Elsevier Editorial System(tm) for Plant Physiology and Biochemistry
Manuscript Draft

Manuscript Number:

Title: Desulfo-glucosinolate sulfotransferases isolated from several Arabidopsis thaliana ecotypes differ in their sequence and enzyme kinetics

Article Type: Research Paper

Keywords: Arabidopsis; ecotype; enzyme kinetics; genotype; glucosinolate; secondary structure.

Corresponding Author: Jutta Papenbrock,

Corresponding Author's Institution: University of Hannover

First Author: Sören Luczak

Order of Authors: Sören Luczak; Fabio Forlani; Jutta Papenbrock

Abstract: The goal was to investigate whether the diverse glucosinolate (Gl) profiles described for different Arabidopsis thaliana (L.) Heynh. ecotypes are at least partially shaped by the kinetic properties of sulfotransferases (SOTs) (EC 2.8.2.-) catalyzing the final step in Gl core structure biosynthesis. Homologues of AtSOT18 proteins were characterized in this study, which was inspired by earlier findings on SOTs from ecotypes Col-0 and C24 differing in two amino acids (aa) and specific enzyme activities. Could there be a correlation of AtSOT18 enzyme activities and differences in Gl profiles between the ecotypes? SOT18 sequences from eight Arabidopsis ecotypes with highly diverse Gl patterns differed in two aa at various positions in the protein sequence. The SOT18 sequence from Col-0 showed the highest similarity to the largest number of other sequences in the alignment. The small differences in the primary sequence lead to important structural changes in secondary and tertiary structure that might be the key of different kinetic activities towards a broad range of substrates. All recombinant AtSOT18 proteins showed low substrate specificity with an indolic Gl, while the specificity for aliphatic substrates varied. There is no correlation in the kinetic behavior with the major ds-Gl contents or with the ratio of C3/C4 ds-Gl in the respective ecotype. Therefore, is it unlikely that ds-Gl AtSOT18 proteins play a major role in shaping the Gl profile in Arabidopsis.

***Highlights (for review)**

Bullet points

- Homologous ds-Gl SOT from different ecotypes differ in their sequence
- Amino acids replacements occur at different positions
- Replacement of single amino acids in ds-Gl SOT changes the secondary structure
- Homologous ds-Gl SOT from different ecotypes differ in their substrate specificity
- No coherency among amino acids sequence and Gl profile could be observed

Desulfo-glucosinolate sulfotransferases isolated from several *Arabidopsis thaliana* ecotypes differ in their sequence and enzyme kinetics

Sören Luczak[1], Fabio Forlani[2], Jutta Papenbrock[1]*

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

[2]Dipartimento di Scienze per gli Alimenti, la Nutrizione e l'Ambiente, Università degli Studi di Milano, Via Celoria, 2, 20133 Milano, Italy

Corresponding author: Jutta Papenbrock
Leibniz University Hannover, Institute of Botany
Herrenhäuserstr. 2, 30419 Hannover
Tel. 49 511 762 3788 Fax 49 511 762 19262
Jutta.Papenbrock@botanik.uni-hannover.de

Abstract

The goal was to investigate whether the diverse glucosinolate (Gl) profiles described for different *Arabidopsis thaliana* (L.) Heynh. ecotypes are at least partially shaped by the kinetic properties of sulfotransferases (SOTs) (EC 2.8.2.-) catalyzing the final step in Gl core structure biosynthesis. Homologues of AtSOT18 proteins were characterized in this study, which was inspired by earlier findings on SOTs from ecotypes Col-0 and C24 differing in two amino acids (aa) and specific enzyme activities. Could there be a correlation of AtSOT18 enzyme activities and differences in Gl profiles between the ecotypes? SOT18 sequences from eight *Arabidopsis* ecotypes with highly diverse Gl patterns differed in two aa at various positions in the protein sequence. The SOT18 sequence from Col-0 showed the highest similarity to the largest number of other sequences in the alignment. The small differences in the primary sequence lead to important structural changes in secondary and tertiary structure that might be the key of different kinetic activities towards a broad range of substrates. All recombinant AtSOT18 proteins showed low substrate specificity with an indolic Gl, while the specificity for aliphatic substrates varied. There is no correlation in the kinetic behavior with the major ds-Gl contents or with the ratio of C_3/C_4 ds-Gl in the respective ecotype. Therefore, is it unlikely that ds-Gl AtSOT18 proteins play a major role in shaping the Gl profile in *Arabidopsis.*

Keywords: *Arabidopsis*, ecotype, enzyme kinetics, genotype, glucosinolate, secondary structure.

Abbreviations: aa, amino acid(s); ds, desulfo; DTT, dithiothreitol; Gl, glucosinolate(s); I3M, indole-3-ylmethyl Gl; ITC, isothiocyanate; 4MTB, 4-methylthiobutyl Gl; 7MTH, 7-methylthioheptyl Gl; 8MTO, 8-methylthiooctyl Gl; 3MTP 3-methylthiopropyl Gl; 5MTP, 5-methylthiopentyl Gl; PAPS, 3'-phosphoadenosine 5'-phosphosulfate; SOTs, sulfotransferases.

Introduction

Sulfotransferases (SOTs) (EC 2.8.2.-) catalyze the transfer of a sulfate group from the co-substrate 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to a hydroxyl group of suitable sulfate acceptors among them desulfo-Gl (ds-Gl) with parallel formation of 3'-phosphoadenosine 5'-phosphate. All SOT sequences are characterized by a conserved SOT domain (PF00685). Members of the multi-protein family can be found in almost all organisms except in Archaea. In *Arabidopsis thaliana* (L.) Heynh. ecotype Col-0 18 sequences containing a SOT domain have been identified and classified into subgroups [1]. Later three more SOT sequences containing a less conserved SOT domain were added to the SOT family [2]. For many of the SOT proteins in plants the respective substrate it not known yet (for an overview see [2]). One subgroup consisting of three cytosolic SOT proteins from *Arabidopsis* (AtSOT) catalyzes the transfer of a sulfate group to ds-Gl (AtSOT16 to 18). Their expression pattern in various conditions in the plant is almost identical. Therefore we assume that they take over different functions within the cell based on their different kinetic affinities to various ds-Gl substrates in vitro [3-6].

Gl are a group of over 130 nitrogen- and sulfur-containing natural products found in vegetative and reproductive tissues of 16 plant families. Gl share a core structure containing a β-D-glucopyranose residue linked via a sulfur atom to a (*Z*)-*N*-hydroxyimino sulfate ester, and are distinguished from each other by a variable R group derived from one of several amino acids (aa), mainly tryptophan, phenylalanine, and methionine. The Gl pattern of *Arabidopsis* varies within the plant as well as among *Arabidopsis* ecotypes. Significant differences could be detected in both quality and quantity of Gl [7, 8]. In 39 *Arabidopsis* ecotypes 34 different Gl have been identified so far [9]. Gl are involved in defense against herbivores and maybe also pathogens [10, 11]. Some of the breakdown products of several Gl are supposed to act anti-cancerogen, whereas others cause goiter formation. Therefore the interest in analyzing the Gl pattern in plants, the regulation of their biosynthesis and the manipulation is also high from an applied point of view [12].

Gl biosynthesis can be divided into three stages: 1) precursor aa, such as methionine, is elongated by one or several methylene groups, 2) precursor aa is converted into parent Gl (here defined as Gl without secondary modifications of the side chains), 3) finally, parent Gl can be secondarily modified. The last step in the biosynthesis of parent Gl is catalyzed by ds-Gl SOTs. The Gl occurring in *Arabidopsis* are synthesized from methionine and tryptophan [13, 14]. Primary sulfur assimilation is closely connected to the so called secondary

metabolism of plants because the intermediate adenosyl-5-phospho-sulfate (APS) is part of the cysteine biosynthetic pathway and of the PAPS biosynthetic pathway. In *Arabidopsis* APS kinase mutants the Gl pattern was drastically changed [15, 16].

This study is inspired by the fact that AtSOT18 proteins from two ecotypes C24 and Col-0 differing in two aa cause a five times difference in the apparent V_{max} and changes in the substrate affinity of the respective recombinant proteins [17]. The goal was to investigate whether the diverse Gl profiles described for different *Arabidopsis* ecotypes are at least partially influenced by the kinetic properties of the AtSOT catalyzing the final step in ds-Gl core structure biosynthesis. We wanted to know whether AtSOT18 proteins from more *Arabidopsis* ecotypes also show sequence deviations of two or even more aa in comparison to the Col-0 sequence. If there is variation in the sequences among ecotypes do the aa variations occur at the same position within the sequence? This would indicate an evolutionary pressure towards these aa. Does the sequence variation result in a modified affinity of the respective recombinant proteins towards ds-Gl? If one can observe a different kinetic behavior is there any correlation with the major ds-Gl contents or with the ratio of C_3/C_4 ds-Gl in the respective ecotype?

Results

AtSOT18 sequence analysis of different ecotypes

In the past we could demonstrate that the comparison of specific activities of two AtSOT18 proteins from different ecotypes, namely from ecotype C24 and Col-0, differed significantly although both protein sequences differ only in two aa [5]. Therefore we decided to analyze AtSOT18 sequences from more ecotypes to investigate whether this observation was an exception or normality and could explain differences in the Gl patterns. Based on a large data set analyzing the Gl composition and contents in 39 *Arabidopsis* ecotypes and revealing 34 different Gl [9], ecotypes revealing extreme amplitudes were selected for AtSOT18 sequence analysis (Table 1). Originally, it was planned to analyze 10 different ecotypes, two from our stock at the LUH and eight from the Nottingham Arabidopsis Stock Centre (NASC, Norwich, UK). However, the seeds obtained did not germinate although many conditions were tried. Finally, the NASC conducted germination experiments and confirmed the low germination rates (Table 2). Due to time limitation and due to the fact that several other ecotypes show a similar broad range in the Gl composition additional ecotypes were ordered and germinated to a higher percentage (Table 2).

Finally, the AtSOT18 sequences from eight different ecotypes were analyzed in detail. Two of them, C24 and Col-0, and intermediate mutants ($C24G_{301}D$ and $C24N_{339}K$) have been already partially characterized [4, 6, 17]. When the intronless AtSOT18 sequences from the two ecotypes originally chosen for analysis, C24 and Col-0, are compared with the newly analyzed ecotypes, the Col-0 shows the highest similarity to the largest number of other AtSOT18 sequences in the alignment and constitutes therefore a kind of a consensus sequence (Fig. S1). The AtSOT18 sequences from Pi-0 and Sorbo are identical to Col-0. Also when sequences are randomly selected from a larger data set, for this study genomes from 19 *Arabidopsis thaliana* ecotypes have been sequenced (http://mus.well.ox.ac.uk/19genomes/, 28.11.2011), the Col-0 sequence also constitutes a consensus sequence (data not shown). AtSOT18 sequences from Bla-10, Cvi-0, and Ws differed in one aa and sequences from the ecotypes Ler-1 and C24 differed in two aa from Col-0 (Fig. 1, Fig. S1). However, the aa position and the aa replacements are different in each sequence (Fig. 2). Some of the replacements are within one of the regions I to IV as Bla-10 and Ler-1, others are between the regions as in Ws, Ler-1, C24 and Cvi-0. The physico-chemical properties of the replaced aa in comparison to the Col-0 AtSOT18 sequence show a broad variety (Fig. 2). In the AtSOT18 sequence from ecotype Ws for example the acidic anionic glutamic acid in position 31 is

replaced by the basic, cationic lysine residue and in Bla-10 the aromatic uncharged aa tyrosine 231 is replaced by the basic, cationic histidine residue. Also the physico-chemical differences among the replaced aa in Col-0 and the C24 sequences are large and therefore also in the intermediate mutants, C24G_{301}D and C24N_{339}K. These divergent replacements might have impact on the overall structure and the catalytic activity of the AtSOT18 proteins from different ecotypes.

The *AtSOT16* and *AtSOT17* sequences have been also isolated from the same ecotypes as AtSOT18. Again the Col-0 sequence has the highest similarity to most of the sequences in the alignment. The AtSOT16 and 17 from C24 differ in two aa from Col-0 AtSOT16 and 17, respectively, Ler-1 AtSOT16 in two aa and Bla-10 AtSOT17 in one aa (Fig. S2A and S2B). All other investigated ecotypes (Cvi-0, Pi-0, Sorbo, Ws) and the ten randomly chosen ecotypes from the databases possess the SOT18 sequences identical to SOT18 from Col-0.

Impact of amino acid replacements on the secondary structure

Therefore the secondary structure was analyzed experimentally and by using prediction programs. Circular dichroism (CD) relies on the differential absorption of left and right circularly polarised radiation by chromophores which either possess intrinsic chirality themselves or are placed in chiral environments. Proteins contain a number of chromophores which can give rise to CD signals. In the far UV region (180 to 240 nm) the absorbing band is principally the peptide bond. Far UV CD analysis can be used to quantitatively assess the overall secondary structure content of the protein because the different forms of regular secondary structure found in proteins generate different spectra [18]. The spectra of AtSOT18 proteins from C24, C24G_{301}D, and Col-0 obtained by CD spectroscopy are very similar (Fig. 3A). However, when deconvolution analyses were carried out as shown in the histogram (Fig. 3B) the results reveal that the exchange of two aa among AtSOT18 from Col-0 and C24 has effects on the composition of the secondary structure of both proteins. The percentage of anti-parallel β-sheets is low in C24 in comparison to Col-0. The replacements of one of the two different aa, from glycine to asparagine, in neighbourhood to region IV has an impact on the experimentally determined secondary structure: The low amount of anti-parallel sheets in C24 increased from 10% to about 25% in the mutant C24G_{301}D which is very close to the value in Col-0 (23%) (Fig. 3B). These results are in agreement with the enzyme activity measurements where C24G_{301}D and Col-0 show almost the same apparent V_{max} activities with ds-benzyl Gl

as substrate (440 and 420 pkatal mg^{-1} protein in $C24G_{301}D$ and Col-0, respectively, versus 15 pkatal mg^{-1} protein in C24, data from [6]).

The bioinformatic prediction of the secondary structure was done by PredictProtein (Table 3) and Domain Annotation (data not shown). Both prediction programs revealed similar results. In summary, the overall content of α-helices amounts to about 33%, of β-sheets to about 9% and the coil structure to about 58% in proteins from all ecotypes including the mutants. The algorithm in Domain Annotation calculates the amount of α-helices to about 40%, of β-sheets 12% and of coils to 48% (data not shown). The analysis of changes in the number of secondary structure elements in comparison to Col-0 shows that there are major differences among the AtSOT18 proteins from different ecotypes, varying between 11 (Bla-10) and 20 (Cvi-0) indicating that the exchange of one aa has already a large impact on the secondary pattern in SOT proteins. The prediction of the three-dimensional structure for all ecotypes and mutants analyzed using Swiss-Model with the deposited structure 1Q44 (AtSOT12) as template demonstrates that all replacements of aa in comparison to Col-0 are localized close to the surface of the protein molecule (data not shown).

Determination of enzyme activities and substrate specificities

Each recombinant protein was purified four to six times and the average protein concentration varied between 6.64 ± 1.28 g L^{-1} (Bla-10) and 1.16 ± 0.26 g L^{-1} (Ws) of expressed protein concentration per bacterial culture. After purification independent determination of the protein concentration all proteins were diluted to 0.1 µg $µL^{-1}$ and stored frozen at -70°C. The stability of the protein was tested after several days and weeks and was shown to be stable for several weeks when stored at -70°C (data not shown).

Apart from differences in side chain length, natural variation in Gl profiles among different ecotypes can be mostly attributed to secondary modifications. We assume that SOTs convert only ds-Gl with unmodified side chains as described in [17]. Therefore these modifications are not relevant in this context and only the following parent Gl were tested as substrates: ds-3-methylthiopropyl Gl (ds-3MTP, glucoiberverin), ds-4-methylthiobutyl Gl (ds-4MTB, glucoerucin), ds-5-methylthiopentyl Gl (ds-5MTP, glucoberteroin), ds-7-methylthioheptyl Gl (ds-7MTH), ds-8-methylthiooctyl Gl (ds-8MTO), ds-indole-3-ylmethyl Gl (I3M, glucobrassicin).

The ds-substrates and the sulfated products can be separated by HPLC and detected using a UV detector at 229 nm. Based on the protocols for the enzyme assays already developed [6]

the HPLC conditions were optimized with respect to the run time and resolution. Then the linear range for the determination of apparent V_{max} was determined for each protein with each substrate. To analyze the kinetic parameters the apparent V_{max} and K_m values were determined using the aliphatic C_3 ds-Gl ds-3MTP and C_4 ds-4MTB as substrates (Table 4). The differences in the apparent V_{max} of the ecotypes for ds-3MTP were between 99 pkatal mg^{-1} protein and 6933 pkatal mg^{-1} protein and for ds-4MTB between 131 pkatal mg^{-1} protein and 4783 pkatal mg^{-1} protein. The apparent K_m for ds-3MTP were between 55 µM and 100 µM and between 55 µM and 100 µM for ds-4MTB. Bla-10 showed the highest apparent V_{max} with the $C3_4$ ds-3MTP and Ler-1 the highest apparent V_{max} with the C_4 ds-4MTB. Col-0 had the highest affinity for ds-3MTP and ds-4MTB in comparison to the other ecotypes, C24 the lowest.

Also AtSOT16 from both ecotypes Col-0 and C24 was expressed and purified. Both proteins showed the highest activity with ds-I3M. The activity of AtSOT16 from Col-0 is approximately two times higher compared to AtSOT16 from C24 (2870 pkatal mg^{-1} protein and 1103 pkatal mg^{-1} protein). Measurements with the short aliphatic ds-3MTP revealed lower activities (about 50 pkatal mg^{-1} protein), as did measurements with the long aliphatic ds-8MTO (between 136 and 210 pkatal mg^{-1} protein). Regarding the aliphatic ds-Gls, the activity seems to depend on the chain length. Activity with short ds-Gls is higher than that with long chain ds-Gls. In principle, proteins of both ecotypes C24 and Col-0 showed the same behaviour with the exception of the two-fold higher value for the indolic ds-I3M (data not shown).

Comparison of the apparent V_{max} determined with all parent Gl

To obtain an overview about the activity of the recombinant SOT18 proteins from the different ecotypes with all six parent Gl the V_{max} values were determined. We assume that the recombinant proteins from Pi-0 and Sorbo with identical SOT18 sequences as Col-0 have the same kinetic properties as the protein from Col-0. Therefore the recombinant Pi-0 and Sorbo proteins were not included into the analysis. All recombinant proteins showed some enzyme activity with all substrates offered (Fig. 4). However, the enzyme activity with regard to single substrates differed largely, especially between C24 and $C24G_{301}D$. The highest enzyme activity of all genotypes was determined for mutant $C24G_{301}D$ with aliphatic Gl. The lowest activity was measured with the indolic substrate ds-I3M, with C24, $C24N_{339}K$ and Ws being the lowest. In comparison to the activity of the Col-0 consensus protein the mutant $C24G_{301}D$

showed a two to four times increased apparent V_{max} with respect to C_3 ds-3MTP Gl, 2.5 times increased apparent V_{max} for ds-5MTP Gl and a 1.5 times apparent V_{max} to C_8 ds-8MTO. C24 showed with all substrates the lowest activity with the exception. Genotypes Bla-10, Ws, and Ler-1 showed an increasing activity with increasing chain length of the aliphatic ds-Gl. The mutants C24G_{301}D and C24N_{339}K showed the same specificity pattern but at different activity levels.

Discussion

Polymorphisms in AtSOT sequences from different ecotypes lead to changes on the amino acid level

During our detailed analysis of SOT proteins we found by chance that AtSOT18 proteins from two different ecotypes showed different behavior on the enzymatic level, although they differ only in two out of 350 aa [6]. We hypothesized that ds-Gl SOT proteins isolated from different ecotypes varying in their sequence have different substrate affinities for parent Gl and influence therefore, in addition to other enzyme activities and regulation, the complete Gl profile of the respective ecotype. Therefore, we decided to investigate this phenomenon in more detail for AtSOT18, including enzyme kinetics, whereas for AtSOT16 and 17 only sequences from different ecotypes have been analyzed. The sequence analysis of six further *Arabidopsis* ecotypes selected based on their Gl pattern revealed four additional AtSOT18 genotypes. Also the extension of this approach to randomly chosen *AtSOT* sequences from a larger sequencing project of *Arabidopsis* ecotypes (http://mus.well.ox.ac.uk/19genomes/) underlines that the AtSOT Col-0 sequence seems to represent the consensus sequence of all 750 known *Arabidopsis* accessions. Regions in Africa and central Asia are underrepresented in the *Arabidopsis* collections and might provide more AtSOT genotypes. The three ecotypes with Col-0 sequence (Col-0, Pi-0 and Sorbo) originate from three different continents and from different ecosystems (Table 1). Therefore the evolutionary pressure, if there is any, cannot be instantaneously deduced from the environmental conditions.

Variation in one or two amino acids cause major structural changes in the protein molecules

The varying aa in the ecotypes analyzed in this study are distributed over the whole sequence length (Fig. 2). One could assume that they are part of defined binding domains. The structure of SOT proteins in general has been analyzed and four different regions have been defined for either substrate (ds-Gl) or co-substrate (PAPS) binding [19]. The pattern KSGTTW in the N-terminus might be responsible for interaction with 5`-phosphate [20]. In all AtSOT18 proteins the pattern KTGTTW was identified probably forming a similar binding. Additional putative PAPS binding sites have been suggested to be localized in AtSOT18 in the following positions: Arginine 177, serine 185, cysteine 282, phenylalanine 284 and in the pattern at the C-terminus RKG [20]. Both patterns, KTGTTW and RKG, are localized in or close by the regions I and IV [21]. However, none of these aa putatively involved in PAPS binding was modified in one of the genotypes investigated. Impact on the enzyme activity could be either

assumed directly by the replaced aa being part of the ds-substrate binding sites or the overall changes in protein conformation affect the activity. The secondary structure was analyzed by CD spectroscopy and several prediction methods were applied. Surprisingly, the results of both approaches are quite different. The prediction algorithms indicate only small changes among the composition of secondary structures. Also the comparison of the CD spectra reveals only minor changes. However, deconvolution analysis indicates changes in the secondary structure composition (Fig. 3B). We also included the secondary structure composition of the most closely related protein with a resolved three-dimensional structure which has been also used for homology modeling: the AtSOT12 has even a higher α-helix content of 47% α-helices and 9% β-sheets almost in agreement with the prediction (Table 3). Prediction programs use training sets of solved protein structures; this might explain the high accordance. In conclusion, i) the changes evidenced in the CD analysis should not be taken numerically; ii) changes were confirmed with different aliquots of purified protein; iii) deconvolution by the CDNN program analyzes differently from prediction algorithms considering that the α-helices, β-sheets and coil structures of prediction become α-helices, parallel β-sheets, antiparallel β-sheets, β-turn and coil in CDNN; iiii) maybe the protein suspension used for CD analysis contained some contaminating proteins which might have influenced the measurement disproportionately strong.

Usually the core structure is highly conserved and responsible for the catalytic activity. The PDB model 1Q44 (AtSOT12) used for homology modeling shows a sequence identity to AtSOT18 of 38% which is not enough for reliable modeling. In addition, the three-dimensional model we obtained reveals only changes on the surface of the protein molecule. It may be possible that the residues are responsible for substrate binding. For final conclusions, an x-ray analysis of protein crystals including substrate and (inactive) co-substrate and elucidation of the structure is necessary to analyze binding and catalytic sites.

Different AtSOT genotypes influence enzyme activity and substrate specificity

The results of the enzyme assays of this study including many genotypes were compared with the Gl pattern in different *Arabidopsis* ecotypes [9]. Correlation might indicate an influence of AtSOT proteins on the Gl pattern, maybe dependent from the ecological niche of the respective ecotype and might give indications for the evolution of the ds-Gl AtSOT enzymes. In Table 5 the C_3 and the C_4 Gl contents and their ratios in the leaves of different *Arabidopsis* ecotypes are summarized. Bla-10, Col-0, Cvi-0, Sorbo and C24 contain mainly C_4 Gl, Ler-1, Pi-0, and Ws contain mainly C_3 Gl. There are only minor differences in the C_7/C_8 ratios in all ecotypes selected [9]. The ecotypes Col-0, Pi-0 and Sorbo which contain the same AtSOT18

genotype reveal quite different contents of Gl with different chain length (3MTP, 4MTB, 7MTH, and 8MTO) in the leaves. The substrate specificity indicated that AtSOT18 genotypes Bla-10 and Ler-1 prefer long-chain substrates. On the other hand the Gl spectrum of both ecotypes shows rather high contents of short-chain Gl 3MTP (Ler-1 9.24 µmol g DW^{-1}) and 4MTB (Bla-10 9.91 µmol g DW^{-1}) whereas the long-chain Gl are accepted with high affinity accumulated to maximal 1.53 µmol g DW^{-1} (Bla-10) and 0.1 µmol g DW^{-1} (Ler-1). In our analysis only the parent ds-Gl offered as single substrates were investigated. It was found in a competitive enzyme assay for AtSOT17 that ds-7MTH and ds-8MTO were preferred whereas in single substrate tests mainly ds-8MTO was sulfated [17]. The parallel occurrence of many ds-Gl in the cell, parent and derived ds-Gl, might influence the substrate specificities.

The differences in specific in vitro activities may not be biologically meaningful because they could be compensated for by different expression levels in the different ecotypes. So far mainly the expression of SOTs in Col-0 and C24 ecotypes has been analyzed. Only minor differences in the expression of AtSOT16, 17 and 18 could observed ([6], own analysis in GENEINVESIGATOR, https://www.genevestigator.com/). Therefore, theoretically some of the homologues compared in this study may not be expressed in leaves at all because all were cloned from genomic DNA. Some AtSOT18 homologues show a preference for substrates with certain chain lengths (Fig. 4). Chain length distribution seems to be primarily determined by the methylthioalkylmalate synthases (MAM) [22].

Also AtSOT16 and 17 contribute to the overall catalytic activity of SOT in the cell and might influence the Gl composition and pattern. The spectrum of Gl was determined in *Arabidopsis* seeds and leaves at a certain developmental stage [9]. The expression and protein content [6] and maybe also the enzyme activity are under developmental control. The concentration of ds-Gl in the cytosol was not exactly determined so far. The ratio of short-chain ds-Gl might be switched to long-chain ds-Gl and therefore a higher affinity of SOT proteins for long-chain ds-Gl could be explained. However, at least for genotype Col-0 the ratio of short-chain ds-Gl to long-chain ds-Gl was not changed as demonstrated by [8].

Other factors probably have a strong influence on the Gl pattern, for example the concentration of the co-substrate PAPS. It was shown that in a Col-0 T-DNA knock-out mutant of adenosine 5'-phosphosulfate kinase 2 (APK2), catalyzing the biosynthesis of PAPS from APS, the biosynthesis of 12-hydroxy-jasmonate, a compound also sulfated by SOT proteins, and Gl was reduced up to 12 times [23]. In a biotechnological approach to enable efficient Gl engineering in plants, none of the three AtSOT proteins alleviated the bottleneck

of the last step in Gl biosynthesis but that co-expression of APK2 alone reduced an accumulation of ds-Gl precursor and its derivatives by more than 98% and increased Gl accumulation 16-fold [24].

In APK double knock-out mutants (*apk1 apk2*) the transcript levels for genes involved in Gl synthesis, such as MAM3, AtSOT17, or C-S-lyase, were increased [23]. However, the steady-state levels of these three mRNAs were significantly increased in AKP lines were bacterial APK was directed to the cytosol, but only MAM3 and AtSOT17 transcripts were elevated in plants with APK directed to the plastids [16] indicating that the abundance of the three AtSOT transcripts is differentially regulated. However, it was also shown that transcriptional profiling of three R2R3 MYB transcription factor knock-out mutants revealed that Gl metabolite levels are uncoupled from the level of transcript accumulation for aliphatic Gl biosynthetic genes. This uncoupling of chemotypes from biosynthetic transcripts suggests that a modular system in which transcription factors cause similar chemotypes via non-overlapping regulatory patterns [25].

In summary, AtSOT18 proteins from many *Arabidopsis* ecotypes show sequence deviations of one or two but not more aa in comparison to the Col-0 sequence. The positions of the aa replacements are different in each sequence. The recombinant proteins show modified ds-Gl, however, there is no correlation in the kinetic behavior with the major ds-Gl contents or with the ratio of C_3/C_4 ds-Gl in the respective ecotype. Therefore, it is unlikely that ds-Gl SOT proteins play a major role in shaping the Gl profile in *Arabidopsis.* The different approaches concerning the regulation of concentration and composition of Gl in plants are still not consistent to fully explain or even predict the levels of Gl in different *Arabidopsis* ecotypes and in different environmental conditions.

Material and Methods

Plant material

The *Arabidopsis thaliana* (L.) Heynh. ecotypes were selected based on a very high diversity of Gl according to [9]. The seeds of the respective ecotypes were obtained from NASC except for seeds of the ecotypes C24 and Wassilewkija (Ws). The ecotypes ordered are listed in Table 1. The sterilization of a small amount of *Arabidopsis* seeds (ca. 10 mg) was carried out by shaking in 1 ml ethanol (70%) for 7 min. Ethanol was replaced by 6% sodium hypochlorite, incubated for 10 min, centrifuged and the supernatant was discarded. Afterwards five washing steps with sterile H_2O followed. The seeds were resuspended in sterile H_2O and plated on MS medium [26]. Seeds were germinated in climatic chambers at 22°C and 12 h light. After generation of roots the seedlings were transferred to VM-uniform soil (Arthur Stangenberg GmbH, Sinntal-Jossa, Germany) and sand in a ratio of 2:1, and grown at 22 to 24°C in the greenhouse until sufficient material for DNA isolation was available.

DNA cloning

The gene At1g74090 (GI 177666948) encoding AtSOT18 in *Arabidopsis* ecotype C24 and AtSOT18* (At1g74090, GI 30698988) from ecotype Col-0 were cloned as described in [6] and [17]. The production of the mutants used in this study was described in a previous publication [6]. The nomenclature used is given in [1]. The AtSOT18 sequences from the other *Arabidopsis* ecotypes were amplified in the following way: 0.5 µl Dream*Taq*™ polymerase (5 U μl^{-1}) (MBI Fermentas, St. Leon-Rot, Germany), template (1 µg μl^{-1}): 1 µl, primer forward/reverse (10 µM): 2 µl, dNTP-Mix (10 mM): 5 µl, buffer (10x): 5 µl, H_2O: 34.5 µl, T_A: 53.3°C) or Phusion® Hot Start II (Finnzymes, Espoo, Finnland; Polymerase (2 U μl^{-1}) 0.5 µl, template (1 µg μl^{-1}): 1 µl, primer forward/reverse (10 µM): 2.5 µl, dNTP-Mix (10 mM): 5 µl, buffer (10x): 5 µl, H_2O: 33.5 µl, T_A: 54.7°C) using the following with restriction sites coupled primers: forward (*Bam*HI) -5'-GGATCCGAATCAGAAACCCTA-3', reverse (*Hin*dIII) 5'-AAGCTTTTTACCATGTTCAAGC-3'. The PCR fragments were either cloned into the pGEM-T vector (Promega, Madison, USA) or into the pJET1.2/blunt vector (MBI Fermentas) and sent for sequencing from both directions with T7 and SP6 primers (GATC Biotech Konstanz, Germany). At least two independent PCRs were performed per ecotype to reach 100% identity. If it was not possible to get two identical amplicons the consensus sequence of at least three independent PCRs was assembled to identify the AtSOT18

sequence. For expression of recombinant AtSOT18 proteins the PCR amplicons identical to the consensus sequence was cloned into the pQE-30 (Qiagen, Hilden, Germany). The AtSOT16 and 17 sequences from the *Arabidopsis* ecotypes Bla-10, Cvi-0, Ler-1, and Ws were amplified with Dream*Taq*™ at the conditions described above but differed in primer sequences, annealing temperatures and restriction sites. The primers used for amplification of AtSOT16 (T_A: 51°C) were forward primer (*Kpn*I) 5'-GGTACCGAATCAAAGACAACC-3', reverse primer (*Pst*I) 5'-CTGCAGGTTATCATGTTGAAGC-3' and for AtSOT17 (T_A: 49°C) forward primer (*Bam*HI) 5'-GGATCCGAATCCAAAACCATA-3', reverse primer (*Sal*I): 5'-GTCGACTGATTTTGTAGAAAG-3'. The conditions for further analysis of AtSOT16 and 17 sequences were the same as for AtSOT18.

Expression and purification of AtSOT proteins

All AtSOT18 proteins were expressed in *E. coli* XL1-Blue. A 50 ml overnight culture was used for the inoculation of 200 ml 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 OD_{600} of 0.6 followed by induction of the T5-promotor with IPTG and incubation at 30°C for 3 h. The induction was terminated by centrifugation at 4,000 x *g* and 4°C. Pellets were stored at -20°C before protein extraction. The pellets were dissolved in 6 ml lysis buffer (20 mM sodium phosphate buffer, 0.5 M NaCl, 20 mM imidazole, pH 7.4 using HCl) and lysozyme (1 mg ml^{-1}). Lysis itself was carried out by sonification. Afterwards the solution was centrifuged for 30 min at 32,000 x *g* and the supernatant transferred to fresh tubes. The AtSOT18 proteins from all ecotypes could be successfully expressed as soluble proteins in *E. coli* in large amounts although some of the recombinant proteins remained insoluble in the pellet. The total amount of recombinant AtSOT18 proteins after 3 h induction with IPTG was estimated between 25 and 35% as a percentage of the total protein.

The purification of recombinant AtSOT18 proteins by affinity chromatography was optimized and standardized by using the purification system ÄKTAprime™ plus (GE Healthcare, Freiburg, Germany). The application "affinity purification any HiTrap" was used for washing the nickel column (HiTrap™ FF crude; GE Healthcare) containing affinity-bound proteins with lysis buffer and elution buffer (20 mM sodium phosphate buffer, 0.5 M NaCl, 0.5 M imidazole, pH 7.4 using HCl). Only one fraction of a volume of 1 ml containing highly purified and concentrated AtSOT18 protein was used for desalting in the presence of DTT. The desalting of recombinant proteins was performed by the purification system

ÄKTAprime™ plus equipped with the HiTrap™ Desalting Column (GE Healthcare) using desalting buffer (20 mM Tris pH 8.0, 1 mM DTT) and the application "desalting any HiTrap". The purity of the recombinant proteins was analyzed by SDS-polyacrylamide gel electrophoresis and subsequent Coomassie Brilliant Blue stain. The impurities were always less than 10% and were considered in the calculation of the final protein concentration. Protein contents were determined by the method of [27] with bovine serum albumin (Roth, Karlsruhe, Germany) as a protein standard.

Preparation of substrates

The ds forms of the parent Gl derived from methionine and tryptophan were prepared as described by [28]. The following Gl were used in the experiments in their ds forms: 3-methylthiopropyl Gl (3MTP, C08412) from *Erysimum pumillum*, 4-methylthiobutyl Gl (4MTB, C08409) from *Eruca sativa*, 5-methylthiopentyl Gl (5MTP, C08401) from *Arabidopsis thaliana*, 7-methylthioheptyl Gl (7MTH, C17252) from *Nasturtium officinale*, 8-methylthiooctyl Gl (8MTO, C17254) from *Arabis stelleri*, indol-3-ylmethyl Gl (I3M, C05837) from *Isatis tinctoria*. In brackets the KEGG COMPOUND numbers are given (http://www.genome.jp/dbget/) and the structures of all *Arabidopsis* Gl are also shown in [29].

Enzyme activity measurements and HPLC analysis

Enzyme activity measurements were done in the following way: The enzyme assays with recombinant proteins were set up as described by [6]. For the determination of substrate specificities the amount of the recombinant protein as well as the incubation time were optimized for each protein/substrate combination. Substrate and co-substrate concentrations were set constant at 60 µM.

Generally, for the determination of the specific activities the 150 µl assay contained 80 mM Tris/HCl, pH 9.0, 9.2 mM $MgCl_2$, 60 µM PAPS (obtained from Professor H. R. Glatt, Institute of Human Nutrition, Berholz-Rehbruecke, Germany), the respective substrate (60 µM), and 0.25 to 7.5 µg purified protein. For the determination of the apparent K_m values of ds-3MTP and ds-4MTB, at least eight different substrate concentrations were used (10 µM to 210 µM). PAPS concentration was kept constant at 60 µM, when it was not limiting. The 150 µl assay for the determination of apparent K_m values contained 80 mM Tris/HCl, pH 9.0, 9.2 mM $MgCl_2$, ds-3MTP or ds-4MTB (10 µM to 210 µM), 2 to 7.5 µg purified protein and 60 µM PAPS. The reactions were started by the addition of PAPS, incubated for 10 or 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) as described [6]. The apparent K_m values were determined using the Enzyme Kinetics module in Sigma Plot.

CD measurements

Far UV CD spectra were recorded in 20 mM TRIS/HCl, pH 8.0, 1 mM DTT as described previously [30] at 25°C in a Jasco J-810 spectropolarimeter (Jasco, Great Dunmow, UK) at a scan rate of 5 nm x min^{-1}, a response of 0.25 s, and a bandwidth of 1 nm and corrected for the buffer baseline. Data were collected at 0.1 nm intervals from 190 to 260 nm. Analyses were performed at a protein concentration of 5 µM. Concentration and different numbers of aa were taken in consideration. At least three accumulations were done for each spectrum.

Bioinformatic analysis

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. The sequences of the ecotypes on DNA level were aligned by Clone Manager 9© using the application "align multiple sequences" (evaluation matrix: standard linear, mismatch penalty: 2, gap open penalty: 4, gap extension penalty: 1). For the comparison on aa level European Bioinformatics Institutes application "ClustalW2-multiple sequence alignment" (evaluation matrix: gonnet, gap open penalty: 10, gap extension penalty: 0.2, gap distance penalty: 5) was used (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Both alignment methods are based on neighbor joining algorithms. AtSOT18 from additional *Arabidopsis* ecotypes were taken from http://mus.well.ox.ac.uk/19genomes/.

The secondary structure prediction was performed by two open source applications, PredictProtein [31] and Domain Annotation [32] available on the Swiss-Model server (http://swissmodel.expasy.org/, 2.10.2011). To create a tertiary structure of the AtSOT18 Col protein a homology modeling was performed on the Swiss-Model server. First a suitable template of the Protein Data Bank (http://www.pdb.org/pdb/home/home.do, 2.10.2011) was identified by multiple sequence alignment with the application Template Identification. The quality of the model was estimated by the QMEAN Z-score [33]. The analysis of the model itself was performed using the Swiss-PDB Viewer, also available on the Swiss-Model server.

Nucleotide sequence data are available in the GenBank database under the accession numbers: AtSOT16: Bla-10, JQ478640; Cvi-0, JQ478641; Ler-1, JQ478642; Ws, JQ478643. AtSOT17: Bla-10, JQ614416; Ler-1, JQ614417; Ws, JQ614418. AtSOT18: Bla-10, JN648506; Cvi-0, JN648502; Ler-1, JN648501; Pi-0, JN648504; Sorbo, JN648505; Ws, JN648503.

Acknowledgements

We would like to thank Julia Volker, Hannover, and Dr. Michael Reichelt, MPI for Chemical Ecology, Jena, Germany, for the preparation of the substrates, and Prof. Dr. Jonathan Gershenzon, Jena, for constant support. We are grateful to Melina Henne, Hannover, who did the CD measurements. Dr. Marion Klein started the sulfotransferase topic in our laboratory as her PhD project and many methods applied in this publication where originally implemented by her. The project was funded by the DFG PA 764/10-1.

References

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

[2] M. Klein, J. Papenbrock, Sulfotransferases and their role in glucosinolate biosynthesis, in: N.A. Khan, S. Singh, S. Umar (Eds.), Sulfur Assimilation and Abiotic Stress in Plants, Springer Verlag, Heidelberg, 2008, pp. 149-166.

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

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

[5] M. Piotrowski, A. Schemenewitz, A. Lopukhina, A. Müller, T. Janowitz, E.W. Weiler, C. Oecking, Desulfoglucosinolate sulfotransferases from *Arabidopsis thaliana* catalyze the final step in the biosynthesis of the glucosinolate core structure, J. Biol. Chem. 279 (2004) 50717-50725.

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

[7] B.K. Petersen, S. Chen, C.H. Hansen, C.E. Olsen, B.A. Halkier, Composition and content of glucosinolates in developing *Arabidopsis thaliana*, Planta 214 (2002) 562-571.

[8] P.D. Brown, T.G. Tokuhisa, M. Reichelt, J. Gershenzon, Variation of glucosinolate accumulation among different organs and developmental stages of *Arabidopsis thaliana*, Phytochemistry 62 (2003) 471-481.

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

[10] U. Wittstock, D.J. Kliebenstein, V. Lambrix, M. Reichelt, J. Gershenzon, Glucosinolate hydrolysis and its impact on generalist and specialist insect herbivores, Phytochemistry 37 (2003) 101-125.

[11] K.L. Falk, J.G. Tokuhisa, J. Gershenzon, The effect of sulfur nutrition on plant glucosinolate content: physiology and molecular mechanisms, Plant Biol. 9 (2007) 573-581.

[12] I.E. Sønderby, F. Geu-Flores, B.A. Halkier, Biosynthesis of glucosinolates - gene discovery and beyond, Trends Plant Sci. 15 (2010) 283-290.

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

[14] B.A. Halkier, J. Gershenzon, Biology and biochemistry of glucosinolates, Annu. Rev. Plant Biol. 57 (2006) 303-333.

[15] R. Yatusevich, S.G. Mugford, C. Matthewman, T. Gigolashvili, H. Frerigmann, S. Delaney, A. Koprivova, U.I. Flügge, S. Kopriva, Genes of primary sulfate assimilation are part of the glucosinolate biosynthetic network in *Arabidopsis thaliana*, Plant J. 62 (2010) 1-11.

[16] S.G. Mugford, B.R. Lee, A. Koprivova, C. Matthewman, S. Kopriva Control of sulfur partitioning between primary and secondary metabolism, Plant J. 65 (2011) 96-105.

[17] M. Klein, J. Papenbrock, Kinetic parameters of desulfoglucosinolate sulfotransferases, Physiol. Plant. 135 (2008) 140-149.

[18] S.M. Kelly, N.C. Price, The use of circular dichroism in the investigation of protein structure and function, Curr. Protein Pept. Sci. 1 (2000) 349-384.

[19] D.M. Otterness, R.M. Weinshilboum, Human dehydroepiandrosterone sulfo-transferase-molecular cloning of cDNA and genomic DNA, Chem-Biol. Interact. 92 (1994) 145-159.

[20] Y. Kakuta, L.G. Pedersen, C.W. Carter, M. Negishi, L.C. Pedersen, Crystal structure of estrogen sulphotransferase, Nature Struct. Biol. 4 (1997) 904-908.

[21] F. Marsolais, L. Varin, Identification of amino acid residues critical for cataly-sis and cosubstrate binding in the favonol 3-sulfotransferase, J. Biol. Chem. 270 (1995) 30458-30463.

[22] J. Kroymann, S. Textor, J.G. Tokuhisa, K.L. Falk, S. Bartram, J. Gershenzon, T. Mitchell-Olds, A gene controlling variation in *Arabidopsis* glucosinolate composition is part of the methionine chain elongation pathway. Plant Physiol. 127 (2001) 1077-1088

[23] S.G. Mugford, N. Yoshimoto, M. Reichelt, M. Wirtz, L. Hill, S.G. Mugford, Y. Nakazato, M. Noji, H. Takahashi, R. Kramell, T. Gigolashvili, U.I. Flügge, C. Wasternack, J. Gershenzon, R. Hell, K. Saito, S. Kopriva, Disruption of adenosine-5'-

phosphosulfate kinase in Arabidopsis reduces levels of sulfated secondary metabolites, Plant Cell 21 (2009) 910-927.

[24] M.E. Møldrup, F. Geu-Flores, C.E. Olsen, B.A. Halkier, Modulation of sulfur metabolism enables efficient glucosinolate engineering, BMC Biotechnol. 11 (2011) 12-19.

[25] I.E. Sønderby, M. Burow, H.C. Rowe, D.J. Kliebenstein, B.A. Halkier, A complex interplay of three R2R3 MYB transcription factors determines the profile of aliphatic glucosinolates in *Arabidopsis*, Plant Physiol. 153 (2010) 348-363.

[26] T. Murashige, F. Skoog, A revised medium for rapid growth and bio assays with tobbaco tissue cultures, Physiol. Plant. 15 (1962) 473-479.

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

[28] G. Graser, N.J. Oldham, P.D. Brown, U. Temp, J. Gershenzon, The biosynthesis of benzoic acid glucosinolate esters in *Arabidopsis thaliana*, Phytochemistry 57 (2001) 23-32.

[29] M. Reichelt, P.D. Brown, B. Schneider, N.J. Oldham, E. Stauber, J. Tokuhisa, D.J. Kliebenstein, T. Mitchell-Olds, J. Gershenzon, Benzoic acid glucosinolate esters and other glucosinolates from *Arabidopsis thaliana*, Phytochemistry 59 (2002) 663-671.

[30] A. Bartels, F. Forlani, S. Pagani, J. Papenbrock, Conformational studies on *Arabidopsis* sulfurtransferase *At*Str1 with spectroscopic methods, Biol. Chem. 388 (2007) 53-59.

[31] B. Rost, Predicting one-dimensional protein structure by profile-based neural networks, Meth. Enzymol. 266 (1996) 525-539.

[32] D.T. Jones, Protein secondary structure prediction based on position-specific scoring matrices, J. Mol. Biol. 292 (1999) 195-202.

[33] P. Benkert, M. Biasini, T. Schwede, Toward the estimation of the absolute quality of individual protein structure models, Bioinformatics 27 (2011) 343-350.

[34] G. Le Gall, S.B. Metzdorff, J. Pedersen, R.N. Bennett, I.J. Colquhoun, Metabolite profiling of *Arabidopsis thaliana* (L.) plants transformed with an antisense chalcone synthase gene. Metabolomics 1 (2005) 181-198.

Supporting information

Fig. S1. Multiple sequence alignment of all AtSOT18 genotypes by ClustalW2.

Fig. S2A and S2B. Phylogenetic tree of all AtSOT16 and AtSOT17ecotypes by ClustalW2 using the neighbor-joining algorithm.

Table 1. Origin of *Arabidopsis* ecotypes.

Name	NASC number	Country	Altitude	Size (cm)
Bla-10	N982	Blanes/Gerona (Spain)	1-100 m	21-27
C24	Stock Institute of Botany	No country (laboratory breeding)	-	25-30
Col-0	N1092	Columbia (USA)	1-100 m	15-24
Cvi-0	N902	Cape Verdi Islands	1200 m	30
Ler-1	N1642	Landsberg am Lech (Germany)	-	16-25
Pi-0	N1454	Pitztal (Austria)	1000-1500 m	18-35
Sorbo	N931	Khurmatov, Tadjikistan	mountains	-
Wassilewskija (Ws)	Stock Institute of Botany	Vasljevici (Wassilewskija)/Dnjepr (USSR)	100-200 m	33-38

Table 2. Germination rates of seeds from different *Arabidopsis* ecotypes.

Germinated	**Not germinated (germination capacity <1.5%, source: NASC)**
Bla-10 (N982)	Ag-0 (N901)
C24	Bl-1 (N968)
Col-0 (N1092)	Ema-1 (N1637)
Cvi-0 (N902)	Hodja-Obi-Garm (N922)
Ler-1 (N1642)	Ita-0 (N1244)
Pi-0 (N1454)	Kas-1 (N903)
Sorbo (N931)	
Ws	

Table 3. Secondary structure prediction by PredictProtein for AtSOT18 proteins from different ecotypes. The secondary structures were differentiated in α-helices, β-sheets, and coils.

Ecotypes	α-helices (%)	β-sheets (%)	coils (%)
Bla-10	32.57	9.14	58.29
C24	32.29	9.14	58.57
C24G$_{301}$D	34.57	9.14	56.29
C24N$_{339}$K	33.14	9.14	57.71
Col-0	34.29	8.86	56.86
Cvi-0	33.14	10.00	56.86
Ler-1	32.00	9.43	58.57
Ws	31.71	9.14	59.14

Table 4. Determination of apparent V_{max} (pkatal mg^{-1} protein) and K_m values (µM) for the recombinant AtSOT proteins isolated from different ecotypes using ds-3MTP and ds-4MTB as substrates. For the determination of the K_m values at least eight data points were included (10 µM to 210 µM). PAPS concentration was kept constant at 60 µM. Each determination was carried out two times. Ds-3MTP, desulfo-3-methylthiopropyl Gl; ds-4MTB, desulfo-4-methylthiobutyl Gl. Values from C24 and Col-0 were taken from [17].

	V_{max} (pkatal mg^{-1} protein)		K_m (µM)	
Ecotypes	**ds-3MTP**	**ds-4MTB**	**ds-3MTP**	**ds-4MTB**
Bla-10	6933 ± 375	3550 ± 183	67 ± 3	58 ± 2
C24	99 ± 17	131 ± 17	100 ± 1	130 ± 14
Col-0	2763 ± 266	1883 ± 314	55 ± 7	43 ± 6
Cvi-0	2183 ± 52	3266 ± 140	60 ± 1	96 ± 1
Ler-1	1438 ± 65	4783 ± 396	90 ± 1	81 ± 3
Ws	703 ± 12	1133 ± 80	60 ± 1	60 ± 1

Table 5. Gl in the leaves of different *Arabidopsis* ecotypes. The data are taken from [9]. Quantities are given in µmol per g dry weight. C_3, sum of three carbon aliphatic Gl; C_4, sum of four carbon aliphatic Gl; C_4 per, $C_4/(C_3 + C_4)$.

Ecotypes	**C_3**	**C_4**	**C_4 per**
Bla-10	1.03	9.91	0.902
Col-0	0.40	5.93	0.937
Cvi-0	10.03	24.40	0.709
Sorbo	4.55	20.96	0.822
C24	Low[a]	High	
Ler-1	9.24	0.02	0.002
Pi-0	2.07	0.08	0.037
Ws	High[b]	Low	

[a]C24 contains mainly 3-butenyl Gl, a C4 Gl ([17] and unpublished results, Michael Reichelt, Jena, Germany). [b]Ws-0 contains mainly allyl Gl, C_3 Gl,according to [34].

Figure legends

Fig. 1. Phylogenetic tree of all AtSOT18 ecotypes by ClustalW2 using the neighbor-joining algorithm. The values in percent show the distance to the ecotype Col-0.

Fig. 2. Schematical view on the Col-0 sequence. The conserved amino acids (aa) and regions are marked in black lines and boxes. The exchanged aa of the analyzed ecotypes and mutants are marked in red lines. In the box below the scheme are the aa of Col-0 in black and the replacing aa in red.

Fig. 3. Secondary structure determination by CD spectroscopy for the genotypes C24 and Col-0, and the mutant $C24G_{301}D$. **A** Smoothed CD spectra. **B** Calculated secondary structure percentages divided in α-helices, β-sheets (divided in anti-parallel and parallel), β-turns, and random coils. Grey, C24; green, $C24G_{301}D$; blue, Col-0.

Fig. 4. Substrate specificity of the recombinant AtSOT18 protein from different ecotypes in pkatal mg^{-1} protein. Data for C24 and Col-0 are taken from [17].

Supplementary figures

Fig. S1. Multiple sequence alignment of all AtSOT18 genotypes by ClustalW2. Amino acids (aa) in red letters are small and hydrophobic inclusive the aromatic Y (AVFPMILW), aa in blue letters are acidic (DE), aa in magenta letters are alkaline (RK) and aa in green letters are hydroxyl, sulfanyl, amine and G (STYHCNGQ). A *, the aa at this position are identical all over the sequence. A :, the replacement of a very similar aa took place. A ., the replacement of a less similar aa took place.

Fig. S2A and S2B. Phylogenetic tree of all AtSOT16 and AtSOT17ecotypes by ClustalW2 using the neighbor-joining algorithm. The values in percent show the distance to the ecotype Col-0. A. AtSOT16 B. AtSOT17.

Figure(s)

Figure 1

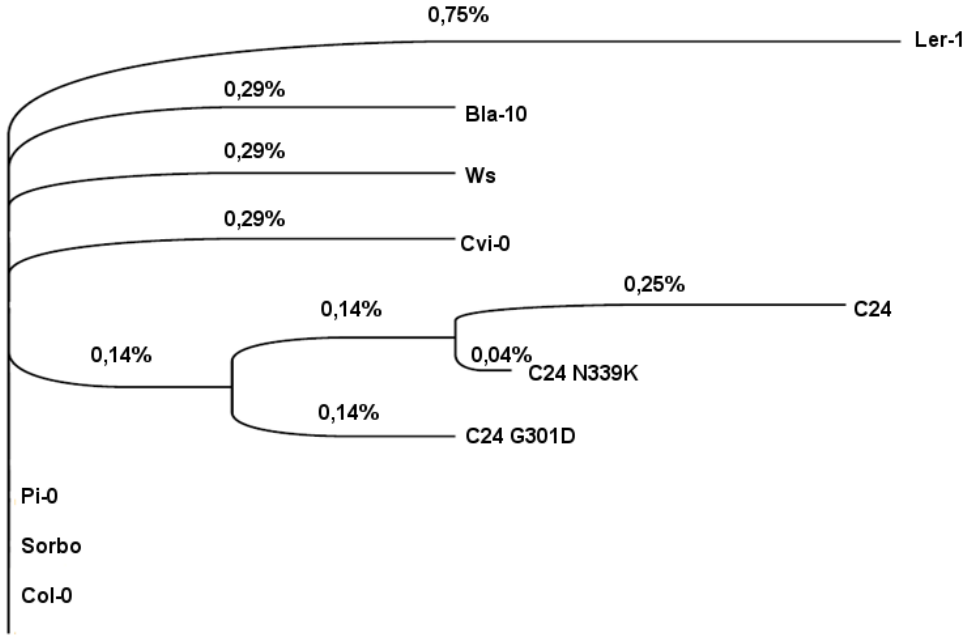


Figure 2

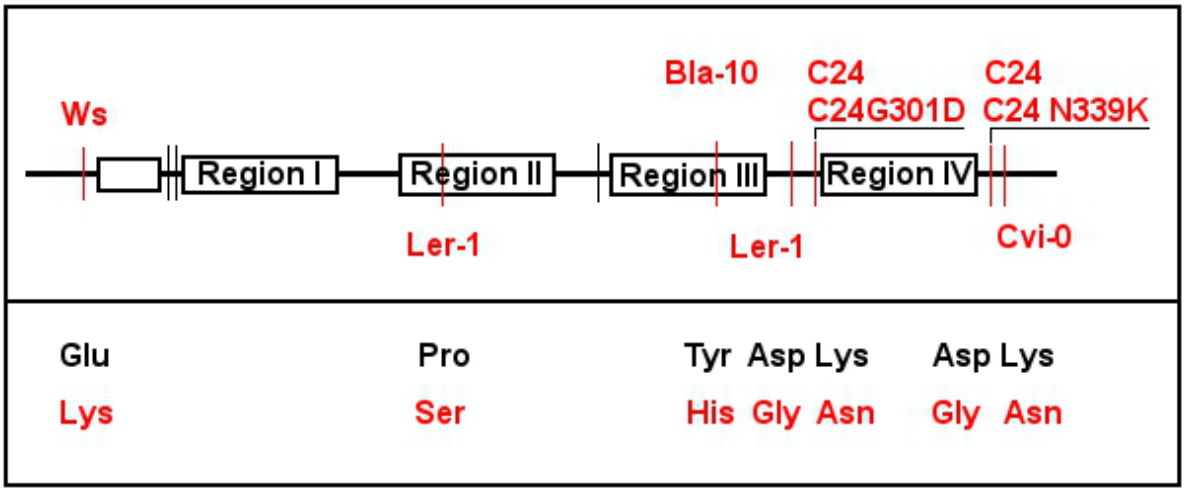

Bla-10
C24
C24
Ws
C24G301D
C24 N339K
Region I
Region II
Region III
Region IV
Ler-1
Ler-1
Cvi-0
Glu
Pro
Tyr Asp Lys
Asp Lys
Lys
Ser
His Gly Asn
Gly Asn

Figure 3A

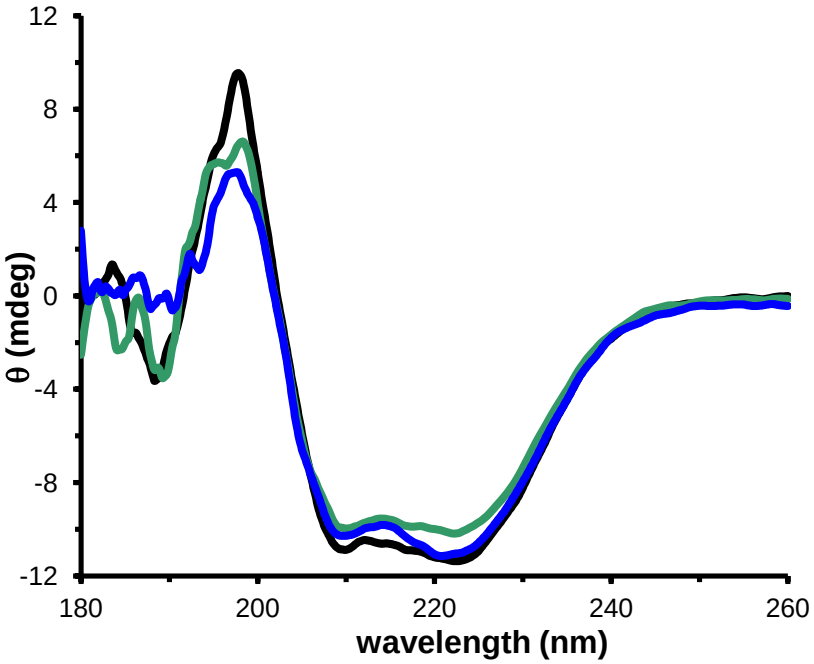


Figure 3B

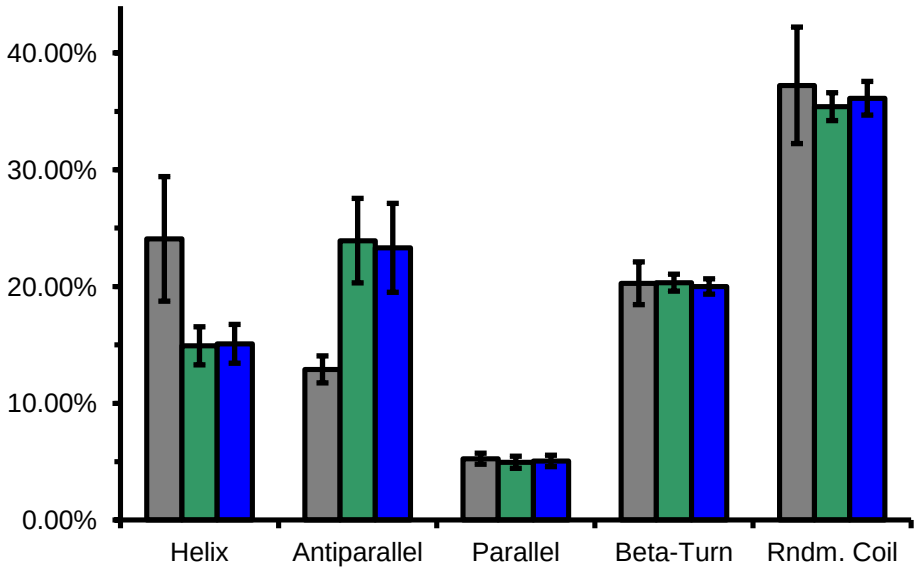
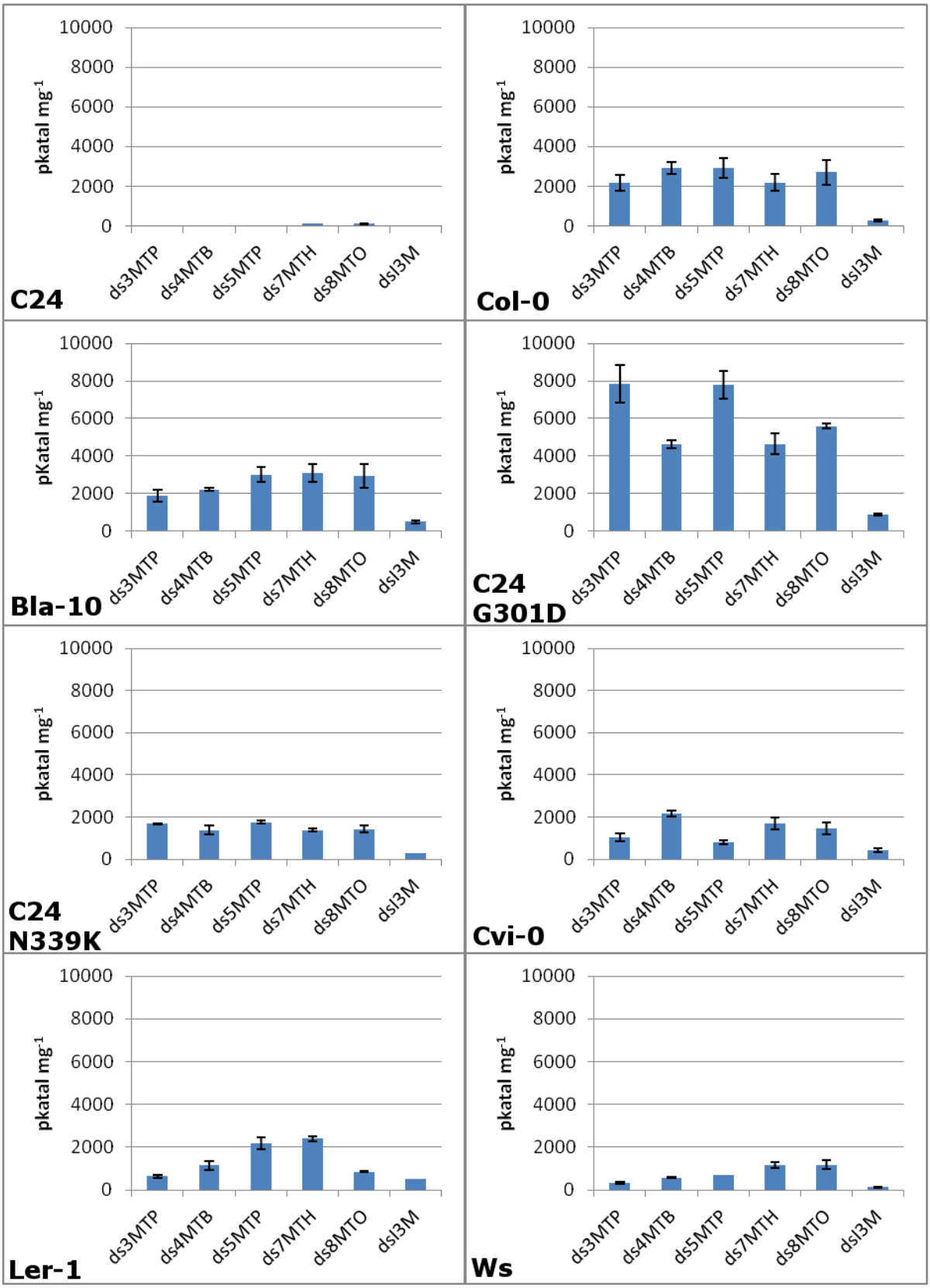


Figure 4


C24
Col-0
Bla-10
C24 G301D
C24 N339K
Cvi-0
Ler-1
Ws
pkatal mg-1
pKatal mg-1
10000
8000
6000
4000
2000
0
ds3MTP
ds4MTB
ds5MTP
ds7MTH
ds8MTO
dsI3M