Metabolic engineering of dhurrin in transgenic *Arabidopsis* plants with marginal inadvertent effects on the metabolome and transcriptome

Charlotte Kristensen*†, Marc Morant*‡, Carl Erik Olsen‡§, Claus T. Ekstrøm§, David W. Galbraith¶, Birger Lindberg Møller*‡, and Søren Bak*‡‖

*Plant Biochemistry Laboratory, Department of Plant Biology, and ‡Center for Molecular Plant Physiology and §Department of Natural Sciences, Royal Veterinary and Agricultural University, 40 Thorvaldsensvej, DK-1871 Frederiksberg C, Copenhagen, Denmark; and ¶Department of Plant Sciences, University of Arizona, 303 Forbes Building, Tucson, AZ 85721

Communicated by Eric E. Conn, University of California, Davis, CA, December 20, 2004 (received for review September 20, 2004)

Focused and nontargeted approaches were used to assess the impact associated with introduction of new high-flux pathways in *Arabidopsis thaliana* by genetic engineering. Transgenic *A. thaliana* plants expressing the entire biosynthetic pathway for the tyrosine-derived cyanogenic glucoside dhurrin as accomplished by insertion of *CYP79A1*, *CYP71E1*, and *UGT85B1* from *Sorghum bicolor* were shown to accumulate 4% dry-weight dhurrin with marginal inadvertent effects on plant morphology, free amino acid pools, transcriptome, and metabolome. In a similar manner, plants expressing only *CYP79A1* accumulated 3% dry weight of the tyrosine-derived glucosinolate, *p*-hydroxybenzylglucosinolate with no morphological pleitropic effects. In contrast, insertion of *CYP79A1* plus *CYP71E1* resulted in stunted plants, transcriptome alterations, accumulation of numerous glucosides derived from detoxification of intermediates in the dhurrin pathway, and in loss of the brassicaceae-specific UV protectants sinapoyl glucose and sinapoyl malate and kaempferol glucosides. The accumulation of glucosides in the plants expressing CYP79A1 and CYP71E1 was not accompanied by induction of glycosyltransferases, demonstrating that plants are constantly prepared to detoxify xenobiotics. The pleiotrophic effects observed in plants expressing sorghum CYP79A1 and CYP71E1 were complemented by retransformation with *S. bicolor UGT85B*. These results demonstrate that insertion of high-flux pathways directing synthesis and intracellular storage of high amounts of a cyanogenic glucoside or a glucosinolate is achievable in transgenic *A. thaliana* plants with marginal inadvertent effects on the transcriptome and metabolome.

DNA microarrays | metabolic engineering | metabolite profiling | channeling | substantial equivalence

Metabolic engineering offers the possibility to design plants that produce desired compounds (1–3). Knowledge of the impact of insertion of metabolic pathways on preexisting metabolic pathways is typically scarce, and is assessed by targeted approaches, but not by more global and unbiased approaches. We have previously reported transfer of the entire pathway for synthesis of the tyrosine-derived cyanogenic glucoside dhurrin from *Sorghum bicolor* to *Arabidopsis thaliana* by coexpression of three *S. bicolor* cDNAs: two multifunctional cytochromes P450, *CYP79A1* and *CYP71E1*, and a family 1 glucosyltransferase, *UGT85B1* (Fig. 1). The transgenic *A. thaliana* plants accumulated up 4% of plant dry matter of dhurrin (4). Transgenic *A. thaliana* plants in which only *CYP79A1* was introduced accumulated up to 3% dry matter of the tyrosine derived *p*-hydroxybenzylglucosinolate, a glucosinolate that does not occur naturally in *A. thaliana* (5). The accumulation of tyrosine derived *p*-hydroxybenzylglucosinolate in *A. thaliana* is the result of metabolic crosstalk, where the *p*-hydroxyphenylacetaldoxime produced by CYP79A1 is efficiently used by the post-oxime-metabolizing enzymes CYP83B1 or CYP83A1 in the endogenous glucosinolate biosynthetic pathway (5, 6) (Fig. 1).

In this study, we have used nontargeted and targeted approaches to investigate the effect of insertion of the entire high-flux pathway for dhurrin synthesis in *A. thaliana* on the transcriptome and metabolome, and compared these effects with those accompanying insertion of only parts of the pathway. We demonstrate that it is possible to engineer plants to express high-flux biosynthetic pathways related to cyanogenic glucosides and glucosinolates production and to accumulate large amounts of these types of natural products without unexpected and significant changes in the transcriptome or metabolome. When incomplete pathways were inserted, metabolic crosstalk or detoxification reactions were found to induce significant changes in plant morphology, the transcriptome, and metabolome.

Materials and Methods

Cultivation of Plants. All plant lines used were homozygous for the transgenes. Plants were grown in 9-cm diameter plastic pots in autoclaved soil supplemented with vermiculite (10% vol/vol) in an insect-free growth chamber at 22°C, with 70% humidity, a photosynthetic flux of 100 μmol photons $m^{-2}{\cdot}s^{-1}$, and a 16-h light:8-h dark regime. One week after germination, seedlings were thinned out, leaving three plants equally spaced in each pot. Plants were watered every fifth day. Rosette leaves were harvested directly into liquid nitrogen at ≈30 day after germination or at similar developmental stages, i.e., just as the inflorescence started to appear in the case of the 2*x* lines. To diminish plant to plant variation, eight leaves from 8–35 individual plants were combined. Wild-type and transgenic lines were grown side by side.

LC-MS Analysis. Plant material was extracted (2 min) in 85% (vol/vol) boiling methanol. The solvent was evaporated by lyophilization. The residue was dissolved in water and extracted three times with *n*-pentane and analyzed as described (7). The mass spectrometer was run in both positive and negative modes. Compounds were identified based on mass spectra and retention times for authentic standards.

Oligonucleotide Microarrays. The focused array was a custom designed 50-mer oligonucleotide array, with 350-μm element spacing, spotted by MWG Biotech using a single pin on epoxy-coated glass slides. The array contained probes for 453 selected *A. thaliana* genes spotted in duplicate (Data Set 1, which is published as supporting information on the PNAS web site). The 50-mer oligonucleotides were designed by MWG Biotech, essentially as described in ref. 8.

Freely available online through the PNAS open access option.

Abbreviations: SAM, significance analysis of microarrays; rt, retention time; UGT, glycosyltransferase.

†Present address: Poalis A/S, Bülowsvej 25, 1870 Frederiksberg C, Denmark.

‖To whom correspondence should be addressed. E-mail: bak@kvl.dk.

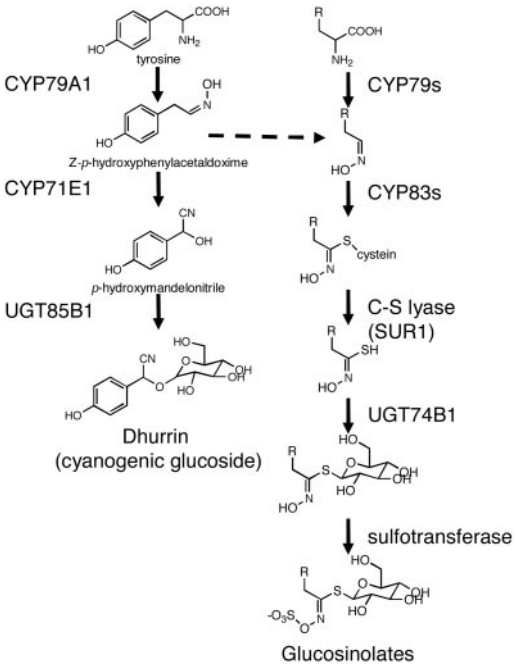

Fig. 1. Biosynthesis of the sorghum-derived cyanogenic glucoside dhurrin and glucosinolates in transgenic *A. thaliana* plants. CYP79A1, CYP71E1, and UGT85B1 are enzymes encoded by the introduced sorghum transgenes. The dashed arrow illustrates metabolic crosstalk resulting from interaction of sorghum CYP79A1 with preexisting endogenous post-oxime-metabolizing enzymes (CYP83B1, SUR1, a sulfotransferase, and UGT74B1), enabling use of *p*-hydroxyphenylacetaldoxime produced by CYP79A1 by the post-oxime-metabolizing enzymes in the preexisting glucosinolate biosynthetic pathway. The presence of CYP71E1 prevents this interaction.

The global 70-mer oligonucleotide array probes were produced by Qiagen-Operon (which can be accessed at http://oparray.operon.com/arabidopsis/index.php) and printed at 180-μm spacing on Telechem superamine aminosilane-coated slides. A complete listing of probes can be accessed at www.ag.arizona.edu/microarray. Isolation and labeling of mRNA and hybridization of arrays were performed as described (9). The arrays were scanned by using an Affymetrix (Santa Clara, CA) GMS 418 Array Scanner using four different photomultiplier gains, while keeping the laser power at 30. IMAGENE Version 5.6.0 (Biodiscovery, Marina del Rey, CA) was used to analyze the scanned TIFF files. Scatterplots were obtained with GENESIGHT Version 4.0 (Biodiscovery) using LOWESS normalization, but with omission of the background subtraction step. For the statistical analysis, LOWESS-normalized data were analyzed with significance analysis of microarrays (SAM) (10) Version 1.21.

Results

Morphological Phenotypes and Metabolite Profiling. Three different types of transgenic *A. thaliana* plants were compared with wild-type plants (Fig. 2). The plants designated 1*x* ectopically express sorghum *CYP79A1* (5), 2*x* lines express sorghum *CYP79A1* and *CYP71E1* (6), whereas 3*x* lines express all three sorghum genes involved in dhurrin biosynthesis: *CYP79A1*, *CYP71E1*, and *UGT85B1* (4). The impact on the metabolome was assessed by LC-MS analysis of extracts of rosette leaves. To limit variation between plants due to different environmental challenges, extensive care was taken to grow the plants under controlled conditions, to eliminate abiotic and biotic stresses, and to collect plant samples at precisely the same time of the day, and at the same developmental stage.

The 1*x* line used in this study was selected among the original 18 1*x* transgenic *A. thaliana* lines expressing sorghum CYP79A1. This line accumulated 2–3% of its dry matter as tyrosine-derived *p*-hydroxybenzylglucosinolate, as determined by GC-MS and HPLC analyses of the glucosinolate content after conversion into desulfoglucosinolates by sulfatase treatment (5). In the current LC-MS analysis (table 1, Fig. 2), *p*-hydroxybenzylglucosinolate was detected directly and gave rise to two peaks when analyzed in positive mode (Fig. 2). These two peaks correspond to intact *p*-hydroxybenzylglucosinolate [no. 7, retention time (rt) = 4.0] and desulfo *p*-hydroxybenzylglucosinolate (no. 8, rt = 6.5 min), as the result of a partial loss of the sulfate group in the mass spectrometer during the electrospray ionization ionization process (data not shown). The 1*x* plant line also accumulated minute amounts of glucosides derived from detoxification of *p*-hydroxyphenylacetaldoxime (Fig. 2, Table 1, and refs. 5 and 6). Compared with wild-type plants, the level of sinapoylglucose (no. 14, rt = 18.4 min), the shared precursor for sinapoylcholine and sinapoylmalate, was found to be ≈7 fold down-regulated. Two metabolites, nos. 9 and 10, with unknown

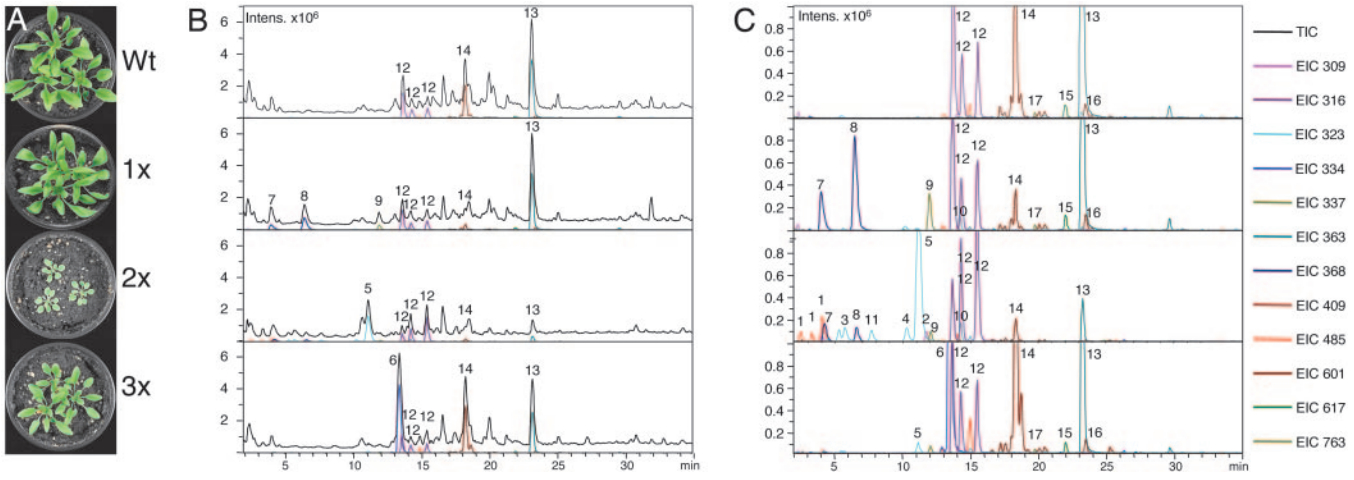

Fig. 2. Comparison of plant morphology and metabolite composition in wild-type *A. thaliana* and transgenic lines. Plants expressing sorghum CYP79A1 are designated 1*x*, plants expressing sorghum CYP79A1 and CYP71E1 are designated 2*x*, and plants expressing CYP79A1, CYP71E1, and UGT85B1 are designated 3*x*. Morphological phenotype (*A*) and metabolite profile as monitored as total ion trace (TIC) and extracted ion chromatographs (EIC) in the three transgenic lines are shown (*B*). (*C*) Close-ups of the extracted ion chromatograms to facilitate visualization of minor components. Compound numbers and absolute and relative changes are tabulated in Table 1.

Table 1. Metabolic profiling of wild-type and transgenic *A. thaliana* lines as monitored by extracted ion chromatography from LC-MS (Fig. 2)

No.	Metabolite	rt, min	$[M+Na]^+$	M^+	1*x*	2*x*	3*x*	wt
1	*p*-glucosyloxy-benzylglucose*	2.5	485	462	–	+	–	–
		3.4	485	462	–	+	–	–
		4.1	485	462	–	++	–	–
2	*p*-glucosyloxy-phenylmethanol*	11.7	309	286	–	+	–	–
3	*p*-glucosyloxy-benzoic acid*	5.8	323	300	–	+	–	–
4	*p*-glucosyloxy-phenylethanol†	10.3	323	300	+	+	–	–
5	*p*-hydroxybenzoylglucose*	11.1	323	300	–	+++	+	–
6	Dhurrin*	13.5	334	311	–	–	+++	–
7	*p*-hydroxybenzylglucosinolate†	4.0	368	345	+++	+	–	–
8	Desulfo *p*-hydroxybenzylglucosinolate†	6.5	368	345	+++	+	–	–
9	Unknown glucoside	12.0	337	314	++	+	+	–
10	Unknown compound	14.2	–	323	+	+	–	–
11	Unknown compound	7.7	–	323	–	+	–	–
12	Unknown	13.7		316	+++	+	+++	+++
	Unknown	14.3	–	316	+++	++	+++	+++
	Unknown	15.5		316	+++	++++	+++	+++
13	Sinapoyl malate	23.2	363	339	+++	+	+++	+++
14	Sinapoyl glucose	18.4	409	385	+	+	+++	+++
15	Kaempferol-3-*O*-glucoside-7-*O*-rhamnoside	22.0	617	594	++	+	++	++
16	Kaempferol-3-*O*-rhamnoside-7-*O*-rhamnoside	23.6	601	578	++	+	++	++
17	Kaempferol-3-*O*-[rhamnosyl(1→2)glucoside]-7-*O*-rhamnoside	19.7	763	740	++	+	++	++

+, Metabolite present with relative amounts indicated by the number of plus signs. –, metabolite was not detected; rt, retention time.
*Metabolites marked derived from *p*-hydroxyphenylacetaldoxime.
†Metabolites marked derived from *p*-hydroxymandelonitrile (6).

structures, were identified at rt = 12.0 and 14.2 min, respectively. In comparison with the wild-type *A. thaliana* plants, the 1*x* plants did not display any apparent morphological phenotype.

In contrast to the wild-type morphological phenotype of the 1*x* plant line, the 2*x* plant line appeared stressed and displayed a stunted morphological phenotype (Fig. 2 and Fig. 5, which is published as supporting information on the PNAS web site). Among the original 10 2*x* transgenic lines, this stressed and stunted phenotype correlated with the expression level of CYP79A1 and CYP71E1 (6) in that the highest expressers had the most severe phenotype. While the number of rosette leaves appeared similar to those of wild-type plants, they were rounder, and the petioles and the leaves did not elongate to the same extent. The plants were fertile, albeit with a much reduced seed set. In agreement with our previously published data (4, 6), the current LC-MS analyses (Fig. 2 and Table 1) identified three major glucosides derived from *p*-hydroxybenzoate: *p*-glucosyloxy-benzoic acid (no. 3, rt = 5.8 min), *p*-hydroxybenzoylglucose, (no. 5, rt = 11.1 min) (6), and *p*-glucosyloxybenzoylglucose (no. 1, rt = 4.1 min) (4) with *p*-hydroxybenzoylglucose (no. 1) being the most prominent. In addition to the diglucoside *p*-glucosyloxybenzoylglucose (no. 1, rt = 4.1 min) (4), two minor diglucosides of *p*-hydroxybenzoic acid, with an unknown glucose linkage pattern, were identified (no. 1, rt = 2.5 and 3.4 min, respectively). Whereas the apparent expression level of CYP79A1 is similar to the 1*x* line as estimated by the total amount of *p*-hydroxyphenylacetaldoxime-derived metabolites (Fig. 2 and Table 1), the 2*x* plant line accumulated ≈6-fold less *p*-hydroxybenzylglucosinolate (nos. 7 and 8) in comparison with the 1*x* plant line. A range of minor glucosides were detected in addition to those previously identified (Table 1) (6). Compared with that of wild type, the level of both sinapoylglucose (no. 14) and sinapoylmalate (no. 13) in 2*x* plants was reduced 10-fold (Fig. 2 and Table 1). This unexpected decrease in sinapoylglucose and sinapoylmalate levels led us to analyze the impact on the three flavonols known to be present in small amounts in *A. thaliana*: kaempferol-3-*O*-glucoside-7-*O*-rhamnoside (no. 15), kaempferol-3-*O*-rhamnoside-7-*O*-rhamnoside (no. 16), and kaempferol-3-*O*-[rhamnosyl(1→2)glucoside]-7-*O*-rhamnoside (no. 17) (11–13). In the 2*x* plant line, the levels of all three kaempferol glucosides were reduced in comparison with wild-type, 1*x*, and 3*x* plants (Table 1 and Fig. 2). In addition, the relative ratios of a triplet of ions (no. 12) with m/z 316, rt = 13.7, 14.3, and 15.5 min, respectively, of unknown composition, are altered compared with wild-type, 1*x*, and 3*x* plants.

Throughout its life cycle, the 3*x* plant line displayed a normal morphological phenotype, and the LC-MS profile was nearly indistinguishable from wild-type plants, except for the predicted accumulation of large amounts of dhurrin (Fig. 2, no. 6, rt = 13.5 min). The metabolites that accumulated in 2*x* plants were absent in 3*x* plants. Likewise, no traces of *p*-hydroxybenzylglucosinolate (nos. 7 and 8) were detected in the 3*x* plants, whereas the levels of sinapoylmalate (no. 13), sinapoylglucose (no. 14), and kaempferol glucosides (nos. 15, 16, and 17) were comparable with wild-type plants. Because the 3*x* lines are directly derived from the 2*x* lines (ref. 4 and Fig. 5), the morphological phenotype and the metabolite profiling data demonstrate that the glycosyltransferase (UGT) UGT85B1 complements the 2*x* phenotype (Fig. 2).

The amino acid tyrosine is the substrate for sorghum CYP79A1. To determine whether the high levels of dhurrin synthesis in the 3*x* plants affected free amino acid pools, the relative (mol%) of free amino acids in rosette leaves of wild-type and 3*x* plants were determined and normalized to the total concentration of amino acids (Fig. 6, which is published as supporting information on the PNAS web site). No significant changes were observed in the 3*x* plants when compared with wild-type plants.

Changes in the Transcriptome. Our LC-MS analyses of the transgenic lines revealed both expected and unexpected impacts on the metabolome. To study the impact on the transcriptome, and to examine whether changes in the metabolome may be correlated to changes in the transcriptome, the transgenic and wild-type *A. thaliana* plants were analyzed by using both a focused and a global spotted-oligonucleotide microarray. The focused array was designed to enable detection of key changes at the transcriptional level within primary and secondary metabolism that might take place in the 1*x*, 2*x*, and 3*x* plants. The focused array contained 453 selected 50-mer oligonucleotide probes, including all 246 full-length *A. thaliana* cytochromes P450 (ref. 14, which can be accessed at www.P450.kvl.dk) except *CYP84A4* and all 112 family 1 UGTs (15), except for members of the *UGT80* and *UGT81* families, because

PLANT BIOLOGY

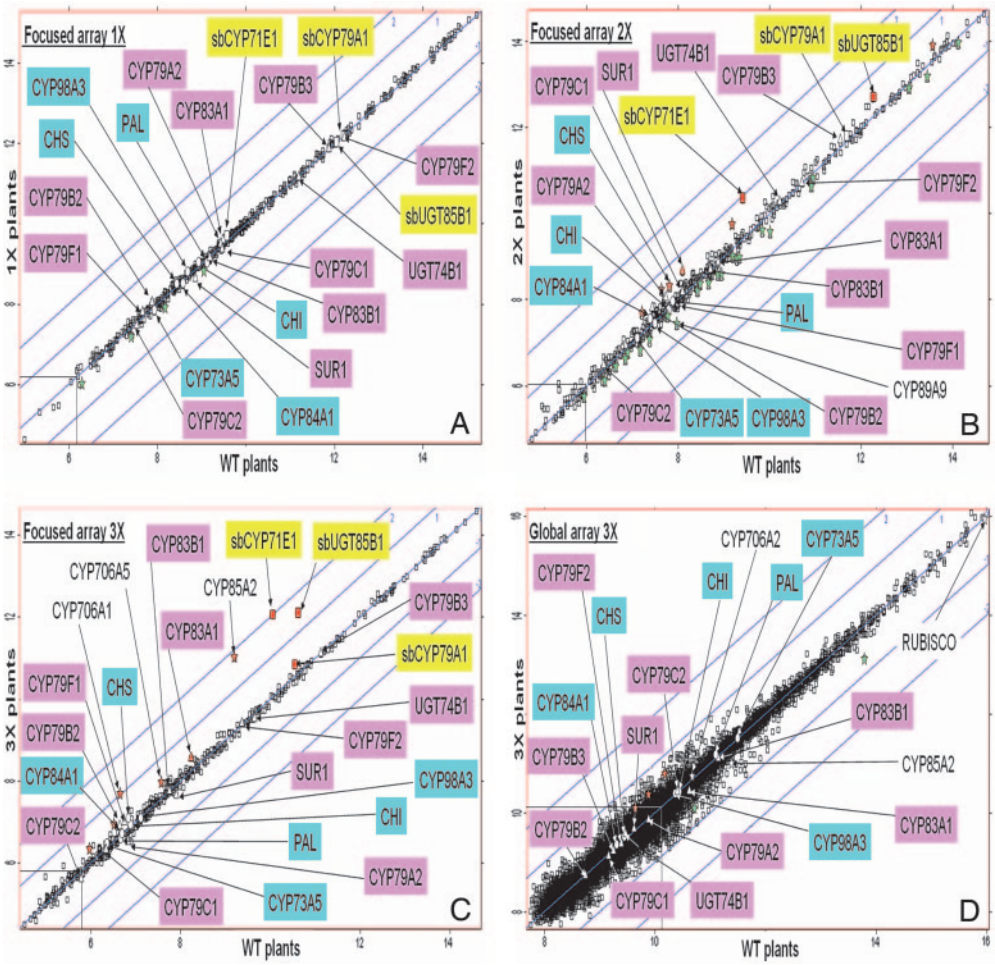

Fig. 3. Transcriptome analyses of wild-type versus transgenic *A. thaliana* lines. Scatterplots of log2-transformed signal intensities from Cy3- and Cy5-labeled mRNA isolated from wild-type (WT) or transgenic *A. thaliana* lines. (*A*) 1*x* lines. (*B*) 2*x* lines. (*C*) 3*x* lines. The focused array (*A–C*) contains probes for 453 selected genes, and the global array (*D*) contains 27,216 probes. Signals below two times the average local background are boxed. Phenylpropanoid marker genes are light blue and represented by circles. PAL, phenylalanine ammonia-lyase; CYP73A5, cinnamate 4-hydroxylase; CYP98A3, coumarate 3-hydroxylase; CYP84A1, ferulate 5-hydroxylase; CHS, chalcone synthase; CHI, chalcone isomerase. Glucosinolate marker genes (CYP79s, CYP83A1, CYP83B1, SUR1, and UGT74B1) are pink and represented by triangles. The Sorghum transgenes are yellow and represented by squares. Genes scored by the SAM analysis as up-regulated are red and genes scored as down-regulated are green (Table 2). Blue lines indicate 2- or 4-fold up- or down-regulated expression levels.

they are sterol and lipid glucosyltransferases, respectively (ref. 15, which can be accessed at www.P450.kvl.dk). The cytochrome P450 and UGT enzyme families have been subject to multiple recent duplication events (15–18). Under these circumstances, oligonucleotide arrays are advantageous and superior to arrays based on longer probes or cDNA because they offer an opportunity to minimize the impact of cross hybridization from closely related paralogs (Fig. 7, which is published as supporting information on the PNAS web site, and refs. 19 and 20). In addition to cytochromes P450 and family 1 UGTs, the focused array contained 36 probes for genes involved in aromatic amino acid biosynthesis, 48 probes comprising additional genes involved in secondary metabolism, marker genes for stress responses, genes controlling indoleacetic acid homeostasis, and 16 probes for constitutively expressed genes (for details see Data Set 1). Finally, the focused array contained probes corresponding to the three sorghum genes: *CYP79A1*, *CYP71E1*, and *UGT85B1*, to enable detection of the introduced transgenes.

The 1*x*, 2*x*, and 3*x* transgenic lines were first analyzed by using the focused array. Arbitrarily, a twofold induction or reduction has traditionally been used to select differentially expressed genes in microarray experiments. Using this criterion, only two endogenous genes were selected as differentially expressed in the 3*x* line, and no genes were selected in the 1*x* and 2*x* lines (Table 2, which is published as supporting information on the PNAS web site, and Figs. 3 and 4). Accordingly, the threshold was lowered to allow detection of differentially expressed genes. In the 1*x* plant line, only four endogenous genes were identified as slightly differentially expressed in comparison with wild type (1.15- to 1.23-fold induction), albeit with a false discovery rate (FDR) of 23%; $\delta = 0.16$, as analyzed by using the SAM program (10) (Figs. 3 and 4 and Table 2). Similarly, in the 3*x* plant line, only six endogenous genes were scored as up-regulated (1.22- to 3.41-fold induction) with a FDR of 10%; $\delta = 0.31$ (Figs. 3 and 4 and Table 2). In contrast to the minute changes observed in the 1*x* and 3*x* lines, the 2*x* line scored six endogenous genes were scored as up-regulated (1.23- 1.44-fold), and 24 genes were scored as down-regulated (1.20- to 1.43-fold) by SAM analysis with an FDR of 2.3%; $\delta = 0.37$ (Figs. 3 and 4 and Table 2). To enable a comparison between the three data sets, the data set was analyzed by fixing the threshold for differentially expressed genes (δ value) to 0.16 (Fig. 4) as used for the 1*x* plant line. A total of 4, 136, and 19 genes, including the transgenes, were scored as differentially expressed for the 1*x*, 2*x*, and 3*x* plant lines, respectively, underpinning that an impact on the transcriptome was mainly observed in the 2*x* plant line. The 3*x* and wild-type plants were subsequently compared by using a global oligonucleotide array that contained 70-mer oligonucleotide probes for 27,216 *A. thaliana* genes. Only three genes were detected as up-regulated (1.37- to 1.58-fold) and two were detected as down-regulated (1.32- and 1.56-fold), with an FDR of 20%; $\delta = 0.11$ (Fig. 3 and Table 2).

The focused array proved to be superior to the global array because a larger fraction of the data points were above the average local background and suitable for SAM analysis. On the focused array, <1% of the data points were below the average local mean background and <13% were below twice the average local mean background. The corresponding values for the global array were 10% and 55%, respectively. On the global array, probes for highly expressed genes like rubisco small subunit (At1g67090) were in-

cluded (Fig. 3), and consequently required a laser setting avoiding signal saturation of such highly expressed genes at the expense of sensitivity toward low expressed genes (9). As a consequence, signals of all known genes in glucosinolate biosynthesis and of markers for phenylpropanoid biosynthesis were more than two times the average local background on the focused array (Fig. 3), whereas on the global array, only *CYP83A1* and *CYP83B1* offered signals that were two times higher than the average local background (compare Fig. 3 *C* and *D*). Mainly because of higher spot density in combination with relatively higher background and higher signal range, 42% of the probes on the global array were excluded from the SAM analysis, whereas <13% were excluded on the focused array, despite the fact that six replicates were used for the global array and only four replicates were used for the focused array.

Discussion

Metabolon Formation Ensures Channeling. The 1*x* and 3*x* plants share three common characteristics: (*i*) no obvious morphological phenotype (Fig. 1), (*ii*) accumulation of large amounts of a tyrosine-derived natural product (Fig. 2 and Table 1), and (*iii*) no evident impact of transgene insertion and expression on the metabolome (Fig. 2 and Table 1) and transcriptome (Figs. 3 and 4). In *S. bicolor*, CYP79A1, CYP71E1, and UGT85B1 have been argued to form a dhurrin-producing metabolon, defined as a multienzyme complex that facilitates channeling of the intermediates in dhurrin synthesis (21, 22). Transient expression studies in *S. bicolor* epidermal cells using various GFP and YFP fusions of CYP79A1, CYP71E1, and UGT85B1 indicated that UGT85B1 colocalizes with CYP79A1 and CYP71E1 (ref. 23 and K. A. Nielsen, D. B. Tattersall, and B.L.M., unpublished data). Our metabolite-profiling analyses further corroborate these interpretations (Fig. 2). The high levels of *p*-hydroxybenzylglucosinolate in 1*x A. thaliana* plants (nos. 7 and 8) (5, 6), would support that CYP79A1 associates, interacts, or at least colocalizes with the endogenous post-oxime metabolizing enzymes in glucosinolate synthesis i.e., CYP83A1 or CYP83B1 (Fig. 1). When CYP79A1 was expressed in *Nicotiana tabacum* (6), a plant not producing glucosinolates nor cyanogenic glucosides, only very low expression levels of CYP79A1 could be achieved, and the reactive and unstable *p*-hydroxyphenylacetaldoxime produced was swiftly subjected to detoxification reactions (6). In the 2*x A. thaliana* plant line, CYP71E1 most likely associates and interacts with CYP79A1, and thereby inhibits CYP79A1 interaction with the endogenous post-oxime-metabolizing enzymes. Consequently, only low levels of *p*-hydroxybenzylglucosinolate (nos. 7 and 8) accumulated in 2*x* plants, whereas the cyanohydrin produced (Fig. 1) is subjected to detoxification reactions and glucosylations (ref. 6 and Table 1). In the 3*x* plant line, no *p*-hydroxybenzylglucosinolate is detectable. This finding augments the proposition that UGT85B1 interacts or colocalizes with CYP79A1 and CYP71E1, thereby excluding *p*-hydroxybenzylglucosinolate formation (23). The 3*x* plant lines were derived from the 2*x* lines by retransformation of the 2*x* plant lines with a 35S::UGT85B1 construct (4). Accordingly, introduction of UGT85B1 facilitated not only rescue of the stressed and stunted phenotype but also eliminated the large majority of novel metabolites observed to accumulate in the 2*x* plants (Fig. 2 and Table 1), and the impact on the transcriptome (Figs. 3 and 4 and Table 2).

In the 1*x* plants, the total levels of glucosinolates were increased 4-fold compared with wild type. The level of *p*-hydroxybenzylglucosinolate accounted for the increase (5). The increased flux through the glucosinolate pathways was not accompanied by up-regulation of the steady-state levels of the known post-oxime genes *CYP83A1*, *CYP83B1*, *SUR1*, or *UGT74B1* within the glucosinolate biosynthetic pathway. In studies of dhurrin synthesis in sorghum, the initial step catalyzed by CYP79A1 has been shown to be rate limiting for the overall pathway and to be regulated at the transcriptional level (24). The transcriptome and metabolome data obtained with the 1*x* plants strongly suggest that the same step is rate limiting in glucosinolate synthesis.

The steady-state transcriptional level of the three sorghum genes *CYP79A1*, *CYP71E1*, and *UGT85B1* was not detected as highly elevated over background levels in the transgenic plants (Fig. 3). High expression levels of the three introduced transgenes would have been expected, because the transgenes were under regulatory control of the strong CaMV 35S promoter, and because high levels of tyrosine derived products were observed to accumulate in the transgenic lines (Table 1). The relatively high signal intensities detected with the sorghum probes in wild-type plants (Fig. 3) indicated that these probes were crossreacting with endogenous *A. thaliana* transcripts. Whereas sorghum *CYP71E1* and *UGT85B1* were clearly overexpressed in 3*x* plants and *CYP71E1* was clearly overexpressed in 2*x* plants, expression of sorghum *CYP79A1* was only marginally above background level in 1*x*, 2*x*, and 3*x* plants (Fig. 3 and Table 2). CYP79A1 occupies a branch point between primary and secondary metabolism, and CYP79A1 has a relatively high turnover number (K_m = 0.21 mM and k_{cat} = 200 min^{-1}) (25). Accordingly, very low steady-state levels of *CYP79A1* transcript may enable production of sufficient amounts of CYP79A1 enzyme to turn over substantial amounts of tyrosine. This proposition is supported by the fact that dhurrin accounts for 30% of the dry matter of the upper part of etiolated *S. bicolor* seedlings (26), although CYP79A1 only constitutes ≈0.2% of the microsomal protein (25). It also concurs with the inability to detect CYP79A1 and CYP71E1 protein in the 1*x*, 2*x*, and 3*x* transgenic lines by Western blotting. Accordingly, a strong selection pressure against high expression must have been encountered in the generation of the transgenic lines.

Pleiotropic Effects. Surprisingly, metabolite profiling of the 2*x* plants revealed that sinapoylmalate (no. 13), sinapoylglucose (no. 14), and kaempferol glucoside (nos. 15, 16, and 17) levels were decreased compared with wild-type plants (Fig. 2 and Table 1). Sinapoylmalate and sinapoylglucose serve as UV protectants and constitute the major sinapate esters in *A. thaliana* leaves (27). Several mutants with impaired sinapoylmalate levels, designated reduced epidermal fluorescence (*ref*) mutants have been characterized (28). The *ref2* mutant has decreased lignin content and ≈3-fold-reduced sinapate ester levels (29). The *ref2* mutant is defective in *CYP83A1*, one of the two oxime-metabolizing cytochrome P450s in glucosinolate biosynthesis (30, 31). It has been suggested that accumulation of oximes or their degradation products perturbs the biosynthesis of phenylpropanoids (29). In the 1*x* plant line, the only major new metabolite is *p*-hydroxybenzylglucosinolate (nos. 7 and 8) (Fig. 2 and Table 1). This finding implies that sorghum CYP79A1 colocalized or associated and interacted with the preexisting post-oxime biosynthetic enzymes, e.g., CYP83's, SUR1, and UGT74B1, of the endogenous glucosinolate pathway in *A. thaliana*. Minute amounts of *p*-hydroxyphenylacetaldoxime detoxification products did accumulate and their presence may explain the reduced content of sinapoyl glucose (Table 1). The reduction of sinapoylmalate and sinopoylglucose and the three kaempferol glucosides in the 2*x* line was not accompanied by a reduction in steady-state transcript levels of UGT84s, UGT73C6, and UGT78D1, the corresponding sinapate and kaempferol glucosyltransferases (32, 33).

The microarray analyses revealed that in 2*x* and 3*x* plants, *CYP85A2* was up-regulated 1.44- and 3.41-fold, respectively (Fig. 3 and Table 2). CYP85A2 has been predicted to be involved in brassinolide biosynthesis (34). Similarly, *CYP706A1* and *CYP706A5* were up-regulated, 2.02- and 1.29-fold, respectively, in 3*x* plants (Table 2). The function of CYP706s in *A. thaliana* is not known. Induction of *CYP85A2* and *CYP706A5* in 3*x* plants was not confirmed by using the global array. *CYP706A1* is not represented on the global array, whereas *CYP85A2* is. Real-time PCR was used to study the precise expression levels of *CYP83B1*, *CYP706A1*, and *CYP85A2* in the 3*x* lines relative to wild type and normalized to *actin*

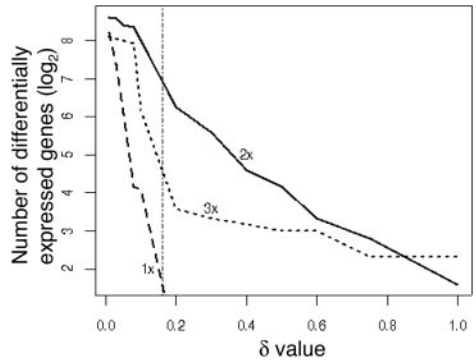

Fig. 4. Number of differentially expressed genes (log_2-transformed) as a function of the threshold for selection of differentially expressed genes (δ value) in the three transgenic lines (1*x*, 2*x*, and 3*x*) as analyzed by SAM analysis of the focused oligonucleotide array. The vertical line indicates a δ value of 0.16. A δ value of 1 is equivalent to a 2-fold change in ratio.

3 (data not shown). In agreement with the data obtained from the focused array, *CYP85A2* was slightly but significantly induced (2.5-fold), whereas *CYP83B1* was not (1.1-fold induction). *CYP706A1* was not induced, which might indicate that the induction observed in the focused microarray analysis (2.02-fold) represented a false-positive.

Surprisingly, no UGT mRNAs were detected as being up-regulated in 2*x* plants, although these plants accumulated high levels of tyrosine-derived glucosides (Fig. 2, Table 1, and ref. 6). *A. thaliana* UGTs that glucosylate *p*-hydroxybenzoic acid *in vitro* have been identified (35), but, as shown for UGT85B1, *in vivo* metabolon formation (23) may restrict the *in vitro*-observed substrate specificity (7). Our transcriptome analyses demonstrated that plants appear constantly prepared to detoxify xenobiotics. Recently, a family 1 UGT, UGT1A1, expressed in human colon cells has been shown to be posttranscriptionally regulated by phosphorylation (36). A similar scenario in plants would explain our microarray results.

Increased metabolic flux from tyrosine into dhurrin in the 3*x* plants was neither accompanied by changes in the transcriptome nor in the pool of free amino acids. Aromatic amino acids are derived from the shikimate acid pathway and are precursors for a vast array of secondary metabolites (37, 38). Up to 20% of the carbon flow in plants passes through the shikimate pathway (39). Aromatic amino acid biosynthesis in plants is allosterically regulated at the enzyme level to accommodate rapid changes in flux and demands for aromatic amino acid-derived natural products (39, 40).

Our combined analyses of the morphological phenotypes, and of metabolite and transcriptome profiles of the transgenic plants used in this study, demonstrate that insertion of an entire high-flux pathway into a transgenic plant is achievable without pleiotropic side effects. The flexibility of the shikimate pathway is demonstrated by the fact that we have increased the dry-weight matter of a tyrosine-derived metabolite to 4% with no impact on free amino acid pools, and without changing the steady-state transcriptional level of genes encoding enzymes in aromatic amino acid biosynthesis. Our data support that formation of metabolons serve to facilitate metabolic channeling to prevent escape of toxic and labile intermediates and to avoid interference with other parts of basic metabolism at all levels (22). The 1*x* and 2*x* plants used in this study illustrate this phenomenon. In the 2*x* plants, the predicted result of gene insertion would be formation of a labile and toxic cyanohydrin. The 2*x* plant counteracts the accumulation of the cyanohydrin by metabolism and detoxification reactions, as is evident from the altered phenotype, the altered metabolite profile showing accumulation of detoxification products, and the changes in transcriptome profile. In comparison with the metabolic and transcriptome changes that have typically been encountered as a result of changed growth conditions or mutations (e.g., refs. 41–44), the changes in metabolite and transcriptome profiles in the 2*x* plants were minor. Thus, targeted introduction of traits by genetic engineering based on a solid knowledge of plant metabolism offers the opportunity to generate cultivars that more strictly adhere to the principle of substantial equivalence than cultivars generated by classical breeding procedures.

We thank Dr. B. A. Halkier (Royal Veterinary and Agricultural University) for providing transgenic 1x seeds. This work was supported in part by National Science Foundation Grant DBI 0211857 (to D.W.G.), a grant from the Danish National Research Foundation to the Center for Molecular Plant Physiology (to C.K., M.M., C.E.O., B.L.M., and S.B.), and Danish Agricultural and Veterinary Research Council Grant 23-02-0095.

1. Hughes, E. H. & Shanks, J. V. (2002) *Metab. Eng.* **4,** 41–48.
2. Verpoorte, R. & Memelink, J. (2002) *Curr. Opin. Biotechnol.* **13,** 181–187.
3. Morant, M., Bak, S., Møller, B. L. & Werck-Reichhart, D. (2003) *Curr. Opin. Biotechnol.* **14,** 151–162.
4. Tattersall, D. B., Bak, S., Jones, P. R., Olsen, C. E., Nielsen, J. K., Hansen, M. L., Høj, P. B. & Møller, B. L. (2001) *Science* **293,** 1826–1828.
5. Bak, S., Olsen, C. E., Petersen, B. L., Møller, B. L. & Halkier, B. A. (1999) *Plant J.* **20,** 663–672.
6. Bak, S., Olsen, C. E., Halkier, B. A. & Møller, B. L. (2000) *Plant Physiol.* **123,** 1437–1448.
7. Hansen, K. S., Kristensen, C., Tattersall, D. B., Jones, P. R., Olsen, C. E., Bak, S. & Møller, B. L. (2003) *Phytochemistry* **64,** 143–151.
8. Kane, M. D., Jatkoe, T. A., Stumpf, C. R., Lu, J., Thomas, J. D. & Madore, S. J. (2000) *Nucleic Acids Res.* **28,** 4552–4557.
9. Ekstrøm, C. T., Bak, S., Kristensen, C. & Rudemo, M. (2004) *Bioinformatics* **20,** 2270–2278.
10. Tusher, V. G., Tibshirani, R. & Chu, G. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 5116–5121.
11. Veit, M. & Pauli, G. F. (1999) *J. Nat. Prod.* **62,** 1301–1303.
12. Rohde, A., Morreel, K., Ralph, J., Goeminne, G., Hostyn, V., De Rycke, R., Kushnir, S., Van Doorsselaere, J., Joseleau, J. P., Vuylsteke, M., *et al.* (2004) *Plant Cell* **16,** 2749–2771.
13. Jones, P., Messner, B., Nakajima, J., Schäffner, A. R. & Saito, K. (2003) *J. Biol. Chem.* **278,** 43910–43918.
14. Werck-Reichhart, D., Bak, S. & Paquette, S. (2002) *The Arabidopsis Book: Cytochromes P450*, eds. Somerville, C. R. & Meyerowitz, E. M. (Am. Soc. Plant Biol., Rockville, MD).
15. Paquette, S. M., Møller, B. L. & Bak, S. (2003) *Phytochemistry* **62,** 399–413.
16. Moore, R. C. & Purugganan, M. D. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 15682–15687.
17. Paquette, S. M., Bak, S. & Feyereisen, R. (2000) *DNA Cell Biol.* **19,** 307–317.
18. Nelson, D. R., Schuler, M. A., Paquette, S. M., Werck-Reichhart, D. & Bak, S. (2004) *Plant Physiol.* **135,** 756–772.
19. Relógio, A., Schwager, C., Richter, A., Ansorge, W. & Valcárcel, J. (2002) *Nucleic Acids Res.* **30,** e51.
20. Xu, W., Bak, S., Decker, A., Paquette, S. M., Feyereisen, R. & Galbraith, D. W. (2001) *Gene* **272,** 61–74.
21. Møller, B. L. & Conn, E. E. (1980) *J. Biol. Chem.* **255,** 3049–3056.
22. Winkel, B. S. J. (2004) *Annu. Rev. Plant Biol.* **55,** 85–107.
23. Nielsen, K. A & Møller, B.L. in *Cytochrome P450: Structure, Mechanism and Biochemistry*, ed. Ortiz de Montellano, P. R. (Kluwer Academic/Plenum, New York), 3rd. Ed., pp. 553–583.
24. Busk, P. K. & Møller, B. L. (2002) *Plant Physiol.* **129,** 1222–1231.
25. Sibbesen, O., Koch, B., Halkier, B. A. & Møller, B. L. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 9740–9744.
26. Halkier, B. A. & Møller, B. L. (1989) *Plant Physiol.* **90,** 1552–1559.
27. Milkowski, C., Baumert, A., Schmidt, D., Nehlin, L. & Strack, D. (2004) *Plant J.* **38,** 80–92.
28. Ruegger, M. & Chapple, C. (2001) *Genetics* **159,** 1741–1749.
29. Hemm, M. R., Ruegger, M. O. & Chapple, C. (2003) *Plant Cell* **15,** 179–194.
30. Bak, S. & Feyereisen, R. (2001) *Plant Physiol.* **127,** 108–118.
31. Naur, P., Petersen, B. L., Mikkelsen, M. D., Bak, S., Rasmussen, H., Olsen, C. E. & Halkier, B. A. (2003) *Plant Physiol.* **133,** 63–72.
32. Milkowski, C., Baumert, A. & Strack, D. (2000) *FEBS Lett.* **486,** 183–184.
33. Lim, E. K., Li, Y., Parr, A., Jackson, R., Ashford, D. A. & Bowles, D. J. (2001) *J. Biol. Chem.* **276,** 4344–4349.
34. Bancos, S., Nomura, T., Sato, T., Molnar, G., Bishop, G. J., Koncz, C., Yokota, T., Nagy, F. & Szekeres, M. (2002) *Plant Physiol.* **130,** 504–513.
35. Lim, E. K., Doucet, C. J., Li, Y., Elias, L., Worrall, D., Spencer, S. P., Ross, J. & Bowles, D. J. (2002) *J. Biol. Chem.* **277,** 586–592.
36. Basu, N. K., Kole, L. & Owens, I. S. (2003) *Biochem. Biophys. Res. Commun.* **303,** 98–104.
37. Celenza, J. L. (2001) *Curr. Opin. Plant Biol.* **4,** 234–240.
38. Knaggs, A. R. (2003) *Nat. Prod. Rep.* **20,** 119–136.
39. Herrmann, K. M. (1995) *Plant Cell* **7,** 907–919.
40. Schimd, J. & Amrhein, N. (1995) *Phytochemistry* **39,** 737–749.
41. Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R. N. & Willmitzer, L. (2000) *Nat. Biotechnol.* **18,** 1157–1161.
42. Rizhsky, L., Liang, H. & Mittler, R. (2002) *Plant Physiol.* **130,** 1143–1151.
43. Yamauchi, Y., Ogawa, M., Kuwahara, A., Hanada, A., Kamiya, Y. & Yamaguchi, S. (2004) *Plant Cell* **16,** 367–378.
44. Hirai, M. Y., Yano, M., Goodenowe, D. B., Kanaya, S., Kimura, T., Awazuhara, M., Arita, M., Fujiwara, T. & Saito, K. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 10205–10210.