Received: 10 March 2021 | Revised: 28 May 2021 | Accepted: 21 June 2021

DOI: 10.1002/eng2.12440

RESEARCH ARTICLE

Engineering Reports Open Access WILEY

Improvement of an *Escherichia coli* whole-cell biocatalyst for geranyl glucoside production using directed evolution

Julian Rüdiger | **Wilfried Schwab**

Biotechnology of Natural Product, Technical University Munich, Freising, Germany

Correspondence
Wilfried Schwab, Biotechnology of Natural Product, Technical University Munich, Liesel-Beckmann-Str. 1, Freising 85354, Germany.
Email: wilfried.schwab@tum.de

Abstract

Biotechnological production of glycosides is an economically competitive manufacturing alternative to classical chemical synthesis. Due to continuous improvement in production, glycosides can now be used in low-cost products by various industries. However, many production systems still suffer from low yields. Directed evolution, coupled with a suitable screening method, can tackle this challenge. We generated glycosyltransferase mutants through error-prone-PCR and screened the library using a small-scale whole-cell biotransformation system to identify highly productive strains. The screening of only 176 colonies yielded three putative candidates. Detailed analysis revealed that the reason for the increase in product titer was mainly due to different expression effects of the mutant genes rather than improved enzyme kinetics. An up to 60-fold increase in whole-cell product quantity was achieved. Therefore, in addition to the quality of the mutant library, an efficient and stable expression system is crucial to achieve high concentrations of active enzyme and product, as formation of inclusion bodies and other inactive forms of the biocatalyst reduces productivity.

KEYWORDS

directed evolution, *Escherichia coli*, glycosylation, random mutagenesis, whole-cell biocatalysis

1 | INTRODUCTION

Glycosides are conjugates of one or more sugars and a secondary molecule, which is referred to as the aglycone. The aglycone gains polarity due to the sugar moiety, resulting in higher water solubility, and altered bioavailability of the glycoside.[1] Therefore, glycosylation modulates the effect of pharmaceuticals.[2] Glycosides are currently used as sweeteners,[3] and as drugs.[4,5] The use as industrial water treatment agent was also proposed.[6]

Classical chemical synthesis can only be used for the production of high-value glycosides because the synthesis requires multiple protection and deprotection steps. Protective group chemistry coupled with long reaction times and low yields results in high production costs.[7] Low-value glycosides are not available to industry. However, as biotechnological production offers a cost-effective alternative to classical chemical synthesis, low-cost glycosides are gaining importance for numerous applications.[8] Identification and characterization of new glycosyltransferases (GTs) provide more efficient enzymes for their production[9] and enables a growing variety of possible glycoside products.[10]

IGSSE project GLUTAIL

GTs transfer an activated sugar-moiety to an acceptor molecule. Activated sugars are expensive fine chemicals and have to be recycled in order to establish an economical process.[11,12] One of the largest groups of GTs are the UDP-glycosyltransferases (UGTs), which use uridine diphosphate (UDP)-sugars as a donor.[13] A number of plant UGTs have already been characterized and they produce economically interesting glycosides such as flavonoids, terpenols, and phenolic compounds.[14,15]

UGTs can be produced by, and purified from genetically modified *Escherichia coli* cells. The in vitro production of glycosides has been improved by adding additional enzymes for UDP-sugar regeneration.[8,16] Nevertheless, the in vitro production of glycosides still requires purified enzymes and starting concentrations of UDP-glucose or UDP. Direct use of UGT-expressing *E. coli* cells as whole-cell biocatalysts may provide an even more cost-effective solution.[17] However, some whole-cell biocatalysts show inefficient conversion rates.[8] Poor transformation rates can be a result of low substrate availability, toxicity of the substrate or product, low enzyme concentration, or the use of inefficient enzymes.[18–20] Enzyme activity can be improved through directed evolution coupled with an efficient screening system.[21–23] Whole-cell biocatalysts offer unique advantages, including endogenous cofactor regeneration, and are widely used for the efficient production of fine and bulk chemicals and active pharmaceutical ingredients. In addition, advances in synthetic biology and metabolic engineering have led to a renaissance in whole-cell biocatalysis, which can enable the manufacture of value-added products from cheap feedstock.[24] Genetic modification of the rhamnose biosynthesis pathway by overexpression of dTDP-rhamnose synthesis genes (*rmlABCD*) yielded 4.3 g/L quercetin 3-O-α-L-rhamoside when rhamnose was fed to *E. coli* MG1655 harboring GtfC.[25] Only 0.35 g/L cyanidin 3-*O*-β-D-glucoside was obtained from catechin by whole-cell biotransformation although the intracellular UDP-α-D-glucose pool was augmented,[26] while 3.9 g/L quercetin-3-*O*-β-glucoside was produced by *E. coli* MEC367 (MG1655 pgi) expressing UGT73B3 from quercetin.[27] Furthermore, *E. coli* cells transformed with *GtUF6CGT1* yielded only 42 mg/L chrysin 6-C-β-D-glucoside from chrysin, even though the UDP-α-D-glucose pool was genetically enhanced by the incorporated of several carbohydrate metabolism genes.[28] The highest titer of 7.2 g/L was obtained for maple furanone ß-D-glucoside by *E. coli* Waksman cells transformed with *UGT71K3a*. Because genetic modification of the UDP-α-D-glucose pool was not necessary to achieve this titer, the UDP-α-D-glucose pool appears not to be the rate-limiting factor in glucoside formation in *E. coli*.[29] Optimization of microbial cell factories concentrate on improving heterologous pathway flux, precursor supply, cofactor balance, as well as other aspects of cellular metabolism, to enhance the efficiency of biocatalysts.[30] When the whole-cell biocatalyst product is exported directly into the surrounding medium, the efficiency of the cell factories can be easily determined by analytical methods, making the system very convenient for screening mutants.

In this study, a biocatalyst composed of *E. coli* BL21(DE3) and *VvGT14a* from grape (*Vitis vinifera*) was optimized using a directed evolution approach to improve the product titer of terpenol glucosides by whole-cell biotransformation. VvGT14a shows activity towards a variety of terpenols and other alcohols.[31] The codon-optimized gene *VvGT14ao* was subjected to error-prone PCR (epPCR) according to a published protocol[32]. The PCR product was cloned into pRSFDuet-1 and *E. coli* BL21(DE3) was transformed with the resulting library. This library was screened by medium-throughput screening of the whole-cell biocatalysts coupled with LC-UV as readout. The goal of this screening was to find mutant strains with improved geranyl glucoside yields. The strain *E. coli* BL21(DE3) /pRSFDuet-1_VvGT14ao was used as a wild type reference during the screening. The results show that not only improvements in the activity of the protein, but also optimizations of the host are equally important to increase the product yield of the whole cell approach.

2 | MATERIALS AND METHODS

2.1 | Chemicals, enzymes, and expression vectors

Chemicals were purchased from Sigma-Aldrich (Taufkirchen, Germany), Merck (Darmstadt, Germany), VWR (Darmstadt, Germany), Acros Organics (Geel, Belgium), and Carl Roth (Karlsruhe, Germany). DNA oligos were purchased from Eurofins Genomics Germany GmbH (Ebersberg, Germany; Table S1) and metabion international AG (Planegg, Germany; Table S2). Q5® High-Fidelity DNA Polymerase was obtained from New England Biolabs (Frankfurt am Main, Germany), FastAP Thermosensitive Alkaline Phosphatase, FastDigest *Aat*II, FastDigest *Bam*HI, FastDigest *Bsa*I, FastDigest *Eco*31I, FastDigest *Not*I, FastDigest *Ssp*I, and T4 DNA polymerase were from Thermo Fisher Scientific (Planegg, Germany), and T4 DNA ligase was purchased from Promega (Mannheim, Germany). Expression

vectors pET-29a(+) (pET29a), pET-32a(+) (pET32a) and pRSFDuet-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).

2.2 | Nucleic acid extraction

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

2.3 | Creation of pET32aK and pGEX-K

pET32aK and pGEX-K were created from pET32a and pGEX-4T-1, respectively. Their ampicillin resistance cassette was exchanged with the kanamycin resistance cassette from pRSFDuet-1.[33]

2.4 | Cloning of VvGT14ao, and VvGT14ao-S168N-W353R in pET29a, pET32aK, pGEX-K, and pRSFDuet-1

The gene *VvGT14ao* was amplified by PCR using primers with *Bam*HI- and *Not*I-extensions (for pRSFDuet-1 the forward primer GT14ao-Bam_+1_fw, for pET29a, pET32aK, and pGEX-K forward primer GT14ao-BamHI-fwd; reverse primer GT14ao-NotI_rev; Table S1). The PCR product and the empty vectors were digested by *Bam*HI and *Not*I. The linearized vectors were dephosphorylated by using FastAP Thermosensitive Alkaline Phosphatase to prevent self-ligation. The ligation was performed by T4 DNA ligase with a vector-to-insert ratio of 1:3 (n/n). The cloning of the gene coding for VvGT14ao-S168N-W353R in pET29a, and pGEX-K was carried out in the same way.

2.5 | Cloning of VvGT14ao in pET-SUMO

The pET-SUMO_VvGT14ao construct was created using ligation-independent cloning.[33]

2.6 | Cloning of VvGT14ao-S168N and VvGT14ao-W353R

Point mutations were created by using the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, USA). Primers SDM-VvGT14a-SN-fwd and SDM-VvGT14a-SN-rev were used to obtain the S168N variant of VvGT14ao, and SDM-VvGT14a-WR-fwd and SDM-VvGT14a-WR-rev were used for the W353R variant.

2.7 | Cloning of SN, WR, and SN-WR mutants of VvGT15a, and VvGT16

The SN variant was created by two PCR steps. The first PCR step was used to amplify two fragments and introduce the point mutation in the overlapping region. Primers VvGT15a-Fwd and VvGT15a-SN-rev were used for the first fragment of VvGT15a, and VvGT15a-SN-fwd and VvGT15a-Rev were used for the second fragment. The two fragments were applied in the second PCR together with the end primers VvGT15a-Fwd and VvGT15a-Rev. The second PCR reaction yielded the complete gene with the SN mutation. The entry vector pGEX-4T-1 was cut and subsequently dephosphorylated by FastAP Thermosensitive Alkaline Phosphatase. The cloning of the DNA fragment was carried out with restriction using *Bam*HI and *Not*I, and ligation using T4 DNA ligase with a vector-to-insert ratio of 1:3 (n/n). The cloning of the WR mutation was carried out analogously, but with VvGT15a-WR-rev and VvGT15a-WR-fwd primers instead of VvGT15a-SN-rev and VvGT15a-SN-fwd, respectively. The SN-WR variant was also

created by two PCR steps. However, the first PCR step was used to amplify three fragments and introduce both point mutations in the overlapping regions. Primers VvGT15a-Fwd and VvGT15a-SN-rev were used for the first fragment of VvGT15a, VvGT15a-SN-fwd, and VvGT15a-WR-rev were used for the second fragment, and VvGT15a-WR-fwd and VvGT15a-Rev were applied for the third fragment. The three fragments were used in the second PCR together with the end primers VvGT15a-Fwd and VvGT15a-Rev. Like before, the second PCR reaction yielded the finished gene. The cloning into the entry vector was carried out analogously to the single mutants. The cloning of the VvGT16 variants was similar to that of the VvGT15a variants, except that VvGT16 primers were used instead of the VvGT15a primers.

2.8 | Sequencing

New constructs were sequenced by Eurofins Genomics GmbH (Ebersberg, Germany) and confirmed by DNA sequence alignment by Serial Cloner (http://serialbasics.free.fr/Serial_Cloner.html; version 2.6.1). The genome sequencing was performed by Eurofins Genomics GmbH and the sequencing of the RT-PCR products by GENEWIZ Germany (Leipzig, Germany).

2.9 | Minimal medium

M9 minimal medium was produced according to a published protocol (https://www.helmholtz-muenchen.de/fileadmin/PEPF/Protocols/M9-medium_150510.pdf).

2.10 | Creation of the mutant library

The mutant library was created by using the GeneMorph II Random Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, USA). As template 3.7 µg pRSFDuet-1_VvGT14ao were used for the epPCR, containing 1000 ng of target DNA. The PCR reaction was performed with 30 cycles, using the primers GM2_RSF-14ao_fwd and GM2_RSF-14ao_rev. The PCR product was gel-purified and cloned in pRSFDuet-1 using restriction with *Bam*HI and *Not*I followed by ligation using T4 DNA ligase.

2.11 | Library screening using hitplate cultivation

In vivo glycosylation was performed using hitplate cultivation as described.[33] One milliliter sample was transferred from each well to a microreaction tube and was centrifuged. The supernatant was analyzed by LC-UV as described.[33]

2.12 | Plasmid curing

In order to remove plasmids from an *E. coli* strain, a cultivation under temperature stress without selective pressure was carried out. Lysogeny broth (LB medium; 10 g/L tryptone, 5 g/L yeast extract, and 10 g/L sodium chloride) was used as culture medium, and the cultivation was carried out at 42°C without the use of antibiotics.[34] A small quantity (approximately 200 cells) from each passage was spread onto agar plates without antibiotics. Single colonies were collected on two collection plates, one with and one without the previously needed antibiotics. The antibiotic-free plates yielded one colony for each picked one. The agar plate with antibiotics served as control and showed which culture still contained the plasmid. More than 10 passages were required to cure the strains.

2.13 | Protein production

Enzymes for in vitro experiments were purified from crude cell extracts using affinity chromatography according to.[31]

2.14 | In vitro analysis

The glycosylation reaction was analyzed with the UDP-Glo™ Glycosyltransferase Assay (Promega, Mannheim, Germany) according to the manufacturer's protocol. The luminescence of the samples was measured after an incubation period of 60 min in the dark using the CLARIOstar plate reader (BMG LABTECH GmbH, Ortenberg, Germany).

2.15 | SDS gel

The protein expression was determined by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE). The stacking and resolving gel contained 5% and 12% acrylamide, respectively. Equal volumes of the culture and 2x Laemmli Buffer (Bio-Rad, Munich, Germany) were mixed and boiled at 95°C for 5 min. The gels were loaded with 10 µL sample per well and run at 100 V for 3 h. The gels were stained overnight with a colloidal coomassie staining solution (50 g/L aluminum sulfate-18-hydrate, 200 mL/L ethanol, 1 g/L coomassie blue G250, and 23.5 mL/L ortho-phosphoric acid) and destained with water for a few hours. The glutathione S-transferas (GST)-GT bands were confirmed via western blot using anti-GST antibody and quantified using ImageJ (https://fiji.sc/).

2.16 | Reverse transcription polymerase chain reaction

The RT-PCR was performed with the SuperScript IV One-Step RT-PCR System from Thermo Fisher Scientific (Planegg, Germany). The RT-PCR products were analyzed by gel electrophoresis on a gel containing 1% (w/v) agarose.

2.17 | Complementary DNA synthesis

The complementary DNA (cDNA) was synthesizes using the iScript™ cDNA Synthesis Kit from Bio-Rad Laboratories (Feldkirchen, Germany).

2.18 | Real-time quantitative PCR

The real-time quantitative PCR (qPCR) was performed with the OneStepPlus Real-Time PCR system from Applied Biosystems (Foster City, CA, USA) using the SensiFAST™ SYBR® Hi-ROX Kit from Bioline GmbH (Luckenwalde, Germany).

3 | RESULTS

A guided evolution approach was employed to increase the geranyl glucoside yield from a whole-cell biocatalyst using VvGT14a from grape (*V. vinifera*). The gene *VvGT14ao* was subjected to an epPCR using 1000 ng of target DNA and 30 amplification cycles. The DNA library was obtained by cloning the PCR product into pRSFDuet-1 and *E. coli* NEB10β was transformed with the library. The plasmid DNA of several transformants was sequenced. The mutation rate was analyzed at the protein level and was described as amino acid exchanges per gene (AAE). The mutant library showed an average mutation rate of 0.63 AAE, resulting in a distribution of 37% 0 AAE, 43% 1 AAE, 17% 2 AAE, and 3% 3 AAE variants in the library.

E. coli BL21(DE3) was transformed with the DNA library for the mutant screening. The transformants were distributed on collection plates with petri dish stickers displaying a 100-square grid. Each collection plate contained 99 putative mutants and one colony of *E. coli* BL21(DE3) /pRSFDuet-1_VvGT14ao as reference. The transformation and collection was repeated several times, to have fresh cell material at hand for the screening. The putative mutants were tested by hitplate cultivation for their ability to produce geranyl glucoside. The hitplates were loaded with the reference strain and 11 putative mutants, each in duplicate. The geranyl glucoside concentration was quantified in the supernatant by LC-UV. In this way, the performance of the whole-cell process was investigated and not merely the catalytic activity of the mutant protein alone. The screening results were sorted by decreasing product yield (Figure 1).

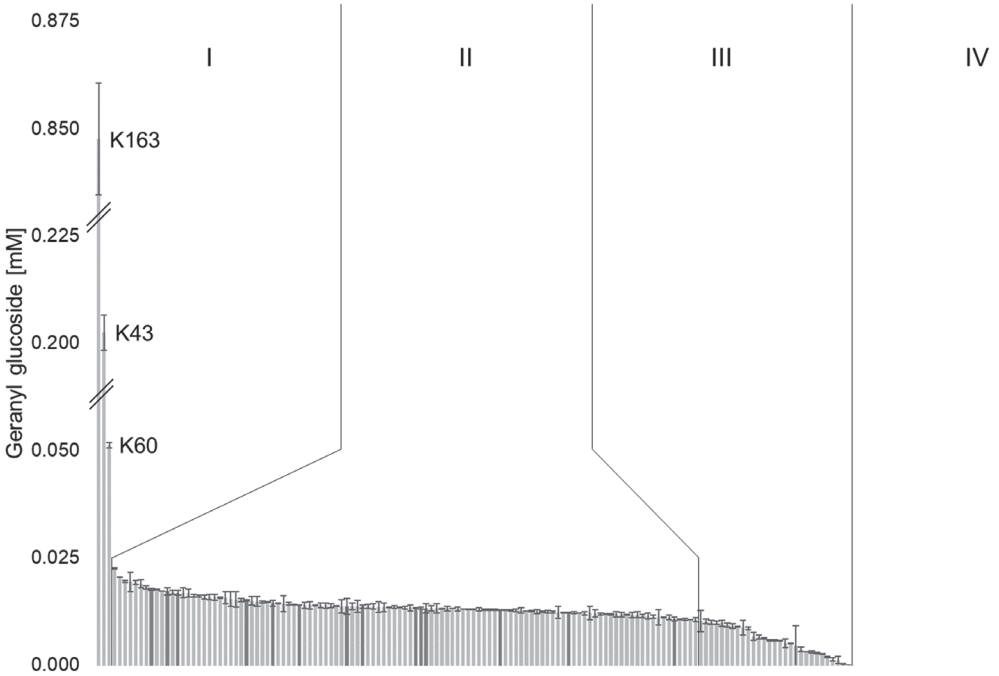

FIGURE 1 Screening of the mutant library. The geranyl glucoside yield of 192 colonies was determined in 16 screening plates each containing 11 putative mutants (light gray) and one wild type reference (dark gray) in duplicate. In total 176 putative mutants were screened. The error bars show the SD of the replicates. Superior mutants (I), wild type strains and neutral mutants (II), inferior mutants (III), and loss-of-function mutants and empty vector strains (IV)

The resultant strains could be categorized into four distinct groups. Group I (superior mutants) consisted of three putative mutants that produced significantly elevated concentrations of geranyl glucoside. Group II (wild type strains and neutral mutants) comprised most of the cultures studied. These candidates showed an activity similar to the wild type references. The mutations occurring in these candidates had no effect on the geranyl glucoside yield. Group III (inferior mutants) contained candidates with a lower activity than the wild type strains. Therefore, the mutations present in those strains were unfavorable and resulted in an inferior production system. The members of group IV (loss-of-function mutants and empty vector strains) showed no measurable geranyl glucoside production. These candidates either contained a mutation, which destroyed the activity of VvGT14a, or they harbored an expression vector without the GT gene (empty vector strains).

The reference strain *E. coli* BL21(DE3) /pRSFDuet-1_VvGT14ao produced 0.014 mM of geranyl glucoside on average. The three screening hits from group I, K60, K43, and K163, produced 0.051, 0.203, and 0.848 mM of geranyl glucoside in the initial screening, respectively. The plasmids from group I were isolated and sequenced to elucidate the specific mutations.

3.1 | Screening hit K60

K60 contained a mixture of expression vectors. Two independent sequencing analyses showed the empty vector sequence and the elevated product yield was confirmed by two independent biotransformation experiments. The strain was cured of the expression vectors, and retransformed with pRSFDuet-1_VvGT14ao. This strain showed only the activity of the wild type strain. The GT gene was detected in the original K60 strain by colony PCR (cPCR) using *VvGT14ao* specific primers. Sequencing of the cPCR product showed a mutation within *VvGT14ao* resulting in the variant VvGT14ao-T398S. This mutation is located two positions downstream of the plant secondary product glycosyltransferase (PSPG) motif in the C-terminal direction of the enzyme. The PSPG motif is responsible for binding of the sugar donor.[17,35] Furthermore, the sequencing chromatogram revealed a mixture of the variant and the wild type gene in the cPCR product. Initial sequencing identified the empty vector sequence; therefore, we concluded that the majority of the plasmids within K60 were empty vectors. According to the cPCR sequencing data the second most abundant plasmid was pRSFDuet-1_VvGT14ao-T398S

followed by the wild type vector. The T398S variant was cloned and analyzed with *E. coli* BL21(DE3) as a host. However, the product concentration was roughly 6.5 times lower than the yield of the wild type strain.

3.2 | Screening hit K43

K43 showed the wild type *VvGT14ao* sequence without mutations. Resequencing confirmed the initial result and the increased activity of K43 was verified by further experiments. The strain K43 was also cured of the expression vector. The cured strain, *E. coli* BL21(DE3)K43, showed no sign of product formation. Upon retransformation with pRSFDuet-1_VvGT14ao, the product yield again reached the elevated levels, indicating a beneficial mutation in the genomic background of the production host. The genome of *E. coli* BL21(DE3)K43 was sequenced and compared with the wild type laboratory strain. The sequencing reaction revealed several single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) located in five loci. The large number of genomic mutations could be a result of the curing process, which may have added some genomic mutations. However, the mutation attributing to the increase in product yield must be within the five loci as the retransformed strain also showed the increase in yield.

Locus 1 contained one SNP and was located at the position 110,370 in the *E. coli* BL21(DE3) genome (AM946981; http://www.ebi.ac.uk/Tools/dbfetch/expasyfetch?AM946981). The SNP was between *lpxC* (an essential zinc-dependent deacetylase of bacterial lipid A synthesis) and *secM* (regulates production of cotranslated SecA), but not within any identified regulatory element. This mutation should therefore not affect the glycosylation ability of the host.

The second locus contained six SNPs and one insertion in the region between the positions 2,929,276 and 2,929,309. The insert was rather small, with only three base pairs added. The locus is downstream of two coding sequences, *yggW* (probably acts as a heme chaperone, transferring heme to an unknown acceptor) and *yggM* (uncharacterized). Similar to locus 1, the second locus was also not within any known regulatory element and should therefore be irrelevant for the glycosylation process.

Locus 3 only contained one SNP and was located at position 2,964,343, between the coding sequences *pulK* (minor pseudopilin, Type II secretion system protein K) and *yghE* (putative type II secretion system L-type protein). Because there is no promoter or terminator between the two coding sequences, this mutation should not affect the glycosylation efficiency of *E. coli* BL21(DE3)K43.

The fourth locus was the largest with 41 SNPs and six indels. This locus was located within the coding sequence of *yghJ*, between 2,975,823 and 2,977,059 in the *E. coli* BL21(DE3) genome. The expressed protein yghJ or acfD (accessory colonization factor) is a lipoprotein anchored in the cell membrane. This protein contains a M60 protease domain and showed mucinase activity in pathogenic *E. coli* strains.[36,37] Thirty-two SNPs and four indels were located within the M60 metalloprotease domain. Even though the gene contained multiple SNPs and indels, it still coded for an uninterrupted yghJ variant. The variant enzyme contained 17 AAE compared with the laboratory strain. There is no apparent link between the mucinase activity and the GT activity.

The last locus was at position 3,857,938 and contained a large insertion of 1338 base pairs. The insert contained the promoter region and coding sequence from an IS4-like element ISVsa5 family transposase (NCBI, Standard Nucleotide BLAST; 100% identity according to https://www.ncbi.nlm.nih.gov/nucleotide/CP040667.1?report=genbank&log$=nuclalign&blast_rank=1&RID=RCHP6UZM01R&from=590392&to=591597). The transposase gene was inserted into the unpaired loop of the *rho* transcription terminator, rendering it useless. Therefore, we concluded that the following operon could be cotranscribed with *rho*. This operon is approximately 13 kb long and contained 12 genes: *rfe*, *wzzE*, *rffE*, *rffD*, *rffG*, *rffH*, *rffC*, *rffA*, *wzxE*, *rffT*, *wzyE*, and *rffM*. The encoded enzymes are involved in the synthesis and transport of lipid III, and the formation of the cyclic enterobacterial common antigen[38]. RT-PCR showed the cotranscription of *rho* and the following operon in *E. coli* BL21(DE3)K43, but also in the wild type strain (Figure S1). In addition, qPCR proved that the transcripts of the following operon were much more abundant in the K43 strain than in the wild type strain (Figure 2(A)). By contrast, the abundance of the GT transcript was reduced by a factor of 1.6 in the K43 strain compared with the wild type strain (Figure 2(B)).

The new glycoside production host, *E. coli* BL21(DE3)K43, was transformed with a variety of different expression vectors to further test the beneficial genomic background mutation. The tested expression vectors include pET29a_VvGT14ao, pET32aK_VvGT14ao, pET-SUMO_VvGT14ao, and pGEX-K_VvGT14ao. The K43 strains were tested against the corresponding *E. coli* BL21(DE3) strains (Figure 3). Except for pGEX-K_VvGT14ao, all other

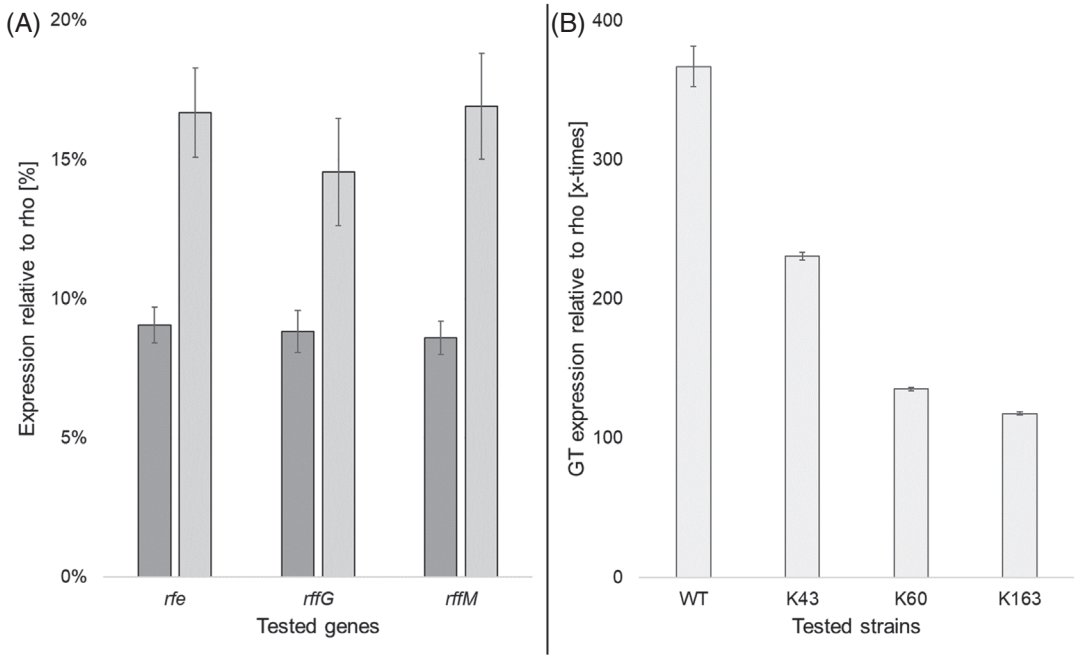

FIGURE 2 Quantitative real-time PCR analysis. Gene expression levels were calculated relative to rho expression. (A) Quantification of rfe, rffG, and rffM transcripts of *E. coli* BL21(DE3) (dark gray) and of *E. coli* BL21(DE3)K43 (medium gray). (B) Quantification of VvGT14ao transcripts (light gray) in *E. coli* BL21(DE3)/pRSFDuet-1_VvGT14ao (WT), and the three screening hits K43, K60, and K163. Samples were taken 20 h after IPTG induction. The error bars represent the SD of the duplicates

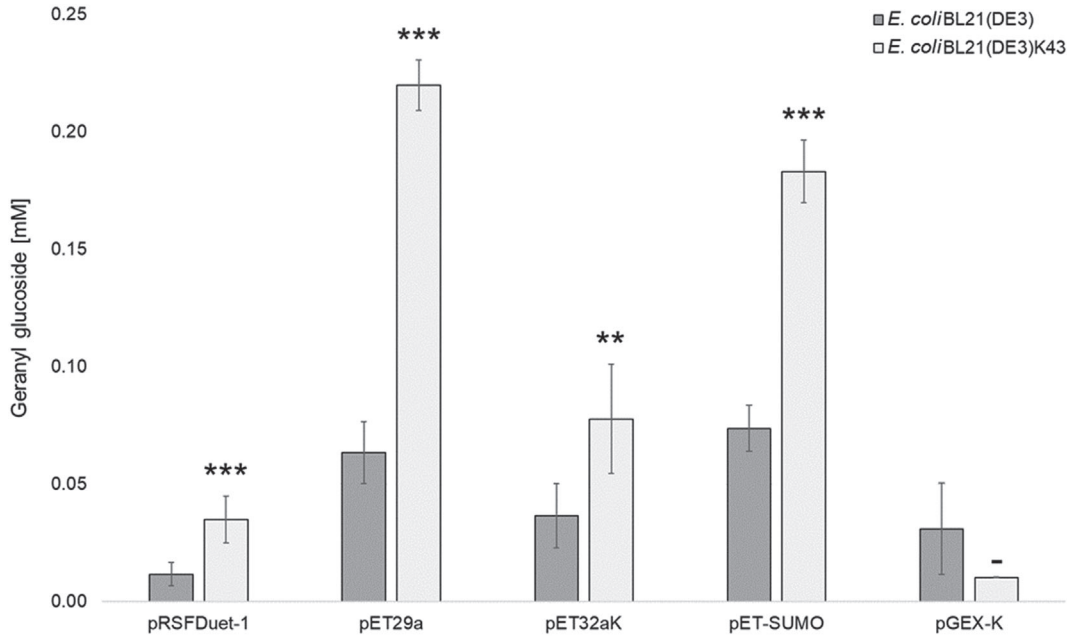

FIGURE 3 Glucoside yield of *E. coli* BL21(DE3) and *E. coli* BL21(DE3)K43 strains combined with different expression vectors containing VvGT14ao. The error bars represent the confidence intervals calculated from 2 to 10 data points. The statistical analysis was carried out with an unpaired *t* test. – not statistically significant, *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$

expression vectors provided higher geranyl glucoside titers when introduced into *E. coli* BL21(DE3)K43 instead of *E. coli* BL21(DE3). The expression vectors, pET29a_VvGT14ao, pET32aK_VvGT14ao, and pET-SUMO_VvGT14ao, showed improved product yields in *E. coli* BL21(DE3)K43 with 0.220, 0.078, and 0.183 mM of geranyl glucoside, respectively (Figure 3).

3.3 | Screening hit K163

Sequencing of the plasmid of K163 revealed two distinct mutations, one exchanged base pair in a codon from AGC to AAC, and one from TGG to CGG. These mutations resulted in two AAE, creating the enzyme variant VvGT14ao-S168N-W353R. While S168 is located in an outward-facing peptide, W353 is the first amino acid of the highly conserved PSPG motif.[35] The single mutants S168N and W353R were prepared to clarify which mutation resulted in the significant increase in yield. The wild type, the two single mutants, and the double mutant sequences were cloned into three vectors, pRSFDuet-1, pET29a, and pGEX-K, for further testing. All tested W353R variants and double mutants showed an increased product yield compared with the wild type enzyme (Figure 4). However, the double mutant showed even higher product yields than the W353R mutant in the pRSFDuet-1 and pET29a vector. The S168N mutation on its own did not seem to be beneficial, except in pRSFDuet-1.

Because of this ambiguous in vivo result, the purified enzyme variants were tested in vitro for their glycosylation activity. The variants were expressed in *E. coli* BL21 using the pGEX-K vector, resulting in GST-tagged enzymes. They were purified with GST-binding resin, and the kinetic data were determined in vitro using the UDP-Glo assay. All three mutants showed a reduced catalytic activity (Table 1; Figure S4). The specific activity v_{max} decreased from 2100 U/mg of the wild type to 581, 219, and 58 U/mg for the S168N, double, and W353R mutant, respectively. The K_m value was less affected by the AAE but showed the highest value for the double mutant S168N-W353R with 230 μM while the K_m value for the wild type was 72 μM.

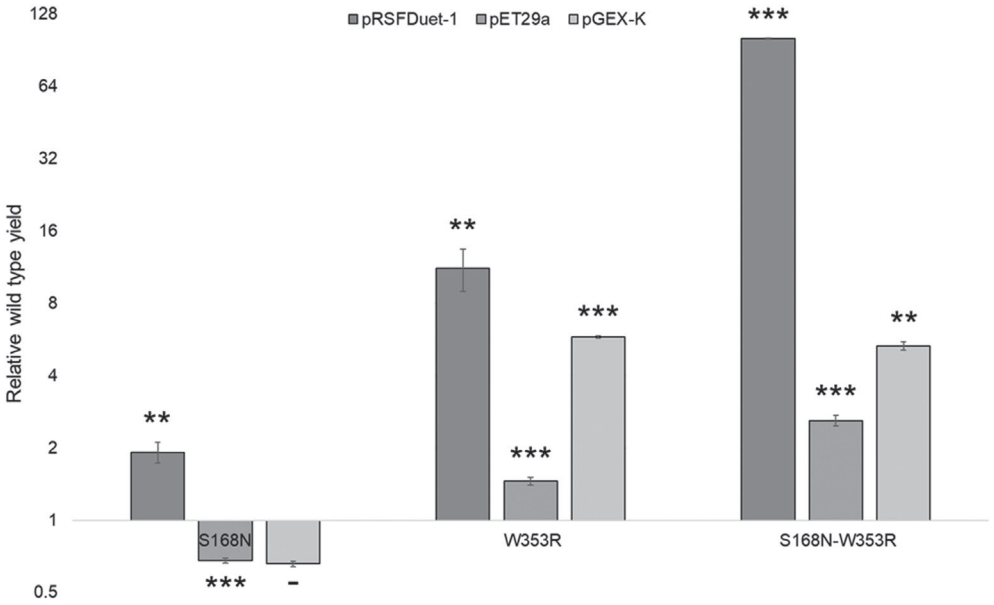

FIGURE 4 Relative product yield of the S168N, W353R, and double mutant (S168N-W353R) of VvGT14a when analyzed by whole-cell biotransformation. Product yield of the wild type enzyme was set to 1. Direct comparison of the VvGT14a variants in different plasmids (pRSFDuet-1, pET29a, and pGEX-K) using *E. coli* BL21(DE3). The average geranyl glucoside yields are shown with SDs using two to three data points; a logarithmic scale was used. The statistical analysis was carried out with an unpaired *t* test. – not statistically significant, $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$

TABLE 1 Kinetic data of GST-VvGT14a variants

GST-VvGT14a variant	K_M (μM)		v_{max} (U/mg)		k_{cat} (s)	k_{cat}/K_M (s/μM)
Wild type	72	±49	2100	±1180	1850	25.5
S168N	64	±6	581	±39	512	8.0
W353R	44	±12	58	±11	51	1.2
S168N-W353R	230	±148	219	±129	193	0.8

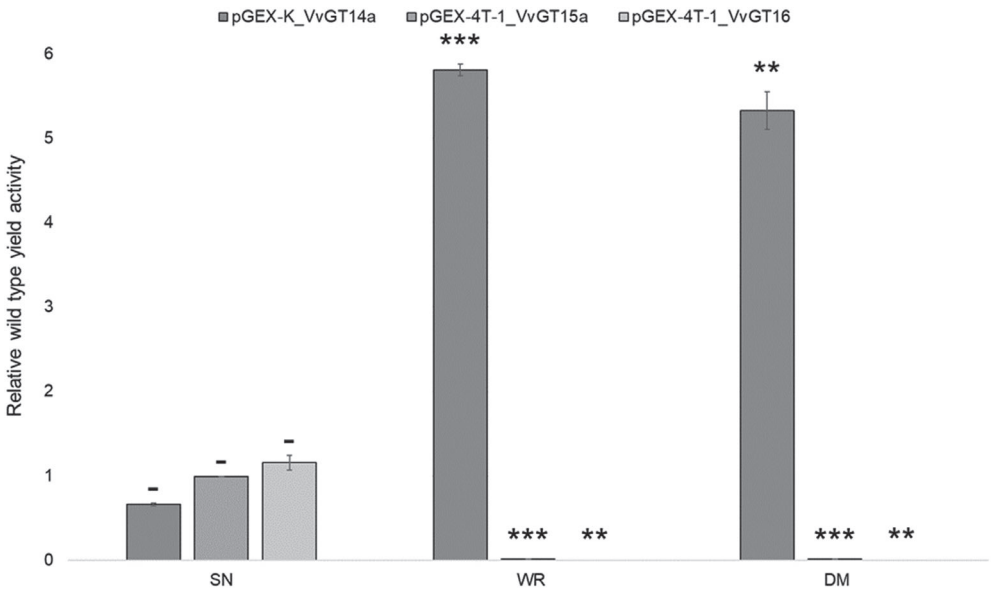

FIGURE 5 Relative product yield of the SN, WR, and double mutant of VvGT14a, VvGT15a, and VvGT16 in a whole-cell biotransformation experiment. Product yield of the respective wild type enzymes was set to 1. Direct comparison of the VvGT14a, VvGT15a, and VvGT16 variants in pGEX-vectors using *E. coli* BL21. The average glucoside yield (geranyl glucoside for VvGT14a and VvGT15a; benzyl glucoside for VvGT16) is shown with SDs of the duplicates. The statistical analysis was carried out with an unpaired *t* test. – not statistically significant, $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$

The S to N, W to R, and SW to NR AAE were also performed in other GTs from *V. vinifera* to investigate the transferability of the result obtained with VvGT14a. The corresponding amino acids where replaced in VvGT15a and VvGT16, the genes inserted in pGEX-4T-1 and introduced into *E. coli* BL21 to test the glycosylation abilities of the mutants in the HP cultivation system by whole-cell biotransformation. Production of geranyl glucoside was quantified in the case of VvGT15a because this biocatalyst glucosylated terpenols similar to VvGT14a, while the formation of benzyl glucoside was determined for VvGT16.[31] The SN mutation in VvGT15a, and VvGT16 did not significantly alter the product yield (Figure 5). By contrast, the WR mutation completely deactivated the biocatalysts (Figure 5). Therefore, the results obtained for VvGT14a were not transferrable to other enzymes.

3.4 | Combination of effects

To further increase product yield, the modifications identified for K43 and K163 were combined. The gene coding for VvGT14ao-S168N-W353R was ligated into the expression vectors pRSFDuet-1, pET29a, and pGEX-K, and the vectors were introduced into plasmid-cured K43. Each improvement measure showed a positive effect on the product yield alone. However, only the production system using the pET29a vector benefited synergistically from both measures combined (Figure 6). The concentration of geranyl glucoside increased from 0.064 mM produced by the wild type strain *E. coli* BL21(DE3) /pET29a_VvGT14ao to 0.656 mM achieved by the *E. coli* strain BL21(DE3)K43 /pET29a_VvGT14ao-S168N-W353R. All other K43 strains containing the double mutant produced less geranyl glucoside than the *E. coli* BL21(DE3) hosts with the double mutant. Overall, *E. coli* BL21(DE3)/pRSFDuet-1_VvGT14ao-S168N-W353R was a superior production system with a geranyl glucoside yield of 0.666 mM.

4 | DISCUSSION

A directed evolution approach was carried out to improve the product yield of an *E. coli* whole-cell biocatalyst. The approach was comprised of random mutagenesis coupled with a suitable screening system. VvGT14a was chosen as the target enzyme because of its broad terpene substrate spectrum. The codon-optimized *VvGT14ao* gene was subjected to

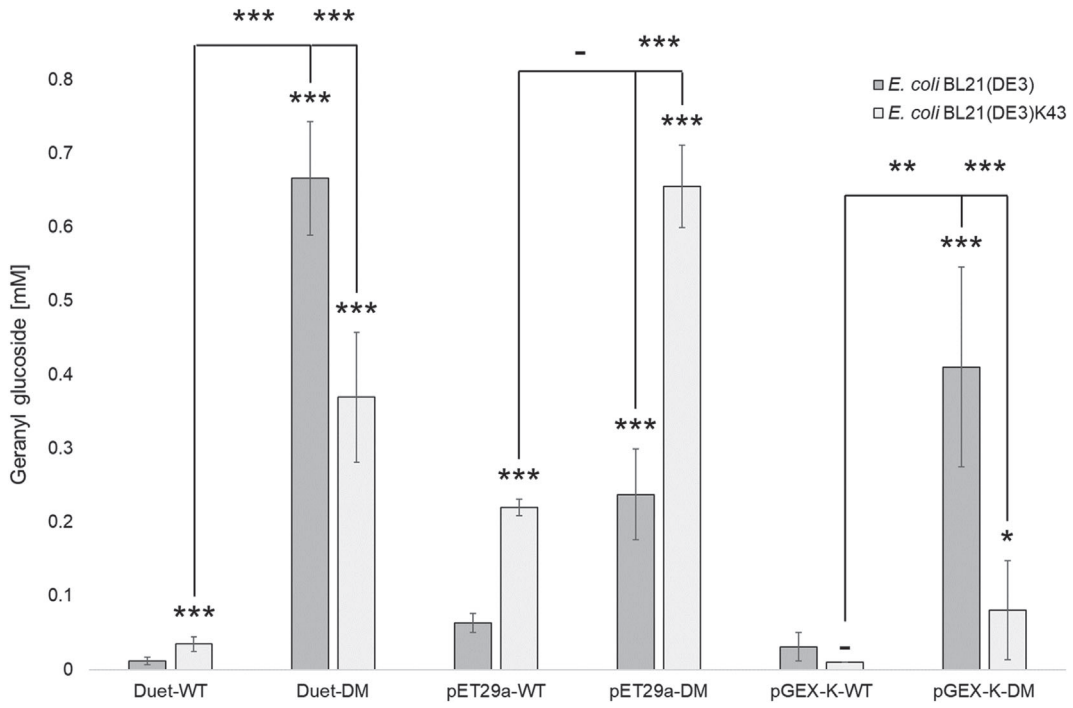

FIGURE 6 Absolute product yield of wild type and double mutant VvGT14a (DM) in different production systems. Direct comparison of VvGT14a wild type and double mutant in different vectors (Duet, pET29a, and pGEX-K) using *E. coli* BL21(DE3) (dark gray bars) and *E. coli* BL21(DE3)K43 (light gray bars) as hosts. The error bars represent the confidence intervals calculated from 2 to 10 data points. The statistical analysis was carried out with an unpaired *t* test. – not statistically significant, *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$

epPCR and the mutant library was screened by medium-throughput screening of the whole-cell biocatalysts. By screening only 176 colonies, the approach yielded a candidate strain (pRSFDuet-1_VvGT14ao-S168N-W353R in *E. coli* BL21(DE3)), which showed an up to 60-fold higher product yield than the initial strain (pRSFDuet-1_VvGT14ao in *E. coli* BL21(DE3)). In comparison, directed evolution of the GT OleD provided an enzyme with a more than 100-times higher catalytic activity towards the glycosylation of novobiocic acid by screening of 1000 and 300 putative mutants [23,32]. However, it should be noted that our attained value refers to the whole-cell process, while the second value of the cited study refers to the catalytic activity of an enzyme.

The analysis of the library showed that on average 37% of the putative mutants still contained the wild type enzyme, which reduced the probability to find highly productive mutants. A high quality mutant library should have a mutation rate between one and two AAE.[39] Our library showed 43% 1 AAE, 17% 2 AAE, and 3% 3 AAE and was therefore of decent quality.

The screening detected three outstanding mutants (Figure 1). Considering the quality of the mutant library and the small number of colonies tested (176), the probability of a hit was not high.[40] The largest group of candidates comprised strains harboring the wild type gene and neutral mutants. The high number was expected as 37% of the vectors coded for the wild type enzyme. The third group consisted of candidates with deleterious mutations and the last group contained mutants with premature stop codons, frame shifts, or other loss-of-function mutations. However, group IV in this screening also comprised a considerable number of strains carrying the pRSFDuet-1 vector without the target gene (up to 25%). All parameters of the restriction and dephosphorylation reactions were within the recommended limits. Nevertheless, the reactions were incomplete. The share of recombinants could be increased two- to three-fold by optimizing the digestion and dephosphorylation reaction.[41]

The highly effective strains K60, K43, and K163 achieved 0.051, 0.203, and 0.848 mM geranyl glucoside by whole-cell biotransformation, respectively, significantly exceeding the average yield of 0.014 mM of the wild type strain. The initial sequencing results of the candidates' plasmids revealed that K60 harbored an empty vector, K43 contained the wild type sequence and K163 carried the double mutant S168N-W353R.

The results for K43 and K60 were confirmed by repeating the biotransformation, starting from a single colony. The sequencing was also repeated with freshly extracted plasmid DNA from the same colonies used in the biotransformation. K43, and K60 were again identified as wild type, and empty vector strain, respectively. Both strains were cured of

their plasmid in order to find the source of the increased product yield. The cured strains *E. coli* BL21(DE3)K43 and *E. coli* BL21(DE3)K60 were unable to glycosylate geraniol. Retransformation of both strains with the wild type vector pRSFDuet-1_VvGT14ao gave two different results. *E. coli* BL21(DE3)K43 /pRSFDuet-1_VvGT14ao showed again the increased product formation in comparison to *E. coli* BL21(DE3) /pRSFDuet-1_VvGT14ao. However, the strain *E. coli* BL21(DE3)K60 /pRSFDuet-1_VvGT14ao produced only the same glucoside concentration as the reference strain. The enhanced glycosylation activity of K60 must be due to its plasmid(s), because the increased activity was lost along with the expression vectors. Therefore, the plasmid(s) contained in K60 were reanalyzed in more detail. cPCR using GT-specific primers (Figure S2) detected the GT gene *VvGT14ao*. The data suggested that K60 contained at least two versions of pRSFDuet-1, one with the *VvGT14ao* gene and one without an insert. As sequencing of the extracted plasmids provided the empty vector sequence with good quality, the majority of the plasmids of K60 must consist mainly of the empty vector. In addition, the sequencing of the cPCR product also revealed a variant, *VvGT14ao-T398S* and the wild type *VvGT14ao*. Therefore, the K60 strain contained at least three different vectors. The T398S variant was cloned and tested in vivo, but produced only 0.15 times the geranyl glucoside of the wild type strain. It is likely that the increased product formation is a result of an expression effect. The high-copy vector pRSFDuet-1 could lead to the overproduction of the target protein, which cannot be folded properly because of the fast production. Therefore, the host could benefit from a lower copy number of *VvGT14ao* and the resulting lower transcription and translation as decreasing *VvGT14ao* transcript levels led to higher glucoside titers (Figure S3). In vivo, protein folding occurs while the ribosome translates mRNA sequences into a growing polypeptide. Cotranslational folding is regulated by specific amino acid interactions as well as by the rate of synthesis relative to folding. Therefore, misinteractions and inappropriate rates of amino acid addition can contribute to protein misfolding events.[42]

The increased productivity of K43 was attributed to the genomic background of the strain instead of the expression vector. In order to test the transferability of this positive feature, the cured strain *E. coli* BL21(DE3)K43 was transformed with other expression vectors and their glucoside yields were compared with those of the corresponding *E. coli* BL21(DE3) strains (Figure 3). The advantageous property of K43 genomic background could be combined with the vectors pRSFDuet-1_VvGT14ao, pET29a_VvGT14ao, pET32aK_VvGT14ao, and pET-SUMO_VvGT14ao. The only vector that did not benefit from the K43 background was pGEX-K_VvGT14ao. The plasmid pGEX-K_VvGT14ao is the only vector that does not use the T7 promoter and the T7 RNA polymerase.

In order to illuminate the unplanned mutation in *E. coli* BL21(DE3)K43, the whole genome of the cured strain was sequenced. The genome sequencing showed an insert in the *rho* terminator hairpin loop of 1338 bp, belonging to a transposable element. This gene jumped to this location by chance and impeded the termination efficiency. The insert thereby increased the transcription of the following 13-kb operon in comparison with *rho* (Figure 2(A)). The enzymes encoded by the upregulated operon are involved in the synthesis of glycosylated membrane lipids and the cyclic polysaccharide ECAcyc (Figure S5). Most of the reactions catalyzed by the encoded enzymes require NDP-sugars that are formed from glucose-1-phosphate, meaning they are in competition with the GT for resources. Because the operon does not contain *glmU,* which codes for N-acetylglucosamine-1-phosphate uridyltransferase, the product of this enzyme, UDP-N-acetylglucosamine, is needed as precursor in the lipid III synthesis. Thus, the enhanced production of the encoded proteins of this operon cannot explain the increased production of geranyl glucosides by K43.

About half of the transcriptions in *E. coli* are terminated by Rho,[43] therefore the transcription of rho cannot be weak. It was proposed that Rho autoregulates its expression.[44] The increased expression of the 13-kb operon, which follows *rho*, is in competition with the expression of other genes for the additionally needed transcription and translation resources. The function of the terminator of rho should be impaired, because the insertion is located within the hairpin loop. However, the insertion caused the increased transcription of the adjacent operon, but not in equal amounts (Figure 2(A)). Consequently, the transcription and translation of the heterologous enzyme could be reduced. The decreased transcription of the target enzyme was experimentally confirmed (Figure 2(B)). Because a reduced amount of *GT* transcript was beneficial to the final product concentration, it seemed that too fast protein production is detrimental. The plant enzyme could be misfolded in *E. coli* and accumulate in inclusion bodies. As a result, the reduced transcription and translation of the GT might result in a higher proportion of correctly formed, active, and soluble enzyme. Chaperones support the correct folding of proteins, and it has been shown that coexpression of GroEL-GroES and VvGT14a led to significantly enhanced glucoside concentrations (unpublished results).

Some plant proteins can only be produced as active enzymes in prokaryotes at low temperatures, that is, when transcription and translation are slowed down. This illustrates that the rate of protein formation is important for proper folding into the catalytically active conformer. The variable dependence of protein production of different *E. coli* strains on expression temperature has been described several times.[45] In addition, *E. coli* strains with different genomic

backgrounds produce recombinant target proteins with varying degrees of success. Therefore, numerous strains are offered based on their different properties and protein production efficiencies.[46,47] In this regard, K43 is an *E. coli* derivative that is a superior host for *VvGT14ao*. We attempted to characterize this strain by genome sequencing but were unable to fully elucidate the cause of enhanced product formation although clearly the genomic background could be attributed to the increased amount of product. Protein yield in transgenic *E. coli* has been improved almost exclusively by targeted engineering approaches. However, there are few reported examples of isolation of *E. coli* protein production strains by evolutionary approaches.[48] K43 can be considered an example of such an optimal expression host for *VvGT14ao*.

Further analysis revealed that the increase in geranyl glucoside production efficiency of the plasmids in the strain background of *E. coli* BL21(DE3)K43 seemed to be inversely related to the length of the transcription and translation product (Figure 7). The more amino acids added to VvGT14a, the smaller the positive impact of the K43 strain background. In case of the pGEX-K vector, where 226 amino acids were added to VvGT14a, a minor effect of the alternative host could be observed (Figures 6 and 7). By contrast, a major effect was achieved with the pRSFDuet-1 vector. The pGEX-K strains differ considerably from the other candidates because pGEX-K depends on an inherent *E. coli* RNA polymerase. Therefore, the transcription of the gene of interest competes with all other transcription events in the cell, including the transcription of the upregulated *rfe*-[…]-*rffM* operon. The other expression vectors use the exclusive T7 RNA polymerase and circumvent this competition, which might explain the poor performance of pGEX-K in the K43 genomic background.

K163 delivered the double mutant S168N-W353R. The single mutants were created and were transformed in different expression vectors to analyze their effects on product titer. The W353R and S168N-W353R mutants resulted in superior product yields compared with the corresponding wild type strains (Figure 4) but the catalytic activities of the isolated enzymes were reduced (Table 1). Because the positive effect could not be transferred to other GTs (Figure 5), we therefore assume that the increased product formation by K163 is probably the result of an expression effect caused by the double mutant. To test different expressions levels of the mutant proteins in comparison to the wild type, SDS-PAGE and western blot analysis was performed (Figure S6). A faint band for GST-VvGT14ao was visible in the western blot while intensive bands for GST-VvGT14ao-W353R and GST-VvGT14ao-S168N-W353R were observed (Figure S5). The gene coding for VvGT14a was used in the experiments in its codon-optimized form to assure optimal translation in *E. coli*. However, too much and too fast protein production oftentimes leads to the formation of inclusion bodies and misfolded proteins.[42,49] The replacement of two seemingly optimal codons could slow the translation down and result in higher amounts of correctly formed, soluble enzyme. This potentially positive effect competes with the reduced activity of the enzymes and could not be transferred to other GTs (Figure 5). The GTs lost almost all their activity in the whole-cell biotransformation experiments when the WR mutation was introduced, which was in agreement with the in vitro data of GST-VvGT14a.

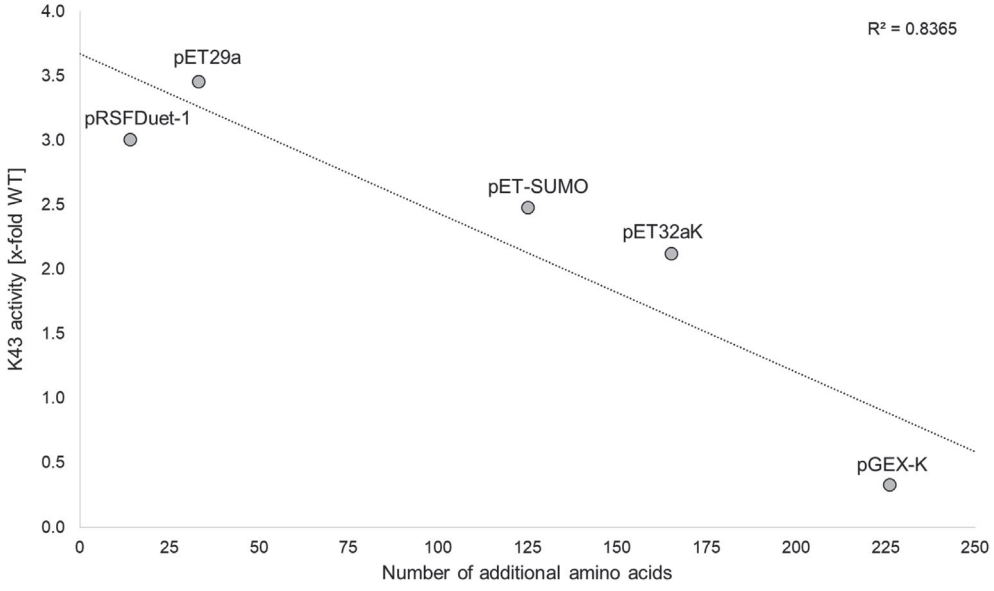

FIGURE 7 Relative geranyl glucoside production activity of different constructs according to the protein tag length. Data points from left to right: pRSFDuet-1, pET29a, pET-SUMO, pET32aK, and pGEX-K. All expression vectors contain the gene for VvGT14ao.
R^2 = coefficient of correlation

The increase of product formation by the double mutant is therefore limited to the whole-cell biotransformation using *VvGT14ao*.

During the execution of the experiments, it turned out that the vector pRSFDuet-1 is not an ideal vector for mutant library screening because the expression system is extremely leaky and not accurately regulated by IPTG (Figure S7). The used vector was not favorable, as lower expression, or a decrease in translation of VvGT14a positively affected the product yields. Nevertheless, this screening was a success as it brought forward one enzyme variant, which produced superior glucoside titers in vivo and revealed a novel production host, which showed improved glucoside production capabilities.

Our experimental set-up primarily yielded improved production strains but no enzyme showing increased catalytic activity. However, whole cell screening can also be used to identify enzyme variants with higher catalytic activities. In this case, to avoid false positive mutants due to background mutations or translational effects, certain steps should be taken or controlled in advance: (1) The recombinant gene should be resynthesized in a codon-harmonized form to obtain optimal transcript levels. (2) The expression system, consisting of host, vector, and target gene, should produce the target enzyme preferably in soluble form. (3) Inducer concentration should be optimized. (4) The mutant library should not contain a high proportion of wild-type strains and empty vector strains.

The directed evolution approach is generally applied in biosciences to improve the enzymatic activity of proteins. It is assumed that enzymes are important actors to increase productivity of a biotechnological process.[50] We were interested in the optimization of a whole-cell biocatalytic process and used intact *E. coli* cells for the screening of beneficial biocatalysts rather than isolated mutant proteins due to simplicity and reproducibility of the screening system. Although no improvement in catalytic efficiency of the enzyme was achieved, the productivity of the system could be increased by a factor of 60. The economic impact of this increase is remarkable, when considering the current price of geranyl glucoside (625 $ per 25 mg, https://www.carbosynth.com). In addition to the catalytic activity of the enzyme, the expression system including transcription, translation, and correct folding of the enzyme was crucial. Therefore, if whole-cell biocatalysts should be developed, intact cells should be screened rather than isolated mutant proteins so that productive enzyme-host combinations are not overlooked.

Optimization of whole-cell biocatalysts has been done exclusively by metabolic engineering using rational design[30] and resulted in moderate to good enhancement of product concentrations.[26–28] In our study, although we could not increase the catalytic activity of the protein by the product-directed screening of whole-cell biocatalysts we obtained a highly efficient and stable catalyst. We have here, admittedly unintentionally, taken advantage of the biological variability of the *E. coli* strain and isolated an efficient random mutant. We therefore conclude that in addition to metabolic engineering, that is, optimizing metabolite flux in the cell, transcriptional and translational events must also be adjusted.[48] Our results show that these phenomena are so far only very incompletely understood but can lead to a substantial increase in product titer.

ACKNOWLEDGMENTS

Special thanks to Prof. Jon S. Thorson for providing invaluable insight into the field of enzyme engineering. This work was supported by the International Graduate School for Science and Engineering (IGSSE) of the Technical University of Munich.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/eng2.12440.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interests.

AUTHOR CONTRIBUTIONS

Julian Ruediger: Formal analysis; investigation; methodology; writing-original draft. **Wilfried Schwab:** Conceptualization; resources; supervision; writing-review and editing.

ORCID

Wilfried Schwab https://orcid.org/0000-0002-9753-3967

REFERENCES

1. Graefe EU, Wittig J, Mueller S, et al. Pharmacokinetics and bioavailability of quercetin glycosides in humans. *J Clin Pharmacol*. 2001;41(5):492-499.
2. Kren V, Martínková L. Glycosides in medicine: εthe role of glycosidic residue in biological activityε. *Curr Med Chem*. 2001;8(11):1303-1328.
3. Dusek J, Carazo A, Trejtnar F, et al. Steviol, an aglycone of steviol glycoside sweeteners, interacts with the pregnane X (PXR) and aryl hydrocarbon (AHR) receptors in detoxification regulation. *Food Chem Toxicol*. 2017;109:130-142.
4. Kytidou K, Artola M, Overkleeft HS, Aerts JMFG. Plant glycosides and glycosidases: a treasure-trove for therapeutics. *Front Plant Sci*. 2020;11:357.
5. Tolstikova T, Bryzgalov A, Sorokina I, et al. Increase in pharmacological activity of drugs in their clathrates with plant glycosides. *Lett Drug Des Discov*. 2007;4(3):168-170.
6. Lozano-Rivas W, Whiting K, Gómez-Lahoz C, Rodríguez-Maroto J. Use of glycosides extracted from the fique (Furcraea sp.) in wastewater treatment for textile industry. *Int J Environ Sci Technol*. 2016;13(4):1131-1136.
7. Nicolaou K, Mitchell HJ. Adventures in carbohydrate chemistry: new synthetic technologies, chemical synthesis, molecular design, and chemical biology. *Angew Chem Int Ed*. 2001;40(9):1576-1624.
8. De Bruyn F, Maertens J, Beauprez J, Soetaert W, De Mey M. Biotechnological advances in UDP-sugar based glycosylation of small molecules. *Biotechnol Adv*. 2015;33(2):288-302. https://doi.org/10.1016/j.biotechadv.2015.02.005
9. Song C, Gu L, Liu J, et al. Functional characterization and substrate promiscuity of UGT71 glycosyltransferases from strawberry (*Fragaria*× *ananassa*). *Plant Cell Physiol*. 2015;56(12):2478-2493.
10. Jugdé H, Nguy D, Moller I, Cooney JM, Atkinson RG. Isolation and characterization of a novel glycosyltransferase that converts phloretin to phlorizin, a potent antioxidant in apple. *FEBS J*. 2008;275(15):3804-3814.
11. Gutmann A, Lepak A, Diricks M, Desmet T, Nidetzky B. Glycosyltransferase cascades for natural product glycosylation: use of plant instead of bacterial sucrose synthases improves the UDP-glucose recycling from sucrose and UDP. *Biotechnol J*. 2017;12(7):1600557.
12. Lairson L, Henrissat B, Davies G, Withers S. Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem*. 2008;77:221-555.
13. Ross J, Li Y, Lim E-K, Bowles DJ. Higher plant glycosyltransferases. *Genome Biol*. 2001;2:1-6.
14. Irmisch S, Jo S, Roach CR, et al. Discovery of UDP-glycosyltransferases and BAHD-acyltransferases involved in the biosynthesis of the antidiabetic plant metabolite Montbretin A. *Plant Cell*. 2018;30(8):1864-1886.
15. Tiwari P, Sangwan RS, Sangwan NS. Plant secondary metabolism linked glycosyltransferases: an update on expanding knowledge and scopes. *Biotechnol Adv*. 2016;34(5):714-739.
16. Schmolzer K, Gutmann A, Diricks M, Desmet T, Nidetzky B. Sucrose synthase: a unique glycosyltransferase for biocatalytic glycosylation process development. *Biotechnol Adv*. 2016;34(2):88-111.
17. Schwab W, Fischer T, Wüst M. Terpene glucoside production: improved biocatalytic processes using glycosyltransferases. *Eng Life Sci*. 2015;15(4):376-386.
18. Inoue Y, Shiraishi A, Hada T, Hirose K, Hamashima H, Shimada J. The antibacterial effects of terpene alcohols on staphylococcus aureus and their mode of action. *FEMS Microbiol Lett*. 2004;237(2):325-331.
19. Martins MA, Silva LP, Ferreira O, Schröder B, Coutinho JA, Pinho SP. Terpenes solubility in water and their environmental distribution. *J Mol Liq*. 2017;241:996-1002.
20. Schmid J, Heider D, Wendel NJ, Sperl N, Sieber V. Bacterial glycosyltransferases: challenges and opportunities of a highly diverse enzyme class toward tailoring natural products. *Front Microbiol*. 2016;7:182.
21. Gantt RW, Peltier-Pain P, Singh S, Zhou M, Thorson JS. Broadening the scope of glycosyltransferase-catalyzed sugar nucleotide synthesis. *Proc Natl Acad Sci*. 2013;110(19):7648-7653.
22. McArthur JB, Chen X. Glycosyltransferase engineering for carbohydrate synthesis. *Biochem Soc Trans*. 2016;44(1):129-142.
23. Williams GJ, Goff RD, Zhang C, Thorson JS. Optimizing glycosyltransferase specificity via εhot spotε saturation mutagenesis presents a catalyst for novobiocin glycorandomization. *Chem Biol*. 2008;15(4):393-401.
24. Garzón-Posse F, Becerra-Figueroa L, Hernández-Arias J, Gamba-Sánchez D. Whole cells as biocatalysts in organic transformations. *Molecules*. 2018;23:1265.
25. Ruprecht C, Bönisch F, Ilmberger N, et al. High level production of flavonoid rhamnosides by metagenome-derived glycosyltransferase C in *Escherichia coli* utilizing dextrins of starch as a single carbon source. *Metab Eng*. 2019;55:212-219.
26. Lim CG, Wong L, Bhan N, et al. Development of a recombinant *Escherichia coli* strain for overproduction of the plant pigment anthocyanin. *Appl Environ Microbiol*. 2015;81(18):6276-6284.
27. Xia T, Eiteman MA. Quercetin glucoside production by engineered *Escherichia coli*. *Appl Biochem Biotechnol*. 2017;182:1358-1370.
28. Shrestha A, Pandey RP, Dhakal D, Parajuli P, Sohng JK. Biosynthesis of flavone C-glucosides in engineered *Escherichia coli*. *Appl Microbiol Biotechnol*. 2018;102:1251-1267.
29. Effenberger I, Hoffmann T, Jonczyk R, Schwab W. Novel biotechnological glucosylation of high-impact aroma chemicals, 3(2H)- and 2(5H)-furanones. *Sci Rep*. 2019;9:10943.
30. Lin B, Tao Y. Whole-cell biocatalysts by design. *Microb Cell Fact*. 2017;16:106.
31. Bönisch F, Frotscher J, Stanitzek S, et al. Activity-based profiling of a physiologic aglycone library reveals sugar acceptor promiscuity of family 1 UDP-glucosyltransferases from grape. *Plant Physiol*. 2014;166(1):23-39. https://doi.org/10.1104/pp.114.242578
32. Williams GJ, Zhang C, Thorson JS. Expanding the promiscuity of a natural-product glycosyltransferase by directed evolution. *Nat Chem Biol*. 2007;3(10):657-662. https://doi.org/10.1038/nchembio.2007.28

33. Rüdiger J, Schwab W. Improving an *Escherichia coli*-based biocatalyst for terpenol glycosylation by variation of the expression system. *J Ind Microbiol Biotechnol*. 2019;46(8):1129-1138. https://doi.org/10.1007/s10295-019-02184-4
34. Trevors J. Plasmid curing in bacteria. *FEMS Microbiol Rev*. 1986;1(3–4):149-157.
35. Paquette SM, Jensen K, Bak S. A web-based resource for the Arabidopsis P450, cytochromes b5, NADPH-cytochrome P450 reductases, and family 1 glycosyltransferases (http://www.P450.kvl.dk). *Phytochemistry*. 2009;70(17–18):1940-1947. https://doi.org/10.1016/j.phytochem.2009.08.024
36. Belousov MV, Bondarev SA, Kosolapova AO, et al. M60-like metalloprotease domain of the *Escherichia coli* YghJ protein forms amyloid fibrils. *PLoS One*. 2018;13(1):e0191317. https://doi.org/10.1371/journal.pone.0191317
37. Tapader R, Bose D, Pal A. YghJ, the secreted metalloprotease of pathogenic *E. coli* induces hemorrhagic fluid accumulation in mouse ileal loop. *Microb Pathog*. 2017;105:96-99.
38. Erbel PJA, Barr K, Gao N, Gerwig GJ, Rick PD, Gardner KH. Identification and biosynthesis of cyclic Enterobacterial common antigen in *Escherichia coli*. *J Bacteriol*. 2003;185(6):1995-2004. https://doi.org/10.1128/jb.185.6.1995-2004.2003
39. Williams GJ, Thorson JS. A high-throughput fluorescence-based glycosyltransferase screen and its application in directed evolution. *Nat Protoc*. 2008;3(3):357.
40. Zhao J, Kardashliev T, Joëlle Ruff A, Bocola M, Schwaneberg U. Lessons from diversity of directed evolution experiments by an analysis of 3,000 mutations. *Biotechnol Bioeng*. 2014;111(12):2380-2389.
41. Chalhoub B, Belcram H, Caboche M. Efficient cloning of plant genomes into bacterial artificial chromosome (BAC) libraries with larger and more uniform insert size. *Plant Biotechnol J*. 2004;2(3):181-188.
42. Kramer G, Boehringer D, Ban N, Bukau B. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nature Struct Mol Biol*. 2009;16(6):589-597.
43. Mitra P, Ghosh G, Hafeezunnisa M, Sen R. Rho protein: roles and mechanisms. *Annu Rev Microbiol*. 2017;71(1):687-709. https://doi.org/10.1146/annurev-micro-030117-020432
44. Kung H-F, Bekesi E, Guterman SK, Gray JE, Traub L, Calhoun DH. Autoregulation of the rho gene of *Escherichia coli* K-12. *Mol Gen Genet*. 1984;193(2):210-213. https://doi.org/10.1007/bf00330669
45. Mühlmann M, Forsten E, Noack S, Büchs J. Optimizing recombinant protein expression via automated induction profiling in microtiter plates at different temperatures. *Microb Cell Fact*. 2017;16:220.
46. Jia B, Jeon CO. High-throughput recombinant protein expression in *Escherichia coli*: current status and future perspectives. *Open Biol*. 2016;6:160196.
47. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol*. 2014;5:172.
48. Schlegel S, Genevaux P, de Gier J-G. Isolating *Escherichia coli* strains for recombinant protein production. *Cell Mol Life Sci*. 2017;74:891-908.
49. Singh A, Upadhyay V, Upadhyay AK, Singh SM, Panda AK. Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. *Microb Cell Fact*. 2015;14(1):41.
50. Arnold FH. Directed evolution: bringing new chemistry to life. *Angew Chem Int Ed*. 2018;57(16):4143-4148. https://doi.org/10.1002/anie.201708408

SUPPORTING INFORMATION

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

How to cite this article: Rüdiger J, Schwab W. Improvement of an *Escherichia coli* whole-cell biocatalyst for geranyl glucoside production using directed evolution. *Engineering Reports*. 2021;3(12):e12440. https://doi.org/10.1002/eng2.12440