SIMB
Society for Industrial Microbiology and Biotechnology

FERMENTATION, CELL CULTURE AND BIOENGINEERING - ORIGINAL PAPER

Check for updates

Improving an *Escherichia coli*-based biocatalyst for terpenol glycosylation by variation of the expression system

Julian Rüdiger[1] · Wilfried Schwab[1]

Abstract

Glycosides are becoming increasingly more relevant for various industries as low-cost whole-cell-biocatalysts are now available for the manufacture of glycosides. However, there is still a need to optimize the biocatalysts. The aim of this work was to increase the titre of terpenyl glucosides in biotransformation assays with *E. coli* expressing *VvGT14ao*, a glycosyltransferase gene from grape (*Vitis vinifera*). Seven expression plasmids differing in the resistance gene, origin of replication, promoter sequence, and fusion protein tag were generated and transformed into four different *E. coli* expression strains, resulting in 18 strains that were tested for glycosylation efficiency with terpenols and a phenol. *E. coli* BL21(DE3)/pET-SUMO_VvGT14ao yielded the highest titres. The product concentration was improved 8.6-fold compared with *E. coli* BL21(DE3)pLysS/pET29a_VvGT14ao. The selection of a small solubility-enhancing protein tag and exploitation of the T7 polymerase-induction system allowed the formation of increased levels of functional recombinant protein, thereby improving the performance of the whole-cell biocatalyst.

Keywords Whole-cell biocatalysis · Glycosylation · Fusion protein · Glycosyltransferase · Terpenol

Introduction

Enzymes offer tremendous opportunities for the chemical industry as they can efficiently and economically convert molecules and replace well-known and established industrial processes with more sustainable ones. Traditionally, enzymes are used by food and feed industry, but are increasingly implemented in emerging markets such as pharmaceutical and fine chemicals industries [18]. The production of catalytically active proteins in heterologous microbial and fungal hosts leads to lower costs and high yields, which is crucial for the economy of an industrial process. *Escherichia coli* is the host of choice if post-translational modifications of the enzyme are not required for its activity [19]. The use of *E. coli* as whole-cell biocatalyst offers additional benefits

such as cofactor regeneration, high catalytic efficiency and mild and simple reaction conditions [24]. However, the production of misfolded and inactive aggregates such as inclusion bodies is problematic when recombinant proteins are over-expressed in bacteria. Strategies for overcoming this disadvantage of heterologous expression are the optimization of the expression system, e.g. by using tightly regulated promoters or by attaching solubility-enhancing fusion tags. When using whole-cell biocatalysts for biocatalysis, substrate and product transmembrane transport and cofactor regeneration capacity of the used strain must similarly be considered. For the moment, the most promising method for whole-cell biocatalysis seems to be the use of engineered *E. coli*, which offers the potential for large-scale and cost-effective production.

Glycosylation plays a central role in cellular communication, transport and storage of metabolites in organisms. Recently, the glycosylation reaction has attracted much attention because volatility, poor water solubility and instability limit the application of many small molecules in consumer products [7, 28]. The transfer of sugars to small molecules makes them non-volatile, more water-soluble and improves their stability, bioavailability and functionality. Volatile terpenols become non-volatile by glycosylation,

✉ Wilfried Schwab
wilfried.schwab@tum.de

1 Biotechnology of Natural Product, Technical University Munich, Liesel-Beckmann-Str. 1, 85354 Freising, Germany

allowing them to be applied as slow release aroma chemicals [29] and the water solubility of resveratrol increases by a factor of 1700 when sugars are attached, which results in improved bioavailability [21]. Glycosylated vitamin C and anthocyanidins are more stable than their non-glycosylated precursors [38, 39], alteration of the sugar residues of quercetin affects pharmacological properties [17], and transfer of sugars to the diterpenol steviol results in sweet tasting steviosides [10].

In nature, glycosylation reactions are primarily catalysed by nucleotide-sugar-dependent glycosyltransferases [30]. These enzymes transfer carbohydrate residues from activated sugar donors to a multitude of acceptor molecules, including proteins, saccharides, lipids and small molecules. Glycosylation of their –OH, –NH$_2$, –SH, and –COOH groups results in O-, N-, S-glucosides, and sugar esters, respectively [16]. Uridine diphosphate sugar-dependent glycosyltransferases (UGT) form the largest group and are characterized by a unique C-terminal consensus sequence representing the sugar donor site [22]. Plant genomes are rich sources for UGT genes, which code, among others for biocatalysts that can transfer sugars to small molecules and on-going sequencing projects continuously deliver new nucleotide sequences (http://supfam.org/). These UGTs are capable of efficiently glycosylating various plant metabolites from different chemical classes in a stereo- and regioselective manner [23]. VvGT14a from grapevine (*Vitis vinifera*) is a promiscuous enzyme that can efficiently glucosylate not only terpenols such as geraniol, citronellol and nerol, but also eugenol [2]. Thus, plant glycosyltransferases are ideal biocatalysts for the biochemical and biotechnological production of small molecule glucosides but utilize costly cosubstrates such as UDP-glucose, which requires the use of whole-cell biocatalysts for cosubstrate recycling and efficient production [7].

In the development of an in vivo biocatalytic glycosylation process, a number of issues have to be considered including the availability of suitable UGTs, amount of the catalytically active protein, the import of substrate, the export of product and the toxicity of the starting material [8]. Multiple glycosides can already be produced by biocatalysis as numerous UGTs have been functionally characterized and protein engineering studies have identified crucial amino acids that are important for acceptor and sugar donor preferences and regioselectivity [14]. However, apart from producing oligosaccharides that reached titres up to 188 g/l, the product yields are unsatisfactory and improvements are urgently needed [7].

The productivity of a whole-cell bioprocess is highly dependent on the effective production of functional enzymes encoded on expression plasmids or microbial chromosomes. The potential strain activity is the product of the enzyme concentration and its correctly formed proportion. Enzyme level and productivity could be increased by influencing the expression of the gene of interest, by using an efficient origin of replication (ori) [24], and by choosing appropriate selection markers (e.g. antibiotic resistance gene). Product titre was also raised by utilizing additional beneficial genes (e.g. the lac-repressor lacI) [34], by optimizing the ribosomal binding site [37], or by fusing the enzyme to a solubility enhancing partner [33]. In addition, catalytic efficiency of biocatalysts was increased by random mutagenesis followed by a screening for better performing variants [9] and the lifetime of enzymes was extended by modifying the C-terminus [27]. Furthermore, additional expression of chaperones promoted proper folding and prevented the accumulation of inactive enzyme in inclusion bodies [36].

Therefore, optimization of the vector for improved process productivity involves the selection of a suitable resistance marker, ori, promoter, and fusion tag. Ampicillin was not suitable for high cell-density fermentation as it was rapidly degraded in the periplasm [1]. Kanamycin was more stable and better suited for high cell-density fermentations [31, 32]. The ori determines the plasmid copy number (PCN). Plasmids with the pBR322 ori resulted in a PCN of approximately 15–20 [3], whereas the high-copy ori RSF yielded PCN of over 100 copies [6]. The promoter affects the expression strength and tac and T7 are considered strong, inducible promoters [4, 35]. A gene containing a T7 promoter is only transcribed by T7 RNA polymerase encoded by the lambda DE3 prophage insertion. Therefore, only strains with the (DE3) insertion transcribe genes harbouring a T7 promoter [15]. The small fusion tags, His-tag and S-tag, are mainly used for protein purification purposes [40] while larger fusion tags are said to improve solubility. The SUMO- and the trxA-tags are 100 and 109 amino acids long, respectively, and the glutathione S-transferase (GST)-tag has a length of 211 amino acids [26]. Other fusion tags like NusA and MBP are also often used to facilitate higher solubility, however they are even bigger than GST with 495 and 396 amino acids, respectively [7]. NusA and MBP were not investigated in this study due to their enormous sizes. The aim of an optimal fusion protein is to facilitate higher solubility without impeding the strain's metabolism.

To increase the glucoside titres in a biotransformation process using *E. coli* whole-cell biocatalysts expressing the codon-optimized glucosyltransferase VvGT14ao, seven plasmids were generated, which differed in the antibiotic resistance gene, ori, promoter, and fusion tag (Supplemental Table S3) and were used to transformed in four common different host strains. Since there is a large number of expression vectors and *E. coli* expression hosts, only a few frequently used expression vectors could be investigated in this study. Parallel biotransformations were performed in a 24-well format and levels of geranyl, neryl, citronellyl, and eugenyl glucoside quantified by liquid chromatography

(LC). The construction of an optimized expression system increased the productivity of the bioprocess considerably.

Methods

Chemicals, enzymes, and expression vectors

Thiamine, iron(III) chloride, nerol and citronellol were purchased from Sigma-Aldrich (Taufkirchen, Germany), boric acid from Merck (Darmstadt, Germany) and manganese(II) chloride from Acros Organics (Geel, Belgium). All other chemicals were obtained from Carl Roth (Karlsruhe, Germany). The T4 DNA Ligase was purchased from Promega (Mannheim, Germany) and all other enzymes from Thermo Fisher Scientific (Planegg, Germany). Expression vectors pET-29a(+) (pET29a), pET-32a(+) (pET32a) and pRSF-Duet-1 were obtained from Merck (Darmstadt, Germany), pGEX-4T-1 from GE Healthcare Life Sciences (Solingen, Germany) and pET His6 Sumo TEV LIC cloning vector (1S) (pET-SUMO) was a gift from Scott Gradia (Addgene plasmid # 29659; http://n2t.net/addgene:29659; RRID:Addgene_29659).

Nucleic acid extraction

Plasmid DNA was extracted from *E. coli* by using the PureYield™ Plasmid Miniprep System (Promega, Mannheim, Germany) according to the manufacturer's instruction. DNA fragments from PCR reactions and agarose gels were purified by using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel, Düren, Germany) following the manufacturer's protocol.

Construction of plasmids

The cloning strategies and vector maps are shown in Supplemental Figures S1 to S8. The glucosyltransferase gene *VvGT14a* was isolated from *Vitis vinifera* and the encoded protein biochemically characterized [2]. *VvGT14ao* was cloned in frame with the N-terminal GST tag into the pGEX-4T-1 expression vector (Amersham Bioscience), which was transformed into *E. coli* BL21 (DE3) pLysS cells. Eurofins Genomics (Ebersberg, Germany) synthesized the codon-optimized gene *VvGT14ao*. The expression vector pET-SUMO_VvGT14ao was produced by ligation independent cloning (LIC) [12]. The entry vector was linearized by using *Ssp*I and subsequently treated with T4 DNA polymerase and dGTP. The gene of interest was amplified by PCR using primers with the LIC-site extension (Supplemental Table S1). The purified PCR product was treated with T4 DNA polymerase and dCTP. The annealing was performed with a vector-to-insert ratio of 1:3 regarding to

the amount of substance. All other constructs were produced by standard cloning procedures. For the cloning of the *VvGT14ao* gene, the entry vectors were digested with *Bam*HI and *Not*I and dephosphorylated with FastAP. The gene of interest was amplified by PCR using primers with the corresponding *Bam*HI-/*Not*I-extensions (Supplemental Table S1). Where applicable, the primers were additionally expanded with spacer bases to assure an in-frame insertion. The cleavage was also done with *Bam*HI and *Not*I. The constructs were finished with a ligation. The vectors pGEX-K and pET32aK are modified versions of pGEX-4T-1 and pET32a. Their ampicillin cassettes were exchanged with the kanamycin resistance cassette KanR2 from pRSFDuet-1. For the cloning of pET32aK two parts, pET32a backbone and the kanamycin resistance gene, were amplified by PCR using primers with an *Eco*31I + linker-extension (Supplemental Table S1). Both parts were cleaved with the IIS nuclease *Eco*31I, creating two unique overhangs. The construct was finished with a ligation. For the cloning of pGEX-K, the entry vector pGEX-4T-1 was digested with *Aat*II and *Bsa*I and dephosphorylated by using FastAP. Subsequently, the DNA fragment was gel purified. The kanamycin resistance gene was amplified by PCR using primers with the corresponding *Aat*II/*Bsa*I-extensions (Supplemental Table S1). The cleavage of both parts was performed with *Aat*II and *Bsa*I. The construct was finished after ligation. All ligations were performed by T4 DNA ligase (Promega) with a vector-to-insert ratio of 1:3 by weight. The codon-optimized UGT gene *VvGT14ao* was cloned by a standard restriction-ligation procedure into respective plasmids to obtain pET29a_VvGT14ao, pET32a_VvGT14ao, pET32aK_VvGT14ao, pGEX-4T-1_VvGT14ao, pGEX-K_VvGT14ao and pRSFDuet-1_VvGT14ao (Supplemental Figure S1). The restriction of the PCR product (insert) and the target vectors was also done with *Bam*HI and *Not*I. The constructs were finished with a ligation.

Sequencing

The DNA sequences of new constructs were analysed by Eurofins Genomics GmbH and confirmed with DNA sequence alignment using Serial Cloner (http://serialbasics.free.fr/ Serial_Cloner.html; version 2.6.1).

Transformation

The different vectors were combined with suitable hosts according to their promotors. Vectors containing the T7 promoter were transformed into *E. coli* BL21(DE3) and *E. coli* BL21(DE3)pLysS. Vectors with the tac promoter were additionally combined with *E. coli* BL21 and *E. coli* Waksman. The resulting 18 production strains were tested under the

same conditions with the following four substrates: geraniol, nerol, citronellol and eugenol.

Minimal medium

The M9 minimal medium was prepared after the recipe from Arie Geerlof (https://www.helmholtz-muenchen.de/fileadmin/PEPF/Protocols/M9-medium_150510.pdf). One litre of medium was produced by adding 100 mL of M9 salt solution 10x ($Na_2HPO_4 \times 2\ H_2O$ 75.2 g/L, KH_2PO_4 30 g/L, NaCl 5 g/L, and NH_4Cl 5 g/L), 20 mL 20% (*w/v*) glucose, 1 mL 1 M $MgSO_4$, 0.3 mL 1 M $CaCl_2$, 1 mL biotin solution (1 mg/mL), 1 mL thiamine solution (1 mg/mL), and 10 mL of trace element solution 100X (EDTA 5 g/L, $FeCl_3 \times 6\ H_2O$ 0.83 g/L, $ZnCl_2$ 84 mg/L, $CuCl_2 \times 2\ H_2O$ 13 mg/L, $CoCl_2 \times 2\ H_2O$ 10 mg/L, H_3BO_3 10 mg/L, and $MnCl_2 \times 4\ H_2O$ 1.6 mg/L) to 867 mL of sterile water. All components were sterilized beforehand by autoclaving (water, M9 salt solution 10X, 20% glucose solution, 1 M $MgSO_4$, and 1 M $CaCl_2$) and by a 0.22-μM filter (biotin solution, thiamine solution, and 100X trace elements solution).

In vivo glycosylation using HitPlate cultivation

The method was adapted from a shaking flask method and optimized in terms of culture medium, and method of inoculation and induction [13]. For each test culture one well of the HitPlate 25 (24 well microtitre plate; HJ-BIOANALYTIK GmbH, Germany; Fig. 1) was filled with 2 mL fresh M9 medium and the appropriate concentration of antibiotic. The wells were inoculated with fresh cells either from an agar plate or from a liquid culture. This pre-culture plate was incubated overnight at 37 °C and 250 rpm on an orbital shaker with a shaking stroke of 25 mm. For the main culture, each well was filled with 2 mL M9 medium and the needed antibiotics. The wells were inoculated with 50 μL of pre-culture (2.5%) and incubated for 4 h at 37 °C and 250 rpm on an orbital shaker with a shaking stroke of 25 mm. After

Fig. 1 Production of glucosides by whole-cell biocatalysts using HitPlates as a cultivation system (**a**) and demonstration of the reproducibility (**b**). The production process includes the preparation of the pre-culture from fresh material (1), main culture (2), protein expression (3), biotransformation (4), sampling (5), HPLC analysis (6), and calculation of the product titre (7). Different strains were tested with this cultivation system to test its reproducibility. The cultivation was repeated four times in duplicate using geraniol as substrate. Geranyl glucoside was quantified by LC. Strains 1–5: *E. coli* BL21/pGEX-K_VvGT14ao, *E. coli* BL21(DE3)pLysS/pET29a_VvGT14ao, *E. coli* BL21(DE3)pLysS/pET32aK_VvGT14ao, *E. coli* BL21/pGEX-4T-1_VvGT14ao, and *E. coli* BL21(DE3)pLysS/pET32a_VvGT14ao. The average geranyl glucoside yields are shown with the 95% confidence intervals
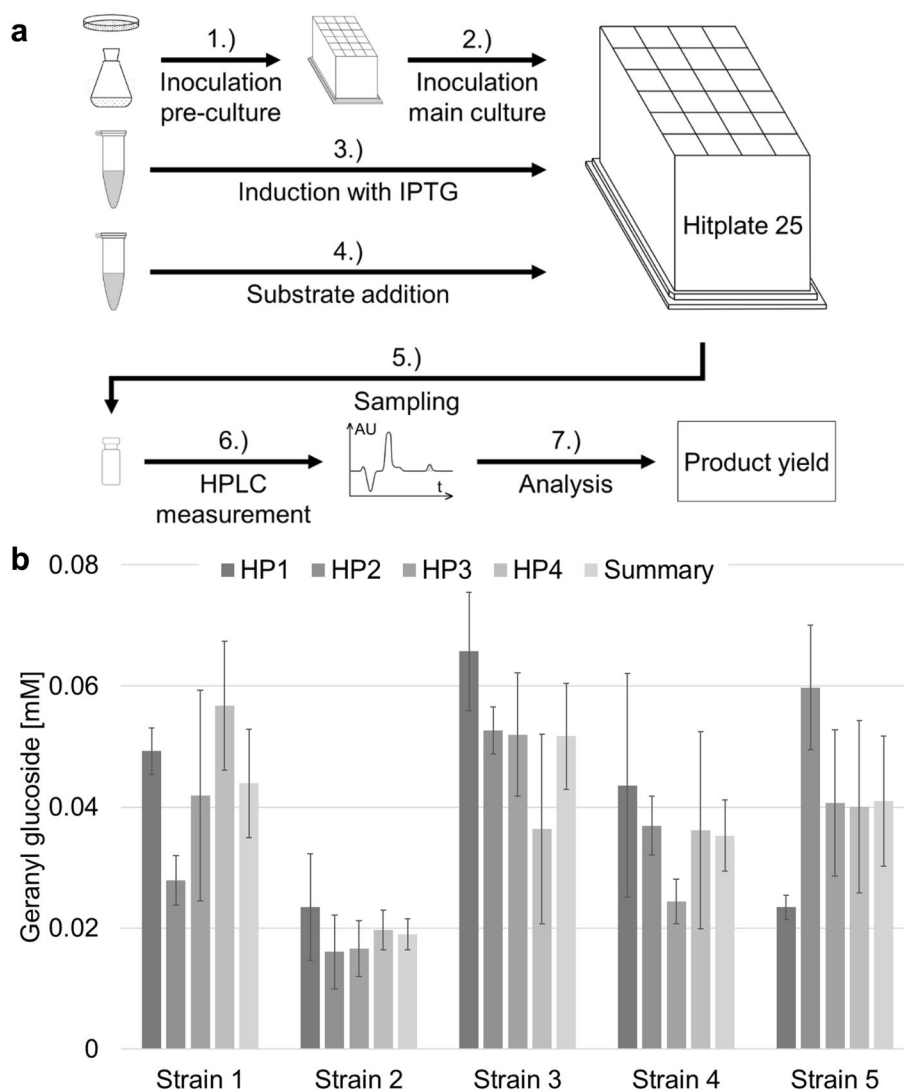
incubation the cultures were induced by adding 25 μL of 0.1 M isopropyl β-D-1-thiogalactopyranoside (IPTG) solution to each well resulting in an IPTG concentration of 1.25 mM. Subsequently, the cultures were incubated for protein production for 20 h at 18 °C and 500 rpm on an orbital shaker with a shaking stroke of 10 mm. After protein production, the cultures were supplemented with 500 μL of fresh M9 medium containing the required antibiotics and the substrate 0.1% (v/v) for the biotransformation. After substrate addition, the cultivation was continued for an additional 24 h under the same conditions. Finally, the optical density at 600 nm was measured from each well to determine the final cell concentration. From each well, a 1-mL sample was transferred to a micro-reaction tube and centrifuged. The supernatant was transferred to an HPLC vial for product analysis.

LC analysis

All LC machine parts were from Jasco (interface: LC-Net II/ADC; degasser: DG-1580-54; mixing chamber: LG-1580-04; pump: PU-1580; UV detector: UV-1575; autosampler: AS-1555). The column was a Luna® 3 μm C18 LC column 150×2 mm from Phenomenex (Aschaffenburg, Germany). The sample ran under isocratic conditions with 45% (v/v) methanol at 0.2 mL/min for 8 min, except for the analysis of citronellyl glucoside, which ran for 12 min. The substances were measured with the UV detector at 210 nm. The products geranyl glucoside, neryl glucoside, citronellyl glucoside and eugenyl glucoside eluted at 5.4 min, 5.5 min, 8.3 min and 3.2 min, respectively. The calibration was done with authentical reference material. The statistically analysis was done with Microsoft Office Excel.

Results and discussion

Validation of the experimental set-up

A small-scale production system based on microtitre plates (HitPlates) was utilized to reliably quantify the titres obtained with the production strains generated in this study (Fig. 1). The plates allowed the downscaling from shaking flasks with 50 mL of culture to a working volume of 2 mL culture per well. Each well had a total volume of 25 mL, which allowed for high rotational speeds (Supplemental Table S2). The formation of a wave inside the wells could only be observed at rotational speeds upwards of 200 rpm. The maximum wave height was 6, 13, and 16 mm above the resting liquid level at rotational speeds of 300, 400, and 500 rpm, respectively. The formation of a wave increases the liquids surface and improves the surface/volume ratio further. This was necessary for good aeration and thus for

unhindered cell growth. Experiments could be performed in parallel and in multiple determination across the 24 wells (4 × 6) of the plate. Multiple repetitions of biotransformation experiments confirmed the reproducibility of the experimental results (Fig. 1). Initially, the cell densities were adjusted before induction. Since the cell densities were only slightly different, but the workload involved was high, the laborious adaptation was omitted. Since the glucosyltransferase VvGT14ao (codon-optimized) from grapevine, which was used to develop the whole-cell biocatalysts exhibited promiscuous activity toward terpenols and phenols [2], four alcohols (geraniol, nerol, citronellol, and eugenol) were used as substrates. In total, 18 strains were generated and analysed (Figs. 2 and 3). The experiments were carried out three times and each culture measured in duplicate.

pET29a_VvGT14ao

Escherichia coli BL21(DE3)pLysS/pET29a_VvGT14ao (Fig. 3, strain #4) was the initial strain for the biotransformation and was used as a reference. The widely used expression vector pET29a provided a S-tag sequence upstream of the multiple cloning site (MCS) and thus, the resulting protein was fused to an N-terminal S-tag. The titres achieved with strain #4 ranged from 0.007 to 0.017 mM for eugenyl glucoside and citronellyl glucoside, respectively. Since no clear substrate preference was observed, the average titre of all substrates was determined (Fig. 3b). When pET29a_VvGT14ao was expressed in *E. coli* BL21(DE3) (Fig. 3, strain #3), the product concentration increased significantly ranging from 0.058 to 0.097 mM for geranyl glucoside and eugenyl glucoside, respectively. The pLysS plasmid, which produces T7 lysozyme to reduce basal level expression of the gene of interest, appeared to be interfering with the production of glucosides.

pRSFDuet-1_VvGT14ao

The pRSFDuet-1 vector can be used for co-expression of two genes and was used in this study as alternative expression plasmid offering the possibility to co-express supportive enzymes. Both strains, *E. coli* BL21(DE3) and *E. coli* BL21(DE3)pLysS, combined with pRSFDuet-1_VvGT14ao produced only a small amount of glucosides of less than 0.009 mM (Fig. 3, strains #1 and #2). There was no significant difference between the two strains in terms of product titre. *E. coli* BL21(DE3)pLysS/pRSFDuet-1_VvGT14ao (Fig. 3, strain #2) only grew to an optical density (OD) of 0.13 on average, while strain #1 showed an OD of 2.4 (Supplemental Figure S9). Taking the cell concentration into account, strain #2 is over 10-fold more productive than strain #1. However, strain #2 is not suitable for glucoside production due to the extremely low cell growth. Overall,
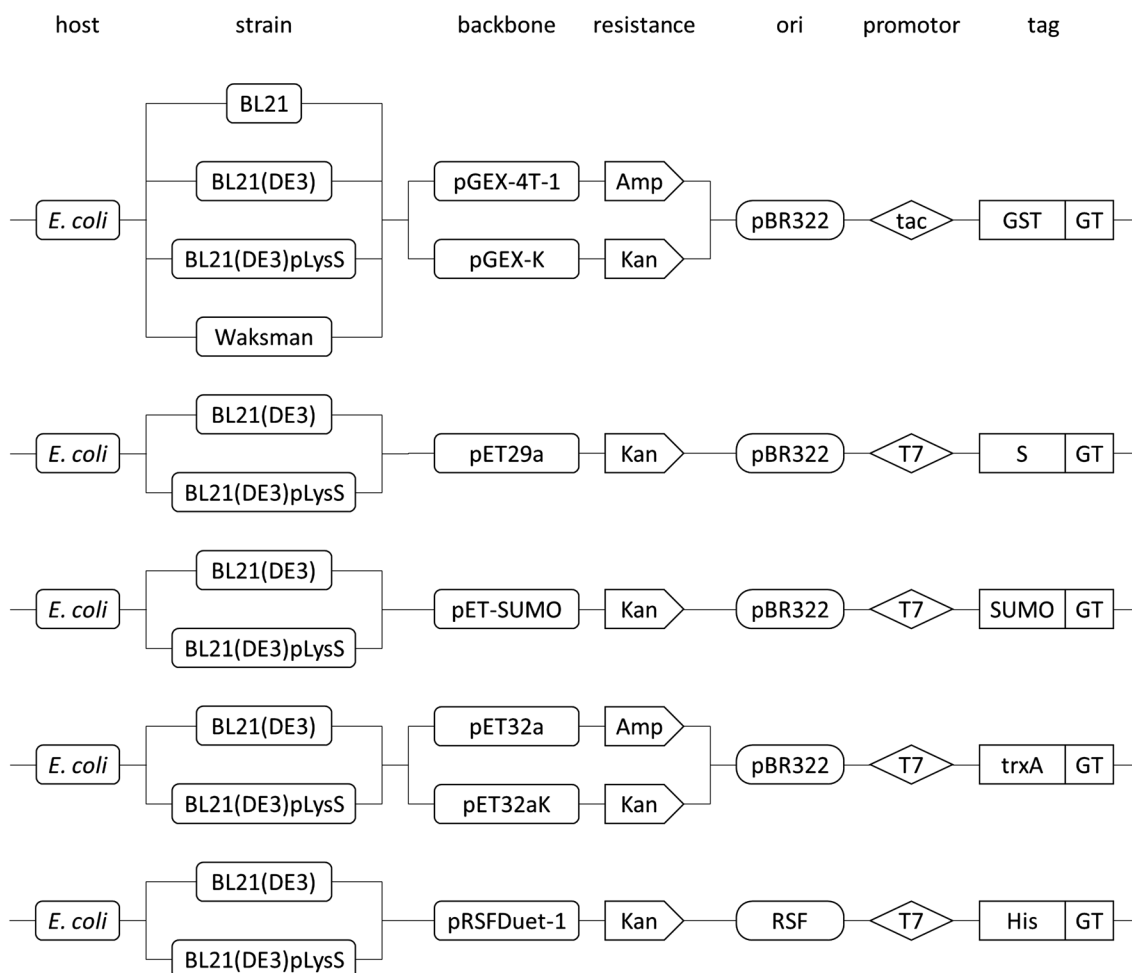
Fig. 2 Tested combination of strains and vectors. Four *E. coli* strains (BL21, BL21(DE3), BL21(DE3)pLysS, and Waksman) were combined with seven plasmids differing in antibiotic resistance (Amp and Kan), ori (pBR322 and RSF), promotor (tac and T7) and protein tag (GST, S, SUMO, trxA, and His). The combination of all suitable strains and vectors resulted in a total of 18 different strains. Each vector contained the gene coding for the glycosyltransferase *VvGT14ao* from *Vitis vinifera*. The gene was codon-optimized

pRSFDuet-1 offered no advantage in terms of productivity over the pET29a vector.

pET32aK_VvGT14ao

The pET32aK vector harboured the trxA sequence coding for the thioredoxin protein upstream of the MCS resulting in a fusion protein with the thioredoxin protein on the N-terminus. This fusion partner should promote solubility and lead to a correctly folded, active enzyme. Both strains, *E. coli* BL21(DE3) and *E. coli* BL21(DE3)pLysS, combined with pET32aK_VvGT14ao (Fig. 3, strains #5 and #6, respectively) produced on average 0.057 mM and 0.052 mM terpenyl glucosides. The glucoside titres between strain #5 and #6 did not differ significantly, but were considerably increased in comparison with the reference strain #4. Thus, the fusion protein thioredoxin has a positive effect on the productivity of the whole-cell biocatalyst.

pET-SUMO_VvGT14ao

The pET-SUMO vector carried the nucleotide sequence coding for the small ubiquitin-like modifier (SUMO) upstream of the LIC site and thus, the resulting protein contained an N-terminal SUMO-tag. The vector was chosen because the fusion product was reported to be more soluble and to increase the activity of the biocatalyst. *E. coli* BL21(DE3), and *E. coli* BL21(DE3)pLysS combined with pET-SUMO_VvGT14ao (Fig. 3, strains #7 and #8, respectively) produced high concentrations of glucosides but differed significantly in their glucosylation capabilities. The product levels ranged from 0.048 to 0.081 mM of eugenyl and citronellyl glucoside for strain #8, respectively, and from 0.072 to 0.120 mM of geranyl and eugenyl glucoside for strain #7, respectively. The SUMO fusion tag resulted in significantly increased product titres. The productivity of the bioprocess was improved by a factor
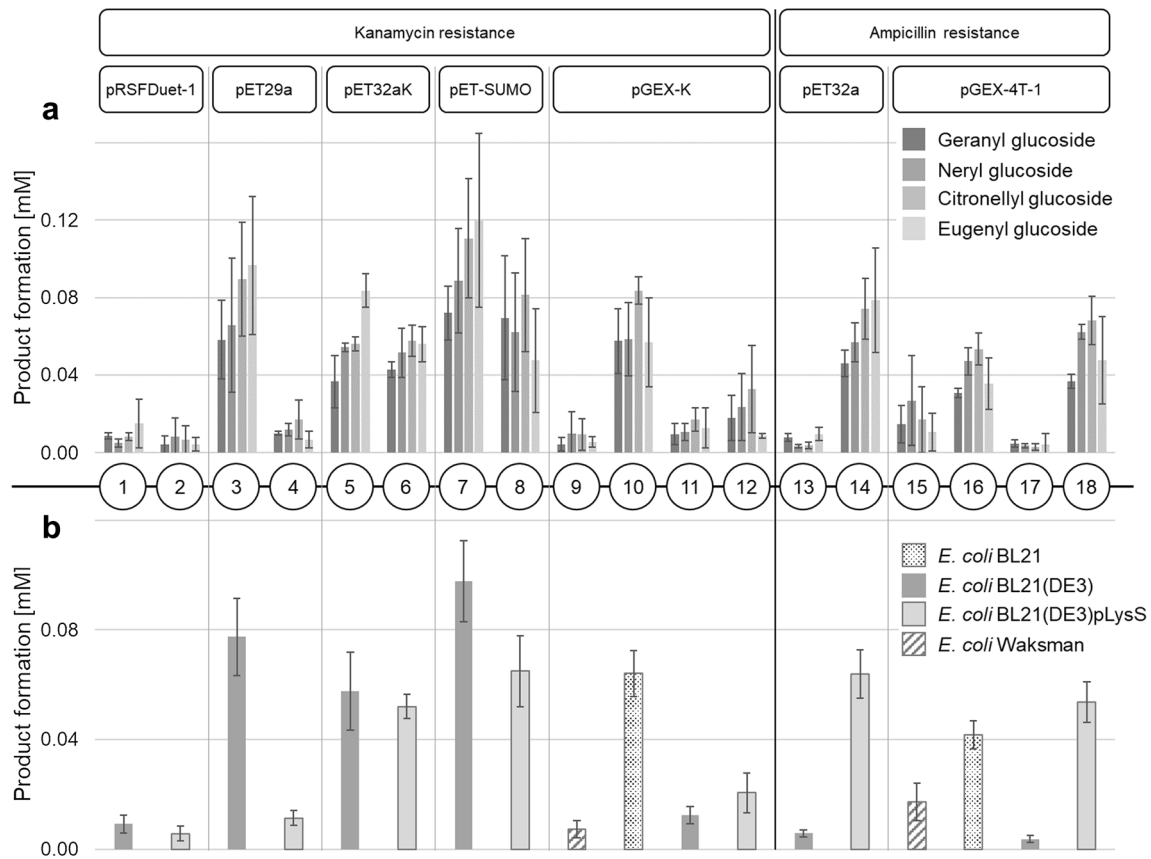
Fig. 3 Product formation of whole-cell biocatalysts separated by vector, substrate and strain. Product formation in [mM] of the 18 different strains separated by substrate (**a**). Products from left to right are geranyl glucoside, neryl glucoside, citronellyl glucoside, and eugenyl glucoside. Each experiment was done in duplicate and repeated three times. Combined product formation in [mM] (**b**). The different strains are separated by different bar styles: doted bars for *E. coli* BL21, dark grey bars for *E. coli* BL21(DE3), light grey bars for *E. coli* BL21(DE3)pLysS, and diagonally striped bars for *E. coli* Waks-

man. Strains 1 through 12 contained a kanamycin resistance gene while strains 13 through 18 contained an ampicillin resistance gene (**a**). The subcategories represent the different expression vectors and are separated by vertical grey lines. The expression vectors from left to right are pRSFDuet-1, pET29a, pET32aK, pET-SUMO, pGEX-K, pET32a, and pGEX-4T-1. Each expression vector contained the gene coding for VvGT14a from *Vitis vinifera*. The average glucoside yields are shown with the 95% confidence intervals

of 8.6 compared with the level obtained with the reference strain #4.

pGEX-K_VvGT14ao

The vector pGEX-K harboured the gene coding for GST upstream of the MCS. GST is a relatively large fusion protein and enhances the solubility of its fusion partner. Four different *E. coli* strains were transformed with the pGEX-K_VvGT14ao vector resulting in strains #9–#12 (Fig. 3). *E. coli* BL21 does not carry the gene for T7 RNA polymerase and thus is only suitable for expression from promoters recognized by the *E. coli* RNA polymerase. The *E. coli* Waksman strain is a very fast growing strain, produces low amounts of acetate and can be grown to high cell density during fed-batch culture with relative ease, and has good tolerance for environmental stresses. *E. coli* Waksman, *E.*

coli BL21(DE3), and *E. coli* BL21(DE3)pLysS combined with pGEX-K_VvGT14ao (Fig. 3, strains #9, #11, and #12, respectively) only showed low product levels. Strain #9 was the weakest producer, followed by strain #11, and then strain #12. Levels never exceeded 0.033 mM (citronellyl glucoside, strain #12). In contrast, *E. coli* BL21/pGEX-K_VvGT14ao (Fig. 3, strain #10) produced significant product titres ranging from 0.057 to 0.084 mM of eugenyl glucosides and citronellyl glucoside, respectively. The fusion protein pGEX-K_VvGT14ao produced high levels of glucosides, but only in the genetic background of *E. coli* BL21.

pET32a_VvGT14ao

Next, the effect of the antibiotic resistance gene on productivity of the whole-cell biocatalyst was determined. *E. coli* cells were transformed with suitable vectors and

biotransformations were performed in the presence of ampicillin instead of kanamycin (Fig. 3). *E. coli* BL21(DE3) and *E. coli* BL21(DE3)pLysS combined with pET32a_VvGT14ao (Fig. 3, strains #13 and #14, respectively), the vector expressing a VvGT14ao fusion protein with the thioredoxin protein on the N-terminus, exhibited significantly different glucosylation activities. Strains #13 and #14 produced 0.010 and 0.079 mM eugenyl glucoside, respectively. In this case, the pLysS plasmid, which produced T7 lysozyme appeared to enhance the production of glucosides.

A comparison of the titres produced by trxA_VvGT14ao fusion proteins in *E. coli* cells exhibiting kanamycin (strains #5 and #6) and ampicillin resistance (strains #13 and #14) showed similar glucoside levels, except for pET32a_VvGT14ao (strain #13), which generated only minor amounts of the products.

pGEX-4T-1_VvGT14ao

Finally, the productivity of whole-cell biocatalysts expressing GST fusion proteins in ampicillin-resistant *E. coli* strains was determined. *E. coli* Waksman and *E. coli* BL21(DE3) combined with pGEX-4T-1_VvGT14ao (Fig. 3, strains #15 and #17, respectively) produced only small amounts of glucosides (e.g. less than 0.027 mM of neryl glucoside, strain #15). In contrast, *E. coli* BL21 and *E. coli* BL21(DE3)pLysS expressing pGEX-4T-1_VvGT14ao (Fig. 3, strains #16 and #18, respectively) produced substantial amounts of products, in case of citronellyl glucoside up to 0.068 mM (strain #18). Similarly, GST_VvGT14ao fusion proteins yielded the highest titres in the kanamycin-resistant *E. coli* BL21 (strain #10) and BL21(DE3)pLysS (strain #12) genetic background.

Effect of pLysS and antibiotic resistance

To find the most suitable whole-cell biocatalyst for the biotechnological production of small molecule glucosides various combinations of hosts and expression vectors were tested that differed in resistance gene, origin of replication, promoter sequence, and fusion protein tag. Focusing on the kanamycin-resistant strains, a trend was observed. The majority of *E. coli* BL21(DE3)pLysS strains (strains #2, #4, #6, and #8) produced lower glucoside titres than their *E. coli* BL21(DE3) counterparts (strains #1, #3, #5, and #7). In both hosts, the expression of *VvGT14ao* prior to IPTG induction was inhibited by the lac-repressor lacI, which was constitutively expressed. In *E. coli* BL21(DE3)pLysS the plasmid pLysS constitutively expressed T7 lysozyme, which suppressed the basal expression of the target protein even further by inhibiting the T7 RNA polymerase. The tendency of *E. coli* BL21(DE3)pLysS strains to produce fewer glucosides was therefore explicable, as the residual basal expression of *VvGT14ao* should not be a heavy burden on the cell.

The low expression of LacI had a rather small impact on the metabolism of the cell unlike the propagation of an extra plasmid such as pLysS. The extra tight expression provided by pLysS is only needed for the production of toxic proteins.

Looking at the ampicillin-resistant strains, the above trend seemed to be reverse. Here, the *E. coli* BL21(DE3) strains formed less product than their counterparts with pLysS. This effect could also be seen with the strains combined with pGEX-4T-1_VvGT14ao (Fig. 3; strains #17 and #18) and pGEX-K_VvGT14ao (strains #11 and #12) and was also observed during the production of the chicken anaemia virus capsid protein VP1 (VP1). The *E. coli* strain BL21(DE3) pLysS yielded 6-times as much VP1 compared with *E. coli* BL21(DE3) using the same vector [20]. The pGEX-4T-1 and pGEX-K vectors did not carry the T7 promoter (Fig. 2) and neither the T7 RNA polymerase nor the T7 lysozyme could affect the transcription of *VvGT14ao*. Therefore, there might be a more complex interaction between pLysS and the heterologous protein production.

Effect of fusion tags

The fusion of VvGT14ao with a solubility-enhancing peptide tag appeared to have a positive effect on product titre, compared to the His-tag (strain #1 and #2). Maximum product yields were achieved with *E. coli* BL21(DE3)/pET-SUMO_VvGT14ao (strain #7), a host expressing *VvGT14ao* fused to the gene coding for SUMO. The addition of SUMO has already been proven to increase the yield of a target protein in *E. coli* such as the SARS-CoV 3CL protease [41] and of the rotavirus VP6 inner capsid protein (VP6) [5]. Similarly, the SUMO fusion tag could improve the solubility, yield and antigen-binding activity of the variable domains of the heavy chain of the heavy-chain antibodies [25]. *E. coli* BL21(DE3)/pET-SUMO_VvGT14ao (strain #7) produced 8 times more glucosides than the same host expressing the GST_VvGT14ao fusion (strain #11). However, the fusion tag sequence was not the only difference between the two expression vectors used. They also carried different promoters and had a widely different vector backbone. Therefore, the product increase was probably not due solely to the different fusion tag.

Effect of promoter

The pGEX vectors in *E. coli* BL21 (strains #10 and #16) produced considerably more glucosides than the same vectors in *E. coli* BL21(DE3) (strains #11 and #17). For example *E. coli* BL21/pGEX-K_VvGT14ao (strain #10) produced 5 times the amount of glucosides compared to *E. coli* BL21(DE3)/pGEX-K_VvGT14ao (strain #11). *E. coli* BL21 does not contain the (DE3) insertion and does not express the T7 RNA polymerase when induced, which poses less metabolic burden on the cell

metabolism. Since all other vectors contained the T7 promoter, which required the (DE3) insertion, this hypothesis could not be confirmed with additional data.

Cell growth

All strains grew to an optical density above 2.0 except for four strains: *E. coli* BL21(DE3)pLysS/pRSFDuet-1_VvG-T14ao (Supplemental Figure S9; Fig. 3, strain #2), *E. coli* Waksman/pGEX-K_VvGT14ao (Fig. 3, strain #9), *E. coli* BL21(DE3)/pGEX-K_VvGT14ao (Fig. 3, strain #11), and *E. coli* BL21(DE3)pLysS/pGEX-K_VvGT14ao (Fig. 3, strain #12). The three strains containing pGEX-K_VvGT14ao (Fig. 3, strains #9, #11, and #12) grew to a density of 50 – 70% of that of *E. coli* BL21 with pGEX-K_VvGT14ao (Fig. 3, strain #10). Even if the cell concentration is taken into account, *E. coli* BL21/pGEX-K_VvGT14ao (Fig. 3, strain #10) still exceeded the productivity of strains #9, #11, and #12. *E. coli* BL21(DE3)pLysS with pRSFDuet-1_VvG-T14ao (Fig. 3, strain #2) only grew to an optical density of barely above 5% of the most productive strain #7 (*E. coli* BL21(DE3)/pET-SUMO_VvGT14ao). Taking the cell concentration into account, *E. coli* BL21(DE3)pLysS/pRSF-Duet-1_VvGT14ao (Fig. 3, strain #2) would be slightly more productive than *E. coli* BL21(DE3)/pET-SUMO_VvGT14ao (Fig. 3, strain #7). However, the extremely low cell growth of strain #2 makes this strain unsuitable for the production of glucosides. The pRSFDuet-1 vector uses the RSF ori and is the only high-copy vector tested here. The pRSFDuet-1 vector does not seem suitable for large-scale glucoside production since both strains did not perform well. It has already been shown, that high-copy vectors are not suitable for protein expression unless the basal expression is suppressed adequately [11].

Suitable combinations

Since the conversions of various substrates did not show a clear trend toward a preferred acceptor molecule, the product titres of the four glucoside products were pooled (Fig. 3b), allowing a clearer distinction between the different strains. The different strains could be categorized into three groups according to their measured product yields. The categories are low performing (0–0.04 mM), average performing (0.04–0.08 mM), and high performing strains (0.08 mM and above).

Conclusions

Escherichia coli BL21 was a superior host for glucoside production using pGEX-K_VvGT14ao as an expression vector (strain #10) while *E. coli* BL21(DE3) was the preferred host for the production of glucosides using the remaining four expression plasmids containing a kanamycin resistance gene (pRSFDuet-1_VvGT14ao, pET29a_VvGT14ao, pET32aK_VvGT14ao, and pET-SUMO_VvGT14ao). The highest product titres were generated from *E. coli* BL21(DE3)pLysS when vectors carrying an ampicillin resistance (pET32a_VvGT14ao, and pGEX-4T-1_VvGT14ao) were used. The *E. coli* Waksman strain only produced small amounts of glucosides under the applied testing conditions. In summary, it is anticipated that the expression of a SUMO-VvGT14ao fusion protein in *E. coli* BL21 would result a highly efficient whole-cell biocatalyst for the production of terpenyl and phenyl β-D-glucopyranosides. The results of the HP cultivation can be employed to select promising production strains for scale-up experiments in shaking flasks. The system can also be adapted to test other production strains for various whole-cell biotransformation applications.

References

1. Bangen K, Hong N, Louie A, Mei G, Trewartha J (2004) The Effects of Incubation with Ampicillin and Tetracycline on the Expression of the bla and tetA genes of pBR322. J Exp Microbiol Immunol 5:23–28
2. Bönisch F, Frotscher J, Stanitzek S, Ruhl E, Wust M, Bitz O, Schwab W (2014) Activity-based profiling of a physiologic aglycone library reveals sugar acceptor promiscuity of family 1 UDP-glucosyltransferases from grape. Plant Physiol 166:23–39
3. Boros I, Pósfai G, Venetianer P (1984) High-copy-number derivatives of the plasmid cloning vector pBR322. Gene 30:257–260
4. Brosius J, Erfle M, Storella J (1985) Spacing of the -10 and -35 regions in the tac promoter. Effect on its in vivo activity. J Biol Chem 260:3539–3541
5. Bugli F, Caprettini V, Cacaci M, Martini C, Paroni Sterbini F, Torelli R, Della Longa S, Papi M, Palmieri V, Giardina B, Posteraro B, Sanguinetti M, Arcovito A (2014) Synthesis and characterization of different immunogenic viral nanoconstructs from rotavirus VP6 inner capsid protein. Int J Nanomed 9:2727–2739
6. Chakravartty V, Cronan JE (2015) A series of medium and high copy number arabinose-inducible *Escherichia coli* expression vectors compatible with pBR322 and pACYC184. Plasmid 81:21–26
7. De Bruyn F, Maertens J, Beauprez J, Soetaert W, De Mey M (2015) Biotechnological advances in UDP-sugar based glycosylation of small molecules. Biotechnol Adv 33:288–302
8. de Carvalho CC (2017) Whole cell biocatalysts: essential workers from Nature to the industry. Microb Biotechnol 10:250–263
9. Dror A, Shemesh E, Dayan N, Fishman A (2014) Protein engineering by random mutagenesis and structure-guided consensus of Geobacillus stearothermophilus lipase T6 for enhanced stability in methanol. Appl Environ Microbiol 80:1515–1527
10. Geuns JM (2003) Stevioside. Phytochemistry 64:913–921
11. Gruber DF, Pieribone VA, Porton B, Kao H-T (2008) Strict regulation of gene expression from a high-copy plasmid utilizing a dual vector system. Protein Expr Purif 60:53–57

12. Haun RS, Serventi IM, Moss J (1992) Rapid, reliable ligation-independent cloning of PCR products using modified plasmid vectors. Biotechniques 13:515–518

13. Huang FC, Hinkelmann J, Schwab W (2015) Glucosylation of aroma chemicals and hydroxy fatty acids. J Biotechnol 216:100–109

14. Jackson R, Knisley D, McIntosh C, Pfeiffer P (2011) Predicting flavonoid UGT regioselectivity. Adv Bioinform 2011:506583

15. Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi SH, Couloux A, Lee SW, Yoon SH, Cattolico L, Hur CG, Park HS, Segurens B, Kim SC, Oh TK, Lenski RE, Studier FW, Daegelen P, Kim JF (2009) Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). J Mol Biol 394:644–652

16. Jones P, Vogt T (2001) Glycosyltransferases in secondary plant metabolism: tranquilizers and stimulant controllers. Planta 213:164–174

17. Kaşıkcı MB, Bağdatlıoğlu N (2016) Bioavailability of quercetin. Curr Res Nutr Food Sci J 4:146–151

18. Kirk O, Borchert TV, Fuglsang CC (2002) Industrial enzyme applications. Curr Opin Biotechnol 13:345–351

19. Knorre D, Kudryashova N, Godovikova T (2009) Chemical and functional aspects of posttranslational modification of proteins. Acta Naturae 3:29–51

20. Lee M-S, Hseu Y-C, Lai G-H, Chang W-T, Chen H-J, Huang C-H, Lee M-S, Wang M-Y, Kao J-Y, You B-J (2011) High yield expression in a recombinant *E. coli* of a codon optimized chicken anemia virus capsid protein VP1 useful for vaccine development. Microb Cell Fact 10:56

21. Lepak A, Gutmann A, Kulmer ST, Nidetzky B (2015) Creating a water-soluble resveratrol-based antioxidant by site-selective enzymatic glucosylation. ChemBioChem 16:1870–1874

22. Li Y, Baldauf S, Lim E-K, Bowles DJ (2001) Phylogenetic analysis of the UDP-glycosyltransferase multigene family of *Arabidopsis thaliana*. J Biol Chem 276:4338–4343

23. Liang D-M, Liu J-H, Wu H, Wang B-B, Zhu H-J, Qiao J-J (2015) Glycosyltransferases: mechanisms and applications in natural product development. Chem Soc Rev 44:8350–8374

24. Lin B, Tao Y (2017) Whole-cell biocatalysts by design. Microb Cell Fact 16:106

25. Z-h Liu, Huang D, X-j Fu, Cheng P, E-q Du (2018) Comparison of three commonly used fusion tags for the expression of nanobodies in the cytoplasm of *Escherichia coli*. Biotechnol Biotechnol Equip 32:462–469

26. Marblestone JG, Edavettal SC, Lim Y, Lim P, Zuo X, Butt TR (2006) Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. Protein Sci 15:182–189

27. Parsell DA, Silber KR, Sauer RT (1990) Carboxy-terminal determinants of intracellular protein degradation. Genes Dev 4:277–286

28. Schwab W, Fischer T, Wüst M (2015) Terpene glucoside production: improved biocatalytic processes using glycosyltransferases. Eng Life Sci 15:376–386

29. Schwab W, Fischer TC, Giri A, Wüst M (2015) Potential applications of glucosyltransferases in terpene glucoside production: impacts on the use of aroma and fragrance. Appl Microbiol Biotechnol 99:165–174

30. Härtl K, MacGraphery K, Rüdiger J, Schwab (2018) Tailoring Natural Products with Glycosyltransferases. In: Schwab W, Lange BM, Wüst M (eds) Biotechnology of natural products, 1st edn. Springer, Heidelberg, pp 219–263

31. Sivashanmugam A, Murray V, Cui C, Zhang Y, Wang J, Li Q (2009) Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. Protein Sci 18:936–948

32. Soufi L, Hoseini-Alfatemi SM, Sharifi-Rad M, Iriti M, Sharifi-Rad M, Hoseini M, Sharifi-Rad J (2015) Recombinant proteins in *Escherichia coli*: optimizing production strategies. Am Eurasian J Agric Environ Sci 15:2149–2159

33. Steinmetz EJ, Auldridge ME (2017) Screening fusion tags for improved recombinant protein expression in *E. coli* with the Expresso® solubility and expression screening system. Curr Protoc Protein Sci 90:5.27.1–25.27.20

34. Stetz MA, Carter MV, Wand AJ (2016) Optimized expression and purification of biophysical quantities of Lac repressor and Lac repressor regulatory domain. Protein Expr Purif 123:75–82

35. Tegel H, Ottosson J, Hober S (2011) Enhancing the protein production levels in *Escherichia coli* with a strong promoter. FEBS J 278:729–739

36. Thomas JG, Ayling A, Baneyx F (1997) Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from *E. coli*. Appl Biochem Biotechnol 66:197–238

37. Wilson B, Kautzer C, Antelman D (1994) Increased protein expression through improved ribosome-binding sites obtained by library mutagenesis. Biotechniques 17:944–953

38. Yamamoto I, Muto N, Murakami K, Suga S, Yamaguchi H (1990) L-Ascorbic acid α-glucoside formed by regioselective transglucosylation with rat intestinal and rice seed α-glucosidases: its improved stability and structure determination. Chem Pharm Bull 38:3020–3023

39. Zhao CL, Chen ZJ, Bai XS, Ding C, Long TJ, Wei FG, Miao KR (2014) Structure–activity relationships of anthocyanidin glycosylation. Mol Divers 18:687–700

40. Zhao X, Li G, Liang S (2013) Several affinity tags commonly used in chromatographic purification. J Anal Methods Chem 2013:581093. https://doi.org/10.1155/2013/581093

41. Zuo X, Mattern MR, Tan R, Li S, Hall J, Sterner DE, Shoo J, Tran H, Lim P, Sarafianos SG (2005) Expression and purification of SARS coronavirus proteins using SUMO-fusions. Protein Expr Purif 42:100–110