Investigating the heterologous expression of plant secondary metabolic enzymes to produce curcuminoids and curcuminoid derivatives

Keir Sam Bailey

PhD

University of York

Biology

August 2016

Abstract

Curcuminoids are highly bioactive polyketide chemicals produced in the rhizomes of the turmeric plant (*Curcuma longa*). In addition to giving turmeric powder its distinctive yellow colour and flavour, these molecules elicit a plethora of medicinal effects. This project aimed to biosynthesize curcuminoids and their derivatives using the metabolic engineering of Baker's yeast (*Saccharomyces cervisiae*) and tobacco (*Nicotiana benthamiana*). Phenylpropanoids, sustainably derived from bio-refinery waste, were to be used as starting materials. To convert phenylpropanoids into curcuminoids three enzymatic steps were needed: this involved the cloning and expression of 4-coumarate CoA ligase 5 (4CL5) from *Arabidopsis thaliana* and diketide-CoA synthase (DCS) and curcuminoid synthase 1 (CURS1) from *C. longa*. Protein extract from yeast expressing 4CL5 was able to convert both natural and non-natural phenylpropanoids to their corresponding phenylpropanoyl-CoA esters. However, protein extracts from yeast transformed with either DCS or CURS1 did not yield the corresponding enzyme activity. Furthermore, it was concluded that tobacco was not an optimal chassis for curcuminoid production as both phenylpropanoid substrates and curcuminoid products were metabolised by endogenous enzymes in the leaf tissue.

Despite their bioactivities, curcuminoids are poorly bioavailable reducing their medicinal potential. Regiospecific glycosylation, performed by UDP-glycosyltransferases (UGTs), is a known strategy to improve a molecule's solubility. Diglucosylated curcuminoids are 10 000 times more soluble in water than the aglycone. Using metabolomics and ^{14}C radiolabelling assays *C. longa* was investigated as a source of novel curcuminoid UGTs. No evidence of curcuminoid glycosylation or other bio- conjugation was found and it was concluded that, like the ginger plant, *C. longa* is likely to sequester curcuminoids within sesquiterpene oil droplets. Instead, CaUGT2, a UGT from *Catharanthus roseus* known to glycosylate curcuminoids, was heterologously expressed in *S. cerevisiae*. Previous work with CaUGT2 by Kaminaga *et al.*[1], gave only a 4 % yield of curcumin monoglucoside in *E.coli*. In this project, when yeast culture expressing CaUGT2 was fed curcumin, both mono- and diglucosides were produced. The bioreaction was successfully scaled up 200 times from a shake flask to a 5 L bioreactor and was estimated to produce 32 mg L^{-1} curcumin monoglucoside (60 % yield). To date, this is the best performing biosynthetic production of curcumin glucosides and with optimisation is expected to surpass chemical methods.

Contents

List of Figures

List of Tables

Acknowledgements

If only I could write a thesis on how grateful I am to everyone who's helped me throughout my PhD.

Firstly, thank you to my Supervisor, Ian Graham for giving me the chance to work in his lab. You are such an inspirational person and scientist and it's been great to work with you and have your help, support and guidance. Secondly, thank you to everyone in CNAP, specifically Ian's lab, for teaching me so much and their patience. Special thanks to Fabian for helping me to be more structured and thorough- basically more scientific. Special thanks also to Andy, Ali, Tony and Swen for their help in the lab, as well as Judith keeping me on track out of the lab. Huge thanks to Nia Bryant, Bob White and Aga Urbanek for stirring up the scientific excitement. Neil Bruce and Gideon Grogan, I cannot say how much I appreciated their constructive support and pastoral advice. Thirdly, thank you to the whole of the York Biology Department. I am lucky to have worked in such a welcoming and inspiring place: from Julie in the Post Grad office, to Jared in the TF, thank you. I could not have asked to work in a better research facility surrounded by such kind, intelligent people. I must also thank the BBSRC IBTI club for funding my work and providing such fun dissemination meetings. Thank you to my old lab who are now at Newcastle University: to Rob, for creating such an exciting project, and for Melissa and Fede for showing me the ropes.

Extra special thanks to my SPSs Emma, Rebecca and Rhi. Without you guys to share lunch, shots and Sharknado my PhD would not have been the same. I love you all so much. As excited as I am about what comes next, I will miss you all terribly.

Extra, extra special thanks to my mum and dad, for believing in me, talking sense and putting things into perspective. You're the absolute best. And to Kim, my amazing boyfriend. I couldn't have made it without you. Thank you for your understanding and pulling me out of my lows, and for sharing with me the most amazing, exciting highs. I can't wait for the next challenge with you.

"I need some kinda voodoo
I need some kinda luck
The shape shifting witch's brew
Is what I'll drink with you
Luck from a rabbit's foot
Killed by a cross-eyed look
I found a little, a little good luck"

Turbowolf, Two Hands

Declaration

The work presented in this thesis is the sole effort of the author, except where explicitly stated. Reference to the work of others and collaboration has been duly acknowledged. No portion of this work has been submitted for any other degree.

Chapter 1. Introduction

For thousands of years, humans across the globe have relied upon plants and the diverse range of chemicals they produce as part of everyday life. Not only do plants form an intrinsic part of our diet, clothing and building materials, but they can also act as remedies to help cure ailments when applied as a poultice or sipped as an infusion. In the first century AD modern pharmacology began; Dioscorides, a Greek physician and botanist wrote the first pharmacopeia, *De Materia Medica*, a vast encyclopaedia of the medicinal qualities of plants he encountered whilst travelling with the Roman army. However, there is evidence that plants have been used medicinally for much longer, with records indicating that Neanderthals, 60 000 years ago, used plants for their therapeutic properties[2], derivatives of which are still used today. According to the World Health Organisation, 80 % of the world's population rely upon medicines derived from plant sources [3] and approximately a quarter of all pharmaceuticals used today are still sourced from plants [4].

The medicinal properties of plants are mainly due to their secondary metabolite constituents. Kossel was the first to define these metabolites, over a century ago, as being distinct from the function of primary metabolites involved in basic life functions such as respiration, cell growth and reproduction. Secondary metabolites, or natural products, are often less than 1 % of the total carbon and function to allow plants to adapt to their environment, having roles in pathogen protection, anti-feedants, UV protection and pollinator attraction [5]. They are classified into different types depending on the chemical structures, related starting materials and functional groups; these include polyphenolics, alkaloids and terpenoids [5]. Humans have found many uses for this diverse range of bioactive chemicals, from common pain relief, such as aspirin to licit and illicit opioid drugs.

For mankind to make use of the plethora of plant secondary metabolites, they need to be readily available; this is often limited by the small quantities synthesized and extractable from the plant. Furthermore, some plants originating from certain ecosystems do not readily grow in the environments where they are needed. Therefore, using retrosynthetic chemistry, organic chemists try to decipher which sequence of chemical reactions are necessary to produce such intricate and potent molecules. One of the most notable examples of retrosynthesis of a plant natural product was the total synthesis of taxol, an anticancer agent from Pacific yew (*Taxus brevifolia*), in which two research groups competed to be the first to complete the total synthesis [6,7]. Although many total syntheses like this have been successful in producing plant natural products, they are often protracted, economically impractical and low yielding. An attractive alternative is to utilise the plant system itself, creating elite cultivars or culturing plant cells. Another alternative is to use

metabolic engineering to reconstruct the plant biosynthetic pathway in a chassis organism with the objective to develop a bio-based production system that is amenable to product extraction. This thesis will focus upon the bioactivity, biosynthesis and modification of plant chemicals, in particular, curcuminoids.

1.1. Biosynthesis of plant natural products

Although organic chemistry offers a wide range of chemical conversions, and a vast array of total syntheses exist for many complex plant natural products, the yields are often low. Additionally, chemical syntheses often rely on unsustainable starting materials from petroleum and expensive, toxic catalysts. In comparison, enzymes can frequently use naturally derived substrates from waste or sustainable sources. Utilising enzymes in biosynthetic reactions also allows milder reaction conditions and produces fewer side products, meaning the target compound is easier to extract and purify. Enzymatic syntheses also allow the reduction of catalytic steps as enzymes often target one specific functional group, avoiding the need for protection and deprotection. However, enzymes can sometimes produce a mixture of isomers and epimers which compromises biological activity leading to a decrease in interest from pharmaceutical companies [8].

Both endogenous and recombinant enzymes can be utilised through metabolic engineering to produce the desired target molecule. In 1991, James Bailey defined metabolic engineering as 'the direct improvement of production, formation, or cellular properties through the modification of specific biochemical reactions or the introduction of new ones with the use of recombinant DNA technology' [9]. The Synthetic Biology Roadmap for the UK defines synthetic biology as 'the design and engineering of biologically based parts, novel devices and systems as well as the redesign of existing, natural biological systems' [10]. There are multiple biosynthetic platforms which are amenable to synthetic biology and can be metabolically engineered for the production of natural products, including *in vivo* and *in vitro* biosynthesis in micro-organisms, plant cell culture and plant biotransformations. To date, substantial efforts have been made to produce multiple classes of natural products, their intermediates and novel analogues such as terpenes, flavonoids, and alkaloids in a range of host organisms [11]. This project will focus upon the biosynthesis of pharmaceutical plant natural products using microbial and plant platforms.

1.1.1 Microbial biosynthesis of plant natural products

Compared to field grown plants or plant cell culture, microorganisms can be grown quickly and are easily genetically manipulated. Microorganisms can be used in three ways to produce or modify plant natural products. Firstly, biotransformations can be carried out using endogenous enzymes. Secondly, using genetic engineering, heterologous enzymes can be expressed in the

microorganism, allowing previously impossible biosynthetic reactions. Thirdly, the microorganism can be used to produce enzymes (endogenous or heterologous) which can be extracted on scale and utilised in *in vitro* bioconversions. These three approaches will be considered in the context of *in vivo* and *in vitro* reactions, with emphasis upon Baker's yeast (*Saccharomyces cerevisiae*).

1.1.1.1 *In vivo* biosynthesis of natural products

If performing *in vivo* biosynthesis, it is important to consider the host organism and how the substrates and products interact with it [12]. Is the substrate taken up by the cell, is it degraded or utilised by an endogenous pathway, what by-products might be formed, is the product sequestered? Whole-cell biocatalysis contains two main categories, fermentation and biotransformation, depending on whether the starting material was part of native or non-native metabolism [13]. In the early 1990s, the first industrial scale whole cell biocatalyses relied upon native enzymes to convert non-native substrates, such as the conversion of methyl substituted heteroarenes to the corresponding carboxylic acids by *Pseudomonas putida* [14]. Another successful strategy relied on knocking-out certain native genes, resulting in accumulation of endogenous pathway intermediates such as the synthesis of long chain dicarboxylic acids from fatty acids in *Candida* yeast strains [15]. Although successful, the scope of utilising native pathways is limited by the discovery and natural abundance of industrially relevant reactions.

An increase in the ease of DNA manipulation and the popularity of systems and synthetic biology has seen a growth in *de novo* heterologous biotransformations. As genes from different organisms can be combined, it means the products are not limited by natural pathways and therefore allows a broader range of substrates [16]. It also means retrobiosynthetic molecule design can be applied and longer biosynthetic pathways have been able to be explored, including the synthesis of plant natural products [17]. There are a range of microbial hosts in which plant biosynthetic pathways have been heterologously expressed, however research has mainly focussed around *S. cerevisiae* and *Escherichia coli* due to their well-known physiology, biochemistry and ease of use in the laboratory. *E. coli* has a relatively simple metabolism, high genetic tractability and grows quickly. Whilst similar to *E. coli*, *S. cerevisiae* is eukaryotic offering larger cells with intracellular organelles, higher tolerance to pH fluctuations and the ability to perform post-translational modifications. Replicating complex biosynthetic pathways requires considered cloning approaches, with promoter and terminator strength and gene expression being finely tuned. Using typical cloning and transformation methods, can often prove time consuming and unproductive, therefore significant efforts are being made to create automated, one-pot DNA construction methods utilising homologous recombination in *S. cerevisiae* [18,19]. Coupling computational and systems biology to metabolic engineering has allowed the production of improved strains which

can efficiently channel carbon flux towards the desired product. For example, the engineering of gene deletions and network modifications in *E. coli* to increase the intracellular concentration of the intermediate malonyl-coenzyme A (CoA), increased recombinant naringenin production by 660 % [20]. Although there are many strategies to increase production of phytochemicals in these recombinant microorganisms, very few have resulted in commercial processes due to low yields.

Biosynthesis of natural products using *S. cerevisiae*

Baker's yeast has been used for centuries in traditional fermentation processes such as making bread, beer and wine and is now the organism of choice for synthetic biology and industry as a metabolically engineered 'cell factory'. In addition to *S. cerevisiae*'s robustness to harsh conditions its GRAS (generally regarded as safe) status awarded by the FDA (U.S. Food and Drug Administration) has made it preferential to *E. coli* in the industrial fermentation arena, despite generally giving lower yields [21]. This is why yeast is so often chosen for metabolic engineering; from bioethanol production [22], to synthesising secondary metabolites of greater purity than can be isolated naturally, such as vanillin and resveratrol [23,24]. Nielsen and Hong [25] evaluate the potential of *S. cerevisiae* as a synthetic biology chassis by reviewing its extensive use in industry and the numerous patents that have been recently filed for the production of biofuels, proteinaceous drugs and bulk and fine chemicals.

The most notable example of heterologous production of a plant biosynthetic pathway in a microorganism is that of the sesquiterpene lactone endoperoxide artemisinin. Artemisinin is produced by the *Artemisia annua* L. plant and exhibits excellent antimalarial activity towards the multi-drug-resistant parasite *Plasmodium falciparum* [26]. As part of a large ongoing project by Keasling and co-workers they first used *E. coli* to produce artemisinin, then moved to Baker's yeast as it could tolerate higher product concentrations [27,28]. This required the recombinant expression of thirteen genes to allow the conversion of acetyl-CoA to dihydroartemisinic acid [29] or artemisinic acid [30], but required a final photochemical reaction step to yield artemisinin [31]. At the time, artemisinin itself could not be biosynthesized due to the last enzymatic step being unidentified in *A. annua*. Therefore, this process is a good example of a 'hybrid' strategy in which biological limitations can be overcome by chemistry. To increase yields, the mevalonate pathway was up-regulated in yeast and FPP consumption was down-regulated and the heterologous *A. annua* pathway was coupled to the native terpenoid metabolism via farnesyl pyrophosphate. After optimisation, the process could yield 25 g L^{-1} artemesinic acid [27]. This was a fantastic biotechnological feat, hailed as a synthetic biology triumph, showcasing the potential of natural product biosynthesis [28]. In combination with Amyris Biotechnologies the pharmaceutical company Sanofi began industrial production of artemisinin theoretically allowing the annual production of 50– 60 tons of artemisinin per year, a third of the global demand. However, it is one hurdle to produce a proof-of-principle biosynthetic pathway, it is another to be able to make

the procedure economically viable and competitive. The cost of the photochemical reaction needed to produce artemisinin resulted in the costs Sanofi's semi-synthetic artemisinin being much higher than that of the naturally extracted product, $350– 400 per kilogram compared to $250 per kilogram [32]. Therefore, in 2015, Sanofi did not produce any artemisinin and the manufacturing plant is now closing [32]. Nonetheless, being the first effort of its kind, there is still much to learn and develop, and metabolic engineering and synthetic biology production of fine chemicals is still high upon academic and industrial interests.

The success of production of the artemisinin precursor artemisinic acid at such high concentrations in yeast, spurred others on to recreate large secondary metabolic pathways in yeast. One recent example is the production of natural and semisynthetic opioids, by Smolke's laboratory [33,34]. By expressing genes from *Papaver somniferum* and *Pseudomonas putida* M10, and optimising co-substrate availability and gene copy number, they were able to produce opioid titres of up to 131 mg L^{-1}. Nonetheless, *P. somniferum* is commercially grown around the world for approximately $250 per kilogram [35], meaning much work is needed before the biosynthesis of morphine can rival extraction from the plant. It is important to remember that, although such proof-of-concept studies have great potential, there remain challenges involving optimisation and scale-up before the biosynthesis of secondary metabolic chemicals become a viable bioeconomy. It is therefore worth noting, that whilst the above are excellent examples, microbial biosynthetic efforts would be more beneficial if applied to secondary metabolites that are recalcitrant to extraction from commercially grown plants.

One way in which metabolic engineering is being used to improve biosyntheses, is the extensive modification of endogenous metabolism to allow an increase in endogenous intermediate availability and therefore an increase in final product yield [25]. Nielsen [36] and others [37,38] have applied rational engineering to create yeast strains capable of producing high levels of important precursors such as acetyl-CoA and malonyl-CoA. By adopting push-pull-block systems, in which upstream endogenous genes are over expressed, heterologous genes are over expressed and downstream genes are knocked out they managed to increase the intracellular pools of these precursors allowing greater substrate availability for heterologous enzymes. These high producing precursor strains have been shown to be valuable in the production of terpenoids, such as α-santalene, an important constituent of sandalwood oil, and fatty acid ethyl esters increasing yields four and three times, respectively [37,38].

1.1.1.2 *In vitro* biotransformations of natural products

Enzymes, heterologous or endogenous, can be extracted from a host and used in *in vitro* biocatalysis or transformations. In comparison to *in vivo* catalysis, this allows the isolation of

metabolic pathways and use of non-conventional substrates which may be toxic to whole cells [17]. It also reduces competition with endogenous reactions, incidence of side reactions and production of derailment products. Furthermore, enzyme subcellular localisation and transport issues of the substrate and product are negated. However, *in vitro* systems often require the production and purification of large quantities of stable and active enzyme, in addition to the need for expensive cofactors such as NADPH or CoA.

As *in vitro* systems are often a single reaction step, the enzymes that lend themselves to this process are often tailoring enzymes which decorate the core molecular structure enhancing the complexity. They include transferases, halogenases, oxidoreductases and cyclases. Glycosyltransferases are common tailoring enzymes used for glycodiversification. An interesting example of *in vitro* chemical biosynthesis and the use of glycodiversification is exhibited by the one-pot synthesis of thymidine diphosphate-D-ravidosamine, a potent antitumour compound, using five enzymes [39].

In vitro biosynthesis offers the opportunity to combine enzymatic reactions with chemical synthesis, known as cooperative catalysis. It has proven to be especially useful for the production of chiral chemicals where an enantioselective enzyme can be used in combination with a chemical reagent. Although cooperative catalysis reaction development can be difficult due to mutual deactivation of the catalyst and enzyme, there are examples of this being overcome. Using a biphasic solvent system of dioctyl phthalate and buffer, a Ru(I) cross-metathesis catalyst and a cytochrome P450 enzyme from *Bacillus megaterium* allowed the selective epoxidation of cross-metathesis products at a greater yield than the two separate reactions [40].

As is apparent, *in vitro* biosynthesis offers a great opportunity for powerfully combining chemical syntheses with biological reactions. It is expected that as strategies improve and knowledge is gained about their synergistic effects, cooperative catalysis and *in vitro* biosynthesis could be powerful tools for biotransformations, allowing diversification of natural products to produce novel analogues.

1.1.2 Natural product biosynthesis using plant platforms

Plants as biosynthetic chassis have one major advantage over other systems, in that they can use atmospheric CO_2 and sunlight as starting materials. Therefore, plant metabolic engineering is being exploited to a range of ends from CO_2 fixation to protein production, as well as chemical biosynthesis. Compared to microbial chasses, plant platforms offer multiple tissue types and intracellular organelles. This can be advantageous, but can also complicate the localisation of heterologous protein expression and complicate metabolite analysis [41]. This highlights the importance of considering transporters and membrane channels in plant metabolic engineering.

Compared to microbial metabolic engineering, plant metabolic engineering is still in its infancy. Utilising plants to produce natural products and important biochemical intermediates is being studied on two fronts: plant cell culture and whole plant systems.

Plant cell culture in some aspects is analogous to producing chemicals in eukaryotic microbial hosts. Both involve cell suspensions within a bioreactor and both can exploit endogenous biosynthetic pathways or support heterologous enzymes. Plant cell cultures are generated from the dedifferentiation of actively dividing root, shoot or embryo tissue to generate mitotically reactivated cells [42]. Plant cell cultures that rely on endogenous biosynthesis require the whole pathway to be expressed in the cultured cells, for example, the production of pharmaceuticals by *Catharanthus roseus* cell suspension [43]. Although almost two decades ago, the greatest successes have been the production of rosmarinic acid and taxol using *Rosmarinus oficinalis* and *Taxus media* cells, respectively. The cell culture of *Rosmarinus oficinalis* was able to produce approximately 5 times more rosmarinic acid than the whole intact plant [44]. *Taxus media* cells were able to produce 0.6 % (dry weight) taxol, when biosynthesis was stimulated by the addition of methyl jasmonate [45]. In fact, the largest producer of taxol, Phyton Biotech, uses a 75 000 L plant cell culture [46]. However, since then, plant cell culture production platforms have mainly focussed around the production of pharmaceutical proteins, such as vaccines, instead of secondary metabolites [47]. This may be due to the fact that secondary metabolite synthesis is typically considered to take place when the culture enters the static growth phase. Additionally, as plant cell cultures do not always maintain photoauxotrophic growth, productivity has been limited [48]. A recent breakthrough by Loake *et al*. [49,50] involved the isolation of cambial meristematic cells from the cambium tissue of plant secondary meristems. These innately undifferentiated cells performed better on all fronts compared to typical dedifferentiated plant cells. They aggregated less, showed lowering sheering damage and were able to produce secondary metabolites at a much higher yield. Undifferentiated *Taxus cuspidata* cambial meristematic cells in a 20 L bioreactor were shown to produce 268 mg kg^{-1} taxol compared to dedifferentiated *T. cuspidata* embryonic cells which didn't produce any [49]. Therefore, the use of cambial meristematic cells offers much potential to the field of plant cell culture biosyntheses.

Some desired secondary metabolic pathways are not expressed in plant cell culture, requiring the biosynthetic pathway to be introduced by bacterial mediated transformation involving *Agrobacterium*. The transformation can be transient or stable, lasting several days until the cells senesce or indefinitely being passed on to descendants [51]. *Agrobacterium rhizogenes* is used for creating transgenic hairy root culture [52], whereas infection with *Agrobacterium tumefaciens* allows the generation of shoot culture [48]. Unlike plant cell culture, shoot and hairy root cultures are able to stably produce secondary metabolites concomitant with growth [53]. This was demonstrated by Mendoza *et al*. [54], in which more than 500 hairy root lines from *Datura*

stramonium produced stable amounts of tropane alkaloids over 5 years. An advantage over plant cell suspension culture is that organ cultures are generally less sensitive to shearing forces, and can be made even more so by immobilising to matrix [55]. To date, there is only one company, ROOTec using hairy root cultures on an industrial scale [11].

Whole plants offer a simpler production process compared to plant cell cultures, with significantly lower production costs due to existing horticulture expertise and crop management facilities. Although the most obvious approach, host plants take time to grow and are often underdeveloped as crops. The production of natural products in whole plants currently takes two main forms: breeding to increase yields or the introduction of heterologous pathways. There has been much success in breeding improved varieties of plants with increased yields of secondary metabolites. Again, a notable example is that of artemisinin production. By performing in-depth transcriptome sequencing of *A. annua*, Graham *et al*. [56], were able to identify key genes and markers for the fast-track breeding of high yielding lines. This allowed the creation of a hybrid (Hyb8001r) which is able to produce a maximum artemisinin yield of 21.8 kg per acre, significantly higher than the previous elite variety and much higher than Keasling's yeast [57]. Not only does studying the crop in such detail enable natural increases in yield, but also increases the understanding of biosynthetic pathways necessary for heterologous biosynthesis.

Popular examples of heterologous biochemical synthesis in whole plants are in the production of natural products important in food nutrition. In 2000, Ye *et al*. [58], created 'Golden rice', an engineered rice variety using stable *Agrobacterium* mediated transformation, to produce provitamin A in the rice endosperm to help address the problem of vitamin A deficiency in Asia. The yield was increased 23 times further, to 37 µg/g, by investigating and replacing the rate limiting phytoene synthase enzyme [59]. Another example is that of the purple tomato, in which tomatoes were engineered by Martin *et al*. [60], to accumulate health-promoting anthocyanins. Interestingly, in contrast to the majority of examples outlined thus far, instead of introducing heterologous biosynthetic enzymes, Martin *et al*. [60], instead introduced transcription factors, which in turn caused the upregulation of anthocyanin biosynthetic enzymes. Although these examples of heterologous secondary metabolite production are for direct consumption and not extraction, they illustrate the successful production of secondary metabolites in plants.

Interestingly, a heterologous plant system offers the most recent advance in the battle for artemisinin production. By developing a new synthetic biology approach, which they called COSTREL (combinatorial supertransformation of transplastomic recipient lines), Fuentes, *et al*. [61], were able to transfer the *A. annua* genes necessary for artemisinic acid production into the chloroplast of tobacco (*Nicotiana tabacum*). Tobacco is a high-biomass crop, having multiple harvests per year, which can be grown easily on a large scale. In tandem they introduced

'accessory' genes into the nuclear genome allowing for improved regulation of the heterologous pathway. They were able to produce tobacco which could synthesize 4.8 kg artemisinic acid per acre, which is impressive for bioengineering, but it is still approximately 5 times lower than that of Graham *et al*.'s *A. annua* hybrid Hyb8001r [56]. Furthermore, the work highlighted how little is still known about plant synthetic biology, exemplified by the nuclear 'accessory' genes which although aided yield, functioned in a 'largely unknown manner'. Nonetheless, the approach of Fuentes, *et al*. [61], exemplifies how developments in methods and techniques allows progression in synthetic biology

When comparing whole plant, plant cell culture and micro-organism platforms for the production of natural products there is are many advantages and disadvantages to be considered. Depending on the nature of the desired product, how it is synthesized and its market demand dictates the efficiency of the synthesis and the economics of its production. However, at a time in which DNA synthesis and assembly is rapidly become automated, it is likely that we are only at the very beginning of an exciting age of natural product microbial biosynthesis.

1.2. Curcuminoid chemicals: activity, synthesis and bioavailability

Ayurveda, from the Indian words *'Ayar'* (life) and *'veda'* (science) meaning the Science of Life, is thought to be the oldest traditional medicine, predating even Chinese traditional medicine. One of a plethora of Ayurvedic medicinal plants is the turmeric plant (*Curcuma longa* L.). Rhizomes of the *Curcuma* species have been used for centuries due to their medicinal properties, as well as to add colour to fabrics and flavour food. The ground up rhizomes gives rise to turmeric powder, which is now commonly used by industry as a flavouring, preservative, and colouring agent. *Curcuma* is the third largest genus within the Zingiberaceae family, containing around 120 species [62]. Of these, the type species *C. longa* L. is the best known and the main source of turmeric powder [63]. *C. longa* is a perennial monocotyledonous plant which grows to around a metre in height and is distributed throughout South and Southeast Asia, with the most diversity in India and Indochina.

The potency of the *C. longa* rhizomes as a medicine has been attributed to the synergistic effects of its many secondary metabolites [64]. However, one of the major bioactive constituents of the turmeric plant rhizomes are the curcuminoid chemicals. Curcuminoids are bis-α,β-unsaturated diketones which are synthesized from phenylalanine or tyrosine *via* the phenylpropanoid and polyketide pathways in the rhizomes of *C. longa* [65,66]. There are three main curcuminoids: curcumin, demethoxycurcumin (DMC) and bisdemethoxycurcumin (BDMC) (Figure 1) which make up 3-5 % of *Curcuma* rhizomes, with *Curcuma zedoaria* having the highest curcuminoid

content. The major curcuminoid is curcumin (~ 77 %), followed by DMC (~ 18%), then BDMC (~ 5 %) [67]. Curcuminoids are highly conjugated molecules, which cause the characteristic yellow colour of the turmeric rhizomes. Under physiological conditions, curcumin exists in equilibrium between two tautomers, either the keto- or the enol- form (Figure 2); in acidic conditions, the keto- form predominates, whereas the enol- form predominates in alkaline conditions.

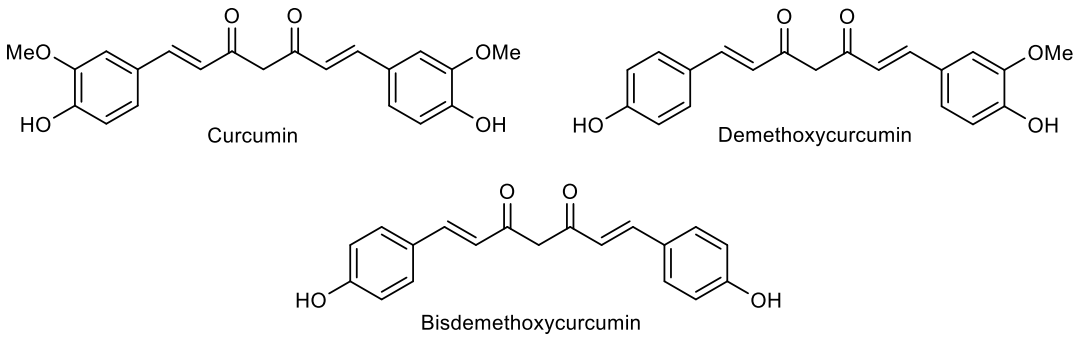


Figure 1. Curcuminoids: curcumin, demethoxycurcumin (DMC) and bisdemethoxycurcumin (BDMC).

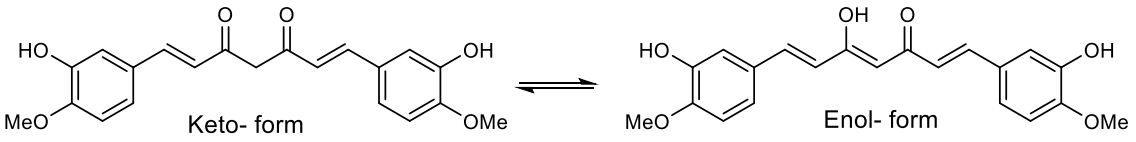


Figure 2. Keto-enol tautomerisation of curumin.

1.2.1 The health benefits and bioactivity of curcuminoids

Curcuminoids exhibit a plethora of health benefits, with thousands of publications claiming their efficacy against a vast range of diseases. In 1949, curcumin was first shown to act as an antibacterial agent towards *Staphylococcus aureus* [68]. Curcumin is also well known for its antioxidant properties which was reported in 1975 by Sharma [69]. Not only can curcumin scavenge free radicals [70], it can reduce the production of reactive oxygen species *in vivo* [71]. Curcumin's antioxidant ability has been shown to reduce oxidised proteins in Alzheimers amyloid pathology in rat models [72]. Cardiovascular disease, including atherosclerosis and myocardial infarction has also been shown to be attenuated by diets supplemented with curcumin [73]. Currently, the largest interest in curcuminoid bioactivity is their chemotherapeutic and chemopreventative effects towards a range of different cancers. Signal transducer and activator of transcription 3 (STAT3) and nuclear factor (NF)-κB are both active components in the signalling of cancer cell development. Curcumin, and chemically synthesized analogues, have been shown to down-regulate both of these signalling pathways [74,75]. For an extensive review of the biological targets and effects of curcumin(oids) see Prasad *et al*. [76].

As well as evaluating the pharmacological effect of natural curcuminoids, many analogues have also been tested for their efficacy against different diseases [77]. This has allowed conclusions to be made about structure-activity relationships. The bioactivity of curcuminoids is believed to be due to the molecular ability to act as a Michael acceptor at the enone moiety, meaning they can react with soft nucleophiles such as thiols (e.g. glutathione) [78]. There are differing bioactivities between the three main curcuminoids curcumin, DMC and BDMC with curcumin exhibiting the greatest potency suggesting the importance of the methoxy groups on the phenyl ring [79]. The positioning of the ring substituents is also significant: molecules with an electron donating group *ortho* to the *para* hydroxyl group are the most efficacious [80], with the *para* hydroxyl being important for antioxidant activities [81].

1.2.2 Curcuminoid biosynthesis

1.2.2.1 Polyketide Synthases

Polyketide synthases (PKSs) catalyse the decarboxylative condensation of carboxylic acids to produce polyketides, a class of secondary metabolite. According to their catalytic mechanism, structure and primary sequence, polyketide synthases are subdivided into classes: I, II and III. Similar to fatty acid synthases, Type I PKSs constitute a large assembly of catalytic polypeptides, including essential ketosynthases, acyl transferases, and acyl carrier proteins, as well as non-essential domains such as ketoreductases, dehydratases and methyltransferases. The presence or absence of these domains determines the chemical structure of the polyketide that is synthesized.

In bacteria, the domains are normally used once, known as non-iterative polyketide synthesis, whereas, in fungi domains are repeated, giving iterative polyketide synthesis. Type II PKSs function to synthesize aromatic polyketide and are large multi-enzyme complexes, each enzyme having a distinct function [82]. Type III PKSs are homodimers, of which the active site in each monomer is responsible for iterative priming, extension, and sometimes cyclisation to form polyketide products [82].

Plant type III PKSs share 21–31 % amino acid sequence identity with bacterial type III PKSs and 35- 90 % identity with one another [83]. The homodimer is 40- 45 kDa in size and catalyses iterative decarboxylation condensations of malonyl-Coenzyme A (malonyl-CoA) with CoA activated substrates from the phenylpropanoid pathway. The linear chain formed is often cyclised to form products including chalcones and stilbenes [83]. Type III PKSs share a common three-dimensional structure and utilise a catalytic triad of cysteine, histidine and asparagine within the active site; the shape and volume of which determines the enzymes selectivity [83]. The cysteine residue is the catalytic residue functioning as a nucleophile to cleave the C–S CoA thioester bond and initiate the addition of a two carbon acetyl unit via C–C bond formation (Figure 3). The histidine and asparagine form an oxyanion hole to stabilise the tetrahedral transition state of nucleophilic attack. Due to the orientation in which chain elongation occurs, type III PKS condensation is commonly referred to act via a head-to-tail model. The CoA activated substrates range from acetyl-CoA to aliphatic and aromatic CoA-esters, with the number of malonyl-CoA condensations ranging from one to eight, although it is usually three. If cyclisation occurs it can be carried out in three ways: Claisen condensation-type cyclisation, which is performed by acridone synthase (ACS) [84]; aldol-type cyclisation, utilised by stilbene synthase (STS) [85] or lactonisation as carried out by 2-pyrone synthase (2PS) [86] (Figure 4). Other type III PKSs, such as benzalacetone synthase (BAS) does not perform cyclisation (Figure 4).

The most well characterised family of type III PKSs in plants are the ubiquitous chalcone synthase (CHS) superfamily, which catalyse the first step of flavonoid biosynthesis. CHS catalyses the condensation of *p*-coumaroyl-CoA with three malonyl-CoA extender units to form the aromatic triketide naringenin (see Figure 4). Once the naringenin flavoinoid scaffold is assembled, the skeleton is diversified by tailoring enzymes which perform acylation, hydroxylation, prenylation, O-methylation and glycosylation yielding over 6 000 different flavonoids [83].

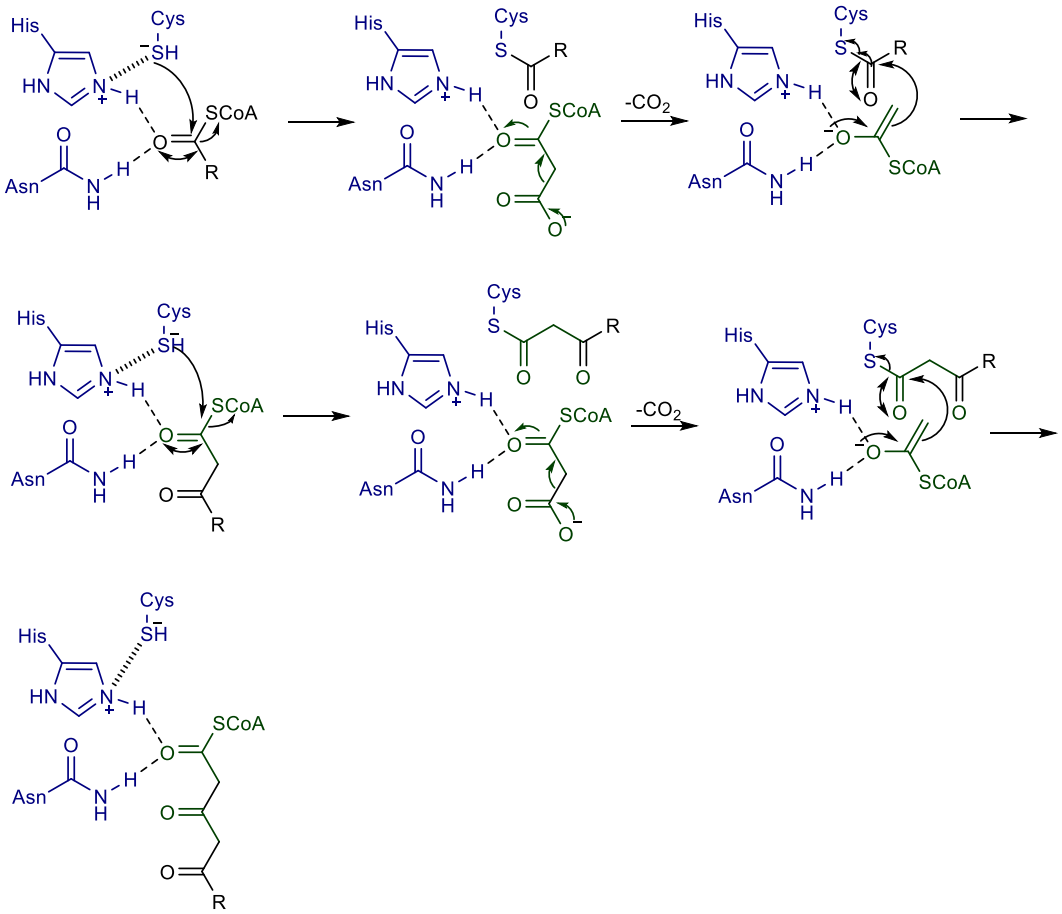


Figure 3. Mechanism for type III PKS chain elongation.
Substrate loading is followed by malonyl-CoA chain extension and decarboxylation forming the diketide intermediate, further extension by malonyl-CoA and decarboxylation forms of the triketide product. Shown in blue are the catalytic triad, cysteine, histidine and asparagine in the active site; shown in green is the malonyl-CoA extender and shown in black is the CoA-esterified substrate. Adapted from Abe *et al*. [83].

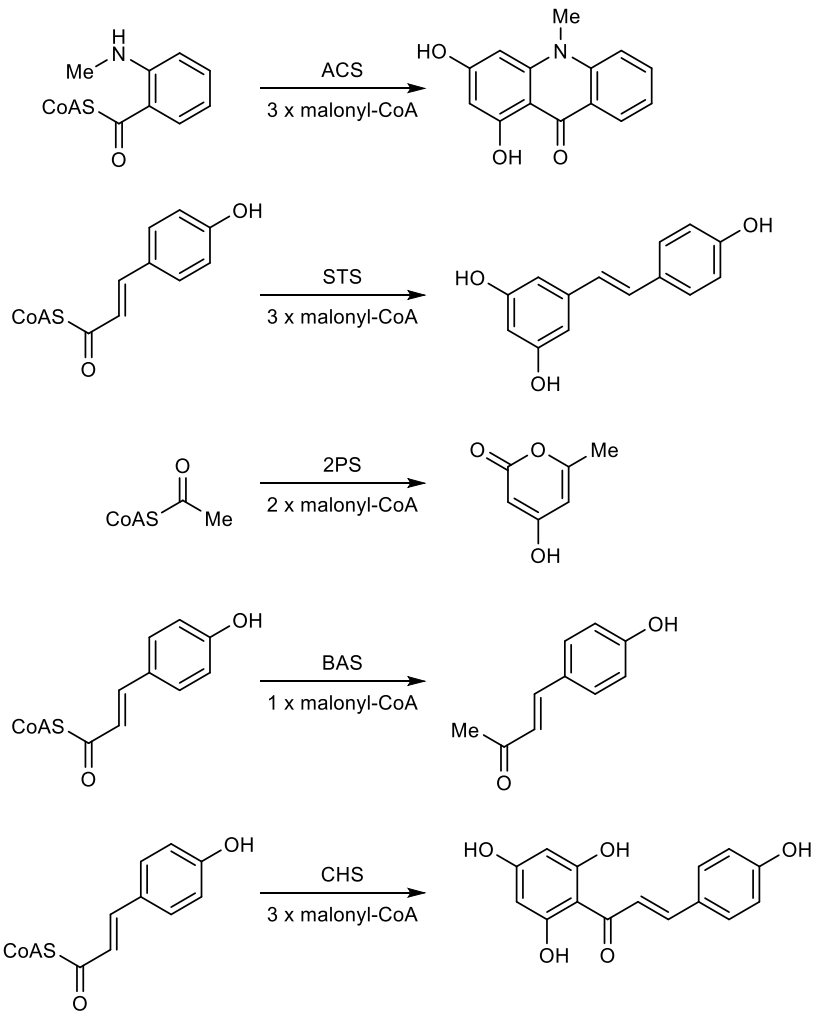


Figure 4. Examples of type III PKS reactions.
Type III PKS synthase reactions shown are catalysed by acridone synthase (ACS), stilbene synthase (STS), 2-pyrone synthase (2PS), benzalacetone synthase (BAS) and chalcone synthase (CHS). Adapted from Abe, *et al*. [83].

1.2.2.2 Curcuminoid polyketide synthases

In 1997, Joachim Schröder predicted that curcuminoids were synthesized by type III PKSs [87], based upon radio tracer study evidence [88]. Despite this prediction it took ten years for the first candidate enzymes to be identified in *C. longa*. It was contested whether the methyl ethers were formed from cinnamoyl-CoA units, then subsequently hydroxylated and methoxylated or formed from already methoxylated cinnamoyl-CoAs, such as feruloyl-CoA [89].

The first enzyme capable of producing curcuminoids was identified in rice (*Oryza sativa*) despite there being no evidence of rice producing curcuminoids. Katsuyama *et al*.[90] discovered it by conducting a BLAST search in the genome of *Oryza sativa* which identified 30 type III PKSs. All 30 were recombinantly expressed in *E. coli* and incubated with various CoA-esters; the products were then analysed using mass spectrometry. From this, one of the putative type III

PKSs was found to synthesize curcuminoids and is now known as curcuminoid synthase (CUS). *In vitro*, CUS produces BDMC from the condensation of two molecules of *p*-coumaroyl-CoA and one malonyl-CoA (Figure 5). Interestingly however, *Oryza sativa* does not produce curcuminoids so the *in vivo* function of CUS remains unknown. CUS was the first example of a type III PKS to break the conventional head-to-tail rule of PKSs, as the β-keto acid intermediate functions as the extender in the next step of the reaction, as opposed to the substrate [90]. Two crystal structures have been solved for CUS, one at 2.5 Å resolution [91] and the other at 2.0 Å resolution [92]. The large active site discovered, was unusual for type III PKSs and was believed to allow the unusual head-to-head condensation where the β-keto acid acted as a second extender substrate.

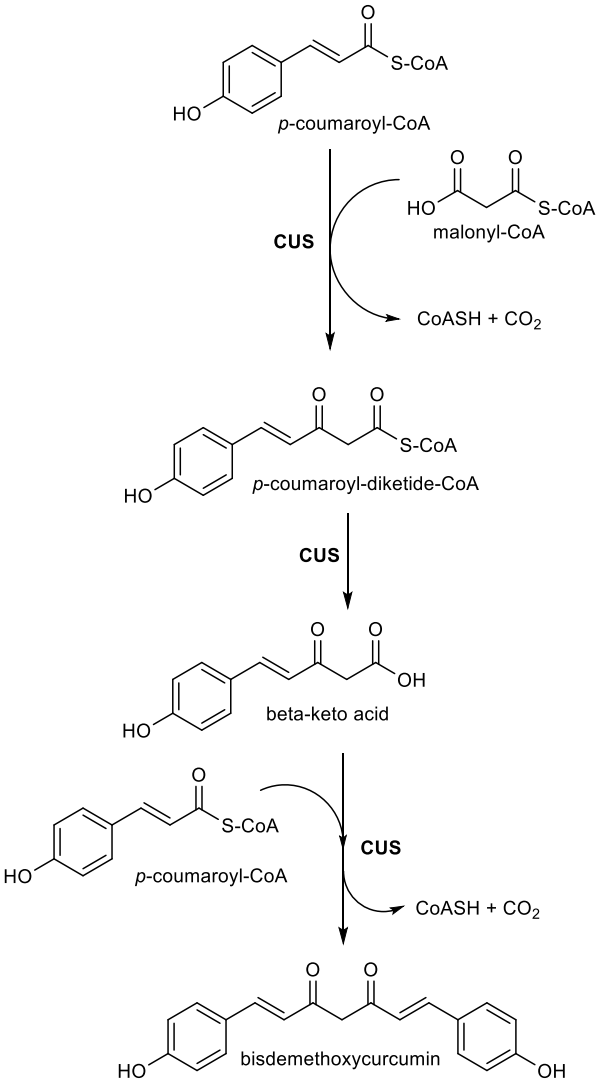


Figure 5. Curcuminoid synthase (CUS) reaction scheme.
CUS forms bisdemethoxycurcumin from *p*-coumaroyl-CoA and malonyl-CoA, as demonstrated by Katsuyama *et al.*[90].

Two years after the discovery of CUS, Katsuyama's group identified two type III PKSs from *C. longa* involved in curcuminoid biosynthesis: diketide CoA synthase (DCS, AB495006.1 see Appendix 1.1) and curcumin synthase 1 (CURS1, AB495007.1 see Appendix 1.2) [93]. In contrast to CUS, the single enzyme from *O. sativa*, which converts *p*-coumaroyl-CoA to BDMC, DCS first preferentially converts feruloyl-CoA to feruloyl-diketide-CoA using AB495006.1 malonyl-CoA as an extender substrate (Figure 6). DCS can also accept *p*-coumaroyl-CoA, but at a lower efficiency [93]. The optimum pH and temperature for DCS activity were determined to be 6.5-7.5 and 25-35˚C, respectively. CURS1 was shown to produce curcuminoids when incubated with the chemically synthesized feruloyl-diketide-CoA mimic cinnamoyldiketide-N-acetylcysteamine and feruloyl-CoA. Like CUS from *O. sativa*, CURS1 performed an unusual head-to-head condensation reaction. The optimum pH and temperature for CURS1 catalysis were calculated to be a particularly high 9 and 50˚C, respectively. The co-incubation of DCS and CURS1 with *p*-coumaroyl-CoA, feruloyl-CoA and malonyl-CoA allowed the production of curcumin, DMC and BDMC (see Figure 6), therefore it was concluded that together these two type III PKS functioned to synthesize curcuminoids in *C. longa*. There are reports of a second isoform of DCS, DCS2, however it has only been mentioned in passing in the literature (unpublished data, AB535216, Kita *et al*. [94]). Two additional isoforms of curcumin synthases, CURS2 and CURS3, were later identified that had different substrate specificity to CURS1 accounting for the mixture of curcuminoids found in *C. longa* [95]. CURS2 shared 78 % amino acid homology to CURS1, whereas CURS3 shared 81 %. CURS2 was shown to prefer feruloyl-CoA as a substrate, whereas CURS3 utilised feruloyl-CoA and *p*-coumaroyl-CoA with a similar efficiency. RT-PCR studies showed CURS1 and CURS2 were primarily expressed in the rhizomes, whereas CURS3 was expressed equally between the leaves and roots. Using an RNA-Sequencing approach, the expression of DCS and CURS1-3 was shown to correlate with the abundance of each of the three curcuminoids produced in *C. longa* and *C. aromatica* [96]. Using the ArREST (Aromatic Rhizome EST) database, a turmeric EST database curated by David Gang's research group [97], Resmi *et al*. [98] identified two other type III PKSs in *C. longa*, ClPKS9 and ClPKS10, however, they have not been shown to participate in curcuminoid biosynthesis. Although only the PKSs have been identified in the *C. longa* curcuminoid biosynthetic pathway, biosynthesis is believed to begin with the amino acids tyrosine and phenylalanine (Figure 7). The amino acids are then deaminated by PAL/TAL and proceed through the phenylpropanoid biosynthetic pathway to produce phenylpropanoyl-CoAs, which are then differentially accepted by the different isoforms of DCS and CURS to produce all three curcuminoids (Figure 7).

Using laser micro-dissection of *C. longa* rhizomes and ultra-high performance liquid chromatography- mass spectrometry, Jaiswal *et al*. [99], identified the cork and cortex of the rhizome to contain the highest abundance of curcuminoids. Conversely, using Direct Analysis in

Real Time (DART) mass spectrometry, Rahman *et al*. [100] concluded that the majority of the curcuminoids accumulate within the pith of the rhizome. Therefore, to date, there is no conclusive evidence suggesting the key rhizome tissue(s) involved in curcuminoid biosynthesis. There is also a lack of data indicating the subcellular localisation of the type III PKSs (DCS and CURS) involved in curcuminoid biosynthesis or how the curcuminoid products are stored within the cell. Plant natural products usually undergo bio-conjugation or sequestration to reduce their potent bioactivities within the cell. However, ginger (*Zingiber officinale*), a fellow member of the Zingiberaceae family has been shown to sequester its polyphenolic metabolites, including 6-gingerol, within terpenoid liposomes [101]. It is possible that *C. longa* adopts a similar strategy for storing curcuminoids.

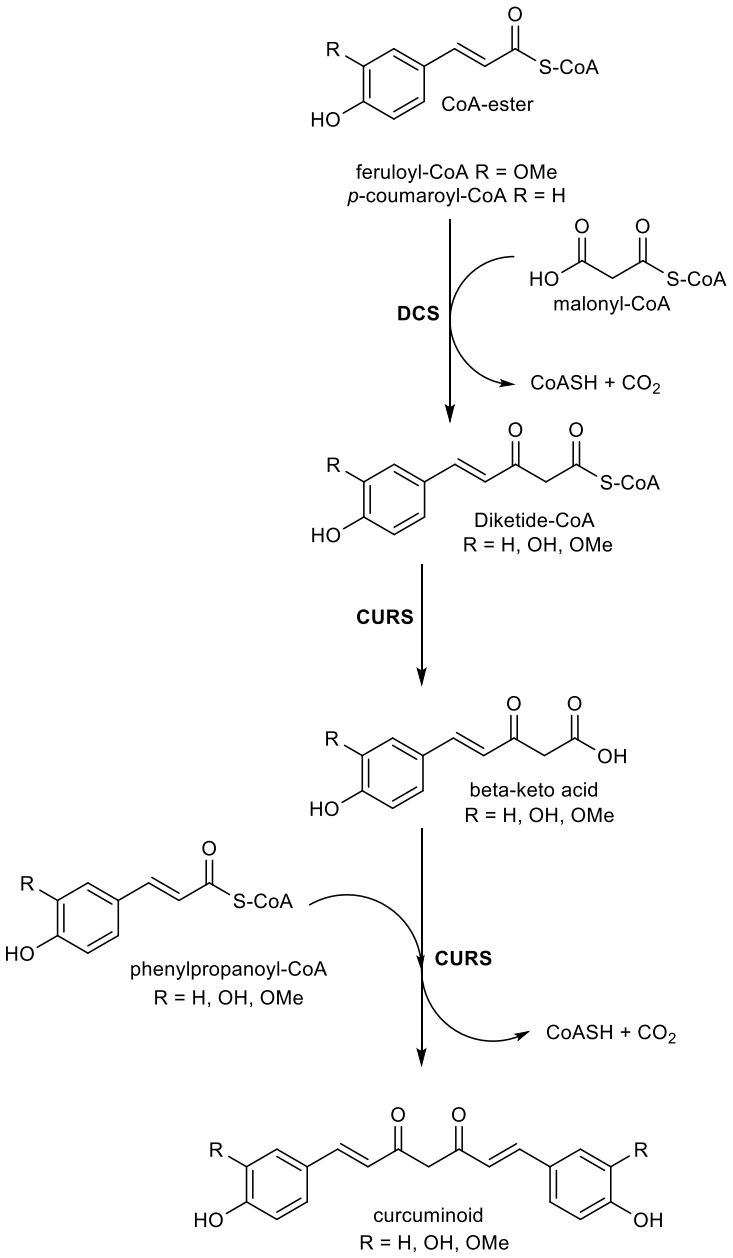


Figure 6. Diketide CoA synthase (DCS) and curcumin synthase (CURS) reaction scheme.
DCS and CURS synthesize curcuminoids from *p*-coumaroyl-CoA or feruloyl-CoA and malonyl-CoA, as demonstrated by Katsuyama *et al*. [93].

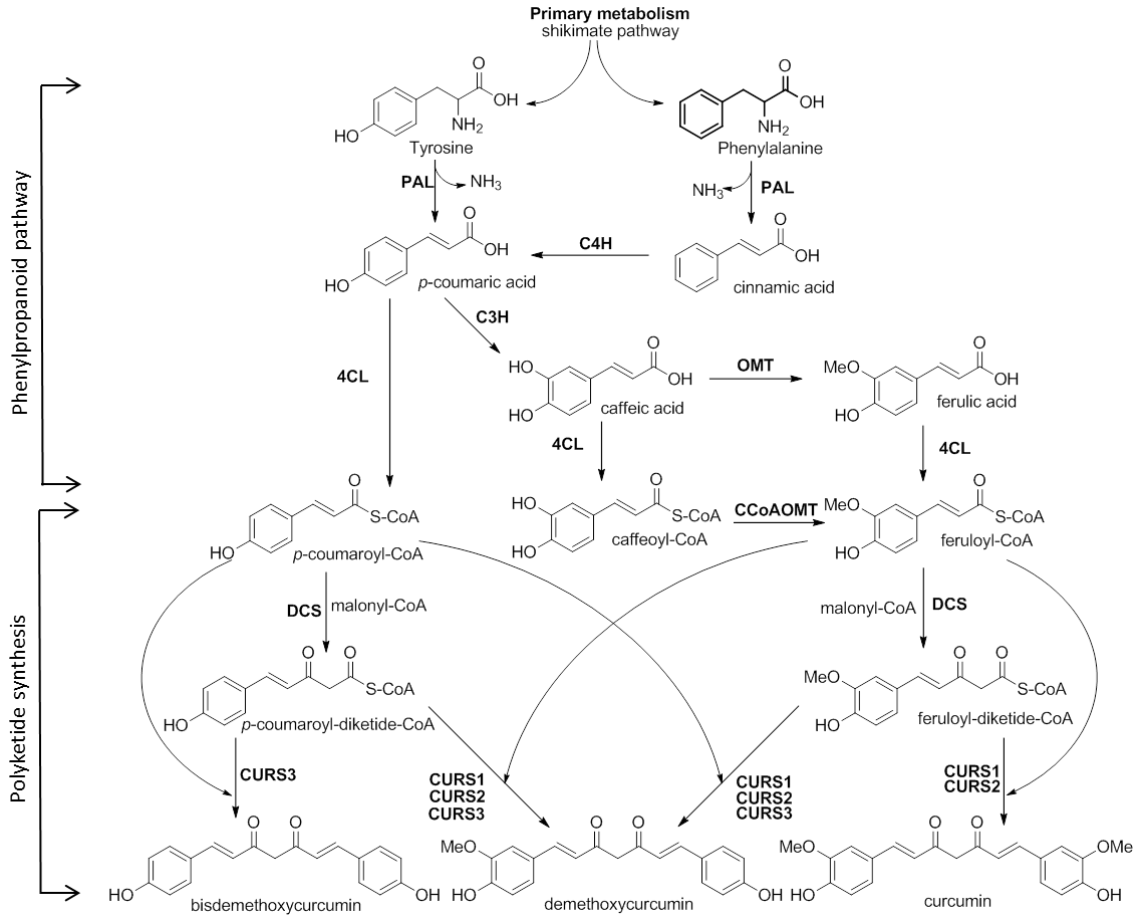


Figure 7. Predicted schematic of curcuminoid biosynthesis in *C. longa* rhizomes.

Tyrosine or phenylalanine are deaminated by tyrosine/phenylalanine ammonium lyase (T/PAL) to form phenylpropanoic acids cinnamic acid and *p*-coumaric acid. Cinnamic acid is converted to *p*-coumaric acid by C-4 hlydorxylase. *p*-Coumaric acid can be converted to caffeic acid by C-3 hydroxylase (C3H), the meta-hydroxyl group of which can be methylated by caffeoyl-CoA O-methyltransferase (CCoAOMT) to give ferulic acid. The phenylpropanoic acids *p*-coumaric acid, ferulic acid and caffeic acid are then activated by CoA esterification by a 4-coumarate CoA ligase (4CL). The CoA esters are then condensed with malonyl CoA by diketide-CoA synthase (DCS) to produce the diketide-CoA, which is condensed with another phenylpropanoyl-CoA to yield the curcuminoid products curcumin, demethoxycurcumin and bisdemethoxycurcumin.

1.2.3 Chemical and biological production of curcuminoids

Due to the bioactivity of curcuminoids, much work has been carried out to synthesize these molecules. This began in 1918 with Lampe's five step synthesis starting from ethyl acetoacetate and carbomethoxyferuloyl chloride (Figure 8) [102]. Other chemical routes have also been explored by Sabitha *et al*. [103] and Pabon *et al*. [104].

With the discovery of the PKS enzymes involved in their synthesis, heterologous expression of the curcuminoid biosynthetic pathway has been used in an effort to sustainably produce curcuminoids. After identifying CUS in *Oryza sativa*, Katsuyma's group produced an artificial curcuminoid biosynthetic pathway that utilised this enzyme.[28] The curcuminoid pathway was engineered in *E. coli* and comprised phenylalanine ammonia-lyase (PAL) from the yeast *Rhodotorula rubra*, 4-coumarate CoA ligase (4CL) from *Lithospermum erythrorhizon* and CUS. The metabolically engineered *E. coli* was fed with tyrosine or phenylalanine and produced three different curcuminoids with the highest yield being 107 ± 18 mg L^{-1} of dicinnamoylmethane when fed with phenylalanine. However, triketide pyrone derailment products were also produced in quantities up to 209 ± 18 mg L^{-1}. In an attempt to minimise this, they repeated the feeding study without the PAL deamination step and fed the bacteria with phenylpropanoic acids instead. This increased the curcuminoid yield approximately three times and reduced triketide pyrone production by up to 10 %. Using this rationale, Katsuyama *et al*. [105] adopted a precursor directed strategy to produce novel curcuminoids by feeding with non-natural phenylpropanoic acids. *Le*4CL and CUS were able to utilise these non-natural substrates and produce novel curcuminoid derivatives such as di-4-chlorocinnamoylmethane, di-2-furylacryloylmethane and di-3-indoleaclyloylmethane (Figure 9).

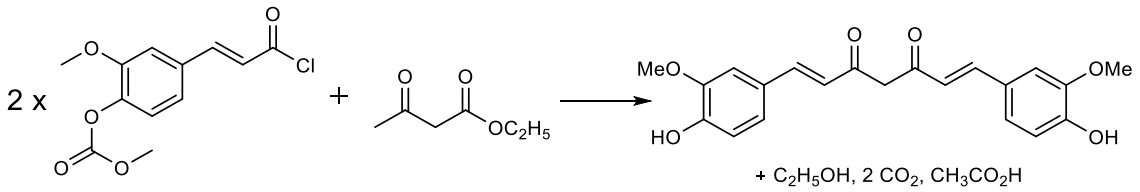


Figure 8. Chemical synthesis of curcumin.
Scheme of curcumin synthesis as carried out by Lampe [102], using carbomethoxyferuloyl chloride and ethyl acetoacetate.

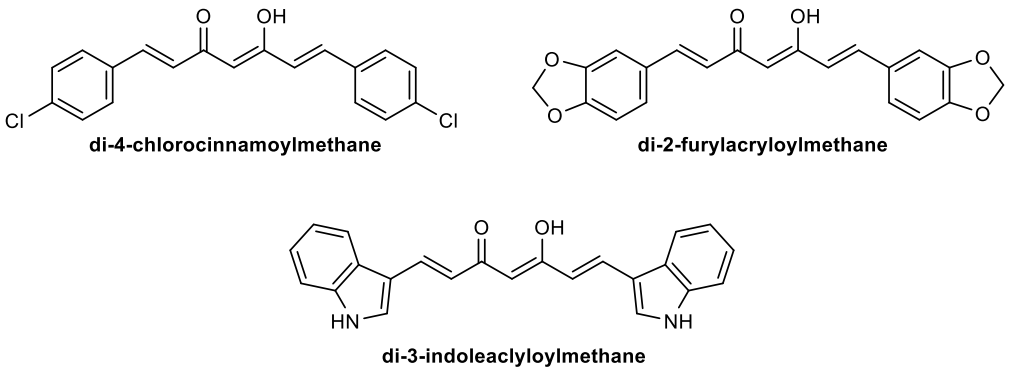


Figure 9. Chemical structures of three non-natural curcuminoids.
Katsuyama *et al.*[105] produced a range of non-natural curcuminoids, including di-4-chlorocinnamoylmethane, di-2-furylacryloylmethane and di-3-indoleaclyloylmethane from non-natural phenylpropanoids, using the heterolgous expression of *Le*4CL and CUS in *E. coli*.

Wang *et al.*[106] took advantage of curcuminoids characteristic yellow colour, using CUS as part of a reporter assay in *E. coli*. By co-expressing *At*4CL1 and CUS with a range of PALs and feeding the cells phenylalanine they used the amount of curcuminoids produced (measured by colour intensity) to determine the PAL with the best functionality in *E. coli*, which could then be used in other heterologous polyphenolic biosyntheses.

In 2012, at the beginning of this project, DCS and CURS1 had not been used in an effort to biosynthesize curcuminoids. However, in 2015, Rodrigues *et al.*[107] used DCS and CURS1 in conjunction with *At*4CL1 to produce curcuminoids in *E. coli*. They compared the curcuminoid production using DCS and CURS1 with that of CUS and saw that DCS and CURS1 yielded almost thirty times more curcumin than CUS (187.9 μM titre compared to 6.7 μM titre). Rodrigues *et al.*[107] also demonstrated curcuminoid production from tyrosine via caffeic acid by expressing tyrosine ammonia lyase (from *Rhodotorula glutinis*) 4-coumarate 3-hydroxylase (from *Saccharothrix espanaensis*) and caffeoyl-CoA 3-O-methyl-transferase (from *Medicago sativa*) in combination with *At*4CL1, DCS and CURS1, although this resulted in lower curcuminoid yields.

1.2.4 Increasing curcuminoid bioavailability

A significant limitation to the pharmaceutical use of curcuminoids is their susceptibility to hydrolytic degradation in aqueous, physiological solutions reducing their efficacy as bioactive compounds [108]. This degradation can be minimised by maintaining the *p*H of the solution below 7, however, this drastically reduces the molecules' solubility [109]. This further decreases their bioavailability as drug molecules as they do not meet all of the ADMET criteria [110,111]. The ADMET criteria stipulates that to be an effective drug the molecule must be easily Absorbed into the body, Distributed quickly and to the correct organs, Metabolised and Excreted at an appropriate pace and have low Toxicity. For example, when human volunteers orally ingested 2

g of curcumin, after an hour only 6 ng of curcumin could be detected per ml of blood [112]. Intravenous delivery of curcumin in rats also results in its rapid metabolism to glucuronide and sulphate conjugates before being excreted [113,114].

Many different approaches have been investigated to address the poor bioavailability of curcuminoids including altering how they are formulated to modifying their chemical structure [115]. Co-administering curcuminoids with an adjuvant, such as piperine, showed a 2000 % increase in bioavailability in humans [112]. Liposome formulation has also been considered as a way of increasing the stability of curcuminoids, encapsulating them in polymeric micelles increased their stability up to 500 times in physiological conditions [116]. Silica coated liposomes encapsulating curcumin saw the gastrointestinal absorption increase even further compared to standard lipid micelles [117]. Further formulation development utilising lipids saw production of curcumin loaded solid lipid nanoparticles, prolonging the release of curcumin and increasing its stability in *in vitro* studies [118]. Another way to overcome the low bioavailability of curcuminoids is to create molecular analogues with improved pharmacokinetic properties. This area has been extensively explored with a vast range of analogues being synthesized from varying the heptanoid chain length, changing ring substituents and altering the enone moiety [66]. One focus of the current thesis is curcuminoid glycoside analogues.

1.3. Glycosyltransferases and curcuminoid glycosylation

One way to increase the bioavailability of curcuminoids is to create a glycoconjugate via glycosylation. Glycosylation can be carried out enzymatically or chemically, however, enzymatic reactions offer greater stereospecificity and regiocontrol [119]. Glycosylation of small molecules increases their hydrophilicity and stability, generally giving the chemicals more desirable pharmacokinetic properties [120]. Enzymatic glycosylation is often preferable to chemical glycosylation as it is regiospecific and can be performed using less harsh reaction conditions without the need for heavy metal catalysts [119]. Enzymatic glycosylation is mainly carried out using glycosidases or glycosyltransferases. Glycosidases reversibly catalyse the hydrolysis of glycosidic bonds, therefore under appropriate conditions, they can be made to produce glycoconjugates. Some glycosidases have been engineered to become glycosynthases by removing their ability to hydrolyse the glycosidic bond [121]. Conversely, glycosyltransferases naturally evolved to perform glycosylations and catalyse the transfer of the sugar from an activated donor (such as the nucleotide uridine di-phosphate (UDP)) to the acceptor. Although glycosidases are robust enzymes, they generally exhibit lower regioselectivity and yields compared to the rarer and more expensive glycosyltransferases [119].

1.3.1 Plant glycosyltransferases

The variety of secondary metabolites produced by plants is due to the diverse chemical skeletons which are decorated by tailoring enzymes to tune their activity. One example of tailoring enzymes are glycosyltransferases. Glycosyltransferases play an important role in modulating the plasticity of plant responses to their environment. Glycosylation of secondary metabolites usually functions to stabilise and reduce their activity meaning they can be safely stored until needed. A common example is that of cyanogenic glycosides which are typically stored in the vacuole. When the plant is under attack by herbivory, the vacuole is damaged releasing the cyanogenic glycoside and allowing the glycosidic bond to be hydrolysed, this releases toxic hydrogen cyanide deterring further feeding [122]. Glycosylation can also be utilised by plants as a crucial step in detoxifying xenobiotics, such as herbicides, leading to their breakdown, sequestration or transport out of the cell [123].

Glycosyltransferases that catalyse the transfer of a hexose sugar from an activated donor to the acceptor molecule are known as UDP-glycosyltransferases (UGTs). UDP-glucose is the typical sugar donor, but UDP- galactose, UDP- rhamnose and UDP- glucuronic acid have also been identified as donors [124]. The acceptor site is usually a hydroxyl group (Figure 10), however, glycosylation can occur at many different nucleophilic functional groups such as carboxylic acids, amines, thiols and also at carbon atoms [124]. UGTs usually function as the last biosynthetic step, but in some pathways, such as that which produces flavonoid glycosides, glycosylation occurs during the synthesis of the backbone [125]. UGTs can be retaining or inverting with the anomeric carbon keeping the same stereochemistry or being inverted (see Figure 10).

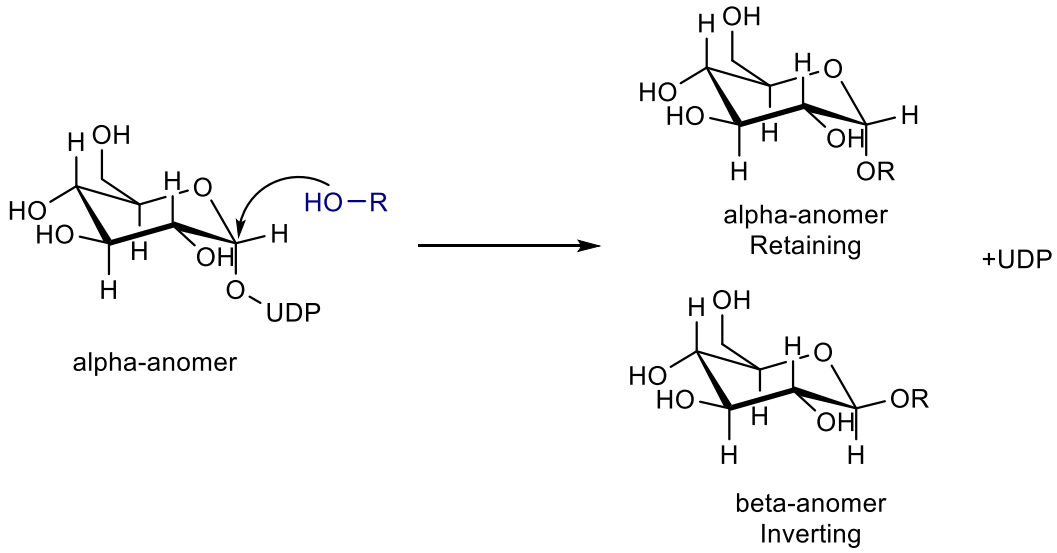


Figure 10. Glucosylation reaction.
Nucleophilic attack of acceptor (dark blue) upon sugar donor (UDP-α-D-glucose) (black) leads to the formation of a glycosidic bond. The reaction can either retain the configuration of anomeric carbon centre or invert it.

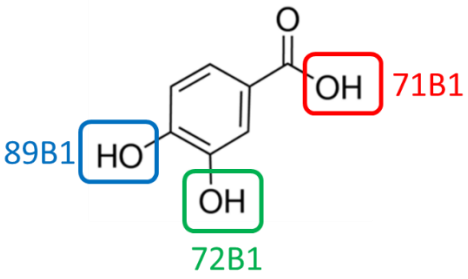


Figure 11. Glycosylation of 3,4-dihydroxybenzoic acid.
Three Arabidopsis UGTs, 71B1, 72B1 and 89B1 all have different regiospecificity for the glycosylation of 3,4-dihydroxybenzoic.

UGTs are made up of a large multi-gene family, comprising over 78 families [126]. Family 1 UGTs are defined by the presence of the plant secondary product glycosyltransferases (PSPG)-box; a 44 amino acid motif, enabling them to transfer sugars to a wide variety of secondary metabolites [127]. The PSGP-box is C-terminally located, where amino acid residues are involved in donor binding. Phylogenetic analysis of the N-terminal of Family 1 UGTs, shows that this terminal is more diverse and has a role in determining the binding specificity of the acceptor substrates [128]. Utilising a functional genomics approach, recombinant *Arabidopsis thaliana* Family 1 UGTs, were divided into 14 groups (A to N) [126,128], with the selectivity of the enzymes varying across the groups. UGTs from other plants also aligned to these 14 groups [123]. UGTs are highly regiospecific as exemplified by the differential glycosylation of the simple molecule 3,4-dihydroxybenzoic acid by three

Arabidopsis UGTs: group L 75B1 glycosylates the carboxylic acid; group E 72B1 glycosylates the *meta* hydroxyl group and group B 89B1 glycosylates the *ortho* hydroxyl group [129] (Figure 11). Family 1 UGTs carry out an inverting glycosylation mechanism resulting in the β-anomer [120].

Although more work has been carried out with microbial and mammalian GTs, work with plant UGTs is increasing. Family 1 UGTs have been used as biocatalysts to produce glycosides of many different small molecules. As with other bioconversions, they can be used *in vitro* as purified recombinant proteins or in whole plant or microbial cell systems. As Arabidopsis Family 1 UGTs are the most well described, these have seen the greatest attention for heterologous expression in microbial systems. For example, heterologous expression of UGT73B3 and UGT71C1 in *E. coli* allowed the production of a range of quercetin glycosides at a 6 L fermenter scale [130]. *S. cerevisiae* has also been used as a successful host for the heterologous production of glycosides. An interesting example, highlighting the unpredictable and complex nature of recombinant bioconversions, was shown by Strucko *et al.*, again utilising a UGT from Arabidopsis [131]. They constructed identical heterologous vanillin glycoside biosynthetic pathways in a laboratory and industrial strain of yeast: surprisingly the laboratory strain produced 10 times more product, emphasising the importance of choosing the correct strain and host for biocatalysis.

The use of whole cell plant systems allows endogenous enzymes to be used, negating the need for heterologous expression. The major advantage of utilising whole cell systems over *in vitro* catalysis is that the cells are able to regenerate the sugar donors, reducing the need for supplementation which is costly. Plant cells offer not just one glycosyltransferase but potentially multiple UGTs which may have different regioselectivities. However, due to UGTs having a role in detoxification, the glycosylated product may end up being targeted to and stored in the vacuole, making recovery and extraction more difficult. A prominent cell culture is that of *Catharanthus roseus* used for many glycosylation reactions such as the production of glycosylated capsaicin and Δ^9-tetrahydrocannabinol [132,133].

1.3.2 Glycosylation of curcuminoids

Due to the highly bioactive, yet poorly bioavailable nature of curcuminoid chemicals, different routes to glycosylate these molecules have been investigated. The first successes were observed using chemical syntheses, but more recent advances have seen the use of enzymes.

1.3.2.1 Chemical glycosylation of curcuminoids

The first industrially relevant curcuminoid glycosylation was seen in 2001, when Hergenhahn *et al.* [134], patented their route to synthesize a range of curcumin oligosaccharides. This was an involved chemical reaction scheme, requiring bromination at the anomeric carbon of the sugar

and protection by acetylation of the other hydroxyl groups of the sugar(s) to ensure regioselective glycosylation when reacted with curcumin in the presence of $Et_3BnN^+Br^-$. With glycosylation, Hergenhahn *et al*. [134] observed an increase in water solubility compared to the curcumin aglycone, as well as the ability to reduce inflammation and act as a tumour suppressor [134]. However, after de-protection, the yields were poor, giving only 3 % curcumin 4',4"-*O*-β-D-diglucoside (curcumin diglucoside) (Figure 12). Utilising a different rationale, expanding on the chemical synthesis of Pabon *et al*. [104] and Whiting *et al*. [84], Mohri *et al*.[109] achieved better yields of 34 % curcumin diglucoside. Instead of using curcumin as a starting material, they condensed two already glucosylated vanillin molecules using a boron catalyst [109]. Nevertheless, this reaction scheme still required protection and deprotection steps which lowered the overall yield. Benassi *et al*. [111], also utilised vanillin and bromoacetogluose as starting materials and a boron catalyst, however, they additionally carried out thorough computational analyses of the products crystal structures and partion co-efficients. They showed that not only did glycosylation of curcuminoids increase the molecules' bioavailability, but it also increased the molecules' ability to act as bidentate iron chelating agents which may be useful in iron overload therapy [111]. The most recent and highest yielding chemical synthesis of curcumin glucosides utilised a biphasic reaction medium and a tetrabutylammonium bromide phase transfer catalyst under simple and ecofriendly conditions, yielding 71 % curcumin diglucoside [135]. However, once again, a loss in yield was observed due to protection and deprotection to make to reaction regioselective. The curcumin glucoside was shown to exhibit enhanced antioxidant and antimicrobial activities [135].

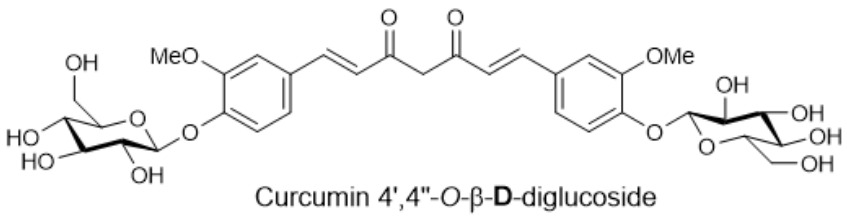


Figure 12. Curcumin diglucoside.

1.3.2.2 Biological glycosylation of curcuminoids

Despite the fact that curcumin glycosides are yet to be discovered in nature, biological syntheses of curcuminoid glycosides have been successfully explored. The main aim of investigating biological systems was to achieve greater regioselective glycosylation and avoid the protection and de-protection steps necessary for chemical syntheses. Two types of enzyme have been shown to perform curcumin glucosylation, glucosidases and glycosyltransferases.

The two instances of glucosidase catalysed curcuminoid glycosylation both utilised commercially available enzyme preparations. The first evidence of glucosidase catalysed curcumin glycosylation was presented by Vijayakumar *et al*. [136] in which they used an α-glucosidase from *Rhizopus*. Usually used for the hydrolysis of α-D-glucosides, by optimising the pH and stoichiometry of reagents, the highest yield of curcumin-bis-α-D-glucoside was achieved at 48 %. Similarly, Shimoda *et al*. [137] used an almond β-D-glucosidase, to produce curcumin monoglucoside, however, the yield was much lower at 19 %.

The first observation of a glycosyltransferase catalysed glycosylation of curcuminoids was in 2003: Kaminaga *et al*. [138] reported that *Catharanthus roseus* cell suspension cultures converted exogenously applied curcumin to five different glucosides (curcumin 4'-*O*-β-D-glucoside; curcumin 4'-*O*-β-D-gentiobioside; curcumin 4',4"-*O*-β-D-diglucoside; curcumin 4'-*O*-β-D-gentiobiosyl 4"-*O*-β-D-glucoside and 4',4"-*O*-β-D-digentiobioside (Figure 13)). They showed that curcumin was not only glycosylated once, but up to four times forming curcumin 4',4"-*O*-β-D-digentiobioside. This high level of glycosylation had not been seen before, with most plant cell cultures converting exogenous compounds with multiple hydroxyls to only the monoglucoside [139,140]. The water solubility of curcumin 4',4"-*O*-β-D-digentiobioside was shown to be an impressive 20 million fold greater than curcumin (6.6×10^{2} μmol ml^{-1} compared to 3.0×10^{-5} μmol ml^{-1}) [138]: the other glucosides also showed greatly improved hydrophilicity. Addition of a stress elicitor, methyl jasmonate, to the *C. roseus* cells, saw the yield of glucoside production increase to 2.5 μmol g^{-1} fresh cells (32 %) [138].

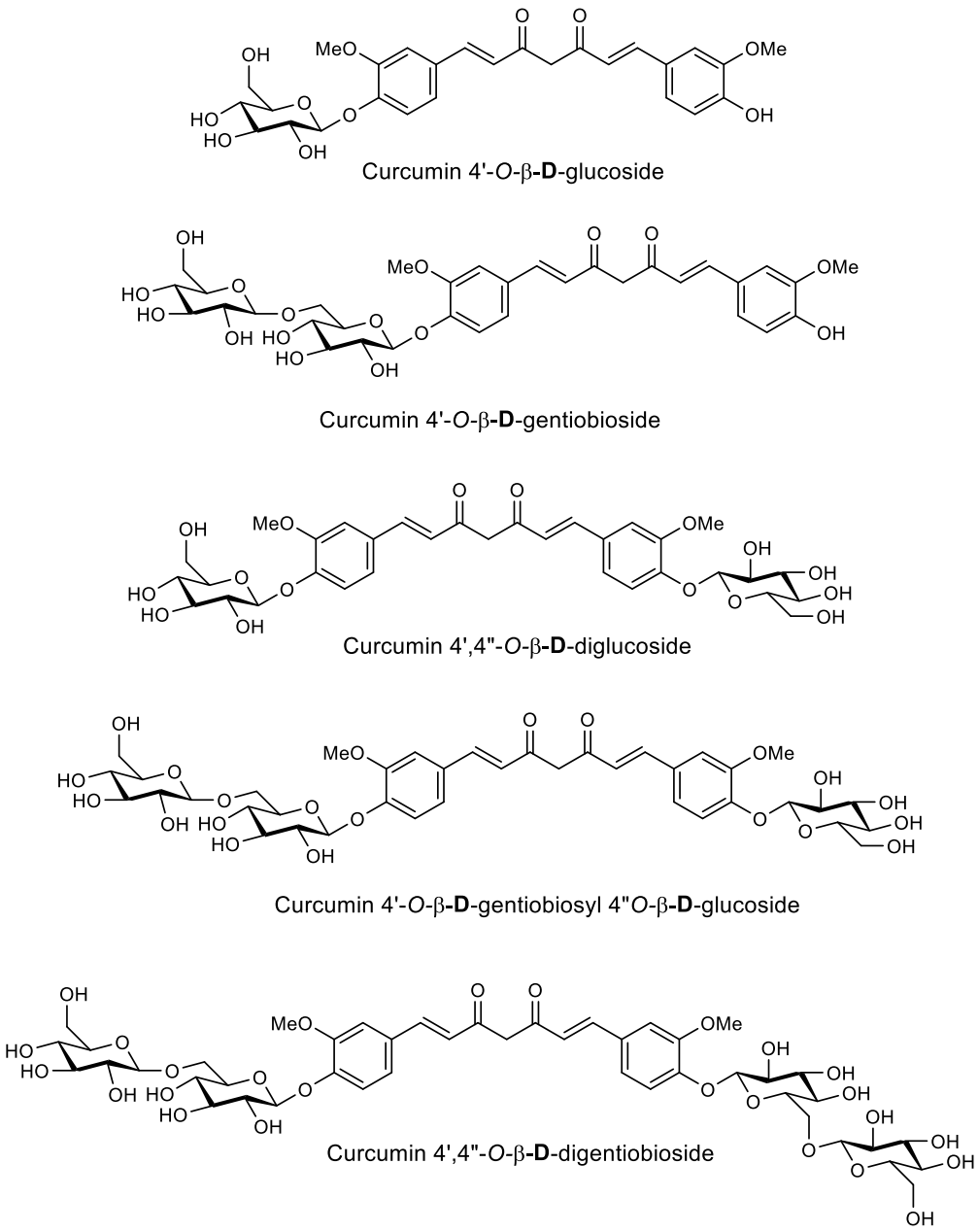


Figure 13. Curcumin glucosides produced by *C. roseus* cell culture.
Five curcumin glycosides were produced by *C. roseus* cell culture: curcumin 4'-*O*-β-D-glucoside; curcumin 4'-*O*-β-D-gentiobioside; curcumin 4',4"-*O*-β-D-diglucoside; curcumin 4'-*O*-β-D-gentiobiosyl 4"-*O*-β-D-glucoside and 4',4"-*O*-β-D-digentiobioside.

The UGT responsible for mono- and di- glucosylation was characterised the following year as CaUGT2 (GenBank accession number AB159213, Appendix 1.3) [1]. Expression of CaUGT2 in *E. coli* showed that it preferentially converted curcumin to curcumin monoglucoside, followed by catalysis of curcumin monoglucoside to curcumin diglucoside (Figure 14). CaUGT2 showed optimal activity between pH 7.5-8. Phylogenetic analysis of CaUGT2, placed it within group D of the Family 1 UGTs, along with Arabidopsis UGT73C6 and other diverse UGTs [1]. When

incubated with curcumin and UDP-glucose *in vitro* for three hours, CaUGT2 was able to produce curcumin diglucoside in a 74 % yield. When the analogous whole cell assay was performed in *E. coli*, only 4 % curcumin monoglucoside could be detected [1]. CaUGT2 has also been shown to glycosylate a range of phenolics including carvacrol, thymol, eugenol and vanillin but at a lower efficiency than curcumin glycosylation [141]. A specific amino acid, Cys377, has been shown to have an important role in the catalytic activity of CaUGT2, although this residue has little influence in other UGTs tested, so is thought to be unique to CaUGT2 [127].

Being the first characterised glycosyltransferase to allow the production of curcumin glucosides, CaUGT2 was subsequently used by Masada *et al*. [142], to create an *in vitro* curcuminoid glucoside production platform. There are problems with using UGTs as *in vitro* catalysts as UDP-glucose is an expensive starting material and the reaction by-product, UDP, inhibits the glycosylation reaction, therefore as the reaction proceeds, it slows down. Masada *et al*. [142] combatted these problems by co-expressing a recombinant sucrose synthase from *Arabidopsis* (*At*SUS1) alongside CaUGT2, which could regenerate UDP-glucose. By creating this one-pot recycling system, the yield of curcumin glucosides increased nine-fold.

In 2009, the identity of the sugar-sugar glycosyltransferase that formed the curcumin gentiobiosides and digentiobiosides in *C. roseus* plant culture was published as CaUGT3 (accession number AB443870) [143]. CaUGT3, purified from *C. roseus* plant cell culture, catalysed the conversion of curcumin 4'-*O*-β-D-glucoside and curcumin 4',4"-*O*-β-D-diglucoside to curcumin 4'-*O*-β-D-gentiobioside and 4',4"-*O*-β-D-digentiobioside, respectively (see Figure 14) [144]. As well as being able to glycosylate curcumin glycosides, it also showed activity towards other flavonoid glycosides such as quercetin 3-*O*-glucoside and apigenin 7-*O*-glucoside [143]. To date, CaUGT3 has not been explored for the bioproduction of curcuminoid gentiobiosides.

Despite no previous reports of *Parthenocissus tricuspidata* cell culture glycosylating exogenous substrates, Shimoda *et al*. [145] showed that it was able to glycosylate exogenous curcuminoids [145]. Although unrelated to *C. roseus*, *P. tricuspidata* cell culture was also able to produce a range of glycosylated curucuminoids as shown in Figure 13. Currently, the enzymes responsible for these conversions have not been identified.

As has been shown, utilising the glycosyltransferase enzymes present in plants to modify exogenous substrates is an effective way to produce glycosyl-conjugates of curcuminoids. The enzymatic reactions have greater specificity and utilise milder conditions compared to chemical glucosylation. However, the scope for scaling the enzymatic production of curcuminoid glycosides has not been explored.

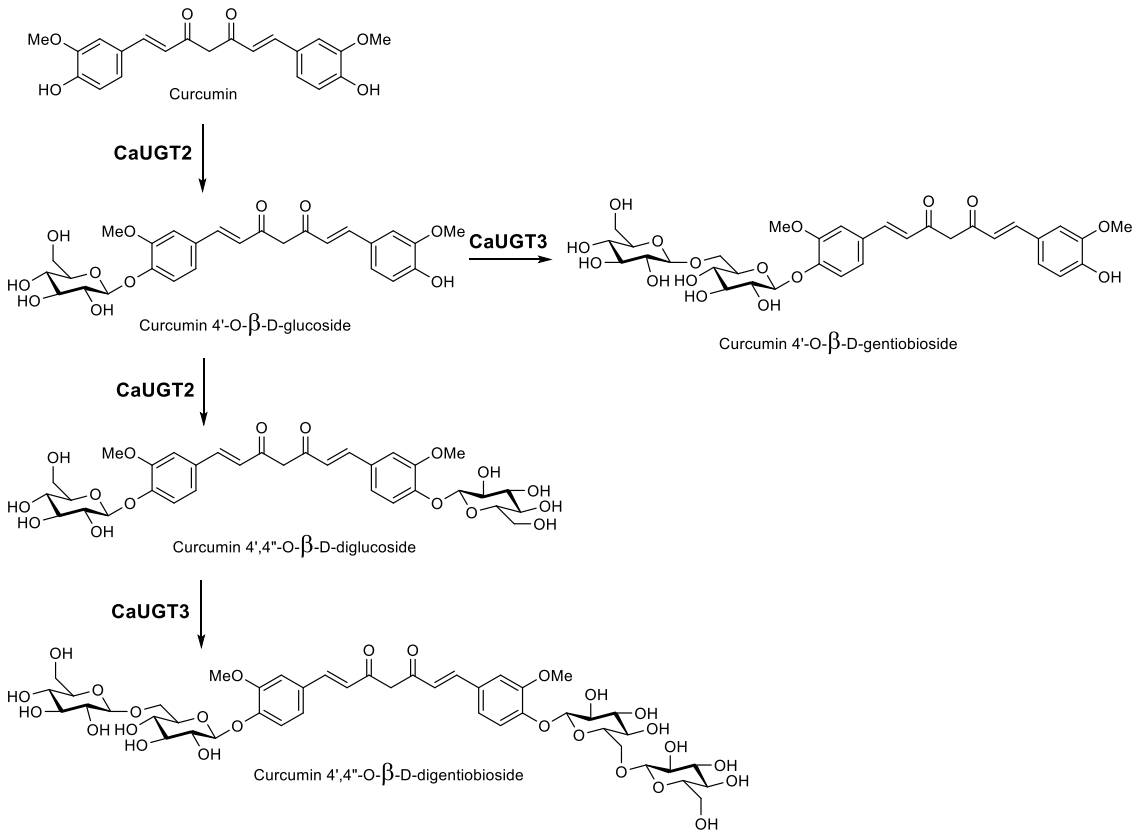


Figure 14. Schematic of curcumin glucosylation as catalysed by CaUGT2 and CaUGT3.

1.4. Project Aims

This study aimed to explore the heterologous expression of plant secondary metabolic enzymes to produce and improve the bioavailability curcuminoid chemicals. Both a microorganism and plant host, *S. cerevisiae* and *Nicotiana benthamiana*, were to be investigated as chassis organisms. As curcuminoids are poorly bioavailable, glycosyltransferases were to be explored as a way of synthesizing more bioavailable conjugates. Whilst previous curcuminoid biosyntheses have utilised enzymes from plant sources which themselves do not produce curcuminoids, this study aimed to discover and utilise enzymes from *C. longa* which synthesizes curcuminoids.

To identify new enzymes, rapid amplification of cDNA ends (RACE) PCR, metabolite profiling and protein activity assays were used. Chapter 2 details the attempts to obtain a full-length coding sequence of a 4CL from *C. longa*. It also describes the expression and activity of *At*4CL5, DCS and CURS1 in *S. cerevisiae* and *N. benthamiana* in the attempt to produce curcuminoids from phenylpropanoic acids.

In Chapter 3, the search for a curcuminoid UGT in *C. longa* is described. Additionally, the development of CaUGT2 as a biocatalyst to produce bioavailable curcuminoid glycosides and the scale up of the bioconversion is investigated.

Chapter 4 outlines the methods which were used and then Chapter 5 summarises the conclusions of the studies, possible future work and discusses this works contribution to the wider literature.

Chapter 2. Biosynthesis of curcuminoid chemicals

2.1 Introduction

Curcuminoids are highly bioactive diarylheptanoids produced in the rhizomes of the turmeric plant (*Curcuma longa*). The three main curcuminoids; curcumin, demethoxycurcumin and bisdemethoxycurcumin (see Figure 1) are made via the polyketide synthase pathway and are structurally similar yet have different biological and physicochemical properties [79,93]. The ratio of the three curcuminoids is determined by the abundance of starting materials and isoforms of enzymes present [95,96]. The main aim of this project was to individually biosynthesise the three curcuminoid chemicals in a heterologous host.

Although curcuminoids have been heterologously produced in *E. coli* (yield ~ 100 mg L^{-1}), unwanted derailment products were also produced due to the use of curcuminoid synthase (CUS) from *O. sativa* [146]. In this study different enzymes were to be used to reduce derailment products and *Saccharomyces cerevisiae* (Bakers' yeast) was chosen as a primary host. *S. cerevisiae* was chosen because of its enhanced ability to express eukaryotic proteins over *E. coli*, similar intracellular compartments to those of plant cells and wide use in biotechnological processes [38,147,148]. Tobacco (*Nicotiana benthamiana*) was also to be investigated and compared as a chassis platform for curcuminoid production.

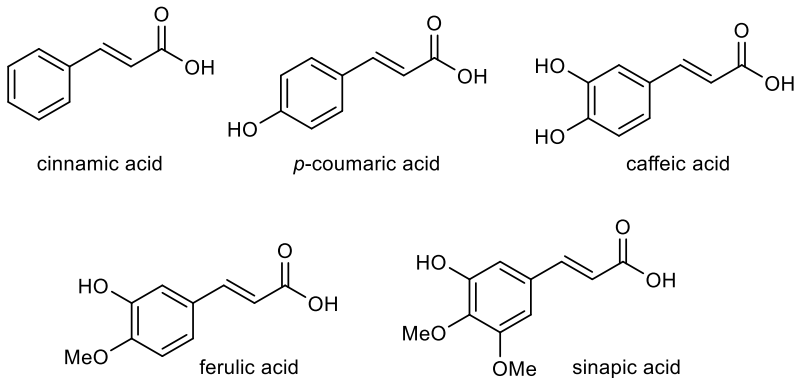


Figure 15. The main phenylpropanoic acids present in plant biomass.

To produce curcuminoids, phenylpropanoids (or phenylpropanoic acids) were chosen as starting materials due to their abundance in bio-refinery waste, allowing for the sustainable production of curcuminoids from a low value feedstock (Figure 15). They have been successfully used in the bio-production of other aromatic, high value chemicals such as vanillin [149,150]. Using ferulic acid as a substrate would theoretically allow curcumin production; *p*-coumaric acid would yield BDMC and DMC could be produced from a mix of both. Caffeic and cinnamic acid could be used as starting materials to produce less abundant natural curcuminoids such as dicinnamoylmethane. Sinapic acid is not a natural starting material for curcuminoid biosynthesis. Unpublished work carried out by Prof Rob Edwards' group determined that phenylpropanoic acids do not negatively affect the growth of *S. cerevisiae* at the concentrations to be used. In yeast, phenylpropanoic acids have also been shown to be metabolised by PAD1 (phenylacrylic acid decarboxylase 1) and FDC1 (ferulic acid decarboxylase 1). Some bioengineering studies of *S. cerevisiae* utilising phenylpropanoids, knocked out these competing pathways [151], however, the literature showed their rate of catabolism to be minimal [152] so is not always necessary if the introduced heterologous pathway can outcompete the endogenous metabolism. In tobacco, exogenous phenylpropanoids are anticipated to have no detrimental effects at low concentrations due to being present for endogenous lignin biosynthesis [153]. The effect of exogenous curcuminoids upon tobacco leaf tissue was unknown.

To catalyse the biotransformation of phenylpropanoids to curcuminoids a 4-coumarate coenzyme A ligase (4CL), to activate the substrate, and polyketide synthase(s), to condense the activated substrates, were required (Figure 16). To date, the polyketide synthases required for the condensation reactions have been identified in *C. longa* as isoforms of DCS and CURS [93], but the 4CL has not. Nonetheless, curcuminoid biosynthesis is believed to proceed via a 4CL as shown in Figure 7. Therefore, to heterologously produce curcuminoids, a 4CL would need to be identified from *C. longa* or an available coding sequence chosen from another organism.

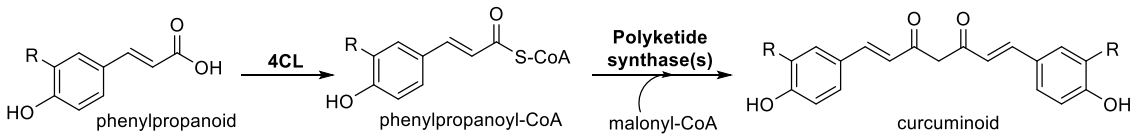


Figure 16. The enzymatic steps needed to produce curcuminoids from phenylpropanoids.

4-coumarate CoA ligases

4CLs play a key role in plant metabolism by helping to shuttle the carbon flux from primary metabolism along the phenylpropanoid pathway into secondary metabolism. They are found in all plants examined to date [154]. The phenylpropanoid pathway begins with the deamination of the amino acids phenylalanine or tyrosine (derived from the shikimate pathway) by phenylalanine/tyrosine ammonium lyase (PAL/ TAL) (see Figure 7). If phenylalanine is the starting material, C4-hydoxylase hydroxylates the *para* position of benzene ring producing *p*-coumaric acid. *p*-Coumaric acid can then be hydroxylated again by C3H or methylated by O-methyltransferase or CCoAOMT giving the four main phenylpropanoids cinnamic acid, *p*--coumaric acid, ferulic acid and caffeic acid. The phenylpropanoids are then activated by 4CL forming CoA esters. Different 4CL isoforms within, and between, plant species exhibit different phenylpropanoid substrate preferences; however, *p*-coumaric, ferulic and caffeic acid are generally preferred with little activity shown towards cinnamic acid [153]. There are also a few examples of 4CLs accepting sinapic acid as a substrate [155]. The activated phenylpropanoyl-CoA esters can then be used as precursors to form numerous compounds including lignin, flavonoids and coumarins which play important roles in growth and responding to biotic and abiotic stresses [153,156].

4CLs use coenzyme A (CoA) to convert phenylpropanoic acids, into their respective CoA thiol esters. 4CL catalysis is a two-step process which utilises ATP and Mg^{2+} and proceeds via a 4-hydroxycinnamoyl-adenylate intermediate [155,157]. This places 4CLs into a large superfamily of enzymes known as adenylate forming enzymes which have divergent functions and include firefly luciferases, acyl-CoA synthetases and peptide synthetases. 4CL peptide sequences contain two highly conserved regions box I (SSGTTGLPKGV) and box II (GEICIRG) which flank the region involved in substrate recognition [158].

In tobacco, parsley and hybrid poplar, all the 4CL isoforms show similar activity towards *p*-coumaric, ferulic, caffeic and cinnamic acid [154,159–161], whereas Arabidopsis, poplar, rice and switchgrass 4CLs have differential substrate utilisation profiles [162–165]. This has been taken as evidence that the 4CLs that exhibit different substrate specificities, function in different metabolic pathways [166]. In the dicotyledons studied, Arabidopsis and aspen, the 4CLs can be split into two groups: type I and type II depending on their involvement in phenolic metabolism or lignin biosynthesis [162,163]. The monocotyledons, rice and switchgrass, have a different phenolic and monolignol composition, with their 4CLs grouping into type III and type IV, respectively [164,167,168].

The most extensively studied 4CLs are from the model plant, Arabidopsis. There are four characterised 4CLs: *At*4CL1, 2, 3 and 4 (although named *At*4CL4 in original studies, it is now

referred to as *At*4CL5 [166]). *At*4CL1, 2 and 5 belong to type I whereas *At*4CL3 belongs to type II; not only differing in their substrate specificity but, also their expression patterns [162]. *At*4CL1 prefers *p*-coumaric and caffeic acid; *At*4CL2 accepts just *p*-coumaric acid and At4CL3 has the best activity with *p*-coumaric and ferulic acid [162]. *At*4CL5 was the second 4CL identified to CoA esterify sinapic acid [166], a rarely catalysed 4CL reaction, the first was from soyabean [169]. *At*4CL5 shows the lowest K_m for sinapic acid followed by ferulic acid and *p*-coumaric acid [166]. It is understood to be the size and hydrophobicity of the substrate binding pocket that determines the substrate binding specificity [170–172].

The best described monocotyledonous 4CLs are from *O.sativa* (rice), of which there are five known in the genome: *Os*4CL1-5, all exhibiting different catalytic properties. *Os*4CL1 has a low turnover rate, *Os*4CL2 prefers ferulic acid, *Os*4CL3 has the highest turnover and together with *Os*4CL4 shows high affinity for ferulic acid and *p*-coumaric acid and *Os*4CL5 has a broad activity as well as being able to accept sinapic acid [173]. Of the five rice 4CLS, *Os*4CL3 was shown to be the most highly expressed in all tissues. *Os*4CL2 is the only rice 4CL which belongs to the type IV clade involved in flavonoid biosynthesis, whereas the other four are associated with type III lignin formation [164]. Other monocotyledonous 4CLs, such as those from switchgrass (*Panicum virgatum*) are currently being studied, as silencing 4CLs leads to a decrease in the lignin content of the plant improving it as feedstock for biofuel [165].

To date, no full length 4CL from *C. longa* has been identified. This project aimed to address this using the available transcriptome data and rapid amplification of cDNA ends polymerase chain reaction (RACE PCR) [174]. The newly identified 4CL coding sequence (CDS) was to be used in the heterologous curcuminoid biosynthetic pathway.

Curcuminoid type III polyketide synthases

To perform the condensation of the activated CoA ester with malonyl-CoA to produce curcuminoids, there were two options for choosing polyketide synthases. There was the single enzyme curcuminoid synthase (CUS) from *O. sativa* or diketide CoA synthase (DCS) and a curcumin synthase (CURS), from *C. longa* (see Figure 5 and Figure 6). CUS has substrate preference for *p*-coumaroyl-CoA resulting in BDMC being the main product [90]. There are multiple isoforms of DCS and CURS which have differing substrate preferences [95].

DCS and CURS were chosen over CUS as the latter has no *in vivo* function in rice, shows low substrate promiscuity and produces unwanted products when expressed heterologously [93]. There are two isoforms of DCS, 1 and 2, however DCS2 is not described in the literature (unpublished

data, Kita, *et al*. [94]), therefore DCS1 was chosen; it is most commonly referred to as DCS. From the three isoforms of CURS (1, 2 and 3), CURS1 was chosen as it has the broadest substrate specificity and is the most well characterised [95].

For expression in yeast, the chosen coding sequences, 4CL, DCS and CURS1, were to be cloned into yeast expression vectors. The vectors, pYES2 and pESC were chosen due to previous use within Prof Rob Edward's and Prof Ian Graham's research groups [175,176]. Yeast strains, BJ2168, G175 and BY4722 were used for these proof of concept studies, although other production strains, such as CEN.PK2, are more commonly used by industry and biotechnology [148]. A variety of strains were used throughout the project due to differing auxotrophies, phenotype and availability (see Materials and Methods). The enzymes were to be cloned and transformed into yeast in two ways: individually into compatible yeast expression vectors and altogether in a single open reading frame construct separated by the 2A polypeptide CDS. The curcuminoid biosynthetic pathway was to be reproduced in tobacco using *Agrobacteria tumefaciens* mediated transient expression [177]. This required cloning the CDSs of 4CL, DCS and CURS1 into the pFGC5941 binary vector. This would allow a comparison of the yeast and tobacco production platforms.

2.2 Results

2.2.1 Investigating the heterologous expression of curcuminoid biosynthetic enzymes in *S. cerevisiae*

2.2.1.1 The heterologous expression of a 4CL

Choosing a 4CL

As there is no known coding sequence for a 4CL from *C. longa*, it was decided to search the *C. longa* NCBI transcript database. Using the NCBI tBLASTn expressed sequence tags (EST) search, five distinct ESTs with the highest homology (63-76 % sequence identity) to rice *Os*4CL3 were identified: DY382977.1, DY391029.1, DY382750.1, DY390122.1 and DY391030.1 (see Appendix 2.1) [173]. A 4CL from rice was chosen as the query sequence because rice is a monocotyledon, like *C. longa*, and it has the best described 4CLs. *Os*4CL3 was specifically chosen as it has the highest turnover rate of the five 4CLs in rice, is the most highly expressed 4CL in all rice tissues, and shows good activity towards ferulic and *p*-coumaric acid [173].

When analysed phylogenetically, rice and *Panicum virgatum* (switchgrass) monocotyledon 4CLs (*Os*4CL1-5 and *Pv*4CL1-2, respectively) were aligned as previously shown by Sun *et al.* [164]: *Os*4CL2 and *Pv*4CL2 belong to type IV whereas the rest belong to type III (Figure 17). Type IV

*Os*4CL2 and *Pv*4CL2 have both been described to be involved in flavonoid biosynthesis [165,173], whereas the type III 4CLs are involved in lignin biosynthesis [164]. The similarity of the five putative *C. longa* 4CLs to the members of type III suggests that they are likely to be type III 4CLs.

The five *C. longa* ESTs did not cover the full-length *Os*4CL3 CDS with both the N and C terminal sequences missing (Figure 18). Of the five ESTs, the longest, DY382977.1 (839 nucleotides in length, compared to full-length 4CLs of ~1650 nucleotides) gave 47 % coverage and 70 % amino acid homology to *Os*4CL3. However, when analysing the key substrate binding region bordered by the conserved peptide motifs, Box I and Box II, the latter part of the substrate binding region was missing (Figure 19).

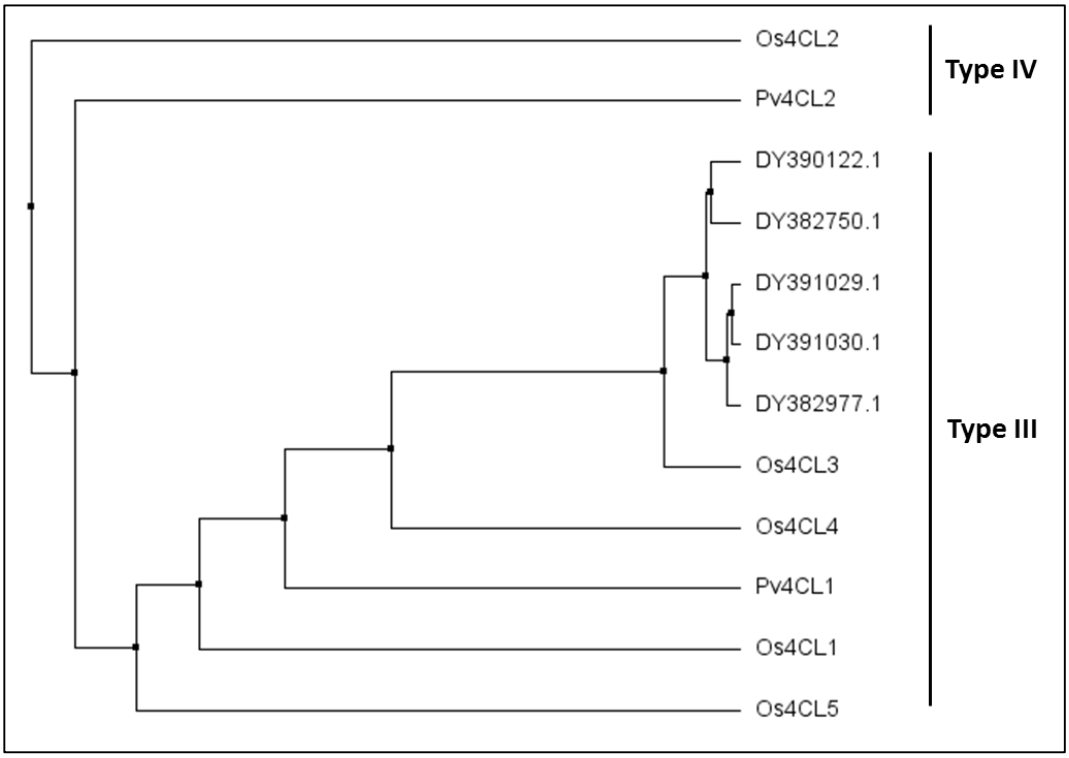


Figure 17. Phylogenetic analysis of five potential 4CL ESTs from *C. longa*.
The five potential 4CL ESTs from *C. longa* are compared with other monocotyledon 4CLs: *Os*4CL1-5 from rice and *Pv*4CL1 and *Pv*4CL2 from switchgrass. They arrange into two groups based on their sequence identity as shown by Sun *et al.* [164]. *Os*4CL2 and *Pv*4CL2 belong to type IV, whereas *Os*4CL1, 3-5 and *Pv*4CL1 belong to type III. Similarity of the *C. longa* 4CL ESTs to the type III 4CLs suggests they are likely to belong to the type III class of 4CLs.

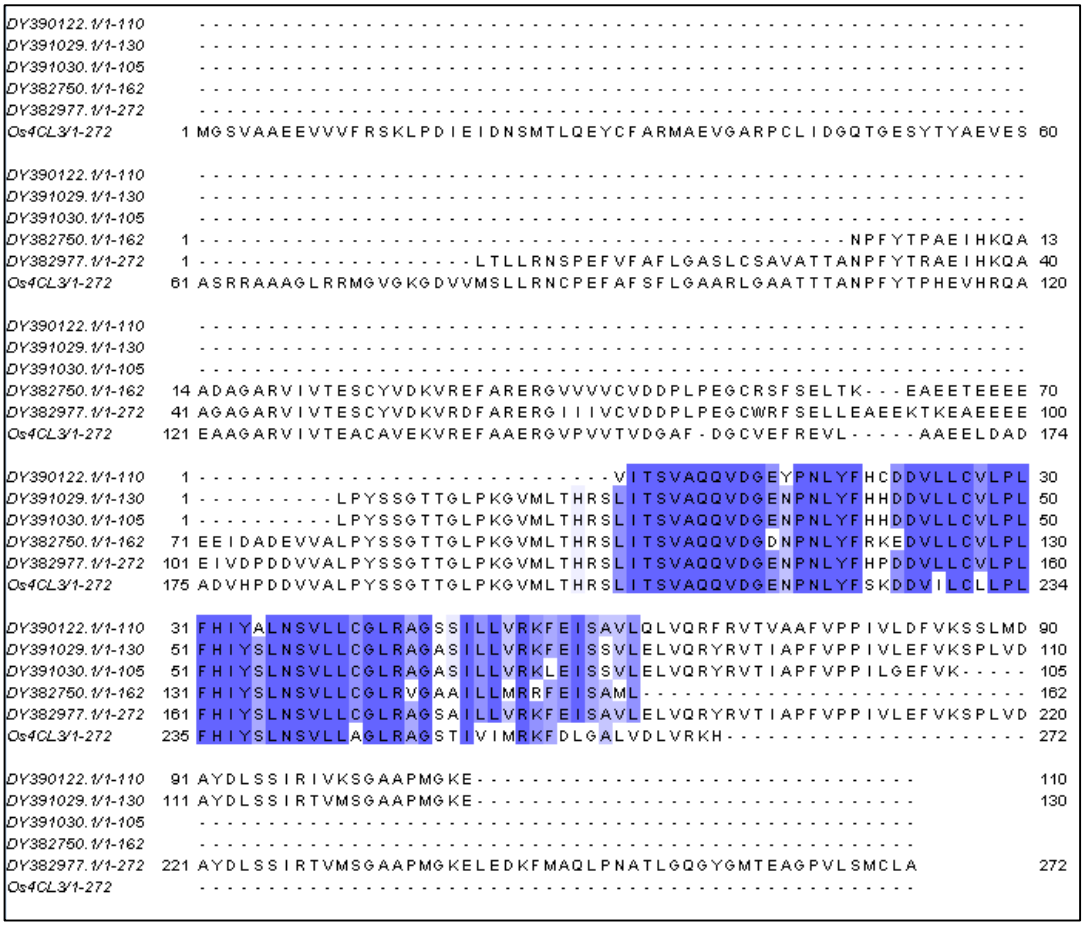


Figure 18. Amino acid sequence alignment of five potential *C. longa* 4CL ESTs and *Os*4CL3.
The homology of DY382977.1, DY391029.1, DY382750.1, DY390122.1 and DY391030.1 and *Os*4CL3 is compared: the darker the shade of blue, the higher the homology between the amino acid sequences.

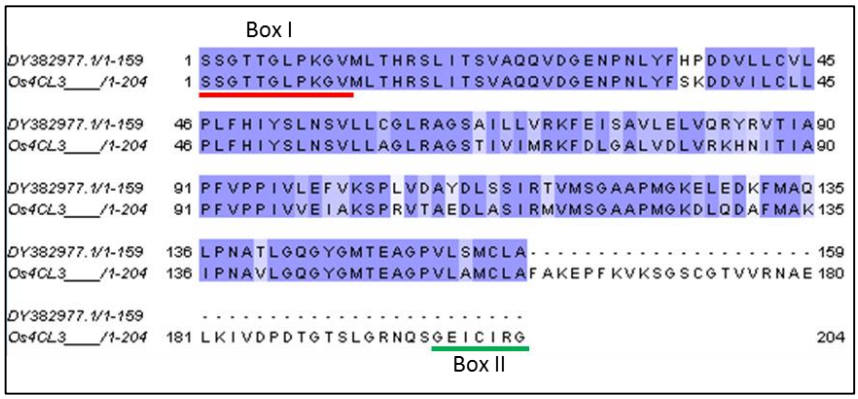


Figure 19. 4CL substrate binding domain comparison between *Os*4CL3 and DY382977.1.
The sequence is reduced to the region flanked by the conserved peptide motifs box I (underlined red) and box II (underlined green). The darker the shade of blue, the greater the homology between the two sequences. The sequences show 70 % homology, although the C-terminal end of the substrate binding region of DY382977.1 is missing. The denoted numbering is according to each single polypeptide, as found in the database.

The ability to detect the five chosen ESTs in *C. longa* RNA was first checked. RNA was extracted from *C. longa* rhizomes using a method specific for tissue high in polyphenolics and lignin (see Materials and Methods) [178]. cDNA synthesis was then carried out, upon which polymerase chain reaction (PCR) and nested PCR was carried out using primers designed to amplify the five ESTs (see Primer Table in Materials and Methods): only DY382977.1 and DY391029.1 could be amplified giving the products of the expected sizes, 0.8 kB and 0.4 kB, respectively (Figure 20A). The PCR products were sequenced and compared to the expected EST sequence from the NCBI transcript database (Figure 20B and C). DY382977.1 cDNA amplified from RNA isolated from rhizomes showed more variation to the NCBI transcript database than DY391029.1, with multiple single nucleotide polymorphisms and insertions. Despite this, as DY382977.1 was the longest detectable mRNA, it was chosen as the sequence to work with. As both the C and N terminal were missing (see Figure 18) 3', 5' RACE PCR was required to obtain a full-length, functional sequence. In spite of multiple attempts (using larger quantities of rhizome tissue, freshly extracted RNA, newly synthesized cDNA and a range of different PCR primers and PCR conditions) the full-length DY382977.1 CDS was not amplified.

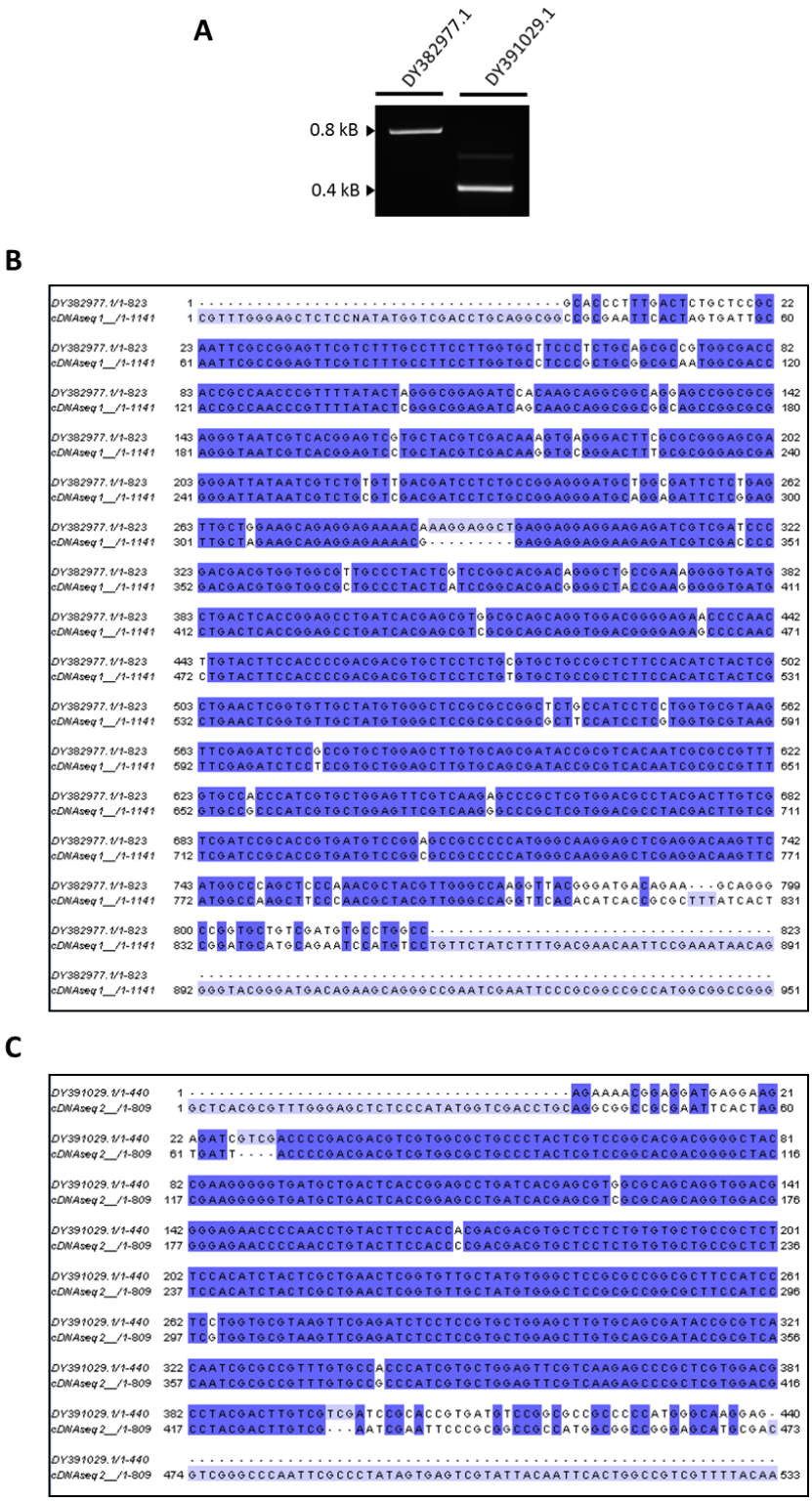


Figure 20. Verifying expression of DY382977.1 and DY391029.1.
(**A**) Agarose gel showing PCR products, using DY382977.1 (0.8 kB) and DY391029.1 (0.4 kB) specific primers to amplify cDNA, giving bands at expected sizes. The nucleotide sequence of DY382977.1 and DY391029.1 from cDNA synthesized from RNA extracted from *C. longa* rhizomes is compared to the expected NCBI EST sequences: (**B**) sequence alignment of DY382977.1 PCR product (cDNAseq1) with expected DY382977.1 database sequence; (**C**) sequence alignment of DY391029.1 PCR product (cDNAseq2) with expected DY391029.1 database sequence.

Expression and activity of 4CL5 in *S. cerevisiae*

To allow progression with the heterologous production of curcuminoids, it was decided to use a 4CL from a different source. *Os*4CL3 had exhibited poor enzymatic activity in *S. cerevisiae* (Dr F. Sabbadin, Prof Rob Edward's Lab, unpublished data) so a 4CL from Arabidopsis was chosen, as these are the most well described group of 4CLs. *At*4CL5 (AT3G21230.1, see Appendix 2.2, henceforth denoted as 4CL5), was chosen as, of the Arabidopsis 4CLs tested in proof of concept studies, it had shown the best enzymatic activity in yeast (Dr H. Housden, Edward's Lab, unpublished data). 4CL5 also has the widest substrate specificity of the Arabidopsis 4CLs [179].

4CL5 CDS was PCR amplified from *Arabidopsis thaliana* RNA and cloned into the pYES2 vector translationally fused to a *Strep*-tag using restriction and ligation (kindly donated by Dr H. Housden, Prof Rob Edward's Lab) (Figure 21A). *S. cerevisiae* cells were transformed with pYES2-4CL5 as verified by PCR, giving the expected product at 1.7 kB (Figure 21B). Transcription of 4CL5 from the pYES2-4CL5 vector was verified by reverse transcription-PCR (RT-PCR), giving the expected product at 0.9 kB (Figure 21C). Western blot analysis confirmed protein expression of *Strep*-tagged 4CL5, with the protein detected at 63 kD, as expected (Figure 21D). By carrying out whole cell (*in vivo*) and protein extract (*in vitro*) assays, it was anticipated that heterologous protein activity and endogenous yeast activity could be distinguished.

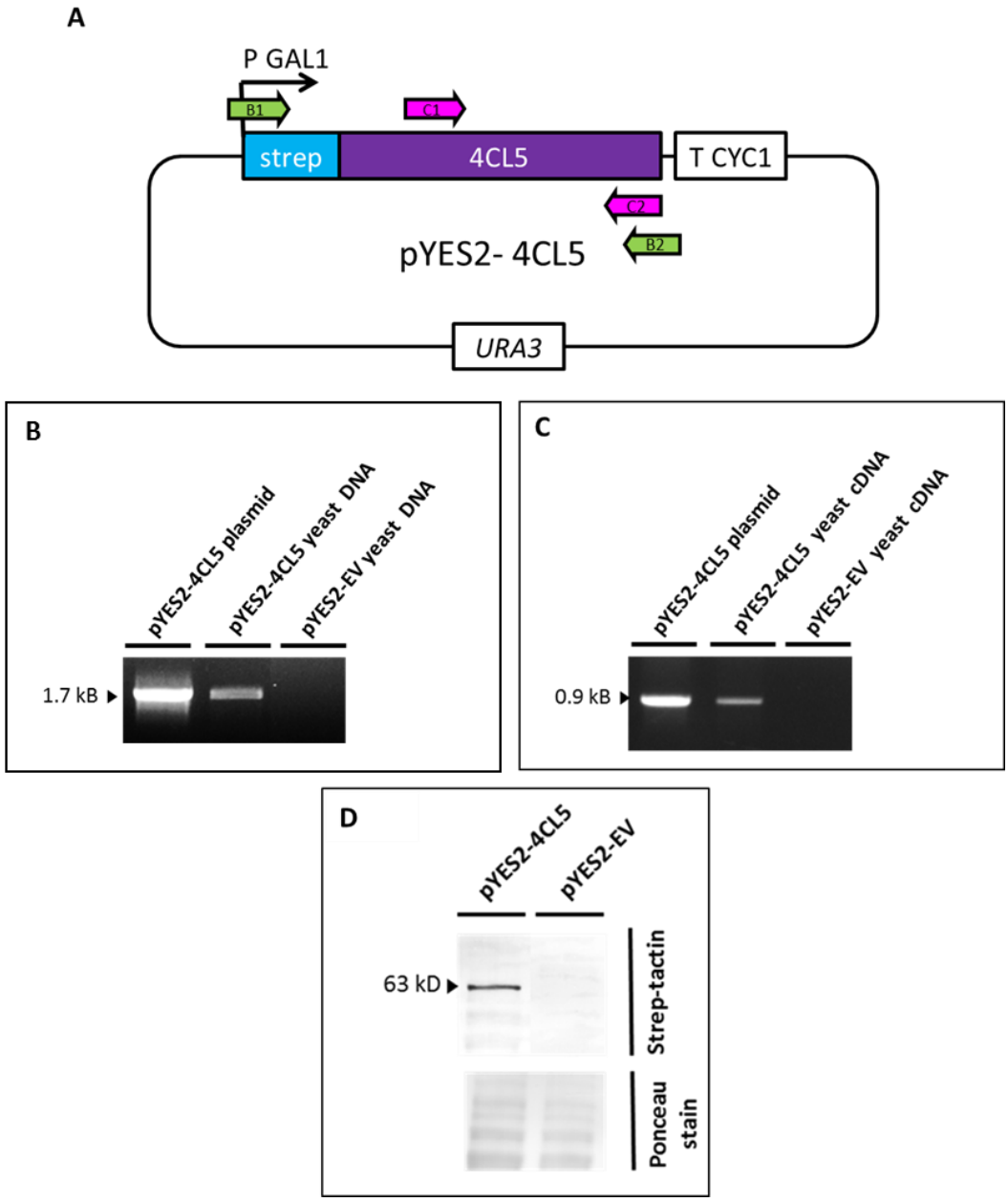


Figure 21. Cloning, transformation and expression of pYES2-4CL5.
(**A**) pYES2-4CL5 plasmid, for the expression of 4CL5 in *S. cerevisiae* under the control of a galactose inducible promoter. CYC1- transcriptional terminator for efficient termination; *URA3*- selection for transformants in yeast host strains with a *ura3* genotype. Primers are denoted by arrows for DNA and cDNA amplification: B1- F_Asc1_4Cl5; B2- R_Xba1_4CL5; C1- Haz_4CL5_int_L; C2- Haz_4CL5_int_R, see Materials and Methods for primer sequences. (**B**) Agarose gel electrophoresis showing diagnostic PCR products from PCR performed upon DNA extracted from transformed yeast, using primers B1 and B2 to determine positive transformation of yeast cells. The expected product, at 1.7 kB was amplified; pYES-4CL5 plasmid was used as a positive control template. (**C**) Agarose gel electrophoresis showing diagnostic PCR products from PCR performed upon cDNA synthesized from RNA extracted from transformed yeast cells, using primers C1 and C2 to determine transcription of 4CL5 from the pYES2-4CL5 vector. The expected product, at 0.9 kB was amplified; pYES-4CL5 plasmid was used as a positive control template. (**D**) Western blot analysis using *Strep*-Tactin to detect *Strep*-tagged 4CL5 protein (63 kD) in protein extracted from yeast cells transformed with the pYES2-4CL5 plasmid. Ponceau stained membrane is shown as a loading control. pYES2-EV is an empty vector control.

Whole yeast cells, expressing 4CL5, were fed *p*-coumaric acid and LC-MS analysis was used to observe the expected appearance of the product, *p*-coumaroyl-CoA. After 24 hours, *p*-coumaroyl-CoA could not be detected in the yeast cells or in the media, however, *p*-coumaric acid in the medium was consumed (Figure 22). As the consumption of *p*-coumaric acid is notably higher in yeast transformed with pYES2-4CL5 than it is in the empty vector control (pYES2-EV) (90 % reduction compared to 27 %, after 24 hours) it was concluded that 4CL5 was enzymatically active, despite the product not being detected.

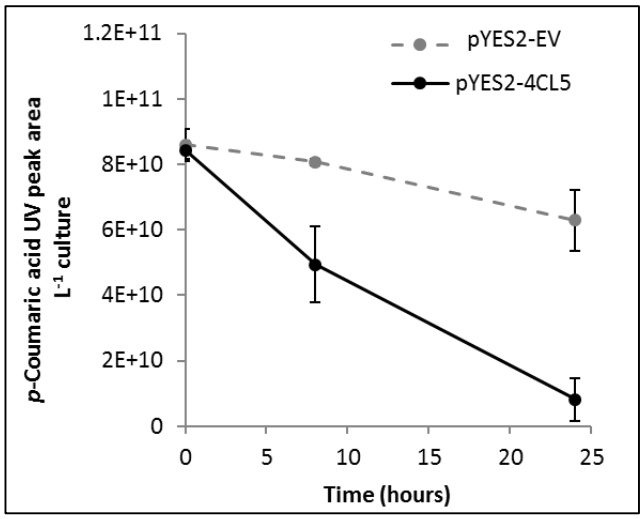


Figure 22. *In vivo p*-coumaric acid consumption by pYES2-4CL5 transformed yeast cells.
The HPLC (260 nm) peak area was used for quantification of *p*-coumaric acid, see Appendix 2.3 for *p*-coumaric acid chromatogram. The pYES2-4CL5 yeast culture consumed *p*-coumaric acid significantly faster than that transformed with pYES2-EV. The data are means ± standard deviations (n = 3).

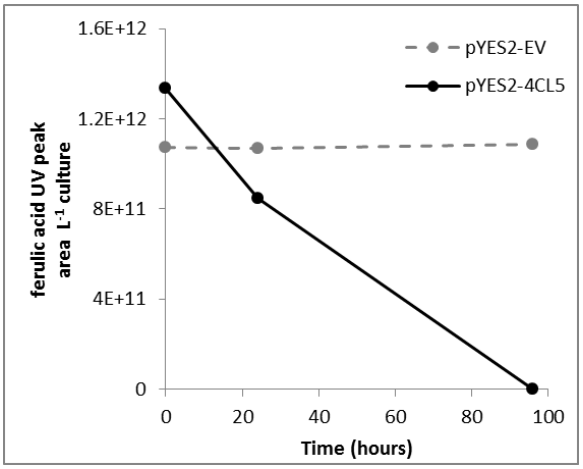


Figure 23. *In vivo* ferulic acid consumption by pYES2-4CL5 transformed yeast cells.
The HPLC (260 nm) peak area was used for quantification of ferulic acid, see Appendix 2.4 for ferulic acid chromatogram. The pYES2-4CL5 yeast culture consumed ferulic acid significantly faster than that transformed with pYES2-EV empty vector control.

The *in vivo* feeding study was then carried out with ferulic acid for comparison. Additionally, a new quadripartite analytical LC method, adapted for improved detection of aromatic CoA esters, based on Larson *et al.* [180], was used (see Materials and Methods). Like *p*-coumaric acid (see Figure 22), ferulic acid was consumed at a quicker rate by the pYES2-4CL5 expressing yeast cells compared to that of pYES2-EV (Figure 23). However, it took 96 hours for ferulic acid to be completely consumed, not ~24 hours as seen with *p*-coumaric acid. The production of feruloyl-CoA was searched for using the new LC-MS method: however, while feruloyl-CoA standard could be detected (see Appendix 2.5 for spectral data), production of feruloyl-CoA by pYES2-4CL5 yeast culture could not (Figure 24).

Despite a decrease in substrate, the CoA ester products from *in vivo* 4CL5 catalysis could not be detected by LC-MS, so *in vitro* assays were carried out. This was to ascertain whether 4CL5 was carrying out the expected catalysis in yeast or producing unexpected products. Protein was extracted from yeast cells expressing pYES2-4CL5 and pYES2-EV and assayed for activity towards cinnamic, *p*-coumaric, caffeic and ferulic acid. No phenylpropanoyl-CoA esters were produced at 0 or 20 hours by yeast transformed with pYES2-EV, as expected (Figure 25A to D). However, phenylpropanoyl-CoA esters were produced by pYES2-4CL5 yeast protein extract after 20 hours (Figure 25A to D), with parental mass ions of the CoA esters of cinnamic, *p*-coumaric, caffeic and ferulic acid detected at *m/z* 897.5, 913.5, 929.5 and 897.5, respectively (see Appendices 2.6-2.8 for spectral data).

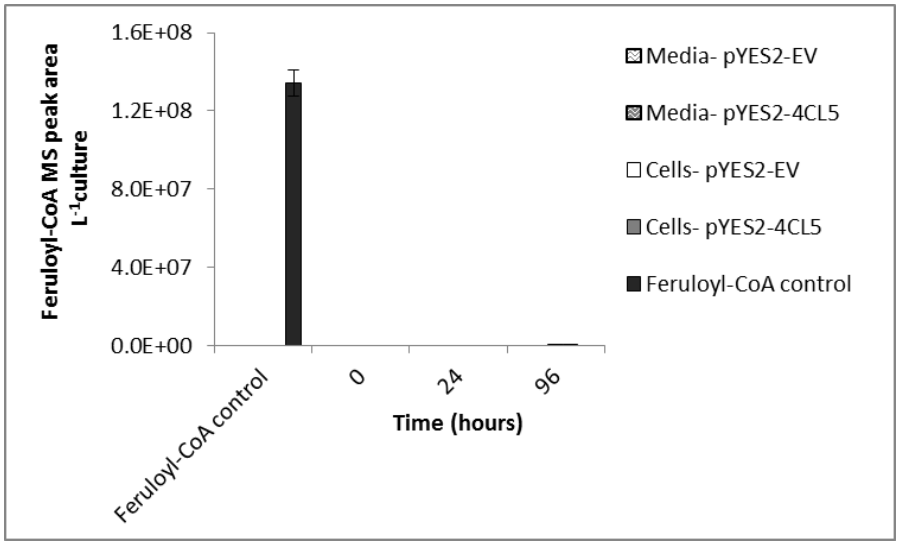


Figure 24. *In vivo* feruloyl-CoA production by yeast cells transformed with pYES2-4CL5.
The cultures were fed with ferulic acid and feruloyl-CoA production was quantified by MS. No feruloyl-CoA, could be detected using LC-MS, whether in the cells or media. Detection of the feruloyl-CoA standard shows the analytical method was functioning. The data are means ± standard deviations (n = 3).

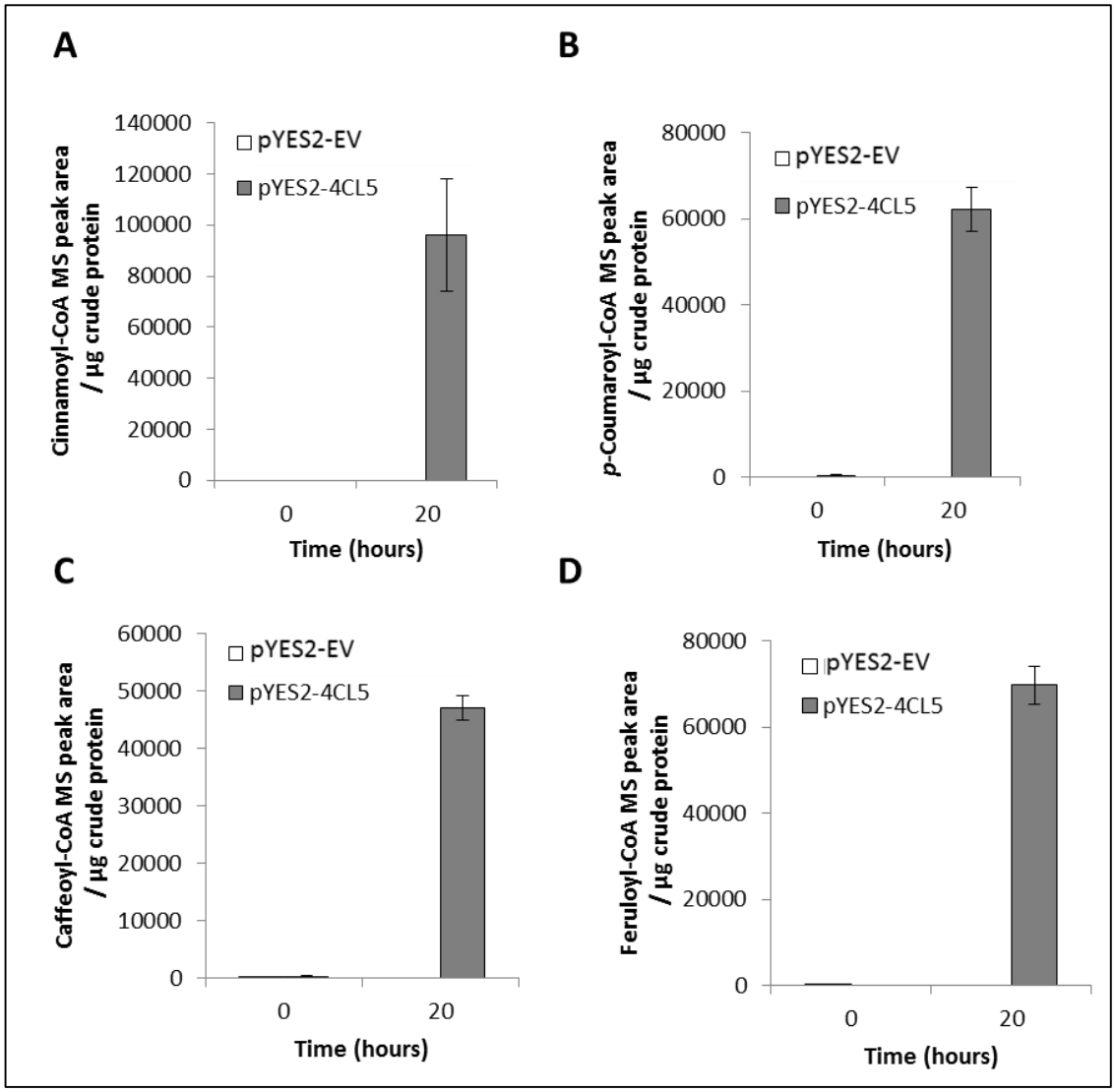


Figure 25. *In vitro* phenylpropanoyl-CoA ester production by pYES2-4CL5 yeast protein extract.
Activity of pYES2-4CL5 and pYES2-EV crude protein extract, at 0 and 20 hours, to produce phenylpropanoyl-CoA esters from cinnamic, *p*-coumaric, caffeic and ferulic acid as quantified by MS peak areas: (**A**) ($[M+H]^+$ 897.5), (**B**) *p*-coumaroyl-CoA ($[M+H]^+$ 913.5), (**C**) caffeoyl-CoA ($[M+H]^+$ 929.5) and (**D**)feruloyl-CoA ($[M+H]^+$ 943.5) respectively. See Appendices 2.6- 2.8 for spectral data. The data are means ± standard deviations (n = 3).

As *in vitro* 4CL5 activity was detected towards cinnamic, *p*-coumaric, caffeic and ferulic acid, its activity towards novel substrates, which may offer the first step in the biosynthesis of non-natural curcuminoids, was also tested (Figure 26). 4-Methoxycinnamic acid, 3,4-dimethoxycinnamic acid and 4-dimethylaminocinnamic acid were chosen because of their *para* electron donating groups, which are important for curcuminoid bioactivity [80]. 4-Fluorocinnamic acid was chosen because of the improved pharmacokinetic properties of fluorinated curcuminoids compared to curcumin [181,182].

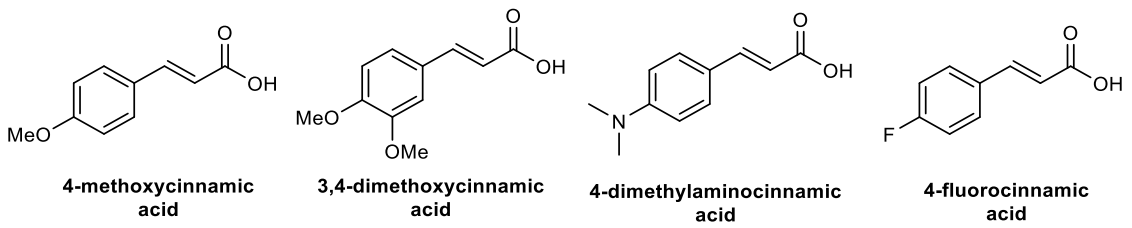


Figure 26. Non-natural phenylpropanoids.
They were assayed with yeast crude protein extract containing 4CL5, to ascertain whether 4CL5 can accept these molecules as substrates.

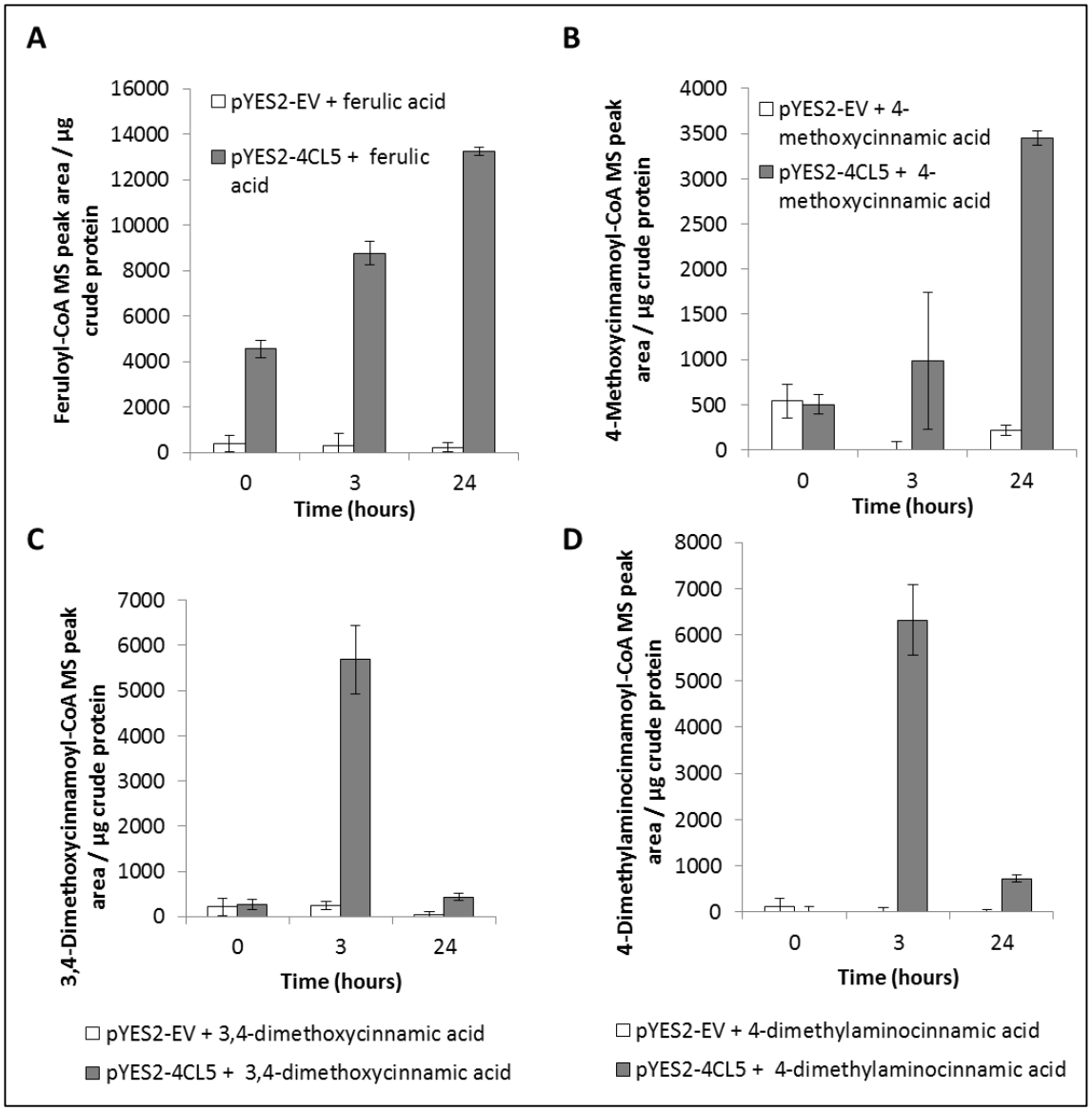


Figure 27. *In vitro* production of non-natural phenylpropanoyl-CoA esters by pYES2-4CL5 yeast protein extract.
Activity of pYES2-4CL5 and pYES2-EV crude protein extract to produce non-natural phenylpropanoyl-CoA esters from ferulic, 4- methoxycinnamic, 3,4-dimethoxycinnamic and 4-dimethylaminocinnamic acid was quantified at 0, 3 and 24 hours, using MS peak area: (**A**) feruloyl-CoA ($[M+H]^+$ 943.5), (**B**) 4- methoxycinnamoyl-CoA ($[M+H]^+$ 927.5), (**C**) 3,4-dimethoxycinnamoyl-CoA ($[M+H]^+$ 958.5) and (**D**) 4-dimethylaminocinnamoyl-CoA ($[M+H]^+$ 940.5). See Appendices 2.9- 2.11 for spectral data. The data are means ± standard deviations (n = 3).

The activity of 4CL5 towards the novel substrates was assayed *in vitro* (3 and 24 hours) in an analogous manner to the natural phenylpropanoids. Ferulic acid was used as a positive control for the assay: feruloyl-CoA was only produced in the presence of 4CL5 crude protein as expected (Figure 27A). pYES2-4CL5 crude protein catalysed the conversion of novel substrate 4-methoxycinnamic acid to 4-methoxycinnamoyl-CoA, with 4-methoxycinnamoyl-CoA increasing over time (Figure 27B). 3,4-Dimethoxycinnamic acid and 4-dimethylaminocinnamic acid were also accepted as substrates by 4CL5. However, instead of the CoA esters accumulating, as shown with feruloyl-CoA and 4-methoxycinnamoyl-CoA, the amount of CoA esters peaked at 3 hours, then decreased to virtually zero by 24 hours (Figure 27C and Figure 27D). 4-Fluorocinnamoyl-CoA could not be detected in the mass spectra when 4-fluorocinnamic acid was incubated with 4CL5 crude protein, indicating that this was the only novel substrate that 4CL5 did not accept as a substrate.

2.2.1.2 Investigating the heterologous expression of DCS

To investigate the activity of the second biosynthetic enzyme, the DCS coding sequence (AB495006.1, from *C. longa* rhizome, Appendix 1.1) was ordered from GenScript USA in the pUC57 plasmid. It was subsequently cloned into the pESC yeast expression vector to allow the translational fusion with a myc tag (Figure 28A) as verified by Sanger sequencing. *S. cerevisiae* cells were transformed with this construct and transformation was verified by PCR (Figure 28B). Transcription and expression of DCS was validated by RT-PCR and western blotting using an anti-myc primary antibody (Figure 28C and Figure 28D).

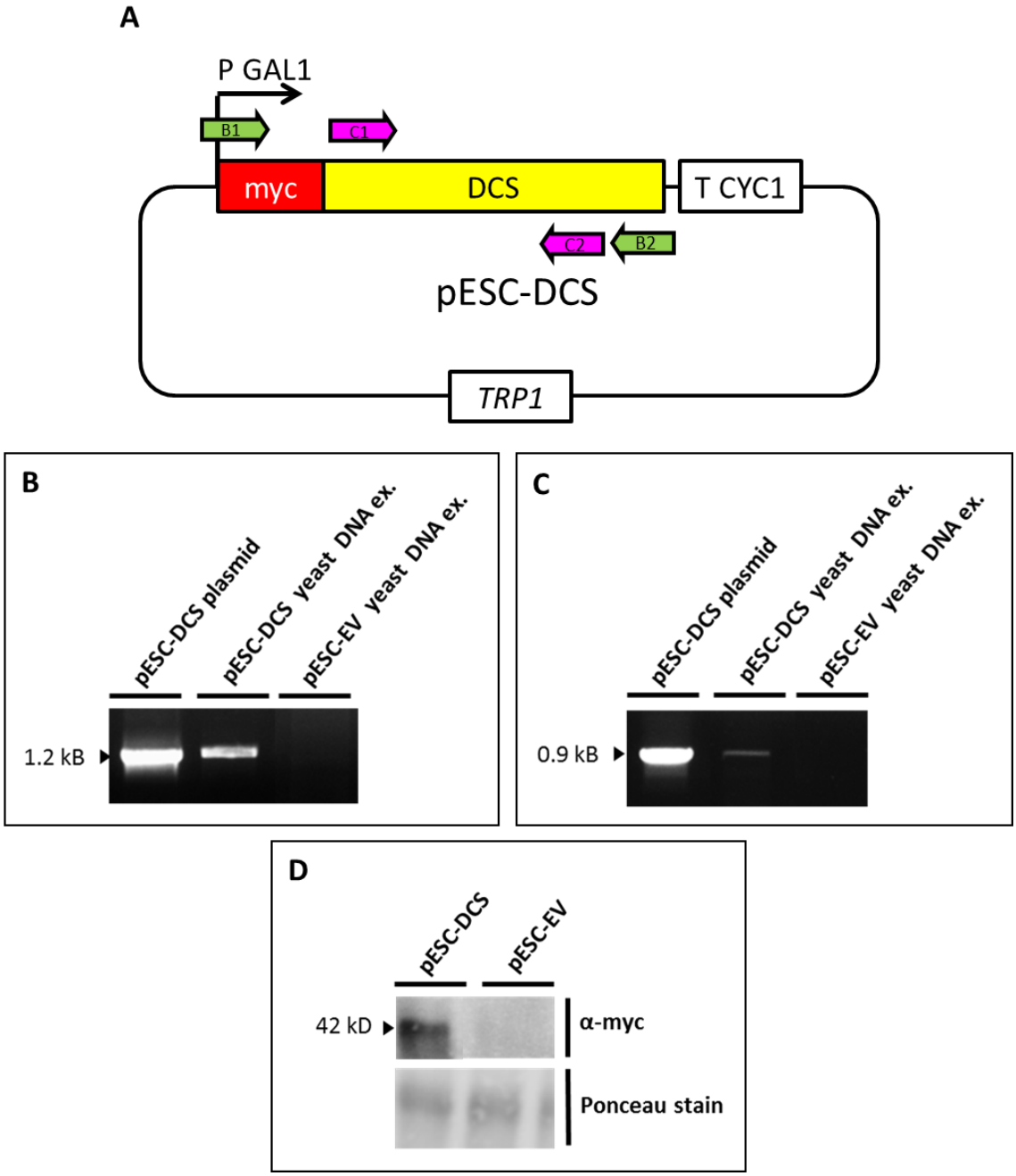


Figure 28. Cloning, transformation and expression of pESC-DCS.
(**A**) pESC-DCS plasmid, for the expression of DCS in yeast under the control of a galactose inducible promoter GAL1. CYC1 transcriptional terminator for efficient termination of mRNA; *TRP1* gene offered selection for transformants in yeast host strains with a *trp1* genotype. Primers are denoted by arrows for DNA and cDNA amplification: B1- F_Asc1_DCS; B2- R_Xba1_DCS; C1- F_DCS_160; C2- DCS_1044_R, see Materials and Methods for primers sequences. (**B**) Agarose gel electrophoresis showing diagnostic PCR products from PCR performed upon DNA extracted from transformed yeast cells, using primers B1 and B2 to determine positive transformation of yeast cells. The expected product, at 1.2 kB was amplified; pESC-DCS plasmid positive control template. (**C**) Agarose gel electrophoresis showing diagnostic PCR products from performed upon cDNA synthesized from RNA extracted from transformed yeast cells, using primers C1 and C2 to determine transcription of DCS from the pESC-DCS vector. The expected product, at 0.9 kB was amplified; pESC-DCS positive control template. (**D**) Western blot analysis using an anti-myc primary antibody to detect myc-tagged DCS protein (42 kD) in protein extracted from yeast cells transformed with the pESC-DCS plasmid. Ponceau stained membrane is shown as a loading control. pESC-EV is an empty vector control.

DCS preferentially utilises feruloyl-CoA as a substrate with which it condenses malonyl-CoA to produce diketide CoA esters [93], therefore feruloyl-CoA was chosen as the model substrate to test the activity of DCS in yeast. The activity of DCS was first tested in *in vivo* cell culture assays. 4CL5 activity was to provide feruloyl-CoA for larger *in vivo* assays as feruloyl-CoA cannot permeate the yeast cell membrane, therefore needed to be made within the cell [183]. Although the phenylpropanoyl-CoA products of 4CL5 catalysis could not be detected in *in vivo* assays, it had been shown that the enzyme was active (see Figure 25), so it was anticipated that DCS would be able to out-compete the endogenous metabolism of the CoA esters and produce feruloyl diketide-CoA. Yeast cell cultures expressing pYES2-4CL5 alone, pESC-DCS alone and co-expressing both pYES2-4CL5 and pESC-DCS were fed ferulic acid (*S. cerevisiae* naturally produces malonyl-CoA so it was not supplemented) [184]. Samples of the cells and medium were taken after 24 and 96 hours and analysed by LC-MS to determine whether feruloyl diketide-CoA had been formed by the culture co-expressing pYES2-4CL5 and pESC-DCS. The positive molecular ion, $[M+H]^+$ 986.7 (feruloyl diketide-CoA M_r = 985.7 g mol^{-1}), was searched for, however, there was no presence of this ion. A decrease in ferulic acid over time, compared to pESC-EV, suggested that 4CL5, whether expressed alone or in combination with DCS, was consuming substrate and therefore likely to be producing feruloyl-CoA as a substrate for DCS (Figure 29). The depletion of ferulic acid was not significantly different between the culture expressing pYES2-4CL5 alone or pESC-DCS and pYES2-4CL5 in combination. As the phenylpropanoyl-CoA ester produced by 4CL5 is believed to be further metabolised by the yeast cells, it may be that the recombinant DCS protein cannot compete with the endogenous pathway. Equally, as ferulyol-CoA production by 4CL5 could not be detected in *in vivo* assays (see Figure 24), it is possible that the product of DCS, feruloyl diketide-CoA, may also be undetectable.

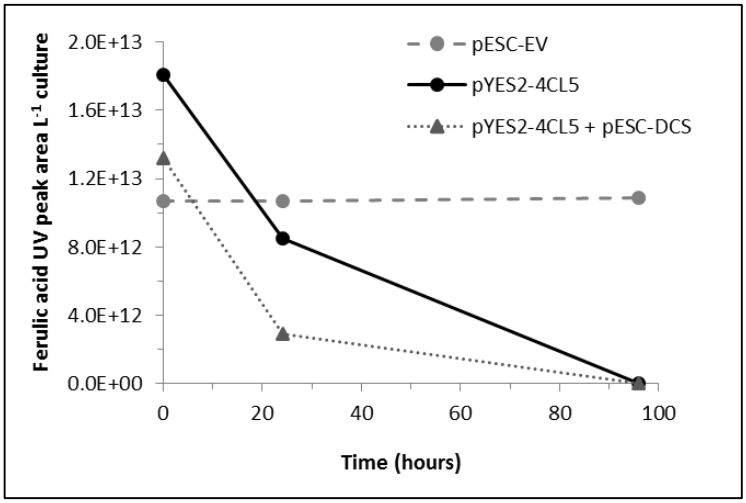


Figure 29. Ferulic acid consumption by pYES-4CL5 and pESC-DCS transformed yeast cells.
Ferulic acid in the culture medium was quantified by HPLC UV (260 nm) peak area. pESC-EV yeast culture showed no consumption of ferulic acid, whereas those transformed with pYES2-4CL5 and both pYES2-4CL5 and pESC-DCS did, at similar rates.

To investigate DCS activity further, and establish if a lack of substrate availability was indeed an issue, an *in vitro* assay was carried out. CoA esters are costly to purchase so efforts were made to synthesize feruloyl-CoA using 1,1'-carbinoyldiimidazole and CoA trilithium salt [185]. Although synthesis of feruloyl-CoA was successful, purification was problematic so feruloyl-CoA was finally purchased for use *in vitro* assays in which lower quantities of substrate are required. Crude protein was extracted from pESC-DCS transformed yeast cells and assayed for DCS activity by incubation with feruloyl-CoA and malonyl-CoA as described by Katsuyama *et al.*[93] (see Materials and Methods). LC-MS was used to determine whether DCS had condensed feruloyl-CoA and malonyl-CoA by searching for the molecular ion $[M+H]^+$ 986.7 (feruloyl diketide- CoA M_r = 985.7 g mol^{-1}): alike to *in vivo* assays, there was no presence of this ion. Although, with no available molecular standard, it cannot be concluded with 100 % certainty. Katsuyama *et al.* [93] reported that DCS was also able to produce derailment products benzalacetone and dehydrozingerone. Evidence for these compounds was looked for in the mass spectra data, without success. To confirm that DCS, or endogenous yeast enzymes, were not converting feruloyl-CoA into something else, the amount of this starting material was quantified by mass spectrometry (Figure 30). There was no discernible difference in the amount of feruloyl-CoA detected after 1 hour incubation with crude protein extract from yeast cells transformed with pESC-EV or pESC-DCS, inferring that feruloyl-CoA was not being consumed by DCS.

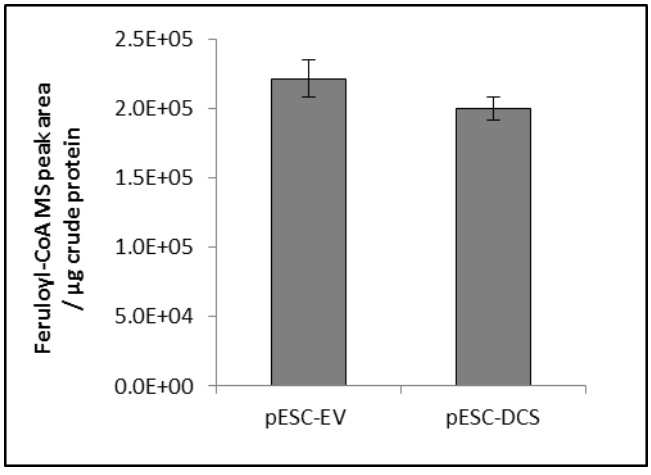


Figure 30. *In vitro* feruloyl-CoA consumption by pESC-DCS yeast protein extract.
After 1 hour incubation with crude protein extract from yeast cells transformed with pESC-EV (empty vector control) or pESC-DCS there is no significant difference between the amount of feruloyl-CoA remaining in the empty vector control compared with the pESC-DCS protein extract. Feruloyl-CoA consumption was quantified by MS peak area. The data are means ± standard deviations (n = 3).

Addressing the codon bias between *S. cerevisiae* and DCS

In an attempt to troubleshoot the lack of diketide-CoA products produced by DCS protein in *S. cerevisiae*, a focussed collaboration with the research teams led by Prof Nia Bryant and Prof Bob White in the Biology Department at the University of York was initiated. Prof Nia Bryant and Prof Bob Whites' research focuses upon heterologous protein expression in yeast and tRNAs, respectively. To heterologously produce a recombinant protein, the nucleotide coding sequence must be able to be 'read' by the host's translational machinery and matched with complementary tRNAs carrying the correct synonymous amino acids. However, the codon usage varies greatly between organisms and even between differentially expressed proteins in the same organism [186]. As an organism's codon bias correlates with the abundance of cognate tRNAs [187], differential codon bias between a heterologous gene and a host can have a profound effect upon heterologous protein production [188]. When synthesized by GenScript, the nucleotide sequence of DCS was not codon optimised; this was so it could be heterologously expressed in both *S. cerevisiae* and tobacco. Additionally, it has been shown that although codon optimisation can increase translation, it can alter mRNA secondary structure, therefore stability, and overall decrease gene expression and protein production [189].

A possible rationale for the lack of detectable DCS activity was hypothesized to be that the codon usage of *S. cerevisiae* is not optimal for correct DCS expression. Not only does a mismatch in codon bias reduce protein quantity, but also quality, as the incidence of frame shifting and misincorporation of amino acids increases during translation [188]. DCS has previously been heterologously expressed by Katsuyama's group; this was in *E. coli* (BL21) and the gene was not codon-optimised [93], therefore, the main difference to the work described in this project is the heterologous host.

Although DCS protein was produced in yeast expressing pESC-DCS (see Figure 28D), this could be due to codon wobble allowing non-synonymous tRNAs to translate heterologous codons for which there were no, or little, endogenous tRNAs [190]. It was hypothesized that creating a yeast strain expressing tRNAs to resolve the issue of codon bias with DCS, would allow the quality and quantity of heterologous DCS protein to increase [188]. To test this hypothesis, the number of copies of each tRNA in *S. cerevisiae* was compared to the codon bias of DCS, as well as the functionally active 4CL5 (Figure 31). Arabidopsis 4CL5 was expressed and functionally active in yeast (see Figure 21C and Figure 25) without codon optimisation. Which may be due to Arabidopsis and *S. cerevisiae* sharing a similar codon usage space [186]. Therefore, the codon usage of 4CL5 was used to further inform which codons may be impeding DCS expression and activity. Codons which were of a higher abundance in DCS than 4CL5 and had little or no yeast tRNAs were searched for. Codons which met these criteria were considered as potential targets for increasing DCS

expression. Of the 9 codons identified (CGC, CTC, GCC, GCG, CTG, GTC, TAT, CCC and CGG) 3 codons were chosen which were anticipated to be beneficial for DCS expression: CTG, CTC and GCG (see Figure 31). Of the 390 DCS codons, 60 codons were one of the three chosen codons which either had no (CTC, GCG) or only had one (CTG) corresponding *S. cerevisiae* tRNA. The chosen codons were shown to be spread throughout the DCS coding sequence (Figure 32). Although none of the codons correspond to the four crucial active site residues (highlighted in green), they are in close proximity to these sites so may cause frame slipping and incorrect base incorporations at these crucial positions.

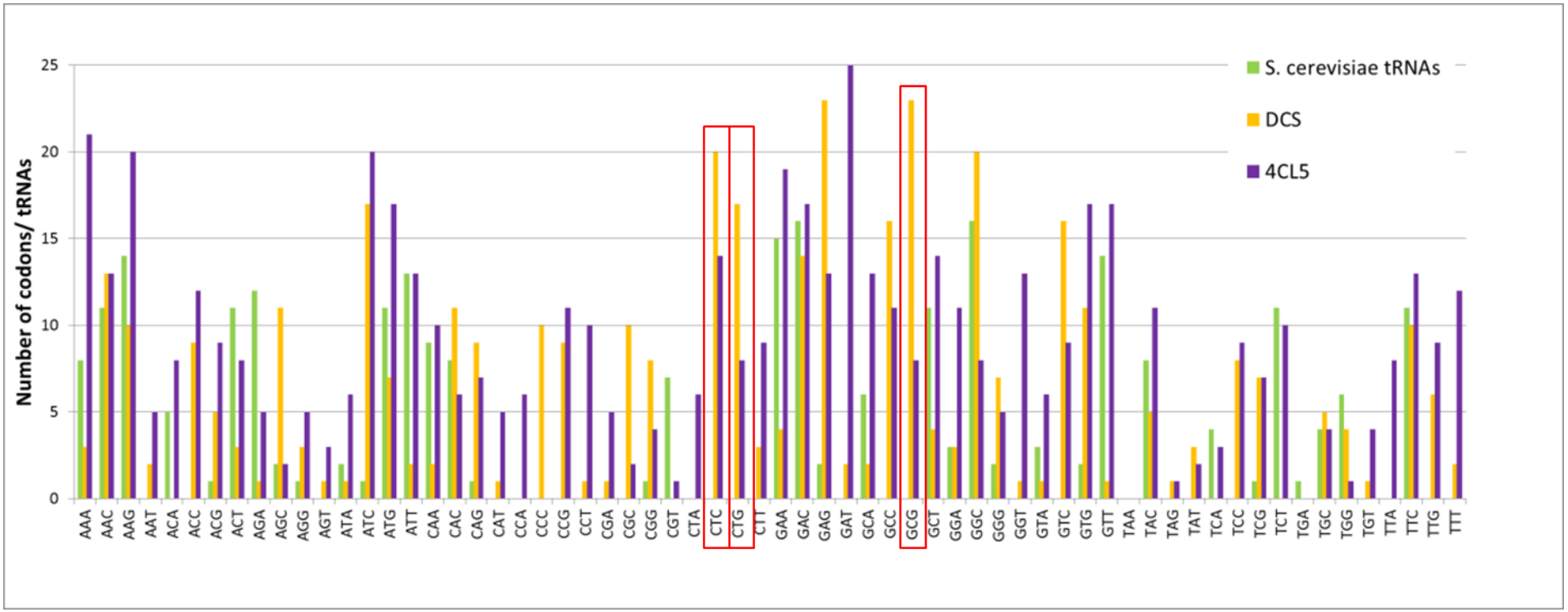


Figure 31. Comparison of *S. cerevisiae* tRNA bias and the codon usage of *C. longa* DCS and Arabidopsis 4CL5.
Highlighted in red boxes are three codons that were chosen for optimisation. They have no (CTC, GCG) or one (CTG) *S. cerevisiae* tRNA(s), and DCS contains more of that particular codon than 4CL5.

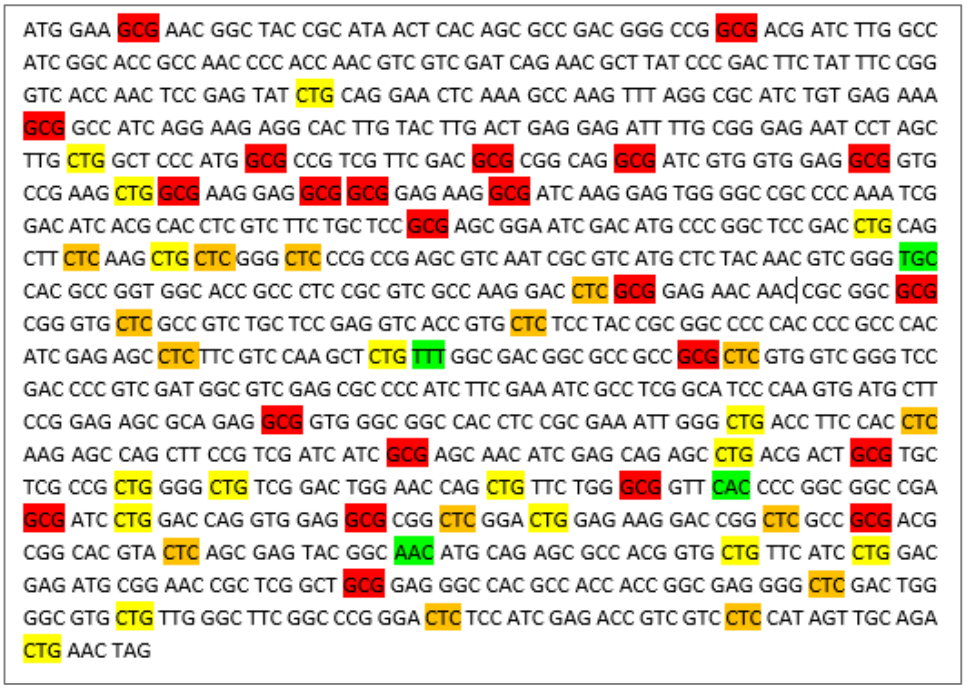


Figure 32. DCS coding sequence showing codons that have no/low number of tRNAs in *S. cerevisiae*. GCG- highlighted in red, CTC- highlighted in orange, CTG- highlighted yellow, highlighted in green are the codons that correspond to the residues in the DCS active site: TGC- Cys 163, TTT- Phe 214, CAC- His 302 and AAC- Asn 335 [191].

To rebalance the codon bias between *S. cerevisiae* and DCS, tRNAs for the three chosen codons, from either *Schizsaccharomyces pombe* or *S. cerevisiae*, were identified (Table 1). The three chosen tRNA nucleotide sequences were identified in their respective chromosomes with approximately 250 bp of flanking sequence to help sub-cloning (Appendix 2.12). The three tRNA coding sequences were synthesized by Invitrogen in the pMA vector, then subsequently cloned into yeast expression vectors pRS416 and pRS426. The vectors pRS416 and pRS426 were chosen due to being low copy centromeric and high copy episomal plasmid copies of one another. A high and a low copy plasmid were used for two reasons. One, tRNA expression in eukaryotes is determined by copy number [186], so a high copy plasmid was theorised to give greater tRNA expression. Two, a low copy plasmid was used in case high copy expression of the tRNAs over-loaded the yeast cell's tRNA capacity. The tRNA nucleotide sequences were cloned individually and all three together into the pRS416 and the pRS426 vectors using restriction and ligation (Figure 33). Cloning with the low copy pRS416 plasmid was carried out as part of this PhD project; cloning with the high copy pRS426 plasmid was carried out by Dr Aga Urbanek (Prof Nia Bryant's research group, University of York). Once correct cloning into both plasmid types was verified by Sanger sequencing, all eight vectors were co-transformed into *S. cerevisiae* cells with pESC-DCS as well as the corresponding empty vector controls.

Table 1. tRNAs chosen to be expressed in *S. cerevisiae.*

Codon	Abundance in DCS	Number tRNAs in *S. cerevisiae*	Amino acid	Substitute tRNA
CTG	17	1	Leucine	*Schizsaccharomyces pombe* (Leu1)
CTC	20	0	Leucine	*Saccharomyces cerevisiae* (Leu2)
GCG	23	0	Alanine	*Schizsaccharomyces pombe* (Ala)

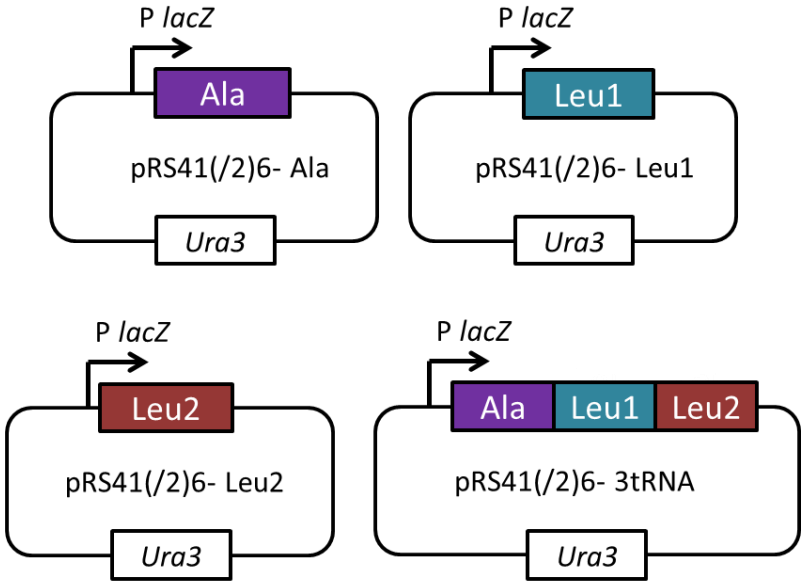


Figure 33. Yeast pRS416 and pRS426 tNRA expression vectors.
The three chosen tRNAs were cloned individually and together into pRS416 and pRS426 vectors. pRS416 is a low copy centromeric plasmid, pRS426 is a high copy episomal plasmid. Ala- *Schizsaccharomyces pombe*, alanine; Leu1- *Schizsaccharomyces pombe*, leucine; Leu2- *Saccharomyces cerevisiae*, leucine.

Yeast cells, transformed with pESC-DCS and each of the tRNA vectors, were induced by the addition of 2 % galactose. The heterologous tRNAs did not require induction as eukaryotic tRNA expression is controlled by two intragenic regions [192,193]. Protein extraction was performed and the quantity of DCS was analysed by western blotting (Figure 34). The western blot analysis of yeast cells transformed with the pRS416 set of plasmids was inconclusive as there was a contaminating band at the size DCS protein is expected, 42 kD, so only results from the pRS426 plasmids is shown. DCS was expressed at low levels from pESC-DCS. Expression of all three tRNAs increased the amount of DCS expressed (see Figure 34). The addition of *S. pombe* leucine tRNA (Leu1) alone showed no effect upon the quantity of DCS protein expressed, whereas *S. cerevisiae* leucine tRNA (Leu2) increased the amount of DCS expressed. This was evident when expressed alone or in combination with the other tRNAs (see Figure 34). Individually, the effect of *S. pombe* alanine (Ala) could not be tested due to problems transforming yeast cells with the

pRS426-Ala vector. It was assumed *S. pombe* alanine tRNA did not have an effect upon DCS expression, as DCS expression was equivalent when expressed alongside both pRS426-Leu2 and pRS426-3tRNA. This suggests that only Leu2 allows an increase in DCS expression.

As the quantity of DCS protein could be increased with the co-expression of *S. cerevisiae* leucine tRNA (pRS426-Leu2), the effect of *S. cerevisiae* leucine tRNA upon DCS protein quality was tested by carrying out an enzymatic assay. It was hypothesized that, in addition to allowing an increase in the quantity of protein expression, Leu2 may also allow a crucial improvement in protein quality allowing DCS enzyme activity to be detected. Crude protein was extracted from yeast transformed with pESC-DCS alone, or co-transformed with pRS426-Leu2, or pRS426-3tRNA. The protein was then incubated with feruloyl-CoA and malonyl-CoA and assayed for the production of feruloyl diketide-CoA. As was the case with previous DCS performance in *in vitro* assays, no desired product could be detected by LC-MS. To ascertain the availability of feruloyl-CoA and investigate the possibility that DCS may be converting the substrate into unexpected products, feruloyl-CoA was quantified by MS peak area (Figure 35). The addition of pRS426-Leu2 had no observable effect upon DCS activity, inferred from no consumption of feruloyl-CoA, whether expressed alone (pRS426-Leu2) or in combination with Leu1 and Ala (pRS426-3tRNA). This observation lead to the conclusion that the lack of detectable DCS activity is not likely to be due to a codon mismatch between DCS and the *S. cerevisiae* heterologous host.

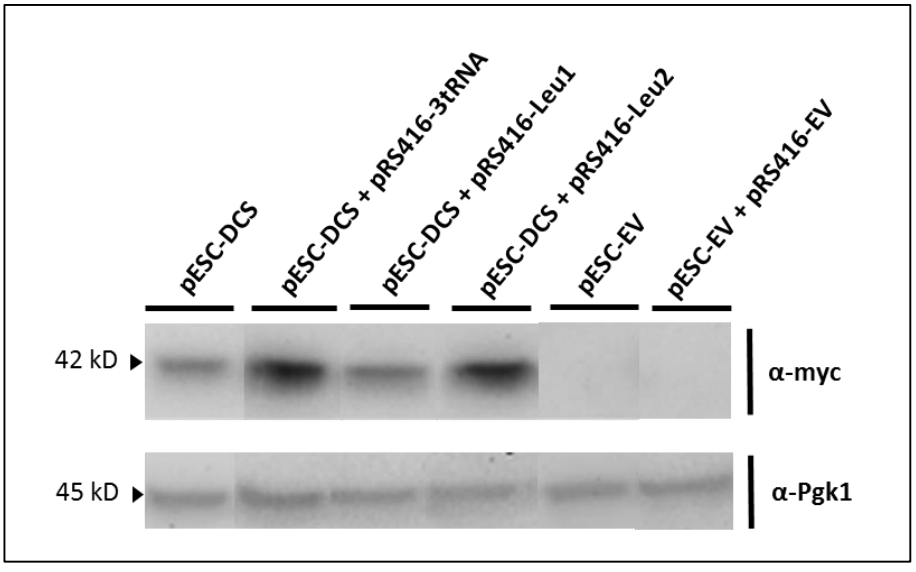


Figure 34. Western blot analysis of DCS protein expression.
Western blot analysis using anti-myc primary antibody (α-myc) to detect myc-tagged DCS protein expected at 42 kD. DCS was cloned into the pESC vector such that it was myc-tagged. DCS is expressed in low levels from pESC-DCS. DCS protein increases when expressed in yeast cells co-transformed with pRS426-3tRNA. This appears to be due to Leu2 and not Leu1. Anti-3-phosphoglycerte kinase 1 (α-Pgk1) detection was used as a loading control. Western blot analysis courtesy of Dr Aga Urbanek, Bryant research group, University of York.

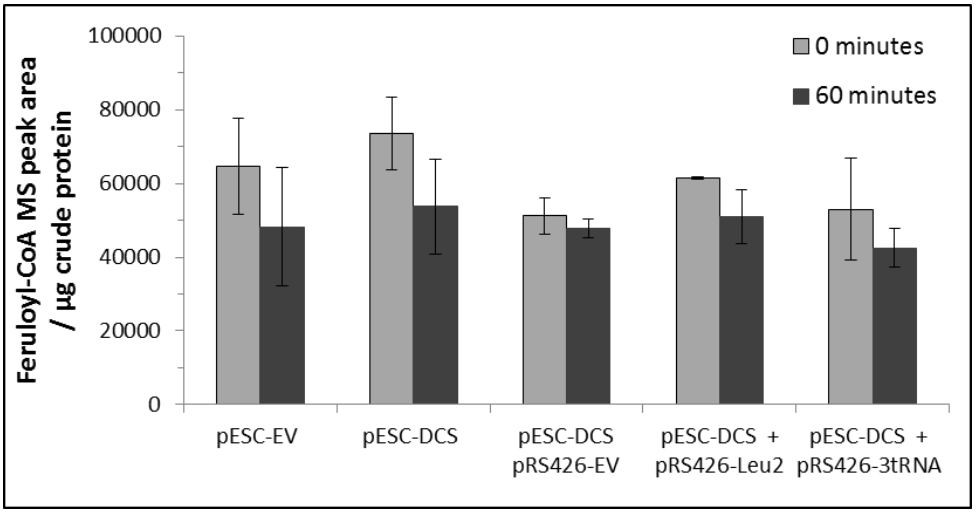


Figure 35. Feruloyl-CoA consumption by protein extracted from yeast cells transformed with pESC-DCS and supplemental tRNAs.
MS peak area quantification of feruloyl-CoA after 60 minutes incubation with malonyl-CoA and crude protein extract from yeast cells transformed with pESC-EV, pESC-DCS, pESC-DCS and pRS426-EV, pESC-DCS and pRS426- Leu2, pESC-DCS and pRS426-3tRNA. There is no significant difference between the amount of feruloyl-CoA remaining in the pESC-EV empty vector control assay compared to that of pESC-DCS. The addition of pRS426-Leu2 has no effect upon DCS activity, measured by consumption of feruloyl-CoA, whether expressed alone (pRS426-Leu2) or in combination with Leu1 and Ala (pRS426-3tRNA). The data are means ± standard deviations (n = 3).

2.2.1.3 Investigating the heterologous expression of CURS1

CURS1 (AB495007.1, Appendix 1.2, from *C. longa* rhizome tissue) catalyses the condensation of a phenylpropanoyl-CoA ester with a diketide-CoA ester to produce curcuminoids (see Figure 6). Due to the diketide-CoA ester substrate of CURS1 being unavailable commercially and very difficult to synthesize chemically, assaying the activity of CURS1 relied upon the successful expression and activity of 4CL5 and DCS to produce the substrate.

To investigate the activity of the third curcuminoid biosynthetic enzyme, CURS1 was synthesized by GenScript USA in the pUC57 plasmid. It was subsequently cloned into the pESC-DCS dual expression vector where it was translationally fused to a FLAG tag (Figure 36A). Yeast cells were transformed with this construct as verified by PCR, showing the expected product size of 1.2 kB (Figure 36B). Transcription of CURS1 was determined by RT-PCR on RNA extracted from yeast cells induced to express CURS1 from pESC-DCS-CURS1. PCR upon the cDNA derived from reverse transcription of the RNA gave a product at the expected size of 1.0 kB (Figure 36C). However, CURS1 protein could not be detected by western blotting of protein extracts from the same yeast cells (Figure 36D). Furthermore, when *in vitro* assays were performed using protein from yeast transformed with pYES2-4CL5 and pESC-DCS-CURS1, the expected curcumin product could not be detected (Figure 37).

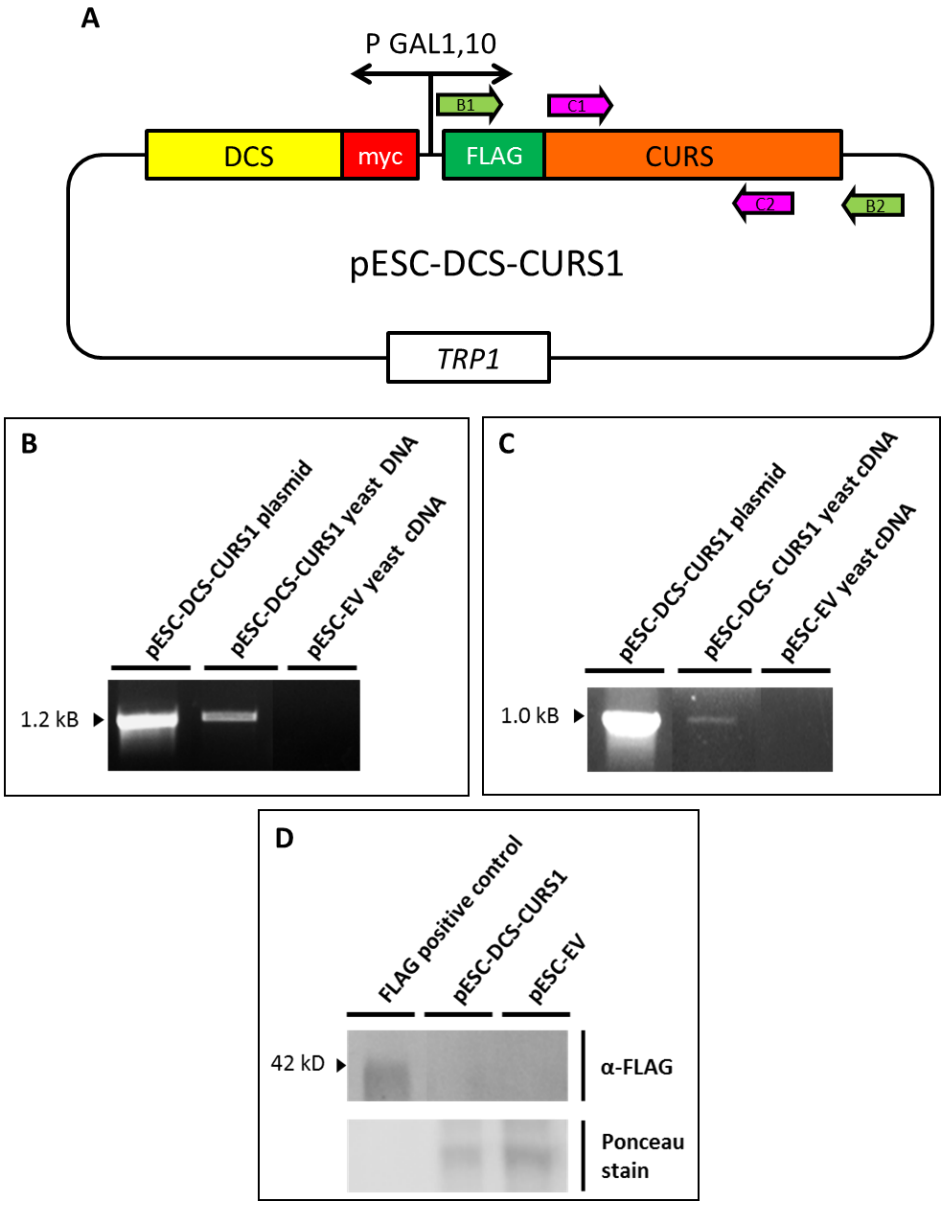


Figure 36. Cloning, transformation and expression of pESC-DCS-CURS1.
(**A**) pESC-DCS -CURS1 plasmid, for the expression of CURS1 in yeast under the control of the GAL10 galactose inducible promoter. CYC1-transcriptional terminator for efficient termination; *TRP1*- selection for transformants in yeast host strains with a *trp1* genotype. Primers are denoted by arrows for DNA and cDNA amplification: B1- F_Asc1_CURS1; B2- GAL10R; C1- CURS1_F; C2- CURS1048_R, see Materials and Methods for primer sequences; (**B**) Agarose gel electrophoresis showing diagnostic PCR products from PCR performed upon DNA extracted from transformed yeast, using primers B1 and B2, to determine positive transformation of yeast cells. The PCR product was at the expected size of 1.2 kB; pESC-DCS-CURS1 plasmid was used as a positive control template (see Materials and Methods for primers); (**C**) Agarose gel electrophoresis showing diagnostic PCR products from PCR performed upon cDNA synthesized from RNA extracted from transformed yeast cells, using priemrs C1 and C2 to determine transcription of CURS1 from the pESC-DCS-CURS1 vector. The PCR product was at the expected size of 1.0 kB; pESC-DCS-CURS1 plasmid was used as a positive control template (see Materials and Methods for primers); (**D**) Western blot analysis using an anti-FLAG primary antibody to detect FLAG-tagged CURS1 protein (42 kD) in protein extracted from yeast cells transformed with the pESC-DCS -CURS1 plasmid. Ponceau stained membrane is shown as a loading control. pESC-EV is an empty vector control. The FLAG-tagged control protein (AbCam Positive Control Lysate) is evident (41 kD) when western blot analysis using anti-FLAG is used, whereas CURS1 is not, indicating no CURS1 protein expression.

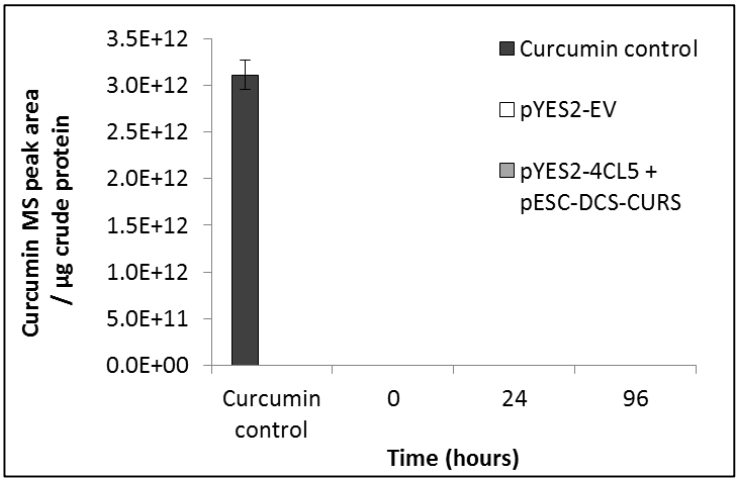


Figure 37. Curcumin production by yeast protein extracted from pYES2-4CL5 and pESC-DCS-CURS1 transformed cells.
Quantification by MS of curcumin produced *in vitro* by pYES2-EV and pYES2-4CL5 + pESC-DCS-CURS1 protein extract fed ferulic acid. See Appendix 2.13 for curcumin mass spectrum. No curcumin production could be detected using LC-MS although detection of the curcumin standard shows the analytical method was functioning. The data are means ± standard deviations (n = 3).

2.2.1.4 Expression of curcuminoid biosynthetic enzymes mediated by the 2A polypeptide

The 2A polyprotein is found in the cardio- and aphtho- (or foot and mouth disease virus; FMDV) genera of the picornavirus family. It is a 19 amino acid long sequence that causes 'ribosome skipping' giving multiple translational products from one open reading frame [194]. Since its discovery, it has proven to be active in all eukaryotic systems tested [195,196] and has been used to a multitude of ends, from the stoichiometric production of therapeutic proteins to reconstituting biosynthetic pathways [197–199]. The most commonly used 2A sequence is from FMDV (F2A) and it has been successfully used in Bakers' yeast [175,200]. The ability of the F2A polypeptide to produce stoichiometric amounts of protein was to be utilised in this study to create a tuneable expression system. In tandem with investigating the activity of 4CL5, DCS and CURS1 when expressed individually, work was carried out into the expression and activity of these enzymes when linked by the F2A polypeptide in one ORF (pYES2-4CL5-F2A-DCS-F2A-CURS1) (Figure 38). If one of the enzymes proved rate limiting, additional copies of that gene could be added to the F2A mediated construct to increase the amount of enzyme translated and in turn increase metabolic flux. To allow the pathway to be broken down, DCS and CURS1 were also linked by the F2A polypeptide (pYES2-DCS-F2A-CURS1) (Figure 38).

Transformation of yeast cells with pYES2-4CL5-F2A-DCS-F2A-CURS1 (pYES2-4DC) was verified by PCR upon DNA extracted from yeast colonies (Figure 38B). Primers were used for amplification of each of the genes giving products of the expected sizes: 4CL5 at 1.7 kD; DCS at

1.2 kD and CURS1 at 1.2 kD. Western blot analysis of yeast transformed with pYES2-4DC was unsuccessful due to contaminating bands at the expected sizes of the heterologously expressed proteins. However, western blot analysis of yeast cells transformed with pYES2-DCS-F2A-CURS1 was successful (Figure 38C). When expressed from pYES2-DCS, DCS appears at the expected molecular weight, 42 kD. When expressed from pYES2-DCS-F2A-CURS1, DCS appears at ~ 47 kD, this is thought to be due to the addition of F2A (2.3 kD). The majority of the DCS protein appears to be cleaved from CURS1 as there is negligible protein detectable at 85 kD (the molecular weight of DCS-F2A-CURS1 protein). Therefore, it was concluded that the F2A polypeptide, linking DCS and CURS1, performed as expected, with the majority of the protein present in the cleaved form.

Although western blot analysis was inconclusive, yeast cell culture transformed with pYES2-4DC was assayed for its ability to convert ferulic acid to curcumin (Figure 39). pYES2-DCS-CURS1 was not investigated as pYES2-4DC allowed the assay to be conducted with readily available ferulic acid which could be CoA esterified by the 4CL5. As seen with the individually expressed enzymes, it was only possible to detect activity from the 4CL5 enzyme: yeast transformed with pYES2-4DC consumed ferulic acid faster than the pYES2-EV (see Figure 39). Although pYES2-4DC appeared to consume ferulic acid faster than 4CL5 alone, no curcumin or feruloyl-CoA or feruloyl-diketide-CoA intermediates could be detected by LC-MS.

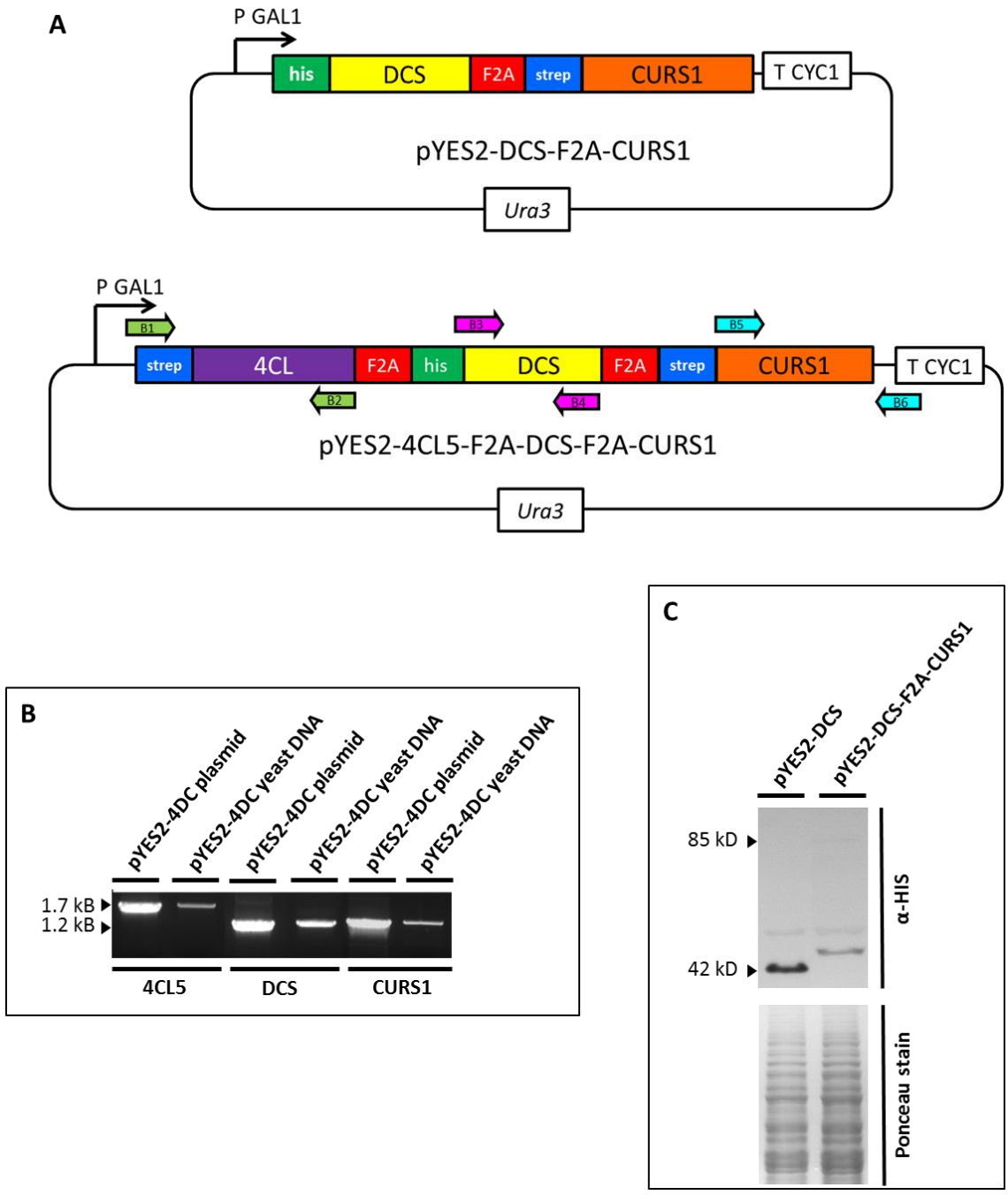


Figure 38. Cloning, transformation and expression of pYES2-DCS-F2A-CURS1 and pYES2-4CL5-F2A-DCS-F2A-CURS1.
(**A**) pYES2-DCS-F2A-CURS1 and pYES2-4CL5-F2A-DCS-F2A-CURS1 plasmids for the 2A polyprotein mediated expression of DCS, CURS1 and 4CL5 in yeast under the control of the GAL1 galactose inducible promoter. CYC1-transcriptional terminator for efficient termination; *TRP1*- selection for transformants in yeast host strains with a *trp1* genotype. Primers are denoted by arrows for DNA and cDNA amplification: B1- F_Asc1_4CL5; B2- R_XbaI_4CL5; B3- F_Asc1_DCS; B4- R_XbaI_DCS; B5- CURS1_F; B6- pYESrseq, see Materials and Methods for primer sequences. (**B**) 0.8 % Agarose gel showing diagnostic PCR products verifying yeast transformation with pYES2-4CL5-F2A-DCS-F2A-CURS1 (pYES2-4DC) using primers B1 and B2 for 4CL5 detection, B3 and B4 for DCS detection and B5 and B6 for CURS1 amplification. PCR carried out upon pYES2-4CL5-F2A-DCS-F2A-CURS1 plasmid control was used as a control. (**C**) Western blot using anti-HIS primary antibody to visualise F2A polypeptide cleavage of protein expressed from pYES2-DCS-F2A-CURS1. Excellent cleavage is shown with the majority of DCS detected as single DCS protein (42 kD) and not linked to CURS1 (85 kD). DCS protein from yeast expressing pYES2-DCS-F2A-CURS1, appears at a slightly higher molecular weight than expected due to the presence of the F2A linker. The band at approximately 50 kD, in both lanes, was not identified. Ponceau stain loading control is shown below.

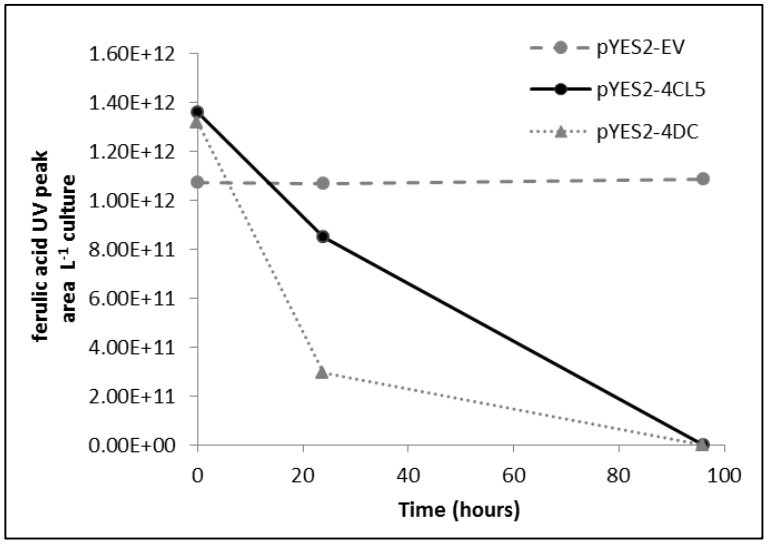


Figure 39. Ferulic acid consumption by pYES2-4CL5-F2A-DCS-F2A-CURS1 yeast cells.
Ferulic acid was quantified by HPLC UV (260 nm) in the culture medium. pYES2-EV yeast cells showed no consumption of ferulic acid, whereas, over 96 hours, those transformed with pYES2-4CL5 and pYES2-4DC (pYES2-4CL5-F2A-DCS-F2A-CURS1) did.

2.2.2 Investigating curcuminoid biosynthesis in *Nicotiana benthamiana*

With their ability to produce organic compounds from sunlight and water via photosynthesis, plants are able to function as sustainable chassis organisms for the recombinant production of secondary metabolites. Plant metabolic pathways occur in multiple subcellular locations, exemplified by the monoterpene indole alkaloid biosynthesis pathway which takes place in five different subcellular compartments [201]. The idea of compartmentalising biosynthetic reactions was successfully used to produce the cyanogenic glycoside dhurrin in *N. benthamiana* chloroplasts [202]. Therefore, it was hypothesized that the curcuminoid biosynthetic enzymes may function as expected in a more 'natural' plant host such as tobacco, compared to *S. cerevisiae*. To test this, *A. tumefaciens* mediated transient expression of the three curcuminoid biosynthetic enzymes was planned.

The fate of curcuminoid related compounds in tobacco

Before carrying out assays with the heterologous enzymes, the fate of the model phenylpropanoid starting material, *p*-coumaric acid, and curcuminoid product, curcumin, was investigated. For curcumin biosynthesis in tobacco to be optimal the curcuminoid biosynthetic pathway would need to be orthologous, with both *p*-coumaric acid and curcumin not being metabolised by tobacco tissue allowing accumulation of the curcumin product.

First, the fate of the model phenylpropanoid starting material, *p*-coumaric acid was assayed. Leaf discs were cut from tobacco leaves and incubated in 1.0 mM and 2.5 mM *p*-coumaric acid solutions. A leaf disc assay was carried out so that the leaf tissue could be easily supplied substrates (see Materials and Methods). After 24 hours, the leaf discs incubated in the 2.5 mM *p*-coumaric acid solution began to turn yellow and senesced, whereas those in the 1.0 mM p-coumaric acid solution remained green and healthy until 48 hours, as monitored by visual inspection. Hence, all subsequent assays using *p*-coumaric acid were carried out in 1.0 mM solutions for 48 hours. As monitored by reverse phase HPLC over 48 hours, the tobacco tissue caused a decrease in *p*-coumaric acid shown by a decrease in the UV (250 nm) peak area (Figure 40). The appearance of new peaks in the HPLC chromatogram at 13.9 and 15.3 minutes, after 24 and 48 hours incubation with 1.0 mM *p*-coumaric acid, but not in the control lead to the conclusion that *p*-coumaric acid was being metabolised or causing the synthesis of new metabolites (Figure 41). Looking at the MS negative ion analysis, the *m/z* of these peaks was 325.2 and 976.2 respectively. Without further analysis it is postulated that the peak with *m/z* 325.2 is the glucoside of *p*-coumaric acid and the peak with *m/z* 976.2 is composed of three *p*-coumaric glucosides. As neither of these chemicals showed a fragmentation ion corresponding to *p*-coumaric acid, *m/z* 163.0, this suggests the peaks are new chemicals synthesized due to the presence of *p*-coumaric acid, not chemical conjugations or modifications of *p*-coumaric acid itself.

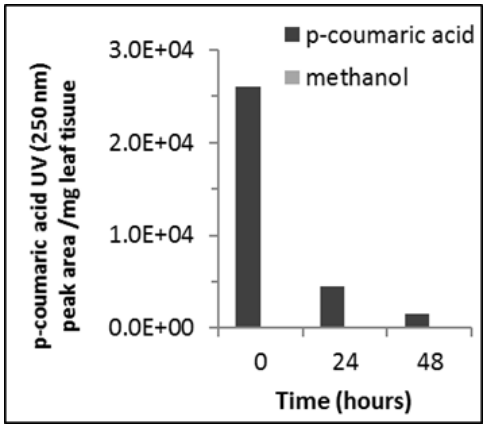


Figure 40. *p*-Coumaric acid decreased when incubated with tobacco leaf tissue.
p-coumaric acid was quantified by HPLC UV (250 nm) peak area at 0, 24 and 48 hours; the most significant decrease was shown over the first 24 hours. Data is from the combined analysis of 10 leaf discs, from 3 individual plants.

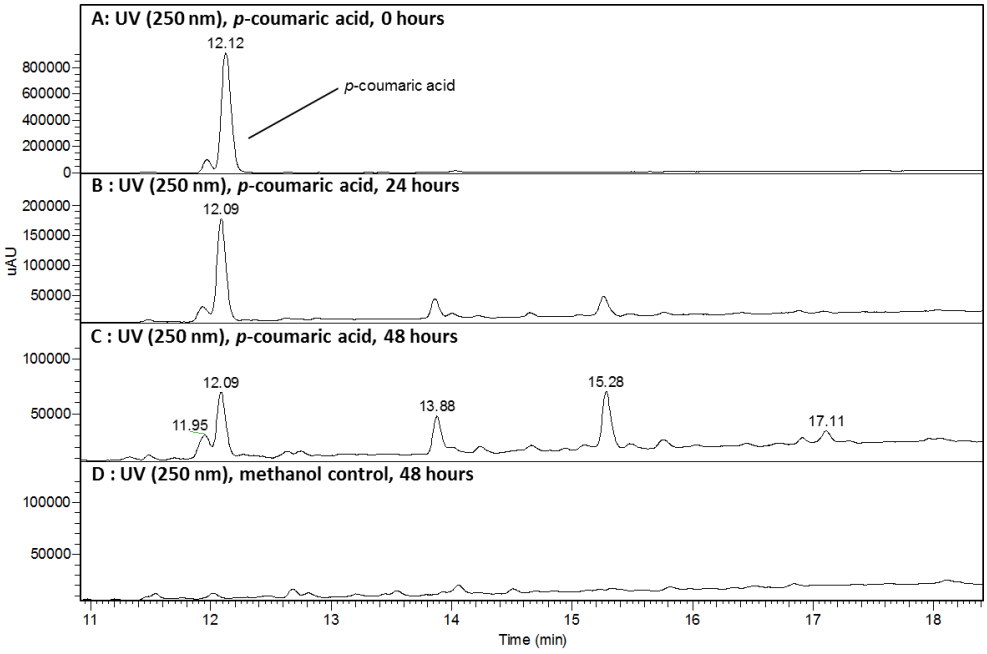


Figure 41. Reverse phase HPLC UV (250 nm) chromatograms showing decrease in *p*-coumaric acid.
p-Coumaric acid in the tobacco leaf incubation solution was monitored by HPLC UV (260 nm) at (**A**) 0 hours, (**B**) 24 hours and (**C**) 48 hours; (**D**) methanol negative control. The most significant decrease was shown over the first 24 hours. New peaks were apparent after 24 and 48 hours at RT 13.9 and 15.3 minutes. Data is from combined analysis of 10 leaf discs, from 3 individual plants.

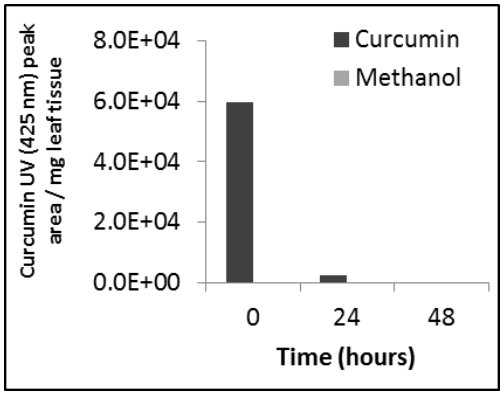


Figure 42. Curcumin decreased when incubated with tobacco leaf tissue.
Curcumin was quantified by HPLC UV (425 nm) peak area at 0, 24 and 48 hours. After 24 hours, the majority of curcumin was no longer detectable. Data is from combined analysis of 10 leaf discs, from 3 individual plants.

Secondly, the fate of curcumin incubated with tobacco leaf tissue was investigated. Tobacco leaf discs were incubated in 0.2 mM curcumin solution for 48 hours. Curcumin was monitored and quantified by HPLC (Figure 42): after 24 hours, no curcumin was detectable. As there were no new peaks in the HPLC trace which could elude to the metabolism of curcumin, the MS total ion current (TIC) trace was examined further: the auto-oxidation degradation product of curcumin, bicyclopentadione *m/z* 399.1 was apparent [203]. This suggested that as well as tobacco metabolising curcumin it was also degrading independently.

Unless the metabolism of *p*-coumaric acid and curcumin in tobacco could be prevented, tobacco would be unlikely to prove a successful chassis for curcuminoid biosynthesis. Turmeric rhizomes are highly specialised for the production of secondary metabolites such as curcuminoids: in an investigation by Koo *et al.* [174], they showed that the branchpoint enzyme to curcuminoid biosynthesis, PAL, is only expressed in the rhizomes and is not detected at all in the leaves. Due to curcuminoids only being produced in the underground rhizomes of the turmeric plant, it was hypothesized that light may affect their accumulation. This hypothesis was supported in the current study by the detection of bicyclopentadione, an auto-oxidative degradation product of curcumin [110,204], which is produced in the presence of free radicals. Therefore, the leaf disc incubations with *p*-coumaric acid and curcumin were repeated in the dark.

After 48 hours in solution without tobacco tissue, *p*-coumaric acid remains at initial levels in both light and dark conditions (Figure 43). However, when incubated with leaf tissue, *p*-coumaric acid is degraded. Initially *p*-coumaric acid degradation occurs faster in the light than in the dark, but after 48 hours the extent of degradation is comparable.

After 48 hours in light conditions, curcumin in solution is degraded by ~ 70 %; when leaf discs are present, this increases drastically with virtually no curcumin remaining (0.34 %) (Figure 44). In the dark, curcumin degradation is barely evident after 48 hours in solution; with 96 % remaining. When leaf discs are added to the dark incubation, curcumin levels drop to 42 % indicating metabolism by the leaf tissue. Therefore, while light alone causes curcumin breakdown, endogenous activity of leaf tissue results in significant light-independent degradation.

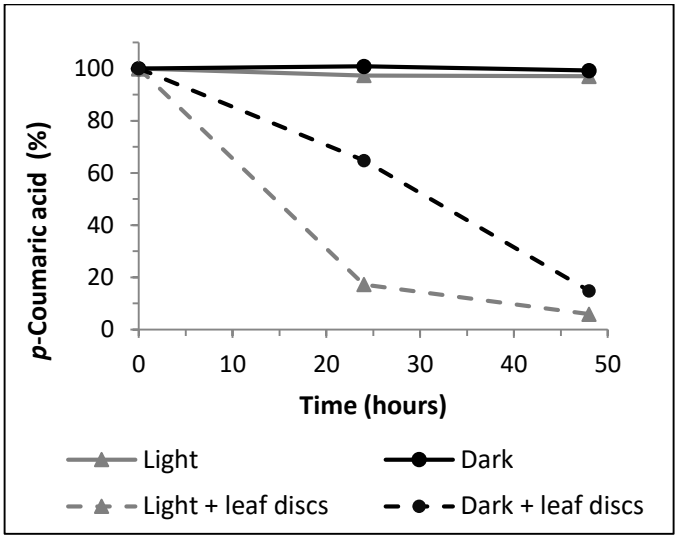


Figure 43. Percentage change in quantified *p*-coumaric acid when incubated with leaf tissue in the light and dark.
p-Coumaric acid in incubation solution was quantified by HPLC UV peak area (250 nm) after 0, 24 and 48 hours with and without tobacco leaf discs in the light and dark. Data is from combined analysis of 10 leaf discs, from 3 individual plants.

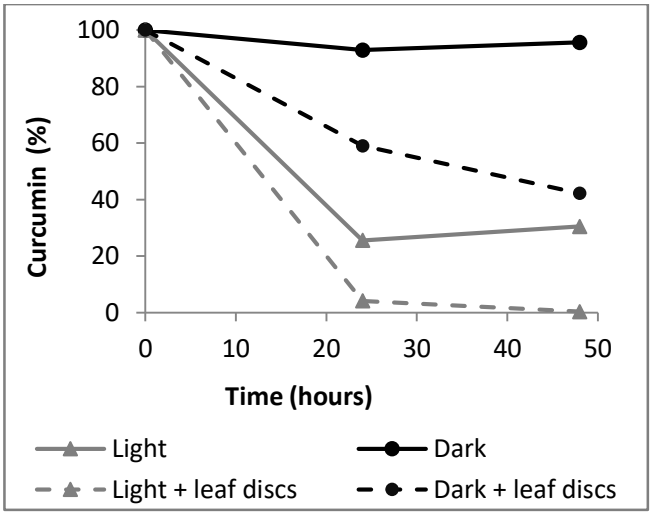


Figure 44. Percentage change in quantified curcumin when incubated with leaf tissue in the light and dark. Percentage of curcumin, measured by HPLC UV peak area (425 nm) after 0, 24 and 48 hours after incubation with and without tobacco leaf discs in the light and dark. Data is from combined analysis of 10 leaf discs, from 3 individual plants.

Exploring the ability of curcuminoid biosynthetic heterologous enzymes to outcompete the endogenous metabolism of *p*-coumaric acid

To replicate the curcuminoid biosynthesic pathway in tobacco, the same three CDSs as used for expression in yeast, 4CL5, DCS and CURS1, were cloned into the pFGC5941 binary vector (Figure 45). When the binary vector is transformed into *Agrobacterium tumefaciens*, the bacterial suspension can be infiltrated into leaf tissue allowing transient expression of the cloned gene (agro-infiltration). The extent to which the heterologous enzymes could compete with the endogenous and degradative pathways of *p*-coumaric acid and curcumin was to be assessed using LC-MS analysis.

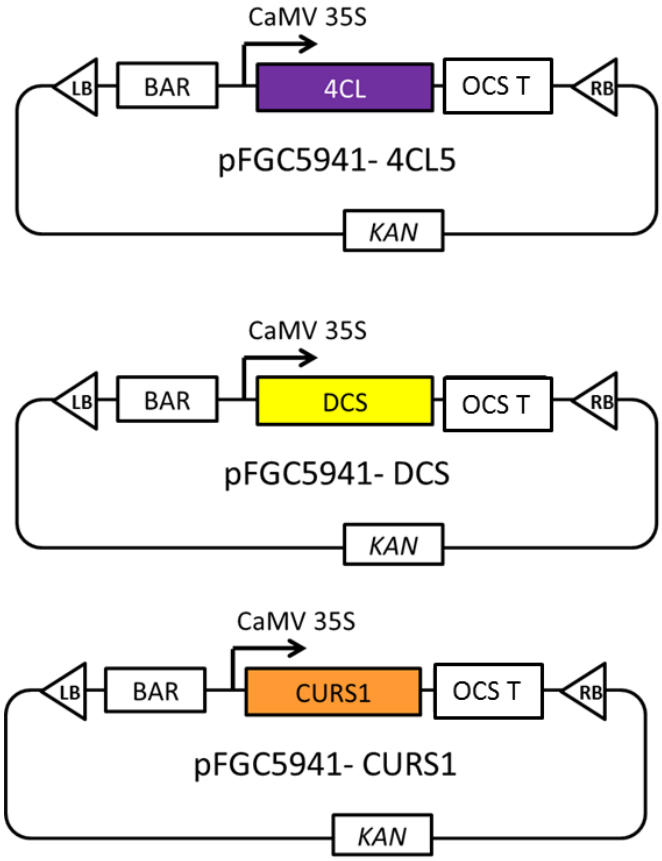


Figure 45. pFGC5941 binary vectors for expression of curcuminoid biosynthetic enzymes in tobacco.
pFGC5941 binary vectors for transformation of *Agrobacterium tumefaciens* for the transient expression of the curcuminoid biosynthetic enzymes 4CL5, DCS and CURS1 in tobacco leaf tissue. The region of the vector which transforms the plant tissue is known as transfer, or T-DNA and is flanked by the left border (LB) and the right border (RB). The octopine synthase terminator (OCS T) was used. Basta resistance gene (BAR) is used for stable expression in plants.

With the expectation that heterologous 4CL5 may be able to out-compete the metabolism of *p*-coumaric acid, leaf disc tissue was harvested from plants agro-infiltrated with a series of differently transformed *A. tumefaciens* culture combinations: pFGC5941-4CL5; both pFGC5941-4CL5 and pFGC5941-DCS; and pFGC5941-4CL5, pFGC5941-DCS and pFGC5941-CURS1. This was to allow each step of curcuminoid biosynthesis to be monitored. When heterologous expression of the curcuminoid biosynthetic enzymes was quantified by quantitative PCR (qPCR), 4CL5 was the only transcript that could be detected in elevated levels: 1500 times greater than the control (Figure 46). There were no detectable transcripts of DCS or CURS1 from agro-infiltration with pFGC5941-DCS or pFGC5941-CURS1 respectively.

Heterologous enzyme activity was monitored by LC-MS analysis. The only enzyme that appeared to be active was 4CL5, and activity could only be detected when assayed in the light (Figure 47): there is a faster decrease in *p*-coumaric acid when incubated with leaf tissue transformed with

4CL5, than untransformed tissue. However, this is only significant at 24 hours (61 % utilised after 24 hours compared to 26 %) and furthermore, no *p*-coumaroyl-CoA ester product could be detected. When *p*-coumaric acid was incubated in the dark with pFGC5941-4CL5 agro-infiltrated leaf tissue, after 48 hours there was no change in its abundance. This suggests that the activity of 4CL5 is dependent upon light, whether this is a direct or indirect effect remains to be addressed. 4CLs are known to be induced by stress [156], but light has not been described as a regulator of 4CL5 activity or expression. The lack of detection of DCS and CURS1 metabolites is most likely due to an absence of enzyme production since an increase in the corresponding transcripts was not detected.

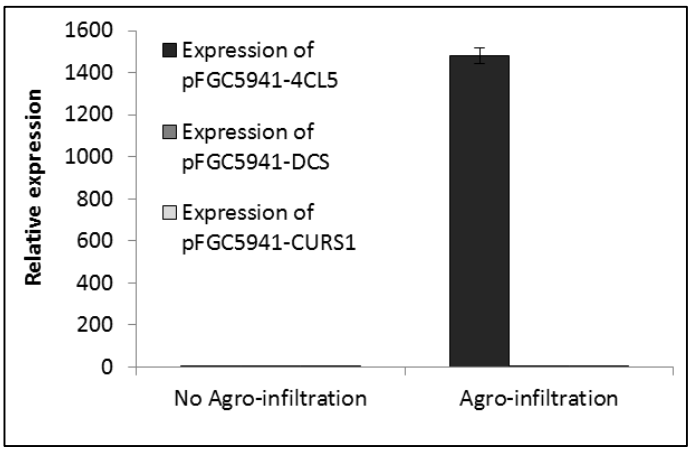


Figure 46. qPCR quantifying 4CL5, DCS and CURS1 expression in tobacco.
qPCR was performed upon RNA extracted from un-infiltrated tobacco leaf tissue and leaf tissue agro-infiltrated with pFGC5941-4CL5, pFGC5941-DCS or pFGC5941-CURS1. The only transcript apparent is that of 4CL5 from expression of pFGC5941-4CL5. There is no difference in DCS or CURS transcript abundance whether the leaf tissue is agro-infiltrated with pFGC5941-DCS, pFGC5941-CURS1 or not.

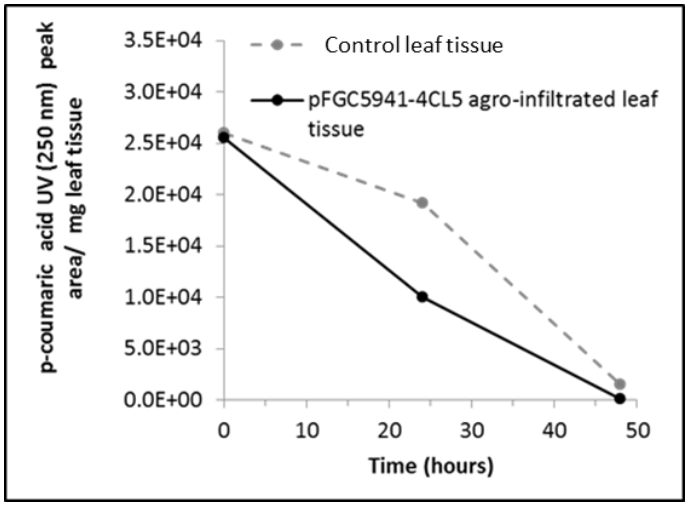


Figure 47. Quantification of *p*-coumaric acid when incubated with leaf tissue transformed with pFGC5941-4CL5.
p-Coumaric acid was quantified by HPLC peak area, UV (250 nm) after 0, 24 and 48 hours after incubation with control tobacco leaf discs or with pFGC5941-4CL5. Data is from combined analysis of 10 leaf discs, from 3 individual plants.

2.3 Discussion and Future Work

Due to the bioactivity of the three curcuminoids, curcumin, DMC and BDMC, the initial objective of this work was to evaluate platforms for their heterologous biosynthesis. *S. cerevisiae* and *N. benthamiana* were selected for evaluation. To produce curcuminoids in these systems the coding sequences for the three biosynthetic enzymes necessary to convert phenylpropanoids to curcuminoids, 4CL5, DCS and CURS1, were amplified and cloned into appropriate expression vectors and transformed into yeast cells or agro-infiltrated into *N. benthamiana* leaf tissue. In yeast, the genes were transformed either individually or as a single unit linked by the 2A polypeptide. In both systems, the only enzyme for which evidence of activity was demonstrated was 4CL5. Consequently, it has not been possible to biosynthesise curcuminoids in either host.

Investigating *S. cerevisiae* as a heterologous host for curcuminoid biosynthesis

A CDS of a 4CL from *C. longa* was to be used as the first enzyme in the biosynthetic pathway. The best described monocotyledonous 4CLs are from rice, and they belong to one of two classes, III or IV [164] (class I and II are in dicots [162]). Using the CDS of monocotyledonous *Os*4CL3, five partial coding sequences of 4CLs were identified from *C. longa* transcriptome data. Phylogenetic analysis of these partial sequences showed them all to align with the type III 4CLs from rice and switchgrass and are therefore likely to be involved in lignin biosynthesis. Due to the abundance of curcuminoids in the rhizome tissue, this was unexpected, as it would have been anticipated that some of the 4CL ESTs from rhizome tissues would be involved in secondary metabolism. Despite multiple attempts it was not possible to amplify full length *C. longa* 4CL from *C. longa* rhizome RNA. As an alternative, a known and characterised 4CL was evaluated.

4CL5 was chosen due to it exhibiting the widest substrate specificity of the four Arabidopsis 4CLs [172,179,205]. In addition to being able to produce natural phenylpropanoyl-CoA ester products of all natural phenylpropanoyl substrates tested *in vitro* (see Figure 25), 4CL5 was also able to produce CoA esters of non-natural substrates 4- methoxycinnamic acid, 3,4-dimethoxycinnamic acid and 4-dimethylaminocinnamic acid (see Figure 27). The accumulation of the novel CoA esters varied: 4- methoxycinnamoyl-CoA abundance increased over time, like feruloyl-CoA, however 3,4-dimethoxycinnamoyl-CoA 4-dimethylaminocinnamoyl-CoA peaked at 3 hours. This may be due to a lower stability of 3,4-dimethoxycinnamoyl-CoA and 4-dimethylaminocinnamoyl-CoA. The versatility of 4CL5 to accept a variety of substrates has since been demonstrated by Eudes *et al.* [151]; its ability to do so was successfully utilised to produce over 160 different anthranilates in *S. cerevisiae*. The flexible substrate specificity of 4CL5 is thought to be due to its larger substrate binding pocket compared to other Arabidopsis 4CLs: amino acid 418 in the substrate binding pocket is a smaller valine, compared to isoleucine [171,205]. Therefore, although 4CL5 was unable to

be used towards producing curcuminoids in this study, it has shown the potential to be a valuable enzyme for the synthesis of a variety of natural and non-natural polyphenolics.

The lack of detectable phenylpropanoyl-CoA esters in 4CL5 *in vivo* cell culture assays was a problem as their availability was required for the subsequent steps of curcuminoid biosynthesis (see Figure 16). One possible explanation for the lack of accumulation of phenylpropanoyl-CoA esters *in vivo* is that they are broken down in a process involving peroxisomal catabolism similar to that of fatty acid CoA esters. In *S. cerevisiae,* fatty acids destined for metabolism *via* β-oxidation first require transport into the peroxisome [206]. Fatty acids cross the peroxisomal membrane in two ways, either passively as free fatty acids or actively transported by the PXA1/PXA2 heterodimeric ABC transporter as CoA esters [207–209]. Therefore, blocking transport of CoA esters into the peroxisome, could avert the metabolism of the phenylpropanoyl-CoA esters.

Individual disruption of PXA1 or PXA2 in yeast exhibits impaired growth on oleic acid by ~50 % [207,208]. Based on the assumption that, as with oleoyl-CoA esters, phenylpropanoyl-CoA ester transport into peroxisomes (and therefore subsequent catabolism) is dependent on PXA1 and PXA2, the corresponding gene-disruption mutants (*pxa1* and *pxa2*) were obtained from GE Healthcare. The plan was to transform these with pYES2-4CL5 and test for accumulation of phenylpropanoyl-CoA esters produced by 4CL5. The phenotypes of *pxa1* and *pxa2* were first tested by growing on agar plates in which oleic acid was the sole carbon source. Wild type *S. cerevisiae* (BY4742) can utilise oleic acid as the sole carbon source and grown normally whereas growth should be impaired in *pxa1* and *pxa2* [210]. However, the mutants that were obtained did not exhibit this phenotype, that is they utilised oleic acid as effectively as wild type (Appendix 2.14), and thus they were considered not suitable to test for accumulation of phenylpropanoyl-CoA esters.

It may not be necessary to knockout phenylpropanoyl-CoA ester utilisation in yeast, as long as the heterologous enzyme using the CoA ester, in this case DCS, can out-compete the endogenous yeast system. 4CL5 successfully competes with the endogenous yeast enzyme PAD1 for phenylpropanoic acids and previous work creating heterologous biosynthetic pathways in *S. cerevisiae* comprising 4CLs had previously proven successful [211,212]. Therefore, despite no detectable production of CoA esters in cell culture by 4CL5, it was considered possible that DCS could out-compete endogenous yeast metabolism for CoA ester substrate availability. Despite DCS protein being detected by western blot analysis (see Figure 28) enzymatic activity *in vivo* and *in vitro*, as analysed by LC-MS, was not. In collaboration with Professors Bryant and White, it was hypothesized that due to a codon mismatch between DCS and *S. cerevisiae* translational machinery, the DCS protein may not be translated correctly yielding a non-functional enzyme.

This was addressed by co-expressing three tRNAs from *S. cerevisiae* or *S. pombe* that were absent, or present in very low levels in *S. cerevisiae*, but whose codons were abundant in DCS. The *S. cerevisiae* leucine tRNA that corresponded to the CTC codon allowed an increase in protein production (see Figure 34), but the increase in protein quantity did not result in detectable enzyme activity (see Figure 35). The other two tRNAs (for codons CTG and CGC) had no effect upon protein abundance. This may be because these tRNAs were from *S. pombe*, rather than *S. cerevisiae*: for transcription, many *S. pombe* tRNAs contain an essential upstream promoter element unlike the tRNAs studied from *S. cerevisiae* [213,214]. To investigate expression of the tRNAs further, a Northern blot should be carried out to ascertain the levels of tRNA. It should also be considered that over expressing a tRNA may not necessarily allow greater protein expression as tRNAs require modifications and many other cellular components to be functional. In fact, under-acetylated tRNAs can actually cause a decrease in translational fidelity as missense substitutions increase [186].

Although expressing tRNAs to balance the codon usage of Bakers' yeast with DCS did not result in enzyme activity being detected, creating a yeast strain that is engineered to produce tRNAs to address codon bias for heterologous gene expression of non-optimised sequences may prove useful. Such a strain would enable yeast to produce high quantity and high quality non-codon optimised recombinant proteins, an idea which has already shown to be beneficial and marketable as evidenced by Novagen's Rosetta *E. coli* strains. It has the potential to allow increased expression of cDNA libraries in *S. cerevisiae* and reduce the costs associated with codon optimisation and gene synthesis.

There are at least two possible reasons why DCS enzyme activity was not detected: either the protein was not correctly assembled or the internal environment of the yeast cell was not conducive for catalysis. Katusuyama's enzymatic studies of DCS determined the enzyme's optimum conditions to be 25- 35˚C and pH 6.5- 7.5 [95]. Yeast *in vivo* assays were carried out at 30˚C and pH was maintained at 6.0 [215], consequently DCS would be expected to be active. Using the Technical University of Denmark's Centre for Biological Sequence Prediction Server the amino acid sequence of DCS was analysed to predict potential subcellular localisation by ascertaining whether it contained any N-terminal presequences such as a chloroplast transit peptide, mitochondrial targeting peptide or secretory pathway signal peptide. The algorithm did not predict DCS to contain any N-terminal signal peptides, so it is unlikely that DCS was non-functional because of incorrect subcellular localisation. However, polyketide synthases are known to associate with the membrane of the endoplasmic reticulum and that the presence of multi-enzyme complexes localised here, facilitate metabolic reactions [216]. This indeed could be a key factor towards the lack of detectable DCS activity. Other avenues such as incorrect protein folding should be investigated.

The final enzyme in the heterologous pathway, CURS1, was not fully investigated in this study due to the lack of substrate availability. Future work would require the synthesis of the substrate by DCS or using chemical synthesis. In publications by Katsuyama [93,95], where the activity of CURS1 was assayed, cinnamoyldiketide-N-acetylcysteamine, a substrate analogous to the natural substrate was used instead [218].

Stoichiometric co-expression of the three different genes was also attempted using the F2A polypeptide as a linker. Exceptional F2A cleavage was ascertained between DCS and CURS1 (see Figure 38D) when expressed from pYES2-DCS-F2A-CURS1, with little uncleaved protein being detected by western blot analysis. Although enzymatic activity was not apparent, the highly efficient cleavage illustrated the potential of F2A for yeast biosynthetic pathway production.

Investigating *N. benthamiana* as a heterologous host for curcuminoid biosynthesis

The three necessary biosynthetic curcuminoid enzymes, 4CL5, DCS and CURS1 were successfully cloned into the pFGC5941 vector and transformed into *Agrobacterium tumefaciens* to allow their transient expression in *N. benthamiana* via agro-infiltration. Before investigating the biosynthetic pathway, the fate of *p*-coumaric acid and curcumin in *N. benthamiana* was evaluated since the metabolism of curcumin would be highly undesirable by a chassis system for curcuminoid production. Unfortunately, over 48 hours, both *p*-coumaric acid and curcumin were metabolised by tobacco (see Figure 43 and Figure 44). Exogenous *p*-coumaric acid caused the appearance of two new detectable compounds, whose identities were not ascertained. There are known endogenous 4CLs in tobacco [219], so it is plausible that *p*-coumaric acid was CoA-esterified and taken up into phenylpropanoid metabolism or lignin biosynthesis [220]. In animals, exogenous curcumin is reduced to tetra- or hexa-hydrocurcumin and conjugated with glucuronic acid or sulfate or cleaved to ferulic acid and vanillin before being excreted [114,221]. However, there are no known pathways for the metabolism of exogenous curcumin in plants. Plant metabolism of xenobiotics usually proceeds via three phases: biotransformation, bioconjugation and deposition in the vacuole or incorporation into the cell wall [222]. In solution, curcumin's minor degradation products are vanillin, ferulic acid, and feruloylmethane and the major degradation product is a bicyclopentadione, the result of auto-oxidative incorporation of O_2 and rearrangement [110,204]. Only the latter was detected in the mass spectrum after tobacco tissue had been incubated with curcumin solution. This may be due to the sequestration of curcumin conjugates for which the extraction method was insufficient.

Despite tobacco metabolising *p*-coumaric acid, it was hypothesized that, as in yeast, heterologous 4CL5 would be able to compete with the endogenous enzymes and shuttle the substrate into curcuminoid biosynthesis. This was indeed observed, as leaf tissue agro-infiltrated with pFGC5941-4CL5 caused a decrease in *p*-coumaric acid quicker than the control (see Figure 47).

However, no *p*-coumaroyl-CoA production by 4CL5 could be detected. Further leaf disc incubation assays should be carried out with *p*-coumaroyl-CoA to attempt to ascertain its fate in tobacco.

As curcuminoid biosynthesis occurs in underground rhizomatous tissue, the effect of dark conditions upon the degradation of *p*-coumaric acid and curcumin was investigated. Assays carried out in the dark showed a reduction in the degradation and metabolism of both *p*-coumaric acid and curcumin. However, curcumin degradation was mainly caused by light alone (see Figure 44). Although light deprivation slowed the metabolism of *p*-coumaric acid, 4CL5 did not exhibit any activity as it did in the light suggesting a dependency upon light (see Figure 47). This could be explained by a reduction of ATP produced in the dark (ATP is involved in the first step of the two step CoA-esterification reaction) due to a reduction in photosynthesis. Another avenue which could be investigated to reduce *p*-coumaric acid consumption by tobacco leaf tissue is utilising RNAi to silence genes whose enzyme products are involved in utilising *p*-coumaric acid for lignin biosynthesis. However, repression of certain lignin biosynthetic enzymes, such as O-methyltransferases, strongly affected plant development so much consideration of target genes is needed [223].

Enzymatic activity of the second two enzymes of the curcuminoid biosynthetic pathway, DCS and CURS1, could not be detected in tobacco. This is likely due to a lack of expression of the enzymes as ascertained by qPCR (see Figure 46). The problems with expressing and detecting activity of the two type III polyketide synthases in both hosts needs to be addressed further. It would be valuable to compare the expression and activity of the non-codon optimised DCS and CURS1 used in this project, with the codon optimised versions successfully utilised by Rodrigues *et al*. [107] in *E. coli*.

In both systems the only enzyme to exhibit activity was 4CL5. A lack of chemical standards for the product of DCS catalysis (and substrate of CURS1) made assays and further analysis difficult. Both *p*-coumaric acid and curcumin were metabolised in the tobacco leaf system making it a poor host for production of curcuminoids. The fact that *C. longa* rhizomes can accumulate curcuminoids, suggests that curcuminoid biosynthesis may be compartmentalised. Ginger, a fellow member of the Zingiberaceae family, accumulates secondary metabolites within large lipid bodies [101]. This environment may be conducive to DCS and CURS1 activity whereas heterologous yeast and tobacco are not. Furthermore, curcumin showed a high rate of degradation in the solution applied to the tobacco leaf tissue. In conclusion, one can identify multiple reasons why curcuminoid biosynthesis was not achieved in either heterologous host system.

Chapter 3. The production of glucosylated curcuminoids

3.1 Introduction

Curcuminoids exhibit a plethora of medicinal benefits, however, their poor bioavailability is a drawback to their efficacy as drug molecules [115]. Consequently, there is a lot of interest in increasing the bioavailability of curcuminoids such as creating liposomal or nanoparticulate formulations, administration with adjuvants or creating structural analogues with improved biophysical properties [112,118,224,225]. Glucosylation of hydrophobic molecules is a well-known strategy to enhance water solubility [226]: curcumin monoglucoside is 230 times more soluble in water than curcumin and curcumin diglucoside is 10 million times more soluble in water than curcumin [138]. Compared to chemical glucosylation, enzymatic glucosylation offers greater regio- and stereo specificity and allows milder reaction conditions [119]. One aim of this project was to use biocatalysis to produce more bioavailable, glucosylated curcuminoids.

UDP-glycosyltransferases (UGTs) catalyse the glucosylation of small molecules, with the majority using uridine diphosphate-glucose (UDP-glucose) as a sugar donor. In plants, they function to mediate the stability, solubility and activity of endogenous secondary metabolites as well as detoxifying xenobiotics such as herbicides [226]. By doing so they maintain intracellular homeostasis whilst responding to abiotic and biotic stresses.

The diverse range of substrates, along with the enantio- and regioselectivity of UGTs, offers a huge potential for biocatalysis. Biotechnology has explored plant cell cultures for their ability to glucosylate a wide variety of industrially relevant aromatic compounds such as vanillin [227], hydroquinone [228] and capsaicin [229]. All these processes utilised *Catharanthus roseus* plant cell culture and this is to date, the only plant from which UGTs that glucosylate curcuminoids have been characterised [138]. Although, there is evidence that *P. tricuspidata* cell culture can also glucosylate curcumin, there is no sequence data for the enzymes involved [145]. In both systems, when exogenous curcumin was applied to plant cells extracts, a series of glucoside, di-glucoside, gentiobioside, and digentiobiosides were produced, all of which were β-anomers (see Figure 13). The only instance of curcumin-α-glucoside production by enzyme catalysis was using an amyloglucosidase from the fungus *Rhizopus* [136].

The cDNA sequence for the two UGTs responsible for curcumin glucosylation by *C. roseus* cell culture were isolated and the corresponding enzymes characterised [1,143]. The two enzymes were designated CaUGT2 and CaUGT3. CaUGT2 converts both curcumin and curcumin monoglucoside to curcumin monoglucoside and curcumin diglucoside, respectively; the K_Ms of

which are 19 and 63 μM, respectively. However, its ability to glucosylate curcuminoids other than curcumin was not studied. CaUGT3, a sugar-sugar glycosyltransferase which was able to add an additional sugar group onto the existing sugar moiety of curcumin glucosides, produced curcumin gentiobiosides (see Figure 14).

This project aimed to express a curcuminoid UGT in *S. cerevisiae* to produce curcuminoids glucosides, with enhanced bioavailability. It was expected to be an improvement upon current proof-of-concept biosyntheses by being easily scalable, therefore allowing the creation of an industrially relevant, greener alternative to the chemical synthesis of curcuminoid glucosides. To date, no UGTs have been identified in *C. longa* or its Zingiberaceae family member ginger (*Z. officinale*) which produces secondary metabolites similar to curcuminoids such as 6-gingerol. This study aimed to investigate *C. longa* as a source of a novel curcuminoid UGTs that could be used in a heterologous yeast system.

3.2 Results

3.2.1 Searching for curcuminoid glucosides in *C. longa* metabolite extracts

To produce glucosylated curcuminoids in yeast, a UGT enzyme which could glucosylate curcuminoids was needed. Although CaUGT2 from *C. roseus* could be used, the discovery of a novel UGT from *C. longa* was first explored. As *C. longa* rhizome tissue is the primary tissue in which curcuminoid biosynthesis occurs [174], evidence of curcuminoid UGT activity was looked for in these organs. Using LC-MS analysis, it was first investigated whether curcumin glucosides could be identified in *C. longa* metabolite extracts.

LC-MS metabolite analysis of extracts from 13 samples (n = 30) of *C. longa* rhizome tissues collected from 9 different locations across Vietnam (Table 2) had previously been performed in the CNAP metabolite profiling unit by Dr Tony Larson in collaboration with Kieu Oahn Nguyen, University of Science and Technology, Hanoi, Vietnam. The extraction and LCMS detection conditions were suitable for detection of curcuminoids and other molecules (see Materials and Methods).

The five most abundant metabolites extracted from the *C. longa* rhizome tissues were quantified by their mass spectrum peak areas (Figure 48). Due to the size of their parent ions, these were identified as the three main curcuminoids (curcumin, DMC and BDMC (see Figure 1)) and two sesquiterpenes, ar-turmerone and turmerone (Figure 49). The most abundant secondary metabolites were the turmerone sesquiterpenes. In all the samples curcumin was the most prevalent curcuminoid, with lower amounts of DMC and BDMC present at similar levels. There was a clear variability in the amount of secondary metabolites within and between the samples.

The two samples from the Lạng Sơn province (HLLS and BSLS) overall contained the greatest amount of secondary metabolites. BVNH, LNBG, TLHB and YTHB contained approximately tenfold less metabolites than the rest of the samples.

Table 2. *C. longa* rhizome samples gathered from across Vietnam.

Abbreviation	**Location (Province)**
BVNH	Ba Vì (Hà Nội)
HLLS	Hữu Lũng (Lạng Sơn)
BSLS	Bắc Sơn (Lạng Sơn)
LDI	Garden 48 (Lâm Đồng)
LDII	Garden 56 (Lâm Đồng)
LGBG	Lạng Giang (Bắc Giang)
LNBG	Lục Nam (Bắc Giang)
KTNA	Kẻ La Thượng (Nghệ An)
DHQT	Đông Hà (Quảng Trị)
ATTQ	An Tường (Tuyên Quang)
XMHB	Xuân Mai (Hòa Bình)
YTHB	Yên Thủy (Hòa Bình)
TLHB	Tân Lạc (Hòa Bình)

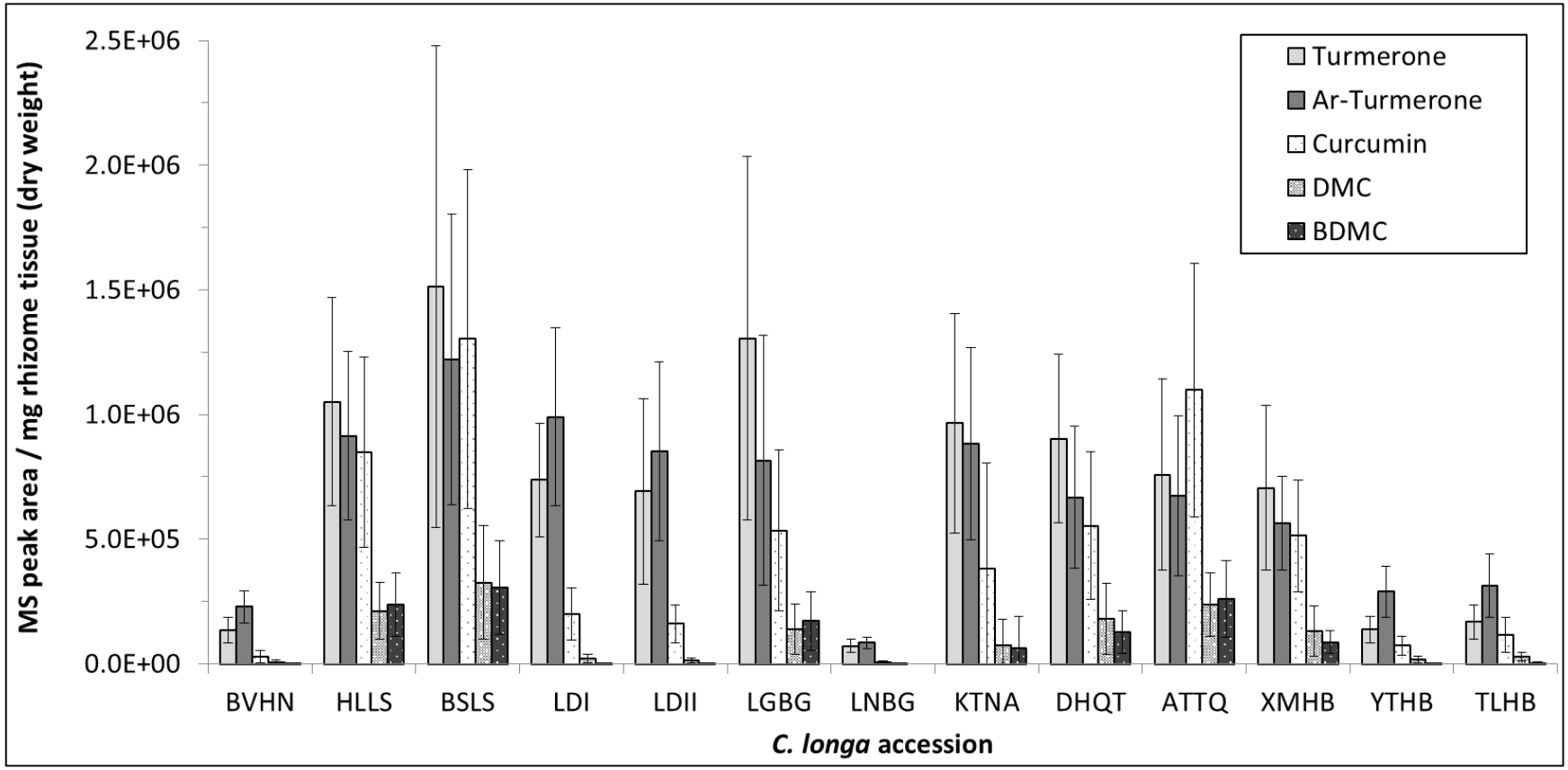


Figure 48. Quantification of the five most abundant secondary metabolites in Vietnames *C. longa* samples.
MS peak area was used to quantify the five most abundant secondary metabolites present in *C. longa* rhizomes from 13 samples from across Vietnam. MS peak area: turmerone *m/z* 219.17, Ar-turmerone *m/z* 217.16, curcumin *m/z* 369.13, DMC *m/z* 339.12 and BDMC *m/z* 309.11. The data are means ± standard deviations (n = 30). With thanks to Kieu Oahn Nguyen for collecting and performing extractions upon the samples and Dr Tony Larson for providing access to the structured mass spectra dataset.

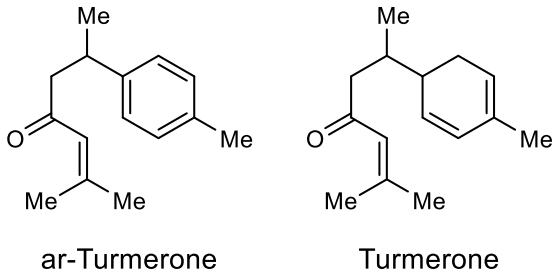


Figure 49. Chemical structures of sesquiterpenes ar-turmerone and turmerone.

The abundance of curcumin in each of the samples was plotted on the basis of collection site location (Figure 50A). The north of Vietnam, in the Lạng Sơn province, provided rhizomes with the greatest curcumin abundance (BSLS and HLLS), with the second most abundant, LTTQ, nearby in Tuyên Quang (Figure 50B). This may be due to their location in the fertile Red Delta, and the fact that the north of Vietnam has a more humid tropical climate, favouring *C. longa* growth compared to that of southern Vietnam. However, the accessions with the lowest abundance, LNBG and BVHN, were also located in the Red Delta area. Due to the small number of accessions and lack of detailed information on the harvest sites (such as age of plant, nitrogen abundance, soil pH, presence of heavy metals and pest populations) no firm conclusion could be made about the geographical influence upon secondary metabolite abundance.

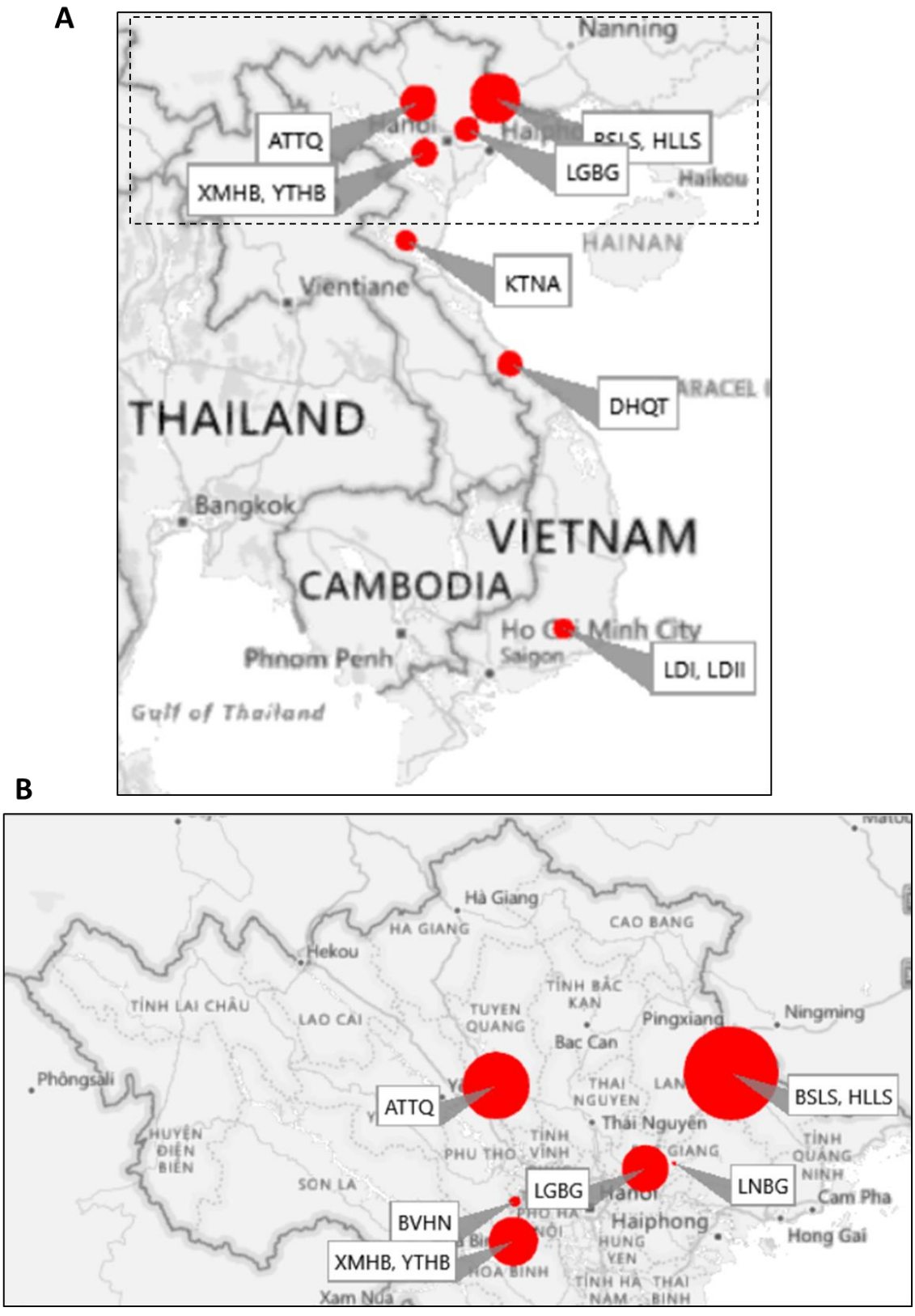


Figure 50. Map of Vietnam showing the sampling locations of the *C. longa* rhizome samples.
The size of each red point mirrors the abundance of curcumin, quantified by MS peak area, in the turmeric rhizome sample gathered from that area. (**A**) All of Vietnam. (**B**) The dotted region in (A) is shown in more detail in to highlight the lowest abundance samples.

The LC-MS data was further analysed for the presence of glucosylated curcuminoids. No detectable mass ions corresponding to the expected sizes of glucosylated curcumin, DMC or BDMC were present in the dataset. Glutathionylated and sulfonated bioconjugates of curcuminoids were also searched for, but could not be found. The lack of curcuminoid conjugates, not only glucosides, was surprising as curcuminoids are extremely abundant in the rhizome tissue. This lead to the suggestion that the plant most likely utilises another strategy to modulate curcuminoid activity and storage. A comparison of the abundance of the secondary metabolites extracted from the rhizome tissues showed that the amounts of curcuminoids correlated as expected, with curcumin and DMC showing an almost perfect Pearson correlation (r (28) = 0.98, $p < 0.01$ (Table 3)). Interestingly, the abundance of the turmerones and curcuminoids also showed a high Pearson correlation co-efficient e.g. the correlation between curcumin and turmerone abundance was r (28) = 0.78, $p < 0.01$. This correlation wasn't expected as the sesquiterpene and polyketide pathways are distinct. Ginger, a fellow member of the Zingiberaceae family, utilises hydrophobic secondary metabolites such as sesquiterpenes to form liposomal droplets around its polyphenolic bioactives [101]. Due to the correlation of sesquiterpene and curcuminoid abundance, it is possible that *C. longa* utilises a similar strategy to ginger; using turmerones to dissolve and encapsulate curcuminoids without the need for glucosylation or bioconjugation.

Table 3. Pearson correlation co-efficients between the abundance of the top five most prevalent metabolites in *C. longa* rhizome tissue as quantified by parent ion MS peak area.
The darker the colour, the greater the correlation. Degrees of freedom = 28, $p < 0.01$.

	Turmerone	Ar-turmerone	Curcumin	DMC	BDMC
Turmerone	1.00	0.91	0.78	0.78	0.78
Ar-turmerone	0.91	1.00	0.66	0.62	0.60
Curcumin	0.78	0.66	1.00	0.98	0.98
DMC	0.78	0.62	0.98	1.00	0.97
BDMC	0.78	0.60	0.98	0.97	1.00

As there were no curcuminoid glucosides detected in the LC-MS dataset of the 13 *C. longa* rhizome samples from Vietnam, it was decided to perform a different, more intensive metabolite extraction. A methanolic Soxhlet extraction was performed upon fresh *C. longa* rhizomes and a liquid-liquid extraction was carried out with the expectation of separating any curcumin glucosides from other metabolites. LC-MS analysis was carried out upon all extracts, yet again no curcumin glucosides could be detected in the LC chromatogram or mass spectrum. To further investigate the question of curcuminoid glucosylation, a glycosidase assay was performed upon the extract using a *Trichoderma reesei* cellulase and products were detected by HPLC analysis (Figure 51). The rationale of this experiment was that the cellulase would cleave any glycosidic bonds causing a disappearance or shift of HPLC chromatogram peaks. This was observed for the monoglucosylated curcumin positive control (Figure 51A): the curcumin monoglucoside peak, *m* (RT 8.9 minutes), disappears with glycosidase treatment. The peak for the aglycone curcumin, *c* (RT 12.4 minutes), remained unaffected, as expected. Analysis of the methanolic rhizome extract showed there was no change in the chromatogram before and after glycosidase treatment, leading to the conclusion that there were no glucosylated curcuminoids in the rhizome methanolic extract (Figure 51B).

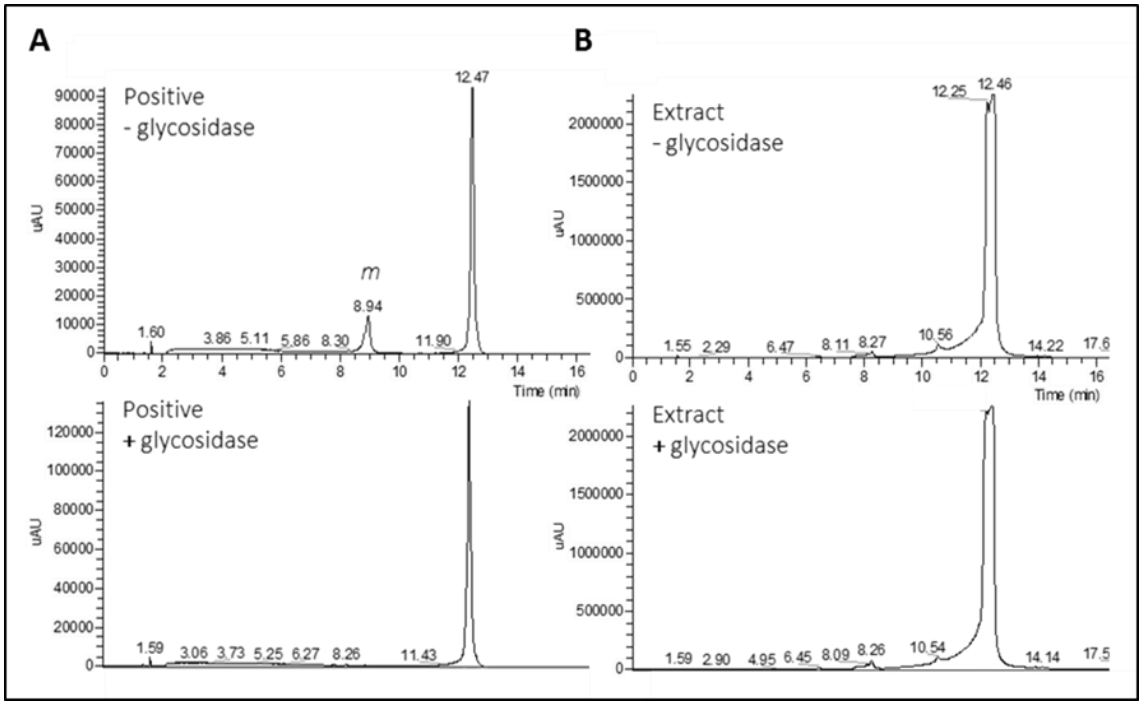


Figure 51. Glycosidase assay of *C.longa* metabolite extract as monitored by HPLC.
HPLC chromatograms (425 nm) of *C. longa* rhizome methanol extract treated with and without cellulase. The positive control (**A**) shows a disappearance in the curcumin monoglucoside peak (*m*, RT 8.9 minutes) in the presence of glycosidase. (**B**) There was no difference in the HPLC traces of the methanol extract treated with, and without glycosidase. Curcumin (*c*, RT 12. 4 minutes), the major peak in all the samples, remained unaffected by glycosidase treatment.

3.2.2 Searching for evidence of curcuminoid UGTs in *C. longa*

To investigate whether the lack of detectable glucosides in *C. longa* was due to a lack of UGTs, protein was extracted from *C. longa* rhizomes and assayed for UGT activity. Protein extract from Arabidopsis (*A. thaliana*) and rice (*O. sativa*) was also investigated for UGT activity towards curcuminoids. Radiolabelled ^{14}C-UDP-glucose was incubated with the protein extracts and potential acceptor substrates curcumin, DMC, BDMC (see Figure 1) and the positive control substrates 2,4,5- trichlorophenol (TCP) and quercetin (Figure 52). TCP is a degradation product of many xenobiotics and persists in the environment, therefore is a typical UGT substrate [230]. Quercetin is a flavonoid which has been shown to be glycosylated by many different UGTs [130].

Transfer of the ^{14}C-glucose radiolabel to the potential acceptors was quantified by scintillation counting of the ethyl acetate layer after liquid-liquid extraction (Figure 53) (see Materials and Methods). There was no UGT activity evident towards methanol as expected. Activity towards the model substrates TCP and quercetin showed that the radiolabelling assay was functioning and UGTs in the plant protein extracts were functioning and could be detected. *C. longa* rhizome protein extract exhibited the greatest activity towards TCP (Figure 53A) whereas Arabidopsis protein extract showed the greatest UGT activity towards quercetin (Figure 53B). *O. sativa* protein extract showed the greatest glucosylation activity of the three plant protein extracts tested and exhibited comparable activity towards both TCP and quercetin (Figure 53C). There was however no evidence of UGT activity towards curcumin, DMC or BDMC by any of the plant protein extracts.

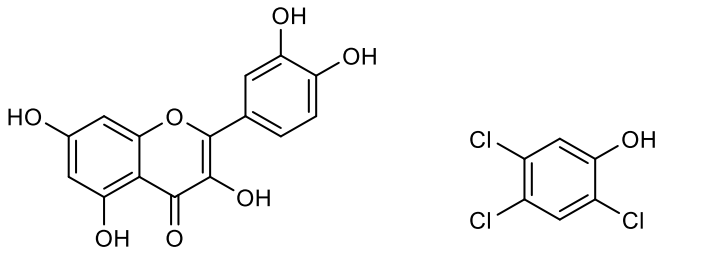


Figure 52. Chemical structures of quercetin and 2,4,5-trichlorophenol.

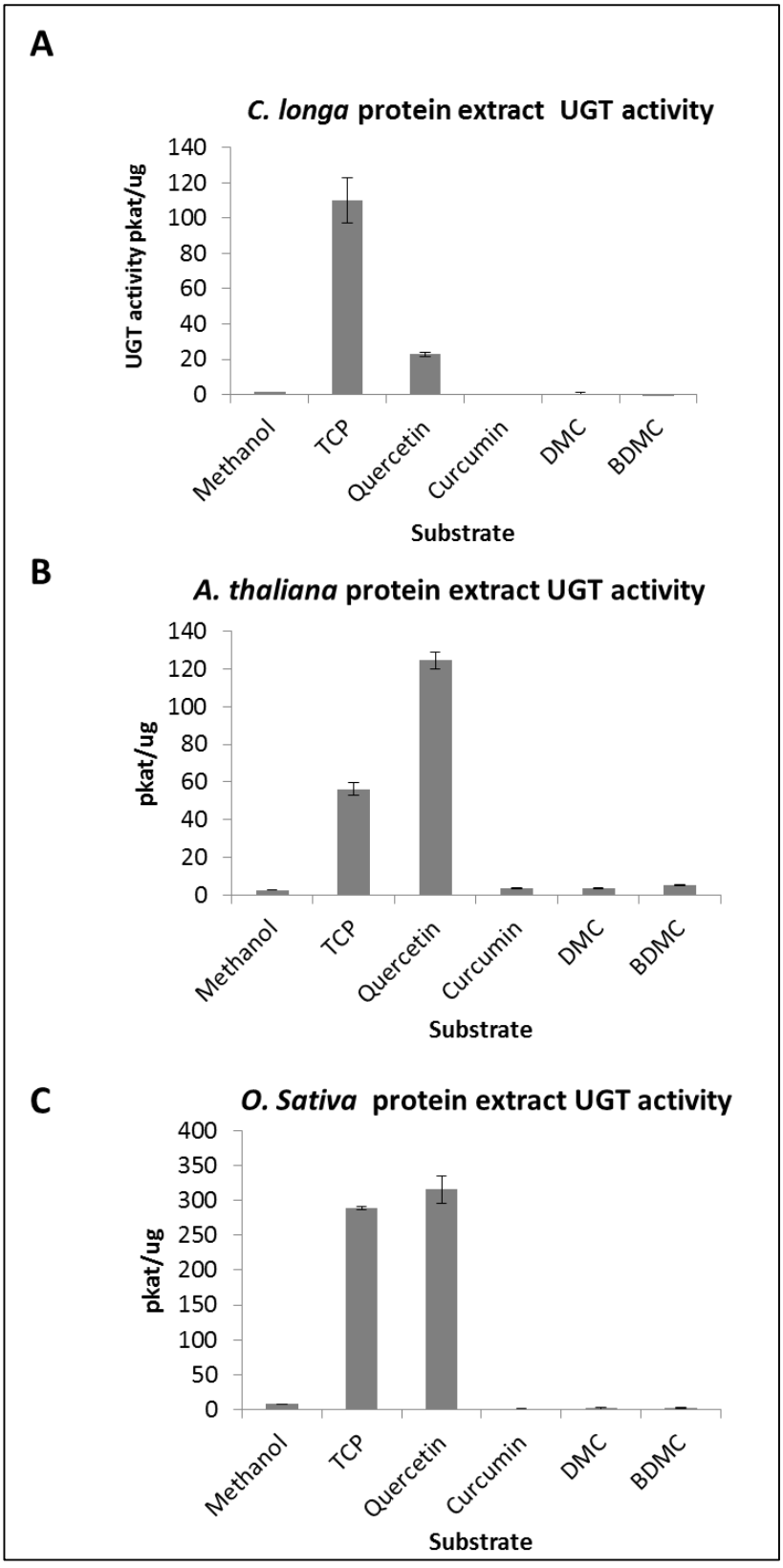


Figure 53. UGT activity of *C. longa*, *A. thaliana* and *O. sativa* protein extracts.
The UGT activity of the extracted protein from (**A**) *C. longa*, (**B**) *A. thaliana* and (**C**) *O. sativa* protein towards curcumin, DMC and BDMC was determined by transfer of ^{14}C-glucose and measured by scintillation counting. TCP and quercetin were used as positive controls; methanol was the negative control. The data are means ± standard deviations (n = 3).

To date, the only characterised UGT to glucosylate curcumin is CaUGT2 from *Catharanthus roseus* [1]. Therefore, to search for novel curcuminoid UGTs in *C. longa*, the amino acid sequence of CaUGT2 was used as a query for a tBLASTn search of the NCBI *C. longa* rhizome transcriptome database. Two potential full length CDS were found (JW777458.1 and JW773795.1, Appendix 3.1) which although both showed 83 % query coverage, shared only 39 % (Figure 54A) and 37 % (Figure 54B) sequence identity with CaUGT2. Due to the lack of evidence of curcuminoid glucosylation in *C. longa* together with the low homology between JW777458.1, JW773795.1 and CaUGT2, it was decided not to further characterise these cDNAs.

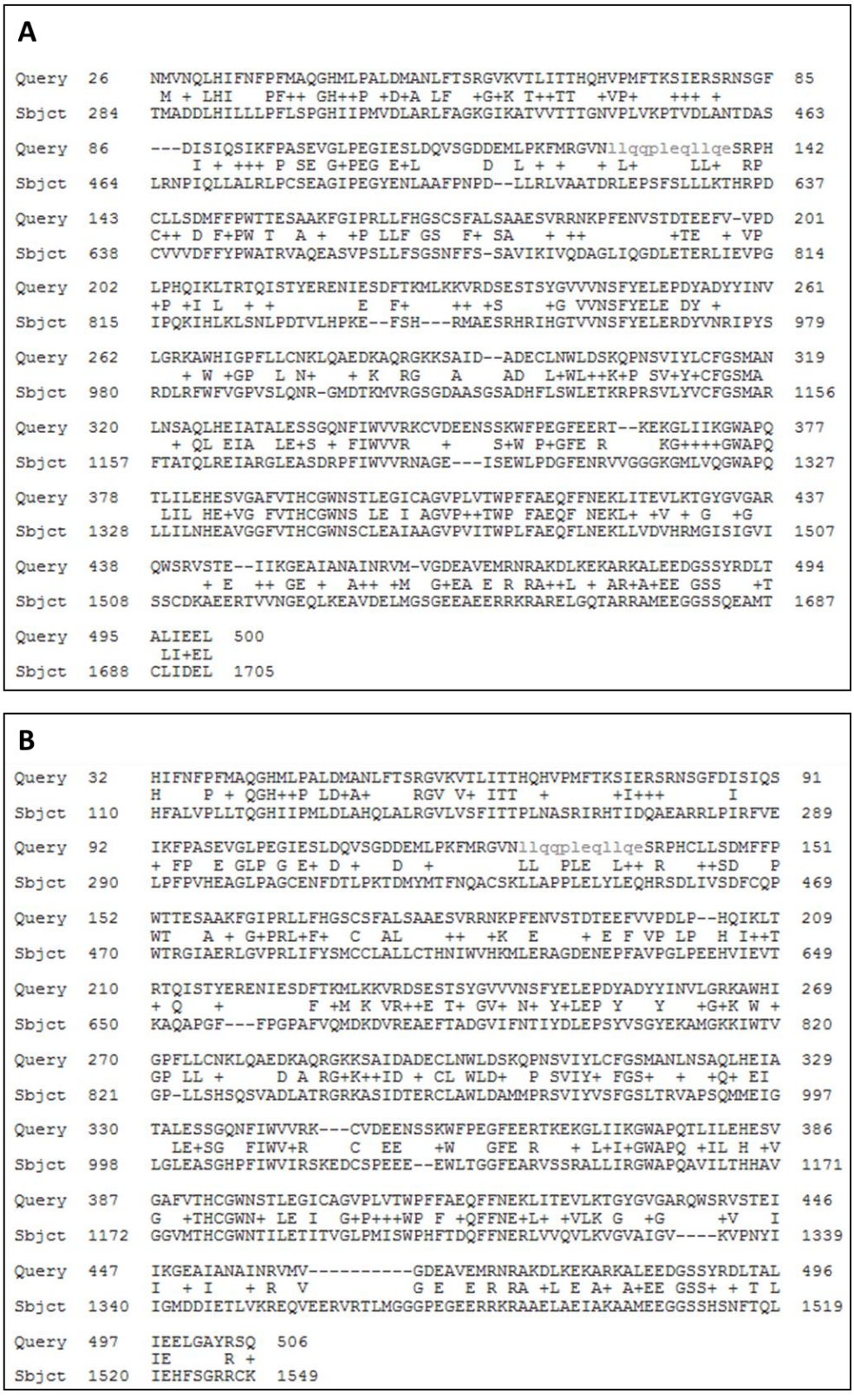


Figure 54. Amino acid sequence alignment between CaUGT2, JW777458.1 and JW773795.1.
(**A**) Subject JW777458.1, Query CaUGT2, 83 % coverage, 39 % identity. (**B**) Subject JW773795.1, Query CaUGT2, 83 % coverage, 37 % identity.

3.2.3 Investigating the activity of recombinant CaUGT2

The apparent lack of *C. longa* curcuminoid UGTs led to the decision to use CaUGT2 from *C. roseus* for development of a biocatalytic platform for curcuminoid glucosylation in *S. cerevisiae*. CaUGT2 had been shown to produce curcumin mono- and diglucosides in *in vitro* experiments, but only curcumin monoglucoside when *in vivo* experiments were carried out in *E. coli* [1,231].

The nucleotide sequence of CaUGT2 (AB159213.1, Appendix 1.3) was synthesized by GenScript and cloned into the pUC57 vector. To validate CaUGT2's ability to glucosylate curcumin, and to test its activity towards DMC and BDMC, which had not been done previously, its activity was initially checked by expression in *E. coli*. The CaUGT2 nucleotide sequence was cloned into the pET vector where it was translationally fused to a *Strep*-tag (Figure 55A). pET-CaUGT2 was transformed into *E. coli* and protein was extracted for *in vitro* analysis. Protein expression was assessed by western blot analysis using *Strep*-tactin (Figure 55B). No protein was detected from the empty vector control (pET-EV) as expected, whereas CaUGT2, translationally fused to the *Strep*-tag, was detected at the expected size, 55 kD. In fact, CaUGT2 expression was so high, it was visible by Ponceau staining. The crude protein extract was used in an *in vitro* assay testing for CaUGT2's ability to glucosylate curcuminoids and phenylpropanoids using ^{14}C-UDP-glucose (Figure 56). Scintillation counting showed protein extracted from *E. coli* transformed with pET-CaUGT2 produced glucosides of the positive controls TCP and quercetin, as well as the three curcuminoids, curcumin, DMC and BDMC. This is the first time CaUGT2 has been shown to glucosylate DMC and BDMC. The four main phenylpropanoids, the building blocks of curcuminoids, were also tested in the *in vitro* assay. Contrary to work by Huang *et al.* [141], CaUGT2 was able to glucosylate caffeic acid, but only to approximately one tenth of the level of the curcuminoids. The other three phenylpropanoids, cinnamic, ferulic and *p*-coumaric acid, were not glucosylated (see Figure 56).

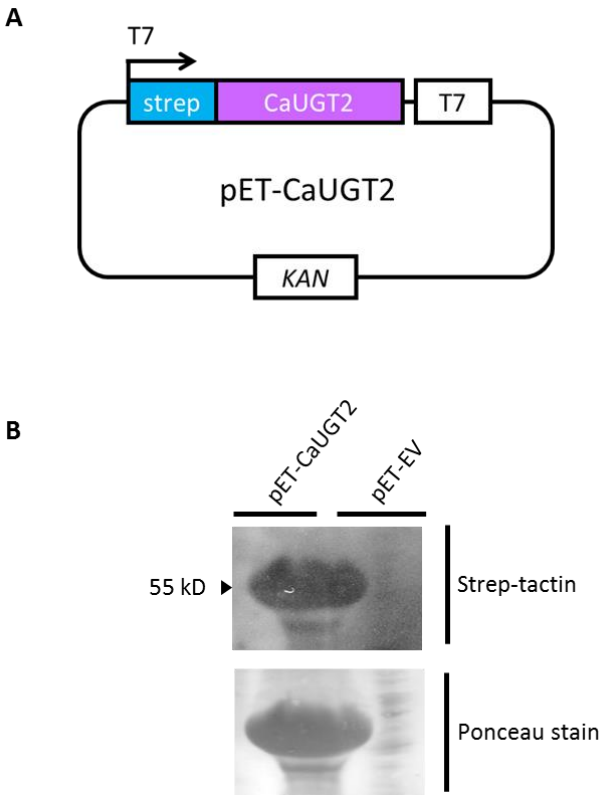


Figure 55. Cloning and expression of CaUGT2 in *E. coli*.
(**A**) CaUGT2 was cloned into the pET vector and translationally fused to the *Strep*-tag nucleotide sequence allowing CaUGT2 protein production in *E. coli* to be analysed by western blot using *Strep*-tactin (**B**). CaUGT2 was observed at the expected size of 55 kD. No *Strep*-tagged protein was evident in the pET-EV, empty vector control. The *Strep*-tactin western blot is shown above the ponceau stain loading control.

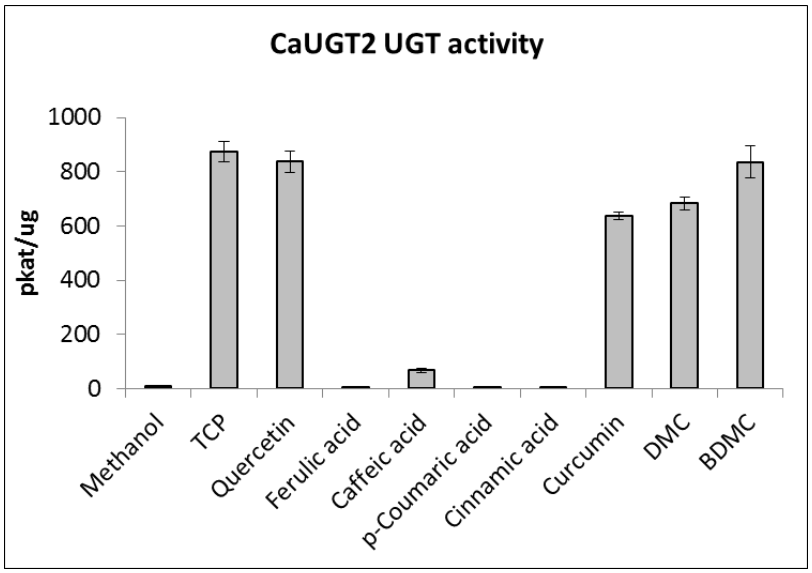


Figure 56. *In vitro* CaUGT2 activity towards curcuminoids and phenylpropanoids.
UGT activity was determined by transfer of ^{14}C-glucose to the substrate which was measured by scintillation counting. 2,4,5-Trichlorophenol (TCP) and quercetin were used as positive controls, methanol was the negative control. The data are means ± standard deviations (n = 3).

Having confirmed CaUGT2 expression in *E. coli* the CaUGT2 CDS was cloned into the pBEVY yeast expression vector using Gibson Assembly (Figure 57A) and transformed into *S. cerevisiae*. Transformation was verified by PCR analysis of DNA extracted from yeast cells (Figure 57B): there is no PCR product for the empty vector control (pBEVY-EV) when using a CaUGT2 specific primer in combination with a vector primer, but there is the correct sized product when DNA extracted from pBEVY-CaUGT2 yeast was used as a template (1.5 kB). *Strep*-tagged CaUGT2 protein could be detected by western blot analysis using *Strep*-tactin, whereas, no protein was detected in empty-vector control extracts (Figure 57C).

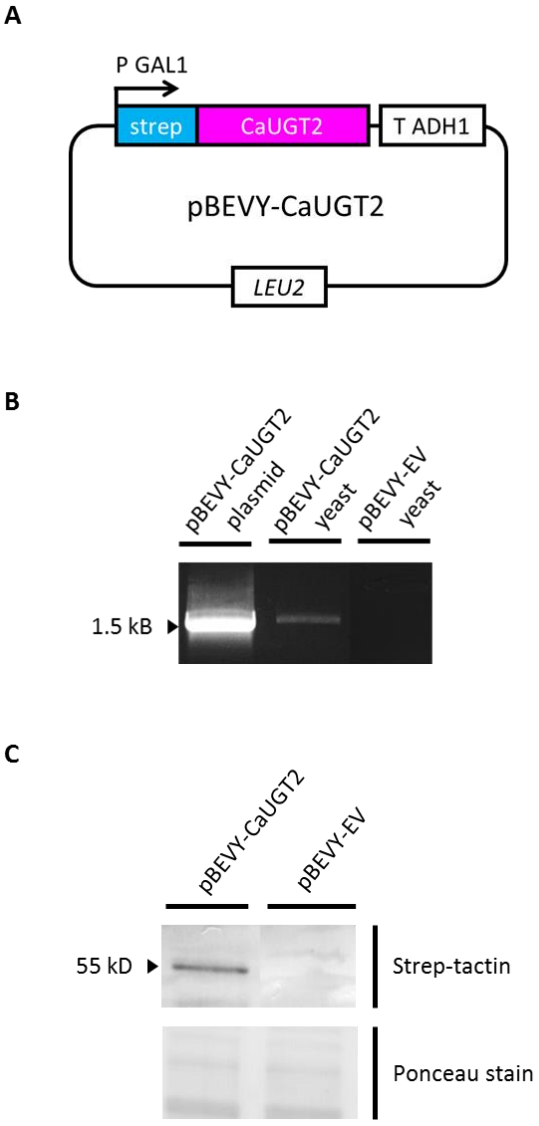


Figure 57. Cloning and expression of pBEVY-CaUGT2.
(**A**) CaUGT2 was translationally fused to the *Strep*-tag nucleotide sequence in the pBEVY vector. (**B**) Agarose gel showing PCR products, when CaUGT2 specific primers (see Primer Table in Material and Methods) were used to amplify DNA extracted from yeast cells transformed with pBEVY-CaGUT2, pBEVY-EV (empty vector control) or the pBEVY-CaUGT2 plasmid. (**C**) Western blot analysis, using *Strep*-tactin upon protein extracted from yeast cells expressing pBEVY-CaUGT2 or pBEVY-EV. *Strep*-tagged CaUGT2 protein is seen at the expected size of 55 kD. Ponceau stain of the nitrocellulose membrane is used as a loading control.

In an analogous experiment to that carried out in *E. coli*, the *in vitro* activity of CaUGT2 in crude protein extract from *S. cerevisiae* carrying pBEVY-CaUGT2 was carried out. Instead of using radiolabelled ^{14}C-UDP glucose, cold UDP-glucose was used and LC-MS analysis was carried out to ascertain activity towards curcumin, DMC and BDMC. All three curcuminoids were glucosylated by *S. cerevisiae* protein extract containing CaUGT2, as can be seen by the accumulation of the respective monoglucosides, measured by the area of their parental mass peaks in the mass spectrum (Figure 58A, B and C). The identity of the glucosides was verified by their fragmentation loss of 162 mass units which is characteristic of glucosides due to the loss of glucose, as exemplified in Figure 59 with curcumin glucoside. Curcumin monoglucoside has a parental mass ion of 528.8, whereas DMC and BDMC have parental mass ions 498.9 and 468.9, respectively (see Appendices 3.2 and 3.3 for spectral data). Over the 60 minutes assayed, the monoglucosides of all three curcuminoids steadily increased in abundance. There was no evidence for production of diglucosides in these *in vitro* assays of protein extracts. Despite CaUGT2 being described by Kaminaga *et al.* [1] as being able to mono- and diglucosylate curcumin in *E.coli in vitro* assays, in these yeast *in vitro* experiments CaUGT2 was only observed to produce monoglucosides.

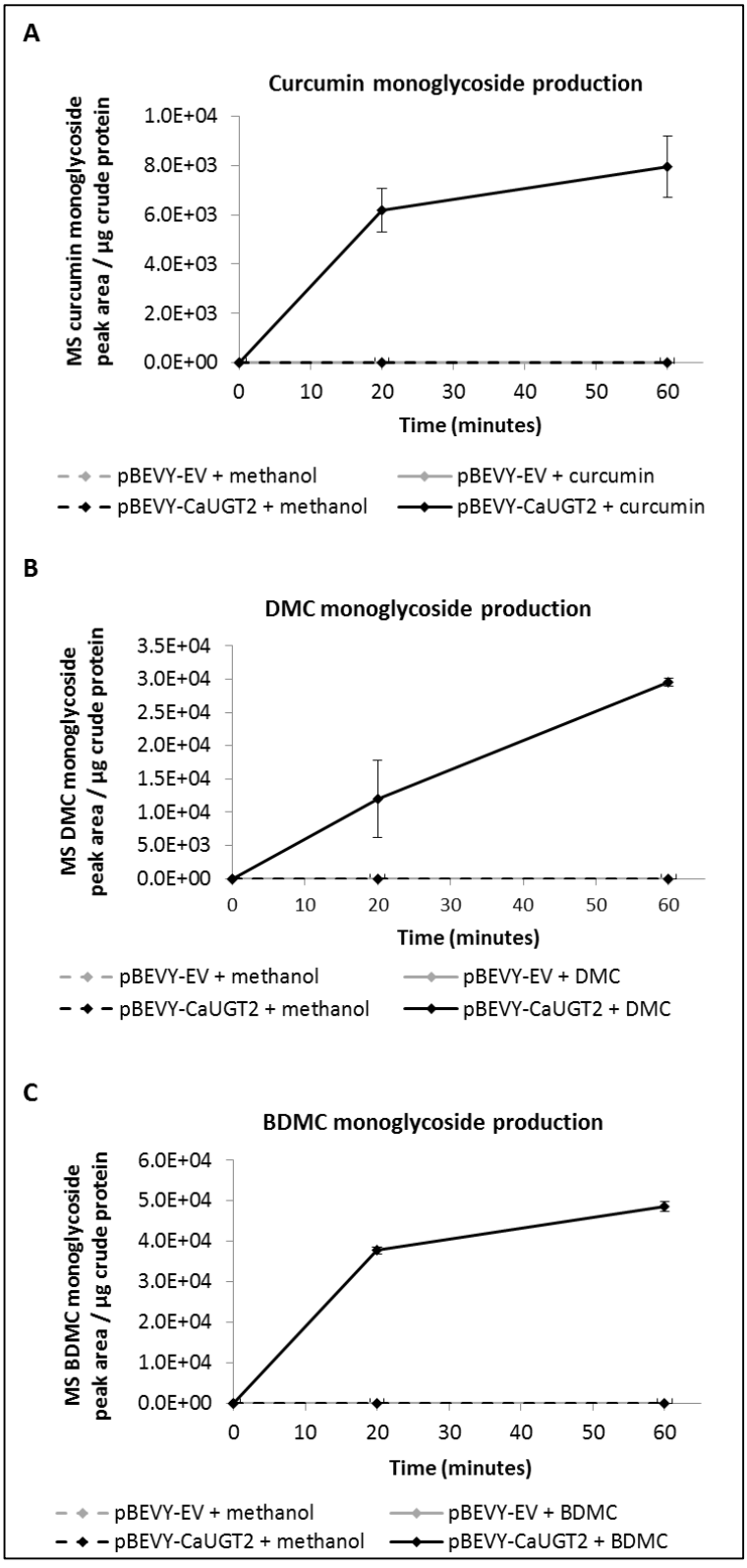


Figure 58. Curcuminoid monoglucoside production by *S. cerevisiae* pBEVY-CaUGT2 protein extract.
LC-MS analysis of protein extracted from yeast expressing pBEVY-CaUGT2 or pBEVY-EV. Production of monoglucosides was quantified by MS peak area: (**A**) curcumin monoglucoside $[M-H]^-$ 528.8, (**B**) DMC monoglucoside $[M-H]^-$ 498.8 and (**C**) BDMC monoglucoside $[M-H]^-$ 468.

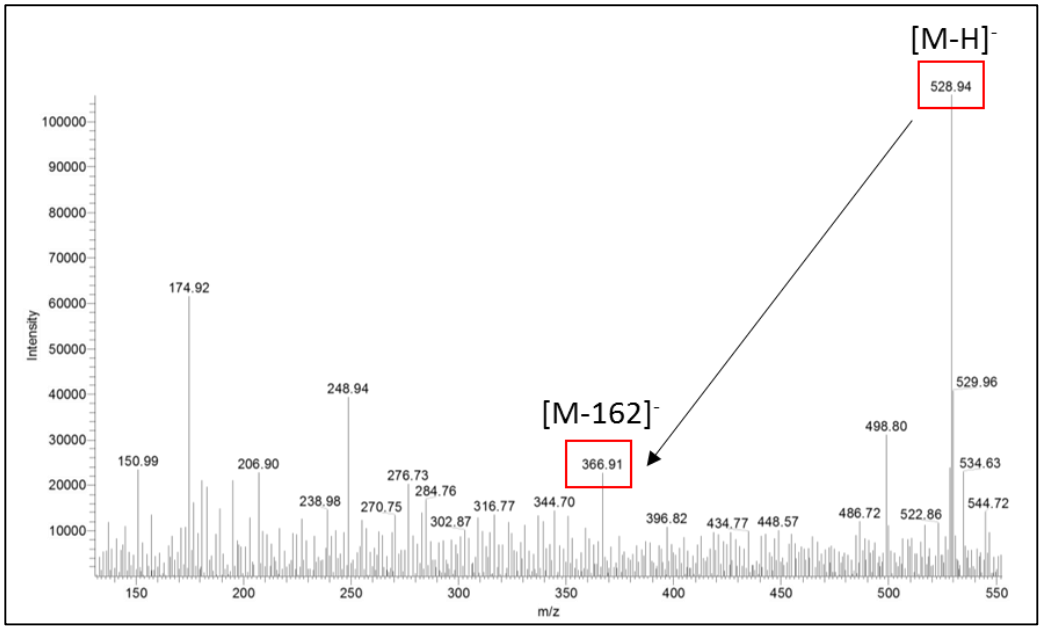


Figure 59. Mass spectrum of curcumin monoglucoside, m/z 528.9 $[M-H]^-$.
There is a characteristic mass loss of a glucose unit at m/z 366.9. $[M-162]^-$.

As monoglucosides of all three curcuminoids were produced in *in vitro* assays, *in vivo* yeast whole cell assays were performed. As only monoglucosides were produced *in vitro* and only monoglucosides were produced when Kaminiga *et al.*[1], performed whole cell *E. coli* assays, it was uncertain whether curcumin diglucosides would be produced in yeast whole cell assays. Due to the financial cost of DMC and BDMC, they were not used as substrates for these larger cell culture assays; only curcumin was used.

When 25 ml *S. cerevisiae* cell cultures, transformed with the empty vector control (pBEVY-EV) were fed with curcumin, no curcumin glucosides were produced, as detected and quantified by LC-MS analysis (Figure 60). When yeast cell culture expressing pBEVY-CaUGT2 was fed with curcumin both mono- and diglucosylated curcumin was produced (see Appendices 3.4 and 3.5 for spectral data). The HPLC retention times of the glucosylated products were comparable with literature values: curcumin monoglucoside eluted at 11.7 minutes and curcumin diglucoside eluted at 7.9, comparable with literature values of 11.8 and 7.4 minutes [1], respectively using the same HPLC column and solvent system (Figure 61). It was apparent from the HPLC chromatograms that the majority of the curcumin substrate was associated with the yeast cells, whereas the mono- and diglucoside products were associated with both the media and the cells (Figure 62). The LC-MS peak area indicated that the majority of the monoglucoside was associated with the cells, whereas the majority of the diglucoside appeared to be in the media (see Figure 60). Curcumin monoglucoside production was rapid, with curcumin monoglucoside being detected at the earliest time point, 0 hours, within the yeast cells. At 24 hours, curcumin

monoglucoside appeared in the media, but at a lower level to what was observed in the cells. In contrast to curcumin monoglucoside, curcumin diglucoside appeared later, at 24 hours, and the levels detected were always higher in the media than in the yeast cells (Figure 60).

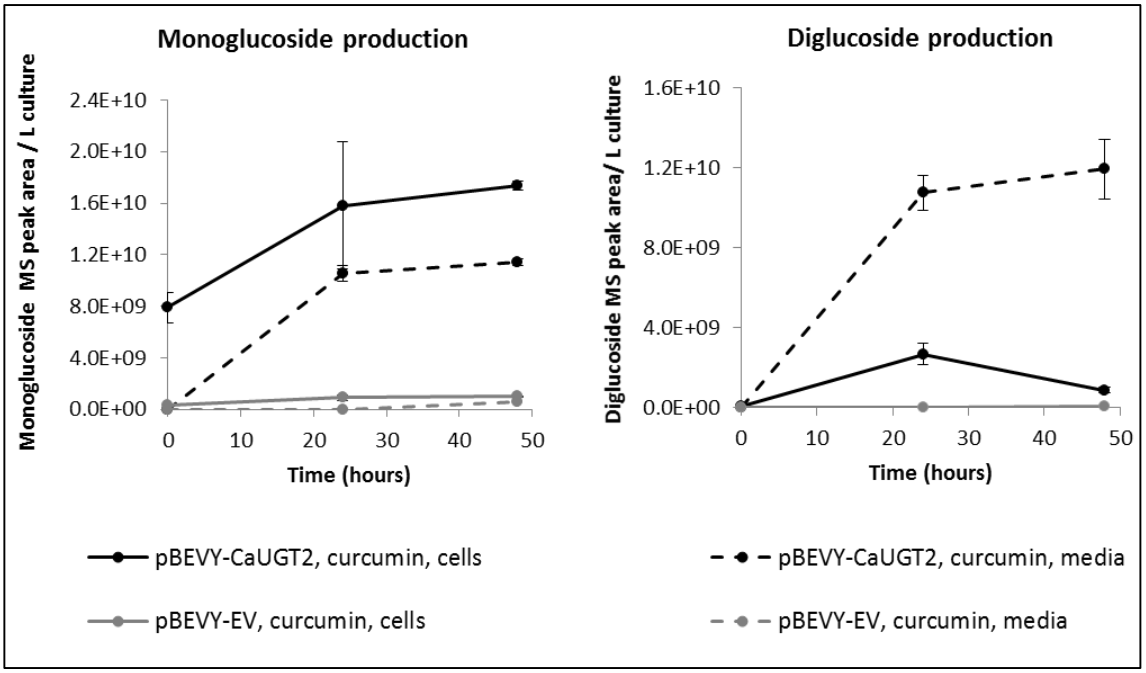


Figure 60. Curcumin mono- and diglucoside production by yeast cells expressing pBEVY-CaUGT2.
Curcumin glucosides were quantified by MS parent ion peak area. Cells transformed with pBEVY-EV did not allow the production of curcumin mono- or diglucosides. Curcumin monoglucosides were produced by cells expressing pBEVY-CaUGT2, the majority of which accumulated within, or was associated with, the cell. Curcumin diglucosides were produced by cells expressing pBEVY-CaUGT2, the majority of which accumulated in the media.

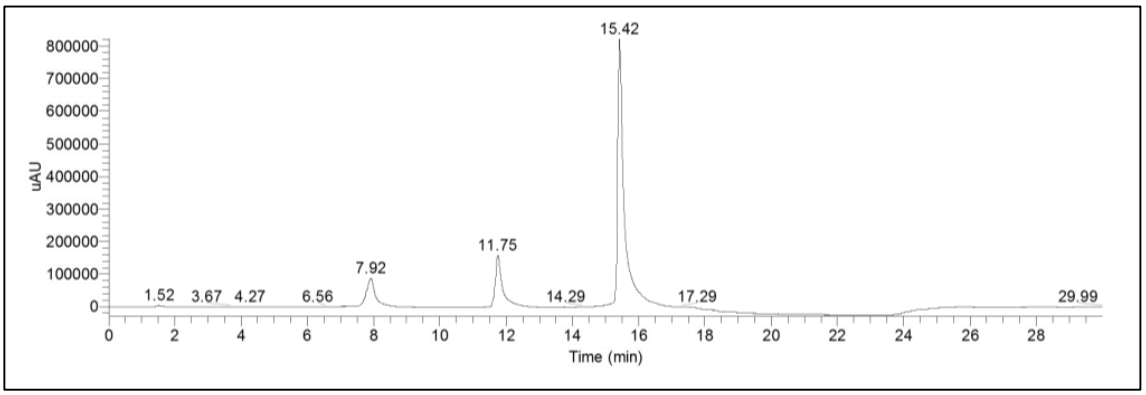


Figure 61. HPLC chromatogram (UV 425 nm) showing curcumin, curcumin monoglucoside and curcumin diglucoside. Curcumin (RT 15.4 minutes), curcumin monoglucoside (RT 11.7 minutes) and curcumin diglucoside (RT 7.9 minutes); the RTs are comparable with literature values using the same HPLC solvent system, at 15.8, 11.8 and 7.4 minutes respectively[1].

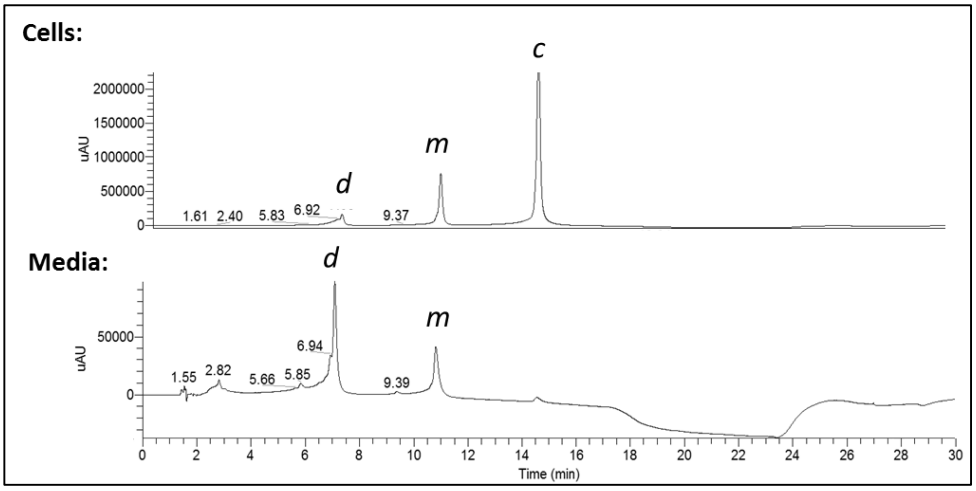


Figure 62. HPLC chromatogram (UV 425 nm) of pBEVY-CaUGT2 cell and media extracts.
24 hours after addition of curcumin to cell culture curcumin (*c*) is most prevalent in the cells, whereas the curcumin monoglucoside (*m*) and curcumin diglucoside (*d*) are in the both the cells and media.

3.2.4 A simple spectroscopic analysis of curcumin glucoside production

Curcumin's characteristic yellow colour quickly depleted after it had been added to the yeast cell culture as curcumin is taken up by the cells [232]. As curcumin was glucosylated and became more water soluble, some of it passed back into the media causing the culture to become significantly more yellow. The difference in intensity of yellow colour between a curcumin fed yeast culture transformed with the empty vector control (pBEVY-EV) and that expressing CaUGT2 (pBEVY-CaUGT2) is apparent in Figure 63.

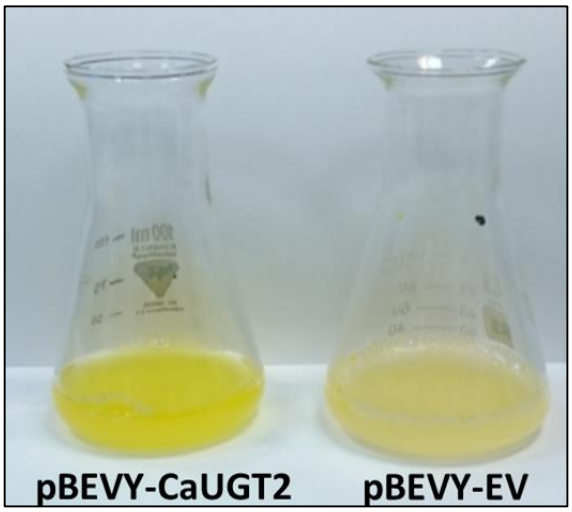


Figure 63. Expression of pBEVY-CaUGT2 in yeast causes the cell culture to become curcumin's characteristic yellow colour.
A visual difference in the intensity of yellow colour between yeast cultures transformed with either pBEVY-CaUGT2 or pBEVY-EV after 48 hours incubation with curcumin is apparent: the pBEVY-CaUGT2 displays a much brighter yellow colour.

As the presence of curcumin glucosides in the media, caused the culture to become yellow, it was investigated whether this colour change could be measured spectrophotometrically to allow simple tracking of the CaUGT2 glucosylation reaction. The absorbance maxima of the curcumin glucosides are the same as curcumin, 425 nm (see Appendix 2.13). pBEVY-CaUGT2 or pBEVY-EV yeast cultures were fed either curcumin, or methanol as control, and the culture media was monitored by its absorption at 425 nm (Figure 64). The methanol fed pBEVY-CaUGT2 showed no absorbance at 425 nm throughout the time course due to the lack of curcumin present. At 0 hours, both of the curcumin-fed cultures exhibited a high initial absorbance around 0.3, due to the presence of curcumin in the media. This decreased over the first 3 hours as curcumin was presumably taken up by the yeast cells, giving an absorbance lower than 0.1. For the rest of the experiment, the pBEVY-EV culture exhibited this low absorbance around 0.1. For the pBEVY-CaUGT2 culture, this low absorbance remained constant until 20 hours but then increased up to 0.4 at 50 hours and remained at this level up to 72 hours. This was presumed to be due to the production of curcumin glucosides and their increased solubility in the media.

To validate that the change in absorbance was due to the presence of curcumin glucosides in the media, the correlation between absorbance and quantifiable curcumin glucosides was determined. Extracts from the media were analysed by LC-MS and the abundance of glucosides, as quantified by HPLC peak area, was compared to the absorbance of the media (Figure 65). Both curcumin monoglucoside (Figure 65A) and diglucoside abundance (Figure 65B) strongly correlated with the absorption of the media measured using a spectrophotometer; $r\,(16) = 1.00$, $p < 0.01$ and $r\,(16) = 0.97$, $p < 0.01$, respectively. Due to these near perfect Pearson correlation values between the absorbance of the media and the abundance of glucosides in the media, it was concluded that the production of curcumin glucosides could be monitored simply using absorbance measurements of the media instead of carrying out HPLC analysis.

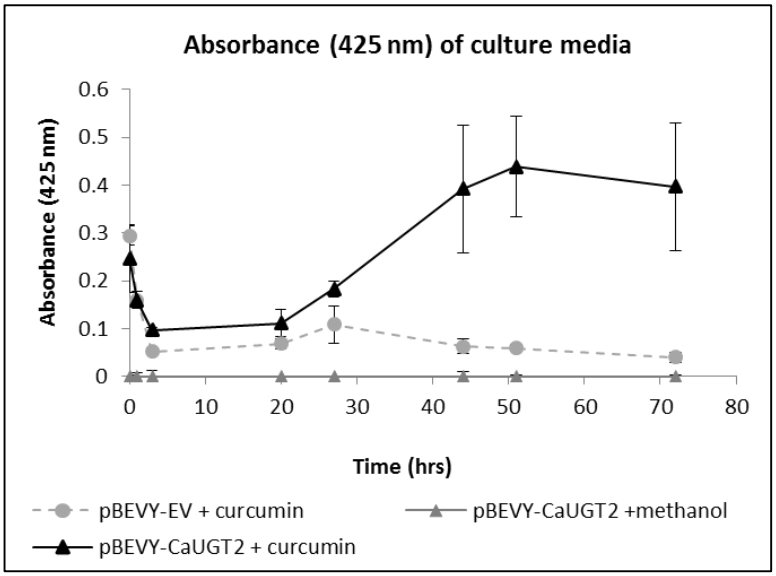


Figure 64. Absorbance (425 nm) of media from yeast cultures expressing pBEVY-CaUGT2 or pBEVY-EV. When pBEVY-CaUGT2 yeast culture was fed methanol, the A_{425} nm reading was 0.0. When pBEVY-EV yeast culture was fed curcumin, the initial A_{425} nm reading is 0.3 but decreased over time as curcumin was absorbed by the cells. When pBEVY-CaUGT2 was fed curcumin, the initial A_{425} nm decreased from 0.3, for the first 20 hours, but then increased due to the presence of curcumin glucosides in the media.

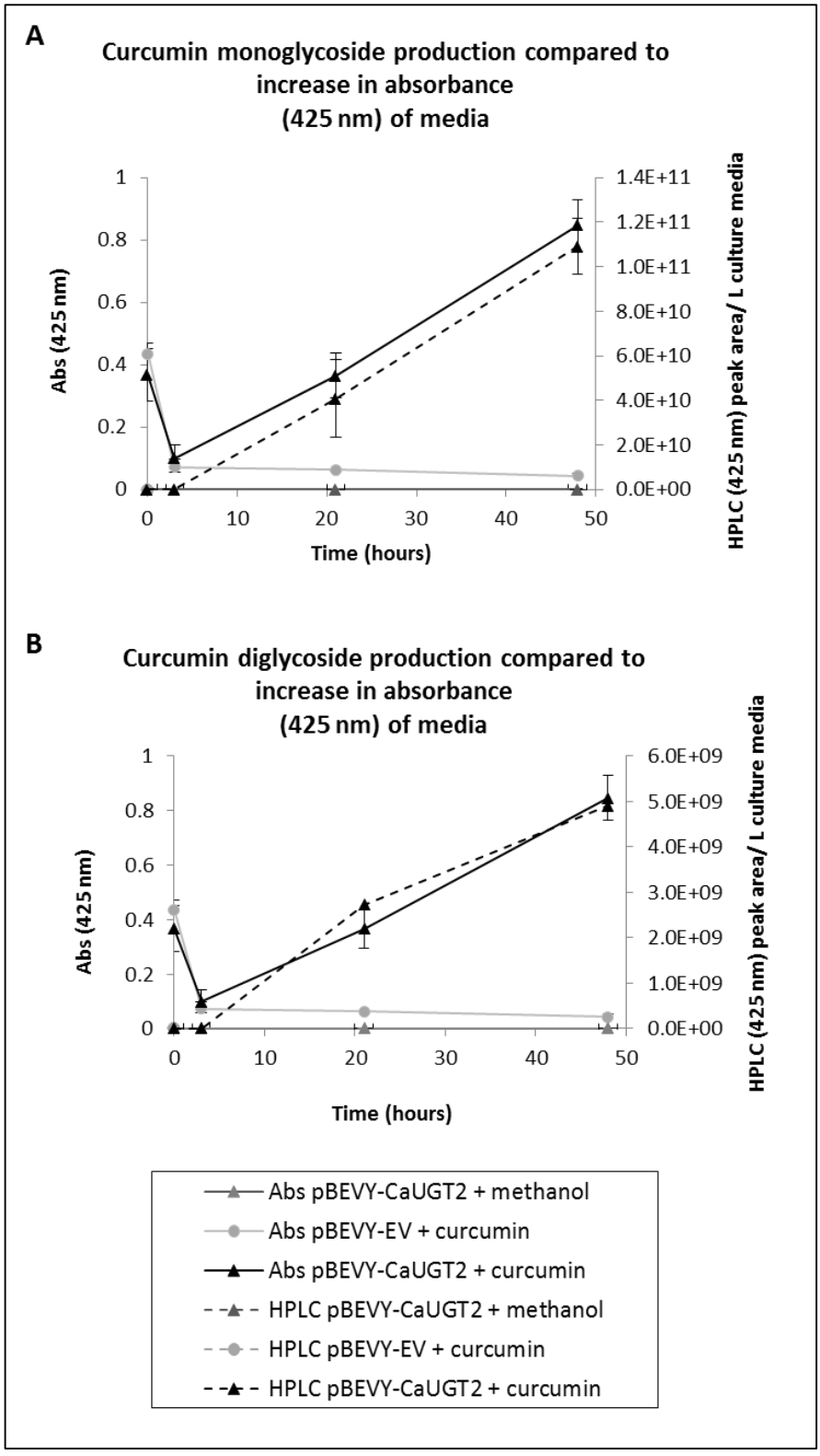


Figure 65. Change in colour of media correlates with presence of curcumin glucosides.
Correlation between abundance of (**A**) curcumin monoglucoside and (**B**) curcumin diglucoside as quantified by HPLC (UV 425 nm) and absorbance (425 nm) of the culture media. Continuous lines show absorbance of the media, measured spectrophotometrically, dashed lines show peak areas of glucoside as quantified by HPLC. Both curcumin monoglucoside (A) and curcumin diglucoside (B) show a very strong correlation between the abundance of the glucosides present in the media and the colour change of the media.

3.2.5 Large-scale production of curcumin glucosides

As yeast cell culture expressing pBEVY-CaUGT2 successfully produced curcumin mono- and diglucosides and this could be simply monitored spectrophotometrically, the reaction was scaled up 200 times from a 25 ml shake flask to a 5 L bioreactor. This required the addition of antifoam but all other media components were kept at the same concentration (see Materials and Methods). The bioreaction was monitored over 72 hours using Applikon software as well as manual measurements of the cell culture growth (OD_{600}) and glucoside production (Abs 425 nm) (Figure 66). The fermentation temperature was kept at 30˚C and the dissolved oxygen was kept above 40 % using the sparger and baffles. For the first 4 hours, up until the addition of curcumin, the culture grew rapidly. Subsequently, for the next 20 hours after curcumin addition, the culture showed very little change with minimal cell growth. However, after 30 hours, growth increased up to 50 hours, shown by the increase in optical density (OD) (600 nm) and decrease in dissolved oxygen caused by the increased demand for oxygen. Throughout the bioreaction, the pH decreased from the pH of the media, 4.5, to 2.5; the most rapid decrease occurred during the period of rapid cell growth. Curcumin glucoside production correlated with culture growth as monitored by absorbance at 425 nm. When the fermentation was ended, the OD and Abs (425 nm) were still increasing suggesting there is potential for further increases in yield.

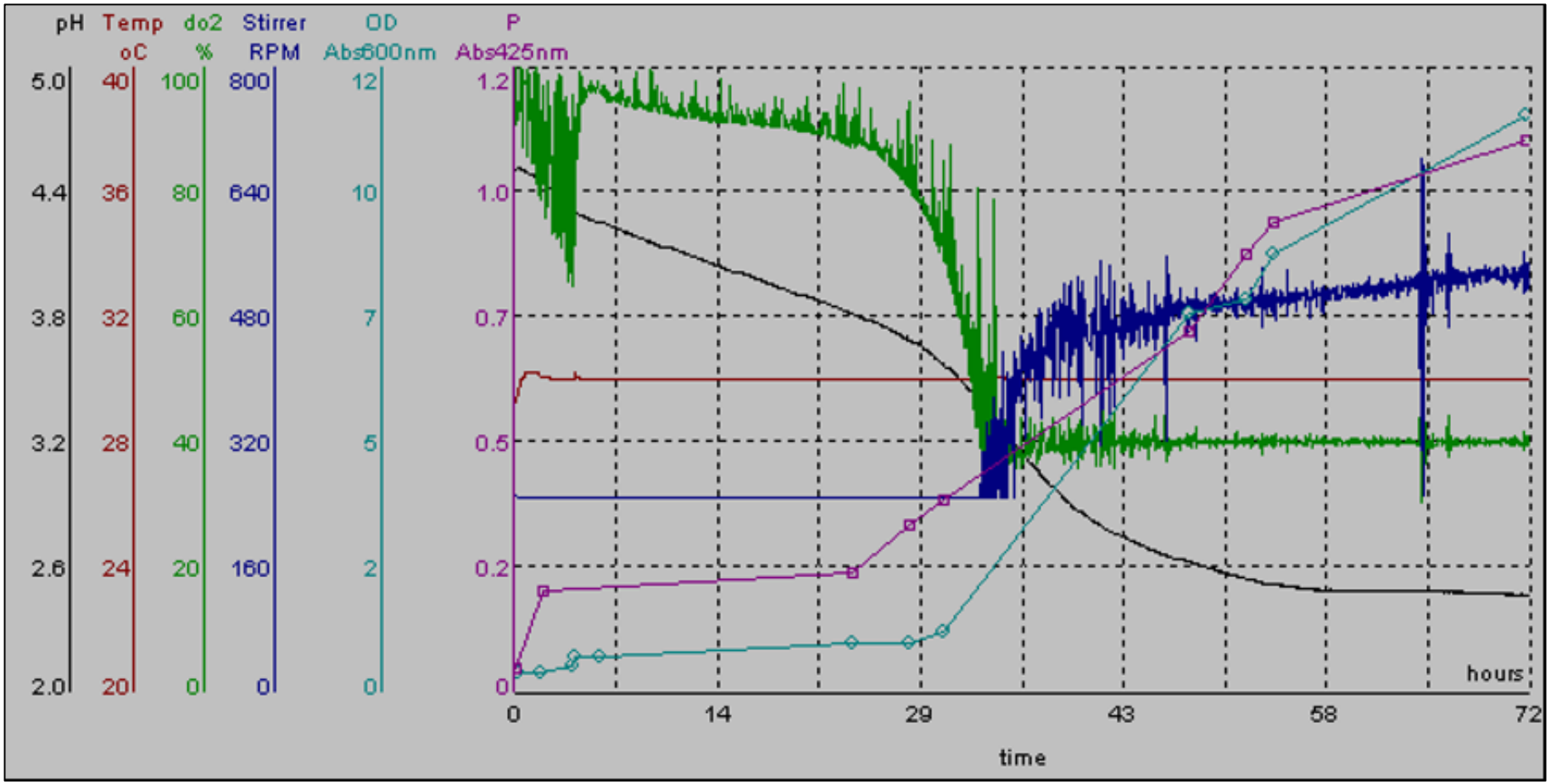


Figure 66. 72-hour pBEVY-CaUGT2 *S. cerevisiae* bioreaction profile.
Recorded and generated by Applikon fermenter software. pH (black), temperature (burgundy), dissolved oxygen (green) and stirrer speed (blue) were automatically monitored, with readings every hour. Optical density 600 nm (OD) (teal) and glucoside production (magenta) were measured manually.

To obtain chemical standards of the glucosides and allow an approximate yield of the large scale bioreaction to be calculated, the aim was to purify the glucosides from the media. To reduce curcumin aglycone recovery and aid purification of glucoside product, *n*-butanol was chosen as a more polar solvent to perform liquid-liquid extraction of the glucosides over ethyl acetate. The butanol extract was concentrated by rotary evaporation and the resultant crude product was analysed by LC-MS: all glucosides had been extracted from the media (Figure 67A). However, curcumin aglycone and other media components were also extracted from the media along with the glucosides (Figure 67B). The majority of the curcumin glucoside isolated appeared to be in the monoglucosylated form. To purify the glucosides from one another and the curcumin aglycone, flash chromatography was considered. Elution conditions were first tested using a 100 mg solid phase extraction (SPE) column. Using different percentages of methanol, curcumin mono- and diglucoside could be separated from each other as well as from contaminants: therefore, a simple methanol gradient was used for flash chromatography. The resultant fractions were collected and those containing curcumin mono- or diglucoside were combined (Figure 68A and Figure 68B). Although the glucosides were successfully separated, there were still contaminants in the samples. For further purification prep-HPLC was carried out. As less curcumin diglucoside was recovered from flash chromatography compared to curcumin monoglucoside (64 mg compared to 168 mg), prep-HPLC was only carried out upon curcumin monoglucoside. Unfortunately, the majority of the curcumin monoglucoside appeared to elute at an earlier RT than expected (6.5 minutes, instead of ~11 minutes) and appeared as various conjugates, such as complexing with sodium formate ($[M–H+68]^-$). Only one fraction contained pure curcumin monoglucoside which eluted at the expected RT and with the correct parental mass ion (Figure 69). As a result, the amount of curcumin monoglucoside purified was low, 9.3 mg. Using the purified curcumin monoglucoside, a standard curve was plotted relating the amount of curcumin monoglucoside with the spectrophotometric absorbance at 425 nm (Figure 70). From this, the amount of curcumin monoglucoside produced during the fermentation was estimated to be 160 mg (32 mg L^{-1} culture), an approximate yield of 60 %.

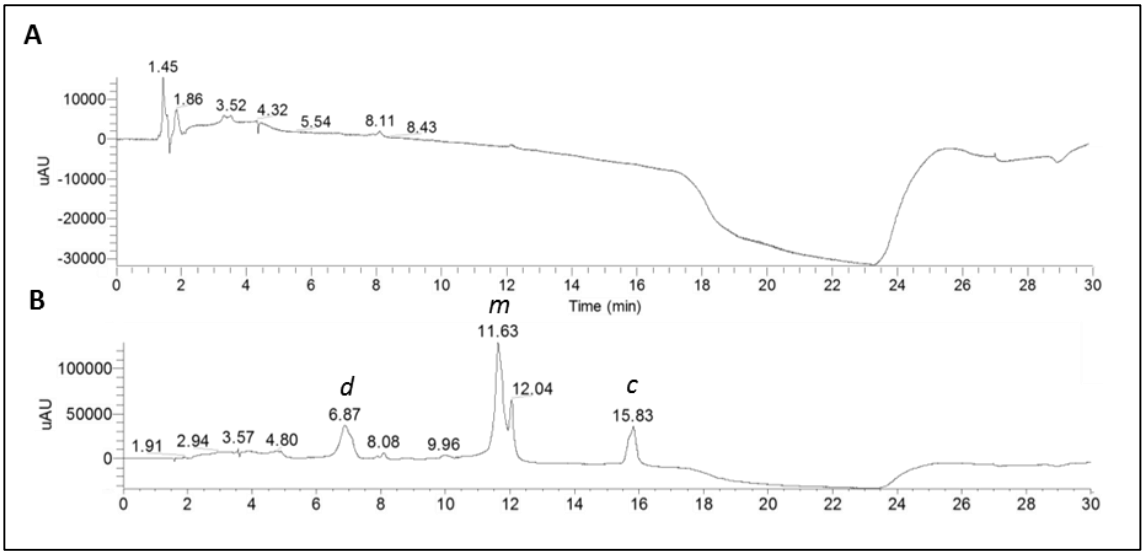


Figure 67. Curcumin glucoside extraction from media.
HPLC chromatograms (425 nm) showing: (A) fermentate after butanol extract; (B) butanol extract containing curcumin (*c*, RT 15.8 minutes), curcumin monoglucoside (*m*, RT 11.6 minutes) and curcumin diglucoside (*d*, RT 6.9 minutes).

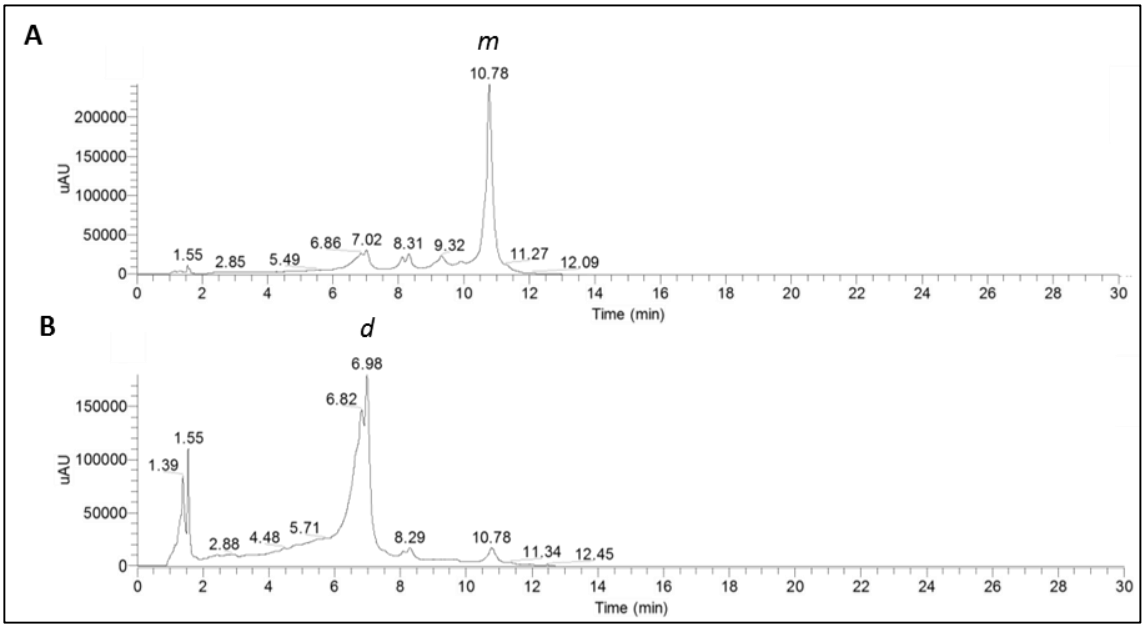


Figure 68. Curcumin glucoside flash chromatography purification.
HPLC chromatograms (425 nm) showing: (A) pooled curcumin monoglucoside (*m*, RT 10.8 minutes) fractions from flash chromatography; (B) pooled curcumin diglucoside (*d*, RT 7.0 minutes) fractions from flash chromatography. Contaminants are apparent in both pools.

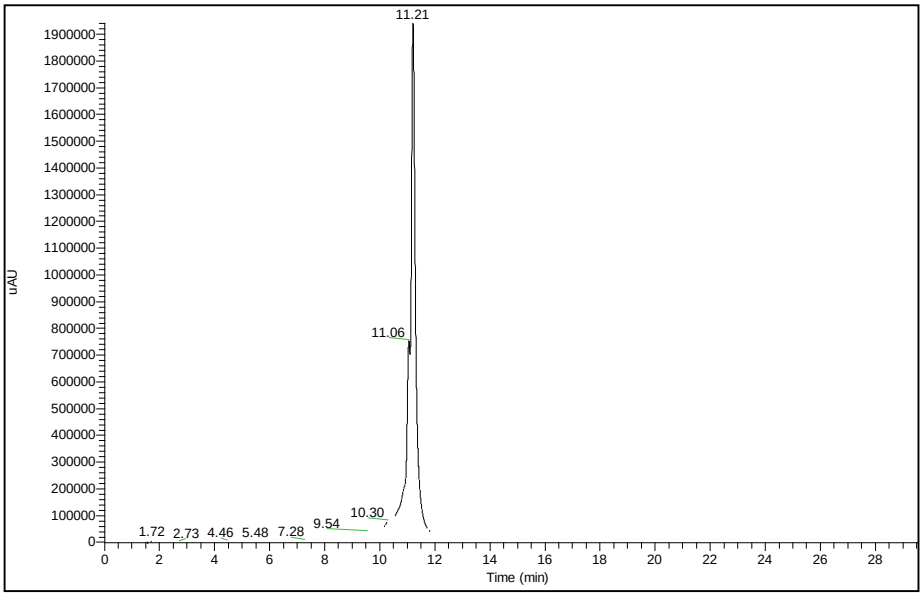


Figure 69. HPLC UV (425nm) chromatogram of pure curcumin monoglucoside.
Curcumin monoglucoside elutes at RT 11.21 minutes. The peak at 11.06 minutes is also curcumin monoglucoside.

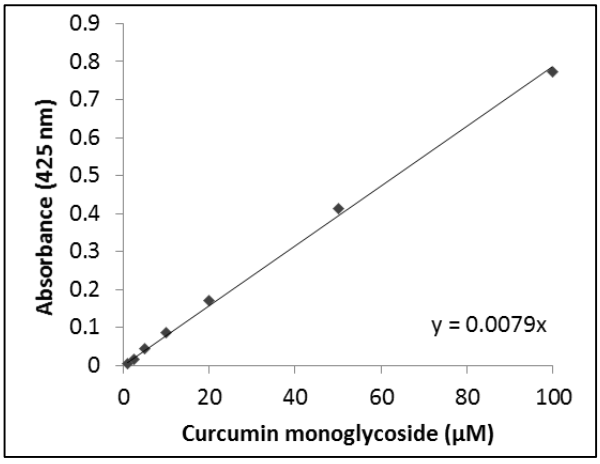


Figure 70. Linear relationship between the concentration of curcumin monoglcyoside in solution and the absorbance (425 nm) measured spectrophotometrically.
The data are technical means ± standard deviations (n = 3).

3.2.6 Optimising media conditions for curcumin glucoside production

During pBEVY-CaUGT2 yeast whole cell assays, cell growth was retarded for approximately 24 hours after curcumin was added to the bioreaction, as well as a concomitant lag in detectable glycoside production (see Figure 66). As it appeared that the production of curcumin glucosides increased with cell growth, different conditions were investigated to ascertain their affect upon yeast cell growth and curcumin glucoside production.

The effect of different carbon sources was initially investigated. The standard carbon source used to grow the pBEVY-CaUGT yeast culture was raffinose. Because the expression of CaUGT2

from the pBEVY vector is induced by galactose and repressed by glucose [233], these two sugars were investigated first. To assist in optimisation of the growth media a series of questions were considered: Was induction tightly controlled? Was addition of galactose necessary? What happens when glucose, an inhibitor of induction, was added to the media? Answering such questions would allow an optimal and cost-effective sugar to be chosen as a carbon source for the fermentation.

Initially, shake flask cultures expressing pBEVY-CaUGT2 or pBEVY-EV were fed curcumin or methanol, plus or minus 2 % (w/v) galactose. It was expected that induced and non-induced cultures may grow similarly, but the production of curcumin glucosides should be much higher in the induced culture as this should contain much more expressed enzyme. As expected, the induced and non-induced cultures grew at a similar rate and to a similar extent (Figure 71A). Despite the prediction that curcumin glucoside production would increase due to increased CaUGT2 expression, glucosides were detected to a similar level irrespective of whether galactose induction was carried out (Figure 71B). A possible explanation for this observation is that CaUGT2 expression is occurring in the absence of galactose, possible due to a leaky GAL1 promoter. To test for this, RNA was extracted from induced and non-induced cells, from which cDNA was synthesized. PCR, using CaUGT2 specific primers was carried out upon the cDNA and the products were separated using agarose gel electrophoresis (Figure 72). There were no PCR products evident for the empty vector controls, whether induced or not, as expected. The positive plasmid control, gave a product at the expected size, 0.7 kB. Induced pBEVY-CaUGT2 transformed cells also gave a PCR product as anticipated. If expression of CaUGT2 from the pBEVY vector was tightly controlled by galactose induction, CaUGT2 RNA should not be detectable from the non-induced culture. However, a PCR product was detected thus indicating that expression of CaUGT2 occurs in the absence of galactose. When the results are corrected for template abundance using the constitutive actin control, CaUGT2 mRNA is more abundant in the culture that had been induced with galactose. Nonetheless, even though levels of CaUGT2 mRNA are greater in the induced cultures, CaUGT2 expression in the non-induced cultures allows a similar levels of curcumin glucosylation.

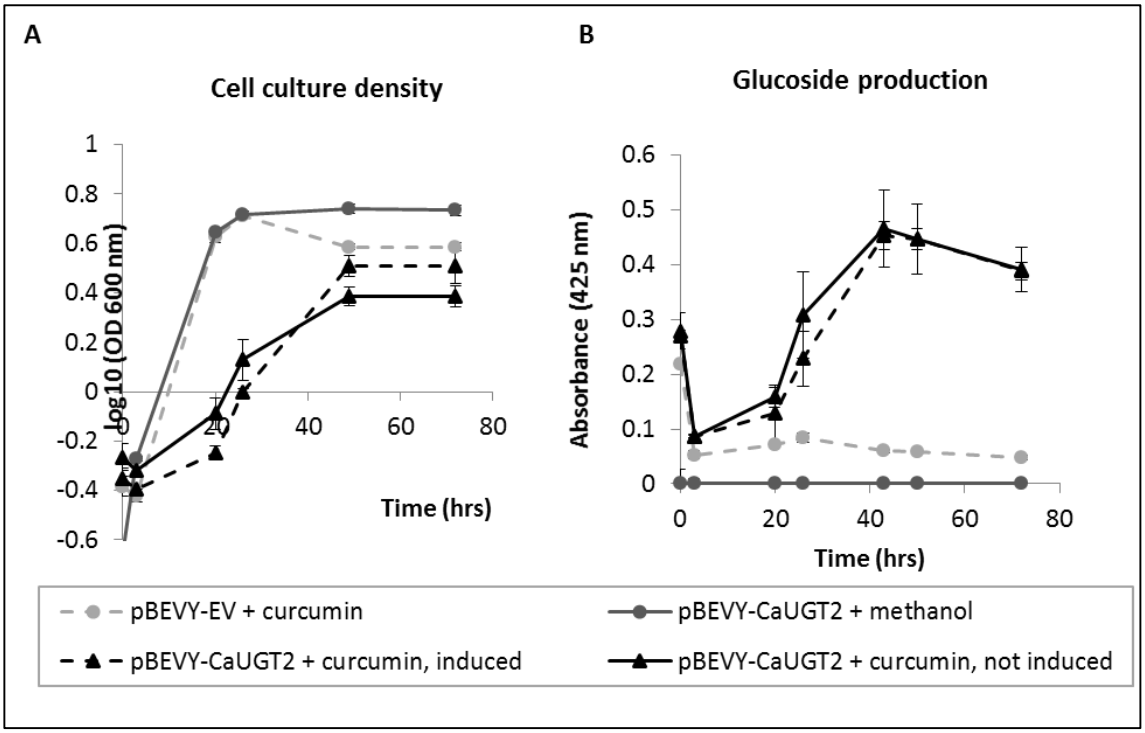


Figure 71. Investigating the induction of CaUGT2 expression upon cell growth and glucoside production.
Comparison of the cell culture growth (A) and curcumin glucoside production (B) of induced (with galactose) and non-induced (without galactose) yeast cultures expressing pBEVY-CaUGT2 or pBEVY-EV. Methanol was used as a control substrate. The data are means ± standard deviations (n = 3).

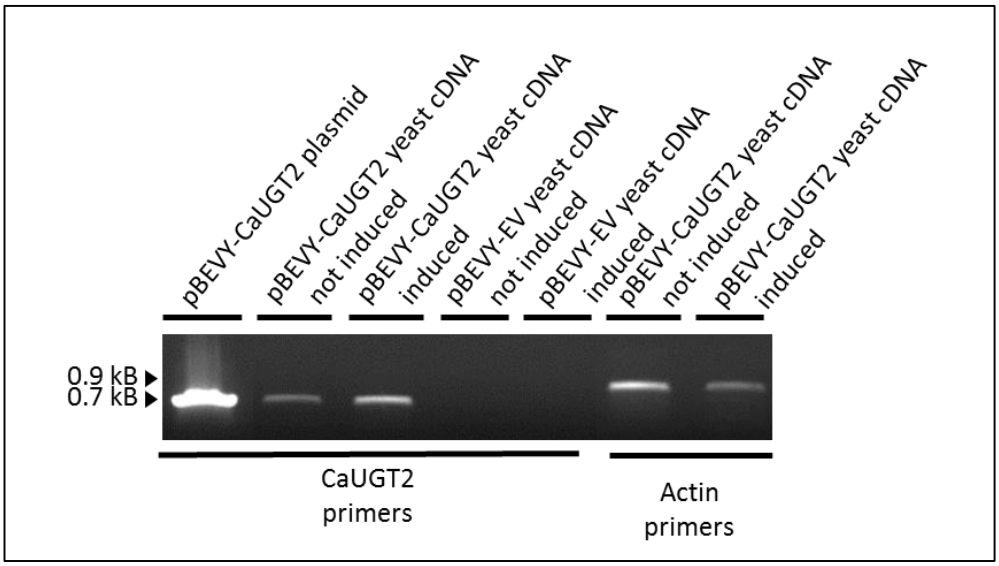


Figure 72. Expression of CaUGT2 when induced and non-induced.
Agarose gel showing 0.7 kB PCR products when using CaUGT2 specific primers upon cDNA derived from pBEVY-CaUGT2 or pBEVY-EV transformed cells that were, or were not, induced. PCR products are evident for pBEVY-CaUGT2 derived cDNA, whether expression was induced by addition of galactose or not. Actin was used as a control, the PCR product of which is shown at 0.9 kB.

As curcumin glucosides were produced without galactose induction (see Figure 71B) the effect of glucose repression upon pBEVY-CaUGT2 expression was investigated. Glucose represses expression from the GAL1 promoter [233], therefore it was suggested that addition of glucose to the media may reduce the amount of CaUGT2 expressed and the amount of curcumin glycoside produced. However, adding 2 % (w/v) glucose to the culture medium was also expected to increase cell growth, as glucose is the preferred carbon source in *S. cerevisiae* [234]. When glucose replaced raffinose in the media there was no difference in pBEVY-CaUGT2 cell culture growth (Figure 73A). There was also no difference in curcumin glucoside production compared to the raffinose control, again eluding to a possible lack of control of CaUGT2 expression by the GAL1 (Figure 73B).

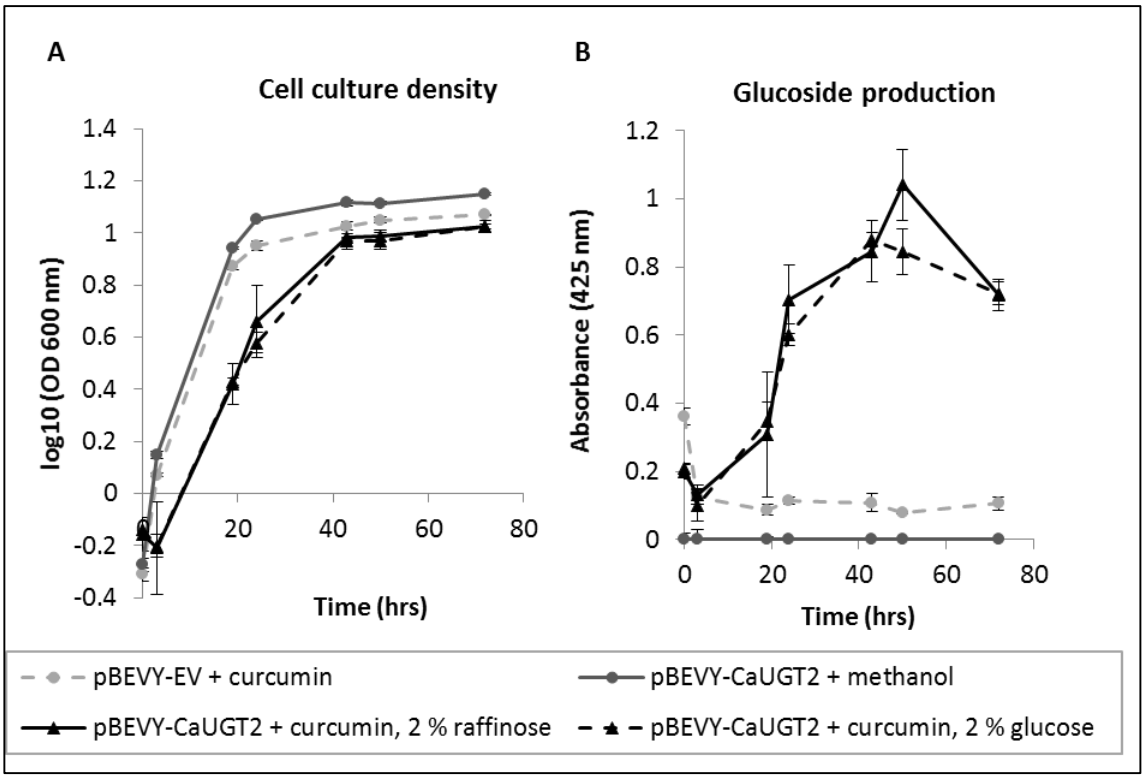


Figure 73. Investigating the repression of CaUGT2 expression upon cell growth and glucoside production.
Comparison of the cell culture growth (A) and curcumin glucoside production (B) of yeast cultures expressing pBEVY-CaUGT2 or pBEVY-EV growing in 2 % (w/v) raffinose yeast media (pBEVY-CaUGT2 + curcumin, 2 % raffinose), or 2 % (w/v) glucose yeast media (pBEVY-CaUGT2 + curcumin, 2 % glucose). Methanol was used as a control substrate. The data are means ± standard deviations (n = 3).

The effect of adding supplemental raffinose to the pBEVY-CaUGT2 culture medium was also investigated. It was reasoned that by increasing the carbon source, cell growth rates should be promoted and hence increase the production of curcumin glucosides. 1 ml of additional 40 % (w/v) raffinose was added to the 25 ml cell cultures twice, at 24 hours 48 hours, giving an overall total of 6.5 % (w/v) raffinose. Cell growth and curcumin glucoside production was again monitored spectrophotometrically (Figure 74). There was no increase in pBEVY-CaUGT2 cell growth (Figure 74A) however, an increase in curcumin glucoside production compared to cultures with the basal amount of raffinose was observed (Figure 74B).

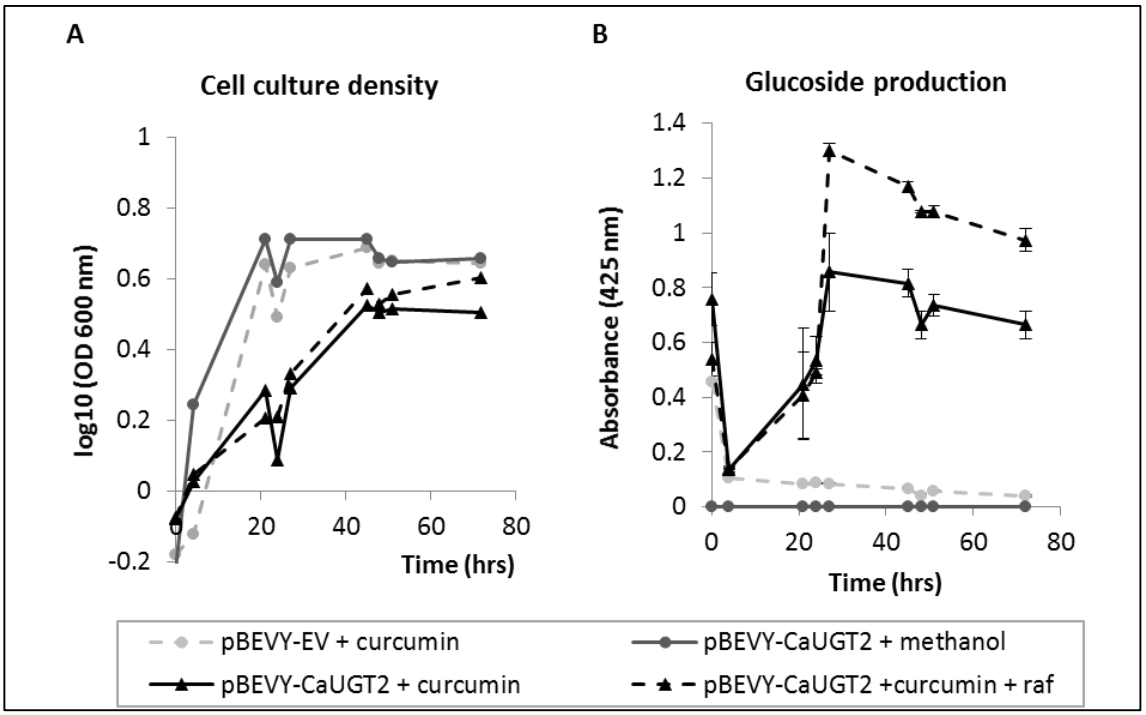


Figure 74. The effect of supplemental raffinose in the yeast media upon cell growth and glucoside production. Comparison of the cell culture growth (A) and curcumin glucoside production (B) of yeast cultures expressing pBEVY-CaUGT2 or pBEVY-EV grown in standard yeast media (pBEVY-CaUGT2 + curcumin, 2 % raffinose), or media with additional raffinose (pBEVY-CaUGT2 + curcumin, 6.5 % raffinose). Methanol was used as a control substrate. The data are means ± standard deviations (n = 3).

As increasing, or changing, the carbon source did not affect cell growth (Figure 71 to Figure 74), it was hypothesized that other factors in the media were limiting growth and curcumin glucoside production. It had been reported that curcumin accumulates in the endoplasmic reticulum of *S. cerevisiae* where it chelates iron, depleting intracellular iron and reducing growth rate [232]. It was theorised that since glucosylated curcumin is considered to have a greater ability to chelate iron leading to its use in iron overload therapy [111], this increased iron chelating effect could be retarding the growth of pBEVY-CaUGT2 *S. cereviseae* cells. To test this, cultures were set up with supplementary iron (50 µM $FeSO_4$) and their growth and curcumin glucoside production were monitored. The iron treatment did allow the curcumin-fed pBEVY-CaUGT2 culture to grow at a similar rate and extent as the pBEVY-EV control cultures (Figure 75A), although additional iron did not allow an increase in glucoside production (Figure 75B).

These results suggest that glucoside production is not limited by culture growth, but by some other factor. This may be something that limits the enzymatic reaction, such as the abundance of UDP-glucose, which once overcome, should allow curcumin glucoside production to increase with cell density. Such an explanation is also consistent with the fact that the induced and non-induced cultures produce the same amount of curcumin glucosides even though the induced culture produces more UGT mRNA and probably more protein.

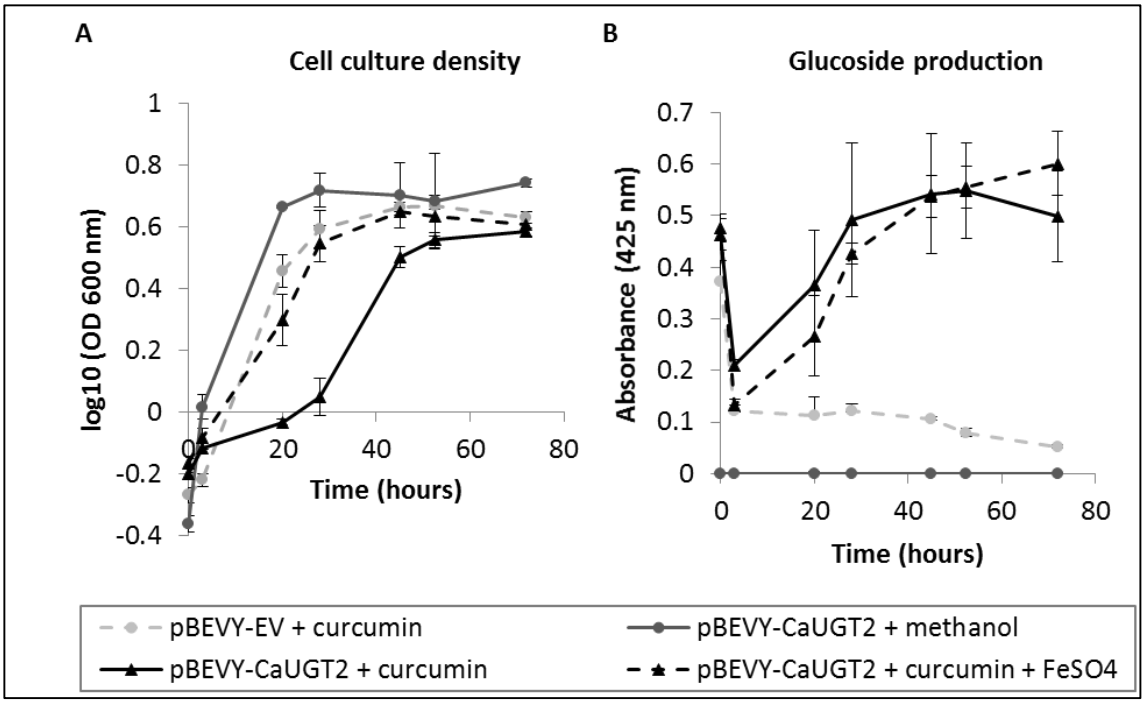


Figure 75. The effect of supplemental iron ($FeSO_4$) in the culture media upon cell growth and glucoside production. Comparison of the curcumin glucoside production (A) and cell culture growth (B) of yeast cultures expressing pBEVY-CaUGT2 or pBEVY-EV growing in yeast media (pBEVY-CaUGT2 + curcumin), or iron supplemented yeast media (CaUGT2 + curcumin + $FeSO_4$). Methanol was used as a control substrate. The data are means ± standard deviations (n = 3).

3.3 Discussion and future work

Glucosylation offers a route to increase a molecule's solubility in water therefore increasing its bioavailability and drug commercialisation potential. Due to the poor bioavailability of curcuminoids, glucosylation has been considered as a method to improve their water solubility and therefore their pharmacokinetic properties. Previous work utilising recombinantly expressed CaUGT2 from *C. roseus* had shown that curcumin diglucoside was only biosynthesized in cell-free *in vitro* assays [1,235]. This chapter aimed to develop a scalable platform for the production of both mono- and diglucosylated curcuminoids.

Initially, *C. longa* was investigated as a source of a novel curcuminoid UGT. Despite two hits in the NCBI *C. longa* transcriptome database, based on sequence similarity to CaUGT2, there was no enzymatic or chemical evidence of *C. longa* being able to produce glucosylated curcuminoids. In-depth LC-MS analysis of metabolite extract from 13 accessions of *C. longa* from 9 provinces across Vietnam also showed no evidence of any other curcuminoid bio-conjugates. The data did however show a strong correlation in curcumin, DMC and BDMC abundance (see Table 3, r (28) = 0.97-1.00, $p < 0.01$). This was expected due to being metabolites from the same pathway, as has been shown in previous research by Ramirez-Ahumada *et al.*, 2006 [236] and Xie *et al.*, 2009 [237]. Additionally, there was a strong correlation in the amount of turmerones and curcuminoids (see Table 3). This lead to the proposal that, like ginger, *C. longa* may sequester its phenolic bioactive metabolites within liposomes of oily secondary metabolites, in *C. longa*'s case, turmerones. Metabolite data gathered by Xie *et al.* [237] from rhizomes subjected to 16 different growth conditions showed the presence of distinct metabolic modules, in which the curcuminoid production profiles also appeared to correlate with that of the turmerones, ar-turmerone and curlone. To further validate this hypothesis, microscopy could be carried out to determine the presence of such liposomal capsules. Additionally, the co-regulation of the *C. longa* curcuminoid biosynthetic pathway and sesquiterpene synthases involved in turmerone synthesis, as identified by Koo *et al.* [238], could be investigated using qPCR.

Since it appears that *C. longa* does not naturally produce glucosylated curcuminoids, CaUGT2 offered the possibility to produce these in a heterologous system [1]. CaUGT2 had been shown to mono- and diglucosylate curcuminoids in recombinant *E. coli* protein extracts but only produce monoglucosides in equivalent whole cell studies [1,138]. In the current study, when CaUGT2 was heterologously expressed in *E. coli*, it glucosylated curcumin, as previously reported [1], and it was also found to glucosylate DMC and BDMC (see Figure 56). This was the first time it has been demonstrated that CaUGT2 has activity to curcuminoids other than curcumin. This made it an excellent candidate for the production of curcuminoid glucosides in a scalable microorganism such as *S. cerevisiae*.

In vitro assays using *S. cerevisiae* protein extract containing CaGUT2 only produced the curcumin monoglucoside (see Figure 58) whereas *in vivo* shake flask assays allowed the production of both curcumin mono- and diglucosides (see Figure 61). Curcumin monoglucosides were detected earlier than curcumin diglucosides which is theorised to be due to CaUGT2 having a lower K_m for curcumin than curcumin monglucoside: 19.0 µM compared to 63.3 µM [1]. The differential production of curcumin mono- and diglucosides depending on the heterologous host and *in vivo* or *in vitro* conditions is very interesting. When added to the fact that in yeast cell cultures assays, the glucoside products are present in the media (see Figure 60 and Figure 62) it seems to be that compartmentalisation may be important for diglucoside production. The fact that diglucosides of curcumin are more soluble in aqueous media than curcumin monoglucosides may be a factor [138]. Membrane bound organelles, or an as yet unknown transport system, may be important for the production of curcumin diglucosides and this could explain the difference in the ability of *E. coli* and *S. cereviseae* to produce the diglucoside when expressing CaUGT2. Transport and organelle specific biosynthesis of glucosylated curcuminoids could be investigated further as a means of increasing glucoside yield or directing the production of one glucoside (mono- or di-) over the other. Interestingly, the highest yielding synthetic chemistry approach to synthesizing curcuminoid diglucosides utilised a biphasic solvent system which separated substrates and products [135]. This rationale could be applied to a multiphase extractive fermentation to aid catalysis and extraction. Notably, this strategy proved successful for Paddon *et al.* [27], for the production of artemisinic acid.

As both curcumin mono- and diglucosides could be produced by yeast cells expressing CaUGT2, the culture was successfully scaled up 200 times from a 25 ml shake flask to a 5 L bioreactor. It is usually scaling up from shake flasks to small bioreactors that is the most problematic stage of scaling bioreactions [239]. The 5 L bioreaction gave a greater production of glucoside per ml of culture compared to 25 ml shake flasks as shown by a higher absorbance (425 nm) measurement: 1.1 compared to an average of 0.7 (see Figure 64 and Figure 66). Furthermore, 25 ml shake flask cultures reached peak glucoside production after approximately 48 hours, whereas glucoside production was still increasing after 72 hours in the 5 L bioreactor. Therefore, to achieve the maximum potential yields, longer bioreactions should be investigated to determine a definite end of fermentation.

Using purified curcumin monoglucoside, the yield of the 5 L bioreaction was estimated to be 60 % or 32 mg L^{-1}. This is an approximate yield as the amount of purified curcumin monoglucoside was low. Efforts need to be made to optimise the purification of the glucosides from the culture media. This could involve the previously mentioned biphasic extractive fermentation technique, investigating other liquid-liquid extractions, large scale SPE or silica column chromatography or different flash chromatography and prep-HPLC methods. Reduction of sodium formate curcumin

monoglucoside adducts could also be investigated by using a different prep-HPLC solvent additive instead of formic acid. The addition of formic acid may also have caused acid catalysed glycosidic bond hydrolysis reducing the amount of recoverable glucosides.

The fact that curcumin glucoside products could be detected in the pBEVY-CaUGT2 yeast fermentate media, allowed a spectrophotometric assay to be developed to allow simple monitoring of the accumulation of product. This allowed the effect of different media compositions upon glucoside production to be monitored in order to optimise the process. Different carbon sources were explored in an attempt to increase cell growth since it is possible that this could also lead to an increase in glucoside production. The standard yeast media contained 2 % (w/v) raffinose, as this does not interfere with the expression of CaUGT2 from the pBEVY vector, which under the control of the GAL1 promoter is induced by galactose and repressed by glucose. First, galactose induction was investigated. There was no difference between induced and non-induced cultures (Figure 71) so galactose expression of CaUGT2 was hypothesized to be leaky; this was verified by RT-PCR (see Figure 72). However, to be certain that the expression of pBEVY-CaUGT2 results in CaUGT2 protein production, western blot analysis should be carried out.

Secondly, to further examine the expression of CaUGT2, the outcome of growth and curcumin glucoside production was monitored when 2 % (w/v) glucose was added to the yeast media. Glucose is known to repress expression from the GAL1 promoter [233], so the expected balance between increased cell growth and CaUGT2 expression repression was unknown. Using glucose instead of raffinose as a carbon source had no effect on the growth of the culture or curcumin glucoside production (see Figure 73). This inferred something else was likely to be limiting cell growth and enzyme catalysis.

Thirdly, the effect of supplemental raffinose was explored. Interestingly, supplemental raffinose did not increase cell growth, but did increase curcumin glucoside production (see Figure 74). This was not expected as in yeast raffinose is hydrolysed by invertase (β-fructosidase) to fructose and melibiose [240]. Fructose can be utilised for respiration, but the disaccharide melibiose cannot be broken down by *S. cerevisiae* as it lacks α-galactosidase [241]. Therefore, it was unknown how supplemental raffinose increased the production of curcumin glucosides.

Curcumin is known to inhibit the growth of *S. cerevisiae* by increasing the G1 phase of the cell cycle due to intracellular iron chelation within the endoplasmic reticulum [232]. Iron starvation was suspected to be heightened by curcumin glucosides as they exhibit greater effectiveness as an iron overload therapy compared to the curcumin aglycone [111]. Supplementing the pBEVY-CaUGT2 culture media with $FeSO_4$ allowed an increase in yeast cell culture growth (see Figure 75A).

However, an increase in culture growth did not translate to an increased production of curcumin glucosides (see Figure 75B).

As the amount of curcumin glucosides produced was the same, regardless of induction media, it was concluded that the amount of enzyme was not limiting the production of glucoside. This was also evident when supplemental iron increased growth but not glucoside production (see Figure 75) and supplemental raffinose increased glucoside production without an increase in cell growth (see Figure 74). Consequently, other factors which may affect the enzymatic reaction should be investigated such as varying the amount of UDP-glucose, especially as the by-product of the glucosylation, UDP, is an end-product inhibitor of UGT activity [242,243]. Masada *et al.* [142], created a glucosylation system in *E. coli* which reduced UDP inhibition and increased UDP-glucose substrate by expressing Arabidopsis sucrose synthase 1 (catalyses breakdown of sucrose to UDP-glucose and fructose) which increased the glucoside yield several fold. A similar recycling system, which relies upon the supplementation of sucrose could be used in yeast and would allow a cheaper carbon source, which does not affect expression of CaUGT2, to be used. Once the enzymatic reaction is no longer limiting, increasing cell growth, by addition of iron and using a favourable carbon source should then allow an even greater increase in glucoside production.

There are of course, a range of other questions, the answers to which could improve curcumin glucoside production. What is the optimum amount of curcumin substrate for the growth of the culture? What transporters are involved in the influx of curcumin and efflux of curcumin glucosides to and from the yeast cell? What pH, temperature and oxygen concentrations are optimal? Furthermore, the system could be expanded, by adding other glycosyltransferases such as CaUGT3 that can form curcumin gentiobiosides with even greater water solubility [143]. The best chemical synthesis of curcumin diglucoside yields 71 % [135]. With optimised bioreaction and extraction conditions, the yield of curcumin glucoside production by *S. cerevisiae* expressing CaUGT2 is expected to compete with the top performing curcumin glucosylation reactions documented to date.

Chapter 4. Materials and Methods

4.1 Materials

All chemicals were purchased from Sigma-Aldrich unless stated otherwise. Phenyl-propanoyl-CoA esters were purchased at HPLC grade from PlantMetaChem (TransMIT Gesellschaft für Technologietransfer, Gießen, Germany). Molecular biology reagents, enzymes and kits were purchased from New England Biolabs, Promega, Thermo Scientific or Qiagen. Oligonucleotide primers were synthesized by and purchased from Eurofins MWG Operon. Gene synthesis was carried out by GenScript or Invitrogen. Reagents were used as solids or formulated in ultrapure water (18.2 MΩ cm-1) or HPLC grade methanol. All growth media was autoclaved or filter sterilised (0.42 µm) prior to use.

4.2 Instrumentation

Polymerase chain reactions were performed using a Bio-Rad DNA Engine PTC-200 Peltier Thermal Cycler. Nucleic acid quantification was carried out using a NanoDrop 1000 spectrophotometer. Spectrophometric measurements were taken using a Thermo Scientific Evolution 60S UV-Visible spectrophotometer. q-PCR was performed using a Bio-Rad My-IQ Single Color RT-PCR detection system. High-performance liquid chromatography was performed using a Thermo Finnigan Surveyor Plus HPLC System equipped with Xcalibur data analysis software. Mass spectrometry was performed using a Thermo Scientific LCQ Fleet ion trap mass spectrometer. Flash chromatography was carried out using a Biotage Isolera Prime system. Preparative HPLC was performed using an Agilent Varian Modular Analytical HPLC System. 5 L bioreaction was carried out using a 5 L Applikon bioreactor. Radioactivity scintillation counting was performed using a Packard Trib-Carb 2900TR Liquid Scintillation Analyzer.

4.3 Methods

4.3.1 Investigating *C. longa* 4CLs

Phylogenetic analysis of five putative 4CL ESTs from *C. longa*

Phylogenetic analysis was carried out upon the following sequences: *Os*4CL1 (UniProt P17814), *Os*4CL2 (UniProt Q42982), |*Os*4CL3 (UniProt Q6ETN3), *Os*4CL4 (UniProt Q67W82), *Os*4CL5 (UniProt Q6ZAC1), *Pv*4CL1 (ACD02135.1), *Pv*4CL2 (AHY94891.1), DY382977.1, DY391029.1, DY382750.1, DY390122.1 and DY391030.1. The phylogenetic tree was made in Jalview using an average distance tree from ClustalWS alignment.

Growing *C. longa*

C. longa rhizomes were potted in a shallow tray and just covered with John Innes compost, silica sand and grit. They were watered well and covered with a propagator lid to give high humidity and left to grow at 37˚C. Six weeks later, shoots from the rhizomes were visible and the propagator lid was removed. As the plants grew, they frequently needed to be split and re-potted.

Extracting and purifying RNA from *C. longa* rhizome tissue

For the extraction of RNA from *C. longa* rhizomes a method from Dong *et al*.,[178] was adapted. Solutions were autoclaved or filter sterilised. Plastic ware was autoclaved. The pestle and mortars were bake at 200˚C overnight. Per sample, 0.8 g of rhizome tissue was ground in a pre-cooled mortar with a pestle, in the presence of liquid nitrogen. The fine tissue powder was transferred with a precooled spatula into a 50 ml sterile falcon tube containing 4 ml extraction buffer (50 mM Tris-HC1 (pH 8.0), 300 mM sodium chloride, 5 mM EDTA, 2% (w/v) SDS, 0.5% (w/v) soluble polyvinylpyrrolidone (M.W. 360,000), 0.5 mM aurintricarboxylic acid and 14.3 mM β-mercaptoethanol) and then vigorously mixed. The mixture was incubated in a water bath at 65˚C for 10 minutes with occasional shaking. The mixture was then transferred into microcentrifuge tubes, and centrifuged for 15 minutes, 10 000 g, room temperature. The supernatant was moved into 2 fresh 2 ml microcentrifuge tubes and 93 µl of 3 M potassium acetate (pH 4.8) was added to each sample. After complete mixing, the tubes were placed on ice for 30 minutes, followed by centrifugation at 10 000g for 10 minutes at 4˚C. The supernatant was transferred into fresh 2 ml microcentrifuge tubes, followed by addition of 667 µl 8 M lithium chloride. The mixture was left overnight at 4˚C. The tube was then centrifuged at 12,000 g for 30 minutes at 4˚C, the supernatant was removed and the was washed twice with 1.33 ml 3 M sodium acetate (pH 5.2) by resuspending the pellet and further centrifugation at 10,000g for 10 minutes at room temperature. The final pellets were each dissolved in 667 µl RNase-free glass distilled water. Each crude RNA solution was mixed with an equal volume 667 µl of phenol (saturated with Tris-HCl, pH 8.0). The mixture was centrifuged at 10 000 g for 10 minutes at room temperature. The upper aqueous phase was transferred into a fresh 2 ml microcentrifuge tube. The aqueous phase was extracted once with an equal volume 667 µl of phenol/chloroform/isoamyl alcohol (24/23/1, v/v/v). The aqueous phase was further extracted once with an equal volume 667 µl of chloroform/isoamyl alcohol (23/1, v/v). The aqueous phase was then transferred into a fresh microcentrifuge tube, to which was added 66.7 µl 3 M sodium acetate (pH 5.2) and 1.73 ml absolute ethanol. The mixture was left overnight at -20˚C. The mixture was then centrifuged at 10 000g for 10 minutes at 4˚C after which the pellet was washed once with 70 % (v/v) ethanol (1.73 ml). After 5 minutes air drying, the RNA pellet was dissolved in 50 µl 60˚C RNase-free water and the concentration of RNA was measured using a NanoDrop spectrophotometer.

cDNA synthesis from *C. longa* RNA

To obtain 4CL cDNA, 6 different primers were used, a generic polyA annealing primer and 5 primers specific to each of the 4CL EST sequences: 10 mM Og2, Cl4CL1_R, Cl4CL2_R, Cl4CL3_R, Cl4CL4_R and Cl4CL5_R (see 4.3.10 Primer table). cDNA was synthesized using ThermoFisher Scientific SuperScript II Reverse Transcriptase. 0.5 ug of RNA was incubated with 1 µl 10 µM dNTPs and the desired primer at 65˚C for 5 minutes. 4 µl First strand buffer, 2 µl 0.1 M DTT and 1 µl RNAse Out were then added to each reaction before incubation at 42˚C for 2 minutes. 1 µl SuperScript II Reverse Transcriptase was then added and cDNA synthesis was allowed to proceed for 2 hours at 42˚C.

4.3.2 Polymerase chain reactions (PCR)

For primers used see 4.3.10 Primer table

NEB High Fidelity Phusion Polymerase PCR

For cloning or RACE PCR where high fidelity of amplified sequence is required

Component	**Volume**
Water	35.5 µl
Buffer (HF/GC)	10 µl
10 µM dNTPs	1 µl
10 mM F-primer	1 µl
10 mM R-primer	1 µl
Template	1 µl
Phusion polymerase	0.5 µl
Total	**50 µl**

Thermocycler conditions	**Temperature**	**Time**
1. Denature	98˚C	30 s
2. Denature	98˚C	10 s
3. Anneal	Primer melting temp.	30 s
4. Extension	72˚C	30 s /kb
5. Cycle		(Steps 2-4) x 25-35
6. Extension	72˚C	5 minutes

For hot start PCR, the reaction was held at the initial denaturing temperature and the polymerase was added. When DMSO was added to aid strand separation, it was carried out at 3 %.

NEB TAQ polymerase PCR

For diagnostic PCR when fidelity isn't as crucial.

Component	Volume
Water	41.75 µl
10 x Taq buffer	5 µl
10 µM dNTPs	1 µl
10 mM F-primer	1 µl
10 mM R-primer	1 µl
Template	2 µl
Taq polymerase	0.25 µl
H_2O	41.75 µl
Total	**50 µl**

Thermocycler conditions:	Temperature	Time
1. Denature	95˚C	30 s
2. Denature	95˚C	10 s
3. Anneal	Primer melting temp.	30 s
4. Extension	68˚C	1 min /kb
5. Cycle		(Steps 2-4) x 25-35
6. Extension	68˚C	5 minutes

Colony PCR

Component	Volume
Water	11.925 µl
10 x Taq buffer	1.5 µl
10 µM dNTPs	0.3 µl
10 mM F-primer	0.3 µl
10 mM R-primer	0.3 µl
Taq polymerase	0.075 µl
Total	**15 µl**

Thermocycler conditions:	Temperature	Time
1. Denature	95	30 s
2. Denature	95	10 s
3. Anneal	Primer melting temp.	30 s
4. Extension	68	1 min /kb
5. Cycle		(Steps 2-4) x 25-35
6. Extension	68	5 minutes

Separation of nucleic acids using agarose gel electrophoresis

To 1 x TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8.0) 0.8 % (w/v) agarose was added and microwaved for ~1.5 minutes, until dissolved. Once cooled to ~60˚C 1 µl ethidium bromide was added, then the gel was cast. Once set, samples (mixed with an appropriate volume of 6 x loading buffer (10 mM Tris-HCl, 50 mM EDTA, 15 % (w/v) Ficoll 400, 0.03 % (w/v) bromophenol blue, 0.03 % (w/v) xylene cyanol) and ladder were loaded and electrophoresis was performed at 125 V for 25 minutes or to desired DNA resolution. The gel was imaged using a UVP GelDoc-It Imaging System.

4.3.3 Cloning and transformation of *E.coli* and *S. cerevisiae*

***E. coli* and *S. cerevisiae* strains**

***E. coli* strain**	**Genotype**	**Use**
XL-10 Gold	*Tet^{r}Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F′proAB lacI^{q}ZΔM15 Tn10 (Tetr) Amy Cam^r]*	Cloning
Tunetta	*F$^-$ ompT hsdS$_B$(r_B^- m_B^-) gal dcm (DE3) pRARE (Cam^R)*	Protein expression
BL21	*fhuA2 [lon] ompT gal [dcm] ΔhsdS*	Protein expression
SoloPack	*Tet^{r}Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F′proAB lacI^{q}ZΔM15 Tn10 (Tetr) Amy Cam^r]*	StrataClone cloning

Yeast strain	**Genotype**
BJ2168	*MATa leu2 trp1 ura3-52 prb1-1122 pep4-3 prc1-407 gal2*
G175	*MATα ADE2 MET his3 leu2 ura3 trp1*
BY4742	*MATα his3 leu2 lys2 ura3*

***E.coli* LB broth**

Luria-Bertani (LB) broth was prepared as follows: 1 % (w/v) peptone/ tryptone, 1 % (w/v) NaCl, 0.5 % (w/v) yeast extract, pH 7 in distilled water. LB-agar was prepared as above with the addition of 1.5 % (w/v) agar. The media was autoclaved for sterility. 1 µM IPTG was added for protein induction.

***E.coli* NZ amine- yeast extract (NZY) broth**

NZY broth was prepared as follows: 0.5 % (w/v) NaCl, 0.2 % (w/v) $MgSO_4$, 0.5 % (w/v) yeast extract, 1 % (w/v) NZ amine pH 7.5 in distilled water. The media was autoclaved for sterility.

Yeast extract- peptone- dextrose (YPD) medium

YPD medium was prepared as follows: 2 % (w/v) peptone/ tryptone, 1 % (w/v) yeast extract, 2 % (w/v) glucose. YPD-agar was prepared as above with the addition of 1.5 % (w/v) agar. The media was autoclaved for sterility.

Yeast synthetic dropout medium (YSD)

YSD medium was made according to the prototrophy of the transformed yeast strain as follows, for 1 L:

Component	**Amount**
Yeast nitrogen w/o amino acids	0.67 % (w/v)
Synthetic drop out medium:	w/o Ura 0.192 % (w/v)
(Sigma Y1771 w/o Ura, Leu, Trp)	w/o Trp 0.192 % (w/v)
	w/o Leu 0.162 % (w/v)
Uracil	0.0076 % (w/v)
Tryptophan	0.0076 % (w/v)
Leucine	0.038 % (w/v)
H_2O	To 950 ml

The medium was autoclaved for sterility, then filter sterilised 40 % (w/v) raffinose was added to give a 2 % (w/v) raffinose solution (50 ml for 1 L of medium). For induction 40 % (w/v) galactose was added to give a 2 % (w/v) galactose solution.

To make 40 % (w/v) sugar solutions the desired amount of sugar was weighed to which, approximately half the desired volume of water was added. A microwave on medium heat was used to dissolve the sugar then the solution was topped up with water to the desired volume. The solution was then filter sterilised (0.42 µm) into a pre-autoclaved duran flask.

Plasmid DNA Cloning

Cloning method	Plasmids cloned
pGEM T-easy ligation	Putative *C. longa* 4CLs, pGEM vector
Restriction and ligation by Dr H. Housden	pYE2-4CL5
Gibson Assembly	pESC-DCS, pECS-DCS-CURS1, pYES2-4CL5-F2A-DCS-F2A-CURS1, pBEVY-CaUGT2
In-Fusion Restriction and ligation	pYES2-DCS-F2A-CURS1, pRS416- Ala, pRS416- Leu1, pRS416- Leu2, pRS416- 3tRNA, pFGC5941-4CL5, pFGC5941-DCS, pFGC5941-CURS1, pET-CaUGT2

Cloning reactions were transformed into XL-10 Gold competent cells (see transformation protocol). Cells were plated onto LB-agar selective medium and incubated at 37˚C overnight. Colony PCR was performed upon the colonies that grew to check for correct clones. If the PCR product was the expected size, selective LB medium was inoculated with the colony, grown overnight at 37˚C with shaking, then DNA was extracted following the Qiagen mini-prep plasmid preparation protocol. The plasmid was then sent to GATC Biotech, with the correct sequencing primer for Sanger sequencing to verify correct cloning. Further details below.

pGEM-T Easy cloning

A 1:3 vector: insert cloning ratio was used and 50 ng of pGEM-T vector. Inserts were generated by PCR. The following reaction was set up and incubated at room temperature for hour:

Component	Volume
2 x pGEM-T Buffer	5 µl
pGEM-T easy vector (50 ng)	1 µl
Insert	3 x no. of moles of vector (µl)
T4 DNA Ligase	1 µl
H_2O	2 µl – volume of insert (µl)
Total	**10 µl**

Restriction and ligation

NEB and Promega restriction enzymes were used. Restriction reactions were set up to generate the backbone and insert. Restriction enzymes were chosen based upon presence in the multiple cloning site of the vector backbone, absence in the cloning sequences and buffer compatibility. The restriction reaction was set up as below and incubated at the enzymes specific incubation temperature for 1-4 hours.

Component	**Volume**
Water	16.3 µl
10 x restriction enzyme buffer	2.0 µl
DNA (~ 1 µg/µl)	1.0 µl
Restriction enzyme 1	0.5 µl
Restriction enzyme 2	0.5 µl
Total	**20.0 µl**

The reaction was then subjected to agarose gel electrophoresis to determine if restriction was successful and allow purification of the correct DNA fragment(s). Purification of DNA from agarose gels was carried out using Promega's Wizard SV purification kit. Once purified, if necessary, and quantified, ligation was carried out, using ~ 100 ng of vector, and the chosen amount of insert depending on the ratio chosen (usually 1:3 vector: insert) as follows:

DNA Ligation	**Volume**
T4 ligation buffer	1 µl
Vector ~100 ng	x µl
Insert (correct molar ratio to vector)	y µl
H_2O	8 – (+y) µl
T4 DNA Ligase	1 µl
Total	**10 µl**

The ligation reaction was left to proceed for 3 hours at room temperature or 4˚C overnight. 2 µl of the reactions were then transformed into XL-10 Gold competent *E. coli* cells.

Gibson Assembly

DNA inserts for Gibson Assembly were amplified by PCR using specially designed primers to create homologous regions between the ends of the inserts and the backbone. The insert was DpnI treated, then purified using Promega's Wizard SV purification kit (see below). Vector backbones were linearised by restriction, as above for restriction and ligation cloning. The total amount of DNA insert and vector backbone fragments was kept between between 0.02- 0.5 pmols.

Gibson Assembly reaction	**Volume**
50-100 ng vector	x µl
100- 300 ng (2-3 x more, use 5 x more if < 200 bp)	y µl
2 x Gibson Assembly master mix	10 µl
Deionised water	10 – (x+y) µl
Total	**20 µl**

2 µl of the reactions were then transformed into XL-10 Gold competent *E. coli* cells.

In-Fusion cloning

DNA inserts for In-Fusion cloning were amplified by PCR using specially designed primers to create homologous regions between the ends of the inserts and the backbone. The insert was DpnI treated, then purified using Promega's Wizard SV purification kit (see below). Vector backbones were linearized by restriction, as above for restriction and ligation cloning, or created by PCR.

In-Fusion cloning reaction	**Volume**
5 x In-Fusion HD enzyme premix	2 µl
Vector	x µl
Insert	y µl
Deionised water	8 – (x+y) µl
Total	**10 µl**

2 µl of the reactions were then transformed into competent XL-10 Gold *E. coli* cells.

DpnI DNA treatment

To remove template DNA from PCR reactions to remove background colonies after cloning and transformation, DpnI digest was carried out:

DpnI restriction digest	Volume
DpnI	1 µl
DNA (~ 1 µg)	10 µl
10 x CutSmart buffer	5 µl
H_2O	34 µl
Total	**50 µl**

Promega Wizard SV Gel and PCR Clean-Up System

To purify DNA from PCR reactions or excised agarose gel DNA bands the Promega Wizard SV Gel and PCR Clean-Up kit was used. Following electrophoresis, excised DNA band was placed in a1.5ml microcentrifuge tube. 10 µl membrane binding solution, per 10 mg of gel slice, was added and incubated at 50– 65˚C until the gel is dissolved. To PCR amplification reactions, an equal volume of membrane binding solution was added to the PCR amplification. The dissolved gel mixture or prepared PCR product was transferred to the minicolumn and incubated at room temperature for 1 minute before being centrifuged at 16 000 g, 1 minute. The minicoloumn was then washed with 700 µl membrane wash solution and centrifuged at 16 000 g for 1 minute. This wash was repeated with 500 µl membrane wash solution, then centrifuged at 16 000 g, 5 minutes. The minicolumn was transferred to a clean 1.5 ml microcentrifuge tube and pure DNA was eluted with 50 µl of nuclease-free water. The concentration of DNA was measured using a NanoDrop 1000 spectrophotometer.

***E.coli* transformation**

All steps were carried out under sterile conditions and on ice unless stated. Resultant colonies were screened for correctly cloned plasmids using colony PCR. Different cells were used for different transformations, see *E. coli* and *S. cerevisiae* strains.

XL-10 Gold ultracompetent cells

To a thawed 50 µl aliquot of XL-10 Gold cells in a 1.5 ml microcentrifuge tube, 2 µl of β-mercaptoethanol was added and incubated on ice for 10 minutes, with swirling every 2 minutes. 2 µl of cloning reaction was added, then incubated for 5 minutes on ice. The cells were then heat shocked for 1 minute at 42˚C followed by 5 minutes on ice. 500 µl NZY medium was added before incubation at 37˚C with shaking at 225-250 rpm for 1 hour. The cell culture was centrifuged for 1 minute at 5000 rpm and the supernatant was removed leaving ~100 µl. The pellet

was resuspended by flicking then spread on selective LB-agar plates; colonies were allowed to grow overnight at 37˚C.

Agilent SoloPack Gold competent cells

SoloPack Gold cells were thawed on ice, to which 0.1- 50 ng of cloning reaction DNA was added, swirled, then incubated on ice for 30 minutes. The tubes were then heat shocked in a 42˚C water bath for 60 seconds then incubated on ice for 2 minutes. 175 µl of preheated (42˚C) NZY broth was added then the tubes were incubated at 37˚C for 1 hour with shaking at 225-250 rpm. 200 µl of the transformation mixture was plated out on LB agar plates containing the appropriate selective antibiotic. Colonies were allowed to grow overnight at 37˚C.

BL 21 cells

BL21 cells were thawed on ice before 5– 10 ng of cloning reaction DNA, in a volume of 1– 5 µl, was added to the cells and mixed by tapping gently. The microcentrifuge(s) were incubated on ice for 30 minutes. The cells were heat shocked by incubating for 30 seconds at 42˚C from which they were quickly placed on ice. 250 µL of pre-warmed NZY medium was added before being placed in a shaking incubator at 37˚C for 1 hour at 225 rpm. The transformation mixture was plated onto LB agar plates containing the appropriate antibiotic and colonies were allowed to grow overnight at 37˚C.

Qiagen Mini-prep DNA extraction from *E. coli*

1-5 ml bacterial overnight culture was pelleted by centrifugation at 6800 x g for 3 minutes at room temperature. The pellet was resuspended in 250 µl Buffer P1 and transferred to a microcentrifuge tube. 250 µl Buffer P2 was added and mixed by inversion until clear. 350 µl Buffer N3 was added and mixed immediately centrifuged for 10 minutes at maximum speed. The supernatant was applied to the QIAprep spin column and centrifuged for 30– 60 s, then washed by adding 0.75 ml buffer PE and centrifuging for 30– 60 s, then a further minute to remove residual wash buffer. To elute DNA, 50 µl water was applied to the centre of the QIAprep spin column, let stand for 1 minute, then centrifuged for 1 minute. The concentration of DNA was measured using a NanoDrop 1000 spectrophotometer.

DNA Sanger sequencing

For DNA Sanger sequencing, 5 µl of the purified plasmid DNA (80- 100 ng/µl) to be sequenced, combined with 5 µl of 5 µM sequencing primer was sent to GATC Biotech. Resultant sequence data was analysed using Jalview.

4.3.4 Yeast manipulation

Yeast chemical transformation

The yeast strain to be transformed was patched (2 cm^2) onto the correct selective agar and incubated overnight at 30˚C. The transformation reagents were autoclaved. An aliquot of carrier single-stranded DNA (deoxyribonucleic acid sodium salt from salmon testes) in TE buffer (10 mM Tris-HCl pH 8.0, 1.0 mM EDTA)) was heated in a boiling water bath for 5 minutes then chilled. A small portion of yeast was scraped from the agar plate and suspended in 1.0 ml of sterile water in a 1.5 ml microcentrifuge tube. The cells were pelleted at top speed in a microcentrifuge for 30 seconds and the supernatant was discarded. The following components were added to the pellet, then incubated in a water bath at 42˚C for 45 minutes.

Component	Volume (µl)
PEG 3500 50% (w/v)	240 µl
LiAc 1.0 M (pH 8.4 - 8.9)	36 µl
Boiled SS-Carrier DNA (2 mg/ml)	50 µl
Plasmid DNA (0.1 to 1 µg) in water	34 µl
Total	**360 µl**

The transformation mix was centrifuged at top speed for 30 seconds and the supernatant removed. The cell pellet was washed with 1.0 ml of sterile water before being resuspended in 1.0 ml of sterile water. 10 and 100 µl samples were plated onto YSD-agar plates of correct selection and incubated at 30˚C for 3-4 days until colonies grew.

DNA extraction from *S. cerevisiae*

DNA was extracted from yeast cells to check transformation using a modification of the Qiagen MiniPrep kit. 10 ml yeast overnight culture was pelleted by centrifugation at 3000 g for 5 minutes. The pellet was resuspended in 800 µl Qiagen Miniprep buffer P1 and transferred to a screw top microcentrifuge tube. The tube was filled with 100 µm acid washed glass beads and the yeast cell suspension was subjected to tissue lysis using the Qiagen TissueLyserII for 5 minutes. The disrupted yeast cell suspension was transferred to a 1.5 ml microcentrifuge tube and centrifuged at maximum speed for 3 minutes. The supernatant was applied to the QIAprep spin column and centrifuged for 30–60 s, then washed by adding 0.75 ml Buffer PE and centrifuged for 30–60 s, then a further minute to remove residual wash buffer. To elute DNA, 50 µl water was applied to the centre of the QIAprep spin column, let stand for 1 minute, and centrifuged for 1 minute.

RNA extraction from *S. cerevisiae*

RNA was extracted from yeast using the Qiagen RNeasy Mini kit. Per sample, a 10 ml yeast overnight culture was pelleted by centrifugation at 1000 x g for 5 minutes at 4˚C. The pellet was resuspended in 600 µl Buffer RLT and transferred to a TissueLyser tube containing 600 µl acid washed glass beads. The sample was agitated at top speed in the Qiagen TissueLyserII homogeniser. The lysate, after the beads settled, was transferred to a new microcentrifuge tube and centrifuged for 2 minutes at full speed. The supernatant was transferred to a new microcentrifuge tube and 1 volume of 70 % ethanol was added to the homogenized lysate, and mixed well by pipetting. The sample was transferred to an RNeasy spin column placed in a 2 ml collection tube and centrifuged for 15 seconds at 8000 x g. The flow through was discarded and 700 µl Buffer RW1 was added to the RNeasy spin column and centrifuged for 15 seconds at 8000 x g. The flow through was discarded and 500 µl Buffer RPE was added to the RNeasy spin column and centrifuged for 15 seconds at 8000 x g. The flow through was discarded and the RNeasy spin column was placed in a new 1.5 ml collection tube and 30–50 µl RNase-free water was applied directly to the spin column membrane and centrifuged for 1 minute at 8000 x g to elute the RNA. RNA was quantified using a NanoDrop 1000 spectrophotometer ready for DNase treatment and cDNA synthesis.

To remove any uunwanted DNA from the extracted RNA, Promega DNase was used. To 8 µl RNA in water, 1 µl RQ1 RNase-Free DNase 10 x Reaction Buffer and 1 µl RQ1 RNase-Free DNase was added and incubated at 37˚C for 30 minutes. Then 1 µl of RQ1 DNase Stop Solution was added and incubated at 65˚C for 10 minutes to inactivate the DNase.

For cDNA synthesis, Invitrogen SuperScript II reverse transcriptase (RT) was used. To the DNase treated RNA, 1 µl Oligo$(dT)_{12\text{-}18}$ (500 µg/mL) and 1 µl dNTP mix (10 mM each) was added then incubated at 65˚C for 5 minutes then chilled on ice. Then 4 µl 5 x First-Strand Buffer, 2 µl 0.1 M DTT and 1 µl RNaseOUT incubated at 42˚C for 2 minutes. Then 1 µl SuperScript™ II RT was added and incubated at 42˚C for 50 minutes, then inactivated by heating at 70˚C for 15 minutes.

4.3.5 Protein extraction and analysis

Ammonium sulphate precipitation of crude protein from plant tissue

All steps were performed on ice where possible. ~4 g of plant tissue was ground under liquid nitrogen using a pestle and mortar to a fine powder, then transferred to a beaker chilled on ice. Immediately, 3 x (w/v) extraction buffer (50 mM Tris-HCl, 2 mM EDTA, 1 mM DTT, pH 7.5) and 50 g L^{-1} polyvinyl polypyrrolidont (PVPP) were added and mixed thoroughly. The slurry was filtered through a double layer of miracloth into chilled beaker. The filtrate was transferred to a centrifuge tube and centrifuged at 10 000g for 15 minutes at 4˚C. The proteinaceous supernatant

was retained. The volume of supernatant was measured and ammonium sulphate, to give 80 % saturation, was added and stirred at 4˚C for 30 minutes. The mixture was then centrifuged at 4000 x g for 20 minutes at 4˚C. The supernatant was discarded and the pellet was resuspended in working buffer (20 mM Tris-HCl, pH 8.0, 2 mM DTT). A 2 ml Zeba spin desalting column was filled with a water slurry of Sephadex-G25, placed in an empty 15 ml falcon tube then centrifuged at 2000 RPM for 2 minutes at 4˚C. The column was washed three times with working buffer before the protein was applied then centrifuged. The flow through was collected and the protein concentration was measured using the Thermo Scientific BCA Protein Assay Kit or the Bio-Rad Bradford Protein Assay according to the manufacturer's instructions.

Protein extraction from *E. coli*

Selective LB medium was inoculated with *E. coli*. The culture was grown overnight at 37˚C with shaking at 225- 250 RPM; protein expression was induced by addition of IPTG when the culture reached an OD of 0.4, to a working condition of 1 µM. After overnight growth, 5 ml of the culture was centrifuged at 3000 RPM for 5 minutes at 4˚C. The pellet was resuspended in 1 ml extraction buffer (20 mM Tris-HCl pH 8.0, 2 mM DTT) and sonicated (3 x 15 seconds with 5 second breaks). The suspension was centrifuged at maximum speed for 3 minutes and the protein in the supernatant was then quantified using the Thermo Scientific BCA Protein Assay Kit or the Bio-Rad Bradford Protein Assay according to the manufacturer's instructions.

Protein extraction from *S. cerevisiae*

YSD medium was inoculated with the desired yeast strain and grown overnight at 37˚C with shaking at 200 RPM. Once the culture reached OD 0.4, 40 % galactose (w/v) was added to the medium to give 2 % (w/v). 4 hours after induction, the culture was pelleted by centrifugation at 3000 x g for 5 minutes at 4˚C. The pellet was weighed, for every 100 mg, 250 µl Sigma Y-PER Reagent was added with Sigma protease inhibitor cocktail P8215 (1 ml per 20 g wet yeast cells plus 1 M DTT to give 7.5 mM). The suspension was agitated at 200 RPM for 30 minutes, then centrifuged at 14 000 × g for 10 minutes. The supernatant was reserved for analysis. Protein concentration was quantified using the BCA Protein Assay Kit (Thermo Scientific) or the Bradford Protein Assay (Bio-Rad) according to the manufacturer's instructions.

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), Coomassie staining and western blot analysis.

SDS-PAGE analysis was performed using a discontinuous gel system, with Bio-Rad Mini-Protean Tetra cell apparatus. Bio-Rad Mini-Protean TGX stain-free 15 µl well precast gels were used to run samples. The tank apparatus was assembled and filled with Laemmli running buffer (25 mM Tris, 192 mM glycine, 0.1 % (w/v) SDS, pH 8.3). Protein samples (~ 50 µg) for analysis were mixed (4:1) with 4 x SDS loading buffer (83.3 mM Tris-HCl, 16.7 % (v/v) glycerol, 2.67 %

(w/v) SDS, 6.67 % (v/v) β-mercaptoethanol, 0.003 % (w/v) bromophenol blue, pH 6.8) and boiled for 5 minutes, before being loaded into the wells. 10 µl Bio-Rad pre-stained broad-range protein marker was also loaded to allow the estimation of the molecular weight of proteins. Electrophoresis was carried out at 200 V for 30 minutes. To visualise protein on the SDS gel, it was stained with Coomassie Brilliant Blue stain (0.1% Coomassie Brilliant Blue 50% methanol, 10% acetic acid). To de-stain the gel, an acetic acid solution was used: 50 % water, 40 % methanol, and 10 % acetic acid.

To allow western blot analysis, the protein was transferred from the SDS gel to a nitrocellulose membrane. To do this, nitrocellulose membrane was incubated in pre-chilled Towbin transfer buffer (25 mM tris-HCl, 192 mM glycine, 20 % (v/v) methanol) for 25 minutes at 4 °C. The SDS gel was also equilibrated in pre-chilled Towbin transfer buffer, for 15 minutes at 4 °C. The gel, membrane and protecting filter paper were arranged according to the Trans-Blot SD Semi-Dry Transfer Cell protocol and placed in the transfer cell which was operated at 10 V for 1 hour. Protein transfer was confirmed by staining in Ponceau solution (0.2 % Ponceau (w/v), 3 % trichloroacetic acid (v/v). To remove the stain, the membrane was washed in TBS (25 mM tris-HCl, 150 mM NaCl, pH 7.5).

For western blot analysis, the membrane was first blocked for 1 hour with TBS buffer plus 3 % (w/v) skimmed milk, then incubated with the primary antibody for 1 hour, according to the table below. The membrane was then washed 3 times for 5 minutes in TBST buffer (TBS 25 mM tris-HCl, 150 mM NaCl, 0.05 % Tween-20 (v/v), pH 7.5). The membrane was then incubated with the secondary antibody for at least 1 hour according to the table below. Again, the membrane was washed 3 times for 5 minutes in TBST buffer before detection according to the table below was carried out. If the secondary antibody was an alkaline phosphatase (AP) conjugate, nitro-blue tetrazolium (NBT) and 5-bromo-4-chloro-3'-indolyphosphate (BCIP) were used for chromogenic blot analysis: the membrane was incubated for 2 minutes in 100 mM Tris-HCl, pH 9.5, then NBT/ BCIP solution (33 ul each per 10 ml in 100 mM Tris-HCl, pH 9.5) until staining was observed. If the secondary antibody was a horse radish peroxidase (HRP) conjugate: a 4-Chloro-1-naphthol tablet (4CN) was dissolved in 10 ml methanol, 2 ml of this was added to 10 ml of triethanolamine buffer saline and 5 ul of 30% H_2O_2, then used to develop stain.

Epitope	**Primary antibody**	**Secondary antibody**	**Detection**
Strep-tag	1:5000 *Strep*-Tactin HRP conjugate	-	4CN
Anti-GFP	1:5000 Anti-GFP mouse	1:1500 Anti-mouse-AP	NBT+BCIP
Anti-HIS tag	1:3000 Anti-HIS HRP conjugate	-	4CN
Anti-myc tag	1:10000 Anti-c-MYC mouse	1:1500 Anti-mouse-AP	NBT+BCIP
Anti-FLAG	1:1500 Anti-FLAG rabbit	1:1500 Anti-rabbit-HRP	4CN

4.3.6 Activity assays

***Trichoderma reesei* cellulase assay**

200 µl of *C. longa* rhizome methanolic extract was dried using a Genevac centrifugal evaporator. The residue was resuspended in 0.5 ml 0.15 M citric phosphate buffer (pH 5.0) to which 1 mg ml^{-1} cellulase (*Trichoderma reesei* ATCC 26921) was added and incubated at 30˚C overnight. The sample was extracted twice with 0.5 ml ethyl acetate which was dried using a Genevac centrifugal evaporator and resuspended in methanol for LC-MS analysis (see Curcumin glucoside LC-MS analysis method).

^{14}C-UDP-glucose radiolabelling assay

Protein was extracted from the organism to be assayed for UGT activity in biological triplicate. An aliquot of ^{14}C-UDP glucose was measured using scintillation counting to determine the volume needed to give a reading of 50 000 dpm. The assay was set up as follows, with ^{14}C-UDP glucose being added last:

Component	Volume
^{14}C-UDP Glucose (50, 000 dpm)	x
1 mM substrate in Methanol	5
Crude protein (50-100 ug)	y
BSA (20 mg/ml)	2.5
20 mM Tris-HCl (pH 8.0, 2 mM DTT)	75 - x - y
Total	**75 ul**

The reaction was incubated at 30˚C then quenched with 125 µl 0.3 M HCl, extracted with 200 µl water saturated ethyl acetate and centrifuged at maximum speed for 1 minute. 100 µl supernatant was added to 4 ml Ecoscint solution in scintillator tubes and scintillation counting was carried out using a Packard Trib-Carb 2900TR Liquid Scintillation Analyzer.

***In vitro* crude protein assays**

Protein extracted from yeast transformed with the desired vector(s) was quantified using the Thermo Scientific BCA Protein Assay Kit or the Bio-Rad Bradford Protein Assay according to the manufacturer's instructions.

4-coumarate CoA ligase 5 (4CL5) activity

Crude protein extract containing 4CL5 (100 µg), was made up to 50 µl in 100 mM Tris-HCl buffer (pH 8.0) containing 10 mM DTT. To this was added 547 µl of 100 mM Tris-HCl buffer pH 8.0 containing 1.6 mM coenzyme A, 8.3 mM ATP and 8.3 mM $MgCl_2$. 3 µl 100 mM phenylpropanoic

acid substrate (in methanol) was added to give a concentration of 0.5 mM. The reaction was then incubated at 30˚C. The reaction was stopped by adding an equal volume of methanol and the samples were centrifuged (16 000 g, 5 minutes), to pellet precipitated protein. The supernatant was analysed by LC-MS (see Phenylpropanoyl-CoA ester LC-MS analysis method 1).

Diketide-CoA synthase (DCS) activity

Crude protein extract containing DCS, were assayed as carried out by Katsuyama, *et al*.[93]. To 50 µg crude protein (in 40 µl Y-PER lysis buffer), 59 µl reaction buffer (100 mM phosphate buffer (pH 7.5), 50 mM malonyl-CoA) was added. Then 1 µl 10 mM feruloyl CoA was added and the reaction was incubated at 37 °C for 1 hour before quenching with 20 µl of 6 M HCl. The samples were centrifuged (16 000 g, 5 minutes), to pellet precipitated protein. The supernatant was analysed by LC-MS (see Phenylpropanoyl-CoA ester LC-MS analysis method 1).

Curcuminoid synthase 1 (CURS1) activity

Crude protein extract containing CURS1 (50 µg), was made up to 50 µl in 100 mM Tris-HCl buffer (pH 8.0) containing 10 mM DTT. To this was added 547 µl of 100 mM Tris-HCl buffer pH 8.0 containing 1.6 mM coenzyme A, 8.3 mM ATP, 8.3 mM $MgCl_2$ and 0.5 mM malonyl-CoA. 3 µl 100 mM phenylpropanoic acid substrate (in methanol) was added to give a concentration of 0.5 mM. The reaction was then incubated at 30˚C. The reaction was stopped by adding an equal volume of methanol and the samples were centrifuged (16 000 g, 5 minutes), to pellet precipitated protein. The supernatant was analysed by LC-MS (see Phenylpropanoyl-CoA ester LC-MS analysis method 2).

***Catharanthus roseus* UDP-glycosyltransferase 2**

Crude protein extract containing CaUGT2 (50 µg), was made up to 100 µl in 50 mMTris–HCl buffer (pH 7.5), 2 mM UDP-glucose and 250 µM curcuminoid substrate. The reaction was incubated at 30˚C for 30 minutes. The reaction was terminated by adding 200 µl methanol. After centrifugation at 16 000 x g for 10 minutes, the reaction products were analysed by LC-MS using a Phenomenex reversed phase Luna 5 µm C18(2) 100 Å, LC Column 250 x 4.6 mm (see curcumin glycoside method).

***In vivo* yeast cell culture assay**

The same assay was performed for investigating the *in vivo* activity of 4CL5, DCS and CaUGT2 expressing yeast cells. 25 ml 2 % (w/v) raffinose YSD medium, made according to the prototrophy of the transformed yeast strain to be investigated in a shake flask, was inoculated, and the yeast was grown over night at 30˚C with shaking at 200 RPM. The OD_{600} was measured and adjusted to 0.4 and 40 % galactose was added to give 2 % (w/v) and induce protein production. After 4 hours induction, 25 µl of 0.1 M substrate was added and the culture was grown at 30˚C with shaking at 200 RPM. The cultures were samples for analysis over 72-96 hours. To analyse,

1 ml of the culture was removed and centrifuged at 16 000 g for 3 minutes. 0.5 ml of supernatent was combined with 0.5 ml ethanol for analysis of the culture medium. The remaining supernatant was discarded and the pellet was washed with 1 ml of water, then repelleted. To the pellet, 200 µl methanol and 50 µl acid washed beads was added and lysed by vortexing for 30 seconds then centrifuged for 16 000 g for 3 minutes. The cell contents and media were separately analysed using the appropriate LC-MS method.

Spectroscopic determination of curcumin glycoside production

25 ml 2 % (w/v) raffinose YSD medium, made according to the prototrophy of the yeast transformed with pBEVY-CaUGT2, was inoculated, and the yeast was grown over night at 30˚C with shaking, 200 RPM. The OD_{600} was measured and altered to 0.4 and the medium was adjusted to the conditions to be investigated (such as supplemental raffinose, or adding $FeSO_4$). 125 µl 0.1 mM curcumin (in methanol) was added. As the cultures grew, at 30˚C with shaking, 200 RPM, they were sampled and spectrophotometrically analysed over 72 hours. To sample, 1 ml was removed from the shake flask and centrifuged at 16 000 g for 3 minutes. The absorbance of the supernatant was then measured at 425 nm and recorded.

5 L Fermentation

25 ml 2 % (w/v) raffinose YSD medium, made according to the prototrophy of the yeast transformed with pBEVY-CaUGT2, was inoculated, and the yeast was grown over night at 30˚C with shaking, 200 RPM. 475 ml 2 % (w/v) raffinose YSD medium was inoculated with the 25 ml culture and grown over night at 30˚C with shaking, 200 RPM. 4.5 L 2 % (w/v) raffinose, 2 % (w/v) galactose YSD medium, with 1 ml Sigma antifoam C, was added to the 5 L Applikon bioreactor and autoclaved. The media was then calibrated and set up for fermentation: 30˚C, 250 RPM stirring, ≥40 % dissolved oxygen. These parameters were monitored by the Applikon *ez*control unit. The bioreactor was inoculated with the 500 ml culture and left to proceed under the control of the Applikon *ez*control unit. After 4 hours, 25 ml 20 mM curcumin (in methanol) was added. Culture density (OD_{600}) and glycoside production (by spectrophotometric analysis at 425 nm) were manually monitored.

4.3.7 Heterologous expression in tobacco

Agrobacterium transformation

Agrobacterium tumefaciens, strain GV3101, was used for the Agrobacterium mediated transformation of tobacco leaf tissue. To a defrosted 100 µl aliquot of cells on ice, 2-3 µl (~ 300 ng) of plasmid was added, then immediately frozen in liquid nitrogen. The cells were then defrosted for 5 minutes at 37˚C then incubated on ice for 30 minutes. 250 µl LB was added and the cells were recovered at 28˚C, using shaking at 200 RPM, for 1-3 hours. The cells were then

spread on LB- agar plates containing gentamicin and kanamycin to select for the agrobacterium transformed with the pFGC5941 plasmid. The plates were incubated at 28˚C until colonies formed.

Agro-infiltration and tobacco leaf disc assay

Transformed *A. tumefaciens* was grown for 2 days at 28˚C with shaking at 180 RPM, after which the cells were resuspended in 10 mM $MgCl_2$ to give an OD_{600} of 1.0. Solutions of *A. tumefaciens* carrying each plasmid (pFGC-4CL5/DCS/CURS1) were combined as required. Using a 1 ml syringe, 4 week old *Nicotiana benthamiana* leaves were infiltrated with the $MgCl_2$ cell solutions. After 5 days, 10 leaf discs (~ 200 mg tissue), representing the different enzyme combinations, were harvested and incubated in 1 ml of each of the solutions to be tested and incubated at 18˚C, 16 hours of light per day, 40 % rH. Samples were collected after 0, 24, and 48 hours. The plant tissue and solutions were added to 10 x ice cold acetone (w/v) and ground in a mortar and pestle using glass beads. The slurry was filtered through two layers of filter paper. Any remaining solids were further ground in 10 x ice cold 1:1 acetone:methanol (w/v). Again the slurry was filtered through two layers of filter paper and the filtrates were combined and evaporated to dryness using a Genevac centrifugal evaporator. The residue was resuspended in 200 µl of methanol and analysed by LC-MS (see Tobacco analysis LC-MS method).

qPCR

RNA was extracted from 100 mg transformed tobacco tissue using the Qiagen RNeasy Mini kit, the protocol of which was followed exactly. To remove any DNA from the extracted RNA, Promega DNase was used: to 8 µl RNA in water, 1 µl RQ1 RNase-Free DNase 10 x Reaction Buffer and 1 µl RQ1 RNase-Free DNase were added and incubated at 37˚C for 30 minutes. Then 1 µl of RQ1 DNase Stop Solution was added to terminate the reaction and incubated at 65˚C for 10 minutes to inactivate the DNase.

For cDNA synthesis Invitrogen SuperScript II reverse transcriptase (RT) was used: to the DNase treated RNA, 1 µl $Oligo(dT)_{12\text{-}18}$ (500 µg/mL) and 1 µl dNTP mix (10 mM of each dNTP) were added then incubated at 65˚C for 5 minutes then chilled on ice. Then 4 µl 5 x First-Strand Buffer, 2 µl 0.1 M DTT and 1 µl RNaseOUT were added and incubated at 42˚C for 2 minutes. Next, 1 µl SuperScript II RT was added and incubated at 42˚C for 50 minutes, then inactivated by heating at 70˚C for 15 minutes.

For qPCR, the cDNA was diluted ten-fold with RNase free water and the reaction was set up as follows:

Component	Volume
SYBR green	10 ul
Primer 1	1.6 ul
Primer 2	1.6 ul
RNase free water	4.8 ul
cDNA	2 ul
Total	**20 ul**

qPCR for each sample was performed in triplicate in a 96 well plate according to the Bio-Rad MyIQ protocol. Data analysis was carried out using the Bio-Rad analysis module.

4.3.8 Extraction of metabolites from *C. longa* rhizome tissue

Extraction of metabolites from Vietnamese *C. longa* rhizome samples

As performed by Kieu Oahn Nguyen (University of Science and Technology, Hanoi, Vietnam). The gathered samples were cleaned in water then freeze dried using an Edwards Modulyo Freeze Dryer. The sample was ground to a fine powder using a pestle and mortar. Extraction and detection had been optimised after investigating different solvents (100 % methanol, 50 %, acetone, 50 %, hexane), sample weight (20, 50, 100, 150 and 200 mg) and dilution (10, 20, 50 and 100 fold). 50 mg powdered sample was mixed with 1 ml methanol and sonicated at 70 ˚C for 1 hour before being cooled at 4˚C for 30 minutes and centrifuged at 10 000 g, 15 minutes. The supernatant was analysed by UPLC-MS (UPLC-LTQ Orbitrap MS analysis) and the data was analysed by Dr Tony Larson.

Soxhlet extraction of metabolites from fresh *C. longa* rhizome tissue

Fresh *C. longa* rhizome tissue (3 g) was ground to a powder using liquid nitrogen and a pestle and mortar. The powder was transferred to an extraction thimble and placed within the Soxhlet apparatus. 125 ml of methanol was heated allowing it to pass through the apparatus for 10 cycles, extracting metabolites from the ground rhizome tissue. The methanol extract was dried using a rotary evaporator and the residue was dissolved in water by sonication for 15 minutes. The water solution was subjected to liquid-liquid extraction with ethyl acetate three times. The ethyl acetate layers were combined and dried using a rotary evaporator. All layers were analysed by LC-MS (see Curcumin glycoside LC-MS analysis method).

4.3.9 Liquid Chromatography-Mass Spectrometry (LC-MS) Analysis

Samples for LC-MS were prepared by centrifugation at 16 000 g for 10 minutes, before decanting 100 µl into a vial for analysis. Data was analysed using Thermo Scientific Xcalibur software.

Phenylpropanoyl-CoA ester LC-MS analysis method 1

Using a quadripartite solvent system and positive electrospray ionisation (ESI), 10 µl of sample was injected for analysis using a Phenomenex reversed phase Kinetex 5µm C18 100 Å, 50 x 2.1 mm column at 30˚C. The method was adapted from Larson and Graham, 2001 [244]. The solvent gradient was as follows:

Time (minutes)	Flow (ml/min)	0.05 % TEA	90 % acetonitrile, 0.05 % TEA	90 % acetonitrile, 1 % acetic acid	1 % acetic acid
0.0	0.75	0.0	0.0	10.0	90.0
5.0	0.75	0.0	0.0	80.0	20.0
5.1	0.75	80.0	0.0	0.0	0.0
7.0	0.75	97.0	3.0	0.0	0.0
10.0	0.50	95.0	5.0	0.0	0.0
10.1	0.50	95.0	5.0	0.0	0.0
30.0	0.50	60.0	40.0	0.0	0.0
30.1	0.50	0.0	100.0	0.0	0.0
32.0	0.50	0.0	100.0	0.0	0.0
32.1	0.50	0.0	100.0	0.0	0.0
35.0	0.50	0.0	100.0	0.0	0.0
35.1	0.50	0.0	0.0	10.0	0.0

Phenylpropanoyl-CoA ester LC-MS analysis method 2

Using a quadripartite solvent system and positive ESI, 10 µl of sample was injected for analysis using a Phenomenex reversed phase Kinetex 5µm C18 100 Å, 50 x 2.1 mm column at 30˚C. The method was adapted from Larson and Graham, 2001 [244]. The solvent gradient was as follows:

Time (minutes)	Flow (ml/min)	0.25 % TEA	90 % acetonitrile, 0.05 % TEA	90 % acetonitrile, 1 % acetic acid	1 % acetic acid
0.0	0.75	0.0	0.0	10.0	90.0
5.0	0.75	0.0	0.0	80.0	20.0
5.1	0.75	80.0	0.0	0.0	0.0
7.0	0.75	97.0	3.0	0.0	0.0
10.0	0.50	95.0	5.0	0.0	0.0
10.1	0.50	95.0	5.0	0.0	0.0
50.0	0.50	60.0	40.0	0.0	0.0
50.1	0.50	0.0	100.0	0.0	0.0
52.0	0.50	0.0	100.0	0.0	0.0
52.1	0.50	0.0	100.0	0.0	0.0
55.0	0.50	0.0	100.0	0.0	0.0
55.1	0.50	0.0	0.0	10.0	0.0

Tobacco analysis LC-MS method

Using a two solvent system and negative ESI, 10 µl of sample was injected for analysis using a Phenomenex reversed phase Synergi Polar-RP 80 Å 250 x4.6 mm column at 30˚C. The solvent gradient was as follows:

Time (minutes)	Flow (ml/min)	95 % water, 5 % acetonitrile	100 % acetonitrile
0.0	1.00	5.0	95.0
2.0	1.00	5.0	95.0
15.0	1.00	50.0	50.0
25.0	1.00	55.0	45.0
32.0	1.00	100.0	0.00
34.0	1.00	100.0	0.0
34.1	1.00	5.0	95.0
37.0	1.00	5.0	95.0

UPLC-LTQ Orbitrap LC-MS analysis method

Using a two phase solvent system and positive APCI, 10 µl of sample was injected for analysis using a Waters Acquity UPLC BEH Phenyl 1.7 µm 2.1x100 mm column at 60˚C. The solvent gradient was as follows:

Time (minutes)	Flow (ml/min)	95 % water, 5 % methanol 0.1 % Formic acid	100 % methanol 0.1 % Formic acid
0.0	0.50	45.0	55.0
2.50	0.50	45.0	55.0
4.00	0.50	0.0	100.0
4.40	0.50	0.0	100.0
4.41	0.50	45.0	55.0
5.00	0.50	45.0	55.0
15.0	0.50	45.0	55.0

Curcumin glycoside LC-MS analysis method

Using a two phase solvent system and negative ESI, 10 µl of sample was injected for analysis using a Phenomenex reversed phase Luna 5 µm C18(2) 100 Å, 250 x 4.6 mm column at 30˚C. The solvent gradient was as follows:

Time (minutes)	Flow (ml/min)	95 % water, 5 % methanol	100 % methanol
0.0	1.00	60.0	40.0
14.0	1.00	21.0	79.0
15.0	1.00	0.0	100.0
20.0	1.00	0.0	100.0
23.0	1.00	100.0	0.0
25.0	1.00	100.0	0.0
27.0	1.00	60.0	40.0
30.0	1.00	60.0	40.0
30.0	1.00	60.0	40.0

Flash chromatography

Flash chromatography was carried out using a Biotage Isolera Prime system and a Reveleris C18 Reversed-Phase 40 g cartridge. The sample, for purification, was dissolved in a minimum volume of 10 % methanol. After priming the system according to the Biotage manual guidelines, the sample was manually loaded onto the cartridge. Using a stepped methanol gradient, as below, 25

ml fractions were collected. The fractions were analysed by LC-MS (Curcumin glycoside LC-MS analysis method) to verify their contents, then pooled accordingly and concentrated using a Genevac centrifugal evaporator.

Column volumes	Flow (ml/min)	100 % water	100 % methanol
2.0	50.0	90.0	10.0
1.0	50.0	90.0	10.0
1.0	50.0	90.0- 80.0	10.0- 20.0
1.0	50.0	80.0	20.0
2.0	50.0	80.0- 60.0	20.0- 40.0
1.0	50.0	60.0	40.0
3.0	50.0	60.0- 0.0	40.0- 100.0
4.0	50.0	0.0	100.0

Preparative HPLC

Preparative HPLC was carried out using an Agilent Varian Modular Analytical HPLC System and a Waters SunFire C18 OBD Prep Column, 100Å, 5 µm, 19 mm X 150 mm. The sample was dissolved in the lowest volume and lowest percentage methanol possible. After priming the system according to the Agilent manual guidelines, the sample was manually loaded onto the cartridge. Using a methanol gradient, as below, 12 ml fractions were collected. The fractions were analysed by LC-MS (Curcumin glycoside LC-MS analysis method) to verify their contents, then pooled accordingly and concentrated using a Genevac centrifugal evaporator.

Time (minutes)	Flow (ml/min)	100 % water, 0.1 % formic acid	100 % acetonitrile, 0.1 % formic acid
0.0	10.00	70.0	30.0
1.0	10.00	70.0	30.0
19.0	10.00	0.0	100.0
20.0	10.00	0.0	100.0

4.3.10 Primer table

Name	Purpose	Sequence
Cl4CL1_R	Amplification of *C. longa* 4CL cDNA DY382977.1	CCAGGCACATCGACAGCAC
Cl4CL1_F	Amplification of *C. longa* 4CL cDNA DY382977.1	GCACCCTTTGACTCTGCTCC
Cl4CL2_R	Amplification of *C. longa* 4CL cDNA DY382750.1	CGCACCGACGTCCTTTACTCTAG
Cl4CL2_F	Amplification of *C. longa* 4CL cDNA DY382750.1	CGGAGATCTCGAACCTGCG
Cl4CL3_R	Amplification of *C. longa* 4CL cDNA DY391029.1	GACATCACGGTGCGGATCG
Cl4CL3_F	Amplification of *C. longa* 4CL cDNA DY391029.1	GGAGGATGAGGAAGAGATCGTCG
Cl4CL4_R	Amplification of *C. longa* 4CL cDNA DY390122.1	GGACTTGACGATGCGGATCG
Cl4CL4_F	Amplification of *C. longa* 4CL cDNA DY390122.1	GCACCTTGTGATCACGAGCG
Cl4CL5_R	Amplification of *C. longa* 4CL cDNA DY391030.1	GGCCTCGTGCCGAATTCG
Cl4CL5_F	Amplification of *C. longa* 4CL cDNA DY391030.1	AACTCCCCTAAGATGGGTGGC
Odt	Amplification of cDNA	TTTTTTTTTTTTTTTTT
Cl4CL1nest_F	Nested amplification of *C. longa* 4CL cDNA DY382977.1	GCAATTCGCCGGAGTTCG
Cl4CL1nest_R	Nested amplification of *C. longa* 4CL cDNA DY382977.1	CGGCCCTGCTTCTGTCA
Cl4CL2nest_F	Nested amplification of *C. longa* 4CL cDNA DY382750.1	CCTCTTGAAACCATATGAACATCT CCACG
Cl4CL3nest_F	Nested amplification of *C. longa* 4CL cDNA DY391029.1	ACCCCGACGACG CG

Cl4CL3nest_R	Nested amplification of *C. longa* 4CL cDNA DY391029.1	CGACAAGTCGTAGGCGTCC
3raceCL1F1	3’ RACE PCR of *C. longa* 4CL cDNA DY382977.1	GCAATTCGCCGGAGTTCGTC
3raceCL1F2	3’ RACE PCR of *C. longa* 4CL cDNA DY382977.1	CCGCTCTTCCACATCTACTCGC
Haz_4CL5_int_L	4CL5 internal primer	CTTCTAAAACCTAAAATATCGCCGGAAG
Haz_4CL5_int_L	4CL5 internal primer	CCCTTGAGATCTTGCTACAAATGCTACT
pYESfseq	Sequencing pYES2 insert	CCTTATTTCTGGGGTAATTAATCAGCGA
pYESrseq	Sequencing pYES2 insert	AGAGGGTTAGGGATAGGCTTACC
T7 forward	Sequencing	TAATACGACTCACTATAGGG
F_Infu_HisDCS	In-Fusion Cloning pYES2-DCS-F2A-CURS1	ATGCATCATCATCATCATCATGAAGCGAACGGCTACCGCATAAC
R_Infu_DCS	In-Fusion Cloning pYES2-DCS-F2A-CURS1	GTCAAAATTCAACAGCTGTACGTAGTTCAGTCTGCAACTATGGAGGACGAC
2A_F	In-Fusion Cloning pYES2-DCS-F2A-CURS1	CAGCTGTTGAATTTTGACCTTCTTAAGCTTGCG
2A*Strep*_infu_R	In-Fusion Cloning pYES2-DCS-F2A-CURS1	GTGGAGGTTGGCCATGTGGTTAATCAAACCTTTTTCAAATTGTGGATGACTCCA
CURS1_F	In-fusion cloning pYES2-DCS-F2A-CURS1	ATGGCCAACCTCCACGCG
CURS1_infu_R	In-Fusion Cloning pYES2-DCS-F2A-CURS1	GCAGAATTCGCCCTTCTACAGTGGCATACTGCGCAGTAC
pYES2_F:	In-Fusion Cloning pYES2-DCS-F2A-CURS1	AAGGGCGAATTCTGCAGATATCCATCAC
pYES2_Infu_R:	In-Fusion Cloning pYES2-DCS-F2A-CURS1	CGCTTCATGATGATGATGATGATGCATATGCCCTATAGTGAGTCGTATTACAGCTGC

4CL52A_Gib_F	Gibson assembly of pYES2-4CL5-F2A-DCS-F2A-CURS1	CTGTAATACGACTCACTATAGGGCATATGGCTAGCTGGAGCCACCCGCAGTTC
4CL52A_Gib_R	Gibson assembly of pYES2-4CL5-F2A-DCS-F2A-CURS1	GCTTCATGATGATGATGATGATGCATGGGCCCAGGGTTGGACTCGACGTC
DCS_482_F	Sequencing DCS	GCGTCATGCTCTACAACGTC
DCS_993_F	Sequencing DCS	GACGCGGCACGTACTCAG
CURS1_286_F	Sequencing CURS1	GACATCGTGGTGGAGGAGATA
pESC_DCS_GAF	Gibson assembly of pESC-DCS	TACGACTCACTATAGGGCCCGGGCGATGGAAGCGAACGGCTACCGCATAA
pESC_DCS_GAR	Gibson assembly of pESC-DCS	TTCGGAAATCAACTTCTGTTCCATGTAGTTCAGTCTGCAACTATGGAGGACGAC
pESCseqF	Sequencing pESC insert	ATTTTCGGTTTGTATTACTTC
pESCseqR	Sequencing pESC insert	GTTCTTAATACTAACATAACT
pESC-CURS_F	Gibson assembly of pESC-DCS-CURS1	TCGTCATCCTTGTAATCCATCGATAACAGTGGCATACTGCGCAGTACAAC
pESC-CURS_R	Gibson assembly of pESC-DCS-CURS1	ATGGCCAACCTCCACGCG
Gal10F	Sequencing pESC insert	GGTGGTAATGCCATGTAATATG
Gal10R	Sequencing pESC insert	GGCAAGGTAGACAAGCCGACAAC
GAL1_F	Sequencing pESC insert	ATTTTCGGTTTGTATTACTTC
GAL1_R	Sequencing pESC insert	GTTCTTAATACTAACATAACT
ACT1_F	Actin PCR control	CTGAGGTTGCTGCTTTGGTT
ACT1_R	Actin PCR control	CGGTGATTTCCTTTTGCATT
F_AscI_4CL5	4CL5 primer	GCGCGCGGCGCGCCATGGCTAGCTGGAGCCAC
R_XbaI_4CL5	4CL5 primer	GCGCGCTCTAGACTATTTAGAGCACATGGTTTC
F_AscI_DCS	DCS primer	GCGCGCGGCGCGCCATGGAAGCGAACGGCTAC
R_XbaI_DCS	DCS primer	GCGCGCTCTAGACTAGTTCAGTCTGCAACT

F_PacI_CaUGT	pET-CaUGT2 cloning	GCGCGCTTAATTAAATGGTTAATCAGCTC
R_SalI_CaUGT	pET-CaUGT2 cloning	GAAAGCAACAAGACTAGTCTTGTTGCTTTC
pBEVY-CaUGT2_GAF	pBEVY-CaUGT2 cloning	AGTCCAAAGCTTGCATGCCTGCAGGCTAGTCTTGTTGCTTTCTTTCAACTTGAGAACGATATGC
pBEVY-CaUGT2_GAR	pBEVY-CaUGT2 cloning	CTTAGTTTCGACGGATCCTCTAGAGATGGCTAGCTGGAGCCACCCGCAGT
GAP_F	Sequencing	CACAAGGCAATTGACCCACG
M13_F	Sequencing	GTAAAACGACGGCCAGT
M13_R	Sequencing	AACAGCTATGACCATG
DCS_qPCR_F	qPCR	GAGAAGGCGATCAAGGAGTG
DCS_qPCR_R	qPCR	GTAGAGCATGACGCGATTGA
4CL5_qPCR_F	qPCR	ATCCCCTGAGACGGAGAGAT
4CL5_qPCR_R	qPCR	GATCACAGTCCCACATGCAC
CURS_qPCR_F	qPCR	CTCCCAACCTCTACGAGCAG
CURS_qPCR_R	qPCR	ATCTCCTCCACCACGATGTC
p19_qPCR_F	qPCR	TCACTGCACAGAGTCCTTGG
p19_qPCR_R	qPCR	AGACCGAATTGCCATCTCAC
3' Nb actin	qPCR	GATACGGGGAGCTAATGCAG
5' Nb actin	qPCR	CGGGAAATTGTTAGGGATGT

Chapter 5. Final Discussion

The biosynthetic production of plant secondary metabolites is a sustainable and green alternative to total chemical syntheses. It can also offer greater yields and simpler extraction processes than purifying natural products from the host plant. This is reflected in the dawn of synthetic and systems biology and the increase in heterologous biosyntheses using metabolic engineering of chassis organisms. With DNA synthesis becoming automated, cheaper and quicker, it is expected that the use of heterologous biosynthetic pathways to produce plant natural products will become ever more prevalent.

Compared to extraction from *C. longa*, synthetic biology and metabolic engineering offered a route to individually produce the highly bioactive curcuminoids, curcumin, DMC and BDMC, with increased purity [245]. Other polyketides synthesized by type III polyketide synthases (PKSs), namely naringenin and resveratrol, have been successfully biosynthesized via a multitude of heterologous enzyme pathways in both *E. coli* and *S. cerevisiae*, as reviewed by Lussier *et al*.[184] This study explored the less investigated heterologous biosynthesis of curcuminoids in *S. cerevisiae* and *N. benthamiana*, as well as the production of curcuminoid analogues by using non-natural starting materials and glycosyltransferase tailoring enzymes. Before the start of the study, curcuminoids had been biosynthesized in *E. coli* by Katsuyama *et al.*[146]. A 4CL from *L. erythrorhizon* and CUS, from rice, was able to produce curcuminoids from exogenous phenylpropanoids giving a yield of 100 mg L^{-1} [146]. Katsuyama *et al*.[105] then used this system to produce non-natural curcuminoids such as di-4-chlorocinnamoylmethane, di-2-furylacryloylmethane and di-3-indoleaclyloylmethane. Unfortunately, this processes also yielded triketide pyrone derailment products.

The current study aimed to utilise phenylpropanoic acids as feedstock due to their presence in bio-refinery waste [246], meaning they would be a sustainable starting material. The conversion of phenylpropanoic acids to curcuminoids required activation by a 4CL, then decarboxylative condensation with two malonyl-CoA units, as catalysed by polyketide synthase(s). Due to CUS allowing the production of derailment products [146] and having no known function in rice [90], two type III polyketide synthases involved in curcuminoid biosynthesis in *C. longa*, DCS and CURS1, were used. Before the start of this study, DCS and CURS1 had not been used to attempt the heterologous production of curcuminoids. However, during the project, Rodrigues *et al*.[107] used DCS and CURS1, along with 4CL1 from Arabidopsis, to produce curcuminoids in *E. coli* from *p*-coumaric acid and ferulic acid.

During this study, a 4CL from *C. longa* was to be identified before being heterologously expressed in yeast and tobacco. Despite a BLAST search yielding five distinct ESTs with high homology to the query sequence of *Os*4CL3, none of the sequences were full-length (see Figure 18). Phylogenetic analysis identified these five potential *C. longa* 4CLs as type III 4CLs, involved in lignin biosynthesis (see Figure 17). Due to the abundance of curcuminoids in the rhizome tissue, it was surprising that none of the putative 4CLs identified as type IV, more commonly associated with secondary metabolite biosynthesis. This was thought to be due to the fact that the type III *Os*4CL3 was used as the BLAST query. However, when type IV *Os*4CL2 was used as a BLAST query, the top 5 BLAST hits were the same as identified as with *Os*4CL3. To obtain a full-length *C. longa* 4CL coding sequence, multiple attempts at RACE were carried out, but this proved unsuccessful within the time constraints of the project.

4CL5, from Arabidopsis was chosen as the first enzyme in the heterologous curcuminoid biosynthetic pathway. This was due to preliminary work within Prof Robert Edwards' research group showing it exhibited better activity when expressed in yeast compared to *Os*4CL3 and other Arabidopsis 4CLs. 4CL5 also had the widest substrate specificity of the known Arabidopsis 4CLs [166]. Since the decision was made to use 4CL5 near the beginning of this project, its ability to accept many different novel cinnamic acid derivatives was shown by Eudes, *et al*. [151]. In this study, 4CL5 was cloned into the pYES2 vector for recombinant expression in yeast (see Figure 21). In *in vivo* pYES2-4CL5 yeast assays, no phenylpropanoyl-CoA ester products were detected, although the phenylpropanoid substrate was consumed quicker than in the negative control (see Figure 22). *In vitro* assays confirmed that 4CL5 was catalysing the desired CoA esterification of the four natural phenylpropanoic acids (see Figure 25). It therefore led to the question as to what was happening to the phenylpropanoyl-CoA ester product within the yeast cell. As phenylpropanoyl-CoA esters do not naturally exist in yeast [183], it was hypothesized that they may be metabolised by the fatty acid β-oxidation pathway, a route that needs to be explored further using knock-outs of CoA ester transporters. Utilising the 4CL5 *in vitro* assay, 4CL5 was shown to exhibit activity towards previously untested substrates 4- methoxycinnamic acid, 3,4-dimethoxycinnamic acid, 4-dimethylaminocinnamic acid and 4-fluorocinnamic-acid chosen for their phenolic substitutions which could give curcuminoids synthesized from these bioactivities (see Figure 27) [80,181,182]. The current work has developed the potential of 4CLs, including 4CL5, for the heterologous biosynthesis of secondary metabolites derived from the phenylpropanoid pathway.

The type III polyketide synthases DCS and CURS1, chosen for curcuminoid biosynthesis, were correctly cloned into pESC and transformed into yeast (see Figure 28 and Figure 36). However, LC-MS analysis could not detect product formation by either enzyme, whether *in vivo* or *in vitro* analysis was performed. According to the literature, many type III PKSs are recalcitrant to

heterologous expression, and aggregate in insoluble inclusion bodies, requiring fusion to solubility tags such as maltose binding protein [247]. However, expression from pESC-DCS appeared to yield soluble DCS, whereas CURS1 protein production was not confirmed. As DCS and CURS1 share 62 % amino acid homology [93], this may contribute to the fact that neither enzyme appeared to be functional. Work utilising these enzymes, by Katsuyama *et al*. [93] and Rodrigues *et al*. [107], both used *E. coli* as a host, and whether codon optimised or not, they catalysed the desired reactions. Yeast is often considered a superior organism for heterologous protein expression, therefore it is interesting that DCS and CURS1 were not found to be active in the current work. Only a few metabolic engineering studies have integrated the subcellular localisation of enzymes into the heterologous production of biosynthetic pathways [41]; it is therefore an avenue which should be explored. The work described here, offers a reminder that there is still much to learn when it comes to heterologous expression of biosynthetic pathways, and that a thorough analysis of the natural system should be carried out before attempting to re-construct the pathway in a heterologous host.

It was hypothesized that, although expressed (see Figure 28), DCS may be inactive due to a codon mismatch between the *C. longa* coding sequence and *S. cerevisiae* translational machinery resulting in an incorrect amino acid sequence due to ribosomal slipping during translation. A comparison of the codon usage of DCS, with the tRNA abundance of *S. cerevisiae* gave three candidate tRNAs which were prevalent in DCS (15 % of the coding sequence) but only had one, or no, analogous *S. cerevisiae* tRNAs (Table 1). These tRNAs were expressed in yeast using the pRS426 vector to increase or introduce the desired tRNA. Addition of *S. cerevisiae* leucine tRNA allowed a doubling in the quantity of DCS protein produced (see Figure 34); however, there was no change in the inactivity of DCS (see Figure 35). Nonetheless, the production of a yeast strain that is engineered to produce tRNAs to rebalance codon bias for heterologous gene expression of non-optimised sequences is a very useful tool: it has the potential to allow simple expression of cDNA libraries and reduce the costs associated with codon optimisation.

During this project, a 2A polypeptide from foot and mouth disease virus (F2A) was investigated for its ability to allow multiple protein production from one polycistron when expressed in Bakers' yeast. The 2A polypeptide has been used as a synthetic biology tool in multiple biosynthetic productions of natural products [248], for example *Thosea asigna* 2A was successfully used in the biosynthesis of β-carotene derivatives in *S. cerevisiae* [198]. F2A was successfully cloned as a linker sequence between the three curcuminoid biosynthetic genes 4CL5, DCS and CURS1 (see Figure 38). It was envisaged that F2A could be used to modulate the stoichiometry of heterologous enzymes production and therefore influence the flux of chemicals through the curcuminoid biosynthetic pathway. When linking DCS and CURS1, F2A gave the desired polypeptide cleavage (see Figure 38), however, its influence on the chemical pathway could not

be ascertained as, alike to when singly expressed, the activities of DCS and CURS1 could not be detected (see Figure 39).

In parallel to curcuminoid biosynthesis in yeast, work was carried out to transiently express the pathway in tobacco using Agrobacterium mediated transformation. 4CL5, DCS and CURS1 were successfully cloned into the pFGC5941 vector and transformed into Agrobacterium (see Figure 45). As was found for to the heterologous expression of the three curcuminoid biosynthetic enzymes in yeast, 4CL5 was the only enzyme activity detected (see Figure 47). Again, there were difficulties with the expression/activity of DCS and CURS1, as ascertained by qPCR (see Figure 46), highlighting the need to answer certain questions about these enzymes in *C. longa*. Where are they localised? How are they regulated? Do they need any specific modifications? Work with tobacco also highlighted known problems with curcumin such as its photo-degradation in light. Additionally, the metabolism and degradation of *p*-coumaric acid by the tobacco leaf tissue illustrates the need for an orthologous heterologous host. However, the fact that tobacco does contain endogenous phenylpropanoic acids and phenylpropanoyl-CoA esters means that there is the possibility to link the endogenous metabolic network with the heterologous curcuminoid production. This would negate the need for supplying a starting material and would yield a more sustainable process.

Despite the diverse range of bioactivities associated with curcuminoids, their physical properties hold them back for drug development. They are readily hydrolysed under basic and acidic conditions, degraded by light and sparingly soluble in water, meaning they are poorly bioavailable under physiological conditions [249–251]. There have been many strategies explored to overcome the poor bioavailability of curcuminoids such as formulating curcumin as nanoparticles or within liposomes [118,252]; administering with adjuvants or synthesizing structural analogues with altered physicochemical properties [66,112]. Before this study, curcumin glucosides had been synthesized both chemically and enzymatically. Chemical synthesis, using tetrabutylammonium bromide, afforded the best yield of curcumin diglucoside, 71 % [135]. Plant cell cultures of *C. roseus* and *P.tricuspidata*, had been shown to produce a diverse range of curcumin glucosides and gentiobiosides using endogenous enzymes, however, this was only performed on a small scale, using only around 1 g of cells [145,231]. Whole cell *E. coli* culture, expressing CaUGT2 had been briefly studied for its ability to glucosylate curcumin: but, it afforded a poor yield of 4 % and only produced curcumin monoglucoside [1]. Part of this work aimed to increase the bioavailability of curcuminoids, by expressing a UGT in *S. cerevisiae* which when fed curcuminoids could produce water soluble glucoside conjugates.

LC-MS metabolite data, from 13 *C. longa* rhizome samples gathered from across Vietnam (Table 2), were analysed for the presence of curcuminoid glucosides and other biological conjugates.

Enzymatic assays were also performed upon protein extracts from *C. longa* rhizomes searching for curcuminoid UGT activity (see Figure 53). There was no detectable evidence suggesting that *C. longa* contained any curcuminoid bioconjugates or curcuminoid UGTs. By studying prevalent secondary metabolites in *C. longa* and literature concerning other Zingiberaceae, it is suggested that instead of using conjugation to sequester curcuminoids, it is likely that they are encapsulated within liposomes of oily compounds such as the sesquiterpenes ar-turmerone and turmerone. Ginger is known to utilise this method to sequester its primary bioactives, such as gingerol, within essential oil lipid bodies [101]. Additionally, detailed metabolite analysis of *C. longa* by Xie at al. [253], suggests the presence of co-regulated metabolic modules, including curcumin and ar-turmerone, which is not observed between other pathways such as monoterpenes and caffeoyl derived curcuminoids. To validate this hypothesis, further studies into the storage of curcuminoids within *C. longa* should be carried out, such as immunolabelling studies, microscopy and searching for evidence of pathway regulators.

CaUGT2's ability to glucosylate the three main curcuminoids was first tested in an *E. coli in vitro* study, as previous work had only investigated its ability to glycosylate curcumin [1]: it was able to glucosylate both DMC and BDMC, as well as curcumin (see Figure 56). CaUGT2 was then cloned into the pBEVY vector and heterologously expressed in yeast (see Figure 57). Although *in vitro* studies showed only curcumin monoglucoside production, whole yeast cell assays showed both mono- and diglucoside production. The difference in products formed depending on *in vivo* or *in vitro* studies should be investigated further and potentially utilised to preferentially produce just one of the glucosides over the other. The glucoside conjugates produced by CaUGT2 expressing cells, were observed within the cell and the media (see Figure 60). The detection of the curcumin glucosides in the culture media is promising from a biotechnological perspective as cell lysis is not necessary to retrieve the products making any large scale up simpler. The CaUGT2 yeast cell culture was scaled up from a shake flask to a 5 L bioreactor and was able to produce 32 mg L^{-1} curcumin monoglucoside (60 % yield).

The change in colour of the culture, as the glucosides appeared in the media (see Figure 65), allowed spectrophotometric monitoring of the process, enabling culture parameters to be altered and easily monitored for optimisation. The effect of different sugars in the medium (galactose, glucose and raffinose) was first explored (see Figure 71 to Figure 74). Despite galactose inducing expression from the pBEVY vector, and glucose repressing expression [233], the addition of either of these sugars showed no effect upon the growth of the culture or the production of curcumin glucosides. Utilising RT-PCR, it was determined that expression of CaUGT2 from the pBEVY vector was leaky so may be the reason the different sugars had no discernible effect upon the promoter. Supplemental raffinose was expected to allow greater cell growth and in turn more

curcumin glycoside production due to an increased abundance of CaUGT2. However, there was an increase in curcumin glycoside production, but without the concomitant increase in cell density. This was unexpected as raffinose is not directly involved in CaUGT2 catalysis, therefore, this result needs further analysis. Addition of iron ($FeSO_4$) to the media allowed an increase in cell growth, believed to be due to overcoming the iron sequestration in the yeast ER caused by curcumin glucosides, but did not allow an increase in curcumin glucoside production (see Figure 75). These findings indicate that something else, such as the amount of available UDP-glucose, may be limiting glucosylation. A systems biology approach could play an important role in the optimisation of CaUGT2 mediated glucosylation of curcumin. Computational analysis should be used to choose and design the optimal number of experiments necessary to converge upon the best media components. Additionally, metabolic modelling of potentially influential yeast metabolic pathways and networks, would allow informed decisions to be made about which endogenous genes could be deleted or over-expressed to fine-tune glucoside product formation. Such analysis was performed in *E. coli* for the heterologous production of naringenin, allowing a 600% increase in yield [20].

As well as altering the pBEVY-CaUGT2 yeast medium to optimise favourable conditions for growth and catalysis, there are many synthetic biology avenues that could be explored that may also allow an increase in yield and greater control over the process. Many studies have been carried out into the effects of using different promoters for increased culture growth, strong protein expression and increased metabolic flux. For example, utilising constitutive promoters such as pTEF1 (transcriptional elongation factor EF-1α) that are regulated by glucose compared to galactose are preferable as yeast prefers glucose as a carbon source [254]. Additionally, altering the terminator is also shown to have positive effects on mRNA stability and protein expression [255]. Effective terminators such as PRM9 and DIT1 should be compared to the current use of CYC1 (Cytochrome c, isoform 1) [255,256].

The functional expression of plant secondary metabolic enzymes in a heterologous host can be a challenging task. Plant natural product pathways are highly complex and each step can take place in a variety of subcellular locations. Although unable to biosynthesize curcuminoids from phenylpropanoids, the first step of the pathway, catalysed by 4CL5, was deemed active in both Bakers' yeast and tobacco. Furthermore, 4CL5 was demonstrated to be a valuable enzyme for the production of novel CoA-esters which could be used for further biosynthesis of non-natural polyketides. This study also allowed the large scale production of curcumin glucosides within a bioreactor, via the heterologous expression of CaUGT2. The yield of curcumin monglucoside produced was 15 fold higher than previous microbial platforms utilising CaUGT2. With further work, to optimise and scale this process, the novel yeast system offers a sustainable route to

produce more bioavailable, medically relevant analogues of curcumin. As curcuminoids are abundant within *C. longa*, further biotechnology processes which concentrate upon their modification to improve solubility and stability would be beneficial to make curcuminoids more biologically relevant.

Appendices

Appendix 1.1 DCS (AB495006.1) cDNA sequence from *C. longa*, 1170 bp

ATGGAAGCGAACGGCTACCGCATAACTCACAGCGCCGACGGGCCGGCGACGATCTTGGCCATCGGCACCG
CCAACCCCACCAACGTCGTCGATCAGAACGCTTATCCCGACTTCTATTTCCGGGTCACCAACTCCGAGTA
TCTGCAGGAACTCAAAGCCAAGTTTAGGCGCATCTGTGAGAAAGCGGCCATCAGGAAGAGGCACTTGTAC
TTGACTGAGGAGATTTTGCGGGAGAATCCTAGCTTGCTGGCTCCCATGGCGCCGTCGTTCGACGCGCGGC
AGGCGATCGTGGTGGAGGCGGTGCCGAAGCTGGCGAAGGAGGCGGCGGAGAAGGCGATCAAGGAGTGGGG
CCGCCCCAAATCGGACATCACGCACCTCGTCTTCTGCTCCGCGAGCGGAATCGACATGCCCGGCTCCGAC
CTGCAGCTTCTCAAGCTGCTCGGGCTCCCGCCGAGCGTCAATCGCGTCATGCTCTACAACGTCGGGTGCC
ACGCCGGTGGCACCGCCCTCCGCGTCGCCAAGGACCTCGCGGAGAACAACCGCGGCGCGCGGGTGCTCGC
CGTCTGCTCCGAGGTCACCGTGCTCTCCTACCGCGGCCCCCACCCCGCCCACATCGAGAGCCTCTTCGTC
CAAGCTCTGTTTGGCGACGGCGCCGCCGCGCTCGTGGTCGGGTCCGACCCCGTCGATGGCGTCGAGCGCC
CCATCTTCGAAATCGCCTCGGCATCCCAAGTGATGCTTCCGGAGAGCGCAGAGGCGGTGGGCGGCCACCT
CCGCGAAATTGGGCTGACCTTCCACCTCAAGAGCCAGCTTCCGTCGATCATCGCGAGCAACATCGAGCAG
AGCCTGACGACTGCGTGCTCGCCGCTGGGGCTGTCGGACTGGAACCAGCTGTTCTGGGCGGTTCACCCCG
GCGGCCGAGCGATCCTGGACCAGGTGGAGGCGCGGCTCGGACTGGAGAAGGACCGGCTCGCCGCGACGCG
GCACGTACTCAGCGAGTACGGCAACATGCAGAGCGCCACGGTGCTGTTCATCCTGGACGAGATGCGGAAC
CGCTCGGCTGCGGAGGGCCACGCCACCACCGGCGAGGGGCTCGACTGGGGCGTGCTGTTGGGCTTCGGCC
CGGGACTCTCCATCGAGACCGTCGTCCTCCATAGTTGCAGACTGAACTAG

Appendix 1.2 CURS1 (AB495007.1) cDNA sequence from *C. longa*, 1170 bp

ATGGCCAACCTCCACGCGTTGCGCAGGGAGCAGAGGGCTCAAGGTCCTGCCACCATCATGGCCATCGGGA
CCGCCACCCCTCCCAACCTCTACGAGCAGAGCACCTTCCCGGACTTCTACTTCCGCGTCACCAACTCCGA
CGACAAGCAGGAGCTCAAGAAAAAGTTCCGCCGCATGTGCGAGAAGACGATGGTGAAGAAGCGGTACCTG
CACTTGACCGAGGAGATCCTGAAGGAGAGGCCCAAGCTCTGCTCCTACAAGGAGGCGTCGTTCGACGACC
GGCAGGACATCGTGGTGGAGGAGATACCGAGATTGGCTAAGGAAGCGGCGGAGAAGGCCATCAAGGAGTG
GGGGCGGCCCAAATCGGAGATCACCCACCTGGTCTTCTGCTCCATCAGCGGGATCGACATGCCCGGCGCC
GACTACCGCCTCGCCACGCTCCTCGGGCTCCCTCTCACCGTCAACCGCCTCATGATCTACAGCCAGGCCT
GCCACATGGGCGCCGCCATGCTCCGCATCGCCAAGGACCTCGCCGAGAACAACAGGGGCGCGCGCGTGCT
GGTGGTCGCCTGCGAGATCACCGTGCTCAGCTTCCGCGGCCCGAACGAGGGCGACTTCGAGGCGCTCGCG
GGGCAGGCCGGCTTCGGCGACGGCGCGGGGGCCGTCGTCGTCGGGGCCGACCCGCTGGAAGGAATTGAAA
AACCCATCTACGAGATCGCGGCGGCGATGCAGGAGACGGTGGCGGAGAGCCAGGGGGCGGTGGGCGGCCA
CCTGCGGGCCTTCGGCTGGACGTTCTACTTCCTGAACCAGCTGCCGGCGATCATCGCCGACAACCTCGGG
AGGAGCCTGGAGCGGGCGTTGGCGCCGCTGGGGGTGAGGGAGTGGAACGACGTCTTCTGGGTGGCGCACC
CGGGCAACTGGGCCATCATTGACGCCATCGAAGCCAAGCTGCAGCTGAGCCCGGACAAGCTCAGCACCGC
CCGCCACGTCTTCACAGAGTACGGCAACATGCAGAGCGCCACCGTGTACTTCGTGATGGATGAGCTGAGG
AAGCGGTCGGCGGTGGAGGGGCGGAGCACCACCGGCGACGGCTTGCAGTGGGGAGTTCTCCTCGGTTTTG
GGCCGGGCCTCAGCATCGAGACCGTTGTACTGCGCAGTATGCCACTGTAG

Appendix 1.3 CaUGT2 (AB159213.1) cDNA sequence from *C. roseus*, 1720 bp

CACTGTACCTCAATTCCATCTTCAATTTTTCCATTCATTTCATCATTTTGAGGTAGAAGAAGAAGAAGA
GGCCATTAATATGGTTAATCAGCTCCATATTTTCAACTTCCCATTCATGGCACAGGGCCATATGTTACCC
GCCTTAGACATGGCCAATCTATTCACTTCTCGTGGAGTCAAAGTAACATTAATCACAACCCATCAACATG
TTCCCATGTTTACAAAATCCATAGAAAGGAGCAGAAATTCTGGATTTGATATATCCATTCAATCCATCAA
ATTCCCAGCTTCAGAAGTTGGTTTACCTGAAGGAATCGAAAGTCTAGATCAAGTTTCAGGGGACGACGAA
ATGCTTCCTAAGTTCATGAGAGGAGTTAATTTACTCCAACAACCTCTCGAACAACTATTGCAAGAATCTC
TCCTCATTGTCTTCTTTCTGATATGTTCTTCCCTTGGACTACTGAATCTGCTGCTAAATTTGGTATTCCC
AGATTGCTTTTTCATGGGTCCTGTTCCTTTGCCCTCTCTGCAGCTGAAAGTGTGAGAAGAAATAAACCTT
TCGAGAATGTTTCCACAGACACAGAGGAATTTGTTGTGCCTGATCTTCCCCACCAAATTAAATTAACCAG
AACACAAATTTCAACATACGAAAGGGAAAATATTGAGTCAGATTTTACCAAAATGCTGAAGAAAGTTAGG
GATTCAGAATCCACATCTTACGGAGTTGTAGTCAATAGTTTCTATGAACTTGAACCAGATTATGCCGATT
ATTACATCAACGTTTTGGGAAGAAAAGCATGGCATATAGGGCCTTTTTTGCTTTGTAACAAATTACAAGC

TGAAGATAAAGCCCAAAGGGGGAAGAAATCAGCAATTGATGCAGACGAATGTTTAAATTGGCTTGATTCG
AAACAACCAAATTCCGTAATTTATCTCTGTTTCGGAAGTATGGCCAATTTAAATTCTGCCCAATTACACG
AAATTGCAACAGCCCTTGAATCCTCCGGCCAAAATTTCATCTGGGTTGTTAGAAAATGTGTGGACGAAGA
AAACAGTTCAAAATGGTTTCCAGAAGGATTCGAAGAAAGAACAAAAGAAAAAGGGCTAATTATAAAGGGA
TGGGCACCACAAACCCTAATTCTTGAACACGAATCAGTAGGAGCATTTGTTACCCATTGTGGTTGGAATT
CAACTCTTGAAGGAATCTGCGCAGGGGTTCCTCTGGTGACTTGGCCTTTCTTTGCTGAGCAATTTTTCAA
TGAGAAATTGATTACAGAGGTACTGAAAACGGGATACGGAGTTGGGGCTCGGCAATGGAGTAGAGTTTCA
ACAGAGATTATAAAAGGAGAAGCCATAGCTAATGCTATTAATCGAGTAATGGTGGGTGATGAAGCTGTTG
AGATGAGAAACAGAGCAAAAGATTTGAAGGAAAAGGCAAGAAAAGCTTTGGAAGAAGATGGATCTTCTTA
TCGTGATCTTACTGCTCTTATTGAAGAATTGGGGGCATATCGTTCTCAAGTTGAAAGAAAGCAACAAGAC
TAGGAAATAAGAAGAGAGAGAGAATTCTCATTTATATTATTTTGTCCATTTAATCAAATTACTATATTC
TGTAAAATTTGCCATTTTTATTATGGAATCTAGAAGGCAGTCTATATATCTTCTTGCCAGGCACTAAAG
AAATATTTGTCATTTCAAGTTTAAAAAAAAAAAAAAAAAAA

The CDS was significantly longer than other known UGTs. Alignment with other Group D UGTs, TOGT1, UBGT, UGT73B3, UGT73B4 and *Vm*UFGT2 showed the AB159213.1 nucleotide sequence was longer than necessary (Figure 76). Therefore, the region underlined, which most likely corresponded to the coding sequence, was used for work in this thesis. The underlined CDS is 1464 bp.

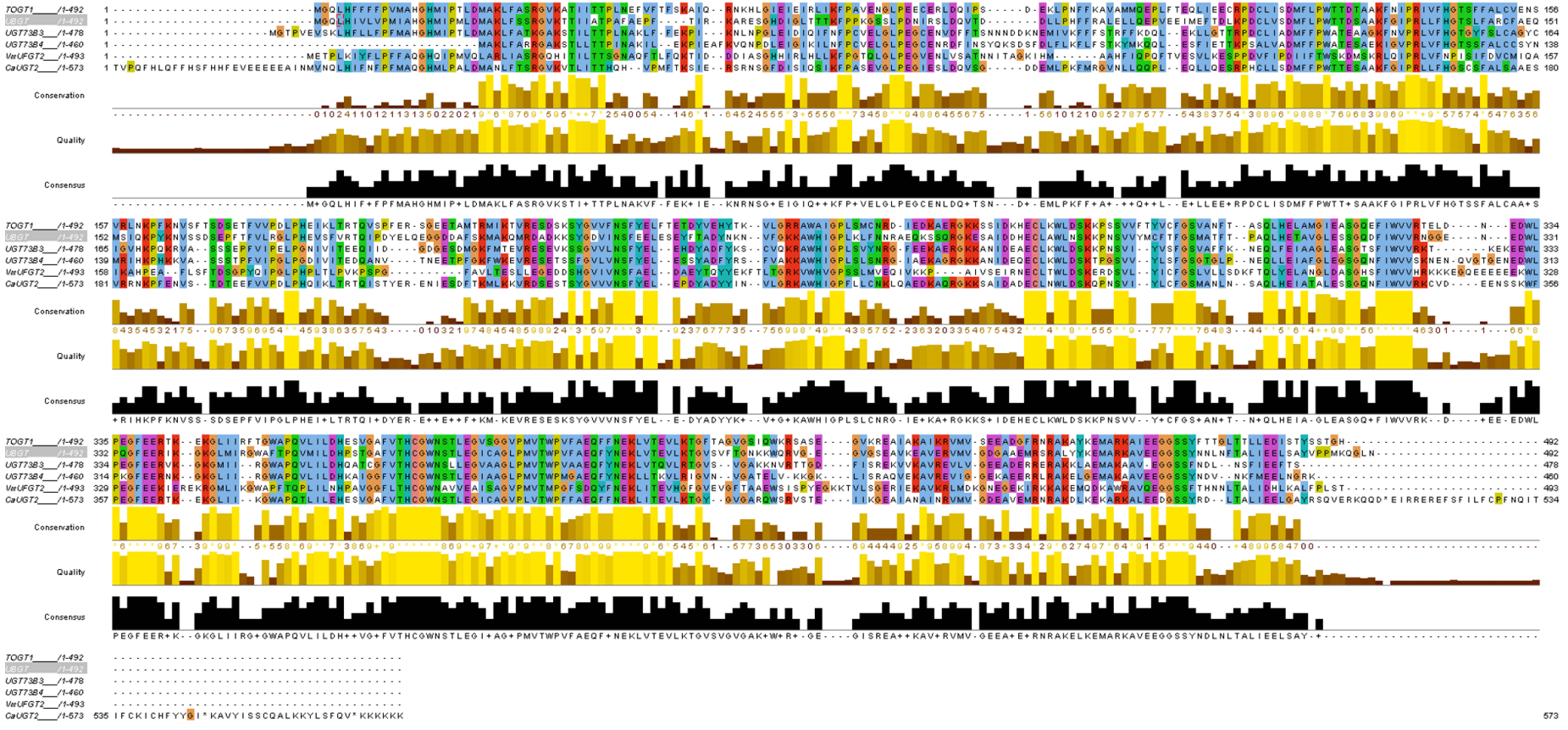


Figure 76. Sequence alignment of CaUGT2 with Group D UGTs.
Alignment of CaUGT2 with other Family 1 Group D UGTs from Arabidopsis, *Nicotiana tabacum* and *Vigna mungo*: *Nt*TOGT1, UBGT, *At*UGT73B3, *At*UGT73B4, *Vm*UFGT2 using JalView. Amino acids are shown in different colours, the conservation and consensus of amino acids is shown below. The brighter the yellow colour, the greater the homology between the amino acid sequences. CaUGT2 is significantly longer than the other sequences.

Appendix 2.1 Five potential 4CL ESTs identified from *C. longa*

DY382977.1

GCACCCTTTGACTCTGCTCCGCAATTCGCCGGAGTTCGTCTTTGCCTTCCTTGGTGCTTCCCTCTGCAGC
GCCGTGGCGACCACCGCCAACCCGTTTTATACTAGGGCGGAGATCCACAAGCAGGCGGCAGGAGCCGGCG
CGAGGGTAATCGTCACGGAGTCGTGCTACGTCGACAAAGTGAGGGACTTCGCGCGGGAGCGAGGGATTAT
AATCGTCTGTGTTGACGATCCTCTGCCGGAGGGATGCTGGCGATTCTCTGAGTTGCTGGAAGCAGAGGAG
AAAACAAAGGAGGCTGAGGAGGAGGAAGAGATCGTCGATCCCGACGACGTGGTGGCGTTGCCCTACTCGT
CCGGCACGACAGGGCTGCCGAAAGGGGTGATGCTGACTCACCGGAGCCTGATCACGAGCGTGGCGCAGCA
GGTGGACGGGGAGAACCCCAACTTGTACTTCCACCCCGACGACGTGCTCCTCTGCGTGCTGCCGCTCTTC
CACATCTACTCGCTGAACTCGGTGTTGCTATGTGGGCTCCGCGCCGGCTCTGCCATCCTCCTGGTGCGTA
AGTTCGAGATCTCCGCCGTGCTGGAGCTTGTGCAGCGATACCGCGTCACAATCGCGCCGTTTGTGCCACC
CATCGTGCTGGAGTTCGTCAAGAGCCCGCTCGTGGACGCCTACGACTTGTCGTCGATCCGCACCGTGATG
TCCGGAGCCGCCCCCATGGGCAAGGAGCTCGAGGACAAGTTCATGGCCCAGCTCCCAAACGCTACGTTGG
GCCAAGGTTACGGGATGACAGAAGCAGGGCCGGTGCTGTCGATGTGCCTGGCC

DY391029.1

AGAAAACGGAGGATGAGGAAGAGATCGTCGACCCCGACGACGTCGTGGCGCTGCCCTACTCGTCCGGCAC
GACGGGGCTACCGAAGGGGGTGATGCTGACTCACCGGAGCCTGATCACGAGCGTGGCGCAGCAGGTGGAC
GGGGAGAACCCCAACCTGTACTTCCACCACGACGACGTGCTCCTCTGTGTGCTGCCGCTCTTCCACATCT
ACTCGCTGAACTCGGTGTTGCTATGTGGGCTCCGCGCCGGCGCTTCCATCCTCCTGGTGCGTAAGTTCGA
GATCTCCTCCGTGCTGGAGCTTGTGCAGCGATACCGCGTCACAATCGCGCCGTTTGTGCCACCCATCGTG
CTGGAGTTCGTCAAGAGCCCGCTCGTGGACGCCTACGACTTGTCGTCGATCCGCACCGTGATGTCCGGCG
CCGCCCCCATGGGCAAGGAG

DY382750.1

CCAACATGGCGGAGATCTCGAACCTGCGCATGAGGAGGATGGCGGCGCCGACGCGGAGGCCGCAGAGCAG
CACCGAGTTGAGGGAGTAAATGTGGAAGAGAGGCAGGACGCAGAGCAGCACGTCTTCCTTCCGGAAGTAG
AGGTTCGGGTTGTCGCCGTCCACCTGCTGCGCCACGCTCGTGATCAGGCTCCGGTGAGTCAGCATCACCC
CCTTGGGAAGCCCCGTCGTCCCCGATGAGTACGGCAGCGCCACTACTTCGTCCGCGTCGATCTCCTCCTC
CTCCTCCTCTGTTTCCTCCGCTTCCTTCGTTAACTCCGAGAAGCTCCTGCAGCCCTCCGGCAGAGGATCG
TCCACACAGACGACCACCACCCCTCGCTCACGCGCGAACTCCCGCACCTTGTCCACGTAGCACGATTCCG
TGACGATCACCCTCGCTCCGGCGTCCGCGCCTGCTTGTGAATCTCTGCGGGGGTGTAGAACGGATTCGCC
CCTCGTGCCGAATTCGGCACGAGGCTTAATTCCTTTGGAGTCAGGATCACGTCAATAATAAGATCGTCAA
TTCTCCAAGAATTGCTTTTGTTTTAATAGCTGGAGCTGGTTTTATGAGTATGTTTCAGGAACTCGTGGAG
ATGTTCATATGGTTTCAAGAGGTTCTAGAGTAAAGGACGTCGGTGCG

DY391030.1

GGNTCTTGACGAACTCCCCTAAGATGGGTGGCACAAACGGCGCGATTGTGACGCGGTATCGCTGCACAAG
CTCCAGCACGGAGGAGATCTCCAACTTACGCACCAGGAGGATGGAAGCGCCGGCGCGGAGCCCACATAGC
AACACCGAGTTCAGCGAGTAGATGTGGAAGAGCGGCAGCACACAGAGGAGCACGTCGTCGTGGTGGAAGT
ACAGGTTGGGGTTCTCCCCGTCCACCTGCTGCGCCACGCTCGTGATCAGGCTCCGGTGAGTCAGCATCAC
CCCCTTCGGTAGCCCCGTCGTGCCGGACGAGTAGGGCAGCGCCACGACGTCGTCGGGGTCGACGATCTCT
TCCTCCTCCTCCGTTTTCTCCTCGTGCCGAATTCGGCACGAGGCCTCGTGC

DY390122.1

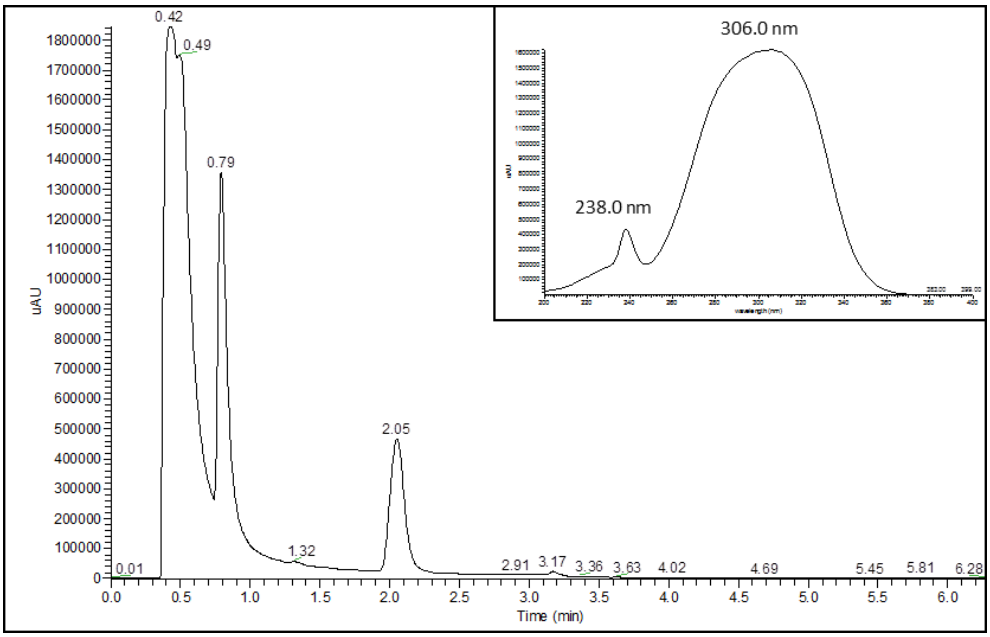
GCACCTTGTGATCACGAGCGTGGCGCAGCAGGTGGACGGCGAGTACCCGAACCTGTACTTCCACTGCGAC
GACGTGCTCCTCTGCGTGCTGCCTCTCTTCCACATCTACGCGCTGAACTCGGTCCTGCTCTGCGGGCTCC
GCGCCGGCAGCTCCATCCTCTTAGTGCGCAAGTTCGAGATCTCCGCCGTGCTGCAGCTGGTGCAGCGCTT
CCGCGTCACGGTGGCCGCCTTCGTTCCGCCCATCGTGCTGGATTTCGTCAAGAGCTCGCTCATGGACGCC
TACGACCTGTCGTCGATCCGCATCGTCAAGTCCGGCGCCGCCCCCATGGGCAAGGAA

Appendix 2.2 4CL5 (AT3G21230) cDNA sequence from *A. thaliana*, 1720 bp

ATGGTGCTCCAACAACAAACGCATTTTCTCACAAAGAAGATAGGTCAAGAAGACGAAGAAGAAGAACCTT
CTCATGATTTCATTTTCCGATCTAAGCTTCCTGATATCTTCATCCCTAACCACCTCCCTCTCACCGATTA
CGTCTTCCAAAGGTTCTCCGGCGATGGCGACGGAGATTCCTCAACAACCTGTATCATCGACGGTGCCACT
GGCCGTATTTTGACCTACGCCGATGTGCAAATCAACATGCGGAGGATAGCTACCGGAATCCACCGGCTCG
GGATCCGCCACGGTGACGTCGTAATGCTTCTCCTCCCGAACTCGCCGGAGTTTGCTCTGTCTTTTCTAGC
CGTTGCATACCTTGGAGCCGTTTCCACTACCGCGAATCCGTTCTACACTCAGCCGGAGATCGCTAAACAG
GCTAAAGCCTCCGCCGCGAAGATGATCATCACGAAGAAATGTTTAGTCGATAAACTAACAAACCTCAAAA
ACGACGGCGTTTTGATCGTTTGCCTAGACGATGACGGTGACAACGGCGTTGTTTCGTCGAGTGATGATGG
TTGCGTGAGTTTCACGGAACTGACTCAAGCGGACGAAACAGAACTTCTAAAACCTAAAATATCGCCGGAA
GACACGGTGGCGATGCCGTACTCTTCCGGAACAACGGGACTTCCAAAGGGAGTGATGATTACTCACAAGG
GTTTAGTTACAAGCATCGCTCAGAAAGTCGACGGAGAAAACCCTAACCTCAACTTCACGGCCAATGACGT
CATCCTCTGTTTTCTCCCAATGTTTCACATTTACGCTCTCGACGCGTTGATGCTTTCGGCGATGAGAACC
GGTGCAGCGTTACTGATCGTGCCGAGGTTCGAGTTAAATCTGGTGATGGAGCTGATTCAAAGGTACAAAG
TCACGGTTGTTCCGGTGGCTCCTCCGGTGGTTTTAGCATTCATTAAATCCCCTGAGACGGAGAGATACGA
CTTGAGCTCGGTGAGAATTATGCTATCAGGTGCAGCTACACTAAAGAAGGAGCTTGAAGACGCCGTGCGA
CTCAAGTTTCCTAATGCCATTTTTGGTCAGGGTTATGGAATGACCGAGTCCGGAACGGTAGCTAAGTCAT
TAGCGTTTGCAAAGAACCCATTTAAAACCAAGTCCGGTGCATGTGGGACTGTGATCAGAAACGCCGAGAT
GAAAGTGGTCGATACCGAAACCGGAATCTCCTTGCCGCGCAACAAATCTGGCGAGATCTGCGTCCGAGGT
CATCAGCTCATGAAAGGTTATTTGAATGATCCGGAAGCTACTGCACGAACCATAGACAAAGACGGGTGGT
TACATACAGGAGATATTGGGTTTGTGGATGATGATGATGAGATCTTCATTGTTGATCGGTTGAAGGAACT
GATCAAATTCAAAGGCTACCAAGTGGCTCCAGCTGAGCTTGAAGCATTGCTTATTTCTCATCCTTCTATC
GATGATGCCGCAGTTGTTGCAATGAAGGATGAAGTAGCTGATGAGGTTCCAGTAGCATTTGTAGCAAGAT
CTCAAGGGTCTCAGTTAACTGAAGATGATGTCAAGAGTTACGTAAACAAACAGGTGGTTCACTACAAGCG
GATTAAGATGGTGTTTTTCATAGAAGTTATACCAAAAGCAGTTTCGGGCAAGATTCTGAGGAAGGATCTG
CGAGCTAAATTGGAAACCATGTGCTCTAAATAG

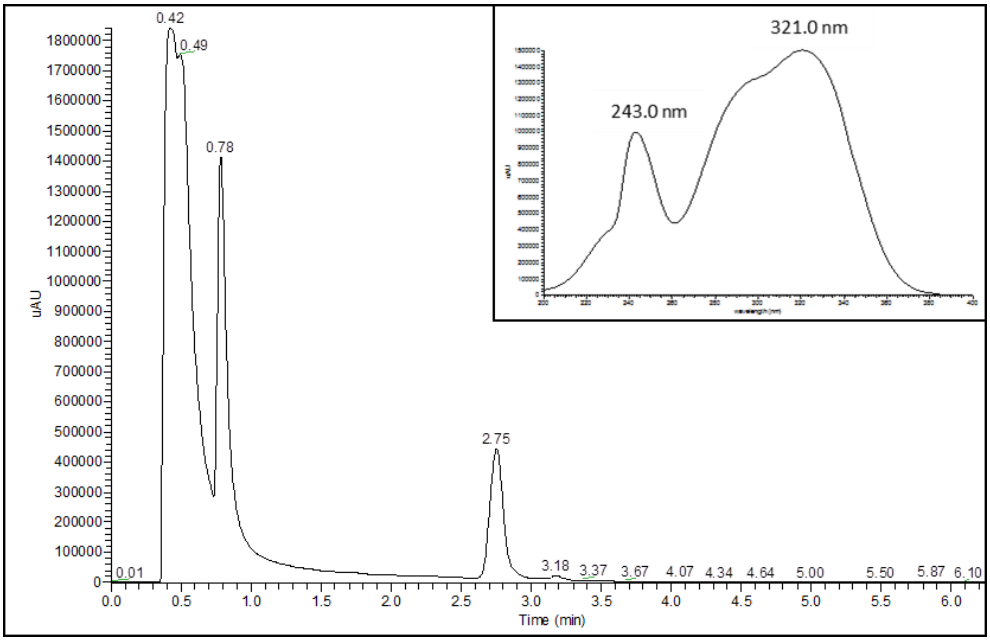
The cDNA sequence was amplified from Arabidopsis (unpublished work, Dr H. Housden, Prof Rob Edward's Lab). Underlined bases are 10 point mutations which differ to the NCBI database. They cause three amino acid substitutions.

Appendix 2.3 *p*-Coumaric acid HPLC chromatogram



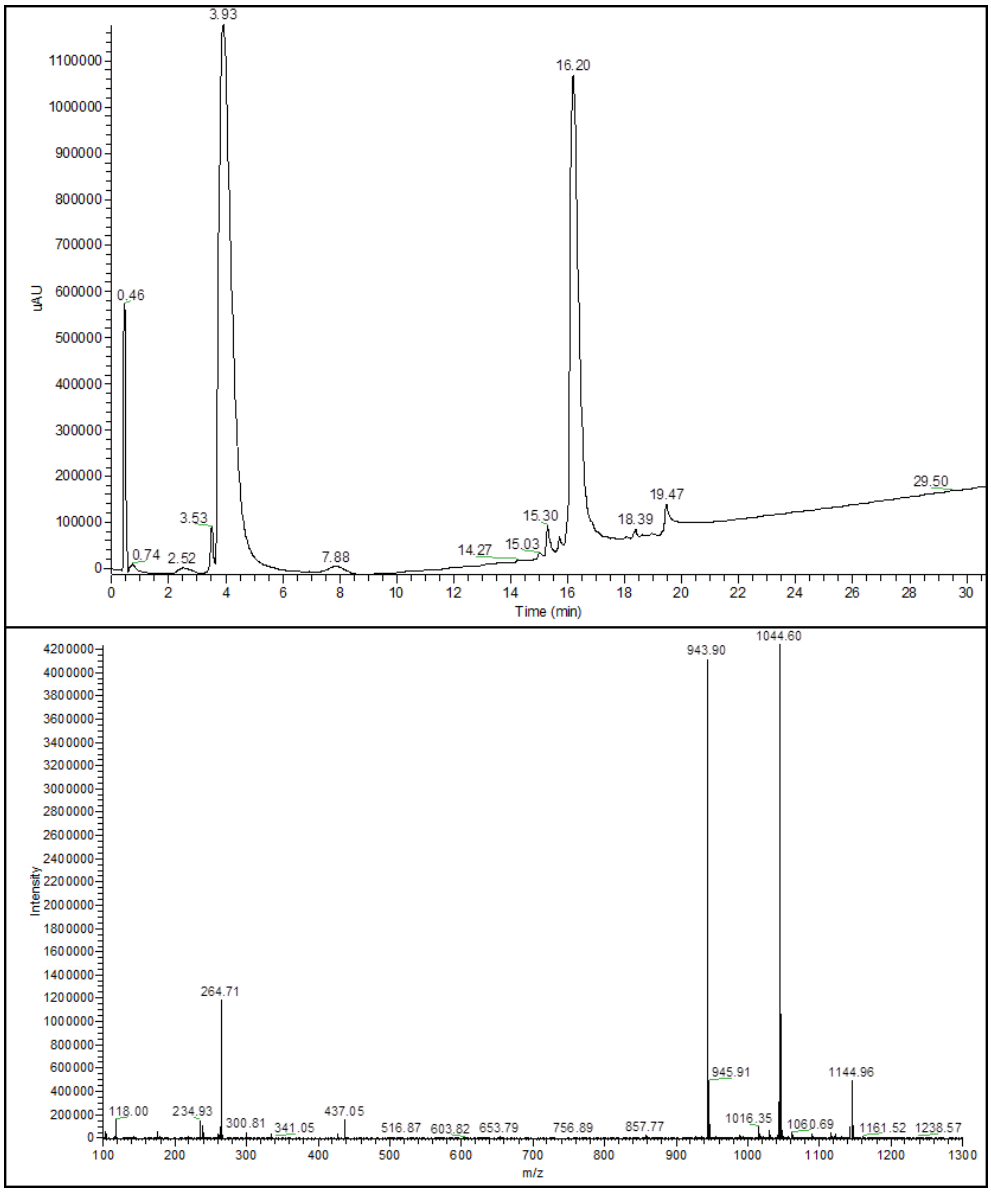
Using phenylpropanoyl-CoA ester LC-MS analysis method 1, to analyse *p*-coumaric acid. HPLC chromatogram (260 nm), RT 2.05, UV max 306 nm.

Appendix 2.4 Ferulic acid HPLC chromatogram



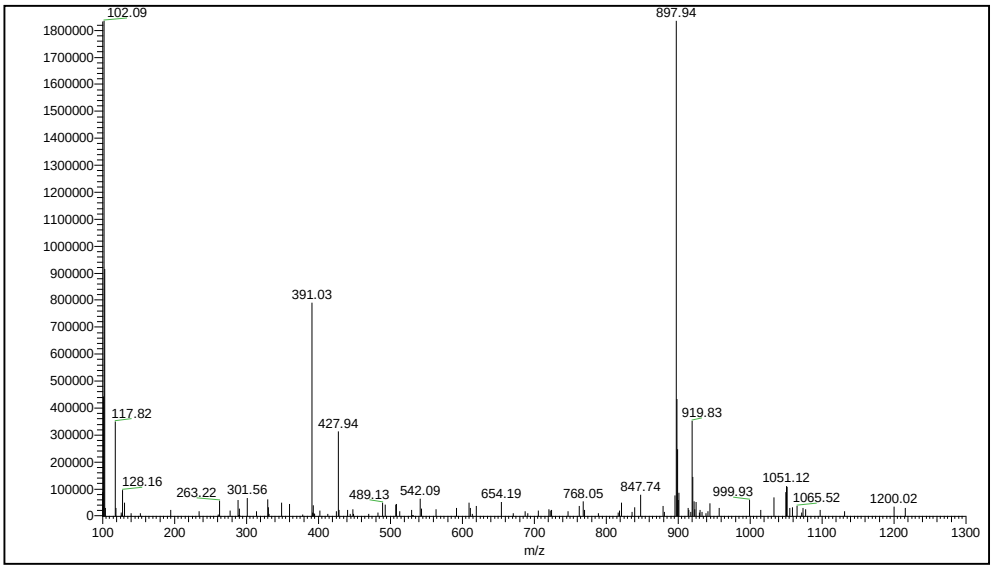
Using phenylpropanoyl-CoA ester LC-MS analysis method 1, to analyse ferulic acid. HPLC chromatogram (260 nm), RT 2.75, UV max 321.0 nm.

Appendix 2.5 Feruloyl-CoA HPLC chromatogram and mass spectrum



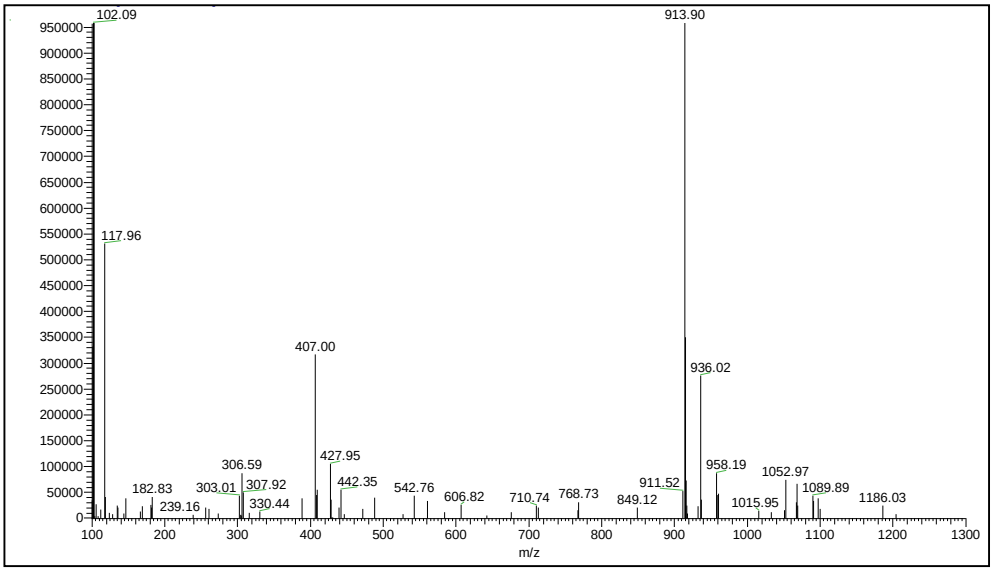
Using phenylpropanoyl-CoA ester LC-MS analysis method 2, to analyse feruloyl-CoA (M_r 943.70 g mol^{-1}). HPLC chromatogram (260 nm), RT 16.2. MS: *m/z* 943.9 $[M]^+$, *m/z* 437.1 $[M\text{-}ADP]^+$.

Appendix 2.6 Mass spectrum of cinnamoyl-CoA



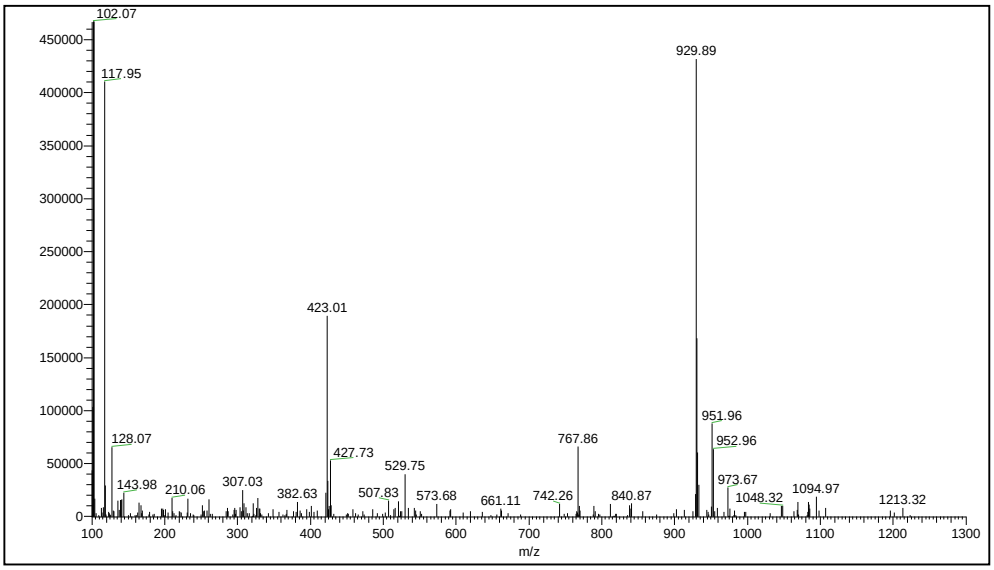
Using phenylpropanoyl-CoA ester LC-MS analysis method 2, to analyse caffeoyl-CoA (M_r 897.68g mol^{-1}). MS: *m/z* 897.9 $[M]^+$, *m/z* 427.9 $[AMP]^+$, *m/z* 391.0 $[M\text{-}ADP]^+$.

Appendix 2.7 Mass spectrum of *p*-coumaroyl-CoA



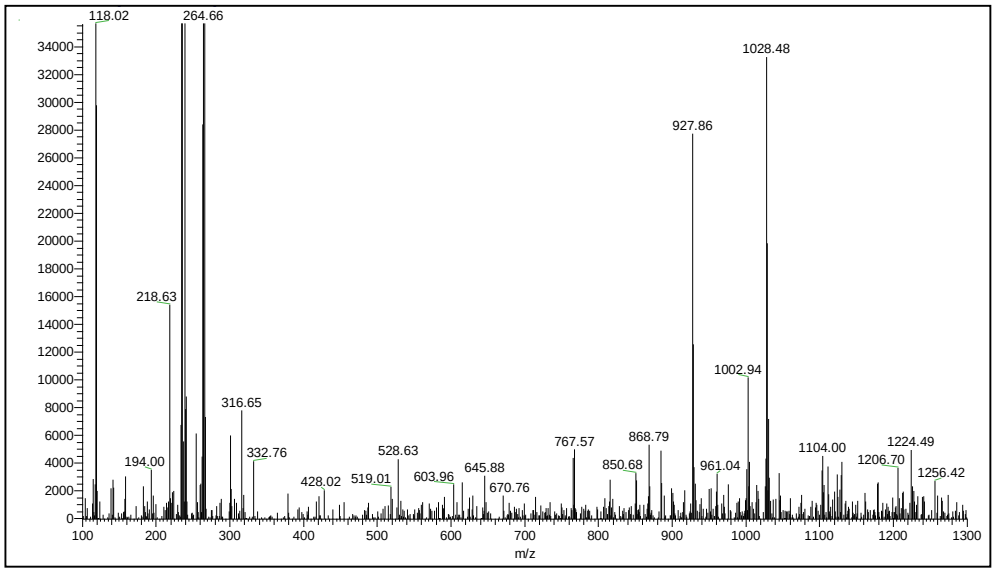
Using phenylpropanoyl-CoA ester LC-MS analysis method 2, to analyse *p*-coumaroyl-CoA (M_r 913.67 g mol^{-1}). MS: *m/z* 913.9 $[M]^+$, *m/z* 428.0 $[AMP]^+$, *m/z* 423.0 $[M\text{-}ADP]^+$.

Appendix 2.8 Mass spectrum of caffeoyl-CoA



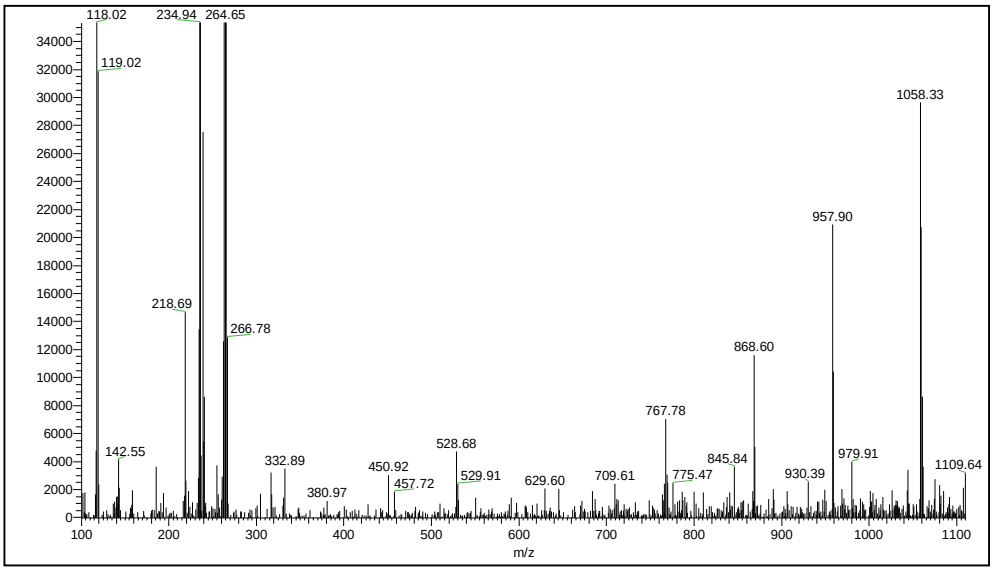
Using phenylpropanoyl-CoA ester LC-MS analysis method 2, to analyse caffeoyl-CoA (M_r 929.68 g mol^{-1}). MS: *m/z* 927.9 $[M]^+$, *m/z* 767.9 [M-caffeic acid]$^+$, *m/z* 529.8 [M-AMP]$^+$, *m/z* 427.7 $[AMP]^+$, *m/z* 423.0 [M-ADP]$^+$.

Appendix 2.9 Mass spectra of 4-methoxycinnamoyl-CoA



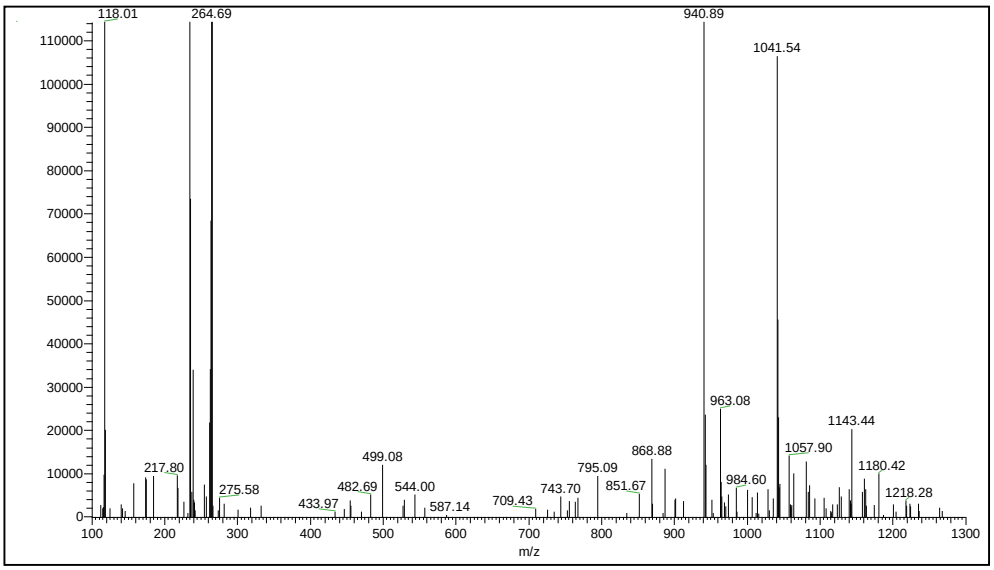
Using phenylpropanoyl-CoA ester LC-MS analysis method 2, to analyse 4-methoxycinnamoyl-CoA (M_r 927.73 g mol^{-1}). MS: *m/z* 927.9 $[M]^+$. Major peaks below 300 (m/z 118.0, 264.7) are hypothesized to be TEA and water complexes.

Appendix 2.10 Mass spectrum of 3,4-dimethoxycinnamoyl-CoA



Using phenylpropanoyl-CoA ester LC-MS analysis method 2, to analyse 3,4-dimethoxycinnamoyl-CoA (M_r 957.73 g mol^{-1}). MS: *m/z* 957.9 $[M]^+$. Major peaks below 300 (m/z 118.0, 264.7) are hypothesized to be TEA and water complexes.

Appendix 2.11 Mass spectra of 4-dimethylaminocinnamoyl-CoA



Using phenylpropanoyl-CoA ester LC-MS analysis method 2, to analyse 4-dimethylaminocinnamoyl-CoA (M_r 940.74 g mol^{-1}). MS: *m/z* 940.9 $[M]^+$. Major peaks below 300 (m/z 118.0, 264.7) are hypothesized to be TEA and water complexes.

Appendix 2.12 Supplemental yeast tRNA coding sequences

S. pombe Alanine tRNA (Ala)

ChrI:4796800-4797111
TAGTCAATAAATAAAACAAATGTGGATAAAAAGGAAAAATAGTTTCTAACATTGGAGATGCCGGGAATCG
AACCCGGGACCTTTTCAAGGTGCGAACTATGCGAATGAAACGTTATACCCCTAAACTACATCCCCAGTCG
ATATGATGTTATAATTTAACCCATTATTAATAAAAGGTTAAAAAAAAGAACGAATGTCAAGCTATAAAAT
AAAATAATCATCATCATCATTGATTAATTTAGCAAGGGTGCCTGCATCTAGTTTTTTTATCAGTCAATAA
CTCATACATTAACCGCATTCATAAGAGGCTAT

S. pombe Leucine tRNA (Leu1)

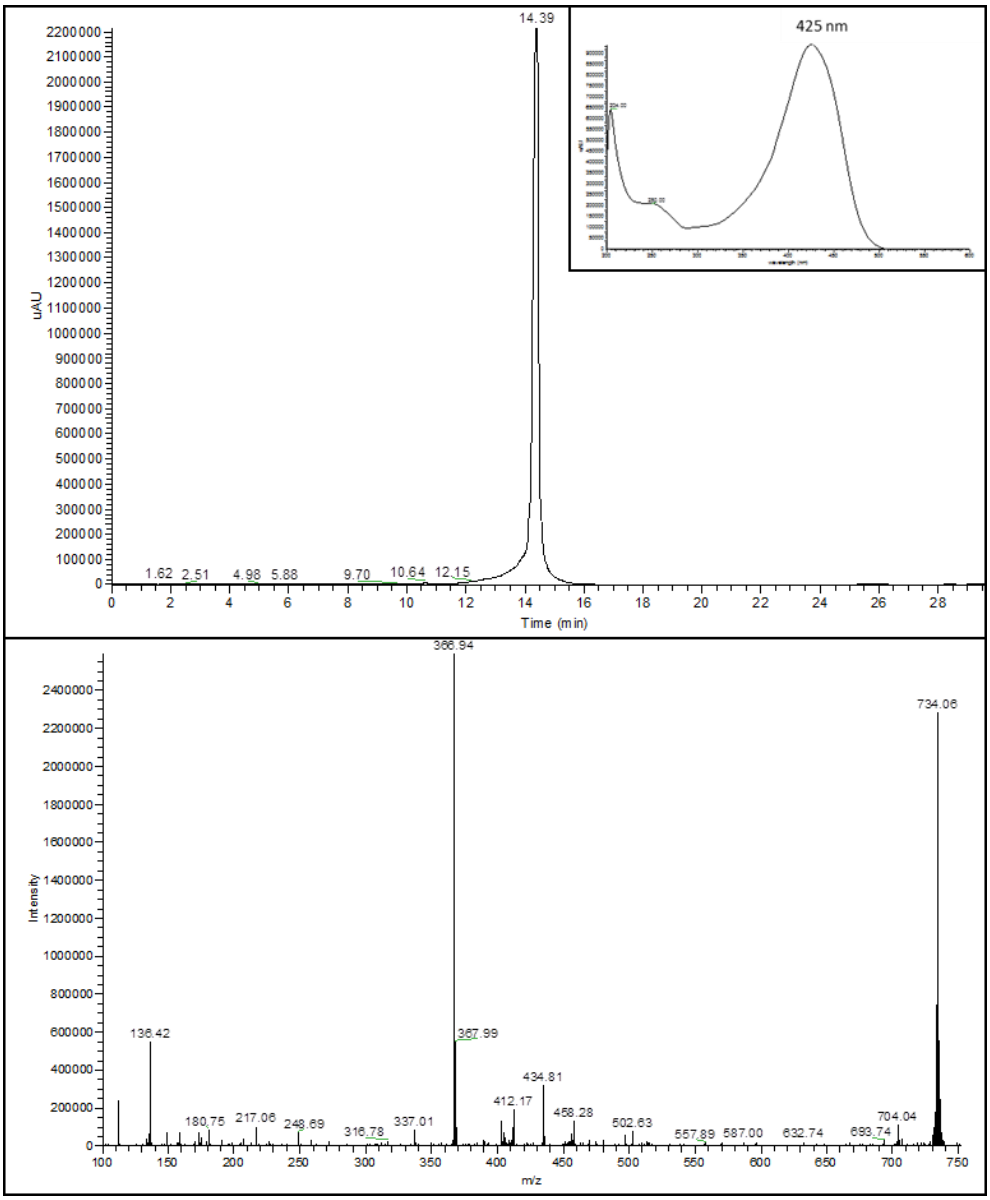
ChrI:1527361-1527050
TTTTCGTTCAATCAAGATATTTTATCACATTATTTCAAGGTTCAATGAATCATTCGAAGAATAGTAACGA
ATTCAGATATATTAAATAAAGGCACACATAATCTCTAAAGATTACAGTTTAATGGCTCTTTCAAAGTGTT
TAGTAATATATACACTAAATTTCTAAAATTCAGTAGTTATGGCGAAGTGGCCGAGTGGTCTATGGCGCTA
GCTTCAGGTGAAAGCAATTGCTAGTCTACGTATGTGGGCGTGGGTTCGAACCCCACCTTCGTCAATATAA
TTTTTTATATCATGTTAATGGAATTGAGAAAA

S. cerevisiae Leucine tRNA (Leu 2)

ChrVII:700500-700800
TCTGCGTGGATTGCCACTTATAGGGCTTTTCAAATTAGTATTTTAAAGGAGTTAGATCGGTTTCAAAAGG
TCATGCTCCGGAGTAATTGACTTGATTTTGCAAAGCTTCACTTAAATGGCGTATCTTCCTTTTATAAATT
GCGTTTGCTTTATTCCCTTACCCAACAGATGAATTGGTACTCTGGCCGAGTGGTCTAAGGCGTCAGGTCG
AGGTCCTGATCTCTTCGGAGGCGCGGGTTCAAACCCCGCGGGTATCAATATTTTTTTGAATCTACGCACT
TCAATAAGAATTAGGCTGTT

The whole nucleotide sequences above were cloned into the pRS41/26 vectors. Underlined are the nucleotides that correspond to the tRNA.

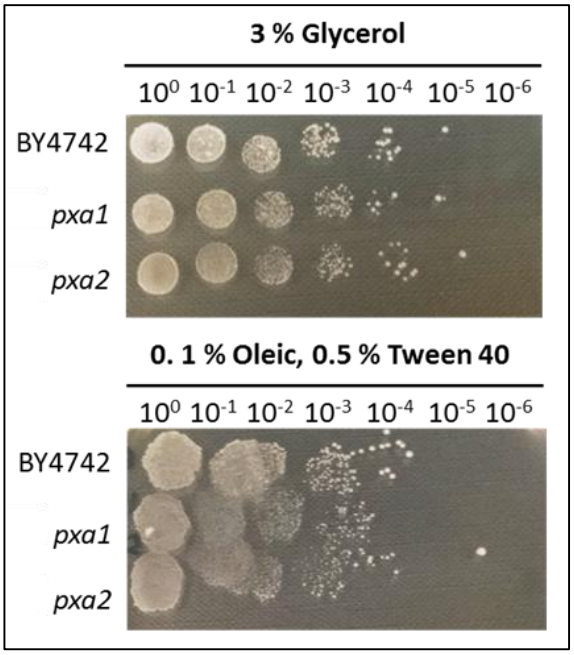
Appendix 2.13 Curcumin HPLC chromatogram and mass spectrum



Using curcumin glycoside LC-MS analysis method, to analyse curcumin (M_r 368.39 g mol^{-1}). HPLC chromatogram (425 nm), RT 14.4, UV max 424 nm. MS: *m/z* 366.9 $[M-H]^-$, *m/z* 734.1 $2[M-H]^-$.

Appendix 2.14 Growth of *pxa1* and *pxa2* yeast gene-disruption mutants

Growth of yeast strains on sugar containing (3 % glycerol) and 0.1 % oleic acid agar plates. Wildtype BY4742, *pxa1* and *pxa2* grew equally as well on both types of carbon source, therefore were concluded to not exhibit the expected phenotype.



Appendix 3.1 C. *longa* UGT BLAST hits JW777458.1 and JW773795.1

JW777458.1

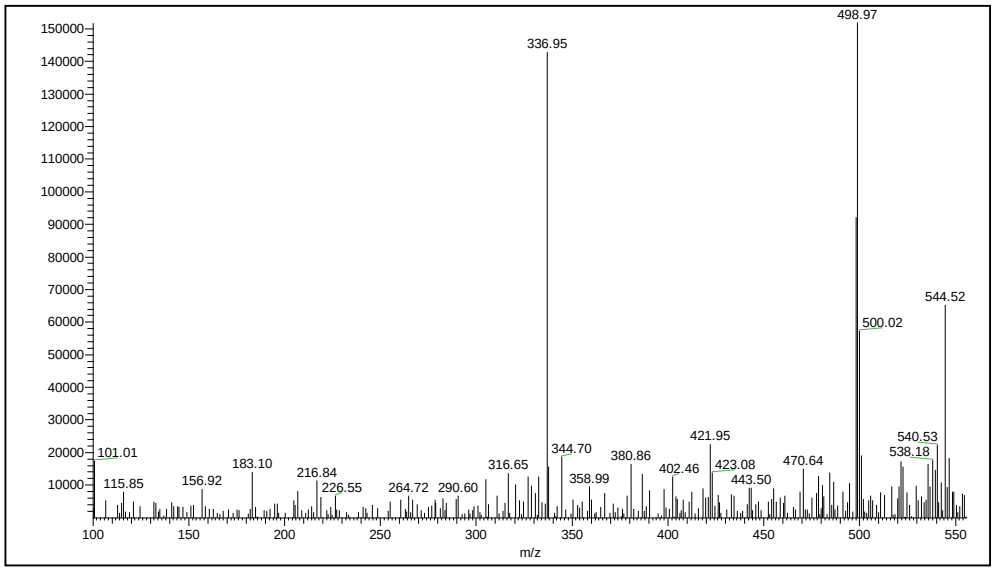
CTCTGATTGTACCCAGCTTTGGATTCCCTTTCAATCACTACTGTATAATTGAATTGTGTCGTCGTGGAAA
GCTGTGGAGAGGCTTTCTAGAAGCCCATTTTGCATTTTTTTTCGTGAGCTGGTGACAGTAATGAATCGAG
GGAAGGAAAAAGACGTCTTTAGCTTAACAGACGTCATTGCAAAGGAGATCGAAGAAGTAAGAAACGCAAA
CCTAGTCTTTGCTTACGCCGTCCTCAGTATTTAATGGATAGAAGCGCACACAGTTTGCAAGCCCATGCTC
AATACCATGGCGGATGATCTGCATATACTCTTGCTGCCCTTTCTCTCCCCAGGCCACATCATCCCCATGG
TCGACTTAGCACGGCTTTTTGCAGGTAAAGGAATCAAAGCCACCGTCGTCACCACCACCGGCAACGTGCC
TCTCGTCAAACCTACCGTTGATCTCGCCAACACCGACGCTTCCCTCCGCAACCCTATCCAACTCCTCGCC
CTCCGCCTCCCCTGCTCAGAAGCTGGAATCCCCGAGGGCTACGAGAACCTCGCCGCCTTCCCTAACCCCG
ACTTGTTACGGTTGGTGGCCGCCACCGACAGGCTTGAGCCCTCCTTCTCGCTGCTGCTCAAAACCCACCG
CCCTGACTGCGTCGTCGTCGATTTCTTCTATCCTTGGGCCACCCGCGTCGCCCAAGAAGCCTCCGTTCCT
TCGCTCCTCTTTAGTGGAAGCAACTTCTTCAGCTCCGCGGTTATTAAAATCGTACAAGACGCCGGGCTGA
TCCAGGGAGACTTGGAGACCGAGCGATTGATCGAGGTTCCCGGCATACCGCAGAAGATCCACCTCAAGCT
CTCTAATCTCCCCGACACCGTCCTTCATCCCAAGGAGTTCTCCCACCGGATGGCCGAAAGCCGTCACCGG
ATCCACGGCACGGTCGTGAACAGCTTCTACGAGCTGGAACGTGACTACGTCAATCGAATCCCCTACTCGC
GAGACCTCAGGTTCTGGTTCGTCGGCCCAGTCTCGCTGCAGAACCGAGGCATGGATACGAAGATGGTGCG
AGGGAGCGGCGACGCCGCTTCGGGGAGCGCCGATCATTTCTTGAGTTGGCTGGAAACCAAGAGGCCTAGG
TCGGTCCTCTACGTATGCTTCGGGAGCATGGCCCGGTTCACGGCCACTCAGCTACGCGAGATCGCGCGAG
GTCTCGAGGCGTCCGACCGGCCGTTCATCTGGGTCGTGCGGAACGCCGGAGAGATATCGGAATGGCTGCC
GGACGGGTTCGAGAACAGGGTGGTCGGCGGCGGGAAGGGGATGCTGGTGCAAGGATGGGCGCCGCAGCTT
CTCATACTGAATCACGAAGCGGTGGGAGGCTTCGTGACTCACTGCGGGTGGAACTCGTGCCTAGAAGCGA
TTGCTGCCGGAGTTCCGGTGATCACTTGGCCGTTGTTTGCCGAGCAATTCCTGAACGAGAAGCTGTTGGT
CGACGTGCACAGGATGGGGATCTCGATCGGCGTGATTTCCAGCTGCGATAAAGCAGAGGAGCGGACGGTG
GTGAACGGTGAGCAACTGAAGGAGGCAGTGGATGAGCTGATGGGTAGCGGAGAGGAAGCCGAAGAGAGGA

GGAAGAGGGCGAGGGAGCTGGGGCAGACGGCGAGAAGGGCAATGGAAGAAGGTGGATCTTCTCAGGAAGC
GATGACTTGCCTGATAGACGAGTTAATTCGATTGAAAACTGTTGATTAAGCAGGGAGAATGGAGACGAAG
TGAAATAACGACAGTCTTCTTTGCACATCAATTCGAATTCATGTTGCCTTAATTATTTTACTTTATCAAT
TAGAATCCGTGTTGCATTCATTACAATCACATGCATGTACTTTATCGAGAATGCATCCAAAGGCACACAA
ACTAAGAACGACCTGTCTGCTTCTTCAAAAAAA

JW773795.1

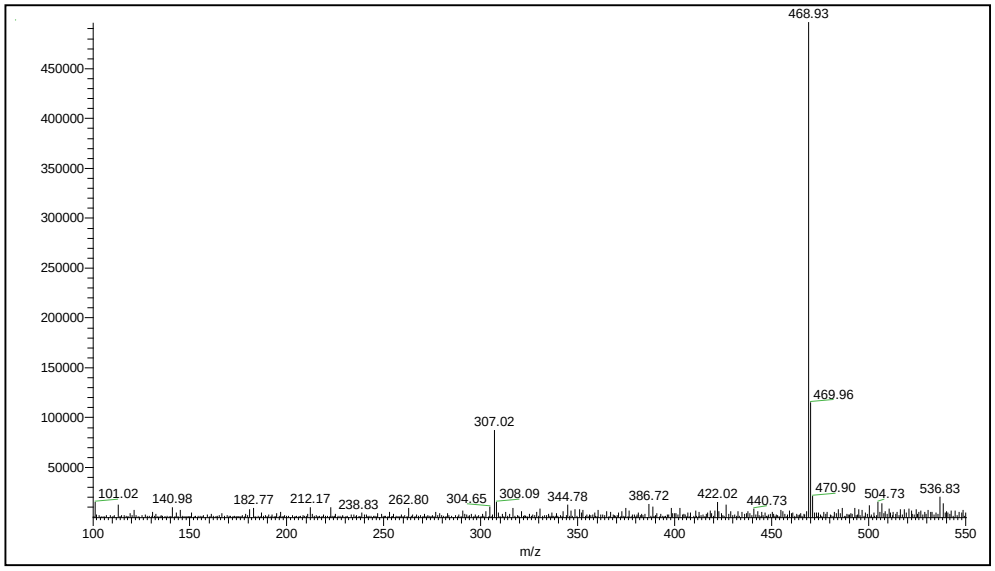
CCAGTAGGTGTGATTGTTTAGGCACGCACGCCGACGGCACTTCGCTAGTCTGCCTCGATCCCCACCGTCC
ATGGAGGAGGAACTTCACCGGGAACGACCGCCGGCGGCGCACTTCGCCTTGGTTCCCCTGTTGACGCAGG
GCCACATCATCCCCATGCTCGACCTCGCCCACCAACTCGCCCTCCGCGGCGTCCTCGTCAGCTTCATCAC
CACCCCCTTGAACGCCTCCCGCATCCGCCACACCATCGACCAAGCCGAAGCCCGCCGTCTCCCCATCCGA
TTCGTCGAGCTCCCCTTCCCAGTCCACGAGGCTGGCCTCCCTGCCGGCTGCGAGAACTTCGACACCCTCC
CCAAAACAGATATGTACATGACCTTCAACCAAGCCTGCTCGAAGCTCCTGGCTCCCCCGCTGGAGCTCTA
CCTGGAGCAGCACCGATCCGATCTCATCGTCTCCGACTTCTGCCAGCCGTGGACGCGCGGGATAGCTGAG
AGACTCGGAGTTCCGCGGCTCATCTTCTATAGCATGTGCTGCTTGGCGCTCCTCTGTACCCACAACATTT
GGGTACACAAGATGTTGGAGAGAGCAGGCGACGAGAACGAGCCGTTCGCGGTTCCGGGATTGCCGGAGGA
GCACGTAATCGAGGTGACGAAGGCCCAGGCGCCGGGGTTCTTCCCTGGCCCTGCTTTCGTGCAGATGGAT
AAAGACGTGCGCGAAGCGGAGTTCACCGCCGACGGCGTGATTTTCAACACTATCTACGATTTGGAACCTT
CCTACGTTTCCGGGTACGAAAAGGCCATGGGGAAGAAGATTTGGACGGTCGGTCCGTTGCTGTCGCACAG
CCAGAGCGTCGCCGACTTGGCCACGAGGGGGAGGAAGGCTTCCATAGACACGGAGCGATGCTTGGCTTGG
CTCGACGCGATGATGCCGCGGTCGGTGATCTACGTCAGCTTCGGGAGCCTCACACGCGTGGCGCCGTCGC
AGATGATGGAGATCGGGCTGGGACTAGAGGCGTCAGGGCATCCGTTCATTTGGGTGATCAGGAGCAAGGA
GGATTGCTCGCCGGAGGAGGAAGAATGGCTGACCGGAGGATTCGAGGCGAGAGTGAGCTCGAGGGCGCTG
TTGATCAGAGGGTGGGCGCCGCAAGCCGTCATCCTGACTCACCATGCCGTCGGAGGAGTGATGACGCACT
GCGGGTGGAACACGATCCTGGAGACGATCACGGTGGGGCTGCCGATGATCTCGTGGCCTCATTTTACGGA
CCAGTTCTTTAACGAGAGGCTGGTCGTCCAAGTGCTGAAGGTTGGAGTTGCGATTGGAGTGAAGGTGCCA
AATTACATTATTGGGATGGATGACATAGAGACGTTGGTGAAGAGGGAGCAGGTGGAGGAGCGGGTGAGGA
CTCTCATGGGTGGAGGACCGGAAGGTGAAGAAAGGAGGAAGAGGGCCGCGGAGTTAGCGGAGATAGCGAA
GGCCGCCATGGAAGAAGGGGGTTCGTCCCACTCCAATTTCACGCAACTGATCGAGCATTTCTCAGGGAGA
AGATGCAAATGAAAGAAAATCGCGTGCACTTCTCTGTGGCATTCAGGTAGGCTCTGTTCTCTGAAATCGG
ATGCACTTTTTTCCGCTTGTAATTTATTAATACTGCACTGTTCTTCAATATATATCAAAGCTCTAATGTT
TGAAATCTTGAGAGAGAGAGAGTTTCACTTCAG

Appendix 3.2 Demethoxycurcumin monoglucoside mass spectrum



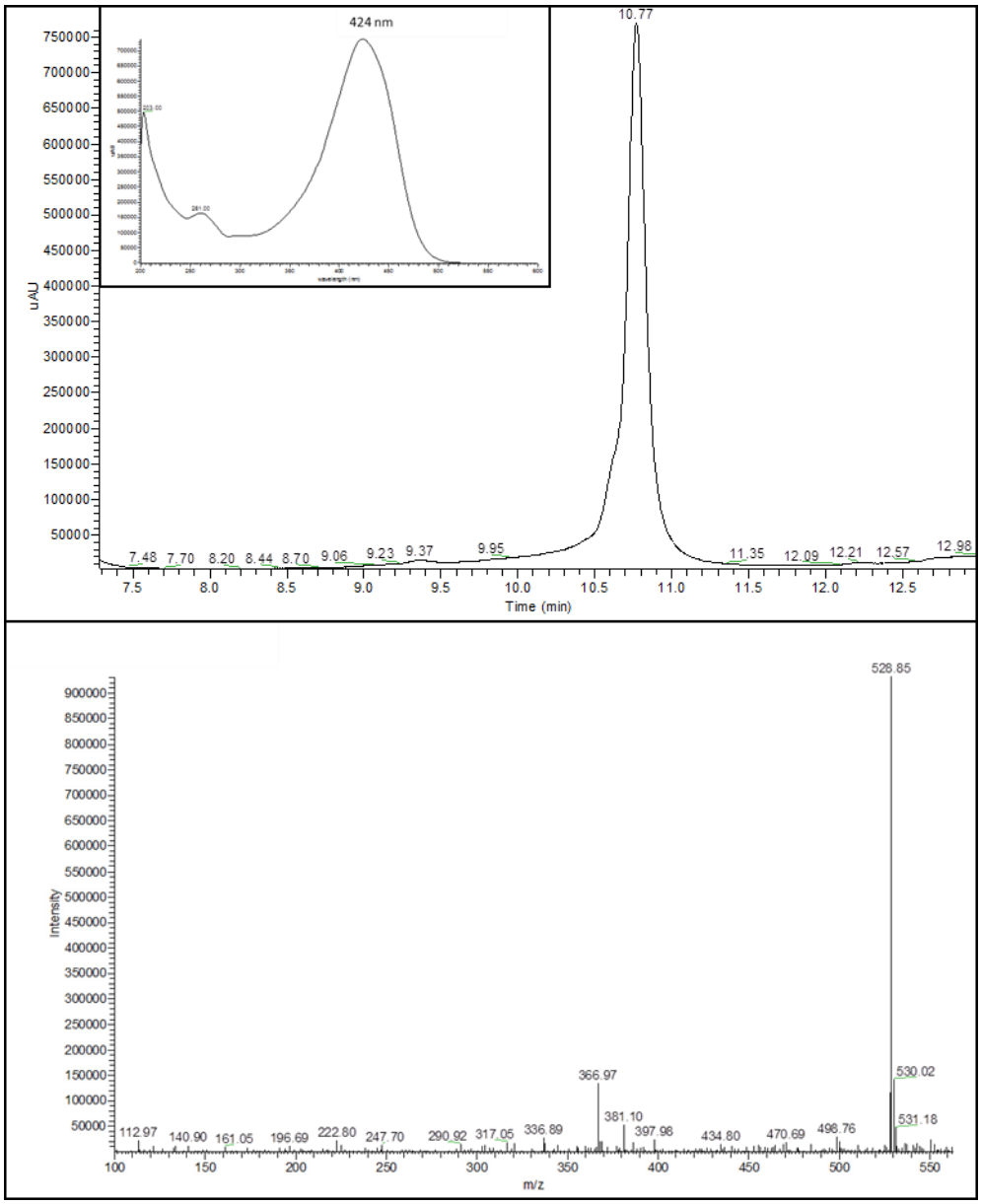
Using curcumin glycoside LC-MS analysis method, to analyse demethoxycurcumin monoglucoside (M_r 500.48 g mol^{-1}). MS: *m/z* 499.0 $[M-H]^-$, m/z 337.0 $[M-162]^-$ DMC.

Appendix 3.3 Bisdemethoxycurcumin monoglucoside mass spectrum



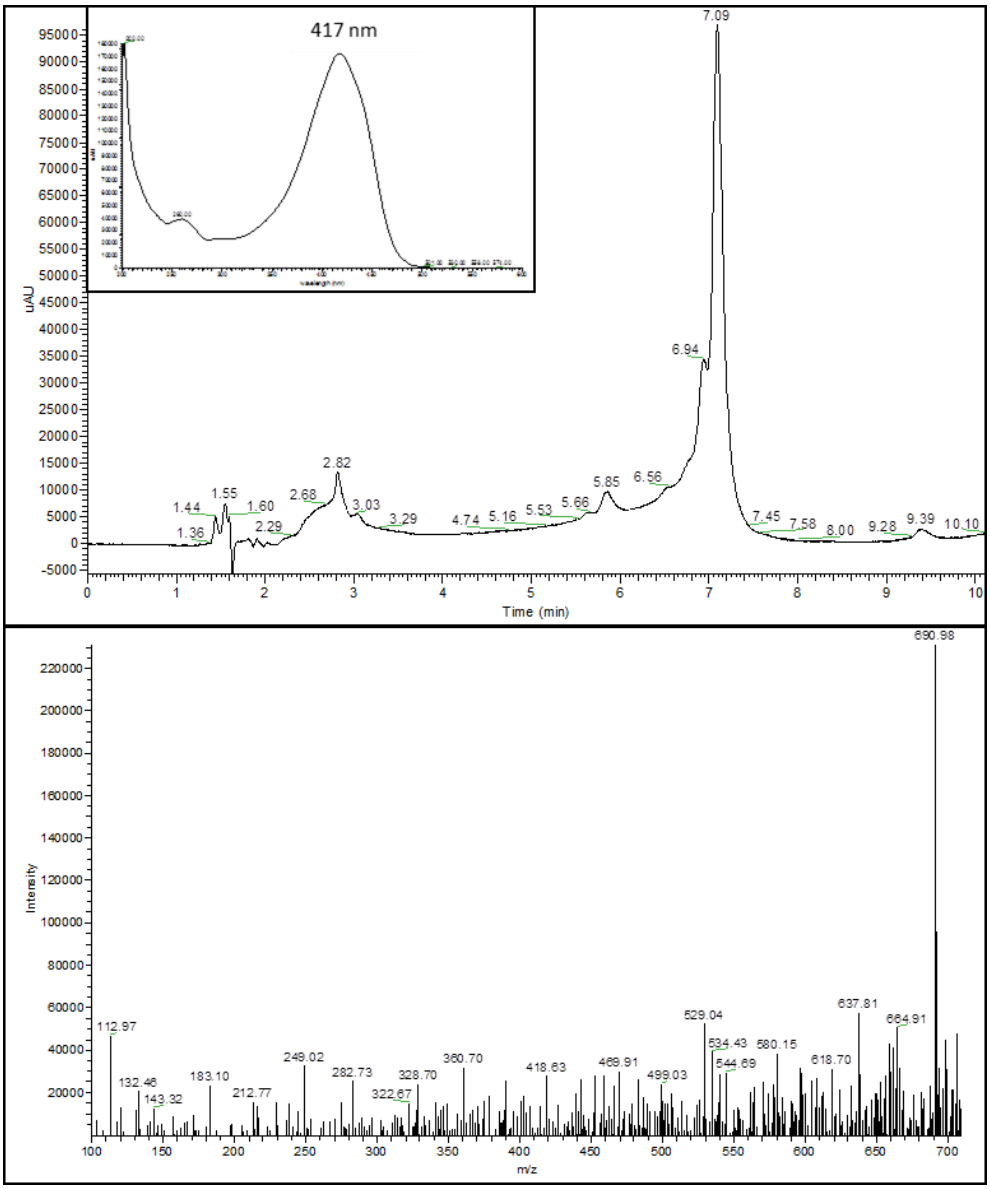
Using curcumin glycoside LC-MS analysis method, to analyse bisdemethoxycurcumin monoglucoside (M_r 470.46 g mol^{-1}). MS: *m/z* 468.9 $[M-H]^-$, m/z 307.0 $[M-162]^-$ BDMC.

Appendix 3.4 Curcumin monoglucoside HPLC chromatogram and mass spectrum



Using curcumin glycoside LC-MS analysis method, to analyse curcumin monoglucoside (M_r 530.52 g mol^{-1}). HPLC chromatogram (425 nm), RT 10.8, UV max 424 nm. MS: *m/z* 528.9 $[M-H]^-$, *m/z* 367.0 $[M-162]^-$ curcumin.

Appendix 3.5 Curcumin diglucoside HPLC chromatogram and mass spectrum



Using curcumin glycoside LC-MS analysis method, to analyse curcumin diglucoside (M_r 691.61 g mol^{-1}). HPLC chromatogram (425 nm), RT 7.1, UV max 417 nm. MS: *m/z* 691.0 $[M-H]^-$.

Abbreviations

2PS	2-Pyrone synthase
4CL	4-coumarate CoA ligase
4CN	4-chloro-1-naphthol
A	Absorbance
ACS	Acridone synthase
ADMET	Absorption, distribution, metabolism, excretion, toxicity
ADP	Adenosine diphosphate
Ala	Schizsaccharomyces pombe, alanine tRNA
AMP	Adenosine monophosphate
AP	Alkaline phpsphatase
APCI	Atmospheric pressure chemical ionization
ArREST	Aromatic Rhizome EST
At	*Arabidopsis thaliana*
ATP	Adenosine triphosphate
BAS	benzalacetone synthase
BCIP	5-bromo-4-chloro-3'-indoylphosphate
BDMC	Bisdemethoxycurcumin
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
C3H	C-3 hydroxylase
CaUGT2/3	Catharanthus roseus UGT2/3
CCoAOMT	Caffeoyl-CoA O-methyltransferase
CDS	Coding sequence
CHS	Chalcone synthase

Cl	*Curcuma longa*
CoA	Coenzyme A
COSTREL	Combinatorial supertransformation of transplastomic recipient lines
CURS	Curcuminoid synthase
CUS	Curcumin synthase
CYC1	Cytochrome c, isoform 1
DART	Direct Analysis in Real Time
DCS	Diketide CoA synthase
DMC	Demethoxycurcumin
ESI	Electrospray ionisation
EST	Expressed sequence tags
F2A	2A sequence from FMDV
FDC1	Ferulic acid decarboxylase 1
FMDV	Foot and mouth disease virus
HPLC	High performance liquid chromatography
HRP	Horse-radish peroxidase
kB	Kilo base pairs
kD	Kilo daltons
LC-MS	Liquid chromatography- mass spectrometry
Le	*Lithospermum erythrorhizon*
Leu1	*Schizsaccharomyces pombe*, leucine tRNA
Leu2	*Saccharomyces cerevisiae*, leucine tRNA
Mr	Molecular mass
NADPH	Nicotinamide adenine dinucleotide phosphate
NBT	Nitro-blue tetrazolium
NCBI	National Center for Biotechnology Information

NF	Nuclear factor
Os	*Oryza sativa*
PAD1	Phenylacrylic acid decarboxylase 1
PAL	Phenylalanine ammonium lyase
PCR	Polymerase chain reaction
PKS	Polyketide synthase
PSPG	Plant secondary product glycosyltransferases
Pv	*Prunella vulgaris*
PVPP	Polyvinyl polypyrrolidone
PXA1/2	Peroxisome ABC membrane transporter 1/2
pYES2-4DC	pYES2-4CL5-F2A-DCS-F2A-CURS1
qPCR	Quantitive PCR
RACE	Rapid amplification of cDNA ends
RNAi	RNA interference
RPM	Rotations per minute
RT	Retention time in minutes
RT-PCR	Reverse transcription PCR
STAT3	Signal transducer and activator of transcription 3
STS	Stilbene synthase
SUS1	Sucrose synthase 1
TAL	Tyrosine ammonium lyase
TCP	2,4,5-trichlorophenol
TEA	Triethylamine
TIC	Total ion current
UDP	Uridine diphosphate
UGT	UDP-glycosyltransferase
UPLC	Ultra Performance Liquid Chromatography

UV	Ultra violet

References

1. Kaminaga, Y., Sahin, F. P. and Mizukami, H. Molecular cloning and characterization of a glucosyltransferase catalyzing glucosylation of curcumin in cultured *Catharanthus roseus* cells. *FEBS Lett.* **567,** 197–202 (2004).

2. Lietava, J. Medicinal plants in a Middle Paleolithic grave Shanidar IV. *J. Ethnopharmacol.* **35,** 263–6 (1992).

3. Ekor, M. The growing use of herbal medicines: issues relating to adverse reactions and challenges in monitoring safety. *Front. Pharmacol.* **4,** 177 (2014).

4. Gurib-Fakim, A. Medicinal plants: traditions of yesterday and drugs of tomorrow. *Mol. Aspects Med.* **27,** 1–93 (2006).

5. Hanson, J. R. *Natural Products : The Secondary Metabolites*. (Royal Society of Chemistry, 2003). doi:10.1039/9781847551535.

6. Nicolaou, K. C., Yang, Z., Liu, J. J., Ueno, H., Nantermet, P. G., Guy, R. K., Claiborne, C. F., Renaud, J., Couladouros, E. A. and Paulvannan, K. Total synthesis of taxol. *Nature* **367,** 630–4 (1994).

7. Holton, R. A., Somoza, C., Kim, H. B., Liang, F., Biediger, R. J., Boatman, P. D., Shindo, M., Smith, C. C. and Kim, S. First total synthesis of taxol. 1. Functionalization of the B ring. *J. Am. Chem. Soc.* **116,** 1597–1598 (1994).

8. Baker, D. D., Chu, M., Oza, U. and Rajgarhia, V. The value of natural products to future pharmaceutical discovery. *Nat. Prod. Rep.* **24,** 1225–44 (2007).

9. Bailey, J. E. Toward a science of metabolic engineering. *Science (80-.).* **252,** 1668–1675 (1991).

10. Research Councils UK. *A Synthetic Biology Roadmap for the UK*. (2012).

11. Mora-Pale, M., Sanchez-Rodriguez, S. P., Linhardt, R. J., Dordick, J. S. and Koffas, M. A. G. Biochemical strategies for enhancing the in vivo production of natural products with pharmaceutical potential. *Curr. Opin. Biotechnol.* **25,** 86–94 (2014).

12. Porro, D., Branduardi, P., Sauer, M. and Mattanovich, D. Old obstacles and new horizons for microbial chemical production. *Curr. Opin. Biotechnol.* **30,** 101–106 (2014).

13. Schrewe, M., Julsing, M. K., Bühler, B. and Schmid, A. Whole-cell biocatalysis for selective and productive C-O functional group introduction and modification. *Chem. Soc. Rev.* **42,** 6346–77 (2013).

14. Kiener, A. Enzymatic Oxidation of Methyl Groups on Aromatic Heterocycles: A Versatile Method for the Preparation of Heteroaromatic Carboxylic Acids. *Angew. Chemie Int. Ed. English* **31,** 774–775 (1992).

15. Picataggio, S., Rohrer, T., Deanda, K., Lanning, D., Reynolds, R., Mielenz, J. and Eirich, L. D. Metabolic engineering of Candida tropicalis for the production of long-chain dicarboxylic acids. *Biotechnology. (N. Y).* **10,** 894–8 (1992).

16. Lopez-Gallego, F. and Schmidt-Dannert, C. Multi-enzymatic synthesis. *Curr. Opin. Chem. Biol.* **14,** 174–183 (2010).

17. Mora-Pale, M., Sanchez-Rodriguez, S. P., Linhardt, R. J., Dordick, J. S. and Koffas, M. A. G. Metabolic engineering and in vitro biosynthesis of phytochemicals and non-natural analogues. *Plant Sci.* **210,** 10–24 (2013).

18. Shao, Z. and Zhao, H. DNA assembler: a synthetic biology tool for characterizing and engineering natural product gene clusters. *Methods Enzymol.* **517,** 203–24 (2012).

19. Shao, Z., Zhao, H. and Zhao, H. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* **37,** e16 (2009).

20. Fowler, Z. L., Gikandi, W. W. and Koffas, M. A. G. Increased Malonyl Coenzyme A biosynthesis by tuning the *Escherichia coli* metabolic network and its application to flavanone production. *Appl. Environ. Microbiol.* **75,** 5831–5839 (2009).

21. Nevoigt, E. Progress in metabolic engineering of *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **72,** 379–412 (2008).

22. Valadi, H., Larsson, C. and Gustafsson, L. Improved ethanol production by glycerol-3-phosphate dehydrogenase mutants of *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* **50,** 434–439 (1998).

23. Hansen, E. H., Moller, B. L., Kock, G. R., Buenner, C. M., Kristensen, C., Jensen, O. R., Okkels, F. T., Olsen, C. E., Motawia, M. S. and Hansen, J. De Novo Biosynthesis of Vanillin in Fission Yeast (*Schizosaccharomyces pombe*) and Baker's Yeast (*Saccharomyces cerevisiae*). *Appl. Environ. Microbiol.* **75,** 2765–2774 (2009).

24. Becker, J., Armstrong, G., Vandermerwe, M., Lambrechts, M., Vivier, M. and Pretorius,

I. Metabolic engineering of for the synthesis of the wine-related antioxidant resveratrol. *FEMS Yeast Res.* **4,** 79–85 (2003).

25. Hong, K. K. and Nielsen, J. Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. *Cell. Mol. Life Sci.* **69,** 2671–2690 (2012).

26. Enserink, M. Infectious diseases. Source of new hope against malaria is in short supply. *Science* **307,** 33 (2005).

27. Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M. D., Tai, A., Main, A., Eng, D., Polichuk, D. R., Teoh, K. H., Reed, D. W., Treynor, T., Lenihan, J., Jiang, H., Fleck, M., Bajad, S., Dang, G., *et al.* High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496,** 528–532 (2013).

28. Paddon, C. J. and Keasling, J. D. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat. Rev. Microbiol.* **12,** 355–67 (2014).

29. Westfall, P. J., Pitera, D. J., Lenihan, J. R., Eng, D., Woolard, F. X., Regentin, R., Horning, T., Tsuruta, H., Melis, D. J., Owens, A., Fickes, S., Diola, D., Benjamin, K. R., Keasling, J. D., Leavell, M. D., McPhee, D. J., Renninger, N. S., Newman, J. D. and Paddon, C. J. Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proc. Natl. Acad. Sci.* **109,** E111–E118 (2012).

30. Ro, D.-K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C. Y., Withers, S. T., Shiba, Y., Sarpong, R. and Keasling, J. D. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440,** 940–943 (2006).

31. Kopetzki, D., Lévesque, F. and Seeberger, P. H. A Continuous-Flow Process for the Synthesis of Artemisinin. *Chem. - A Eur. J.* **19,** 5450–5456 (2013).

32. Peplow, M. Synthetic malaria drug meets market resistance. *Nature, News Focus* **530,** 390 (2016).

33. Thodey, K., Galanie, S. and Smolke, C. D. A microbial biomanufacturing platform for natural and semisynthetic opioids. *Nat. Chem. Biol.* **10,** 837–844 (2014).

34. Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M., Smolke, C. D., Seya, M. J., Gelders, S. F., Achara, O. U., Milani, B., Scholten, W. K., Reed, J. W., Hudlicky, T., Siddiqui, M. S., Thodey, K., Trenchard, I., Smolke, C. D., Ro, D. K., Paradise, E. M.,

Ouellet, M., *et al.* Complete biosynthesis of opioids in yeast. *Science* **349,** 1095–100 (2015).

35. *World Drug Report Opium/Heroin market*. (2009).

36. Krivoruchko, A. and Nielsen, J. Production of natural products through metabolic engineering of *Saccharomyces cerevisiae*. *Curr. Opin. Biotechnol.* **35C,** 7–15 (2014).

37. Chen, Y., Daviet, L., Schalk, M., Siewers, V. and Nielsen, J. Establishing a Platform Cell Factory through Engineering of Yeast Acetyl-CoA Metabolism. *Metab. Eng.* **15,** 48 (2012).

38. Shi, S., Chen, Y., Siewers, V. and Nielsen, J. Improving production of malonyl coenzyme A-derived metabolites by abolishing Snf1-dependent regulation of Acc1. *MBio* **5,** e01130-14 (2014).

39. Kharel, M. K., Lian, H. and Rohr, J. Characterization of the TDP-d-ravidosamine biosynthetic pathway: one-pot enzymatic synthesis of TDP-d-ravidosamine from thymidine-5-phosphate and glucose-1-phosphate. *Org. Biomol. Chem.* **9,** 1799 (2011).

40. Denard, C. A., Huang, H., Bartlett, M. J., Lu, L., Tan, Y., Zhao, H. and Hartwig, J. F. Cooperative Tandem Catalysis by an Organometallic Complex and a Metalloenzyme. *Angew. Chemie Int. Ed.* **53,** 465–469 (2014).

41. Heinig, U., Gutensohn, M., Dudareva, N. and Aharoni, A. The challenges of cellular compartmentalization in plant metabolic engineering. *Curr. Opin. Biotechnol.* **24,** 239–246 (2013).

42. Laux, T., Brand, U., Fletcher, J. C., Hobe, M., Meyerowitz, E. M., Simon, R., Brand, U., Grunewald, M., Hobe, M., Simon, R., Clark, S. E., Doetsch, F., Petreanu, L., Caille, I., Garcia-Verdugo, J., Alvarez-Buylla, A., Fletcher, J. C., Brand, U., Running, M. P., *et al.* The stem cell concept in plants: a matter of debate. *Cell* **113,** 281–283 (2003).

43. Schlatmann, J. E., Nuutila, A. M., Van Gulik, W. M., Ten Hoopen, H. J., Verpoorte, R. and J Heijnen, J. Scaleup of ajmalicine production by plant cell cultures of *Catharanthus roseus*. *Biotechnol. Bioeng.* **41,** 253–262 (1993).

44. Whitaker, R. J., Hashimoto, T. and Evans, D. A. Production of the Secondary Metabolite, Rosmarinic Acid, by Plant Cell Suspension Cultures. *Ann. N. Y. Acad. Sci.* **435,** 364–366 (1984).

45. Yukimune, Y., Tabata, H., Higashi, Y. and Hara, Y. Methyl jasmonate-induced

overproduction of paclitaxel and baccatin III in Taxus cell suspension cultures. *Nat. Biotechnol.* **14,** 1129–1132 (1996).

46. Exposito, O., Bonfill, M., Moyano, E., Onrubia, M., Mirjalili, M., Cusido, R. and Palazon, J. Biotechnological Production of Taxol and Related Taxoids: Current State and Prospects. *Anticancer. Agents Med. Chem.* **9,** 109–121 (2009).

47. Ma, J. K.-C., Drake, P. M. W. and Christou, P. Genetic modification: The production of recombinant pharmaceutical proteins in plants. *Nat. Rev. Genet.* **4,** 794–805 (2003).

48. Bourgaud, F., Gravot, A., Milesi, S. and Gontier, E. Production of plant secondary metabolites: a historical perspective. *Plant Sci.* **161,** 839–851 (2001).

49. Lee, E.-K., Jin, Y.-W., Park, J. H., Yoo, Y. M., Hong, S. M., Amir, R., Yan, Z., Kwon, E., Elfick, A., Tomlinson, S., Halbritter, F., Waibel, T., Yun, B.-W. and Loake, G. J. Cultured cambial meristematic cells as a source of plant natural products. *Nat. Biotechnol.* **28,** 1213–1217 (2010).

50. Ochoa-Villarreal, M., Howat, S., Jang, M. O., Kim, I. S., Jin, Y.-W., Lee, E.-K. and Loake, G. J. Cambial meristematic cells: a platform for the production of plant natural products. *N. Biotechnol.* **32,** 581–587 (2015).

51. Ziemienowicz, A. Agrobacterium-mediated plant transformation: Factors, applications and recent advances. *Biocatal. Agric. Biotechnol.* **3,** 95–102 (2013).

52. Ludwig-Müller, J., Jahn, L., Lippert, A., Püschel, J. and Walter, A. Improvement of hairy root cultures and plants by changing biosynthetic pathways leading to pharmaceutical metabolites: Strategies and applications. *Biotechnol. Adv.* **32,** 1168–1179 (2014).

53. Bhadra, R., Morgan, J. A. and Shanks, J. V. Transient studies of light-adapted cultures of hairy roots of *Catharanthus roseus*: Growth and indole alkaloid accumulation. *Biotechnol. Bioeng.* **60,** 670–678 (1998).

54. Maldonado-Mendoza, I. E., Ayora-Talavera, T. and Loyola-Vargas, V. M. Establishment of hairy root cultures of *Datura stramonium*. Characterization and stability of tropane alkaloid production during long periods of subculturing. *Plant Cell. Tissue Organ Cult.* **33,** 321–329 (1993).

55. Taya, M., Yoyama, A., Kondo, O., Kobayashi, T. and Matsui, C. Growth characteristics of plant hairy roots and their cultures in bioreactors. *J. Chem. Eng. JAPAN* **22,** 84–89 (1989).

56. Graham, I. A., Besser, K., Blumer, S., Branigan, C. A., Czechowski, T., Elias, L., Guterman, I., Harvey, D., Isaac, P. G., Khan, A. M., Larson, T. R., Li, Y., Pawson, T., Penfield, T., Rae, A. M., Rathbone, D. A., Reid, S., Ross, J., Smallwood, M. F., *et al.* The genetic map of *Artemisia annua* L. identifies loci affecting yield of the antimalarial drug artemisinin. *Science* **327,** 328–31 (2010).

57. ArtemisiaF1seed. *Centre for Novel Agricultural Products, East-West Seed International* (2001). Available at: http://www.artemisiaf1seed.org/. (Accessed: 10th August 2016)

58. Ye, X., Al-Babili, S., Klöti, A., Zhang, J., Lucca, P., Beyer, P., Potrykus, I., Sommer, A., West, K. P., Howard, G. R., Sommer, A., Humphrey, J. H., West, K. P., Sommer, A., Pirie, A., Burkhardt, P., Schledz, M., Bonk, M., Misawa, N., *et al.* Engineering the Provitamin A (β-Carotene) Biosynthetic Pathway into (Carotenoid-Free) Rice Endosperm. *Science (80-.).* **287,** 96–305 (2000).

59. Paine, J. A., Shipton, C. A., Chaggar, S., Howells, R. M., Kennedy, M. J., Vernon, G., Wright, S. Y., Hinchliffe, E., Adams, J. L., Silverstone, A. L. and Drake, R. Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nat. Biotechnol.* **23,** 482–487 (2005).

60. Butelli, E., Titta, L., Giorgio, M., Mock, H.-P., Matros, A., Peterek, S., Schijlen, E. G. W. M., Hall, R. D., Bovy, A. G., Luo, J. and Martin, C. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat. Biotechnol.* **26,** 1301–1308 (2008).

61. Fuentes, P., Zhou, F., Erban, A., Karcher, D., Kopka, J. and Bock, R. A new synthetic biology approach allows transfer of an entire metabolic pathway from a medicinal plant to a biomass crop. *Elife* **5,** (2016).

62. Záveská, E., Fér, T., Šída, O. and Krak, K. Phylogeny of Curcuma (Zingiberaceae) based on plastid and nuclear sequences: Proposal of the new subgenus Ecomata. *Taxon* **61,** 747–763 (2012).

63. Aggarwal, B.B., Surh, Y.J., Shishodia, S. *In The Molecular Target and Therapeutic Uses of Curcumin in Health and Disease*. (Advances in experimental medicine and biology, Springer, 2007).

64. Lantz, R. C., Chen, G. J., Solyom, A. M., Jolad, S. D. and Timmermann, B. N. The effect of turmeric extracts on inflammatory mediator production. *Phytomedicine* **12,** 445–452 (2005).

65. Shishodia, S., Sethi, G. and Aggarwal, B. B. Curcumin: getting back to the roots. *Ann. N. Y. Acad. Sci.* **1056,** 206–217 (2005).

66. Agrawal, D. K. and Mishra, P. K. Curcumin and Its Analogues: Potential Anticancer Agents. *Med. Res. Rev.* **30,** 818–860 (2010).

67. Basnet, P. and Skalko-Basnet, N. Curcumin: An Anti-Inflammatory Molecule from a Curry Spice on the Path to Cancer Treatment. *Molecules* **16,** 4567–4598 (2011).

68. Schraufstatter, E. and Bernt, H. Antibacterial Action of Curcumin and Related Compounds. *Nature* **164,** 456–457 (1949).

69. Sharma, O. P. Antioxidant activity of curcumin and related compounds. *Biochem. Pharmacol.* **25,** 1811–1812 (1976).

70. Ruby, A. J., Kuttan, G., Dinesh Babu, K., Rajasekharan, K. N. and Kuttan, R. Anti-tumour and antioxidant activity of natural curcuminoids. *Cancer Lett.* **94,** 79–83 (1995).

71. Joe, B. and Lokesh, B. R. Role of capsaicin, curcumin and dietary n - 3 fatty acids in lowering the generation of reactive oxygen species in rat peritoneal macrophages. *Biochim. Biophys. Acta - Mol. Cell Res.* **1224,** 255–263 (1994).

72. Lim, G. P., Chu, T., Yang, F., Beech, W., Frautschy, S. A. and Cole, G. M. The curry spice curcumin reduces oxidative damage and amyloid pathology in an Alzheimer transgenic mouse. *J. Neurosci.* **21,** 8370–8377 (2001).

73. Sahebkar, A. Are Curcuminoids Effective C-Reactive Protein-Lowering Agents in Clinical Practice? Evidence from a Meta-Analysis. *Phyther. Res.* **28,** 633–642 (2014).

74. Anthwal, A., Thakur, B. K., Rawat, M. S. M., Rawat, D. S., Tyagi, A. K. and Aggarwal, B. B. Synthesis, characterization and in vitro anticancer activity of C-5 curcumin analogues with potential to inhibit TNF-α-induced NF-κB activation. *Biomed Res. Int.* **53,** 8260–8273 (2014).

75. Chiablaem, K., Lirdprapamongkol, K., Keeratichamroen, S., Surarit, R. and Svasti, J. Curcumin suppresses vasculogenic mimicry capacity of hepatocellular carcinoma cells through STAT3 and PI3K/AKT inhibition. *Anticancer Res.* **34,** 1857–1864 (2014).

76. Prasad, S., Gupta, S. C., Tyagi, A. K. and Aggarwal, B. B. Curcumin, a component of golden spice: From bedside to bench and back. *Biotechnol. Adv.* **32,** 1053–1064 (2014).

77. Agrawal, D. K. and Mishra, P. K. Curcumin and Its Analogues: Potential Anticancer

Agents. *Med. Res. Rev.* **30,** (2010).

78. Anand, P., Sung, B., Kunnumakkara, A. B., Rajasekharan, K. N. and Aggarwal, B. B. Suppression of pro-inflammatory and proliferative pathways by diferuloylmethane (curcumin) and its analogues dibenzoylmethane, dibenzoylpropane, and dibenzylideneacetone: Role of Michael acceptors and Michael donors. *Biochem. Pharmacol.* **82,** 1901–1909 (2011).

79. Sandur, S. K., Pandey, M. K., Sung, B., Ahn, K. S., Murakami, A., Sethi, G., Limtrakul, P., Badmaev, V. and Aggarwal, B. B. Curcumin, demethoxycurcumin, bisdemethoxycurcumin, tetrahydrocurcumin and turmerones differentially regulate anti-inflammatory and anti-proliferative responses through a ROS-independent mechanism. *Carcinogenesis* **28,** 1765–1773 (2007).

80. Gafner, S., Lee, S.-K., Cuendet, M., Barthélémy, S., Vergnes, L., Labidalle, S., Mehta, R. G., Boone, C. W. and Pezzuto, J. M. Biologic evaluation of curcumin and structural derivatives in cancer chemoprevention model systems. *Phytochemistry* **65,** 2849–2859 (2004).

81. Nurfina, A. N., Reksohadiprodjo, M. S., Timmerman, H., Jenie, U. A., Sugiyanto, D. and van der Goot, H. Synthesis of some symmetrical curcumin derivatives and their anti-inflammatory activity. *Eur. J. Med. Chem.* **32,** 321–328 (1997).

82. Yu, D., Xu, F., Zeng, J. and Zhan, J. Type III polyketide synthases in natural product biosynthesis. *IUBMB Life* **64,** 285–295 (2012).

83. Abe, I. and Morita, H. Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. *Nat. Prod. Rep.* **27,** 809–838 (2010).

84. Junghanns, K. T., Kneusel, R. E., Baumert, A., Maier, W., Grӧger, D. and Matern, U. Molecular cloning and heterologous expression of acridone synthase from elicited Ruta graveolens L. cell suspension cultures. *Plant Mol. Biol.* **27,** 681–692 (1995).

85. Tropf, S., Karcher, B., Schroder, G. and Schroder, J. Reaction Mechanisms of Homodimeric Plant Polyketide Synthases (Stilbene and Chalcone Synthase). *J. Biol. Chem.* **270,** 7922–7928 (1995).

86. Eckermann, S., Schröder, G., Schmidt, J., Strack, D., Edrada, R. A., Helariutta, Y., Elomaa, P., Kotilainen, M., Kilpeläinen, I., Proksch, P., Teeri, T. H. and Schröder, J. New pathway to polyketides in plants. *Nature* **396,** 387–390 (1998).

87. Schroder, J. A family of plant-specific polyketide synthases: Facts and predictions. *Trends Plant Sci.* **2,** 373–378 (1997).

88. Roughley, P. J. and Whiting, D. A. Experiments in Biosynthesis of Curcumin. *J. Chem. Soc. Trans. 1* 2379–2388 (1973).

89. Kita, T., Imai, S., Sawada, H., Kumagai, H. and Seto, H. The biosynthetic pathway of curcuminoid in turmeric (*Curcuma longa*) as revealed by (13)C-labeled precursors. *Biosci. Biotechnol. Biochem.* **72,** 1789–1798 (2008).

90. Katsuyama, Y., Matsuzawa, M., Funa, N. and Horinouchi, S. *In vitro* synthesis of curcuminoids by type III polyketide synthase from *Oryza sativa*. *J. Biol. Chem.* **282,** 37702–37709 (2007).

91. Morita, H., Wanibuchi, K., Nii, H., Kato, R., Sugio, S. and Abe, I. Structural basis for the one-pot formation of the diarylheptanoid scaffold by curcuminoid synthase from *Oryza sativa*. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 19778–19783 (2010).

92. Miyazono, K., Um, J., Imai, F. L., Katsuyama, Y., Ohnishi, Y., Horinouchi, S. and Tanokura, M. Crystal structure of curcuminoid synthase CUS from *Oryza sativa*. *Proteins Struct. Funct. Bioinforma.* **79,** 669–673 (2011).

93. Katsuyama, Y., Kita, T., Funa, N. and Horinouchi, S. Curcuminoid Biosynthesis by Two Type III Polyketide Synthases in the Herb *Curcuma longa*. *J. Biol. Chem.* **284,** 11160–11170 (2009).

94. Kita, T., Komatsu, K., Zhu, S., Iida, O., Sugimura, K., Kawahara, N., Taguchi, H., Masamura, N. and Cai, S.-Q. Development of intron length polymorphism markers in genes encoding diketide-CoA synthase and curcumin synthase for discriminating *Curcuma* species. *Food Chem.* **194,** 1329–1336 (2016).

95. Katsuyama, Y., Kita, T. and Horinouchi, S. Identification and characterization of multiple curcumin synthases from the herb *Curcuma longa*. *Febs Lett.* **583,** 2799–2803 (2009).

96. Donghan, L., Ono, N., Sato, T., Sugiura, T., Altaf-Ul-Amin, M., Ohta, D., Suzuki, H., Arita, M., Tanaka, K., Ma, Z. and Kanaya, S. Targeted Integration of RNA-Seq and Metabolite Data to Elucidate Curcuminoid Biosynthesis in Four *Curcuma* Species. *Plant Cell Physiol.* **56,** 843–851 (2015).

97. Koo, H. J. PAVE: The ArREST Database. (2013). Available at: http://www.agcol.arizona.edu/cgi-bin/pave/GT/index.cgi. (Accessed: 30th June 2013)

98. Resmi, M. S. and Soniya, E. V. Molecular cloning and differential expressions of two cDNA encoding Type III polyketide synthase in different tissues of *Curcuma longa* L. *Gene* **491,** 278–283 (2012).

99. Jaiswal, Y., Liang, Z., Ho, A., Chen, H. and Zhao, Z. Tissue-specific metabolite profiling of Turmeric by using laser micro-dissection, ultra-high performance liquid chromatography-quadrupole time of fight-mass spectrometry and liquid chromatography-tandem mass spectrometry. *Eur. J. Mass Spectrom. (Chichester, Eng).* **20,** 383–393 (2014).

100. Rahman, A. F. M. M., Angawi, R. F. and Kadi, A. A. Spatial localisation of curcumin and rapid screening of the chemical compositions of turmeric rhizomes (*Curcuma longa* Linn.) using Direct Analysis in Real Time-Mass Spectrometry (DART-MS). *Food Chem.* **173,** 489–494 (2015).

101. Rafael, Z. and Michael, Y. Studies of the Cellular Localization of the Phenolic Pungent Principle of Ginger, *Zingiber officinale* Roscoe on JSTOR. *The New Phytologist* 295–300 (1994).

102. Lampe, V. Synthesis of curcumin. *Berichte Der Dtsch. Chem. Gesellschaft* **51,** 1347–1355 (1918).

103. Sabitha, G., Srinivas, C., Reddy, T. R., Yadagiri, K. and Yadav, J. S. Synthesis of gingerol and diarylheptanoids. *Tetrahedron-Asymmetry* **22,** 2124–2133 (2011).

104. Pabon, H. J. J. Synthesis of curcumin and related compounds. *Recl. Des Trav. Chim. Des Pays-Bas-Journal R. Netherlands Chem. Soc.* **83,** (1964).

105. Katsuyama, Y., Hirose, Y., Funa, N., Ohnishi, Y. and Horinouchi, S. Precursor-Directed Biosynthesis of Curcumin Analogs in *Escherichia coli*. *Biosci. Biotechnol. Biochem.* **74,** 641–645 (2010).

106. Wang, S., Zhang, S., Zhou, T., Zeng, J. and Zhan, J. Design and application of an *in vivo* reporter assay for phenylalanine ammonia-lyase. *Appl. Microbiol. Biotechnol.* **97,** 7877–7885 (2013).

107. Rodrigues, J. and Araújo, R. Production of curcuminoids from tyrosine by a metabolically engineered *Escherichia coli* using caffeic acid as an intermediate. *Biotechnol. …* **10,** 1–11 (2015).

108. Tonnesen, H. H. and Karlsen, J. Studies on curcumin and curcuminoids. 5. Alkaline

degradation of curcumin. *Zeitschrift Fur Leb. Und-forsch.* **180,** 132–134 (1985).

109. Mohri, K., Watanabe, Y., Yoshida, Y., Satoh, M., Isobe, K., Sugimoto, N. and Tsuda, Y. Synthesis of glycosylcurcuminoids. *Chem. Pharm. Bull. (Tokyo).* **51,** 1268–1272 (2003).

110. Wang, Y. J., Pan, M. H., Cheng, A. L., Lin, L. I., Ho, Y. S., Hsieh, C. Y. and Lin, J. K. Stability of curcumin in buffer solutions and characterization of its degradation products. *J Pharm Biomed Anal* **15,** 1867–1876 (1997).

111. Benassi, R., Ferrari, E., Grandi, R., Lazzari, S. and Saladini, M. Synthesis and characterization of new beta-diketo derivatives with iron chelating ability. *J. Inorg. Biochem.* **101,** 203–213 (2007).

112. Shoba, G., Joy, D., Joseph, T., Majeed, M., Rajendran, R. and Srinivas, P. S. Influence of piperine on the pharmacokinetics of curcumin in animals and human volunteers. *Planta Med.* **64,** 353–356 (1998).

113. Wahlström, B. and Blennow, G. A study on the fate of curcumin in the rat. *Acta Pharmacol. Toxicol. (Copenh).* **43,** 86–92 (1978).

114. Ravindranath, V. and Chandrasekhara, N. Absorption and tissue distribution of curcumin in rats. *Toxicology* **16,** 259–265 (1980).

115. Anand, P., Kunnumakkara, A. B., Newman, R. A. and Aggarwal, B. B. Bioavailability of curcumin: Problems and promises. *Mol. Pharm.* **4,** 807–818 (2007).

116. Naksuriya, O., van Steenbergen, M. J., Torano, J. S., Okonogi, S. and Hennink, W. E. A Kinetic Degradation Study of Curcumin in Its Free Form and Loaded in Polymeric Micelles. *AAPS J.* **18,** 777–787 (2016).

117. Li, C., Zhang, Y., Su, T., Feng, L., Long, Y. and Chen, Z. Silica-coated flexible liposomes as a nanohybrid delivery system for enhanced oral bioavailability of curcumin. *Int. J. Nanomedicine* **7,** 5995–6002 (2012).

118. Tiyaboonchai, W., Tungpradit, W. and Plianbangchang, P. Formulation and characterization of curcuminoids loaded solid lipid nanoparticles. *Int. J. Pharm.* **337,** 299–306 (2007).

119. Kren, V. and Thiem, J. Glycosylation employing bio-systems: from enzymes to whole cells. *Chem. Soc. Rev.* **26,** 463–473 (1997).

120. Lim, E. K. Plant glycosyltransferases - Their potential as novel biocatalysts. *Chem. Eur.*

J. **11,** 5486–5494 (2005).

121. Williams, S. and Withers, S. Glycosynthases: mutant glycosidases for glycoside synthesis. *Aust. J. Chem.* **55,** 3–12 (2002).

122. Vetter, J. Plant cyanogenic glycosides. *Toxicon* **38,** 11–36 (2000).

123. Bowles, D., Isayenkova, J., Lim, E.-K. and Poppenberger, B. Glycosyltransferases: managers of small molecules. *Curr. Opin. Plant Biol.* **8,** 254–263 (2005).

124. Bowles, D., Lim, E.-K., Poppenberger, B. and Vaistij, F. E. Glycosyltransferases of lipophilic small molecules. *Annu. Rev. Plant Biol.* **57,** 567–597 (2006).

125. Lewinsohn, E., Britsch, L., Mazur, Y. and Gressel, J. Flavanone Glycoside Biosynthesis in Citrus: Chalcone Synthase, UDP-Glucose:Flavanone-7-O-Glucosyl-Transferase and -Rhamnosyl-Transferase Activities in Cell-Free Extracts. *Plant Physiol.* **91,** 1323–1328 (1989).

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

127. Masada, S., Terasaka, K. and Mizukami, H. A single amino acid in the PSPG-box plays an important role in the catalytic function of CaUGT2 (Curcumin glucosyltransferase), a Group D Family 1 glucosyltransferase from *Catharanthus roseus*. *FEBS Lett.* **581,** 2605–2610 (2007).

128. Lim, E. K., Baldauf, S., Li, Y., Elias, L., Worrall, D., Spencer, S. P., Jackson, R. G., Taguchi, G., Ross, J. and Bowles, D. J. Evolution of substrate recognition across a multigene family of glycosyltransferases in Arabidopsis. *Glycobiology* **13,** 139–145 (2003).

129. Lim, E.-K., Doucet, C. J., Li, Y., Elias, L., Worrall, D., Spencer, S. P., Ross, J. and Bowles, D. J. The activity of Arabidopsis glycosyltransferases toward salicylic acid, 4-hydroxybenzoic acid, and other benzoates. *J. Biol. Chem.* **277,** 586–92 (2002).

130. Lim, E. K., Ashford, D. A., Hou, B. K., Jackson, R. G. and Bowles, D. J. Arabidopsis glycosyltransferases as biocatalysts in fermentation for regioselective synthesis of diverse quercetin glucosides. *Biotechnol. Bioeng.* **87,** 623–631 (2004).

131. Strucko, T., Magdenoska, O. and Mortensen, U. H. Benchmarking two commonly used Saccharomyces cerevisiae strains for heterologous vanillin-β-glucoside production.

Metab. Eng. Commun. **2,** 99–108 (2015).

132. Shimoda, K., Kwon, S., Utsuki, A., Ohiwa, S., Katsuragi, H., Yonemoto, N., Hamada, H. and Hamada, H. Glycosylation of capsaicin and 8-nordihydrocapsaicin by cultured cells of *Catharanthus roseus*. *Phytochemistry* **68,** 1391–1396 (2007).

133. Akhtar, M. T., Mustafa, Na. R. and Verpoorte, R. Hydroxylation and glycosylation of Δ9-tetrahydrocannabinol by *Catharanthus roseus* cell suspension culture. *Biocatal. Biotransformation* **33,** 279–286 (2016).

134. Hergenhahn Manfred, Bertram Barbara, Wiessler Manfred, S. B. Curcumin derivatives with improved water solubility compared to curcumin and medicaments containing the same. US20030153512 A1. (2001).

135. Bhaskar Rao, A., Prasad, E., Deepthi, S. S. and Ansari, I. A. Synthesis and Biological Evaluation of Glucosyl Curcuminoids. *Arch. Pharm. (Weinheim).* **347,** 1–6 (2014).

136. R Vijayakumar, G. and Divakar, S. Synthesis of guaiacol-alpha-D: -glucoside and curcumin-bis-alpha-D: -glucoside by an amyloglucosidase from Rhizopus. *Biotechnol. Lett.* **27,** 1411–1415 (2005).

137. Shimoda, K. and Hamada, H. Enzymatic synthesis and anti-allergic activities of curcumin oligosaccharides. *Biochem. Insights* **3,** 1–5 (2010).

138. Kaminaga, Y., Nagatsu, A., Akiyama, T., Sugimoto, N., Yamazaki, T., Maitani, T. and Mizukami, H. Production of unnatural glucosides of curcumin with drastically enhanced water solubility by cell suspension cultures of *Catharanthus roseus*. *FEBS Lett.* **555,** 311–316 (2003).

139. Tabata, M., Ikeda, F., Hiraoka, N. and Konoshima, M. Glucosylation of phenolic compounds by *Datura innoxia* suspension cultures. *Phytochemistry* **15,** 1225–1229 (1976).

140. Orihara, Y., Noguchi, T. and Furuya, T. Biotransformation of (+)-camphor by cultured cells of *Eucalyptus perriniana*. *Phytochemistry* **35,** 941–945 (1994).

141. Huang, F.-C., Hinkelmann, J. and Schwab, W. Glucosylation of aroma chemicals and hydroxy fatty acids. *J. Biotechnol.* **216,** 100–109 (2015).

142. Masada, S., Kawase, Y., Nagatoshi, M., Oguchi, Y., Terasaka, K. and Mizukami, H. An efficient chemoenzymatic production of small molecule glucosides with in situ UDP-glucose recycling. *FEBS Lett.* **581,** 2562–2566 (2007).

143. Masada, S., Terasaka, K., Oguchi, Y., Okazaki, S., Mizushima, T. and Mizukami, H. Functional and structural characterization of a flavonoid glucoside 1,6-glucosyltransferase from *Catharanthus roseus*. *Plant Cell Physiol.* **50,** 1401–1415 (2009).

144. Oguchi, Y., Masada, S., Kondo, T., Terasaka, K. and Mizukami, H. Purification and characterization of UDP-glucose : curcumin glucoside 1,6-glucosyltransferase from *Catharanthus roseus* cell suspension cultures. *Plant Cell Physiol.* **48,** 1635–1643 (2007).

145. Shimoda, K. Production of Curcumin Oligosaccharides by Glycosylation with *Parthenocissus tricuspidata*. *Japanese J. Plant Sci.* **2,** 43–45 (2008).

146. Katsuyama, Y., Matsuzawa, M., Funa, N. and Horinouch, S. Production of curcuminoids by *Escherichia coli* carrying an artificial biosynthesis pathway. *Microbiology-Sgm* **154,** 2620–2628 (2008).

147. Borodina, I. and Nielsen, J. Advances in metabolic engineering of yeast S*accharomyces cerevisiae* for production of chemicals. *Biotechnol. J.* **9,** 609–620 (2014).

148. Hong, K.-K. and Nielsen, J. Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. *Cell. Mol. Life Sci.* **69,** 2671–2690 (2012).

149. Di Gioia, D., Sciubba, L., Setti, L., Luziatelli, F., Ruzzi, M., Zanichelli, D. and Fava, F. Production of biovanillin from wheat bran. *Enzyme Microb. Technol.* **41,** 498–505 (2007).

150. Torres, B. R., Aliakbarian, B., Torre, P., Perego, P., Domínguez, J. M., Zilli, M. and Converti, A. Vanillin bioproduction from alkaline hydrolyzate of corn cob by *Escherichia coli* JM109/pBB1. *Enzyme Microb. Technol.* **44,** 154–158 (2009).

151. Eudes, A., Teixeira Benites, V., Wang, G., Baidoo, E. E. K., Lee, T. S., Keasling, J. D. and Loqué, D. Precursor-Directed Combinatorial Biosynthesis of Cinnamoyl, Dihydrocinnamoyl, and Benzoyl Anthranilates in *Saccharomyces cerevisiae*. *PLoS One* **10,** e0138972 (2015).

152. Adeboye, P. T., Bettiga, M., Aldaeus, F., Larsson, P. T. and Olsson, L. Catabolism of coniferyl aldehyde, ferulic acid and p-coumaric acid by *Saccharomyces cerevisiae* yields less toxic products. *Microb. Cell Fact.* **14,** 149–163 (2015).

153. Douglas, C. J. Phenylpropanoid metabolism and lignin biosynthesis: from weeds to trees. *Trends Plant Sci.* **1,** 171–178 (1996).

154. Cukovic, D., Ehlting, J., VanZiffle, J. A. and Douglas, C. J. Structure and evolution of 4-coumarate:coenzyme A ligase (4CL) gene families. *Biol. Chem.* **382,** 645–654 (2001).

155. Knobloch, K. H. and Hahlbrock, K. 4-Coumarate:CoA ligase from cell suspension cultures of *Petroselinum hortense* Hoffm. Partial purification, substrate specificity, and further properties. *Arch. Biochem. Biophys.* **184,** 237–248 (1977).

156. Dixon, R. A. and Paiva, N. L. Stress-Induced Phenylpropanoid Metabolism. *Plant Cell* **7,** 1085–1097 (1995).

157. Becker-André, M., Schulze-Lefert, P. and Hahlbrock, K. Structural comparison, modes of expression, and putative cis-acting elements of the two 4-coumarate: CoA ligase genes in potato. *J. Biol. Chem.* **266,** 8551–8559 (1991).

158. Ehlting, J., Shin, J. and Douglas, C. Identification of 4-coumarate: coenzyme A ligase (4CL) substrate recognition domains. *Plant J.* **27,** 455–465 (2001).

159. Lee, D. Two divergent members of a tobacco 4-coumarate:coenzyme A ligase (4CL) gene family. cDNA structure, gene inheritance and expression, and properties of recombinant proteins. *PLANT Physiol.* **112,** 193–205 (1996).

160. Lozoya, E., Hoffman, H., Douglas, C., Schulz, W., Scheel, D. and Hahlbrock, K. Primary structures and catalytic properties of isoenzymes encoded by the two 4-coumarate: CoA ligase genes in parsley. *Eur. J. Biochem.* **176,** 661–667 (1988).

161. Allina, S. M. 4-Coumarate:Coenzyme A Ligase in Hybrid Poplar . Properties of Native Enzymes, cDNA Cloning, and Analysis of Recombinant Enzymes. *PLANT Physiol.* **116,** 743–754 (1998).

162. Ehlting, J., Buttner, D., Wang, Q., Douglas, C. J., Somssich, I. E. and Kombrink, E. Three 4-coumarate:coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *Plant J.* **19,** 9–20 (1999).

163. Hu, W.-J., Kawaoka, A., Tsai, C.-J., Lung, J., Osakabe, K., Ebinuma, H. and Chiang, V. L. Compartmentalized expression of two structurally and functionally distinct 4-coumarate:CoA ligase genes in aspen (*Populus tremuloides*). *Proc. Natl. Acad. Sci.* **95,** 5407–5412 (1998).

164. Sun, H., Li, Y., Feng, S., Zou, W., Guo, K., Fan, C., Si, S. and Peng, L. Analysis of five rice 4-coumarate:coenzyme A ligase enzyme activity and stress response for potential roles in lignin and flavonoid biosynthesis in rice. *Biochem. Biophys. Res. Commun.* **430,** 1151–1156 (2013).

165. Xu, B., Escamilla-Treviño, L. L., Sathitsuksanoh, N., Shen, Z., Shen, H., Percival Zhang,

Y.-H., Dixon, R. A. and Zhao, B. Silencing of 4-coumarate:coenzyme A ligase in switchgrass leads to reduced lignin content and improved fermentable sugar yields for biofuel production. *New Phytol.* **192,** 611–625 (2011).

166. Hamberger, B. and Hahlbrock, K. The 4-coumarate:CoA ligase gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 2209–2214 (2004).

167. Hatfield, R., Ralph, J. and Grabber, J. H. A potential role for sinapyl *p*-coumarate as a radical transfer mechanism in grass lignin formation. *Planta* **228,** 919–928 (2008).

168. Hatfield, R. D., Marita, J. M., Frost, K., Grabber, J., Ralph, J., Lu, F. and Kim, H. Grass lignin acylation: p-coumaroyl transferase activity and cell wall characteristics of C3 and C4 grasses. *Planta* **229,** 1253–1267 (2009).

169. Lindermayr, C., Möllers, B., Fliegmann, J., Uhlmann, A., Lottspeich, F., Meimberg, H. and Ebel, J. Divergent members of a soybean (*Glycine max* L.) 4-coumarate:coenzyme A ligase gene family. *Eur. J. Biochem.* **269,** 1304–1315 (2002).

170. Stuible, H.-P., Büttner, D., Ehlting, J., Hahlbrock, K. and Kombrink, E. Mutational analysis of 4-coumarate:CoA ligase identifies functionally important amino acids and verifies its close relationship to other adenylate-forming enzymes. *FEBS Lett.* **467,** 117–122 (2000).

171. Stuible, H. P. and Kombrink, E. Identification of the substrate specificity-conferring amino acid residues of 4-coumarate:coenzyme A ligase allows the rational design of mutant enzymes with new catalytic properties. *J. Biol. Chem.* **276,** 26893–26897 (2001).

172. Schneider, K., Hövel, K., Witzel, K., Hamberger, B., Schomburg, D., Kombrink, E. and Stuible, H.-P. The substrate specificity-determining amino acid code of 4-coumarate:CoA ligase. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 8601–8606 (2003).

173. Gui, J., Shen, J. and Li, L. Functional characterization of evolutionarily divergent 4-coumarate:coenzyme a ligases in rice. *Plant Physiol.* **157,** 574–586 (2011).

174. Koo, H. J., McDowell, E. T., Ma, X., Greer, K. A., Kapteyn, J., Xie, Z., Descour, A., Kim, H., Yu, Y. and Kudrna, D. Ginger and turmeric expressed sequence tags identify signature genes for rhizome identity and development and the biosynthesis of curcuminoids, gingerols and terpenoids. *BMC Plant Biol.* **13,** 27–44 (2013).

175. Brazier-Hicks, M. and Edwards, R. Metabolic engineering of the flavone-C-glycoside

pathway using polyprotein technology. *Metab. Eng.* **16,** 11–20 (2012).

176. Winzer, T., Kern, M., King, A. J., Larson, T. R., Teodor, R. I., Donninger, S. L., Li, Y., Dowle, A. A., Cartwright, J., Bates, R., Ashford, D., Thomas, J., Walker, C., Bowser, T. A. and Graham, I. A. Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein. *Science (80-.).* **349,** 309–312 (2015).

177. Křenek, P., Šamajová, O., Luptovčiak, I., Doskočilová, A., Komis, G. and Šamaj, J. Transient plant transformation mediated by *Agrobacterium tumefaciens*: Principles, methods and applications. *Biotechnol. Adv.* **33,** 1024–1042 (2015).

178. Dong, J.-Z. and Dunstan, D. A reliable method for extraction of RNA from various conifer tissues. *Plant Cell Rep.* **15,** 516–521 (1996).

179. Hamberger, B. and Hahlbrock, K. The 4-coumarate:CoA ligase gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 2209–2214 (2004).

180. Larson, T. R. and Graham, I. A. Technical Advance: A novel technique for the sensitive quantification of acyl CoA esters from plant tissues. *Plant J.* **25,** 115–125 (2008).

181. Padhye, S., Banerjee, S., Chavan, D., Pandye, S., Swamy, K. V., Ali, S., Li, J., Dou, Q. P. and Sarkar, F. H. Fluorocurcumins as Cyclooxygenase-2 Inhibitor: Molecular Docking, Pharmacokinetics and Tissue Distribution in Mice. *Pharm. Res.* **26,** 2438–2445 (2009).

182. Padhye, S., Yang, H., Jamadar, A., Cui, Q. C., Chavan, D., Dominiak, K., McKinney, J., Banerjee, S., Dou, Q. P. and Sarkar, F. H. New Difluoro Knoevenagel Condensates of Curcumin, Their Schiff Bases and Copper Complexes as Proteasome Inhibitors and Apoptosis Inducers in Cancer Cells. *Pharm. Res.* **26,** 1874–1880 (2009).

183. Eudes, A., Baidoo, E. E. K., Yang, F., Burd, H., Hadi, M. Z., Collins, F. W., Keasling, J. D. and Loqué, D. Production of tranilast [N-(3',4'-dimethoxycinnamoyl)-anthranilic acid] and its analogs in yeast *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* **89,** 989–1000 (2011).

184. Lussier, F.-X., Colatriano, D., Wiltshire, Z., Page, J. E. and Martin, V. J. J. Engineering Microbes for Plant Polyketide Biosynthesis. *Comput. Struct. Biotechnol. J.* **3,** e201210020 (2012).

185. Kawaguchi, A., Yoshimura, T. and Okuda, S. A new method for the preparation of acyl-CoA thioesters. *J. Biochem.* **89,** 337–339 (1981).

186. Gustafsson, C., Govindarajan, S. and Minshull, J. Codon bias and heterologous protein expression. *Trends Biotechnol.* **22,** 346–353 (2004).

187. Ikemura, T. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J. Mol. Biol.* **151,** 389–409 (1981).

188. Kane, J. F. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* **6,** 494–500 (1995).

189. Presnyak, V., Alhusaini, N., Chen, Y.-H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K. E., Graveley, B. R. and Coller, J. Codon Optimality Is a Major Determinant of mRNA Stability. *Cell* **160,** 1111–1124 (2015).

190. Crick, F. H. C. Codon—anticodon pairing: The wobble hypothesis. *J. Mol. Biol.* **19,** 548–555 (1966).

191. Ferrer, J. L., Jez, J. M., Bowman, M. E., Dixon, R. A. and Noel, J. P. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nat. Struct. Biol.* **6,** 775–784 (1999).

192. DeFranco, D., Dingermann, T., Johnson, D. L., Sharp, S. and Söll, D. Eukaryotic tRNA gene transcription is controlled by signals within and outside the mature coding sequence. *Princess Takamatsu Symp.* **12,** 63–72 (1982).

193. Sharp, S., Dingermann, T. and Söll, D. The minimum intragenic sequences required for promotion of eukaryotic tRNA gene transcription. *Nucleic Acids Res.* **10,** 5393–5406 (1982).

194. Ryan, M. D., King, A. M. Q. and Thomas, G. P. Cleavage of foot-and-mouth-disease virus polyprotein is mediated by residues located within a 19 amino-acid sequence. *J. Gen. Virol.* **72,** 2727–2732 (1991).

195. Ryan, M. D. and Drew, J. Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *Embo J.* **13,** (1994).

196. Donnelly, M. L. L., Luke, G., Mehrotra, A., Li, X. J., Hughes, L. E., Gani, D. and Ryan, M. D. Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. *J. Gen. Virol.* **82,** 1013–1025 (2001).

197. Ha, S.-H., Liang, Y. S., Jung, H., Ahn, M.-J., Suh, S.-C., Kweon, S.-J., Kim, D.-H., Kim, Y.-M. and Kim, J.-K. Application of two bicistronic systems involving 2A and IRES sequences to the biosynthesis of carotenoids in rice endosperm. *Plant Biotechnol. J.* **8,** 928–938 (2010).

198. Beekwilder, J., van Rossum, H. M., Koopman, F., Sonntag, F., Buchhaupt, M., Schrader, J., Hall, R. D., Bosch, D., Pronk, J. T., van Maris, A. J. A. and Daran, J.-M. Polycistronic expression of a β-carotene biosynthetic pathway in *Saccharomyces cerevisiae* coupled to β-ionone production. *J. Biotechnol.* **192,** 383–392 (2014).

199. Szymczak, A. L., Workman, C. J., Wang, Y., Vignali, K. M., Dilioglou, S., Vanin, E. F. and Vignali, D. A. A. Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. *Nat. Biotechnol.* **22,** 589–594 (2004).

200. de Felipe, P., Hughes, L. E., Ryan, M. D. and Brown, J. D. Co-translational, intraribosomal cleavage of polypeptides by the foot-and-mouth disease virus 2A peptide. *J. Biol. Chem.* **278,** 11441–11448 (2003).

201. Verma, P., Mathur, A. K., Srivastava, A. and Mathur, A. Emerging trends in research on spatial and temporal organization of terpenoid indole alkaloid pathway in Catharanthus roseus: a literature update. *Protoplasma* **249,** 255–268 (2012).

202. Nielsen, A. Z., Ziersen, B., Jensen, K., Lassen, L. M., Olsen, C. E., Møller, B. L. and Jensen, P. E. Redirecting Photosynthetic Reducing Power toward Bioactive Natural Product Synthesis. *ACS Synth. Biol.* **2,** 308–315 (2013).

203. Gordon, O. N., Luis, P. B., Sintim, H. O. and Schneider, C. Unraveling curcumin degradation: autoxidation proceeds through spiroepoxide and vinylether intermediates en route to the main bicyclopentadione. *J. Biol. Chem.* **290,** 4817–4828 (2015).

204. Schneider, C., Gordon, O. N., Edwards, R. L. and Luis, P. B. Degradation of Curcumin: From Mechanism to Biological Implications. *J. Agric. Food Chem.* **63,** 7606–7614 (2015).

205. Costa, M. A., Bedgar, D. L., Moinuddin, S. G. A., Kim, K.-W., Cardenas, C. L., Cochrane, F. C., Shockey, J. M., Helms, G. L., Amakura, Y., Takahashi, H., Milhollan, J. K., Davin, L. B., Browse, J. and Lewis, N. G. Characterization in vitro and in vivo of the putative multigene 4-coumarate:CoA ligase network in Arabidopsis: syringyl lignin and sinapate/sinapyl alcohol derivative formation. *Phytochemistry* **66,** 2072–2091 (2005).

206. Kunau, W. β-Oxidation of fatty acids in mitochondria, peroxisomes, and bacteria: A century of continued progress. *Prog. Lipid Res.* **34,** 267–342 (1995).

207. Shani, N., Watkins, P. A. and Valle, D. PXA1, a possible *Saccharomyces cerevisiae* ortholog of the human adrenoleukodystrophy gene. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 6012–6016 (1995).

208. Shani, N. and Valle, D. A *Saccharomyces cerevisiae* homolog of the human adrenoleukodystrophy transporter is a heterodimer of two half ATP-binding cassette transporters. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 11901–11906 (1996).

209. Verleur, N., Hettema, E. H., Roermund, C. W. T., Tabak, H. F. and Wanders, R. J. A. Transport of Activated Fatty Acids by the Peroxisomal ATP-binding-cassette Transporter Pxa2 in a Semi-Intact Yeast Cell System. *Eur. J. Biochem.* **249,** 657–661 (1997).

210. van Roermund, C. W. T., Waterham, H. R., Ijlst, L. and Wanders, R. J. A. Fatty acid metabolism in *Saccharomyces cerevisiae*. *Cell. Mol. Life Sci.* **60,** 1838–1851 (2003).

211. Jiang, H. X., Wood, K. V and Morgan, J. A. Metabolic engineering of the phenylpropanoid pathway in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **71,** 2962–2969 (2005).

212. Eudes, A., Juminaga, D., Baidoo, E. E. K., Collins, F. W., Keasling, J. D. and Loqué, D. Production of hydroxycinnamoyl anthranilates from glucose in *Escherichia coli*. *Microb. Cell Fact.* **12,** 62–72 (2013).

213. Huang, Y. and Maraia, R. J. Comparison of the RNA polymerase III transcription machinery in *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae* and human. *Nucleic Acids Res.* **29,** 2675–2690 (2001).

214. Dieci, G., Percudani, R., Giuliodori, S., Bottarelli, L. and Ottonello, S. TFIIIC-independent in vitro transcription of yeast tRNA genes. *J. Mol. Biol.* **299,** 601–613 (2000).

215. Fowler, Z. L. and Koffas, M. A. G. Biosynthesis and biotechnological production of flavanones: current state and perspectives. *Appl. Microbiol. Biotechnol.* **83,** 799–808 (2009).

216. Hrazdina, G. and Wagner, G. J. Metabolic pathways as enzyme complexes: Evidence for the synthesis of phenylpropanoids and flavonoids on membrane associated enzyme complexes. *Arch. Biochem. Biophys.* **237,** 88–100 (1985).

217. Katsuyama, Y., Kita, T. and Horinouchi, S. Identification and characterization of multiple curcumin synthases from the herb *Curcuma longa*. *Febs Lett.* **583,** (2009).

218. Gilbert, I. H., Ginty, M., O'Neill, J. A., Simpson, T. J., Staunton, J. and Willis, C. L. Synthesis of β-keto and α,β-unsaturated N-acetylcysteamine thioesters. *Bioorg. Med.*

Chem. Lett. **5,** 1587–1590 (1995).

219. Lee, D. and Douglas, C. J. Two divergent members of a tobacco 4-coumarate:coenzyme A ligase (4CL) gene family. cDNA structure, gene inheritance and expression, and properties of recombinant proteins. *Plant Physiol.* **112,** 193–205 (1996).

220. Ralph, J., Hatfield, R. D., Quideau, S., Helm, R. F., Grabber, J. H. and Jung, H.-J. G. Pathway of p-Coumaric Acid Incorporation into Maize Lignin As Revealed by NMR. *J. Am. Chem. Soc.* **116,** 9448–9456 (1994).

221. Holder, G. M., Plummer, J. L. and Ryan, A. J. The Metabolism and Excretion of Curcumin (1,7-Bis-(4-hydroxy-3-methoxyphenyl)-1,6-heptadiene-3,5-dione) in the Rat. *Xenobiotica* **8,** 761–768 (2008).

222. Sandermann, H. Plant metabolism of xenobiotics. *Trends Biochem. Sci.* **17,** 82–84 (1992).

223. Pinçon, G., Maury, S., Hoffmann, L., Geoffroy, P., Lapierre, C., Pollet, B. and Legrand, M. Repression of O-methyltransferase genes in transgenic tobacco affects lignin synthesis and plant growth. *Phytochemistry* **57,** 1167–1676 (2001).

224. Shaikh, J., Ankola, D. D., Beniwal, V., Singh, D. and Kumar, M. N. V. R. Nanoparticle encapsulation improves oral bioavailability of curcumin by at least 9-fold when compared to curcumin administered with piperine as absorption enhancer. *Eur. J. Pharm. Sci.* **37,** 223–230 (2009).

225. Liu, Z., Tang, L., Zou, P., Zhang, Y., Wang, Z., Fang, Q., Jiang, L., Chen, G., Xu, Z., Zhang, H. and Liang, G. Synthesis and biological evaluation of allylated and prenylated mono-carbonyl analogs of curcumin as anti-inflammatory agents. *Eur. J. Med. Chem.* **3,** 671–682 (2013).

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

227. Sato, D., Eshita, Y., Katsuragi, H., Hamada, H., Shimoda, K. and Kubota, N. Glycosylation of vanillin and 8-nordihydrocapsaicin by cultured *Eucalyptus perriniana* cells. *Molecules* **17,** 5013–5020 (2012).

228. Inomata, S., Yokoyama, M., Seto, S. and Yanagi, M. High-level production of arbutin from hydroquinone in suspension cultures of *Catharanthus roseus* plant cells. *Appl. Microbiol. Biotechnol.* **36,** 315–319 (1991).

229. Kometani, T., Tanimoto, H., Nishimura, T., Kanbara, I. and Okada, S. Glucosylation of

Capsaicin by Cell Suspension Cultures of *Coffea arabica*. *Biosci. Biotechnol. Biochem.* **57,** 2192–2193 (2014).

230. Sharma, H. A., Barber, J. T., Ensley, H. E. and Polito, M. A. A comparison of the toxicity and metabolism of phenol and chlorinated phenols by *Lemna gibba*, with special reference to 2,4,5-trichlorophenol. *Environ. Toxicol. Chem.* **16,** 346 (1997).

231. Kaminaga, Y., Nagatsu, A., Akiyama, T., Sugimoto, N., Yamazaki, T., Maitani, T. and Mizukami, H. Production of unnatural glucosides of curcumin with drastically enhanced water solubility by cell suspension cultures of *Catharanthus roseus*. *FEBS Lett.* **555,** 311–316 (2003).

232. Minear, S., O'Donnell, A. F., Ballew, A., Giaever, G., Nislow, C., Stearns, T. and Cyert, M. S. Curcumin inhibits growth of *Saccharomyces cerevisiae* through iron chelation. *Eukaryot. Cell* **10,** 1574–1581 (2011).

233. Flick, J. S. and Johnston, M. Two systems of glucose repression of the GAL1 promoter in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **10,** 4757–4769 (1990).

234. Turcotte, B., Liang, X. B., Robert, F. and Soontorngun, N. Transcriptional regulation of nonfermentable carbon utilization in budding yeast. *FEMS Yeast Res.* **10,** 2–13 (2010).

235. Masada, S., Kawase, Y., Nagatoshi, M., Oguchi, Y., Terasaka, K. and Mizukami, H. An efficient chemoenzymatic production of small molecule glucosides with in situ UDP-glucose recycling. *FEBS Lett.* **581,** 2562–2566 (2007).

236. Ramirez-Ahumada, M. D., Timmermann, B. N. and Gang, D. R. Biosynthesis of curcuminoids and gingerols in turmeric (Curcuma longa) and ginger (Zingiber officinale): Identification of curcuminoid synthase and hydroxycinnamoyl-CoA thioesterases. *Phytochemistry* **67,** 2017–2029 (2006).

237. Xie, Z., Ma, X. and Gang, D. R. Modules of co-regulated metabolites in turmeric (*Curcuma longa*) rhizome suggest the existence of biosynthetic modules in plant specialized metabolism. *J. Exp. Bot.* **60,** 87–97 (2009).

238. Koo, H. J. and Gang, D. R. Suites of terpene synthases explain differential terpenoid production in ginger and turmeric tissues. *PLoS One* **7,** e51481 (2012).

239. Humphrey, A. Shake flask to fermentor: what have we learned? *Biotechnol. Prog.* **14,** 3–7 (1998).

240. Winge, O. and Roberts, C. The relation between the polymeric genes for maltose raffinose,

and sucrose fermentation in yeasts. **25,** 141–171 (1952).

241. Liljestrom-Suominen, P. L., Joutsjoki, V. and Korhola, M. Construction of a Stable {alpha}-Galactosidase-Producing Baker's Yeast Strain. *Appl. Envir. Microbiol.* **54,** 245–249 (1988).

242. Koster, A. S. and Noordhoek, J. Kinetic properties of the rat intestinal microsomal 1-naphthol:UDP-glucuronosyl transferase. Inhibition by UDP and UDP-N-acetylglucosamine. *Biochim. Biophys. Acta* **761,** 76–85 (1983).

243. Fujiwara, R., Nakajima, M., Yamanaka, H., Katoh, M. and Yokoi, T. Product inhibition of UDP-glucuronosyltransferase (UGT) enzymes by UDP obfuscates the inhibitory effects of UGT substrates. *Drug Metab. Dispos.* **36,** 361–367 (2008).

244. Larson, T. R. and Graham, I. A. Technical Advance: a novel technique for the sensitive quantification of acyl CoA esters from plant tissues. *Plant J.* **25,** 115–125 (2001).

245. Aggarwal, B. B., Sundaram, C., Malani, N. and Ichikawa, H. Curcumin: The Indian solid gold. *Mol. Targets Ther. Uses Curcumin Heal. Dis.* **595,** 1–75 (2007).

246. Hernanz, D., Nuñez, V., Sancho, A. I., Faulds, C. B., Williamson, G., Bartolomé, B. and Gómez-Cordovés, C. Hydroxycinnamic Acids and Ferulic Acid Dehydrodimers in Barley and Processed Barley. *J. Agric. Food Chem.* **49,** 4884–4888 (2001).

247. Weng, J.-K. and Noel, J. P. Structure-function analyses of plant type III polyketide synthases. *Methods Enzymol.* **515,** 317–335 (2012).

248. Luke, G. A., Escuin, H., De Felipe, P. and Ryan, M. D. 2A to the Fore - Research, Technology and Applications. *Biotechnol. Genet. Eng. Rev. Vol 26* **26,** 223–260 (2010).

249. Pfeiffer, E., Höhle, S., Solyom, A. M. and Metzler, M. Studies on the stability of turmeric constituents. *J. Food Eng.* **56,** 257–259 (2003).

250. Bernabé-Pineda, M., Ramírez-Silva, M. T., Romero-Romo, M., González-Vergara, E. and Rojas-Hernández, A. Determination of acidity constants of curcumin in aqueous solution and apparent rate constant of its decomposition. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **60,** 1091–1097 (2004).

251. Tønnesen, H. H. Solubility, chemical and photochemical stability of curcumin in surfactant solutions. Studies of curcumin and curcuminoids, XXVIII. *Pharmazie* **57,** 820–824 (2002).

252. Agarwal, N. B., Jain, S., Nagpal, D., Agarwal, N. K., Mediratta, P. K. and Sharma, K. K. Liposomal formulation of curcumin attenuates seizures in different experimental models of epilepsy in mice. *Fundam. Clin. Pharmacol.* **27,** 169–172 (2013).

253. Xie, Z., Ma, X. and Gang, D. R. Modules of co-regulated metabolites in turmeric (*Curcuma longa*) rhizome suggest the existence of biosynthetic modules in plant specialized metabolism. *J. Exp. Bot.* **60,** 87–97 (2009).

254. Partow, S., Siewers, V., Bjørn, S., Nielsen, J. and Maury, J. Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*. *Yeast* **27,** 955–964 (2010).

255. Curran, K. A., Karim, A. S., Gupta, A. and Alper, H. S. Use of high capacity terminators in *Saccharomyces cerevisiae* to increase mRNA half-life and improve gene expression control for metabolic engineering applications. *Metab Eng* **19,** 88–97 (2013).

256. Ito, Y., Yamanishi, M., Ikeuchi, A., Imamura, C., Tokuhiro, K., Kitagawa, T. and Matsuyama, T. Characterization of five terminator regions that increase the protein yield of a transgene in *Saccharomyces cerevisiae*. *J. Biotechnol.* **168,** 486–492 (2013).