The multi-protein family of *Arabidopsis* sulphotransferases and their relatives in other plant species

Marion Klein and Jutta Papenbrock*

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

Received 9 March 2004; Accepted 21 April 2004

Abstract

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

Key words: *Brassica napus*, *Flaveria* species, functional analysis, intracellular localization, substrate.

Introduction

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

\[
PAPS + R-OH \xrightarrow{\text{SOT}} PAP + R-\text{OSO}_3 \]

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

* To whom correspondence should be addressed. Fax: +49 511 762 3992. Jutta.Papenbrock@botanik.uni-hannover.de
In Arabidopsis thaliana Heyn. 18 SOT genes have been identified through alignment search tools (described in this article), that may encode SOT proteins based on sequence identities, but little is known about their functions. The first SOT gene identified in A. thaliana (RA047, A2g03760) was only expressed in shoots and in A. thaliana cell cultures, and its expression was enhanced by challenging plants with isolates of Xanthomonas campestris pathovar campestris 147 and of Pseudomonas syringae pathovar maculicola M2 undergoing an incompatible interaction (Lacomme and Roby, 1996).

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

What are sulphotransferases not?

Sulphotransferases are different from sulphurtransferases. The names of both groups of enzymes, sulphotransferases and sulphurtransferases, are very similar; both groups of proteins were often mixed up in the past. To avoid confusion about the catalytic activity of members of each protein family, the difference will be defined.

Sulphurtransferases/thiadenases are a group of enzymes widely distributed in all three phyla that catalyse the transfer of reduced sulphur from a donor, such as thiosulphate or 3-mercaptopyruvate, to a thiophilic acceptor substrate (Westley, 1973). In the reaction catalysed by sulphurtransferase proteins the transferred sulphur is highly reduced (oxidation state –II), whereas the SOT proteins catalyse the transfer of sulphur that is highly oxidized (oxidation state VI). The sequences of both protein families are not related. In A. thaliana the family of sulphurtransferases also consists of 18 members (Bauer and Papenbrock, 2002).

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

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

Biosynthesis of APS and PAPS

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

$$\text{ATP} + \text{SO}_4^{2-} \xrightarrow{\text{ATP sulphurylase}} \text{APS} + \text{PP}_i$$

The energetics of the sulphate adenylylation reaction favours ATP formation. Therefore, the reaction products, APS and pyrophosphate (PPi), must be maintained at a low concentration by the enzymes, inorganic pyrophosphatase that hydrolyses PPi, and APS reductase (EC 1.8.4.9), and APS kinase (EC 2.7.1.25) that metabolize APS. APS reductase catalyses the first step of sulphate reduction. APS kinase catalyses the ATP-dependent phosphorylation on the 3'-position of APS. The product PAPS is the substrate for the SOT proteins.

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

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

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

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

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

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

### Distribution of sulphotransferases in mammals and plants

The well-studied mammalian SOTs contribute important information to understanding about plant SOTs. Due to the structural similarity among SOTs in general (protein family characteristics are defined in KOG1584, http://genome.jgi-psf.org/chlre2/kog/168755.html, and PF00685, http://pfam.wustl.edu/hmmsearch.shtml), knowledge about mammalian SOTs may be useful. Mammalian SOTs catalyse the sulphate conjugation of many hormones, neurotransmitters, drugs, and xenobiotic compounds. These reactions lead to increased renal excretion of sulphonated products. This results in a decrease in biological activity; but sulphonation is required to activate molecules as well (Weinshilboum and Otterness, 1994; Weinshilboum et al., 1997). According to their cellular localization, SOTs sulphonate different kinds of molecules. The so-called cytosolic soluble SOTs sulphonate small molecules, while Golgi membrane-associated SOTs sulphonate large molecules such as proteins and glycosaminoglycans (Habuchi, 2000; Honke and Taniguchi, 2002). Genes for human enzymes, as well as those of other mammalian soluble enzymes, show a high degree of structural homology (Weinshilboum and Otterness, 1994; Weinshilboum et al., 1997). To date, nine soluble SOTs have been identified from humans (Sugahara et al., 2003). The sulphonation reaction and the formation of its substrates seem to be very important for life: a naturally-occurring defect in the synthesis of the co-substrate PAPS is lethal in humans (Superfi-Furga, 1994). A large number of membrane SOTs produce numerous biological substrates essential for regulatory processes in life such as the action of steroid hormones (Negishi et al., 2001).

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

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

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

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

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

Evidence exists that plants, like mammals, use steroids to regulate their growth and development. In B. napus a SOT protein was characterized that catalysed the sulphonation of brassinosteroids and mammalian estrogenic steroids. The sulphonation abolsishes specifically the biological activity of 24-epibrassinolide. Treatment with salicylic acid, a signal molecule in plant defence, leads to increased expression of the B. napus SOT gene. This suggests an involvement of at least one SOT protein in plant responses to pathogen infection (Rouleau et al., 1999).

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

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

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

In recent years the scientific community was provided with valuable information about the model plant A. thaliana.
Extensive use of all sources available will help to analyse and differentiate between the members of protein families. Therefore, the aim of this study was to identify all genes and gene products which might be classified as SOT in *A. thaliana*. The fully sequenced genome of *A. thaliana* was searched for SOT sequences applying the BLAST program with the already isolated SOT RaR047 protein sequence from *A. thaliana* (Lacomme and Roby, 1996). 18 SOT protein sequences showing high similarity to already known SOT protein sequences and to each other were identified. The phylogenetic tree of the family of *A. thaliana* SOTs shows the relationships among these 18 sequences (Fig. 1). The protein sequences were divided into seven groups according to their sequence similarities; the results are displayed in Table 1. The table shows an overview of the complete SOT family with additional information including gene identification, numbers of amino acids, number of ESTs identified, and intracellular localization predictions (as explained later). Apart from two proteins, all members of this family consist of an approximately equal number of amino acids of at least 310 residues. Computer analysis of *A. thaliana* amino acid sequences indicates that all 18 SOTs might be soluble and none of the SOT proteins contains a transmembrane region as indicated by hydropathy plots in SOSUI (http://sosui.proteome.bio.tuat.ac.jp).

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

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

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

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

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

AtSOT14 in group VI has been characterized biochemically. The recombinant AtSOT14 protein sulphonated 11- and 12-hydroxyjasmonate, whereas for the closely related AtSOT15 protein (87% identity with AtSOT14) no activity was observed with these substrates (Gidda et al., 2003). Both SOTs in group VI show a smaller correspondence of only 42–43% compared with the flavonol SOT of F. chloraeofolia (F4-ST) and B. napus SOTs (steroidST1-3, 41–43%). Therefore, other substrates related in structure to jasmonates have to be tested to determine the biochemical function of AtSOT15.

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

Group II shows a slightly better correspondence with the B. napus SOTs (steroidST1-3, 43–49%) than with the F. chloraeofolia SOT (F4-ST, 39–43%). Despite these results, At5g45070 has been shown to accept a number of flavonols and flavone aglycones as well as their sulphonated derivates (Marsolais et al., 2000). The remaining groups I, III, and IV show an average similarity of about 44% with B. napus SOTs (steroidST1-3). This value is only slightly higher than the respective F. chloraeofolia SOT (F4-ST) value (40–41%). In summary, the sequence/function analysis for the SOT protein family did not reveal clear results. The relatively weak similarities and the small differences among the groups make a biochemical analysis to identify the natural substrate for each SOT protein essential.

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

<table>
<thead>
<tr>
<th>Group</th>
<th>Name</th>
<th>Gene ID</th>
<th>No. aa</th>
<th>EST</th>
<th>Localization prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>AtSOT1</td>
<td>At5g43690</td>
<td>331</td>
<td>2</td>
<td>Microbody (peroxisome) (0.748)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>At3g51210</td>
<td>67</td>
<td>-</td>
<td>ER (membrane) (0.550)</td>
</tr>
<tr>
<td>II</td>
<td>AtSOT3</td>
<td>At4g26280</td>
<td>314</td>
<td>-</td>
<td>Cytoplasm (0.450)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>At2g27570</td>
<td>273</td>
<td>-</td>
<td>Nucleus (0.980)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>At3g45070</td>
<td>323</td>
<td>4</td>
<td>Microbody (peroxisome) (0.429)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>At3g45080</td>
<td>329</td>
<td>-</td>
<td>Microbody (peroxisome) (0.575)</td>
</tr>
<tr>
<td>III</td>
<td>AtSOT7</td>
<td>At1g28170</td>
<td>326</td>
<td>-</td>
<td>Microbody (peroxisome) (0.572)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>At1g13420</td>
<td>331</td>
<td>2</td>
<td>Cytoplasm (0.650)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>At1g13430</td>
<td>351</td>
<td>2</td>
<td>Mitochondrial matrix (0.100)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>At2g14920</td>
<td>333</td>
<td>-</td>
<td>Cytoplasm (0.450)</td>
</tr>
<tr>
<td>IV</td>
<td>AtSOT11</td>
<td>At2g03750</td>
<td>351</td>
<td>9</td>
<td>Microbody (peroxisome) (0.392)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>At2g03760 (RaR047)</td>
<td>326</td>
<td>5</td>
<td>Microbody (peroxisome) (0.291)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>At2g03770</td>
<td>324</td>
<td>-</td>
<td>Microbody (peroxisome) (0.622)</td>
</tr>
<tr>
<td>VI</td>
<td>AtSOT14</td>
<td>At5g07000</td>
<td>347</td>
<td>6</td>
<td>Microbody (peroxisome) (0.602)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>At5g07010</td>
<td>359</td>
<td>15</td>
<td>Microbody (peroxisome) (0.640)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>At5g07010</td>
<td>359</td>
<td>15</td>
<td>Microbody (peroxisome) (0.640)</td>
</tr>
<tr>
<td>VII</td>
<td>AtSOT16</td>
<td>At1g74100</td>
<td>338</td>
<td>25</td>
<td>Microbody (peroxisome) (0.640)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>At1g18590</td>
<td>346</td>
<td>8</td>
<td>Microbody (peroxisome) (0.640)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>At1g74090</td>
<td>350</td>
<td>6</td>
<td>Microbody (peroxisome) (0.566)</td>
</tr>
</tbody>
</table>

1814 Klein and Papenbrock
The rules for subdivision into families and subfamilies according to the percentage of their sequence identity (Weinshilboum et al., 1997) are not very useful for the plant SOT families. The detailed comparison above shows that high sequence similarity alone does not necessarily indicate specificity for the same chemical group of substrates. Even from high sequence identity of more than 85% among two SOT proteins, one cannot conclude the same substrate specificity. Probably SOT proteins specific for a group of substrates evolved independently on more than one occasion. Thus, for each SOT protein the in vitro and finally the in vivo substrate specificity has to be detected.

The crystal structures of a number of mammalian SOT proteins are available which might be used for three-dimensional modelling of the active site and putative substrates from ligand libraries.

### Alignment of the highly conserved regions

Cytosolic soluble SOTs from mammalian species and plant SOT proteins have high structural similarities. All SOTs have conserved amino acid motives which are involved in PAPS binding (regions I and IV) (Marsolais and Varin, 1995; Weinshilboum et al., 1997). Figure 3 shows a partial amino acid alignment of the putative PAPS-binding regions of the A. thaliana SOT protein family. Region I is localized near the N-terminus and region IV at the C-terminus. This alignment indicates that the typical binding site for PAPS exists in all 18 SOTs. However, a comparison of the consensus sequences for SOT proteins from a broad spectrum of species described previously (region I, TYPKSGTxW; region IV, RKGxxGDWKxxFT) (Weinshilboum et al., 1997) are different from the consensus sequences in the PAPS-binding regions of A. thaliana SOT proteins (region I, PKxGTxT; region IV, FRKGxVGDWxxLT). In at least 14 A. thaliana SOTs, the amino acids of these consensus sequences are identical; there are not more than three different amino acids at one position among all 18 SOT proteins. The amino acids in the A. thaliana consensus sequence printed in bold are identical with the overall consensus sequence. In the first published A. thaliana RaR047 sequence (Lacomme and Roby, 1996), there are a number of sequence deviations compared with the sequence published later (The Arabidopsis genome initiative, 2000), including in the consensus sequence of region IV.

### Intracellular localization of A. thaliana SOT

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

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

One method to demonstrate the intracellular localization of nuclear-encoded proteins is the transient expression of fusion constructs with the green fluorescent protein (GFP) (Bauer et al., 2004; Nowak et al., 2004).

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

### Table 1.

<table>
<thead>
<tr>
<th>Protein</th>
<th>N-term Protein</th>
<th>Mitochondrial Pre-sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtSOT1</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>AtSOT5</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>AtSOT6</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>AtSOT11</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>AtSOT12</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

1816 Klein and Papenbrock
One example of this approach is shown in Fig. 4. According to the intracellular localization prediction, AtSOT18 is either localized in the peroxisomes (0.640) or in the cytoplasm (0.450). The full-length cDNA sequence encoding AtSOT18 was cloned in a frame downstream (pGFP-C) or upstream (pGFP-N) of the GFP reading frame. A. thaliana protoplasts were transiently transformed with the GFP constructs according to standard procedures. Bright-field images are shown to visualize the protoplast’s cell borders and the chloroplasts. Fluorescent images of the same protoplasts were done using a fluorescence microscope. As a control, the protoplasts were transformed with the pGFP-C and pGFP-N vector without additional insertion. According to the results, the AtSOT18 protein is cytoplasmic. These studies will be extended to obtain a complete set of data of the intracellular localization of all A. thaliana SOT proteins.

Expression studies

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

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

How to identify the respective substrates and their function?

It is a challenging task to analyse the respective substrates for each SOT. The study of the structure/function relationship of SOT proteins in order to elucidate the mechanism of sulphurate transfer, and to define the amino acids responsible for substrate binding and catalysis will help to identify putative substrates. Using site-directed mutagenesis of the flavonol 3-SOT, several amino acids required for catalysis and co-substrate binding were mapped, while the construction of
chimeric proteins allowed definition of the domain responsible for substrate specificity (Varin et al., 1995). A domain was found which is involved in substrate binding, designated as domain II. Within this domain, two subdomains of high divergence were identified, probably participating in the recognition and binding of different acceptor substrates (Marsolais and Varin, 1997, 1998; Varin et al., 1995). These domains of high divergence could be used for modelling ligands which might serve as substrates using docking programs.

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

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

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

**Experimental evidence for an involvement of a SOT in phytosulphokine biosynthesis**

Phytosulphokine-α (PSK-α) is a sulphonated pentapeptide (Tyr(SO₃H)-Ile-Tyr(SO₃H)-Thr-Gln) and was shown to act as a plant growth factor (Matsubayashi and Sakagami, 1996). The biosynthetic pathway of a preprophytosulphokine was partially elucidated in Oryza sativa L. and Arabidopsis thaliana (Yang et al., 1999, 2001). The PSK-α is only active in its sulphonated form, thus the identification of a tyrosyl peptide SOT (EC 2.8.2.20) is of high importance for the understanding of PSK-α and related peptides in the plant organism. Protein tyrosyl O-sulphation is one of the post-translational modifications that occurs with many secretory and membrane proteins in animal cells. In mammals, a tyrosyl protein SOT was localized specifically in the trans-Golgi-network (Lee and Huttner, 1983). In cell cultures of a number of plant species the existence of a tyrosyl protein SOT activity was demonstrated also in the Golgi-network (Hanaï et al., 2000). However, the sequence of the respective protein has not been determined so far (Youji Sakagami, Nagoya, Japan, personal communication). The sequence of the mammalian tyrosyl protein SOT does not show any similarity to a DNA or protein sequence from a plant species. Therefore, it might be possible that one of the 18 Arabidopsis thaliana SOT acts as tyrosyl protein SOT.

**Involvement of sulphotransferases in glucosinolate biosynthesis**

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

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

Acknowledgements

We would like to thank P von Trzebiatowski for her excellent technical assistance and M Bauer for performing most of the transient expression experiments with GFP.

References


Matsubayashi Y, Sakagami Y. 1996. Phytosulphokine, sulphated peptides that induce the proliferation of single mesophyll cells of...