ELSEVIER

Divergence in structure and function of tau class glutathione transferase from *Pinus tabulaeformis*, *P. yunnanensis* and *P. densata*

Qing-Yin Zeng [a,b], Xiao-Ru Wang [a,b,*]

[a] *State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, 100093 Beijing, China*
[b] *National Institute for Working Life, SE-90713 Umeå, Sweden*

Abstract

Glutathione transferases (GSTs) are a family of enzymes that play important roles in stress tolerance and detoxification in plants. The plant GSTs are divided into four classes (phi, tau, zeta and theta), among which tau is the most numerously represented. To date, studies on GSTs in plants have focused largely on crop species. There is extremely little information on the molecular characteristics of GSTs in gymnosperms. Generalization on GST characteristics unique to gymnosperms and the patterns of GST evolution in plants cannot be made before more members of the gene family in conifers are described. In this study we report three new GSTs from *Pinus tabulaeformis*, *Pinus densata* and *Pinus yunnanensis*. Structural and phylogenetic analyses placed these three GSTs in tau class. The tau GST class is subdivided into three clades and this subdivision seems an ancient event that may have pre-dated the gymnosperm and angiosperm split. Sequence analysis revealed a highly conserved N-terminal domain in contrast to a highly variable C-terminal domain. Mutations even outside the critical glutathione-binding site in N-terminal domain can have pronounced effect on GST catalytic property. Thus, sequence similarity does not parallel functional specificity. The high diversity in C-terminal domain determines a wide range of substrate selectivity and specificity among tau GSTs. Thus the a few conserved residues in C-terminal domain seem essential to maintain the structure of the domain and the protein dimer. More extensive data on GST family organization and a thorough gene-by-gene analysis in conifers are needed to advance our understanding of the true diversity and evolution of GST in structure and function in plants.
© 2006 Elsevier Ltd. All rights reserved.

Keywords: Pines; Protein structure; Sequence diversity; Enzyme assay

1. Introduction

Glutathione transferases (GSTs, EC 2.5.1.18) are a family of enzymes that catalyze the conjugation of the tripeptide glutathione (GSH) to a range of hydrophobic and electrophilic compounds. GSTs are ubiquitously

distributed in eukaryotes and are involved in the cellular detoxification and excretion of many physiological and xenobiotic substances (Armstrong, 1991; Edwards et al., 2000). Based on protein sequence similarity and immunological cross-reactivity, eight classes of mammalian GSTs have been recognized: Alpha, Kappa, Mu, Pi, Theta, Sigma, Zeta and Omega (Frova, 2003). The plant GSTs are typically divided into four classes (phi, tau, zeta and theta) (Edwards et al., 2000). The zeta and theta GSTs are found both in plants and animals, but the phi and tau GSTs are unique to plants and are the most represented. GSTs are soluble proteins with typical molecular masses of around 50 kDa, each of which is composed of two subunits that form a dimeric structure (Dixon et al., 2002). Each subunit contains two distinct domains (the N-terminal domain and C-terminal domain), and have an active site consisting of a GSH-binding site (the G-site) and a hydrophobic substrate-binding site (the H-site). Whereas the structure of the G-site is well conserved among GSTs, the H-site varies widely in different classes, leading to differences in substrate selectivity (Sheehan et al., 2001; Ivarsson et al., 2003). Recently, plant GSTs have been ascribed roles as diverse as oxidative stress tolerance, detoxification of xenobiotic and heavy metal toxins, transportation of toxic secondary products, and cell signalling during stress responses (Kampranis et al., 2000; Loyall et al., 2000; Mueller et al., 2000; Thom et al., 2002; Kilili et al., 2004).

Among the four plant GST classes, tau is the most numerous. In both *Arabidopsis* and maize 28 tau GST genes have been identified (McGonigle et al., 2000; Wagner et al., 2002), and 20 and 40 in soybean and rice, respectively (McGonigle et al., 2000; Soranzo et al., 2004). The functions of tau GSTs have been speculated as generic cell protectants, and play essential roles in plant development and in buffering environmental stresses (Edwards et al., 2000). A large variety of tau GSTs may be important to enable the plant to cope with changing environmental conditions. Members of tau GSTs overlap in function, differ in their substrate specificity, thus providing the plant a broad range of protection, although the molecular mechanisms for various functions are largely unrevealed. Phylogenetic relationships among the GST classes have been presented in several reviews (Snyder and Maddison, 1997; Sheehan et al., 2001; Dixon et al., 2002; Marco et al., 2004). Many of the sequences used in phylogeny have come from studies lacking an analysis of the activity of each GST protein. It has been shown that GST function cannot be inferred from their high sequence similarity to other members of known function (Alfenito et al., 1998; McGonigle et al., 2000). Thus, to fully understand the evolution of GSTs in structure and function, each member should be characterized individually by biochemical and genetic analysis.

To date, GST studies on plants have concentrated mainly on *Arabidopsis* and agricultural crops such as rice, wheat, maize and soybean. In contrast, there is virtually no information on the molecular characteristics of GSTs in gymnosperms. Gymnosperms represent a large group of plants with a long evolutionary history. Most of the gymnosperms such as conifers are important forest-forming species. Unlike annual agricultural plants, the perennial conifers have long generation times and wide distributions over heterogeneous environments. Over their lifespans the conifers have to cope with severe stresses including temperature fluctuations, drought, elevated ozone and UV levels, halogenated hydrocarbons and other air pollutants. The presence of enzymes that can detoxify these compounds and respond to natural stresses is clearly beneficial for the adaptation of conifer trees. However, data on GSTs in conifer trees are very scarce. Only recently a tau GST from a pine *Pinus tabulaeformis* has been characterized in molecular structure and biochemical property (Zeng et al., 2005; Zeng and Wang, 2005). Generalization on GST properties and characteristics unique to gymnosperms and the patterns of GST evolution in plants cannot be made before more members of the gene family in conifers are described.

In this study, we report three new GSTs isolated from *P. tabulaeformis*, *Pinus yunnanensis* and *Pinus densata*. The three pines form extensive forest from northern to southern China. *P. tabulaeformis* is distributed from northern to central China covering cold temperate and drought regions. It is replaced by *P. yunnanensis* in southern China. *P. densata* is a high mountain pine native to Tibetan plateau. It grows from 2700 to 4000 m above sea level (Wang et al., 2001). Through enzyme assays, phylogenetic analysis and structural modelling we examine the differences among GSTs from pines and from other plants in structure and function. This type of analysis provides insights into the significance of the observed mutations in a biochemical context. Unlike the model plants, the lack of comprehensive genome information on conifers makes the comprehensive characterization of GST family in conifers difficult. Nevertheless, accumulation of individual GST gene information from more gymnosperms and non-model angiosperms would advance our understanding of the true diversity and evolution of the GST classes.

2. Materials and methods

2.1. Molecular cloning, expression and purification of recombinant protein

Based on a putative *Picea mariana* GST gene sequence (GenBank AF051238), two primers (5′-CAATA-CATTTTCCGTGCAA-3′ and 5′-CCGACATGGCTTGCCCT-3′) were designed to amplify GST cDNA sequences from *P. tabulaeformis*, *P. densata* and *P. yunnanensis*. Total RNA was isolated from needles of the three pines using an Aurum Total RNA Kit (Bio-Rad, Hercules, CA, USA). RT-PCR was performed as described by Zeng et al. (2005). The GST cDNA from *P. tabulaeformis*, *P. yunnanensis* and *P. densata* were named PtGSTU2 (DQ062681), PyGSTU2 (DQ062679) and PdGSTU2 (DQ062680), respectively.

For cloning, the open reading frame (ORF, including the stop codon) of the cDNA encoding PtGSTU2 and PyG-STU2 were amplified by PCR with primers GST-U1 5′-TA*GGATCC*ATGGCGAGGGAGGGAGAAA-3′ and GST-D1 5′-T*AAGCTT*TTACTAATTGCCAGTAGGTGGA-3′ (*Bam*HI and *Hind*III sites underlined, respectively). The ORF encoding PdGSTU2 was amplified using the primers GST-U1 and GST-D2 5′-T*AAGCTT*TTACTAATTGGCAG-TAGGTGGA-3′ (*Hind*III sites underlined). The PCR products were digested with *Bam*HI and *Hind*III and subcloned into the plasmid expression vector pET30a (Novagen, Inc., Madison, WI, USA). The proteins were expressed in *Escherichia coli* BL21 with 0.1 mM isopropyl-β-D-thiolactopyranoside (IPTG) induction at 25 °C. The recombinant proteins were purified using Ni Sepharose High Performance column (Amersham Pharmacia Biotech, Uppsala, Sweden).

2.2. Enzyme assays and kinetic analysis

Classical GST substrates 1-chloro-2,4-dinitrobenzene (CDNB), ethacrynic acid (ECA), 4-nitrophenyl acetate (4-NPA) and 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole (NBD-Cl) were used in GST enzymatic assays. GST activities towards CDNB, ECA and 4-NPA were measured as described by Habig et al. (1974), and activity towards NBD-Cl as described by Ricci et al. (1994). All assays were carried out at 25 °C. The molecular weight of the purified recombinant PtGSTU2, PdGSTU2 and PyGSTU2 were estimated by gel filtration on a Sephadex G-75 column. The enzyme's kinetic parameters were determined by steady-state kinetic analysis using GSH and CDNB as substrates (Zeng et al., 2005). Thermal stability of the three GSTs was analysed according to Zeng and Wang (2005). All assays were repeated three times.

2.3. Alignment and sequence analysis

Amino acid sequences of the pine GSTs were aligned with 70 full length tau, 22 phi, 16 zeta and 6 theta plant GSTs retrieved from GenBank, using ClustalX (Thompson et al., 1997) and further modified using BioEdit (Hall, 1999). Phylogenetic relationships among the GSTs were evaluated using the neighbour-joining method (Saitou and Nei, 1987) with 1000 bootstrap replicates using MEGA 3 software (Kumar et al., 2004). Pairwise gap deletion and amino acid Poisson correction were applied. Protein sequence similarities among the tau GSTs were analysed with GeneDoc version 2.6 (Nicholas et al., 1997).

2.4. Structure modelling

To generate structural models, amino acid sequences of the three pine GSTs were submitted to SwissModel (Schwede et al., 2003) using the First Approach Method set at default parameters. A homology model was generated using the known X-ray structures of the wheat TaGSTU4 and rice OsGSTU1 (Protein Data Bank code Nos. 1gwc and 1oyj, respectively) as templates. Swiss-Pdb viewer version 3.7 and MOLMOL were used to generate a three-dimensional image.

3. Results

3.1. Protein sequence analysis

The three pine GST cDNAs contained a 711 bp ORF (excluding the stop codon), which encoded a 237 amino acid protein. PdGSTU2 and PyGSTU2 had almost identical protein sequence with only one amino acid difference (Fig. 1
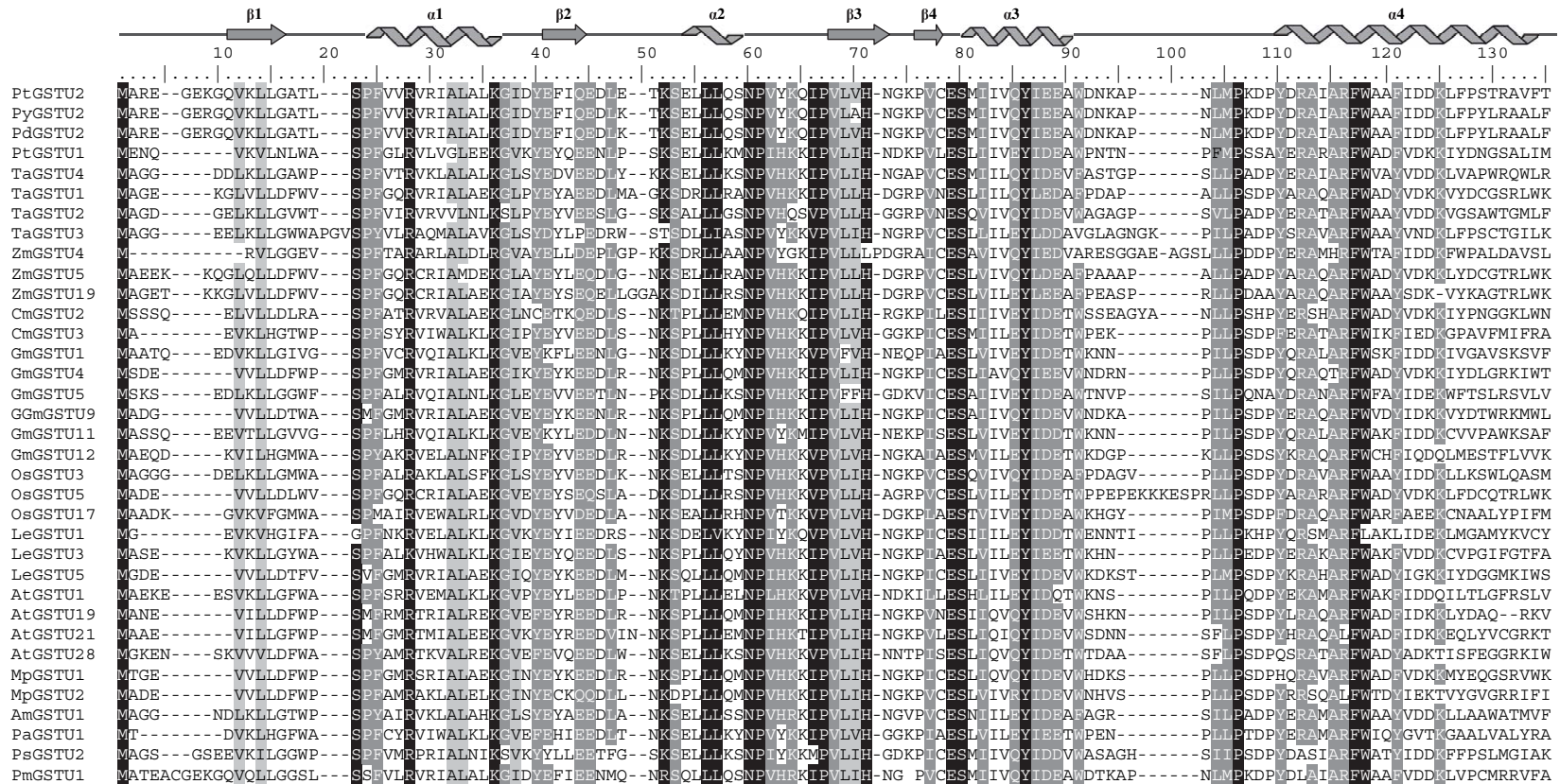
Fig. 1. Amino acid sequence alignment of 35 tau GSTs. The predicted secondary structure elements of PtGSTU2 are marked on top of the matrix. Alpha helices and beta strands are represented as helices and arrows, respectively. Highly conserved residues are shaded. The GenBank accession numbers for all the sequences are listed in the Table 1.

```
                                α5                                          α6                       α7           α8          α9

            140       150       160       170       180       190       200       210       220       230       240       250       260
            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|.

PtGSTU2   G---QGEQLQKAVEDSVTNFLLIEEALRTNRCFAG---KAYFGG--DEIGLLDIALGGMSTFIKALEKATNTILI-DPKKMPLLTAWMDRFCKSD-GVKEVLPD-PVKHAEYLCARRASMISPPTGN---
PyGSTU2   G---QGEQLQKAVEDSVTNFLLIEEALRTNRCFAG---KAYFGG--DEIGLLDIALGGMSTFIKAVEKATDSVLI-DPEKMPLLTAWMDRFCKSD-GVKEVLPD-PVKHAEYLCARRASMVSPPTGN---
PdGSTU2   G---QGEQLQKAVEDSVTNFLLIEEALRTNRCFAG---KAYFGG--DEIGLLDIALGGMSTFIKAVEKATDSVLI-DPEKMPLLTAWMDRFCKSD-GVKEVLPD-PVKHAEYLCARRASMVSPPTAN---
PtGSTU1   KC--KGEAQEEAKRNMLEYLGLLEGALDEL--SGG--IKPYFGG--EKFGYMDIAFIPFASWFQAWEVMGNWKIP-LETQFPRLHEWVNACMERE-SVKKVLPH-PEKVAEFAMQIRQRFVGSD------
TaGSTU4   GK--TEEEKSEGKKQAFAAVGVLEGALRE--CSKG---GGFFGG--DGVGLVDIALGGVLSMKVTEALSGDKIF-DAAKTPLIAAWVERFIELD-AAKAALPD-VGRLLEFAKAREAAAAASK-----
TaGSTU1   L---KGEPQAQARAEMLDILKTLDGALGD---------KPFFGG--DKFGFVDAAFAPFTAWFHSYERYGEFSL---PEVAPKIAAWAKRCGERE-SVAKSLYS-PDKVYDFIGLLKKKYGIE-------
TaGSTU2   SCK-TEEERAEAVSRAVAALETLEGAFAE--CSKG---KAFFGG--DAIGFVDVLGGYLGWFGAIDKIIGRRLI-DPARTPLIARWEERFRAAD-AAKGVVPDDADKMLEFLPTVLA--WIAGKAK---
TaGSTU3   TT--KQEERAGKMEETLSGLRHLEAVMAE--CSEGEAEAPFFGG--DAIGFLDIALGCYLPWFEAAGRLAGLGPIIDPARTPKLAAWAERFSVAE-PIKALLPG-VDKLEEYITTALYPKWNIAVTGN--
ZmGSTU4   AP--TPGARAQAAEDTRAALSLLLEAFKD--RSNG---RAFFSGDAAPGLLDIALGCFLPALRACERLHGLSLI-DASATPLIDGWSQRFAAHP-AAKRVLPD-TEKVVQFTRFLQVQAQFR-VHVS--
ZmGSTU5   L---KGDGQAQARAEMVEILRTLEGALGD---------GPFFGG--DALGFVDVALVPFTSWFLAYDRFGGVSV---EKECPKIAAWRCAERP-SVAKNLYP-PEKVYDFVCGMKKRLGIE-------
ZmGSTU19  L---RGDARAQARAEIVQVVRNLDGELGD---------KAFFGG--EAFGFVDVALVPPVPWLPSYERYGDFSV---AEIAPRLAAWARRCAQRE-SVARTLHP-PEKVDEFINLLKKTYGIE------
CmGSTU2   KKGLTEEEKEAAMKELLGSFKQLEEELGD---------KPYFGG--QSFGFLDISLIPFYSMFLAFKLLGNLDL---EAECPKILEWAKRCCLRE-TVAKNLPT-QQEVYDVVLDIIKRH---------
CmGSTU3   K---SDEEQENAKKQSLEMLRTLEWGFRGGDGR---ERKFFGG--DEIGMTDLAFGWIAHWIGVIEEVVNVKLL-DANEFPKVYAWTERFKDAA-VIRDNLPD-RNLMAAAFRRRREQLLQST------
GmGSTU1   TV--DEKEREKNVEETYEALQFLENELKD---------KKFFGG--EEFGLVDIAAVFIAFWIPIFQEIAGLQLF-TSEKFPILYAWSQEFLNHP-FVHEVLPP-RDPLFAYFKARYESLSASK------
GmGSTU4   S---KGEEKEAAKKEFIEALKLLEEQLGD---------KTYFGG--DNLGFVDIALVPPFYTWFKAYETFGTLNI---ESECPKFIAWAKRCLQKE-SVAKSLPD-QQKVYEFIMDLRKKLGIE------
GmGSTU5   AE--DDEAKKPHFEQAEEGLERLEEVFNKYSEG-----KAYFGG--DSIGFIDIGFGSFLSWMRVIEEMSGRKLL-DEKKHPGITQWAETFAADP-AVKGILPE-TDKLVEFAKILQLKWTAAAAAAAK-
GGmGSTU9  S---KGEEHEAGKKEFISIFKQLEETLSD---------KAYYCS--DTFGFLDIGLIPFYSWFYTFETYGNFKM---EEECPKIAWAKRCMQRE-AVSKSLS--------------------------
GmGSTU11  MT--DEKEKEKAKEELFEALSFLENELK---------GKFFGG--EEFGFVDIA--AVLI-PIIQEIAGLQLF--TSEKFPKLSKWSQDFHNHP-VVNEVLPP-KDQLFAYFKARAQSFVAKRKN----
GmGSTU12  T---DGEAQQKAIDHVYEKLKVLDGMKTYLGE-G--NAIISCV-ENNFGILDIVFCALYGAYKAHEEVIGLKFI-VPEKFPVLFSWLMAIAEVE-AVKIATPP-HEKTVGILQLFRLSALKSSSATE--
OsGSTU3   CK--TEQEKAAAMKETFAAVANLEAAFKE--CSKG---KPFFGG--DAVGYVDVTLGAVIGWVRVGEAVHGMRLF-DASRSPLIDAWLDRFAALD-AAKAVLPD-TGRLAEYAKMKQAEWAAAATN----
OsGSTU5   LRA-GDAAHEQAKRDMAEALGTLEAELGE---------GDYFGG--EAFGYLDVVLVPFVAWFHAYERLAGFAV---AEICPRLVAWGERCKGRD-SVAKTLTD-PEKVYEFALYLKAKFGAK-------
OsGSTU17  T---TGEEQRKLVHEAQQCLKTLETALE---------GKKFFGG--DAFGYLDVTGWFAYLWFVIEEACGVEVV-TDEALPKLAVKAVPP-RDKLVALNKARREQILSA------
LeGSTU1   G---KGEEREKGCDETFEVLKYLDNELQ---------NKKFFGG--DSIGFVDIVASYIALWFGAIQEAIGMELL-TEQKFPKLSKWNIDEFLCCR-IVKENLPN-REVLVLLYKAQFAAATQKASS----
LeGSTU3   K---VGVEQQKIAKEARENLKILEDELG---------KKHFFGD--AKIGFMDVTSAWIICWAQIVEEVVDIRLI-DAEEMPSLVSWFQNVLEAAPILKECTPP-KDLLLEHNKGFHEMLVASAPS----
LeGSTU5   S---KVEEHKTANKDFIECLKVLEGELGD---------KPYFDG--KNFGLVDMAFIPYYSWFPVYKKLSNLNI---EAECPKFVAWAKRCMQKE-SVSKTLVD-PDKIYEFIVLFRQKIGVA-------
AtGSTU1   K---AEKGREVAIEETRELLMFLKEVT---------GKDFFGG--KTIGFLDIIAGSMIPCLARLWKGIGIDMIPEEKFPELNRWIKNLEEVE-AVRGCLPP-REKQIERMTKIAETIKSA-------
AtGSTU19  WAT-KGEEQEAGKKDFIEILKTLSELGD---------KPYFSG--DDFGYVDIALIGFYTWFPAYEKFANFSI---ESECPKIIAWVKKCLQRE-SVAKSLPD-PEKVTEFVSELRKKFVPE-------
AtGSTU21  WAT-KGEELEAANKEFIEILKTLQCELGE---------KPFFGG--DKFGFVDLVLIGFYSWFPAYQKFGNFSI---EPECLKIIAWGKRCMQRE-SVAKALPD-SEKVVGYVLQLKKLYGIE------
AtGSTU28  GNK-KGEEQEKGKKEFLESLKVLEAELGD---------KSYFGG--ETFGYVDITLVPPFYSWFYALEKCGDFSV---EAECKIVAWGKRCVERN-SVAATLPE-SEKVYQQVLKLRQIFGVE------
MpGSTU1   T---KGEEQATAKKEFIESLKLLEGELGA---------SLTFGG--ENLGYVDGVLVPFYSWVYAYEKCGNFNI---EPECPKIIAWAKRCMEKE-SVAKSLPD-QEKVYDFVLQLKKIYGIE------
MpGSTU2   A---TGEEQKKAKEELIQFLKTILETELGE--------KLYFGG--DSIGFVDIALLPFTAWFYTYETCGNFSI---ESECPKLIBWVKRCTQKE-SVSKALAH-PHKIYDFFLEYKEKLGGN------
AmGSTU1   KGK-TEEEKLEGKKALFAALETLEGALAK--CSDG---KDFFGG--DTVGLVDVVLGSHLSFLKATEAMAGEEIL-RSDRTQLLAAWMARFSELD-AAKAALPD-VDRVVEFVKMRQARLAAAAAASNN
PaGSTU1   ----SGEELEKAVKEVVEVLRVLEEQGLG--------DKKFFGG--DSINLVDISYGLFTCWFAAMEEAMGVKVL-EPSTLPRLHAWAQNFIEVP-LIKENLPD-NDKLLLFMKGFREKLVNE-------
PsGSTU2   SK--DAEEKKAAIEQAIAAFGILEEAYQK--TSKG---KDFFGG--EKIGFVDIAFGCYVGWIRVTEKMNGIKLF-DEEKVPGITKWAEKFCADE-TVKSVVPE-TDALMEFAKKIFGSKPPPSN-----
PmGSTU1   G---HGEQLQKEAEDLVTNFHLLEEALRTNSCFSG---KAYFGG--DKIGLVDIALGGMLAVLKGLEKATDTVII-DPEKMPLLSAWMDRFCQSN-GVKEVLPD-PAKQLESLSARRARLAS-PAGN---
```
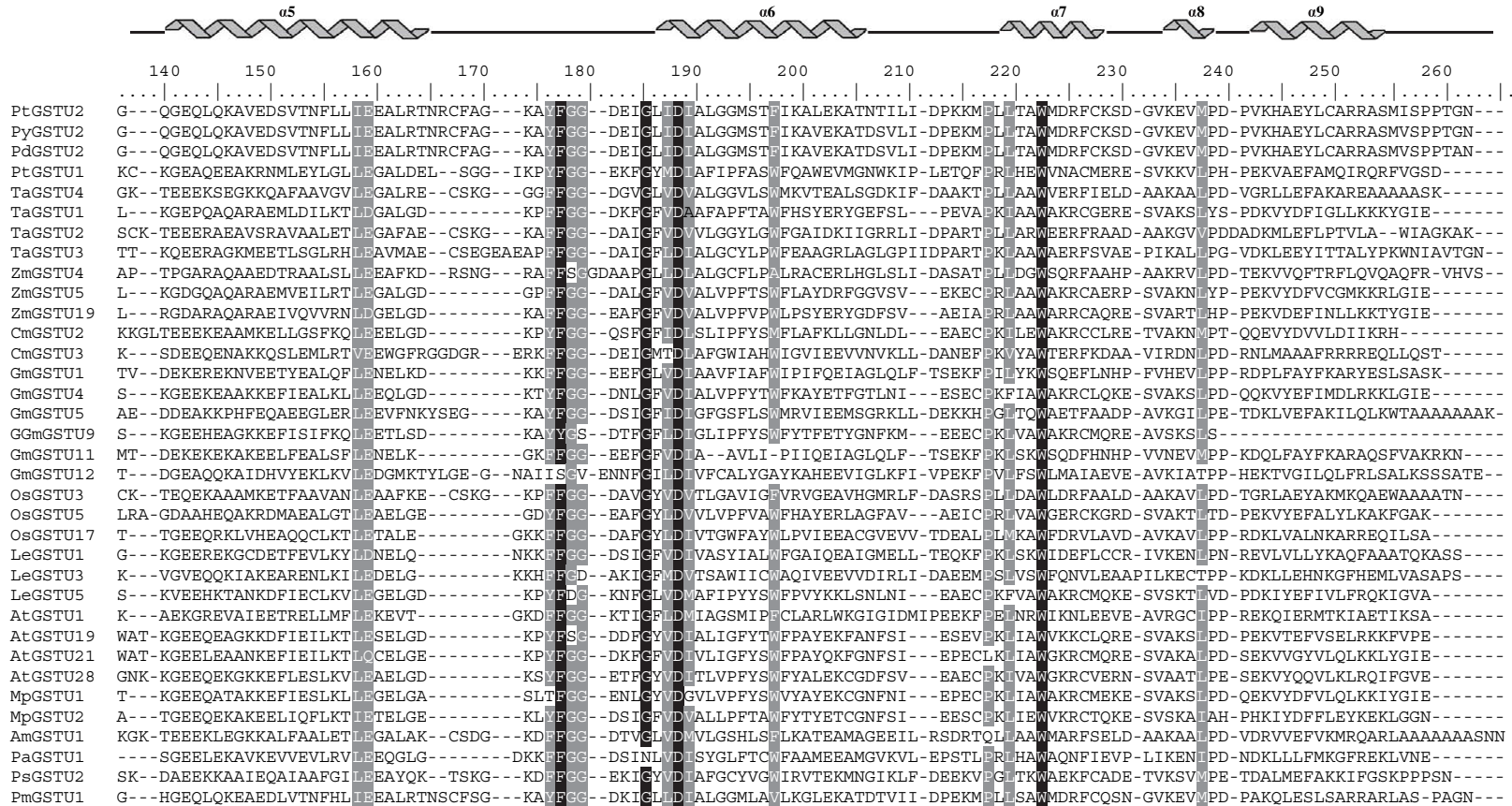
Fig. 1 (continued).

alignment position 70). The sequence of PtGSTU2 showed 94% identity to PdGSTU2 and PyGSTU2. Thus, the three sequences represent the same gene in the three pines. These three GSTU2 are similar to PmGSTU1 (AF051238) isolated from *P. mariana* with sequence identity of 74%, but divergent from PtGSTU1 (AY651784) isolated from *P. tabulaeformis* with sequence identity of 38%.

The protein sequence homology analysis was performed with the BLAST program at the NCBI web site (www.ncbi.nlm.nih.gov/BLAST) on all databases. The sequences of PtGSTU2, PdGSTU2 and PyGSTU2 showed a high similarity to the plant tau class GSTs. When aligned with tau, phi, theta and zeta GSTs, the three pine GSTU2 clearly positioned within tau class (data not shown). The diversification of the four GST classes was pronounced and each class formed a separate clade with high bootstrap support, with one exception that a phi GmGSTF23 (AF243378) clustered into theta GSTs (data not shown). The classification of this GST should be re-examined. To simplify the presentation, the 70 tau GSTs used in the initial phylogenetic analysis were reduced to 32 by selecting representatives from each clade. Their origins are listed in Table 1. An alignment of PtGSTU2, PdGSTU2 and PyGSTU2 with these 32 tau GSTs is shown in Fig. 1. Phylogenetic analysis of these 35 tau GSTs revealed three clades (Fig. 2A). The three pine GSTs from this study and PmGSTU1 from *P. mariana* were grouped together within clade I. PtGSTU1 was distinctly divergent from them and clustered into clade II. There is no clear tendency of grouping the GSTs by species origin. Members of GSTs from the same species can spread out over the clades. This topology differs from the tree based on the corresponding gene nucleotide sequences where all the GST genes from monocots clustered into a separate clade (III, Fig. 2B), and the dicots and the conifers were mixed into another two clades.

Table 1
List of 35 tau GSTs included in the phylogenetic analysis

| GST	Species	GenBank accession no.
PtGSTU2	*Pinus tabulaeformis*	DQ062681
PdGSTU2	*Pinus densata*	DQ062680
PyGSTU2	*Pinus yunnanensis*	DQ062679
PtGSTU1	*Pinus tabulaeformis*	AY651784
TaGSTU1	*Triticum aestivum*	AJ414699
TaGSTU2	*Triticum aestivum*	AJ414700
TaGSTU3	*Triticum aestivum*	AJ414701
TaGSTU4	*Triticum aestivum*	AF479764
ZmGSTU4	*Zea mays*	U14599
ZmGSTU5	*Zea mays*	Y12862
ZmGSTU19	*Zea mays*	AF244684
CmGSTU2	*Cucurbita maxima*	AB055117
CmGSTU3	*Cucurbita maxima*	AB055118
GmGSTU1	*Glycine max*	J03197
GmGSTU4	*Glycine max*	AF048978
GmGSTU5	*Glycine max*	AF243360
GmGSTU9	*Glycine max*	AF243364
GmGSTU11	*Glycine max*	AF243366
GmGSTU12	*Glycine max*	AF243367
OsGSTU3	*Oryza sativa*	AF309379
OsGSTU5	*Oryza sativa*	AF309377
OsGSTU17	*Oryza sativa*	AF402804
LeGSTU1	*Lycopersicon esculentum*	AY007558
LeGSTU3	*Lycopersicon esculentum*	AY007560
LeGSTU5	*Lycopersicon esculentum*	AY007562
AtGSTU1	*Arabidopsis thaliana*	AJ010448
AtGSTU19	*Arabidopsis thaliana*	AY087032
AtGSTU21	*Arabidopsis thaliana*	NM_106483
AtGSTU28	*Arabidopsis thaliana*	NM_104246
MpGSTU1	*Malva pusilla*	AY206001
MpGSTU2	*Malva pusilla*	AY206002
AmGSTU1	*Alopecurus myosuroides*	AJ010448
PaGSTU1	*Populus alba*	AB190440
PsGSTU2	*Papaver somniferum*	AF118925
PmGSTU1	*Picea mariana*	AF051238
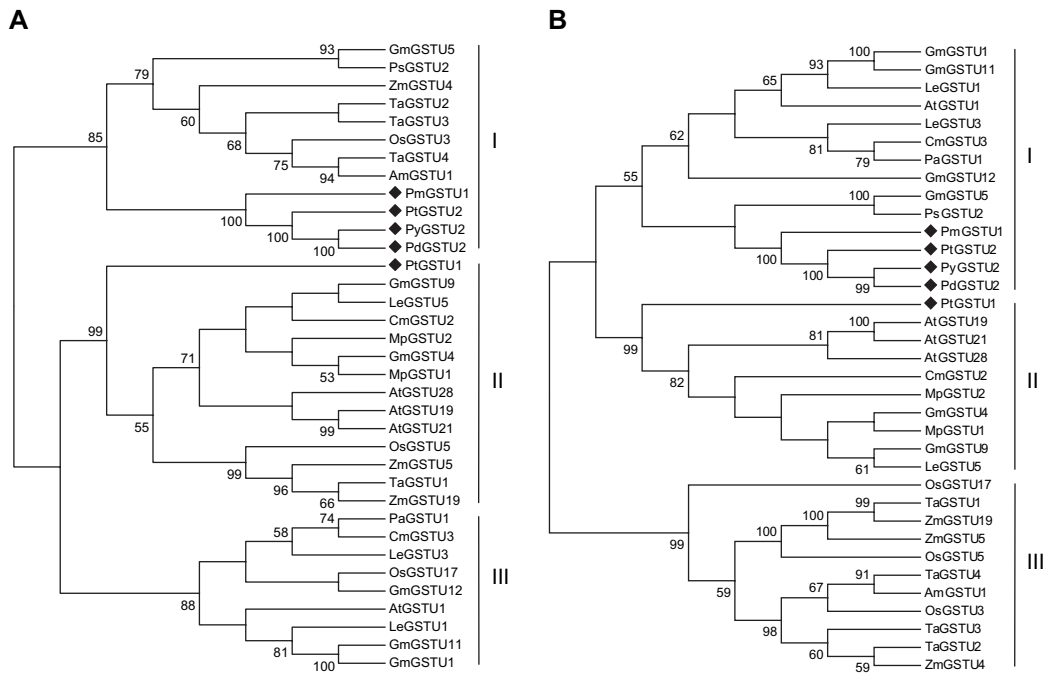
Fig. 2. Unrooted neighbour-joining phylogenetic tree obtained for the 35 tau GST protein sequences (A) and the corresponding gene nucleotide sequences (B). Bootstrap values higher than 50% are indicated on the branches. The five GSTs from conifers are marked with ◆.

Assignment of the secondary structure elements to the alignment matrix is described in the next section. The N-terminal domain of the tau GSTs spans to position 90 in the alignment matrix and the C-terminal domain from 110 to 265. In between is a loop of 11–19 residues that links the two domains. In the following calculations, we included the linker into the C-terminal domain. The N-terminal domain is highly conserved with a few length variations occurred in the beginning and in the two loops connecting β1 and α1, and β2 and α2. Fifteen residues showed to be absolutely conserved in this short domain. The C-terminal domain is much more variable, only seven residues are absolutely conserved among the 35 analysed GSTs. Protein sequence identities are plotted in Fig. 3. In N-terminal domain, 80% of the pairwise sequences scored 36–60% identity, while in C-terminal domain 80% of the sequences scored lower than 34% identity. At the full sequence length, 80% of the GSTs had 28–42% identity. PtGSTU2, PdGSTU2 and PyGSTU2 showed 30–42% sequence identity to the angiosperm tau GSTs. The most divergent members among the 35 GSTs had 26% identity (Table 2).
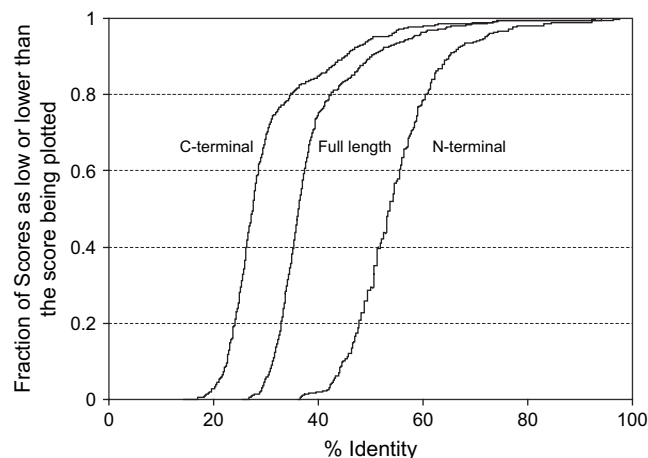
Fig. 3. Protein sequence identity plot of the 35 tau GSTs.

Table 2
Protein sequence identities (%) among the analysed tau GSTs

	N-terminal	C-terminal	Full length
Clade I	39−98	28−99	34−99
Clade II	46−92	31−69	37−75
Clade III	55−77	22−63	37−68
All 35 GSTs	37−98	17−99	26−99
3 pine GSTU2 and PmGSTU1	75−98	73−99	74−99

3.2. Structure modelling of PtGSTU2, PdGSTU2 and PyGSTU2

The three-dimensional structure of PtGSTU2, PdGSTU2 and PyGSTU2 were modelled based on the X-ray structure of a wheat tau GST TaGSTU4 and rice OsGSTU1. The three pine GSTU2 shared the same structure pattern (Fig. 4). The secondary structure elements for this GST are marked on the alignment matrix (Fig. 1). The monomer of the three GSTU2 constitutes two distinct domains; a smaller thioredoxin-like N-terminal domain (residues 1−82) and a larger helical C-terminal domain (residues 83−237). The N-terminal domain forms a β1α1β2 motif connected to a β3β4α3 motif by a short helix α2 and two loops (Fig. 4). At the end of helix α3 Trp83 forms the beginning of a 12-residue linker (residues 83−95) that joins the N- and C-terminal domains. The C-terminal domain is made of a α-helical bundle of six helices. The core of the C-terminal domain is a bundle of four helices (α4α5α6α7). The α5 and α6 helix are disrupted by two (Asp133 and Ser134) and three (Gly171, Met172 and Ser173) amino acids, respectively. To keep the naming of the secondary structure consistent with other tau GSTs the two segments of each α5 and α6 were named α5′-α5″ and α6′-α6″, respectively.

3.3. Biochemical characterization of the recombinant PtGSTU2, PyGSTU2 and PdGSTU2

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS−PAGE) showed that the three recombinant proteins were expressed mainly as soluble protein in *E. coli* BL21 at 25 °C. The molecular weight of the purified recombinant PtGSTU2, PyGSTU2 and PdGSTU2 were estimated to be approximately 60 kDa according to Sephadex G-75 gel filtration (data not shown). The subunit weight of PtGSTU2, PyGSTU2 and PdGSTU2 was 31.8 kDa (including the His-tag) according to SDS-PAGE analysis (Fig. 5). Thus, PtGSTU2, PyGSTU2 and PdGSTU2 are considered to be homodimers similar to the other plant tau GSTs (Thom et al., 2002; Kilili et al., 2004; Zeng et al., 2005).

The catalytic activities of the three recombinant proteins towards different substrates are presented in Table 3. PtGSTU2 and PdGSTU2 showed higher activities towards CDNB and NBD-Cl than to ECA. PyGSTU2 had
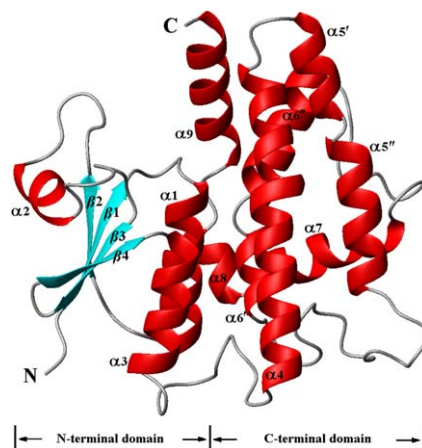

Fig. 4. Predicted three-dimensional structure of PtGSTU2. Alpha helices and beta strands are represented as red helices and blue arrows, respectively. N: N-terminus; C: C-terminus. The three-dimensional structural image was generated using the program MOLMOL.
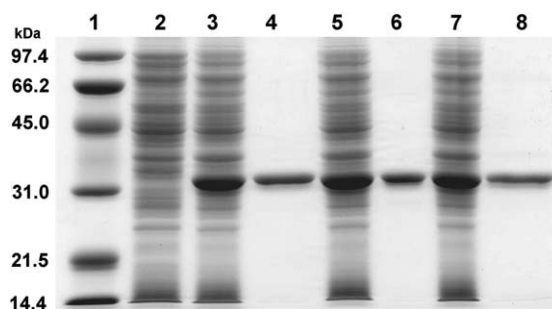
Fig. 5. SDS–PAGE analysis of the expression and purification of recombinant PtGSTU2, PdGSTU2 and PyGSTU2. Lane 1, molecular mass markers with the sizes shown on the left in kDa; lane 2, total cellular extract from *E. coli* BL21; lanes 3, 5 and 7, total cellular extracts from induced bacteria containing pET30a/PtGSTU2, pET30a/PdGSTU2 and pET30a/PyGSTU2; lanes 4, 6 and 8, the purified recombinant PtGSTU2, PdGSTU2 and PyGSTU2.

noticeably lower activity towards CDNB and NBD-Cl as compared to PtGSTU2 and PdGSTU2. All the three GSTs did not reveal any activity towards substrate 4-NPA.

The kinetic parameters of the three GSTs are presented in Table 4, and compared to that of PtGSTU1. All the three GSTU2 had similar affinity for GSH (K_m^{GSH}) (significance test $P > 0.05$). The affinity for CDNB (K_m^{CDNB}), however, varied by 2-fold among the three GSTU2. The highest catalytic efficiency (k_{cat}/K_m) for CDNB and GSH were found in PdGSTU2. These values in PdGSTU2, however, were still much lower than that of PtGSTU1. Thus, PtGSTU1 is more active in GSH conjugation reactions towards these substrates than the three GSTU2.

The thermal stability of the PtGSTU2, PdGSTU2 and PyGSTU2 is illustrated in Fig. 6. PtGSTU2, PdGSTU2 and PyGSTU2 retained 85%, 59% and 43% of their initial activity after 15 min incubation at 55 °C, respectively. PyGSTU2 almost lost its activity at 65 °C, whereas PtGSTU2 and PdGSTU2 still retained 16% and 7% of their initial activity, respectively. Thus, PyGSTU2 is more thermolabile than PtGSTU2 and PdGSTU2.

4. Discussion

GSTs comprise a widespread, multifunctional gene superfamily that is thought to have evolved from ancient stress-related proteins (Sheehan et al., 2001). The plant GST structure, function, genome organization and evolution have been the subject of several reviews (Edwards et al., 2000; Sheehan et al., 2001). However, all the available data to date are from a few crop plants. Genomic and biochemical characteristics of GSTs in gymnosperms remain unclear, and so are their phylogenetic relationships to angiosperms.

An analysis involving 96 tau GSTs from two monocots (rice and maize) and two dicots (soybean and *Arabidiopsis*) revealed a strict clustering between the two groups indicating rapid evolution of tau GSTs following the monocot–dicot split (Soranzo et al., 2004). By including more species into the analysis, our phylogenetic tree did not show the clear monocot–dicot clustering, except the tandemly duplicated genes in a genome often occupied adjacent positions on the tree (data not shown). Rather the tau class is divided into three subfamilies each with GSTs from different species scattered over. The five conifer GSTs (PtGSTU1, PtGSTU2, PyGSTU2, PdGSTU2, and PmGSTU1) did not show

Table 3
Specific activities of the three pine GSTU2 towards different substrates

GST	Specific activity (μmol min⁻¹ protein mg⁻¹)			
	CDNB	ECA	NBD-Cl	4-NPA
PtGSTU2	5.63 ± 0.28	1.18 ± 0.06	5.95 ± 0.29	<0.001
PdGSTU2	8.05 ± 0.41	2.16 ± 0.11	7.74 ± 0.38	<0.001
PyGSTU2	1.20 ± 0.06	1.58 ± 0.08	2.30 ± 0.11	<0.001

The values shown are means ± S.D., calculated from three replicates.

Table 4
Steady-state kinetic constants of PtGSTU2, PdGSTU2 and PyGSTU2

GST	GSH			CDNB		
	K_m (mM)	k_{cat} (S^{-1})	k_{cat}/K_m ($mM^{-1} S^{-1}$)	K_m (mM)	k_{cat} (S^{-1})	k_{cat}/K_m ($mM^{-1} S^{-1}$)
PtGSTU2	0.67 ± 0.15	96.98	144.75	1.70 ± 0.18	150.26	88.39
PdGSTU2	0.49 ± 0.09	117.20	239.18	0.81 ± 0.08	142.47	175.89
PyGSTU2	0.53 ± 0.12	15.10	28.49	1.25 ± 0.12	16.19	12.95
PtGSTU1	0.83 ± 0.14	402.40	484.80	0.47 ± 0.08	362.90	772.10

Data are means ± S.D, calculated from three replicates. Measurements for PtGSTU1 Zeng et al. (2005) are included for comparison.

clear divergence from the angiosperms to form a separate clade. They are distributed in two clades with other mono-cots and dicots. Although data from conifers are still very limited, our result nonetheless implies that the subdivision of the class is a more ancient event and may have pre-dated the split of gymnosperm and angiosperm. Different sub-families may have different functional specialization that was already evolved in the ancestor and further amplified through rapid evolution in different groups of plants. The topology of the phylogenetic trees based on protein and nu-cleotide sequences are different (Fig. 2A, B). This difference is likely a reflection of the functional constrains and/or specialization at the protein level.

Despite of low protein sequence identity among the analysed tau GSTs, certain structural conservation is evident. The GSH-binding N-terminal domain is highly conserved. All the highly conserved residues in this domain should have specific functions. In a previous study, Ser13, Lys40, Ile54, Glu66 and Ser67 of PtGSTU1 (alignment positions 23, 52, 66, 79 and 80 in Fig. 1, respectively) have been identified as G-site residues (Zeng et al., 2005). These res-idues are conserved across tau GSTs and are critical components that contribute to the enzyme's catalytic activity and structural stability. Replacement of these residues in PtGSTU1, through site-directed mutagenesis, significantly im-paired the enzyme's catalytic activity and protein folding (Zeng and Wang, 2005). In a natural occurring LeGSTU1 where the Ser13 is replaced by Gly (Kilili et al., 2004, Fig. 1), the enzyme showed very poor activity in GSH conjugation reactions with CDNB (0.07 μmol min^{-1} protein mg^{-1}) and NBD-Cl (0.15 μmol min^{-1} protein mg^{-1}) (Kilili et al., 2004), which further confirms the essential role of this residue in GST catalysis. Apart from the G-site residues, Asn53, Pro54 and Pro60 of PtGSTU2 (alignment positions 60, 61 and 67 in Fig. 1, respectively) are also absolutely conserved. Structure modelling showed that Asn53 and Pro54 together with the two following residues form a beta turn that connects α2 to β3. This specific loop structure has been suggested to play an important role in the recognition and orientation of GSH at the G-site (Dirr et al., 1994). Pro60 contributes to the sharp bending
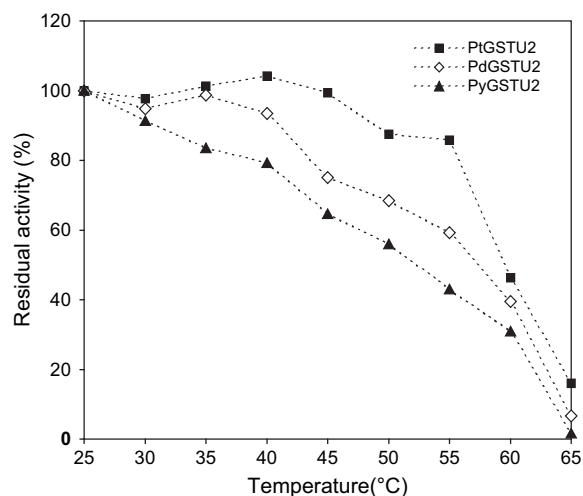

Fig. 6. Thermal stability of PtGSTU2, PdGSTU2 and PyGSTU2, based on the retention of enzymatic activity towards the substrate CDNB fol-lowing heat treatment.

conformation in β3, and is thought to be important for the proper folding and packing of G-site substructures (Dirr et al., 1994).

In contrast to the conserved N-terminal domain, the C-terminal domain is considerably more variable. This high diversity in sequence (and hence structure) determines a wide range of substrate selectivity and specificity among tau GSTs (Armstrong, 1997; Thom et al., 2002). The X-ray structure of TaGSTU4 indicates that Trp113, Arg114, Leu117, Arg118, Val168, Trp171, Phe218, Ala219 and Arg222 of TaGSTU4 (alignment positions 130, 131, 134, 135, 194, 197, 247, 248 and 251 in Fig. 1, respectively) are H-site residues in C-terminal domain. However, these residues are not conserved among members of tau GSTs both within and between species. This diversity in the H-site residues contributes to the wide array of substrate recognition among tau GSTs, thus provide plants a broad protections. Among the few conserved residues (alignment positions 106, 117, 118, 177, 185, 188 and 222) in this domain, little is known about their specific functions although they suggest functional and structural importance. The predicted three-dimensional structure of PtGSTU2 indicated that the conserved residues Phe103, Trp104 and Tyr107 (alignment positions 117, 118 and 121) in α4 helix form hydrophobic interactions with Val55 and Tyr56 (alignment positions 62 and 63) of the opposite monomer. Additional hydrophobic interaction is formed between Tyr96 (alignment positions 110) of helix α4 and Tyr78 (alignment positions 86) of α3 of the opposite monomer. The X-ray structure of TaGSTU4 and OsGSTU1 showed a similar conformation involving these residues (Thom et al., 2002). Thus, Tyr78, Tyr96, Phe103, Trp104 and Tyr107 are likely to play important roles in maintaining the dimmer structure. Structure modelling of PtGSTU2 also suggests that the conserved Asp166 forms hydrogen bonds with the main chain amide groups of Phe157 and Gly163, and the indole nitrogen of Trp199. These four residues Phe157, Gly163, Asp166 and Trp199 (alignment positions 177, 185, 188 and 222) could be essential to conserve the basic structure of the C-terminal domain. Thus, it seems all the conserved residues in C-terminal domain act as the corner-stones that maintain the proper structural frame of the domain and the protein dimer.

Compared to other characterized tau GSTs, the four pine GSTs showed a higher activity towards ECA (1.18−2.16 μmol min^{-1} protein mg^{-1}) than the GSTs from angiosperms (0.01−0.26 μmol min^{-1} protein mg^{-1}) (Zeng et al., 2005; McGonigle et al., 2000; Thom et al., 2002; Kilili et al., 2004; Dixon et al., 1999; Yu and Facchini, 2000; Fujita and Hossain, 2003). ECA is a phenylacetic acid derivative that contains an electrophilic group similar to the α-β-alkenals. The biological significance of this increased catalytic activity towards ECA in pines is unclear. More GSTs from conifers need to be characterized before speculations on its biological implication can be made. In general, members of tau GSTs vary largely in activity towards CDNB and NBD-Cl (Zeng et al., 2005; McGonigle et al., 2000; Thom et al., 2002; Kilili et al., 2004). This variation is neither correlated with sequence similarity nor with the species origin. PdGSTU2 and PyGSTU2 had nearly identical protein sequence only Val63 in PdGSTU2 is replaced with Ala63 in PyGSTU2. Their activities towards CDNB and NBD-Cl, however, differed by 7- and 3-fold, respectively. In addition, PyGSTU2 had only about 7.4% of the CDNB catalytic efficiency (k_{cat}/K_m) of PdGSTU2, and is more thermolabile. These demonstrate the importance of Vla63 in catalyzing GSH conjugation of CDNB and in maintaining the stability of PdGSTU2. Both PtGSTU1 and PtGSTU2 are from *P. tabulaeformis*, but PtGSTU1 showed about 4-fold higher enzymatic activity towards CDNB. Similar phenomenon is observed among tau GSTs isolated from soybean and maize, where GSTs with sequence identity higher than 89% can vary dramatically in activity towards different substrates (McGonigle et al., 2000). GmGSTU9 from soybean is much shorter than other tau GSTs (Fig. 1, McGonigle et al., 2000). Structure modelling revealed that GmGSTU9 lacks the α9 helix. This GST structure, however, apparently does not impair the enzyme's catalytic function as it still maintained relatively high activity towards CDNB (2.58 μmol min^{-1} protein mg^{-1}). This supports the theory that the variations in specificity determinants can arise without disruption of GSH binding and catalytic activity (Frova, 2003).

The natural occurring forms of the enzyme provide good opportunity for detailed analysis of how specific amino acid contributes to GST's biochemical property. Until now most of the GST characterizations focused on N-terminal domain. Our knowledge about the specific functions of the C-terminal domain is limited. Characterization of enzymatic activities should be extended to include more cosubstrates of diverse chemistries, coupled with sequence and structural information, to reveal the mechanisms underlying substrate specificity, and thus the link of sequence and structural diversity and functional significance. The patterns of GST family organization and diversity in gymnosperm are interesting areas that await extensive analysis. Large scale genomic and EST projects on conifers and a thorough gene-by-gene analysis will undoubtedly help to better clarify the evolution of GSTs in structure and function in plants.

Acknowledgments

References

Alfenito, M.R., Souer, E., Goodman, C.D., Buell, R., Mol, J., Koes, R., Walbot, V., 1998. Functional complementation of anthocyanin seques-tration in the vacuole by widely divergent glutathione S-transferases. Plant Cell 10, 1135–1149.

Armstrong, R.N., 1991. Glutathione S-transferases: reaction mechanism, structure, and function. Chem. Res. Toxicol. 4, 131–140.

Armstrong, R.N., 1997. Structure, catalytic mechanism, and evolution of the glutathione transferases. Chem. Res. Toxicol. 10, 2–18.

Dirr, H., Reinemer, P., Huber, R., 1994. X-ray crystal structures of cytosolic glutathione S-transferases. Implications for protein architecture, sub-strate recognition and catalytic function. Eur. J. Biochem. 220, 645–661.

Dixon, D.P., Cole, D.J., Edwards, R., 1999. Dimerisation of maize glutathione transferases in recombinant bacteria. Plant Mol. Biol. 40, 997–1008.

Dixon, D.P., Lapthorn, A., Edwards, R., 2002. Plant glutathione transferases (REVIEWS3004). Genome Biol. 3, 1–10.

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

Frova, C., 2003. The plant glutathione transferase gene family: genomic structure, functions, expression and evolution. Physiol. Plant 119, 469–479.

Fujita, M., Hossain, M.Z., 2003. Molecular cloning of cDNAs for three tau-type glutathione S-transferases in pumpkin (*Cucurbita maxima*) and their expression properties. Physiol. Plant 117, 85–92.

Habig, W.H., Pabst, M.J., Jakoby, W.B., 1974. Glutathione S-transferases. The first enzymatic step in mercapturic acid formation. J. Biol. Chem. 249, 7130–7139.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.

Ivarsson, Y., Mackey, A.J., Edalat, M., Pearson, W.R., Mannervik, B., 2003. Identification of residues in glutathione transferase capable of driving functional diversification in evolution – a novel approach to protein redesign. J. Biol. Chem. 278, 8733–8738.

Kampranis, S.C., Damianova, R., Atallah, M., Toby, G., Kondi, G., Tsichlis, P.N., Makris, A.M., 2000. A novel plant glutathione S-transferase/peroxidase suppresses Bax lethality in yeast. J. Biol. Chem. 275, 29207–29216.

Kilili, K.G., Atanassova, N., Vardanyan, A., Clatot, N., Al-Sabarna, K., Kanellopoulos, P.N., Makris, A.M., Kampranis, S.C., 2004. Differential roles of tau class glutathione S-transferases in oxidative stress. J. Biol. Chem. 279, 24540–24551.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform. 5, 150–163.

Loyall, L., Uchida, K., Braun, S., Furuya, M., Frohnmeyer, H., 2000. Glutathione and a UV light-induced glutathione S-transferase are involved in signaling to chalcone synthase in cell cultures. Plant Cell 12, 1939–1950.

Marco, A., Cuesta, A., Pedrola, L., Palau, F., Marin, I., 2004. Evolutionary and structural analyses of GDAP1, involved in Charcot-Marie-Tooth disease, characterize a novel class of glutathione transferase-related genes. Mol. Biol. Evol. 21, 176–187.

McGonigle, B., Keeler, S.J., Lau, S.M., Koeppe, M.K., O'Keefe, D.P., 2000. A genomics approach to the comprehensive analysis of the gluta-thione S-transferase gene family in soybean and maize. Plant Physiol. 124, 1105–1120.

Mueller, L.A., Goodman, C.D., Silady, R.A., Walbot, V., 2000. AN9, a petunia glutathione S-transferase required for anthocyanin sequestration, is a flavonoid-binding protein. Plant Physiol. 123, 1561–1570.

Nicholas, K.B., Nicholas, J.H.B., Deerfield, I.D.W., 1997. GeneDoc: analysis and visualization of genetic variation. EMBNEW. News 2, 14.

Ricci, G., Caccuri, A.M., Lo Bello, M., Pastore, A., Piemonte, F., Federici, G., 1994. Colorimetric and fluorometric assays of glutathione trans-ferase based on 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole. Anal. Biochem. 218, 463–465.

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

Schwede, T., Kopp, J., Guex, N., Peitsch, M.C., 2003. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res. 31, 3381–3385.

Sheehan, D., Meade, G., Foley, V.M., Dowd, C.A., 2001. Structure, function and evolution of glutathione transferases: implications for classifi-cation of non-mammalian members of an ancient enzyme superfamily. Biochem. J. 360, 1–16.

Snyder, M.J., Maddison, D.R., 1997. Molecular phylogeny of glutathione-S-transferases. DNA Cell Biol. 16, 1373–1384.

Soranzo, N., SariGorla, M., Mizzi, L., De Toma, G., Frova, C., 2004. Organisation and structural evolution of the rice glutathione S-transferase gene family. Mol. Genet. Genomics 271, 511–521.

Thom, R., Cummins, I., Dixon, D.P., Edwards, R., Cole, D.J., Lapthorn, A.J., 2002. Structure of a tau class glutathione S-transferase from wheat active in herbicide detoxification. Biochemistry 41, 7008–7020.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The Clustal × windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24, 4876–4882.

Wagner, U., Edwards, R., Dixon, D.P., Mauch, F., 2002. Probing the diversity of the arabidopsis glutathione S-transferase gene family. Plant Mol. Biol. 49, 515–532.

Wang, X.R., Szmidt, A.E., Savolainen, O., 2001. Genetic composition and diploid hybrid speciation of a high mountain pine, *Pinus densata*, native to the Tibetan plateau. Genetics 159, 337−346.

Yu, M., Facchini, P.J., 2000. Molecular cloning and characterization of a type III glutathione S-transferase from cell suspension cultures of opium poppy treated with a fungal elicitor. Physiol. Plant 108, 101−109.

Zeng, Q.Y., Lu, H., Wang, X.R., 2005. Molecular characterization of a glutathione transferase from *Pinus tabulaeformis* (Pinaceae). Biochimie 87, 445−455.

Zeng, Q.Y., Wang, X.R., 2005. Catalytic properties of glutathione-binding residues in a tau class glutathione transferase (PtGSTU1) from *Pinus tabulaeformis*. FEBS Lett. 579, 2657−2662.