# A Glutathione Transferase from *Agrobacterium tumefaciens* Reveals a Novel Class of Bacterial GST Superfamily

## Katholiki Skopelitou<sup>1</sup>, Prathusha Dhavala<sup>2</sup>, Anastassios C. Papageorgiou<sup>2</sup>, Nikolaos E. Labrou<sup>1</sup>\*

1 Laboratory of Enzyme Technology, Department of Agricultural Biotechnology, Agricultural University of Athens, Athens, Greece, 2 Turku Centre for Biotechnology, University of Turku, Åbo Akademi University, BioCity, Turku, Finland

### Abstract

In the present work, we report a novel class of glutathione transferases (GSTs) originated from the pathogenic soil bacterium *Agrobacterium tumefaciens* C58, with structural and catalytic properties not observed previously in prokaryotic and eukaryotic GST isoenzymes. A GST-like sequence from *A. tumefaciens* C58 (*Atu*3701) with low similarity to other characterized GST family of enzymes was identified. Phylogenetic analysis showed that it belongs to a distinct GST class not previously described and restricted only in soil bacteria, called the Eta class (H). This enzyme (designated as *Atu*GSTH1-1) was cloned and expressed in *E. coli* and its structural and catalytic properties were investigated. Functional analysis showed that *Atu*GSTH1-1 exhibits significant transferase activity against the common substrates aryl halides, as well as very high peroxidase activity towards organic hydroperoxides. The crystal structure of *Atu*GSTH1-1 was determined at 1.4 Å resolution in complex with *S*-(p-nitrobenzyl)-glutathione (Nb-GSH). Although *Atu*GSTH1-1 adopts the canonical GST fold, sequence and structural characteristics distinct from previously characterized GSTs were identified. The absence of the classic catalytic essential residues (Tyr, Ser, Cys) distinguishes *Atu*GSTH1-1 from all other cytosolic GSTs of known structure and function. Site-directed mutagenesis showed that instead of the classic catalytic residues, an Arg residue (Arg34), an electron-sharing network, and a bridge of a network of water molecules may form the basis of the catalytic mechanism. Comparative sequence analysis, structural information, and site-directed mutagenesis in combination with kinetic analysis showed that Phe22, Ser25, and Arg187 are additional important residues for the enzyme's catalytic efficiency and specificity.

Citation: Skopelitou K, Dhavala P, Papageorgiou AC, Labrou NE (2012) A Glutathione Transferase from Agrobacterium tumefaciens Reveals a Novel Class of Bacterial GST Superfamily. PLoS ONE 7(4): e34263. doi:10.1371/journal.pone.0034263

Editor: Sotirios Koutsopoulos, Massachusetts Institute of Technology, United States of America

Received January 4, 2012; Accepted February 24, 2012; Published April 4, 2012

**Copyright:** © 2012 Skopelitou et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was partially supported by the grant THALES. The grant THALES falls under the Operational Programme "Education and Lifelong Learning" and is co-financed by the European Social Fund and National Resources. Dr. Papageorgiou thanks the Academy of Finland for financial support (grant number 121278). Access to EMBL Hamburg (c/o DESY) was provided by the European Community's Seventh Framework Programme (FP7/2007-2012) under grant agreement no. 226716. No additional external funding received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

\* E-mail: lambrou@aua.gr

### Introduction

Glutathione transferases (GSTs; EC 2.5.1.18) are phase II detoxification enzymes that metabolize a wide range of hydrophobic toxic compounds by catalyzing the conjugation of glutathione (GSH) to the hydrophilic centre of the toxic substances [1-4]. GSTs are known as promiscuous enzymes capable of catalyzing the conjugation of GSH with a broad range of electrophilic substrates [5-7]. Several members of the GST family are selectively induced by biotic and abiotic stress treatments and play important roles in the regulation of redox homeostasis as well as in endogenous metabolism [3,4]. GSTs can also bind hydrophobic compounds that are not their substrates [6]. This non-substrate binding (termed 'ligandin' function) is possibly associated with the sequestration, storage, and transportation of drugs, hormones, and other metabolites [6]. GSTs, therefore, are able to participate in various unrelated biological processes and may be considered as 'moonlighting' proteins [5].

GSTs form a highly diverse protein family and, therefore, have been subdivided into a number of subfamilies associated with

different functionalities and enzymatic properties [8–12]. GSTs are divided into at least four major families of proteins, namely cytosolic GSTs, mitochondrial GSTs, microsomal GSTs, and bacterial fosfomycin-resistance proteins [5,7,8]. GSTs that are grouped into different classes usually have different general substrate profiles, while members of the same class have fewer differences in substrate recognition [2,7]. All cytosolic GSTs have the same protein folding, which comprises two domains. The N-terminal domain (domain I) adopts  $\alpha/\beta$  topology and provides the GSH-binding site (G-site). The C-terminal domain (domain II) is an all- $\alpha$ -helical structure and provides the structural element for recognition of a broad range of hydrophobic co-substrate (H-site). The H-site lies adjacent to the G-site and defines the substrate specificity of the enzyme [7–12].

Like eukaryotic organisms, bacteria are characterized by multiple GST genes of widely divergent sequences and unknown biological function [8]. In bacteria, four different classes of GSTs have been identified: beta, chi, theta and zeta. Most of the bacterial GSTs identified to date belong to the bacterial-specific beta class and the crystal structures of several representatives of this class have been determined and characterized, such as *Proteus mirabilis* GST (*Pm*GST) [13] and *Ochrobactrum anthropi* GST (*Oa*GST) [14].

Agrobacterium tumefaciens is a ubiquitous soil borne pathogen 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 [15]. We have recently reported the identification and functional analysis of the GST family of enzymes from *A. tumefaciens* C58 [16]. In the present study, we report the kinetic characterization and crystal structure determination of *Atu*3701 protein from *A. tumefaciens*. Sequence and structural analysis indicate that *Atu*3701 defines a novel GST class distinct from other previously characterized GSTs.

#### **Results and Discussion**

# Identification and bioinformatics analysis of a new class of GSTs

In silico homology searches of Agrobacterium tumefaciens strain C58 genomic sequence revealed the presence of several sequences corresponding to putative GST homologues [16]. A putative sequence with NCBI accession number AAK89703 (ORF name Atu3701, AtuGST [16]) which shares low sequence homology, and therefore significant evolutionary distance, to other prokaryotic and eukaryotic GST classes was identified and selected for further study. AtuGST4 contains an open reading frame of 693 bp, coding for a polypeptide of 230 amino acid residues with a predicted molecular mass of 26,140 Da (residues 1–230) and a theoretical pI of 6.33. The gene is located in a linear chromosome of A. tumefaciens, between 779,833–780,525 bp [17].

BLAST analysis revealed that AtuGST4 has the highest identity (~64–68%) with unclassified GSTs from proteobacteria species (e.g. *Stigmatella, Mesorhizobium, Sinorhizobium, Bradyrhizobium)*. Interestingly, several close homologs of AtuGST4 were found in a set of environmental sequences determined recently by the environmental (marine metagenome) sequencing project carried out by the Whole Genome Shotgun (WGS) sequencing project (www.ncbi. nlm.nih.gov/projects/WGS/WGSprojectlist.cgi). This sequence, therefore, is likely to belong to a larger family. The size of this family is expected to increase as the existing sequence databases expand.

GSTs that share greater than 40% sequence identity are generally included in the same class, and those that possess less than 20–30% sequence identity are assigned to separate classes [5,8,18]. As shown in Figure 1 and Table S1, AtuGST4 exhibits only 17.2 to 26.1% sequence identity with representatives of all the available different GST classes, which supports the grouping of this enzyme into a new class. The AtuGST4 shows the highest identity with the bacterial Chi (26.1%) and plant Phi (24.2%) class enzymes.

In order to examine the genetic relationship between this enzyme and GSTs from all known classes, a phylogenetic analysis was created (Figure 2). The results showed that the *Atu*GST4 sequence is clearly separate from all GST classes presented in the phylogenetic tree even from those representing bacterial-specific classes (e.g. beta, chi) [12,19]. *Atu*GST4 branch extends separately from the clades of GSTB and GSTX and appears to be more ancient than them. All the above evidences point to the conclusion that *Atu*GST4 belongs to a new GST class, distinct from previously characterized GSTs. According to the available GSTs nomenclature and classification system we propose that *Atu*GST4 belongs to the Eta class (H) and may be designated as *Atu*GSTH1-1.

#### Purification and kinetic analysis

In order to characterize the AtuGSTH1-1 protein, the fulllength sequence was cloned, expressed in *E. coli* BL21 (DE3) cells, and purified. The enzyme did not bind adequately to the classical affinity adsorbents (GSH-Sepharose or hexyl-GSH-Sepharose) that are widely used for the purification of recombinant as well as native GSTs. This indicates differences in the G-site topology of AtuGSTH1-1 compared to the majority of other GST classes that are efficiently purified using GSH-based affinity adsorbents. AtuGSTH1-1 was purified (>98% purity) in a single-step procedure by metal-chelate affinity chromatography on Ni-NTA affinity adsorbent.

Steady-state kinetic analysis using CDNB and GSH was carried out and the k<sub>cat</sub>, and K<sub>m</sub> parameters were determined (Table 1). The K<sub>m</sub> values for GSH and CDNB were determined as 0.29 mM and 1.5 mM, respectively. Initial screening has shown that AtuGSTH1-1 exhibits high GSH-dependent peroxidase activity (GPOX) towards organic hydroperoxides such as cumene hydroperoxide and tert-butyl hydroperoxide [16]. Organic hydroperoxides can be formed both nonenzymatically by reaction of free radicals with polyunsaturated fatty acids and enzymatically by lipoxygenase- or cyclooxygenase-catalyzed oxidation of linoleic acid and arachidonic acid. AtuGSTH1-1 exhibits very high peroxidase activity (specific activity with cumene hydroperoxide 23.6 U/mg). With cumene hydroperoxide and tert-butyl peroxide as electrophile substrates, AtuGSTH1-1 exhibits high catalytic efficiency  $(k_{cat}/K_m)$  (Table 1), suggesting that hydroperoxides may be the 'natural' substrates for AtuGSTH1-1.

AtuGSTH1-1 exhibited significant thioltransferase activity using the 2-hydroxyethyl disulfide (HED) as a substrate. The  $k_{cat}$  and  $K_m$  values for HED were determined as 2.4 min<sup>-1</sup> and 4.1 mM, respectively (Table 1). In cases of oxidative stress, when there is a lack of GSH, some protein thiols are S-thiolated making proteinthiol disulfides [20]. This modification affects the activity of the proteins or enzymes, suggesting that AtuGSTH1-1 may play an important regulatory role in stress defence mechanism [21].

As shown in Table 1, the  $K_m$  values for GSH are dependent on the electrophilic substrate used. For example, the  $K_m^{\rm GSH}$  varies between 0.3 to 1.7 mM. Probably, this is the result of the rapid equilibrium random sequential bi-bi mechanism with intrasubunit modulation between the GSH binding site and electrophile binding site that is operated by GSTs [9,10].

#### Structural characterization of AtuGSTH1-1

Quality of the structure. The crystal structure of AtuGSTH1-1 was determined to 1.4 Å resolution with  $R_{work}$  and  $R_{\rm free}$  of 17.2% and 18.6%, respectively (Table 2). The final refined structure contains 213 residues, 313 water molecules, 1 S-(pnitrobenzyl)-glutathione (Nb-GSH) molecule, and 1 phosphate ion. The first 13 and the last 4 residues are flexible and were not modeled in the structure. Lys14 and Trp141 lack side-chains owing to their high flexibility. Ten residues were modeled in alternative conformations. The structure exhibits good stereochemistry with root mean square deviation (r.m.s.d) in bond lengths and bond angles of 0.008 Å and  $1.15^{\circ}$ , respectively. The (phi, psi) plot shows 92.5% of the non-Gly and non-Pro residues in the most favored regions and no residues in disallowed regions. One residue (Glu85) is found in the generously allowed region, possibly as a result of its interaction with Nb-GSH. The coordinate error as deduced by the diffraction precision indicator is 0.06 Å.

**Description of the structure.** The structure of AtuGSTH1-1 exhibits the characteristic overall fold of GSTs that comprises an N-terminal thioredoxin-like domain and a C-terminal all  $\alpha$ -helical

		20		40		60		80		
AtuGST4	<b>MSNIETVPAS</b>	IEMKPNPTIT	VFERSPDGGR	GLARDMPVRW	ALEEVG - QPY	HVRRLSFEAM	KEAS HLAY	Q PF	GQ1	73
GSTA AAA16572	MS GKPVLH	YA	N	TRGRMESVRW	LLAAAGVEFE	EKELE	KKEDLOKL	KSDGSLLE	voov	55
GSTR CAR42030	Maaaak	I YYTPG		- SCSI SPHI	VIRETGIDES	IERIDI RTKK	TESGKDELAL	NessePKee		52
COTD_CAR42000	MAD	VV	LBC	CODC DOVIM	TAKAVOVE	NKKILMIPAC	ENIKDEELKI	N DO	UTI	62
COTE EDV/66071	Meee CIV		CTD	I SDCVDAVKI	TI KALNID Y	EVKEYNLOTG	ELMOEEELDK	N BO	нту	57
COTE AAA22460	MAD		CAV	MOMMUT DCAT	ALEEACED Y	EIVOINEATA	EHKODEHLVD	DE	COV	55
GSTF_AAA33409	MAP	L1	GAV	MOWNLINGAT	ALEEAGSD - T	EIVPINFAIA	EHKSPEHLVK	NPF	CLARK D COUCH	55
GSTK_EDM15501	MGP			APRVLE	LEYDVLSPYS	WEGFEVECRY	QHEWNIKEKE	RPALLA	GIMKDSGNQP	55
GSTL_BT051648	MATIGVKPV-	LPP	PLISTSQPPP	LFDGTTRLYV	SYSCPFA	QRTWITRNYK	GLQNNIHLVP	IDLQNRPAWY	KEKVYLENKV	79
GSTM_AAC17866	MSCESSMVLG	YW	• • • • • • • • D	IRGLAHAIRL	LLEFTDTSYE	EKRYTCGEAP	DYDRSQWLDV	KFKLDLDF	PNL	64
GSTO_EDL42044	MSGDLSRCLG	KGSCPPGPVP	EGVIRIYSMR	FCPYSHRARL	VLKAKGI	RHEVININLK	SKPDWYYTK -		HPFGQI	72
GSTP_AAP72967	MP PYTVV	YF	•••••P	VRGRCAALRM	LLADQGQSWK	EEVVTV	ETWQEG	SLKASCLY	GQL	53
GSTR_ABV24478	MA QNMM	LY	WGS	GSPPCWRVMI	ALEEKLLQGY	KHKLLSFDKK	EHQGAEVKAL	N PR	VQL	57
GSTS_EAA45010	MPDYKVYYFN			VKALGEPLRF	LLSYGNLPFD	DVRITREEWP	ALKPTMPM		GQM	51
GSTT_BAB39498	MG	LEL	YLDLLSQP	CRAVYI	FAKKNDIPFE	LR IVDLIK	GQ HL SD	AFAQVNPL	KKV	54
GSTU ABF99228	MAEEKELVLL	DFW		VSPFGQRCRI	AMAEKGLEFE	YRE EDL	GNKSDLLLRS	N	PVHRKI	56
GSTX EAW33767	M LK	LY	GGA	FS - RASIVOW	YLEEISIP-Y	EFVLLDMKAG	AHLEPDYLAI	N P	GKV	52
GSTZ P28342	MSSSETOKMQ	LY	SFS	LSSCAWRVRI	ALHLKGLD - F	EYKAVDLFKG	EHLTPEFLKL	N PL	GYV	60
100%										
Conservation		-		~ ~	A					
0%										
	100		120		140		160		180	
AtuGST4	PSVE-OGDL I	LEESGALVMH	LAOHH	GLIPEDOL RR	ARTVAWMEAA	INTIE	PSILNETTVW	LEERNERWHE	ARI ARTKEOL	153
GSTA AAA16572	PMVE - LDGMK	MYOTRALLNY	LAGKY	NIVGKDIKER	ALL	MYVEGLADLY	- EL LMMNVVO	PADKKEEHLA	NALD	126
CSTR CAR42030	PVI OL DNCDI	LTECVALVOY	ADI KRORNI	LADDKALEDY	HOL	MITTEGLADET	EWINELASE	VHKCYSPIES	SPIDESVIDV	124
COTD EDW/42479	PTLV DNCEA	LINESDALOWY	LADERPORTE	LYPKCPKKP	AVINORIVED	M OTLY	OSEANYYYDO	VEAKADADD	EAEV	124
G310_EDW42470	PILV-DIGFA	LWESKATOVI	LVERTGRIDS	-LIPKCPKKK	AVINGRETED	MGILI	DOLANYVADE	WTTCCTEVPO	EAFR	129
GSTE_EDV550/1	PVLE-DNGIV	LWDSHATAAT	LVDKTAKSDD	-LTPRULVKR	ATTNORLFFE	ASVIT	PGLANVVAPF	WIIGCIEVPQ	EKLD	134
GSTF_AAA33469	PALQ - DGDLY	LFESRAICKY	AAKK N	NPE-LLRE	GNLEEAAMVD	VVIEVEANQY	TAALNPILFQ	VLISPMLGGT	DUKVVDENL	136
GSTK_EDM15501	PAMVPHKGQY	ILKEIPLLKQ	LFQVPMSVPK	DFFGEHV-KK	GTVNAMRFLT	AVSMEQPEML	EKVSRELWMR	IWSRDEDITE	SQNILSAAEK	144
GSTL_BT051648	PSLE-HNGKV	LGESLDLIKY	IDANF - DGPP	L - FPND - PAK	KEFAEQLLSH	V DTFT	KELLVSL	KGDT	V Q	143
GSTM_AAC17866	PYLL - DGKNK	ITQSNAILRY	I ARKH	NMCGETEEEK	IRV D	IIENQVMDFR	TQLIRLCYSS	DHEKLKPQYL	EELP	136
GSTO_EDL42044	PVLENSQCQL	VYESVIACEY	LDDVYPGR	KLFPYDPYER	A	R	QKMLLELFCK	VPPLSKECLI	ALRCGRDCTD	142
GSTP_AAP72967	PKFQ-DGDLT	LYQSNTILRH	LGRTL	GLYGKDQQEA	ALV D	MVNDGVEDLR	CKYISLIYTN	-YEAGKDDYV	KALP	124
GSTR_ABV24478	PTFK-HGDLI	VNESFAACLY	LESAFKSQGT	RLIPDDPAEQ	ALVYQRM-FE	T NN LQ	QKMYDVAFYE	WYIPEGERHE	SALKRNKENL	140
GSTS EAA45010	PVLEVDG - KK	VHQSVAMSRY	LANQV	GLAGADDWEN	LMI D	TVVDTVNDFR	LKIAIVAYEP	DDMVKEKKMV	T L	121
GSTT BAB39498	PALK - DGDFT	LTESVAILLY	LTRKY-KVPD	YWYPODLOAR	ARVDEYLAWQ	HTTLR	RSCLRALWHK	VMFPVFLGEP	VSPQ	131
GSTU ABE99228	PVL - LHAGRP	VSESLVILOY	LDDAEPGTPH	LIPPANSGDA	DAAYARATAR	EWADYVDRKI	YDCGSRLW		- RI KGEPOAA	132
GSTX FAW33767	PALV-DGDVK	LWESGGILLY	LADKYGKMPD	SPEKR	GELYO	-WV-LEGNST	LA NGVELE	T	NREKETPRLM	118
GST7 P28342	PVI V - HGDIV	LADSLALIMY	LEEKEDEND.	- I I PODI OKP	ALNYOAANIV	Teressio	P-LONLAVIN	VIEEKIGSD	EKI SWAKHHI	140
100%		TADOLATIMI	LEGARFENT.	- LEF QDE QKK	ALITICAANTY	TTTTTTTTTT	FILGHEATEN	TILLEREGOD	ERESMARNIT	140
Conservation									100	
Conservation										
0%								~~~		
0%		200		220		240		260		
Atucesta		200 1	D PEWI	220 1				260 1		215
AtuGST4	LK-RLDE	200 - LSAWLG	-DREWL	EGSFSA	ADILMICV	240 I LRRLESSGI -		260     LLAYVE	RGKARPAFKR	215
AtuGST4 GSTA_AAA16572	LK-RLDE	200 I -LSAWLG VFEKVLKDH-	-DREWL -GHDFL	220 I EGS FSA VGNKL SRADV	ADILMICV	240 I LRRLESSGI- ILAVEES	KPDALAKFPL	260   LLAYVE LQSFKA	RGKARPAFKR	215 196
AtuGST4 GSTA_AAA16572 GSTB_CAR42930	LK-RLDE -K-AANRYFP VKNKLKSKFV	200 I - LSAWLG VFEKVLKDH- YINDVLS	-DREWL -GHDFL -KQKCV	220 I EGSFSA VGNKLSRADV CGDHFTVADA	ADILMICV HLLET YLF	240 I LRRLESSGI - ILAVEES TLSQWAPH	KPDALAKFPL VALDLTDLSH	260   LLAYVE LQSFKA LQDYLA	RGKARPAFKR RTSNIPNIKK RIAQRPNVHS	215 196 193
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478	LK-RLDE -K-AANRYFP VKNKLKSKFV KIESAFE	200   -LSAWLG VFEKVLKDH- YINDVLS FLNTFLE	-DREWL -GHDFL -KQKCV -GQEYA	220 I EGS FSA VGNKL SRADV CGDHFT VADA AGDSLT	ADILMICV HLLET YLF	240 I LRRLESSGI- 	KPDALAKFPL VALDLTDLSH -GFEISKYAN	200   	RGKAR PAFKR RTSNIPNIKK RIAQR PNVHS NAKK	215 196 193 189
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071	LK-RLDE -K-AANRYFP VKNKLKSKFV KIESAFE SIHRGLK	200 J VFEKVLKDH- YINDVLS FLNTFLE LLESFLH	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL	220 I EGSFSA VGNKLSRADV CGDHFTVADA AGDSLT AGDSLT	ADILMICV HLLET YLF	240 I LRRLESSGI- ILAVEES TLSQWAPH VASVSTFEVA GPTVSALRAA	VALDLTDLSH GFEISKYAN VDIEPVEFPK	200   LLAYVE LQSFKA LQDYLA VNKWYE VSAWLD	RGKARPAFKR RTSNIPNIKK RIAGRPNVHS NAKK RLNRLPFYKA	215 196 193 189 201
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AAA33469	LK-RLDE -K-AANRYFP VKNKLKSKFV KIESAFE SIHRGLK -E-KLKKVLE	200 I VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT	- D REWL - G HDFL - K QKCV - G QEYA - S SSYL - K CKYL	2200 I EGSFSA VGNKLSRADV CGDHFTVADA AGDSLT AGDSLT AGDFLS	ADILMICV HLLET YLF	240 I LRRLESSGI- ILAVEES TLSQWAPH VASVSTFEVA GPTVSALRAA TLCLFATPYA	KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH	200 I LLAYVE LQSFKA LQDYLA VNKWYE VSAWLD VKAWWS	RGKARPAFKR RTSNIPNIKK RIAGRPNVHS NA	215 196 193 189 201 205
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AA33469 GSTK_EDM15501	LK-RLDE -K-AANRYFP VKNKLKSKFV KIESAFE SIHRGLK -E-KLKKVLE AGMATAQAQH	200 I VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT LLNKISTELV	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL KSKLRETT	220 I EGSFSA VGNKLSRADV CGDHFTVADA AGDSLT AGDSLT AGDFLS -GAACKYGAF	AD I L MI C V H L L E T	240 IRRLESSGI- ILAVEES TLSQWAPH VASVSTFEVA GPTVSALRAA TLCLFATPYA TTVAHV	LKDYGN KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS	200 I LLAYVE LQSFKA LQDYLA VNKWYE VSAWVD VKAWWS DRMELLAYLL	RGKAR PAFKR RTSNIPNIKK RIAGR PNVHS NAKK RLNRL PFYKA GLMER PSVQK GE KWMGP VPP	215 196 193 189 201 205 220
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AAA33469 GSTK_EDM15501 GSTL_BT051648	LK-RLDE -K-AANRYFP VKNKLKSKFV KIESAFE SIHRGLK -E-KLKKVLE AGMATAQAQH QSSPTFE	200 I VFEKVLKDH- VINDVLS FLNTFLE LLESFLH VYEARLT LLNKISTELV FLENALGKFD	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL KSKLRETT DGPFL	220 I EGSFSA VGNKLSRADV CGDHFTVADA AGDSLT AGDSLT AGDFLS -GAACKYGAF LG.QLS	ADILMICV HLLET YLF	240 I LRRLESSGI- ILAVEES -TLSQWAPH VASVSTFEVA GPTVSALRAA TLCLFATPYA TVAHV VERFHIVLA-	LKDYGN KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE	200 I LLAYVE LQSFKA LQDYLA VNKWYE VSAWLD VKAWS DRMELLAYLL GRPKLATWIE	RGKARPAFKR RTSNIPNIKK RIAGRPNVHS NAKK RLNRLPFYKA GLMERPSVGK GEKWMGPVPP ELNKIDA	215 196 193 189 201 205 220 214
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AAA33469 GSTK_EDM15501 GSTL_BT051648 GSTM_AAC17866	LK-RLDE -K-AANRYFP VKNKLKSKFV SIHRGLK -E-KLKKVLE AGMATAQAQH QSPTFE -G-QLKQ	200 I VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT LLNKISTELV FLENALGKFD -FSMFL	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL KSKLRETT- DGPFL GKFSWF	220 I EGSFSA VGNKLSRADV CGDHFTVADA AGDSLT AGDSLT AGDFLS -GAACKYGAF LG-QLS AGEKLTFVDF	ADILMICV HLLET	240 I LRRLESSGI- ILAVEES TLSQWAPH VASVSTFEVA GPTVSALRAA TLCLFATPYA TTVAHV VERFHIVLA- LDQNRIF	KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLD - AYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN	200 I LLAYVE LQSFKA LQDYLA VNKWYE VSAWUD VKAWWS DRMELLAYLL GRPKLATWIE LKAFMC	RGKARPAFKR RTSNIPNIKK RIAQRPNVHS NAKK RLNRLPFYKA GEKWMGPVPP ELNKIDA RFEALEKIAA	215 196 193 189 201 205 220 214 200
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDV42478 GSTE_EDV55071 GSTF_AAA33469 GSTK_ED015501 GSTL_BT051648 GSTM_AAC17866 GSTO_EDL42044	LK-RLDE -K-AANRYFP VKNKLKSKFV 	LSAWLG. VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT LLNKISTELV FLENALGKFD -FSMFL MMEEILE	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL -KSKLRETT DGPFL GKFSWF YQNTFF	2200 I EGS FSA VGNKL SRADV CGDHFTVADA AGDSLT AGDSLT AGDSLT AGDSLT AGDSLT AGDSLT AGDSLT AGDSLT AGDSLT AGDSLT AGDSLT AGDSLT AGDSLT AGDSLT AGDSLT AGDSLT AGDSLT AGDSLT AGDSLT	ADILMICV HLLET	240 LRRLESSGI- 	KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN LADCVN	200 LLAYVE LQSFKA LQDYLA VKAWYE VSAWUD VKAWWE USAWUG UKAWUG GRPKLAYLL GRPKLAYLL GRPKLATWIE LKAFMC HTPMLRLWIA	RGKARPAFKR RTSNIPNIKK RIAGRPNYHS NAKK RLNRLPFYKA GLMERPSVQK GEKWMGPVPP ELNKIDA RFEALEKIAA SMKQDPAVCA	215 196 193 189 201 205 220 214 200 215
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AA33469 GSTK_EDM15501 GSTL_BT051648 GSTM_AAC17866 GSTO_EDL42044 GSTP_AAP72967	LK-RLDE -K-AANRYFP VKNKLKSKFV KIESAFE SIHRGLK -E-KLKKVLE AGMATAQAQH QSSPTFE -G-QLKQ LKVALRQELC	200 I VFEKVLKDH- YINDVLS FLNTFLE LLSFLH VYEARLT VYEARLT FELNALGKFD -FSMFL FETLLSQNQ	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL KSKLRETT. DGPFL GKFSWF YQNTTF GGKTF1	220 I EGSFSA VGNKLSRADV CGDHFTVADA AGDSLT AGDSLT GAACKYGAF LG-QLS AGEKLTFVDF GGDCIS VGDQISFADY	- ADILMICV HLLET VADIAL - LADLLS - LADLNHVSV GLP LVDIAYIPF LTYDI - MIDYLVWPW NLLDL	240 LRRLESSGI 	KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN LADCVN APGCLDAFPL	200 LLA YVE LQS FKA LQD YLA VNK WVE VSA WVE VKA WVE VKA FMC GRPKLATWIE LKA FMC HTPMLRLWIA HSA YVG	RGKARPAFKR RTSNIPNIKK NAKK RLNRLPFYKA GLMERPSVQK GEKWMGPVPP ELNKIDA RFEALEKIAA SMKQDPAVCA RLSARPKLKA	215 196 193 189 201 205 220 214 200 215 192
AluGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTE_AAA33469 GSTK_EDM15501 GSTL_BT051648 GSTM_AAC17866 GSTO_EDL42044 GSTP_AAP72967 GSTR_ABV24478	LK-RLDE -K-AANRYFP VKNKLKSKFV SIHRGLK -E-KLKKVLE AGMATAQAQH QSPTFE -G-QLKQ LKVALRQELC -G-QLKQ VT-ELKLWDG	200 I VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT LNKISTELV FLENALGKFD -FSMFL NMEEILE -FETLLSQNQ YLEKMGK	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL KSKLRETT- DGPFL GKFSWF YQNTTFF GGKTFI -GS.YL	2200 I EGS FSA VGNKL SRADV GDHFTVADA AGDSLT AGDFLS GACK YGAF LG-QLS VGDQ ISFADY VGDQ ISFADY AGKNFT	ADILMICV HLLET	240 I LRRLESSGI- ILAVEES ILAVEES ILSQWAPH VASVSTFEVA GPTVSALRAA TLCLFATPYA TVAHV VERFHIVLA- LDQNRIF FERLDVYG- LLIHEVL FP-VVAYFPR	KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLD-AYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN LADCVN APGCLDAFPL LHCPKEKCPR	200 I LLAYVE LQSFKA LQDYLA VNKWLP VSAWLD VKAWWS DRMELLAYLL GRPKLATWIE LKAFMC HTPMLRLWIA LSAYVG LMEYVE	RGKARPAFKR RTSNIPNIKK RIAQRPNVHS NAKK GEKERPSVQK GEKWMGPVPP ELNKIDA RFEALEKIAA SMKQDPAVCA RLSARPKLKA MVKDRPSIKA	215 196 193 189 201 205 220 214 200 215 192 207
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDV42478 GSTE_EDV55071 GSTF_AAA33469 GSTK_EDM15501 GSTL_BT051648 GSTM_AAC17866 GST0_EDL42044 GSTP_AAP72967 GSTR_ABV24478 GSTS_EA445010	LK-RLDE -K-AANRYFP VKNKLKSKFV KIESAFE SIHRGLK -E-KLKKVLE -G-QLKQ UKVALRQELC -G-QLKP VT-ELKLWDG NNEVIPFYLT	LESAWLG. VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT LLNKISTELV FLENALGKFD -FSMFL -FETLLSONG YLEKMGK KLNVIAKENN	-D REWL -G HDFL -K QKCV -G QEYA -S SSYL -K CKYL -K CKYL GK FSWF GG KTF1 -G S -YL GHLVLGKPTW	2200 I EGS FSA VGNKL SRADV CGDHFTVADA AGDSLT AGSLT AGSLT AGSLT AGSLT AGSLT	ADILMICV HLLET VADIAL -LADLLS -LADLNHVSV GLP -LVDIAYIPF LTYDI -MIDYLVWPW NLLDL -MADVVC GILDYLNYLT	240 LRRLESSGI- 	KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN APGCLDAFPL LHCPKEKCPR LENFPN	200 LLAYVE LQSFKA LQDYLA VNKWYE VSAWYE VKAWYE USAWUD VKAWYE LKAFMC LKAFMC LKAYVG LMEYYE LQEVVQ	RGKARPAFKR RTSNIPNIKK RIAGRPNYHS NAKK RLNRLPFYKA GLMERPSVGK GEKWMGPVPP ELNKIDA RFEALEKIAA SMKQDPAVCA RLSARPKLKA MVKDRPSIKA KVLDNENVKA	215 196 193 189 201 205 220 214 200 215 192 207 193
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AA33469 GSTK_EDM15501 GSTL_BT051648 GSTM_AAC17866 GSTO_EDL42044 GSTP_AAP72967 GSTR_ABV24478 GSTS_EAA45010 GSTT_BAB39498	LK-RLDE -K-AANRYFP VKNKLKSKFV KIESAFE SIHRGLK -E-KLKKVLE AGMATAQAQH QSSPTFE -G-QLKQ LKVALRQELC -G-QLKP VT-ELKLWDG NNEVIPFYLT TLAATLA	200 I VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT VYEARLT FENFL FETLSQNQ YLEKMGK KLNVIAKENN ELDVTLALE	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL KSKLRETT DGPFL GKFSWF YQNTTF -GGKTF1 -GS.YL GHLVLGKPTW DKFLQNKAFL	220 I EGSFSA VGNKLSRADV CGDHFTVADA AGDSLT AGDSLT AGDFLS GAACKYGAF GGDCIS AGEKLTFVDF GGDCIS ADVYFA ADVYFA	- ADILMICV HLLET VADIAL -VADIAL -LADLNHVSV GLP -UVDIAYIPF LTYDI -MIDYLVWPW NLLDL -MADVVC GILDYLNYLT -LADLVAI	240 LRRLESSGI 	KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN LADCW APGCLDAFPL LHCPKEKCPR LENFPN CQVFE	200 LLA YVE LQS FKA LQD YLA VNK WVE VSA WVD VKA WVD VKA WVD DRMELLAYLL GRPKLATWIE LSA YVG LME YVG LME YVG LQE V VVQ GRPKLATWRO	RGKARPAFKR RTSNIPNIKK RIAQRPNVHS NAKK RLNRLPFYKA GLMERPSVQK GEKWMGPVPP ELNKIDA RFEALEKIAA SMKQDPAVCA RLSARPKLKA MVKDRPSIKA KVLDNENVKA RVEAAVGEDL	215 196 193 201 205 220 214 200 215 192 207 193 206
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AAA33469 GSTK_EDM15501 GSTL_BT051648 GST0_EDL42044 GSTD_AAP72967 GSTR_ABV24478 GSTS_EAA45010 GSTT_BAB39498 GSTU_ABF99228	LK-RLDE -K-AANRYFP VKNKLKSKFV 	LSAWLG- VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT LLNKISTELV FLENALGKFD -FSMFL -FETLLSQNQ VEKMGK KLNVIAKENN ELDVTLQLLE LEAELGD	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL KSKLRETT- DGFSWF -YQNTTFF GGKTFI -GS.YL GHLVLGKPTW DKFLONKAFL 	2200 I EGS FSA VGNKL SRADV CGDHFTVADA AGDSLT AGDSLT AGDFLS AGDFLS AGEKLTFVDF GGDCIS VGDQISFADY AGKNFT ADVYFA TGPHIS	ADILMICV HLLET	240 LRRLESSGI- 	LKDYGN KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN LADCVN APGCLDAFPL LHCPKEKCPR LENFPN CQVFE GGFSVEE	200 LLAYVE LQSFKA LQDYLA VKAWVE VKAWVE VKAWUS DRMELLAYLL GRPKLATWIE LKAFMC HTPMLRLWIA LSAYVE LQEVYVE GRPKLATWAR VAPRLAAWAR	RGKARPAFKR RTSNIPNIKK RIAGRPNVHS NAKK RLNRLPFYKA GLMERPSVGK GEKWMGPVPP ELNKIDA RFEALEKIAA SMKQDPAVCA RLSARPKLKA RVKDRPSIKA KVLDNENVKA RVEAAVGEDL RCGRIDSVVK	215 196 193 201 205 220 214 200 215 192 207 193 206 209
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AAA33469 GSTK_EDM15501 GSTL_ED0142044 GSTD_AE172967 GSTP_AAP72967 GSTR_ABU24478 GSTS_EAA45010 GSTT_BAB39498 GSTU_AB799228 GSTX_EAW33767	LK-RLDE -K-AANRYFP VKNKLKSKFV 	200 I VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT VYEARLT FENFL FETLLSQNQ YLEKMGK KLNVIAKENN ELDVTLQLLE TLEAELGD	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL KSKLRETT DGPFL GKFSWF GGKTFI -GS.YL GHLVLGKPTW DKFLQNKAFL -REFF -KQPY1	220 I EGS FSA VGNKL SRADV CGDHF TVADA AGDSLT	AD I L MI CV H L L E T Y V AD I AL - L AD L LS - L AD L NH VS V GLP - L VD I AY I PF L T VD I A - MI DY L VWPW N LLD L - MAD VV G G I L DY LNY LT - LAD L VAI GFV DV AL VPF - VAD I AI	240 LRRLESSGI- 	KPDALAKFPL VALDLTDLSH GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN APGCLDAFPL LHCPKEKCPR GGFSVEEE GGFSVEEE GGFSVEEE	200 LLAYVE LQSFKA LQDYLA VNKWYE VSAWUD VKAWVE VKAWVS DRMELLAYLL GRPKLATWIE LKAFMC LMEYVE LMEYVE GRPKLATWRQ VAPRLAAWAR VIDYTK	RGKARPAFKR RTSNIPNIKK RIACRPNVHS NAKK RLNRLPFYKA GLMERPSVQK GEKWMGPVPP ELNKIDA RFEALEKIAA MVKDRPSIKA RVSARPKLKA RVEAAVGEDL RCGRIDSVVKA RVEAAVGEDL	215 196 193 201 205 220 214 200 215 192 207 193 206 209 179
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_FAA33469 GSTK_EDM15501 GSTL_BT051648 GSTM_AAC17866 GSTO_EDL42044 GSTP_AAP72967 GSTR_ABV24478 GSTS_EAA45010 GSTT_BAB39498 GSTU_ABF99228 GSTU_ABF99228 GSTV_EAW33767	LK-RLDE -K-AANRYFP VKNKLKSKFV KIESAFE SIHRGLK -E-KLKKVLE AGMATAQAQH QSSPTFE -G-QLKQ LKVALRQELC -G-QLKP VT-ELKLWDG NNEVIPFYLT TLAATLA AGREMAEILR -T-ALNEIFQ 	200 I VFEKVLKDH- YINDVLS FLNTFLE ULESFLH VYEARLT VYEARLT FETLSQNQ YLEKMGK FETLSQNQ YLEKMGK KLNVIAKENN KLNVIAKENN ALEKLLK	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL KSKLRETT DGFSWF YQNTTF GGKTF1 -GS.YL GHLVLGKPTW DKFLONKAFL REFF -KQPY1 	220 I EGS FSA VGNKL SRADV CGDHFTVADA AGDSLT AGDSLT GAACKYGAF LG-QLS AGEKLTFVDF GGDCIS ADVFA GGG.SGGRL MGDEFN GGG-GGGRL	AD I LMI CV HLLET VADIAL - LADLLS - LADLNHVSV GLP - LVDIAYIPF LTYDI - MIDYLVWPW NLLD - MADVVC GILDYLNYLT - GFVDVALVPF - VADIAI GLAD LFL	240 LRRLESSGI- ILAVEES LSQWAPH VASVSTFEVA GPTVSALRAA TLCLFATPYA VERFHIVLA- LQNRIF FERLDVYG- LIHEVL FP-VVAYFPR KTNL- TELMHPVGAG TAWFYSYERC GSILAYIPMM APQIIASITG	KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN LADCW APGCLDAFPL LHCPKEKCPR LENFPN GGFSVEE LKLDLSAYPA FGMDMAEFPL	200 LLAYUE LQSFKA LQDYLA VNKWIE VKAWUE VKAWUE VKAWUE VKAWUE UKAFMC HTPMLRLWIA LSAYUE LQEVYUE LQEVYUE VAPRLAAWAR VIDYTK KSLNDAYLK	RGKARPAFKR RTSNIPNIKK RIACRPNVHS NAKK RLNRLPFYKA GLMERPSVQK GEKWMGPVPP ELNKIDA RFEALEKIAA MKODPAVCA RLSARPKLKA MVKDRPSIKA KVLDNENVKA RVEAVGEDL RCGRIDSVVK RVCERPAFQK	215 196 193 201 205 220 214 200 215 192 207 193 206 209 179 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AAA33469 GSTL_BT051648 GSTM_AAC17866 GST0_ED142044 GSTP_AAP72967 GSTR_ABV24478 GSTS_EAA45010 GSTT_BAB39498 GSTU_AB799228 GSTZ_P28342 GSTZ_P28342 0572_P28342	LK-RLDE -K-AANRYFP VKNKLKSKFV 	LESAWLG. VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT LLKKISTELV FLENALGKFD -FSMFL MMEEILE -FETLLSONQ YLEKMGK KLNVIAKENN ELDVTLQLLE TLEAELGD ALEKLLK	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL GKFSWF -YQNTFF GGKTF1 -GS.YL GHLVLGKPTW DKFLONKAFL -REFF -KQPY1 -GH.	2200 I EGS FSA VGNKL SRADV CGDHFTVADA AGDSLT AGGSLTS GGDCIS GGC AGSLT ADVYFA AGGS AGSLT AGSLT AGSLT AGSLT AGSLT AGSLT AGSLT AGSLT AGSLT AGSLT AGSLT AGSLT	ADILMICV HLLET	240 LRRLESSGI- 	KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN APGCLDAFPL LHCPKEKCPR ENFPN CQVFE GGFSVEE GGFSVE	200 LLAYVE LQSFKA LQDYLA VKAWVE VSAWUD VKAWVE USAWVE LKAFMC HTPMLRWIA LSAYVG RPKLATWRQ VAPRLAAWAR VIDYTK LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RIAGRPNVHS NAKK RLNRLPFYKA GLMERPSVGK GEKWMGPVPP ELNKIDA RFEALEKIAA SMKQDPAVCA RVEALAVGEDL KVLDNENVKA RVEAAVGEDL RCGRIDSVVK RVCERPAFQK YQHFRMRCQR	215 196 193 201 205 220 214 200 215 192 207 193 206 209 179 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AAA33469 GSTK_EDM15501 GSTL_ED51648 GST0_EDL42044 GSTD_AAP72967 GSTR_ABV24478 GSTS_EAA45010 GSTT_BAB39498 GSTV_ABM3767 GSTZ_P28342 1000 COnservation	LK-RLDE -K-AANRYFP VKNKLKSKFV 	200 I VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT VYEARLT VTEARLT FSMFL FETLLSQNQ YLEKMGK KLNVIAKENN ELDVTLQLLE TLEAELGD ALEKLLK	-DREWL -GHDFL -KQKCV -GQKCV -GQKYA -SSSYL -KCKYL KSKLRETT DGPFL GKFSWF -YQNTTFF -GGKTFI -GS.YL GHLVLGKPTW DKFLQNKAFL REFF -KQPYI -GH	220 I EGS FSA VGNKL SRADV CGDHF TVADA AGDSL T AGDSL T AGDSL T AGDK YGAF GGDC IS VGDQ ISFADY AGKNF T TGPH IS GGG - GGGRL MGDEFN AGKY ATGDEV	AD ILMICV HLLET VADIAL -VADLLS -LADLNHVSV GLP -UVDIAYIPF LTYDI -MIDYLVWPW NLLDL GILDYLVYLT -LADLVAI GFVDVALVPF -VADIAI GLADLFL	240 LRRLESSGI- 	KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN APGCLDAFPL LHCPKEKCPR GGFSVEE	200 LLAYVE LQSFKA LQDYLA VNKWYE VSAWUD VKAWVE DRMELLAYLL GRPKLATWIE LKAFMC LGEVVQ GRPKLATWRQ VAPRLAAWAR VIDYTK LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RIACPPVHS NAKK RLNRLPFYKA GLMERPSVQK GEKWMGPVPP ELNKIDA RFEALEKIAA SMKQDPAVCA RLSARPKLKA MVKDRPSIKA KVLDNENVKA RVEAVGEDL RCGRIDSVVK YQHFRMRCQR	215 196 193 201 205 220 214 200 215 192 207 193 206 209 179 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTK_EDU55071 GSTK_EDU55071 GSTK_ED145001 GSTL_BT051648 GST0_EDL42044 GSTD_AAP72967 GSTR_ABV24478 GSTS_EAA45010 GSTT_BAB39498 GSTL_EAW33767 GSTZ_P28342 GSTZ_P28342 SGSTZ_EAW33767 GSTZ_P28342 SGSTZ_P283443 SGSTZ_P28342 SGSTZ_P283443 SGSTZ_P283443443 SGSTZ_P283443443 SGSTZ_P283443444444444444444444444444444444444	LK-RLDE -K-AANRYFP VKNKLKSKFV VKNKLKSKFV 	LSAWLG VFEKVLKDH. YINDVLS FLNTFLE LLESFLH VYEARLT LLKISTELV FLENALGKFD .FSMFL MEEILE FETLLSQNQ VLEKMGK KLNVIAKENN ELDVTLQLLE LEAELGD ALEKLLK	-DREWL -GQKCV -GQEYA -SSSYL -KCKYL KSKLRETT DGFSWF GGFSWF -YQNTTFF GGKTFI -GS.YL GHLVLGKPTW DKFLONKAFL -KQPYI -GH.	2200 I EGS FSA VGNKL SRADV CGDHFTVADA AGDSLT AGDSLT AGDSLT GACKYGAF LG-QLS AGEKLTFVDF GGDCIS VGDQISFADY AGKNFT AGVYFA TGPHIS AGG-GGGRL MGDEFN AGKYATGDEV	AD ILMICV HLLET	240 LRRLESSGI- ILAVEES ILAVEES ILSQWAPH VASVSTFVA GPTVSALRAA GPTVSALRAA GPTVSALRAA GPTVSALRAA FEXLDVIG- LIHEVL FEXLDVIG- LIHEVL FP-VVAYFPR KTNL GSILAYIPMM APQIIASITG	LKDYGN KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN LADCVN APGCLDAFPL LHCPKEKCPR LENFPN CQVFE GGFSVEE LKLDLSAYPA FGMDMAEFPL	200 LLAYVE LQSFKA LQDYLE VKAWYE VSAWLD VKAWLD GRPKLATWLE LKAFMC HTPMLRLWIA LSAYVE GRPKLATWIE LKEYVE GRPKLATWA VAPRLAAWAR VIDYTK LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RIAGRPNVHS NAKK RLNRLPFYKA GLMERPSVGK GEKWMGPVPP ELNKIDA RFEALEKIAA SMKQDPAVCA RLSARPKLKA RVEAPVED KVLDNENVKA RVEAAVGEDL RCGRIDSVVK RVCERPAFQK YQHFRMRCQR	215 196 193 201 205 220 214 200 215 192 207 193 206 209 179 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_EAA33469 GSTK_EDW15501 GSTL_BT051648 GSTM_AAC17866 GSTD_EAU42044 GSTP_AAP72967 GSTF_AA845010 GSTS_EA445010 GSTS_EA445010 GSTS_EA445010 GSTS_EA445010 GSTS_P28342 GSTZ_P28342 000000000000000000000000000000000000	LK-RLDE -K-AANRYFP VKNKLKSKFV 	LESAWLG VFEKVLKDH- YINDVLS FLNTFLE ULESFLH VYEARLT VYEARLT MMEEILE -FETLLSQNQ YLEKMGK KLNVIAKENN ELDVTLQLLE TLEAELGD ALEKLLK	-DREWL -GQKCV -GQKCV -GQEYA -SSSYL -KCKYL -KSKLRETT- DGPFL GKFSWF GGKTFI -GS.YL OHLVLGKPTW DKFLQNKAFL -KQPY1 -GH.	220 I EGS FSA VGNKL SRADV CGDHFTVADA AGDSLT AGSLT AGDSLT AGSLT	AD I L MI CV H L L E T VAD I AL -LAD L LS -LAD L NH VS V GLP -LVD I AY I PF LTYD I -MI DY LVWPW N LLD L -MAD VVC GI LD YLNYLT -LAD LVAI GFVD VAL VPF -VAD I AI GLAD LFL	240 LRRLESSGI- ILAVEES ILAVEES TLSQWAPH VASVSTFEVA GPTVSALRAA TLCLFATPYA VERFHIVLA. LUAVYG- LLIHEVL FP-VVAYFPR TELMHPVGAG TAWFYSYEG GSILAYIPMM APQIIASITG	KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN APGCLDAFPL LHCPKEKCPR CGVFE GGFSVE KKLDLSAYPA FGMDMAEFPL	200 LLAYVE LQSFKA LQDYLA VNKWYE VSAWLD VKAWYE USAWLG VKAWYE CALAYLL GRPKLATWIE LKAFMC LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RIACRPNYHS NAKK RLNRLPFYKA GLMERPSVQK GEKWMGPVPP ELNKIDA RFEALEKIAA MVKDRPSIKA KVLDRPSIKA RVEAAVGEDL RCGRIDSVVK RVCERPAFQK YQHFRMRCQR	215 196 193 201 205 220 214 200 215 192 207 193 206 209 179 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AA33469 GSTK_EDM15501 GSTL_ED051648 GST0_EDL42044 GST0_EDL42044 GST0_EDL42044 GST0_EAAP72967 GSTR_ABV24478 GST3_EAA45010 GSTT_BAB39498 GST3_EAA93767 GST2_P28342 000 Conservation 0% AtuGST4	LK-RLDE -K-AANRYFP VKNKLKSKFV 	200 I VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT VYEARLT VTEARLT FINFL NMEEILE -FETLLSQNQ YLEKMGK KLNVIAKENN ELDVTLQLLE TLEAELGD ALEKLK	-DREWL -GQKCV -GQKCV -GQKCV -GQKYA -SSSYL -KCKYL KSKLRETT DGPFL GKFSWF -YQNTTFF -GGKTFI -GS.YL GHLVLGKPTW DKFLONKAFL -KQPYI -GH.	220 I EGS FSA VGNKL SRADV CGDHF TVADA AGDSLT AGDSLT AGDKLS AGCK YGAF GGDC IS AGVFA GGG - GGGRL MGDEFN AGKYATGDEV ASKN 230	AD ILMICV HLLET VADIAL -VADLLS -LADLNHVSV GLP -MIDYLVWPW NLLD -MADVVC GILDYLNYLT -VADIAI GLADLFL	240 LRRLESSGI- 	KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN APGCLDAFPL LHCPKEKCPR CQVFE GGFSVEE LKLDLSAYPA FGMDMAEFPL	200 LLAYVE LQSFKA VNKYLA VNKWUD VKAWUD VKAWUS DRMELLAYLL GRPKLATWIE LKAFMC LGEVYVG LME	RGKARPAFKR RTSNIPNIKK RIAQRPNVHS NAKK RLNRLPFYKA GLMERPSVQK GEKWMGPVPP ELNKIDA RFEALEKIAA SMKQDPAVCA RESARPKLKA MVKDRPSIKA KVLDNENVKA RVEAVGEDL RCGRIDSVVK RVCERPAFQK YQHFRMRCQR	215 196 189 201 205 220 214 200 215 192 207 193 206 209 179 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AAA33469 GSTL_BT051648 GSTM_AAC17866 GST0_EDL42044 GSTP_AAP72967 GSTR_ABV24478 GSTS_EAA45010 GSTT_BAB39498 GSTL_AB45928 GSTL_AB45928 GSTL_AB45928 GSTL_AB46572	LK-RLDE -K-AANRYFP VKNKLKSKFV 	Z00 I VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT VYEARLT FSMFL FSMFL FETLLSONQ YLEKMGK KLNVIAKENN ELDVTLQLE TLEAELGD ALEKLLK	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL -KPFL GGPFL GGFSWF -YQNTFF GGKTF1 -GS.YL GHLVLGKPTW DKFLONKAFL -KQPY1 -GH 3000 1 AQLAVFTA	220 I EGS FSA VGNKL SRADV CGDHFT VADA AGDSLT AGDSLT AGDSLT AGDSLT AGDSLT AGDKLT AGDKLT AGCKIS AGKYATGDEV ASKN 230 ALFH 221	AD I L MI CV H L L E T	240 LRRLESSGI- ILAVEES TLSQWAPH VASVSTFEVA GPTVSALRAA TLCLFATPYA VERFHIVLA- LIHEVL FP-VVAYFPR TELMHPVGAG TAWFYSYERC GSILAYIPMM APQIIASITG	LKDYGN KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN APGCLDAFPL LHCPKEKCPR ENFPN CQVFE GGFSVEE GGFSVEE GGFSVE	200 LLAYVE LQSFKA LQDYLA VKAWVE VSAWUE VKAWVE UKAWVE LKAFMC HTPMLRLWIA LSAYVG RPKLATWRQ VAPRLAAWAR VIDYTK LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RIAGRPNVHS RLNRLPFYKA GLMERPSVGK GEKWMGPVPP ELNKIDA RFEALEKIAA SMKQDPAVCA RLSARPKLKA MVKDRPSIKA KVLDNENVKA RVEAAVGEDL RCGRIDSVVK RVCERPAFQK YQHFRMRCQR	215 196 193 189 201 205 220 214 200 215 192 207 193 206 209 179 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AAA33469 GSTK_EDM15501 GSTL_ED51648 GST0_ED142044 GSTD_AAP72967 GSTT_AAP72967 GSTT_AAP72967 GSTT_AAP72967 GSTT_BAB39498 GSTJ_AB45010 GSTT_BAB39498 GSTJ_AB45010 GSTT_AB45010 GSTZ_P283420 Conservation % AtuGST4 GSTA_AAA16572 GSTE_AA242030	LK-RLDE -K-AANRYFP VKNKLKSKFV 	200 I VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT VYEARLT FETLLSQNQ YLEKMGK KLNVIAKENN ELDVTLQLLE TLEAELGD ALEKLLK	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL KSKLRETT DGPFL GKFSWF -GGKTFI -GS.YL GHLVLGKPTW GHLVLGKPTW GHLVLGKPTW -REFF -KQPYI -GH	220 I EGS FSA VGNKL SRADV CGDHF TVADA AGDSLT AGDSLT AGDSLT AGDSLT AGDSLT VGDQ ISFADY AGKNFT TGPH IS TGPH IS AGKYATGDEV ASKN 230 AIFH 221 IKE 203	AD I LMI CV HLLET VADIAL VADIAL LAD GLP 	240 LRRLESSGI- 	KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN APGCLDAFPL LHCPKEKCPR GGFSVEEE GGFSVEEE GGFSVEEE GGMDMAEFPL	200 LLAYVE LQSFKA LQDYLA VNKWYE VSAWVE VKAWVE DRMELLAYLL GRPKLATWIE LKAFMC LKSYVG LMEYVG GRPKLATWRQ VAPRLAAWAR VIDYTK LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RIACPPVHS NAKK RLNRLPFYKA GLMERPSVQK GEKWMGPVPP ELNKIDA RFEALEKIAA SMKQDPAVCA RUSARPKLKA MVKDRPSIKA RVEAVGEDL RCGRIDSVVK YQHFRMRCQR	215 196 193 201 205 220 214 200 215 192 207 193 206 209 179 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AAA33469 GSTL_BT051648 GSTU_BT051648 GST0_EDL42044 GSTD_AAP72967 GSTR_ABV24478 GSTS_EAA45010 GSTT_BAB39498 GSTZ_P28342 GSTZ_P28342 GSTZ_P28342 GSTZ_P28342 GSTZ_P28342 00% Conservation 0%	LK-RLDE -K-AANRYFP VKNKLKSKFV VKNKLKSKFV 	ZOO - LSAWLG VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT LLKISTELV FLENALGKFD -FSMFL -FETLLSQNQ VEKMGK KLNVIAKENN ELDVTLQLLE LEAELGD ALEKLLK	-DREWL -GQKCV -GQEYA -SSSYL -KCKYL KSKLRETT DGFSWF GGFSWF -YQNTTFF GGKTFI -GS.YL GHLVLGKPTW DKFLQNKAFL -KQPYI -GH. - GALAVFT AQLAVFT AQLAVFT C	220 I EGS FSA VGNKL SRADV CGDHFTVADA AGDSLT AGDSLT AGDFLS AGDFLS AGCKIGAF LG-QLS AGEKLTFVDF GGDCIS AGEKLTFVDF GGDCIS AGKYFT AGKYATGDEV ASKN 230 AIFH 221 -IKE 200	AD I L MI C V HLLET VADIAL -LADLLS CADLNHVSV GLP -LVDIAYIPF TYDI MIDYLVWPW NLLDL -MADVVC GILDYLNYLT -LADLVAI GFVDVALVPF -VADIAI GLADLFL	240 LRRLESSGI- 	LKDYGN KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN LADCVN APGCLDAFPL LHCPKEKCPR E GGFSVEE LKLDLSAYPA FGMDMAEFPL	200 LLAYVE LQSFKA LQDYLE VKAWZE VSAWLD VKAWZE USAYUE GRPKLATWIE LKAFMC HTPMLRLWIA LSAYVG GRPKLATWRE VAPRLAAWAR VIDYTK LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RIAGRPNVHS MAKK RLNRLPFYKA GLMERPSVGK GEKWMGPVPP ELNKIDA RFEALEKIAA SMKQDPAVCA RLSARPKLKA MVKDPPSIKA KVLDNENVKA RVEAAVGEDL RCGRIDSVVK RVCERPAFQK	215 196 193 189 201 205 220 214 200 215 192 207 193 206 209 179 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_EAA33469 GSTK_EDW15501 GSTL_BT051648 GSTM_AAC17866 GST0_EDL42044 GSTP_AAP72967 GSTR_AB424478 GSTS_EA45010 GSTT_BAB34948 GSTJ_ABF9228 GSTJ_ABF9228 GSTJ_AB4572 GSTZ_P28342 Conservation % AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478	LK-RLDE -K-AANRYFP VKNKLKSKFV 	200 I VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT VYEARLT LLNKISTELV FLENALGKFD -FSMFL -FETLLSQNQ YLEKMGK KLNVIAKENN ELDVTLQLLE TLEAELGD ALEKLLK A.	-DREWL -GQKCV -GQKCV -GQKYA -SSSYL -KCKYL -KCKYL -FSWF GGFSWF GGFSWF GGKTFI -GS.YI -GHLVLGKPTW DKFLQNKAFL -KQPY1 -GH 3000 1 AQLAVFTA LEEKDIPRLM GCLEFK	220 I EGS FSA VGNKL SRADV CGDHFTVADA AGDSLT AGDSLT AGDSLT AGDSLT AGDSLT AGDVFA GGC 1S VGDQ 1SFADY AGKNFT GGC - GGGRL MGDEFN AGKYATGDEV ASKN 230 AIFH 221 -IFKE 203 KFFE 203 KFFE 203	AD I L MI CV HL LET Y - VAD I AL - LAD LLS - LAD L NH VS V GLP - LVD I AY I PF LTYD I - MI DY LVWPW NLLDL - MAD VV C GI LD YLNYLT - LAD LVAI GFVD VAL VPF - VAD I AI GLAD LFL	240 LRRLESSGI- 	LKDYGN KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN APGCLDAFPL LHCPKEKCPR ENFPN CQVFE GGFSVE FGMDMAEFPL	200 LLAYVE LQSFKA LQDYLA VNKWYE VSAWLD VKAWYE USAWLG VKAWYE LKAFMC LKAFMC LKAYVG CMFLATWIE LKAYVG GRPKLATWIE LCEVVQ GRPKLATWRQ VAPRLAAWAR VIDYTK LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RIAGRPNYHS NAKK RLNRLPFYKA GLMERPSVQK GEKWMGPVPP ELNKIDA RFEALEKIAA MVKDRPSIKA KVLDNENVKA RVEAAVGEDL RCGRIDSVVK RVCERPAFQK YQHFRMRCQR	215 196 193 189 201 215 220 215 192 207 193 206 209 179 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AA33469 GSTK_EDM15501 GSTL_ED51648 GST0_EDL42044 GST0_EDL42044 GST0_AAP72967 GSTC_AAP72967 GSTC_AAP72967 GSTC_AB439498 GSTU_AB799228 GSTX_EAW33767 GSTZ_P28342 000000000000000000000000000000000000	LK-RLDE -K-AANRYFP VKNKLKSKFV 	200 I VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT LLNKISTELV FSMFL NMEEILE FETLLSQNQ YLEKMGK KLNVIAKENN ELDVTLQLLE TLEAELGD ALEKLLK R Y	-DREWL -GHDFL -KQKCV -GQKCV -GQKCV -SSSYL -KCKYL KSKLRETT DGPFL GKFSWF -YQNTTFF -GGKTF1 -GS.YL GHLVLGKPTW DKFLQNKAFL -KQPY1 -GH. - - 	220 I EGS FSA VGNKL SRADV CGDHF TVADA AGDSLT AGDSLT AGDSLT AGDSLT AGDX SA GGC IS AGCK YGAF GGDC IS TGPH IS AGK YATGDEV ASKN 230 AIFH 221 -IKE 209 KLGA 223 KFFE 209 KLGA 224	ADILMICV HLLET VADIAL -VADLLS -LADLNHVSV GLP -MIDYLVWPW NLDL GILDYLVYLT -MADVVC GILDYLNYLT -VADIAI GLADLFL	240 LRRLESSGI- 	KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN APGCLDAFPL LHCPKEKCPR CQVFE GGFSVEE LKLDLSAYPA FGMDMAEFPL	200 LLAYVE LQSFKA LQDYLA VNKWYE VSAWYE VSAWYE VKAWYE CRPKLATWIE LKAFMC LGEVVQ GRPKLATWRQ VAPRLAAWAR VIDYTK LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RIACPPVHS NAKK RLNRLPFYKA GLMERPSVQK GEKWMGPVPP ELNKIDA RFEALEKIAA SMKQDPAVCA RLSARPKLKA MVKDRPSIKA KVLDNENVKA RVEAPVGEDL RCGRIDSVVK RVCERPAFQK YQHFRMRCQR	215 196 193 189 201 220 214 200 215 192 207 193 206 179 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AAA33469 GSTL_BT051648 GSTM_EDL42044 GSTP_AAP72967 GSTR_ABV24478 GSTS_EAA45010 GSTT_BAB39498 GSTZ_P28342 GSTZ_P28342 GSTZ_P28342 GSTZ_P28342 GSTZ_P28342 GSTZ_CAR42930 GSTB_CAR4930 GSTB_CAR42930 GSTB_CAR42930 GSTB_CAR44040 GSTB_CAR44040 GSTB_CAR44040 GSTB_CAR44040 GSTB_CAR44040 GSTB_CAR4400 GSTB_CAR4400 GSTB_CAR4400 GSTB_CAR	LK-RLDE -K-AANRYFP VKNKLKSKFV VKNKLKSKFV SIHRGLK -E-KLKKVLE -G-QLKQ LKVALRQELC -G-QLKQ UKVALRQELC -G-QLKQ VT-ELKLWDG NNEVIPFYLT 	Z00 - LSAWLG- VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT VYEARLT FSMFL FSMFL FETLLSONQ YLEKMGK KLNVIAKENN ELDVTLQLLE TLEAELGD ALEKLLK A V	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL GGPFL GGPFL GGFSWF -YQNTTFF GGKTFI -GS.YL GHLVLGKPTW DKFLQNKAFL -KQPYI -GH. - CLEFK - GCLEFK AFLRSKWT. 	2200 I EGS FSA VGNKL SRADV CGDHFTVADA AGDSLT	ADILMICV HLLET	240 LRRLESSGI- ILAVEES TLSQWAPH VASVSTFVA GPTVSALRAA TLCLFATPVAHV VERFHIVLA- LIHEVL FERLDVYG LIHEVL KTNL- TELMHPVGAG GSILAYIPMM APQIIASITG	LKDYGN KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN LACVN APGCLDAFPL LHCPKEKCPR 	200 LLAYVE LQSFKA LQDYLA VKAWVE VSAWUE VSAWUE VKAWVE LKAFMC HTPMLRLWIA LSAYVG GRPKLATWRQ GRPKLATWRQ VIDYTK LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RIAGRPNVHS AKK RLNRLPFYKA GLMERPSVQK GEKWMGPVPP ELNKIA RFEALEKIAA SMKQDPAVCA RVEARYELKA MVKDPPSIKA KVLDNENVKA RVEAAVGEDL RCGRIDSVVK RVCERPAFQK	215 196 193 189 201 202 214 200 215 192 207 193 206 209 2179 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AA33469 GSTK_EDM15501 GSTL_ED0142044 GSTD_ED142044 GSTD_AAP72967 GSTD_AAP72967 GSTZ_P28342 GSTL_AB45010 GSTT_BAB39498 GSTU_AB799228 GSTL_P28342 Conservation % AtuGST4 GSTZ_P28342 Conservation % AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTD_EDW42478 GSTD_EDW42478 GSTD_EDW42478 GSTD_EDW42478	LK-RLDE -K-AANRYFP VKNKLKSKFV 	200 I VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT LLKISTELV FENALGKFD -FSMFL MEEILE -FETLLSQNQ YLEKMGK KLNVIAKENN ELDVTLQLLE TLEAELGD ALEKLLK A V	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL KSKLRETT DGPFL GKFSWF -YQNTTF GGKTFI -GS.YL OHLVLGKPTW DKFLQNKAFL 	220 I EGS FSA VGNKL SRADV CGDHFTVADA AGDSLT AGDSLT AGDSLT AGDSLT AGDVFA VGDQISFADY AGKNFT AGKYATGDEV AGKYATGDEV ASKN 230 AIFH 221 -IKE 203 KFFE 209 KLGA 223 KFFE 209 KLGA 223 KFFE 209	ADILMICV HLLET VADIAL -LADLLS -LADLNHVSV GLP -MIDYLVWPW NLLDL -MADVVC GILDYLNYLT -LADLVAI GFVDVALVPF -VADIAI GLADLFL	240 LRRLESSGI- 	KPDALAKFPL VALDLTDLSH GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN APGCLDAFPL LHCPKEKCPR GGFSVEEE GGFSVEEE GGFSVEEE GGFSVEEE	200 LLAYVE LQSFKA LQDYLA VNKWYE VSAWUD VKAWVE USAWVE GRPKLATWIE LKAFMC GRPKLATWIE LKAYVE LMEYVE GRPKLATWRQ VAPRLAAWAR VIDYTK LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RIACRPNVHS NAKK RLNRLPFYKA GLMERPSVQK GEKWMGPVPP ELNKIDA RFEALEKIAA MVKDRPSIKA RVSARPKLKA RVEAAVGEDL RCGRIDSVVK RVCERPAFQK YQHFRMRCQR	215 196 193 189 201 205 220 214 200 215 192 207 193 206 209 179 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AAA33469 GSTL_BT051648 GSTM_EDL42044 GSTD_AAP72967 GSTZ_DL42044 GSTD_AAP72967 GSTZ_PA872967 GSTZ_PA8349498 GSTZ_FAA45010 GSTT_BA839498 GSTZ_FAA45010 GSTT_BA839498 GSTZ_FAA45010 GSTZ_P28342 00% Conservation % AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AAA33469 GSTL_EDV55071 GSTF_AAA33469 GSTL_EDV55071	LK-RLDE -K-AANRYFP VKNKLKSKFV VKNKLKSKFV -S.HRGLK -E-KLKKVLE AGMATAQAQH QSSPTFE -G-QLKQ LKVALRQELC -G-QLKQ LKVALRQELC -G-QLKQ UT-ELKWDG NNEVIPFYLT TLAATLA AGREMAEILR -T-ALNEIFQ 	ZOO - LSAWLG V FEK VLKDH- Y IND VLS FLNTFLE LLESFLH VYEARLT LLKISTELV FLENALGKFD - FSMFL - FETLLSQNQ V LEKMGK KLNVIAKENN ELDVTLQLLE R ALEKLLK ALEKLLK	-DREWL -GQKCV -GQEYA -SSSYL -KCKYL KSKLRETT DGPFL GKFSWF -YQNTTFF GGKTFI -GS.YL GHLVLGKPTW DKFLQNKAFL GHLVLGKPTW DKFLQNKAFL -KQPYI -GH. 300 AQLAVFTA AQLAVFTA AFLRSKWT 	220 I EGS FSA VGNKL SRADV CGDHFTVADA AGDSLT AGDFLS AGDFLS AGDFLS AGCKYGAF LG-QLS VGDQ ISFADY GGDC IS VGDQ ISFADY AGKNFT AGKYFT AGKYFT AGKYATGDEV AGKATG	AD ILMICV HLLET VADIAL -LADLLS LADLNHVSV GLP -LVDIAYIPF - MIDYLVWPW NLLDL -MADVVC GILDYLNYLT -LADLVAI GFVDVALVPF -VADIAI GLADLFL	240 LRRLESSGI- 	LKDYGN KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN LADCVN APGCLDAFPL LHCPKEKCPR COVFE GGFSVEE LKLDLSAYPA FGMDMAEFPL	200 LLAYVE LQSFKA LQDYLE VKAWYE VSAWS DRMELLAYLL GRPKLATWIE LKAYVG GRPKLATWIE LQEVYQ GRPKLATWAR VIDYTK LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RIACRPNVHS AKK RLNRLPFYKA GLMERPSVGK GEKWMGPVPP ELNKIDA RFEALEKIAA SMKQDPAVCA RLSARPKLKA MVKDPPSIKA KVLDNENVKA RVEAAVGEDL RCGRIDSVVK RVCERPAFQK	215 196 193 201 205 220 215 207 193 206 209 179 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDV42478 GSTE_EDV55071 GSTF_AAA33469 GSTL_BT051648 GSTM_EDM15501 GSTL_BT051648 GST0_ED142044 GST9_AAP72967 GST8_AA45010 GST7_BAB34948 GST2_P28342 GST2_P28342 GST2_P28342 GST2_P28342 GST2_P28342 GST2_P28342 GST2_CAR42930 GST0_EDV42478 GST8_CAR42930 GST6_ED455071 GST8_CAR42930 GST6_ED455071 GST8_CAR42930 GST6_ED455071 GST8_CAR42930 GST6_ED455071 GST8_AAA33469 GST4_ED4516488 GST4_CAR42930 GST6_ED4516488 GST4_CAR42930 GST6_ED4516484 GST4_CAR42930 GST6_ED4516484 GST4_CAC17866	LK-RLDE -K-AANRYFP VKNKLKSKFV 	200 I VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT VYEARLT ILNKISTELV FLENALGKFD -FSMFL MMEEILE -FETLLSQNQ YLEKMGK KLNVIAKENN ELDVTLQLLE TLEAELGD ALEKLLK A. V- VUL P	-DREWL -GQKCV -GQEYA -SSSYL -KCKYL KSKLRETT DGPFL GGKTFI -GS.YL GHLVLGKPTW DKFLONKAFL 	220 I EGS FSA VGNKL SRADV CGDHFTVADA AGDSLT AGDSLT AGDSLT AGDSLT AGDVFA GGCCIS VGDQISFADY AGKNFT AGVYFA GGG-GGGRL MGDEFN AGKYATGDEV ASKN 230 AIFH 221  IFE 203 KFFE 203 KFFE 203 KFPSA 214  CSA COMPARIANCE	ADILMICV HLLET YLF LADLLS -LADLNHVSV GLP -LVDIAYIPF LTYDI -MADVVC GILDYLNYLT -LADLVAI GFVDVALVPF -VADIAI GLADLFL	240 LRRLESSGI- 	KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN APGCLDAFPL LHCPKEKCPR ENFPN CQVFE GGFSVEE FGMDMAEFPL	200 LLAYVE LQSFKA LQDYLA VKWYE VSAWUE VKAWYE USAWUE GRPKLATWIE LKAFMC LKSAYVG GRPKLATWRQ VAPRLAAWAR VIDYTK LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RIAGRPNYHS RLNRLPFYKA GLMERPSVQK GEKWMGPVPP ELNKIDA RFEALEKIA AVKDRPSIKA KVLDNENVKA RVEAVGEDL RCGRIDSVVK RVCERPAFQK YQHFRMRCQR	215 196 193 201 205 220 215 207 193 206 209 179 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AA33469 GSTK_EDM15501 GSTL_ED51648 GST0_ED142044 GST0_ED142044 GST0_ED142044 GST0_AB799228 GST1_EAA5010 GSTT_EAA5010 GSTT_EAA5010 GSTT_EAA5010 GSTT_AA972967 GST2_P28342 Conservation % AtuGST4 GSTA_AAA16572 GST2_P28342 GSTD_EDW42478 GSTD_EDW42478 GSTD_EDW42478 GSTL_ETD51648 GSTK_EDM15501 GSTK_EDM15501 GSTK_ET7866 GSTM_AAC17866 GSTM_AC17866	LK-RLDE -K-AANRYFP VKNKLKSKFV 	200 I VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT LLKISTELV FENALGKFD -FSTLLSQNQ YLEKMGK KLNVIAKENN ELDVTLQLLE TLEAELGD ALEKLLK R A V - FLNLYFQNNP	-DREWL -GHDFL -KQKCV -GQKCV -GQKCV -GQKCV -SSSYL -KCKYL KSKLRETT DGPFL GKFSWF -GGKTFI -GS.YL OKFLQNKAFL REFF GGH - 	220 I EGS FSA VGNKL SRADV CGDHF TVADA AGDSLT AGDSLT AGDSLT AGDSLT VGDQ ISFADY AGKNFT TGPH IS GGC GGC GGC ADV FFA TGPH IS AGKYATGDEV AGKYATGDEV ASKN 230 AIFH 221 	ADILMICV HLLET VADIAL -LADLLS -LADLNHVSV GLP -MIDYLVWPW NLLDL -MADVVC GILDYLVV-IAI GFVDVALVPF -VADIAI GLADLFL	240 LRRLESSGI- 	KPDALAKFPL VALDLTDLSH GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN APGCLDAFPL LHCPKEKCPR GGFSVEEE GGFSVEEE GGFSVEEE GGFSVEEE	200 LLAYVE LQSFKA VNKWE VSAWUD VKAWWE VKAWWE DRMELLAYLL GRPKLATWIE LKAFMC LMEYVE LMEYVE GRPKLATWRQ VAPRLAAWAR VIDYTK LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RIACRPVYHS RAKK RLNRLPFYKA GLMERPSVQK GEKWMGPVPP ELNKIDA RFEALEKIAA MVKDRPSIKA RVEAPKLKA RVEAPKLKA RVEAPKIKA RVEAPAFQK YQHFRMRCQR	215 196 193 201 205 220 215 192 207 193 206 209 179 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AAA33469 GSTL_BT051648 GSTM_ED015501 GSTL_BT051648 GSTM_AAA16572 GSTR_ABV24478 GSTZ_P28342 GSTZ_P28342 GSTZ_P28342 GSTZ_P28342 GSTZ_P28342 GSTZ_P28342 GSTZ_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTB_CAR42930 GSTB_CAR42930 GSTB_CAR42930 GSTB_CAR42930 GSTB_CAR42930 GSTB_CAR42930 GSTB_CAR42930 GSTB_CAR42930 GSTB_CAR42930 GSTB_CAR42930 GSTF_FAA33469 GSTK_EDM15501 GSTL_BT051648 GSTM_AAC17866 GSTO_EDL42044	LK-RLDE -K-AANRYFP VKNKLKSKFV VKNKLKSKFV 	200 - LSAWLG	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL GKFSWF -YQNTTFF GGFSWF -YQNTTFF GGKTFI -GS.YL GHLVLGKPTW DKFLQNKAFL -KQPYI -GH. - GCLEFK ACLAVFTA EEKDIPRLM - GCLEFK AFLRSKWT 	220 I EGS FSA VGNKL SRADV CGDHFTVADA AGDSLT AGDSLT AGDSLT AGDSLT AGDSLT AGDVIS AGDVIS VGDQISFADV GGDCIS VGDQISFADV AGKNFT AGKYATGDEV AGKNFT AGKYATGDEV ASKN 230 AIFH 221 -IKE 203 KFFE 209 KLGA 223 KPSA 214  L 226  PIIR 248  210	ADILMICV HLLET	240 LRRLESSGI- 	LKDYGN KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN LENFPN CQVFE GGFSVEE LKLDLSAYPA FGMDMAEFPL	200 LLAYVE LQSFKA LQDYLA VKAWYE VSAWUE VKAWYE LKAFMC HTPMLRLWIA LSAYVG GRPKLATWRQ GRPKLATWRQ VIDYTK LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RLACRPNVHS AKK RLNRLPFYKA GLMERPSVQK GEKWMGPVPP ELNKIA RFEALEKIAA SMKQDPAVCA RLSARPKLKA MVKDPAVCA RVEAAVGEDL RVCERPAFQK YQHFRMRCQR	215 196 193 201 205 220 214 207 193 206 209 179 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AA33469 GSTK_EDM15501 GSTL_BT051648 GSTM_AAC17866 GSTO_EDL42044 GSTP_AAP72967 GSTR_ABV24478 GSTA_AAA5100 GSTT_BAB34948 GSTJ_ABF9228 GSTJ_ABF9228 GSTJ_AB4501 GSTZ_P28342 	LK-RLDE -K-AANRYFP VKNKLKSKFV 	200 I VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT VYEARLT LLNKISTELV FLENALGKFD -FSMFL -FETLLSQNQ YLEKMGK KLNVIAKENN ELDVTLQLLE TLEAELGD ALEKLLK R IVDL- P FLNLYFQNNP P	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL KSKLRETT DGPFL GKFSWF GGKTFI -GSYL -YQNTFF GGKTFI -GSYL OKFLQNKAFL -KQPYI -GH GCLEFK AFLRSKW-T - AFLRSKW-T - - AAL-M FKNRFLPK INNKMAQWGN CAFDFG-LCN	220 I EGS FSA VGNKL SRADV CGDHFTVADA AGDSLT AGDSLT AGDSLT AGDSLT AGDVFA GGC 1S VGDQ ISFADY AGKNFT GGC -GGGRL MGDEFN AGKYATGDEV ASKN 230 AIFH 221 -IFK 203 KFFA 203 KFFA 203 KFFA 203 KFFA 203 KFSA 214 L 226 PILR 248 210 NL 226	ADILMICV HLLET VADIAL -LADLLS -LADLNHVSV GLP -LVDIAYIPF LTYDI -MIDYLVWPW NLLDL -MADVVC GILDYLNYLT -LADLVAI GFVDVALVPF -VADIAI GLADLFL	240 LRRLESSGI- 	LKDYGN KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN APGCLDAFPL LHCPKEKCPR ENFPN CQVFE GGFSVE FGMDMAEFPL	200 LLAYVE LQSFKA LQDYLA VNKWVE VSAWUD VKAWVE USAWVE GRPKLATWIE LKAFMC LMEYVE LQEVVQ GRPKLATWRQ VAPRLAAWAR VIDYTK LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RIAGRPVYHS RLNRLPFYKA GLMERPSVGK GEKWMGPVPP ELNKIDA RFEALEKIAA MVKDRPSIKA KVLDNENVKA RVEAAVGEDL RCGRIDSVVK RVCERPAFQK YQHFRMRCQR	215 196 193 189 201 205 220 214 200 215 192 207 193 206 209 179 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AAA33469 GSTK_EDM15501 GSTL_ET051648 GST0_EDL42044 GST0_EDL42044 GST0_EAA72267 GSTR_ABV24478 GST2_P28342 1000 Conservation % Conservation Conservation % AtuGST4 GSTA_CAR46512 GSTB_CAR42930 GSTD_EDW42478 GSTA_CAR46501 GSTK_EDM15501 GSTK_EDM15501 GSTK_EDM15501 GSTK_EDM15501 GSTK_EDM15501 GSTK_EDM5501 GSTK_EDM5501 GSTK_EDM15501 GSTK_EDM15501 GSTK_EDM15501 GSTK_EDM15501 GSTK_EDM15501 GSTK_EDM15501 GSTK_EDM15501 GSTK_EDM15501 GSTK_EDM24478 GSTM_AAC17866 GST0_EDL42044 GSTF_AAP72967 GSTR_ABV24478 GSTS_EAA45010	LK-RLDE -K-AANRYFP VKNKLKSKFV VKNKLKSKFV VKNKLKSKFV 	200 - LSAWLG V FEK VLKDH- Y IND VLS FLNTFLE LLESFLH VYEARLT LLKISTELV FLENALGKFD - FSMFL - FETLLSQNO V ELKMGK KLNVIAKENN ELDVTLQLE ELEAELGD ALEKLLK ALEKLLK - FULLYFQNNP P V V	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL -KFSWF -YQNTTFF GGFSWF -YQNTTFF GGKTFI -GS.YL GHLVLGKPTW DKFLQNKAFL GHLVLGKPTW -KQPYI -GH. - GCLEFK AGLAVFT-A - AFLRSKW-T - - - - - - - - - - - - - - - - - -	220 I EGS FSA VGNKL SRADV CGDHFTVADA AGDSLT AGDFLS AGDFLS AGDFLS AGCKITFVDF GGDCIS VGDQISFADY AGKNFT AGKY FA AGKY FA AGKY AT GDEV AGKNFT AGKY AT GDEV AGKNFT AGKY AT GDEV ASKN 230 AIFH 221 -IKE 209 KIGA 223 KFFE 209 KIGA 223 KFE 209 KIGA 225 FIIR 248 	AD ILMICV HLLET VADIAL -LADLLS LADLNHVSV GLP -LVDIAYIPF LTYDI -MDDYLVWPW NLLDL -MADVVC GILDYLNYLT -LADLVAI GFVDVAUVPF -VADIAI GLADLFL	240 LRRLESSGI- ILAVEES ILSQWAPH VASVSTFVA GPTVSALRAA GPTVSALRAA GPTVSALRAA TLCLFATPYA LIHEVL FERLDVYG LIHEVL FP.VVAYFPR KTNL- GSILAYIPMM APQIIASITG	LKDYGN KPDALAKFPL VALDLTDLSH GFEISKYAN VDIEPVEFPK SVLD-AYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN LADCVN APGCLDAFPL LHCPKEKCPR CQVFE GGFSVEE LKLDLSAYPA FGMDMAEFPL	200 LLAYVE LQSFKA LQDYLE VKAWYE VSAWYE VSAWYE USAYWS DRMELLAYLL GRPKLATWIE LKAYVG GRPKLATWIE LQEVYQ GRPKLATWAR VIDYTK LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RLACRPNVHS AKK RLNRLPFYKA GLMERPSVGK GEKWMGPVPP ELNKIDA SMKQDPAVCA RFEALEKIAA SMKQDPAVCA RVEAPKLKA KVLDNENVKA RVEAAVGEDL RCGRIDSVVK RVCERPAFQK YQHFRMRCQR	215 196 193 201 205 220 214 200 215 192 207 193 206 209 179 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTL_EDV55071 GSTF_AAA33469 GSTL_STP_AAA716572 GSTR_ABV24478 GSTA_ABV15501 GSTD_AB4729677 GSTR_ABV24478 GSTS_EAA45010 GSTT_BAB39498 GSTX_EAW33767 GSTZ_P28342 GSTZ_P28342 GSTZ_P28342 GSTZ_P28342 GSTZ_P28342 GSTZ_P28342 GSTZ_CAR42930 GSTD_EDW42478 GSTE_AAA316572 GSTB_CAR42930 GSTB_CAR42930 GSTB_CAR42930 GSTB_CAR42930 GSTB_CAR42930 GSTF_AAA346572 GSTF_AAA346572 GSTF_AAA346572 GSTF_CAR42930 GSTF_CAR49	LK-RLDE -K-AANRYFP VKNKLKSKFV 	200 - LSAWLG. VFEKVLKDH. YINDVLS FLNTFLE LLESFLH VYEARLT VFERALGKFD. -FSMFL -FETLLSQNQ YLEKMGK KLNVIAKENN ELDVTLQLE TLEAELGD ALEKLLK R Y IVDL P. TLNLYFQNNP P. V AKDFPPADPT	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL GKFSWF -YQNTFF GGKTFI -GS.YL GHLVLGKPTW DKFLONKAFL 	220 I EGS FSA VGNKL SRADV CGDHFTVADA AGDSLT AGDSLT AGDSLT AGDXIS AGCKYGAF LG-QLS VGDQISFADY AGKNFT AGVYFA TGPHIS AGKYATGDEV AGKN 230 AIFH 221 -IKE 203 KFFE 209 KLGA 223 KPSA 214 L 226 Q 237 KPVC 225 PIIR 248 L 226 V 203 AMIR 240	ADILMICV HLLET	240 LRRLESSGI- 	LKDYGN KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN APGCLDAFPL LHCPKEKCPR ENFPN CQVFE GGFSVEE GGFSVEE KLDLSAYPA FGMDMAEFPL	200 LLAYVE LQSFKA LQDYLA VKAWVE VSAWUE VKAWYE LKAFMC HTPMLRWIA LSAYVE LQEVVQ GRPKLATWRQ VAPRLAAWAR VIDYTK LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RIACRPVYHS GLMERPSVQK GELWKIGFYVP ELNKIDA MKQDPAVCA RFEALEKIAA MVKDRPSIKA KVLDNENVKA RVEAAVGEDL RCGRIDSVVK RVCERPAFQK YQHFRMRCQR	2155 196 193 189 201 215 192 207 214 200 215 207 207 211 3206 209 179 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_EAA33469 GSTK_EDM15501 GSTL_BT051648 GST0_EDL42044 GSTP_AAP72967 GST2_P28342 GST2_P28342 GST2_P28342 GST2_P28342 GST2_P28342 Conservation 0% AtuGST4 GST2_P28342 GST2	LK-RLDE -K-AANRYFP VKNKLKSKFV 	200 I VFEKVLKDH- YINDVLS- FLNTFLE LLESFLH VYEARLT LLKSISTELV FSMFL INKISTELV FETALSQNQ YLEKMGK KLNVIAKENN ELDVTLQLLE TLEAELGD ALEKLLK ALEKLLK IVDL FLNLYFQNNP P FURUFPADPT FVGVLKKKYG	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL KSKLRETT DGPFL GKFSWF GGKTFI -GS.YL OHLVLGKPTW DKFLQNKAFL REFF -KQPY1 -GH 	220 I EGS FSA VGNKL SRADV CGDHFTVADA AGDSLT AGDSLT AGDSLT AGDSLT AGDSLT VGDQISFADY AGKNFT AGKYATGDEV AGKNFT AGKYATGDEV AGKNFT AGKYATGDEV ASKN 230 AIFH 221 AIFH 221 AGKYATGDEV ASKN 230 AIFH 221 AIFH 221 ASKN 230 AIFH 221 AIFH 221 ASKN 230 AIFH 221 AIFH 221 AIF	ADILMICV HLLET VADIAL -LADLLS -LADLNHVSV GLP -MIDYLVWPW NLLDL -MADVVC GILDYLNYLT -LADLVAI GFVDVALVPF -VADIAI GLADLFL	240 LRRLESSGI- 	KPDALAKFPL VALDLTDLSH GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN APGCLDAFPL LHCPKEKCPR GGFSVEEE GGFSVEEE GGFSVEEE GGFSVEEE GGDMAEFPL	200 LLAYVE LQSFKA LQDYLA VNKWYE VSAWUD VKAWYE QRPKLATWIE LKAFMC QRPKLATWIE LKAYVE LMEYVE LMEYVE GRPKLATWRQ VAPRLAAWAR VIDYTK LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RIACRPVYHS RAKK RLNRLPFYKA GLMERPSVQK GEKWMGPVPP ELNKIDA RFEALEKIAA MVKDRPSIKA RVEAPKLKA RVEAPKLKA RVEAPKUGEDL RCGRIDSVVK YQHFRMRCQR	215 196 193 189 201 202 214 200 215 192 207 193 206 209 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AA33469 GSTL_BT051648 GSTM_ED015501 GSTL_BT051648 GSTM_AAC17866 GSTD_2D420478 GSTS_EAA45010 GSTT_BAB39498 GSTZ_P28342	LK-RLDE -K-AANRYFP VKNKLKSKFV VKNKLKSKFV 	Z00 - LSAWLG- VFEKVLKDH- YINDVLS- FLNTFLE- LLESFLH- VYEARLT- VFERALGKFD -FSMFL -FETLLSQNQ VLEKMGK KLNVIAKENN ELDVTLQLE TLEAELGD- ALEKLLK ALEKLLK P- FUNLYFQNNP P- V- AKDFPPADPT FVGVLKKKYG	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL GGFSWF -YQNTTFF GGFSWF -YQNTTFF GGKTFI -GS.YL GHLVLGKPTW DKFLQNKAFL -KQPYI -GH. - GCLEFK ACLAVFTA SO GCLEFK AFLRSKWT AALM FKNRFLPK INGNGKQ GQDTLK	220 I EGS FSA VGNKL SRADV CGDHFTVADA AGDSLT AGDSLT AGDSLT AGDSLT AGDSLT AGDVIS AGDVIS VGDQISFADV GGDCIS VGDQISFADV GGDCIS CGGGGGRL MGDEFN AGKYATGDEV AGKNFT AGKYATGDEV ASKN 230 AIFH 221 -IKE 203 KFFE 209 KLGA 223 KFFE 209 KLGA 226 KFFE 209 KLGA 226 KFFE 209 KLGA 226 KFFE 209 KLGA 226 KFFE 209 KLGA 200 KFFE 209 KFFE 209	ADILMICV HLLET	240 LRRLESSGI- ILAVEES 	LKDYGN KPDALAKFPL VALDLTDLSH -GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN LENFPN CQVFE GGFSVEE LKLDLSAYPA FGMDMAEFPL	200 LLAYVE LQSFKA LQDYLE VKAWYE VKAWYE LKAFMC HTPMLRLWIA LSAYVG GRPKLATWRQ GRPKLATWRQ VAPRLAAWAR VIDYTK LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RIAGRPNVHS AKK RLNRLPFYKA GLMERPSVQK GEKWMGPVPP ELNKIDA RFEALEKIAA SMKQDPAVCA RLSARPKLKA MVKDPSIKA KVLDNENVKA RVEAAVGEDL RCGRIDSVVK RVCERPAFQK YQHFRMRCQR	215 196 193 189 201 200 214 200 215 207 193 206 209 179 211
AtuGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AA33469 GSTK_EDM15501 GSTL_BT051648 GSTM_AAC17866 GSTO_EDL42044 GSTP_AAP72967 GSTR_ABV24478 GSTS_EAA45010 GSTT_BAB39498 GSTL_BAB39498 GSTL_BAB39498 GSTL_BAB39498 GSTL_AA33469 GSTL_EDV55071 GSTE_AAA33469 GSTL_EDV55071 GSTE_AAA33469 GSTL_EDV55071 GSTE_AAA33469 GSTL_EDV55071 GSTE_AAA33469 GSTL_EDV55071 GSTE_AAA33469 GSTL_EDV55071 GSTE_AAA33469 GSTL_EDV55071 GSTL_AAA33469 GSTL_EDV55071 GSTL_AAA33469 GSTL_AAA17866 GSTD_EDL42044 GSTP_AAA72867 GSTR_ABV24478 GSTS_EAA45010 GSTL_BAB39498 GSTL_AB99228 GSTL_AB93428 GSTL_AEN33767 GSTL_P28342	LK-RLDE -K-AANRYFP VKNKLKSKFV 	200 I VFEKVLKDH- YINDVLS- FLNTFLE LLESFLH VYEARLT VYEARLT ILNKISTELV FLENALGKFD -FSMFL -FETLLSQNQ YLEKMGK KLNVIAKENN ELDVTLQLE TLEAELGD- ALEKLLK ALEKLLK IVDL- P V- V- AKDFPPADPT FVGVLKKKYG	-DREWL -GHDFL -KQKCV -GQEYA -SSSYL -KCKYL -KSKLRETT- DGPFL GGKTFI -GSYL -YQNTFF GGKTFI -GSYL 	220 I EGS FSA VGNKL SRADV CGDHFTVADA AGDSLT AGDSLT AGDSLT AGDSLT AGDVFA GGC 1SFADY AGKNFT AGVYFA GGG - GGGRL MGDEFN AGVYFA TGPHIS GGG - GGGRL MGDEFN AGKYATGDEV AIFH 221  KPSA 214  C223 KPSA 214  C223 KPSA 214  C225 PIIR 248  C237 KPVC 225 PIIR 248  C237 KPVC 225 PIIR 248  C237 KPVC 225 PIIR 248  C237 KPVC 225 PIIR 248  C237 KPVC 225 C25 C25 C25 C25 C25 C25 C25	ADILMICV HLLET	240 LRRLESSGI- 	LKDYGN KPDALAKFPL VALDLTDLSH GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN APGCLDAFPL LHCPKEKCPR CQVFENFPN CQVFE GGFSVEL FGMDMAEFPL	200 LLAYVE LQSFKA LQDYLA VNKWYE VSAWVE VKAWYE LKAFMC LKAFMC LKAYVG LMEYVE LQEVVQ GRPKLATWRQ VAPRLAAWAR VIDYTK LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RIAGRPNYHS RLNRLPFYKA GLMERPSVQK GEKWMGPVPP ELNKIDA RFEALEKIAA MVKDRPSIKA KVLDNENVKA RVEAAVGEDL RCGRIDSVVK RVCERPAFQK YQHFRMRCQR	215 196 193 189 201 220 214 200 215 207 192 207 192 209 179 211
AluGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTE_EDV55071 GSTF_AAA33469 GSTK_EDM15501 GSTL_ET051648 GST0_EDL42044 GSTD_AAP72967 GSTC_AB45010 GSTT_BAB39498 GSTU_AB45010 GSTT_BAB39498 GSTU_AB45010 GSTZ_P28342 Conservation % AluGST4 GSTA_AAA16572 GSTB_CAR42930 GSTD_EDW42478 GSTD_EDW42478 GSTA_AAA16572 GSTB_CAR42930 GSTL_ET051648 GSTL_BA17866 GST0_EDL42044 GSTF_AAA33469 GSTK_EDM15501 GSTL_BA17866 GST0_EDL42044 GSTD_AA772967 GSTZ_P28342 GSTS_EAA45010 GSTT_BAB39498 GSTL_BA99228 GSTX_EAW33767	LK-RLDE -K-AANRYFP VKNKLKSKFV 	200 I VFEKVLKDH- YINDVLS FLNTFLE LLESFLH VYEARLT LLKISTELV FENALGKFD -FSMFL MEEILE MEEILE FETLLSQNQ YLEKMGK KLNVIAKENN ELDVTLQLLE TLEAELGD ALEKLLK ALEKLLK IVDL FLNLYFQNNP P FLNLYFQNNP P V	-DREWL -GHDFL -KQKCV -GQKCV -GQKCV -GQKCV -SSSYL -KCKYL -KQNTF -GFSWF -YQNTFF -GS.YL -YQNTFF -GKTFI -GS.YL -YQNTFF -GKTFI -GREFF -KQPY1 -GH. - - 	220 I EGS FSA VGNKL SRADV CGDHFTVADA AGDSLT AGDFLS AGDFLS GGDCIS VGDQISFADY GGDCIS VGDQISFADY GGCCIS TGPHIS AGKYATGDEV AGKNFT	ADILMICV HLLET VADIAL -LADLLS -LADLNHVSV GLP -LVDIAYIPF -MIDYLVWPW NLLDL -MADVVC GILDYLNYLT -LADLVAI GFVDVALVPF -VADIAI GLADLFL	240 LRRLESSGI 	KPDALAKFPL VALDLTDLSH GFEISKYAN VDIEPVEFPK SVLDAYPH DGKTYMLFGS -EVFKHDITE DPKCLDEFPN APGCLDAFPL LHCPKEKCPR GGFSVEEE GGFSVEEE GGFSVEEE GGFSVEEE GGFSVEEE	200 LLAYVE LQSFKA VNKWE VSAWUD VKAWWE VKAWWE DRMELLAYLL GRPKLATWIE LKAFMC LMEYVE LMEYVE GRPKLATWRQ VAPRLAAWAR VIDYTK LKSLNDAYLK	RGKARPAFKR RTSNIPNIKK RIAGRPNVHS AKK RLNRLPFYKA GLMERPSVGK GEKWMGPVPP ELNKIDA SMKQDPAVCA RFEALEKIAA SMKQDPAVCA RFEALEKIAA KVLDNENVKA RVEAAVGEDL RCGRIDSVVK RVCERPAFQK YQHFRMRCQR	2155 1963 189 2011 2052 215 1922 207 193 206 209 179 211

**Figure 1. Multiple sequence alignment of** *Atu***GST4 with representative GST sequences.** Alpha, (GSTA, AAA16572); beta, (GSTB, CAR42930); delta, (GSTD, EDW42478); epsilon, (GSTE, EDV55071); phi, (GSTF, AAA33469); kappa, (GSTK, EDM15501); lambda, (GSTL, BT051648); mu, (GSTM, AAC17866); omega, (GSTO, EDL42044); pi, (GSTP, AAP72967); ro, (GSTR, ABV24478); sigma, (GSTS, EAA45010); theta, (GSTT, BAB39498); tau, (GSTU, ABF99228), chi, (GSTX, EAW33767); and zeta, (GSTZ, P28342). NCBI accession numbers are in parentheses. The degree of conservation is shown below the alignments in green. Amino acids are colored according to polarity or charge (red for negative charged, blue for positive charged, black for neutral and green for uncharged polar amino acids). doi:10.1371/journal.pone.0034263.q001

domain (Figure 3) [1,5,9,10]. In total, ten  $\alpha$ -helices (H1–H10), 2 3<sub>10</sub>-helices and 4  $\beta$ -strands ( $\beta$ 1– $\beta$ 4) were located in the structure. The N-terminal domain comprises two large  $\alpha$ -helices (H1,

residues 35–45; H4, residues 85–95), two short  $\alpha$ -helices (H2, residues 56–59; H3, residues 62–65) and a four-stranded mixed  $\beta$ -sheet with a left-handed twist formed by strands  $\beta$ 2 (residues 50–



Figure 2. Hypothetical evolutionary history of *Atu*GST4. Phylogeny tree was constructed using representative members from all known GST classes: GSTA, (AAA16572, DAA16513 EDL26376); GSTB, (CAR42930 EFE52214 EDN73431); GSTD, (EDS36584 CAB03592 EDW42478); GSTE, (EDV55071 AAF64647 EDV36040); GSTF, (ABQ96852 CAI51314 AAA33469), GSTK, (EDL13490 EDM15501 AAS01180); GSTL, (BT051648 AED90518 ACH63212); GSTM, (AAC17866 AAK28508 P46419); GSTO, (AAF71994 CAI17224 EDL42044); GSTP, (AAP72967 EDL32992 AAF01323); GSTR, (CAK10882 ABV24478 ABD67511); GSTS, (EAA45010 CAA86859 AAA92066); GSTTtheta, (XP001089367 BAB39498 DAA20393); GSTU, (ABF99228 AAC05216 BAF27055), GSTX, (EAW33767 BAC07760); and GSTZ, (P57108 AAO61856 AAN39918 P28342) and the AtuGST4 from *A. tumefaciens* C58. NCBI accession numbers are in parentheses. doi:10.1371/journal.pone.0034263.g002

54),  $\beta 1$  (residues 18–22),  $\beta 3$  (residues 75–78) and  $\beta 4$  (residues 81– 84). Pro73 at the beginning of  $\beta 3$  adopts a cis-configuration and creates a characteristic turn essential for GSH binding. A 10residue linker region (residues 96–105) that adopts an extended structure connects the N- terminal domain with the larger Cterminal domain. The latter (residues 105 to 224) has an all- $\alpha$ structure with the  $\alpha$ -helices arranged in a right-handed spiral.  $\alpha$ -Helix H5 exhibits a sharp kink at its centre (Thr121) that splits it into two smaller helices, namely H5a (residues 105–120) and H5b (residues 122–135).  $\alpha$ -Helix H5a is straight and oriented nearly parallel to  $\alpha$ -helix H4, while  $\alpha$ -helix H5b has a bent appearance and projects over the active site located in the N-terminal domain. The C-terminal end of H7 takes a 3<sub>10</sub>-helix conformation (residues 185–190). Helices H8 (residues 197–207) and H9 (194–197) correspond closely to similar regions in most of the other GST classes. H10 (residues 203–212) folds back over the top of the N-terminal domain and against helix H1.

In the C-terminal domain, AtuGSTH1-1 possesses two local structural motifs, an N-capping box and a hydrophobic staple motif at the beginning of  $\alpha$ -helix H6 in the hydrophobic core of the molecule, similar to other cytosolic GSTs [22,23]. Both motifs are located between amino acids 172–177 (Phe-Ser-Ala-Ala-Asp-Ile). The N-capping box (Ser-Ala-Ala-Asp) consists of a reciprocal hydrogen bonding interaction of Ser173 with Asp176, whereas the hydrophobic staple motif consists of a hydrophobic interaction between Phe172 and Ile177. In mammalian GSTs and in beta

Table 1. Steady-state I	kinetic analysis	of AtuGSTH1-1.
-------------------------	------------------	----------------

Substrate	K <sub>m</sub> (mM)	k <sub>cat</sub> (min <sup>−1</sup> ) <sup>a</sup>	$k_{cat}/K_m$ (m $M^{-1}$ ·min <sup>-1</sup> )
GSH	0.3±0.03	31.9±0.05	21.4
CDNB	$1.5 \pm 0.09$		
GSH	0.9±0.07	339.4±6.05	123.9
CuOOH	2.7±0.21		
GSH	$1.1 \pm 0.05$	149.7±5.21	95.3
t-BuOOH	1.6±0.07		
GSH	1.7±0.1	2.4±0.05	0.6
HEDS	4.1±0.12		

 $^{a}k_{cat}$  values were calculated for the substrates CDNB, CuOOH, and HEDS. doi:10.1371/journal.pone.0034263.t001

class bacterial *Pm*GST these structural elements are critical for protein folding, stability, and catalytic function [8,22,23].

**Structural comparison with other proteins.** A Dali search [24] showed an r.m.s.d with other GSTs between 2.3–3.5 Å and a 20–25% sequence identity. The GST-like protein YfcG (PDB id 3gx0) [25] an *E. coli* GST homologue with disulfide-bond reductase activity, was identified as the closest structural neighbor of *Atu*GSTH1-1 (Z = 22.9, r.m.s.d = 1.9 Å, 22% sequence identity). The second structure in the Dali list was that of *Rhodobacter sphaeroides* GST (PDB id 3lsz; Z = 21.7, r.m.s.d = 2.5 Å, 28% sequence identity).

Subunit-subunit interactions. The structure of *Atu*GSTH1-1 contains one molecule in the asymmetric unit. The functional dimer found in GSTs was generated by the symmetry operator -x, y,  $-z+\frac{1}{2}$  of the C222<sub>1</sub> space group (Figure 3B). The interface involves 49 residues from each monomer and the buried surface area is ~1645 Å<sup>2</sup> from each monomer (about 15% of the total solvent accessible area of each monomer), which is within the values found in most other GST families [1,7,9,10]. The main regions involved in subunit interactions are residues 65-72 (part of helix H3), 80-85 (strand \$4), 86-96 (helix H4), 105-128 (part of helix H5), 139-143 and 154-162 from helix H6. Close inspection shows that the formation of the dimer follows the 'lock-and-key' mode that is also found in the phi, alpha, mu and pi classes of GSTs [9,26]. The "lock-and-key" motif plays important functional and structural roles and is generally considered important for dimerization. The "key" is an aromatic residue in one subunit and the "lock" is a cluster of hydrophobic residues from the other interacting subunit. Indeed, the side chain of Phe70 acts as the 'key' that locks into a hydrophobic pocket consisting of Ile122', Leu160', Leu170', Met181', Leu200', and Trp114' from the second subunit. Six hydrogen bonds (three from each subunit) with distances from 2.5 to 3.5 Å contribute further to the stability of the interface: Arg156 NH1-Phe70 O 3.4 Å; Thr121 OG1-Glu85 OE2 2.6 Å; Arg148 NH2-Glu139 OE1 3.0 Å; Arg148 NH1-Glu139 OE2 2.8 Å. Glu85, in particular, is involved in Nb-GSH binding through its OE1 atom whereas its OE2 atom makes a hydrogen bond with Thr121 OG1 (distance 2.6 Å) from subunit B at the subunit interface. This interaction might also induce the kink of helix H5.

**GSH Binding Site (G-site).** A molecule of Nb-GSH was found bound in the active site of *Atu*GSTH1-1 (Figure 3C). The glutathione portion of Nb-GSH is located in a region formed by the beginning of helices H1 and H4 and part of the  $\beta$ -turn between H3 and  $\beta$ 3. The  $\gamma$ -Glu portion makes hydrogen bonds through the oxygen atoms O11 and O12 with Glu85 and Ser86. Table 2. Data collection and refinement statistics.

Data collection	
Space group	C222 <sub>1</sub>
Cell dimensions (Å)	49.4×96.0×88.4
Number of molecules	1
Resolution range (Å)	20.0-1.40 (1.5-1.4)#
Number of measured reflections	247406 (45974)
Unique reflections	41008 (7614)
Completeness (%)	99.4 (99.9)
Mosaicity (°)	0.2
/o(I)	19.9 (4.1)
R <sub>merge</sub> (%)	5.6 (49.7)
R <sub>meas</sub> (%) <sup>&amp;</sup>	6.2 (54.2)
Wilson B-factor (Å <sup>2</sup> )	20.8
Refinement	
Reflections (working/test)	41008 (38919/2089)
R <sub>work</sub> /R <sub>free</sub> (%)	17.2/18.6
Number of protein atoms	1757
Number of waters	313
RMS deviation from ideal geometry	
Bond lengths (Å)	0.008
Bond angles (°)	1.15
Ramachandran plot	
Residues in most favoured regions (%)	92.5
Residues in additional allowed regions (%)	7.0
Residues in generously allowed regions (%)	0.5
Average B factors (Ų)	
Main chain/Side chain	14.9/19.8
Waters	31.6
S-(p-nitrobenzyl)-glutathione	16.2

#Numbers in parenthesis correspond to the highest resolution shell. Redundancy-independent R-value [54]. doi:10.1371/journal.pone.0034263.t002

In addition, O11 makes two indirect contacts with main chain O of Pro74 and side-chain NE2 of Gln68. The N1 atom interacts with the side-chain atoms of Glu84 and also with two residues Thr121 OG1 and Asn120 NE2 from the symmetry-related subunit that forms the functional dimer. The conserved SNAIL/TRAIL motif in the N-terminal domain that is present in most GST classes and contributes polar functional groups to the GSH binding site is absent in *Atu*GSTH1-1 [27,28]. However, a putative SNAIL/TRAIL-like motif (SGAIV) was found at amino acid positions 86–90 (Figure 1). The hydroxyl group of Ser86 makes a hydrogen bond with the  $\gamma$ -Glu portion of GSH. The other residues of the motif are not directly involved in GSH binding.

**Electrophilic Binding Site (H-site).** The H-site in GSTs is characterized by low conservation that reflects its role in substrate specificity. In contrast to other GSTs where the H-site involves C-terminal domain residues, interactions in *Atu*GSTH1-1 are mainly provided by Arg187 and the long turn between strand  $\beta$ 1 and helix H1 (residues 25–33) (Figure 4). Compared to Nb-GSH binding in tau class *Gm*GSTU4-4 [9] the orientation is different with the *Atu*GSTH1-1 4-nitrobezyl group more buried than the *Gm*GSTU4-4/4-nitrobenzyl group which points towards the bulk solvent. This might be caused by the presence of bulky Trp163 in



**Figure 3. Crystal structure of** *Atu***GSTH1-1. A.** Ribbon diagram of *Atu*GSTH1-1 monomer. Assignment of secondary structure elements was carried out by DSSP [55]. The bound Nb-GSH and phosphate ion are shown as sticks coloured according to atom type. **B.** Ribbon diagram of the dimeric *Atu*GSTH1-1 structure. The 2-fold axis is perpendicular to the plane of the page. Subunit A is colored in cyan and subunit B is in lemon yellow. The inhibitor Nb-GSH is represented with sticks and coloured according to atom type. **C.** Stereo view of the mFo-DFc electron density omit map for the bound Nb-GSH contoured at  $3\sigma$ . The figures were produced using the CCP4 molecular graphics program [56].

doi:10.1371/journal.pone.0034263.g003

*Gm*GSTU4-4 whereas in *Atu*GSTH1-1 the structural equivalent position of Trp163 is occupied by Arg187. Further comparison shows that several hydrophobic residues from the C-terminal helix in other GSTs are absent in *Atu*GSTH1-1 as a result of the different position of the C-terminal helix (H10) in *Atu*GSTH1-1

away from the active site. In the case of alpha GSTA1-1 [29], tau *Gm*GSTU4-4 [9], and pi class GSTP1-1 [30], the C-terminal helix is longer and acts as a lid over the substrate binding site, thus creating a more restricted binding site entrance. The absence of such a feature in *Atu*GSTH1-1 may explain the ability of this enzyme to accommodate a diverse range of substrates at the H-site [16]. Salt bridges between helix H1 residues Arg34, Glu43, and Glu44 with helix H10 Arg214 and Arg209 may contribute to the stabilization of H10 position away from the active site.

#### Catalytic mechanism and site-directed mutagenesis

It is widely accepted that GSTs achieve catalysis mainly through the involvement of an active site residue that interacts with and activates the sulfhydryl group of GSH to generate the catalytically active thiolate anion [1,9-11,31,32]. This residue in the alpha, mu, pi, sigma classes is a Tyr. In the delta, epsilon, theta, tau and zeta GSTs, the active site residue is a Ser. In omega, beta and lambda is instead a catalytically essential Cys, which is involved in forming a mixed disulfide with GSH. Analysis of the structure of AtuGSTH1-1 showed the absence of any functional side-chain (Ser, Tyr, Cys) in hydrogen bond distance with the cysteinyl moiety of the bound Nb-GSH (Figure 4). This observation distinguishes AtuGSTH1-1 from all other cytosolic GSTs of known structure and function. However, structural analysis indicated that the side-chains of Phe22, Ser25, Arg34 and Arg187 are oriented towards the ligand binding-site and may be important in substrate binding and/or catalysis. Phe22 and Ser25 are located at the beginning of  $\alpha$ -helix H2 whose structural and functional role has been established in numerous publications [9,10,31]. The guanidium groups of Arg187 and Arg34 are adjacent to the sulfur atom of Nb-GSH.

To investigate the role of Phe22, Ser25, Arg34 and Arg187, these residues were mutated to Ala and the mutant enzymes (Phe22Ala, Ser25Ala, Arg34Ala and Arg187Ala) were expressed in E. coli BL21(DE3), purified as the wild-type enzyme, and subjected to kinetic analysis. The kinetic parameters k<sub>cat</sub> and K<sub>m</sub> toward the two model substrates CDNB and cumene hydroperoxide were determined by steady-state kinetic analysis, and the results are listed in Table 3. The results showed that in the case of CDNB/ GSH system the mutants Phe22Ala and Ser25Ala exhibit moderate differences in K<sub>m</sub> values for CDNB, compared to the wild-type enzyme, indicating that the mutations do not change appreciably the affinity of the H-site for the CDNB. Small differences were also observed for the k<sub>cat</sub> values. It is noteworthy that mutant Ser25Ala and Phe22Ala showed decreased  $\mathrm{K}_{\mathrm{m}}$  values for GSH, suggesting that these residues are involved in GSH binding in the G-site. The mutant enzyme Arg187Ala exhibits larger reduction in catalytic efficiency and shows about 3-fold lower k<sub>cat</sub> value and 5-fold increase K<sub>m</sub> value for CDNB, compared to the wild-type enzyme. These results suggest that Arg187 may contribute significantly either to the rate-limiting step or to the chemistry of the catalytic reaction. The mutation of Arg34 had the most detrimental effect on activity. Indeed, the Arg34Ala mutant was inactive ( $k_{cat}$  approximately 0.01 min<sup>-1</sup>), indicating that Arg34 may represent an important catalytic residue.

The effect of mutations using CuOOH/GSH as substrates appears to be significantly different from that seen in the CDNB/GSH system. Phe22 and Ser25 seem to play an important role in determining the  $K_m$  values for CuOOH since a significant increase was observed (6.4–14.7-fold). Interestingly, both mutants show also significant increase in  $k_{cat}$  values towards CuOOH. Probably, the structural integrity or flexibility of the loop where Phe22 and Ser25 are located has been altered in the mutated form



**Figure 4. Close-up stereo view of the active site. Hydrogen-bonds (<4.0 Å) between Nb-GSH and the enzyme are shown as dashed lines.** W304 and W117 from the proposed electron-sharing network are depicted. The orientation of Nb-GSH is the same as in Figure 3C. The figure was produced using the CCP4 molecular graphics program [56]. doi:10.1371/journal.pone.0034263.g004

of the enzyme. A plot of the crystallographic B-factors along the polypeptide chain, which can give an indication of the relative flexibility of the protein portions, indicates that this region undergo large conformational changes (data not shown). The perturbation of loop's flexibility or the loss of specific interactions may lead to structural perturbation of helix H2 with concomitant effect the alterations in  $K_{\rm m}$  and  $k_{\rm cat}$  values.

The mutant enzyme Arg187Ala displays moderate differences in kinetic constants, compared to the wild-type enzyme. On the other hand, the mutation Arg34 to Ala abolishes enzyme activity  $(k_{cat}$  approximately 0.02 min<sup>-1</sup>) using the CuOOH/GSH substrate system, providing additional evidence for the catalytic role of Arg34 (Table 3).

The effect of viscosity on the kinetic parameters was measured in order to analyze the rate-limited step of the catalytic reaction. A decrease of  $k_{cat}$  by increasing the medium viscosity should indicate that the rate-limiting step of reaction is related to the product release or to diffusion-controlled structural transitions of the protein [9,33–35]. A plot of the inverse relative rate constant  $k_{cat}^{\circ}$ /

 $k_{cat}$  ( $k_{cat}^{o}$  is determined at viscosity  $\eta^{o}$ ) versus the relative viscosity  $\eta/\eta^{\circ}$  should be linear, with a slope equal to unity when the product release is limited by a strictly diffusional barrier or a slope approaching zero if the catalytic reaction chemistry is ratelimiting. The inverse relative rate constant  $k_{cat}^{o}/k_{cat}$  for AtuGSTH1-1 for the CDNB/GSH substrates system shows linear dependence on the relative viscosity with a slope  $0.151 \pm 0.003$ (Table 4). The observed intermediate value of the slope  $(0 \le \text{slope} \le 1)$  indicates that the rate-limiting step in the enzyme is not dependent on a diffusional barrier (i.e. product release) and other viscosity-dependent motions or conformational changes of the protein contribute to the rate-limiting step of the catalytic reaction. The effect of viscosity was also evaluated using CuOOH. The slope obtained was determined to be equal to  $0.339 \pm 0.008$ supporting the results obtained using CDNB as substrate. The mutants Phe22, Ser25 and Arg187 exhibit kcat-viscosity slopes with slight differences compared to the wild type enzyme (Table 4). This suggests that the mutations may contribute to catalysis through modulation of specific conformational changes in the

Table 3. Kinetic parameters of mutant enzymes for the CDNB/GSH and CuOOH/GSH reactions catalyzed by AtuGSTH1-1.

Enzyme	K <sub>m</sub> , (mM) (GSH)	K <sub>m</sub> (mM) (CDNB)	$\mathbf{k}_{cat}$ (CDNB) (% of the wild-type)
Wild-type	0.30±0.03	1.5±0.09	100
Phe22Ala	0.085±0.002	3.2±0.39	90.5
Ser25Ala	0.14±0.01	1.9±0.18	78.2
Arg34Ala	ND <sup>a</sup>	ND <sup>a</sup>	0.01
Arg187Ala	0.34±0.08	7.1±0.76	33.5
Enzyme	K <sub>m</sub> , (mM) (GSH)	K <sub>m</sub> , (mM) (CuOOH)	k <sub>cat</sub> (CuOOH) (% of the wild-type)
<b>Enzyme</b> Wild-type	K <sub>m</sub> , (mM) (GSH) 0.9±0.07	K <sub>m</sub> , (mM) (CuOOH) 2.7±0.21	k <sub>cat</sub> (CuOOH) (% of the wild-type)
Enzyme Wild-type Phe22Ala	K <sub>m</sub> , (mM) (GSH) 0.9±0.07 2.9±0.56	K <sub>m</sub> , (mM) (CuOOH) 2.7±0.21 39.6±2.81	k <sub>cat</sub> (CuOOH) (% of the wild-type) 100 553.0
<b>Enzyme</b> Wild-type Phe22Ala Ser25Ala	K <sub>m</sub> , (mM) (GSH) 0.9±0.07 2.9±0.56 17.3±1.11	K <sub>m</sub> , (mM) (CuOOH) 2.7±0.21 39.6±2.81 17.4±1.61	k <sub>cat</sub> (CuOOH) (% of the wild-type) 100 553.0 298.1
Enzyme Wild-type Phe22Ala Ser25Ala Arg34Ala	K <sub>m</sub> , (mM) (GSH) 0.9±0.07 2.9±0.56 17.3±1.11 ND <sup>a</sup>	K <sub>mr</sub> (mM) (CuOOH) 2.7±0.21 39.6±2.81 17.4±1.61 ND <sup>a</sup>	k <sub>cat</sub> (CuOOH) (% of the wild-type) 100 553.0 298.1 0.02

<sup>a</sup>ND: Non determined.

doi:10.1371/journal.pone.0034263.t003

**Table 4.** The effect of viscosity on  $k_{cat}$  for the CDNB/GSH and CuOOH/GSH reactions catalyzed by *Atu*GSTH1-1 and its mutants.

Enzyme	Slope (CDNB/GSH)	Slope (CuOOH/GSH)
Wild-type	0.151±0.003	0.339±0.008
Phe22Ala	0.189±0.003	0.263±0.003
Ser25Ala	$0.325 \pm 0.008$	0.318±0.005
Arg187Ala	0.140±0.005	0.271±0.002

The slopes for the wild-type and the mutant enzymes were derived from the linear plot of the relative turnover number ( $k^{o}_{cat}/k_{cat}$ ) as a function of relative viscosity ( $\eta/\eta^{o}$ ) using glycerol as co-solvent. doi:10.1371/journal.pone.0034263.t004

enzyme without excluding the possibility of their involvement in the reaction chemistry (i.e. Arg187).

Recently, a conserved electron-sharing network that assists the glutamyl  $\gamma$ -carboxylate of GSH to act as a catalytic base accepting the proton from the -SH thiol group of GSH, forming an ionized GSH was investigated in GSTs [36]. This electron-sharing network is created by residues that form ionic bridge interactions between the negatively-charged glutamyl carboxylate group of GSH, a positively-charged residue (primarily Arg) and a negatively-charged residue (Glu or Asp) stabilized by hydrogenbonding networks with surrounding residues (Ser, Thr) and/or water-mediated contacts. This network has been suggested to contribute to the "base-assisted deprotonation" model postulated to be essential for the GSH ionization step of the catalytic mechanism [36]. In the AtuGSTH1-1/Nb-GSH complex, the conserved residues Arg34, Glu85, Ser86, Gln68 and Asn120' appear to form the proposed electron-sharing network. Based on Ouantum mechanics/Molecular mechanics (OM/MM) calculations it was recently proposed [37] that the GSH activation by GSTs is accomplished by a water-assisted proton-transfer mechanism that takes into account the suggested roles of the GSH γ-glutamyl carboxylate group and the active-site water molecules. According to this mechanism, a water molecule acting as a bridge is able to transfer the proton from the GSH thiol group to the GSH y-glutamyl carboxylate group. Dourado et al. have resorted to density functional theory and to potential of mean force calculations to determine the GSH activation mechanism of GSTP1-1 and GSTM1-1 isoenzymes [37]. For the GSTP1-1 enzyme, they have demonstrated that a water molecule can assist a proton transfer between the GSH cysteine thiol and the GSH glutamate alpha carboxylate groups. In the case of GSTM1-1 enzyme, two water molecules positioned between the GSH-SH and the N atom of His107, working like a bridge, are able to promote the proton transfer between these two active groups. Arg34 in AtuGSTH1-1 occupies two alternative positions and exhibits high crystallographic temperature factors, indicating significant flexibility. In particular, in one conformation its guanidium group interacts with the  $\gamma$ -glutamyl carboxylate of GSH forming an electrostatic interaction, while in the second conformation with the sulfur atom of Nb-GSH (3.4 Å), the water molecules W117 and W304, and the backbone carbonyl group of bound Nb-GSH. Arg187 interacts with the water molecule W304 and forms a  $\pi$ -cation interaction with the benzyl group of Nb-GSH. Hence, Arg34 and Arg187 appear to work as a bridge that connects the two water molecules 304 and 117 (Figure 4). Wat304 might be a crucial element in the catalytic mechanism. In the structure, Wat304 was found fixed by the guanidium group of Arg187 with a hydrogen bond of 2.7 Å and with Arg34 with one weak hydrogen bond (3.8 Å). The sulfur atom of Nb-GSH is 5.1 Å away from Wat304. The residues Arg187 and Arg34 could, therefore, function as a 'clamp' to grip Wat304 in a position to form a hydrogen bond with the sulfonate group. Based on the above analysis, in the case of AtuGSTH1-1, a putative bridge of a network of water molecules in the region of an electron-sharing network does exist as shown in Figure 4. Consequently, Arg34 may act as a catalytic residue for GSH activation.

In conclusion, in the present study we showed the structural and functional characterization of the Atu3701 protein from A. tumefaciens. Sequence and structural analysis indicated that Atu3701 defines a new GST class. Based on the available GSTs nomenclature and classification system the new class was classified as the Eta class (H) and accordingly the enzyme was named AtuGSTH1-1. Members of this class were found in soil bacteria and more recently in a set of environmental sequences. Thus, this structure most likely represents a larger family, whose size is expected to grow further as the existing sequence databases expand. AtuGSTH1-1 exhibits wide substrate specificity although analysis of the catalytic efficiency  $(k_{cat}/K_m)$  suggests that hydroperoxides may be its 'natural' substrates, indicating that the enzyme may play important role in counteracting oxidative stress conditions. Investigation of the crystal structure of AtuGSTH1-1 in complex with Nb-GSH indicated that although the enzyme adopts the canonical GST fold it lacks the classic catalytic essential residues in GSTs (e.g. Tyr, Ser, Cys). This characteristic distinguishes AtuGSTH1-1 from all other cytosolic GSTs of known structure and function. Site-directed mutagenesis showed that Arg34 may represent the catalytic residue. This residue together with an electron-sharing network and a bridge of water molecules are proposed to form the basis of the catalytic mechanism.

#### **Materials and Methods**

#### Materials

Reduced glutathione, 1-chloro-2,4-dinitrobenzene (CDNB), Nb-GSH and all other enzyme substrates and chemicals were obtained from Sigma-Aldrich, USA. Molecular biology reagents were purchased from Invitrogen, USA.

## Cloning, expression, and purification of *Atu*GSTH1-1 in *E. coli*

Cloning and expression of AtuGSTH1-1 in E. coli BL21(DE3) cells was carried out as described previously [16]. Purification of AtuGSTH1-1 was carried out as following: after expression, E. coli BL21(DE3) cells were harvested by centrifugation at 10,000 g for 10 min (4°C), resuspended in potassium phosphate buffer (50 mM, pH 8.0, 9 ml) containing sodium chloride (0.3 M), sonicated, and centrifuged at 10,000 g for 20 min. The supernatant was collected and was loaded to a column of Ni-NTA adsorbent (1 ml), which was previously equilibrated with potassium phosphate buffer (50 mM, pH 8.0) containing sodium chloride (0.3 M). Non-adsorbed protein was washed off with 10 ml equilibration buffer, followed by 20 ml of potassium phosphate buffer (50 mM, pH 6.2) containing sodium chloride (0.3 M) and glycerol (10%, v/v). Bound AtuGSTH1-1 was eluted with equilibration buffer containing imidazole in gradually increasing concentrations of 5 mM, 20 mM, 0.1 M, 0.2 M and 0.5 M (total volume of 10 ml). Collected fractions (2 ml) were assayed for GST activity and protein (Bradford assay). Fractions with AtuGSTH1-1 activity were pooled and dialysed overnight against appropriate

buffer and was used for kinetics and structural analysis. Protein purity was judged by SDS-PAGE.

#### **Bioinformatic analysis**

Multiple sequence alignment and phylogenetic analysis were carried out as described by Skopelitou et al. (2012) [16].

#### Assay of enzyme activity and protein

Enzyme assays were performed according to Skopelitou et al. [16]. Observed reaction velocities were corrected for spontaneous reaction rates when necessary. All initial velocities were determined in triplicate in buffers equilibrated at constant temperature. Turnover numbers were calculated on the basis of one active site per subunit. One unit of enzyme activity is defined as the amount of enzyme that catalyses the turnover of  $1 \ \mu mol$  of substrate per min. Specific activity is expressed in  $\mu$ mol  $\cdot$  min<sup>-1</sup> per mg of protein. Protein concentration was determined by the Bradford assay using bovine serum albumin (fraction V) as standard. Steadystate kinetic measurements for the wild-type enzyme were performed at 37°C in 0.1 M potassium phosphate buffer, pH 6.5, over 10-fold varied substrate concentrations. Steady-state data were fitted to the Michaelis-Menten equation by nonlinear regression analysis using the GraFit (Erithacus Software Ltd.) computer program.

#### Viscosity dependence of kinetic parameters

The effect of viscosity on kinetic parameters was assayed in 0.1 M potassium phosphate buffer, pH 6.5, containing variable glycerol concentrations. Viscosity values  $(\eta)$  were calculated as described in Wolf et al [38].

#### Site-Directed Mutagenesis

Site-directed mutagenesis was performed according to Ho et al [39]. The pairs of oligonucleotide primers used in the PCR reactions were as follows: for the Ser25Ala mutation, 5'-CGTTTTTGAACGCGCGCCCGATGGCGG-3' and 5'- CC-GCCATCGGGCGCGCGCGTTCAAAAACG-3' for the Phe22Ala mutation, 5'-CGATCACCGTTGCGGAACGCTCTCC-3' and 5'-GGAGAGCGTTCCGCAACGGTGATCG-3', for the Arg34Ala mutation, 5'- GGTCTCGCGGCGGATATGCCG-3' and 5'-CGGCATATCCGCCGCGAGACC-3', for the Arg187Ala mutation, 5'- CGTCTTACGCGCGCTGGAATCG-3' and 5'-CGATTCCAGCGCGCGTAAGACG-3'. All mutations were verified by DNA sequencing. The mutant enzymes were expressed and purified as described for the wild-type enzyme.

#### Crystallization

Prior to crystallization, AtuGSTH1-1 was concentrated to 4.85 mg/ml in buffer Tris-HCl 15 mM, pH 7.0 and mixed with a 100 mM stock solution of S-(p-nitrobenzyl)-glutathione (10 mM final concentration). Crystals were grown with the hanging drop vapor diffusion method. The drops contained 2 µl of the protein solution mixed with 2 µl of a well solution (1.4 M Na/K phosphate, pH 8.3). The drops were equilibrated against 800 µl of well solution at 16°C.

#### Structure determination and refinement

An initial data set to 1.4 Å resolution was collected on station X13 at EMBL-Hamburg c/o DESY from a single crystal soaked

### References

any clear solution as judged by the low Z-scores (below 5) in PHASER [42] and the poor quality of the resultant electron density maps. Initial phases were obtained by Br-SAD from a single AtuGSTH1-1 crystal soaked with 1 M KBr for 45 seconds in cryoprotectant solution. The crystal was subsequently flash-cooled to 100 K in a stream of gaseous N2. A total of 600 diffraction images were collected ( $\lambda = 0.9$  Å) to 2.01 Å resolution on the BW7A beamline at EMBL-Hamburg (c/o DESY) using a rotation angle of 0.5°, exposure time of 1 sec per image, and a MARCCD detector. Data were processed with DENZO and Scalepack [43]. The search for Br atoms was performed with SHELX [44], which identified an anomalous signal of 1.2 up to 2.4 Å resolution and located 8 Br ions. Phasing with AutoSHARP [45] resulted in a phasing power of 1.125 and an initial figure-of-merit of 0.3. Following solvent flattening and density modification, ARP/ wARP [46] was able to build 202 residues in 3 chains out of 227 residues in total in the aminoacid sequence. Refinement was initially carried out with REFMAC [47] and slowly extended to 1.4 Å in small steps of 0.2 Å. At the final stages of refinement, the program PHENIX [48] was employed. No anisotropic B-factor refinement was performed as the drop in  $R_{\rm free}$  was insignificant. The structure was visualized and rebuilt using COOT [49]. MOLPROBITY [50] and PROCHECK [51] were used to validate the structure. Structural superpositions were performed with SSM [52] and analysis of interfaces with PDBePISA [53].

#### Protein Data Bank accession code

The final coordinates and the structure factors have been deposited with the Protein Data Bank under the accession code 2ycd.

#### Supporting Information

Table S1 Aminoacid sequence homology between AtuGST4 and representative GST sequences from **classes:** alpha, (Q08392); beta, (P15214); delta, (B4HHD9); epsilon, (B3NMR7); phi, (P12653); kappa, (P24473); lambda, (B7FHT3); mu, (P21266); omega, (Q8K2Q2); pi, (P09211); ro, (Q0GZP3); sigma, (P46428); theta, (P30711); tau, (Q10CE7), chi, (O8DMB4); and zeta, (P28342). (DOC)

## Author Contributions

Conceived and designed the experiments: NEL ACP. Performed the experiments: KS PD. Analyzed the data: KS PD NEL ACP. Contributed reagents/materials/analysis tools: KS PD NEL ACP. Wrote the paper: NEL ACP.

Armstrong RN (1997) Structure, catalytic mechanism, and evolution of the 1. glutathione transferases. Chem Res Toxicol 10: 2-18.

for a few seconds in crystallization solution supplemented with

20% v/v glycerol as cryoprotectant. The crystal was subsequently placed in a gaseous nitrogen stream and flash-cooled directly at

100 K. A total of 250 images were recorded on a MARCCD

detector using a rotation angle of  $0.5^{\circ}$  and an exposure time of

10 seconds. Data were processed with XDS [40]. Crystals of AtuGSTH1-1 were found to belong to the C2221 space group with unit cell dimensions 49.4×96.0×88.4 Å. Assuming one molecule

in the asymmetric unit, the Matthews coefficient [41] is 2.3  $Å^3/$ 

 $Da^{-1}$ , corresponding to 46.5% solvent content. Attempts to

determine the structure by molecular replacement did not produce

- Hayes JD, McLellan LI (1999) Glutathione and glutathione-dependent enzymes represent a co-ordinately regulated defence against oxidative stress. Free Rad Res 31: 273–300.
- Dixon DP, Skipsey M, Edwards R (2010) Roles for glutathione transferases in plant secondary metabolism. Phytochemistry 71: 338–50.
- Gupta MN, Kapoor M, Majumder A, Singh V (2011) Isozymes, moonlighting proteins and promiscuous enzymes. Current Science 100: 1152–1162.
- Axarli IA, Rigden DJ, Labrou NE (2004) Characterization of the ligandin site of maize glutathione S-transferase I. Biochem J 382: 885–893.
- Honaker MT, Acchione M, Sumida JP, Atkins WM (2011) Ensemble perspective for catalytic promiscuity: calorimetric analysis of the active site conformational landscape of a detoxification enzyme. J Biol Chem 286: 42770–42776.
- Allocati N, Federici L, Massulli M, Di Ilio C (2008) Glutathione transferases in bacteria. FEMS J 276: 58–75.
- Axarli I, Dhavala P, Papageorgiou AC, Labrou NE (2009) Crystallographic and functional characterization of the fluorodifen-inducible glutathione transferase from *Glycine max* reveals an active site topography suited for diphenylether herbicides and a novel L-site. J Mol Biol 385: 984–1002.
- Axarli I, Dhavala P, Papageorgiou AC, Labrou NE (2009) Crystal structure of *Glycine max* glutathione transferase in complex with glutathione: investigation of the mechanism operating by the Tau class glutathione transferases. Biochem J 422: 247–256.
- Norrgard MA, Hellman U, Mannervik B (2011) Cys-X scanning for expansion of active-site residues and modulation of catalytic functions in a glutathione transferase. J Biol Chem 286: 16871–8.
- Federici L, Masulli M, Di Ilio C, Allocati N (2010) Characterization of the hydrophobic substrate-binding site of the bacterial beta class glutathione transferase from *Proteus mirabilis*. Protein Eng Des Sel 23: 743–750.
- Rossjohn J, Polekhina G, Feil SC, Allocati N, Masulli M, et al. (1998) A mixed disulfide bond in bacterial glutathione transferase: functional and evolutionary implications. Structure 15: 721–734.
- Allocati N, Federici L, Masulli M, Favaloro B, Di Ilio C (2008) Cysteine 10 is critical for the activity of *Ochrobachum anthropi* glutathione transferase and its mutation to alanine causes the preferential binding of glutathione to the H-site. Proteins 71: 16–23.
- Pitzschke A, Hirt H (2010) New insights into an old story: Agrobacteriuminduced tumour formation in plants by plant transformation. EMBO J 29: 1021–1032.
- Skopelitou K, Muleta A, Pavli O, Skarakis GN, Flemetakis E, et al. (2012) Overlapping protective roles for glutathione transferase gene family members in chemical and oxidative stress response in *Agrobacterium tumefaciens*. Funct & Integr Genom 12: 157–172.
- Wood DW, Setubal JC, Kaul R, Monks DE, Kitajima JP, et al. (2001) The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. Science 294: 2317–2323.
- Mannervik B (1996) Evolution of glutathione transferases and related enzymes for the protection of cells against electrophiles. Biochem Soc Trans 24: 878–880.
- Wiktelius E, Stenberg G (2007) Novel class of glutathione transferases from cyanobacteria exhibit high catalytic activities towards naturally occurring isothiocyanates. Biochem J 406: 115–123.
- Hurd TR, Costa NJ, Dahm CC, Beer SM, Brown SE, et al. (2005) Glutathionylation of mitochondrial proteins. Antioxid Redox Signal 7: 999–1010.
- Sies H, Dafré AL, Ji Y, Akerboom TP (1998) Protein S-thiolation and redox regulation of membrane-bound glutathione transferase. Chem Biol Interact 24: 177–185.
- Stenberg G, Dragani B, Cocco R, Mannervik B, Aceto A (2000) A conserved "hydrophobic staple motif" plays a crucial role in the refolding of human glutathione transferase P1-1. J Biol Chem 7: 10421–10428.
- Cocco R, Stenberg G, Dragani B, Rossi Principe D, Paludi D, et al. (2001) The folding and stability of human alpha class glutathione transferase A1-1 depend on distinct roles of a conserved N-capping box and hydrophobic staple motif. J Biol Chem 24: 32177–32183.
- Holm L, Sander C (1997) Dali/FSSP classification of three-dimensional protein folds. Nucleic Acids Res 25: 231–234.
- Wadington MC, Ladner JE, Stourman NV, Harp JM, Armstrong RN (2009) Analysis of the structure and function of YfcG from *Escherichia coli* reveals an efficient and unique disulfide bond reductase. Biochemistry 48: 6559–6561.
- Hegazy UM, Hellman U, Mannervik B (2006) Replacement surgery with unnatural amino acids in the lock-and-key joint of glutathione transferase subunits. Chem Biol 13: 929–936.
- Pemble SE, Wardle AF, Taylor JB (1996) Glutathione S-transferase class Kappa: characterization by the cloning of rat mitochondrial GST and identification of a human homologue. Biochem J 1: 749–754.
- Allocati N, Masulli M, Pietracupa M, Federici L, Di Ilio C (2006) Evolutionarily conserved structural motifs in bacterial GST (glutathione S-transferase) are involved in protein folding and stability. Biochem J 394: 11–17.

- Grahn E, Novotny M, Jakobsson E, Gustafsson A, Grehn L, et al. (2006) New crystal structures of human glutathione transferase A1-1 shed light on glutathione binding and the conformation of the C-terminal helix. Acta Crystallogr D Biol Crystallogr 62: 197–207.
- Micaloni C, Kong GK, Mazzetti AP, Nuccetelli M, Antonini G, et al. (2003) Engineering a new C-terminal tail in the H-site of human glutathione transferase P1-1: structural and functional consequences. J Mol Biol 3: 111–122.
- Ricci G, Caccuri AM, Lo Bello M, Rosato N, Mei G, et al. (1996) Structural flexibility modulates the activity of human glutathione transferase P1-1. Role of helix 2 flexibility in the catalytic mechanism. J Biol Chem 271: 16187–16192.
- Balogh LM, Le Trong I, Kripps KA, Shireman LM, Stenkamp RE, et al. (2010) Substrate specificity combined with stereopromiscuity in glutathione transferase A4-4-dependent metabolism of 4-hydroxynonenal. Biochemistry 23: 1541–1548.
- 33. Axarli I, Georgiadou C, Dhavala P, Papageorgiou AC, Labrou NE (2010) Investigation of the role of conserved residues Ser13, Asn48 and Pro49 in the catalytic mechanism of the tau class glutathione transferase from *Glycine max*. Biochim Biophys Acta 1804: 662–667.
- Johnson WW, Liu S, Ji X, Gilliland GL, Armstrong RN (1993) Tyrosine 115 participates both in chemical and physical steps of the catalytic mechanism of a glutathione S-transferase. J Biol Chem 268: 11500–11511.
- Sampson NS, Knowles JR (1992) Segmental motion in catalysis: investigation of a hydrogen bond critical for loop closure in the reaction of triosephosphate isomerase. Biochemistry 31: 8488–8494.
- Winayanuwattikun P, Ketterman AJ (2005) An electron-sharing network involved in the catalytic mechanism is functionally conserved in different glutathione transferase classes. J Biol Chem 280: 31776–31782.
- Dourado DFA, Fernandes PA, Mannervik B, Ramos MJ (2008) Glutathione transferase: New model for glutathione activation. Chem Eur J 14: 9591–9598.
- Wolf AV, Brown MG, Prentiss PG In Handbook of Chemistry and Physics Weast RC, Astle MJ, Beyer WH, eds. pp D-219–D-269, CRC Press, Inc., Boca Raton, F.L.).
- Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77: 51–59.
- 40. Kabsch W (2010) XDS. Acta Crystallogr D Biol Crystallogr 66: 125-132.
- Matthews BW (1968) Solvent content of protein crystals. J Mol Biol 33: 491–497.
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) Phaser crystallographic software. J Appl Crystallogr 40: 658–674.
- Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. In Methods in Enzymology CW. Carter, J. & Sweet RM, eds. pp 307–326.
- 44. Sheldrick GM (2008) A short history of SHELX. Acta Crystallogr A 64: 112–122.
- Vonrhein C, Blanc E, Roversi P, Bricogne G (2006) Automated Structure Solution With autoSHARP. Methods Mol Biol 364: 215–230.
- Morris RJ, Perrakis A, Lamzin VS (2003) ARP/wARP and automatic interpretation of protein electron density maps. Methods Enzymol 374: 229–244.
- Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 53: 240–255.
- Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, et al. (2002) PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr 58: 1948–1954.
- Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60: 2126–2132.
- Lovell SC, Davis IW, Arendall WB, 3rd, de Bakker PI, Word JM, et al. (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins 50: 437–450.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: A program to check the stereochemical quality of protein structures. J Appl Crystallogr 26: 283–291.
- Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr 60: 2256–2268.
- Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372: 774–797.
- Diederichs K, Karplus PA (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. Nat Struct Biol 4: 269–275.
- Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22: 2577–2637.
- Pterton L, McNicholas S, Krissinel E, Gruber J, Cowtan K, et al. (2004) Developments in the CCP4 molecular-graphics project. Acta Crystallogr D Biol Crystallogr 60: 2288–2294.