De novo production of benzyl glucosinolate in *Escherichia coli*

Annette Petersen, Christoph Crocoll, Barbara Ann Halkier*

DynaMo Center, Copenhagen Plant Science Centre, Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark

ARTICLE INFO

Keywords:
- Glucosinolates
- *Escherichia coli*
- Metabolic engineering
- Targeted proteomics

ABSTRACT

Microbial production of plant specialised metabolites is challenging as the biosynthetic pathways are often complex and can contain enzymes, which function is not supported in traditional production hosts. Glucosinolates are specialised metabolites of strong commercial interest due to their health-promoting effects. In this work, we engineered the production of benzyl glucosinolate in *Escherichia coli*. We systematically optimised the production levels by first screening different expression strains and by modification of growth conditions and media compositions. This resulted in production from undetectable to approximately 4.1 μM benzyl glucosinate, but also approximately 3.7 μM of desulfo-benzyl glucosinate, the final intermediate of this pathway. Additional optimisation of pathway flux through entry point cytochrome P450 enzymes and PAPS-dependent sulfotransferase increased the production additionally 5-fold to 20.3 μM (equivalent to 8.3 mg/L) benzyl glucosinate.

1. Introduction

A continuously increasing number of plant specialised metabolites are considered high value compounds due to e.g. pharmaceutical, nutraceutical or cosmetic properties. Production through natural sources is often not sustainable because of low abundance of compound and/or source as well as expensive purification processes (Ikram et al., 2015). The demand has inspired researchers to apply metabolic engineering and synthetic biology approaches to microorganisms. The efforts have led to innumerable reports on heterologous production of plant specialised metabolites in expression hosts such as *Escherichia coli* and *Saccharomyces cerevisiae* (Chae et al., 2017; Cho et al., 2015; Huang et al., 2008; Pontrelli et al., 2018; Siddiqui et al., 2012). Only few compounds have reached commercially relevant production levels and production is most commonly reported in the μg/L to low mg/L range (Chubakov et al., 2016; Pontrelli et al., 2018). In pursuit of still higher titles, researchers continue to develop technologies in the fields of omics, high-throughput analytics and computational engineering. Today, strategies to increase yields often include multiple rounds of optimisation and large mutant library screens - an approach also known as systems metabolic engineering (Chae et al., 2017; Cho et al., 2015; Yang et al., 2017). An emerging approach to identify bottlenecks and optimise flux through a pathway is mass spectrometry-based targeted proteomics by selected reaction monitoring (SRM). By using isotopically labelled synthetic peptides as internal references in the samples, it is possible to determine not only the presence but also the relative or absolute quantities of a given protein of interest e.g. over time or at different conditions (Batth et al., 2012). This method has been used successfully for increasing production titres of several products in *E. coli* (Alonso-Gutierrez et al., 2015; Batth et al., 2012; Brunk et al., 2016; Dahl et al., 2013; George et al., 2015, 2014; Juminaga et al., 2012; Redding-Johanson et al., 2011; Singh et al., 2012; Tan et al., 2016).

One group of plant specialised metabolites with commercial interest is glucosinolates (GLSs). They are sulfur-rich compounds characteristic of brassicaceous plants, including the model plant *Arabidopsis thaliana* and several agriculturally important crops (Fahey et al., 2001). In planta, GLSs constitute a line of defence, which becomes activated upon attack from pathogens or herbivores (Jeschke and Burow, 2018). For humans, the intake of brassicaceous vegetables has been associated with reduced risks of cardiovascular diseases and several types of cancer (reviewed in (Traka, 2016)), and with better insulin resistance for type 2 diabetes (Bahadoran et al., 2012). These health beneficial effects have made GLSs interesting as dietary supplements and primed a desire to engineer microbial production as rich source of GLSs. GLSs are derived from amino acids and synthesised through a core structure pathway consisting of seven enzymatic steps (Fig. 1). Some GLSs, e.g. the aliphatic methionine-derived GLSs, require chain

**Abbreviations:** dsBGLS, desulfo-benzyl glucosinolate; BGLS, benzyl glucosinolate; I3M, indol-3-ylmethyl glucosinolate; 4MSB, 4-methylsulfinylbutyl glucosinolate; APS, adenosine 5′-phosphosulfate; PAPS, 3′-phosphoadenosine-5′-phosphosulfate; GLS, glucosinolate

*Corresponding author. University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark.

E-mail address: bah@plen.ku.dk (B.A. Halkier).

https://doi.org/10.1016/j.ymben.2019.02.004
Received 4 December 2018; Received in revised form 11 February 2019; Accepted 24 February 2019
Available online 01 March 2019
1096-7176/ © 2019 The Authors. Published by Elsevier Inc. on behalf of International Metabolic Engineering Society. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).
incorporating steps. The feasibility of engineering the GLS pathway in a heterologous host was first shown by transient expression of the benzyl GLS (BGLS) genes in *Nicotiana benthamiana* (Geu-Flores et al., 2009b). This work identified an additional enzyme - a γ-glutamyl peptidase, GGPI - in the pathway (Geu-Flores et al., 2009a), and showed that the reduced sulfur atom was derived from glutathione (Geu-Flores et al., 2011). The final step of the pathway – the conversion of desulfo-BGLS (dsBGLS) to BGLS - was identified as a metabolic bottleneck. This conversion is catalysed by a sulfotransferase, which requires 3’-phosphoadenosine-5’-phosphosulfate (PAPS) as co-factor. The metabolic bottleneck was alleviated by co-expressing the Arabidopsis kinase APK2 responsible for PAPS production (Meldrup et al., 2011). Having identified and overcome these basic metabolic bottlenecks in *N. benthamiana* provided a basis for engineering of GLSs in microbial hosts. The tryptophan-derived GLS indol-3-ylmethyl GLS (I3M) was produced at 1.07 mg/L in *S. cerevisiae* proving for the first time that GLSs can be produced in a microbial organism (Mikkelsen et al., 2012). Recently, two reports of engineering of GLS biosynthetic genes in *E. coli* have been published. In 2016, the chain elongation pathway for methionine was successfully expressed in *E. coli* resulting in the production of di-homo-methionine, which is precursor to 4-methylsulfinylbutyl GLS (4MSB) (Mirza et al., 2016). In 2018, 4MSB was produced in *E. coli* although production levels were not quantified (Yang et al., 2018).

In this work, we set out to establish microbial production of the phenylalanine-derived BGLS and systematically optimise the yields. *E. coli* was chosen as production host, as previous reports showed it was capable of expressing a GLS pathway despite its complexity (Yang et al., 2018). The choice of expression strain, medium, cultivation conditions and construct designs were all factors shown to have major effects on BGLS yield in our production system. Targeted proteomics was used to monitor the relative protein expression levels. We optimised production levels to 20.3 μM (or 8.3 mg/L) BGLS.

2. Materials and methods

2.1. Construct design and generation

The GLS biosynthetic gene sequences were found at [www.arabidopsis.org](http://www.arabidopsis.org) and codon optimised for *E. coli* expression using the Codon Optimisation Tool from IDT (Integrated DNA Technologies Inc, USA) [http://eu.idtdna.com/CodonOpt]. The optimised gene sequences were ordered as gBlocks from IDT (Integrated DNA Technologies Inc, USA) and cloned into expression vectors by Genewiz (New Jersey, USA). The *E. coli* sulfate assimilation gene sequences *cysC* (adenylyl-sulfate kinase), *cysW* (subunit of sulfamate adenlytransferase) and *cysD* (subunit of sulfate adenlytransferase) were found in UniProtKB database and PCR amplified with Phusion® High Fidelity Polymerase (New England Biolabs, USA) from genomic DNA of the BL21 (DE3) *E. coli* strain (Table 1).

![Fig. 1. Biosynthetic pathway of benzyl glucosinolate in *Arabidopsis thaliana*. CYP79A2 and CYP83B1 are cytochrome P450s. GSTF9, GGPI, SUR1 and UGT74B1 are a glutathione-S-transferaseF9, γ-glutamyl peptidase1, C-S lyase and UDP-glucosyltransferase74B1. The SOT16 and SOT18 are sulfotransferases. The pathway compounds are 1) phenylalanine, 2) phenylacetaldoxime, 3) phenylacetocyanide oxide, 4) S-(Z)-phenylacetohydroximoyl-L-glutathione, 5) Cys-Gly conjugate, 6) phenylacetothiohydroximic acid, 7) desulfo-benzyl glucosinolate (dsBGLS) and 8) benzyl glucosinolate (BGLS).](image-url)
For this work pET-52b (Novagen®, Merck, #71554), pCDF-1b (Novagen®, Merck, #71330) and pCOLADuet™-1 (Novagen®, Merck, #71406) were used as expression vectors. GGPI, SUR1 and ATR1 were expressed from pCDF-1b (∼20–40 copies/cell). The remaining core structure genes were expressed from pET-52b (∼40 copies/cell). The E. coli sulfate assimilation pathway genes leading to PAPS, i.e. the two subunits of sulfate adenylyltransferase (cysD and cysN) and the adenylyl-sulfate kinase (cysC) were expressed from pCOLADuet (∼20–40 copies/cell). The construct is referred to as cysCND (Fig. 2C). All genes were organised in separate operons using four different T7 promoters with varying strengths: consensus, G6, H10 and C4 (naming as in (Jones et al., 2015)). In the constructs indicated, the first eight residues of the N-terminus anchors on the two P450 enzymes were exchanged with a modified anchor of bovine CYP17A1 (hereafter bovine-modified) optimised for E. coli expression (Barnes et al., 1991). DNA fragments encoding for the modified P450 anchor, the SOT18 enzyme as well as the promoters for CYP79A2 and SOT enzymes were exchanged by ligation cloning with T4 DNA ligase (Fig. 2 and Supplemental Fig. S1). The sequences were ordered as gBlocks with appropriate restriction sites (Integrated DNA Technologies Inc, USA) and ligated into the expression vectors.

Fig. 2. Design of multigene constructs for expression of BGLS pathway and support genes in E. coli. Promoters and ribosomal binding site sequences are represented as grey arrows. Biosynthetic genes, terminator sequences and P450 bovine anchor are boxes coloured blue, red and yellow, respectively. A) The two constructs containing the BGLS biosynthetic pathway with choice of promoter indicated. B) For optimisation changes to the two constructs included bovine modification of N-terminal anchor of P450 enzymes, change of promoters as well as use of different sulfotransferase genes. C) Construct used to overexpress the E. coli PAPS synthesis pathway (adenylyl-sulfate kinase (cysC)) and the two subunits of sulfate adenylyltransferase (cysD and cysN)), referred to as cysCND. D) T7 promoters used in this study (Table modified from Jones et al. (2015)). Sequence differences are highlighted and the expression levels relative to the consensus T7 promoter are listed. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

For this work pET-52b (Novagen®, Merck, #71554), pCDF-1b (Novagen®, Merck, #71330) and pCOLADuet™-1 (Novagen®, Merck, #71406) were used as expression vectors. GGPI, SUR1 and ATR1 were expressed from pCDF-1b (∼20–40 copies/cell). The remaining core structure genes were expressed from pET-52b (∼40 copies/cell). The E. coli sulfate assimilation pathway genes leading to PAPS, i.e. the two subunits of sulfate adenylyltransferase (cysD and cysN) and the adenylyl-sulfate kinase (cysC) were expressed from pCOLADuet (∼20–40 copies/cell). The construct is referred to as cysCND (Fig. 2C). All genes were organised in separate operons using four different T7 promoters with varying strengths: consensus, G6, H10 and C4 (naming as in (Jones et al., 2015)). In the constructs indicated, the first eight residues of the N-terminus anchors on the two P450 enzymes were exchanged with a modified anchor of bovine CYP17A1 (hereafter bovine-modified) optimised for E. coli expression (Barnes et al., 1991). DNA fragments encoding for the modified P450 anchor, the SOT18 enzyme as well as the promoters for CYP79A2 and SOT enzymes were exchanged by ligation cloning with T4 DNA ligase (Fig. 2 and Supplemental Fig. S1). The sequences were ordered as gBlocks with appropriate restriction sites (Integrated DNA Technologies Inc, USA) and ligated into the expression vectors.

2.2. E. coli: strains, growth conditions and transformation

All cloning and plasmid amplifications were done in NEB® DH10B strain (New England Biolabs, #C3019H). Cells were cultivated in Luria-Bertani broth (LB Lennox, Duchefa, #L1703) containing 10 g tryptone, 5 g yeast extract and 5 g sodium chloride per litre, or Terrific Broth (TB) containing 12 g tryptone, 24 g yeast extract, 5% glycerol (w/v), 0.17 M KH₂PO₄ and 0.72 M K₂HPO₄ per litre. Transformations were performed by heat-shock and the cells recovered in LB at 37 °C for 60 min before plating on LB agar plates with appropriate selection (50 μg/ml carbenicillin, spectinomycin or kanamycin).

The constructs containing the core structure genes were transformed into seven different expression strains (Table 2): four different BL21 (DE3) strains inducible by isopropyl β-D-thiogalactopyranoside (IPTG), as well as one BL21-AI strain and two cloning strains with the T7 RNA polymerase expressed from a plasmid (pTARA, (Wycuff and Matthews, 2000)) and inducible by arabinose. Five of the strains contain an extra plasmid (pLysS, Rosetta, pRI952 or pTARA), which is compatible with co-expression of the three expression plasmids (pET-52b, pCDF-1b and pCOLA-duet).

For expression cultures, LB media was inoculated with single
colonies and incubated over night at 37 °C, 200 rpm. The expression cultures were cultivated in LB or TB as indicated and with the appropriate antibiotics (50 μg/mL carbenicillin, spectinomycin and/or kanamycin, 15 μg/mL chloramphenicol). Experiments were performed in 24-well plates with 4 ml culture. Fresh media with appropriate selection was inoculated to an OD600 of approximately 0.1 with the overnight 24-wellplates with 4 ml culture. Fresh media with appropriate selection was inoculated to an OD600 of approximately 0.1 with the overnight culture and allowed to grow at 37 °C and 200 rpm to an OD600 of approximately 0.6. At this point the cultures were induced with 0.5 mM IPTG or 0.1% (w/v) arabinose and grown at 18 °C and 120 rpm for 24 or 48 h before harvesting. Three individual colonies from each strain were tested in triplicates for the metabolic analysis. When the cultures were supplemented with various intermediates, the following concentrations were added: 25 μM phenylacetaldoloxime, 75 mg/L S-aminolevulinic acid, 20 μM glutathione, 20 μM cysteine, 1 mM thiamine or 20 μM magnesium sulfate. When the strain overexpressing the cysCND construct was supplemented with sulfur either 1 mM cysteine and/or 50 μM magnesium sulfate were used.

2.3. Metabolic extraction and LC-MS analysis

The supernatant was collected after centrifugation at 13,000 × g for 5 min. An aliquot of the supernatant was diluted 10-fold with milliQ grade water and finally 1:10 (v/v) mixed with a stock solution containing 10 μg/mL 13C-, 15N-labelled amino acids (Algal amino acids 13C, 15N, Isotec, Miamisburg, US) and 25 μM sinigrin (PhytoLab, Vestenbergsreuth, Germany). The resulting 100-fold diluted samples were filtered (Durapore® 0.22 μm PVDF filters, Merck Millipore, Tullagreen, Ireland) and used directly for LC-MS analysis.

The LC-MS analysis was performed as previously described (Mirza et al., 2016) with changes as detailed below to monitor amino acid levels simultaneously with products and intermediates of the BGLS biosynthesis pathway. Briefly, chromatography was performed on an Advance UHPLC system (Bruker, Bremen, Germany). Separation was achieved on a Zorbax Eclipse XDB-C18 column (100 × 3.0 mm, 1.8 μm, Agilent Technologies, Germany). Formic acid (0.05% (v/v)) in water and acetonitrile (supplied with 0.05% (v/v) formic acid) were employed as mobile phases A and B, respectively. The elution profile was: 0–1.2 min 3% B; 1.2–4.3 min 3–65% B; 4.3–4.4 min 65–100% B; 4.4–4.9 min 100% B; 4.9–5.0 min 100% B and 5.0–6.0 min 3% B. Mobile phase flow rate was 500 μl/min and column temperature was maintained at 40 °C. The liquid chromatography was coupled to an EVOQ Elite TripleQuad mass spectrometer (Bruker, Bremen, Germany) equipped with an electrospray ionisation source (ESI). Instrument parameters were optimised by infusion experiments with pure standards. The ionspray voltage was maintained at 3000 V or −4000 V in positive or negative ionisation mode, respectively. Cone temperature was set to 300 °C and cone gas flow to 20 psi. Heated probe temperature was set to 400 °C and probe gas flow set to 50 psi. Nebulising gas was set to 60 psi and collision gas to 1.6 mTorr. Nitrogen was used as both cone gas and nebulising gas and argon as collision gas.

Multiple reaction monitoring (MRM) was used to monitor analyte parent → product ion transitions: MRM for the 13C-, 15N-labelled phenylalanine was chosen as previously described (Docimo et al., 2012). MRMs for dsBGLS, BGLS and sinigrin were chosen as previously described (Crocchi et al., 2016). Both Q1 and Q3 quadrupoles were maintained at unit resolution. Bruker MS Workstation software (Version 8.2.1, Bruker, Bremen, Germany) was used for data acquisition and processing. BGLS and dsBGLS were quantified using sinigrin and 13C, 15N-Phe, respectively, as internal standards, except for Fig. 3A, where quantification was performed based on a dilution series (linear regression). Response factors were calculated based on dilution series of the respective analytes. Internal standards were chosen based on matching ionisation mode with the analyte of interest (i.e. negative ionisation mode for glucosinolates and positive ionisation mode for desulfo-glucosinolates and amino acids). Matrix effects from media and cultivating E. coli were found to impair correct quantification of the analytes. The correction factors for matrix effects have been calculated into the response factors given in Supplemental Table S1. The presence of the phenylacetaldoloxime and the GSH-conjugated intermediate were monitored without quantification. For analytes where multiple transitions were monitored, the transition used for quantification is marked as quantifier (Q). Further details for transitions and collision energies can be found in Supplemental Table S1. Visualisation and statistical analysis were performed in RStudio v1.0.153 (R version 3.4.1) (RStudio Team, 2015) using (Garnier, 2018; Wickham, 2016, 2011, 2007).

2.4. Targeted proteomics

2.4.1. Design of reference peptides

For each protein of interest, a set of proteotypic peptides was designed and selected as stable isotopically labelled synthetic peptides (JPT, SpikeTides™). These were used as internal references in the samples. Recently, a set of peptides covering most of the BGLS pathway was designed and generously provided to us by Meike Burow and Daniel Vík. We designed peptides for the remaining enzymes of the pathway (ATRI1 and GSTF9) and housekeeping enzymes from E. coli. A list of peptides was obtained from each protein through an in silico tryptic digest in Skyline 4.2 (MacLean et al., 2010). Suitable peptides were selected based on the following criteria: 1) less than 1250 m/z, 2) no neighbouring arginine or lysine residue to the cleavage site, and 3) avoidance of methionine and cysteine residues if possible. Each enzyme was covered by a minimum of two non-neighbouring peptides (Table 3).

2.4.2. Protein extraction and tryptic digest

Pellets from 10 μl culture were harvested by centrifugation and kept at −20 °C until extraction. Total protein extraction was performed as previously described (Wessel and Flügge, 1984) with one exception: centrifugation steps were done at 21,000 × g for 1 min at 4 °C. The protein pellet was dried in Speed-Vac (30–60 min, 1000 rpm, max 35 °C). The downstream protocol was modified from previously described methods (Batth et al., 2012). The protein pellet was extracted twice in 100 μl 100 mM ammonium bicarbonate buffer with 10% (v/v) methanol by sonicating for 2–5 min in ultrasonic bath followed by a
brief centrifugation (1.5 min at 1,000 × g). Supernatant of both extractions was combined and protein concentration determined by Pierce™ BCA Protein Assay Kit (ThermoFisher, #23225). For tryptic digest, 100 μl of 0.5 mg/ml protein extraction was used (50 μg total protein). The extraction was incubated with 1 μl 10 mM DTT for 30 min at room temperature. Subsequently, 25 μl 50 mM iodoacetamide was added and the samples incubated in the dark for 20 min at room temperature. An additional 100 μl 100 mM ammonium bicarbonate buffer with 10% (v/v) methanol was added with 1 μg trypsin/Lys-C mix (Promega, #V5073) and the samples incubated overnight at 37°C. The digest was stopped by adding 20 μl 10% (v/v) trifluoroacetic acid (TFA) and samples were diluted up to 1.5 ml in buffer A (2% (v/v) acetonitrile and 0.1% (v/v) formic acid) and centrifuged (20,000 × g for 15 min). Peptides were purified over Sep-Pak C-18 columns (Waters, Sep-Pak® Vac 1 cc 100 mg, #WAT023590). Columns were attached to vacuum manifold and equilibrated with 1 ml buffer B (65% (v/v) acetonitrile and 0.1% (v/v) formic acid) followed by 1 ml buffer A. Next, the peptide digest was loaded, followed by three washes with 1 ml buffer A and finally eluted with two times 0.5 ml buffer B. Purified peptides were dried by Speed-Vac® (2–4 h, 1000 rpm, max 35°C). Dried peptides were redissolved in 25 μl buffer C (2% (v/v) acetonitrile, 0.5% (v/v) formic acid and 0.1% (v/v) TFA), spiked with 20 nM isotopically labelled peptide standards (JPT, SpikeTides™), and filtered through 0.22 μm centrifugal filters (#FC30GV00, Merck, Darmstadt, Germany).

2.4.3. LC-MS method

The gradient was adopted from Barth et al. (2014) and Percy et al. (2012) with modifications. Briefly, formic acid (0.1% (v/v)) in water and acetonitrile (supplied with 0.1% (v/v) formic acid) were employed as mobile phases A and B, respectively. The elution profile was: 0–1.0 min 5–10% B; 1.0–3.0 min 10–11% B; 3.0–13.0 min 11–19% B; 13.0–21.0 min 19–27.5% B; 21.0–21.7 min 27.5–24% B; 21.7–22.5 min 34–42% B; 22.5–23.5 min 42–90% B; 23.5–26.9 min 90% B; 26.9–30.0 min 90–5% B and 30.0–34.0 min 5% B. Mobile phase flow rate was 500 μl/min and column temperature was maintained at 55°C. Peptide separation was achieved on an Aeris PEPTIDE, XB-C18 column (1.7 μm, 2.1 × 150 mm, Phenomenex, Palo Alto, USA) on an Advance UHPLC-OLE (Bruker Daltonics, Bremen, Germany). The injection volume was 10 μl.

Source settings for heated electrospray ionisation were as follows: spray voltage 3200 V, in positive ionisation mode; cone temperature 300°C; cone gas flow 20 psi; heated probe temperature 300°C; probe gas flow 40 and nebuliser gas flow 50. Nitrogen was used as cone and probe gas and argon as collision gas. The triple quadrupole mass spectrometer (EVOQ Elite, Bruker Daltonics, Bremen, Germany) was set to scan for parention → product ion transitions for individual peptides within scheduled 3 max windows. Resolution of the first and third quadrupole was set to ± 1 Da. Detailed information on peptides including retention times, transitions selected for detection and quantification and collision energies are summarised in Supplemental Table S2. The acquired chromatograms were manually inspected and ratios between endogenous light and synthetic heavy peptides were obtained through Skyline 4.2 for each peptide (MacLean et al., 2010). Relative quantification was performed by normalisation to heavy peptides, followed by normalisation to the housekeeping protein ICD and plotted relative to the CSaro control. Undetected peptides (N/A) were treated as zero values for analysis. Additional statistical analyses and visualisation were performed in RStudio v1.0.153 (R version 3.4.1) (RStudio Team, 2015) using (de Mendiburu, 2017; Wickham, 2016, 2011, 2007).

3. Results

3.1. Establishment of dsBGLS production in E. coli

The biosynthetic pathway of phenylalanine-derived BGLS consisting of seven genes - CYP79A2, CYP83B1, GSTF9, GGP1, SUR1, UGT74B1 and SOT16 - was chosen for engineering and optimising GLS production in E. coli. The NADPH cytochrome P450 reductase, ATR1, from A. thaliana was included, as E. coli does not efficiently support the function of cytochrome P450 enzymes (Park et al., 2013). The pathway was expressed from plasmids to enable efficient and rapid modifications of the construct design. The genes were divided between two medium to high copy-number plasmids and all genes were expressed from IPTG- or arabinose-inducible T7 promoters (Fig. 2A). Different T7 promoters were used as a previous study showed that choice of T7 promoters can affect production levels substantially (Jones et al., 2015). Based on previous findings on the evolutionarily related cyanogenic glucoside biosynthetic pathway (Laursen et al., 2016) and on the low frequency by which CYP79s compared to CYP83s are found in classical shotgun proteomics approaches in Arabidopsis, we expected CYP83s to be present in higher levels than CYP79s in planta (Baerenfaller et al., 2011; Laursen et al., 2016). We therefore chose to express CYP83B1 from a consensus T7 promoter and the other genes from the weaker G6 promoter.

First, the widely used BL21 (DE3) expression strain was engineered with the BGLS pathway. Besides the desired product BGLS, three intermediates of the pathway were monitored: phenylacetaldoxime, the intermediate of the pathway (Fig. 1), this result indicated functionality of all but the ultimate sulfotransferase step. The OD600 at the time of induction was 18°C, dsBGLS was produced, but still none of the other monitored products were detected (data not shown). As dsBGLS is the final intermediate of the pathway, besides the desired product BGLS, three intermediates were monitored: phenylacetaldoxime, the intermediate of the pathway (Fig. 1), this result indicated functionality of all but the ultimate sulfotransferase step. The OD600 at the time of induction proved important to titres with optimal production after an induction at OD600 = 0.6 (data not shown). Harvesting 48 h after induction resulted in 4.74 μM dsBGLS, but still no BGLS was detected. Based on these results, subsequent experiments were performed by inducing expression at OD600 = 0.6 and allowing production cultures to grow at 18°C for 48 h after induction.

3.2. Screening of different E. coli strains for production

The plasmids containing the BGLS genes were transformed into six additional E. coli strains to investigate whether strain specificity would result in BGLS production. The BL21 (DE3) strain was included as control. Strains pRI952 and Rosetta are engineered with rare tRNAs to help expression. BL21-AI has the T7 RNA polymerase expressed from the araBAD promoter enabling repression of potential toxic, leaky expression by glucose (Miyada et al., 1984). The pLysS strain represses expression until IPTG induction through the T7 lysozyme (Studier, 1991). These five strains are all of the classical B-lineage of E. coli. Two K-12-derived strains, DH5a and DH10B, were included as an alternative. These were transformed with the pTARA plasmid carrying the T7 RNA polymerase expressed from the araBAD promoter like the BL21-AI strain (Wyccuf and Matthews, 2000).

When seven different E. coli strains harbouring the BGLS biosynthetic genes were grown for 48 h at 18°C, differences in dsBGLS production were observed (Fig. 3A). The highest-producing strains were the standard strain BL21 (DE3) and the pRI952. The latter produced 1.6-fold more dsBGLS than BL21 (DE3). These two strains were selected for further optimisations.
3.3. Optimisation of medium composition

3.3.1. Supplying additives or intermediates of the pathway

The first two enzymes of the BGLS pathway are the heme-containing P450 proteins. Supplementation of medium with heme precursor 5-aminolevulinic acid was previously shown to improve activity of P450 enzymes in E. coli (Hansen et al., 2001; Pritchard et al., 1997; Wittstock and Halkier, 2000). When the expression cultures were fed with 5-aminolevulinic acid to investigate whether co-factor supplement would result in BGLS production only little increase in dsBGLS levels and still no BGLS was observed (Fig. 3B). When feeding the first intermediate of the BGLS pathway, phenylacetaldoxime, production of dsBGLS was increased 10-fold and 7-fold in strains BL21 (DE3) and pRI952, respectively. This indicated that the entry point to the pathway was a metabolic bottleneck.

The BGLS pathway contains two sulfur-incorporating enzymatic steps; the glutathione-conjugation reaction by GSTF9 and the transfer of sulfate by SOT16 (Fig. 1). When the cultures were supplemented with glutathione, no effect on production levels was observed suggesting that glutathione is not limiting (Fig. 3B). The second sulfur-incorporation step is responsible for the conversion of dsBGLS to BGLS, and hence the step of the pathway, which appeared non-functional. Sulfotransferase (SOT) enzymes transfer a sulfate moiety from PAPS onto dsGLSs (Klein and Papenbrock, 2009). PAPS is a product of the sulfate assimilation pathway (Kopriva and Koprivova, 2004). To increase the SOT activity and obtain BGLS, different sulfur sources were fed to the expression cultures in the form of cysteine, thiamine and magnesium sulfate. This did not result in any BGLS production, and both cysteine and thiamine appeared to hamper production of dsBGLS (Fig. 3B). The pRI952 strain proved consistently better than the BL21 (DE3) and subsequent experiments were carried out using only this strain.

3.3.2. Production of BGLS using TB medium

Instead of supplementing individual compounds to the medium, the LB medium was exchanged with the more nutritious TB medium that predominantly differs by containing approximately 5-fold more yeast extract and by being buffered by potassium phosphate (KPi). BGLS was produced in TB to levels of approximately 4.1 μM (Fig. 4A). This is comparable to the production of I3M in S. cerevisiae (2.38 μM, Mikkelsen et al., 2012). Only dsBGLS was detected in LB medium. When KPi buffer was added to LB medium to investigate if the phosphate buffer and thus the control of pH was the determining factor for dsBGLS to BGLS conversion in TB medium, still only dsBGLS was produced. This suggests that component(s) in the yeast extract are critical for SOT activity (Fig. 4A).

3.4. Towards elimination of metabolic bottlenecks

3.4.1. Overexpression of E. coli PAPS-producing genes

Approximately one third of the produced dsBGLS is converted to the desired product, BGLS, when cultivating the pRI952 strain in TB medium. As mentioned above, the sulfotransferase responsible for this conversion needs PAPS as a co-factor. In E. coli, the genes responsible for producing PAPS in the sulfate assimilatory pathway are cysC encoding for an adenylyl-sulfate kinase and cysN and cysD encoding for two subunits of a sulfate adenylyltransferase (Kopriva and Koprivova, 2004). To increase the levels of available PAPS, a construct harbouring these genes driven by different T7 promoters was generated (Fig. 2C, construct cysCND). When cultures of pRI952 strain expressing the BGLS pathway genes and overexpressing cysCND were supplemented with

---

**Fig. 3.** Production level of dsBGLS in LB medium of E. coli expressing the BGLS biosynthetic genes. A) Seven different E. coli strains were grown for 48 h at 18 °C and the media were analysed for GLS production. B) The BL21 (DE3) and pRI952 strains were grown in media supplemented with either 25 μM phenylacetaldoxime, 75 mg/L 5-aminolevulinic acid or different sulfur sources (20 μM glutathione, 20 μM cysteine, 1 mM thiamine or 20 μM magnesium sulfate). Data represent the average and standard deviation of three biological replicates each grown in three technical replicates.

**Fig. 4.** Medium levels of dsBGLS and BGLS produced by pRI952 E. coli strain grown in different media and with sulfur supplements. A) Production of dsBGLS and BGLS in medium by the pRI952 strain grown in LB, LB with KPi buffer or TB medium. B) as A) except the TB medium is supplemented with different sulfur sources. C) as B) except the strain harbours the cysCND construct in addition to the BGLS biosynthesis genes. Data represent the average and standard deviation of three biological replicates each grown in three technical replicates.
magnesium sulfate or cysteine or both as sulfur sources, no increase in the BGLS levels was observed, on the contrary BGLS levels were reduced (Fig. 4B and C). Similarly, in the strain expressing only the BGLS pathway genes and fed with added cysteine, BGLS levels were also reduced, showing that BGLS production is reduced when cysteine is supplemented, despite overexpression of the PAPS producing genes. However, supplementing with cysteine more than doubled the production of dsBGLS by the wildtype strain compared to the cysCND construct, which could possibly be explained by higher levels of glutathione as a result of endogenous regulation to establish cellular equilibrium. This regulation may have been disrupted in strains expressing the cysCND construct, which could explain the lack of increased dsBGLS production (Fig. 4C).

3.4.2. Effect of choice of promoters and P450 N-terminal anchors

The increase in dsBGLS upon feeding with the first intermediate of the pathway, phenylacetaldoxime, shows that the CYP79A2 activity is limiting (Figs. 4 and 5). A series of new constructs were designed as an alternative approach to increase BGLS production when compared to the unmodified construct. This limitation was improved by protein engineering and increased promoter strength for the gene.

The other bottleneck resulting in dsBGLS accumulation and thus involving the SOT enzyme was addressed with a similar strategy. Increasing the strength of the promoter driving SOT16 expression (Opt 3) did improve the conversion of dsBGLS to BGLS when compared to the unmodified construct. In strains expressing SOT18 from the H10 (Opt 6), the conversion of dsBGLS to BGLS is significantly improved compared to the unmodified construct, but not compared to Opt 3. The construct with highest SOT18 expression (Opt 7) also resulted in the most favourable dsBGLS/BGLS ratio (Fig. 5).

3.4.3. Protein expression levels of BGLS biosynthetic enzymes after optimisation

A targeted proteomics approach was applied to monitor effects of the different optimisation strategies on the protein expression levels. For all proteins, the level was determined from a minimum of two peptides, with the exception of GGP1 for which only one peptide was quantifiable (Fig. 6 and Supplemental Fig. S2). Higher levels of CYP79A2 but not CYP83B1 were detected when the N-terminal anchors were bovine-modified (Fig. 6B and C, Opt 1, 4 & 5), but not when the stronger consensus T7 promoter was used for CYP79A2 (Fig. 6B, Opt 2). When the SOT16 and SOT18 were expressed from H10 and C4 promoters, it resulted in increased SOT levels when compared to expression from the G6 promoter (Fig. 6H and I, Opt 3, 6 and 7). GGP1, SUR1 and ATR1 showed similar protein levels in all samples (Fig. 6A, E & F). Unexpectedly, modifying the anchors on the two P450 enzymes resulted in increased GSTF9 levels, which is unchanged at all other conditions (Fig. 6D, Opt 1, 4, & 5).

4. Discussion

In this work, we report the first production of BGLS in a microbial host and the first quantified GLS production in E. coli. We found that growing the cultures at 18 °C rather than 28 °C and having a cell density of OD$_{600}$ at 0.6 at the time of induction was crucial to obtain production. These factors are known to influence production from heterologous pathways and can vary greatly for different compounds (Galloway et al., 2003; Huyen et al., 2015; San-Miguel et al., 2013; Sørensen and Mortensen, 2005). Previous reports have shown that expression temperatures below 20 °C improved production of
phenylacetaldoxime by CYP79A2 (Miki and Asano, 2014), which corresponds well with our results. Additionally, harvesting after 48 h compared to 24 h as well as choosing the pRI952 strain compared to a standard BL21 (DE3) (Fig. 3A) improved our production 21.5- and 1.6-fold, respectively. The most crucial optimisation for BGLS production was changing to a more nutritious medium composition. In standard LB medium only dsBGLS was produced but upon cultivation of the E. coli strains harbouring the BGLS biosynthetic pathway in TB medium, approximately one third of the dsBGLS was converted to BGLS (Fig. 4A). Both, dsBGLS and BGLS were detected in the medium without first disrupting the cells. This corresponds well to previously published microbial GLS engineering studies, which all found the products extra-cellularly (Mikkelsen et al., 2012; Mirza et al., 2016; Yang et al., 2018).

Metabolic bottlenecks identified in this study are associated with the P450 and SOT enzymes. This is based on an increased production when feeding the CYP79A2 product (phenylacetaldoxime) and optimising the CYP79A2 enzyme expression, as well as improved dsBGLS/ BGLS ratio when optimising the SOT enzymes expression. Many studies have focused on optimising activity of eukaryotic P450 enzymes in E. coli. Being membrane-associated proteins, overexpression can cause severe growth inhibition and influence the energy metabolism of E. coli (Wagner et al., 2007). Particularly, plant P450 enzymes can have problems with proper folding, membrane translocation and toxicity to the host cell (Chemler and Koffas, 2008). In this work, the BGLS-producing E. coli strains were not affected in growth compared to strains with empty vector control, and both CYP79A2 and CYP83B1 - representing the entry point to the engineered pathway - were functionally expressed. When the N-terminals on CYP79A2 and CYP83B1 were modified with the bovine anchor known to improve P450 expression in E. coli (Barnes et al., 1991; Goder and Spiess, 2003; Ichinose and Wariishi, 2013; Miki and Asano, 2014), protein levels of CYP79A2, but not CYP83B1, increased (Fig. 6B and C) along with the production of both dsBGLS and BGLS (Fig. 5), reflecting increased flux through the pathway. This modification also increased levels of GSTF9 (Fig. 6D) that is expressed just downstream of CYP79A2 in the construct. This apparent linked co-expression between CYP79A2 and GSTF9 could be explained by an increased transcription of CYP79A2, which mistakenly continues as a read-through to the neighbouring GSTF9. This is, however, unlikely as the nucleotide sequences are identical apart from the modified CYP anchor and the increased protein levels of GSTF9 might be the result of unknown cellular regulatory mechanisms.

For the metabolic bottleneck associated with the final PAPS-dependent sulfotransferase step converting dsBGLS to BGLS, even the best-producing strain converted little more than half of the dsBGLS to BGLS. As the conversion requires incorporation of sulfate, BGLS production likely affects sulfur metabolism of the host. In N. benthamiana, the same step was identified as a metabolic bottleneck in BGLS production, but could be alleviated by co-expressing an Arabidopsis PAPS kinase, APK2 (Moldrup et al., 2011). Overexpression of E. coli adenylylsulfate kinase (cysC) and sulfate adenylyltransferase (cysND) designed to increase the PAPS pool, did not improve conversion of dsBGLS to BGLS (Fig. 4C). In fact, overexpression of these genes appears to be a disadvantage when the cultures are supplemented with additional cysteine. Although BGLS levels were reduced at any level of cysCND expression when cysteine is supplemented, dsBGLS was produced in approximately twice as high amounts in wildtype strains compared to cysCND-overexpressor strains (Fig. 4B and C). In the former, the additional cysteine may increase the availability of glutathione that is required for the earlier sulfur-incorporating step leading to the thio-glucose moiety in the GLS structure. In the cysCND-overexpressor strain regulation of the entire sulfate assimilation pathway may be compromised as most of the pathway is regulated at a transcriptional level (Rossi et al., 2014). This in turn might hamper the cells ability to channel the additional sulfur efficiently towards glutathione and thus dsBGLS production.

Uptake of sulfur from the environment is tightly regulated and dependent on available as well as preferred sources (Guédon and Martin-Verstraete, 2007). For E. coli, the main sulfur source from the environment is sulfate, while cysteine if present is the preferred source (Kredich, 2008). Sufficient cysteine in the environment represses expression of the sulfate assimilation pathway. This could well explain the complete loss of BGLS when cysteine was fed to the expression
cultures, as available PAPS would decrease or vanish. Conversely, when cysteine is scarce, expression of the assimilation pathway as well as import of sulfate are increased (Kredich, 2008; Rossi et al., 2014). In the assimilation pathway, both PAPS and its precursor APS (adenosine 5’-phosphosulfate) are signalling molecules. APS signals sufficient organic sulfate source and will thus promote the assimilation pathway (Güédon and Martin-Verstraete, 2007). PAPS has broader signalling effects, which are not fully elucidated, but it has been shown to affect global gene transcription through inhibition of nucleotide diphosphate kinase (NdK) (Rossi et al., 2014). Thus, if overexpression of cysCND leads to accumulation of PAPS, this could negatively affect expression of the BGLS pathway genes.

The most promising result with regards to dsBGLS to BGLS conversion was seen in cultures with higher SOT expression. Increasing the SOT enzyme levels by expressing the gene from the strong C4 promoter improved the BGLS/dsBGLS ratio significantly when compared to any other construct. This suggests that the E. coli cells contain a PAPS pool, which the SOT enzymes cannot access and/or utilise efficiently.

5. Conclusion

Optimisation strategies included test of different expression strains, cultivation conditions and media compositions, as well as modification of flux through the exogenous BGLS pathway and the endogenous sulfur metabolism. A targeted proteomics approach uncovered how changes in genetic regulatory elements affected levels of the BGLS biosynthetic enzymes and showed that increasing promoter strength did not necessarily yield higher protein levels, but could in fact result in decreased levels. Similarly, modifying N-terminal anchors on the P450 enzymes according to previously reported bovine modification had different effects depending on the enzyme in question. Interestingly, increased CYP79A2 levels caused several folds increase in the neighbouring GSTP9. Generally, the modifications to the P450 and SOT genes, which are co-located on the pET-52b plasmid, appear to affect the expression levels of other genes expressed from that plasmid. In contrast, genes (GGP1, SUR1, and ATR1) expressed from the pCDF-1b plasmid did not alter protein levels under the different optimisation approaches, with the exception of ATR1 in Opt 5, which is mildly reduced.

The metabolic bottlenecks identified through our work were related to the expression of the two cytochrome P450s and the two sulfur-incorporating steps. The P450 expression was improved by increasing promoter strength for the genes. For the sulfur incorporation, solutions to previously alleviated the same bottleneck in N. benthamiana, did not translate into the microbial host system. Feeding sulfur sources increased dsBGLS production but at the cost of BGLS. This increased dsBGLS production indicates that both sulfur-incorporating steps are limiting for the flux through the pathway. Increased expression of the SOT18 protein improved the ratio of dsBGLS/BGLS. However, no condition or strain tested enabled complete conversion of dsBGLS to BGLS. In summary, our work identified and addressed metabolic bottlenecks and optimised the production to 20.3 μM BGLS accompanied by 13.5 μM dsBGLS.

Author contributions

All authors took part in designing the study. AP performed all experiments and CC performed metabolite and proteomics LC-MS method development and analysis. All authors wrote the manuscript based on a draft provided by AP.

Competing financial interests

The authors declare no competing financial interests.

Acknowledgements

This work was supported by the Danish National Research Foundation [DNRF99]; and the Novo Nordisk Fonden [NNF1400011253]. The authors would like to thank Daniel Vik and Meike Burow for providing isotopically labelled peptides for the targeted proteomics analysis. AP thanks the Department of Plant and Environmental Sciences (University of Copenhagen) for a PhD stipend.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymgen.2019.02.004.

References


RStudio Team, 2015. RStudio. Integrated Development for R. Boston, MA, USA.


