A Genomics Approach to the Comprehensive Analysis of the Glutathione *S*-Transferase Gene Family in Soybean and Maize

Brian McGonigle, Sharon J. Keeler, Sze-Mei Cindy Lau, Mary K. Koeppe, and Daniel P. O'Keefe*

Nutrition and Health (B.M.), Central Research and Development (S.J.K., S.-M.C.L., D.P.O.), and Crop Protection (M.K.K.), E.I. du Pont de Nemours and Company, Experimental Station, P.O. Box 80328, Wilmington, Delaware 19880–0328

By BLAST searching a large expressed sequence tag database for glutathione *S*-transferase (GST) sequences we have identified 25 soybean (*Glycine max*) and 42 maize (*Zea mays*) clones and obtained accurate full-length GST sequences. These clones probably represent the majority of members of the GST multigene family in these species. Plant GSTs are divided according to sequence similarity into three categories: types I, II, and III. Among these GSTs only the active site serine, as well as another serine and arginine in or near the "G-site" are conserved throughout. Type III GSTs have four conserved sequence patches mapping to distinct structural features. Expression analysis reveals the distribution of GSTs in different tissues and treatments: Maize GSTI is overall the most highly expressed in maize, whereas the previously unknown GmGST 8 is most abundant in soybean. Using DNA microarray analysis we observed increased expression among the type III GSTs after inducer treatment of maize shoots, with different genes responding to different treatments. Protein activity for a subset of GSTs varied widely with seven substrates, and any GST exhibiting greater than marginal activity with chloro-2,4 dinitrobenzene activity also exhibited significant activity with all other substrates, suggesting broad individual enzyme substrate specificity.

Gene families arise through gene duplications and their persistence through time relies on their ability to fulfill one or both of two functions: to provide a level of control of expression that a single gene is unable to provide or to provide proteins with differing functionality (Durbin et al., 2000). The gene family that encodes the glutathione *S*-transferases (GSTs) is a good example of a family that appears to exist to fulfill both functions. To begin to understand this gene family and the functional differences between individual members we have attempted to clone all of the members of this family from the monocotyledonous species maize (*Zea mays*) and the dicotyledonous species soybean (*Glycine max*). Furthermore, we have initiated a study of the protein functionality of selected members of this family by overexpressing the proteins and using purified proteins in in vitro enzymatic assays.

GSTs (EC 2.5.1.18) catalyze the nucleophilic attack of the thiol group of the tripeptide glutathione (GSH) to various electrophilic molecules. GSTs function as dimers composed of either homogeneous or heterogeneous subunits. Mammalian GSTs are categorized into six classes (alpha, mu, pi, sigma, theta, and zeta) based on amino acid identity, immunocrossreactivity, and substrate specificity. Several plant GSTs have been crystallized and their structure characterized (Reinemer et al., 1996; Neuefeind et al., 1997a, 1997b).

Plant GSTs are typically divided into three types (Droog et al., 1995) with type I and type III being similar to the mammalian theta class and type II being similar to the mammalian zeta class (Board et al., 1997). The overall structure of the plant GSTs shows a high degree of structural homology to the structure of animal GSTs. One distinct difference is that whereas mammalian GSTs share a conserved catalytic Tyr residue, it appears that plant, as well as non-plant theta and zeta class GST enzymes lack this conserved Tyr molecule with it likely being replaced by a Ser (Board et al., 1995; Board et al., 1997).

Plant GSTs have recently been reviewed (Marrs, 1996; Droog, 1997; Edwards et al., 2000). GSTs are present at every stage of plant development from early embryogenesis to senescence and in every tissue type examined. GSTs presumably function to protect the cell from oxidative damage by quenching reactive molecules with the addition of GSH. However definitive identification of endogenous GST substrates in plants has been difficult (Edwards and Dixon, 1991; Marrs et al., 1995; Li et al., 1997; Edwards et al., 2000). So although it is clear that GST activity levels frequently increase in response to stimuli that cause oxidative damage, the mechanisms involved in protection are not clear. Molecules that have been conjugated with GSH are efficiently imported into vacuoles via ATP-binding cassette transporters (Martinoia et al., 1993; Rea, 1999). The import of GSH-conjugated compounds into the vacuole acts to limit the effects of GST end-product inhibition and

---

to further protect plant cells from danger by sequestration of those compounds whose conjugation with GSH does not cause detoxification (Rea et al., 1998).

In addition, it seems likely that GSTs may also function as a reversible ligand (Zettl et al., 1994) and it is with this function that they may play a role in hormonal regulation. The Arabidopsis GST Atpm24.1 was shown to bind to the photoaffinity analog of indole-3-acetic acid. However, this enzyme appears to be unable to conjugate GSH to indole-3-acetic acid so it was suggested that the indole-3-acetic acid analog may bind at a second binding site distinct from the active center of the enzyme in a manner similar to that of steroid and porphyrin derivatives (Bhargava et al., 1978; Boyer, 1986). It is as a ligand-binding protein that maize BZ2 has been proposed to interact with cyanadin-3-glucoside (Edwards et al., 2000) rather than catalyzing a GSH conjugation reaction.

In addition to their endogenous functions, GSTs play an important role in xenobiotic degradation (McGonigle et al., 1997; Neuefeind et al., 1997c; Dixon et al., 1998b) and it is in this role that they have been particularly well studied. Several major classes of herbicides including sulfonylureas (chlorimuron ethyl [Brown and Neighbors, 1987], trisulfuron methyl [Wittenbach et al., 1994], and flupyrsulfuron-methyl [Koeppe et al., 1998]), triazines (atrazine), chloroacetanilides (alachlor [Shimabukuro et al., 1971] and metolachlor [Cottingham and Hatzios, 1992]), thiocarbamate sulfoxides (S-ethyl dipropyl-thiocarbamate sulfoxide [Cottingham et al., 1993]), and diphenylethers (flurodifen) are found as a GSH conjugate and this conjugate is (typically) no longer toxic to the target enzyme. More direct proof of the importance of GSTs in protection from xenobiotics is that the expression of maize GSTIV in tobacco provides protection from metolachlor (Jepson et al., 1997).

In addition to the functions that have been ascribed to GSTs they have been well studied because of their notable expression patterns. The maize GSTIV responds to safeners and a variety of herbicidal stresses (Jepson et al., 1994; Holt et al., 1995). In soybean, GH2/4 (also known as Gmhsp26-A) was cloned independently as a heat shock protein (Czarnecka et al., 1988) and an auxin-induced protein (Hagen et al., 1984). Only later was the protein identified as a GST: first on the basis of homology to other cloned GSTs and later by showing that the protein is able to conjugate GSH to the model substrate 1-chloro-2,4-dinitrobenzene (CDNB; Ulmasov et al., 1995). Besides heat shock and auxin, a wide range of chemical agents including abscisic acid, kinetin, gibberellic acid, polyethylene glycol, canavine, KCl, NaF, and heavy metals induce GH2/4 message levels. GSTs in other species have been described that are transcriptionally induced by a range of different environmental stimuli including fungal attacks, dehydration stress, ethylene, and wounding (Marrs, 1996). Fur-

thermore, overexpression of a cDNA encoding a protein with GSTs and GSH peroxidase activity enhances the growth of transgenic tobacco seedlings during chilling and salt stress conditions (Roxas et al., 1997). It is clear that GSTs play an important role in the response of plants to changing environmental conditions.

We have studied the GST multigene family by identifying GST sequences in soybean and maize expressed sequence tag (EST) databases using BLAST searches. Using this strategy we have identified 25 soybean and 42 maize GST sequences that represent the majority of expressed GSTs in these species. The identification of the various members of a multigene family allows a more complete understanding of the functions of that gene family.

RESULTS

Cloning and Distribution of GSTs in Maize and Soybean

Full-length sequences were aligned and classified according to the criteria in Droog et al. (1995). We found 12 maize type I GSTs including the three previously described in the literature (Moore et al., 1986; Shah et al., 1986; Jepson et al., 1994), two maize type II GSTs, and 28 type III GSTs including the four previously described in the literature (Marrs et al., 1995; Dixon et al., 1998a, 1999) for a total of 42 GSTs in maize (Table I; Fig. 1). We found four soybean type I GSTs, one soybean type II GST, and 20 type III GSTs including the four previously described in the literature (Ulmasov et al., 1995; Andrews et al., 1997; McGonigle and O'Keefe, 1998; Skipsey et al., 2000) for a total of 25 GSTs in soybean (Table II; Fig. 2). Although present in the phylogenies shown in Figures 1 and 2, we have not categorized in2-1 and Gmin2-1 in these GST categories.

In naming the new plant GSTs shown here we have adapted the nomenclature scheme of Dixon et al. (1997), and have retained nomenclature for any previously discovered sequences. Thus, new names for maize GSTs begin with ZmGST 8, and with GmGST 5 for soybeans, although alternative systematic nomenclature for the soybean GSTs GH2/4 (renamed GmGST 1) and GSTa (renamed GmGST 4) are suggested in Table II. We avoid the use of Roman numerals as originally proposed (Dixon et al., 1997) for the sake of simplicity. The numbering in individual species is independent. The dimer structure of the proteins under this system is easily indicated, so a homodimer of GmGST 20 can be written GmGST 20/20; a heterodimer with GmGST 40 would be GmGST 20/40 for example.

Because the tissue used to create the EST libraries comes from many different portions of the plant grown under a wide diversity of environmental conditions it is possible to obtain a broad sense of the numbers of each type of GST in each species; that is,

Table I. *GSTs of maize*

Type	Name	NCBI Accession No.
Type I GSTs	GSTI	M16901
	GSTIII	X04455
	GSTIV	X79515
	ZmGST 8	AF244673
	ZmGST 9	AF244674
	ZmGST 10	AF244675
	ZmGST 11	AF244676
	ZmGST 12	AF244677
	ZmGST 13	AF244678
	ZmGST 14	AF244679
	ZmGST 15	AF244680
	ZmGST 16	AF244681
Type II GSTs	ZmGST 17	AF244682
	ZmGST 18	AF244683
Type III GSTs	BZ2	U14599
	GST5	Y12862
	GST6	AJ010439
	GST7	AJ010440
	ZmGST 19	AF244684
	ZmGST 20	AF244685
	ZmGST 21	AF244686
	ZmGST 22	AF244687
	ZmGST 23	AF244688
	ZmGST 24	AF244689
	ZmGST 25	AF244690
	ZmGST 26	AF244691
	ZmGST 27	AF244692
	ZmGST 28	AF244693
	ZmGST 29	AF244694
	ZmGST 30	AF244695
	ZmGST 31	AF244696
	ZmGST 32	AF244697
	ZmGST 33	AF244698
	ZmGST 34	AF244699
	ZmGST 35	AF244700
	ZmGST 36	AF244701
	ZmGST 37	AF244702
	ZmGST 38	AF244703
	ZmGST 39	AF244704
	ZmGST 40	AF244705
	ZmGST 41	AF244706
	ZmGST 42	AF244707

which types of GSTs are most highly expressed. In maize, 54% of individual cDNAs are type I GSTs, 4% of individual cDNAs are type II GSTs, whereas 41% of individual cDNAs are type III GSTs (Fig. 3). Within this distribution certain individual cDNAs are particularly prominent. Eighty percent of maize type I GSTs are made up of a combination of GSTI (42%), GSTIII (13%), and GSTIV (24%), whereas 26% of maize type III GSTs are GST5. It is notable that these most abundant maize GSTs are those that were originally discovered by traditional protein purification techniques. In contrast to maize, 6% of individual cDNAs in soybean are type I GSTs, 2% of individual cDNAs are type II GSTs, whereas 92% of individual cDNAs are type III GSTs (Fig. 3). In the case of the soybean, only one GST sequence is par-

ticularly predominant with 33% of soy type III GSTs being GmGST 8.

Some of the libraries have been sequenced sufficiently such that a distribution of the more abundant GST cDNAs within that library can be compared. In Table III and Table IV we show the results with some of these libraries in addition to the overall leaf versus root distribution found. A select set of GST sequences is found predominantly or only in libraries made from "induced" tissue. In maize this includes GSTIV, GST 7, ZmGST 8, ZmGST 20, and ZmGST 31, all of which are induced by various chemical treatments. In soybean, GmGST 1, GmGST 7, GmGST 15, and GmGST 19 are all induced by infection with the fungus *Sclerotinia sclerotiorum*. Somewhat surprisingly, the mature soybean leaf library contained no detectable GST sequences (Table IV), whereas it should have had >10 if expression was at the same level as the comparable maize leaf (Table III). This result suggests that expression of GSTs is very low in mature soybean leaf. We have no further data about expression in younger leaves. The GSTs identified here are reflective of the types of tissues used to create the EST libraries. For instance, there were over 50,000 ESTs of maize created from chemically treated plants, but no maize ESTs from maize infected with fungi.

DNA microarray data has not been widely used for determining absolute expression levels of individual gene products, but because the GST target and probe sequences are of similar length and G + C content, we have attempted to use microarray analysis to determine relative expression levels of the entire gene family in a specific tissue. In Figure 4A, data are presented as normalized raw values to show relative levels of expression of each GST gene target. At the extremes of this data, GSTI, known to be the most abundant maize GST by a variety of criteria, shows maximum expression; *Bz2*, not predicted to be expressed in these tissues, shows the minimum expression level. The *Bz2* signal, about 2× the blank (control) in absolute signal intensity, is the smallest of all the true target signals and most likely represents the background level of non-specific cross-hybridization. The high GSTI and very low *Bz2* expression are the only independent calibration we have applied to the data in Figure 4A, and we cannot verify the linearity or uniformity of response for the other targets. With that caveat in mind, the highest expression levels in Figure 4A are in general agreement with the frequency of appearance in maize libraries (Table III); i.e. GSTI, GSTIII, GSTIV, and GST5 are all strong signals, although the relative expression level of each is not directly comparable to the library data, as we have no data for etiolated seedlings (the source of the microarray expression data). The data in Figure 4A suggests that many of the previously unknown GSTs are expressed in this tissue.

Expression data has also been presented as ratios (fold induction) of each gene in response to dichlor-

Figure 1. Phylogenetic tree showing relationship of all known maize GST protein sequences. The horizontal scale shows the number of differences per 100 residues derived from the Clustal V alignment. Details of construction are in "Materials and Methods" and NCBI accession numbers are shown in Table I.
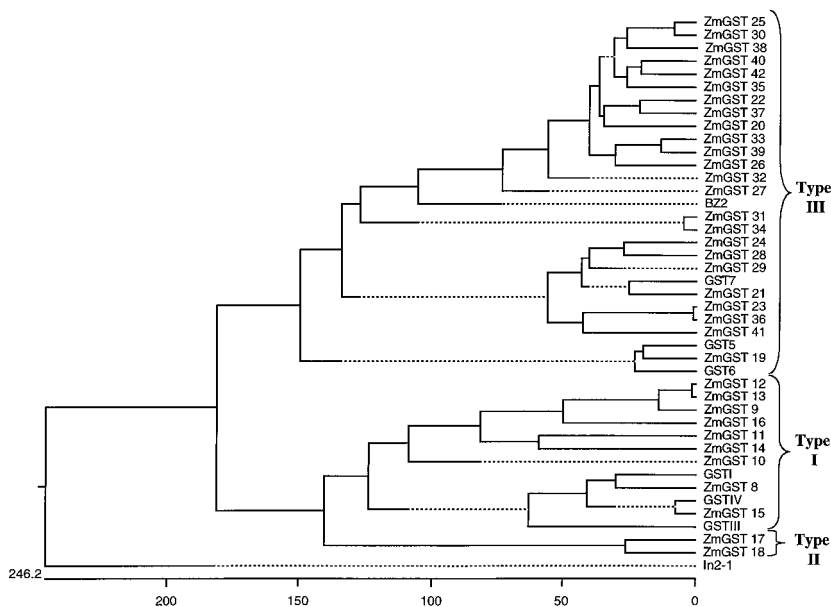
mid (Fig. 4B) or to ethanol (Fig. 4C). Treatment with dichlormid increases the rate of GSH conjugation of xenobiotics (Jepson et al., 1994; Holt et al., 1995; Dixon et al., 1997), and treatment with ethanol, while not necessarily associated with changes in GST activity, has also been used to increase the rate of herbicide metabolism (Frear et al., 1991). Expression level increases as small as 1.3-fold are detectable by microarray analysis (Lockhart and Winzeler, 2000), and we have found that increases need to be larger than 1.5-fold to be reproducible. By this criterion, there are 15 GST genes induced in response to dichlormid treatment and 18 induced by treatment with ethanol. Cross-hybridization between related sequences could result in apparent induction of genes as determined by this method, and we have carried out a detailed evaluation to determine the extent of this problem.

It is possible to develop general guidelines for how cross-hybridization of closely related sequences effects the interpretation of expression data in Figures 4B and 4C, first by establishing that the strongest signals are independent, and then by comparing the responses of the most highly related sequences. In Figure 4B, the two strongest responding signals, GST7 and ZmGST 31, are only 45.0% similar to each other. There are several examples of sequences on this array, which are >45.0% related to GST7 (ZmGST 23 and ZmGST 36) or ZmGST 31 (ZmGST 22 and ZmGST 29), but do not appear to be significantly induced by this treatment. This suggests that GST7 and ZmGST 23 respond independently to dichlormid treatment. In Figure 4C, the two strongest signals are from 36.7% similar sequences (GST7 and ZmGST 30), and because of this low sequence similarity and the presence of more closely related sequences that do not respond to ethanol treatment, we conclude that they are also responding independently. Table V shows the 11 most

closely related maize GST nucleotide sequences arranged in order of decreasing similarity. The closest relatives, the ZmGST 23 and ZmGST 36, undergo very similar responses in Figures 4B and 4C, as expected since cross-hybridization should occur. In the next closest related pairs, ZmGST 25 and ZmGST 30, ZmGST 30 is the strongest responding signal to ethanol treatment (Fig. 4C) and ZmGST 25 also exhibits

Table II. *GSTs of soybean*

| Type	Name	NCBL Accession No.
Type I GSTs	GmGST 21	AF243376
	GmGST 22	AF243377
	GmGST 23	AF243378
	GmGST 24	AF243379
Type II GSTs	GmGST 25	AF243380
Type III GSTs	GH2/4 (Gmhsp26-A or GmGST 1)	J03197, M20363
	GmGST 2	Y10820
	GmGST 3	X68819
	GST a (GmGST 4)	AF048978
	GmGST 5	AF243360
	GmGST 6	AF243361
	GmGST 7	AF243362
	GmGST 8	AF243363
	GmGST 9	AF243364
	GmGST 10	AF243365
	GmGST 11	AF243366
	GmGST 12	AF243367
	GmGST 13	AF243368
	GmGST 14	AF243369
	GmGST 15	AF243370
	GmGST 16	AF243371
	GmGST 17	AF243372
	GmGST 18	AF243373
	GmGST 19	AF243374
	GmGST 20	AF243375
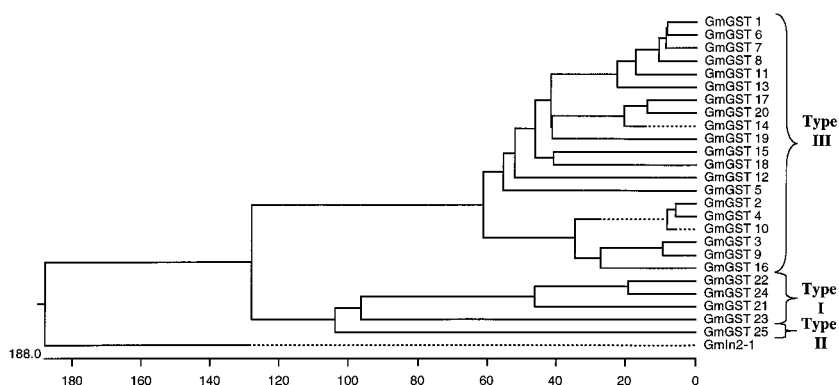
Figure 2. Phylogenetic tree showing relationship of all known soybean GST protein sequences. The horizontal scale shows the number of differences per 100 residues derived from the Clustal V alignment. Details of construction are in "Materials and Methods" and NCBI accession numbers are shown in Table II.

a substantial response. This is consistent with the induction of the stronger signal (ZmGST 30) leading to cross-hybridization with the ZmGST 25 target and an "apparent" induction. In the third most closely related pair, ZmGST 34 and ZmGST 31, one member (ZmGST 31) is strongly induced by dichlormid (Fig. 4B). In this case the signal from ZmGST 34 is also >2.5× increased, and is also consistent with the induction of the stronger signal ZmGST 30 leading to cross-hybridization with the ZmGST 34 target and an "apparent" induction. In the next three entries in Table V, starting with ZmGST 12 and ZmGST 13, the signals in Figures 4B and 4C exhibit behavior that suggests they are responding independently; i.e. in either dichlormid or ethanol treatment, the two targets respond in opposite directions with one response >1.0 and one response <1.0. Although this analysis shows that the responses of target sequences being 78.9% similar are consistent with cross-hybridization, it is also possible that both genes in the highly related pairs are being induced. The analysis clearly implies that as the sequence similarity drops below 78.9%, cross-hybridization among the related GST targets does not significantly interfere with their individual responses. Similar results have been obtained with other sets of related genes where the targets respond independently when the sequence similarity drops below approximately 80% (data not shown).

Based on the above considerations, it is possible to interpret the data in Figure 4 with precautions to carefully evaluate the six targets, which may be cross hybridizing (ZmGST 23, ZmGST 36, ZmGST 25, ZmGST 30, ZmGST 34, and ZmGST 31). We conclude from Figure 4B that dichlormid most strongly induces ZmGST 31 and GST7, but the response of ZmGST 34 may be artificially elevated. Ethanol treatment results in strongest induction of ZmGST 30, GST5, and GST7, and the response of ZmGST 25 may be artificially elevated. Other GST genes appear to be induced to a lesser extent by these treatments. The majority of induced genes are type III GSTs. This induction pattern suggests a general mechanism for increasing GST activity not dependent on specific GST gene products. A generalized mechanism for the

induction of a group of GSTs supports the development of metabolism-based resistance to multiple herbicides (Cummings et al., 1999). Many of the Type I GSTs appear to be down-regulated in these experiments, particularly under ethanol treatment. We have not investigated this finding any further.

Previous reports have shown that the gene for GSTIV is strongly transcriptionally induced by the herbicide safener dichlormid (Holt et al., 1995; Jepson et al., 1994). However, our microarray analysis does not show an induction of GSTIV by dichlormid. The apparent contradiction has several explanations. Jepson et al. (1994) showed that the GSTIV message was present in treated and untreated roots and was induced only in the aerial portion of light grown plants. Because the microarrays were probed with mRNA extracted from the entire seedlings of etiolated plants, the basal level will be significant due to the presence of roots. It has not been reported whether or not GSTIV is inducible in etiolated plants. In addition, Jepson et al. (1994) show that at 4 h there
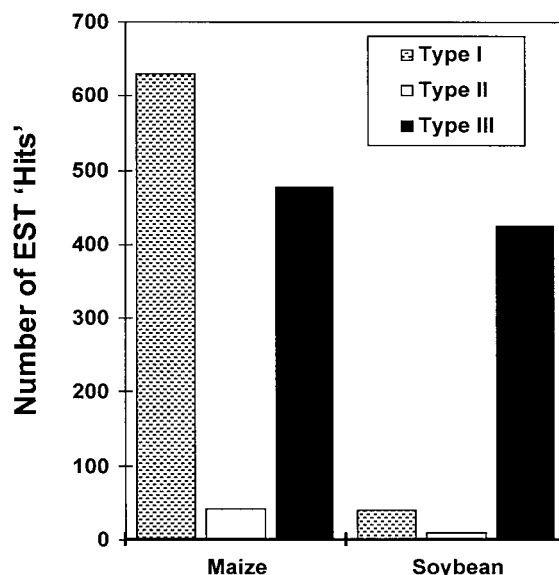
Figure 3. A bar graph showing the distribution of type I, type II, and type III GSTs within maize and soybean.

Table III. *Distribution of sequences in selected maize libraries*

Description of Library	Total No. of Clones Sequenced	GSTI	GSTIII	GSTIV	GST 5	Other GSTs (no. of occurrences)
Corn leaf and sheath from 5 weeks plant	4,433	11	0	0	0	ZmGST 20 (1)
						ZmGST 9 (2)
Corn root from 7-d-old seedlings	2,847	7	3	5	2	ZmGST 9 (1)
						ZmGST 11 (2)
						ZmGST 21 (2)
Corn silk	3,391	9	5	0	1	
BMS cells	1,047	1	0	0	0	ZmGST 8 (1)
BMS cells treated with chemicals related to membrane ionic force	3,609	8	0	2	7	ZmGST 20 (1)
						ZmGST 18 (1)
						ZmGST 8 (3)
						ZmGST 22 (1)
						ZmGST 23 (4)
						Gst 7 (8)
						ZmGST 36 (1)
All corn libraries	350,000	285	165	89	150	38 Others (597)

is only minimal induction; thus our 5-h time point may precede significant induction. Finally, there is likely to be significant cultivar differences in GST induction. For example, Dixon et al. (1997) show that in one maize cultivar dichlormid treatment caused an increase in activity toward several herbicides, but in another cultivar the increase in activity was only noticed for a subset of those herbicides.

Structural Features of GSTs

Sequence comparisons of the 66 sequences from maize and soybean revealed some notable features (Fig. 5). We have excluded GmGST21 from this anal-

ysis because it is clearly truncated at the 5′ end. With this and the in2-1 and Gmin2-1 as exceptions, there are three amino acids that are absolutely conserved in all remaining sequences, and these are Ser-12, Arg-17, and Ser-68, using the GSTI numbering (these amino acids are indicated in the consensus sequences in Fig. 5). The first of these is the active site Ser (Ser-12 in maize GSTI). This amino acid is thought to activate the GSH sulfhydryl by lowering its pKa. In most animal GSTs this function is carried out by a Tyr residue, but in the theta class of GSTs and, so far in all plant GSTs this residue is a Ser (Board et al., 1995). The two in2-1 sequences have a Cys and not a

Table IV. *Distribution of sequences in selected soybean libraries*

Description of Library	Total No. of Clones Sequenced	GmGST 8	GmGST 1	Other GSTs (no. of occurrences)
Soybean root	6,804	15	0	GmGST 5 (1)
				GmGST 12 (1)
				GmGST 13 (1)
Soybean mature leaf	3,276	0	0	None
Soybean (cv Manta) infected with *S. sclerotiorum* mycelium	12,272	4	46	GmGST 4 (3)
				GmGST 2 (2)
				GmGST 7 (8)
				GmGST 13 (2)
				GmGST 14 (2)
				GmGST 23 (2)
				GmGST 15 (8)
				GmGST 18 (1)
				GmGST 22 (1)
				GmGST 16 (1)
				GmGST 24 (1)
Soybean immature flower	1,035	9	0	GmGST 11 (2)
				GmGST 25 (1)
				GmGST 22 (2)
				GmGST 16 (1)
				GmGST 9 (1)
				GmGST 2 (1)
All soybean libraries	150,000	144	109	24 Others (230)

Figure 4. Expression of maize GST genes by DNA microarray analysis. The names of the new GSTs presented in this study have been abbreviated so that only the numerical designation is shown (e.g. ZmGST 13 is 13). A, Relative intensity values for each gene spot after hybridization to Cy3 labeled cDNA probe prepared from mRNA isolated from untreated etiolated maize seedlings. B, Fold induction of GST genes due to treatment of etiolated maize seedlings with 5 μL L^{-1} dichlormid. Fold induction equals the intensity value for each gene target with a Cy5 labeled cDNA probe derived from treated tissue divided by the intensity value for each gene target with a Cy3-labeled cDNA probe derived from untreated tissue. C, Fold induction of GST genes due to treatment of etiolated maize seedlings with 10% (v/v) ethanol.

Ser in this position, although the soybean Gmin2-1 has an adjacent Ser in the alignment, and both have a Tyr two residues before the Cys. The other conserved Arg and Ser are present in these sequences.

In the available crystal structure data for an Arabidopsis GST, maize GSTI and maize GSTIII, Glu-67, Ser-68, and Arg-69 (GSTI numbering) are all involved in binding charged groups on the γ-glutamyl portion of the substrate GSH, as part of the "G-site" (Reinemer et al., 1996; Neuefeind et al., 1997a, 1997b). The complete conservation of Ser-68 in all maize and soybean GSTs is consistent with a critical role in substrate binding. Although Arg-17 does appear to be part of the G-site in GSTI and it is a nearest neighbor of Ser-68 and Arg-69 (Neuefeind et al., 1997a), a specific role of Arg-17 in binding ligand has not been established. The G-site Glu-Ser-Arg trio provides a notable distinction between type I and type III GSTs. In type I, the first position Glu can have the conservative substitution Asp, (although in one case, ZmGST 8, it is Asn), and Ser-Arg is found in all of the

type I sequences. In the type III the first two residues are always Glu-Ser, and the Arg in the third position is rarely present (two out of 46), usually replaced by Leu. We speculate that in type III GSTs the totally conserved Arg-17 may substitute for the missing Arg in the G-site trio.

There is also a Trp residue (Trp-98 in GSTI or Trp-115 or Trp-114 in the consensus sequences in Fig. 5, A and C, respectively) that is conserved in all of the type I and type III GSTs, but is missing in the type II GSTs. This Trp is located in the region of the GSTI structure that forms the interface between the two subunits of the dimer, but it is not in close enough contact to contribute to the hydrophobic interactions between the two subunits. The role of this amino acid remains unclear.

One feature that is readily apparent from a sequence comparison of the maize and soybean GSTs is shown in Figure 5C. When all of the type III GSTs are compared, it is evident that there are four distinct segments of homology that are a strong feature of the

Table V. *Maize GSTs with cDNA sequences ≥ 60% similar*

GST Comparison	Nucleotide Similarity
	%
ZmGST 23–ZmGST 36	97.5
ZmGST 25–ZmGST 30	81.0
ZmGST 34–ZmGST 31	78.9
ZmGST 12–ZmGST 13	76.8
GST IV–ZmGST 15	68.1
ZmGST 22–ZmGST 37	67.0
GST7–ZmGST 28	62.8
GST7–ZmGST 21	61.1
ZmGST 14–ZmGST 11	60.5
ZmGST 25–ZmGST 38	60.2
GST7–ZmGST 24	60.0

type III GSTs. The four strongly conserved segments consist of S20-E38, K49-H68, E76-E86, and L101-W114. By contrast, the type I GSTs have several strongly conserved residues, but they are more widely distributed around the protein. To illustrate where these regions fall in a known GST structure, in Figure 6 we show a color-coded representation of the GSTI dimer structure. Because GSTI is not a type III GST, we also show a computed homology model for a monomer of its nearest relative in the type III type, ZmGST 24. The regions of type III strong homology correspond to distinct structural features in the model. The first of these is an α helix that begins with the active site Ser and ends with a turn and beginning of a β sheet. The second is the latter one-half of a 3_{10} helical segment, followed by a sharp turn (a Pro in this position is conserved in all type III and in the majority of other sequences) and another β sheet strand. This region contains a flexible loop that is thought to be important in induced substrate fit in the active site (Neuefeind et al., 1997a). The third and fourth regions are two antiparallel α helices that appear to be arranged in a four-helix bundle with their counterparts on the other subunit of the dimer. The poorly conserved sequence between these segments is the linker segment between the N-terminal domain and the helix-rich C-terminal domain.

The Tested GSTs Show a Range of Broad and Overlapping Specificities

We have expressed a partial set of these plant GSTs in *Escherichia coli*, and recovered purified enzyme for functional analysis. Table VI shows the results of an activity screening using seven different substrates and 27 of the GSTs. These substrates were chosen because three of them represent structurally diverse classes of herbicides for which GSH conjugation is important for crop safety, and the remaining four are common model GST substrates representing a variety of GSH conjugation reactions. For the most part, the proteins in Table VI were expressed as His-tagged fusion proteins, which are simply purified by a single affinity chromatography step on a Ni col-

umn. We initially attempted to purify some proteins using GSH affinity resins, however it was necessary to experimentally determine the appropriate resin for each protein, and some of the proteins did not purify on any of these resins. For GSTI we have compared the activity with the His-tagged protein and the natural sequence (purified by GSH affinity column). Table VI shows that although there is somewhat lower activity, the His-tagged GSTI still retains the general pattern of activity with the various substrates as the natural GSTI.

The activities in Table VI demonstrate that the majority of these proteins are active as GSTs when cloned and expressed. In general these proteins can be categorized according to their CDNB activity. If this value is >50 nmol min^{-1} mg^{-1}, the enzymes also exhibit substantial activity with most of the other substrates. Enzymes with CDNB activity < 50 nmol min^{-1} mg^{-1} are more difficult to categorize. Activity this low, especially if they are also nearly inactive with other substrates, suggests that some of the enzymes are non-functional as GSTs. For GmGST 21 we have already discussed that the truncation at or near the active site Ser should result in lack of GST activity. This also provides a reference for what a non-functional enzyme would look like in this survey. For the maize in2-1 and soybean Gmin2-1, we previously discussed the lack of clear evidence for the appropriate active site Ser residue. We have expressed these proteins with and without His-tag, and although they bind to GSH affinity resin, we conclude from the lack of activity that they are not GSTs, although they exhibit significant sequence similarity to this class of proteins. BZ2 is another example of an enzyme in this low activity category. Recent reports (Edwards et al., 2000) suggest that the physiological role of this protein does not involve the catalysis of a GSH substitution reaction, so the lack of any activity in these assays is consistent with that hypothesis. For the remainder of the very low activity proteins in Table VI (ZmGST 8, ZmGST 9, ZmGST 10, GSTIV, ZmGST 17, GmGST 11, GmGST 21, and GmGST 25), there are a number of possible explanations; these include intrinsically low activity, substrate specificity outside the range of compounds tested in this experiment, negative effects of the His-tag construction, problems with solubility, or an absolute requirement for a different partner in a heterodimer.

DISCUSSION

We have characterized 42 distinct GSTs from maize and 25 distinct GSTs from soybean. With only one exception, each GST sequence is less than 95% similar to the other GST sequences in the same species, suggesting that each GST represents a different gene and not alleles of the same gene (Clegg et al., 1997). ZmGST 23 and ZmGST 36 are 97.5% identical at the nucleotide level. By this somewhat arbitrary criterion
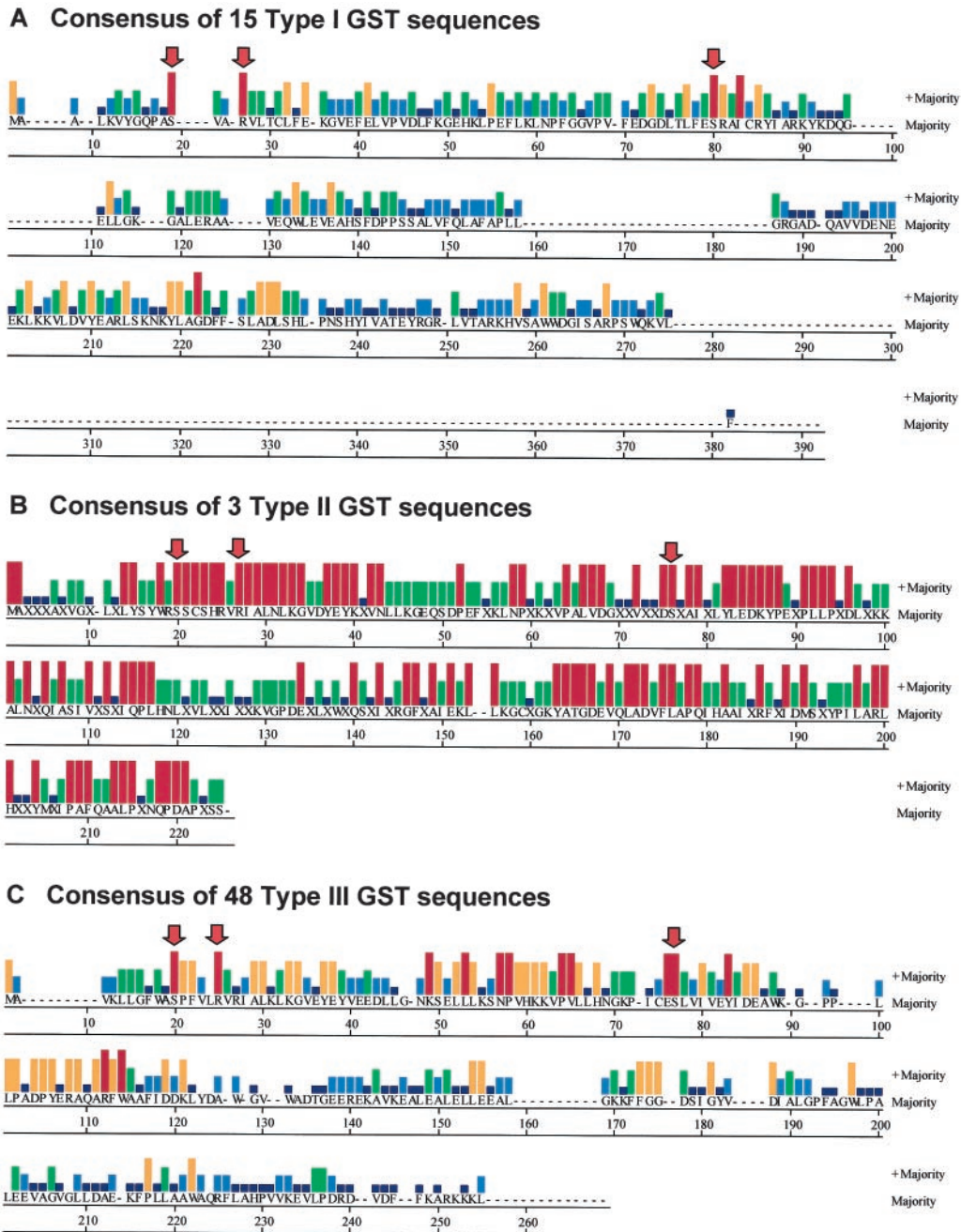
Figure 5. The consensus sequence report of all known maize and soybean GSTs. The amino acid sequences in each of the three classes of GSTs were aligned using the Clustal V method in the Megalign module of DNASTAR. Only the consensus sequence is shown. The histograms show the relative abundance of each amino acid in the group of sequences according to the consensus strength score that is a whole number from 0 to 5 (0, no bar; 1, dark blue; 2, light blue; 3, green; 4, orange; and 5, red). The red arrows indicate the three amino acids that are totally conserved in all of the GSTs reported here. A, The consensus report from all known maize and soybean type I GSTs. Note that a single sequence ZmGST 16 at a length of 299 amino acids is substantially longer than all of the other GSTs and accounts for the long stretch of unique sequence at the C terminus of the consensus. B, The consensus report from all known maize and soybean type II GSTs. Note that there are only three members of this class and they are quite closely related. C, The consensus report from all known maize and soybean type III GSTs.

they could be alleles of the same gene, although this will not be certain until the genes are mapped or the maize genome is sequenced. We inadvertently included them both in this study and have included the

results for an example of a very similar pair of sequences. Of interest is whether we have found all GSTs encoded by the soybean and maize genomes, respectively. It is clear that we have cloned enough

different members of the GST gene family to account for the many bands that have been found during protein purification (e.g. Fuerst et al., 1993; Dixon et al., 1997). However, the definitive answer this question awaits the completion of the genomic sequence of these species. We have surveyed a large number of ESTs from libraries constructed from many different types of tissue and each clone that we have analyzed is represented by between four and 225 ESTs, and therefore, we expect that we have, if not all, the large majority of GST sequences from soybean and maize. To our knowledge this study is the first to attempt to catalog all GSTs from a given plant species. Many of the GSTs known in the literature have been found because of the investigator's interest in specific stresses (e.g. dehydration or cold), not in studies of GSTs. Those investigators who did focus on GSTs typically focused on those involved with xenobiotic metabolism. We have attempted to take a more comprehensive approach, although still with an emphasis on xenobiotic metabolism; however, other general aspects of plant GST sequence, structure, and activity have become apparent in this study.

The numbers of GST genes in a given species may be a complex process reflective of large scale DNA duplication events and the evolutionary pressures existing for a given species. We show that maize, a monocot, has somewhat similar numbers of type I and type III genes and soybean, a dicot, has considerably more type III genes than type I genes. However, this may not reflect a monocot/dicot difference. Arabidopsis, a dicot, has been used extensively as a model system and a large, public EST database exists. In Arabidopsis the majority of (published) GSTs are type I (Marrs, 1996). Wheat, a hexaploid with a very complex set of GSTs, seems to have somewhat similar numbers of type I and type III GSTs (B. McGonigle and D.P. O'Keefe, unpublished data). Thus the distribution between type I and type III GSTs in a given species is difficult to predict and further characterizations of the complete set of GSTs from various species will be necessary before generalizations can be made about the differences between gene types in monocots and dicots.

Type II GSTs seem to be relatively poorly represented in a variety of individual genes (two in maize and one in soybean) and in absolute expression levels (only 2%–4% of individual cDNAs representing type II GSTs). This low-level expression was confirmed by microarray analysis. Among plant species, type II GSTs have previously only been cloned from carnation (Meyer et al., 1991; Itzhaki and Woodson, 1993). The limited number of soybean and maize type II sequences found in this study could result from either the scarcity of "target" sequences for BLAST searches or the absolute scarcity of these genes and
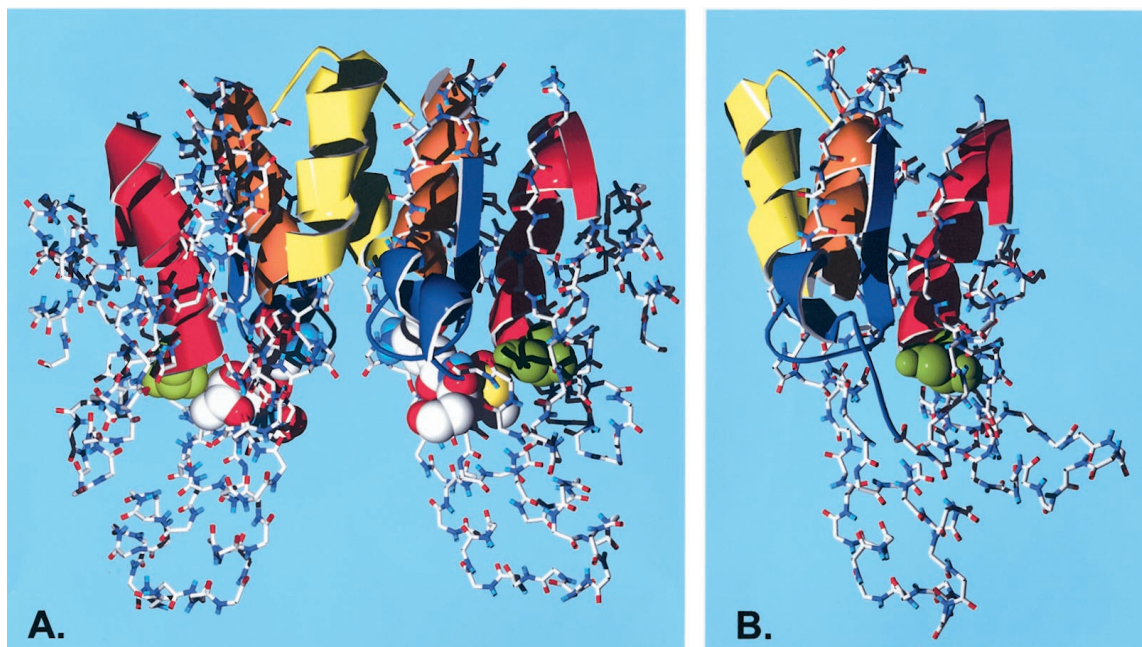
Figure 6. Three-dimensional structure of plant GSTs with the strongly conserved type III features mapped. The active site Ser is shown in green as a space-filled model. The conserved patches in the Type III consensus sequence are shown as ribbons and colored as red, S20-L38; blue, K49-H68; orange, E76-E86; and yellow, L101-W114. A, The lactoylglutathione complex of a GSTI dimer taken from Neuefeind et al. (1997a). The substrate analog is shown as a space-filled model using Corey, Pauling, and Koltun colors. The regions of GSTI that are homologous to the type III conserved patches are S11-E29 (red), K41-N58 (blue), E66-R76 (orange), and R84-W98 (yellow). B, A homology model of ZmGST 24 prepared as described in the text. The monomer is shown in the same orientation as the GSTI dimer. The conserved patches in the ZmGST 24 sequence are S11-E29 (red), K38-H57 (blue), E64-E74 (orange), and L85-W98 (yellow).

Table VI. *Activity of the GST enzymes with seven substrates, assayed as described in the text*

GST	GST Type	6-His Tag[a]	Activity						
			Chlorimuron ethyl	Alachlor	Atrazine	CDNB	Ethacrynic acid	trans-Stilbene oxide	1,2-Epoxy-3-(p nitrophenoxy) propane
			$nmol\ min^{-1}\ mg^{-1}$						
ZmGST 20	III	Yes	0.1	8	0.02	1,348	20	1.25	43
GSTV	III	Yes	0.4	18	0.01	3,939	102	0.01	30
ZmGST 19	III	Yes	1.9	27	0.08	2,136	117	0.02	14
BZ2	III	Yes	0.2	0	0.00	15	23	0.05	0
ZmGST 10	I	Yes	0.1	0	0.00	15	5	0.00	0
ZmGST 9	I	Yes	0.1	0	0.00	30	9	0.00	0
ZmGST 8	I	Yes	0.2	0	0.00	15	13	0.00	0
GSTI	I	No	0.4	77	0.60	46,485	32	0.98	92
GSTI-his tag	I	Yes	0.3	40	0.75	12,879	19	0.90	68
GSTIII	I	No	0.3	3	0.05	1,803	1	0.31	28
GSTIV	I	Yes	0.3	1	0.00	15	13	0.00	0
ZmGST 17	II	Yes	0.1	0	0.00	45	17	0.00	1
in2-1		No	0.0	0		15			
GmGST 6	III	Yes	0.1	1	0.19	2,364	13	0.06	1
GH2/4	III	No	0.5	104	0.13	6,030	8	7.93	33
GmGST 7	III	Yes	0.2	10	1.40	515	17	4.04	12
GmGST 8	III	Yes	0.3	111	0.46	2,545	14	0.12	10
GmGST 11	III	Yes	0.1	0	0.00	45	9	0.00	1
GmGST 13	III	Yes	0.1	4	0.03	1,394	13	0.49	19
GmGST 12	III	Yes	0.1	7	0.03	470	14	0.02	47
GmGST 4	III	No	0.5	71	0.03	1,924	109	0.06	22
GmGST 10	III	Yes	1.4	166	0.00	2,030	11	0.06	4
GmGST 2	III	Yes	0.5	8	0.76	1,379	4	0.07	9
GmGST 9	III	Yes	0.9	30	0.00	2,576	68	0.16	10
GmGST 5	III	Yes	4.4	168		14,364	1	0.07	20
GmGST 21	I	Yes	0.1	0	0.00	15	11	0.00	0
GmGST 25	II	Yes	0.0	0	0.00	15	5	0.04	2
Gmin2-1		Yes	0.0	0	0.00	30	3	0.15	0

[a] Proteins with His tags were purified on nickel columns as described in the text. Proteins that were not His-tagged were purified on glutathione-agarose affinity columns.

messages. Significant amino acid sequence homology exists with type III GSTs, particularly in the GSH-binding domain, so the scarcity of BLAST target sequences is probably not a significant factor. It seems more likely that there are very few type II GST genes and they are not highly expressed. It should, however, be noted that no libraries of ethylene-treated tissue were sampled, and as the carnation genes are ethylene inducible, it may be that an increase in the number of genes or levels of expression would be found in this type of library. It has recently been shown that humans and *Caenorhabditis elegans* have a GST similar to the plant type II GSTs known as zeta class GSTs and the high degree of conservation observed between these species over a long evolutionary period suggests a common biological role, yet unknown, for type II GSTs in many living cells (Board et al., 1997).

It is possible to use the data in Table VI to examine the role of specific enzymes in physiological functions, herbicide detoxification for example. GSTI, known to be a major GST component in many maize tissues, has the highest activity of all the maize proteins against alachlor. This confirms that it plays a major role in alachlor detoxification in maize. On the other hand, GmGST 5 has the highest activity of all the enzymes with chlorimuron ethyl (known to be detoxified by a GST in soybeans). However, this protein has not been identified in libraries other than from soybean embryo, which suggests it is not a major contributor to soybean tolerance to chlorimuron ethyl. The most widely distributed soybean GST, GmGST 8, has less than one-tenth the activity of GmGST5 against chlorimuron ethyl. This suggests that the most active enzymes for a particular xenobiotic substrate may not be the one that is primarily responsible for metabolism; abundance is a very important feature as well.

For structure/function analysis, the group of type III soybean GSTs, GmGST 4, GmGST 10, and GmGST 2 provide an interesting comparison of activities within a very narrow range of primary structure. The amino acid sequence of these three proteins differ by no more than 11.5%, yet their activity varies dramatically with substrate. A different GST from this group has the highest activity with each of the substrates, alachlor, atrazine, and ethacrynic acid. The difference between most active and least active is at least 20-fold. This group of proteins may provide a unique opportunity for further detailed comparison of how

specific amino acids contribute to GST substrate specificity.

GST function does not correspond to the classifications of Droog et al. (1995), which are based on sequence identity. The maize *bz2* (a type III) GST is functionally complemented by the petunia GST an9 and maize GSTIII (both type I GSTs) and GH2/4 (a soybean type III GST), but not by several other GSTs of type I and III (Alfenito et al., 1998). The function of a given GST cannot be predicted from the primary structure and must be experimentally determined. We found that among substrates susceptible to GSH conjugation the specificity of the individual GSTs in this collection is quite broad. This suggests that GST-mediated herbicide tolerance may not be a function of a single GST gene, but instead is reflective of the expression characteristics and the functionality of all of the GSTs present in a given species. In agreement with this observation we also observed induction of multiple GST genes with the safener dichlormid and ethanol. We propose that a combination of expression and activity studies similar to those performed here, but comparing crop and weed species, will form the basis of a very powerful technique for evaluation of the role of individual GSTs in the metabolism of new xenobiotics.

It should also be noted that all of the functional assays we carried out measured the ability of homodimers to perform certain reactions. It is known that at least some native GSTs are composed of heterodimers. The strong segmented homology corresponding to the backbone of the dimer structure (Fig. 6) in the type III GSTs may result in a higher capability to form heterodimers among the type III GSTs. This is supported by the in vitro dimerization of recombinant ZmGST 6 and ZmGST 7 (Dixon et al., 1999). However, it is unclear that a heterodimer would have different specificities than its component subunits as each of the two active sites per dimer is formed exclusively by each subunit, and there is little evidence to support subunit cooperativity (positive or negative) among plant GSTs. Substrate cooperativity could be the basis of a vast range of possible specificity since the number of possible heterodimers is so high. In addition to the 42 homodimers, maize has 861 possible combinations of heterodimers (soybean has 25 homodimers and 300 possible heterodimers). Many of the monomers may never occur in the same tissue or at the same time and microarray expression analysis can be used to narrow the scope of possibilities. Still, there exists a large number of possibilities for unique specificities that would not be uncovered using our current approach. In addition, the functional assays we carried out used the thiol GSH. In soybean, the majority of the free thiol is homoglutathione (Klapheck, 1988) except in nodules where GSH is also present (Matamoros et al., 1999). Studies for two soybean GSTs have shown that the enzymes discriminate between GSH and hGSH de-

pending upon the second substrate (McGonigle et al., 1998; Skipsey et al., 1997). The complex 2-substrate kinetics precluded any further analysis of homoglutathione effects in this study.

It is of interest to question why there are so many different GSTs in a single species considering that their substrate specificities seem to be overlapping. The number of GSTs simply may be reflective of the process of homologous recombination and exon shuffling that leads to gene diversity and not the need for specific GSTs to fulfill specific roles. On the other hand, we have very limited knowledge of endogenous GST substrates and it may be that these compounds exhibit distinctly higher specific activities with a given GST. Plants are typically sessile and committed throughout their life span to a particular location. A large variety of GSTs may be important to be able to deal with changing environmental conditions, including allochemicals and xenobiotics. From a practical standpoint, the diversity of GSTs and the differences in specific forms and expression levels between different species has enabled the development of many selective herbicides, which are widely used in modern agriculture.

MATERIALS AND METHODS

Construction of cDNA Libraries and Identification of GST Clones

cDNA libraries were constructed using standard methods (Sambrook et al., 1989) typically using the lambda zap II kit (Stratagene, La Jolla, CA). mRNA representing a variety of tissue types was isolated from maize (*Zea mays*) and soybean (*Glycine max*) grown under various conditions. Libraries were converted into plasmid libraries according to the protocol provided by Stratagene. cDNA inserts from randomly picked bacterial colonies containing recombinant pBluescript plasmids were amplified via PCR using primers specific for vector sequences flanking the inserted cDNA sequences, or plasmid DNA was purified from randomly selected colonies using R.E.A.L. Prep 96 System (Qiagen, Valencia, CA). Amplified insert DNAs or plasmid DNAs were sequenced in either dye-primer sequencing or dye terminator reactions to generate partial cDNA sequences (ESTs; Adams et al., 1991). The resulting ESTs were analyzed using a fluorescent sequencer (Model 377, Perkin Elmer, Norwalk, CT). Three hundred to 10,000 clones were sequenced per library. Over 215 maize cDNA libraries and 62 soybean cDNA libraries were sampled. It was typical that libraries continued to be sampled at least until the percentage of novel genes was less then 30%. A subset of the libraries was also normalized prior to sequencing using the techniques of Bonaldo et al. (1996). Over 350,000 ESTs of maize and over 150,000 ESTs of soybean were created.

All sequences were used to query the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/Entrez/) using the BLAST program. An approximate 85% of the sequences were analyzed

using BLASTX and BLASTN (Altschul et al., 1990), and 15% of the sequences were analyzed using Gapped BLASTX and Gapped BLASTN (Altschul et al., 1997). All queries that were returned using as subjects the words "glutathione" and "transferase" with a score greater then 100 and a plog greater then 4 were examined. Over 1,357 maize sequences and 586 soybean sequences were examined. Each sequence was aligned using Megalign (DNAS-TAR, Madison, WI) and an individual sequence representing the most complete sequence from each contig was chosen for further sequencing. Plasmid DNA was purified using QIAFilter cartridges (Qiagen) according to the manufacturer's instructions. Additional sequence was generated on an ABI (Sunnyvale, CA) automatic sequencer using dye terminator technology using a combination of vector and insert-specific primers. Sequence editing was performed using Seqman (DNASTAR). All sequences presented here represent coverage at least two times in both directions. Upon further sequencing some clones clearly represented chimeric cDNAs or unspliced messages and these clones were not analyzed further.

Sequences were conceptually translated using Seqman (DNASTAR), the protein sequences were aligned, and phylogenetic trees were created using Megalign. The alignment was created using the Clustal V algorithm (Higgins and Sharp, 1989) set to default parameters, and the resulting alignment was then used by the Saitou and Nei (1987) algorithm to generate a phylogenetic tree. To obtain phylogenetic trees with strong alignments, a sequence representing an outgroup was included in the analysis. For the alignment of the maize sequences, in2-1 (NCBI accession no. X58573; Hershey and Stoner, 1991) was used. For the alignment of the soybean sequences, we cloned a soybean homolog of in2-1 and deposited the sequence into the NCBI database (accession no. AF249913).

Preparation and Analysis of DNA Arrays

Amplification of GST gene sequences was performed using 2 ng of plasmid as template, primers flanking the insert, and 20% GCmelt (CLONTECH, Palo Alto, CA) along with the PCR amplification mixture per manufacturer's protocol. PCR products were purified using Qiaquick 96 well vacuum purification system (Qiagen). Eluted PCR products were dried under vacuum and dissolved in 20 μL of 6 M NaSCN. DNA solutions were spotted on silanized microscope slides (Molecular Dynamics, Sunnyvale, CA) using Molecular Dynamic GenII arrayer under >40% humidity. Each slide had eight spots of each of the GST genes as targets for hybridization.

For production of probes, total RNA was extracted using TRIzol reagent (Life Technologies, Rockville, MD) from 1 to 2 g of 4-d-old etiolated maize seedlings (B73 inbred, Illinois Foundation Seed, Champagne, IL; or Pioneer hybrid 3394, Pioneer Hi-Bred International, Johnston, IA), with the kernel removed, treated with water or 5 μL L^{-1} dichlormid in water for 5 h, or 10% (v/v) ethanol for 3 h. Poly(A)$^+$ RNA was purified from total RNA using Pharmacia QuickPrep mRNA purification kit (Amersham Phar-

macia Biotech, Arlington Heights, IL) The preparation was quantitated by UV absorption. Probe labeling was performed by combining 1 μg of purified poly(A)$^+$ RNA with 0.25 μg/μL anchored oligo-(dT)$_{25}$ primer; 1$\times$ Superscript II reaction buffer (Life Technologies); 0.01 M dithiothreitol; 0.05 mM dATP, dGTP, or dTTP; 0.025 mM dCTP; and 0.025 mM fluorescent Cy3 or Cy5-dCTP in a final volume of 19 μL. Two hundred units of Superscript reverse transcriptase (Life Technologies) was added and the reaction incubated at 42°C for 2 h. The reaction was terminated by heating in boiling water for 3 min. The RNA strands were denatured by adding 1 μL of 5 M NaOH and incubating at 37°C for 15 min, and the reaction was neutralized by adding 1 μL of 5 M HCl and 5 μL of 1 M Tris [tris(hydroxymethyl)-amino-methane]-HCl. The cDNA was purified using QiaQuick PCR purification kit (Qiagen), vacuum dried, and resuspended in hybridization buffer. Cy3 label was used for control (0 time, untreated) samples. Cy5 label was used for variables (treated, 3- or 5-h time point).

Prior to prehybridization, spotted slides were submerged in boiling deionized water for 5 min. The slides were subsequently incubated in prehybridization buffer (3.5$\times$ SSC, 0.2% [w/v] SDS, and 1% [w/v] bovine serum albumin) at 60°C for 20 min, rinsed in deionized water and then in isopropanol at room temperature, and dried. The purified probe was dissolved in hybridization solution (5$\times$ SSC, 100 μg/mL sheared salmon sperm DNA [Life Technologies], 0.1% [w/v] SDS, 100 μg/mL oligo[dA]$_{80}$ [Operon Technologies, Alameda, CA], and 50% [v/v] deionized formamide [Sigma, St. Louis]) and denatured at 95°C for 3 min. Hybridization under a coverslip was at 42°C overnight with high humidity. After hybridization, slides were washed with 2$\times$ SSC and 0.1% (w/v) SDS once at 37°C for 5 min, 0.1$\times$ SSC and 0.1% (w/v) SDS once at 37°C for 5 min, and then three times at room temperature in 0.1$\times$ SSC for 1 min each. Slides were dried with compressed nitrogen gas.

Slides were scanned with a confocal laser scanner (Molecular Dynamics) at 532 nm with a photomultiplier tube voltage of 700 V for Cy3 and 633 nm with a photomultiplier tube voltage of 800 V for Cy5. Array images were analyzed using Array Vision software (version 4.0, Molecular Dynamics, Imaging Research Inc., Ontario, Canada). Integral intensities were obtained for each spot and slide background was subtracted. The eight replicated spots per gene were averaged for Cy3 and separately for Cy5. The average intensity per gene was determined for Cy3 or Cy5. A scaling factor was used so that the overall average intensities (fold induction) between Cy3 and Cy5 were equivalent. These normalized values were used to determine the ratio of intensities between treated (Cy5) and control (Cy3) values for each gene.

Sequence-Structure Comparisons and Homology Modeling

Structural modeling was primarily based on the known crystal structures for maize GSTI (Neufeind et al., 1997a). A Megalign sequence alignment of all maize and soybean type III GSTs and maize GSTI was used to map regions of

type III GST homology onto the sequence of GSTI. This also showed that ZmGST 24 at 19.2% identical is the closest type III GST to GSTI, and for this reason, ZmGST 24 was chosen for creating a homology model. Submission of the ZmGST 24 sequence to the Swiss-model program (Guex and Peitsch, 1997; http://www.expasy.ch/swissmod/) revealed that the degree of homology to existing templates was too low for a model to be generated. To obtain a model, three intermediate models were produced by creating artificial chimeric sequences of GSTI (a known template) with increasing amounts of ZmGST 24 sequence inserted. Step 1, substitute P20-K113 from ZmGST 24 for W13-Q105 of GSTI; step 2, substitute M1-G18 from ZmGST 24 for M1-M11 of GSTI, and C114-L147 from ZmGST 24 for Y106-L150 of GSTI; and step 3, substitute Q148-E189 from ZmGST 24 for T151-A186 of GSTI. The model generated at each step was used as template for the subsequent step, and the model at step 3 was used as template for the complete ZmGST24 protein, yielding a homology model from P9-V201 of this protein.

Protein Expression and Activity Assays

Sequences for GSTI, GSTIII, and GmGst 1 (GH2/4) were obtained by reverse transcriptase-PCR and verified by sequencing; in2-1 was a kind gift from Dr. Howard Hershey (DuPont, Wilmington, DE) and GmGST 4 (GSTa) was described previously (McGonigle and O'Keefe, 1998). These sequences were cloned into a pET vector (Novagen, Madison, WI), protein was expressed according to the manufacturer's instructions, and proteins were affinity purified using GSH agarose (Sigma). The *Bz2* clone was obtained from Dr. Virginia Walbot (Stanford University, Stanford, CA). All other sequences were obtained from cDNA libraries as described above, and proteins (including BZ2) were expressed using the pET-30 LIC system (Novagen) according to the manufacturer's instructions. Expressed protein was purified using the HIS binding kit (Novagen) according to the manufacturer's instructions. Purified protein was examined on 15% to 20% SDS-Phast Gels (Bio-Rad Laboratories, Medina, OH) and quantitated either spectrophotometrically using bovine serum albumin as a standard, or using the sequence-derived extinction coefficient and the UV absorbance of the isolated protein. A subset of the proteins formed inclusion bodies (ZmGST 9, ZmGST 10, ZmGST 12, and BZ2) and these proteins were co-expressed with a GroESL background to produce soluble protein (Goloubinoff et al., 1989).

GST activity was measured essentially as described previously (McGonigle et al., 1998), using an HP1050 HPLC with a diode array detector to quantitate the formation of a single GSH conjugate peak from the substrates chlorimuron ethyl, alachlor, atrazine, CDNB, ethacrynic acid, trans-stilbene oxide, and 1,2-epoxy-3-(*p*-nitrophenoxy) propane.

ACKNOWLEDGMENTS

LITERATURE CITED

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Venter JC (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science **252:** 1651–1656

Alfenito MR, Souer E, Goodman CD, Buell R, Mol J, Koes R, Walbot V (1998) Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione *S*-transferases. Plant Cell **10:** 1135–1149

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol **215:** 403–410

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:** 3389–3402

Andrews CJ, Jepson I, Skipsey M, Townson JK, Edwards R (1997) Nucleotide sequence of a glutathione transferase (accession no. Y10820) from soybean with activity towards herbicides. Plant Physiol **113:** 1005

Bhargava MM, Listowsky I, Arias IM (1978) Ligandin: bilrubin binding and glutathione-*S*-transferase activity are independent processes. J Biol Chem **253:** 4112–4115

Board PG, Baker RT, Chelvanayagam G, Jermiin LS (1997) Zeta, a novel class of glutathione transferases in a range of species from plants to humans. Biochem J **328:** 929–935

Board PG, Coggan M, Wilce MCJ, Parker MW (1995) Evidence for an essential serine residue in the active site of the theta class glutathione transferases. Biochem J **311:** 247–250

Bonaldo MF, Lennon G, Soares MB (1996) Normalization and subtraction: two approaches to facilitate gene discovery. Genome Res **6:** 791–806

Boyer TD (1986) Covalent labeling of the nonsubstrate ligand-binding site of glutathione *S*-transferases with bilirubin-Woodward's reagent K. J Biol Chem **261:** 5363–5367

Brown HM, Neighbors SM (1987) Soybean metabolism of chlorimuron ethyl: physiological basis for soybean selectivity. Pestic Biochem Physiol **29:** 112–120

Clegg MT, Cummings MP, Durbin ML (1997) The evolution of plant nuclear genes. Proc Natl Acad Sci USA **94:** 7791–7798

Cottingham CK, Hatzios KK (1992) Basis of differential tolerance of two corn hybrids (*Zea mays*) to metolachlor. Weed Sci **40:** 359–363

Cottingham CK, Hatzios KK, Meredith SA (1993) Comparative responses of selected corn (*Zea mays* L.) hybrids to EPTC and metolachlor. Weed Res **33:** 161–170

Cummings I, Cole DJ, Edwards R (1999) A role for glutathione transferases functioning as glutathione peroxidases in resistance to multiple herbicides in black grass. Plant J **18:** 285–292

Czarnecka E, Nagao RT, Key JL, Gurley WB (1988) Characterization of *Gmhsp26-A*, a stress gene encoding a divergent heat shock protein of soybean: heavy-metal-induced inhibition of intron processing. Mol Cell Biol **8:** 1113–1122

Dixon DP, Cole DJ, Edwards R (1997) Characterization of multiple glutathione transferases containing the GSTI subunit with activities toward herbicide substrates in maize (*Zea mays*). Pestic Sci **50:** 72–82

Dixon DP, Cole DJ, Edwards R (1998a) Purification, regulation and cloning of a glutathione transferase (GST) from maize resembling the auxin-inducible type-III GSTs. Plant Mol Biol **36:** 75–87

Dixon DP, Cole DJ, Edwards R (1999) Dimerisation of maize glutathione transferases in recombinant bacteria. Plant Mol Biol **40:** 997–1008

Dixon DP, Cummings I, Cole DJ, Edwards R (1998b) Glutathione-mediated detoxification systems in plants. Curr Opin Plant Biol **1:** 258–266

Droog F (1997) Plant glutathione *S*-transferases, a tale of theta and tau. J Plant Growth Regul **16:** 95–107

Droog FNJ, Hooykaas PJJ, van der Zaal BJ (1995) 2,4-Dichlorophenoxyacetic acid and related chlorinated compounds inhibit two auxin-regulated type-III tobacco glutathione *S*-transferases. Plant Physiol **107:** 1139–1146

Durbin ML, McCaig B, Clegg MT (2000) Molecular evolution of the chalcone synthase multigene family in the morning glory genome. Plant Mol Biol **42:** 79–92

Edwards R, Dixon DP, Walbot V (2000) Plant glutathione *S*-transferases: enzymes with multiple functions in sickness and in health. Trends Plant Sci **5:** 193–198

Edwards R, Dixon RA (1991) Glutathione *S*-cinnamoyl transferases in plants. Phytochemistry **30:** 79–84

Frear DS, Swanson HR, Thalacker FW (1991) Induced microsomal oxidation of diclofop, triasulfuron, chlorsulfuron, and linuron in wheat. Pestic Biochem Physiol **41:** 274–287

Fuerst EP, Irzyk GP, Miller KD (1993) Partial characterization of glutathione *S*-transferase isozymes induced by the herbicide safener benoxacor in maize. Plant Physiol **102:** 795–802

Goloubinoff P, Gatenby AA, Lorimer GH (1989) GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli*. Nature **337:** 44–47

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis **18:** 2714–2723

Hagen G, Kleinschmidt AJ, Guilfoyle TJ (1984) Auxin-regulated gene expression in intact soybean hypocotyl and excised hypocotyl sections. Planta **162:** 147–153

Hershey HP, Stoner TD (1991) Isolation and characterization of cDNA clones for RNA species induced by substituted benzenesulfonamides in corn. Plant Mol Biol **17:** 679–690

Higgins DG, Sharp PM (1989) Fast and sensitive multiple sequence alignments on a microcomputer. CABIOS **5:** 151–153

Holt DC, Lay VJ, Clarke ED, Dinsmore A, Jepson I, Bright SWJ, Greenland AJ (1995) Characterization of the safener-induced glutathione *S*-transferase isoform II from maize. Planta **196:** 295–302

Itzhaki H, Woodson WR (1993) Characterization of an ethylene-responsive glutathione *S*-transferase gene cluster in carnation. Plant Mol Biol **22:** 43–58

Jepson I, Holt DC, Roussel V, Wright SY, Greenland AJ (1997) Transgenic plant analysis as a tool for the study of maize glutathione *S*-transferases. *In* KK Hatzios, ed, Regulation of Enzymatic Systems Detoxifying Xenobiotics in Plants. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 313–323

Jepson I, Lay VJ, Holt DC, Bright SWJ, Greenland AJ (1994) Cloning and characterization of maize herbicide safener-induced cDNAs encoding subunits of glutathione *S*-transferase isoforms I, II and IV. Plant Mol Biol **26:** 1855–1866

Klapheck S (1988) Homoglutathione: isolation, quantification and occurrence in legumes. Physiol Plant **74:** 727–732

Koeppe MK, Barefoot AC, Cotterman CD, Zimmerman WT, Leep DC (1998) Basis of selectivity of the herbicide flupyrysulfuron-methyl in wheat. Pestic Biochem Physiol **59:** 105–117

Li Z-S, Alfenito M, Rea PA, Walbot V, Dixon RA (1997) Vacuolar uptake of the phytoalexin medicarpin by the glutathione conjugate pump. Phytochemistry **45:** 689–693

Lockhart DJ, Winzeler EA (2000) Genomics, gene expression and DNA arrays. Nature **405:** 827–836

Marrs KA (1996) The functions and regulation of glutathione *S*-transferases in plants. Annu Rev Plant Physiol Plant Mol Biol **47:** 127–158

Marrs KA, Alfenito MR, Lloyd AM, Walbot V (1995) A glutathione *S*-transferase involved in vacuolar transfer encoded by the maize gene *Bronze-2*. Nature **375:** 397–400

Martinoia E, Grill E, Tommasini R, Kruz K, Amrhein N (1993) ATP-dependent glutathione *S*-conjugate "export" pump in the vacuolar membrane of plants. Nature **364:** 247–249

Matamoros MA, Moran JF, Iturbe-Ormaetxe I, Rubio MC, Becana M (1999) Glutathione and homoglutathione synthesis in legume root nodules. Plant Physiol **121:** 879–888

McGonigle B, Lau S-MC, Jennings LD, O'Keefe DP (1998) Homoglutathione selectivity by soybean glutathione *S*-transferases. Pestic Biochem Physiol **62:** 15–25

McGonigle B, Lau S-MC, O'Keefe DP (1997) Endogenous reactions and substrate specificity of herbicide metabolizing enzymes. *In* KK Hatzios, ed, Regulation of Enzymatic Systems Detoxifying Xenobiotics in Plants. Kluwer

Academic Publishers, Dordrecht, The Netherlands, pp 9–18

McGonigle B, O'Keefe DP (1998) GSTa, a 2,4-D-inducible glutathione S-transferase from soybean (*Glycine max* cv Williams 82) (accession no. AF048978) (PGR 98–079). Plant Physiol **117:** 332

Meyer RC Jr, Goldsbrough PB, Woodson WR (1991) An ethylene responsive flower senescence-related gene from carnation encodes a protein homologous to glutathione S-transferases. Plant Mol Biol **17:** 277–281

Moore RE, Davies MS, O'Connell KM, Harding EI, Wiegand RC, Tiemeier DC (1986) Cloning and expression of a cDNA encoding a maize glutathione S-transferase in *E. coli*. Nucleic Acids Res **14:** 7227–7235

Neuefeind T, Huber R, Dasenbrock H, Prade L, Bieseler B (1997a) Crystal structure of herbicide-detoxifying maize glutathione S-transferase-I in complex with latoylglutathione: evidence for an induced-fit mechanism. J Mol Biol **274:** 577–587

Neuefeind T, Huber R, Reinemer P, Knäblein J, Prade L, Mann K, Bieseler B (1997b) Cloning, sequencing, crystallization and x-ray structure of glutathione S-transferase-III from *Zea mays* var. *mutin*: a leading enzyme in detoxification of maize herbicides. J Mol Biol **274:** 446–453

Neuefeind T, Reinemer P, Bieseler B (1997c) Plant glutathione S-transferases and herbicide detoxification. Biol Chem **378:** 199–205

Rea PA (1999) MRP subfamily ABC transporters from plants and yeast. J Exp Bot **50:** 895–913

Rea PA, Li Z-S, Lu Y-P, Drozdowicz YM, Martinoia E (1998) From vacuolar GS-X pumps to multispecific ABC transporters. Annu Rev Plant Physiol Plant Mol Biol **49:** 727–760

Reinemer, Prade L, Hof P, Neuefeind T, Huber R, Zettl R, Palme K, Schell J, Koelln I, Barunik HD, Bieseler B (1996) Three-dimensional structure of glutathione S-transferase from *Arabidopsis thaliana* at 2.2 Å resolution: structural characterization of herbicide-conjugating plant glutathione S-transferases and a novel active site architecture. J Mol Biol **255:** 289–309

Roxas VP, Smith RK Jr, Allen ER, Allen RD (1997) Over-expression of glutathione S-transferase/glutathione peroxidase enhances the growth of transgenic tobacco seedlings during stress. Nat Biotechnol **15:** 988–991

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol **4:** 406–425

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning: A Laboratory Manual, Ed 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Shah DM, Hironaka CM, Eiegand RC, Harding E, Krivi GG, Tiemeier DC (1986) Structural analysis of a maize gene coding for glutathione-S-transferase involved in herbicide detoxification. Plant Mol Biol **6:** 203–211

Shimabukuro RH, Frear DS, Swanson HR, Walsh WC (1971) Glutathione conjugation: an enzymatic basis for atrazine resistance in corn. Plant Physiol **47:** 10–14

Skipsey M, Andrews CJ, Townson JK, Jepson I, Edwards R (1997) Substrate and thiol specificity of a stress-inducible glutathione transferase from soybean. FEBS Lett **409:** 370–374

Skipsey M, Andrews CJ, Townson JK, Jepson I, Edwards R (2000) Cloning and characterization of glyoxalase I from soybean. Arch Biochem Biophys **374:** 261–268

Ulmasov T, Ohmiya A, Hagen G, Guilfoyle T (1995) The soybean GH2/4 gene that encodes a glutathione S-transferase has a promoter that is activated by a wide range of chemical agents. Plant Physiol **108:** 919–927

Wittenbach VA, Koeppe MK, Lichtner FT, Zimmerman WT, Reiser RW (1994) Basis of selectivity of triflusulfuron methyl in sugar beets (*Beta vulgaris*). Pestic Biochem Physiol **49:** 72–81

Zettl R, Schell J, Palme K (1994) Photoaffinity labeling of *Arabidopsis thaliana* plasma membrane vesicles by 5-azido-[7-^{3}H] indole-3-acetic acid: identification of a glutathione S-transferase. Proc Natl Acad Sci USA **91:** 689–693