PROTEINS: Structure, Function, and Bioinformatics 65:527–537 (2006)

SHORT COMMUNICATION

Comparative Structural Analysis of a Novel Glutathione *S*-Transferase (Atu5508) from *Agrobacterium tumefaciens* at 2.0 Å Resolution

Mickey Kosloff,[1†] Gye Won Han,[2,3†] S. Sri Krishna,[2,4,5] Robert Schwarzenbacher,[2,5] Marc Fasnacht,[1,6] Marc-André Elsliger,[2,3] Polat Abdubek,[2,7] Sanjay Agarwalla,[2,7] Eileen Ambing,[2,7] Tamara Astakhova,[2,5] Herbert L. Axelrod,[2,8] Jaume M. Canaves,[2,5] Dennis Carlton,[2,3] Hsiu-Ju Chiu,[2,8] Thomas Clayton,[2,3] Michael DiDonato,[2,7] Lian Duan,[2,5] Julie Feuerhelm,[2,7] Carina Grittini,[2,3] Slawomir K. Grzechnik,[2,5] Joanna Hale,[2,7] Eric Hampton,[2,7] Justin Haugen,[2,7] Lukasz Jaroszewski,[2,4,5] Kevin K. Jin,[2,8] Hope Johnson,[2,3] Heath E. Klock,[2,7] Mark W. Knuth,[2,7] Eric Koesema,[2,7] Andreas Kreusch,[2,7] Peter Kuhn,[2,3] Inna Levin,[2,8] Daniel McMullan,[2,7] Mitchell D. Miller,[2,8] Andrew T. Morse,[2,5] Kin Moy,[2,3] Edward Nigoghossian,[2,7] Linda Okach,[2,7] Silvya Oommachen,[2,8] Rebecca Page,[2,3] Jessica Paulsen,[2,7] Kevin Quijano,[2,7] Ron Reyes,[2,8] Christopher L. Rife,[2,8] Eric Sims,[2,3] Glen Spraggon,[2,7] Vandana Sridhar,[2,3] Raymond C. Stevens,[2,3] Henry van den Bedem,[2,8] Jeff Velasquez,[2,3] Aprilfawn White,[2,7] Guenter Wolf,[2,8] Qingping Xu,[2,8] Keith O. Hodgson,[2,8] John Wooley,[2,5] Ashley M. Deacon,[2,8] Adam Godzik,[2,4,5] Scott A. Lesley,[2,3,7] and Ian A. Wilson[2,3*]

[1]*Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York*
[2]*Joint Center for Structural Genomics*
[3]*The Scripps Research Institute, La Jolla, California*
[4]*Burnham Institute for Medical Research, La Jolla, California*
[5]*Center for Research in Biological Systems, University of California, San Diego, La Jolla, California*
[6]*Howard Hughes Medical Institute*
[7]*Genomics Institute of the Novartis Research Foundation, San Diego, California*
[8]*Stanford Synchrotron Radiation Laboratory, Stanford University, Menlo Park, California*

ABSTRACT **Glutathione *S*-transferases (GSTs) comprise a diverse superfamily of enzymes found in organisms from all kingdoms of life. GSTs are involved in diverse processes, notably small-molecule biosynthesis or detoxification, and are frequently also used in protein engineering studies or as biotechnology tools. Here, we report the high-resolution X-ray structure of Atu5508 from the pathogenic soil bacterium *Agrobacterium tumefaciens* (atGST1). Through use of comparative sequence and structural analysis of the GST superfamily, we identified local sequence and structural signatures, which allowed us to distinguish between different GST classes. This approach enables GST classification based on structure, without requiring additional biochemical or immunological data. Consequently, analysis of the atGST1 crystal structure suggests a new GST class, distinct from previously characterized GSTs, which would make it an attractive target for further biochemical studies. Proteins 2006;65:527–537.**

Key words: glutathione transferase; X-ray crystallography; consensus sequence; local structural motif; active site; dimer interface; specificity; JCSG; structural genomics

INTRODUCTION

Agrobacterium tumefaciens is a pathogenic soil organism that is responsible for crown gall, the plant disease that causes large tumor-like growth in over 90 families of plants, and results in major agronomical losses. The *Atu5508* gene from *Agrobacterium tumefaciens* encodes a protein with a molecular weight of 25,904 Da (residues 1–236) and a calculated isoelectric point of 4.8. Atu5508 has been annotated as a hypothetical protein in NCBI sequence databases and as a putative glutathione *S*-transferase (GST) in the Joint Center

Grant sponsor: National Institutes of Health, Protein Structure Initiative; Grant numbers: P50 GM62411 and U54 GM074898; Grant sponsor: National Institute of General Medical Sciences.

[†]Mickey Kosloff and Gye Won Han contributed equally to this work.

*Correspondence to: Dr. Ian Wilson, JCSG, The Scripps Research Institute, BCC206, 10550 North Torrey Pines Road, La Jolla, CA 92037. E-mail: wilson@scripps.edu

Received 6 February 2006; Revised 2 June 2006; Accepted 13 June 2006

Published online 20 September 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21130

for Structural Genomics (JCSG) database, based on statistically significant sequence similarity detected by PSI-BLAST.[1] In general, GSTs catalyze the transfer of the tripeptide glutathione (γ-glutamyl-cysteinyl-glycine; GSH) to a cosubstrate that contains a reactive electrophilic centre to form a polar *S*-glutathionylated reaction product.

GSTs comprise a large superfamily of enzymes that are ubiquitous in organisms ranging from bacteria to humans. This protein family has been the focus of intensive research not only because of its important biological roles, but also due to its multiple biomedical and biotechnology applications. The multiple roles and applications of GSTs encompass: (1) Major components of small-molecule detoxification pathways in humans, especially for electrophilic compounds that include mutagens, carcinogens, and drugs; (2) Conferral of antibiotic resistance in bacteria and herbicide resistance in plants; (3) Biosynthesis of metabolites; (4) Biotechnology tools, notably used in fusion with other proteins of interest; and (5) Candidates for directed evolution and design of new enzymatic catalysts.[2,3] GSTs form a very diverse protein family and, therefore, have been subdivided into an ever-increasing number of subfamilies (referred to here as "GST classes"), associated with different functionalities and enzymatic properties.[4,5] This classification has usually been based on a combination of criteria, such as biochemical properties, primary, tertiary, and quaternary structure and immunological reactivity.[3]

Here, we report the crystal structure of Atu5508 from *Agrobacterium tumefaciens* that was determined using the semiautomated high-throughput pipeline of the JCSG.[6] Our analysis of the sequence and structure of Atu5508 indicate that it defines a new GST class, distinct from other previously characterized GSTs.

MATERIALS AND METHODS

Protein Production and Crystallization

A putative glutathione S-transferase, Atu5508 (GI: 15162326, Swissprot: Q8UJG9) was amplified by PCR from genomic DNA from *A. tumefaciens*, using TAQ polymerase and primers corresponding to the predicted 5′- and 3′-ends. The PCR product, representing residues 1–236 of Atu5508 was cloned into plasmid pMH1, which encodes an expression and purification tag (MGSDKIHHHHHH) at the amino terminus of the full-length protein. The cloning junctions were confirmed by DNA sequencing. Protein expression was performed in selenomethionine-containing medium using the *E. coli* methionine auxotrophic strain DL41. Bacteria were lysed by sonication in lysis buffer (50 m*M* KPO_4, pH 7.8, 300 m*M* NaCl, 10% glycerol, 5 m*M* imidazole, 0.25 m*M* TCEP, benzonase, PefaBloc protease inhibitor cocktail). Immediately after sonication, the cell debris was pelleted by ultracentrifugation at 158,000*g* for 60 min (4°C). The soluble fraction was applied to a gravity flow metal chelate column (Talon resin charged with cobalt; Clontech) equilibrated in lysis buffer. The column was then washed with seven column volumes (CV) of wash buffer (25 m*M* Tris, pH 7.8, 15 m*M* NaCl, 5 m*M* imidazole) and eluted with 3 CV of elute buffer (25 m*M* Tris, pH 7.8, 15 m*M* NaCl, 150 m*M* imidazole). The protein-containing fraction was loaded onto a 1 × 10 cm column packed with Poros HQ anion exchange resin and equilibrated in 25 m*M* Tris, pH 7.8, 0.25 TCEP. Following a 2 CV wash, the protein was eluted in a 15 CV linear 0–1*M* NaCl gradient. Pooled fractions were buffer exchanged into crystallization buffer (10 m*M* Tris, pH 7.8, 100 m*M* NaCl, 0.25 m*M* TCEP) and concentrated to 36 mg/mL by centrifugal ultrafiltration (Orbital). The protein was either frozen in liquid nitrogen for later use or used immediately for crystallization trials. The protein was crystallized using the nanodroplet vapor diffusion method[7] with standard JCSG crystallization protocols.[6] The crystallization solution contained 18% PEG 3350, 0.2*M* potassium thiocyanate, and 10% glycerol with an additive of 10 m*M* spermine tetra-HCL. The cryosolution also contained 10% MPD. The crystals were indexed in monoclinic space group C2 (Table I).

Data Collection

Anomalous diffraction data were collected at the Advanced Photon Source (APS, Chicago, USA) on beamline 19-BM at wavelengths corresponding to peak, remote, and inflection (λ_1, λ_2, and λ_3), of a selenium multiple-wavelength anomalous diffraction (MAD) experiment (Table I). The data sets were collected at 100 K using an APS1 CCD detector. Data were integrated, reduced, and scaled using the HKL2000 suite.[10] Data statistics are summarized in Table I.

Structure Solution and Refinement

Phases to 2.5 Å resolution were calculated with the CCP4 suite,[8] SHELXD,[11] and AUTOSHARP[12] using 11 Se positions. Automatic model building was performed with RESOLVE[13] and ARP/wARP.[14] Model completion and refinement were performed using the programs O,[15] COOT,[16] and REFMAC5.[8] The final refinement at 2.0 Å resolution was performed using the peak data (λ_1). Refinement statistics are summarized in Table I.

Structure Analysis and Deposition

Analysis of the stereochemical quality of the model was accomplished using AutoDepInputTool,[17] MolProbity,[18] SFcheck 4.0,[19] and WHAT IF 5.0.[20] The protein quaternary structure was analyzed using the PQS server.[21] Figure 1(B) was adapted from a PDBsum[22] analysis and all other figures were prepared using PyMOL (DeLano Scientific). Atomic coordinates and experimental structure factors of Atu5508 (gi:15162326) have been deposited in the PDB and are accessible under the code 2fno.

Phylogenetic Tree Construction

To assess the evolutionary relationship of atGST1 (Atu5508) to other members of the GST superfamily in SCOP, we performed a phylogenetic analysis. We selected a

TABLE I. Summary of Crystal Parameters, Data Collection, and Refinement Statistics for Atu5508 (PDB: 2fno)

Data collection	λ_1MADSe	λ_2MADSe	λ_3MADSe
Wavelength (Å)	0.9791	0.9641	0.9793
Resolution range (Å)	44.7–2.00	48.2–2.00	44.7–1.90
Number of observations	116,425	127,502	69,197
Number of reflections	37,755	31,807	28,241
Completeness (%)	95.0 (78.1)[a]	80.2 (28.5)[a]	61.6 (6.1)[a]
Mean $I/\sigma(I)$	18.9 (5.1)[a]	19.2 (3.9)[a]	16.3 (2.7)[a]
R_{sym} on I	0.08 (0.20)[a]	0.09 (0.17)[a]	0.07 (0.20)[a]
Highest resolution shell (Å)	2.07–2.00	2.07–2.00	1.97–1.90
Model and refinement statistics			
Resolution range (Å)	44.7–2.00		
No. of reflections (total)	37,729[b]		
No. of reflections (test)	1,871		
Completeness (% total)	95.0		
Data set used in refinement	λ_1		
Cutoff criteria	$\|F\| > 0$		
R_{cryst}	0.181		
R_{free}	0.219		
Deviation from ideal geometry (RMS)			
Bond length	0.018 Å		
Bond angle	1.51°		
Average isotropic B-value protein	14.5 Å^2		
Average isotropic B-value ions	11.9 Å^2		
Average isotropic B-value water	20.8 Å^2		
ESU based on R value	0.17 Å		
Protein residues/atoms	473/3,617		
Ions/atoms	4/12		
Water molecules	394		

Space group is C2.
The unit cell parameters are $a = 110.20$ Å, $b = 50.26$ Å, $c = 117.66$ Å, $\beta = 115.72°$.
ESU = Estimated overall coordinate error.[8,9]
$R_{sym} = \Sigma|I_i - \langle I_i \rangle| / \sum |I_i|$ where I_i is the scaled intensity of the ith measurement, and $\langle I_i \rangle$ is the mean intensity for that reflection.
$R_{cryst} = \Sigma| \ |F_{obs}| - |F_{calc}| \ | \ / \Sigma |F_{obs}|$ where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively.
R_{free} = as for R_{cryst}, but for 5.0% of the total reflections chosen at random and omitted from refinement.
[a]Highest resolution shell in parentheses.
[b]Typically, the number of unique reflections used in refinement is slightly lesser than the total number that were integrated and scaled. Reflections are excluded because of systematic absences, negative intensities, and rounding errors in the resolution limits and cell parameters.

subset of proteins from each of the GST structural classes in this superfamily and performed a multiple-sequence alignment using the T-coffee program.[23] Sequence alignments were generated using both the complete protein sequences and those of only the N-terminal domain (which is considered less divergent and can be more reliably aligned across classes). The PDB IDs of the proteins selected from each GST class are alpha: 1ev4, 1f3a, 1ml6, 1gsd, 1guk, 1gul, 1oe7; mu: 1c72, 2gtu, 6gsv, 3gtu; pi: 11gs, 4gss, 1bay, 2gsrA; sigma: 1pgt, 1iyh, 1gsq; beta: 1a0f, 2pmt, 1f2e; delta: 1jlv, 1jlw; phi: 1aw9, 1bx9, 1axd; theta: 1ljr; zeta: 1e6b, 1fw1; omega: 1eem; tau: 1gwc, 1oyj; Yeast prion protein Ure2p, nitrogen regulation fragment: 1g6w; GST-like domain of elongation factor 1-gamma: 1nhy; Glutaredoxin 2: 1g7o; Chloride intracellular channel 1: 1k0m; and pfGST: 1okt. We then used the ClustalW software[24] to build neighbor joining trees. To estimate the reliability of the segregations in the tree, we used the bootstrap method with 1000 trials, as implemented in the ClustalW package. Graphical representations of the trees were generated using the Treeview software.[25]

Profile–Profile Alignments

Atu5508 (atGST1) was aligned with all other members of the GST superfamily in the SCOP structure database (version 1.65),[26] using hybrid profile-profile alignments, as implemented in HMAP.[27,28] HMAP uses hybrid multidimensional profile alignment that combines sequence, secondary and tertiary structure to facilitate the detection of remote homologs. GST structures that were not in the SCOP database were also aligned with atGST1 using HMAP.

Multiple-Structure Comparisons

Representative GST structures that had similar architecture and significant sequence similarity (by HMAP profile-

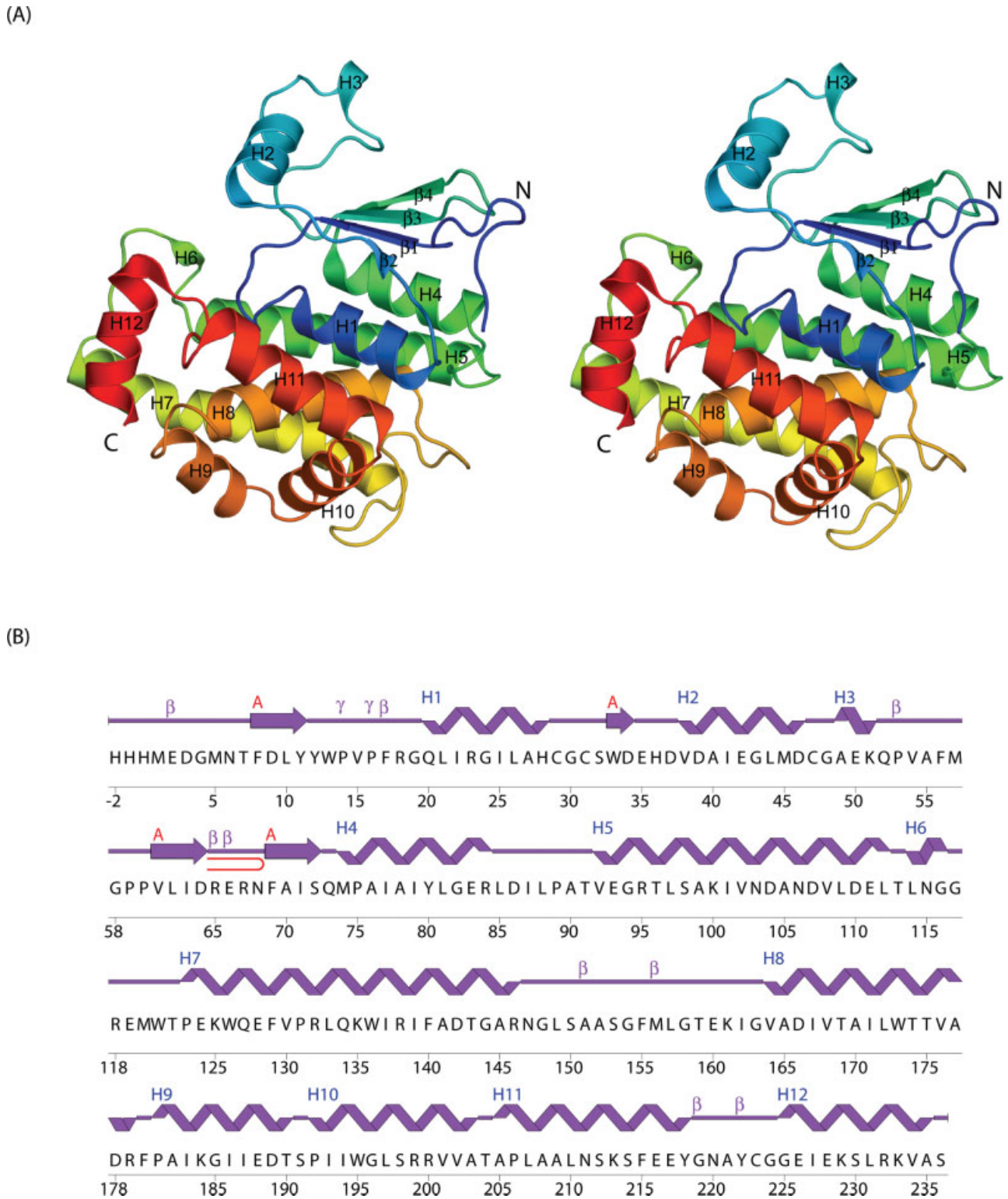

Fig. 1. Crystal structure of atGST1: (**A**) Stereo ribbon diagram of atGST1 monomer color-coded from N-terminus (blue) to C-terminus (red). Helices (H1–H12) and β-strands (β1–β4) are labeled. (**B**) Diagram showing the secondary structural elements in atGST1 superimposed on its primary sequence. The α-helices, 3_{10}-helix, β-strands of sheet A, β-bulges, and γ-turns are indicated. The β-hairpin is depicted as a red loop.

profile alignment) were chosen from the individual families in the SCOP GST superfamily (http://scop.mrc-lmb.cam.ac.uk/scop/data/scop.b.b.fd.b.A.html) and are listed in Table II. In general, the representative structure from each SCOP family with the highest sequence identity to atGST1 was chosen.

Multiple-structure superimpositions were performed using the Grasp2 package,[29] which combines secondary structure elements (SSEs) superimposition with C_α RMSD minimization. GST class-specific motifs or residues that had been identified in the literature were compared both in the superimposed structures and in the structure-based sequence alignments. Motifs that distinguish between the different classes at the structural level were identified (Table II) and, for each motif, atGST1 was compared with each class to assess similarities and differences.

RESULTS AND DISCUSSION

The crystal structure of atGST1 [Fig. 1(A)] was determined at 2.0 Å by the MAD method. Data collection, model, and refinement statistics are summarized in Table I. The final model includes two monomers in the asymmetric unit. Each monomer contains 236 residues (residue 1 is disordered in chain A and residues 1–3 are disordered in chain B), 4 SCN^- ions, and 394 water molecules. Electron density also was observed for three residues of the expression/purification tag in chain A (His −2, His −1, and His 0). No electron density was observed for the side chains of His(−2), Ser32, Lys51, Arg67, Arg118, Gln127, Glu160, Lys213, Lys233 in chain A, and residues Asp3, Ser32, Lys51, Arg118, Gln127, Glu160, Lys213, Glu216, Glu217 in chain B.

The Matthews' coefficient (V_m)[30] is 2.61 $Å^3$/Da and the estimated solvent content is 52.4%. The Ramachandran plot, produced by MolProbity,[18] shows that 98.25%, 99.56%, and 0.44% of the residues are in favored, allowed and disallowed regions, respectively. The only residue in the disallowed regions of the Ramachandran plot corresponds to Gln73 of the A and B monomers, but has unambiguous electron density. The GST monomer consists of four β-strands (β1–β4), ten α-helices (H1–H2, H4–H5, H7–12), and two 3_{10}-helices (H3, H6) [Fig. 1(A,B)]. The total β-strand, α-helical, and 3_{10}-helical content is 5.9%, 57.3%, and 2.5%, respectively. Similar to other GSTs, atGST1 is composed of two domains: a smaller N-terminal domain (residues 2–85) followed by a short linker region (residues 86–92) and a larger C-terminal domain (residues 93–236) [Fig. 1(A)]. The N-terminal domain belongs to the thioredoxin fold,[31,32] while the C-terminal domain belongs to the "GST-C-terminal domain" α-helical fold. In most GSTs, the glutathione substrate binds in the so-called "G-site", while the electrophilic cosubstrate that becomes covalently conjugated to the glutathione binds in the usually hydrophobic "H-site". As the two binding sites are in the dimer interface, many of the features of this interface are inevitably related to the binding and catalytic properties of GSTs.

Analysis of the crystallographic packing of Atu5508 using the PQS server[21] identified a crystallographic dimer, for which a significant portion of the interface is formed by interactions of α-helices H3–H4 and two short loop regions. This interface buries a surface area of 1900 $Å^2$ upon complex formation [Fig. 2(B)]. Comparison of this interface with other biologically relevant GST dimers [e.g. PDB 6gsv, Fig. 2(A)] indicates that this crystallographic dimer is the biologically relevant oligomeric form. The interaction of the so-called "key" (Phe56), after rotation of this side-chain to the position seen in 6gsv [Fig. 2(A)], would result in its interaction with the patch of hydrophobic residues in the adjacent monomer (the "lock") and further extend this dimer interface. The asymmetric unit itself contains two distinct monomers that interact through α-helices H9–H10 and two short loop regions. This second noncrystallographic interface features a much smaller buried surface area of 358 $Å^2$ per monomer and is, therefore, less likely to be biologically relevant.

A search using the DALI server[33] found structural similarities to members from different GST classes. The top hit was pi GST (PDB: 1pgt, $Z = 20.7$) with an RMSD of 2.6 Å for this structural alignment of 204 residues (C_α atoms) with 19% sequence identity. Other high scoring DALI hits include sigma GST (PDB 1pd2, $Z = 18.3$), phi GST (PDB 1hqo, $Z = 17.6$) and beta GST (PDB 1b8x, $Z = 16.8$). Significant structural similarities to proteins from other GST classes (tau, omega, zeta, theta) are also among the high scoring hits ($Z > 9$).

Although atGST1 can be reliably classified as a GST based on sequence analysis, neither the sequence nor the structure searches found a homolog with a sequence identity of >20%. It is generally accepted that GSTs with <30% sequence identity are assigned to a separate class.[3] However, in the absence of any biochemical data, it was not clear whether atGST1 could be classified into any of the already defined GSTs classes or whether it should define a new class based on sequence comparison alone. Compared with its distant homologs, atGST1 seemed to have some features of the alpha class and the malaria parasite pfGST, but lacks some of the defining features of each of these.

These observations were corroborated by phylogenetic analysis [Fig. 3(A,B)], which indicate that atGST1 forms a new branch, not grouped within any GST families (as classified by SCOP), indicating divergence from previously characterized GST classes. It should be noted, however, that the bootstrap value for the node separating atGST1 from the neighboring branches (containing GSTs from the alpha, mu, pi, and sigma families) was 635 out of 1000 trials, with even lower values elsewhere in the tree, indicating that the significance of this separation is low. The bootstrap value increased to 735 out of 1000 trials, if only the N-terminal domain of these proteins (that are usually less divergent between classes and, therefore, can be more reliably aligned) were used in the alignment.

We then analyzed representative structures of the different GST classes using profile-based methods and multiple-structure alignments. Together with the extensive literature available on GSTs, we identified local struc-

TABLE II. GST Class-Specific Structural Motifs and Their Corresponding Residues in atGST1

		I	II	III	IV	V	VI	VII
GST class	Representative PDB	N′catalytic tyrosine[a]	Mu loop (β2-α2)	Hydrophobic "Key" α2–β3[b]	SNAIL/TRAIL motif (α3)[c]	Catalytic Asp	Catalytic histidine (bacteria)	C′ extension[d]
Pi	1bay	**Y7**	—	(Y49)	[65]SNAIL	**D98**	—	(+) 202–209
Mu	3gtu	**Y10**	38–44	**F60**	[76]SNAIL	**D109**	—	(+) 210–224
Beta	1a0f	**Y5**	—	—	**[[66]GVAIM]**	(E104)	H106	—
Phi	1axd	(S11)	—	—	[67]SRAIC	—	—	—
Zeta	1e6b	(S13)	—	(M57)	[73]SFAII	—	—	—
Omega	1eem	(C32)	—	—	[86]SAITC	—	—	**(+, α)** 222–241
Alpha	1f3a	**Y8**	—	**F51**	[67]TRAIL	**D100**	—	**(+, α)** 206–221
Sigma	1gsq	**Y7**	—	—	[63]SMCIA	**D96**	—	—
Tau	1gwc	(S15)	—	—	[69]SMIIL	**D107**	—	**(+, α)** 215—225
Theta	1ljr	(S11)	—	—	[67]SSAIL	**D104**	—	**(+, α)** 213—244
pfGST	1okt	**Y9**	37–41	**F56**	[72]SQAIV	**D105**	H107	—
atGST1	**2fno**	**Y12**	—	**F56**	**[[74]MPAIA]**	**D107**	—	**(+, α) 219–236**

Motifs marked in boldface are also found in atGST1.

[a]Residues in round brackets mark residues in a structure that differ (in identity and position) from the noted motif but, in accordance with the literature, might substitute for its catalytic function.

[b]The hydrophobic lock residues are in α4 and α5 of the opposing monomer. The key motif (with a tyrosine substituting for the phenylalanine) is present in the Pi class, according to the literature, but its dimer interface is different than in the other GST classes. However, it cannot be visualized in 1bay because of missing residues. The corresponding residue in PDB id 13gs (another representative of the same family) is Y49.

[c]The SNAIL/TRAIL motif is present in most GSTs and contributes polar functional groups to the glutathione (G) binding site, located in the dimer interface. Square brackets denote the corresponding residues in a structure, which do not match the conserved functional motif.

[d]A "+" denotes the presence of a C-terminal extension, "α" corresponds to an α-helical extension.

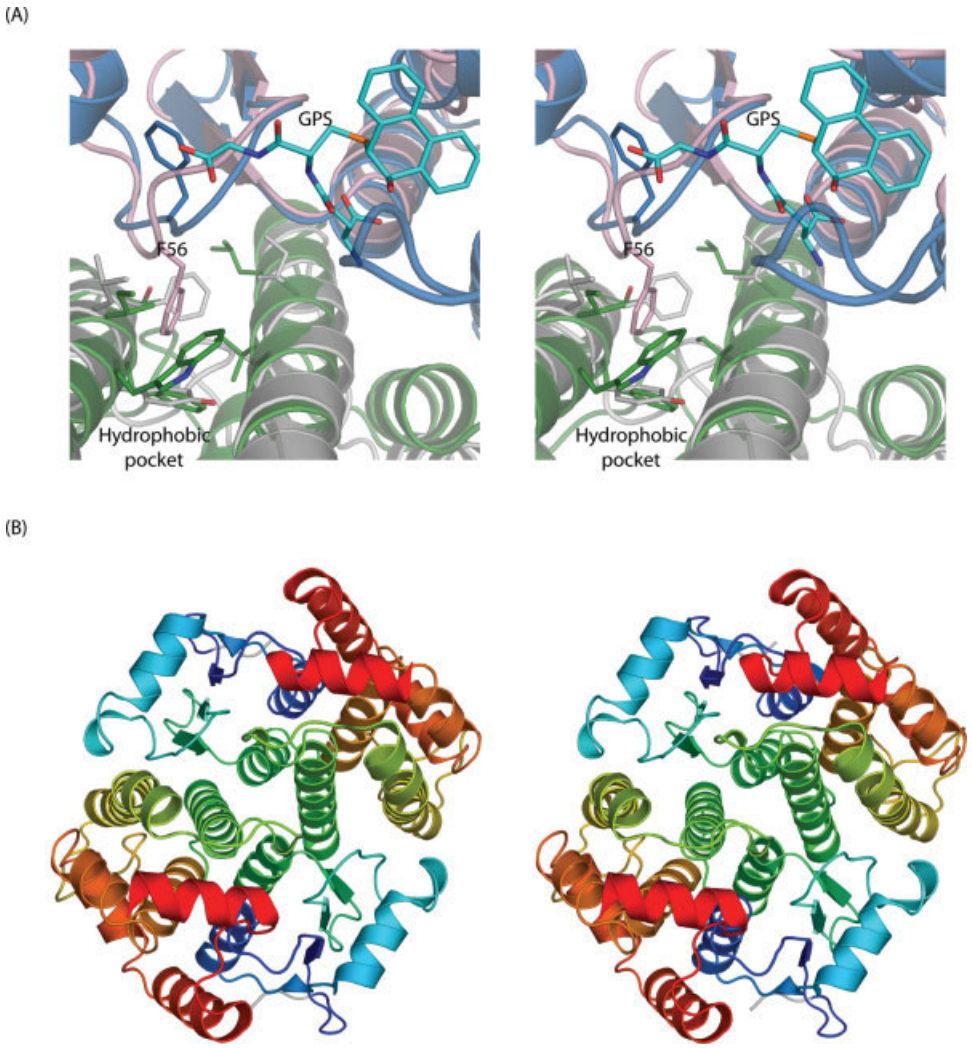

Fig. 2. Dimerization of atGST1: (**A**) Stereo diagram of the hydrophobic key (Phe56) and lock (Leu97, Ile101, Ala105, Trp136, Ile139, Phe140, and Thr143) motif. The atGST1 lock-and-key motif is shown superimposed on the corresponding regions from μ-GST (6gsv). The ligand in the active site of 6gsv, (9S,10S)-9-(*S*-glutathionyl)-10-hydroxy-9,10-dihydrophenanthrene (GPS) is shown in ball-and-stick. The atGST1 dimer is colored in forest green and blue, while the μ-GST dimer is colored in grey and pink. (**B**) Stereo diagram of the biologically relevant dimer of atGST1, color-coded from N-terminus (blue) to C-terminus (red).

tural signatures that define each class, most of which consist of single residues or short, but not necessarily contiguous, structural motifs (Fig. 4 and Table II). By combining these characteristics, we could differentiate the previously characterized GST classes from one another based only on the presence or absence of these motifs. These structural signatures have corresponding functional significances, such as differences in catalytic properties or selective dimer formation only between members of a specific GST class. These motifs, detailed in Table II and shown in Figure 4, include (i) a tyrosine (Tyr12 in atGST1) in the N-terminal domain that is a conserved catalytic residue in several GST classes,[34] (ii) the μ-loop (not present in atGST1), (iii) the hydrophobic "lock-and-key" motif, (iv) the SNAIL/TRAIL motif (not present in atGST1), (v) a catalytic aspartic acid at the beginning of the C-terminal domain, (vi) a catalytic histidine (not present in atGST1) and (vii) a divergent C-terminal region, which in some classes assumes an α-helical conformation. A conserved *cis*-proline[35] found in all representative GSTs, is similarly conserved as Pro60 in atGST1.

Although *Agrobacterium* is pathogenic to plants, atGST1 is not closely related to GSTs in other bacterial or plant organisms in which GSTs of the pi, sigma, phi, and beta class are found. Comparison of the class-specific motifs of the various GST classes revealed that atGST1 is different from all of these classes. For example, comparison with beta (bacterial) GSTs showed that catalytic residues conserved in this class (Cys10 and

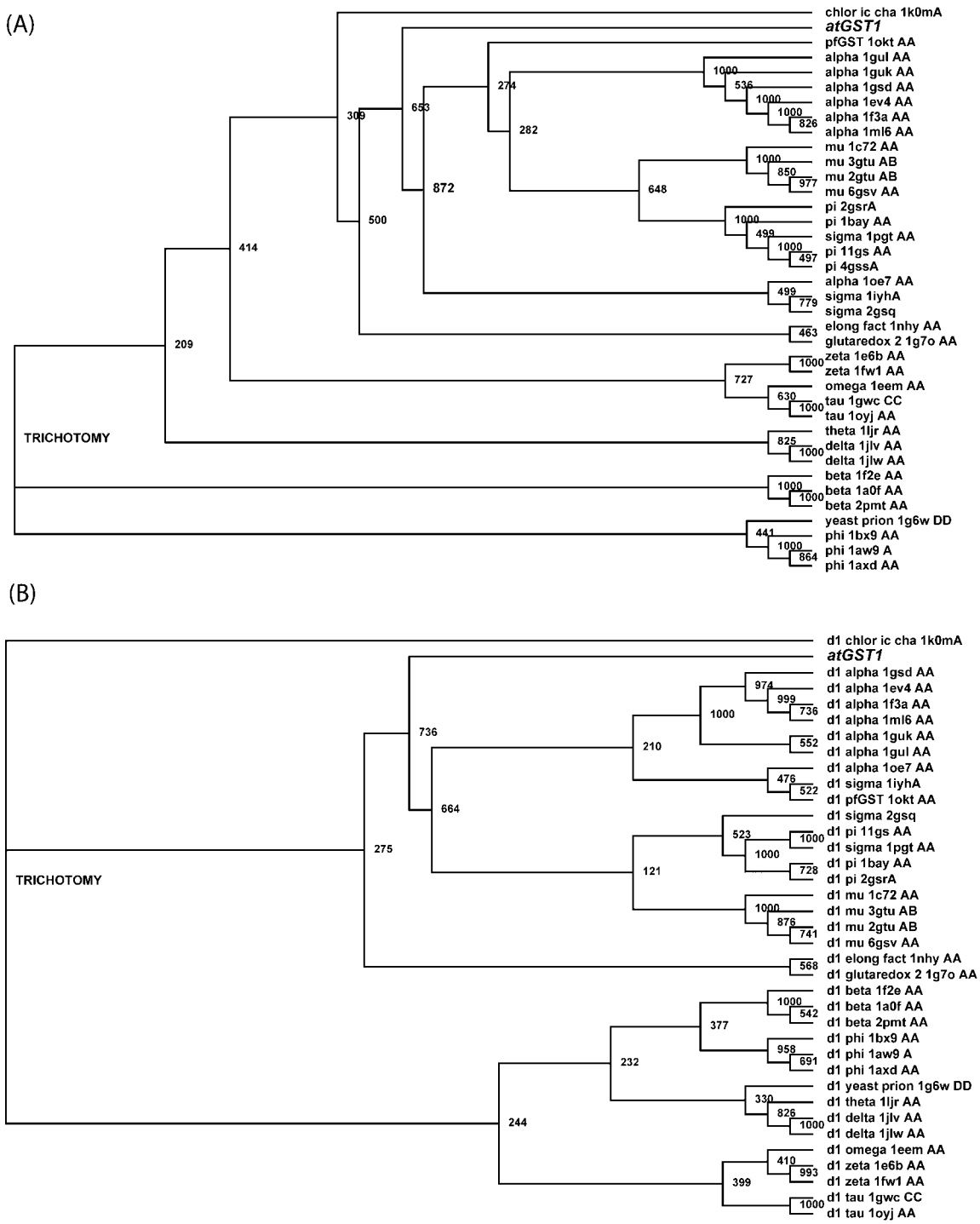

Fig. 3. Phylogenetic trees of representative GSTs and atGST1: (**A**) Phylogenetic tree from a T-coffee alignment of the full-length GST sequences shown in rectangular cladogram form with bootstrap values. (**B**) Phylogenetic tree from a T-coffee alignment of the GST N-terminal domain sequences shown in rectangular cladogram form with bootstrap values.

His106 of PDB 1a0f) are absent in atGST1. Also, beta GSTs do not contain the hydrophobic lock-and-key motif that is found in atGST1.

An HMAP profile–profile alignment determined that pfGST from the malaria parasite was one of the most similar GST structures to atGST1 (Sequence identity of ~20%). The structure of pfGST (1okt) was previously shown to define a novel GST class.[36] Comparing this structure with atGST1 revealed that key functional and structural features are not conserved between the two

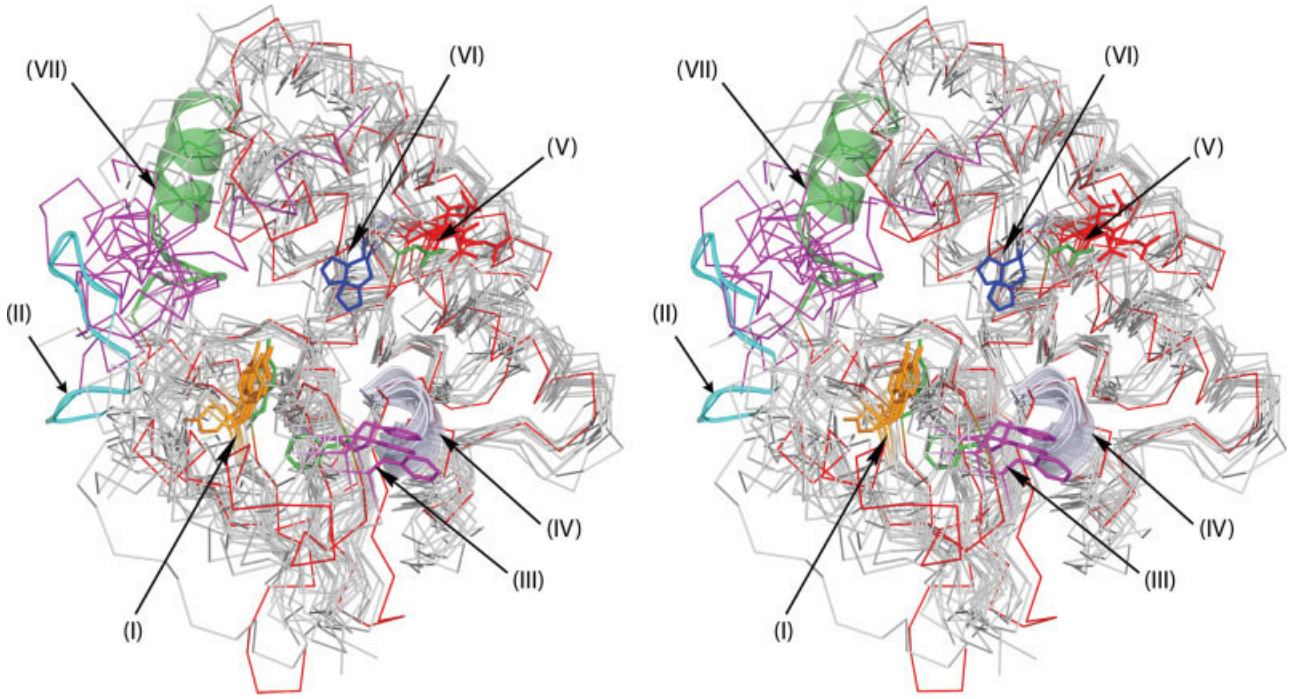

Fig. 4. GST class-specific motifs shown in the context of a multiple-structure alignment of representative GSTs. Stereo representations of the C_α trace of atGST1 is colored red and all other GSTs are shown in grey. The seven motifs that define the various GST classes are labeled: (I) N′ catalytic tyrosine (ball-and-stick, orange), (II) mu loop (cyan), (III) hydrophobic "Key" (ball-and-stick, magenta), (IV) SNAIL/TRAIL motif (grey helix), (V) catalytic Asp that binds glutathione across the dimer interface (ball-and-stick, red), (VI) catalytic histidine (ball-and-stick, blue), and (VII) C' extension (C_α trace, magenta). For clarity, the corresponding motifs, when present in atGST1, are shown in forest green and arrows indicate the general positions of these motifs.

proteins; pfGST contains a variation of the "SNAIL motif", has a short "mu loop", and a putative catalytic histidine in position 107, all of which are not present in atGST1.

Quite a few of the class-specific motifs are located in the GST dimer interface. Dimerization is essential for stability and function,[37] and therefore, analysis of the dimer interface is crucial in understanding GST function and specificity. One such feature, generally considered important for dimerization, is the hydrophobic "lock-and-key" motif [Figs 2(A) and 4]. This motif is conserved in alpha, mu, and pi GSTs[3] and in pfGST. The "key" is an aromatic residue in one monomer and the "lock" is a cluster of hydrophobic residues from the other interacting subunit. atGST1 seems to have this conserved feature (Phe56 corresponds to the "key" residue) in the appropriate structural location, but the aromatic side-chain points away from the dimeric interface, possibly because of the absence of bound ligands [Fig. 2(A)]. This observation is consistent with previous investigations of hGST-A1-1, a member of the alpha class, where dimer formation was still maintained after mutations of this aromatic residue to a serine, but its catalytic function and dimer stability were severely reduced.[38,39]

As indicated previously, Gln73 from both the A and B chains lies in the disallowed region of the Ramachandran plot. Interestingly, a structural alignment of the best scoring DALI hit (1pgt) reveals that this glutamine (Gln64 in the 1pgt structure) is oriented toward the substrate cavity, within hydrogen bonding distance to the substrate. The glutamine side-chain adopts a similar orientation in our structure. Similar side-chain orientations are observed in the structures from the pi, mu, alpha, sigma, and pfGST classes, suggesting a conserved functional role of Gln73 in substrate binding.

Almost all of the structural motifs analyzed above are functional determinants of atGST1, namely they play a significant role in catalytic function and substrate specificity. In the N-terminal domain, the complement of functional motifs that enable glutathione binding and catalysis[4] are indeed present in atGST1 (Table II), including the conserved *cis*-proline (Pro60), the catalytic N-terminal tyrosine (Tyr12), and the hydrophobic "key" (Phe56). The higher conservation of the N-terminal domain and the presence of these functional motifs affirm the functionality of glutathione binding in the G site. Indeed, after testing the GST function of atGST1 using a standard functional GST assay (35–141 μg of protein, 25°C, as described in Habig and Jakoby[40]) with GSH and *p*-nitrobenzyl chloride, we measured a specific activity of 0.22 ± 0.03 μmol/min/mg.

Interestingly, the presence of both the N-terminal tyrosine and the lock-and-key motifs is the hallmark of the alpha, mu, pi, and pfGST classes (which are found in mammals or in mammalian parasites). The lock-and-key motif is not characteristic of previously described bacte-

rial or plant GSTs. Moreover, an arginine residue (Arg18) is present in atGST1 in the same location and orientation as the catalytic arginine that coordinates the glutathione sulfur in the alpha class.[4,41] This suggests that the catalytic mechanism for nucleophilic activation of glutathione in atGST1 is similar to that of alpha GSTs. Notably, pfGST has a lysine residue, which is important for catalysis,[42] at the position corresponding to the catalytic arginine.[36]

On the other hand, the C-terminal domain of atGST1 diverges significantly from other GSTs, suggesting differences in the substrate specificity of the H site.[3,4,41] In particular, the extreme C-terminal region of atGST1 (residues 214–236), which is dissimilar in sequence and conformation to the corresponding regions of other GSTs, deserves further consideration (Fig. 4, forest green, motif VII). A corresponding (but divergent) C-terminal helix (α9) in the alpha family plays a role in the catalytic cycle and confers distinctive substrate specificity.[4,40,43–45] The positioning of this C-terminal helix, which is located near the H site, alters the size of the substrate-binding pocket and, therefore, influences the specificity for the electrophilic substrate. In atGST1, the C-terminal region is longer than in most GSTs and adopts a distinctive loop-helix (H12) structure, suggesting unique catalytic properties and perhaps specific substrate recognition. Indeed, atGST1 catalyzes the conjugation of GSH to *p*-nitrobenzyl chloride, as do many other GSTs.[40] However, no GST activity was detected with the following substrates: 1-chloro-2,4-dinitrobenzene, 1,2-dichloro-4-nitrobenzene, *p*-nitrophenethyl bromide, and *trans*-4-phenyl-3-buten-2-1. Thus, the structural and functional characterization of atGST1 presented here should help direct further studies of the catalytic properties and electrophilic substrate specificity of this novel GST.

In conclusion, we have combined various structural motifs identified in the known GST classes that enable us to classify GSTs at the structural level, without requiring additional biochemical or immunological results. By comparing the structural similarities and differences of atGST1 to other GSTs (as detailed in Table II), we conclude that the atGST1 structure represents a novel GST class, distinct from previously defined GST classes. Surprisingly, the catalytically relevant structural motifs of atGST1 show a closer evolutionary relationship to mammalian GSTs than to previously characterized bacterial or plant GSTs. Its distinctiveness, therefore, makes atGST1 an attractive target for further biochemical studies that will aid in a better understanding of the evolution of GSTs. In particular, the substantial differences between its active site and those of plant GSTs suggest that atGST1 may be a promising target for designing specific inhibitors against *A. tumefaciens*. Interestingly, we found several close homologs of atGST1 in the set of environmental sequences determined recently by the environmental sequencing project carried out by the Craig Venter Institute (https://research.venterinstitute.org/moore). Therefore, it is clear that atGST1 is not a single member of this branch—this structure likely represents a larger family, which will grow as the existing sequence databases expand. Models for GST homologs can be accessed at http://www1.jcsg.org/cgi-bin/models/get_mor.pl?key=15162326.

ACKNOWLEDGMENTS

Portions of this research were carried out at the Advanced Photon Source (APS, Chicago, USA) and the Stanford Synchrotron Radiation Laboratory (SSRL). The SSRL is a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL structural molecular biology program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health (National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences). Use of the Argonne National Laboratory Structural Biology Center beamlines at the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Biological and Environmental Research, under Contract No. W-31-109-ENG-38. Mickey Kosloff is a postdoctoral fellow of the Human Frontiers Science Program.

REFERENCES

1. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
2. Salinas AE, Wong MG. Glutathione *S*-transferases—a review. Curr Med Chem 1999;6:279–309.
3. Sheehan D, Meade G, Foley VM, Dowd CA. Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. Biochem J 2001;360:1–16.
4. Armstrong RN. Structure, catalytic mechanism, and evolution of the glutathione transferases. Chem Res Toxicol 1997;10:2–18.
5. Armstrong RN. Mechanistic imperatives for the evolution of glutathione transferases. Curr Opin Chem Biol 1998;2:618–623.
6. Lesley SA, Kuhn P, Godzik A, Deacon AM, Mathews I, Kreusch A, Spraggon G, Klock HE, McMullan D, Shin T, Vincent J, Robb A, Brinen LS, Miller MD, McPhillips TM, Miller MA, Scheibe D, Canaves JM, Guda C, Jaroszewski L, Selby TL, Elsliger MA, Wooley J, Taylor SS, Hodgson KO, Wilson IA, Schultz PG, Stevens RC. Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. Proc Natl Acad Sci USA 2002;99:11664–11669.
7. Santarsiero BD, Yegian DT, Lee CC, Spraggon G, Gu J, Scheibe D, Uber DC, Cornell EW, Nordmeyer RA, Kolbe WF, Jin J, Jones AL, Jaklevic JM, Schultz PG, Stevens RC. An approach to rapid protein crystallization using nanodroplets. J Appl Crystallogr 2002;35:278–281.
8. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50:760–763.
9. Tickle IJ, Laskowski RA, Moss DS. Error estimates of protein structure coordinates and deviations from standard geometry by full-matrix refinement of γB- and βB2-crystallin. Acta Crystallogr D Biol Crystallogr 1998;54:243–252.
10. Borek D, Minor W, Otwinowski Z. Measurement errors and their consequences in protein crystallography. Acta Crystallogr D Biol Crystallogr 2003;59:2031–2038.
11. Schneider TR, Sheldrick GM. Substructure solution with SHELXD. Acta Crystallogr D Biol Crystallogr 2002;58:1772–1779.
12. Bricogne G, Vonrhein C, Flensburg C, Schiltz M, Paciorek W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta Crystallogr D Biol Crystallogr 2003;59:2023–2030.

13. Terwilliger T. SOLVE and RESOLVE: automated structure solution, density modification and model building. J Synchrotron Radiat 2004;11:49–52.
14. Cohen SX, Morris RJ, Fernandez FJ, Ben Jelloul M, Kakaris M, Parthasarathy V, Lamzin VS, Kleywegt GJ, Perrakis A. Towards complete validated models in the next generation of ARP/wARP. Acta Crystallogr D Biol Crystallogr 2004;60:2222–2229.
15. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A 1991;47: 110–119.
16. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. \Acta Crystallogr D Biol Crystallogr 2004;60:2126–2132.
17. Yang H, Guranovic V, Dutta S, Feng Z, Berman HM, Westbrook JD. Automated and accurate deposition of structures solved by X-ray diffraction to the protein data bank. Acta Crystallogr D Biol Crystallogr 2004;60:1833–1839.
18. Davis IW, Murray LW, Richardson JS, Richardson DC. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res 2004;32: W615–W619.
19. Vaguine AA, Richelle J, Wodak SJ. SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. Acta Crystallogr D Biol Crystallogr 1999;55:191–205.
20. Vriend G. WHAT IF: a molecular modeling and drug design program. J Mol Graph 1990;8:52–56, 29.
21. Henrick K, Thornton JM. PQS: a protein quaternary structure file server. Trends Biochem Sci 1998;23:358–361.
22. Laskowski RA, Chistyakov VV, Thornton JM. PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. Nucleic Acids Res 2005;33:D266–D268.
23. Notredame C, Higgins DG, Heringa J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol 2000;302:205–217.
24. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997;25:4876–4882.
25. Page RD. TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 1996;12:357, 358.
26. Andreeva A, Howorth D, Brenner SE, Hubbard TJ, Chothia C, Murzin AG. SCOP database in 2004: refinements integrate structure and sequence family data. Nucleic Acids Res 2004;32: D226–D229.
27. Tang CL, Xie L, Koh IY, Posy S, Alexov E, Honig B. On the role of structural information in remote homology detection and sequence alignment: new methods using hybrid sequence profiles. J Mol Biol 2003;334:1043–1062.
28. Petrey D, Xiang Z, Tang CL, Xie L, Gimpelev M, Mitros T, Soto CS, Goldsmith-Fischman S, Kernytsky A, Schlessinger A, Koh IY, Alexov E, Honig B. Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling. Proteins 2003;53(Suppl 6):S430–S435.
29. Petrey D, Honig B. GRASP2: visualization, surface properties, and electrostatics of macromolecular structures and sequences. Methods Enzymol 2003;374:492–509.
30. Matthews BW. Solvent content of protein crystals. J Mol Biol 1968;33:491–497.
31. Martin JL. Thioredoxin—a fold for all reasons. Structure 1995; 3:245–250.
32. Qi Y, Grishin NV. Structural classification of thioredoxin-like fold proteins. Proteins 2005;58:376–388.
33. Holm L, Sander C. Dali: a network tool for protein structure comparison. Trends Biochem Sci 1995;20:478–480.
34. Kong KH, Nishida M, Inoue H, Takahashi K. Tyrosine-7 is an essential residue for the catalytic activity of human class PI glutathione *S*-transferase: chemical modification and site-directed mutagenesis studies. Biochem Biophys Res Commun 1992;182: 1122–1129.
35. Nathaniel C, Wallace LA, Burke J, Dirr HW. The role of an evolutionarily conserved *cis*-proline in the thioredoxin-like domain of human class α glutathione transferase A1-1. Biochem J 2003; 372:241–246.
36. Fritz-Wolf K, Becker A, Rahlfs S, Harwaldt P, Schirmer RH, Kabsch W, Becker K. X-ray structure of glutathione *S*-transferase from the malarial parasite *Plasmodium falciparum*. Proc Natl Acad Sci USA 2003;100:13821–13826.
37. Abdalla AM, Bruns CM, Tainer JA, Mannervik B, Stenberg G. Design of a monomeric human glutathione transferase GSTP1, a structurally stable but catalytically inactive protein. Protein Eng 2002;15:827–834.
38. Sayed Y, Wallace LA, Dirr HW. The hydrophobic lock-and-key intersubunit motif of glutathione transferase A1-1: implications for catalysis, ligandin function and stability. FEBS Lett 2000; 465:169–172.
39. Alves CS, Kuhnert DC, Sayed Y, Dirr HW. The intersubunit lock-and-key motif in human glutathione transferase A1-1: role of the key residues Met51 and Phe52 in function and dimer stability. Biochem J 2006;393:523–528.
40. Habig WH, Jakoby WB. Assays for differentiation of glutathione *S*-transferases. Methods Enzymol 1981;77:398–405.
41. Bjornestedt R, Stenberg G, Widersten M, Board PG, Sinning I, Jones TA, Mannervik B. Functional significance of arginine 15 in the active site of human class α glutathione transferase A1-1. J Mol Biol 1995;247:765–773.
42. Hiller N, Fritz-Wolf K, Deponte M, Wende W, Zimmermann H, Becker K. *Plasmodium falciparum* glutathione *S*-transferase—structural and mechanistic studies on ligand binding and enzyme inhibition. Protein Sci 2006;15:281–289.
43. Micaloni C, Kong GK, Mazzetti AP, Nuccetelli M, Antonini G, Stella L, McKinstry WJ, Polekhina G, Rossjohn J, Federici G, Ricci G, Parker MW, Lo Bello M. Engineering a new C-terminal tail in the H-site of human glutathione transferase P1-1: structural and functional consequences. J Mol Biol 2003;325:111–122.
44. Dirr HW, Wallace LA. Role of the C-terminal helix 9 in the stability and ligandin function of class α glutathione transferase A1-1. Biochemistry 1999;38:15631–15640.
45. Bruns CM, Hubatsch I, Ridderstrom M, Mannervik B, Tainer JA. Human glutathione transferase A4-4 crystal structures and mutagenesis reveal the basis of high catalytic efficiency with toxic lipid peroxidation products. J Mol Biol 1999;288:427–439.