Research Article   **Open Access**

Snap-shot of Serine Carboxypeptidase-like Acyltransferase Evolution: The Loss of Conserved Disulphide Bridge is Responsible for the Completion of Neo-functionalization

Felix Stehle[1], Franziska Götsch[1], Victor Wray[3], Jürgen Schmidt[2], Dieter Strack[1]* and Wolfgang Brandt[2]*

[1]Department of Secondary Metabolism, Leibniz Institute of Plant Biochemistry (IPB), Weinberg 3, D-06120 Halle (Saale), Germany
[2]Department of Bioorganic Chemistry, Leibniz Institute of Plant Biochemistry (IPB), Weinberg 3, D-06120 Halle (Saale), Germany
[3]Helmholtz Centre for Infection Research (HCI), Inhoffenstraße 7, D-38124 Braunschweig, Germany

Abstract

In this work, it is shown that the *At2g23010* gene product encodes 1-*O*-sinapoyl-β-glucose:1-*O*-sinapoyl-β-glucose sinapoyltransferase (SST). In contrast to all other functional characterized acyltransferases, the SST protein is highly specific towards this reaction only, and the substrate specificity was correlated to one amino acid substitution. Detailed sequence analysis revealed the lack of the disulfide bond S1 (C78 and C323 in the SMT (sinapoylglucose:malate sinapoyltransferase), that is in SST C80 and D327). The reconstitution of this disulfide bond led to an enzyme accepting many different substrates including disaccharides. Interestingly, the overall changes within the model structures are not very dramatic, but nevertheless, the enzyme models provide some explanations for the broadened substrate specificity: the reconstitution of the disulfide bond provoked more space within the substrate binding pocket simultaneously avoiding electrostatic repulsion. As the SST sequence of *A. lyrata* also showed the same mutation, the loss of the disulfide bond should has arisen at least 10 mya ago. A *Ka/Ks* ratio ≤ 1 supports the hypothesis that the loss of this disulfide bond was rather a specification towards a certain reaction than the beginning of a gene death. At the same time, this is also associated with the fixation in the genome.

Introduction

Intensive discussions on models addressing evolutionary fates of new gene copies are aimed at answering two basic questions: (i) "How do new gene copies originate in the genome?" and (ii) "How do they evolve and become fixed within the population?" Haldane and Muller first proposed that gene duplication could lead to new genes [1,2]. Later, the role of gene duplication has been reinforced and suggested it is most important for producing new genes [3]. Arguello et al. [4] suggested that gene duplication probably arises by unequal crossing over, as a result of homologous recombination between paralogous sequences or non-homologous recombination by replication-dependent chromosome cleavages. Beside gene duplication, new genes can additionally arise by retro position [5], horizontal gene transfer [6,7], and/or *de novo* origination from non-coding sequences [8]. A newly duplicated gene most likely becomes inactive (pseudogenization or non-functionalization) due to the accumulation of degenerative mutations. Only a few genes become fixed by mutations in the coding sequence or by changes in regulatory elements. The duplicated genes can achieve new functions (neo-functionalization) or maintain ancestral functions (sub-functionalization), thereby acting in a (partially) redundant manner [9-11].

Enzyme evolution can be studied by analysis of the coding sequences arranged in gene clusters to study the impact of mutations between "twins". The focus of this study was the gene product of the *At2g23010* gene, which is one of 51 Serine Carboxy Peptidase-Like (SCPL) genes encoded by the *Arabidopsis* genome [12], and one of five SCPL genes arranged in a tandem cluster on chromosome II. The corresponding enzymes share homology with Serine Carboxy Peptidases (SCPs), but are characterized by the lack of protease activity. These enzymes are able to acylate natural products, giving rise to the name SCPL-acyltransferases [13]. Sequence analysis showed that 19 of the 51 SCPL genes encode acyltransferases [14]. In contrast to the most common acyl transfer reactions catalysed by acyl CoA-utilizing

enzymes [15], belonging to the BAHD acyltransferase family [16], in sinapate metabolism, the energy-rich metabolites are represented by 1-*O*-acyl-glucose esters (β-acetal esters).

One of the neighbouring SCPL genes of *At2g23010* encodes the acyltransferase SMT (sinapoylglucose: malate sinapoyltransferase) [17-19], so far the best characterized SCPL acyltransferase. Based on the analysis of *Arabidopsis* mutants, it was proposed that the *At2g23010* gene encodes for a SST (1-*O*-sinapoyl-β-glucose:1-*O*-sinapoyl-β-glucose sinapoyltransferase) (Figure 1), whereas the *At2g2300* gene encodes for a SAT (1-*O*-sinapoyl-β-glucose:anthocyanin sinapoyltransferase [20]. The functions of the two remaining sequences of the cluster are still unknown. Beside this, only one other enzyme, the SCT (sinapoylglucose: choline sinapoyltransferase) [21,22] was characterized. For a further putative acyltransferase, the gene product of *At3g12203*, a benzoylation and sinapoylation activity of hydroxylated glucosinolates was recently suggested [23]. All these enzymes use sinapoylglucose as acyl donor, but differ in their acyl acceptor specificities [20,24-26].

Furthermore, the SCPL-dependent acyltransferases are not only restricted to the sinapate ester metabolism of Brassicaceaes. The isobutyroyl transferase from wild tomato (*Lycopersicon penellii*) represents the first identified SCPL-acyltransferase [13], in other species, and recently, the spreading of this acyltrasferases type was extended by the AsSCPL1 (SAD7), the first member of a new subfamily of monocot-specific SCPL acyltransferases [27].

The identity of the 5 SCPL acyltransferases on chromosome 2 in *Arabidopsis* regarding the amino acid sequence lies between 71 and 78% [20], indicating that they have evolved from a common ancestor probably by gene duplication. Thus, the structure–function relationships, reaction mechanism and sequence evolution of these enzymes represent a good tool to investigate molecular changes required to impart acyltransferase activity to hydrolytic enzymes.

This opens the opportunity to study the gain of new functions, as well as the further development of these new functions during evolution.

The focus of this work was the *At2g23010* gene product, which was expressed heterologously in *Saccharomyces cerevisiae*, exhibited a disproportionated activity towards 1,2-di-*O*-sinapoyl-β-glucose. Substrate specificity tests showed a high specific activity of the SST protein towards this reaction, and a conversion of other natural substrates including L-malate was not observable.

Furthermore, a detailed sequence comparisons with other SCPL acyltransferases revealed the lack of the disulphide bond S1 (C78 and C323 in the SMT, C80 and D327 in SST) [28]. The reconstitution of this disulphide bond led to an enzyme with a broad substrate spectrum, accepting several different substrates including disaccharides. To substantiate the differences between the wild type enzyme with two disulphide bridges and the mutant variant with three disulphide

bridges (Figure 2), both enzyme structures were modeled. The model revealed no dramatic changes within the overall enzyme structures, but docking studies showed enhanced space within the substrate binding pocket after disulphide bridge formation, thereby decreasing spatial and electrostatic repulsion caused by D327 and broadening of the substrate specificity. Sequence alignments with *Arabidopsis lyrata* additionally proved that the loss of the disulphide bridge within the SST gene occured at least 10 mya ago. Together with a *Ka/Ks* ratio smaller than one, it must be concluded that this loss is a move towards specific reactions, associated with the fixation in the genome. Taken together, these results gain further insights into the evolution and specification of the SCPL acyltransferase family.

Materials and Methods

Constructs for expression of SMT and SST in *S. cerevisiae*

The cDNA variants designed for expression in *S. cerevisiae* were amplified by PCR with primers attaching restriction sites for *Hin*dIII and *Xba*I to the 5'- and 3'-ends of the product. By cloning as *Hin*dIII-*Xba*I fragments into the expression vector pDIONYSOS [25], the PCR products were transcriptionally fused to the galactose-inducible yeast GAL1 promoter. Modification of the 5'-UTR was introduced *via* PCR by a modified forward primer. Site-directed mutagenesis was performed with the QuickChange© XL Site-Directed Mutagenesis
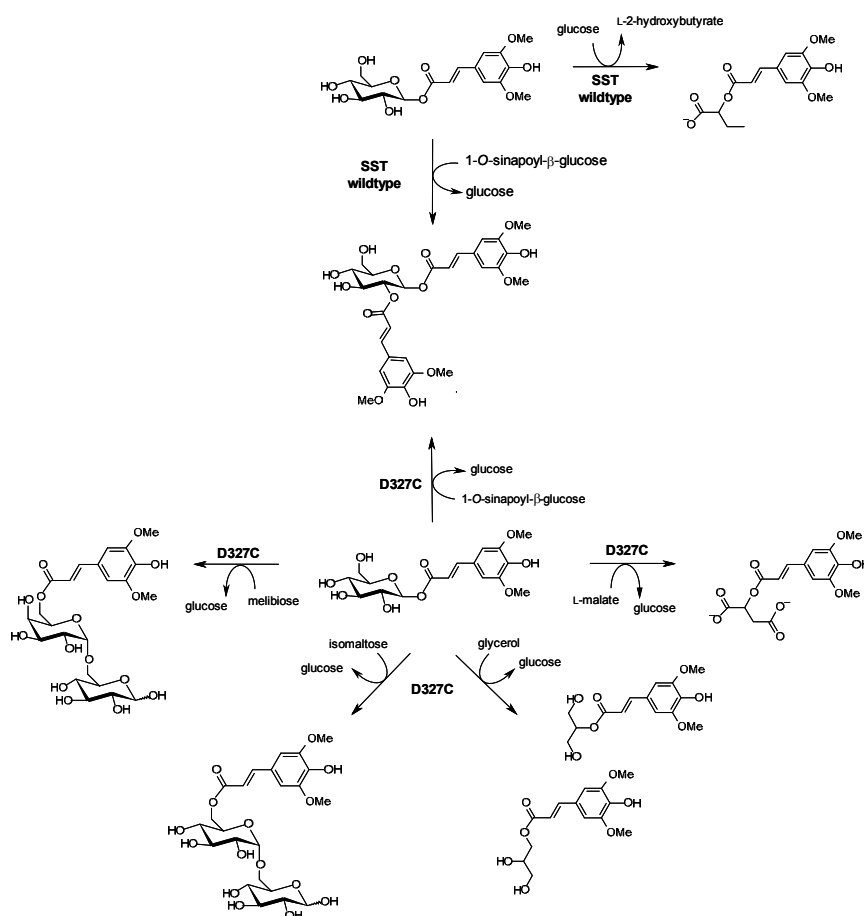
Figure 1: Scheme of the wild type SST (1-*O*-sinapoyl-β-glucose:1-*O*-sinapoyl-β-glucose sinapoyltransferase) and the SST-D327C variant catalyzed reactions dependent on 1-*O*-sinapoyl-β-glucose. The SST wild type enzyme catalyzes the formation of 1,2-di-*O*-sinapoyl-β-glucose and sinapoyl-L-2-hydroxybutyrate. The SST D327C variant additionally catalyzes the formation of 1,2-di-*O*-sinapoyl-β-glucose. Furthermore, it converts the disaccharides isomaltose and melibiose, as well as L-malate and glyceol forming 6-*O*-sinapoylisomaltose, 6-*O*-sinapoylmelibiose, sinapoyl-L-malate and sinapoylglycerol, respectively.

Figure 2: SST and SST-D327C activities (pkat L^{-1} culture protein; mean ± s.d.) at pH 6.0.
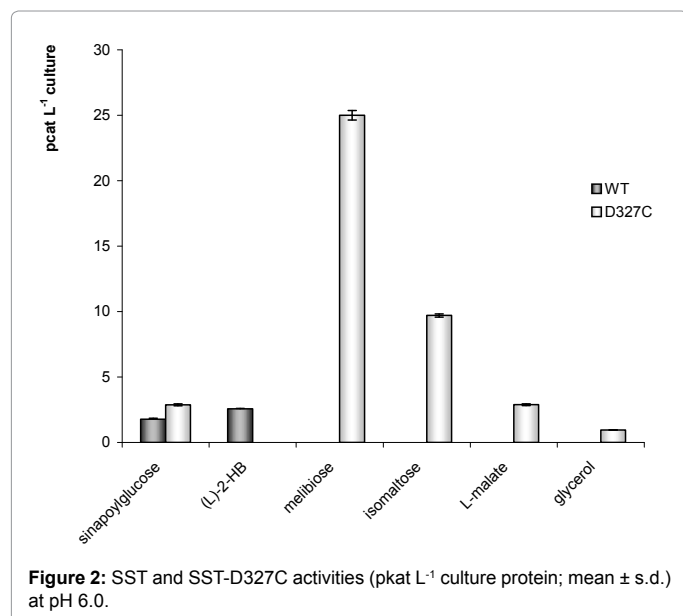
Kit (Agilent Technologies, Inc., Santa Clara, CA, USA), according to the protocol given by the supplier. Design and synthesis of the ySST sequence adapted to the codon usage of *S. cerevisiae* was performed by GeneArt (Life Technologies GmbH, Darmstadt).

Expression of SMT and SST in *S. cerevisiae*

For transformation, competent cells of *S. cerevisiae* INVSc1 (Life Technologies GmbH, Darmstadt) and BY4741 *Δgal1* (MATa *his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; YBR020w::kanMX4;* Euroscarf, Frankfurt) were prepared using the *S. cerevisiae* EasyCom Kit (Life Technologies GmbH, Darmstadt), and transformed according to the protocol given by the supplier. *S. cerevisiae* cells harbouring AtSMT or AtSST expression plasmids, were grown in synthetic drop-out medium without uracil or leucine to an attendance of 1 at $D600$ nm, respectively. After harvesting (1500 g, room temperature, 5 min), the cells were resuspended in drop-out media without uracil or YPD media, respectively. Induction of AtSMT and AtSST expression was initiated by adding galactose and raffinose to a final concentration of 2% (w/v) (INSc1 cells) or 0.5% (w/v) (BY4741 cells), and 4% (w/v) for raffinose. Cells were cultivated in the presence of the inductor galactose for an additional 36 h, and then harvested and disrupted, as described previously [28].

Enzyme assays

The enzyme assays with the recombinant SMT protein contained 1 mM 1-*O*-sinapoyl-β-glucose as substrate in the presence or absence of 10 mM L-malate in a total volume of 100 μL 100 mM MES buffer (pH 6.0), containing 5% (v/v) DMSO. The enzyme activity assays with the recombinant SST proteins were performed with 2.5 mM 1-*O*-sinapoyl-β-glucose for detection of the disproportionate reaction. For monitoring other activities, additionally 60 mM melibiose, 60 mM isomaltose, 90 mM L-malate or 60 mM L-2-hydroxybutyrate were present in the assay, respectively. After incubation at 30°C for 30 min, the reaction was terminated by adding TFA to a final concentration of 10% (v/v). Product formation was analyzed by HPLC. After centrifugation, assay aliquots were injected onto a Nucleosil C18 column (3 μm; 120×4 mm i.d.; Macherey-Nagel, Düren). Product separation was performed using a 5-min linear gradient at a flow rate of 0.7 mL min^{-1} from 10% to 40% MeCN in 1% (v/v) formic acid. Compounds were photometrically detected (maxplot between 210 and 500 nm), using a Waters (Eschborn)

2996 photodiode array detector. The quantitative data, calculated as equivalents of 1-*O*-sinapoyl-β-glucose, were represented as mean values (± s.d.) from three independent replicates. The pH optimum was determined in citrate buffer (pH 3.0), sodium acetate buffer (pH 4.0 and 5.0), MES buffer (pH 6.0), Tris buffer (pH 7.0-9.0) and CAPS buffer (pH 10.0), at a concentration of 100 mM for each buffer system. The protein samples were dialysed against the corresponding buffers, and then used in enzyme assays.

Sequence comparisons and phylogenetic analysis

Alignment of 162 plant SCPL proteins was performed using the BioEdit 7.0.4 software package (http://www.mbio.ncsu.edu/bioedit/bioedit.html), with default parameters. Therefore, predicted mature protein sequences were used in phylogenetic analysis, excluding signal peptides and linker regions. Construction of neighbour-joining phylogeny and assessment of sequence diversity was performed using the Mega 4.0 software package [29]. Evolutionary distances were computed using the Poisson correction method. All positions containing gaps and missing data were eliminated from the data set (complete deletion option).

To study the selection pressure on AlSST, *Ka/Ks* ratios (non-synonymous substitution rate/synonymous substitution rate) between AlSST and AtSST were calculated using the method described by Goldman and Yang [30], using the software package KaKs_Calculator [31].

Homology modelling

The 3D-structure of SST was modelled with the molecular modelling software YASARA [32]. YASARA identified six templates based on alignment scores and low E-values suitable for homology modelling of SST. Altogether, 46 models were automatically created and subsequently refined. The model based on the X-ray structure of wheat serine carboxypeptidase II, deposited in the pdb-database (3SC2) [33], resulted with the best quality Z-score of -1.718. The model was refined with the md-refinement tool of YASARA. The quality of the final model was checked with PROSA II [34], and PROCHECK [35]. The graphical analysis with PROSA II showed two small loop areas within the positive energy range, but the combined energy z-score of −9.37 clearly indicate a native like folded structure. All analysis with PROCHECK evaluated the model insight or better for all stereo-chemical parameters, e.g. the Ramachandran plot quality (89.3% of the backbone dihedral angles in most favoured areas). The ligands were docked using the GOLD-suite 5.0.1 software package [36,37], also allowing rotation of amino acid side chains, as described in the results chapter. For each ligand, 30 docking runs with GOLD-score and standard settings of GOLD were performed. The structural model of the D327C mutant was generated simply by mutation of Asp327 to Cys, and subsequent formation of the disulphide bond with Cys80 by using the SYBYL8.0 software package (Tripos International, 1699 South Hanley Road, St. Louis, MO 63144-2319 USA), followed by an energy optimization for these two residues. The formed disulphide bond coincides exactly with the one Cys56-Cys303 occurring in the template protein (3sc2).

LC–ESI-MS/MS

The ESI mass spectra were obtained from a TSQ Quantum Ultra AM system (Thermo Electron), equipped with a hot ESI source (HESI), electrospray voltage 3.0 kV, sheath gas: nitrogen; vaporizer temperature: 50°C; capillary temperature: 250°C. The MS system is coupled with a Surveyor Plus micro-HPLC (Thermo Electron), equipped with an Ultrasep ES RP18E column (5 μm, 1×100 mm, SepServ). HPLC separation was achieved by using a 15-min linear gradient at a flow rate

of 50 µL min^{-1} from 10% to 95% MeCN in 0.2% aq. HOAc, with the latter held at 95% MeCN for another 15 min.

The collision-induced dissociation (CID) mass spectra were recorded during the HPLC run with the indicated collision energies (collision gas: argon; collision pressure: 1.5 mTorr). The corresponding ESI-MS/MS data of sinapic acid, 1-O-sinapoyl-β-D-glucose and 1,2-di-O-sinapoyl-β-D-glucose were previously published [24].

6-O-Sinapoylmelibiose: RT^{HPLC}=7.3 min; 15 eV positive ion CID mass spectrum (m/z, rel. int. (%)): 549 ([M+H]$^+$, -), 369 (46), 207 (sinapoyl cation, 100).

Sinapoyl-L-2-hydroxybutyrate: RT^{HPLC}=21.5 min, 15 eV positive ion ESI-CID mass spectrum (m/z, rel. int. (%)): 311 ([M+H]$^+$, 1), 207 (sinapoyl cation, 100), 175 (22); 15 eV negative ion ESI-CID mass spectrum (m/z, rel. int. (%)): 309 ([M-H]$^-$, 4), 223 (sinapate anion, 42), 205 ([m/z 223-H2O], 100), 190 [8].

NMR

All 1D and 2D ^{1}H NMR spectra were recorded at 300 K on a Bruker AVANCE DMX 600 NMR spectrometer locked to the major deuterium signal of the solvent, CD3OD. Chemical shifts are given in ppm relative to the residual solvent signal at 3.35 ppm and coupling constants in Hz.

^{1}H NMR 6'-sinapoyl-α-isomaltose (CD3OD): 7.70 (d, 1H, J=15.9 Hz, H-7''), 6.97 (s, 2H, H-2''/6''), 6.44 (d, 1H, J=15.9 Hz, H-8''), 5.16 (d, 1H, J=3.7 Hz, H-1), 4.88 (d, 1H, J=3.7 Hz, H-1'), 4.50 (dd, 1H, J=11.8, 2.1 Hz, H-6'A), 4.35 (dd, 1H, J=11.8, 5.8 Hz, H-6'B), 4.00 (m, 1H, H-5'), 3.93 (s, 3''/5''-OCH3), 3.44, 3.41 (m, H-2, H-2'), 4.05-3.30 (remaining protons of glucose units).

^{1}H NMR 6'-sinapoyl-β-isomaltose (CD3OD): 7.69 (d, 1H, J=15.9 Hz, H-7''), 6.97 (s, 2H, H-2''/6''), 6.45 (d, 1H, J=15.9 Hz, H-8''), 4.88 (d, 1H, J=3.7 Hz, H-1'), 4.55 (d, 1H, J=7.8 Hz, H-1), 4.48 (dd, 1H, J=11.9, 2.1 Hz, H-6'A), 4.38 (dd, 1H, J=11.9, 5.4 Hz, H-6'B), 4.00 (m, 1H, H-5'), 3.93 (s, 3''/5''-OCH3), 3.44 (m, H-2'), 3.19 (dd, 1H,J=9.1, 7.9 Hz, H-2), 4.05-3.30 (remaining protons of glucose units).

^{1}H NMR 6'-sinapoyl-α-melibiose (CD3OD): 7.71 (d, 1H, J=15.9 Hz, H-7''), 6.96 (s, 2H, H-2''/6''), 6.42 (d, 1H, J=15.9 Hz, H-8''), 5.16 (d, 1H, J=3.7 Hz, H-1), 4.92 (d, 1H, J=3-4 Hz, H-1'), 4.40 (dd, 1H, J=11.3, 7.7 Hz, H-6'A), 4.25 (m, 1H, ΣJ=~ 14 Hz, H-5'), 3.93 (s, 3''/5''-OCH3), 3.38 (m, H-2 or H-2'), 4.05-3.30 (remaining protons of glucose/galactose units).

^{1}H NMR 6'-sinapoyl- -melibiose (CD3OD): 7.70 (d, 1H, J=15.9 Hz, H-7''), 6.96 (s, 2H, H-2''/6''), 6.43 (d, 1H, J=15.9 Hz, H-8''), 4.93 (d, 1H, J=3-4 Hz, H-1'), 4.55 (d, 1H, J=7.8 Hz, H-1), 4.42 (dd, 1H, J=11.4, 7.5 Hz, H-6'A), 4.34 (dd, 1H, J=11.3, 4.7 Hz, H-6'B), 4.25 (m, 1H, ΣJ=~14 Hz, H-5'), 3.93 (s, 3''/5''-OCH3), 3.38* (m, H-3), 3.18 (dd, 1H, J=9.0, 8.0 Hz, H-2), 4.05-3.30 (remaining protons of glucose/galactose units).

Results and Discussion

Expression of the SST protein in *S. cereviaiae*

The full-length *At2g23010* cDNA was used to adopt the optimizations found to be essential for the expression of the SMT gene in Baker's yeast [25]. However, *Saccharomyces cerevisiae* INVSc1 cells carrying the codon usage optimized AtSST cDNA fused to PEP4 signal peptide did not develop detectable SST activities.

For this reason, a further optimization of the expression in yeast was necessary. Since the SST activity was biochemically not yet proven, the expression was optimized based on the SMT construct. For this

purpose, the sequence motif near the ATG translation initiation codon of AtSMT was adopted according to the consensus sequence proposed by Hamilton et al. [38], to achieve a better initiation of translation. In the resulting sequence (AAA AAA ATG TCT), the corresponding amino acids were not modified and are unaltered compared to the unoptimized construct. The construct was named yUTR-PEP4-Ysmt, and led to an intracellular SMT activity in the range of 1300 pkat L^{-1} culture, indicating a 1.6-fold increase compared to the reference construct (PEP4-ySMT) (Supplementary Figure 1).

As INVSc1 yeast cells metabolize the inductor galactose, a yeast strain carrying a mutation in the *GAL1*-gene was tested to achieve a permanent high gene induction. Therefore, both constructs (PEP4-ySMT and yUTR-PEP4-ySMT) were transformed in the yeast strain *BY4741 Δgal.* In both cases, the SMT activity assayed in the crude protein extract from these cells indicated a 12.5-fold higher SMT yield compared to the INVSc1 based expression of the corresponding constructs (Supplementary Figure 1).

Taken together, these optimizations led to a 20-fold higher protein yield, and these improvements were transformed to the SST construct. The expression of the resulting construct in the *GAL1*-deficient cells resulted in the formation of a new compound with absorption characteristics of sinapate esters (RT, 2.5 min; max, 330 nm). Co-chromatography with standard compounds, as well as LC–ESI-MS/MS analysis, identified this compound as 1,2-di-O-sinapoyl-β-glucose. The negative ion ESI mass spectrum of 1,2-di-O-sinapoyl-β-glucose ([M-H]$^-$) at m/z 591 showed the key ions at m/z 367 ([M-H–sinapic acid]$^-$) and m/z 223 (sinapate anion) [24]. The activity of this enzyme was calculated to be approximately 2 pkat L^{-1}. Nevertheless, these result demonstrated that the *At2g23010* gene encodes an enzyme converting two 1-O-sinapoyl-β-glucose molecules forming 1,2-di-O-sinapoyl-β-glucose.
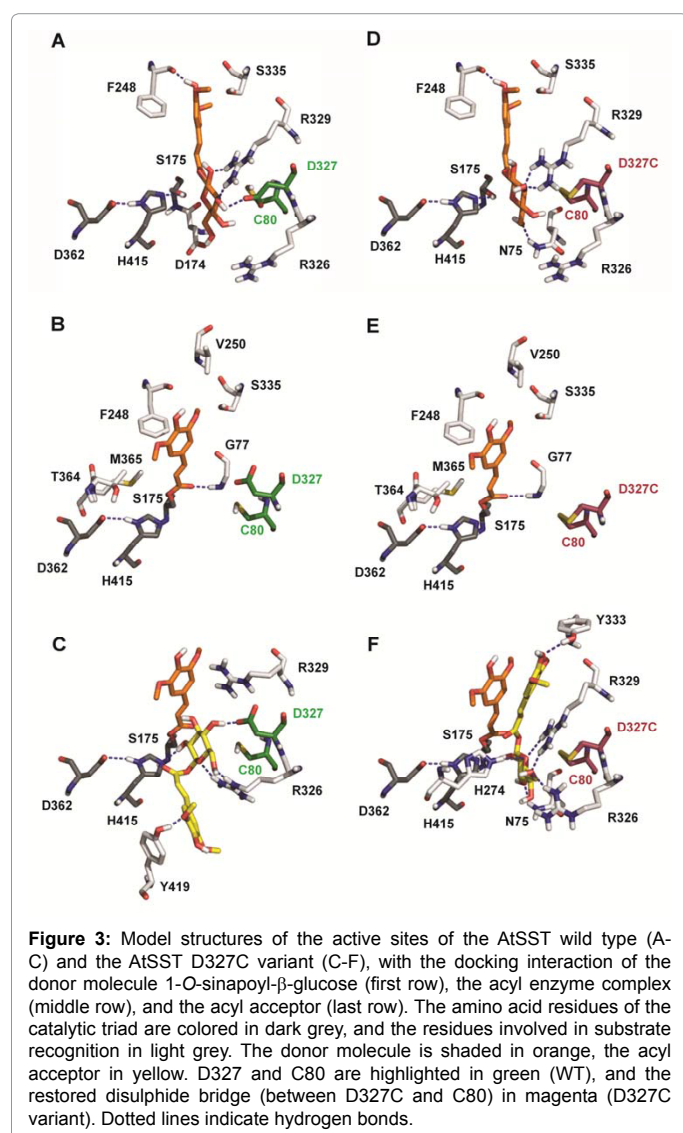
Characterization of the SST protein

The relative low specific activity of the SST protein of 2 pkat L^{-1} prompted us to analyze its amino acid sequence. The SST shows a homology of 78% on amino acid level compared to the SMT. It is worthy of note that almost all residues essential for substrate recognition in the SMT are conserved in the SST protein. Based on these findings, L-malate was tested as acyl donor, but no sinapoylmalate formation could be detected. In addition, L-2-hydroxybutyrate and D-3-hydroxybutyrate, which are structurally related to L-malate, were also tested in an SST activity assay. These molecules were not used as SMT substrates, but show inhibitory properties due to the absence of the second carboxyl group [25]. Unexpectedly, the SST showed an activity towards the formation of sinapoyl-L-2-hydroxybutyrate (Supplementary Figures 2 and 3), resulting in an activity of 0.04 pcat mg^{-1}. The positive ion ESI mass spectrum showed a [M+H]$^+$ ion at m/z 311, and in its CID mass spectrum, a prominent sinapoyl cation at m/z 207 (base peak), indicating a sinapoylated hydroxybutyric acid (see experimental section). However, with D-3-hydroxybutyric acid no conversion was observed.

The activity of *Arabidopsis* SST was examined at pH values ranging from 4.0 to 10.0 at 30°C. The maximum activity was observed in the pH range 8.0 to 10.0. The pH values for half maximal activity were pH 6.3 and 9.6, respectively.

Lack of a disulphide bond is responsible for the high substrate specificity. A sequence alignment between the SMT and the SST protein showed the lack of the disulphide bond S1 formed in the SMT between C78 and C323 [28]. In the SST protein, the corresponding

disulphide bridge was not formed due to the amino acid substitution in position 327 (C327D). Interestingly, the disulphide bridge S1 is highly conserved among the serine carboxypeptidase and the SCPL-acyltransferase family [24]. Hence, the substitution in the SST provides a possible reason for the low, but very specific activity of this enzyme. To test this hypothesis, the SST mutant variant D327C was generated. The expressed variant showed an activity of 3 pcat L^{-1} for the conversion of two molecules 1-O-sinapoyl-β-glucose, which is 1.5-fold larger than the wild type activity (Supplementary Figures 2 and 3).

Whether to test the amino acid substitution in position 327 towards a cysteine (D327C) results in the formation of an additional disulphide bond between Cys80 and the newly introduced Cys327, a thermal denaturation curve was recorded. The SST-D327C variant and the wild type enzyme were incubated for 10 min at different temperatures, and subsequently subjected to enzyme activity assays. The resulting relative activities were proportional to the fraction of denaturated protein f_D. The midpoint of denaturation (T_m) was determined to be approximately 61°C for the wild type enzyme and 68°C for the mutant variant, respectively. The significant higher T_m value of almost 7°C for the D327C variant indicated the formation of an additional disulphide bond compared to the wild type enzyme (Supplementary Figure 2).

Figure 3: Model structures of the active sites of the AtSST wild type (A-C) and the AtSST D327C variant (C-F), with the docking interaction of the donor molecule 1-O-sinapoyl-β-glucose (first row), the acyl enzyme complex (middle row), and the acyl acceptor (last row). The amino acid residues of the catalytic triad are colored in dark grey, and the residues involved in substrate recognition in light grey. The donor molecule is shaded in orange, the acyl acceptor in yellow. D327 and C80 are highlighted in green (WT), and the restored disulphide bridge (between D327C and C80) in magenta (D327C variant). Dotted lines indicate hydrogen bonds.

Enzyme assays with the crude extract of the recombinant expressed variant revealed the formation of a new highly converted compound. Analysis *via* mass spectrometry suggested that the sinapoyl residue is linked to a disaccharide. The structure elucidation by NMR spectroscopy indicated this new compound was 6-O-sinapoylmelibiose. The new activity was found to be 14 times higher compared to the disproportionation (Supplementary Figures 2 and 3). This could be confirmed by an enzyme activity assay using the desalted crude extract of the SST-D327C variant and melibiose as acyl acceptor. This prompted us to test other disaccharides as acyl acceptors. Using isomaltose as acyl acceptor resulted in a product with characteristic sinapic acid derived UV absorption spectrum, and a retention time longer than sinapoylmaltose. The mass spectrometric analysis of this compound yielded a hint to a sinapoyl-disaccharide, and the structure elucidation using NMR spectroscopy identified it as 6-O-sinapoylisomaltose.

The 1D ^{1}H NMR spectrum of the unknown product (Peak 3, assayed as containing isomaltose) indicated a compound with a free terminal anomeric unit as two signals with characteristic chemical shifts and coupling constants, α- (5.16 ppm, J=3.7 Hz) und β-protons (4.55 ppm, J=7.8 Hz), in a ratio of 1:1.5 were observed. A second strong signal of an α-anomeric proton was also observed (4.88, J=3.7 Hz), arising from the α-proton of the internal glucose unit of both forms of the isomaltose unit. Characteristic signals and their corresponding intensities confirmed the presence of the sinapoyl moiety, which must be attached to C-6' of the isomaltose unit from the low field shift of the corresponding protons (H-6'AB). A similar analysis of the product (double peak (2) see 3.1.1), containing melibiose unambiguously identified this as the anomeric forms of 6'-sinapoylmelibiose.

However, all other mono- and disaccharides tested did not result in product formation. Both, melibiose and isomaltose are 1,6 linked disaccharides, whereas galactose or glucose is linked with a glucose molecule, respectively. Isomaltose where glucose is linked with 1,6 to fructose was not accepted as acyl acceptor.

Given that the SST-D327C variant possesses almost all conserved residues responsible for catalysis in the SMT, as well as a functional disulphide bridge S1, this mutant variant was tested for SMT activities. First, the primary activity of SMT, the formation of sinapoyl-L-malate, was tested. Within the HPLC diagram of an assay containing 1-O-sinapoyl-β-glucose and L-malate as substrates, a peak of sinapoyl-L-malate was detected. Furthermore, this could be confirmed by liquid chromatography and mass spectrometric analysis. Second, the side activity of the SMT, the formation of 1,2-di-O-sinapoyl-β-glucose [24] was tested. The presence of 1-O-sinapoyl-β-glucose, serving as both acyl donor and acceptor, a 1,2-di-O-sinapoyl-β-glucose peak was detected and confirmed by a mass spectrometric analysis. These data clearly showed that the SST-D327C variant is capable of synthesizing sinapoyl-L-malate and 1,2-di-O-sinapoyl-β-glucose, thereby displaying activities of the SMT protein. Nevertheless, the SST variant shows substantial differences compared with SMT. Not only the significantly lower specific activity towards the formation of these products discriminate between both enzymes, but also the fact that the SMT is not able to accept melibiose and isomaltose as substrates (data not shown). It is also worth mentioning that a product formation using glycerol as acyl acceptor could be detected (not shown). Nevertheless, this activity was just above the detection limit and not further investigated.

Model structures

The docking results of the substrate 1-O-sinapoyl-β-glucose showed perfect docking arrangements in both the wild type, as well as in the D327C variant, to allow catalytic cleavage of the ester bond

supported by the catalytic triad S175, H415 and D362 (Figures 3a-3c). The glucose moiety is well recognized by forming a hydrogen bond with R329 and D174. In the case of the wild type enzyme, an additional hydrogen bond is formed with D327. The sinapoyl group formed a T-shaped van der Waals interaction with the side chain of F248, weak hydrophobic interactions with P332 (not shown), and a hydrogen bond with the backbone carbonyl group of F248. The oxygen atom of the active serine is almost ideally placed perpendicular to the plane of the ester bond to be cleaved, and the carbonyl group of the ester bond formed a hydrogen bond with the backbone amide of G77 to stabilize the intermediate tetrahedral oxanion intermediate (see also figure 3b representing the formed acyl enzyme intermediate after removal of the glucose). So far, there were no striking differences between the wild type enzyme and the D327C variant. However, the analysis of the docking results of the second substrates to the acyl enzyme (assuming a ping pong mechanisms), some interesting differences obviously explaining the experimental results could be detected.

The docking arrangement of 1-*O*-sinapoyl-β-glucose as a second substrate significantly differed between both enzymes. Whereas the sinapoyl moiety interacts with Y419 within the wild type enzyme (pointed to the bottom in Figure 3c), the arrangement in the D327C variant was just the opposite, i.e. sinapoyl was forming a hydrogen bond with the hydroxyl group of Y333. Despite these differences, the C4-OH group of glucose forms a hydrogen bond with the active side histidine (to abstract the C4-OH proton from the acyl donor), perpendicular to the ester bond of the acyl enzyme in both cases. In wild type enzyme, glucose was recognized by forming hydrogen bonds with R326 and
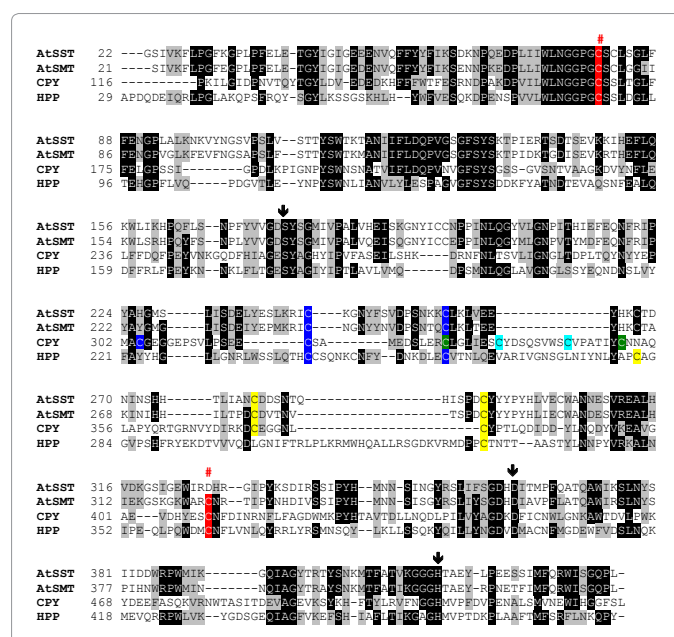

Figure 5: Alignment of amino acid sequences of SCPL acyltransferases with serine carboxypeptidases. N-terminal leader pepties were removed. Fully conserved residues are shaded in black. Gray shading indicates conservation of at least 70%. The disulphide bridges are coloured. The number of disulphide bridges varies between three in SST, SMT and HPP, and five in CPY. The disulphide bridge designated as S1 (red) [28] is highly conserved among the SCPL enzymes. The missing S1 bride in the SST is marked by a red rhomb. The disulphide bridge S2 is marked in blue and S3 in yellow. The two additional disulphide bonds in the CPY are shaded in green and cyan. Amino acid residues forming the catalytic triad are marked with a black arrow. AtSST (*A. thaliana* SST; SwissProt Accession O64811) was aligned with AtSMT (*A. thaliana* SMT; Q8RUW5), CPY (*S. cereviaiae*; P00729)and HPP (human protective protein; P10619) by CLUSTALW using the BLOSUM62 matrix [43,44].
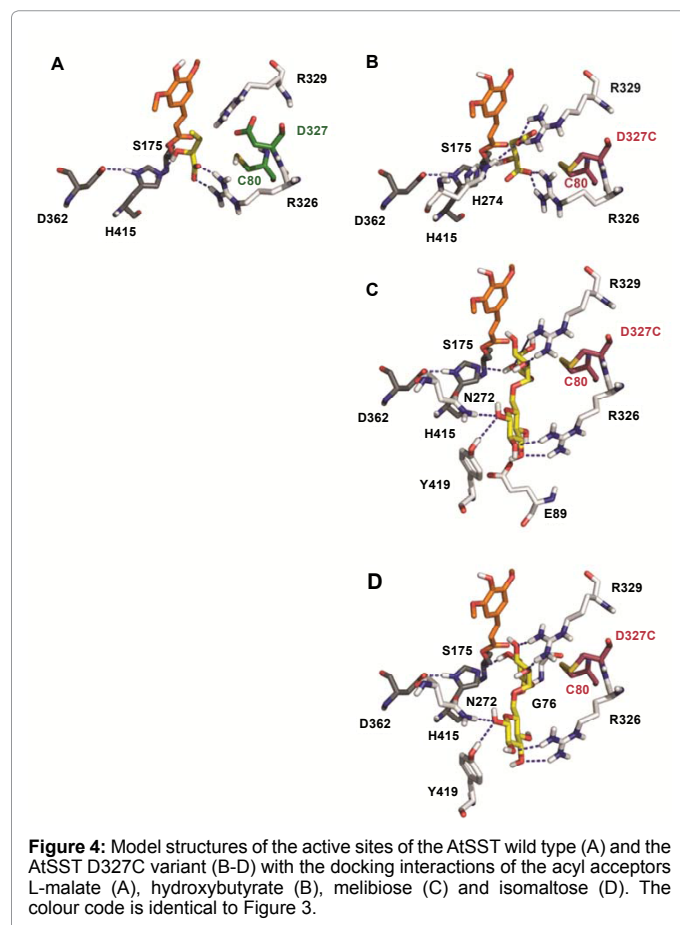
Asp327 (Figure 3a), whereas it formed hydrogen bonds with R329 in the D327C variant instead. A closer look towards the different docking arrangements verified the 2-fold higher conversation rate of the D327C variant in case of the disproportionation reaction. There was no steric hindrance for the intermediate glucose anion perpendicular to the carbonyl group of the acyl ester bond for the D327C variant, whereas in the neighbouring C3-OH group, pointing also to the carbonyl group, might have caused steric clashes for the wild type enzyme.

The most striking differences between both enzymes, however, were the acceptance of L-malate, melibiose, isomaltose and glycerol (with a low rate) (not shown), as second substrates in the case of the D327C variant. In contrast, the wild type enzyme displayed substrate specificity, accepting only the artificial L-2-hydroxybutyrate as further acyl acceptor beside 1-*O*-sinapoyl-β-glucose L-2-hydroxybutyrate formed a salt bridge with the side chain of R326, and was oriented well for activation of the hydroxyl group by H415 to allow reaction (Figure 4a and Figure 4b). In the case of L-malate, however, the negatively charged D327 prevents correct docking of this compound due to the second carboxylic acid group compared to hydroxybutyrate. Within the D327C variant C327 forms a disulphide bridge with C80, whereby providing space and avoiding electrostatic repulsion for optimal docking of malate forming two salt bridges with R326 and R329 as well (Figure 4c).

The same statements hold true for the docking arrangements of melibiose (Figure 4c) and isomaltose (Figure 4d), within the D327C variant. Small conformational changes (induced fit), especially of the side chain of R326 allowed the recognition of both sugar moieties, either one by R326 and R329 with optimal hydrogen bond of the

Figure 4: Model structures of the active sites of the AtSST wild type (A) and the AtSST D327C variant (B-D) with the docking interactions of the acyl acceptors L-malate (A), hydroxybutyrate (B), melibiose (C) and isomaltose (D). The colour code is identical to Figure 3.

appropriate hydroxyl group to H416. This stable docking arrangement might explain why these two substrates display higher conversion rates, with regard of the acyl acceptors melibose and ismaltose in comparison to 1-*O*-sinapoyl-β-glucose as acyl acceptor.

Functional diversification of the SST

The enzymes encoded by the gene cluster on chromosome II exhibited partly overlapping specificities. Beside the SST, the At2g22980 gene product and the AtSMT exhibited minor SST activities [20,24]. In general, this points to a common principle: functional diversification is neo-functionalization or sub-functionalization with redundant activities. For the SST, a specification took place. Our studies clearly demonstrated that the loss of the disulphide bridge led to a highly specific enzyme only accepting 1-*O*-sinapoyl-β-glucose as acyl acceptor. The disulphide bond S1 between the C80 and the C327 is highly conserved in serine carboxypeptidases (Figure 5).

To prove that this was not only an evolutionary accident in *A. thaliana,* but a real specification of the SST protein, the genome of the related *A. lyrata*, which diverged from a common ancestor about 10 million years ago [39-41] was studied. This resulted in the identification of 45 SCPL genes in *A. lyrata,* whereas 12 cluster with functional proven acyltransferases (Figure 6). These putative acyltransferases all showed the hallmark of SCPL-acyltransferase: the pentapeptide motif Gly-Asp-Ser*-Tyr-Ser around the catalytic serine (*) [26]. The phylogenetic analysis revealed three SCPL genes that have no homologous gene in *A. thaliana,* leading to their classification as *A. lyrata*-specific genes. On the other hand, there are 9 genes that have no equivalent ones in *A. thaliana*.

The pairwise sequence comparison of the genomic SST region showed a similar exon-intron pattern, both have 15 exons and 14 introns (Figure 7). Interestingly, the size of the introns 4 and 9
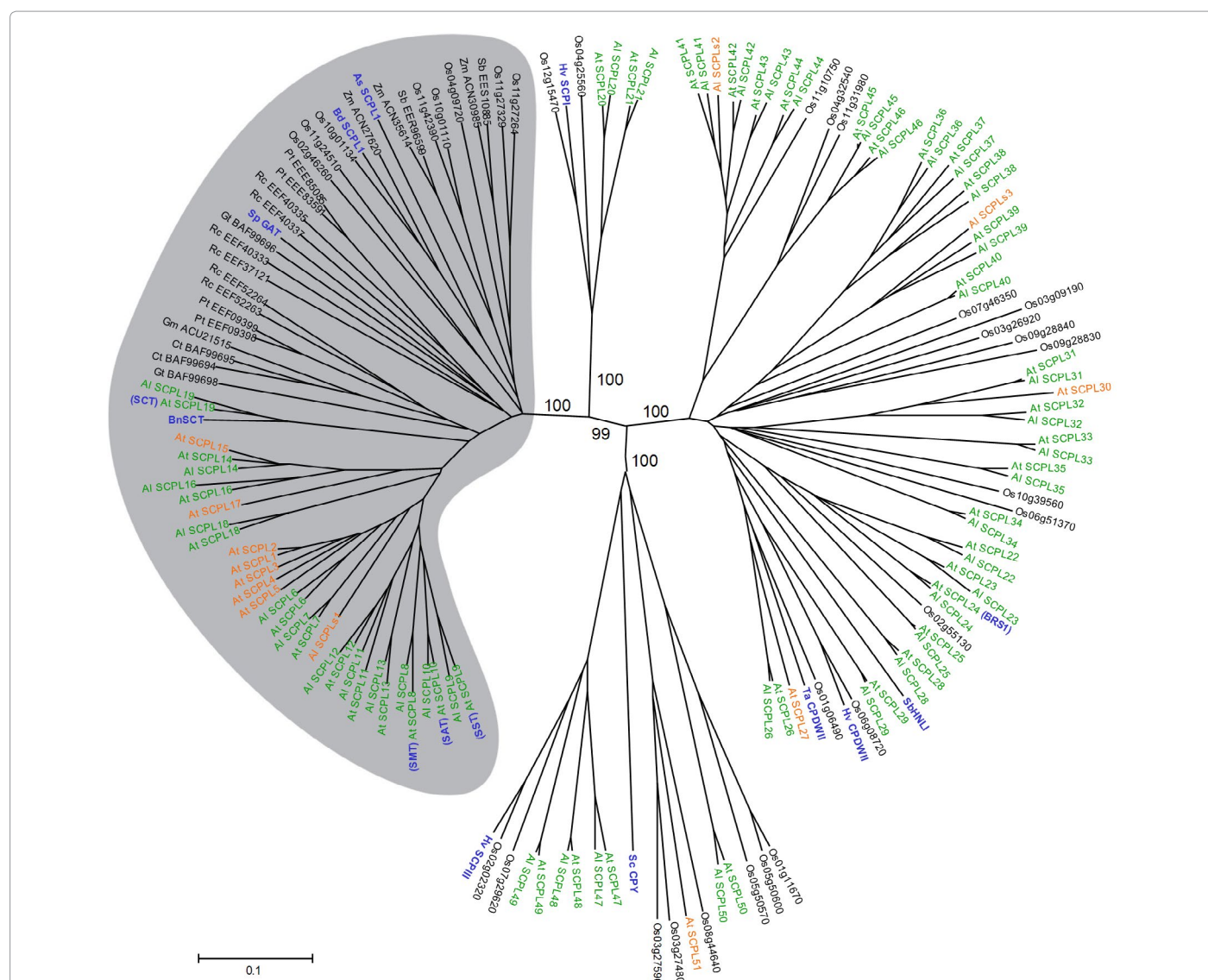

Figure 6: Neighbour-joining phylogenetic analysis of SCPL proteins from *A. thaliana* and *A. lyrata* (green and orange). To obtain more distinguished branches, other SCPL protein sequences were included (black). Bootstrap values are presented for key branches. The scale bar indicates 0.1 substitutions per site. The highlighted clade shows SCPL-acyltransferases. Blue marked are functionally characterized enzymes. Green marked enzymes are present in *A. thaliana* and *A. lyrata*, whereas orange marked enzymes have no homolog in *A. thaliana* or *A. lyrata*, respectively. Protein sequence accession numbers, sequence alignment, and further details of the sequences used for this phylogenetic analysis are provided in the supplemental data.
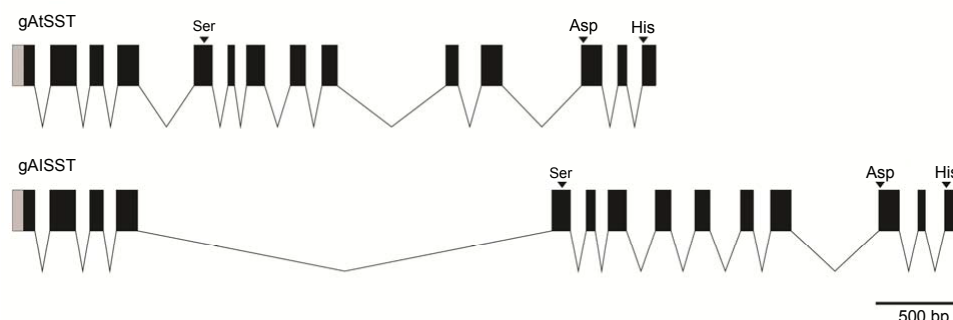
Figure 7: Exon-intron pattern of genomic AtSST and AlSST. Exons are coloured in black boxes, signal peptides in grey. Arrows indicate the positions of the triplets coding for the catalytic triade. Exon-intron pattern are very similar, except the size of the introns 4 and 9, leading to a similarity of only 45% of the genomic sequences compared to 91% of the cDNA sequences. Graphic was made on http://wormweb.org/exonintron.

dramatically differ in size, resulting in an overall similarity of only 45% and a 2700 bps longer AlSST intron region. In contrast, the two SST cDNAs exhibit an identity of 91%. Furthermore, the important amino acid change from cysteine to aspartate, resulting in disulphide bridge loss in the AtSST is also present in the AlSST sequence. This suggests the mutation of the cysteine occurred at least 10 myo years ago. To verify this hypothesis, the ratio of the number of non-synonymous substitutions per non-synonymous site (Ka) to the number of synonymous substitutions per synonymous site (Ks) were determined. This is an effective method to detect the selection on a gene or a gene region [42-44]. If the ratio is significant less than 1, purifying selection is inferred, while positive selection is evoked if the ratio is significant greater than 1. An estimated ratio close to 1 indicates the presence of neutral evolution. We found a Ka<Ks (0.42; p-value (Fisher)=4.35e-10; GY [30] indicating no modification/evolution for SST (purifying selection), as well as its fixation within the two genomes. It is worthy to note that the altered substrate selectivity represents a loss-of-function, as well as an accomplished neo-functionalization.

References

1. Haldane JBS (1932) The causes of evolution. Longmans and Green, London, UK.

2. Muller HJ (1935) The origination of chromatin deficiencies as minute deletions subject to insertion elsewhere. Genetica 17: 237-252.

3. Ohno S (1970) Evolution by gene duplication. Springer-Verlag, New York, USA.

4. Arguello JR, Fan C, Wang W, Long M (2007) Origination of chimeric genes through DNA-level recombination. Genome Dyn 3: 131-146.

5. Brosius J (1991) Retroposons--Seeds of evolution. Science 251: 753.

6. Dunning Hotopp JC, Clark ME, Oliveira DC, Foster JM, Fischer P, et al. (2007) Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science 317: 1753-1756.

7. Mower JP, Stefanović S, Young GJ, Palmer JD (2004) Plant genetics: Gene transfer from parasitic to host plants. Nature 432: 165-166.

8. Begun DJ, Lindfors HA, Kern AD, Jones CD (2007) Evidence for de novo evolution of testis-expressed genes in the *Drosophila yakuba/Drosophila erecta* clade. Genetics 176: 1131-1137.

9. Clark AG (1994) Invasion and maintenance of a gene duplication. Proc Natl Acad Sci U S A 91: 2950-2954.

10. Walsh JB (1995) How often do duplicated genes evolve new functions? Genetics 139: 421-428.

11. Force A, Lynch M, Pickett FB, Amores A, Yan YL, et al. (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151: 1531-1545.

12. Fraser CM, Rider LW, Chapple C (2005) An expression and bioinformatics analysis of the *Arabidopsis* serine carboxypeptidase-like gene family. Plant Physiol 138: 1136-1148.

13. Steffens JC (2000) Acyltransferases in protease's clothing. Plant Cell 12: 1253-1256.

14. Milkowski C, Strack D (2004) Serine carboxypeptidase-like acyltransferases. Phytochemistry 65: 517-524.

15. Strack D, Mock HP (1993) Hydroxycinnamic acids and lignins. In: Methods in Plant Biochemistry, Academic Press, USA 45-97.

16. D'Auria JC (2006) Acyltransferases in plants: a good time to be BAHD. Curr Opin Plant Biol 9: 331-340.

17. Tkotz N, Strack D (1980) Enzymatic synthesis of sinapoyl-L-malate from 1-sinapoylglucose and L-malate by a protein preparation from *Raphanus sativus* cotyledons. Naturforsch 35: 835-837.

18. Lehfeldt C, Shirley AM, Meyer K, Ruegger MO, Cusumano JC, et al. (2000) Cloning of the *SNG1* gene of *Arabidopsis* reveals a role for a serine carboxypeptidase-like protein as an acyltransferase in secondary metabolism. Plant Cell 12: 1295-1306.

19. Lorenzen M, Racicot V, Strack D, Chapple C (1996) Sinapic acid ester metabolism in wild type and a sinapoylglucose-accumulating mutant of arabidopsis. Plant Physiol 112: 1625-1630.

20. Fraser CM, Thompson MG, Shirley AM, Ralph J, Schoenherr JA, et al. (2007) Related *Arabidopsis* serine carboxypeptidase-like sinapoylglucose acyltransferases display distinct but overlapping substrate specificities. Plant Physiol 144: 1986-1999.

21. Strack D, Knogge W, Dahlbender B (1983) Enzymatic synthesis of sinapin from 1-O-sinapoyl-glucose and choline by a cell-free system from developing seeds of red radish (*Raphanus sativus* L var sativus). Z Naturforsch 38: 21-27.

22. Shirley AM, McMichael CM, Chapple C (2001) The *sng2* mutant of *Arabidopsis* is defective in the gene encoding the serine carboxypeptidase-like protein sinapoylglucose:choline sinapoyltransferase. Plant J 28: 83-94.

23. Lee S, Kaminaga Y, Cooper B, Pichersky E, Dudareva N, et al. (2012) Benzoylation and sinapoylation of glucosinolate R-groups in *Arabidopsis.* Plant J 72: 411-422.

24. Stehle F, Brandt W, Schmidt J, Milkowski C, Strack D (2008) Activities of *Arabidopsis* sinapoylglucose:malate sinapoyltransferase shed light on functional diversification of serine carboxypeptidase-like acyltransferases. Phytochemistry 69: 1826-1831.

25. Stehle F, Stubbs MT, Strack D, Milkowski C (2008) Heterologous expression of a serine carboxypeptidase-like acyltransferase and characterization of the kinetic mechanism. FEBS J 275: 775-787.

26. Stehle F, Brandt W, Stubbs MT, Milkowski C, Strack D (2009) Sinapoyltransferases in the light of molecular evolution. Phytochemistry 70: 1652-1662.

27. Mugford ST, Qi X, Bakht S, Hill L, Wegel E, et al. (2009) A serine carboxypeptidase-like acyltransferase is required for synthesis of antimicrobial compounds and disease resistance in oats. Plant Cell 21: 2473-2484.

28. Stehle F, Brandt W, Milkowski C, Strack D (2006) Structure determinants and substrate recognition of serine carboxypeptidase-like acyltransferases from plant secondary metabolism. FEBS Lett 580: 6366-6374.

29. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596-1599.

30. Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11: 725-736.

31. Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, et al. (2006) KaKs_Calculator: Calculating Ka and Ks through model selection and model averaging. Genomics Proteomics Bioinformatics 4: 259-263.

32. Krieger E, Joo K, Lee J, Lee J, Raman S, et al. (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. Proteins 77: 114-122.

33. Liao DI, Remington SJ (1990) Structure of wheat serine carboxypeptidase II at 3.5-A resolution. A new class of serine proteinase. J Biol Chem 265: 6528-6531.

34. Sippl MJ (1990) Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. J Mol Biol 213: 859-883.

35. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) Procheck-A program to check the stereochemical quality of protein structures. J Appl Crystallogr 26: 283-291.

36. Hartshorn MJ, Verdonk ML, Chessari G, Brewerton SC, Mooij WT, et al. (2007) Diverse, high-quality test set for the validation of protein-ligand docking performance. J Med Chem 50: 726-741.

37. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD (2003) Improved protein-ligand docking using GOLD. Proteins 52: 609-623.

38. Hamilton R, Watanabe CK, de Boer HA (1987) Compilation and comparison of the sequence context around the AUG startcodons in *Saccharomyces cerevisiae* mRNAs. Nucleic Acids Res 15: 3581-3593.

39. Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, et al. (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science 327: 92-94.

40. Wright SI, Lauga B, Charlesworth D (2002) Rates and patterns of molecular evolution in inbred and outbred *Arabidopsis*. Mol Biol Evol 19: 1407-1420.

41. Beilstein MA, Nagalingum NS, Clements MD, Manchester SR, Mathews S (2010) Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. Proc Natl Acad Sci U S A 107: 18724-18728.

42. Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. Trends Ecol Evol 15: 496-503.

43. Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A 89: 10915-10919.

44. Henikoff S, Henikoff JG (1993) Performance evaluation of amino acid substitution matrices. Proteins 17: 49-61.