Identification and biochemical characterization of the UDP-glycosyltransferases UGT73C10 and UGT73C11 in biosynthesis of saponins in *Barbarea vulgaris* ssp *arcuata*.
Preface

This master thesis project was performed at the faculty of LIFE, department of Plant Biology and Biotechnology at the University of Copenhagen. It is a part of my MSc Biology and counts for 39 ECTS, in accordance with circa a half year study.

Above all, I would like to thank my supervisors Søren Bak, Jörg M. Augustin and Peter de Jong for their guidance and help during my thesis. I would like to thank Jörg for being so patient and helpful with me and for the interesting discussions. I’m proud to have gained even a part of his knowledge, which will be valuable for me in any field within Biology. I’m also very grateful to Søren Bak for guiding me through the writing process and believing in me even when I did not. Moreover, I would like to thank Peter de Jong for his support on a longer distance, who always replied with some inspiring and friendly words.

For the experimental part of my thesis I would like to thank Carl Erik Olsen for the LCMS and NMR analyses and Esben Hansen from Evolva his help with the large scale production of monoglucosides.

At last, I would like to thank my colleagues at the department for being so friendly and social from the very start. Many thanks my office mates, who have been a great help and inspiration during this thesis. It was a pleasure to be part of this for half a year and I hope we will meet again soon.

_____________________

Sylvia Drok

Copenhagen, June 2010
## Table of contents

**Preface**

**Summary**

1. **Introduction**
   1.1 Secondary metabolites in insect resistance.  
   1.2 The interaction between *Barbarea vulgaris* ssp *arcuata* and *Phyllostreta nemorum*  
   1.3 Saponins-natural soap compounds  
   1.4 Saponin modes of resistance  
   1.5 The biosynthetic pathway of hederagenin and oleanolic acid  
   1.6 Glycosyltransferases  
   1.7 Thesis project

2. **Materials and methods**
   2.1 Cloning the sequences  
   2.2 Heterologous expression of the UGTs  
   2.3 Characterization

3. **Results**
   3.1 Phylogenetic analysis of the UGTcDNA and amino acid sequences  
   3.2 UGT sequence analysis  
   3.3 Cloning and expression of the UGTs  
   3.4 Characterization of the UGTs  
   3.5 UGT substrate specificity  
   3.6 Kinetics

4. **Conclusions and discussion**

5. **Perspectives**

6. **References**

7. **Appendix I**
Abstract

Plants are sessile organisms and in order to defend themselves against pathogens and pest they have developed a wide variety of chemical and mechanical ways of defense. The chemical defense of *Barbarea vulgaris* ssp *arcuata* against the flea beetle *Phyllotreta nemorum* was correlated with the presence of four saponins, among which hederagenin and oleanolic acid celllobioside (Kuzina et al., 2009). Both saponins were present in the resistant G-type *Barbarea* but absent in the susceptible P-type plants. The proposed biosynthetic pathway of the two saponins shares its first steps with the phytosterol pathway, but branches with an alternative set of cyclization steps of 2,3 oxisqualene into β-amyrin, from which both oleanolic acid and hederagenin are derived. Glycosylation of these aglycones is performed by family 1 UDP- glycosyltransferases (UGTs). A UGT was identified that could transfer the first glucose moiety to hederagenin and oleanolic acid aglyones, as in the first glucylation step in the biosynthesis of *Barbarea vulgaris var. variegata*. A search for orthologs in *B.v.arcuata* revealed five close homologous sequences, two derived from the G-type and three from the P-type plant. In this master thesis project these five sequences were cloned, expressed and characterized. UGT73C10, UGT73C11, UGT73C12 and UGT73C13 utilized oleanolic acid and hederagenin as acceptor substrates, and NMR analysis revealed that a hederagenin monoglucoside was glycosylated at the C3 position. UGT73C9 was only active on 2,4,6-trichlorophenol (TCP). UGT73C10 and UGT73C11 were highly specific towards sapogenins, in vitro, suggesting that they could be involved in the biosynthesis pathway of oleanolic acid and hederagenin celllobioside. The difference in the production of saponins in the G and P plant is most likely not due to biochemical differences of UGTs involved in the first glycosylation step.
1. Introduction

Plants are autotrophic organisms, meaning they are able to produce their own organic compounds from carbon dioxide and water, using light as an energy source. Therefore they are at the bottom of the food chain and serve as a putative food source for a broad range of herbivores, pests and pathogens. Because plants are sessile organisms, they cannot escape their environment and have to cope with the accompanying biotic and abiotic stress. Plants have developed several ways to counteract these threats, e.g. by trying to avoid their enemies or by defending themselves using mechanical and chemical defenses. The chemical defense compounds are typically secondary metabolites: compounds which are not directly involved in growth, development or reproduction of the plant but have important roles in the interaction between the plant and its environment (Albers et al., 1996). Secondary metabolites are usually derived from the basic metabolic pathway; and are considered to originate from random mutations in that pathway which had accidental beneficial effects.

1.1 Secondary metabolites in plant defense

There are three major classes within plant secondary metabolites: terpenes (or terpenoids), phenolic compounds and nitrogen containing compounds. Terpenoids are derived from isoprenoids building blocks which contain 5 carbons. Nomenclature of the terpenoids is based on the amount of 5-carbon isoprenoids, namely monoterpenes (2 isoprenoids, 10 carbons), sesquiterpenes (15 carbons), diterpenes (20 carbons), triterpenes (30 carbons), etc. Terpenes are often toxins and feeding deterrents that function as a defense mechanism widespread in the plant kingdom. Monoterpenes accumulate in the resin in conifers. Mono and sesquiterpenes are the essential oils present in many plant species and well-known to have insect repellent properties. Within the triterpenes there is a compound, azadirachtin, which is perhaps the most powerful insect deterrent known as it works at a dose of 50 parts per billion in some cases (Aerts, 1997).

Phenolic compounds are derived from aromatic amino acids, often phenylalanine. They are a big class within the secondary metabolites with a wide variety of biological functions. Anthocyanins are thought to play a role in the attraction of pollinators while other phenolic compounds may play a role in UV protection of the plant. Some compounds are suggested to
Introduction

act as allelopathic compounds thus influencing the neighboring plants e.g. when released in
the soil. Last, a group of phenolics, isoflavonoids and tannins are known to have
antimicrobial and feeding deterrent activities on herbivores.

The third major class of secondary metabolite is the nitrogen containing compounds, which
function as chemical defense as well. Nitrogen containing compounds like alkaloids are often
found to be toxic themselves. Other nitrogen containing compounds e.g. cyanogenic
glucosides and glucosinolates are glycosides of otherwise toxic compounds. Glucosinolates
are an important group within the chemical defense of crucifers. They are wide spread over
the brassicales, and are known to act as antibiotics, fungal growth inhibitor and toxins to
nematodes as well as a wide range of generalist herbivores (Renwick, 2002). Therefore
glucosinolates are considered to be the first line of defense in crucifers. However, crucifer
specialist insects have adapted to these glucosinolates and are able to detoxify or sequester
them, or even use them as a trigger for oviposition on the host plant (reviewed by Renwick,
2002). As a counter adaptation some crucifers have developed a second line of defense
which includes compounds acting as repellents or are toxic to their specific attacker. Such
second line of defense can be found in the form of saponins in Barbarea, a genus of the
crucifer family. Saponins are a wide ranging class of compounds but within the crucifer
family they only occur in the Barbarea genus. In Barbarea vulgaris two saponins are found to
be toxic to the diamond back moth which is a common pest on rape seed (Agerbirk et al.,
2003a). Recently its toxicity to the flee beetle Phyllotreta nemorum has also been shown
(Kuzina et al., 2009).

1.2 The interaction between Barbarea vulgaris ssp arcuata and Phyllotreta nemorum

The interaction between both winter cress and Barbarea vulgaris ssp arcuata L. and the flee
beetle Phyllotretata nemorum is a unique model system to study plant insect interactions.
The plant, B. vulgaris ssp arcuata (from now on referred to as Barbarea) is polymorphic with
respect to insect resistance. It has two genotypes: a resistant G-type (with glabrous leaves)
and a pubescent P-type which is susceptible to the flee beetles. In experiments in vitro no
larvae could survive on G-type whereas 91% of the larvae survived on the P-type (Kuzina et
al., 2009). It was observed that the flee beetle larvae often started feeding on the G- type
but stopped feeding after a few bites and starved to death. Whether their death is caused by
the lack of food intake or the toxicity of the G-type leaves themselves remains unclear. However, in some populations in Denmark flee beetles were collected which were able to survive on the G type of *Barbarea*. The *R* gene that confers this virulence (resistance towards the plant) is dominant as both *RR* and *Rr* genotypes confer resistance to G-type *Barbarea* while the *rr* genotype does not (Nielsen, 1997). Thus, *Barbarea* plants are polymorphic in insect resistance and the beetles are polymorphic with respect to resistance toward *Barbarea*, making it a unique model system for the understanding of plant-insect interactions and speciation.

*Barbarea vulgaris* grows naturally Europe and is naturalized in North-America. Several ecotypes exist of which two, ssp *vulgaris* and ssp *arcauta*, occur in Denmark. Ssp *arcauta* is more common than ssp *vulgaris*. The P and G genotypes within *B. vulgaris* ssp *arcauta* are morphologically, cytologically and genetically different and deviate by glucosinolate profile, leaf pubescence and flee beetle resistance (Kuzina et al., 2009; Agerbirk et al., 2003b). In this thesis also a third subspecies is mentioned, *B. vulgaris* var. *variegata*, which most likely is a cultivated form of *Babarea vulgaris* and is commercially available, used as decoration in gardens and eventually in salads. It has partially white leaves, which are glabrous and confer resistance towards the Diamont Back moth, *Plutella xylostella* (Shinoda et al., 2002).

Using untargeted metabolite profiling and clustering analyses, Kuzina et al. (2009) identified four saponins to be correlated with resistance against flee beetles. Two of those saponins were novel compounds. The other two were oleanolic acid cellobioside and hederagenin cellobioside which both are known, in *Barbarea*, to confer resistance to the diamond back moth *Plutella xylostella* L. (Agerbirk et al., 2003a). Both saponins occur in the resistant G-type but are absent, or only present in very low concentration in the P-type plants. Hederagenin cellobioside is more abundant than the oleanolic acid cellobioside (Kuzina et al., 2009) and shows a strong negative correlation with flee beetle survival (Nielsen et al., 2010). This correlation is weaker and sometimes not significant at all for oleanolic cellobioside (Nielsen et al., 2010). Therefore hederagenin cellobioside is considered to be the most important saponin in resistance towards flee beetles. The glucosinolate content of the P and G type plant show no correlation with resistance (Agerbirk et al., 2003b).
Introduction

1.3 Saponins – natural soap compounds

Saponins are the glycosides of an isoprenoidal originated aglycone (also ‘sapogenin’). Saponins can have a steroidal or triterpenoid backbone, depending on the folding of the 30C backbone in the biosynthesis. Saponins occur all over the plant kingdom but can also be found in ancient animals like seacucumbers (holodurentia) and starfish (asteroidea). The name saponin is derived from their soap-like abilities (sapo = Latin for soap). Various saponins are commercially used for a variety of purposes including drugs, medicines, precursors for hormone synthesis, adjuvants, sweeteners, taste modifiers and cosmetics (Osbourn, 1996). Furthermore anti-inflammatory, anti-molluscial, anti-microbial and anti-fungal properties have been described for these compounds (Sparg et al. 1987).

The sugar moieties of triterpenoid aglycones commonly contain glucose, arabinose, xylose, rhamnose or glycoronic acid (Vincken et al., 2007) and these sugar moieties can be attached to O (OH or COOH), S and N atoms. The amount, origin and pattern (e.g. branching) of the sugars leads to enormous variation. An aglycone with one or more sugars attached to one position (usually the C3 position of triterpenoid aglycones) is referred to as a monodesmoside. Bidesmosides saponins have their sugars attached to two different positions on the aglycone, typically at the C26 (steroidal) or C28 (triterpenoid) position. Didesmosidic triterpenoid saponins often lack the bioactivity of the corresponding monodesmosides, but can be converted back into their monodesmoside by removal of the C26 or C28 sugar chain (Osbourn, 1996). Glycosylation (also for monoglycosides) is important for storage but it also allows the plant to accumulate potentially toxic compounds in high concentration so when needed they can be released in such concentrations (Jones and Vogt, 2001).

Various other roles of glycosylation have been described, e.g. as signaling molecules for plant-microbe interactions, plant-to-plant signaling and transportation (Jones and Vogt, 2001). Glycosylation often results in a loss of bioactivity and can be used for storage and detoxification of a compound. In contrast, in the interaction of Barbarea vulgaris and Phyllotreta nemorum the glycosylation of the aglycone causes the bioactivity of the toxins. The diglucosides of hederagenin and oleanolic acid have shown to cause plant resistance to
the flee beetle, while the aglycone itself does not show any effect on flee-beetle survival (Nielsen et al., 2010).

1.4 Saponin modes of resistance
The mechanism behind saponin resistance has been mostly studied for fungi. This resistance is believed to be due to a complex formation between saponins and sterols in the cell membrane of the fungus which leads to a loss of membrane integrity. The precise mechanism of complex formation and membrane leakage is not fully understood. Electron microscope analyses suggest a formation of membrane pores, while other studies suggest that the complexes interfere with membrane integrity by extracting the sterols from the cell membrane (figure 1). (Morrissey and Osbourn, 1999) 

![Figure 1: two models of the interaction of saponins with membrane sterols, using pore formation (A) or sterols extraction (B), leading to cell leakage (Morrissey and Osbourn, 1999).](image)

The sugar chain attached to the C-3 plays a crucial role in the complex formation as it mediates the aggregation of the sterols in the membrane. The sugar is therefore important and when removed it leads to loss of anti-fungal activity (Armah et al., 1999). The complex formation with sterols is a rather general mechanism, and therefore potentially toxic for every organism with sterols in their membranes. To protect itself from its toxic compounds the plant sequesters the saponins in the vacuole; the vacuole membrane may therefore have a low sterol content or different types of sterols that are less suitable for complex formation (Osbourn, 2003).
Introduction

Comparable strategies are described in which a pathogen can counter this mechanism of saponin resistance. Resistance of oomycetes towards saponins has been associated with a lack of sterols in the oomycete cell membranes. Also experiments with sterol-deficient mutants of *Neospora crassia* and *Fusarium solanum* show an increase of resistance towards the steroidal alkaloid α-tomatine (Prakash et al., 1999; Defago and Kern, 1983). A second way of resistance to saponins may be by degradation of saponins from the host plant, which is found in a number of fungi. Degradation of saponins is usually performed by hydrolysis of sugar molecules, often the ones at the C-3 position of the saponin (Osbourn, 1996). Although not much is known about the saponin resistance to insects, the resistance in flee beetle towards the saponins of *B. vulgaris* could be based on the same principle, as the *R* genes code for a beta-glucosidase (Ilkink, 2008).

1.5 The biosynthetic pathway of hederagenin cellobioside and oleanolic cellobioside.

Triterpenoid saponins are derived from metabolites produced during the anabolism of phytosterols, therefore they share the first steps of the phytosterol pathway. This includes the linkage (head-tail) of two famesildisphosphate (FPP) molecules (15 carbon) into squalene by a squalene synthase (figure 2) and the subsequent oxidization of squalene into 2,3, oxisqualene (Yendo, 2010).

Figure 2: A) synthesis of squalene from condensaion of FFP B) cyclisation of 2,3, xisqualene into β-amyrin by a β-amyrin synthase. (Yendo, 2010)
Further cyclization steps of 2,3-Oxisqualene by oxysqualene synthases (OCSs) form a branch point between the biosynthetic pathway of sterols and triterpenoid saponins. OSCs utilize 2,3-oxisqualene as substrate and depending on the type of OSC, different compounds will be released. In the sterol pathway 2,3, oxisqualene is converted into lanosterol (in fungi or animals) or cycloartenol (in plants) (Vincken et al., 2007; Haralampidis, 2002; Phillips et al., 2006; (Yendo, 2010). In the triterpenoid saponin pathway an OSC converts 2,3, oxisqualene into β-amyrin through diverse intermediates (figure 2) (Yendo, 2010). An OSC can be specific for one end product or multifunctional and release several end products. It is not known if the sterol and triterpenoid pathway are connected through one multifunctional OSC or if each pathway is facilitated by one OSC. From β-amyrin all kinds of aglycone backbones can be derived through oxidation and substitution. Examples are shown in figure 3 (Yendo, 2010).

![Figure 3: β-amyrin is a precursor for many sapogenin backbones. Glycosylation of hederagenin by UGT73K1 in Medicago truncatula is shown in G. (Yendo, 2010)](image)

The enzymes involved in cyclization of of 2,3 oxisqualene are largely uncharacterized. The first OCS, that could cyclize and release cycloartenol, was found in Arabidopsis by screening of extracts in a yeast lanosterol deficient mutant (Corey et al., 1993). Other OSCs have been identified based on sequence homology. β-amyrin producing OSCs have been characterized
Introduction

for trees (Betula platyphylla), monocots (avena strigosa), Euphorbia tirucalli and several legumes like Panax ginseng, Medicago truncatula and Pisum sativa (Phillips et al., 2006). Four multifunctional OSCs have been found in Arabidopsis thaliana and could, among other compounds, also release β-amyrin (Phillips et al., 2006). No β-amyrin releasing OSCs have been found in Barbarea yet.

β-amyrin is a precursor for a wide variety of saponin aglycones (figure 3). Structural changes that confer these different backbones are catalyzed by a class of enzymes named chytochrome p450 mono-oxidases. The p450s involved in the pathway of oleanolic acid and hederagenin, have not been characterized yet. Intermediates in the proposed pathway of hederagenin and oleanolic acid in Barbarea vulgaris (figure 4), are produced by one or possibly more p450s. The last step, the oxidation of the C23 position, will most likely be performed by a different p450, as this modification is done on a different Carbon atom than the first set of modifications.

In the final step of the biosynthetic pathway hederagenin and oleanolic acid are glycosylated. There is accumulating evidence that the glycosylation occurs by one sugar moiety at a time, but it cannot be excluded that a chain of sugars can be transferred in one step (Haralampidis, 2002).

Analysis of metabolite extracts of the G-type plant did not reveal the presence of β-amyrin or any of the intermediates. This may be explained by a high turnover rate of the intermediates, which causes that the reaction intermediates are not freely available in the cytosol. P450s are membrane bound enzymes, while UGTs are generally cytosolic. In many metabolic pathways, e.g., alkaloid, phenylpropanoids, cyanogenic glycoside and the first part of the isoprenoid pathway enzymes have been found to form a multi-enzyme complex (metabolon) and form a channel for the pathway. Channeling of a pathway results in concentration of the substrates, decreases the transit time for intermediates because the active sites of the enzymes are close together, and avoid metabolite cross-talk and avoid toxic intermediates to diffuse and cause problems elsewhere (Jorgensen et al., 2005; (Bak, 2006).
Figure 4: Proposed biosynthetic pathway of hederagenin and oleanolic acid cellobioside in *Barbarea vulgaris arcuata.*
Introduction

1.6 Glycosyltransferases

Enzymes that transfer a donor sugar molecule to an acceptor substrate are glycosyltransferases (GT). The hierarchical classification is divided in superfamilies, families and subfamilies. At present time 91 superfamilies have been identified (Campbell et al., 1997). The plant GTs belongs to superfamily 1, a class of enzymes, which use UDP-sugar as sugar donor. The classification into families and subfamilies are based on sequence homology (Mackenzie et al., 1997; Campbell et al., 1997). However, a high sequence identity does not automatically mean that the enzymes share same substrate specificity. Thus UGTs within one family can accept all kinds of different substrates.

Despite a low sequence identity, the secondary and tertiary structure is conserved in most of the UGTs, and some intrinsic structural features are conserved within a superfamily (Coutinho et al., 2003). The tertiary structure can have a GT-A and GT-B fold (figure 5). Within both folds an enzyme can be inverting (clan I and II) or retaining (clan II and IV). The
type of bold and the feature of being either an inverting or retaining enzyme are conserved within the GT superfamilies.

The term inverting and retaining refers to the putative change of conformation of the sugar moiety during glycosylation. A sugar moiety can be either in α or β conformation, depending on the stereochemical properties of the OH group on the C1 position of the sugar. If the OH group points in the opposite direction of the methyl group on the C6 position than the sugar is in α conformation, and if it points in the same direction (upwards) the sugar is in a β conformation. In an UDP-sugar, the sugar moiety is attached to the uridine-diphosphate in the α-conformation. When attaching the sugar to the aglycone backbone, the sugar conformation can either be retained (retaining) or inverted (inverting) into a β conformation (figure 6). The first sugar moiety of hederagenin cellobioside and oleanolic acid cellobioside has a β conformation so the UGT responsible for this step would be an inverting enzyme.

Figure 6: UDP-glucosyltransferases can be inverting or retaining. The OH group at the C1 position is indicated in red.

The GT-B fold separates two subunits placed at the C and N terminal end, respectively. The C domain forms mainly interactions with the donor substrate, whereas the N domain mainly interacts with the acceptor substrate (Wang, 2009; Hansen et al., 2009). The sequence identity within Arabidopsis UGTs is rather low, ca 10 percent (Vogt and Jones, 2000). However, alignment of these sequences revealed a 44 amino acids long motif containing very conserved amino acid sites. This motif was named the Plant Secondary Product Glycosyltransferase (PSPG) motif (figure 7) (Vogt and Jones, 2000; Hughes and Hughes, 1994).
Introduction

Crystal structures indicate that the sugar is located in a tunnel in the C domain of the UGT and interacts with amino acids in the PSPG motif (Hughes and Hughes 1999). Directed mutagenesis of amino acids show that one amino acid substitution can change the donor specificity e.g. from UDP-glucuronic acid to UDP-glucose (Noguchi et al., 2009).

![The PSPG motif. A red color indicates highly conserved (identity >80%) amino acids. Blue is moderately conserved (identity >50%) and black not conserved (>50%). (Vogt and Jones, 2000)](image)

The mechanism behind the acceptor site is even more complex and less understood. Most interaction between enzyme and acceptor substrate occur in the N-domain of the UGT, but no conserved motif has been identified. A lot of progress has been made to understand the mechanism for acceptor binding in UGTs using crystal structures, homology based modeling in combination with site directed mutagenesis and even swopping of N domains among different UGTs (Hansen et al., 2009). The state of knowledge about UGT modeling is recently reviewed in Osmani et al., 2009. Mutagenesis studies found that a single amino substitution can lead to a change of substrate specificity (Wang, 2009). Despite promising developments in this research field, the substrate specificity is, at present time, not predictable from the UGT amino acid sequence.

In the past decade a lot of progress has been made on the identification and biochemical characterization of UGTs. Lots of UGTs are found to glycosylate a wide variety of acceptor substrates, mostly to form monoglucosides or bidesmosedic diglycosides. Most of the work on characterization of triterpenoid saponins has been done on UGTs in legumes such as soybean (*Glycine max*), alfalfa (*Medicago sativa*), ginseng (*Panax ginseng*) and the model plant *Medicago truncatula*. These studies revealed several UGTs that transfer the sugar to the aglycone, yielding in a monoglucoside. Not much is known about glycosyltransferases that can catalyze the second step of adding a second sugar moiety to a monoglucoside to yied a monodesmosedic diglucoside. In soybean UGTs have been identified of which one can catalyse transfer of a UDP-galactose onto the sugar moiety of the soyasaponin (UGT73P2) and another UGT could transfer a UDP-rhamnose moiety to the latter soyasaponin diglycoside (UGT91H4) (Shibuya, 2010). They are the first examples of GTs which transfer the
second and third sugar to form monodesmosedic tri-glucosides of a triterpene saponin. However, these UGTs utilize UDP-rhamnose and UDP-galactose and no UGT is characterized that can transfer a glucose moiety to a glucose-monoglucoside as in the second step of the biosynthesis pathway of saponins in *Barbarea vulgaris*.

Dr Tetsuro Shinoda, a Japanese collaborator, identified a UDP-glycosyltransferase (UGT) from *B. vulgaris variegata* that was able to transfer the first glucose molecule to the hederagenin and oleanolic acid aglycone (Shinoda, unpublished). Orthologues of that gene in *Barbarea vulgaris arcuata* are the focus of this master thesis project.

### 1.7 This master thesis project

Jörg M. Augustin, a PhD student who is working on the biosynthetic pathway in *B. v. arcuata* and was the daily supervisor in this thesis, searched for close homologs of Shinoda’s UGT in the Danish *B. v. arcuata*. Five sequences were isolated of which two were from the G plant and three were from the P plant (table 1). The five sequences were named by the UGT Nomenclature Committee. The gene from *B.v. variegata* has not been named yet and is referred to as UGT73C<sub>shi</sub> as it is a gene from the UGT73C subfamily isolated by Shinoda. According to their nucleotide and amino acid sequence identity UGT73C9, UGT73C10 and UGT73C11 are referred to as orthologs and UGT73C12 and UGT73C13 are referred to as paralogs of UGT73C<sub>shi</sub>. To make the thesis better understandable, the genes derived from the P type will be marked in blue and the genes from the G type in red.

<table>
<thead>
<tr>
<th>Orthologs</th>
<th>P-type (non-resistant)</th>
<th>G-type (resistant)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orthologs</td>
<td>UGT73C9, UGT73C10</td>
<td>UGT73C11</td>
</tr>
<tr>
<td>Paralogs</td>
<td>UGT73C12</td>
<td>UGT73C13</td>
</tr>
</tbody>
</table>

The aim of this master thesis is to clone, express and characterize these five genes. The main purpose of the biochemical characterization is to elucidate the biosynthetic pathway of saponins in *Barbarea vulgaris*. In addition the possible role of these UGTs in plant resistance is examined. Small amino acid differences can have large effects on the enzyme function (Wang, 2009; (Hansen et al., 2009). Such differences in functions between the UGTs might
Introduction

therefore explain the different saponin content and resistance found in G and P type of *B. v. arcuata*.

Thus, the research questions for this thesis are as follows: i) Are the five UGTs capable of transferring the first sugar moiety to the oleanolic acid and hederagenin aglycones? ii) How specific are the UGTs for these aglycones? iii) Is there a difference in biochemical properties between the enzymes originating from P and G plants? iv) Is there a difference in biochemical properties between the orthologs and paralogs?

The starting point of this thesis will be the cDNAs encoding the five respective UGTs obtained by Jorg M. Augustin. These cDNAs will be cloned into *E.coli*, expressed and characterized to answer the above-mentioned questions. The characterization of the UGTs will be based on the activity assays. All enzymes will be tested for glycosylation of oleanolic acid and hederagenin. Initial activity assays will be analysed by Thin Layer Chromatography (TLC), a fast and cheap way to determine glycosylation activity of the UGTs. TLC will also be used to optimize the activity assay protocols when needed. Liquid Chromatography Mass Spectrometry (LCMS) will be used to determine the mass of the oleanolic acid and hederagenin glycosides. Nuclear Magnetic Resonance (NMR) spectroscopy can be used to determine the position and conformation (α or β) of the sugar moiety.

The specificity of the UGTs towards different acceptor substrates is studied. The substrates include the sapogenins, flavonols and sterols. The results will be analyzed by TLC and/or kinetics. Kinetic parameters offer a good way to compare the ‘specificity’ of a UGT. For enzymes following Michaelis–Menten kinetics the most important parameters are the \( V_{\text{max}} \), the maximum velocity and the \( K_M \), a constant that represents the affinity of an enzyme towards a substrate. A low \( K_M \) indicates a high affinity for the substrate, as it reaches its own top speed already at low substrate concentrations. The amount of substrate molecules per enzyme molecule per time is represented by the \( K_{\text{cat}} \).
2. Materials and methods

2.1 Cloning the five sequences
The five UGTs sequences, provided by Jörg M. augustin, served as a template for Polymerase
Chain Reaction (PCR) amplification. The primers were designed by Jörg and were extended
at the 5’end, containing a BamHI or NheI restriction site for the reverse and forward primers,
respectively, which is shown in blue and red in table 2. Two forward and two reverse primes
were designed to cover the four sequences.

Table 2: primers used for amplification of the sequences

<table>
<thead>
<tr>
<th>Sequence 1</th>
<th>Sequence 2</th>
<th>Primer 1</th>
<th>Primer 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>UGT73C9, UGT73C10, UGT73C11, UGT73C12</td>
<td>forward</td>
<td>5'-ATATGGCTAGCATGTTTCCGAAATCACCCA-3'</td>
<td></td>
</tr>
<tr>
<td>UGT73C13</td>
<td>forward</td>
<td>5'-ATATGGCTAGCATGTTTCAGAAATACCCAT-3'</td>
<td></td>
</tr>
<tr>
<td>UGT73C9, UGT73C11 UGT73C10</td>
<td>reverse</td>
<td>5'-AATTCGGATCTCAATTATAGATTGTCAGTTGC-3'</td>
<td></td>
</tr>
<tr>
<td>UGT73C12, UGT73C13</td>
<td>reverse</td>
<td>5'-AATTCGGATCTCAATTATGGATTGTCAGTTGC-3'</td>
<td></td>
</tr>
</tbody>
</table>

For the PCR reaction, the primers were in 0.5μM concentration, mixed together with 2U/μl
Phusion® High-Fidelity DNA polymerase, 0.4mM dNTP, 1μl template in 1x Phusion® HC buffer. To find the optimal annealing temperature of the primers, a gradient PCR
was performed with an initial denaturation step of 30 sec at 98˚C, 26 cycles of 10 sec at 98˚C
denaturation), a 56˚C gradient for 20 sec (annealing); 90 sec at 72˚C for (extension);
followed by final 10min 72˚C extension. In order to obtain sufficient amounts of the
fragments, the PCR was subsequently performed in 50μl reactions with the same conditions
but with 56˚C annealing temperature. The PCR amplicons were separated according to size
by gel electrophoreses (TAE, 1.2 % agarose). For estimation of the DNA size, a 1 Kb plus DNA
Ladder (invitrogen™) was loaded flanking the DNA samples.

Ligation
10ml of an E.coli culture containing the pET-28c vector was grown overnight (37˚C) and the
plasmids were isolated using Qiagen MiniPrep Spin kit. The vectors and fragments were
restricted with BamHI and NheI restriction enzymes and subsequently cleaned up using
Nucleospin Extract II. A molar ratio 3:1 between insert:vector in a total of 100ng in a 10μl
reaction was used. The corresponding formula (ng_vector * kb_insert)/kb_vector *3 = ng_insert, was used
to estimate the optimal ligation conditions. The length of the vector and insert were resp
5370bp and 1490bp, therefore 45ng insert and 55ng vector were used for the ligation. The
concentration of the vector and fragment solutions was determined with Nanodrop.
The ligation was performed in 10μl with 3 Weiss units pGEM® DNA T4 ligase overnight at 4˚C.

Transformation
10-20ng plasmids per 50μl competent E. coli XJb autolysis™ culture were used for
transformation. Transformations were performed either through electroporation or a 50
seconds heat shock on 42˚C. After a heat shock, the were incubated with 200μl SOC medium
for 1h 37˚C to recover and subsequently plated on LB+50μM kanamycin overnight 37˚C.
Insert containing colonies were selected by colony PCR with T7 primers flanking the multiple
cloning site of the pEt28c vector. Concentrations of the T7 primers and Hotmaster Taq
Materials and Methods

Polymerase were respectively 2,5μM and 5U/μl, with 0.2mM dNTPs in 1x Hotmaster Taq buffer. A PCR program with initial denaturation of 3 min 94˚C, 40 cycles of 94˚C 20 sec, annealing 51˚C 20sec, extension 2min 65˚C, and a final extension of 65˚C 10min as final step was used for the colony PCR. Test restrictions, incubating 100ng plasmid, 0.3μl of each restriction enzyme together for 1h at 37˚C, were performed to identify the inserted fragments (table 3). Plasmids containing the right sequence were sequenced (MWG/Eurofins). Stock cultures were made out of single colonies. UGT73C10 was cloned and transformed later by Jörg Augustin and included in the expression and characterization studies.

Table 3: restriction positions of PStI, EcoRV and PCiI on pET28+ with UGT73C9, UGT73C11, UGT73C12, UGT73C13 inserts.

<table>
<thead>
<tr>
<th></th>
<th>PStI</th>
<th>EcoRV</th>
<th>PCiI</th>
<th>expected length of the fragments (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UGT73C9</td>
<td>6034</td>
<td>5776; 3796</td>
<td>4592; 1979; 259</td>
<td></td>
</tr>
<tr>
<td>UGT73C11</td>
<td>6034</td>
<td>3796</td>
<td>4592; 2238</td>
<td></td>
</tr>
<tr>
<td>UGT73C12</td>
<td>3796</td>
<td></td>
<td>6830</td>
<td></td>
</tr>
<tr>
<td>UGT73C13</td>
<td>3796</td>
<td>688</td>
<td>3722; 3108</td>
<td></td>
</tr>
</tbody>
</table>

Expression of the UGTs

Crude extracts

For expression of the recombinant UGTs, 1.5ml LB medium (+50µM kanamycin) was inoculated with 10μl transformed E. coli culture and grown for 12h at 30˚C, 350rpm. Subsequently, 3ml TB medium and arabinose (final concentration 3mM) were added and gene expression was induced by 0.5mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Incubation was continued for 24h at 15˚C. E. coli cells were harvested by spinning (4°C, 14000xg, 20 minutes), resuspension of the pellet in 750µl 50mM Tris HCL pH7.5 and 1x Complete protease inhibitor (EDTA free), and stored in -80 until further usage.

A 250μl sample of every culture was taken after both incubation steps, and the OD$_{600}$ was measured with an Amersham® Ultrospec 3100 pro spectrophotometer.

After freezing at -80˚C, the cultures were thaw and spunned for 20min at 20000xg. 75μl 1% [W/v] Protamine Sulphate was added to remove remaining DNA, and after spinning another 20min 20000xg the supernatant was used as crude extract. These crude extracts are used for the activity assays. The presence of the UGTs in the crude extract was analyzed by a SDS Page and in some cases by Western Blotting.

SDS sample preparation

A solution containing 10% glycerol, 2% SDS, 5% β-mercaptoethanol 62.5mM Tris pH 6.8 and a spatula tip Bromphenol Blue was used as 4x concentrated sample buffer. Prior to loading the samples were reduced and denatured by heating for 5-10 min 95˚C with 1-2x sample buffer. To study the level of expression of the heterologous expressed genes, samples of E.coli cultures were analyzed by SDS-PAGE. The volume of these culture samples was normalized using the OD$_{0.6}$, whereby 1 OD$_{600}$ corresponded with 30μl culture sample. For the corresponding crude extracts 8.5 OD$_{600}$ corresponded with 6μl crude extract. When the
purpose was to analyze differences among the crude extracts, the volume crude extract used as a sample was constant among the samples (e.g. 5μl).

**SDS-Poly Acrylamide Gel Electrophoresis and Western blotting**

The samples were loaded on 12% or 14% SDS-polyacrylamide gels and run for 1.5h at maximum 200V and 100A (per gel) in electrophoresis buffer containing 0.05M in Tris, 192mM glycine and 3.47mM SDS. For SDS-PAGE analysis, the gels were stained with 10% acetic acid, 50% ethanol and 0.2 % (w/v) coomassie brilliant R, by incubation for 60-90 sec in the microwave (700W) until the gels colored dark-blue. Destaining was performed with 7% acetic acid in 2x 4 minutes in the microwave. To obtain maximum contrast, the gels could be further destained in 7% acetic acid overnight, by gently shaking at 20°C.

Western blotting was performed on unstained SDS pages. Proteins were transferred to a PVDF membrane (Immuno-blot™) using a Biorad™ criterion blotter. Blotting was done at 350mA for 1.5h. The transfer buffer contained 25mM Tris and 192mM Glycine, and was cooled during blotting. Blocking of the membrane was performed with 5% skimmed milk in PSB-Tween buffer (8.0mM K$_2$HPO$_4$, 0.15M NaCl, 3.9mM KH$_2$PO$_4$) for 0.5-3 h 20°C or overnight 4°C. After washing with PSB Tween buffer, the membrane was incubated with anti-his antibodies in a 5% skimmed milk in PSB-Tween buffer. As secondary antibody HRP anti-mouse (1:1000) was incubated at RT for 2 hours 20°C or 4°C overnight. The secondary antibody was visualized using Super Signal West® Dura Extended Duration Substrate and the Auto Chemi System.

**Activity assays**

**UDP-sugar assays (developed by Jörg M. Augustin)**

In the initial protocol a mixture (total volume 20μl) of 100mM Tris HCL pH 7.19, 1μl UDP-glucose (1.85 MBq/2mL, Amersham®), 10mM Dithiothreitol (DTT), 0.1mM acceptor substrate (solved in 80% ethanol) and 5μl crude extract was incubated for 30min at 37°C. The reactions were stopped by adding 20μl Ethyl Acetate, and spunned for 5min 20000xg to achieve phase separation. The organic phase (ca 20μl) was dried using a Scanvac vacuum centrifuge, resolved in 6μl ethyl acetate and subsequently transferred to a TLC silica gel 60 F$_{254}$ plate. The eluent was a mixture of 32:9:1 CHCl$_3$/MeOH/H$_2$O (as Shinoda, 2002). The TLC plates were processed in a phosphor-imager screen for 24h. Radio labeled compounds were depicted on the imager-screen and visualized with a phosphor imager (Molecular™ Dynamics Storm 860).

Subsequently, the above protocol was optimized. Both the concentration and pH of the Tris buffer were modified into 50mM Tris-HCL pH7.5. The DTT added to the lyses buffer, prior to freezing -80, instead of to the assays themselves and the assays were performed at 30°C instead of 37°C. When the purpose of the assays was to study substrate specificity, only 10μM substrate was used. For assays with only hederagenin and oleanolic acid as substrates, the 32:9:1 CHCl$_3$/MeOH/H$_2$O mixture was used as eluent. For assays with flavonols a mixture of 5:2:1:1 EtAc/MeOH/HCOOH/H$_2$O (Kurosawa, 2002) and for activity assays with saponins (incl. β-amyрин), flavonols, sterols and TCP a mixture of 7.5:0.5:1:1 EtAc/MeOH /HCOOH/H$_2$O was used.
Materials and Methods

Quantification of the TLC results
The phospho-imager data were analyzed with ImageQuant 5.0 software. For quantification the brightness/contrast ratio was optimalized and the spots were selected manually. The intensity of these selected regions was analyzed by the ImageQuant 5.0 software. The average intensity of the spots was used as a measure of for the amount of end product detected.

LC-MS assays
For LC-MS assays the above reaction as for the UDP-sugar assays was used, but substituted the radiolabeled UDP-glucose with 1mM unlabeled UDP-Glucose as sugar donor. The reaction was stopped with 100μl methanol. After spinning at 10000xg the supernatant was purified with 0.45nm filter and used diluted 5x in 50% methanol.

Solvent A in the LCMS analysis was: 0.1%HCOOH, 50uM NaCl and B: 80% MeCN, 0.1% HCOOH. Three programs were used to detect the monogluosides. The TOF analysis used a gradient from 88%A and 12%B to 0%A and 100%B during 19 minutes. The second program, used for the MS/MS techniques run the samples on a gradient from 63%A and 37%B to 20%A and 80%B in 9 minutes. A third program, used to detect flavonol and TCP monoglucosides, a 7 min gradient from 98%A and 2%B to 60%A and 40%B was used. The data were analyzed with Bruker Daltonics Data Analysis software.

Up scaling gene expression for NMR (protocol provided by Evolva).
For NMR, a sufficient amount (>2 mg) of the monogluoside was obtained using a 1,5L bacterial culture and a three day 50ml activity assay. 500ml NZYM medium was inoculated with 1-2 transformed culture and grown overnight at 30˚C. 1000 ml of the same medium was added, together with arabinose (final concentration 3mM) and IPTG (final concentration of 0,1mM), and incubation continued at 15˚C for 24h. Cells were harvested in a 50mL tube by spinning for 7min at 6500xg at 4˚C and resuspension in a lysis buffer (10mM tris, 5mM MgCl2, 1mM Ca 3 tablets/100 ml Complete® protease inhibitor) to total volume of 30mL. The cultures were stored at -80˚C.

After thawing the cultures to room temperature, 250μl DNaSI solution (1.4mg/ml) was added. After the viscosity had decreased sufficiently 1/3 volume of 4x binding buffer (80mM Tris-Hcl pH7.4; 2M NaCl) was added. After centrifugation for 15min 156000xg 4˚C, the supernatant was transferred to a fresh tube and spunned for 20min at 26000xg to remove the remaining DNA. 3mL of His-select (sigma P661) resin was incubated with the supernatant for 2h at 4˚C, subsequently centrifuged for 4min at 2000xg and 4˚C; the supernatant was discarded. The resin was washed 2-3 times with 1x binding buffer, with a final volume of 10-15 ml.

The final reaction had a total volume of 40-50 ml containing the resin, 100mM Tris-HCL pH8.0, 5mM DTT, 5mM MgCl2, 1mM hederagenin, 1mM UDP-glucose and 1% Fermentas FastAP phosphatase 1U/μl. The reaction was incubated on 30˚C under gently stirring, and was monitored by TLC analysis and staining with 10% sulphuric acid, which showed both aglycone and monoglucoside. When the ratio aglycone:monoglucoside did not notably change anymore, the reaction was stored at 4˚C until further usage. Prior to HPLC analysis,
supernatant was filtrated with a 0.22μM and filter (syringe) and the pH was adjusted to pH3-4 by addition of Trifluoroacetid acid (TFA). HLPC purification was performed by Evolva.

**NMR**

The NMR analysis was performed and interpreted by Carl Erik Olsen. NMR spectra were recorded in methanol-d$_4$ on a Bruker Avance 400 instrument using TMS as internal standard. 

$^1$H-NMR (methanol-d$_4$, 400 MHz) $\delta$: 0.71 (3H, s, CH$_3$), 0.81 (3H, s, CH$_3$), 0.91 (3H, s, CH$_3$), 0.94 (3H, s, CH$_3$), 0.98 (3H, s, CH$_3$), 1.17 (3H, s, CH$_3$), 0.7-2.1 (overlapping multiplets from steroid skeleton), 2.84 (1H, dd, J=4.2 & 13.9 Hz, H-18), 3.16 (1H, dd, J= 7.8 & 8.7 Hz, H-2’), 3.25-3.36 (m, overlapped by solvent), 3.60-3.69 (3H, m), 3.82 (1H, dd, J=2.1 & 12.0 Hz, H-6’), 4.38 (1H, d, J=7.9 Hz, H-1’), 5.24 (1H, broad t, J=3.5 Hz, H-12). $^{13}$C-NMR (methanol-d$_4$, 100.6 MHz) $\delta$: 13.4 (C-24), 16.4 (CH$_3$), 17.8 (CH$_3$), 18.9 (CH$_2$), 24.0 (CH$_3$), 24.1 (CH$_2$), 24.6 (CH$_2$), 26.3 (CH$_3$), 26.5 (CH$_3$), 28.9 (CH$_2$), 31.6, 33.5 (CH$_2$), 33.6 (CH$_3$), 33.9 (CH$_3$), 35.0 (CH$_2$), 37.7, 39.5 (CH$_2$), 40.6, 42.8, 43.0, 43.9, 47.3 (CH$_3$), 47.7, 48.2, 62.8 (CH$_2$, C-6’), 64.9 (CH$_2$, C-23), 71.6 (C-4’), 75.7 (C-2’), 77.8 (C-5’), 78.4 (C-3’), 83.5 (C-3), 105.8 (C-1’), 123.6 (C-12), 145.3 (C-13), 181.9 (C-28).

**His-tag purification**

His-tag purification was performed with the native and denaturized protein. This purification method is based on binding of the His-tag of the protein to a nickel-containing resin. Therefore 750μl of a crude extract was incubated with resin for 1h at 4˚C, spun through a NucleoSpin® RNA Plant filter afterwards and washed twice with 750μl washing buffer. Elution of the protein was performed in two steps with 250μl elution buffer and a final clean up step was performed with 750μl of a strong buffer (e.g. 1M EDTA) to make sure all the proteins were unbound from the resin. The fractions of all steps were collected and run on a SDS page.

For the native enzyme purification, the conditions for the resin binding step were 300mM NaCl, 10mM Immidazol, 50mM DTT and 50mM Tris pH 7.5. The washing buffer contained 300NaCl, 20mM Immidazol, 50mM DTT and 50mM Tris pH 7.5. Elution was obtained with a buffer of 300NaCl, 50mM Tris-HCl pH 7.5 and 50mM DTT and 150mM Immidazol. The Clean up was performed with 1M NaCl. In some cases EDTA was used to elute the protein, with an elution buffer containing 200mM and 2M EDTA as clean up buffer. EDTA attracts nickel and removes the nickel from the resin, thus competing with the resin.

For denaturized purification the 750μl crude was incubated with the resin in 8M urea and 50mM Tris. The washing step contained 8M urea and 10mM Tris-HCl, with a pH6.3. Elution steps were performed with the same buffer but the pH was respectively 5.9 and 4.5 for the two steps. Clean up was performed with 600μl NaAC –HCL pH3.

The eluted fractions were concentrated using to one IVSS vivaspin 500 centrifugal concentrat (centricon) and spinned for 5 min at 12500xg, or until a residual volume of 50μl was obtained. The centrifon was washed 2-3x with an appropriate buffer (e.g. 100mM Tris-HCl pH7.5 and 10mM Tris). The final volume of the purified solutions was ca 50μl.
Materials and Methods

Protein measurements
The concentrations of the purified enzyme were measured with Nanodrop at A280; performing Pierce BCA assays or by analyzing the intensity on SDS of Western with Image J. A standard curve was made with known concentrations of BSA, essentially according to the following the Pierce BCA Assay kit instructions, but using 5μl sample, 95μl sample buffer in a 96 multi-wells plate. After incubation the samples were kept on ice until they were analyzed with Nano-drop, using the corresponding BCA-program.

Standard curves with both BSA or known concentrations purified UGT were made with SDS-PAGEs and Western blots (performed as described above) and the gels were analyzed with Image J freeware.

Sequence analysis
Alignments of the sequences and corresponding phylogenetic trees and similarity tables were produced using MEGA4©, a program that makes use of a Custal W algorithm. Penalty for Gap was set on 10 and Gap extention 0.5.

The crystal structures were obtained from the Protein data bank (www.pbd.org) and used for analysis I with Pymol 1.3 (source: www.pymol.org.) The structures used were UGT71G1 containing UDP-glucose: PBD 2ACW and VvGT: PBD 2C9Z.
3. Results

3.1 Phylogenetic analysis of the UGT cDNA and amino acid sequences

The sequences of all known members of the UGT73C subfamily, including *B. vulgaris* variegata UGT73C<sub>shi</sub> and the five UGT derived from *B. vulgaris* arcuta were aligned with Clustal W. The parameters were a gap penalty of 10 and a gap extension penalty of 0.5 for both the pairwise and multiple alignments. A Neighbor Joining tree was constructed based on this alignment, for both the nucleotide and amino acid sequence (figure 8). As statistical bootstrapping was used (1000x). As an out group UGT71G1 is used, a family 1 glycosyltransferase from *Medicago truncutala* that is able to glycosylate isoflavones and flavonols and the sapogenin medicagenic acid (He et al., 2008; Achnine et al., 2005). The p-distance, defined as ‘the proportion (p) of amino acid (or nucleotide for DNA sequences) sites at which the two sequences to be compared are different’ (Mega4 2010), is used to calculate the percentage pairwise identity (1- p-distance) as shown in table 4.

The five UGT sequences from *Barbarea vulgaris* ssp arcuta form together with the gene from *B. vulgaris* ssp variegata a distinct clade within the UGT73C subfamily (figure 8). Two
Results

subclades can be distinguished: UGT73C12 and UGT73C13 form one subclade and UGT73C9, UGT73C10 and UGT73C11 and UGT73Cshi cluster together in the second. Orthologs and paralogs cluster together in a QTL analysis of Barbarea vulgaris (Kuzina, Augustin, unpublished). Their neighboring positions imply gene duplication.

UGT73C11 is the closest homolog to B.v. variegata UGT73Cshi, they share only one amino acid substitution of an aspartic acid (UGT73Cshi) into a glutamic acid (UGT73C11) at position 338 (figure 9). Therefore UGT73C11 will be referred to as the ortholog for UGT73Cshi in the G type plant of B vulgaris ssp arcuata. The NJ analysis on amino acid level indicates that UGT73C10 is the ortholog in the P-type plant (98% similar to UGT73C11). UGT73C12 and UGT73C13 share 98% (table 4) of their amino acid sequence and cluster together in both nucleotide and amino acid bootstrap consensus trees. UGT73C12 was isolated from the P plant and UGT73C13 from the G plant, indicating that there are orthologs towards one another but paralogs of the orthologs of UGT73Cshi from B.v. variegata.

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>C8</th>
<th>C9</th>
<th>C10</th>
<th>C11</th>
<th>C12</th>
<th>C13</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>0.30</td>
<td>0.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>0.29</td>
<td>0.65</td>
<td>0.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4</td>
<td>0.28</td>
<td>0.66</td>
<td>0.76</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C5</td>
<td>0.29</td>
<td>0.68</td>
<td>0.73</td>
<td>0.76</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C6</td>
<td>0.27</td>
<td>0.65</td>
<td>0.70</td>
<td>0.74</td>
<td>0.73</td>
<td>0.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C7</td>
<td>0.26</td>
<td>0.50</td>
<td>0.51</td>
<td>0.52</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.48</td>
<td>0.40</td>
<td>0.40</td>
<td>0.40</td>
</tr>
<tr>
<td>C8</td>
<td>0.27</td>
<td>0.44</td>
<td>0.41</td>
<td>0.44</td>
<td>0.42</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
<td>0.41</td>
<td>0.48</td>
<td>0.40</td>
<td>0.40</td>
<td>0.40</td>
</tr>
<tr>
<td>C9</td>
<td>0.29</td>
<td>0.67</td>
<td>0.69</td>
<td>0.71</td>
<td>0.72</td>
<td>0.77</td>
<td>0.75</td>
<td>0.48</td>
<td>0.40</td>
<td>0.49</td>
<td>0.41</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>C10</td>
<td>0.29</td>
<td>0.67</td>
<td>0.70</td>
<td>0.73</td>
<td>0.73</td>
<td>0.79</td>
<td>0.76</td>
<td>0.49</td>
<td>0.41</td>
<td>0.95</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>C11</td>
<td>0.29</td>
<td>0.67</td>
<td>0.70</td>
<td>0.72</td>
<td>0.73</td>
<td>0.79</td>
<td>0.76</td>
<td>0.49</td>
<td>0.41</td>
<td>0.95</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>C12</td>
<td>0.28</td>
<td>0.68</td>
<td>0.70</td>
<td>0.74</td>
<td>0.74</td>
<td>0.83</td>
<td>0.79</td>
<td>0.51</td>
<td>0.43</td>
<td>0.90</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>C13</td>
<td>0.28</td>
<td>0.67</td>
<td>0.70</td>
<td>0.74</td>
<td>0.74</td>
<td>0.83</td>
<td>0.79</td>
<td>0.51</td>
<td>0.43</td>
<td>0.90</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>UGT73Cshi</td>
<td>0.29</td>
<td>0.67</td>
<td>0.69</td>
<td>0.72</td>
<td>0.73</td>
<td>0.79</td>
<td>0.76</td>
<td>0.49</td>
<td>0.40</td>
<td>0.95</td>
<td>0.98</td>
<td>1.00</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Table 4: percentage amino acid sequence identity of UGT73C subfamily members
The closest related known subfamily members of the UGTs derived from *B. vulgaris* are UGT73C5 and UGT73C6 from *Arabidopsis thaliana*. They are ca 80% percent identical to the UGTs derived from *Barbarea vulgaris*. UGT73C5 has been characterized as a gene involved in the glucosylation of brassinosteroids, a class plant hormones that are involved cell elongation and differentiation (Poppenberger et al., 2005). Brassinosteroids are triterpene structures, derived from the phytosterol pathway, with campesterol as precursor. Glucosylation of brassinosteroids has a regulatory function as it inactivates the brassinosteroids, and over expression of UGT73C5 in Arabidopsis led to a dwarf phenotype.

Besides glycosylation of brassinosteroids, UGT73C5 is also capable of glycosylating the cytokinins (Poppenberger et al., 2005) trans-zeatin and dihydrozin (Hou et al., 2004) and both UGT73C5 and UGT73C6 can utilize zearelone, a myco-toxin with estrogenic activity (Poppenberger et al., 2006). UGT73C6 has shown to be *in vivo* involved in the glycosylation of quercetin at the 3 and 7 positions (Jones et al., 2003).

UGT73C3, UGT73C4, UGT73C1 and UGT73C are ca 70% identical, UGT73C3 and UGT73C4 closer related than UGT73C2 and UGT73C1. Little is known about the function or activity of these four UGTs. All are derived from *A. thaliana*. Like UGT73C5, UGT73C1 was capable of glycosylating of trans-zeatin and dihydrozeatin. UGT73C3 and UGT73C4, also included in this research did not show activity on these substrates (Hou et al., 2004). Within the UGT73C subfamily the *A. thaliana* UGT73C7 and *M. truncatula* UGT73C8 are the least similar to the UGT from *Barbarea vulgaris*, with a sequence identity of 40-50 percent.
Results

3.2 UGT sequence analysis

UGT73C9 shows an altered solubility and activity towards oleanolic acid and hederagenin (see 3.3) compared to the other four UGTs. To gain more insight in the amino acid substitutions underlying these differences, the sequences were aligned with UGT71G1, a UGT that can utilize flavonols and saponins and of which a crystal structure is available. The sites that differ from the consensus sequence are highlighted in figure 9.

Figure 9: CLC alignment of UGT73C9, UGT73C10, UGT73C11, UGT73C12, UGT73C13, UGT73C14, and UGT71G1 using Clustal W algorithm. Amino acids that differ from the consensus sequence are marked in pink. The PSPG box is underlined in black.
The crystal structure of UGT71G1 was used to study the amino acid substitutions in UGT73C9 according to the alignment above. A crystal structure of UGT71G1 containing the UDP-sugar was available (PDB 2ACW), and aligned with VvGT (PBD 2C9Z) (Offen et al., 2006) containing quercetin and UDP as substrates. Thus in figure 10 both substrates are visualized, derived from different crystal structures. The substrates are shown as sticks and are stained by elements in which carbon is green for the UDP-sugar in the UGT71G1 model, whereas quercetin is stained by elements but with purple carbon. The corresponding amino acids of the positions in UGT71G1 where UGT73C9 differed from other Barbarea UGT73C are shown as sticks and marked in red. This visualization can indicate how the amino acids of the substitutions in UGT73C9 can interact with UDP-glucose and an acceptor substrate. The active center is shown in figure 10.

![Figure 10: Alignment of the crystal structure of UGT71G1 with UDP-glucose, and VvGT containing quercetin and UDP. Only the active center of the enzyme is shown. The substrates are shown as sticks and are stained by elements with O in red, N in blue and P in orange. However, Carbon atoms are shown white in UDP, green in UDP glucose and purple in quercetin. UGT71C1 and VvGT are overlaid and showed as cartoon in which UGT71G1 is orange and VvGT is blue. Amino acid of UGT71G1 which have a unique amino acid substitution in UGT73C9 are shown as sticks and indicated in red.](image-url)
Results

Table 5 one summarized the amino acid substitutions that are indicated in red in figure 10. At positions 210, 297, and 379 corresponding with position 202, 286, and 395 of UGT71G1, UGT73C9 conferred unique amino acid substitutions towards compared with UGT73C10 and UGT73C11. Even though these sites are in the active center, these sites do not seem to have in direct contact with the substrate. However, an effect of these amino acid substitutions on activity cannot be excluded as the amino acids are so close to the active center. The Tyr202 of UGT71G1, however, might well interact with the acceptor substrate (figure 10). In the functional UGT73C10 this tyrosine is replaced by a lysine and in the non-functional UGT73C9 by an arginine. Lysine and arginine are very similar amino acids, but the extra length of the arginine in UGT73C9 may have a steric effect to the acceptor substrate, as it changes the size and shape of the cavity that fits the acceptor substrates.

Table 5: amino acid substitution

<table>
<thead>
<tr>
<th>UGT71G1 position</th>
<th>UGT71G1</th>
<th>UGT73C10 position</th>
<th>UGT73C10 &amp; 11</th>
<th>UGT73C9</th>
</tr>
</thead>
<tbody>
<tr>
<td>194</td>
<td>Asn</td>
<td>202</td>
<td>Val</td>
<td>Ala</td>
</tr>
<tr>
<td>202</td>
<td>Tyr</td>
<td>210</td>
<td>Lys</td>
<td>Arg</td>
</tr>
<tr>
<td>286</td>
<td>Met</td>
<td>297</td>
<td>Ile</td>
<td>Asn</td>
</tr>
<tr>
<td>379</td>
<td>Tyr</td>
<td>395</td>
<td>Phe</td>
<td>Ile</td>
</tr>
<tr>
<td>380</td>
<td>Ala</td>
<td>396</td>
<td>Gly</td>
<td>Val</td>
</tr>
</tbody>
</table>

Interesting is an amino acid substitution in the PSPG box, even the site itself is not very conserved according to the PSPG box in Vogt and Jones 2000. All B. vulgaris except UGT73C9 contain an alanine on that position, and UGT71G1 a glycine. UGT73C9, however, contains a valine, which contains has an extra CH₃ group. Figure 10 shows this alanine is located exactly in between donor and acceptor and the bigger size of the extra methyl group in valine may form a steric barrier between donor and acceptor substrate, and thereby blocking the activity of the enzyme. This could explain the lack of activity on sapogenins and the low activity on TCP. Further studies will be needed to study the effect of this amino acid substitution.

However, the altered solubility of UGT73C9 can be casual by amino acid substitutions outside the active center. For example, the substitution of an arginine into an a cysteine on position 167 introduces an extra S group on this position which might interfere with the folding of the enzyme.
3.3 Cloning and heterologous expression of the UGTs

The cDNA sequences of UGT73C9, UGT73C11, UGT73C12 and UGT73C13 were amplified by PCR, purified, ligated into a pET28c vector and transformed into DH5α or Xjb strains of E. coli. Insert containing colonies were selected by colony of culture PCR. The five UGTs could be individually distinguished on their EcoRV, PstI and PciI restriction pattern. Plasmids containing the right insert were selected and sequenced. Often the sequences contained point mutations compared to the original sequence, and were not used in this thesis. The UGT73C10 sequence was cloned and transformed by Jörg M. Augustin and included in the expression and characterization studies.

To study the expression of the recombinant UGTs, samples of the bacterial culture before and after expression, the crude extracts and corresponding insoluble phase were analyzed by SDS-PAGE figure 11. The SDS-PAGE analysis reveals induction of proteins with a size of approximately 55kDa, which is within the expected UGT size region. UGT expression is found in all cultures besides the empty vector culture which reveals one or more proteins with approximately the same size as a background in the SDS-PAGE analysis. All the expressed proteins could be seen in the soluble phase except for UGT73C9, which could not be distinguished from the background (figure 11C) but is present in the insoluble phase as

![Figure 11: SDS-PAGE analysis of bacterial cultures before (A) and after (B) induction with IPTG. C) The corresponding according crude extract (soluble phase) D) insoluble phase. The size (kDa) of the fragments in the marker lane is indicated on the left. The size region of the recombinant UGT is indicated with a black arrow.](image-url)
Results

inclusion bodies (figure 11D).

To study the formation of inclusion bodies in UGT73C9, a sequential dilution series of the IPTG concentration was used for induction of expression (figure 12A,B). Western Blot analysis revealed the presence of UGT73C9 in the crude extract (figure 12D). However, the concentration of UGT73C9 in the crude extract is obviously much lower than the concentration of the other UGTs.

Figure 12: SDS-page (A,B,C) and Western Blot (D) of UGT73C9. Bacterial culture before (A) and after (B) induction by IPTG. (C) crude extract, and D) the corresponding Western Blot analysis. The concentration IPTG used to induce gene expression is indicated below each lane. The arrows indicated the position of the heterologous expressed UGT.
3.4 Characterization of the UGTs

To study the activity of the heterologous expressed UGTs, activity assays were performed with radioactively $^{14}$C labeled UDP-glucose as donor and oleanolic acid and hederagenin as acceptor substrates. The radioactive monoglucosides could be detected with a phosphor-imager. The TLC analysis revealed glycosylation activity for UGT73C10, UGT73C11, UGT73C12, UGT73C13 and UGT73C$_{shi}$ (positive control) for the substrates oleanolic acid and hederagenin (figure 13). No monoglucosides were detected for the empty vector crude extract. Also UGT73C9 did not show any activity on these aglycones, which can be due to various reasons such as a low concentration of UGT73C9 in the crude extract, an inactive form of the protein in the crude extract, a very low affinity of UGT73C9 for hederagenin and oleanolic acid, or a combination of the above mentioned reasons. However, in activity assays with different substrates reveal that UGT73C9 is able to utilize 2,4,6-trichlorophenol (TCP) (figure 14D), thus UGT73C9 is active in the crude extract of UGT73C9.

Activity assays were performed to study the activity of the UGTs under different conditions regarding pH, temperature, buffer concentration, etc. The TLC results indicate that the amount of end product correlates with the pH of the Tris-HCl buffer, with an optimal pH of
Results

pH7.5 (figure 14A). No pattern was found in LCMS analysis of assays of UGT73C12 with 20, 50 and 100mM Tris concentrations. Interestingly, the concentration of hederagenin and oleanolic acid in a range of 10-100µM did not seem to have an effect on the amount of monoglucoside produced in the assays, suggesting that these concentration of oleanolic acid and hederagenin are within the $V_{\text{max}}$ range of UGT73C11. 10 times dilution of the crude extract, however, correlated with a 2.5 times smaller TLC dot intensity. LCMS data of UGT73C12 revealed that the enzyme seemed to have a similar activity well on 30 and 37°C, however no monoglucosides could be detected after 30 min incubation at 20°C. Activity assays using fresh crude extract and the same crude extract after storage on ice (0°C) overnight indicated that both UGT73C12 and UGT73C10 are stable on ice for at least 20h, as no difference in the amount of monoglucoside could be detected.

Figure 14: TLC analysis of UGT73C members of B.v.arcuata in varying assay conditions. Assays were in principle preformed with 100µM Tris pH 7.2 and 100µM substrate concentration, and varying conditions treatments are indicated for each lane the. A) UGT73C11 activity with varying pH and concentrationTris buffer. B) Activity of UGT73C11 with 10-100µM oleanolic acid and hederagenin concentration. C) activity of UGT73C9, UGT73C10 and UGTC12 on TCP before and after 24h storage on ice. Below: UGT739 (D), UGT73C10 (E), UGT73C12 (F) tested for activity on different substrates: ol = oleanolic acid, he = hederagenin, qu = quercetin, ka = kaempferol, TCP = 2,4,6, tri-chlorophenol (UGT73C11 and UGT73C13 not shown).
In contrast, after freezing (-20) the activity of UGT73C12 gradually decreased (figure 15). Addition of up to 50% glycerol limited the effect of freezing to some extent, but it was not sufficient to maintain the original activity. UGT73C11 did not show loss of activity after freeze thawing, with or without addition of glycerol. The percentage glycerol itself did not have a notable effect on enzyme activity (figure 15).

![Figure 15: The TLC intensity of activity assays UGT73C13 and UGT73C12 with addition 0, 10, 20, 30, 40 and 50% glycerol to the crude extracts after one, two and 3 freeze-thaw cycles (respectively 0, 1, 4 and 5 days in -20°C). The volume of the crude extract used in the activity assays was adjusted to the percentage glycerol in the crude extract. The percentage glycerol is indicated below each lane and the amount of freeze-thaw cycles is represented by the bar colour.](image)

The activity assays with oleanolic acid and hederagenin as acceptor substrate and were analyzed by LCMS to determine the mass of the compounds detected by the TLC analysis. The predicted mass of the monoglucosides is 619 for oleanolic acid and 634 for hederagenin, based on the masses of the oleanolic acid (mw 457) and hederagenin (mw 472) aglycone. Different LCMS programs were applied to detect the monoglucosides, using either positive or the negative detection mode of the program.

An LCMS analysis of activity assays of UGT73C10 in negative mode is shown in figure 16. The main compound in the peak at 6.1 min in assays with hederagenin has a weight of 679 m/z, which corresponds to the hederagenin formic acid adducts (figure 16). The fragments of this compound in MS/MS analysis represent the monoglucoside (m/z 634), and the difference in mass is represented by the loss of a formic acid ion. The MS/MS/MS revealed a compound with the mass of the hederagenin aglycone (m/z 472) is found. The loss of the sugar moiety is represented by a loss in m/z 162.
The oleanolic acid formic acid adduct was found at 8.3 min with a mass of m/z 619 that represented the monoglucoside, and the aglycone was represented by a fragment of m/z 456 in the MS/MS. In the positive mode 50μM NaCl (mw 23) was added to the mobile phase to facilitate the detection of the monoglucosides, and the masses we found were m/z 619 and m/z 656 representing the Na+ adducts of the hederagenin and oleanolic acid monoglucosides.
Results

An NMR analysis was performed on the hederagenin monoglucoside to determine the position of glucosylation and the conformation of the sugar. For this purpose >2 mg of the hederagenin monoglucoside was required. Therefore activity assays were performed on a large scale with 1.5L UGT73C11 containing culture and a 50ml activity assay for two days. Ca 24mg hederagenin aglycone was used for the activity assay; however, only 2.1mg of the monoglucoside was purified. The unexpected low amount of purified monoglucoside is probably due to a loss of monoglucoside during the pH adjustment and 0.21μm filtration of the activity assay prior to the HPLC purification, as precipitation of one of the compounds in the assay was observed. The NMR analysis of the hederagenin monoglucoside revealed that the glucose moiety was attached to the C3 position of the aglycone and the sugar was in β-conformation (figure 17).

3.5 UGT substrate specificity

UGT73C9, UGT73C10, UGT73C11, UGT73C12, UGT73C13 and also A. thaliana UGT73C5 UGT73C6, were tested for their ability to utilize different compounds as substrate. A relatively low substrate concentration was used (10μM) so that the differences in efficiency of the enzyme per substrate would be detectable. The acceptor substrates included oleanolic acid and hederagenin, the substrates for the presumed biosynthetic step. Also β-amyrin, the presumed precursor of oleanolic acid and hederagenin, was tested. Other acceptor substrates included flavonols, quercetin and kaempferol, and phytosterols (sitosterol, campesterol, sigmasterol, obtusifoliol) (figure 18). Quercetin and kaempferol are the major flavonols in model plant A.thaliana and many UGTs that utilize sapogenins are also active on quercetin aglycones. The saponin biosynthesis pathway is derived from the phytosterol pathway, thus it would be relevant to also test sterols as substrate. Sitosterol, sigmastero, campesterol and obtusifoliol were available as substrates. 2,4,6-trichlorophenol
Results

(TCP) served as a positive control for activity of the UGTs and was used in 100μM concentration. TLC results of these activity assays are shown in figure 19 and summarized in table 6.

Figure 18: Substrates used in the activity assays. The tripenoid sapogenin aglycones hederagenin, oleanolic acid and β-amyrin are marked in blue, the two flavonols kaempferol and quercetin in green, the four sterols sigmasterol, sitosterol, campesterol, obtusifoliol in brown and 2,4,6 trichlorophenol (TCP) in black.

The activity assays reveal that UGT73C10, UGT73C11, UGT73C12 and UGT73C13 can utilize oleanolic acid and hederagenin as a substrate. Two glycosylation patterns were found, one for UGT73C10, UGT73C11 and one for UGT73C12, UGT73C13. Therefore only one representative of each (UGT73C10 and UGT73C12) is shown in figure 19. The spot intensity of all is summarized in table 6. UGT73C10 and UGT73C11 were clearly active towards oleanolic acid, hederagenin, and β-amyrin aglycones. A smear in the lanes of quercetin and kaempferol indicated also activity for the flavonols, but much weaker than the activity on sapogenins. Also UGT73C12 and UGT73C13 were active on oleanolic acid and hederagenin, but the spots were less intense than those of UGT73C10 and UGT73C11 suggesting a lower efficiency of UGT73C12 and UGT73C13 under these conditions. The activity of UGT73C12 on TCP was similar to that of UGT73C10.

Whereas UGT73C10 and UGT73C11 seem to utilize β-amyrin, oleanolic acid and hederagenin equally efficient, it is not clear if the R_f 0.73 compound in UGT73C12 (and UGT73C13) activity assays is a real β-amyrin monoglucoside. (Impurities of) 80% ethanol appears to be glucosylated and the products gave a R_f 0.28 and 0.73, which is the same retention time as
the β-amyrin monoglycoside (which is solubilized in 80% ethanol). LCMS analysis could neither detect β-amyrin in assays with UGT73C12 and UGT73C13 nor assays in with UGT73C10 and UGT73C11. Thus, whether UGT73C12 and UGT73C13 can utilize β-amyrin as a substrate could not be determined. A weak spot is detected in the lane of UGT73C9 and TCP, but not for any other substrate tested.

Activity towards kaempferol and quercetin was detected for all four UGTs, although it was very weak compared to oleanolic acid and hederagenin. The more intense spot in UGT73C6 reveals a quercetin monoglucoside with an $R_f$ of 0.58 (data not shown). In the lanes with the sterols as acceptor substrate, vague spots are detected with an $R_f$ of 0.67. An additional
Results

smear is detected in the lane with UGT73C12 and sigmasterol. However, the monoglucosides are solubilized in 1%tween instead of 80% ethanol, and a control experiment with 1% tween as substrate is lacking to determine if the compounds on Rf 0.67 or 0.73 are sterol-glucosides.

The intensity of the main compounds for each lane is summarized in table 6. The intensity values in table supports the difference in glycosylation for UGT73C10-UGT73C11 and UGT73C12-UGT73C13 as mentioned earlier. However, the enzyme concentration is not taken into account, as we did not find a feasible way to determine the amount of enzyme in the crude extract in this master (see section 3.6). Correlations between the intensity of the UGT band on SDS-PAGE and the activity on either the positive control TCP or any other compound in the assays were not found (compare table 6 with figure 20). However, figure 14 shows that a 10x dilution of a crude extract does affect the amount of end products formed. The affinity for TCP may differ among the enzymes, and also the percentage enzymes in an active state may differ between the UGTs, which could explain why this correlation between the SDS-PAGE and the difference in activity towards TCP or any other substrate was not found.

Table 6: summary of the spot intensity of the TLC analysis with different acceptor substrates

<table>
<thead>
<tr>
<th></th>
<th>oleanolic acid</th>
<th>hederagenin</th>
<th>quercetin</th>
<th>kaempferol</th>
<th>obtusifoliol</th>
<th>sitosterol</th>
<th>sigmasterol</th>
<th>campesterol</th>
<th>β-amyrin</th>
<th>TCP*</th>
</tr>
</thead>
<tbody>
<tr>
<td>UGT73C9</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>13</td>
</tr>
<tr>
<td>UGT73C10</td>
<td>480</td>
<td>717</td>
<td>33</td>
<td>24</td>
<td>17</td>
<td>13</td>
<td>12</td>
<td>11</td>
<td>439</td>
<td>292</td>
</tr>
<tr>
<td>UGT73C11</td>
<td>507</td>
<td>893</td>
<td>23</td>
<td>31</td>
<td>14</td>
<td>10</td>
<td>nd</td>
<td>329</td>
<td>471</td>
<td></td>
</tr>
<tr>
<td>UGT73C12</td>
<td>91</td>
<td>85</td>
<td>nd</td>
<td>28</td>
<td>13</td>
<td>13</td>
<td>15</td>
<td>13</td>
<td>26</td>
<td>298</td>
</tr>
<tr>
<td>UGT73C13</td>
<td>70</td>
<td>49</td>
<td>12</td>
<td>10</td>
<td>11</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>31</td>
<td>190</td>
</tr>
<tr>
<td>UGT73C5</td>
<td>76</td>
<td>20</td>
<td>nd</td>
<td>44</td>
<td>14</td>
<td>16</td>
<td>15</td>
<td>16</td>
<td>29</td>
<td>459</td>
</tr>
<tr>
<td>UGT73C6</td>
<td>nd</td>
<td>nd</td>
<td>285</td>
<td>56</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>563</td>
</tr>
</tbody>
</table>

* The TCP concentration was 100μM, whereas 10μM was used for the other substrates.

** nd= not detected.

Figure 20: Crude extracts used for the activity assays in figure 19 and table 6. The recombinant UGTs are indicated below each lane and their size is indicated by the black arrow. NC is the negative control, a crude extract containing an empty vector. UGT73C5 is less intense because of a pipetting mistake.
Even though TCP could not be used to normalize the UGT concentration and compare crude extracts among each other, the ratio in which a UGT can utilize the substrates could be used to study efficiency of a UGT towards different substrates. For example, UGT73C10 and UGT73C11 are more efficient in utilizing oleanolic acid and hederagenin than they are with the flavonols and sterols. Moreover, the difference between the capacity of UGT73C12 and UGT73C13 to utilize sapogenins or flavonols is smaller than that of UGT73C10 and UGT73C11. This indicates that UGT73C10 and UGT73C11 are more specific to oleanolic acid and hederagenin than UGT73C12 and UGT73C13. To predict differences in $K_M$ values of the enzymes towards the different substrates, information is needed about the $V_{\text{max}}$ of the UGTs for every substrate. The $K_M$ for UGT73C11 towards oleanolic acid and hederagenin is most likely <10μM, as the spot intensity did not decrease when 10μM instead of 100μM of substrate concentration was used.

### Table 7: spot intensity relative to TCP

<table>
<thead>
<tr>
<th></th>
<th>oleanolic acid</th>
<th>hederagenin</th>
<th>quercetin</th>
<th>kaempferol</th>
<th>obtusifoliol</th>
<th>sitosterol</th>
<th>sigmasterol</th>
<th>campesterol</th>
<th>β-amyrin</th>
<th>TCP*</th>
</tr>
</thead>
<tbody>
<tr>
<td>UGT73C9</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>1.00</td>
</tr>
<tr>
<td>UGT73C10</td>
<td>1.64</td>
<td>2.45</td>
<td>0.11</td>
<td>0.08</td>
<td>0.06</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>1.50</td>
<td>1.00</td>
</tr>
<tr>
<td>UGT73C11</td>
<td>1.08</td>
<td>1.90</td>
<td>0.05</td>
<td>0.07</td>
<td>0.03</td>
<td>0.02</td>
<td>nd</td>
<td>nd</td>
<td>0.70</td>
<td>1.00</td>
</tr>
<tr>
<td>UGT73C12</td>
<td>0.30</td>
<td>0.29</td>
<td>0.00</td>
<td>0.09</td>
<td>0.04</td>
<td>0.04</td>
<td>0.05</td>
<td>0.04</td>
<td>0.09</td>
<td>1.00</td>
</tr>
<tr>
<td>UGT73C13</td>
<td>0.37</td>
<td>0.26</td>
<td>0.07</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
<td>0.05</td>
<td>0.16</td>
<td>1.00</td>
</tr>
<tr>
<td>UGT73C5</td>
<td>0.17</td>
<td>0.04</td>
<td>nd</td>
<td>0.10</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.06</td>
<td>1.00</td>
</tr>
<tr>
<td>UGT73C6</td>
<td>nd</td>
<td>nd</td>
<td>0.10</td>
<td>0.10</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

* The TCP concentration was 100μM, whereas 10μM was used for the other substrates.

** nd= not detected.

### 3.6 Kinetics

A more convenient way to study the substrate specificity of an enzyme is to determine kinetic parameters like the $K_M$, $V_{\text{max}}$, and $k_{\text{cat}}$ for each substrate. The $V_{\text{max}}$ is the maximum velocity in which an enzyme can utilize its substrates. With low concentrations of substrate the amount of formed product is linear related to the amount of substrate available for the enzyme, as there are ‘free’ enzymes enough to catalyze the reaction. In higher substrate concentrations the enzymes get saturated with substrates and the amount of product formed is then dependent on the maximum speed of the enzyme ($V_{\text{max}}$). The substrate concentration on which the protein reaches half of the $V_{\text{max}}$ is the $K_M$, and is a measure for the efficiency of the enzyme towards a substrate. A high $K_M$ implies that the enzyme reaches...
Results

its maximum velocity at a high concentration and therefore a low efficiency towards the
substrate. To determine the $K_M$ and $V_{\text{max}}$ of the enzyme, information on the concentration of
the enzyme is not necessary as long it is constant between the assays. The $k_{\text{cat}}$ is the
turnover number, thus the number of times each enzyme site converts substrate to product
per unit time. Therefore, precise information about the enzyme concentration and substrate
concentration is inevitable.

To explore the optimal conditions for kinetics a test time series was made with UGT73C11
and hederagenin (figure 21). The time series indicated that the linear range of the graph lies,
under these conditions, within the first few minutes of the times series. Kinetics should be
performed with the linear range of the graph, thus the conditions need to be optimized
regarding enzyme and substrate concentration. A difficulty in finding these conditions is that
the enzyme concentration differs between the crude extracts; therefore the optimal
conditions also differ per crude extract.

![Figure 21: Time series of UGT73C11 with 1mM UDP-glucose 100μM hederagenin in Tris-HCL pH 7.19. The peak
area on the Y axis is derived from LCMS data.](image)

Therefore the recombinant UGTs were purified by his-tag purification. However, in
*B.v.variegata* UGT73C_{shi} a loss of activity after purification was found (Augustin, personal
communication) Activity assays of UGT73C10 and UGT73C12 did not show an obvious loss of
activity (figure 21), but one should keep in mind that the enzyme concentration after
purification is likely to be higher than in the crude extract (recombinant UGTs from 750μl
crude extract were concentrated into 50μl purified solution). Therefore it was preferred to
use the untreated crude extract for kinetics and activity assays. Several methods have been
explored to determine the UGT concentration circumventing purification.
For the first approach two 2ml samples derived from one culture were used, using one batch as crude extract while determining the total amount of UGT in the other batch by his-tag purification and protein measurements. These batches contained an equal amount of the same bacterial culture and therefore theoretically an equal total amount of the UGT. Determining the total amount of protein in one batch would enable us to calculate back the concentration of this protein in the crude extract of the other batch. However, the His-tag enzyme purification (purification of native enzyme using immidazol or EDTA elution step as well as purification based on the denaturized enzyme) always resulted in a loss of enzyme in both the resin binding and elution step and was therefore unsuitable for this purpose.

For the second approach the intensity of purified UGT on an SDS-PAGE was analyzed, using Image J. BSA was used to produce a standard curve to determine the concentration of the purified UGT and the UGT sample in each lane was used to explore the sensitivity and reliability of this method. The standard curve was used in a range of 200-800μg/ml BSA, using 2,5μl samples, and resulted in an $R^2$ of 0,994 (figure 24). The standard deviation of 7 repetitions of the purified UGT was ca 10% of the total value, and 4% after discarding the most extreme values. It was however not feasible to compare the standard curve of purified enzyme (BSA) with the UGT bands in the crude extract, due to the high background in the SDS-PAGE analysis of the crude extracts (figure 23).
However, the purified and known concentration of UGT (determined by SDS-PAGE analysis) could be used to make a standard curve for Western Blotting. Anti-his antibodies were used so that only the recombinant UGTs were detected, thus circumventing the previous background problems of the in SDS-PAGE analysis.
An initial Western Blot was performed on a series of 250, 500, 750, 1000 and 1250μg/ml purified UGT. After black and white inversion, the same Image J analysis technique was used to determine the band intensity as used for the SDS-PAGE analysis. Western Blotting is a sensitive method and the lane with 1000μg/ml was clearly overloaded. However, in the enzyme range of 250-750μg/ml, the correlation between band intensity and enzyme concentration was linear, with an $R^2=0.99$, but more measuring points are necessary to obtain a more reliable result. Moreover the enzyme concentration of the (4x diluted) crude extract was below this range of 250-750μg/ml. Because of time limitations this strategy was not further optimizations during this thesis.

Figure 23: Correlation between the intensity of the peak area in Image J analysis of the SDS page (see figure 23) and the concentration BSA.
4. Conclusions and discussion

This thesis is the first characterization of a candidate UGT in the saponin biosynthesis pathway of *Barbarea vulgaris arcuata*. The main focus was to study if the genes could be involved in the biosynthetic pathway of saponins, but also to study differences between genes with respect to their role in resistance of *Barbarea* towards the flea beetle. Therefore four research focus questions were proposed, which will be the structure of following discussion. These research questions are: i) Are the five UGTs from *B. vulgaris arcuata* capable of glycosylating the hederagenin and oleanolic acid aglycone on C3 position? ii) How specific are these UGTs towards oleanolic acid and hederagenin as acceptor substrate? iii) Is there a difference in activity between the UGTs from the P and the G type plant? iv) Is there a difference in activity between the ortholog the paralogs?

i) *Are the five UGTs from B. vulgaris arcuta capable of glucosylation of hederagenin and oleanolic acid aglycone at the 3 position?*

The TLC analysis of figure 13 reveals activity of UGT73C10, UGT73C11, UGT73C12 and UGT73C13 towards oleanolic acid and hederagenin. More than one glycosylation product was present in each lane and no hederagenin and oleanolic acid 3-O-glucosides were available to find out the Rf value of these monoglucosides to compare which spot represents the 3-O-glucoside. Stainings with 10% sulphuric acid, which gave different color reactions for both aglycones, the hederagenin stock solution turned out to be not 100% pure and contained minor amounts of oleanolic acid. Two other compounds with Rf 0.73 and Rf 0.28 also appeared in a control lane with 80% alcohol and are apparently glycosides of ethanol or impurities thereof. However, the extra spots presumably derived from ethanol and impurities of hederagenin with oleanolic acid (and perhaps vice versa) do not account for all the spots shown in the lanes in figure 13.

In contrast, LCMS analysis of the activity assays revealed only one extra peak for each monoglucoside. Lacking (synthetically produced) 3-O-monoglucosides as reference for LCMS, we assumed that the extra peak in LCMS analysis corresponds with the most intense spots on TLC analysis. The fragments found in activity assays represented the oleanolic acid and
Conclusions and discussion

Hederagenin monoglucosides and the corresponding aglycones. Thus, UGT73C10, UGT73C11, UGT73C12 and UGT73C13 can utilize oleanolic acid and hederagenin as a substrate to produce oleanolic acid and hederagenin monoglucosides. TLC and LCMS data however do not provide any information about the position or conformation of the sugar moiety.

The UGT73C subfamily belongs to superfamily 1, a class of inverting enzymes (figure 5). This means that the sugar moiety of UDP-glucose, which is in α-conformation, will be inverted to a β-conformation glucose moiety when attached to the aglycone. The NMR analysis of a hederagenin-monoglucoside produced with UGT73C11 confirmed that the sugar moiety in that monoglucoside was in β-conformation, and we have no reason to believe that the highly similar UGT73C10, UGT73C12 and UGT73C13 do not share this feature.

The classification of the UGT73C does not make predictions about the position of glycosylation or substrates specificity. However, B.v.variegata UGT73C_{Shi} was shown to glucosylate oleanolic acid and hederagenin at the C3 position, and the high similarity (over 90% identity on amino acid level) between UGT73C_{Shi} and the B.v.arcuata UGTs suggested that the latter may also glycosylate the C3 position of sapogenins. This was supported by the high activity of UGT73C10 and UGT73C11 towards β-amyrin, a sapogenin with only one OH group, located at the C3-position. UGTs can be rather promiscous towards their substrate specificity and are usually rather regiospecific than substrate specific (Vogt and Jones, 2000). Thus, an enzyme which utilized β-amyrin is like to be able to utilize the 3-OH group on oleanolic acid, hederagenin and other sapogenins. The NMR analysis revealed that the hederagenin monoglucoside produced by UGT73C11 is indeed a 3-O-glucoside. Considering the regiospecificity and the high similarity to UGT73C_{Shi} from B.v. variegata, I would expect the oleanolic acid monoglucoside to be a 3-O-glucoside as well.

Moreover, TLC and LCMS analysis of the oleanolic acid and hederagenin monoglucosides produced by UGT73C10, UGT73C11, UGT73C12 and UGT73C13 could not detect any differences in the monoglucosides produced by UGT73C11 and the other three UGTS. This indicates that also the oleanolic acid and hederagenin monoglucosides produced by these UGTs are 3-O-monoglucoside.
UGT73C9 does not show any activity towards oleanolic acid and hederagenin. The Western blot in figure 12 revealed the presence of the enzyme in the crude extract and the activity towards TCP in figure 14 shows that at least a percentage of the enzyme is present in an active form. However, it is not easy to determine if the lack of oleanolic acid and hederagenin monoglucosides is due to the low concentration of (active) UGT73C9 in the crude extract, a low affinity for these substrates or both of the above mentioned reasons. On the other hand, the affinity of UGT73C10, UGT73C11, UGT73C12 and UGT73C13 towards oleanolic acid and hederagenin was either similar or higher than the affinity for TCP. Thus, if the lack of detection of activity of UGT73C9 would be only due to its low concentration in the crude extract, we should have detected some oleanolic acid and hederagenin monoglucosides TLC analyses of UGT73C9 activity assays, as we can see the TCP monoglucoside (figure 14).

The aberrant activity of UGT73C9 lies within the little amino acid substitutions between UGT73C9 and e.g. UGT73C10 (95 % identity). That makes these enzymes very suitable for further studies to unravel the mechanism of enzyme activity. Homology studies with e.g. the crystallized UGT71G1, which is also known to glycosylate flavonols and sapogenins, revealed that a few of the amino acid substitutions that were unique for UGT73C9 were located within the active site and therefore may be involved in the enzyme-substrate binding. However, further studies (e.g. site directed mutagenesis) are beyond the scope and time limits of this master thesis.

**ii How specific are these UGTs towards oleanolic acid and hederagenin as acceptor substrate?**

The most common way to study the acceptor specificity of an enzyme is by determining kinetic parameters like $K_M$, $V_{max}$ and $k_{cat}$. Several conditions are required to perform kinetics. For example, the enzyme concentration is required to determine some kinetic parameters (e.g. the $k_{cat}$) and to optimize the activity assay conditions prior to kinetics. Furthermore the reaction should only one end product, and ideally no reverse reaction should occur. The activity assays of the UGTs studied in this thesis did not seem to meet these requirements, as the TLC analysis of activity assays of hederagenin and oleanolic acid always showed more than one product, and more importantly because the methods to
Conclusions and discussion

determine the UGT concentration in the crude extract were either not accurate enough or too time consuming for this purpose. Therefore it was decided not to pursue kinetics but to study the enzyme specificity by TLC analysis instead.

No difference was found within the activity of UGT73C11 towards oleanolic acid and hederagenin in 10μM, 20μM, 50μM and 100μM concentrations. For the flavonol substrates, however, the difference of end product resulting from activity assays is much lower in a 10μM substrate concentration assay than 100μM. This indicated that the K_M of UGT73C enzymes towards both flavonols is higher than 10μM, whereas the K_M towards flavonols could be below 10μM. However, the intensity of the spots in figure 14 with figure 19 and table 6 may be well comparable as these activity assays as the crude extract and buffer pH differs from the assays in figure 19.

The intensity of the main spots was shown in table 6. The spots on the TLC were manually selected and the average intensity of each region was calculated by ImageQuant 5.0 and shown in the table below. Thus, these numbers have no value itself but represent the intensity of the spots on the TLC plates. The background value of each TLC plate was between 5 and 7, and was comparable among TLC plates. The size of the selected region of the homogeneous background had no influence on the average value of that region, but the way of selecting the spot as region had a large influence on the average value. A ‘too large’ spot, including background, would decrease the average value and a ‘too small’ spot would overestimate the average value of a spot. However, the ratios between the spots would be maintained as long as the way of selecting the spot regions is the kept constant.

The relative activity to TCP is shown in table. To compare the ratio between enzymes, the activity towards TCP should be equal for all enzymes under these conditions. No apparent relation was found between the estimated UGT concentration in the crude extract on SDS-PAGE and the amount of TCP end products detected in TLC analysis. This means that either the affinity for TCP differs among enzymes under these substrate concentrations, or that the percentage of active enzyme differs. It also means that in the ratio’s can only be compared within enzymes and not among enzymes.
UGT73C10 and UGT73C11 seem to have a higher efficiency towards the sapogenin aglycones than to any of the other tested substrates on 10μM concentration. UGT73C10 and UGT73C11 produced 1-2 times as much monoglucoside of oleanolic acid (10μM) than they do using (100μM) TCP as substrate. Their activity on flavonols and sterols is low, generally 1-10% of their activity towards TCP. Not that the concentration of the TCP in these assays is 10 times higher than the other aglycones and TCP serves purely as a positive control.

UGT73C12 and UGT73C13 seem to be both less efficient and less specific towards oleanolic acid and hederagenin. The spot intensity of (table 6) is generally 5x lower than that of UGT73C10 and UGT73C11. However spot intensity could be influences by other factors like a difference in the active enzyme in the crude extract or even a difference in the optimal conditions between these enzymes, as those have only been tested for UGT73C11. However, the ratio saponin:flavonol and sterols is relatively low in UGT73C12 and UGT73C13 compared to UGT73C10 and UGT73C11. The activity of UGT73C12 and UGT73C13 on oleanolic acid and hederagenin is 30% of the activity on TCP and the activity on the flavonols is also 5-10%. The difference in efficiency on these compounds is relatively small (compared to UGT73C10 and UGT73C11). Thus, despite possible differences in the UGT concentration in the assays etc. we can still conclude that UGT7312 and UGT73C13 are less specific for oleanolic acid and hederagenin than UGT73C10 and UGT73C11 are for these substrates.

The pattern of activity of UGT73C5 towards saponins, flavonols and sterols is similar to that of UGT73C12 and UGT73C13, but the enzyme might be even less specific for these compounds. This is not surprising, as literature already mentions that UGT73C5 utilizes wide spectrum of triterpene structures, such as cytokinins and, brassinosteroids. It would be interesting to see if UGT73C12 and UGT73C13 are also capable of utilizing those aglycones.

Interestingly, within the sapogenins UGT73C10 and UGT73C11 appear to glycosylate oleanolic acid, hederagenin and β-amyrin evenly well. As β-amyrin is the precursor of both oleanolic acid and hederagenin, it would be inconvenient to glycosylate β-amyryin instead of the aglycones. Not much is known about the function of a β-amyrin 3-O-glycoside. However, evidence is accumulating that metabolic pathways in the plant are highly channeled.
Conclusions and discussion

(reviewed in Jørgensen, 2005) and therefore UGT73C10 and UGT73C11 may not interact physically with β-amyrin.

The high specificity of UGT73C10 and UGT73C11 for sapogenins indicates that they are involved in the saponin pathway. However, in vitro assays do not necessarily describe the in vivo function. The expression in planta, interaction with other genes, availability of the substrate and many more factors all contribute to the function of a gene in the plant. Blast of the sequences against pyrosequencing database showed that all four genes are transcribed but not much have been tested for furthermore.

Stronger evidence could be by gained by knocking down the UGT of interest, using e.g. RNAi to target all copies of the gene and it is possible to knock-down (preferably out) both orthologs and paralogs at once. In combination with a line of over expression of these UGTS would provide more insight in the function of the UGTs in the biosynthetic pathway of saponins. This would, however, require that the plant can be transformed and that has not been done yet for this plant species.

Kinetics
Quantification of the end products of activity assays would be much easier to interpret when the amount of enzyme in the activity assay would be constant among the assays. In this thesis, different approaches ways have been explored to measure the UGT concentration in the crude extract. This was based on the assumption that UGTs partially lose activity during the process of purification, which itself was based on observations of J.M. Augustin during pilot experiments. However, no obvious loss of activity detected in figure 22 for both paralog (UGT73C10) and ortholog (UGT73C12). Crucial is the comparison of both UGT concentrations before and after purification in this case, which could have been estimated by SDS-PAGE analysis.

Whether or not the assumption about loss of activity after purification is justified, in this thesis I preferred to work with fresh crude extract. In particular because of the loss of activity of UGT73C13 after storage at -20°C, and purification every time to have freshly purified enzymes (UGT73C12 and UGT73C13) is time consuming. The method which makes
use of purified UGT to make standard curve using Western blot (see results) seemed promising. However, Western Blotting is a very time consuming technique and even though the *B. vulgaris* UGTs seem to be stable on ice (figure), a faster method is preferred to determine the amount of UGT in a crude extract. A well-established method is using an S-tag fusion of the UGTs for quantification. S-tag assays are a labor intensive but sensitive measurement, but the assays only take on hour, making them suitable for both a repetition of the specificity assays and to continue kinetics.

*i is there a difference in activity towards oleanolic acid and hederagenin between the UGTs isolated from the P and the G type plant?*

This question was posed with respect to the resistance of the *B. v. arcuata* towards the flea beetles. The G type of *B. v. arcuata* is resistant towards the flea beetles, whereas the P type is not. A difference in functionality between enzymes derived from the P type and G type could contribute a difference in the presence of the saponins and therefore the resistance of *B. v. arcuata* towards flees beetles. Especially the activity of hederagenin cellobioside, because that is the most abundant and most biologically active compound (Nielsen et al., 2010).

The data show only little, if any, difference between UGT73C10 and UGT73C11, the two UGTs that are most likely involved in the saponin biosynthesis pathway. The intensity of the spots is comparable for the activity on oleanolic acid (respectively 480 and 505 for UGT7310 and UGT73C11) but UGT73C11, derived from the G plant, seems to show a slightly higher efficiency on hederagenin than UGT73C10 (intensity of respectively 893 and 713 for UGT73C11 and UGT73C10). However, this difference would definitely not explain the difference in the presence of hederagenin monoglucoside in G and the P plant, as UGT73C10 is still very efficient with hederagenin.

Using protein extracts of G and P plants instead of crude extracts for activity assays, both G and P protein extracts were able to glycosylate hederagenin and oleanolic acid (Augustin, unpublished). The cellobioside, which is shown to be bio-active against flea beetles, was found in both the G and P type plants. Mapping of the UGT73C members of *B. v. arcuata* and
Conclusions and discussion

*variegata* revealed that the UGTs are not located within the QTL for resistance (Kuzina et al., 2009). These results indicate that the difference in production of saponins in G and P is due to a difference earlier in the saponin biosynthesis pathway. The higher functionality of UGT73C11 (if significant at all) than indicates that UGT73C11 is under a selection pressure towards efficiency of utilizing hederagenin to produce cellobiosides, whereas UGT73C10 is not.

iv) *Is there a difference in activity between the orthologs (of UGT73C_{Shi} in B.v.variegata) and their paralogs?*

The genes of *Barbarea vulgaris* var. *variegata* and both the G and P type of *Barbarea vulgaris* ssp *arcuata* are homologs, meaning that they have a common ancestor. Within homology there are separate terms for genes were sequence divergence follows speciation (orthologs) or gene duplication (paralogs) after speciation (Fitch, 2000). The most recent common ancestor of the taxa is called the cenancestor. Assuming that the separation between species and subspecies or variants is an earlier event than separation of one subspecies into types, the UGTs derived from the G and the P type plant would have a more recent cenancestor than the UGT from the two subspecies of *Barbarea vulgaris*.

However, the high sequence identity suggest that UGT73C_{Shi} may be extremely close related to UGT73C11 or that it even may be considered as the same gene. The difference between the varient *variegata* seems to be smaller than the difference between the G and P type, which is unlikely to be explained either by very high selection pressure of the beetles. First of all the beetles selection pressure has not been validated and second a high number of silent mutations would have been present in both sequences. It is more likely that the current classification of the P and G type and variant variegata do not reflect their genetic distances. The origin of *B.v. variegata* is referred to as a variant, and is probably a cultivar of *B.vulgaris*. The high % identity suggests it may well be a cultivar of the G type of *B.v.arcuata*. In contrast, recent microsatellite analysis of G and P plants and different subspecies of *Barbarea vulgaris* suggests that the P and G type plants should be classified as different subspecies or even species, rather than two types within one subspecies (Hauser et al, unpublished).
The high similarity of $\text{UGT73C}_{\text{shi}}$, $\text{UGT73C11}$ and $\text{UGT73C10}$ suggest that these genes are the orthologs in $B. v. \text{variegata}$, the G and P plant respectively. They most likely are derived from one gene in the cenancestor of $B. v. \text{arcuata}$ and $B. v. \text{variegata}$. Within the subspecies $\text{arcuata}$, $\text{UGT73C9}$ and $\text{UGT73C10}$ in the P type are paralogs of each other, in which $\text{UGT73C10}$ is the functional ortholog of $\text{UGT73C11}$ from the G type. $\text{UGT73C12}$ and $\text{UGT73C13}$ are paralogs to $\text{UGT73C10}$, but to be able to say whether they are orthologs or paralogs towards $\text{UGT73C}_{\text{shi}}$, in the way we used the terms in this thesis, information about the event of gene duplication in $\text{variegata}$ is required. The $\text{UGT73C12}$ and $\text{UGT73C13}$ have not been found in $\text{variegata}$ because we haven’t looked there yet. However, if $\text{variegata}$ is indeed a cultivar of the G type, it will most likely also contain an ortholog of $\text{UGT73C12}$ and $\text{UGT73C13}$. Therefore, and also to facilitate readability, in this thesis $\text{UGT73C12}$ and $\text{UGT73C13}$ are referred to as paralogs to $\text{UGT73C}_{\text{shi}}$, even though strictly spoken we need to search for $\text{UGT73C12}$ orthologs in $\text{variegata}$ to confirm these relations.

The gene duplication that led to the presence of both $\text{UGT73C9}$ and $\text{UGT73C10}$ may have occurred after separation into the P and G type plant, as no UGT has been isolated in the G-plant that is closest to $\text{UGT73C9}$. However, inconsistencies with the sequence of the PCR products of the $\text{UGT73C11}$ sequence in the G plant indicate that there are more copies of $\text{UGT73C11}$ present in the G-type plant (Augustin, personal communication). Because all inconsistencies were silent mutations, this extra copy (or copies) was never isolated.

So far we have seen that the plants contain orthologs and paralogs of the UGT from $B. v. \text{variegata}$, and the research question was if, and how, these paralogs and orthologs differ in activity and biochemical properties. Most of the experiments regarding optimal conditions for activity assays were conducted with $\text{UGT73C11}$ only, under the assumption that all enzymes would respond the same on the Tris buffer concentration, pH, ect. However, we did find a difference in activity after freeze-thawing between the ortholog $\text{UGT73C11}$ and paralog $\text{UGT73C13}$. $\text{UGT73C13}$ lost partially activity after storage at -20°C, while this effect was not detectable for $\text{UGT73C11}$. Moreover a difference in activity was found between the orthologs and the paralogs. The genes that are the orthologs to $B. v. \text{variegata}$ $\text{UGT73C}_{\text{shi}}$ had a higher specific towards sapogenin substrates than the paralogs.
Conclusions and discussion

Because of the terms orthologs, paralogs, and the high efficiency of the orthologs for oleanolic acid and hederagenin it is tempting to say that the orthologs are the ‘original’ genes from which by an event of duplication the paralogs are derived, which than by lack of selection pressure lost their function and became less specific for the saponin biosynthesis pathway. However, the term orthologs and paralogs depend on the point of reference, which in this case is the UGT from *B.v.variegata*. This point of reference was chosen from a biochemical point of view, to elucidate the saponin biosynthesis pathway, but has no meaning from an evolutionary point of view. Without a reference point the P and G plant contain (at least) two copies of a gene and the question arises which is the copy of which.

Secondary metabolites are often derived from primary metabolites; copies of genes in the primary metabolite pathway could get mutations and might change function, whereby the newly produced metabolite ‘accidentally’ provides a beneficial effect on the interaction with the plant and its environment. **UGT73C12** and **UGT73C13**, orthologs to one another, show in their glucosylation pattern a great similarity with the *A.thaliana* UGT73C5, which is a UGT involved in the regulation of Brassinolide, a plant hormone involved in growth and development (Poppenberger et al., 2005). Thus, perhaps **UGT73C12** and **UGT73C12** are orthologs to *A.thaliana* UGT73C5 and shared a common ancestor further back in the Brassicacea that possibly had a function in hormone regulation. Than **UGT73C10** and **UGT73C11** are the result of a duplication of **UGT73C12** and gained a new function in the metabolic pathway of saponins, rather than that **UGT73C10** lost their function. This theory is, however, purely speculative. However, the thought that the characterization of these genes could bring us closer to the origin of resistance of *Barbarea vulgaris* against flee beetles is fascinating.
5. Perspectives

This thesis describes the characterization of what could be the first characterized gene involved in the biosynthetic genes in the pathway of *Barbarea vulgaris* ssp *arcuata*. However, the experiments are purely in vitro experiments and therefore provide only information about the capacity of the UGTs under lab conditions.

Within the in vitro study, there are some gaps in the data. The data regarding optimization of the activity assays should be interpreted with care, as both repetitions is lacking and the range of the conditions is limited. For example, the pH should be tested over a larger range and be repeated just to make sure that pH 7.5 was not just an accidental high value. Moreover, the activity might changes with the buffer composition and optimum pH can change among buffers (Hou et al., 2004). Moreover, the variation among assays is not tested here. The intensity of the spots is influence by the duration time of an assay, the percentage active enzyme present but also by the precision of the extraction. To limit the latter variation, the specificity assays were always extracted twice.

The results obtained from TLC assays do suggest differences in specificity among the enzymes. However, kinetic parameters are still necessary because they are more exact, provide more detailed information and allow comparison with other publications. To perform these kinetic assays, an S-tag fusion of the proteins would be a suitable approach, because the methods explored in this thesis are very time consuming and might have a high inaccuracy.

Interesting point would be to test the UGTs for more substrates. Other saponins, soybeansaponin and medicagenin would give insight in the regiospecificity and substrate specificity, and brassinosteroids substrates would be interesting to test possible similarities in function towards UGT73C5. Since the NMR we now have purified hederagenin monoglucoside available. It would be interesting to test the four recombinant UGTs on their activity on this monoglucoside, as in fig 12 spots appear on Rf ≈0.3 which could be hederagenin and oleanolic diglicosides but which were not reproducible within this thesis. Using the purified monoglucoside the C3 of hederagenin is no longer avaiable and the
activity of the UGT for other positions can be tested. Moreover, the monoglucoside can be used to test candidate UGTs for the second step of glycosylation yielding the cellobiosides. Identification of these UGTs would be very interesting, as it would be the first enzyme known to transfer a sugar onto a monoglycoside.

The large-scale activity assay unfortunately yielded in a low amount of purified monoglucosides. However, the monoglucoside would be very interesting to test for its biological activity towards flea beetles. Earlier bioassays with flea beetles revealed bioactivity for the cellobioside but no bioactivity for the monoglucoside. As the resistance of flea beetles might be due to a β-glucosidase, it would be really interesting to see if the monoglucoside has bioactivity towards the common type of susceptible flea beetles.

To get more insight in the phylogenetic relations between the subspecies and types within *Barbarea vulgaris*, one should screen *B.v.*variegata for orthologs of e.g. UGT73C12 or UGT73C13 and UGT73C9. Those data would provide more insight in the events of gene duplication among these subspecies and types and contribute to the (currently ongoing) revision of the status of the P and G type and *B.v.*variegata.

The in vitro studies performed in this thesis do not prove the involvement of the UGTs in the biosynthetic pathway of saponins. The role of the UGTs in the saponin pathway can be studied by studying plants over expression or lacking expression of the UGT. RNAi could be a suitable method to knock-down the UGTs as it can target all five UGTs in the plant. To produce a real knock out, e.g. T-DNA, may be not feasible to produce as the amount of copies of these UGTs are unknown, but are at least two or three copies per plant. Moreover, the copies map together and may be very difficult to cross into each other to generate a knock-out line. However, this RNAI technique requires transformation of the plant and that has not been done yet for *B. vulgaris*.

The biological function of a UGT in the plant is influenced by many factors, e.g. the expression level of the gene, regulation, localization and substrate availability. The expression levels between the four five genes can be studied with RT-PCR. Localization studies (roughly using RT-PCR in different plant organs or more detailed e.g. using GUS-
staining) could provide more insight into the localization and therefore in the function of these proteins. In addition, it would be interesting to study which factors influence gene expression in *Barbarea*. Often UGT expression in the plant can be induced by Methyl Jasmonate, a plant hormone that is released with plant stress, caused by cell damage (Osbourn, 2003). As the cellobiosides have a function of insect resistance, it would be interesting to see if (imitation of) insect damage increases genes involved in the saponin pathway.

Moreover, the resistance of *Babarea vulgaris* towards the flea beetles declines in the autumn (Agerbirk et al., 2001) which seems to be due to a decline of saponin content in the same period (Nielsen, personal communication). In addition, resistance of the plants has been shown to be inducible with light and temperature conditions. It would be very interesting to study how the pathway is regulated and which genes are inducible by which factors. However, the saponin pathway may be regulated at a different place in the pathway, for instance at the OSC or p450 level. The other way around, studying gene expression before and after induction of saponins using microarray can be used to search for candidate genes.

By date the P and G type plant have been sequenced using Pyrosequencing and candidate OSCs, p450s, and UGTs involved in the second glycosylation step, are being selected. Candidate OSC genes will be screened by heterologous expression in a lanosterol sterol deficient yeast strain, as used in (Corey, 1993).

The resistance of *B. vulgaris* towards the flea beetles is conferred by the two saponin cellobiosides in which hederagenin cellobioside is the most bioactive and most abundant compound (Nielsen et al., 2010). Although the precise mechanism is not completely elucidated yet, studies agree on complex formation between saponins and the sterols in the membrane of the pest or pathogen. This mode of resistance is a general mechanism it can confer resistance against many other insect species. Therefore genes in this pathway could be used for crop engineering, once the pathway is elucidated. *Barbarea vulgaris* is closely related to *Brassica napus* (rape seed) and moreover the Brassica family contains many eatable crop species.
Crystal structures of UGTs became available to gain more insight into the precise mechanism of glucosylation and with directed mutagenesis the role of individual amino acids in donor and acceptor binding have been elucidated. The knowledge about donor and acceptor substrate binding in UGTs is expanding rapidly and will be used soon to modify UGTs to exactly the function that you want them to have. UGT73C9 and UGT73C10 would be an interesting pair of UGTs to include in this research field. The amino acid sequence is highly similar (95% identity), yet UGT73C9 has an altered (solubility and) activity on hederagenin and oleanolic acid. Homology studies of revealed that UGT of amino acids located in the activity site, which could have an effect on the activity of UGT73C9. For the altered solubility, other amino acids might be involved. Because of their many medicinal or soap like, antiflammatory and other abilities, saponins are of great interest for the pharmaceutical purposed, thus so are the GTs producing them. Using glycosyltransferases to produce monoglucosides is relatively easy compared to chemical synthesis. A better understanding of the mechanism behind acceptor-binding lead to efficient enhancement of UGT and the ability to ‘design’ UGTs for any substrate could be in the near future.
6. References


References


Nielsen, J. K. 1997. Variation in defences of the plant Barbarea vulgaris and in counteradaptations by the flea beetle Phyllotreta nemorum. Entomologia Experimentalis Et Applicata, 82, 25-35.


Alignment of the UGT73C subfamily

Appendix
Appendix

UGT73C11 AGGC--TCA- ---------- ----------/TCTCA TCTCA-ATAT CACATCCCTG CTAGAAAGCA TAATGCAA
UGT73C12 AGGC--TCA- ---------- ----------/TCTCA TCTCA-ATAT CACATCCCTT CTAGAAAGCA TAATGCAA
UGT73C9 AGGC--TCA- ---------- ----------/TCTCA TCTCA-ATAT CACATCCCTG CTAGAAAGCA TAATGCAA
UGT73C8 TGGA--TCT- ---------- ---------- -----TCT CA CTCTA-ATAT TTCTTTGTTC ATCCAAGATA TCATG-AA
UGT73Cshi AGGC--TCA- ---------- ---------- -----TCTCA TCTCA-ATAT CTCTTGTTC CTAGAAAGCA TAATGCAA

UGT71G1 GAATAAAAGT TGAATGTGGA ATTAGTAGCA TATTTGTG TA TAGCAAATTT AATCAAGCTA GCACATGTGC CTATCTTT
UGT73C1 TTAGAACAAC CCAA----GA AATTGATTGTA ACCTTTCT T- ---------- -------AGA TTG--TAA-- ------CT
UGT73C2 GAATTTAAT CCAA----GA CTCTA-ATAT TGACACAGA TTTAGTTAAT TAATGCAAGA TCCAGAAGCA TAATGCAA
UGT73C3 TTAGAACAAT CCAA----GA ATTGAGTATA TGTCATCTA T--TTAAAGG ATTTAAAAA TTAAATAG-- ------CT
UGT73C4 CAATGACAAG CCAA----GA ATCAATTTTA AATTTAACAT TAATGCAAGA TCCAGAAGCA TAATGCAA
UGT73C5 TTTTATTTCA GTACTGCTTT TCTTTGGAGG GTTGTTTA TA TAATATTTTT TTATTAACAG CTGAACTAAT GAAGT---
UGT73C6 TTTCTTTTGT ATTACCCAAA AAAAATGATT GTGACGTTG TTCT------ ---------- ---------- --------
UGT73C7 TTTGATTTCT ATATTGGAGA AATTTAAATC AGAGCCTT TG TTAAACACGT GGATAATGAA TCAGAAGAAG ATAAATCA
UGT73C8 GAAAAATAA- ---A----GA TATGATGTCA T------- -- ---------- ---------- ------CA

UGT71G1 TTTATTTTCA GTACTGCTTT TCTTGGAGG GTTGTTTA TA TAATATTTTT TTATTAACAG CTGAACTAAT GAAGT---
UGT73C1 TTTTATTTTGT ATTACCCAAA AAAAATGATT GTGACGTTG TTCT------ ---------- ---------- --------
UGT73C2 TTTTATTTTGT ATTACCCAAA AAAAATGATT GTGACGTTG TTCT------ ---------- ---------- --------
UGT73C3 TTTTATTTTGT ATTACCCAAA AAAAATGATT GTGACGTTG TTCT------ ---------- ---------- --------
UGT73C4 TTTTATTTTGT ATTACCCAAA AAAAATGATT GTGACGTTG TTCT------ ---------- ---------- --------
UGT73C5 TTTTATTTTGT ATTACCCAAA AAAAATGATT GTGACGTTG TTCT------ ---------- ---------- --------
UGT73C6 TTTTATTTTGT ATTACCCAAA AAAAATGATT GTGACGTTG TTCT------ ---------- ---------- --------
UGT73C7 TTTTATTTTGT ATTACCCAAA AAAAATGATT GTGACGTTG TTCT------ ---------- ---------- --------
UGT73C8 TTTTATTTTGT ATTACCCAAA AAAAATGATT GTGACGTTG TTCT------ ---------- ---------- --------
UGT73C9 TTTTATTTTGT ATTACCCAAA AAAAATGATT GTGACGTTG TTCT------ ---------- ---------- --------
UGT73C10 TTTTATTTTGT ATTACCCAAA AAAAATGATT GTGACGTTG TTCT------ ---------- ---------- --------
UGT73C11 TTTTATTTTGT ATTACCCAAA AAAAATGATT GTGACGTTG TTCT------ ---------- ---------- --------
UGT73C12 TTTTATTTTGT ATTACCCAAA AAAAATGATT GTGACGTTG TTCT------ ---------- ---------- --------
UGT73C13 TTTTATTTTGT ATTACCCAAA AAAAATGATT GTGACGTTG TTCT------ ---------- ---------- --------
UGT73C14 TTTTATTTTGT ATTACCCAAA AAAAATGATT GTGACGTTG TTCT------ ---------- ---------- --------
UGT73C15 TTTTATTTTGT ATTACCCAAA AAAAATGATT GTGACGTTG TTCT------ ---------- ---------- --------