Microbial Biosynthesis of Grapevine Specialised Metabolites

Emi Kaye Schutz

Supervisors:
Assoc. Prof. Christopher Ford
Dr John Bruning

The University of Adelaide
School of Agriculture, Food and Wine
February 2019

Thesis submitted for the degree of Master of Philosophy

Table of Contents

Abstract

Plants are capable of producing a wide variety of molecules, and these fall broadly into two categories: primary metabolites and specialised metabolites. Primary metabolites are critical for growth and survival. Specialised metabolites, which are present in some plant species but not others, do not have a role in the central metabolism of all plants. Specialised metabolites can play many roles to allow plants to fill specific ecological niches, such as chemical defence or attraction of pollinators. Specialised metabolites may also be useful to humans as drugs, flavours or fragrances.

The research described in this thesis investigates the potential of applying microbial biosynthesis techniques to produce grapevine (*Vitis vinifera*) specialised metabolites. Two classes of specialised metabolites were focussed on: tartaric acid and its biosynthetic intermediates, and sesquiterpenoids. Tartaric acid accumulates in grape berries during development and is essential to the winemaking process as it lowers the pH of the wine must, preventing discolouration and microbial spoilage. Sesquiterpenoids are a large class of compounds, many of which have unique aromas, and some have been found to contribute significantly to the character of wine.

Bioinformatics were used to identify candidate 2-keto-L-gulonate reductases that may be involved in the biosynthesis of tartaric acid. Two candidate enzymes were expressed and purified for *in vitro* characterisation. It was found that these candidates are capable of utilising 2-keto-L-gulonate as a substrate but do not produce L-idonate, the next step in the tartaric acid pathway. The candidate enzymes exhibited low-level, broad-spectrum reductase activity. A previously identified 2-keto-L-gulonate reductase was used in the development of an *E. coli* cell factory for the whole cell biocatalysis of 2-keto-L-gulonate to L-idonate.

A yeast strain specially engineered to overproduce the sesquiterpene precursor farnesyl pyrophosphate (FPP) was used for the synthesis of a range of grapevine sesquiterpenoid compounds. Six grapevine sesquiterpene synthases were expressed by the engineered yeast strain alongside a promiscuous P450 monooxygenase to produce potential grapevine sesquiterpenoid products *in vivo*. The compounds were identified by gas-chromatography-mass-spectrometry. A total of seventeen unique compounds were confidently identified, with a number of uncharacterised compounds also produced. These compounds were found to possess a variety of aromas, and some compounds have beneficial health effects, which may be relevant for wine making and marketing.

Finally, homology modelling and docking studies were used to compare three grapevine P450 monooxygenases and investigate the mechanism of the formation of the sesquiterpenoid (-)-rotundone, the pepper aroma compound found in some wines. Computational docking studies of (-)-rotundone inside the P450 binding sites suggests the oxidation of α-guaiene to (-)-rotundone is due to the unique shape of the binding site of *V. vinifera* Sesquiterpene Oxidase 2 (VvSTO2). Two other P450s, VvSTO4 and VvSTO6 are able to accept α-guaiene as a substrate but do not form rotundone, due to different positioning of α-guaiene within the P450 active sites.

The investigation of these specialised metabolites may provide new insights into the biosynthesis of tartaric acid and sesquiterpenoids in grapevine, and assist in the identification and production of novel wine aromas and flavours.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

23/01/2019

Emi Schutz Date

Acknowledgements

First of all, thank you to my supervisor Associate Professor Christopher Ford and my co-supervisor Dr John Bruning. Chris, thank you for your encouragement, advice, feedback, and for allowing me time and flexibility to explore this topic and make the project my own. John, thank you for making time to show me the hands-on side of protein purification. Thank you also to Associate Professor Stephen Bell for lending his expertise and guidance to the project.

Thank you to members of the Wine Microbiology lab and the Bruning lab for welcoming me into your workspace and assisting me with any questions. Thanks in particular to Michelle Walker, who guided me through the wild world of yeast microbiology, and Nick van Holst Pellekaan, whose technical expertise and assistance was invaluable. Many thanks also to Tony Hall for running my GC-MS samples and for sharing his extensive knowledge with me.

During my studies I was fortunate to have the opportunity to travel to Calgary, Canada and I must acknowledge the Australian Government's Endeavour Leadership Program for providing me with funding and support for this incredible experience. Thank you to Dae-Kyun Ro for welcoming me into your country and your laboratory, and to the members of the Ro lab for their support, advice and friendship, both within and outside of the lab.

Finally, thank you to my family who have supported me through my years of study, my friends, and especially to my fiancé, Kieran. I could not have done it without you all.

Chapter 1 Introduction

Specialised Metabolism in Plants

Plants are capable of synthesising a wide range of chemical compounds for many purposes, from basic metabolism to growth and survival. Primary metabolites are defined as compounds synthesised by all or most plant species, and it is estimated that there are less than 10,000 primary metabolites (Pichersky & Lewinsohn 2011). Specialised metabolites are compounds present in some plant species but not others, and do not have a role in the central, primary metabolism of all plants. Specialised metabolites can play many roles to allow plants to fill specific ecological niches. Some act as chemical attractants for pollination, or as defence against other organisms (Pichersky & Lewinsohn 2011). The number of specialised metabolites made by plant species has been estimated at approximately 200,000 (Dixon & Strack 2003). The functions of many specialised metabolites are yet to be discovered.

Specialised metabolites can be useful to humans. For example, many active drug compounds are natural products or natural product derivatives (Cragg, Newman & Snader 1997). Specialised metabolites can also contribute to the flavour and aroma of food. The wine grape (*Vitis vinifera)* produces many specialised metabolites that confer its unique properties and potential for winemaking (Ali et al. 2010). Grapes produce a large number of terpene compounds, predominantly monoterpenes and sesquiterpenes (Drew et al. 2016). In particular, the sesquiterpenoid rotundone, found in some varieties of grapevine, most notably Shiraz/Syrah, has been shown to have a potent peppery aroma and is an important contributor to wine character (Wood et al. 2008). Another important contributor to the organoleptic properties of wine is tartaric acid. Although tartaric acid is widespread throughout the plant kingdom, it accumulates in only very few species, and thus its biosynthesis can be considered specialised metabolism.

Tartaric acid in plants

Tartaric acid (TA) is a four-carbon dicarboxylic acid. It has a first dissociation constant (pK_1) of 2.98 and a second dissociation constant (pK_2) of 4.34, making it comparatively more acidic than other common organic acids including citric acid, malic acid, and ascorbic acid (Table 1-1) (Belitz, Grosch & Schieberle 2009).

Table 1-1 Dissociation constants for tartaric, citric, malic and ascorbic acids

Acid	**pK_1**	**pK_2**	**pK_3**
Tartaric	2.98	4.34	N/A
Citric	3.09	4.74	5.41
Malic	3.40	5.05	N/A
Ascorbic	4.17	11.57	N/A

TA exists as three stereoisomers, namely L-(+)-tartaric acid, D-(-)-tartaric acid and meso-tartaric acid (Figure 1-1). All three stereoisomers have been found in nature. L-TA is the most common form, found in members of the Vitaceae and Geraniaceae families (Stafford 1959). D-TA has been found in *Bauhinia reticulata,* and as caeffic acid esters in *Cichoriumintybus L., C. endivia L. and Lactuca sativa L.* (Wagner, Yang & Loewus 1975). Meso-TA occurs in spinach (*Spinacia oleracea*) as an ester of paracoumaryl alcohol (Suzuki et al. 1970). While TA has been found in a large number of species, only members of the Vitaceae, Geraniaceae and Leguminosae plant families accumulate significant amounts (Stafford 1959). These species accumulate only the L isomer of TA (Wagner, Yang & Loewus 1975).

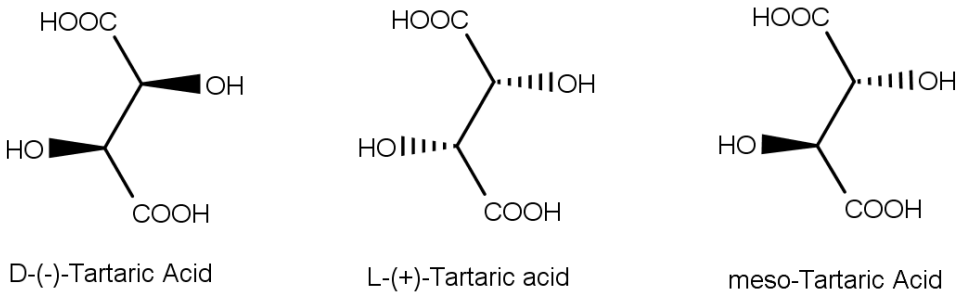


Figure 1-1 Tartaric acid exists as three stereoisomers: the optically active D and L forms and the optically inactive meso form

TA can be used as a carbon source by some species of yeast and bacteria (Fonseca 1992; Klasen, Bringer-Meyer & Sahm 1992). L-(+)-TA is most widely accessible but a small number of microorganisms can utilise D-(-)-TA and meso-TA (Fonseca 1992; Gutnick et al. 1969). However; only the biosynthetic pathway of L-(+)-TA has been investigated (Loewus 1999). Therefore, only L-(+)-TA will be discussed from here.

Sesquiterpenoids in Plants

Terpenes are one of the largest classes of plant natural products, comprising at least 30,000 compounds, and are incredibly structurally diverse (Connolly & Hill 1984). Sesquiterpenoids are 15-carbon terpene derivatives, and have a wide scope of activities in

plants. Many act as semiochemicals for plant signalling or defence (Bohlmann & Keeling 2008). Sesquiterpenoids are often important contributors to plant aroma and components of plant essential oils (Dudareva & Pichersky 2000). Sesquiterpenes may have acyclic, mono-, bi-, tri-, or tetracyclic structures. Thousands of sesquiterpenoids have been identified in plants, although not all have been structurally elucidated (Duhamel et al. 2018).

Rotundone in plants

(-)-Rotundone (Figure 1-2) is a sesquiterpene derivative responsible for the peppery aroma in a number of herbs and other plants. Rotundone was first identified in the tubers of *Cyperus rotundus* (nutgrass) (Kapadia et al. 1967) but its potent aroma was not discovered until the 2000s (Wood et al. 2008). In high concentrations it has a harsh, vinegar-like odour, which becomes spicy and peppery on dilution. Rotundone is present in high levels in pepper (*Piper nigrum*) and nutgrass (*C. rotundus*), and is also present in marjoram (*Origanum majorana*), oregano (*Origanum vulgare*), basil (*Ocimum basilicum*), thyme (*Thymus vulgaris*), rosemary (*Rosmarinus officinalis*), saltbush (*Atriplex cinerea*) and geranium (*Pelargonium alchemilloides*) (Wood et al. 2008). Rotundone is responsible for the spicy, peppery aroma of some Shiraz wines (Davies et al. 2015). It has also been detected in other wine varieties including Mourvedre and Durif (Wood et al. 2008).

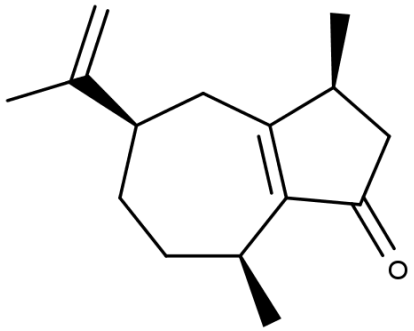


Figure 1-2 (-)-Rotundone, a sesquiterpene with a potent peppery aroma

Grape berry development

Grapevine is a perennial plant that flowers and fruits during spring and summer and is dormant throughout winter. It takes a grapevine at least three years to reach full maturity and produce wine-quality grapes, after which it can remain productive for many decades. The annual growth cycle of the grapevine progresses from budding to formation of shoots, cap fall, flowering, fruit set, veraison, harvest and, finally, leaf fall and winter dormancy (Lorenz et al. 1995). Berry development occurs immediately after flowering. This process

is characterised by rapid cell division and growth during berry formation, then a lag phase of minimal growth, and another period of cell growth during berry ripening (Coombe 1960). During the berry formation phase, TA and malic acid accumulate and are stored in the vacuole of pericarp cells (Ollat et al. 2002). Tartaric acid accumulation begins soon after anthesis (flowering) and is linked to rapid cell division. During ripening, TA content plateaus and concentration decreases slightly due to dilution (Saito & Kasai 1968). Experiments using radiolabelled carbon dioxide show TA synthesis is restricted to the early stages of ripening and is not degraded in the cell (Saito & Kasai 1968). Organic acid levels peak 1-2 weeks before berries change colour, and at this time the levels of tartrate and malate in the berry are approximately equal (Kliewer, Howarth & Omori 1967).

'Veraison' indicates initiation of berry ripening and is hallmarked by acquisition of colour (in red fruit varieties) and softening of berry tissue (Coombe 1960). Before veraison, most of the sugars supplied to the berry are catabolised to provide energy for the rapidly growing cells. After the onset of ripening, hexoses accumulate in pericarp cells at a high rate and the berry grows due to expansion of cell size and uptake of water. The rates of sugar and water accumulation are tightly linked (Ollat et al. 2002). During early veraison, malate is metabolised via the TCA cycle, so TA levels are higher than malic acid levels in ripe grapes (Kliewer, Howarth & Omori 1967; Sweetman et al. 2009).

During veraison the berry also accumulates specialised metabolites including aromatic compounds and polyphenolics. Sesquiterpenes are synthesised in the berry exocarp (May, Lange & Wüst 2013). Sesquiterpene precursors are synthesised in both the plastids and the cytoplasm, and transported to the cytoplasm where they act as a substrate for cytoplasmic sesquiterpene synthases (Martin et al. 2012; May, Lange & Wüst 2013). The levels of most sesquiterpenes continuously increase during the intermediate ripening phase through to post-ripening (Zhang et al. 2016b). At full ripeness the majority of sesquiterpenes are found in the berry exocarp, with only minor amounts in the mesocarp (May, Lange & Wüst 2013).

TA synthesis occurs in both leaf and berry tissue (Saito & Kasai 1968). In both leaf and berry, TA synthesis occurs rapidly during early stages of development and plateaus in mature tissue (Ruffner 1982). Early work suggested organic acids were synthesised in the leaf and transported to the fruit but there is little evidence to support this (Hale 1962). In fact, when radiolabelled TA was injected into leaf tissue, there was no detection of radioactivity in the stem or berry tissue, indicating that it is unlikely that TA is transported into the berry (Ruffner 1982). Work by Hale (1962) established the berry as an important site of TA synthesis. There are several lines of evidence that suggest TA

synthesis occurs in the cytoplasm. Firstly, the TA biosynthetic precursor, ascorbic acid, is found in relatively high concentrations in the cytoplasm (Davey et al. 2000). Secondly, L-idonate dehydrogenase (IdnDH), which catalyses the rate limiting step of the primary TA biosynthetic pathway in grapes, is mainly located in the cytoplasm (Wen et al. 2010). TA and malate are stored in the vacuole, which is characterised by a pH below 3.0 in the pre-veraison berry (Diakou et al. 1997; Moskowitz & Hrazdina 1981). TA and malate are transported into the vacuole via tonoplast vesicles and likely compete for the same transporter (Terrier et al. 1998). Within the vacuole, TA is stored as potassium and calcium salts and free acid (Iland & Coombe 1988).

Levels of TA and salts may reach up to 10 mg per berry (Iland & Coombe 1988; Kliewer, Howarth & Omori 1967). TA is important during the winemaking process as it decreases the pH of the must, preventing microbial spoilage and discolouration (Kalathenos, Sutherland & Roberts 1995). Exogenous TA may be added to the wine must to raise titratable acidity and stabilise the wine during the fermentation and maturation processes. TA is the ideal compound to lower pH because it is not broken down during winemaking, unlike malic acid which is degraded to the weaker lactic acid by malo-lactic fermentation (Liu 2002). TA has a 'zingy' taste and is an important contributor to the palatability, flavour and mouthfeel of wine, while malic acid has an unpleasant, metallic taste (Plane, Mattick & Weirs 1980).

Over 90 different sesquiterpenoids have been identified in grapes and wine products (Duhamel et al. 2018). These exist in varying quantities, ranging from trace amounts up to several micrograms per litre (Duhamel et al. 2018; Parker et al. 2007). Concentrations of rotundone in Shiraz grapes can reach up to 5.44 μg/kg (Caputi et al. 2011). Rotundone concentration is variable between vintages and depends on a number of factors including temperature and soil properties (Scarlett, Bramley & Siebert 2014; Zhang et al. 2015). Rotundone yield during winemaking is relatively low, with approximately only 6% recovered in the bottled wine (Caputi et al. 2011). Rotundone concentration in Shiraz wine has been found in levels varying between 29 ng/L to 145 ng/L (Wood et al. 2008). While the concentration of rotundone in Shiraz is relatively low compared to other species that synthesise rotundone, it is an important contributor to wine character. Rotundone is extremely potent, with a threshold of detection as low as 8 ng/L in water and 16 ng/L in red wine (Wood et al. 2008). This gives the wine a 'spicy' or 'peppery' character.

Tartaric acid biosynthesis in *Vitaceae*

Early efforts to elucidate the biosynthetic pathway of tartaric acid in grapes identified L-ascorbic acid (Asc) as a precursor. Saito and Kasai (1969) fed young grape berries with radiolabelled L-ascorbic acid-1-^{14}C for 24 hours and discovered 72% of the radioactivity was found in TA. Combined with previous results which showed radioactivity from Asc-6-^{14}C was not incorporated into TA (Loewus & Stafford 1958), this gave evidence that TA was formed from the C_1-C_4 moiety of Asc (Figure 1-3). To confirm this, immature grape berries and leaves were fed with L-ascorbic acid-4-^{14}C. 60% of soluble ^{14}C was found in TA, and of this virtually all of the ^{14}C was present in the terminal carboxyl groups (Williams & Loewus 1978). This indicates that the C_1-C_4 moiety of Asc is directly converted into TA.

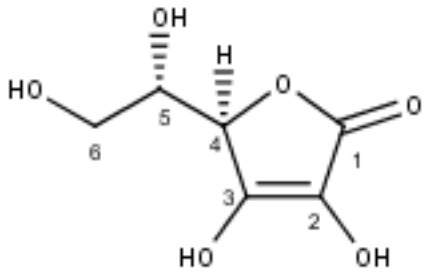


Figure 1-3 L-ascorbic acid

To investigate intermediates in the TA synthesis pathway, Saito and Kasai (1982) performed a time course experiment where slices of immature grape berry were fed with L-ascorbic acid-1-^{14}C and the metabolites analysed after 0-5 hours. Based on ^{14}C incorporation, L-idonic acid, L-idono-γ-lactone and 2-keto-L-idonic acid (2-keto-L-gulonic acid) were identified as metabolic products of ascorbic acid, but there was insufficient evidence to conclude their position in the pathway from ascorbic acid to tartaric acid. A fourth, unknown compound, later identified as 5-keto-L-idonic acid (5-keto-D-gluconic acid), was also observed to contain ^{14}C before the accumulation of labelled tartaric acid. Subsequent experiments by Saito and Kasai (1984) introduced specifically labelled L-idonic acid, L-idono-γ-lactone, 2-keto-L-gulonic acid (2KLG) and 5-keto-D-gluconic acid (5KDG) into grape slices and tissue to analyse the incorporation of ^{14}C into tartaric acid and the involvement of each metabolite in the tartaric acid biosynthesis pathway. Based on the incorporation of a large amount of label (86-90%) from the substrates into tartaric acid, Saito and Kasai (1984) proposed a pathway from ascorbic acid to 2KLG, then L-idonic acid, then 5KDG and, via unknown intermediate(s), to tartaric acid. This pathway was supported by further studies (DeBolt, Cook & Ford 2006; Malipiero, Ruffner & Rast 1987). Incorporation of radiolabel from 5-keto-1-^{14}C-gluconic acid but not from 5-keto-6-

^{14}C-gluconic acid indicated a C4/C5 cleavage of 5-keto-D-gluconic acid. The four-carbon product of this cleavage is L-threo-tetruronate (tartaric acid semialdehyde) which is oxidised to tartaric acid (Saito 1992). The two carbon product is likely glycolaldehyde, which is recycled back into the hexose pool (Saito & Loewus 1979). Kinetic studies suggest the rate limiting step of the pathway is the oxidation of L-idonate to 5KDG (Malipiero, Ruffner & Rast 1987) (Figure 1-4).

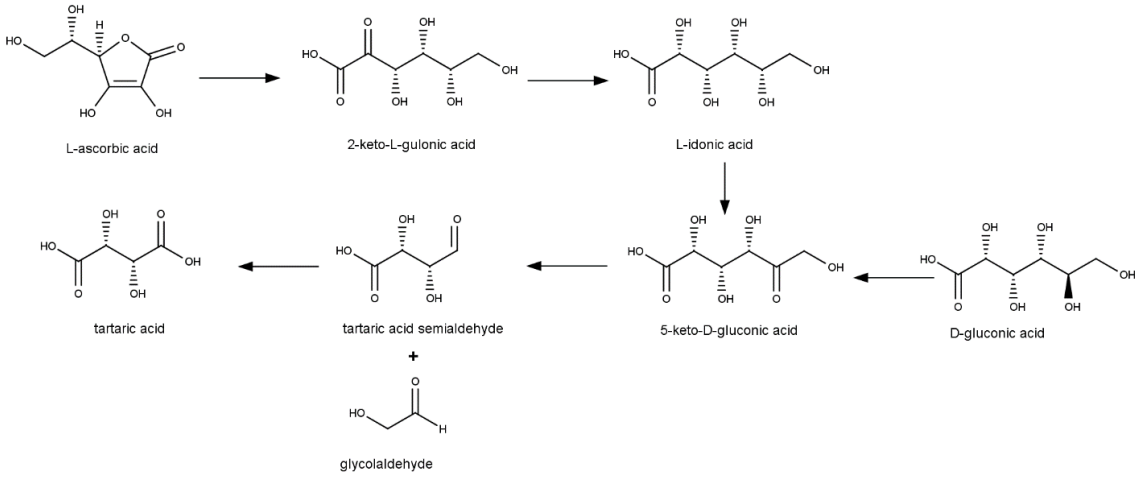


Figure 1-4 The primary pathway of tartaric acid biosynthesis in Vitaceae

As well as the primary pathway of TA biosynthesis from an Asc precursor, a secondary pathway from D-gluconic acid is also present in Vitaceae. The primary pathway accounts for up to 99% of TA synthesis in the light, and between 68-74% of TA synthesis in the dark (Saito & Loewus 1989). The secondary pathway utilises the C_1-C_4 backbone of 6-carbon glucose and proceeds via D-gluconate to 5KDG and follows the subsequent pathway shown in Figure 1-4 (Saito & Loewus 1989).

L-Ascorbic acid can also act as a precursor for oxalic acid. When grape berries were fed with L-ascorbic acid-1-^{14}C, 21% of radiolabel was observed in oxalic acid, and 53% of radiolabel was found in tartaric acid (DeBolt et al. 2004). This indicates both C2/C3 cleavage and C4/C5 cleavage of ascorbic acid occurs in grapevine.

Enzymes involved in the biosynthesis of Tartaric Acid in *Vitaceae*

The stereospecificity of the metabolic intermediates suggests each step is likely to be enzyme catalysed. Selective oxidation of 5KDG to L-TA can occur in acidic and alkaline conditions with a vanadate catalyst (Matzerath et al. 1995), but ^{18}O labelling experiments suggest the C4/C5 cleavage of 5KDG is likely to be catalysed by a hydrolase (Saito, Ohmoto & Kuriha 1997). Furthermore, TA is produced as a by-product of 5KDG fermentation in *E.coli* in the absence of vanadium, so this conversion is likely to be enzyme-catalysed (Salusjärvi et al. 2004). Several *E. coli* transketolases and commercial

yeast transketolases accept 5KDG as a substrate. Salusjärvi et al. (2004) propose 5KDG is cleaved by a ketolase to TA semialdehyde which is oxidised by a dehydrogenase. This mechanism is consistent with the observation that ^{18}O is incorporated into TA from water (Saito, Ohmoto & Kuriha 1997), but the corresponding enzymes in *E. coli* have not been identified. Nevertheless, this evidence supports enzyme catalysis of each step in the TA synthesis pathway in grapevine.

To date, very few of the enzymes in the pathway have been identified. An L-idonate dehydrogenase (L-IdnDH) gene was identified by comparing the transcript abundance of TA biosynthesis candidate genes in TA accumulating and non-accumulating species respectively. The L-IdnDH gene was not present in a non-TA-accumulating species, but was expressed in TA-accumulating grape species, and its gene expression pattern was consistent with TA synthesis in developing berries (DeBolt, Cook & Ford 2006). The candidate gene was found to share 31% amino acid identity with an *E. coli* oxidoreductase that catalyses the interconversion of L-idonate and 5-keto-D-gluconate. *In vitro* assays confirmed the enzyme was able to specifically catalyse the interconversion of L-idonate and 5-keto-gluconate using NAD^+/NADH as a cofactor (step 3 in Figure 1-4). Kinetic analysis gave a K_m of 2.2 mM for L-idonate in the forwards reaction, and 12.5 mM for 5KDG in the reverse direction. This provides *in vitro* and, through analysis of a non-tartaric acid accumulating species, *in vivo* evidence the L-IdnDH catalyses the rate-limiting step of tartaric acid biosynthesis (DeBolt, Cook & Ford 2006). Subcellular localisation studies using antibodies against L-IdnDH showed L-IdnDH is located mainly in the cytoplasm, with a small portion in the cell wall, secondary cell wall and chloroplasts in developing berries (21 days after full bloom) (Wen et al. 2010). After 60 days, L-IdnDH localises to the cytoplasm and vacuole, which suggests tartaric acid synthesis occurs in the cytoplasm, and possibly the vacuole, although vacuolar localisation may be due to protein degradation (Wen et al. 2010). Further work has supported the role of L-IdnDH in tartaric acid synthesis. CRISPR/Cas9-mediated disruption of an L-IdnDH gene in Chardonnay led to a decrease in TA content of the transgenic cell mass (Ren et al. 2016). Sequencing of the grapevine genome led to the identification of a further three isoforms of L-IdnDH, one of which is likely to be a sorbitol dehydrogenase (Jia et al. 2015). Therefore, there may be several, closely related, isoforms of each enzyme involved in TA synthesis.

Since the sequencing of the grape genome, further bioinformatic approaches have been utilised to identify putative enzymes in the tartaric acid synthesis pathway. Burbidge (2011) found a number of 2KLG reductase candidates through a sequence homology search with an *Escherichia coli* enzyme capable of catalysing 2KLG to L-idonate (Yum et al.

1998). Three candidates were expressed for further analysis and one of these was found to have activity against 2KLG with NADH and NADPH, with a K_m of 4.7 mM for 2KLG using NADH as the cofactor. Although activity was highest against 2KLG, activity was also observed with D-gluconate and L-ascorbate substrates, although to a lower extent (Burbidge 2011). While there is evidence to support the annotation of this enzyme as a 2KLG reductase, the K_m of this enzyme is fairly high, which is inconsistent with the low levels of TA biosynthetic intermediates present in immature grapes (Saito & Loewus 1979). Further evidence, including *in planta* experiments, is required to confirm its role in TA synthesis.

A limiting factor in the study of tartaric acid biosynthesis is the limited commercial availability of L-idonate. Cheap and efficient production of L-idonate would greatly benefit further research in this field.

Fungal 2-keto-L-gulonate reductases

Comparison between grapevine enzymes and enzymes from other species with known activity can help identify potential catalysts in the TA biosynthesis pathway. Recently, two fungal 2-keto-L-gulonate reductases (2KGR) capable of interconverting 2KLG and L-idonate (step 2 in Figure 1-4) were identified in *Aspergillus niger* (Kuivanen, Arvas & Richard 2017; Kuivanen et al. 2016). These 2KGRs are involved in the catabolism of D-glucuronate and show unique enzyme activity. Previously, Yum et al. (1998) identified an *E. coli* 2-ketoaldonate reductase with low specificity that can also catalyse the conversion of 2KLG to IA, but these fungal enzymes are the first substrate specific 2KGRs to be reported. The GluC protein uses NAD^+/NADH as a cofactor and has a K_m of 30 mM and 20 mM respectively for 2KLG and L-idonate (Kuivanen et al. 2016). The GluD protein uses $NADP^+$/NADPH as a cofactor with a Km 25.3 and 12.6 mM for 2KLG and L-idonate respectively (Kuivanen, Arvas & Richard 2017). The amino acid sequences of these 2KGRs can be compared to proteins present in grapevine using a Basic Local Alignment Search Tool (BLAST). Grapevine enzymes with high similarity, as well as similar conserved domains and cofactor binding sites, are likely to have the same 2KGR activity.

Sesquiterpene biosynthesis in *Vitaceae*

All terpenes are synthesised from 5-carbon building blocks of isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) (Eisenreich et al. 2004). IPP and DMAPP can be synthesised by two separate pathways in plants (Figure 1-5). The mevalonate pathway is common to all eukarya, and also present in archaea and some bacteria, and occurs in the cytoplasm (Eisenreich et al. 2004). The non-mevalonate pathway, or MEP pathway is present only in plants, bacteria and some archaea. In plants, this pathway

occurs in the plastids (Eisenreich et al. 2004). Radiolabelling studies have shown that both pathways contribute to sesquiterpene synthesis in grapevine, indicating transport of IPP/DMAPP from plastid to cytosol (May, Lange & Wüst 2013). In the cytosol, two IPP units and one DMAPP unit combine to form the 15-carbon sesquiterpene precursor farnesyl pyrophosphate (FPP). The sesquiterpene backbone is created through modification of FPP by sesquiterpene synthases (STS). The sesquiterpene can be further functionalised or modified to a sesquiterpenoid by regio- and stereo-selective cytochrome P450 monooxygenases (CYPs) or various transferases (Nguyen, MacNevin & Ro 2012; O'Maille et al. 2008).

Rotundone is synthesised via a sesquiterpene synthase that converts FPP to α-guaiene in a single step, followed by CYP oxidation at C2 (Figure 1-5) (Drew et al. 2016; Kumeta & Ito 2010; Takase et al. 2016). The grapevine sesquiterpene synthase that converts FPP to α-guaiene, VvGuaS, was identified as an allelic variant of the *VvTPS24* which encodes VvPNSeInt, a selinene-type sesquiterpenes synthase (Drew et al. 2016). Two key polymorphisms close to the active site were implicated in the difference in products between the two enzymes. The products of the VvGuaS enzyme are α-guaiene (44%), δ-guaiene 35% and a number of minor products (Drew et al. 2016). α-Guaiene is oxidised at the C2 position by VvSTO2, an α-guaiene 2-oxidase, which is a cytochrome P450 in the CYP71BE subfamily (Takase et al. 2016). VvSTO2 is highly specific for α-guaiene and (+)-valencene, although the latter is not present in grapevine in significant levels, so VvSTO2 acts solely as an α-guaiene oxidase in grapevine (Takase et al. 2016). The major product of VvSTO2 is rotundone, with two minor products which are probably 2R- and 2S-rotundol, likely intermediates in the oxidation of α-guaiene to rotundol (Takase et al. 2016). Auto-oxidation of α-guaiene to rotundone has also been shown to occur spontaneously in air over a number of days to weeks (Huang et al. 2014). This demonstrates that exposure to air during winemaking may also play a role in rotundone synthesis. Therefore, rotundone is a product of both enzymatic control and winemaking techniques.

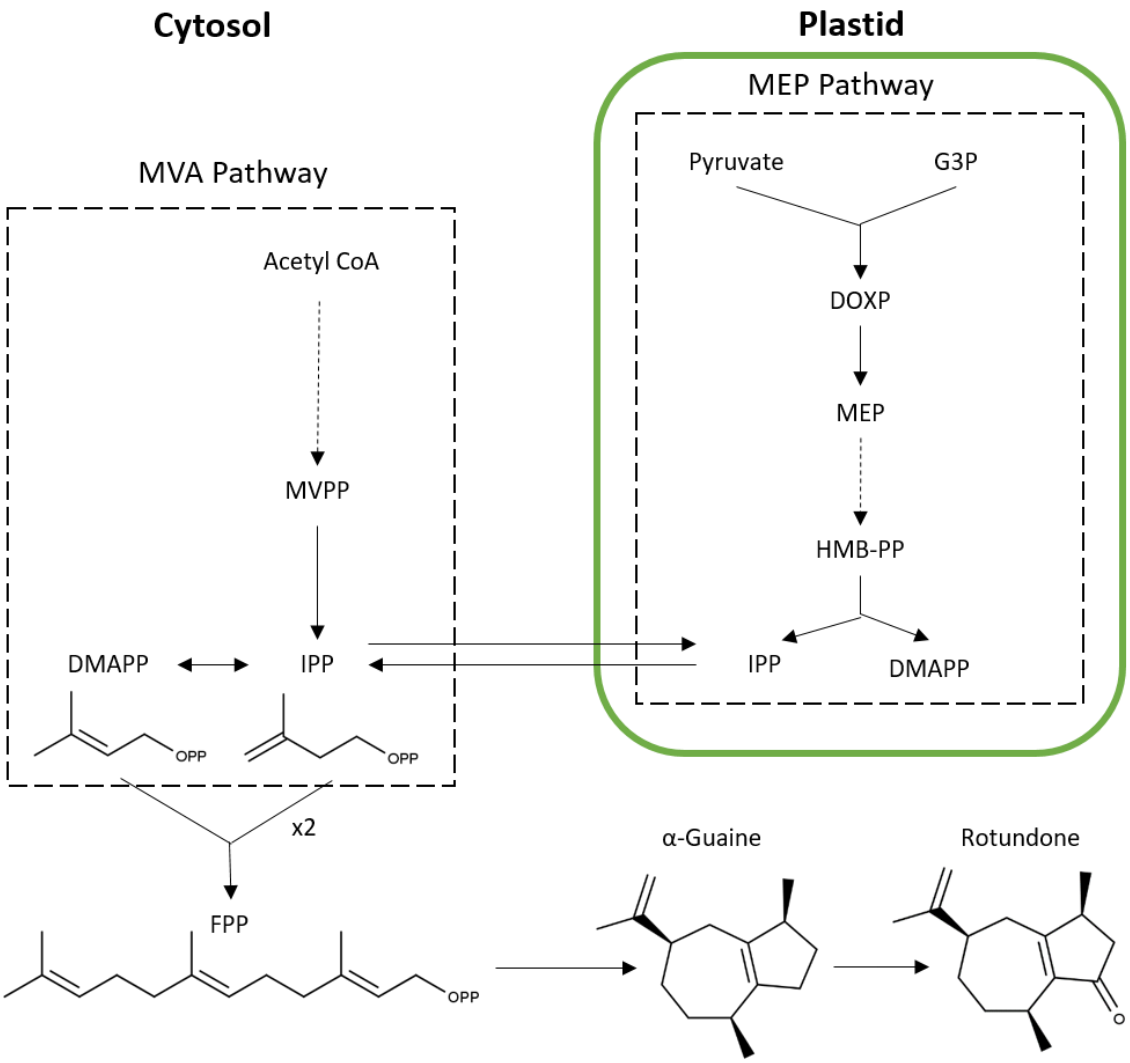


Figure 1-5 Rotundone synthesis in Vitaceae*. Adapted from Martin et al. (2012). The following metabolites are shown: G3P glyceraldehyde 3-phosphate, DOXP 1-Deoxy-D-xylulose 5-phosphate, MEP 2-C-methyl-D-erythritol 4-phosphate, HMB-PP (E)-4-hydroxy-3-methyl-but- 2-enyl pyrophosphate, IPP isopentenyl pyrophosphate, DMAPP dimethylallyl pyrophosphate, MVPP mevalonate-5-pyrophosphate, FPP farnesyl pyrophosphate*

Microbial Biosynthesis

Microbial biosynthesis is an efficient and versatile method of obtaining useful products. It has many advantages over traditional chemical synthesis. Firstly, microbial biosynthesis circumvents the need for toxic organic solvents and heavy metals, reducing the environmental impact of synthesis. Secondly, synthesis is usually high-yielding as it is stereoselective and avoids the formation of by-products (Du, Shao & Zhao 2011). Finally, synthesis can be achieved from cheap feedstocks such as glucose, glycerol, or xylose (Markham & Alper 2015; Mazumdar, Clomburg & Gonzalez 2010). Microbial synthesis has been used in the production of a wide range of useful chemicals including pharmaceuticals, biofuels, polymers, and vitamins (Abdel-Rahman et al. 2011; Bremus et al. 2006; Buchholz & Seibel 2008; Jeandet et al. 2013).

There are three main categories of microbial biosynthesis: fermentation, bioconversion with purified enzymes, and whole-cell biocatalysis. Fermentation requires the engineering of metabolic pathways to produce the desired product from a cheap, readily available feedstock. Fermentation generally requires low-cost raw materials, but may be low-yielding. Additionally, since growth rate is affected by product synthesis, costly supplementation may be required (Bückle-Vallant et al. 2014). Alternatively, bioconversion from substrate to product can be achieved by purified soluble or immobilised enzymes. Generally yields are higher and purification is easier, especially with immobilised enzymes. However, enzymatic bioconversion is expensive due to the cost of protein purification and immobilisation, and the need to supply cofactors (Song et al. 2016). Furthermore, the price is dependent on the cost of the substrate. Whole-cell biocatalysis is a third option that uses metabolically dormant cells containing the desired enzyme activity. This method is estimated to be ten times cheaper than using purified enzyme catalysts (Tufvesson et al. 2011). Another advantage is that cofactors are supplied by the cell and can, to some extent, be regenerated by existing cellular mechanisms. If supplementary cofactor regeneration is required it is usually straightforward to implement (Wachtmeister & Rother 2016). The cellular structure also provides protection for the catalysts, allowing harsher conditions and non-aqueous media to be used in order to maximise product yield. While the cost is still dependent on the price of the substrate, whole-cell biocatalysis is generally an efficient and inexpensive method of producing target metabolites.

Microbial biosynthesis can be used to achieve efficient stereoselective synthesis of compounds. Fermentation has been used to produce L-(+)-lactic acid with a greater than 99.5% enantiomeric excess from xylose, an abundant sugar (Abdel-Rahman et al. 2011). Maintenance of stereochemistry is one of the biggest challenges in the chemical synthesis of L-idonate and sesquiterpenes due to their high number of stereocentres. Additionally, the separation of the product from a mixture of enantiomers poses further challenges. Using a biosynthetic tactic to produce L-idonate and rotundone with high optical purity will avoid this problem.

Microbial Biosynthesis of Keto Acids

While the microbial synthesis of L-idonate is novel, the technique has been used to produce related compounds. As well as being an intermediate in the TA biosynthesis pathway in grapevine, 2KLG is also a precursor for the commercial synthesis of ascorbic acid. 2KLG has been produced by fermentation from glucose in *Erwinia citreus* with a conversion efficiency of 49.4% (Grindley et al. 1988). *E. citreus* is naturally able to oxidise

D-glucose to 2,5-diketo-D-gluconate (2,5-DKG) via D-gluconate and 2-keto-D-gluconate. By expressing a gene for a 2,5-DKG reductase from *Corynebacterium* sp. in *E. citreus*, the organism was able to produce 2KLG from glucose. Yields were improved for selecting for mutants that were not able to utilise 2,5-DKG or 2KLG as a sole carbon source (Grindley et al. 1988). More recently, 2KLG production from 2,5-DKG has been achieved by coupling 2,5-DKG reductase to a NADPH-regenerating glucose dehydrogenase (Kaswurm et al. 2013). Modelling was used to optimise process parameters and double the yield of 2KLG (Kaswurm et al. 2013). This is an example of *in situ* cofactor regeneration using glucose as a sacrificial substrate. It also demonstrates the usefulness of mathematical modelling as a tool for designing bio-catalytic systems. Many other keto acids, including α-ketoglutarate and 2,5-DKG, have been produced by microbial biosynthesis, sometimes with yields over 90% (Song et al. 2016). Therefore, microbial biosynthesis represents a viable method of producing L-idonate cheaply and efficiently. Whole-cell biocatalysis is ideal because several 2KLG reductases have already been identified (Burbidge 2011; Kuivanen, Arvas & Richard 2017; Kuivanen et al. 2016; Yum et al. 1998) and the substrate, 2KLG, is inexpensive in bulk quantities.

The 2KGRs that catalyse the interconversion of 2KLG and L-idonate utilise either NADH or NADPH as a cofactor. The cofactor is required in stoichiometric quantities, so *in situ* regeneration of $NAD(P)^+$ to NAD(P)H is preferable to adding large amounts of cofactor. Several regeneration methods are possible. Electrochemical regeneration using a cathode is environmentally friendly because no co-substrate is required and no by-products are produced, but requires specialist equipment, and current technology is not efficient enough to be economically feasible (Kara, Schrittwieser & Hollmann 2013). The use of H_2 as a reducing agent, catalysed by hydrogenases has been explored but is limited by the oxygen-sensitivity of hydrogenase enzymes and poor water solubility of H_2 (Kara, Schrittwieser & Hollmann 2013). The use of a sacrificial co-substrate is the most widely used method. This requires a catalyst to oxidise the co-substrate which produces a by-product that must be separated from the reaction mixture. Common co-substrates include formates, phosphites, alcohols and glucose, with a corresponding dehydrogenase to catalyse their oxidation (Kara, Schrittwieser & Hollmann 2013). The choice of co-substrate depends on the cofactor required (NADH or NADPH) and the selectivity of the reductase. It is important that the co-substrate cannot act as a substrate for the reductase that produces the target product as this would decrease the yield and efficiency of the process.

Microbial Biosynthesis of Sesquiterpenes

Starting materials for sesquiterpene synthesis such as FPP are often expensive, making bioconversion uneconomical. Instead, a strain of yeast (*Saccharomyces cerevisiae* EPY300) that produces a high level of FPP was engineered for the fermentation of sesquiterpenes from a simple sugar feedstock (Ro et al. 2006). Sesquiterpene biosynthesis can be achieved by expression of at least one sesquiterpene synthase (STS) and, optionally, a sesquiterpene-modifying P450.

An example of the use of this yeast strain is for the production of capsidiol, a hydroxylated sesquiterpenoid that is an important plant defence molecule. EPY300 was transformed with a pESC-Leu2d plasmid containing a 5-*epi*-aristolochene STS with a 5-*epi*-aristolochene-hydroxylating P450 and its corresponding redox partner. The yeast was cultured on media containing galactose and methionine. The strain is engineered so that in the presence of galactose and methionine, the mevalonate pathway is upregulated and FPP production increases. Capsidiol production was analysed by GC-MS, with a yield of approximately 250 μg/mL (Nguyen, MacNevin & Ro 2012).

Grapevine sesquiterpenoids may be produced in a similar way. By expressing grapevine sesquiterpene synthases with a promiscuous P450 in EPY300 yeast, sesquiterpenoid synthesis could be achieved from a glucose/galactose feedstock.

Research Objectives

Tartaric acid is an important chemical in the winemaking process as it prevents microbial spoilage and oxidation during fermentation and contributes to the organoleptic properties of the wine. However, its biosynthesis is poorly understood. Research undertaken for this project will identify genes that may be involved in the TA biosynthesis pathway in grapevine.

Sesquiterpenes are important contributors to the flavour and aroma of wine. In particular, rotundone is a key aromatic compound in Shiraz wine and has been shown to have a significant effect on wine character. This project will characterise enzymes involved in the synthesis of sesquiterpenoids, including rotundone, and establish methods to synthesise these compounds cheaply and efficiently.

The specific aims of this study are to:

- Identify putative 2-keto-L-gulonate reductase enzymes involved in the TA biosynthesis pathway in grapevine
- Characterise the substrate specificity and enzyme kinetics of the above candidates

- Develop a microbial biosynthesis system capable of producing L-idonate from 2-keto-L-gulonate
- Develop a microbial biosynthesis system capable of producing grapevine sesquiterpenoids from simple sugar feedstocks
- Develop a microbial biosynthesis system to investigate the mechanism of rotundone formation

This will lead to an overall better understanding of how organic acids and aromatic molecules are synthesised in grapes. Potentially this will allow grape breeders in the future to better control and select for tartaric acid or rotundone production in grapevine and produce high-quality wines.

Chapter 2 Methods

Materials, Chemicals and Equipment

Substrates, reagents, growth media and buffers

Substrates, reagents and solvents were supplied by Sigma-Aldrich (USA), Bio-Rad (USA) or Roche (Germany) as required. 2-Keto-L-gulonate was purchased from Sigma-Aldrich and L-Idonate was purchased from Omicron Biochemicals. All media and buffers used are listed in Table 2-1 and Table 2-2 respectively. Antibiotics were added to LB as required. Ampicillin was added to a concentration of 100 mg μL^{-1}. Liquid media was sterilised by filtration or autoclaving. Solid media was prepared by the addition of 2% bacteriological agar and sterilised by autoclaving.

Table 2-1 Growth media components

Medium	Components (L^{-1})
LB	tryptone (10 g), yeast extract (5 g), NaCl (5 g)
SOC	tryptone (20 g), yeast extract (5 g), $MgCl_2$ (1 g), NaCl (0.5 g), KCl (0.2 g), glucose (2 g)
YPD	glucose (20 g), yeast extract (10 g), bacto peptone (20 g)
SC	Yeast Nitrogen Base without amino acids (1.7 g), ammonium sulfate (5 g), adenine (21 mg), 16 standard amino acids (without histidine, leucine, methionine and tryptophan) (85.6 mg each), leucine (173.4 mg), histidine (85.6 mg), methionine (85.6 mg), tryptophan (85.6 mg), uracil (85.6 mg). Amino acids omitted as appropriate.

Table 2-2 Buffer components

Buffer	Components
Buffer A	Tris 8.0 (20 mM), NaCl (0.5 M), 2-mercaptoethanol (BME) (2 mM), imidazole (10 mM)
Buffer B	Tris 8.0 (20 mM), NaCl (0.5 M), BME (2 mM), imidazole (250 mM)
Dialysis Buffer	Tris 8.0 (20 mM), NaCl (50 mM), BME (2 mM), Dithiothreitol (DTT) 1mM
Tris Buffer	Tris(hydroxymethyl)aminomethane 50 mM (pH adjusted by addition of HCl)
TE Buffer pH 8.0	Tris 10 mM, Ethylenediaminetetraacetic acid (EDTA) 1mM (pH adjusted by addition of HCl)
PBS pH 7.4	NaCl (137 mM), KCl (2.7 mM), Na_2HPO_4 (10 mM), KH_2PO_4 (1.8 mM)
HEPES Buffer	HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) 100 mM (pH adjusted by addition of NaOH)
SDS Loading Buffer	Tris pH 6.5 (100 mM), sodium dodecyl sulphate (SDS) (4% v/v), glycerol (20% v/v), bromophenol blue (0.2% w/v), BME (5% v/v).

Equipment

High Performance Liquid Chromatography (HPLC) was performed on an Agilent 1100 series pump equipped with an auto injector connected to a Phenomenex ROA-Organic Acid Column (300 x 7.8 mm).

Gas Chromatography-Mass Spectrometry (GC-MS) for sesquiterpenoid analysis was performed on an Agilent 7890B GC using a 5977B MS detector and a DB-5 MS fused silica column (30 m x 0.25 mm, 0.25 µm).

GC-MS analysis for α-guaiene analysis was performed on an Agilent 6890 using an Agilent 5973 N mass spectrometer and a DB-5 MS fused silica column (30 m x 0.25 mm, 0.25 µm).

Bioinformatics

The genome sequences for *Vitis vinifera* were accessed from the National Center for Biotechnology Information (NCBI) database. Using this database, candidates were identified by protein Basic Local Alignment Search Tools (BLAST) (https://blast.ncbi.nlm.nih.gov/). Protein sequence analysis and classification was performed using InterPro (https://www.ebi.ac.uk/interpro/). Expression patterns were analysed using Expression Atlas (https://www.ebi.ac.uk/gxa/home). Subcellular localisation and transmembrane structure analysis was performed using CBS Prediction

Services (http://www.cbs.dtu.dk/services/). Local secondary structure prediction was performed using Spider 2 (http://sparks-lab.org/server/SPIDER2/).

Enzyme Expression, Purification and Characterisation

Plasmids and Strains

E. coli strain DH5α (New England Biolabs) was used for cloning and plasmid construction. Plasmids used in this study are listed in Table 2-3. Putative 2KGR genes were codon-optimised for *E. coli*, synthesised and cloned into pET-11a vectors using *NdeI* and *BamHI* restriction sites by GenScript (US). For Gateway® cloning, the 2KGR genes were PCR amplified using the primers in Table 2-4. PCR was performed in a 50 µL reaction with 2 units VELOCITY DNA Polymerase (Bioline), 1 x HiFi Reaction Buffer (Bioline), 10 mM dNTP, 10 µM forward and reverse primers, and 2 ng template plasmid. PCR was performed on a BioRad T100 thermal cycler with a standard cycle of: 98 °C for 2 minutes, 30 cycles of 98 °C for 30 seconds, 55 °C for 30 seconds and 72 °C for 35 seconds, followed by a final extension at 72 °C for 4 minutes. PCR products were visualised with gel electrophoresis. The donor vector was pDONR221 and the destination vector was pDEST57. For the BP reaction, a 5 µL reaction consisting of 50 ng PCR product, 500 ng donor vector, 2 µL TE buffer pH 8.0 and 1x BP Clonase Enzyme Mix (Invitrogen) was incubated at room temperature for 2 hours followed by incubation with 0.5 µL Protein Kinase K (Invitrogen) at 37 °C for 10 minutes. For the LR reaction, a 5 µL reaction mixture containing 100 ng each of donor and destination vector, 2 µL TE buffer and 1x LR Clonase Enzyme Mix (Invitrogen) was incubated at 25 °C for 1.5 hours followed by incubation with 1 µL Protein Kinase K (Invitrogen) at 37 °C for 10 minutes.

Table 2-3 Plasmids used for protein expression and purification

Plasmid	Description	Source
pET-11a	Expression vector, Amp^R, P_{T7}, 3851 ori	Novagen
pET-14b	Expression vector, Amp^R, P_{T7}, 2845 ori	Novagen
pDONR-221	Gateway® donor vector, Kan^R, P_{T7}, pUC ori	Novagen
pDEST-57	Gateway® destination vector for expressing proteins tagged at the N-terminus with NusA, Amp^R, P_{T7}, pBR322 ori	Novagen
pET-11a-2KGR.1	Candidate XP_002284520.1 inserted into NdeI/BamHI sites of pET-11a	This study
pET-11a-2KGR.2	Candidate XP_002281980.1 inserted into NdeI/BamHI sites of pET-11a	This study
pDEST-57-2KGR.1	Candidate XP_002284520.1 inserted into *attR* sites of pDEST-57	This study
pDEST-57-2KGR.2	Candidate XP_002281980.1 inserted into *attR* sites of pDEST-57	This study
pET-14b-Vv2KGR.0	Gene *Vv2KGR.0* inserted into NdeI/BamHI sites of pET-14b	Jia (2015)

Table 2-4 Primers used for Gateway® Cloning

Primer	Sequence
Candidate 1 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTY YGAAAACCTGTATTTTCAGGGAATGGAAAA CATCTGTGTCTTATTGAC
Candidate 1 R	GGGGACCACTTTGTACAAGAAAGCTGGGTY TTAAATTACTGGTGTTAACACGGGT
Candidate 2 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTY YGAAAACCTGTATTTTCAGGGAATGACTGC TATGGATGAGTTGC
Candidate 2 R	GGGGACCACTTTGTACAAGAAAGCTGGGTY TTAGTCAAGATTAATGGGACTAAGTAATG

Sequencing was performed by the Australian Genome Research Facility. Purified DNA and primers were supplied according to AGRF guidelines.

E. coli strain BL21 (DE3) (New England Biolabs) was used for all protein expression and feeding experiments.

Transformation

E. coli strains DH5α and BL21 (DE3) were transformed according to the following protocol. Approximately 50 ng plasmid DNA was added to 50 µL competent cells. After 30 minutes on ice, the cells were heat-shocked at 42 °C for 10 seconds and placed back on ice for 5 minutes. Sterile SOC medium (300 µL) was added and the cells incubated at 37 °C with shaking for 1 hour. The cells were then plated on LB agar with appropriate antibiotic and incubated at 37 °C overnight.

Protein Expression and Purification

BL21(DE3) was transformed with pDONR57 derived vectors. Single colonies were picked and inoculated in 50 mL LB at 37°C overnight on a rotary shaker. 20 mL of pre-inoculum was transferred to 4 L LB and grown in a baffled flask at 37°C until the OD_{600} reached 0.6-0.8. Protein expression was induced by the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and cultures incubated at 16 °C overnight. Cells were harvested by centrifugation (5000 x *g*, 20 minutes) and lysed using a French press. The lysis solution was centrifuged at 4 °C 10,000 x *g* for 60 minutes. The supernatant was loaded onto a BioScale Mini Profinity IMAC 5 mL Cartridge (BioRad) using a LKB pump (Pharmacia). The protein was eluted into 45 mL Buffer B and dialysed at 4 °C overnight in Dialysis Buffer. The enzyme solution was flash-frozen with liquid nitrogen and stored at -80 °C for enzyme assays.

Protein purity was confirmed by visualisation using Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE). 10 µL samples were added to SDS loading buffer and boiled for 5 minutes. Samples were then loaded onto SurePAGE Bis-Tris 10 x 8 (GenScript) gels in Tris-MOPS-SDS running buffer (GenScript) and electrophoresed for 30 minutes at 200 V. Gels were stained for 10 minutes using Coomassie blue stain (Coomassie blue stain (40% (v/v) ethanol, 7% glacial acetic acid, 0.1% (w/v) Coomassie blue). Excess Coomassie blue stain was removed and gels were immersed in Coomassie destain (10% (v/v) glacial acetic acid, 40% methanol (v/v)) overnight.

Proteins were quantified as required using a NanoDrop 2000 spectrophotometer (Thermo Scientific, USA) using an extinction coefficient of 45000 cm^{-1} M^{-1}.

In vitro Enzyme Assays

Enzyme activity was analysed by monitoring the oxidation or reduction of the cofactor NADH at 340 nm using a 96-well plate (Costar) with a FLUOStar OPTIMA plate reader (BMG Labtech). Activity was tested over varying pH using both HEPES and Tris-HCl buffers (100 mM) to ensure changes in activity were due to pH and not buffer composition. A typical reaction consisted of 186 µL buffer, 2 µL enzyme (0.5 µg final), 2 µL coenzyme solution (0.25 mM final). The reaction was initiated by addition of 10 µL substrate solution to a total reaction volume of 200 µL. The reaction was monitored for 40 minutes at 37 °C with absorbance readings taken every 90 seconds. Negative controls included no enzyme, no substrate and no cofactor controls. Data was analysed using GraphPad Prism version 8.0.0.

The rate of enzyme activity was calculated using a modified Beer-Lambert formula:

$$\frac{slope}{\varepsilon l} \times \frac{1000}{mg/mL\ Protein} = \mu moles/min/mg\ protein$$

Where l is the path length and ε is the extinction coefficient of NADH. The extinction coefficient of NADH used was 6200 L mol^{-1} cm^{-1} at 340 nm.

Whole Cell Biocatalysis

Whole cell biocatalysis experiments were conducted using *E. coli* BL21 (DE3) and transformants. Strains were grown at 37 °C overnight in 1 mL LB. 500µL of overnight culture was transferred to 50 mL fresh LB and grown at 37 °C to an OD_{600} of 0.6. Protein expression was induced by the addition of 0.5 mM IPTG, the temperature was lowered to 16 °C and the cells were grown for a further 18 hours. Cells were collected by centrifugation at 5000g for 10 minutes, and washed twice with 50 mM Tris-HCl buffer (pH

7.5) or PBS (pH 7.4). Cells (0.2 g wet weight) were resuspended in 20 mL 50 mM Tris-HCl buffer (pH 7.5) or PBS and 0.25M NaCl (pH 7.4) with 10 mM 2-keto-L-gulonate. The biotransformation reaction was performed at 30 °C for 24 h and monitored by taking 1 mL aliquots for HPLC analysis.

HPLC Analysis

Samples (1 mL) were centrifuged at 14,800 rpm for 10 min and the supernatant filtered through a 0.45 μm filter and used for analysis. The concentration of 2-keto-L-gulonate was determined by HPLC using a ROA-Organic Acid Column (300 x 7.8 mm, Phenomenex) with 5.0 mM H_2SO_4 as eluent and a flow rate of 0.5 mL min^{-1}. The column was maintained at 55 °C. Peaks were detected at 210 nm.

Production and Analysis of Sesquiterpenoid Compounds

Plasmid Construction

Yeast plasmids used were derived from pESC-LEU2d. All plasmids used for production of sesquiterpenoid compounds are outlined in Table 2-5. All sesquiterpene synthases were cloned from *Vitis vinifera* cv. Shiraz. For α-guaiene production, TPS24 was codon optimised and gene synthesised by Thermo Fisher.

Table 2-5 Plasmids used for production and analysis of sesquiterpenoid compounds

Plasmid	Description	Source
pESC-LEU2d	Expression vector containing GAL1-GAL10 divergent promotor, 2μ origin, LEU2 marker	Agilent
pESC-LEU2d-TPS24	*TPS24* gene inserted into HindIII/NheI sites of pESC-LEU2d	This study
pESC-LEU2d-BM3WT	Wild-type $P450_{BM3}$ gene inserted into XhoI/BamHI sites of pESC-LEU2d	This study
pESC-LEU2d-BM3Mut	Mutant $P450_{BM3}$ gene inserted into XhoI/BamHI sites of pESC-LEU2d	This study
pESC-LEU2d-TPS07-BM3WT	*TPS07* gene inserted into NotI/PacI sites of pESC-LEU2d-BM3WT	This study
pESC-LEU2d-TPS07-BM3Mut	*TPS07* gene inserted into NotI/PacI sites of pESC-LEU2d-BM3Mut	This study
pESC-LEU2d-TPS24-BM3WT	*TPS24* gene inserted into NotI/PacI sites of pESC-LEU2d-BM3WT	This study
pESC-LEU2d-TPS24-BM3Mut	*TPS24* gene inserted into NotI/PacI sites of pESC-LEU2d-BM3Mut	This study
pESC-LEU2d-TPS26-BM3WT	*TPS26* gene inserted into NotI/PacI sites of pESC-LEU2d-BM3WT	This study
pESC-LEU2d-TPS26-BM3Mut	*TPS26* gene inserted into NotI/PacI sites of pESC-LEU2d-BM3Mut	This study
pESC-LEU2d-TPS27-BM3WT	*TPS27* gene inserted into NotI/PacI sites of pESC-LEU2d-BM3WT	This study
pESC-LEU2d-TPS27-BM3Mut	*TPS27* gene inserted into NotI/PacI sites of pESC-LEU2d-BM3Mut	This study
pESC-LEU2d-TPSY1-BM3WT	*TPSY1* gene inserted into NotI/PacI sites of pESC-LEU2d-BM3WT	This study
pESC-LEU2d-TPSY1-BM3Mut	*TPSY1* gene inserted into NotI/PacI sites of pESC-LEU2d-BM3Mut	This study
pESC-LEU2d-TPSY2-BM3WT	*TPSY2* gene inserted into NotI/PacI sites of pESC-LEU2d-BM3WT	This study
pESC-LEU2d-TPSY2-BM3Mut	*TPSY2* gene inserted into NotI/PacI sites of pESC-LEU2d-BM3Mut	This study

Yeast Strains and Cultivation

Sesquiterpenoids were microbially synthesised by the yeast strain EPY300 (S288C, MATα his3Δ1 leu2Δ0 P_{GAL1}–tHMG1::δ1 P_{GAL1}–upc2-1::δ2 erg9::P_{MET3}–ERG9::HIS3 P_{GAL1}–ERG20::δ3 P_{GAL1}–tHMG1::δ4). EPY300 has been metabolically engineered to overproduce the sesquiterpene precursor farnesyl pyrophosphate (FPP) (Nguyen, MacNevin & Ro 2012). EPY300 was freshly streaked out from glycerol stocks and revived on YPD media. It was selected for on SC-His-Met media. pESC-Leu transformants were selected for on SC-His-Met-Leu media. Yeast strains were grown in appropriate SC media with 2% glucose as the carbon source. To induce sesquiterpene synthesis, transformants were grown in SC

media with 1.8% galactose, 2% glucose and 1 mM methionine. All yeast liquid cultures were grown at 28° C, 140 rpm and all plates grown at 30° C.

Yeast Transformation

The transformation protocol used was based on the LiAc/SS carrier DNA/PEG method (Gietz & Schiestl 2007). EPY300 was cultured overnight in 1 mL SC-His-Met 2% glucose at 28 °C. 500 µL of pre-inoculum was transferred to fresh SC-His-Met (25 mL) and grown at 28 °C for 4 hours. Cells were harvested by centrifugation and washed in 0.1 M Lithium Acetate (25 mL), then centrifuged again and resuspended in 0.4 mL 0.1 M Lithium Acetate. 100 ng of DNA were added to a transformation mixture consisting of the following: 50 µL EPY300 cells, 240 µL 50 % Polyethylene Glycol, 36 µL 1 M Lithium Acetate, 50 µL salmon sperm DNA (prepared by boiling 10 minutes then immediately chilling on ice for 5 minutes) and 29 µL sterile water. The transformation mixture was incubated at 30 °C for 30 minutes then 42 °C for 30 minutes. The cells were collected by centrifugation, washed with sterile saline (0.5 mL) and resuspended in 50 µL sterile water. The cell suspension was plated on appropriate selective agar and incubated at 30 °C for 3 days.

Metabolite Analysis

EPY300 – pESC LEU transformants were used for the production of sesquiterpenoids *in vivo*. A single colony was picked and inoculated in 5 mL appropriate SC media and grown overnight (15 – 18 hours) at 28 °C with shaking (140 rpm). 1 mL of pre-inoculum was transferred to 30 mL fresh SC media with 1.8% galactose, 0.2% glucose and 1 mM methionine. Cultures were incubated at 28 °C with shaking (140 rpm) for 3 days. The medium (30 mL) was extracted with ethyl acetate (5 mL) by vortexing for 20 seconds, followed by centrifugation to separate the organic and aqueous phases. The organic fraction was collected for GC-MS analysis. Samples (1 mL) were concentrated to 250 µL using a gentle flow of N_2 gas.

For analysis of products of $P450_{BM3,}$ cultures of EPY300 expressing $P450_{BM3}$ were cultured as above and the culture medium spiked with 0.5 µg/mL α-caryophyllene or β-caryophyllene. The medium (30 mL) was extracted with ethyl acetate (5 mL) as described above but samples were not concentrated before GC-MS analysis.

Sesquiterpene production was analysed by GC-MS using helium as carrier gas at a flow rate 1 mL/ min. The injector was in a splitless mode with an inlet temperature set at 300 °C. The GC oven temperature was initially 50 °C and held for 1 minute then ramped at 8 °C/min to 300 °C and held for 17.75 min. Mass fragments were detected by normal scanning over a range of m/z 33 to 500. Detection was carried out in positive ion mode.

Compounds produced were identified by comparison of retention times and mass spectra to those of authentic standards or by reference spectra in the National Institute of Standards and Technology library.

α-Guaiene production was analysed using GC-MS using helium as carrier gas at a flow rate 1 mL/ min. The injector was in splitless mode with an inlet temperature set at 250 °C. The GC oven temperature was initially 100 °C, then ramped at 20 °C/min to 250 °C, then ramped at 30 °C/min to 300 °C and held for 3 minutes.

Homology Modelling

Protein-protein Basic Local Alignment Search Tools (BLAST) were used to identify a protein in the Protein Data Bank (PDB) database with homology to *V. vinifera* P450s. The coordinates for Human Cytochrome P450 CYP17A1 were downloaded from the PDB (accession number 3RUK). Protein structure was modelled by satisfaction of spatial restraints using MODELLER 9.19 (Sali & Blundell 1993). Models were evaluated using VADAR (Volume, Area, Dihedral Angle Reporter) (Willard et al. 2003). Docking was performed using AutoDock 4.2.6 (Morris et al. 2009). The charge on the porphyrin iron was manually set to +0.400 *e*, corresponding to Iron(II), and the charges on the ligating nitrogens set to -0.348 *e* to compensate. The grid box (60 x 60 x 60 points with a grid spacing of 0.375 Å) was centred around the haem iron. A total of 150 docking runs were performed using a Lamarckian Genetic Algorithm with default parameters. Resulting conformations were ranked by binding energy and evaluated by determining distance between hydroxylation sites and the haem iron. All figures were generated using PyMOL (DeLano Scientific LLC, San Francisco, CA, USA). Multiple sequence alignment was performed by ClustalW and visualised using BioEdit.

Chapter 3 Identification and Characterisation of Potential Grapevine 2-Ketogulonate Reductases

Introduction

Tartaric acid (TA) is found in many species of plants, but only accumulates in a few species, including grapevine (*Vitis vinifera)* (Stafford 1959). During grape berry development, TA levels increase rapidly during the first four weeks after flowering and then remain stable after ripening, reaching levels of up to 10 mg per berry (Iland & Coombe 1988; Saito & Kasai 1968). TA is essential to the winemaking process as it raises the acidity of the wine must which prevents microbial spoilage and discolouration (Kalathenos, Sutherland & Roberts 1995). Importantly, is not broken down during fermentation and maturation, so exogenous TA is often added during the winemaking process. TA also contributes to the palatability, flavour and mouthfeel of wine (Plane, Mattick & Weirs 1980). No physiological role however has yet been attributed to its formation and accumulation in grapevine.

Despite the importance of TA to the winemaking industry, relatively little is known about its biosynthesis. Early experiments have shown that most TA is synthesised from ascorbic acid via a C4/C5 cleavage pathway (Saito & Loewus 1989). Radiolabelling experiments revealed that ascorbic acid fed to grapevine is converted to 2-keto-L-gulonate, then L-idonate, then to 5-keto-D-gluconate, which is cleaved to form tartaric acid semialdehyde, the direct precursor to tartaric acid (Saito 1992; Saito & Kasai 1984). Kinetic studies suggest the rate-limiting step of the reaction is the conversion of L-idonate to 5-keto-D-gluconate (Malipiero, Ruffner & Rast 1987).

Although the pathway intermediates have been well-established, few of the enzymes involved in the pathway have been identified. L-idonate dehydrogenase (IdnDH), which catalyses the conversion of L-idonate to 5-keto-D-gluconate, was the first enzyme in this pathway to be characterised (DeBolt, Cook & Ford 2006). This enzyme was found to be highly specific, with no activity detected against alternative substrates tested. Additionally, a grapevine 2-keto-L-gulonate reductase (2KGR) identified *via* bioinformatics has been partially characterised *in vitro* (Burbidge 2011; Jia 2015). This enzyme is capable of interconverting 2-keto-L-gulonate and L-idonate, but also acts as a reductase on a number of other substrates, including ascorbate, D-gluconate and L-idonate (Burbidge 2011). Further evidence, including *in planta* experiments, is required to confirm its involvement in the TA biosynthetic pathway. A limiting factor in the study of

tartaric acid biosynthesis is the limited commercial availability of L-idonate. Cheap and efficient production of L-idonate would greatly benefit further research in this field.

This study focuses on the identification and characterisation of new 2KGR candidates, and the development of a system for microbial production of L-idonate.

Comparison between grapevine enzymes and enzymes from other species with known activity can serve as a foundation for identification of enzymes involved in the TA biosynthesis pathway. The conversion of 2-keto-L-gulonate to L-idonate is a fairly unusual reaction within nature, and only a handful of enzymes have been identified with this activity. The previously characterised Vv2KGR was identified based on similarity to an *Escherichia coli* enzyme that is able to reduce 2-keto-L-gulonate to L-idonate, but with low activity and specificity (Burbidge 2011; Yum et al. 1998). More recently, two fungal 2KGRs capable of interconverting 2-keto-L-gulonate and L-idonate were identified in *Aspergillus niger* (Kuivanen, Arvas & Richard 2017; Kuivanen et al. 2016). These enzymes, GluC and GluD, are the first substrate-specific 2KGRs to be reported and may offer new avenues for the identification of previously unrecognised 2KGRs within the grapevine genome.

If a suitable new 2KGR is found, it may be used for the microbial production of L-idonate. Otherwise, the previously identified Vv2KGR, or a 2KGR from another species, may be used. Microbial biosynthesis is a versatile method of obtaining useful products, and offers several advantages over chemical synthesis (Du, Shao & Zhao 2011). Most importantly for this purpose, microbial biosynthesis has the ability to be regio- and stereoselective, which will increase product yield and facilitate straightforward purification. 2-keto-L-gulonate is an intermediate in ascorbic acid synthesis, making it a relatively cheap compound that can be used as a feedstock for the bioconversion of 2-keto-L-gulonate to L-idonate (Grindley et al. 1988). This reaction can be performed *in vitro* using purified soluble or immobilised enzymes, or *via* whole-cell biotransformation, using metabolically dormant cells containing the desired enzyme activity (Faber 2004). Whole-cell biotransformation is preferred over the use of a cell-free extract as it eliminates the need for protein purification and improves enzyme stability. Furthermore, whole-cell systems circumvent the need to supply exogenous cofactor and can be engineered to provide facile cofactor regeneration (Duetz, Van Beilen & Witholt 2001).

In this study, four potential grapevine 2KGRs were identified based on homology to the *A. niger* 2KGRs GluC and GluD. Two of these enzymes were expressed in *E. coli* and purified, and the kinetic parameters of the recombinant proteins were investigated *in vitro*. Additionally, the feasibility of constructing a whole-cell biotransformation system for the production of L-idonate was explored.

Results

To search for potential grapevine 2-keto-L-gulonate reductases (2KGRs), a protein-protein BLAST search (BLASTP) was performed against the *V. vinifera* genome on the NCBI database using the fungal 2KGRs GluC and GluD to identify homologues. GluC gave 24 hits (17 with score >80) (Table 3-1), and GluD gave 29 hits (23 with score >80) (Table 3-2). All of the hits from GluC were also included in the GluD results for a total of 29 unique hits. Several of these candidates had been previously identified by Burbidge (2011) but only one (XP_003632860.1) was expressed and purified. XP_003632860.1 was found to have 2KGR activity *in vitro* and will be referred to in this study as *Vitis vinifera* 2KGR 0 (Vv2KGR0).

Table 3-1 BLASTP hits of Aspergillus niger *strain ATCC protein GluC against* Vitis vinifera *proteome. The accession number is from the NCBI database. The max score describes the highest alignment score between the query sequence and the database sequence segment, the query cover describes the percent of the query length included in the aligned segments and the identity describes the percentage of exact matches between the aligned segments. Highlighted sequences were previously investigated by Burbidge (2011).*

Accession	Description	Max Score	Query Cover	Identity
XP_002281980.1	PREDICTED: glyoxylate/hydroxypyruvate reductase HPR3	105	74%	29%
XP_002284520.1	PREDICTED: hydroxyphenylpyruvate reductase	102	78%	29%
XP_002282078.1	PREDICTED: glyoxylate/hydroxypyruvate reductase HPR3	99	74%	30%
XP_002285358.1	PREDICTED: D-3-phosphoglycerate dehydrogenase 1, chloroplastic	96.7	74%	29%
XP_002283022.1	PREDICTED: D-3-phosphoglycerate dehydrogenase 3, chloroplastic	96.7	74%	29%
CBI19470.3	unnamed protein product	96.7	74%	29%
XP_003632860.1	PREDICTED: hydroxyphenylpyruvate reductase	92.4	85%	30%
CAN64820.1	hypothetical protein VITISV_009548	89.4	50%	31%
CBI38817.3	unnamed protein product	91.3	75%	25%
XP_003633919.1	PREDICTED: glyoxylate/hydroxypyruvate reductase HPR3 isoform X1	89.4	73%	25%
XP_010661599.1	PREDICTED: glyoxylate/hydroxypyruvate reductase HPR3 isoform X2	87.4	73%	25%
CBI36125.3	unnamed protein product	86.3	73%	31%
XP_002273552.1	PREDICTED: D-3-phosphoglycerate dehydrogenase 3, chloroplastic	86.3	74%	27%
CAN65023.1	hypothetical protein VITISV_020147	81.3	73%	25%
XP_002279281.1	PREDICTED: glycerate dehydrogenase HPR, peroxisomal isoform X1	81.3	73%	25%
CBI40269.3	unnamed protein product	80.9	50%	32%
XP_019074661.1	PREDICTED: glycerate dehydrogenase isoform X2	80.1	72%	25%

Table 3-2 BLASTP hits of Aspergillus niger *strain ATCC protein GluD against* Vitis vinifera *proteome. The accession number is from the NCBI database. Highlighted sequences were previously investigated by Burbidge (2011).*

Accession	Description	Max Score	Query Cover	Identity
XP_002284520.1	PREDICTED: hydroxyphenylpyruvate reductase	154	81%	37%
XP_002281980.1	PREDICTED: glyoxylate/hydroxypyruvate reductase HPR3	141	74%	36%
XP_002282078.1	PREDICTED: glyoxylate/hydroxypyruvate reductase HPR3	141	74%	38%
XP_010661599.1	PREDICTED: glyoxylate/hydroxypyruvate reductase HPR3 isoform X2	139	79%	35%
XP_003632860.1	PREDICTED: hydroxyphenylpyruvate reductase	135	81%	35%
XP_003633919.1	PREDICTED: glyoxylate/hydroxypyruvate reductase HPR3 isoform X1	136	74%	35%
CBI19470.3	unnamed protein product	138	67%	36%
XP_002283022.1	PREDICTED: D-3-phosphoglycerate dehydrogenase 3, chloroplastic	139	67%	36%
XP_002273552.1	PREDICTED: D-3-phosphoglycerate dehydrogenase 3, chloroplastic	139	67%	37%
XP_002285358.1	PREDICTED: D-3-phosphoglycerate dehydrogenase 1, chloroplastic	137	67%	36%
CBI36125.3	unnamed protein product	129	70%	36%
CBI38817.3	unnamed protein product	135	78%	33%
CAN68604.1	hypothetical protein VITISV_036580	124	61%	37%
CAN64820.1	hypothetical protein VITISV_009548	116	55%	38%
CBI25279.3	unnamed protein product	115	49%	39%
CAN77384.1	hypothetical protein VITISV_006350	111	51%	38%
CAN65023.1	hypothetical protein VITISV_020147	110	67%	31%
XP_019074661.1	PREDICTED: glycerate dehydrogenase isoform X2	110	67%	31%
XP_002279281.1	PREDICTED: glycerate dehydrogenase HPR, peroxisomal isoform X1	109	67%	31%
CBI40269.3	unnamed protein product	109	49%	37%
XP_002278444.1	PREDICTED: formate dehydrogenase, mitochondrial	94	67%	30%
CAN71454.1	hypothetical protein VITISV_036417	94	67%	30%
CBI34779.3	unnamed protein product	79.7	74%	30%

The families, sites and domains of the top 10 BLAST results were analysed for similarities to GluC and GluD using InterPro (Table 3-3). Predicted chloroplastic proteins were not analysed as TA synthesis is most likely to occur in the cytoplasm (Wen et al. 2010). The top five BLAST results showed the most similarity to the GluC and GluD domains. All proteins analysed shared the same NAD(P)H binding domain signature and a D-isomer specific 2-hydroxyacid dehydrogenase catalytic domain. Lower-ranked BLAST results had additional domains that were not present in the GluC and GluD proteins, including an S-adenosyl-L-homocysteine hydrolase domain, an NAD binding domain and an ACT domain. These proteins were classified as members of the D-3-phosphoglycerate dehydrogenase family and were discounted from further analysis. No protein families were identified for the earlier results. The top four results were selected for additional investigation due to their strong homology and similarity of binding domains with GluC and GluD. The fifth candidate, Vv2KGR0 also shared a majority of domains with GluC and GluD but has already been studied *in vitro.* The DNA sequences and amino acid sequences of the candidates investigated are shown in Figure 3-1.

Table 3-3 Summary of domains present in A. niger *2KGRs and* V. vinifera *homologues. Domains were identified using InterPro (http://www.ebi.ac.uk/interpro/). No families were identified except for those with additional s-adenosyl-L-homocysteine hydrolase, NAD binding and ACT domains. These proteins were classed as members of the D-3-phosphoglycerate dehydrogenase family. The shaded cells represent the proteins selected for further analysis.*

Name	Signature	GluC	GluD	XP_002284520.1	XP_002281980.1	XP_002282078.1	XP_010661599.1	XP_003632860.1	XP_003633919.1	CBI19470.3	XP_002283022.1	CBI38817.3	CAN68604.1
NAD(P)H	G3DSA:3.40.50.720	x	x	x	x	x	x	x	x	x	x	x	x
	SSF51735	x	x	x	x	x	x	x	x	x	x	x	x
D-isomer specific 2-hydroxyacid dehydrogenase, catalytic domain	PF00389 (2-Hacid_dh)	x	x	x	x	x	x	x	x	x	x	x	x
D-isomer specific 2-hydroxyacid dehydrogenase, NAD-binding domain	PF02826 2-Hacid_dh_C	x	x	x	x	x	x	x	x	x	x	x	x
	D 2 Hydroxyacid Dh 1	x	x	x				x		x	x		x
	D 2 Hydroxyacid Dh 2									x	x		x
	D 2 Hydroxyacid Dh 3	x	x							x	x		x
Additional domains*											x	x	x

*S-adenosyl-L-homocysteine hydrolase, NAD binding domain; ACT domain

XP_002284520.1

A
ACTTCCCAGCCTAACCTAACGTCCTAGATCTAAACATGGCCTCCTCTCCCAAGTTTAAAATATGCACTTCAATTCTTTCGTTCAGGAAC
GACGCCGTTTGTTAGGACGCAGTGAAGAACCCTAACCCTAATTAGTGTTCGCATTCATTGTAACGATGGAAAACATATGCGTTTTGTTG
ACCTATCCAGTTCCAGAATACCTTGTACAAAAACTCGAAAAACGATTTACAGTCTTCAAGTTCCGTGAAGTTGCTTCCAACCCTCAGCT
CCTAAGAGAAATCTCAAATTCAATAAGAGCAATCGTCGGCACCTCCGTATGCGGCGCCGATGCCGGACTCATTGATGCATTGCCGAAGT
TGGAGATTGTGGCGAGTTACAGCGTAGGTTTCGACAAGATCGATTTGGTGAAGTGCAAGGAGAGAGGCATTACCGTTACAAACACGCCC
GATGTGCTTACGGACGATGTGGCGGACTCGGCGATCGGGTTGGCTTTGGCAACGTTAAGGCGAATGTGCGTCTGTGATCGGTTTGTGAG
AAGCGGGAAATGGAAGAAAGGCGATTTTGAATTGACCACGAAGTTCAGTGGTAAATCCATTGGAATTGTTGGGTTGGGCAGGATAGGTT
CAGCAATTGCAAAGAGAGCTGAAGCTTTTGGCAGCTCAATTAGTTACCACTCCAGATCAGAGAAACCAGAATCGAATTACAAATATTAT
TCTAACATAATCGACTTGGCCACCAACTGTCAAATCCTCTTTGTGGCATGTGCTCTGACCAAAGAAACGCACCATATTGTTGACCGCAA
AGTCATTGATGCACTGGGTCCAAAAGGCATTATCATCAACATTGGACGGGGTGCACATATTGACGAACCTGAGCTTGTGTCTGCACTGC
TTGAAGGCCGGTTGGCTGGTGCAGGGCTGGACGTGTTTGAACATGAGCCAGAGGTACCTGAAGAACTGTTGGGGCTTGAAAATGTAGTC
CTCCAACCTCATGCGGGAAGCGATACCGTGGAAACGAGTGTGGCAATGTCAGACCTTGTAATCGATAACCTAGAGGCATGCTTCCAAAA
CAAACCTGTCTTAACTCCAGTCATTTGATTTGTCGTAAGGTATGTGAACATGATCACTTAAAGTCATTGAGATGGGATTTGAATCATTT
GATTTACCGTCTGAA

B
MENICVLLTYPVPEYLVQKLEKRFTVFKFREVASNPQLLREISNSIRAIVGTSVCGADAGLIDALPKLEIVASYSVGFDKIDLVKCKER
GITVTNTPDVLTDDVADSAIGLALATLRRMCVCDRFVRSGKWKKGDFELTTKFSGKSIGIVGLGRIGSAIAKRAEAFGSSISYHSRSEK
PESNYKYYSNIIDLATNCQILFVACALTKETHHIVDRKVIDALGPKGIIINIGRGAHIDEPELVSALLEGRLAGAGLDVFEHEPEVPEE
LLGLENVVLQPHAGSDTVETSVAMSDLVIDNLEACFQNKPVLTPVI

XP_002281980.1

A
GTGGATCCAATCAATCTTAAGTGACGAAGACATTCGAACTTGTGAAACAGGAATTGAGGGTTAAAAGAGTAAACACGGACTTGCAGGTA
CTGAAATGACCGCTATGGATGAACTGCCTCTGGTTCTTGTCCATGTCTTGCCGCCATTCGAGATCCCGTTCAAGGGCCGACTTCAGAGT
CGATTCCAACTCATCGACTCCTCCGATTCAACTTTCTCCCCACACGCCAGCGTCCTGCTCTGTGTTGGCCCCGCTCCGGTCAGTTCCGA
CACCCTCCGCCCACCTCCCCTCTCTCCAGTGCATCGTCGGCTCAAGCGCCGGCGTAGACCATATTGACCTAGAGGAGTGCCGCCGCCGCG
GCATCACTGTCACCAATGCTGGCTCCTCCTTCTGCGAGGACGGCGCTGATTTCGCCATCGGGCTTCTGATCGATGTTCTCCGACGGATA
TCGGCCGCCGATCGGTACGTTCGAGCCGGTTTGTGGCCGATGAAGGGAGATTACCCACTAGGTTCCAAGTTAGGGGGAAAGCGAGTTGG
AATTGTTGGACTGGGGAAAATTGGCTCCGAAATTGCCAAAAGGCTCGTGGCATTTGGCTGCAGAATCGCCTACAACTCCAGGAATAAAA
AGTCATCAGTTTCATTCCCATACTATGCAAATATTTGCAACCTTGCTGCTAATAGTGACATACTCATTATTTGTTGTGCCTTGACTAAG
GAAACTCACCACTTGATTGACAAGGATGTCATGACAGCATTGGGAAAGGAAGGAGTCATCATCAATGTTGGTCGCGGGGGTCTCATCAA
TGAAAAGGAATTGGTGCAGTGTCTTGTGCAAGGTCAAATTAGAGGTGCTGGTCTTGATGTGTTTGAGAATGAGCCTGATGTGCCAAAAG
AGCTATTTGAATTGGAGAATGTTGTGCTATCTCCACATAAAGCTATTGCGACCCTAGAGTCCTTAGCATCATTGCAAGAGCTGATCGTG
GGCAACTTGGAGGCATTCTTCTCTAATAAACCTTTGCTTTCCCCAATAAACCTTGACTAAACAGCTTATTTTTTCTTGGACCCTGGAGG
ATTATCTCTACTTCCATGTTGACGAACAGGTTGGCTCCTTGCTTCCAATAAAGATCTCACGTTTAATTTGGGATCATGTTGAAAATTAC
CATATTAATTCTTGAATCCTCACAAAAAATGGAGTTGCTAATGTGTTTGTGTAAA

B
MTAMDELPLVLVHVLPPFEIPFKGRLQSRFQLIDSSDSTFSPHASVLLCVGPAPVSSDTLRHLPSLQCIVGSSAGVDHIDLEECRRRGI
TVTNAGSSFCEDGADFAIGLLIDVLRRISAADRYVRAGLWPMKGDYPLGSKLGGKRVGIVGLGKIGSEIAKRLVAFGCRIAYNSRNKKS
SVSFPYYANICNLAANSDILIICCALTKETHHLIDKDVMTALGKEGVIINVGRGGLINEKELVQCLVQGQIRGAGLDVFENEPDVPKEL
FELENVVLSPHKAIATLESLASLQELIVGNLEAFFSNKPLLSPINLD

XP_002282078.1

A
ACCCGTGTATAACAAGAGAATCCATACGTCAACCAAGGCCAGATTGAATAATCCCAACGCTTCTCCTTCCCAATCCCTACAATAAGTTA
GCAAAAGGTGGAAAATGTTAAACGGAGAAACACCATTGCCGCAGAATGGACACCGGAGAAACTCCATGGCTGCTATCACTGGTGAGGCA
GAAGCAGAAGCACCAGGGGTGGTATTTGTTCATGGCAGCCCACCGTTTGGCCTCCCTTTCAAGGACCGGTTACTGAGTCGGTTCCAACT
CATCCACATGTCTGAGTTACCTGAATCCTCCCACGTCAAGGTCATGCTCTGCATGGATCACACTCCGGTCACCTCCCAGACTCTATATA
AACTCCCTTCACTTGAGTGCATCGTGGCATCAAGCGCTGGCGTTGACCACATTGACCTCACGACTTGCCGCCTCCGTGGGATTGCGGTC
GCCAACGGCAGTCAGGCTTTCTCTGAGGACGTGGCTGATTACGCCGTCGCCCTCTTGATGGATGTCCTCAGAAAGATTTCGGCGGGCGA
TAGGTATCTGCGCTCCGGTTTGTGGTCCACGAAGGGAGATTATCCGCTTGGTTGGAAGCTAGGGGGAAAGCGAGTTGGAATTGTGGGAC
TGGGGAACATTGGCTCTGAAGTTGCCAAAAGGCTCGTGGCCTTTGGCTGCGCCATTGCCTACAACTCCAGGAAGAAAAGGTCATCTGTT
TCATTCCCATACTACGCAGATGTGTGCGACCTCGCGGCTAACAGCGACATTCTCGTTATTTGTGGTGCATTGACAAGTGAAACCCACCA
CATAATCAACAAGGATGTGATGACAGCACTGGGGAAGGAAGGAGTCATCATCAACGTTGGCCGTGGATCTCTCATTAATCAGAAGGAAC
TGGTGCAGTTTCTGGTGGAAGGTCAGATTAGAGGTGCGGGGCTTGATGTGTTTGAGAATGAACCCATTGTGCCTAGAGAACTACTTGAG
TTGGATAATGTTGTGTTGTCTCCACATAATGCTGTTGTGACACCAGAAGCCTTTGAAGCTATGCAAGAGCTGGCCATATCCAACTTGGG
GGCCTTCTTCTCTAACAAACCTCTGCTATCACCTATCTAGCATGTATGATACTTGGTATTCATTGGGCCATCGGATTATTTTACTCATT
CTAATCCAACTTTTCACATGTAGAATTCACTATCAATCGCCTTGAACCTTCAGGCATTCACATAA

B
MLNGETPLPQNGHRRNSMAAITGEAEAEAPGVVFVHGSPPFGLPFKDRLLSRFQLIHMSELPESSHVKVMLCMDHTPVTSQTLYKLPSL
ECIVASSAGVDHIDLTTCRLRGIAVANGSQAFSEDVADYAVALLMDVLRKISAGDRYLRSGLWSTKGDYPLGWKLGGKRVGIVGLGNIG
SEVAKRLVAFGCAIAYNSRKKRSSVSFPYYADVCDLAANSDILVICGALTSETHHIINKDVMTALGKEGVIINVGRGSLINQKELVQFL
VEGQIRGAGLDVFENEPIVPRELLELDNVVLSPHNAVVTPEAFEAMQELAISNLGAFFSNKPLLSPI

XP_010661599.1

A
TTAAATATGTCATGCTAATTGATGAATTATATTAGTTCAAATAAAATCTAATTTAATGTAATTATTAAATAGAATTGATAATTAAATGA
GTTATATTATTTTTTATTTTGGTCATTGGGATACGACAGGAGCTACCGACAAGAATTCCACCTGGTTTGCATGGTCACATCCTAAGCCA
CATGGAATGGAATCTATCACTCCCAAATCCCAATAACAACCAATTCCTCGATTGTTGGAATACGCAATAGTGAATTAGCCCTTCTTTCG
TCCATAATCGTCCATGGCGGATCAGCTCCCGCAAGTTTTAGTTCTTCGTCCACCCCCAGTTTTCACTCTGTTCGAAACTCAATTCTCCC
AGAAATTCCATTTCCTCAGAGCCTGGGAATCTCCACTGCCCACCGCTGAGTTTCTTGCTACCCATGCGGCTTCCGTCAAGGCCGTTCTC
TGCTCGGGCAGTACTCCCATCACGGCCGACATCCTCCGCCACCTACCTTCCCTCCAACTCATCGTCACCACCAGCGCTGGTCTCAACCA
CATCAACCTCCCTGAGTGCCGGCGCCGCAGTATCTCCATCGCCAACGCCGGAGAGATCTTCTCCGATGATTGCGCCGACCTGGCGGTCG
GACTTCTGATGGACGTTCTCCGAAAAATATCAGCCGCTGATCGGTTTATTCGAGCCGGCCTCTGGCCGATCAGAGGAGACTATCCTCTT
GGCTCCAAGTTGGGAGGCAAGCGAGTTGGGATTGTTGGACTAGGAAGCATTGGCTTAGAAGTTGCAAAAAGACTTGAAGCTTTTGGCTG
CATTATCTTATACAACTCAAGGAGGAAGAAGGCAAACATATCCTACCCTTTCTACTCAAATGTCTGTGAACTTGCAGCCAATAGCAATG
CCCTCATAATTTGCTGTGCATTAACTGATGAAACCCGCCACATGATTAACAAGGAAGTCATGAAAGCATTGGGAAAGGAAGGAGTCATC
ATCAACATTGGACGTGGGGCTATCATTGATGAGAAGGAACTGGTGCAGTGTTTGGTGCAAGGAGAGATAGGGGGTGCTGGTTTGGATGT
GTTTGAGAATGAACCTGATGTGCCCAAGGAGCTGTTCACATTGGACAATGTTGTGCTATCACCACATGTGGCTGTTTTTACACAAGAAT
CCTTTTCAGACTTGTACGATCTTATGGTGGGGAATTTGGAAGCTTTCTTCTCAAATAAAACCTTACTTTCTCCGGTCTTGGATGAATAA
TCAGATTAGCTCCATATTTTCAGGTTAGTTTCTTCATCATGTTAAAGCATTTTCACTTAAGACTCTGTATCAAACAGATTTCCGGGTTG
GACATGTATGTTCCAATTAGCTGTCTCATGAATAAATGCTCTGATAATTTTGCTTATGTATGAATTAGATGACGACTCAGGCACAAACT
A

B
MADQLPQVLVLRPPPVFTLFETQFSQKFHFLRAWESPLPTAEFLATHAASVKAVLCSGSTPITADILRHLPSLQLIVTTSAGLNHINLP
ECRRRSISIANAGEIFSDDCADLAVGLLMDVLRKISAADRFIRAGLWPIRGDYPLGSKLGGKRVGIVGLGSIGLEVAKRLEAFGCIILY
NSRRKKANISYPFYSNVCELAANSNALIICCALTDETRHMINKEVMKALGKEGVIINIGRGAIIDEKELVQCLVQGEIGGAGLDVFENE
PDVPKELFTLDNVVLSPHVAVFTQESFSDLYDLMVGNLEAFFSNKTLLSPVLDE

Figure 3-1 DNA sequences (A) and amino acid sequences (B) of the four potential 2KGRs selected for further investigation

To determine if the candidate 2KGR expression patterns are likely to be temporally and spatially consistent with TA synthesis, gene expression was analysed using Expression Atlas (Petryszak et al. 2016). Limited data was available from a small number of studies using different varieties of grapes. One study by Palumbo et al. (2014) measured the gene expression in berries at different stages of development. These data are most relevant since TA synthesis occurs in the berry during development. The expression data is presented in Table 3-4. The relative expression of the four 2KGR candidates was compared to previously identified L-idonate dehydrogenase (LIDH) and Vv2KGR0 proteins.

Table 3-4 Relative expression of V. vinifera *proteins at various stages of berry development. Expression was analysed using Expression Atlas (https://www.ebi.ac.uk/gxa/home/). LIDH: L-idonate dehydrogenase; Vv2KGR0:* V. vinifera *2-keto-L-gulonate reductase*

	Expression in Berry			
	Pea-sized	Beginning to touch	Soft	Ripe for harvest
XP_002284520.1	Low	Not detected	Low	Low
XP_002281980.1	Med	Low	Low	Low
XP_002282078.1	Not detected	Med	Med	Med
XP_010661599.1	Low	Low	Low	Low
LIDH	Med	Med	Med	Med
Vv2KGR0	Med	Med	Med	Med

To further characterise the potential candidates, TargetP (Emanuelsson et al. 2007) was used to predict their subcellular localisation (Table 3-5). XP_002282078.1 was predicted to localise to the mitochondria. The other three candidates analysed did not contain a signal peptide for mitochondria, chloroplasts or secretory pathways. None of the candidates were predicted to contain transmembrane helices.

Finally, Spider2 (http://sparks-lab.org/server/SPIDER2/) was used to predict regions of secondary structure. This was to check for long unstructured regions at the beginning or end of the protein sequence, which could pose problems for expressing, purifying or crystallising the protein in future investigations. XP_002282078.1 was predicted to have an unstructured region of approximately 30 amino acids at the N terminus. All other candidates were predicted to have regions of secondary structure near the N and C termini.

Table 3-5 Chromosome location, subcellular localisation prediction and length of 2KGR candidates. Subcellular localisation and transmembrane helices were predicted using CBS Prediction Services (http://www.cbs.dtu.dk/services/).

	Chromosome	Subcellular Localisation	Transmembrane Helices	Length (aa)
XP_002284520.1	9	other*	0	313
XP_002281980.1	2	other*	0	314
XP_002282078.1	2	mitochondria	0	334
XP_010661599.1	15	other*	0	321

*other denotes location other than chloroplast, mitochondria or secretory pathway

Based on these results, candidates XP_002284520.1 (UniProt D7U0H8) and XP_002281980.1 (UniProt F6HUJ6) were chosen for expression and purification. Henceforth, they will be referred to as 2KGR1 and 2KGR2 respectively.

The two 2KGR candidates were cloned into pET-11a and expressed in *E. coli* BL21. 2KGR1 was predicted to be 34.0 kDa and 2KGR2 was predicted to be 33.9 kDa. High levels of both proteins were produced; however, they showed poor solubility in Buffer A (Figure 3-2). Expression tests were performed at 37 °C, 30 °C and 16 °C and it was found that expression of 2KGR1 was highest at 37 °C after 4 hours and expression of 2KGR2 was highest at 30 °C after 6 hours. However, expression under optimised conditions failed to produce a large amount of soluble protein. Solubility tests were performed using a range of buffers, but neither 2KGR1 nor 2KGR2 were sufficiently soluble for purification (Supplementary Figure 3-1).

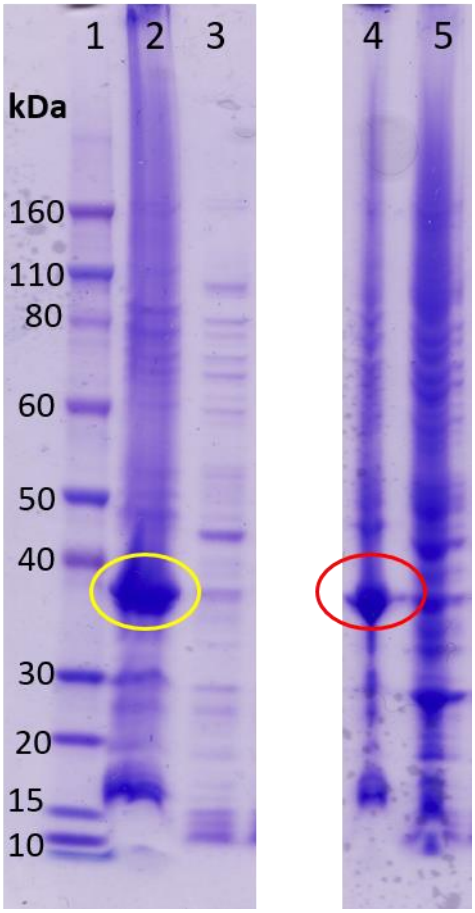


Figure 3-2 Expression of 2-keto-L-gulonate reductase candidates in E. coli *visualised by SDS PAGE. 1. Protein Standard Lanes 2-3 are from 2KGR1expression: 2. Insoluble fraction 3. Soluble fraction; Lanes 4-5 are from 2KGR2expression: 4. Insoluble fraction 5. Soluble fraction; 2KGR1 and 2KGR2 are circled in yellow and red respectively.*

To determine if the 2KGR1 and 2KGR2 proteins were catalytically active *in vivo*, whole-cell biotransformation of 2-keto-L-gulonate was attempted using *E. coli* BL21 strains expressing each of the 2KGR candidates (Figure 3-3). No difference in 2-keto-L-gulonate concentration was detected between the control strain and the 2KGR-expressing strains.

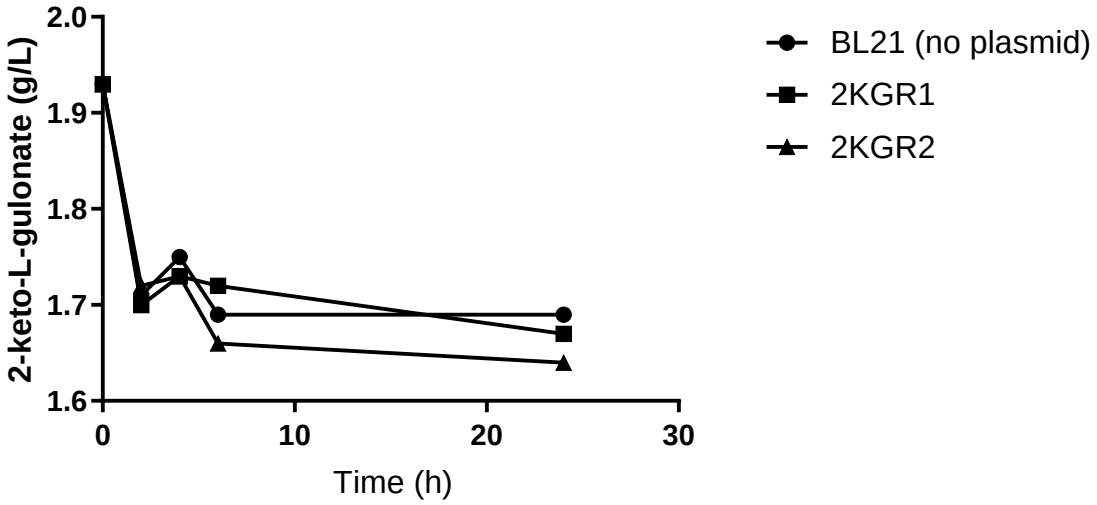


Figure 3-3 Whole-cell biotransformation of 2-keto-L-gulonate was attempted using E. coli *expressing the 2KGR candidates. Assays were performed using 0.3g cell wet weight with 10 mM 2-keto-L-gulonate substrate in 50 mM tris-HCl buffer (pH 7.5) at 30 °C. There was no evidence of enzyme activity compared to the no plasmid control.*

In order to obtain soluble protein for purification, 2KGR1 and 2KGR2 were each fused with NusA. The size of the 2KGR1-NusA and the 2KGR2-NusA fusion proteins were each predicted to be 94 kDa. The fusion proteins were expressed in *E. coli* BL21 and showed high levels of expression and good solubility in Buffer A (Figure 3-4). 2KGR1-NusA and 2KGR2-NusA were expressed and purified for further analysis.

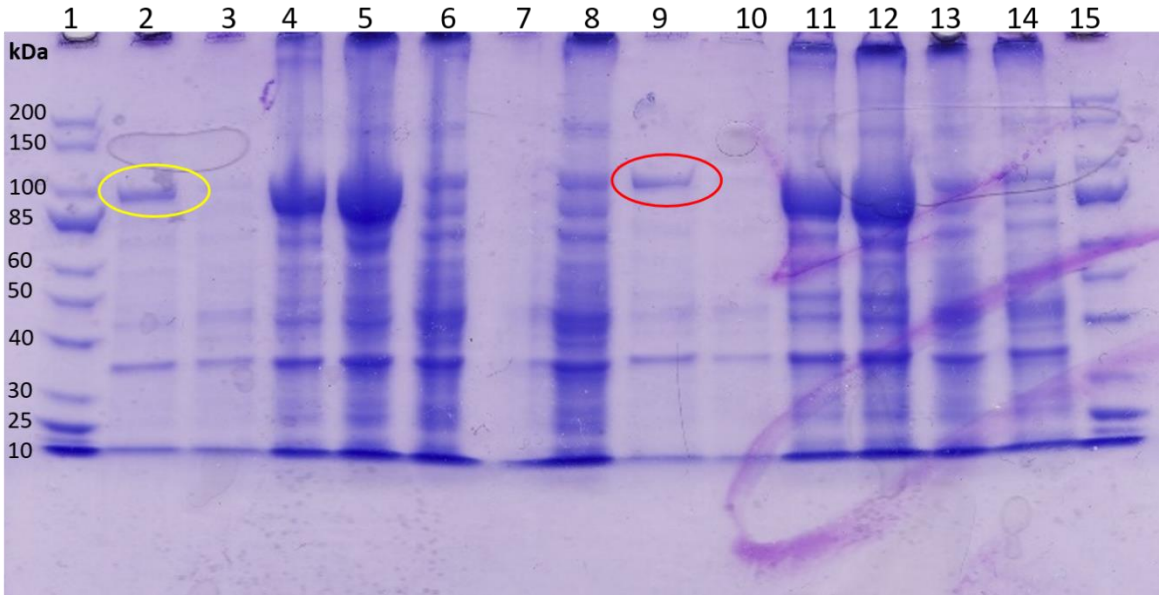


Figure 3-4 Expression of 2KGR-NusA fusion proteins visualised by SDS-PAGE. Lanes 1 and 15 are protein standard. Lanes 2-8 are from 2KGR1-NusA expression: 2. Soluble fraction from IPTG-induced cells (+IPTG) 3. Soluble fraction from non-induced cells (-IPTG) 4,5. Insoluble fraction +IPTG, 6,7. Insoluble fraction –IPTG.. Lanes 9-14 are from 2KGR2-NusA expression: 9. Soluble fraction +IPTG 10. Soluble fraction -IPTG 11,12.. Insoluble fraction +IPTG, 13,14 Insoluble fraction –IPTG. 2KGR1 and 2KGR2 in the soluble fraction are circled in yellow and red respectively.

Purified 2KGR1-NusA and 2KGR2-NusA were used in enzyme assays to explore the potential of 2-keto-L-gulonate reductase activity. The pH conditions under which each enzyme was most active were investigated using two different buffers to ensure differences in activity were due to pH and not buffer composition. Both enzymes were found to be most active at pH 7.0 and favoured Tris buffer over HEPES (Figure 3-5). Tris 7.0 buffer was used for all future experiments.

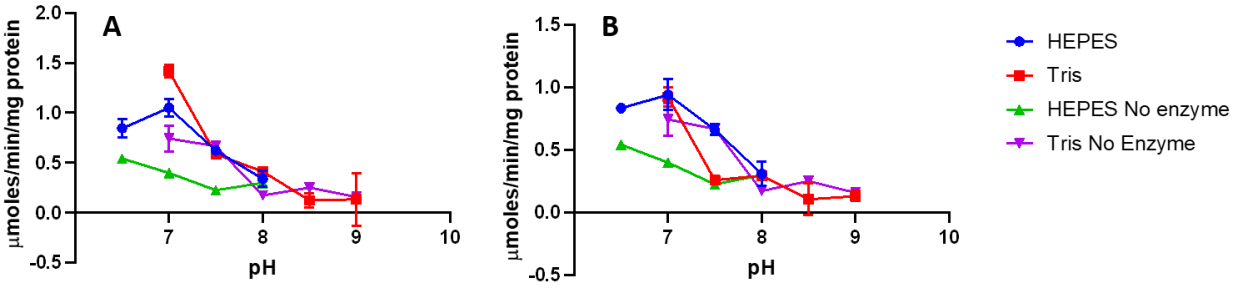


Figure 3-5 Activity of recombinant 2KGR1 and 2KGR2 at various pH in 100 mM Tris and HEPES buffers. Assays were performed with 0.5 µg purified enzyme, 40 mM 2-keto-L-gulonate and 0.25 mM NADH at 37° C. Plotted values represent the mean of duplicate assays with error bars representing the Standard Error of the Mean.

2KGR1 and 2KGR2 are hypothesised to be involved in the conversion of 2-keto-L-gulonate to L-idonate in the tartaric acid biosynthesis pathway. Therefore their activity was analysed using 2-keto-L-gulonate as substrate. 2KGR1 and 2KGR2 activity was tested against varying concentrations of 2-keto-L-gulonate (Figure 3-6, Figure 3-7). A hyperbolic curve was fitted to the data, however, neither enzyme reached saturation, so the kinetic model is undetermined. Assuming Michaelis–Menten kinetics, 2KGR2 has a V_{max} of 0.44 ± 0.4 µmoles/min/mg protein and a K_M of 6.38 ±0.97 mM. The V_{max} and K_M of 2KGR1 could not be determined with the available data.

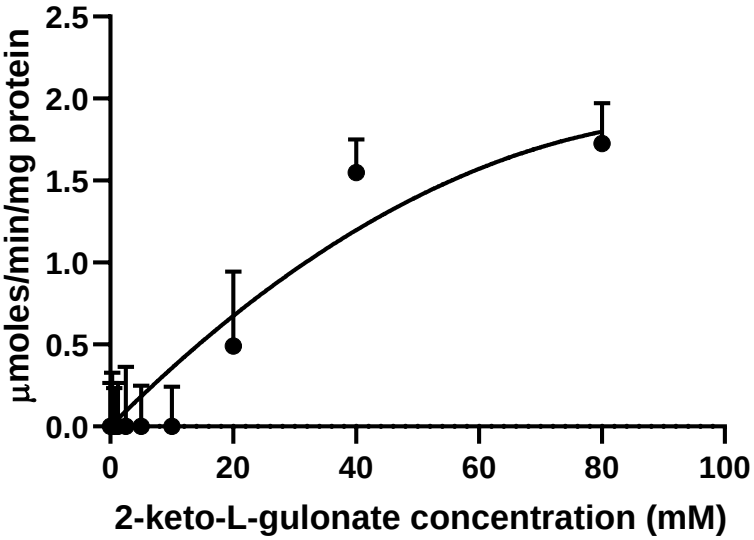


Figure 3-6 Activity of 2KGR1 against varying concentrations of 2-keto-L-gulonate. Assays were performed with 0.5 µg purified enzyme and 0.25 mM NADH in 100 mM Tris 7.0 at 37° C. Plotted values represent the mean of triplicate assays with error bars representing the Standard Error of the Mean.

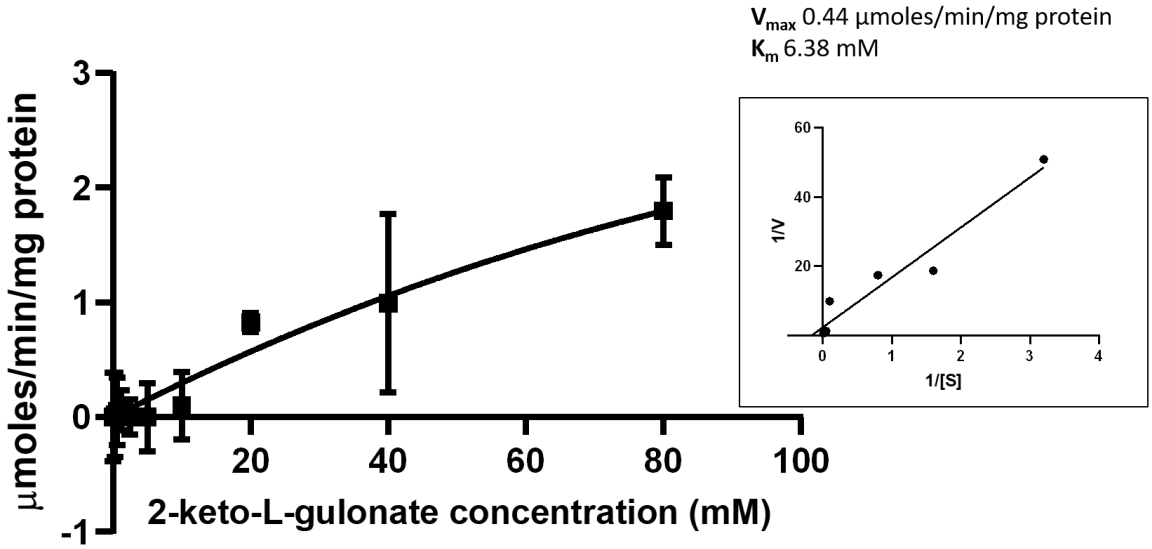


Figure 3-7 Activity of 2KGR2 against varying concentrations of 2-keto-L-gulonate. Assays were performed with 0.5 µg purified enzyme and 0.25 mM NADH in 100 mM Tris 7.0 at 37° C. Plotted values represent the mean of triplicate assays with error bars representing the Standard Error of the Mean. The trendline represents the hyperbolic curve of Michaelis-Menten kinetic analysis fitted using GraphPad prism 8.0. Inset: double-reciprocal Lineweaver-Burk plot.

To determine the products of 2KGR1 and 2KGR2 using 2-keto-L-gulonate as a substrate, an *in vitro* enzyme reaction was performed using 5 µg purified enzyme, 100 mM Tris 7.0, 40 mM 2-keto-L-gulonate (substrate) and 0.25 mM NADH in 1 mL total volume. The reaction mixture was analysed by HPLC (Figure 3-8). No new product peak was present and there was no evidence for the formation of L-idonate. L-idonate would be expected to

elute at 11.8 min (Supplementary Figure 3-2). The concentration of 2-keto-L-gulonate did not significantly decrease after 2 hours reaction time compared to the no enzyme control.

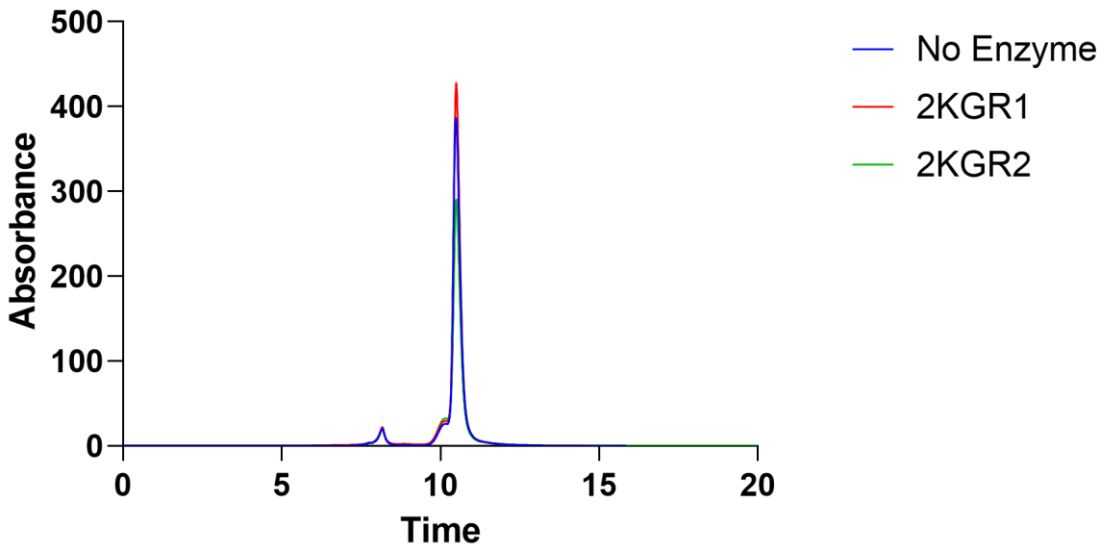


Figure 3-8 HPLC Chromatogram of in vitro *enzyme reaction. The reaction mixture consisted of 0.5 µg purified enzyme, 100 mM Tris 7.0, 40 mM 2-keto-L-gulonate (substrate) and 0.25 mM NADH. The reaction mixture was incubated at 37°C for 2 hours. No new product is present in the recombinant enzyme samples compared to the no enzyme control.*

The activity of 2KGR1 and 2KGR2 was tested against a range of substrates, which were selected due to their presence in *V. vinifera* metabolic pathways (Figure 3-9). 2KGR1 displayed activity against 2-keto-L-gulonate, ascorbate and glyoxylate, and no activity against 5-keto-D-gluconate. Likewise, 2KGR2 displayed a similar, but slightly lower, level of activity against each substrate.

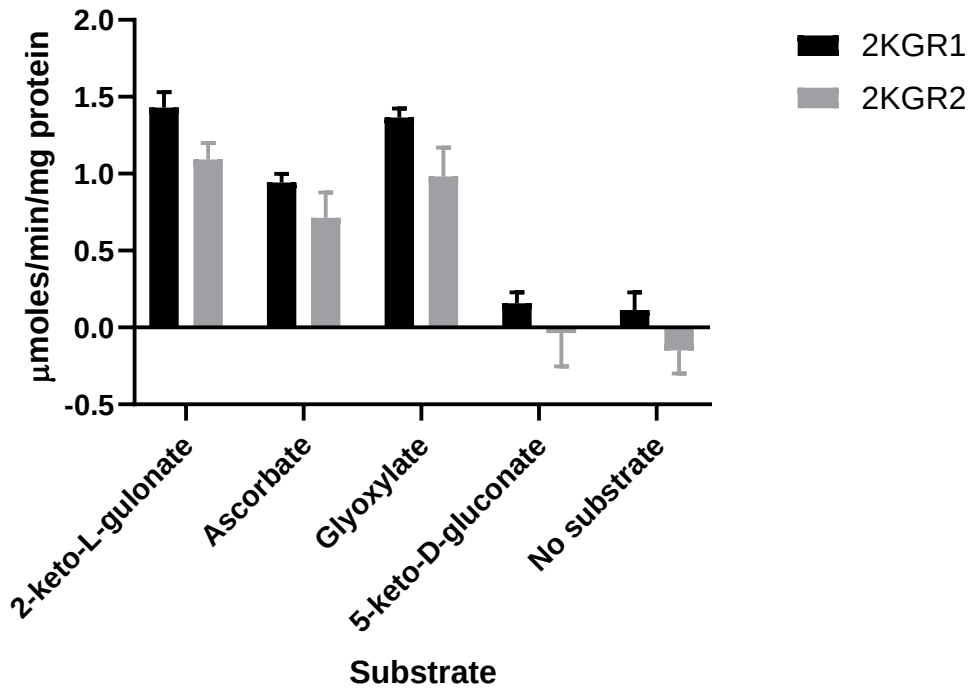


Figure 3-9 Substrate activity of purified 2KGR1 and 2KGR2. Assays were performed with 0.5 µg purified enzyme, 40 mM substrate and 0.25 mM NADH in 100 mM Tris 7.0 at 37° C. The bars represent the mean of triplicate assays with error bars representing the Standard Error of the Mean.

Since there was no evidence that either 2KGR1 or 2KGR2 produced L-idonate from 2-keto-L-gulonate, these enzymes were determined unsuitable for the whole-cell production of L-idonate. Instead, the previously identified Vv2KGR0 (Burbidge 2011; Jia 2015) was used. The whole-cell reaction was performed using *E. coli* BL21 expressing Vv2KGR0 with 0.090 g cell wet weight in 1 x PBS (pH 7.4) containing 0.25 M NaCl and 10 mM 2-keto-L-gulonate at 30 °C. The products were analysed by HPLC (Figure 3-10). After 24 hours, no production of L-idonate was observed, nor was there a significant decrease in 2-keto-L-gulonate concentration compared to the no enzyme control strain (BL21).

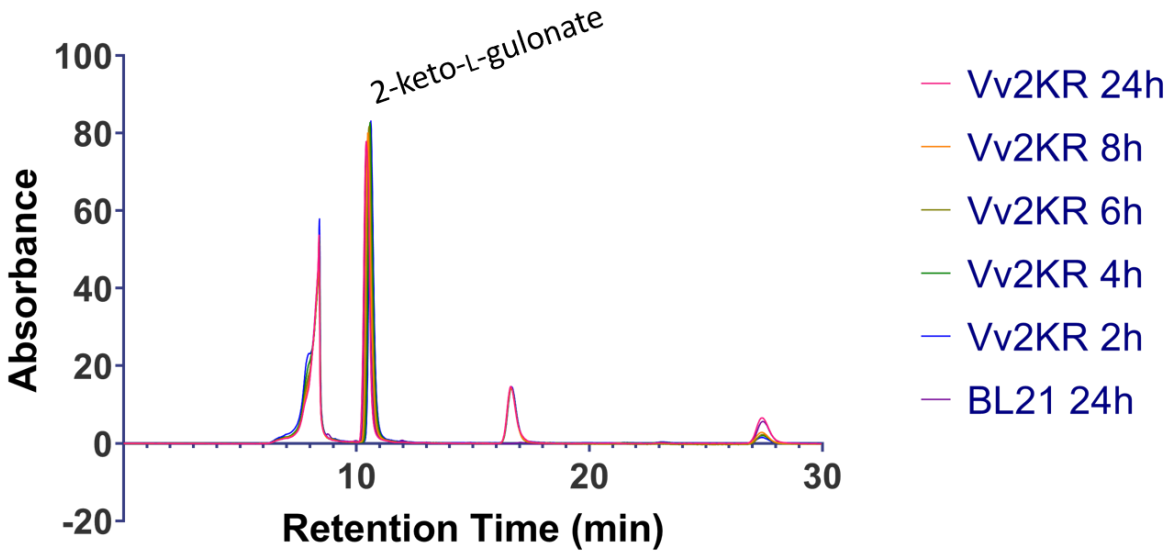


Figure 3-10 Overlay of HPLC chromatograms of whole-cell biotransformation using the previously identified Vv2KGR0. The reaction was performed using 0.090 g cell wet weight in 1 x PBS (pH 7.4) containing 0.25 M NaCl and 10 mM 2-keto-L-gulonate at 30 °C. Other peaks in the chromatogram are due to the PBS solution.

Discussion

Tartaric acid (TA) accumulates in grape berries during ripening and is essential to the winemaking process. It lowers the pH of the wine must, preventing discolouration and microbial spoilage, and confers a 'sharp' acidic taste to the wine (Kalathenos, Sutherland & Roberts 1995). Despite the importance of tartaric acid, its biosynthetic pathway has not been fully elucidated. Two enzymes have been identified in the pathway: L-idonate dehydrogenase, which catalyses the interconversion of L-idonate and 5-keto-D-gluconate (DeBolt, Cook & Ford 2006), and a 2-ketogulonate reductase (2KGR) (Burbidge 2011). Research in this area has been hindered by a lack of commercially available L-idonate. The aim of this study was to identify potential novel isoforms of 2KGR and construct a microbial system to biosynthesise L-idonate.

To identify potential *V. vinifera* 2KGRs, a BLASTp search was performed against two fungal 2KGRs, GluC and GluD to compare amino acid sequences. Although some candidates had previously been identified by Burbidge (2011), including Vv2KGR0 which was confirmed to have 2-keto-L-gulonate reductase activity *in vitro*, a number of new candidates were found. The top hits had between 35-39% identity over the open reading frame (ORF). In general, evidence suggests that sequences that share more than 40% identity are highly likely to share functional similarity, but it is difficult to reliably infer functional similarity at greater evolutionary distances (Pearson 2013). The 35-39% shared sequence identity over the evolutionary distance between fungi and grapevine

indicates these enzymes are possibly distantly related and are likely to share function. Amino acid alignments of GluD with the four 2KGR candidates selected for further investigation show short homologous regions of up to 10 amino acids in length, possibly denoting key catalytic motifs (Figure 3-11).

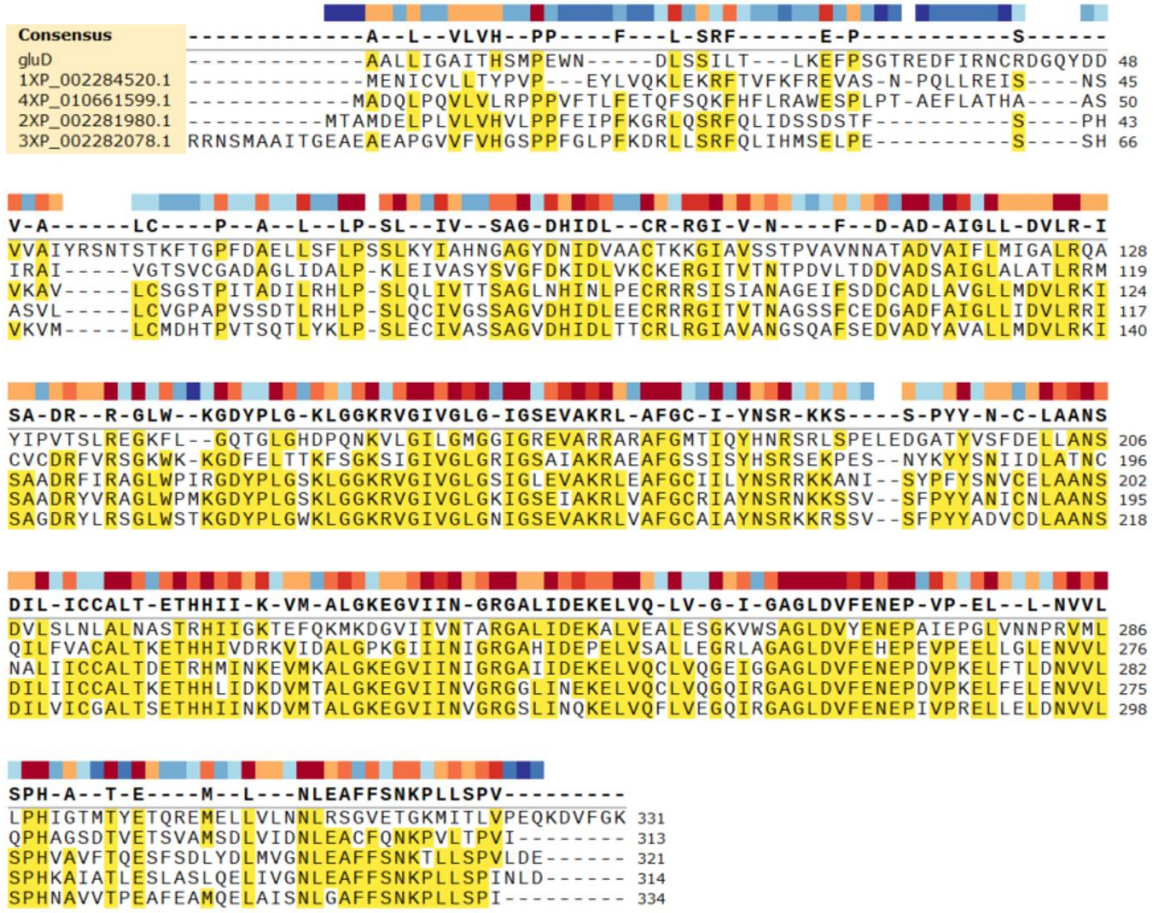


Figure 3-11 Sequence alignment of GluD and the top four 2KGR candidates selected for further investigation.

Many genes have low or no expression *in vivo,* so gene expression databases were analysed to examine the expression patterns and levels of the candidate genes. Tartaric acid (TA) accumulates in berry and leaf tissue and accumulates rapidly during early development, plateauing after ripening (Ruffner 1982; Saito & Kasai 1968). Thus, enzymes involved in TA synthesis would be expected to be expressed throughout development in the berries. Expression data for the candidate enzymes is limited, and there is little consistency across experiments in terms of tissues and time points sampled, making it hard to combine data from multiple studies. Most studies available on Expression Atlas analysed expression in different varieties or tissues rather than berry stages. The expression data presented in Table 3-4 is based on a single study by Palumbo et al. (2014) that looked at expression differences across different stages of berry development. The expression is ranked on a scale of Low-Med-High, so this data is limited

and can be used for comparative purposes only. None of the candidates displayed an expression level as high as the previously identified LIDH and Vv2KGR0 enzymes, which surprisingly were still expressed in berries ripe for harvest, a time when TA synthesis is minimal.

Current evidence suggests tartaric acid synthesis occurs in the cytoplasm. Firstly, ascorbic acid, the precursor to TA, is found in relatively high concentrations in the cytoplasm (Davey et al. 2000). Secondly, L-idonate dehydrogenase (IdnDH), the rate limiting enzyme of the primary TA biosynthetic pathway in grapes, is mainly located in the cytoplasm (Wen et al. 2010). Finally, since TA intermediates are found only in low concentrations inside the cell, the enzymes for TA synthesis are likely to co-localise to increase the local concentration gradient of substrate (Weeks & Chang 2011). Several candidates were predicted to localise to the chloroplast by the NCBI database. These proteins were discounted from future analysis as they are spatially unlikely to be involved in TA synthesis. Further characterisation of candidates was performed using TargetP to predict subcellular localisation. Candidate XP_002282078.1 was predicted to localise to the mitochondria and thus also unlikely to be involved in TA synthesis. TargetP accurately predicts the localisation of 80-90% of chloroplastic and mitochondrial proteins in *Arabidopsis thaliana*, so this prediction is likely to be valid (Leister 2003). None of the candidates were predicted to contain transmembrane helices, so they are unlikely to be membrane-bound.

These analyses suggest, based on homology to previously identified 2KGRs, expression patterns, and predicted subcellular localisation, that XP_002284520.1 and XP_002281980.1 are good candidates for 2-keto-l-gulonate reductase activity in *V. vinifera.* These candidates were selected for *in vitro* characterisation and further study (2KGR1 and 2KGR2).

Expression of 2KGR1 and 2KGR2 in *E. coli* BL21 did not produce sufficient quantities of soluble protein for purification (Figure 3-2). Even after expression conditions were optimised, neither 2KGR1 nor 2KGR2 was found to be soluble in a range of different buffers. Whole-cell biotransformation of 2-keto-L-gulonate was attempted using *E. coli* BL21 strains expressing each of the 2KGR candidates, but no difference was observed between the control strain and the 2KGR-expressing strains. (Figure 3-3). Therefore, it is likely that the proteins were catalytically inactive and sequestered in inclusion body aggregates.

It is possible to recover aggregated proteins from inclusion bodies by solubilisation with a denaturant followed by removal of the denaturant under conditions optimal for protein

folding (Lilie, Schwarz & Rudolph 1998). However, the specific folding conditions must be optimised for each protein, and the final yield of correctly folded protein may still be low (Harrison 2000). An alternative approach is to use a fusion tag to enhance the solubility of the expressed protein. An affinity-tagged (6 His) *E. coli* N-utilization substance A (NusA) protein was chosen for this purpose as it is predicted to be over 95% soluble and has been evaluated against a range of proteins (Davis et al. 1999). Expression of the 2KGR1-NusA and 2KGR2-NusA fusion proteins in *E. coli* BL21 was successful and produced high yields of soluble protein for purification (Figure 3-4).

The NusA fusion tag was not removed from the purified protein before enzymatic analysis. It is possible in some cases to use a protease to cleave the linker between the solubility tag and the protein of interest, but this can cause problems including incomplete cleavage or re-aggregation of the target protein (Esposito & Chatterjee 2006). As NusA is a large protein, its presence may however affect the activity of the 2KGRs.

Purified 2KGR1-NusA and 2KGR2-NusA were used to investigate the activity of the candidate 2KGRs against 2-keto-L-gulonate. Cofactory (Geertz-Hansen et al. 2014) was used to predict the cofactor preference of the two 2KGRs. Both were predicted to contain domains specific to NADH only. Therefore, NADH was used for analysis of enzyme activity. The reaction was monitored by measuring the change in absorbance due to oxidation of NADH. A decrease in absorbance was observed only after the addition of substrate, indicating that oxidation was due to 2KGR activity. The optimal pH conditions for each enzyme were investigated using two different buffers to ensure differences in activity were due to pH and not buffer composition. Both enzymes were found to be most active at pH 7.0 in Tris buffer (Figure 3-5). Other reaction conditions, including temperature and substrate and cofactor concentrations, were based on previous optimisation for kinetic studies on the similar Vv2KGR0 protein (Burbidge 2011).

The activities of 2KGR1 and 2KGR2 against increasing concentrations of 2-keto-L-gulonate were analysed (Figure 3-6, Figure 3-7). However, at lower substrate concentrations, the change in absorbance was too low to precisely measure, and it appears the enzymes did not reach substrate saturation. It may be necessary to use a higher concentration of enzyme and greater concentration of substrate in order to accurately model the kinetics of these enzymes. Graphpad Prism 8.0 was used to fit a Michaelis-Menten kinetic model to the data. The model had an R^2 value of 0.91 for the 2KGR1 data and 0.96 for the 2KGR2 data. While this model is a relatively good fit, the V_{max} and K_m values for 2KGR1 could not be determined within a narrow confidence interval due to lack of data at higher 2-keto-L-gulonate concentrations. A double reciprocal Lineweaver-Burk plot was used to

determine a V_{max} of 0.44 μmoles/min/mg protein and a K_m of 6.38 mM for 2KGR2. While these 2KGR enzymes appear to display Michaelis-Menten kinetics, there is evidence that the previously identified Vv2KGR0 operates under an allosteric model, although with minimal cooperation effects (Burbidge 2011). Time-course studies have shown that the dehydration of L-idonate is the rate-limiting step of tartaric acid synthesis (Malipiero, Ruffner & Rast 1987), although there may be multiple points of control along the biosynthetic pathway. In order to accurately determine the kinetics of the 2KGR1 and 2KGR2 candidates, the analysis should be repeated with higher 2-keto-L-gulonate concentrations to ensure both enzymes reach saturation. However, 40mM of substrate was sufficient to saturate the previously investigated Vv2KGR, suggesting that if higher substrate concentrations are needed to investigate the candidate 2KGR activities, the reduction of 2-keto-L-gulonate may not be their primary reaction.

The K_m of 2KGR2 was calculated as 6.38 mM ± 0.97. This is of the same magnitude as the calculated K_m of 4.67 mM for the previously investigated Vv2KGR (Burbidge 2011) and the K_m for L-idonate dehydrogenase of 2.2 mM in the rate-limiting reaction (DeBolt, Cook & Ford 2006). This suggests 2KGR2 has a similar affinity for 2-keto-L-gulonate to enzymes involved in the TA biosynthesis pathway.

To identify the products of 2KGR1 and 2KGR2 using 2-keto-L-gulonate as a substrate, an *in vitro* enzyme reaction was performed and the reaction mixture analysed by HPLC (Figure 3-8). 2-Keto-L-gulonate and L-idonate were identified based on authentic standards. Neither the 2KGR1 nor the 2KGR2 reaction produced a detectable new product peak, and the concentration of 2-keto-L-gulonate remained virtually unchanged after two hours reaction time. This suggests the 2KGR candidates have low activity against 2-keto-L-gulonate under the conditions tested. The reaction conditions may need to be optimised or performed for a longer amount of time. However, the chosen reaction conditions were based on the activity of the previously identified Vv2KGR, so it is expected that the candidate 2KGR enzymes would have performed under similar conditions.

2KGR1 was computationally predicted to be a hydroxyphenylpyruvate reductase, and 2KGR2 was predicted to be a glyoxylate/hydroxypyruvate reductase (Table 3-2). Therefore the activity of these enzymes was tested over a range of substrates found in *V. vinifera*, including glyoxylate. 2KGR1 and 2KGR2 exhibited a similar substrate range, showing activity against three out of the four substrates tested (Figure 3-9). Both 2KGR1 and 2KGR2 showed similar levels of activity against glyoxylate as 2-keto-L-gulonate, and approximately 50% lower activity against ascorbate. However, activity levels were low, suggesting these enzymes are capable only of low-level reactions, or primary substrate(s)

were not tested. Both enzymes were able to utilise ascorbic acid, the tartaric acid precursor, as substrate. To investigate if the product of this reaction is a biosynthetic precursor of tartaric acid, HPLC could be used to analyse the reaction mixture. The observed activity of both 2KGR1 and 2KGR2 against 5-keto-D-gluconate was not significant.

These results suggest both 2KGR1 and 2KGR2 display low activity as 2-keto-L-gulonate reductases, but it may not be their primary function. No evidence of L-idonate formation was found, indicating that it is unlikely that these 2KGR candidates are involved in the ascorbic acid pathway of tartaric acid biosynthesis. Both 2KGR1 and 2KGR2 display low-level, broad-spectrum reductase activity and may be involved as reductases in other metabolic pathways. Further characterisation with a wider range of substrates is required to confirm this. However, both 2KGR1 and 2KGR2 were fused with NusA to enhance their solubility, which may have affected their catalytic properties. In order to accurately characterise the activity of these 2KGRs, it may be necessary to cleave the NusA tag. Additionally, reaction conditions may be further optimised. For example, these assays were performed at 37 °C based on results from the previously characterised Vv2KGR0, but the optimal temperature for the enzymes may in fact be lower, considering they originate from a plant which would grow at a lower ambient temperature.

As neither 2KGR1 nor 2KGR2 were capable of producing large amounts of L-idonate, these enzymes were deemed unsuitable for whole-cell biotransformation of 2-keto-L-gulonate to L-idonate. A limiting factor in tartaric acid biosynthesis research to date has been the limited availability of L-idonate, a key intermediate in the pathway. Whole-cell biotransformation of 2-keto-L-gulonate to L-idonate was investigated as an alternative to synthetic approaches for the production of L-idonate as 2-keto-L-gulonate is a readily available, cheap substrate, and enzymes are able to achieve stereoselective synthesis for ease of purification. Vv2KGR0 was expressed in *E. coli* BL21 for whole-cell biotransformation. Initial attempts revealed the concentration of 2-keto-L-gulonate in the reaction mixture did not decrease after 24 hours in both the Vv2KGR0 strain and the no plasmid control strain. This was unexpected, as Vv2KGR0 has been shown to have 2-keto-L-gulonate reductase activity *in vitro*, and *E. coli* is capable of metabolising 2-keto-L-gulonate to a small degree (Yum et al. 1998). It was hypothesised that biocatalysis was restricted due to cell permeability limitations. *E. coli* is a gram-negative bacterium that possesses an outer membrane and an inner membrane. While hydrophobic molecules are able to diffuse relatively quickly though the outer lipid bilayer, transport of hydrophilic molecules is regulated by specific transport proteins on the inner membrane (Neidhardt,

Ingraham & Schaechter 1990). Sugars, amino acids, and other small hydrophilic molecules essential to growth rely on passage through porins to cross the membrane barrier. These channels are lined with positively charged amino acid residues (Nikaido 2001). This may present a cell entry challenge for 2-keto-L-gulonate.

Enhanced cell permeability was attempted using salt stress. Addition of NaCl has successfully been used to improve the uptake of crotonobetaine and D(+)-carnitine, both small, charged molecules, by *E. coli*. Cánovas et al. (2003) found production of L-(-)carnitine improved with an increase in NaCl in the reaction buffer up to a concentration of 0.5 M. However, when transformation of 2-keto-L-gulonate to L-idonate was re-attempted with the addition 0.25 M NaCl, no improve in substrate uptake was observed (Figure 3-10). This suggests that cell permeability may be a significant challenge in the whole-cell production of L-idonate from 2-keto-L-gulonate. There are several other methods that can be used to enhance cell permeability (Chen 2007). Cells may be treated with an organic solvent such as ethanol or toluene, but this can remove endogenous cofactor which needs to be re-supplied to the cell. Detergent treatment is another approach, but is harsher and may lead to cell lysis. Freeze and thaw is a simple approach but time-consuming when multiple cycles are required. Electropermeabilisation may be applied to ionic products, but requires specialised equipment. Alternatively, molecular engineering approaches may be used to modify the membrane structures or display catalysts on the cell surface, circumventing the need for cell entry. These are all viable options to enhance 2-keto-L-gulonate cell entry, but the process may require extensive optimisation. Once cell entry is achieved, L-idonate production may be further increased by knocking out metabolic pathways that consume L-idonate (Bausch et al. 1998) and considering *in situ* cofactor regeneration (Kara, Schrittwieser & Hollmann 2013).

Another approach could be to use a cell-free system. Although a whole cell system is preferred since it circumvents the need for extensive extraction and purification and the addition of expensive coenzyme, an *in vitro* reaction with purified enzyme may be more straightforward in this case. Finally, if low yields are experienced using Vv2KGR0, *A. niger* GluD or GluC may be substituted to optimise the reaction. More work is required to successfully achieve microbial biosynthesis of L-idonate, but the number of viable avenues are plentiful.

Conclusions

Two potential *V. vinifera* 2-keto-L-gulonate reductases were identified based on homology to recently characterised *A. niger* 2-keto-L-gulonate reductases that are able to produce L-idonate, a key intermediate in tartaric acid biosynthesis. The two 2KGRs were found to

have low level activity using 2-keto-L-gulonate as a substrate, but did not produce L-idonate. These enzymes exhibited low-level broad-spectrum reductase activity, although their primary substrate(s) may not have been tested. It is unlikely that these enzymes are involved in the biosynthesis of tartaric acid from ascorbate in grapevine.

Further research into the biosynthesis of tartaric acid in grapevine may be assisted by providing access to L-idonate. Whole-cell biotransformation of 2-keto-L-gulonate to L-idonate may be a viable option for the production and purification of L-idonate, but the process requires further testing and optimisation to overcome the challenge of cell entry. Alternatively, i*n vitro* production of L-idonate with purified or resin-bound enzyme may be a faster and more efficient option for L-idonate synthesis from 2-keto-L-gulonate.

Supplementary Information for Identification and Characterisation of Potential Grapevine 2-Ketogulonate Reductases

Vv2KGR.0 (Burbidge 2011)

TTTGGAGGAGCAGCATCCCTTCCCTTTGCCTTGCCTTCCATTTATACCGATGAGTCGCCAGCTT
ATCTAAAACCAGCGAAGTCCCCAAAGATATCGGACGGCTCATCCGCCCCACCGTCACCGCCGCT
GCACTCCCACCTCCGTGCGCCGCCGAGTGTATAGTGCGTCTTTCACAAAGCGAAACCATGGAAA
GCATCGGGGTACTGTTGACTTGCCCAATGAACCCATACCTGGAACAGGAACTGGACAAGCGCTT
CAAGCTCTTCCGCTTCTGGGACTTTCCAAGCGCCAACGATCTTTTCAGGGAGCATTCAAATTCG
ATCCGAGCTGTGGTTGGAAACTCCTTCATCGGCGCCGACGCCCAGATGATCGAGGCGTTGCCCA
AGATGGAGATTGTGTCGAGTTTCAGCGTTGGGTTGGACAAGATCGATTTGGTGAGGTGCAAGGA
GAAGGGAATTAGGGTTACGAACACTCCGGATGTGCTGACGGAGGACGTGGCGGACTTGGCACTT
GCTTTGATTTTGGCGACTCTGAGACGTATTTGTGAAAGTGATCGTTATGTGAGGAGTGGGTCGT
GGAAGAAAGGGGATTTCAAGTTGACTACCAAGTTCACTGGAAAATCAGTTGGCATTATAGGGTT
GGGTAGGATTGGCTCAGCAATTGCCAAGAGAGCCGAGGGATTTAGCTGTCCAATTAGTTACCAT
TCCAGAACAGAGAAACCAGGGACAAACTACAAGTACTATCCTAGTGTCGTTGAATTGGCCTCCA
ACTGTCAAATCCTGGTTGTTGCTTGCGCGTTAACACCAGAAACCCGCCACATCATCAACCGTGA
AGTCATCAATGCACTGGGTCCAAAGGGTGTGGTCATCAACATCGGAAGGGGATTACATGTGGAT
GAACCTGAGCTTGTATCCGCACTGGTTGAAGGCCGGTTGGGAGGTGCTGGACTTGATGTGTTTG
AAAATGAGCCTAATGTACCTGAAGAGCTGTTAGCAATGGACAATGTAGTCCTTTTGCCTCATGT
AGGAAGCGGAACGGTGGAAACCCGGAAAGACATGGCTGACCTGGTACTTGGAAACTTAGAGGCT
CACTTTCTGAACAAACCACTGTTAACTCCAGTGGTTTAATTGTCATGAGAAGGTATGTAGCTCA
TCATTCATTGAAGTCTTTTCTTAATGTTGCTCAATGAACCACTCGATGTTTTCTCTGCATTGTT
TCCTCAACAACTTGGTATTTATGTTATGATGAATGGAAGACCTTAATGACTACTAAATTGTCTC
TTAATGTTTCA

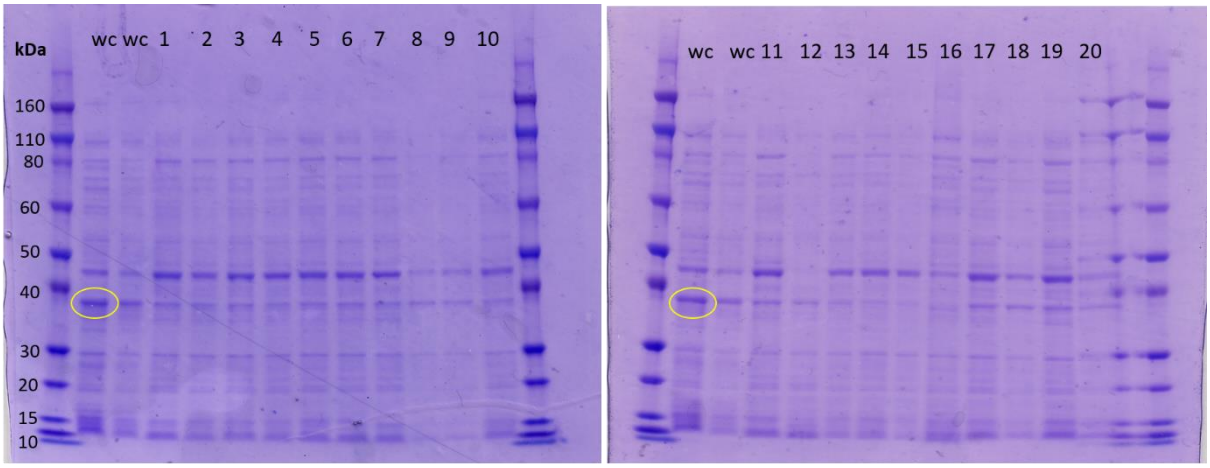
Vv2KGR.1 Gene Sequence (Codon Optimised, 6-His-tagged, Ordered)

CATATGGAAAACATCTGTGTCTTATTGACGTACCCAGTGCCCGAATACTTGGTCCAAAAATTGG
AGAAACGGTTCACCGTCTTTAAATTCCGGGAAGTCGCAAGTAACCCACAGCTGCTGCGCGAAAT
CAGCAATAGCATAAGAGCCATTGTAGGTACCAGTGTATGCGGAGCGGATGCCGGGCTTATCGAT
GCCCTGCCCAAATTAGAGATTGTGGCGTCGTACAGTGTAGGTTTCGACAAAATAGACTTGGTAA
AATGTAAGGAAAGAGGTATTACAGTAACAAACACGCCAGACGTACTGACGGACGACGTTGCGGA
TTCTGCTATCGGATTAGCACTTGCTACGCTGCGCCGTATGTGTGTTTGCGATCGCTTTGTTAGA
AGTGGTAAGTGGAAAAAGGGCGACTTTGAGTTGACAACGAAGTTTTCCGGCAAGAGCATAGGGA
TCGTCGGCCTTGGAAGAATTGGGTCTGCCATTGCCAAGAGAGCTGAGGCCTTCGGCTCTTCAAT
ATCATACCACAGCCGTTCGGAAAAACCAGAGTCAAATTATAAGTATTACTCGAACATTATAGAC

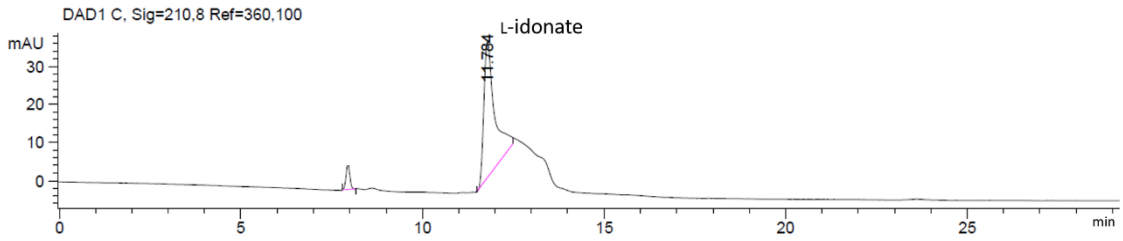
TTAGCTACGAATTGTCAGATATTGTTCGTCGCGTGTGCCCTGACTAAGGAGACTCATCACATTG
TAGATCGTAAAGTCATCGACGCGCTGGGACCCAAGGGTATTATAATAAACATTGGGCGTGGCGC
TCATATAGACGAGCCTGAGTTGGTGAGCGCACTGCTTGAAGGAAGATTGGCCGGAGCCGGTTTG
GATGTCTTCGAACACGAGCCGGAGGTCCCTGAAGAGCTGTTGGGTCTTGAGAACGTTGTACTTC
AGCCACACGCCGGGTCTGACACTGTCGAAACGTCTGTAGCAATGTCGGATCTTGTCATAGACAA
CTTGGAGGCGTGTTTCCAAAATAAACCCGTGTTAACACCAGTAATTCATCATCACCATCATCAC
TAAGGATCC

Vv2KGR.2 Gene Sequence (Codon Optimised, 6-His-tagged, Ordered)

CATATGACTGCTATGGATGAGTTGCCTCTTGTCTTGGTTCATGTCTTGCCTCCTTTTGAAATTC
CGTTCAAAGGCCGTTTGCAAAGCCGTTTTCAACTGATCGACTCGAGTGATTCGACGTTTAGTCC
GCACGCCTCTGTGCTGTTGTGTGTCGGTCCTGCTCCTGTTTCAAGCGATACCTTACGTCACTTA
CCTTCGTTGCAATGTATAGTGGGCAGTTCAGCAGGAGTAGACCATATTGACTTGGAGGAGTGTC
GCCGTCGCGGTATAACTGTGACTAACGCGGGCTCGTCCTTCTGTGAAGACGGGGCAGACTTTGC
GATTGGGCTTCTGATCGATGTGTTACGCCGCATTTCGGCTGCTGATCGCTACGTGAGAGCGGGA
TTGTGGCCCATGAAAGGAGACTACCCGCTTGGGTCTAAGCTGGGAGGCAAGCGTGTGGGCATTG
TTGGGTTGGGGAAGATCGGGTCAGAAATAGCAAAAAGACTGGTAGCTTTCGGTTGTCGGATCGC
CTACAATAGTCGTAACAAGAAGTCGTCAGTGTCATTTCCCTACTACGCCAACATTTGCAATCTT
GCTGCGAACAGCGACATCCTTATAATATGCTGCGCTCTTACCAAGGAAACACACCACCTTATCG
ATAAGGATGTGATGACCGCATTAGGCAAAGAAGGTGTCATTATCAACGTTGGACGCGGCGGTTT
AATCAACGAGAAAGAGTTAGTGCAATGTCTTGTACAGGGCCAAATACGGGGAGCGGGCCTTGAC
GTCTTCGAGAACGAGCCGGATGTTCCGAAGGAATTATTCGAATTAGAGAATGTTGTATTGTCCC
CTCATAAAGCCATAGCGACCTTAGAATCGTTGGCGTCCTTGCAGGAGCTGATAGTTGGTAATTT
GGAAGCGTTCTTTTCCAATAAGCCATTACTTAGTCCCATTAATCTTGACCATCACCATCACCAT
CACTAAGGATCC



Supplementary Figure 3-1 Solubility test for 2KGR1. WC: whole cell samples (insoluble fraction). 1-20 are different buffers used to test solubility. 1: 20 mM Tris 8.0, 0.5 M NaCl, 10 mM imidazole, 2mM β-mercaptoethanol (BME). 2: 20 mM Tris 8.0, 0.5 M NaCl, 10 mM imidazole, 2mM BME, 1mM zinc sulphate. 3: 20 mM Tris 8.0, 0.5 M NaCl, 10 mM imidazole, 2mM BME, 1mM MgCl2 4: 20 mM Tris 8.0, 0.5 M NaCl, 10 mM imidazole, 2mM BME, 1mM CaCl2. 5: 20 mM Tris 8.0, 0.5 M NaCl, 10 mM imidazole, 2mM BME, 1mM NADH. 6: 20 mM Tris 8.0, 0.5 M NaCl, 2mM BME. 7: 20 mM Tris 8.0, 0.5 M NaCl, 15mM BME. 8: 20 mM Tris 8.0, 0.5 M NaCl, 10mM imidazole, 2mM BME, 20% glycerol. 9: 20 mM Tris 8.0, 0.5 M NaCl, 10mM imidazole, 2mM BME, 10% glycerol. 10: 20 mM HEPES 7.0, 0.5 M NaCl, 2mM BME. 11: 20 mM Sodium Citrate 5.5, 0.5 M NaCl, 2mM BME. 12: 50 mM Sodium Acetate, 0.5 M NaCl, 2mM BME. 13: 20 mM Tris 8.0, 0.2 M NaCl, 10 mM imidazole, 2mM BME. 14: 20 mM Tris 8.0, 20 mM NaCl, 10 mM imidazole, 2mM BME. 15: 20 mM Tris 8.0, 2mM BME. 16: 100 mM Sodium Phosphate 7.5, 0.5 M NaCl, 10 mM imidazole, 2mM BME. 17: 20 mM Tris 8.0, 0.5 M NaCl, 10 mM imidazole, 2mM BME, 0.1% tween. 18: 20 mM Tris 8.0, 0.5 M NaCl, 10 mM imidazole, 2mM BME, 0.5% sucrose. 19: 20 mM Tris 8.0, 0.5 M NaCl, 10 mM imidazole, 2mM BME, 0.1 M urea. 20: 50 mM NaH2PO4 pH 8.0, 300 mM NaCl, 10 mM imidazole, 15% glycerol. 2KGR1 is circled in yellow.



Supplementary Figure 3-2 HPLC chromatogram for L-idonate standard (1.0 mg/mL).

Chapter 4 Production of Grapevine Sesquiterpenes and Sesquiterpenoids in Yeast

Introduction

Terpenoids are one of the largest classes of plant metabolites and perform many ecological roles, including plant defence, attraction of pollinators, and primary plant metabolism (Bohlmann & Keeling 2008). Terpenes are biosynthesised from five-carbon prenyldiphosphate precursors, and these simple building blocks are transformed into a large number of diverse structures through the action of terpene synthases (May, Lange & Wüst 2013). The hydrocarbon terpene backbone may be further diversified by the action of cytochrome P450 monoxygenases or various transferases (O'Maille et al. 2008).

Sesquiterpenoids are 15-carbon terpene derivatives that are synthesised by sesquiterpene synthases using farnesyl pyrophosphate (FPP) as substrate. Thousands of sesquiterpenoids have been identified, although not all have been fully structurally elucidated (Duhamel et al. 2018). Sesquiterpenoids are important contributors to the flavour and aroma of many plants (Dudareva & Pichersky 2000, 2006). A wide variety of sesquiterpenoids have been identified in grapevine (*Vitis vinifera)* and wine (Coelho et al. 2006; Schreier, Drawert & Junker 1976). However, the contributions of many of these compounds to the aroma and character of wine are as yet unrecognised. The exception is rotundone, a sesquiterpene ketone derived from α-guaiene, which has a potent spicy, peppery aroma, and has been found in some cultivars of Shiraz and several other varieties (Wood et al. 2008). Additionally, there has been recent interest in the potential health benefits of wines containing sesquiterpenoids (Rocha et al. 2006; Vinholes 2013). The immense structural diversity of sesquiterpenoids and their unique properties suggest there is more to be learned about sesquiterpenoids in grapes and wine.

A number of *V. vinifera* sesquiterpene synthases have been previously identified through genome mining and transcriptome analysis (Martin et al. 2010; Martin et al. 2012; Sweetman et al. 2012); however, not all have been characterised *in vivo.* Previous investigations have shown that occasionally sesquiterpene synthases may produce different products *in vitro* than *in vivo* (Salvagnin et al. 2016), so it is important to use a number of systems to accurately characterise these enzymes. By characterising the products of grapevine sesquiterpene synthases it may be easier to identify sesquiterpenoids in wine.

However, *in planta* characterisation requires skill, time and resources to create and grow transgenic plants, and sensitive analysis techniques to separate and identify sesquiterpenoids amongst other plant metabolites. An alternative is to use a microbial expression system that is able to produce the sesquiterpene precursor FPP, and overexpress the sesquiterpene synthase of interest. The advantages of using microbial biosynthesis are a greater concentration of product, allowing facile analysis and identification, and rapid investigation, allowing a number of enzymes to be analysed and compared. Furthermore, the cellular milieu mimics the native protein environment, making the results more applicable to *in planta* conditions than *in vitro* experiments using recombinant protein. EPY300 is a yeast strain based on the S*accharomyces cerevisiae* strain S228C that has been genetically engineered to overproduce FPP (Ro et al. 2006). This strain may be used to express sesquiterpene synthases and is capable of product titres in the range of several mg/L in optimised strains (Rodriguez et al. 2014). This platform allows co-expression of P450 monooxygenases capable of modifying the sesquiterpene hydrocarbon backbone. Investigation of grapevine sesquiterpene synthases using microbial systems in conjunction with P450s may lead to the discovery of new products and novel wine aroma compounds.

The aim of this study was to investigate and identify products of six grapevine sesquiterpene synthases and their potential oxygenated derivatives. To achieve this, the sesquiterpene synthase genes were expressed in EPY300 alongside $P450_{BM3}$. $P450_{BM3}$ is unusual within the P450 family as it is fused with a reductase, negating the need to express an additional, third enzyme and increasing the catalytic efficiency of the P450 (Govindaraj & Poulos 1995; Nguyen, MacNevin & Ro 2012). A mutant $P450_{BM3}$ was used as it is a promiscuous enzyme that can modify many substrates and may assist in identifying sesquiterpenoids able to be produced *in planta* by grapevine P450s.

Results

Six sesquiterpene synthases were cloned from *Vitis vinifera* cv. Shiraz for use in this study (Table 4-1). To identify products of the sesquiterpene synthases and investigate the reactivity of the mutant $P450_{BM3}$ with those products, each sesquiterpene synthase was expressed in EPY300 yeast with either the mutant $P450_{BM3,}$ or the wild-type $P450_{BM3}$ as a control. The empty pESC-Leu2d vector and vectors containing only the wild-type $P450_{BM3}$ and mutant $P450_{BM3}$ respectively were used as controls.

Table 4-1 Accession numbers of genes characterised in this study.

Gene	Accession #	Functional Gene ID
TPS07	HM807377	VvGwGerD
TPS24	XM_002282452	VvGuaS
TPS26	HM807407	VvPNCuCad
TPS27	HM807374	VvGwECar2
TPSY1	XM_003634648.1	
TPSY2	XM_002263544.2	

Sesquiterpenes were identified in the yeast culture organic extracts by GC-MS analysis. Although the concentrations of sesquiterpenoids in samples were relatively low, chromatography showed clear separation and sharp peaks. Total Ion Chromatograms (TIC) of sample extracts were compared to control extracts to identify compounds unique to the sample culture (Figure 4-1). All sample TICs are available in Supplementary Figures.

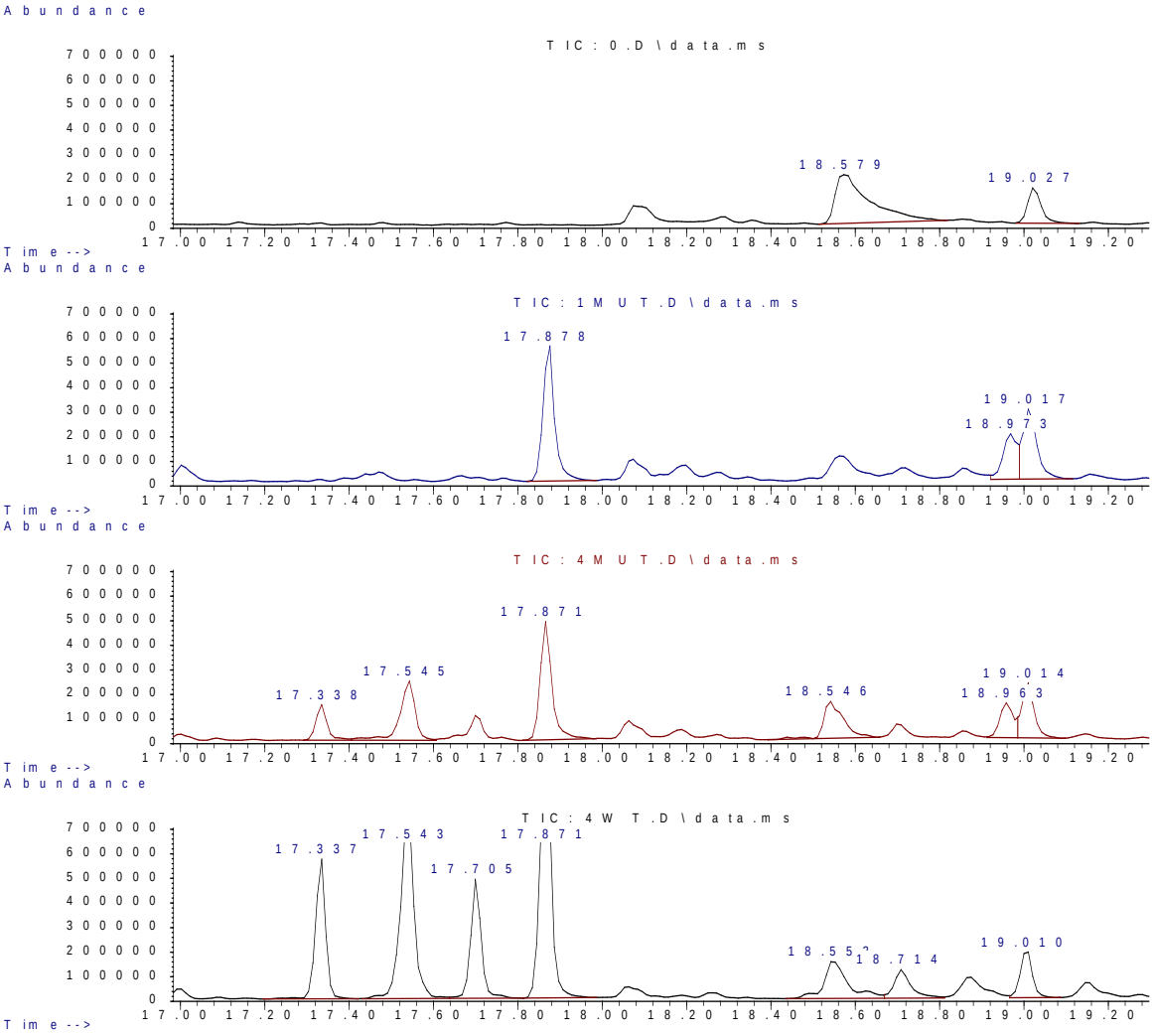


Figure 4-1 Total Ion Chromatograms (TIC) of empty vector (top) and P450$_{BM3}$ mutant only (second from top) culture extracts compared to sample culture extracts (second from bottom and bottom). The comparison reveals

unique products that are present in the sample culture extracts. Chromatography showed good separation and sharp peaks.

Sesquiterpenoid compounds were identified by comparison of the mass spectra to the National Institute of Standards and Technology (NIST) library using NIST MS Search 2.2. The NIST match factors and probability values were used to assess library hit results. In general, a value of 900 or greater is considered an excellent match and 800-900 a good match (*NIST Standard Reference Database 1A User's Guide* 2008). Match factor values are given in brackets after the name of the library match compound.

The major product of *TPS07* was germacrene D, but cadinene-type structures and caryophyllene-type structures were also identified (Table 4-2). Several sesquiterpene oxides present in the sample were not found in the NIST library, including one product produced only by the $P450_{BM3}$ mutant.

Table 4-2 Sesquiterpenoid products of EPY300 expressing TPS07. *The NIST match values for library compounds are given in brackets.*

Retention Time (min)	Sample	Library Match
14.34	BM3 Wild Type and BM3 Mutant	germacrene D (916); isogermacrene D (880)
17.70	BM3 Wild Type and BM3 Mutant	α-cadinol (909)
19.72	BM3 Wild Type and BM3 Mutant	$C_{15}H_{24}O$
20.03	BM3 Mutant only	$C_{15}H_{24}O$
20.48	BM3 Wild Type and BM3 Mutant	caryophyllene oxide-like[1] $C_{15}H_{24}O$

α-Bisabolol was identified in the TPS24 sample. (Table 4-3). Pogostol was produced by the $P450_{BM3}$ mutant.

Table 4-3 Sesquiterpenoid products of EPY300 expressing TPS24. *The NIST match values for library compounds are given in brackets.*

Retention Time (min)	Sample	Library Match
17.72	BM3 Mutant only	pogostol (842)
18.06	BM3 Wild Type and BM3 Mutant	α-bisabolol (922)

TPS24 has been previously characterised as a guaiene synthase (Drew et al. 2016). However, neither α-guaiene nor δ-guaiene were found in the sample. As these are volatile

[1] 'Caryophyllene oxide-like' refers to a compound that shares similar mass spectral features to caryophyllene oxide but is not a close enough match to any compound in the NIST library to be positively identified. Due to the similarity in the mass spectra, it is likely these compounds contain similar substructures as caryophyllene oxide.

sesquiterpenoids, EPY300 culture was repeated using a dodecane overlay (10% of culture volume) to trap volatile compounds. α-Guaiene, δ-guaiene, and a third guaiene-like sesquiterpene were present in the sample, as well as pogostol and α-bisabolol, which were identified in the non-dodecane overlayed culture (Table 4-4).

Table 4-4 Sesquiterpenoid products of dodecane overlay from culture of EPY300 expressing TPS24. The NIST match values for library compounds are given in brackets.

Retention Time (min)	Sample	Library Match
14.74	BM3 Wild Type and BM3 Mutant	α-guaiene (951)
15.68	BM3 Wild Type and BM3 Mutant	δ-guaiene (940)
15.88	BM3 Mutant only	guaiene-like $C_{15}H_{24}$
17.72	BM3 Mutant only	pogostol (909)
18.06	BM3 Wild Type and BM3 Mutant	α-bisabolol (935)

TPS26 produced mainly cadinene and cubebene-type structures, with two unidentified oxygenated sesquiterpenoids present in the mutant $P450_{BM3}$ extract (Table 4-5).

Table 4-5 Sesquiterpenoid products of EPY300 expressing TPS26. *The NIST match values for library compounds are given in brackets.*

Retention Time (min)	Sample	Library Match
15.76	BM3 Wild Type and BM3 Mutant	δ-cadinene (905)
16.10	BM3 Mutant only	$C_{15}H_{24}O$
17.33	BM3 Wild Type and BM3 Mutant	di-epi-1,10-cubenol (919)
17.54	BM3 Wild Type and BM3 Mutant	epicubenol (924)
17.70	BM3 Wild Type and BM3 Mutant	α-cadinol (930)
22.20	BM3 Mutant only	$C_{15}H_{24}O_2$

TPS27 produced predominantly caryophyllene-like structures and α-cadionol (Table 4-6). The $P450_{BM3}$ mutant produced four oxygenated sesquiterpenoids.

Table 4-6 Sesquiterpenoid products of EPY300 expressing TPS27. *The NIST match values for library compounds are given in brackets.*

Retention Time (min)	Sample	Library Match
14.20	BM3 Wild Type and BM3 Mutant	β-caryophyllene (955)
14.73	BM3 Wild Type and BM3 Mutant	α-caryophyllene (945)

17.70	BM3 Wild Type and BM3 Mutant	α-cadinol (907)
19.82	BM3 Mutant only	caryophyllene oxide-like $C_{15}H_{24}O$
20.01	BM3 Mutant only	$C_{15}H_{24}O$
20.69	BM3 Mutant only	clovanediol-like $C_{15}H_{26}O_2$
21.38	BM3 Mutant only	caryophyllene oxide-like $C_{15}H_{24}O$

To confirm the identity of the peaks at 14.20 min and 14.73 min, authentic standards of β-caryophyllene and α-caryophyllene were analysed using the same GC-MS method. The elution times of these compounds and the mass spectra matched those of the peaks at 14.20 min and 14.73 min respectively (Supplementary Information).

To investigate whether the subsequent peaks were products of the sesquiterpene synthase or the result of caryophyllene oxidation, cultures of EPY300 expressing $P450_{BM3}$ were spiked with 0.5 μg/mL α-caryophyllene or β-caryophyllene. This produced a number of caryophyllene oxide derivatives, many of which were not present in the TPS27 sample in significant concentrations (Figure 4-2).

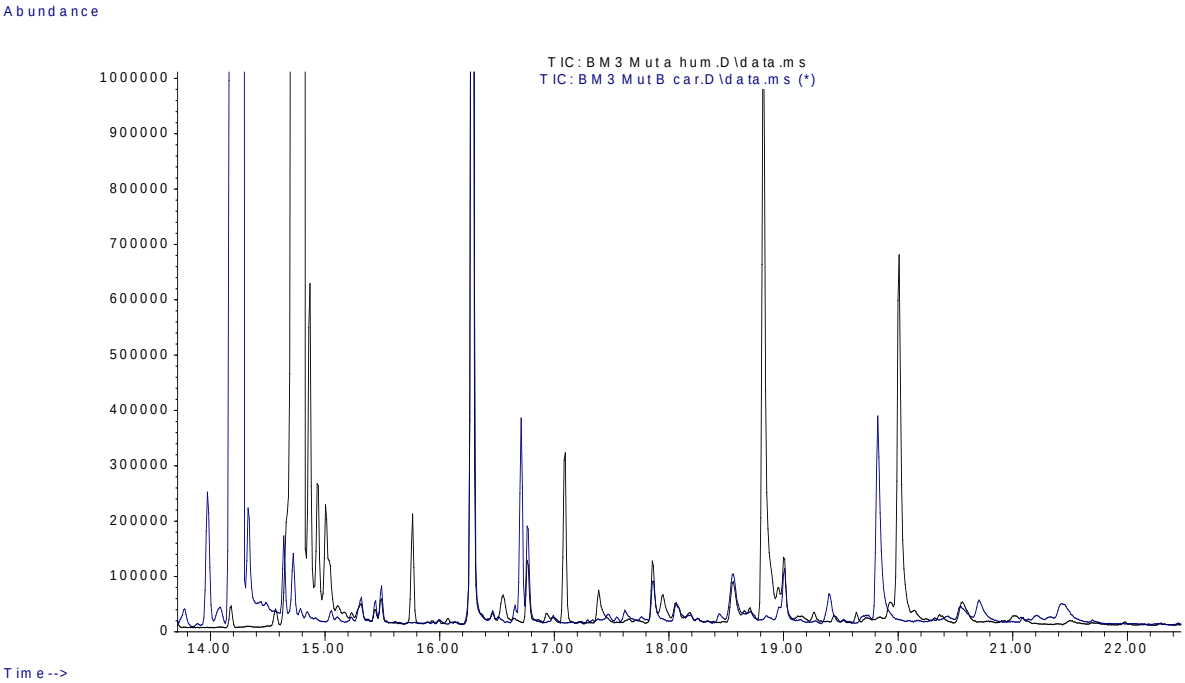


Figure 4-2 Overlay of gas chromatogram culture extract from EPY300 expressing BM3 spiked with α-caryophyllene (black) and β-caryophyllene (blue)

The compounds present at 17.70 min, 20.01 min 20.69 min and 21.38 min were not found in either the α-caryophyllene- nor β-caryophyllene-spiked samples. A peak at 19.82 min

was present in the β-caryophyllene-spiked sample. The mass spectra of the peaks at 19.82 min in the β-caryophyllene-spiked sample and the TPS27 sample were found to share significant similarities (Figure 4-3).

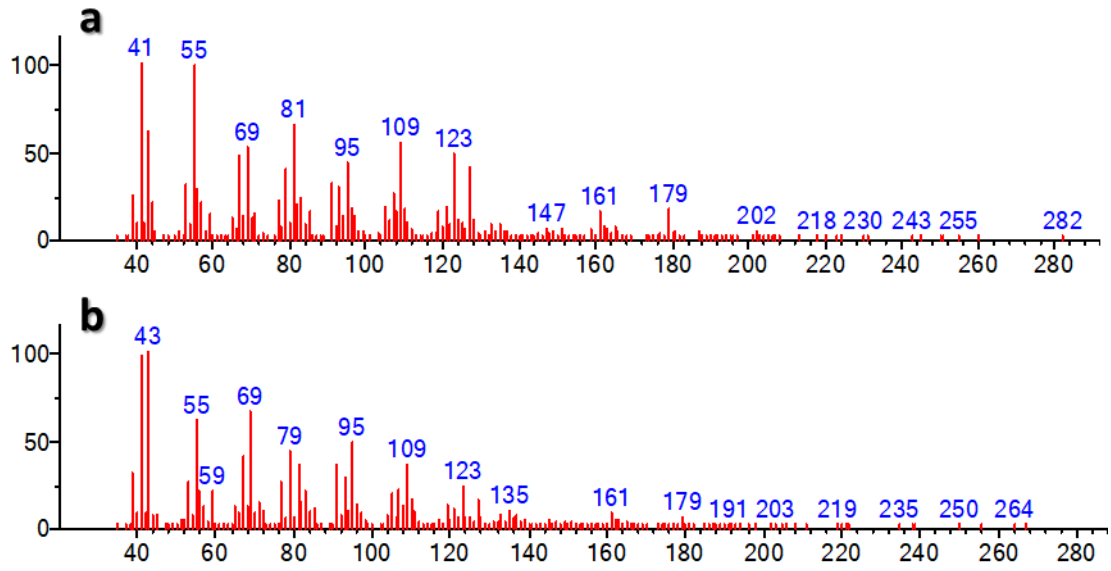


Figure 4-3 Comparison of mass spectra at 19.82 min from the extract of EPY300 expressing Mutant $P450_{BM3}$ spiked with β-caryophyllene (a) and the extract of EPY300 expressing TPS27 with Mutant $P450_{BM3}$ (b).

TPSY1 produced a range of structurally varied compounds, including cubebene-like, germacrene-like and cadinene-like sesquiterpenoids (Table 4-7). All compounds were found in both the mutant $P450_{BM3}$ and wild-type $P450_{BM3}$ extracts.

Table 4-7 Sesquiterpenoid products of EPY300 expressing TPSY1. *The NIST match values for library compounds are given in brackets.*

Retention Time (min)	Sample	Library Match
14.16	BM3 Wild Type and BM3 Mutant	β-ylangene (896)
14.33	BM3 Wild Type and BM3 Mutant	β-cubebene (914); β-copaene (901)
14.58	BM3 Wild Type and BM3 Mutant	isogermacrene D (882)
14.71	BM3 Wild Type and BM3 Mutant	germacrene D-like $C_{15}H_{24}$
15.15	BM3 Wild Type and BM3 Mutant	germacrene D (936)
15.77	BM3 Wild Type and BM3 Mutant	δ-cadinene (883)
17.25	BM3 Wild Type and BM3 Mutant	juneol (923)
17.52	BM3 Wild Type and BM3 Mutant	α-cadinol-like $C_{15}H_{26}O$
17.70	BM3 Wild Type and BM3 Mutant	α-cadinol (940)

20.02	BM3 Wild Type and BM3 Mutant	aromadendrene oxide-like $C_{15}H_{24}O$
20.47	BM3 Wild Type and BM3 Mutant	aromadendrene oxide-like $C_{15}H_{24}O$

TPSY2 also produced a range of compounds including caryophyllenes, cadinene and cubenols (Table 4-8). All compounds were present in both the mutant $P450_{BM3}$ and wild-type $P450_{BM3}$ extracts.

Table 4-8 Sesquiterpenoid products of EPY300 expressing TPSY2. *The NIST match values for library compounds are given in brackets.*

Retention Time (min)	Sample	Library Match
14.18	BM3 Wild Type and BM3 Mutant	β-caryophyllene (931) or isocaryophyllene (926)
14.73	BM3 Wild Type and BM3 Mutant	α-caryophyllene (942)
15.75	BM3 Wild Type and BM3 Mutant	δ-cadinene (923)
17.33	BM3 Wild Type and BM3 Mutant	di-epi-1,10-cubenol (906)
17.54	BM3 Wild Type and BM3 Mutant	epicubenol (907)

Discussion

In this study, the products of several *Vitis vinifera* sesquiterpene synthases and their derivatives were investigated and identified using a yeast *in vitro* assay. A promiscuous P450, $P450_{BM3,}$ was included to identify oxygenated derivatives of the sesquiterpene synthase products. This may provide insight into sesquiterpenoids that can be produced through oxidation *in planta* by grapevine P450s. A wild-type, non-promiscuous $P450_{BM3}$ was used as a control. Since wild-type $P450_{BM3}$ is selective, the compounds produced by the sesquiterpene synthases are not expected to be oxidised, so the products would be essentially the same as if the sesquiterpene synthases were expressed without a P450. The aim of the study was to investigate and identify products of six sesquiterpene synthases and their potential oxygenated derivatives.

Previous studies have shown that using the metabolically engineered EPY300 yeast strain with the high-copy pESC Leu2d plasmid can produce significant amounts of sesquiterpenes and oxygenated sesquiterpenes, although yield can vary dramatically (Nguyen, MacNevin & Ro 2012). Here, the engineered EPY300 strain was able to produce sesquiterpenes in detectable amounts, but the yield was relatively low. Previously, EPY300 cultures have been overlayed with dodecane to capture volatile sesquiterpene

hydrocarbons(Rodriguez et al. 2014), but in this study a dodecane overlay was not used for every sample as it has been observed to interfere with the production of oxygenated sesquiterpenoids (Nguyen, MacNevin & Ro 2012). This may have led to the loss of volatile sesquiterpenes and decreased the product yield. If this method were to be used for the production and purification of sesquiterpenoids on a large scale, modifications could be made to improve the yield. Codon-optimisation or protein modifications could be made to improve the expression and stability of the sesquiterpene synthase in yeast. Rich media (YPD) may be used in place of synthetic media, which would increase the rate of cell growth and metabolism, but could lead to plasmid loss over time due to lack of selection pressure (Nguyen, MacNevin & Ro 2012). Alternatively, sesquiterpene synthase genes could be genomically integrated. This would reduce the gene copy number but increase stability and allow for the engineering of highly-stable, productive, sesquiterpenoid cell factories. Nevertheless, using the current method, sufficient amounts of product were present in the extracts to enable GC-MS analysis and identification with a good signal-to-noise ratio (Supplementary Figures).

The NIST library was used to identify sesquiterpenoids based on their mass spectra (MS). The search compares the mass spectrum of the unknown compound to compounds in the NIST library and assigns a match score based on how closely the mass-to-charge ratios and relative abundance of each peak are related. A score of 1000 correlates to a perfect match, ranging to zero for nothing in common (Stein 1994). In general, a value of 900 or greater is considered an excellent match, 800-900 a good match, 700–800 a fair match and less than 600 a very poor match. Each match is also assigned a probability value which estimates the probability that the match is correct, based on the similarity of the other matches in the hit list and the probability of the unknown compound existing in the library (*NIST Standard Reference Database 1A User's Guide* 2008; Stein 1994).

The match value and probability value were used to assess the library hit results. Match values over 900 generally indicate a positive hit, and it is possible to be confident in the result. However, compounds listed with match values under 900 did show noticeable differences between the unknown MS and the library MS (e.g. Supplementary Figure 4-5). It is possible that compounds with match values less than 900 are very similar to the library match but vary slightly in their structure. Currently there are 191,436 spectra in the main library (*NIST Standard Reference Database 1A User's Guide* 2008). Due to the size of the library, it provides a good basis for the identification of common compounds, but is not an exhaustive list, so some compounds remain unidentified. Furthermore, retention times were not taken into account for product identification, as they vary depending on

the experiment set up. Overall, the library is a useful tool that can lead to the positive identification of common products and closely-related structures but should be used with caution.

TPS07 produced germacrene D, with cadinene-type structures and caryophyllene-type structures also present (Table 4-2). TPS07 is similar to VvGwGerD, which has been shown to produce germacrene D as a major product and δ-cadinene as a minor product (8%) *in vitro* (Lücker, Bowen & Bohlmann 2004; Martin et al. 2010). This confirms that germacrene D is likely correctly identified in the sample, and supports the presence of cadinene-type structures. The caryophyllene-type structure is likely not correctly identified as it is not consistent with previously identified products of TPS07, and the true structure may be a germacrene D or cadinene derivative with formula $C_{15}H_{24}O$ (MW 220). Two other sesquiterpenoid compounds with MW 220 and formula $C_{15}H_{24}O$ were also present in the extract but no corresponding library structure was found. Substructures identified on the basis of the MS fragmentation pattern suggest these compounds may be epoxide derivatives of germacrene D or δ-cadinene. Compounds with similar MW and retention times were also identified in the TPSY1 extract.

TPS24 produced α-bisabolol. The mutant $P450_{BM3}$ produced pogostol and a second unidentified product with formula $C_{15}H_{26}O$. TPS24 has two known alleles with distinct product profiles: VvPNSeInt and VvGuaS (Drew et al. 2016). VvPNSeInt has been shown to produce δ-selinene, selinene-type products, intermedeol and α-guaiene *in vivo* in metabolically engineered *E. coli* (Martin et al. 2010). VvGuaS expressed in *Nicotiana benthamiana* leaves produced α-guaiene and δ-guaiene as its major products, with several minor products, including pogostol (3%). The TPS24 allele used for this study was VvGuaS cloned from Shiraz; however, neither α-guaiene nor δ-guaiene were detected in the TPS24 sample extract. This could be due to a number of reasons. Firstly, a dodecane overlay was not used initially on the yeast culture, so α-guaiene and δ-guaiene, being volatile sesquiterpene hydrocarbons, may have evaporated from the sample. Secondly, TPS24 may have low activity in EPY300.The gene was not codon-optimised for yeast so it may have low levels of expression, resulting in low and undetectable product concentrations. However, the presence of pogostol in the mutant $P450_{BM3}$ sample suggests the existence of other compounds with 5,7-bicyclic ring structures. Pogostol is likely a derivative of delta guaiene, given their structural similarity, (Figure 4-4). This suggests δ-guaiene present in the culture may have been oxidised by $P450_{BM3}$ to pogostol, a more hydrophilic and less volatile compound that would not have escaped from the culture solution as easily. This

could be confirmed by feeding EPY300 expressing $P450_{BM3}$ only with δ-guaiene to determine if pogostol is produced by the P450.

To test whether α-guaiene and δ-guaiene had evaporated from the sample, culture of EPY300 expressing TPS24 was repeated using a layer of dodecane over the aqueous culture media in order to trap volatile sesquiterpenes. The dodecane was analysed by GC-MS using the same method as the culture extracts. Pogostol at 17.72 min and α-bisabolol at 18.06 min were present in the dodecane overlay, which is consistent with the data from the TPS24 culture extract. Importantly, α-guaiene and δ-guaiene were both present in the dodecane overlay. This confirms that these volatile sesquiterpenes were produced but had escaped from culture and is consistent with the previous characterisation of TPS24 (Drew et al. 2016). Additionally, a third guaiene-like sesquiterpenoid was present at 15.68 min. However, rotundone, the pepper aroma compound formed by oxidation of α-guaiene (Huang et al. 2014) was not present in the sample, which suggests either α-guaiene was sequestered from the yeast culture before oxidation could occur or $P450_{BM3}$ is unable to oxidise α-guaiene to rotundone.

The presence of α-bisabolol is surprising, as it is not structurally related to the expected products, which all share a 5,7-bicyclic ring. The match score for the MS spectrum of the peak at 18.06 min with α-bisabolol was 922, which is considered an excellent match, and the mass spectra are nearly identical (Supplementary Figures). This indicates there is a high probability that the unknown compound is α-bisabolol, but considering that α-bisabolol has never been previously found among products of TPS24 and does not share structural properties with other products of TPS24, more conclusive evidence is required. The unknown product could be a novel product that is similar to α-bisabolol but not present in the library. The mass spectrum of the unknown product does not match that of rotundol or any downstream products from the aerial oxidation of α-guaiene (Huang et al. 2015). In order to confirm the identity of the product, a standard of α-bisabolol could be run using the same GC-MS protocol to confirm that the retention time matches that of the unknown product. Alternatively, the yeast culture volume could be scaled up and the products purified and analysed by NMR spectroscopy.

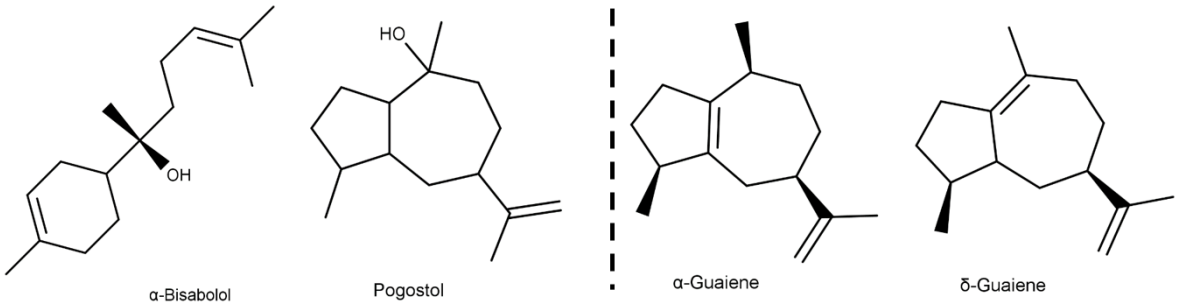


Figure 4-4 α-Bisabolol and Pogostol were present in extracts of EPY300 expressing TPS24. α-Guaiene and δ-Guaiene were expected products of TPS24 but were not found in the extract.

TPS26 produced δ-cadinene, di-epi-1,10-cubenol, epicubenol and α-cadionol. The mutant $P450_{BM3}$ also produced two unidentified sesquiterpenoid derivatives (Table 4-5). TPS26 (functional gene ID VvPNCuCad) has previously been characterised *in vivo* using metabolically engineered *E. coli* and was found to be a Cubebol/δ-Cadinene synthase with major products cubebol, δ-cadinene, α-cubebene, α-copaene, α-gurjunene, γ-cadinene and β-cubebene (Martin et al. 2010). This is consistent with the observed production of δ-cadinene and cadinene and cubebene derivatives. Again, not all of the previously identified products of TPS26 were found in the yeast culture extract, and this could be due to loss of volatile sesquiterpene hydrocarbons from the culture. All of the products that were identified in the TPS26 yeast culture extract had match factors greater than 900, indicating an excellent match with the library compound. Given the high match factors and the consistency with previous results, it is highly likely that these products have been correctly identified. The products identified in the yeast culture extract show many structural similarities (Figure 4-5). All share a 6,6-bicyclic ring structure with 1-isopropyl and 4,7-dimethyl groups. The structural similarities suggest that the products are formed by a common, non-enantioselective mechanism (Yoshikuni, Ferrin & Keasling 2006).

The unidentified sesquiterpenoids produced by $P450_{BM3}$ were not found in any of the other samples. The two products have molecular formulas of $C_{15}H_{24}O$ (MW 220) and $C_{15}H_{24}O_2$ (MW 236) respectively. These unidentified products are likely singly- and doubly- oxygenated derivatives of δ-cadinene or cubebene. The double oxygenation of the unknown compound at 22.20 min also explains its unusually long retention time, as this compound would be more polar and less volatile.

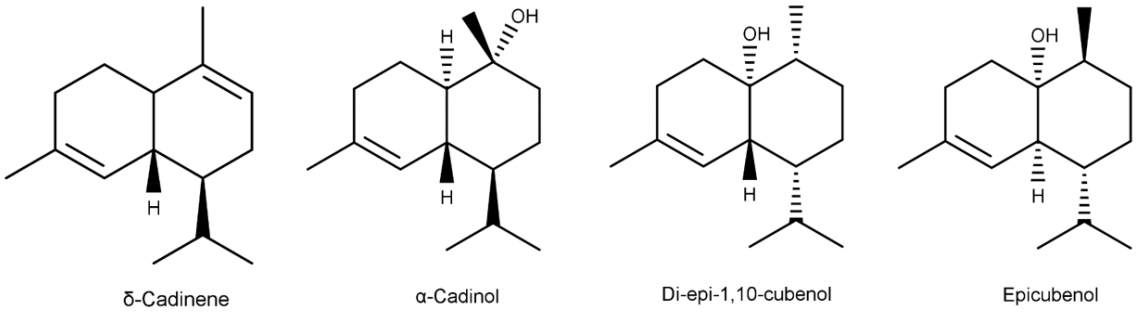


Figure 4-5 Products of TPS26 share structural similarities.

TPS27 produced β-caryophyllene, α-caryophyllene (α-humulene) and α-cadinol. Co-expression of TPS27 and the mutant $P450_{BM3}$ resulted in the formation of three additional, unidentified sesquiterpenoid products (Table 4-6). Standards of β-caryophyllene and α-caryophyllene are both commercially available, so these were compared to the sample peaks. The elution times of and mass spectra of the standards matched those of the sample peaks at 14.20 min and 14.73 min respectively. This indicates that NIST library comparison was able to be used to accurately identify these compounds, and justifies the use of this method for other compounds for which standards are not so readily available.

TPS27 (functional gene ID VvGwECar2) has previously been functionally characterised both *in vitro* by feeding recombinant protein with farnesyl pyrophosphate (FPP) as substrate and *in planta* in both transgenic *Arabidopsis thaliana* and grapevine. *In vitro*, TPS27 catalysed the cyclisation of FPP to produce β-caryophyllene, α-caryophyllene and a small amount of germacrene D (Martin et al. 2010; Matarese et al. 2014). Expressed in *A. thaliana,* TPS27 produced β-caryophyllene, α-caryophyllene and thujopsene, but no germacrene D was detected. Thujopsene appears to be a by-product of the enzyme outside the homologous expression system as only β-caryophyllene and α-caryophyllene were detected when TPS27 was overexpressed in *V. vinifera* (Matarese et al. 2014). These results are consistent with the identification of β-caryophyllene and α-caryophyllene within the TPS26 yeast culture extract and suggest that the production of α-cadinol may be an artefact of the expression system for this particular enzyme. The presence of α-cadinol is unexpected as it does not share similar structural properties with β-caryophyllene and α-caryophyllene, but interestingly it does share a 6,6-bicyclic ring structure and dimethyl group with thujopsene (Figure 4-6).

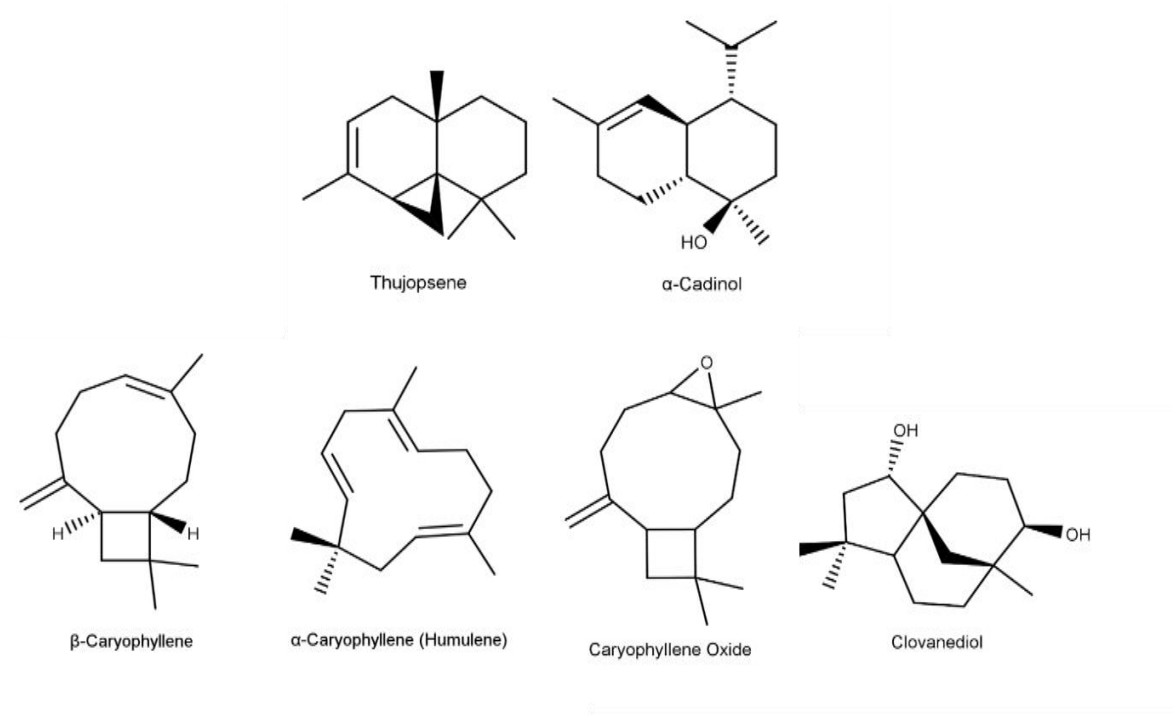


Figure 4-6 Products of TPS27. Thujopsene and α-cadinol share structural similarities but are not closely structurally related to β-caryophyllene and α-caryophyllene.

Co-expression of the mutant $P450_{BM3}$ led to the formation of four additional products. The product at 20.69 min had a molecular formula of $C_{15}H_{26}O_2$ (MW 238). The closest library match was clovanediol, which was a good match (match factor 800-900), but not identical. This suggests the compound may share some features in common with clovanediol, but it is not an exact match. Similarly, the product at 21.38 min had a molecular formula of $C_{15}H_{24}O$ (MW 220) and was a good match, but not identical, to caryophyllene oxide. Caryophyllene oxide would be an expected product of β-caryophyllene oxidation, so it is possible that the unknown compound is very similar to caryophyllene oxide.

To investigate the origin of these products, EPY300 expressing the mutant $P450_{BM3}$ only was cultured with either β-caryophyllene or α-caryophyllene in order to identify which peaks arise as the result of the presence of either β-caryophyllene or α-caryophyllene. GC-MS analysis of both β-caryophyllene and α-caryophyllene culture extracts revealed a number of peaks not present in the TPS27 culture. These appear to be caryophyllene oxides and epoxides formed by providing $P450_{BM3}$ with a high concentration of substrate. The compounds present at 17.70 min, 20.01 min 20.69 min and 21.38 min were not found in either of the caryophyllene-spiked samples. Since α-cadinol (17.70 min) was found in both the BM3 Wild Type and BM3 Mutant culture extracts, it is likely to be a direct product of TPS27. The compounds at 20.01 min 20.69 min and 21.38 min were only found in the BM3 Mutant samples and not in either of the caryophyllene-spiked samples, so they

may be $P450_{BM3}$-oxygenated derivatives of α-cadinol. To confirm this, EPY300 expressing the mutant $P450_{BM3}$ only could be cultured with α-cadinol in a similar way to the caryophyllene experiments.

A peak at 19.82 min was present in the β-caryophyllene-spiked sample. The mass spectra of the peaks at 19.82 min in the β-caryophyllene-spiked sample and the TPS27 sample were found to share significant similarities, so this is likely to be the same compound (Figure 4-3). This indicates the caryophyllene-oxide like compound at 19.82 min is a β-caryophyllene derivative. However, conclusive structural characterisation of the product would require NMR spectroscopy of the purified compound since no library match is available.

TPSY1 produced a range of structurally varied products, including β-ylangene, β-cubebene or β-copaene, isogermacrene D, germacrene D, δ-cadinene, juneol, α-cadinol and germacrene-like and cadinene-like oxygenated sesquiterpenoids (Table 4-7). Broadly, these can be classified as germacrene-like structures and structures containing 6,6-bicyclic rings (Figure 4-7). It is likely that these products are formed through a common mechanism via germacrene D (Bülow & König 2000; López-Gallego, Wawrzyn & Schmidt-Dannert 2010).

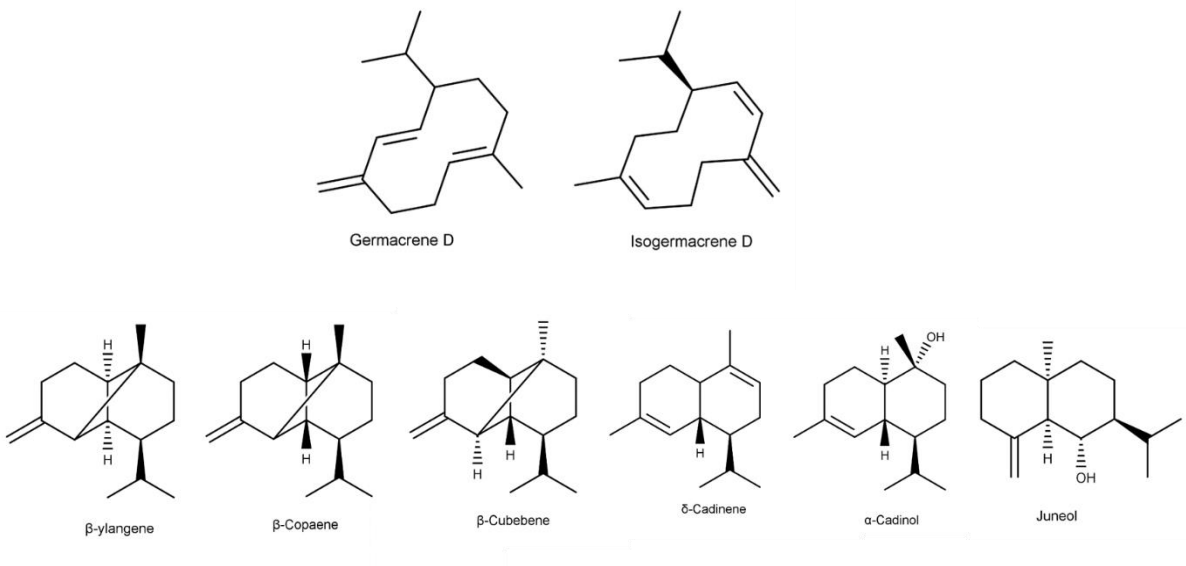


Figure 4-7 Products of TPSY1

The peaks at 20.02 min and 20.47 min, corresponding to oxygenated sesquiterpenoids, match the retention times of peaks present in the TPS07 sample. Comparison of the mass spectra between the TPS07 and TPSY1 samples reveal these products are likely the same in both samples (Figure 4-8). These peaks do not have a close library match but are suggested to be caryophyllene-oxide or aromadendrene-oxide like with molecular

formula $C_{15}H_{24}O$. Since TPS07 is a germacrene D synthase, and germacrene D is also produced by TPSY1, it is hypothesised that both compounds are germacrene D mono-oxygenated derivatives. Substructure reports suggest these compounds are epoxides of germacrene D.

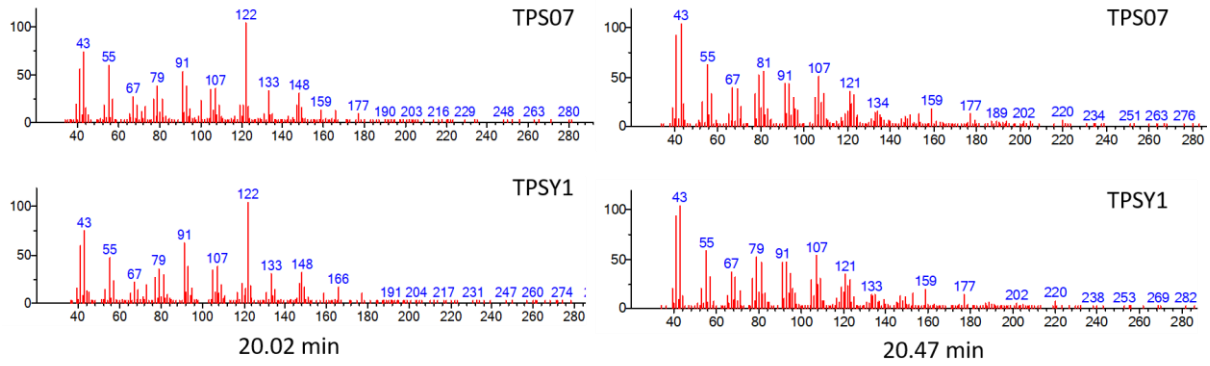


Figure 4-8 The mass spectra of peaks found in the TPS07 extract match those of peaks found in the TPSY1 extract

TPSY1 has not been previously characterised; however, transcriptome analysis shows it is only expressed in young berries (3-weeks post anthesis) in *Vitis vinifera* cv. Shiraz (Sweetman et al. 2012). Of the products identified in the TPSY1 yeast culture extract, δ-cadinene is the only one that has been identified so far in young berries (4-weeks post anthesis) (Zhang et al. 2016a). The other products found in the TPSY1 yeast culture extract may not have been detected in grapes previously due to low concentrations or because they are not major products, compared to the abundance of more common sesquiterpenes like δ-cadinene. It is reasonable to conclude that these products have been accurately identified due to their high library match factors (>900). Furthermore, the structural similarity of the products means it was reasonable that they were synthesised by the same enzyme. Nevertheless, further characterisation of TPSY1 is required to confirm these products are produced *in vitro* by the enzyme using FPP as a substrate and *in planta* in an overexpression system to determine whether these products can be found in grapevine tissues.

TPSY2 produced β-caryophyllene/isocaryophyllene, α-caryophyllene, δ-cadinene, di-epi-1,10-cubenol and epicubenol. All of the products of TPSY2 could be confidently identified using the NIST library, with every compound having a match factor over 900. The identity of the product at 14.18 min was ambiguous between β-caryophyllene and isocaryophyllene, as both had high match factors and almost identical structures (Figure 4-9). The products of TPSY2 have not been previously characterised. Like TPSY1, it is expressed in *V. vinifera* cv. Shiraz exclusively at the young berry stage(Sweetman et al. 2012). β-Caryophyllene, α-caryophyllene, δ-cadinene, and epicubenol have all been

identified in extracts of young berries (4 weeks post anthesis) (Zhang et al. 2016a). Therefore, it is reasonable to conclude this is likely an accurate characterisation of the TPSY2 products.

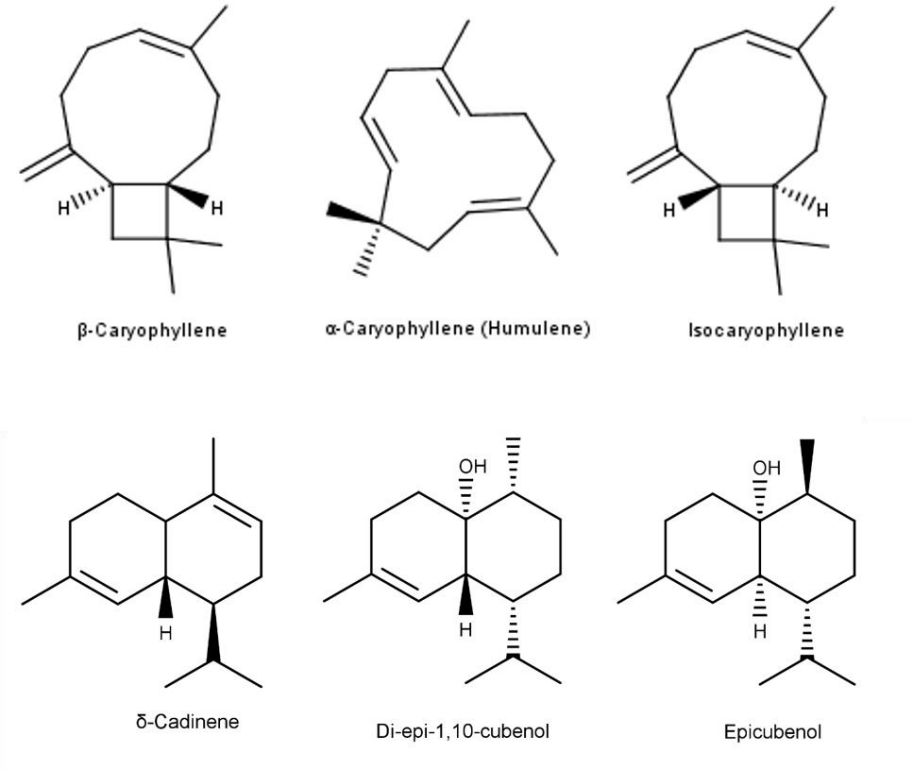


Figure 4-9 Products of TPSY2

TPSY2 shares products in common with both TPS26 (δ-cadinene, di-epi-1,10-cubenol, epicubenol) and TPS27 (β-caryophyllene, α-caryophyllene). TPSY2 shares 58.9% sequence identity with TPS26 and 70.9% sequence identity with TPS27 at the amino acid level. The similarities in amino acid sequence may account for the common products. Additionally, β-caryophyllene is involved in chemical defence and is biologically active against the pest European Grapevine Moth (*Lobesia botrana*) (Tasin et al. 2005). Therefore, expressing multiple β-caryophyllene synthases may be advantageous to the plant.

The six sesquiterpene synthase expression systems produced a large number of structurally varied sesquiterpenoids. In total, seventeen unique sesquiterpenoids were confidently identified as products of the six sesquiterpene synthases. Additionally, a number of sesquiterpene hydrocarbons acted as substrates for $P450_{BM3}$; however, many of these products were not positively identified due to the limitations of the available library. Oxygenation can change the biological properties of sesquiterpenes. For example, valencene is a sesquiterpene hydrocarbon with an orange-citrus aroma. Oxidation to a

ketone yields nootkatone, which has a distinctive grapefruit aroma (Hunter & Brogden JR. 1965). Another sesquiterpene hydrocarbon, α-guaiene, has a woody, balsamic aroma (Acree & Arn 2004), and may be oxidised to rotundone, which is responsible for the peppery character of some wines (Wood et al. 2008). Therefore, further investigation into identifying the origins and the structures of these oxygenated sesquiterpenoids may provide insight into the flavour or aroma of wine. This may include gas chromatography analysis with olfactometric detection (GC-O) to investigate whether these oxygenated sesquiterpenoids have detectable aromas, or large-scale culture and purification to positively identify the products by NMR.

Most of the sesquiterpene synthases produced multiple products, and several had products in common. Terpenoids are important in nature as they perform a multitude of ecological roles including direct defence against pathogens and pests (Heiling et al. 2010), attraction of pollinators (Dudareva & Pichersky 2000), and signals for plant-plant communication (Arimura et al. 2000). The core active site of sesquiterpene synthases is preserved across the greater family of terpene synthases, with the active site cavity acting as a template for FPP binding (Lesburg et al. 1997). It is thought that sesquiterpene synthases evolved from a common, promiscuous ancestor and have the ability to adopt novel or altered functions via a few amino acid substitutions (Yoshikuni, Ferrin & Keasling 2006). Phylogenetic analysis suggests that most grapevine sesquiterpene synthases form clusters of paralogous genes that have arisen due to gene duplications (Martin et al. 2010). This accounts for the product commonality between the sesquiterpene synthases. Product diversity can arise from a small number of amino acid substitutions (Drew et al. 2016), and given the wide range of ecological activities of sesquiterpenoids, product diversity is likely to confer an evolutionary advantage to the organism.

Sesquiterpenoids are also highly valuable to humans as fragrances and flavours, and may have medicinal benefits (Babalola, Anetor & Adeniyi 2001; Bohlmann & Keeling 2008). Sesquiterpenoids are important for the aroma and character of wine (Mayr et al. 2014; Wood et al. 2008), and recent research has explored the potential health benefits of wines containing sesquiterpenoid compounds (Duhamel et al. 2018). Several of the sesquiterpenoids identified in the EPY300 culture extracts have distinctive aromas or bioactivity (Table 4-9).

Table 4-9 Many of the sesquiterpenoids found in the yeast culture extracts have distinctive aromas or bioactivity

Sesquiterpenoid	Aroma	Bioactivity
Germacrene D	wood, spice[1]	
δ-Cadinene	thyme, medicine, wood[1]	
Caryophyllene Oxide	herb, sweet, spice[1]	analgesic, anti-inflammatory[2], preservative, anti-fungal[3]
Pogostol		anti-emetic[4]
α-Bisabolol	spice, flower[1]	wound healing[5] ,anti-microbial[6], anti-proliferative[7]
α-Cadinol	herb, wood[1]	anti-mite[8]
Epicubenol	spice, herb, green tea[1]	
β-Caryophyllene	wood, spice[1]	targets CB2 receptor of endocannabinoid system, anti-anxiety, anti-depressant [9], analgesic [10], local anaesthetic [11]
α-Caryophyllene	wood[1]	anti-inflammatory[12], cytotoxic[13]
β-Cubebene	citrus, fruit[1]	

[1] (Acree & Arn 2004), [2] (Chavan, Wakte & Shinde 2010), [3] (Yang, D et al. 2000), [4] (Yang, Y et al. 1999), [5] (Villegas et al. 2001), [6] (Forrer et al. 2013), [7] (Mendes et al. 2017), [8] (Chang et al. 2001), [9] (Bahi et al. 2014), [10] (Klauke et al. 2014), [11] (Ghelardini et al. 2001) [12] (Fernandes et al. 2007), [13] (Legault et al. 2003)

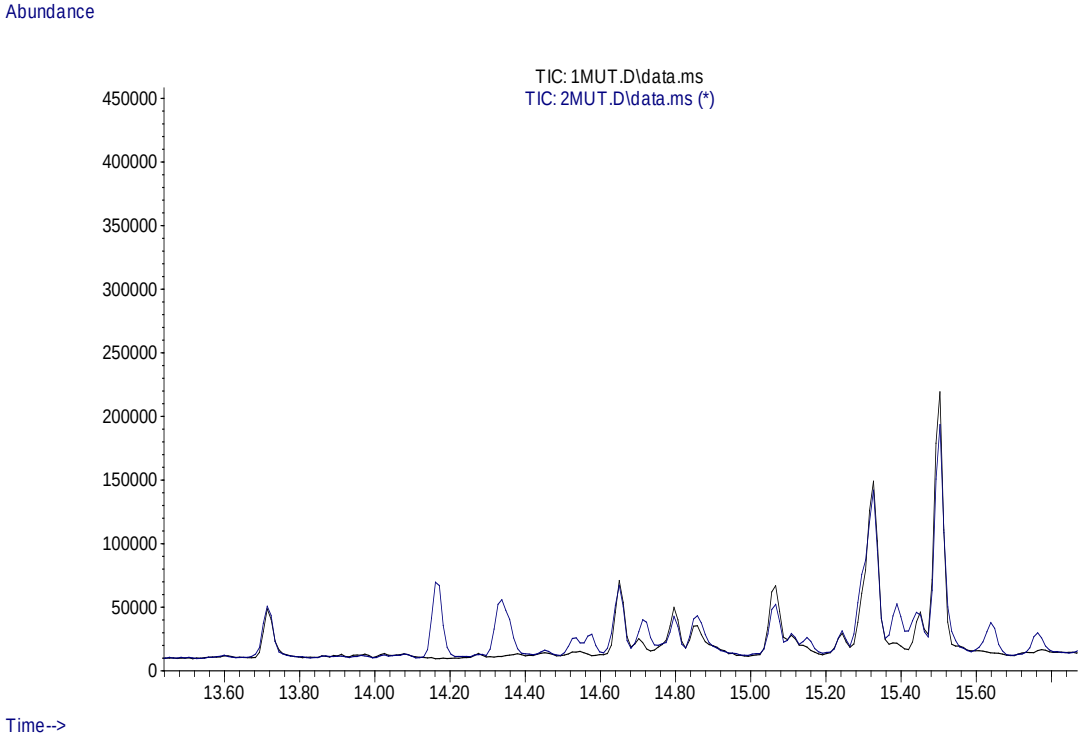
Many of the sesquiterpenoid products identified in the EPY300 culture extracts have been described in the viticulture and oenology literature (Schreier, Drawert & Junker 1976; Zhang et al. 2016b). However, the effect of some compounds on wine aroma and character have not been studied. Given the range of compounds and variety of aromas, the products of the terpene synthases studied may contribute significantly to wine character and quality. Furthermore, a number of these compounds show beneficial bioactivities, which can serve as a marketing platform for wines containing sesquiterpenoids. The molecules identified in this work may serve as a useful reference framework for future studies into wine composition, aroma, and health benefits.

Conclusions

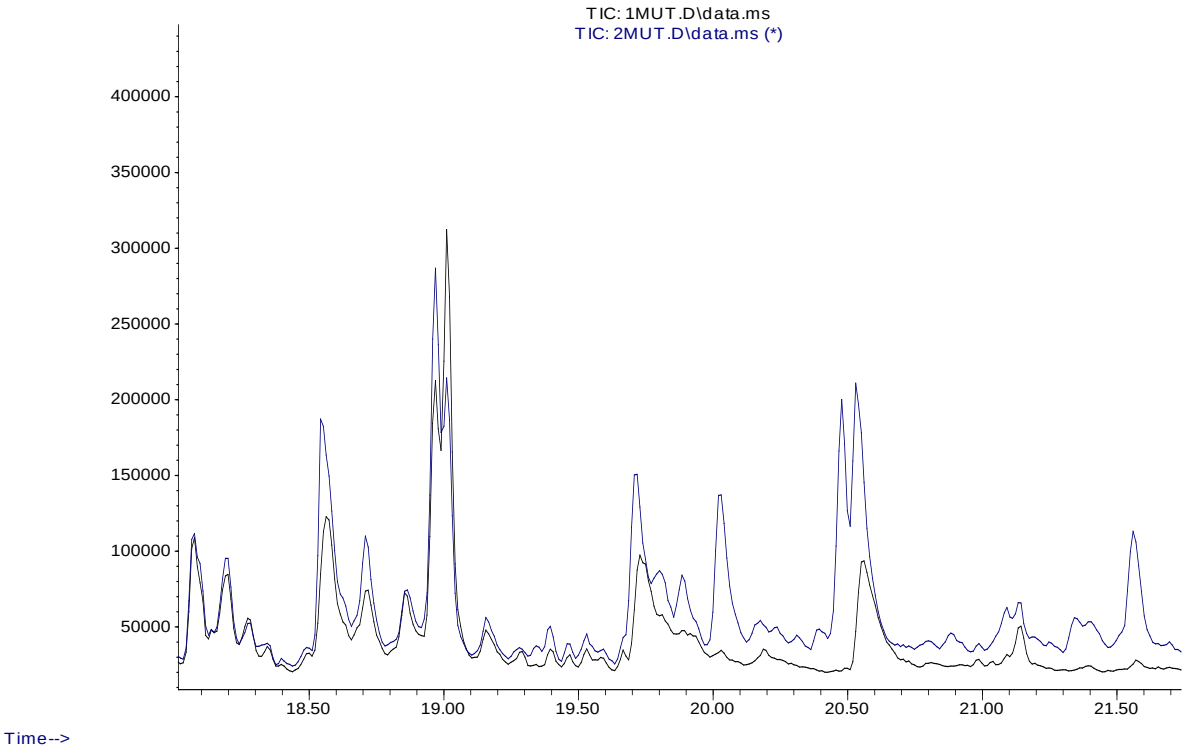
A number of sesquiterpenes and oxygenated derivatives were identified as products of the six *V. vinifera* sesquiterpene synthases investigated in this study. This demonstrates the diversity of sesquiterpene synthases and sesquiterpenoid products that may be found in grapes and wine. This work may be used as a platform for further identification of sesquiterpenoids and investigation into their contributions to the flavour, aroma and health benefits of wine.

Supplementary Information for Production of Grapevine Sesquiterpenes and Sesquiterpenoids in Yeast

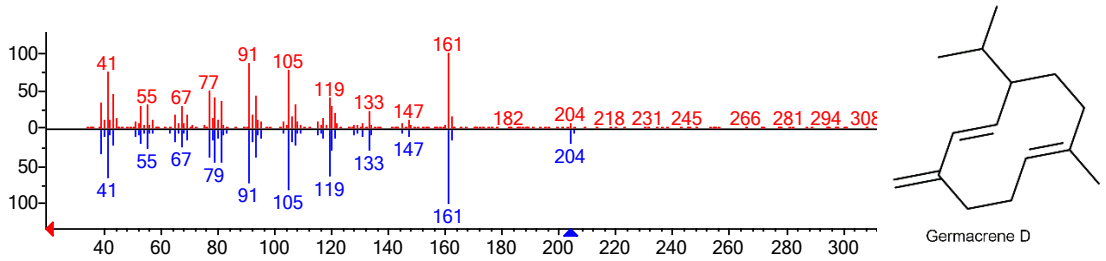
GC-MS Data for *TPS07*



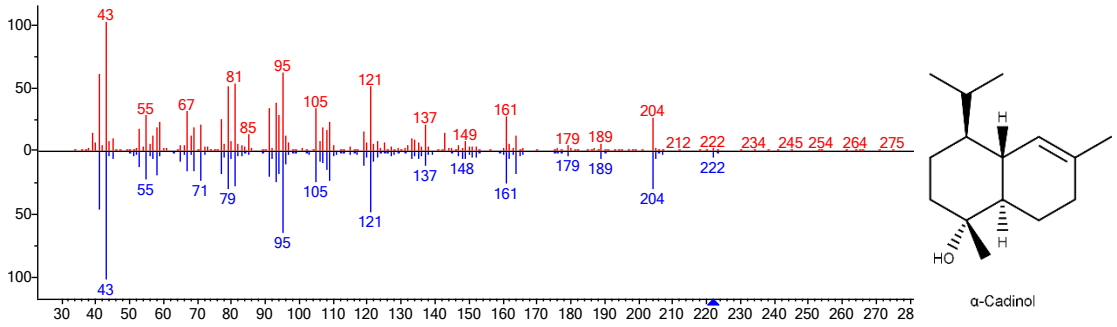
Supplementary Figure 4-1 Total Ion Chromatogram (TIC) of extract from EPY300 expressing TPS07 with Mutant P450$_{BM3}$ (blue) overlayed with TIC of extract from the Mutant P450$_{BM3}$ only control (black). New peaks are present at 14.19 min and 14.34 min.



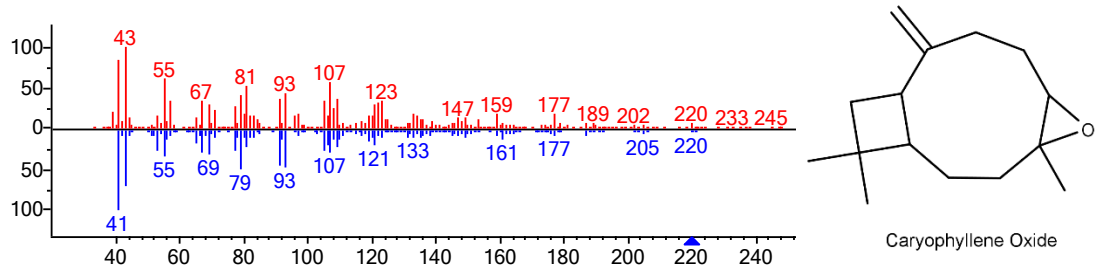
Supplementary Figure 4-2 TIC of extract from EPY300 expressing TPS07 with Mutant P450$_{BM3}$ (blue) overlayed with TIC of extract from the Mutant P450$_{BM3}$ only control (black). New peaks are present at 20.03 min and 20.48 min.



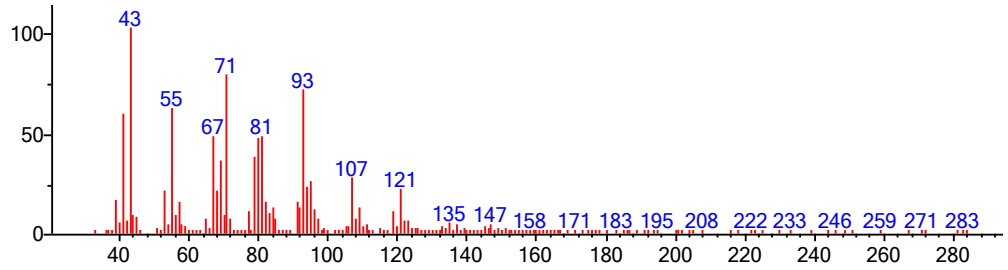
Supplementary Figure 4-3 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPS07 with Mutant P450$_{BM3}$ at 14.34 min (red) with the library spectrum of Germacrene D (blue).



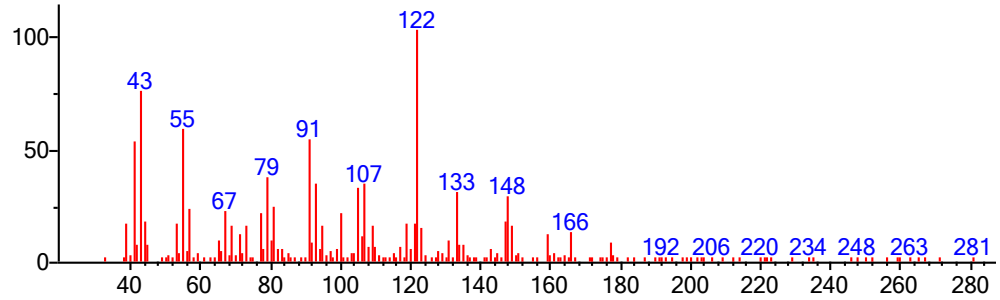
Supplementary Figure 4-4 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPS07 with Mutant P450$_{BM3}$ at 17.70 min (red) with the library spectrum of α-cadinol.(blue).



Supplementary Figure 4-5 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPS07 with Mutant P450$_{BM3}$ at 20.48 min (red) with the library spectrum of caryophyllene oxide (blue).

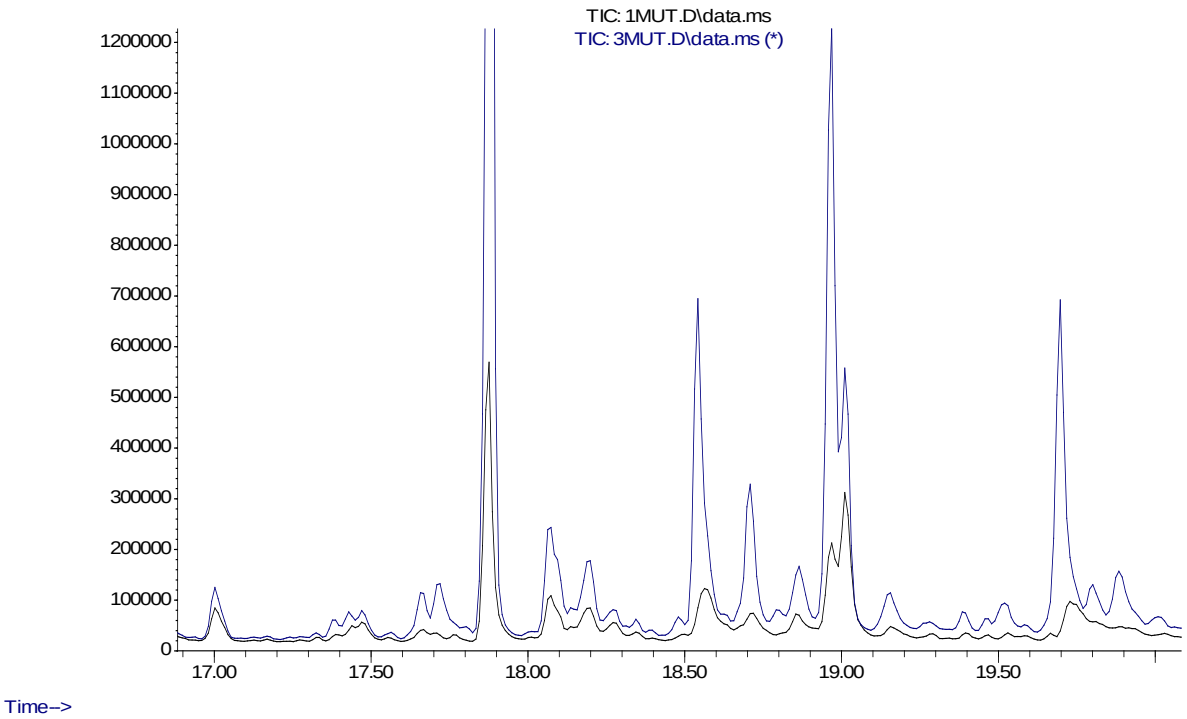


Supplementary Figure 4-6 Mass spectra from the extract of EPY300 expressing TPS07 with Mutant P450$_{BM3}$ at 19.72 min.

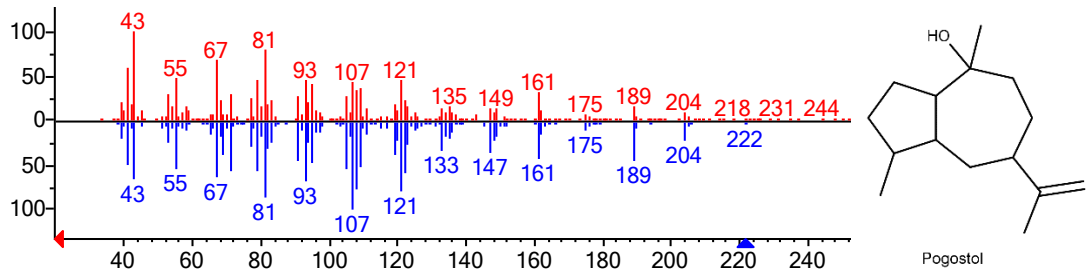


Supplementary Figure 4-7 Mass spectra from the extract of EPY300 expressing TPS07 with Mutant P450$_{BM3}$ at 20.03 min.

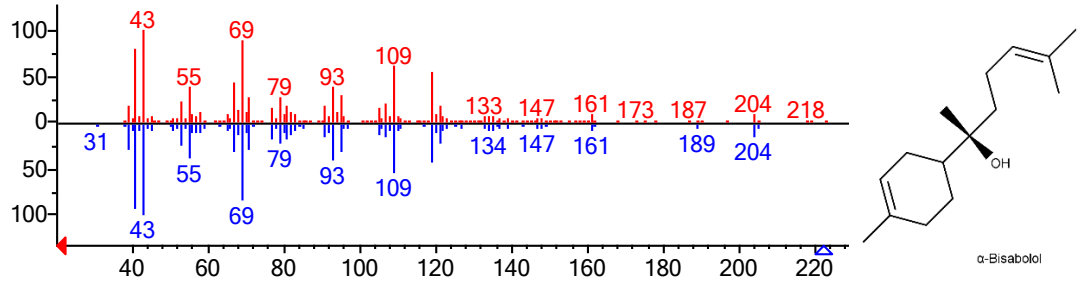
GC-MS Data for *TPS24*



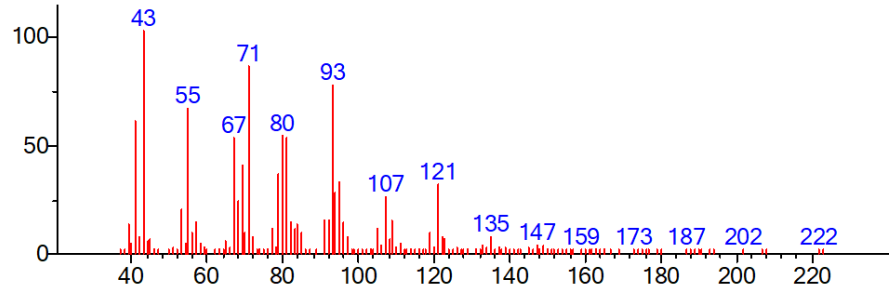
Supplementary Figure 4-8 TIC of extract from EPY300 expressing TPS24 with Mutant P450$_{BM3}$ (blue) overlayed with TIC of extract from the Mutant P450$_{BM3}$ only control (black). New peaks are present at 17.72 min, 18.06 min and 19.70 min.



Supplementary Figure 4-9 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPS24 with Mutant P450$_{BM3}$ at 17.72 min (red) with the library spectrum of pogostol (blue).

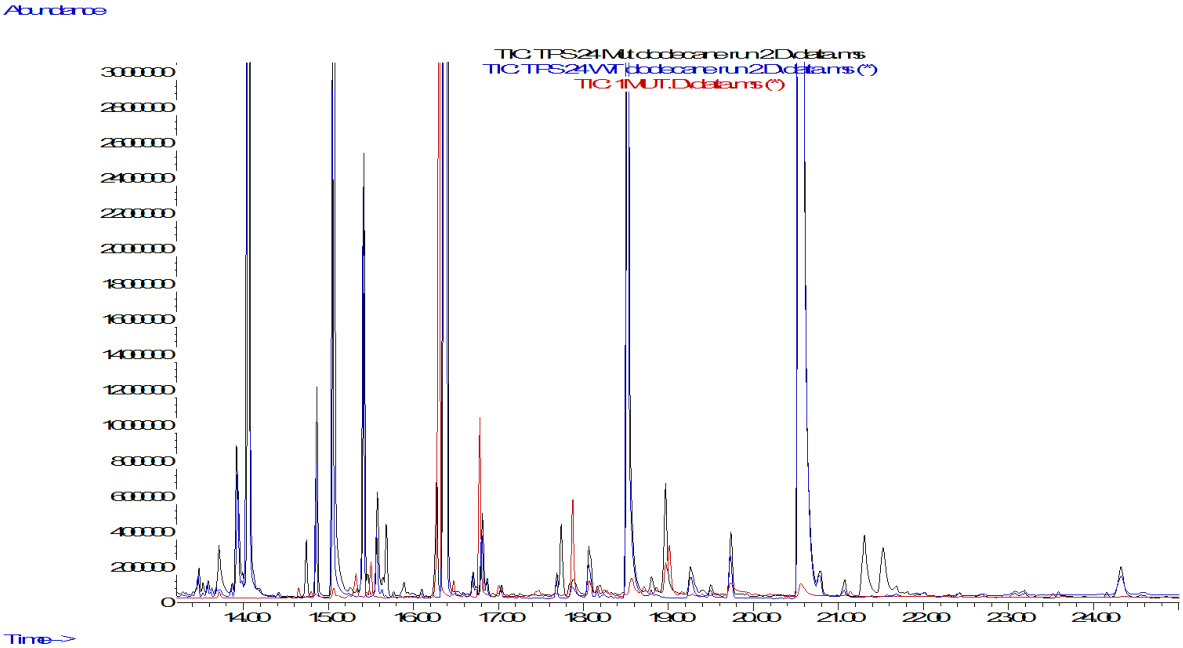


Supplementary Figure 4-10 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPS24 with Mutant P450$_{BM3}$ at 18.06 min (red) with the library spectrum of α-bisabolol (blue).

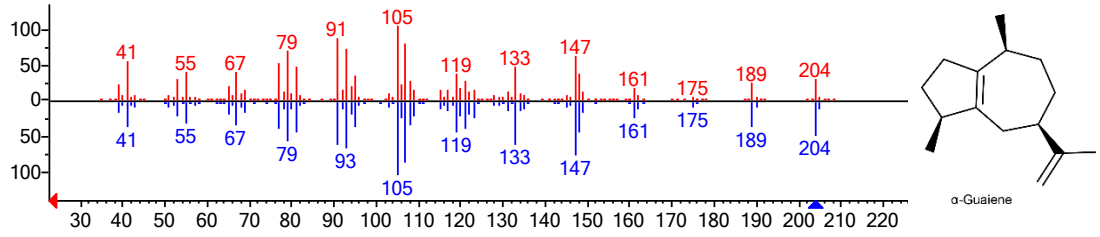


Supplementary Figure 4-11 Mass spectra from the extract of EPY300 expressing TPS24 with Mutant P450$_{BM3}$ at 19.70 min.

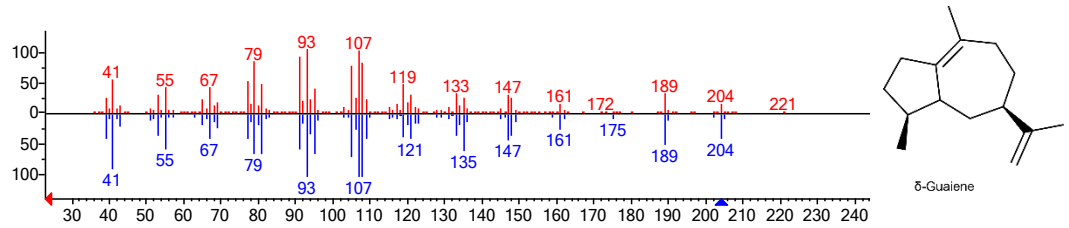
GC-MS Data for *TPS24* with dodecane overlay



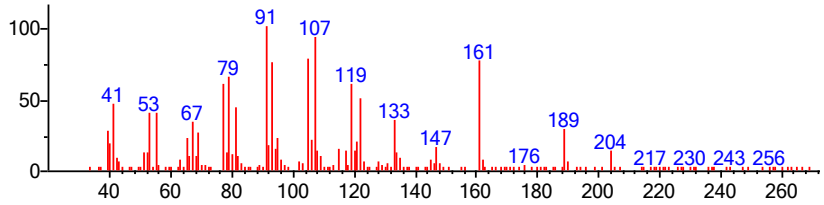
Supplementary Figure 4-12 TIC of dodecane from EPY300 expressing TPS24 with Mutant P450BM3 (black) overlayed with TIC of dodecane from EPY300 expressing TPS24 with wild-type P450BM3 (blue) and TIC of extract from the Mutant P450BM3 only control (red). Many new peaks are dodecane derivatives. New terpene peaks are present at 14.74 min, 15.68 min, 15.88 min, 17.72 min and 18.06 min.



Supplementary Figure 4-13 Head-to-tail comparison of mass spectra from the dodecane overlay of EPY300 expressing TPS24 with Mutant P450BM3 at 14.74 min (red) with the library spectrum of α-guaiene (blue).

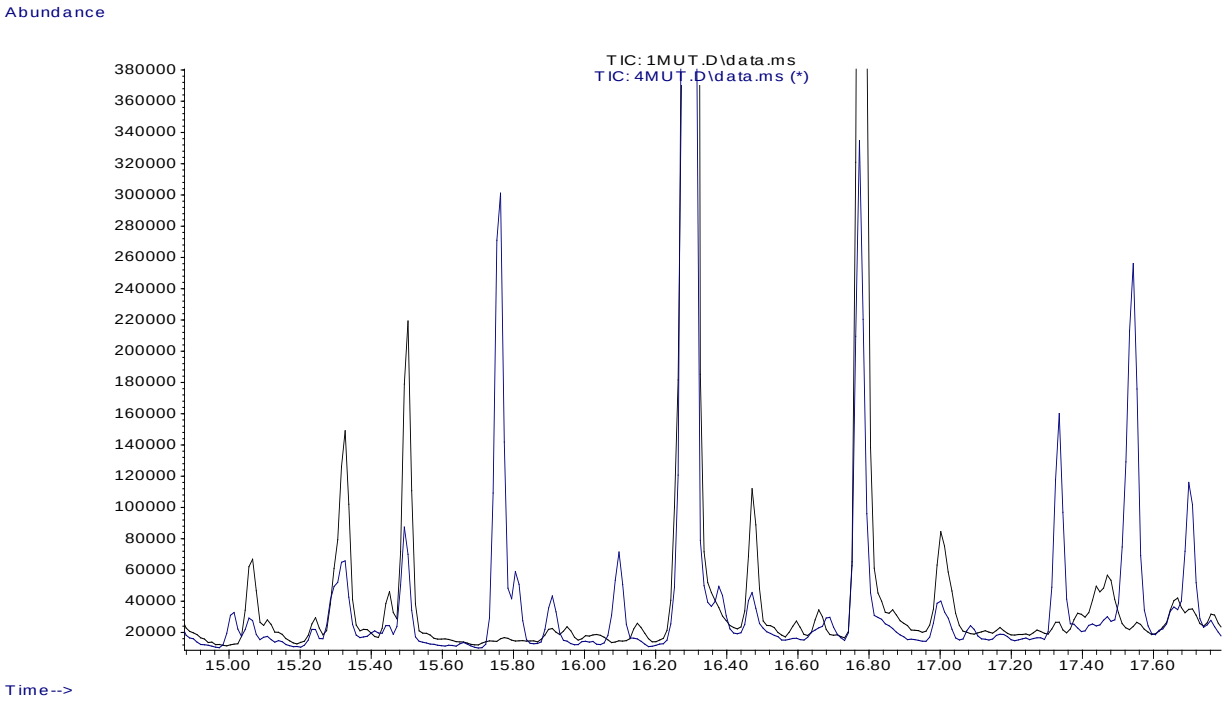


Supplementary Figure 4-14 Head-to-tail comparison of mass spectra from the dodecane overlay of EPY300 expressing TPS24 with Mutant P450BM3 at 15.68 min (red) with the library spectrum of δ-guaiene (blue).

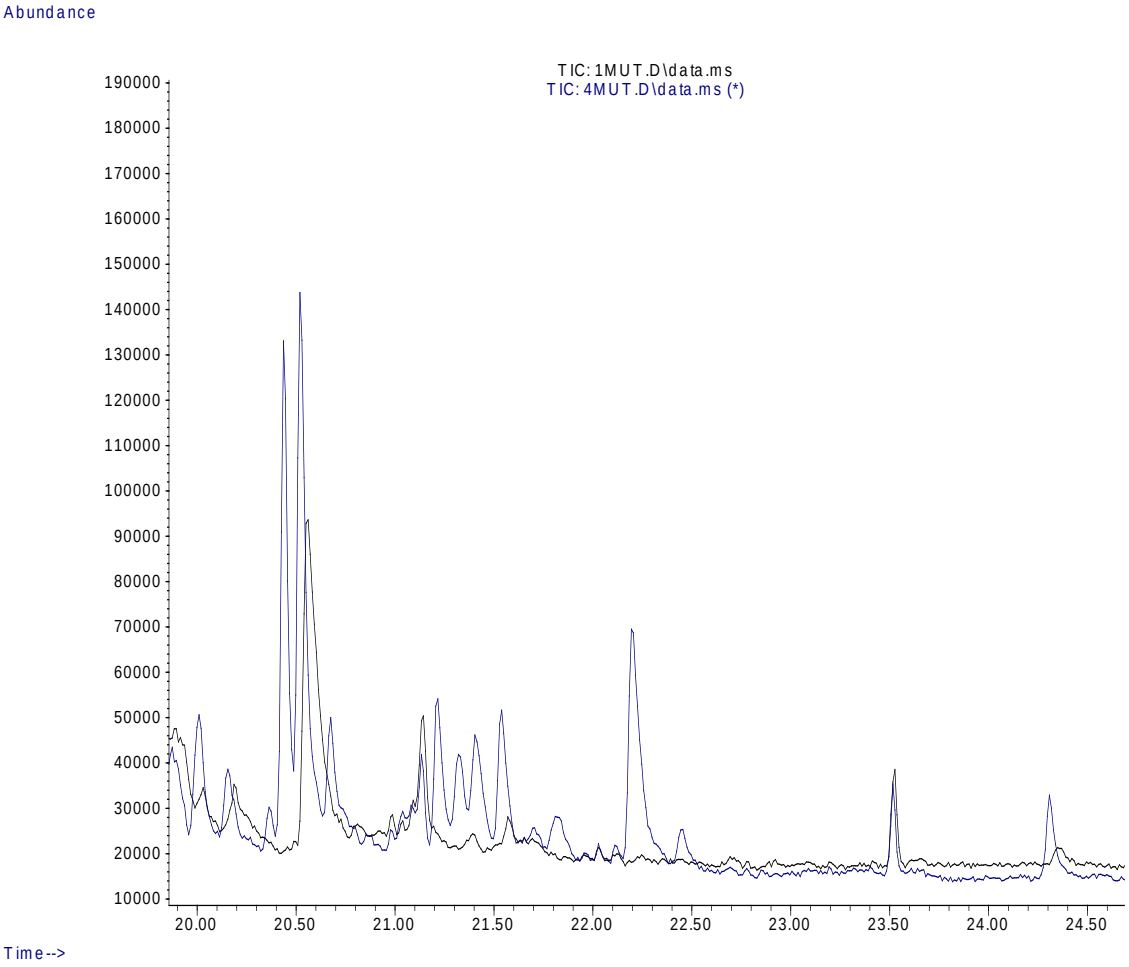


Supplementary Figure 4-15 Mass spectra from the dodecane overlay of EPY300 expressing TPS24 with Mutant P450BM3 at 15.88 min.

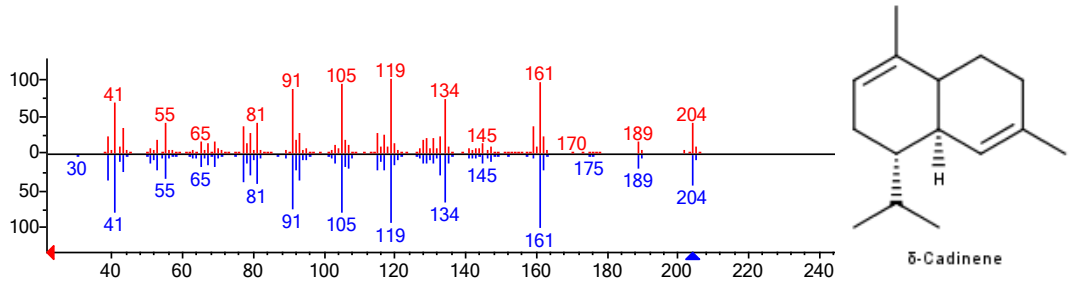
GC-MS Data for *TPS26*



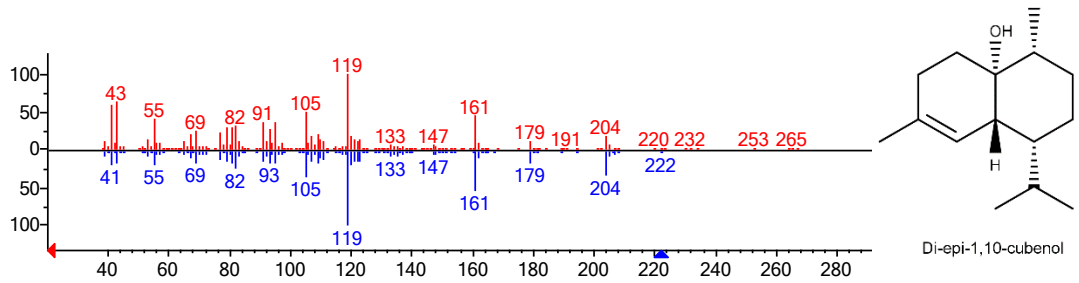
Supplementary Figure 4-16 TIC of extract from EPY300 expressing TPS26 with Mutant P450$_{BM3}$ (blue) overlayed with TIC of extract from the Mutant P450$_{BM3}$ only control (black). New peaks are present at 15.76 min, 16.10 min, 17.33 min, 17.52 min and 17.70min.



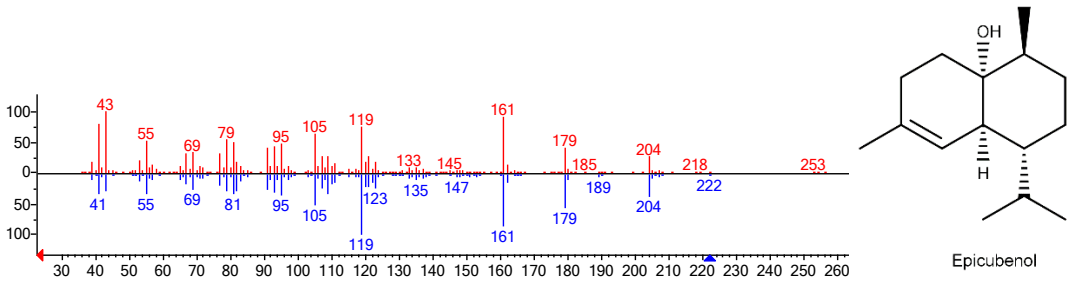
Supplementary Figure 4-17 TIC of extract from EPY300 expressing TPS26 with Mutant P450$_{BM3}$ (blue) overlayed with TIC of extract from the Mutant P450$_{BM3}$ only control (black). A new peak is present at 22.20 min.



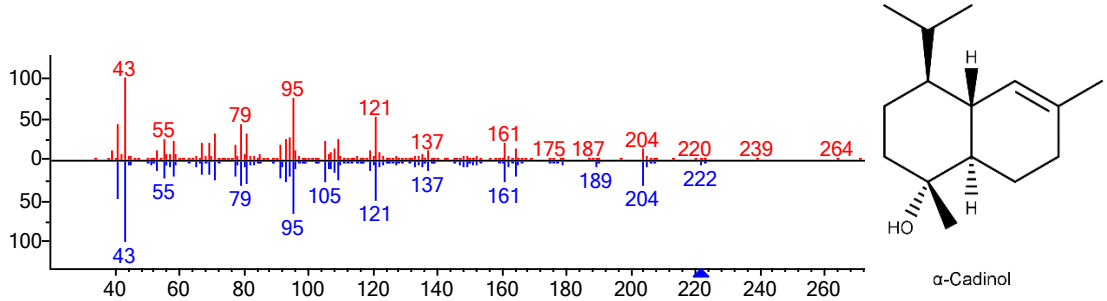
Supplementary Figure 4-18 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPS26 with Mutant P450$_{BM3}$ at 15.76min (red) with the library spectrum of δ-cadinene (blue).



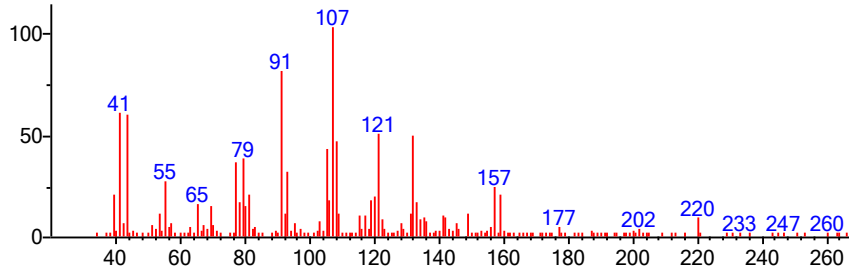
Supplementary Figure 4-19 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPS26 with Mutant P450$_{BM3}$ at 17.33 min (red) with the library spectrum of di-epi-1,10-cubenol (blue).



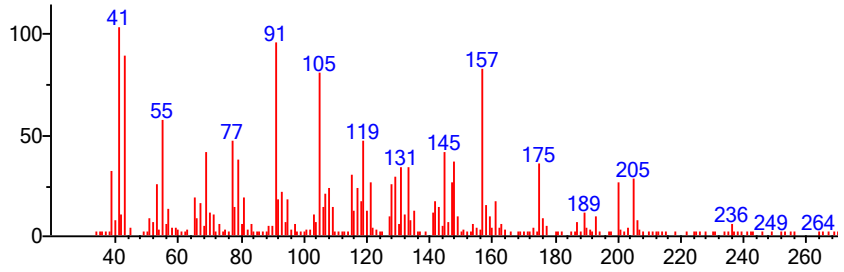
Supplementary Figure 4-20 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPS26 with Mutant P450$_{BM3}$ at 17.54 min (red) with the library spectrum of epicubenol (blue).



Supplementary Figure 4-21 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPS26 with Mutant P450$_{BM3}$ at 17.70 min (red) with the library spectrum of α-cadinol (blue).

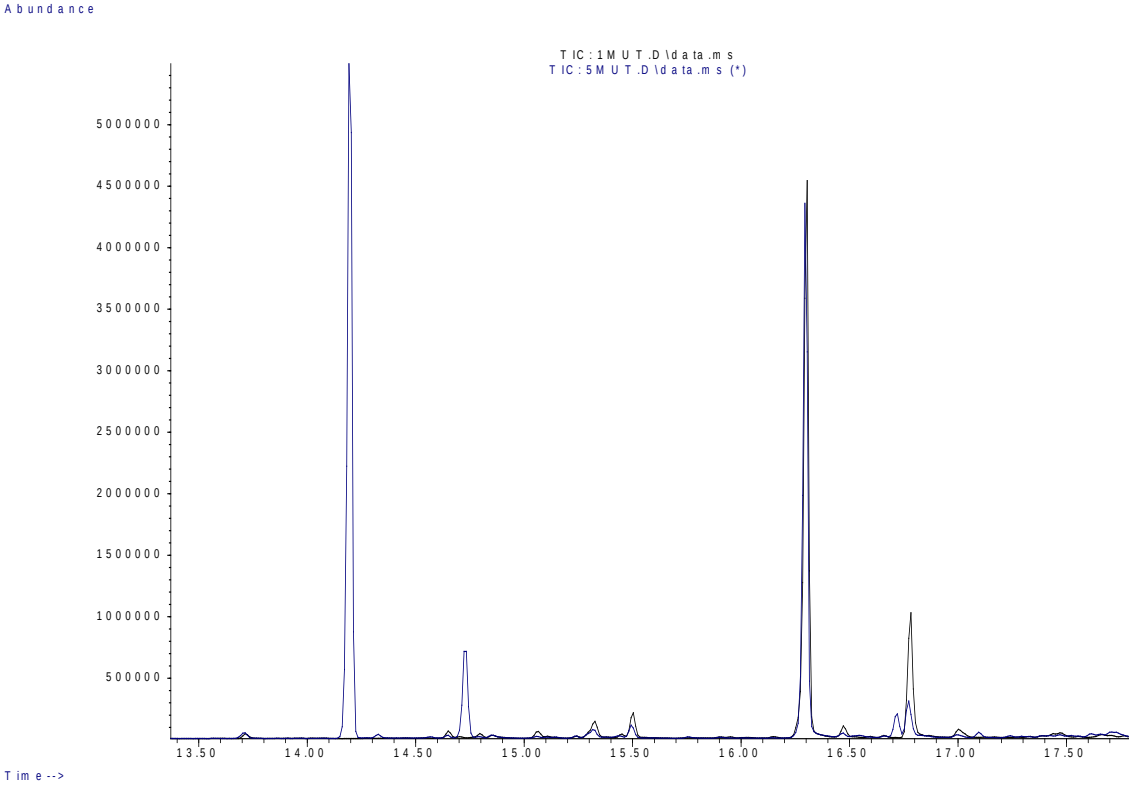


Supplementary Figure 4-22 Mass spectra from the extract of EPY300 expressing TPS26 with Mutant P450$_{BM3}$ at 16.10 min.

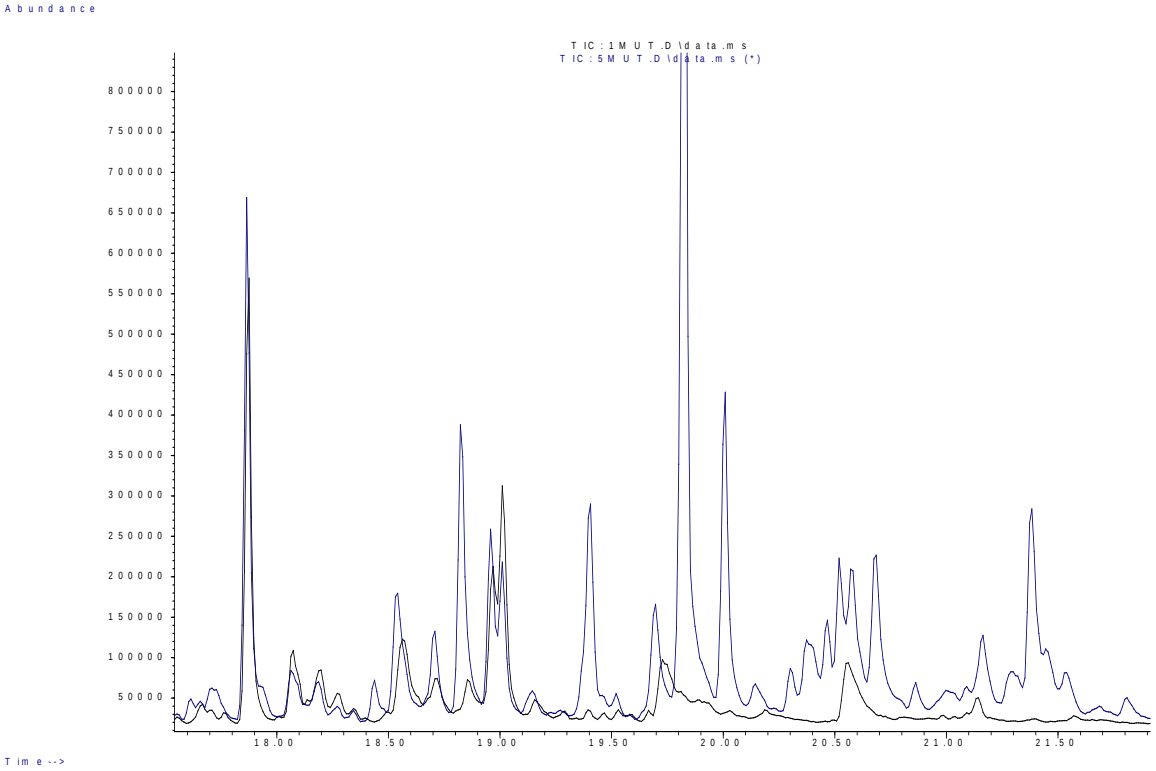


Supplementary Figure 4-23 Mass spectra from the extract of EPY300 expressing TPS26 with Mutant P450$_{BM3}$ at 22.20 min.

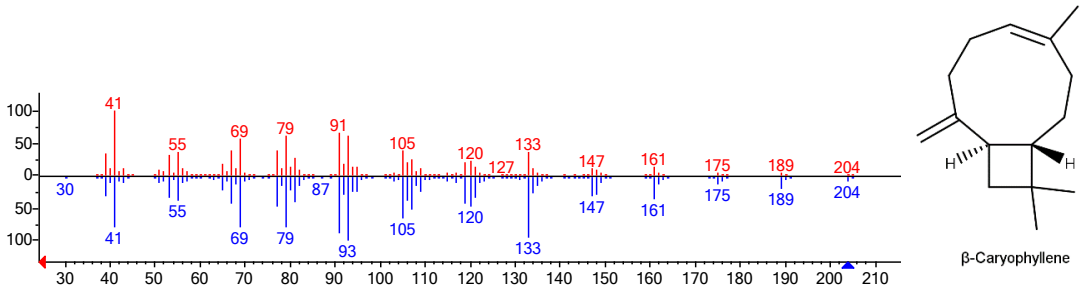
GC-MS Data for *TPS27*



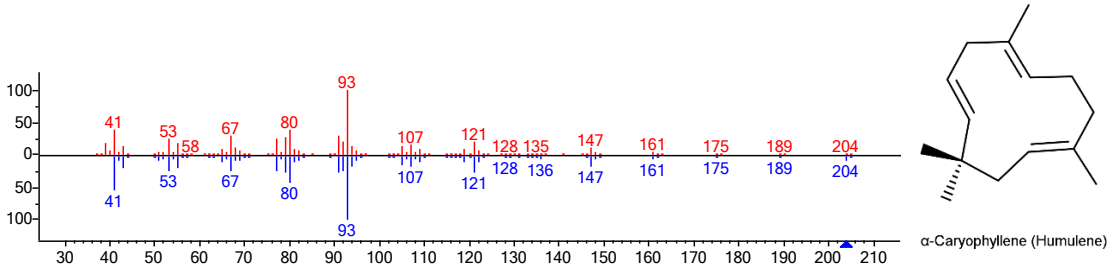
Supplementary Figure 4-24 TIC of extract from EPY300 expressing TPS27 with Mutant P450$_{BM3}$ (blue) overlayed with TIC of extract from the Mutant P450$_{BM3}$ only control (black). New peaks are present at 14.20 min and 14.73 min.



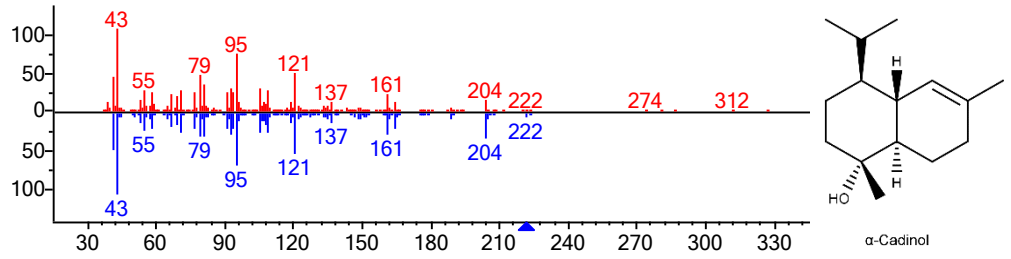
Supplementary Figure 4-25 TIC of extract from EPY300 expressing TPS27 with Mutant P450$_{BM3}$ (blue) overlayed with TIC of extract from the Mutant P450$_{BM3}$ only control (black). New peaks are present at 17.70 min, 19.82 min, 20.02 min, 20.69 min and 21.38 min.



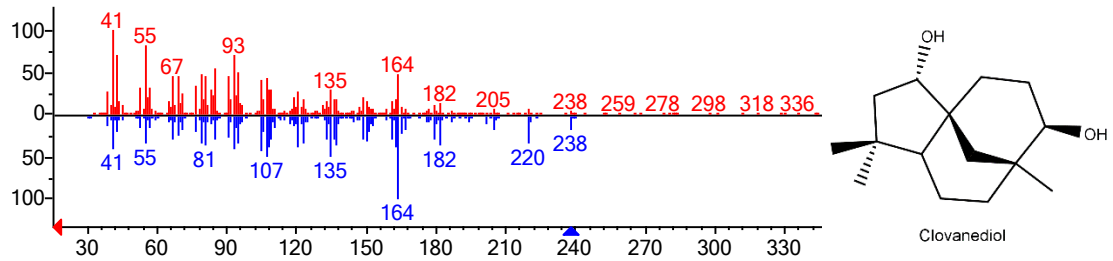
Supplementary Figure 4-26 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPS27 with Mutant P450$_{BM3}$ at 14.20min (red) with the library spectrum of β-caryophyllene (blue).



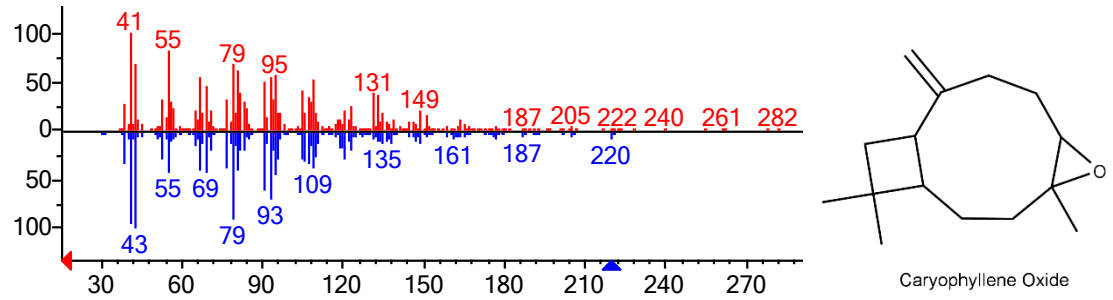
Supplementary Figure 4-27 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPS27 with Mutant P450$_{BM3}$ at 14.73 min (red) with the library spectrum of α-caryophyllene (humulene) (blue).



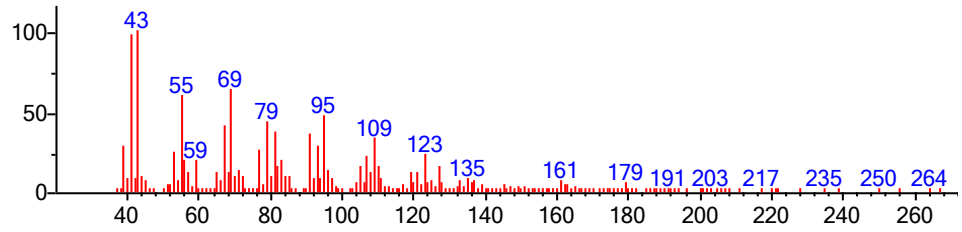
Supplementary Figure 4-28 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPS27 with Mutant P450$_{BM3}$ at 17.70 min (red) with the library spectrum of α-cadinol (blue).



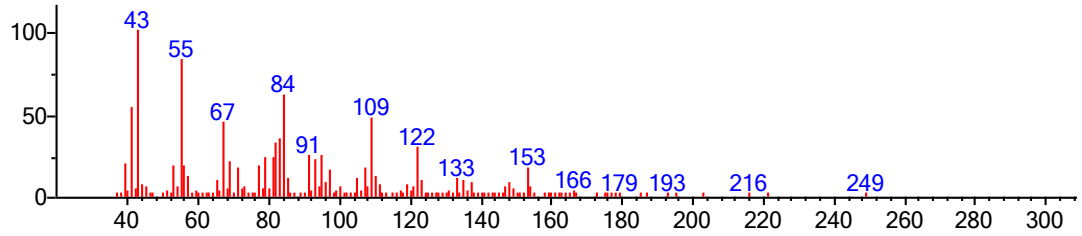
Supplementary Figure 4-29 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPS27 with Mutant P450$_{BM3}$ at 20.69 min (red) with the library spectrum of clovanediol (blue).



Supplementary Figure 4-30 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPS27 with Mutant P450$_{BM3}$ at 21.38 min (red) with the library spectrum of caryophyllene oxide (blue).

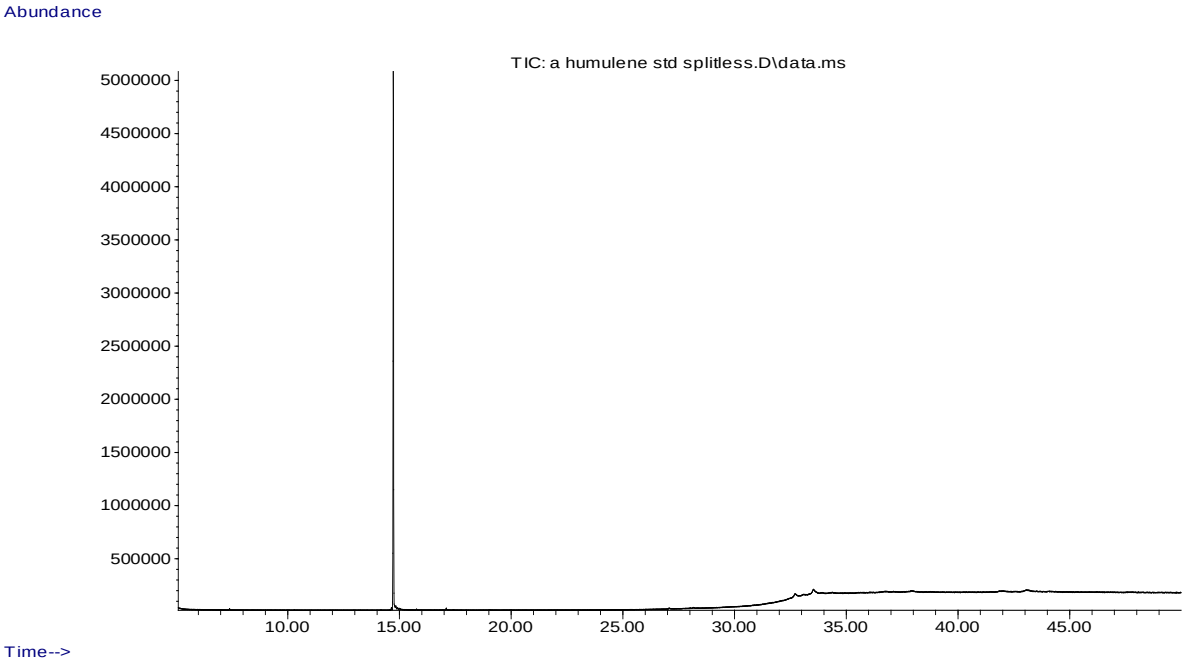


Supplementary Figure 4-31 Mass spectra from the extract of EPY300 expressing TPS27 with Mutant P450$_{BM3}$ at 19.82 min.

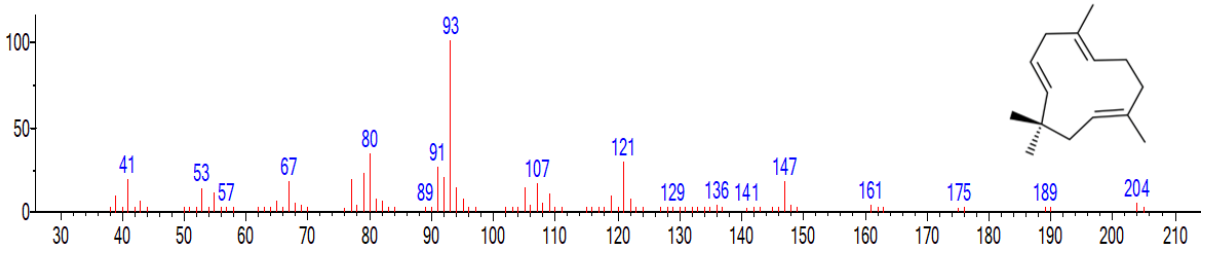


Supplementary Figure 4-32 Mass spectra from the extract of EPY300 expressing TPS27 with Mutant P450BM3 at 20.02 min.

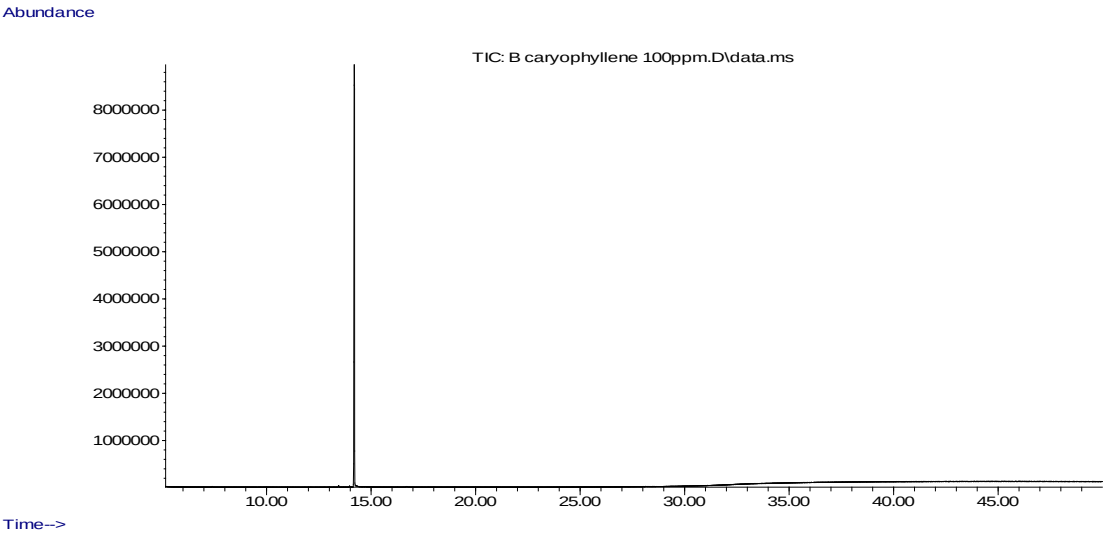
GC-MS Data for Authentic Standards



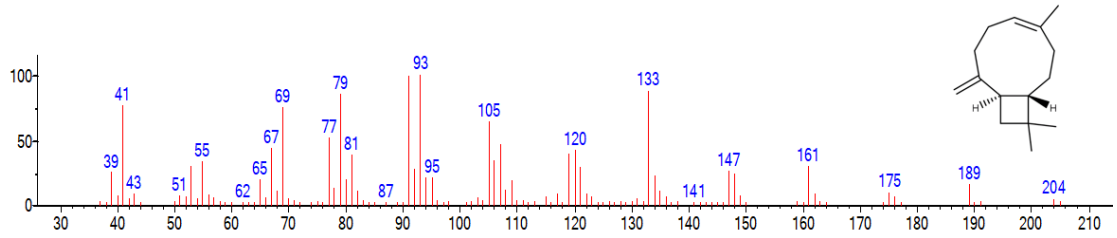
Supplementary Figure 4-33 TIC of α-caryophyllene standard (100ppm)



Supplementary Figure 4-34 Mass spectrum of peak at 14.73 min of α-caryophyllene standard

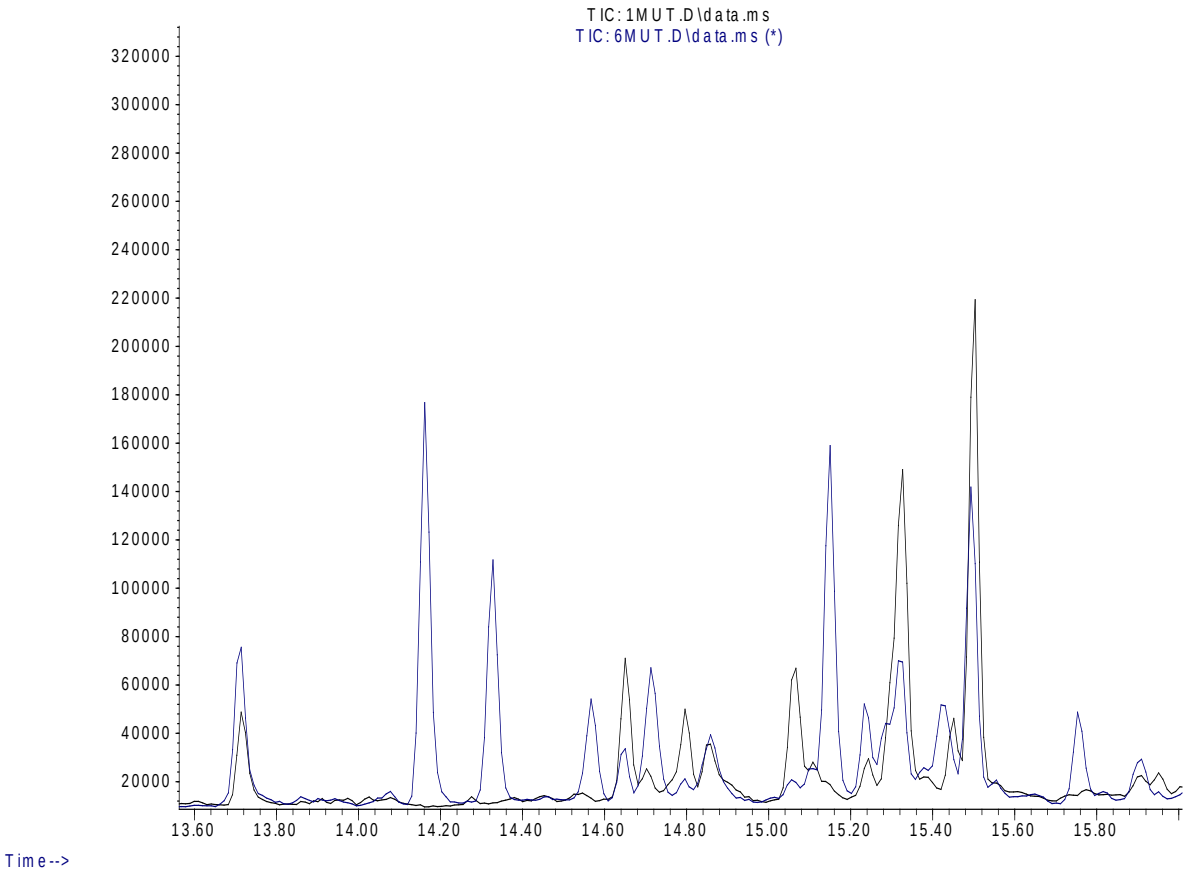


Supplementary Figure 4-35 TIC of β-caryophyllene standard (100ppm)

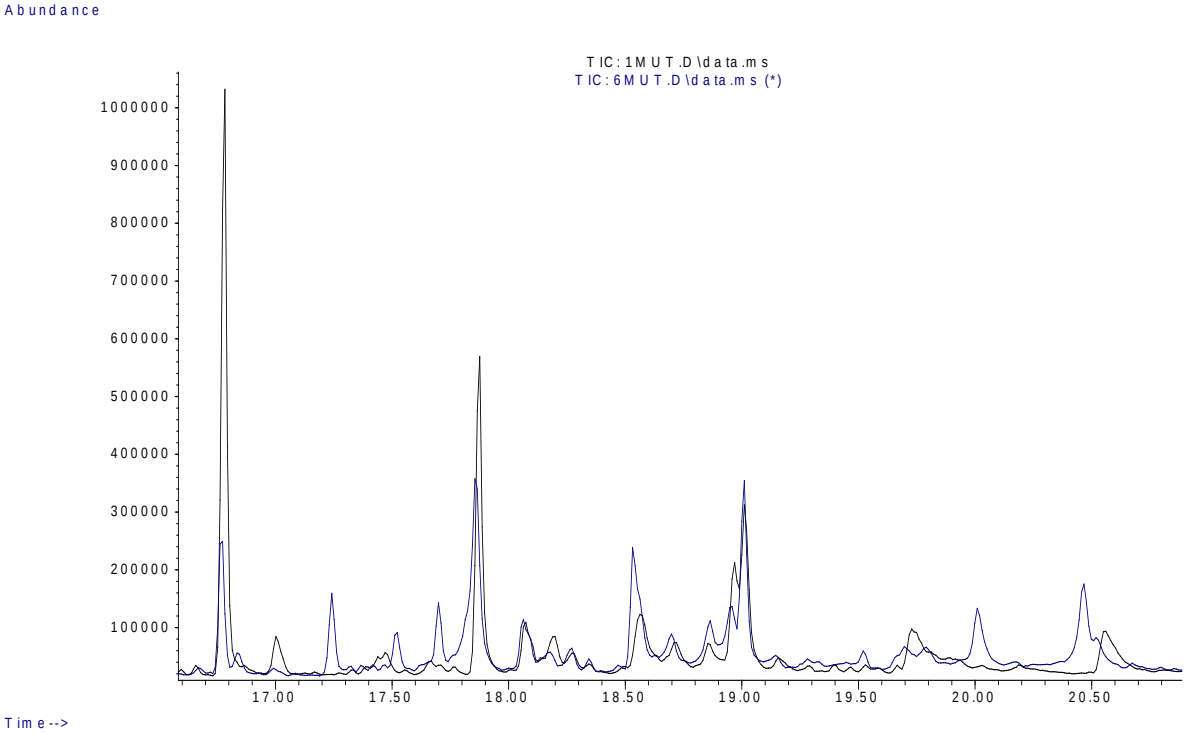


Supplementary Figure 4-36 Mass spectrum of peak at 14.20 min of β-caryophyllene standard

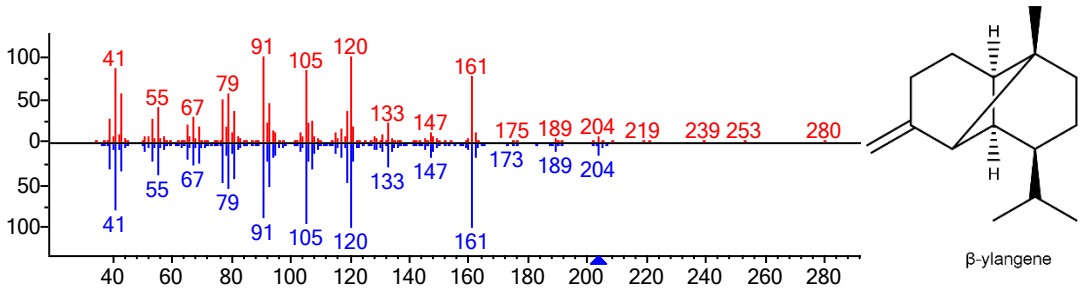
GS-MS Data for *TPSY1*



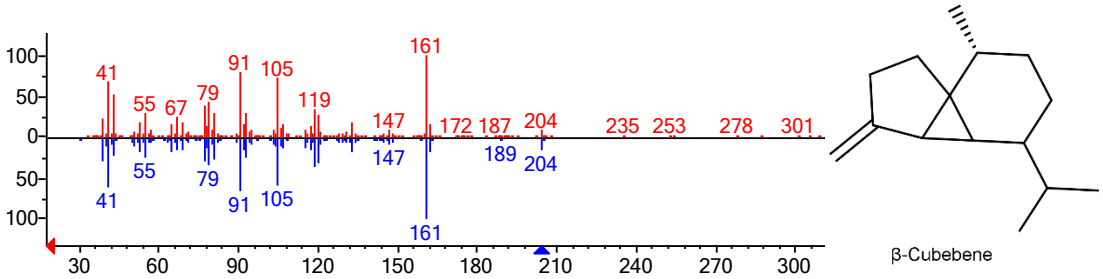
Supplementary Figure 4-37 TIC of extract from EPY300 expressing TPSY1 with Mutant P450$_{BM3}$ (blue) overlayed with TIC of extract from the Mutant P450$_{BM3}$ only control (black). New peaks are present at 14.16 min, 14.33 min, 14.58 min, 14.17 min, 15.15 min and 15.77 min.



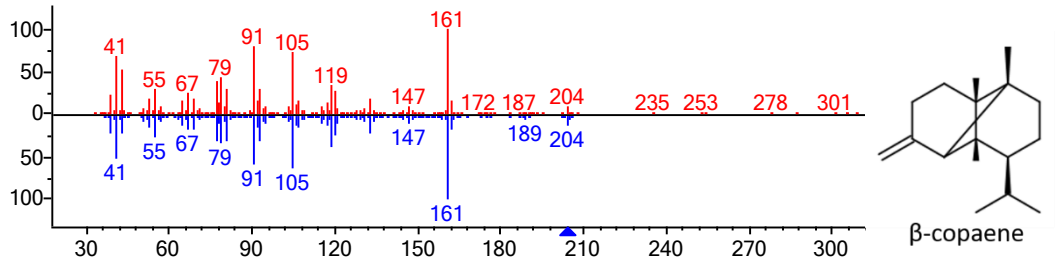
Supplementary Figure 4-38 TIC of extract from EPY300 expressing TPSY1 with Mutant P450$_{BM3}$ (blue) overlayed with TIC of extract from the Mutant P450$_{BM3}$ only control (black). New peaks are present at 17.25 min, 17.52 min, 17.70 min, 20.02 min and 20.47 min.



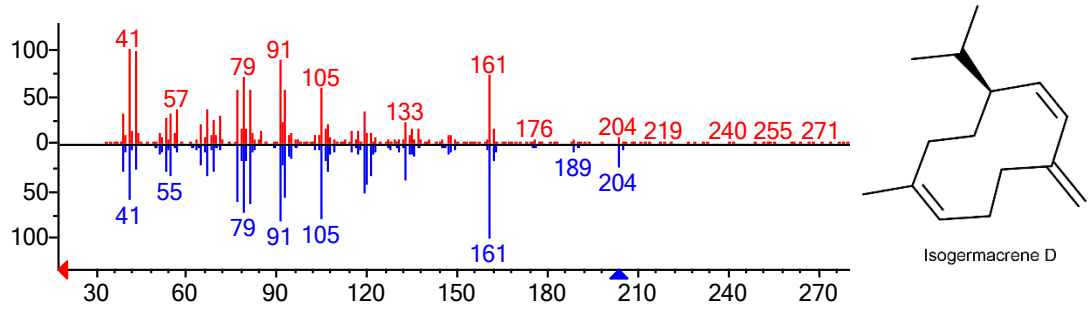
Supplementary Figure 4-39 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPSY1with Mutant P450$_{BM3}$ at 14.16 min (red) with the library spectrum of β-ylangene (blue).



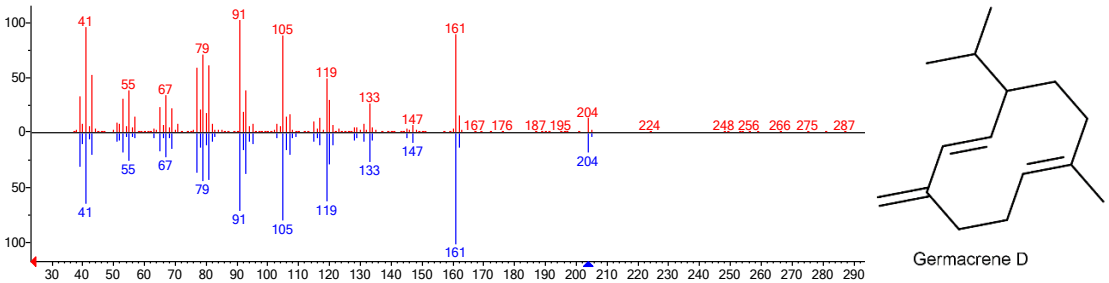
Supplementary Figure 4-40 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPSY1with Mutant P450$_{BM3}$ at 14.33 min (red) with the library spectrum of β-cubebene(blue).



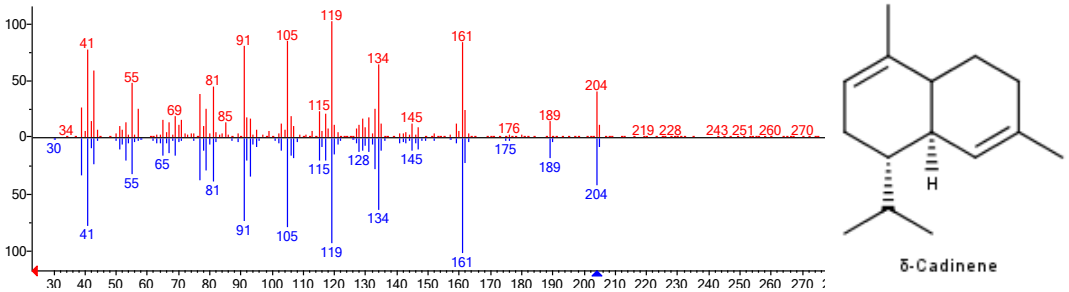
Supplementary Figure 4-41 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPSY1with Mutant P450$_{BM3}$ at 14.33 min (red) with the library spectrum of β-copaene (blue).



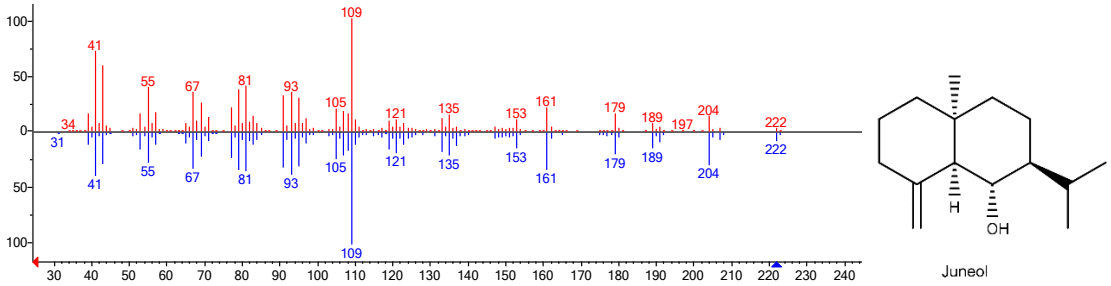
Supplementary Figure 4-42 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPSY1with Mutant P450$_{BM3}$ at 14.58 min (red) with the library spectrum of isogermacrene D (blue).



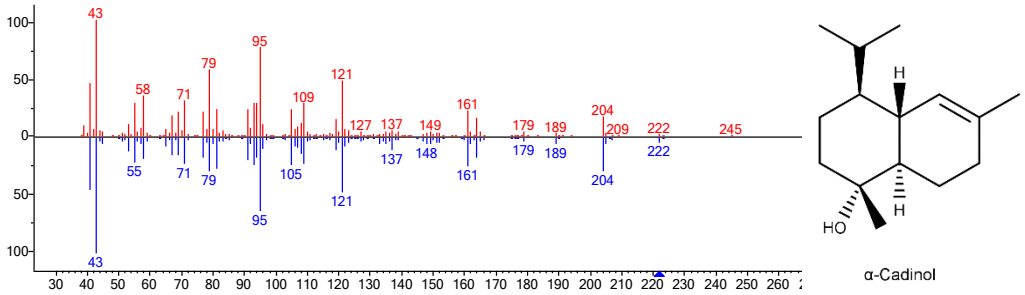
Supplementary Figure 4-43 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPSY1with Mutant P450$_{BM3}$ at 15.15 min (red) with the library spectrum of germacrene D (blue).



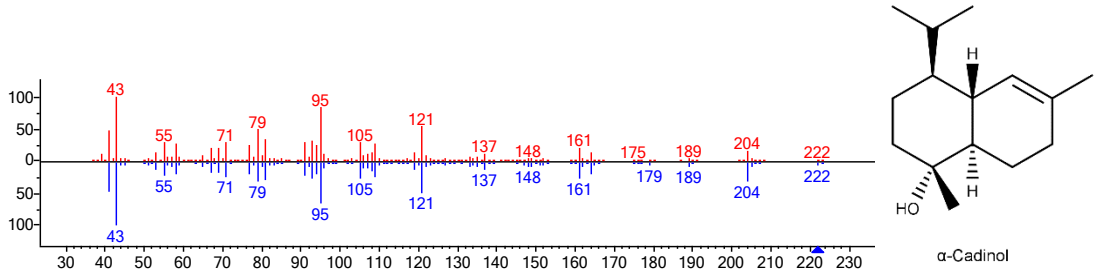
Supplementary Figure 4-44 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPSY1with Mutant P450$_{BM3}$ at 15.77 min (red) with the library spectrum of δ-cadinene (blue).



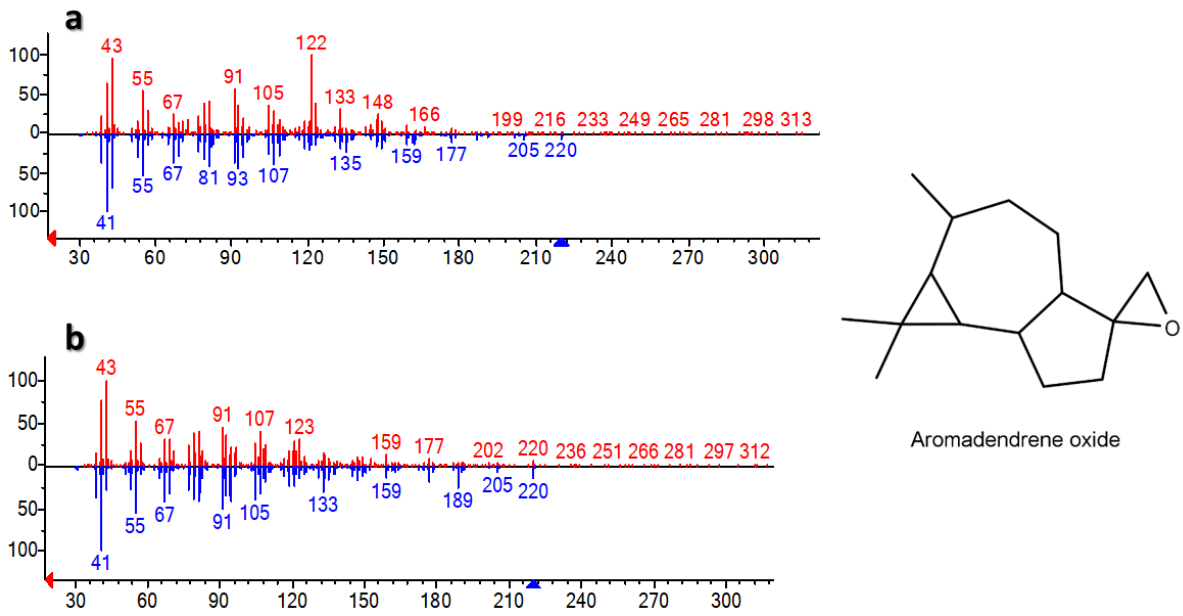
Supplementary Figure 4-45 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPSY1with Mutant P450$_{BM3}$ at 17.25 min (red) with the library spectrum of juneol (blue).



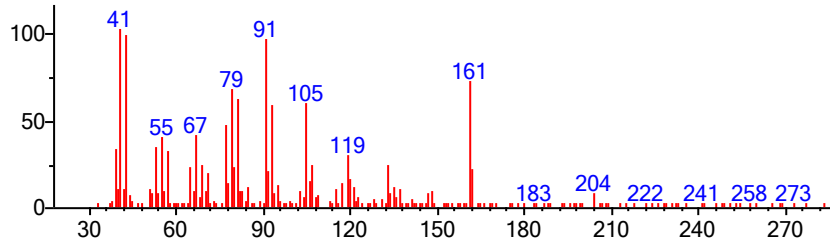
Supplementary Figure 4-46 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPSY1with Mutant P450$_{BM3}$ at 17.52 min (red) with the library spectrum of α-cadinol (blue).



Supplementary Figure 4-47 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPSY1with Mutant P450$_{BM3}$ at 17.70 min (red) with the library spectrum of α-cadinol (blue).

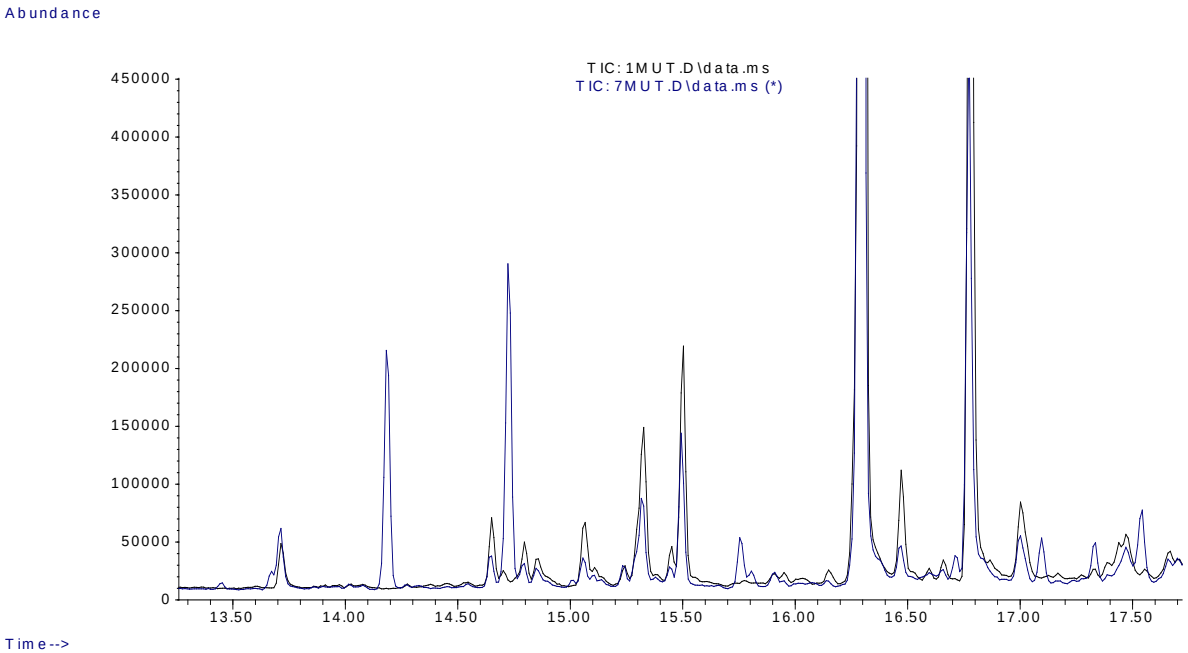


Supplementary Figure 4-48 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPSY1with Mutant P450$_{BM3}$ at a) 20.02 min and b) 20.47 min (red) with the library spectrum of aromadendrene oxide (blue).

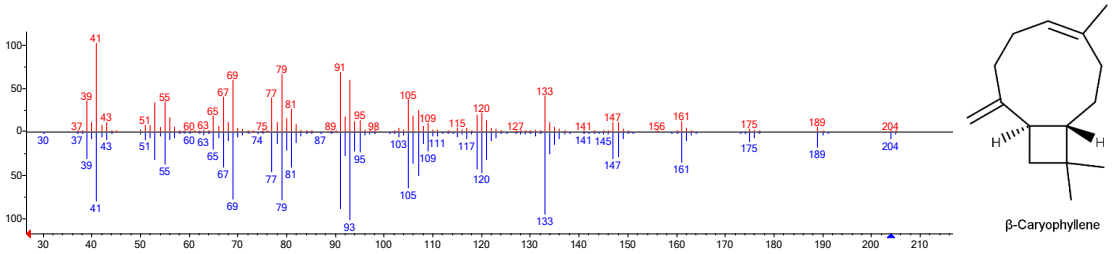


Supplementary Figure 4-49 Mass spectra from the extract of EPY300 expressing TPS27 with Mutant P450$_{BM3}$ at 14.71 min.

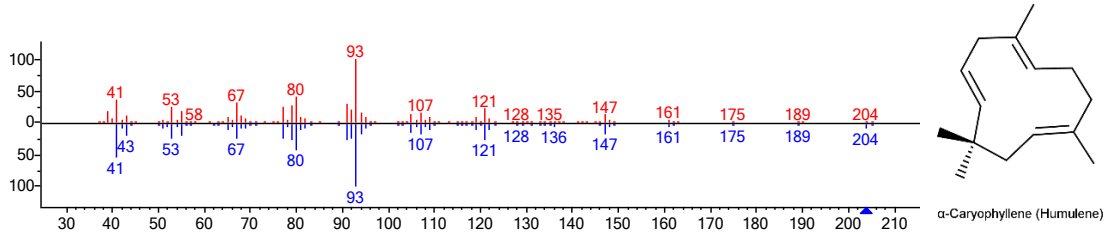
GC-MS Data for *TPSY2*



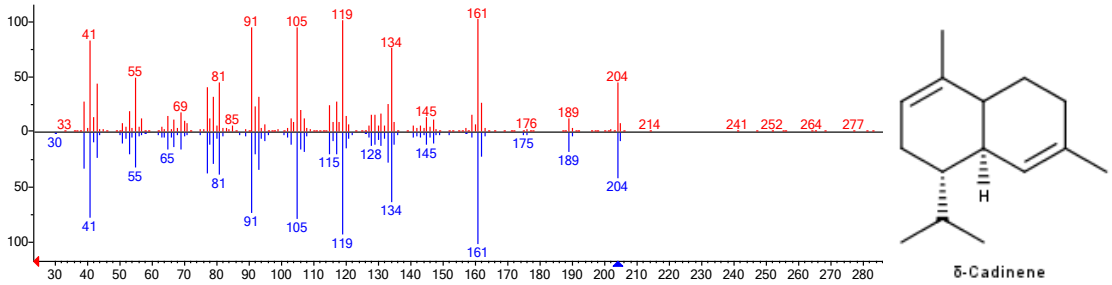
Supplementary Figure 4-50 TIC of extract from EPY300 expressing TPSY2 with Mutant P450$_{BM3}$ (blue) overlayed with TIC of extract from the Mutant P450$_{BM3}$ only control (black). New peaks are present at 14.18 min, 14.73 min, 15.75 min, 17.33 min and 17.54 min.



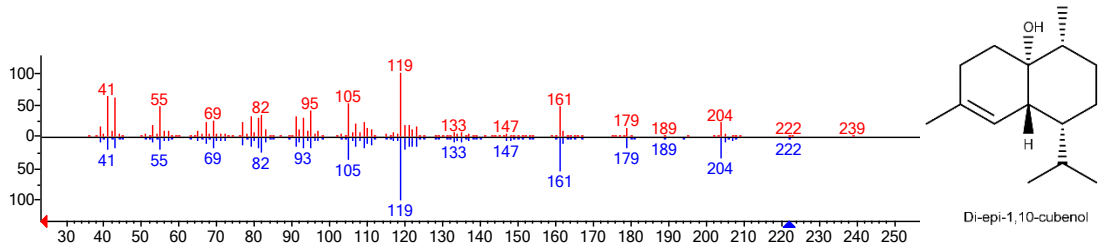
Supplementary Figure 4-51 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPSY2 with Mutant P450$_{BM3}$ at 14.18 min (red) with the library spectrum of β-caryophyllene (blue).



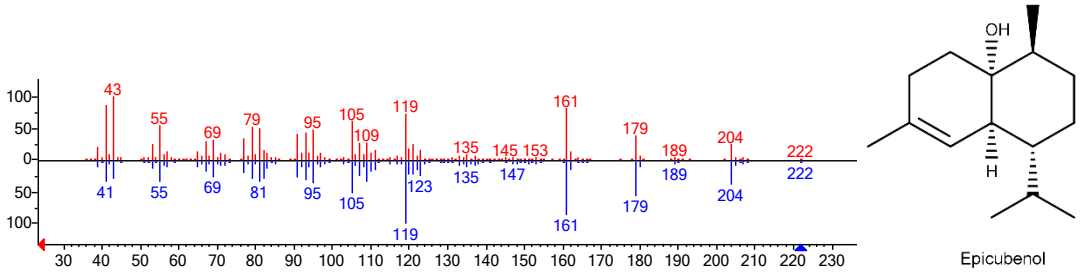
Supplementary Figure 4-52 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPSY2 with Mutant P450$_{BM3}$ at 14.73 min (red) with the library spectrum of α-caryophyllene (humulene) (blue).



Supplementary Figure 4-53 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPSY2 with Mutant P450$_{BM3}$ at 15.75 min (red) with the library spectrum of δ-cadinene (blue).



Supplementary Figure 4-54 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPSY2 with Mutant P450$_{BM3}$ at 17.33 min (red) with the library spectrum of di-epi-1,10-cubenol (blue).



Supplementary Figure 4-55 Head-to-tail comparison of mass spectra from the extract of EPY300 expressing TPSY2 with Mutant P450$_{BM3}$ at 17.54 min (red) with the library spectrum of epicubenol (blue).

Chapter 5 Molecular Modelling of Grapevine Cytochrome P450s

Introduction

Cytochrome P450s are a class of haemoproteins that are capable of oxygenating a wide variety of substrates and catalysing a range of interesting reactions, including hydroxylation, dehydrogenation, double-bond rearrangement, and carbon-bond cleavage (Manikandan & Nagini 2018). P450s primarily act as mono-oxygenases. They catalyse the cleavage of molecular dioxygen to insert a single oxygen atom onto a substrate, while the second oxygen atom is reduced to water. The reducing equivalents are provided by NADH or NADPH, which is facilitated by a redox partner that varies depending on the system. Overall, Cytochrome P450 systems catalyse the following reaction:

$$RH + O_2 + NAD(P)H + H^+ \rightarrow ROH + H_2O + NAD(P)^+$$

Cytochrome P450s are involved in the biosynthesis of many plant specialised metabolites, and are responsible for the production of a number of valuable plant compounds. For example, CYP71AV1 from *Artemisia annua* catalyses the three-step oxidation of amorpha-4,11-diene to artemisinic acid, a precursor to the anti-malarial drug artemisinin (Ro et al. 2006). Several P450s, including CYP71AV8 from *Cichorium intybus*, have been shown to convert (+)-valencene, a cheap and abundant sesquiterpene, to (+)-nootkatone, a natural compound that is valuable to the fragrance and flavouring industries (Cankar et al. 2011). As P450s are able to convert a broad range of substrates, they are valuable catalysts for biotechnology, and the identification and understanding of P450s has the potential to lead to the discovery of novel flavour and aroma molecules.

The wine grape (*Vitis vinifera*) produces a number of P450-oxygenated metabolites that contribute to the flavour and aroma of wine, including monoterpenoids, flavonoids and sulphur compounds (Rapp & Mandery 1986; Yao et al. 2004). (-)-Rotundone, an oxygenated sesquiterpene, has been identified as a potent peppery aroma molecule (Wood et al. 2008). Rotundone is biosynthesised from the sesquiterpene α-guaiene in a two-step oxidation at the C2 position via rotundol (Figure 5-1) (Huang et al. 2014; Takase et al. 2016). Although this reaction can occur through aerial oxidations (Huang et al. 2014), Takase et al. (2016) have identified a cytochrome P450, *Vitis vinifera* Sesquiterpene Oxidase 2 (VvSTO2), capable of adding a ketone moiety on C2 position of α-guaiene backbone to produce (-)-rotundone. Rotundone has been known to be more abundant in the grape cultivar Syrah than other cultivars (Wood et al. 2008).

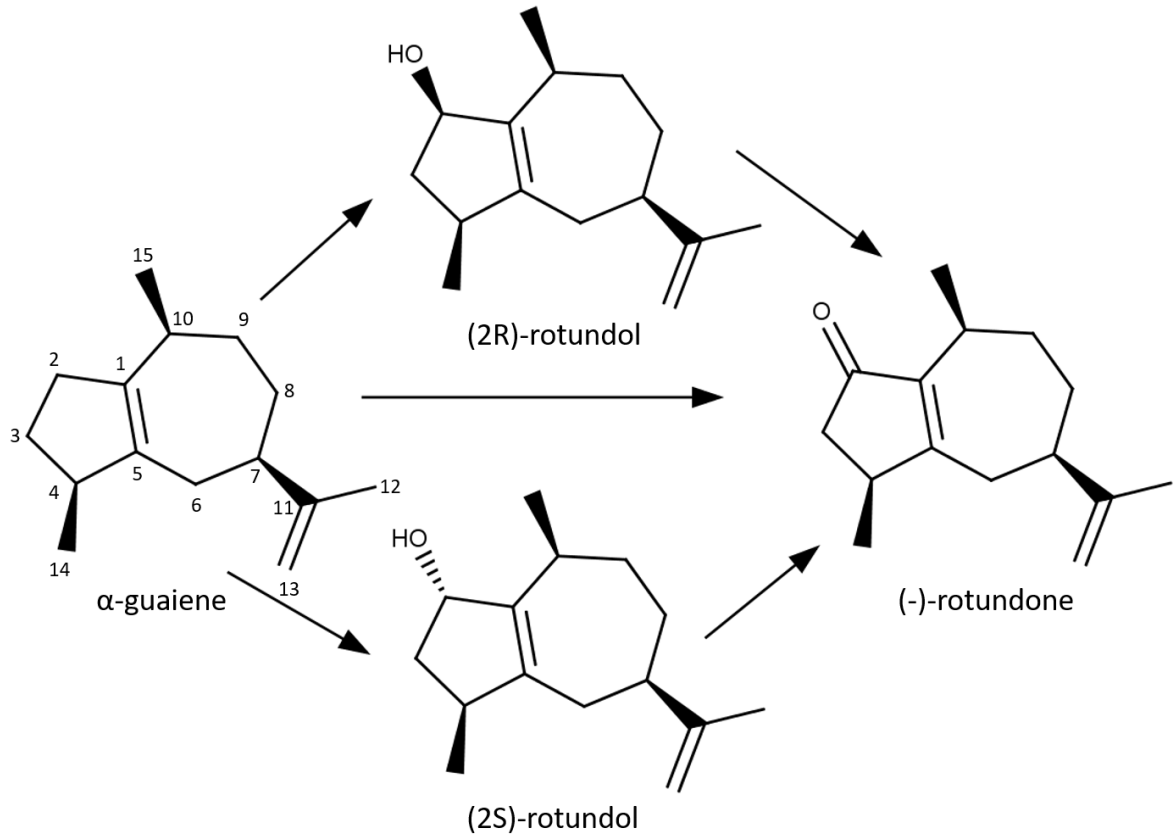


Figure 5-1 Biosynthetic pathway of (-)-rotundone from α-guaiene

Along with STO2, Takase et al. (2016) identified two other P450s, STO4 and STO6, in the same CYP71BE subfamily that accept α-guaiene as a substrate. STO4 and STO6 share 71% amino acid sequence identity with each other and 61-62% sequence identity with STO2. Despite their similarities, STO4 and STO6 are unable to convert α-guaiene to (2R)-rotundol, (2S)-rotundol or (-)-rotundone. This suggests STO2 is unique within the CYP71BE subfamily. However, feeding α-guaiene to microsomal STO4 and STO6 proteins resulted in the formation of several, as yet unidentified products, including two products in common between the two P450s (Takase et al. 2016). Comparison of the structural and functional characteristics of STO2, STO4 and STO6 can lead to a better understanding of the substrate specificity and regio-specificity of STO2. As STO2 catalyses the formation of a valuable aroma molecule, understanding its mechanism could have implications for biotechnology and allow large-scale production of the molecule. Furthermore, investigating the roles of cytochrome P450s in grapevine could lead to the discovery of novel flavours and aromas so far unidentified in grapevine and wine.

Here, comparative modelling and docking techniques were used to investigate the binding of α-guaiene within the binding sites of STO2, STO4 and STO6. Key residues that influence the binding orientation of α-guaiene and P450 reaction specificity were identified. Finally,

a yeast strain for the production of α-guaiene was created that can be used to test and analyse STO mutants.

Results

To investigate the mechanism and binding of P450s STO2, STO4 and STO6, α-guaiene was docked into a model structure of each enzyme. No crystal structures are available for STO2, STO4 or STO6 so it was necessary to use a homology modeling approach. Using protein-protein BLAST, Human Cytochrome P450 CYP17A1 (PDB accession 3RUK) was identified as a suitable template due to homology with the target P450s (27-28% for each of STO2, STO4 and STO6). Additionally, 3RUK is in complex with Abiraterone, a rigid, cyclic ligand, which may increase the speed and accuracy of docking α-guaiene, due to its similar structural properties. Models were evaluated using VADAR (Volume Area Dihedral Angle Reporter) (Willard et al. 2003). The Stereo/Packing and 3D Profile Quality indices are available in Supplementary Information (Supplementary Figure 5-1 and Supplementary Figure 5-2).

Haem was docked into each of the grapevine P450 models using Autodock 4.2.6. Subsequently, α-guaiene was docked into each of the P450-haem complexes. As α-guaiene is relatively rigid, with only two rotatable bonds, the default docking parameters were sufficient for this application. Docking results were validated by repetition, and the results were found to be reproducible. The docking result conformation information is displayed in Table 5-1. The binding energy is calculated as the sum of the intermolecular energy and the torsional free-energy penalty.

Table 5-1 Binding energy of the most stable conformations of α-guaiene docked into the binding sites of STO2, STO4 and STO6. Binding energy was calculated by Autodock as the sum of the intermolecular energy and the torsional free-energy penalty.

Docking Result	Binding Energy (arbitrary units)
STO2	-8.60
STO4	-6.94
STO6-A	-8.18
STO6-B	-8.11

Docking of α-guaiene into the model of STO2 shows the 5-membered ring in closest proximity to the porphyrin ring (Figure 5-2). C3 is nearest to the haem iron (3.0 Å). C2, which is oxidised to form (-)-rotundone, is positioned at a distance of 4.2 Å from the haem iron.

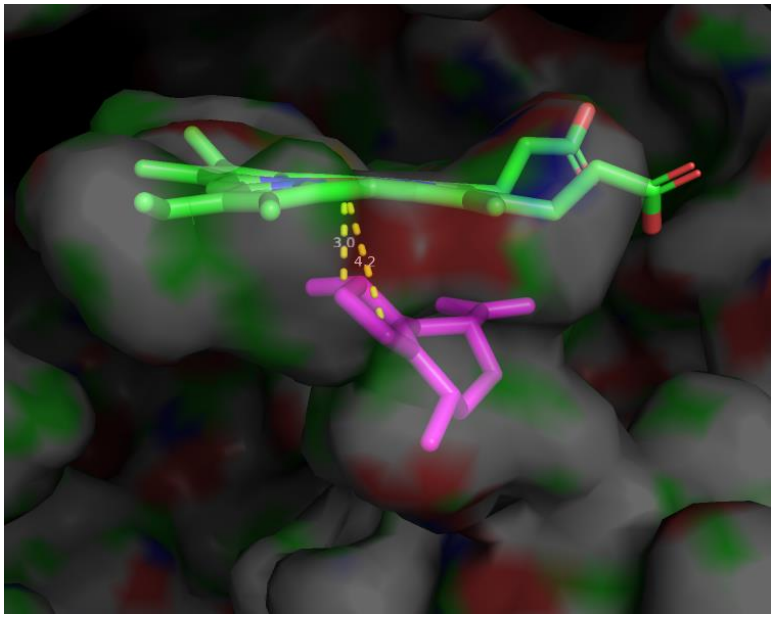


Figure 5-2 Docking of α-guaiene into the model structure of VvSTO2 shows C3 in closest proximity to the haem iron (3.0 Å). C2, which is oxidised to form (-)-rotundone is positioned 4.2 Å from the haem iron.

Docking of α-guaiene into the model of STO4 showed the substrate oriented such that C12 and C13 were closest to the porphyrin ring (3.4 Å and 4.0 Å) (Figure 5-3).

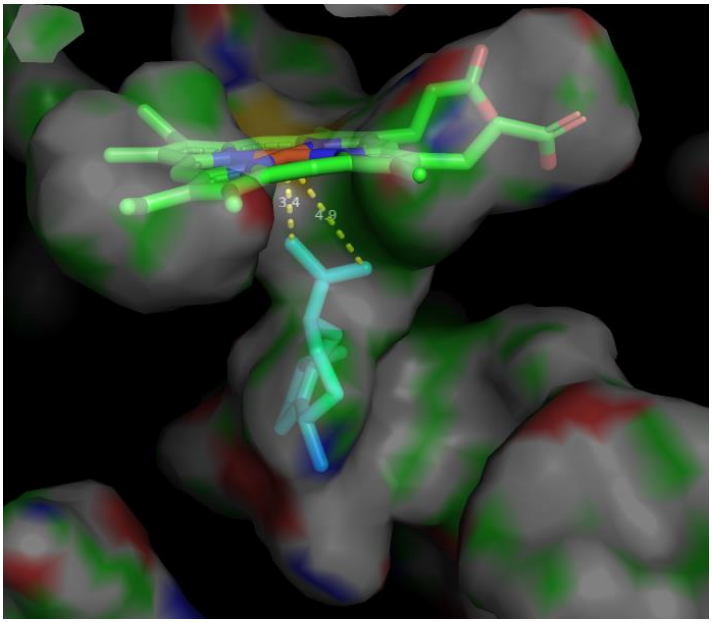


Figure 5-3 Docking of α-guaiene into the model structure of VvSTO4 shows a very different conformation to that of VvSTO2. Here, C12 and C13 are positioned closest to the haem iron, at distances of 3.4 and 4.0 Å.

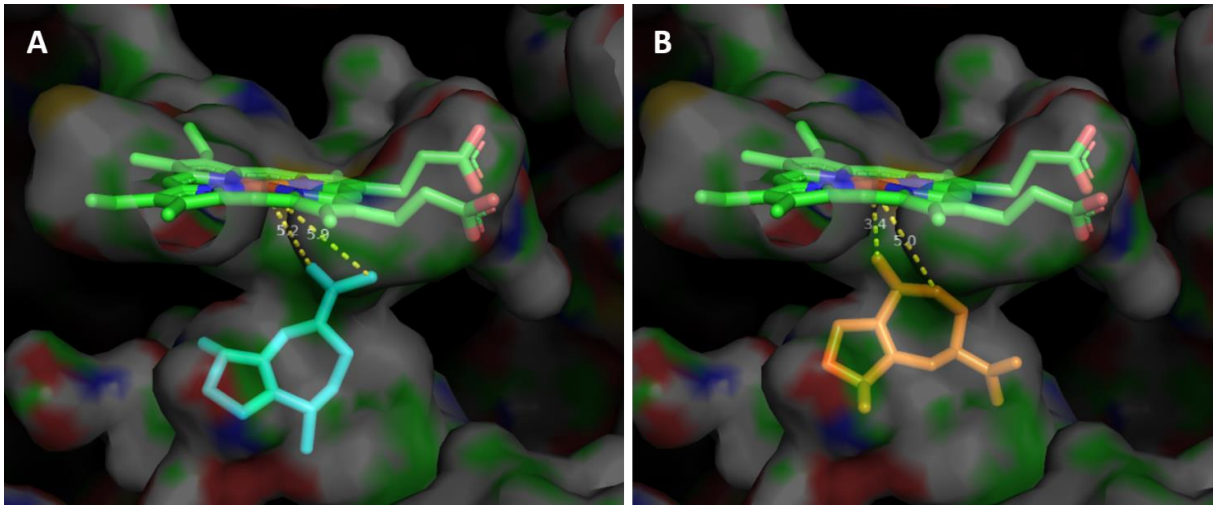


Figure 5-4 Docking of α-guaiene into the model structure of VvSTO6 gives two distinct conformations. A) α-guaiene is positioned with C12 and C13 in closest proximity to the haem iron, at 5.0 and 5.2 Å. B) α-guaiene is positioned with C9 and C15 in closest proximity to the porphyrin ring (3.4 and 5.0 Å respectively).

Docking of α-guaiene into the model of STO6 produced two different stable conformations (Figure 5-4). In the first conformation, α-guaiene is positioned in a similar orientation to that of the STO4 binding, with C12 and C13 closest to the porphyrin ring (5.0 Å and 5.2 Å) (Figure 5-4A). In the second conformation, α-guaiene is positioned such that the 7-membered ring is closest to the porphyrin ring with C9 and C15 in closest proximity to the haem iron (3.4 and 5.0 Å respectively, Figure 5-4B).

As these docking results show a clear difference in the positioning of α-guaiene within the binding site, it was hypothesised that key residues in the P450 binding pocket would alter the orientation of α-guaiene within the active site and, subsequently, the regioselectivity of the reaction. Multiple sequence alignment of the grapevine P450s was performed using ClustalW (Figure 5-5). Residues within 4 Å of the docked α-guaiene substrate were identified using PyMol. From this, three residues in particular were identified as being significantly different between STO2 and STO4 and STO6 (Table 5-2).

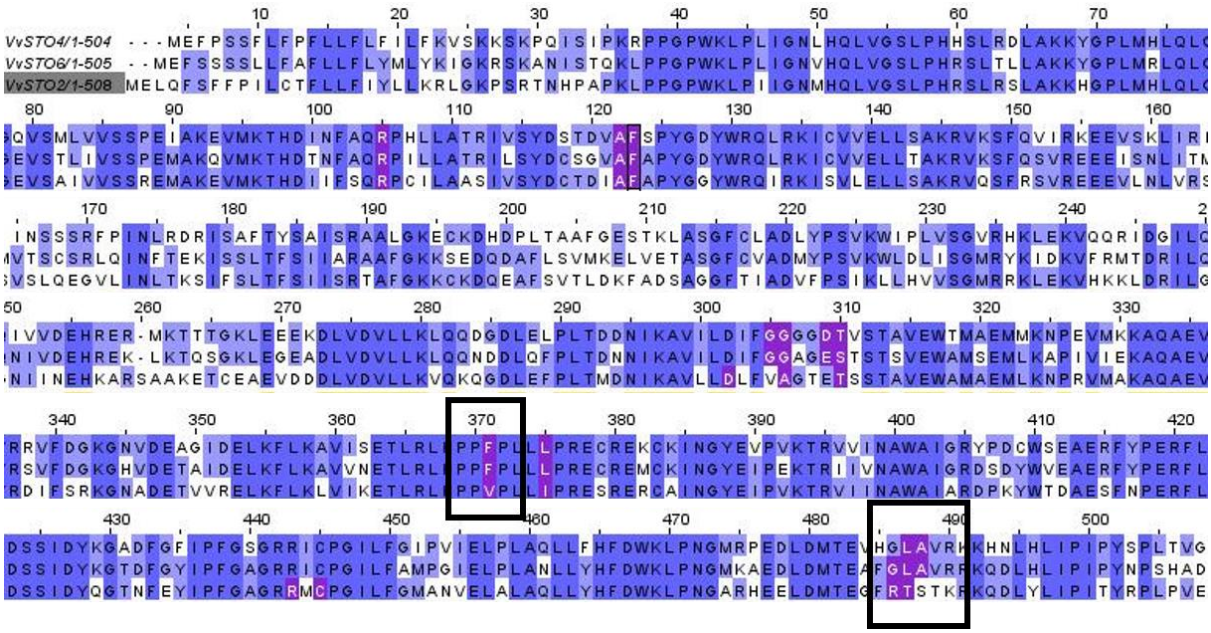


Figure 5-5 Amino acid sequence alignment of VvSTO4, VvSTO6 and VvSTO2. Multiple sequence alignment was performed by ClustalW and visualised using BioEdit. Similar amino acids are highlighted in blue. Amino acids within 4 Å of the docked substrate are highlighted in purple. Differing amino acids within 4 Å of the binding sites are outlined.

Table 5-2 Amino acids within 4 Å of the docked α-guaiene substrate. Docking was performed using Autodock 4.2.6 and measurements performed using PyMol. Dissimilar amino acids are highlighted in green.

STO2	**STO4**	**STO6**
A122		R103
F123	F120	F121
	G301	G302
	G302	G303
D302	D305	E306
A306		
T310	T307	S308
V371	F367	F368
I375	L371	L372
R443		
C445		
		G483
R486	L483	L484
T487	A484	A485

It is proposed that mutation of V371, R486 and T487 may alter the regioselectivity of the α-guaiene oxidation reaction by STO2 and lead to the formation of novel products. Furthermore, mutation of STO4 and STO6 at their corresponding amino acid positions may enable these enzymes to convert α-guaiene to rotundone.

In order to test this, it is necessary to obtain a source of α-guaiene. α-Guaiene is not available commercially, although it can be extracted from agarwood (Wood et al. 2008). An alternative method is to construct a yeast strain for the production of α-guaiene that can subsequently be used to test different P450 mutants *in vivo*.

To produce α-guaiene in yeast, EPY300 was transformed with pESC-Leu containing a codon-optimised synthetic gene *Vv*TPS24 (Supplementary Information), a grapevine

sesquiterpene synthase that produces α-guaiene as one of its major products (Drew et al. 2016). The culture was analysed by GC-MS for sesquiterpene products. The presence of α-guaiene was confirmed by comparing the sample mass spectrum to a standard mass-fragmenting pattern of α-guaiene in NIST EI-MS library. (Figure 5-6).

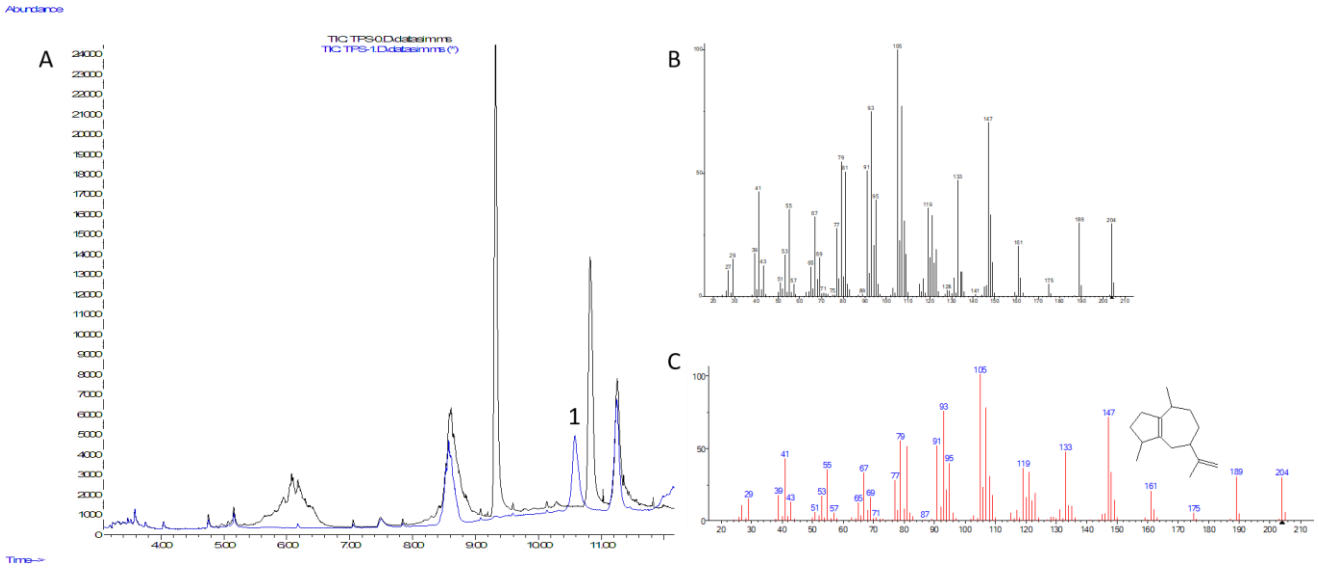


*Figure 5-6 GC-MS data from EPY300 pESC-LEU2d-*TPS24 *strain. A) Overlay of total ion chromatograms (TIC) from control strain (black) and experimental strain (blue). B) Mass spectrum of peak 1. C) Mass spectrum of α-guaiene, from NIST library.*

Discussion

Modelling the binding of the α-guaiene substrate with the different *V. vinifera* P450s provides insight into the reaction specificity and the possible resulting products. Since experimental crystal structures of STO2, STO4 and STO6 are not available, the most reliable tool for obtaining structural data on these targets is to use computational modelling based on the three-dimensional structure of a homologous protein (Suresh Kumar, Thomas & Poornima 2018). Human Cytochrome P450 CYP17A1 (3RUK) was chosen as a template, and shares a sequence homology of 27-28% with the grapevine P450s. Sequence homology is not the only consideration when choosing a template; the functional context of the structure must also be taken into account (Schmidt, Bergner & Schwede 2014). *Naegleria fowleri* CYP51 (PDB accession 5TL8) was also identified as a suitable template based on sequence homology, but it is in complex with posaconazole, a large, flexible, linear ligand. Alternatively, 3RUK is in complex with abiraterone, a smaller and more rigid compound that bears greater structural resemblance to α-guaiene (Figure 5-7). Therefore, using 3RUK as a template is more likely to give fast and accurate results

when the P450 models are used to dock α-guaiene. Although the sequence homology between the STOs and 3RUK is relatively low, the structural folding pattern of distantly related P450s is surprisingly conserved (Poulos & Johnson 2015; Schmidt, Bergner & Schwede 2014), so the 3RUK crystal structure can still be a reliable template. However, the crystal structure is limited to 2.6 Å resolution, which will further limit the resolution of the model.

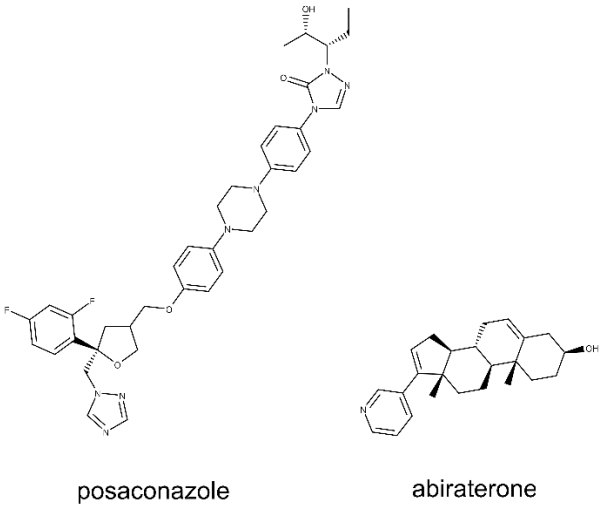


Figure 5-7 The structures of posaconazole, the 5TL8 ligand, and abiraterone, the 3RUK ligand. Abiraterone has a smaller and more rigid structure that bears greater resemblance to α-guaiene, so 3RUK was chosen as a more appropriate template for homology modelling.

Although computational modelling of proteins is limited by the resolution of the template and the similarity of the target and template proteins, it has been shown to be an appropriate and reliable technique for the study of P450 structure and function. There are many examples in the literature where P450 modelling has provided insights to structure/function relationships that are supported by experimental mutation studies (Eichler et al. 2016; Luirink et al. 2018; Nikolaus et al. 2017). Therefore, computational modelling of P450s is a reliable method that can be used to gain insights into enzyme function and provide initial directions for experimental work. The STO models were assessed using VADAR and the results show reasonable quality indices. The 3D profile index assesses the local environment, packing and hydrophobic energy for the given structure (Willard et al. 2003). The majority of residues have quality indices which fall within the typical range of 5-8, and problem areas (scores below 4) are not located around the active site (Supplementary Figure 5-1). The stereo/packing index assesses backbone angles, and the presence of packing defects such as atomic overlap or large gaps (Willard et al. 2003). The majority of residues have stereo/packing indices above the benchmark of 6, although some defects are present (Supplementary Figure 5-2).

Nevertheless, the quality indices show that in general, local protein structure and folds are spatially reasonable, confirming that these models are appropriate for docking and binding site analysis.

Docking α-guaiene into the substrate recognition sites of each of the three STOs may improve understanding of the product specificity of these P450s. The docking binding energy can be used to compare the stability of the docked conformations and assess the most stable conformation (Table 5-1). The binding energy for all three STOs was negative, indicating binding of α-guaiene to the proteins is energetically favourable. Docking α-guaiene with STO2 and STO6 resulted in similar binding energies, but docking α-guaiene into STO4 gives a less favourable binding energy, indicating that STO4 may have lower binding affinity for α-guaiene. There is little difference in stability between the two different STO6 docked conformations, suggesting both binding orientations would readily arise.

Within the model structure of STO2, α-guaiene is positioned such that C2, which is selectively oxidised by STO2 to form (-)-rotundone, is positioned 4.2 Å from the haem iron (Figure 5-2). This distance is similar to iron-substrate distances reported for co-crystallised structures of other P450-substrate complexes. For example, $P450_{cam}$ and $P450_{eryF}$ have been observed to bind fatty acids with a distance of 4.2-4.8 Å between the atom that is hydroxylated and the porphyrin iron (Li & Poulos 1997). The docking model shows that C2 is positioned within ideal distance of the catalytic iron atom, explaining the formation of rotundone by STO2. It should be noted that C3 is modelled to be closest to the haem iron, at a distance of 3.0 Å. However, this distance is relatively short, and not within the typical catalytic distance of P450s. C3 may bind in this position but avoid oxidation due to steric hindrance, or the measured distance could be incorrect due to the limitations of the model.

Despite the limitations of the model, the modelled binding orientation of α-guaiene within the binding pocket of STO2 shows a clear contrast to the STO4 and STO6 docking models. The distinctly different orientations of the substrate within the binding site explain why STO4 and STO6 do not produce rotundone. α-Guaiene docks within the STO4 binding pocket such that C12 and C13 are in closest proximity to the haem iron (Figure 5-3), so it is expected that oxidation would occur at these sites. Docking of α-guaiene into the model structure of VvSTO6 gives two distinct conformations (Figure 5-4): one where α-guaiene is positioned with C12 and C13 in closest proximity to the haem iron, and another with α-guaiene positioned with C9 and C15 in closest proximity to the porphyrin ring. In all three of these binding orientations, C2 is positioned further away from the haem iron and is

therefore not readily oxidised by STO4 and STO6, hence STO4 and STO6 do not form (-)-rotundone or (-)-rotundol.

Previously, STO4 and STO6 have been observed to oxidise α-guaiene to form several, currently unidentified, compounds, including two compounds produced by both enzymes (Takase et al. 2016). The similarities between the orientation of α-guaiene within the STO4 binding site (Figure 5-3) and the first orientation of α-guaiene within the STO6 binding site (Figure 5-4A) may explain the two products in common, as in both these orientations C12 and C13 are in closest proximity to the haem iron. In the STO6 model, C12 and C13 are positioned at a greater distance from the haem iron (5.0 and 5.2 Å), but P450 oxidation has still been shown to occur at this distance. X-ray crystallography and NMR data have shown $P450_{BM3}$ binds substrates at a distance of 7.5-7.9 Å from the porphyrin ring (Li & Poulos 1997), so the distance observed between the STO6 haem iron and the α-guaiene substrate is not catalytically unreasonable. Both STO4 and STO6 would be expected to oxidise α-guaiene at C12 and C13 and form the potential products shown in Figure 5-8. While the masses of these products are consistent with the GC-MS data obtained by Takase et al. (2016) for the unidentified peaks, further confirmation such as direct comparison of the GC-MS with a standard or NMR analysis is needed to accurately identify the products.

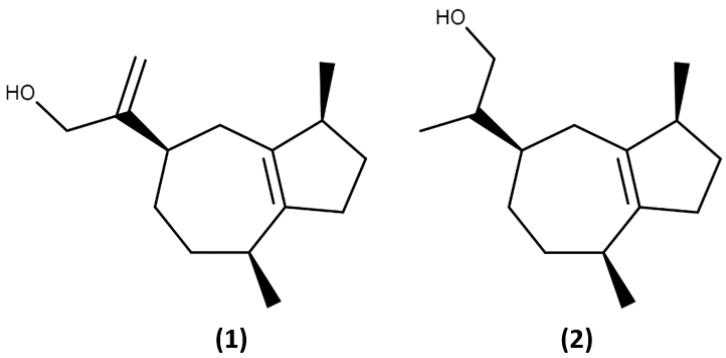


Figure 5-8 Potential products formed through the oxidation of α-guaiene by STO4 and STO6

Previous studies have shown that mutations in the substrate recognition site of plant P450s can alter the substrate selectivity, product specificity and regiospecificity of the oxidation reaction (Kahn et al. 2001; Komori et al. 2013; Takahashi et al. 2005). Since STO2, STO4 and STO6 share high homology (60-71%), it was hypothesised that key residues within the substrate binding sites could account for the difference in products formed by these enzymes. Residues within 4 Å of the docked substrate were compared (Table 5-2). While STO4 and STO6 share highly similar binding residues, there are three key differences identified within the STO2 binding site. It is expected that mutation of

these residues would alter the binding of α-guaiene and the products formed. The change from valine at position 371 to phenylalanine would introduce steric effects that change the shape of the binding pocket, and introduce a new hydrogen bond donor. Mutation of the arginine in position 486 to leucine would disrupt electrostatic interactions with the positively-charged residue. Finally, replacing threonine with alanine at position 487 would introduce another hydrogen bond donor and further change the shape of the binding site and the interactions with the substrate. However, experimental site-specific mutation studies are required to confirm these assumptions.

In order to conduct site-specific mutation studies, it is necessary to provide a source of α-guaiene to feed to the target enzymes. α-Guaiene can be extracted from agarwood (Wood et al. 2008) or synthesised from guaiol (Takase et al. 2016). Previously, *in vitro* enzyme assays have been performed by the addition of α-guaiene to microsomal proteins (Takase et al. 2016). An alternative approach is to perform *in vivo* assays using a yeast strain engineered to produce α-guaiene. Mutant P450s could be expressed directly by the engineered strain, allowing rapid analysis of products and straightforward genetic manipulation and screening of mutants.

The strain EPY300 has been engineered to overproduce FPP, the sesquiterpene precursor (Ro et al. 2006). This is achieved by upregulation of the mevalonate pathway and the repression of squalene synthesis. EPY300 has been previously used for the *de novo* synthesis of plant sesquiterpenes and oxygenated sesquiterpenoids with typical yields in the range of up to 10 mg (Nguyen, MacNevin & Ro 2012). Several plant sesquiterpene synthases that produce α-guaiene have been identified. Faraldos et al. (2010) discovered a patchoulol synthase from *Pogostemon cablin* that produces α-guaiene as 26% of a mixture of products. Two sesquiterpene synthases were identified from cultured cells of *Aquilaria*, one that produced α-guaiene as its major product (44%) and another that produced α-guaiene as a minor product (18%) (Kumeta & Ito 2010). A particular allele of the *VvTPS24* gene has been shown to encode the sesquiterpene synthase VvGuaS that produces α-guaiene as its major product (44%) (Drew et al. 2016). *VvTPS24* was chosen for expression in EPY300 as it produces a high proportion of α-guaiene and is naturally involved in (-)-rotundone biosynthesis in grape. Furthermore, the use of a grapevine sesquiterpene synthase means the system has the potential to be used to identify novel oxygenated sesquiterpenoids found in grapevine that could be produced from VvTPS24 minor products. *VvTPS24* was expressed in EPY300 and the resulting strain was shown to produce α-guaiene in culture (Figure 5-6). P450 mutants could be tested using this system by transforming the yeast strain with a vector containing the P450 and its redox partner

and analysing the culture for novel products using GC-MS. Product identification may be performed by comparing the GC-MS data against a library or by purification of the product and NMR.

Conclusions

This study has provided preliminary insight into the mechanism and binding of three grapevine P450s, including STO2, which is responsible for the production of rotundone, a valuable peppery aroma in wine. Modelling suggests that differences in the binding orientation of α-guaiene within the P450 binding sites are responsible for the variation in products formed by STO2, STO4 and STO6. Three key residues have been identified within the STO2 binding site that may affect product specificity, but experimental site-specific mutation studies are needed to confirm these hypotheses. A platform for conducting these experiments has been created by constructing a yeast strain capable of producing α-guaiene.

P450s are versatile and valuable catalysts that are capable of producing a large range of compounds in nature. Understanding the mechanism and binding of P450s in grapevine may assist in the discovery of novel flavours and aromas. Furthermore, identification of key residues within the binding sites of P450s allows targeted mutation and has the potential to lead to the formation of new products. Understanding key residues within the substrate recognition sites may assist in identification of novel P450 alleles within plant genomes and the discovery of unique metabolites.

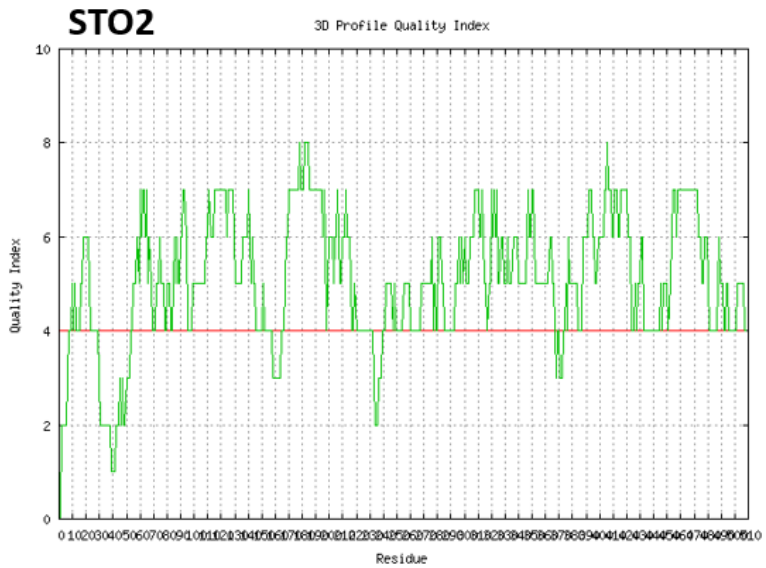
Acknowledgements

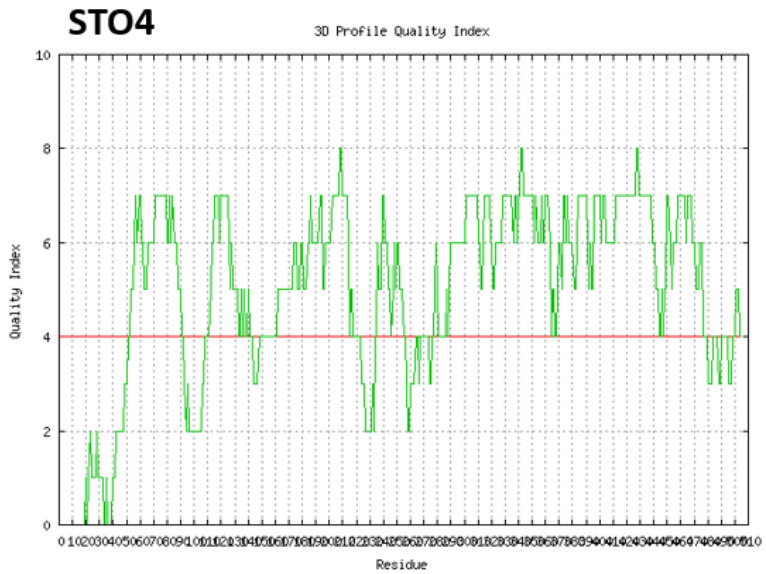
This work was conducted at the University of Calgary under the supervision of Dr Dae-Kyun Ro as part of an Endeavour Research Fellowship. Funding was provided by the Australian Government.

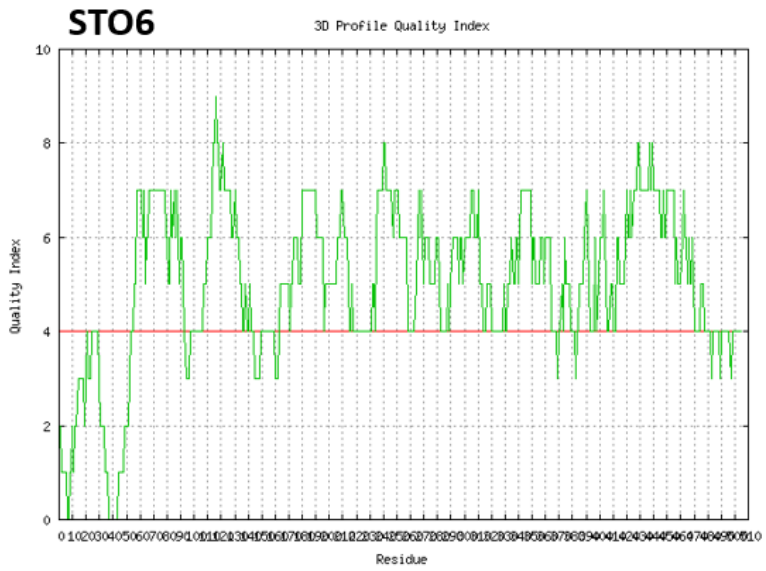
Supplementary Information for Molecular Modelling of Grapevine Cytochrome P450s

Codon Optimised TPS24 (Ordered from Thermo Fisher)

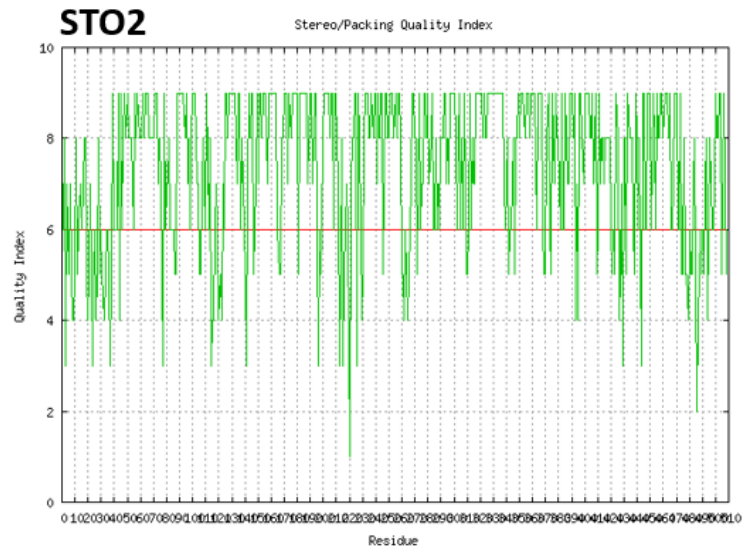
GAAGAAGACCTCGAGGCTATGTCTGTTCCATTGTCTGTTTCTGTTACCCCAATCTTGTCCCAAA
GAATTGATCCTGAAGTTGCTAGACATGAAGCTACTTACCATCCAAATTTTTGGGGTGATAGATT
CTTGCACTACAACCCAGATGATGATTTCTGTGGTACTCATGCTTGCAAAGAACAACAGATCCAA
GAGTTGAAAGAAGAGGTCAGAAAATCTTTGGAAGCTACTGCTGGTAATACCTCTCAGTTGTTGA
AGTTGATCGACTCCATTCAAAGATTGGGTTTAGCCTACCATTTCGAAAGGGAAATTGAAGAGGC
TTTGAAGGCTATGTACCAAACTTACACTTTGGTTGATGATAACGATCACTTGACCACCGTTTCT
TTGTTGTTTAGGTTGTTGAGACAAGAGGGTTACCACATTCCATCTGACGTTTTCAAAAAGTTCA
TGGACGAAGGTGGTAACTTCAAAGAATCATTGGTTGGTGATTTGCCAGGTATGTTGGCATTATA
TGAAGCTGCTCATTTGATGGTTCACGGTGAAGATATTTTGGATGAAGCTTTGGGTTTTACTACC
GCTCACTTGCAATCTATGGCTATCGATTCTGATAACCCATTGACCAAGCAAGTTATCAGAGCTT
TGAAAAGGCCAATCAGAAAAGGTTTGCCTAGAGTTGAAGCCAGACATTACATTACCATCTACCA
AGAGGATGACTCCCATAACGAATCCTTGTTGAAATTGGCTAAGCTGGACTACAATATGTTGCAG
TCCTTGCACAGAAAAGAGTTGTCTGAAATTACCAAGTGGTGGAAAGGTTTGGATTTCGCTACAA
AATTGCCATTCGCTAGAGATAGAATCGTCGAAGGTTACTTTTGGATCTTGGGTGTTTATTTCGA
ACCCCAATATTACTTGGCTAGGCGTATATTGATGAAGGTTTTCGGTGTTTTGTCCATCGTTGAT
GACATCTATGATGCTTACGGTACTTTCGAGGAATTGAAGTTGTTCACCGAAGCTATTGAAAGAT
GGGATGCCTCTTCTATAGATCAATTGCCAGATTACATGAAGGTCTGCTATCAAGCTTTGTTGGA
TGTCTACGAAGAAATGGAAGAAGAGATGACAAAACAGGGCAAGTTGTACAGAGTTCATTATGCT
CAAGCTGCCTTGAAGAGACAAGTTCAAGCTTATTTGTTGGAAGCCAAGTGGTTGAAGCAAGAGT
ATATTCCAACTATGGAAGAGTACATGTCCAACGCTTTGGTTACTTCTGCTTGTTCTATGTTGAC
TACCACTTCTTTTGTTGGTATGGGTGATATGGTTACCAAAGAAGCTTTTGACTGGGTTTTCTCT
GATCCAAAGATGATTAGAGCCTCTAACGTCATCTGTAGATTGATGGATGATATCGTGTCCCACG
AATTCGAACAAAAAAGAGGTCACGTTGCTTCAGCTGTTGAATGCTACATGAAGCAATACGGTGT
TTCAAAAGAAGAAGCCTACGACGAATTCAAGAAGCAAGTTGAATCTGCTTGGAAGGACAACAAT
GAAGAAGTTTTACAACCTACCGCTGTTCCAGTTCCATTATTGACTAGAGTCTTGAACTTCTCCA
GAATGGTTGACGTCTTGTATAAGGATGAAGATGAGTATACCTTGGTCGGTCCTTTAATGAAGGA
TTTGGTTGCTGGTATGTTGATTGATCCAGTGCCAATGTAACTAAGATCCGCTCTAAC

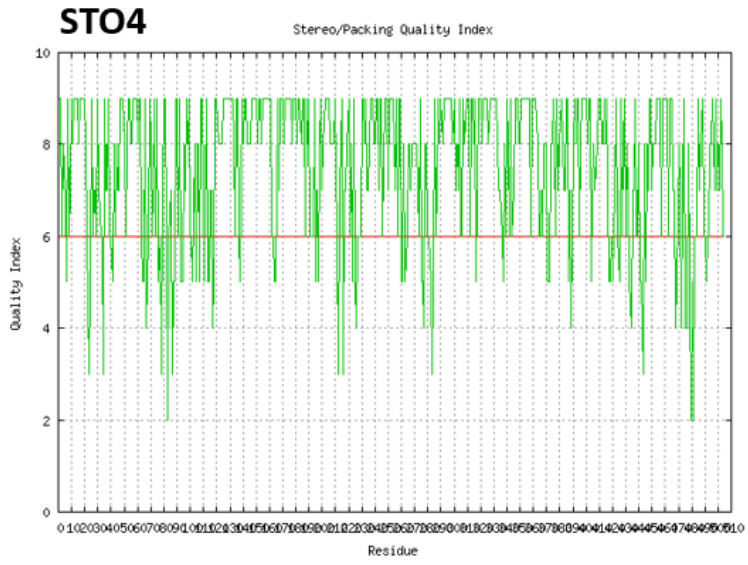


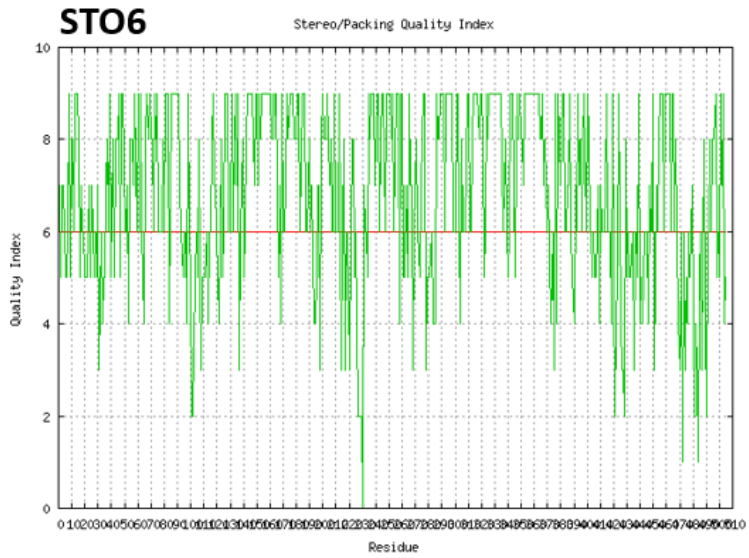




Supplementary Figure 5-1 3D Profile Quality Index of grapevine cytochrome P450 models. Quality indices were calculated by VADAR.







Supplementary Figure 5-2 Stereo/Packing Quality Index of grapevine cytochrome P450 models. Quality indices were calculated by VADAR.

Chapter 6 Thesis Summary

The focus of this work was to use microbial biosynthesis to investigate two classes of grapevine specialised metabolites: tartaric acid (TA) and its biosynthetic intermediates, and sesquiterpenoids, in particular the pepper aroma rotundone.

The first aim was to produce L-idonate, the substrate for the rate-limiting reaction in TA biosynthesis, *via* the microbial biotransformation of 2-keto-L-gulonate. Two potential *Vitis vinifera* 2-ketogulonate reductases (2KGRs) were identified using bioinformatics based on homology to two specific 2KGRs identified in *Aspergillus niger*. *In vitro* characterisation of these proteins showed they were capable of reducing 2-keto-L-gulonate, but no evidence was found for the formation of L-idonate. Substrate specificity tests suggest the enzymes possess broad-spectrum, low-level reductase activity, although their primary substrate may not have been tested. It is unlikely that these enzymes are involved in the biosynthesis of TA from ascorbic acid. To fully characterise these enzymes, their activity could be tested against a wider range of substrates, and reaction conditions, including temperature, enzyme concentration, substrate concentration and cofactor concentration, could be optimised. Enzyme activity may also be tested using $NADP^+$/NADPH as cofactor, and assays could be performed to investigate the affinity of each enzyme for each cofactor.

Since these enzymes are unlikely to play a role in the biosynthesis of TA, there is increased evidence for the involvement of the previously identified Vv2KGR in this pathway. However, *in planta* evidence is required to directly confirm the role of this enzyme in TA synthesis.

As the two putative 2KGRs identified did not produce high levels of L-idonate, instead the previously identified Vv2KGR was used to investigate the viability of a whole-cell biotransformation system for the production of L-idonate. Although the 2KGR activity of Vv2KGR has previously been confirmed *in vitro*, cell permeability proved a significant challenge, and *E. coli* cells expressing Vv2KGR were not able to convert 2-keto-L-gulonate to L-idonate in the whole-cell system. Future research may investigate other methods of increasing cell permeability, such as treatment with solvents or detergents, or bioengineering. Alternative host organisms may also be investigated, such as a gram-positive bacterium or eukaryotic organism. Alternatively, the bioconversion of 2-keto-L-gulonate may be achieved *in vitro* with purified soluble or resin-bound enzyme. However, this could be significantly expensive, as exogenous cofactor would need to be supplied, and Vv2KGR prefers NADPH over NADH (Jia 2015). Using the *A. niger* enzyme GluC, an

NADH-dependent 2KGR, may be a more economical option for cell-free biotransformation.

Chapters 4 and 5 describe experiments in which the biosynthesis and microbial synthesis of sesquiterpenoids were investigated. Six *V. vinifera* sesquiterpene synthases were expressed in EPY300, a yeast strain engineered to overproduce farnesyl pyrophosphate (FPP), the sesquiterpene precursor. Co-expression of these sesquiterpene synthases with a promiscuous cytochrome P450 monooxygenase in EPY300 enabled the microbial synthesis of complex sesquiterpenoids from a simple sugar feedstock. The six sesquiterpene synthases collectively produced a large number of structurally diverse sesquiterpenoids. A total of seventeen unique compounds were confidently identified. A number of uncharacterised compounds were also found in the cell extract samples. To identify these compounds, yeast cultures could be scaled up and the extracts purified by chromatography to allow the collection of several milligrams of purified product for NMR analysis. This could lead to the identification of novel sesquiterpenoids, which are not only relevant to winemaking but also to the flavour and aroma industries. A number of the identified compounds also possessed previously identified bioactivity, which may serve as a marketing platform for wines containing these compounds. This work may serve as a useful platform for the future identification of sesquiterpenoids in grapevine and their contributions to the flavour, aroma, and health benefits of wine.

Chapter 5 focussed on the sesquiterpenoid rotundone, which is responsible for the 'peppery' character of some wines (Wood et al. 2008). Rotundone is formed by oxidation of the sesquiterpene hydrocarbon α-guaiene, which is catalysed by Sesquiterpene Oxidase 2 (STO2) in *V. vinifera.* Two other, closely related P450s, STO4 and STO6 are also capable of utilising α-guaiene as substrate but do not produce rotundone (Takase et al. 2016). Computational homology modelling of STO2, STO4 and STO6 revealed that differences in the binding orientation of α-guaiene within the P450 active site are responsible for the differences in products formed by these three enzymes. It was hypothesised that rotundone is formed by STO2 due to the positioning of the C3 carbon of α-guaiene in close proximity to the STO2 haem. Three key residues within the STO2 binding site were identified that may affect product specificity, and by altering these residues it may be possible to produce novel sesquiterpenoids. Experimental site-specific mutation studies are needed to confirm these hypotheses. To do this, a yeast strain capable of producing α-guaiene was constructed. This strain can be used to express STO mutants, allowing *in vivo* production of sesquiterpenoids and rapid screening and analysis of mutant enzymes. New products may be isolated by extraction of large-scale culture followed by purification of

the extraction mixture. The products can then be identified by NMR spectroscopy. Mutagenesis studies will increase understanding of the mechanism and binding of these P450s in grapevine, and may assist in the discovery of novel aromas or flavours. Understanding the role of key residues within the P450 active site may facilitate the identification of novel P450 alleles within plant genomes and the discovery of unique metabolites.

Specialised metabolites play many roles that allow plants to gain evolutionary advantages, but many are also useful to humans. Plant specialised metabolites have been used by humans for millennia as fragrances, flavours and drugs. Specialised metabolites in grapevine are of particular importance to the wine industry, as the composition of specialised metabolites confer unique flavours and characteristics to the wine. Furthermore, there has been a recent growth in interest in the potential health benefits of consuming certain compounds in wine.

Microbial biosynthesis has been shown to be a useful method for the investigation of these metabolites as it allows facile, rapid and stereoselective synthesis of natural compounds. Although the biocatalysis of L-idonate requires further optimisation, the microbial biosynthesis of sesquiterpenoids using EPY300 is well-established and has facilitated the rapid identification of grapevine sesquiterpenoids. The investigation of these specialised metabolites has provided insights into the biosynthesis of tartaric acid and sesquiterpenoids in grapevine, and may assist in the identification and production of novel wine aromas and flavours, and the development of varietals selected for the presence of these important metabolites.

References

Abdel-Rahman, MA, Tashiro, Y, Zendo Hanada, K, Shibata, K & Sonomoto, K 2011, 'Efficient homofermentative L(+)-lactic acid production from xylose by a novel lactic acid bacterium, *Enterococcus mundtii* QU25', *Applied Environmental Microbiology*, vol. 77.

Acree, T & Arn, H 2004, 'Flavornet', DATU Inc., <http://www.flavornet.org>.

Ali, K, Maltese, F, Choi, YH & Verpoorte, R 2010, 'Metabolic constituents of grapevine and grape-derived products', *Phytochemistry Reviews*, vol. 9, no. 3, pp. 357-378.

Arimura, G-i, Ozawa, R, Shimoda, T, Nishioka, T, Boland, W & Takabayashi, J 2000, 'Herbivory-induced volatiles elicit defence genes in lima bean leaves', *Nature*, vol. 406, no. 6795, p. 512.

Babalola, OO, Anetor, JI & Adeniyi, FA 2001, 'Amelioration of carbon tetrachloride-induced hepatotoxicity by terpenoid extract from leaves of *Vernonia amydgalina*', *African journal of medicine and medical sciences*, vol. 30, no. 1-2, pp. 91-93.

Bahi, A, Al Mansouri, S, Al Memari, E, Al Ameri, M, Nurulain, SM & Ojha, S 2014, 'β-Caryophyllene, a CB_2 receptor agonist produces multiple behavioral changes relevant to anxiety and depression in mice', *Physiology & Behavior*, vol. 135, pp. 119-124.

Bausch, C, Peekhaus, N, Utz, C, Blais, T, Murray, E, Lowary, T & Conway, T 1998, 'Sequence analysis of the GntII (subsidiary) system for gluconate metabolism reveals a novel pathway for L-idonic acid catabolism in *Escherichia coli*', *Journal of Bacteriology*, vol. 180, no. 14, pp. 3704-3710.

Belitz, H-D, Grosch, W & Schieberle, P 2009, 'Food Additives', in *Food Chemistry*, 4 edn, Springer, Berlin, DOI 10.1007/978-3-540-69934-7, <https://link-springer-com.proxy.library.adelaide.edu.au/book/10.1007/978-3-540-69934-7/page/1>.

Bohlmann, J & Keeling, CI 2008, 'Terpenoid biomaterials', *The Plant Journal*, vol. 54, no. 4, pp. 656-669.

Bremus, C, Herrmann, U, Bringer-Meyer, S & Sahm, H 2006, 'The use of microorganisms in L-ascorbic acid production', *Journal of Biotechnology*, vol. 124, no. 1, pp. 196-205.

Buchholz, K & Seibel, J 2008, 'Industrial carbohydrate biotransformations', *Carbohydrate Research*, vol. 343, no. 12, pp. 1966-1979.

Bückle-Vallant, V, Krause, FS, Messerschmidt, S & Eikmanns, BJ 2014, 'Metabolic engineering of *Corynebacterium glutamicum* for 2-ketoisocaproate production', *Applied Microbiology and Biotechnology*, vol. 98, no. 1, pp. 297-311.

Bülow, N & König, WA 2000, 'The role of germacrene D as a precursor in sesquiterpene biosynthesis: investigations of acid catalyzed, photochemically and thermally induced rearrangements', *Phytochemistry*, vol. 55, no. 2, pp. 141-168.

Burbidge, C 2011, 'Identification and characterisation of the enzymes involved in the biosynthetic pathway of tartaric acid in *Vitis vinifera*', School of Biological Sciences, Doctor of Philosophy thesis, Flinders University, Adelaide.

Cankar, K, van Houwelingen, A, Bosch, D, Sonke, T, Bouwmeester, H & Beekwilder, J 2011, 'A chicory cytochrome P450 mono-oxygenase CYP71AV8 for the oxidation of (+)-valencene', *FEBS letters*, vol. 585, no. 1, pp. 178-182.

Cánovas, M, Torroglosa, T, Kleber, H-P & Iborra, JL 2003, 'Effect of salt stress on crotonobetaine and D(+)-carnitine biotransformation into L(−)-carnitine by resting cells of *Escherichia coli*', *Journal of Basic Microbiology*, vol. 43, no. 4, pp. 259-268.

Caputi, L, Carlin, S, Ghiglieno, I, Stefanini, M, Valenti, L, Vrhovsek, U & Mattivi, F 2011, 'Relationship of changes in rotundone content during grape Ripening and Winemaking to Manipulation of the 'Peppery' Character of Wine', *Journal of Agricultural and Food Chemistry*, vol. 59, no. 10, pp. 5565-5571.

Chang, S-T, Chen, P-F, Wang, S-Y & Wu, H-H 2001, 'Antimite Activity of Essential Oils and Their Constituents from *Taiwania cryptomerioides*', *Journal of Medical Entomology*, vol. 38, no. 3, pp. 455-457.

Chavan, MJ, Wakte, PS & Shinde, DB 2010, 'Analgesic and anti-inflammatory activity of Caryophyllene oxide from *Annona squamosa* L. bark', *Phytomedicine*, vol. 17, no. 2, pp. 149-151.

Chen, RR 2007, 'Permeability issues in whole-cell bioprocesses and cellular membrane engineering', *Applied Microbiology and Biotechnology*, vol. 74, no. 4, pp. 730-738.

Coelho, E, Rocha, SM, Delgadillo, I & Coimbra, MA 2006, 'Headspace-SPME applied to varietal volatile components evolution during *Vitis vinifera* L. cv. 'Baga' ripening', *Analytica Chimica Acta*, vol. 563, no. 1, pp. 204-214.

Connolly, JD & Hill, RA 1984, *Dictionary of Terpenoids*.

Coombe, BG 1960, 'Relationship of Growth and Development to Changes in Sugars, Auxins, and Gibberellins in Fruit of Seeded and Seedless Varieties of Vitis Vinifera', *Plant Physiology*, vol. 35, no. 2, pp. 241-250.

Cragg, GM, Newman, DJ & Snader, KM 1997, 'Natural Products in Drug Discovery and Development', *Journal of Natural Products*, vol. 60, no. 1, pp. 52-60.

Davey, MW, Montagu, MV, Inzé, D, Sanmartin, M, Kanellis, A, Smirnoff, N, Benzie, IJJ, Strain, JJ, Favell, D & Fletcher, J 2000, 'Plant L-ascorbic acid: chemistry, function, metabolism,

bioavailability and effects of processing', *Journal of the Science of Food and Agriculture*, vol. 80, no. 7, pp. 825-860.

Davies, C, Nicholson, EL, Böttcher, C, Burbidge, CA, Bastian, SEP, Harvey, KE, Huang, AC, Taylor, DK & Boss, PK 2015, 'Shiraz Wines Made from Grape Berries (*Vitis vinifera*) Delayed in Ripening by Plant Growth Regulator Treatment Have Elevated Rotundone Concentrations and "Pepper" Flavor and Aroma', *Journal of Agricultural and Food Chemistry*, vol. 63, no. 8, pp. 2137-2144.

Davis, GD, Elisee, C, Newham, DM & Harrison, RG 1999, 'New fusion protein systems designed to give soluble expression in *Escherichia coli*', *Biotechnology and Bioengineering*, vol. 65, no. 4, pp. 382-388.

DeBolt, S, Cook, DR & Ford, CM 2006, 'L-tartaric acid synthesis from vitamin C in higher plants', *Proceedings of the National Academy of Science USA*, vol. 103, no. 14, pp. 5608-5613.

DeBolt, S, Hardie, JIM, Tyerman, S & Ford, CM 2004, 'Composition and synthesis of raphide crystals and druse crystals in berries of *Vitis vinifera* L. cv. Cabernet Sauvignon: Ascorbic acid as precursor for both oxalic and tartaric acids as revealed by radiolabelling studies', *Australian Journal of Grape and Wine Research*, vol. 10, no. 2, pp. 134-142.

Diakou, P, Moing, A, Svanella, L, Ollat, N, Rolin, DB, Gaudillere, M & Gaudillere, JP 1997, 'Biochemical comparison of two grape varieties differing in juice acidity', *Australian Journal of Grape and Wine Research*, vol. 3, no. 3, pp. 1-10.

Dixon, RA & Strack, D 2003, 'Phytochemistry meets genome analysis, and beyond', *Phytochemistry*, vol. 62, no. 6, pp. 815-816.

Drew, DP, Andersen, TB, Sweetman, C, Møller, BL, Ford, C & Simonsen, HT 2016, 'Two key polymorphisms in a newly discovered allele of the *Vitis vinifera* TPS24 gene are responsible for the production of the rotundone precursor α-guaiene', *Journal of Experimental Botany*, vol. 67, no. 3, pp. 799-808.

Du, J, Shao, Z & Zhao, H 2011, 'Engineering microbial factories for synthesis of value-added products', *Journal of Industrial Microbiology & Biotechnology*, vol. 38, no. 8, pp. 873-890.

Dudareva, N & Pichersky, E 2000, 'Biochemical and molecular genetic aspects of floral scents', *Plant Physiology*, vol. 122, no. 3, pp. 627-634.

Dudareva, N & Pichersky, E 2006, 'Floral Scent Metabolic Pathways: Their Regulation and Evolution', *Biology of Floral Scent*, pp. 55-78.

Duetz, WA, Van Beilen, JB & Witholt, B 2001, 'Using proteins in their natural environment: potential and limitations of microbial whole-cell hydroxylations in applied biocatalysis', *Current Opinion in Biotechnology*, vol. 12, no. 4, pp. 419-425.

Duhamel, N, Slaghenaufi, D, Pilkington, LI, Herbst-Johnstone, M, Larcher, R, Barker, D & Fedrizzi, B 2018, 'Facile gas chromatography–tandem mass spectrometry stable isotope dilution method for the quantification of sesquiterpenes in grape', *Journal of Chromatography A*, vol. 1537, pp. 91-98.

Eichler, A, Gricman, Ł, Herter, S, Kelly, PP, Turner, NJ, Pleiss, J & Flitsch, SL 2016, 'Enantioselective Benzylic Hydroxylation Catalysed by P450 Monooxygenases: Characterisation of a P450cam Mutant Library and Molecular Modelling', *ChemBioChem*, vol. 17, no. 5, pp. 426-432.

Eisenreich, W, Bacher, A, Arigoni, D & Rohdich, F 2004, 'Biosynthesis of isoprenoids via the non-mevalonate pathway', *Cellular and Molecular Life Sciences CMLS*, vol. 61, no. 12, pp. 1401-1426.

Emanuelsson, O, Brunak, S, von Heijne, G & Nielsen, H 2007, 'Locating proteins in the cell using TargetP, SignalP and related tools', *Nat. Protocols*, vol. 2, no. 4, pp. 953-971.

Esposito, D & Chatterjee, DK 2006, 'Enhancement of soluble protein expression through the use of fusion tags', *Current Opinion in Biotechnology*, vol. 17, no. 4, pp. 353-358.

Faber, K 2004, 'Biotransformations in organic chemistry: a textbook', Berlin: Springer.

Faraldos, JA, Wu, S, Chappell, J & Coates, RM 2010, 'Doubly Deuterium-Labeled Patchouli Alcohol from Cyclization of Singly Labeled [2-$^{2}H_{1}$]Farnesyl Diphosphate Catalyzed by Recombinant Patchoulol Synthase', *Journal of the American Chemical Society*, vol. 132, no. 9, pp. 2998-3008.

Fernandes, ES, Passos, GF, Medeiros, R, da Cunha, FM, Ferreira, J, Campos, MM, Pianowski, LF & Calixto, JB 2007, 'Anti-inflammatory effects of compounds alpha-humulene and (−)-trans-caryophyllene isolated from the essential oil of *Cordia verbenacea*', *European Journal of Pharmacology*, vol. 569, no. 3, pp. 228-236.

Fonseca, A 1992, 'Utilization of tartaric acid and related compounds by yeasts: taxonomic implications', *Canadian Journal of Microbiology*, vol. 38, no. 12, pp. 1242-1251.

Forrer, M, Kulik, EM, Filippi, A & Waltimo, T 2013, 'The antimicrobial activity of alpha-bisabolol and tea tree oil against *Solobacterium moorei*, a Gram-positive bacterium associated with halitosis', *Archives of Oral Biology*, vol. 58, no. 1, pp. 10-16.

Geertz-Hansen, HM, Blom, N, Feist, AM, Brunak, S & Petersen, TN 2014, 'Cofactory: sequence-based prediction of cofactor specificity of Rossmann folds', *Proteins*, vol. 82, no. 9, pp. 1819-1828.

Ghelardini, C, Galeotti, N, Di Cesare Mannelli, L, Mazzanti, G & Bartolini, A 2001, 'Local anaesthetic activity of β-caryophyllene', *Il Farmaco*, vol. 56, no. 5, pp. 387-389.

Gietz, RD & Schiestl, RH 2007, 'High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method', *Nature protocols*, vol. 2, no. 1, p. 35.

Govindaraj, S & Poulos, TL 1995, 'Role of the Linker Region Connecting the Reductase and Heme Domains in Cytochrome $P450_{BM-3}$', *Biochemistry*, vol. 34, no. 35, pp. 11221-11226.

Grindley, JF, Payton, MA, Van de Pol, H & Hardy, KG 1988, 'Conversion of glucose to 2-keto-L-gulonate, an intermediate in L-ascorbate synthesis, by a recombinant strain of *Erwinia citreus*', *Applied and Environmental Microbiology*, vol. 54, no. 7, pp. 1770-1775.

Gutnick, D, Calvo, JM, Klopotowski, T & Ames, BN 1969, 'Compounds Which Serve as the Sole Source of Carbon or Nitrogen for *Salmonella typhimurium* LT-2', *Journal of Bacteriology*, vol. 100, no. 1, pp. 215-219.

Hale, CR 1962, 'Synthesis of Organic Acids in the Fruit of the Grape', *Nature*, vol. 195, no. 4844, pp. 917-918.

Harrison, R 2000, 'Expression of soluble heterologous proteins via fusion with NusA protein', *inNovations*, vol. 11, pp. 4-7.

Heiling, S, Schuman, MC, Schoettner, M, Mukerjee, P, Berger, B, Schneider, B, Jassbi, AR & Baldwin, IT 2010, 'Jasmonate and ppHsystemin regulate key malonylation steps in the biosynthesis of 17-hydroxygeranyllinalool diterpene glycosides, an abundant and effective direct defense against herbivores in *Nicotiana attenuata*', *The Plant cell*, vol. 22, no. 1, pp. 273-292.

Huang, AC, Burrett, S, Sefton, MA & Taylor, DK 2014, 'Production of the pepper aroma compound, (-)-rotundone, by aerial oxidation of alpha-guaiene', *Journal of Agricultural and Food Chemistry*, vol. 62, no. 44, pp. 10809-10815.

Huang, AC, Sefton, MA, Sumby, CJ, Tiekink, ER & Taylor, DK 2015, 'Mechanistic Studies on the Autoxidation of α-Guaiene: Structural Diversity of the Sesquiterpenoid Downstream Products', *Journal of Natural Products*, vol. 78, no. 1, pp. 131-145.

Hunter, GLK & Brogden JR., WB 1965, 'Conversion of Valencene to Nootkatone', *Journal of Food Science*, vol. 30, no. 5, pp. 876-878.

Iland, PG & Coombe, BG 1988, 'Malate, Tartrate, Potassium, and Sodium in Flesh and Skin of Shiraz Grapes During Ripening: Concentration and Compartmentation', *American Journal of Enology and Viticulture*, vol. 39, no. 1, pp. 71-76.

Jeandet, P, Vasserot, Y, Chastang, T & Courot, E 2013, 'Engineering Microbial Cells for the Biosynthesis of Natural Compounds of Pharmaceutical Significance', *BioMed Research International*, vol. 2013, p. 13.

Jia, Y 2015, 'Molecular and structural characterization of the candidate enzymes responsible for tartaric acid synthesis in the grapevine', School of Agriculture, Food and Wine, Doctor of Philosophy thesis, The University of Adelaide, Adelaide.

Jia, Y, Wong, DCJ, Sweetman, C, Bruning, JB & Ford, CM 2015, 'New insights into the evolutionary history of plant sorbitol dehydrogenase', *BMC Plant Biology*, vol. 15, no. 1.

Kahn, RA, Le Bouquin, R, Pinot, F, Benveniste, I & Durst, F 2001, 'A conservative amino acid substitution alters the regiospecificity of CYP94A2, a fatty acid hydroxylase from the plant Vicia sativa', *Archives of Biochemistry and Biophysics*, vol. 391, no. 2, pp. 180-187.

Kalathenos, P, Sutherland, JP & Roberts, TA 1995, 'Resistance of some wine spoilage yeasts to combinations of ethanol and acids present in wine', *Journal of Applied Bacteriology*, vol. 78, no. 3, pp. 245-250.

Kapadia, VH, Naik, VG, Wadia, MS & Dev, S 1967, 'Sesquiterpenoids from the essential oil of *cyperus rotundus*', *Tetrahedron Letters*, vol. 8, no. 47, pp. 4661-4667.

Kara, S, Schrittwieser, JH & Hollmann, F 2013, 'Strategies for Cofactor Regeneration in Biocatalyzed Reductions', in *Synthetic Methods for Biologically Active Molecules*, Wiley-VCH Verlag GmbH & Co. KGaA, pp. 209-238.

Kaswurm, V, van Hecke, W, Kulbe, KD & Ludwig, R 2013, 'Engineering of a bi-enzymatic reaction for efficient production of the ascorbic acid precursor 2-keto-L-gulonic acid', *Biochemical Engineering Journal*, vol. 79, pp. 104-111.

Klasen, R, Bringer-Meyer, S & Sahm, H 1992, 'Incapability of Gluconobacter oxydans to produce tartaric acid', *Biotechnology and Bioengineering*, vol. 40, no. 1, pp. 183-186.

Klauke, AL, Racz, I, Pradier, B, Markert, A, Zimmer, AM, Gertsch, J & Zimmer, A 2014, 'The cannabinoid CB_2 receptor-selective phytocannabinoid beta-caryophyllene exerts analgesic effects in mouse models of inflammatory and neuropathic pain', *European Neuropsychopharmacology*, vol. 24, no. 4, pp. 608-620.

Kliewer, WM, Howarth, L & Omori, M 1967, 'Concentrations of Tartaric Acid and Malic Acids and Their Salts in *Vitis vinifera* Grapes', *American Journal of Enology and Viticulture*, vol. 18, no. 1, p. 42.

Komori, A, Suzuki, M, Seki, H, Nishizawa, T, Meyer, JJM, Shimizu, H, Yokoyama, S & Muranaka, T 2013, 'Comparative functional analysis of CYP71AV1 natural variants reveals an important residue for the successive oxidation of amorpha-4, 11-diene', *FEBS letters*, vol. 587, no. 3, pp. 278-284.

Kuivanen, J, Arvas, M & Richard, P 2017, 'Clustered Genes Encoding 2-Keto-L-Gulonate Reductase and L-Idonate 5-Dehydrogenase in the Novel Fungal D-Glucuronic Acid Pathway', *Frontiers in Microbiology*, vol. 8, no. 225.

Kuivanen, J, Sugai-Guerios, MH, Arvas, M & Richard, P 2016, 'A novel pathway for fungal D-glucuronate catabolism contains an L-idonate forming 2-keto-L-gulonate reductase', *Science Reports*, vol. 6, p. 26329.

Kumeta, Y & Ito, M 2010, 'Characterization of δ-Guaiene Synthases from Cultured Cells of *Aquilaria*, Responsible for the Formation of the Sesquiterpenes in Agarwood', *Plant Physiology*, vol. 154, no. 4, pp. 1998-2007.

Legault, J, Dahl, W, Debiton, E, Pichette, A & Madelmont, J-C 2003, 'Antitumor Activity of Balsam Fir Oil: Production of Reactive Oxygen Species Induced by α-Humulene as Possible Mechanism of Action', *Planta Med*, vol. 69, no. 05, pp. 402-407.

Leister, D 2003, 'Chloroplast research in the genomic age', *TRENDS in Genetics*, vol. 19, no. 1, pp. 47-56.

Lesburg, CA, Zhai, G, Cane, DE & Christianson, DW 1997, 'Crystal Structure of Pentalenene Synthase: Mechanistic Insights on Terpenoid Cyclization Reactions in Biology', *Science*, vol. 277, no. 5333, pp. 1820-1824.

Li, H & Poulos, TL 1997, 'The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid', *Nature Structural Biology*, vol. 4, p. 140.

Lilie, H, Schwarz, E & Rudolph, R 1998, 'Advances in refolding of proteins produced in E. coli', *Current Opinion in Biotechnology*, vol. 9, no. 5, pp. 497-501.

Liu, SQ 2002, 'Malolactic fermentation in wine – beyond deacidification', *Journal of Applied Microbiology*, vol. 92, no. 4, pp. 589-601.

Loewus, FA 1999, 'Biosynthesis and metabolism of ascorbic acid in plants and of analogs of ascorbic acid in fungi', *Phytochemistry*, vol. 52, no. 2, pp. 193-210.

Loewus, FA & Stafford, HA 1958, 'Observations on the incorporation of C14 into tartaric acid and the labelling pattern of D-glucose from an excised grape leaf administered L-ascorbic acid-6-C14', *Plant Physiology*, vol. 33, no. 2, pp. 155-156.

López-Gallego, F, Wawrzyn, G & Schmidt-Dannert, C 2010, 'Selectivity of Fungal Sesquiterpene Synthases: Role of the Active Site's H-1α Loop in Catalysis', *Applied and Environmental Microbiology*, vol. 76, no. 23, pp. 7723-7733.

Lorenz, DH, Eichhorn, KW, Bleiholder, H, Klose, R, Meier, U & Weber, E 1995, 'Growth Stages of the Grapevine: Phenological growth stages of the grapevine (*Vitis vinifera* L. ssp. vinifera)—Codes and descriptions according to the extended BBCH scale', *Australian Journal of Grape and Wine Research*, vol. 1, no. 2, pp. 100-103.

Lücker, J, Bowen, P & Bohlmann, J 2004, '*Vitis vinifera* terpenoid cyclases: functional identification of two sesquiterpene synthase cDNAs encoding (+)-valencene synthase and (−)-germacrene D synthase and expression of mono- and sesquiterpene synthases in grapevine flowers and berries', *Phytochemistry*, vol. 65, no. 19, pp. 2649-2659.

Luirink, RA, Dekker, SJ, Capoferri, L, Janssen, LFH, Kuiper, CL, Ari, ME, Vermeulen, NPE, Vos, JC, Commandeur, JNM & Geerke, DP 2018, 'A combined computational and

experimental study on selective flucloxacillin hydroxylation by cytochrome P450 BM3 variants', *Journal of Inorganic Biochemistry*, vol. 184, pp. 115-122.

Malipiero, U, Ruffner, HP & Rast, DM 1987, 'Ascorbic to tartaric acid conversion in grapevines', *Journal of Plant Physiology*, vol. 129, no. 1, pp. 33-40.

Manikandan, P & Nagini, S 2018, 'Cytochrome P450 Structure, Function and Clinical Significance: A Review', *Current Drug Targets*, vol. 19, no. 1, pp. 38-54.

Markham, KA & Alper, HS 2015, 'Synthetic Biology for Specialty Chemicals', *Annual Reviews of Chemical and Biomolecular Engineering*, vol. 6, pp. 35-52.

Martin, DM, Aubourg, S, Schouwey, MB, Daviet, L, Schalk, M, Toub, O, Lund, ST & Bohlmann, J 2010, 'Functional Annotation, Genome Organization and Phylogeny of the Grapevine (Vitis vinifera) Terpene Synthase Gene Family Based on Genome Assembly, FLcDNA Cloning, and Enzyme Assays', *BMC Plant Biology*, vol. 10, no. 1, p. 226.

Martin, DM, Chiang, A, Lund, ST & Bohlmann, J 2012, 'Biosynthesis of wine aroma: transcript profiles of hydroxymethylbutenyl diphosphate reductase, geranyl diphosphate synthase, and linalool/nerolidol synthase parallel monoterpenol glycoside accumulation in Gewürztraminer grapes', *Planta*, vol. 236, no. 3, pp. 919-929.

Matarese, F, Cuzzola, A, Scalabrelli, G & D'Onofrio, C 2014, 'Expression of terpene synthase genes associated with the formation of volatiles in different organs of *Vitis vinifera*', *Phytochemistry*, vol. 105, pp. 12-24.

Matzerath, I, Kläui, W, Klasen, R & Sahm, H 1995, 'Vanadate catalysed oxidation of 5-keto-D-gluconic acid to tartaric acid: the unexpected effect of phosphate and carbonate on rate and selectivity', *Inorganica Chimica Acta*, vol. 237, no. 1, pp. 203-205.

May, B, Lange, BM & Wüst, M 2013, 'Biosynthesis of sesquiterpenes in grape berry exocarp of *Vitis vinifera* L.: Evidence for a transport of farnesyl diphosphate precursors from plastids to the cytosol', *Phytochemistry*, vol. 95, pp. 135-144.

Mayr, CM, Geue, JP, Holt, HE, Pearson, WP, Jeffery, DW & Francis, IL 2014, 'Characterization of the Key Aroma Compounds in Shiraz Wine by Quantitation, Aroma Reconstitution, and Omission Studies', *Journal of Agricultural and Food Chemistry*, vol. 62, no. 20, pp. 4528-4536.

Mazumdar, S, Clomburg, JM & Gonzalez, R 2010, 'Escherichia coli strains engineered for homofermentative production of D-lactic acid from glycerol', *Applied Environmental Microbiology*, vol. 76.

Mendes, FB, Bergamin, LS, Dos Santos Stuepp, C, Braganhol, E, Terroso, T, Pohlmann, AR, Guterres, SS & Battastini, AMO 2017, 'Alpha-bisabolol Promotes Glioma Cell Death by Modulating the Adenosinergic System', *Anticancer Research*, vol. 37, no. 4, pp. 1819-1823.

Morris, GM, Huey, R, Lindstrom, W, Sanner, MF, Belew, RK, Goodsell, DS & Olson, AJ 2009, 'AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility', *Journal of computational chemistry*, vol. 30, no. 16, pp. 2785-2791.

Moskowitz, AH & Hrazdina, G 1981, 'Vacuolar Contents of Fruit Subepidermal Cells from Vitis Species', *Plant Physiology*, vol. 68, no. 3, pp. 686-692.

Neidhardt, FC, Ingraham, JL & Schaechter, M 1990, *Physiology of the bacterial cell: a molecular approach*, vol. 20, Sinauer Associates Sunderland, MA.

Nguyen, T-D, MacNevin, G & Ro, D-K 2012, 'De Novo Synthesis of High-Value Plant Sesquiterpenoids in Yeast', *Methods in Enzymology*, vol. 517, pp. 261-278.

Nikaido, H 2001, 'Preventing drug access to targets: cell surface permeability barriers and active efflux in bacteria', in *Seminars in Cell & Developmental Biology,* Elsevier, vol. 12, pp. 215-223.

Nikolaus, J, Nguyen, KT, Virus, C, Riehm, JL, Hutter, M & Bernhardt, R 2017, 'Engineering of CYP106A2 for steroid 9α- and 6β-hydroxylation', *Steroids*, vol. 120, pp. 41-48.

NIST Standard Reference Database 1A User's Guide 2008, The NIST Mass Spectrometry Data Center Gaithersburg, MD.

O'Maille, PE, Malone, A, Dellas, N, Hess Jr, BA, Smentek, L, Sheehan, I, Greenhagen, BT, Chappell, J, Manning, G & Noel, JP 2008, 'Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases', *Nature chemical biology*, vol. 4, no. 10, p. 617.

Ollat, N, Carde, J-P, Gaudillère, J-P, Barrieu, F, Diakou-Verdin, P & Moing, A 2002, 'Grape berry development : A review', *Journal international des sciences de la vigne et du vin*, vol. 36, no. 3, p. 23.

Palumbo, MC, Zenoni, S, Fasoli, M, Massonnet, M, Farina, L, Castiglione, F, Pezzotti, M & Paci, P 2014, 'Integrated network analysis identifies fight-club nodes as a class of hubs encompassing key putative switch genes that induce major transcriptome reprogramming during grapevine development', *The Plant cell*, vol. 26, no. 12, pp. 4617-4635.

Parker, M, Pollnitz, AP, Cozzolino, D, Francis, IL & Herderich, MJ 2007, 'Identification and Quantification of a Marker Compound for 'Pepper' Aroma and Flavor in Shiraz Grape Berries by Combination of Chemometrics and Gas Chromatography–Mass Spectrometry', *Journal of Agricultural and Food Chemistry*, vol. 55, no. 15, pp. 5948-5955.

Pearson, WR 2013, 'An Introduction to Sequence Similarity ("Homology") Searching', *Current protocols in bioinformatics*, vol. Chapter 3.

Petryszak, R, Keays, M, Tang, YA, Fonseca, NA, Barrera, E, Burdett, T, Füllgrabe, A, Fuentes, AM-P, Jupp, S, Koskinen, S, Mannion, O, Huerta, L, Megy, K, Snow, C, Williams, E, Barzine, M, Hastings, E, Weisser, H, Wright, J, Jaiswal, P, Huber, W, Choudhary, J, Parkinson, HE &

Brazma, A 2016, 'Expression Atlas update—an integrated database of gene and protein expression in humans, animals and plants', *Nucleic Acids Research*, vol. 44, no. D1, pp. D746-D752.

Pichersky, E & Lewinsohn, E 2011, 'Convergent Evolution in Plant Specialized Metabolism', in SS Merchant, WR Briggs & D Ort (eds), *Annual Review of Plant Biology*, vol. 62, pp. 549-566.

Plane, RA, Mattick, LR & Weirs, LD 1980, 'An Acidity Index for the Taste of Wines', *American Journal of Enology and Viticulture*, vol. 31, no. 3, pp. 265-268.

Poulos, TL & Johnson, EF 2015, 'Structures of Cytochrome P450 Enzymes', in PR Ortiz de Montellano (ed.), *Cytochrome P450: Structure, Mechanism, and Biochemistry*, Springer International Publishing, Cham, pp. 3-32.

Rapp, A & Mandery, H 1986, 'Wine aroma', *Experientia*, vol. 42, no. 8, pp. 873-884.

Ren, C, Liu, X, Zhang, Z, Wang, Y, Duan, W, Li, S & Liang, Z 2016, 'CRISPR/Cas9-mediated efficient targeted mutagenesis in Chardonnay (*Vitis vinifera* L.)', *Scientific Reports*, vol. 6, p. 32289.

Ro, D-K, Paradise, EM, Ouellet, M, Fisher, KJ, Newman, KL, Ndungu, JM, Ho, KA, Eachus, RA, Ham, TS, Kirby, J, Chang, MCY, Withers, ST, Shiba, Y, Sarpong, R & Keasling, JD 2006, 'Production of the antimalarial drug precursor artemisinic acid in engineered yeast', *Nature*, vol. 440, no. 7086, pp. 940-943.

Rocha, S, Coelho, E, Vinholes, J & Coimbra, M 2006, 'Grapes and wine from *Vitis vinifera* L. as a potential source of sesquiterpenoids', *Natural Products, Series Recent Progress in Medicinal Plants*, vol. 15, pp. 253-272.

Rodriguez, S, Kirby, J, Denby, CM & Keasling, JD 2014, 'Production and quantification of sesquiterpenes in *Saccharomyces cerevisiae*, including extraction, detection and quantification of terpene products and key related metabolites', *Nature protocols*, vol. 9, no. 8, pp. 1980-1996.

Ruffner, HP 1982, 'Metabolism of tartaric and malic acids in Vitis: A review- Part A', *Vitis*, vol. 21, no. 3, pp. 247-259.

Saito, K 1992, 'Metabolism of L-threotetruronic acid by Pelargonium crispum', *Phytochemistry*, vol. 31, no. 4, pp. 1219-1222.

Saito, K & Kasai, Z 1968, 'Accumulation of Tartaric Acid in the Ripening Process of Grapes', *Plant and Cell Physiology*, vol. 9, no. 3, pp. 529-537.

Saito, K & Kasai, Z 1969, 'Tartaric acid synthesis from L-ascorbic acid-1-14C in grape berries', *Phytochemistry*, vol. 8, no. 11, pp. 2177-2182.

Saito, K & Kasai, Z 1982, 'Conversion of L-Ascorbic Acid to L-Idonic Acid, L-Idono-γ-lactone ane 2-Keto-L-idonic Acid in Slices of Immature Grapes', *Plant and Cell Physiology*, vol. 23, no. 3, pp. 499-507.

Saito, K & Kasai, Z 1984, 'Synthesis of L-(+)-Tartaric Acid from L-Ascorbic Acid via 5-Keto-D-Gluconic Acid in Grapes', *Plant Physiology*, vol. 76, no. 1, pp. 170-174.

Saito, K & Loewus, FA 1979, 'The metabolism of L-[6-^{14}C]ascorbic acid in detached grape leaves', *Plant and Cell Physiology*, vol. 20, no. 8, pp. 1481-1488.

Saito, K & Loewus, FA 1989, 'Formation of tartaric acid in vitaceous plants - relative contributions of L-ascorbic acid-inclusive and acid-noninclusive pathways', *Plant and Cell Physiology*, vol. 30, no. 6, pp. 905-910.

Saito, K, Ohmoto, J & Kuriha, N 1997, 'Incorporation of ^{18}O into oxalic, L-threonic and L-tartaric acids during cleavage of L-ascorbic and 5-keto-D-gluconic acids in plants', *Phytochemistry*, vol. 44, no. 5, pp. 805-809.

Sali, A & Blundell, TL 1993, 'Comparative protein modelling by satisfaction of spatial restraints', *J Mol Biol*, vol. 234, no. 3, pp. 779-815.

Salusjärvi, T, Povelainen, M, Hvorslev, N, Eneyskaya, EV, Kulminskaya, AA, Shabalin, KA, Neustroev, KN, Kalkkinen, N & Miasnikov, AN 2004, 'Cloning of a gluconate/polyol dehydrogenase gene from Gluconobacter suboxydans IFO 12528, characterisation of the enzyme and its use for the production of 5-ketogluconate in a recombinant Escherichia coli strain', *Applied Microbiology and Biotechnology*, vol. 65, no. 3, pp. 306-314.

Salvagnin, U, Carlin, S, Angeli, S, Vrhovsek, U, Anfora, G, Malnoy, M & Martens, S 2016, 'Homologous and heterologous expression of grapevine E-(*β*)-caryophyllene synthase (VvGwECar2)', *Phytochemistry*, vol. 131, pp. 76-83.

Scarlett, NJ, Bramley, RGV & Siebert, TE 2014, 'Within-vineyard variation in the 'pepper' compound rotundone is spatially structured and related to variation in the land underlying the vineyard', *Australian Journal of Grape and Wine Research*, vol. 20, no. 2, pp. 214-222.

Schmidt, T, Bergner, A & Schwede, T 2014, 'Modelling three-dimensional protein structures for applications in drug design', *Drug Discovery Today*, vol. 19, no. 7, pp. 890-897.

Schreier, P, Drawert, F & Junker, A 1976, 'Identification of volatile constituents from grapes', *Journal of Agricultural and Food Chemistry*, vol. 24, no. 2, pp. 331-336.

Song, Y, Li, J, Shin, H-d, Liu, L, Du, G & Chen, J 2016, 'Biotechnological production of alpha-keto acids: Current status and perspectives', *Bioresource Technology*, vol. 219, pp. 716-724.

Stafford, HA 1959, 'Distribution of Tartaric Acid in the Leaves of Certain Angiosperms', *American Journal of Botany*, vol. 46, no. 5, pp. 347-352.

Stein, SE 1994, 'Estimating probabilities of correct identification from results of mass spectral library searches', *Journal of the American Society for Mass Spectrometry*, vol. 5, no. 4, pp. 316-323.

Suresh Kumar, P, Thomas, J & Poornima, V 2018, 'Structural insights on bioremediation of polycyclic aromatic hydrocarbons using microalgae: a modelling-based computational study', *Environmental Monitoring and Assessment*, vol. 190, no. 2, p. 92.

Suzuki, Y, Shimada, M, Tadera, K, Kawai, F & Mitsuda, H 1970, 'Studies on a New p-Coumaryl Derivative Isolated from Spinach Leaves', *Agricultural and Biological Chemistry*, vol. 34, no. 4, pp. 511-522.

Sweetman, C, Deluc, LG, Cramer, GR, Ford, CM & Soole, KL 2009, 'Regulation of malate metabolism in grape berry and other developing fruits', *Phytochemistry*, vol. 70, no. 11, pp. 1329-1344.

Sweetman, C, Wong, DC, Ford, CM & Drew, DP 2012, 'Transcriptome analysis at four developmental stages of grape berry (*Vitis vinifera* cv. Shiraz) provides insights into regulated and coordinated gene expression', *BMC Genomics*, vol. 13, no. 1, p. 691.

Takahashi, S, Zhao, Y, O'Maille, PE, Greenhagen, BT, Noel, JP, Coates, RM & Chappell, J 2005, 'Kinetic and molecular analysis of 5-epiaristolochene 1, 3-dihydroxylase, a cytochrome P450 enzyme catalyzing successive hydroxylations of sesquiterpenes', *Journal of Biological Chemistry*, vol. 280, no. 5, pp. 3686-3696.

Takase, H, Sasaki, K, Shinmori, H, Shinohara, A, Mochizuki, C, Kobayashi, H, Ikoma, G, Saito, H, Matsuo, H, Suzuki, S & Takata, R 2016, 'Cytochrome P450 CYP71BE5 in grapevine (*Vitis vinifera*) catalyzes the formation of the spicy aroma compound (−)-rotundone', *Journal of Experimental Botany*, vol. 67, no. 3, pp. 787-798.

Tasin, M, Anfora, G, Ioriatti, C, Carlin, S, De Cristofaro, A, Schmidt, S, Bengtsson, M, Versini, G & Witzgall, P 2005, 'Antennal and behavioral responses of grapevine moth *Lobesia botrana* females to volatiles from grapevine', *Journal of chemical ecology*, vol. 31, no. 1, pp. 77-87.

Terrier, N, Deguilloux, C, Sauvage, F-X, Martinoia, E & Romieu, C 1998, 'Proton pumps and anion transport in *Vitis vinifera*: The inorganic pyrophosphatase plays a predominant role in the energization of the tonoplast', *Plant Physiology and Biochemistry*, vol. 36, no. 5, pp. 367-377.

Tufvesson, P, Lima-Ramos, J, Nordblad, M & Woodley, JM 2011, 'Guidelines and Cost Analysis for Catalyst Production in Biocatalytic Processes', *Organic Process Research & Development*, vol. 15, no. 1, pp. 266-274.

Villegas, LF, Marçalo, A, Martin, J, Fernández, ID, Maldonado, H, Vaisberg, AJ & Hammond, GB 2001, '(+)-epi-α-Bisbolol Is the Wound-Healing Principle of *Peperomia galioides*: Investigation of the *in Vivo* Wound-Healing Activity of Related Terpenoids', *Journal of Natural Products*, vol. 64, no. 10, pp. 1357-1359.

Vinholes, JR 2013, 'Volatile terpenoids and C13 norisoprenoids in wines: development of rapid methods of analysis and evaluation of the sesquiterpenoids biological potential', Chemistry Department, PhD thesis, Universidade de Aveiro, <http://hdl.handle.net/10773/12029>.

Wachtmeister, J & Rother, D 2016, 'Recent advances in whole cell biocatalysis techniques bridging from investigative to industrial scale', *Current Opinion in Biotechnology*, vol. 42, pp. 169-177.

Wagner, G, Yang, JC & Loewus, FA 1975, 'Stereoisomeric Characterization of Tartaric Acid Produced during L-Ascorbic Acid Metabolism in Plants', *Plant Physiology*, vol. 55, no. 6, pp. 1071-1073.

Weeks, AM & Chang, MCY 2011, 'Constructing de Novo Biosynthetic Pathways for Chemical Synthesis inside Living Cells', *Biochemistry*, vol. 50, no. 24, pp. 5404-5418.

Wen, Y-Q, Li, J-M, Zhang, Z-Z, Zhang, Y-F & Pan, Q-H 2010, 'Antibody Preparation, Gene Expression and Subcellular Localization of L-Idonate Dehydrogenase in Grape Berry', *Bioscience, Biotechnology, and Biochemistry*, vol. 74, no. 12, pp. 2413-2417.

Willard, L, Ranjan, A, Zhang, H, Monzavi, H, Boyko, RF, Sykes, BD & Wishart, DS 2003, 'VADAR: a web server for quantitative evaluation of protein structure quality', *Nucleic Acids Research*, vol. 31, no. 13, pp. 3316-3319.

Williams, M & Loewus, FA 1978, 'Biosynthesis of (+)-Tartaric Acid from l-[4-14C]Ascorbic Acid in Grape and Geranium', *Plant Physiology*, vol. 61, no. 4, pp. 672-674.

Wood, C, Siebert, TE, Parker, M, Capone, DL, Elsey, GM, Pollnitz, AP, Eggers, M, Meier, M, Vössing, T, Widder, S, Krammer, G, Sefton, MA & Herderich, MJ 2008, 'From Wine to Pepper: Rotundone, an Obscure Sesquiterpene, Is a Potent Spicy Aroma Compound', *Journal of Agricultural and Food Chemistry*, vol. 56, no. 10, pp. 3738-3744.

Yang, D, Michel, L, Chaumont, J-P & Millet-Clerc, J 2000, 'Use of caryophyllene oxide as an antifungal agent in an in vitro experimental model of onychomycosis', *Mycopathologia*, vol. 148, no. 2, pp. 79-82.

Yang, Y, Kinoshita, K, Koyama, K, Takahashi, K, Tai, T, Nunoura, Y & Watanabe, K 1999, 'Anti-emetic principles of *Pogostemon cablin* (Blanco) Benth', *Phytomedicine*, vol. 6, no. 2, pp. 89-93.

Yao, LH, Jiang, YM, Shi, J, Tomás-Barberán, FA, Datta, N, Singanusong, R & Chen, SS 2004, 'Flavonoids in Food and Their Health Benefits', *Plant Foods for Human Nutrition*, vol. 59, no. 3, pp. 113-122.

Yoshikuni, Y, Ferrin, TE & Keasling, JD 2006, 'Designed divergent evolution of enzyme function', *Nature*, vol. 440, p. 1078.

Yum, DY, Lee, BY, Hahm, DH & Pan, JG 1998, 'The yiaE gene, located at 80.1 minutes on the Escherichia coli chromosome, encodes a 2-ketoaldonate reductase', *Journal of Bacteriology*, vol. 180, no. 22, pp. 5984-5988.

Zhang, P, Barlow, S, Krstic, M, Herderich, M, Fuentes, S & Howell, K 2015, 'Within-Vineyard, Within-Vine, and Within-Bunch Variability of the Rotundone Concentration in Berries of Vitis vinifera L. cv. Shiraz', *Journal of Agricultural and Food Chemistry*, vol. 63, no. 17, pp. 4276-4283.

Zhang, P, Fuentes, S, Siebert, T, Krstic, M, Herderich, M, Barlow, EW & Howell, K 2016a, 'Comparison data of common and abundant terpenes at different grape development stages in Shiraz wine grapes', *Data Brief*, vol. 8, pp. 1127-1136.

Zhang, P, Fuentes, S, Siebert, T, Krstic, M, Herderich, M, Barlow, EWR & Howell, K 2016b, 'Terpene evolution during the development of *Vitis vinifera* L. cv. Shiraz grapes', *Food Chemistry*, vol. 204, pp. 463-474.