

Towards engineering glucosinolates into non-cruciferous plants

Fernando Geu-Flores · Carl Erik Olsen ·
Barbara Ann Halkier

Received: 22 July 2008 / Accepted: 11 September 2008 / Published online: 2 October 2008
© Springer-Verlag 2008

Abstract Glucosinolates are amino acid-derived secondary metabolites present in cruciferous plants. Glucosinolates and their hydrolysis products are involved in defence against insects and pathogens, but are also known for their characteristic flavor and their cancer-preventive and anti-bacterial properties. This wide range of bioactivities has prompted a desire to engineer glucosinolates into non-cruciferous plants. We report the one-step transfer of the last three steps of the benzylglucosinolate pathway (comprising the C–S lyase, glycosyltransferase and sulfotransferase) from *Arabidopsis* to tobacco. This was achieved using an expression construct consisting of a single 2A polycistronic open reading frame, which allowed the expression of the three coding-sequences from a single promoter. When compared to wildtype plants, transgenic tobacco lines showed increased ability to convert the intermediate phenyl-

acetothiohydroxamate to benzylglucosinolate upon in vivo feeding. Enzymatic assays using plant extracts demonstrated that the individual activities required for this conversion were enhanced in the transgenic plants. The relatively high conversion by wildtype plants in feeding assays supports the hypothesis that the last part of the glucosinolate pathway was recruited from existing detoxification reactions. Immunoblots confirmed that individual proteins were being successfully produced from the 2A polycistronic open reading frame, albeit fusion proteins could also be detected. In summary, we transferred the last three steps of the benzylglucosinolate pathway to tobacco as a first step towards engineering glucosinolates into non-cruciferous plants.

Keywords 2A-ORFs · Co-expression · Engineering · Glucosinolates · Pathway · Tobacco

Abbreviations

ORF	Open reading frame
BGLS	Benzylglucosinolate
dBGLS	Desulfobenzylglucosinolate
PATH	Phenylacetothiohydroxamate
ST	Sulfotransferase
GT	Glycosyltransferase

Introduction

Glucosinolates are amino acid-derived secondary metabolites present in the Brassicales order, which includes the agriculturally important cruciferous vegetables (e.g. cabbage, broccoli, and oilseed rape) and the model plant *Arabidopsis thaliana* (Fahey et al. 2001). Upon tissue damage, glucosinolates are hydrolyzed by myrosinases to produce a

Electronic supplementary material The online version of this article (doi:10.1007/s00425-008-0825-y) contains supplementary material, which is available to authorized users.

F. Geu-Flores · B. A. Halkier (✉)
Plant Biochemistry Laboratory,
VKR Research Centre Pro-Active Plants,
Department of Plant Biology and Biotechnology,
Faculty of Life Sciences, University of Copenhagen,
40 Thorvaldsensvej, 1871 Frederiksberg C, Denmark
e-mail: bah@life.ku.dk

F. Geu-Flores
e-mail: feg@life.ku.dk

C. E. Olsen
Department of Natural Sciences, Faculty of Life Sciences,
University of Copenhagen, 40 Thorvaldsensvej,
1871 Frederiksberg C, Denmark
e-mail: ceo@life.ku.dk

series of bioactive compounds, mainly isothiocyanates and nitriles (Bones and Rossiter 2006). Glucosinolates and their hydrolysis products play a defensive role against generalist insects (Wittstock et al. 2003) and pathogens (Brader et al. 2006). In addition, they are known for their characteristic pungent flavor (e.g. mustard and horse radish) and most importantly for their cancer-preventive (Bianchini and Vainio 2004; Fimognari and Hrelia 2007) and antibacterial properties (Fahey et al. 2002).

In agriculture, the use of crucifers of the *Brassica* genus as break crops (i.e. crops that break the life cycle of pathogens and pests in rotation systems) is referred to as biofumigation (Kirkegaard et al. 2000). The suppression of soil-borne pathogens and pests by biofumigation is attributed to the release of glucosinolate-derived isothiocyanates into the soil (Kirkegaard et al. 2000; Smith et al. 2004). This suggests that the transfer of glucosinolates to non-cruciferous crop plants by genetic engineering can increase pest resistance and reduce pesticide use as an important step towards sustainable agriculture.

Glucosinolates are divided into three classes depending on the nature of the precursor amino acids (Halkier and Gershenzon 2006). The aliphatic are derived from chain-elongated derivatives of methionine, the phenolic from tyrosine and phenylalanine, and the indolic from tryptophan. Each class has a corresponding core structure pathway that consists of at least five enzymatic steps that catalyze the conversion of the amino acid to the corresponding glucosinolate. These include two monooxygenation reactions by cytochromes P450 from the CYP79 and CYP83 families, which, respectively, catalyze the conversion of precursor amino acids to the corresponding oximes, followed by the conversion of oximes to reactive compounds (Halkier and Gershenzon 2006). In a sulfur donation step, which may be either non-enzymatic or involve a glutathione-*S*-transferase-type of enzyme (Hansen et al. 2001), *S*-alkylthiohydroxamates are formed. The last three enzymatic steps include a *C*-*S* lyase, which converts *S*-alkylthiohydroxamates to thiohydroxamates (Mikkelsen et al. 2004), a glucosyltransferase, which glucosylates thiohydroxamates to yield desulfoglucosinolates (Grubb et al. 2004), and a sulfotransferase, which adds a sulfate moiety to produce glucosinolates (Piotrowski et al. 2004; Klein et al. 2006).

Metabolic engineering of multiple gene pathways into plants may lead to transgene silencing if a single promoter is used repeatedly (Morant et al. 2007). A method to achieve co-expression of different transgenes while avoiding the repeated use of promoters is the 2A system, which allows the production of several proteins from a single promoter (de Felipe et al. 2006). In this co-expression system, multiple protein-coding sequences are fused in frame with intervening virus-derived 2A-coding sequences into a

single open reading frame (ORF). During translation of the long transcript, the approximately 19 amino acid-long 2A sequences mediate a ‘ribosomal skip’ at their C-terminus, releasing each protein as a discrete translation product C-terminally fused to 2A (Donnelly et al. 2001; Atkins et al. 2007).

In our engineering project, we chose benzylglucosinolate (BGLS) because, unlike many aliphatic glucosinolates, it is directly derived from a protein amino acid (phenylalanine). Furthermore, in contrast to engineering indolic glucosinolates, engineering BGLS is not expected to interfere with auxin homeostasis (Halkier and Gershenzon 2006). In this manuscript, we report the transfer of the three last steps of the BGLS pathway into tobacco by using a single 2A polycistronic ORF, both in a tagged and untagged version. The successful engineering is evidenced by the high production of BGLS upon *in vivo* feeding of an intermediate and by measurement of individual enzyme activities. The efficiency of the 2A system is discussed in the light of these assays and of immunoblots that show the protein products derived from the tagged 2A polycistronic ORF.

Materials and methods

Chemical synthesis of phenylacetothiohydroxamate

The synthesis of phenylacetothiohydroxamate (PATH) was achieved by the method of Ettlinger and Lundeen (1957), also presented by Walter and Schaumann (1971). The product was crystallized as the organic acid in benzene/hexane, and the crystals were analyzed by ¹H-, ¹³C-, ¹³C DEPT-, and 2D (¹H and ¹³C) COSY-NMR (data not shown). For a cautionary note on the synthesis of PATH by the method of Doszczak and Rachon (2002), see Supplementary Material.

Generation of expression constructs

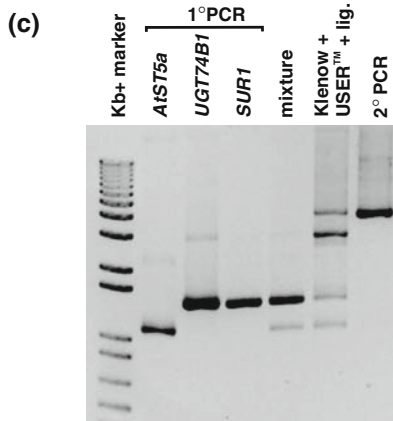
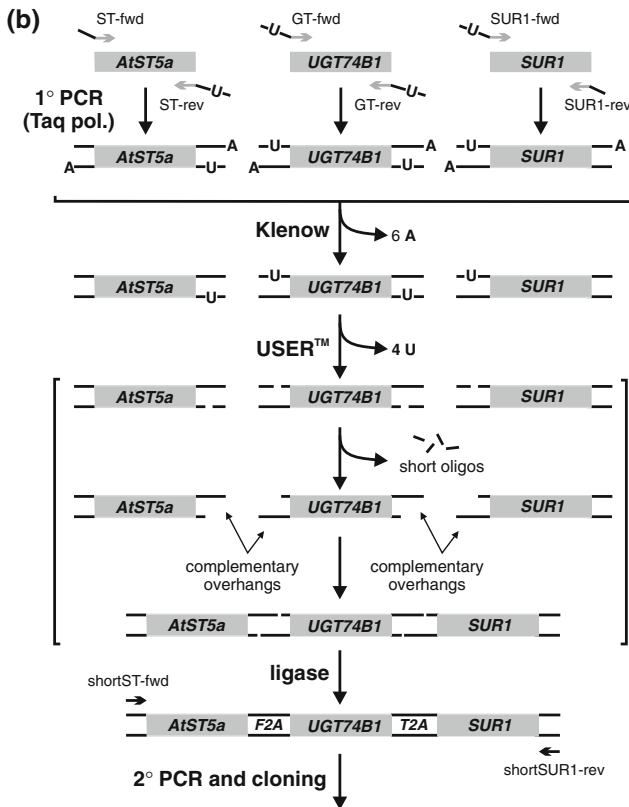
The deoxyuridine excision-based assembly procedure is schematized in Fig. 1a and the sequences of the oligonucleotide primers used for PCR amplification are detailed in Supplementary Table 1.

The coding sequences of *AtSOT16* (At1g74100, 1,014 bp without the stop codon), *UGT74B1* (At1g24100, 1,380 bp without the stop codon) and *SUR1* (At2g20610, 1,389 bp including the stop codon) were independently amplified by PCR from RAFL clones (RAFL05-13-F01 for *AtSOT16* and RAFL04-19-M06 for *UGT74B1*, RIKEN BioResource Center) or from a published clone (for *SUR1*; Mikkelsen et al. 2004) using HotMaster™ Taq DNA Polymerase (Eppendorf). For the assembly of ORF2 (Fig. 1a), the primer pairs used were ST-fwd/ST-rev, GT-fwd/GT-rev, and SUR1-fwd/SUR1-rev, respectively. The F2A

(a) ORF2



ORF2nat



coding sequence was included as two separate but overlapping fragments in primers ST-rev and GT-fwd, and the T2A coding sequence was similarly included in primers GT-rev and SUR1-fwd. Furthermore, in these primers, a single deoxyuridine residue (instead of a deoxythymidine one) was located 7–9 residues downstream of a 5' terminal

Fig. 1 a–c Molecular strategy. **a** Structure of the 2A polycistronic open reading frames ORF2 and ORF2nat. The coding sequences of genes *AtST5a*, *UGT74B1*, and *SUR1* were linked by F2A- and T2A-coding sequences. In ORF2, the encoded *AtST5a* protein was N-terminally His-tagged, and *UGT74B1* was N-terminally Strep[®]-tagged. **b** Scheme of the assembly ORF2. After PCR amplification using deoxyuridine-containing primers and Taq polymerase, the PCR fragments were polished with Klenow enzyme and treated with USER[™] enzyme mix to produce 7–9 nucleotide 3' overhangs. The two pairs of complementary overhangs annealed to each other and yielded a linear hybridization product that was ligated to give the full-length product. **c** Steps in the assembly of ORF2 visualized by agarose gel electrophoresis. Lane 1 Kb + marker; lanes 2–4 primary PCR products (~1,100 bp for *AtST5a*, ~1,400 bp for *UGT74B1*, and ~1,400 bp for *SUR1*); lane 5 mixed primary PCR products; lane 6 ligation products, after Klenow and USER[™] treatment. Apart from the full-length product (uppermost band ~4 kb), the ligation product of *AtST5a* and *UGT74B1* (second uppermost band) and the fusion product of *UGT74B1* and *SUR1* (third uppermost band) can be clearly seen; lane 7 secondary PCR product

deoxyadenosine. Equimolar amounts of PCR products were mixed and potential A tails were removed using Klenow fragment (New England Biolabs). After heat inactivation of the Klenow enzyme, the mixture was treated with the deoxyuridine-excising USER[™] enzyme mix (New England Biolabs) for 30 min at 37°C at a final concentration of 0.1 U/μL. Removal of the primer-derived deoxyuridines in the PCR products generated short oligonucleotides that dissociated to create relatively long overhangs. The 7-nt overhang created by partial removal of primer ST-rev was designed to complement the 7-nt overhang created by partial removal of primer GT-fwd, and a similar design had been made for the 9-nt overhangs created by partial removal of primers GT-rev and SUR1-fwd. Ligation was performed overnight (Fig. 1b). A second PCR amplification was performed using primer pair shortST-fwd/shortSUR1-rev (shorter versions of ST-fwd and SUR1-rev) and Phusion[™] High Fidelity DNA Polymerase (Finnzymes). The gel-purified full-length PCR product was cut with *EcoRI* and *SacI*, ligated to similarly treated pGEM4Z, and transformed into *E. coli*. Sequencing revealed a pGEM4Z + ORF2 clone with only two silent PCR-introduced errors. For the assembly of ORF2nat, the exact same procedure was followed, except that primer STnat-fwd replaced primer ST-fwd, and that primer GTnat-fwd replaced primer GT-fwd. All sequenced pGEM4Z + ORF2nat clones contained at least one non-silent PCR-introduced error. An error-free version of pGEM4Z + ORF2nat was assembled from three different individual clones by restriction digestion/ligation using unique restriction sites.

To obtain the plant expression vectors, ORF2 and ORF2nat were first subcloned from pGEM4Z into pRT-101 using *EcoRI* and *KpnI*. Fragments including the 35S promoter and terminator were subcloned into pCAMBIA2300 using *PstI*.

In vitro transcription/translation

ORF2 and ORF2nat were transcribed from the pGEM4Z construct versions using SP6 polymerase (Promega) and translated using the rabbit reticulocyte lysate system (Promega) in the presence of L-[U-¹⁴C]-leucine. Both processes were done according to the manufacturer's instructions. A negative control was prepared in parallel starting with water instead of plasmid solution. The translation reactions (15 µL each) were run on a 12% polyacrylamide gel. The gel was stained with Coomassie Brilliant Blue and analyzed by phosphorimaging on a Storm[®] 860 Phosphorimager[®] (Molecular Dynamics). The intensities of the radioactive protein bands were quantified with the ImageQuant 5.0 software. To calculate the molar ratios between the different protein species, the intensities of the bands were normalized based on the number of leucine residues per species (His-AtST5a-F2A: 36 leu/molecule; Strep-UGT74B1-T2A: 48 leu/molecule; SUR1: 46 leu/molecule).

Plant transformation

ORF2 and ORF2nat were transformed independently into tobacco (*Nicotiana tabacum* L. cv. Xanthi; gift from Lilli Sander, Biotechnology Group, University of Aarhus) by *Agrobacterium*-mediated transformation of leaf explants (Horsch et al. 1985) using the pCAMBIA2300 construct versions. Regenerants selected on kanamycin were tested for the presence of ORF2 or ORF2nat by PCR on genomic DNA using primers ATGGAATCAAAGACAACCCAA and GGTCTCGTACCTAAGGAACA (the expected fragment of ~800 bp was part of *AtST5a*). A control PCR was carried out using primers GGAGTCTTTCAGCATGGAGCAA and ATGTCGCAAGGACGTAAGCCCA to rule out *Agrobacterium* contamination (the expected fragment of ~450 bp was part of the *VirD1* gene).

Confirmed transformants were periodically subcultivated in vitro in MS media supplemented with 3% sucrose, 100 µg/L kanamycin and 300 µg/L cefotaxime. Selected transformants were grown under greenhouse conditions to obtain T2 seeds. Sixteen T2 seeds of each of three selected lines were grown in the greenhouse to generate T3 seeds. Approximately, 50 T3 seeds of each individual T2 plant were germinated on 1/2 MS media with 100 µg/L kanamycin to assess heterozygosity of the mother plants.

Southern blot

DNA was extracted from ~1 g of leaf material by a modified CTAB protocol (Doyle and Doyle 1987) and treated with Ribonuclease A (Sigma). A measure of 25 µg of DNA was digested overnight with 10 U of HindIII and used for Southern-blot analysis following the manufacturer's protocol

for detection with a DIG-labeled probe (Roche, www.rocke-applied-science.com). All the materials and reagents used for the blotting were purchased from Roche, including the positively charged nylon membrane and the DIG Easy Hyb solution. The probe was made by PCR amplification of a ~650-bp fragment from the *nptIII* (*kan^R*) gene in pCAMBIA2300 using primers GGCTATTCGGCTATGACTGG and CAGCAATATCACGGGTAGCCA in the presence of DIG-labeled dUTP at a ratio of 1:1 (labeled/unlabeled dUTP). Detection was achieved by chemiluminescence using anti-digoxigenin-AP conjugate as antibody, CSPD ready-to-use as substrate, and Lumi-Film Chemiluminescent Detection Film as light-sensitive film.

Feeding assays

In vitro-grown primary transformants were used for the feeding assays. Six or seven leaves per line were fed PATH, and a single leaf per line was fed water by immersing the cut end of a petiole in 30 µL of either 1 mM NaPATH_(aq) or water. The cut leaves were of different sizes and weighed between 20 and 270 mg. Immediately after the initial uptake of liquid, 30 µL of water were fed to ensure complete uptake of PATH. After uptake of these last 30 µL, each leaf was left overnight with 300 µL of water at room temperature and under constant light conditions. Twenty-four hours after commencement of the feeding, the leaves were lyophilized and analyzed for glucosinolates as described by Hansen et al. (2007).

Enzymatic assays

Approximately 1 g of leaf material from in vitro-grown homozygous T3 plants was homogenized in the presence of 4 mL of extraction buffer [250 mM sucrose, 100 mM Tris-HCl, 50 mM NaCl, 5 mM DTT, 2 mM EDTA, 5% PVPP, and 1 × complete protease inhibitor cocktail (EDTA-free, Roche)]. The homogenate was filtered through a nylon mesh and the filtrate was centrifuged at 20,000g and 4°C for 20 min. The supernatant ('soluble protein extract') was collected and assayed for enzyme activity.

The glucosyltransferase assay (GT assay) was performed using 50 µg of soluble protein in a total volume of 100 µL. The reaction mixtures contained 50 mM MOPS pH 6.0, 5 mM Mg₂SO₄, 0.065% β-mercaptoethanol, 1.0 mM UDP-Glc and 1.0 mM PATH. The reactions were started by addition of the protein. After incubation at 30°C, they were stopped by addition of 2.4 µL of 100% trichloroacetic acid (TCA). The products were extracted with 200 µL ethyl acetate, and 100 µL of the extract were evaporated, redissolved in water and analyzed by HPLC (Hansen et al. 2007). For quantification, desulfobenzylglucosinolate (dBGLS) standard solutions were prepared

by enzymatic desulfation of benzylglucosinolate (BGLS, Calbiochem).

For the sulfotransferase assay (ST assay), dBGLS (to be used as substrate) was prepared and quantified as described for the GT assay. The ST assay was also carried out using 50 µg of soluble protein and in a total volume of 100 µL. The reaction mixtures contained 50 mM Bicine pH 9.0, 0.2 mM PAPS (Calbiochem), and 0.2 mM dBGLS. The reactions were started by addition of the protein. After incubation at 37°C, they were stopped by addition of 200 µL of methanol. The entire mixture was run through 45-µL DEAE-Sephadex columns. The glucosinolate analysis was continued as described previously (Hansen et al. 2007). For quantification, BGLS standards were subjected to the same analytical procedure.

In both assays, a selected transformant, a WT plant, and a control (with water instead of extract) were compared at four different time points (0, 45, 90 and 135 min) with three replicates each. For $t = 0$ min, TCA (GT assays) or methanol (ST assays) was added before the soluble protein extract.

Western blots

Soluble protein extracts were prepared as described for the enzymatic assays. For Western blots using an anti-His antibody, small-scale purification of His-tagged proteins from the extracts was performed using His SpinTrap columns (GE Healthcare). The composition of the soluble protein extracts was adjusted to reach 500 mM NaCl, 20 mM imidazole (for a more specific binding), and 3 mM MgCl (to counteract the chelating effect of the EDTA present in the extraction buffer). The purification was then carried out following the manufacturer's instructions. The binding buffer used consisted of 100 mM Tris-HCl pH 7.5, 500 mM NaCl, and 20 mM imidazole, whereas the elution buffer consisted of 100 mM Tris-HCl pH 7.5, 500 mM NaCl, and 500 mM imidazole. Nine microgram of protein from the first eluate fraction were run on a 12% polyacrylamide gel. The gel was blotted on a Protran[®] nitrocellulose membrane (Whatman) using a Criterion[™] Blotter (BioRad). After blocking with 5% skim milk (in PBS-Tween 20), the membrane was incubated with monoclonal anti-polyhistidine-peroxidase (Sigma, diluted in PBS-Tween 20). The detection was carried out using SuperSignal West Dura Extended Duration substrate (Pierce) and an AutoChem[™] System (UVP) for imaging.

For Western blots using an anti-Strep[®]-tag antibody, small-scale purification of Strep-tagged proteins from the extracts was performed using Strep-tactin[®] magnetic beads (Qiagen). The composition of the soluble protein extracts was adjusted to reach 300 mM NaCl, and 100 µg/mL avidin (from egg white, Invitrogen). Avidin binds to biotin and

biotinylated proteins, and prevents them from binding to Strep-tactin[®]. The purification was carried out following the manufacturer's instructions. A measure of 1 µg of protein from the first eluate fraction was run on a 12% polyacrylamide gel. The gel was blotted on a Protran[®] nitrocellulose membrane (Whatman) using a Criterion[™] Blotter (BioRad). After sequential blocking with 3% BSA and 2 µg/mL avidin (both in PBS-Tween 20), the membrane was incubated with QIAexpress[®] Strep-tag[®] mouse monoclonal antibody (Qiagen) and with a dilution of stabilized goat anti-mouse HRP-conjugated antibody (Pierce) in PBS-Tween 20. The detection was carried out using SuperSignal West Dura Extended Duration Substrate (Pierce) and an AutoChem[™] System (UVP) for imaging.

Results

Generation of expression constructs

The three last steps in the biosynthesis of BGLS include SUR1 as C-S lyase (Mikkelsen et al. 2004), UGT74B1 as glucosyltransferase (Grubb et al. 2004), and AtST5a as sulfotransferase (Piotrowski et al. 2004; Klein et al. 2006). We assembled a tagged and an untagged 2A polycistronic ORF, respectively called ORF2 and ORF2nat, each carrying the coding sequences of the three mentioned enzymes along with intervening 2A-coding sequences. ORF2 coded for a His-tag at the N-terminus of AtSOT16, and for a Strep[®]-tag at the N-terminus of UGT74B1 (Fig. 1a).

For the assembly procedure, the individual coding sequences were PCR-amplified using deoxyuridine-containing primers. This allowed creation of long complementary overhangs upon treatment with the deoxyuridine-excising USER[™] enzyme mix. The ORFs were obtained by ligation of the treated PCR products followed by a second PCR amplification and conventional cloning (Fig. 1b, c).

In vitro transcription/translation

In vitro transcription/translation of ORF2 and ORF2nat yielded three polypeptides whose molecular weight corresponded to the desired individual proteins (His-)AtST5a-F2A, (Strep-)UGT74B1-T2A, and SUR1 (Fig. 2). A polypeptide of higher molecular weight corresponding to the fusion of the two upstream proteins, (His-)AtST5a-F2A and (Strep-)UGT74B1-T2A was observed. Quantification of the different polypeptides derived from ORF2 showed that the protein fusion accounted for ~1% of the molar polypeptide amount and that the three individual proteins were not produced in equimolar amounts, but in an estimated ratio of 12:4:3 [His-AtST5a-F2A/Strep-UGT74B1-T2A/SUR1].

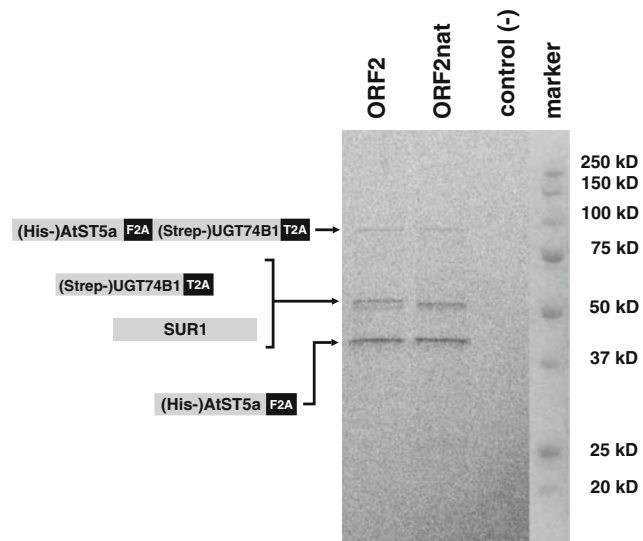


Fig. 2 Analysis of in vitro transcription/translation products of ORF2 and ORF2nat by SDS-PAGE. Radiolabeled proteins were visualized by Phosphorimaging[®]. *Lane 1*: ORF2. The three individual proteins His-AtST5a-F2A (~42 kDa), Strep-UGT74B1-T2A (~54 kDa), and SUR1 (~51 kDa), and a band corresponding to the fusion of His-AtST5a-F2A and Strep-UGT74B1-T2A (~96 kDa) have been marked with arrows. *Lane 2* ORF2nat. While AtST5a-F2A (~41 kDa) was clearly seen, UGT74B1-T2A (~53 kDa) and SUR1 (~51 kDa) migrated as a double band. A band corresponding to the fusion between AtST5a-F2A and UGT74B1-T2A (~95 kDa) was also seen

Plant transformation

ORF2 and ORF2nat were introduced into tobacco plants by *Agrobacterium*-mediated transformation of leaf explants. Regenerants selected on kanamycin were analyzed by PCR on genomic DNA to confirm the presence of the ORFs. This yielded 13 transgenic lines for ORF2 and 9 for ORF2nat. The lines were subjected to further analysis by Southern blot. Four ORF2 lines and four ORF2nat lines were confirmed to have a single T-DNA insertion in their genomes (Suppl. Fig. 1). No visible phenotype was observed for these plants when grown in vitro or under greenhouse conditions.

Feeding assays

Transgenic lines having a single T-DNA insertion and wildtype plants were subjected to assays in which PATH (substrate of UGT74B1) was fed to leaves through cut petioles. Twenty-four hours after commencement of the feeding, the leaves were analyzed for BGLS. All eight transgenic lines produced higher average levels of BGLS when compared to wildtype plants. For six of the eight lines, the increased conversion was significant ($P < 0.005$ on a Student's *t* test) (Fig. 3). Wildtype leaves converted

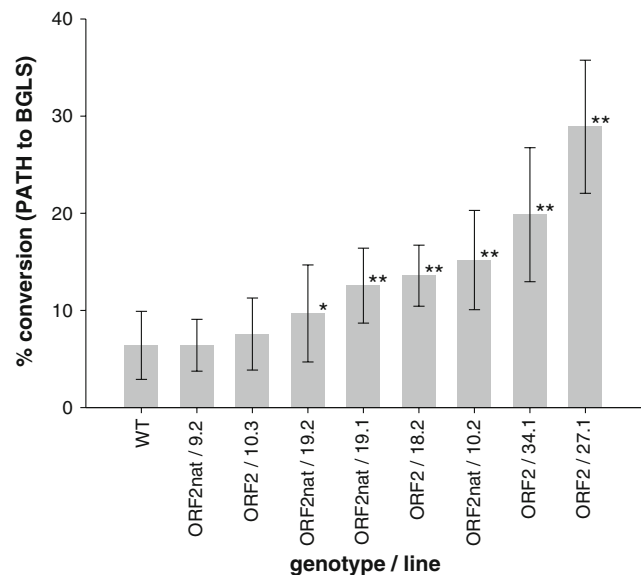


Fig. 3 Analysis of the production of BGLS in transgenic lines upon in vivo feeding. Individual leaves were fed PATH through cut petioles; ~24 h after commencement of the feeding, leaves were lyophilized and analyzed for BGLS. Graph bars show the mean percentage of conversion of 7 leaves of a particular genotype, and error bars represent standard deviation. A Student's *t* test was performed to analyze the significance of the differences between each transgenic line and wildtype plants (WT). Columns marked with asterisk represent lines with $0.001 < P < 0.005$, whereas double asterisks represent $P < 0.001$

around 6% of the administered PATH to BGLS, whereas leaves of the highest-converting transgenic line (ORF2/27.1) converted around 29%. No BGLS was detected in leaves in which water was fed instead of PATH. No correlation was found between leaf size and the level of conversion (data not shown). Shorter incubation times or increased PATH concentrations did not change the relative conversion levels in leaves of ORF2/27.1 in relation to wildtype leaves (data not shown).

Enzymatic assays

Individual glucosyltransferase (GT) and sulfotransferase (ST) assays were performed using soluble protein extracts from leaves of line ORF2/27.1 and from wildtype leaves. In both assays, the substrate to product conversion was linear over the length of the assay for the extracts from both line ORF2/27.1 and wildtype, although for the latter, conversions were close to detection limits (Fig. 4a, b). Control reactions with no protein added showed negligible conversions. For the GT assay, linear regression revealed conversion rates of 0.004 nmol/min for WT plants and 0.116 nmol/min for line ORF2/27.1. For the ST assay, a similar analysis revealed conversion rates of 0.001 nmol/min for WT plants and 0.033 nmol/min for line ORF2/27.1.

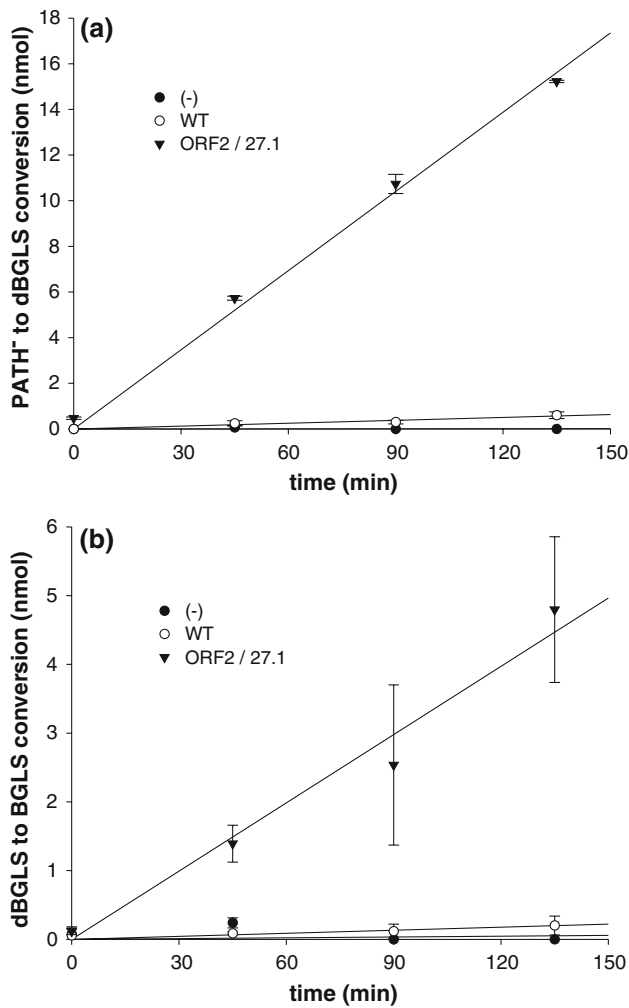


Fig. 4 **a, b** Analysis of glucosyltransferase (GT) and sulfotransferase (ST) enzymatic activity in soluble protein extracts from leaves of line ORF2/27.1 and wildtype leaves (WT). The graphs represent the substrate to product conversion of 50 µg of crude protein at four different time points: 0, 45, 90 and 135 min. As negative control (–), no protein was added to a set of reactions. **a** GT assay. The reaction volume was 100 µL, and the substrate concentration 1.0 mM. Linear regression revealed conversion rates of 0.004 nmol/min for WT and 0.116 nmol/min for line ORF2/27.1. **b** ST assay. The reaction volume was 100 µL and the substrate concentration 0.2 mM. Linear regression revealed conversion rates of 0.001 nmol/min for WT and 0.033 nmol/min for line ORF2/27.1

Western blots

The His-tag at the N-terminus of AtST5a and the Strep®-tag at the N-terminus of UGT74B1 allowed immuno-detection of these proteins in extracts from leaves of line ORF2/27.1. Small-scale affinity purification of the extracts was performed to enrich for either His- or Strep-tagged proteins. His-purified extracts from ORF2/27.1 were subjected to Western blotting using an anti-polyhistidine antibody, while the analogous Strep-purified extracts were subjected to Western blotting using an anti-Strep®-tag antibody. The

His-purified extract of ORF2/27.1 clearly presented two main bands with molecular weight corresponding to the individual protein His-AtST5a-F2A and the fusion of His-AtST5a-F2A and Strep-UGT74B1-T2A (Fig. 5a). A faint third band with a molecular weight corresponding to the triple fusion of His-AtST5a-F2A, Strep-UGT74B1-T2A and SUR1 was observed. For the Strep-purified extract of ORF2/27.1, a clear band corresponding to the individual Strep-UGT74B1-T2A was seen (Fig. 5b). A fainter band, possibly a double one, was also observed. This is likely to correspond to the fusion of His-AtST5a-F2A and Strep-UGT74B1-T2A and/or the fusion of Strep-UGT74B1-T2A and SUR1. An extract of wildtype leaves was subjected to identical procedures and did not present any clear bands in the blots (Fig. 5a, b).

Discussion

The many different bioactivities of glucosinolates (and their degradation products) have made the engineering of these

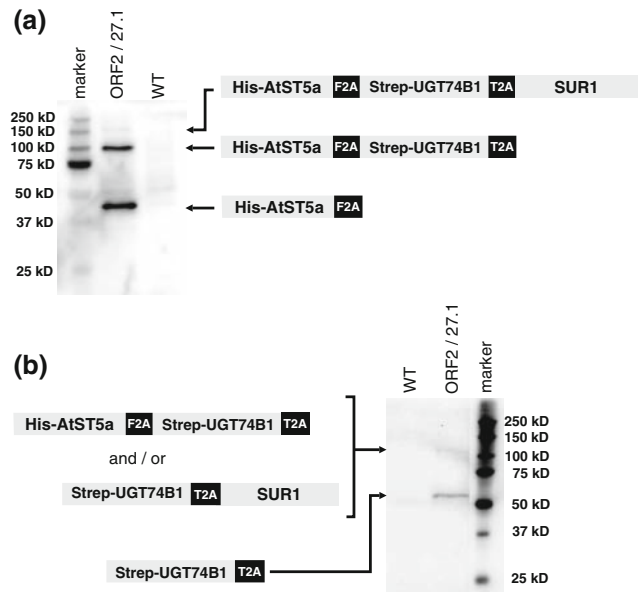


Fig. 5 **a, b** Western-blot analysis of tagged proteins. **a** Immunoblot with an anti-polyhistidine antibody. *Lane 1* molecular weight marker; *lane 2* His-purified extract from leaves of ORF2/27.1. Two main bands with molecular weights corresponding to the individual protein His-AtST5a-F2A and the fusion of His-AtST5a-F2A and Strep-UGT74B1-T2A have been marked with *arrows*. A third, much fainter band corresponding to the triple fusion of His-AtST5a-F2A, Strep-UGT74B1-T2A and SUR1 has also been marked; *lane 3* His-purified extract of wildtype leaves. **b** Immunoblot with an anti-Strep®-tag antibody. *Lane 1* Strep-purified extract of wildtype leaves; *lane 2* Strep-purified extract from leaves of ORF2/27.1. A band corresponding to the Strep-UGT74B1-T2A and a fainter band corresponding to the fusion of His-AtST5a-F2A and Strep-UGT74B1-T2A and/or the fusion of Strep-UGT74B1-T2A and SUR1 have been marked; *lane 3* molecular weight marker

compounds into heterologous host plants desirable. This is biotechnologically challenging, as the pathway involves at least five different enzymatic steps. We describe the transfer of the last three steps in the biosynthesis of BGLS into tobacco plants using an expression construct consisting of a single 2A polycistronic ORF. The rationale behind engineering the last part of the BGLS pathway as the initial step was that the ‘late’ enzymes would not encounter their natural substrates in the transgenic plants, whereas the ‘early’ enzymes could produce pathway intermediates (or byproducts) which may compromise subsequent transformations (Tattersall et al. 2001).

The 2A co-expression system was used as a means to achieve co-expression of the three transgenes from a single promoter. In our design, we considered previous experiments in which several synthetic 2A polycistronic ORFs were analyzed by *in vitro* transcription/translation (Donnelly et al. 2001). In these experiments, polypeptides encoded upstream of a single 2A-coding sequence accumulated at higher levels than the ones encoded downstream. The molar ratios varied from 2:1 to 10:1, depending on the sequence upstream of the 2A-coding sequence, and on the type and batch of translation system used. This prompted us to arrange the different coding sequences in our 2A polycistronic ORFs using the following rule: the further downstream the enzyme in the pathway, the more upstream the coding sequence in the ORF. This ensured no molar excess of any protein in the pathway compared to its immediate downstream protein. Therefore, when compared to the opposite arrangement or to a random one, this particular arrangement had the lowest chances of leading to accumulation of potentially toxic intermediates. This holds true regardless of the individual enzyme kinetics, but provided that the short 2A sequences do not interfere with the activities.

The procedure used for the assembly of ORF2 and ORF2nat involved the use of deoxyuridine-containing primers for primary PCR amplification and a commercial enzyme mix (USER™) for deoxyuridine excision and consequent creation of 7–9 bp long complementary overhangs. These custom-made overhangs enabled the formation of a hybridization product that could be readily ligated. However, the method suffered from one main disadvantage: only a non-proof reading polymerase could be used for the primary PCR amplification to ensure compatibility with deoxyuridine-containing primers (Lasken et al. 1996; Sakaguchi et al. 1996). This introduced the need for Klenow treatment of the primary PCR products and, above all, created a high risk of PCR-introduced errors. Nevertheless, since the production of ORF2 and ORF2nat, a proof-reading polymerase that is compatible with the method (Pfu Turbo Cx, Stratagene) has been identified (Nour-Eldin et al. 2006). Moreover, the method has been further developed to combine the

assembly and cloning in a single step, eliminating the need for restriction digestions, ligations or secondary amplifications (Geu-Flores et al. 2007).

In vitro transcription/translation of ORF2 and ORF2nat showed the successful production of three individual proteins, but also showed an imbalance in protein production. For ORF2, the protein encoded upstream was about three times more abundant than the one encoded in the middle, and about four times more abundant than the one encoded further downstream. Furthermore, the analysis revealed the production of a fusion protein encoded by the upstream and middle coding sequences. However, this fusion protein accounted only for a small percentage of the produced protein.

Transgenic lines carrying ORF2 and ORF2nat were tested by assays in which PATH was fed to leaves and BGLS was analyzed. Only lines carrying a single T-DNA insertion were tested since single-insertion lines have reduced chances of containing disrupted endogenes and of developing silencing of the transgenes. In these feeding assays, the conversion of PATH to BGLS was a measure of the combined activity of relevant glucosyltransferases and sulfotransferases in planta. The C–S lyase activity was not tested because the proposed substrate (a cysteinyl-S-thiohydroxamate) is highly unstable (Hansen et al. 2001). The different transgenic lines presented a range of PATH to BGLS conversions that reflected differential position effects. Out of eight lines, two resembled wildtype plants and six presented a significantly higher conversion, which for the best line represented 29% of the administered PATH. This is remarkably high since it was found early in the elucidation of the biosynthesis of glucosinolates that *Tropaeolum majus* L., a BGLS-containing plant, converted only 36–38% of the administered PATH into BGLS (Underhill and Wetter 1969). Surprisingly, the feeding assays showed that wildtype plants were able to convert an average of 6% of the administered PATH into BGLS.

As mentioned above, the feeding assays measured *combined* GT and ST activities, including endogenous activities. Even when most of the transgenic lines presented higher conversions than wildtype plants, the possibility existed that one of the two activities was not actually being enhanced by the transgenes, and that increased conversions were due to only one of the two activities being enhanced and the other one being present at wildtype levels. However, when performing separate GT and ST assays using leaf extracts, wildtype plants presented very low background activities, while a selected transgenic line presented conversions that were 30 and 20 times higher, respectively. This proved that each of the individual activities was enhanced in the transgenic line and suggested that the high background conversion of wildtype plants in feeding assays was due to induction of endogenous non-specific GT and ST enzymes.

This assay-dependent induction may have been triggered by wounding (when cutting the leaf petioles) and/or by the toxicity of the administered PATH (Grubb et al. 2004). The latter supports the theory that the post-oxime enzymes in the glucosinolate pathway were recruited from an existing detoxification machinery (Hansen et al. 2001).

The tags encoded in ORF2 (for the sulfotransferase and the glucosyltransferase) enabled us to perform Western blots on a selected transgenic line and show the in planta production of individual proteins from the 2A polycistronic ORF. Additionally, the blots showed the production of protein fusions, as was also seen in in vitro transcription/translation experiments. This confirmed that 2A-mediated ribosomal skips do not occur with 100% efficiency from synthetic ORFs. The ratio of individual/fused protein was much lower in the selected transgenic line than in the in vitro transcription/translation experiments. Prominent protein fusions delivered from synthetic 2A constructs have been seen in planta before (Cruz et al. 1996; Gopinath et al. 2000; Mlotshwa et al. 2002).

The last steps of the BGLS pathway were engineered first in order to prevent the accumulation of intermediates that can hamper future transformations. From the eight transgenic lines having single T-DNA insertions, none presented a visible phenotype in vitro or under greenhouse conditions. Having selected a single-insertion line with optimal BGLS production potential (line ORF2/27.1), we now plan to introduce the initial steps of the pathway as the next step towards engineering glucosinolates in non-cruciferous plants.

Acknowledgments The authors would like to thank Douglas Grubb, Brandon Zipp (University of California, Davis, USA) and Leszek Doszczak (Universität Würzburg, Germany) for advice on the synthesis of PATH; Claire Halpin (University of Dundee, Scotland) for advice on the 2A system and Bodil Jørgensen (University of Copenhagen, Denmark) for advice on transformation of tobacco. Niels Agerbirk (University of Copenhagen, Denmark) is acknowledged for the kind gift of benzylglucosinolate. This work was supported by the Danish International Developmental Agency (DANIDA, project 91175) and by the Villum Kann Rasmussen Fond through its support to the VKR Research Centre Pro-Active Plants.

References

- Atkins JF, Wills NM, Loughran G, Wu CY, Parsawar K, Ryan MD, Wang CH, Nelson CC (2007) A case for “StopGo”: Reprogramming translation to augment codon meaning of GGN by promoting unconventional termination (Stop) after addition of glycine and then allowing continued translation (Go). *RNA* 13:803–810
- Bianchini F, Vainio H (2004) Isothiocyanates in cancer prevention. *Drug Metab Rev* 36:655–667
- Bones AM, Rossiter JT (2006) The enzymic and chemically induced decomposition of glucosinolates. *Phytochemistry* 67:1053–1067
- Brader G, Mikkelsen MD, Halkier BA, Tapio Palva E (2006) Altering glucosinolate profiles modulates disease resistance in plants. *Plant J* 46:758–767
- Cruz SS, Chapman S, Roberts AG, Roberts IM, Prior DA, Oparka KJ (1996) Assembly and movement of a plant virus carrying a green fluorescent protein overcoat. *Proc Natl Acad Sci USA* 93:6286–6290
- de Felipe P, Luke GA, Hughes LE, Gani D, Halpin C, Ryan MD (2006) E unum pluribus: multiple proteins from a self-processing polyprotein. *Trends Biotechnol* 24:68–75
- Donnelly MLL, Luke G, Mehrotra A, Li X, Hughes LE, Gani D, Ryan MD (2001) Analysis of the aphthovirus 2A/2B polyprotein ‘cleavage’ mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal ‘skip’. *J Gen Virol* 82:1013–1025
- Doszczak L, Rachon J (2002) S-thioacyldithiophosphates in the synthesis of thiohydroxamic acids and O-thioacylhydroxylamines. *Synthesis* 8:1047–1052
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11–15
- Ettlinger MG, Lundeen AJ (1957) First synthesis of a mustard oil glucoside; the enzymatic Lossen rearrangement. *J Am Chem Soc* 79:1764–1765
- Fahey JW, Zalcman AT, Talalay P (2001) The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* 56:5–51
- Fahey JW, Haristoy X, Dolan PM, Kensler TW, Scholtus I, Stephenson KK, Talalay P, Lozniewski A (2002) Sulforaphane inhibits extracellular, intracellular, and antibiotic-resistant strains of *Helicobacter pylori* and prevents benzo[*a*]pyrene-induced stomach tumors. *Proc Natl Acad Sci USA* 99:7610–7615
- Fimognari C, Hrelia P (2007) Sulforaphane as a promising molecule for fighting cancer. *Mutat Res/Rev Mutat* 635:90–104
- Geu-Flores F, Nour-Eldin HH, Nielsen MT, Halkier BA (2007) USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. *Nucleic Acids Res* 35:e55
- Gopinath K, Wellink J, Porta C, Taylor KM, Lomonosoff GP, van Kammen A (2000) Engineering cowpea mosaic virus RNA-2 into a vector to express heterologous proteins in plants. *Virology* 267:159–173
- Grubb CD, Zipp BJ, Ludwig-Muller J, Masuno MN, Molinski TF, Abel S (2004) *Arabidopsis* glucosyltransferase UGT74B1 functions in glucosinolate biosynthesis and auxin homeostasis. *Plant J* 40:893–908
- Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. *Annu Rev Plant Biol* 57:303–333
- Hansen BG, Kliebenstein DJ, Halkier BA (2007) Identification of a flavin-monoxygenase as the S-oxygenating enzyme in aliphatic glucosinolate biosynthesis in *Arabidopsis*. *Plant J* 50:902–910
- Hansen CH, Du L, Naur P, Olsen CE, Axelsen KB, Hick AJ, Pickett JA, Halkier BA (2001) CYP83B1 is the oxime-metabolizing enzyme in the glucosinolate pathway in *Arabidopsis*. *J Biol Chem* 276:24790–24796
- Horsch RB, Fry JE, Hoffmann NL, Eichholtz D, Rogers SG, Fraley RT (1985) A simple and general method for transferring genes into plants. *Science* 227:1229–1231
- Kirkegaard JA, Sarwar M, Wong PTW, Mead A, Howe G, Newell M (2000) Field studies on the biofumigation of take-all by *Brassica* break crops. *Aust J Agric Res* 51:445–456
- Klein M, Reichelt M, Gershenzon J, Papenbrock J (2006) The three desulfoglucosinolate sulfotransferase proteins in *Arabidopsis* have different substrate specificities and are differentially expressed. *FEBS J* 273:122–136
- Lasken RS, Schuster DM, Rashtchian A (1996) Archaeobacterial DNA polymerases tightly bind uracil-containing DNA. *J Biol Chem* 271:17692–17696
- Mikkelsen MD, Naur P, Halkier BA (2004) *Arabidopsis* mutants in the C–S lyase of glucosinolate biosynthesis establish a critical role for indole-3-acetaldoxime in auxin homeostasis. *Plant J* 37:770–777

- Mlotshwa S, Verver J, Sithole-Niang I, Gopinath K, Carette J, Kammen Av, Wellink J (2002) Subcellular location of the helper component-proteinase of cowpea aphid-borne mosaic virus. *Virus Genes* 25:207–216
- Morant A, Jørgensen K, Jørgensen B, Dam W, Olsen C, Møller B, Bak S (2007) Lessons learned from metabolic engineering of cyanogenic glucosides. *Metabolomics* 3:383–398
- Nour-Eldin HH, Hansen BG, Norholm MHH, Jensen JK, Halkier BA (2006) Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res* 34:e122
- Piotrowski M, Schemenewitz A, Lopukhina A, Muller A, Janowitz T, Weiler EW, Oecking C (2004) Desulfoglucosinolate sulfotransferases from *Arabidopsis thaliana* catalyze the final step in the biosynthesis of the glucosinolate core structure. *J Biol Chem* 279:50717–50725
- Sakaguchi AY, Sedlak M, Harris JM, Sarosdy MF (1996) Cautionary note on the use of dUMP-containing PCR primers with Pfu and VentR DNA polymerases. *Biotechniques* 21:368–370
- Smith BJ, Kirkegaard JA, Howe GN (2004) Impacts of *Brassica* break-crops on soil biology and yield of following wheat crops. *Aust J Agric Res* 55:1–11
- Tattersall DB, Bak S, Jones PR, Olsen CE, Nielsen JK, Hansen ML, Hoj PB, Møller BL (2001) Resistance to an herbivore through engineered cyanogenic glucoside synthesis. *Science* 293:1826–1828
- Underhill LEW, Wetter LR (1969) Biosynthesis of mustard oil glucosides: sodium phenylacetothiohydroximate and desulfobenzylglucosinolate, precursors of benzylglucosinolate in *Tropaeolum majus*. *Plant Physiol* 44:584–590
- Walter W, Schaumann E (1971) The chemistry of thiohydroxamic acids. *Synthesis* 3:111–130
- Wittstock U, Kliebenstein DJ, Lambrix VM, Reichelt M, Gershenzon J (2003) Glucosinolate hydrolysis and its impact on generalist and specialist insect herbivores. In: Romeo JT (ed) *Integrative phytochemistry: From ethnobotany to molecular ecology*, vol 37. Elsevier, Amsterdam, pp 101–125