

TECHNISCHE UNIVERSITÄT MÜNCHEN

Professur Biotechnologie der Naturstoffe

# Biosynthesis of Acylphloroglucinol Glucosides in Strawberry Fruit

Chuankui Song

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für  
Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung  
des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation

Vorsitzende(r): Univ.-Prof. Dr. K.-H. Engel

Prüfer der Dissertation:

1. Univ.-Prof. Dr. W. Schwab

2. Univ.-Prof. Dr. B. Poppenberger-Sieberer

Die Dissertation wurde am 27.07.2015 bei der Technischen Universität München eingereicht  
und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung  
und Umwelt 03.09.2015 angenommen



## Acknowledgements

Finally, after a long period, I came to the end of my PhD here in Germany, and I would like to acknowledge all the people that were involved during these three years with both my scientific and personal aspects of my doctorate.

The decision to perform a PhD in Munich was largely influenced by the recommendation I received from the head of the group Biotechnology of Natural Products, professor Wilfried Schwab (WILLI). WILLI, I would appreciate you for offering me an opportunity to pursue my PhD in your laboratory and consistent supporting me during these past three years. WILLI, thank you very much for your support, encouragement and giving me the freedom to pursue various projects within a promising framework. You have also provided insightful discussions and suggestions about the research and these are the biggest contributors that drove this project to be a complete story.

The major people who provided practical experiences and professional skills, especially at the beginning of my research are Dr. Thomas Hoffmann and Dr. Fong-Chin Huang. Tom, thank you for teaching me how to use the LC-MS, GC-MS, helpful discussions in RNAi experiments and statistical analysis throughout the PhD period. Fong-Chin, I use the word "thanks!" to express my appreciation to you. You helped me a lot with both my scientific (clone, vector construction) and personal aspects (encouraged me, helped resolving conflicts) throughout my study.

I would like to thank my colleagues (Heike Adamski, Dr. Ludwig Ring, Dr. Christopher Fuchs, Dr. Doreen Schiller, Dr. Friedericke Bönisch, Dr. Anja Preuß, Katrin Franz, Katja Schulenburg, Fatma Besbes, and Lucia Witasari). Heike, thank you, you helped me a lot with my registration and accommodation. Ludwig and Christopher, thank you for helping me with the isolation of novel compounds and I still remember the beer I drank together with you. It is the only time I got drunk..... Doreen, Friedericke, Katrin, Katja, Fatma and Lucia, I would like to say "Danke" to all of you for helping me a lot with my project and giving me lots of useful suggestion. Special thanks to Katja for helping me translate the long summary! Sorry for my bad English, you had to explain a lot even on the very simple things. It was great to have you here both as friends and colleagues in the lab. Also, I would like to thank Dr. Ruth Habegger, Mechthild Mayershofer, Kilian Skowranek and Hannelore Meckl, thank you each of you for your assistance in organizing the greenhouse space and strawberry availability.

During these years, several students spent time working with me on different aspects of the project. I would like to thank Shuai Zhao, Le Gu, Xiaotong Hong, and Jingyi Liu! All of you supported the work with great enthusiasm and the desire to learn and to do experiments. It was pleasure and fun to guide you all. I hope the time with me was of help for your present and future scientific career. I will also say “Thank you” to Ludwig Friedrich from the Chair of Food Chemistry and Molecular Sensory Science of Technische Universität München for providing authentic reference material from hop and the high-resolution mass spectrometry analysis.

Further, I would like to thank some Chinese friends whom I met in Freising (Jin Huang, Qin Sai, Tingting, Chen, Shun Li, Kai Li and so on). With all of you here, I felt not alone. We encouraged each other and shared ideas. In the second year of my PhD, I met Dr. Shengqing Shi (Chinese Academy of Forestry), a Chinese guest who has been visiting Freising for one year, and we spent a nice time together.

To my family in China who followed physically far away, but actually very close the progress of my PhD. Mother, Father, elder brother, young sister, and my girlfriend Yanli Wang (Wageningen University, The Netherlands)! Along all the years, I knew that you are always there to support me, encourage me, help me, and also wait for me. Thank you!

Also, I would like to thank Prof. Dr. Klaus Palme (University of Freiburg) for helping me to apply for the scholarship and Prof. Dr. Iryna Smetanska (Berlin University of Technology) who accepted me for a short stay and helped me to extend my visa at the beginning of my PhD when I was in trouble. Sorry I could not stay longer there, but I will remember your kind helps forever.

The last but not the least, I would like to thank for the financial support by the China Scholarship Council. Now, when I evaluate this period and achievements, I am very satisfied and happy with the brave decision to conduct my PhD in Germany.

*Chuankui*

Freising, April, 2015

## Table of Contents

Abbreviations .....	V
Summary .....	VII
Zusammenfassung .....	VII
1. Introduction .....	1
1.1 Strawberry ( <i>Fragaria sp</i> ) .....	1
1.1.1 Genetic background of strawberry .....	1
1.1.2 Functional genomics in strawberry .....	1
1.2 Phenolic compounds .....	2
1.2.2 Phloroglucinol and its derivatives <i>in planta</i> .....	3
1.2.3 Biosynthesis of phenolic compounds <i>in planta</i> .....	3
1.3 Glucosyltransferase .....	5
1.3.1 Physiological roles of glycosylation .....	5
1.3.2 Glucosyltransferases in strawberry .....	6
1.4 Volatiles in strawberry .....	7
1.5 Scope of current study .....	8
2. Material and Methods .....	9
2.1 Plant material .....	9
2.2 Chemicals and reagents .....	9
2.3 Plasmid construction .....	9
2.4 Transfection of strawberry fruit by agroinfiltration .....	9
2.5 Isolation of nucleic acids and qPCR analysis .....	10
2.6 Metabolite analysis .....	10
2.7 Preparative isolation of unknown metabolites .....	10
2.8 Nuclear Magnetic Resonance (NMR) Spectroscopy .....	11
2.9 Analysis of the NGS data and candidate genes selection .....	11
2.10 Cloning of full-length genes .....	12
2.10.1 Cloning of full-length <i>CHS</i> genes .....	12
2.10.2 Cloning of full-length <i>UGTs</i> genes .....	12
2.11 Construction of the expression vector .....	13
2.11.1 Construction of the expression vector pGEX-4T1- <i>CHS</i> .....	13

---

2.11.2	Construction of the expression vector pGEX-4T1-UGTs .....	13
2.12	Heterologous expression and partial purification of the recombinant protein .....	13
2.13	Preparation of starter CoA esters .....	14
2.14	Enzyme assay .....	14
2.14.1	Enzyme assay of CHS .....	14
2.14.2	Activity assay of FaGTs .....	15
2.15	Analyses of volatiles by gas chromatography-mass spectrometry (GC-MS) .....	16
2.16	Stable isotope labeling .....	16
2.17	Untargeted metabolite analysis .....	17
2.18	Preparation of an aglycone library .....	17
2.19	Activity based profiling using a physiologic aglycone library .....	17
2.20	Enzymatic total synthesis of APGs glucoside <i>in vitro</i> .....	17
2.21	Site-directed mutagenesis .....	18
3.	Results.....	19
3.1	Acylphloroglucinol biosynthesis in strawberry fruit .....	19
3.1.1	Selection of ripening-related genes.....	19
3.1.2	Gain or loss-of -function phenotype .....	21
3.1.3	Metabolite profiling analysis .....	23
3.1.4	Identification of unknown compounds.....	26
3.1.5	Chalcone synthase genes from <i>Fragaria vesca</i> .....	30
3.1.6	Catalytic activity of putative FvCHS enzymes.....	33
3.1.7	Kinetic properties of FvCHS enzymes and starter-CoA preference.....	39
3.1.8	Downregulation of FvCHS enzymatic activity in a transient system and a stable transgenic line.....	42
3.1.9	Production of volatiles in transiently <i>FaCHS</i> -silenced fruits and fruits of a stable transgenic line.....	44
3.1.10	Stable isotope labelling experiments .....	46
3.2	Substrate promiscuity of glucosyltransferases from strawberry .....	50
3.2.1	Selection of candidate glucosyltransferases .....	50
3.2.2	Protein expression and purification .....	51
3.2.3	Enzymatic activity.....	52
3.2.4	Identification of reaction products.....	54

## Table of Contents

3.2.5	FaGT24225a and b show contrasting regioselectivity .....	58
3.2.6	Kinetic properties and substrate preference .....	60
3.2.7	Identification of the natural substrates of FaGTs .....	61
3.2.8	Expression analysis in <i>Fragaria x ananassa</i> .....	64
3.3	Characterization of a phloroglucinol-glucosyltransferase from strawberry .....	65
3.3.1	Selection of candidate GTs and proteins purification .....	66
3.3.2	Substrate screening .....	67
3.3.3	Identification of reaction products .....	69
3.3.4	Regioselectivity of selected FaGTs .....	69
3.3.5	Substrate preference and kinetic parameters .....	71
3.3.6	FaGTs expression correlates with accumulation of APGs .....	73
3.3.7	Screening of phloroglucinol glucosyltransferase activity .....	74
3.3.8	Total enzymatic synthesis APG glucosides .....	75
3.3.9	Identification of a novel APG-glucoside based on the catalyzed reaction .....	77
3.4	UGT71C3 glucosylates the key flavor compound 4-hydroxy-2,5-dimethyl-3(2H)- furanone (HDMF) .....	79
3.4.1	Cloning and functional expression of putative FaGTs .....	79
3.4.2	Screening of volatile substrates by ten glycosyltransferases .....	79
3.4.3	FaGT07876 shows HDMF glycosylation activity .....	81
3.4.4	Glucosylation of EHMF and HMF .....	82
3.4.5	Kinetic properties of the recombinant FaGT07876a and b .....	83
3.4.6	Site-directed mutagenesis of FaGT24224 .....	84
3.4.7	Mutagenesis of residue 383 of FaGT07876b .....	86
4.	Discussion .....	88
4.1	Acylphloroglucinol biosynthesis in strawberry fruit .....	88
4.1.1	Function of candidate genes in anthocyanin accumulation .....	88
4.1.2	Untargeted analysis revealed novel strawberry metabolites .....	89
4.1.3	<i>CHS</i> genes are involved in the biosynthesis of APGs .....	90
4.1.4	<i>In planta</i> functional analysis of <i>CHS2</i> genes .....	91
4.1.5	APG pathway in strawberry fruit .....	92
4.1.6	Evolutionary relevance .....	94
4.2	Substrate promiscuity of glucosyltransferases from strawberry .....	95

4.2.1	Candidates UGTs selection and phylogenetic tree analysis.....	95
4.2.2	Biochemical characterization of selected UGTs from <i>Fragaria x ananassa</i> .....	96
4.2.3	<i>In vitro</i> versus <i>in planta</i> substrates.....	97
4.3	Characterization of a phloroglucinol-glucosyltransferase from strawberry.....	98
4.3.1	Substrate promiscuity of selected GTs.....	99
4.3.2	Regioselectivity.....	99
4.3.3	Phloroglucinol GTs in strawberry.....	100
4.3.4	APG glucoside pathway in strawberry.....	100
4.4	UGT71C3 glucosylates the key flavor compound 4-hydroxy-2,5-dimethyl-3(2H)- furanone (HDMF).....	101
4.4.1	HDMF glucosylation activity.....	101
4.4.2	Screening of volatile substrates.....	101
4.4.3	Site-directed mutagenesis.....	102
5.	References.....	104
	Curriculum vitae.....	115
	Appendix.....	116

## Abbreviations

4CL	4-coumaroyl-CoA ligase
APG	Acylphloroglucinols
BLAST	Basic local alignment search tool
2-EHMF	5-Ethyl-4-hydroxy-2-methyl-3(2H)-furanone (homofuraneol)
BSA	Bovine serum albumin
C3H	p-Coumaroyl shikimate 3'-hydroxylase
CoA	Coenzyme A
CHS	Chalcone synthase
COSY	Correlation spectroscopy
cv.	Cultivated variety
DMHF	2, 5-Dimethyl-4-hydroxy-3[2H]-furanon (Furaneol)
DMSO	Dimethyl sulfoxide
EST	Expressed sequence tag
GC-MS	Gas chromatography-mass spectrometry
HPLC	High performance liquid chromatography
ihp RNA	Intron-hairpin RNA
HMBC	Heteronuclear multiple-bond correlation spectroscopy
HMQC	Heteronuclear multiple quantum coherence
IPTG	Isopropyl- $\beta$ -D-thiogalactopyranoside
kcat	Turnover number
kb	Kilo-base pair
kDa	Kilodalton
$K_M$	Michaelis constant
LB	Luria-Bertani
HMF	4-Hydroxy-5-methyl-3[2H]-furanon (Norfuraneol)
LC-UV-ESI-MSn	Liquid chromatography ultraviolet electro spray ionization mass spectrometry
$m/z$	Mass-to-charge ratio
MES	Morpholino ethanesulfonic acid
MMLV-RT	Moloney murine leukemia viruse-reverse transcriptase

MQ	Water Milli Q water
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
MS-Salt	Murashige and Skoog basal salt mixture
NMR	Nuclear magnetic resonance spectroscopy
PAGE	Polyacrylamide gel electrophoresis
PIBP	Phloroisobutyrophenone
PIVP	Phlorisovalerophenone
PSPG	(plant secondary product glycosyltransferases)-box
RNAi	RNA interference
qRT-PCR	Quantitative real-time PCR
rpm	Rounds per minute
RT	Reverse transcription
RT-PCR	Reverse transcription polymerase chain reaction
SOC	Super optimal broth with catabolite repression
SPME	Solid phase microextraction
SDS	Sodium dodecyl sulfate
PCR	polymerase chain reaction
cDNA	Complimentary DNA
ORF	Open reading frame
NCBI	National center for biotechnology information
UGT	UDP-dependent glycosyltransferase
UV	Ultraviolet
$v_{\max}$	Maximum reaction rate
VPS	Phlorisovalerophenone synthase
X-gal	5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside

## Summary

Phenolics are a major group of metabolites in fruit crops and have multiple biological activities and health-promoting properties. In plants phenolic compounds originate from the shikimate, phenylpropanoid, flavonoid, and the lignin pathways. Genes and enzymes of the basic biosynthetic pathway leading to anthocyanins are known and remarkable progress has been achieved in understanding the regulation of this pathway. However, the regulation of phenolics accumulation and flux through the pathway is not that well established. In a recent study, an examination of the transcriptome of different strawberry fruit genotypes by microarray analyses, coupled with targeted metabolite profiling by LC-MS, was undertaken to reveal genes whose expression levels correlate with altered phenolics composition. The survey led to the identification of candidate genes that might control accumulation of phenolic compounds in strawberry fruit.

Through reverse genetic analysis of the functions of four ripening-related genes in the octoploid strawberry, *Fragaria ×ananassa*, we discovered four acylphloroglucinol (APG)-glucosides as native strawberry fruit metabolites whose levels were differently regulated in the transgenic fruits. The biosynthesis of the APG aglycones was investigated by examination of the enzymatic properties of three recombinant *F. vesca* chalcone synthase (CHS) proteins. CHS is involved in anthocyanin biosynthesis during ripening. The diploid strawberry enzymes readily catalyzed the condensation of two intermediates in branched-chain amino acids metabolism, isovaleryl-CoA and isobutyryl-CoA, with three molecules of malonyl-CoA to form phlorisovalerophenone and phlorisobutyrophenone, respectively, and formed naringenin chalcone when 4-coumaroyl-CoA was used as starter molecule. Isovaleryl-CoA was the preferred starter substrate of FvCHS2-1. Suppression of CHS activity in transient CHS-silenced *F. ×ananassa* fruit and stable transgenic lines resulted in a substantial decrease in the levels of APG glucosides and anthocyanins, and enhanced levels of volatiles derived from branched-chain amino acids. The proposed APG pathway was confirmed by feeding isotopically labeled amino acids. Thus, strawberry plants have the capacity to synthesize pharmaceutically important APGs using dual functional CHS/(phloriso)valerophenone synthases (VPS) that are expressed during fruit ripening. Duplication and adaptive evolution of CHS is the most probable scenario for this observation

and might be generally applicable to other plants. The results highlight, that important promiscuous gene function may be missed when annotation relies solely on *in silico* analysis. Glycosylation determines complexity and diversity of plant natural products, increases their solubility and accumulation, and regulates their subcellular localization and bioactivity. To functionally characterize fruit ripening-related glycosyltransferases in strawberry, we mined the publicly available *F. vesca* genome sequence and found 199 apparently distinct genes that were annotated as glycosyltransferase (GT). Candidate GT genes whose expression levels were strongly upregulated in three *F. vesca* varieties during fruit ripening were selected, cloned from *Fragaria xananassa* and five were successfully expressed in *Escherichia coli* and functionally characterized *in vitro*. FaGT26479 showed very strict substrate specificity and glucosylated only galangin out of 33 compounds. Unlike FaGT26479, the other recombinant enzymes exhibited a broad substrate tolerance *in vitro*, accepting numerous flavonoids, hydroxycoumarins, and naphthols. FaGT22709 showed highest activity towards 1-naphthol, while GT24224, 24225a, b and 24226 preferred 3-hydroxycoumarin and formed 3- and 7-O-glucosides as well as a diglucoside from flavonols. FaGT22709 could not glucosylate the hydroxyl group at position 7 of flavonols. Screening of a strawberry physiologic aglycone library identified kaempferol and quercetin, as well as three unknown natural compounds as putative *in planta* substrates of FaGT24225a and 22709 in strawberry. The study clearly shows that both, generalist and specialist GTs are expressed during strawberry fruit ripening.

APG glucosides are novel physiologically active metabolites in strawberry fruit. To functionally characterize APG GTs in strawberry, candidate GT genes were selected from a transcriptome data set of *F. vesca* and heterologously expressed in *E. coli*. Both allelic proteins of FaGT07876 were able to catalyze the glucosylation of commercially available phloroglucinol as well as acylphloroglucinol (APG) aglycons, recently identified in strawberry fruit. APG glucosylation was confirmed by total enzymatic synthesis of APG glucosides from isovaleryl-CoA/isobutyryl-CoA, malonyl-CoA and UDP-glucose by co-incubation with bifunctional chalcone/valerophenone synthase and FaGT07876. Additional substrate screening by *in vitro* assays revealed a high degree of substrate promiscuity and regioselectivity of ripening-related FaGTs. The FaGT07876 allelic proteins catalyze the last step of APG glucoside biosynthesis in *Fragaria* and thus provide the foundation for the

breeding of strawberry with improved health benefits and for the biotechnological production of bioactive natural products.

Strawberries emit several hundreds of volatiles of which only a dozen compounds truly contribute to the overall aroma perception of the ripe fruit. Among them, 4-hydroxy-2,5-dimethyl-3(2H)-furanone (HDMF) is the most important flavor compound in strawberry due to its low odor threshold and attractive flavor properties. However, in strawberry fruit HDMF is metabolized to the odorless HDMF glucoside. Thus, we analyzed strawberry fruit ripening-related glucosyltransferases (FaGTs) that function in the glycosylation of volatile metabolites. Some of the selected FaGTs show a rather broad substrate tolerance and glucosylate a range of strawberry aroma compounds *in vitro*, but others have a more limited substrate spectrum. FaGT07876a, a homologue of *Arabidopsis thaliana* UGT71C3 and its allelic protein FaGT07876b catalyze the glucosylation of the key strawberry flavor compound HDMF. Both proteins also convert the commercially important structural homologue 2 (or 5)-ethyl-4-hydroxy-5 (or 2)-methyl-3(2H)-furanone (EHMF) to the  $\beta$ -D-glucoside with a similar efficiency, but none of them could convert 4-hydroxy-5-methyl-3(2H)-furanone (HMF) *in vitro*. The functional characterization of FaGT07876 as UDP-glucose:HDMF glucosyltransferase provides the foundation for improvement of strawberry flavor and the biotechnological production of HDMF-glucosid.

## Zusammenfassung

Phenolische Verbindungen sind wichtige Sekundärmetabolite in verschiedenen Nutzpflanzen und weisen zahlreiche biologische und ernährungsphysiologische Eigenschaften auf. Sie tragen wesentlich zur Farbe, zum Geruch und Geschmack von Früchten bei und besitzen zudem gesundheitsfördernde Eigenschaften, weshalb sie für den Menschen von besonderem Interesse sind. Die Biosynthese phenolischer Komponenten ist in der Pflanze über verschiedene Stoffwechselwege reguliert. Sie können dem Shikimat-, Phenylpropanoid-, Flavonoid- sowohl als auch dem Lignin-Stoffwechsel entspringen. Der Biosyntheseweg, der der Entstehung von Anthocyanen zugrunde liegt, wurde in der Vergangenheit intensiv beforscht, weshalb nun auf ein solides wissenschaftliches Grundverständnis über die Wirkungsweise beteiligter Gene und Enzyme zugegriffen werden kann. Wenig bekannt ist jedoch über die Regulierung der Akkumulation der Anthocyane in pflanzlichen Geweben. In einer kürzlich durchgeführten Studie wurde das Transkriptom von Früchten verschiedener Erdbeer-Varietäten mittels einer Mikroarray-Analyse untersucht und gleichzeitig eine gezielte Metaboliten-Profilierung mittels LC-MS durchgeführt. Hierbei sollten Gene identifiziert werden, deren Expressionslevel mit der Veränderung der Gehalte an phenolischen Komponenten korreliert. Durch diese Studie wurden Kandidatengene identifiziert, die vermutlich die Akkumulation phenolischer Substanzen in der Erdbeerfrucht regulieren.

In der vorliegenden Arbeit wurde die Funktion von vier reife-korrelierten Genen der oktoploiden Erdbeere *Fragaria xananassa* durch eine „revers genetische“ Analyse untersucht. Es konnten damit erstmals vier Acylphloroglucinol (APG)-Glukoside als native Erdbeerfrucht-Metabolite nachweisen, deren Gehalte in den transgenen Früchten unterschiedlich reguliert waren. Durch die Bestimmung der enzymatischen Eigenschaften von drei rekombinanten *F. vesca* Chalkon-Synthase (CHS) Proteinen war es möglich die Biosynthese der zugehörigen APG-Aglyka zu erklären. Während der Fruchtreife ist CHS an der Biosynthese von Anthocyanen beteiligt. Die Enzyme der diploiden Erdbeere katalysierten bereitwillig die Kondensation von zwei Intermediaten des Metabolismus der verzweigtkettigen Aminosäuren, Isovaleryl-CoA und Isobutyryl-CoA mit drei Molekülen Malonyl-CoA wodurch Phlorisovalerophenon und Phlorisobutyrophenon gebildet werden. Naringenin-Chalcon entsteht, wenn 4-Coumaroyl-CoA als Starter-Molekül eingesetzt wird.

Isovaleryl-CoA war das bevorzugte Starter-Molekül von FvCHS2-1. Eine Unterdrückung der CHS Aktivität führte sowohl in transient „stillgelegten“ *F. xananassa* Früchten als auch in Früchten stabil transgener Linien zu einer erheblichen Verminderung des Gehalts an APG-Glukosiden und Anthocyanen, wohingegen der Gehalt an flüchtigen Derivaten verzweigt-kettiger Aminosäuren anstieg. Der postulierte APG-Biosyntheseweg wurde zusätzlich durch Markierungsversuche mittels isotopenmarkierter Aminosäuren und den Nachweis ihrer Folgeprodukte bestätigt. Somit konnte eindeutig gezeigt werden, dass Erdbeerpflanzen die Fähigkeit besitzen pharmakologisch aktive APGs zu synthetisieren. Die Pflanzen bedienen sich hierbei der dual agierenden CHS/(Phloroiso)valerophenon Synthase (VPS), die während der Fruchtreifung exprimiert wird. Genduplikation und adaptive Evolution der CHS sind höchstwahrscheinlich der Grund für diese zweite Funktion. Grundsätzlich ist diese Beobachtung auch für andere Pflanzensysteme denkbar. Unsere Ergebnisse belegen, dass wichtige promiskuitive Genfunktionen unentdeckt bleiben können, wenn sich Annotationen ausschließlich auf eine *in silico* Analyse stützen.

Glykosylierungen bestimmen die Komplexität und Vielfältigkeit natürlicher Pflanzenprodukte. Sie erhöhen Löslichkeit und Akkumulation der Aglyka und regulieren deren subzelluläre Lokalisation und Bioaktivität. Um reife-korrelierte Glykosyltransferasen (GTs) der Erdbeerpflanze funktionell zu charakterisieren, wurde die veröffentlichte *F. vesca* Genomsequenz genutzt und 199 Kandidatengene, die als Glykosyltransferasen annotiert waren identifiziert. Potentielle GTs, deren Transkriptgehalt während der Fruchtreife stark anstieg, wurden ausgewählt und aus *F. xananassa* kloniert. Fünf Gene konnten erfolgreich in *Escherichia coli* exprimiert und *in vitro* charakterisiert werden. FaGT26479 zeigte eine spezifische enzymatische Aktivität und glykosylierte von 33 eingesetzten Substraten ausschließlich Galangin. Im Gegensatz zu FaGT26479 besaßen die anderen rekombinanten Enzyme *in vitro* eine eher breite Substrattoleranz. Sie setzten zahlreiche Flavonoide, Hydroxycoumarine und Naphthole um. FaGT22709 katalysierte bevorzugt die Glukosylierung von 1-Naphthol. Bei GT24224, 24225a, b und 24226 hingegen war die Aktivität gegenüber 3-Hydroxycoumarin am höchsten, es wurden jedoch ebenfalls diverse 3- bzw. 7-O-Glykoside gebildet als auch ein Flavonol-Diglykosid. Das Durchmustern einer physiologischen „Erdbeer-Aglyka-Bibliothek“ lieferte Kaempferol, Quercetin und drei weitere unbekannte natürliche Stoffe als potentielle *in planta* Substrate der Glykosyltransferasen 24225a und 22709. Unsere Ergebnisse zeigen somit, dass während der

Fruchtreife nicht nur GTs mit breiter Substrattoleranz (Generalisten), sondern auch GTs mit sehr spezifischer Aktivität (Spezialisten) exprimiert werden.

Im Rahmen dieser Thesis konnten APG-Glukoside als neue physiologisch aktive Metabolite der Erdbeerfrucht identifiziert werden. Um APG spezifische GTs der Erdbeere funktionell zu charakterisieren, wurden Kandidatengene aus einem Transkriptom Datensatz der Walderdbeere *F. vesca* ausgewählt und in *E. coli* heterolog exprimiert. Beide allelische Versionen des FaGT07876 Proteins katalysierten die Glukosylierung käuflicher Phloroglucinol- und APG Aglyka. Durch Co-Inkubation von FaGT07876 und der CHS/VPS konnte die enzymatische Synthese von APG Glukosiden aus den Substraten Isovaleryl-CoA/Isobutyryl-CoA, Malonyl-CoA und UDP-Glukose bestätigt werden. Ein *in vitro* Screening ließ auf ein hohes Maß an Substratpromiskuität und Regioselektivität der reife-korrelierten FaGTs schließen. Die allelischen Versionen der FaGT07876 katalysieren den finalen Schritt der APG Glukosid-Biosynthese. Sie liefern somit die Grundlage für die Züchtung von Erdbeerpflanzen mit verbessertem gesundheitlichem Potential und die Produktion von bioaktiven Naturstoffen.

Erdbeeren (*F. ×ananassa*) emittieren eine Vielzahl leichtflüchtiger Stoffe ab. Jedoch trägt nur ein Dutzend dieser Verbindungen wirklich zum Aroma der reifen Frucht bei, das vom Menschen wahrgenommen werden kann. Unter ihnen ist 4-Hydroxy-2,5-dimethyl-3(2H) -furanon (HDMF) die Wichtigste, aufgrund ihres niedrigen Geruchsschwellenwertes und ihren attraktiven Eigenschaften. Im Rahmen dieser Thesis wurden GTs der Erdbeerfrucht (faGTs) charakterisiert, die flüchtige Metabolite glykosylieren. Einige der untersuchten FaGTs zeigen eine breite Substrattoleranz und glykosylieren eine große Anzahl an Erdbeeraroma Komponenten *in vitro*. Andere dagegen akzeptieren ein limitiertes Spektrum an Substraten. FaGT07876a, ein Homologes der UGT71C3 aus *Arabidopsis thaliana* und das allelische Protein FaGT07876b katalysieren die Glykosylierung von HDMF, der zentralen Aroma-Komponente der Erdbeerfrucht. Beide Proteine setzen *in vitro* außerdem das wirtschaftlich bedeutsame 2 (bzw. 5)-Ethyl-4-hydroxy-5 (bzw. 2)-methyl-3(2H)-furanon (EHMF) zu einem  $\beta$ -D-Glukosid um. Nicht umgesetzt wurde hingegen 4-Hydroxy-5-methyl-3(2H)-furanon (HMF). Die funktionelle Charakterisierung der ersten UDP-Glukose:HDMF Glukosyltransferase FaGT07876 aus Erdbeeren liefert die Grundlage zur Verbesserung des Erdbeeraromas in Lebensmitteln und die Kenntnis kann zur biotechnologischen Produktion von HDMF-Glukosid genutzt werden.

# 1. Introduction

## 1.1 Strawberry (*Fragaria sp*)

### 1.1.1 Genetic background of strawberry

Strawberry is one of the most economically important fruit plant and is a member of the Rosaceae family (Potter et al., 2007). The only octoploid cultivated species, *F. × ananassa* which emerged from hybridization between *F. chiloensis* and *F. virginiana* (Darrow, 1966) is grown widely in Europe. The *F. × ananassa* genome ( $2n=8x=56$ ) harbors eight sets of chromosomes, which derived from four different diploid relatives. The complex genetic background of the octoploid strawberry has limited its application in molecular, genetic, and functional studies. Thus, the diploid woodland strawberry, *F. vesca*, the closest diploid relative to the octoploid *F. × ananassa*, based on an analysis of chloroplast DNA and nuclear DNA (Potter et al., 2000), has emerged as a model system for genomic and genetic studies (Slovin et al., 2009) of the cultivated octoploid strawberry and Rosaceae family due to its small physical size, small ( $2n = 14$ , 240 Mb) and sequenced genome, short life cycle, and facile transformation (Shulaev et al., 2011).

### 1.1.2 Functional genomics in strawberry

Functional genomics studies are currently performed by two fundamentally different genetic approaches. The forward genetics approach begins with the observation of a mutant phenotype, either natural or induced by mutagenesis. Analysis of the inheritance of a variation in a mapping population will allow the identification of the genes responsible for this mutant. The limiting factors for this approach are the time and effort required for creating the mapping population and fine mapping of the mutant locus (Radhamony et al., 2005; Peters et al., 2003).

Reverse genetics rely upon sequence information (Peters et al., 2003). This approach begins with a candidate gene and identifies the mutant phenotypes that results from disruption (Mathur et al., 1998). The reverse genetics is one way to discover the biological roles of the thousands of new genes. The availability of the complete genome sequence of *F. vesca* (Shulaev et al., 2011) makes reverse genetics an ideal method to study the function of

candidate genes. Ribonucleic acid interference (RNAi) of plant genes, initiated by the delivery of double-stranded RNA (dsRNA), is an attractive new reverse-genetics tool for the study of gene function (Waterhouse and Helliwell, 2003).

## **1.2 Phenolic compounds**

### **1.2.1 Phenolic compounds in strawberry**

Phenolic compounds constitute one of the most numerous and ubiquitous groups of plant secondary metabolites and have attracted much attention due to their reputed beneficial effects on human health protection (Scalbert et al, 2005). Phenolics are reported to play a role, among others, in the prevention of cardiovascular disease, diabetes mellitus, and neurodegenerative disease (Scalbert et al, 2005), inhibition of the growth of human tumor cell lines (Zhang et al, 2008), and against UV-A irradiation damage on human fibroblasts (Giampieri et al., 2012).

Strawberry fruits are consumed in high quantities and represent a valuable source of phenolics such as flavonoids, comprising anthocyanins, flavonols and flavan-3-ols including proanthocyanidins, phenolic acids, and ellagitannins (Aaby et al, 2005; Seeram et al, 2006). The anthocyanins that give the fruits the intense red color, mainly consist of pelargonidin-3-glucoside, contributing 60-95% of the total anthocyanin content in ripe fruits, in addition to pelargonidin-3-glucoside-6'-malonyl (0-33%), pelargonidin-3-rutinoside (0-7%), and cyanidin-3-glucoside (1-6%; Aaby et al., 2007). Glucosides and glucuronides of kaempferol and quercetin are the major flavonols (Buendía et al., 2010; Fischer et al., 2014). Proanthocyanidins with an average degree of polymerization of 3.4 to 6.3 units are abundant in strawberry fruit but (+)-catechin is the only flavan-3-ol monomer present (Buendía et al., 2010; Fischer et al., 2014). Glucose esters of cinnamic acid, 4-coumaric acid, and ferulic acid are the main phenolic acid derivatives and agrimoniin and casuarictin are the major ellagitannins (Lunkenbein et al., 2006a; Gasperotti et al., 2013).

In recent years liquid chromatography mass spectrometry (LC-MS) has become a powerful tool to analyze the phenolics in different plant tissues, including strawberry fruit and flower (Aaby et al, 2007; Hanhineva et al., 2008; Seeram et al., 2006). These compounds play several important functions in plants as the red colored anthocyanins may attract frugivores that help to disperse seeds, flavonols act as protective UV-B absorbing chemicals in fruit skin

and proanthocyanidins contribute to defense and stress resistance (Cheynier et al., 2013). The glucose esters of phenylpropanoic acids serve as energy-rich substrates in plant secondary metabolism, and ellagic acid may play a role in protection from predation and in plant growth regulation (Cheynier et al., 2013).

### **1.2.2 Phloroglucinol and its derivatives *in planta***

Phloroglucinol and its derivatives are phenolics that have attracted much attention due to their reputed biological activities, such as antibacterial activity, cytotoxic, antiproliferative, and antiangiogenic effects (Schmidt et al., 2012). They have been widely used in medicine, cosmetics, pesticides, paints, cements and dyes (Singh et al., 2010). In addition to the chemical synthetic pathways, phloroglucinols also include about 700 naturally occurring compounds, of which acylphloroglucinols (APGs) comprise the largest group of natural phloroglucinol compounds and show various biological activities (Singh et al., 2010).

More than 50 glycosides of phloroglucinol and its derivatives have been identified from natural sources (Singh et al., 2010). In plants, APG glucosides are reported to be the prominent secondary metabolites of the genus *Hypericum* (Shiu et al., 2012; Crispin et al., 2013) and *Humulus* (Bohr et al., 2005) and have been detected in *Phyllanthus emblica* (Zhang et al., 2002), *Jatropha multifida* (Kosasi et al., 1989) and *Curcuma comosa* (Suksamrarn et al., 1997) but are rarely found in other plant species.

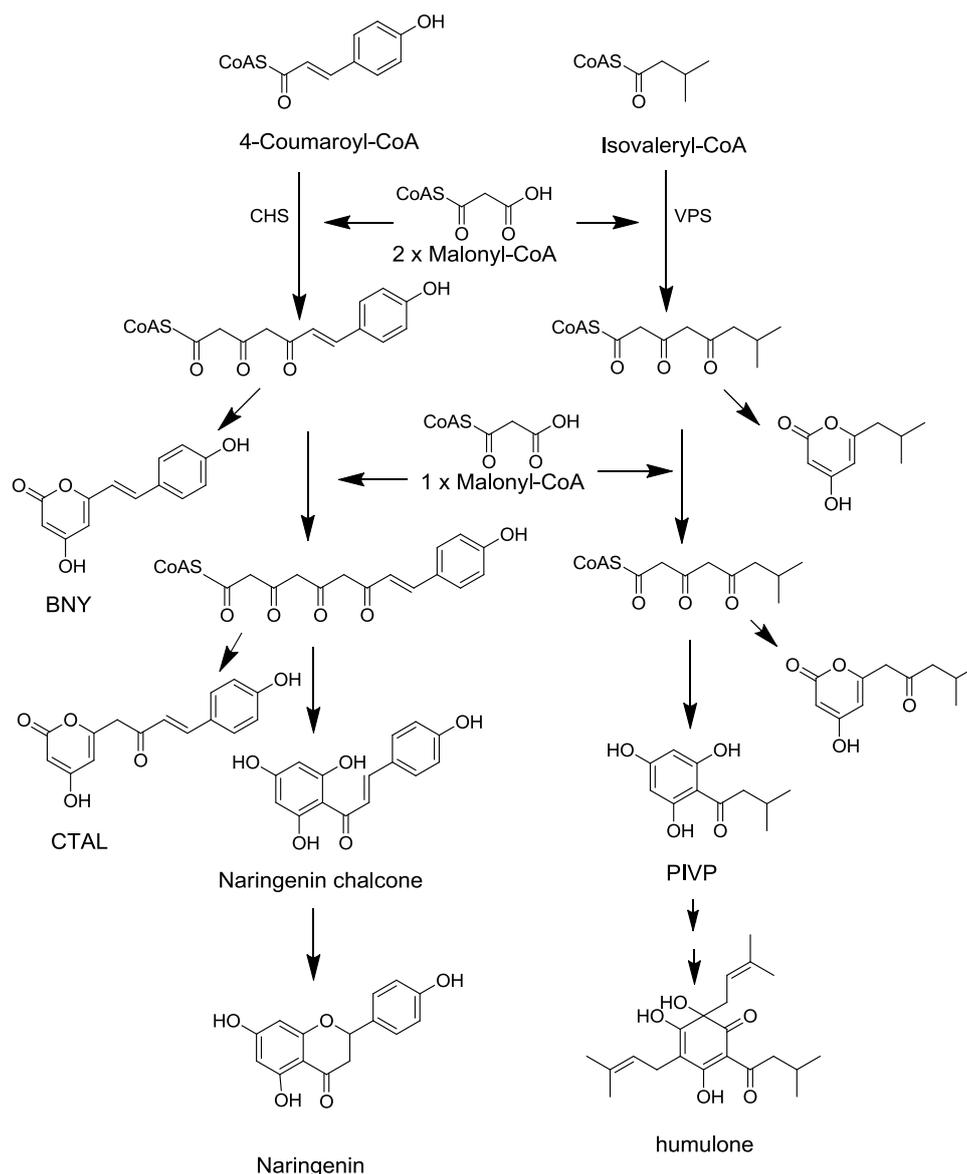
### **1.2.3 Biosynthesis of phenolic compounds *in planta***

In plants, phenolic compounds originate from the shikimate, phenylpropanoid, flavonoid, and the lignin pathways (Vogt, 2010). The shikimate pathway provides aromatic amino acids and may deliver gallic acid, the presumed precursor of ellagitannins, whereas in most plants, the biosynthesis of the phenolics starts with 4-coumaric acid formation from the primary metabolite phenylalanine. The phenylpropanoids are further modified in many ways, including the elongation and cyclization by the sequential addition of three molecules of malonyl-CoA to form flavonoids that are finally converted to anthocyanidins. Genes and enzymes of the basic biosynthetic pathway leading to anthocyanins are known (Ververidis et al., 2007) and remarkable progress has been made in understanding the regulation of this pathway (Boudet, 2007; Allan et al., 2008). However, the regulation of their accumulation and flux through the pathway is not that well established. In a recent study, an examination of the transcriptome of different strawberry fruit genotypes by microarray analyses, coupled

with targeted metabolite profiling by LC-MS, was undertaken to reveal genes whose expression levels correlate with altered phenolics composition (Ring et al., 2013). The survey led to the identification of candidate genes that might control accumulation of phenolic compounds in strawberry fruit.

The biosynthesis of phloroglucinol and its derivatives has also been extensively studied. In bacteria, the pHID (type III polyketide synthase) gene has been reported to be involved in phloroglucinol formation (Bangera and Thomashow, 1999; Achkar et al., 2005). The aglycones (phloroisovalerophenone, PIVP and phloroisobutyrophenone, PIBP) of bioactive acylphloroglucinol glucosides which are found in hop are generated by the enzyme phloroisovalerophenone synthase (VPS) during the biosynthesis of the hop bitter acids (Paniego et al., 1999). Condensation of three malonyl-CoA units and one 4-coumaroyl-CoA or isovaleryl-CoA molecule gives rise to naringenin chalcone (catalyzed by chalcone synthase, CHS) and PIVP (catalyzed by VPS), respectively (Figure 1; Paniego et al., 1999). PIBP is formed when isobutyryl-CoA is used instead of isovaleryl-Co. Phloroglucinol and [U-<sup>14</sup>C]-glucose feeding experiment in leaf disks of *Pelargonium hortorum* yielded [U-<sup>14</sup>C]-labeled phloroglucinol glucoside (Hutchinson et al., 1958), which indicated that plants have the capacity to glycosylate phloroglucinol using glucose as a donor substrate.

Recent structural and functional studies of CHS have elucidated the basic chemical mechanism for polyketide formation (Ferrer et al 1999; Jez et al., 2000). Three essential catalytic amino acids, Cys164, His303, and Asn336 (Jez et al., 2000, 2001, 2002) are conserved in all known CHS-related enzymes. In addition, a site-directed mutagenesis research guided by the three-dimensional structure of CHS revealed that two Phe residues (Phe215 and Phe265), situated at the active site entrance, are important in determining the substrate specificity of CHS (Jez et al., 2002).



**Figure 1.** Proposed formation pathway of naringenin chalcone and phloroisovalerophenone, (PIVP) from 4-coumaroyl-CoA and isovaleryl-CoA by CHS and VPS, respectively. BNY bisnoryangonin, CTAL 4-coumaroyltriacetic acid lactone.

## 1.3 Glucosyltransferase

### 1.3.1 Physiological roles of glycosylation

Glycosylation is a widespread modification of plant secondary metabolites. In plant, UDP-dependent glucosyltransferases (UGTs) transfer activated sugar molecules to a broad range of substrates. The glycosylation of low-molecular-weight compounds in plants usually changes acceptors in terms of increasing solubility and accumulation, regulating their subcellular localization and bioactivity, such as antioxidant and anticancer activity (Bowles

et al., 2005; Kramer et al., 2003). In addition, glycosylation is also involved in the detoxification of xenobiotics (Gachon et al., 2005; Lim and Bowles, 2004), in the regulation of the active levels of several hormones (Xu et al., 2012; Ghose et al., 2014; Poppenberger et al., 2005) as well as in the chemical stabilization of secondary metabolites. For instance glycosylation at the 3-OH position of anthocyanins is crucial for their stability (Kroon et al., 1994; Brugliera et al., 1994; Boss et al., 1996; Fukuchi-Mizutani et al., 2003), and glycosylation of plant hormones inactivates their biological activity (Kleczkowski et al., 1995). Plants have many UGTs in their genome. More than 120 UGT encoding genes have been identified in *Arabidopsis thaliana* (Gachon et al., 2005; Osmani et al., 2009). Numerous studies on UGTs revealed that single or multiple glycosylation of the acceptors can occur at -OH, -COOH, -NH<sub>2</sub>, -SH, and C-C groups (Bowles et al., 2005).

Plant UGTs capable of transferring sugars to a wide range of small molecules belong to family 1 UGTs, and are defined by the presence of a 44 amino acid C-terminal signature motif designated as the PSPG (plant secondary product glycosyltransferases)-box (Masada et al., 2007). Understanding the mechanism of the glycosylation reaction and the physiological roles of glycosides would be of great importance for synthesizing valuable glycosides *in vitro*, and for metabolic engineering of important agronomic traits of crops (Kristensen et al., 2005; Bowles et al., 2005; Weis et al., 2008). Although the UGT family has been studied intensively for many years, to date only a handful UGTs have been characterized in plants (Gachon et al., 2005).

### **1.3.2 Glycosyltransferases in strawberry**

Strawberry is one of the most popular fruit crops worldwide (Giovannoni, 2001) and contains high levels of antioxidants such as anthocyanins and flavonoids. Quercetin and kaempferol, two important flavonols in strawberry fruit, occur as glucosides and glucuronides (Häkkinen et al., 1999). Their glycosylation pattern is quite complex in strawberry. Quercetin 3-O-, 7-O-, and 4'-O-glucoside, kaempferol 3-O-, and 7-O-glucoside, as well as isorhamnetin 7-O-glucoside have been reported to be present in strawberry fruits (Griesser et al., 2008b). In recent years, several strawberry UGTs have been isolated and functionally characterized, including one UGT that preferentially glucosylates biochanin A (Cheng et al., 1994), two anthocyanidin glycosyltransferases (Almeida et al., 2007; Griesser et al., 2008a), and two flavonol-O-glucosyltransferases (Griesser et al., 2008b) which accept

numerous flavonoids as substrates *in vitro*. Only limited information is available about the *in planta* substrates (Griesser et al., 2008a). There are approximately 200 candidate UGTs annotated in the genome sequence of the woodland strawberry *Fragaria vesca* (Shulaev et al., 2011) but more than 95% of them are uncharacterized and their functions are unknown.

#### **1.4 Volatiles in strawberry**

Plants have the capability to synthesize, accumulate and emit volatiles that may act as aroma and flavor molecules due to interactions with human receptors. They are commercially important for the food, pharmaceutical, agricultural and chemical industries as flavorants, drugs, pesticides and industrial feedstocks (Schwab et al., 2008). Strawberries emit several hundreds of volatiles of which only a dozen compounds truly contribute to the aroma perception of the ripe fruit (Schieberle and Hofmann, 1997). The most important is 4-hydroxy-2,5-dimethyl-3(2H)-furanone (HDMF, furaneol) due to their low odor threshold and attractive flavor properties (Schwab, 2013). HDMF exhibits a caramel-like aroma similar to its structural homologues 4-hydroxy-5-methyl-3(2H)-furanone (HMF, norfuraneol) and 2-ethyl-4-hydroxy-5-methyl-3(2H)-furanone or the tautomer 5-ethyl-4-hydroxy-2-methyl-3(2H)-furanone (2-EHMF or 5-EHMF, respectively, homofuraneol). All these molecules are based on a cyclic dicarbonyl compound with a planar enol-tautomeric substructure capable of forming strong hydrogen bonds (Schiefner et al., 2013).

Incorporation experiments using radio-labeled precursors and substrates labeled with stable isotopes revealed that D-fructose-1, 6-diphosphate is an efficient biogenetic precursor of HDMF and provided initial evidence for the enzymatic formation of this important aroma compound in strawberries (Schwab, 1998). In strawberry and tomato, the hexose diphosphate is converted by an as yet unknown enzyme to 4-hydroxy-5-methyl-2-methylene-3(2H)-furanone, which serves as the substrate for an enone oxidoreductase (FaEO) that catalyzes the final biosynthetic step to HDMF (Klein et al., 2007; Raab et al., 2006; Schiefner et al., 2013). In strawberry fruit HDMF is further metabolized by FaOMT (*Fragaria x ananassa* O-methyltransferase) to its methyl ether 2,5-dimethyl-4-methoxy-3(2H)-furanone, its  $\beta$ -D-glucoside and, subsequently, to the malonylated derivative of the glucoside (Roscher et al., 1997; Wein et al., 2002). HDMF  $\beta$ -D-glucoside is the naturally occurring glycosidically bound form of HDMF (Roscher et al., 1996).

## 1.5 Scope of current study

The aim of the current work was to confirm the relationship between the expression pattern of the candidate genes and the accumulation of phenolics and to structurally identify novel phenolic metabolites whose levels are affected by the transcript levels of the candidate genes. Through reverse genetic analysis of four ripening-related genes in octoploid strawberry, we discovered four biologically active acylphloroglucinol (APG)-glucosides as native strawberry metabolites.

The formation of the APG aglycones should be investigated by examination of the enzymatic properties of three recombinant *F. vesca* chalcone synthase (CHS) proteins. The dual CHS/VPS function should be confirmed by activity assays and suppression of CHS catalytic activity in transient *CHS*-silenced strawberry fruit and a stable antisense *chs* transgenic line as well as by tracer experiments using isotopically labeled precursor amino acids.

Furthermore fruit ripening-related glucosyltransferases should be functionally characterized. Thus, ripening related UGTs were selected based on their transcription levels in three *F. vesca* varieties during fruit maturation. Substrate promiscuity should be functionally characterized *in vitro* after heterologous expression of the UGTs in *Escherichia coli*.

The study yielded the first APG and HDMF UGTs and total enzymatic synthesis of APG glucosides was achieved for the first time.

## 2. Material and Methods

### 2.1 Plant material

Strawberry plants (*F. x ananassa* cv. Elsanta, Mara des Bois, Senga Sengana, Calypso, and a stable transgenic *FaCHS* antisense Calypso line, all belong to the octaploid cultivars; Lunkenbein et al., 2006b) were cultured in a green house in Freising, Germany. Growing conditions were maintained at 25°C and a 16-h photoperiod under 120  $\mu\text{mol m}^{-2} \text{sec}^{-1}$  irradiance provided by Osram Fluora lamps (München, Germany). For genetic and molecular analyses, fruits were injected 14 d after pollination and harvested 10 – 14 d after agroinfiltration. Seasonal flowering short-day strawberry plants (diploid *F. vesca*) and perpetual flowering *F. vesca* accession Hawaii-4 were used for cloning and were also cultured in the green house.

### 2.2 Chemicals and reagents

Isovaleryl-CoA, isobutyryl-CoA, malonyl-CoA, L-isoleucine- $^{13}\text{C}_6$ ,  $^{15}\text{N}$  (98+%  $^{13}\text{C}$ , 98+%  $^{15}\text{N}$ ), L-leucine- $^{13}\text{C}_6$ ,  $^{15}\text{N}$  (98 %  $^{13}\text{C}$ , 98 %  $^{15}\text{N}$ ), and L-valine-2,3,4,4,4,5,5,5,-D<sub>8</sub> (98 atom %D) were purchased from Sigma (Taufkirchen, Germany). UDP-[U- $^{14}\text{C}$ ] glucose (300 mCi  $\text{mmol}^{-1}$ , 0.1 mCi  $\text{mL}^{-1}$ ) was obtained from American Radiolabeled Compounds. All chemicals and solvents were obtained from Sigma, Fluka and Aldrich (all Taufkirchen, Germany), Carl Roth (Karlsruhe, Germany) and VWR International (Darmstadt, Germany), unless otherwise noted. Authentic phloroglucinol derivatives were kindly provided by the Chair of Food Chemistry and Sensory Analysis, Technische Universität München.

### 2.3 Plasmid construction

The fragments (200-400 bp) corresponding to the candidate genes *expansin-A8-like* (gene21343) and *ephrin-A1-like* (gene33865) were PCR amplified from *F. x ananassa* cv. Elsanta DNA using primers shown in Appendix Table 1, and cloned into the P9U10 vector. Similarly, the full length DNA corresponding to the *SRG1-like* (gene10776) and *defensin-like* (gene00897) genes were PCR amplified and cloned into the PBI121 vector.

### 2.4 Transfection of strawberry fruit by agroinfiltration

The *Agrobacterium tumefaciens* strains AGL0 containing the p9U10-*gene21343*, p9U10-*gene33865*, pBI121-*gene10776*, pBI121-*gene00897*, pBI-Intron-*GUS*, and pBI-*CHS* were

grown at 28 °C in LB medium with appropriate antibiotics according to Hoffmann et al. (2006). When the culture reached an OD<sub>600</sub> of about 0.8, *Agrobacterium* cells were harvested and re-suspended in a modified MMA medium (MS salts, 10 mM MES pH 5.6, 20 g/L sucrose). The *Agrobacterium* suspension was evenly injected throughout the entire attached fruit about 14 d after pollination by using a sterile 1-ml hypodermic syringe.

### **2.5 Isolation of nucleic acids and qPCR analysis**

Ten to 14 days after injection, agroinfiltrated and the wild type fruits were harvested and individually freeze-dried. For candidate gene expression analysis, wild type fruits (*F. x ananassa* cv. Elsanta) of different ripening stages (green, white, red), root, stem and leaves were harvested and freeze-dried. After grinding to a fine powder, 200 mg of the powder from each sample were used for RNA extraction according to the protocol reported (Liao et al., 2004), followed by DNase I (Fermentas, St. Leon-Rot, Germany) treatment and RT-PCR (Promega, Mannheim, Germany). Real-time PCR was performed with a StepOnePlus real-time PCR system (Applied Biosystems) using Fast SYBR Green Master Mix (Applied Biosystems) to monitor double-stranded cDNA synthesis according to Ring et al. (2013). At least five biological replicates for each sample were used for the quantitative RT-PCR analysis, and at least two technical replicates were analyzed for each biological replicate. An interspacer gene was used as an internal control for normalization. The gene-specific primers used to detect the transcripts are listed in Appendix Table 1.

### **2.6 Metabolite Analysis**

Metabolites in individual strawberry fruits were analyzed about 14 d after injection in the ripe stage. The materials were ground in liquid nitrogen and kept at -80 °C for further use. For metabolite analysis, an aliquot of 50 mg was used for each of the three biological replicates. Samples were worked-up and analyzed by LC-MS according to Ring et al. (2013).

### **2.7 Preparative isolation of unknown metabolites**

Frozen *F. x ananassa* cv Elsanta fruits (2.0 kg) were extracted with 1 L of methanol. After vortexing and sonication for 10 min the sample was centrifuged at 16,000 g for 10 min. The supernatant was removed and the residue was re-extracted with 500 mL methanol. The supernatants were combined, concentrated to dryness in a vacuum concentrator and re-dissolved in 10 mL water. After 1 min vortexing, 10 min sonication and 10 min centrifugation at 16,000g the clear supernatant was used for preparative fractionation. Separations were

carried out using a RP18 column (25 cm × 46 mm, particle size 5 µm, Phenomenex) at room temperature, connected to a Jasco PU-1580 LC system and Jasco UV-1575 detector (both Jasco, Gross-Umstadt, Germany). Metabolites were separated by the following gradient of water containing 1% formic acid (A) and methanol containing 1% formic acid (B): 0-5 min, 0% B; 5-30 min, 0-40% B; 30-35 min, 100% B. The injection volume was 2 mL and the flow rate 10 mL/min. One fraction was collected every min. Fractions containing the target compounds were further purified by the gradient: 0-5 min, 20% B; 5-30 min, 20-40% B. Furthermore, a normal phase column (SeQuant ZIC®-HILIC 200Å, 5µ, 250 x 10 mm, Merck, Darmstadt) was used with water containing 1% formic acid (A) and acetonitrile containing 1% formic acid (C). The gradient was 0-30 min, 95-80% C; 30-35 min, 80-0% C; 35-45 min, 0% C. Final purification was achieved by analytical LC (RP18 column, 25 cm x 4.0 mm, particle size 5 µm, Phenomenex) applying the following gradient: 0-5 min, 0% B; 5-35 min, 0-100% B; 35-50 min, 100% B. Purity was analyzed by LC-MS as described by Ring et al. (2013.). High resolution mass spectra of the compounds were measured on a Bruker Micro-TOF (Bruker Daltronics, Bremen, Germany) mass spectrometer and referenced to sodium formate (Intelmann et al., 2011).

## **2.8 Nuclear Magnetic Resonance (NMR) Spectroscopy**

The samples were evaporated and dissolved in methanol-D<sub>4</sub>, (99.8%) containing 0.03% v/v TMS. <sup>1</sup>H NMR spectra were recorded at 500.13 MHz with a Bruker DRX 500 spectrometer (Bruker, Karlsruhe, Germany). The chemical shifts were referred to the solvent signal. The one-dimensional and two-dimensional COSY, HMQC, and HMBC spectra were acquired and processed with standard Bruker software (XWIN-NMR).

## **2.9 Analysis of the NGS data and candidate genes selection**

Candidate genes were selected based on their transcript levels in strawberry (*F. vesca*) achenes and receptacle of different developmental stages (green, white and red). Transcript abundances were determined by next generation sequencing (RNAseq, Schulenburg and Franz-Oberdoerf, unpublished results). GT genes whose expression levels showed a ripening related expression pattern were selected for cloning.

## **2.10 Cloning of full-length genes**

### **2.10.1 Cloning of full-length *CHS* genes**

Total RNA was isolated from mature fruit of *F. vesca* using the method described by Liao et al. (2004). First-strand cDNA was synthesized from 1 µg of DNase I (Fermentas)-treated total RNA by Moloney murine leukemia virus reverse transcriptase H2 (Promega) and random hexamer primer oligo(dT) primer (5'-GCT GTC AAC GAT ACG CTA CGT AAC GGC ATG ACA GTG T(18)-3'). The cDNA fragments of *CHS2* were amplified by PCR with the cDNA templates of *F. vesca* fruit. The specific primers were designed according to the *F. vesca* mRNA (*gene26825* and *gene26826*) sequence (Appendix Table 1). The full-length coding sequences were cloned from cDNA by PCR using proof-reading Pfu DNA polymerase (Promega, Mannheim, Germany). The PCR was carried out in a 25-µL total reaction volume and the temperature program used was 5 min at 95 °C, one cycle; 45 s at 95 °C, 45 s at 55 °C, 2 min at 72 °C, 35 cycles, and the final extension at 72 °C for 10 min. The PCR products were A-tailed and cloned into the pGEM-T Easy (Promega, Mannheim, Germany) vector and the ligation product transformed into *E. coli* NEB10 beta (New England Biolabs, Frankfurt, Germany). The identity of the cloned genes was confirmed by sequencing the complete insert (MWG Biotech, Ebersberg, Germany) from both sides and by restriction enzyme digest with BamH1 and EcoRI.

### **2.10.2 Cloning of full-length UGTs genes**

The full length UGTs were amplified by PCR from *F. x ananassa* red fruit cDNA using a pair of gene specific primers (Appendix Table 2 and 3), designed according to the *F. vesca* mRNA sequences. The full-length coding sequences were amplified using proof-reading Phusion DNA polymerase (Promega, Mannheim, Germany). The PCR was carried out in a 25-µL total reaction volume and the temperature program used was 5 min at 95 °C, one cycle; 45 s at 95 °C, 45 s at 55 °C, 2 min at 72 °C, 35 cycles; final extension at 72 °C for 10 min. The PCR products was A-tailed and cloned into the pGEM-T Easy (Promega, Mannheim, Germany) vector and the ligation product transformed into *E. coli* NEB10 beta (New England Biolabs, Frankfurt, Germany). The identity of the cloned gene was confirmed by sequencing the complete insert (MWG Biotech, Ebersberg, Germany) from both sides and by restriction enzyme digest with BamH1 and EcoRI.

## **2.11 Construction of the expression vector**

### **2.11.1 Construction of the expression vector pGEX-4T1-CHS**

To obtain recombinant GST fusion proteins for functional characterization, the *CHS* sequences were cloned into the expression vector pGEX-4 T-1 (Amersham Biosciences, Freiburg, Germany). After digestion with two restriction enzymes (EcoRI and BamH), the full-length *CHS* genes were sub-cloned into the pGEX-4T1 vector, which had been digested by the same enzymes. The recombinant plasmid (pGEX-4T1-*CHS*) was then introduced into *E. coli* NEB10 beta, plasmid DNA was purified and sequenced to check for correct insertion.

### **2.11.2 Construction of the expression vector pGEX-4T1-UGTs**

The amplified full length sequences were digested with EcoRI and BamH1 and the resulting gene fragments were cloned into the expression vector pGEX-4 T-1 (Amersham Biosciences, Freiburg, Germany). The recombinant plasmids (pGEX-4T1-*UGTs*) were then introduced into *E. coli* NEB10 beta, plasmid DNAs were purified and sequenced to check for correct insertions.

## **2.12 Heterologous expression and partial purification of the recombinant protein**

Expression constructs were transformed into *E. coli* strain BL21 (DE3) pLysS. A single colony of *E. coli* strain BL21(DE3) pLysS cells harboring the pGEX-4T1-*CHS* plasmid or pGEX-4T1-*UGTs* was introduced and cultured overnight at 37 °C in Luria–Bertani (LB) liquid medium containing ampicillin (50 µg/mL) and chloramphenicol (50 µg/mL). The following day the culture was diluted 1:40 with LB medium containing the antibiotics and grown under the same conditions as above until the optical density (OD<sub>600</sub>) of the cultured cells reached 0.8. After inducing expression by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM the culture was incubated at 16 °C and 150 rpm. The next day cells were harvested by centrifugation (5,000g, 10 min) and subjected to a purification procedure using glutathione S-transferase (GST) bind resin (Novagen, Darmstadt, Germany), following the manufacturer's protocol with slight modifications. All steps were performed at 4 °C with pre-chilled buffers to maintain enzyme activity. The harvested cells were frozen at -80 °C for 15 min and the pellet re-suspended in 10 ml GST wash buffer (43 mM Na<sub>2</sub>HPO<sub>4</sub>, 14.7 mM KH<sub>2</sub>PO<sub>4</sub>, 1.37 M NaCl, 27 mM KCl, pH 7.3). The freeze-thaw cycle supported cell disruption, which was further enhanced by subsequent sonication in three intervals of 30 s at 50% power (Sonopuls UW2200, Bandelin electronic, Berlin, Germany). The lysate was

centrifuged (10,000g, 20 min) and incubated for at least 60 min (soft shaking) with GST bind resin, previously equilibrated with GST wash buffer. The liquid was removed by gravity flow and the resin was washed two times with 5 mL of GST wash buffer. Then, the resin was incubated 5 min at room temperature with 200  $\mu$ L of 1x elution buffer (50 mM Tris-Cl, pH 8.0, and 10 mM reduced glutathione) and the flow through collected. This step was repeated 3 times. The resulting fractions were assessed using SDS-PAGE. A BL21 strain and a BL21 strain harboring the pEXT-4T1 empty vector were induced and subjected to the same procedure and served as controls. Protein concentration was determined by the photometric method of Bradford, 1976.

### **2.13 Preparation of starter CoA esters**

Enzymatic synthesis of hydroxycinnamoyl-CoA (cinnamoyl-CoA, 4-coumaroyl-CoA, caffeoyl-CoA, and feruloyl-CoA) was carried out with purified 4-coumarate-CoA ligase (4CL), according to the method of Beuerle and Pichersky (2002). To purify the hydroxycinnamoyl-CoAs, 0.8 g of ammonium acetate was added to the reaction mixture. Before the reaction mixture was loaded onto a 6 mL Isolute C8 (EC) SPE column (Biotage, UK), the column was washed successively with methanol, distilled water, and 4% ammonium acetate solution. After loading, the column was washed with 4% ammonium acetate solution until free CoA was detectable, which was determined spectrophotometrically at 259 nm. The CoA esters were obtained by elution with MQ water. Fractions (2 mL) were collected and lyophilized. The reaction products were dissolved in MQ water after freeze-drying, and the concentration of the CoAs was determined spectrophotometrically by measuring the absorbance at the absorption maxima of the products according to Beuerle and Pichersky (2002).

### **2.14 Enzyme assay**

#### **2.14.1 Enzyme assay of CHS**

The reactions were performed according to Zuurbier et al. (1998) with modifications. The standard assay for determining CHS activity was conducted in a total volume of 100  $\mu$ L containing 100 mM potassium phosphate buffer (pH 7.0), 300  $\mu$ M malonyl-CoA, and 100  $\mu$ M starter CoA. The mixtures were incubated at 30 °C for 30 min. Reactions were initiated by addition of enzyme and were quenched with 5% (vol/vol) acetic acid. The products were extracted twice with 200  $\mu$ L ethyl acetate and analyzed by HPLC and LC-MS. Optimum

reaction temperature was determined in assays as before which were carried out in the range of 10 to 50 °C at pH 7.0. Before adding the enzyme and substrate, the mixture was equilibrated to the tested temperature. The pH optimum was tested in the range from pH 3 to pH 10. Citric acid, sodium phosphate and Tris-HCl were used for pH 3 to 6, pH 6 to 8, and pH 8 to pH 10, respectively. At least two biological replicates (different preparations) were carried out, and the total product formation was measured by LC-MS analysis. LC-MS was performed using a Nucleosil C18 (15 cm x 4.0 mm, particle size 4 µm, Macherey Nagel, Düren, Germany) column with a flow rate of 0.1 mL/min at UV 280 nm detection, according to Ring et al. (2013). Gradient elution was performed with water (A) and methanol (B), both containing 0.1% formic acid: 0-5 min, 0-70% B; 5-25 min, 70-100% B; 25-30 min, 100% B. Products were identified by literature data and comparison of the retention times and MS data with those of authentic references. The amount of each product was calculated as mg-equivalent of naringenin using a standard curve. Steady-state kinetic constants were determined from initial velocity measurements (Jez et al., 2000). All kinetic experiments were conducted in the enzyme assay buffer at pH 7.0, 300 µM malonyl-CoA, and various concentrations of starter CoA, and the reactions were incubated at 40 °C for 10 min. Reactions were initiated by addition of enzyme and were quenched with 5% (v/v) acetic acid, then the products were extracted twice with 200 µl ethyl acetate for 2 min, and the upper layer was analyzed by LC-MS. At least seven different substrate concentrations covering the range of 1–300 µM were used. Data were fitted to the Michaelis–Menten equation using a non-linear regression program (Enzyme Kinetics, Sigmaplot) to calculate  $v_{\max}$  and  $K_M$  values.

#### **2.14.2 Activity assay of FaGTs**

In the initial screening, each reaction mixture (200 µL in total) contained Tris-HCl buffer (100 mM, pH 7.0, 10% glycerol, and 10 mM 2-mercaptoethanol), 37 pmol of UDP-[U-<sup>14</sup>C] glucose (0.01 µCi), substrate (10 µL of a 10 mM stock solution), and purified protein (2-5 µg per reaction). The reaction mixture was incubated at 30 °C for 1 hour. The assays were stopped by extraction with 800 mL of water-saturated 1-butanol. The organic phase was mixed with 2 mL of Pro Flow P+ cocktail (Meridian Biotechnologies), and radioactivity was determined by liquid scintillation counting (Tri-Carb 2800TR; Perkin Elmer). Optimum reaction temperature was determined in the range of 10 to 50 °C at pH 7.0. The pH optimum was tested in the range from pH 3 to pH 10. Citric acid, sodium phosphate and Tris-HCl buffers

were used for pH 3 to 6, 6 to 8, and 8 to 10, respectively. At least two biological replicates (different preparations) were carried out, and radioactivity of products was determined by liquid scintillation counting. Assay conditions used for determining the kinetic data were essentially the same except only 2  $\mu\text{g}$  recombinant protein was used. At least seven different substrate concentrations covering the range of 1–500  $\mu\text{M}$  were used. Data were fitted to the Michaelis–Menten equation using a non-linear regression program (Enzyme Kinetics, SigmaPlot) to calculate  $v_{\text{max}}$  and  $K_{\text{M}}$  values.

### **2.15 Analyses of volatiles by gas chromatography-mass spectrometry (GC-MS)**

Volatiles released by fruits of the *CHS*-silenced transgenic lines (agroinfiltrated fruits and fruits of a stable *chs* transgenic line; Lunkenbein et al., 2006b) and wild type strawberry fruits were sampled by solid phase micro extraction (SPME). Intact fruits of approximately the same size at mature stages were placed under a glass funnel with caps to create a headspace, as described by Aharoni et al. (2000). The volatile compounds collected from the headspace were analyzed by a Thermo Finnigan Trace DSQ mass spectrometer coupled to a BPX520M fused silica capillary column with a 30 m x 0.25 mm inner diameter. Helium (1.1 ml/min) was used as carrier gas. The injector temperature was 250 °C, the ion source and interface temperatures were kept at 250 and 280 °C, respectively. The temperature program was 40 °C for 3 min, increased to 280 °C at a rate of 5 °C/min. The EI–MS ionization voltage was 70 eV (electron impact ionization). Mass data were acquired in the range of  $m/z$  50–650. Compounds were identified by comparing their mass spectra and retention indices to the NIST mass spectra library and authentic reference compounds. Relative concentration of volatile esters was performed according to Sinz and Schwab, 2011.

### **2.16 Stable isotope labeling**

L-isoleucine- $^{13}\text{C}_6$ , L-leucine- $^{13}\text{C}_6$ , and L-valine-2,3,4,4,4,5,5,5- $\text{D}_8$  were evenly injected throughout the entire strawberry (*F. x ananassa* cv. Elsanta, Mara des Bois) fruits, which were still attached to the plant by a 1mL syringe according to Wein et al. (2001). Mature fruits of similar sizes were selected for the application. Solutions (500  $\mu\text{L}$ ) containing the isotopically labeled compounds (10 - 50 mM in water) were injected into the strawberry fruits from the top. As a control experiment, 500  $\mu\text{L}$  of an aqueous solution containing 10 - 50 mM of unlabeled L-isoleucine, L-leucine, and L-valine were injected into strawberry fruits. The experiments were repeated at least twice. The fruits were harvested after 1 day and

stored at -20 °C until they were analyzed by LC-MS. Besides, equal amounts (10 mM) of L-leucine-<sup>13</sup>C<sub>6</sub> and L-isoleucine-<sup>13</sup>C<sub>6</sub> were administered to strawberry fruit of the same ripening stage and similar weight. After 1, 2, 4, 5 and 7 days, individual fruits were harvested and stored at -20 °C until they were analyzed by LC-MS.

### **2.17 Untargeted metabolite analysis**

Analysis by metaXCMS was performed according to Tautenhahn et al. (2011).

### **2.18 Preparation of an aglycone library**

Fifty mg of strawberry glycosides isolated by XAD solid phase extraction was dissolved in 50 mL 250 mM MES buffer (pH 5.84). Fifty mg Rapidase AR 2000 (DSM Food Specialties Beverage ingredients, Delft, Netherlands) was added and incubated for 24 hours at room temperature. A strawberry extract without the addition of AR 2000 was used as a control. The liberated aglycones were extracted by 200 mL ethyl acetate for two times. The organic phases were combined, concentrated to dryness in a vacuum concentrator and the residue re-dissolved in 500 µL methanol and 500 µL water. The extracts were combined.

### **2.19 Activity based profiling using a physiologic aglycone library**

Each solution (200 µL in total) contained 5 µg purified enzyme, 100 µL Tris-HCl buffer (100 mM, pH 7.0, 10% glycerin), 0.1 µL UDP-[U-<sup>14</sup>C] glucose (0.01 µCi), 44.9 µL unlabeled UDP-glucose (0.2 mM) and 10-20 µL aglycone library extract (dissolved in methanol/water 1:1, v/v). The reaction mixture was incubated at 30°C overnight and was stopped by adding 1 mL of 24% (v/v) trichloroacetic acid. Products were extracted with 800 mL of water-saturated 1-butanol. The organic phase was mixed with 2 mL of Pro Flow P+ cocktail (Meridian Biotechnologies), and radioactivity was determined by liquid scintillation counting (Tri-Carb 2800TR; Perkin Elmer). Alternatively, reactions were performed with only unlabeled UDP-glucose and the products analyzed by LC-MS.

### **2.20 Enzymatic total synthesis of APGs glucoside *in vitro***

Five microgram FvCHS2 (encoded by gene26826) and FaGT07876 was incubated with isovaleryl-CoA/isobutyryl-CoA (100 µM), malonyl-CoA (300 µM), and 5 mM UDP-glucose in a volume of 200 µL of 100 mM Tris-HCl buffer (pH 7). Enzyme assay was performed at 30° C overnight. As a control, BL21 (DE3) pLysS cells were transformed with an empty pGEX-4T-1 vector and the respective protein extract was assayed under the same condition. The

reaction solutions were extracted twice with 200  $\mu$ L ethyl acetate, and the organic solvent was vaporized, redissolved in 100  $\mu$ L water/methanol (1:1, v/v) and analyzed by LC-MS. The products were identified by comparison of the MS and MS2 spectra, and the retention times with those data of the APGs glucosides identified in strawberry fruit.

### **2.21 Site-Directed Mutagenesis**

The site-directed mutagenesis of FaGT24224 (D259R, D343E, A389V, K458R, D445E, G433A, I434V) and FaGT07876 (V383A) was conducted using the Quick-Change strategy (Novagen). The primers were designed as described in the manual (Appendix Table 4). The cDNAs were amplified using proof-reading Phusion DNA polymerase (Promega, Mannheim, Germany). The PCR was carried out in two 20- $\mu$ L total reaction volume with either forward primer or reverse primer and the temperature program used was 30 s at 98 °C, one cycle; 30 s at 98 °C, 4.5 minutes at 72 °C, 15 cycles; final extension at 72 °C for 5 min. Then, the two reactions were combined, mixed and used for further reaction and use the same temperature program. Mutations were verified by sequencing. The mutant proteins were expressed and purified using the same procedures described above for native protein.

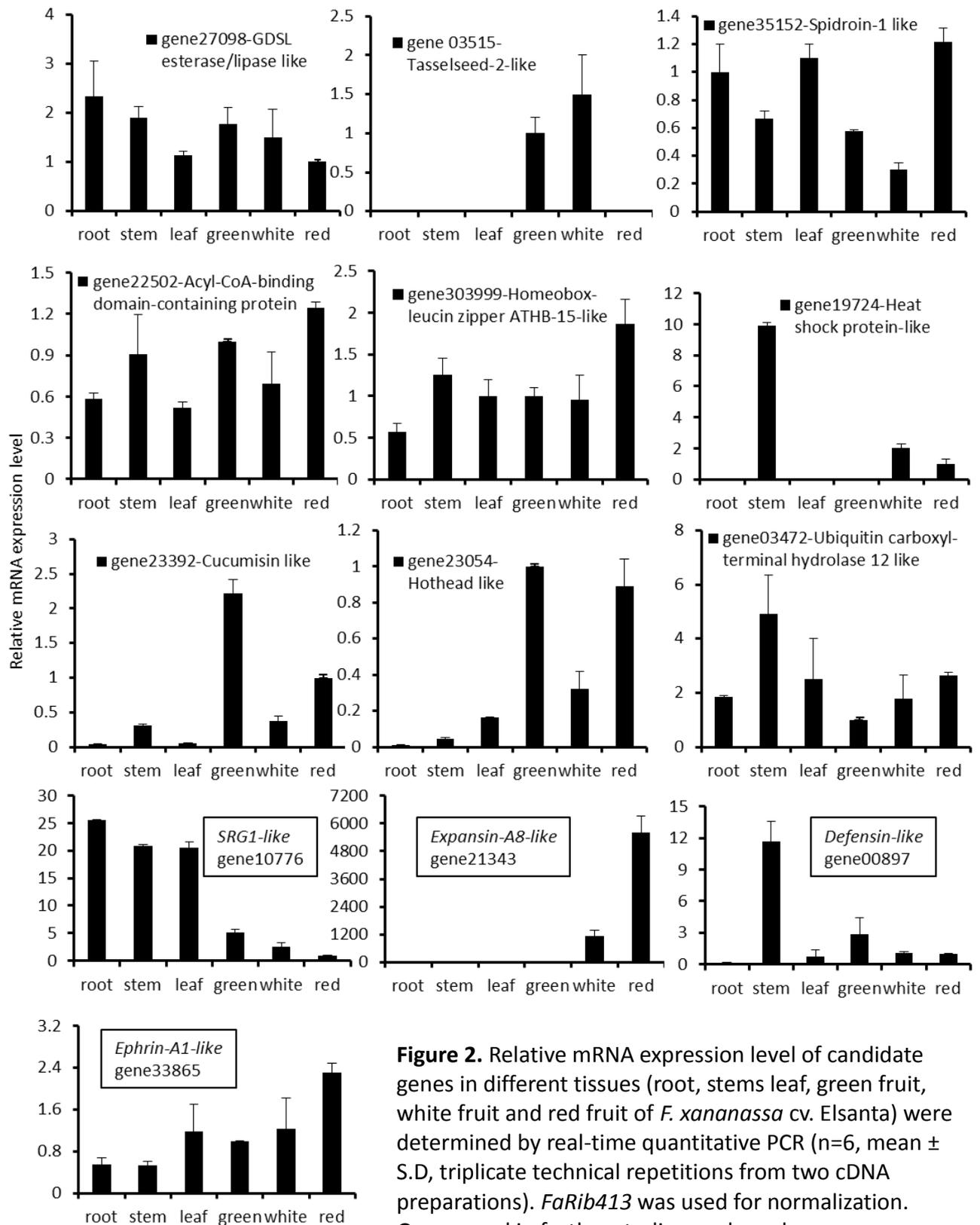
## 3. Results

### 3.1 Acylphloroglucinol biosynthesis in strawberry fruit

#### 3.1.1 Selection of Ripening-related Genes

In a previous study, comparison of the transcript patterns of different strawberry genotypes by microarray analysis combined with a metabolite profiling study revealed novel candidate genes that might affect the accumulation of flavonoids and anthocyanins in strawberry fruit during ripening (Ring et al., 2013). Thus, the mRNA transcription levels of eleven candidates were determined by quantitative real-time PCR in vegetative tissues (root, stems, and leaves) and in fruits at different developmental stages to select genes that show a ripening-related expression pattern (Figure 2). Primers are listed in Appendix Table 1. The data indicate that *expansin-A8*-like transcripts are highly abundant in the red ripe fruit and the expression level strongly increases during ripening (Figure 2). *Ephrin-A1*-like mRNA is only slightly expressed in the different tissues but the relative transcript level increases during fruit development. On the contrary, gene *SRG1*-like transcripts are highly expressed in vegetative tissue and the level decreases with the development of the fruit. Similarly, *defensin*-like mRNA is abundant in stem tissue but in fruit the level falls during ripening. Thus, these genes were used for further analyses. The other genes did not show a ripening-related expression pattern or have very low expression levels.

## Results

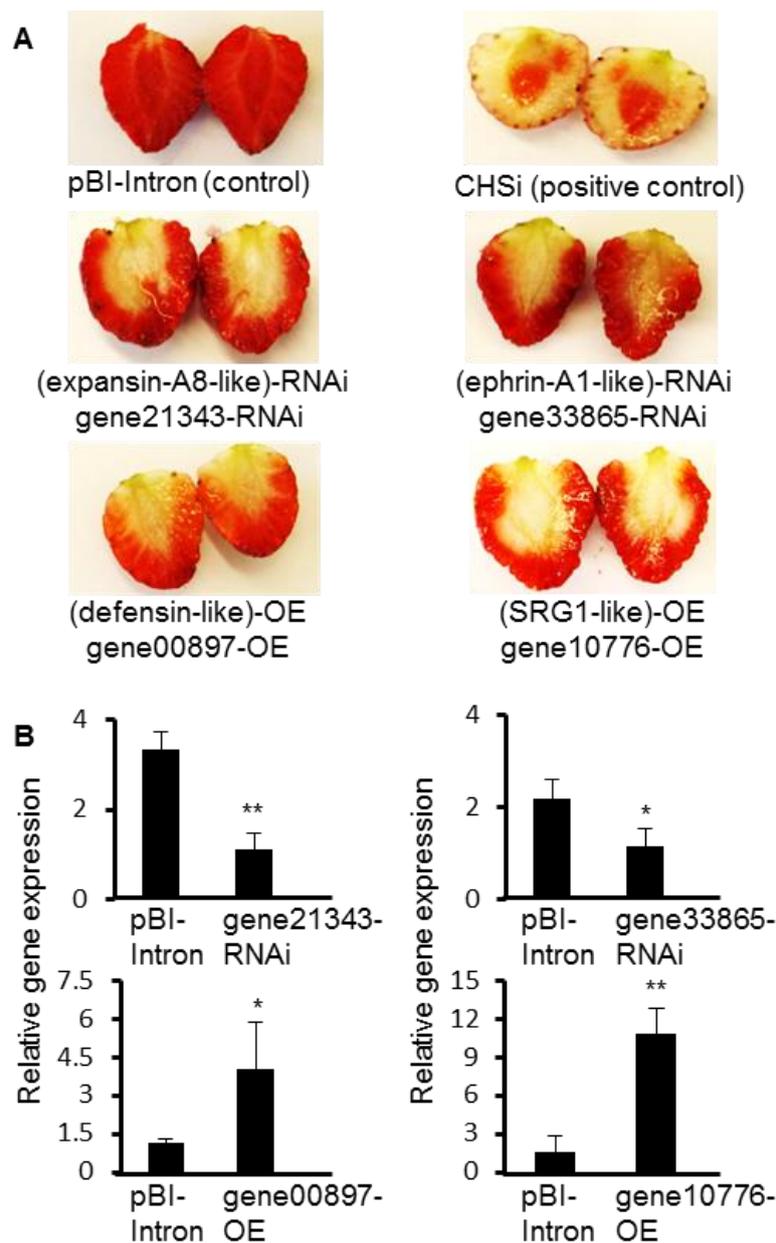


**Figure 2.** Relative mRNA expression level of candidate genes in different tissues (root, stems leaf, green fruit, white fruit and red fruit of *F. xananassa* cv. Elsanta) were determined by real-time quantitative PCR (n=6, mean  $\pm$  S.D, triplicate technical repetitions from two cDNA preparations). *FaRib413* was used for normalization. Genes used in further studies are boxed.

### 3.1.2 Gain or loss-of -function phenotype

To answer the question whether the transcription of the ripening-related genes affect the accumulation of phenolic compounds in strawberry fruit, gain- and loss-of-function phenotypes were generated by transient overexpression or silencing of the candidate genes. *Agrobacterium tumefaciens* strains AGL0 containing overexpression (pBI121-*gene10776*, pBI121-*gene00897*) and silencing (p9U10-*gene21343*, p9U10-*gene33865*) constructs were evenly injected throughout the entire fruit of cultivated strawberry *F. x ananassa* cv. Mara des Bois about 14 d after pollination. Fruits remained attached to the plants. *CHS*, the first gene in the flavonoid pathway, was chosen as a positive reporter gene in this study. Down-regulation of *CHS* transcript levels can lead immediately to loss of pigmentation in strawberry fruit (Lunkenbein et al., 2006b; Hoffmann et al., 2006) and is thus easy to detect (Figure 3).

Transcript levels of the *expansin-A8*- and *ephrin-A1*-like genes were successfully down-regulated by agroinfiltration and resulted in loss-of-function phenotypes that showed white regions (Figure 3), a clear sign of impaired anthocyanin accumulation. On the other hand, up-regulation of *SRG1* and *defensin*-like mRNA levels produced gain-of-function phenotypes with white regions, indicative that *SRG1* and *defensin*-like might be negative regulators of the anthocyanin pathway.



**Figure 3.** Strawberry fruit phenotypes (**A**) and relative mRNA expression levels of candidate genes in agroinfiltrated fruit (**B**). pBI-Intron control, CHSi fruits agroinfiltrated with pBI-CHSi as positive control and fruits agroinfiltrated with gene-silencing (RNAi) and overexpression (OE) constructs of defined genes. Transcript expression levels were determined by quantitative real-time qPCR. An interspacer gene was used as an internal control for normalization, and pBI-Intron (gene00897 and gene10776), gene21343 or gene33865 fruit was set at 1 as a reference (means  $\pm$  SE six replicates with two sets of cDNAs). Relative changes are shown.

### 3.1.3 Metabolite profiling analysis

Each of the candidate genes analyzed in this study might be involved in different metabolic pathways and mechanisms, however, their expression levels putatively correlated with phenolics accumulation (Ring et al., 2013). Thus, we hypothesized that these four candidates might affect the formation of some common metabolites of the phenolics pathway. To understand the altered metabolism when mRNA levels are up- or down-regulated we compared the metabolite profiles of the transgenic fruits with those of the control fruits (pBI-Intron). Targeted metabolite profile analysis showed that the levels of anthocyanins (pelargonidin and cyanidin derivatives), flavonoids (naringenin, kaempferol, quercetin, afzelechin and (epi)catechin derivatives) and phenylpropanoids (cinnamoyl, 4-coumaroyl, caffeoyl, and feruloyl glucose) were differently affected in the transgenic fruit in comparison to the controls (Figure 4A). Consistent with the gain- and loss-of-function phenotypes (Figure 3A) it is evident that all selected candidate genes are involved in the accumulation of phenolics in strawberry fruit as changes in their transcript levels alter the pool sizes of metabolites in the different branches of the phenolics pathway.

Pairwise untargeted comparisons provide physiologically relevant data but often result in hundreds of differences (Patti et al., 2012). To facilitate the extraction of interesting metabolites from our large untargeted LC-MS data sets before the time-consuming step of structural identification we performed second-order (meta-) analysis by metaXCMS (Gowda et al., 2014). In this investigation, second-order comparison was applied using a tolerance of  $m/z$  0.01 and 60 s retention time to find out interesting metabolite features associated with the candidate pathways. In the negative MS ion model, the positive control (*CHS* gene) possess the most unique metabolite features (15 unique signals; Table 1), followed by the *defensin*-like, *ephrin-A1*-like, *DRG1*-like and *expansin-A8*-like gene that own 11, 5, 4 and 2 unique metabolites in this analysis, respectively (Figure 4B).



## Results

**Table 1.** Metabolites (*m/z* negative ionization and retention time) that were found to be differentially accumulated in strawberry fruit after down-regulation of *CHS* (*CHSi*) when compared with levels in the pBI-Intron-*GUS* control fruit (pBI).

<i>m/z</i>	retention time	CHSi vs PBI	p-value
715	34.28	-17.1	5.00E-05
435	37.17	-5.3	8.00E-05
309	32.95	-5.8	0.00036
565	34.30	5.4	0.00042
709	32.49	23.9	5.00E-04
355	32.93	-6.7	0.00062
356	32.96	-6.7	0.00109
661	34.25	-16.7	0.00173
347	33.00	-8.1	0.00193
345	32.96	-5.5	0.0028
371	38.27	-5.6	0.00305
389	32.43	7.5	0.00308
399	32.42	15	0.00404
709	32.47	26.9	0.00771
664	18.05	-23.3	0.00926
371	37.10	-6.4	0.01726

**Table 2.** Metabolites (*m/z* and retention times) that were found to be differentially accumulated in strawberry fruit after down-regulation of *CHS* (*CHSi*), gene21343, and gene33865, and overexpression of gene10776 and gene00897 when compared with levels in the pBI-Intron-*GUS* control fruit (pBI).

<i>m/z</i>	retention time	CHSi_vs_BI.Intron		9U10.7_vs_BI.Intron		BI.10_vs_BI.Intron		BI.10_vs_BI.Intron		BI.19_vs_BI.Intron	
		Fold change	p-value	Fold change	p-value	Fold change	p-value	Fold change	p-value	Fold change	p-value
357	34.4	-17.1	0.00005	-6.2	0.00001	-5.0	0.00014	-5.4	0.00003	-5.9	0.00001
323	33.1	-28.1	0.00049	-118.8	0.00046	-9.1	0.00034	-9.5	0.00011	-109	0.00045
188	17.9	-54.6	0.00238	-81.7	0.00229	-24.7	0.00252	-27.6	0.00229	-85.8	0.00231
321	36.8	-184.2	0.00261	-62.2	0.00263	-6.2	0.00401	-5.3	0.00315	-109.2	0.00262
664	18.0	-65.5	0.00275	-55.3	0.00272	-5.5	0.00273	-20.9	0.00264	-255.8	0.00264
313	33.1	-31.1	0.00281	-77.9	0.00262	-16.9	0.00302	-21.5	0.00283	-88.3	0.00262
455	31.7	-33.8	0.00397	-43.3	0.00387	-7.3	0.00547	-8.6	0.00307	-59.9	0.00379
771	40.2	-86.8	0.00608	-38.1	0.00637	-20.2	0.00685	-360.4	0.0059	-72.4	0.00612
496	18.1	-15.3	0.00764	-17.1	0.00701	-5.1	0.00959	-7.7	0.00726	-32.3	0.00675

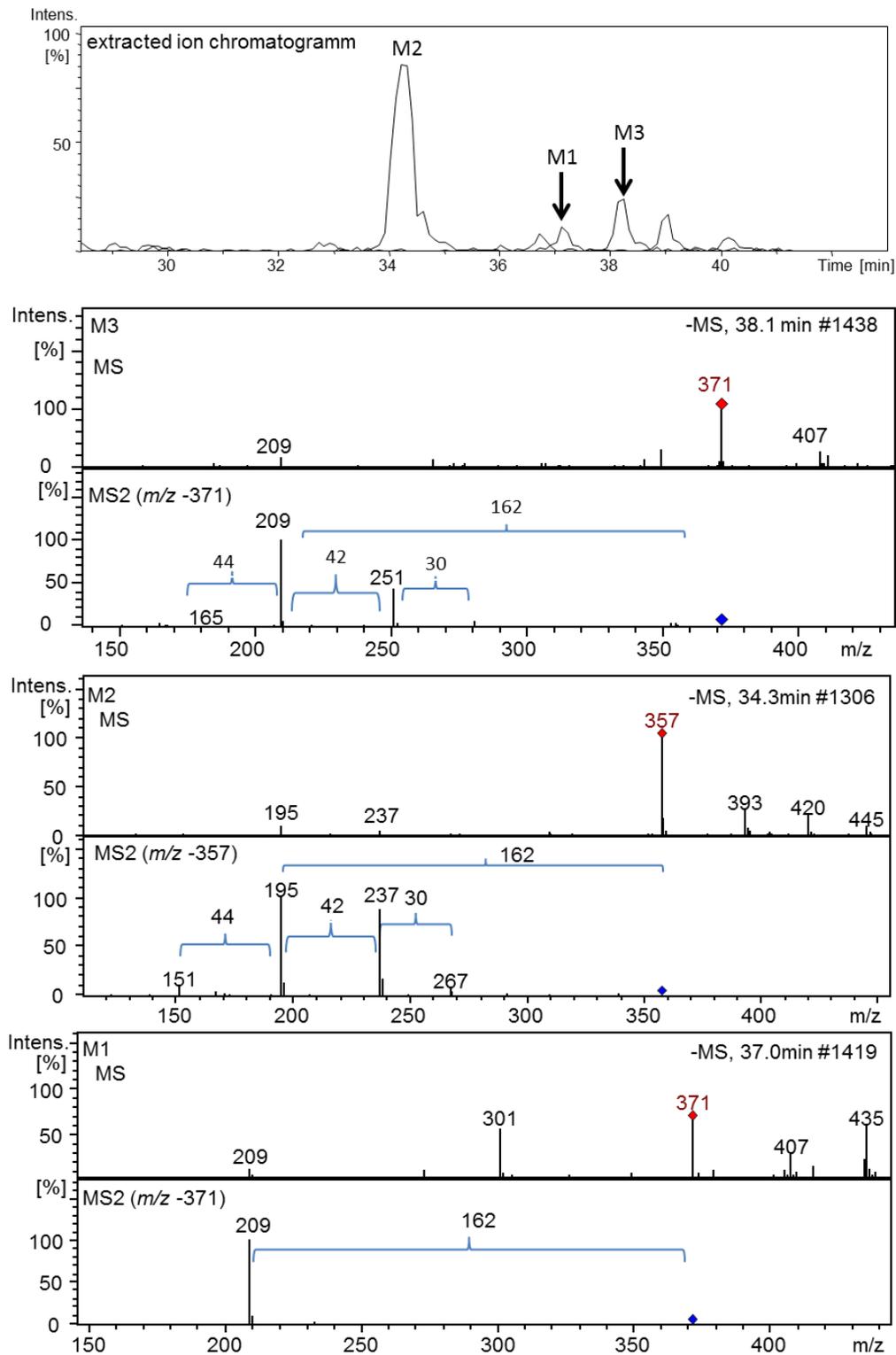
Eleven features were found to be differentially regulated in all five samples, among them are 9 metabolites that could be confirmed manually and assigned to LC-MS signals (retention times and  $m/z$  of the pseudo-molecular ions, Table 2). Thus, there are common metabolites whose levels are regulated by the candidate genes in strawberry fruit. Known strawberry metabolites were not among them. Consequently, isolation and elucidation of the chemical structure of shared and unique metabolites were undertaken.

### 3.1.4 Identification of unknown compounds

The mass spectra of metabolite 3 (M3, *CHSi* unique metabolite) and 2 (M2, common feature) show a similar fragmentation pattern (Figure 5), in which mass differences of  $m/z$  30, 42, 44, and 162 indicate a loss of  $\text{CH}_2\text{O}$ ,  $\text{C}_2\text{H}_2\text{O}$ ,  $\text{CO}_2$ , and a glucose residue, respectively. The mass difference between their pseudomolecular ions  $m/z$  371  $[\text{M-H}]^-$  and 357 equals 14, the mass of a  $-\text{CH}_2-$  unit.

Metabolite 1 (M1) is an isomer of M3, sharing the same pseudomolecular ion ( $m/z$  371  $[\text{M-H}]^-$ ) but shows a different product ion spectrum (MS2). Preparative liquid chromatography yielded pure M1 which was identified as 1-[(2-methylbutyryl)-phlorogluciny]-2-*O*- $\beta$ -D-glucopyranoside by  $^1\text{H}$  NMR, HMQC, and HMBC data (Figure 6; Table 3). The  $^1\text{H}$  NMR spectrum shows one methyl doublet at  $\delta$  1.15, d (3H), one methyl triplet at  $\delta$  0.90, t (3H), a methine signal at  $\delta$  3.93 m (H-2') and a geminally coupled methylene spin system ( $\delta$  1.81, m, H-3') which indicates a 2-methylbutyryl moiety (Kosasi et al., 1989). Besides, the  $^1\text{H}$  NMR spectrum reveals two meta-coupled aromatic doublets ( $\delta$  5.98,  $\delta$  6.20, each 1H). The HMQC data suggests that these hydrogens are connected to carbons at  $\delta$  99.3 (C-4) and  $\delta$  96.4 (C-6), indicating an asymmetric substituted phloroglucinol moiety. Therefore, the sugar residue must be attached to C-1 of the phloroglucinol, which was supported by the upfield resonances of C-4 and C-6. The sugar moiety was identified by comparison of its  $^1\text{H}$  NMR, HMQC and HMBC data with those of known phloroglucinol glucosides (Kosasi et al., 1989; Tsukamoto et al., 2004; Bohr et al., 2005).

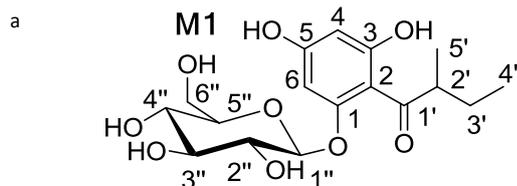
## Results



**Figure 5.** Extracted ion chromatogram ( $m/z$  357 and 371, superimposed), MS and MS2 spectra of metabolites M1, M2 and M3.

**Table 3**  $^1\text{H}$ - and  $^{13}\text{C}$ -NMR data derived from HSQC and HMBC data of isolated metabolite M1 and data from literature<sup>a</sup>.

Atom	$\delta\text{H}$	Splitting	Bohr et al., 2005	$\delta\text{C}$	HMBC	Bohr et al., 2005
1				162.9		161.8
2				107.0		106.8
3				167.3		167.4
4	5.98	D	5.94 d 1H	99.3	C-6,C-5,C-3	98.3
5				165.6		165.6
6	6.20	D	6.17 d 1H	96.4	C-2,C-1,C-5	95.3
1'				211.8	C-2',C-3';C-5'	211.8
2'	3.93	M	3.90 m 1H	47.0	C-3',C-4;C-5'	47.0
3'	1.81, 1.40	M	1.80, 1.38 m 1H	28.3	C-4;C-5'	28.3
4'	0.90	T	0.87 t 3H	12.0	C-2',C-3;C-5'	12.0
5'	1.15	D	1.12 d 3H	16.8		16.8
1''	5.01	D	5.03 d 1H	101.7	C-2''	101.7
2''-5''	3.5-4.0	M	3.38 - 3.91 m	70-78		71.2 - 78.7
6''	3.70	Dd	3.71	62.5		62.5

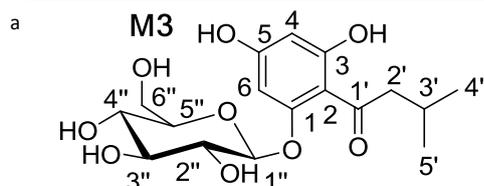


M1 had been previously reported as “multifidol glucoside” from the latex of *Jatropha multifida* L (Kosasi et al., 1989) and hop (*Humulus lupulus*; Bohr et al., 2005). The structure of M1 was confirmed by comparison of its LC retention time, MS data and high-resolution MS data with those of M1 isolated from hop.

The  $^1\text{H}$ -NMR spectrum of M3 differed only in the signals of the acyl side chain from M1 (Figure 6; Table 4) and was kindly provided by Ludwig Ring (Biotechnology of Natural Products, TUM). The LC-MS data of M3 of  $m/z$  371  $[\text{M}-\text{H}]^-$  supported the structure of 1-[(3-methylbutyryl)-phloroglucinyl]-2-*O*- $\beta$ -D-glucopyranoside, which has earlier been found in *F. x ananassa* cv. Tochiotome (Tsukamoto et al., 2004), hop (*Humulus lupulus*, Bohr et al., 2005), *Indigofera hetrantha* (Aziz-ur-Rehman et al., 2005), and roots of *Lysidice rhodostega* (Gao et al., 2004a,b).

**Table 4.**  $^1\text{H-NMR}$  data of isolated metabolite M3 and data from literature<sup>a</sup>.

M3				
H-Atom	Chemical shift $\delta\text{H}$	Splitting	Bohr et al., 2005	Tsukamoto et al., 2004
4	5.98	1H d	5.94 d 1H	5.93 d 1H
6	6.20	1H d	6.17 d 1H	6.15 d 1H
2'	3.14	2H 2dd	2.88 dd 1H, 3.17 dd 1H	2.87 dd 1H, 3.16 dd 1H
3'	1.82	1H sept	2.24 sept 1H	2.24 sept 1H
4' u.5'	0.91	6H dd	0.93 d 3H, 0.96 d 3H	0.92 d 3H, 0.96 d 3H
1''	5.05	1H d	5.01 d 1H	5.01 d 1H
2''-6''	3.5 - 4.0	M	3.39 - 3.91 m	3.39 - 3.91 m



Similarly, M2 was identified as 1-[(2-methylpropanoyl)-phlorogluciny]-2-O- $\beta$ -D-glucopyranoside by comparison of its LC-MS and high-resolution MS data with those of authentic reference material (Intelmann et al., 2011; Bohr et al., 2005). Besides, the metabolites of *F. x ananassa* cv. *Senga Sengana* were analyzed to identify additional phloroglucinol derivatives in strawberry fruit.

**Table 5.** Relative concentration of APG glucosides in in strawberry fruit (*Fragaria* sp) of different varieties.

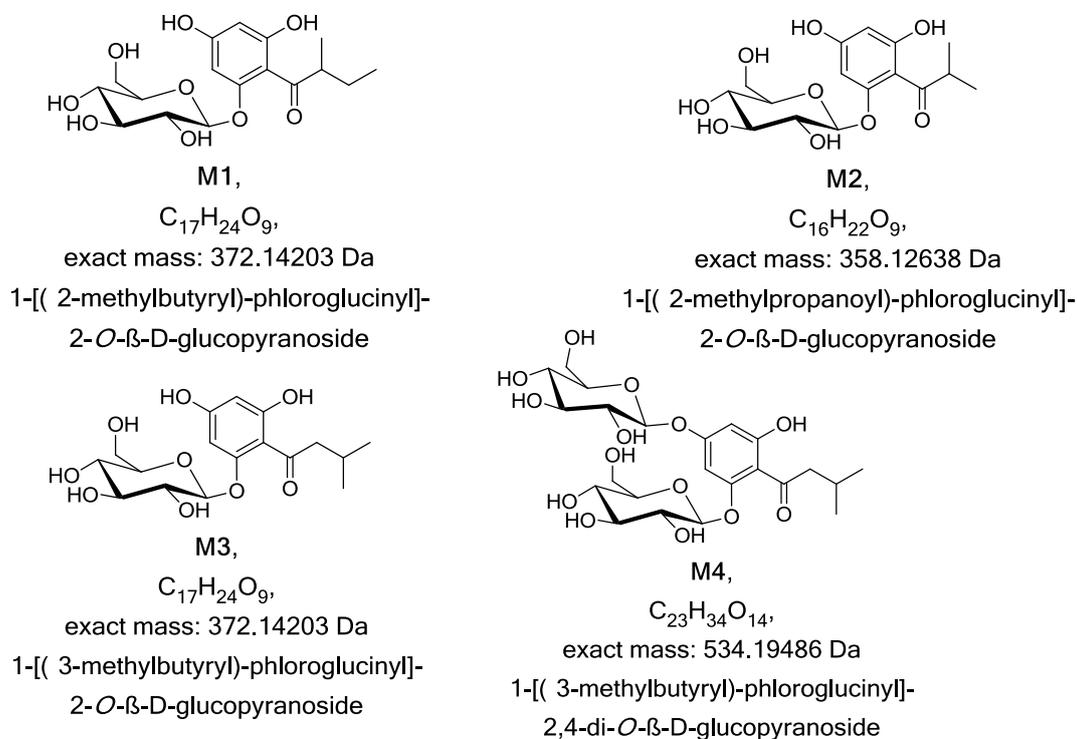
Compound	Elsanta <sup>a</sup> (Red)	Senga Sengana (Red)	Mara des Bois (Red)	Calypso (Red)	HW4 (White)
M1	0.134 $\pm$ 0.052 <sup>b</sup>	0.094 $\pm$ 0.021	0.036 $\pm$ 0.007	0.220 $\pm$ 0.103	N.D
M2	0.084 $\pm$ 0.023	0.379 $\pm$ 0.023	0.233 $\pm$ 0.023	0.766 $\pm$ 0.296	N.D
M3	0.287 $\pm$ 0.081	0.208 $\pm$ 0.085	0.018 $\pm$ 0.006	0.332 $\pm$ 0.111	N.D
M4	0.002 $\pm$ 0.001	0.003 $\pm$ 0.001	N.D	N.D	N.D

<sup>a</sup> Elsanta, Senga Sengana, Mara des Bois and Calypso are cultivars of the octoploid hybrid *F. x ananassa* whereas HW4 (Hawaii 4) is a white-fruited variety of diploid *F. vesca*.

<sup>b</sup> mg-equ. g<sup>-1</sup> lyophilized strawberry fruit powder (mean  $\pm$  Std). N.D not detected by LC-MS, internal standard biochanin A.

The study yielded 1-[(3-methylbutyryl)-phlorogluciny]-2,4-di- $\beta$ -D-glucopyranoside ( $m/z$  533 [M-H]<sup>-</sup>, MS2  $m/z$  323; Gao et al., 2004a, b) as novel strawberry fruit metabolite (M4; Figure 6) by comparing its retention time and mass spectra (MS and MS2) and high resolution mass spectra with those of authentic compounds (Intelmann et al., 2011). Metabolites M1 – M4

were detected in fruit of *F. x ananassa* cv. Elsanta and Senga Sengana, and M1 –M3 in cv. Mara des Bois and Calypso. However, the white-fruited diploid *F. vesca* var. Hawaii 4 (HW4) genotype lacks the APG metabolites (Table 5).



**Figure 6.** Structures of APGs identified in strawberry fruit (*F. x ananassa* cv. Elsanta, Senga Sengana, Mara des Bois, and Calypso). M1, 1-[(2-methylbutyryl)-phloroglucinyl]-2-*O*- $\beta$ -D-glucopyranoside, M2 1-[(2-methylpropanoyl)-phloroglucinyl]-2-*O*- $\beta$ -D-glucopyranoside, M3 1-[(3-methylbutyryl)-phloroglucinyl]-2-*O*- $\beta$ -D-glucopyranoside, M4 1-[(3-methylbutyryl)-phloro-glucinyl]-2,4-di-*O*- $\beta$ -D-glucopyranoside.

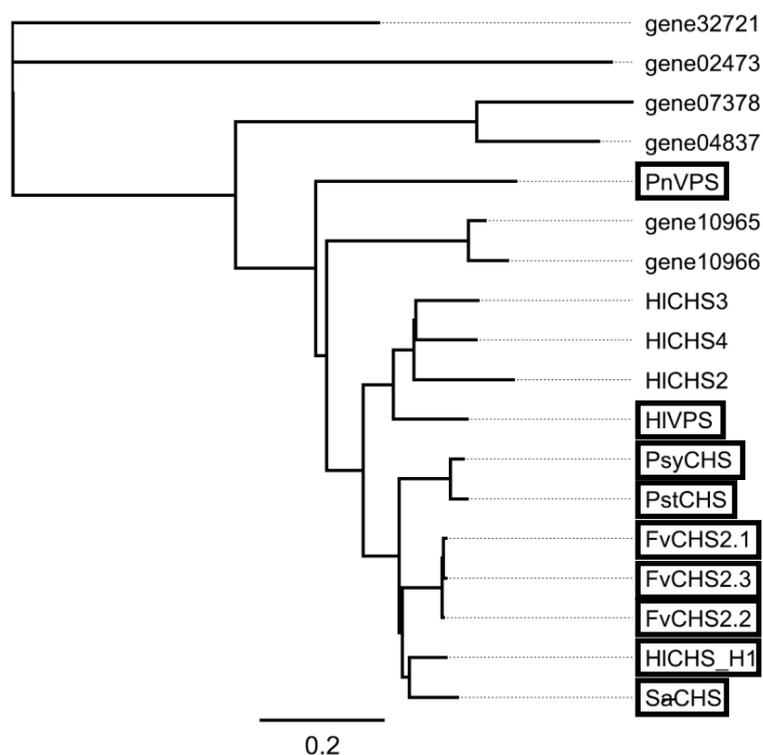
### 3.1.5 Chalcone synthase genes from *Fragaria vesca*

Untargeted metabolite profiling analysis (Table 2) showed that the level of M2 (*m/z* 357) was decreased in response to the down-regulation/overexpression of all five candidate genes, whereas the level of M1 and 3 declined due to silencing of *CHS*. It has been demonstrated that the aglycones of M1-M4 are formed by (phloroiso)valerophenone synthase (*VPS*, Okada and Ito, 2001) in hop (*Humulus lupulus*). However, a *VPS* gene has not been annotated in the *F. vesca* genome sequence (Shulaev et al., 2011). Since the basic catalytic mechanisms of *VPS* and *CHS* are similar we presumed that a dual function *CHS* may also act as *VPS* in strawberry fruit. To verify the hypothesis we cloned *CHS* genes from *F. vesca* to characterize the enzymatic activities of the encoded proteins. Primers were

designed based on the cDNA sequences of gene26825 and gene26826 from *F. vesca* (Appendix Table 1) both are annotated as *CHS2-like*. They are the closed homologues of *FaCHS* which is highly expressed in *F. x ananassa* fruit (Almeida et al., 2007; Lunkenbein et al., 2006b). It is assumed that the diploid woodland strawberry *F. vesca* and the cultivated octoploid *F. x ananassa* share a common ancestor (Shulaev et al., 2011). Eight putative *CHS* genes have been annotated in the *F. vesca* genome sequence, whereas two of them seem to be truncated. In accordance with data from *F. x ananassa* (Lunkenbein et al., 2006b), RNAseq data of *F. vesca* showed that gene26825 and gene26826 are the only *CHS* genes expressed in strawberry fruit tissue (Kang et al., 2013; Appendix Figure 1).

Three full-length *CHS* sequences (*FvCHS2.1*, *2.2*, and *2.3*) were obtained from *F. vesca* which all consist of 1170-bp and encode proteins of 389 amino acids. The molecular masses of *FvCHS2.1*, *2.2* and *2.3* are predicted to be 42.63, 42.65, and 42.66 kDa, respectively. Alignment of the three deduced protein sequences revealed that *FvCHS2.1* is identical with gene26835 and *FvCHS2.2* with gene26836 whereas *CHS2.3* bears the N-terminal sequence of *CHS2.1* and the C-terminus of *CHS2.2* (Appendix Figure 2). The genomic DNA of *FvCHS2.1* and *2.2* contains a 213-bp intron at 180 bp from the start codon. The deduced amino acid sequence of *FvCHS2.3* shows 97% identity to each of gene26825 and gene26826. The strictly conserved CHS active site residues, Cys164, His303, and Asn336 (Ferrer et al., 1999), as well as the highly conserved CHS signature sequence, G372FGPG (Suh et al., 2000) were found in all of the three *FvCHS* proteins (Appendix Figure 2). The two Phe residues (Phe215 and Phe265), important in determining the substrate specificity of CHS (Jez et al, 2002) were also found.

A phylogenetic tree built from amino acid sequences of CHS/VPS enzymes and homologous proteins from *F. vesca* and *H. lupulus* as well as functionally characterized CHS proteins with VPS promiscuous catalytic activity shows that *FvCHS2.1*, *2.2* and *2.3* cluster with H1CHS-H1 from *H. lupulus* and three other proteins (*PsyCHS*, *PstCHS*, *SaCHS*) that perform both the function of CHS and VPS, reflecting their close evolutionary relationship (Figure 7).



**Figure 7.** Phylogenetic tree of CHS/VPS enzymes and homologues from *F. vesca* and *Humulus lupulus* as well as functionally characterized CHS proteins with VPS promiscuous catalytic activity, generated by the Geneious (Pro 5.5.4) Tree Builder (Jukes Cantor genetic distance model and neighbor-joining method). The scale bar indicates the average number of amino acid substitutions per site. FvCHS2.1 *F. vesca* chalcone synthase 2.1 (XM\_004306495.1; gene 26825), FvCHS2.2 *F. vesca* chalcone synthase 2.2 (XM\_004306494.1; gene26826), FvCHS2.3 *F. vesca* chalcone synthase 2.3 (chimera of 2.1 and 2.2), PsyCHS *Pinus sylvestris* chalcone synthase (P30079), PstCHS *Pinus strobus* chalcone synthase (O65872), SaCHS *Sinapis alba* chalcone synthase (P13416), PnVPS *Psilotum nudum* valerophenone synthase (Q9SLX9), HICHS2-4 *Humulus lupulus* chalcone synthase 2-4 (AB061020, AB061022, CAD23044), HIVPS *Humulus lupulus* valerophenone synthase (O80400), HICHS\_H1 *Humulus lupulus* chalcone synthase\_H1 (CAC19808), gene02473, 04837, 07378, 10965, 10966, and 32721 are translated protein sequences of putative *F. vesca* chalcone synthases (Shulaev et al., 2011). Proteins with biochemically verified VPS activity are boxed.

### 3.1.6 Catalytic activity of putative FvCHS enzymes

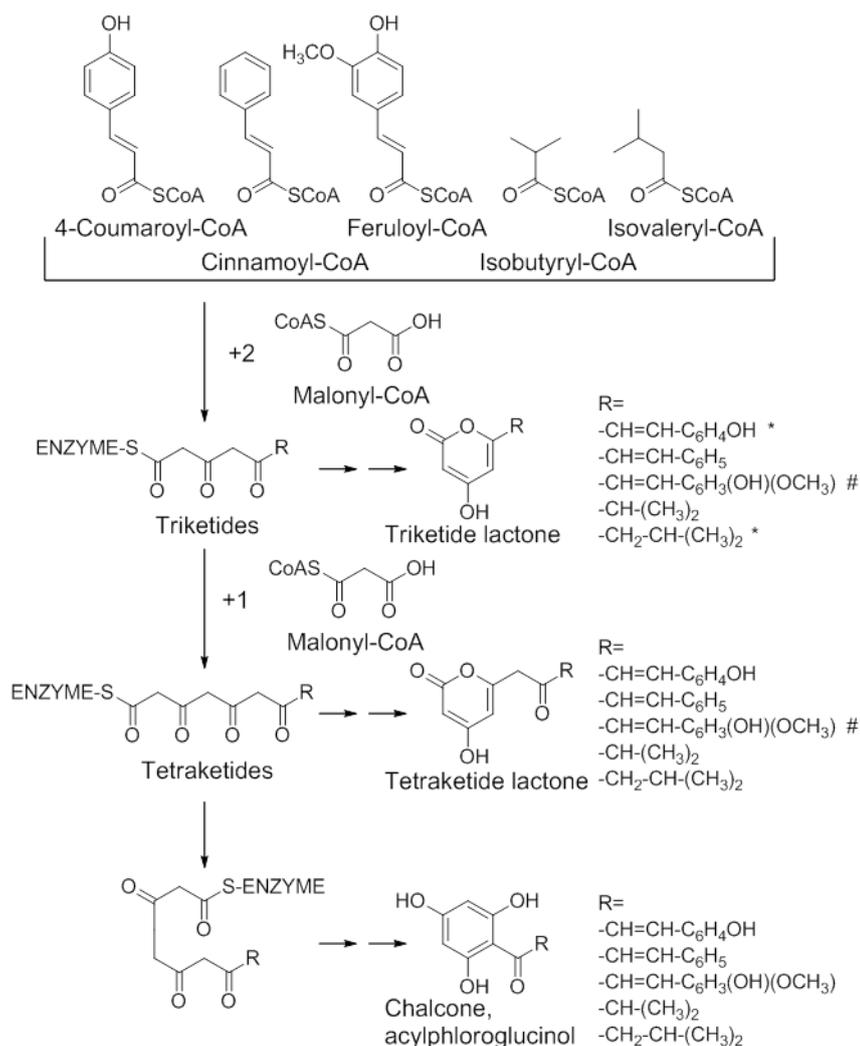
*FvCHS2.1*, *2.2* and *2.3* genes were introduced into *E. coli* strain BL21 (DE3) pLysS cells and the recombinant proteins were expressed and affinity purified. The enzyme assays were carried out using malonyl-CoA and the starter substrates isovaleryl-CoA, isobutyryl-CoA, 4-coumaroyl-CoA, feruloyl-CoA, caffeoyl-CoA, and cinnamoyl-CoA under standard conditions. Analysis of the products was performed by LC-MS (Yamazaki et al., 2001; Akiyama et al., 1999; Jez et al., 2001). In the presence of 4-coumaroyl-CoA, all three enzymes FvCHS2.1, 2.2 and 2.3 produced 4-coumaroyl tetraketide lactone (the derailed lactone after three condensations) and naringenin-chalcone (Table 6; Figure 8).

**Table 6.** LC-MS analysis of products formed by CHS2.1-2.3 and obtained product ratio using negative electrospray ionization mode (ESI-).

Starter CoA	Product <sup>a</sup>	Elemental formula	MS ( <i>m/z</i> )	MS/MS ( <i>m/z</i> )	product ratio [%]		
					CHS2.1	CHS2.2	CHS2.3
Isovaleryl-CoA	APG PIVP <sup>b</sup>	C <sub>11</sub> H <sub>14</sub> O <sub>4</sub>	[M-H] <sup>-</sup> 209	165,125	40-45	70-80	10-30
	Tetraketide lactone	C <sub>11</sub> H <sub>14</sub> O <sub>4</sub>	[M-H] <sup>-</sup> 209	165,125	25-30	10-15	<10
	Triketide lactone	C <sub>9</sub> H <sub>12</sub> O <sub>3</sub>	[M-H] <sup>-</sup> 167	123	30-35	15-20	65-85
Isobutyryl-CoA	APG PIBP	C <sub>10</sub> H <sub>12</sub> O <sub>4</sub>	[M-H] <sup>-</sup> 195	151,131	40-45	70-80	<10
	Tetraketide lactone	C <sub>10</sub> H <sub>12</sub> O <sub>4</sub>	[M-H] <sup>-</sup> 195	151,125	40-45	20-30	90-95
	Triketide lactone	C <sub>8</sub> H <sub>10</sub> O <sub>3</sub>	[M-H] <sup>-</sup> 153	109	10-15	<10	<10
4-Coumaroyl-CoA	Naringenin	C <sub>15</sub> H <sub>12</sub> O <sub>5</sub>	[M-H] <sup>-</sup> 271	177,151,107	25-30	70-75	80-90
	Tetraketide lactone	C <sub>15</sub> H <sub>12</sub> O <sub>5</sub>	[M-H] <sup>-</sup> 271	227,201,125	70-75	25-30	15-20
Cinnamoyl-CoA	Chalcone	C <sub>15</sub> H <sub>12</sub> O <sub>4</sub>	[M-H] <sup>-</sup> 255	211,151	80-85	80-90	10-20
	Tetraketide lactone	C <sub>15</sub> H <sub>12</sub> O <sub>4</sub>	[M-H] <sup>-</sup> 255	211,187	5-15	10	10-20
	Triketide lactone	C <sub>13</sub> H <sub>10</sub> O <sub>3</sub>	[M-H] <sup>-</sup> 213	169	<10	<10	70-75
Feruloyl-CoA	Triketide lactone	C <sub>16</sub> H <sub>14</sub> O <sub>6</sub>	[M-H] <sup>-</sup> 259	215	100	n.d	100

<sup>a</sup> Chemical structures are shown in Figure 8

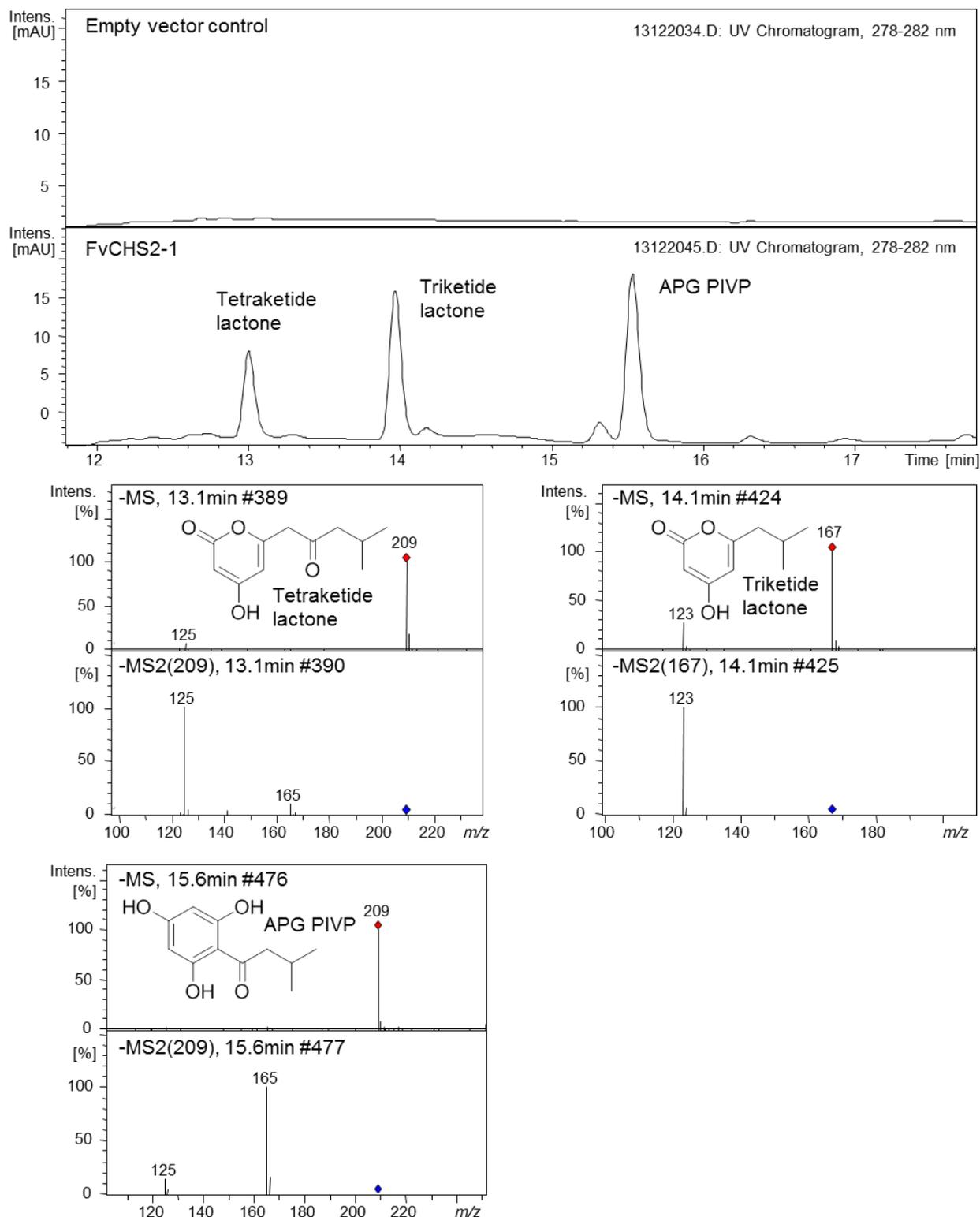
<sup>b</sup> APG acylphloroglucinol, PIVP phloroisovalerophenone, PIBP phloroisobutyrophenone



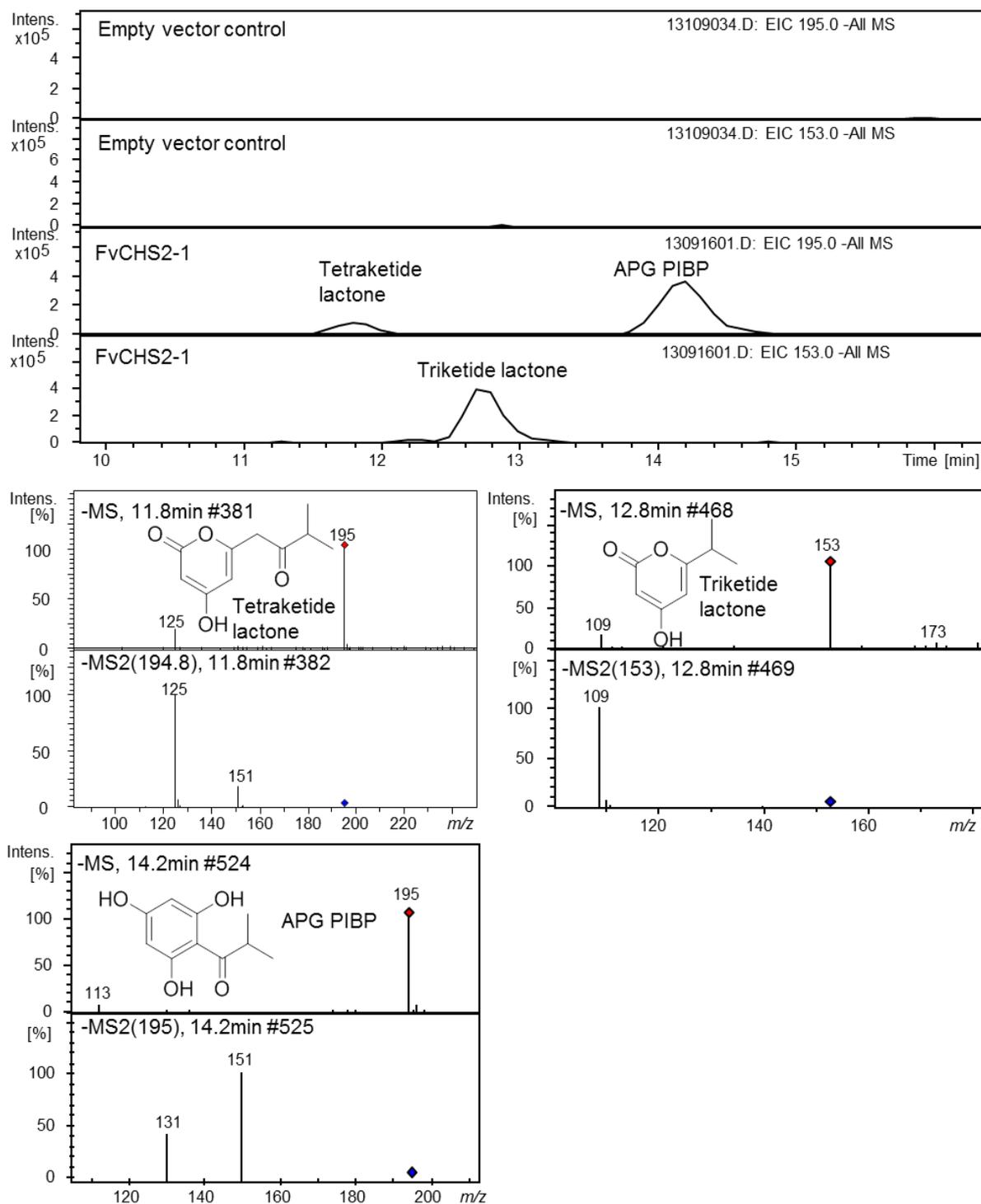
**Figure 8.** Proposed formation mechanism and *in vitro* products formed by FvCHS2.1-2.3 when 4-coumaroyl-CoA, cinnamoyl-CoA, feruloyl-CoA, isobutyryl-CoA and isovaleryl-CoA were used as starter molecules. # Product was not detected by LC-MS; \* product represents less than 10% of total products.

The predominant product was the chalcone, except for CHS2.1, indicating that FvCHS2.1 - 2.3 can perfectly perform the function of CHS. When isovaleryl-CoA and isobutyryl-CoA were used as starter substrates instead of 4-coumaroyl-CoA, FvCHS2.1 - 2.3 produced the APGs PIVP and PIBP, respectively, in addition to a considerable portion of prematurely terminated products (Table 6; Figure 8-10).

## Results

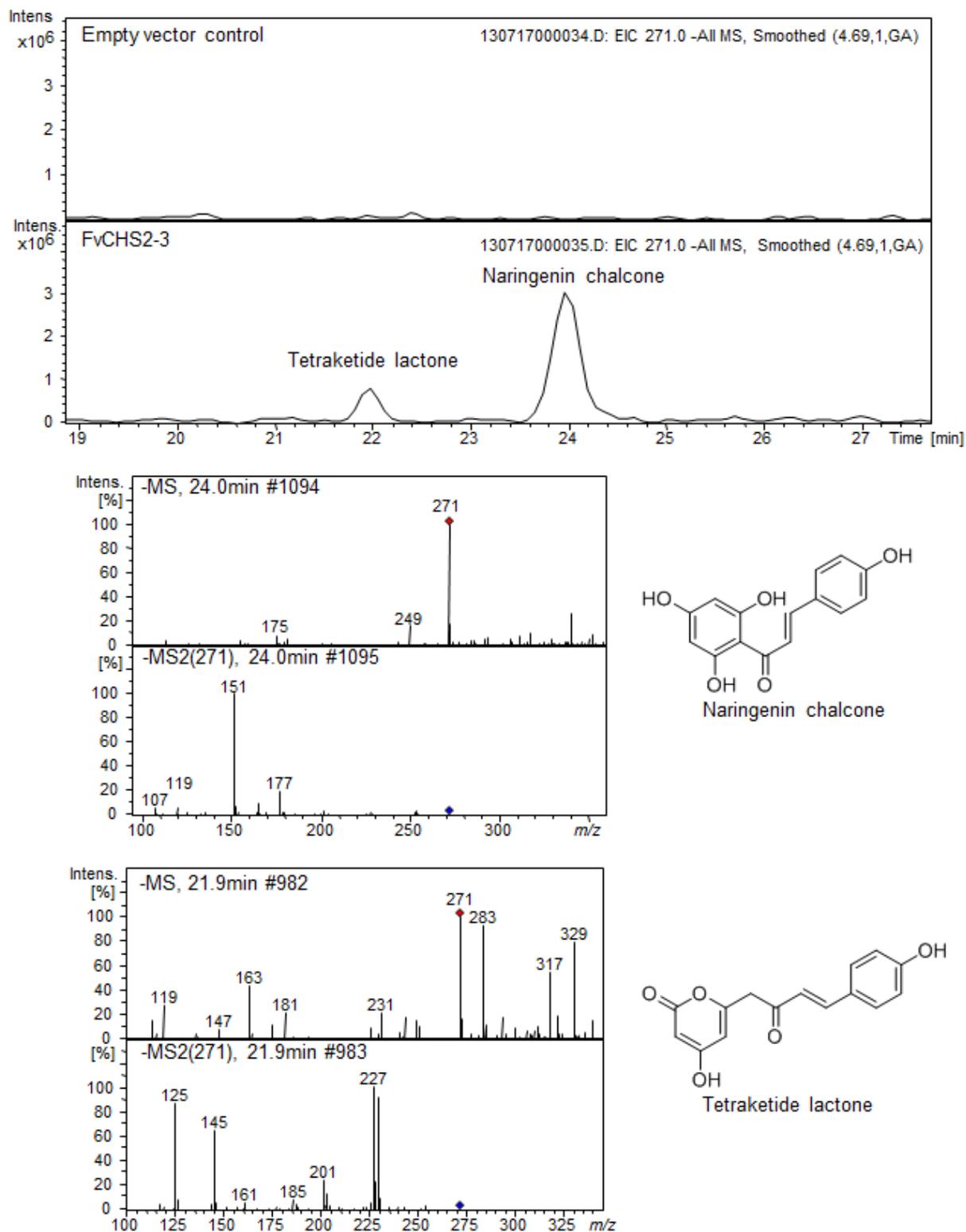


**Figure 9.** LC-UV-MS analysis of products formed by the empty vector (UV trace) control and FvCHS2-1 from the starter molecule isovaleryl-CoA (UV trace). MS and MS2 spectra of tetraketide lactone, triketide lactone and APG (acylphloroglucinol). PIVP phlorisovalerophenone.

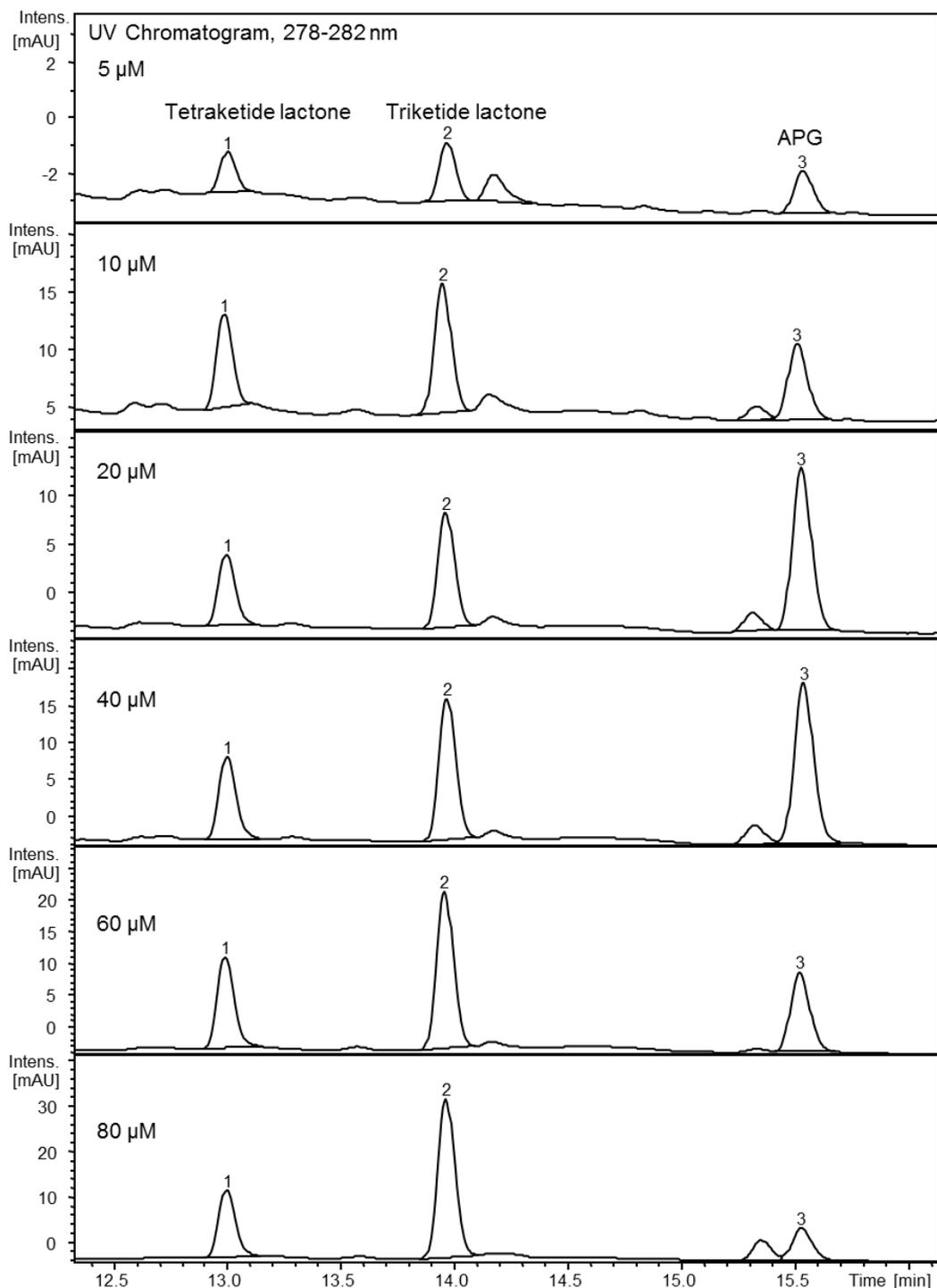


**Figure 10.** LC-MS analysis of products formed by the empty vector control and FvCHS2-1 from the starter molecule isobutyryl-CoA (extracted ion chromatograms, EIC). MS and MS2 spectra of tetraketide lactone, triketide lactone and APG (acylphloroglucinol). PIBP phlorisobutyrophenone.

## Results



**Figure 11.** LC-MS analysis of products formed by the empty vector control and FvCHS2-3 from the starter molecule 4-coumaroyl-CoA (extracted ion chromatograms, EIC). MS and MS2 spectra of naringenin chalcone and tetraketide lactone.



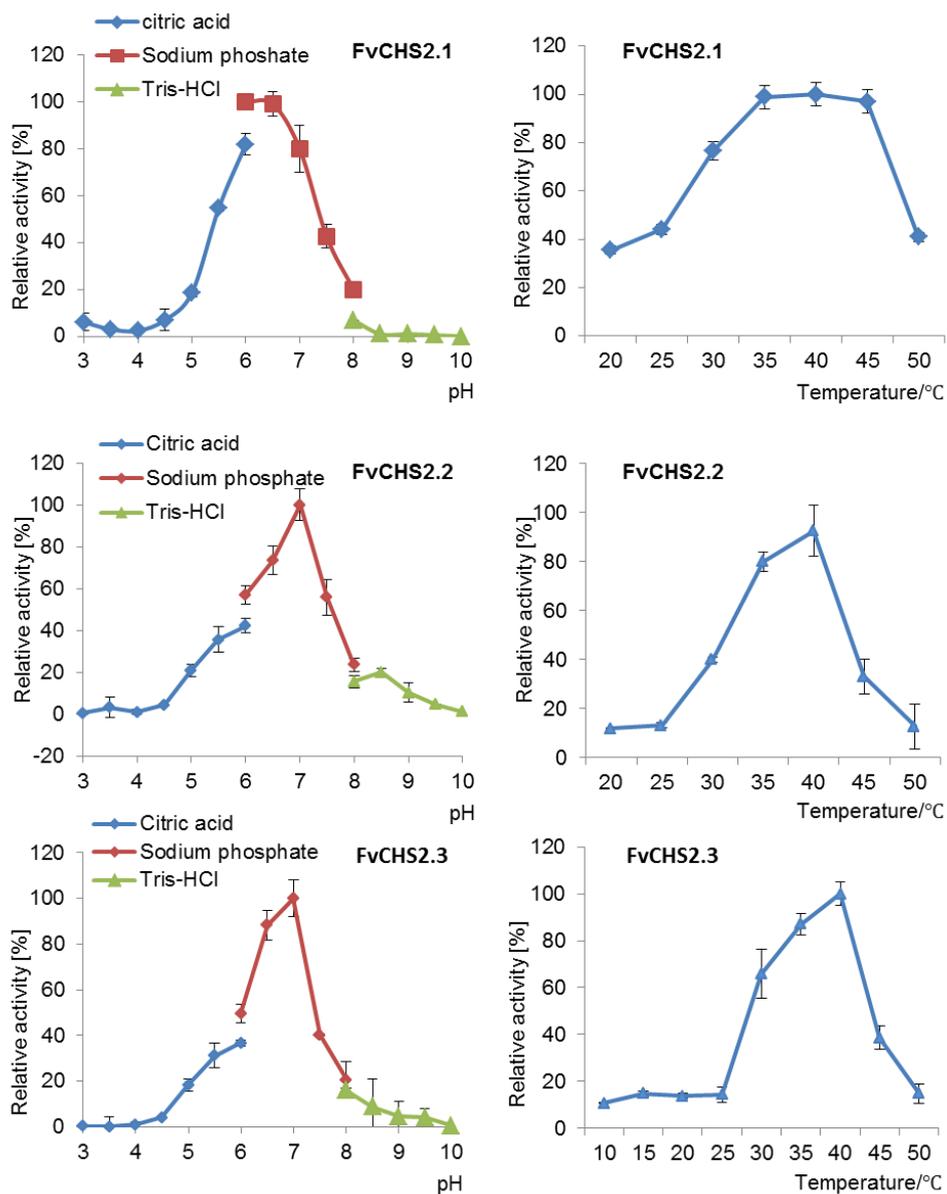
**Figure 12.** Products (tetraketide lactone 1, triketide lactone 2 and APG 3; signals at 14.3 and 15.4 min are unknowns) formed by FvCHS2.1 using 300  $\mu\text{M}$  malonyl-CoA and different concentration of isovaleryl-CoA (20, 40, 60, 80, and 100  $\mu\text{M}$ ). The ratio of APG 3 is highest when the concentration of isovaleryl-CoA is 20 - 40  $\mu\text{M}$ .

The product profiles formed by FvCHS2.1, 2.2 and 2.3 in the presence of the starter CoAs isovaleryl-CoA and isobutyryl-CoA were virtually identical to those of recombinant VPS from

hop (*H. lupulus* L) and PnP (*Psilotum nudum* phloroisovalerophenone synthase) (Okada et al., 2004; Yamazaki et al., 2001). The most abundant products were identified as triketide and tetraketide lactone when CHS2.3 was used in the assay, but the APGs PIVP and PIBP were the major products formed by CHS2.1 and CHS2.2 (Table 6). In contrast, CHS2.3 formed mainly naringenin chalcone from 4-coumaroyl-CoA (Figure 11). These properties indicate that FvCHS enzymes, especially CHS2.1 and CHS2.2 act as VPS *in vitro*. Besides, the relative ratio of isovaleryl-CoA/malonyl-CoA and isobutyryl-CoA/malonyl-CoA affects the composition of the products and thus the PIVP and PIBP content (Figure 12). In addition to isovaleryl-CoA, isobutyryl-CoA, and 4-coumaroyl-CoA, FvCHS2.1 and 2.3 also accept feruloyl-CoA as starter molecule but form only the triketide lactone (Jez et al., 2001; Appendix Figure 3; Table 6). With cinnamoyl-CoA, all three enzymes yield varying amounts of the tetraketide and triketide lactone and the chalcone (Appendix Figure 4; Table 6). No products were detected when caffeoyl-CoA was used as starter substrate.

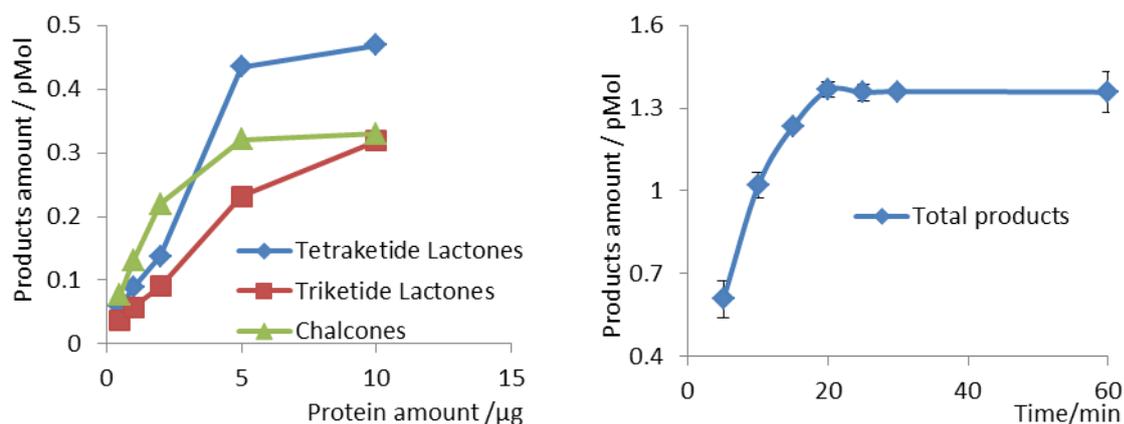
### 3.1.7 Kinetic properties of FvCHS enzymes and starter-CoA preference

Activity of the FvCHS enzymes was determined at pH values from pH 3-10. The highest activity of FvCHS2.1, 2.2 and 2.3 was detected at optimum pH 6.0, 7.0 and 7.0, respectively (Figure 13). Temperature optimum of the enzymes was analyzed in the range of 10 to 50 °C (Figure 13). The optimal temperature was 40°C for all enzymes using either 4-coumaroyl-CoA or isovaleryl-CoA as the starter substrate. Kinetic properties of the FvCHS enzymes were determined in the linear range of the enzymatic reaction (Figure 14). Two µg of protein, 10 min reaction time at 40°C, optimal pH value, 300 µM malonyl-CoA and various concentrations of the starter substrates were chosen.



**Figure 13.** The pH and temperature optima. (a-d) The purified enzymes (FvCHS2.1 – 2.3) were incubated with 100  $\mu$ M isovaleryl-CoA at 30 °C and different pH values (left column) and at pH 7.0 and different temperature values (right column). Citric acid buffer was used for pH 3 to 6, sodium phosphate buffer was used for pH 6 to 8, Tris-HCl buffer was used for pH 8 to pH 10. Temperature optima were determined in the range of 10 to 50 °C and were carried out at pH 7.0 for 10 min. Product formation was quantified by LC–MS analysis allowing the calculation of Michaelis–Menten equation by hyperbolic regression and quantified as activity ( $\text{pmol mg}^{-1} \text{s}^{-1}$ ).

## Results



**Figure 14.** The effect of different amounts of protein (0.5-10 µg) and incubation time (5 – 60 min) on the product formation of FvCHS2.3 using cinnamoyl-CoA as the starter substrate. Naringenin (10 µg/mL) was used as an internal standard for quantification by LC-MS.

**Table 7.** Steady-state kinetic constants for FvCHS2.1, FvCHS2.2, FvCHS2.3 of *F. vesca* with different starter substrates. Values shown are means (n = 3).

Enzyme	Starter substrate	$K_M$ [µM]	$k_{cat}$ [min <sup>-1</sup> ]	$k_{cat}/K_M$ [M <sup>-1</sup> s <sup>-1</sup> ]
CHS2.1	Isovaleryl-CoA	14.7± 2.3	19.1±1.5	21700
	Isobutyryl-CoA	86.6± 11.6	4.4±0.5	847
	4-coumaroyl-CoA	6.4±1.7	0.8±0.1	2021
CHS2.2	Isovaleryl-CoA	19.7± 6.3	5.1±0.5	4355
	Isobutyryl-CoA	88.6± 14.5	1.0±0.1	179
	4-coumaroyl-CoA	10.4±3.8	1.3±0.2	2045
CHS2.3	Isovaleryl-CoA	18.0 ± 2.7	5.8 ± 0.2	5431
	Isobutyryl-CoA	62.3 ± 13.4	1.2 ± 0.1	313
	4-coumaroyl-CoA	10.2 ± 2.5	2.42 ± 0.12	3957
	Cinnamoyl-CoA	77.5 ± 17.2	4.6± 0.5	981
	Feruloyl-CoA	16.1 ± 3.3	2.3± 0.1	2364

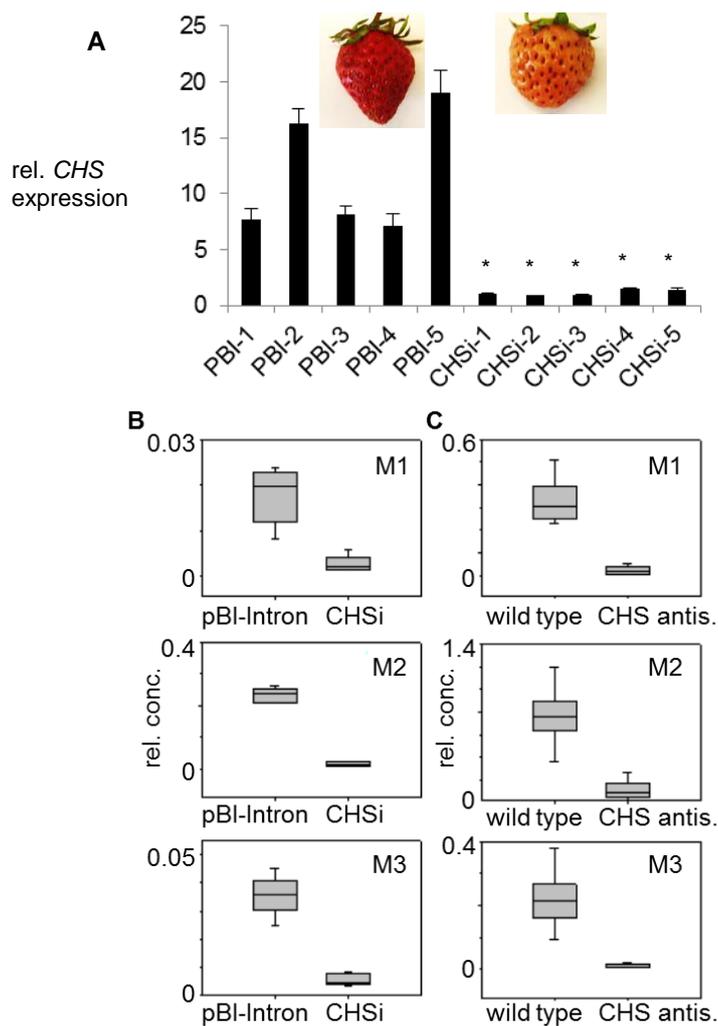
Isovaleryl-CoA is the preferred substrate of FvCHS2.1 ( $k_{cat}/K_M$  21700 M<sup>-1</sup> s<sup>-1</sup>) whereas a 5-fold and 4-fold lower  $k_{cat}/K_M$  value was calculated for this starter substrate in the case of FvCHS2.2 and FvCHS2.3, respectively (Table 7). The apparent  $K_M$  values for 4-coumaroyl-CoA of FvCHS2.1, 2.2 and 2.3 were 6.35, 10.35, and 10.21 µM respectively. These values are the minimum  $K_M$  values among the three starter molecule of all three enzymes. The  $k_{cat}$  value for isobutyryl-CoA of FvCHS2.1 is 5-fold and 4-fold higher than the corresponding values of FvCHS2.2 and 2.3, respectively yielding a maximum specificity constant  $k_{cat}/K_M$  value for FvCHS2.1 of 847 M<sup>-1</sup> s<sup>-1</sup> (Table 7).

### 3.1.8 Downregulation of FvCHS enzymatic activity in a transient system and a stable transgenic line

To further corroborate the hypothesis that *FvCHS* genes play a role in the biosynthesis of APGs in strawberry, metabolite profiling analysis of transiently silenced *CHS2* fruits (*F. x ananassa* cv Mara des Bois) and fruits of a stable transgenic *FaCHS* antisense line (*F. x ananassa* cv Calypso; Lunkenbein et al., 2006b) was performed. A suspension of *Agrobacterium tumefaciens* harboring the pBI-*CHSi* construct was injected in green strawberry fruits about 14 days after pollination. The fruits remained attached to the plants. Levels of metabolites were determined by LC-MS 14 days after infiltration in *CHS2*-silenced fruits and in fruits infiltrated with *Agrobacterium* containing a pBI-*GUS* control vector (Hoffmann et al., 2006).

The relative *CHS* mRNA expression levels (*FvCHS2.1-2.3*) were significantly suppressed in *CHS*-silenced fruits compared with levels in the pBI-Intron-*GUS* controls (Figure 15A). Fruits infiltrated with the control vector turned completely red like the wild type fruits however, agroinfiltration of pBI-*CHSi* resulted in slightly red colored (orange) strawberry fruits, a clear sign of impaired anthocyanin accumulation (Figure 15A). Metabolite analyses revealed that *CHS2*-silenced receptacles not only produced significantly lower levels of anthocyanins (Hoffmann et al., 2006) but also of M1 (10% of control), M2 (8% of control) and M3 (14% of control) when compared with the levels in pBI-Intron-*GUS* controls (Figure 15B). M4 was not detected in fruit of *F. x ananassa* cv. Mara de Bois (Table 5).

## Results

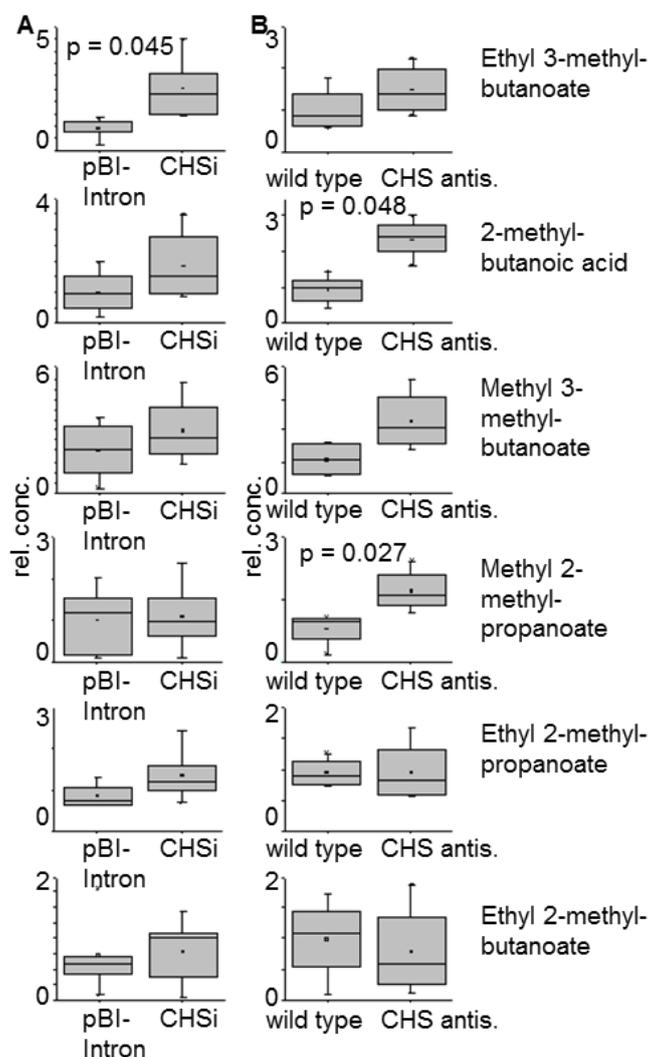


**Figure 15.** Strawberry fruit phenotypes and *CHS* gene expression levels in agroinfiltrated fruits (**A**) and effect of *CHS* gene downregulation on APGs (M1 - M3) in transiently *CHS2*-silenced fruits (*CHSi*, *F. xananassa* cv Mara des Bois) (**B**) and a stable *CHS* antisense transgenic line (*CHS antis.*, *F. xananassa* cv Calypso) (**C**). Fourteen days after pollination, green strawberry fruit were infiltrated with *Agrobacterium* transformed with a construct encoding *CHSi*. Levels of metabolites were determined by LC-MS 14 days after infiltration in *CHS2*-silenced fruits (*CHSi*; n =5) and fruits infiltrated with *Agrobacterium* containing the control vector (PBI, pBI-Intron; n =5). Down-regulation of the *CHS* genes results in a decrease of APGs in the *CHSi* fruits and fruits of the stable transgenic line (*CHS antis.*, *F. xananassa* cv Calypso) compared with the control fruits (pBI-Intron and wild type; for each n =5). Relative concentration (rel. conc.) is expressed in mg-equ. internal standard g<sup>-1</sup>. Asterisks indicate significant differences at p=0.01 (\*).

Besides, levels of APG derivatives were also quantified in fruits of a stable transgenic *FaCHS* line (*F. x ananassa* cv Calypso; Lunkenbein et al., 2006b) where the *CHS* transcript level is less than 5% of control fruit and in fruits of a wild type control. Again, concentrations of APGs (M1 – M3) were significantly reduced in response to the down-regulation of the *CHS* function (Figure 15C) which unambiguously confirms the biochemical role of *FaCHS* enzymes in the biosynthesis of APGs in strawberry fruit.

### **3.1.9 Production of volatiles in transiently *FaCHS*-silenced fruits and fruits of a stable transgenic line**

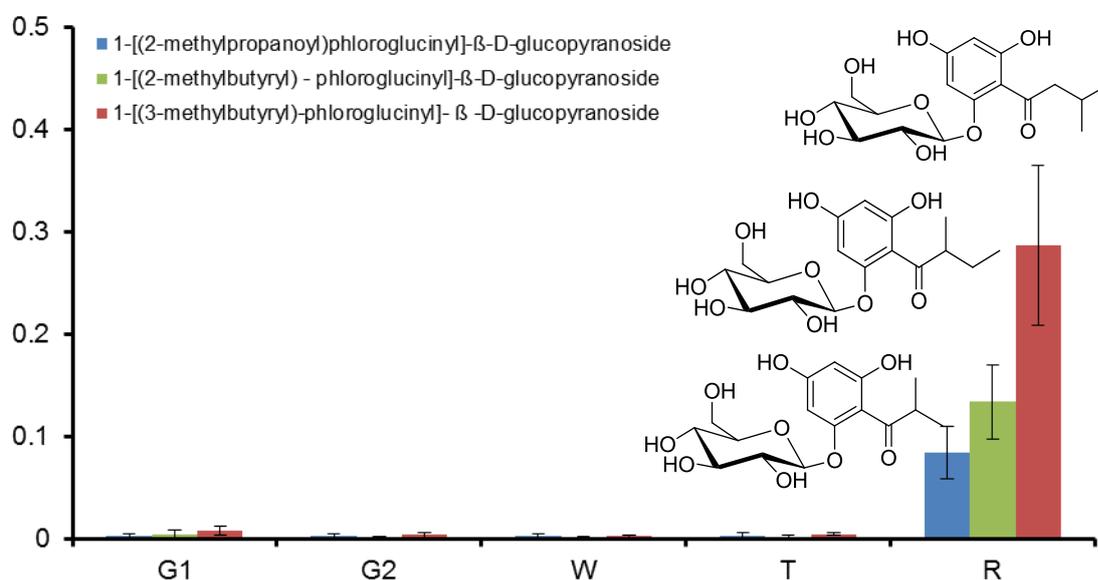
Since isovaleryl-CoA and isobutyryl-CoA, the precursor molecules of APGs, also act as substrates for alcohol acyl-CoA transferase enzymes, that form aroma esters during strawberry fruit ripening (Cumplido-Laso et al., 2012) we analyzed the production of volatile esters in RNAi-mediated *CHS* silenced fruits, and alternatively, in fruits of a stable *CHS* antisense transgenic line by solid phase microextraction GC-MS. The relative content of ethyl 3-methylbutanoate was significantly increased ( $p=0.045$ ) by 2.2-fold in fruits after transient silencing of *CHS2*, whereas levels of 2-methylbutanoic acid and methyl 2-methylpropanoate are significantly enhanced by 2.5-fold ( $p=0.048$ ) and 2.1 fold ( $p=0.027$ ), respectively, in fruits of the stable transgenic line in comparison with the values of the wild type fruits (Figure 16). Besides, most of the methyl-branched esters show higher concentrations in fruits in which the *CHS* transcript levels have been down-regulated but the differences were not statistically significant due to the high biological variation of the values, indicated by the sizes of the boxes.



**Figure 16.** Relative concentration of ester in *CHSi* agroinfiltrated fruits (*F. x ananassa* cv. Elsanta) (A) and a stable transgenic *CHS* antisense (*CHS* antis.) line (*F. x ananassa* cv Calypso) (B). Metabolite levels were determined by GC-MS fourteen days after infiltration with *Agrobacterium* transformed with a construct encoding *CHS*-ihpRNA or with pBI-*Intron*. Metabolite levels in fruits of the *CHS* antisense line (Lunkenbein et al., 2006) and wild type *F. x ananassa* cv Calypso fruits were determined at the mature ripening stage. Identity of the compounds was confirmed by authentic references. n=5-7. Relative concentration (rel. conc.) in mg-equ. internal standard kg<sup>-1</sup> is shown.

### 3.1.10 Stable isotope labelling experiments

Finally, the APG pathway was traced by injection of isotopically labeled L-isoleucine- $^{13}\text{C}_6$  (the biogenic precursor of 2-methylbutyryl-CoA), L-leucine- $^{13}\text{C}_6$  (precursor of isovaleryl-CoA), and L-valine-2,3,4,4,4,5,5,5- $\text{D}_8$  (precursor of isobutyryl-CoA) throughout entire strawberry fruits still attached to the plants. Strawberry fruits (*F. x ananassa* cv Mara des Bois) in the turning ripening stage were injected because APGs start to accumulate after this period (Figure 17). Labeled products were determined by LC-MS, one and four days after incubation. Aqueous solutions containing the same amount of the unlabeled substrates were used as controls. The isotopically labeled products M1 (36.5 min, MS  $m/z$  376, MS2  $m/z$  214) and M3 (37.4 min, MS  $m/z$  376, MS2  $m/z$  256 and 214) were identified by comparison of their retention times and molecular weights, as well as the MS2 fragment ions with those of unlabeled M1 (36.5 min, MS  $m/z$  371, MS2  $m/z$  209) and M3 (37.4 min, MS  $m/z$  371, MS2  $m/z$  251 and 209), originating from the natural pool (Figure 18 and 19). Labeled M2 (33.9 min, MS  $m/z$  362) was identified by comparison of the retention time and pseudomolecular ion with those of unlabeled M2 (33.9 min, MS  $m/z$  352) originating from the natural pool.



**Figure 17.** Relative content of M1, M2 and M3 measured by LC-MS ( $n=3-5$ ) and relative *CHS* expression (gene 26825 and gene 26826) determined by quantitative real time PCR during strawberry fruit ripening (small green G1, large green G2, white W, turning T, red R).

**Table 8.** Isotopically labeled APGs detected one day after injection of different concentration of L-isoleucine- $^{13}\text{C}_6$ , L-leucine- $^{13}\text{C}_6$ , and L-valine-2,3,4,4,4,5,5,5- $\text{D}_8$  into strawberry fruits (*F. x ananassa* cv. Mara des Bois).

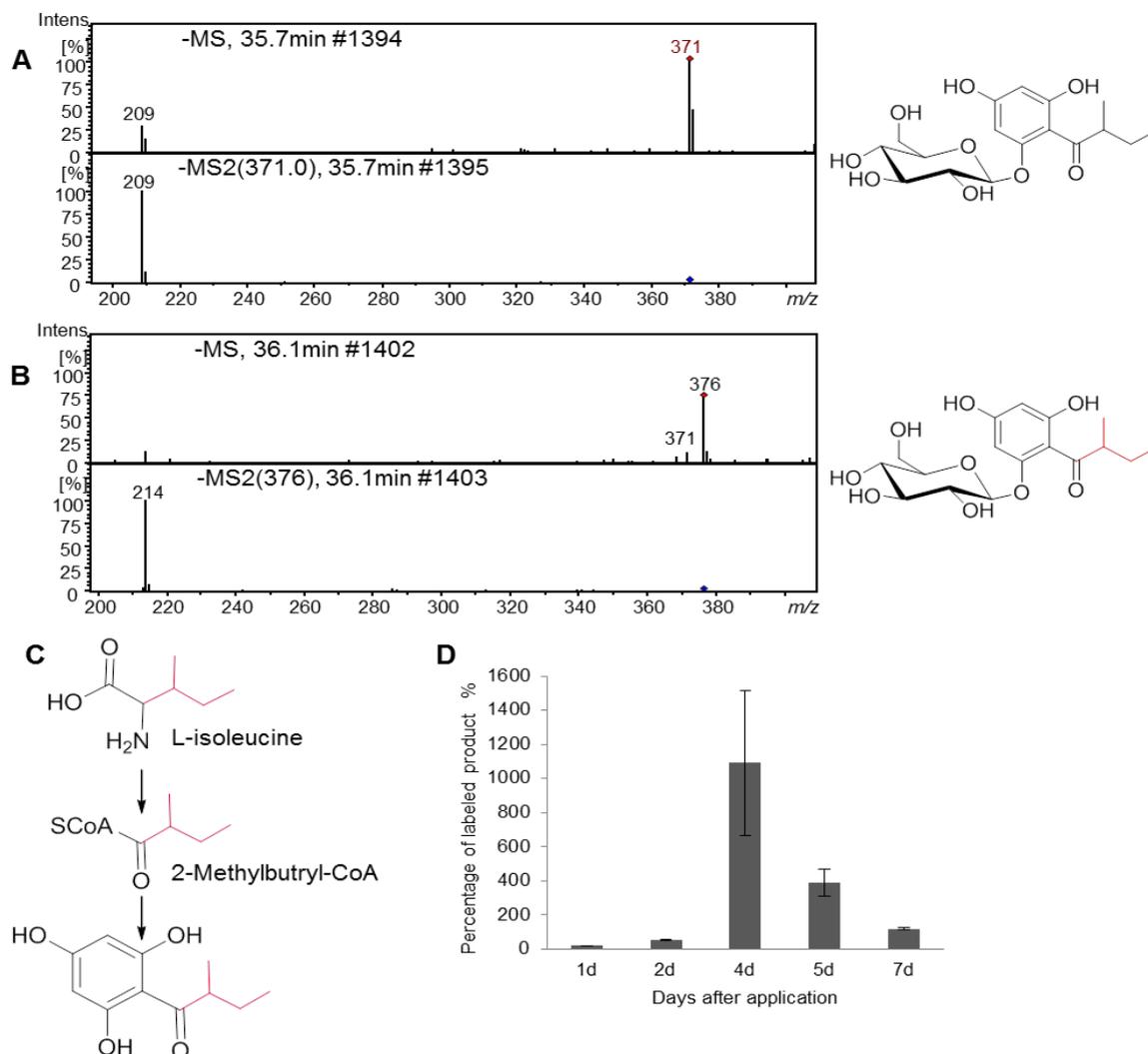
Concentration applied [mM]	Substrates	Compound	Products [%]	
			unlabeled	labeled
10	L-leucine	M3	100 <sup>a</sup>	0 <sup>b</sup>
50	L-leucine	M3	100	0
10	L-leucine- $^{13}\text{C}_6$	M3	100	19.7±1.3
50	L-leucine- $^{13}\text{C}_6$	M3	100	78.7±6.3
10	L-isoleucine	M1	100	0
50	L-isoleucine	M1	100	0
10	L-isoleucine- $^{13}\text{C}_6$	M1	100	15.9±5.3
50	L-isoleucine- $^{13}\text{C}_6$	M1	100	40.4±0.3
10	L-valine	M2	100	0
50	L-valine	M2	100	0
10	L-valine-2,3,4,4,4,5,5,5- $\text{D}_8$	M2	100	22.9±4.7
50	L-valine-2,3,4,4,4,5,5,5- $\text{D}_8$	M2	100	48.0±24.9

<sup>a</sup> integrated peak areas of pseudomolecular ions  $m/z$  371, 371, and 357 (after application of L-leucine, L-isoleucine, and L-valine, respectively) were set as 100%

<sup>b</sup> percentages of labeled products were calculated from the integrated peak areas of the pseudomolecular ions of isotopically labeled products  $m/z$  376, 376, and 364 (after application of L-leucine- $^{13}\text{C}_6$ , L-isoleucine- $^{13}\text{C}_6$ , and L-valine-2,3,4,4,4,5,5,5- $\text{D}_8$ , respectively).

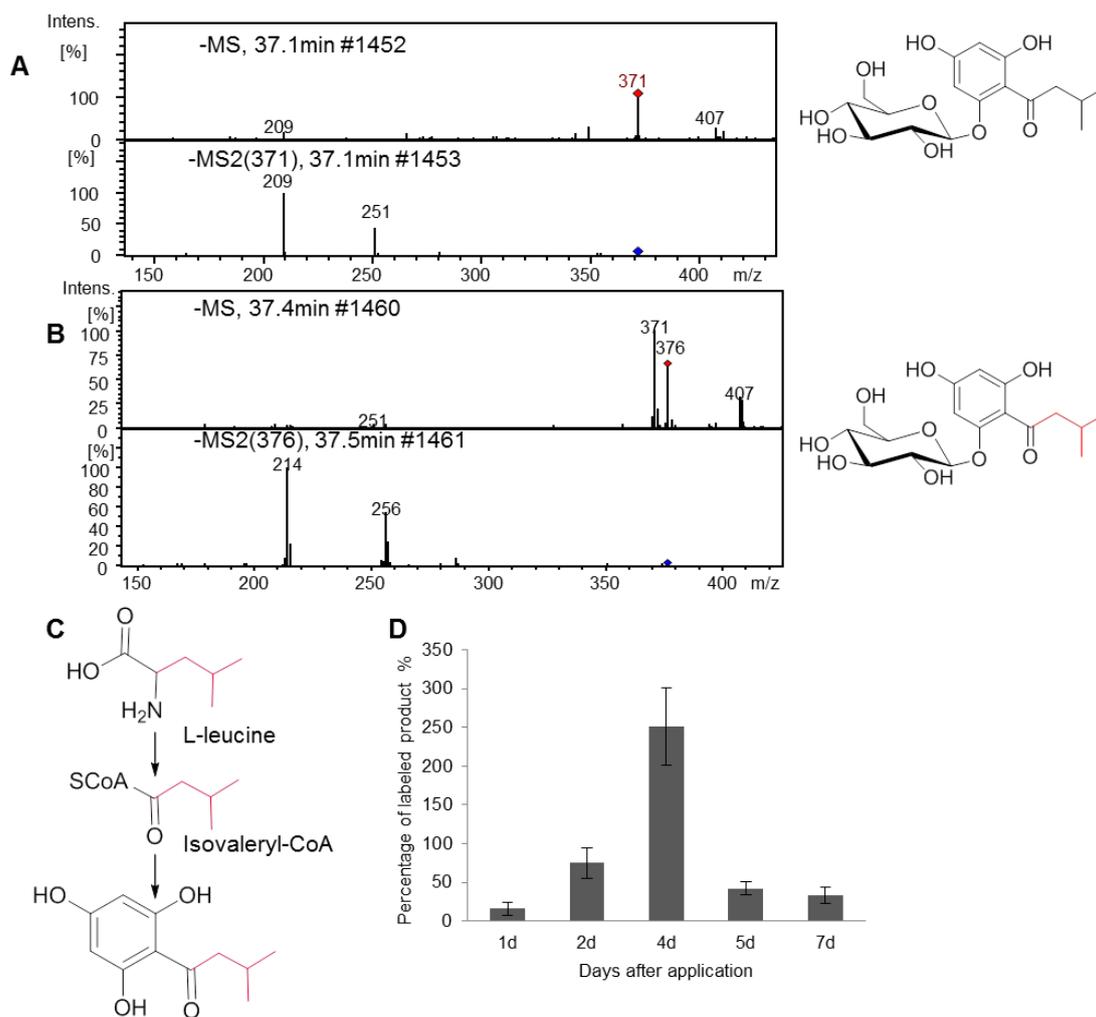
Different amounts of L-leucine- $^{13}\text{C}_6$  (10 and 50 mM), L-isoleucine- $^{13}\text{C}_6$  and L-valine-2,3,4,4,4,5,5,5- $\text{D}_8$  were administered into strawberry fruit of the same ripening stage and similar weight. Incorporation of the isotope labels into M1, M2 and M3 was shown by comparison of the integrated peak areas of the pseudomolecular ions of unlabeled M1 and M3 ( $m/z$  371) and M2 ( $m/z$  357) with the integrated peak area of the pseudomolecular ions of the labeled compound M1 and M3 ( $m/z$  376) and M2 ( $m/z$  364). The levels of isotopically labeled APG products were calculated as percentage of the unlabeled naturally occurring metabolites (Table 8). One day after the application of 10 mM L-leucine- $^{13}\text{C}_6$  (1.9 mg), L-isoleucine- $^{13}\text{C}_6$  (1.9 mg), and L-valine-2,3,4,4,4,5,5,5- $\text{D}_8$  (1.8 mg), the level of isotopically labeled M3, M1 and M2 accounted for 20, 16, and 23% of the unlabeled metabolites, respectively. These values increased 4-, 2.5-, and 2.1-fold when the amounts of the labeled precursors were enhanced by a factor of 5 (500  $\mu\text{L}$  of a 50 mM solution), indicating that strawberry fruit readily form APGs from amino acids, in particular L-leucine which is

converted to 1-[(3-methylbutyryl)-phlorogluciny]- $\beta$ -D-glucopyranoside M3, in the mature fruit.



**Figure 18.** Mass spectrum (MS) and MS2 of unlabeled M1 (**A**) and isotopically labeled M1 (**B**) after application of L-isoleucine- $^{13}\text{C}_6$ , proposed pathway (**C**), as well as the effect of the incubation period on the accumulation of labeled product (**D**) (percentage of labeled product was calculated from the integrated peak area of the pseudomolecular ion of isotopically labeled product  $m/z$  376; integrated peak area of pseudomolecular ion  $m/z$  371 of unlabeled product was set to 100%).

## Results



**Figure 19.** Mass spectrum (MS) and MS2 of unlabeled M3 (**A**) and isotopically labeled M3 (**B**) after application of L-leucine- $^{13}\text{C}_6$ , proposed pathway (**C**), as well as the effect of the incubation period on the accumulation of labeled product (**D**) (percentage of labeled product was calculated from the integrated peak area of the pseudomolecular ion of isotopically labeled product  $m/z$  376; integrated peak area of pseudomolecular ion  $m/z$  371 of unlabeled product was set to 100%).

The effect of longer incubation periods, up to 7 days, on the accumulation of labeled APGs was also studied. The contents of isotopically labeled APG products increased constantly and reached the highest levels at 4 days after injection (Figure 18D and 19D).

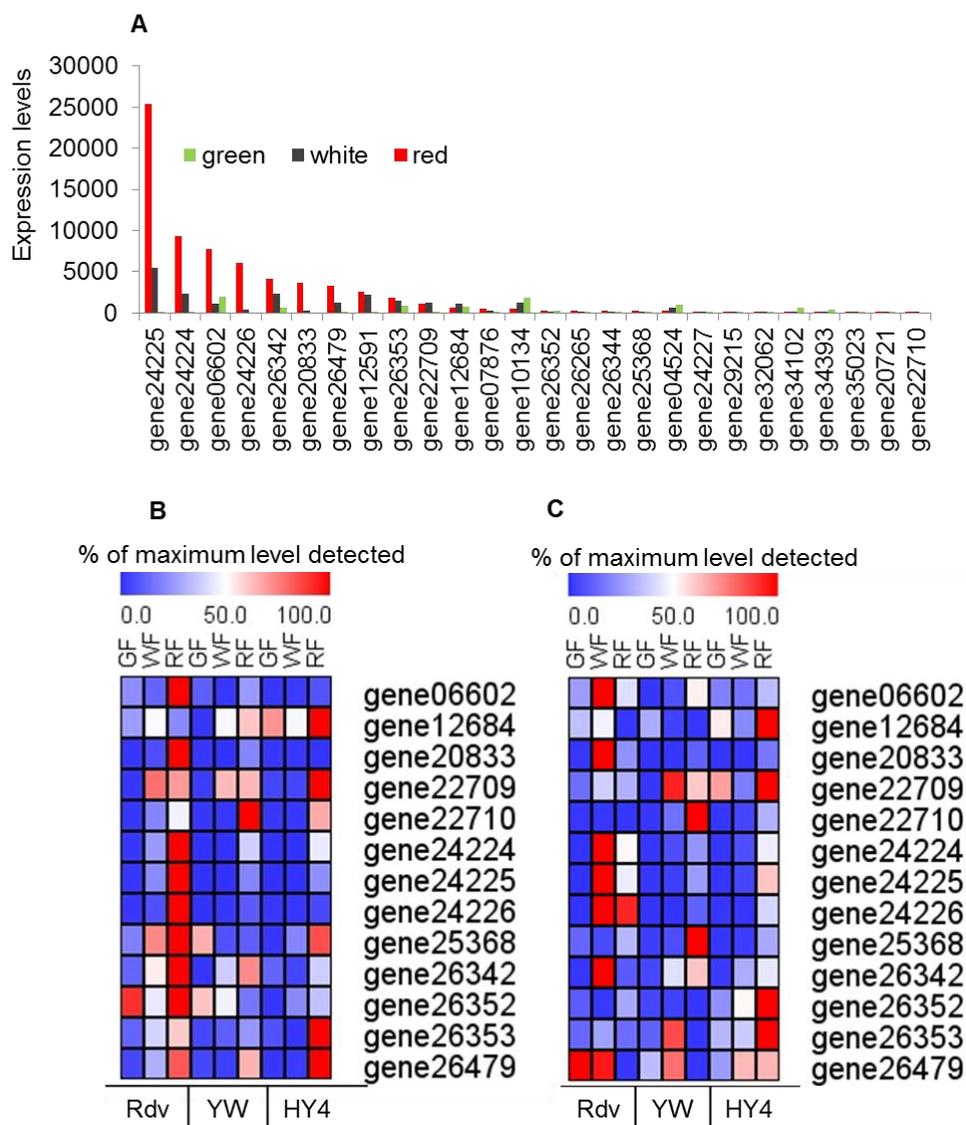
## 3.2 Substrate promiscuity of glucosyltransferases from strawberry

Genome-wide transcriptional analysis is a powerful tool to accelerate the pace of candidate gene discovery (Achnine et al., 2005). The achievement of the genome sequence of the diploid *F. vesca* makes it possible to mine candidate genes and monitor their transcription during fruit ripening on a large scale.

### 3.2.1 Selection of candidate glucosyltransferases

Mining of the publicly available *F. vesca* genome sequence yielded 199 apparently distinct putative UGT genes (Shulaev et al., 2011). To functionally characterize fruit ripening-related strawberry UGTs, transcript levels of the putative UGTs were analyzed in a transcriptomic data set obtained from fruits of different developmental stages (green, white and red) of three *F.vesca* varieties (Reine de Vallées, Yellow Wonder and Hawaii 4; Schulenburg and Franz-Oberdorf, unpublished results). Putative UGT genes of *F. vesca* var. Reine des Vallées were sorted in decreasing order of their expression levels in receptacle (Figure 20A). Reine des Vallées (Rdv) is an aromatic cultivar of *F. vesca* with an attractive red-fruit genotype. Only the transcript levels of 24 candidates increased significantly during strawberry ripening (Figure 20).

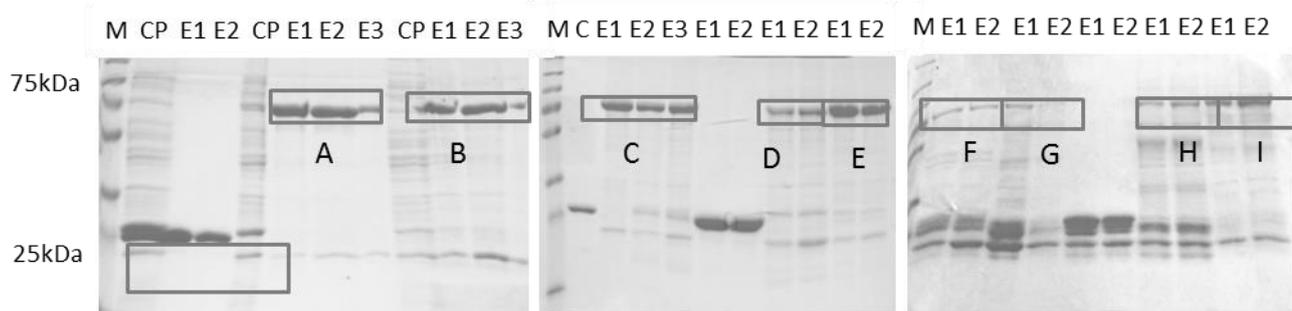
Family 1 UGTs usually utilize small-molecular-weight compounds as acceptor substrates and UDP-sugars as donors (Li et al., 2001). They commonly possess a carboxy terminal consensus sequence (plant secondary product glycosyltransferase PSPG box) believed to be involved in binding to the UDP moiety (Vogt and Jones, 2000). To narrow down the number of candidates we considered only UGTs containing a functional PSPG box which reduced the total number of candidates to 19. Finally, we took into account the transcripts levels of putative UGTs from two white-fruit genotypes of *F. vesca* (Yellow Wonder and Hawaii 4) and selected gene 06602, gene20833, gene22709, gene22710, gene24224, gene24225, gene24226, gene26342, gene26353, and gene26479 for further study (Figure 20B and C).



**Figure 20.** Gene expression levels of candidates UGTs in receptacle of fruits of Reine des Vallées (**A**) and relative expression levels of selected UGTs in receptacle (**B**) and achene (**C**) of three *F. vesca* varieties from the RNAseq data. GF: green fruits, WF: white fruits, RF: red fruits.

### 3.2.2 Protein expression and purification

Full length sequences of selected UGTs were amplified using the red fruit cDNA (*F. x ananassa* cv. Elsanta) as a template. Except for gene22710 and gene06602, which was annotated as indole-3-acetate-UGT, eight genes (gene20833, 22709, 24224, 24225, 24226, 26342, 26353, and 26479) were successfully amplified and ligated into the pGEX-4T1 vector. Two nucleotide sequences were obtained for gene24225 (a and b).



**Figure 21.** SDS-Page of GST fusion proteins expressed in *E. coli* after purification with GST affinity columns. E1 to E3: elution fraction 1-3, CP: crude proteins, M: marker proteins.

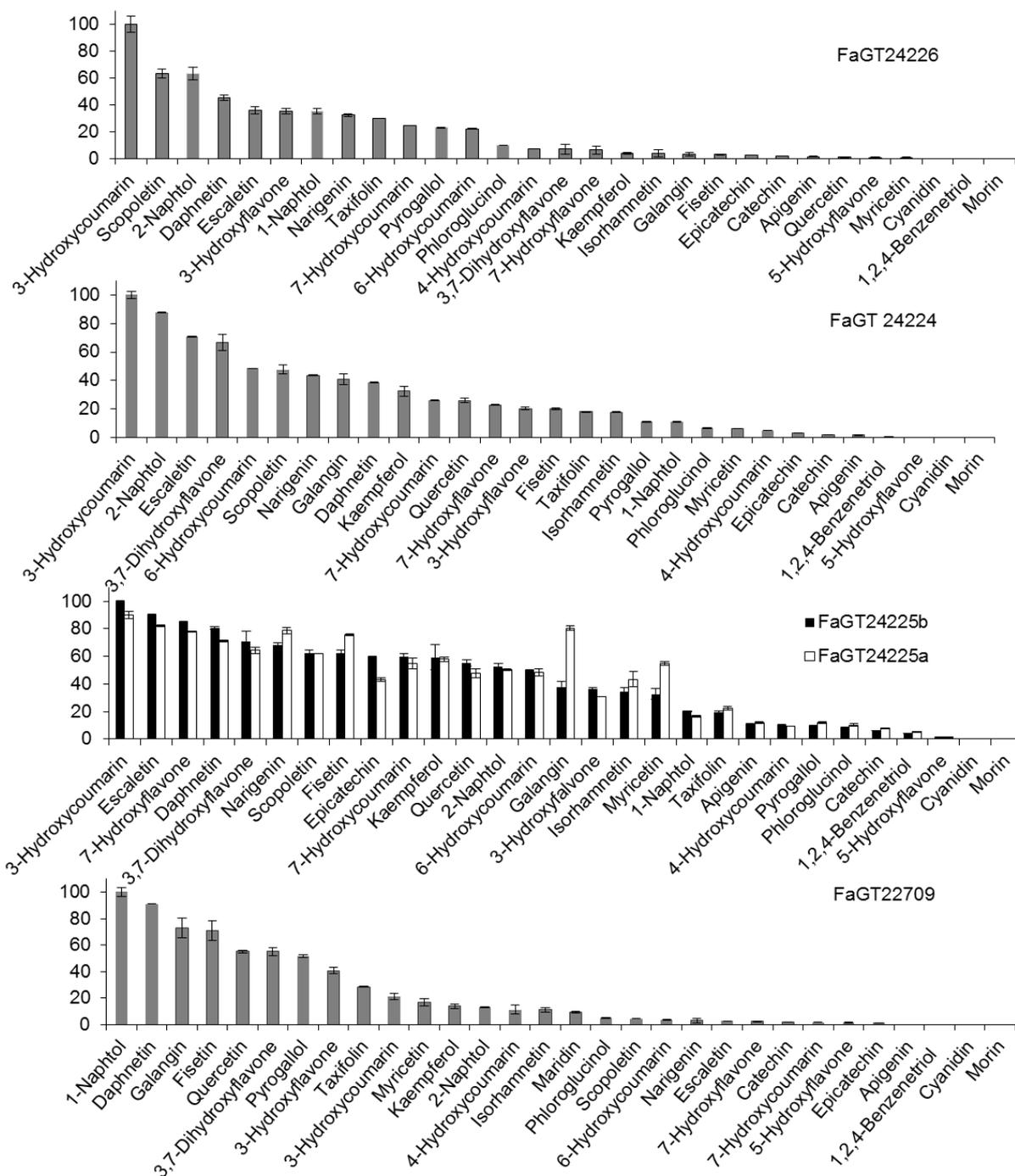
Proteins A-I was encoded by genes 24224, 24225a, 24225b, 24226, 22709, 26479, 20833, 26342 and 26353 in this order.

The resulting ORFs of the nine putative UGTs are from 1365 to 1494 bp in length and encode proteins of 455-498 amino acids. The nine candidate UGTs were successfully expressed in *E. coli* BL21 with an N-terminal GST-tag, affinity purified and verified by SDS-PAGE (Figure 21). The recombinant UGT proteins were used for further analysis.

### 3.2.3 Enzymatic activity

Enzymatic activities of the nine purified recombinant proteins were analyzed with UDP-[U-<sup>14</sup>C] glucose and various potential substrates present in strawberry (anthocyanin, flavonols, furaneol) and others not reported (naphthols, trihydroxybenzene isomers and hydroxycoumarins). Proteins encoded by gene20833, 26342 and 26353 could not convert any of the tested substrates and were not studied further. The protein encoded by gene24224 (named FaGT24224; the names of other proteins were similarly derived from the numbers of the genes) readily showed activity with most flavonols tested (3-, 7-, 3,7-dihydroxyflavone, galangin, kaempferol, quercetin, isorhamnetin), flavanones (naringenin and taxifolin) but not with 5- hydroxyflavone, cyanidin and morin (Figure 22). In addition, FaGT24224 glycosylated several hydroxycoumarins (3-, 4-, 6-, and 7-hydroxycoumarin, esculetin, daphnetin, and scopoletin), naphthols (1-naphthol, 2-naphthol), and trihydroxybenzene isomers (pyrogallol and phloroglucinol) to some extent. Other substrates such as menthol, 1,2,4-benzenetriol and gallic acid were not converted at all.

## Results



**Figure 22.** Relative specific activities (the value of the first column was set 100%) of FaGT24226, 24224, 24225a/b, and 22709 towards putative substrates as determined by radiochemical analysis with UDP-[U-<sup>14</sup>C] glucose.

FaGT24225a and FaGT24225b, encoded by two alleles of gene24225, showed activity towards 3-, 4-, and 7-hydroxycoumarin, 3-, 7-, and 3,7-hydroxyflavone, daphnetin, escaletin, naringenin, scopoletin, fisetin, kaempferol, quercetin, galangin, isohamnetin, myricetin, 2-naphthol, as well as epicatechin (> 50%). In addition, FaGT24225a,b glucosylated apigenin, pyrogallol, phloroglucinol, 1,2,4-benzenetriol and catechin (Figure 22).

Of the 33 substrates tested, 27 substrates were accepted by FaGT24226, of which 3-hydroxycoumarin, scopoletin, and 2-naphthol showed the highest activity (> 50%). FaGT24226 did not accept cyanidin, 1,2,4-benzenetriol and morin.

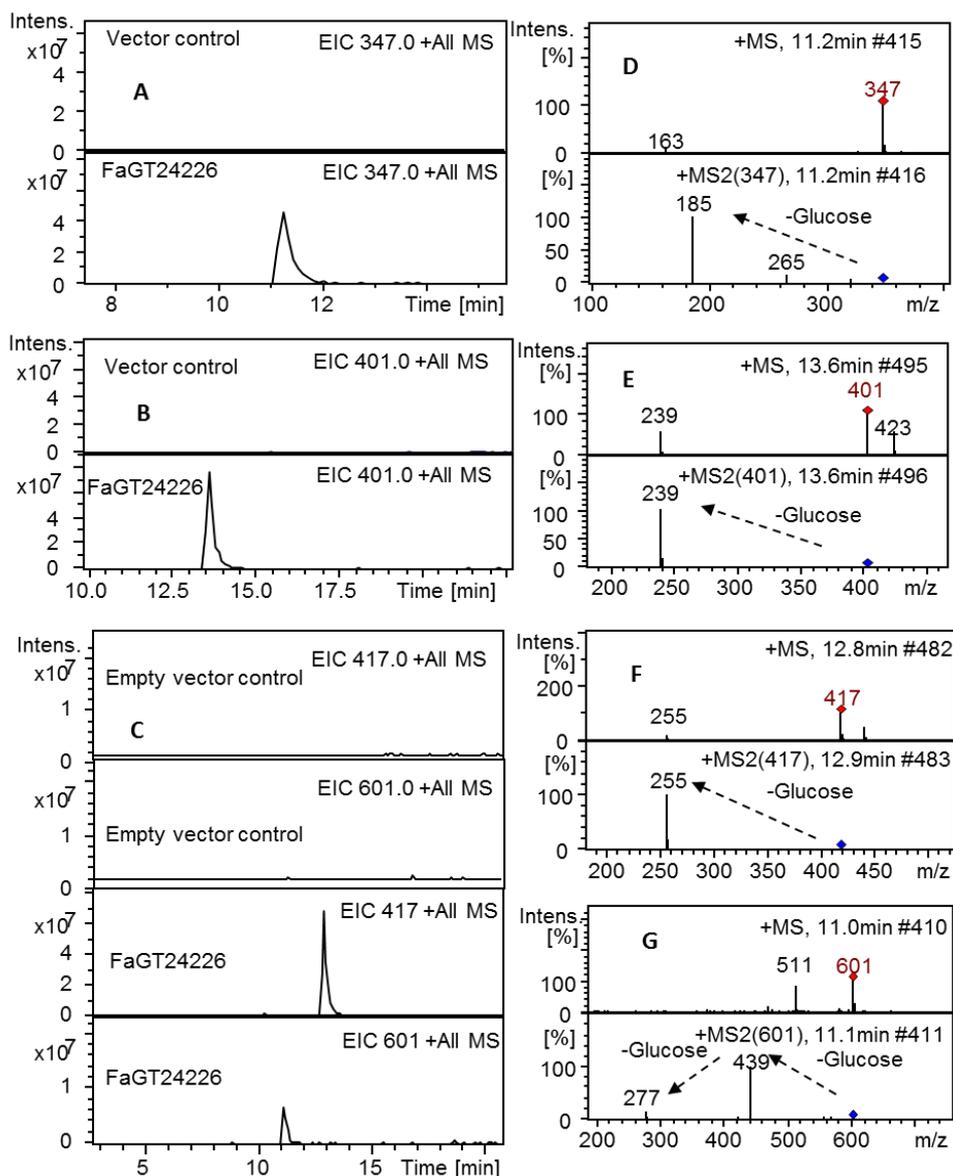
FaGT22709 showed highest activity towards 1-naphthol and is distinct from FaGT24224, 24225a/b and 24226 which all showed highest activity with 3-hydroxycoumarin (Figure 22). It should be noted that all of the five GTs have the capacity to accept a broad range of natural and xenobiotic substrates *in vitro*.

In contrast, the protein FaGT26479 showed very high substrate specificity as it could only efficiently convert galangin and trace amount of quercetin.

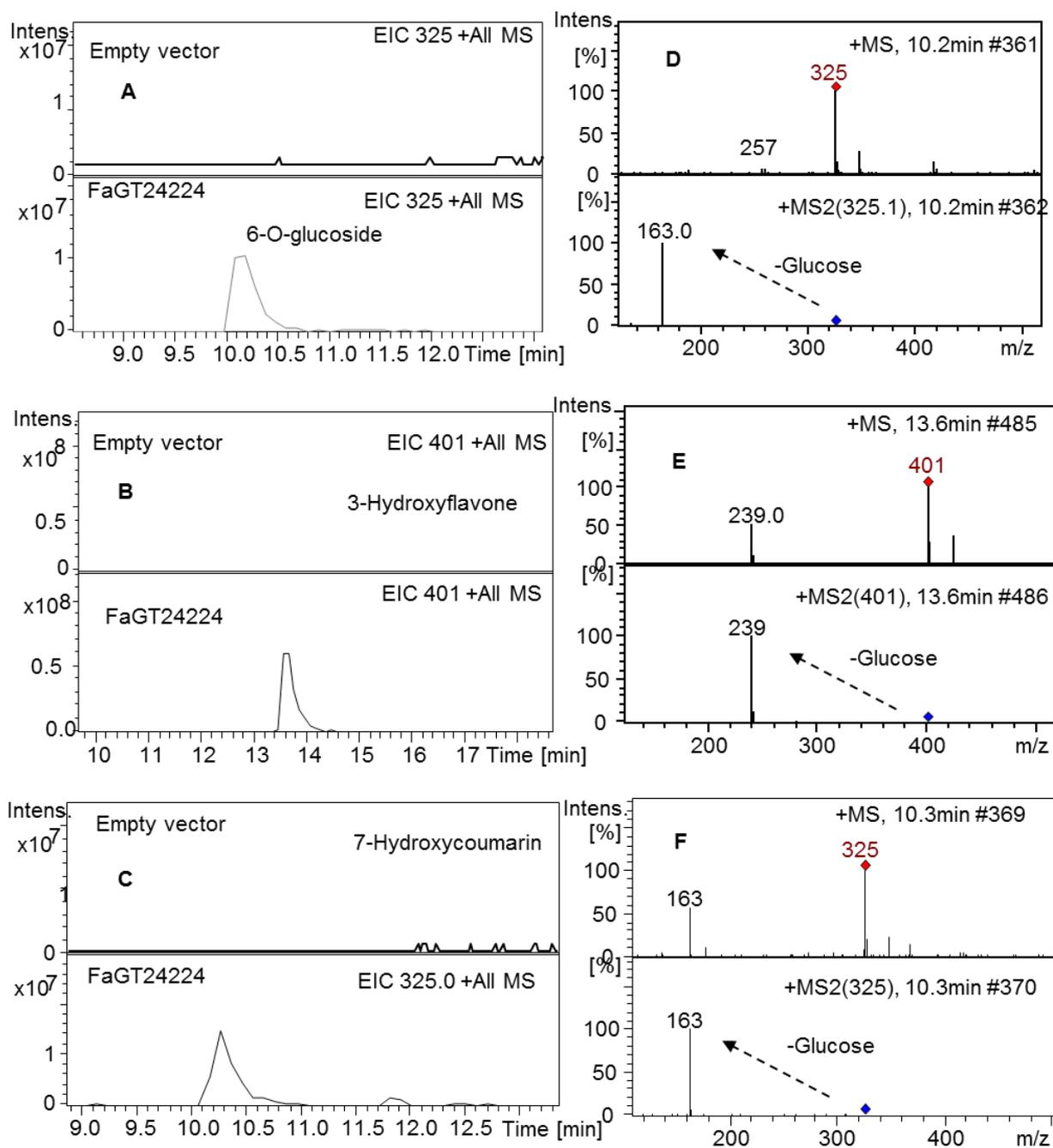
#### 3.2.4 Identification of reaction products

The products formed by FaGT24224, 24225a/b, 24226, 22709, and 26479 were identified by LC-MS (Figures 23-27). Substrate screenings already showed that FaGT24224, 24225a/b, 24226 and 22709 transform 3-hydroxycoumarin and 3-hydroxyflavone (Figure 22). The products formed from 3-hydroxyflavone show a loss of a glucose substituents ( $m/z$  162), which indicate that the hydroxyl group at position 3 was successfully glucosylated by FaGT24224, 24225a/b, 24226 and 22709 (Figures 23-25; 27A). To identify more complex products, products formed by FaGT24226 and 22709 were analyzed in detail. The substrate screening already showed that FaGT24226 could act on the hydroxyl group at position 3 and 7 of flavonols, whereas FaGT22709 could only act on hydroxyl group at position 3 (Figure 22). When 3,7-dihydroxyflavone was used as substrate, both FaGT24226 and 22709 produced one major monoglucoside, whereas only FaGT24226 also formed a diglucoside (Figure 23C), consistent with the substrate screening results (Figure 22). The diglucoside was putatively identified as 3,7-diglucoside because the product ions  $m/z$  439 and 277 were detected in the product ion spectrum of the pseudomolecular ion  $[M+H]^+$   $m/z$  601, indicating the loss of two glucose moiety (Figure 23G; Ablajan et al., 2006).

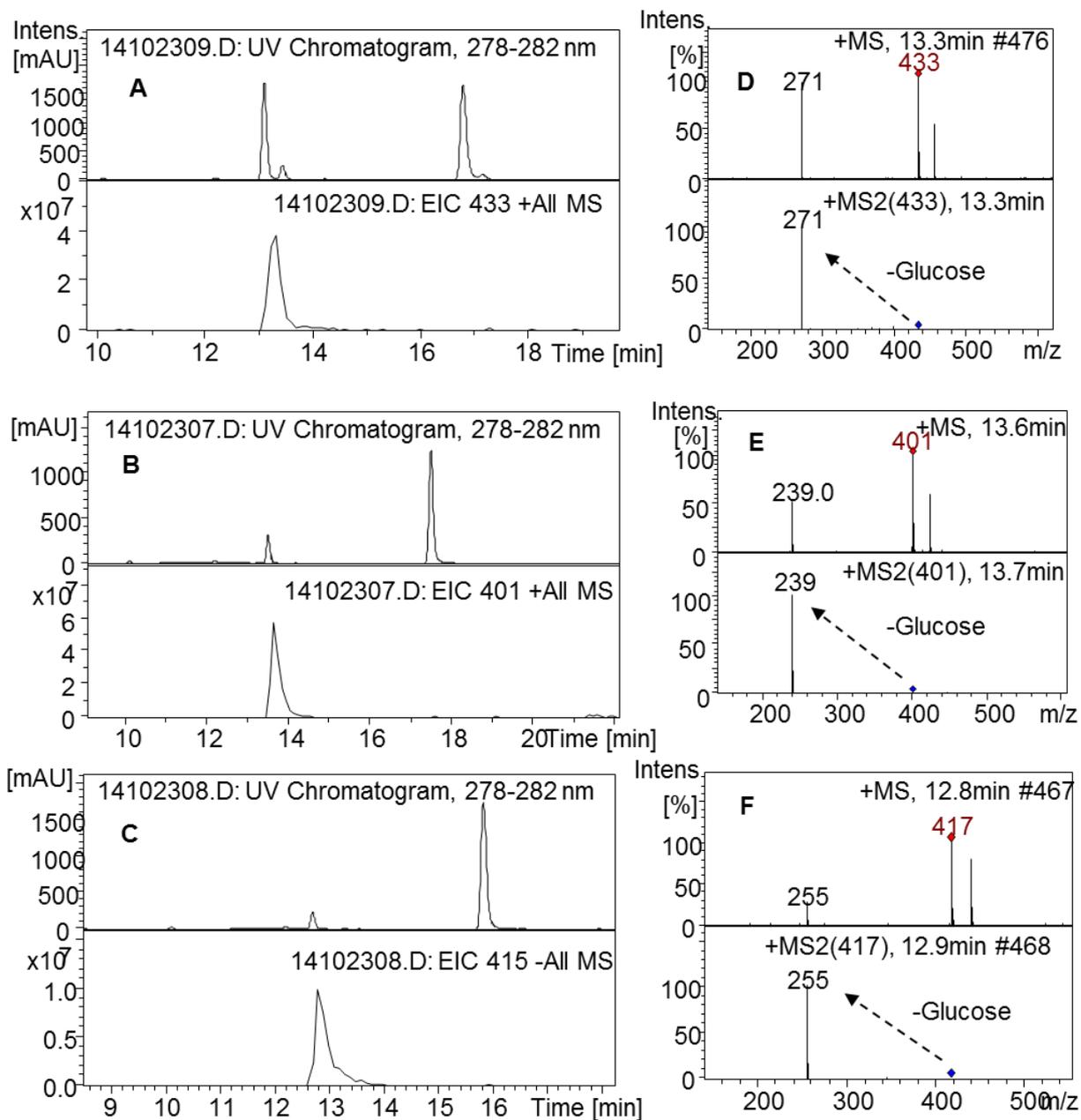
## Results



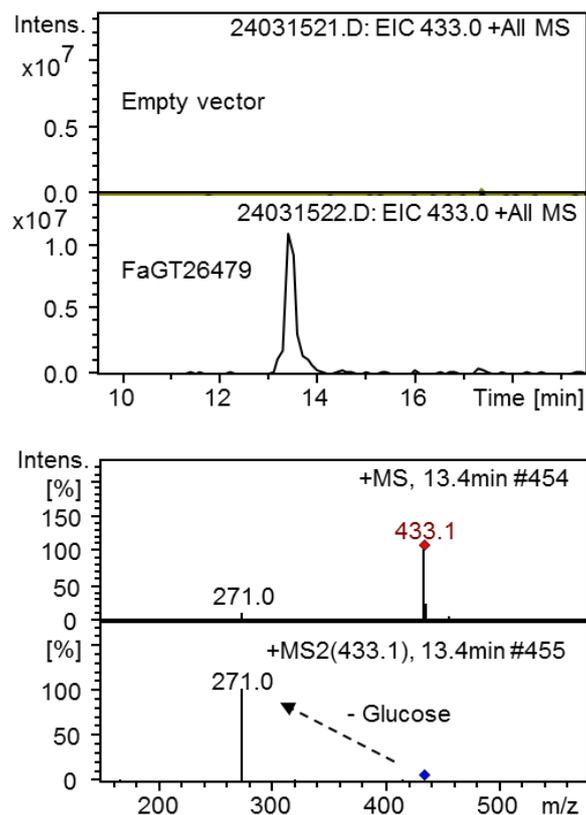
**Figure 23.** LC-MS analysis of products formed by FaGT24226 from 3-hydroxycoumarin (A), 3-hydroxyflavone (B) and 3,7-dihydroxyflavone (C). MS and MS2 spectra of glycosylated products of 3-hydroxycoumarin (D), 3-hydroxyflavone (E) and 3,7-dihydroxyflavone (F and G) clearly show the loss of one or two glucoses moiety.



**Figure 24.** LC-MS analysis of products formed by FaGT24224 from 6-hydroxycoumarin (A), 3-hydroxyflavone (B) and 7-hydroxycoumarin (C). MS and MS2 spectra of glycosylated products of 6-hydroxycoumarin (D), 3-hydroxyflavone (E) and 7-hydroxycoumarin (F) clearly show the loss of one glucose moiety.



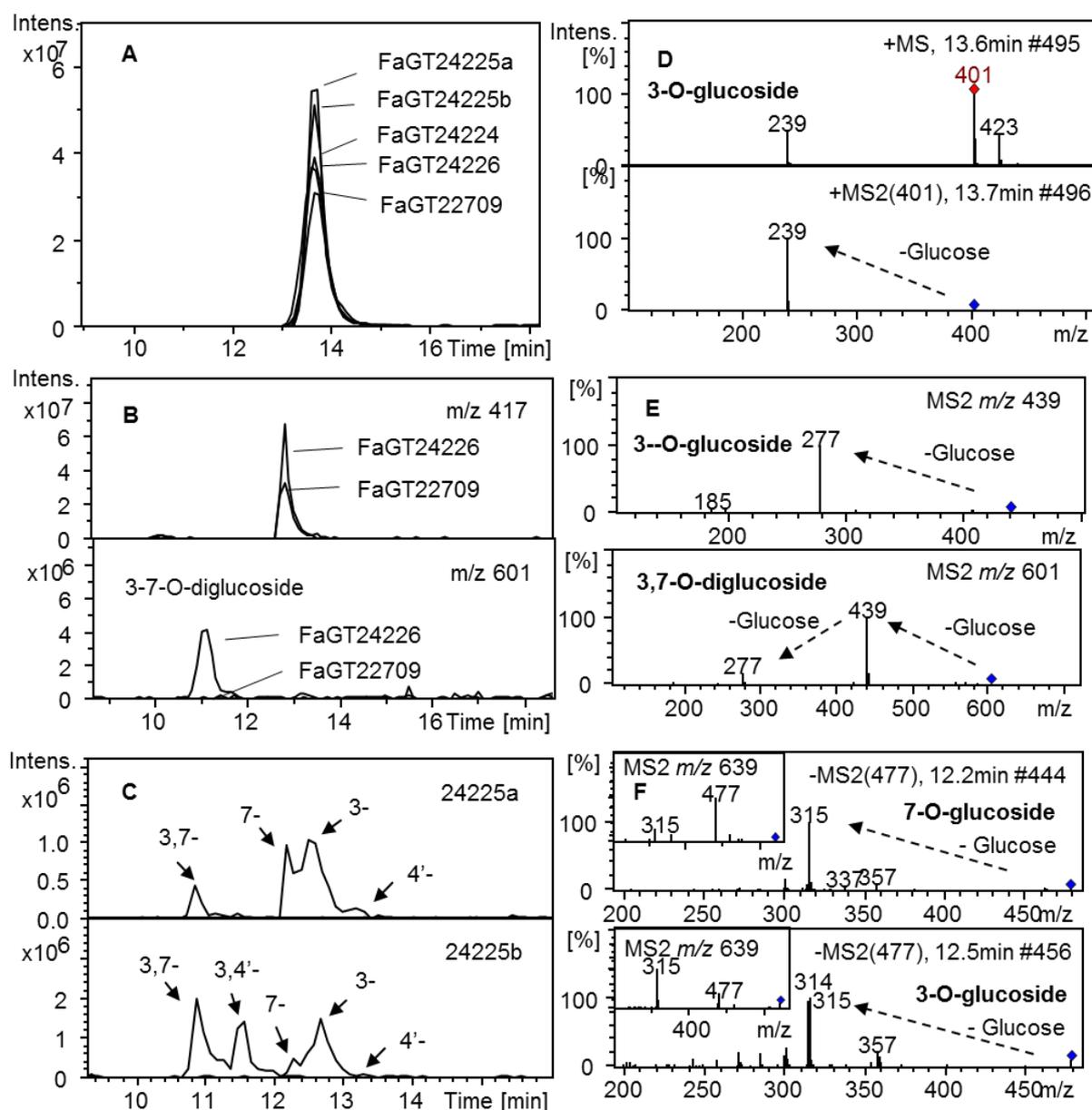
**Figure 25.** LC-MS analysis of products formed by FaGT22709 from galangin (**A**), 3-hydroxyflavone (**B**) and 7-hydroxyflavone (**C**). MS and MS2 spectra of glycosylated products of galangin (**D**), 3-hydroxyflavone (**E**) and 7-hydroxyflavone (**F**) clearly show the loss of one glucose moiety.



**Figure 26.** LC-MS analysis of products formed by FaGT26479 from galangin. MS and MS2 spectra of the glycosylated product of galangin clearly show the loss of one glucose moiety.

### 3.2.5 FaGT24225a and b show contrasting regioselectivity

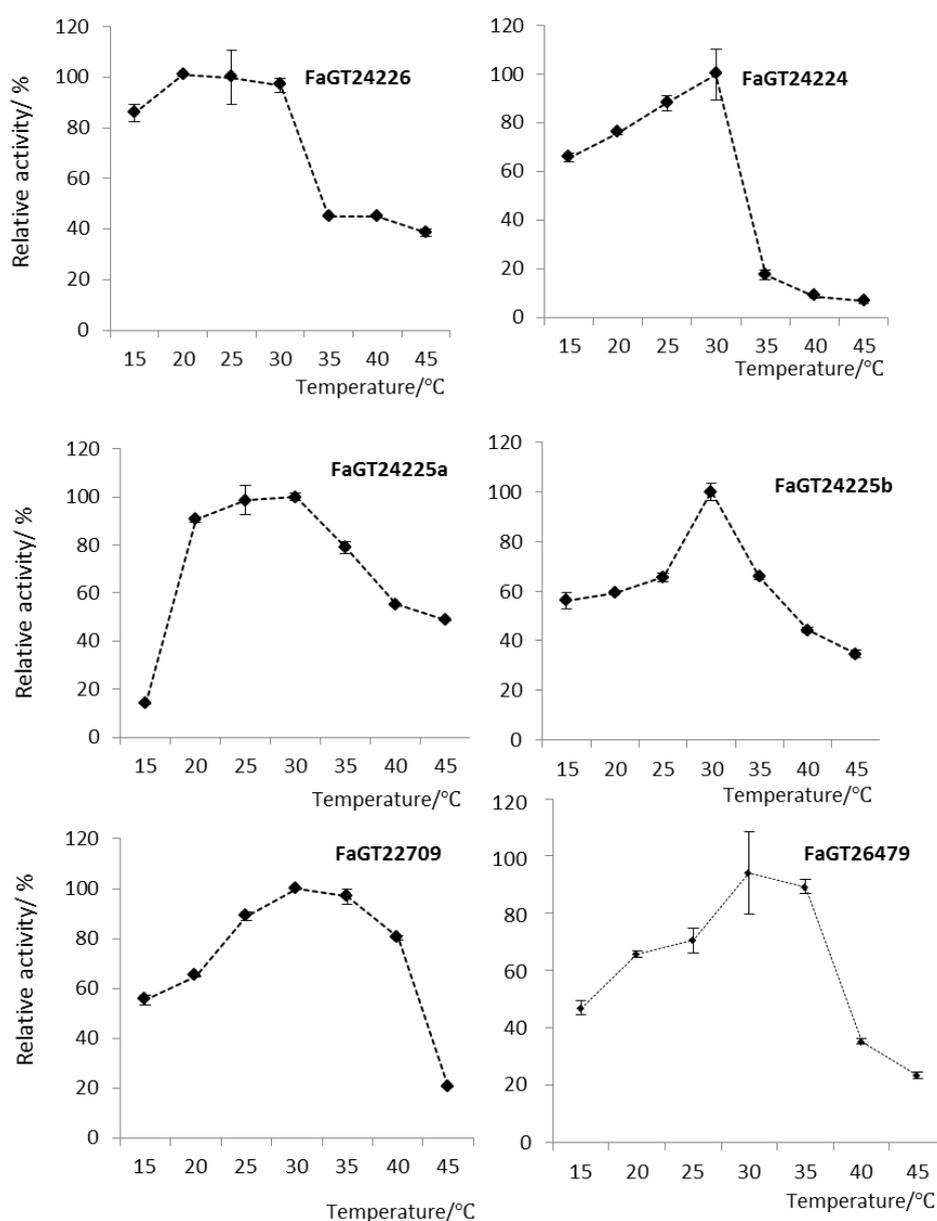
Monoglucosides and diglucoside were formed by both FaGT24225a and b when isorhamnetin was used as substrate. Three monoglucosides were putatively identified as 3-, 7-, 4'-glucoside (Figure 27C), according to the published elution order of the isomers (Shao et al., 2005; Lim et al., 2004) and their mass spectral data (Ablajan et al., 2006; Griesser et al., 2008b). The products formed by FaGT24225a and 24225b from isorhamnetin differ significantly. FaGT24225a produced two major mono-glucosides (3-, and 7-) and only one diglucoside. However, FaGT24225b formed two diglucosides but only one mono-glucoside (3-) (Figure 27C). FaGT24225a/b should be used to study the mechanism of regioselectivity because these two proteins show contrasting regioselectivity but differ in only a few amino acids.



**Figure 27.** LC-MS analysis of products formed by FaGT24224, 24225a/b, 24226, and 22709. 3-hydroxyflavone (**A**), 3,7-dihydroxyflavone (**B**), and isorhamnetin (**C**) were incubated with UDP-glucose and recombinant FaGTs. MS and MS2 spectra of glycosylated products of 3-hydroxyflavone (**D**), 3,7-dihydroxyflavone (**E**), and isorhamnetin (**F**) clearly show the loss of one glucose moiety.

### 3.2.6 Kinetic properties and substrate preference

The optimum reaction conditions for FaGT24224, 24225a/b and 24226 were determined with 3-hydroxycoumarin as acceptor, whereas the assay condition for FaGT22709 and 26479 was optimized using 1-naphtol and galangin as substrate, respectively. The highest activities of all recombinant FaGTs were detected at pH 7.0, and the optimal temperature for FaGT24224, 24225a/b, 24226 and 22709 was 30, 30, 30, 20, and 30°C, respectively (Figure 28).



**Figure 28.** Temperature optima of selected FaGTs.

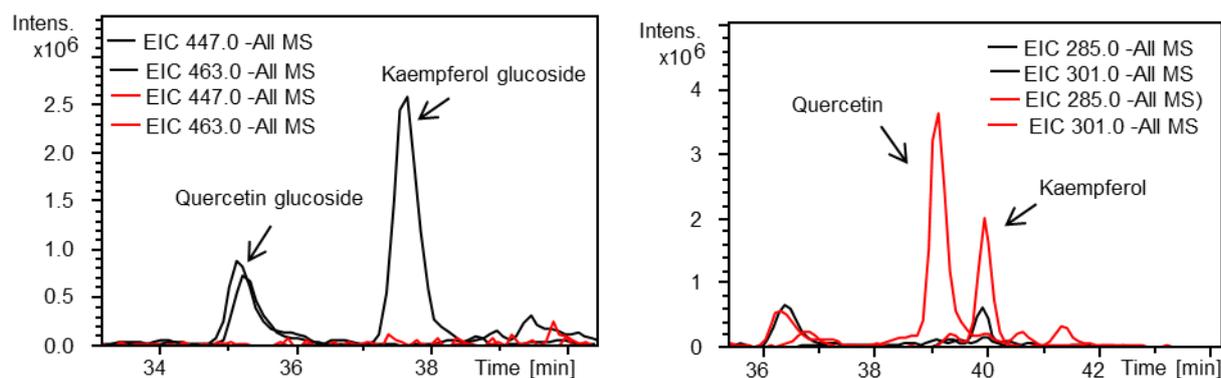
Kinetic properties were determined in the linear range of the enzymatic reaction (2  $\mu\text{g}$  of protein, 30 min reaction time). The apparent  $K_M$  values for naringenin of FaGT24224, 24225a/b, and 24226 were 77.9, 30.1, 23.2, and 31.7  $\mu\text{M}$ , respectively (Table 9). The  $V_{\text{max}}$  value of FaGT24225b with 3-hydroxycoumarin was 2-, 6- and 12-fold higher than that of FaGT24225a, FaGT24224, and FaGT24226, respectively. FaGT24225b show a maximum specificity constant  $k_{\text{cat}}/K_M$  value for 3-hydroxycoumarin and kaempferol of 3837 and 8141  $\text{M}^{-1} \text{s}^{-1}$ , respectively. The  $V_{\text{max}}$  value of FaGT26479 for galangin is 4-fold higher than that of FaGT22709 (Table 9).

**Table 9.** The kinetic data of candidates FaGTs.

protein	substrate	$K_M$ ( $\mu\text{M}$ )	$V_{\text{max}}$ (nKat $\text{mg}^{-1}$ )	$k_{\text{cat}}/K_M$ ( $\text{M}^{-1} \text{s}^{-1}$ )
FaGT24224	3-Hydroxycoumarin	54.55 $\pm$ 4.83	0.35 $\pm$ 0.01	481
	Naringenin	77.92 $\pm$ 18.25	0.12 $\pm$ 0.01	116
	Kaempferol	3.12 $\pm$ 0.33	0.086 $\pm$ 0.002	2067
FaGT24225a	3-Hydroxycoumarin	35.19 $\pm$ 3.58	1.02 $\pm$ 0.02	2174
	Naringenin	30.06 $\pm$ 1.90	0.33 $\pm$ 0.004	823
	Kaempferol	5.2 $\pm$ 1.25	0.32 $\pm$ 0.02	4615
FaGT24225b	3-Hydroxycoumarin	39.09 $\pm$ 5.53	2.00 $\pm$ 0.07	3837
	Naringenin	23.19 $\pm$ 3.37	0.66 $\pm$ 0.02	2135
	Kaempferol	7.37 $\pm$ 2.36	0.80 $\pm$ 0.06	8141
FaGT24226	3-Hydroxycoumarin	168.86 $\pm$ 14.30	0.17 $\pm$ 0.01	76
	Naringenin	31.74 $\pm$ 14.73	0.02 $\pm$ 0.003	47
	1-Naphtol	77.78 $\pm$ 28.54	0.02 $\pm$ 0.002	19
FaGT22709	3-Hydroxycoumarin	94.19 $\pm$ 63.83	0.0037 $\pm$ 0.001	3
	1-Naphtol	127.85 $\pm$ 6.45	0.28 $\pm$ 0.01	164
	Kaempferol	77.83 $\pm$ 10.21	0.017 $\pm$ 0.001	16
	Quercetin	17.26 $\pm$ 5.63	0.0056 $\pm$ 0.001	24
	Galangin	10.54 $\pm$ 1.79	0.22 $\pm$ 0.01	1565
FaGT26479	Galangin	9.32 $\pm$ 2.79	0.81 $\pm$ 0.02	6517

### 3.2.7 Identification of the natural substrates of FaGTs

Activity based metabolite profiling of aglycone libraries provides a versatile tool to uncover the enzymatic activities encoded by genes of unknown function (Bönisch et al., 2014). In this study, an enzymatically hydrolyzed strawberry glycoside extract was used as a physiologic library to reveal the natural substrates of selected FaGT proteins. LC-MS analysis of the aglycone library showed that the glycosides in the extract were successfully enzymatically hydrolyzed (Figure 29).



**Figure 29.** LC-MS analysis of a strawberry glycoside extract (black line) and an enzymatically hydrolyzed glycoside extract (red line). Signals for kaempferol glucoside, quercetin glucoside (left), as well as the aglycone kaempferol and quercetin (right) are shown.

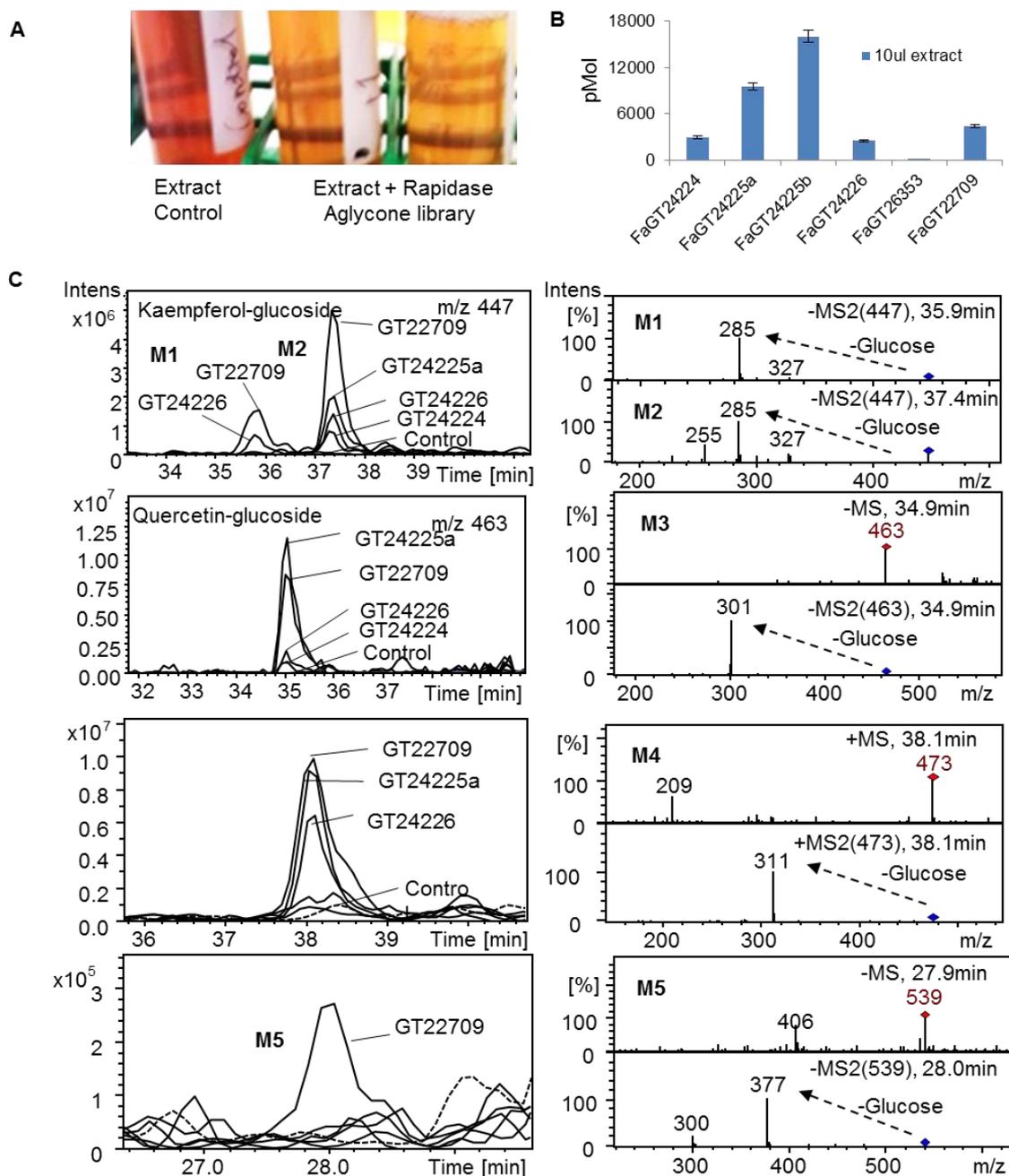
The physiologic library which contained potential biological substrates of UGTs was screened with recombinant FaGTs and either radiochemical labeled or unlabeled UDP-glucose. Initially, the aglycone extract was incubated with purified recombinant enzymes and UDP-[U-<sup>14</sup>C] glucose and products were extracted. Products formed by FaGT24224, 24225a/b, and 22709 became radioactively labeled (Figure 30B), which indicated that the FaGTs could utilize the natural aglycones in the hydrolyzed extract as substrates. FaGT24226 showed low activity towards the aglycones. FaGT26353, which could not convert any of the tested substrates *in vitro*, was also inactive with the strawberry aglycones.

The formed glycosides were identified by LC-MS (Figure 30C). The glycosylated products M2 and M3 were identified as kaempferol glucoside and quercetin glucoside, respectively (Figure 30C). Additional analyses revealed that FaGT24224, 24225a, 24226 and 22709 could form kaempferol glucoside but quercetin glucoside was mainly synthesized by FaGT24225a and FaGT22709.

Additionally, pairwise comparison of the MS data of the products formed by the FaGTs and the corresponding data of the empty-vector control was performed. This approach yielded three novel glucosides (Figure 30C). Structure identification of these glucose conjugates would be helpful to reveal biologically relevant substrates of the FaGTs. The pseudomolecular ions [M-H]<sup>-</sup> of M1 (*m/z* 447; retention time 35.9 min), [M+H]<sup>+</sup> of M4 (*m/z* 473; 38.1 min), and [M-H]<sup>-</sup> of M5 (*m/z* 539; 27.9 min) show a neutral loss of 162 dalton (loss of glucose) in the corresponding MS2 (Figure 30C) indicating that all metabolites are modified by a glucose residue. In other words, the aglycones of M1, M4, and M5 are

## Results

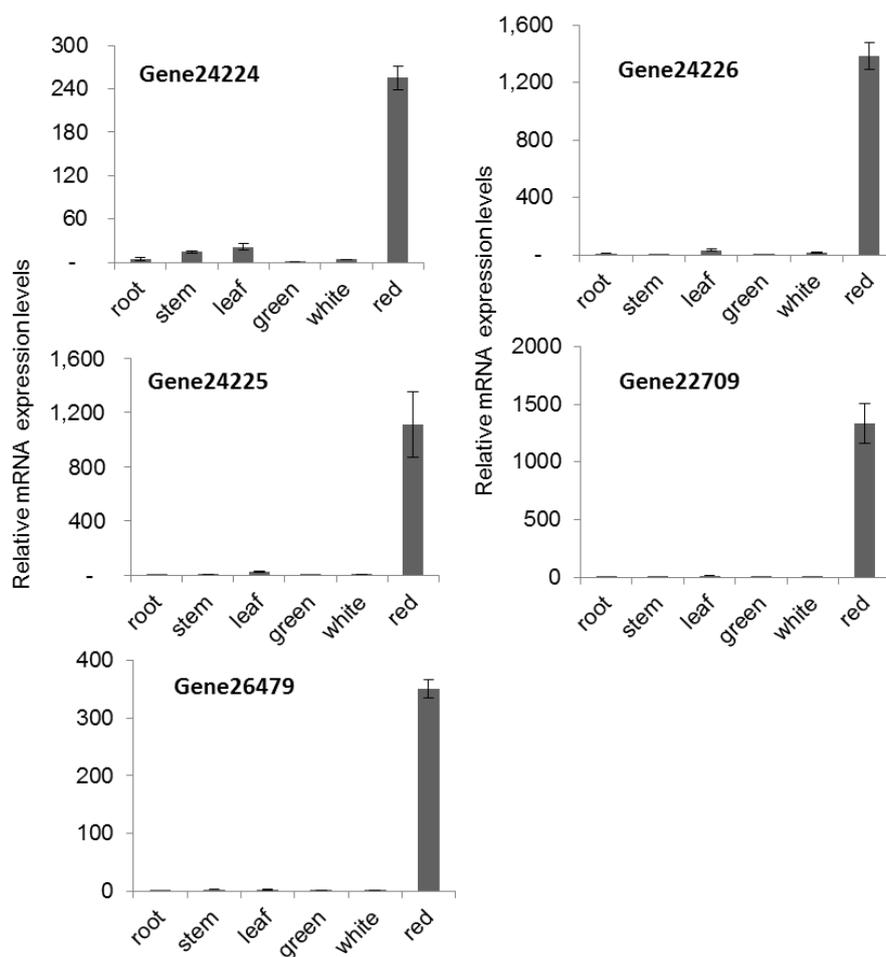
potential biological substrates of FaGTs. Isolation and structural identification of the products could provide novel substrates and reveal the function of the FaGT proteins (Figure 30C).



**Figure 30.** Preparation of an aglycone library from glycosides isolated from strawberry (A). Screening of the aglycone library with FaGT proteins and UDP-[U-<sup>14</sup>C] glucose (B). LC-MS analysis of products M1 – M5 formed by screening of the aglycone library with FaGT24224, 24225a, 24226, 26353 and 22709.

### 3.2.8 Expression analysis in *Fragaria x ananassa*

The mRNA transcription levels of the genes which encode the active FaGTs were determined by quantitative real-time PCR in vegetative tissues (root, stems, and leaves) and in fruits at different developmental stages (Appendix Table 2) of *F. x ananassa* (cv Elsanta). Gene24224 and 26479 showed a 255- and 350-fold higher expression level, respectively at the red stage when compared to the green stage (Figure 31). In particular, transcripts of gene24225, 24226 and 22709 are highly abundant in the ripe red fruit. Their expression levels strongly increase during ripening by 1113-, 1386-, and 1332-fold, respectively. Except for gene24224, the transcripts of the other four FaGTs are barely expressed in vegetative tissues, such as the root, stems, and leaves (Figure 31). Because these genes are predominantly expressed in fully ripened fruit receptacles it is assumed that these FaGTs are involved in metabolic pathways strongly activated in ripe receptacle tissue.

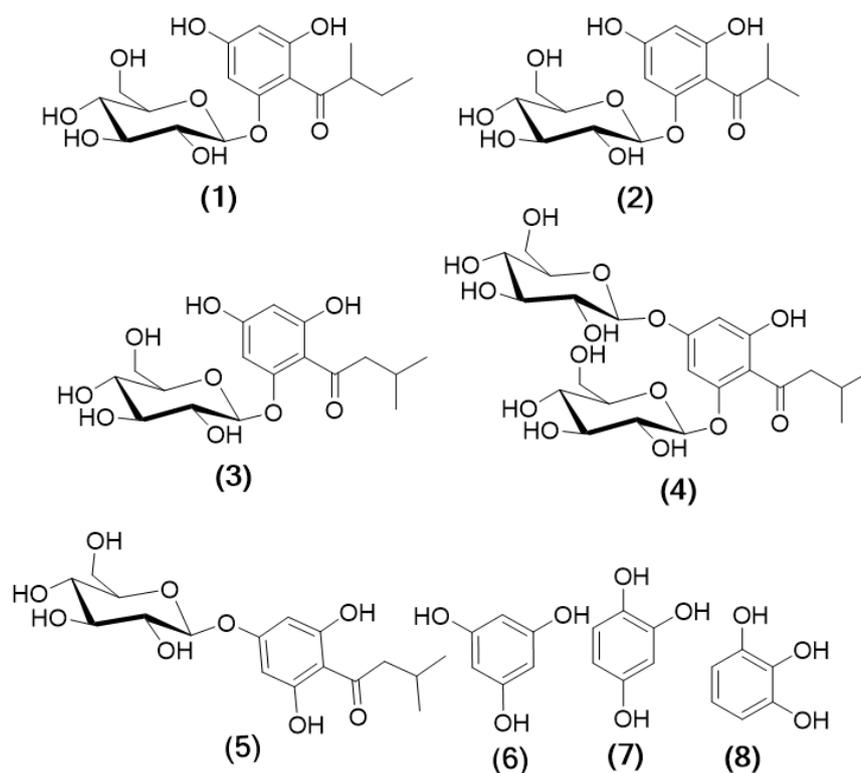


**Figure 31.** Relative mRNA expression levels of the candidate genes in different tissues (root, stems, leaves, green fruit, green fruit and red fruit) were determined by real-time PCR with

*FaRib413* used for normalization. mRNA abundances were normalized to the level in green fruit (means  $\pm$  SE triplicate technical repetitions from two cDNA preparations)

### 3.3 Characterization of a phloroglucinol-glycosyltransferase from strawberry

Through reverse genetic analysis of four ripening-related genes in octoploid strawberry, we discovered biologically active acylphloroglucinol (APG)-glucosides as native strawberry metabolites (Figure 32). In hop and strawberry, APGs are produced by valerophenone synthase (Paniego et al., 1999; Okada et al., 2004a), and a dual functional chalcone synthase 2 (CHS2, refer to 3.1), respectively. However, little is known about the enzymes catalyzing the glycosylation of phloroglucinols.

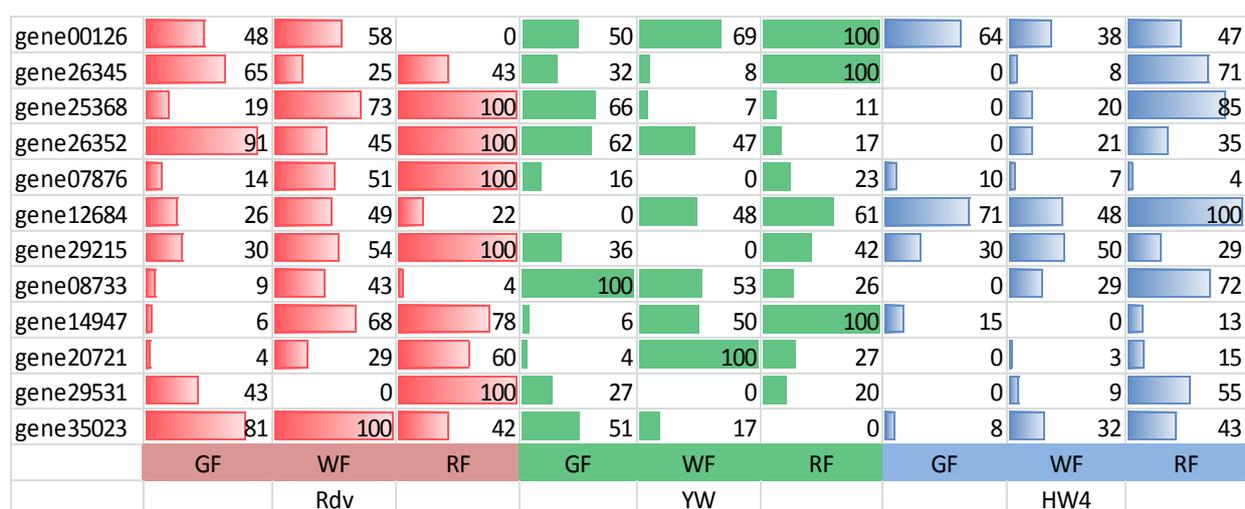


**Figure 32.** The structures of glycosylated APGs found in strawberry fruit (1-5) and commercially available trihydroxybenzenes (phloroglucinol **6**, 1,2,4-benzenetriol **7**, and pyrogallol **8**) used in this study.

### 3.3.1 Selection of candidate GTs and proteins purification

Candidate UGT genes were selected from transcriptome data sets (kindly provided by Katja Schulenburg and Katrin Franz-Oberdorf) obtained from fruit receptacles and achenes of three *F. vesca* varieties (Reine des Valeés, Yellow Wonder, and Hawaii 4) of different developmental stages (green, white and ripe). Gene00126, 07876, 08733, 12684, 14947, 20721, 26345, 26352, 29215, 29531, 25368 and 35023 (Shulaev et al., 2011) were selected because they show a ripening related expression pattern in at least one of the tissue of one of the three *F. vesca* varieties (Figures 33 and 34). All selected UGTs possess the carboxy terminal consensus sequence of plant secondary product GTs (PSPG box) assumed to be involved in the binding of the sugar donor.

Full-length sequences of selected GTs were amplified using cDNA from red fruit of *F. x ananassa* cv. Elsanta as a template and ligated into the pGEX-4T1 vector (Appendix Table 3). Only four of them, including two alleles of a gene (gene07876a/b, 26345 and 00126) were successfully expressed in *E. coli* BL21 with an N-terminal GST-tag. The proteins were further affinity purified and verified by SDS-PAGE. The resulting ORFs of genes 07876a/b, 26345 and 00126 are 1437, 1437, 1434 and 1434 bp in length and code for proteins with 479, 479, 478 and 478 amino acids, respectively with a calculated molecular mass of 52.59, 52.63, 53.30 and 53.57 kD, respectively.



**Figure 33.** Relative transcript levels of selected genes in strawberry receptacle of different developmental stages (green, white and red/ripe stage) from three *F. vesca* varieties (Reine des Vallées Rdv, Yellow Wonder YW, and Hawaii 4 HW4). GF: green fruit, WF: white fruit, RF: red fruit or ripe fruit.

## Results

Gene	GF	WF	RF	GF	WF	RF	GF	WF	RF
gene00126	58	45	92	0	22	70	35	25	100
gene26345	0	100	85	7	40	68	13	52	81
gene25368	12	6	32	3	7	100	0	3	29
gene26352	10	1	29	5	4	0	23	52	100
gene07876	45	100	63	33	26	0	40	94	35
gene12684	35	47	0	30	4	0	54	21	100
gene29215	5	32	64	9	28	100	0	100	34
gene08733	14	100	45	17	6	0	8	15	19
gene14947	0	100	84	8	39	41	6	34	29
gene20721	0	97	0	100	67	0	29	0	48
gene29531	0	0	5	36	100	0	56	5	16
gene35023	77	0	39	100	75	0	86	42	24
	GF	WF	RF	GF	WF	RF	GF	WF	RF
		Rdv			YW			HW4	

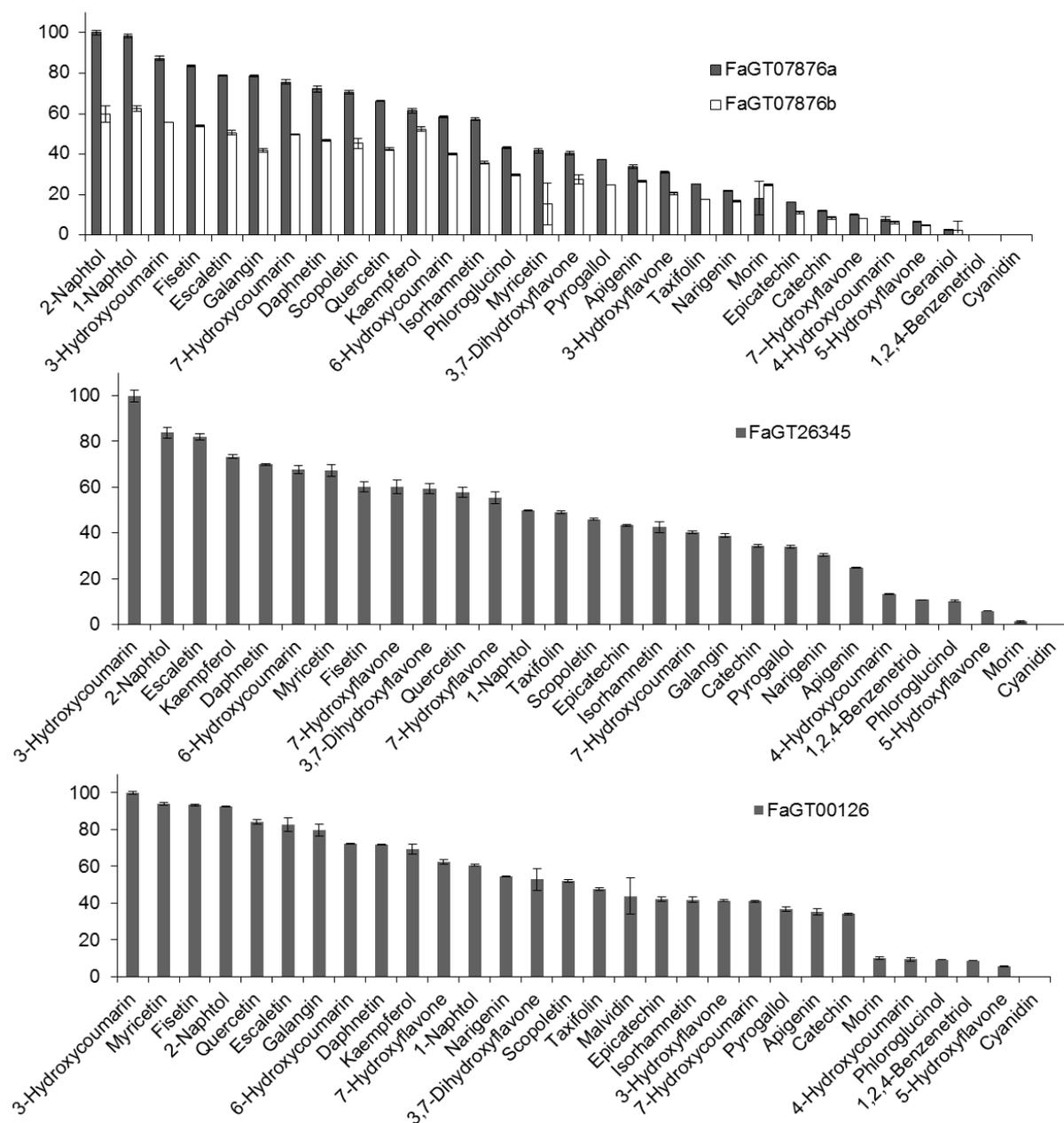
**Figure 34.** Relative transcript levels of selected genes in strawberry achenes of different developmental stages (green, white and red/ripe stage) from three *F. vesca* varieties (Reine des Vallées Rdv, Yellow Wonder YW, and Hawaii 4 HW4). GF: green fruit, WF: white fruit, RF: red fruit or ripe fruit.

### 3.3.2 Substrate screening

Substrate specificity of the recombinant proteins was determined *in vitro* with UDP-glucose as the donor substrate and more than 30 acceptor substrates including compounds present in strawberry fruit such as anthocyanidins, and flavonols as well as others not reported in strawberry (Figure 35). Both allelic proteins of FaGT07876 (a/b) preferred 2-naphthol and 1-naphthol but also efficiently (> 50% relative activity) glucosylated several hydroxycoumarins (3-, 6-, and 7-hydroxycoumarin, esculetin, daphnetin, and scopoletin) and flavonols (fisetin, galangin, quercetin kaempferol and isorhamnetin). Additionally, FaGT07876a/b showed catalytic activity towards flavanones (naringenin, taxifolin), trihydroxybenzene (phloroglucinol, pyrogallol) and flavonols (3-, 7-hydroxyflavone, 3,7-dihydroxyflavone, myricetin, apigenin and morin), epicatechin and catechin. Phenols such as 5-hydroxyflavone and 1,2,4-benzenetriol were poor substrates and cyanidin was not converted (Figure 35). In general, the allelic protein FaFT07576b showed lower activity towards the substrates, except for morin, when compared with the values obtained for FaGT07876a (Figure 35).

FaGT00126 and 26245 showed a similar substrate tolerance and accepted most of the phenols that were tested such as flavonols (3-, 7-hydroxyflavone, 3,7-dihydroxyflavone, galangin, kaempferol, quercetin, and isorhamnetin), flavanones (naringenin and taxifolin), hydroxycoumarins (3-, 6-, and 7-hydroxycoumarin, esculetin, daphnetin, and scopoletin),

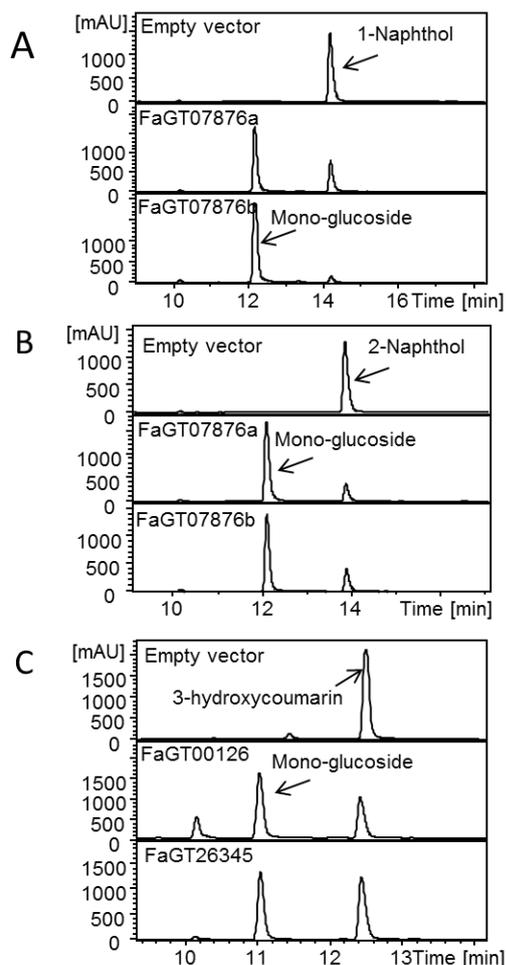
naphthols (1- and 2-naphthol), pyrogallol, epicatechin and catechin. Low enzymatic activity was observed with 5-hydroxyflavone, 1,2,4-benzenetriol, phloroglucinol, 4-hydroxycoumarin and morin (Figure 35). Both proteins showed highest activity towards 3-hydroxycoumarin, in contrast to FaGT07876a/b (Figure 35).



**Figure 35.** Relative enzymatic activity (the value of the first column was set 100%) of Fa07876a/b, FaGT026345, and FaGT00126 proteins from strawberry towards putative substrates as determined by radiochemical analysis with UDP-[U-<sup>14</sup>C] glucose.

### 3.3.3 Identification of reaction products

Major products formed by the enzymes were identified by LC-MS (Figure 36). When 1-naphthol and 2-naphthol were used as substrates, more than 80% of the tested substrates were converted to the corresponding glucosides by FaGT07876a/b (Figure 36). Similarly, more than half of 3-hydroxycoumarin were successfully catalyzed to its corresponding mono-glucosides by FaGT26345 and FaGT00126 (Figure 36).



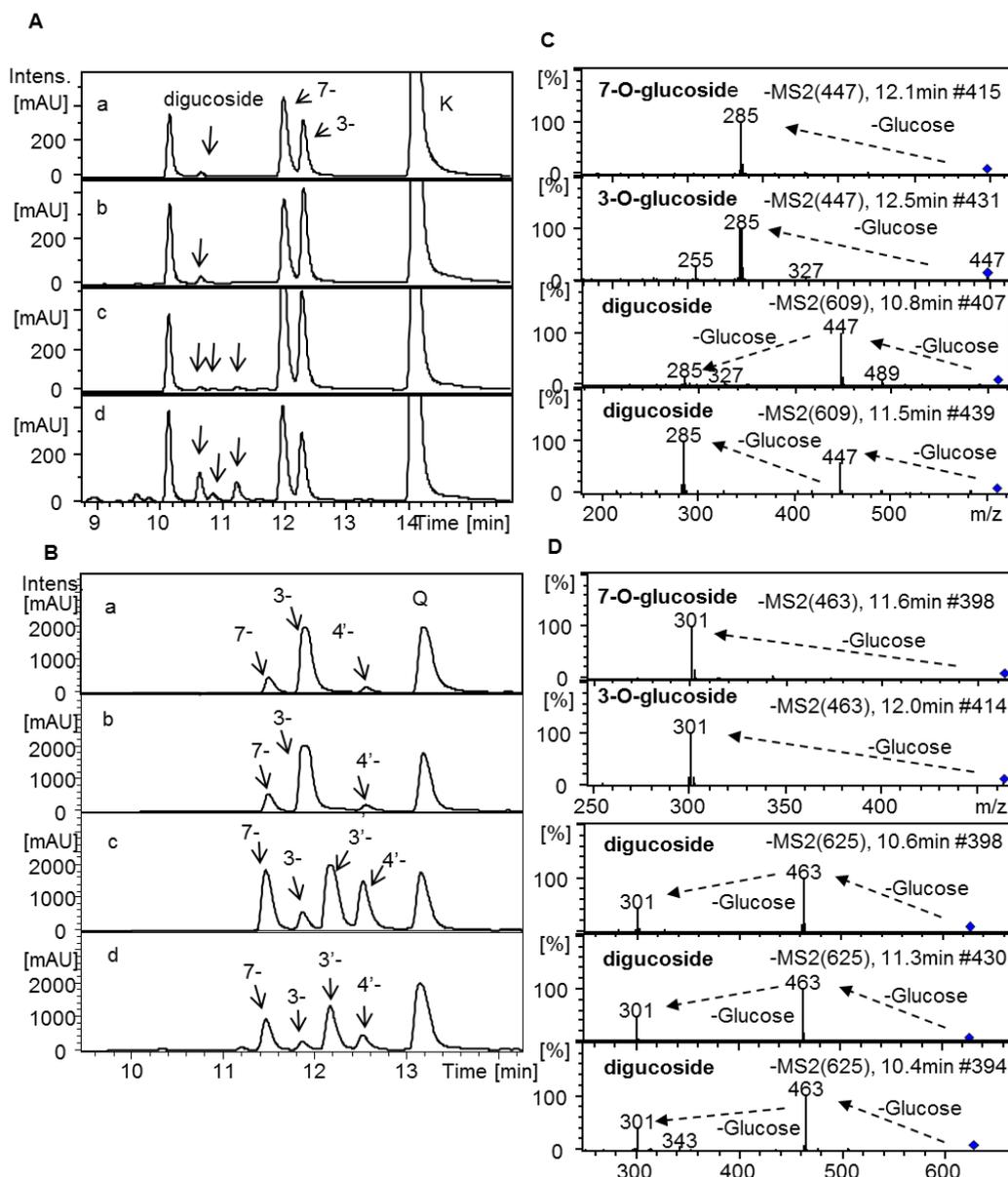
**Figure 36.** Identification of enzymatically formed products by LC-MS. 1-Naphthol (A), and 2-naphthol (B) were incubated with UDP-glucose and recombinant FaGT07876a/b and empty vector control. 3-Hydroxycoumarin (C) was incubated with UDP-glucose and recombinant FaGT00126 and FaGT26345 and empty vector control.

### 3.3.4 Regioselectivity of selected FaGTs

Substrate specificity of UGTs is not absolute but they show regioselectivity (Vogt and Jones, 2000). This regioselectivity offers an important means to overcome the limitations of chemical synthesis of small molecule glycosides (Lim et al., 2004). To explore the regioselective glucosylation of FaGTs, we used kaempferol and quercetin as substrates because their glucosides have important medicinal properties and the precursor

metabolites provide a complex structure for regioselective glucosylation (Figure 37). The substrate screenings indicated that FaGT07876a/b, 26345 and 00126 probably react on the hydroxyl group at position 3, 6 and 7 of flavonols. All four FaGTs formed two major products which were identified as the 3-, and 7-glucoside (Figure 37A). FaGT07876a, similar to FaGT26345 and FaGT00126, produced less 3-than the 7-glucoside, but FaGT07876b formed more 3-glucoside when kempferol was used as substrate (Figure 37A). Additional analyses showed that only trace amounts of diglucosides were produced by FaGT07876a/b and FaGT00126, whereas FaGT26345 readily formed three diglucosides (Figure 37A). The product ion spectra of the diglucosides revealed that both glucose moieties are located at different hydroxy groups of the flavonols, because only in this case the product ions  $m/z$  477 and 463 can be detected (Ablajan et al., 2006).

Using authentic reference compounds, the major product formed by both FaGT07876a/b was identified as the 3-glucoside while the minor products were assigned as 7-, 3'-, and 4'-glucoside when quercetin was used as substrate (Figure 37B) (Lim et al., 2004; Shao et al., 2005). However, a totally different product pattern was obtained for FaGT26345 and FaGT00126 which formed only a trace amount of the 3-glucoside. Quercetin-7-, and 3'-glucoside were identified as the major products, in accordance with the results of the substrate screenings. Further, the abundance of the 4'-O-glucoside produced by FaGT00126 was much higher than that of the other three proteins (Figure 37B).

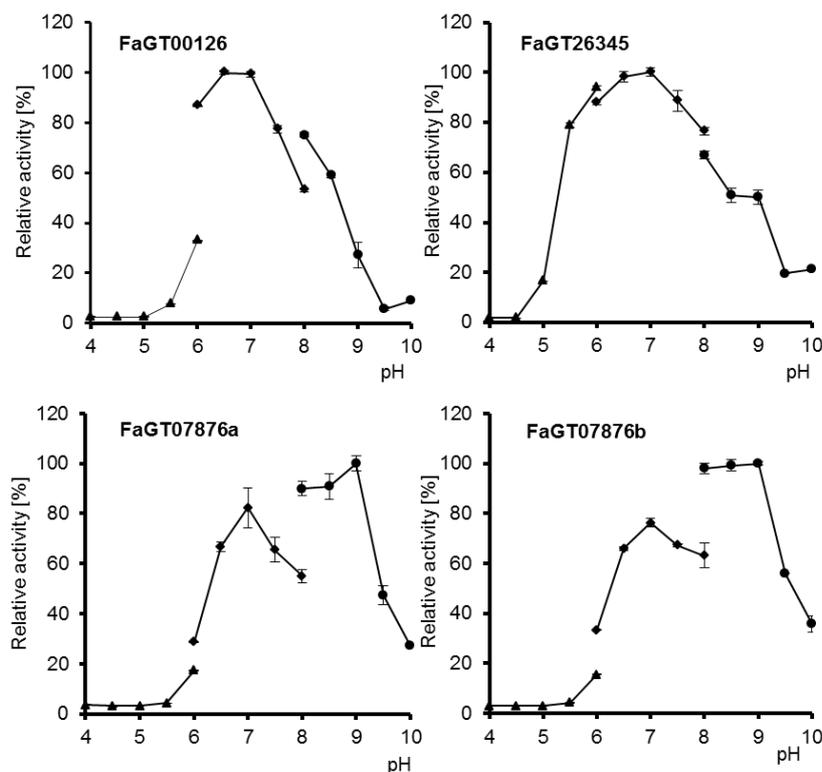


**Figure 37.** Regioselective glucosylation by FaGTs. Kaempferol (**A**), and quercetin (**B**) was incubated with UDP–glucose and recombinant FaGT07876a (a), FaGT07876b (b), FaGT00126 (c) and FaGT26345 (d) and subsequently analysed by LC-MS. 3-O-glucoside and 7-O-glucoside and diglucosides of kaempferol (**C**), and quercetin (**D**) were clearly distinguishable by their product ion spectra (MS2) according to Ablajan et al. (2006). Peak at 10.1 min is an impurity of kaempferol.

### 3.3.5 Substrate preference and kinetic parameters

The optimum reaction conditions for FaGT26345 and FaGT00126 were established with 3-hydroxycoumarin as acceptor substrate, whereas the assay conditions for FaGT07876a/b were optimized using 1-naphtol as substrate. The highest activity of FaGT00126 and

FaGT26345 was detected at optimum pH 6.5 and 7.0, respectively (Figure 3.8). Both FaGT07876a and b showed two different pH optima (pH 7.0 and 9.0) depending on the buffer. The optimal temperature for FaGT00126, FaGT26345, FaGT07876a and b was 35, 40, 30, and 30°C, respectively.



**Figure 38.** The pH optima. The purified enzymes (FaGT00126, FaGT07876a/b, and FaGT26345) were incubated at optimal temperature and different pH values. Citric acid, sodium phosphate, and Tris-HCl buffer was used for pH 3 to 6, pH 6 to 8, and pH 8 to pH 10, respectively. Product formation was determined by radiochemical analysis with UDP-[U-<sup>14</sup>C] glucose.

Kinetic properties were determined for selected substrates in the linear range of the enzymatic reaction (2 µg of protein, 30 min reaction time). The apparent  $K_M$  values of FaGT00126, 26345, 07876a and b for kaempferol were 110.8, 22.4, 199.1 and 108.8 µM, respectively (Table 10). The specificity constant  $k_{cat}/K_M$  of FaGT00126 for quercetin ( $K_M$  of  $1.06 \pm 0.03 \mu\text{M}$ ) is 32-fold, 22-fold and 65-fold higher than that of FaGT26345, FaGT07876a and b, respectively. The  $V_{max}$  value of FaGT07876a for phloroglucinol is 3-fold higher than that of FaGT07876b but both enzymes show a similar specificity constant  $k_{cat}/K_M$  value ( $75.6$  and  $68.6 \text{ M}^{-1} \text{ s}^{-1}$ ) for this substrate (Table 10).

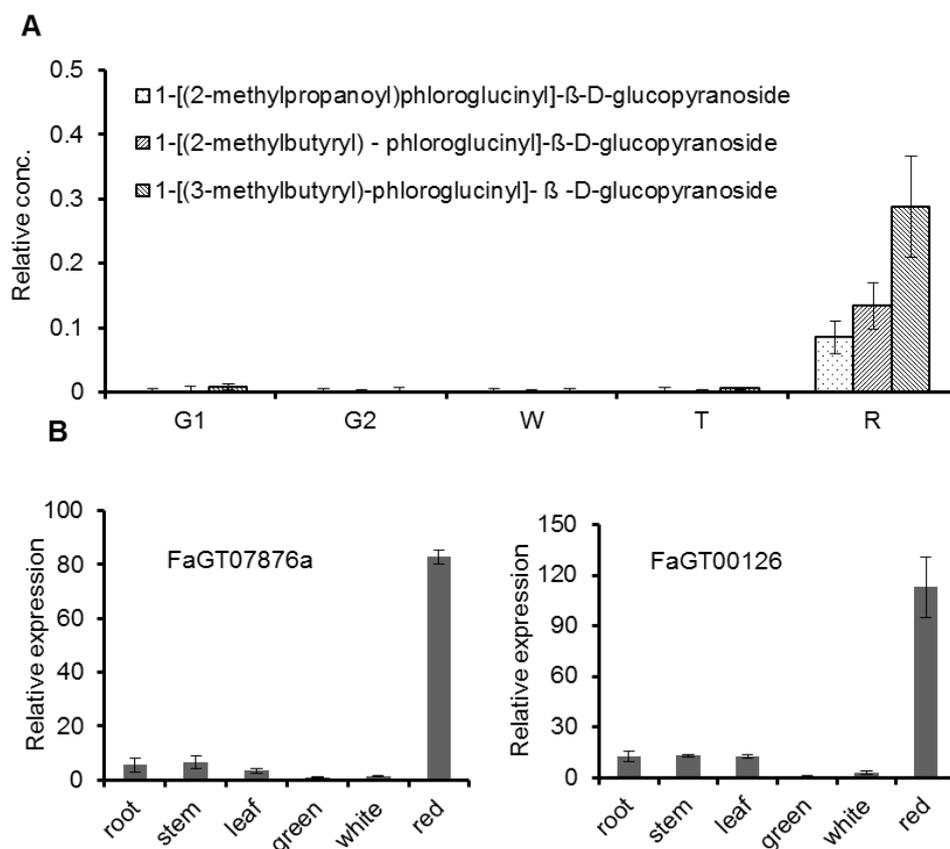
## Results

Table 10. Kinetic data of FaGT00126, FaGT07876a/b, and FaGT26345 with selected substrates.

Proteins	substrate	$K_M(\mu\text{M})$	$V_{\text{max}}(\text{nKat}/\text{mg})$	$K_{\text{cat}}/K_M(\text{M}^{-1}\text{s}^{-1})$
FaGT26345	3-Hydroxycoumarin	38.79±7.40	0.84±0.04	1633.60
	Kaempferol	110.79±39.93	1.40±0.31	947.74
	Quercetin	24.03±15.41	0.36±0.11	1123.60
FaGT00126	3-Hydroxycoumarin	217.11±11.62	1.34±0.03	462.90
	Kaempferol	22.36±3.15	1.46±0.09	4897.14
	Quercetin	1.06±0.03	0.51±0.01	36084.91
FaGT07876a	1-Naphtol	42.50±6.80	0.60±0.02	1058.82
	Kaempferol	199.05±49.02	2.45±0.22	923.13
	Quercetin	37.54±15.27	0.82±0.14	1638.25
	Furaneol	1582.55±197.76	0.33±0.03	15.64
	Phloroglucinol	744.56±52.27	0.75±0.03	75.55
FaGT07876b	1-Naphtol	35.84±2.48	0.61±0.01	1276.51
	Kaempferol	108.50±23.97	0.89±0.06	615.21
	Quercetin	98.75±23.72	0.73±0.05	554.43
	Furaneol	283.00±33.52	0.12±0.01	31.80
	Phloroglucinol	284.38±23.07	0.26±0.01	68.57

### 3.3.6 FaGTs expression correlates with accumulation of APGs

FaGT00126 and FaGT07876a show very similar enzymatic activities (Figure 35) and only very few amino acids exchanges (Appendix Figure 8 and 9) when compared with FaGT26345 and FaGT07876b, respectively. Thus, only the transcript level of gene07876 and 00126 was determined by real-time PCR in vegetative tissues (root, stems, and leaves) and in fruits at different developmental stages in *F. x ananassa* (Figure 39). The transcript levels of gene07876 and 00126 strongly increase during ripening and are highly abundant in the red ripe fruit, showing an 83- and 113-fold higher expression in the ripe fruit when compared with the green stage (Figure 39). Both genes are poorly expressed in vegetative tissues, such as root, stems, and leaves. The observation that both genes are predominantly expressed in fully ripe fruit receptacles suggests that the encoded FaGTs are associated with metabolic pathways strongly activated in ripe receptacle tissue. LC-MS analysis showed that APG glucosides accumulate during strawberry fruit ripening but cannot be detected in immature fruits (Figure 39). The amount of APG glucosides peaks in totally mature fruit receptacles and strongly correlates with the expression of gene07876 and 00126 in *F. x ananassa* (Figure 39).

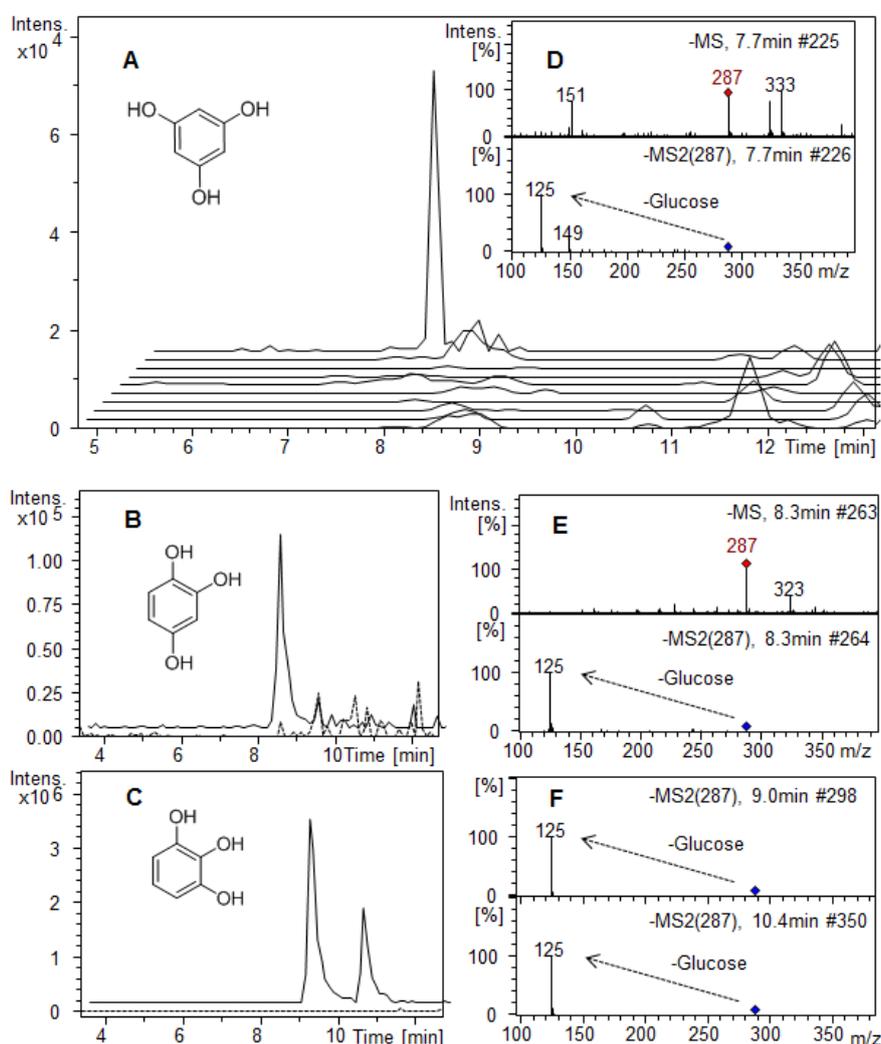


**Figure 39.** Relative concentration of APG glucosides in strawberry fruit of different developmental stages (**A**) and relative mRNA expression levels of FaGT00126 and FaGT07876a in root, stems, leaves, green, white and red/ripe fruit (**B**) determined by real-time PCR with *FaRib413* used for normalization. The expression level in green fruit was set at 1 as a reference (means  $\pm$  SE six replicates with two sets of cDNAs).

### 3.3.7 Screening of phloroglucinol glucosyltransferase activity

To assess whether FaGT07876 is involved in the APG biosynthesis pathway, we tested the ability of recombinant FaGT07876 to glucosylate commercially available phloroglucinol (1,3,5- trihydroxybenzene), 1,2,4-benzenetriol, and pyrogallol (1,2,3-trihydroxybenzene), structural homologues of APG aglycones. The enzymatically formed products were analyzed by LC-MS. The FaGT07876a/b proteins successfully glucosylated all three trihydroxybenzene isomers (Figure 40) as the fragmentation of the pseudomolecular ions  $[M-H]^-$   $m/z$  287 of the products proceeded by loss of the glucose substituents (neutral loss of 162 dalton) and yielded the fragment at  $m/z$  125 (negatively charged ion of phloroglucinol, 1,2,4-benzenetriol and pyrogallol). When pyrogallol was used as substrate, two main products

were formed. The glucosides showed similar mass spectra but eluted at different retention times (Figure 40).

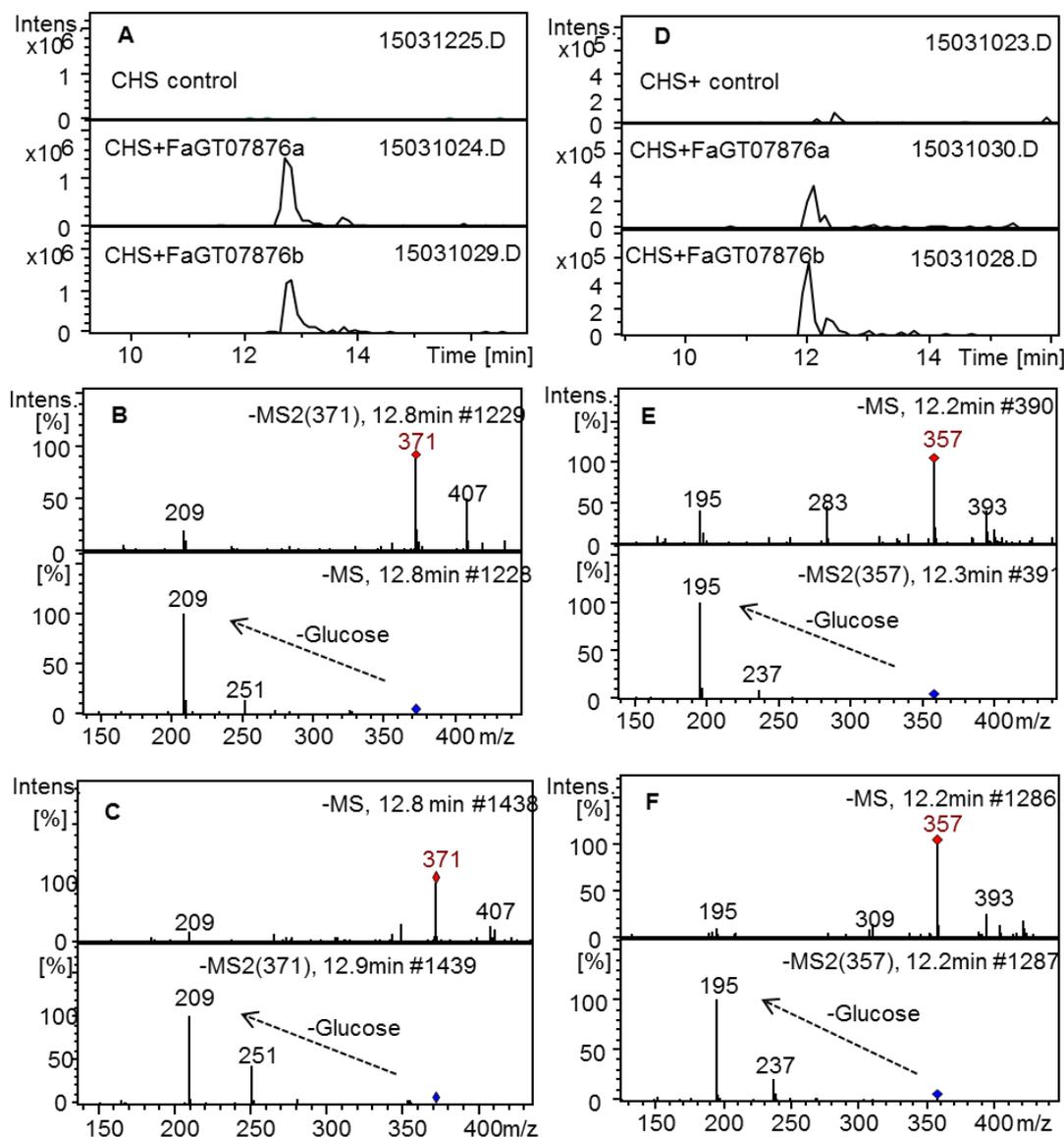


**Figure 40.** Enzyme activity screening of FaGT00126, FaGT07876a, and FaGT26345 towards phloroglucinol (**A**). FaGT07876 is the only enzymes tested which has the capacity to glucosylate phloroglucinol. LC-MS analysis of products formed from 1,2,4-benzenetriol (**B**), and pyrogallol (**C**), after incubation with UDP–glucose and recombinant FaGT07876a. The fragmentation of the pseudomolecular ions  $[M-H]^-$   $m/z$  287 of the products proceeded by loss of the glucose substituents (neutral loss of 162 dalton) and yielded the fragment at  $m/z$  125 (negatively charged ion of phloroglucinols **D**, 1,2,4-benzenetriol **E**, and pyrogallol **F**).

### 3.3.8 Total enzymatic synthesis APG glucosides

In strawberry receptacle, APG aglycones phloroisovalerophenone (PIVP) and phloroisobutyrophenone (PIBP) are formed by bifunctional chalcone synthase (CHS2) from

malonyl-CoA and either isovaleryl-CoA or isobutyryl-CoA, respectively (refer to 3.1) and become glucosylated. Because PIVP and PIBP are not commercially available we incubated recombinant CHS2 and FaGT07876a or b with isovaleryl-CoA/isobutyryl-CoA, malonyl-CoA and UDP-glucose to test whether FaGT07876a/b can glucosylate APG aglycones *in vitro*.

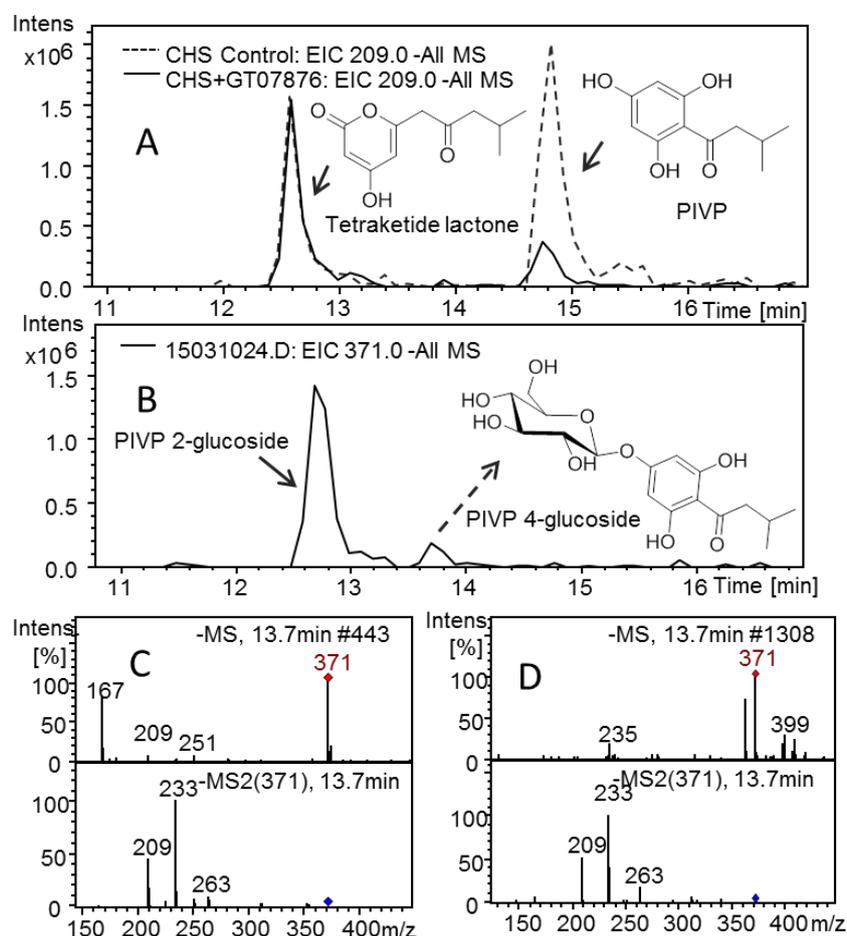


**Figure 41.** Total enzymatic synthesis of APG glucosides. Ion traces  $m/z$  371 (A) and  $m/z$  357 (D) of products formed by CHS2 without (CHS control) and CHS2 with FaGT07876a and b, malonyl-CoA, and isovaleryl-CoA (A) or isobutyryl-CoA (D), and UDP-glucose (reaction time 20 h). Mass spectra and product ion spectra of  $m/z$  371 and  $m/z$  357 of the enzymatically formed product PIVP 2-glucoside (B) and PIBP 2-glucoside (E) and of authentic APG glucosides from strawberry fruit (PIVP 2-glucoside, C and PIBP 2-glucoside F).

After incubation overnight, PIVP 2-glucoside and PIBP 2-glucoside were successfully detected by LC-MS (Figure 41A,D). The mass spectral data and retention times were identical with those of the APG glucosides isolated and identified in strawberry fruit (Figure 41B, C, E, and F). The outcome of this experiment provides strong evidence that FaGT07876a/b catalyze the last step of APG glucoside biosynthesis.

### 3.3.9 Identification of a novel APG-glucoside based on the catalyzed reaction

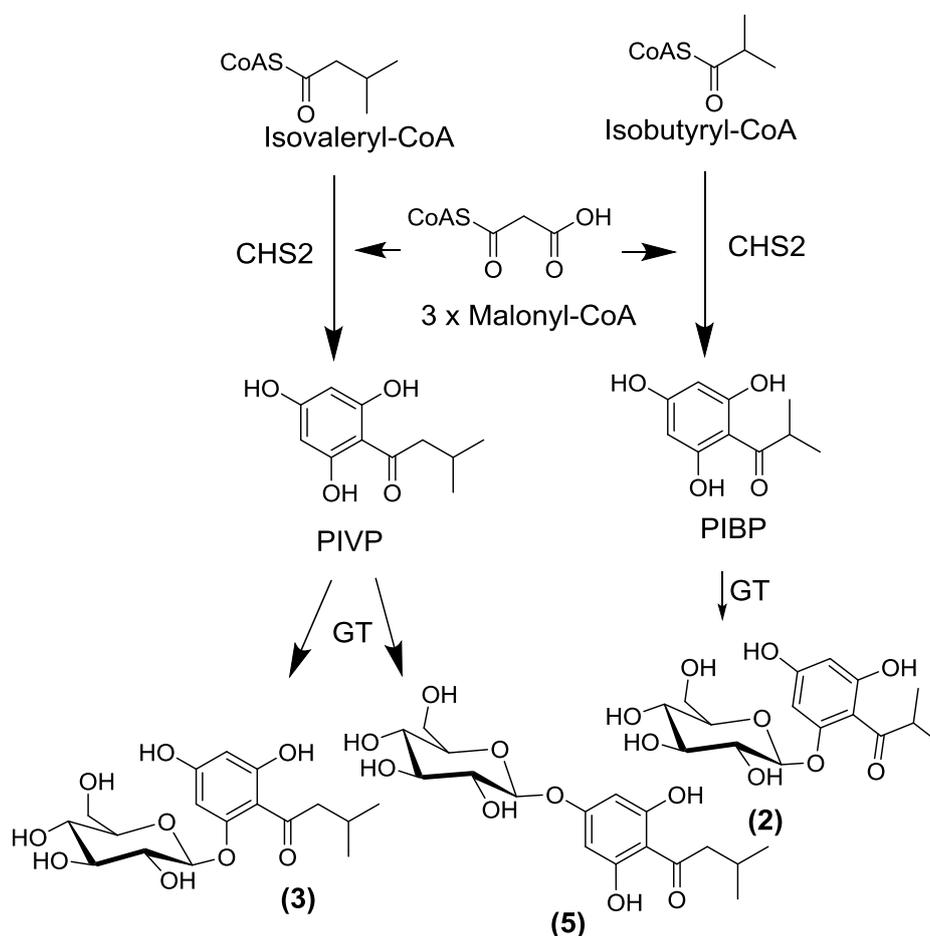
In addition to PIVP 2-glucoside an additional APG-glucoside was found when isovaleryl-CoA, malonyl-CoA and UDP-glucose were co-incubated with CHS2 and FaGT07876a/b (Figure 42). The MS, MS2 and retention time (13.7 min) of the additional product was identical with the data of a metabolite isolated from strawberry fruit (Figure 42C and D).



**Figure 42.** Identification of a novel enzymatically formed product by total enzymatic synthesis of APG glucosides. Isovaleryl-CoA, malonyl-CoA and UDP-glucose was incubated with only CHS2 (A, dashed line) and with both CHS and FaGT07876a (A, continuous line). PIVP 2-glucoside and a novel product, putatively identified as PIVP 4-glucoside was identified at

ion trace  $m/z$  371 by LC-MS (**B**). MS, MS2 and retention time of the enzymatically formed PIVP-4-glucoside (**C**) is identical to the data of a metabolite isolated from strawberry (**D**).

Because CHS formed tetraketide lactone and PIVP (Figure 42A dashed line) in the first step of the reaction sequence but only PIVP was consumed by FaGT07876a/b (Figure 42A continuous line) we assumed that the extra product is PIVP 4-glucoside in accordance with its MS and MS2 data (Figure 42C) that are similar to those of PIVP 2-glucoside. Since glucosylation increases the polarity of the aglycones associated with the sugar moiety the tetraketide lactone was excluded as aglycone (Figure 42A and B; Aherne and O'Brien, 2002). PIVP 4-glucoside has not been reported in plants, yet. The proposed pathway of APG glucosides in strawberry is shown in Figure 43.



**Figure 43.** APG biosynthesis pathway in strawberry

### **3.4 UGT71C3 glucosylates the key flavor compound 4-hydroxy-2,5-dimethyl-3(2H)-furanone (HDMF)**

Although the ubiquitous UGT family has been intensively studied for many years (Gachon et al., 2005), and enzymes involved in the glycosylation of plant products such anthocyanidins, flavonoids, naphthols, terpenes (Cheng et al 1994; Almeida et al., 2007; Griesser et al., 2008a,b , Bönisch et al., 2014a/b), and plant hormones (Poppenberger et al., 2005) were functionally characterized little is known about the glycosylation of strawberry aroma chemicals.

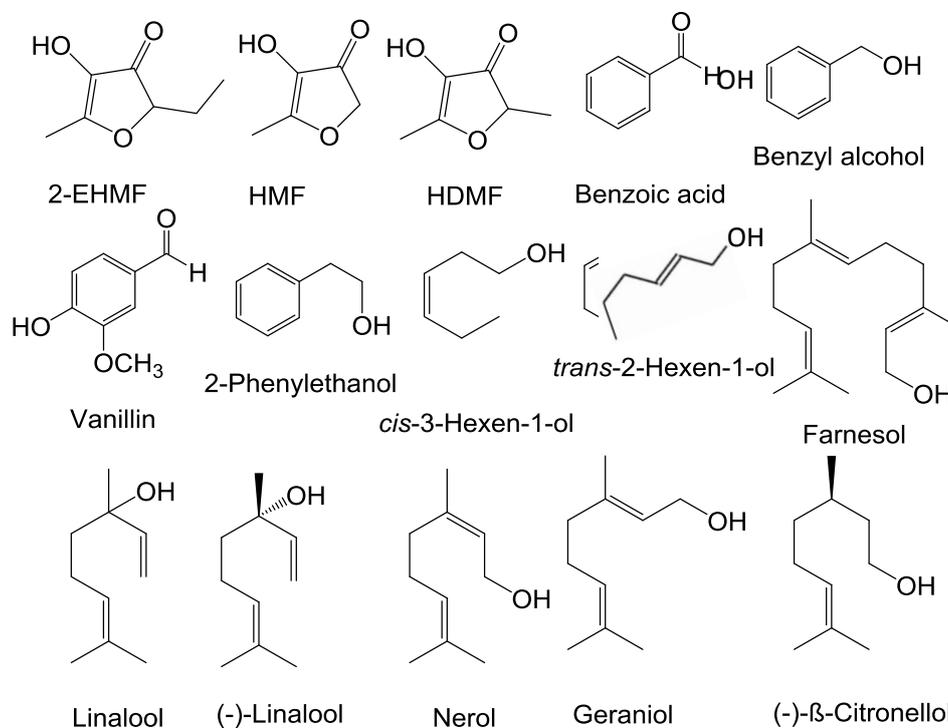
#### **3.4.1 Cloning and functional expression of putative FaGTs**

In order to find UGTs function in glucosylation of aroma chemicals in strawberry, eight ripening related UGT genes were selected from transcriptome data sets (kindly provided by Katja Schulenburg and Katrin Franz-Oberdorf) obtained from fruit receptacles and achenes of three *F. vesca* varieties (Reine des Valeés, Yellow Wonder, and Hawaii 4) of different developmental stages (green, white and ripe), which all showed a ripening related expression pattern in at least one of the tissue of one of the three *F. vesca* varieties. Full-length sequences of selected UGTs, including two alleles of two genes (FaGT07876a/b, FaGT24225a/b), were amplified using cDNA from red fruit of *F. x ananassa* cv. Elsanta as a template and ligated into the pGEX-4T1 vector. Their corresponding proteins (FaGT07876a/b, FaGT24225a/b, FaGT00126, FaGT26345, FaGT26342, FaGT22709, FaGT24224 and FaGT24226) were heterologous expressed as a thioredoxin-GST-Tag-fusion protein in *E. coli*. FaGT24225a/b and FaGT07876a/b are allelic proteins. By comparing the deduced amino acid sequences of FaGT24225a and FaGT07876a enzyme, we find that there are 12 and 2 amino acid differences with their corresponding allelic protein FaGT24225b and FaGT07876b, respectively (Appendix Figure 7 and Appendix Figure 9).

#### **3.4.2 Screening of volatile substrates by ten glycosyltransferases**

Substrates screening of FaGT07876a/b, FaGT24225a/b, FaGT00126 , FaGT26345 , FaGT22709, FaGT24224 and FaGT24226 was performed *in vitro* with UDP-glucose as donor substrate and 13 volatile acceptors, which are known to be glycosidically bound and present in strawberry (Ubeda et al., 2012), or of commercial importance (Figure 44). Vanillin was the

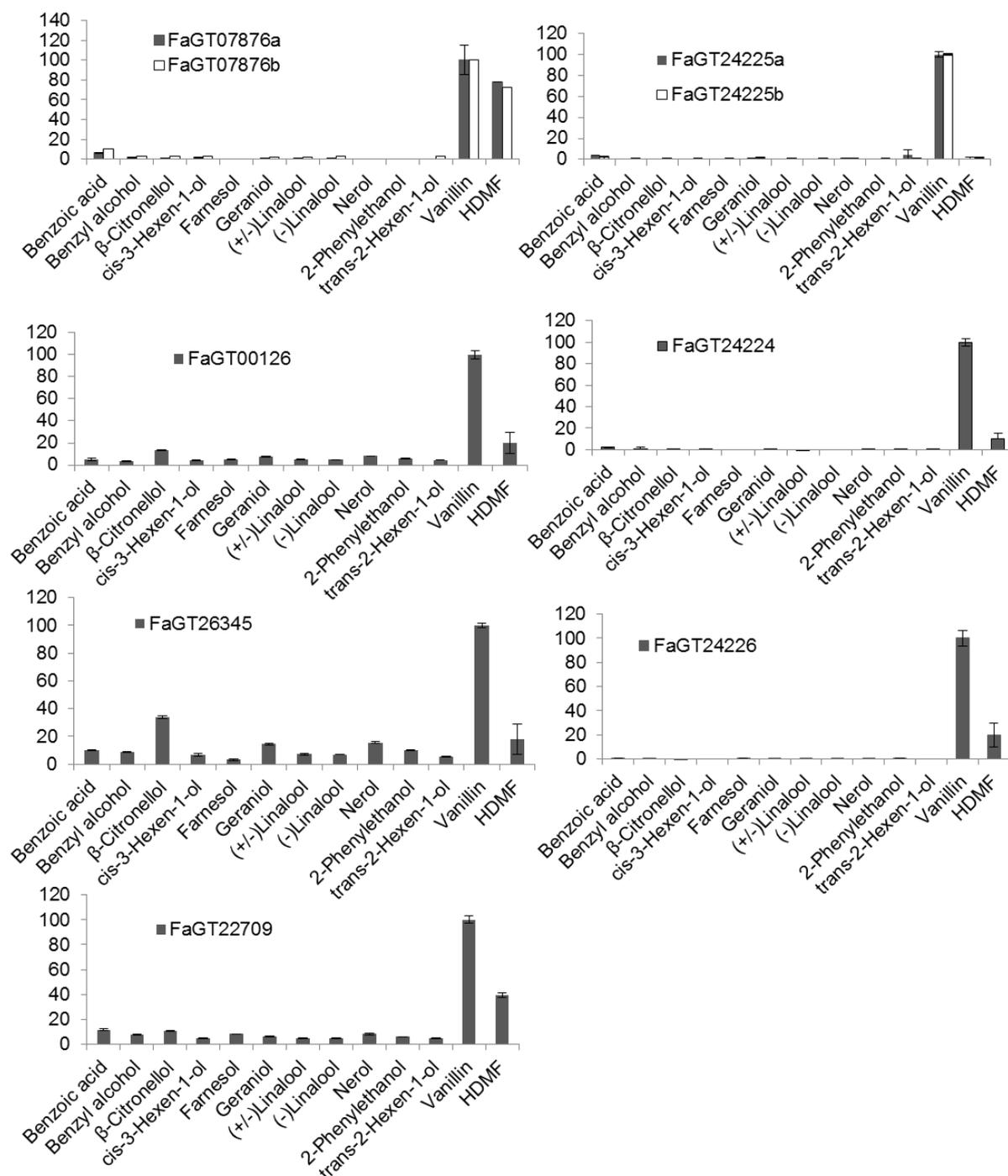
preferred substrate of all nine recombinant FaGT enzymes but the allelic proteins of FaGT07876a/b FaGT00126, FaGT22709, and FaGT26345 also glucosylated HDMF (furanol) (Figure 45).



**Figure 44.** Chemical structures of 2-ethyl-4-hydroxy-5-methyl-3(2H)-furanone (2-EHMF; one of two tautomers of homofuraneol), 4-hydroxy-5-methyl-3-furanone (HMF; norfuraneol), and 4-hydroxy-2,5-dimethyl-3(2H)-furanone (HDMF, furaneol), benzoic acid and additional aroma chemicals used for the substrate screening of FaGTs.

Besides, FaGT07876a/b showed minor catalytic activity towards benzoic acid (6- 10% relative activity), however, the other alcohols were not converted (< 3%). Three active proteins FaGT00126, FaGT26345 and FaGT22709 showed a similar substrate spectrum. In addition to vanillin and HDMF, FaGT00126, FaGT26345 and FaGT22709 were also able to use benzyl alcohol, β-citronellol, cis-3-hexen-1-ol, farnesol, geraniol, (+/-)-linalool, (-)-linalool, nerol, 2-phenylethanol and trans-2-hexen-1-ol as acceptor molecules but with low efficiency (Figure 45). FaGT24225a/b, 24224 and 24226 showed a more limited substrate spectrum as they preferred vanillin but could not efficiently glycosylated any other tested substrate (< 3%).

## Results

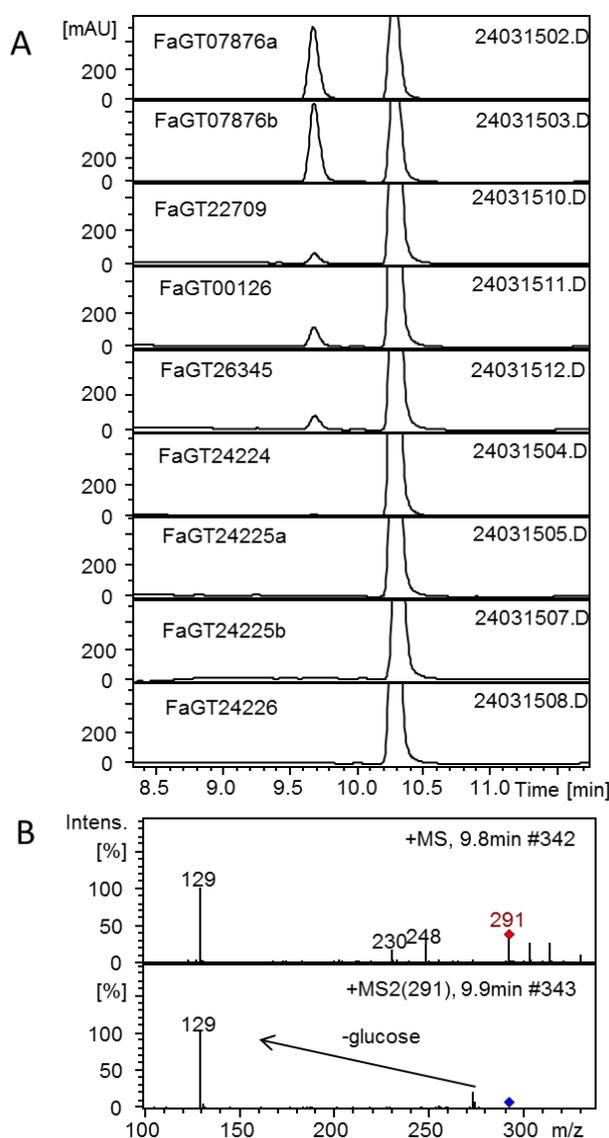


**Figure 45.** Screening of recombinant FaGT proteins encoded by ripening related genes from strawberry fruits (*F. x ananassa*) with aroma chemicals. Relative activity is shown; the activity of vanillin was set as 100%

### 3.4.3 FaGT07876 shows HDMF glycosylation activity

To verify the formation of HDMF-glucoside by LC-MS, FaGT07876a/b, FaGT24225a/b, FaGT00126, FaGT26345, FaGT22709, FaGT24224 and FaGT24226 were affinity purified and incubated with HDMF overnight. FaGT07876a/b efficiently glucosylated HDMF (Figure 46A).

The main product was identified as HDMF-glucoside by LC-MS (Figure 46B). Besides, FaGT00126, 26345 and 22709 also formed low amounts of HDMF-glucoside consistent with the results of the substrate screening. Thus, FaGT07876a/b was used for further study.

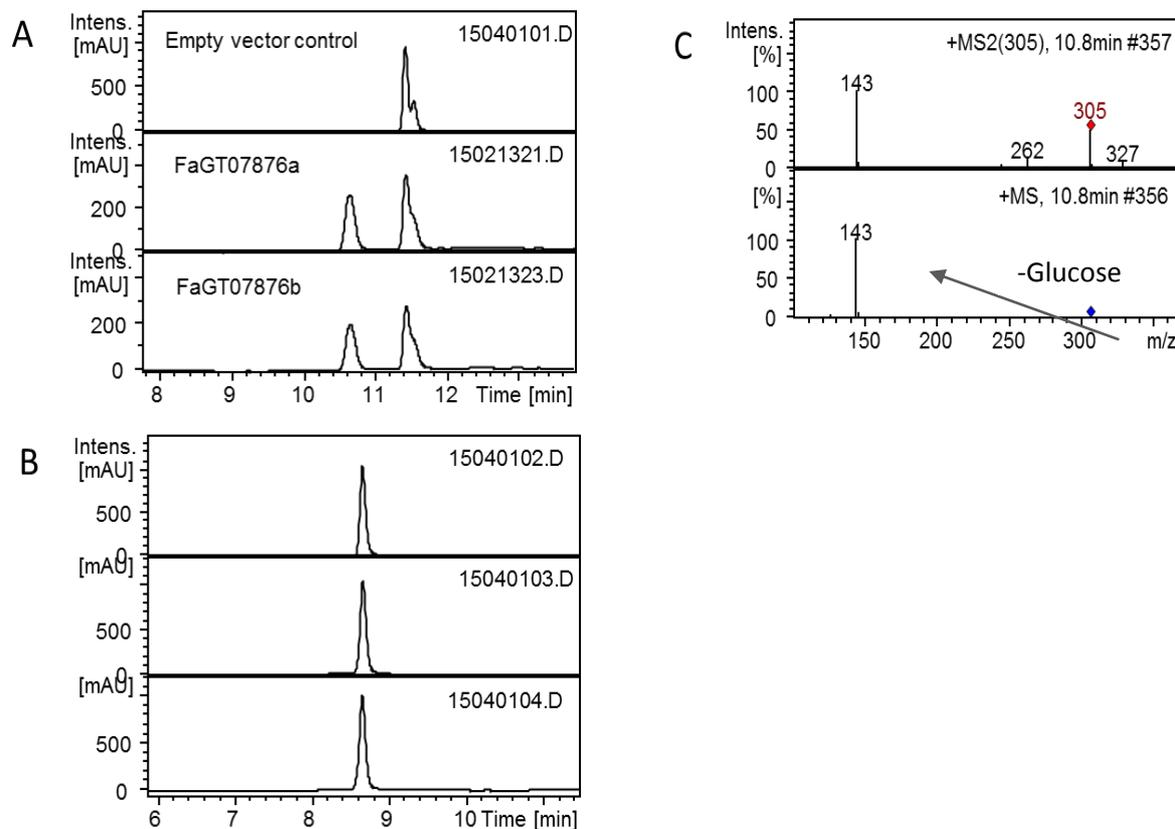


**Figure 46.** Screening of nine putative FaGT enzymes from strawberry fruit (*F. x ananassa*) with HDMF. HDMF-glucoside was identified by LC-MS (UV trace of at 280 nm (A) and MS and MS2 of HDMF-glucoside (B).

#### 3.4.4 Glucosylation of EHMf and HMF

The furanones EHMf and HMF are structural homologues of HDMF (Figure 44) and exhibit a caramel-like flavor similar to the key strawberry flavor compound HDMF. They were also tested as substrates because of their structural similarity and commercial importance. In addition to HDMF both FaGT07876a and b have the capacity to glucosylate EHMf with similar efficiency *in vitro* (Figure 47A). The formed glucoside was identified by LC-MS (Figure

47C). However, both FaGT07876a and b could not convert the structural homologue HMF (Figure 47B).

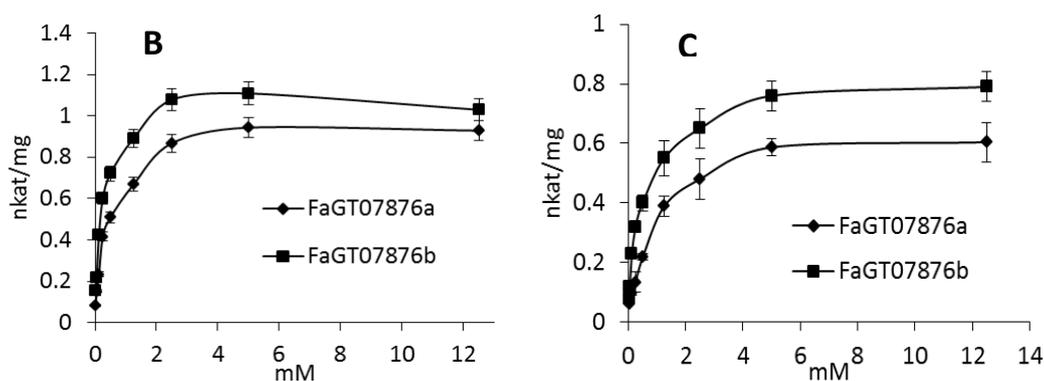


**Figure 47.** Identification of enzymatically formed products formed by recombinant FaGT07876a and b from EHMf (**A**) and HMF (**B**). MS and MS2 of EHMf-glucoside(**C**).

### 3.4.5 Kinetic properties of the recombinant FaGT07876a and b

Kinetic properties of FaGT07876a and b were determined in the linear range of the enzymatic reaction. Two  $\mu\text{g}$  of the proteins, 2 hours reaction time at  $30^\circ\text{C}$  at the optimal pH 7.0 in Tris-HCl buffer and various concentrations of HDMF and EHMf were used (refer to 3.3). The products formed from HDMF and EHMf were measured by LC-MS. The apparent  $K_M$  value for HDMF of FaGT07876a and b is 899.8 and 405.8  $\mu\text{M}$ , respectively (Figure 48). The specificity constant  $k_{\text{cat}}/K_M$  of FaGT07876b for EHMf is 2-fold higher than that of FaGT07876a, due to the  $K_M$  value of 213.2  $\mu\text{M}$ . The  $V_{\text{max}}$  value of FaGT07876b for both HDMF and EHMf is higher than those of FaGT07876a (Figure 48B, C).

A	Proteins	Substract	$K_M(\mu\text{M})$	$V_{\max}(\text{nKat}/\text{mg})$	$K_{\text{cat}}/K_M(\text{M}^{-1}\text{S}^{-1})$
	FaGT07876a	HDMF	$899.8\pm 124.9$	$0.6629\pm 0.0261$	55.254
		EHMF	$396.8\pm 52.8$	$0.9725\pm 0.0315$	183.814
	FaGT07876b	HDMF	$405.8\pm 58.8$	$0.7882\pm 0.0279$	145.675
		EHMF	$213.2\pm 26.9$	$1.0993\pm 0.0315$	386.714

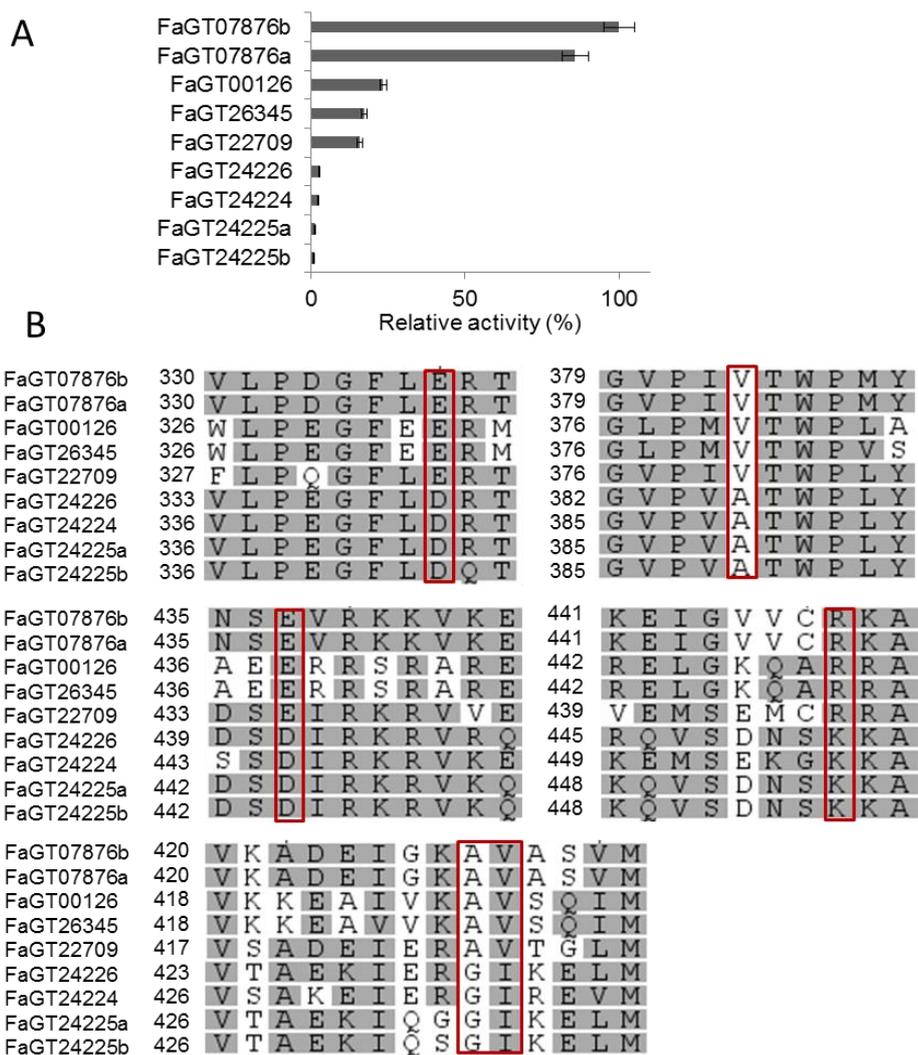


**Figure 48. Kinetic data of FaGT07876a and b (A) for EHMF (B) and HDMF (C).**

### 3.4.6 Site-directed mutagenesis of FaGT24224

To examine the molecular basis for the HDMF glucosylation activity of plant UGTs, we compared the amino acid sequences of five FaGTs that showed apparently HDMF glucosylation activity (>15% relative activity; Figure 49A) with four FaGTs that exhibited negligible HDMF glucosylation activity (<3% relative activity) using ClustalW multiple alignment. Interestingly, a highly conserved Val residue in the center of the PSPG box of those five FaGTs with significant HDMF glucosylation activity was replaced with Ala in all FaGTs with negligible HDMF glucosylation activity (Figure 49B). Similarly, five additional conserved Glu, Glu, Arg, Ala and Val residues in all FaGTs with significant HDMF glucosylation activity were replaced by Asp, Asp, Lys, Gly and Ile, respectively at the equivalent position of four FaGTs with negligible HDMF glucosylation activity (Figure 49B). These loci are of special interest, since these differences are assumed to be related to the HDMF glucosylation specific enzyme activity of FaGTs.

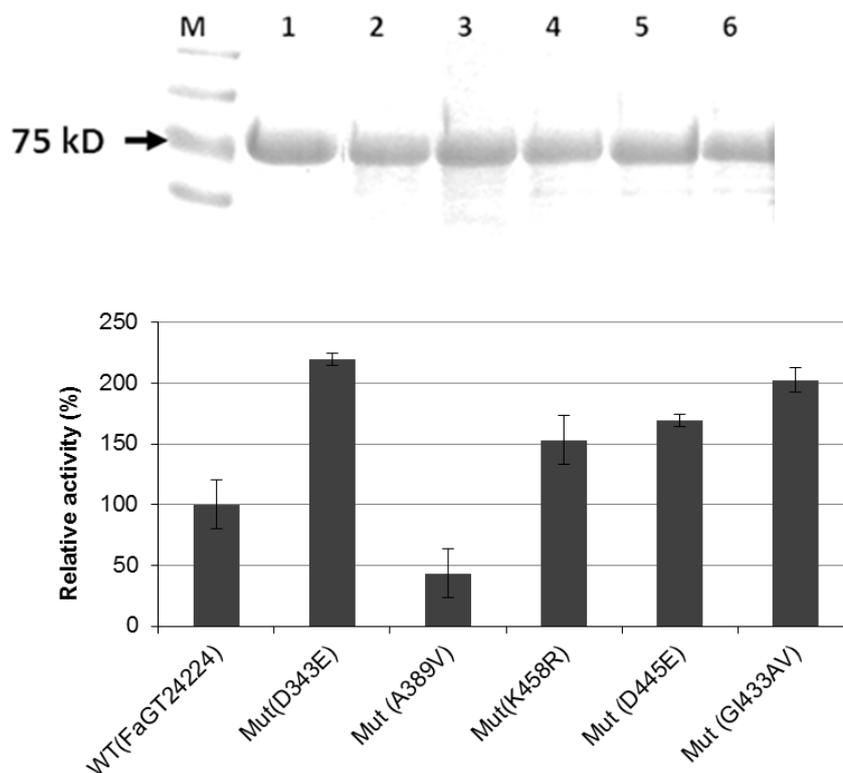
## Results



**Figure 49.** Relative HDMF glucosylation activity (**A**) and ClustalW multiple alignments of FaGTs with apparently or negligible HDMF glucosylation activity (**B**).

To validate the importance of these unique residues for the HDMF glucosylation activity of FaGTs, single site-directed mutagenesis of D343E, A389V, K458R, D445E, and double mutagenesis of G433A I434V was carried out and catalytic efficiency of both wildtype FaGT24224 and mutant enzymes was compared. A double mutant of G433A I434V was generated because GI and AV are close neighbors (Figure 49). The results showed that single substitutions D343E, K458R, D445E and the double mutant of G433A I434V all displayed enhanced HDMF glucosylation activity, especially mutant D343E and the double mutant G433A I434V exhibited a more than 2-fold higher activity towards HDMF than that of the wildtype FaGT24224 (Figure 50). By contrast, a single mutation in the center of the PSPG box, where Ala 389 of FaGT24224 was replaced with Val (the corresponding amino acid is

conserved in FaGTs with apparently HDMF glucosylation activity) resulted in a decrease of the HDMF glucosylation activity (43% of the wildtype control; Figure 50).

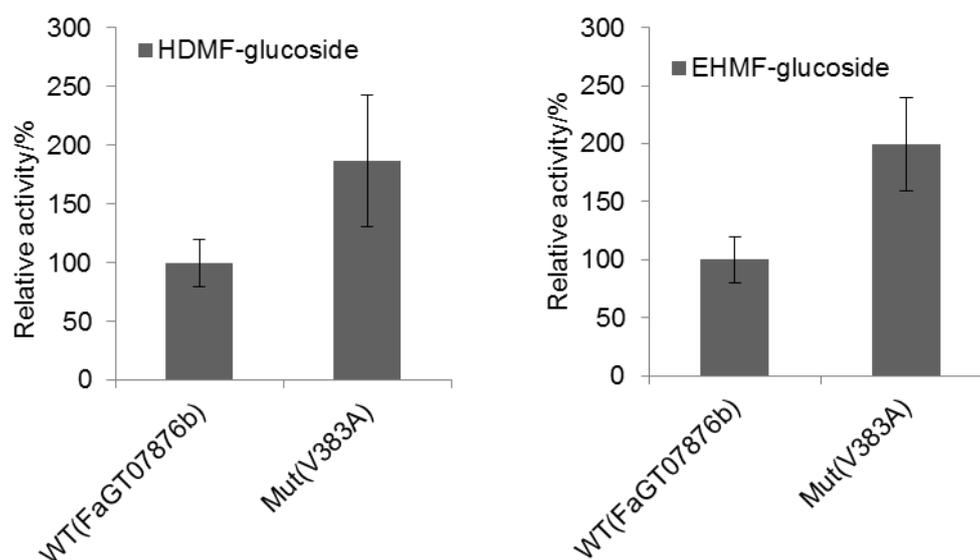


**Figure 50.** SDS-PAGE analysis of the wild type protein and mutants (A) and relative activity (B). M: Marker; 1: wild type FaGT24224; 2: Mutant D343E (FaGT24224); 3: A389V (FaGT24224); 4: K458R (FaGT24224); 5: D445E (FaGT24224); 6: Double mutant G433A I434V (FaGT24224).

### 3.4.7 Mutagenesis of residue 383 of FaGT07876b

Mutation A389V localizes within the PSPG-box, the highly conserved region that is considered to code the sugar-binding site of UGTs (Vogt and Jones, 2000). A single substitution of FaGT24224, where Ala 389 was replaced with a Val residue, resulted in a significant decrease of the HDMF glucosylation activity (Figure 50). Thus, we inferred that a Val-389 substitution of FaGT24224 appeared to play an adverse impact on HDMF glucosylation activity. On the other side, the most potent UGT for producing HDMF-glucoside (FaGT07876b; Figure 49) has a Val residue at the equivalent position which might limit the catalytic function of FaGT07876b towards HDMF.

In order to investigate whether the Val residue within the PSPG-box of FaGT07876b affects the enzymatic activity towards HDMF, site-directed mutagenesis and catalytic activity of both wildtype FaGT07876b and mutant enzyme was carried out. The result shows that the V383A mutant indeed displayed enhanced HDMF glucosylation activity. The V383A mutant of FaGT07876b was about 1.87- fold more active than the wild-type enzyme in the glucosylation of HDMF (Figure 51). Similarly, the V383A mutation also enhanced the glucosylation activity towards EHMf 2-fold compared with that of the wild type (Figure 51).



**Figure 51.** Relative activity of wildtype FaGT07876b and mutant enzyme (V383A) in towards HDMF (left) and EHMf (right).

## 4. Discussion

### 4.1 Acylphloroglucinol biosynthesis in strawberry fruit

Plants produce numerous structurally diverse metabolites that have beneficial effects on human health (De Luca et al., 2012; Saito et al., 2010). Phenolic compounds are the most widely distributed secondary metabolites in the plant kingdom, and there is increasing evidence that consumption of a variety of phenolic compounds may lower the risk of serious health disorders (Visioli et al., 2011). Although the basic reactions of the phenolics biosynthetic pathways in plants have been intensively analyzed the regulation of their accumulation and flux through the pathway is not that well established. Recent research revealed novel candidate genes that might affect accumulation of flavonoids and anthocyanins in strawberry fruit by comparing the transcript patterns of different *F. x ananassa* genotypes combined with metabolite profiling analysis (Ring, et al., 2013).

#### 4.1.1 Candidate genes function in anthocyanin accumulation

Four genes (*expansin-A8-like*, *SRG1-like*, *ephrin-A1-like*, and *defensin-like*) were selected for further analysis as they show a ripening-related expression pattern (Figure 1). Transient up- or RNAi-mediated down-regulation of the candidate genes by agroinfiltration (Hoffmann et al., 2006; Schwab et al., 2011) clearly confirmed the correlation of transcript abundance and flavonoid and anthocyanin accumulation (Figure 3). Loss-of-function phenotypes of the *expansin-A8-like* and *ephrin-A1-like* gene as well as the gain-of-function phenotypes of the *SRG1-like* and *defensin-like* gene showed impaired anthocyanin accumulation indicating that the first two genes might be positive and the last two genes negative regulators in the anthocyanin synthesis pathway.

Expansins were originally identified as cell wall-loosening proteins and are now considered as key regulators of cell wall breakdown and softening in processes such as fruit ripening, pollination, germination and abscission (Li et al., 2003). *FaExp2* was identified as ripening-regulated gene in strawberry fruit, but its expression is largely unaffected by auxin (Civello et al., 1999), contrarily to an *expansin* gene from tomato (Catalá et al., 2000). Nearly all previously identified ripening-regulated genes in strawberry fruit are negatively regulated by auxin (Harpster et al., 1998; Raab et al., 2006). *SRG1* denotes a protein-coding *Senescence Related Gene 1*, found in *A. thaliana* (Callard et al., 1996). Amino acid sequence analysis

shows that the putative SRG1 protein is a new member of the Fe (II)/ascorbate oxidase superfamily. Some members of this family catalyze the oxidation of intermediates of the flavonoid pathway (Almeida et al., 2007). Ephrins are receptor protein-tyrosine kinases and have been implicated in mediating developmental events (Wilkinson, 2000). They regulate cellular responses and their ectodomain is an eight-stranded  $\beta$ -barrel with topological similarity to plant nodulins and phytocyanins (Toth et al., 2001). Plant defensins are cationic peptides that belong to a large superfamily of antimicrobial peptides found in several organisms (De Oliveira Carvalho and Gomes, 2011). They present numerous biological activities, such as inhibiting protein synthesis, mediating abiotic stress, and altering the ascorbic acid redox state.

Flavonoids and anthocyanins are synthesized by highly complex but coordinated pathways during fruit ripening that are regulated at different levels such as signal reception (Chai et al., 2011; Jia et al., 2011), signal transduction (Jia et al., 2013), transcription factors (Medina-Puche et al., 2014), and structural genes (Griesser et al., 2008a). In addition to this, a complex regulatory network of positive and negative feedback mechanisms controlling anthocyanin synthesis and fruit ripening has been described (Petroni and Tonelli, 2011) such as the strawberry fruit Fra a allergen, a member of the pathogenesis related 10 (PR10) proteins which also functions in flavonoid biosynthesis (Muñoz et al., 2010). As interference at any regulatory level may perturb anthocyanin accumulation further investigations are needed to elucidate the precise functions of the four candidate genes that are essential for pigment formation in the ripe fruit. It is comprehensible that although the four candidate genes are probably not directly involved in flavonoid formation, their function may be required for the coordinate progression of the ripening process and they may have different roles for the different classes of phenolics (Figure 4).

#### **4.1.2 Untargeted analysis revealed novel strawberry metabolites**

Mass spectrometry-based untargeted metabolite analysis of the data sets was performed, followed by second-order comparisons to reduced and identify shared disturbances among the obtained phenotypes. Nine common metabolites were found to be differentially regulated by all five genes (*CHS*-silenced fruit as a positive control), one of which was identified as APG M2. Besides, pBI-*CHSi* fruits accumulated significantly lower levels of the structurally related metabolites M1 and M3 in comparison with control fruits. LC-MS

screening revealed one additional APG M4. Except for M3, the metabolites have not been reported from strawberry fruit, yet (Tsukamoto et al., 2004).

APG are prominent secondary metabolites of the genus *Hypericum* (Hypericaceae; Shiu et al., 2012; Crispin et al., 2013) and *Humulus* (Cannabinaceae; Bohr et al., 2005) and have been detected in *Phyllanthus emblica* (Zhang et al., 2002), *Jatropha multifida* (Kosasi et al., 1989) and *Curcuma comosa* (Suksamrarn et al., 1997) but are rarely found in other plant species. The structural diversity among APGs leads to various pharmacological activities *in vitro* and *in vivo*. APGs show significant antibacterial activity, cytotoxic, antiproliferative and antiangiogenic effects (Schmidt et al., 2012). The altered concentrations of APGs in response to changed expression levels of the candidate genes provide further evidence that the five genes studied are involved in the biosynthesis of phenolic compounds, but the role of the bioactive APGs in strawberry fruit remains unclear.

#### **4.1.3 CHS genes are involved in the biosynthesis of APGs**

The aglycones of M1-M4, which have been identified in hop (Bohr et al., 2005), were proposed to be generated by VPS, a key enzyme in the bitter acid biosynthesis pathway of hop (Zuurbier et al., 1998; Okada et al., 2004; Okada and Ito, 2001). Since *VPS* genes were not annotated in the *F. vesca* genome sequence (Shulaev et al., 2011) we searched for related polyketide synthase genes whose products may catalyze the VPS reaction (Abe and Morita, 2010). Eight *CHS* genes were detected but only two of them (gene26825 and 26826) are transcribed during strawberry fruit ripening (Appendix Figure 1; Kang et al., 2013). Three *CHS2* genes were cloned from *F. vesca*, including a novel *CHS* sequence, named FvCHS2.3. Comparison of the deduced amino acid sequences showed that the N-terminal part of FvCHS2.3 is identical with gene26825 (FvCHS2.1) whereas the C-terminus contains the sequence of gene26826 (FvCHS2.2). Thus, we assume that FvCHS2.3 is an artifact formed during PCR.

Recent structural and functional studies have elucidated the basic chemical mechanism for polyketide formation in CHS (Ferrer et al., 1999; Jez et al., 2000). Three essential catalytic aminoacids, Cys164, His303, and Asn336 (Jez et al., 2000, 2001, 2002) are conserved in all known CHS-related enzymes and are all well conserved in the FvCHS2 proteins (Appendix Figure 2). In addition two Phe residues (Phe215 and Phe265), important in determining the substrate specificity of CHS (Jez et al., 2002) are also fully conserved in the three enzymes.

Earlier studies on recombinant CHS enzymes from *Pinus sylvestris*, *P. strobus* and *Sinapis alba* showed that they can perform the function of VPS, but not perfectly, because the majority of the products were released from the polyketide synthases after two condensation reactions (Zuurbier et al., 1998). Additional research on CHS homologues from hop revealed that VPS and CHS\_H1s were able to form naringenin chalcone and APGs, but the reactions of CHS2 and CHS4 were prematurely terminated when isovaleryl-CoA and isobutyryl-CoA were utilized as substrates (Novák et al., 2006). Overall, CHS\_H1 and VPS from hop efficiently catalyze the VPS reaction but only CHS\_H1 has true CHS activity (Novák et al., 2006; Table 11).

**Table 11.** Apparent kinetic constants of VPS and CHS\_H1 from *Humulus lupulus* L. with different substrates. Data ( $K_M$  and  $V_{lim}$ ) were taken from Novák et al., 2006; Values of  $k_{cat}/K_M$  were calculated

	starter substrate	$K_M$ [ $\mu\text{M}$ ]	$V_{lim}$ [ $\text{pKat } \mu\text{g}^{-1}$ ]	$k_{cat}/K_M$ $\text{M}^{-1} \text{s}^{-1}$
HIVPS	isovaleryl-CoA	$5.0 \pm 0.2$	$65.2 \pm 6.6$	2414
	isobutyryl-CoA	$14.6 \pm 0.1$	$57.3 \pm 4.4$	2087
	4-coumaroyl-CoA	$29.0 \pm 0.2$	$5.6 \pm 1.0$	222
HICHS_H1	isovaleryl-CoA	$8.0 \pm 0.2$	$196.8 \pm 20.0$	7287
	isobutyryl-CoA	$14.9 \pm 0.2$	$106.5 \pm 10.3$	3878
	4-coumaroyl-CoA	$40.9 \pm 0.1$	$256.5 \pm 28.1$	10189

In contrast, CHS2.1, 2.2 and 2.3 from *F. vesca* form naringenin chalcone from 4-coumaroyl-CoA with similar efficiency ( $k_{cat}/K_M$  2021, 2045, 3957  $\text{M}^{-1} \text{s}^{-1}$ , respectively; Table 7) but FvCHS2.1 shows superior VPS activity with isovaleryl-CoA as starter molecule ( $k_{cat}/K_M$  21700  $\text{M}^{-1} \text{s}^{-1}$ ). FvCHS2.1, 2.2, and 2.3 also accept cinnamoyl-CoA and feruloyl-CoA, except CHS2.2 as starter molecules, which confirms the CHS activity of the strawberry enzymes (Table 6). FvCHS2.3 is probably an artifact and resembles CHS2.2 in its catalytic properties (Table 6). The *in vitro* reactions are accompanied by the formation of byproducts as has been observed for CHS\_H1 and VPS from hop (Figure 8; Table 6; Novák et al., 2006). Thus, CHS2.2 and in particular CHS2.1 are able to efficiently catalyze the formation of the aglycones of M1-M4 *in vitro*.

#### 4.1.4 *In planta* functional analysis of CHS2 genes

Analysis of the CHS2 function *in planta* was performed by RNAi-mediated down-regulation of CHS2 transcript abundance by agroinfiltration in receptacles of *F. x ananassa* cv Mara des

Bois (Hoffmann et al., 2006). The *CHS2*-silenced fruits showed white regions, accumulated lower levels of flavonoids and anthocyanins and contained less *CHS2* transcripts than control fruits (Hoffmann et al., 2006; Figure 15A). Besides, down-regulation of *CHS2* also resulted in significantly reduced levels of APGs (M1 – M3; Figure 15B) in comparison with the values of the pBI-Intron control fruit.

In addition, plants of a stable transgenic *chs* antisense line (*F. x ananassa* cv Calypso) were available to analyze the effects of *CHS* down-regulation on the accumulation of M1 – M3 in fruit receptacles (Lunkenbein et al., 2006b). The pink colored fruit of the transgenic line, a sign of the impaired anthocyanin production, contained significantly lower concentrations of APGs than strawberry fruit of *F. x ananassa* cv Calypso wild type plants (Figure 15C). The results of both reverse genetic approaches clearly confirm the biochemical function of the FvCHS2 enzyme in the anthocyanin but also in the APG biosynthesis pathway *in vivo*.

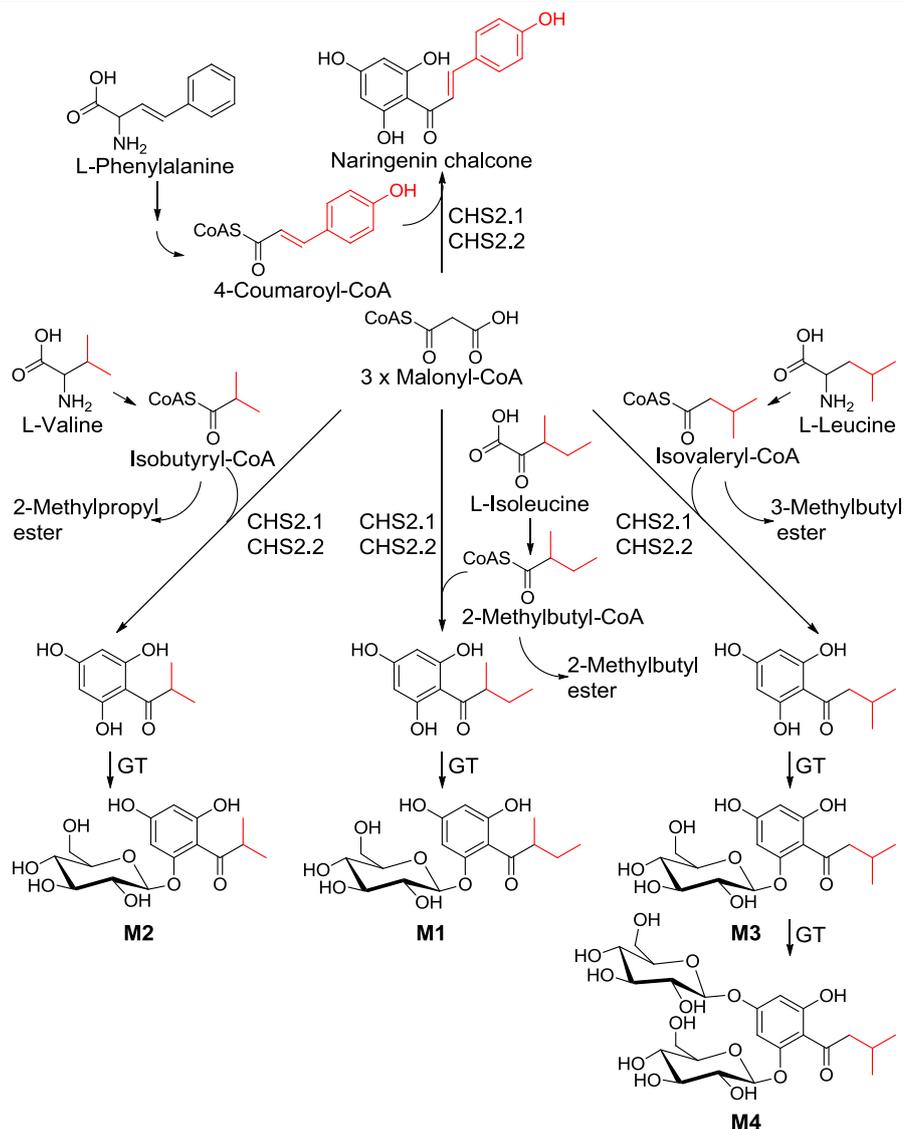
#### 4.1.5 APG pathway in strawberry fruit

Polyketide synthases play major roles in the biosynthesis of diverse secondary metabolites as they generate the backbones of chalcones, stilbenes, phloroglucinols, resorcinols, benzophenones, biphenyls, dibenzyls, chromones, acridones, pyrones, and curcuminoides (Abe and Morita, 2010). CHS is the best studied plant-specific polyketide synthase that catalyzes sequential decarboxylative condensation of 4-coumaroyl-CoA with three molecules of malonyl-CoA. The functional diversity derives from the differences in the selection of the starter molecule, the number of malonyl-CoA condensations, and the mechanisms of the cyclization reaction. In enzyme reactions *in vitro*, triketide and tetraketide derivatives are also obtained as early-release derailment by-products (Abe and Morita, 2010; Figure 8). CHS shows extremely broad substrate promiscuity and catalytic potential as it accepts a variety of CoA thioesters, to produce a series of structurally divergent polyketides (Zuurbier et al., 1998).

Similarly, FvCHS2.1, 2.2, and 2.3 can use aromatic and methyl branched-chain aliphatic CoAs (Table 6) as starter substrates that are derived from the transformation of aromatic and branched-chain amino acids, respectively (Xu et al., 2013). Isovaleryl-CoA, isobutyryl-CoA and 2-methylbutyryl-CoA are produced by transamination of Leu, Val, and Ile, respectively followed by the oxidative decarboxylation of the ketoacid intermediates (Xu et al., 2003). Thus, we propose Ile, Val, and Leu as biogenetic precursors of M1, 2 and 3, respectively,

whereas M4 is formed by glucosylation of M3 (Figure 52). The transformation of branched chain amino acids to M1–M3 was confirmed by stable isotope tracer experiments in which 40 to 79% of the APG products were labeled with the isotopes after one day when solutions of labeled precursor amino acids were injected into fruits (Table 8). The degree of labeling was already high in the APG products after one day but a calculation showed that the amount of labeled Val, Leu and Ile (2.9, 3.3. and 3.3 mg/10 g fresh fruit, respectively) which was injected into fruit exceeded the amount of the naturally unlabeled amino acids (1.0 - 2.0, 0.4, and 1.1 – 1.4 mg/100 g fresh weight of Val, Leu and Ile, respectively; Perez et al., 1992; Keutgen and Pawelzik, 2008) in the fruit by at least a factor of 15 (1500%). Thus, a substantial amount of the labeled precursors is probably degraded by other pathways or was not yet transformed. In fact, maximum labeling was achieved after 4 days (Figures 18 and 19) peaking at 1100% (factor of 11) in the case of Ile.

One of the main sources of substrates for volatile ester production is the metabolism of amino acids, generating alcohols and acids, either aliphatic, branched-chain, or aromatic (Pérez et al., 2002). These esters contribute, and in some cases are determining, to the primary aroma of many fruits. The fate of Val and Ile in relation to aroma biogenesis has already been studied in strawberry fruit and showed that feeding of Ile resulted in a substantial increase in the sum of 2-methylbutanoate esters and 2-methylbutyl esters compared to those of control fruits (Pérez et al., 2002). Similarly, levels of volatile, branched-chain aliphatic esters were increased in pBI-*CHSi* fruits in comparison to controls (Figure 16) which demonstrates that higher concentrations of precursor CoA thioesters are available for ester formation in pBI-*CHSi* fruits due to the down-regulation of the CHS function. Thus, the pathways producing branched-chain aliphatic esters and APGs compete for their common CoA thioester substrates (Figure 52).



**Figure 52.** The proposed APG biosynthesis pathway catalyzed by CHS2.1 and 2.2 in strawberry fruit. GT= glucosyl transferase.

#### 4.1.6 Evolutionary relevance

Gene duplication has been proposed to be an important process in the generation of evolutionary novelty (Hughes, 2005). Neofunctionalization, as an adaptive process where one copy of the original gene mutates into a function that was not present in the ancestral gene, is one mechanism that can lead to the retention of both copies. In the case of *FvCHS2.1* and *2.2* (*CHS2.3* is considered as artifact), duplication of an ancestral *CHS2* gene appears to have been occurred only recently, on evolutionary scales, as the nucleotide sequences of *CHS2.1* and *2.2* are still highly similar (95 and 98% nucleotide and amino acid identity, respectively) and both genes show identical expression profiles (Appendix Figure 1). However, enzymatic activity of *FvCHS2.1* for the starter substrate isovaleryl-CoA has been

significantly improved (Table 7). Thus, subfunctionalization, as a neutral process where a paralog specializes one (here VPS function) of several ancestral functions best describes the fact that both *CHS2* copies are retained in the genome of *F. vesca* (Lynch and Force, 1999). The results of the studies on *FvCHS2* confirm that when multifunctionality precedes gene duplication, it is straightforward for duplicates to specialize by sharing the ancestral function (Hughes, 2005).

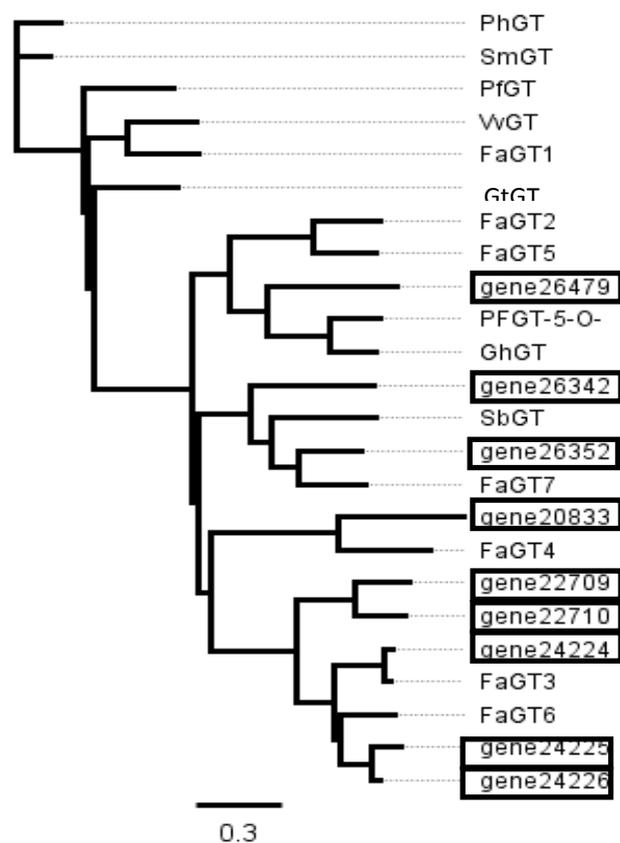
## 4.2 Substrate promiscuity of glucosyltransferases from strawberry

Glycosylation is a key mechanism that determines the chemical complexity and diversity of plant natural products (Gachon et al., 2005; Osmani et al., 2009), ensures their chemical stability, water solubility while reducing toxicity (Bowles et al., 2005), as well as facilitates their intercellular transport, storage and accumulation in plant cells (Wang and Hou, 2009). Members of the UGT superfamily have been classified into 94 families where family 1 refers to the uridine glucosyltransferases (UGTs) (Yonekura-Sakakibara and Hanada, 2011; Caputi et al., 2012). Given the biological, pharmacological and agronomic relevance of secondary metabolites, UGTs have attracted considerable interest for decades but, only a handful of them have been characterized.

### 4.2.1 Candidates UGTs selection and phylogenetic tree analysis

Genome-wide transcriptional analysis is a powerful tool to discover candidate genes (Achnine et al., 2005). With the aim to characterize ripening-related strawberry UGTs functionally, UGT genes were selected which showed increasing transcript levels in three *F. vesca* varieties during fruit ripening. All selected candidate GTs display the characteristic conserved PSPG box responsible for the interaction with the sugar donor.

The protein sequence analysis showed that FaGT24224 is similar to FaGT3 but differs in 20 amino acids. FaGT24225/24226 and FaGT22709/22710 are highly similar to each other (Figure 53). FaGT26352 shows closest relationship with FaGT7 which accepts numerous flavonoids and hydroxycoumarins (Griesser et al., 2008b). The transcripts of the functionally characterized FaGT24226, 24224, 24225, 22709 and 26479 are predominantly expressed in fully ripe fruit receptacles but show very weak expression in vegetative tissues of *F. x ananassa* (Figure 31). A similar expression pattern was observed for a (hydroxy)cinnamate GT from strawberry (FaGT2, Lunkenbein et al., 2006a).



**Figure 53.** Phylogenetic tree of UGT proteins from *F. x ananassa* and functionally characterized UGTs from different plants, generated by the Geneious (Pro 5.5.4) Tree Builder (Jukes Cantor genetic distance model and neighbor-joining method). The scale bar indicates the average number of amino acid substitutions per site. PfGT *Perilla frutescens* flavonoid 3-GT (AB002818), GtGT *Gentiana triflora* flavonoid-3-GT (D85186), SmGT *Solanum melongena* GT (X77369), PhGT *Petunia x hybrida* anthocyanidin 3-GT (AB027454), VvGT *Vitis vinifera* flavonoid 3-GT (AF000371), PFGT-5-GT *Perilla frutescens* anthocyanin 5-GT (AB013596), FaGT7 *F. x ananassa* UDP-glucose GT (ABB92749), FaGT6 *F. x ananassa* UDP-glucose GT (ABB92748), FaGT3 *F. x ananassa* UDP-glucose GT (AAU09444), FaGT2 *F. x ananassa* UDP-glucose GT (AAU09443), FaGT5 *F. x ananassa* UDP-glucose GT (ABB92747), FaGT4 *F. x ananassa* UDP-glucose GT (AAU09445), FaGT24224, 24225, 24226, 22709, 22710, 20833, 26342, 26352, and 26479 are translated protein sequences of putative *F. vesca* UGTs (Shulaev et al., 2011). Proteins selected in this study are boxed.

#### 4.2.2 Biochemical characterization of selected UGTs from *Fragaria x ananassa*

The substrate screening showed that FaGT24226, 24224, 24225, and 22709 are able to glucosylate a number of phenols such as quercetin, kaempferol, and isorhamnetin that occur naturally in strawberry (Griesser et al., 2008b) and form multiple products. Plant GTs have been found to exhibit a rather strict regioselectivity towards the position of the sugar attachment (Vogt and Jones, 2000). Radiochemical analyses revealed that FaGT24226, 24224, and 24225 could react on the hydroxyl group at position 3 and 7 of flavonols,

whereas FaGT22709 was unable to catalyze glucosylation at position 7. In contrast to FaGT24224, 24225a/b, 24226, and 22709 which all exhibited a broad substrate tolerance *in vitro*, FaGT26479 showed strict substrate specificity and could only efficiently glucosylate galangin *in vitro*. FaGT7 was reported to exhibit a broad substrate tolerance *in vitro* (Griesser et al., 2008b) but FaGT26352, its closest homologue could not use any of the substrates tested (Figure 53).

Two flavonoid GTs, UGT73A4 and UGT71F1, from *Beta vulgaris* glucosylate numerous substrates but also show distinct regioselectivity. The hydroxylation pattern in ring B of the acceptor molecule was reported to influence product specificity (Isayenkova et al., 2006). If a hydroxy group is present at position 3', 4'-glucosides are preferentially formed. If the 3'-hydroxy group is missing, the enzyme produces 7-glucosides (Isayenkova et al., 2006). Likewise, FaGT24225a formed almost equal amounts of the 3- and 7-glucoside but rarely 4'-glucoside when isorhamnetin was used as substrate, where the hydroxyl group at position 3' is transformed by an O-methylation. However, in the case of FaGT24225b, the 3-glucoside of isorhamnetin is the main product. GTs from *Allium cepa* and *Arabidopsis thaliana* show a similar regioselectivity forming several monoglucosides and diglucosides of flavonol substrates (Kramer et al., 2003; Lim et al., 2004). Likewise, FaGT24225a and b formed contrasting amounts of mono- and diglucosides (Figure 27) although there are only few amino acids exchanges in the protein sequences.

Sugar conjugation increases stability and water solubility of natural products (Jones and Vogt, 2001) but UGTs might also be involved in the inactivation and detoxification of xenobiotics and harmful compounds (Pedras et al., 2001). All enzymes tested in this study accepted a series of metabolites not present in strawberry fruits and anthropogenic substrates. FaGT22709 showed highest activity towards 1-naphtol, and FaGT24224, 24225, 24226 towards 3-hydroxycoumarin. Both, 1-naphtol and 3-hydroxycoumarin have not been detected in strawberry, up to now. This indicates that the FaGTs may not only function in flavonol biosynthesis, but perhaps play a role in xenobiotic metabolism as well. Similarly, FaGT6 and FaGT7 also accept secondary metabolites not occurring in strawberry fruits (Griesser et al., 2008b).

#### **4.2.3 *In vitro* versus *in planta* substrates**

The broad substrate specificity exhibited by most UGTs *in vitro* hinders the identification of their real substrates *in vivo* (Achnine et al., 2005). A physiologic aglycone library which is enriched in naturally occurring aglycones (Bönisch et al., 2014) can be used to reveal the natural substrates of GTs. Thus, an enzymatically hydrolyzed strawberry glycosides extract was used as a physiologic library in this study to search for *in vivo* substrates of FaGT24224, 24225a/b, 24226 and 22709. This physiologic library was co-incubated with purified recombinant FaGTs, UDP-[U-<sup>14</sup>C] glucose or unlabeled UDP-glucose. FaGT24224, FaGT24225a/b, and FaGT22709 could utilize biological substrates occurring in the strawberry aglycone library. LC-MS analysis identified kaempferol as putative natural substrate for FaGT24224, 24225a, 24226 and 22709 and quercetin for FaGT24225a and FaGT22709.

Large-scale analytical technologies are ideal tools to look for candidates (Vogt and Jones, 2000; Achnine et al., 2005) that fulfill a given function. Activity-based profiling of aglycone libraries could be used as a large-scale analytical technology to reveal the function of UGTs and to uncover novel biologically relevant substrates of small molecule GTs that often show broad sugar acceptor promiscuity (Bönisch et al., 2014). In addition to the identification of quercetin and kaempferol-3-glucoside, LC-MS analysis reveals three additional glucose conjugates that were formed by the FaGTs. Thus, the strawberry aglycone library still contains biological substrates which remained to be structurally identified. The screening of aglycone libraries with recombinant GT enzymes in combination with LC-MS and NMR analyses will greatly facilitate the functional analyses of GTs in the future.

### **4.3 Characterization of a phloroglucinol-glucosyltransferase from strawberry**

More and more pharmaceutically important phloroglucinol glucosides and its derivatives are reported from natural sources (Singh et al., 2010). However, little is known about the glycosylation of phloroglucinols and no GT has been reported, up to now, that could glycosylate phloroglucinols. To functionally characterize fruit ripening-related UGTs in strawberry, candidate GTs were selected based on their transcription levels in fruits of three *F. vesca* genotypes during fruit development. Substrate screenings yielded two UDP-glucose:phloroglucinol glucosyltransferases.

### 4.3.1 Substrate promiscuity of selected GTs

Many GTs show relatively broad substrate tolerance such as the multifunctional UGT enzyme arbutin synthase, which accepted 45 out of 74 tested natural and synthetic substrates (Hefner et al., 2002). In strawberry, a multifunctional enzyme, FaGT2, is involved in the metabolism of both natural and xenobiotic compounds, resulting in the formation of O- and S-glucose esters, as well as O-glucosides (Landmann et al., 2007; Lunkenbein et al., 2006a). However, there are also examples of UGTs that are quite specific (Fukuchi-Mizutani et al., 2003; Jugdé et al., 2008).

Similarly to arbutin synthase and FaGT2, FaGT00126, FaGT07876a/b, and FaGT26345 showed broad substrate tolerance *in vitro*, accepting numerous flavonoids, hydroxycoumarins, and naphthols. Additionally, FaGT07876a/b also glucosylated trihydroxybenzene isomers and APG aglycones and may function in APG glucosides biosynthesis in plants.

### 4.3.2 Regioselectivity

Regioselectivity of GTs offers an important means to overcome the limitations of chemical synthesis of small molecule glycosides (Lim et al., 2004). Family 1 glycosyltransferases predominantly recognize low-molecular-weight compounds, such as phenolics, with a high regioselectivity (Vogt and Jones, 2000; Lim et al., 2002). Out of the 91 recombinant GTs from *Arabidopsis thaliana*, 29 enzymes expressed catalytic activity toward quercetin, 14 enzymes recognized only a single site, 11 of those glycosylated the C3-OH, whereas 3 GTs glycosylated the C7-OH (Lim et al., 2004). UGT73A4 and UGT71F1 isolated from *Beta vulgaris* also showed distinct regioselectivity, UGT73A4 showed a preference for the 4- and 7-OH position in the flavonoids, whereas UGT71F1 preferentially glycosylated the 3- or the 7-OH position (Isayenkova et al., 2006) and steviol GTs also behave in a regioselective manner (Richman et al., 2005). We showed that FaGT07876a, similar to FaGT26345 and 00126, produces less 3-glucoside than 7-glucoside, however, the product ratio was reversed for FaGT07876b when kaempferol was used as substrate. Additional analyses revealed that only trace amount of diglucosides were produced by FaGT07876a/b and 00126 but three diglucosides were readily formed by FaGT26345.

It is assumed that regiospecificity of UGTs differentiated prior to speciation (Noguchi et al., 2009) and the ability of GTs to glycosylate particular sites on the ring was found to be

strongly affected by the presence or absence of additional hydroxyl groups at positions relative to the initial glycosylation site (Lim et al., 2002). Similarly, FaGT00126, FaGT07876a/b, and FaGT26345 produced different amounts of 3- and 7-glucosides from kaempferol and quercetin. Additionally, distinct product patterns were observed as FaGT07876a/b and FaGT26345 formed mainly quercetin 7- and 3'-monoglucosides but FaGT00126 produced also remarkable levels of the 4'-glucoside.

#### 4.3.3 Phloroglucinol GT in strawberry

Pharmaceutically important phloroglucinols are widely used in medicine, cosmetics, pesticides, paints, cements and dyes. More than 50 structure including non-acylated, mono- and di-acylated glycosides wherein the sugar moiety is attached to the aglycone through an O- or C-glycosidic linkage have been identified from natural sources (Singh et al., 2010). [<sup>14</sup>C]-phloroglucinol glucoside was synthesized in *Pelargonium* by feeding with phloroglucinol and labelled glucose (Hutchinson et al., 1958).

FaGT07876a and b readily catalyze the glycosylation of commercially available phloroglucinol and 1,2,4-benzenetriol. Two main products were formed when pyrogallol was used as substrate. UGTs have attracted considerable interest because the glycosylation of low-molecular-weight compounds usually changes acceptors in terms of increasing solubility and accumulation, as well as regulating their bioactivity, such as antioxidant or anticancer activity (Bowles et al., 2005; Kramer et al., 2003a). Therefore, the identification of UDP-glucose:phloroglucinol GT is of great significance for synthesizing pharmaceutically bioactive phloroglucinol glycosides *in vitro*, and for metabolic engineering of plants which have the capability to biosynthesize phloroglucinols and structurally similar natural products such as APGs (Kristensen et al., 2005; Bowles et al., 2005; Weis et al., 2008).

#### 4.3.4 APG glucoside pathway in strawberry

APGs comprise the largest group of naturally occurring phloroglucinol compounds and more than 100 simple APGs have been reported (Singh et al., 2010). This study shows that strawberry plants are able to synthesize and accumulate biologically active APG glucosides, which have been only detected in a limited number of plants, up to now (Crispin et al., 2013; Bohr et al., 2005). The APG aglycones PIVP and PIBP are formed by *valerophenone synthase* (VPS) in hop (Paniego et al., 1999; Okada et al., 2004b), and by a dual functional *chalcone synthase 2* (CHS2) in strawberry fruits (refer to 3.1 ; Figure 52).

Numerous groups have endeavoured to synthesize APG glucosides owing to the vast biological activities of these compounds. Acid catalyzed direct condensation of phloroglucinol with unprotected sugars yielded phloroglucinol C- $\beta$ -D-glucoside when D-glucose was used as a glycosylating agent at temperatures below 80°C. The yield was only 11% (Onodera et al., 1983). In contrast, FaGT07876a and b readily form O-glucosides from trihydroxybenzenes and APG glucosides such as 2- and 4-PIVP glucoside whereas the expression level of gene07876 correlates with APG glucoside accumulation in strawberry fruit (*F. × ananassa*). Furthermore, co-incubation of CHS2 and FaGT07876 together with isovalery-CoA/isobutyryl-CoA, malonyl-CoA, and UDP-glucose yielded PIVP and PIBP glucoside. The result provides evidence that FaGT07876a/b can catalyze the last step of APG glucoside formation (Figure 52).

#### **4.4 UGT71C3 glucosylates the key flavor compound 4-hydroxy-2,5-dimethyl-3(2H)-furanone (HDMF)**

Although GTs have been studied for many years (Gachon et al., 2005), no glucosyltransferase has been reported up to now which catalyzes the glucosylation of the key strawberry flavor compound HDMF.

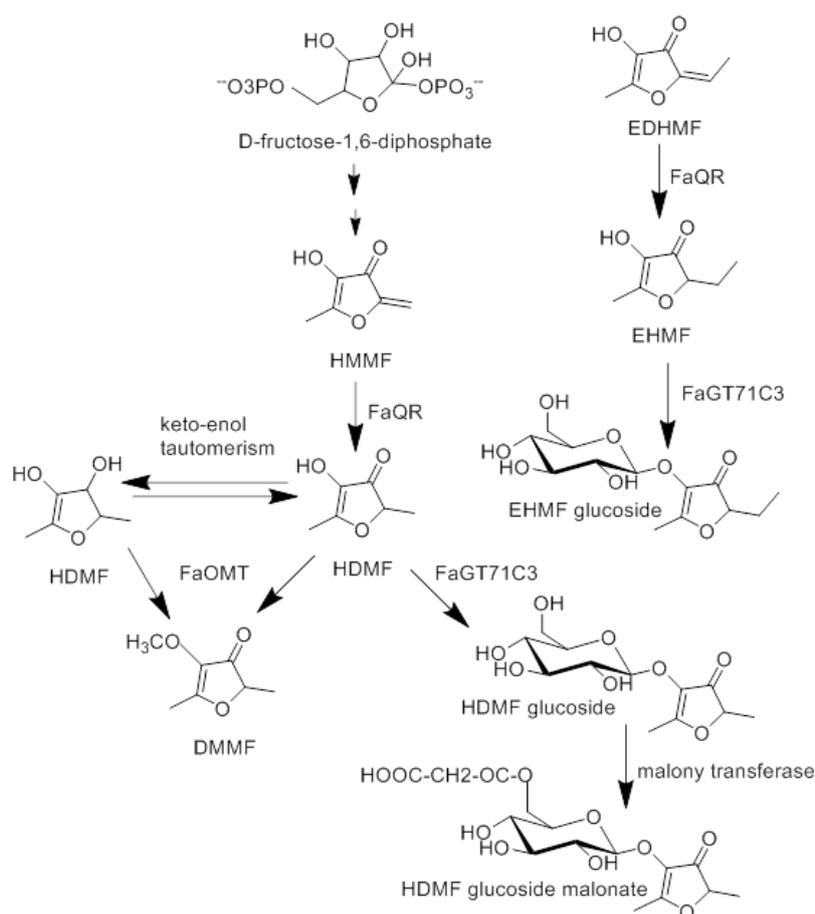
##### **4.4.1 HDMF glucosylation activity**

A screening of putative, ripening related GTs with different aroma chemicals (Figure 44) revealed two allelic proteins FaGT07876a and b that accept HDMF as substrate and formed HDMF–glucoside (Figure 45). HDMF was first reported as a product of the Maillard reaction (Hodge, 1963) and was subsequently isolated from different fruit including pineapple, strawberry, and tomato (Rodin et al., 1965; Ubeda et al., 2012; Buttery et al., 2001; Schwab, 2013). HDMF exhibits a caramel-like aroma similar to its structural homologues HMF and EHMF, which have been identified in tomato and melon fruits (Lignou et al., 2014), respectively, but their biogenetic pathways remain unknown. FaGT07876a/b also glycosylated EHMF but not HMF *in vitro* (Figure 47), implicating that melon express a similar UGT enzyme.

##### **4.4.2 Screening of volatile substrates**

Substrate screenings of selected FaGTs also showed that all proteins tested in this study were able to efficiently form vanillyl glucoside. Vanillin is one of the most widely used flavor

agents in the world, and has been reported to occur in many fruits and fruit products. Vanillyl- $\beta$ -D-glucoside has been found in addition to “free” vanillin in mango (Sakho et al., 1997), in the fresh pod of the vanilla orchid (Negishi et al., 2009), in orange juice (Marin et al., 1992) and strawberries (Pyysalo et al., 1979). Seven UGTs have already been identified which possess high *in vitro* catalytic activity toward vanillin, namely UGT71C2, UGT72B1, UGT72E2, UGT84A2, and UGT89B1 from *A. thaliana* (Hansen et al., 2009), UGT85B1 from *Sorghum bicolor* (Jones et al., 1999), and arbutin synthase from *Rauwolfia serpentine* (Jones, 1998). Similarly, we found seven ripening related GTs in strawberry that formed vanillyl glucoside.



**Figure 54.** Biosynthesis and metabolism of HDMF in strawberry (left). FaQR (identical with FaEO) and FaGT71C3 (identical with FaGT07876a/b) also form and glucosylate the structural homologue EHMf.

#### 4.4.3 Site-directed mutagenesis

Site-directed mutagenesis is a powerful method for determining the specific amino acid residues in substrate recognition or catalytic function. Phylogenetic analysis and comparison

of substrate recognition patterns among family 1 UGTs have indicated that amino acid residues in the N-terminal half of the proteins were responsible for acceptor binding (including the PSPG-box), whereas those in the C-terminal half were involved mainly in interactions with donor substrates (Paquette et al., 2003; Hu and Walker, 2002; Ross et al., 2001; Lim et al., 2003; Li et al., 2001). 3D-structures of betanidin 5-O-glucosyltransferase from *Dorotheanthus bellidiformis* (Hans et al., 2004) and cyanohydrin glucosyltransferase from *Sorghum bicolor* (Thorsøe et al., 2005) by homology modeling, and of isoflavonoid 3'-O-glucosyltransferase from *Medicago truncatula* (Shao et al., 2005; He et al., 2006) and flavonoid 3-O-glucosyltransferase from *Vitis vinifera* (Offen et al., 2006) revealed the role of specific conserved amino acid residues in the PSPG-box that constitute the donor-sugar binding pockets. However, the roles of less well conserved amino acids within the motif that may determine the characteristics unique to particular enzymes such as substrate recognition and catalytic potential have been less closely examined.

Although the targets of site-directed mutagenesis are usually highly conserved residues, less conserved amino acids can also be of interest since these may decide the characteristics unique to particular enzymes. Such as it was reported that a non-conserved residue, Cys377, in the PSPG-box of CaUGT2 plays an essential role in determining the catalytic function of CaUGT2 (Masada et al., 2007). In the current study, we show that single substitutions D343E, K458R, D445E and a double mutant of G433A I434V of FaGT24224 all displayed enhanced HDMF glucosylation activity, especially mutant D343E and the double mutant of G433A I434V exhibited a more than 2-fold higher activity than that of the wildtype enzyme FaGT24224. By contrast, a single mutant in the center of the PSPG box, where Ala 389 of FaGT24224 was replaced by a Val residue, resulted in a decrease of the HDMF glucosylation activity. Consequently, the HDMF glucosylation activity of FaGT07876b was further increased by replacing valine at position 383 by alanine.

In summary, we have identified the first dual functional HDMF/EHMF UGTs (Figure 54) from strawberry and nine FaGTs that transfer glucose onto vanillin. This knowledge can be applied in strawberry breeding for the improvement of fruit quality and for the biotechnological production of aroma glucosides which find use in food and cosmetics industry as flavor and aroma precursors.

## 5. References

- Aaby, K., Ekeberg, D., and Skrede, G.** (2007). Characterization of phenolic compounds in strawberry (*Fragaria x ananassa*) fruits by different HPLC detectors and contribution of individual compounds to total antioxidant capacity. *J. Agric. Food Chem.* **55**: 4395–406.
- Aaby, K., Skrede, G., and Wrolstad, R.E.** (2005). Phenolic composition and antioxidant activities in flesh and achenes of strawberries (*fragaria ananassa*). *J. Agric. Food Chem.* **53**: 4032–4040.
- Abe, I. and Morita, H.** (2010). Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. *Nat. Prod. Rep.* **27**: 809–38.
- Ablajan, K., Abliz, Z., Shang, X.-Y., He, J.-M., Zhang, R.-P., and Shi, J.-G.** (2006). Structural characterization of flavonol 3,7-di-O-glycosides and determination of the glycosylation position by using negative ion electrospray ionization tandem mass spectrometry. *J. Mass Spectrom.* **41**: 352–360.
- Achkar, J., Xian, M., Zhao, H., and Frost, J.W.** (2005). Biosynthesis of phloroglucinol. *J. Am. Chem. Soc.* **127**: 5332–3.
- Achnine, L., Huhman, D. V., Farag, M. a, Sumner, L.W., Blount, J.W., and Dixon, R. a** (2005). Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*. *Plant J.* **41**: 875–87.
- Aharoni, A., Keizer, L.C., Bouwmeester, H.J., Sun, Z., Alvarez-Huerta, M., Verhoeven, H.A., Blaas, J., van Houwelingen, A.M., De Vos, R.C., van der Voet, H., Jansen, R.C., Guis, M., Mol, J., Davis, R.W., Schena, M., van Tunen, A.J., O'Connell, A.P.** (2000). Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. *Plant Cell* **12**: 647–662.
- Aherne, S.A. and O'Brien, N.M.** (2002). Dietary flavonols: Chemistry, food content, and metabolism. *Nutrition*. **18**. 75–81.
- Akiyama, T., Shibuya, M., Liu, H.M., and Ebizuka, Y.** (1999). p-Coumaroyltriacyclic acid synthase, a new homologue of chalcone synthase, from *Hydrangea macrophylla* var. thunbergii. *Eur. J. Biochem.* **263**: 834–839.
- Allan, A.C., Hellens, R.P., and Laing, W.A.** (2008). MYB transcription factors that colour our fruit. *Trends Plant Sci.* **13**: 99–102.
- Almeida, J.R.M., D'Amico, E., Preuss, A., Carbone, F., de Vos, C.H.R., Deiml, B., Mourgues, F., Perrotta, G., Fischer, T.C., Bovy, A.G., Martens, S., and Rosati, C.** (2007). Characterization of major enzymes and genes involved in flavonoid and proanthocyanidin biosynthesis during fruit development in strawberry (*Fragaria x ananassa*). *Arch. Biochem. Biophys.* **465**: 61–71.
- Aziz-Ur-Rehman, M., Riaz, N., Ahmad, H., Nawaz, S.A., and Choudhary, M.I.** (2005). Lipooxygenase Inhibiting Constituents from *Indigofera hetrantha*. *Chem. Pharm. Bull* **53**: 263–266.
- Bangera, M.G. and Thomashow, L.S.** (1999). Identification and characterization of a gene cluster for synthesis of the polyketide antibiotic 2,4-diacetylphloroglucinol from *pseudomonas fluorescens* Q2-87. *J. Bacteriol.* **181**: 3155–3163.
- Beuerle, T. and Pichersky, E.** (2002). Enzymatic synthesis and purification of aromatic coenzyme a esters. *Anal. Biochem.* **302**: 305–312.
- Bohr, G., Gerhäuser, C., Knauff, J., Zapp, J., and Becker, H.** (2005). Anti-inflammatory acylphloroglucinol derivatives from Hops (*Humulus lupulus*). *J. Nat. Prod.* **68**: 1545–8.

- Bönisch, F., Frotscher, J., Stanitzek, S., Rühl, E., Wüst, M., Bitz, O., and Schwab, W.** (2014). Activity based profiling of a physiologic aglycone library reveals sugar acceptor promiscuity of family 1 UDP-glucosyltransferases from *Vitis vinifera*. *Plant Physiol.* **166**:23-39.
- Boss, P.K., Davies, C., and Robinson, S.P.** (1996). Expression of anthocyanin biosynthesis pathway genes in red and white grapes. *Plant Mol. Biol.* **32**: 565–569.
- Boudet, A.-M.** (2007). Evolution and current status of research in phenolic compounds. *Phytochemistry* **68**: 2722–2735.
- Bowles, D., Isayenkova, J., Lim, E.K., and Poppenberger, B.** (2005). Glycosyltransferases: Managers of small molecules. *Curr. Opin. Plant Biol.* **8**: 254–263.
- Bradford, M.M.** (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**: 248–254.
- Bravo, L.** (1998). Polyphenols: chemistry, dietary sources, metabolism, and nutritional significance. *Nutr. Rev.* **56**: 317–333.
- Brugliera, F., Holton, T.A., Stevenson, T.W., Farcy, E., Lu, C.Y., and Cornish, E.C.** (1994). Isolation and characterization of a cDNA clone corresponding to the Rt locus of *Petunia hybrida*. *Plant J.* **5**: 81–92.
- Buendía, B., Gil, M.I., Tudela, J.A., Gady, A.L., Medina, J.J., Soria, C., López, J.M., and Tomás-Barberán, F.A.** (2010). HPLC-MS analysis of proanthocyanidin oligomers and other phenolics in 15 strawberry cultivars. *J. Agric. Food Chem.* **58**: 3916–3926.
- Buttery, R.G., Takeoka, G.R., Naim, M., Rabinowitch, H., and Nam, Y.** (2001). Analysis of Furanol in tomato using dynamic headspace sampling with sodium sulfate. *J. Agric. Food Chem.* **49**: 4349–4351.
- Callard, D., Axelos, M., and Mazzolini, L.** (1996). Novel molecular markers for late phases of the growth cycle of *Arabidopsis thaliana* cell-suspension cultures are expressed during organ senescence. *Plant Physiol.* **112**: 705–715.
- Caputi, L., Malnoy, M., Goremykin, V., Nikiforova, S., and Martens, S.** (2012). A genome-wide phylogenetic reconstruction of family 1 UDP- glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. *Plant J.* **69**: 1030–1042.
- Catalá, C., Rose, J.K., and Bennett, A.B.** (2000). Auxin-regulated genes encoding cell wall-modifying proteins are expressed during early tomato fruit growth. *Plant Physiol.* **122**: 527–534.
- Chai, Y.M., Jia, H.F., Li, C.L., Dong, Q.H., and Shen, Y.Y.** (2011). FaPYR1 is involved in strawberry fruit ripening. *J. Exp. Bot.* **62**: 5079–5089.
- Cheng, G.W., Malencik, D.A., and Breen, P.J.** (1994). UDP-glucose: Flavonoid O-glucosyltransferase from strawberry fruit. *Phytochemistry* **35**: 1435–1439.
- Cheynier, V., Comte, G., Davies, K.M., Lattanzio, V., and Martens, S.** (2013). Plant phenolics: Recent advances on their biosynthesis, genetics, and ecophysiology. *Plant Physiol. Biochem.* **72**: 1–20.
- Civello, P.M., Powell, A.L., Sabehat, A., and Bennett, A.B.** (1999). An expansin gene expressed in ripening strawberry fruit. *Plant Physiol.* **121**: 1273–1280.
- Crispin, M.C., Hur, M., Park, T., Kim, Y.H., and Wurtele, E.S.** (2013). Identification and biosynthesis of acylphloroglucinols in *Hypericum gentianoides*. *Physiol. Plant.* **148**: 354–370.

- Cumplido-Laso, G., Medina-Puche, L., Moyano, E., Hoffmann, T., Sinz, Q., Ring, L., Studart-Wittkowski, C., Caballero, J.L., Schwab, W., Muñoz-Blanco, J., and Blanco-Portales, R.** (2012). The fruit ripening-related gene FaAAT2 encodes an acyl transferase involved in strawberry aroma biogenesis. *J. Exp. Bot.* **63**: 4275–4290.
- Darrow, G.M. (1966).** *The Strawberry: History, Breeding and Physiology.* Holt Rinehart & Winston, New York.
- Da Silva, F.L., Escribano-Bailón, M.T., Pérez Alonso, J.J., Rivas-Gonzalo, J.C., and Santos-Buelga, C. (2007).** Anthocyanin pigments in strawberry. *LWT - Food Sci. Technol.* **40**: 374–382.
- De Luca, V., Salim, V., Atsumi, S.M., and Yu, F. (2012).** Mining the Biodiversity of Plants: A Revolution in the Making. *Science (80- )*. **336**: 1658–1661.
- De Oliveira Carvalho, A. and Moreira Gomes, V. (2011).** Plant Defensins and Defensin-Like Peptides - Biological Activities and Biotechnological Applications. *Curr. Pharm. Des.* **17**: 4270–4293.
- Ferrer, J.L., Jez, J.M., Bowman, M.E., Dixon, R.A., and Noel, J.P. (1999).** Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nat. Struct. Biol.* **6**: 775–784.
- Fischer, T.C., Mirbeth, B., Rentsch, J., Sutter, C., Ring, L., Flachowsky, H., Habegger, R., Hoffmann, T., Hanke, M.-V., and Schwab, W. (2014).** Premature and ectopic anthocyanin formation by silencing of anthocyanidin reductase in strawberry (*Fragaria × ananassa*). *New Phytol.* **201**: 440-451.
- Fukuchi-Mizutani, M., Okuhara, H., Fukui, Y., Nakao, M., Katsumoto, Y., Yonekura-Sakakibara, K., Kusumi, T., Hase, T., and Tanaka, Y. (2003).** Biochemical and molecular characterization of a novel UDP-glucose:anthocyanin 3'-O-glucosyltransferase, a key enzyme for blue anthocyanin biosynthesis, from gentian. *Plant Physiol.* **132**: 1652–1663.
- Gachon, C.M.M., Langlois-Meurinne, M., and Saindrenan, P. (2005).** Plant secondary metabolism glycosyltransferases: the emerging functional analysis. *Trends Plant Sci.* **10**: 542–9.
- Gao, S., Feng, N., Yu, S., Yu, D., and Wang, X. (2004b).** Vasodilator constituents from the roots of *Lysidice rhodostega*. *Planta Med.* **70**: 1128–1134.
- Gao, S., Yu, S.S., Yu, D. (2004a).** Two new phloroglucinol glycosides from *Lysidice rhodostega* Hance. *Chin. Chem. Lett.* **15**: 313-315.
- Gasperotti, M., Masuero, D., Guella, G., Palmieri, L., Martinatti, P., Pojer, E., Mattivi, F., and Vrhovsek, U. (2013).** Evolution of ellagitannin content and profile during fruit ripening in *Fragaria spp.* *J. Agric. Food Chem.* **61**: 8597–8607.
- Ghose, K., Selvaraj, K., McCallum, J., Kirby, C.W., Sweeney-Nixon, M., Cloutier, S.J., Deyholos, M., Datla, R., and Fofana, B. (2014).** Identification and functional characterization of a flax UDP-glycosyltransferase glucosylating secoisolariciresinol (SECO) into secoisolariciresinol monoglucoside (SMG) and diglucoside (SDG). *BMC Plant Biol.* **14**: 82.
- Giampieri, F., Alvarez-Suarez, J.M., Tulipani, S., Gonzàles-Paramàs, A.M., Santos-Buelga, C., Bompadre, S., Quiles, J.L., Mezzetti, B., and Battino, M. (2012).** Photoprotective potential of strawberry (*Fragaria × ananassa*) extract against UV-A irradiation damage on human fibroblasts. *J. Agric. Food Chem.* **60**: 2322–7.
- Giovannoni, J. (2001).** Molecular Biology of Fruit Maturation and Ripening. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**: 725–749.

- Gowda, H., Ivanisevic, J., Johnson, C.H., Kurczy, M.E., Benton, H.P., Rinehart, D., Nguyen, T., Ray, J., Kuehl, J., Arevalo, B., Westenskow, P.D., Wang, J., Arkin, A.P., Deutschbauer, A.M., Patti, G.J., and Siuzdak, G. (2014). Interactive XCMS online: Simplifying advanced metabolomic data processing and subsequent statistical analyses. *Anal. Chem.* **86**: 6931–6939.
- Griesser, M., Hoffmann, T., Bellido, M.L., Rosati, C., Fink, B., Kurtzer, R., Aharoni, A., Muñoz-Blanco, J., and Schwab, W. (2008a). Redirection of flavonoid biosynthesis through the down-regulation of an anthocyanidin glucosyltransferase in ripening strawberry fruit. *Plant Physiol.* **146**: 1528–1539.
- Griesser, M., Vitzthum, F., Fink, B., Bellido, M.L., Raasch, C., Munoz-Blanco, J., and Schwab, W. (2008b). Multi-substrate flavonol O-glucosyltransferases from strawberry (*Fragaria x ananassa*) achene and receptacle. *J. Exp. Bot.* **59**: 2611–25.
- Häkkinen, S.H., Kärenlampi, S.O., Heinonen, I.M., Mykkänen, H.M., and Törrönen, A.R. (1999). Content of the flavonols quercetin, myricetin, and kaempferol in 25 edible berries. *J. Agric. Food Chem.* **47**: 2274–2279.
- Hanhineva, K., Rogachev, I., Kokko, H., Mintz-Oron, S., Venger, I., Kärenlampi, S., and Aharoni, A. (2008). Non-targeted analysis of spatial metabolite composition in strawberry (*Fragaria x ananassa*) flowers. *Phytochemistry* **69**: 2463–81.
- Hans, J., Brandt, W., and Vogt, T. (2004). Site-directed mutagenesis and protein 3D-homology modelling suggest a catalytic mechanism for UDP-glucose-dependent betanidin 5-O-glucosyltransferase from *Dorotheanthus bellidiformis*. *Plant J.* **39**: 319–333.
- Hansen, E.H., Møller, B.L., Kock, G.R., Bünner, C.M., Kristensen, C., Jensen, O.R., Okkels, F.T., Olsen, C.E., Motawia, M.S., and Hansen, J. (2009). De novo biosynthesis of Vanillin in fission yeast (*Schizosaccharomyces pombe*) and baker's yeast (*Saccharomyces cerevisiae*). *Appl. Environ. Microbiol.* **75**: 2765–2774.
- Harpster, M.H., Brummell, D.A., and Dunsmuir, P. (1998). Expression analysis of a ripening-specific, auxin-repressed endo-1, 4-beta-glucanase gene in strawberry. *Plant Physiol.* **118**: 1307–1316.
- He, X.Z., Wang, X., and Dixon, R.A. (2006). Mutational analysis of the Medicago glycosyltransferase UGT71G1 reveals residues that control regioselectivity for (iso)flavonoid glycosylation. *J. Biol. Chem.* **281**: 34441–34447.
- Hefner, T., Arend, J., Warzecha, H., Siems, K., and Stöckigt, J. (2002). Arbutin synthase, a novel member of the NRD1 $\beta$  glycosyltransferase family, is a unique multifunctional enzyme converting various natural products and xenobiotics. *Bioorganic Med. Chem.* **10**: 1731–1741.
- Hodge, J.E., Fisher, B.E., and Nelson, H.. (1963). Dicarbonyls, reductones, and heterocyclics produced by reaction of reducing sugars with secondary amine salts. *Am. Soc. Brew. Chem. Proc.* **83**: 84–92.
- Hoffmann, T., Kalinowski, G., and Schwab, W. (2006). RNAi-induced silencing of gene expression in strawberry fruit (*Fragaria x ananassa*) by agroinfiltration: a rapid assay for gene function analysis. *Plant J.* **48**: 818–26.
- Hu, Y. and Walker, S. (2002). Remarkable structural similarities between diverse glycosyltransferases. *Chem. Biol.* **9**: 1287–1296.
- Hughes, A.L. (2005). Gene duplication and the origin of novel proteins. *Proc. Natl. Acad. Sci. U. S. A.* **102**: 8791–8792.

- Hutchinson, A., Roy, C., Towers, G.H.N.** (1958). Synthesis of phlorin and other phenolic glucosides by plant tissues. *Nature* **181**: 841–842.
- Intelmann, D., Haseleu, G., Dunkel, A., Lagemann, A., Stephan, A., and Hofmann, T.** (2011). Comprehensive sensomics analysis of hop-derived bitter compounds during storage of beer. *J. Agric. Food Chem.* **59**: 1939–1953.
- Isayenkova, J., Wray, V., Nimtz, M., Strack, D., and Vogt, T.** (2006). Cloning and functional characterisation of two regioselective flavonoid glucosyltransferases from *Beta vulgaris*. *Phytochemistry* **67**: 1598–1612.
- Jez, J.M., Austin, M.B., Ferrer, J., Bowman, M.E., Schröder, J., and Noel, J.P.** (2000). Structural control of polyketide formation in plant-specific polyketide synthases. *Chem. Biol.* **7**: 919–30.
- Jez, J.M., Bowman, M.E., and Noel, J.P.** (2002). Expanding the biosynthetic repertoire of plant type III polyketide synthases by altering starter molecule specificity. *Proc. Natl. Acad. Sci. U. S. A.* **99**: 5319–24.
- Jez, J.M., Bowman, M.E., and Noel, J.P.** (2001). Structure-Guided Programming of Polyketide Chain-Length Determination in chalcone synthase. *Biochemistry* **40**: 14829–14838.
- Jez, J.M. and Noel, J.P.** (2000). Mechanism of chalcone synthase. pKa of the catalytic cysteine and the role of the conserved histidine in a plant polyketide synthase. *J Biol Chem.* **275**:39640-39646.
- Jia, H.-F., Chai, Y.-M., Li, C.-L., Lu, D., Luo, J.-J., Qin, L., and Shen, Y.-Y.** (2011). Abscisic Acid plays an important role in the regulation of strawberry fruit ripening. *Plant Physiol.* **157**: 188–199.
- Jia, H.-F., Lu, D., Sun, J.-H., Li, C.-L., Xing, Y., Qin, L., and Shen, Y.-Y.** (2013). Type 2C protein phosphatase ABL1 is a negative regulator of strawberry fruit ripening. *J. Exp. Bot.* **64**: 1677–87.
- Jones, D.A.** (1998). Why are so many food plants cyanogenic? *Phytochemistry* **47**: 155–162.
- Jones, P. and Vogt, T.** (2001). Glycosyltransferases in secondary plant metabolism: Tranquilizers and stimulant controllers. *Planta* **213**: 164–174.
- Jones, P.R., Møller, B.L., and Høj, P.B.** (1999). The UDP-glucose:p-hydroxymandelonitrile-O-glucosyltransferase that catalyzes the last step in synthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolor*. Isolation, cloning, heterologous expression, and substrate specificity. *J. Biol. Chem.* **274**: 35483–35491.
- Jugdé, H., Nguy, D., Moller, I., Cooney, J.M., and Atkinson, R.G.** (2008). Isolation and characterization of a novel glycosyltransferase that converts phloretin to phlorizin, a potent antioxidant in apple. *FEBS J.* **275**: 3804–3814.
- Kang, C., Darwish, O., Geretz, A., Shahan, R., Alkharouf, N., and Liu, Z.** (2013). Genome-scale transcriptomic insights into early-stage fruit development in woodland strawberry *Fragaria vesca*. *Plant Cell* **25**: 1960–78.
- Keutgen, A.J. and Pawelzik, E.** (2008). Contribution of amino acids to strawberry fruit quality and their relevance as stress indicators under NaCl salinity. *Food Chem.* **111**: 642–647.
- Kleczkowski, K., Schell, J., and Bandur, R.** (1995). Phytohormone Conjugates: Nature and Function. *CRC. Crit. Rev. Plant Sci.* **14**: 283–298.
- Klein, D., Fink, B., Arold, B., Eisenreich, W., and Schwab, W.** (2007). Functional characterization of enone oxidoreductases from strawberry and tomato fruit. *J. Agric. Food Chem.* **55**: 6705–11.

- Kosasi, S., Van der Sluis, W.G., and Labadie, R.P. (1989). Multifidol and multifidol glucoside from the latex of *Jatropha multifida*. *Phytochemistry* **28**: 2439–2441.
- Kramer, C.M., Prata, R.T.N., Willits, M.G., De Luca, V., Steffens, J.C., and Graser, G. (2003). Cloning and regiospecificity studies of two flavonoid glucosyltransferases from *Allium cepa*. *Phytochemistry* **64**: 1069–1076.
- Kristensen, C., Morant, M., Olsen, C.E., Ekstrøm, C.T., Galbraith, D.W., Møller, B.L., and Bak, S. (2005). Metabolic engineering of dhurrin in transgenic Arabidopsis plants with marginal inadvertent effects on the metabolome and transcriptome. *Proc. Natl. Acad. Sci. U. S. A.* **102**: 1779–1784.
- Kroon, J., Souer, E., de Graaff, A., Xue, Y., Mol, J., and Koes, R. (1994). Cloning and structural analysis of the anthocyanin pigmentation locus Rt of *Petunia hybrida*: characterization of insertion sequences in two mutant alleles. *Plant J.* **5**: 69–80.
- Landmann, C., Fink, B., and Schwab, W. (2007). FaGT2: A multifunctional enzyme from strawberry (*Fragaria x ananassa*) fruits involved in the metabolism of natural and xenobiotic compounds. *Planta* **226**: 417–428.
- Li, Y., Baldauf, S., Lim, E.K., and Bowles, D.J. (2001). Phylogenetic analysis of the UDP-glycosyltransferase multigene family of *Arabidopsis thaliana*. *J. Biol. Chem.* **276**: 4338–4343.
- Li, Y., Jones, L., and McQueen-Mason, S. (2003). Expansins and cell growth. *Curr. Opin. Plant Biol.* **6**: 603–610.
- Liao, Z., Chen, M., Guo, L., Gong, Y., Tang, F., Sun, X., and Tang, K. (2004). Rapid isolation of high-quality total RNA from taxus and ginkgo. *Prep. Biochem. Biotechnol.* **34**: 209–214.
- Lignou, S., Parker, J.K., Baxter, C., and Mottram, D.S. (2014). Sensory and instrumental analysis of medium and long shelf-life Charentais cantaloupe melons (*Cucumis melo* L.) harvested at different maturities. *Food Chem.* **148**: 218–229.
- Lim, E.K., Ashford, D.A., Hou, B., Jackson, R.G., and Bowles, D.J. (2004). Arabidopsis glycosyltransferases as biocatalysts in fermentation for regioselective synthesis of diverse quercetin glucosides. *Biotechnol. Bioeng.* **87**: 623–631.
- Lim, E.K., Baldauf, S., Li, Y., Elias, L., Worrall, D., Spencer, S.P., Jackson, R.G., Taguchi, G., Ross, J., and Bowles, D.J. (2003). Evolution of substrate recognition across a multigene family of glycosyltransferases in Arabidopsis. *Glycobiology* **13**: 139–145.
- Lim, E.-K. and Bowles, D.J. (2004). A class of plant glycosyltransferases involved in cellular homeostasis. *EMBO J.* **23**: 2915–2922.
- Lim, E.K., Doucet, C.J., Li, Y., Elias, L., Worrall, D., Spencer, S.P., Ross, J., and Bowles, D.J. (2002). The activity of Arabidopsis glycosyltransferases toward salicylic acid, 4-hydroxybenzoic acid, and other benzoates. *J. Biol. Chem.* **277**: 586–592.
- Lunkenbein, S., Bellido, M., Aharoni, A., Salentijn, E.M.J., Kaldenhoff, R., Coiner, H.A., Muñoz-Blanco, J., and Schwab, W., (2006a). Cinnamate Metabolism in Ripening Fruit. Characterization of a UDP-Glucose : Cinnamate Glucosyltransferase from Strawberry. *Plant Physiol.* **140**: 1047–1058.
- Lunkenbein, S., Coiner, H., De Vos, C.H.R., Schaart, J.G., Boone, M.J., Krens, F.A., Schwab, W., and Salentijn, E.M.J. (2006b). Molecular characterization of a stable antisense chalcone synthase phenotype in strawberry (*Fragaria x ananassa*). *J. Agric. Food Chem.* **54**: 2145–2153.
- Lynch, M., and Force, A. (1999). The probability of duplicate gene preservation by subfunctionalization. *Genetics*. **154**: 450–473.

- Marin, A.B., Acree, T.E., Hotchkiss, J.H., and Nagy, S.** (1992). Gas-chromatography olfactometry of orange juice to assess the effects of plastic polymers on aroma character. *J. Agric. Food Chem.* **40**: 650–654.
- Masada, S., Terasaka, K., and Mizukami, H.** (2007). A single amino acid in the PSPG-box plays an important role in the catalytic function of CaUGT2 (Curcumin glucosyltransferase), a Group D Family 1 glucosyltransferase from *Catharanthus roseus*. *FEBS Lett.* **581**: 2605–2610.
- Mathur, J., Szabados, L., Schaefer, S., Grunenberger, B., Lossow, A., Jonas-Straube, E., Schell, J., Koncz, C., and Koncz-Kálmán, Z.** (1998). Gene identification with sequenced T-DNA tags generated by transformation of Arabidopsis cell suspension. *Plant J.* **13**: 707–716.
- Medina-Puche, L., Cumplido-Laso, G., Amil-Ruiz, F., Hoffmann, T., Ring, L., Rodríguez-Franco, A., Caballero, J.L., Schwab, W., Muñoz-Blanco, J., and Blanco-Portales, R.** (2014). MYB10 plays a major role in the regulation of flavonoid/phenylpropanoid metabolism during ripening of *Fragaria x ananassa* fruits. *J. Exp. Bot.* **65**: 401–17.
- Muñoz, C., Hoffmann, T., Escobar, N.M., Ludemann, F., Botella, M. a, Valpuesta, V., and Schwab, W.** (2010). The strawberry fruit Fra a allergen functions in flavonoid biosynthesis. *Mol. Plant* **3**: 113–24.
- Negishi, O., Sugiura, K., and Negishi, Y.** (2009). Biosynthesis of vanillin via ferulic acid in *vanilla planifolia*. *J. Agric. Food Chem.* **57**: 9956–9961.
- Noguchi, A., Horikawa, M., Fukui, Y., Fukuchi-Mizutani, M., Iuchi-Okada, A., Ishiguro, M., Kiso, Y., Nakayama, T., and Ono, E.** (2009). Local differentiation of sugar donor specificity of flavonoid glycosyltransferase in Lamiales. *Plant Cell* **21**: 1556–72.
- Novák, P., Krofta, K., and Matoušek, J.** (2006). Chalcone synthase homologues from *Humulus lupulus* : some enzymatic properties and expression. **50**: 48–54.
- Offen, W., Martinez-Fleites, C., Yang, M., Kiat-Lim, E., Davis, B.G., Tarling, C.A., Ford, C.M., Bowles, D.J., and Davies, G.J.** (2006). Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *EMBO J.* **25**: 1396–1405.
- Okada, Y. and Ito, K.** (2001). Cloning and analysis of valerophenone synthase gene expressed specifically in lupulin gland of hop (*Humulus lupulus* L.). *Biosci. Biotechnol. Biochem.* **65**: 150–155.
- Okada, Y., Sano, Y., Kaneko, T., Abe, I., Noguchi, H., and Ito, K.** (2004). Enzymatic reactions by five chalcone synthase homologs from hop (*Humulus lupulus* L.). *Biosci. Biotechnol. Biochem.* **68**: 1142–1145.
- Onodera, J., Takano, M., Kishi, Y., Yokoyama, N., and Ishida, R.** (1983). Direct condensation of polyhydric phenols with glucose. *Chem. Lett.*: 1487–1488.
- Osmani, S.A., Bak, S., and Møller, B.L.** (2009). Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling. *Phytochemistry* **70**: 325–347.
- Paniego, N.B., Zuurbier, K.W., Fung, S.Y., van der Heijden, R., Scheffer, J.J., and Verpoorte, R.** (1999). Phlorisovalerophenone synthase, a novel polyketide synthase from hop (*Humulus lupulus* L.) cones. *Eur. J. Biochem.* **262**: 612–6.
- Paquette, S., Møller, B.L., and Bak, S.** (2003). On the origin of family 1 plant glycosyltransferases. *Phytochemistry* **62**: 399–413.
- Patti, G.J., Tautenhahn, R., and Siuzdak, G.** (2012). Meta-analysis of untargeted metabolomic data from multiple profiling experiments. *Nat. Protoc.* **7**: 508–16.

- Pedras, M.S., Zaharia, I.L., Gai, Y., Zhou, Y., and Ward, D.E.** (2001). In planta sequential hydroxylation and glycosylation of a fungal phytotoxin: Avoiding cell death and overcoming the fungal invader. *Proc. Natl. Acad. Sci. U. S. A.* **98**: 747–752.
- Perez, A., Rios, J., Sanz, C., and Olias, J.** (1992). Aroma components and free amino acids in strawberry variety Chandler during ripening. *J. Agric. Food Chem.* **4**: 2232–2235.
- Pérez, A.G., Olias, R., Luaces, P., and Sanz, C.** (2002). Biosynthesis of strawberry aroma compounds through amino acid metabolism. *J. Agric. Food Chem.* **50**: 4037–4042.
- Peters, J.L., Cnudde, F., and Gerats, T.** (2003). Forward genetics and map-based cloning approaches. *Trends Plant Sci.* **8**: 484–491.
- Petroni, K. and Tonelli, C.** (2011). Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Sci.* **181**: 219–229.
- Poppenberger, B., Fujioka, S., Soeno, K., George, G.L., Vaistij, F.E., Hiranuma, S., Seto, H., Takatsuto, S., Adam, G., Yoshida, S., and Bowles, D.** (2005). The UGT73C5 of *Arabidopsis thaliana* glucosylates brassinosteroids. *Proc. Natl. Acad. Sci. U. S. A.* **102**: 15253–15258.
- Potter, D., Eriksson, T., Evans, R.C., Oh, S., Smedmark, J.E.E., Morgan, D.R., Kerr, M., Robertson, K.R., Arsenault, M., Dickinson, T.A., and Campbell, C.S.** (2007). Phylogeny and classification of Rosaceae. *Plant Systematics and Evolution.* **266**: 5–43.
- Potter, D., Luby, J.J., and Harrison, R.E.** (2000). Phylogenetic Relationships among Species of *Fragaria* (Rosaceae) Inferred from Non-Coding Nuclear and Chloroplast DNA Sequences. *Syst. Bot.* **25**: 337.
- Pyysalo, T., Honkanen, E., and Hirvi, T.** (1979). Volatiles of Wild Strawberries, *Fragaria vesca* L., Compared to Those of Cultivated Berries, *Fragaria x ananassa* cv. Senga Sengana. *J. Agric. Food Chem.* **27**: 19–22.
- Raab, T., López-Ráez, J.A., Klein, D., Caballero, J.L., Moyano, E., Schwab, W., and Muñoz-Blanco, J.** (2006). FaQR, required for the biosynthesis of the strawberry flavor compound 4-hydroxy-2,5-dimethyl-3(2H)-furanone, encodes an enone oxidoreductase. *Plant Cell* **18**: 1023–1037.
- Radhamony, R.N., Prasad, A.M., and Srinivasan, R.** (2005). T-DNA insertional mutagenesis in *Arabidopsis*: A tool for functional genomics. *Electron. J. Biotechnol.* **8**: 82–106.
- Richman, A., Swanson, A., Humphrey, T., Chapman, R., McGarvey, B., Pocs, R., and Brandle, J.** (2005). Functional genomics uncovers three glucosyltransferases involved in the synthesis of the major sweet glucosides of *Stevia rebaudiana*. *Plant J.* **41**: 56–67.
- Ring, L., Yeh, S.Y., Hücherig, S., Hoffmann, T., Blanco-Portales, R., Fouche, M., Villatoro, C., Denoyes, B., Monfort, A., Caballero, J.L., Muñoz-Blanco, J., Gershenson, J., and Schwab, W.** (2013). Metabolic interaction between anthocyanin and lignin biosynthesis is associated with peroxidase FaPRX27 in strawberry fruit. *Plant Physiol.* **163**: 43–60.
- Rodin, J.O., Himel, C.M., Silverstein, R.M., Leeper, R.W., and Gortner, W.A.** (1965). Volatile flavor and aroma components of pineapple. I. Isolation and tentative identification of 2,5-dimethyl-4-hydroxy-3(2h)-furanone. *J Food Sci* **30**: 280–285.
- Roscher, R., Herderich, M., Steffen, J.P., Schreier, P., and Schwab, W.** (1996). 2,5-dimethyl-4-hydroxy-3[2H]-furanone 6'O-malonyl-beta-D- glucopyranoside in strawberry fruits. *Phytochemistry* **43**: 155–159.
- Roscher, R., Schreier, P., and Schwab, W.** (1997). Metabolism of 2,5-dimethyl-4-hydroxy-3(2H)-furanone in detached ripening strawberry fruits. *J. Agric. Food Chem.* **45**: 3202–3205.

- Ross, J., Li, Y., Lim, E., and Bowles, D.J. (2001). Higher plant glycosyltransferases. *Genome Biol.* **2**: REVIEWS3004.
- Saito, K. and Matsuda, F. (2010). Metabolomics for functional genomics, systems biology, and biotechnology. *Annu. Rev. Plant Biol.* **61**: 463–489.
- Sakho, M., Chassagne, D., and Crouzet, J. (1997). African Mango Glycosidically Bound Volatile Compounds. *J. Agric. Food Chem.* **45**: 883–888.
- Scalbert, A., Manach, C., Morand, C., Rémésy, C., and Jiménez, L. (2005). Dietary polyphenols and the prevention of diseases. *Crit. Rev. Food Sci. Nutr.* **45**: 287–306.
- Schieberle, P. and Hofmann, T. (1997). Evaluation of the Character Impact Odorants in Fresh Strawberry Juice by Quantitative Measurements and Sensory Studies on Model Mixtures. *J. Agric. Food Chem.* **45**: 227–232.
- Schiefner, A., Sinz, Q., Neumaier, I., Schwab, W., and Skerra, A. (2013). Structural basis for the enzymatic formation of the key strawberry flavor compound 4-hydroxy-2,5-dimethyl-3(2H)-furanone. *J. Biol. Chem.* **288**: 16815–16826.
- Schijlen, E.G.W.M., de Vos, C.H.R., Martens, S., Jonker, H.H., Rosin, F.M., Molthoff, J.W., Tikunov, Y.M., Angenent, G.C., van Tunen, A.J., and Bovy, A.G. (2007). RNA interference silencing of chalcone synthase, the first step in the flavonoid biosynthesis pathway, leads to parthenocarpic tomato fruits. *Plant Physiol.* **144**: 1520–1530.
- Schmidt, S., Jürgenliemk, G., Skaltsa, H., and Heilmann, J. (2012). Phloroglucinol derivatives from *Hypericum empetrifolium* with antiproliferative activity on endothelial cells. *Phytochemistry* **77**: 218–225.
- Schwab, W. (1998). Application of stable isotope ratio analysis explaining the bioformation of 2,5-dimethyl-4-hydroxy-3(2H)-furanone in plants by a biological Maillard reaction. *J. Agric. Food Chem.* **46**: 2266–2269.
- Schwab, W. (2013). Natural 4-hydroxy-2,5-dimethyl-3(2h)-furanone (furaneol®). *Molecules* **18**: 6936–6951.
- Schwab, W., Davidovich-Rikanati, R., and Lewinsohn, E. (2008). Biosynthesis of plant-derived flavor compounds. *Plant J.* **54**: 712–732.
- Schwab, W., Hoffmann, T., Kalinowski, G., and Preuß, A. (2011). Functional Genomics in Strawberry Fruit through RNAi-mediated Silencing. *Gene, Genomes and Genomics* **5**: 91–101.
- Seeram, N.P., Lee, R., Scheuller, H.S., and Heber, D. (2006). Identification of phenolic compounds in strawberries by liquid chromatography electrospray ionization mass spectroscopy. *Food Chem.* **97**: 1–11.
- Shao, H., He, X., Achnine, L., Blount, J.W., Dixon, R.A., and Wang, X. (2005). Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula*. *Plant Cell* **17**: 3141–3154.
- Shiu, W.K.P., Rahman, M.M., Curry, J., Stapleton, P., Zloh, M., Malkinson, J.P., and Gibbons, S. (2012). Antibacterial acylphloroglucinols from *hypericum olympicum*. *J. Nat. Prod.* **75**: 336–343.
- Shulaev, V., Sargent, D.J., Crowhurst, R.N., Mockler, T.C., Folkerts, O., Delcher, A.L., Jaiswal, P., Mockaitis, K., Liston, A., Mane, S.P., Burns, P., Davis, T.M., Slovin, J.P., Bassil, N., Hellens, R.P., Evans, C., Harkins, T., Kodira, C., Desany, B., Crasta, O.R., Jensen, R.V., Allan, A.C., Michael, T.P., Setubal, J.C., Celton, J.M., Rees, D.J., Williams, K.P., Holt, S.H., Ruiz Rojas, J.J., Chatterjee, M., Liu, B., Silva, H., Meisel, L., Adato, A., Filichkin, S.A., Troggio, M., Viola, R., Ashman, T.L., Wang, H., Dharmawardhana, P., Elser, J., Raja, R., Priest, H.D., Bryant, D.W. Jr, Fox, S.E., Givan, S.A., Wilhelm, L.J.,

- Naithani, S., Christoffels, A., Salama, D.Y., Carter, J., Lopez Girona, E., Zdepski, A., Wang, W., Kerstetter, R.A., Schwab, W., Korban, S.S., Davik, J., Monfort, A., Denoyes-Rothan, B., Arus, P., Mittler, R., Flinn, B., Aharoni, A., Bennetzen, J.L., Salzberg, S.L., Dickerman, A.W., Velasco, R., Borodovsky, M., Veilleux, R.E., and Folta, K.M. (2011). The genome of woodland strawberry (*Fragaria vesca*). *Nat. Genet.* **43**: 109–116.
- Singh, I.P., Sidana, J., Bharate, S.B., and Foley, W.J. (2010). Phloroglucinol compounds of natural origin: synthetic aspects. *Nat. Prod. Rep.* **27**: 393–416.
- Sinz, Q., and Schwab, W. (2012). Metabolism of amino acids, dipeptides and tetrapeptides by *Lactobacillus sakei*. *Food Microbiol.* **29**: 215–223.
- Slovin, J.P., Schmitt, K., and Folta, K.M. (2009). An inbred line of the diploid strawberry *Fragaria vesca* f. *semperflorens* for genomic and molecular genetic studies in the Rosaceae. *Plant Methods* **5**: 15.
- Suh, D.Y., Fukuma, K., Kagami, J., Yamazaki, Y., Shibuya, M., Ebizuka, Y., and Sankawa, U. (2000). Identification of amino acid residues important in the cyclization reactions of chalcone and stilbene synthases. *Biochem. J.* **350 Pt 1**: 229–35.
- Suksamrarn, A., Eiamong, S., Piyachaturawat, P., and Byrne, L.T. (1997). A phloracetophenone glucoside with choleric activity from *Curcuma comosa*. *Phytochemistry* **45**: 103–105.
- Tautenhahn, R., Patti, G.J., Kalisiak, E., Miyamoto, T., Schmidt, M., Lo, F.Y., McBee, J., Baliga, N.S., and Siuzdak, G. (2011). metaXCMS: second-order analysis of untargeted metabolomics data. *Anal. Chem.* **83**: 696–700.
- Thorsøe, K.S., Bak, S., Olsen, C.E., Imberty, A., Breton, C., and Lindberg Møller, B. (2005). Determination of catalytic key amino acids and UDP sugar donor specificity of the cyanohydrin glycosyltransferase UGT85B1 from *Sorghum bicolor*. Molecular modeling substantiated by site-specific mutagenesis and biochemical analyses. *Plant Physiol.* **139**: 664–673.
- Toth, J., Cutforth, T., Gelinis, A.D., Bethoney, K.A., Bard, J., and Harrison, C.J. (2001). Crystal Structure of an Ephrin Ectodomain. *Dev. Cell* **1**: 83–92.
- Tsukamoto, S., Tomise, K., Aburatani, M., Onuki, H., Hirorta, H., Ishiharajima, E., and Ohta, T. (2004). Isolation of cytochrome P450 inhibitors from strawberry fruit, *Fragaria ananassa*. *J. Nat. Prod.* **67**: 1839–41.
- Ubeda, C., San-Juan, F., Concejero, B., Callejón, R.M., Troncoso, A.M., Morales, M.L., Ferreira, V., and Hernández-Orte, P. (2012). Glycosidically bound aroma compounds and impact odorants of four strawberry varieties. *J. Agric. Food Chem.* **60**: 6095–6102.
- Ververidis, F., Trantas, E., Douglas, C., Vollmer, G., Kretzschmar, G., and Panopoulos, N. (2007). Biotechnology of flavonoids and other phenylpropanoid-derived natural products. Part I: Chemical diversity, impacts on plant biology and human health. *Biotechnol. J.* **2**: 1214–1234.
- Visioli, F., De La Lastra, C.A., Andres-Lacueva, C., Aviram, M., Calhau, C., Cassano, A., D'Archivio, M., Faria, A., Favé, G., Fogliano, V., Llorach, R., Vitaglione, P., Zoratti, M., Edeas, M. (2011). Polyphenols and Human Health: A Prospectus. *Crit. Rev. Food Sci. Nutr.* **51**: 524–546.
- Vogt, T. (2010). Phenylpropanoid biosynthesis. *Mol. Plant* **3**: 2–20.
- Vogt, T. and Jones, P. (2000). Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. *Trends Plant Sci.* **5**: 380–386.
- Wang, J. and Hou, B. (2009). Glycosyltransferases: Key players involved in the modification of plant secondary metabolites. *Front. Biol. China* **4**: 39–46.

- Waterhouse, P.M. and Helliwell, C.A.** (2003). Exploring plant genomes by RNA-induced gene silencing. *Nat. Rev. Genet.* **4**: 29–38.
- Wein, M., Lavid, N., Lunkenbein, S., Lewinsohn, E., Schwab, W., and Kaldenhoff, R.** (2002). Isolation, cloning and expression of a multifunctional O-methyltransferase capable of forming 2,5-dimethyl-4-methoxy-3(2H)-furanone, one of the key aroma compounds in strawberry fruits. *Plant J.* **31**: 755–765.
- Wein, M., Lewinsohn, E., and Schwab, W.** (2001). Metabolic fate of isotopes during the biological transformation of carbohydrates to 2,5-dimethyl-4-hydroxy-3(2H)-furanone in strawberry fruits. *J. Agric. Food Chem.* **49**: 2427–2432.
- Weis, M., Lim, E.K., Bruce, N.C., and Bowles, D.J.** (2008). Engineering and kinetic characterisation of two glucosyltransferases from *Arabidopsis thaliana*. *Biochimie* **90**: 830–834.
- Wilkinson, D.G.** (2000). Eph receptors and ephrins: regulators of guidance and assembly. *Int. Rev. Cytol.* **196**: 177–244.
- Xu, F., Li, L., Zhang, W., Cheng, H., Sun, N., Cheng, S., and Wang, Y.** (2012). Isolation, characterization, and function analysis of a flavonol synthase gene from *Ginkgo biloba*. *Mol. Biol. Rep.* **39**: 2285–96.
- Xu, H., Zhang, F., Liu, B., Huhman, D. V., Sumner, L.W., Dixon, R. a, and Wang, G.** (2013). Characterization of the formation of branched short-chain fatty acid:CoAs for bitter acid biosynthesis in hop glandular trichomes. *Mol. Plant* **6**: 1301–17.
- Yamazaki, Y., Suh, D.-Y., Sitthithaworn, W., Ishiguro, K., Kobayashi, Y., Shibuya, M., Ebizuka, Y., and Sankawa, U.** (2001). Diverse chalcone synthase superfamily enzymes from the most primitive vascular plant, *Psilotum nudum*. *Planta* **214**: 75–84.
- Yonekura-Sakakibara, K. and Hanada, K.** (2011). An evolutionary view of functional diversity in family 1 glycosyltransferases. *Plant J.* **66**: 182–193.
- Zhang, Y., Seeram, N.P., Lee, R., Feng, L., and Heber, D.** (2008). Isolation and identification of strawberry phenolics with antioxidant and human cancer cell antiproliferative properties. *J. Agric. Food Chem.* **56**: 670–5.
- Zhang, Y.-J., Abe, T., Tanaka, T., Yang, C.-R., and Kouno, I.** (2002). Two new acylated flavanone glycosides from the leaves and branches of *Phyllanthus emblica*. *Chem. Pharm. Bull. (Tokyo)*. **50**: 841–843.
- Zuurbier, K.W.M., Leser, J., Berger, T., Hofte, A.J.P., Schröder, G., Verpoorte, R., and Schröder, J.** (1998). 4-hydroxy-2-pyrone formation by chalcone and stilbene synthase with nonphysiological substrates. *Phytochemistry* **49**: 1945–1951.

## Curriculum vitae

Name: Song Chuankui (Male)  
School: Technische Universität München, Germany  
Institute: Biotechnology of Natural Products  
Phone +49-176-9986-3324;  
e-mail: [sckfriend@163.com](mailto:sckfriend@163.com)/[ch.song@mytum.de](mailto:ch.song@mytum.de)

### Education

2012/04-now: PhD student in the group of Biotechnology of Natural Products, Technische Universität München, Munich, Germany

2008/09-2011/07: MSc Tea science (Biochemistry of tea science), Northwest A&F University, Yangling, China

2009/09-2011/01: Joint training master students (Detoxification function of tea), Peking University, Beijing, China

2004/09-2008/07: BSc Horticulture, Northwest A&F University, Yangling, China

### Publication

1. **Chuankui Song**, Ludwig Ring, Thomas Hoffmann, Fong-Chin Huang, Janet P. Slovin, and Wilfried Schwab, Acylphloroglucinol biosynthesis in strawberry fruit. *Plant Physiology* (2015), doi: 10.1104/pp.15.00794
2. **Chuankui Song**, Yanli Wang, Zhengcao Xiao, Bin Xiao. Protective effects of Green Tea Polyphenols against Benzo[a]pyrene-induced reproductive and developmental toxic effects in Japanese Medaka (*Oryzias latipes*). *Journal of Functional Foods*, 14 (2015): 354–362
3. Yan-Fang Sun, **Chuankui Song**, Helmut Viernstein, Frank Unger, Zong-Suo Liang. Apoptosis of human breast cancer cells induced by microencapsulated betulinic acid from sour jujube fruits through the mitochondria transduction pathway. *Food Chemistry*, 138(2013): 1998-2007.

## Appendix

**Appendix Table 1.** Primers used for real-time PCR, cloning, RNAi and overexpression

Name	Sequence (5'->3')	Purpose
Gene21343F	TCTTGCTACGAAATGCGATG	RT-qPCR
Gene21343R	TTCACACAAGCAACTCTTCTGA	RT-qPCR
Gene10776F	GGTCCGAACACATCCTGAGT	RT-qPCR
Gene10776R	ATGATGGGGTATACCGATGG	RT-qPCR
Gene33865F	GATACAGAACATTCGGTGGGTTT	RT-qPCR
Gene33865R	CAAGCATATGACTACTGGCATGAT	RT-qPCR
Gene00897F	CACACCCACAGCTAGATAAGAGG	RT-qPCR
Gene00897R	AAACAGATGCACAGTTGCTTTCT	RT-qPCR
Gene23054F	AAAACGGGAACCAACATCAA	RT-qPCR
Gene23054R	CTACTGCTAGGGACAAAGATTGC	RT-qPCR
Gene35152F	TTGTTGGGCTATGGATTGT	RT-qPCR
Gene35152R	AATCCCGCGATACCTCTACC	RT-qPCR
Gene03515F	ATTCAACTACGATATGCTCATGG	RT-qPCR
Gene03515R	GTTCAGCAATGGTGTATCAACAG	RT-qPCR
Gene27098F	CTGCATGGGTAAATGATAAGAGG	RT-qPCR
Gene27098R	CCAGTAACAGCATTAAAGGAGAGC	RT-qPCR
Gene19724F	TTGCAATCTCTCTGGATCCTT	RT-qPCR
Gene19724R	CTCTCTGCGCTAACATGAAGAAC	RT-qPCR
Gene22502F	GTTGGACTTGGTCAAAGATTGAG	RT-qPCR
Gene22502R	AAGTGTGTAGATCAAACGCCTTC	RT-qPCR
Gene30399F	TTGGGTAAAGGGGATAGGTAGAG	RT-qPCR
Gene30399R	TGGTTAGGATGCTGATTCAAAAC	RT-qPCR
Gene03472F	CCAGGTTGATTTCTCTTCGTAAC	RT-qPCR
Gene03472R	GGCGGTATCCATCTTAGAGAAAGT	RT-qPCR
Gene21343Fo	CGC GGATCCTCTTGCTACGAAATGCGATG	RNAi
Gene21343Re	CCCAAGCTTTTTACACAAGCAACTCT TCTGA	RNAi
Gene33865Fo	CGCGGATCCCACCG GTCCACCTT ATG	RNAi
Gene33865Re	CCC AAGCTTGCCCTTCCGTCAATTCCT	RNAi
Gene10776Fo	CGCGGATCCATGGCTGCAGTTGCTCAAGCG	Over expression(OE)
Gene10776Re	CGAGCTCCTAAGTTCGGATGTGATC	Over expression(OE)
Gene00897Fo	CGCGGATCCATGGAGAGTTTCATGCGT	Over expression(OE)
Gene00897Re	CGAGCTCTTAACAGTGTGGTGCAGAA	Over expression(OE)
Gene21343Fo-RNAi	AGA GGC ACC TAT GCC GAC TA	RT-qPCR(RNAi)
Gene21343Re-RNAi	CAT CGC ATT TCG TAG CAA GA	RT-qPCR(RNAi)
Gene33865Fo-RNAi	TCG ACG TGT GAT TCT AAG CG	RT-qPCR(RNAi)
Gene33865Re-RNAi	CAA GCA TAT GAC TAC TGG CAT GA	RT-qPCR(RNAi)
Gene10776Fo-OE	ATGATGGGGTATACCGATGG	RT-qPCR(OE)
Gene10776Re-OE	GGTCCGAACACATCCTGAGT	RT-qPCR(OE)
Gene00897Fo-OE	CACACCCACAGCTAGATAAGAGG	RT-qPCR(OE)
Gene00897Re-OE	AAACAGATGCACAGTTGCTTTCT	RT-qPCR(OE)
CHS2.1F(Gene26825)	CGCGGATCCATGGTGACCGTTGAGGAA	Whole Length CHS
CHS2.1R(Gene26825)	CCGGAATTCTCAAGCAGATACACTGTG	Whole Length CHS
CHS2.2F(Gene26826)	CGCGGATCCATGGTGACCGTCGAGGAA	Whole Length CHS
CHS2.2R(Gene26826)	CCGGAATTCTCAAGCAGCCACACTGTG	Whole Length CHS
CHS2.3-F	CGCGGATCCATGGTGACCGTCGAGGAA	Whole Length CHS
CHS2.3-R	CCGGAATTCTCAAGCAGCCACACTGTG	Whole Length CHS

**Appendix Table 2.** Primers used for real-time PCR, cloning and RNAi of FaGTs

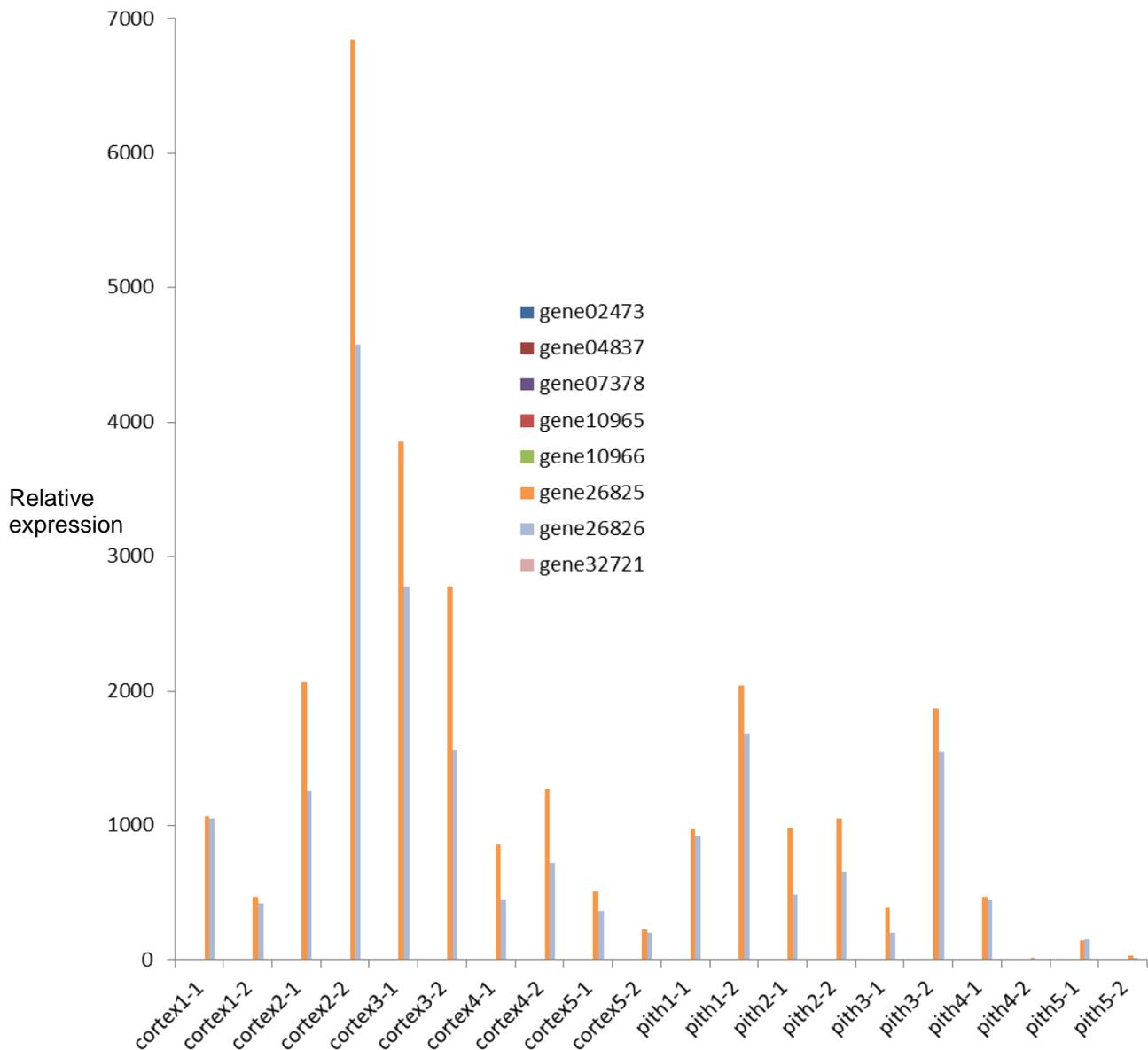
Primer	Sequence (5'→3')	Purpose
sGT7-1-qPCR-F	AAT TGC GGG AGT GGT TAT TG	RT- PCR
sGT7-1-qPCR-R	CCG ACT AGG TCG TAC CGT TC	
sGT2-8-qPCR-F	CCC CCA TAT CAC AGC GTA TC	RT- PCR
sGT2-8-qPCR-R	TCG TCT TGG AGT CCC TGA GT	
sGT3-5-qPCR-F	TCG AAA AGG AGG GAG GAA AC	RT- PCR
sGT3-5-qPCR-R	CTT GCC ATC GGA ACT GAG AG	
sGT6-9-qPCR-F	TAA TAA CCA CCC CAG CCA AT	RT- PCR
sGT6-9-qPCR-R	CGG TTT TAC CCC AGT CGT TA	
sGT4-3-qPCR-F	CTA CAT ACG CGT TTT TCA CGT C	RT- PCR
sGT4-3-qPCR-R	CCA AAT CAG TAG CCG AGT CC	
Gene06602-Sall-F	ACG CGT CGA CTC ATG CTC TCC TAT CTT CAC CT	Clone
06602-Not1-R	ATTTGCGGCCGCTTAGGCTAGAAAACA ACCAT	Clone
26479-BamH1-F	CGCGGATCCATGGAGAGTAGGAACCACC	Clone
26479-Sma1-R	TCC CCC GGG TCA GGC CAA ACT TCT GAC G	Clone
22709-Sma1-F	TCC CCC GGG TAT GAA GAG AGC AGA GCT C	Clone
22709-Sall -R	ACG CGT CGA CTC AGT TAT TCT CAA GGT T	Clone
22710-BamH1-F	CGC GGA TCC ATG GAG AGA GCA GAG CTG	Clone
22710-Sma1-R	TCC CCC GGG TCA AGT GAT TTT TAG GCT CC	Clone
20833-BamH1-F	CGC GGA TCC ATG GAC TCC ATC ACT GTC	Clone
20833-Sma1-R	TCC CCC GGG TCA TGT TGA ACG ACA TCC T	Clone
24224- BamH1-F	CGC GGA TCC ATG GAG AAA CCT GCA GAG	Clone
24224-Sma1-R	TCC CCC GGG TTAGAT CTG ATC AATAAA ATG TC	Clone
24225- BamH1-F	CGC GGA TCC ATG AAG CAA TGG GTA GAG A	Clone
24225- Sma1-R	TCC CCC GGG TTA AAT TTG ATC AAT GAA GTG	Clone
24226- BamH1-F	CGC GGA TCC ATG AAG CAA TCG GCA GAG C	Clone
24226- Sma1-R	TCC CCC GGG TTA AAT TTG ATC AAT AAA GTG T	Clone
26342- BamH1-F	CGC GGA TCC ATG GAT TCA GAA CCT CCG	Clone
26342- Sma1-R	TCC CCC GGG TCA GTT CTT CTT CAG TGA C	Clone
26353- BamH1-F	CGC GGA TCC ATG GAA ATC AAA ACT CAT CA	Clone
26353- Sma1-R	TCC CCC GGG TCA AGA CCC AAA TGA CCT C	Clone
22709-RNAi-F	CGC GGA TCC GGTACGACCTAGTCGGCAAA	RNAi
22709-RNAi-R	CCC AAG CTT TCACCCTTGGTGTCAATCAA	RNAi

**Appendix Table 3.** Primers used for real-time PCR, cloning and RNAi of FaGTs

Primers	Sequence (5'->3')	Purpose
N07876-BamH1-F	CGC GGA TCC ATG AAG AAA GCA GAG CTA GT	Clone
N07876-SmaI-R	TCC CCC GGG TTA CTC AGA ACC AAA ATG ATT	Clone
N29215 -BamH1-F	CGC GGA TCC ATG CTT CTC AAA GGT GCT	Clone
N29215 -SmaI-R	TCC CCC GGG TCA CAT CAC ATC TTC ATT GTC	Clone
N00126-BamH1-F	CGC GGA TCC ATG GGT AGC GAA AGC CAT GA	Clone
N00126-SmaI-R	TCC CCC GGG TTA TCC TAG TTG TTG GGA TTT C	Clone
N08733-BamH1-F	CGC GGA TCC ATG GCA TCA CCA CCG TTA	Clone
N08733-SmaI-R	TCC CCC GGG TCA ACT CCT GCT TCT GCG	Clone
N12684-BamH1-F	CGC GGA TCC ATG GAG ATG AAG TCC AAA GAT	Clone
N12684-SmaI-R	TCC CCC GGG TCA AGA CGT TGA AGC TGG	Clone
N14947-BamH1-F	CGC GGA TCC ATG AGC TCT TCC TCT GCA A	Clone
N14947-SmaI-R	TCC CCC GGG TCA AAC CTC ATG ATG CGC	Clone
N20721-BamH1-F	CGC GGA TCC ATG AGC TTT CTC AGT TCT G	Clone
N20721-SmaI-R	TCC CCC GGG TCA AAC CTC ATT ATG CAC A	Clone
N26345-BamH1-F	CGC GGA TCC ATG GGT AGC GAA TGC CAT	Clone
N26345-SmaI-R	TCC CCC GGG TTA TCC TAG TTG TTG GGA TT	Clone
N26352-BamH1-F	CGC GGA TCC ATG GAA ACC AAA ACT CAT C	Clone
N26352-SmaI-R	TCC CCC GGG TCA AGA TAA AGA CCT CAA CTC	Clone
N29531-BamH1-F	CGC GGA TCC ATG GCG GAG ATT AGC TCA	Clone
N29531-SmaI-R	TCC CCC GGG TCA AAC TTT CTG GCC TTG C	Clone
N25368-BamH1-F	CCG GAA TTC ATG GAG AAA GAA CAG AGA GA	Clone
N25368-SmaI-R	TCC CCC GGG TTA ATG CTT AAC CAA CGT C	Clone
N35023-BamH1-F	CCG GAA TTC ATG GAA AAA GAG CTG TTT C	Clone
N35023-SmaI-R	TCC CCC GGG CTA CTC TCC TTC TTC TTC CA	Clone
GT00126-qPCR-F	AAC CAC AGT TCG AGG AGC TT	RT-PCR
GT00126-qPCR-R	GTC CGA AGC ACA CAT AGC AA	RT-PCR
GT07876-qPCR-F	GTC AGG CTC GGT CCA ATT TA	RT-PCR
GT07876-qPCR-R	GGC CCA GAG CTA TCT CCT TC	RT-PCR
GT07876 -RNAi-F	CGC GGA TCC TGA TTG GTT AGT CCC CGG TA	RNAi
GT07876 -RNAi-R	CCC AAG CTT GCC AGC TGA TGA TCT TGT CA	RNAi

**Appendix Table 4.** Primers used for mutagenesis

Primer	Sequence
D259R-F	CTAAACAGTAATGAAAGTAGTGTACGTTCCGGACGAGGTTAAGAAG
D259R-R	CTTCTTAACCTCGTCCGAACGTACACTACTTTCATTACTGTTTAG
D343E-F	CCTGAAGGGTTCCTCGAGCGAACAGTTGGGATTGG
D343E-R	CCAATCCCAACTGTTGCTCGAGGAACCCTTCAGG
A389V-F	CACGGTGTGCCGTTGTGACGTGGCCGTTGTAC
A389V-R	GTACAACGGCCACGTCACAACCGGCACACCGTG
K458R-F	GAAATGAGTGAGAAGGGCAGGAAAGCTTTGATGGATGGC
K458R-R	GCCATCCATCAAAGCTTTCCTGCCCTTCTCACTCATTTT
D445E-F	GGAACTTGACAGTAGTGAGATAAGGAAGAGAGTGAAAG
D445E-R	CTTTCACTCTCTTCCTTATCTCACTACTGTCAAGTTCC
GI433AV-F	GCGAAAGAAATAGAGAGAGCTGTAAGAGAGGTGATGGAAGTTG
GI433AV-R	CAAGTTCCATCACCTCTCTTACAGCTCTCTCTATTTCTTTTCGC
V383A-F	GGCATGGCGTGCCGATTGCGACATGGCCTATGTATGC
V383A-R	GCATACATAGGCCATGTGCGCAATCGGCACGCCATGCC



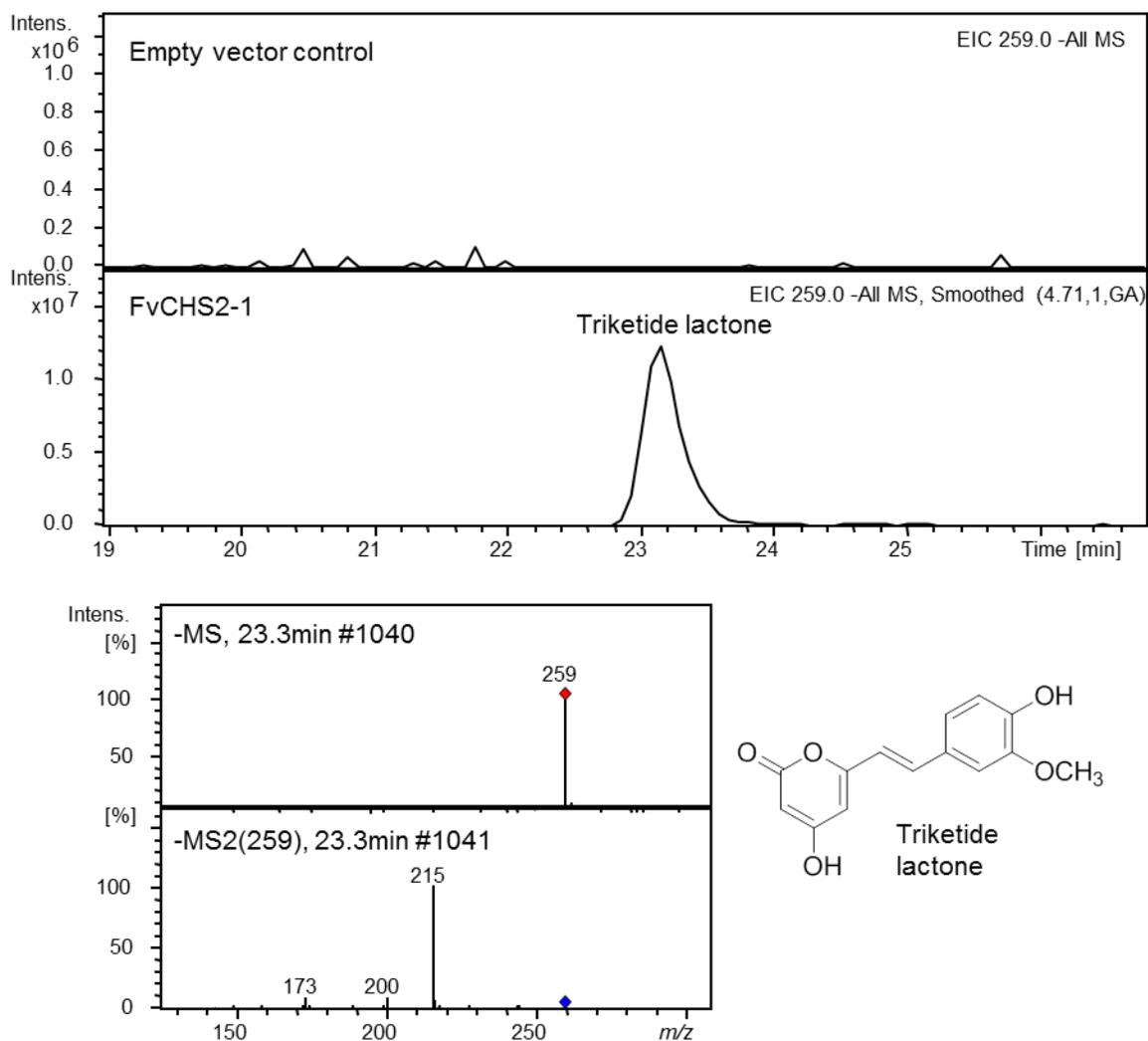
**Appendix Figure 1.** Relative expression levels of putative chalcone synthase genes in strawberry (*F. vesca* Hawaii-4) fruit tissue (cortex and pith) during early stages of fruit development. Samples are named as “tissue\_stage\_replicate”. Such as “cortex1-1” means “cortex-stage1-replicate1”. Stage 1 : prefertilization stage , Stage 2 : 2 to 4 d postanthesis, stage 3 : complete loss of anthers, and a heart stage embryo inside each seed, stage 4: embryos adopt torpedo or walking stick morphology, stage 5 :embryos and achenes maturation

[http://bioinformatics.towson.edu/strawberry/newpage/Search\\_By\\_Gene\\_Desc.aspx](http://bioinformatics.towson.edu/strawberry/newpage/Search_By_Gene_Desc.aspx)

## Appendix

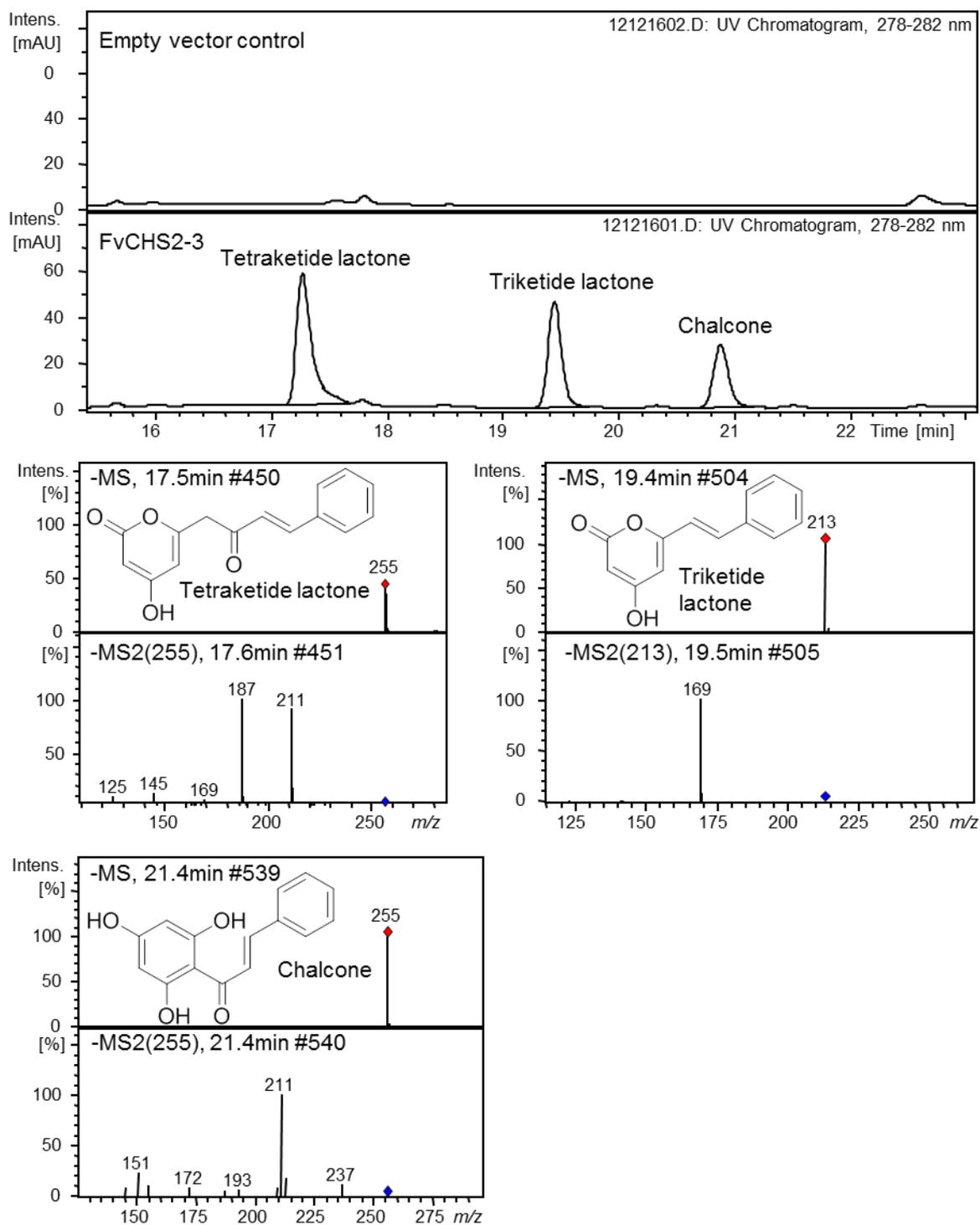
		10	20	30	40	50
FvCHS2-1	1	MVTVEEVKRA	QRAEGPATVL	AIGTATPPNC	IDQSTYPDYY	FRITNSEHKA
FvCHS2-1	1	.....	.....	.....	.....	.....
FvCHS2-3	1	.....	.....	.....	.....	.....
		60	70	80	90	100
FvCHS2-1	51	ELKEKFQRM	DKSMIKKRYM	YLTEEILKEN	PSMCEYMAPS	LDARQDMVVV
FvCHS2-1	51	.....	.....	.....	.....	.....
FvCHS2-3	51	.....	.....	.....	.....	.....
		110	120	130	140	150
FvCHS2-1	101	EIPKLGKDA	VKAIKEWGQP	KSRITHLVFC	TTSGVDMPGA	DYQLTKLLGL
FvCHS2-1	101	.....E..	.....	..K.....	.....	.....
FvCHS2-3	101	.....	.....	.....	.....	.....
		160	170	180	190	200
FvCHS2-1	151	RPSVKRLMY	QQGCFAGGTV	LRLAKDLAEN	NRGARVLVVC	SEITAVTFRG
FvCHS2-1	151	.....	.....	.....	.....	.....
FvCHS2-3	151	.....	.....	.....	.....	.....
		210	220	230	240	250
FvCHS2-1	201	PSDTHLDSL	GQALF <sup>372</sup> GDGAA	AIIVGSDPLP	EVERPLFELV	SAAQTILPDS
FvCHS2-1	201	.....	.....	.....	.....	.....
FvCHS2-3	201	.....	.....	.....	.....	.....
		260	270	280	290	300
FvCHS2-1	251	DGAIDGHLRE	VGLTFHLLKD	VPGLISKNIE	KSLNEAFKPL	NITDWNLSLFW
FvCHS2-1	251	.....	.....	.....	.....	.....
FvCHS2-3	251	.....	.....	.....	.....	.....
		310	320	330	340	350
FvCHS2-1	301	IAHPGGPAIL	DQVEAKLALK	PEKLEATRHI	LSEYGNMSSA	CVLFILDEVR
FvCHS2-1	301	.....	.....	.....	.....	.....
FvCHS2-3	301	.....	.....	.....	.....	.....
		360	370	380	390	
FvCHS2-1	351	RKSAANGHKT	TGEGLEWGV	FG <sup>372</sup> FGPGLTVE	TVVLHVSVA*	
FvCHS2-1	351	.R.....	..K.....	.....	.....A.*	
FvCHS2-3	351	.R.....	..K.....	.....	.....A.*	

**Appendix Figure 2.** Comparison of the deduced amino acid sequences of CHS enzymes from *F. vesca*. FvCHS2-1 (gene26825) and FvCHS2-2 (gene26826) are chalcone synthases from *Fragaria vesca*. FvCHS2-3 is a third protein whose corresponding gene was cloned from *F. vesca* fruit. The sequences were aligned using the ClustalW program. The highly conserved active site loop of CHS enzymes, G<sup>372</sup>FGPG and two Phe residues (Phe215 and Phe265), important in determining the substrate specificity of CHS are boxed.

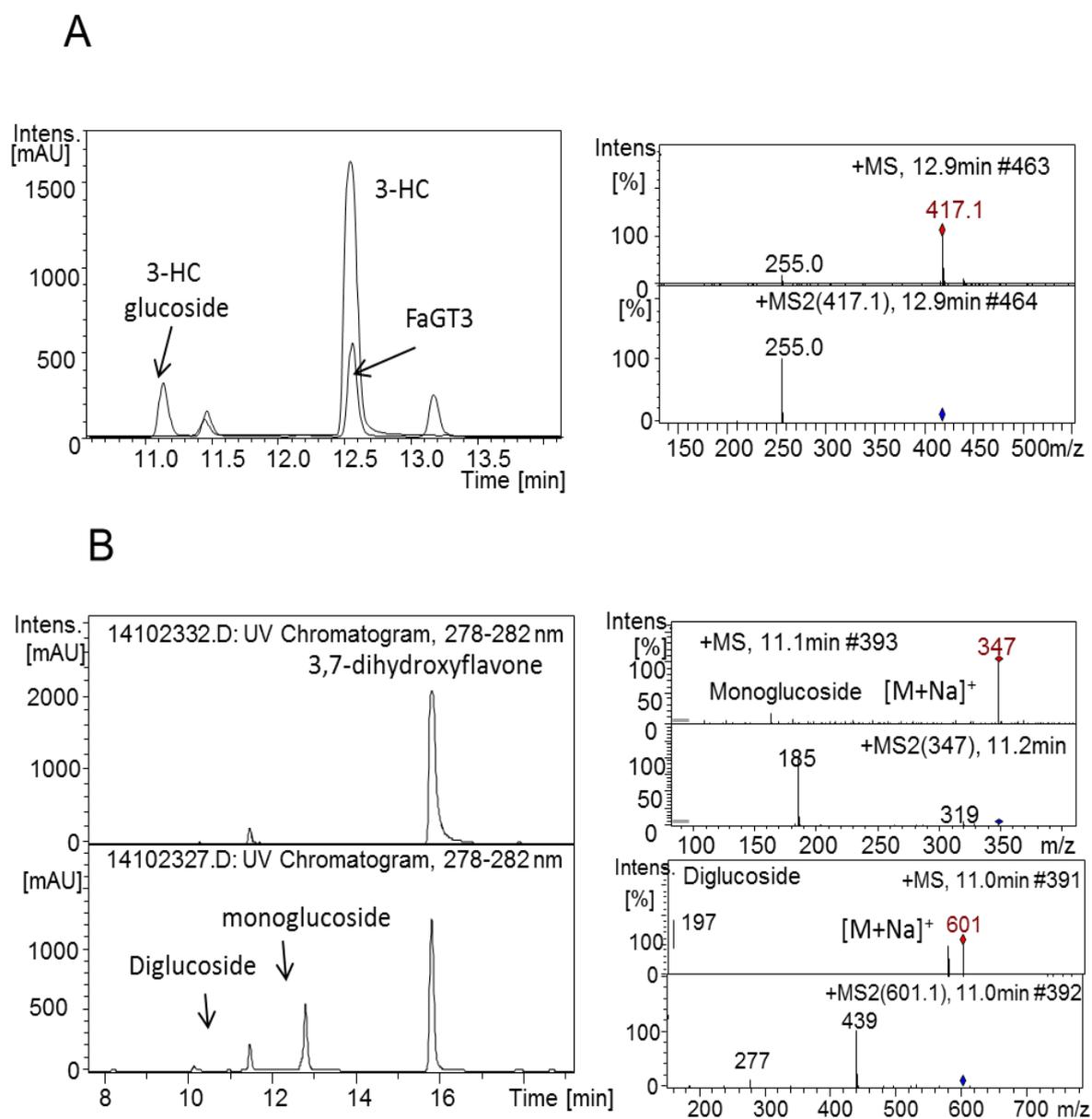


**Appendix Figure 3.** LC-MS analysis of the product formed by the empty vector control and FvCHS2-1 from the starter molecule feruloyl-CoA. MS and MS2 spectra of the triketide lactone.

## Appendix

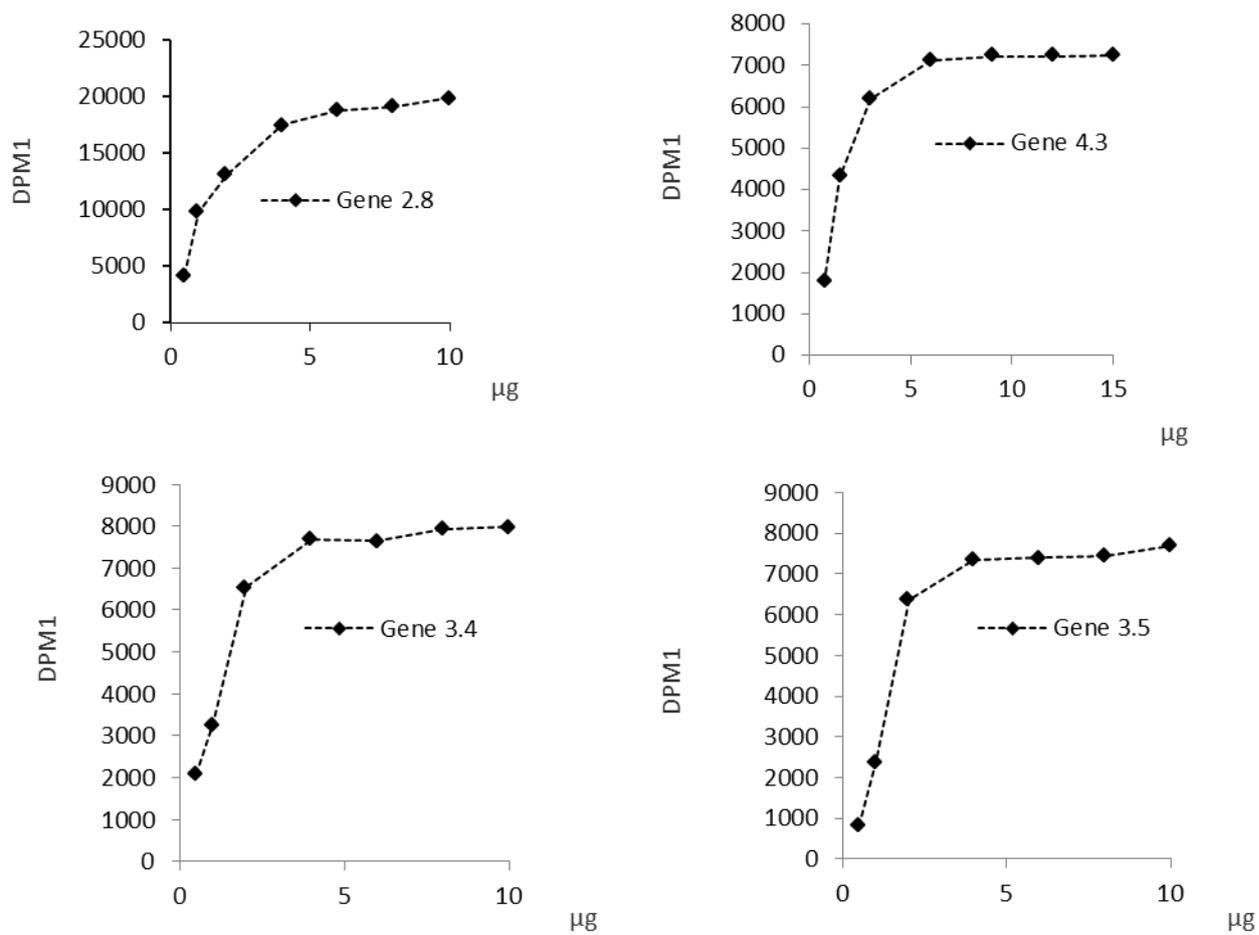


**Appendix Figure 4.** LC-MS analysis of products formed by the empty vector control and FvCHS2-3 from the starter molecule cinnamoyl-CoA. MS and MS2 spectra of tetraketide lactone, triketide lactone and chalcone.



**Appendix Figure 5.** LC-MS analysis of products formed by FaGT24224 using 3-hydroxycoumarin (A) and 3,7-dihydroxyflavone (B) as substrate.

## Appendix



**Appendix Figure 6.** The effect of different amounts of proteins (0.5-10 µg) on the product formation of protein FaGT24224 (gene2-8), FaGT24225a (gene3-4), FaGT24225b (gene3-5), FaGT24226 (gene4-3).

## Appendix

		10	20	30	40	50	
FaGT24225a	1	..... ..... ..... ..... ..... ..... ..... ..... ..... .....	MKQWVEIVFIPSPGIGHLVSTVEVAKLLLSRDDRLFITVLMKFPFSSDP	50			
FaGT24225b	1	..... ..... ..... ..... ..... ..... ..... ..... ..... .....	.....L.....L.....	50			
		60	70	80	90	100	
FaGT24225a	51	..... ..... ..... ..... ..... ..... ..... ..... ..... .....	IDAYIESFADSSISHRIKFINLPQQNIETQGNSTINFLDFSGSQKTNVKD	10			
FaGT24225b	51	..... ..... ..... ..... ..... ..... ..... ..... ..... .....	.....N...T.....	10			
		110	120	130	140	150	
FaGT24225a	101	..... ..... ..... ..... ..... ..... ..... ..... ..... .....	VVAKLIESKTETRLAGFVIDMFCTSMIDVANELGVPTYVFFTSAAASLGV	15			
FaGT24225b	101	..... ..... ..... ..... ..... ..... ..... ..... ..... .....	.....	15			
		160	170	180	190	200	
FaGT24225a	151	..... ..... ..... ..... ..... ..... ..... ..... ..... .....	LLHLQALRDDQNKDCLEFNDSTADLVIPSYANPLPVRVLPGLFEKVGGN	20			
FaGT24225b	151	..... ..... ..... ..... ..... ..... ..... ..... ..... .....	.....Y.....A.....E...	20			
		210	220	230	240	250	
FaGT24225a	201	..... ..... ..... ..... ..... ..... ..... ..... ..... .....	GFLNLAKRFRDVKGILINTMTELESHALLSLSSDGKLPVYPVGPILNVK	25			
FaGT24225b	201	..... ..... ..... ..... ..... ..... ..... ..... ..... .....	.....I.....	25			
		260	270	280	290	300	
FaGT24225a	251	..... ..... ..... ..... ..... ..... ..... ..... ..... .....	SDDNNDQVDSKQSKQTSIDILKWLDDQPPSSVVFLCFGSMGFSSEDQVKEI	30			
FaGT24225b	251	..... ..... ..... ..... ..... ..... ..... ..... ..... .....	.....	30			
		310	320	330	340	350	
FaGT24225a	301	..... ..... ..... ..... ..... ..... ..... ..... ..... .....	ARALEQGGFRFLWSLRQPPPKGKNGVPSDYADHTGVLPEGFLDRTAGVGK	35			
FaGT24225b	301	..... ..... ..... ..... ..... ..... ..... ..... ..... .....	.....I.....Q.....	35			
		360	370	380	390	400	
FaGT24225a	351	..... ..... ..... ..... ..... ..... ..... ..... ..... .....	VIGWAPQVAILSHPAVGGFVSHCGWNSTLESLEWFGVPVATWPLYAEQQQN	40			
FaGT24225b	351	..... ..... ..... ..... ..... ..... ..... ..... ..... .....	.....	40			
		410	420	430	440	450	
FaGT24225a	401	..... ..... ..... ..... ..... ..... ..... ..... ..... .....	AFQLVRELGIAVEIDMSYRKDGPIVVTAEKIQGGIKELMELDSDIRKRVK	45			
FaGT24225b	401	..... ..... ..... ..... ..... ..... ..... ..... ..... .....	.....S.....D.....	45			
		460	470				
FaGT24225a	451	..... ..... ..... ..... ..... ..... ..... ..... ..... .....	QVSDNSKKALMDGGSSYASLGHFIDQI*	478			
FaGT24225b	451	..... ..... ..... ..... ..... ..... ..... ..... ..... .....	.....*	478			

**Appendix Figure 7.** Comparison of the deduced amino acid sequences of FaGT24225a and FaGT24225b enzymes from *F. x ananassa*

## Appendix

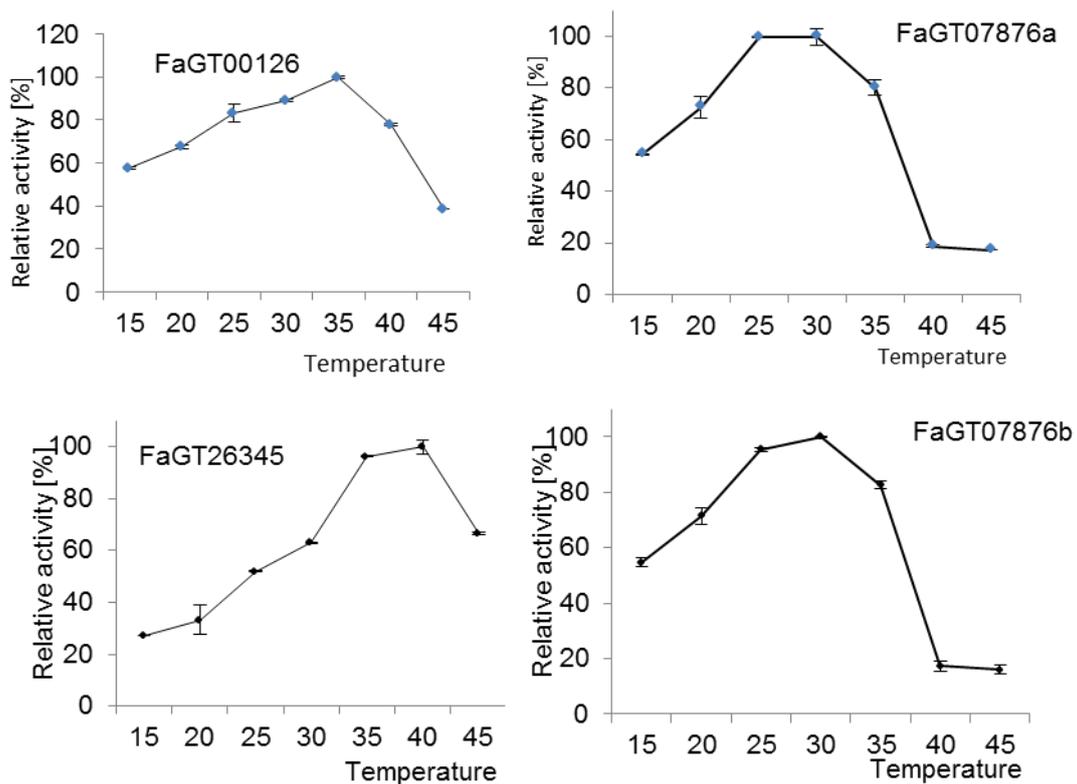
			10	20	30	40	50	
gene00126	1	..... ..... ..... ..... ..... ..... ..... .....						
gene26345	1	.....C.G.....T.....Y.....						50
			60	70	80	90	100	
gene00126	51	..... ..... ..... ..... ..... ..... ..... .....						
gene26345	51	.....D.....						10
			110	120	130	140	150	
gene00126	101	..... ..... ..... ..... ..... ..... ..... .....						
gene26345	101	.....F.....V.....S.....						15
			160	170	180	190	200	
gene00126	151	..... ..... ..... ..... ..... ..... ..... .....						
gene26345	151	.....L.....S.V..E.....						20
			210	220	230	240	250	
gene00126	201	..... ..... ..... ..... ..... ..... ..... .....						
gene26345	201	.....						25
			260	270	280	290	300	
gene00126	251	..... ..... ..... ..... ..... ..... ..... .....						
gene26345	251	.....L.....						30
			310	320	330	340	350	
gene00126	301	..... ..... ..... ..... ..... ..... ..... .....						
gene26345	301	.....V.....M.....						35
			360	370	380	390	400	
gene00126	351	..... ..... ..... ..... ..... ..... ..... .....						
gene26345	351	.....LA.....A.....						40
			410	420	430	440	450	
gene00126	401	..... ..... ..... ..... ..... ..... ..... .....						
gene26345	401	.....I.....						45
			460	470				
gene00126	451	..... ..... ..... ..... ..... ..... ..... .....						
gene26345	451	.....*.....						478

**Appendix Figure 8.** Comparison of the deduced amino acid sequences of FaGT00126 and FaGT26345 enzymes from *F. x ananassa*

			10	20	30	40	50	
FaGT71C3a	1	..... ..... ..... ..... ..... ..... ..... ..... ..... .....						
FaGT71C3a	1	MKKAELVFIPAPGAGHLVSALQFGKRLLRDDRISITVLAIKSAAPSSLG						50
FaGT71C3b	1	..... ..... ..... ..... ..... ..... ..... ..... ..... .....						50
			60	70	80	90	100	
FaGT71C3a	51	..... ..... ..... ..... ..... ..... ..... ..... ..... .....						
FaGT71C3a	51	SYTEALVASESRLQLIDVPQAEPLPEFAKSPAKFFILNIENHVPNVREA						100
FaGT71C3b	51	..... ..... ..... ..... ..... ..... ..... ..... ..... .....						100
			110	120	130	140	150	
FaGT71C3a	101	..... ..... ..... ..... ..... ..... ..... ..... ..... .....						
FaGT71C3a	101	LTNYVSSKQDSVPIVGVVLDFFCVSMIDVVNEFNLP SYLFMTSNAGYLSF						150
FaGT71C3b	101	..... ..... ..... ..... ..... ..... ..... ..... ..... .....						150
			160	170	180	190	200	
FaGT71C3a	151	..... ..... ..... ..... ..... ..... ..... ..... ..... .....						
FaGT71C3a	151	KFHFP AQDSRTGRPPKDSDPDWLVP GIVPPVPTKVLVSLTDGSYYNYLG						200
FaGT71C3b	151	..... ..... ..... ..... ..... ..... ..... ..... ..... .....						200
			210	220	230	240	250	
FaGT71C3a	201	..... ..... ..... ..... ..... ..... ..... ..... ..... .....						
FaGT71C3a	201	VASRFREAKGIIANTCVELETHAFNSFAEDQTT PPVYPVGPVLDLNDGQA						250
FaGT71C3b	201	..... ..... ..... ..... ..... ..... ..... ..... ..... .....						250
			260	270	280	290	300	
FaGT71C3a	251	..... ..... ..... ..... ..... ..... ..... ..... ..... .....						
FaGT71C3a	251	RSNLNQAQRDKIISWLDDQPEESVVF LCFGSMGSFTEAQVKEIALGLEQS						300
FaGT71C3b	251	..... ..... ..... ..... ..... ..... ..... ..... ..... .....						300
			310	320	330	340	350	
FaGT71C3a	301	..... ..... ..... ..... ..... ..... ..... ..... ..... .....						
FaGT71C3a	301	GQRFLWSLRLTPPKGSKSLSPVDCSNLEEVLP DGFLERTREKGLICGWAP						350
FaGT71C3b	301	..... ..... ..... ..... ..... ..... ..... ..... ..... .....						350
			360	370	380	390	400	
FaGT71C3a	351	..... ..... ..... ..... ..... ..... ..... ..... ..... .....						
FaGT71C3a	351	QVDVLSHKATGGFVSHCGWNSILESLWHGVPIV TWPMYAEQQLNAFRLVK						400
FaGT71C3b	351	..... ..... ..... ..... ..... ..... ..... ..... ..... .....						400
			410	420	430	440	450	
FaGT71C3a	401	..... ..... ..... ..... ..... ..... ..... ..... ..... .....						
FaGT71C3a	401	EMGLGLEMLD YKRGDEVVKADEIGKAVASVMENSEVRKKVKEIGVVCR						450
FaGT71C3b	401	..... ..... ..... ..... ..... ..... ..... ..... ..... .....						450
			460	470				
FaGT71C3a	451	..... ..... ..... ..... ..... ..... ..... ..... ..... .....						
FaGT71C3a	451	KAVEDGGSSSVSLGRFIEDV MRNHFGSE*						479
FaGT71C3b	451	..... ..... ..... ..... ..... ..... ..... ..... ..... .....						* 479

**Appendix Figure 9.** Comparison of the deduced amino acid sequences of FaGT71C3a (gene07876a) and FaGT71C3b (gene07876b) enzymes from *F. x ananassa*

## Appendix



**Appendix Figure 10.** The effect of different temperatures on the product formation of protein FaGT00126 (gene00126), FaGT26345 (gene26345), FaGT07876a (gene07876a; renamed as FaGT71C3a), FaGT07876b (gene07876b; renamed as FaGT71C3b).

## Appendix

Amino acid sequence of FaGTs from *F. x ananassa* which have been functionally characterized in this study

### **FaGT00126** (encoded by gene 00126; Gene 8-1)

MGSECHGSVHIFLPPFMAYGHMIPVSDMAKLFASHGVKITIVTTPLNAIRFAQTTQSSKFNIQIKAIEFPSE  
EAGLPKGCENVDTLPSPNLVNPFFKATRLLQPQFEELLKEFKPTCIVADMFFPWATEAAAKFGIPRLVFHG  
TSFFAMCASDCVKVYEPYNKVSSDTEPFVPHPLPGKIELTRAQVPDFIKNNVLNDVTQLLKEAREAEKLSFG  
IIMNSFYELEPVYADFYRNELGRRAWHIGPVSLCNRETEEKVQRGKEATIDEHECLKWLDSKKPDSVVYV  
CFGSVADFNSTQLKEIAMALEAAGQDFIWVVRKKGKDEWDEWLPEGFEERMEGKGLIIRGWAPQVLILD  
HPSVGGFVTHCGWNSTLEGISAGLPMVTWPLAAEQFYNEKLVAQVLKIGVGVGTQKVVRLFGDSVKK  
EAIKAVSQIMVGEEAEERRSRARELGKQARRAVEEGSSYQDFNKLIQELKSQQLG\*

### **FaGT26345** (encoded by gene 26345; Gene z3-5)

MGSESHDSVHIFLPPFMAHGHMIPVSDMAKLFASHGVKITIVTTPLNAIRFSQTTQSSKFNIQIKAIEFPSE  
EAGLPKGCENVDTLPSPNLVNPFFKATRLLQPQFEELLKEFKPTCIVADMFFPWATEAAAKFGIPRLVFHG  
TSFFAMCASDCVKVYEPYNKVSSDTEPFVPHPLPGKIELTRAQVPDFIKNNVLNDVTQLLKEAREAEKLSFG  
IIMNSFYELEPVYADFYRNELGRKAWHIGPVSQCNRETEEKVQRGKEATIDEHECLKWLDSKKPDSVVYV  
CFGSVADFNSTQLKEIAMALEAAGQDFIWVVRKKGKDEMDEWLPEGFEERMEGKGLIIRGWAPQVLILD  
HPSVGGFVTHCGWNSTLEGISAGLPMVTWPVSAEQFYNEKLVTVLQKIGVGVGTQKVVRLFGDSVKKE  
AVKAVSQIMVGEEAEERRSRARELGKQARRAVEEGSSYQDFNKLIQELKSQQLG\*

### **FaGT07876a** (encoded by gene 07876; Gene 1-5; FaGT71C3a)

MKKAELVFIPAPGAGHLVSALQFGKRLLRDDRISITVLAIKSAAPSSLGSYTEALVASESRLQLIDVPQAE  
PPLEFAKSPAKFFILNIENHVPNVREALTNYVSSKQDSVPIVGVVLDFFCVSMIDVVNEFNLPYLFMTSNA  
GYLSFKFHFAQDSRTGRPPKDSDPDWLVPVGPVPPVPTKVLVPSLTDGSSYNYLGVASRFREAKGIIANTC  
VELETHAFNSFAEDQTTTPVYPVGPVLDLNDGQARSNLNQAQRDKIISWLDDQPEESVVFLCFGSMGSF  
TEAQVKEIALGLEQSGQRFLWSLRLTPPKGSKSLSPVDCSNLEEVLPDGFLETRKGLICGWAPQVDVLS  
HKATGGFVSHCGWNSILESLWHGVPIVTWPMYAEQQLNAFRLVKEMGLGLEMRDLYKRGGDEVVKA  
DEIGKAVASVMENSEVRKKVKEIGVVCRAVEDGGSSSVSLGRFIEDVMRNHFGSE\*

### **FaGT07876b** (encoded by gene 07876; Gene 1-13; FaGT71C3b)

MKKAELVFIPAPGAGHLVSALQFGKRLLRDDRISITVLAIKSAAPSSLGSYTEALVASESRLQLIDVPQAE  
PPLEFAKSPAKFFILNIENHVPNVREALTNYVSSKQDSVPIVGVVLDFFCVSMIDVVNEFNLPYLFMTSNA  
GYLSFKFHFAQDSRTGRPPKDSDPDWLVPVGPVPPVPTKVLVPSLTDGSSYNYLGVASRFREAKGIIANTC  
VELETYAFNSFAEDQTTTPVYPVGPVLDLNDGQARSNLNQAQRDKIISWLDDQPEESVVFLCFGSMGSF  
TEAQVKEIALGLEQSGQRFLWSLRLTPPKGSKSLSPVECSNLEEVLPDGFLETRKGLICGWAPQVDVLS  
HKATGGFVSHCGWNSILESLWHGVPIVTWPMYAEQQLNAFRLVKEMGLGLEMRDLYKRGGDEVVKA  
DEIGKAVASVMENSEVRKKVKEIGVVCRAVEDGGSSSVSLGRFIEDVMRNHFGSE\*

### **FaGT22709** (encoded by gene 22709; Gene 7-1)

MKRAELVFIPTSTGHLVSTIEFSKRLDRCDQFSVTILLMKSPFGVAADQSLPAASNTNIKLIHLPNINPPIK  
LDSVEKFLSDYIETYKHHVKDTILNQVLPNSSRIAGVVIDMFCTTMIDIANELKVPSFLFFTSGAAFLGLLCL  
PERYDLVGKEFVHSDPDSIVPSYVNPVPTNVLPGFVFNNGGYVSFASHARRFKETKGVIIINTLVELESHAV  
HSIFRVGEGDQSDQPWPAVYPVGPLIDTKGEHQVRSRDRIMEFLDNQPPKSVVFLCFGSGFSFDEAQL  
REIAIGLEKSGHRFLWSVRQRPKPKTEFPGEYKNYEDFLPQGFLERTKGVGMLCGWAPQVEVLGHKST

GGFVSHCGWNSILES LWYGVPIVTWPLYAEQQVNAFLIARDLGLGVELRLDYVYVYSGDFVSADEIERAVT  
GLMVG DSEIRKRVVEMSEMCRRAVDDGSSSTSLGSLIKVLSQNLENN\*

**FaGT24226** (encoded by gene 24226; Gene 4-3)

MKQSAELVFIPCPGIGHLVSTVEVAKLLLSRDDRLFITVLMKFPFSDPTDAYIESFADSSISQRIKFINLPQQ  
NIETQGNRTINFFNFIDSQQTNVKD VVVKLIESKTETQLAGFVIDMFCTSMIDVANELGVPTYAFFTSTAA  
MLGLMFHLQALRDDH NKHCIEFKDSATDLVVP SYSHPLPAARVLP SVLLDKEASNR FVN LAKRLRDV KGI  
VINTFTELESHAFSLSSDGELPPVYPVGPILNVKSDDNNDQVNSKKKSDILNWLDDQPPSSVFLCFGSM  
GSFSEDQVKEIACALEQGGFRFLWSLRQPPPKGKNGVPSDYADHTEVLPEGFLDRTAGLGK VIGWAPQV  
AILAHPAVGGFVSHCGWNSTLES LWFGVPVATWPLYAEQQ LNAFQLVKELGIAVEIDMSYRKDGPVVV  
TAEKIERGIKELMELSDIRKRV RQVSDNSKKALMDGGSSYASLGHFIDQI\*

**FaGT24225a** (encoded by gene 24225; Gene 3-4)

MKQWVEIVFIPSPGIGHLVSTVEVAKLLLSRDDRLFITVLMKFPFSSDPIDAYIESFADSSISHRIKFINLPQQ  
NIETQGNSTINFLDFSGSQKTNVKD VVAKLIESKTETRLAGFVIDMFCTSMIDVANELGVPTYVFFTSSAAS  
LGVLLHLQALRDDQNKDCLEFNDSTADLVIPSYANPLPVRVLP GILFEKVGGNGFLN LAKRFRDVKGILINT  
MTELESHALLS SDGKLPVYPVGPILNVKSDDNNDQVDSKQSKQTS DILKWLDDQPPSSVFLCFGSM  
GSFSEDQVKEIARALEQGGFRFLWSLRQPPPKGKNGVPSDYADHTGVLPEGFLDRTAGVGVK VIGWAPQ  
VAILSHPAVGGFVSHCGWNSTLES LWFGVPVATWPLYAEQQ QNAFQLVRELGI AVEIDMSYRKDGPV  
VTAEKIQQGIKELMELSDIRKRVKQVSDNSKKALMDGGSSYASLGHFIDQI\*

**FaGT24225b** (encoded by gene 24225; Gene s3-5)

MKQWVEIVFIPSPGIGHLVSTVELAKLLLSRDDRLITVLMKFPFSSDPIDAYIESFADSSISHRIKFINLPQQ  
NIETQGNSTINFLNFGTQKTNVKD VVAKLIESKTETRLAGFVIDMFCTSMIDVANELGVPTYVFFTSSAAS  
LGVLLHLQALRDDQNKDYLEFNDSTADLVIPSYANPLPARVLP GILFEKEGGNGFLN LAKRIRDVKGILINT  
MTELESHALLS SDGKLPVYPVGPILNVKSDDNNDQVDSKQSKQTS DILKWLDDQPPSSVFLCFGSM  
GSFSEDQVKEIARALEQGGFRFLWSLRQPPPKGKIGVPSDYADHTGVLPEGFLDQTAGVGVK VIGWAPQV  
AILSHPAVGGFVSHCGWNSTLES LWFGVPVATWPLYAEQQ QNAFQLVRELGI AVEIDMSYRKDGPVVT  
AEKIQSGIKELMDLSDIRKRVKQVSDNSKKALMDGGSSYASLGHFIDQI\*

**FaGT24224** (encoded by gene 24224; Gene 2-8; FaGT3)

MEKPAELVLIPSPGIGHLVSTLEIAKLLVSRDDQLFITVLMHFPVAVSKGTESYVQSLADSSSPISQRIKFINLP  
HTNMDHTEGSVRNSLVGFVESQQPHVKDAVTKLRDSKTTRLAGFVDMFCTTMIDVANQFGVPSYVFF  
TSSAATLGLMFHLQEMRDQYNKDCTEFKDSAEIIPSFNPLPAKVLPGRILVKDSAESFLNVIKRYRDTK  
GILVNTFTDLESHALHALSSDAEIPVYPVGPLLNLSNESSVDSDEVKKN DILKWLDDQPPLSVFLCFG  
SMGSFDENQVREIANALEHAGHRFLWSLRPPPTGKVAFPDYYDHSGVLP EGFDRTVGIGK VIGWAP  
QVAVLAHPSVGGFVSHCGWNSTLES LWHGVPVATWPLYAEQQ LNAFQLVRELELGVEIDMSYRSESPV  
LVSAKEIERGIREVMELSDSDIRKRVKEMSEKGGKALMDGSSSYTSLGHFIDQI\*

**FaGT26342** (encoded by gene 26342; Gene 5-8; abscisate beta-glucosyltransferase-like)

MDSEPPVEMYFFPFVGGGHQIPMIDTARVFAAHGAKSTILSTLSNALRFCNSIHRDQTHNRLISIHVLDL  
PNDVPPDTSMSAAPFTDTSVFKQLRHFLTQHPPDCIVIDVFHRWASVIDSLGIRRVFNNGFFSRC  
VMQNVGKFAPQEKVGS DSEPFVVPGLPDRVELTKSQLPVFARNKSGPDKFGQLEDKSFV VVNSFYELE  
SKYVDYFKKDLGKKA WIGIPVSLCNRDEADKVERGQAASVDEEKLK WCLDWLDSQEPDSV VYISFGSLA  
RLSYKQLIEIAHG VVNSTNCFVWVVGK VSENDGQSH EDEENWLLDFEKRMR ESERGVVIRGWAPQILM  
LEHKAVGGFVSHCGWNSTLES VCAGVPMITWPLSAEQFSNEKLITDVLGIGVQVGSKEWTSWNMERK  
EVIGREKVEAAVRKV VGGGDEAVEMRKRARDLAEKAKKAVEEGSSYAEVDALISELRS LKKN\*

**FaGT26479** (encoded by gene 26479; Gene 10-9)

MESRNHHFLVISCAGQGHLNPSLQLAKRLIDIGSSHVTFVTNIHGLTQIKSLPSLEGLSFASFSDGFDDGVH  
PANDPEHIMSELKRAGESLAALIEKISTSDERGPITFLIYTILLPWAAEVASSFGIPSAFLCITSATSFAICGHY  
FKDYYSQSSLPFPSCITIDGLPPFASDELPSYLLPTSPHVSILPTLQEHYQILEKNPNSCVLLNTFDGLEAAAI  
RDMRVHMNLITVGPLIKSAEVCCDLFDKSGDDYLQWLDSRADCTVVYVSFGSMVVLSSGQMEEILHGL  
VDSGLPVLWVIRKSGNEGDQETENLINNTLKEQGLIVPWCSQVEVLSHKSVGCCVLHCGWNSTVESLA  
AGVPVVGCPHFADQTTNAKLVEGLWGTGVRARANEEGVIEREEIKRCLEVVMGDGVRGEEMRRNAQK  
WKSLAMKAVNESGSSDDNLRNFVRSLA\*

**FaGT26353** (encoded by gene 26353; Gene 6-6)

MEIKTHQQLHIFFLPFMGQGHTLPLIDIAKLFASRGEKSSIITTPANAPLFTKAIQTSRSSGLEIEILLIKFPSSE  
VGLPEGIESSNDWGKTAIEIAEKFFKALTLLKHQVEQLLHQYHPHCLVASSLFHWSTDLAAKFGIPRLIFQG  
PGFFSLCAAMSVTLYQPHMKVASDSESFIVPNLPHEIKMTRNELPSFLKQEGETELMKLLRECRETEKSSY  
GIIINSFYELEPDYADHYRMAFGRRSWHIGPVSLCNTAENDKLARGREGSVDEVHECLQWLNSKKPSSVV  
YVCFGSLNTFSDSVLLEIALGLEASQQQFIWVVKENNNQEEWLPKGFQRMKGKGLIIRGWAPQLLILQ  
HEAVGAFALTHCGWNSILEGVTAGVPMITWPLFADQFYNEKLVLTQILGIGVSVGSQKXSEDGGVKSEARVK  
WEAIKAVTEIMEGDKAAEIRSKAAALGEIARSAVEEGSSYSDLTALIGELRSFGS\*