

Characterization of desulfo-glucosinolate sulfotransferases by sequence- and 3D- structure analysis

Von der Naturwissenschaftlichen Fakultät

der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades

Doktor der Naturwissenschaften

Dr. rer. nat.

genehmigte Dissertation

von

Felix Hirschmann, M. Sc.

geboren am 30.12.1986, in Oldenburg (Oldb)

2017

Referentin: Prof. Dr. rer. nat. Jutta Papenbrock

Korreferent: Prof. Dr. rer. nat. Bernhard Huchzermeyer

Tag der Promotion: 11. Oktober 2016

Erklärung kumulative Dissertation

aus:

Gemeinsame Ordnung für die Promotion zur Doktorin der Naturwissenschaften oder zum Doktor der Naturwissenschaften (Dr. rer. nat.) an der Gottfried Wilhelm Leibniz Universität Hannover (25.03.2013)

§ 8 Dissertation

A: (3)

...² Es ist eine ausführliche Darstellung voranzustellen, die eine kritische Einordnung der Forschungsthemen und wichtigsten Erkenntnisse aus den Publikationen in den Kontext der wissenschaftlichen Literatur zum Thema vornimmt ...

Die voranzustellende ausführliche Darstellung ist in dieser Arbeit aufgeteilt in die Kapitel 1 und 5.

B: (3)

...vornimmt sowie die individuellen eigenen Beiträge und ggf. die Beiträge weiterer Autoren an den jeweiligen Publikationen darlegt.

Publikation (Kapitel 2):

Hirschmann, F., Krause, F., Papenbrock, J. (2014). The multi-protein family of sulfotransferases in plants: Composition, occurrence, substrate specificity, and functions. *Frontiers in Plant Science* 5, 556. doi: 10.3389/fpls.2014.00556

- Literaturrecherche zu sämtlichen Themen des Reviews
- Erstellung des Manuskripts für die Veröffentlichung

Publikation (Kapitel 3):

Hirschmann, F. and Papenbrock, J. (2015). The fusion of genomes leads to more options: A comparative investigation on the desulfo-glucosinolate sulfotransferases of *Brassica napus* and homologous proteins of *Arabidopsis thaliana*. *Plant Physiology and Biochemistry* 91, 10-19. doi: 10.1016/j.plaphy.2015.03.009

- Durchführung sämtlicher Experimente
- Erstellung des Manuskripts für die Veröffentlichung

Publikation (Kapitel 4):

Hirschmann, F., Krause, F., Baruch, P., Chizhov, I., Mueller, J.W., Manstein, D.J., Papenbrock, J., Fedorov, R. The structural and biochemical study of sulfotransferase 18 from *Arabidopsis thaliana* explains its substrate specificity and reaction mechanism. *Scientific Reports* (manuscript for publication)

- Erstellung der Expressionskonstrukte
- Expression und Aufreinigung der rekombinanten Proteine
- Durchführung der Kristallisationsstudien
- Analyse der Strukturen
- Erstellung der Homologie Modelle
- Durchführung der Normal-Mode-Analysen
- Durchführung der Mutationsstudien
- Durchführung der Enzym Aktivitätstests
- Erstellung des Manuskripts für die Veröffentlichung

Zusammenfassung

Sulfotransferasen (SOTs) (EC 2.8.2.-) katalysieren den Transfer eines Sulfonatrestes vom Donor 3'-Phosphoadenosin 5'-phosphosulfat (PAPS) zu Hydroxylgruppen verschiedener Stoffe. In *Arabidopsis thaliana* benutzen drei SOTs, AtSOT16, AtSOT17 und AtSOT18, desulfo-(ds) Glucosinolate (Gls) als Substrate. Im Vorfeld wurde gezeigt, dass AtSOT16 indolische ds-Gls präferiert, während AtSOT17 und AtSOT18 aliphatische Gls bevorzugen.

In dieser Arbeit wurde untersucht, ob es möglich ist die Substratspezifitäten von ds-Gls im kürzlich sequenzierten *Brassica napus* Genom vorherzusagen. Insgesamt wurden in *B. napus* 71 vermeintliche SOTs identifiziert und elf davon wurden suggeriert ds-Gl SOTs zu sein. Neben den AtSOT16-18 Homologen, haben phylogenetische Analysen eine neue ds-Gl SOT Subfamilie offenbart, welche nicht in *A. thaliana* präsent ist. *In vitro* Charakterisierungen haben gezeigt, dass drei der fünf rekombinant exprimierten und aufgereinigten BnSOTs ähnliche Substratspezifitäten wie ihre *A. thaliana* Homologe haben. Zwei der getesteten Proteine, welche als SOT18 Homologe identifiziert wurden, zeigten keinerlei Aktivität.

Um ein besseres Verständnis über ds-Gl SOT Katalyse und Substratbindung zu erhalten, wurden hochauflösende Kristallstrukturen von AtSOT18 mit 3'-Phosphoadenosin 5'-phosphat (PAP) alleine und zusammen mit Sinigrin im aktiven Zentrum gelöst. Residuen im aktiven Zentrum, welche essentiell für die Substratbindung und Katalyse sind, wurden identifiziert und ihre individuellen Aufgaben durch Mutationsstudien analysiert. Mit Hilfe der neuen strukturellen Erkenntnisse, kann die Inaktivität der BnSOT18 Homologe, durch natürliche Variationen von katalytischen und PAP-bindenden Residuen, erklärt werden. Des Weiteren wurde die PAP Inhibition durch die Anwendung eines zweidimensionalen nichtlinearen Modells kinetischer Daten untersucht. Insgesamt wurde eine hohe Konservierung gängiger Eigenschaften zwischen Säugtier-SOTs und pflanzlichen SOTs offenbart.

Schlagwörter: *Arabidopsis thaliana*, *Brassica napus*, Glucosinolate, Inhibition, Katalyse, Proteinstruktur, Sequenzanalyse, Substratspezifität, Sulfotransferase

Abstract

Sulfotransferases (SOTs) (EC 2.8.2.-) catalyse the transfer of a sulfate moiety from the donor 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to hydroxyl groups of various compounds. In *Arabidopsis thaliana*, three SOTs, AtSOT16, AtSOT17 and AtSOT18, use desulfo-(ds) glucosinolates (Gls) as substrates. Previously, it was shown that AtSOT16 prefers indolic ds-Gls, while AtSOT17 and AtSOT18 prefer aliphatic ds-Gls.

In this work, it was investigated, if it is possible to predict ds-Gl SOT substrate specificities in the recently sequenced *Brassica napus* genome. In total, 71 putative SOTs were identified in *B. napus* and eleven were suggested to be ds-Gl SOTs. Besides the homologs of AtSOT16-18, phylogenetic analysis revealed a new subfamily of ds-Gl SOTs, which is not present in *A. thaliana*. *In vitro* characterization showed that three of the five recombinantly expressed and purified BnSOTs showed similar substrate specificities as their *A. thaliana* homologs. Two of the tested proteins that were predicted to be SOT18 homologs did not show any activity.

In order to gain a better understanding of ds-Gl SOT catalysis and substrate binding, high-resolution crystal structures of AtSOT18 with 3'-phosphoadenosine 5'-phosphate (PAP) alone and together with sinigrin in the active site were solved. The active site residues essential for substrate binding and catalysis were identified and their individual roles were confirmed by mutational studies. With the help of the new structural insights, the inactivity of the BnSOT18 homologs could be explained, due to natural variations of the catalytic and PAPS-binding residues. Furthermore, PAP inhibition was studied by applying a two-dimensional non-linear model of kinetical data. Overall, a high conservation of common features between mammalian and plant SOTs was revealed.

Keywords: *Arabidopsis thaliana*, *Brassica napus*, catalysis, glucosinolate, inhibition, protein structure, sequence analysis, substrate specificity, sulfotransferase

Contents

CHAPTER 1 – General Introduction	1
The economic relevance of <i>Brassica napus</i> and its large genome	2
Connection between primary and secondary sulfur metabolism.....	4
Glucosinolates.....	5
The glucosinolate biosynthesis can be divided in three stages.....	7
The sulfotransferase enzyme family	9
The broad variety of sulfotransferases and enzymes.....	10
Is it possible to predict ds-GI SOT specificity?.....	11
Aims of this thesis	13
References.....	14
CHAPTER 2 – The multi-protein family of sulfotransferases in plants: composition, occurrence, substrate specificity and functions	21
Abstract.....	22
Introduction.....	22
Primary structure of SOTs, PAPS binding regions, and alignment of the highly conserved regions.....	23
Sulfotransferase families in different plant genomes	25
Substrates for sulfotransferases	25
Biosynthesis of the co-substrate PAPS.....	25
Substrates for plant SOTs	26
Glucosinolates: products of the SOT reaction	27
Sulfotransferases are involved in sulfation of desulfo-glucosinolates.....	28
In <i>A. thaliana</i> ecotypes SOT18 proteins differ in their sequence and substrate specificity	28
Expression of sulfotransferases	29
What is known about the reaction mechanism of sulfotransferases.....	30
How to identify the substrate specificity	30
Chances and restrictions of modelling.....	30
<i>A. thaliana</i> as a model plant – suited for the elucidation of all SOT functions?	31
Future challenges.....	31
Acknowledgments	32
References.....	32

CHAPTER 3 – The fusion of genomes leads to more options: A comparative investigation on the desulfo-glucosinolate sulfotransferases of <i>Brassica napus</i> and homologous proteins of <i>Arabidopsis thaliana</i>	35
Abstract.....	36
Introduction.....	36
Material and methods	37
Plant material	37
DNA cloning.....	37
Expression and purification of AtSOT proteins	37
Preparations of substrates	38
Enzyme activity measurements and HPLC analysis.....	38
Database screen and bioinformatic analysis	38
Results	38
BnSOT identification and amplification.....	38
Amplification of ds-GI BnSOTs	39
The BnSOTs analyzed have similar substrate specificities as their <i>A. thaliana</i> homologs	40
Discussion.....	41
Conclusion	43
Contributions	43
Acknowledgment.....	43
References.....	43
Supplementary data	46
CHAPTER 4 – The structural and biochemical study of sulfotransferase 18 from <i>Arabidopsis thaliana</i> explains its substrate specificity and reaction mechanism	48
Abstract	49
Introduction	50
Results	53
Structure of AtSOT18.....	53
Analysis of substrate binding sites in the AtSOT18 complexes.....	55
Identification of the amino acids involved in the enzymatic reaction	58
Enzyme kinetics and inhibition tests of AtSOT18	60
The source of ds-GI SOT specificity	63
Discussion	65
Biology of SOTs.....	65
Mechanism of SOTs and kinetics	66

Analysis of the catalytic center.....	66
The regulatory effect of PAP inhibition	67
The substrate specificity of ds-GI SOTs.....	68
Methods.....	69
Expression, purification and crystallization	69
Data collection and refinement.....	70
Homology modelling and NMA.....	72
Preparation of substrates.....	72
Enzyme activity measurements and analysis of the kinetic parameters.....	72
Acknowledgements	73
Author contributions	73
References	74
Supplementary Material	79
CHAPTER 5 – General Discussion.....	84
Combination of structure and enzyme assay results.....	85
Are ds-GI SOTs a good target for breeding?	86
The influence of PAPS and PAP concentrations on SOTs.....	88
References.....	91
Abbreviations.....	95
Selbständigkeitserklärung	97
List of Publications.....	98
Acknowledgments.....	99
Curriculum Vitae	100

CHAPTER 1

General Introduction

The economic relevance of *Brassica napus* and its large genome

Brassica napus, called rapeseed or oilseed rape, is the world's third largest source of vegetable oil after palm oil and soybean (United States Department of Agriculture, 2016). Rapeseed is used for animal feeds, edible vegetable oils and biodiesel. The world production has been constantly increasing and reached more than 70 million tons of commodity in the year 2014 (<http://faostat3.fao.org>, accessed 16.03.2016) (Fig. 1). Germany is the world's fourth biggest producer after Canada, China, and India, with more than 6 million tons produced. The increasing importance of rapeseed in Germany becomes clear when comparing the usage of arable land from 1961 to 2014 (Fig. 2). While the used area for rye, potatoes and sugar beet has declined, it has constantly risen for maize and wheat. Interestingly, the agricultural crop land for *B. napus* has been comparably low, until it drastically increased from the mid-1980s to the mid-2000s.

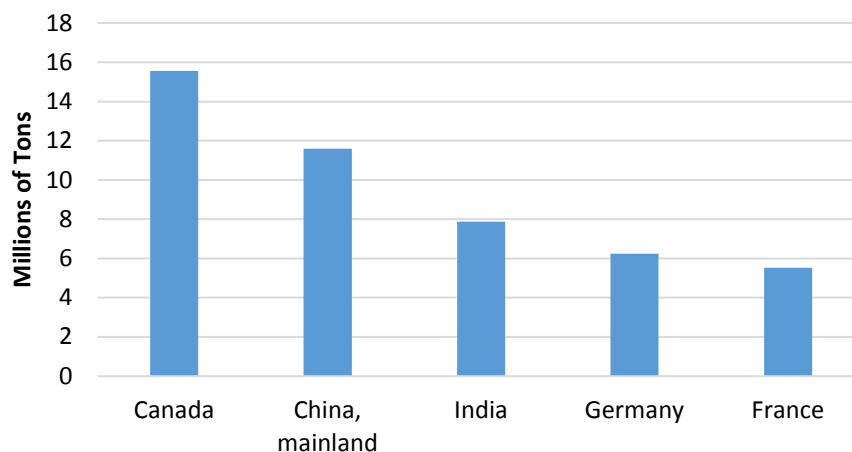


Fig. 1. Rapeseed production of the world's top five producing countries in 2014. Data from the Food and Agriculture Organization of the United Nations, Statistic Division, accessed 16.03.2016.

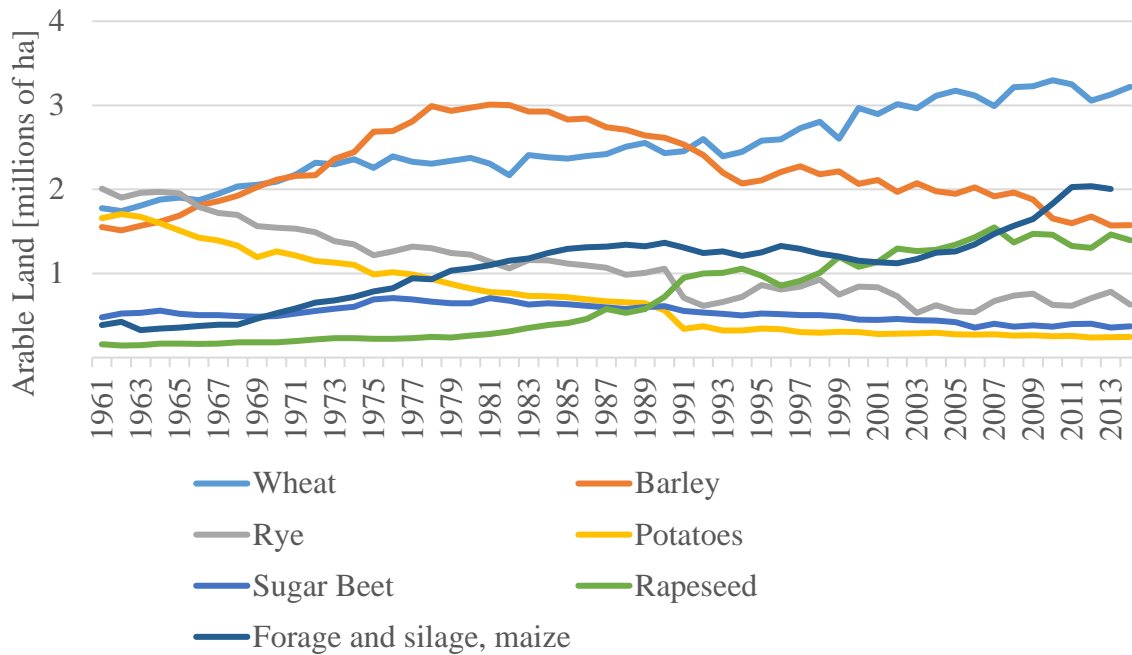


Fig. 2. Arable land used for different crops in Germany from 1961 to 2014. Data from the Food and Agriculture Organization of the United Nations, Statistic Division, accessed 16.03.2016.

The sudden increase in *B. napus* production in the 1980s is the result of the introduction of so called 00 or “double low” cultivars. Before that, *B. napus* contained high concentrations of the antinutritional glucosinolates (Gls) and toxic erucic acid, which made the rapeseed oil and cake less valuable as animal feed and unfeasible for human consumption. The oil of modern *B. napus* cultivars contains less than $30 \mu\text{mol g}^{-1}$ Gls and less than 2% erucic acid and therefore is safe to consume (Canola Council of Canada, 1990).

Another factor for the rising demand of *B. napus*, is the increasing usage of biodiesel. In Europe, it is the main oil source for the production of biodiesel, which emphasizes its importance also in respect to future renewable energy supply. In 2014, 30 billion liters of biodiesel were produced worldwide, with the United States of America being the main producer with 16% of the global total, followed by Brazil and Germany both with 11%. In Europe, the biodiesel production has increased by 9% in 2014 relative to 2013 (REN21, 2015).

In order to cover the future demand of rapeseed, it is necessary to conduct further research in increasing its production, for which a deeper understanding of its biology is fundamental. A major difficulty in *B. napus* research is its big genome. The 1130-Mb genome is the result of an allopolyploidization event of the ancestors *B. oleracea* ($2n = 2 \times 9 = 18$, genome CC) and *B. rapa* ($2n = 2 \times 10 = 20$, genome AA), and has been sequenced in 2014 (Chalhoub et al., 2014). The impact of the polyploidization on the physiology is poorly understood and therefore it is

questionable to what degree findings from related plants, such as *Arabidopsis thaliana*, can be transferred to *B. napus*.

Connection between primary and secondary sulfur metabolism

An important characteristic of *B. napus* is its high sulfur requirement. It needs with 30-50 kg ha⁻¹ about twice as much sulfur as wheat (Bloem et al., 2012). Sulfur plays a crucial role in plant stress resistance, and its deficiency can lead to reduced crop quality and yield (Scherer, 2001). Therefore, it is necessary to have a detailed understanding of the role of sulfur and the sulfur-containing compounds in *B. napus*, especially the secondary sulfur metabolism, where sulfotransferases (SOTs) play a key role in the formation of sulfur containing secondary metabolites.

Sulfur is taken up as sulfate from the soil by sulfate transporters (Fig. 3). Sulfate transporters are membrane proteins, which are divided into four different groups depending on their localization and affinity to sulfate (Takahashi et al., 2011). After the uptake, sulfate is activated by complexation with adenosine triphosphate (ATP) to adenosine phosphosulfate (APS) by ATP sulfurylase. The APS is then either further reduced to sulfide and incorporated in cysteine in the primary sulfur metabolism, or phosphorylated to 3'-phosphoadenosine 5'-phosphosulfate (PAPS) in the secondary sulfur metabolism. The PAPS can then be used as a substrate by SOTs for the sulfation of various compounds (Abuelsoud et al., 2016). Hence, SOTs are the connecting link between the sulfur metabolism and numerous other physiological functions and pathways. Various SOT substrates have been identified (Fig. 3), but many more can be expected. Identification of further substrates would give valuable insights, not only in the sulfur metabolism directly, but also in connected pathways and the functionality of sulfation.

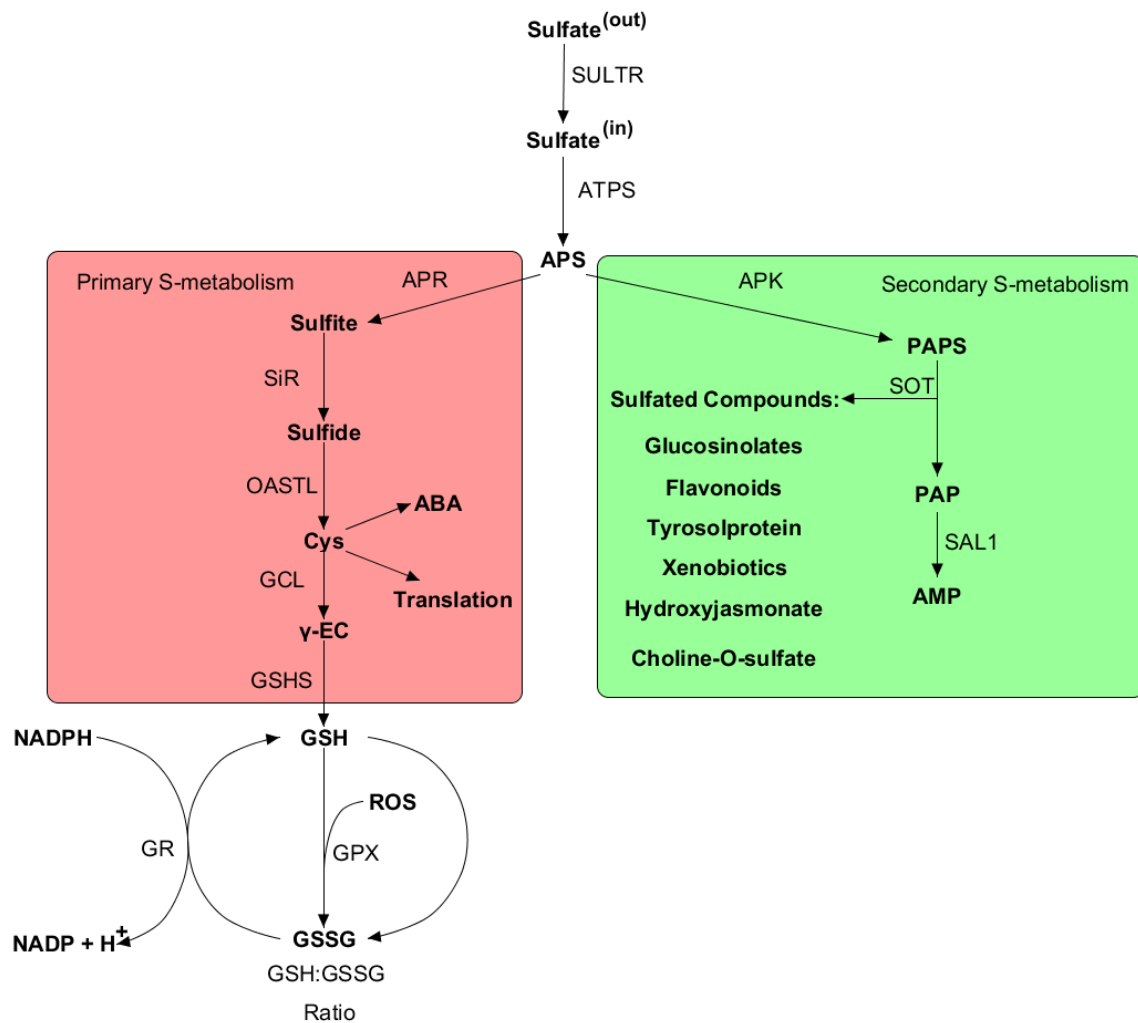


Fig. 3. Schematic overview of plant sulfur metabolism, modified from Abuelsoud et al. (2016). Sulfate is taken up by SULTRs and complexed to ATP to form APS. APS can either be reduced to sulfite as part of the primary sulfur metabolism (red box), or phosphorylated to PAPS as part of the secondary sulfur metabolism (green box). Cosubstrates and products are not displayed. Abbreviations: ATPS, ATP sulfurylase; APR, APS reductase; APK, APS kinase; SiR, sulfite reductase; OASTL, O-acetylserine thiol lyase; Cys, cysteine; ABA; abscisic acid; GCL, Glutamate cysteine lyase; γ -EC, γ -glutamyl cysteine; GSH, reduced glutathione; ROS, reactive oxygen species; GSSG, glutathione disulfide; GSHS, GSH synthetase; GR, glutathione reductase; GPX, glutathione peroxidase; PAP, 5'-phosphoadenosine phosphate; SOT, sulfotransferase, AMP, adenosine monophosphate.

Glucosinolates

Gls (*S*-glucopyranosyl thiohydroxymates) are one of the best studied secondary plant metabolites, mainly found in Brassicaceae, including *A. thaliana* and the economically relevant crops *B. napus*, *B. rapa* and *B. oleracea*. To date, there are almost 200 different Glcs (Fig. 4) described. Glcs consist of a sulfated isothiocyanate group, which is conjugated to thioglucose,

and a further R-group (Clarke, 2010). Depending on their precursor amino acid, Gls can be divided into three groups: aliphatic Gls derive from either alanine, leucine, isoleucine, valine or methionine, while indolic Gls derive from tryptophan and aromatic from either phenylalanine or tyrosine (Ishida et al., 2014).

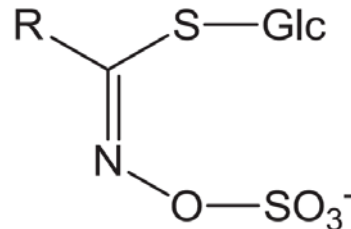


Fig. 4. General GI structure. A thioglucose moiety is linked via a sulfur atom to a (*Z*)-*N*-hydroxyimino sulfate ester with a variable R group, which is derived from a precursor amino acid.

Gls and their breakdown products were suggested to have functions in numerous plant processes, such as sulfur storage, water transport, heat tolerance, stomatal regulation, apoptosis, growth inhibition, and signalling (Bones et al., 2015). However, what is most studied are their roles in plant defence against biotic stress, as part of the GI-myrosinase system. After biosynthesis, Gls are stored in the vacuole, hardly hydrolyzed and biologically inactive. However, after cell disruption, Gls are hydrolyzed by cytoplasm-localized β -thioglucosidases, also called myrosinases (Bones and Rossiter, 2006). The GI breakdown products, mainly isothiocyanates, but also nitriles, epithionitriles and thiocyanates are the biological active compounds (Fig. 6). In numerous studies they were shown to be involved in plant defense against insects, pathogens and herbivores (Manici et al., 1999; Rask et al., 2000; Tierens et al., 2001; Agrawal and Kurashige, 2003; Hopkins et al., 2009).

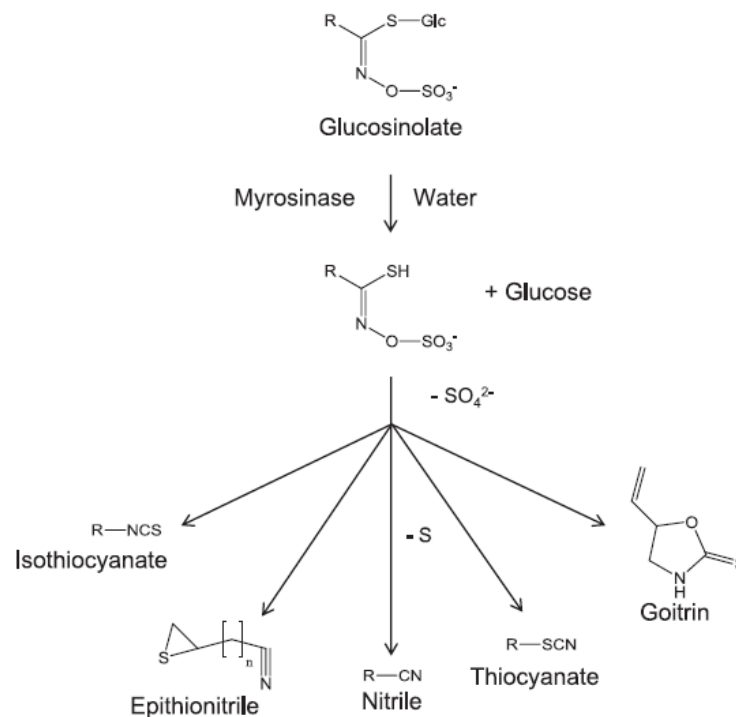


Fig. 5. The GI breakdown products isothiocyanate, epithionitrile, nitrile, thiocyanate, and goitrin (Ishida et al., 2014).

Despite their functions in plant defence and stress response, GIs and their breakdown products have gained much attention due to their cancer chemopreventive properties, especially the isothiocyanate sulforaphane, which is found in high concentrations in broccoli. Sulforaphane is involved in multiple anticarcinogenic mechanisms against several kinds of cancer (Kaufman-Szymczyk et al., 2015). On the other hand, there are antinutritious GIs, such as goitrin, which causes goiter in animals (Stoewsand, 1995). Due to the contrary properties in the wide variety of GIs, the breeding towards enhanced GI profiles of Brassicaceae is of great interest.

The glucosinolate biosynthesis can be divided in three stages

GI biosynthesis, which partly takes place in the cytosol and partly in the chloroplast, can generally be divided into three stages. In the first stage, the precursor amino acid can be elongated by the addition of further methylene groups to the side chain. This only occurs in the biosynthesis of aliphatic or aromatic GIs (Sønderby et al., 2010). For chain elongation, the precursor amino acids are deaminated by branched-chain amino acid aminotransferases (BCAT) (Schuster et al., 2006). The elongation of the resulting 2-oxo acids is then catalyzed by methylthioalkylmalate synthases (Kroymann et al., 2001), isopropylmalate isomerases (Knill et al., 2009) and isopropylmalate dehydrogenases (He et al., 2009). After a cycle of

elongation steps, the 2-oxo acids are turned into the corresponding precursor amino acids by BCAT (Knill et al., 2008).

In the second stage the core Gl is formed. The precursor amino acids are converted to aldoximes by cytochromes P450 and then activated through oxidation by CYP83 (Sønderby et al., 2010). Afterwards, the activated aldoximes are turned into thiohydroximates by conjugation to a sulfur donor, followed by the *C-S* lyase (SUR1) reaction (Mikkelsen et al., 2004). The thiohydroximates are then S-glucosylated by S-glucosyltransferases to form desulfated-(ds) GlS (Grubb et al., 2004). In the final step of core-Gl biosynthesis, the ds-Gls are sulfated by ds-Gl SOTs (Fig. 6).

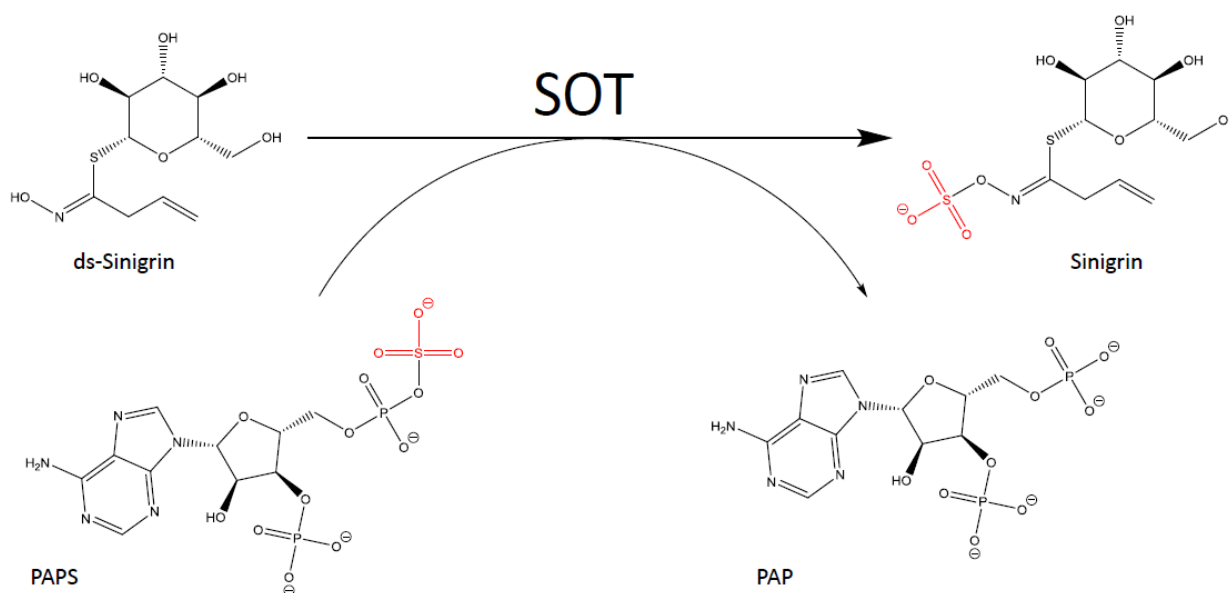


Fig. 6. SOT catalyzed reaction. The red colored sulfate moiety (RSO_3^-) is transferred from PAPS to the hydroxyl group of the example substrate the desulfated Gl sinigrin.

In the third stage of Gl biosynthesis, the side chains of core GlS undergo further modifications, such as oxygenation, hydroxylation, alkenylation, benzoylation, and methoxylation, which results in the great variety of GlS (Ishida et al., 2014).

As a secondary metabolite model, GlS also offer a good opportunity to research the influence of SOT specificity on the outcome of complex secondary metabolite biosynthesis. A better understanding of ds-Gl SOT specificity on a molecular level, could enable its sequence-based prediction and maybe even manipulation, not only for the model plant *A. thaliana*, but also for other crops such as *B. napus*.

The sulfotransferase enzyme family

Eukaryotic SOTs (EC 2.8.2.-) catalyze the transfer of a sulfate group from the universal donor PAPS to hydroxyl or primary amine groups of various compounds (Fig. 4). Only some bacterial PAPS-independent SOTs catalyze the sulfo transfer from phenolic sulfate esters to another phenol. These SOTs neither share other common eukaryotic SOT characteristics regarding sequence, structure and mechanism (Malojčić et al., 2008; Malojčić and Glockshuber, 2010). Therefore, this section will focus on eukaryotic PAPS-dependent SOTs.

SOTs can generally be divided in two groups: soluble SOTs, which accept small organic molecules as substrates, such as xenobiotics, steroids and flavonoids; and membrane associated SOTs, which sulfate proteins, peptides and glycosaminoglycans. The soluble SOTs have four conserved characteristic regions (region I - IV) (Varin et al., 1992), that are mainly involved in PAPS binding (Chapman et al., 2004).

Eukaryotic soluble SOTs share a common spherical α/β fold with four or five central β -sheets surrounded by several α -helices. While the PAPS-binding site is conserved, there are variations in the acceptor binding site, which is gated by two or three flexible loops (Chapman et al., 2004; Tibbs et al., 2015). These loops were recently reported to have a major impact on substrate selectivity (Rohn et al., 2012; Rakers et al., 2016). In mammals, the short dimerization motif KxxxTVxxxE is suggested to be responsible for the formation of SOT homodimers (Petrotchenko et al., 1999), while plant SOTs are expected to be monomers (Klein and Papenbrock, 2004; Smith et al., 2004).

Eukaryotic soluble SOTs are proposed to follow a sequential mechanism, however it is still unclear whether the binding of substrates follows a specific or random order (Chapman et al., 2004; Tibbs et al., 2015). A highly conserved histidine at the beginning of region II and a lysine in region I have fundamental functions in catalysis. It is suggested that the histidine may abstract a proton from the hydroxyl group of the respective substrate, allowing the substrate to attack the PAPS sulfur atom. A shift of the conserved lysine may then complete the sulfate dissociation from the 5'-phosphoadenosine phosphate (PAP) (Fig. 7). Whether the in-line attack of the nucleophile follows a S_N1 or S_N2 mechanism is still a matter of debate. While kinetic isotope effect studies on a human estrogen SOT designated a S_N1 -like mechanism (Hoff et al., 2006), structural analysis of a mouse catecholamine SOT with PAPS and substrate bound indicated a S_N2 -like in-line displacement mechanism (Teramoto et al., 2009).

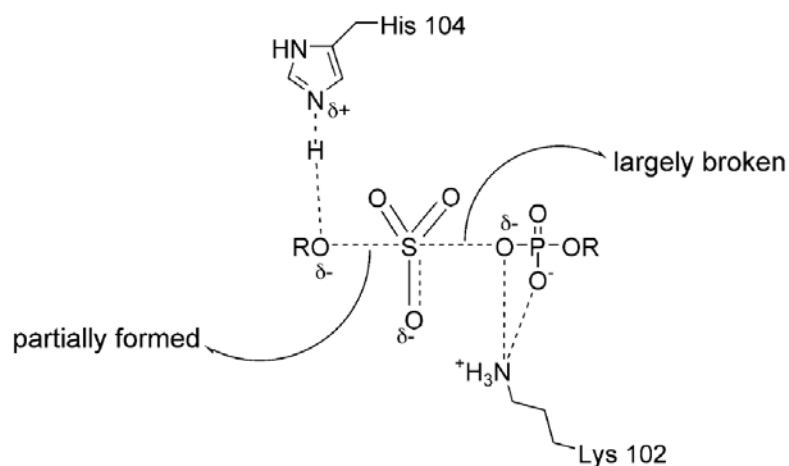


Fig. 7. The highly conserved histidine and lysine in the active site are proposed to play a crucial part in sulfation catalysis (Chapman et al., 2004).

Another SOT characteristic is the inhibition by the co-product PAP. There have been several mechanisms suggested to explain PAP-inhibition, such as allosteric regulation (Zhang et al., 1998), binding of multiple substrates in the active-site cavity (Gamage et al., 2003), dead-end complex formation (Gamage et al., 2005; Sun and Leyh, 2010; Gulcan and Duffel, 2011) and gating (Lu et al., 2008). However, the most detailed study to date on SOT mechanism and PAP-inhibition states that PAP-inhibition is caused by the formation of a dead-end complex, with PAP-release being the rate limiting step (Wang et al., 2014).

The broad variety of sulfotransferases and substrates

The wide range of sulfated compounds also leads to a wide range of SOT functions, such as cell communication, growth and development, and defence (Negishi et al., 2001). Generally, sulfation results in a higher polarity of the substrate, and therefore a better solubility and alteration of its target affinity (Kotov et al., 1999; Cook et al., 2009). The best studied SOT family is from humans, where they are mainly involved in homeostatic control of several hydrophobic signalling molecules, such as estrogens, and detoxification (Tibbs et al., 2015).

In the plant model organism *A. thaliana*, 18 soluble SOTs, including one pseudo-gene, have been identified and were numbered AtSOT1 to AtSOT18 (Klein and Papenbrock, 2004). Based on BLAST results, three more sequences were assumed to be putative SOTs (AtSOT19 – AtSOT21) (Klein and Papenbrock, 2008), but the sequences do not show any other SOT characteristics and are therefore questionable to encode those. Additionally, a Golgi-membrane

localized tyrosylprotein SOT has been identified and characterized, which interestingly does not show sequence similarities to any other SOTs (Komori et al., 2009).

The identified SOTs also cover a wide variety of substrates (Fig. 3), including signalling molecules, such as brassinosteroids (Marsolais et al., 2007) and hydroxyjasmonate (Gidda et al., 2003), but also ds-Gls (Fig. 4) (Piotrowski et al., 2004), flavonoids (Gidda and Varin, 2006) and proteins (Komori et al., 2009). While some AtSOTs are highly specific, others use numerous compounds as substrates, including a variety of xenobiotics, such as bacterial-produced toxin cycloheximide, hinting towards a general detoxification function (Chen et al., 2015).

AtSOT16, AtSOT17 and AtSOT18 were identified as ds-Gl SOTs, catalyzing the last step of Gl core biosynthesis (Fig. 6) (Piotrowski et al., 2004). Detailed analysis of substrate specificities revealed that AtSOT16 highly prefers indolic Gls as substrates, while AtSOT17 and AtSOT18 prefer aliphatic Gls (Klein et al., 2006; Klein and Papenbrock, 2009). It was also shown that mutation of one or two amino acids in AtSOT18 can lead to dramatic changes in activity (Klein et al., 2006) and substrate specificity (Luczak et al., 2013).

It still remains unclear why some SOTs are highly specific, while others accept numerous compounds. It is also not understood how a SOT distinguishes between the large amounts of possible substrates. In order to address questions about substrate specificity, ds-Gl SOTs are particularly interesting, because they only accept ds-Gls as substrates, but distinguish very specifically in between those.

Is it possible to predict ds-Gl SOT specificity?

Generally, it is possible to identify SOTs by their amino acid sequence, either by four conserved regions (regions I – IV), including a highly conserved histidine at the beginning of region II, or by one of the seven Pfam SOT domains (Hirschmann et al., 2014). Several studies have tried to predict substrate preferences by primary sequence analysis, but failed to order SOTs according to their substrate specificities. In *A. thaliana*, the only exception were the three Gl-SOTs, SOT16, SOT17 and SOT18, which were constantly found in a separate branch of phylogenetic trees (Klein and Papenbrock, 2004; Hernández-Sebastiá et al., 2008; Klein and Papenbrock, 2008; Labonne et al., 2009; Hashiguchi et al., 2013). But how reliable are sequence-based predictions of ds-Gl SOTs? Are they strong enough for identification of ds-Gl

SOTs in other plant species, such as *B. napus*? Is it even possible to distinguish between the three ds-Gl SOTs, SOT16, SOT17 and SOT18, just by analysis of the primary structure? These questions shall be addressed by identification and characterization of *B. napus* ds-Gl SOTs.

In order to understand how the ds-Gl SOTs' specificity is encoded in the primary structure, one has to understand how the substrate selectivity is controlled on a molecular level. The determination of three dimensional structures is a powerful tool for these kind of investigations. Previously, the structure of AtSOT12 has been solved, but only incompletely and without any bound ligands (Smith et al., 2004). A structure with bound substrates and/or products could give further insights on what part of the protein is involved in substrate binding, selectivity and catalysis. Furthermore, it could show how conserved plant SOTs are among the SOT enzyme family. Most eukaryotic SOT characteristics were determined by analysis of mammalian or insect SOTs. The structure of AtSOT12 could already demonstrate that plant SOTs also have the common overall structure, but failed to confirm other characteristics, such as the functions of the conserved regions, mechanism, catalytic residues, inhibition and substrate specificity. Therefore, the structure of AtSOT18 with bound product sinigrin and co-product PAP was solved and analysed. Furthermore, catalytic residues were identified, inhibition studies were performed and a molecular mechanism controlling substrate selectivity was suggested.

Aims of this thesis

- Revision of the current state of knowledge of plant SOTs
- Identification of all putative SOTs in the economically relevant crop plant *B. napus*
- Characterization of *B. napus* ds-GI SOTs for confirmation of phylogenetic analysis and test for possible SOT knowledge transfer from *A. thaliana* to other Brassicaceae
- Crystallization and determination of the three-dimensional molecular structure of AtSOT18, in order to address substrate specificity of ds-GI SOTs
- Investigation of AtSOT18 mutants activity for deeper understanding of the catalytic mechanism, PAP-inhibition and structural conservation of plant SOTs

References

- Abuelsoud, W. A., Hirschmann, F., Papenbrock, J. (2016). “Sulfur metabolism and drought stress tolerance in plants“ in *Drought Stress Tolerance in Plants, Vol 1: Physiology and Biochemistry*, ed. M. A. Hossain, S. H. Wani, S. Bhattachajee, D. J. Burritt, L. S. Phan Trans (New York, NY: Springer US)
- Agrawal, A. A., and Kurashige, N. S. (2003). A role for isothiocyanates in plant resistance against the specialist herbivore *Pieris rapae*. *Journal of Chemical Ecology* 29, 1403–1415.
- Bloem, E., Haneklaus, S., Kesselmeier, J., Schnug, E. (2012). Sulfur fertilization and fungal infections affect the exchange of H₂S and COS from agricultural crops. *Journal of Agricultural and Food Chemistry* 60, 7588–7596. doi: 10.1021/jf301912h
- Bones, A. M., Hara, M., Rossiter, J. T., Kissen, R. (2015). Physiology of isothiocyanates and other glucosinolate degradation products in plants. *Frontiers in Plant Science* 6. doi: 10.3389/fpls.2015.01105
- Bones, A. M., and Rossiter, J. T. (2006). The enzymic and chemically induced decomposition of glucosinolates. *Phytochemistry* 67, 1053–1067. doi: 10.1016/j.phytochem.2006.02.024
- Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A. P., Tang, H., Wang, X., et al. (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345, 950–953. doi: 10.1126/science.1253435
- Chapman, E., Best, M. D., Hanson, S. R., Wong, C.-H. (2004). Sulfotransferases: structure, mechanism, biological activity, inhibition, and synthetic utility. *Angewandte Chemie (International ed. in English)* 43, 3526–3548. doi: 10.1002/anie.200300631
- Chen, J., Gao, L., Baek, D., Liu, C., Ruan, Y., Shi, H. (2015). Detoxification function of the *Arabidopsis* sulphotransferase AtSOT12 by sulphonation of xenobiotics. *Plant, Cell & Environment* 38, 1673–1682. doi: 10.1111/pce.12525
- Clarke, D. B. (2010). Glucosinolates, structures and analysis in food. *Analytical Methods* 2, 310. doi: 10.1039/b9ay00280d
- Cook, I. T., Duniec-Dmuchowski, Z., Kocarek, T. A., Runge-Morris, M., Falany, C. N. (2009). 24-hydroxycholesterol sulfation by human cytosolic sulfotransferases: formation of monosulfates and disulfates, molecular modeling, sulfatase sensitivity, and inhibition of liver x receptor activation. *Drug Metabolism and Disposition: the Biological Fate of Chemicals* 37, 2069–2078. doi: 10.1124/dmd.108.025759
- Food and Agricultural Organization of the United Nations, Statistic Division, Online Database. <http://faostat3.fao.org/home/E>

- Gamage, N. U., Duggleby, R. G., Barnett, A. C., Tresillian, M., Latham, C. F., Liyou, N. E., et al. (2003). Structure of a human carcinogen-converting enzyme, SULT1A1. Structural and kinetic implications of substrate inhibition. *The Journal of Biological Chemistry* 278, 7655–7662. doi: 10.1074/jbc.M207246200
- Gamage, N. U., Tsvetanov, S., Duggleby, R. G., McManus, M. E., Martin, J. L. (2005). The structure of human SULT1A1 crystallized with estradiol. An insight into active site plasticity and substrate inhibition with multi-ring substrates. *The Journal of Biological Chemistry* 280, 41482–41486. doi: 10.1074/jbc.M508289200
- Gidda, S. K., Miersch, O., Levitin, A., Schmidt, J., Wasternack, C., Varin, L. (2003). Biochemical and molecular characterization of a hydroxyjasmonate sulfotransferase from *Arabidopsis thaliana*. *The Journal of Biological Chemistry* 278, 17895–17900. doi: 10.1074/jbc.M211943200
- Gidda, S. K., and Varin, L. (2006). Biochemical and molecular characterization of flavonoid 7-sulfotransferase from *Arabidopsis thaliana*. *Plant Physiology and Biochemistry : PPB / Société française de physiologie végétale* 44, 628–636. doi: 10.1016/j.plaphy.2006.10.004
- Grubb, C. D., Zipp, B. J., Ludwig-Müller, J., Masuno, M. N., Molinski, T. F., Abel, S. (2004). *Arabidopsis* glucosyltransferase UGT74B1 functions in glucosinolate biosynthesis and auxin homeostasis. *The Plant Journal for Cell and Molecular Biology* 40, 893–908. doi: 10.1111/j.1365-313X.2004.02261.x
- Gulcan, H. O., and Duffel, M. W. (2011). Substrate inhibition in human hydroxysteroid sulfotransferase SULT2A1: Studies on the formation of catalytically non-productive enzyme complexes. *Archives of Biochemistry and Biophysics* 507, 232–240. doi: 10.1016/j.abb.2010.12.027
- Hashiguchi, T., Sakakibara, Y., Hara, Y., Shimohira, T., Kurogi, K., Akashi, R., et al. (2013). Identification and characterization of a novel kaempferol sulfotransferase from *Arabidopsis thaliana*. *Biochemical and Biophysical Research Communications* 434, 829–835. doi: 10.1016/j.bbrc.2013.04.022
- He, Y., Mawhinney, T. P., Preuss, M. L., Schroeder, A. C., Chen, B., Abraham, L., et al. (2009). A redox-active isopropylmalate dehydrogenase functions in the biosynthesis of glucosinolates and leucine in *Arabidopsis*. *The Plant Journal for Cell and Molecular biology* 60, 679–690. doi: 10.1111/j.1365-313X.2009.03990.x
- Hernández-Sebastiá, C., Varin, L., Marsolais, F. (2008). “Sulfotransferases from Plants, Algae and Phototrophic Bacteria,” in *Sulfur Metabolism in Phototrophic Organisms*, eds. R. Hell, C. Dahl, D. Knaff, and T. Leustek (Dordrecht: Springer Netherlands), 111–130.

- Hirschmann, F., Krause, F., Papenbrock, J. (2014). The multi-protein family of sulfotransferases in plants: Composition, occurrence, substrate specificity, and functions. *Frontiers in Plant Science* 5, 556. doi: 10.3389/fpls.2014.00556
- Hoff, R. H., Czyryca, P. G., Sun, M., Leyh, T. S., Hengge, A. C. (2006). Transition State of the Sulfuryl Transfer Reaction of Estrogen Sulfotransferase. *Journal of Biological Chemistry* 281, 30645–30649. doi: 10.1074/jbc.M604205200
- Hopkins, R. J., van Dam, N. M., van Loon, Joop J A (2009). Role of glucosinolates in insect-plant relationships and multitrophic interactions. *Annual Review of Entomology* 54, 57–83. doi: 10.1146/annurev.ento.54.110807.090623
- Ishida, M., Hara, M., Fukino, N., Kakizaki, T., Morimitsu, Y. (2014). Glucosinolate metabolism, functionality and breeding for the improvement of Brassicaceae vegetables. *Breeding Science* 64, 48–59. doi: 10.1270/jsbbs.64.48
- Kaufman-Szymczyk, A., Majewski, G., Lubecka-Pietruszewska, K., Fabianowska-Majewska, K. (2015). The Role of Sulforaphane in Epigenetic Mechanisms, Including Interdependence between Histone Modification and DNA Methylation. *International Journal of Medical Science* 16, 29732–29743. doi: 10.3390/ijms161226195
- Klein, M., and Papenbrock, J. (2004). The multi-protein family of *Arabidopsis* sulphotransferases and their relatives in other plant species. *Journal of Experimental Botany* 55, 1809–1820. doi: 10.1093/jxb/erh183
- Klein, M., and Papenbrock, J. (2008). “Sulfotransferases and Their Role in Glucosinolate Biosynthesis,” in *Sulfur Assimilation and Abiotic Stress in Plants*, eds. N. A. Khan, S. Singh, and S. Umar (Berlin, Heidelberg: Springer Berlin Heidelberg), 149–166.
- Klein, M., and Papenbrock, J. (2009). Kinetics and substrate specificities of desulfoglucosinolate sulfotransferases in *Arabidopsis thaliana*. *Physiologia Plantarum* 135, 140–149. doi: 10.1111/j.1399-3054.2008.01182.x
- Klein, M., Reichelt, M., Gershenzon, J., Papenbrock, J. (2006). The three desulfoglucosinolate sulfotransferase proteins in *Arabidopsis* have different substrate specificities and are differentially expressed. *The FEBS Journal* 273, 122–136. doi: 10.1111/j.1742-4658.2005.05048.x
- Knill, T., Reichelt, M., Paetz, C., Gershenzon, J., Binder, S. (2009). *Arabidopsis thaliana* encodes a bacterial-type heterodimeric isopropylmalate isomerase involved in both Leu biosynthesis and the Met chain elongation pathway of glucosinolate formation. *Plant Molecular Biology* 71, 227–239. doi: 10.1007/s11103-009-9519-5

- Knill, T., Schuster, J., Reichelt, M., Gershenzon, J., Binder, S. (2008). *Arabidopsis* branched-chain aminotransferase 3 functions in both amino acid and glucosinolate biosynthesis. *Plant Physiology* 146, 1028–1039. doi: 10.1104/pp.107.111609
- Komori, R., Amano, Y., Ogawa-Ohnishi, M., Matsubayashi, Y. (2009). Identification of tyrosylprotein sulfotransferase in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* 106, 15067–15072. doi: 10.1073/pnas.0902801106
- Kotov, A., Falany, J. L., Wang, J., Falany, C. N. (1999). Regulation of estrogen activity by sulfation in human Ishikawa endometrial adenocarcinoma cells. *The Journal of Steroid Biochemistry and Molecular Biology* 68, 137–144. doi: 10.1016/S0960-0760(99)00022-9
- Kroymann, J., Textor, S., Tokuhisa, J. G., Falk, K. L., Bartram, S., Gershenzon, J., et al. (2001). A gene controlling variation in *Arabidopsis* glucosinolate composition is part of the methionine chain elongation pathway. *Plant Physiology* 127, 1077–1088.
- Labonne, J. J. D., Goultiaeva, A., Shore, J. S. (2009). High-resolution mapping of the S-locus in *Turnera* leads to the discovery of three genes tightly associated with the S-alleles. *Molecular Genetics and Genomics* 281, 673–685. doi: 10.1007/s00438-009-0439-5
- Lu, L.-Y., Hsieh, Y.-C., Liu, M.-Y., Lin, Y.-H., Chen, C.-J., Yang, Y.-S. (2008). Identification and characterization of two amino acids critical for the substrate inhibition of human dehydroepiandrosterone sulfotransferase (SULT2A1). *Molecular Pharmacology* 73, 660–668. doi: 10.1124/mol.107.041038
- Luczak, S., Forlani, F., Papenbrock, J. (2013). Desulfo-glucosinolate sulfotransferases isolated from several *Arabidopsis thaliana* ecotypes differ in their sequence and enzyme kinetics. *Plant Physiology and Biochemistry* 63, 15–23. doi: 10.1016/j.plaphy.2012.11.005
- Malojčić, G., and Glockshuber, R. (2010). The PAPS-independent aryl sulfotransferase and the alternative disulfide bond formation system in pathogenic bacteria. *Antioxidants & Redox Signaling* 13, 1247–1259. doi: 10.1089/ars.2010.3119
- Malojčić, G., Owen, R. L., Grimshaw, J. P. A., Brozzo, M. S., Dreher-Teo, H., Glockshuber, R. (2008). A structural and biochemical basis for PAPS-independent sulfuryl transfer by aryl sulfotransferase from uropathogenic *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 105, 19217–19222. doi: 10.1073/pnas.0806997105
- Manici, L. M., Leoni, O., Lazzeri, L., Galletti, S., Palmieri, S. (1999). Fungitoxic activity of four thio-functionalised glucosinolate enzyme-derived products on ten soil-borne pathogens. *Pesticide Science* 55, 486–488.

- Marsolais, F., Boyd, J., Paredes, Y., Schinas, A.-M., Garcia, M., Elzein, S., et al. (2007). Molecular and biochemical characterization of two brassinosteroid sulfotransferases from *Arabidopsis*, AtST4a (At2g14920) and AtST1 (At2g03760). *Planta* 225, 1233–1244. doi: 10.1007/s00425-006-0413-y
- Mikkelsen, M. D., Naur, P., Halkier, B. A. (2004). *Arabidopsis* mutants in the C-S lyase of glucosinolate biosynthesis establish a critical role for indole-3-acetaldoxime in auxin homeostasis. *The Plant Journal for Cell and Molecular Biology* 37, 770–777.
- Negishi, M., Pedersen, L. G., Petrotchenko, E., Shevtsov, S., Gorokhov, A., Kakuta, Y., et al. (2001). Structure and Function of Sulfotransferases. *Archives of Biochemistry and Biophysics* 390, 149–157. doi: 10.1006/abbi.2001.2368
- Petrotchenko, E. V., Doerflein, M. E., Kakuta, Y., Pedersen, L. C., Negishi, M. (1999). Substrate gating confers steroid specificity to estrogen sulfotransferase. *The Journal of Biological Chemistry* 274, 30019–30022.
- Piotrowski, M., Schemenewitz, A., Lopukhina, A., Müller, A., Janowitz, T., Weiler, E. W., et al. (2004). Desulfoglucosinolate sulfotransferases from *Arabidopsis thaliana* catalyze the final step in the biosynthesis of the glucosinolate core structure. *The Journal of Biological Chemistry* 279, 50717–50725. doi: 10.1074/jbc.M407681200
- Rakers, C., Schumacher, F., Meinl, W., Glatt, H., Kleuser, B., Wolber, G. (2016). *In Silico* Prediction of Human Sulfotransferase 1E1 Activity Guided by Pharmacophores from Molecular Dynamics Simulations. *Journal of Biological Chemistry* 291, 58–71. doi: 10.1074/jbc.M115.685610
- Rask, L., Andréasson, E., Ekbom, B., Eriksson, S., Pontoppidan, B., Meijer, J. (2000). Myrosinase: gene family evolution and herbivore defense in Brassicaceae. *Plant Molecular Biology* 42, 93–113.
- Rohn, K. J., Cook, I. T., Leyh, T. S., Kadlubar, S. A., Falany, C. N. (2012). Potent Inhibition of Human Sulfotransferase 1A1 by 17 β -Ethinylestradiol: Role of 3'-Phosphoadenosine 5'-Phosphosulfate Binding and Structural Rearrangements in Regulating Inhibition and Activity. *Drug Metabolism and Disposition* 40, 1588–1595. doi: 10.1124/dmd.112.045583
- Scherer, H. W. (2001). Sulphur in crop production — invited paper. *European Journal of Agronomy* 14, 81–111. doi: 10.1016/S1161-0301(00)00082-4
- Schuster, J., Knill, T., Reichelt, M., Gershenzon, J., Binder, S. (2006). Branched-chain aminotransferase4 is part of the chain elongation pathway in the biosynthesis of methionine-derived glucosinolates in *Arabidopsis*. *The Plant Cell* 18, 2664–2679. doi: 10.1105/tpc.105.039339

- REN21 (2015). The Renewables Global Status Report, Paris, REN21 Secretariat
- Smith, D. W., Johnson, K. A., Bingman, C. A., Aceti, D. J., Blommel, P. G., Wrobel, R. L., et al. (2004). Crystal structure of At2g03760, a putative steroid sulfotransferase from *Arabidopsis thaliana*. *Proteins* 57, 854–857. doi: 10.1002/prot.20258
- Sønderby, I. E., Geu-Flores, F., Halkier, B. A. (2010). Biosynthesis of glucosinolates – gene discovery and beyond. *Trends in Plant Science* 15, 283–290. doi: 10.1016/j.tplants.2010.02.005
- Stoewsand, G. S. (1995). Bioactive organosulfur phytochemicals in *Brassica oleracea* vegetables--a review. *Food and Chemical Toxicology* 33, 537–543.
- Sun, M., and Leyh, T. S. (2010). The human estrogen sulfotransferase: a half-site reactive enzyme. *Biochemistry* 49, 4779–4785. doi: 10.1021/bi902190r
- Takahashi, H., Buchner, P., Yoshimoto, N., Hawkesford, M. J., Shiu, S.-H. (2011). Evolutionary relationships and functional diversity of plant sulfate transporters. *Frontiers in Plant Science* 2, 119. doi: 10.3389/fpls.2011.00119
- Teramoto, T., Sakakibara, Y., Liu, M.-C., Suiko, M., Kimura, M., Kakuta, Y. (2009). Snapshot of a Michaelis complex in a sulfuryl transfer reaction: Crystal structure of a mouse sulfotransferase, mSULT1D1, complexed with donor substrate and acceptor substrate. *Biochemical and Biophysical Research Communications* 383, 83–87. doi: 10.1016/j.bbrc.2009.03.146
- Tibbs, Z. E., Rohn-Glowacki, K. J., Crittenden, F., Guidry, A. L., Falany, C. N. (2015). Structural plasticity in the human cytosolic sulfotransferase dimer and its role in substrate selectivity and catalysis. *Drug Metabolism and Pharmacokinetics* 30, 3–20. doi: 10.1016/j.dmpk.2014.10.004
- Tierens, K. F., Thomma, B. P., Brouwer, M., Schmidt, J., Kistner, K., Porzel, A., et al. (2001). Study of the role of antimicrobial glucosinolate-derived isothiocyanates in resistance of *Arabidopsis* to microbial pathogens. *Plant Physiology* 125, 1688–1699.
- United States Department of Agriculture (2016). Oilseeds: World Markets and Trade <http://apps.fas.usda.gov/psdonline/circulars/oilseeds.pdf>.
- Varin, L., deLuca, V., Ibrahim, R.K., Brisson, N. (1992). Molecular characterization of two plant flavonol sulfotransferases. *Proceedings of the National Academy of Sciences of the United States of America* 89, 1286–1290.
- Wang, T., Cook, I., Falany, C. N., Leyh, T. S. (2014). Paradigms of sulfotransferase catalysis: the mechanism of SULT2A1. *The Journal of Biological Chemistry* 289, 26474–26480. doi: 10.1074/jbc.M114.573501

Zhang, H., Varmalova, O., Vargas, F. M., Falany, C. N., Leyh, T. S. (1998). Sulfuryl Transfer: The Catalytic Mechanism of Human Estrogen Sulfotransferase. *Journal of Biological Chemistry* 273, 10888–10892. doi: 10.1074/jbc.273.18.10888

CHAPTER 2

The multi-protein family of sulfotransferases in plants: composition, occurrence, substrate specificity, and functions

Hirschmann, F., Krause, F., Papenbrock, J. (2014). The multi-protein family of sulfotransferases in plants: Composition, occurrence, substrate specificity, and functions. *Frontiers in Plant Science* 5, 556. doi: 10.3389/fpls.2014.00556



The multi-protein family of sulfotransferases in plants: composition, occurrence, substrate specificity, and functions

Felix Hirschmann, Florian Krause and Jutta Papenbrock *

Institute of Botany, Leibniz University Hannover, Hannover, Germany

Edited by:

Stanislav Kopriva, University of Cologne, Germany

Reviewed by:

Masami Yokota Hirai, RIKEN Plant Science Center, Japan
Tamara Gigolashvili, University of Cologne – Biocenter, Germany
Frédéric Marsolais, Agriculture and Agri-Food Canada, Canada

*Correspondence:

Jutta Papenbrock, Institute of Botany, Leibniz University Hannover, Herrenhäuser Straße 2, D-30419 Hannover, Germany
e-mail: jutta.papenbrock@botanik.uni-hannover.de

All members of the sulfotransferase (SOT, EC 2.8.2.-) protein family transfer a sulfuryl group from the donor 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to an appropriate hydroxyl group of several classes of substrates. The primary structure of these enzymes is characterized by a histidine residue in the active site, defined PAPS binding sites and a longer SOT domain. Proteins with this SOT domain occur in all organisms from all three domains, usually as a multi-protein family. *Arabidopsis thaliana* SOTs, the best characterized SOT multi-protein family, contains 21 members. The substrates for several plant enzymes have already been identified, such as glucosinolates, brassinosteroids, jasmonates, flavonoids, and salicylic acid. Much information has been gathered on desulfo-glucosinolate (dsGI) SOTs in *A. thaliana*. The three cytosolic dsGI SOTs show slightly different expression patterns. The recombinant proteins reveal differences in their affinity to indolic and aliphatic dsGIs. Also the respective recombinant dsGI SOTs from different *A. thaliana* ecotypes differ in their kinetic properties. However, determinants of substrate specificity and the exact reaction mechanism still need to be clarified. Probably, the three-dimensional structures of more plant proteins need to be solved to analyze the mode of action and the responsible amino acids for substrate binding. In addition to *A. thaliana*, more plant species from several families need to be investigated to fully elucidate the diversity of sulfated molecules and the way of biosynthesis catalyzed by SOT enzymes.

Keywords: *Arabidopsis thaliana*, glucosinolate, histidine residue, phosphoadenosine 5'-phosphosulfate, sulfotransferase

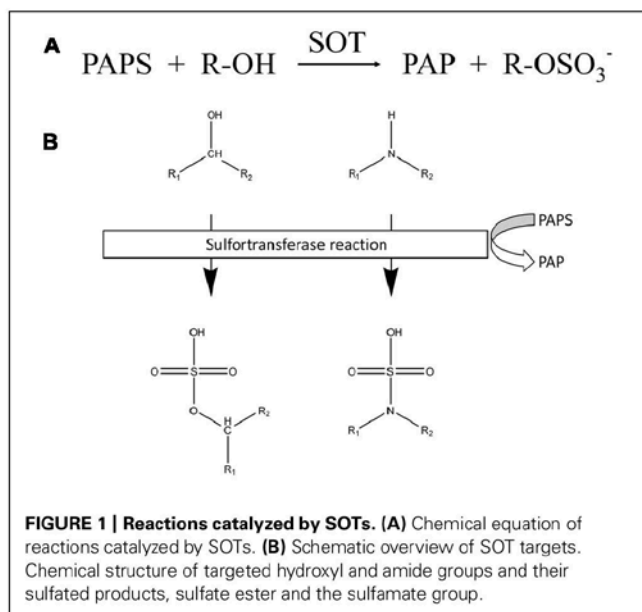
INTRODUCTION

Members of the sulfotransferase (SOT) family have been found in all organisms investigated to date. All of these enzymes catalyze the transfer of a sulfuryl group from 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to an appropriate hydroxyl group (Figure 1), hydroxyl amine or unprotonated amine of various substrates with the parallel formation of PAP.

The SOTs catalyze the sulfation of a wide range of compounds and produce sulfate esters, sulfamates, and sulfate conjugates (Klaassen and Boles, 1997). A sulfate conjugate is more water soluble than a non-sulfated molecule (Weinshilboum and Otterness, 1994), thus facilitating excretion and bioactivation.

Due to the unifying use of the co-substrate PAPS, all SOT proteins are characterized by a histidine residue in the active site, defined PAPS binding sites and a defined SOT domain (Pfam: PF00685; Finn et al., 2014). Proteins with this SOT domain occur in all organisms from all three domains investigated so far, usually as a multi-protein family. Originally, the SOT proteins in mammals 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 glycosaminoglycans (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. In plants, the best criteria for forming subgroups within the multi-protein family is still a matter of debate, because either sequence identity/similarity or their substrate specificity could be chosen. Several compounds have been found in different plant species, such as: brassinosteroids, coumarins, flavonoids, gibberellic acids, glucosinolates (GIs), phenolic acids, sulfate esters such as choline-*O*-sulfate, and terpenoids that might be sulfated by SOT proteins. However, only for some of these substrates has the catalyzing SOT protein been identified. Not all sulfated compounds are necessarily sulfated by SOTs. Sulfolipids contain a 6-deoxy-6-sulfolipid sugar head group, referred to as sulfoquinovose. The sulfoquinovose precursor UDP-sulfoquinovose is biosynthesized from UDP-glucose by a UDP-sulfoquinovose synthase associated with a ferredoxin-dependent glutamate synthase using sulfite as cosubstrate (Shimajima et al., 2005). Much information has been gathered on desulfo-glucosinolate (dsGI) SOTs in *Arabidopsis*, differing in their affinity to indolic and aliphatic dsGIs. However, determinants of substrate specificity and the exact reaction mechanism still need to be clarified. Probably, the three-dimensional structures of more plant proteins have to be solved to analyze the mode of action and the responsible amino acids for substrate binding.



PRIMARY STRUCTURE OF SOTS, PAPS BINDING REGIONS, AND ALIGNMENT OF THE HIGHLY CONSERVED REGIONS

Generally, SOTs can be divided into membrane-bound proteins and soluble cytosolic proteins. So far, only a few membrane-bound SOTs have been characterized in plants. They are either bound to the plasma membrane, as shown for the gallic acid glucoside SOT from *Mimosa pudica* L. (Varin et al., 1997a), or localized in the Golgi apparatus, as shown for the tyrosylprotein SOTs (TPSTs) from *Asparagus officinalis* L. (Hanai et al., 2000) and *Arabidopsis thaliana* (L.) Heynh. (Komori et al., 2009). The term cytosolic SOT might indicate a localization in the cytoplasm, yet the name implies that the proteins can be purified from plant cells and kept in solution (Hernández-Sebastiá et al., 2008). The exact localization of most cytosolic plant SOTs still remain unknown.

Sequence alignments of eleven cytosolic SOTs from plants, animals, and bacteria resulted in the identification of four highly conserved regions I to IV (Marsolais and Varin, 1995; Figure 2). Further analyses showed that especially the regions I and IV are highly conserved, for example throughout the SOT family of *A. thaliana* (Klein and Papenbrock, 2004). The regions I, II, and IV are responsible for the binding of the co-substrate PAPS (Varin et al., 1997b). The first structural approach to clarify the relevance of the regions for PAPS binding was determined by X-ray crystallography analyses of a mouse estrogen SOT (Kakuta et al., 1997). Region I is localized close to the N-terminus and includes the PAPS binding domain (PSB domain) that interacts with the 5'-phosphate of PAPS. Region II starts with a characteristic highly conserved histidine, responsible for proton acceptance during the sulfur transfer (Kakuta et al., 1998). In the C-terminal part of region II the two amino acids Arg130 and Ser138 are responsible for the binding of the 3'-phosphate of PAP and form a 3' P-motif (Kakuta et al., 1997). This motif can be found in 18 SOTs out of 22 from *A. thaliana* and from almost all other plant SOTs. In many plant SOTs, a conserved hydrophilic

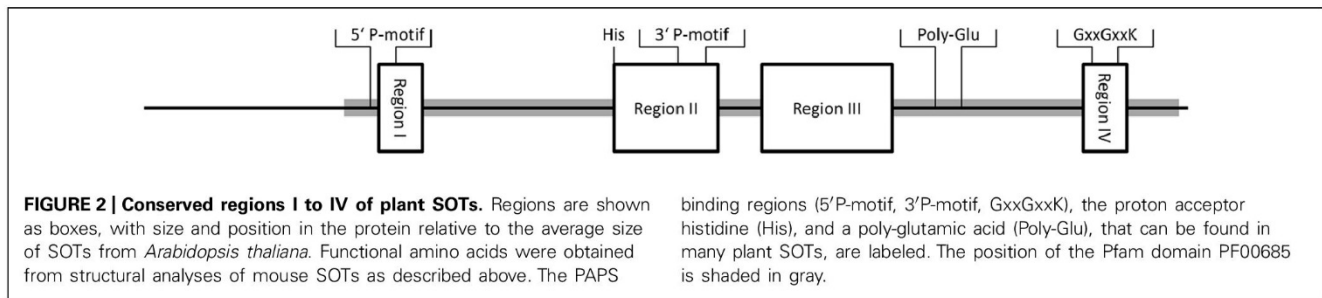
site containing poly-glutamic acid (Poly-Glu) of unknown function can be found between region III and IV. Similar motifs have been found in a human chondroitin 6-SOT, but at a different position (Fukuta et al., 1998). Region IV is localized at the C-terminus and contains a P-loop related GxxGxxK motif (Figure 2).

In *A. thaliana*, 18 protein sequences with high similarity to known SOTs have been originally identified by BLAST approaches (Klein and Papenbrock, 2004). Later, another three SOTs were added (Klein and Papenbrock, 2008). These were formerly annotated in NCBI as “nodulation-related protein” and are now annotated as “P-loop containing triphosphate hydrolase family protein.” In addition to these SOTs, a TPST has been identified (Komori et al., 2009). Furthermore, a not yet literarily mentioned protein, Q9SCR3, with a *Sulfotransfer_1* domain (PF00685), is available in the Pfam database [<http://pfam.xfam.org/protein/Q9SCR3> (accessed 23.06.2014)]. About 75% of the amino acid sequence is identical to *A. thaliana* SOT (AtSOT19). Therefore, it might be a redundant entry or a product of a different splicing process. Interestingly, AtTPST is exceptional in its structure compared to the remaining AtSOTs. With 500 amino acids, it is not only bigger, but is also the solely identified transmembrane, *cis*-Golgi localized AtSOT. Furthermore, it shows no sequence similarity to human TPSTs; and no other typical features like the regions I to IV and the characteristic highly conserved histidine were identified (Komori et al., 2009). It is also the only *Arabidopsis* SOT that contains a *Sulfotransfer_2* domain (PF03567), instead of a *Sulfotransfer_1* domain (PF00685). Hence, it is only associated by function and not by sequence. Excluding the pseudogenic sequence AtSOT2 and TPST, the SOT protein lengths range between 273 and 403 amino acids with an average length of 321 amino acids. Only seven out of 21 AtSOTs contain introns.

There are several nomenclatures for *A. thaliana* SOTs used in the literature. The most common ones are listed in Table 1, including information about the preferred SOT substrates. In this review, the nomenclature first introduced by Klein and Papenbrock (2004) is used.

So far, only one SOT from plants (AtSOT12 from *A. thaliana*) was structurally solved (Smith et al., 2004). Therefore, most SOT proteins lack structural analyses and detailed enzymological characterizations. The identified motifs only give a hint on the proteins' general function as a SOT, but no information about their specificity and affinity toward certain substrates.

Most proteins identified as putative SOTs contain at least one out of seven related Pfam motifs that are based on Hidden Markov Models (HMM). The most important HMMs referring to SOTs are the SOT domains *Sulfotransfer_1* (PF00685), *Sulfotransfer_2* (PF03567), and *Sulfotransfer_3* (PF13469), which have an average length of 230.1, 218.3, and 224 amino acids, respectively. According to the model information in the Pfam database, the SOT domain 1 shows an average coverage of its contributing protein sequences of 64%. An average of 16% of all amino acid residues that are covered by the HMM are identical to it. The average coverage of the SOT domains 2 and 3 in their respective sequences are 67 and 47%, with average sequence identities of 15 and 14%. In addition, HMMs for more specific SOT subfamilies have been

**Table 1 | Summary of the members of the SOT family in *Arabidopsis* and their putative substrates.**

NCBI accession	<i>Arabidopsis</i> gene ID	Nomenclature			Amino acids	Preferred substrate	Reference
		I	II	III			
NP_199182	AT5G43690	AtSOT1		AtSULT202B4	331		
NP_190689	AT3G51210	AtSOT2		Pseudogene	67		
NP_194358	AT4G26280	AtSOT3		AtSULT202C1	314		
NP_180325	AT2G27570	AtSOT4		AtSULT202B3	273		
NP_190093	AT3G45070	AtSOT5	AtST3a	AtSULT202B1	323	Flavonol	Gidda and Varin (2006), Hashiguchi et al. (2013)
NP_190094	AT3G45080	AtSOT6	AtST3b	AtSULT202B2	329		
NP_174139	AT1G28170	AtSOT7		AtSULT202B8	326		
NP_172799	AT1G13420	AtSOT8	AtST4b	AtSULT202B7	331	Flavonol glycosides	Hashiguchi et al. (2014)
NP_172800	AT1G13430	AtSOT9	AtST4c	AtSULT202B5	351		
NP_179098	AT2G14920	AtSOT10	AtST4a	AtSULT202B6	333	Brassinosteroids	Marsolais et al. (2007)
NP_565305	AT2G03750	AtSOT11		AtSULT202D1	351		
NP_178471	AT2G03760	AtSOT12	AtST1	AtSULT202A1	326	Flavonone, brassinosteroids, salicylic acid	Lacomme and Roby (1996), Marsolais et al. (2007), Baek et al. (2010), Hashiguchi et al. (2013)
NP_178472	AT2G03770	AtSOT13		AtSULT202E1	324	Flavonol	Hashiguchi et al. (2013)
NP_196317	AT5G07000	AtSOT14	AtST2b	AtSULT203A2	347		
NP_568177	AT5G07010	AtSOT15	AtST2a	AtSULT203A1	359	Hydroxyjasmonate	Gidda et al. (2003)
NP_177550	AT1G74100	AtSOT16	AtST5a	AtSULT201B3	338	Phenylalanine and tryptophan derived dsGIs	Piotrowski et al. (2004), Klein et al. (2006)
NP_173294	AT1G18590	AtSOT17	AtST5c	AtSULT201B2	346	Benzyl and methionine derived dsGIs	Piotrowski et al. (2004), Klein et al. (2006)
NP_177549	AT1G74090	AtSOT18	AtSTb	AtSULT201B1	350	Phenylalanine and methionine derived dsGIs	Piotrowski et al. (2004), Klein et al. (2006), Luczak et al. (2013)
NP_190631	AT3G50620	AtSOT19			340		
NP_179175	AT2G15730	AtSOT20			344		
NP_195168	AT4G34420	AtSOT21			403		
NP_563804	AT1G08030	TPST			500	Tyrosylprotein	Komori et al. (2009)

I: Nomenclature used in this review introduced by Klein and Papenbrock (2004). II: Nomenclature introduced by Piotrowski et al. (2004). III: Nomenclature introduced by Hashiguchi et al. (2013). The term TPST was introduced by Komori et al. (2009).

deposited in Pfam. Generally, they show a lower number of hits in the database, with mostly increased sequence coverage and a higher average sequence identity as compared to the more general SOT domains 1–3. The aryl SOT domains *Arylsulfotrans_1* and *Arylsulfotrans_2* (PF05935 and PF14269) and the Stf0 SOT domain *Sulphotransf* (PF09037) represent groups of sequences with more specific occurrence, especially in prokaryota. The two aryl SOT domains show an average coverage of 83 and 57%, with an identity of 30 and 28%, respectively. For the Stf0 SOT domain the average sequence coverage is 81%, with an average identity of 33%. The galactose-3-O-SOT domain (PF06990) shows coverage of 78% with 24% identity.

SULFOTRANSFERASE FAMILIES IN DIFFERENT PLANT GENOMES

Sulfotransferases have a broad range of substrates and therefore many functions. In previous literature it has been stated that SOTs are present in all kingdoms except in Archaea (Klein and Papenbrock, 2008; Chen et al., 2012). Nevertheless, according to the protein family database Pfam [http://pfam.xfam.org/ (accessed 23.06.2014)] there are Archaea sequences with a characteristic conserved SOT domain.

There are only few studies that aim to identify all SOTs of a plant species. A requirement to do this is a fully sequenced genome, but due to their eclectic functions, we assume that SOTs are present in almost every plant species. The Pfam database already stores 538 putative plant SOT sequences, 459 of which have a *Sulfotransfer_1* domain (PF00685), 49 a *Sulfotransfer_2* domain (PF03567), 16 a *Sulfotransfer_3* domain (PF13469), and 10 a *Sulphotransf* domain (PF09037). The actual number might be less, because of redundant entries. The *Sulfotransfer_3* (PF13469) and the *Gal-3-O_sulfotr* domain (PF06990) are only present in algae. The *Arylsulfotrans* domain (PF05935) is not present in plants, while *Arylsulfotran_2* (PF14269) is only found in a single *Ricinus communis* L. sequence.

While so far 22 putative *A. thaliana* SOTs were identified, 35 genes coding proteins with a SOT domain were reported in *Oryza sativa* L., including six genes likely to be pseudogenes (Chen et al., 2012). In phylogenetic analyses, they are clustered into seven subfamilies. However, microarray data revealed that the genes within subfamilies are expressed in a different manner, indicating individual functions. When 17 AtSOTs were added to the distance trees, they did not group together with any of the *O. sativa* genes. This was taken as a hint for independent evolution of *O. sativa* and *A. thaliana* SOTs by gene duplication or loss. This was supported by the finding that half of the *O. sativa* SOTs contain introns, which is hardly the case for AtSOTs (Klein and Papenbrock, 2004).

Comparative genomics studies were conducted in *Brassica rapa* L. with *A. thaliana*, in order to identify all Gl biosynthesis genes (Zang et al., 2009; Wang et al., 2011). Thirteen putative *desulfo-glucosinolate* SOT (*dsGl* SOT) genes were identified. Two genes are paralogs of AtSOT16, 1 of AtSOT17, and 10 of AtSOT18. One AtSOT18 paralog appears to be nonfunctional, because of transposon insertion, and one carries a frame shift. None of the genes contains introns, as it is the case for AtSOT genes. All paralogs share at least 70% sequence identity with their AtSOT counterparts, with the exception of one SOT from *B. rapa* (*BrSOT18*, 68%; Wang et al., 2011). The higher number of BrSOTs is explained

by the triplication of the *B. rapa* genome and later duplication, transposition, or tandem duplication of the genes.

In *Brassica napus* L., so far only twelve putative genes encoding SOTs were identified (Rouleau et al., 1999; Marsolais et al., 2000). Additionally, there are at least five isoforms of dsGl SOTs in *B. napus*, which have similar substrate affinities as their *A. thaliana* homologs (own unpublished results). Regarding that *B. napus* is an allotetraploid species formed by the hybridization of *B. rapa* and *B. oleracea*, a much higher number of SOT genes can be expected.

To group these diverse enzymes into families and subfamilies remains a difficult task. Klein and Papenbrock (2004, 2008) ordered 21 *A. thaliana* SOTs in eight groups, according to their amino acid sequence identity. However, already characterized SOTs with the same substrate specificity did not group together, and even high sequence identity of more than 85% among two SOTs did not reveal equal enzymological characteristics. Neither sequence identity, nor generated trees ordered already characterized SOTs in groups according to their substrate specificities. Three dsGl SOTs were on one separate branch, but flavonoid and brassinosteroid SOTs could not be distinguished.

Hernández-Sebastiá et al. (2008) generated a phylogenetic tree including 78 SOTs from 13 different plant species. This approach faced the same problems as the one by Klein and Papenbrock (2004, 2008) and it was again concluded that the prediction of SOT substrates by high primary sequence identities is limited. For example, it was speculated that AtSOT13 was a brassinosteroid SOT, because of its close distance to AtSOT12, but Hashiguchi et al. (2013) showed that AtSOT13 uses flavonoids as preferred substrates.

Another attempt included, besides 17 *A. thaliana*, also *B. napus* and *Flaveria* spp. sequences (Hashiguchi et al., 2013). AtSOT2 was excluded, because it is most likely a pseudogene. According to a dendrogram, three families were defined with two, three and five subfamilies, respectively. The families had an amino acid sequence identity of at least 45% and the subfamilies of at least 60%. But again, except for the dsGl SOTs, the SOTs did not group together according to their substrate specificities.

Labonne et al. (2009) tried to identify a putative SOT of *Turnera krapovickasii* Arbo (Passifloraceae) by phylogenetic analysis. The sequence was aligned with 28 SOTs from *A. thaliana*, *B. napus*, *Vitis vinifera* L., *O. sativa*, *Hordeum vulgare* L., *Populus trichocarpa* Hook. The SOT from *T. krapovickasii* was on a branch by itself and alignments with characterized SOTs revealed low sequence identity. Therefore, it was not possible to identify the function of the respective SOT.

Overall, past attempts indicate that it is difficult to order plant SOTs according to their amino acid sequence. Only for dsGl SOTs does it seem to be possible, because they are clustered together on a separate branch in all approaches. Therefore, only enzymatic assays with additional mutational studies can give reliable information about substrate specificity and function.

SUBSTRATES FOR SULFOTRANSFERASES

BIOSYNTHESIS OF THE CO-SUBSTRATE PAPS

3'-phosphoadenosine 5'-phosphosulfate is an obligate co-substrate for sulfation reactions catalyzed by SOTs. In plants, PAPS does not represent an intermediate of reductive sulfate assimilation

as in fungi and some bacteria, but it seems to play an exclusive role as a sulfuryl donor for sulfation reactions. PAPS is synthesized from ATP and sulfate in a two-step reaction (Figure 3). In the first step, ATP sulfurylase (EC 2.7.7.4) catalyzes sulfate activation. The enzyme hydrolyses the bond between the β - and the γ -phosphates of ATP and then adds sulfate to the γ -phosphate. The activation step is necessary, because sulfate is metabolically inert. The energy is stored in the phosphoric acid-sulfuric anhydride bond of the reaction product, adenosine 5'-phosphosulfate (APS), allowing sulfate to undergo further reactions. The energetic balance of the sulfate adenylation reaction favors ATP formation. Therefore, the reaction products, APS, and pyrophosphate (PP_i), need to be maintained at a low concentration by the enzymes inorganic pyrophosphatase that hydrolyses PP_i , APS reductase (EC 1.8.4.9) and APS kinase (EC 2.7.1.25; AKN) that metabolize APS. APS reductase catalyzes the first step of sulfate reduction. APS kinase catalyzes the ATP-dependent phosphorylation on the 3'-position of APS. *In vitro* tests have shown that excess APS inhibits APS kinase. The product PAPS is the substrate for the SOT proteins.

In general, the availability of PAPS for sulfation *in vivo* depends on its synthesis, transport, degradation, and utilization as investigated in mammals (Klaassen and Boles, 1997). Recently, it was shown that the transporter PAPT1 in the chloroplast envelope membrane is not only involved in the provision of PAPS for the extraplastidic sulfation reactions, but is also capable to transport PAP in an antiport manner. The loss of PAPT1 leads to a decreased production of sulfated compounds like Gl, increased production of dsGl, and the modulation of primary sulfate assimilation, another indication for the strong interconnectedness of primary and secondary sulfur metabolism (Gigolashvili et al., 2012). The by-product of the sulfation reaction, PAP, has gene regulatory attributes. In turn, PAP is regulated by the adenosine bisphosphate phosphatase SAL1 that dephosphorylates PAP to adenosine monophosphate. 3'-Phosphoadenosine 5'-phosphate accumulates at drought stress and high light conditions. Mutational studies indicated that PAP inhibits 5'-3' exoribonucleases in the cytosol and nucleus, which causes changes in expression of stress-responsive genes. It was suggested that a PAP-SAL1 retrograde pathway alters gene expression as part of the stress response

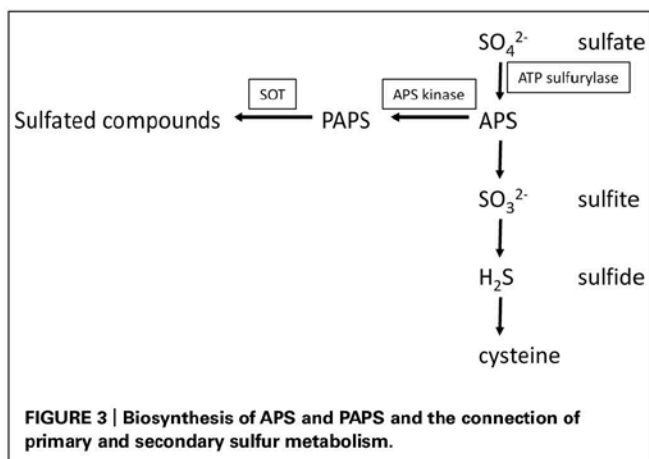


FIGURE 3 | Biosynthesis of APS and PAPS and the connection of primary and secondary sulfur metabolism.

(Estavillo et al., 2011). Additionally, a correlation between the increases of PAP with changes in the sulfur metabolism was reported. Further analysis of *sal1* knock out mutants led to the conclusion that changes of gene expression due to sulfur limitation is triggered by internal sulfur deficiency and not by low external sulfur levels. PAP accumulation also resulted in an increase of enzymatic oxygenation of fatty acids, an increase of jasmonic acid synthesis and a decrease of Gls. Possible explanations for the Gl decrease could be inhibition of dsGl SOTs or the disruption of PAPS transport from plastids to the cytosol (Gigolashvili et al., 2012; Lee et al., 2012).

SUBSTRATES FOR PLANT SOTs

The first isolated and characterized plant SOTs were flavonol 3'- and flavonol 4'-SOT of *Flaveria chloraefolia* (Varin et al., 1992) and later of *F. bidentis* (L.) Kuntze (Varin et al., 1997b). These SOTs sequentially sulfate specific hydroxyl groups of the flavonol quercetin to quercetin tetrasulfate. Flavonol biosynthesis was demonstrated to be regulated by auxin and ethylene. In turn, the flavonol quercetin and quercetin sulfates affect root development processes, such as the basipetal root auxin transport, elongation growth, and gravitropism (Faulkner and Rubery, 1992; Lewis et al., 2011).

So far, four flavonoid SOTs have been characterized in *A. thaliana*: AtSOT5, AtSOT8, AtSOT12, and AtSOT13 (see Table 1 for details). Hashiguchi et al. (2013) compared characteristics and substrate specificities of AtSOT5, AtSOT12, and AtSOT13.

AtSOT13 and AtSOT5 showed the highest activity with the flavonol galangin (3,5,7-trihydroxy-2-phenyl-4H-chromen-4-one), while AtSOT12 showed the highest activity for the flavanone naringenin [(2S)-5,7-dihydroxy-2-(4-hydroxyphenyl)-2,3-dihydro-4H-chromen-4-one] and was the only SOT that sulfates anthocyanidin. Interestingly, the AtSOTs showed no or comparably low activity with quercetin. It was speculated that the position-3 hydroxyl group of quercetin inhibits the catalytic activity. It was also shown that only AtSOT12 is able to use 3-hydroxyflavone as substrate, while 7-hydroxyflavone is used by all three AtSOTs. After comparisons of K_m -values using kaempferol as substrate, it was concluded that particular hydroxyl groups of kaempferol are specifically sulfated by the AtSOTs.

AtSOT8 was also characterized by Hashiguchi et al. (2014). The pH optimum at 5.5 was lower than for previously characterized flavonoid SOTs. Thus it was speculated that AtSOT8 might be located in the vacuole. Comparison of V_{max}/K_m -values showed that AtSOT8 prefers flavonol glycosides instead of their aglycone counterparts as substrates. Also, there was only activity to flavonoids with a hydroxyl group at position 7. Hence, it was suggested that AtSOT8 might be a flavonol glucoside-7 SOT. Surprisingly, neither sulfated glucoside flavonoids could be detected *in vivo* in *A. thaliana* by LC/MS, nor were there any accordant database entries. Possible explanations for the non-detected sulfated glucoside flavonoids could be low or condition-dependent occurrence.

AtSOT10 showed activity with brassinosteroids (Marsolais et al., 2007), specifically brassinosteroid biosynthetic end products. In summary, it was speculated that it inactivates brassinosteroids and therefore is involved in plant development processes. In

numerous studies, overexpression of brassinosteroid catabolic genes led to brassinosteroid-deficient phenotypes. However, overexpression and T-DNA insertion in null mutants of AtSOT10 did not show brassinosteroid-deficient phenotypes, emphasizing difficulties of transferring *in vitro* results to *in vivo* insights (Sandhu and Neff, 2013).

Of all investigated *A. thaliana* SOTs so far, AtSOT12 has the broadest substrate specificity. Besides using flavonoids as substrates, it was also shown to be active with brassinosteroids and salicylic acid. Within the brassinosteroids, it showed preference for 24-epibrassinosteroids (Marsolais et al., 2007). It was stereospecific for 24-epibrassinosteroids and accepted mammalian hydroxysteroids and estrogens, too. The most preferred substrate was the metabolic precursor 24-epicathasterone ($K_m = 6.9 \mu\text{M}$), which showed inhibitory effects above $5 \mu\text{M}$. The K_m -value for salicylic acid is comparably high ($440 \mu\text{M}$; Baek et al., 2010). Salicylic acid is a signal molecule in plant defense, and cellular concentrations increased up to $40 \mu\text{M}$ after pathogen infection indicating that sulfation of salicylic acid is a response to pathogen attack. This theory was supported by the fact that *atsot12* knock out mutants were less resistant to the pathogen *Pseudomonas syringae*, while AtSOT12 overexpressing lines showed a higher resistance.

Two *B. napus* brassinosteroid isoforms, BNST3 and BNST4, were enzymatically characterized. Recombinant BNST3 stereospecifically sulfated 24-epibrassinosteroids and preferred 24-epicathasterone ($K_m = 1.4 \mu\text{M}$), which is a biosynthetic intermediate of 24-epibrassinolide. Because of the biological inactivity of 24-epibrassinolide sulfate, it was hypothesized that BNST3 is involved in brassinosteroid inactivation (Rouleau et al., 1999). BNST4 also preferred 24-epibrassinosteroids ($K_m = 4.9 \mu\text{M}$), but also showed a broad substrate specificity with other steroids, also indicating a role in detoxification (Marsolais et al., 2004). Overall, they showed similar substrate specificities toward brassinosteroids as AtSOT12 (Marsolais et al., 2007).

AtSOT15 specifically sulfates 11- and 12-hydroxyjasmonate, which is a signaling molecule in plant defense and development. K_m -values indicate a higher affinity to 12- than to 11-hydroxyjasmonate ($10 \mu\text{M}$ and $50 \mu\text{M}$, respectively). 12-hydroxyjasmonate naturally occurs in *A. thaliana* and it was suggested that sulfation might function in inactivation of 12-hydroxyjasmonic acid (Gidda et al., 2003).

Komori et al. (2009) identified a 62 kDa, Golgi-localized, transmembrane protein, that sulfates tyrosylproteins in *A. thaliana*. The recombinantly expressed TPST sulfated tyrosine residues of precursor polypeptides of the “plant peptide containing sulfated tyrosine 1” (PSY1) and phytosulfokine (PSK). PSY1 and PSK are peptide hormones, which promote growth and cell proliferation (Matsubayashi and Sakagami, 1996; Amano et al., 2007). The activity with both substrates indicates broad substrate specificity. TPST showed a higher activity with PSY1, which was explained by a closer proximity of an acidic region to the sulfated tyrosine residue. TPST loss-of-function mutants showed numerous abnormal attributes, which led to the conclusion that sulfated peptides or proteins are involved in plant growth and development. Previously, in microsomal membrane preparations from carrot cells, rice, and asparagus TPST activity was shown

(Hanai et al., 2000). In rice, the K_m -value was $71 \mu\text{M}$ at a pH of 7.0–8.5 in the presence of manganese ions. The enzyme kinetic values such as K_m and V_{\max} *A. thaliana* TPST remain to be determined. TPST also sulfates peptide root meristem growth factors (RGFs), which are involved in postembryonic root development. Loss-of-function *tpst-1* mutants showed reduction in root meristem size and loss of coordination between cell elongation and expansion in the elongation–differentiation zone. Addition of RGF restored the meristem activity to ~70% and addition of RGF, PSK, and PSY1 restored the activity comparable to the wild-type. Sulfation of RGFs was found to be critical for its function. Further experiments showed that RGFs positively regulate the expression of PLETHORA transcription factors that mediate the pattern of the root stem niche (Matsuzaki et al., 2010).

Other examples of already characterized SOTs like a choline-*O*-sulfate SOT of the halophytic *Limonium* species (Rivoal and Hanson, 1994) and a plasma membrane-associated gallic acid SOT of *M. pudica* L. (Varin et al., 1997a) undercut the diversity of substrates and functions of these enzymes. Choline sulfate is an osmolyte that accumulates under saline conditions. The respective choline-*O*-SOT showed a fourfold higher activity under high salinity. The choline-*O*-SOT had its pH optimum at 9.0 and the K_m -value for choline was $25 \mu\text{M}$. The 42 kDa membrane bound gallic acid SOT might be involved in the regulation of the seismostatic response. It showed strict substrate specificity and a K_m -value of only $3.0 \mu\text{M}$.

Further studies indicate the existence of more SOTs, even though they were not especially isolated or characterized. It was shown *in vivo* that poplar trees convert hydroxylated metabolites of polychlorinated biphenyls (PCBs) into sulfated PCB. It was suggested that SOTs catalyze this reaction (Zhai et al., 2013). Sulfated polysaccharides occur in marine angiosperms, mangroves (Aquino et al., 2005, 2011), freshwater plants (Dantas-Santos et al., 2012), and algae (Ngo and Kim, 2013), which are likely to be sulfated by not yet identified SOTs.

GLUCOSINOLATES: PRODUCTS OF THE SOT REACTION

Glucosinolates are a group of over 200 nitrogen- and sulfur-containing natural products found in vegetative and reproductive tissues of 16 plant families within the Capparales (Clarke, 2010). They are well-known as the major secondary metabolites in agriculturally important crop plants of the Brassicaceae family, such as oilseed rape (*B. napus*), fodder and vegetables (e.g., broccoli and cabbage). The model plant *A. thaliana*. Gl share a core structure containing a β -D-glucopyranose residue linked via a sulfur atom to a (*Z*)-*N*-hydroximino sulfate ester. They are distinguished by a variable R group derived from one of several amino acids, mainly tryptophan, phenylalanine and methionine (Mithen, 2001). The Gl pattern varies among the plant species and among *A. thaliana* ecotypes. In 39 *A. thaliana* ecotypes, 34 different Gl have been identified. Quantity and composition of Gl depend on the developmental stage of the plants and on the plant organ (Kliebenstein et al., 2001a).

Intact Gl are not toxic to cells. However, after cell damage Gl are hydrolyzed, catalyzed by thioglucosidase enzymes (“myrosinase”), to produce a variety of volatile hydrolysis

products, such as thiocyanates, isothiocyanates, and nitriles. Only these breakdown products have a wide range of biological activities including both negative and positive effects (Fenwick and Heaney, 1983). In several studies these breakdown products were shown to be involved in plant defense against pathogens and herbivores. Thus, Gls are the best-characterized preformed defense compounds in the Brassicaceae and contribute to the protection against pathogens of the generalist type (Rausch and Wachter, 2005).

The last step in the Gl core structure biosynthesis of the different aliphatic, aromatic, and indole desulfo (ds) Gls is catalyzed by members of the SOT family. Glendening and Poulton (1990) partially purified a protein from *Lepidium sativum* L. that had PAPS-dependent dsGl SOT activity; however, at that time no molecular data was available. Later it was shown that three SOT proteins from *A. thaliana* are involved in Gl biosynthesis catalyzing the sulfation of dsGls to the intact Gls (Varin and Spertini, 2003; Piotrowski et al., 2004; Hirai et al., 2005).

Sulfotransferases involved in sulfation of desulfo-glucosinolates

The three dsGl AtSOT proteins (AtSOT16, AtSOT17, and AtSOT18) were predicted and then verified by different means (screening of many sulfated compounds, combining of knowledge, and integration of metabolomics and transcriptomics) for being responsible for the sulfation of dsGl (Varin and Spertini, 2003; Piotrowski et al., 2004; Hirai et al., 2005). Up till now, it was not unambiguously demonstrated why multiple *dsGl* SOT genes have been conserved during evolution in *A. thaliana* and in other Brassicaceae species.

The glucosylation and the sulfation reactions were assumed to be non-specific with respect to the side chain (Halkier, 1999). It is also hypothesized that first the side chains are elongated to synthesize so-called parent Gls, then the glycone moiety is developed and finally, several side chain modifications take place to produce the respective daughter Gl (Wittstock and Halkier, 2002). However, it is not clarified when the dsGls are sulfated by SOT proteins and whether there is a specificity for certain parent or daughter Gls. As the Gl pattern differs among *A. thaliana* ecotypes (Kliebenstein et al., 2001a), the investigation of the three dsGl SOTs from ecotype C24, which shows the broadest variety of Gls in comparison to other ecotypes, was most rational. In addition, one exemplary SOT from the fully sequenced ecotype Col-0 was investigated. To determine if and how these three dsGl SOT proteins might influence the Gl pattern, different *in vitro* enzyme assays were performed. Substrate specificity varies among the three proteins in the same ecotype (C24) and between ecotypes (C24 versus Col-0). AtSOT16 (C24) has the broadest substrate specificities. Tryptophan and phenylalanine-derived dsGl are the most preferred substrates, but it also accepts methionine-derived dsGl of chain length C3, C4, C5, C6, C7, and C8, although at much lower activities. AtSOT17 (C24) has narrow substrate specificities and does not act upon tryptophan-derived dsGl. Phenylalanine-derived benzyl dsGl is the most preferred substrate, but it also accepts methionine-derived dsGls, but has a strong preference for longer side chains, C6, C7, and C8. AtSOT18 (C24) also has narrow substrate specificities, does not act upon tryptophan-derived dsGl.

It accepts phenylalanine-derived dsGl and methionine-derived dsGl, but has a strong preference for longer side chains, C6, C7, and C8. In summary, these three enzymes differ significantly in their affinity for the investigated substrates and the co-substrate PAPS (Klein et al., 2006; Klein and Papenbrock, 2009). It was speculated that the differences between AtSOT16–18 could be an explanation for the different Gl patterns between organs, developmentally stages and growth environments reported by Brown et al. (2003). Anyhow, Møldrup et al. (2011) transformed genes involved in Gl biosynthesis into tobacco, thus successfully enabling it to synthesize Gls. In this approach, they could show that SOTs are not the bottleneck of Gl synthesis, but the supply of the co-substrate PAPS. Therefore, regulation of AtSOT16–18 could be, only taken together with other Gl and PAPS biosynthesis genes, partly responsible for Gl variation.

Up to now, the knowledge on secondary modifications of parent Gls is limited (Graser et al., 2001). In future work it could be interesting to verify the general acceptance that parent dsGls are sulfated before secondary modifications of Gls take place (Kliebenstein et al., 2001b). However, assuming 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 *in vitro*, but with no *in vivo* relevance.

In *A. thaliana* ecotypes SOT18 proteins differ in their sequence and substrate specificity

It was shown that AtSOT18 proteins from two different *A. thaliana* ecotypes differ in their kinetic parameters as well as their substrate specificities. The primary structure of AtSOT18 proteins from the ecotypes Col-0 and C24 differ in two amino acids (Klein et al., 2006; Klein and Papenbrock, 2009). One could assume that there could be a correlation of AtSOT18 enzyme activities and differences in Gl profiles between these ecotypes. Therefore, AtSOT18 sequences from eight *A. thaliana* ecotypes with highly diverse Gl patterns were investigated: The AtSOT18 sequence from Col-0 showed the highest similarity to the largest number of other sequences in the alignment. The AtSOT18 proteins showed sequence deviations of maximal two amino acids in comparison to the AtSOT18 sequence from Col-0. The positions of the amino acid replacements were different in each sequence. The small differences in the primary sequence lead to important structural changes in secondary and tertiary structure that might be the key for different kinetic activities toward a broad range of substrates (Luczak et al., 2013). All recombinant AtSOT18 proteins showed low substrate specificity with an indolic Gl, while the specificity for aliphatic substrates varied. There was no correlation in the kinetic behavior with the major dsGl contents or with the ratio of C₃/C₄ dsGl in the respective ecotype. Therefore, it is unlikely that dsGl AtSOT18 enzymes play a major role in shaping the Gl profile in *A. thaliana* (Luczak et al., 2013). Interestingly, in humans, inter-individual variation in sulfation capacity may be important in determining an individual's response to xenobiotics, and recent studies have begun to suggest roles for SOT polymorphism in disease susceptibility (Gamage et al., 2006). Variations in concentration and composition of Gls in *A. thaliana* ecotypes

and different environmental conditions can still neither be fully explained, nor predicted.

EXPRESSION OF SULFOTRANSFERASES

Sulfated compounds are mainly linked to biotic and abiotic stress response. This is supported by several expression studies of *SOTs*. So far, the mRNA levels of most characterized *A. thaliana SOTs* were rather low under normal growth conditions (Lacomme and Roby, 1996; Gidda et al., 2003; Piotrowski et al., 2004). This is supported by the fact that there is a relatively low number of *SOTs* in EST databases with the exception of *AtSOT15* and *AtSOT16* (Klein and Papenbrock, 2008). However, the expression of *AtSOT12*, *AtSOT15*, *AtSOT16*, and *AtSOT17* was significantly increased by treatment with jasmonate (Lacomme and Roby, 1996; Gidda et al., 2003; Piotrowski et al., 2004).

Only 17 *AtSOTs* were found to be present on 24 k Affymetrix chips and for many of those the absolute signal was quite low (Klein and Papenbrock, 2008). Hashiguchi et al. (2013, 2014) reported that microarray database research suggested that *AtSOT8* is mainly expressed in roots, while *AtSOT13* is expressed in the early stages of the embryonic development. Interestingly, Hashiguchi et al. (2013) cloned *AtSOT13* from 2-week old seedlings.

Transcripts of *AtSOT10* were mainly detected in roots. Transcript levels were repressed 4 h after *trans*-zeatin treatment. After 8 h no transcripts were detectable anymore by qRT-PCR (Marsolais et al., 2007).

Northern Blot analysis revealed that *AtSOT12*, the encoded protein uses brassinosteroids, flavonoids, and salicylic acid as substrates, is moderately expressed in roots and leaves, and highly in flowers, while no expression was detected in stems and siliques. Furthermore, it was strongly induced by salt and sorbitol and slightly by cold, ABA, auxins, cytokinins, methyl jasmonate, salicylic acid, and interactions with bacterial pathogens (Lacomme and Roby, 1996; Baek et al., 2010). These results strongly indicate a function of *AtSOT12* in stress and hormone response. Similar results were obtained for the respective homologous genes in *B. napus*. *BNST3* and *BNST4* mRNA levels were quite low, but increased after treatment with salicylic acid, ethanol, xenobiotics, low oxygen stress, and the herbicide safener naphthalic anhydride (Rouleau et al., 1999; Marsolais et al., 2004). *BNST3* and *BNST4* induction also indicates a function in stress response and detoxification.

The protein encoded by *AtSOT15* uses hydroxyl jasmonate as substrate. Expression was induced upon methyljasmonate and 12-hydroxyjasmonate treatment. Probably, it inactivates the function of jasmonic acids and therefore enhances the hypocotyl growth (Gidda et al., 2003). Yamashino et al. (2013) showed that *AtSOT15* transcription is also regulated by an external coincidence mechanism. Database research indicated that *AtSOT15* might be a target of the phytochrome interacting transcriptional factors PIF4 and PIF5. PIF4 and PIF5 are controlled by the circadian clock, but also independently influenced by light and temperature. Further qRT-PCR analysis showed that *AtSOT15* was diurnally regulated by PIF4 and PIF5 at the end of a short day dark phase and/or high temperatures. Accordingly, *AtSOT15* is induced under conditions when hypocotyl growth takes place.

At first, the *dsGl SOTs AtSOT16-18* were reported to be constitutively expressed in all leaves, flowers, and siliques (Varin and Spertini, 2003). Later Northern Blot analysis revealed that *AtSOT16* mRNA level increased after treatment with coronatine (an analog of octadecanoid signaling molecules), jasmonic acid precursor 12-oxophytodienonic, ethylene precursor ACC and after treatment with jasmonic acid. UV-C illumination and wounding also induced *AtSOT16* expression. *AtSOT17* mRNA increased 2.4 fold and 1.2 fold, respectively, while *AtSOT17* expression only slightly increased (1.3 fold) after coronatine treatment (Piotrowski et al., 2004). Regarding the developmental stages, *AtSOT16* and *AtSOT17* mRNA levels were highest in two week old seedlings and lowest in flowering plants. In contrary, *AtSOT18* levels were quite low in young plants and slightly increased after 5–6 weeks. Only *AtSOT17* expression was influenced by a 12 hour dark / 12 hour light cycle. It was the highest at the end of the light phase and the lowest at the end of the dark phase. No differences in any of the three mRNA levels were detected, when *A. thaliana* was grown in media with tenfold sulfate concentration (Klein et al., 2006).

Huseby et al. (2013) investigated how the Gl biosynthesis is controlled by light and the diurnal rhythm. By qRT-PCR analyses, it was shown that *AtSOT16*, *AtSOT17*, and *AtSOT18* are up regulated in light and down regulated in darkness. Further experiments indicated that the three *dsGl AtSOTs* are controlled by different transcriptional factors. In *A. thaliana* mutants, lacking the transcription regulator HY5, *AtSOT18* was less up-regulated than in the wild-type, indicating the HY5 is in control of *AtSOT18*. Interestingly, HY5 not only promotes numerous genes, but also seemed to repress MYBs. MYBs are a group of transcription factors, which are also involved in the control of Gl biosynthetic genes (Gigolashvili et al., 2007a,b, 2008; Hirai et al., 2007; Sønderby et al., 2007; Malitsky et al., 2008; Sønderby et al., 2010; Li et al., 2013; Jensen et al., 2014). *AtSOT16* was significantly down regulated in *myb34 myb51 myb122-2* triple mutant, revealing the specific control of these transcription factors (Frerigmann and Gigolashvili, 2014). Furthermore, *MYB51*, an indolic Gl metabolism specific transcription factor, was found to be down regulated in the dark, resulting in repression of indolic *dsGl* specific *AtSOT16*. Another indolic Gl transcription factor, *MYB34*, was up regulated after re-illumination (Celenza et al., 2005). This was not the case for *MYBs* controlling aliphatic Gl biosynthesis. It was concluded that *MYB* factors controlling biosynthesis of indolic Gl have a specific function in light regulation of their target gene, unlike the aliphatic group of *MYB* (Huseby et al., 2013).

Even so, the interaction and hierarchy of HY5 and MYBs still remains unclear. *AtSOT16*, *AtSOT17*, and *AtSOT18* were also up regulated in *apk1 apk2* double mutants. Hence, a reduction in PAPS supply and therefore reduction in Gl concentration leads to an up regulation *dsGl SOTs* (Mugford et al., 2009).

The expression of the twelve putative *dsGl SOTs* in *B. rapa* was investigated by qRT-PCR (Zang et al., 2009). Two genes are paralogs of *AtSOT16*, one of *AtSOT17*, and ten of *AtSOT18*. Generally, *BrSOT16s* were most strongly expressed, followed by *BrSOT18s* and then the *BrSOT17s*. With the exception of one *BrSOT18*, all of them were expressed in all examined tissue

types. One *BrSOT16* was expressed in all tissue types, except in the stamen, while the other one was strongly expressed in the stamen, but weakly in the floral bud and carpel. Some *BrSOT18s* were strongly expressed in the carpel and others in the stamen. Hence, the expression was not tissue-specific, but there was great variation in between tissue types. The expression of some *BrSOT18s* was developmentally regulated, but not of *BrSOT16s*. Again it was concluded that the expression could influence the Gl content, since SOTs play a crucial role in Gl biosynthesis.

The *TPST* gene is expressed in the whole plant, which was shown by analyzing *A. thaliana* *TPST-GUS* transformants, but especially strong in the root apical meristem and in the lateral root primordium and vascular tissues (Komori et al., 2009).

Expression of the 35 *O. sativa* SOTs was investigated by microarray database analysis (Chen et al., 2012). The overall expression was reported to be considerably low. Low expression levels were in the apical meristem and young leaves. Higher expression was found in the stigma, ovary and roots. Treatment with IAA and BAP led to up and down regulation of several SOTs also with differences in respect to tissue types and seedlings age. Furthermore, expression of eleven SOTs reacted to abiotic stress, such as high and low temperatures and dehydration. It was concluded that the individual responses of SOTs indicate functions in stress response and plant development.

Overall, SOT expressions suggest functions in plant defense, stress response, signaling and developmental regulation. Sulfation can either lead to activation or deactivation of the according substrate. SOT expression takes place basically in all organs and many stages in plant development. Interestingly, all SOTs studied so far, were induced by several conditions or stress signaling compounds, indicating a general stress response. Additionally, in the case of *dsGl AtSOTs* and *AtSOT15*, a diurnal and circadian control was detected. It seems plausible that this could be the case for other SOTs, too.

WHAT IS KNOWN ABOUT THE REACTION MECHANISM OF SULFOTRANSFERASES

So far, the reaction mechanism of plant SOTs remains largely unknown. Kinetic and inhibition studies of a flavonol 3'-SOT from *F. chloraefolia* A. Gray led to the hypothesis of an ordered Bi-Bi mechanism (Varin and Ibrahim, 1992). However, the few conducted experiments are not sufficient enough for a definite conclusion.

More information is available about human SOTs. By pre-steady state binding studies, isotopic trapping, quenched-flow, and classic inhibition studies, Wang et al. (2014) completely solved the kinetic mechanism of the human SOT SULT2A1. SULT2A1 sulfates dehydroepiandrosterone and regulates binding of steroids to their receptors and detoxifies steroid-like xenobiotics. The according mechanism was found to be rapid equilibrium random. In this mechanism, substrates are bound and products are released in a random order. The ligands are bound in separate binding sites and released independently of the presence of its partner, hence without contribution of sulfuryl-group interactions. Ligand-binding rate constants also indicated that ligand-protein interactions, which enable the chemical reaction,

are either established prior to addition of the second substrate and/or they are engaged as the system moves toward the transition state. Furthermore, it was shown that the release of the PAP nucleotide is the rate-determining step of the reaction. Substrate inhibition was explained by trapping of PAP in a dead end complex (enzyme with bound PAP and substrate), which decreases the release of PAP. Since closely related enzymes often share the same mechanism, it was speculated that this could also be the case for other human SOTs. Anyhow, this cannot be done for plant SOTs without further experimental analysis. This is already illustrated, when regarding that SULT2A1 is a half-site reactive dimer, while yet investigated plant SOTs are monomers.

The mechanism of a monomeric SOT Stf0 from *Mycobacterium tuberculosis* was analyzed by electrospray ionization mass spectrometry and Fourier transform ion cyclotron resonance mass spectrometry (Pi et al., 2005). Stf0 forms trehalose sulfate, which is the core disaccharide of the potential virulence factor sulfolipid-1. Interestingly, the results also indicated a rapid equilibrium random mechanism, at which the sulfuryl group is transferred in the ternary complex. Again, there is one binding site for products and one independent binding site for substrates. Results also indicated that PAPS binding was competitively inhibited by PAP.

Further studies of non-herbal SOTs, human estrogen SOT (Zhang et al., 1998) and insect retinol dehydratase (Vakiani et al., 1998), also indicated random Bi-Bi mechanisms. So far, only for a *Rhizobium meliloti* NodH SOT a hybrid random ping-pong mechanism was suggested (Pi et al., 2004). Therefore, the investigation of the complete kinetic mechanisms of plant SOTs remains an interesting task, which could also give new insights of the overall evolution of SOTs.

HOW TO IDENTIFY THE SUBSTRATE SPECIFICITY? CHANCES AND RESTRICTION OF MODELING

Simple online tools like SWISS-MODEL do not lead to satisfying Z-scores and therefore unreliable models. However, Cook et al. (2013) generated significant models of human SOTs by using more advanced programs such as MODELLER, GOLD, GROMACS, and AMBER. Models of human SOTs SULT1A1 and SULT1A2, which are Phase II detoxifying enzymes, were used for *in silico* docking studies. As substrates, 1455 small molecule drugs were tested. For SULT1A1, 76 substrates were predicted, of which 53 were already known substrates. Of the remaining 23 putative substrates, 21 were tested in enzyme assays and all of them were accepted as substrates. Of 22 predicted substrates for SULT2A1, eight were not previously mentioned in literature. Enzyme assays were carried out with four of the eight newly identified substrates, and all of them were accepted as substrates. For both SOTs neither a single false positive nor a false negative prediction occurred. Furthermore, 136 SULT1A1 and 35 SULT2A1 inhibitors were predicted. Two of those were exemplary tested in classical inhibition studies and both showed inhibitory effects.

But can these techniques be transferred to plant SOTs? Principally they could be transferred to plant SOTs, but it has to be kept in mind that SULT1A1 and SULT1A2 are extensively studied SOTs. Building reliable models for *in silico* docking studies requires

knowledge about structure and mechanism of the protein. For example, SULT1A1 and SULT1A2 have a site cap, which regulates substrate specificity (Cook et al., 2012). This also had to be considered when generating the *in silico* models. Furthermore, it was shown that human SOTs have a high plasticity (Allali-Hassani et al., 2007; Berger et al., 2011) and that PAP binding leads to dramatic conformational changes, such as pre-formation of the acceptor binding pocket (Bidwell et al., 1999; Dajani et al., 1999; Berger et al., 2011). Hence, without prior structural knowledge about at least some of the plant SOTs, *in silico* modeling is still restricted. Nevertheless, *in silico* modeling of SOTs is a promising approach, especially because of the limitations in substrate identification based on phylogenetic analyses.

A. THALIANA AS A MODEL PLANT – SUITED FOR THE ELUCIDATION OF ALL SOT FUNCTIONS?

Elucidation of all SOT functions in *A. thaliana* as a model plant is difficult for several reasons. Until now, ten out of 22 identified putative *A. thaliana* SOTs have been enzymatically characterized *in vitro*. The identified substrates were peptides, flavonoids, brassinosteroids, GIs, hydroxyjasmonate, and salicylic acid. As discussed before, prospects of phylogenetic analyses are very limited for SOTs. Already small changes in the sequence can lead to wide variations in substrate specificity. Even reliable predictions of yet uncharacterized SOTs in the organism *A. thaliana* are not possible. Therefore, reliable function prediction of SOTs in other plant species on the basis of *A. thaliana* sequences seems very unlikely.

Even when comparing SOTs that use the same class of substrates from different plant species, not only differences in kinetic values, but also variation of specificity toward different substrates and specific hydroxyl groups are noticed. For example, flavonol SOTs (AtSOT5, AtSOT8, AtSOT12, AtSOT13) from *A. thaliana* prefer kaempferol or flavonol glycosides as substrate and sulfate the hydroxyl groups at 3- and 7-position (Hashiguchi et al., 2013, 2014). But flavonol SOTs from *F. chloraefolia* and *F. bidentis* (L.) Kuntze prefer quercetin as substrates and sulfate at 3'- and 4'-position (Varin et al., 1992, 1997b). Furthermore, all so far characterized *A. thaliana* SOTs sulfate a broad range of substrates. Most are functionally and biochemically related, but for example in case of AtSOT12, substrates with a wide range of biological functions are accepted as substrates.

Another difficulty is that SOTs are part of secondary metabolism and therefore fulfill species-specific functions. Hence, it is unlikely that all types of SOTs occur in *A. thaliana*. This is supported by the Pfam database research described in chapter 3.1. In *A. thaliana*, only the TPST contains a *Sulfotransfer_2* domain (PF03567). The remaining AtSOTs all contain a *Sulfotransfer_1* domain (PF00685), while the *Sulfotransfer_3* (PF13469) and *Gal-3-O_sulfotr* domain (PF06990) are only present in algae. *Arylsulfotran_2* domain (PF14269) is only found in a single *Ricinus communis* sequence. In addition, SOT homologues in different plant species differ in their number of paralogs. For example, there were nine homologues of AtSOT18 found in *B. rapa* (Zang et al., 2009), which could differ in their characteristics.

All in all, it remains an important future task to clarify the biological functions and characteristics of the remaining

A. thaliana SOTs, not only by *in vitro* enzymatic assays, but in consideration of mutation, expression and localization studies, as well as metabolomics. Findings could at least be partly transferred and give valuable hints about specific SOTs in other species. Since *A. thaliana* is the most studied plant, complete characterization of all AtSOTs could also give more information about the connection of primary and secondary metabolisms in plants in general.

FUTURE CHALLENGES

Plant SOT research still remains a biological field with many open questions, especially in comparison with mammalian SOTs. In the model organism *A. thaliana*, only ten out of 22 SOTs have been enzymatically characterized *in vitro* so far. In many of these cases, the *in vivo* function is not elucidated yet. Some compounds, which were found to be sulfated by SOTs *in vitro*, could not be detected *in vivo*, as it was the case for sulfated glucoside flavonoids, sulfated by AtSOT8 (Hashiguchi et al., 2014). Furthermore, the function of sulfation or the sulfated compound is often not completely understood. Hence, for a deeper understanding it is advisable to follow *in vitro* enzymatic characterization with mutation, expression and localization studies.

For the remaining twelve putative *A. thaliana* SOTs, disregarding the pseudogene AtSOT2, no accepted substrates have been identified so far. Due to the enormous number of putative substrates and the restricted reliability of phylogenetic analyses, complete functional elucidation of all AtSOTs is an ambitious goal. Hence, recombinant expression and offering randomly chosen substrates seems like looking for a needle in a haystack. A more promising approach could be to feed wild-type and mutant plants with ³⁵S, followed by mass spectrometry analysis. This can facilitate the identification of newly sulfated compounds *in vivo*. Next steps could be the isolation of these compounds, depending on its availability and chemical properties. If possible, the compounds could be bound to a column and used for affinity chromatography of total protein preparations. This would be a very systematic approach and was already partly used for the successful identification of TPST (Komori et al., 2009).

Especially in pharmaceutical research, *in silico* analysis has become a powerful tool (Song et al., 2009). With the help of three dimensional structures of the target molecules, computational drug design becomes more and more promising. The three dimensional structure of one *A. thaliana* SOT (AtSOT12) has already been solved, but without including substrates into the crystals (Smith et al., 2004). A deeper understanding could be reached with the help of more solved structures with and without substrates and with additional knowledge about the enzymatic mechanism. Definitely identified binding sites, combined with protein modeling could give more specific hints about putative substrates of SOTs.

Another interesting field would be the elucidation of SOTs from more plant species, especially highly specialized ones. In order to cope with additional stress, plants growing in challenging environments often biosynthesize specific compounds. Its properties are often promising from a biological point of view, for a better understanding of stress response, but also interesting for medical or biotechnological applications. Zosteric acid [*p*-(sulfo-oxy) cinnamic acid] from the seagrass *Zostera*

marina, for example, has anti-fouling properties (Newby et al., 2006). In a patent, SOT involvement in biosynthesis was suggested, but not proven yet (Alexandratos, 1999). Additionally, in *Zostera*, *Halophila*, and *Thalassia* seagrass, the existence of sulfated flavones was indicated (Harborne and Williams, 1976). Furthermore sulfated polysaccharides were detected in seagrass (Aquino et al., 2005, 2011), freshwater plants (Dantas-Santos et al., 2012) and algae (Ngo and Kim, 2013). While the sulfation of polysaccharides is well-studied in humans (Kusche-Gullberg and Kjellén, 2003), no polysaccharide SOTs have been studied in plants yet. It is hypothesized that sulfated polysaccharides modify the cell wall in halophytes in order to increase salt tolerance (Aquino et al., 2005). They are also interesting for human nutrition and pharmaceutical products, because of their antioxidant, anti-allergic, anti-human immunodeficiency virus, anti-cancer and anticoagulant properties (Ngo and Kim, 2013).

Overall, substrate specificities, regulations, and catalytic mechanisms of plant SOTs are still poorly understood. Considering the large number of possible functions, further research on these enzymes remains a challenging field.

ACKNOWLEDGMENTS

We acknowledge support by Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of Leibniz Universität Hannover. Experimental work in our laboratory is funded by the Deutsche Forschungsgemeinschaft (PA 764/10-1).

REFERENCES

- Alexandratos, S. D. (1999). Synthesis and purification of zosteric acid. US Patent 5990336.
- Allali-Hassani, A., Pan, P. W., Dombrowski, L., Najmanovich, R., Tempel, W., Dong, A., et al. (2007). Structural and chemical profiling of the human cytosolic sulfotransferases. *PLoS Biol.* 5:e97. doi: 10.1371/journal.pbio.0050097
- Amano, Y., Tsubouchi, H., Shinohara, H., Ogawa, M., and Matsubayashi, Y. (2007). Tyrosine-sulfated glycopeptide involved in cellular proliferation and expansion in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18333–18338. doi: 10.1073/pnas.0706403104
- Aquino, R. S., Grativol, C., Mourão, P. S., and Meyer, P. (2011). Rising from the sea: correlations between sulfated polysaccharides and salinity in plants. *PLoS ONE* 6:e18862. doi: 10.1371/journal.pone.0018862
- Aquino, R. S., Landeira-Fernandez, A. M., Valente, A. P., Andrade, L. R., and Mourão, P. A. (2005). Occurrence of sulfated galactans in marine angiosperms: evolutionary implications. *Glycobiology* 15, 11–20. doi: 10.1093/glycob/cwh138
- Baek, D., Pathange, P., Chung, J.-S., Jiang, J., Gao, L., Oikawa, A., et al. (2010). A stress-inducible sulphotransferase sulphonates salicylic acid and confers pathogen resistance in *Arabidopsis*. *Plant Cell Environ.* 33, 1383–1392. doi: 10.1111/j.1365-3040.2010.02156.x
- Berger, I., Guttman, C., Amar, D., Zarivach, R., and Aharoni, A. (2011). The molecular basis for the broad substrate specificity of human sulfotransferase 1A1. *PLoS ONE* 6:e26794. doi: 10.1371/journal.pone.0026794
- Bidwell, L. M., Mcmanus, M. E., Gaedigk, A., Kakuta, Y., Negishi, M., Pedersen, L., et al. (1999). Crystal structure of human catecholamine sulfotransferase. *J. Mol. Biol.* 293, 521–530. doi: 10.1006/jmbi.1999.3153
- Brown, P. D., Tokuhisa, J. G., Reichelt, M., and Gershenzon, J. (2003). Variation of glucosinolate accumulation among different organs and developmental stages of *Arabidopsis thaliana*. *Phytochemistry* 62, 471–481. doi: 10.1016/S0031-9422(02)00549-6
- Celenza, J. L., Quiel, J. A., Smolen, G. A., Merrikkh, H., Silvestro, A. R., Normanly, J., et al. (2005). The *Arabidopsis* ATR1 Myb transcription factor controls indolic glucosinolate homeostasis. *Plant Physiol.* 137, 253–262. doi: 10.1104/pp.104.054395
- Chen, R., Jiang, Y., Dong, J., Zhang, X., Xiao, H., Xu, Z., et al. (2012). Genome-wide analysis and environmental response profiling of SOT family genes in rice (*Oryza sativa*). *Genes Genom.* 34, 549–560. doi: 10.1007/s13258-012-0053-5
- Clarke, D. B. (2010). Glucosinolates, structures and analysis in food. *Anal. Methods* 2, 310–325. doi: 10.1039/B9AY00280D
- Cook, I., Wang, T., Falany, C. N., and Leyh, T. S. (2012). A nucleotide-gated molecular pore selects sulfotransferase substrates. *Biochemistry* 51, 5674–5683. doi: 10.1021/bi300631g
- Cook, I., Wang, T., Falany, C. N., and Leyh, T. S. (2013). High accuracy in silico sulfotransferase models. *J. Biol. Chem.* 288, 34494–34501. doi: 10.1074/jbc.M113.510974
- Dajani, R., Cleasby, A., Neu, M., Wonacott, A. J., Jhoti, H., Hood, A. M., et al. (1999). X-ray crystal structure of human dopamine sulfotransferase, SULT1A3: molecular modeling and quantitative structure-activity relationship analysis demonstrate a molecular basis for sulfotransferase substrate specificity. *J. Biol. Chem.* 274, 37862–37868. doi: 10.1074/jbc.274.53.37862
- Dantas-Santos, N., Gomes, D. L., Costa, L. S., Cordeiro, S. L., Costa, M. S. S. P., Trindade, E. S., et al. (2012). Freshwater plants synthesize sulfated polysaccharides: heterogalactans from water hyacinth (*Eichhornia crassipes*). *Int. J. Mol. Sci.* 13, 961–976. doi: 10.3390/ijms13010961
- Estavillo, G. M., Crisp, P. A., Pomsiriwong, W., Wirtz, M., Collinge, D., Carrie, C., et al. (2011). Evidence for a SAL1-PAP chloroplast retrograde pathway that functions in drought and high light signaling in *Arabidopsis*. *Plant Cell* 23, 3992–4012. doi: 10.1105/tpc.111.091033
- Faulkner, I. J., and Rubery, P. H. (1992). Flavonoids and flavonoid sulphates as probes of auxin-transport regulation in *Cucurbita pepo* hypocotyl segments and vesicles. *Planta* 186, 618–625. doi: 10.1007/BF00198044
- Fenwick, G. R., and Heaney, R. K. (1983). Glucosinolates and their breakdown products in cruciferous crops, foods and feedingstuffs. *Food Chem.* 11, 249–271. doi: 10.1016/0308-8146(83)90074-90072
- Finn, R. D., Bateman, A., Clements, J., Coghill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucl. Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223
- Frerigmann, H., and Gigolashvili, T. (2014). MYB34, MYB51, and MYB122 distinctly regulate indolic glucosinolate biosynthesis in *Arabidopsis thaliana*. *Mol. Plant* 7, 814–828. doi: 10.1093/mp/ssu004
- Fukuta, M., Kobayashi, Y., and Uchimura, K. (1998). Molecular cloning and expression of human chondroitin 6-sulfotransferase. *Biochim. Biophys. Acta* 1399, 57–61.
- Gamage, N., Barnett, A., Hempel, N., Duggleby, R. G., Windmill, K. F., Martin, J. L., et al. (2006). Human sulfotransferases and their role in chemical metabolism. *Toxicol. Sci.* 90, 5–22. doi: 10.1093/toxsci/kfj061
- Gidda, S. K., Miersch, O., Levitin, A., Schmidt, J., Wasternack, C., and Varin, L. (2003). Biochemical and molecular characterization of a hydroxyjasmonate sulfotransferase from *Arabidopsis thaliana*. *J. Biol. Chem.* 278, 17895–17900. doi: 10.1074/jbc.M211943200
- Gidda, S. K., and Varin, L. (2006). Biochemical and molecular characterization of flavonoid 7-sulfotransferase from *Arabidopsis thaliana*. *Plant Physiol. Biochem.* 44, 628–636. doi: 10.1074/jbc.M211943200
- Gigolashvili, T., Berger, B., Mock, H. -P., Muller, C., Weisshaar, B., and Flügge, U.-I. (2007a). The transcription factor HAG1/MYB51 regulates indolic glucosinolate biosynthesis in *Arabidopsis thaliana*. *Plant J.* 50, 886–901. doi: 10.1111/j.1365-313X.2007.03099.x
- Gigolashvili, T., Yatushevich, R., Berger, B., Müller, C., and Flügge, U.-I. (2007b). The R2R3-MYB transcription factor HAG1/MYB28 is a regulator of methionine-derived glucosinolate biosynthesis in *Arabidopsis thaliana*. *Plant J.* 51, 247–261. doi: 10.1111/j.1365-313X.2007.03133.x
- Gigolashvili, T., Engqvist, M., Yatushevich, R., Müller, C., and Flügge, U.-I. (2008). HAG2/MYB76 and HAG3/MYB29 exert a specific and coordinated control on the regulation of aliphatic glucosinolate biosynthesis in *Arabidopsis thaliana*. *New Phytol.* 177, 627–642. doi: 10.1111/j.1469-8137.2007.02295.x
- Gigolashvili, T., Geier, M., Ashykhmina, N., Frerigmann, H., Wulfert, S., Krueger, S., et al. (2012). The *Arabidopsis* thylakoid ADP/ATP carrier TAAC has an additional role in supplying plastidic phosphoadenosine 5-phosphosulfate to the cytosol. *Plant Cell* 24, 4187–4204. doi: 10.1105/tpc.112.101964
- Glendening, T. M., and Poulton, J. E. (1990). Partial purification and characterization of a 3'-phosphoadenosine 5'-phosphosulfate: desulfoglucosinolate

- sulfotransferase from cress (*Lepidium sativum*). *Plant Physiol.* 94, 811–818. doi: 10.1104/pp.94.2.811
- Graser, G., Oldham, N. J., Brown, P. D., Temp, U., and Gershenzon, J. (2001). The biosynthesis of benzoic acid glucosinolate esters in *Arabidopsis thaliana*. *Phytochemistry* 57, 23–32. doi: 10.1016/S0031-9422(00)00501-X
- Halkier, B. A. (1999). “Glucosinolates,” in *Naturally Occurring Glycosides: Chemistry, Distribution and Biological Properties*, ed. R. Ikan (New York: John Wiley & Sons Ltd), 193–223.
- Hanai, H., Nakayama, D., Yang, H., Matsubayashi, Y., Hirota, Y., and 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. doi: 10.1016/S0014-5793(00)01299-0
- Harborne, J. B., and Williams, C. A. (1976). Occurrence of sulphated flavones and caffeic acid esters in members of the fluviales. *Biochem. Syst. Ecol.* 4, 37–41. doi: 10.1016/0305-1978(76)90007-7
- Hashiguchi, T., Sakakibara, Y., Hara, Y., Shimohira, T., Kurogi, K., Akashi, R., et al. (2013). Identification and characterization of a novel kaempferol sulfotransferase from *Arabidopsis thaliana*. *Biochem. Biophys. Res. Commun.* 434, 829–835. doi: 10.1016/j.bbrc.2013.04.022
- Hashiguchi, T., Sakakibara, Y., Shimohira, T., Kurogi, K., Yamasaki, M., Nishiyama, K., et al. (2014). Identification of a novel flavonoid glycoside sulfotransferase in *Arabidopsis thaliana*. *J. Biochem.* 155, 91–97. doi: 10.1093/jb/mvt102
- Hernández-Sebastiá, C., Varin, L., and Marsolais, F. (2008). “Sulfotransferases from plants, algae and phototrophic bacteria,” in *Sulfur Metabolism in Phototrophic Organisms*, eds R. Hell, C. Dahl, D. Knaff, and T. Leustek (Amsterdam: Springer), 111–130.
- Hirai, M. Y., Klein, M., Fujikawa, Y., Yano M., Goodenowe, D. B., Yamazaki, Y., et al. (2005). Elucidation of gene-to-gene networks in *Arabidopsis* by integration of metabolomics and transcriptomics. *J. Biol. Chem.* 280, 25590–25595. doi: 10.1074/jbc.M502332200
- Hirai, M. Y., Sugiyama, K., Sawada, Y., Tohge, T., Obayashi, T., Suzuki, A., et al. (2007). Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6478–6483. doi: 10.1073/pnas.0611629104
- Huseby, S., Koprivova, A., Lee, B.-R., Saha, S., Mithen, R., Wold, A.-B., et al. (2013). Diurnal and light regulation of sulphur assimilation and glucosinolate biosynthesis in *Arabidopsis*. *J. Exp. Bot.* 64, 1039–1048. doi: 10.1093/jxb/ers378
- Jensen, L. M., Halkier, B. A., and Burow, M. (2014). How to discover a metabolic pathway? An update on gene identification in aliphatic glucosinolate biosynthesis, regulation and transport. *Biol. Chem.* 395, 529–543. doi: 10.1515/hsz-2013-0286
- Kakuta, Y., Pedersen, L. G., Carter, C. W., Negishi, M., and Pedersen, L. C. (1997). Crystal structure of estrogen sulphotransferase. *Nature Struct. Biol.* 4, 904–908. doi: 10.1038/nsb1197-904
- Kakuta, Y., Petrotchenko, E. V., Pedersen, L. C., and Negishi, M. (1998). The sulfuryl transfer mechanism. *J. Biol. Chem.* 273, 27325–27330. doi: 10.1074/jbc.273.42.27325
- Klaassen, C. D., and Boles, J. W. (1997). The importance of 3'-phosphoadenosine 5'-phosphosulfate (PAPS) in the regulation of sulfation. *FASEB J.* 11, 404–418.
- Klein, M., and Papenbrock, J. (2004). The multi-protein family of *Arabidopsis* sulphotransferases and their relatives in other plant species. *J. Exp. Bot.* 55, 1809–1820. doi: 10.1093/jxb/erh183
- Klein, M., and Papenbrock, J. (2008). “Sulfotransferases and their role in glucosinolate biosynthesis,” in *Sulfur Assimilation and Abiotic Stress in Plants*, eds N. Khan, S. Singh, and S. Umar (Berlin: Springer), 149–166. doi: 10.1007/978-3-540-76326-0_7
- Klein, M., and Papenbrock, J. (2009). Kinetic parameters of desulfoglucosinolate sulfotransferases. *Physiol. Plant.* 135, 140–149. doi: 10.1111/j.1399-3054.2008.01182.x
- Klein, M., Reichelt, M., Gershenzon, J., and Papenbrock, J. (2006). The three desulfo-glucosinolate sulfotransferase proteins in *Arabidopsis* have different substrate specificities and are differentially expressed. *FEBS J.* 273, 122–136. doi: 10.1111/j.1742-4658.2005.05048.x
- Kliebenstein, D. J., Kroymann, J., Brown, P., Figuth, A., Pedersen, D., Gershenzon, J., et al. (2001a). Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. *Plant Physiol.* 126, 811–825. doi: 10.1104/pp.126.2.811
- Kliebenstein, D. J., Lambrix, V. M., Reichelt, M., and Gershenzon, J. (2001b). Gene duplication in the diversification of secondary metabolism: Tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in *Arabidopsis*. *Plant Cell* 13, 681–693. doi: 10.1105/tpc.13.3.681
- Komori, R., Amano, Y., Ogawa-Ohnishi, M., and Matsubayashi, Y. (2009). Identification of tyrosylprotein sulfotransferase in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15067–15072. doi: 10.1073/pnas.0902801106
- Kusche-Gullberg, M., and Kjellén, L. (2003). Sulfotransferases in glycosaminoglycan biosynthesis. *Curr. Opin. Struct. Biol.* 13, 605–611. doi: 10.1016/j.sbi.2003.08.002
- Labonne, J. J. D., Goultiaeva, A., and Shore, J. S. (2009). High-resolution mapping of the S-locus in *Turnera* leads to the discovery of three genes tightly associated with the S-alleles. *Mol. Genet. Genomics* 281, 673–685. doi: 10.1007/s00438-009-0439-5
- Lacomme, C., and 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. doi: 10.1007/BF00020810
- Lee, B. R., Huseby, S., Koprivova, A., Chételat, A., Wirtz, M., Mugford, S. T., et al. (2012). Effects of *fou8/fry1* mutation on sulfur metabolism: is decreased internal sulfate the trigger of sulfate starvation response? *PLoS ONE* 7:e39425. doi: 10.1371/journal.pone.0039425
- Lewis, D. R., Ramirez, M. V., Miller, N. D., Vallabhaneni, P., Ray, W. K., Helm, R. F., et al. (2011). Auxin and ethylene induce flavonol accumulation through distinct transcriptional networks. *Plant Physiol.* 156, 144–164. doi: 10.1104/pp.111.172502
- Li, Y., Sawada, Y., Hirai, A., Sato, M., Kuwahara, A., Yan, X., et al. (2013). Novel insights into the function of *Arabidopsis* R2R3-MYB transcription factors regulating aliphatic glucosinolate biosynthesis. *Plant Cell Physiol.* 54, 1335–1344. doi: 10.1093/pcp/pct085
- Luczak, S., Forlani, F., and Papenbrock, J. (2013). Desulfo-glucosinolate sulfotransferases isolated from several *Arabidopsis thaliana* ecotypes differ in their sequence and enzyme activities. *Plant Physiol. Biochem.* 63, 15–23. doi: 10.1016/j.plaphy.2012.11.005
- Malitsky, S., Blum, E., Less, H., Venger, I., Elbaz, M., Morin, S., et al. (2008). The transcript and metabolite networks affected by the two clades of *Arabidopsis* glucosinolate biosynthesis regulators. *Plant Physiol.* 148, 2021–2049. doi: 10.1104/pp.108.124784
- Marsolais, F., Boyd, J., Paredes, Y., Schinas, A.-M., Garcia, M., Elzein, S., et al. (2007). Molecular and biochemical characterization of two brassinosteroid sulfotransferases from *Arabidopsis*, AtST4a (At2g14920) and AtST1 (At2g03760). *Planta* 225, 1233–1244. doi: 10.1007/s00425-006-0413-y
- Marsolais, F., Gidda, S. K., Boyd, J., and Varin, L. (2000). Structural and functional similarity with mammalian enzymes. *Rec. Adv. Phytochem.* 34, 433–456. doi: 10.1371/journal.pone.0104594
- Marsolais, F., Sebastià, C. H., Rousseau, A., and 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. doi: 10.1016/j.plantsci.2004.01.019
- Marsolais, F., and 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. doi: 10.1074/jbc.270.51.30458
- Matsubayashi, Y., and Sakagami, Y. (1996). Phytosulfokine, sulfated peptides that induce the proliferation of single mesophyll cells of *Asparagus officinalis* L. *Proc. Natl. Acad. Sci. U.S.A.* 93, 7623–7627. doi: 10.1073/pnas.93.15.7623
- Matsuzaki, Y., Ogawa-Ohnishi, M., Mori, A., and Matsubayashi, Y. (2010). Secreted peptide signals required for maintenance of root stem cell niche in *Arabidopsis*. *Science* 329, 1065–1067. doi: 10.1126/science.1191132
- Mithen, R. (2001). Glucosinolates - biochemistry, genetics and biological activity. *Plant Growth Regul.* 34, 91–103. doi: 10.1023/A:1013330819778
- Moldrup, M. E., Geu-Flores, F., Olsen, C. E., and Halkier, B. A. (2011). Modulation of sulfur metabolism enables efficient glucosinolate engineering. *BMC Biotechnol.* 11:12. doi: 10.1186/1472-6750-11-12.
- Mugford, S. G., Yoshimoto, N., Reichelt, M., Wirtz, M., Hill, L., Mugford, S. T., et al. (2009). Disruption of adenosine-5'-phosphosulfate kinase in *Arabidopsis* reduces levels of sulfated secondary metabolites. *Plant Cell* 21, 910–927. doi: 10.1105/tpc.109.065581
- Newby, B.-M. Z., Cutright, T., Barrios, C., and Xu, Q. (2006). Zosteriac acid—An effective antifoulant for reducing fresh water bacterial attachment on coatings. *JCT Res.* 3, 69–76. doi: 10.1007/s11998-006-0007-4

- Ngo, D.-H., and Kim, S.-K. (2013). Sulfated polysaccharides as bioactive agents from marine algae. *Int. J. Biol. Macromol.* 62, 70–75. doi: 10.1016/j.ijbiomac.2013.08.036
- Niehrs, C., Beisswanger, R., and Huttner, W. B. (1994). Protein tyrosine sulfation, 1993 – an update. *Chem. Biol. Interact.* 92, 257–271. doi: 10.1016/0009-2797(94)90068-X
- Pi, N., Hoang, M. B., Gao, H., Mougous, J. D., Bertozzi, C. R., and Leary, J. A. (2005). Kinetic measurements and mechanism determination of Stf0 sulfotransferase using mass spectrometry. *Anal. Biochem.* 341, 94–104. doi: 10.1016/j.ab.2005.02.004
- Pi, N., Yu, Y., Mougous, J. D., and Leary, J. A. (2004). Observation of a hybrid random ping-pong mechanism of catalysis for NodST: a mass spectrometry approach. *Protein Sci.* 13, 903–912. doi: 10.1110/ps.03581904
- Piotrowski, M., Schemenewitz, A., Lopukhina, A., Muller, A., Janowitz, T., Weiler, E. W., et al. (2004). Desulfoglucosinolate sulfotransferases from *Arabidopsis thaliana* catalyze the final step in the biosynthesis of the glucosinolate core structure. *J. Biol. Chem.* 279, 50717–50725. doi: 10.1074/jbc.M407681200
- Rausch, T., and Wachter, A. (2005). Sulfur metabolism: a versatile platform for launching defence operations. *Trends Plant Sci.* 10, 503–509. doi: 10.1016/j.tplants.2005.08.006
- Rivoal, J., and Hanson, A. D. (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.
- Rouleau, M., Marsolais, F., Richard, M., Nicolle, L., Voigt, B., Adam, G., et al. (1999). Inactivation of brassinosteroid biological activity by a salicylate-inducible steroid sulfotransferase from *Brassica napus*. *J. Biol. Chem.* 274, 20925–20930. doi: 10.1074/jbc.274.30.20925
- Sandhu, K. S., and Neff, M. M. (2013). The *Arabidopsis* gene ATST4a in not a typical brassinosteroid catabolic gene. *Plant Signal. Behav.* 8, 10. doi: 10.4161/psb.26847
- Shimajima, M., Hoffmann-Benning, S., Garavito, R. M., and Benning, C. (2005). Ferredoxin-dependent glutamate synthase moonlights in plant sulfolipid biosynthesis by forming a complex with SQD1. *Arch. Biochem. Biophys.* 436, 206–214. doi: 10.1016/j.abb.2005.02.005
- Smith, D. W., Johnson, K. A., Bingman, C. A., Aceti, D. J., Blommel, P. G., Wrobel, R. L., et al. (2004). Crystal structure of At2g03760, a putative steroid sulfotransferase from *Arabidopsis thaliana*. *Proteins* 57, 854–857. doi: 10.1002/prot.20258
- Sønderby, I. E., Burrow, M., Rowe, H. C., Kliebenstein, D. J., and Halkier, B. A. (2010). A complex interplay of three R2R3 MYB transcription factors determines the profile of aliphatic glucosinolates in *Arabidopsis*. *Plant Physiol.* 153, 348–363. doi: 10.1104/pp.109.149286
- Sønderby, I. E., Hansen, B. G., Bjarnholt, N., Ticconi, C., Halkier, B. A., and Kliebenstein, D. J. (2007). A system biology approach identifies a R2R3 MYB gene subfamily with distinct and overlapping functions in regulation of aliphatic glucosinolates. *PLoS ONE* 2:e1322. doi: 10.1371/journal.pone.0001322
- Song, C. M., Lim, S. J., and Tong, J. C. (2009). Recent advances in computer-aided drug design. *Brief. Bioinf.* 10, 579–591. doi: 10.1093/bib/bbp023
- Vakiani, E., Luz, J. G., and Buck, J. (1998). Substrate specificity and kinetic mechanism of the insect sulfotransferase, retinol dehydratase. *J. Biol. Chem.* 273, 35381–35387. doi: 10.1074/jbc.273.52.35381
- Varin, L., Chamberland, H., Lafontaine, J. G., and Richard, M. (1997a). 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. doi: 10.1046/j.1365-3113X.1997.12040831.x
- Varin, L., Marsolais, F., Richard, M., and Rouleau, M. (1997b). Biochemistry and molecular sulfotransferases biology of plant sulfotransferases. *FASEB J.* 11, 517–525.
- Varin, L., Deluca, V., Ibrahim, R. K., and Brisson, N. (1992). Molecular characterization of two plant flavonol sulfotransferases. *Proc. Natl. Acad. Sci. U.S.A.* 89, 1286–1290. doi: 10.1073/pnas.89.4.1286
- Varin, L., and Ibrahim, R. K. (1992). Novel flavonol 3-sulfotransferase. Purification, kinetic properties, and partial amino acid sequence. *J. Biol. Chem.* 267, 1858–1863.
- Varin, L., and Spertini, D. (2003). Desulfoglucosinolate sulfotransferases, sequences coding the same and uses thereof for modulating glucosinolate biosynthesis in plants. Patent WO 2003010318.
- Wang, H., Wu, J., Sun, S., Liu, B., Cheng, F., Sun, R., et al. (2011). Glucosinolate biosynthetic genes in *Brassica rapa*. *Gene* 487, 135–142. doi: 10.1016/j.gene.2011.07.021
- Wang, T., Cook, I., Falany, C. N., and Leyh, T. S. (2014). Paradigms of sulfotransferase catalysis – the mechanism of SULT2A1. *J. Biol. Chem.* 289, 26474–2680. doi: 10.1074/jbc.M114.573501
- Weinshilboum, R. M., and Otterness, D. M. (1994). "Sulfotransferase enzymes," in *Handbook of Experimental Pharmacology*, ed. Kauffman (Berlin: Springer-Verlag), 45–78.
- Wittstock, U., and Halkier, B. A. (2002). Glucosinolate research in the *Arabidopsis* era. *Trends Plant Sci.* 7, 263–270. doi: 10.1016/S1360-1385(02)02273-2
- Yamashino, T., Kitayama, M., and Mizuno, T. (2013). Transcription of ST2A encoding a sulfotransferase family protein that is involved in jasmonic acid metabolism is controlled according to the circadian clock and PIF4/PIF5-mediated external coincidence mechanism in *Arabidopsis thaliana*. *Biosci. Biotechnol. Biochem.* 77, 2454–2460. doi: 10.1271/bbb.130559
- Zang, Y.-X., Kim, H. U., Kim, J. A., Lim, M.-H., Jin, M., Lee, S. C., et al. (2009). Genome-wide identification of glucosinolate synthesis genes in *Brassica rapa*. *FEBS J.* 276, 3559–3574. doi: 10.1111/j.1742-4658.2009.07076.x
- Zhai, G., Lehmler, H.-J., and Schnoor, J. L. (2013). Sulfate metabolites of 4-monochlorobiphenyl in whole poplar plants. *Environ. Sci. Technol.* 47, 557–562. doi: 10.1021/es303807f
- Zhang, H., Varmalova, O., Vargas, F. M., Falany, C. N., and Leyh, T. S. (1998). Sulfuryl transfer: the catalytic mechanism of human estrogen sulfotransferase. *J. Biol. Chem.* 273, 10888–10892. doi: 10.1074/jbc.273.18.10888

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 August 2014; accepted: 28 September 2014; published online: 16 October 2014.

Citation: Hirschmann F, Krause F and Papenbrock J (2014) The multi-protein family of sulfotransferases in plants: composition, occurrence, substrate specificity, and functions. *Front. Plant Sci.* 5:556. doi: 10.3389/fpls.2014.00556

This article was submitted to *Plant Physiology*, a section of the journal *Frontiers in Plant Science*.

Copyright © 2014 Hirschmann, Krause and Papenbrock. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

CHAPTER 3

The fusion of genomes leads to more options: A comparative investigation on the desulfo-glucosinolate sulfotransferases of *Brassica napus* and homologous proteins of *Arabidopsis thaliana*

Hirschmann, F. and Papenbrock, J. (2015). The fusion of genomes leads to more options: A comparative investigation on the desulfo-glucosinolate sulfotransferases of *Brassica napus* and homologous proteins of *Arabidopsis thaliana*. *Plant Physiology and Biochemistry* 91, 10-19. doi: 10.1016/j.plaphy.2015.03.009



The fusion of genomes leads to more options: A comparative investigation on the desulfo-glucosinolate sulfotransferases of *Brassica napus* and homologous proteins of *Arabidopsis thaliana*



Felix Hirschmann, Jutta Papenbrock*

Institute of Botany, Leibniz University Hannover, Herrenhäuserstr. 2, D-30419 Hannover, Germany

ARTICLE INFO

Article history:

Received 3 March 2015

Accepted 25 March 2015

Available online 26 March 2015

Keywords:

Brassica napus

Copy number variation

Glucosinolate

Sulfotransferase

ABSTRACT

Sulfotransferases (SOTs) (EC 2.8.2.-) play a crucial role in the glucosinolate (GI) biosynthesis, by catalyzing the final step of the core glucosinolate formation. In *Arabidopsis thaliana* the three desulfo (ds)-GI SOTs AtSOT16, AtSOT17 and AtSOT18 were previously characterized, showing different affinities to ds-GIs. But can the knowledge about these SOTs be generally transferred to other GI-synthesizing plants? It was investigated how many SOTs are present in the economically relevant crop plant *Brassica napus* L., and if it is possible to predict their characteristics by sequence analysis. The recently sequenced *B. napus* is a hybrid of *Brassica rapa* and *Brassica oleracea*. By database research, 71 putative functional BnSOT family members were identified and at least eleven of those are putative ds-GI SOTs. Besides the homologs of AtSOT16 – 18, phylogenetic analyses revealed new subfamilies of ds-GI SOTs, which are not present in *A. thaliana*. Three of the *B. napus* ds-GI SOT proteins were expressed and purified, and characterized by determining the substrate affinities to different ds-GIs. Two of them, BnSOT16-a and BnSOT16-b, showed a significantly higher affinity to an indolic ds-GI, similarly to AtSOT16. Additionally, BnSOT17-a was characterized and showed a higher affinity to long chained aliphatic GIs, similarly to AtSOT17. Identification of homologs to AtSOT18 was less reliable, because putative SOT18 sequences are more heterogeneous and confirmation of similar characteristics was not possible.

© 2015 Elsevier Masson SAS. All rights reserved.

1. Introduction

Members of the sulfotransferase (SOT) (EC 2.8.2.-) protein family catalyze the transfer of a sulfate group from the co-substrate 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to a hydroxyl group of different kinds of substrates (Lyon and Jakoby, 1980). In almost all investigated organisms SOTs are found. Most plant SOTs can be identified by one of the three Pfam motifs, *Sulfotransfer_1* (PF00685), *Sulfotransfer_2* (PF03567), and *Sulfotransfer_3* (PF13469), and have four highly conserved regions (Region I–IV). Region I, II and IV are suggested to be involved in PAPS binding (Weinshilboum et al., 1997; Marsolais and Varin, 1995). A highly

conserved histidine at the beginning of Region II is responsible for proton acceptance during sulfonyl transfer (Kakuta et al., 1998). The singly known exception is AtTPST, which is only associated to the SOT family from *Arabidopsis thaliana* (L.) Heynh by function and not by sequence (Komori et al., 2009).

Hirschmann et al. (2014) reported 22 genes in *A. thaliana* that are likely to encode SOTs. For ten of those the used substrates have been determined (Hirschmann et al., 2014). Eighteen SOTs (AtSOT1 – 18) were previously divided in seven groups and share about 51% sequence identity on amino acid level (Klein and Papenbrock, 2009). Except for AtSOT3, 4 and 10, none of those contain introns. Several phylogenetic attempts, trying to order and to predict plant SOTs by substrate specificity, based on sequence identity led to unsatisfying results (Dantas-Santos et al., 2012; Hashiguchi et al., 2013; Klein and Papenbrock, 2004, 2008). Therefore the prediction of SOT substrates by sequence analysis seems to be not reliable and has to be confirmed by enzymatical assays.

A subgroup (EC 2.8.2.24) of *A. thaliana* SOTs, AtSOT16 to AtSOT18, uses different desulfo- (ds) glucosinolates (GI) as substrates. Glucosinolates are nitrogen and sulfate containing

Abbreviations: 2PE, 2-phenylethyl GI; 3MTP, 3-methylthiopropyl GI; 4MTB, 4-methylthiobutyl GI; 5MTP, 5-methylthiopentyl GI; DFCI, Dana–Farber Cancer Institute Gene Index; ds, desulfo; EST, expressed sequence tags; GI, glucosinolate; I3M, indol-3-yl-methyl GI; IPTG, isopropyl-β-D-thiogalactopyranosid; LB, lysogeny broth; PAPS, 3'-phosphoadenosine 5'-phosphosulfate; SOT, sulfotransferase.

* Corresponding author.

E-mail address: Jutta.Papenbrock@botanik.uni-hannover.de (J. Papenbrock).

secondary metabolites, which are largely found in the order Brassicales. The Gl core structure consists of a β -D-glucopyranose residue linked via a sulfur atom to a (Z)-N-hydroximino sulfate ester and a variable rest, derived from one of eight amino acids. The latter allows the classification of GlS (Fahey et al., 2001). There are over 200 different described GlS with a wide variety of properties (Clarke, 2010). It was suggested that their major function is plant defense against pathogens and herbivores (Chew, 1988; Louda and Mole, 1991; Wittstock et al., 2003). For humans and animals Gl breakdown products were reported to have positive and negative effects. Goitrine, found in several *Brassica* species, for example, was reported to cause goiter when fed to cattle, while indolic GlS, found in many genera within the Brassicales, show anti-carcinogenic properties (Agerbirk et al., 2009).

The last step of core Gl biosynthesis is catalyzed by AtSOT16 to AtSOT18. Core GlS are the precursors for further secondary modifications, which lead to the wide range of GlS. Even though SOTs can use GlS with secondary modifications (Klein et al., 2006), it is assumed that *in planta* ds-core GlS are mainly used as substrates (Sønderby et al., 2010). Characterization of AtSOT16 to 18 revealed that AtSOT16 prefers indolic ds-GlS, such as indol-3-yl-methyl Gl (I3M), as substrate, followed by aliphatic ds-GlS, such as ds-3-methylthiopropyl Gl (3MTP) and ds-4-methylthiobutyl Gl (4MTB). The preferences of AtSOT17 and 18 are long chained aliphatic methylthio GlS, such as ds-7-methylthioheptyl Gl and ds-8-methylthiooctyl Gl (Klein et al., 2006; Klein and Papenbrock, 2009; Piotrowski et al., 2004). Detailed analysis of AtSOT18 from different ecotypes showed that the exchange of only one or two amino acids can lead to different affinities to the substrate and PAPS (Klein et al., 2006; Klein and Papenbrock, 2009; Luczak et al., 2013). The questions remain, whether AtSOT16 to 18 are specific for *A. thaliana*, or if there are homologous SOTs in other Gl-synthesizing plants. And is it possible to predict the respective AtSOT homolog by sequence analysis?

Recently sequenced *Brassica napus* is the closest related major economic crop plant to *A. thaliana*; therefore, it was chosen for comparison of presence and characteristics of ds-Gl SOTs. So far, four SOT isoforms (BNST1 - 4) were isolated from *B. napus*, and two of those were characterized. They use brassinosteroids as substrates, thus abolishing the biological activity of 24-epibrassinolide (Marsolais et al., 2000; Rouleau et al., 1999). *B. napus* is an allotetraploid species formed by the hybridization of *Brassica rapa* L. and *Brassica oleracea* L. The genome consists of 38 chromosomes: 20 derived from *B. rapa* and 18 from *B. oleracea* (Nagahara, 1935).

In the formation of *B. napus*, the progenitor *A. thaliana* has undergone a whole genome triplication approximately 20–40 million years ago (Parkin et al., 2002) followed by the allopolyploidization event less than 12,500 years ago (Chalhoub et al., 2014). This leads to the conclusion that each gene in *B. napus* is present in up to six homologs (Parkin et al., 2010). Random extensive gene loss and genetic drift after genome triplication (Cheung et al., 2009; Town et al., 2006; Yang et al., 2006) as well as homologs variation were typically observed after polyploidization (Gaeta and Chris Pires, 2010; Quijada et al., 2006) and affected orthologous gene copy number in both subgenomes. Gene expression investigations showed that in average 4.4 functional gene copies are present in *B. napus* (Parkin et al., 2010).

In the 1970s, breeding programs developed so-called 00-varieties of *B. napus*, also known as canola. The oil of canola contains less than 2% erucic acid and less than 30 $\mu\text{mol g}^{-1}$ of GlS in the seeds (Canola Council of Canada (1990)). Glucosinolates are the major secondary metabolites in *B. napus* and 80%–92% are aliphatic (Velasco et al., 2008). Concentration in the seed is 3.8–7.1 times as high as concentration in the leaf, depending on the crop. In 00 varieties the leaf Gl content is the same as in conventional *B. napus*

(Mithen, 1992), but the seed content is reduced by 80–90 % (Fieldsend and Milford, 1994). Only twelve different GlS were detected in *B. napus* (Velasco et al., 2008). In comparison, 34 GlS were found in certain genotypes of *A. thaliana*, which were mainly methionine-derived (Kliebenstein et al., 2001). Several studies showed that the Gl content in *B. napus* increases as a stress response (Bodnaryk, 1992, 1994; Doughty et al., 1991, 1995; Hopkins et al., 2009), which is also the case in *A. thaliana* (Falk et al., 2007; Wittstock et al., 2003). A comparative genomics study of *B. rapa* with *A. thaliana* identified 56 putative biosynthetic and regulator genes (Zang et al., 2009) involved in Gl biosynthesis. Twelve were putative ds-Gl SOTs. Their expression was depending on the tissue type and partially on environmental conditions. It was concluded that the biosynthetic Gl pathway in *B. rapa* is more complex than in *A. thaliana* (Zang et al., 2009).

The reasons for the different Gl patterns in *A. thaliana* and *B. napus* remain unknown. Polyploidization is considered a major restriction in knowledge transfer from *A. thaliana*. What are the effects of the allotetraploidy and the breeding programs of *B. napus*? Does the biosynthetic Gl pathway in *B. napus* differ from *A. thaliana*? How is it influenced by SOTs? In order to address these questions, proteins of the *B. napus* SOT family have been identified including putative ds-Gl SOT homologs. The sequences were analyzed to gain new insights into the ds-Gl SOTs present in *B. napus* and five were recombinantly expressed and enzymatically characterized. Two BnSOTs showed no activity at all, while the active homologs showed equal substrate affinities as their *A. thaliana* counterparts.

2. Material and methods

2.1. Plant material

Seeds from winter rape (*B. napus* L.) genotype 4 (Burandt et al., 2001) were germinated and grown in the greenhouse at 22 °C–24 °C. Young leaves were harvested and stored at –70 °C.

2.2. DNA cloning

The SOT nomenclature was used according to Klein and Papenbrock (2004). The BnSOT sequences were amplified in the following way: 5 U DreamTaq™ polymerase (Thermo Fisher Scientific™, Waltham, USA); 1 μg template DNA; 2 nM primer forward/reverse; 1 mM dNTP-Mix; DreamTaq™ polymerase buffer using the primers listed in Table 1. The PCR fragments were cloned into the pGEM-T vector (Promega, Madison, USA) and sent for sequencing from both directions with T7 and SP6 primers (GATC Biotech, Konstanz, Germany). For expression of recombinant BnSOT proteins the PCR amplicons identical to the consensus sequence were cloned into the pQE-30 expression vector containing an N-terminal in frame 6 \times His-tag (Qiagen, Hilden, Germany).

2.3. Expression and purification of AtSOT proteins

The SOT proteins were expressed in *Escherichia coli* XL1-Blue. A 50 mL overnight culture was used for the inoculation of 200 mL lysogeny broth (LB) medium (bacto tryptone 1% w/v, yeast extract 0.5% w/v, NaCl 1% w/v, pH 7.0 using NaOH). *E. coli* was grown at 37 °C up to an OD₆₀₀ of 0.6–0.8 followed by induction of the T5-promotor with 1 mM isopropyl- β -D-thiogalactopyranoside (IPTG) and incubation at 37 °C for 3 h. Expression of BnSOT18-a and BnSOTa8-b was conducted with several variations of the conditions: expression temperature ranged between 30 °C and 37 °C, time of expression between 1 h and 24 h, 0.5–1 mM IPTG was used for induction, *E. coli* XL1-Blue and BL21-CodonPlus (DE3)-RIL

Table 1
Primers for the amplification of BnSOTs.

SOT	Sequence (5' → 3')	Additional restriction site	T _A (°C)
BnSOT16-a	GGTACCGAACCAACCACGACCCAG	<i>KpnI</i>	60
BnSOT16-c	CCCGGGCTGATCATGTCAAGCAAGCC	<i>Cfr9I (XmaI)</i>	
BnSOT16-b	GGTACCGAACCAACCACGACCCAA	<i>KpnI</i>	60
	CCCGGGCTGAATCATGTTGAAGCAAGCC	<i>Cfr9I (XmaI)</i>	
BnSOT17-a	GGATCCGAATCCAAATCCGAAAACGA	<i>BamHI</i>	60
BnSOT17-b	CCCGGGTGGTGTGAAGCAAGAAAGC	<i>Cfr9I (XmaI)</i>	
BnSOT18-a	GGATCCTCTCCAATCTTCTCTGTATCT	<i>BamHI</i>	60
	AAGCTTTTCACCGTTTCAAGCAAAC	<i>HindIII</i>	
BnSOT18-b	GGATCCGAATCAGAACCCTAACCGATAC	<i>BamHI</i>	58.8
	CTGCAGTAGAATCGAACTGACAGATGGAG	<i>PstI</i>	

strains (Agilent Technologies Deutschland GmbH, Böblingen, Germany) were used. The induction was terminated by centrifugation at $4000 \times g$ at 4 °C. Pellets were stored at –20 °C before protein extraction. The pellets were dissolved in 6 mL lysis buffer (20 mM sodium phosphate, 0.5 M NaCl, 20 mM imidazole, (pH 7.4 using HCl) and 1 mg mL⁻¹ lysozyme. Lysis itself was carried out by sonification. Afterwards the solution was centrifuged for 30 min at $32,000 \times g$ and the supernatant transferred to fresh tubes. The BnSOT16-a, BnSOT16-b, and BnSOT17-a proteins could be successfully expressed as soluble proteins in *E. coli* in large amounts although a part of the recombinant proteins remained insoluble in the pellet. The total amount of recombinant proteins after 3 h induction with IPTG was determined between 12 and 27% as a percentage of the total protein.

The purification of recombinant BnSOT16-a, BnSOT16-b, and BnSOT17-a proteins by affinity chromatography was performed according to Luczak et al. (2013). 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 proteins, which was considered in the calculation of the final protein concentration, was analyzed by SDS-polyacrylamide gel electrophoresis 3 h after induction of expression, and was calculated with the program Aida Image Analyzer 3.2 (Raytest, Straubenhardt, Germany). Total protein contents were determined by the method of Bradford (1976) with bovine serum albumin (Roth, Karlsruhe, Germany) as a protein standard.

2.4. Preparation of substrates

The ds forms of the core GIs were prepared as described by Graser et al. (2001). The following GIs were used in the experiments in their ds forms: 3-methylthiopropyl GI (3MTP, C08412) from *Erysimum pumillum*, 4-methylthiobutyl GI (4MTB, C08409) from *Eruca sativa*, 5-methylthiopentyl GI (5MTP, C08401) from *A. thaliana*, indol-3-ylmethyl GI (13M, C05837) from *Isatis tinctoria* and 2-phenylethylglucosinolate (2 PE, C08417) from *Nasturtium officinale*. In brackets the KEGG COMPOUND numbers are given (<http://www.genome.jp/dbget/>) and the structures of all GIs are also shown in Reichelt et al. (2002).

2.5. Enzyme activity measurements and HPLC analysis

Enzyme activity measurements and HPLC analysis was performed as described (Luczak et al., 2013). For the determination of the apparent K_m values of ds-3MTP and ds-4MTB, ten different substrate concentrations were used (20 μM–220 μM). The PAPS concentration was kept constant at 60 μM. The 150 μL assay for the determination of apparent K_m values contained 80 mM Tris/HCl, pH 9.0, 9.2 mM MgCl₂, ds-3MTP or ds-4MTB (20 μM–220 μM), 0.5–12 μg purified protein, and 60 μM PAPS. The reactions were started by the addition of PAPS, incubated for 20 min at 37 °C, and

stopped by incubation at 95 °C for 10 min. The formation of the respective sulfated product was analyzed at 229 nm using a UV detector connected to an HPLC system (Knauer, Berlin, Germany) with a SUPELCOSIL™ LC-18-DB HPLC Column (Sigma–Aldrich, St. Louis, USA) as described in Klein et al. (2006). The apparent K_m values were determined using the Enzyme Kinetics module in Sigma Plot 12.5 (Systat Software Inc., San Jose, USA). Tukey-HSD test ($\alpha = 0.05$) for identification of substrate specificities was performed in R 2.15.1.

2.6. Database screen and bioinformatic analysis

The databases NCBI (<http://www.ncbi.nlm.nih.gov/>), accessed 2nd January 2015 and The Dana–Farber Cancer Institute Gene Index Project (DFCI) (<http://compbio.dfci.harvard.edu/tgi/>), accessed 18th February 2015) were used for BLAST searches. The used query sequences are listed in Table 2 (AtSOT2 was excluded, because it is considered a pseudogene). *B. oleracea* sequences were additionally searched in BolBase (<http://ocri-genomics.org/bolbase/>), accessed 4th February 2015.

The results of the bidirectional sequencing were assembled by Clone Manager 9© (Scientific & Educational Software, Cary, USA) using the application “simple sequence assembly” with an overlap score of 200 bp. Alignments (MUSCLE) and phylogenetic trees were constructed with MEGA6. The best model was calculated by MEGA6. Afterwards Maximum Likelihood trees were created. For the test of reliability for tree branches the “Bootstrap method” application was chosen with 1000 repetitions. Sequence Identity Matrix was calculated in BioEdit 7.2.5. Primers were designed with the online tool Oligo Perfect™ Designer (<http://tools.invitrogen.com/content.cfm?pageid=9716>).

For the prediction of protein localization, the programs TargetP (<http://www.cbs.dtu.dk/services/TargetP/>), SignalP (<http://www.cbs.dtu.dk/services/SignalP/>), Mitoprot II (http://ihg.gsf.de/ihg/mito_prot.html), ChloroP (<http://www.cbs.dtu.dk/services/ChloroP/>) and ProtComp (<http://linux1.softberry.com/berry.phtml?topic=protcomppl&group=programs&subgroup=proloc>) were used.

3. Results

3.1. BnSOT identification and amplification

In order to identify and isolate novel ds-GI SOTs from *B. napus*, BLASTp searches were performed within several online databases. Twenty AtSOTs were used as query sequences (Table 2). In total, 112 non-redundant amino acid sequences with a SOT PFAM domain were identified (Table S1). An additional sequence (BnSOT18-a) was amplified by PCR with highest sequence identity to AtSOT18 (71%). It was assumed that a functional SOT must contain all four typically conserved regions to be enzymatically functional, with exception of ATPST homologs. Only 67 sequences contained, at least partially,

Table 2
All *B. napus*, *B. rapa* and *B. oleracea* ds-GI SOTs and previously identified *A. thaliana* SOTs (Klein and Papenbrock, 2009; Komori et al., 2009) with accession numbers.

<i>A. thaliana</i>		<i>B. napus</i>		<i>B. rapa</i>		<i>B. oleracea</i>	
Name	Accession	Name	Accession	Name	Accession	Name	Accession
AtSOT1	NP_199182	BnSOT16-a	KP055788	BrSOT16-a	ACH41750	BoSOT16-a	Bol039395
AtSOT2	NP_190689	BnSOT16-b	KP055789	BrSOT16-b	XP_009128026	BoSOT16-b	Bol026200
AtSOT3	NP_194358	BnSOT16-c	KP055790	BrSOT16-c	ACH41745	BoSOT17-a	Bol030757
AtSOT4	NP_180325	BnSOT16-d	CDY45856	BrSOT16-d	ACR10265	BoSOT18-a	Bol026202
AtSOT5	NP_190093	BnSOT17-a	KP055791	BrSOT17-a	ACR10273	BoSOT18-b	Bol026201
AtSOT6	NP_190094	BnSOT17-b	KP055792	BrSOT18-a	XP_009106065		
AtSOT7	NP_174139	BnSOT18-a	KP055793	BrSOT18-b	XP_009106068		
AtSOT8	NP_172799	BnSOT18-b	KP055794	BrST5b-3	XP_009113327		
AtSOT9	NP_172800	BnSOT18-c	CDX96514	BrST5b-4	Pseudogene		
AtSOT10	NP_179098	BnSOT18-d	CDX73049	BrST5b-5	ACR10269		
AtSOT11	NP_565305	BnSOT18-e	CDX96512	BrST5b-6	Pseudogene		
AtSOT12	NP_178471	BnSOT18-f	CDX73045	BrST5b-7	XP_009113161		
AtSOT13	NP_178472			BrST5b-8	XP_009113160		
AtSOT14	NP_196317			BrST5b-9	ACR10272		
AtSOT15	NP_568177						
AtSOT16	NP_177550						
AtSOT17	NP_173294						
AtSOT18	NP_177549						
AtSOT19	NP_190631						
AtSOT20	NP_179175						
AtSOT21	NP_195168						
TPST	NP_563804						

all four conserved regions. By BLASTp searches, four AtTPST homologs were identified, leading to a total of 71 BnSOTs. Therefore, 3.9 times more putative SOTs were identified in *B. napus* than in *A. thaliana*.

Alignment of the 71 identified *B. napus* SOTs with 18 *A. thaliana* SOTs (AtSOT2 and AtSOT19–21 were excluded, because they miss regions I – IV) suggested the existence of up to 23 ds-GI SOT candidates (Fig. 1). Four sequences clustered with AtSOT16 (BnSOT16-a – BnSOT16-d), two with BnSOT17 (BnSOT17-a and BnSOT17-b), and nine with AtSOT18. Another set of eight sequences were in close proximity to the ds-GI AtSOTs, but did not cluster directly with any of those. The 23 candidate sequences were further aligned with the 17 AtSOTs. The resulting phylogenetic tree allowed a more detailed assignment of the BnSOT homologs (Fig. S1). While the same sequences mentioned before clustered with AtSOT16 and AtSOT17, only five sequences formed a cluster with AtSOT18 (BnSOT18-b to BnSOT18-f). Comparison in a sequence identity matrix showed that out of the twelve remaining candidate sequences, including AtSOT18-a, that could not surely be related to a specific ds-GI AtSOT, had the highest identity with AtSOT18 (Table S2). Therefore, they are referred to as SOT18-like. The only exception is CDX93182, which is closer related to AtSOT7. The matrix also showed that SOT18 homologs have less similarity to their *A. thaliana* counterparts (74%–84%) than the SOT16 and SOT17 homologs (89%–92% and 86%, respectively). The SOT18-likes share even less sequence identities with AtSOT18 (57%–75%). In phylogenetic analyses they formed two individual clusters and one single branch, hence they could be considered as a new subgroup of ds-GI SOTs, which is not present in *A. thaliana* (Fig. S1). In *B. rapa* and *B. oleracea* seven and five, respectively, ds-GI SOTs were identified by BLASTp searches in NCBI and phylogenetic analyses (Table S1). They were aligned to the 23 putative ds-GI BnSOTs, in order to gain more information about the ancestry of ds-GI SOTs in *B. napus* (Fig. S2). Non-redundant *B. rapa* ds-GI SOT sequences (BrST5b-3 to BrST5b-9), previously identified by Zang et al. (2009), were added to the alignment. For eight out of the eleven BnSOT16 – 18, an ancestor sequence from either *B. rapa* or *B. oleracea* could be identified. Five sequences derived each from *B. rapa* and from *B. oleracea*, only for BnSOT18-b no clear ancestor could be identified (Fig. 2). The SOT18-like sequences clustered on separate branches

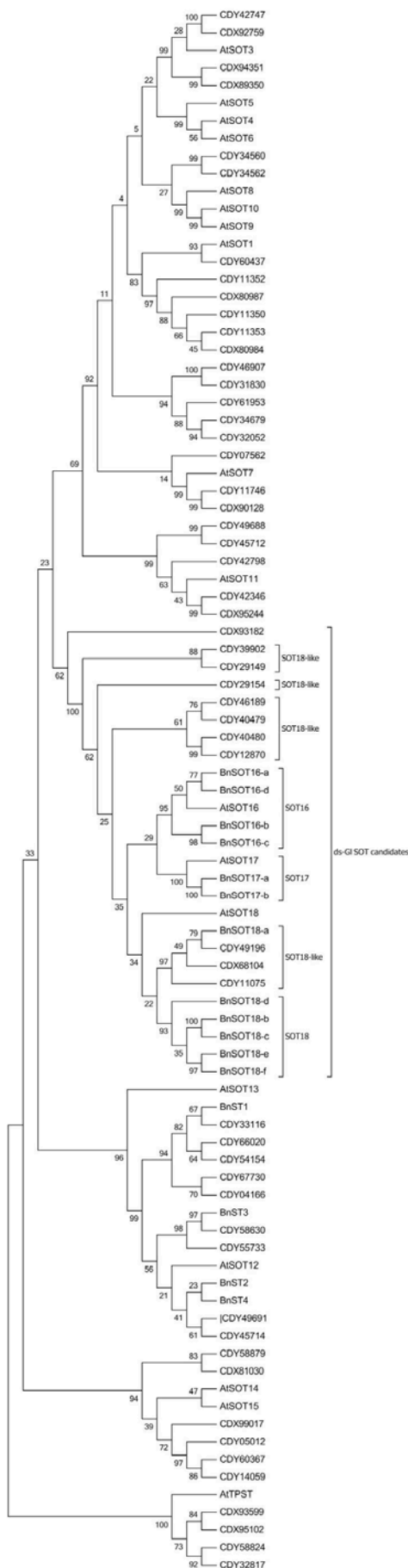
together with the added *B. rapa* sequences from Zang et al. (2009), but no homologous *B. oleracea* sequences could be identified (Fig. S2). Neither did additional BLASTp searches with the SOT18-like sequences as query result in the identification of any *B. oleracea* homologs.

In several alignments (data not shown), BnSOT18-a showed several non-aligned amino acids at the N-terminus and BnSOT18-b at the C-terminus. The online tools TargetP, SignalP, Mitoprot II, ChloroP, and ProtComp were used to screen for target peptide sequences, but all programs predicted localization in the cytoplasm (data not shown).

In summary, the phylogenetic analysis revealed that 11 out of the 23 candidate sequences encoded most likely an AtSOT16 – 18 homolog. Therefore, the gene copy number of ds-GI SOTs was 3.7. Parkin et al. (2010) discovered in a genome expression study that the average gene copy number in *B. napus* was with 4.4 slightly higher. Furthermore, 11 SOT candidate sequences showed high similarity to AtSOT18 and were classified as SOT18-like sequences. The candidate sequence CDX93182 had the highest identity to AtSOT7 and was therefore not considered as a putative ds-GI SOT sequence.

3.2. Amplification of ds-GI BnSOTs

Sequence amplification was conducted before the publication of the *B. napus* genome (Chalhoub et al., 2014). Sequences from already assembled ESTs from DFCI (TC180059, TC165797, TC202195, TC170591, TC206772) were used for the primer design. Due to high sequence similarities at the 5'- and 3'-ends of the sequences, several sequences were amplified by the same primer pair, leading to difficulties with the identification of the correct PCR fragments. Therefore, three to 16 PCRs were run in order to gain the correct nucleotide sequence, which were then used for the recombinant expression of the BnSOT proteins. At least two independent PCRs were performed per SOT to reach 100% identity. If it was not possible to get two identical amplicons the consensus sequence of at least three independent PCRs were assembled to identify the sequence. In this way BnSOT16-a to BnSOT16-c, BnSOT17-a and BnSOT17-b, and BnSOT18-a and BnSOT18-b could be amplified. Later, the sequences were verified in NCBI: The sequences of BnSOT16-a to BnSOT16-c,



BnSOT17-a and BnSOT18-a were confirmed in NCBI with 100% identity. The BnSOT17-a sequence differed in three amino acids. No identical sequence of BnSOT18-a was found in NCBI, but the sequence TC170591 in DFCI was identical to the PCR product. The differences between our amplified sequences and the database entries could be either explained by incomplete, respectively incorrect database entries, or because of differences in used *B. napus* cultivars. Chalhouh et al. (2014) fully sequenced cultivar Darmorbzh, while in this work winter rape genotype 4 (Burandt et al., 2001) was used. In the above described phylogenetic analyses the amplified genotype 4 sequences were used.

3.3. The BnSOTs analyzed have similar substrate specificities as their *A. thaliana* homologs

Enzyme assays with the purified BnSOT16-a, BnSOT16-b, and BnSOT17-a were performed. Purified BnSOT18-a and BnSOT18-b showed no activity either due to wrong folding or because they could not use the tested ds-GIs as substrates. Methionine-derived ds-3MTP, ds-4MTB, ds-5MTP, indolic ds-I3M and the aromatic ds-2PE were used as substrates. The substrates chosen for enzymatic assays were the core GIs of the GIs reported to be present in *B. napus* (Velasco et al., 2008). The respective core GIs were researched in online databases BioCyc (www.biocyc.org, Karp et al., 2005) and KEGG (<http://www.genome.jp/kegg/>, Kanehisa and Goto, 2000). The enzymes BnSOT16-a, BnSOT16-b, and BnSOT17-a were active with all offered core GIs (Fig. 3). BnSOT16-a and BnSOT16-b showed a significantly higher activity with the indolic I3M compared to 3MTP, 4MTB, 5MTP, and 2PE. BnSOT16-a had generally a higher activity with all substrates than BnSOT16-b. After analyzing BnSOT16-a and BnSOT16-b, it was assumed that BnSOT16-c and BnSOT16-d would show similar substrate specificities and therefore they were not analyzed. The same was applied for BnSOT17-b after analyzing BnSOT17-a. BnSOT17-a preferred long-chained methionine-derived ds-GIs 4MTB and 5MTP. These results correspond with the substrate specificities to the AtSOT homologs (Klein and Papenbrock, 2009). Therefore, knowledge about AtSOT16 and AtSOT17 could generally be transferred to *Brassica* species. Furthermore, the results showed that the identification of SOT16 and SOT17 homologs by sequence analysis was possible. This could not be shown for SOT18 homologs, as no activity could be detected for BnSOT18-a and BnSOT18-b. Furthermore the identification was more problematic, as described in chapter 3.1.

K_m and V_{max} values were exemplary determined for the substrates ds-3MTP and ds-4MTB. These substrates were chosen for a better comparison, since the respective AtSOT values were determined previously (Klein and Papenbrock, 2009; Luczak et al., 2013; Piotrowski et al., 2004). The determined K_m values ranged from $90 \pm 43 \mu\text{M}$ to $281 \pm 193 \mu\text{M}$, while the V_{max} values ranged from $58 \pm 18 \text{ pkat mg}^{-1}$ to $638 \pm 291 \text{ pkat mg}^{-1}$ (Table 3). The lower K_m value for BnSOT17-a with ds-4MTB ($90 \mu\text{M}$) compared to ds-3MTP ($271 \mu\text{M}$) and the higher V_{max} value (ds-4MTB: 595 pkat mg^{-1} , ds-3MTP 174 pkat mg^{-1}) supported the results of the substrate specificity assays (Fig. 3). Thus BnSOT17-a showed a higher affinity to long-chain methionine-derived GIs. The V_{max} and K_m of AtSOT17 showed similar characteristics as BnSOT17-a. Comparison of BnSOT16-a and BnSOT16-b with AtSOT16 showed different tendencies. AtSOT16 had a lower V_{max} value for ds-4MTB and a slightly

Fig. 1. Phylogenetic tree of all putative functional *B. napus* and *A. thaliana* SOTs, after MUSCLE alignment to amino acid level, based on the Maximum likelihood method and Jones–Taylor–Thornton model with gamma distribution, including 1000 bootstrap values. The tree is unrooted and is displaying topology only.

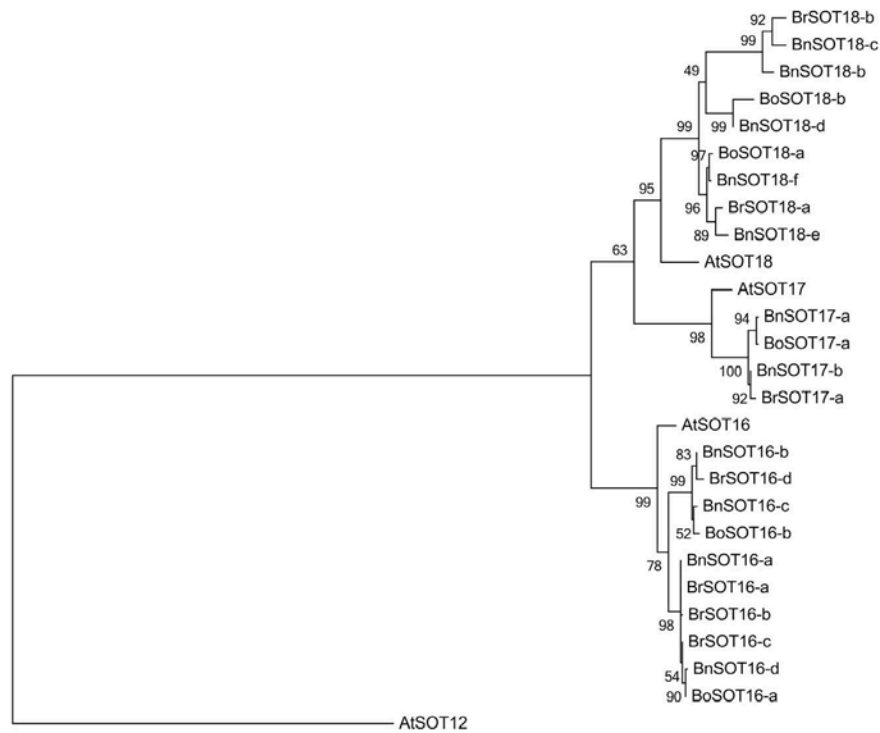


Fig. 2. Phylogenetic tree of all identified ds-Gl SOTs from *B. napus*, *A. thaliana*, *B. rapa* (including sequences identified by Zang et al., 2009), and *B. oleracea*, after MUSCLE alignment on amino acid level, based on the Maximum likelihood method and Jones–Taylor–Thornton model with gamma distribution and invariant sites, including 1000 bootstrap values. The tree is rooted on AtSOT12.

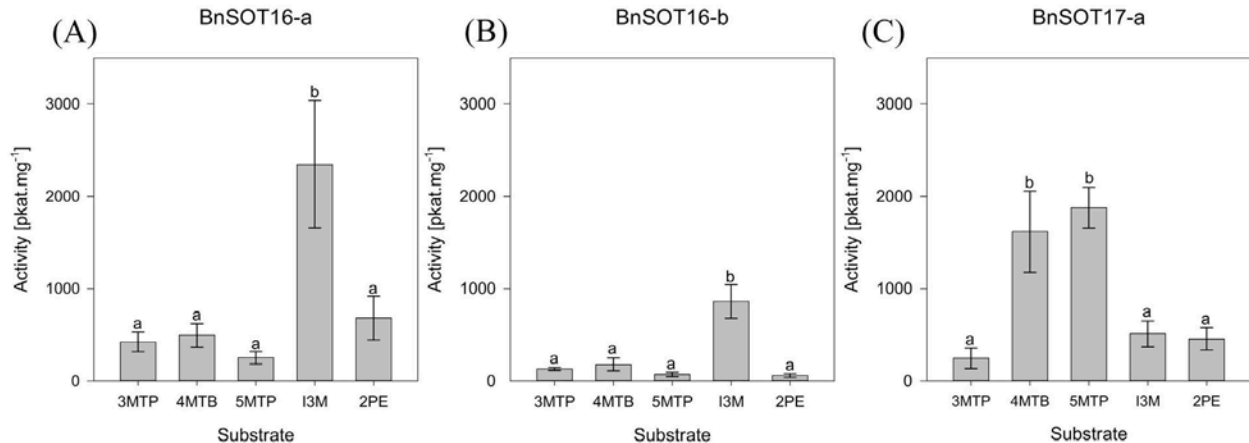


Fig. 3. Determination of substrate specificities of BnSOT proteins using different core ds-Gls as substrates. The 150 μ L assays contained 80 mM Tris/HCl, pH 9.0, 9.2 mM MgCl₂, 60 μ M of the respective ds-Gl substrates, between 0.5 and 12 μ g purified protein (amount determined in pre-tests to investigate enzyme activity always in the linear range), and 60 μ M PAPS. The reactions were started by the addition of PAPS, incubated for 20 min at 37 °C, and stopped by incubation at 95 °C for 10 min. The formation of the respective sulfated product was analyzed by HPLC at 229 nm. The specific activities are given in pkat mg⁻¹ protein. Shown are measurements of (A) BnSOT16-a, (B) BnSOT16-b, and (C) BnSOT17-a with core Gls. The standard deviation was calculated (n = 5–9). Groups of statistical significance (a,b) were determined after Tukey-HSD test ($\alpha = 0.05$). Abbreviations: 3MTP, ds-3-methylthiopropyl Gl; 4MTB, ds-4-methylthiobutyl Gl; 5MTP, ds-5-methylthiopentyl Gl; I3M, ds-indol-3-yl-methyl Gl; 2 PE, 2-phenylethyl Gl.

higher K_m value compared to ds-3MTP, indicating a higher affinity to ds-4MTB. In contrast, BnSOT16-a had a higher V_{max} value for ds-4MTB compared to ds-3MTP and a lower K_m value for ds-4MTB compared to ds-3MTP. Also noticeable is that the V_{max} value of BnSOT16-b was more than tenfold smaller than the respective value of BnSOT16-a.

4. Discussion

In previous work we identified and characterized AtSOT16 – 18 (Klein et al., 2006; Klein and Papenbrock, 2004, 2009). The results

demonstrated that all three SOTs use Gls as substrates, but have different affinities to certain Gls. Until now, only ds-Gl SOTs from *A. thaliana* were characterized. In this work, ds-Gl SOTs from *B. napus* were characterized, in order to investigate if characteristics of ds-Gl AtSOTs are specific for *A. thaliana* or if the knowledge about ds-Gl SOTs and the Gl biosynthetic pathway can be adapted to other plant species.

Marsolais et al. (2007) characterized *A. thaliana* brassinosteroid SOTs AtST1 and AtST4a. Comparison with previously characterized *B. napus* brassinosteroid SOTs BNST3 (Rouleau et al., 1999) and BNST4 (Marsolais et al., 2004) showed that AtST1 like BNST3 used

Table 3

K_m and V_{max} values of BnSOTs and AtSOTs for ds-3MTP and ds-4MTB. AtSOT values were taken from Klein and Papenbrock (2009). For determination of the K_m values, up to 10 data points were included (20–220 mM). PAPS concentration was constant at 60 μ M. Each determination was carried out twice.

SOT	K_m (μ M)		V_{max} (pkat mg ⁻¹ protein)	
	ds-3MTP	ds-4MTB	ds-3MTP	ds-4MTB
BnSOT16-a	113 \pm 36	280 \pm 193	378 \pm 58	638 \pm 291
BnSOT16-b	145 \pm 58	90 \pm 67	133 \pm 28	58 \pm 18
AtSOT16	70 \pm 0	80 \pm 0	623 \pm 67	496 \pm 15
BnSOT17-a	271 \pm 180	90 \pm 43	174 \pm 74	595 \pm 120
AtSOT17	88 \pm 16	65 \pm 7	439 \pm 214	575 \pm 16
AtSOT18	100 \pm 0	130 \pm 14	99 \pm 17	131 \pm 17

hydroxysteroids and estrogens as substrates, while BNST4 only uses hydroxysteroids. AtST4a, belonging to a different subfamily of SOTs (Piotrowski et al., 2004), showed no activity with hydroxysteroids or estrogens, but with biologically active end products of the brassinosteroid biosynthetic pathway. Accordingly, we speculated that the BnSOTs have similar characteristics as their AtSOT homologs.

Furthermore it was tested, if it is possible to predict the preferences to certain GIs by sequence analysis. It was previously demonstrated that AtSOT18 proteins from different *A. thaliana* ecotypes, which only differ in two amino acids, showed different substrate affinities (Klein et al., 2006; Klein and Papenbrock, 2009; Luczak et al., 2013). How reliable are predictions based on sequence alignments? Comparison of several previous phylogenetic analyses revealed that the general prediction of substrate specificities of SOTs is limited, but ds-GI SOTs always formed a distinct group (Hirschmann et al., 2014).

Recent publication of the complete *B. napus* genome (Chalhoub et al., 2014) enabled a detailed investigation of its complete SOT protein family. Database research revealed the presence of at least 67 functional SOTs (Fig. 1). It was possible to identify four homologs of AtSOT16, two of AtSOT17, and five of AtSOT18, thus revealing individual copy number variations. Additionally, 11 more sequences were identified that have high sequence identity to ds-GI AtSOTs (Table S2, Fig. S1, S2). Interestingly, phylogenetic analysis showed that these 11 proteins formed new ds-GI subfamilies, which are not present in *A. thaliana*. Due to genome triplication and allopolyploidization of *B. napus*, three homologs from each ancestor genome can theoretically be expected, but in average there are 4.4 functional gene copies present (Parkin et al., 2010). Variation of the gene copy number in *B. napus* was previously reported for several proteins, such as proline dehydrogenase (Faes et al., 2014), cytokinin receptors (Kuderova et al., 2015), and flowering-time regulatory genes (Schiessl et al., 2014). In the process of polyploidization, *B. napus* has undergone a high degree of genome and gene duplications and genome rearrangements (Zmienko et al., 2014), including homoeologous recombination, which leads to gene copy variation and presence/absence variation (Gaeta and Chris Pires, 2010; Udall et al., 2006). Furthermore the genetic changes of interspecific hybridization lead to an increase of transposal activity, which is a key part in adaptive plant evolution (Lisch, 2013). The high number of homologous genes leads to higher robustness against mutation and loss of function risks (Bekaert et al., 2011) and a better adaptation to the environment (Chen, 2010). Simultaneous transcription enables a faster and increased reaction capacity and a better fine tuning in physiological pathways (Schiessl et al., 2014).

In order to confirm the sequence based predictions, BnSOT16-a, BnSOT16-b, and BnSOT17-a were recombinantly expressed and characterized. The comparison of substrate affinity tests of

B. napus with *A. thaliana* (Fig. 3) revealed that the BnSOTs have similar substrate affinities as their *A. thaliana* counterparts. Kinetic values of BnSOT17-a and AtSOT17 behave similarly and therefore support the findings of the substrate affinity tests. The V_{max} and K_m values of AtSOT16 compared to the ones of BnSOT16-a and BnSOT16-b (Table 3) showed different tendencies in their preferences for 3MTP and 4MTB. But generally the substrate specificity determinations demonstrated similar substrate specificities, supporting the indications of the sequence analysis. Hence, this work demonstrated that it is principally possible to predict SOT16 and SOT17 homologs and their general substrate affinities, but not the kinetic parameters of the respective SOTs. Function of SOTs with similar substrate specificities, but different kinetics could be used for fine regulation of the GI pathway. That was also shown for the *A. thaliana* aspartate metabolism, where various protein isoforms with different kinetic parameters contribute in different ways to the flux and regulation of the pathway (Curien et al., 2009).

Investigations of SOT16 and SOT17 homologs led to explicit results, since they could be easily identified and characterized. Unfortunately, the recombinantly expressed BnSOT18-a and BnSOT18-b showed no activity *in vitro*, which could have several reasons: (1) the proteins might have been folded incorrectly in the *E. coli* expression system. This is a common problem in recombinant protein expression. In order to address this problem different bacteria strains, expression times and temperatures, media compositions and purifications protocols have been tried, but without success. Several highly related ds-GI SOTs were previously successfully expressed and characterized in our system without major difficulties. (2) BnSOT18-a and BnSOT18-b could be wrongly identified as ds-GI SOTs. The SOT18 homologs are more heterogeneous in their sequence composition (Table S2), leading to difficulties in identification. Homologs were identified based on phylogenetic trees, but it has to be kept in mind that these trees hardly ever completely reflect true relation. This is apparent when comparing the number of AtSOT18 homolog candidates in different phylogenetic analyses (eight in Fig. 1 and five in Fig. S2). These problems could explain a misidentification in between the three ds-GI AtSOTs, since the sequence identities are close (64%–75%), but the highest identities to the other AtSOTs are significantly lower (BnSOT18-a to AtSOT8: 36%; BnSOT18-b to AtSOT5: 39%). Therefore, it seems unlikely that BnSOT18-a and BnSOT18-b were wrongly identified as ds-GI SOTs. Furthermore it has to be kept in mind that the sequence of BnSOT18-a could only be verified in DFCI, but not in NCBI. This could be explained either by incompleteness in NCBI, or by sequence differences in the cultivars Darmor-bzh (NCBI) and the winter rape genotype 4 (used in this study), or because of amplification mistakes in the PCR. (3) BnSOT18-a and BnSOT18-b are not only inactive after recombinant expression *in vitro*, but also *in vivo*. After duplication, redundant genes often undergo neofunctionalization, subfunctionalization or pseudogenization (Conant and Wolfe, 2008). The high copy number of AtSOT18 homologs as well as the high number of sequences showing highest sequence identity to AtSOT18, but forming clusters apart from ds-GI AtSOTs (Table S2, Figs. 1 and S1) would support this hypothesis. Furthermore, products of whole genome duplications are more likely to show haplo-insufficiency than other duplicates (Wapinski et al., 2007). Since the whole genome duplication event happened only recently in *B. napus*, the coding genes might still be in the process of pseudogenization. A possibility of subfunctionalization could be the sulfation of already synthesized core GIs. It is generally agreed that ds-GI SOTs in *A. thaliana* use precursors of core GIs as substrates and catalyze the last step of their biosynthesis (Sønderby et al., 2010), but it was demonstrated that *in vitro* ds-GI SOTs can also use core GIs as substrates (Klein et al., 2006).

The new insights encourage speculations about the impact of the ds-GI SOTs in *B. napus*, *B. rapa* and *B. oleracea*. When comparing GI profiles, especially when taken from different studies, it has to be kept in mind that the GI content varies based on length of growing season, environmental conditions, and time of planting (Cartea et al., 2008; Charron et al., 2005; Rosa and Heaney, 1996). The total GI contents vary in the same range, though *B. napus* shows the smallest maximal GI content in leaves (*B. napus*: 14–24 $\mu\text{mol g}^{-1}$ DW, *B. rapa*: 7.5–74 $\mu\text{mol g}^{-1}$ DW, *B. oleracea*: 11–53 $\mu\text{mol g}^{-1}$ DW). Overall in all three species aliphatic GIs are predominant (Cartea et al., 2008; Padilla et al., 2007; Velasco et al., 2008). One reason for the high quantity of aliphatic GIs in *B. napus* (80%–92% are aliphatic) and quality (seven out of twelve GIs are aliphatic) could be the high number of SOT18 genes (Fig. 1), if the BnSOT18s and the similar candidates are assumed to be active *in vivo* with similar characteristics to AtSOT18, as SOT18s usually prefer methionine-derived ds-GIs (Klein and Papenbrock, 2009). While the gene inventory of GI biosynthetic genes is almost completely identified in *A. thaliana* (Sønderby et al., 2010), only little information about the GI biosynthesis in *B. napus* is available, though all ancestral GI catabolism genes were identified (Chalhoub et al., 2014). Zang et al. (2009) conducted a comparative genomics study with *B. rapa* and *A. thaliana* and concluded that the *B. rapa* GI pathway is collinear, but more complex, due to the high number of gene copies. Accordingly, it can be assumed that the pathway in *B. napus* might be similar, but even more complex. Furthermore, it has to be considered that agriculturally used *B. napus* was especially bred for low GI content (Abadi and Leckband, 2011). Hence, major, but yet unknown, changes on a cell physiological level can be expected. Chen et al. (2011) used a bioinformatical approach to identify genes, which could be crosslinked or be components of the GI pathway in *A. thaliana*. In total 330 genes were identified to have a high potential to be related to the GI pathway. A selection of 64 genes was used to construct a molecular network, which was then validated by the evaluation of microarray data of nine mutated candidate genes. Mutation led to changes in the expression profiles of known GI biosynthesis genes (Chen et al., 2011). These findings point out the complexity of this secondary pathway, in which SOTs are just one piece of a big puzzle.

Having this in mind, the question arises, whether ds-GI SOTs are a promising target for breeding in regard of GI content. In a associative transcriptomic study, Lu et al. (2014) identified SOT17 and SOT18 as two out of 24 most significant candidate genes that might be correlated with seed GI content. Klein and Papenbrock (2009) speculated that the GI composition could be changed by overexpressing specific SOTs. Overexpression of SOT17 and SOT18, for example, could lead to more nitriles and epithionitriles, due to hydrolysis of the respective GIs and thereby contribute to a better plant defense. An overexpression of SOT16 could lead to a higher indolic GI concentration, which hydrolysis products are assumed to have anti-carcinogenic properties (Brew et al., 2009). The relatively high number of SOT16s in *B. rapa* could lead to the assumption of a higher indolic GI concentration, but it was shown that in general *B. oleracea* have the highest indolic GI content (*B. rapa*: 1.58 $\mu\text{mol g}^{-1}$ DW, *B. oleracea*: 7.99 $\mu\text{mol g}^{-1}$ DW). Interestingly, Chinese cabbage *B. rapa* L. ssp. *pekinensis* was shown to have relatively high contents of indolic GIs (32.8%, Kim et al., 2010). GI biosynthesis is fueled by the co-substrates nicotinamide adenine dinucleotide phosphate, glutathione, uridine diphosphoglucose, and PAPS (Sønderby et al., 2010), at which PAPS provision is a limiting factor (Møldrup et al., 2011). Hence, limiting breeding efforts towards a single GI biosynthesis step might not be promising enough and other factors and key regulators, such as HAG1 (Harper et al., 2012), should be considered, too.

5. Conclusion

In *B. napus*, 71 putative functional SOTs were identified by database research. Phylogenetic analyses revealed copy number variations in the ds-GI SOT subfamily members, as four AtSOT16, two AtSOT17, and five AtSOT18 homologs were identified. Another eleven sequences, referred to as AtSOT18-like, showed high similarity to AtSOT18 (57%–75% sequence identity). *In vitro*, substrate specificity assays of BnSOT16-a, BnSOT16-b, and BnSOT17-a revealed that the *B. napus* homologs have similar substrate preferences as their *A. thaliana* counterparts. For the recombinantly expressed AtSOT18-a and AtSOT18-b no activity could be measured, either due to wrong folding, misidentification or loss-of-function mutation.

Contributions

Felix Hirschmann has conducted all experiments. Jutta Papenbrock has planned, organized and supervised all experiments.

Acknowledgment

We would like to thank Julia Volker, Hannover, and Dr. Michael Reichelt, MPI for Chemical Ecology, Jena, Germany, for the preparation of the substrates. The project was funded by the German Research Foundation (PA 764/10-1).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.plaphy.2015.03.009>.

References

- Abadi, A., Leckband, G., 2011. Rapeseed breeding for oil content, quality, and sustainability. *Eur. J. Lipid Sci. Tech.* 113, 1198–1206.
- Agerbirk, N., de Vos, M., Kim, J.H., Jander, G., 2009. Indole glucosinolate breakdown and its biological effects. *Phytochem. Rev.* 8, 101–120.
- Bekaert, M., Edger, P.P., Pires, J.C., Conant, G.C., 2011. Two-phase resolution of polyploidy in the *Arabidopsis* metabolic network gives rise to relative and absolute dosage constraints. *Plant Cell* 23, 1719–1728.
- Bodnaryk, R.P., 1992. Effects of wounding on glucosinolates in the cotyledons of oilseed rape and mustard. *Phytochemistry* 31, 2671–2677.
- Bodnaryk, R.P., 1994. Potent effect of jasmonates on indole glucosinolates in oilseed rape and mustard. *Phytochemistry* 35, 301–305.
- Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254.
- Brew, C.T., Aronchik, I., Kosco, K., McCammon, J., Bjeldanes, L.F., Firestone, G.L., 2009. Indole-3-carbinol inhibits MDA-MB-231 breast cancer cell motility and induces stress fibers and focal adhesion formation by activation of Rho kinase activity. *Int. J. Cancer* 124, 2294–2302.
- Burandt, P., Papenbrock, J., Schmidt, A., Bloem, E., Haneklaus, S., Schnug, E., 2001. Genotypical differences in total sulfur contents and cysteine desulfhydrase activities in *Brassica napus* L. *Phyton* 41, 75–86.
- Canola Council of Canada, 1990. CANOLA Oil and Meal: Standards and Regulations. Canola Council of Canada Publication.
- Cartea, M.E., Velasco, P., Obregón, S., Padilla, G., de Haro, A., 2008. Seasonal variation in glucosinolate content in *Brassica oleracea* crops grown in northwestern Spain. *Phytochemistry* 69, 403–410.
- Chalhoub, B., Denoed, F., Liu, S., Parkin, I.A.P., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., Corréa, M., Da Silva, C., Just, J., Falentin, C., Koh, C.S., Le Clainche, I., Bernard, M., Bento, P., Noel, B., Labadie, K., Alberti, A., Charles, M., Arnaud, D., Guo, H., Daviaud, C., Alameiry, S., Jabbari, K., Zhao, M., Edger, P.P., Chelaifa, H., Tack, D., Lassalle, G., Mestiri, I., Schnel, N., Le Paslier, M., Fan, G., Renault, V., Bayer, P.E., Golick, A.A., Manoli, S., Lee, T., Thi, V.H.D., Chalabi, S., Hu, Q., Fan, C., Tollenaere, R., Lu, Y., Battail, C., Shen, J., Sidebottom, C.H.D., Wang, X., Canaguier, A., Chauveau, A., Bérard, A., Deniot, G., Guan, M., Liu, Z., Sun, F., Lim, Y.P., Lyons, E., Town, C.D., Bancroft, I., Wang, X., Meng, J., Ma, J., Pires, J.C., King, G.J., Brunel, D., Delourme, R., Renard, M., Aury, J., Adams, K.L., Batley, J., Snowden, R.J., Tost, J., Edwards, D., Zhou, Y., Hua, W., Sharpe, A.G., Paterson, A.H., Guan, C., Wincker, P., 2014. Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345, 950–953.

- Charron, C.S., Saxton, A.M., Sams, C.E., 2005. Relationship of climate and genotype to seasonal variation in the glucosinolate-myrosinase system. I. Glucosinolate content in ten cultivars of *Brassica oleracea* grown in fall and spring seasons. *J. Sci. Food Agric.* 85, 671–681.
- Chen, Z.J., 2010. Molecular mechanisms of polyploidy and hybrid vigor. *Trends Plant Sci.* 15, 57–71.
- Chen, Y., Yan, X., Chen, S., 2011. Bioinformatic analysis of molecular network of glucosinolate biosynthesis. *Comput. Biol. Chem.* 35, 10–18.
- Cheung, F., Trick, M., Drou, N., Lim, Y.P., Park, J., Kwon, S., Kim, J., Scott, R., Pires, J.C., Paterson, A.H., Town, C., Bancroft, I., 2009. Comparative analysis between homeologous genome segments of *Brassica napus* and its progenitor species reveals extensive sequence-level divergence. *Plant Cell* 21, 1912–1928.
- Chew, F.S., 1988. Biological effects of glucosinolates, biologically active natural products. *ACS Symp.* 380, 155–181.
- Clarke, D.B., 2010. Glucosinolates, structures and analysis in food. *Anal. Methods* 2, 310.
- Conant, G.C., Wolfe, K.H., 2008. Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* 9, 938–950.
- Curien, G., Bastien, O., Robert-Genthon, M., Cornish-Bowden, A., Cárdenas, M.L., Dumas, R., 2009. Understanding the regulation of aspartate metabolism using a model based on measured kinetic parameters. *Mol. Syst. Biol.* 5, 271.
- Dantas-Santos, N., Gomes, D.L., Costa, L.S., Cordeiro, S.L., Costa, M.S.S.P., Trindade, E.S., Franco, C.R.C., Scortecchi, K.C., Leite, E.L., Rocha, H.A.O., 2012. Freshwater plants synthesize sulfated polysaccharides: heterogalactans from water hyacinth (*Eichhornia crassipes*). *Int. J. Mol. Sci.* 13, 961–976.
- Doughty, K.J., Kiddle, G.A., Pye, B.J., Wallsgrave, R.M., Pickett, J.A., 1995. Selective induction of glucosinolates in oilseed rape leaves by methyl jasmonate. *Phytochemistry* 38, 347–350.
- Doughty, K.J., Porter, A.J.R., Morton, A.M., Kiddle, G., Bock, C.H., Wallsgrave, R., 1991. Variation in the glucosinolate content of oilseed rape (*Brassica napus* L.) leaves. *Ann. Appl. Biol.* 118, 469–477.
- Faës, P., Deleu, C., Ainouche, A., Le Cahérec, F., Montes, E., Clouet, V., Gouraud, A.M., Benjamin Albert, B., Orsel, M., Lassalle, G., Leport, L., Bouchereau, A., Niogret, M.F., 2014. Molecular evolution and transcriptional regulation of the oilseed rape proline dehydrogenase genes suggest distinct roles of proline catabolism during development. *Planta* 241, 403–419.
- Fahey, J.W., Zalcmann, A.T., Talalay, P., 2001. The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* 56, 5–51.
- Falk, K.L., Tokuhisa, J.G., Gershenzon, J., 2007. The effect of sulfur nutrition on plant glucosinolate content: physiology and molecular mechanisms. *Plant Biol.* 9, 573–581.
- Fieldsend, J., Milford, G.F.J., 1994. Changes in glucosinolates during crop development in single- and double-low genotypes of winter oilseed rape (*Brassica napus*): I. Production and distribution in vegetative tissues and developing pods during development and potential role in the recycling of sulphur within the crop. *Ann. Appl. Biol.* 124, 531–542.
- Gaeta, R.T., Chris Pires, J., 2010. Homeologous recombination in allopolyploids: the polyploid ratchet. *New Phytol.* 186, 18–28.
- Graser, G., Oldham, N.J., Brown, P.D., Temp, U., Gershenzon, J., 2001. The biosynthesis of benzoic acid glucosinolate esters in *Arabidopsis thaliana*. *Phytochemistry* 57, 23–32.
- Harper, A.L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., Hattori, C., Werner, P., Bancroft, I., 2012. Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. *Nat. Biotechnol.* 30, 798–802.
- Hashiguchi, T., Sakakibara, Y., Hara, Y., Shimohira, T., Kurogi, K., Akashi, R., Liu, M., Suiko, M., 2013. Identification and characterization of a novel kaempferol sulfotransferase from *Arabidopsis thaliana*. *Biochem. Biophys. Res. Commun.* 434, 829–835.
- Hirschmann, F., Krause, F., Papenbrock, J., 2014. The multi-protein family of sulfotransferases in plants: composition, occurrence, substrate specificity, and functions. *Front. Plant Sci.* 5, 556.
- Hopkins, R.J., van Dam, N.M., van Loon, J.J., 2009. Role of glucosinolates in insect-plant relationships and multitrophic interactions. *Annu. Rev. Entomol.* 54, 57–83.
- Kakuta, Y., Petrotchenko, E.V., Pedersen, L.C., Negishi, M., 1998. The sulfuryl transfer mechanism. *J. Biol. Chem.* 273, 27325–27330.
- Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucl. Acids Res.* 28, 27–30.
- Karp, P.D., Ouzounis, A., Moore-Kochlacs, C., Goldovsky, L., Kaipa, P., Ahrén, D., Tsoka, S., Darzentas, N., Kunin, V., López-Bigas, N., 2005. Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucl. Acids Res.* 33, 6083–6089.
- Kim, J.K., Chu, S.M., Kim, S.J., Lee, D.J., Lee, S.Y., Lim, S.H., Ha, S., Kweon, S.J., Cho, H.S., 2010. Variation of glucosinolates in vegetable crops of *Brassica rapa* L. ssp. *pekinensis*. *Food Chem.* 119, 423–428.
- 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., Papenbrock, J., 2008. Sulfotransferases and their role in glucosinolate biosynthesis. In: Khan, N., Singh, S., Umar, S. (Eds.), *Sulfur Assimilation and Abiotic Stress in Plants*. Springer, Berlin Heidelberg, pp. 149–166.
- Klein, M., Papenbrock, J., 2009. Kinetics and substrate specificities of desulfoglucosinolate sulfotransferases in *Arabidopsis thaliana*. *Physiol. Plant* 135, 140–149.
- 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 J.* 273, 122–136.
- Kliebenstein, D.J., Kroymann, J., Brown, P., Figuth, A., Pedersen, D., Gershenzon, J., 2001. Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. *Plant Physiol.* 126, 811–825.
- Komori, R., Amano, Y., Ogawa-Ohnishi, M., Matsubayashi, Y., 2009. Identification of tyrosylprotein sulfotransferase in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 106, 15067–15072.
- Kudrová, A., Gallová, L., Kuricová, K., Nejedlá, E., Čurdová, A., Mícenková, L., 2015. Identification of AHK2- and AHK3-like cytokinin receptors in *Brassica napus* reveals two subfamilies of AHK2 orthologues. *J. Exp. Bot.* 66, 339–353.
- Lisch, D., 2013. How important are transposons for plant evolution? *Nat. Rev. Genet.* 14, 49–61.
- Louda, S., Mole, S., 1991. Glucosinolates: chemistry and ecology. In: Rosenthal, G.A., Berenbaum, M.R. (Eds.), *Herbivores: Their Interactions with Secondary Plant Metabolites*, second ed., pp. 123–164 San Diego, CA.
- Lu, G., Harper, A.L., Trick, M., Morgan, C., Fraser, F., O'Neill, C., Bancroft, I., 2014. Associative transcriptomics study dissects the genetic architecture of seed glucosinolate content in *Brassica napus*. *DNA Res.* 21, 613–625.
- Luczak, S., Forlani, F., Papenbrock, J., 2013. Desulfo-glucosinolate sulfotransferases isolated from several *Arabidopsis thaliana* ecotypes differ in their sequence and enzyme kinetics. *Plant Physiol. Biochem.* 63, 15–23.
- Lyon, E., Jakoby, W., 1980. The identity of alcohol sulfotransferases with hydroxysteroid sulfotransferases. *Arch. Biochem. Biophys.* 202, 474–481.
- Marsolais, F., Boyd, J., Paredes, Y., Schinas, A., Garcia, M., Elzein, S., Varin, L., 2007. Molecular and biochemical characterization of two brassinosteroid sulfotransferases from *Arabidopsis*, AtST4a (At2g14920) and AtST1 (At2g03760). *Planta* 225, 1233–1244.
- Marsolais, F., Gidda, S.K., Boyd, J., Varin, L., 2000. Plant soluble sulfotransferases: structural and functional similarity with mammalian enzymes. *Rec. Adv. Phytochem.* 34, 433–456.
- Marsolais, F., Sebastião, C.H., 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.
- Mithen, R., 1992. Leaf glucosinolate profiles and their relationship to pest and disease resistance in oilseed rape. *Euphytica* 63, 71–83.
- Møldrup, M.E., Geu-Flores, F., Olsen, C.E., Halkier, B.A., 2011. Modulation of sulfur metabolism enables efficient glucosinolate engineering. *BMC Biotechnol.* 11, 12.
- Nagahara, U., 1935. Genomic analysis of *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Jpn. J. Bot.* 7, 389–452.
- Padilla, G., Carrea, M.E., Velasco, P., Haro, A., Ordás, A., 2007. Variation of glucosinolates in vegetable crops of *Brassica rapa*. *Phytochemistry* 68, 536–545.
- Parkin, I.A.P., Clarke, W.E., Sidebottom, C., Zhang, W., Robinson, S.J., Links, M.G., Karcz, S., Higgins, E.E., Fobert, P., Sharpe, A.G., 2010. Towards unambiguous transcript mapping in the allotetraploid *Brassica napus*. *Genome* 53, 929–938.
- Parkin, I.A.P., Lydiat, D.J., Trick, M., 2002. Assessing the level of collinearity between *Arabidopsis thaliana* and *Brassica napus* for A. *thaliana* chromosome 5. *Genome* 45, 356–366.
- Piotrowski, M., Schemenewitz, A., Lopukhina, A., Müller, A., Janowitz, T., Weiler, E.W., Oecking, C., 2004. Desulfoglucosinolate sulfotransferases from *Arabidopsis thaliana* catalyze the final step in the biosynthesis of the glucosinolate core structure. *J. Biol. Chem.* 279, 50717–50725.
- Quijada, P.A., Udall, J.A., Lambert, B., Osborn, T.C., 2006. Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (*Brassica napus* L.): 1. Identification of genomic regions from winter germplasm. *Theor. Appl. Genet.* 113, 549–561.
- Reichelt, M., Brown, P.D., Schneider, B., Oldham, N.J., Stauber, E., Tokuhisa, J., Kliebenstein, D.J., Mitchell-Olds, T., Gershenzon, J., 2002. Benzoic acid glucosinolate esters and other glucosinolates from *Arabidopsis thaliana*. *Phytochemistry* 59, 663–671.
- Rosa, E., Heaney, R., 1996. Seasonal variation in protein, mineral and glucosinolate composition of Portuguese cabbages and kale. *Anim. Feed Sci. Technol.* 57, 111–127.
- 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.
- Schiessl, S., Samans, B., Hüttel, B., Reinhard, R., Snowden, R.J., 2014. Capturing sequence variation among flowering-time regulatory gene homologs in the allopolyploid crop species *Brassica napus*. *Front. Plant Sci.* 5, 404.
- Sønderby, I.E., Geu-Flores, F., Halkier, B.A., 2010. Biosynthesis of glucosinolates – gene discovery and beyond. *Trends Plant Sci.* 15, 283–290.
- Town, C.D., Cheung, F., Maiti, R., Crabtree, J., Haas, B.J., Wortman, J.R., Hine, E.E., Althoff, R., Arbogast, T.S., Tallon, L.J., Vigouroux, M., Trick, M., Bancroft, I., 2006. Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. *Plant Cell* 18, 1348–1359.
- Udall, J.A., Quijada, P.A., Lambert, B., Osborn, T.C., 2006. Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (*Brassica napus* L.): 2. Identification of alleles from unadapted germplasm. *Theor. Appl. Genet.* 113, 597–609.
- Velasco, P., Soengas, P., Vilar, M., Carrea, M.E., del Rio, M., 2008. Comparison of

- glucosinolate profiles in leaf and seed tissues of different *Brassica napus* crops. *J. Am. Soc. Hortic. Sci.* 133, 551–558.
- Wapinski, I., Pfeffer, A., Friedman, N., Regev, A., 2007. Natural history and evolutionary principles of gene duplication in fungi. *Nature* 449, 54–61.
- Weinshilboum, R.M., Otterness, D.M., Aksoy, I.A., Wood, T.C., Her, C., Raftogianis, R.B., 1997. Sulfation and sulfotransferases 1: sulfotransferase molecular biology: cDNAs and genes. *FASEB J.* 11, 3–14.
- Wittstock, U., Kliebenstein, D.J., Lambrix, V., Reichelt, M., Gershenzon, J., 2003. Glucosinolate hydrolysis and its impact on generalist and specialist insect herbivores. In: Romeo, J.T. (Ed.), *Integrative Phytochemistry: from Ethnobotany to Molecular Ecology*. Elsevier, Amsterdam, pp. 101–125.
- Yang, T., Kim, J.S., Kwon, S., Lim, K., Choi, B., Kim, J., Jin, M., Park, J.Y., Lim, M., Kim, H., Lim, Y.P., Kang, J.J., Hong, J., Kim, C., Bhak, J., Bancroft, I., Park, B., 2006. Sequence-level analysis of the diploidization process in the triplicated flowering locus C region of *Brassica rapa*. *Plant Cell* 18, 1339–1347.
- Zang, Y., Kim, H.U., Kim, J.A., Lim, M., Jin, M., Lee, S.C., Kwon, S., Lee, S., Hong, J.K., Park, T., Mun, J., Seol, Y., Hong, S., Park, B., 2009. Genome-wide identification of glucosinolate synthesis genes in *Brassica rapa*. *FEBS J.* 276, 3559–3574.
- Żmieńko, A., Samelak, A., Kozłowski, P., Figlerowicz, M., 2014. Copy number polymorphism in plant genomes. *Theor. Appl. Genet.* 127, 1–18.

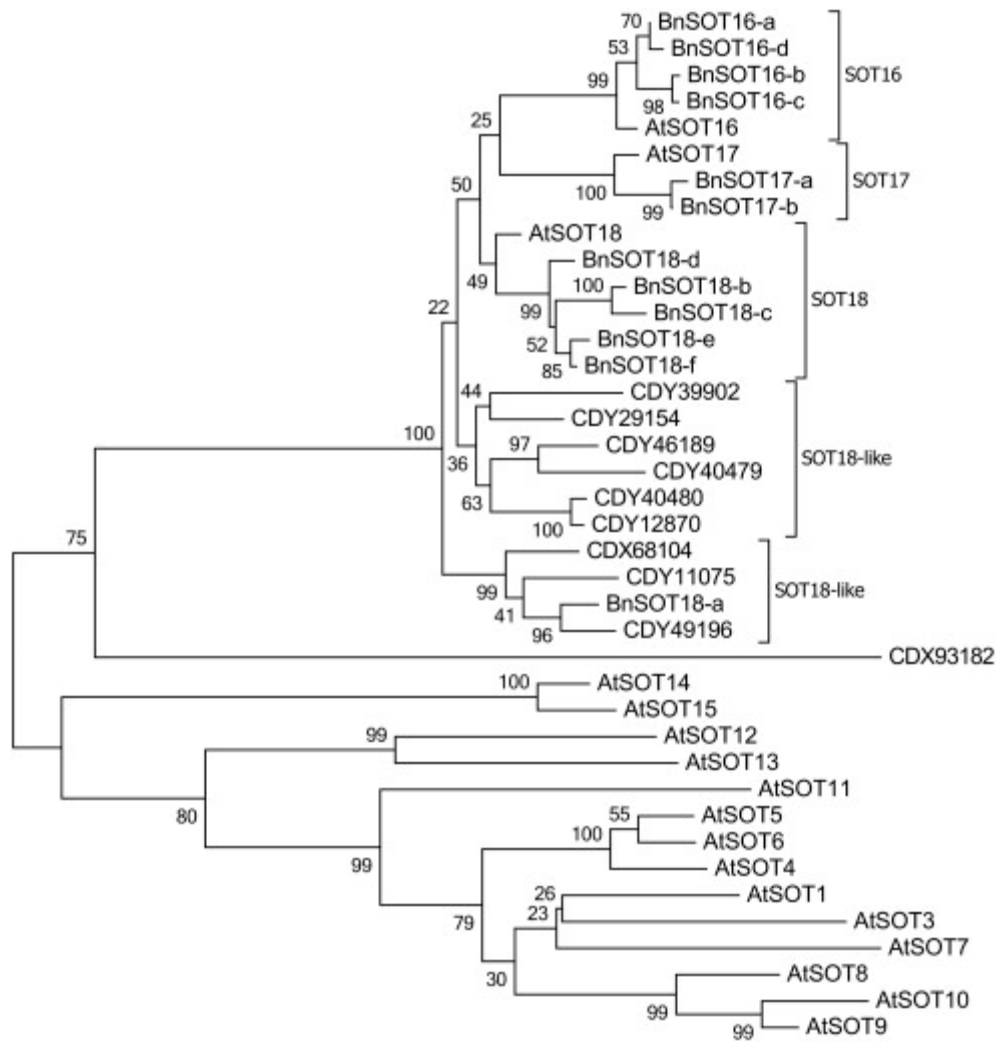


Fig. S1. Phylogenetic tree of all *B. napus* ds-GI SOT candidates and *A. thaliana* SOTs, after MUSCLE alignment on amino acid level, based on the Maximum likelihood method and Jones–Taylor–Thornton model with gamma distribution, including 1000 bootstrap values. The tree is unrooted.

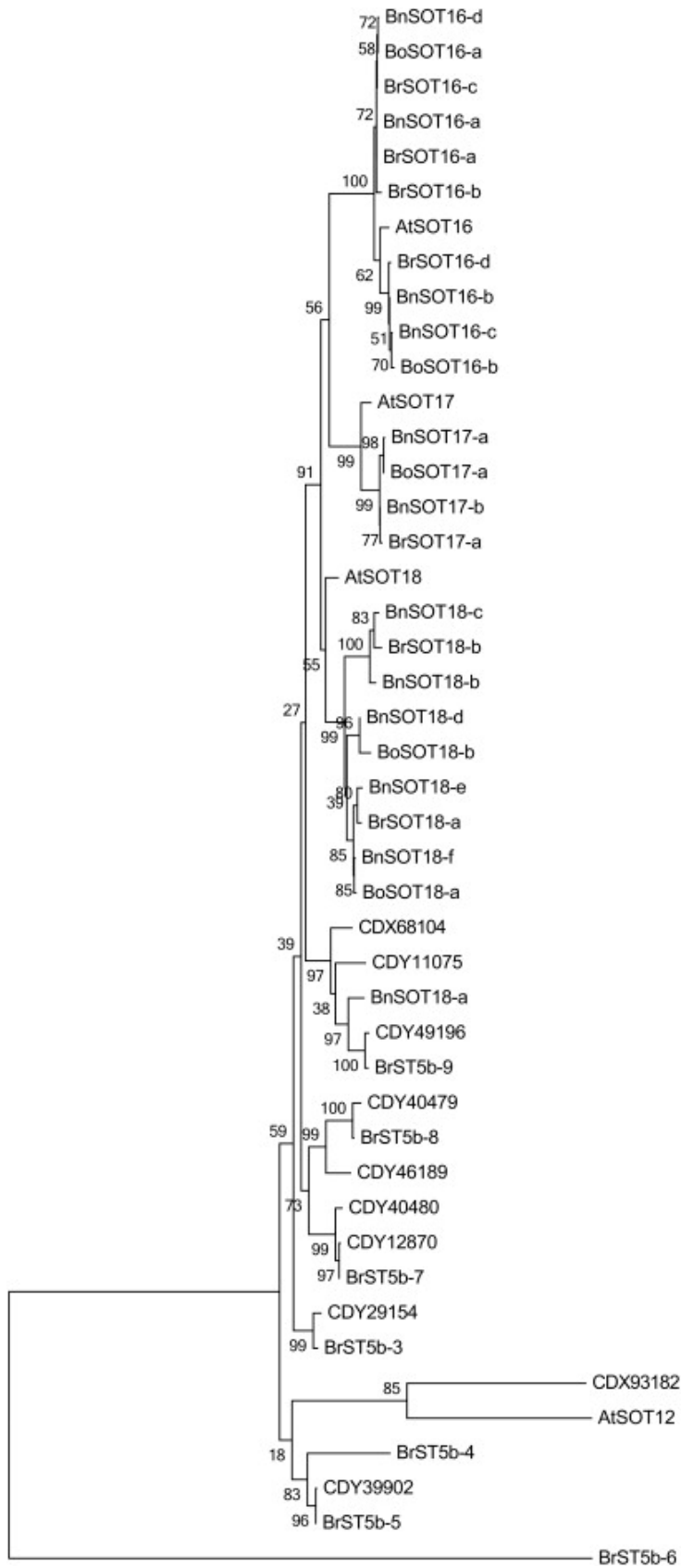


Fig. S2. Phylogenetic tree of all *B. napus*, *A. thaliana*, *B. rapa*, and *B. oleracea* ds-GI SOT candidates, after MUSCLE alignment on amino acid level, based on the Maximum likelihood method and Jones–Taylor–Thornton model with gamma distribution, including 1000 bootstrap values. The tree is unrooted.

CHAPTER 4

The structural and biochemical study of sulfotransferase 18 from *Arabidopsis thaliana* explains its substrate specificity and reaction mechanism

Hirschmann, F., Krause, F., Baruch, P., Chizhov, I., Mueller, J.W., Manstein, D.J., Papenbrock, J., Fedorov, R. The structural and biochemical study of sulfotransferase 18 from *Arabidopsis thaliana* explains its substrate specificity and reaction mechanism. *Scientific Reports* (manuscript for publication)

Abstract

Sulfotransferases (SOTs) catalyze the transfer of a sulfate group from 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to secondary metabolites in the cytosol or peptides in the Golgi apparatus. Twenty-two SOT sequences were identified in *Arabidopsis thaliana* (L.) Heynh. Three of those participate in stress response to herbivores, pathogens and abiotic stress by anticipation in glucosinolate (Gl) formation. The transfer mechanism of SOTs is not conserved throughout all members of the SOT family, but most SOTs from eukaryotic organisms follow an ordered sequential or random sequential Bi Bi catalytic mechanism. To explain the differences in substrate specificity of desulfo (ds)-Gl SOTs and get an insight into the reaction mechanism of plant SOTs, we determined high-resolution crystal structures of the ds-Gl SOT AtSOT18 ecotype "Col-0" in complex with 3'-phosphoadenosine 5'-phosphate (PAP) alone and together with the Gl sinigrin in the active site. The new structural data were supplemented by mutagenesis studies, a two-dimensional substrate-inhibitor titration of enzymatic activity and conformational dynamics analysis. The overall structure of AtSOT18 shows high similarity to mammalian SOTs, which illustrates the high evolutionary conservation of this multifunctional enzyme family. We identified the essential residues for substrate binding and catalysis and demonstrated that the catalytic mechanism is conserved between human and plant enzymes. The structural basis for the functional divergence of SOTs is provided by the fine-tuned conformational control of the unique residues in the individual enzymes. The obtained kinetic data imply that turnover time of AtSOT18 reaction is close to 4 seconds and that the sulfate group of PAPS does not contribute significantly to the binding energy of the donor. The residues in the active site and the adjacent regions do not provide the selectivity source for the ds-Gl SOT proteins. Instead, the non-conserved residues of the three functional loops are most likely to be responsible for substrate selectivity.

Introduction

Members of the sulfotransferase (SOT) (EC 2.8.2.-) protein family, which can be found in all organisms analyzed so far, catalyze the transfer of a sulfate group from the co-substrate 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to a hydroxyl group of different kinds of substrates. In plants the sulfated compounds act as hormones, as secondary metabolites in stress defense and they probably serve as a sulfur reservoir (Klein and Papenbrock, 2004; Hirschmann et al., 2014). Of special interest is the role of SOTs in the sulfation of desulfo-glucosinolates (ds-GI) (Fig. 1), as they are important secondary metabolites involved in defense against herbivores and pathogens in the order Brassicales (Rausch and Wachter, 2005).

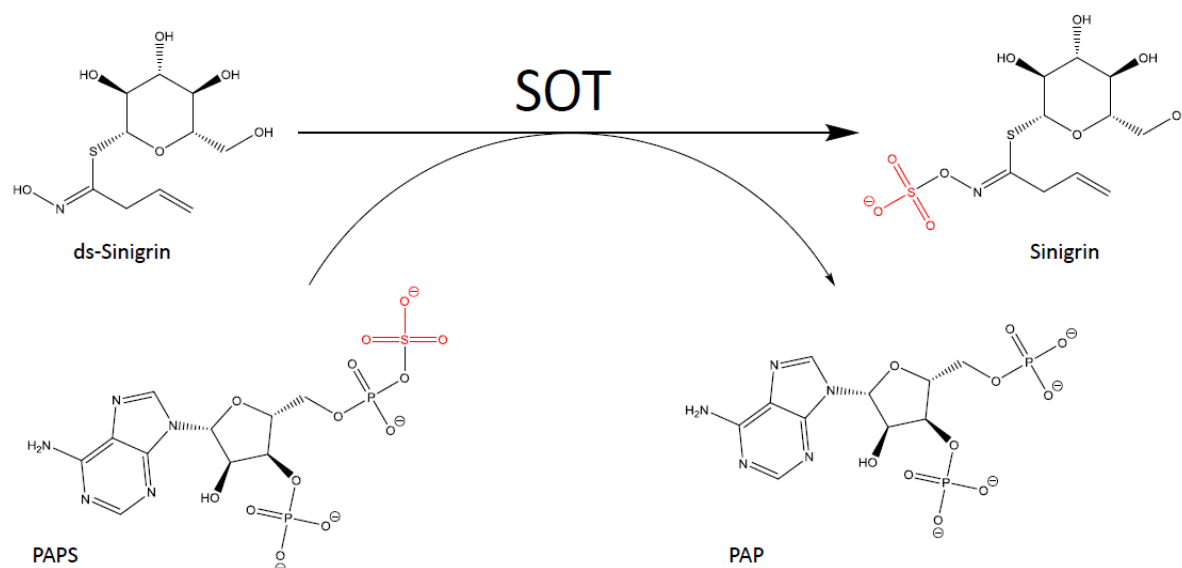


Fig. 1. AtSOT18 catalyzed reaction. The red coloured sulfate moiety (RSO_3^-) is transferred from PAPS to the hydroxyl group of the ds-GI sinigrin.

For humans, GI degradation products play a role as flavor compounds from numerous cabbage, radish and mustard species. As an example, allyl isothiocyanate accounts for the spiciness of horse radish and mustard (Clarke, 2010). Of high interest for the search of new medical compounds is the anticancerogenic activity of GIs (Faulkner et al., 1998; Gupta et al., 2014). Sulforaphan from broccoli and cabbage (1-isothiocyanato-4-methylsulfinyl-butan GI) has been reported to inhibit phase I cytochrome P450 enzymes (Gross-Steinmeyer et al., 2005), to induce phase II detoxification enzymes (Fahey and Talalay, 1999), to cause cell cycle arrest (Wang et al., 2004), to be anti-inflammatory (Konwinski et al., 2004) and to inhibit angiogenesis (Bertl

et al., 2006). However, Gl degradation products also have negative characteristics, e.g. goitrogenicity (Laurberg et al., 2002), and can be toxic to humans (Cartea and Velasco, 2008).

In *Arabidopsis thaliana* the three ds-Gl SOTs AtSOT16, AtSOT17 and AtSOT18 exclusively catalyze the transfer of a sulfate group to different ds-Gls (Piotrowski et al., 2004; Hirai et al., 2005; Klein et al., 2006). All three enzymes are localized in the cytoplasm and their expression pattern in several tested conditions is similar (Klein et al., 2006). These three AtSOTs share at least 72% sequence identity on amino acid level, but they differ remarkably in their catalytic activities. *In vitro* enzyme assays revealed a preference of AtSOT16 for the indolic ds-Gl indol-3-yl-methyl Gl (I3M), AtSOT17 showed an increased specific activity with long-chained ds-Gls derived from methionine and AtSOT18 preferred the long-chained ds-Gls, 7-methylthioheptyl Gl and 8-methylthiooctyl Gl (8MTO), also derived from methionine (Klein and Papenbrock, 2009). Although the substrate specificity of ds-Gl SOTs is well known, its molecular basis is not understood. In previous studies we could show that the mutation of a single amino acid in AtSOT18 leads to dramatic changes in activity (Klein et al., 2006). Furthermore, we could demonstrate that AtSOT18s from different ecotypes vary in their substrate specificities (Luczak et al., 2013). These findings indicate a major function of AtSOT18 in Gl biosynthesis, hence further analysis on a molecular level is necessary

Most SOTs can easily be identified by four conserved regions (I – IV) (Varin et al., 1992) (Fig. 3), including a highly conserved catalytic histidine at the beginning of region II (Kakuta et al., 1997). For plant SOTs the functions of these regions have not yet been identified, but have been suggested to be involved in PAPS binding (Hirschmann et al., 2014). So far, SOT structures from *Mus musculus*, *Homo sapiens* and several prokaryotes have been published (Chapman et al., 2004; Tibbs et al., 2015). Furthermore, the apo-form structure of *A. thaliana* SOT12 has been resolved (Smith et al., 2004). Structurally, all soluble SOT enzymes share a common fold consisting of four central β -strands surrounded by α -helices (Chapman et al., 2004). Three flexible loops, gating the substrate binding site were reported to influence substrate specificity (Tibbs et al., 2015). The conformational properties of these loops in the apo-state of AtSOT12 remain unclear, due to the lack of their structural information (Smith et al., 2004). It was proposed that human SOTs are dimeric units, because of a short and strictly conserved dimerization domain (Petrotchenko et al., 2001).

In order to be able to modulate the sulfation system, it is necessary to understand the exact SOT reaction mechanism, which is still under debate. Eukaryotic soluble SOTs follow a sequential mechanism, in either a specific or independent order (Tibbs et al., 2015), while eukaryotic

membrane-associated (Chapman et al., 2004) and bacterial SOTs (Malojčić et al., 2014) studied so far, follow a ping-pong mechanism. Also for the well-studied human SOTs the order of the nucleophilic substitution is not determined. Kinetic isotope effect studies indicated an S_N1 -like mechanism (Hoff et al., 2006), but crystal structures with PAPS and substrate suggest an S_N2 -like inline displacement mechanism (Teramoto et al., 2009).

In addition to the basic characterization of the two-substrate reaction mechanism of ds-Gl SOTs, which is interesting from a biochemical point of view, a better understanding of the reaction mechanism and the substrate specificity might contribute to develop strategies for manipulating and optimizing the Gl content and composition of crop plants in the Brassicaceae family. Then the medicinal and biotechnological potential of Gl-containing plants as nutraceuticals, and as source of anticancerogenic and antimicrobial compounds could be fully exploited.

A related interesting question concerns regulatory aspects. In former studies we could show, that high concentrations of substrates (ds-Gl and PAPS) led to an inhibition of enzyme activity (unpublished results). Furthermore, investigating the effects of higher co-substrate PAPS and co-product 3'-phosphoadenosine 5'-phosphate (PAP) concentrations could lead to a better understanding of the regulation of the enzyme activity.

The aim of this study was to gain deeper insights into (1) amino acids, that are responsible for the substrate specificity of ds-Gl SOTs and (2) the reaction mechanism of ds-Gl SOTs. To address these questions, the enzyme was crystallized in a binary complex with PAP and a ternary complex with PAP and sinigrin. To gain insights into the reaction mechanism and kinetics of AtSOT18, its inhibition/activation by PAP and PAPS was studied using two-dimensional fit of enzymatic titration data. Furthermore, catalytic residues were identified based on our structural information and mutated to alanine. The respective mutants did not show any activity, which indicates their significance in catalysis. Analysis of the substrate binding site indicated that ds-Gl SOT specificity is controlled outside of the active center, most likely by the interplay of three functional loops shaping the binding pocket.

Results

Structure of AtSOT18

In order to address ds-GI SOT specificity and catalysis, the structure of AtSOT18 with the product GI sinigrin alone and together with the co-product PAP, was solved (Fig. 2). The overall structure of AtSOT18 matches with previously described mammalian ones (Chapman et al., 2004; Tibbs et al., 2015). The globular enzyme consists of four central β -strands forming the characteristic backbone, surrounded by 21 α -helical turns and two additional smaller β -strands (Fig. 2, Fig. 3). Also the three typical flexible loops, including loop 1, which is only found in human SOT1 subfamily proteins, gating the sinigrin binding site could be identified. The highly conserved His155 is localized in the catalytic center right in between the bound PAP and sinigrin (Fig. 5).

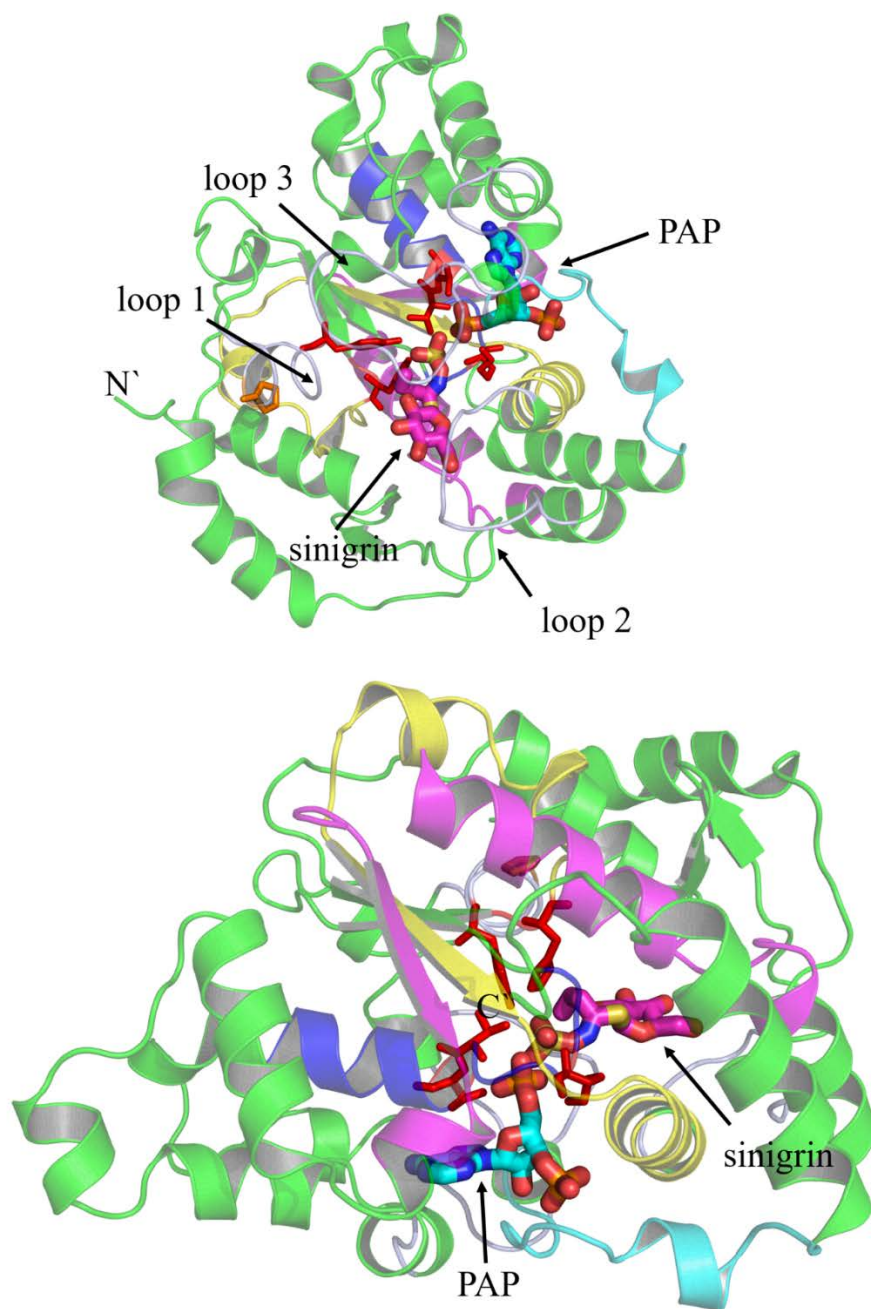


Fig. 2. Overall view of AtSOT18 from two perspectives bound with sinigrin (magenta sticks) and PAP (cyan sticks). Indicated are the four conserved regions (region I: blue; region II: yellow; region III: magenta; region IV: cyan), the three flexible loop regions (grey), the catalytic residues (red) and proline 136 (orange).

located close to the adenyl group of the donor PAP. The electron density for sinigrin was partly diffuse, therefore the electron-rich sulfate and hydroxyl groups were used as reference points for better positioning of the substrate (Fig. S2). The obtained fit was in agreement with the reference structural groups and produced no stereochemical clashes.

Sinigrin binding is facilitated by hydrogen bonds with the residues Arg51, Glu54, Thr96, Tyr130 and Tyr306 (Fig. 4a). The Arg51 guanidinium group interacts with the 6'-hydroxyl group of the glucopyranose of the Gl. The carboxyl group of Glu54 interacts with the 4'- and the 6'-hydroxyl groups. The hydroxyl of Tyr306 forms a hydrogen bond with the oxygen in the glucopyranose ring. The transferred sulfate group is stabilized by a hydrogen bond to Thr96 and Tyr130. The minimal distance between the transferred sulfate group of sinigrin and the 5'-phosphate group of PAP was 4.4 Å.

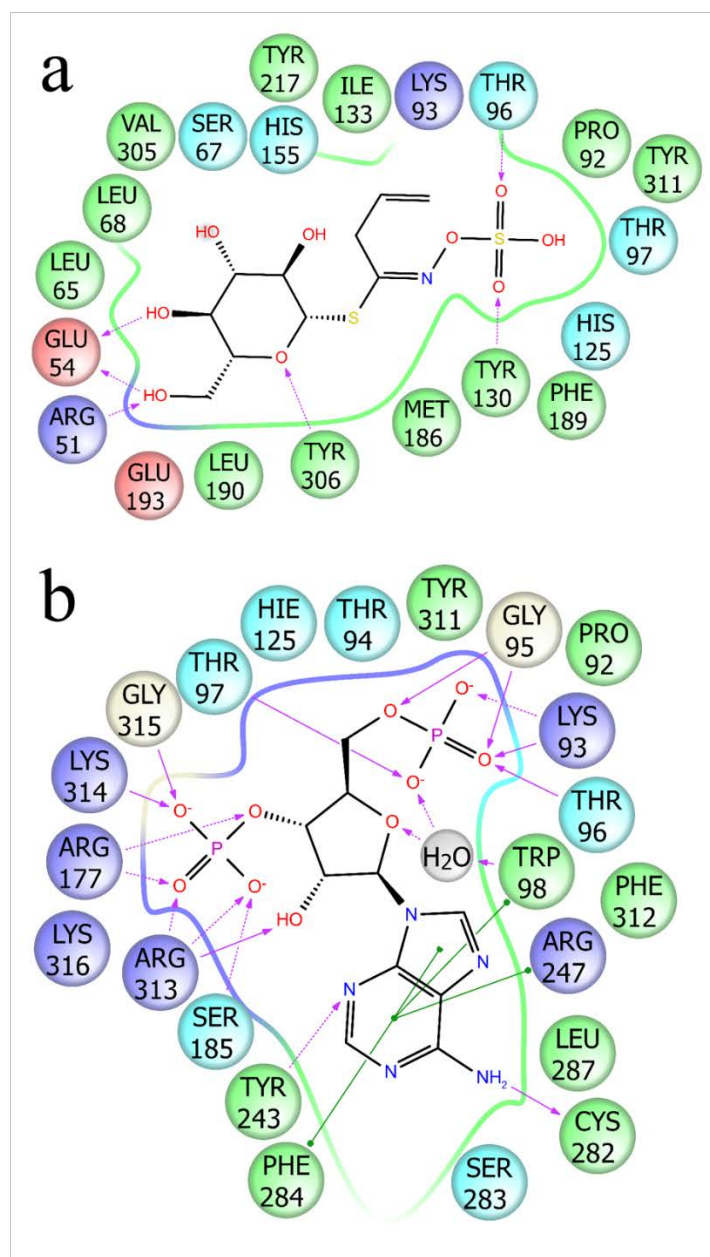


Fig. 4. Ligand binding sites. (a) Sinigrin binding site and (b) PAP binding site with π - π stacking (green lines) interaction and hydrogen bonds to residue backbone (purple arrows) and side chains (dotted purple arrow).

The quality of the electron density allowed an exact determination of the position and stereochemistry of the PAP (Fig. S2). PAP was embedded within the structure and stabilized by several hydrogen bonds and π - π stacking with surrounding amino acids. The four oxygen atoms of the 3'-phosphate group were stabilized by six hydrogen bonds with the side chains of Arg177, Ser185, Arg313, and the main chain of Lys314 and Gly315 (Fig. 4b). The oxygen atoms of the 5'-phosphate group were stabilized by six hydrogen bonds to the side chains of Lys93, the main chain of Gly95, and both the main and the side chains of Thr96 and Thr97.

Arg313 formed a hydrogen bond to the hydroxyl group at the 2'-carbon of PAP. The adenyl group of PAP is stabilized by hydrogen bonds with Cys282 and Tyr243, the stacking interaction with Trp98, and the hydrophobic contacts with Phe284 and aliphatic groups of Arg247. Interestingly, from the 14 residues that contact PAP directly, 11 are within the highly conserved regions I-IV (with the exception of Arg247, Cys282 and Phe284), while only a single (Thr96) of the 5 substrate binding residues is in there (Fig. 3). In the AtSOT18•PAP complex, the main conformation of Met186 was oriented away from the acceptor site (Fig. S3). In the second conformation the methionine side chain is oriented towards the acceptor, providing an additional stabilization to its hydrophobic moiety.

In summary, analysis of the binding sites revealed that the Gl sinigrin is rather loosely bound compared to the tightly bound PAP. Furthermore, it was shown that PAP binding residues are mostly located inside and sinigrin binding residues outside of the conserved regions (Fig. 3). Binding of the comparatively small substrate sinigrin induces only subtle structural changes, while the overall loop conformation remains about the same, except for a conformational change of the Met186 residue. The latter may contribute to the sinigrin stabilization by forming a new hydrophobic contact with the substrate.

Identification of the amino acids involved in the enzymatic reaction

The stereochemical analysis suggests that Lys93, Thr96 and 97, Tyr130 and His155 (Fig. 5), which are located directly in the catalytic center, may be involved in the proton transfer event, as well as in the stabilization of the transition states, which could lower the transition energy barrier of the chemical reaction.

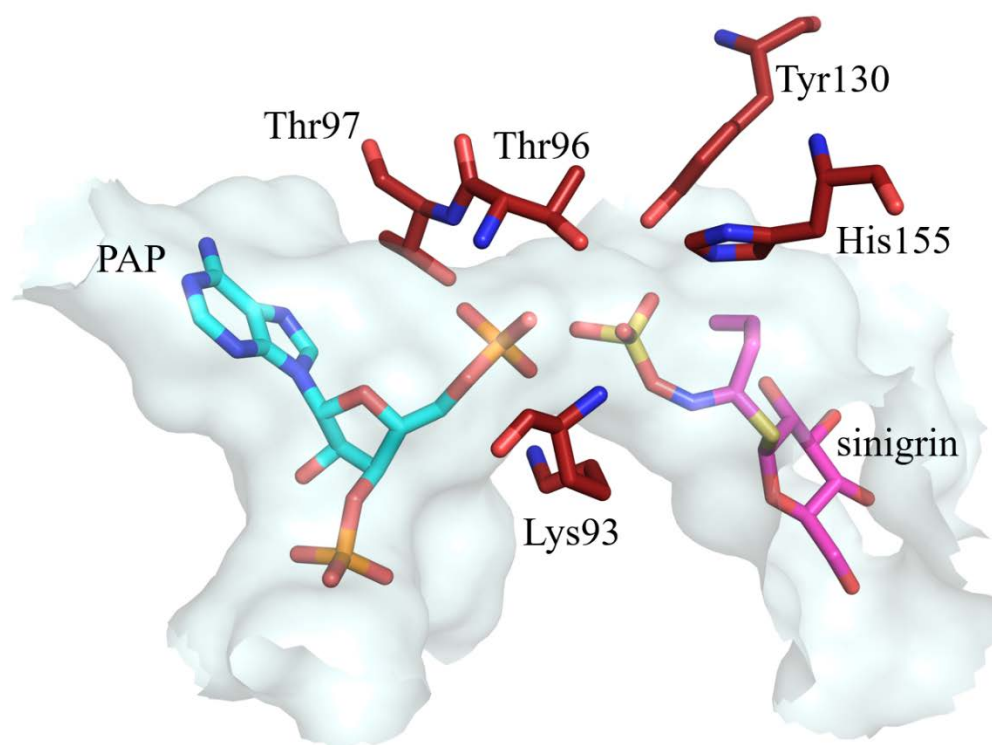


Fig. 5. Close up of the binding site with residues of the catalytic center. Catalytic residues (red) with bound sinigrin (magenta), PAP (cyan) and protein surface (light-grey)

To test the importance of the five residues in the catalytic center we performed mutagenesis and enzymatic activity studies. The selected amino acids are conserved through the AtSOT family, indicating a significant function, except for Tyr130, which is only conserved in 6 out of 17 AtSOTs, including the three ds-GI SOTs.

After the respective point mutations to alanine, the mutants were tested with 3-methylthiopropyl Gl (3MTP), 8MTO and sinigrin as substrates (Table 1). In the assays, the mutants Lys93Ala, Thr97Ala, Tyr130Ala and His155Ala showed strongly reduced activity. Thr96Ala still showed residual activity with the preferred substrates 3MTP and 8MTO (12-fold reduction and 3-fold reduction, respectively), while with sinigrin no more activity was detected (wildtype activity: 879 ± 410 pkatal mg^{-1}), possibly because the signal was under the detection limit. A major significance of the selected amino acids for catalysis is indicated by the absent or strongly reduced activities.

Table 1. Mutational analysis of the AtSOT18 enzyme. Selected amino acids in the catalytic center were mutated to alanine. The activity was tested with short-chained aliphatic 3MTP, long-chained aliphatic Gl 8MTO, co-crystallized sinigrin and indolic Gl I3M. The 150 μ L assays contained 80 mM Tris/HCl, pH 8.0, 9.2 mM MgCl₂, 60 mM of the respective ds-Gl substrates, 1 μ g purified protein, and 60 μ M PAPS. The reactions were started by the addition of PAPS, incubated for 20 min at 37 °C, and stopped by incubation at 95 °C for 10 min. The formation of the respective sulfated product was analyzed by HPLC at 229 nm. The specific activities are given in pkatal mg⁻¹. N.D., not detectable; -, not tested.

AtSOT18	Activity in pkatal mg ⁻¹			
	3MTP	8MTO	Sinigrin	I3M
Wildtype*	1624 \pm 122	1618 \pm 272	879 \pm 410	501 \pm 46
Lys93Ala	N.D.	N.D.	N.D.	-
Thr96Ala	122 \pm 19	536 \pm 73	N.D.	-
Thr97Ala	N.D.	N.D.	N.D.	-
Tyr130Ala	N.D.	N.D.	N.D.	-
His155Ala	N.D.	N.D.	N.D.	-
Pro136Ala	1726 \pm 264	-	-	473 \pm 18

Overall, the residues for substrate binding and catalysis were identified by analysing the active sites. Putative catalytic residues were tested by mutagenesis and enzyme activity study. The mutants showed drastically reduced activity, confirming the high impact of the respective residues.

Enzyme kinetics and inhibition tests of AtSOT18

The AtSOT18 enzyme kinetics and the inhibition of the enzyme by PAP were analyzed, as PAP is an important second messenger molecule in *A. thaliana* (Estavillo et al., 2011; Chan et al., 2013). For better comparison with former studies (Klein et al., 2006; Klein and Papenbrock, 2009; Hirschmann and Papenbrock, 2015), the inhibition tests were performed with 3MTP instead of sinigrin. The product formation was used to determine the respective enzymatic activity. A two-dimensional titration experiment was performed, where product concentration was determined at various concentrations of donor and inhibitor molecules.

For AtSOT18, the analysis of the enzyme kinetics included the following assumptions: According to literature (Chapman et al., 2004; Wang et al., 2014; Tibbs et al., 2015), PAP and

PAPS were expected to bind to the same site in the enzyme resulting in a competitive inhibition. This implies unaltered dissociation constants K_D for PAPS, K_A for 3MTP, and K_I for PAP. We also assumed that the dissociation constants K_D for PAPS, K_A for 3MTP, and K_I for PAP are independent of each other. Finally, the concentration of enzyme molecules E_0 in the reaction was much lower than the initial ligand concentrations of donor D_0 , acceptor A_0 , and inhibitor I_0 . Therefore, the ligand concentrations at equilibrium were assumed to be unchanged as compared to the initial concentrations. Also, the correlation between the product 3MTP and the equilibrium population of AtSOT18•PAPS•ds-3MTP complex was assumed to be linear. In other words: once the enzyme included both ligand molecules, the reaction was catalyzed and two product molecules were released with respective dissociation rates. These assumptions result in a six equilibrium states model depicted in Fig. 6. At the pseudo first order conditions (i.e. when $E_0 \ll A_0 ; D_0 ; I_0$) the following simple system of algebraic equations represent the probabilities to find the enzyme in the particular occupied state:

$$P_{00} + \tilde{P}_{10} + P_{10} + P_{01} + \tilde{P}_{11} + P_{11} = 1$$

$$\frac{P_{10}}{P_{00}} = \frac{P_{11}}{P_{01}} = \frac{D_0}{K_D}$$

$$\frac{P_{11}}{P_{10}} = \frac{P_{01}}{P_{00}} = \frac{\tilde{P}_{11}}{\tilde{P}_{10}} = \frac{A_0}{K_A}$$

$$\frac{\tilde{P}_{10}}{P_{00}} = \frac{\tilde{P}_{11}}{P_{01}} = \frac{I_0}{K_I}$$

where:

$$P_{ij} = \frac{E_{ij}}{E_0}$$

The resulting solution for the productive state P_{11} has the following hyperbolic form:

$$P_{11} = \frac{A_0 D_0}{A_0 D_0 + A_0 K_D + D_0 K_A + K_D K_A + I_0 \frac{(A_0 K_D + K_D K_A)}{K_I}}$$

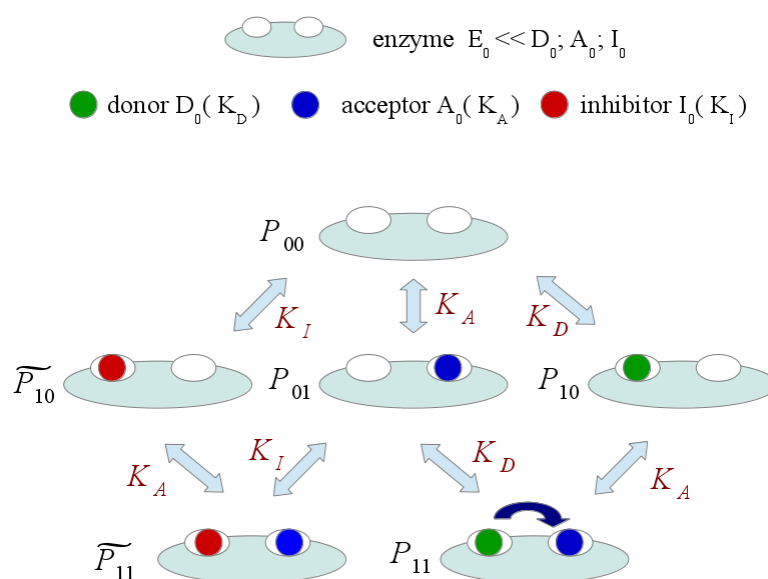


Fig. 6. Model of the competitive inhibition of AtSOT18 by PAP. In equilibrium six different populations of enzyme•ligand complexes are present with their respective probabilities P_{ij} , where i and j represent the status of the donor and acceptor site, respectively. 0 and 1 represent non-occupied and occupied enzyme sites. \tilde{P}_{ij} represents the probability of an enzyme•inhibitor complex species with the inhibitor bound to the donor site. D_0 , A_0 , and I_0 represent the initial concentration of donor PAPS, acceptor ds-3MTP, and inhibitor PAP, respectively. Dissociation constants are displayed as K_D for the donor, K_A for the acceptor, and K_I for the inhibitor in equilibrium, respectively. The sulfate transfer could only be catalyzed within the state P_{11} where donor and acceptor site are each occupied with the substrates.

Our experimental data were therefore approximated by the function $V_{max} * P_{11}$. Since the concentration of acceptor A_0 was fixed at 60 μM and the K_A were chosen equal to 55 μM , based on conditions used in previous studies (Glendening and Poulton, 1990; Klein et al., 2006; Klein and Papenbrock, 2009; Hirschmann and Papenbrock, 2015), this function contains three unknown parameters K_D , K_I , and V_{max} , and depends from the two independent variables: D_0

and I_0 . We have found that the Igor Pro software is the most appropriate data analysis package, since it allowed to perform the non-linear least squares fit of multidimensional data by user-defined functions. The best approximation of experimental data provides the following parameters for wild-type AtSOT18: $V_{max} = 5200 \pm 300$ pkatal mg^{-1} , $K_D = 4.2 \pm 2.6$ μM , $K_I = 3.0 \pm 1.7$ μM , with the RMSD value of 150 pkatal mg^{-1} . The experimental data and fit are shown in Fig. 7.

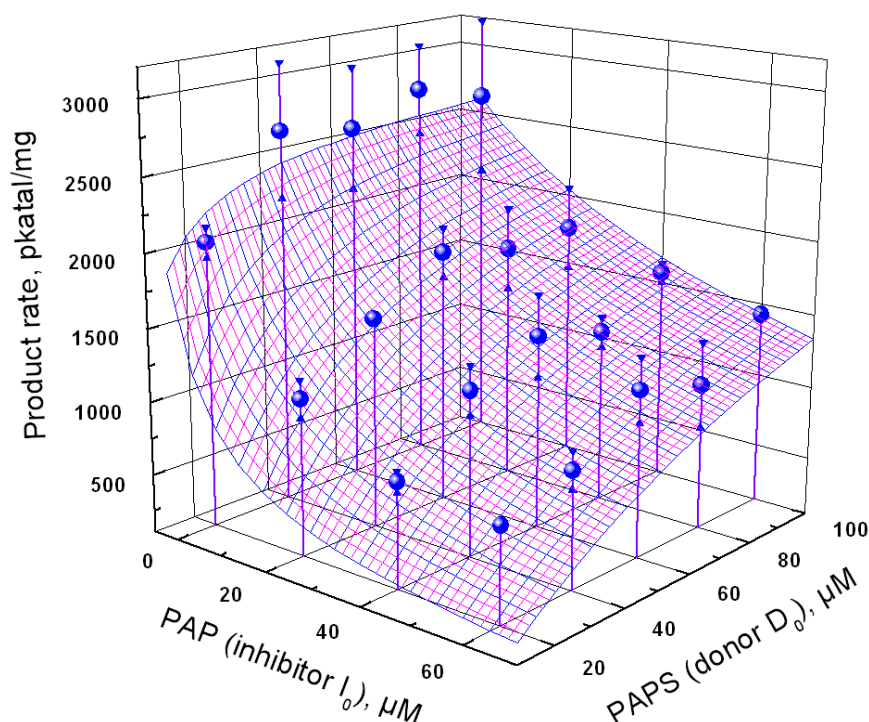


Fig. 7. 3D plot of AtSOT18 kinetic titration data their fit. Blue spheres represent the mean values of three replicates of the enzyme rate in pkatal mg^{-1} . Triangles are the estimated errors of experimental points. The wire frame surface represents the fitted function. Ds-3MTP concentration was kept constant at 60 μM .

In summary, the obtained kinetic data imply that turnover time of AtSOT18 reaction is close to 4 seconds. Furthermore, the sulfate group of PAPS does not contribute significantly to the binding energy of the donor.

The source of ds-GI SOT specificity

Another topic of interest was how the substrate specificity of ds-GI SOTs is regulated on the molecular level. The major differences are between AtSOT18, which hardly accepts indolic ds-GIs, and AtSOT16, which highly prefers these (Klein and Papenbrock, 2009). To investigate

structural differences between AtSOT18 and AtSOT16, a homology model of AtSOT16 was built based on our AtSOT18 structure. Comparison of the experimental AtSOT18 and homology AtSOT16 models revealed that both the amino-acid composition and the geometry of the Gl binding sites are highly conserved in both proteins (Fig. S4). Therefore, we concluded that the substrate specificity of ds-Gl SOTs may be regulated by the residues outside of the immediate active site. Analysis of residues in the second coordination sphere with respect to the Gl binding site revealed that AtSOT18 Pro136 is replaced by an alanine residue (Ala124) in AtSOT16 (Fig. S5). Hence, we hypothesized that the differences of conformational dynamics of helical turn $\alpha 6$ containing proline vs alanine may influence the flexibility of the Gl binding site, thus decreasing the ability of AtSOT18 to bind the comparably large I3M. Normal-Mode-Analysis (NMA) with the original AtSOT18 structure and an AtSOT18 Pro136Ala mutant revealed a bending deformation of $\alpha 6$ in the mutant vs a simple shift in the wild-type protein (Fig. S6), which supported our hypothesis. However, the mutation of Pro136Ala did not lead to the changes in substrate specificity *in vitro* (Table 1). At the same time, the homology models combined with NMA results suggested that the specificity may be controlled by the three functional loops. Amino acid alignment of AtSOT16 and AtSOT18 shows that the loops are non-conserved (Fig. S7). Loop 2 in particular is highly heterogeneous, with 30% of amino acids conserved. Loops 1 and 3 also contain many non-conserved residues. After careful exclusion of other options, we hypothesize the loop regions to be most important for substrate specificity, especially in the absence of other selectivity sources directly in the active site or in the adjacent regions.

Discussion

Biology of SOTs

Several studies have investigated molecular and structural features of SOTs, regarding the mechanism, specificity and function, as reviewed in Chapman et al. (2004) and Tibbs et al. (2015). In most cases, interest was focused on human SOTs (Lu et al., 2010), but also SOTs from mice (Kakuta et al., 1998), insects (Pakhomova et al., 2005) and prokaryotes (Malojčić et al., 2014) were analyzed in detail. In the plant kingdom, SOTs were mainly studied on a physiological level (Hirschmann et al., 2014), and until now, only one plant structure was published, which substrate-binding loops showed no electron density (Smith et al., 2004). Thus, the molecular details of SOTs from plants and their structure-function relationship in comparison with other SOTs remained unclear. Here, we demonstrated that the SOT18 structure from *A. thaliana* shares a classical fold of soluble eukaryotic SOT proteins, including conserved catalytic residues, PAPS binding region and flexible loops surrounding the substrate binding site (Fig. 2). Hence, it can be speculated that many general SOT characteristics are also conserved in plants. A major difference between mammalian and plant SOTs seems to be the absence of a dimerization domain.

Analysing the binding sites for ligand binding residues revealed that the PAP binding residues are mainly located in the regions I-IV, which are conserved though all organisms (Fig. 3). High conservation of the PAPS binding site has been reported for previously solved structures from other organisms (Chapman et al., 2004) and could now also be confirmed for plant SOTs. Hence, it is plausible that residues responsible for the binding of various Gl substrates are located outside of the conserved regions. The high and low degrees of binding site conservation for the sulfate donor and acceptor, respectively, illustrates how SOTs easily adapt to different substrates in various organisms. In mammals, for example, SOTs are involved in detoxification, thus sulfating a broad range of compounds. They also perform various specific tasks, like the homeostatic control of signalling molecules, such as oxysterols and steroids like dehydroepiandrosterone (Mueller et al., 2015; Tibbs et al., 2015). In plants, they also fulfil a broad range of functions, some of general nature, such as AtSOT10 sulfating brassinosteroids (Marsolais et al., 2007), and more specific functions such as the choline-*O*-sulfate SOT from *Limonium* as part of a highly specialized salt stress response (Rivoal and Hanson, 1994). Ds-Gl SOTs, found only within the plant order Capparales, could be considered as a plant order specific SOT. In a previous study we analyzed ds-Gl SOTs in *Brassica napus* and identified a

new subgroup of SOT18-like enzymes, which did not show any activity with the tested ds-Gls (Hirschmann and Papenbrock, 2015). We speculated that due to genome triplication and allopolyploidization events in the evolutionary history from *A. thaliana* to *B. napus*, SOT18 might have undergone pseudo- (loss of function), or neofunctionalization (gain of new function). Regarding the solved AtSOT18 structure it is comprehensible that the natural variation of the identified catalytic residues or residues involved in PAPS binding are likely to result in pseudofunctionalization, while mutations in the substrate binding site, but also in the flexible loops, could lead to neofunctionalization.

Mechanism of SOTs and kinetics

Our kinetic data indicate that the sulfate group of PAPS does not contribute significantly to the binding energy of the donor. At the same time, the structure of the AtSOT18•PAP•sinigrin complex reveals H-bond interactions of sinigrin with Thr96 and Tyr130 stabilizing the sulfate group in the active site. Our mutagenesis study showed that disturbing these interactions leads to a dramatic loss of enzymatic activity (Table 1). This seeming contradiction is resolved by our AtSOT18•PAP complex structure, where the oxygen positions of the sulfate group are occupied by the oxygens of the solvent. The necessity to replace solvent molecules upon binding, the sulfate group leads to a near zero binding enthalpy balance, which explains the experimental kinetic data. The main contribution to the PAPS binding energy is thus provided by the PAP nucleoside moiety, which binds to a deep hydrophobic pocket between helices $\alpha 3$, $\alpha 13$ and $\alpha 16$.

Analysis of the catalytic center

The high-resolution structure of the AtSOT18 complex provided detailed information about the spatial arrangement and conformation of the residues of the catalytic center. Structural analysis suggests two major functions for the active site residues: providing the proximity and orientation effect for the substrate and sulfate donor and providing a direct effect on the catalytic reaction center by charge transfer and stabilization of the transition state geometry. The five residues of the catalytic center (Lys93, Thr96, Thr97, Tyr130, His155 (Fig. 5)) that stabilize the interface between the two substrates were tested by mutagenesis (Table 1). The dramatic

effect of alanine mutations at all these positions on enzymatic activity confirmed the importance of these residues for catalysis.

Based on these experiments and the structural conservation between AtSOT18 and human SOTs, we are able to adopt mechanistic information from human enzymes (Tibbs et al., 2015). In AtSOT18, after charge neutralization of PAPS by conserved residues upon binding, His155 abstracts a proton from the ds-Gl and the PAPS sulfur is attacked by the newly formed nucleophile. For completion of the reaction, the partial participation of the nucleophile leads to a charge build-up on the bridging oxygen, which then results in a shift of Lys93. This shift may then complete the sulfate dissociation from PAP. Transition state could be stabilized by hydrogen bond formation of sulfate oxygens of sinigrin with Thr96 and Tyr130 (Fig. 4a) and oxygens of the PAP 5'-phosphate with Thr96 and Thr97 (Fig. 4b)

Overall findings strongly indicate that the basic catalytic mechanism, regarding deprotonation of the respective substrate and sulfate dissociation from PAP by conserved histidine and lysine residues, respectively, is conserved in between mammals and *A. thaliana*.

The regulatory effect of PAP inhibition

The dissociation constants obtained from our two-dimensional titration experiment K_D (PAPS) = $4.2 \pm 2.6 \mu\text{M}$, K_I (PAP) = $3.0 \pm 1.7 \mu\text{M}$ of AtSOT18 are relatively high compared to human SOTs. For human SOT2A1, 22 individual rate constants were estimated considering a dead-end-complex formation with PAP (Wang et al., 2014). The dissociation constants determined by Wang et al. (2014) for PAPS were $0.2 \mu\text{M}$ and for PAP $0.3 \mu\text{M}$. Furthermore, the K_m values for PAPS were determined for various other human SOTs and ranged between $0.07 \mu\text{M}$ and $1.6 \mu\text{M}$ (Chapman et al., 2004).

By transferring Gl synthesis genes in Tobacco, thus enabling it to synthesize GlS, it could be shown that SOTs are not the bottleneck of synthesis. Instead it was stated that the PAPS supply could be the limiting step for Gl synthesis (Møldrup et al., 2011). A possible biological reason for the comparably low affinity of AtSOT18 for PAPS could be that there are three ds-Gl AtSOTs, hence a reduced affinity would preserve the limited PAPS pool. Since, GlS are transported from the cytoplasm into the vacuole and are only biologically active upon cell disruption, Gl biosynthesis could be considered a foresighted safety mechanism and not a fast immediate stress response. Hence, the limited PAPS would be more available for SOTs that are involved in immediate stress response, such as AtSOT12 and AtSOT15 using salicylic acid and

hydroxyjasmonate as substrates, respectively (Gidda et al., 2003; Baek et al., 2010), if they have a lower K_m value for PAPS than AtSOT18. Respective kinetic data have not been determined for these SOTs yet.

Further to the regulative functions of PAPS, also the co-product of SOT reaction PAP is considered to be a retrograde signal for induction of stress response (Estavillo et al., 2011; Chan et al., 2013). PAP is suggested to move into the nucleus where it inhibits RNA-degrading activity of 5'-3' exoribonucleases, which leads to the prevention of post transcriptional gene silencing of stress response genes. Further mutational studies of PAP catabolic genes led to accumulation of ds-Gls and lower levels of Gls. It was suggested that this is either caused by inhibition of PAPS transport or SOTs (Lee et al., 2012). Here we could demonstrate that ds-Gl SOTs are indeed inhibited by PAP.

The substrate specificity of ds-Gl SOTs

The substrate specificity of SOTs in general, including the ds-Gl SOTs is still hardly understood. SOTs in plants and mammals often have an overlapping substrate spectrum between each other, making it difficult to assign the enzyme's specific function. Various trials to group SOTs according to accepted substrates based on primary sequence analysis were without success (Hirschmann et al., 2014; Tibbs et al., 2015). Comparison of the AtSOT18 Gl binding site with the one in a homology model of AtSOT16 showed no obvious differences that would explain the distinctions in substrate specificity. Also, the extension of our search to the adjacent residues to the catalytic site could not explain the differences in the Gl binding affinities. Hence, we suggest that a specificity source of the ds-Gl SOTs might be provided by the non-conserved functional loops forming the Gl binding site, similar to the human enzymes (Allali-Hassani et al., 2007; Rohn et al., 2012; Rakers et al., 2016). However, substrate specificity cannot be entirely explained by the conformational properties of the gating loops. For example, for human SOT1A1 a molecular clamp mechanism was suggested, where two phenylalanine residues are repositioned in response to preferred substrates in such a way that stabilize the substrate's phenolic residue in a catalytic enhancing position (Cook et al., 2015). Even though we provided the first crystal structure of a plant SOT with bound ligands and complete electron density of the gating loops, further studies, including mutation of the loops, AtSOT16 crystallization and full molecular dynamics studies, are needed to achieve a comprehensive understanding of the ds-Gl SOT selectivity mechanisms.

Methods

Expression, purification and crystallization

The sequence encoding SOT18 from *Arabidopsis thaliana* (AtSOT18, At1g74090) was cloned into pQE-30 (Qiagen, Hilden, Germany) and expressed in *Escherichia coli* as described in Luczak et al. (2013). Mutagenesis was performed as described by Klein et al. (2006). The purification of recombinant AtSOT18 protein by affinity chromatography was performed according to Luczak et al. (2013) with modifications. An additional washing step with 0.12 M imidazole (20% buffer B + 80% buffer A; buffer B: 20 mM NaH₂PO₄, 0.5 M NaCl, 0.5 M imidazole, pH 7.4) was performed to obtain protein in a higher purity of up to 95%. The protein was dialyzed in 20 mM Tris/HCl, pH 8.0 plus 1 mM DTT for enzymatic assays or 20 mM HEPES, pH 8 for crystallization. Previously identified crystallization conditions (done by Prof. Dr. George N. Phillips, Jr., University of Wisconsin-Madison, USA) were used for further optimization by fine screens and additive screens. For the crystallization set-ups the concentrated protein sample was mixed gently with 4 mM PAP, and for the AtSOT18•PAP•sinigrin complex, with 4 mM PAP plus 4 mM sinigrin. During the complex formation the sample remained on ice for 45 min, followed by a centrifugation step at 21,000 x g for 30 min. Fine screening was performed in 24-well plates for hanging and sitting drop plates with a total reservoir volume of 500 µL and 200 µL, respectively. The total droplet size was 1.8 to 2.2 µL. Protein complex and reservoir solution were mixed in a ratio of 1:1. Additive screening was performed in a 96-well sitting drop plate by preparing the desired protein complex and reservoir solution, and mixing the reservoir with 10% volume of the Additive Screen HT™ - HR2-138 (Hampton Research, Aliso Viejo, USA).

Plate incubation and crystal growth documentation was performed using the Ministrel CrystalMation incubation and imaging system (Rigaku, Tokyo, Japan) for standard format plates (SBS format, Society for Biomolecular Screening). The measured crystals grew under 0.1 M MES pH 5.9, 16% PEG₄₀₀₀, 160 mM NaCl, and 4% 1,3-butanediol (AtSOT18•PAP•sinigrin complex) and 0.1 M MES pH 5.9, 16.5% PEG₄₀₀₀, 160 mM NaCl and 5% 1-propanol (AtSOT18•PAP complex) at 18°C.

Data collection and refinement

Before crystal harvest and freezing in liquid nitrogen the crystal was immersed in a cryoprotective solution (cryo buffer was identical to the reservoir conditions plus 20% ethylene glycol). The harvested crystals were irradiated by synchrotron radiation at the European Synchrotron Radiation Facility. Crystal and modeling parameters, and the refinement statistics are summarized in Table 2. The number of protein molecules in the asymmetric unit of the crystal was estimated using the Matthews coefficient in the Collaborative Computational Project No. 4 (CCP4) application `matthews_coef`. Molecular replacement was performed using the CCP4 application AMoRe. To solve the phase problem of the *de novo* generated data, the AtSOT12 model (PDB ID 1Q44) was used in AMoRe. The high and low resolution cutoff parameters and the search sphere radius were optimized for maximal significance of the rotation and translation function and the final fit of the search model. Using the application fast Fourier transformation, the initial electron density was calculated. Using the initial electron density and the known protein sequence of AtSOT18 the initial protein model was generated using ARP/warp Classic.

Table 2. Crystallographic data and refinement statistics for the AtSOT18•PAP•sinigrin and the AtSOT18•PAP complexes. Information on high resolution data referred to the outer 0.1 Å of the resolution shell.

	AtSOT18 complexed with	
	PAP, sinigrin	PAP, no acceptor
PDB code	to be submitted	to be submitted
Crystal Parameters		
space group	P 4 ₃ 2 ₁ 2	P 4 ₃ 2 ₁ 2
cell parameters		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	63.82, 63.82, 209.97	62.74, 62.74, 201.5
α , β , γ (deg)	90, 90, 90	90, 90, 90
Data collection		
ESRF beamline	ID23-1	ID23-1
wavelength (Å)	0.91	0.91
Crystal mosaicity (deg)	0.066	0.073
Wilson <i>B</i> -factor	37.246	32.727
resolution range (Å; total/high)	19.95 - 1.92 / 2.02 - 1.92	19.99 - 1.74 / 1.84 - 1.74
unique reflections (total/high)	34,239 / 4625	42,500 / 6394
completeness (total/high) %	99.63	99.9
$\langle I/\sigma(I) \rangle$ (total/high)	22.68 / 3.21	20.44 / 2.65
R_{sym} (total/high) %	0.065 / 0.544	0.093 / 0.753
Refinement statistics		
included amino acids	26 – 347	26 – 347
number of protein atoms	2652	2698
number of H ₂ O molecules	283	242
R_{cryst} / R_{free} %	16.8/21.7	16.7/20.8
RMS deviation for bonds (Å) / angles (deg)	0.02/2.047	0.022/2.145

Model building was performed using the Crystallographic object-oriented toolkit (COOT) V 0.7.2.1 and the CCP4 program suite V 6.3.0. For an automated overall refinement of the modeled AtSOT18 structures Refmac5 and ARP/warp Classic V 7.3.0 within the CCP4

package were used. For electron density map calculation, the refmac5 output was submitted to ARP/warp classic. The density map was recalculated within 15 iterations and thereafter used with the refmac5 protein coordinates in another manual refinement cycle in COOT. Protein visualization and analysis was performed using Pymol and Schrödinger Maestro. Chemical equations were designed in ChemDraw.

Homology modelling and NMA

Homology models were created in Schrödinger Prime, using AtSOT18 as template. NMA was performed with the eINémo web interface (Suhre and Sanejouand, 2004) and the respective movements were analyzed in Pymol.

Preparation of substrates

The ds forms of the parent GI derived from methionine and tryptophan were prepared as described by Graser et al. (2001). The following GI were used in the experiments in their ds forms: 3MTP (glucoiberberin) from *Erysimum pumillum*, 8MTO from *Arabis stelleri*, I3M (glucobrassicin) from *Isatis tinctoria*. 2-Propenyl GI (sinigrin) and PAP was bought commercially (Sigma-Aldrich, Taufkirchen, Germany). PAPS was obtained from Prof. H. R. Glatt, Institute of Human Nutrition, Berholz-Rehbruecke, Germany.

Enzyme activity measurements and analysis of the kinetic parameters

Enzymatic AtSOT18 assays were performed as described by Hirschmann and Papenbrock (2015). For inhibition experiments, a set of two-dimensional titration experiments was performed to determine the dissociation constants K_D of the donor PAPS and the K_I of the competitive inhibitor PAP. In the experiments the amounts of enzyme AtSOT18 and acceptor ds-3MTP were kept constant at 0.5 μg and 60 μM , respectively. The concentration of the sulfate donor PAPS was varied to 20, 40, 60, 80, and 100 μM and the inhibitor PAP was added at concentrations of 0, 20, 40, and 60 μM . Each reaction was prepared in triplicate. Two-dimensional non-linear least squares fit of the measured data with donor PAPS and inhibitor PAP as variables were performed. IgorPro V4.00 (WaveMetrics Inc., Lake Oswego, USA) was

used for nonlinear least squared Levenberg-Marquard fitting of the two dimensional experimental data (x as inhibitor concentration I_0 and y as donor concentration D_0) to the theoretical hyperbolic function derived from the model assumptions. The dissociation constants for the donor PAPS K_D , and the inhibitor PAP K_I and V_{max} were determined in an iterative minimization of the RMSD to the measured data. The dissociation constant K_A and concentration A_0 of the acceptor ds-3MTP was set to 55 μM according to literature (Klein and Papenbrock, 2009) and 60 μM due to the experimental procedure.

Acknowledgements

We would like to thank Julia Volker, Leibniz University Hannover, for valuable technical assistance. Initial screening for crystallization conditions was done by Prof. Dr. George N. Phillips, Jr., University of Wisconsin-Madison, USA, within the Protein Structure Initiative. We gratefully thank the staff scientists at the synchrotron beamline ID14-1, ESRF/Grenoble, for their assistance during diffraction data collection. This work was partially financed by the German Research Foundation (PA 764/10-1). RF was supported by the German Research Foundation (FE 1510/2-1).

Author contributions

FH prepared expression constructs. FH and FK purified recombinant proteins. FH and FK did kinetic measurements and IC did kinetic modeling. RF, PB, FH and FK performed crystallization trials. RF did the diffraction data collection, and RF and FK performed the structure refinement. FH built the homology models and performed NMA. JWM, DJM, JP and RF provided guidance on experimental design, data interpretation, and manuscript preparation.

References

- Allali-Hassani, A., Pan, P. W., Dombrowski, L., Najmanovich, R., Tempel, W., Dong, A., et al. (2007). Structural and chemical profiling of the human cytosolic sulfotransferases. *PLoS Biology* 5, e97. doi: 10.1371/journal.pbio.0050097
- Baek, D., Pathange, P., Chung, J.-S., Jiang, J., Gao, L., Oikawa, A., et al. (2010). A stress-inducible sulphotransferase sulphonates salicylic acid and confers pathogen resistance in *Arabidopsis*. *Plant, Cell & Environment* 33, 1383–1392. doi: 10.1111/j.1365-3040.2010.02156.x
- Bertl, E., Bartsch, H., Gerhauser, C. (2006). Inhibition of angiogenesis and endothelial cell functions are novel sulforaphane-mediated mechanisms in chemoprevention. *Molecular Cancer Therapeutics* 5, 575–585. doi: 10.1158/1535-7163.MCT-05-0324
- Cartea, M. E., and Velasco, P. (2008). Glucosinolates in Brassica foods: Bioavailability in food and significance for human health. *Phytochemical Reviews* 7, 213–229. doi: 10.1007/s11101-007-9072-2
- Chan, K. X., Wirtz, M., Phua, S. Y., Estavillo, G. M., Pogson, B. J. (2013). Balancing metabolites in drought: the sulfur assimilation conundrum. *Trends in Plant Science* 18, 18–29. doi: 10.1016/j.tplants.2012.07.005
- Chapman, E., Best, M. D., Hanson, S. R., and Wong, C.-H. (2004). Sulfotransferases: structure, mechanism, biological activity, inhibition, and synthetic utility. *Angewandte Chemie (International ed. in English)* 43, 3526–3548. doi: 10.1002/anie.200300631
- Clarke, D. B. (2010). Glucosinolates, structures and analysis in food. *Analytical Methods* 2, 310. doi: 10.1039/b9ay00280d
- Cook, I., Wang, T., Leyh, T. S. (2015). Sulfotransferase 1A1 Substrate Selectivity: A Molecular Clamp Mechanism. *Biochemistry* 54, 6114–6122. doi: 10.1021/acs.biochem.5b00406
- Estavillo, G. M., Crisp, P. A., Pornsiriwong, W., Wirtz, M., Collinge, D., Carrie, C., et al. (2011). Evidence for a SAL1-PAP chloroplast retrograde pathway that functions in drought and high light signaling in *Arabidopsis*. *The Plant Cell* 23, 3992–4012. doi: 10.1105/tpc.111.091033
- Fahey, J. W., and Talalay, P. (1999). Antioxidant functions of sulforaphane: a potent inducer of Phase II detoxication enzymes. *Food and Chemical Toxicology* 37, 973–979.
- Faulkner, K., Mithen, R., Williamson, G. (1998). Selective increase of the potential anticarcinogen 4-methylsulphinylbutyl glucosinolate in broccoli. *Carcinogenesis* 19, 605–609. doi: 10.1093/carcin/19.4.605

- Gidda, S. K., Miersch, O., Levitin, A., Schmidt, J., Wasternack, C., Varin, L. (2003). Biochemical and molecular characterization of a hydroxyjasmonate sulfotransferase from *Arabidopsis thaliana*. *The Journal of Biological Chemistry* 278, 17895–17900. doi: 10.1074/jbc.M211943200
- Glendening, T. M., and Poulton, J. E. (1990). Partial Purification and Characterization of a 3'-Phosphoadenosine 5' -Phosphosulfate: Desulfoglucosinolate Sulfotransferase from Cress (*Lepidium sativum*). *Plant Physiology* 94, 811–818.
- Graser, G., Oldham, N. J., Brown, P. D., Temp, U., Gershenzon, J. (2001). The biosynthesis of benzoic acid glucosinolate esters in *Arabidopsis thaliana*. *Phytochemistry* 57, 23–32.
- Gross-Steinmeyer, K., Stapleton, P. L., Tracy, J. H., Bammler, T. K., Lehman, T., Strom, S. C., et al. (2005). Influence of Matrigel-overlay on constitutive and inducible expression of nine genes encoding drug-metabolizing enzymes in primary human hepatocytes. *Xenobiotica* 35, 419–438. doi: 10.1080/00498250500137427
- Gupta, P., Kim, B., Kim, S.-H., Srivastava, S. K. (2014). Molecular targets of isothiocyanates in cancer: Recent advances. *Molecular Nutrition & Food Research* 58, 1685–1707. doi: 10.1002/mnfr.201300684
- Hirai, M. Y., Klein, M., Fujikawa, Y., Yano, M., Goodenowe, D. B., Yamazaki, Y., et al. (2005). Elucidation of gene-to-gene and metabolite-to-gene networks in *Arabidopsis* by integration of metabolomics and transcriptomics. *The Journal of biological chemistry* 280, 25590–25595. doi: 10.1074/jbc.M502332200
- Hirschmann, F., Krause, F., Papenbrock, J. (2014). The multi-protein family of sulfotransferases in plants: composition, occurrence, substrate specificity, and functions. *Frontiers in Plant Science* 5, 556. doi: 10.3389/fpls.2014.00556
- Hirschmann, F., and Papenbrock, J. (2015). The fusion of genomes leads to more options: A comparative investigation on the desulfo-glucosinolate sulfotransferases of *Brassica napus* and homologous proteins of *Arabidopsis thaliana*. *Plant Physiology and Biochemistry* 91, 10–19. doi: 10.1016/j.plaphy.2015.03.009
- Hoff, R. H., Czyryca, P. G., Sun, M., Leyh, T. S., Hengge, A. C. (2006). Transition State of the Sulfuryl Transfer Reaction of Estrogen Sulfotransferase. *Journal of Biological Chemistry* 281, 30645–30649. doi: 10.1074/jbc.M604205200
- Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Research* 38, W545-9. doi: 10.1093/nar/gkq366
- Kakuta, Y., Pedersen, L. G., Carter, C. W., Negishi, M., Pedersen, L. C. (1997). Crystal structure of estrogen sulphotransferase. *Nature Structural Biology* 4, 904–908.

- Kakuta, Y., Petrotchenko, E. V., Pedersen, L. C., Negishi, M. (1998). The Sulfuryl Transfer Mechanism: Crystal structure of a vanadate complex of estrogen sulfotransferase and mutational analysis. *Journal of Biological Chemistry* 273, 27325–27330. doi: 10.1074/jbc.273.42.27325
- Klein, M., and Papenbrock, J. (2004). The multi-protein family of *Arabidopsis* sulphotransferases and their relatives in other plant species. *Journal of Experimental Botany* 55, 1809–1820. doi: 10.1093/jxb/erh183
- Klein, M., and Papenbrock, J. (2009). Kinetics and substrate specificities of desulfo-glucosinolate sulfotransferases in *Arabidopsis thaliana*. *Physiologia Plantarum* 135, 140–149. doi: 10.1111/j.1399-3054.2008.01182.x
- Klein, M., Reichelt, M., Gershenzon, J., Papenbrock, J. (2006). The three desulfoglucosinolate sulfotransferase proteins in *Arabidopsis* have different substrate specificities and are differentially expressed. *The FEBS Journal* 273, 122–136. doi: 10.1111/j.1742-4658.2005.05048.x
- Konwinski, R. R., Haddad, R., Chun, J. A., Klenow, S., Larson, S. C., Haab, B. B., et al. (2004). Oltipraz, 3H-1,2-dithiole-3-thione, and sulforaphane induce overlapping and protective antioxidant responses in murine microglial cells. *Toxicology Letters* 153, 343–355. doi: 10.1016/j.toxlet.2004.06.006
- Laurberg, P., Andersen, S., Knudsen, N., Ovesen, L., Nøhr, S. B., Pedersen, I. (2002). Thiocyanate in food and iodine in milk: from domestic animal feeding to improved understanding of cretinism. *Thyroid: Official Journal of the American Thyroid Association* 12, 897–902. doi: 10.1089/105072502761016520
- Lee, B.-R., Huseby, S., Koprivova, A., Chételat, A., Wirtz, M., Mugford, S. T., et al. (2012). Effects of fou8/fry1 mutation on sulfur metabolism: is decreased internal sulfate the trigger of sulfate starvation response? *PloS One* 7, e39425. doi: 10.1371/journal.pone.0039425
- Lu, J., Li, H., Zhang, J., Li, M., Liu, M.-Y., An, X., et al. (2010). Crystal structures of SULT1A2 and SULT1A1*3: Insights into the substrate inhibition and the role of Tyr149 in SULT1A2. *Biochemical and Biophysical Research Communications* 396, 429–434. doi: 10.1016/j.bbrc.2010.04.109
- Lu, J.-H., Li, H.-T., Liu, M.-C., Zhang, J.-P., Li, M., An, X.-M., et al. (2005). Crystal structure of human sulfotransferase SULT1A3 in complex with dopamine and 3'-phosphoadenosine 5'-phosphate. *Biochemical and Biophysical Research Communications* 335, 417–423. doi: 10.1016/j.bbrc.2005.07.091

- Luczak, S., Forlani, F., Papenbrock, J. (2013). Desulfo-glucosinolate sulfotransferases isolated from several *Arabidopsis thaliana* ecotypes differ in their sequence and enzyme kinetics. *Plant Physiology and Biochemistry* 63, 15–23. doi: 10.1016/j.plaphy.2012.11.005
- Malojčić, G., Owen, R. L., Glockshuber, R. (2014). Structural and Mechanistic Insights into the PAPS-Independent Sulfotransfer Catalyzed by Bacterial Aryl Sulfotransferase and the Role of the DsbL/DsbI System in its Folding. *Biochemistry* 53, 1870–1877. doi: 10.1021/bi401725j
- Marsolais, F., Boyd, J., Paredes, Y., Schinas, A.-M., Garcia, M., Elzein, S., et al. (2007). Molecular and biochemical characterization of two brassinosteroid sulfotransferases from *Arabidopsis*, AtST4a (At2g14920) and AtST1 (At2g03760). *Planta* 225, 1233–1244. doi: 10.1007/s00425-006-0413-y
- Møldrup, M. E., Geu-Flores, F., Olsen, C. E., Halkier, B. A. (2011). Modulation of sulfur metabolism enables efficient glucosinolate engineering. *BMC Biotechnology* 11, 12. doi: 10.1186/1472-6750-11-12
- Mueller, J. W., Gilligan, L. C., Idkowiak, J., Arlt, W., Foster, P. A. (2015). The Regulation of Steroid Action by Sulfation and Desulfation. *Endocrine Reviews* 36, 526–563. doi: 10.1210/er.2015-1036
- Pakhomova, S., Buck, J., Newcomer, M. E. (2005). The structures of the unique sulfotransferase retinol dehydratase with product and inhibitors provide insight into enzyme mechanism and inhibition. *Protein Science* 14, 176–182. doi: 10.1110/ps.041061105
- Petrochenko, E. V., Pedersen, L. C., Borchers, C. H., Tomer, K. B., Negishi, M. (2001). The dimerization motif of cytosolic sulfotransferases. *FEBS Letters* 490, 39–43. doi: 10.1016/S0014-5793(01)02129-9
- Piotrowski, M., Schemenewitz, A., Lopukhina, A., Müller, A., Janowitz, T., Weiler, E. W., et al. (2004). Desulfoglucosinolate sulfotransferases from *Arabidopsis thaliana* catalyze the final step in the biosynthesis of the glucosinolate core structure. *Journal of Biological Chemistry* 279, 50717–50725. doi: 10.1074/jbc.M407681200
- Rakers, C., Schumacher, F., Meinl, W., Glatt, H., Kleuser, B., Wolber, G. (2016). *In Silico* Prediction of Human Sulfotransferase 1E1 Activity Guided by Pharmacophores from Molecular Dynamics Simulations. *Journal of Biological Chemistry* 291, 58–71. doi: 10.1074/jbc.M115.685610
- Rausch, T., and Wachter, A. (2005). Sulfur metabolism: a versatile platform for launching defence operations. *Trends in Plant Science* 10, 503–509. doi: 10.1016/j.tplants.2005.08.006

- Rivoal, J., and Hanson, A. D. (1994). Choline-O-Sulfate Biosynthesis in Plants (Identification and Partial Characterization of a Salinity-Inducible Choline Sulfotransferase from Species of *Limonium* (*Plumbaginaceae*). *Plant Physiology* 106, 1187–1193.
- Rohn, K. J., Cook, I. T., Leyh, T. S., Kadlubar, S. A., Falany, C. N. (2012). Potent Inhibition of Human Sulfotransferase 1A1 by 17 β -Ethinylestradiol: Role of 3'-Phosphoadenosine 5'-Phosphosulfate Binding and Structural Rearrangements in Regulating Inhibition and Activity. *Drug Metabolism and Disposition* 40, 1588–1595. doi: 10.1124/dmd.112.045583
- Smith, D. W., Johnson, K. A., Bingman, C. A., Aceti, D. J., Blommel, P. G., Wrobel, R. L., et al. (2004). Crystal structure of At2g03760, a putative steroid sulfotransferase from *Arabidopsis thaliana*. *Proteins* 57, 854–857. doi: 10.1002/prot.20258
- Suhre, K., and Sanejouand, Y.-H. (2004). ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Research* 32, 4. doi: 10.1093/nar/gkh368
- Teramoto, T., Sakakibara, Y., Liu, M.-C., Suiko, M., Kimura, M., Kakuta, Y. (2009). Snapshot of a Michaelis complex in a sulfuryl transfer reaction: Crystal structure of a mouse sulfotransferase, mSULT1D1, complexed with donor substrate and acceptor substrate. *Biochemical and Biophysical Research Communications* 383, 83–87. doi: 10.1016/j.bbrc.2009.03.146
- Tibbs, Z. E., Rohn-Glowacki, K. J., Crittenden, F., Guidry, A. L., Falany, C. N. (2015). Structural plasticity in the human cytosolic sulfotransferase dimer and its role in substrate selectivity and catalysis. *Drug Metabolism and Pharmacokinetics* 30, 3–20. doi: 10.1016/j.dmpk.2014.10.004
- Varin L., de Luca, V., Ibrahim, R.K., Brisson, N. (1992). Molecular characterization of two plant flavonol sulfotransferases. *Proceedings of the National Academy of Sciences of the United States of America* 1992, 89:1286–1290.
- Wang, L., Liu, D., Ahmed, T., Chung, F. L., Conaway, C., Chiao, J. W. (2004). Targeting cell cycle machinery as a molecular mechanism of sulforaphane in prostate cancer prevention. *International Journal of Oncology* 24, 187–192.
- Wang, T., Cook, I., Falany, C. N., Leyh, T. S. (2014). Paradigms of sulfotransferase catalysis: the mechanism of SULT2A1. *The Journal of Biological Chemistry* 289, 26474–26480. doi: 10.1074/jbc.M114.573501

Supplementary Material

Table S1. Summary of the amino acids putatively involved in sinigrin and PAP binding, as well as catalysis.

Proposed residue / possible role of the amino acid

E54 / substrate binding (dsG1)

E193 / substrate binding (dsG1)

R51 / substrate binding (dsG1)

Y306 / substrate binding (dsG1)

L190 / substrate binding (dsG1)

F189 / substrate binding (dsG1)

M186 / substrate binding (dsG1)

Y311 / substrate binding (dsG1)

V305 / substrate binding (dsG1)

L65 / substrate binding (dsG1)

S67 / substrate binding (dsG1)

L68 / substrate binding (dsG1)

I133 / substrate binding (dsG1)

K93 / stabilization of transition state, proton or electron transfer

Y130 / stabilization of transition state, proton or electron transfer

H155 / stabilization of transition state, proton or electron transfer

T97 / stabilization of transition state, proton or electron transfer

T96 / stabilization of transition state, proton or electron transfer

R177 / substrate binding (PAPS), proton or electron transfer

S185 / substrate binding (PAPS), proton or electron transfer

R313 / substrate binding (PAPS), proton or electron transfer

F284 / substrate binding (PAPS)

R247 / substrate binding (PAPS), proton or electron transfer

W98 / substrate binding (PAPS)

Y243 / substrate binding (PAPS), proton or electron transfer

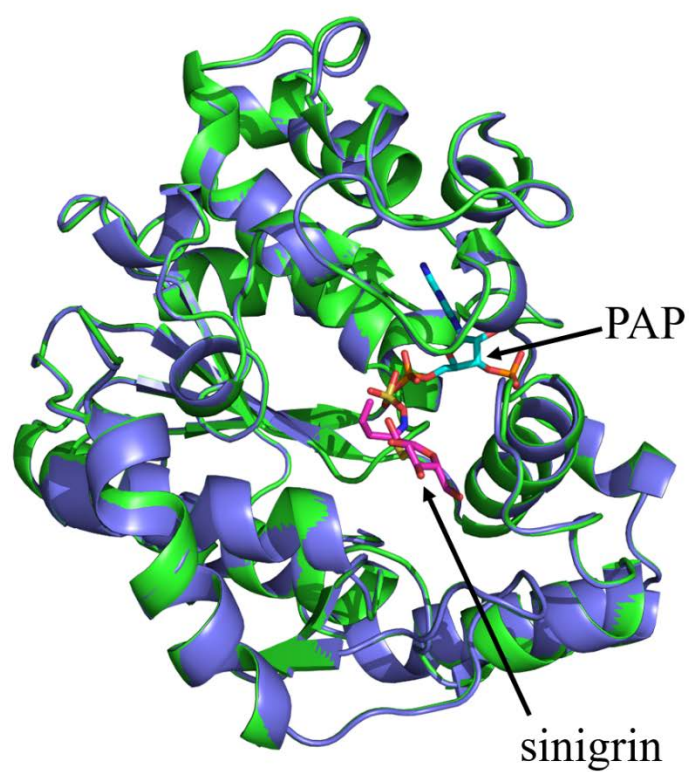


Fig.S1. Superimposition of overall structures of AtSOT18•sinigrin•PAP (green) and AtSOT18•PAP (blue).

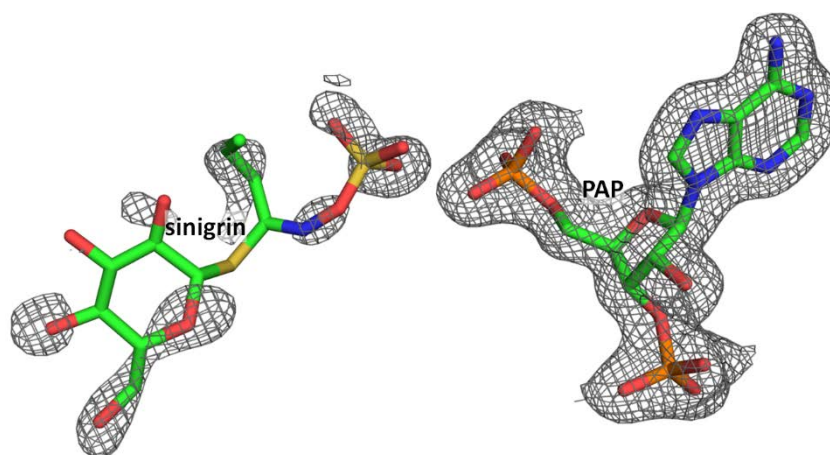


Fig. S2. Ligand electron density of PAP and sinigrin.

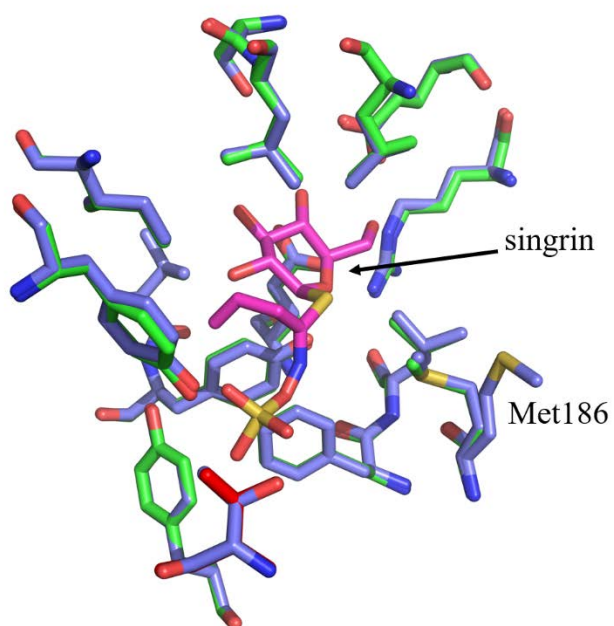


Fig. S3. Comparison of the AtSOT18•PAP•sinigrin (green) and AtSOT18•PAP (blue) complex' sinigrin (magenta) binding site. All sinigrin surrounding residues are in both complexes in the same position, except for Met186, which shows a double confirmation in the AtSOT18•PAP (blue) complex.

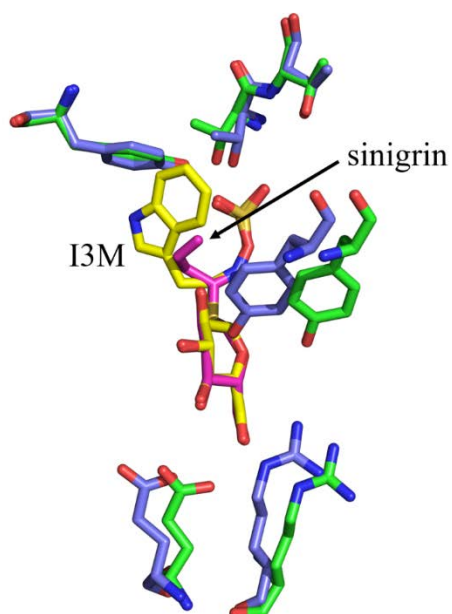


Fig. S4. Structural alignment of the sinigrin (magenta) binding residues of the AtSOT18•PAP•sinigrin (green) and AtSOT16 homology model (blue) with superimposed indol-3-ylmethylglucosinolate (I3M).

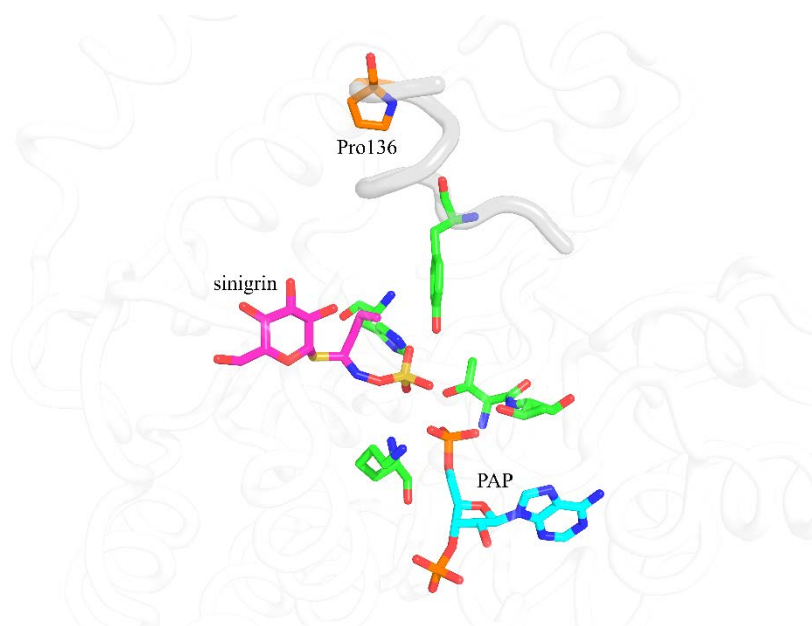


Fig. S5. Location of Pro136 (orange) in $\alpha 6$ in the second coordination sphere, in relation to the active site residues (green) with sinigrin (magenta) and PAP (cyan).

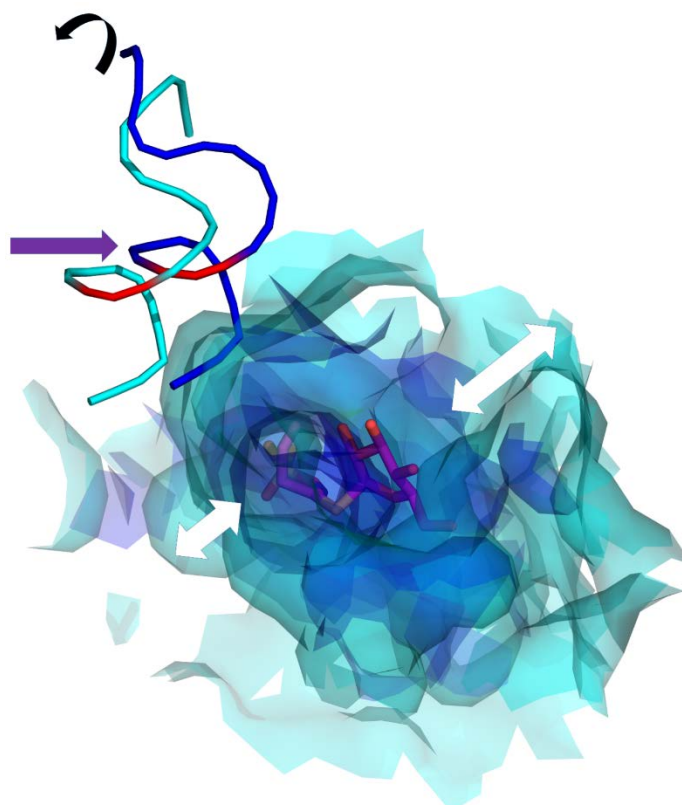


Fig. S6. Visualization of Normal Mode Analyses. The movement of loop 1 localized $\alpha 6$ directly alters the volume of the GI binding site (white arrows). From open (cyan) to closed state (blue) AtSOT18 $\alpha 6$ undergoes a simple shift movement (purple arrow), while AtSOT18 Pro136Ala (red) showed a bending deformation (black arrow).

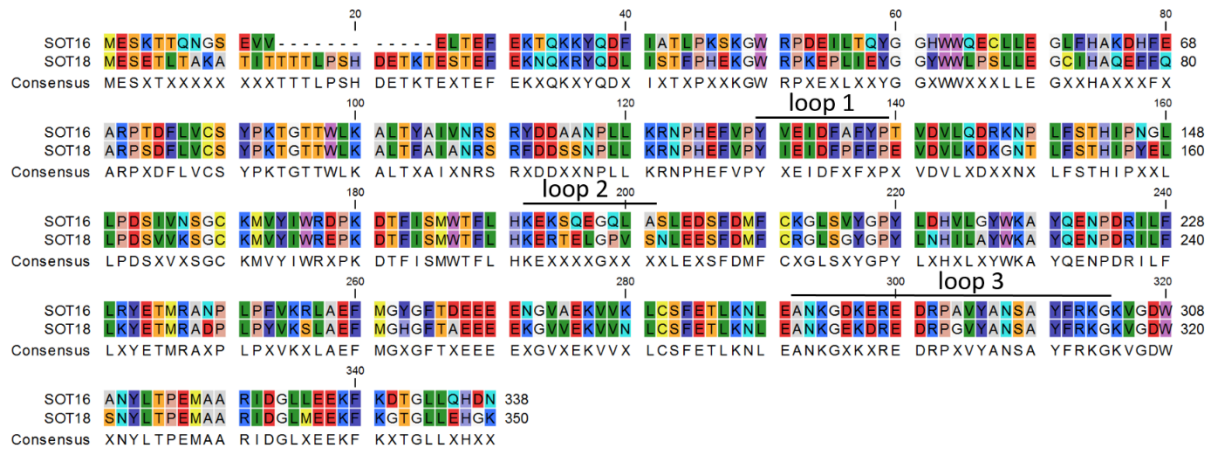


Fig. S7. Alignment of the AtSOT16 and AtSOT18 protein sequences. The three flexible loops forming the ds-GI binding site are highlighted.

CHAPTER 5

General Discussion

Combination of structure and enzyme assay results

The sequence-based identification of SOT substrates is due to their high similarity and overlapping substrate specificities almost impossible and has led in plants (Chapter 2) and mammals (Tibbs et al., 2015) to false predictions. In *A. thaliana*, three ds-Gl SOTs have been identified and characterized that showed different substrate specificities among those (Klein and Papenbrock, 2009). The main aims of this thesis were the sequence-based identification of ds-Gl SOT substrate specificities in *B. napus* (Chapter 3) and the molecular explanation for substrate binding, catalysis and substrate selectivity (Chapter 4). To test whether the sequence-based prediction is possible, *B. napus* was chosen due to its economic relevance. In total, 71 putative functional SOTs were identified in the allotetraploid *B. napus* and eleven of those are likely to encode ds-Gl SOTs. Interestingly, a new subgroup of SOT18-like enzymes could be identified. Several of the newly identified ds-Gl BnSOTs were recombinantly expressed and purified. By *in vitro* characterization it could be demonstrated that the SOT16 and SOT17 homologs show similar substrate specificities as their *A. thaliana* counterparts. These results show, that it is not only principally possible to identify ds-Gls as substrates, but also to predict if they prefer aliphatic or indolic ds-Gls. Surprisingly, the tested BnSOT18-a and BnSOT18-b did not show any activity. It was hypothesised that the inactivity is either caused by wrong folding during the recombinant expression or by natural variation.

By solving the 3D structure of AtSOT18 fundamental residues for catalysis and substrate binding could be identified. Five catalytic residues were identified in AtSOT18 and their mutation to alanine led to inactivation or strong reduction of activity (Chapter 4). Sequence alignments of AtSOT18 and the experimentally analysed BnSOT18s now reveal more detailed explanations for the inactivity (Fig. 1). In BnSOT18-a two of the five catalytic residues (Thr96Ser and Tyr130Phe, AtSOT18 residue numbering) and one Gl-binding residue (Glu54Ser) differ compared to AtSOT18. Hence, it can be speculated that the differences in the BnSOT18-a catalytic residues explain its inactivity, even though the substituting residues are chemically quite similar. In BnSOT18-b, Arg247 (AtSOT18 residue numbering), involved in PAPS binding is replaced by a serine. This substitution could also already inactivate the SOT, as Arg247 stabilises the PAPS adenine group by π - π stacking effects, which is a conserved characteristic in PAPS binding (Chapman et al., 2004).

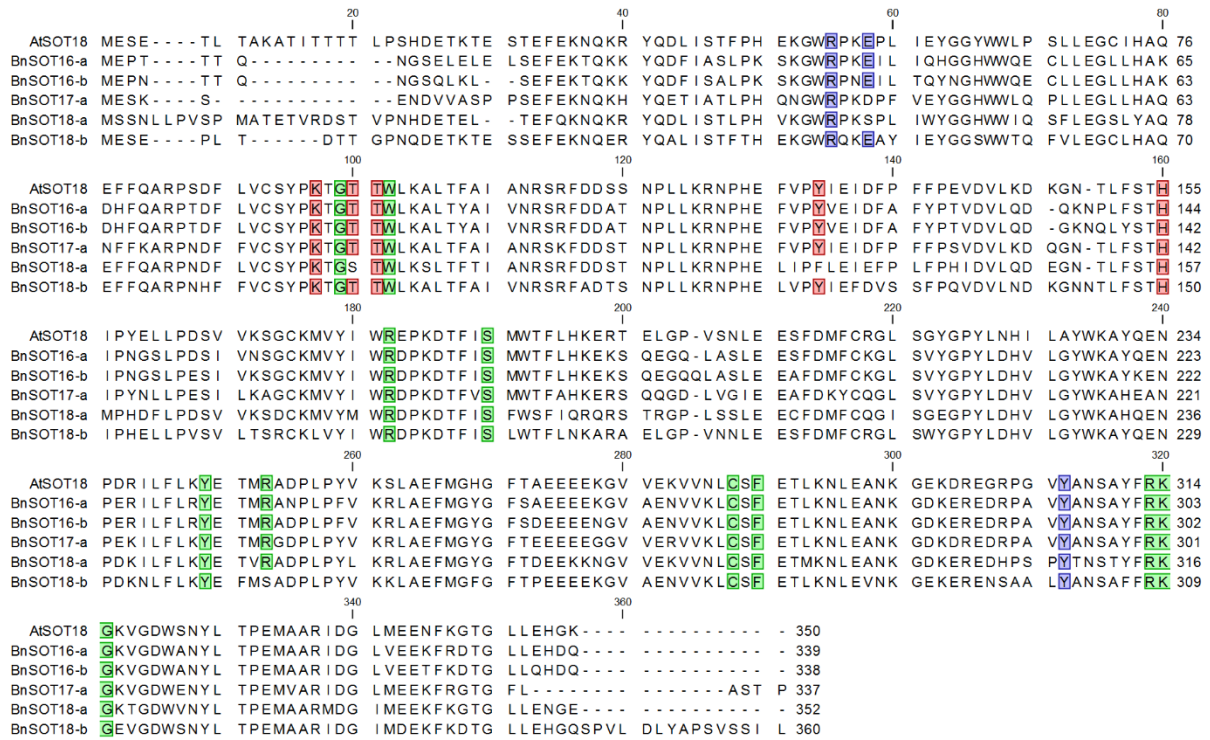


Fig. 1. Alignment of AtSOT18 and the analysed BnSOTs. Residues that were identified by AtSOT18 structure analysis to be involved in catalysis (red), PAPS-binding (green) and GI-binding (blue).

On the other hand, in the active BnSOT17-a, only a glutamic acid residue (54, AtSOT18 residue numbering) involved in GI-binding is replaced by the very similar aspartic acid. Therefore, this replacement is not decisive enough to inactivate the SOT. Furthermore, the active BnSOT16-a and BnSOT16-b do not show any differences to AtSOT18 in the catalytic or substrate binding residues.

Overall, the newly gained AtSOT18 structural information helps to explain results from BnSOT enzyme assays. Hence, the remaining BnSOTs, as well as *B. rapa* and *B. oleracea* SOTs, should be analysed in consideration of the new structural information to see what candidates are likely to be active or inactive. At the same time, enzyme assay (Chapter 3) results support conclusions drawn from the structural analysis (Chapter 4).

Are ds-GI SOTs a good target for breeding?

In this work it was demonstrated that ds-GI SOTs from *B. napus* have similar substrate preferences as their *A. thaliana* homologs (Chapter 3). It was also shown that by sequence analysis, supported by structural data, it is principally possible to predict their preferred

substrates. Furthermore, the analysis of AtSOT18 revealed, which parts of the protein must be mutated for inactivation or changes in substrate specificity (Chapter 4). This leads to the question, how these new insights might help in future breeding programs? Plants could either be bred towards an increase or decrease of total Gl content, a subgroup of GlS, like aliphatic or indolic GlS or a single Gl, like the anticarcinogenic glucoraphanin and glucobrassicin.

In *A. thaliana*, at least 40 genes are involved in Gl biosynthesis (Sønderby et al., 2010) and the ones for manipulation should be chosen depending on the desired Gl alteration. For the in- or decrease of total Gl content, manipulation of ds-Gl SOTs seems inapplicable, because they catalyse only one of many steps in Gl biosynthesis. Instead, it seems more plausible to up- or downregulate transcription factors, which control many Gl biosynthesis genes at the same time. The best characterized Gl transcription factors are members of the MYB family, whose impact on Gl concentration was already demonstrated by overexpression, RNAi and T-DNA insertion studies (Celenza et al., 2005; Gigolashvili et al., 2007b; Gigolashvili et al., 2007a; Hirai et al., 2007; Sønderby et al., 2007; Gigolashvili et al., 2008; Malitsky et al., 2008; Yatusевич et al., 2010). Interestingly, they can be divided into two groups, controlling either aliphatic or indolic Gl biosynthesis. Hence, the manipulation of MYBs enables the alteration of total Gl concentration, as well as the aliphatic or indolic Gl composition, by manipulation of a whole cascade of Gl biosynthesis genes, instead of just a single catalysis step. This becomes even more relevant, when the Gl content of crops and vegetables shall be altered. In this work (Chapter 3), five ds-Gl SOTs were identified in *B. oleracea*, seven in *B. rapa* and eleven in *B. napus*. Considering the overlapping substrate specificities, the control of all homologs seems rather complicated, even with modern biotechnological methods. Furthermore, *in vivo* ds-Gl SOT catalysis was shown to be limited by the PAPS supply (Mugford et al., 2009; Møldrup et al., 2011). This limitation could also be overcome by MYB factor manipulation, since they also regulate genes of the primary sulfur metabolism (Sønderby et al., 2007; Malitsky et al., 2008; Yatusевич et al., 2010), hence a sufficient PAPS concentrations could be provided.

Overall, MYB transcription factors are a more promising target for breeding crops with altered Gl concentrations. Nevertheless, ds-Gl SOTs could be interesting for breeding plants for pharming purposes. In pharming (a coinage of farming and pharmaceutical) plants are genetically engineered to biosynthesize valuable compounds that are either impossible or too expensive to synthesize with classical chemical or biotechnological procedures. Engineering plants to biosynthesize anticarcinogenic GlS like glucoraphanin or glucobrassicin could be an interesting field in the future. A traditionally bred broccoli variety called Beneforté, with 2.5 - 3 times higher glucoraphanin concentrations than standard hybrids (Traka et al., 2013), has

already been introduced to the British, Swedish and Finnish market. Furthermore, by metabolic engineering, the non-cruciferous plant *Nicotiana benthamiana* has already been enabled to biosynthesize Gls (Geu-Flores et al., 2009; Møldrup et al., 2011). Logically, for the biosynthesis of the indolic glucobrassicin, SOT16 should be transferred to the plant. However, the artificial biosynthesis of aliphatic Gls, like glucoraphanin, with a specific side chain length would be more challenging. The number of side chain elongation steps is controlled by the interplay of three *2-isopropylmalate synthase* genes, which is not entirely understood (Jensen et al., 2014). Hence, the side chain length of the final product must be regulated in a followed biosynthetic step. SOT17 and SOT18 could be possible candidates, but only if their specificities were further improved. Even though SOT17 and SOT18 have substrate preferences, they still accept all aliphatic Gls (Klein and Papenbrock, 2009). However, the SOT18 structure could help to create a ds-Gl SOT that is highly specific for a certain chain length. In this work (Chapter 4), it was proposed that indolic and aliphatic Gl distinction is controlled by the dynamics of loop 1, providing the basis for further research regarding the ds-Gl SOT specificity. Since the respective hypothesis could not be confirmed by mutational studies, it was speculated that the specificity in between aliphatic Gls could be controlled by the interplay of all three flexible loops. This could be further tested by solving more structures with different substrates, combined with full molecular dynamics studies and advanced modelling and docking. By applying these methods a deeper understanding of human SOT 1A1, 2A1 and 1E1 was provided, which even enabled substrate prediction from a 1445-compound library without any false predictions (Cook et al., 2013c; Rakers et al., 2016). Nonetheless, the artificial control of SOT specificity would be a very ambiguous goal, for which not only protein structure and mechanism have to be considered, but also substrate, product, PAP and PAPS concentrations.

The influence of PAPS and PAP concentrations on SOTs

SOT inhibition by PAP is a common characteristic (Whittemore et al., 1985; Zhang et al., 1998; Cook et al., 2009; Wang et al., 2014) that has now also been confirmed for AtSOT18 (Chapter 4). Studies on mammalian SOTs indicate that PAP release is the rate-limiting step (Wang et al., 2014). Furthermore, the analysis of several human SOTs revealed a number of compounds that only bind in the presence or absence of PAP. For example, hSOT1B1 only binds 1-naphtol in the presence of PAP, while it binds apomorphine only in its absence (Allali-Hassani et al., 2007). Further studies on mammalian SOTs showed that loop 3, the largest of three flexible loops that overlays the PAPS and substrate binding site, undergoes structural shifts to a closed

state upon PAPS binding, which results in alteration of the active site, hence affecting substrate specificity (Cook et al., 2010). Further molecular dynamics and experimental binding studies suggested that SOTs isomerize between open and closed states and the closed state prevents larger substrates from binding (Cook et al., 2013a, 2013b).

Considering the high similarities between mammalian SOTs and AtSOT18, regarding the overall structure, binding and catalytic sites, as well as PAP inhibition, as shown in this work (Chapter 4), similar effects of PAP and PAPS could also be assumed in plant SOTs. This would have drastic consequences for the interpretation of *in vitro* results, as the *in vivo* PAP and PAPS concentrations would not only regulate SOT activity (Mugford et al., 2009), but also specificity. Hence, in order to understand *in vivo* sulfation, a deep knowledge about the cytosolic PAPS and PAP concentration would be required.

PAPS and PAP concentrations in *A. thaliana* were reported in previous studies, but the results vary, probably due to different growing and harvesting conditions as well as measuring techniques. In Estavillo et al. (2011), 1.1 pmol mg⁻¹ fresh weight and 0.6 pmol mg⁻¹ fresh weight PAPS and PAP, respectively, were reported, while Lee et al. (2012) measured 0.6 nmol PAPS mg⁻¹ fresh weight and < 0.01 nmol PAP mg⁻¹ fresh weight. Despite the big differences in the reports, it also has to be considered that the concentrations in whole leaf tissue were measured and therefore no information is available about the actual cytosolic concentrations.

Assuming that AtSOT18 inhibition leads to a higher preference of smaller substrates, as it was shown for human SOTs (Allali-Hassani et al., 2007), it could be speculated that *in vitro* preference of long-chained over short-chained aliphatic Gl is less distinct *in vivo*. The effect of PAP on ds-Gl SOT substrate specificity could further be investigated by structural approaches combined with enzyme assays, as it has already been done for mammalian SOTs (Cook et al., 2013a, 2013b).

Regarding the impact of PAPS and PAP on SOT activity and possibly also on the specificity, a vast amount of factors have to be taken into account in order to understand the regulation of sulfation. First of all the PAPS supply has to be considered, and therefore the connection to sulfur assimilation and the primary sulfur metabolism. After sulfate is taken up and activated to APS, it can either be used for the formation of sulfite in the primary sulfur metabolism or for PAPS synthesis in the secondary sulfur metabolism (summarised in Abuelsoud et al., 2016). The multiple APS kinase isoforms, catalysing PAPS formation, are localized in the cytosol and plastids. Interestingly, the knockout of the cytosolic APS kinase did not lead to any changes in the phenotype and only a minor contribution to the PAPS synthesis was concluded (Mugford

et al., 2009). Hence, it was stated that PAPS synthesis mainly takes place in the plastids. Logically, the PAPS has to be transported out of the plastids into the cytosol, where the SOT-catalysed sulfation takes place (Klein and Papenbrock, 2004). Therefore, PAPS transport is another factor to be considered regarding *in vivo* SOT activity. One plastidic PAPS transporter has already been described, but the existence of more has been postulated (Gigolashvili et al., 2012).

Furthermore, the catabolism and function of PAP have to be taken into account. After PAP formation, it is probably transported into the plastids in exchange for PAPS (Gigolashvili et al., 2012), where it is metabolized to AMP by a 3'(2'),5'-bisphosphate nucleotidase (Wilson et al., 2009). Knockout of the respective gene leads to a large number of pleiotrophic phenotypes (Chan et al., 2013), as it also encodes an inositol polyphosphate 1-phosphatase (Wilson et al., 2009). Changes in PAP lead to accumulation of ds-Gls and lower levels of Gls (Lee et al., 2012), which could be explained by our results regarding inhibition of SOT by PAP (Chapter 4). PAP is also considered to be a retrograde signal in stress response (Estavillo et al., 2011), altering microRNA processing (Gy et al., 2007; Hirsch et al., 2011), which illustrates its co-functionality with SOTs, as they are also involved in stress responses.

Overall, *in vitro* SOT assays form the basis for an understanding of sulfation, but regarding its *in vivo* function and significance, many other factors have to be considered, as SOTs are only one part of a big network.

References

- Abuelsoud, W. A., Hirschmann, F., Papenbrock, J. (2016). “Sulfur metabolism and drought tolerance in plants“ in *Drought Stress Tolerance in Plants, Vol 1: Physiology and Biochemistry*, ed. M. A. Hossain, S. H. Wani, S. Bhattachajee, D. J. Burritt, and L. S. Phan Trans (New York, NY: Springer US)
- Allali-Hassani, A., Pan, P. W., Dombrowski, L., Najmanovich, R., Tempel, W., Dong, A., et al. (2007). Structural and chemical profiling of the human cytosolic sulfotransferases. *PLoS Biology* 5, e97. doi: 10.1371/journal.pbio.0050097
- Celenza, J. L., Quiel, J. A., Smolen, G. A., Merrikkh, H., Silvestro, A. R., Normanly, J., et al. (2005). The *Arabidopsis* ATR1 Myb transcription factor controls indolic glucosinolate homeostasis. *Plant Physiology* 137, 253–262. doi: 10.1104/pp.104.054395
- Chan, K. X., Wirtz, M., Phua, S. Y., Estavillo, G. M., Pogson, B. J. (2013). Balancing metabolites in drought: the sulfur assimilation conundrum. *Trends in Plant Science* 18, 18–29. doi: 10.1016/j.tplants.2012.07.005
- Chapman, E., Best, M. D., Hanson, S. R., Wong, C.-H. (2004). Sulfotransferases: structure, mechanism, biological activity, inhibition, and synthetic utility. *Angewandte Chemie (International ed. in English)* 43, 3526–3548. doi: 10.1002/anie.200300631
- Cook, I., Wang, T., Almo, S. C., Kim, J., Falany, C. N., Leyh, T. S. (2013a). Testing the sulfotransferase molecular pore hypothesis. *The Journal of Biological Chemistry* 288, 8619–8626. doi: 10.1074/jbc.M112.445015
- Cook, I., Wang, T., Almo, S. C., Kim, J., Falany, C. N., Leyh, T. S. (2013b). The gate that governs sulfotransferase selectivity. *Biochemistry* 52, 415–424. doi: 10.1021/bi301492j
- Cook, I., Wang, T., Falany, C. N., Leyh, T. S. (2013c). High Accuracy *in Silico* Sulfotransferase Models. *Journal of Biological Chemistry* 288, 34494–34501. doi: 10.1074/jbc.M113.510974
- Cook, I. T., Duniec-Dmuchowski, Z., Kocarek, T. A., Runge-Morris, M., Falany, C. N. (2009). 24-hydroxycholesterol sulfation by human cytosolic sulfotransferases: formation of monosulfates and disulfates, molecular modeling, sulfatase sensitivity, and inhibition of liver x receptor activation. *Drug Metabolism and Disposition: the Biological Fate of Chemicals* 37, 2069–2078. doi: 10.1124/dmd.108.025759
- Cook, I. T., Leyh, T. S., Kadlubar, S. A., Falany, C. N. (2010). Structural rearrangement of SULT2A1: effects on dehydroepiandrosterone and raloxifene sulfation. *Hormone Molecular Biology and Clinical Investigation* 1, 81–87. doi: 10.1515/HMBCI.2010.012

- Estavillo, G. M., Crisp, P. A., Pornsiriwong, W., Wirtz, M., Collinge, D., Carrie, C., et al. (2011). Evidence for a SAL1-PAP chloroplast retrograde pathway that functions in drought and high light signaling in *Arabidopsis*. *The Plant Cell* 23, 3992–4012. doi: 10.1105/tpc.111.091033
- Geu-Flores, F., Nielsen, M. T., Nafisi, M., Møldrup, M. E., Olsen, C. E., Motawia, M. S., et al. (2009). Glucosinolate engineering identifies a gamma-glutamyl peptidase. *Nature Chemical Biology* 5, 575–577. doi: 10.1038/nchembio.185
- Gigolashvili, T., Berger, B., Mock, H.-P., Müller, C., Weisshaar, B., Flügge, U.-I. (2007a). The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in *Arabidopsis thaliana*. *The Plant Journal for Cell and Molecular Biology* 50, 886–901. doi: 10.1111/j.1365-313X.2007.03099.x
- Gigolashvili, T., Engqvist, M., Yatusевич, R., Müller, C., Flügge, U.-I. (2008). HAG2/MYB76 and HAG3/MYB29 exert a specific and coordinated control on the regulation of aliphatic glucosinolate biosynthesis in *Arabidopsis thaliana*. *The New Phytologist* 177, 627–642. doi: 10.1111/j.1469-8137.2007.02295.x
- Gigolashvili, T., Geier, M., Ashykhmina, N., Frerigmann, H., Wulfert, S., Krueger, S., et al. (2012). The *Arabidopsis* thylakoid ADP/ATP carrier TAAC has an additional role in supplying plastidic phosphoadenosine 5'-phosphosulfate to the cytosol. *The Plant Cell* 24, 4187–4204. doi: 10.1105/tpc.112.101964
- Gigolashvili, T., Yatusевич, R., Berger, B., Müller, C., Flügge, U.-I. (2007b). The R2R3-MYB transcription factor HAG1/MYB28 is a regulator of methionine-derived glucosinolate biosynthesis in *Arabidopsis thaliana*. *The Plant Journal for Cell and Molecular Biology* 51, 247–261. doi: 10.1111/j.1365-313X.2007.03133.x
- Gy, I., Gascioli, V., Laressergues, D., Morel, J. B., Gombert, J., Proux, F., et al. (2007). *Arabidopsis* FIERY1, XRN2, and XRN3 are endogenous RNA silencing suppressors. *The Plant Cell* 19, 3451–3461. doi: 10.1105/tpc.107.055319
- Hirai, M. Y., Sugiyama, K., Sawada, Y., Tohge, T., Obayashi, T., Suzuki, A., et al. (2007). Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* 104, 6478–6483. doi: 10.1073/pnas.0611629104
- Hirsch, J., Misson, J., Crisp, P. A., David, P., Bayle, V., Estavillo, G. M., et al. (2011). A novel fry1 allele reveals the existence of a mutant phenotype unrelated to 5'-3' exoribonuclease (XRN) activities in *Arabidopsis thaliana* roots. *PloS One* 6, e16724. doi: 10.1371/journal.pone.0016724

- Jensen, L. M., Halkier, B. A., Burow, M. (2014). How to discover a metabolic pathway? An update on gene identification in aliphatic glucosinolate biosynthesis, regulation and transport. *Biological Chemistry* 395, 529–543. doi: 10.1515/hsz-2013-0286
- Klein, M., and Papenbrock, J. (2004). The multi-protein family of *Arabidopsis* sulphotransferases and their relatives in other plant species. *Journal of Experimental Botany* 55, 1809–1820. doi: 10.1093/jxb/erh183
- Klein, M., and Papenbrock, J. (2009). Kinetics and substrate specificities of desulfo-glucosinolate sulfotransferases in *Arabidopsis thaliana*. *Physiologia Plantarum* 135, 140–149. doi: 10.1111/j.1399-3054.2008.01182.x
- Lee, B.-R., Huseby, S., Koprivova, A., Chételat, A., Wirtz, M., Mugford, S. T., et al. (2012). Effects of fou8/fry1 mutation on sulfur metabolism: is decreased internal sulfate the trigger of sulfate starvation response? *PloS One* 7, e39425. doi: 10.1371/journal.pone.0039425
- Malitsky, S., Blum, E., Less, H., Venger, I., Elbaz, M., Morin, S., et al. (2008). The transcript and metabolite networks affected by the two clades of *Arabidopsis* glucosinolate biosynthesis regulators. *Plant Physiology* 148, 2021–2049. doi: 10.1104/pp.108.124784
- Møldrup, M. E., Geu-Flores, F., Olsen, C. E., Halkier, B. A. (2011). Modulation of sulfur metabolism enables efficient glucosinolate engineering. *BMC Biotechnology* 11, 12. doi: 10.1186/1472-6750-11-12
- Mugford, S. G., Yoshimoto, N., Reichelt, M., Wirtz, M., Hill, L., Mugford, S. T., et al. (2009). Disruption of adenosine-5'-phosphosulfate kinase in *Arabidopsis* reduces levels of sulfated secondary metabolites. *The Plant Cell* 21, 910–927. doi: 10.1105/tpc.109.065581
- Rakers, C., Schumacher, F., Meinl, W., Glatt, H., Kleuser, B., Wolber, G. (2016). *In Silico* Prediction of Human Sulfotransferase 1E1 Activity Guided by Pharmacophores from Molecular Dynamics Simulations. *Journal of Biological Chemistry* 291, 58–71. doi: 10.1074/jbc.M115.685610
- Russell M. Whittemore, L. Bruce Pearce, Jerome A. Roth (1986). Purification and kinetic characterization of a phenol-sulfating form of phenol sulfotransferase from human brain. *Archives of Biochemistry and Biophysics* 249, 464–471. doi: 10.1016/0003-9861(86)90023-8
- Sønderby, I. E., Geu-Flores, F., Halkier, B. A. (2010). Biosynthesis of glucosinolates – gene discovery and beyond. *Trends in Plant Science* 15, 283–290. doi: 10.1016/j.tplants.2010.02.005
- Sønderby, I. E., Hansen, B. G., Bjarnholt, N., Ticconi, C., Halkier, B. A., Kliebenstein, D. J. (2007). A systems biology approach identifies a R2R3 MYB gene subfamily with distinct

- and overlapping functions in regulation of aliphatic glucosinolates. *PloS One* 2, e1322. doi: 10.1371/journal.pone.0001322
- Tibbs, Z. E., Rohn-Glowacki, K. J., Crittenden, F., Guidry, A. L., Falany, C. N. (2015). Structural plasticity in the human cytosolic sulfotransferase dimer and its role in substrate selectivity and catalysis. *Drug Metabolism and Pharmacokinetics* 30, 3–20. doi: 10.1016/j.dmpk.2014.10.004
- Traka, M. H., Saha, S., Huseby, S., Kopriva, S., Walley, P. G., Barker, G. C., et al. (2013). Genetic regulation of glucoraphanin accumulation in Beneforté broccoli. *The New Phytologist* 198, 1085–1095. doi: 10.1111/nph.12232
- Wang, T., Cook, I., Falany, C. N., Leyh, T. S. (2014). Paradigms of sulfotransferase catalysis: the mechanism of SULT2A1. *The Journal of Biological Chemistry* 289, 26474–26480. doi: 10.1074/jbc.M114.573501
- Whittemore, R. M., Pearce, L. B., Roth, J. A. (1985). Purification and kinetic characterization of a dopamine-sulfating form of phenol sulfotransferase from human brain. *Biochemistry* 24, 2477–2482. doi: 10.1021/bi00331a013
- Wilson, P. B., Estavillo, G. M., Field, K. J., Pornsiriwong, W., Carroll, A. J., Howell, K. A., et al. (2009). The nucleotidase/phosphatase SAL1 is a negative regulator of drought tolerance in *Arabidopsis*. *The Plant Journal for Cell and Molecular Biology* 58, 299–317. doi: 10.1111/j.1365-313X.2008.03780.x
- Yatusevich, R., Mugford, S. G., Matthewman, C., Gigolashvili, T., Frerigmann, H., Delaney, S., et al. (2010). Genes of primary sulfate assimilation are part of the glucosinolate biosynthetic network in *Arabidopsis thaliana*. *The Plant Journal for Cell and Molecular Biology* 62, 1–11. doi: 10.1111/j.1365-313X.2009.04118.x
- Zhang, H., Varmalova, O., Vargas, F. M., Falany, C. N., Leyh, T. S. (1998). Sulfuryl Transfer: The Catalytic Mechanism of Human Estrogen Sulfotransferase. *Journal of Biological Chemistry* 273, 10888–10892. doi: 10.1074/jbc.273.18.10888

Abbreviations

2PE	2-phenylethyl glucosinolate
3MTP	3-methylthiopropyl glucosinolate
4MTB	4-methylthiobutyl glucosinolate
5MTP	methylthiopentyl glucosinolate
8MTO	8-methylthiooctyl glucosinolate
ABA	abscisic acid
AMP	adenosine monophosphate
APS	adenosine phosphosulfate
APK	adenosine phosphosulfate kinase
APR	adenosine phosphosulfate reductase
Ala	alanine
Arg	arginine
ATP	adenosine triphosphate
ATPS	adenosine triphosphate sulfurylase
BCAT	branched-chain amino acid aminotransferases
Cys	cysteine
ds	desulfo
ds-Gl	desulfo-glucosinolate
DFCI	Dana-Farber Cancer Institute Gene Index
γ -EC	γ -glutamyl cysteine
EST	expressed sequence tags
GCL	glutamate cysteine lyase
Gl	glucosinolate
Glu	glutamic acid
GPX	glutathione peroxidase
GR	glutathione reductase
GSH	reduced glutathione
GSHS	reduced glutathione synthetase
GSSG	glutathione disulphide
His	histidine

HMM	Hidden Markov Model
I3M	indol-3-yl-methyl glucosinolate
IPTG	isopropyl- β -D-thiogalactopyranosid
LB	lysogeny broth
Lys	lysine
NMA	Normal-Mode-Analysis
OASTL	O-acetylserine thiol lyase
PAP	5'-phosphoadenosine phosphate
PAPS	3'-phosphoadenosine 5'-phosphosulfate
PCB	polychlorinated biphenyls
Poly-Glu	poly-glutamatic acid
PP _i	pyrophosphate
PSK	phytosulfokine
PSY1	plant peptide containing sulphated tyrosine 1
RFG	root meristem growth factor
RMSD	root mean square deviation
ROS	reactive oxygen species
Ser	serine
SiR	sulfite reductase
SOT	sulfotransferase
Thr	threonine
Tyr	tyrosine
TPST	tyrosolprotein sulfotransferase

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die Dissertation selbständig verfasst und die benutzten Hilfsmittel und Quellen sowie gegebenenfalls die zu Hilfsleistungen herangezogenen Institutionen vollständig angegeben habe. Ich erkläre auch, dass die Dissertation nicht schon als Masterarbeit, Diplomarbeit oder ähnliche Prüfungsarbeit verwendet worden ist.

Hannover,

List of Publications

Publications in Journals

Hirschmann, F. and Papenbrock, J. (2015). The fusion of genomes leads to more options: A comparative investigation on the desulfo-glucosinolate sulfotransferases of *Brassica napus* and homologous proteins of *Arabidopsis thaliana*. *Plant Physiology and Biochemistry* 91, 10-19. doi: 10.1016/j.plaphy.2015.03.009

Hirschmann, F., Krause, F., Papenbrock, J. (2014). The multi-protein family of sulfotransferases in plants: Composition, occurrence, substrate specificity, and functions. *Frontiers in Plant Science* 5, 556. doi: 10.3389/fpls.2014.00556

Manuscripts for Publications in Journals

Hirschmann, F., Krause, F., Baruch, P., Chizhov, I., Mueller, J. W., Manstein, D. J., Papenbrock, J., Fedorov, R. The structural and biochemical study of sulfotransferase 18 from *Arabidopsis thaliana* explains its substrate specificity and reaction mechanism. *Scientific Reports* (Manuscript for Publication)

Publications in Books

Abuelsoud, W. A., **Hirschmann, F.**, Papenbrock, J. (2016). “Sulfur metabolism and drought stress tolerance in plants“ in *Drought Stress Tolerance in Plants, Vol 1: Physiology and Biochemistry*, ed. M .A. Hossain, S. H. Wani, S. Bhattachajee, D. J. Burritt, and L. S. Phan Trans (New York, NY: Springer US)

Talks on Conferences

“The multi-protein family of sulfotransferases in plants: Composition, occurrence, substrate specificity, and functions” at the *International Symposium Sulfation Pathways*, 28. – 30. 09. 2015 in Greifswald.

Overseas Research Visits

Four weeks, JSPS funded research visit at the Dept. of Biochemistry and Applied Biosciences, Univ. of Miyazaki, Japan, as part of the research project: “Characterization of plant sulfotransferases for a deeper understanding of plant sulfate metabolism.”

Acknowledgments

I would like to thank **Prof. Dr. Jutta Papenbrock** for trusting me with this very interesting and challenging topic, and for always friendly guiding and helping me in in the past years.

I would like to thank **Dr. Roman Fedorov** for taking over the function as second examiner and teaching me many aspects of protein structural biology and biochemistry.

I would like to thank PD **Dr. Carsten Zeilinger** for being the chair of the examination board.

Florian Krause for his contributions to the review and SOT18 structure determination.

Petra Baruch for her help with protein crystallization assays. **Dr. Jonathan Wolf Müller** for his professional and personal advice and **Dr. Igor Chizov** for his help with the evaluation of kinetic data.

Furthermore, I would like to thank **Pamela von Trzebiatowski** and **Julia Volker** for their general help in the laboratory and especially for their help with mutations, protein expression and purification, as well as HPLC handling.

I would also like to thank all present and past members of the **AG Papenbrock** for creating a pleasant work environment.

Finally, I would like to thank **Raudhah Fatin**, my **family** and **friends** for their patience and support.

Curriculum Vitae

Felix Hirschmann

*30th of. December 1986; Oldenburg (Oldb), Germany

Education

- 07.2013 – 06.2016 Graduate studies at the Institute of Botany, Gottfried Wilhelm Leibniz University Hannover, Germany; Topic of the thesis: “Characterization of desulfo-glucosinolate sulfotransferases by sequence- and 3D-structure analysis”
- 10.2010 – 11.2012 Master of Science in Plant Biotechnology, Gottfried Wilhelm Leibniz University Hannover, Germany; Topic of the thesis: “Characterization of desulfo-glucosinolate sulfotransferases in *Brassica napus*“
- 10.2007 – 10.2010 Bachelor of Science in Plant Biotechnology, Gottfried Wilhelm Leibniz University Hannover, Germany; Topic of the thesis: “Untersuchung der Hydrogenase-Aktivität in *Chlamydomonas reinhardtii*“
- 07.2004 – 06.2006 Acquirement of the „Allgemeinen Hochschulreife“ at the Herbartgymnasium Oldenburg, Germany