UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE
DEPARTMENT OF PLANT AND ENVIRONMENTAL SCIENCE

Project outside course scope (PUK)

Jens Sigurd Agger Raabyemagle (cvd676)

Functional Screening of UDP-Glycosyltransferases Potentially Capable of Glycosylating the Sorgoleone Precursor Dihydrosorgoleone

Supervisors: Associate Professor Tomas Laursen, Ph.D. Fellow Kasper Hinz, & Ph.D. Silas Mellor

Submitted on: June 16th 2024

Abstract

Nitrification, the conversion of ammonia into nitrate by soil microorganisms, leads to substantial losses in fertilizer efficiency in agriculture. Nitrification also leads to nitrate leaching into our groundwater, as well as the release of the greenhouse gas N_2O into the atmosphere. These problems are easily mitigated by the use of chemical compounds known as nitrification inhibitors. However, only synthetic nitrification inhibitors are on the market, leaving sustainable agricultural practices, such as organic farming, with reduced fertilizer efficiency. The investigation into utilizing biological nitrification inhibitors such as sorgoleone, produced in *Sorghum bicolor*, to enhance ammonia fertilizer efficiency in organic farming has demonstrated promising potential. However, the challenges associated with the production and handling of sorgoleone, primarily due to its hydrophobicity and toxicity, necessitate innovative solutions. A proposed strategy of glycosylating the sorgoleone precursor dihydrosorgoleone to mitigate its toxicity and enable heterologous production efficiency is attempted here by use of UDP-glycosyltransferases (UGTs). Five UGTs known for their substrate promiscuity and/or hydroquinone affinity were screened with hopes of identifying a candidate capable of glycosylating dihydrosorgoleone. The UGTs were expressed in *Escherichia coli*, and their activity towards the tested substrates was assessed using liquid chromatography-mass spectrometry (LC-MS). Despite successful heterologous expression and enzyme assays, none of the UGTs effectively glycosylated dihydrosorgoleone. This highlights the need for alternative strategies or optimization of the experimental procedure used here to allow high heterologous production yields of glycosylated dihydrosorgoleone.

Table of Contents

1. Introduction

1.1. Nitrification in agriculture

In 1928, the German chemist Fritz Haber, recognized as the forefather of nitrogen fixation (and later chemical warfare), was awarded the Nobel Prize in Chemistry for his development of the Haber-Bosch process. This process enabled the industrial production of nitrogen fertilizer in the form of ammonia, significantly enhancing global agricultural productivity and contributing to the exponential growth of the global population since the early 20th century (Smil, 2001). Today, the global population continues to increase, with some projections estimating it will reach up to 10 billion by 2050 (Lutz & Kc, 2010). This escalating population growth underscores the urgent need to further optimize agricultural practices to ensure future food security.

A major challenge in contemporary agriculture is nitrification, which causes up to 50% of the applied ammonium fertilizer to be lost (Beeckman et al., 2018). Nitrification is the biological conversion of reduced nitrogen forms, such as ammonia (NH_4^+), by soil bacteria and archaea into nitrate (NO_3^-). This nitrate can then further act as substrate in denitrification by denitrifying bacteria, resulting in the atmospheric release of N_2O, a potent greenhouse gas (Figure 1). Additionally, nitrification contributes to nitrate leaching, which adversely affects groundwater quality (Tian et al., 2020). One potential solution to mitigate nitrification is the use of synthetic nitrification inhibitors (SNIs), which effectively reduce nitrification rates (Woodward et al., 2021).

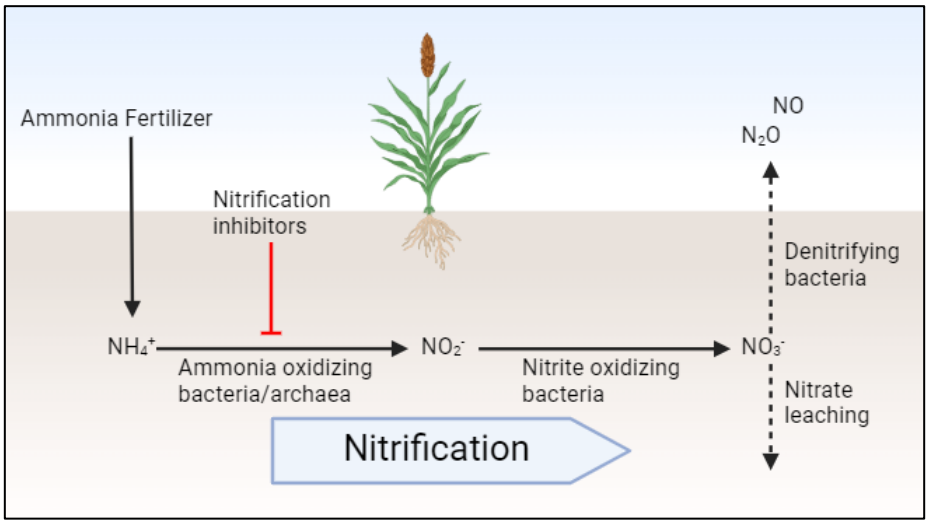


Figure 1: The process of nitrification in soil and potential nitrate sinks.

However, conventional agricultural practices face significant challenges, prompting an increasing demand for more sustainable approaches, such as organic farming. Legislative restrictions on organic farming prohibit the use of SNIs, resulting in fertilizer being less effective in organic agriculture.

1.2. Sorgoleone as a biological nitrification inhibitor

Certain plants produce secondary metabolites that act as biological nitrification inhibitors (BNIs). The use of BNIs offers a potential sustainable solution for improving nitrogen fertilizer efficiency in organic farming systems. A potential BNI is sorgoleone, a secondary metabolite synthesized and exuded through the root hairs of *Sorghum bicolor*. Sorgoleone exhibits phytotoxic and allelopathic properties, as well as inhibition of ammonia-oxidizing bacteria and archaea. Consequently, sorgoleone functions similarly to SNIs by reducing the amount of nitrate being denitrified or leaching into the ground (Sarr et al., 2020).

Sorgoleone comprises an acyl chain attached to a benzoquinone head group, rendering it highly hydrophobic due to its lipid-like tail, and toxic due to the chemical properties of the head group (Tibugari et al., 2020). Sorgoleone is also found in various analogs missing one or more of the double bonds on its acyl chain (Dayan et al., 2010). Plants typically have measures that allow them to store their secondary metabolites intracellularly, where accumulation might distress the plants basic physiology if toxic and/or very hydrophobic (Hefner et al., 2002). Sorgoleone is on the other hand exuded by the root hairs of *S. bicolor* in very small quantities, thereby avoiding interference with its basic metabolism (Pan et al., 2021). Because of the low production in *S. bicolor*, it would be optimal to heterologously express the biosynthetic pathway for sorgoleone, which has been fully elucidated (Figure 2), in a suitable production host and enhance production yields to use sorgoleone effectively as a BNI in organic agriculture. However, due to its toxicity, one would anticipate autotoxicity in a potential host organism once a certain threshold is reached, thereby impeding production.

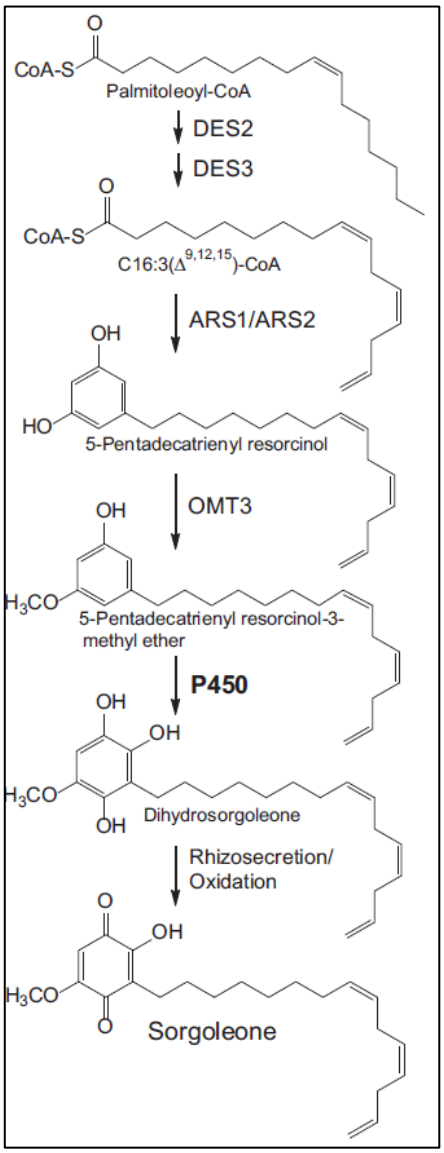


Figure 2: The biosynthetic pathway for the production of sorgoleone in S. bicolor. (Pan et al., 2018)

The final step in the biosynthetic pathway of sorgoleone is the spontaneous oxidation of dihydrosorgoleone. Because of this spontaneous step, dihydrosorgoleone poses an interesting target for detoxification by glycosylation of either or both of the hydroxyl groups in para position which undergo this spontaneous oxidization. This would permit higher heterologous production titers as well as easier handling of the product. Glucose moieties are also easily removable by non-specific β-glucosidases, which could be added to glycosylated dihydrosorgoleone prior to use in organic agriculture.

1.3. Proposed UGTs for glycosylation of dihydrosorgoleone

UDP glycosyltransferases (UGTs) comprise a superfamily of enzymes present in plants, bacteria, mammals, and fungi, playing a crucial role in the metabolism and detoxification of various compounds (Campbell et al., 1997). Their mechanism of action involves conjugating a sugar moiety, such as glucose, glucuronic acid, xylose, galactose, or N-acetylglucosamine, to a small nucleophilic substrate, such as a hydroxyl group on dihydrosorgoleone (Hu et al., 2019). For instance, a UGT from *Arabidopsis thaliana* has been shown to confer resistance to the mycotoxin deoxynivalenol by glycosylating and detoxifying the compound, thereby reducing its bioavailability (Poppenberger et al., 2003). Vanillin, an industrially sought after compound, is another example of a toxic compound, which by means of glycosylation by UGTs can be converted into the less toxic vanillin β-D-glucoside. Introducing the glycosylation step into the biosynthetic pathway for vanillin therefore allows for higher product accumulation in heterologous hosts (Liu et al., 2023).

Since there are no known UGTs capable of glycosylating dihydrosorgoleone, it is necessary to identify UGTs from the literature that can either glycosylate compounds with similar chemical properties to dihydrosorgoleone or exhibit high substrate promiscuity, to realize this task.

1.3.1. UGT72B27

UGT72B27 (Uniprot: D7UCM3) is a glycosyltransferase found in grapevines (*Vitis vinifera*), which are often exposed to volatile bioactive phenolic compounds from bush and forest fire derived smoke. These compounds can affect grapevine health and wine quality. UGT72B27 was identified in a screen for *V. vinifera* glycosyltransferases capable of glycosylating smoke-derived phenolics, such as guaiacol. This enzyme exhibited activity towards more than ten different smoke-derived substrates, demonstrating its promiscuity and specificity towards phenolic compounds, making it a promising candidate for this project (Härtl et al., 2017).

1.3.2. UGT74AN3

UGT74AN3 (Uniprot: AYC35244) is another highly promiscuous glycosyltransferase, originating from *Catharanthus roseus*. Initially characterized for its activity towards cardiotonic steroids and phenolic compounds, later studies revealed its catalytic activity towards 78 out of 92 tested sugar acceptor substrates and six different UDP-sugar donors. UGT74AN3 can produce both mono- and di-glucosides, with acceptor substrates including flavonoids, steroids, coumarins, phenols, terpenoids, and anthraquinones, showcasing its very high substrate promiscuity (Wen et al., 2020; Huang et al., 2024).

1.3.3. UGT84A95

UGT84A95 (Genbank: OQ581042) was identified in a study aimed at finding UGTs capable of glycosylating phenylethanoid glycosides, which are large compounds produced by many medicinal plants. Expressed in *Penstemon barbatus*, UGT84A95 demonstrated high promiscuity, catalyzing the

conjugation of glucose to 23 out of 28 tested acceptor substrates, including phenylethanoid glycosides, flavonoids, terpenoids, coumarins, and simple polyphenols. This UGT was also able to both produce mono- and di-glucosides (Wu et al., 2023).

1.3.4. Arbutin synthase

Arbutin synthase (Uniprot: Q9AR73) is a UGT derived from *Rauvolfia serpentina*, and named so because of its ability to produce arbutin, a skin whitening agent, by glycosylation of hydroquinone. Despite not achieving conversion rates as high as with hydroquinone, arbutin synthase is still able to catalyze the glycosylation of 45 out 74 tested sugar acceptor substrates, including diverse natural and synthetic phenols as well as cinnamyl alcohols (Hefner et al., 2002).

1.3.5. SlUGT5

In tomatoes (Solanum lycopersicum), volatile compounds contributing to aroma and flavor are often stored in a glycosylated form. This led to the identification of SlUGT5 (Uniprot: HM209439), a member of the UGT72 family, which showed activity towards a range of acceptor substrates, including flavonoids, flavonols, and hydroquinone (Baldwin et al., 2000; Louveau et al., 2011).

1.4. Scope of the project and experimental strategy

The aim of this study is to identify a UGT which is be able to glycosylate and thereby detoxify dihydrosorgoleone. This would hopefully permit eventual heterologous production hosts to produce high titers of dihydrosorgoleone-glucoside without experiencing autotoxicity, and mitigate the need for protective measures in downstream handling of the BNI. The experimental strategy for this project is to functionally screen five different UGTs, known from literature to have acceptor substrate promiscuity and/or have an affinity towards hydroquinone properties. These UGTs will then be expressed in model organism *Escherichia coli*. Assays are set up using the recombinant *E. coli* lysates containing the respective UGTs incubated with UDP-glucose and either hydroquinone, hexyl resorcinol, or *S. bicolor* root extract containing dihydrosorgoleone as sugar acceptor substrate, with the hopes of producing glycosylated versions of these (Figure 3). The resulting reaction would then be analyzed by liquid chromatography coupled to mass spectrometry (LC-MS) to assess the assay success and putative glycosylated products.

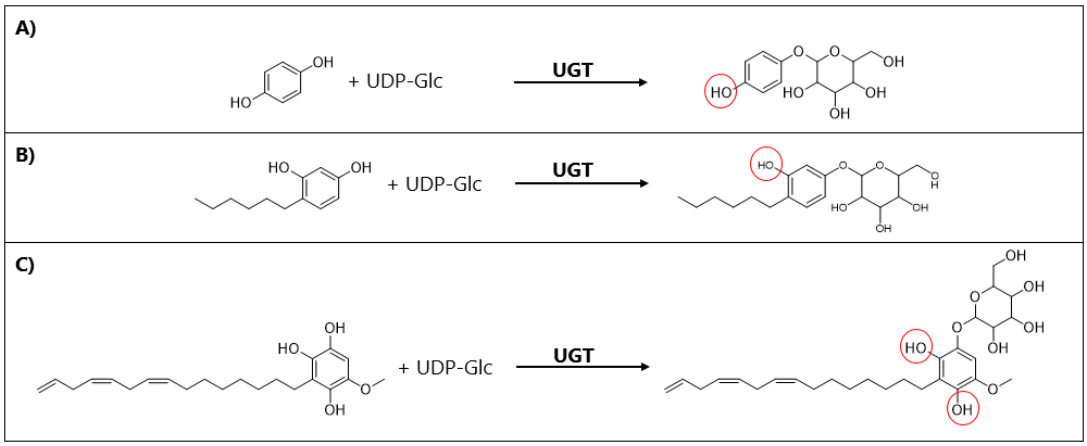


Figure 3: A) UGT mediated glycosylation of hydroquinone. B) UGT mediated glycosylation of hexyl resorcinol. C) UGT mediated glycosylation of dihydrosorgoleone. Circled hydroxyl groups in the products highlight potential glycosylation targets beyond the illustrated sites.

2. Materials and methods

2.1. Dihydrosorgoleone extraction

Approximately 40 to 50 sorghum seeds (BTx623) were imbibed in 40 mL of water and incubated in darkness overnight at 28°C. Subsequently, 20 to 25 of these seeds were placed in disposable plastic boxes lined with three layers of filter paper moistened with 20 mL of water. The seeds were then incubated for five days in darkness at 28°C. Following incubation, all roots were excised from the sprouting seeds, and the proximal segment of each root was immersed in 2 mL of 10% methanol for 30 seconds. The resulting extract was concentrated using a SpeedVac system set to 1000 rpm (Scanspeed 32, Holm & Halby).

2.2. Vector design

The UGT expression vectors used were custom synthesized by Twist Bioscience. Each vector was designed to include the respective UGT genes, with codons optimized for *E. coli* expression, inserted into a pET29b(+) vector backbone.

2.3. Heat-shock transformation of NiCo21(DE3) *E. coli* cells

Chemically competent NiCo21(DE3) *E. coli* cells (New England Biolabs) were thawed on ice. Aliquots of 25 µL of cells were incubated on ice for 30 minutes with 50 ng of plasmid pET29b(+) DNA containing the respective UGT inserts, the positive control (UGT85B1) and the negative control (empty vector). The cells were then heat-shocked at 42°C for 10 seconds, followed by a 5-minute incubation on ice. To enable the expression of kanamycin resistance genes, the cells were allowed to recover in 975 µL super optimal catabolite (SOC) media for 1 hour at 37°C with vigorous shaking at 250 rpm. Subsequently, 50 µL aliquots of the cell cultures were spread onto Luria-Bertani (LB) agar plates containing kanamycin at a concentration of 100 µg/mL.

2.4. UGT expression

UGT expressing *E. coli* colonies were respectively suspended in 4 mL terrific broth (TB) media containing kanamycin at a concentration of 100 µg/mL and incubated overnight at 37°C shaking at 250 rpm. An expression culture of 20 mL TB media containing kanamycin at a concentration of 100 µg/mL was inoculated with overnight cultures to an optical density (OD_{600}) of 0.2 in 100 mL Erlenmeyer flasks. These cultures were then incubated at 37°C shaking at 250 rpm until reaching mid-log phase (OD_{600}=0.4−0.8). Expression was then induced by the addition of 500 µM IPTG, followed by incubation at 37°C for 4 hours shaking at 250 rpm. SDS-PAGE analysis was performed to assess the expression levels of the *E. coli* UGT transformants. Aliquots of 200 µL from each culture were centrifuged for 1 minute at 10,000 rcf, and the supernatant was discarded. The pellets were resuspended in 120 µL of SDS-PAGE loading buffer and lysed at 95°C for 5 minutes. Then, 10 µL of the supernatant containing total soluble *E. coli* protein was loaded onto a Criterion™ TGX Stain-Free™ Precast Gel (Bio-Rad), and electrophoresis was conducted at 300 V for 20 minutes. The resulting protein bands were visualized using a ChemiDoc™ UV transilluminator (Bio-Rad).

A western blot was performed to confirm that the bands observed in the SDS-PAGE gel corresponded to the proteins of interest. Proteins from the gel was transferred to a nitrocellulose membrane using a Trans-Blot Turbo Transfer Pack (Bio-Rad) according to manufacturer's instructions and processing using a Trans-Blot® Turbo™ Transfer System (Bio-Rad) running an integrated program for mixed molecular weights. The membrane was then incubated in 50 mL of 5% blocking buffer overnight at 4°C with shaking at 150 rpm. Subsequently, monoclonal anti-6xHis primary antibodies from mouse were added directly to the blocking buffer at a 1:3000 dilution and incubated at room temperature with shaking at 150 rpm for 1 hour. The buffer was then removed, and the membrane was washed six times with PBS-Tween buffer, each wash lasting 30 seconds, to remove non-bound primary antibodies. The membrane was then incubated with PBS-Tween buffer containing polyclonal rabbit anti-mouse secondary antibodies conjugated to horseradish peroxidase (HRP) at a 1:3000 dilution for 1 hour at room temperature with shaking at 150 rpm. This was followed by five additional washing steps with PBS-Tween buffer as described previously. Equal volumes of two substrate solutions (SuperSignal™ West Dura Extended Duration Substrate, Thermo-Scientific™) were mixed and applied to the membrane, which was then incubated at room temperature for 5 minutes with shaking at 150 rpm. The blot was finally visualized using a ChemiDoc™ UV transilluminator (Bio-Rad).

2.5. UGT functional screen

The functional screening of the five different UGT's, as well as the negative control, was conducted as follows: expression cultures were transferred to 50 mL Falcon tubes and subjected to centrifugation at 4000 rcf for 5 minutes. The supernatant was discarded, and the resulting cell pellet was resuspended in 5 mL of 100 mM Tris-HCl buffer (pH 7.5). Cell suspensions were sonicated using a Sonifier® Cell

Disruptor (Branson) set to a 50% duty cycle and an output level of 2, applied in ten intervals of 30 seconds of sonication followed by 1 minute of cooling on ice.

The enzyme assays were then prepared in 100 µL reactions containing 88-89 µL of crude enzyme extract, 1 mM UDP-glucose (UDPG) (dissolved in deionized water), and 1 mM glucose acceptor substrate (either hydroquinone or hexyl resorcinol) (dissolved in deionized water) or 1 µL of dihydrosorgoleone extract (dissolved in 100 µL of 100% DMSO). The reactions were incubated for 2 hours at the optimal temperatures for each UGT: UGT72B27 at 21°C, UGT74AN3 at 37°C, UGT84A95 at 30°C, Arbutin Synthase at 50°C, and SlUGT5 at 37°C.

To prepare the samples for LC-MS analysis, the reactions were quenched with 300 µL of 100% methanol and incubated on ice for 10 minutes to facilitate compound extraction. The samples were then centrifuged for 10 minutes at 20000 rcf. Subsequently, 20 µL of the supernatant was combined with 150 µL of water and 25 µL of 100% methanol in a filter plate, followed by centrifugation at 3000 rcf for 5 minutes. Finally, the filtered samples were transferred to glass vials with inserts for LC-MS analysis.

2.6. LC-MS analysis of functional screen

Assay samples were subjected to ultrahigh pressure liquid chromatography-mass spectrometry. Filtered samples in volumes of 2µL were injected into a Thermo Scientific UltiMate 3000 UHPLC system. Samples were separated on a 1.7 µm Kinetex® XB-C18 column (Phenomenex) at 40°C with a flow rate of 0.3 mL minute^{-1}. The mobile phase consisted of solvent A (0.005% formic acid in water) and solvent B (0.05% formic acid in 100% acetonitrile) and a gradient of 2%-100% B was applied over a 35 minute elution program. The LC was coupled to a Bruker Daltonics Compact QqToF mass spectrometer, where samples were subject to electrospray ionization in positive mode. Results were analyzed using Bruker Compass DataAnalysis 4.3 software.

3. Results

3.1. Heterologous expression of UGT's

The experimental workflow of this project necessitated a host organism expressing the UGTs of interest at an adequate level. To achieve this, vectors containing the UGT inserts were transformed into *E. coli* cells, which then were cultured with respects to optimal overexpression. Cells were then lysed, and total *E. coli* protein was analyzed by SDS-PAGE to assess the expression profiles of the respective UGTs (Figure 4.A). UGT85B1, known to express well in *E. coli*, was included as a positive control, while an *E.* coli culture transformed with an empty vector was used as a negative control. Bands corresponding to inserts with C-terminal 6xHis tags were expected as follows: UGT72B27 at 52.36 kDa, UGT74AN3 at 54.35 kDa, UGT84A95 at 55.08 kDa, Arbutin synthase at 52.62 kDa, SlUGT5 at 53.32 kDa, and the positive control at 53.75 kDa. There are clear indications of good expression for all of the UGTs. The

bands correspond well with the expected sizes when comparing them to each other, however there seems to be a downshift in expected size by around 5-10 kDa when comparing to the protein ladder.

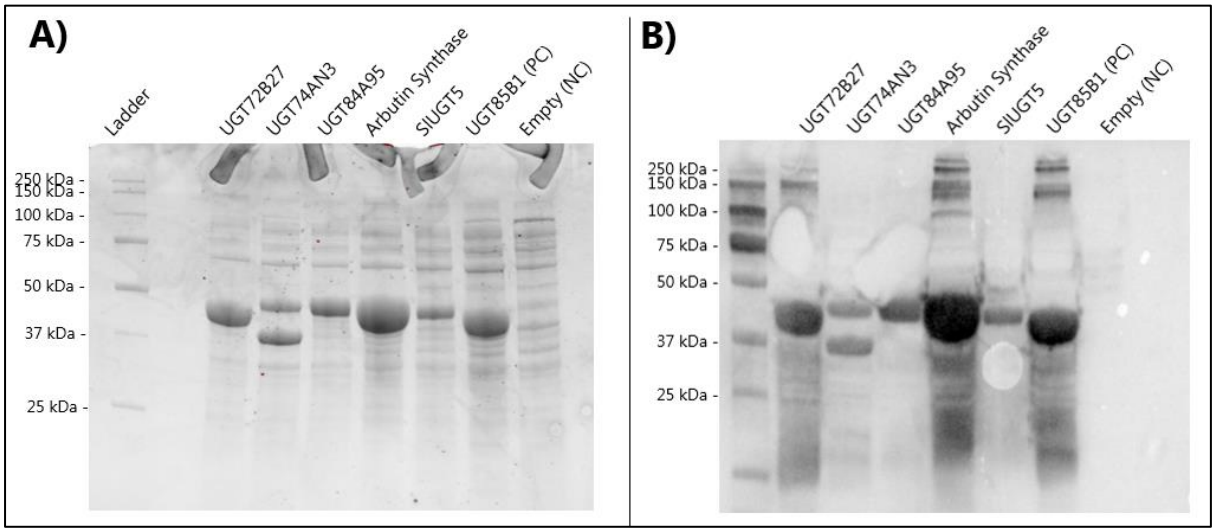


Figure 4: A) SDS-PAGE gel analysis of total E. coli protein. Labels indicate the respective UGT transformants. Positive control (PC) is UGT85B1. Negative control (NC) is E. coli transformed with an empty vector. B) Western blot made with gel from A) with primary anti-6xHis antibodies. Ladder is Precision Plus Protein Standards (BioRad).

To further verify that the bands seen in the SDS-PAGE gel were indeed the desired enzymes, a western blot was conducted, using the same gel (Figure 4.B). Primary anti-6xHis antibodies were used to bind specifically to the His-tagged enzymes. Comparing the samples to the negative control containing an empty vector confirms that the bands are indeed the inserted UGTs, and that they are expressed well. There is some potential aggregation when looking at the top of the blot, as well as there being a lot of unspecific binding. Looking at UGT74AN3, it is evident that there is one band at the expected size as well as another around 10 kDa smaller.

3.2. Functional screening of UGTs

To determine if the proposed UGTs could glycosylate dihydrosorgoleone, a functional screen was established. Crude enzyme extracts were provided with UDP-glucose as the sugar donor and either hydroquinone, hexyl resorcinol, or root extract containing dihydrosorgoleone as the sugar acceptor in equimolar concentrations. The reactions were then incubated at the optimal temperatures for each respective UGT, as specified in the relevant literature.

To evaluate the assay parameters and confirm the proper folding of the enzymes post-expression, a positive control was included. Arbutin synthase, which utilizes hydroquinone as a substrate to produce arbutin, served as the positive control. Additionally, hydroquinone was tested with the other UGTs. Since dihydrosorgoleone possesses a quinone head group, any observed reaction with hydroquinone could indicate the potential for future enzyme engineering to enable glycosylation of dihydrosorgoleone. LC-MS analysis was conducted to assess whether the reactions had produced

glycosylated versions of the provided substrates. The positive control is able to produce detectable levels of arbutin, indicating that the assay parameters are satisfactory (Figure 5). The negative control also does not show any production of arbutin. The extracted ion chromatogram (EIC) for hydroquinone shows a peak corresponding to the hydroquinone molecular ion without addition of adducts at m/z = 110.0362 in all screens, albeit the peak for the standard has a much higher peak. There is a peak for the sodium adduct of arbutin present in samples for arbutin synthase as well as UG72B27. As expected, a second hydroquinone peak is present in the samples able to produce arbutin, caused by the cleavage of the glucose moiety of arbutin because of the relative weakness of the glycosidic bond.

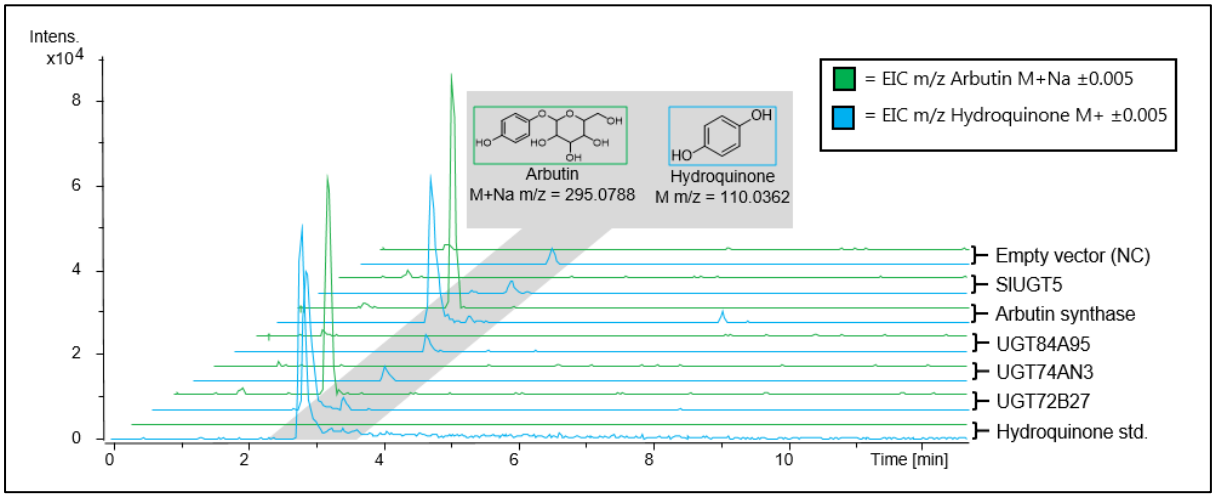


Figure 5: Extracted ion chromatograms of hydroquinone (blue) and the glycosylation product arbutin (green) for the hydroquinone standard (1µM in 20% methanol) and the respective UGT assays conducted with hydroquinone as sugar acceptor substrate. UGT assay samples are representative of triplicates.

The assay was also conducted applying hexyl resorcinol as a sugar acceptor substrate in the reactions. When looking at the EIC for substrate and product in this reaction (Figure 6), it is apparent that there is something wrong. There are no peaks in the chromatogram for hexyl resorcinol in the standard when looking for either the molecular ion, the protonated molecule or the sodium adduct. There is a small peak for hexyl resorcinol at m/z = 195.1380 corresponding to the protonated molecule in the samples for UGT74AN3 and SlUGT5, and the same peak with a much higher relative intensity in the sample containing arbutin synthase. No glycosylated hexyl resorcinol is present when looking at EICs for the molecular ion, the protonated molecule, nor the sodium adduct.

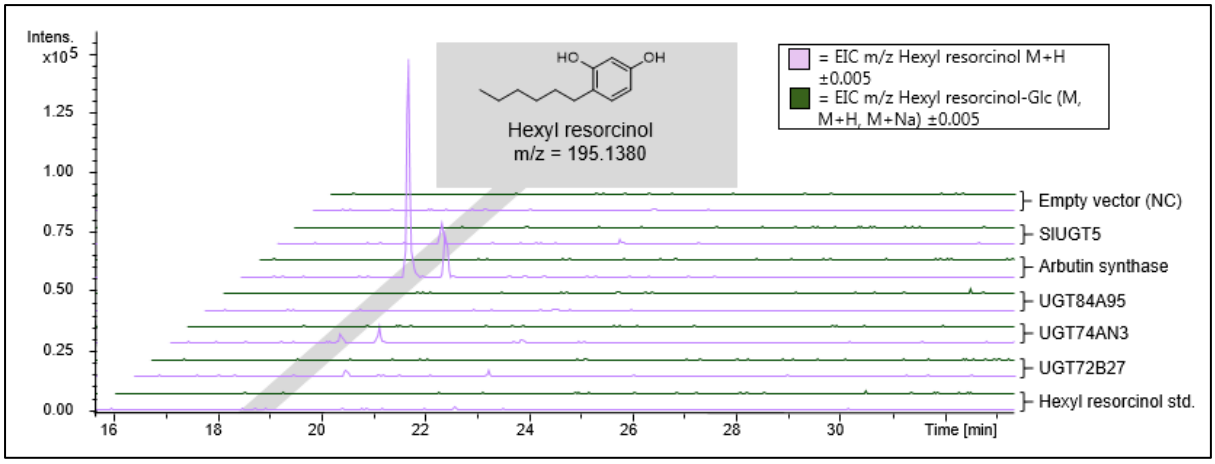


Figure 6: Extracted ion chromatograms of hexyl resorcinol (pink) and glycosylated hexyl resorcinol (dark green) for the hexyl resorcinol standard (1µM in 20% methanol) and the respective UGT assays conducted with hexyl resorcinol as sugar acceptor substrate. UGT assay samples are representative of triplicates.

The final sugar acceptor substrate used in the assay was the dihydrosorgoleone extract, with the hopes of being able to produce a glycosylated dihydrosorgoleone (Figure 7). It was possible to find an ion peak in the EIC corresponding to the m/z value for the sodium adduct of dihydrosorgoleone in all of the samples, albeit at a rather low relative intensity. The dihydrosorgoleone peak for the root extract standard and UGT74AN3 sample seems to elude around 15 seconds before the other peaks. No glycosylated product was able to be detected when looking for either the molecular ion, the protonated molecule, nor the sodium adduct. No di-glucosides were found either (data not shown).

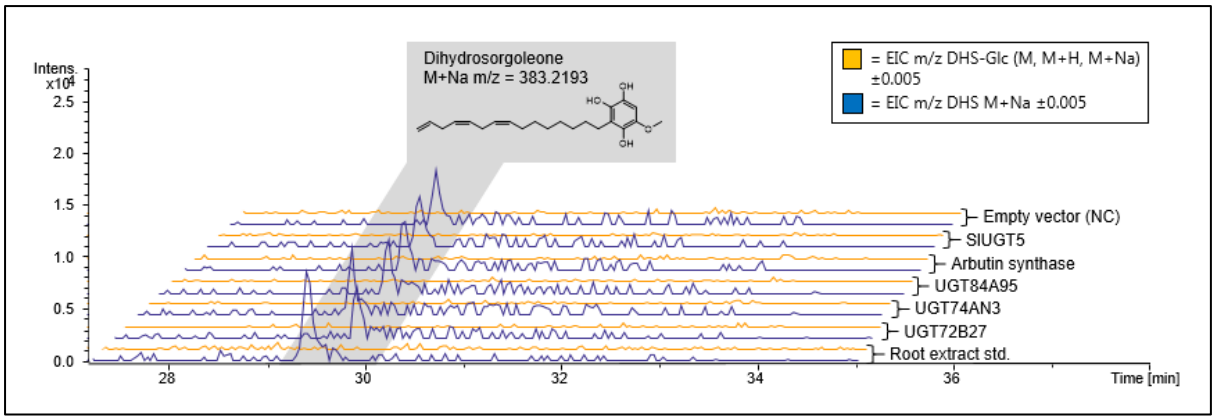


Figure 7: Extracted ion chromatograms of dihydrosorgoleone (DHS) (dark blue) and the glycosylated dihydrosorgoleone (DHS-Glc) (yellow) for the root extract standard (Concentrate dissolved in DMSO) and the respective UGT assays conducted with root extract containing dihydrosorgoleone as sugar acceptor substrate. UGT assay samples are representative of triplicates.

Addressing the challenge of the minor analogs of sorgoleone that may be present, they were searched for by looking for the EIC molecular ion, protonated molecule, and the sodium adduct of sorgoleone, sorgoleone-360, and sorgoleone-362 (so named because of their change in m/z ratio as a result of loss of acyl chain double bonds). The sodium adduct of all three were present in the root extract (Figure 8), but notably the peak for sorgoleone-360 is the same peak seen in figure 7 for dihydrosorgoleone as they

have the same mass to charge ratio. However, neither sorgoleone nor sorgoleone-362 are to be found in the assay samples, which are fed with the root extract sample from figure 8 (Appendix 1).

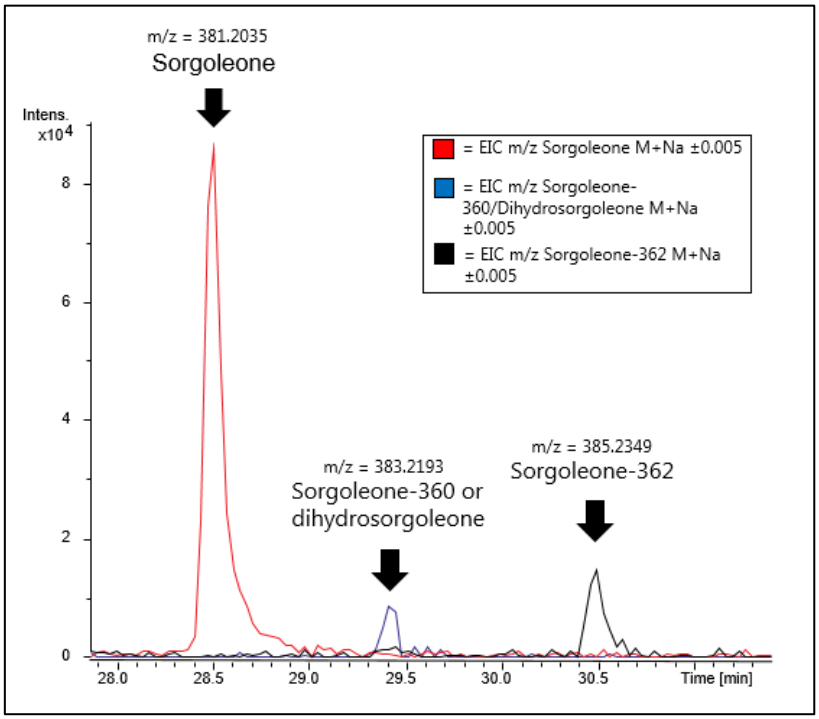


Figure 8: Extracted ion chromatograms for sorgoleone (red), sorgoleone-360 (dark blue), and sorgoleone-362 (black) from the root extract sample used in the assays.

4. Discussion

4.1. UGT expression in *E. coli*

On both the gel and the western blot in figure 4, the respective protein bands do not appear to migrate exactly according to their expected molecular weight. Comparing to the protein ladder, all the proteins exhibited a gel shift of around 10 kDa smaller than expected. Although plant UGTs widely are believed to be cytosolic (Ross et al., 2001), there is no literature on the cellular localization of the UGTs assayed in this project, meaning that them being membrane associated cannot be ruled out. It has been reported that membrane proteins can migrate at rates up to 30% faster than their calculated molecular weights due to the amount of hydrophobic residues which can affect the SDS-PAGE gel migration pattern, thus explaining the shift (Rath et al., 2009). Although, when observing the total *E. coli* protein in the UGT74AN3 transformant, there are two bands that seem equally expressed. Both bands here are present in the western blot, indicating that they likely both are the inserted UGT. An explanation to this could that some of the protein experienced N-terminal proteolytic degradation by proteases released under cell lysis. Here it would be interesting to add protease inhibitor to the cells under culturing and/or before lysis. There is also a lot of unspecific binding in the western blot, indicating either that there was too

few washing steps in the method, or that the smears above and under the bands are aggregates of the proteins of interest and different degradation patterns, respectively. Generally, the western blot confirmed that the contrast full bands in the SDS-PAGE gel were in fact the expected UGTs and that they were expressing well, which led to continuation in the experimental workflow.

4.2. Assay conditions and LC-MS data

Although the expression of the respective UGTs appeared satisfactory based on the western blot results, there remained the challenge of determining if they were folding correctly and functional in a prokaryotic host, which lacks the machinery for post-translational modifications. To evaluate this, hydroquinone was included as an acceptor substrate in the assay, since it is the natural substrate of arbutin synthase, thereby serving as an integrated positive control, on top of it also having similar chemical properties to dihydrosorgoleone. The putative sodium adduct of arbutin, the product of arbutin synthase, was detected in an EIC of the positive control assay sample (Figure 5). Additionally, an extra hydroquinone peak eluted at the same time as arbutin. This is expected, as the glycosidic bond is relatively weak, leading to the loss of the glucose moiety and resulting in the aglycone ion eluting concurrently (March et al., 2004). The UGT72B27 assay samples were also able to glycosylate hydroquinone, which highlights its further potential in future work regarding glycosylation of dihydrosorgoleone. An arbutin standard would be ideal for confirming the exact elution time, so one was prepared and included in the LC-MS run. However, due to the depletion of calibrant solution, the last two samples in the run containing the arbutin standards could not be included in the analysis.

Hexyl resorcinol was also included as a substrate in the assay due to its structural similarity to dihydrosorgoleone and the ability to precisely determine its concentration for the assay, unlike dihydrosorgoleone in the root extract. Despite this, the hexyl resorcinol standard did not show any peaks in the LC-MS data (Figure 6), which was initially puzzling. The stock solution of hexyl resorcinol used in the assay and as the standard was prepared as 50 mM hexyl resorcinol dissolved in water, which then was further diluted in water to applied concentrations. This presented a source of error, as hexyl resorcinol is a hydrophobic compound and should be dissolved in an organic solvent, such as DMSO, as done with the root extract. Since hexyl resorcinol could not be detected in all of the samples, these data are rendered obsolete.

Regarding the assay conducted with root extract as the substrate, several questions arise. Firstly, peaks corresponding to the sodium adduct of dihydrosorgoleone are observed in all samples (Figure 7). Notably, the peaks for the root extract standard and the UGT74AN3 sample elute 15 seconds earlier than those in the other samples. This discrepancy is likely due to a technical anomaly, as a comparison of the elution times in the base peak chromatograms (BPCs) of the UGT74AN3 and UGT72B27 samples reveals a consistent 15-second shift across all peaks (Appendix 2).

Secondly, the presence of an analog, sorgoleone-360, which shares the same m/z ratio as dihydrosorgoleone, complicates the identification of the peak eluting at 29.5 minutes. Examination of the extracted ion chromatograms (EICs) from the root extract (Figure 8) indicates the presence of sorgoleone 362, suggesting that other analogs of sorgoleone may also be present. To distinguish between dihydrosorgoleone and sorgoleone-360, one could analyze the fragmentation pattern of the M+Na ion with m/z = 383.2193 or purify the compound and perform NMR spectroscopy, thereby ensuring that the correct substrate is being used as done in Pan et al. (2018). If the compound is not dihydrosorgoleone, then the results in figure 7 cannot exclude the possibility that any of the five screened UGTs exhibit affinity towards dihydrosorgoleone. However, neither sorgoleone nor sorgoleone-362 is detectable in any of the assay samples (Appendix 1), whereas dihydrosorgoleone/sorgoleone-360 is detectable in all of the assay samples. This could indicate that sorgoleone and its analogs might react with components in the crude enzyme extract, hindering their ionization, or that they interact with cellular debris, preventing their presence in the supernatant of the assay mixture and subsequent injection into the LC-MS system. If this is the case, it could be assumed that sorgoleone-360 behaves similarly, thereby strengthening the possibility that the observed peak corresponds to dihydrosorgoleone.

5. Conclusion and perspectives

The heterologous expression of five UGTs in *E. coli* was successfully achieved, as confirmed by SDS-PAGE and western blot analyses, and all exhibited adequate expression levels, indicating their potential for glycosylation activities. Functional assays with the UGTs using hydroquinone as a model substrate demonstrated that UGT72B27 and Arbutin Synthase were capable of producing glycosylated products, validating the assay conditions and the enzymes' functionality, as well as directing extra attention to these because of their affinity towards hydroquinone. Coupling docking experiments and enzyme engineering of UGTs with these affinities could prove useful in forcing affinity for dihydrosorgoleone.

Despite these positive indications, the assays conducted with root extract containing dihydrosorgoleone did not yield detectable glycosylated products. This outcome, on top of the uncertainty regarding the actual availability of dihydrosorgoleone, suggests that further optimization is required for this assay. Overall, while the goal of detoxifying dihydrosorgoleone through glycosylation was not fully realized, the study establishes a foundational method and identifies areas for further research. Future efforts should focus on refining the selection and engineering of UGTs, improving assay conditions, and potentially through exploring additional UGT candidates. Developing a high-throughput screen that would allow efficient and cost-effective screening of many UGTs at a time would be an ideal way to realize the aim of the project. These advancements will be crucial for the industrial potential of sorgoleone as a sustainable BNI, contributing to the broader goal of optimizing organic agricultural practices to meet the demands of a growing global population.

6. Literature

Baldwin, E. A., Scott, J. W., Shewmaker, C. K., & Schuch, W. (2000). Flavor Trivia and Tomato Aroma: Biochemistry and Possible Mechanisms for Control of Important Aroma Components. *HortScience*, *35*(6), 1013–1022. https://doi.org/10.21273/HORTSCI.35.6.1013

Beeckman, F., Motte, H., & Beeckman, T. (2018). Nitrification in agricultural soils: impact, actors and mitigation. In *Current Opinion in Biotechnology* (Vol. 50, pp. 166–173). Elsevier Ltd. https://doi.org/10.1016/j.copbio.2018.01.014

Campbell, J. A., Davies, G. J., Bulone, V., & Henrissat, B. (1997). Campbell et al. 1997 - A classification of nucleotide diphospho sugar glycosyltransferases based on amino acid sequence similarity. *Biochemical Journal*, *326*(3), 929–942. https://doi.org/10.1042/bj3260929u

Dayan, F. E., Rimando, A. M., Pan, Z., Baerson, S. R., Gimsing, A. L., & Duke, S. O. (2010). Sorgoleone. In *Phytochemistry* (Vol. 71, Issue 10, pp. 1032–1039). Elsevier Ltd. https://doi.org/10.1016/j.phytochem.2010.03.011

Härtl, K., Huang, F. C., Giri, A. P., Franz-Oberdorf, K., Frotscher, J., Shao, Y., Hoffmann, T., & Schwab, W. (2017). Glucosylation of Smoke-Derived Volatiles in Grapevine (Vitis vinifera) is Catalyzed by a Promiscuous Resveratrol/Guaiacol Glucosyltransferase. *Journal of Agricultural and Food Chemistry*, *65*(28), 5681–5689. https://doi.org/10.1021/acs.jafc.7b01886

Hefner, T., Arend, J., Warzecha, H., Siems, K., & Sto¨ckigt, J. S. (2002). Arbutin Synthase, a Novel Member of the NRD1 Glycosyltransferase Family, is a Unique Multifunctional Enzyme Converting Various Natural Products and Xenobiotics. *Bioorganic & Medicinal Chemistry*, *10*, 1731–1741. https://doi.org/10.1016/S0968-0896(02)00029-9

Huang, W., Zhang, X., Li, J., Lv, J., Wang, Y., He, Y., Song, J., Ågren, H., Jiang, R., Deng, Z., & Long, F. (2024). Substrate Promiscuity, Crystal Structure, and Application of a Plant UDP-Glycosyltransferase UGT74AN3. *ACS Catalysis*, *14*(1), 475–488. https://doi.org/10.1021/acscatal.3c05309

Hu, D. G., Hulin, J. ulie A., Nair, P. C., Haines, A. Z., McKinnon, R. A., Mackenzie, P. I., & Meech, R. (2019). The UGTome: The expanding diversity of UDP glycosyltransferases and its impact on small molecule metabolism. In *Pharmacology and Therapeutics* (Vol. 204). Elsevier Inc. https://doi.org/10.1016/j.pharmthera.2019.107414

Liu, Y., Sun, L., Huo, Y. X., & Guo, S. (2023). Strategies for improving the production of bio-based vanillin. In *Microbial Cell Factories* (Vol. 22, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s12934-023-02144-9

Louveau, T., Leitao, C., Green, S., Hamiaux, C., Van Der Rest, B., Dechy-Cabaret, O., Atkinson, R. G., & Chervin, C. (2011). Predicting the substrate specificity of a glycosyltransferase implicated in the production of phenolic volatiles in tomato fruit. *FEBS Journal*, *278*(2), 390–400. https://doi.org/10.1111/j.1742-4658.2010.07962.x

Lutz, W., & Kc, S. (2010). Dimensions of global population projections: What do we know about future population trends and structures? In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 365, Issue 1554, pp. 2779–2791). Royal Society. https://doi.org/10.1098/rstb.2010.0133

March, R. E., Miao, X. S., Metcalfe, C. D., Stobiecki, M., & Marczak, L. (2004). A fragmentation study of an isoflavone glycoside, genistein-7-O-glucoside, using electrospray quadrupole time-of-flight mass spectrometry at high mass resolution. *International Journal of Mass Spectrometry*, *232*(2), 171–183. https://doi.org/10.1016/j.ijms.2004.01.001

Pan, Z., Baerson, S. R., Wang, M., Bajsa-Hirschel, J., Rimando, A. M., Wang, X., Nanayakkara, N. P. D., Noonan, B. P., Fromm, M. E., Dayan, F. E., Khan, I. A., & Duke, S. O. (2018). A cytochrome P450 CYP71 enzyme expressed in Sorghum bicolor root hair cells participates in the biosynthesis of the benzoquinone allelochemical sorgoleone. *New Phytologist*, *218*(2), 616–629. https://doi.org/10.1111/nph.15037

Pan, Z., Bajsa-Hirschel, J., Vaughn, J. N., Rimando, A. M., Baerson, S. R., & Duke, S. O. (2021). In vivo assembly of the sorgoleone biosynthetic pathway and its impact on agroinfiltrated leaves of Nicotiana benthamiana. *New Phytologist*, *230*(2), 683–697. https://doi.org/10.1111/nph.17213

Poppenberger, B., Berthiller, F., Lucyshyn, D., Sieberer, T., Schuhmacher, R., Krska, R., Kuchler, K., Glössl, J., Luschnig, C., & Adam, G. (2003). Detoxification of the Fusarium Mycotoxin Deoxynivalenol by a UDP-glucosyltransferase from Arabidopsis thaliana. *Journal of Biological Chemistry*, *278*(48), 47905–47914. https://doi.org/10.1074/jbc.M307552200

Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G., & Deber, C. M. (2009). Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *PNAS*, *106*(6), 1760–1765. www.pnas.org/cgi/doi/10.1073/pnas.0813167106

Ross, J., Li, Y., Lim, E.-K., & Bowles, D. J. (2001). Protein family review: Higher plant glycosyltransferases. *Genome Biology*, *2*(2), 1–6. http://genomebiology.com/2001/2/2/reviews/3004.1

Sarr, P. S., Ando, Y., Nakamura, S., Deshpande, S., & Subbarao, G. V. (2020). Sorgoleone release from sorghum roots shapes the composition of nitrifying populations, total bacteria, and archaea and determines the level of nitrification. *Biology and Fertility of Soils*, *56*(2), 145–166. https://doi.org/10.1007/s00374-019-01405-3

Smil, V. (2001). *Enriching the Earth: Fritz Haber, Carl Bosch, and the Transformation of World Food Production* (1st ed.). MIT Press.

Tian, H., Xu, R., Canadell, J. G., Thompson, R. L., Winiwarter, W., Suntharalingam, P., Davidson, E. A., Ciais, P., Jackson, R. B., Janssens-Maenhout, G., Prather, M. J., Regnier, P., Pan, N., Pan, S., Peters, G. P., Shi, H., Tubiello, F. N., Zaehle, S., Zhou, F., … Yao, Y. (2020). A comprehensive quantification of global nitrous oxide sources and sinks. *Nature*, *586*(7828), 248–256. https://doi.org/10.1038/s41586-020-2780-0

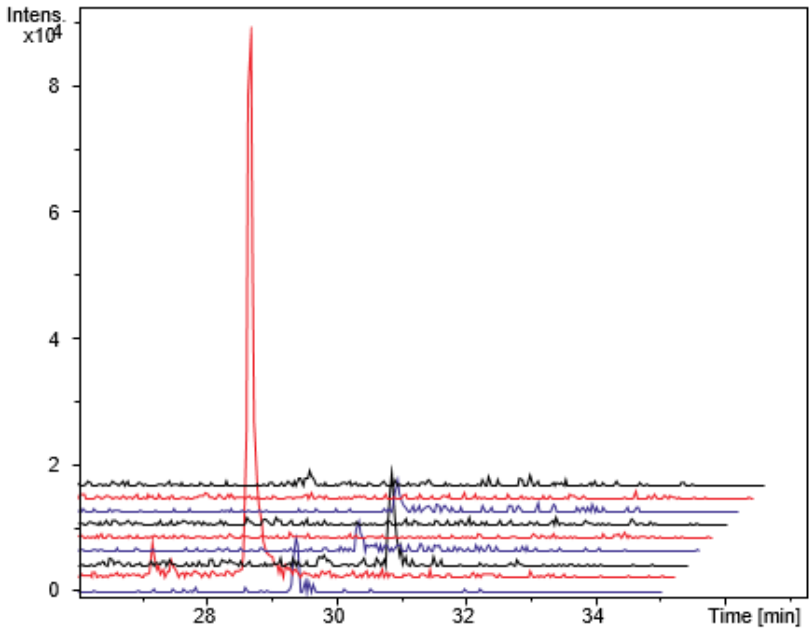
Tibugari, H., Chiduza, C., Mashingaidze, A. B., & Mabasa, S. (2020). High sorgoleone autotoxicity in sorghum (Sorghum bicolor (L.) Moench) varieties that produce high sorgoleone content. *South African Journal of Plant and Soil*, *37*(2), 160–167. https://doi.org/10.1080/02571862.2020.1711539

Wen, C., Huang, W., He, M. M., Deng, W. L., & Yu, H. H. (2020). Cloning and characterization of a glycosyltransferase from Catharanthus roseus for glycosylation of cardiotonic steroids and phenolic compounds. *Biotechnology Letters*, *42*(1), 135–142. https://doi.org/10.1007/s10529-019-02756-5

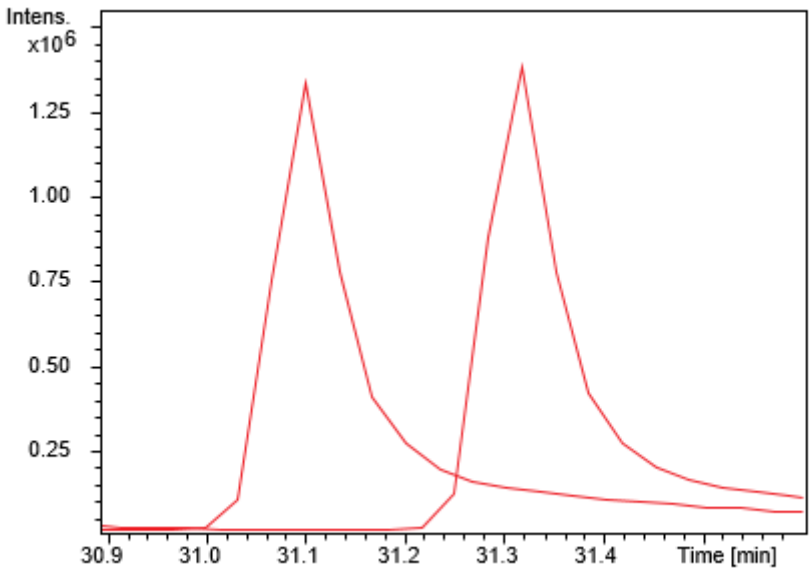
Woodward, E. E., Edwards, T. M., Givens, C. E., Kolpin, D. W., & Hladik, M. L. (2021). Widespread Use of the Nitrification Inhibitor Nitrapyrin: Assessing Benefits and Costs to Agriculture, Ecosystems, and Environmental Health. *Environmental Science and Technology*, *55*(3), 1345–1353. https://doi.org/10.1021/acs.est.0c05732

Wu, Y., Yang, Y., Du, L., Zhuang, Y., & Liu, T. (2023). Identification of a highly promiscuous glucosyltransferase from Penstemon barbatus for natural product glycodiversification. *Organic and Biomolecular Chemistry*, *21*(21), 4445–4454. https://doi.org/10.1039/d3ob00370a

7. Appendix



Appendix 1: Extracted ion chromatograms of the sodium adduct of sorgoleone (red), dihydrosorgoleone/sorgoleone-360 (dark blue), and sorgoleone-362 (black) (±0.005). The front three chromatograms are the same seen in figure 8. The middle three are from assay samples with UGT72B27, and are representative of all assay samples. The back three are from the negative control assay sample (empty plasmid).



Appendix 2: Base peak chromatograms from assay samples with UGT72B27 (right) and UGT74AN3 (left). This represents an approximate shift of 15 seconds present for all equal peaks in the chromatograms.