Characterization of sulfotransferase proteins from *Arabidopsis thaliana*

Von der Naturwissenschaftlichen Fakultät

der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades

Doktorin der Naturwissenschaften

Dr. rer. nat.

genehmigte

Dissertation

von

Dipl.-Biol. Marion Klein

geboren am 07.06.1972 in Rinteln

2007

Referentin: Prof. Dr. Jutta Papenbrock

Korreferent: Prof. Dr. Bernd Huchzermeyer

Tag der Promotion: 21.12.2006

Summary

The aim of this study is the characterization of sulfotransferases (SOTs) from *Arabidopsis thaliana* (L.) Heynh. SOTs are a group of enzymes widely distributed in eubacteria and eukaryotes, but not in archaea. These enzymes catalyze the transfer of a sulfate group from the co-substrate 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to a hydroxyl group of various substrates with parallel formation of 3'-phosphoadenosine 5'-phosphate (PAP).

The fully sequenced genome of *Arabidopsis* was searched for SOT sequences with the already isolated SOT RaR047 protein sequence from *Arabidopsis* using the BLAST program. Eighteen genes encoding SOTs (AtSOT1 to 18) have been identified and were divided into seven groups according to their sequence similarities. To date, the SOT family of *Arabidopsis* has to be enlarged by four proteins. For most of the SOTs the respective substrate specificities and therefore their functions in the organism are not known yet.

In this study, group VII SOTs (AtSOT16 to 18) were investigated in detail. This group was identified as desulfo-glucosinolate specific SOTs (ds-Gl SOTs). Two strategies for this identification as ds-Gls SOTs were applied: combination of transcriptomics and metabolomics on the one hand, as well as sequence comparison and the use of isolated ds-Gls as substrates in *in vitro* enzyme assays on the other hand.

Gls are secondary metabolites found in the order Capparales, including agriculturally important crop plants of the Brassicaceae family, such as oilseed rape (*Brassica napus*) and the model plant *Arabidopsis*. In different *Arabidopsis* ecotypes altogether about 30 Gls were identified. Upon tissue damage, Gls are hydrolysed by myrosinase, resulting in degradation products which show a wide range of biological functions. These hydrolysis products are involved in plant defence against pathogens and herbivores and can act anti-carcinogen. The SOTs of group VII (AtSOT16 to 18) are involved in the modulation of the Gl pattern of *Arabidopsis*, as they catalyze the last step in the biosynthesis of the Gls core structure: due to sulfation of ds-Gls “parent”-Gls are produced. Subsequent secondary modifications lead to the great variety of Gls found *in vivo*.

To investigate the gene expression of AtSOT16, 17 and 18, corresponding mRNA levels were determined by Northern blot analysis. This investigation revealed a slightly differential expression of the three *AtSOT* genes analyzed. Dependency on developmental stages, light conditions, different plant organs and sulfate nutrition were observed. Expression studies with high (500 and 1500 µM) and low (50 and 30 µM) sulfate

concentrations revealed at sulfate deficiency (30 µM) increased mRNA levels of the three *AtSOT* genes.

All three ds-Gl AtSOT proteins are localized in the cytoplasm, as could be shown by fusion and expression of each protein with the green fluorescent protein (GFP) in transiently transformed *Arabidopsis* protoplasts.

To determine the specific activities of the three AtSOT proteins from ecotype C24 as well as AtSOT18* from ecotype Col-0, they were overexpressed in *Escherichia coli* and purified. The recombinant proteins were used for *in vitro* enzyme measurements. The products of *in vitro* measurements were analyzed by HPLC.

Isolated ds-Gls from different species and a mixture of parent ds-Gls identified in *Arabidopsis* were used as substrates in *in vitro* measurements. Kinectic studies and determination of substrate specificities indicate a differing behaviour of the three AtSOT enzymes.

Although both AtSOT18 and AtSOT18* proteins differ only in two out of 350 amino acids, they vary in their kinetic behaviour and in substrate specificity. Replacement of one of the two amino acids of AtSOT18 from ecotype C24 with the respective amino acid of AtSOT18* from Col-0 leads to a substrate specificity comparable with the wild-type AtSOT18* protein.

The aim of this study is to elucidate the function of ds-Gl SOTs *in planta.* Thus, the initial *in vitro* measurements were enhanced. A partially artificial, partially natural test was designed. Same enzyme assays were done with an *in vivo* extract from *Arabidopsis* leaves used as substrate instead of isolated ds-Gls. In conclusion, substrate specificities found in *in vitro* measurements could be confirmed in this partially natural system.

The characterization of mutants could be helpful to elucidate the *in vivo* function of the ds-Gl SOT proteins. Selective knock-out of each of the ds-Gl SOT proteins in *Arabidopsis* could clarify in which manner the Gl pattern is influenced.

Keywords:

Arabidopsis thaliana, function analysis, GFP, glucosinolate, sulfotransferases

Zusammenfassung

Ziel dieser Arbeit ist die Charakterisierung von Sulfotransferasen (SOT) aus *Arabidopsis thaliana* (L.) Heynh. Es handelt sich hier um eine weitverbreitete Proteinfamilie, die in Eukaryoten und Eubakterien, aber nicht in Archaea vorkommt. Diese Enzyme katalysieren den Transfer einer Sulfatgruppe vom Cosubstrat 3'-Phosphoadenosin 5'-Phosphosulfat (PAPS) auf eine Hydroxylgruppe verschiedener Substrate mit einer gleichzeitigen Bildung von 3'-Phosphoadenosin 5'-Phosphat (PAP).

Das vollständig sequenzierte Genom von *Arabidopsis* wurde unter Verwendung von BLAST und einer bereits isolierten SOT-Proteinsequenz (RaR047) aus *Arabidopsis* mit dem Ziel durchsucht, weitere Sequenzen für SOT-Proteine zu identifizieren. Die SOT-Familie aus *Arabidopsis* setzt sich aus 18 potentiellen Mitgliedern (AtSOT1 bis 18) zusammen, die sich nach ihrer Sequenzähnlichkeit in sieben Gruppen gliedern. Inzwischen konnte diese SOT-Familie um vier SOTs erweitert werden. Für nahezu alle SOT-Proteine aus *Arabidopsis* ist die jeweilige Substratspezifität und damit auch die Funktion innerhalb des Organismus noch nicht bekannt.

Diese Arbeit untersucht detailliert die Gruppe VII (AtSOT16 bis 18). Es handelt sich bei dieser Gruppe um Desulfo-Glucosinolat-Sulfotransferasen (ds-Gl-SOTs). Zwei Strategien zur Identifizierung von AtSOT16 bis 18 als ds-Gl-SOTs wurden angewandt: durch die Kombination von Transkriptom- und Metabolomanalyse einerseits sowie anderseits durch Sequenzvergleiche und die Verwendung von ds-Gle als Substrate in *in vitro*-Enzymtests.

Gle sind sekundäre Metabolite, die vor allem in der Ordnung Capparales vorkommen. Diese Ordnung schließt landwirtschaftlich wichtige Nutzpflanzen der Brassicaceae, z. B. Raps (*Brassica napus)*, und die Modellpflanze *Arabidopsis* ein. In verschiedenen *Arabidopsis* Ökotypen wurden ungefähr 30 Gle identifiziert. Durch die Beschädigung des Planzengewebes werden Gle durch das Enzym Myrosinase hydrolysiert. Die entsprechenden Abbauprodukte weisen eine Vielzahl von biologischen Funktionen auf. Sie sind zum Beispiel in die Pflanzenabwehr gegen Pathogene und Herbivore involviert und wirken anti-carcinogen. Die SOT-Proteine der Gruppe VII (AtSOT16 bis 18) katalysieren den letzten Schritt in der Biosynthese der Gl-Kernstruktur und modulieren somit das Gl-Muster von *Arabidopsis*: durch die Sulfatisierung der ds-Gle entstehen die sogenannten „Eltern“-Gle. Nachfolgende sekundäre Modifikationen der „Eltern“-Gle führen zur Vielfältigkeit der *in vivo* vorgefundenen Gle.

Um die Genexpression von AtSOT16, 17 und 18 zu untersuchen wurden die entsprechenden mRNA-Gehalte durch Northern Blot-Analyse bestimmt. Bei den

Expressionsstudien zeigten sich geringe Unterschiede in der Expression der drei *AtSOT*-Gene. Es ergab sich eine Abhängigkeit der Expressionsstärke vom Alter der Pflanzen, von den Lichtverhältnissen, von den verschiedenen Pflanzenorganen und von der Sulfat-Ernährung. Bei Expressionsstudien mit hohen (500 and 1500 µM) und niedrigen (50 and 30 µM) Sulfat-Konzentrationen zeigte sich bei Sulfat-Mangel (30 µM) eine Hochregulierung der mRNA-Gehalte der drei *SOT*-Gene.

Alle drei ds-Gl AtSOT-Proteine sind im Cytoplasma lokalisiert, wie durch Fusion und Expression jedes Proteins mit dem grün fluoreszierenden Protein (GFP) in transient transformierten *Arabidopsis*-Protoplasten gezeigt werden konnte.

Zur Bestimmung der spezifischen Aktivität der drei AtSOT-Proteine des Ökotyps C24 sowie eines Proteins AtSOT18* des Ökotyps Col-0 wurden die Proteine in *Escherichia coli* überexpremiert und gereinigt. Die rekombinanten Proteine wurden für *in vitro*-Enzymtests verwendet. Die entstandenen Produkte der *in vitro*-Tests wurden mit Hilfe der HPLC analysiert.

Aus verschiedenen Pflanzenarten isolierte ds-Gle und eine Mixtur aus in *Arabidopsis* natürlich vorkommenden „Eltern"-Gle wurden als Substrate in *in vitro*-Messungen verwendet. Kinetische Studien und die Bestimmung von Substratspezifitäten zeigten ein unterschiedliches Verhalten der drei AtSOT-Enzyme auf.

Obwohl die beiden Protein-Varianten 18 und 18* sich nur in zwei von 350 Aminosäuren unterscheiden, variieren sie deutlich in ihrem kinetischen Verhalten und in ihrer Substratspezifität. Der Austausch einer dieser beiden Aminosäuren von AtSOT18 des Ökotyps C24 mit der entsprechenden Aminosäure von AtSOT18* von Col-0 führt dabei zu einer Substratspezifität vergleichbar mit dem Wildtyp Col-0.

Ein Ziel dieser Arbeit ist, die Funktion von ds-Gl-SOTs in der Pflanze herauszufinden. Hierzu wurden die anfänglichen *in vitro*-Tests zu einem teils künstlichen, teils natürlichen Ansatz weiterentwickelt. Dabei ersetzte ein *in vivo*-Blattextrakt die isolierten ds-Gle als Substrat in denselben Enzymtests. Zusammenfassend konnten die Ergebnisse aus den *in vitro*-Tests mit diesem partiell natürlichen System bestätigt werden.

Weitere Hinweise auf die Funktion der ds-Gl-SOT-Proteine könnte die Charakterisierung von Mutanten liefern. Durch gezieltes Ausschalten einzelner ds-Gl-SOT-Proteine in *Arabidopsis* könnte geklärt werden, in welcher Art und Weise das Gl-Muster beeinflußt wird.

Schlüsselwörter:

Arabidopsis thaliana, Funktionsanalyse, GFP, Glucosinolate, Sulfotransferasen

Abbreviations

aa	amino acid
ADP	adenosine diphosphate
AMP	adenosine monophosphate
APS	adenosine-5'-phosphosulfate
ATP	adenosine triphosphate
AtSOT	*Arabidopsis thaliana* sulfotransferase
BL-SOM	batch-learning-selforganizing mapping
BSA	bovine serum albumine
CDP	2-chlor-5-(4-methoyxspiro{1,2-dioxetan-3,2’-(5’-chlor) tricycle[3.3.1.13,7]decan}-4-yl-1-phenylphosphate
DIG	digoxigenin
DNA	deoxyribonucleic acid
dNTPs	deoxynucleotide triphosphates
ds	desulfo
DTT	1,4-dithiothreitol
EDTA	ethylenediaminetetraacetic acid
ER	endoplasmic reticulum
ESP	epithiospecifier protein
EST	expressed sequenced tags
GFP	green fluorescent protein
Gl	glucosinolate
IAA	indol-3-acetic acid
IAN	indolyl-3-acetonitrile
I3M	indol-3-yl-methyl glucosinolate
IPTG	isopropyl thio-β-D-galactoside
ITC	isothiocyanate
1MOI3M	1-methoxy-indol-3-yl-methyl glucosinolate

4MOI3M	4-methoxy-indol-3-yl-methyl glucosinolate
4MSOB	4-methylsulfinylbutyl glucosinolate
7MSOH	7-methylsulfinylheptyl glucosinolate
8MSOO	8-methylsulfinyloctyl glucosinolate
3MSOP	3-methylsulfinylpropyl glucosinolate
4MTB	4-methylthiobutyl glucosinolate
7MTH	7-methylthioheptyl glucosinolate
8MTO	8-methylthiooctyl glucosinolate
3MTP	3-methylthiopropyl glucosinolate
5MTP	5-methylthiopentyl glucosinolate
NBT	nitroblue tetrazolium
OD	optical density
ORF	open reading frame
PAP	3'-phosphoadenosine 5'-phosphate
PAPS	3'-phosphoadenosine 5'-phosphosulfate
PCR	polymerase chain reaction
2PE	2-phenylethyl glucosinolate
PPi	pyrophosphate
RNA	ribonucleic acid
S	sulfur
SO_4^{2-}	sulfate
SOT	sulfotransferase
STR	sulfurtransferase

Contents

Chapter 1:
General Introduction

Overview of glucosinolates

What are glucosinolates

Glucosinolates (Gls) are secondary metabolites occuring in the order Capparales, which includes the agriculturally important Brassicaceae family, containing the crop plant rape (*Brassica napus*), and vegetables like broccoli, cabbage, and horseradish and the model plant *Arabidopsis thaliana*. Upon tissue damage, Gls are hydrolysed by the enzyme myrosinase, resulting in many hydrolysis products, such as isothiocyanate, thiocyanate and nitriles. These hydrolysis products have a wide range of biological acitvities; the compounds have positive as well as negative effects. They are involved in plant defense against pathogens and herbivores, but also attract specialized herbivores. For humans, these degradation products can act cancer preventive and are responsible for a certain taste and flavour. On the other hand, specific hydrolysis products can be toxic. E. g., the presence of specific Gls and their degradation products like oxazolidine-2-thione, derived from 2-hydroxy-3-butenyl Gls in seeds of rape reduce the quality of seed meel used as animal food, causing goiter and having further harmful effects on animal nutrition (Griffiths *et al.*, 1998).

Glucosinolate biosynthesis

The biosynthesis of Gls can be divided into three steps and in short proceeds as follows: 1) a precursor amino acid, such as methionine or tryptophan, is elongated by one or several methylene groups 2) precursor amino acids are converted into parent-Gls 3) parent-Gls can be modified secondarily (Wittstock and Halkier, 2002). The sequence of the Gl biosynthesis is supported by *in vivo* and *in vitro* studies (Graser *et al.*, 2001; Kliebenstein *et al.*, 2001*b*).

Gls were grouped according to their chemical properties into aliphatic, aromatic and indol Gls depending on the amino acids they are derived from. Aliphatic precursor amino acids are methionine, alanine, valine, leucine, isoleucine; aromatic amino acids are tyrosine, phenylalanine or tryptophan for indolic Gls (Wittstock and Halkier, 2002). In *Arabidopsis*, methionine and tryptophan are major precursors of Gls.

In detail, the Gl biosynthesis proceeds as follows: the side chain of precursor amino acids may be elongated in one or several elongation cycles, catalyzed by methylthioalkylmalate synthases MAM, MAM-1 and MAM-L. After elongation, the precursor is converted to an aldoxime by a CYP79 enzyme, followed by an oxidation

reaction due to CYP83. Both CYP proteins belong to the cytochrome P450 protein family. The S-alkylthiohydroximate produced is converted to thiohydroximic acid due to C-S lyase. Then, glucose is added by S-glucosyltransferase, forming desulfo-Gl (ds-Gl). The last step in the biosynthesis of the Gl core structure (parent Gl) was suggested to be catalyzed by members of the sulfotransferase (SOT) protein family (EC 2.8.2.-) (Wittstock and Halkier, 2002). In a former study, a partially purified protein from *Lepidium sativum* was shown to have ds-Gl activity with dependence on the co-substrate 3'-phosphoadenosine 5'-phosphosulfate (PAPS) (Glending and Poulton,1990). Unfortunately, at that time no molecular data were available.

The biosynthesis of Gls is only partially investigated: it is assumed that both the glucosylation and the sulfation reactions are non-specific (Halkier, 1999). Therefore, we decided to investigate SOTs of *Arabidopsis* to find out more about the last step of the Gl core structure biosynthesis. We would like to know how many members exist in this protein family and how these enzymes are related. Finally, we would like to find out whether there is more than one SOT active with ds-Gls. In this case, it would be interesting to find out whether the assumption is correct that SOTs in Gl biosynthesis are non-specific (Halkier, 1999).

In order to answer these initial questions, a phylogenetic tree of the SOT protein family of *Arabidopsis* will be generated. The bioinformatical analysis may be helpful to get more information about this protein family.

To get a comprehensive idea of *Arabidopsis* SOTs, a broad spectrum of methods will be included in this study to answer further questions. We would like to know how many ds-Gl SOTs exist in *Arabidopsis*. If more than one ds-Gl SOT is found, it may be interesting to find out whether ds-Gl SOTs are specific for certain ds-Gls. Is it therefore possible to alter the Gl pattern in the plant due to modulation of ds-Gl SOT levels? Are SOTs increasingly expressed under certain conditions or are they expressed constitutively? Are there differences in Gl patterns in mutant analysis? In which cell compartments are SOTs localized? All these information should be helpful to characterize the functions of these enzymes *in vivo* and their regulation.

Glucosinolate diversity in *Arabidopsis*

At least 130 different Gls have been detected in plant species of the order Capparales (Fahey *et al.*, 2001; Mithen, 2001; Mikkelsen *et al.*, 2002). In *Arabidopsis*, about 30 Gls have been described (Hogge *et al.* 1988; Mithen, 2001; Reichelt *et al.*, 2002). This variety of Gls arises due to chain elongation of the precursor amino acids at stage one and secondary modifications of parent Gls at stage three.

The Gl content of various organs of *Arabidopsis* ecotype Col-0 was analyzed at different developmental stages. Differences in concentration and composition were found between organs. Young leaves and reproductive tissues, such as siliques and seeds, contain the highest concentrations, whereas senescing rosette leaves contain the lowest concentration of Gls (Brown *et al.*, 2003).

Gl variations in leaves and seeds of 39 *Arabidopsis* ecotypes were determined. 34 different Gls were identified, most of which are derived from methionine. It could be shown that *Arabidopsis* ecotypes differ in their Gl profiles (Kliebenstein *et al.*, 2001*a*)

Degradation products of glucosinolates

Gls are hydrolyzed by myrosinase upon tissue damage. This reaction leads to the formation of glucose and an unstable aglycone. Depending on the structure of the Gl side chain, on additional proteins and cofactors, different hydrolysis products, such as isothiocyanates, oxazolidine-2-thiones, nitriles, epithionitriles and thiocyanates can be produced (Halkier and Gershenzon, 2006).

Most hydrolysis products of aliphatic and aromatic Gls are nitriles and isothiocyanates. Isothiocyanate is toxic for a broad spectrum of organisms. When the isothiocyanate bears a hydroxyl group at C-2, it is unstable, forming oxazolidine-2-thione. These compound is known to cause goiter (Halkier and Gershenzon, 2006).

Protein factors, such as the epithiospecifier protein (ESP), are involved in nitrile formation *in vivo.* When a Gl side chain has a terminal double bound, ESP promotes the reaction to epithionitrile. ESP, identified in some *Arabidopsis* ecotypes, promotes the formation of epithionitriles and simple nitriles from many Gls (Halkier and Gershenzon, 2006). ESP is expressed in ecotype C24 and not expressed in ecotype Col-0 (Lambrix *et al.*, 2001).

Thiocyanates were only formed from benzyl-, allyl- and 4-methylsulfinylbutyl-Gls. Like nitrile formation, thiocyanate production is associated with protein factors, but these proteins have not been identified yet (Halkier and Gershenzon, 2006).

The hydrolysis of indol Gls is different from other Gls types, because the hydrolysis product isothiocyanate is unstable and is converted into further metabolites.

The hydrolysis of indolic Gls in the presence of ESP leads to elemental sulfur and indol-3-acetonitrile (IAN), which has auxin activity. Further, IAN can be converted to the plant hormone indol-3-acetic acid (auxin IAA) (Mithen, 2001).

Mature Gls are stored in the vacuole and occur in all tissues of all plant organs. Compared to Gl storage, myrosinases are localized in the cytoplasm of specialized scattered myrosin cells, which seem to be Gl free, and also in the cytoplasm of other cells. Thus, Gls and myrosinase are separated on cellular and subcellular level (Thangsatd *et al.*, 1990 and 1991). Additionally, myrosinase can be located in the same compartment as Gls, in the cell vacuole, but in an inactivated form (Bones and Rossiter, 1996). In this context, it would be interesting to find out more about the subcellular localization of ds-Gl SOT proteins, which catalyze the production of Gls.

Modulation of glucosinolate pattern in double zero rape

The Brassicaceae family includes important agricultural plants, like cabbage, broccoli and other *Brassica* vegetables and rape (*Brassica napus*). To date, Gls in rapeseed have been investigated well to decrease their antinutritional effects, resulting in “double zero” (zero erucic acid and zero Gls) rapeseed (Rosa *et al.*, 1997). Due to breeding, almost no Gls are produced by rapeseed anymore. This allows the protein rich seed meel (the remain after oil is expressed) to be used as animal food supplement. In rapeseed the most notable Gl is 2-hydroxy-3-butenyl Gl, its hydrolysis product is oxazolidine-2-thione, which causes goiter and has other harmful effects on animal nutrition (Griffiths *et al.*, 1998). In double zero rape all Gls, those with beneficial values and those with antinutritional effects, were almost removed. Maybe in future it will be possible to produce metabolically engineered plants with improved properties by modulation of the contents of single Gl.

General information about sulfotransferases

Characteristics of sulfotransferases

SOTs are a group of enzymes catalyzing the transfer of a sulfate group from the co-substrate 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to a hydroxyl group of various substrates with parallel formation of 3'-phosphoadenosine 5'-phosphate (PAP).

$$PAPS + R-OH \xrightarrow{SOT} PAP + R-OSO_3^-$$

According to their subcellular localization in mammalian cells SOTs were divided into two groups. The first group comprises membrane-associated SOTs that accept macromolecules as substrates, such as proteins and glucosaminoglycans. Members of this group were found in animals and plants (Niehrs *et al.*, 1994; Bowman and Bertozzi, 1999; Habuchi 2000; Hanai *et al.*, 2000). Members of the second group are soluble SOTs and accept small organic molecules such as flavonoids, steroids, neurotransmitters and xenobiotics (Varin *et al.*, 1992; Weinshilboum and Otterness, 1994; Weinshilboum *et al.*, 1997; Rouleau *et al.*, 1999). Members of this group were divided into two subgroups according to their functions. Enzymes of the first subgroup are involved in detoxification, and members of the second subgroup are involved in metabolic processes, such as steroid inactivation. The first subgroup accepts a broad variety of substrates and the second subgroup shows a high specificity for their substrates (Marsolais *et al.*, 2000).

To date, mammalian SOTs are studied well. It is known that all SOTs from bacteria, plants and mammals exhibit structural similarities, use PAPS as co-substrate and reveal highly conserved domains for PAPS-binding (Varin *et al.*, 1992; Yamazoe *et al.*, 1994). Due to these common features, knowledge from animal SOTs could be helpful to characterize plant SOTs, such as *Arabidopsis* SOTs. In contrast to plant SOTs, a dimerization motif near the C-terminus was identified in mammalian SOTs, named as KTVE motif. The mammalian SOTs mainly occur as homo- and heterodimers (Weinshilboum *et al.*, 1997; Negishi *et al.*, 2001). In plants, SOTs investigated are active as monomers (Varin and Ibrahim, 1989). In mammals, it could be shown that SOTs sulfate hormones, neurotransmitters, drugs and xenobiotics. The sulfation leads to an increased water solubility of most compounds and therefore to their renal excretion, but sulfation can also lead to activated metabolites (Weinshilboum, 1986; Mulder and Jakoby, 1990; Falany, 1991; Weinshilboum and Otterness, 1994).

Plant sulfotransferases

The aim of this study is the characterization of SOT proteins of *Arabidopsis*. The knowledge of already investigated SOTs could be helpful to approach this goal. Thus, the results of research on plant SOTs are summarized subsequently.

SOTs from *Flaveria chloraefolia* were the first plant SOTs which have been characterized (Varin *et al*., 1992). Flavonol SOTs from *Flaveria* species catalyze the sulfation of different flavonols and exhibited a strict specificity for substrates and the hydroxyl positition to be sulfated (Varin *et al*., 1997*b*; Marsolais *et al*., 2000). Flavonols might be involved in adaption to stress and are involved in polar auxin transport (Hahlbrock and Scheel, 1989; Faulkner and Rubery, 1992).

Plants use steroids to regulate their growth and development. In *Brassica napus* steroid SOTs were characterized, designated as BNST3 and BNST4. Both enzymes catalyze the sulfation of 24-epibrassinolide precursors with similar specificity. The expression of *B. napus SOT* gene(s) have been found to be induced by salicylic acid, a signal molecule in plant defence. Furthermore, it could be demonstrated that *BNST* genes are inducible by ethanol, other xenobiotics or low oxygen stress. Thus, is hypothesized that *Brassica* steroid SOTs are involved in growth, development and adaption to stress (Rouleau *et al.*, 1999; Marsolais *et al.*, 2004).

Tyrosylprotein SOTs were investigated in higher plants. Both monocot (rice and asparagus) and dicot (carrot, tomato and tobacco) cells were used to prepare microsomal membrane fractions for *in vitro* measurements. These membrane-bound enzymes catalyze the sulfation of tyrosine of a phytosulfokine-α precursor (Hanai *et al.*, 2000). Phytosulfokine-α, a disulfated pentapeptide, is a plant growth factor.

In conclusion, these prior investigations demonstrate that SOTs sulfate compounds as different as steroids, xenobiotics, hormones, neurotransmitters, drugs, flavonoids and peptides. Therefore, SOTs are involved in a broad spectrum of physiological processes *in vivo*.

Known sulfotransferases of *Arabidopsis* to date

RaR047 is the first cDNA clone found encoding a SOT of *Arabidopsis*. It could be shown that RaR047 mRNA levels are increased due to treatment with salicylic acid and methyl jasmonate, which are hormonal and stress related compounds. In addition, mRNA accumulation could be observed upon infection with bacterial pathogens. Furthermore, an mRNA accumulation was reported during growth of *Arabidopsis* cell cultures and in the aerial parts of young seedlings (Lacomme and Roby, 1996). To date, the substrate of the RaR047 protein is not known yet. Due to comparison of the crystal structure of RaR047 with the already crystalized mammalian steroid SOT and high

sequence similarity with the *Brassica napus* steroid SOT, it is assumed that the RaR047 protein is a steroid SOT (Kakuta *et al*., 1997; Smith *et al.*, 2004).

One aim of this study was to identify all SOT sequences of *Arabidopsis*. Thus, the *Arabidopsis* genome was searched for SOTs with the BLAST program, using the already known SOT sequence of RaR047.

It could be shown that one SOT of *Arabidpsis* (At5g07010) catalyzes the sulfation of 11- and 12-hydroxyjasmonate, but the closely related SOT (At5g07000, 87% identity) is inactive with the same substrates. 12-hydroxyjasmonate is derived from jasmonic acid, a signaling molecule, which is involved in diverse developmental processes and in plant defense. 12-hydroxyjasmonate and its sulfated form are naturally occuring compounds in *Arabidopsis* (Gidda *et al.*, 2003).

For another SOT of *Arabidopsis* (At3g45070), a flavonol activity was found. Compared to flavonol SOTs from *Flaveria* species, the *Arabidopsis* SOT showed strict specificity for position 7 of flavonols, but accepts a number of flavonols. This is surprising, because to date no natural occurrence of sulfated flavonoids in *Arabidopsis* was reported. It is thougt that flavonoids might be involved in plant growth and development (Marsolais *et al.*, 2000).

For a further *Arabidopsis* SOT (At2g14920), a brassinosteroid activity has been shown. This protein shows a preference for castasterone and 28-homocastasterone, the precursor of brassinolide and 28-homobrassinolide. A possible function of this enzyme could be the inactivation of brassinosteroids in *Arabidopsis* due to sulfation, and therefore an involvement in plant development (Marsolais *et al.*, 2000).

This summary about mammalian, plant and especially *Arabidopsis* SOTs demonstrates the broad spectrum of substrates and thus shows, that SOT functions in the organism vary a lot. The information collected on SOTs in general may be helpul in the characterization of the *Arabidopsis* SOT family, especially of ds-Gl SOTs.

Aims of this thesis

- Generation of a phylodendogram to identify all members of the *Arabidopsis* SOT family.
- Heterologous expression in *Escherichia coli* and purification of several SOT proteins from *Arabidopsis* ecotype C24 to use these recombinant proteins for *in vitro* measurements.
- Identification of *in vivo* substrates of AtSOT proteins following two strategies: 1) combination of transcriptomics and metabolomics 2) sequence comparison and usage of isolated ds-Gls as substrate in *in vitro* assays.
- Detailed enzyme kinetic studies with identified ds-Gl SOTs using different ds-Gls as substrate to determine substrate specificities of SOTs with naturally occuring substrates.
- Heterologous expression in *E. coli* and purification of one exemplary ds-Gl SOT protein from another *Arabidopsis* ecotype (Col-0) for *in vitro* measurements to compare *Arabidopsis* ecotypes C24 and Col-0.
- Comparison of the same ds-Gl SOT from these *Arabidopsis* ecotypes (C24 and Col-0) after point mutation with subsequent heterologous expression in *E. coli*, purification and *in vitro* measurements to determine whether replacement of single amino acids leads to differential substrate specificities.
- Expression studies of *ds-Gl SOT* genes under various conditions using Northern blot analysis to find out which conditions lead to overexpression or repression of the respective mRNAs.
- Fusion of the three ds-Gl SOT proteins with the green fluorescent protein (GFP) and expression in *Arabidopsis* protoplasts to identify the subcellular localization.

Chapter 2:
The multi-protein family of *Arabidopsis* sulpho-transferases and their relatives in other plant species

Klein M and Papenbrock J

Journal of Experimental Botany (2004) Vol. 55, No. 404: 1809–1820

Abstract

All members of the sulphotransferase (SOT, EC 2.8.2.-) protein family use 3'-phospho-adenosine 5'-phosphosulphate (PAPS) as the sulphuryl donor and transfer the sulphonate group to an appropriate hydroxyl group of several classes of substrates. These enzymes have highly conserved domains and can be found in eubacteria and eukaryotes. In mammals, sulphate conjugation catalysed by SOTs constitutes an important reaction in the transformation of xenobiotics and in the modulation of the biological activity of steroid hormones and neurotransmitters. In plants, sulphateconjugation reactions seem to play an important role in plant growth, development and adaptation to stress. To date only a few plant SOTs have been characterized in detail. The flavonol 3- and 4'-SOTs from *Flaveria* species (Asteraceae), which catalyse the sulphonation of flavonol aglycones and flavonol 3-sulphates, respectively, were the first plant SOTs for which cDNA clones were isolated. The plasma membrane associated gallic acid SOT of *Mimosa pudica* L. pulvini cells may be intrinsic to signalling events that modify the seismonastic response. In *Brassica napus* L. a SOT catalyses the *O*-sulphonation of brassinosteroids and thereby abolishes specifically the biological activity of 24-epibrassinolide. The fully sequenced genome of *Arabidopsis thaliana* Heynh. contains in total 18 genes that are likely to encode SOT proteins based on sequence similarities of the translated products with an average identity of 51.1%. So far only one SOT from *A. thaliana* (At5g07000) was functionally characterized: the protein was shown to catalyse the sulphonation of 12-hydroxy-jasmonate and thereby inactivate excess jasmonic acid in plants. The substrates and, therefore, the physiological roles of SOTs are very diverse. By using the numerous informative databases and methods available for the model plant *A. thaliana* the elucidation of the functional role of the SOT protein family will be accelerated.

Introduction

Members of the sulphotransferase (SOT) family have been found in most organisms investigated to date, except in Archaea. These enzymes catalyse the transfer of a sulphonate group from 3'-phosphoadenosine 5'-phosphosulphate (PAPS) to an appropriate hydroxyl group of various substrates with the parallel formation of 3'-phosphoadenosine 5'-phosphate (PAP). The SOTs, therefore catalyse the sulphonation of a wide range of compounds and produce sulphate esters and sulphate conjugates.

$$\text{PAPS} + \text{R-OH} \xrightarrow{\text{SOT}} \text{PAP} + \text{R-OSO}_3^-$$

The SOT proteins were classified on the basis of their affinity for different classes of substrates. One group of SOT proteins, mainly membrane-associated, accepts as substrates macromolecules such as proteins and peptides, and glucosaminoglycans (Niehrs *et al*., 1994). The second group, usually soluble proteins, accepts as substrates small organic molecules such as flavonoids, steroids, and xenobiotics with diverse chemical structures. A sulphate conjugation is more water-soluble than the non-sulphonated molecules (Weinshilboum and Otterness, 1994). It is hypothesized that sulphonation, via SOTs, affects the biological activity of certain compounds, thereby modulating physiological processes such as growth, development, and adaptation to stress (Varin *et al.*, 1997*b*; Yang *et al*., 1999; Marsolais *et al*., 2000).

In *Arabidopsis thaliana* Heyn. 18 *SOT* genes have been identified through alignment search tools (described in this article), that may encode SOT proteins based on sequence identities, but little is known about their functions. The first *SOT* gene identified in *A. thaliana* (RaR047, At2g03760) was only expressed in shoots and in *A. thaliana* cell cultures, and its expression was enhanced by challenging plants with isolates of *Xanthomonas campestris* pathovar *campestris* 147 and of *Pseudomonas syringae* pathovar *maculicola* M2 undergoing an incompatible interaction (Lacomme and Roby, 1996).

Structural similarities are present among SOTs from eubacteria, plants, and mammals (Yamazoe *et al*., 1994). SOTs comprise a very important and interesting group of enzymes because they are widely distributed and are involved in a broad spectrum of physiological functions (Weinshilboum and Otterness, 1994; Klaassen, and Boles, 1997; Varin *et al.*, 1997*b*; Hanai *et al*., 2000; Marsolais *et al*., 2000).

What are sulphotransferases not?

Sulphotransferases are different from sulphurtransferases. The names of both groups of enzymes, sulphotransferases and sulphurtransferases, are very similar; both groups of

proteins were often mixed up in the past. To avoid confusion about the catalytic activity of members of each protein family, the difference will be defined. Sulphurtransferases/ rhodaneses are a group of enzymes widely distributed in all three phyla that catalyse the transfer of reduced sulphur from a donor, such as thiosulphate or 3-mercaptopyruvate, to a thiophilic acceptor substrate (Westley, 1973). In the reaction catalysed by sulphur-transferase proteins the transferred sulphur is highly reduced (oxidation state -II), whereas the SOT proteins catalyse the transfer of sulphur that is highly oxidized (oxidation state VI). The sequences of both protein families are not related. In *A. thaliana* the family of sulphurtransferases also consists of 18 members (Bauer and Papenbrock, 2002).

In the past the enzyme in the sulphur-assimilation pathway which reduces adenosine-5'-phosphosulphate (APS) to sulphite was misleadingly called APS sulphotransferase. Only recently, after the reaction mechanism had been elucidated in detail, was the protein renamed APS reductase (Suter *et al*., 2000). In the older literature the enzymes catalysing the APS reduction are always referred to as sulphotransferase. However, they do not belong to the SOT protein family described in this paper and do not possess any sequences similarities with SOT proteins.

The nomenclature for protein families often has a historical background. To minimize confusion between different protein families, it is proposed to abbreviate the sulphotransferases “SOT” instead of the formerly used abbreviation “ST”, which is now broadly used as abbreviation for sulphate transporters. The abbreviation for sulphur-transferases has been changed from the formerly used “ST” into “STR” (The Arabidopsis information resource, http://arabidopsis.org/info/genefamily/STR_gene family.html).

Biosynthesis of APS and PAPS

PAPS is an obligate co-substrate for sulphonation reactions catalysed by SOTs. PAPS is synthesized from ATP and endogenous sulphate in a two-step reaction. In the first step ATP sulphurylase (EC 2.7.7.4) catalyses sulphate activation. The enzyme hydrolyses the bond between the β- and the γ-phosphates of ATP and then adds sulphate to the γ-phosphate. The activation step is necessary because sulphate is metabolically inert. The energy is stored in the phosphoric acid-sulphuric anhydride bond of the reaction product, 5'-adenylylsulphate (APS), allowing sulphate to undergo further reactions.

$$ATP + SO_4^{2-} \xleftrightarrow{\text{ATP sulphurylase}} APS + PP_i$$

The energetics of the sulphate adenylylation reaction favours ATP formation. Therefore the reaction products, APS and pyrophosphate (PP_i), must be maintained at a

low concentration by the enzymes, inorganic pyrophosphatase that hydrolyses PP_i, and APS reductase (EC 1.8.4.9) and APS kinase (EC 2.7.1.25) that metabolize APS. APS reductase catalyses the first step of sulphate reduction. APS kinase catalyses the ATP-dependent phosphorylation on the 3'-position of APS. The product PAPS is the substrate for the SOT proteins.

$$\text{ATP + APS} \xleftrightarrow{\text{APS kinase}} \text{PAPS + ADP}$$

In general the availability of PAPS for sulphonation *in vivo* depends on its synthesis, transport, degradation, and utilization. As the SOT proteins from mammals have been investigated in more detail, the situation in mammals will be outlined initially. In mammals, PAPS synthesis depends directly on sulphate availability which is the limiting factor in the first step. On the other hand, steady-state PAPS concentration cannot be increased by increasing endogenous sulphate concentration. The second part of PAPS biosynthesis depends on both PAPS and APS. PAPS inhibits its own synthesis through negative product feedback. PAPS utilization or transport out of cytoplasm promotes PAPS synthesis (Klaassen and Boles, 1997). *In vitro* tests have shown that excess APS (substrate) also inhibits APS kinase. PAPS is formed in the cytoplasm. Sulphonation of macromolecules takes place in the lumen of the Golgi apparatus (Capasso and Hirschberg, 1984; Mandon *et al.*, 1994). Their sulphonation depends on PAPS synthesis and transport. A PAPS transporter was purified from Golgi vesicles (Mandon *et al*., 1994); the transporter is inhibited by 3',5'-adenosine diphosphate (Zaruba *et al.*, 1988).

In mammals, PAPS is degraded by two different pathways leading to the same end product, 5'-adenosine monophosphate (5'-AMP): PAPS is either desulphated by PAPS sulphohydrolase, forming PAP, which is then dephosphorylated by PAP nucleotidase to yield 5'-AMP, or PAPS is dephosphorylated by 3'-nucleotidase/PAPS-phosphorylase, forming APS, which is degraded by APS sulphohydrolase, forming 5'-AMP. The intermediate products PAP and APS can in turn influence sulphonation: high concentrations of APS inhibit APS kinase activity; PAP inhibits certain SOT proteins and competes with PAPS for transport in the Golgi apparatus (Klaassen and Boles, 1997).

In plants, the influences on and the regulation of the PAPS pool are not very well understood. It was shown for the intracellular localization of ATP sulphurylase that there are two forms, localized in plastids and in the cytoplasm (Rotte and Leustek, 2000). Plastid-localized ATP sulphurylase makes up 70% - 95% of the total enzyme activity in leaves. One could assume that plastid-localized ATP sulphurylase is involved in the assimilative sulphate reduction since at least one of the subsequent enzymes, sulphite reductase, can only be found in plastids. The absence of sulphate-reduction

enzymes in the cytoplasm suggests that cytoplasmic ATP sulphurylase may be responsible for another function, such as providing activated sulphate for the sulphonation reactions catalysed by cytoplasmic SOT proteins.

It has been estimated that the *A. thaliana* genome contains four APS kinase genes and two of them have been functionally characterized (At2g14750, At4g39940). Both proteins presumably represent plastidial isoforms in the plant APS kinase gene family (Leustek, 2002; Lillig *et al*., 2001). The localization predictions for the two other APS kinase proteins (At3g03900, At5g67520) obtained with the programs PSORT and TargetP give very low probabilities for plastid localization and indicate an association with membranes, probably with the plasma membrane (data not shown). The APS kinase 1 (Atakn1, At2g15750) was expressed in *E. coli*. The recombinant protein formed PAPS at a V_{max} of 7.35 U x mg^{-1}, the K_m for APS was 0.14 µM, and for ATP 147 µM. APS caused a severe substrate inhibition (K_i 4.5 µM). The type of inhibition is uncompetitive with respect to MgATP. Chloroplast APS kinase is regulated *in vitro* by the redox charge with thioredoxin as essential activator. More experiments are required to differentiate which of the APS kinase proteins in *A. thaliana* are involved in providing PAPS for sulphonation reactions.

One might ask the question whether PAPS can be used up under certain conditions and, therefore limit the process of sulphonation reactions. It was shown that sulphonation is a high-affinity, low-capacity enzymatic process in which the entire liver content of PAPS can be consumed in less than 2 min (Klaassen and Boles, 1997). For plants there is no information so far on whether sulphate availability can influence the PAPS pool used for sulphonation. The determination of the APS and PAPS pools in plants will be important for the understanding of the regulation of PAPS biosynthesis and, therefore, also the regulation of sulphonation reactions as well.

Distribution of sulphotransferases in mammals and plants

The well-studied mammalian SOTs contribute important information to understanding about plant SOTs. Due to the structural similarity among SOTs in general (protein family characteristics are defined in KOG1584, http://genome.jgi-psf.org/chlre2/kog/168755.html, and PF00685, http://pfam.wustl.edu/hmmsearch.shtml), knowledge about mammalian SOTs may be useful. Mammalian SOTs catalyse the sulphate conjugation of many hormones, neurotransmitters, drugs, and xenobiotic compounds. These reactions lead to increased renal excretion of sulphonated products. This results in a decrease in biological activity; but sulphonation is required to activate molecules as well (Weinshilboum and Otterness, 1994; Weinshilboum *et al.*, 1997). According to their cellular localization, SOTs sulphonate different kinds of molecules. The so-called

cytosolic soluble SOTs sulphonate small molecules, while Golgi membrane-associated SOTs sulphonate large molecules such as proteins and glycosaminoglycans (Habuchi, 2000; Honke and Taniguchi, 2002). Genes for human enzymes, as well as those of other mammalian soluble enzymes, show a high degree of structural homology (Weinshilboum and Otterness, 1994; Weinshilboum *et al.*, 1997). To date, nine soluble SOTs have been identified from humans (Sugahara *et al.*, 2003). The sulphonation reaction and the formation of its substrates seem to be very important for life: a naturally-occurring defect in the synthesis of the co-substrate PAPS is lethal in humans (Superti-Furga, 1994). A large number of membrane SOTs produce numerous biological substrates essential for regulatory processes in life such as the action of steroid hormones (Negishi *et al.*, 2001).

X-ray crystal structures have been produced of four mammalian cytosolic soluble and from one domain of a Golgi membrane-associated SOT. All crystallized SOTs are globular proteins composed of a single α/β domain with a characteristic five-stranded β-sheet. The β-sheet constitutes the core of the PAPS-binding and catalytic sites. A common molecular mechanism reaction of the sulphuryl transfer was determined (Negishi *et al.*, 2001).

Originally, the proteins in the SOT superfamily were classified according to the following scheme: a family contains proteins that share at least 45% amino acid sequence identity (SULT1, phenol SOT; SULT2: hydroxysteroid SOT; SULT3: flavonol SOT family), whereas members of subfamilies further divided in each SOT protein family are more than 60% identical in amino acid sequence. Although these percentages were arbitrary, they are very similar to those that have proved useful in classifying other protein superfamilies, such as the cytochrome P450s (Weinshilboum *et al.*, 1997). Recently, the human SOT families were further subdivided in the following way: the SULT1 phenol SOT family contains four subfamilies, the phenol SOT (SULT1A, EC 2.8.2.-), the Dopa/tyrosine SOT (SULT1B, EC 2.8.2.9), the hydroxyarylamine SOT (SULT1C, EC 2.8.2.3), and the estrogen SOT (SULT1E, EC 2.8.2.15). The SULT2 hydroxysteroid SOT family consists of two subfamilies, the dehydroepiandrosterone SOT (SULT2A) and the cholesterol SOT (SULT2B) (Sugahara *et al.*, 2003; Yoshinari *et al.*, 2001). However, based on the above-mentioned criteria a number of other eukaryotic SOT proteins whose substrate specificity was elucidated could not be classified in the existing scheme. Either the proteins showed less sequence identity although they possessed the same substrate specificity as other members of the subfamily, or they used a substrate chemically very different from the substrate typical for a particular subfamily but showed high sequence identities (Sugahara *et al.*, 2003). Thus, the classification scheme has to be adapted to the latest results with respect to substrate specificity.

Little is known about plant SOTs compared with mammalian SOTs. SOTs of *Flaveria* species and *Brassica napus* L. are well characterized by means of molecular biology and biochemistry. Thus, the SOTs from *Flaveria* species represent a general model for plant SOTs. The flavonol 3- and 4´-SOTs from *Flaveria chloraefolia* were the first plant SOTs for which cDNA clones were isolated and characterized (Varin *et al.*, 1992). Additional SOTs from *Flaveria bidentis* (L.) Kuntze have been characterized. This group of SOTs accept different flavonols as sulphate acceptors (Varin *et al.*, 1997*b*). These enzymes exhibit strict specificity for the substrate and the position of the hydroxyl group to be sulphonated. The biological function of flavonols and their derivatives are not fully understood. Flavonols might be involved in adaptation to stress, for example microbial attack (Hahlbrock and Scheel, 1989). Flavonols may also act as a regulator of polar auxin transport (Faulkner and Rubery, 1992). The common characteristics of flavonol SOTs are as follows: they all do not use divalent cations for sulphonation, they have a similar mass of 35 kDa and the active form of these enzymes is a monomer. K_m values range between 0.2 and 0.4 µM for PAPS and various flavonols (Varin *et al*., 1997*b*).

A very different substrate type for SOT proteins was described in halophytic plants. Choline-*O*-sulphate is a compatible osmolyte accumulated under saline conditions by members of the halophytic genus *Limonium* and other Plumbaginaceae. A choline SOT (EC 2.8.2.6) responsible for the formation of choline-*O*-sulphate was characterized in *Limonium* species. The choline SOT activity was catalysed by a soluble protein and required PAPS as the sulphate donor. Apparent K_m values were 25 µM for choline and 5.5 µM for PAPS. In roots and leaves of *Limonium perezii* (Stapf) F.T. Hubb. the activity was increased at least 4-fold by salinization with 40% (v/v) artificial sea water. Here the sulphonated choline has a role in tolerance against salt stress as a beneficial osmoprotectant. Among the non-accumulators such as barley, maize, sunflower and *Brassica* species, none had significant choline SOT activity (Rivoal and Hanson, 1994). The type and the sequence of the SOT protein catalysing this reaction has not been identified so far.

In *Mimosa pudica* L. a SOT activity was characterized from plasma membrane preparations. The SOT protein catalysed the transfer of sulphate from PAPS to gallic acid glucoside; the reaction product was identical with gallic acid, β-D-glucopyranosyl-6'-sulphate, the periodic leaf movement factor. Therefore the 42 kDa SOT protein analysed in *M. pudica* might be involved in the induction of the seismonastic response movement (Varin *et al*., 1997*a*).

Evidence exists that plants, like mammals, use steroids to regulate their growth and development. In *B. napus* a SOT protein was characterized that catalysed the sulpho-

nation of brassinosteroids and mammalian estrogenic steroids. The sulphonation abolishes specifically the biological activity of 24-epibrassinolide. Treatment with salicylic acid, a signal molecule in plant defence, leads to increased expression of the *B. napus SOT* gene. This suggests an involvement of at least one SOT protein in plant responses to pathogen infection (Rouleau *et al.*, 1999).

The first SOT encoding cDNA clone from *A. thaliana*, RaR047 (At2g03760), was isolated by Lacomme and Roby (1996). However, the gene product was not functionally analysed. The first *A. thaliana* SOT protein (At5g07000) was functionally analysed only recently. A different group of chemical compounds was shown to be sulphonated by this SOT protein. *In vitro* the recombinant *A. thaliana* SOT protein exhibited strong substrate specificity for 11- and 12-hydroxyjasmonate. The K_m value for PAPS was found to be 1 µM. *In vivo* the naturally occurring 12-hydroxyjasmonate was sulphonated in *A. thaliana* (Gidda *et al.*, 2003). Initially 12-hydroxyjasmonate was isolated as a tuber-inducing compound from *Solanum tuberosum* (Yoshihara *et al.*, 1989). Hydroxylation and subsequent sulphonation might be components of a pathway that controls the biological activity of 12-hydroxyjasmonate or inactivates excess jasmonic acid in plants (Gidda *et al.*, 2003).

This short summary demonstrates the diversity of substrates used by SOT proteins in addition to the identical co-substrate PAPS, and the broad spectrum of physiological processes where sulphonated compounds are involved. The different K_m values for PAPS (from 0.1 - 5 µM) provide the plant with a regulatory system for the use of the PAPS pool for different sulphonation reactions.

The multi-protein family of *A. thaliana* sulphotransferases

In recent years the scientific community was provided with valuable information about the model plant *A. thaliana* (The Arabidopsis genome initiative, 2000). Extensive use of all sources available will help to analyse and differentiate between the members of protein families. Therefore, the aim of this study was to identify all genes and gene products which might be classified as SOT in *A. thaliana*. The fully sequenced genome of *A. thaliana* was searched for SOT sequences applying the BLAST program with the already isolated SOT RaR047 protein sequence from *A. thaliana* (Lacomme and Roby, 1996). 18 SOT protein sequences showing high similarity to already known SOT protein sequences and to each other were identified. The phylogenetic tree of the family of *A. thaliana* SOTs shows the relationships among these 18 sequences (Fig. 1). The protein sequences were divided into seven groups according to their sequence similarities; the results are displayed in Table 1. The table shows an overview of the complete SOT family with additional information including gene identification,

numbers of amino acids, number of ESTs identified, and intracellular localization predictions (as explained later). Apart from two proteins, all members of this family consist of an approximately equal number of amino acids of at least 310 residues. Computer analysis of *A. thaliana* amino acid sequences indicates that all 18 SOTs might be soluble and none of the SOT proteins contains a transmembrane region as indicated by hydropathy plots in SOSUI (http://sosui.proteome.bio.tuat.ac.jp).

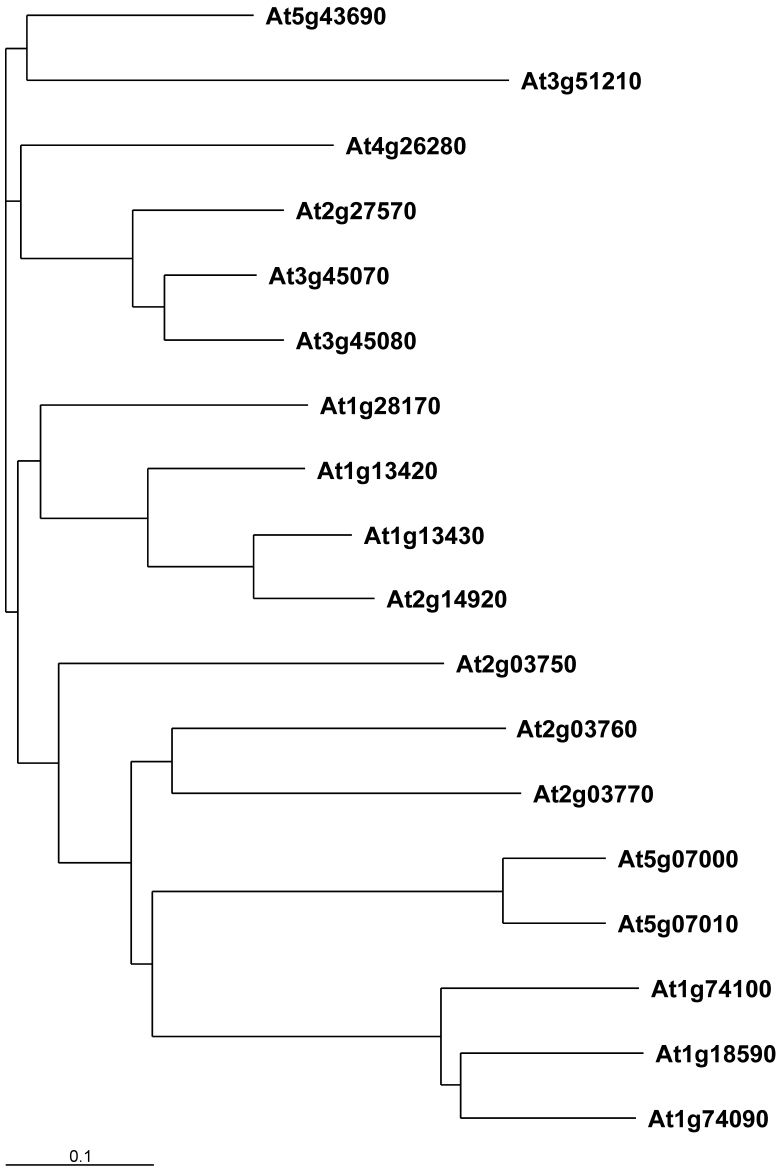


Fig. 1. Phylodendrogram of *Arabidopsis thaliana* sulphotransferases (SOTs). The fully sequenced genome of *A. thaliana* was searched for SOT sequences applying the BLAST program with the already isolated SOT RaR047 protein sequence from *A. thaliana* (At2g03760) (http://www.ncbi.nlm.nih.gov/BLAST/). 18 SOT sequences were identified in *A. thaliana*. The respective proteins were grouped according to their amino acid sequence similarities using the Clustal W program (http://www.ebi.ac.uk/clustalw). The phylogenetic tree shows the relationship among the 18 SOT proteins.

Table 1. Features of the SOT family from *A. thaliana* and localization prediction for the SOT proteins

Protein name, gene identification, number of amino acids, and EST clones for 18 SOTs are summarized. The program PSORT was used for the localization prediction (http://psort.ims.u-tokyo.ac.jp). Numbers in brackets give the certainty of prediction. The number of EST clones was determined on 14 February 2004 (http://mips.gsf.de, http://arabidopsis.org). Abbreviations: aa, amino acids; ER, endoplasmic reticulum; ID, identification; No.,number.

Group	Name	Gene ID	No. aa	EST	Localization prediction
I	AtSOT1	At5g43690	331	2	Microbody (peroxisome) (0.748) Chloroplast stroma (0.200)
	AtSOT2	At3g51210	67	-	ER (membrane) (0.550) Microbody (peroxisome) (0.320)
II	AtSOT3	At4g26280	314	-	Cytoplasm (0.450) Microbody (peroxisome) (0.313)
	AtSOT4	At2g27570	273	-	Nucleus (0.980) Microbody (peroxisome) (0.429)
	AtSOT5	At3g45070	323	4	Microbody (peroxisome) (0.575) Cytoplasm (0.450)
	AtSOT6	At3g45080	329	-	Microbody (peroxisome) (0.522) Cytoplasm (0.450)
III	AtSOT7	At1g28170	326	-	Microbody (peroxisome) (0.572) Mitochondrial matrix (0.100)
	AtSOT8	At1g13420	331	2	Cytoplasm (0.650) Mitochondrial matrix (0.100)
	AtSOT9	At1g13430	351	2	Cytoplasm (0.450) Microbody (peroxisome) (0.405)
	AtSOT10	At2g14920	333	-	Cytoplasm (0.450) Microbody (peroxisome) (0.392)
IV	AtSOT11	At2g03750	351	9	ER (membrane) (0.550) Microbody (peroxisome) (0.291)
V	AtSOT12	At2g03760 (RaR047)	326	5	Microbody (peroxisome) (0.622) Mitochondrial matrix (0.100)
	AtSOT13	At2g03770	324	-	Microbody (peroxisome) (0.705) Mitochondrial matrix (0.100)
VI	AtSOT14	At5g07000	347	6	Microbody (peroxisome) (0.602) Chloroplast stroma (0.200)
	AtSOT15	At5g07010	359	15	Microbody (peroxisome) (0.640) Mitochondrial matrix (0.484)
VII	AtSOT16	At1g74100	338	25	Microbody (peroxisome) (0.640) Cytoplasm (0.450)
	AtSOT17	At1g18590	346	8	Microbody (peroxisome) (0.640) Chloroplast stroma (0.566)
	AtSOT18	At1g74090	350	6	Microbody (peroxisome) (0.640) Cytoplasm (0.450)

In human and mouse SOT sequences, a dimerization motif near the C-terminus was identified, designated as the KTVE motif (Negishi *et al*., 2001). In mammals, cytosolic soluble SOTs are predominantly dimers, both homo- and heterodimers (Weinshilboum *et al*., 1997). In plants, the enzymes characterized so far exist as catalytically active monomers (Varin and Ibrahim, 1989). *A. thaliana* SOT protein sequences do not contain a KTVE motif; therefore they also might occur as monomers.

Interestingly, nearly all of the *A. thaliana* SOT genes do not contain introns (exceptions: AtSOT3, 4, and 10), in contrast to mammalian SOT genes. Genes for human SOTs, as well as for other mammalian SOTs, show a high degree of structural homology with conservation of the locations of most intron/exon splice junctions (Weinshilboum *et al.*, 1997).

Remarkably, the numbers of EST clones available are very low. For the *A. thaliana* genes in group IV, VI, and VII, the number of EST clones identified are in the same range as in other gene families (Bauer and Papenbrock, 2002) whereas in the other groups the numbers are relatively low. For seven out of the 18 putative SOT genes no EST clone has been identified so far. There are several hypothetical explanations: several of the SOT genes might be pseudogenes which are not expressed; the abundance of the SOT mRNA molecules is very low or the mRNAs are not very stable; the genes are only expressed in very specific conditions not included in the EST projects done so far. Fortunately, the coding sequence for most of the AtSOTs could be amplified from genomic DNA because the genes do not contain introns. It will be a challenge to find conditions for the expression of these seven SOT genes.

Sequence/function analysis of *A. thaliana* SOT proteins

A comparison of amino acid sequences of *A. thaliana* SOTs was done with other plant SOTs with known substrate specificities (Fig. 2). The aim of this comparison was to get indications about substrate specificities of *A. thaliana* SOTs. RaRO47 (At2g03760) was the first cDNA clone isolated encoding a SOT from *A. thaliana* and has not yet been characterized biochemically. The clone shows a high similarity of 87% with a SOT of *B. napus* (steroidST-3). Another SOT in the same group (At2g03770, group V) also shows a good correspondence of 57% with steroidST-3. *B. napus* SOTs catalyse the *O*-sulphonation of brassinosteroids and mammalian estrogenic steroids (Rouleau *et al*., 1999). Because of the mentioned close similarity, RaR047 should be tested for catalysis of these substrates as well.

AtSOT14 in group VI has been characterized biochemically. The recombinant AtSOT14 protein sulphonated 11- and 12-hydroxyjasmonate whereas for the closely related AtSOT15 protein (87% identity with AtSOT14) no activity was observed with

these substrates (Gidda *et al*., 2003). Both SOTs in group VI show a smaller correspondence of only 42 - 43% compared with the flavonol SOT of *F. chloraefolia* (F4-ST) and *B. napus* SOTs (steroidST1-3, 41 - 43%). Therefore, other substrates related in structure to jasmonates have to be tested to determine the biochemical function of AtSOT15.

The comparison of sequence similarities to proteins with known substrate specificities was also applied for group I to IV and VII, however, the differences are probably not significant. Group VII shows a better correspondence with the flavonol SOT of *F. chloraefolia* (F4-ST, 46 - 48%) than the comparison with *B. napus* SOTs (steroidST1-3, about 41%). Thus, the substrate for this group of *A. thaliana* SOTs may be a flavonol or a similar substrate. If such a substrate could be identified, the SOT group VII would represent an *A. thaliana* flavonol protein family.

Group II shows a slightly better correspondence with the *B. napus* SOTs (steroidST1-3, 43 - 49%) than with the *F. chloraefolia* SOT (F4-ST, 39 - 43%). Despite these results, At3g45070 has been shown to accept a number of flavonols and flavone aglycones as well as their sulphonated derivates (Marsolais *et al.*, 2000). The remaining groups I, III, and IV show an average similarity of about 44% with *B. napus* SOTs (steroidST1-3). This value is only slightly higher than the respective *F. chloraefolia* SOT (F4-ST) value (40 - 41%). In summary, the sequence/function analysis for the SOT protein family did not reveal clear results. The relatively weak similarities and the small differences among the groups make a biochemical analysis to identify the natural substrate for each SOT protein essential.

The rules for subdivision into families and subfamilies according to the percentage of their sequence identity (Weinshilboum *et al*., 1997) are not very useful for the plant SOT families. The detailed comparison above shows that high sequence similarity alone does not necessarily indicate specificity for the same chemical group of substrates. Even from high sequence identity of more than 85% among two SOT proteins, one cannot conclude the same substrate specificity. Probably SOT proteins specific for a group of substrates evolved independently on more than one occasion. Thus, for each SOT protein the *in vitro* and finally the *in vivo* substrate specificity has to be detected. The crystal structures of a number of mammalian SOT proteins are available which might be used for three-dimensional modelling of the active site and putative substrates from ligand libraries.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22			Group
	57	64	67	66	65	63	64	65	63	56	46	47	42	44	42	41	42	43	43	44	46	1	At5g43690	I
		52	56	54	54	54	54	50	48	47	49	42	35	36	39	41	41	40	42	46	44	2	At3g51210	
			64	65	62	58	58	57	55	51	46	47	40	41	40	40	41	43	43	44	45	3	At4g26280	II
				80	82	59	66	64	60	53	46	46	39	41	42	42	40	39	43	46	45	4	At2g27570	
					86	62	66	63	60	56	49	48	43	42	43	44	43	43	47	49	49	5	At3g45070	
						61	67	63	61	56	48	47	41	41	42	43	42	41	46	48	48	6	At3g45080	
							66	59	59	55	46	44	41	40	40	40	41	39	43	43	45	7	At1g28170	III
								77	74	54	47	42	43	44	40	41	39	40	45	45	46	8	At1g13420	
									85	51	44	41	40	41	40	40	39	40	43	42	43	9	At1g13430	
										52	44	38	38	40	40	39	39	39	43	41	43	10	At2g14920	
											44	47	42	44	39	40	38	41	44	43	45	11	At2g03750	IV
												55	44	44	40	45	42	40	78	82	87	12	At2g03760	V
													43	41	43	43	44	41	52	54	57	13	At2g03770	
														87	39	41	42	43	42	41	42	14	At5g07000	VI
															39	41	39	42	43	41	42	15	At5g07010	
																74	76	46	39	39	42	16	At1g74100	VII
																	77	48	41	43	45	17	At1g18590	
																		46	39	40	42	18	At1g74090	
																			39	38	39	19	F4-ST	
																				74	81	20	steroidST-1	
																					83	21	steroidST-2	
																						22	steroidST-3	

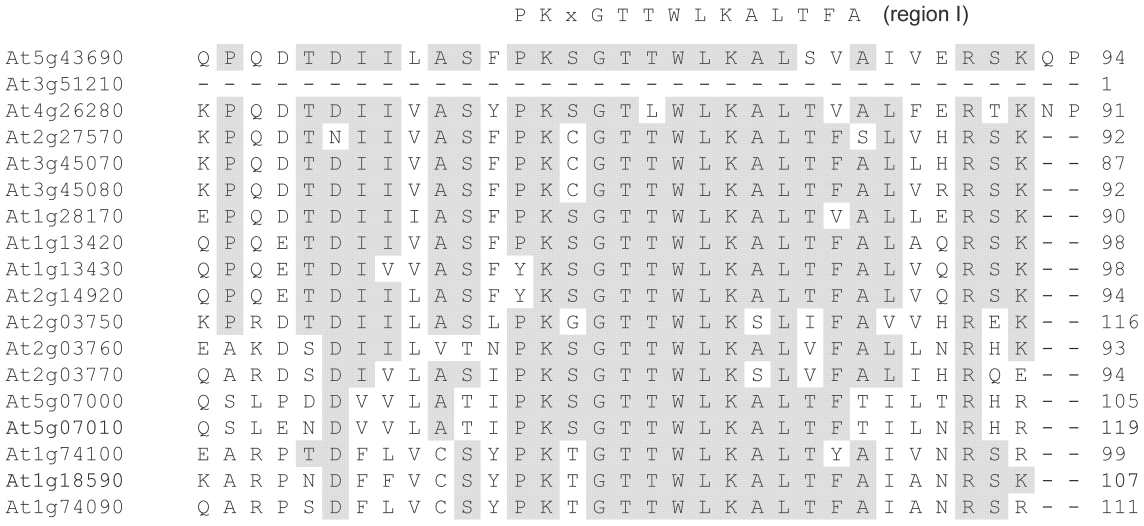
Fig. 2. Comparison of amino acid sequences of *Arabidopsis thaliana* SOTs with well-known plant SOTs from *Flaveria chloraefolia* and *Brassica napus* (Jotun Hein method in MegAlign/DNASTAR, Madison, WI, USA). Gene identifications, with subdivision into groups and abbreviations of known plant SOTs, are listed in the right columns. F4´-ST represents flavonol-4´-SOT from *F. chloraefolia* (Accession no. M84135), steroidST1-3 represent the steroid SOT 1–3 from *B. napus* (Accession no. AF000305–307). Shadowed areas are as mentioned in the text. The values indicate the identity in percent at the amino acid level.

Alignment of the highly conserved regions

Cytosolic soluble SOTs from mammalian species and plant SOT proteins have high structural similarities. All SOTs have conserved amino acid motives which are involved in PAPS binding (regions I and IV) (Marsolais and Varin, 1995; Weinshilboum *et al*., 1997). Figure 3 shows a partial amino acid alignment of the putative PAPS-binding regions of the *A. thaliana* SOT protein family. Region I is localized near the N-terminus and region IV at the C-terminus. This alignment indicates that the typical binding site for PAPS exists in all 18 SOTs. However, a comparison of the consensus sequences for SOT proteins from a broad spectrum of species described previously (region I, TYPKSGTxW; region IV, RKGxxGDWKxxFT) (Weinshilboum *et al*., 1997) are different from the consensus sequences in the PAPS-binding regions of *A. thaliana* SOT proteins (region I, **PK**x**GT**T**W**LKALTFA; region IV, F**RKG**xV**GDW**xxxL**T**). In at least 14 *A. thaliana* SOTs, the amino acids of these consensus sequences are identical; there are not more than three different amino acids at one position among all 18 SOT proteins. The amino acids in the *A. thaliana* consensus sequence printed in bold are identical with the overall consensus sequence. In the first published *A. thaliana* RaR047 sequence (Lacomme and Roby, 1996), there are a number of sequence deviations

compared with the sequence published later (The Arabidopsis genome initiative, 2000), including in the consensus sequence of region IV.

A



B

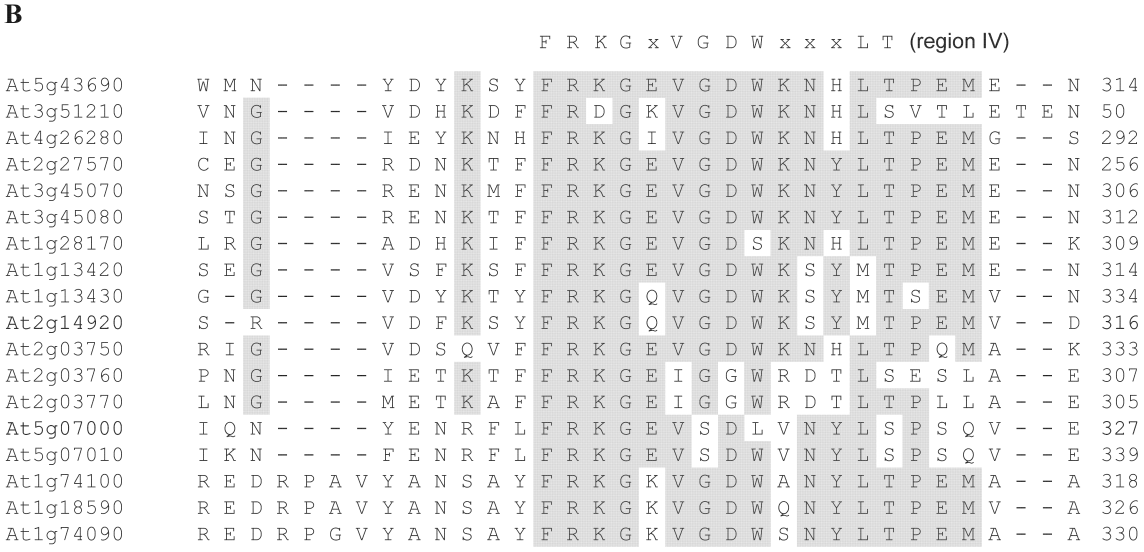


Fig. 3. Partial amino acid alignment of *Arabidopsis thaliana* SOTs. The 18 protein sequences of the identified SOTs were aligned (Jotun Hein method in MegAlign/DNASTAR, Madison, WI, USA). The consensus sequences are listed at the top (regions I and IV). 'x' represents any amino acid. Numbers on the right refer to amino acid position within the protein. Majority values are shadowed, including residues critical for 3'-phosphoadenosine 5'-phosphosulphate (PAPS) binding. (A) Shows the highly conserved region I, and (B) the highly conserved region IV. Both conserved regions, critical for PAPS binding, are present among all 18 SOT proteins.

Intracellular localization of *A. thaliana* SOT

In mammalian species one group of SOT proteins is associated with membranes and accepts as substrates macromolecules such as proteins and glycosylaminoglycans, and a second group of SOT proteins is soluble and accepts small organic molecules as substrates (Habuchi, 2000; Niehrs *et al*., 1994; Sugahara *et al*., 2003). In plants, membrane-associated SOT proteins might be involved in the biosynthesis of phytosulphokines (Hanai *et al*., 2000). A SOT protein characterized in more detail was shown to be localized in the plasma membrane and to be involved in the seismonastic response in *M. pudica*. The size of this membrane SOT protein was 42 kDa, whereas most soluble plant SOT proteins have only a molecular mass of about 35 kDa. The difference might reflect the addition of a transmembrane domain (Varin *et al*., 1997*a*).

For the localization prediction of nuclear-encoded proteins in the cell, several programs have been developed. PSORT, TargetP, and further programs in http://www.expasy.ch/tools were applied. The prediction programs use different algorithms. PSORT is based on an expert system with a knowledge-base and is a collection of 'if-then'-type rules (Nakai and Kanehisa, 1992). TargetP is a neural network-based tool using N-terminal sequence information only. It discriminates between proteins destined for the mitochondrion, the chloroplast, the secretory pathway, and “other” localizations with a calculated success rate of 85% (plant) and 90% (non-plant) in redundancy-reduced test sets (Emanuelsson *et al*., 2000). The results are summarized in Table 1. None of the *A. thaliana* SOT proteins contains an N-terminal transit peptide or a mitochondrial pre-sequence in TargetP, apart from AtSOT11 (At2g03750) which possess a 17 amino acid extension indicating a label for the secretory pathway (probability 0.794). The program specific for the prediction of peroxisomal proteins (http://mendel.imp.univie.ac.at/mendeljsp/sat/pts1/PTS1predictor.jsp) did not recognize and classify any of the *A. thaliana* SOT proteins as peroxisomal, although the PSORT program suggested transport into peroxisomes with probabilities higher than 0.5 for AtSOT1, AtSOT5 to 7, and AtSOT12 to 18. In summary, the computer predictions for the intracellular localization of the *A. thaliana* SOT proteins do not reveal clear and reliable results. Therefore the intracellular localization has to be investigated experimentally. One method to demonstrate the intracellular localization of nuclear-encoded proteins is the transient expression of fusion constructs with the green fluorescent protein (GFP) (Bauer *et al*., 2004; Nowak *et al*., 2004).

One example of this approach is shown in Fig. 4. According to the intracellular localization prediction, AtSOT18 is either localized in the peroxisomes (0.640) or in the cytoplasm (0.450). The full-length cDNA sequence encoding AtSOT18 was cloned in a frame downstream (pGFP-C) or upstream (pGFP-N) of the GFP reading frame.

A. thaliana protoplasts were transiently transformed with the GFP constructs according to standard procedures. Bright-field images are shown to visualize the protoplast's cell borders and the chloroplasts. Fluorescent images of the same protoplasts were done using a fluorescence microscope. As a control, the protoplasts were transformed with the pGFP-C and pGFP-N vector without additional insertion. According to the results, the AtSOT18 protein is cytoplasmic. These studies will be extended to obtain a complete set of data of the intracellular localization of all *A. thaliana* SOT proteins.

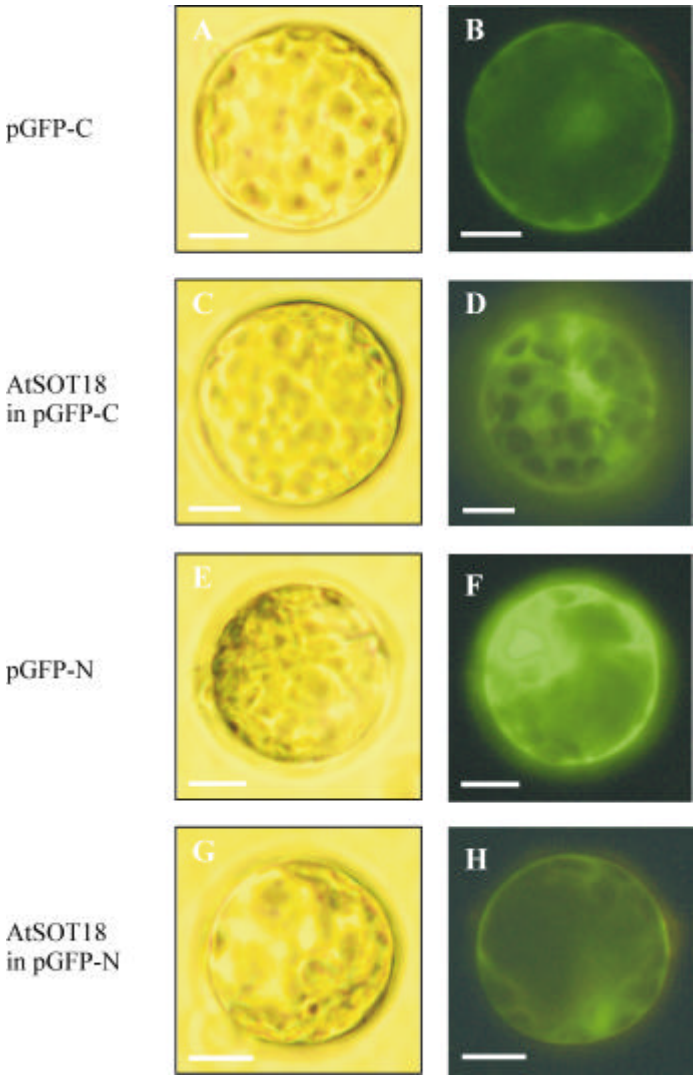


Fig. 4. Intracellular localization of AtSOT18. The full length cDNA encoding AtSOT18 (At1g74090) was amplified from a phage library containing genomic DNA. The *SOT* gene for AtSOT18 does contain an intron. The cDNA sequence was cloned in frame downstream (pGFP-C) or upstream (pGFP-N) of the green fluorescent protein (GFP) reading frame. The gene-cassettes were driven by the CaMV-35S promoter with a double enhancer and the polyA-tail from CaMV-35S. *A. thaliana* protoplasts were transiently transformed with the GFP constructs as described in Bauer and co-workers (2004). Bright field images, shown in (A, C, E, G), were made to visualize the protoplast's cell borders and the chloroplasts. Fluorescence images of the protoplasts, shown in (B, D, F, H), were taken using an Axioskop microscope with filter sets optimal for GFP fluorescence (BP 450–490/LP 520). As a control, the protoplasts, shown in (A, B, E, F), were transformed with the pGFP-C and the pGFPN vector, respectively, without an insert. All scale bars represent 10 µm.

Expression studies

So far the results concerning gene expression and protein steady-state levels of *A. thaliana* SOTs are very limited. Only for RaR047 (AtSOT12, At2g03760) was mRNA accumulation determined in different conditions. The mRNA coding for RaR047/AtSOT12 was expressed during active growth of *A. thaliana* cell cultures and in the aerial parts of young seedlings, but not in roots. Treatment of *A. thaliana* seedlings with hormonal or stress-related compounds showed that mRNA for RaR047/AtSOT12 was induced in response to salicylic acid and methyl jasmonate. Infection with avirulent bacterial pathogens causing an hypersensitive response increased the RaR047/AtSOT12 mRNA levels (Lacomme and Roby, 1996). The expression of AtSOT14 mRNA (At5g07000) was induced following treatment of *A. thaliana* rosette leaves with methyljasmonate and 12-hydroxyjasmonate, the substrate for the AtSOT14 protein. However, the expression of the gene encoding the thionin Thi2.1 protein, specifically induced by wounding, pathogen infection, and methyljasmonate treatment, was not induced by 12-hydroxyjasmonate, indicating two independent response pathways mediated by methyljasmonate and 12-hydroxy-jasmonate (Gidda *et al.*, 2003). So far the experiments described above are the only results on expression levels of SOTs in *A. thaliana.*

There are numerous results of microarray analyses available in the databases performed under a broad variety of conditions. The authors' own microarray data of an experiment with *A. thaliana* seedlings treated with methyljasmonate for several hours indicated differential expression of the *A. thaliana* SOT genes included on the chip. Some genes were up-regulated, others were down-regulated with expression maxima at different time points (R Jost and J Papenbrock, unpublished results). A detailed search of microarray analyses using the 18 *A. thaliana* SOT genes might help to characterize the expression pattern and to form groups of similarly expressed SOT genes.

How to identify the respective substrates and their function?

It is a challenging task to analyse the respective substrates for each SOT. The study of the structure/function relationship of SOT proteins in order to elucidate the mechanism of sulphonate transfer and to define the amino acids responsible for substrate binding and catalysis will help to identify putative substrates. Using site-directed mutagenesis of the flavonol 3-SOT, several amino acids required for catalysis and cosubstrate binding were mapped, while the construction of chimeric proteins allowed definition of the domain responsible for substrate specificity (Varin *et al.*, 1995). A domain was found which is involved in substrate binding, designated as domain II. Within this domain, two subdomains of high divergence were identified, probably participating in the

recognition and binding of different acceptor substrates (Marsolais and Varin, 1997, 1998; Varin *et al*., 1995). These domains of high divergence could be used for modelling ligands which might serve as substrates using docking programs.

To characterize the biochemical function of SOT proteins one could think about all putative substrate candidates which might be found in plants in the desulpho- and in the sulphonated form, such as coumarins, desulphoglucosinolates, flavonoids, gibberellic acids, hydroxyjasmonates, phenolic acids, steroids, sulphate esters, such as choline-*O*-sulphate, and terpenoids, and test them by using recombinant SOT proteins. This idea was followed by building up a substrate library comprising more than 100 desulpho-derivatives of most of the known plant-sulphonated metabolites as well as a collection of metabolites for which no sulphonated metabolites have been reported (Gidda *et al*., 2003).

In the review written by Marsolais *et al.* (2000), the putative substrates and functions of two more *A. thaliana* SOT proteins are mentioned as unpublished results of the same group. The substrates already identified for other SOT proteins were tested using two *A. thaliana* SOT proteins. The purified recombinant AtSOT5 (At3g45070) was found to exhibit strict specificity for position 7 of flavonoids. The natural occurrence of a SOT protein exhibiting high specificity for flavonoids in *A. thaliana* was surprising considering the absence of reports on the presence of flavonoid sulphates in this plant. It was hypothesized that flavonoid sulphates may act as regulators of polar auxin transport (Faulkner and Rubery, 1992). In the same review AtSOT10 (At2g14920) was shown to exhibit strict specificity for brassinosteroids having 22*R*-, 23*R*-hydroxyls, and a 24*S*-methyl or ethyl group on the steroid side chain. Due to the high sequence identity to the *B. napus* SOTs, AtSOT12 was suggested to be involved in stereospecific inactivation of brassinosteroid by sulphonation.

The rare expression of the *A. thaliana* SOT genes might suggest that plants use the sulphonation reaction to modulate the biological activity of hormones and messengers molecules under special conditions where only low amounts of protein are necessary. The presence of the sulphate group might suggest a role in signalling processes, as shown for other sulphate metabolites such as the Nod factors in the interaction between *Rhizobium meliloti* and *Medicago sativa* L. (Truchet *et al*., 1991). However, so far no sulphatase-like sequences could be identified in plant genomes. In the enzyme databases (http://au.expasy.org/enzyme) there are already 29 different SOT protein reactions described giving more ideas for substrates. Thus, there are a number of strategies to follow for the elucidation of the SOT's physiological role.

Experimental evidence for an involvement of a SOT in phytosulphokine biosynthesis

Phytosulphokine-α (PSK-α) is a sulphonated pentapeptide (Tyr(SO_3H)-Ile-Tyr(SO_3H)-Thr-Gln) and was shown to act as a plant growth factor (Matsubayashi and Sakagami, 1996). The biosynthetic pathway of a preprophytosulphokine was partially elucidated in *Oryza sativa* L. and *A. thaliana* (Yang *et al*., 1999; 2001). The PSK-α is only active in its sulphonated form, thus the identification of a tyrosyl peptide SOT (EC 2.8.2.20) is of high importance for the understanding of PSK-α and related peptides in the plant organism. Protein tyrosyl *O*-sulphation is one of the posttranslational modifications that occurs with many secretory and membrane proteins in animal cells. In mammals, a tyrosyl protein SOT was localized specifically in the *trans*-Golgi-network (Lee and Huttner, 1983). In cell cultures of a number of plant species the existence of a tyrosyl protein SOT activity was demonstrated also in the Golgi-network (Hanai *et al*., 2000). However, the sequence of the respective protein has not been determined so far (Youji Sakagami, Nagoya, Japan, personal communication). The sequence of the mammalian tyrosyl protein SOT does not show any similarity to a DNA or protein sequence from a plant species. Therefore, it might be possible that one of the 18 *A. thaliana* SOT acts as tyrosyl protein SOT.

Involvement of sulphotransferases in glucosinolate biosynthesis

Glucosinolates are found in vegetative and reproductive tissues of 16 plant families, but are most well known as the major secondary metabolites in the Brassicaceae (Mithen, 2001; Mikkelsen *et al*., 2002). In *A. thaliana* more than 20 different glucosinolates have been identified (Mithen 2001; Reichelt *et al*., 2002). Glucosinolates and their degradation products have a wide range of biological activities, for example, in plant defense as deterrents against insect and fungi. Their biosynthetic pathways are partially identified (Mikkelsen *et al*., 2002; Reichelt *et al*., 2002). Interestingly, glucosinolates contain two forms of sulphur in different oxidation states. The reduced form is a thioether and is derived from cysteine (Mikkelsen *et al*., 2002; Reichelt *et al*., 2002) whereas the oxidized form is a sulphate ester and is derived from the sulphonation pathway. The last step from the different aliphatic, aromatic, and indole desulphoglucosinolates to the active glucosinolates might be catalysed by members of the SOT family. Glendening and Poulton (1990) partially purified a protein from *Lepidium sativum* L. that had PAPS-dependent desulphoglucosinolate SOT activity, however, no molecular data are available to date. AtSOT16, AtSOT17, and AtSOT18 were suggested as being involved in sulphonation of desulphoglucosinolates and might be used to modulate the glucosinolate pattern of plants (patent WO 03010318-A, L. Varin

and D. Spertini, 06/02/2003, Concordia University, Canada). However, at the moment the results of the substrate specificity of the three SOT proteins mentioned are not publicly available. The approach has great potential for design of metabolically engineered plants with improved pest resistance and increased nutritional value.

A. thaliana as a model plant: suited for the elucidation of all SOT functions?

In the almost fully sequenced genome of *A. thaliana*, 18 SOT encoding genes have been identified while in the close relative *B. napus*, at least 12 genes were detected (Marsolais *et al*., 2000), and the genome of the monocotyledonous plant *O. sativa*, about 3.5-times larger in size than the *A. thaliana* genome, contains 13 SOT genes (M Klein and J Papenbrock, unpublished results). Thus, the SOT gene number is about 1.5-times higher in *A. thaliana*. However, similar genes might have evolved from gene duplications and might be functionally redundant or silent. Evaluation of already available microarray analyses will help to characterize expression patterns. Metabolic profiling of desulpho- and sulphonated compounds in T-DNA insertion mutants available for almost all *A. thaliana* SOT genes and in mutants obtained with the RNAi techniques in combination with exact observation of the physiological phenotype should be very successful for analysing the function of the members of this protein family in *A. thaliana*. On the other hand, detection of the sulphonation reaction of choline in halophytes and of gallic acid glucoside involved in the seismonastic response shows the limitations of the model plant *A. thaliana*. Results of expression and metabolic profiling will be required for the analysis of the functional role of SOT proteins in non-*A. thaliana* species.

Chapter 3:
Elucidation of Gene-to-Gene and Metabolite-to-Gene Networks in *Arabidopsis* by Integration of Metabolomics and Transcriptomics

Hirai MY, Klein M, Fujikawa Y, Yano M, Goodenowe DB, Yamazaki Y, Kanaya S, Nakamura Y, Kitayama M, Suzuki H, Sakurai N, Shibata D, Tokuhisa J, Reichelt M, Gershenzon J, Papenbrock J, Saito K

The Journal of Biological Chemistry (2005) Vol. 280, 25590–25595

Abstract

Since the completion of genome sequences of model organisms, functional identification of unknown genes has become a principal challenge in biology. Postgenomics sciences such as transcriptomics, proteomics, and metabolomics are expected to discover gene functions. This report outlines the elucidation of gene-to-gene and metabolite-to-gene networks via integration of metabolomics with transcriptomics and presents a strategy for the identification of novel gene functions. Metabolomics and transcriptomics data of *Arabidopsis* grown under sulfur deficiency were combined and analyzed by batch-learning self-organizing mapping. A group of metabolites/genes regulated by the same mechanism clustered together. The metabolism of glucosinolates was shown to be coordinately regulated. Three uncharacterized putative sulfotransferase genes clustering together with known glucosinolate biosynthesis genes were candidates for involvement in biosynthesis. *In vitro* enzymatic assays of the recombinant gene products confirmed their functions as desulfoglucosinolate sulfotransferases. Several genes involved in sulfur assimilation clustered with *O*-acetylserine, which is considered a positive regulator of these genes. The genes involved in anthocyanin biosynthesis clustered with the gene encoding a transcriptional factor that up-regulates specifically anthocyanin biosynthesis genes. These results suggested that regulatory metabolites and transcriptional factor genes can be identified by this approach, based on the assumption that they cluster with the downstream genes they regulate. This strategy is applicable not only to plant but also to other organisms for functional elucidation of unknown genes.

Introduction

In the era of post-genomics, a systematic and comprehensive understanding of the complex events of life is a great concern in biology. The first step in this process is to identify all gene functions and gene-to-gene networks as the components of the system, the whole events of life. The metabolome is the final product of a series of gene actions. Hence, metabolomics has a potential to elucidate gene functions and networks, especially when integrated with transcriptomics. A promising approach is pair-wise transcript-metabolite correlation analysis, which can reveal unexpected correlations and bring to light candidate genes for modifying the metabolite content (Urbanczyk-Wochniak *et al*., 2003). Gene functions involved in the specific pathway of interest have been identified by the integration of transcript and targeted metabolic profiling in experimental systems where the pathway was activated (Aharoni *et al.*,2000; Guterman *et al.*, 2002; Goossens *et al.*, 2003; Mercke *et al.*, 2004; Tohge *et al.*, 2005). However, up to now, no gene function has been identified by non-targeted analyses of the transcriptome and metabolome. In this report, we analyzed the non-targeted metabolome and transcriptome of a model plant *Arabidopsis* under sulfur (S)1 deficiency whose genome sequencing has been completed. Our strategy for integrated analyses using batch-learning-selforganizing mapping (BL-SOM) (Kanaya *et al.*, 2001; Abe *et al.*, 2003; Hirai *et al.*, 2004) enabled the identification of gene-to-gene and metabolite-to-gene networks and new gene functions.

Experimental Procedures

Plant Materials

Arabidopsis thaliana ecotype Columbia was grown for 21 days on agar-solidified S-sufficient medium following the methodology as described by Hirai *et al.* (2003). Plants were transferred to S-sufficient or S-deprived medium and grown for up to 1 week (168 h). Rosette leaves and roots were harvested at 3, 6, 12, 24, 48, and 168 h after transfer, immediately frozen with liquid nitrogen, and stored at -80 °C until use.

Metabolome Analyses

Fourier-transform ion cyclotron resonance mass spectrometry analysis was used to conduct non-targeted metabolomic profiling (Tohge *et al.*, 2005). Targeted metabolic profiling of amino acids, *O*-acetyl-L-serine (OAS), anions, organic acids, and sugars was performed by high performance liquid chromatography and capillary electrophoresis as described (Hirai *et al.*, 2004). Extraction of metabolites was conducted in triplicate from each sample. Among 2,123 metabolites detected by targeted and non-

targeted analyses, 84 metabolites whose coefficient of variation was greater than 0.9 were eliminated. For each metabolite the logarithm of the ratio of the average signal intensity of S-starved samples to that of the control samples was calculated.

Transcriptome Analyses

The transcriptomes were analyzed using the Agilent Arabidopsis 2 microarray (Agilent Technologies, Palo Alto, CA), which carries 21,500 *Arabidopsis* genes, according to the manufacturer's specifications. Data were acquired using Agilent Feature Extraction software. Normalization of log ratio of expression intensity between S-starved and control samples was carried out based on MA plot (Dudoit *et al.*, 2002; Quackenbush, 2002). Initially, log ratio M_i [=log(R_i/G_i)] and average of logarithmic intensity A_i [=(logR_i + log G_i)/2] were calculated for *i*th gene. Here, R_i and G_i are differences between mean signal and mean background intensities for Cy5 dye (S-starved sample) and for Cy3 dye (control sample), respectively, obtained by Agilent Feature Extraction software. Normalized log ratio M_i" was estimated as the difference between M_i and baseline M_i'. Here, using a relation between M_i and A_i, (M_i = f(A_i) + $\varepsilon_{I,}$ ε_i was the difference between M_i and f(A_i) for gene *i*) by MA plot; the baseline for *i*th gene was estimated by M_I' = f(A_i). The genes whose signal intensity was regarded as zero were eliminated in the present analysis.

BL-SOM Analyses

BL-SOM analyses were conducted according to Hirai *et al.* (2004). The metabolites and transcripts whose maximum log ratio through time course was <0.1 and minimum log ratio was >0.1 were eliminated. For ~1,000 metabolites and ~10,000 transcripts left after the elimination, the sum of the square of the 6 log ratio values at 6 time points was set equal to 1 to give relative log ratio values, and all data were combined into a single matrix to be subjected to BL-SOM. ~1,000 metabolites and ~10,000 genes were classified into 40 x 29 cells in the lattice formed by BL-SOM based on the time-dependent pattern of accumulation/expression in leaves in response to -S (Fig. 1A). In the same way, ~1,000 metabolites and ~10,000 genes were classified into 40 x 24 cells in the lattice based on the time-dependent pattern of accumulation/expression in roots in response to -S (Fig. 1B). Each cell contained ~10 metabolites and/or genes on the average. Each cell was colored according to the relative log ratio values of metabolites/genes in it. When all of the relative log ratio values of metabolites/genes in the cell were greater or smaller than the average, the cell was colored in pink or pale blue, respectively. Red and blue indicated that at least one of the relative log ratio values was greater than the average + S.D. or smaller than the average - S.D., respectively.

DNA Cloning Techniques

The gene encoding sulfotransferase 18 (At1g74090) from *A. thaliana* (*AtSOT18*) (for nomenclature, see Klein and Papenbrock, 2004) does not contain any introns. A 1053-bp cDNA encoding *AtSOT18* was amplified from the λEMBL3 genomic library of *A. thaliana* ecotype Columbia (Clontech) with primer 5'-GGA TCC GAA TCA GAA ACC CTA-3' extended by a BamHI restriction site and primer 5'-AAG CTT TTT ACC ATG TTC AAG C-3' extended by a HindIII restriction site. The PCR contained 0.2 mM dNTPs (Roth, Karlsruhe, Germany), 0.4 µM each primer (MWG, Ebersberg, Germany), 1 mM $MgCl_2$, 0.75 µl of Red *Taq* DNA polymerase (Sigma), and 1 µg of template DNA in a final volume of 50 µl. Before starting the first PCR cycle, the DNA was denatured for 180 s at 94 °C followed by 28 PCR cycles conducted for 60 s at 94 °C, 60 s at 48 °C, and 60 s at 72 °C. The process was finished with an elongation phase of 420 s at 72 °C. The amplified PCR fragment was ligated into the expression vector pQE-30 (Qiagen, Hilden, Germany), and the vector was introduced into the *Escherichia coli* strain XL1-blue.

Expression and Purification of the AtSOT18 Protein in *E. coli*

The recombinant protein was expressed according to the following protocol. After growth of the *E. coli* culture at 37 °C to an A_{600} of 0.6 in Luria Bertani medium containing ampicillin (100 µg ml^{-1}) (AppliChem, Darmstadt, Germany), induction was carried out for 2 h at 30 °C with 1 mM (final concentration) isopropyl-β-D-1-thio-galactopyranoside (AppliChem). Cell lysis was carried out by adding lysozyme (final concentration 1 mg ml^{-1}) (Roth) and vigorously homogenizing using a glass homogenizer and a pestle. The recombinant AtSOT18 protein was purified under non-denaturing conditions by affinity chromatography with the nickel affinity resin according to the manufacturer's instructions (Qiagen), dialyzed overnight, and used for enzyme activity measurements. The purity of the protein preparation was checked by SDS-polyacrylamide gel electrophoresis and subsequent Coomassie brilliant blue and silver staining.

Substrate Preparation and Enzyme Activity Measurement

The desulfo form of the intact allyl GLS, sinigrin (Sigma), was prepared according to Graser *et al.* (2000 ; 2001). The enzyme assay with recombinant AtSOT18 protein was set up in the manner of Glendening and Poulton (1990). 15 µg of purified recombinant protein was used in each reaction. The 300-µl assays contained: 83 mM Tris/HCl, pH 9.0, 9.2 mM $MgCl_2$, and 58 µM 3'-phosphoadenosine 5'-phosphosulfate (PAPS; Calbiochem), and 6.2 mM desulfo-allyl GLS. The reaction was started by the addition of PAPS, incubated for 30 min at 30 °C, and stopped by incubation for 20 min at 95 °C.

The formation of intact allyl GLS was analyzed by high performance liquid chromatography according to Mellon *et al.* (2002). Separation was achieved with a gradient of 0.1% trifluoroacetic acid and acetonitrile with 0.1% trifluoroacetic acid on an RP-C18 column (Supelcosil LC18-DB, 250 x 4.6 mm, 5 µm). The product was identified by its retention time, absorption spectrum, and mass spectrum as described by Mellon *et al.* (2002). Quantification of the product formation was done with a calibration curve prepared using authentic sinigrin.

Results and Discussion

Time-dependent changes in the metabolome and the transcriptome were analyzed in a non-targeted way after *Arabidopsis* plants were subjected to S deprivation for up to 168 h. ~2,000 metabolites detected by targeted and non-targeted analyses and 21,500 transcripts by DNA microarray were used for calculation, and each was expressed as a vector in sixdimensional space, where six components of the vector were six log ratio values of signal intensities under S deficiency compared with those under control condition at six time points.

For integrated analysis of metabolome and transcriptome, we used BL-SOM, which classified the metabolites and the transcripts according to their time-dependent pattern of changes in accumulation and expression. BL-SOM is a sophisticated form of multivariate analyses that can classify metabolites and transcripts into cells on a two-dimensional lattice; those showing similar patterns are clustered into the same or the neighboring cells. After elimination of the metabolites and the transcripts exhibiting almost no change (see “Experimental Procedures”), the sum of the square of 6 log ratio values was set to 1. This procedure made it possible to classify the metabolites and the transcripts by the shape of the time-dependent changing pattern alone and not by the absolute value of the degree of change. All vectors (corresponding to the metabolites and the transcripts) left after elimination were combined into a single matrix to be subjected to BL-SOM. The results are shown in Fig. 1, A (the changes in leaf) and B (in root). By this analysis, sets of metabolites and transcripts with strong correlations were elucidated.

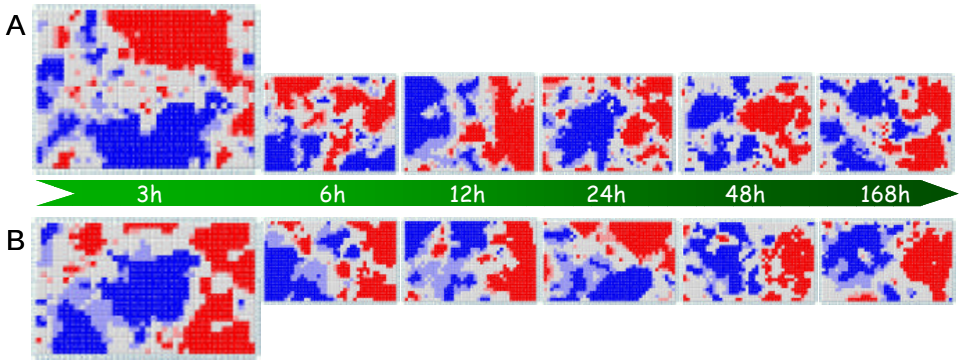


Fig. 1. BL-SOM analysis of the leaf (A) and root (B) samples. The map is a lattice comprised of 40 x 29 (A) and 40 x 24 (B) cells. The metabolites and genes were classified into each cell according to their time-dependent pattern of changes in accumulation and expression. The color of each cell indicates the level of induction/repression of the metabolites and genes under S deficiency at the given time points. When all of the relative log ratio values of metabolites/genes in the cell were greater or smaller than the average, the cell was colored in pink or pale blue, respectively. Red and blue indicated that at least one of the relative log ratio values was greater than the average + S.D. or smaller than the average - S.D., respectively.

Genes Involved in the Same Metabolic Pathway

Six glucosinolates (GLSs) and four isothiocyanates, which are the degradation products of GLSs, clustered into the cells in the specific regions, suggesting that GLS metabolism is coordinately regulated (Fig. 2, A and B). GLSs are sulfur- and nitrogen-containing secondary metabolites found mainly in the Capparales order and are important as defense compounds to pathogens and herbivores and as S storage sources (Wittstock and Gershenzon, 2002; Wittstock and Halkier, 2002). They are synthesized from amino acids such as tryptophan, tyrosine, and methionine homologs with elongated side chains. The core biosynthetic pathway of GLS has been elucidated, and the *Arabidopsis* genes encoding most of the enzymes have been identified or at least assumed, except for those encoding the sulfotransferases (Mikkelsen *et al.*, 2002; Field *et al.*, 2004; Mikkelsen *et al.*, 2004) (Fig. 3).

The genes encoding the GLS biosynthesis enzymes, including those encoding the *MAM*, *CYP79*, and *CYP83* families and *SUR1* were clustered into the cells in the same region (Fig. 2, B and C). It is notable that three putative sulfotransferase genes (At1g74100, At1g18590, and At1g74100 named *AtSOT16*, *AtSOT17*, and *AtSOT18*, respectively, by Klein and Papenbrock (Klein and Papenbrock 2004) were classified into the same region where the known GLS biosynthetic genes were clustered (Fig. 2, B and C). This result strongly suggested that these putative sulfotransferase genes are also involved in GLS biosynthesis. To confirm the function of these putative sulfotransferase gene products, *in vitro* sulfotransferase assays were conducted using the respective recombinant proteins. The recombinant AtSOT18 protein could convert desulfo-allyl GLS to intact, sulfonated allyl GLS in the presence of PAPS (Fig. 2E). The identity of the product was confirmed by mass spectrometry (data not shown). The activity of the

gene products of the two other putative sulfotransferase genes (*AtSOT16* and *AtSOT17*) has been analyzed in the same way, giving comparable results with desulfo-allyl GLS (data not shown). These results confirmed that the three sulfotransferase-like genes, which clustered with known GLS biosynthesis genes after BL-SOM analysis, actually encode the proteins catalyzing PAPS-dependent sulfation of desulfo-GLSs to GLSs. During the reviewing process of this publication, Piotrowski *et al.* (2004) reached the same conclusion by quite different strategy to ours and reported that these three genes encode desulfo-GLS sulfotransferase. It was also the case with At1g24100, which was assumed to encode S-glucosyltransferase involved in GLS biosynthesis (Petersen *et al.*, 2001). This gene (Figs. 2 and 3, *S-GT*) was clustered with known GLS biosynthesis genes, supporting the assumption by Petersen *et al.* (2001). Recently, the function of this gene was confirmed by Douglas Grubb *et al.* (2004). These facts proved that our strategy is effective to elucidate gene functions in the same metabolic pathway at once.

In a similar way, based on the results of BL-SOM the functions of other candidate genes for GLS biosynthesis enzymes (C-S lyase and glutathione *S*-transferase) were also putatively identified. A putative tyrosine aminotransferase gene (At5g36160) and two putative glutathione *S*-transferase (GST) genes (At3g03190 and At1g78370) were also clustered together with the known GLS biosynthesis genes. The putative tyrosine aminotransferase gene At5g36160 may represent a C-S lyase gene involved in GLS biosynthesis. The *SUR1* gene (Fig. 3), whose gene product is the C-S lyase of the core GLS biosynthetic pathway (Mikkelsen *et al.,* 2004), had also been originally misannotated as a tyrosine aminotransferase (Jones *et al.*, 2003). The *sur1* mutant does not accumulate GLS, at least under normal conditions, suggesting no apparent functional redundancy of C-S lyase of GLS biosynthesis (Mikkelsen *et al.*, 2004). However, this mutant occasionally accumulated a trace level of indol-3-ylmethyl GLS (Mikkelsen *et al.*, 2004), and so the product of At5g36160 might also function as C-S lyase in GLS biosynthesis under different environmental or developmental conditions.

As for the putative GST genes clustering with the other GLS biosynthesis genes, it has been suggested that GST-type enzymes may be involved in an enzyme complex formed by CYP83s and C-S lyase. The S-alkylthiohydroximate formed after CYP83-catalyzed aldoxime oxidation and spontaneous conjugation to cysteine (Fig. 3) is cyclized *in vitro* to form a dead-end product. Hence, metabolic channeling aided by GST-type enzymes is postulated *in vivo* to avoid this consequence (Mikkelsen *et al.*, 2004). The two putative GST genes (At3g03190 and At1g78370) could be candidates coding for such an activity. We are now examining the genes clustered with GLSs and isothiocyanates for identification of new factors regulating GLS metabolism.

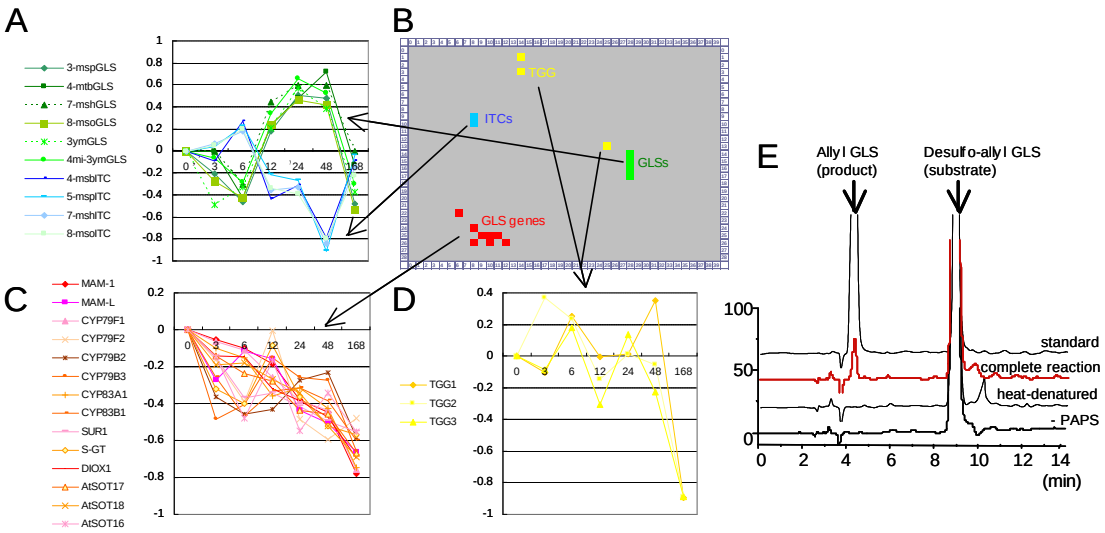


Fig. 2. A, time-dependent changes in accumulation of GLSs (green) and isothiocyanates (blue) in leaves under S deficiency. The ordinate scale indicates the relative log ratio values after the sum of squares of 6 log ratio values were set to 1. 3-msp, 3-methylsulfinylpropyl; 4-mtb, 4-methylthiolbutyl; 7-msh, 7-methylsulfinylheptyl; 8-mso, 8-methylsulfinyloctyl; 3ym, indol-3-ylmethyl; 4mi-3ym, 4-methoxyindol-3-ylmethyl; 4-msb, 4-methylsulfinylbutyl; 5-msp, 5-(methysulfinyl)pentyl.

B, unified map of the leaf samples (the same as Fig. 1A) showing the clustering of GLSs (green), isothiocyanates (blue), GLS biosynthesis genes (red), and the degrading enzyme myrosinase genes (yellow) into regions.

C and D, time-dependent changes in expression of GLSs biosynthesis genes (C) and myrosinase genes *TGG 1–3* (D) under S deficiency. The ordinate scale indicates the relative log ratio values. *DIOX1* is the gene involved in side chain modification of GLS.

E, high performance liquid chromatography trace of product of *in vitro* enzymatic assay of AtSOT18 recombinant protein. Standard, standards of desulfo-allyl GLS (substrate) and intact allyl GLS (product); complete reaction, product formation in the complete reaction mixture; heat-denatured, reaction mixture with heat-denatured AtSOT18 protein; -PAPS, reaction mixture with AtSOT18 protein but without PAPS.

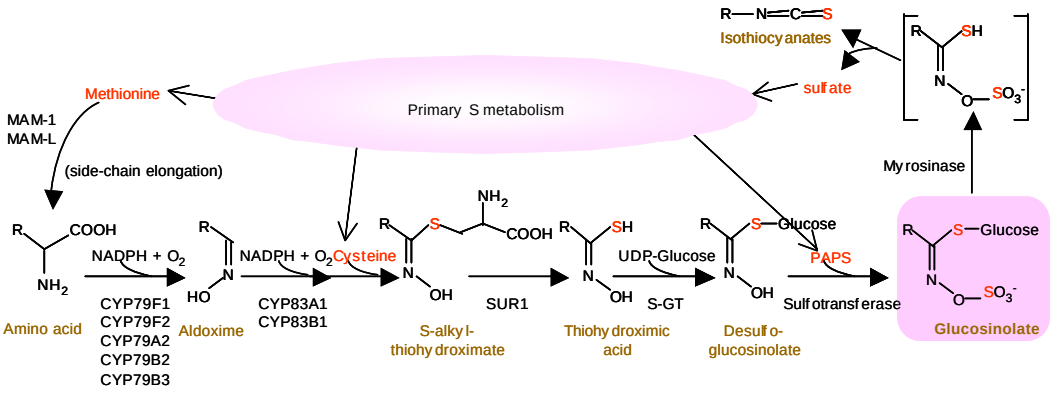


Fig. 3. Outline of GLS biosynthesis and degradation. MAM, methylthioalkylmalate synthase; CYP, cytochrome P450; SUR1, a C-S lyase encoded by *SUPERROOT1* gene; S-GT, S-glucosyltransferase; PAPS, 3'-phosphoadenosine 5'-phosphosulfate. In *Arabidopsis*, methionine and tryptophan are the major precursors of GLSs, but the side chain of methionine is first elongated by a cycle involving MAM-1 and MAM-L. Isoforms of CYP79s and CYP83s catalyze the initial reactions of the core biosynthetic pathway, followed by SUR1, S-GT, and sulfotransferase.

Regulatory Metabolite O-Acetylserine and Genes under Its Regulation

Primary S metabolism is regulated by modulation of gene expression (see below) and of enzymatic activity (Saito, 2004). For example, the activities of serine acetyltransferases (Fig. 4A, *Serat*) are regulated by two mechanisms in an isoformspecific manner: by allosteric feedback inhibition by cysteine and by reversible formation of a protein complex with OAS-(thiol)-lyase (Fig. 4A, *Bsas*). The enzymes involved in primary S metabolism are encoded by gene families in *Arabidopsis* (Fig. 4A). These genes were scattered on the map (Fig. 4B), which was consistent with the fact that primary S metabolism is regulated not only by mRNA accumulation but also by enzymatic activity (Saito, 2004). Among them, however, several genes were clustered together with OAS (Fig. 4, B and C). Sulfate transporter (*Sultr*) and adenosine phosphosulfate reductase (*APR*) genes are known to be induced by S deficiency. OAS, whose content increases under -S, is considered a positive regulator of this induction mechanism (Fig. 4A) (Saito, 2004). The induction of *Sultr* and *APR* genes enables more sulfate ions to be assimilated into cysteine. By BL-SOM, these genes and OAS were clustered into the same region (Fig. 4, B and C), confirming the previous findings. Among the isoforms of ATP sulfurylase (*APS*) and *Serat* genes, *APS3*, *Serat 3;1* and *Serat 3;2* were clustered with OAS (Fig. 4, B and C and supplemental Fig. 1), suggesting their specific functions under S deficiency (see below).

Besides S assimilation genes, the 12-oxophytodienoate reductase 1 gene and the nitrilase 3 gene are known to be induced under S deficiency and by OAS (Kutz *et al.*, 2002; Hirai *et al.*, 2003; Nikiforova *et al.*, 2003). They were also in the same region where OAS clustered (data not shown). Two putative thioglucosidase genes, which are known to be induced by S deficiency (Maruyama-Nakashita *et al.*, 2003), were also in the same region, suggesting that they are coordinately regulated with OAS.

Moreover, the specific function of each member of a gene family could be estimated by this analysis. The ATP sulfurylase isoforms *APS2* and *APS4* among four members of this gene family were clustered with GLS biosynthesis genes (Fig. 4B and supplemental Fig. 1), suggesting that these isoforms may be specialized for the biosynthesis of PAPS required for GLS biosynthesis. On the other hand, as mentioned above, *APS3*, which clustered with OAS, may be involved in the enhancement of cysteine biosynthesis under -S.

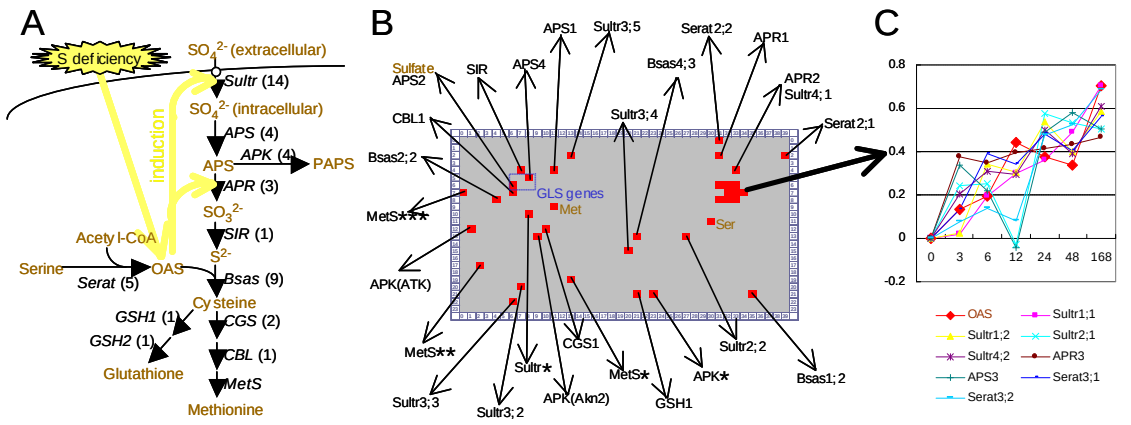


Fig. 4. A, outline of primary S metabolism. *Sultr*, sulfate transporter; *APS*, ATP sulfurylase; *APR*, APS reductase; *APK*, APS kinase; *SIR*, sulfite reductase; *Serat*, serine acetyltransferase; *Bsas*, O-acetylserine(thiol)-lyase (a member of the β-substituted alanine synthase family); *CGS*, cystathionine γ synthase; *CBL*, cystathionine β lyase; *MetS*, methionine synthase; *GSH1*, γ glutamylcysteine synthetase; *GSH2*, glutathione synthetase. The numbers in parentheses indicates the numbers of isoform genes.

B, unified map of the root samples (the same as Fig. 1B) showing clustering of genes encoding enzymes of primary S metabolism. Gene names are shown except for Sultr*, At1g80310; APK*, At5g67520; MetS*, At5g17920; MetS**, At5g20980; and, MetS***, At3g03780. Blue rectangle indicates the region where known GLS biosynthesis genes are clustered.

C, time-dependent changes in OAS accumulation and gene expression. The ordinate scale indicates the relative log ratio values.

Transcription Factor and Downstream Genes

PAP1 is a Myb transcription factor that activates phenylpropanoid/flavonoid biosynthesis (Borevitz *et al.*, 2000). In an activation-tagged mutant of the *PAP1* gene that overaccumulates red-purple pigment anthocyanins, the biosynthesis genes of flavonoids such as phenylalanine ammonia lyase, chalcone synthase, and dehydroflavonol 4-reductase were induced (Borevitz *et al.*, 2000). We clarified that in this mutant line anthocyanin biosynthesis genes were specifically induced among flavonoid biosynthesis genes, suggesting that PAP1 is a transcription factor of anthocyanin biosynthesis genes (Tohge *et al.*, 2005). In this study *PAP1* and the anthocyanin biosynthesis genes were clustered together (Fig. 5, B and C), indicating that transcription factors and the downstream genes regulated by them can be elucidated by these analyses. Our plant materials subjected to S deficiency did not turn red, indicating anthocyanin biosynthesis was not apparently induced. Nevertheless, *PAP1* and the anthocyanin biosynthesis genes were clustered together. This fact implies that a new regulatory network can be clarified by our strategy even when we do not focus on a specific pathway.

For GLS biosynthesis, several putative transcription factor genes were clustered with GLS biosynthesis genes (data not shown). These genes are the candidate genes controlling GLS biosynthesis.

In the present study, we could find gene-to-gene and metabolite-to-gene networks and could identify a new gene function through integrated analysis of metabolomics and transcriptomics. Our strategy may be useful especially in cases where the gene of interest is functionally redundant and thus no visible changes are observed in knocked-out lines of the gene (Bino *et al.*, 2004; Weckwerth *et al.*, 2004). This approach is generally applicable not only to plants but also to other organisms, including bacteria and animals, for the identification of novel gene functions.

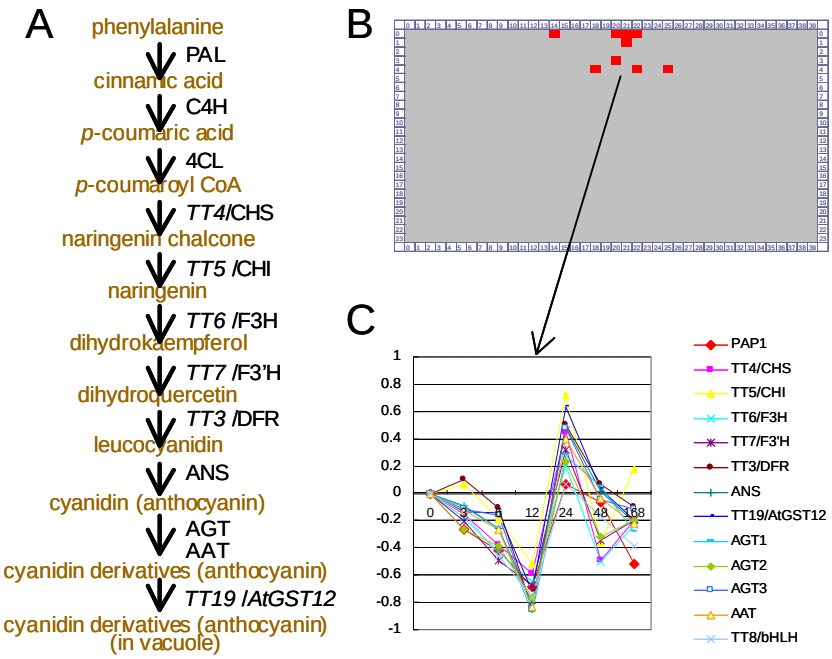


Fig. 5. A, outline of anthocyanin biosynthesis. *PAL*, phenylalanine ammonia lyase; *C4H*, cinnamate 4-hydroxylase; *4CL*, 4-coumarate-CoA ligase; *TT4/CHS*, chalcone synthase; *TT5/CHI*, chalcone isomerase; *TT6/F3H*, flavanone 3-hydroxylase; *TT7/F3'H*, flavonoid 3'-hydroxylase; *TT3/DFR*, dihydroflavonol 4-reductase; *ANS*, anthocyanidin synthase; *AGT 1–3*, anthocyanin glucosyltransferases 1 (At4g14090) (6), 2 (At5g17050) (6), and 3, identified by T. Tohge, Y. Nishiyama, M. Yamazaki, and K. Saito, manuscript in preparation; *AAT*, anthocyanin acyltransferase identified by Y. Nishiyama, T. Tohge, Y. Tanaka, M. Yamazaki, and K. Saito, manuscript in preparation; *TT19/AtGST12*, glutathione S-transferase.

B, unified map of the root samples (the same as Fig. 1B) showing clustering of the *PAP1* transcription factor gene and anthocyanin biosynthesis genes.

C, time-dependent changes in gene expression. The ordinate scale indicates the relative log ratio values TT8/bHLH, At4g09820.

Chapter 4:
The three desulfoglucosinolate sulfotransferase proteins in *Arabidopsis* have different substrate specificities and are differentially expressed

Klein M, Reichelt M, Gershenzon J, Papenbrock J

FEBS Journal (2006) Vol. 273: 122-136

Abstract

Sulfotransferases (SOTs) catalyse the transfer of a sulfate group from 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to an appropriate hydroxyl group of various substrates with the parallel formation of 3'-phosphoadenosine 5'-phosphate. In *Arabidopsis thaliana*, 18 SOT proteins (AtSOT) have been identified. Three of them, AtSOT16, AtSOT17 and AtSOT18, catalyse the sulfation of desulfoglucosinolates. The proteins were expressed in *Escherichia coli*, purified by affinity chromatography and used for enzyme kinetic studies. By establishing two types of enzyme assay using both ^{35}S-labelled and unlabelled PAPS, separation of the products by HPLC, and detection of the products by monitoring radioactivity or UV absorption, the substrate specificities of the three AtSOT proteins were determined. They show different maximum velocities with several desulfoglucosinolates as substrates and differ in their affinity for desulfobenzylglucosinolate and PAPS. The sequences encoding AtSOT18 were amplified from *Arabidopsis* ecotypes C24 and Col0; the two expressed proteins differ in two out of 350 amino acids. These amino-acid variations led to different substrate specificities. Exchange of one of the two amino acids in AtSOT18 from C24 to the respective amino acid in AtSOT18 from Col0 gave the C24 protein the same substrate specificity as the wild-type AtSOT18 protein from Col0. All three desulfo-glucosinolate AtSOT proteins are localized in the cytoplasm, as demonstrated by transient expression of fusion constructs with the green fluorescent protein in *Arabidopsis* protoplasts. Northern blot analysis indicated differential expression of the three *AtSOT* genes in plant organs and tissues at different developmental stages and during a light/darkness cycle. High (500 µM) and low (50 µM) sulfate concentrations in the medium did not influence the levels of expression.

Introduction

Glucosinolates are a group of over 130 nitrogen-containing and sulfur-containing natural products found in vegetative and reproductive tissues of 16 plant families, but are most well known as major secondary metabolites in the Brassicaceae (Mithen, 2001; Mikkelsen *et al.*, 2002). In *Arabidopsis thaliana* (L.) Heynh., nearly 30 different glucosinolates have been described (Hogge *et al.*, 1988; Mithen, 2001; Reichelt *et al.*, 2002) with considerable variation found among ecotypes (Kliebenstein *et al.*, 2001*a*). Glucosinolates and their degradation products have a wide range of biological activities, e.g. in plant defence as deterrents against insects and fungi. They share a core structure containing a β-D-glucopyranose residue linked via a sulfur atom to a (*Z*)-*N*-hydroxy-imino sulfate ester, and are distinguished from each other by a variable R group derived from one of several amino acids. Their biosynthetic pathways have been partially identified (Mikkelsen *et al.*, 2002; Wittstock and Halkier, 2002). The biosynthesis of glucosinolates can be divided into three independent parts. First, some amino acids, such as methionine and phenylalanine, are elongated by one or several methylene groups. Then the precursor amino acids are converted into the parent glucosinolates, and finally, the parent glucosinolates can undergo secondary modifications (Wittstock and Halkier, 2002).

In *Arabidopsis*, methionine and tryptophan are major precursors of glucosinolates. The side chain of methionine is elongated by a chain elongation cycle including the methylthioalkylmalate synthases (MAM) MAM-1 and MAM-L. After elongation, isoforms of cytochrome P450 proteins called CYP79s and CYP83s catalyse the aldoxime-forming and aldoxime-oxidizing reactions with different substrates. Then, the *S*-alkylthiohydroximates are converted into thiohydroximic acids by a C-S lyase, identified as the *SUPERROOT1* gene product SUR1 (Mikkelsen *et al.*, 2004). A UDP-glucose-thiohydroximate glucosyltransferase is responsible for glucosylation (Douglas Grubb *et al.*, 2004). The last step in the biosynthesis of the glucosinolate core structure of the different aliphatic, aromatic, and indole desulfoglucosinolates was suggested to be catalysed by members of the sulfotransferase (SOT) (EC 2.8.2.-) protein family. Glendening & Poulton (1990) partially purified a protein from *Lepidium sativum* L. that had 3'-phosphoadenosine 5'-phosphosulfate (PAPS)-dependent desulfoglucosinolate SOT activity. However, at that time no molecular data were available.Three genes have recently been identified that encode desulfoglucosinolate SOT proteins in *Arabidopsis* (AtSOT) (Piotrowski *et al.*, 2004). The glucosylation and sulfation reactions were assumed to be non-specific with respect to the side chain (Halkier, 1999).

All members of the SOT protein family transfer the sulfate group from PAPS to an appropriate hydroxy group of several classes of substrates. These enzymes have highly

conserved domains and can be found in eubacteria and eukaryotes (Varin *et al.*, 1997*b*; Marsolais *et al.*, 2000). Several SOT proteins have been characterized in *Flaveria* sp. and *Brassica napus* L. They showed substrate specificity for several flavonols, steroids, and brassinosteroids. The genes are regulated by physiological processes, such as growth, development, and adaptation to stress (Varin *et al.*, 1997*b*; Rouleau *et al.*, 1999; Marsolais *et al.*, 2004). In *Arabidopsis*, 18 sequences with high similarity to SOT proteins have been identified (AtSOT1 to AtSOT18) (Klein & Papenbrock, 2004). The *in vivo* substrates for most of the gene products in *Arabidopsis* are not known to date. For one AtSOT the substrate was identified as 12-hydroxyjasmonate, but the *in vivo* function has not been unequivocally clarified (Gidda *et al.*, 2003).

New approaches in plant molecular biology have helped to identify candidate genes for the last step in glucosinolate biosynthesis. Clustering of the metabolomes and the transcriptomes of sulfur-starved *Arabidopsis* plants indicated that the accumulation of glucosinolates correlates with expression of the genes responsible for their biosynthesis (Hirai *et al.*, 2005). Also in a screening for genes induced by the phytotoxin coronatine, a putative SOT protein was identified in *Arabidopsis*. The coronatine-induced protein preferred tryptophan-derived and phenylalanine-derived desulfoglucosinolates, whereas long-chain desulfoglucosinolates derived from methionine are the preferred substrates for two close homologues (Piotrowski *et al.*, 2004). However, these conclusions are based on measurements of substrate disappearance and two-substrate competition assays (Piotrowski *et al.*, 2004), and thus can only approximate real kinetic comparisons.

In this paper, we show that the three AtSOT proteins, AtSOT16, AtSOT17, and AtSOT18, are cytoplasmic with different substrate specificities and gene expression patterns. This detailed knowledge may be useful in the design of metabolically engineered plants with improved pest resistance and increased nutritional value.

Experimental procedures

Growth and harvest of plants

Seeds of *A. thaliana*, ecotype C24 and Col0, were originally obtained from the Arabidopsis stock centre at Ohio State University. Seeds were germinated on substrate TKS1 (Floragard, Oldenburg, Germany). Seedlings were transplanted into pots (7 cm diameter) containing TKS2 substrate (Floragard) after 14 days. Plants were grown in the greenhouse in a 16 h light/8 h darkness rhythm at a temperature of 23°C/21°C. When necessary, additional light was provided for 16 h a day to obtain a constant quantum fluence rate of 300 $\mu mol \cdot m^{-2} \cdot s^{-1}$ (sodium vapour lamps, SON-T Agro 400; Philips, Hamburg, Germany).

To obtain plant material from different tissues, *Arabidopsis* plants were grown for 7 weeks in the greenhouse, counted from the transfer into pots. Plants were harvested and split into roots, mature leaves of the primary rosette, young leaves of the rosette, stems (lower halves of primary and secondary bolts), siliques, and flowers.

To investigate natural senescence, *Arabidopsis* plants were grown in the greenhouse for up to 6 weeks, and the parts above ground were cut every week. The oldest leaves were comparable to the S3 stage as defined by Lohmann *et al*. (1994).

Four-week-old plants were grown in a 12 h light/12 h darkness cycle in a growth chamber at a quantum fluence rate of $50 \cdot \mu mol \cdot m^{-2} \cdot s^{-1}$ (TLD 58W/33; Philips) and a constant temperature of 22°C. To follow one complete diurnal cycle, plant parts above ground were harvested every 4 h for nearly 1.5 days starting 1 h after the onset of light.

To investigate the influence of high and low sulfate concentrations in the growing medium, *Arabidopsis* seeds, ecotype Col0, were germinated under sterile conditions (Schlesinger *et al.*,2003) in MS medium prepared as described by Murashige & Skoog (1962). The medium contained modified sulfate concentrations of 500 µM (high) and 50 µM (low), respectively.

DNA cloning techniques

The three genes under investigation, At1g74100, At1g18590, and At1g74090, encoding SOT16 from *Arabidopsis* (AtSOT16) (previously called AtST5a; explanations for the new nomenclature are given in Klein & Papenbrock (2004), AtSOT17/AtST5c, and AtSOT18/AtST5b, respectively, do not contain any introns. Therefore, the coding sequences could be amplified from genomic DNA isolated from *Arabidopsis* (ecotype C24) leaves with the Qiagen DNA extraction kit (Qiagen, Hilden, Germany). The following primer pair was used to amplify a 1017-bp coding sequence for AtSOT16: primer 258 (5'-GGT ACC GAA TCA AAG ACA ACC-3') extended by a *Kpn*I restriction site, and primer 259 (5'-CTG CAG GTT ATC ATG TTG AAG C-3') extended by a *Pst*I restriction site. For the amplification of the 1041-bp coding sequence for AtSOT17, primer 262 (5'-GGA TCC GAA TCC AAA ACC ATA-3') extended by a *Bam*HI restriction site, and primer 263 (5'-GTC GAC TGA TTT TGT AGA AAG-3') extended by a *Sal*I restriction site were used. A 1053-bp cDNA encoding AtSOT18 was amplified from genomic DNA of *Arabidopsis* (ecotype C24) and also from the λEMBL3 genomic library of *Arabidopsis* (ecotype Col0) (Clontech, Heidelberg, Germany) as described (Hirai *et al.*, 2005).

Two amino acids in AtSOT18 were replaced as described (Burow *et al.*, 2002). For the replacement of G_{301} with D_{301} following primers were used: P1/P260, P2/261, P3/P369 (5'-AAA GAC AGA GAG G**A**T CGT CCT GGT GTT-3') and P4/P370

(5'-AAC ACC AGG ACG A**T**C CTC TCT GTC TTT-3'). For the replacement of N_{339} with $K_{339,}$ the following primers were used: P1/P260, P2'/377 (5'-CTT TAC GGT CTT TAA AAA GGC CGT AAT-3'), P3/P371 (5'-TTA ATG GAA GAG AA**A** TTT AAG GGC ACC-3') and P4/P372 (5'-GGT GCC CTT AAA **T**TT CTC TTC CAT TAA-3').

To obtain fusions with GFP either with the 5' end of the GFP coding sequence (pGFP-N/AtSOT) or with the 3' end (pGFP-C/AtSOT) (Bauer *et al.*, 2004) the following primer pairs were used: for the AtSOT16 (123 bp) fusion with the 5' end of the GFP primer 295 (5'-CCA TGG AAT CAA AGA CAA-3') extended by a *Nco*I restriction site, and primer 296 (5'-AGA TCT ATC TGG TCT CCA GCC-3') extended by a *Bgl*II restriction site, and for fusion of the AtSOT16 cDNA (873 bp) with the 3' end of the GFP primer 297 (5'-CCA TGG GTG GAC ACT GGT GGC AA-3') extended by a *Nco*I restriction site, and primer 298 (5'-AGA TCT GTT ATC ATG TTG AAG-3') extended by a *Bgl*II restriction site. The following primer pairs were used for AtSOT17 (1041-bp) fusions with the GFP sequence: primer 299 (5'-CAA TGG AAT CCA AAA CCA TAA ACG-3') extended by a *Nco*I restriction site, and primer 300 (5'-AGA TCT TGA TTT TGA AGA AAG-3') extended by a *Bgl*II restriction site.

The PCR contained 0.2 mM dNTPs (Roth, Karlsruhe, Germany), 0.4 µM of each primer (MWG Biotech, Ebersberg, Germany), 1 mM $MgCl_2$ (final concentration, respectively), 0.75 µl Red*Taq* DNA-Polymerase (Sigma, Taufkirchen, Germany), and about 1 µg template DNA in a final volume of 50 µl. Before the start of the first PCR cycle, the DNA was denatured for 180 s at 94°C followed by 28 PCR cycles conducted for 60 s at 94°C, 60 s at 48°C (AtSOT16, AtSOT18 and AtSOT18*) or 45°C (AtSOT17), and 60 s at 72°C. The amplified PCR fragments were ligated into the expression vector pQE-30 (Qiagen) or into the pGFP vectors, and the vectors were introduced into the *Escherichia coli* strain XL1-blue. All clones were sequenced at least twice from both sides to obtain overlapping sequences (MWG Biotech). The four different AtSOT proteins [AtSOT16, AtSOT17, AtSOT18 (ecotype C24), and AtSOT18* (ecotype Col0)] were expressed in *E. coli* and purified as described (Hirai *et al.*, 2005). The purity of the protein preparations was checked by SDS/PAGE (Laemmli, 1970) and subsequent Coomassie Brilliant Blue and silver staining. Protein contents were determined by the method of Bradford (1976) with BSA (Roth) as a protein standard.

Preparation of the substrates

The desulfo forms of the parent glucosinolates derived from methionine, tryptophan and phenylalanine were prepared as described by Graser *et al.* (2001). In short, the glucosinolates were extracted by 80% methanol from seeds of different plant species and purified on DEAE-Sephadex A25. The bound glucosinolates were treated with

sulfatase (Sigma) overnight, and the desulfoglucosinolates obtained (Graser *et al.*, 2001) were eluted with 60% methanol. Individual desulfoglucosinolates were isolated by preparative HPLC with fraction collection. The following glucosinolates were used in the experiments in their desulfo forms: 3-butenylglucosinolate from *B. napus*, 3-methylthiopropylglucosinolate from *A. thaliana*, 4-methylthiobutylglucosinolate from *Eruca sativa*, 6-methylthiohexylglucosinolate from *Alyssum maritimum*, indol-3-yl-methylglucosinolate from *Isatis tinctoria*, 2-phenylethylglucosinolate from *Nasturtium officinale*, 4-pentenylglucosinolate from *B. napus*, 5-methylthiopentylglucosinolate from *Arabidopsis thaliana*, 7-methylthioheptylglucosinolate from *A. thaliana*, 8-methylthiooctylglucosinolate from *A. thaliana* and benzylglucosinolate from *Tropaeolum majus*. In addition, sinigrin, also called allylglucosinolate (Sigma) was purchased and desulfated as described above. 4-Pentenylglucosinolate, 5-methylthio-pentylglucosinolate, 7-methylthioheptylglucosinolate and 8-methylthiooctylglucosino-late were only used as substrates in the assays including [^{35}S]PAPS as cosubstrate whereas all other glucosinolates were used in both assay types (see next paragraph). Structures of all *Arabidopsis* glucosinolates are shown in Mellon *et al.* (2002).

Enzyme activity measurements

The enzyme assays with recombinant proteins were set up in a modified way from that described by Glendening & Poulton (1990). During the optimization of the enzyme assay conditions, 5 - 20 µg of the purified recombinant proteins was used. Different pH values (6 - 10 using appropriate buffers for the respective pH range) and the effects of the addition of 9.2 mM $MgCl_2$ were tested. Different PAPS preparations (Calbiochem or obtained from Professor H. R. Glatt, Institute of Human Nutrition, Berholz-Rehbruecke, Germany) were used, and [^{35}S]PAPS (3.7 MBq; Perkin-Elmer, Boston, MA, USA) was mixed in a molar ratio of 1:100 with nonradioactive PAPS. Finally, the 300 µL-assays contained: 83 mM Tris/HCl pH 9.0, 9.2 mM $MgCl_2$, 58 µM PAPS, the respective substrate (58 µM), and 15 µg purified protein. For the determination of the K_m values for desulfobenzylglucosinolate, at least 10 different substrate concentrations were used depending on their inhibitory effects; in these assays the PAPS concentrations were increased by various amounts to avoid limitation of the cosubstrate. The reaction was started by the addition of PAPS, incubated for 60 min at 37°C, and stopped by incubation for 10 min at 95°C. The formation of the respective sulfated product was analysed by HPLC (Knauer, Berlin, Germany) as described by Mellon *et al.* (2002). Separation was achieved with a gradient of 0.1% trifluoroacetic acid and acetonitrile with 0.1% trifluoroacetic acid on an RPC18 column (Supelcosil LC-18-DB, 250 x 4.6 mm internal diameter, 5 µm particle size; Supelco, Taufkirchen, Germany) with a flow rate of 1 mL·min^{-1}. The eluent was monitored by diode array detection

(Knauer) between 190 and 360 nm (2 nm interval). Desulfoglucosinolates and the respective intact glucosinolates were identified by comparison of retention time and UV spectra with those of purified standards previously extracted from *Arabidopsis* and quantified at A_{229} relative to the internal standard (Brown *et al.*, 2003).

When [^{35}S]PAPS was included, the assay products were separated as described above except that the flow was reduced to 0.25 mL·min^{-1} and the column diameter was 2.1 mm. In addition to UV absorption at A_{229} (Agilent HP1100 Series; Agilent, Böblingen, Germany), products were detected by a flow-through radioactivity monitor (Packard Radiomatic 500TR; Hewlett-Packard, Böblingen, Germany) with a 0.5-mL flow cell and Ultima-Flo AP scintillation fluid (Packard) in a ratio of 4:1 to column eluent. All data reported are means of at least three independent measurements.

Transient expression of the GFP fusion constructs in *Arabidopsis* protoplasts

Preparation and transformation of protoplasts from *Arabidopsis* plants were done as described previously (Bauer *et al.*, 2004). The transiently transformed protoplasts were analysed by fluorescence microscopy (Axioskop Zeiss, Jena, Germany) with the following filter set: excitation 450-490 nm (filter BP450-490) and emission 520 nm (filter LP520) for GFP. All images were prepared with Corel Photo Paint 10.

RNA extraction and Northern Blot analysis

Total RNA was extracted essentially as described by Chomczynski & Sacchi (1987). RNA samples (20 µg) were separated on 1% denaturing agarose/formaldehyde gels. Equal loading was controlled by staining the gels with ethidium bromide. After RNA transfer to nylon membranes, the filters were probed with digoxigenin-labelled cDNA probes obtained using the PCR DIG probe synthesis kit (Roche, Mannheim, Germany). Colorimetric or chemiluminescent detection methods with nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate or with CDP-*Star* (Roche) as substrates for alkaline phosphatase were applied.

Bioinformatical analyses and statistics

DNA and amino-acid sequences were analysed with the programs MapDraw and Protean in Lasergene (DNASTAR, Madison, WI, USA). For the prediction of protein localization, different programs were applied (iPSORT, PSORT, SignalP, and TargetP). Several programs were used for the determination of phosphorylation sites and for the prediction of the secondary structure (http://www.expasy.ch/tools). The K_m values were calculated from the nonlinear Michaelis-Menten plot using an enzyme kinetics program (SigmaPlot 7.0).

Results

Analysis of the AtSOT DNA and protein sequences

The glucosinolate pattern differs among the various *Arabidopsis* ecotypes (Kliebenstein *et al.*, 2001*a*). As the ecotype C24 has the broadest variety of glucosinolates representing all chemical groups except aromatic (Michael Reichelt, Max Planck Institute for Chemical Ecology, Jena, Germany, unpublished data), we decided to study the three putative SOT proteins from ecotype C24 and for comparison one SOT protein from ecotype Col0. The DNA sequence encoding the AtSOT18 protein isolated from ecotype Col0 (AtSOT18*) showed 100% identity with the available sequences in the databases (accession number NP_177549). Each sequence encoding the AtSOT16 - 18 proteins from ecotype C24 revealed one transition in comparison with published sequences from ecotype Col0 (The Arabidopsis genome initiative, 2000), and AtSOT17 and 18 also had one transversion.

These replacements led to either one or two changes at the amino-acid level compared with the primary amino-acid sequences in ecotype Col0. In AtSOT16, V_{324} was replaced by L, in AtSOT17 W_{183} was replaced by R and S_{344} by T, and in AtSOT18 D_{301} was replaced by G and K_{339} by N. The replacement of V with L (AtSOT16) and of S with T (AtSOT17) should not cause any significant changes in the tertiary structure of the proteins whereas the replacement of the nonpolar W with the polar R and the charged D with the small G and the polar L with the charged N may affect the tertiary structure or the substrate-binding cavities. The putative PAPS-binding site was not changed. Four conserved regions, I - IV, which were speculated to have functional roles in soluble SOT proteins, have previously been defined (Varin *et al.*, 1992). The amino acids V_{324} in AtSOT16 and S_{344} in AtSOT17 and K_{339} in AtSOT18 are located after region IV at the C-terminus of the protein. The replacement of W_{183} by R in AtSOT17 is located in region II and that of D_{301} by G in AtSOT18 between region III and IV, parts of the protein thought to be responsible for substrate binding.

The predicted molecular masses of the three AtSOT proteins from ecotype Col0 were 39.2 kDa for AtSOT16 consisting of 338 amino acids, 39.9 kDa for AtSOT17 (346 amino acids), and 40.5 kDa for AtSOT18 (350 amino acids). At the amino-acid level, AtSOT16 showed 74% identity with AtSOT17 and 76% identity with AtSOT18; AtSOT17 shared 77% identity with AtSOT18.

Enzyme activity measurements using the recombinant AtSOT proteins

All four recombinant AtSOT proteins could be purified to ≈ 90% homogeneity, as estimated by the analysis of protein fractions on silver-stained SDS polyacrylamide gels

(data not shown). To analyse the substrate specificities *in vitro*, SOT assays were performed as described by Glendening & Poulton (1990). The 300 µL-assays mixtures contained 15 µg purified recombinant protein in 83 mM Tris/HCl, pH 9.0, 9.2 mM $MgCl_2$, and 58 µM PAPS (Calbiochem, Darmstadt, Germany) and 6.2 mM desulfoglucosinolate, and were incubated for 30 min at 30°C. The results of the assays using AtSOT18* have been published (Hirai *et al.*, 2005). Only in the presence of nondenatured AtSOT18* and both substrates, PAPS and desulfoallyl glucosinolate, could the formation of the product, the sulfated allyl glucosinolate, be detected by HPLC analysis (Hirai *et al.*, 2005). The product was analysed by MS, which confirmed its identity (data not shown). V_{max} was calculated as 467 pkatal· (mg protein)$^{-1}$, ≈ 30 times higher than previously obtained using the same substrate (Varin and Spertini, 2003). When the assay conditions described by Glendening & Poulton (1990) were followed exactly, only the recombinant AtSOT18*, out of the four expressed and purified AtSOT proteins, was shown to be active. Furthermore, it was only active with desulfoallylglucosinolate as substrate out of 12 desulfoglucosinolates tested. Therefore two strategies were applied: optimizing the assay conditions for all four recombinant AtSOT proteins and increasing the sensitivity of the detection by using [^{35}S]PAPS.

To determine the pH optimum, activities in the pH range 6 - 10 were tested; the highest activities were measured at pH 9.0. A higher incubation temperature of 37°C instead of 30°C led to the formation of larger product peaks. In the absence of $MgCl_2$, the enzyme activity itself was not impaired, but HPLC separation and peak shape were, for unknown reasons, negatively influenced; therefore $MgCl_2$ was always included in the assay buffer. In the first experiments inhibitory effects of higher PAPS (Calbiochem) concentrations were observed. It was found that the PAPS preparation also contained ≈ 40% 3'-phosphoadenosine 5'-phosphate (PAP), which is known to inhibit the SOT reaction (Klaassen and Boles, 1997). Therefore almost pure PAPS (99.5%) from another source (Institute of Human Nutrition) was used in subsequent experiments. The concentration of the desulfoglucosinolates was reduced from 1 mM to 58 µM, the same concentration as the cosubstrate, because substrate concentrations close to 1 mM were also found to have inhibitory effects. In parallel, we attempted to increase the sensitivity of the enzyme assay by including ^{35}S-labelled PAPS in addition to unlabelled PAPS, and to analyse the products with a radioactivity monitor connected in series with a UV detector. Twelve desulfoglucosinolates were used as substrates in the assay in which [^{35}S]PAPS was included. For all of them, product formation could be unequivocally detected. In the optimized nonradioactive assay, product formation could be shown for eight substrates (see next paragraph). Therefore the sensitivity of the assay using [^{35}S]PAPS was slightly higher than the nonradioactive assay.

The four different desulfoglucosinolate AtSOT proteins may exhibit different enzyme kinetics and/or different substrate specificities. Comparisons of kinetic properties with a single substrate were performed using desulfobenzylglucosinolate, because all four proteins showed catalytic activity with this substrate (Table 1). The kinetic data were determined with either PAPS or desulfobenzylglucosinolate as the variable substrate. For the determination of the K_m values, 10 data points using substrate concentrations in the range 25-250 µM were included. For AtSOT16 the range was extended to 700 µM. The results were obtained by fitting data to the Michaelis-Menten equation. No enzyme inhibition was observed over the ranges of PAPS and desulfobenzylglucosinolate concentration used. Typical hyperbolic saturation was observed when PAPS was the variable substrate (data not shown). The V_{max} for desulfobenzylglucosinolate of AtSOT16 and AtSOT18* are in the same range, whereas that of AtSOT17 is ≈ 6 times lower. The results also reveal that the binding affinity of the donor substrate PAPS and the acceptor substrate desulfobenzylglucosinolate to AtSOT16 and AtSOT18* are the same, whereas PAPS has a higher affinity for AtSOT17. It was not possible to determine a reliable K_m value for AtSOT18 because a sigmoid curve was obtained in the low concentration range. Higher substrate and cosubstrate concentrations were inhibitory.

Table 1. Determination of V_{max} [pkatal·(mg protein)$^{-1}$], K_m values, the catalytic constants k_{cat}, and the quotients k_{cat}/K_m for the recombinant AtSOT proteins using desulfobenzylglucosinolate (ds-benzyl-GL)as substrate. For the determination of the K_m values, 10 data points using substrate concentrations in the range 25 -250 µM were included. For AtSOT16 the range was extended to 700 µM. Each determination was carried out three times, and the means are presented. The molecular masses were calculated for the recombinant protein including the 6xHis tag. n.d., Not determinable (for explanation see Results).

Enzyme	Substrate	V_{max}	K_m (µM)	k_{cat} (s^{-1})	k_{cat}/K_m (μM^{-1} s^{-1})
AtSOT16 C24	ds-benzyl GL[a]	1100	100	44 x 10^{-3}	0.44 x 10^{-3}
	PAPS[b]	692	100	27.7 x 10^{-3}	0.28 x 10^{-3}
AtSOT17 C24	ds-benzyl GL[c]	125	100	5.1 x 10^{-3}	0.05 x 10^{-3}
	PAPS[d]	142	25	5.8 x 10^{-3}	0.23 x 10^{-3}
AtSOT18 C24	ds-benzyl GL[e]	171	n.d.	n.d.	n.d.
	PAPS[f]	27	35	1.1 x 10^{-3}	0.03 x 10^{-3}
AtSOT18* Col-0	ds-benzyl GL[g]	851	50	35.2 x 10^{-3}	0.70 x 10^{-3}
	PAPS[h]	860	60	35.5 x 10^{-3}	0.59 x 10^{-3}

[a]To avoid limitations of the cosubstrate different substrate amounts were used, 350 µM PAPS; [b]250 µM desulfobenzylglucosinolate; [c]250 µM PAPS; [d]250 µM desulfobenzylglucosinolate; [e]58 µM PAPS; [f]100 µM desulfobenzylglucosinolate; [g]350 µM PAPS; [h]250 µM desulfobenzylglucosinolate.

To determine the substrate specificities, enzyme assays were performed under identical conditions with the nonradioactive assay using the four recombinant AtSOT proteins and eight different purified desulfoglucosinolates. As an example, HPLC chromatograms with desulfo-6-mehtylthiohexylglucosinolate as substrate are shown in Fig. 1. The peaks on the right represent the substrate peak (58 µM), the peaks on the left the product peaks. The four recombinant AtSOT proteins show different activities with desulfo-6-mehtylthiohexylglucosinolate as substrate using the same assay conditions.

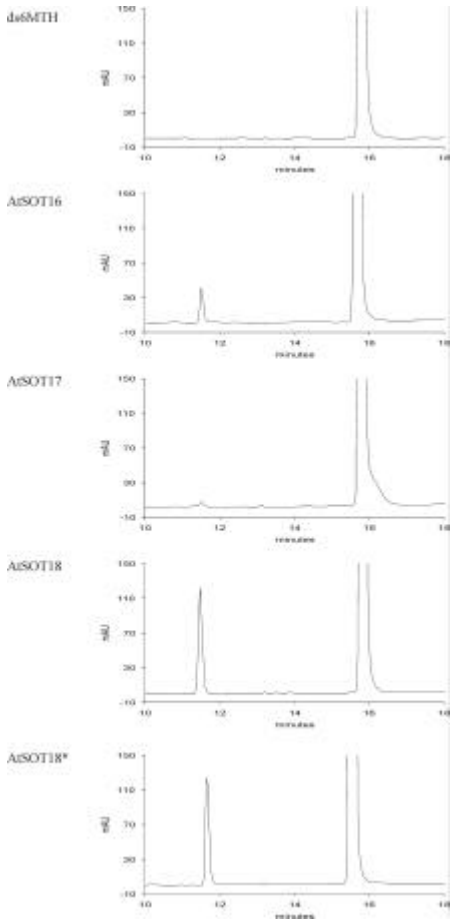


Fig. 1. Representative HPLC analysis results of enzyme assays using desulfo-6-methylthiohexylglucosinolate as substrate. Formation of the sulfated intact 6-methylthio-hexylglucosinolate was analysed by HPLC and is shown for the respective recombinant enzyme from the top to the bottom: Control incubation without enzyme, AtSOT16, AtSOT17, AtSOT18, AtSOT18*. 6-Methylthiohexylgluco-sinolate, left peak; desulfo-6-methylthiohexylglucosinolate, right peak.

The results of the enzyme assays with the full range of substrates are summarized in Table 2. AtSOT16 has a very broad substrate specificity and can sulfate all desulfo-glucosinolates offered. Both the tryptophan-derived and the phenylalanine-derived desulfoglucosinolates are accepted equally. AtSOT17 principally shows activity with the phenylalanine-derived substrate desulfobenzylglucosinolate, and to a much lower extent with the methionine-derived desulfo-6-methylthiohexylglucosinolate. AtSOT18 shows low activity with desulfo-6-methylthiohexylglucosinolate and with both aromatic desulfoglucosinolates tested. The AtSOT18* protein shows the highest activity with desulfobenzylglucosinolate and sulfates all methionine-derived desulfoglucosinolates and also the tryptophan-derived desulfoindol-3-ylmethylglucosinolate. In summary,

AtSOT16 and AtSOT18* possess broad substrate specificities, whereas the AtSOT17 and AtSOT18 are active only with a small number of desulfoglucosinolates.

Table 2. Determination of substrate specificities of desulfoglucosinolate AtSOT proteins. The enzyme assays were performed with the respective recombinant proteins using the respective desulfoglucosinolates in equal concentrations (58 µM). The peak areas of the products were determined and the relative velocities in comparison with the best substrate (set as 100) were calculated. MTP, Methylthiopropylglucosinolate; MTB, methylthiobutylglucosinolate; MTH, methylthiohexylglucosinolate; I3M, indol-3-yl methylglucosinolate; PE, phenylethylglucosinolate.

Enzyme	Best substrate	Other substrates						
AtSOT16	2PE	I3M	Benzyl	6MTH	4MTB	3MTP	Sinigrin	3Butenyl
	100	93	45	34	13	12	3	3
AtSOT17	Benzyl	6MTH	2PE	I3M	4MTB	3MTP	Sinigrin	3Butenyl
	100	84	0	0	0	0	0	0
AtSOT18	6MTH	Benzyl	2PE	I3M	4MTB	3MTP	Sinigrin	3Butenyl
	100	30	24	0	0	0	0	0
AtSOT18*	Benzyl	4MTB	6MTH	3MTP	I3M	Sinigrin	3Butenyl	2PE
	100	98	89	77	65	22	14	0

Enzyme activity of the mutagenized AtSOT18 protein

To investigate the differences between AtSOT18 from ecotype C24 and AtSOT18* from ecotype Col0, the two amino acids in AtSOT18 were replaced separately by the respective amino acids in AtSOT18* resulting in AtSOT18G_{301}D and AtSOT18N_{339}K. The AtSOT18G_{301}D protein showed a 25 times higher specific activity with desulfobenzylglucosinolate than AtSOT18 comparable to the specific activity of AtSOT18* (Fig. 2). The protein sulfated the same desulfoglucosinolates with velocities comparable to AtSOT18* (Table 3). To date, we have been unable to purify the AtSOT18N_{339}K protein under native conditions. The results indicate a critical role for the glycine residue at position 301 in substrate recognition.

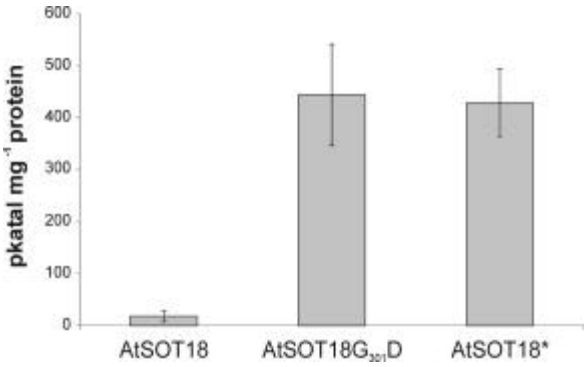


Fig. 2. Determination of the specific activities of different AtSOT18 proteins. Enzyme assays were performed at a PAPS concentration of 58 µM using the three purified recombinant proteins AtSOT18 (ecotype C24), AtSOT18G_{301}D, and AtSOT18* (ecotype Col0). The specific activity is given in pkatal· (mg protein)$^{-1}$. The standard deviation was calculated (*n = 3*).

Table 3. Determination of substrate specificities of the desulfoglucosinolate AtSOT18G_{301}D. The enzyme assays were performed with the recombinant protein using the respective desulfoglucosinolates in equal concentrations (58 µM). The peak areas of the products were determined and the relative velocities in comparison with the best substrate (set as 100) were calculated. MTP, Methylthiopropylglucosinolate; MTB, methylthiobutylglucosinolate; MTH, methylthiohexylglucosinolate; I3M, indol-3-yl-methylglucosinolate; PE, phenylethylglucosinolate.

Enzyme	Best substrate	Other substrates						
AtSOT18G_{301}D	4MTB	Benzyl	6MTH	I3M	3MTP	Sinigrin	3Butenyl	2PE
	100	94	94	80	73	45	18	0

Arabidopsis desulfoglucosinolate SOTs are localized in the cytoplasm

Several publicly available programs have been used to predict the subcellular localization of AtSOT16, 17 and 18 (Klein & Papenbrock, 2004). For all three proteins, localizations in peroxisomes were predicted with the highest probabilities, whereas the second highest scores were given for a localization in the cytoplasm (AtSOT16 and 18) and the chloroplast (AtSOT17). To confirm the *in silico* prediction results of all three AtSOT proteins, in-frame fusions of the sequences encoding the full-length proteins with the green fluorescent protein (GFP) encoding sequence (C) were transformed in *Arabidopsis* protoplasts (Fig. 3A-D). AtSOT18 has already been shown to remain in the cytoplasm (Klein & Papenbrock, 2004). In the present study, AtSOT16 and AtSOT17 also remained in the cytoplasm. Transformations with fusion constructs encoding endoplasmic reticulum (ER)-localized proteins resulted in completely different images (data not shown). Therefore, it can be concluded that none of these proteins contain a targeting peptide at either the N-terminus or C-terminus.

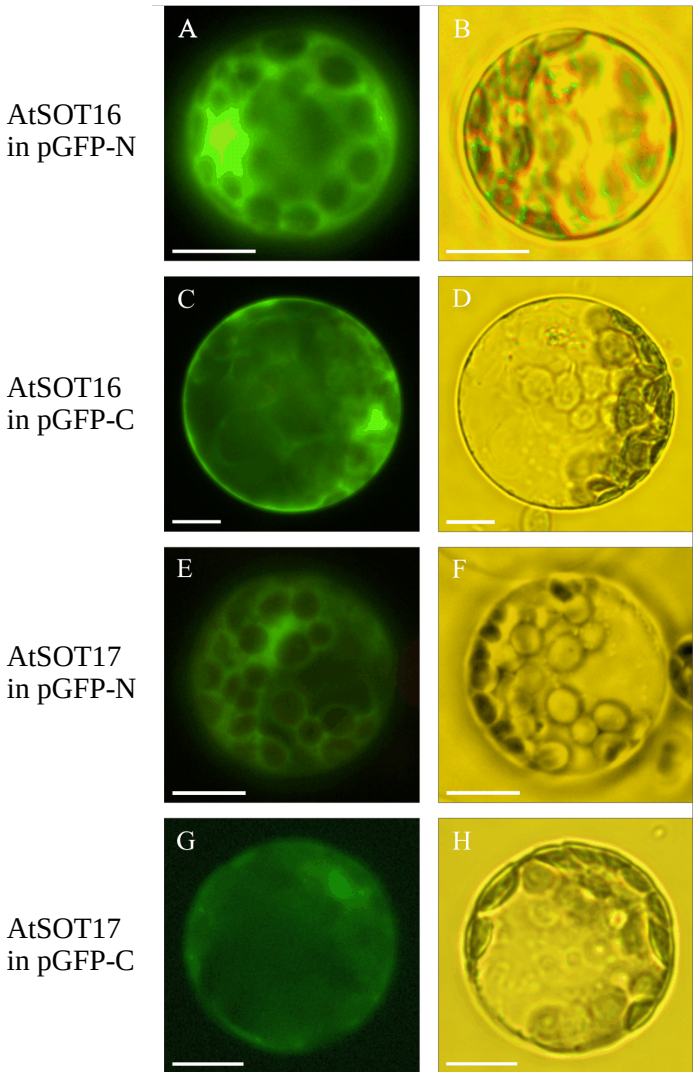


Fig. 3. Intracellular localization of AtSOT-GFP fusion constructs. The AtSOT16 and AtSOT17 encoding cDNAs were ligated in-frame into the pGFP-N and the pGFP-C vectors, respectively. The fusion constructs were introduced into *Arabidopsis* protoplasts. The protoplasts were incubated overnight at room temperature and then analysed with an Axioskop microscope with filter sets optimal for GFP fluorescence (BP450-490/LP520 filters). Fluorescence images of the transformed protoplasts are shown in (A) and (C) (AtSOT16/pGFP-N and AtSOT16/pGFP-C, respectively) and (E) and (G) (AtSOT17/pGFP-N and AtSOT17/pGFP-N, respectively). Bright field images of the same protoplasts were produced to visualize the protoplast's cell membrane and the chloroplasts (B and D, and F and H).

Expression studies at the RNA level

In addition to different substrate specificities, the three *desulfoglucosinolate AtSOT* genes investigated may have different roles in glucosinolate metabolism because of their different expression patterns. mRNA transcript levels were assessed by Northern blotting (Fig. 4A). As the DNA sequences showed a high degree of similarity (Klein & Papenbrock, 2004), the specificity of the probes was tested by analysing cross-hybridization reactions in Southern blots using the respective plasmid DNA. Cross-reactivity was almost undetectable among the three different *desulfoglucosinolate AtSOT* genes (data not shown). The highest mRNA amounts for AtSOT16 and 18 were

found in roots; mRNA coding for AtSOT17 was also easily detectable in roots but the expression was much higher in mature leaves of the primary rosette. In young leaves the mRNA levels for all three genes were about half those of the mature rosette leaves and stems. In flowers the expression levels were lower than in the other tissues analysed and even more reduced in siliques. In summary, the expression patterns of all three *desulfo-glucosinolate AtSOT* genes show small, but distinctive differences.

A comparison of developmental stages of up to 6-week-old *Arabidopsis* plants showed the mRNA levels of *AtSOT16* and *AtSOT17* to be maximum in the 2-week-old plants, whereas in older plants the expression levels decreased, reaching a minimum in flowering plants (weeks 5 and 6; Fig. 4B). In contrast, the mRNA levels of *AtSOT18* were low in the younger plants (up to 4 weeks old) and were slightly elevated in the oldest plants harvested in weeks 5 and 6.

The influence of light on the expression was investigated by following a 12 h light/12 h darkness cycle of 4-week-old *Arabidopsis* plants (Fig. 4C). A cycling of the *AtSOT17* mRNA levels was observed: the highest mRNA levels were determined at the end of the light phase, and the lowest at the end of dark phase.

The effects of a 10-fold different sulfate concentration in the medium were investigated by germinating seeds in MS medium with 500 µM (high) and 50 µM (low) sulfate concentrations and growing them for 18 days. The *Arabidopsis* plants grown at high and low sulfate were phenotypically identical. The levels of mRNA encoding AtSOT16, AtSOT17, and AtSOT18 were almost the same at both sulfate concentrations (Fig. 4D).

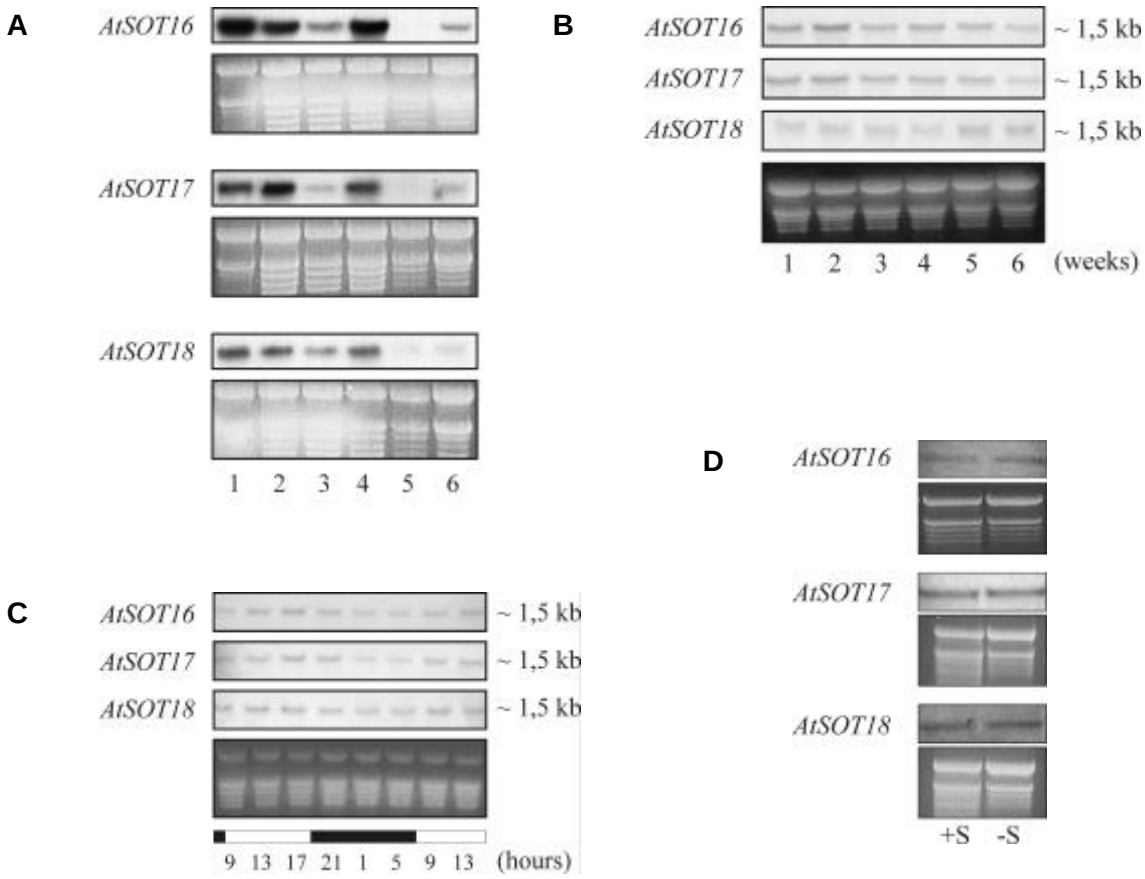


Fig. 4. Variation in mRNA transcript concentrations with organ type, developmental stage, diurnal variation and sulfate concentration. *Arabidopsis* plants were grown under different conditions, and pieces of plant material were cut out and immediately frozen in liquid nitrogen. Total RNA was extracted, 15 µg RNA per lane was separated on 1% denaturing agarose gels, and blotted on to nylon membranes. To confirm equal loading of the extracted RNA, the ethidium bromide-stained gel is shown at the bottom of each figure. *AtSOT16*, *AtSOT17*, and *AtSOT18* cDNAs were labelled with digoxigenin by PCR.

(A) *Arabidopsis* plants were grown for 7 weeks. The tissue-specific expression was analysed in roots (lane 1), mature leaves of the primary rosette (lane 2), young leaves of the rosette (lane 3), stems (lane 4), siliques (lane 5), and flowers (lane 6).

(B) Plants were grown in the greenhouse for 10 - 45 days, counting from the transfer to pots, and all plant tissue above ground was harvested every 7 days and used for the analyses.

(C) 2-week-old *Arabidopsis* plants (counting from the transfer to TKS2 medium) were grown in a 12 h light/12 h darkness cycle and the parts above ground were harvested every 4 h.

(D) *Arabidopsis* seeds were germinated in MS medium with high (500 µM) and low (50 µM) sulfate concentration in the medium. The seedlings were grown for 18 days in the same medium. The shoots were removed and directly frozen in liquid nitrogen. Representative Northern blots from three independent experiments are shown.

Discussion

Different substrate affinities and specificities of the four recombinant AtSOT proteins

The last step of core glucosinolate biosynthesis in *Arabidopsis* is catalysed by a family of small PAPS-desulfoglucosinolate SOT proteins. Three genes are found in the *Arabidopsis* genome, *ATSOT16*, *AtSOT17* and *AtSOT18*. In this study, we compared the properties of the three desulfoglucosinolate AtSOT proteins from *Arabidopsis* ecotype C24 and one from ecotype Col0 to determine how they may influence glucosinolate composition and distribution. It has previously been demonstrated that the AtSOT16-18 proteins are active with desulfoglucosinolates, but not with a wide range of other substrates including steroids, flavonoids and phenolic acids (Varin and Spertini, 2003; Piotrowski *et al.*, 2004). For the comparative kinetic analyses, a desulfoglucosinolate substrate, desulfobenzylglucosinolate, was chosen that was accepted by all four enzymes investigated, although this aromatic desulfoglucosinolate has not yet been identified in *Arabidopsis* ecotype C24. K_m values ranged from 50 to 100 µM for desulfobenzylglucosinolate and 25 to 100 µM for PAPS. *In planta*, the concentrations of desulfoglucosinolates (Brown *et al.*, 2003) are typically in the same range as the K_m values determined for desulfobenzylglucosinolate (Brown *et al.*, 2003). However, the concentration of PAPS is not known in plants so it is difficult to evaluate the significance of the K_m values for this compound. Surprisingly, the AtSOT18 proteins isolated from the two different *Arabidopsis* ecotypes differ in their kinetic behaviour, although it was not possible to determine their K_m for desulfobenzylglucosinolate because over the range of substrate concentrations used a sigmoid curve was obtained. Previously published kinetic analyses gave much lower K_m values for both substrates, between 0.14 and 4.8 µM for the desulfoglucosinolates (using desulfobenzyl-glucosinolate as substrate) and 0.17-1.17 µM for the cosubstrate PAPS (Varin and Spertini, 2003). However, the concentrations of both substrates in the enzyme assay were only about 1/50th of that used in the present study, and the PAPS used was ^{35}S-labelled in contrast with the unlabeled form used in the present work. These and other differences in assay conditions may explain the lower K_m values. Unfortunately, Varin & Spertini (2003) provide no information on the number of replicates and their variance. It is also hard to compare our results with those of Piotrowski et al. (Piotrowski *et al.*, 2004) because that study measured activity by substrate disappearance and estimated substrate specificity from two-substrate competition assays rather than using real kinetic comparisons. In future work the type of reaction mechanism needs to be investigated. Initial rate kinetics analysis may be done to determine whether a sequential or a ping-pong Bi-Bi mechanism is operative in

desulfoglucosinolate SOT proteins. To analyse regulatory aspects, the *in vitro* and *in vivo* effects of PAP as inhibitor could be determined. For a bacterial carbohydrate SOT, the K_i for PAP was in the same range as the $K_{m,}$ indicating comparable binding affinities (Pi *et al.*, 2005).

One could postulate that each SOT protein might specifically sulfate glucosinolates of one of the three glucosinolate groups: aliphatic, aromatic and indole glucosinolates. In previous experiments, desulfoindol-3-ylmethylglucosinolate and desulfobenzyl-glucosinolate were the best substrates for AtSOT16, and it was concluded that this protein preferentially sulfates tryptophan- derived and phenylalanine-derived desulfo-glucosinolates, whereas AtSOT17 and AtSOT18 preferentially sulfate methionine-derived desulfoglucosinolates (best substrates desulfo-8-methylthiooctylglucosinolate and desulfo-7-methylthioheptylglucosinolate) (Piotrowski *et al.*, 2004). Our results confirm a preference of AtSOT16 for aromatic desulfoglucosinolates and also reveal high activity of this protein with longer-chain aliphatic desulfoglucosinolates. For AtSOT17, our results show that this enzyme actually prefers the aromatic desulfobenzylglucosinolate as a substrate more than the aliphatic 6-methylthioheptyl, and that AtSOT18* also shows high activity with phenylalanine- derived and tryptophan-derived substrates. The specific activities of these enzymes are divergent. In the experiments of Varin & Spertini (2003) AtSOT18 showed the highest activities followed by AtSOT17 and AtSOT16. In our experiments the highest specific activities were obtained with AtSOT16 followed by AtSOT18*, 18 and 17. Obviously, the results depend on the assay conditions chosen and the *Arabidopsis* ecotype used.

The results also indicate that some of the side-chain modification reactions could take place before the glucosinolate core structure is completed by sulfation, in contrast with the generally accepted sequence of glucosinolate formation. Previous work has shown that the formation of the complete glucosinolate moiety including the sulfate function precedes side-chain transformations of methionine-derived glucosinolates that convert methylthioalkyl side chains into methylsulfinylalkyl, hydroxylalkyl and alkenyl side chains (Graser *et al.*, 2000; Graser *et al.*, 2001). However, the fact that two alkenyl glucosinolates (desulfosinigrin and desulfo-3-butenylglucosinolate) are sulfated by AtSOT18* at measurable rates compared with other glucosinolates, suggests that sulfation may follow some side-chain modifications.

The fact that sulfation is a terminal step in glucosinolate formation may be of physiological significance to plants as desulfoglucosinolates are not susceptible to myrosinase activity and thus not biologically active, but also not likely to be poisonous to the plant through formation of toxic hydrolysis products, such as isothiocyanates. Hence, it may be advantageous for plants to use desulfoglucosinolates as a transport or

even a storage form under conditions in which the risk of autotoxicity is high (Graser *et al.*, 2001).

Point mutations cause major effects in substrate specificities of AtSOT proteins

The comparison of the AtSOT18 proteins from ecotype C24 and ecotype Col0 revealed an interesting phenomenon: although they vary only in two amino acids and one of the differences is not even in one of the functionally defined regions of the protein, the proteins differ significantly in their kinetic parameters. In addition, the AtSOT18 from ecotype Col0 (AtSOT18*) showed a much broader substrate specificity than that from ecotype C24.

The mutagenesis experiments showed that altering a single amino acid of AtSOT18 to that present in AtSOT18* dramatically changed the substrate specificity to that of AtSOT18*. One could speculate about several reasons for this major difference. Aspartate (D) is a negatively charged amino acid at physiological pH, which may influence local charge distribution, altering substrate binding and recognition. The additional acidic amino acid might also simulate a phosphorylation, although none of the phosphorylation sites in the protein are present in this region. The amino-acid replacement does not impair ordered secondary structures in this region because several programs for the prediction of the secondary structure indicate a coil formation for both wild-type and mutant proteins. Obviously, additional amino acids play an important role in substrate recognition besides the previously defined conserved regions (Varin *et al.*, 1992). It would be very challenging to isolate more AtSOT18 homologues from other ecotypes to compare their amino acid sequences and substrate specificities.

In parallel with biochemical analyses aimed at identifying the specific substrate of an enzyme reaction, determination of the 3D structure and subsequent substrate modelling by docking programs may also be a successful approach to learning more about substrate specificity. To date the 3D structure for one AtSOT12 protein (At2g03760) has been determined which has ≈ 40% identity with the desulfo-glucosinolate AtSOT proteins. The overall 3D structure was nearly identical with already solved structures of steroid SOT proteins from humans. The structures contain a long disordered loop not found in the electron-density map, presumably because the substrates were not present. Whether substrate binding will in part order this region requires prior identification of the appropriate substrates and subsequent crystallo-graphic investigations including the substrates (Smith *et al.*, 2004). Thus, to date structural investigations have not been helpful in determining the putative substrates of AtSOTs.

To analyse the influence of the protein sequence on *in planta* substrate specificities, it might be interesting to replace the endogenous SOT form with mutated ones. However, based on our results, care must be taken in choosing an ecotype background for such experiments as the substrate specificities of desulfoglucosinolate SOTs vary among ecotypes.

Glucosinolate profiles have previously been determined in 39 *Arabidopsis* ecotypes, and altogether 34 different glucosinolates have been identified. A modular genetic control system based on polymorphisms at only five loci seems sufficient to produce the 14 different *Arabidopsis* leaf glucosinolate profiles (Kliebenstein *et al.*, 2001*a*). However, the complete enzymatic basis for the diversity is not yet known, and changes in the substrate specificity of desulfoglucosinolate AtSOT proteins might well influence the relative abundance of different glucosinolates among ecotypes.

Intermediates of glucosinolate biosynthesis have to be transported across membranes

As shown in many previous studies, computer-based predictions for the subcellular localization of nuclearencoded proteins are not reliable (Bauer *et al.*, 2004). However, comparative studies have shown that fusion with GFP results in reliable subcellular localization data (Bauer *et al.*, 2004). All three desulfoglucosinolate AtSOT proteins are now known to remain in the cytoplasm (Klein & Papenbrock, 2004; this work). However, the *in vitro* pH optimum for the SOT reaction is 9.0, which is more alkaline than the average pH in the cytoplasm, which is ≈ 7.5. At pH 7.0, the reaction is expected to proceed at ≈ 50% of the rate at pH 9.0 with variations for each AtSOT protein. Locally the pH in the cytoplasm may be higher.

The other proteins involved in the biosynthesis of glucosinolates are potentially localized in a variety of cell compartments according to results obtained from PSORT and TargetP (data not shown). The proteins that elongate the side chains of methionine are probably localized in the plastids because the MAM proteins contain putative transit peptides for chloroplast import. For the CYP79 proteins which catalyse the formation of aldoximes with different substrates, a localization in the ER, or with a lower probability in mitochondria (e.g. CYP79B2, At4g39950), can be assumed, whereas CYP83 proteins (CYP83A1, At4g13770 and CYP83B1, At4g31500) which catalyse aldoxime-oxidizing reactions may be localized in the ER or, with a lower probability, in the cytoplasm. The subsequent C-S-lyase protein(s) which convert *S*-alkylthiohydroximates into thiohydroximic acids are either localized in the cytoplasm (SUR1, At2g20610) (Mikkelsen *et al.*, 2004) or, a second putative C-S lyase (At5g36160), in the ER (Hirai *et al.*, 2005). The step preceding the sulfotransferase reaction catalysed by UDP-glucose-thiohydroximate glucosyltransferase (At1g24100) is predicted to be localized in

the ER (PSORT, own results). In summary, one has to postulate a reaction sequence via the following compartments: plastids – ER – cytoplasm – ER – cytoplasm. Finally, the mature glucosinolates are stored in the vacuole. Therefore several transport processes have to be assumed for the metabolic flow through the complete glucosinolate biosynthetic pathway. To our knowledge no carrier proteins or transporters have been reported that transport glucosinolate intermediates.

The three desulfoglucosinolate SOT proteins from *Arabidopsis* are differentially expressed

It has been hypothesized that sulfation, by SOTs, affects the biological activity of certain compounds, thereby modulating physiological processes such as growth, development, and adaptation to stress (Varin *et al.*, 1997*b*; Marsolais *et al.*, 2000). In previous experiments, the mRNA levels for the three AtSOT proteins were analysed by RT-PCR, and it was concluded that all three genes were expressed constitutively in leaves, flowers, and siliques (Varin and Spertini, 2003). However, it was shown by Piotrowski *et al.* (2004) and in our experiments that the three *desulfoglucosinolate AtSOT* genes are indeed differentially expressed, with the respective mRNAs accumulating to different extents under various conditions. This may explain the differences in glucosinolate content among organs, developmental stages and environmental conditions.

The tissue contents and specificities of many glucosinolates have been analysed in detail. Dormant and germinating seeds have the highest concentrations, followed by inflorescences, siliques, leaves and roots. The total glucosinolate concentration in the different plant organs of *Arabidopsis*, ecotype Col0, varied nearly 100-fold ranging from the high of 63 µmol per g dry weight in seeds to the low of 0.7 µmol per g dry weight in senescent rosette leaves. The diversity varied considerably among organs, with 20 glucosinolates identified in siliques, 18 in seeds, 15 in inflorescences and cauline leaves, 12 in rosette leaves, and only 10 in roots. Whereas aliphatic glucosinolates predominate in most organs, indole glucosinolates made up nearly half of the total composition in roots and late-stage-rosette leaves (Brown *et al.*, 2003). The gene expression results obtained in this work support the proposal that all three desulfoglucosinolate AtSOT proteins are involved in the biosynthesis of glucosinolates but are active to different extents in the various organs of the plant. To determine the exact specificity of each AtSOT for desulfoglucosinolates in specific organs, it is necessary to investigate many more substrates *in vitro* and analyse the glucosinolate composition in desulfoglucosinolate *AtSOT* mutants. Because there are major differences in glucosinolate composition between seeds and the rest of the plant, it has been proposed that most of the glucosinolate content of the seeds is synthesized *de novo*

in the siliques (Petersen *et al.*, 2002). The level of expression of the *AtSOT* genes was lowest in siliques, but the enzyme activity may still be sufficient to synthesize new glucosinolate.

It was shown previously that, from a developmental perspective, older leaves have lower glucosinolate concentrations than younger leaves, but this was not due to decreasing concentrations in individual leaves with age (glucosinolate accumulation was stable during leaf expansion). Rather, leaves initiated earlier in development simply had much lower rates of glucosinolate accumulation per dry weight gain throughout their lifetimes. During leaf senescence, there were significant declines in glucosinolate concentration (Brown *et al.*, 2003). These glucosinolate contents are in agreement with the expression data of *AtSOT16* and *AtSOT17* (Fig. 4B); however, the mRNA levels of *AtSOT18* increased with age, suggesting that the sulfation of specific desulfoglucosinolates is needed during senescence, e.g. in defence against pathogens.

A high rate of catabolism has been suggested by studies showing dramatic (twofold) changes in glucosinolate concentration during a single diurnal cycle in *Brassica oleracea* (Rosa *et al.*, 1994). Our expression data show only small changes in the three *AtSOT* mRNAs levels. Therefore we postulate (post)translational regulation in the case where desulfoglucosinolate AtSOT proteins are rate-limiting for biosynthesis.

In previous experiments to test the effects of nutrient supply on glucosinolate concentration, the sulfate concentrations applied were 1500 µM (high) and 30 µM (low); in addition the nitrogen concentrations were varied (Hirai *et al.*, 2004; Hirai *et al.*, 2005). It was shown by microarray analysis that all genes involved in glucosinolate biosynthesis had an apparently similar expression pattern under sulfur deficiency (Hirai *et al.*, 2005) suggesting that the expression of these genes is controlled by the same regulatory mechanism. These results are consistent with those of glucosinolate metabolome analysis, implying the co-ordinated regulation of glucosinolate metabolism at the level of gene expression and metabolite accumulation. In contrast, the genes encoding enzymes involved in primary sulfur metabolism, which are known to be regulated not only at the level of mRNA accumulation, but also at enzymatic activity level, showed diverse levels of expression (Hirai *et al.*, 2005). In our experiments, any differences in the expression levels of the three *desulfoglucosinolate AtSOT* genes at different sulfur concentrations were almost undetectable. In comparison with the conditions chosen by Hirai *et al.* (2005) the differences in the sulfur status were much smaller (10 times vs. 50 times). The lower sulfate concentration was chosen because it represents the borderline for normal growth rates and should reflect conditions in the field for sulfurfertilized and nonfertilized *Brassica napus* plants (Ewald Schnug,

Braunschweig, Germany, personal communication). Under biotic or abiotic stress, differences in *desulfoglucosinolate AtSOT* expression may be more pronounced.

The results of hormone effects on the expression of *desulfoglucosinolate AtSOT* genes are diverse. According to Varin & Spertini (2003), expression of the three *desulfoglucosinolate AtSOT* genes was not affected by salicylic acid and jasmonic acid. On the other hand, the *AtSOT16* gene was first identified as an upregulated gene after treatment with the phytotoxin coronatine, which is an analogue of the octadecanoids 12-oxophytodienoic acid and/or jasmonic acid (Lopukhina *et al.*, 2001). It was also shown that the mRNA levels of *AtSOT18* were specifically and transiently increased after treatment with jasmonic acid and the ethylene precursor 1-aminocyclopropane-1-carboxylic acid, whereas abscisic acid, 2,4 dichlorophenoxyacetic acid, gibberellin A3, kinetin and salicylic acid did not alter *AtSOT18* mRNA levels (Piotrowski *et al.*, 2004). *B. napus* responds to methyl jasmonate treatment by increasing the accumulation of indole glucosinolate, whereas similar treatments led to increased accumulation of 2-phenylethylglucosinolate (Marsolais *et al.*, 2004). It would be very interesting to analyse the influence of environmental and endogenous stimuli on all genes and proteins involved in the pathway of glucosinolate biosynthesis and thereby to identify key regulatory steps in the pathway.

The *in vivo* action of the three expressed proteins needs to be analysed in detail by characterizing the respective knock-out and transgenic plants. One can assume that the repression or over-expression of each *desulfoglucosinolate AtSOT* gene will have specific effects on the glucosinolate pattern under certain conditions, even when a constitutive promoter is chosen. The approach has great potential for designing metabolically engineered plants with improved pest resistance and/or increased nutritional value.

Chapter 5:
Kinetics and substrate specificities of the three desulfo-glucosinolate sulfotransferase proteins in *Arabidopsis thaliana*

Klein M, Papenbrock J (2006), *Phytochemistry*, in preparation

Abstract

Sulfotransferases (SOTs) catalyze the transfer of a sulfate group from the co-substrate 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to a hydroxyl group of different substrates. In *Arabidopsis thaliana* three SOTs were identified to catalyze the last step of glucosinolate (Gl) core structure biosynthesis, called AtSOT16, 17 and 18. These enzymes from *Arabidopsis* ecotype C24 were overexpressed in *Escherichia coli* and purified by affinity chromatography. Recombinant proteins were used to determine substrate specificities to investigate the possibility that each of the three desulfo-glucosinolate (ds-Gl) SOTs influences the Gl pattern of *Arabidopsis*. After optimization of the enzyme assay it was possible to measure *in vivo* substrates with non-radioactive PAPS by HPLC analysis of the product. *In vitro* enzyme assays revealed a preference of AtSOT16 for the indolic ds-Gl indol-3-yl-methyl (I3M), AtSOT17 showed an increased specific activity with increasing chain-length of ds-Gls derived from methionine and AtSOT18 preferred the long-chain ds-Gls, 7-methylthioheptyl (7MTH) and 8-methyl-thiooctyl (8MTO), derived from methionine. *In planta* Gls exist side by side, therefore initial results from one-substrate measurements, were verified using a mixture of ds-Gls as substrates. To learn more about these enzymes *in vivo* AtSOT proteins were tested using Gl leaf extracts from *Arabidopsis* ecotype C24 as substrate. This study confirmed the *in vitro* measurements. To compare SOTs from different *Arabidopsis* ecotypes, additionally AtSOT18* from ecotype Col-0 was overexpressed in *E. coli* and purified. The recombinant protein was used for *in vitro* measurements and revealed a different behaviour on biochemical level compared to AtSOT18 from C24.

Introduction

Sulfotransferases (SOTs) catalyse the transfer of a sulfate group from the co-substrate 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to a hydroxyl group of ds-Gl with parallel formation of 3'-phosphoadenosine 5'-phosphate (PAP).

Twenty-two SOTs could be identified in *Arabidopsis thaliana* (L.) Heynh. For most of them, the according substrate specificity and therefore the *in vivo* function are not known yet. Three of them, AtSOT16, 17 and 18 (nomenclature is given in Klein and Papenbrock, 2004) are desulfo-glucosinolate (ds-Gl) SOT enzymes.

Gls are secondary metabolites found in the order Capparales, including agriculturally important crop plants of the Brassicaceae family such as oilseed rape (*Brassica napus*), fodder and vegetables (e.g. broccoli and cabbage) and the model plant *Arabidopsis*.

Intact Gls are non-toxic. Damage to plant tissue results in the hydrolysis of Gls, catalyzed by thioglucosidase ("myrosinase"), to produce a variety of degradation products, such as thiocyanates, isothiocyanates and nitriles. These breakdown products have a wide range of biological activities, which include both negative and positive effects. E.g., they can act as cancer-preventing agents and are involved in plant defense against pathogens and herbivores. Gl level is reduced in the seed of rape varieties due to breeding, because of antinutrional value of specific Gls. This allows to use protein-rich rape seed cake as animal food.

Gl biosynthesis can be divided into three stages: 1) precursor amino acids, such as methionine and tryptophan, are elongated by one or several methylene groups 2) precursor amino acids are converted into parent Gls 3) finally, parent Gls can be secondarily modified (Fig. 1) (Wittstock and Halkier, 2002). This last step in the biosynthesis of parent Gls is catalyzed by ds-Gl SOTs.

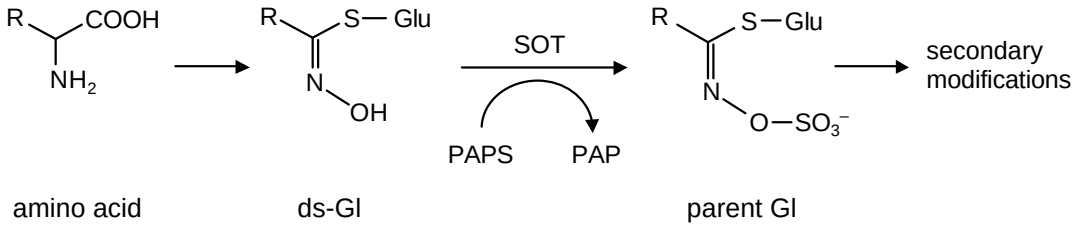


Fig. 1. Glucosinolate (Gl) biosynthesis. The precursor amino acid is elongated by methylen group(s), converted to parent Gl, followed by secondary modifications. ds, desulfo; Gl, glucosinolate; SOT, sulfotransferase; PAPS, 3'-phosphoadenosine 5'-phoshosulfate; PAP, 3'-phosphoadenosine 5'-phosphate.

The Gl pattern of *Arabidopsis* varies within the plant as well as among *Arabidopsis* ecotypes. Gl contents on organ level of *Arabidopsis* were analyzed at different developmental stages. Significant differences could be detected in both quality and quantity of Gls (Brown *et al.*, 2003). Due to determination of Gl content of 39 *Arabidopsis* ecotypes 34 different Gls have been identified in the genus *Arabidopsis* so far. (Kliebenstein *et al.*, 2001*a*).

The aim of this study is to elucidate the role of the three ds-Gl SOTs in organism. We would like to investigate the influence of these proteins on the Gl pattern. To reach this goal, *in vitro* specificities with naturally occuring substrates and *in vivo* specificities with a leaf extract from ecotype C24 including all Gls in the ds form were measured. Finally, the influence of SOT proteins on the *in vivo* Gl pattern was investigated.

Experimental procedures

DNA cloning techniques

The three genes At1g74100, At1g18590, and At1g74090, encoding AtSOT16 to 18 from *Arabidopsis* ecotype C24 and AtSOT18* from ecotype Col-0, respectively were cloned as described in Klein *et al.* (2006). Nomenclature is given in Klein and Papenbrock, 2004.

Expression and purification of *Arabidopsis* sulfotransferase proteins

The four different AtSOT proteins (AtSOT16, AtSOT17, AtSOT18 ecotype C24, and AtSOT18* ecotype Col-0) were overexpressed in *E. coli* and purified as described (Hirai *et al.*, 2005; Klein *et al.*, 2006). The purified recombinant proteins were dialyzed overnight at 4°C in 20 mM Tris / HCl, pH 8.0 plus 1 mM DTT and used for enzyme activity measurements.

The purity of the recombinant protein was analyzed by SDS-polyacrylamide gel electrophoresis and subsequent Coomassie Brilliant Blue stain. Protein contents were determined by the method of Bradford (1976) with bovine rinder albumin (Roth, Karlsruhe Germany) as a protein standard.

Preparation of substrates

The ds forms of the parent Gls derived from methionine and tryptophan were prepared as described by Graser *et al.* (2001).

The following Gls were used in the experiments in their ds forms: 3-methylthiopropyl Gl (3MTP) from *Erysimum pumillum*, 4-methylthiobutyl Gl (4MTB) from *Eruca sativa*, 5-methylthiopentyl Gl (5MTP) from *Arabidopsis thaliana*,

7-methylthioheptyl Gl (7MTH) from *Nasturtium officinale*, 8-methylthiooctyl Gl (8MTO) from *Arabis stelleri*, indol-3-ylmethyl Gl (I3M) from *Isatis tinctoria*. Structures of all *Arabidopsis* Gls are shown in Reichelt *et al.* (2002).

For the production of an *in vivo* ds-Gl extract, leaves from *Arabidopsis* ecotype C24 (five week-old plants) were used. The ds-Gl extract was set up as described by Graser *et al.* (2001).

Enzyme activity measurements and HPLC analysis

One-substrate measurements were done in the following way: The enzyme assays with recombinant proteins were set up as described by Klein *et al.* (2006). For determination at substrate specifities in one-substrate measurements, the concentration of the recombinant protein as well as the incubation time were optimized for each protein/substrate combination. Substrate and co-substrate concentrations were set constant at 60 µM.

ds-Gl mixture: Measurements with mixtures of the isolated ds-Gls contained 3 µg recombinant protein and 60 µM mixture of parent Gls. This ds-Gl mixture containted in equal parts ds3MTP, ds4MTB, 7MTH, 8MTO and dsI3M. PAPS concentration at 60 µM and incubation time (20 min) were constant.

in vivo ds-Gl mixture: For investigations of *Arabidopsis* ecotype C24 leaf extract, the extract was set up as described by Graser *et al.* (2001). This *in vivo* leaf extract contained a mixture of parent Gls and Gls with secondary modifications in ds form. For determination of substrate specificities, 5 µl of leaf extract as substrate and 7.5 µg recombinant protein were used. PAPS concentration at 60 µM and incubation time (20 min) were constant.

Generally, for the determination of the specific activities of the one-substrate measurement, the ds-Gls mixture and the *in vivo* ds-Gl mixture, the 150 µl assay contained: 80 mM Tris / HCl, pH 9.0, 9.2 mM $MgCl_2$, 60 µM PAPS (Calbiochem or obtained from Professor H. R. Glatt, Institute of Human Nutrition, Berholz-Rehbruecke, Germany), the respective substrate (60 µM), and 0.5 to 7.5 µg purified protein.

For the determination of the K_m values of ds3MTP and ds4MTB, at least 10 different substrate concentrations were used (20 µM to 220 µM). PAPS concentration was constant at 60 µM, whereas it was not limiting.

The 150 µl assay for determination of K_m values contained: 80 mM Tris / HCl, pH 9.0, 9.2 mM $MgCl_2$, ds3MTP or ds4MTB (20 µM to 220 µM), 2 to 7.5 µg purified protein and 60 µM PAPS (Calbiochem / Professor H. R. Glatt).

The reactions were started by the addition of PAPS, incubated for 5 to 20 min at 37°C, and stopped by incubation at 95°C for 10 min. The formation of the respective

sulfated product was analysed by HPLC (Knauer, Berlin, Germany) as described (Klein *et al.*, 2006).

Results

Determination of K_m values of two parent glucosinolates using the *Arabidopsis* sulfotransferase proteins

In the comparison of specific activities of both AtSOT18 proteins from different ecotypes, namely from ecotye C24 and Col-0, different behaviours could be observed (Klein *et al.*, 2006). We decided to investigate the phenomenon in more detail. After further optimizing of enzyme measurements it was now possible to determine K_m values with naturally occuring substrates and with non-radioactive PAPS. Due to adding of DTT (a reducing agent) during dialysis after expression and purification of recombinant proteins, the sensitivity of enzyme assays was much higher than in former experiments (Hirai *et al.*, 2005; Klein *et al.*, 2006). The short-chain methylthio Gls ds3MTP and ds4MTB were used as representative substrates for the determination of K_m values using AtSOT18 proteins from both ecotyes C24 and Col-0 (see table 2 for abbreviations for ds-Gls). Additionally, K_m values for AtSOT16 and 17 from ecotype C24 with the same substrates were determined.

The kinetic parameters were deduced from Michaelis-Menten plots. The kinetic analysis revealed a much higher affinity (about twofold) of AtSOT18* from Col-0 for the investigated substrates than AtSOT18 from C24. AtSOT16 and 17 from ecotype C24 showed differences in the kinetic behaviour, but to a smaller extent. The results are summarized in table 1.

Table 1. Determination of V_{max} (pkatal × mg^{-1} protein) and K_m values (µM) for the recombinant AtSOT proteins using ds3MTP and ds4MTB as substrates. For determination of the K_m values, at least 10 data points were included (20 µM to 220 µM). PAPS concentration was constant at 60 µM. Each determination was carried out two times. ds3MTP, desulfo-3-methylthiopropyl Gl; ds4MTB, desulfo-4-methylthiobutyl Gl; Gl, glucosinolate; PAPS, 3'-phosphoadenosine 5'-phoshosulfate; AtSOT, *Arabidopsis thaliana* sulfotransferase.

Enzyme	Substrate	V_{max} (pkatal × mg^{-1} protein)	K_m (µM)
AtSOT16	ds3MTP	623 ± 66,7	70 ± 0
	ds4MTB	496 ± 15	80 ± 0
AtSOT17	ds3MTP	439 ± 214	88 ± 16
	ds4MTB	575 ± 16	65 ± 7
AtSOT18	ds3MTP	99 ± 17	100 ± 0
	ds4MTB	131 ± 17	130 ± 14
AtSOT18*	ds3MTP	2763 ± 266	55 ± 7
	ds4MTB	1883 ± 314	43 ± 6

Identification of *in vivo* substrates from leaves of *Arabidopsis* ecotype C24

Kinetic analysis of AtSOT16 and 17 with short methylthio Gls as substrates described above showed differences in their K_m values, but to a smaller extent. These results raise the question, why are there three ds-Gl SOTs in *Arabidopsis*? In order to answer this question, more *in vivo* substrates need to be investigated. Therefore we focused on Gl content from *Arabidopsis* leaves. Identified Gls are summarized in table 2. They are grouped according to their chemical properties. Precursor amino acids and parent Gls from which identified Gls were deduced are listed in table 2. It is generally agreed that ds-Gl SOTs sulfate parent Gls, secondary modifications take place after producing parent Gls. For this reason we decided subsequently to investigate these parent Gls to see whether ds-Gl SOTs are substrate specific or not.

Table 2. Glucosinolate (Gl) contents from leaf extracts of *Arabidopsis* ecotype C24 (four week-old plants). Identified Gls are grouped according to their chemical properties. Amino acids and parent Gls from which identified Gls are deduced are listed. 3MTP, 3-methylthiopropyl Gl; 4MTB, 4-methylthiobutyl Gl; 5MTP, 5-methylthiopentyl Gl; 7MTH, 7-methylthioheptyl Gl; 8MTO, 8-methylthiooctyl Gl; I3M, indol-3-yl-methyl Gl; 3MSOP, 3-methylsulfinylpropyl Gl; 4MSOB, 4-methylsulfinylbutyl Gl; 7MSOH, 7-methylsulfinylheptyl Gl; 8MSOO, 8-methylsulfinyloctyl Gl; 1MOI3M, 1-methoxy-indol-3-yl-methyl Gl; 4MOI3M, 4-methoxy-indol-3-yl-methyl Gl; Gl(s), glucosinolate(s).

glucosinolate	group	amino acid	parent glucosinolate
allyl (2-propenyl)	alkyl (aliphatic)	methionine	3MTP
3-butenyl			4MTB
R2OH3butenyl			4MTB
S2OH3butenyl			4MTB
4-pentenyl			5MTP
3MSOP	methylsulphinylalkyl (aliphatic)	methionine	3MTP
4MSOB			4MTB
7MSOH			7MTH
8MSOO			8MTO
3MTP	methylthioalkyl (aliphatic)	methionine	yes
4MTB			yes
7MTH			yes
8MTO			yes
I3M	indol	tryptophan	yes
4MOI3M			I3M
1MOI3M			I3M

Are desulfo-glucosinolate sulfotransferase proteins of *Arabidopsis* substrate specific?

To determine the substrate specificities of the four recombinant AtSOT proteins, they were overexpressed in *E. coli*, purified and used for enzyme activity measurements. The *in vitro* assays were set up as described in "experimental procedures". Following parent Gls were used as substrates: ds3MTP, ds4MTB, ds5MTP, ds7MTH, ds8MTO and dsI3M. The products were analysed using HPLC analysis.

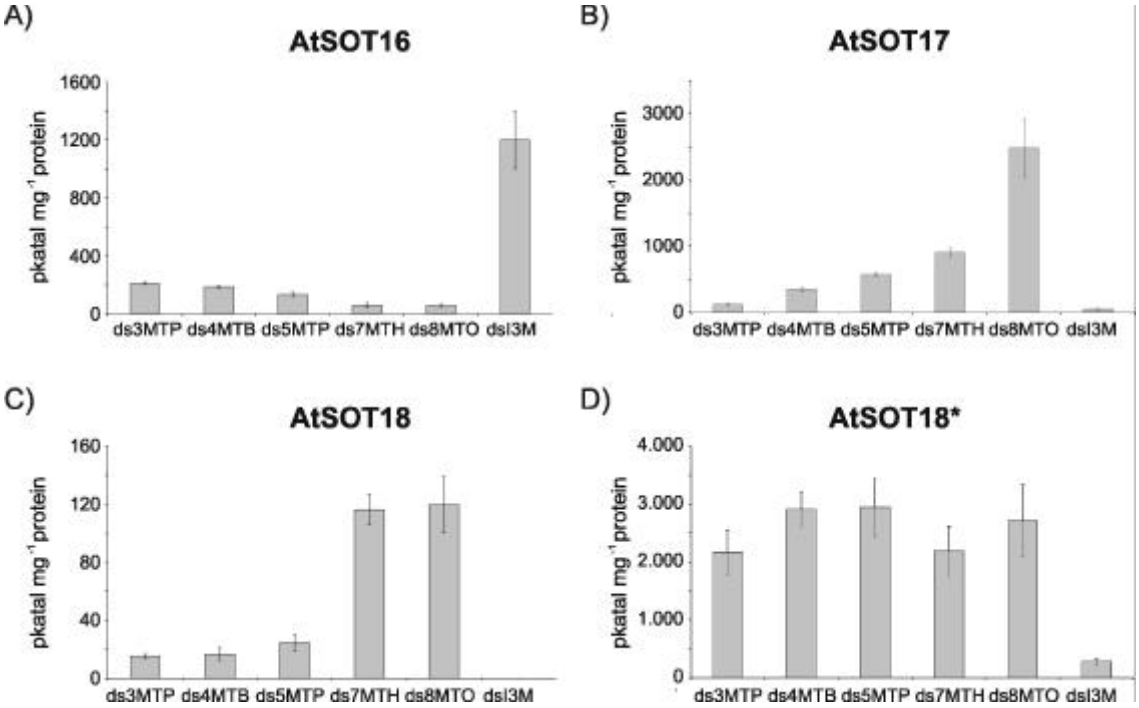


Fig. 2. Determination of substrate specificities of AtSOT proteins using different parent glucosinolates (Gls) as substrates. The specific activities (pkatal × mg^{-1} protein) of A) AtSOT16, B) AtSOT17, C) AtSOT18 (all C24) and D) AtSOT18* (Col-0) with parent Gls. ds3MTP, desulfo-3-methylthiopropyl Gl; ds4MTB, desulfo-4-methylthiobutyl Gl; ds5MTP, desulfo-5-methylthiopentyl Gl; ds7MTH, desulfo-7-methylthioheptyl Gl; ds8MTO, desulfo-8-methylthiooctyl Gl; dsI3M, desulfo-indol-3-yl-methyl Gl; Gl(s), glucosinolate(s); AtSOT, *Arabidopsis thaliana* sulfotransferase. The standard deviation was calculated (n=3).

AtSOT16 from ecotype C24 showed the highest enzymatic activity with the tryptophan-derived dsI3M of 1200 pkatal × mg^{-1} protein (Fig. 2A). AtSOT16 was active with all other offers of parent Gls (methionine-derived) as well, but with lower specific activities. Compared to ds3MTP (about 200 pkatal × mg^{-1} protein), the activity is 6 times lower. Regarding the methionine-derived Gls, the activity seems to depend on the chain-length of Gls. Activity with short Gls is higher than with longer-chain Gls (Fig. 2A).

AtSOT17 from ecotype C24 protein shows a completely different behaviour than AtSOT16. This protein seems to prefer the methionine-derived ds-Gls as substrates. The

specific activity increased with increasing chain-length of the Gls. The highest value is 2486 pkatal × mg^{-1}protein with ds8MTO. The lowest value is 122 pkatal × mg^{-1} protein with ds3MTP (20 times lower) (Fig. 2B).

AtSOT18 from ecotype C24 is active with methionine-derived ds-Gls only. No activity could be observed with the tryptophan-derived ds-Gl I3M. At SOT18 seems to make a distinction between long- and short-methylthio ds-Gls. AtSOT18 prefers the long-chain ds-Gls 7MTH and 8MTO (Fig. 2C). AtSOT17 and 18 favour the long-methylthio Gls, whereas AtSOT17 shows much higher specific activities with all offered Gls. For example the specific activity for AtSOT17/8MTO is 20 times higher than for AtSOT18/8MTO (~ 2500 vs. ~ 120 pkatal × mg^{-1} protein).

AtSOT18* from ecotype Col-0 showed high activity with all offered methionine-derived ds-Gls in the same value range (between 2167 and 2943 pkatal × mg^{-1} protein). Compared to the respective SOT from ecotye C24, there is no difference between short- and long-chain methylthio ds-Gls. Instead, the tryptophan-derived ds-Gl I3M revealed a much lower value of 279 pkatal × mg^{-1} protein. Interestingly, one can observe differences between the ecotypes. AtSOT18 protein from ecotype C24 showed a preference for the long-chain ds-Gls from methionine. AtSOT18* protein from ecotype Col-0 catalyzed all offered metionine-derived ds-Gls. The absolute specific activity is much higher for SOT 18* from Col-0 than for AtSOT18 from C24. An other difference is AtSOT18*'s activity with dsI3M, which is low but measurable (Fig. 2D).

In summary, it seems that the four investigated ds-Gl SOTs are active with all offered parent Gls (except AtSOT18 with dsI3M), but with certain preferences. There is a noticeable distinction between methionine- and tryptophan-derived ds-Gls. In most cases, there is even a subdivision in long- or short-chain ds-Gls within the group of methionine-derived ds-Gls. Finally, these enzymes seem to be substrate specific or at least substrate group specific.

Enzyme activity measurements with a mixture of parent glucosinolates using the *Arabidopsis* sulfotransferase proteins

In vivo ds-Gls exist side by side. To learn more about the behaviour of the four recombinant SOTs under competitive conditions, enzyme assays were set up as described in "experimental procedures". In this study SOTs 16, 17, 18 and 18* were incubated with a mixture of parent ds-Gls.

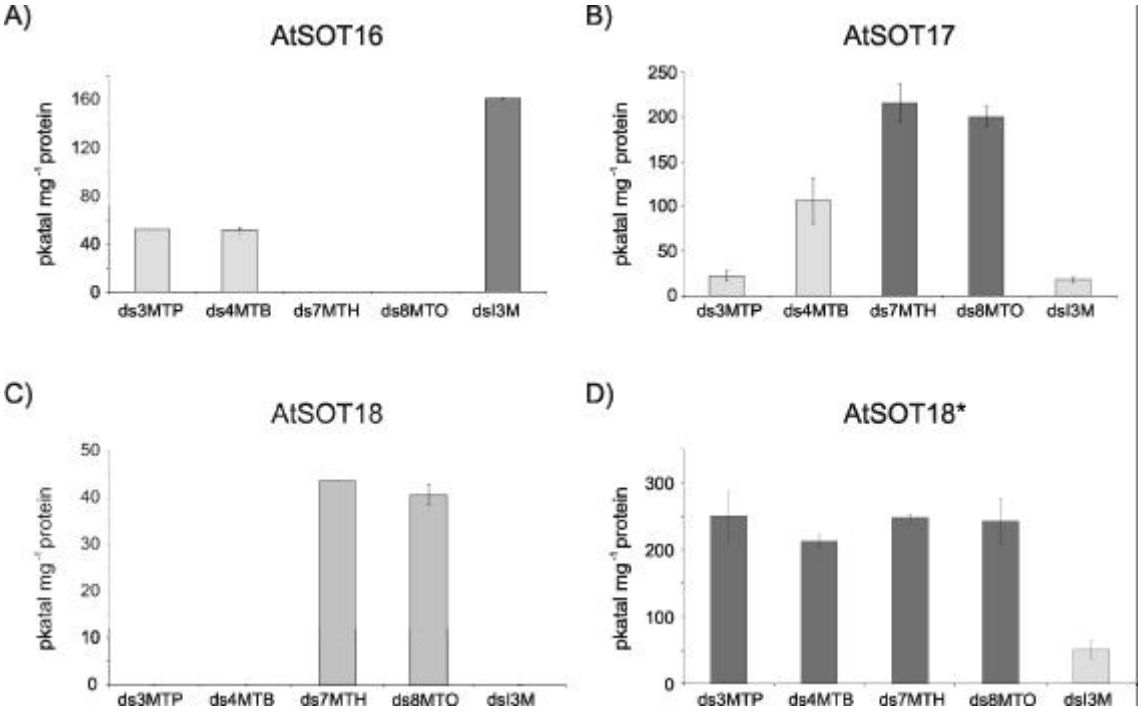


Fig. 3. Determination of substrate specificities of AtSOT proteins using a mixture of parent ds-Gls as substrate. The specific activities are given in pkatal × mg^{-1} protein. Shown are measurements of A) AtSOT16, B) AtSOT17, C) AtSOT18 (all C24) and D) AtSOT18* (Col-0) using a mixture of parent ds-Gls as substrate. Gl(s), glucosinolate(s); ds-Gl(s), desulfo Gl(s); ds3MTP, desulfo-3-methylthiopropyl Gl; ds4MTB, desulfo- 4-methylthiobutyl Gl; ds5MTP, desulfo-5-methylthiopentyl Gl; ds7MTH, desulfo-7-methylthioheptyl Gl; ds8MTO, desulfo-8-methylthiooctyl Gl; dsI3M, desulfo-indol-3-yl-methyl Gl; AtSOT, *Arabidopsis thaliana* sulfotransferase. The standard deviation was calculated (n=2). Dark grey bars represent used up-substrates; light grey bars are for used (but not used up) substrates.

Measurements have been performed with a mixture of the same parent ds-Gls used in the previously described one-substrate measurements. AtSOT16 completely used up dsI3M, followed by ds3MTP and ds4MTB (Fig. 3A).

AtSOT17 used up ds8MTO and ds7MTH. ds4MTB, 3MTP and dsI3M were converted in this order in much lower amounts (Fig. 3B).

To determine which substrates were preferred, AtSOT18 protein was incubated with the same mixture of parent ds-Gls (Fig. 3C). It seems to be that the long-chained methionine-derived ds-Gls (ds7MTH, ds8MTO) are preferred.

Measurements with AtSOT18* from ecotype Col-0 revealed that all methionine-derived (long and short) ds-Gls were used up, dsI3M was converted in a lower extent. This result is in agreement with the one-substrate measurements described above (Fig. 3D).

To analyze the order of converted substrates in this mixture of ds-Gls using AtSOT17, measurements have been performed under limiting conditions, that means it was incubated 10 and 20 min (Fig. 4). This experiment revealed an increased activity

for ds8MTO and ds7MTH at 10 min incubation time compared to the 20 min incubation. The activities for the other ds-Gls (ds3MTP, ds4MTB, dsI3M) remain about the same. No specific order of substrate conversion could be detected. Obviously, the ds-Gls offered in a mixture are used independently from each other.

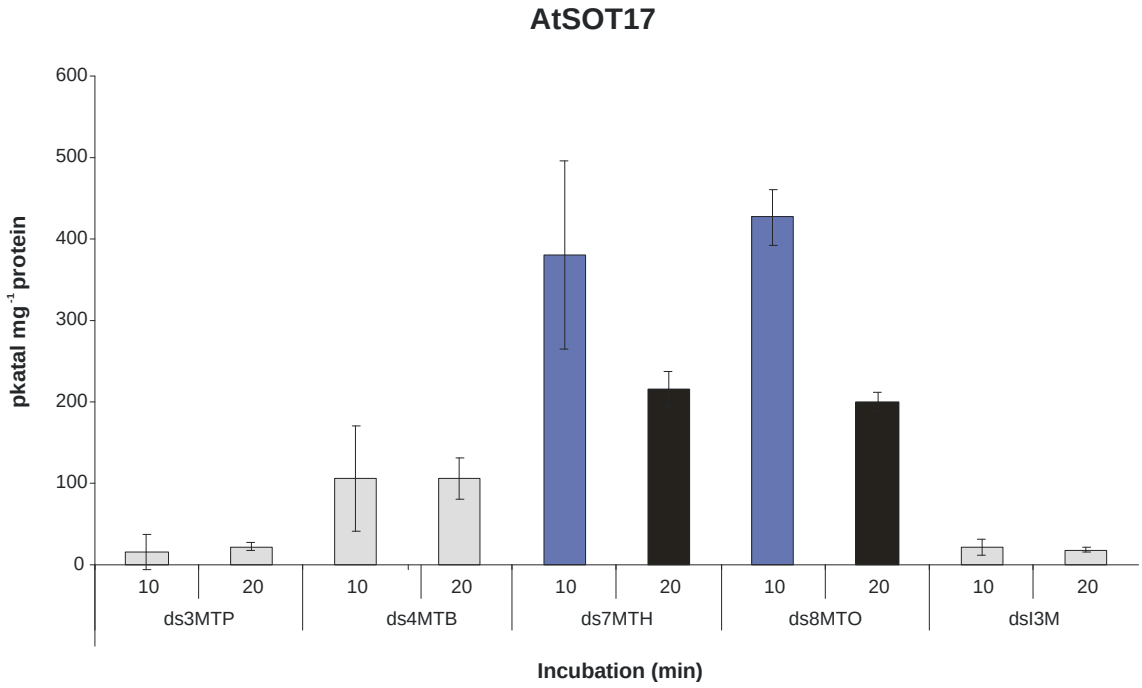


Fig. 4. Measurements of AtSOT17 with a mixture of parent ds-Gls under limiting conditions (10 and 20 min). The specific activity is given in pkatal mg^{-1} protein. Gl(s), glucosinolate(s); ds-Gl(s), desulfo Gl(s) ds3MTP, desulfo-3-methylthiopropyl Gl; ds4MTB, desulfo-4-methylthiobutyl Gl; ds5MTP, desulfo-5-methylthiopentyl Gl; ds7MTH, desulfo-7-methylthioheptyl Gl; ds8MTO, desulfo-8-methylthiooctyl Gl; dsI3M, desulfo-indol-3-yl-methyl Gl; AtSOT, *Arabidopsis thaliana* sulfotransferase. The standard deviation was calculated (n=2). Black marked bars represent used up substrate; blue marked bars represent almost used up substrate; grey bars indicate used (but not used up) substrate.

In summary, one can observe that these results confirm the results of one-substrate measurements displayed above (Fig. 2 and Fig. 3). Investigated SOTs showed a preference for specific ds-Gls. Under limiting competitive conditions, no specific order of converted substrates could be detected (Fig. 4).

Analysis of enzyme activity using C24 leaf extract

To learn more about the functions of the three SOTs *in vivo*, enzyme assays were set up as described in “experimental procedures” using recombinant AtSOT proteins from ecotype C24 and AtSOT18* from ecotype Col-0 with ds-Gl extract from leaves (ecotype C24) as substrate mixture. This leaf extract was used to simulate the *in vivo* situation. The results are summarized in table 3.

Table 3. HPLC analysis of enzyme assays using AtSOT proteins with ds-Gl from leaf extracts of *Arabidopsis* ecotype C24. Identified products (intact Gl) were confirmed by mass spectrometry and UV spectra (marked grey). Amino acids and parent Gls, from which the identified intact Gls are deduced, listed at the top. Intact Gls are grouped according to their chemical properties (at the top). The bottom part shows identified (×) products using AtSOT16, 17, 18 and 18*. Gl(s), glucosinolate(s); ds-Gl(s), desulfo Gl(s); intGl, intact Gl; 4MTB, 4-methylthiobutyl Gl; 7MTH, 7-methylthioheptyl Gl; 8MTO, 8-methylthiooctyl Gl; I3M, indol-3-yl-methyl; 8MSOO, 8-methylsulfinyloctyl Gl; 4MOI3M, 4-methoxy-indol-3-yl-methyl Gl; AtSOT, *Arabidopsis thaliana* sulfotransferase.

group	indol	indol	alkyl (aliphatic)	methyl-thioalkyl (aliphatic)	methyl-thioalkyl (aliphatic)	methyl-sulphinylalkyl (aliphatic)
amino acid	trpytophan	trpytophan	methionine	methionine	methionine	methionine
parent Gl	I3M	I3M	4MTB	7MTH	8MTO	8MTO
intGl (product)	I3M	4MOI3M	3-butenyl	7MTH	8MTO	8MSOO
AtSOT16	´	´	´			
AtSOT17				´	´	´
AtSOT18				´	´	
AtSOT18*			´	´	´	

In the enzyme assay using AtSOT16, I3M, 4MOI3M (both trytophan-derived) and 3-butenyl (methionine-derived) were formed. AtSOT17 preferred the long-chain aliphatic ds-Gls from methionine: 7MTH, 8MTO and 8MSOO. Enzyme assays using AtSOT18 protein produced 7MTH and 8MTO. In contrast, in the enzyme assay using AtSOT18* from ecotype Col-0 long- and short-chain Gls were formed: 7MTH, 8MTO and 3-butenyl (all derived from methionine).

In summary, this experiment is in agreement with the one-substrate measurements and measurements with the mixture of ds-Gls (see above). In all cases, AtSOT16 preferred indolic tryptophan-derived Gls. AtSOT17 and 18 are active with long-chain aliphatic methionine-derived Gls. AtSOT18* from Col-0 showed a broader substrate specificity compared to AtSOT18 from C24 with a preference for aliphatic methionine-derived Gls (long- and short-chain Gls).

Discussion

Kinetic analysis and substrate specificities of desulfo-glucosinolate sulfotransferase proteins

The last step of the Gl biosynthesis is catalyzed by SOTs. It is thought that after formation of a core structure (the parent Gls), secondary modifications take place (Graser *et al.*, 2001; Kliebenstein *et al.*, 2001*b*). To date, little is known about these secondary modifications of parent Gls.

In *Arabidopsis* three ds-Gl SOTs were found, AtSOT16, 17 and 18. As the Gl pattern differs among *Arabidopsis* ecotypes (Kliebenstein *et al.*, 2001*a*), we decided to investigate these three ds-Gl SOTs from ecotype C24, which shows the broadest variety of Gls (Micheal Reichelt, Max Planck Institute for Chemical Ecology, Jena, Germany, unpublished data) as well as one exemplary SOT from ecotype Col-0. Gl composition and concentration vary with the developmental stage in different plant organs (Brown *et al.*, 2003). In the present study, the influence of SOTs on Gl patterns were investigated. To determinate if and how these three ds-Gl SOTs influence the Gl pattern, different *in vitro* enzyme assays were performed.

Both AtSOT18 proteins from ecotypes C24 and Col-0 showed different behaviour on enzymatic level, although they differ only in two out of 350 amino acids (Klein *et al.*, 2006). Therefore, we decided to investigate this phenomenon in more detail. After optimization of the enzyme assay it was possible to determine K_m values from *in vivo* substrates with non-radioactive PAPS. Due to adding of DTT (a reducing agent) during dialysis enzyme assays were much more sensitive. Exemplary ds-Gls ds3MTP and ds4MTB, short aliphatic Gls deduced from methionine, were chosen for the determination of K_m values. The results revealed an about two-times higher affinity of AtSOT18* for ds3MTP und ds4MTB than of AtSOT18 from C24 (table 1).

In previous experiments, kinetic studies with AtSOT16*, 17* and 18* from ecotype Col-0 were performed. For enzyme measurements naturally occuring substrates of *Arabidopsis*, but mainly ds-Gls with secondary modifications, were used. Finally, only few data using naturally occuring substrates from *Arabidopsis* without secondary modifications are available. AtSOT18* from Col-0 revealed for 4MTB a K_m value of 0.52 µM (Varin and Spertini, 2003). Compared to our results, this value is about 100-times lower (table 1). It is hard to compare our results with those of Varin and Spertini (2003), because of different enzyme assay conditions, e.g. lower PAPS/substrate concentrations and use of radioactive PAPS. Unfortunately, Varin and Spertini (2003) provide no standard deviation and there are no details about possible PAPS limitation

(because of the use of only 1 µM PAPS). Taken together, this may lead to the different results observed.

K_m values for AtSOT16 and 17 were determined as well and revealed differences among these enzymes, but to less smaller extents. In conclusion, it shows that there are differences among ecotypes, but only small differences between the three AtSOTs within ecotype C24. Thus, one question remains: why are there three SOTs? In order to answer this question, more *in vivo* substrates need to be measured.

A leaf extract from *Arabidopsis* C24 contains mainly aliphatic Gls, derived from methionine. Aside from aliphic ones, indolic Gls derived from tryptophan were found. These Gls orginated from following parent Gls: 3MTP, 4MTB, 5MTP, 7MTH, 8MTO and I3M (table 2). These naturally occuring parent Gls were investigated.

One-substrate measurements and measurements with a mixture of the mentioned parent Gls revealed a clear preference of AtSOT16 for I3M, the indolic Gl, followed by short methylthio Gls (3MTP and 4MTB). AtSOT17 and 18 favour long methylthio Gls, namely 7MTH and 8MTO. In general, AtSOT17 showed much higher specific activities. In summary, SOTs showed activtiy with (almost) all offered ds-Gls, but with noticable preferences.

To answer the initial question, in this study it could be shown, that the three enzymes have different substrate specifities. Additionally, a slightly differential expression has already been shown on the mRNA level (Klein *et al.*, 2006). Whether these three enzymes are regulated posttranscriptionally is not known yet. Our results lead to the hypothesis that the investigated ds-Gl SOTs influence the Gl pattern *in vivo*.

Comparison of ecotypes C24 and Col-0 shows that AtSOT18* from Col-0 seems to be less specific than AtSOT18 from C24. Obviously, the AtSOT18 proteins from the different ecotypes show a different behaviour. Maybe this contributes to the different Gl patterns within the ecotypes.

A previously published study of the three SOTs of *Arabidopsis* ecotype Col-0 showed different substrate specificities using a substrate dissappearance method under competitive conditions. In the investigation the 18* protein from Col-0 preferred long chain Gls from methionine (7MTH and 8MTO) as substrates (Piotrowski *et al.*, 2004). Measured activity was calculated by substrate dissappearance from two-substrate competition assays. It is hard to compare these findings to our results, because of different assay conditions (other PAPS/substrate concentration, different pH).

Another previous study about substrate specificities using AtSOT18* from Col-0 with different ds-Gls was published. Unfortunately, mainly no *in vivo* substrates from *Arabidopsis* and no ds-Gls with secondary modifications were used. In fact, specific

activities of only two *in vivo* parent Gls are available. AtSOT18* with ds4MTB reavealed a specific activity of 221/268 pkatal × mg^{-1} protein. AtSOT18* with dsI3M showed values of 30/113 pkatal × mg^{-1} protein (using 1 µM and 5 µM substrate for the determination, respectively) (Varin and Spertini, 2003). Compared to our results these specific activities are for 18* with ds4MTB 13 or 11- times lower, respectively. Specific activities for AtSOT18* with dsI3M give 9 or 2-times lower values, respectively (Fig. 2D). As mentioned above, differences in specific activities could be caused by the different assay conditions (lower PAPS and substrate concentrations, use of radioactive PAPS). Unfortunately, no information about possible PAPS limitations is available here as well, as only 1 µM PAPS was used in the experiments.

The final aim of our studies is to elucidate the function of ds-Gl SOTs *in planta*. So far including this study, *in vitro* assays were done to reach our aim. In our experiment, AtSOT proteins and a C24 leaf extract from *Arabidopsis* as substrate mixture were used to simulate the *in vivo* situation. Hence, this is a partially artificial, partially natural system. In conclusion, the substrate specificities found in pure *in vitro* assays could be confirmed. The other finding was that Gls with secondary modifications were sulfated as well. Unfortunately, less is known about secondary modifications of parent Gls (Graser *et al.*, 2001). In future work it could be interesting to verify the general acceptance that parent ds-Gls, not Gls with secondary modifications, are sulfated (Kliebenstein *et al.*, 2001*b*). On the other hand, whether the general acceptance is right, no secondarily modified Gls would exist in a ds form to interact with the SOTs. Therefore, it is possible, that artificially de-sulfated Gls with secondary modifications are sulfated, but with no *in vivo* relevance.

Speculations on modulations of the glucosinolate pattern and their biological effects in case of altered expression of desulfo-glucosinolate sulfotransferase proteins

The Gl-myrosinase system is an important anti-herbivore defense in Gl-containing plants. Upon tissue damage, myrosinase and Gls come into contact and Gls are hydrolysed. Many of these hydrolysis products are toxic to a wide range of organisms, such as bacteria, fungi, nematodes and insects (Chew, 1988; Louda and Mole, 1991; Rask *et al.*, 2000).

In a previously published study 122 *Arabidopsis* ecotypes were analyzed to learn more about Gl hydrolysis products. It was found that each ecotype produced mainly isothiocyanates (ITC) or epithionitriles/nitriles. This polymorphism is correlated with the expression of the epithiospecifier protein (ESP). ESP promotes the production of epithionitrile or nitrile, respectively, depending on the Gl side chain. In absence of ESP

predominately ITCs are produced. Ecotype Col-0 revealed no detectable ESP expression, whereas ecotype C24 showed ESP expression (Lambrix *et al.*, 2001).

The putative hydrolysis products of Gls in leaves of *Arabidopsis* ecotype C24 in the presence of ESP are summarized in table 4 (not experimentally confirmed). One can only speculate about changes in the Gl pattern due to modulation of AtSOT16, 17 and 18. In theory, on overexpression of protein 17 or 18, both preferred long chain ds-Gls from methionine (7MTH and 8MTO) (Fig. 2 and 3), nitrile and epithionitrile production would be increased due to hydrolysis of respective Gls. In contrast the production of oxazolidine-2-thione would be decreased, because the precursor Gl is a short Gl derived from methionine (table 4).

Table 4. Glucosinolate (Gl) content identified in leaves of *Arabidopsis* ecotype C24 (four week-old plants) and the putative hydrolysis products in presence of ESP. Which hydrolysis products are produced in theory is described in Lambrix *et al*., (2001); Mithen (2001) and Halkier and Gershenzon (2006). Gls were divided into groups according to their chemical properties. Parent Gls from which identified Gls are derived are listed. Gl(s), glucosinolate(s); ESP, epithiospecifier protein; IAA, indol-3-acetic acid; IAN, indolyl-3-acetonitrile; S, elemental sulfur; 3MTP, 3-methylthiopropyl Gl; 4MTB, 4-methylthiobutyl Gl; 5MTP, 5-methylthiopentyl Gl; 7MTH, 7-methylthioheptyl Gl; 8MTO, 8-methylthiooctyl Gl; I3M, indol-3-yl-methyl Gl; 3MSOP, 3-methylsulfinylpropyl Gl; 4MSOB, 4-methylsulfinylbutyl Gl; 7MSOH, 7-methylsulfinylheptyl Gl; 8MSOO, 8-methylsulfinyloctyl Gl; 1MOI3M, 1-methoxy-indol-3-yl-methyl Gl; 4MOI3M, 4-methoxy-indol-3-yl-methyl Gl.

glucosinolate	group	parent glucosinolate	putative hydrolysis products (in presence of ESP)
allyl (2-propenyl)	alkyl (aliphatic)	3MTP	Epithionitrile, Thiocyanate
3-butenyl		4MTB	Epithionitrile
R2OH3butenyl		4MTB	Epithionitrile, Oxazolidine-2-thione
S2OH3butenyl		4MTB	Epithionitrile, Oxazolidine-2-thione
4-pentenyl		5MTP	Epithionitrile
3MSOP	methylsulphinylalkyl (aliphatic)	3MTP	Epithionitrile
4MSOB		4MTB	
7MSOH		7MTH	
8MSOO		8MTO	
3MTP	methylthioalkyl (aliphatic)	yes	Nitrile
4MTB		yes	
7MTH		yes	
8MTO		yes	
I3M	indol	yes	IAN + S → IAA
4MOI3M		I3M	
1MOI3M		I3M	

Numerous studies have shown that Gl hydrolysis products are toxic to a wide range of organisms. The biological effects of Gl hydrolysis products on insects were summarized (Wittstock *et al.*, 2003). It was found that predominantly ITC is toxic to insects. Other hydrolysis products, nitriles and thiocyanates, were also found to be toxic to insects, but nothing is known about toxicity of epithionitriles and oxazolidine-2-thione. As oxazolidine-2-thione is known to cause goiter in mammals and has other harmful effects, this compound could act as a defense against mammalian herbivores (Wittstock *et al.*, 2003).

In expression studies an increased mRNA amount of *AtSOT17* was found especially in mature leaves in investigations on plant organ level and under sulfur deficiency. An increased mRNA accumulation could be observed in young plants up to two weeks. Compared to *AtSOT17, AtSOT18* mRNA was accumulated in five to six weeks old *Arabidopsis* plants, especially in the root and under sulfur deficiency (Hirai *et al*., 2005; Klein *et al.*, 2006). In *Arabidopsis* leaves more than 80% of the total Gls were contributed due to aliphatic Gls. The proportion of aliphatic Gls decreased with age (Brown *et al.*, 2003). In conclusion, modulation of the Gls pattern due to repression or overexpression of both AtSOT17 and 18 could help to find out whether the resulting hydrolysis products (nitriles and epithionitriles) are involved in plant defense. Thus, an overexpression of each protein may lead to an improved plant defense.

The hydrolysis of indolic Gls in presence of ESP leads to elemental sulfur and indolyl-3-acetonitrile (IAN), which has auxin activity. Subsequently IAN could be converted to the auxin indol-3-acetic acid (IAA), a plant hormon (Mithen, 2001; Halkier and Gershenzon 2006). Auxin is known to be involved in growth and development.

IAN was found to act as a feeding deterrent against the locust *Schistocerca gregaria* (El Sayed *et al.*, 1996), as a growth inhibitor of the plant-pathogenic fungus *Leptoshaeria maculans* (Mithen and Lewis, 1986) and a plant-pathogenic bacterium *Erwinia carotovora* (Brader *et al.*, 2001).

An icreased expression of *AtSOT16 gene* can be observed at sulfur deficiency in young *Arabidopsis* plants up to two weeks and especially in roots (Hirai *et al*., 2005; Klein *et al.*, 2006). A great proportion of indolic Gls, nearly half of the Gls content, was found in roots (Brown *et al.*, 2003). Taken together, this leads to the hypothesis that overexpression of SOT protein 16 may lead to increased production of IAN or IAA and elemental sulfur in roots. Under sulfur deficiency, increased elemental sulfur could be produced from indolic Gls and could be made available to the plant. In conclusion, hydrolysis products of indolic Gls as well as elemental sulfur can constitute to an improved defense-system against plant-pathogens and herbivores.

Chapter 6:
General Discussion

Why are glucosinolates of interest to humans

Glucosinolates (Gls) are of interest to humans because of their presence in *Brassica* vegetables (broccoli, cabbage) and condiments (mustard). The distinct taste and flavor are mainly due to certain Gl hydrolysis products (isothiocyanate = ITC) (Mithen, 2001; Halkier and Gerzhenzon, 2006).

Additionally, an agricultural interest in Gls exist because of their antinutritional effect of rapeseed caused due to specific Gls hydrolysis products (oxazolidine-2-thione). Due to breeding, almost no Gls are included anymore in the seed of a resulting "double zero" rape (Rosa *et al.*, 1997). This allows the protein rich seed meel (the remains after oil is pressed) to be used as animal food (Mithen, 2001; Halkier and Gerzhenzon, 2006). However, the Gl concentration in the rest of the plant should not decrease to keep up a defense against many herbivores and pathogens.

The importance of Gl research has increased because of the discovery of Gls as cancer-preventing compounds, such as 3-methylsulphinylpropyl and 4-methyl-sulphinylbutyl ITCs (Mithen, 2001; Halkier and Gerzhenzon, 2006). In *Arabidopsis* it was shown that predominantly ITCs or nitriles occur. The presence of epithiospecifier protein (ESP) promotes the formation of nitriles (Lambrix *et al.*, 2001). Therefore, knocking out ESP would support the formation of ITCs with anti-carcinogenic effects. This approach could lead to *Brassica* vegetables with improved nutritional values.

As it could be shown that specific Gl hydrolysis products can act toxic to herbivores and pathogens (Kliebenstein *et al.*, 2003), Gls have a great potential in use as biopesticides.

The significance of sulfotransferase proteins in glucosinolate biosynthesis and subsequent hydrolysis

The biosynthesis of Gls can be divided into tree steps. 1) Certain aliphatic and aromatic amino acids are elongated by methylene groups. 2) The precursor amino acids are converted to the parent glucosinolates. 3) Parent Gls can be modified secondarily (Wittstock and Halkier, 2002). This order of the Gl biosynthesis is supported by *in vivo* and *in vitro* studies (Graser *et al.*, 2001; Kliebenstein *et al.*, 2001*b*).

While the first part of Gl biosynthesis (amino acid → ds-Gl) is well understood, much less is known about the last part (sulfation and secondary modifications of parent

Gls) and the hydrolysis of Gls. Which hydrolysis product is produced depends on the Gl itself and on the presence of additional proteins and cofactors. The best known of those proteins is ESP, which promotes the production of nitriles (Lambrix *et al.*, 2001).

Thus, in future more investigations have to be done concerning the last part of Gl biosynthesis and hydrolysis. We need to know which proteins and cofactors exist next to ESP and how they work. To date, the function of the hydrolysis product ITC is well investigated, but less is known about the remaining Gl hydrolysis products (nitriles, thiocyanates, epithionitriles, oxazolidine-2-thione). Thus, the biological functions of all hydrolysis products have to be investigated in more detail.

This study focused on the SOT family of *Arabidopsis*, especially on AtSOT16 to 18. In summary, it could be shown that these three enzymes are involved in Gl biosynthesis (chapter 3, Hirai *et al.*, 2005), that they are localized in cytoplasm, that they are slightly differentially expressed and that they have different substrate specificities (chapter 2, Klein and Papenbrock, 2004; chapter 4, Klein *et al.*, 2006; chapter 5). Taken together, due to mentioned properties of the investigated SOTs, specific parent Gls are preferably produced. Which in turn results in an increased production of specific Gl hydrolysis products with different biological effects (chapter 5, table 4). Thus, one could hypothesize that specific ds-Gl SOTs contribute to specific attributes of the plant.

The role of sulfotransferase proteins in glucosinolate distribution within the plant

Plants synthesize a broad spectrum of compounds that are toxic to herbivores and pathogens. In plants, two forms of defense mechanisms exist, the induced and the constitutive defense. One example of constitutive defense compounds in plants are Gls. The production of defensive compounds is thought to be cost intensive for the plant, because of usage of resources for biosynthesis and the possible toxicity for the plant itself. One strategy to reduce these costs is to synthesize the defense compounds only after an induction, with the disadvantages in lost time of the plant response (Wittstock and Gershenzon, 2002).

Defense toxins may be toxic to the plant itself. Thus, plants have to store constitutive defense compounds without poisoning themselves. These compounds can be stored as inactive precursors, such as Gls. Gls are seperated from the activating enzyme myrosinase. On tissue damage, Gls and myrosinase come into contact hydrolyzing the Gls. The hydrolysis products are biologically active compounds with a wide range of effects. To date, the detoxification mechanisms in Gls-containing plants are unknown (Wittstock and Gershenzon, 2002). Nitrilase and methyltransferases have been

suggested to be involved in *in planta* detoxification of the hydrolysis products nitriles and thiocyanates, respectively (Attieh *et al.*, 2000*a* and 2000*b*; Vorwerk *et al.*, 2001).

It was shown that the genes for Gl synthesizing enzymes in *Arabidopsis*, *AtSOT16 to 18*, are increasingly transcribed under specific conditions. *AtSOT16* is regulated up after treatment with coronatine, an analog of the octadecanoid signaling molecule, which is known to be involved in plant defense (Piotrowski *et al.* 2004); all three *ds-Gls SOT* genes are increasingly expressed by sulfur deficiency (Hirai *et al.*, 2005). In *Arabidopsis*, Gls vary in composition and concentrations in different plant organs. Highest Gl concentrations are found in seeds, followed by inflorescences, siliques, leaves and roots (Brown *et al.*, 2003).

In summary, knowledge from literature and results of this study indicate:

1. that Gls seem to be synthesized constitutively (Wittstock and Gershenzon, 2002).
2. ds-Gl expression can be upregulated under specfic environmental conditions and endogenous stimuli (chapter 3, Hirai *et al.*, 2005; Piotrowski *et al.*, 2004)
3. the distribution of Gls within the plant is not homogeneous, as reproductive organs seem to be more protected than other plant organs (Brown *et al.*, 2003).
4. *AtSOT16, 17*and *18* are differentially expressed in plant organs. In contrast to Gl concentration *in planta*, the lowest expression of the three genes could be observed in siliques, flowers and young leaves (chapter 4, Klein *et al.*, 2006). Thus, there seems to be no correlation between *SOT* expression and Gl concentration on plant organ level.

In conclusion, there is no direct correlation between plant organs with a high level of Gls and SOT expression. In a former study it was shown that Gls are transported in the phloem in *Arabidopsis*. Phloem exudates were shown to contain sulfated Gls (and not ds-Gls) as transport form (Chen *et al.*, 2001). The fact that highest concentrations of Gls were found in reproductive organs of *Arabidopsis*, while the expression of SOTs in the same organs is significanly decreased compared with other plant organs, supports the idea that Gl sulfation takes place in those plant organs with high ds-Gl SOT expression, followed by transport in the phloem to other parts of the plant. Additionally, from the chemical point of view, due to sulfation the water solubility of Gls is increased.

On the other hand, advantages for the plant of using ds-Gls as transport form could be that Gls in ds form are not subject to myrosinase activity. Thus, the plant is protected against autotoxic formation of toxic Gl hydrolysis products, for example ITC.

Suggestions for further experiments to learn more about the role of sulfotransferases in the plant

To learn more about the functions of ds-Gl SOTs *in vivo*, more investigations need to be performed.

To find out whether SOT proteins are able to alter the Gl pattern, analysis of mutants in *AtSOT16*, *17* and *18* genes will be helpful. By knocking out one of three ds-Gl SOTs, it is possible that another SOT compensates the function of the knocked-out SOT. In this case, it will be maybe necessary to grow the mutants under certain conditions such as sulfur deficiency to prohibit the takeover of functions of another ds-Gl SOT.

Another approach could be to investigate mutants under further stress conditions, such as pathogen infection, to find out whether the three ds-Gl SOTs are involved in plant defense or whether there have other functions.

To date, the function of most Gl hydrolysis products is not investigated well (except ITCs). Therefore, more studies are necessary to get more information about the functions of the remaining Gl hydrolysis products.

Conclusion

The aim of this study is to find out the functions of AtSOT16 to 18 *in vivo*. To reach this goal, expression studies were performed, subcellular localization and substrate specificities were determined. The results revealed that the respective genes are slightly differently expressed, that the three enzymes are localized in cytoplasm and that they have different substrate specificities. This findings indicate that ds-Gl SOTs could be able to modify the Gl pattern in plants und subsequently promote the increase or decrease of specific hydrolysis products. This may lead to metabolic engineered plants with improved pest resistence or vegetables with improved nutritional value.

References

Abe T, Kanaya S, Kinouchi M, Ichiba Y, Kozuki T, Ikemura T (2003) Informatics for unveiling hidden genome signatures. *Genome Res*. **13**, 693–702

Aharoni A, Keizer LC, Bouwmeester HJ, Sun Z, Alvarez-Huerta M, Verhoeven HA, Blaas J, van Houwelingen AM, De Vos RC, van der Voet H, Jansen RC, Guis M, Mol J, Davis RW, Schena M, van Tunen AJ, O'Connell AP (2000) Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. *The Plant Cell* **12**, 647–662

Attieh J, Kleppinger-Sparace KF, Nunes C, Sparace SA, Saini HS (2000*a*) Evidence implicating a novel thiol methyltransferase in the detoxification of glucosinolate hydrolysis products in *Brassica oleracea* L. *Plant Cell Environ*. **23**, 165-174

Attieh J, Sparace SA, Saini HS (2000*b*) Purification and properties of multiple isoforms of a novel thiol methyltransferase involved in the production of volatile sulfur compounds from *Brassica oleracea*. *Arch. Biochem. Biophys*. **380**, 257-266

Bauer M, Dietrich C, Nowak K, Sierralta WD, Papenbrock J (2004) Intracellular localization of sulfurtransferases from *Arabidopsis thaliana*. *Plant Physiol*. **135**, 916-926

Bauer M, Papenbrock J (2002) Identification and characterization of single-domain thiosulfate sulfurtransferases from *Arabidopsis thaliana*. *FEBS Lett*. **532**, 427-431

Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, Nikolau BJ, Mendes P, Roessner-Tunali U, Beale MH, Trethewey RN, Lange BM, Wurtele ES, Sumner LW (2004) Potential of metabolomics as a functional genomics tool. *Trends Plant Sci*. **9**, 418–425

Bones AM, Rossiter JT (1996) The myrosinase-glucosinolate system, its organisation and biochemistry. *Physiol. Plant*. **97**, 194-208

Borevitz JO, Xia Y, Blount J, Dixon RA, Lamb C (2000) Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *The Plant Cell* **12**, 2383–2394

Bowman KG, Bertozzi CR (1999) Carbohydrate sulfotransferases: Mediators of extracellular communication. *Chem. Biol*. **6**, R9-R22

Brader G, Tas E, Palva, ET (2001) Jasmonate-dependent induction of indole glucosinolates in *Arabidopsis* by culture filtrates of the nonspecific pathogen *Erwinia carotovora*. *Plant Physiol*. **126**, 849-860

Bradford MM (1976) A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem*. **72**, 248-254

Brown PD, Tokuhisa JG, Reichelt M, Gershenzon J (2003) Variation of glucosinolate accumulation among different organs and developmental stages of *Arabidopsis thaliana*. *Phytochem*. **62**, 471-481

Burow M, Kessler D, Papenbrock J (2002) Enzymatic activity of the *Arabidopsis* sulfurtransferase resides in the C-terminal domain but is boosted by the N-terminal domain and the linker peptide in the full length enzyme. *Biol. Chem*. **383**, 1363-1372

Capasso JM, Hirschberg CB (1984) Effect of atractylosides, palmitoyl coenzym A, and anion transport inhibitors on translocation of nucleotide sugars and nucleotide sulfate into Golgi vesicles. *J. Biol. Chem*. **259**, 4263-4266

Chen S, Petersen BL, Olsen CE, Schulz A, Halkier BA (2001) Long-Distance Phloem Transport of Glucosinolates in *Arabidopsis*. *Plant Physiol*. **127**, 194-201

Chew FS (1988) Biological effects of glucosinolates. In: *Biologically Active Natural Products* (Cutler HG, ed.), 155-181. American Chemical Society, Washington DC

Chomczynski P, Sacchi N (1987) Single step method of RNA isolation by guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156-159

Douglas Grubb C, Zipp BJ, Ludwig-Muller J, Masuno MN, Molinski TF, Abel S (2004) *Arabidopsis* glucosyltransferase UGT74B1 functions in glucosinolate biosynthesis and auxin homeostasis. *Plant J.* **40**, 893-908

Dudoit S, Fridlyand J, Speed TP (2002) Comparison of discrimination methods for the classification of tumors using gene expression data. *J. Am. Stat. Assoc.* **97**, 77–87

El Sayed G, Louveaux A, Mavratzotis M, Rollin P, Quinsac A (1996) Effects of glucobrassicin, epiprogoitrin and related breadkown products on locusts feeding : *Schouwia purpurea* and desert locust relationships. *Entomol. Exp. Appl.* **78**, 231-236

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005-1016

Fahey JW, Zalcmann AT, Talalay P (2001) The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochem.* **56**, 5-51

Falany CN (1991) Molecular enzymology of human liver cytosolic sulfotransferases. *Trends Pharmacol. Sci.* **12**, 255-259

Faulkner IJ, Rubery PH (1992) Flavonoids and flavonoid sulfates as probes of auxin transport regulation in *Cucurbita pepo* hypocotyl segments and vesicles. *Planta* **186**, 618-625

Field B, Cardon G, Traka M, Botterman J, Vancanneyt G, Mithen R (2004) Glucosinolate and amino acid biosynthesis in *Arabidopsis*. *Plant Physiol.* **135**, 828–839

Gidda SK, Miersch O, Levitin A, Schmidt J, Wasternak C, Varin L (2003) Biochemical and molecular characterization of a hydroxyjasmonate sulfotransferase from *Arabidopsis thaliana*. *J. Biol. Chem.* **278**, 17895-17900

Glendening TM, Poulton JE (1990) Partial purification and characterization of a 3'-phosphoadenosine 5'-phosphosulfate : desulfoglucosinolate sulfotransferase from cress (*Lepidium sativum*). *Plant. Physiol.* **94**, 811-818

Goossens A, Hakkinen ST, Laakso I, Seppanen-Laakso T, Biondi S, De Sutter V, Lammertyn F, Nuutila AM, Soderlund H, Zabeau M, Inze D, Oksman-Caldentey KM (2003) A functional genomics approach toward the understanding of secondary metabolism in plant cells. *Proc. Natl Acad. Sci. USA* **100**, 8595–8600

Graser G, Oldham NJ, Brown PD, Temp U, Gershenzon J (2001) The biosynthesis of benzoic acid glucosinolate esters in *Arabidopsis thaliana*. *Phytochem.* **57**, 23-32

Graser G, Schneider B, Oldham NJ, Gershenzon J (2000) The methionine chain elongation pathway in the biosynthesis of glucosinolates in *Eruca sativa* (Brassicaceae). *Arch. Biochem. Biophys.* **378**, 411-419

Griffiths DW, Birch ANE , Hillman JR (1998) Antinutritional compounds in the Brassicaceae: Analysis, biosynthesis, chemistry and dietary effects. *J. Hortic. Sci. Biotechnol.* **73**, 1-18

Guterman I, Shalit M, Menda N, Piestun D, Dafny-Yelin M, Shalev G, Bar E, Davydov O, Ovadis M, Emanuel M, Wang J, Adam Z, Pich-ersky E, Lewinsohn E, Zamir D, Vainstein A, Weiss D (2002) Rose scent: genomics approach to discovering novel floral fragrance-related genes. *The Plant Cell* **14**, 2325–2538

Habuchi O (2000) Diversity and functions of glycosaminoglycan sulfotransferases. *Biochim*. *Biophy*. *Acta* **1474**, 115-127

Hahlbrock K, Scheel D (1989) Physiology and molecular biology of phenylpropanoid metabolism. *Annu*. *Rev. Plant Physiol*. *Mol*. *Biol*. **41**, 339-367

Halkier BA (1999) Glucosinolates. In: *Naturally Occurring Glycosides: Chemistry, Distribution and Biological Properties* (Ikan R, ed.), 193-223. John Wiley & Sons Ltd, New York

Halkier BA, Gershenzon J (2006) Biology and Biochemistry of Glucosinolates. *Annu. Rev. Plant Biol*. **57**, 303-33

Hanai H, Nakayama D, Yang HP, Matsubayashi Y, Hirota Y, Sakagami Y (2000) Existence of a plant tyrosylprotein sulfotransferase: novel plant enzyme catalyzing tyrosine *O*-sulfation of preprophytosulfokine variants *in vitro*. *FEBS Lett*. **470**, 97-101

Hirai MY, Fujiwara T, Awazuhara M, Kimura T, Noji M, Saito K (2003) Global expression profiling of sulfur-starved *Arabidopsis* by DNA macroarray reveals the role of O-acetyl-l-serine as a general regulator of gene expression in response to sulfur nutrition. *Plant J*. **33**, 651–663

Hirai MY, Klein M, Fujikawa Y, Yano M, Goodenowe DB, Yamazaki Y, Kanaya S, Nakamura Y, Kitayama M, Suzuki H, Sakurai N, Shibata D, Tokuhisa J, Reichelt M, Gershenzon J, Papenbrock J, Saito K (2005) Elucidation of gene-to-gene networks in *Arabidopsis* by integration of metabolomics and transcriptomics. *J*. *Biol*. *Chem*. **280**, 25590-25595

Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuhara M, Arita M, Fujiwara T, Saito K (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **101**, 10205-10210

Hogge LR, Reed DW, Underhill EW, Haughn GW (1988) HPLC separation of glucosinolates from leaves and seeds of *Arabidopsis thaliana* and their identification using thermospray liquid chromatography/mass spectrometry. *J*. *Chromatogr*. *Sci*. **26**, 551-556

Honke K, Taniguchi N (2002) Sulfotransferases and sulfated oligosaccarides. *Medicinal Research Reviews* **22**, 637-654

Jones PR, Manabe T, Awazuhara M, Saito K (2003) A new member of plant CS-lyases. A cystine lyase from *Arabidopsis thaliana*. *J*. *Biol*. *Chem*. **278**, 10291–10296

Kakuta Y, Pedersen LG, Carter CW, Negishi M, Pedersen LC (1997) Crystal Structure of Estrogen Sulphotransferase. *Nat*. *Struct*. *Biol*. **4**, 904-908

Kanaya S, Kinouchi M, Abe T, Kudo Y, Yamada Y, Nishi T, Mori H, Ikemura T (2001) Analysis of codon usage diversity of bacterial genes with a self-organizing map (SOM): characterization of horizontally transferred genes with emphasis on the E. coli O157 genome. *Gene* **276**, 89–99

Klaassen CD, Boles JW (1997) The importance of 3'-phosphoadenosine 5'-phosphosulfate (PAPS) in the regulation of sulfation. *FASEB J*. **11**, 404-418

Klein M, Papenbrock J (2004) The multi-protein family of *Arabidopsis* sulphotransferases and their relatives in other plant species. *J*. *Exp*. *Bot*. **55**, 1809-1820

Klein M, Reichelt M, Gershenzon J, Papenbrock J (2006) The three desulfoglucosinolate sulfotransferase proteins in *Arabidopsis* have different substrate specificities and are differentially expressed. *FEBS Lett*. **273**, 122-136

Kliebenstein DJ, Kroymann J, Brown P, Figuth A, Pedersen D, Gershenzon J, Mitchell-Olds T (2001*a*) Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. *Plant Physiol*. **126**, 811-825

Kliebenstein DJ, Lambrix VM, Reichelt M, Gershenzon J (2001*b*) Gene Duplication in the Divesification of Secondary Metabolism: Tandem 2-Oxoglutarate-Dependent Dioxygenases Control Glucosinolate Biosynthesis in *Arabidopsis*. *The Plant Cell* **13**, 681-693

Kutz A, Muller A, Hennig P, Kaiser WM, Piotrowski M, Weiler EW (2002) A role for nitrilase 3 in the regulation of root morphology in sulphur-starving *Arabidopsis thaliana*. *Plant J*. **30**, 95–106

Lacomme C, Roby D (1996) Molecular cloning of a sulfotransferase in *Arabidopsis thaliana* and regulation during development and in response to infection with pathogenic bacteria. *Plant Mol*. *Biol*. **30**, 995-1008

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685

Lambrix V, Reichelt M, Mitchell-Olds T, Kliebenstein DJ, Gershenzon J (2001) The *Arabidopsis* Epithiospecifier Protein Promotes the Hydrolysis of Glucosinolates to Nitriles and Influences *Trichoplusia ni* Herbivory. *The Plant Cell*. **13**, 2793-2807

Lee RW, Huttner WB (1983) Tyrosine-*O*-sulfated proteins of PC12 pheochromocytoma cells and their sulfation by a tyrosylprotein sulfotransferase. *J*. *Biol*. *Chem*. **258**, 11326-11334

Leustek T (2002) Sulfate metabolism. In: Somerville CR, Meyerowitz EM, eds. *The Arabidopsis Book*. Rockville, MD, USA: American Society of Plant Biologists, Vol. doi/10.1199/tab.0017, http://www.aspb.org/publications/arabidopsis/

Li G, Quiros CF (2002) In planta side-chain glucosinolate modification in *Arabidopsis* by introduction of dioxygenase Brassica homolog BoGSL-ALK. *Theor*. *Appl*. *Genet*.**106**, 1116-1121

Lillig CH, Schiffmann S, Berndt C, Berken A, Tischka R, Schwenn JD (2001) Molecular and catalytic properties of *Arabidopsis thaliana* adenylyl sulfate (APS)-kinase. *Arch*. *Biochem*. *Biophys.* **392**, 303-310

Lohmann KN, Gan S, John MC, Amasino RM (1994) Molecular analysis of natural leaf senescence in *Arabidopsis thaliana*. *Physiol*. *Plant*. **92**, 322-328

Lopukhina A, Dettenberg M, Weiler EW, Holländer-Czytko H (2001) Cloning and characterization of a coronatine-regulated tyrosine aminotransferase from *Arabidopsis*. *Plant Physiol*. **126**, 1678-1687

Louda S, Mole S (1991) Glucosinolates: Chemistry and ecology. In: *Herbivores: Their Interaction with Secondary Plant Metabolites*. Vol. 1 The Chemical Participants (Rosenthal GA, Berenbaum MR, eds.), 123-164. Academic Press, San Diego

Mandon EC, Milla ME, Kempner E, Hirschberg CB (1994) Purification of the golgi adenosine 3´-phosphate 5´-phosphosulfate transporter, a homodimer within the membrane. *Proc*. *Natl Acad*. *Sci*. *USA* **91**, 10707-10711

Marsolais F, Gidda SK, Boyd J, Varin L (2000) Plant soluble sulfotransferases: Structural and functional similarity with mammalian enzymes. *Recent Advances in Phytochem*. **34**, 433-456

Marsolais F, Hernandez Sebastia C, Rousseau A, Varin L (2004) Molecular and biochemical characterization of BNST4, an ethanol-inducible steroid sulfotransferase from *Brassica napus*, and regulation of BNST genes by chemical stress and during development. *Plant Sci*. **166**, 1359-1370

Marsolais F, Varin L (1995) Identification of amino acid residues critical for catalysis and cosubstrate binding in the flavonol 3-sulfotransferase. *J. Biol. Chem.* **270**, 30458-30463

Marsolais F, Varin L (1997) Mutational analysis of domain II of flavonol 3-sulfotransferase. *Eur. J. Biochem.* **247**, 1056-1062

Marsolais F, Varin L (1998) Recent developments in the study of the structure-function relationship of flavonol sulfotransferases. *Chem. Biol. Interact.* **109**, 117-122

Maruyama-Nakashita A, Inoue E, Watanabe-Takahashi A, Yamaya T, Takahashi H (2003) Transcriptome profiling of sulfur-responsive genes in *Arabidopsis* reveals global effects of sulfur nutrition on multiple metabolic pathways. *Plant Physiol.* **132**, 597–605

Matsubayashi Y, Sakagami Y (1996) Phytosulfokine, sulfated peptides that induce the proliferation of single mesophyll cells of *Asparagus officinalis* L. *Proc. Natl Acad. Sci. USA* **93**, 7623-7627

Mellon FA, Bennet RN, Holst B, Williamson G (2002) Intact glucosinolate analysis in plant extracts by programmed cone voltage electrospray LC/MS: Performance and comparison with LC/MS/MS methods. *Anal. Biochem.* **306**, 83-91

Mercke P, Kappers IF, Verstappen FW, Vorst O, Dicke M, Bouwmeester HJ (2004) Combined transcript and metabolite analysis reveals genes involved in spider mite induced volatile formation in cucumber plants. *Plant Physiol.* **135**, 2012–2024

Mikkelsen MD, Naur P, Halkier BA (2004) *Arabidopsis* mutants in the C-S lyase of glucosinolate biosynthesis establish a critical role for indole-3-acetaldoxime in auxin homeostasis. *Plant J.* **37**, 770-777

Mikkelsen MD, Petersen BL, Olsen CE, Halkier BA (2002) Biosynthesis and metabolic engineering of glucosinolates. *Amino Acids* **22**, 279-295

Mithen R (2001) Glucosinolates - biochemistry, genetics and biological activity. *Plant Growth Regul.* **34**, 91-103

Mithen R, Lewis BG (1986) *In vitro* activity of glucosinolates and their products against *Leptosphaeria maculans*. *Trans. Br. Mycol. Soc.* **87**, 433-440

Mulder GJ and Jokob WB (1990) Sulfation. In: *Conjugation Reactions in Drug Metabolism*. 107-161, Taylor and Francis Ltd., New York

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant.* **15**, 473-497

Nakai K, Kanehisa M (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics* **14**, 897-911

Negishi M, Pedersen LG, Petrotchenko E, Shevtsov S, Gorokhov A, Kakuta Y, Pedersen LC (2001) Structure and function of sulfotransferases. *Arch. Biochem. Biophys.* **390**, 149-157

Niehrs C, Beisswanger R, Huttner WB (1994) Protein tyrosine sulfation, 1993 – an update. *Chem. Biol. Interact.* **92**, 257-271

Nikiforova V, Freitag J, Kempa S, Adamik M, Hesse H, Hoefgen R (2003) Transcriptome analysis of sulfur depletion in *Arabidopsis thaliana*: interlacing of biosynthetic pathways provides response specificity. *Plant J.* **33**, 633–650

Nowak K, Luniak N, Meyer S, Mendel RR, Hänsch R (2004) Fluorescent proteins in poplar: A useful tool to study promotor function and protein localization. *Plant Biol.* **6**, 65-73

Petersen BL, Andreasson E, Bak S, Agerbirk N, Halkier BA (2001) Characterization of transgenic *Arabidopsis thaliana* with metabolically engineered high levels of p-hydroxybenzylglucosinolate. *Planta* **212,** 612–618

Petersen BL, Chen S, Hansen CH, Olsen CE, Halkier BA (2002) Composition and content of glucosinolates in developing *Arabidopsis thaliana*. *Planta* **214**, 562-571

Pi N, Hoang MB, Gao H, Mougous JD, Bertozzi CR, Leary JA (2005) Kinetic measurements and mechanism determination of Stf0 sulfotransferase using mass spectrometry. *Anal. Biochem.* **341**, 94-104

Piotrowski M, Schemenewitz A, Lopukhina A, Müller A, Janowitz T, Weiler EW, Oecking C (2004) desulfo-Glucosinolate sulfotransferases from *Arabidopsis thaliana* catalyzing the final step in biosynthesis of the glucosinolate core structure. *J. Biol. Chem.* **49**, 50717-50725

Quackenbush J (2002) Microarray data normalization and transformation. *Nat. Genet*. **32**, (suppl.) 496–501

Rask L, Andreasson E, Ekbom B, Eriksson S, Pontoppidan B, Meijer J (2000) Myrosinase: gene family evolution and herbivore defense in *Brassicaceae*. *Plant Mol. Biol*. **42**, 93-113

Reichelt M, Brown PD, Schneider B, Oldham NJ, Stauber E, Tokuhisa J, Kliebenstein DJ, Mitchell-Olds T, Gershenzon J (2002) Benzoic acid glucosinolate esters and other glucosinolates from *Arabidopsis thaliana*. *Phytochem*. **59**, 663-671

Rivoal J, Hanson AD (1994) Choline-*O*-Sulfate biosynthesis in plants (identification and partial characterization of a salinity-inducible choline sulfotransferase from species of *Limonium* (Plumbaginaceae). *Plant Physiol* **106**, 1187-1193

Rosa EAS, Heaney RK, Rego FC, Fenwick GR (1994) The variation of glucosinolate concentration during a single day in young plants of *Brassica oleracea* var. *acephala* and *capitata*. *J. Sci. Food Agric*. **66**, 457-463

Rotte C, Leustek T (2000) Differential subcellular localization and expression of ATP sulfurylase and 5'-adenylylsulfate reductase during ontogenesis of *Arabidopsis* leaves indicates that cytosolic and plastid forms of ATP sulfurylase may have specialized functions. *Plant Physiol*. **124**, 715-724

Rouleau M, Marsolais F, Richard M, Nicolle L, Voigt B, Adam G, Varin L (1999) Inactivation of brassinosteroid biological activity by a salicylate-inducible steroid sulfotransferase from *Brassica napus*. *J. Biol. Chem*. **274**, 20925-20930

Saito K (2004) Sulfur assimilatory metabolism. The long and smelling road. *Plant Physiol.* **136**, 2443–2450

Schlesinger B, Breton F, Mock HP (2003) A hydroponic culture system for growing *Arabidopsis thaliana* plantlets under sterile conditions. *Plant Mol. Biol. Rep*. **21**, 449-456

Smith DW, Johnson KA, Bingman CA, Aceti DJ, Blommel PG, Wrobel RL, Frederick RO, Zhao Q, Sreenath H, Fox BG, Volkman BF, Jeon WB, Newman CS, Ulrich EL, Hegeman AD, Kimball T, Thao S, Sussman MR, Markley JL, Phillips GN Jr (2004) Crystal Structure of At2g03760, a putative steroid sulfotransferase from *Arabidopsis thaliana*. *Proteins* **57**, 854-857

Sugahara T, Liu CC, Pai G, Collodi P, Suiko M, Sakakibara Y, Nishiyama K, Liu MC (2003) Sulfation of hydroxychlorobiphenyl. Molecular cloning, expression, and functional characterization of zebrafish SULT1 sulfotransferases. *Eur. J. Biochem*. **270**, 2404-2411

Superti-Furga A (1994) A defect in the metabolic activation of sulfate in a patient with achondrogenesis type IB. *Am. J. Hum. Genet*. **55**, 1137-1145

Suter M, von Ballmoos P, Kopriva S, den Camp RO, Schaller J, Kuhlemeier C, Schurmann P, Brunold C (2000) Adenosine 5'-phosphosulfate sulfotransferase and adenosine 5'-phosphosulfate reductase are identical enzymes. *J. Biol. Chem.* **275**, 930-936

Thangsatd OP, Evjen K, Bones A (1991) Immunogold-EM localisation of myrosinase in *Brassicaceae*. *Protoplasma* **161**, 85-93

Thangsatd OP, Inversen TH, Slupphaug G, Bones A (1990) Immunocytochemical localisation of myrosinase in *Brassica napus* L. *Planta* **180**, 245-248

The Arabidopsis genome initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796-815

Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima J, Awazuhara M, Inoue E, Takahashi H, Goodenowe DB, Kitayama M, Noji M, Yamazaki M, Saito K (2005) Functional genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. *Plant J.* **42**, 218–235

Truchet G, Roche P, Lerouge P, Vasse J, Camut S, DeBilly F, Prome JC, Denarie J (1991) Sulphated lipo-oligosaccharide signals of *Rhizobium meliloti* elicit root nodule organogenesis in alfalfa. *Nature* **351**, 670-673

Urbanczyk-Wochniak E, Luedemann A, Kopka J, Selbig J, Roessner- Tunali U, Willmitzer L, Fernie AR (2003) Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. *EMBO Rep.* **4**, 989–993

Varin L, Chamberland H, Lafontaine JG, Richard M. (1997*a*) The enzyme involved in sulfation of the turgorin, gallic acid 4-*O*-(beta-D-glucopyranosyl-6'-sulfate) is pulvini-localized in *Mimosa pudica*. *Plant J.* **12**, 831-837

Varin L, DeLuca V, Ibrahim RK, Brisson N (1992) Molecular characterization of two plant flavonol sulfotransferases. *Proc. Natl Acad. Sci. USA* **89**, 1286-1290

Varin L, Ibrahim RK (1989) Partial purification and characterization of three flavonol-specific sulfotransferases from *Flaveria chloraefolia. Plant Physiol.* **90**, 977-981

Varin L, Marsolais F, Brisson N (1995) Chimeric flavonol sulfotransferases define a domain responsible for substrate and position specificities. *J. Biol. Chem.* **270**, 12498-12502

Varin L, Marsolais F, Richard M, Rouleau M (1997*b*) Biochemistry and molecular biology of plant sulfotransferases. *FASEB J.* **11**, 517-525

Varin L, Spertini D (2003) Desulfoglucosinolate sulfotransferases, sequences coding the same and uses thereof for modulating glucosinolate biosynthesis in plants. Patent WO 03/010318-A1, Concordia University, Canada (February, 6, 2003)

Vorwerk S, Biernacki S, Hillebrand H, Janzik I, Müller A, Weiler EW, Piotrowski M (2001) Enzymatic characterization of the recombinant *Arabidopsis thaliana* nitrilase subfamily encoded by the *NIT2/NIT1/NIT3*-gene cluster. *Planta* **212**, 508-516

Weckwerth W, Loureiro ME, Wenzel K, Fiehn O (2004) Differential metabolic networks unravel the effects of silent plant phenotypes. *Proc. Natl Acad. Sci. USA* **101,** 7809–7814

Weinshilboum RM (1986) Phenol sulfotransferase in humans: Properties, regulation, and function. *Federation Proc.* **45**, 2223-2228

Weinshilboum RM, Otterness DM (1994) Sulfotransferase Enzymes. In: *Handbook of Experimental Pharmacology* (Kauffman FC, ed.), 45-78. Springer, Berlin

Weinshilboum RM, Otterness DM, Aksoy IA, Wood TC, Her C, Raftogianis RB (1997) Sulfotransferase molecular biology: cDNAs and genes. *FASEB J*. **11**, 3-14

Westley J (1973) Rhodanese. *Advances in Enzymology* **39**, 327-368

Wittstock U, Gershenzon J (2002) Constitutive plant toxins and their role in defense against herbivores and pathogens. *Curr. Opin. Plant Biol.* **5,** 300-307

Wittstock U, Halkier BA (2002) Glucosinolate research in the *Arabidopsis* era. *Trends Plant Sci*. **7**, 263-270

Wittstock U, Kliebenstein DJ, Lambrix V, Reichelt M, Gershenzon J (2003) Glucosinolate hydrolysis and its impact on generalist and specialist insect herbivores. *Phytochem*. **37**, 101-125

Yamazoe Y, Nagata K, Ozawa S, Kato R (1994) Structural similarity and diversity of sulfotransferases. *Chem. Biol. Interact*. **92**, 107-117

Yang HP, Matsubayashi Y, Nakamura K, Sakagami Y (1999) *Oryza sativa PSK* gene encodes a precursor of phytosulfokine-α, a sulfated peptide growth factor found in plants. *Proc. Natl Acad. Sci. USA* **96**, 13560-13565

Yang HP, Matsubayashi Y, Nakamura K, Sakagami Y (2001) Diversity of *Arabidopsis* genes encoding precursors for phytosulfokine, a peptide growth factor. *Plant Physiol*. **127**, 842-851

Yoshihara T, Omer EA, Koshino H, Sakamura S, Kikuta Y, Koda Y (1989) Structure of tuber inducing stimulus from potato leaves. *Agric. Biol. Chem*. **53**, 2835-2837

Zaruba ME, Schwartz NB, Tennekoon GI (1988) Reconstitution of adenosine 3´-phosphosulfate transporter from rat brain. *Biochem. Biophys. Res. Commu*. **155**, 1271-1277

Erklärung

Ich versichere, dass die Dissertation selbständig verfasst und die benutzten Hilfsmittel und Quellen, sowie gegebenenfalls die zu Hilfsleistungen herangezogenen Institutionen, vollständig angegeben wurden und die Dissertation nicht schon als Diplomarbeit oder ähnliche Prüfungsarbeit verwendet worden ist.

Hannover, den 24. Oktober 2006

Publikationen

Klein M, Papenbrock J (2004) The multi-protein family of *Arabidopsis* sulphotransferases and their relatives in other plant species. *J. Exp. Bot.* **55**, 1809-1820

Hirai MY, Klein M, Fujikawa Y, Yano M, Goodenowe DB, Yamazaki Y, Kanaya S, Nakamura Y, Kitayama M, Suzuki H, Sakurai N, Shibata D, Tokuhisa J, Reichelt M, Gershenzon J, Papenbrock J, Saito K (2005) Elucidation of gene-to-gene networks in *Arabidopsis* by integration of metabolomics and transcriptomics. *J. Biol. Chem*. **280**, 25590-25595

Klein M, Reichelt M, Gershenzon J, Papenbrock J (2006) The three desulfoglucosinolate sulfotransferase proteins in *Arabidopsis* have different substrate specificities and are differentially expressed. *FEBS* **273**, 122-136

Klein M, Papenbrock J (2006) Kinetics and substrate specificities of the three desulfoglucosinolate sulfotransferase proteins in *Arabidopsis*. *Phytochemistry*, in preparation

Mein herzlicher Dank gilt:

Frau Prof. Dr. Jutta Papenbrock für die Überlassung dieses interessanten Themas sowie für die hervorragende Betreuung und Zusammenarbeit während meiner wissenschaftlichen Tätigkeit im Institut für Botanik. Ihre freundliche und zugleich kompetente Unterstützung waren mir eine große Hilfe!

Herrn Prof. Dr. B. Huchzermeyer für seine Bereitschaft, jederzeit Fragen zu diskutieren sowie für die Übernahme des Korreferats.

Herrn Prof. Dr. A. Schmidt und **Herrn Dr. A. Gau** für ihre reichhaltigen fachlichen Anregungen und für die Gesprächsbereitschaft.

Frau Dr. N. Hausmann, Herrn Dr. M. Bauer, Frau Dr. A. Bartels,
Frau Dr. A. Riemenschneider, Herrn Dipl. Biol. J. Brock,
Frau Dipl. Biol. A. Büttner, Frau Dr. S. Kürkcüoglu, Frau J. Lensing,
Frau Dipl. Biol. A. Steinfelder, Frau B. Lippmann, Frau H. Feierabend,
Herrn Dr. B. Ramani, Frau Dipl. Biol. E. Bonacina, Frau Dipl. Biol. T. Triulzi und **Herrn J. Hartmann** für die super Zusammenarbeit im Institut für Botanik.

Frau P. von Trzebiatowski, Frau J. Volker, Frau A.-L. Becker und **Frau I. Pillukat** für ihre kompetente technische Unterstützung im Labor und die gute Zusammenarbeit.

Frau M. Klunker, Frau C. Hausmann, Frau Y. Leye und **Herrn L. Krüger** für die freundliche Zusammenarbeit und die kompetente Anzucht der Versuchspflanzen.

Den Kooperationspartnern **Herrn Prof. Dr. J. Gershenzon, Herrn Prof. Dr. J. Tokuhisa, Herrn Dr. M. Reichelt** und **Frau Dr. M. Burow** für ihre Hilfsbereitschaft und Gastfreundschaft.

Meiner Familie und **meinen Freunden** für die moralische Unterstützung.

Meinem Mann Tom für sein Interesse an meiner Arbeit und seiner Unterstützung.

Meiner Mutter für ihr Interesse an meiner Arbeit und Unterstützungen jeglicher Art.

Lebenslauf

Persönliche Daten:

Name:	Marion Klein
Geburtsdatum:	07.06.72
Geburtsort:	Rinteln

Schulausbildung:

08/1979 – 07/1981	Grundschule am Ebersberg in Springe
08/1981 – 07/1983	Albert-Schweitzer-Schule in Hannover
08/1983 – 05/1992	Integrierte Gesamtschule Hannover-Linden Abschluß: Abitur

Berufsausbildung:

02/1993 – 01/1995	Ausbildung zur „staatlich geprüften biologisch-technischen Assistentin“ an der Sabine-Blindow-Schule Hannover

Berufliche Tätigkeit, Institut für Angewandte Genetik der Universität Hannover:

04/1995 – 10/1996	BMFT-Projekt „Entwicklung von Mikrosateliten-Markern bei Beta-Rüben für die Aufstellung einer dichten Kopplungskarte“ (befristete Teilzeitstelle)
07/1996 – 10/1996	DFG-Projekt „Molekulare Charakterisierung des Inkompatibilitätsmechanismus beim Roggen“ (befristete Teilzeitstelle)
11/1996 – 10/1997	DFG-Projekt „Molekulare Charakterisierung des Inkompatibilitätsmechanismus beim Roggen“ (befristete Vollzeitstelle)

Studium:

10/1997 – 01/2003	Fachrichtung: Biologie (Dipl.) an der Gottfried Wilhelm Leibniz Universität Hannover Abschluß: Diplom-Biologin

Promotion:

08/2003 – 12/2006	Promotion im Institut für Botanik der Gottfried Wilhelm Leibniz Universität Hannover