by HPLC-UV/MS analysis using the method E (Supplementary Information Table S3). As shown in Fig. 4, 6'-AT-17 $\beta$ -G (**8b**) was detected in LacA-catalyzed bio-conversion of T-17 $\beta$ -G (**8a**), attesting *lacA* encoded a SGA (Fig. 4, upper panel). On the contrary, there were not any new products in Maa-mediated reaction. LacA was thus designated as EcSGA1 (*E. coli* steroidal glycoside acetyltransferase) for convenience hereinafter and submitted to GenBank with an accession number of

MF688777. It is generally accepted that hydrolases and acyltransferases are two classes of enzymes responsible for acylation reactions of SGs<sup>22</sup>. The enzymatic acylations reported now are largely performed by hydrolases like lipases<sup>22</sup>. On the other hand, not any genes encoding SGAs are isolated up to date<sup>16,22</sup>. EcSGA1 is therefore regarded as the first steroidal glycoside acyltransferase catalyzing the attachment of acyl groups into the hydroxyl groups of steroidal 17 $\beta$ -glycosides, to our knowledge.

Also, EcSGA1 was observed to catalyze another steroidal 17 $\beta$ -glycoside, E-17 $\beta$ -G (**9a**), to form corresponding acylate (**9b**, Fig. 5, upper panel). On the other hand, the other glucosides listed in Supplementary Information Fig. S58 could not be acetylated by EcSGA1, testifying EcSGA1 was specific to steroidal 17 $\beta$ -glycosides.

Moreover, the acyl donor promiscuity of EcSGA1 was investigated. T-17 $\beta$ -G (**8a**) or E-17 $\beta$ -G (**9a**) was used as the acyl acceptor to react with different acyl donors (acetyl-CoA, succinyl-CoA, arachidonoyl-CoA, palmitoyl-CoA and acetoacetyl-CoA) under the action of the purified EcSGA1. Results manifested that neither T-17 $\beta$ -G (**8a**) nor E-17 $\beta$ -G (**9a**) could react with these acyl donors except acetyl-CoA, indicating that EcSGA1 had strict donor selectivity.

After careful check of EcSGA1-catalyzed reaction mixture in HPLC profile, we have found several other minor peaks adjacent to the major product **8b** (Fig. 4, upper panel). These minor peaks are so close that we could not distinguish. Therefore, an efficient HPLC method, namely method I (Supplementary Information Table S3), was developed to separate these peaks. As shown in Fig. 4 (lower panel), besides the major product **8b** ( $t_R = 25.860$  min), we observed three other minor peaks at  $t_R = 18.243$ , 19.013



**Figure 5** HPLC profile of EcSGA1-catalyzed acetylation of **9a**. Upper panel, HPLC chromatogram of reaction product of **9a** incubated with purified EcSGA1 (a) or without EcSGA1 (b) using method D (Supplementary Information Table S3). Lower panel, HPLC profile of acetylated products of **9a** separated by chromatographic method (Supplementary Information Table S3).

and 23.680 min, respectively. The LC–MS measurement of these minor peaks showed that all of them have an  $[M + Na]^+$  value of 515.3, thus suggesting their monoacetylated testosterone glucosides (Supplementary Information Fig. S59).

Likewise, E-17 $\beta$ -G (**9a**,  $t_R = 13.587$  min) was observed to form four acetylated glucosides using purified EcSGA1 as the biocatalyst (Fig. 5). Besides the well-characterized 6'-AE-17 $\beta$ -G (**9b**,  $t_R = 15.247$  min), the other three products were determined to be monoacetylated estradiol glucosides based on their MS data (Supplementary Information Fig. S60). It was assumed that EcSGA1 could introduce an acyl group into different hydroxyl groups of steroidal 17 $\beta$ -glycosides, generating varied monoacetylated products (Figs. 4 and 5).

### 2.7. Enzymatic two-step synthesis of AT-17 $\beta$ -Gs from testosterone

To obtain sufficient amount of monoacetylated testosterone glucosides for structural characterization and further cytotoxicity assay, an enzymatic two-step process for AT-17 $\beta$ -Gs (**8b–8e**) was developed (Scheme 1).

Firstly, the whole cell biotransformation for the formation of AT-17 $\beta$ -Gs (**8b–8e**) was exploited due to its simple catalyst preparation. When testosterone (**8**) was incubated with the engineered strain BL21(DE3)[pET28a-OsSGT1+pGro7], not any new products were detected. On the other hand, when T-17 $\beta$ -G (**8a**) was added into the same whole-cell system, 6'-AT-17 $\beta$ -G (**8b**) was present in the reaction mixture (Supplementary Information Fig. S61). These data indicated that testosterone (**8**) could not be transported into the cell while the glycosylation of testosterone (**8**) significantly improved the intercellular transport.

Thus, the formation of AT-17 $\beta$ -Gs (**8b–8e**) from testosterone (**8**) using the single whole-cell biocatalyst is infeasible. A two-step process is therefore established to address this limitation. Specifically, OsSGT1-catalyzed reaction was performed in the membrane-free crude cell extract of BL21 (DE3)[pET28a-OsSGT1+pGro7], while EcSGA1-mediated acetylation was conducted in the whole-cell system of BL21(DE3)[pET28a-EsSGA1].

The optimal pH and temperature of OsSGT1-catalyzed reaction using the cell-free extract of BL21(DE3)[pET28a-OsSGT1+pGro7] as the biocatalyst were first determined to be alkaline pH value of 11 and 50 °C, respectively (Supplementary Information Fig. S62). Next, the 100  $\mu$ L screening scale of OsSGT1-catalyzed glycosylation reaction was scaled to 53 mL scale, in which 152 mg testosterone (**8**) were glycosylated to form 61 mg T-17 $\beta$ -G (**8a**) under optimized conditions (Scheme 1). The resultant T-17 $\beta$ -G (**8a**) was subsequently used as the substrate applied in the scale-up of the whole-cell system of BL21(DE3)[pET28a-EsSGA1] (135 mL) under optimized pH 5.0 and 40 °C. The resulting reaction mixture was subjected to high-performance liquid chromatography–solid phase extraction–NMR spectroscopy (HPLC–SPE–NMR) measurement. Comparison of the  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra of **8c–8e** with those of **8b** suggested that compounds **8c–8e** had the same framework as **8b** and the structural difference might be the position of the acetyl group. The location of acetyl group was determined to be at C-2' based on the HMBC correlations between H-2' ( $\delta$  4.68) and C-1" ( $\delta$  170.3) as shown in Supplementary Information Fig. S63. Thus, compound **8c** was assigned as 2'-AT-17 $\beta$ -G. The isolated glucose proton at  $\delta$  4.91 (H-3') of compound **8d** exhibited long-range correlations with carbonyl carbons at  $\delta$  172.7 (Supplementary Information Fig. S64). Moreover, H-4' ( $\delta$  4.73) of compound **8e** showed long-range

**Table 3**  $^1\text{H}$  and  $^{13}\text{C}$  NMR data for **8c** (600 MHz for  $^1\text{H}$  NMR and 150 MHz for  $^{13}\text{C}$  NMR, methanol- $d_4$ ,  $\delta$  in ppm).

Position	$\delta_{\text{C}}$	$\delta_{\text{H}}$
1	36.7	2.09–2.03 (5H, m, H-2'', 16a, 1a overlap) 1.70–1.61 (5H, m, H-8, 11a, 15a, 16b, 8, 1b overlap)
2	34.7	2.50–2.45 (2H, m, H-2a, 6a overlap) 2.31–2.27 (2H, m, H-2b, 6b overlap)
3	202.3	–
4	124.1	5.71 (1H, s, H-4)
5	175.1	–
6	33.9	2.50–2.45 (2H, m, H-2a, 6a overlap) 2.31–2.27 (2H, m, H-2b, 6b overlap)
7	32.8	1.87–1.84 (2H, m, H-7a, 12a overlap) 1.03–0.96 (3H, m, H-7b, 14, 9 overlap)
8	36.8	1.70–1.61 (5H, m, H-8, 11a, 15a, 16b, 1b overlap)
9	55.4	1.03–0.96 (3H, m, H-7b, 14, 9 overlap)
10	40	–
11	21.7	1.70–1.61 (5H, m, H-8, 11a, 15a, 16b, 1b overlap), 1.50 (1H, m, H-11b)
12	38.3	1.87–1.84 (2H, m, H-7a, 12a overlap) 1.24–1.22 (4H, m, H-19, 12b overlap)
13	44	–
14	51.4	1.03–0.96 (3H, m, H-7b, 14, 9 overlap)
15	24.2	1.70–1.61 (5H, m, H-8, 11a, 15a, 16b, 1b overlap), 1.32 (1H, m, H-15b)
16	29.7	2.09–2.03 (5H, m, H-2'', 16a, 1a overlap) 1.70–1.61 (5H, m, H-8, 11a, 15a, 16b, 1b overlap)
17	90.6	3.69–3.63 (2H, m, H-6'b, 17 overlap)
18	12	0.78 (3H, m, H-18)
19	17.7	1.24–1.22 (4H, m, H-19, 12b overlap)
1'	103	4.46 (1H, d, $J = 8.0\text{Hz}$ , H-1')
2'	75.7	4.68 (1H, dd, $J = 9.6, 8.1\text{Hz}$ , H-2')
3'	76.1	3.49 (1H, m, H-3')
4'	71.6	3.36 (1H, m, H-4')
5'	78	3.27–3.24 (1H, m, H-5')
6'	62.6	3.86 (1H, dd, $J = 11.9, 2.3\text{Hz}$ , H-6'a) 3.69–3.63 (2H, m, H-6'b, 17 overlap)
1''	171.7	–
2''	21.2	2.09–2.03 (5H, m, H-2'', 16a, 1a overlap)

–: not applicable.

correlations with C-1" ( $\delta$  170.7), as revealed by the HMBC spectrum (Supplementary Information Fig. S65). These data supported that the structure of **8d** and **8e** was elucidated as 3'-AT-17 $\beta$ -G and 4'-AT-17 $\beta$ -G, respectively. Hence, the three trace products at  $t_{\text{R}} = 18.243$ , 19.013 and 23.680 min were thus assigned to be 3'-(**8d**), 4'-(**8e**) and 2'-AT-17 $\beta$ -G (**8c**) based on their respective NMR data (Tables 3–5 and Supplementary Information Figs. S63–77). These data indicate that EcSGA1 can effectively introduce the acetyl group into the primary hydroxyl group and each secondary hydroxyl group of T-17 $\beta$ -G (**8a**), yielding four monoacylates without the formation of diacylates (Fig. 4 and Scheme 1). Because the primary C(6')-OH was the most reactive of the four hydroxyl groups in T-17 $\beta$ -G (**8a**), acetylation of T-17 $\beta$ -G (**8a**) took place preferentially at the C(6')-OH, giving 6'-O-acylate predominantly in 82% yield (Fig. 4 and Scheme 1). Also, EcSGA1 can regioselectively acetylate each secondary hydroxyl of T-17 $\beta$ -G (**8a**) in the presence of the primary hydroxyl group, giving 2'-(**8c**), 3'-(**8d**) and 4'-AT-17 $\beta$ -G (**8e**) in 8%, 6% and 4% yield, respectively. These data revealed the reactivity trend of hydroxyls is 6'-OH >> 2'-OH > 3'-OH > 4'-OH.

Likewise, the formation of four monoacylates was also present in EcSGA1-catalyzed acylation of E-17 $\beta$ -G (**9a**,  $t_{\text{R}} = 13.587$  min, Fig. 5). In addition to the well-characterized major product **9b**, there are three trace products **9c–9e**. Because of their trace amount, we did not further enrich these monoacetylated estradiol glucosides for NMR analysis. However, according to the catalytic behavior of EcSGA1 towards T-17 $\beta$ -G (**8a**), it was easy to infer that these products were most likely 2'-(**9c**,  $t_{\text{R}} = 14.513$  min), 3'-(**9d**,  $t_{\text{R}} = 13.873$  min) and 4'-AE-17 $\beta$ -G (**9e**,  $t_{\text{R}} = 14.213$  min, Fig. 5). The order of reactivity of the hydroxyls was determined as 6'-OH >> 2'-OH > 3'-OH > 4'-OH with a yield ratio of 80:10:6:4 (Fig. 5). Regioselective acylation of one of the multiple hydroxyl groups in SGs is the major obstacle to the synthesis of SGEs and direct methods for site-selective acylation of unprotected SGs have rarely been documented. In this contribution, we successfully achieved the regioselective acylation of fully unprotected SGs using EcSGA1 as the biocatalyst, thereby leading to an extremely short-step synthesis of ASGs.

### 2.8. Cytotoxic activity of acetylated steroidal glycosides

Acetylated steroidal glycosides, namely **8b**, **8c**, **8d** and **8e**, together with **9b** were tested for their *in vitro* cytotoxicity against seven human cancer cell lines including HCT116, Bel7402, MGC803, Capan2, NCI-H460, NCI-H1650 and A549. The results indicated that 3'-AT-17 $\beta$ -G (**8d**) exhibited a wide spectrum of cytotoxic activities against the tested cell lines (Table 6). 6'-AT-17 $\beta$ -G (**8b**) displayed much less cytotoxicity than 3'-AT-17 $\beta$ -G (**8d**) but showed a mild cytotoxicity against human non-small cell lung carcinoma cell line NCI-H1650 with  $\text{IC}_{50}$  values of 26.5  $\mu\text{mol/L}$  (Table 6). On the contrary, the control T-17 $\beta$ -G (**8a**) did not display significant cytotoxicity towards these tested cell lines ( $\text{IC}_{50} > 50.0$   $\mu\text{mol/L}$ ). These evidences revealed that the acyl groups of SGEs are of importance to their cytotoxicity and direct regioselective acylation of SGs is thus believed as a powerful tool for the discovery of drug candidates.

## 3. Discussion

Acylated steroidal glycosides have attracted our attentions primarily due to their biological and pharmacological significances<sup>2,10,11,23</sup>. There are two enzymes, namely SGTs and SGAs, responsible for the biosynthesis of ASGs. To synthesize ASGs, the primary premise is to obtain glycosyltransferases capable of catalyzing the formation of SGs from the abundantly available free steroids. *O. saundersiae* is thus selected as the candidate plant for SGTs isolation. *O. saundersiae* is a monocotyledonous plant rich in acylated steroidal glycosides, suggesting that it may contain SGTs and SGAs responsible for the biosynthesis of ASGs<sup>23</sup>. Thus, the transcriptome of *O. saundersiae* was sequenced with the aim to facilitate the genes discovery. OsSGT1 was then isolated from *O. saundersiae* based on the RNA-Seq data. Subsequently, OsSGT1-containing cell-free extract was used as the biocatalyst for glycosylations of 49 structurally diverse drug-like scaffolds. The use of cell-free extract offers a number of advantages. Unlike the ambitious purification procedures, the preparation of cell-free extract was simple and time-saving. Moreover, compared the purified enzymes, the recombinant proteins used in crude extract-based system were more stable.

**Table 4**  $^1\text{H}$  and  $^{13}\text{C}$  NMR data for **8d** (600 MHz for  $^1\text{H}$  NMR and 150 MHz for  $^{13}\text{C}$  NMR, methanol- $d_4$ ,  $\delta$  in ppm).

Position	$\delta_{\text{C}}$	$\delta_{\text{H}}$
1	36.8	2.11–2.03 (6H, m, H-2'', 1a, 12a, 16a overlap), 1.71–1.60 (5H, m, H-15a, 11a, 1b, 16b, 8 overlap)
2	34.7	2.51–2.45 (2H, m, H-2a, 6a overlap)
3	202.4	–
4	124.1	5.71 (1H, s, H-4), 3.44 (1H, m, H-4')
5	175.3	–
6	33.9	2.51–2.45 (2H, m, H-2a, 6a overlap), 2.32–2.28 (2H, m, H-2b, 6b overlap)
7	32.8	1.89 (1H, d, $J = 12.6\text{Hz}$ , H-7a), 1.06–0.95 (3H, m, H-7b, 14, 9 overlap)
8	36.7	1.71–1.60 (5H, m, H-15a, 11a, 1b, 16b, 8 overlap)
9	55.4	1.06–0.95 (3H, m, H-7b, 14, 9 overlap)
10	40	–
11	21.7	1.71–1.60 (5H, m, H-15a, 11a, 1b, 16b, 8 overlap), 1.51 (1H, dd, $J = 12.9, 3.8\text{Hz}$ , H-11b)
12	38.4	2.11–2.03 (6H, m, H-2'', 1a, 12a, 16a overlap), 1.24–1.21 (4H, m, H-19, 12b overlap)
13	44.2	–
14	51.7	1.06–0.95 (3H, m, H-7b, 14, 9 overlap)
15	24.2	1.71–1.60 (5H, m, H-15a, 11a, 1b, 16b, 8 overlap), 1.32 (1H, m, H-15b),
16	29.8	2.11–2.03 (6H, m, H-2'', 1a, 12a, 16a overlap), 1.71–1.60 (5H, m, H-15a, 11a, 1b, 16b, 8 overlap)
17	89.8	3.77 (1H, m, H-17)
18	12	0.89 (3H, s, H-18)
19	17.7	1.24–1.21 (4H, m, H-19, 12b overlap)
1'	104.6	4.42 (1H, d, $J = 7.8\text{Hz}$ , H-1')
2'	73.7	3.29–3.26 (2H, m, H-5', 2' overlap)
3'	79.2	4.91 (1H, d, $J = 9.5\text{Hz}$ , H-3')
4'	69.8	3.44 (1H, m, H-4')
5'	77.6	3.29–3.26 (2H, m, H-5', 2' overlap)
6'	62.4	3.85 (1H, m, H-6'a), 3.67 (1H, m, H-6'b)
1''	172.7	–
2''	21.1	2.11–2.03 (6H, m, H-2'', 1a, 12a, 16a overlap)

–: not applicable.

Steroidal glycosides are one of the main sources of innovative drugs<sup>17</sup>. SGT-catalyzed glycodiversification of steroids could expand the molecular diversification, thereby facilitating the discovery of pharmacological leads. Thus, the search of SGTs with catalytic promiscuity may provide potent biocatalysts for glycodiversification. Therefore, a library containing 49 structural diverse drug-like molecules was utilized to react with the recombinant OsSGT1 with the aim to explore the substrate flexibility of OsSGT1. *In vitro* enzymatic analyses revealed that OsSGT1 was active against various steroids, including physterols (**1–3**, Supplementary Information Figs. S5–7), steroid hormones (**4**, **7–10**, Figs. 2 and 3, and Supplementary Information Figs. S8, S11 and S14), steroidal sapogenin (**5**, Supplementary Information Fig. S9) and cardiac aglycon (**6**, Supplementary Information Fig. S10), exhibiting a wider substrate range than that of previously identified SGTs from plant<sup>18,24</sup>.

To investigate the regioselectivity of OsSGT1, diversified steroids (**1–24**) were selected as the sugar acceptors for OsSGT1-catalyzed glycosylations. As illustrated in Figs. 1–3 and Supplementary Information S5–14, OsSGT1 specifically attacked the hydroxyl groups at C-3 and C-17 positions, but no activities towards hydroxyl groups at C-2 (**13**, **18**), C-7 (**22**), C-11 (**23**), C-12 (**22**), C-14(**6**), C-15 (**14**), C-16 (**15**, **16** and **19**) and C-21 (**23** and **24**). When steroids having two potentially reactive hydroxyl groups, like  $17\alpha$ -hydroxypregnenolone (**7**) or androstenediol (**10**), were used as the substrate for OsSGT1-assisted glycosylation, only glycosides with a glycosyl substituent in C-3 position were detected in the reaction mixture, suggesting OsSGT1 exhibited prominent regioselectivity towards the 3-OH of both substrates (Supplementary Information Figs. S11 and S14). Also, OsSGT1

could catalyze the attachment of a sugar moiety to the hydroxyl group at C-17 position. However, the hydroxyl groups with the stereo-configuration at other positions, like  $2\beta$ - (**13** and **18**),  $15\beta$ - (**14**),  $16\beta$ - (**15** and **19**) and  $16\alpha$ -positions (**16** and **20**), would inhibit this attachment of sugar moieties to C-17OH, resulting in no yields of steroidal glycosides.

The stereoselectivity of OsSGT1 was also assessed in this study. Estradiol (**9**) and  $\alpha$ -estradiol (**12**) differ for the configuration of the hydroxyl group at C-17 position. When each of the two compounds was used to react with OsSGT1, only  $\beta$ -configured glycosides were generated (Fig. 3). Likewise, OsSGT1 showed  $\beta$ -selective glycosylation towards the hydroxyl group at C-3 position. Cumulatively, these evidences revealed that OsSGT1-catalyzed glycosylations were conducted in a region-and stereo-selective fashion.

One of the most striking findings of this study is the characterization of bacterial LacA protein as a steroidal glycoside acyltransferase. It is well known that *lacA* is one of three structural genes (*lacZ*, *lacY* and *lacA*) in *lac* operon<sup>25,26</sup>. The function of LacZ and LacY is well-characterized<sup>25,26</sup>. LacZ encodes a  $\beta$ -galactosidase, catalyzing the cleavage of lactose into glucose and galactose. LacY encodes a lactose permease responsible for lactose uptake<sup>25,26</sup>. The third structural protein encoded by *lacA* gene in *lac* operon was initially inferred to be an acetyltransferase. The exact action of this protein, however, remains in doubt until now. In this investigation, in an effort to identify the function of OsSGT1, we unexpectedly characterized LacA protein from *E. coli* as a SGA. *In vitro* enzymatic analyses revealed that LacA protein could specifically catalyze the attachment of acyl groups into the hydroxyl groups of sugar moieties of steroidal  $17\beta$ -

**Table 5**  $^1\text{H}$  and  $^{13}\text{C}$  NMR data for **8e** (600 MHz for  $^1\text{H}$  NMR and 150 MHz for  $^{13}\text{C}$  NMR, methanol- $d_4$ ,  $\delta$  in ppm).

Position	$\delta_{\text{C}}$	$\delta_{\text{H}}$
1	36.8	2.10–2.03 (6H, m, H-2'', 16a, 1a, 12a overlap), 1.69–1.50 (5H, H-11a, 15a, 16b, 1b, 8 overlap)
2	34.7	2.51–2.45 (2H, m, H-6a, 2a overlap) 2.32–2.27 (2H, m, H-6b, 2b overlap)
3	202.4	–
4	124.1	5.71 (1H, s, H-4)
5	175.2	–
6	33.9	2.51–2.45 (2H, m, H-6a, 2a overlap) 2.32–2.27 (2H, m, H-6b, 2b overlap)
7	32.8	1.88 (1H, m, H-7a) 1.03–0.94 (3H, m, H-7b, 14, 9 overlap)
8	36.7	1.69–1.50 (5H, H-11a, 15a, 16b, 1b, 8 overlap)
9	55.5	1.03–0.94 (3H, m, H-7b, 14, 9 overlap)
10	40.0	–
11	21.8	1.69–1.50 (5H, H-11a, 15a, 16b, 1b, 8 overlap) 1.51 (1H, d, $J = 13.2\text{Hz}$ , H-11b)
12	38.4	2.10–2.03 (6H, m, H-2'', 16a, 1a, 12a overlap) 1.24–1.23 (4H, m, H-19, 12b overlap)
13	44.2	–
14	51.7	1.03–0.94 (3H, m, H-7b, 14, 9 overlap)
15	24.2	1.69–1.50 (5H, H-11a, 15a, 16b, 1b, 8 overlap) 1.34–1.29 (1H, m, H-15b)
16	29.8	2.10–2.03 (6H, m, H-2'', 16a, 1a, 12a overlap) 1.69–1.50 (5H, H-11a, 15a, 16b, 1b, 8 overlap)
17	89.8	3.75 (1H, m, H-17)
18	12.0	0.90 (3H, s, H-18)
19	17.7	1.24–1.23 (4H, m, H-19, 12b overlap)
1'	104.7	4.36 (1H, d, $J = 7.8\text{Hz}$ , H-1')
2'	75.4	3.24 (1H, dd, $J = 9.3, 7.9\text{Hz}$ , H-2')
3'	75.9	3.52 (1H, m, H-3')
4'	72.8	4.73 (1H, m, H-4')
5'	75.8	3.39 (1H, m, H-5')
6'	62.5	3.60 (1H, dd, $J = 12.1, 2.6\text{Hz}$ , H-6'a) 3.50 (1H, m, H-6'b)
1''	172.2	–
2''	20.9	2.10–2.03 (6H, m, H-2'', 16a, 1a, 12a overlap)

–: not applicable.

glycosides (Figs. 4 and 5). Although we have no evidences for the role of LacA protein *in vivo*, these findings in the present work may provide a starting point for identifying the exact activity of LacA protein in lactose metabolism. The bottleneck in enzymatic synthesis of ASGs is the lack of well-characterized SGAs. The successful characterization of LacA made it to be the first SGA and LacA protein was thus designated as EcSGA1.

A novel enzymatic two-step synthesis of AT-17 $\beta$ -Gs and AE-17 $\beta$ -Gs from the abundantly available free steroids under the sequential actions of a steroidal glycosyltransferase OsSGT1 from *O. saundersiae* and EcSGA1 was achieved. The two-step process is characterized by acyltransferase-catalyzed regioselective acylations of all hydroxyl groups of unprotected SGs in the late stage,

thereby significantly streamlining the synthetic route towards ASGs and thus forming four monoacylates.

Regioselective acylation could expand molecular diversity, thereby facilitating the discovery of pharmaceutical leads. In this investigation, EcSGA1-catalyzed acetylation of two steroid 17 $\beta$ -glucosides (T-17 $\beta$ -G and E-17 $\beta$ -G) led to the production of eight new monoacylates. Furthermore, the cytotoxic activities of these monoacylates were tested and 3'-AT-17 $\beta$ -G was observed to display improved activities towards seven human tumor cell lines, suggesting this compound had promisingly pharmacological potential. This study therefore reports for the first time a novel synthetic process for the green preparation of acylated steroidal glycosides with medicinal interest.

#### 4. Conclusions

A steroidal glycosyltransferase OsSGT1 from *O. saundersiae* was identified to be the first plant SGT with selectivity towards both 3 $\beta$ - and 17 $\beta$ -hydroxysteroids. One of the most striking findings of this study is the characterization of bacterial LacA protein as a steroidal glycoside acyltransferase, catalyzing the attachment of acyl groups into the hydroxyl groups of steroidal 17 $\beta$ -glycosides. A novel enzymatic two-step synthesis of AT-17 $\beta$ -Gs and AE-17 $\beta$ -Gs from the abundantly available free steroids under the sequential actions of OsSGT1 and EcSGA1 was achieved. The two-step process is characterized by acyltransferase-catalyzed regioselective acylations of all hydroxyl groups of unprotected SGs in the late stage, thereby significantly streamlining the synthetic route towards ASGs and thus forming four monoacylates.

Moreover, the cytotoxic activities of these monoacylates were tested and 3'-AT-17 $\beta$ -G was observed to display improved activities towards seven human tumor cell lines. This study therefore reports for the first time a novel synthetic process for the green preparation of acylated steroidal glycosides with medicinal interest.

#### 5. Experimental

##### 5.1. Chemicals

In this contribution, four compound libraries, namely sugar acceptor, sugar donor, acyl acceptor and acyl donor libraries, were provided for enzyme-mediated reactions. The compounds listed in Fig. 1 and Supplementary Information Fig. S4 include diverse structures like steroids (1–24), flavonoids (25–31), alkaloids (32–38), triterpenoids (39–42), phenolic acids (43–47) and coumarins (48–49) are used as the sugar acceptors for OsSGT1-catalyzed glycosylation reactions (Fig. 1 and Supplementary Information Fig. S4). The sugar donors consist of seven UDP-activated nucleotides, among which, UDP-D-glucose (UDP-Glc), UDP-D-galactose (UDP-Gal), UDP-D-glucuronic acid (UDP-GlcA) and UDP-N-acetylglucosamine (UDP-GlcNAc) were obtained from Sigma-Aldrich Co., LLC. (St. Louis, MO, USA). UDP-D-xylose (UDP-Xyl), UDP-L-arabinose (UDP-Ara) and UDP-D-galacturonic acid (UDP-GalA) was synthesized by enzyme-mediated reactions in our laboratory<sup>27–29</sup>. The acyl acceptor library is made up of 10 steroidal glucosides (1a–10a) and 13 other glucosides (50–62) listed in Supplementary Information Fig. S58. The acyl donor library includes acetyl-CoA, succinyl-CoA, arachidonoyl-CoA, palmitoyl-CoA and acetoacetyl-CoA, all of which were purchased from Sigma-Aldrich Co., LLC. The other chemicals were either reagent or analytic grade when available.

**Table 6** The cytotoxic activities of monoacylates against human tumor cell lines.

Compd.	IC <sub>50</sub> (μmol/L)						
	HCT116	Bel7402	MGC803	Capan2	NCI-H460	NCI-H1650	A549
T-17β-G	> 50.0	> 50.0	> 50.0	> 50.0	> 50.0	> 50.0	> 50.0
2'-AT-17β-G	> 50.0	> 50.0	> 50.0	> 50.0	> 50.0	> 50.0	> 50.0
3'-AT-17β-G	14.1	16.5	10.6	16.1	8.94	3.62	6.91
4'-AT-17β-G	> 50.0	> 50.0	> 50.0	> 50.0	> 50.0	> 50.0	> 50.0
6'-AT-17β-G	> 50.0	> 50.0	–	> 50.0	26.5	–	–
E-17β-G	> 50.0	> 50.0	–	> 50.0	> 50.0	–	–
6'-AE-17β-G	> 50.0	> 50.0	–	> 50.0	30.5	–	–
Taxol	0.000311	1.08	0.000299	0.00186	0.00754	2.43	0.0135

–: not applicable.

### 5.2. Plant material

*O. saundersiae* plants were grown in experimental pots in Institute of Materia Medica, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China. Fresh bulbs of *O. saundersiae* were collected to frozen in liquid N<sub>2</sub> and then stored at –80 °C for RNA isolation.

### 5.3. Transcriptome sequencing of *O. saundersiae*

The procedure of transcriptome sequencing of *O. saundersiae* was the same as that of *O. caudatum*<sup>28–31</sup>. In brief, total RNA was extracted from the frozen bulbs of *O. saundersiae* using TRIzol reagent (CW BIO Co., Ltd., Beijing, China) according to the manufacturer's instructions. The concentration, the RNA ratio of 28 S to 18 S and RNA integrity number (RIN) were evaluated using an Agilent 2100 Bioanalyzer. The RNAs with RIN value ≥ 8.5 were sent to BGI Tech Solutions Co., Ltd. for cDNA preparation and library construction. Finally, the purified library was sequenced on Illumina HiSeq™ 2000 apparatus in BGI Tech Solutions Co., Ltd. (BGI-Tech). RNA-Seq raw data was thus obtained to form a sequence library.

### 5.4. Data analyses and functional annotations of unigenes

The resulting raw reads from sequence library of *O. saundersiae* was firstly filtered to obtain clean reads, discarding dirty reads with adaptors, unknown or low quality bases. These clean reads were subsequently combined to form longer unigenes by assembling program trinity. These unigenes obtained by *de novo* assembly cannot be extended on either end. Next, these unigenes were aligned by Blast X algorithm to protein databases, such as NR, Swiss-Prot, KEGG and COG (e-value < 0.00001) for functional annotation. The unigenes displaying similarity to SGTs were retrieved for further ORF analysis. In a word, those unigenes with a complete ORF and displaying high similarity to SGTs were selected as the candidate for further investigation.

### 5.5. cDNA isolation and bioinformatics analysis of *OsSGT1*

To verify the authenticity of the candidate unigene, cDNA isolation was performed using gene-specific primers by a nested PCR assay as previously described (Supplementary Information Table S1)<sup>28–31</sup>. The obtained amplicon was inserted into

*pEASY*<sup>TM</sup>-Blunt plasmid (TransGen Co., Ltd., Beijing, China) and then transformed into *E. coli* *Trans1*-T1 competent cells for recombinant plasmid selection (Supplementary Information Table S2). The resultant recombinant plasmid was isolated and subjected to nucleotide sequencing. The obtained cDNA was thus designated as *OsSGT1* for convenience. The bioinformatics analyses of *OsSGT1*, like prediction of physiochemical properties, multiple sequence alignment and phylogenetic analysis, were performed as detailed in our previous reports<sup>27–30,32</sup>.

### 5.6. Prokaryotic expression and preparation of crude cell extract for glycosylation reactions

*OsSGT1* was amplified using gene-specific primers (Supplementary Information Table S1) and the resulting PCR product was ligated into *EcoRI* and *Hind III* sites within the pET-28a(+) vector (Novagen, Madison, USA) using seamless assembly cloning kit (CloneSmarter Technologies Inc., Houston, TX, USA) as described previously<sup>33</sup>. The generated construct pET28a-*OsSGT1* was transformed into *E. coli* strain *Transtetta* (DE3) for expression as described previously<sup>34</sup>. Also, to improve heterologous expression of *OsSGT1*, pET28a-*OsSGT1* was co-transformed into *E. coli* BL21(DE3) strain with a chaperone plasmid pGro7 (Takara Biotechnology Co., Ltd., Dalian, China) as introduced by Yin et al.<sup>27</sup> The expression of *OsSGT1* was induced by IPTG at a final concentration of 0.3 mmol/L. The expressed *OsSGT1* was checked by SDS-PAGE and Western-blot analyses as described by Guo et al.<sup>31</sup> Next, the BL21(DE3) [pET28a-*OsSGT1*+pGro7] suspension cells were disrupted in a high-pressure homogeniser (APV-2000, Albertslund, Denmark) operated at 800 bar. Disrupted cells were centrifuged at 12,000 rpm for 30 min to discard the pellet. The resultant supernatant, namely the membrane-free crude extract, was used as the biocatalyst for steroidal glycosylation.

### 5.7. Assay for steroidal glycosyltransferase activity of *OsSGT1*

After verification of heterologous expression of *OsSGT1*, the crude extract containing the recombinant *OsSGT1* was applied as the biocatalyst to react with various sugar acceptors and donors (Fig. 1 and Supplementary Information Fig. S4). The total reaction mixture was 100 μL contained 20 mmol/L phosphate buffer (pH 8.0), a sugar acceptor (1 mmol/L), a sugar donor (1 mmol/L) and 20 μL crude *OsSGT1* proteins. The reaction mixture was incubated at 50 °C for 1 h. The formation of glycosylated products was

unambiguously determined by a combination of HPLC–UV, HPLC–MS and NMR as described previously<sup>35,36</sup>. The determination conditions for HPLC–UV were summarized in [Supplementary Information Table S3](#).

#### 5.8. Isolation and prokaryotic expression of *E. coli* genes encoding *O*-acetyltransferase

The genome DNA was extracted from *E. coli* strain BL21(DE3) using BacteriaGen DNA kit (CWBio Co., Ltd., Beijing, China) according to the supplier recommendation. The resulting genome DNA was then used as the template of PCR amplification to isolate these candidate SGA genes using gene-specific primers ([Supplementary Information Table S1](#)). The amplified PCR products were inserted into *pEASY*<sup>TM</sup>-Blunt plasmid to generate recombinant vectors for sequencing verification. Next, these *O*-acetyltransferase-encoding genes were heterologously expressed in BL21(DE3) as described above. SDS-PAGE and Western-blot of these recombinant proteins were conducted as that of OsSGT1 (see above). The recombinant *O*-acetyltransferase proteins were subjected to purification with Ni-NTA agarose columns according to the manufacturer's protocol. Purified protein concentrations were determined using Bradford protein assay (Bio-Rad, Hercules, CA, USA).

#### 5.9. Functional characterization of *E. coli* steroidal glycoside acyltransferases (EcSGAs)

The enzymatic activities of EcSGAs were determined in 100  $\mu$ L citrate buffer solution (pH 5.0) containing an acyl acceptor (1 mmol/L) listed in [Supplementary Information Fig. S58](#), an acyl donor (1 mmol/L) summarized in Chemicals section and the purified protein (2.97  $\mu$ g). The reactions were incubated at 37 °C for 1 h. Then 100  $\mu$ L methanol was added to terminate the reaction. The reaction mixture was monitored by HPLC–UV ([Supplementary Information Table S3](#)) and the structure of the generated product was determined by a combination of HPLC–MS and NMR as reported by Liu et al.<sup>35</sup>

#### 5.10. Biotransformation of testosterone and testosterone-17-*O*- $\beta$ -glucoside using the whole cell system

After IPTG induction, the engineered BL21(DE3)[pET28a-OsSGT1+pGro7] cells were harvested by centrifugation at 12,000 rpm for 10 min and then resuspended in M9 medium<sup>37</sup> with cell density of OD<sub>600</sub> value of 0.6. The substrate testosterone (**8**) or testosterone-17-*O*- $\beta$ -glucoside (T-17-*O*- $\beta$ -G, **8a**) with the final concentration of 0.3 mmol/L was added into the M9 medium and continued to incubate at 37 °C overnight. The formation of products was monitored by HPLC analysis as mentioned above.

#### 5.11. Optimization of OsSGT1- and EcSGA1-catalyzed reactions

The effects of pH and temperature on OsSGT1- and EcSGA1-catalyzed reactions were investigated. The crude extract of BL21(DE3)[pET28a-OsSGT1+pGro7] and the purified EcSGA1 protein were applied as the biocatalyst in their respective reactions.

The effects of pH on both reactions were determined at varied buffers including citric acid/sodium citrate buffer (0.1 mol/L, pH 3.0–6.6), Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> (0.1 mol/L, pH 5.8–9.0), Na<sub>2</sub>HPO<sub>4</sub>/NaOH buffer (0.1 mol/L, pH 9.0–12).

The influences of temperature were explored in a range of 20 to 80 °C with intervals of 15 °C (OsSGT1-mediated glycosylation) or 10 °C (EcSGA1-catalyzed acylation) in the standard reaction mixture as described above.

#### 5.12. Enzymatic two-step synthesis and structural characterization of AT-17 $\beta$ -Gs

Scale-up of OsSGT1- and EcSGA1-catalyzed reactions was performed to obtain sufficient AT-17 $\beta$ -Gs for structural characterization and further cytotoxicity assay. Initially, the 100  $\mu$ L OsSGT1-catalyzed reaction was directly scaled to 53 mL, in which 152 mg testosterone (**8**) were added into and then incubated with crude cell extract at optimal pH and temperature for 1 h. The resultant reaction mixture was applied to preparative HPLC to isolate pure T-17 $\beta$ -G, which was then used as the substrate in 135 mL EcSGA1-catalyzed reaction for AT-17 $\beta$ -Gs production. Structure characterization of AT-17 $\beta$ -Gs was performed using HPLC–SPE–NMR technique as described by Liu et al.<sup>35</sup> except some modifications on chromatographic conditions. HPLC separation was carried out on an YMC-Pack Ph column (5  $\mu$ m, 12 nm, 250 mm  $\times$  4.6 mm) with an isocratic elution of 50% water-trifluoroacetic acid (A, 99.9%:0.1%, v/v) and 50% methanol (B) at a flow rate of 1 mL/min.

#### 5.13. Cytotoxicity assay of acetylated testosterone-17-*O*- $\beta$ -D-glucosides

Seven human cancer cell lines, HCT-116 (human colon cancer cell line), Bel7402 (human hepatocellular carcinoma cell line), MGC803 (human gastric carcinoma cell line), Capan 2 (human pancreatic cancer cell line), NCI-H1650, NCI-H460 and A549 (human lung cancer cell lines) were used in the cytotoxicity assay. The viability of the cells after treated with various chemicals was evaluated using the MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide) assay performed as previously reported<sup>7,38</sup>. The inhibitory effects of these tested compounds on the proliferation of cancer cells were reflected by their respective IC<sub>50</sub> (50% inhibitory concentration).

#### Acknowledgments

This work was supported by the CAMS Innovation Fund for Medical Sciences (CIFMS, No. 2016-I2M-3-012) and Beijing Natural Science Foundation (No. 7172143).

#### Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.apsb.2018.04.006](https://doi.org/10.1016/j.apsb.2018.04.006).

#### References

1. Nyström L, Schär A, Lampi AM. Steryl glycosides and acylated steryl glycosides in plant foods reflect unique sterol patterns. *Eur J Lipid Sci Technol* 2012;**114**:656–69.
2. Ito Y, Nakashima Y, Matsuoka S. Rice bran extract containing acylated steryl glucoside fraction decreases elevated blood LDL cholesterol level in obese Japanese men. *J Med Investig* 2015;**62**:80–4.

- Usuki S, Ariga T, Dasgupta S, Kasama T, Morikawa K, Nonaka S, et al. Structural analysis of novel bioactive acylated sterol glucosides in pre-germinated brown rice bran. *J Lipid Res* 2008;**49**:2188–96.
- Yoon NY, Min BS, Lee HK, Park JC, Choi JS. A potent anti-complementary acylated sterol glucoside from *Orostachys japonicus*. *Arch Pharm Res* 2005;**28**:892–6.
- Shimamura M, Hidaka H. Therapeutic potential of cholesteryl *O*-acyl alpha-glucoside found in *Helicobacter pylori*. *Curr Med Chem* 2012;**19**:4869–74.
- Shimamura M. Immunological functions of sterol glycosides. *Arch Immunol Ther Exp (Warsz)* 2012;**60**:351–9.
- Mimaki Y, Kuroda M, Kameyama A, Sashida Y, Hirano T, Oka K, et al. Cholestane glycosides with potent cytostatic activities on various tumor cells from *Ornithogalum saundersiae* bulbs. *Bioorg Med Chem Lett* 1997;**7**:633–6.
- Mimaki Y, Kuroda M, Kameyama A, Sashida Y, Hirano T, Oka K, et al. A new rearranged cholestane glycoside from *Ornithogalum saundersiae* bulbs exhibiting potent cytostatic activities on leukemia HL-60 and molt-4 cells. *Bioorg Med Chem Lett* 1996;**6**:2635–8.
- Kubo S, Mimaki Y, Terao M, Sashida Y, Nikaido T, Ohmoto T. Acylated cholestane glycosides from the bulbs of *Ornithogalum saundersiae*. *Phytochemistry* 1992;**31**:3969–73.
- Morzycki JW, Wojtkielewicz A. Synthesis of a cholestane glycoside OSW-1 with potent cytostatic activity. *Carbohydr Res* 2002;**337**:1269–74.
- Deng S, Yu B, Lou Y, Hui Y. First total synthesis of an exceptionally potent antitumor saponin, OSW-1. *J Org Chem* 1999;**64**:202–8.
- Wimmerová M, Siglerová V, Šaman D, Šlouf M, Kaletová E, Wimmer Z. Improved enzyme-mediated synthesis and supramolecular self-assembly of naturally occurring conjugates of  $\beta$ -sitosterol. *Steroids* 2017;**117**:38–43.
- Jäger SN, Mittelbach M, Cabrera R, Labadie GR. Simple method for high purity acylated sterol glucosides synthesis. *Eur J Lipid Sci Technol* 2016;**118**:827–33.
- Ueda Y, Mishiro K, Yoshida K, Furuta T, Kawabata T. Regioselective diversification of a cardiac glycoside, lanatoside C, by organocatalysis. *J Org Chem* 2012;**77**:7850–7.
- Chaturvedi P, Misra P, Tuli R. Sterol glycosyltransferases—the enzymes that modify sterols. *Appl Biochem Biotechnol* 2011;**165**:47–68.
- Grille S, Zaslowski A, Thiele S, Plat J, Warnecke D. The functions of sterol glycosides come to those who wait: recent advances in plants, fungi, bacteria and animals. *Prog Lipid Res* 2010;**49**:262–88.
- Malik V, Zhang M, Dover LG, Northen JS, Flinn A, Perry JJ, et al. Sterol  $3\beta$ -glucosyltransferase biocatalysts with a range of selectivities, including selectivity for testosterone. *Mol Biosyst* 2013;**9**:2816–22.
- Tiwari P, Sangwan RS, Asha, Mishra BN, Sabir F, Sangwan NS. Molecular cloning and biochemical characterization of a recombinant sterol 3-*O*-glucosyltransferase from *Gymnema sylvestris* R.Br. catalyzing biosynthesis of sterol glucosides. *Biomed Res Int* 2014;**2014**:934351.
- Warnecke D, Erdmann R, Fahl A, Hube B, Müller F, Zank T, et al. Cloning and functional expression of UGT genes encoding sterol glucosyltransferases from *Saccharomyces cerevisiae*, *Candida albicans*, *Pichia pastoris*, and *Dictyostelium discoideum*. *J Biol Chem* 1999;**274**:13048–59.
- Vogt T, Jones P. Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. *Trends Plant Sci* 2000;**5**:380–6.
- Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T. Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. *Appl Environ Microbiol* 1998;**64**:1694–9.
- González-Sabín J, Morán-Ramallal R, Rebolledo F. Regioselective enzymatic acylation of complex natural products: expanding molecular diversity. *Chem Soc Rev* 2011;**40**:5321–35.
- Tang Y, Li N, Duan JA, Tao W. Structure, bioactivity, and chemical synthesis of OSW-1 and other steroidal glycosides in the genus *Ornithogalum*. *Chem Rev* 2013;**113**:5480–514.
- Stucky DF, Arpin JC, Schrick K. Functional diversification of two UGT80 enzymes required for sterol glucoside synthesis in *Arabidopsis*. *J Exp Bot* 2015;**66**:189–201.
- Marbach A, Bettenbrock K. lac operon induction in *Escherichia coli*: systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. *J Biotechnol* 2012;**157**:82–8.
- Roderick SL. The lac operon galactoside acetyltransferase. *Comptes Rendus Biol* 2005;**328**:568–75.
- Yin S, Sun YJ, Liu M, Li LN, Kong JQ. cDNA Isolation and functional characterization of UDP-D-glucuronic acid 4-epimerase family from *Ornithogalum caudatum*. *Molecules* 2016;**21**:E1505.
- Yin S, Kong JQ. Transcriptome-guided discovery and functional characterization of two UDP-sugar 4-epimerase families involved in the biosynthesis of anti-tumor polysaccharides in *Ornithogalum caudatum*. *RSC Adv* 2016;**6**:37370–84.
- Yin S, Kong JQ. Transcriptome-guided gene isolation and functional characterization of UDP-xylose synthase and UDP-D-apiose/UDP-D-xylose synthase families from *Ornithogalum caudatum* Ait. *Plant Cell Rep* 2016;**35**:2403–21.
- Li LN, Kong JQ. Transcriptome-wide identification of sucrose synthase genes in *Ornithogalum caudatum*. *RSC Adv* 2016;**23**:18778–92.
- Guo L, Chen X, Li LN, Tang W, Pan YT, Kong JQ. Transcriptome-enabled discovery and functional characterization of enzymes related to (2S)-pinocembrin biosynthesis from *Ornithogalum caudatum* and their application for metabolic engineering. *Microb Cell Fact* 2016;**15**:27.
- Yin S, Liu M, Kong JQ. Functional analyses of OcRhS1 and OcUER1 involved in UDP-L-rhamnose biosynthesis in *Ornithogalum caudatum*. *Plant Physiol Biochem* 2016;**109**:536–48.
- Kong J, Shi Y, Wang Z, Pan Y. Interactions among SARS-CoV accessory proteins revealed by bimolecular fluorescence complementation assay. *Acta Pharm Sin B* 2015;**5**:487–92.
- Liu M, Li LN, Pan YT, Kong JQ. cDNA isolation and functional characterization of squalene synthase gene from *Ornithogalum caudatum*. *Protein Expr Purif* 2017;**130**:63–72.
- Liu X, Wang ZB, Wang YN, Kong JQ. Probing steroidal substrate specificity of cytochrome P450 BM3 variants. *Molecules* 2016;**21**:760.
- Liu X, Kong JQ. Steroids hydroxylation catalyzed by the monooxygenase mutant 139-3 from *Bacillus megaterium* BM3. *Acta Pharm Sin B* 2017;**7**:510–6.
- Atsumi S, Cann AF, Connor MR, Shen CR, Smith KM, Brynildsen MP, et al. Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metab Eng* 2008;**10**:305–11.
- Kuroda M, Mimaki Y, Yokosuka A, Sashida Y, Beutler JA. Cytotoxic cholestane glycosides from the bulbs of *Ornithogalum saundersiae*. *J Nat Prod* 2001;**64**:88–91.